Camera Calibration and Pose Estimation from Planes

Hamid Bazargani and Robert Laganière

Camera calibration plays a key role in every computer vision application dealing with the problems of recovering a camera's geometry with respect to a 3D world reference, making 3D measurement in a captured scene or extracting 3D data from observed objects. These problems emerge in various applications such as structure from motion, robotics, augmented reality, 3D object recognition, and Simultaneous Localization And Mapping (SLAM). Using a camera as a measuring device is becoming an important trend [1], and as a consequence the need for calibrating this instrument also becomes of prime importance.

Cameras are, by nature, projective devices that map our 3D world onto a 2D image. Their prime use is to create representations (i.e., images) of an environment and its actors as pictured from a particular viewpoint at a precise moment in time. Extracting 3D data from such devices is therefore inherently difficult. The camera measures the intensity of the light emitted or reflected from a certain direction; however, the depth information is lost. The projective process also implies that, in the absence of any external reference, the scale of the observed objects is undeterminable; that is an object of a given size at a given camera distance will produce the same image as an object of twice the size seen at twice the distance. In spite of these limitations, with a good understanding of the camera geometry, 3D reconstruction can become possible under specific circumstances.

In this article, we review some techniques proposed in the literature to parameterize camera metric information, also referred to as camera calibration techniques. We also illustrate the use of calibrated cameras by describing two application examples involving camera pose estimation and distance estimation.

The Pin-Hole Camera Model

A camera is a simple instrument that captures rays of light coming from an observed scene and projects them onto an image sensor in order to produce a picture. This image sensor is composed of a grid of photosites that collect the incoming photons converging from the camera lens. Surprisingly, as complex as the camera technology can be, the image formation process can be accurately described by the rudimentary pin-hole model. This simple, yet realistic model assumes the image sensor is a two-dimensional Euclidean plane and is used as a standard basis to formulize cameras' *intrinsic* and *extrinsic* quantities.

As Fig. 1 shows, the camera model (also known as perspective projection) can be described as a 3 × 4 homogenous transformation that maps a 3D world point $\boldsymbol{M} = [X\ Y\ Z\ 1]^T$ expressed in homogenous coordinates to a 2D pixel $\boldsymbol{m} = [u\ v\ 1]^T$ also in homogenous coordinates. Thus, we can write:

$$s\boldsymbol{m} = \boldsymbol{P}_{3\times 4}\boldsymbol{M} \tag{1}$$

where $\boldsymbol{P}$ is a $\mathbb{R}^3 \rightarrow \mathbb{R}^2$ mapping up to scale factor s with 11 degrees of freedom. To factor out the projection matrix into intrinsic and extrinsic parameters, (1) can be decomposed as:

$$s\boldsymbol{m} = \boldsymbol{K}[\boldsymbol{R} \mid \boldsymbol{t}]M \tag{2}$$

Intrinsic Parameters

Intrinsic parameters denoted as $\boldsymbol{K}$ in (2), embed the internal configuration of a camera. The *focal length f* measured in pixel unit is the shortest distance between pinhole $\boldsymbol{C}$ and the sensor plane. The coordinate of the pixel on which the focal length vector pierces the image plane is the *principal point*. The camera principal point is ideally taken to sit at the center of the image sensor that is represented by $[c_u, c_v]^T$. To more accurately model image sensors with non-square pixels, two scale factors proportional to pixel density per unit distance, namely k_x and k_y, are introduced. A small non-zero parameter γ represents the skewness of the pixel grid and

$$\boldsymbol{K} = \begin{pmatrix} k_u f & \gamma & c_u \\ 0 & k_v f & c_v \\ 0 & 0 & 1 \end{pmatrix} \tag{3}$$

However, under a reasonable assumption [2], γ can be neglected in most normal cameras where image axes are

1094-6969/15/$25.00©2015IEEE

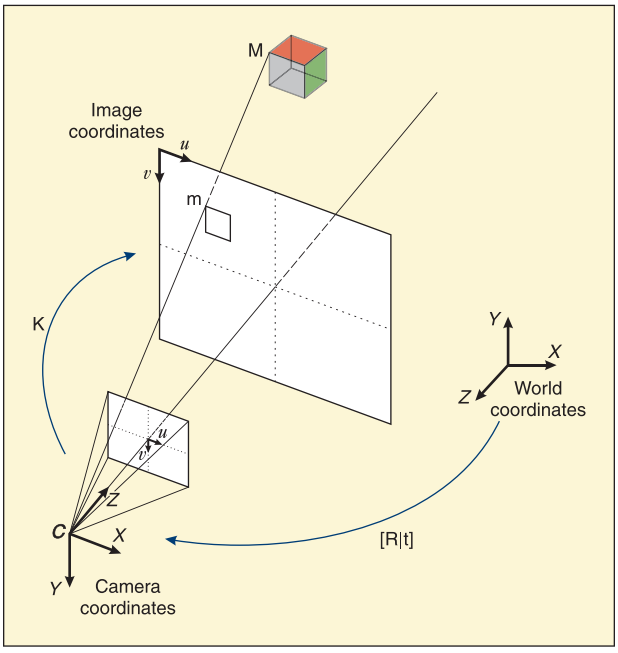

Fig. 1. Pinhole model of a digital camera.

perpendicular. Incidentally, in modern CCD cameras, pixels are usually considered to be square, thus resulting in both k_u and k_v to be equal to 1.

