Towards Auto-calibration of Smart Phones Using Orientation Sensors

Philip Saponaro and Chandra Kambhamettu
Video/Image Modeling and Synthesis (VIMS) Lab,
Dept. of Computer and Information Science, University of Delaware
Newark, Delaware
{saponaro,chandra}@cis.udel.edu

Abstract

In this paper, we address the problem of auto calibration of cameras which can rotate freely and change focal length, and we present an algorithm for finding the intrinsic parameters using only two images. We utilize orientation sensors found on many modern smart phones to help decompose the infinite homography into two equivalent upper triangular matrices based only on the intrinsic parameters. We account for small translations between views by calculating the homography based on correspondences on objects that are far away from the camera. We show results based on both real and synthetic data, and quantify the tolerance of our system to small translations and errors in the orientation sensors. Our results are comparable to other recent auto-calibration work while requiring only two images and being tolerant to some translation.

1. Introduction

Smart phones with sophisticated cameras are now widespread. There are a wide variety of vision applications for smart phones; for example, distributed surveillance networks [2] and augmented reality systems [18]. Many of these applications require the intrinsic parameters of the camera – the focal length, aspect ratio, skew, and principle point. However, due to the fact that many newer smart phones can change focus, zoom, and move freely, traditional off-line calibration techniques are not practical. Therefore, auto-calibration is needed to calibrate the camera on-line.

Most Android phones made after 2010 use multiple sensors, such as accelerometers, gyroscopes, and compasses, to determine the orientation. Accelerometers record linear acceleration and gravity, and thus can give the orientation of the phone. However, the readings from the accelerometer can be quite noisy and change quickly with small hand jittering. Compasses are magnetic field sensors, and thus can be used to get the heading based on the Earth's magnetic field. However, any other magnetic field will interfere with the reading, causing inaccuracies. Gyroscopes sense angular velocity and give very smooth readings, but the readings drift over time, and cannot measure gravity [24]. The fusion of these sensors significantly improves the accuracy and usability of the orientation reading for real applications [27], with up to $0.03°$ accuracy. We developed our algorithms using the Samsung Galaxy S3 running Android 4.0.4 Ice-Cream Sandwich (see Figure 1), which automatically handles the sensor fusion. This allows us to record the orientation of the phone along with an image, and more easily determine the intrinsic parameters of the camera.

Our contributions are as follows. First, we implement an algorithm for finding the intrinsic parameters of a camera in the case of general rotation, zooming, and small translations using only two images. This algorithm is a modification of previous work by Junejo and Foroosh [13]. Secondly, we quantify how translations can affect the final results. We show that if correspondences are found on an object that is 40m away, then small translations, which are normally due to shaking or imperfect rotation about the camera center, are negligible. Finally, we quantify how errors in the orientation measurements affect the final calculated intrinsic parameters.

The remainder of the paper is as follows. In Section 2, we describe related works and give context to our work. In Section 3, we describe our methods for finding the calibration parameters using only two images. In Section 4, we show the results of running our algorithm on real data, and compare the performance to other works. Finally, in Section 5, we conclude the paper and give recommendations for future work.

2. Background and Related Works

Traditional calibration methods [26, 28] achieve high accuracy; however, these methods are offline and require physical calibration patterns to be used in the scene. Thus, they are not practical for use on cameras with dynamically changing intrinsic and extrinsic parameters, but can be used
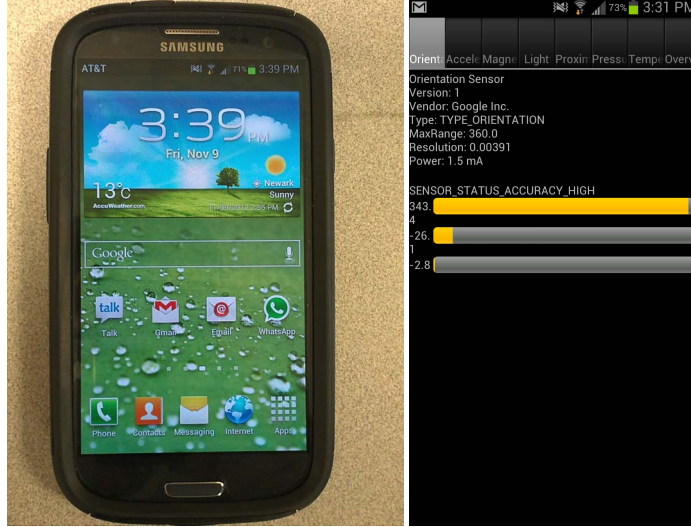
Figure 1: Samsung Galaxy S3 running Android 4.0.4 Ice-Cream Sandwich and Androsens app which shows results of sensor fusion (courtesy of Christian Schneider's Androsens app).

