16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

CALIBRATION AND 3D GEOMETRY ESTIMATION OF A PAN-TILT-ZOOM CAMERA NETWORK

Imran N. Junejo

IRISA/INRIA, Campus de Beaulieu, 35042 Rennes, France
imran.junejo@inria.fr

ABSTRACT

Due to the recent growing need for increased surveillance, we believe that it is necessary to move from single or stationary cameras to a network of multiple cameras. This is useful for various applications, like surveillance or 3D model building. In this paper, we deal with such a dynamic network of pan-tilt-zoom (PTZ) cameras. We allow the camera to freely vary its internal parameters by rotating and zooming. Moreover, these cameras need not have an overlapping *Field of View* (FoV), in order to maximize their area of coverage. A practical framework is proposed to self-calibrate each camera installed in the network. A calibrated camera act as an angle measuring device, allowing for a 3D reconstruction of the scene. It is also shown that only one automatically computed vanishing point and a line lying on any plane orthogonal to the vertical direction is sufficient to infer the dynamic orientation between all the cameras in the network. Using minimal assumptions, we are able to successfully demonstrate promising results on synthetic as well as on real data.

1. INTRODUCTION

Various areas of interest typically demand surveillance from multiple cameras. This is primarily due to the fact that a single camera, even if allowed to rotate or translate, is not sufficient to cover a large area. Examples include a group of mobile robots deployed to perform tasks in a coordinate manner, a group of users interacting in a mixed-reality environment, and a network of surveillance cameras including both stationary and moving cameras (e.g. on roaming security vehicles). By employing multiple cameras with non-overlapping or disjoint FoV, we would like to maximize the operating area in addition to inferring the network configuration. By network configuration we mean the location and orientation of cameras in the network with respect to each other, also known as the network geometry. This change of camera parameters at each time instance induces a *dynamic network geometry*. We propose a framework for auto-calibration of such a dynamic network.

The addressed scenario involves multiple pan-tilt-zoom (PTZ) cameras monitoring an area of interest. In order to allow camera to know their mutual or relative configuration, each camera in the network needs to be calibrated. This calibration process extracts the intrinsic camera parameters, i.e. the parameters that govern how a scene is captured on to the image plane. In this regard we use the PTZ cameras, for their easy of availability and increased area of coverage. Once these parameters are estimated, we use the extracted vanishing point and use the knowledge of lines in the world to estimate the relative geometry of the camera network. By configuring such a camera network we can (i) direct cameras to follow a particular object, (ii) calibrate cameras so that the observations are more coordinated and perform measurements, (iii) solve the camera hand-over problem, (iv) estimate the 3D model of the scene, (v) generate image/video scene mosaic, or (vi) infer network topology.

1.1 Related work and our approach

The first self-calibration method, originally introduced in computer vision by Faugeras et al. [FLM92] involves the use of the Kruppa equations. The Kruppa equations are two-view constraints that involve only the fundamental matrix and the dual image of the absolute conic. Since then various methods have been proposed, some exploiting special motions, and some generalizing to variable intrinsic parameters [AHR00, FK03, Har97, Men01, Tri97, PKG99, SH99].

Recently, tracking across multiple non-overlapping cameras for various applications such as robotics, mixed reality, and surveillance has attracted considerable amount of attention [TDG05, JSS05, KCM03, ZAKS05]. It is well known that due to perspective projection the measurements made from the images do not represent metric data. Thus the obtained object trajectories and consequently the associated probabilities, used in most of the work cited above, represent projectively distorted data, unless we have a calibrated camera. Also, appearance based features exhibit undesirable results under varying lighting conditions. On the other hand, inter-camera relationships can not be correctly established unless dynamic positions and orientations between cameras are known at any point in time.

The most related work to ours is that of Jaynes [Jay04]. Assuming a common ground plane for all cameras, relative rotation of each camera to the ground plane is computed independently. The motion trajectories of objects tracked in each camera are then reprojected on to a plane in front of the camera frame in order to compute corresponding unwarped trajectories. Camera-to-ground-plane rotation and plane-to-plane transform computed from the matched trajectories is then used to compute relative transform between a pair of cameras. This method assumes that all cameras are calibrated, requires motion trajectories on objects, and each camera is considered to be stationary looking at a common ground plane.