Extrinsic Parameters

To complete the perspective projection $\boldsymbol{P}$ matrix, camera extrinsic parameters must be taken into account. In fact, extrinsic parameters describe the position and orientation of the camera in 3D world coordinates comprising of $\boldsymbol{R}_{3\times3}$ and $\boldsymbol{t}_{3\times1}$, respectively. Referring to Fig. 1, the transformation from a 3D point in the camera coordinate system to the world coordinates can be expressed as:

$$M_w = R^{-1}\left(M_c t\right) = R^{-1}M_c - R^{-1}t \tag{4}$$

in which, $-\boldsymbol{R}^{-1}\boldsymbol{t}$ describes the camera center in world coordinates. Therefore, we can alternatively reformulate (2) as follows:

$$sm = K_{3\times3}\begin{pmatrix} R_{3\times3} & -RC \\ \mathbf{0} & \mathbf{1} \end{pmatrix} M \tag{5}$$

Lens Distortion

One aspect that is not taken into account by the pin-hole model is the distortion introduced by the camera lens. Although, it is expected to have straight lines mapped into straight lines, in some practical cases, the camera lens induces a distortion that cannot be neglected. The corresponding distortion is composed of two displacement terms, namely radial distortion and lens shift distortion. Letting x_u and x_d be undistorted and distorted 2D points, respectively, we can write:

$$x_u = x_d - d_{radial} - d_{shift} \tag{6}$$

While the lens shift is usually ignored, the radial distortion can be modeled by the following polynomial expression:

$$d_{radial} = 1 + k_1 r^2 + k_2 r^4 + \ldots, \tag{7}$$

where r is the radial distance from the center of image plane. Lenses of shorter focal length usually exhibit higher image distortion.

The Calibration Process

Camera calibration is the process by which the exact values of a camera's parameters are determined. There is a mass body of literature concerning different calibration techniques. However, we focus here on *calibration using apparatus* methods as being more practical and accurate. In its simplest form, a calibration apparatus is a 2D plane showing a pattern of known geometry. 3D calibration objects could also be used, but these ones are more difficult to manufacture and to handle. The role of the apparatus is to provide the camera to be calibrated with a set of points of known 3D coordinates (with respect to some world reference). These 3D points and their corresponding observed image on the camera sensor are then used to estimate the intrinsic parameters (and the distortion) from the projective relation that the pin-hole camera defines. The most widely used 2D apparatus are the chessboard pattern and the circle pattern (Fig. 2). They present the advantage of having calibration points (corners and circle centers) with image coordinates that can be estimated at subpixel accuracy.

In principle, the more points observed, the more accurate will be the calibration. The calibration points should also cover as much as possible the field of view of the camera as well as being presented at various depths. This requirement cannot be achieved from a single planar calibration pattern; it is therefore a common practice to capture multiple images of the calibration pattern at different positions. It follows that during the parameter estimation process not only the intrinsic will need to be estimated but also multiple extrinsic parameters, that is the position of the pattern with respect to the camera for each captured image. Note that more accurate results can be obtained when a precise motion control platform is available, such that the calibration pattern can be moved in a controlled way.

The Linear Method of Hall

In 1982, Hall *et al.* [3] proposed a direct linear method, based on direct linear transformation (DLT), to estimate elements of projection matrix by forming a set of equations. Equation (1) can be explicitly decomposed into three simultaneous equations, the first two giving expressions for su and sv while the third one gives an expression for s only. We can therefore substitute s in the first two equations by this last expression, and with some variable rearrangements, we can derive:

$$\begin{aligned} X_i p_{11} + Y_i p_{12} + Z_i p_{13} + p_{14} - u_i X_i p_{31} - u_i Y_i p_{32} - u_i Z_i p_{33} - u_i p_{34} = 0 \\ X_i p_{21} + Y_i p_{22} + Z_i p_{23} + p_{24} - v_i X_i p_{31} - v_i Y_i p_{32} - v_i Z_i p_{33} - v_i p_{34} = 0 \end{aligned} \tag{8}$$