for performance comparison.

Auto-calibration, on the other hand, can deal with dynamically changing intrinsic and extrinsic parameters. Auto-calibration of a camera is calibration done with sequence of images instead of a calibration board, and can be performed online if necessary. Brown describes a method of calibrating a close range camera with bundle adjustment in [3, 4]. The first auto calibration using the Kruppa equations is described in [6], and handles the problem with constant intrinsic parameters of a freely moving camera. Other works focus on the problem when the camera can only translate [17] or when it can only rotate [10]. Finally, some works calibrate the camera with varying intrinsic parameters, specifically the focal length [13, 11, 12, 14, 19, 21].

Our work is related the most to [13], which in turn is built off the work in [5, 1]. Junejo and Foroosh [13] show that with only two images, four independent constraints can be found by decomposing the infinite homography using a series of Givens rotations. In our work, however, we do not need to calculate the Givens rotations, and instead use the orientation reading directly from the Android phone, which saves computation time. Unlike [13], we show how to deal with small translations that are expected from taking images while holding a phone.

3. Methods

3.1. Notation and General Methods

Camera calibration is the process of finding the invertible 3x3 camera matrix K such that

$$y \sim Kx \qquad (1)$$

where x is a 3-dimensional vector representing a 3D point relative to the camera, y is the corresponding 2D image point, and $\sim$ is equality within a scale factor. Specifically,

$$K = \begin{pmatrix} \lambda f & s & u \\ 0 & f & v \\ 0 & 0 & 1 \end{pmatrix} \qquad (2)$$

which contains the 5 intrinsic camera parameters: f is the focal length, λ is the aspect ratio, s is the skew, and u, v are the principle point.

Since most cameras generally use CCD (charged-coupled device) or CMOS (complementary metal-oxide-semiconductor) type sensors, the transformation between imaged planes is 2D to 2D, which is called the plane induced homography [9]. This plane induced homography relates pixels in one image to pixels in the other image, and can be applied to an entire image to warp it. The infinite homography is a special type of plane induced homography which maps features of the plane at infinity of one view to another, and also can be used to calibrate a camera using image sequences, as in [15]. In general, we can write the infinite homography as

$$H_{ij}^{\infty} \sim K_i R_{ij}^{-1} K_j^{-1} = K_i R_{ij}^{T} K_j^{-1}. \qquad (3)$$

Our method is summarized as follows. First take two images and record their orientation values as Euler angles, which are used to construct R. Then calculate the infinite homography by obtaining image correspondences between the images using SIFT. Then, by assuming zero skew, we

can solve a nonlinear homogeneous system for the intrinsics. Finally, these intrinsics are further refined using bundle adjustment [25]. Note that if there is any translation, then the calculated homography will not be the infinite homography, and error will be introduced. We show that by limiting our correspondences to those on objects far away from the camera that this error is negligible (see section 4.1.1).