We present a more general solution for registering a network of disjoint cameras. *Our key contributions* include a novel solution for self-calibrating cameras undergoing a general pan-tilt motion, and a method to compute the relative orientation between non-overlapping PTZ cameras using only vertical vanishing point. The translation between cameras is assumed to be established either by a set of surveyed points in 3D space or by using GPS. The target is that each calibrated camera should be able to communicate its intrinsic and ex-

trinsic parameters with other cameras in the network. We demonstrate that a (vertical) vanishing point and the knowledge of a line in a plane orthogonal to the vertical direction is sufficient to perform this task. Our framework consists of two steps of self-calibration of individual PTZ cameras, and recovering inter-camera orientations. Accordingly, the paper is divided into the corresponding sections, with also thorough analysis of singular cases at the end.

2. BACKGROUND AND NOTATIONS

As our work builds on the theory of the projective geometry, a brief introduction to the related concepts is a must. For a pinhole camera model used in this paper, a 3D point $\mathbf{M} = [X\ Y\ Z\ 1]^T$ and its corresponding image projection $\mathbf{m} = [u\ v\ 1]^T$ are related via a 3×4 matrix $\mathbf{P}$ by

$$\mathbf{m} \sim \underbrace{\mathbf{K}[\mathbf{r}_1\ \mathbf{r}_2\ \mathbf{r}_3\ \mathbf{t}]}_{\mathbf{P}}\mathbf{M},\ \ \mathbf{K} = \begin{bmatrix} \lambda f & \gamma & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad (1)$$

where $\sim$ indicates equality up to multiplication by a non-zero scale factor, $\mathbf{r}_i$ are the columns of the rotation matrix $\mathbf{R}$, $\mathbf{t}$ is the translation vector, and $\mathbf{K}$ is a nonsingular 3×3 upper triangular matrix known as the camera calibration matrix including five parameters, i.e. the focal length f, the skew γ, the aspect ratio λ and the principal point at $(u_0,\ v_0)$.

3. PTZ CAMERA SELF-CALIBRATION

This section deals with a single PTZ camera, not the network as a whole. The notations i and j represent any two consecutive frames from a single camera. Each camera in the network is allowed to vary its internal parameters by zooming in/out. As argued by [PKG99, AHR00, Zha01], it is safe to assume zero skew for currently available CCD cameras.

The proposed self-calibration process is divided into two parts: (1) estimating the principal point, and (ii) estimating the remaining intrinsic parameters (focal length and the aspect ratio). The method is described below.

3.1 Principal Point Estimation

Let $\mathbf{K}_1$ and $\mathbf{K}_2$ be the camera calibration matrices for a pair of images obtained by a fixed rotating and zooming camera, differing only in the level of zoom (or focal length). Let also $\mathbf{R}_{12}$ denote the relative rotation between the two orientations of the camera. As is well-known, independently of the scene structure, the two images are related by the infinite homography given by [AHR00]

$$\mathbf{H}_{21} \sim \mathbf{K}_2\mathbf{R}_{21}\mathbf{K}_1^{-1}, \qquad (2)$$

If we rearrange this homogeneous equation as follows

$$\mathbf{K}_2^{-1}\mathbf{H}_{21}\mathbf{K}_1 \sim \mathbf{R}_{21}, \qquad (3)$$

and use the relation that $\mathbf{R}\mathbf{R}^T = \mathbf{I}$ and after rearranging we obtain:

$$(\mathbf{K}_2\mathbf{K}_2^T)^{-1} \sim \underbrace{\mathbf{H}_{21}^{-T}(\mathbf{K}_1\mathbf{K}_1^T)^{-1}\mathbf{H}_{21}^{-1}}_{\mathscr{H}}, \qquad (4)$$

where $(\mathbf{K}_i\mathbf{K}_i^T)^{-1} = \omega_i$ is the Image of the Absolute Conic (IAC) for camera i, containing the intrinsic parameter matrix $\mathbf{K}$. Once ω is obtain, $\mathbf{K}$ is obtained by performing Cholesky decomposition [PFTV88].

The infinite homography matrix $\mathbf{H}_{21}$ has very interesting properties which we exploit for performing calibration. We adopt a *stratified* [HZ04] approach, and first estimate the principal point (u_o, v_o). What we propose can be considered as a *cold-start* for the network i.e. at the start of the network when the cameras are initializing, our method can be easily used to estimate (u_o, v_o).