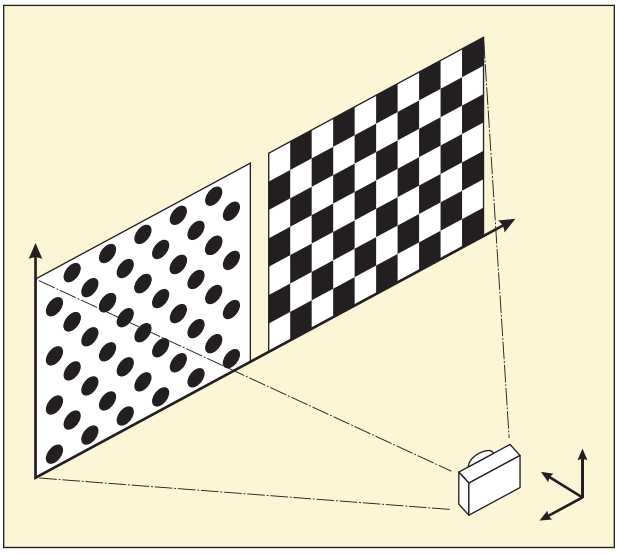

Fig. 2. Circle and chessboard calibration patterns.

where, p_{ij} constitute the 12 parameters of the projection matrix $\boldsymbol{P}$. Since $\boldsymbol{P}$ has 11 degrees of freedom, we can set the very last element of $\boldsymbol{P}$ to unity (i.e., $p_{34} = 1$). Given n pairs of correspondences, we can rewrite (8) in the form of a homogeneous system ($\boldsymbol{Ap} = \boldsymbol{0}$):

$$\begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -u_1X_1 & -u_1Y_1 & -u_1Z_1 & -u_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -v_1X_1 & -v_1Y_1 & -v_1Z_1 & -v_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_n & Y_n & Z_n & 1 & 0 & 0 & 0 & 0 & -u_nX_n & -u_nY_n & -u_nZ_n & -u_n \\ 0 & 0 & 0 & 0 & X_n & Y_n & Z_n & 1 & -v_nX_n & -v_nY_n & -v_nZ_n & -v_n \end{bmatrix} \begin{bmatrix} p_{11} \\ p_{12} \\ \vdots \\ p_{33} \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \tag{9}$$

Vector $\boldsymbol{p}$ is solved such that the algebraic error constrained by calibration points is minimized, that is:

$$\boldsymbol{p} = \text{argmin}_{\boldsymbol{p}} \ \|\boldsymbol{Ap}\|^2 \tag{10}$$

A norm constraint ($\|\boldsymbol{p}\| = 1$) is also imposed to avoid trivial solution. The least square solution is obtained by applying the so-called *Singular Value Decomposition* (SVD), so the right null vector of matrix $\boldsymbol{A}$ corresponds to the smallest eigenvalue, minimizing the error function defined by (10).

The Two-Stage Method of Zhang

Since the camera lens is precisely modeled with some non-linear terms, linear approaches like the Hall approach are inferior when dealing with the non-linear radial distortion. Zhang [4] proposed a flexible two-fold method that also covers the radial distortion. The method initially estimates an explicit solution and then refines it through an iterative scheme based on *Maximum Likelihood Estimation* (MLE). In addition, this method resolves the problem in a way that is more geometrically meaningful.

The idea behind this approach is based on homography transformations between multiple views (at least two) of a planar target. Assuming a planar calibration target to be located at Z = 0 cancels the third row of the projection matrix of (1). It results then in a new relation involving a homography $\boldsymbol{H}$ relating the 3D point with the 2D image, that is:

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \boldsymbol{H}_{3\times3}\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}, \text{where} \quad \boldsymbol{H} = K\,[\boldsymbol{r}_1\,\boldsymbol{r}_2\,\boldsymbol{t}] \tag{11}$$

where $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are the first and second vector of the rotation matrix, respectively. From the given $\boldsymbol{H}$ obtained from the DLT solution and bearing that $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are two orthonormal vectors, the cameras' intrinsic parameters can then be constrained by the following system of equations:

$$h_1^T \overbrace{K^{-T}K^{-1}}^{B} h_2 = 0 \tag{12}$$

$$h_1^T K^{-T}K^{-1}h_1 = h_2^T K^{-T}K^{-1}h_2$$

$\boldsymbol{B}$ is a symmetric matrix with six degrees of freedom, referred as to the projection of the absolute conic. Following the steps described in [4], one can find an initial estimate of intrinsic parameters $\boldsymbol{K}$, as well as the extrinsic $\boldsymbol{R}$ and $\boldsymbol{t}$.