3.2. Camera Calibration with Two Images – General Rotation and Changing Focal Length

As in [13], we need only two images to calibrate the camera. We rearrange Eq. 3 into an equivalency between two upper triangular matrices. This representation provides four constraints, which, together with the assumption that the skew $s = 0$ and that the principle point is constant, will allow us to solve for all the intrinsic parameters. Thus Eq. 3 becomes

$$H_{12}^{\infty} \sim K_1 R_{12}^T K_2^{-1}. \tag{4}$$

Using language from aviation, yaw $= \alpha$ is the rotation around the z axis, pitch $= \beta$ is the rotation around the y axis, and roll $= \gamma$ is the rotation around the x axis. R_{12} is defined as

$$R_{12} = R_z(\alpha)R_y(\beta)R_x(\gamma)$$
$$R_z(\alpha) = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}$$
$$R_y(\beta) = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix}$$
$$R_z(\gamma) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma & -\sin\gamma \\ 0 & \sin\gamma & \cos\gamma \end{pmatrix}. \tag{5}$$

Multiplying both sides of Eq. (4) on the left by K_1^{-1} and R_{12} gives us

$$R_{12}K_1^{-1}H_{12}^{\infty} \sim K_2^{-1}. \tag{6}$$

We can rewrite this matrix as an upper triangular matrix and equate it to the inverse of the second camera matrix, with the assumption that the skew is zero.

$$R_{12}K_1^{-1}H_{12}^{\infty} =$$
$$\begin{pmatrix} c_1 & c_2 & c_3 \\ 0 & c_4 & c_5 \\ 0 & 0 & c_6 \end{pmatrix} \sim \begin{pmatrix} \frac{1}{\lambda f_2} & 0 & -\frac{u}{\lambda f_2} \\ 0 & \frac{1}{f_2} & -\frac{v}{f_2} \\ 0 & 0 & 1 \end{pmatrix} \tag{7}$$

The only unknowns left in this equality are the intrinsic parameters. Thus, each of the c_i are first order polynomials in terms of only the intrinsic parameters. This equality leads to four nonlinear first-order homogeneous constraints on the intrinsic parameters:

$$c_3 + uc_1 = 0$$
$$c_5 + vc_4 = 0$$
$$c_4 - \lambda c_1 = 0$$
$$c_2 = 0 \tag{8}$$

We assumed the skew is zero and the principle point is constant; therefore, these four equations are enough to solve for the four unknowns f, λ, u, v. Moreover, these equations do not rely on f_2, the focal length in the second image. These equations are first order; however, since they are nonlinear, it is possible to get multiple solutions. We take the solutions which yield positive intrinsic parameters. If the focal length of the second image, f_2, is desired, we simply switch the subscripts 1 and 2 in the above derivation since it is symmetric. The assumption that the principle point is constant is reasonable, since [20, 1] have shown performance is not hindered much by these assumptions. Moreover, bundle adjustment is used to refine the results, which can alleviate these assumptions.

4. Experiments and Results

We performed several experiments to quantify the errors of our method, to compare to the method described in [13], and to quantify how inaccuracies in the measured orientation would affect the final calculated parameters. We ran both synthetic experiments and experiments on real data.

4.1. Experiments with Synthetic Data

To measure the accuracy of our system in ideal circumstances, we created synthetic data. 3d points were generated along a quadric surface, and a synthetic camera with $f = 1320.5$, $\lambda = .9996$, $u = 600$, $v = 600$ was placed at the world origin to project those points into a 2d image. A random rotation on each axis from $[-5°, 5°]$ was applied to the camera, then SIFT and RANSAC was run to compute a homography between the two images. Note that the focal length was held constant in this test. Table 1 shows the results. The average percent error from the ground truth was under 1% in all cases.

We then perform a similiar experiment, but with differing focal lengths. The focal length was changed by a random number from $[-300, 300]$, as well as a random rotation

	f	λ	u	u
Ground Truth	1320.5	.9996	600	600
Mean	1319.9	1.003	597.7	599.5
Std Dev	12.9	.006	10.2	5.61
Avg % Error	.04	.3	.4	.07

Table 1: Results from synthetic tests without changing the focal length.

	f	f_2	λ	u	u
Ground Truth			.9996	600	600
Mean			1.002	597.3	598.9
Std Dev	11.8	13.1	.005	10.7	5.89
Avg % Error	.03	.05	.26	.45	.08