First, we allow the camera to pan. For a panning camera i.e. when only θ_y is non-zero, the eigendecomposition of $\mathbf{H}_{21}^{pan}$ yields three fixed points given by the three eigenvectors one of which corresponds to the real eigenvalue, denoted by $\mathbf{e}_p$. This fixed point $\mathbf{e}_p$ is the vanishing point of the axis of rotation of our panning camera. This can be represented as:

$$\mathbf{e}_p = \begin{bmatrix} a \pm ib \\ v_o \\ 1 \end{bmatrix} \qquad (5)$$

where we are only interested in the second component v_o. Thus, if we have a panning camera and we estimate the infinite homography $\mathbf{H}_{21}^{pan}$ between the two views of such a camera and perform its eigendecomposition, we obtain the y-component (v_o) of the principal point from the eigenvector corresponding to the real eigenvalue.

Similarly, we allow the camera to perform tilt motion i.e. only θ_x is non-zero. The eigenvector corresponding to the real eigenvalue of $\mathbf{H}_{21}^{tilt}$ in this case can be given as:

$$\mathbf{e}_t = \begin{bmatrix} u_o \\ c \pm id \\ 1 \end{bmatrix} \qquad (6)$$

where the first component is the x-component of the principal point. Thus, by a simple application of eigendecomposition to the infinite homography, obtained from a PTZ camera when it is allowed to pan and tilt, yields the principal point of the camera.

3.2 Estimating f and λ

Now that we have estimated u_o and v_o, We allow the camera to perform a general pan-tilt-zoom motion. As expressed by (4), the left hand side of the equation, representing the intrinsic parameters of the camera 2 can be given as:

$$\omega_j \sim \begin{bmatrix} \frac{1}{\tau} & 0 & -\frac{u_o}{\tau} \\ * & \frac{\lambda^2}{\tau} & -\frac{\lambda^2 v_o}{\tau} \\ * & * & 1 \end{bmatrix} \qquad (7)$$

where $*$ represent duplicate symmetric values, $\tau = \left(\lambda^2 {f_2}^2 + \lambda^2 {v_o}^2 + {u_o}^2\right)$, and f_2 is the focal length of camera 2. As mentioned above, we assume the aspect ratio is zero, this provides a linear constraint from (4):

$$\mathscr{H}_{(1,2)} = 0 \qquad (8)$$

i.e. the (1,2) entry of $\mathscr{H}$, as defined in (4), is zero. Similarly, we obtain two more constraints:

$$\frac{\mathscr{H}_{(1,3)}}{\mathscr{H}_{(1,1)}} + u_o = 0 \qquad (9)$$

$$\frac{\mathscr{H}_{(2,3)}}{\mathscr{H}_{(2,2)}} + v_o = 0 \qquad (10)$$

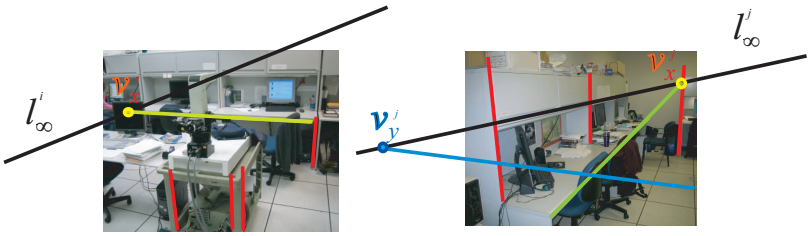


Figure 1: Two views from different non-overlapping directions in a lab: A pair of parallel lines intersect l_∞ at a vanishing point v_x^i in the left image and v_x^j in the right image, respectively. Above, the vanishing line for each view is drawn in black while the lines parallel are drawn in green. The green line in each view intersect the vanishing line at a point. This point is the corresponding vanishing points between the two views. See text for details.

These are three equations for two unknowns (f_2 and λ). Any two of the above two can be used to solve for these unknowns. Once ω_j is obtained, ω_i is easily solved in the same manner by replacing $\mathbf{H}_{21}$ by $\mathbf{H}_{21}^{-1}$ in equation (4). The relative rotation of the camera between the multiple views is obtain by (3).