Refinement Using Maximum Likelihood Estimation

Now that an initial estimate of the camera parameters is available, it can be refined through a Maximum Likelihood Estimation (MLE) that minimizes a geometric distance function. In this case, MLE relies on the set of independently and identically distributed data obtained from the m calibration points observed across n views. Given this set of observations, MLE determines, from a parametric model $f(.\,|\,\boldsymbol{K},\boldsymbol{R}_i,\boldsymbol{t}_i)$, the vector of parameters that exhibits the highest degree of consistency with the observed data. The distance function to be minimized is expressed as:

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\left\|\boldsymbol{m}_{ij} - \boldsymbol{P}_i\boldsymbol{M}_j\right\|^2, \text{where}\, \boldsymbol{P}_i = f\left(\boldsymbol{K},\boldsymbol{R}_i,\boldsymbol{t}_i\right) \tag{13}$$

Given an initial guess $\boldsymbol{P}_0$ from the previously found closed-form solution, an iterative non-linear minimization algorithm such as Levenberg-Marquardt is used to find the refined solution. Note that function f can also include the parameters that model the lens distortion.

Self-Calibration

In cases where an offline calibration process cannot be applied, it is still possible to rely on *self-calibration* methods. These methods require no or little prior information about the camera and can be conducted online during the scene capture process. Although these methods bring more flexibility than

classical methods, they are more complex to deploy and lead to less accurate solutions, often at a higher computational cost.

Source and Impact of Uncertainties in Calibration

When a camera is used to perform measurements, it becomes essential to understand the effect of calibration errors on 3D reconstruction. The uncertainties in the measurement system may stem from an inaccurate camera model, noisy observations or a systematic error associated with the method of calibration. It is important to realize that any error introduced in the calibration process will follow a chain of propagation that will lead to errors in the 3D measurements to be made. Error in calibration target measurements and uncertainties in extracted feature point position introduce errors in calibration parameters which, in turn, introduce errors in 3D estimations. To analyze the impact of these errors, a sensitivity analysis must be performed [5]. This is generally done by computing a covariance matrix on the calibrating objective function, describing how the variations of a noisy input parameter (e.g., feature point position) result in variations of an estimated parameter (e.g., focal length) [6].

Error analysis in vision-based measurements is also made more complex by the fact that its impact also depends on how the camera is used to perform the measurements. For example, an error in the estimate of the camera's principal point will have a more important impact on a forward-moving camera than on a camera measuring the same object but by following a lateral movement [7]**.** It has also been observed that some parameters have more impact than others on the measurement uncertainties. For example, focal length and camera distortion parameter accuracy have been shown to be critical to the production of reliable 3D estimates.

Camera Pose Estimation

Once a camera has been calibrated, it becomes possible to estimate the position of the camera with respect to some 3D objects. This problem is generally formulated as finding the pose, or extrinsic parameters, of a calibrated camera given a set of observed 3D points and their 2D images. This is known as the *Perspective-n-Point* problem (PnP). Theoretically, three points ($n = 3$) called P3P, is the minimal amount of information required to solve the PnP problem [8]. Various methods capable of explicitly solving this problem for $n \geq 3$ have been proposed; those bringing low complexity, yet accurate result are of particular interest.

One of the earliest attempts on this ground has been made by [9] to find an explicit solution for the P3P problem by forming a bi-quadratic polynomial. Following their proposed approach, the transformation is estimated by first computing the distance between 3D object points and the camera center ***C***. In Fig. 3, let L_a, L_b and L_c be the length of three legs that connect the camera center to the 3D object points A, B, and C respectively. Letting a, b and c be the projection of these points on the image plane, the goal is to determine L_a, L_b and L_c.

By applying basic geometry, one can determine the length of the three sides of DABC and its interior angles from 3D object points. Applying the law of cosines yields:

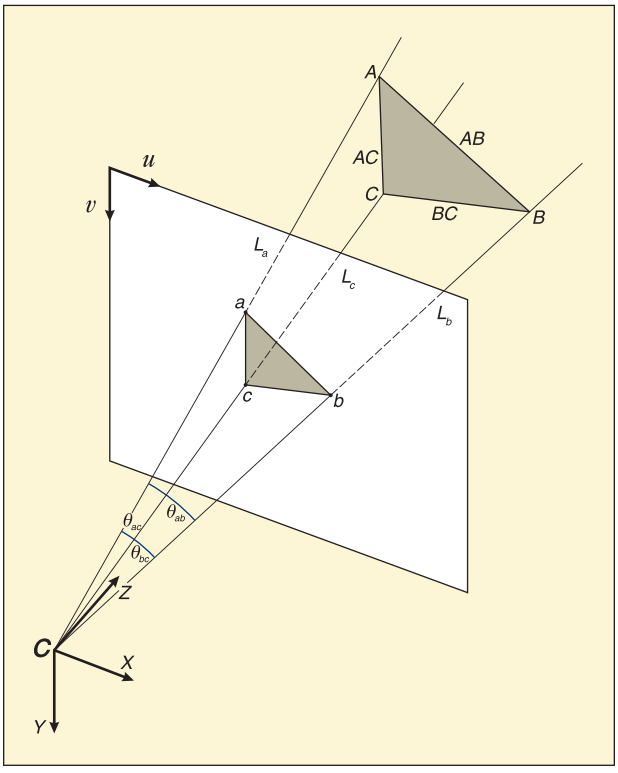

Fig. 3. P3P geometric view.