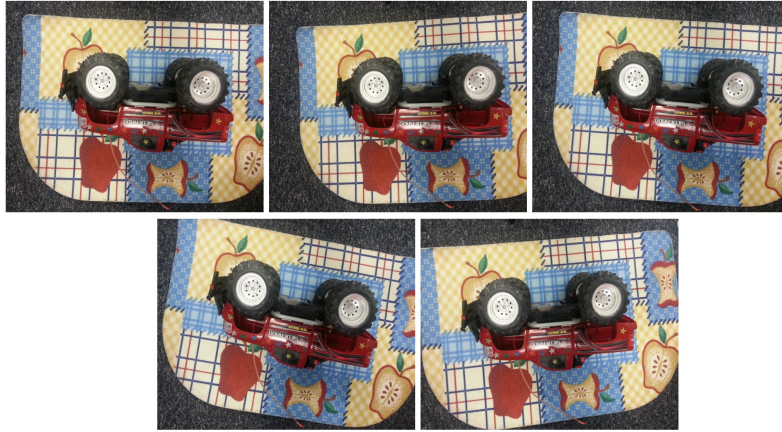Table 2: Results from synthetic tests with changing the focal length.

Figure 2: Images used in Table 3 for calculating intrinsic parameters. The camera was rotated about all 3 axes with minimal translation between pairs of images.

on each axis from $[-5°, 5°]$ was applied. Those results are presented in Table 2.

4.1.1 The Effect of Translation

Translations can be caused by the hand shaking while taking pictures, or by imperfect rotation about the camera center. According to Hartley and Zisserman [8], image points are related by the equation

$$x_2 = K_2 R_{21}^T K_1^{-1} x_1 + \frac{K_2 t}{z} = H_{21}^\infty x + \frac{K_2 t}{z} \quad (9)$$

where x_1 and x_2 are the image correspondences, t is the translation, and z is the depth measured from the first camera. Thus, if there is no translation between the two cameras, the homography we calculate with SIFT is the infinite homography. Furthermore, if we solve this equation using correspondences on objects that lie far from the camera (i.e. $z \to \infty$), the translation term can be negligible.

To test this out, we generated synthetic data. The synthetic camera had the same parameters as in our first experiment: $f = 1320.5$, $\lambda = .9996$, $u = 600$, $v = 600$. An image was created by placing a quadric surface 1m from the camera and applying the calibration matrix to the points on the surface. The camera was rotated and another image was generated. Then, in a loop, without changing the rotation, the camera was translated in small increments. A plot of the relative percent error vs. the translation is given in Figure 3. This experiment was repeated by placing the surface at 20m and 40m from the camera. In all three graphs, the error curves for the parameters are similar; however, note that the x-axis scale changes between the graphs, which shows that as the object moves further away, the error decreases. In fact, at 40m, the camera can move up to 8cm before there is an error of 7%.

	f	f_2	λ	u	v
Mean	578.8628	496.8547	0.9875	325.6590	257.6474
Std Dev	128.7396	66.9009	0.04945	7.5077	51.276
Std Dev [13]	120.5	63.3	0.49	14.29	17.59
Ground Truth	526.8950	526.8950	0.9985	320.5490	255.5600
Percent Error	9.86	5.7	1.1	1.59	0.8