4. GEOMETRY OF NETWORKED CAMERAS

After self-calibrating each PTZ camera in the network, our goal in this section is to demonstrate that one can establish a common world reference frame to recover absolute camera orientations even with non-overlapping FoVs. The key to establishing a common reference frame is the fact that all cameras share the same plane at infinity and, in our case, also the same vertical vanishing point. In addition, we require a line to be visible in each image in order to completely determine the orientation between the cameras with disjoint FoV. The lines in each image need not to be parallel in the world; orthogonal lines can be used as well (explanation follows in the next subsection).

Since we are dealing now with different camera, not with different views from the same camera, let the relative orientation between any two cameras i and j at any time instance t is denoted by $\mathfrak{R}_{i,j}^t$ and the relative translation by $\mathfrak{T}_{i,j}^t$. Hereafter, we assume that the relative translation $\mathfrak{T}_{i,j}^t$ between two disjoint cameras is recovered by either a set of surveyed points for indoor operations, or by using GPS in outdoor environment.

4.1 Relative rotation using vanishing points

Vertical vanishing point (v_z^i) can be readily obtained from most naturally occurring or man-made scenes. Similarly, people or objects in the FoV of each camera can be used to determine v_z^i [KM05]. For a camera i at any time instance, given a vertical vanishing point v_z^i, the vanishing line l_∞^i can be determined by using the pole-polar relationship [HZ04]: $l_\infty^i = \omega_i v_z^i$.

In addition, we require that a line be visible in each image. This line can lie on any plane that is orthogonal to the vertical direction. For example, checkered tiles on the floor, or brick lining on the wall, or other lines abundant in indoor and outdoor setting, can be used to serve our purpose. Two situations, simplified to two-image cases, can occur with such a configuration, as shown in Figure 1:

1 ***When the visible lines are parallel to each other in world***: In this case, intersection of the imaged line, l_i, with the l_∞^i yields a vanishing point orthogonal to v_z^i:

$$v_x^i \sim l_i \times l_\infty^i \tag{11}$$

where v_x^i, without loss of generality, is orthogonal to v_z^i for an image i.

2 ***When the visible lines are perpendicular to each other in world***: The intersection of the imaged line with the line at infinity yields vanishing point in each image that represent mutually orthogonal direction in the world. In addition to Eq. 11, for the second image (j) we get:

$$v_y^j \sim l_j \times l_\infty^j \tag{12}$$

Note that in both cases l_i is visible in the left image and l_j only in right image (or vice versa). But since l_i and l_j are orthogonal in the world, they intersect l_∞ at mutually orthogonal vanishing points.

Absolute rotation w.r.t. the world reference frame: Two vanishing points v_x^i and v_z^i from each view of a single camera, the rotation of camera i with respect to a common world coordinate system can be computed as:

$$r_3 = \pm\frac{\mathbf{K}_i^{-1}v_z^i}{\|\mathbf{K}_i^{-1}v_z^i\|}, \quad r_1 = \pm\frac{\mathbf{K}_i^{-1}v_x^i}{\|\mathbf{K}_i^{-1}v_x^i\|}, \quad r_2 = \frac{r_3 \times r_1}{\|r_3 \times r_1\|}, \tag{13}$$

where r_1, r_2 and r_3 represent three columns of the rotation matrix. The sign ambiguity can be resolved by the cheirality constraint [HZ04] or by known world information, like the maximum rotation possible for the camera.

Relative orientation is obtained from the obtained absolute orientation for each camera view. Care must be taken in using Eq. (11) and Eq. (12). Based on the obtained vanishing points (v_y or v_x), appropriate equations from Eq. (13) must be selected for determining the absolute orientations.