$$AB^2 = L_a^2 + L_b^2 - 2L_aL_b\cos\left(\theta_{ab}\right)$$

$$AC^2 = L_a^2 + L_c^2 - 2L_aL_c\cos\left(\theta_{ac}\right)$$

$$BC^2 = L_b^2 + L_c^2 - 2L_bL_c\cos\left(\theta_{bc}\right) \qquad (14)$$

Equation (14) comprises three second-degree polynomials; therefore, it brings up to eight unique solutions. However, since the polynomials only contain second degree and constant terms, it has a maximum of four positive solutions. If we denote $x = L_a/L_b$ and $y = L_c/L_a$ it is proven in [9] that a bi-quadratic polynomial can be derived from (14).

Although the described method is usually considered as a basic solution, other analytical and iterative solutions have been proposed by computer vision scientists. For example, Lepetit *et al.* in [10] showed that any n 3D points in a world coordinate can be expressed as a weighted sum of four virtual control points. Their so-called *EPnP* solution with $O(n)$ complexity is capable of handling arbitrary values of n. In addition, the virtual control points in EPnP also allow recovering the shape of deformable objects. POSIT developed by Daniel DeMenthon [11], is another well-known iterative method that does not require an initial guess. Instead, it starts with a *Scaled Orthographic Projection* (SOP) as an approximation to the perspective projection. The approximate projection is refined through an iterative process, minimizing the projection error. Unlike many other Newton-based methods, POSIT obtains the optimal solution by iteratively shifting object points and re-calculating a new SOP, until the shifting parameter

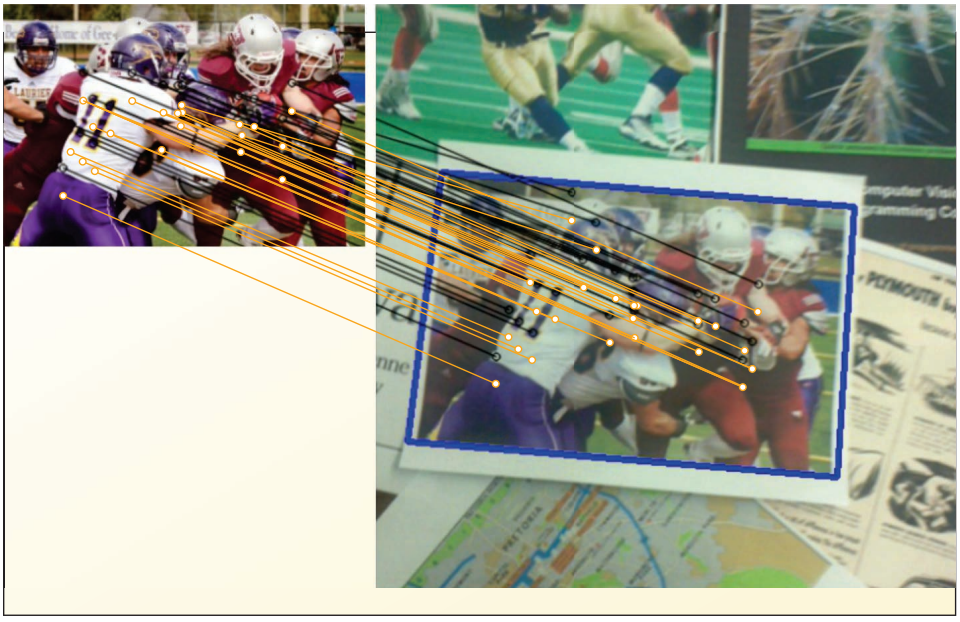

Fig. 4. Matching of local features between a target and an image of it.

remains unchanged from one iteration to the next. The POSIT algorithm has been further modified in [12] for dealing with co-planar configurations. Since an orthographic projection induces two different planes which look the same when being seen from the image plane, the algorithm selects the pose yielding minimum projection error. SoftPOSIT [13] is another variant of the POSIT algorithm that simultaneously establishes correspondences while determining the pose. SoftPOSIT exploits the POSIT strategy integrated with an iterative correspondence assignment technique referred to as softassign technique.

Pose Estimation from Planar Targets

The PnP algorithm allows estimating the pose of a camera from a set of 3D points. But one may wonder how these 3D points can be obtained in practical applications. One technique that is often used, in augmented reality applications for instance, consists in using a predefined planar target. The recent progress in image matching has made possible the reliable detection of a planar target at a high frame rate, which makes this kind of approach very attractive for the real-time pose tracking of a calibrated camera [14].