Table 3: Results from calculating intrinsic parameters over all combinations of images.
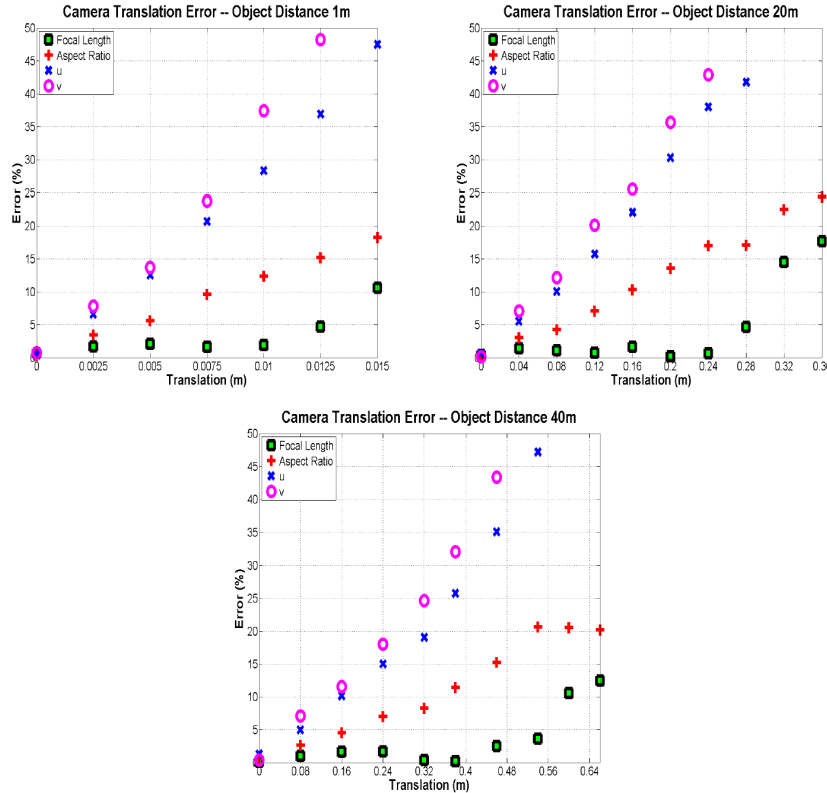

Figure 3: Effect of small translations on the relative percent error.

4.2. Experiments with Real Data

All real data, i.e. orientations readings and images, were collected with the Samsung Galaxy S3 running Android 4.0.4 (Ice-Cream Sandwich) at a resolution of 640x480. The images and corresponding orientations were sent to a desktop computer running 64-bit Windows 7, where the intrinsic parameters were calculated using Matlab 2011. Image feature correspondences were obtained using SIFT [16] and RANSAC [7] for outlier removal. For each pair of images, 150 correspondences were obtained.

We measure the performance of the real data in two ways. The first way to measure the performance is by calibrating for the parameters from a planar checkerboard offline using Callab and Calde from the Institute of Robotics

and Mechatronics in the German Aerospace Center [23, 22]. We took 15 images of the calibration board at different orientations while not changing the zoom of the camera. Using this method, we obtained sub-pixel RMS (Root Mean Square) error, and thus, we use this as the ground truth to compare to.

The second way we measure performance is similar to [13]; we take a sequence of images and calculate the intrinsic parameters using all possible combinations between the images. Then we characterize the uncertainty by the standard deviation over all the calculated parameters.

The results of these methods are compared to the results from [13], and are displayed in Table 3. As can be seen in the table, our results are comparable in error. The relatively high standard deviation for the v parameter was caused by

24

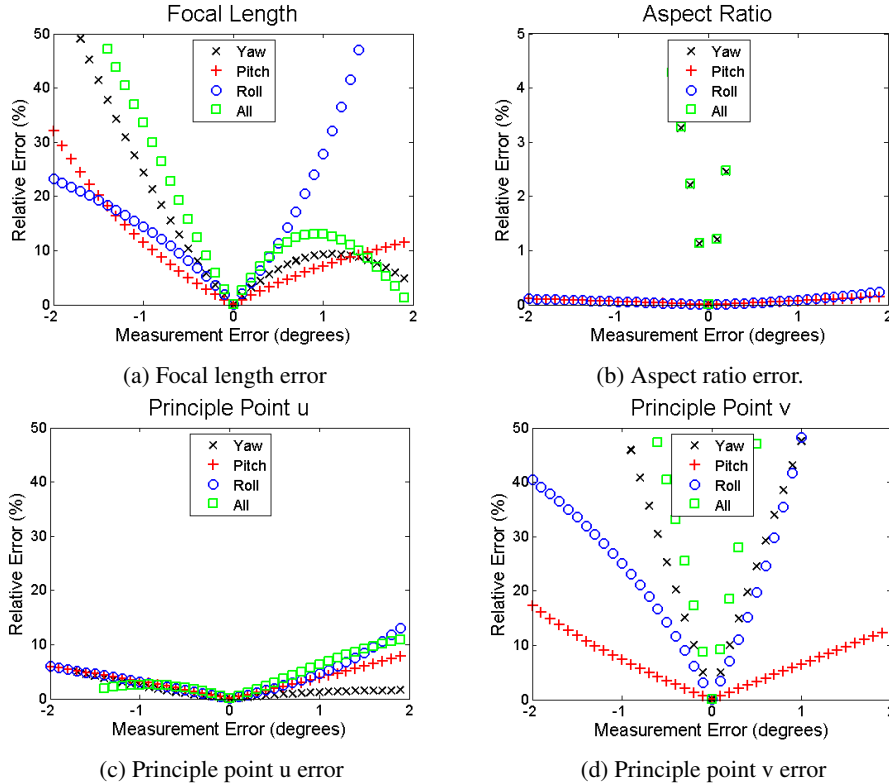|	
(a) Focal length error	(b) Aspect ratio error.
(c) Principle point u error	(d) Principle point v error