5. RESULTS

5.1 Synthetic Data

In order to validate the robustness of the proposed self-calibration method, a point cloud of 100 points was generated inside a unit cube to determine point correspondences. The synthetic cameras parameters were chosen as: $f = 1000, \lambda = 1, \gamma = 0, u_o = 320$, and $v_o = 240$. Rotation and translation

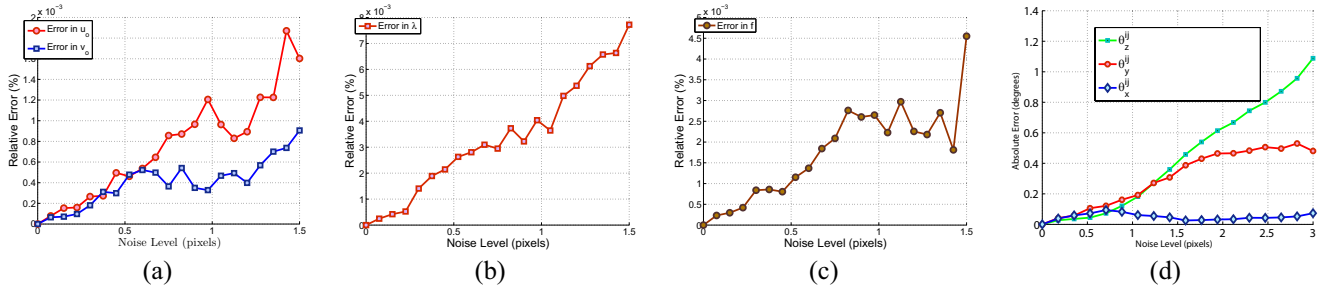


Figure 2: (a)-(c) Performance of the self-calibration method VS. noise level in pixels: (a) shows the relative error in principal point estimation, (b) shows the relative error in estimating the aspect ratio, and (c) shows the relative error in the estimated focal length, respectively. (d) Absolute error in the estimated relative rotation angles.

between views was chosen subjectively to avoid degenerate configurations. Hundred vertical lines of random length and random location are generated to approximate the vertical vanishing points. Similarly, the line (l_i) which is visible in all the images (see Section 4) is also represented by hundred points. Vertical vanishing point is obtained using *SVD* on the vertical lines. Similarly, *SVD* is applied to the points making up l_i to obtain the point of intersection of l_i and l_i^∞. A Gaussian noise with zero mean and standard deviation of $\sigma \leq 1.5$ was added to the data points.

We measure the relative error of estimated f with respect to true f while varying the noise level from 0.01 to 1.5 pixels. For each noise level, we performed 1000 independent trials and the results are shown in Figure 2. For a maximum noise of 1.5 pixels [Zha01], we found that the relative error for f was under 1% (cf. Fig. 2(c)). The absolute error for aspect ratio λ and the relative error for the principal point (u_o, v_o), is also well below 1%. Also notice that with the linear increase of noise to the synthetic data, the error curves also increasing linearly.

Similarly we performed 1000 independent trials for each noise level to estimate the relative orientation between the cameras, the results are shown in Fig. 2(d). The absolute error is found to be less than 1.2° for the maximum noise of 3 pixel in our tests using the method described in Section 4.1. Also, the errors curves increase linearly.

5.2 Real Data

In order to obtain ground truth for relative camera rotations, we employ a SONY® SNC-RZ30N PTZ cameras. The purpose of this demonstration is to verify the accuracy and applicability of the proposed method. Four of the test cases are shown in Figure 3. The ground-truth relative rotation angles are compared to the obtained relative rotation angles. Two PTZ cameras are used for this sequence. The lines which are parallel in the world are drawn in green, while the lines used for the vertical vanishing point are drawn in red. After self-calibrating each rotating camera, as described above, the angles are estimated as described in Section 4. The estimated rotation angles are shown below the figure. The results obtained are very encouraging and close to the ground truth.

Errors and uncertainties in both self-calibration and inter-camera orientation can be attributed to many factors. Main source of error in currently available PTZ cameras is the radial distortion, as visible in the test images. Another important factor is the inherent error present in localizing pixels for determining vanishing points.

6. CONCLUSION

There is a growing trend of installing a number of networked PTZ cameras on a site. In order to better utilize such a camera network, we propose a novel solution to self-calibrate any PTZ camera existing in the network by using the estimated infinite homography obtained from different views taken from the camera. We estimate 4 camera parameters while also estimating the rotation between each view of the camera. Moreover, we propose a method to determine the 3D orientation between all the cameras in the network. We do this by requiring only the vertical vanishing point. We show results on both synthetic data and real data as well and compare it with the ground truth. The error in the estimated parameters is shown to be very low, indicating the applicability of the proposed method.