Approaches for the detection of planar targets are based on interest point detection and matching. The target is chosen, therefore, such that it contains sufficient textural details to produce enough distinctive feature points that could be reliably matched. Interest points are generally obtained using the FAST feature detector [15]. Recent binary descriptors such as BRIEF [16] or FREAK [17] combine both efficiency and distinctiveness, and when integrated to a robust matching scheme they yield to very reliable match results. Fig. 4 illustrates how a reference target is detected in a camera view by matching corresponding image points. When the size of the target image is known, it follows that the 3D coordinates of these points can be obtained in the target reference frame (generally assuming that $Z = 0$). Using one of the PnP algorithms, camera pose with respect to the target can then be obtained.

Augmented reality applications would use this information and add extra virtual points (or more generally 3D objects) inside the target's coordinate system and then back-project this 3D data onto the camera sensor using the computed projection matrix (2). However, performing per-frame pose estimation would lead to camera jittering effects due to the noisy estimation. Most augmented reality systems use different strategies to stabilize the rendering of the virtual objects such as combining estimates from several views or using Kalman filtering and feature tracking to smooth out the successive camera positions.

In [18], pose estimation from a planar target is used for the self-localization of a patrolling robot. Targets are used as natural landmarks accurately positioned in the environment patrolled by the robot. Using odometry and inertial devices, the robot navigates the space. However, as it moves, its positional estimates accumulate errors. Whenever a target falls inside the field of view of the robot's camera, the position of the robot can then be re-established. For instance, reported experiments show an error of 2.3 cm and 2.1 degrees in the estimation of the calibrated camera position from a target viewed at 3 m.

Obstacle Distance Estimation

In the previous example, a camera was freely moved over a planar pattern and its position was estimated. Let us now consider the case where the position of a camera is fixed with respect to the world and the need is now to estimate the position of observed moving objects. This is the case, for example, of a car back-up camera for which you want to estimate the distance to potential obstacles [19].

Fig. 5 illustrates a typical vehicle-mounted camera configuration. Full calibration is required here, which in addition to the intrinsic parameters will provide us with the height h and orientation $\boldsymbol{R}_c$ of the camera with respect to the ground. Without loss of generality, it can be assumed that the camera is located along the Z-axis with ground plane at $Z = 0$, which gives us $\boldsymbol{t}_c = [0, 0, h]^T$. Equation (11) has demonstrated that when a camera observes particular points on a planar surface, an invertible 3×3 homographic relation exists between the world points on the ground and their image.

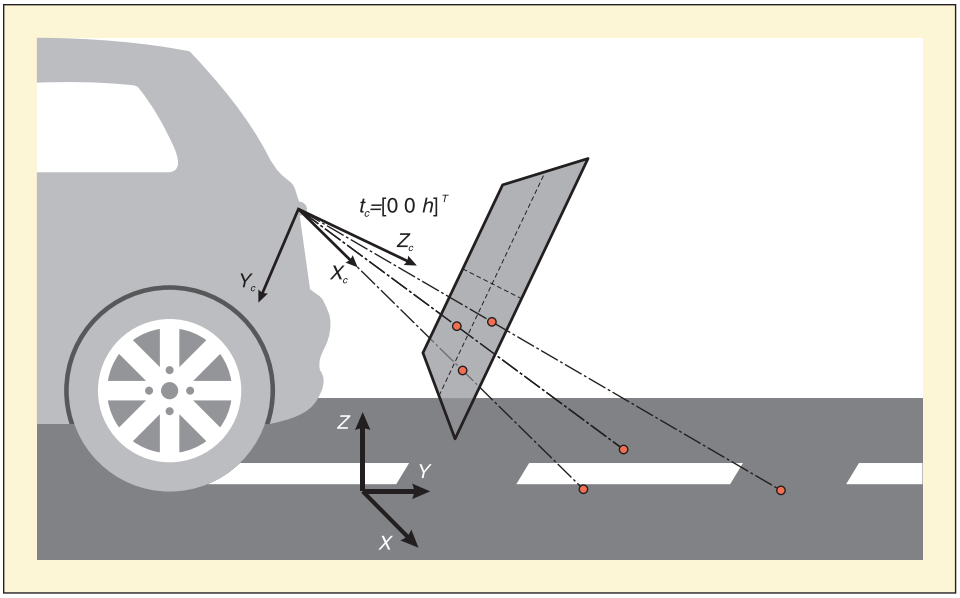

Fig. 5. Coordinate system of a mounted camera on the back of a vehicle.