Figure 4: Results for image pair 1,2 with synthetic error added

a single outlier. Moreover, the distance to the object was around 1m, which, according to Section 4.1.1, means the results are sensitive to translations.

4.3. Measuring the Sensitivity to Orientation Sensor Errors

We test how tolerant our system is to orientation measurement errors. To do this, we take a single image pair and calculate the intrinsic parameters. Then, we introduce synthetic error to the recorded orientations and recalculate the intrinsic parameters. The error we introduce is from $-2°$ to $2°$ in steps of $0.001°$. See Figure 4 for a plot of the results.

For each intrinsic parameter, Figure 4 plots the error introduced (x-axis) vs the relative difference from the original calculation with no error (y-axis). The error is introduced to the individual rotations – yaw, pitch, and roll – and then introduced to all three at the same time.

5. Conclusions and Future Work

We described a method for calibrating a smartphone camera by taking two images at different rotations while tolerating small translations. We show how the amount of error in the smartphone orientation sensors and amount of translation affects the overall results. We demonstrated with synthetic data that our results are very close to the ground

truth, and with real data our results are comparable to other state of the art works.

Our work shows some advantages over other methods. First, since we are using measured rotation angles, our methods do not require us to calculate the rotation between camera positions. Secondly, we require only 2 images and no calibration board. This lends itself to future works which can calibrate a camera in real time. Also, in future work, we would like to remove effects of translation by directly calculating the infinite homography, H^∞, by using vanishing points in the scene. This will allow the calibration to be effective in both indoor environments, which can be more sensitive to translations but have easy to find vanishing points, and outdoor environments.

In conclusion this paper has 3 contributions. First, we implement an algorithm for finding the intrinsic parameters of a camera in the case of general rotation, zooming, and small translations using only two images. This algorithm is a modification of previous work by Junejo and Foroosh [13]. Secondly, we quantify how translations can affect the final results. We show that if correspondences are found on an object that is 40m away, then small translations, which are normally due to shaking or imperfect rotation about the camera center, are negligible. Finally, we quantify how errors in the orientation measurement affects the final calcu-

lated intrinsic parameters.

References

[1] L. Agapito, E. Hayman, and I. Reid. Self-calibration of rotating and zooming cameras. *Int. J. Comput. Vision*, 45(2):107–127, Nov. 2001. 2, 3

[2] P. Bolliger, M. Köhler, and K. Römer. Facet: towards a smart camera network of mobile phones. In *Proceedings of the 1st international conference on Autonomic computing and communication systems*, Autonomics '07, pages 17:1–17:10, ICST, Brussels, Belgium, Belgium, 2007. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering). 1

[3] D. C. Brown. Close-range camera calibration. *PHOTOGRAMMETRIC ENGINEERING*, 37(8):855–866, 1971. 2

[4] D. C. Brown. The bundle adjustment progress and prospects. *Int. Archives Photogrammetry*, 1976. 2

[5] L. de Agapito, R. Hartley, and E. Hayman. Linear self-calibration of a rotating and zooming camera. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 1, pages 2 vol. (xxiii+637+663), 1999. 2

[6] O. Faugeras, Q. Luong, and S. Maybank. Camera self-calibration: Theory and experiments. In G. Sandini, editor, *Computer Vision ECCV'92*, volume 588 of *Lecture Notes in Computer Science*, pages 321–334. Springer Berlin Heidelberg, 1992. 2