REFERENCES

[AHR00] AGAPITO L. D., HAYMAN E., REID I.: Self-calibration of rotating and zooming cameras. *Int. J. Comput. Vision 22*, 11 (2000), 1330–1334.

[FK03] FRAHM J., KOCH R.: Camera calibration with known rotation. In *Proc. IEEE ICCV* (2003), pp. 1418–1425.

[FLM92] FAUGERAS O., LUONG T., MAYBANK S.: Camera self-calibration: theory and experiments. In *Proc. of ECCV* (1992), pp. 321–334.

[Har97] HARTLEY R. I.: Self-calibration of stationary cameras. *Int. J. Comput. Vision 22*, 1 (1997), 5–23.

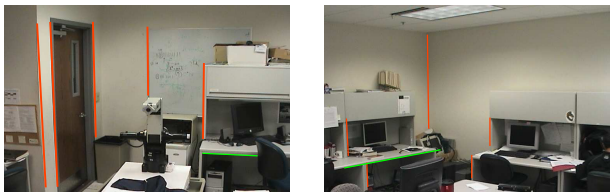
[HZ04] HARTLEY R. I., ZISSERMAN A.: *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, 2004.

[Jay04] JAYNES C. O.: Multi-view calibration from planar motion trajectories. *Image Vision Computing 22*, 7 (2004), 535–550.

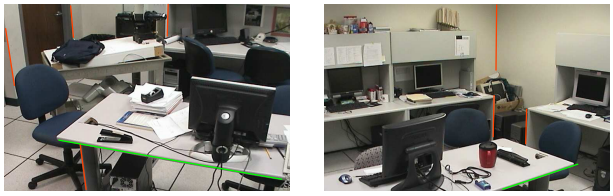
[JSS05] JAVED O., SHAFIQUE K., SHAH M.: Appearance modeling for tracking in multiple non-overlapping cameras. In *IEEE CVPR* (2005).

[KCM03] KANG J., COHEN I., MEDIONI G.: Continuous tracking within and across camera streams. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (2003).

[KM05] KRAHNSTOEVER N., MENDONCA P. R. S.: Bayesian autocalibration for surveillance. In *Tenth IEEE International Conference on Computer Vision* (2005).



(a) Ground Truth $(5°, 0°, 65°)$: Calculated $(4.49°, 1.61°, 63.9°)$



(b) Ground Truth $(11°, 0°, 55°)$: Calculated $(13.13°, 0.82°, 52.98°)$

(c) Ground Truth$(0°, 0°, 80°)$: Calculated $(2.2°, 0.98°, 81.65°)$

(d) Ground Truth $(10°, 0°, 45°)$: Calculated $(12.57°, 1.03°, 47.47°)$

Figure 3: Four of the many test sequences taken from a PTZ camera. The ground truth relative rotation angles are compared to the obtained rotation angles. Green line indicates a common lines parallel in real world) while the lines used to compute the vertical vanishing point are drawn in red.

[Men01] MENDONÇA P. R. S.: *Multiview Geometry: Profiles and Self-Calibration*. PhD thesis, University of Cambridge, Cambridge, UK, 2001.

[PFTV88] PRESS W., FLANNERY B., TEUKOLSKY S., VETTERLING W.: *Numerical Recipes in C*. Cambridge University Press, 1988.

[PKG99] POLLEFEYS M., KOCH R., GOOL L. V.: Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. *Int. J. Comput. Vision 32*, 1 (1999), 7–25.

[SH99] SEO Y., HONG K.: About the self-calibration of a rotating and zooming camera: Theory and practice. In *Proc. IEEE ICCV* (1999), pp. 183–189.

[TDG05] TIEU K., DALLEY G., GRIMSON W. E. L.: Inference of non-overlapping camera network topology by measuring statistical dependence. In *International Conference on Computer Vision* (2005).

[Tri97] TRIGGS B.: Autocalibration and the absolute quadric. In *Proc. IEEE CVPR* (1997), pp. 609–614.

[ZAKS05] ZHAO T., AGGARWAL M., KUMAR R., SAWHNEY H.: Real-time wide area multi-camera stereo tracking. In *IEEE Computer Vision and Pattern Recognition (CVPR)* (2005).

[Zha01] ZHANG Z.: A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence 45*, 2 (2001), 107–127.