Now assuming that the vehicle is moving over the ground surface, the planar motion $[\boldsymbol{R}_p, \boldsymbol{t}_p]$ of the vehicle can then be estimated by observing the image of points on the ground. These points are tracked as the vehicle moves such that their position can be obtained in two image frames. Let $\boldsymbol{m}$ and $\boldsymbol{m}'$ be the images of one of these points, and let $\boldsymbol{M}$ and $\boldsymbol{M}'$ be the 3D coordinates of these image points on the ground plane as obtained using the camera-to-world-plane homography. We can then write:

$$\boldsymbol{M}' = \boldsymbol{R}_p \boldsymbol{M} + \boldsymbol{t}_p \tag{15}$$

since this is a planar motion, the only unknowns are the 2D translation parameters and the rotation angle q of the in-plane vehicle rotation, which allows writing:

$$\begin{aligned} \alpha X_{ip} - \beta Y_{ip} + t_x - X'_{ip} = 0 \\ \beta X_{ip} + \alpha Y_{ip} + t_y - Y_{ip}' = 0 \end{aligned} \tag{16}$$

with $\alpha = \cos(\theta)$ and $\beta = \sin(\theta)$. This system of equations can be solved linearly for n points. This approximate solution can then be refined using ML estimation. Combining the planar motion estimate with the camera calibration information, we thus obtain the full camera projection matrices $\boldsymbol{P}$ and $\boldsymbol{P}'$ of the two camera views.

These motion estimates will be accurate as long as all tracked points are located on the ground plane. When an obstacle is present, then this assumption is no longer valid and some of the tracked points will be located on the visible 3D object. It is then necessary to use a robust estimation scheme like RANSAC that discards the outlier points not moving in a way consistent with the ground plane assumption. The inlier points are then used to estimate the vehicle motion while the outlier points correspond to the obstacle points. The images of each outlier point can then be triangulated to obtain the corresponding 3D position. This can be done by solving (14) but this time with the coordinates of the 3D point as unknowns; the two estimated projection matrices result in four independent linear equations. Fig. 6a demonstrates the ground surface with extracted features while Fig. 6b shows reconstructed obstacle points.

Conclusion

We reviewed in this article some of the important concepts in camera calibration. In particular, we considered the case of calibration using a planar apparatus that remains a method of choice because it allies both simplicity and accuracy. We also

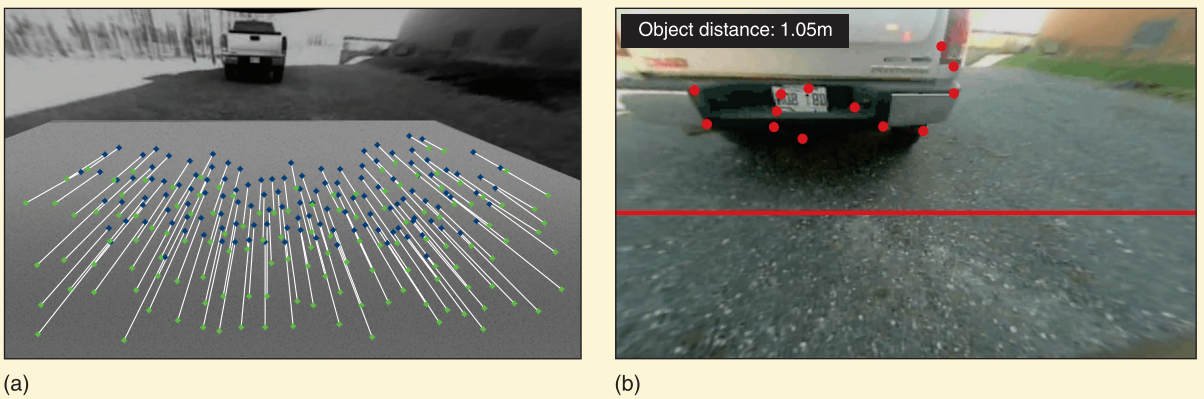

Fig. 6. (a) The tracked feature points to the ground surface. (b) The reconstructed obstacle points.

looked at the pose estimation problem that is also, in essence, a calibration problem but for which the unknowns to be recovered are the extrinsic parameters.

Reliable and accurate calibration relies on the use of good quality apparatus and on sub-pixel accurate image measurements. Refinement of the parameter values from non-linear optimization using physically meaningful geometrical quantities is also essential. Finally, a good calibration approach should also take into account the particular geometry of the configuration used.

Because of its importance in computer vision and vision-based measurement applications, camera calibration remains a very active topic of research. In addition, the advent of new imaging devices and positional sensors such as Omni-directional cameras [20], depth cameras [21], multi-view cameras, laser [22], inertial measurement units [23], also pose new research challenges.

References

[1] S. Shirmohammadi and A. Ferrero, "Camera as the instrument: the rising trend of vision based measurement," *IEEE Instrum. Meas. Mag.*, vol. 17, no. 3, pp. 41-47, 2014.

[2] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, England, UK: Cambridge University Press, 2003.

[3] E. L. Hall, J. B. K. Tio, C. A. McPherson, and F. A. Sadjadi, "Measuring curved surfaces for robot vision," *Computer 12*, no. 15, pp. 42-54, 1982.

[4] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 22, no. 11, pp. 1330–1334, 2000.

[5] L. Zhu, H. Luo, and X. Zhang, "Uncertainty and sensitivity analysis for camera calibration," *Industrial Robot: An International Journal*, vol. 36, no. 3, pp. 238 - 243, 2009.