[7] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981. 5

[8] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, New York, NY, USA, 2 edition, 2003. 4

[9] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge books online. Cambridge University Press, 2004. 2

[10] R. I. Hartley. Self-calibration of stationary cameras. *Int. J. Comput. Vision*, 22(1):5–23, Feb. 1997. 2

[11] R. I. Hartley, E. Hayman, L. Agapito, and I. D. Reid. Camera calibration and the search for infinity. In *Proc. 7th International Conference on Computer Vision, Kerkyra, Greece*, pages 510–517, 1999. 2

[12] A. Heyden and K. Astrom. Flexible calibration: minimal cases for auto-calibration. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 1, pages 350 –355 vol.1, 1999. 2

[13] I. N. Junejo and H. Foroosh. Optimizing ptz camera calibration from two images. *Mach. Vision Appl.*, 23(2):375–389, Mar. 2012. 1, 2, 3, 5, 6

[14] F. Kahl, B. Triggs, and K. Astrom. Critical motions for auto-calibration when some intrinsic parameters can vary. *J. Math. Imaging Vis.*, 13(2):131–146, Oct. 2000. 2

[15] J.-S. Kim and I. S. Kweon. Infinite homography estimation using two arbitrary planar rectangles. In *Proceedings of the 7th Asian conference on Computer Vision - Volume Part II*, ACCV'06, pages 1–10, Berlin, Heidelberg, 2006. Springer-Verlag. 2

[16] D. Lowe. Object recognition from local scale-invariant features. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 2, pages 1150 –1157 vol.2, 1999. 5

[17] T. Moons, L. Van Gool, M. Proesmans, and E. Pauwels. Affine reconstruction from perspective image pairs with a relative object-camera translation in between. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 18(1):77 –83, jan 1996. 2

[18] G. Papagiannakis, G. Singh, and N. Magnenat-Thalmann. A survey of mobile and wireless technologies for augmented reality systems. *Computer Animation and Virtual Worlds*, 19(1):3–22, 2008. 1

[19] M. Pollefeys, R. Koch, and L. V. Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *INTERNATIONAL JOURNAL OF COMPUTER VISION*, pages 7–25, 1999. 2

[20] M. Pollefeys, R. Koch, and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *Proceedings of the Sixth International Conference on Computer Vision*, ICCV '98, pages 90–, Washington, DC, USA, 1998. IEEE Computer Society. 3

[21] K.-T. Song and J.-C. Tai. Dynamic calibration of pan/tilt/zoom cameras for traffic monitoring. *Trans. Sys. Man Cyber. Part B*, 36(5):1091–1103, Oct. 2006. 2

[22] K. Strobl and G. Hirzinger. Optimal hand-eye calibration. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pages 4647 –4653, oct. 2006. 5

[23] K. H. Strobl, W. Sepp, S. Fuchs, C. Paredes, M. Smisek, and K. Arbter. DLR CalDe and DLR CalLab. 5

[24] D. Titterton, J. Weston, D. Titterton, and J. Weston. *Strapdown Inertial Navigation Technology, 2nd Edition*. Iee Radar Series. The Institution of Engineering and Technology, 2004. 1

[25] B. Triggs, P. Mclauchlan, R. Hartley, and A. Fitzgibbon. Bundle adjustment a modern synthesis. In *Vision Algorithms: Theory and Practice, LNCS*, pages 298–375. Springer Verlag, 2000. 3

[26] R. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *Robotics and Automation, IEEE Journal of*, 3(4):323 –344, august 1987. 1

[27] B. van Tonder and J. Wesson. The impact of sensor fusion on tilt interaction in a mobile map-based application. In *Proceedings of the South African Institute of Computer Scientists and Information Technologists Conference on Knowledge, Innovation and Leadership in a Diverse, Multidisciplinary Environment*, SAICSIT '11, pages 249–258, New York, NY, USA, 2011. ACM. 1

[28] Z. Zhang. A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(11):1330 – 1334, nov 2000. 1