[6] G. Di Leo and A. Paolillo, "Uncertainty evaluation of camera model parameters," in *Proc. IEEE 2011 Instrum. Meas. Tech. Conf. (I2MTC),* pp. 1-6, 2011.

[7] A. Steffen and W. Förstner, "Calibration errors in structure from motion," *Mustererkennung 1998.* Berlin, Germany: Springer-Verlag Berlin Heidelberg, pp. 117-124, 1998.

[8] X. S. Gao, X. R. Hou, J. Tang and H. F. Cheng, "Complete solution classification for the perspective-three-point problem," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 25, no. 8, pp. 930–94, 2003.

[9] M. A. Bolles and R. C. Fischler, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM 24, 6*, pp. 381-395, 1981.

[10] V. Lepetit, F. Moreno-Noguer F. and P. Fua, "EPnP: an accurate O(n) solution to the PnP problem," *Int. J. Comput. Vision 81, 2*, pp. 155-166, 2009.

[11] D. DeMenthon and L. S. Davis, "Model-based object pose in 25 lines of code," in *European Conf. on Computer Vision*, 1992, pp. 335–343.

[12] D. Oberkampf, D. DeMenthon, and L. Davis, "Iterative pose estimation using coplanar points," in *Proc. Computer Vision and Pattern Recognition, CVPR'93,* pp. 626-627, 1993.

[13] P. David, D. DeMenthon, R. Duraiswami, and H. Samet, "Softposit: simultaneous pose and correspondence determination," in *Proc. Computer Vision ECCV 2002*, pp. 698-714, 2002.

[14] H. Bazargani, O. Bilaniuk and R. Laganière, "A fast and robust homography scheme for real-time planar target detection," *J. Real-Time Image Processing,* pp. 1-20, 2015.

[15] E. Rosten and T. Drummond, "Machine learning for high-speed corner," in *EECV '06, Lecture Notes in Computer Science*, pp. 430-443, 2006.

[16] M. Calonder, V. Lepetit, C. Strecha and P. Fua, "BRIEF: Binary robust independent elementary features," in *ECCV '10, Lecture Notes in Computer Science*, pp. 778-792, 2010.

[17] A. Alahi, R. Ortiz, and P. Vandergheynst, "FREAK: Fast retina keypoint," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR),* pp. 510-517, 2012.

[18] V. Lepetit, L. Vacchetti, D. Thalmann, and P. Fua, "Fully automated and stable registration for augmented reality applications," in *Proc. 2nd IEEE/ACM Intern. Symp. Mixed and Augmented Reality* (*ISMAR '03*), p. 93, 2003.

[19] J. Lalonde, R. Laganiere, and L. Martel, "Single-view obstacle detection for smart back-up camera systems," in *Proc. IEEE Computer Vision and Pattern Recognition Workshops* (*CVPRW*), pp. 1-8, 2012.

[20] C. Mei and P. Rives, "Single view point omnidirectional camera calibration from planar grids," in *Proc. IEEE 2007 Intern. Conf. Robotics and Automation,* 2007.

[21] R. Macknojia, A. Chávez-Aragón, P. Payeur, and R. Laganière, "Calibration of a network of Kinect sensors for robotic inspection over a large workspace," in *Proc. IEEE Workshop on Robot Vision*, pp. 184-190, 2013.

[22] J. L. L. Galilea, J.-M. Lavest, C. A. L. Vazquez, A. G. Vicente, and I. B. Munoz, "Calibration of a high-accuracy 3-D coordinate measurement sensor based on laser beam and CMOS camera," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 9, pp. 3341-3346, 2009.

[23] G. Panahandeh, M. Jansson and P. Handel, "Calibration of an IMU-camera cluster using planar mirror reflection and its observability analysis," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 1, pp. 75-88, 2015.

Hamid Bazargani (hbaza043@uottawa.ca) received the M.A.Sc. degree in Electrical and Computer Engineering from the University of Ottawa, ON, Canada, in 2014. He has been a computer vision research assistant at the VIVA Research Laboratory, University of Ottawa since 2013. His primary research areas include real-time embedded and mobile vision, object recognition and visual tracking, pattern recognition and robust estimation. He is currently an embedded vision researcher at Synopsys Inc., world leader in electronic design and automation.

Robert Laganière (laganier@eecs.uottawa.ca) received the Ph.D. degree from INRS-Telecommunications, Montreal, QC, Canada, in 1996. He is a Full Professor and a Faculty Member with the VIVA Research Laboratory, School of Electrical Engineering and Computer Science, University of

Ottawa, Ottawa, ON, Canada. He is the author of *OpenCV Computer Vision Application Programming* published by Packt Publishing and the co-author of *Object-Oriented Software Engineering* published by McGraw Hill. His research interests are in computer vision and image processing with applications to visual surveillance, human detection and recognition, embedded and mobile vision, content-based video analysis.