PLANE RECTIFICATION THROUGH ROBUST VANISHING POINT TRACKING USING THE EXPECTATION-MAXIMIZATION ALGORITHM

Marcos Nieto and Luis Salgado

Grupo de Tratamiento de Imágenes - E. T. S. Ing. Telecomunicación
Universidad Politécnica de Madrid - Madrid (Spain)
{mnd, lsa}@gti.ssr.upm.es

ABSTRACT

This paper introduces a new strategy for plane rectification in sequences of images, based on the Expectation-Maximization (EM) algorithm. Our approach is able to compute simultaneously the parameters of the dominant vanishing point in the image plane and the most significant lines passing through it. It is based on a novel definition of the likelihood distribution of the gradient image considering both the position and the orientation of the gradient pixels. Besides, the mixture model in which the EM algorithm operates is extended, compared to other works, to consider an additional component to control the presence of outliers.

Some synthetic data tests are described to show the robustness and efficiency of the proposed method. The plane rectification results show that the method is able to remove the perspective and affine distortion of real traffic sequences without the need to compute two vanishing points.

Index Terms— Plane rectification, vanishing point, Expectation-Maximization, outliers rejection

1. INTRODUCTION

In the field of traffic applications the rectification of the road plane has been extensively used as a great help to perform metric measurements in static surveillance [1], and onboard systems [2].

Plane rectification can be done using the information contained in vanishing points to compute the road-plane to image-plane homography. Typically, two orthogonal vanishing points are detected and used to determine the line at the infinity and thus, the projective distortion can be removed [3]. However, in road scenarios, there is typically information about just one dominant vanishing point, hence there are no reliable methods to compute such rectification. Fortunately, as shown by Schaffalitzky [4], three or more lines that are parallel (thus converging into a single vanishing point) and equally spaced, can determine in closed form the line at the infinity, from which, knowing the camera calibration matrix, an orthogonal vanishing point in the plane can be obtained. This way only one vanishing point determined by such three lines must be detected to perform plane rectification. In the road scenario, three lane markings corresponding to two adjacent lanes satisfy this condition.

The estimation of the dominant vanishing point inside the limits of the image can be of great complexity, considering the reduced amount of information of the scenario and the large proportion of

_____

outliers. Specific methods that work on the image plane, using accumulators [5], or transforms to the polar space [6] lacks the required robustness for this scenario since they do not separate inliers from outliers [7], which may cause very inaccurate estimations. General methods, which work on the projective plane [8][9], do not specifically treat this scenario, and typically work very well for vanishing points far away from the image boundaries. Besides, these strategies typically require the knowledge of the calibration matrix to be able to treat the points at the infinity.

In this paper we propose a robust tracking system, based on the Expectation-Maximization algorithm [10] that is able to determine the position of the dominant vanishing point of the image plane as well as the required set of converging lines to perform the plane rectification at each time instant. The main advantages of the proposal come from the usage of the EM algorithm. On the one hand, it is an iterative procedure that is guaranteed to converge, and that can be used as a tracking system, as each estimate can be used in the next time instant as initialization. On the other hand, a mixture model is defined that considers an implicit classification between inliers and outliers, which makes the method very robust. Besides, the likelihood function uses the information of position and orientation of the gradient pixels, which improves the models of previous approaches only based on position [7].

2. VANISHING POINT TRACKING

This section introduces the likelihood model for gradient pixels, which can be obtained with any method that approximates the gradient, such as the Sobel operator. The EM algorithm leads to the equations that describe the optimization problem, considering also a distribution to handle the presence of outliers.

2.1. Likelihood model

Let $\mathbf{x}_i = \{\mathbf{r}_i, \mathbf{g}_i\}$ be a data sample composed by $\mathbf{r}_i = (x_i, y_i)^\top$, a point of the image plane with significant gradient magnitude, and the gradient vector $\mathbf{g}_i = (g_{xi}, g_{yi})^\top$.

Lines in the image plane are parameterized in the homogeneous form as $\mathbf{l}_j = (a_j, b_j, c_j)^\top$, subject to $a_j^2 + b_j^2 = 1$. The perpendicular distance from a point, $\mathbf{r}_i$, to a line, $\mathbf{l}_j$, is defined as $f_\rho(\mathbf{r}_i, \mathbf{l}_j) = a_j x_i + b_j y_i + c_j$. In the case the line passes through a vanishing point $\mathbf{v} = (v_x, v_y)^\top$, then $c_j = -a_j v_x - b_j v_y$ and the function f_ρ depends on the parameters $\mathbf{l}_j$ and $\mathbf{v}$:

$$f_\rho(\mathbf{r}_i, \mathbf{l}_j, \mathbf{v}) = a_j(x_i - v_x) + b_j(y_i - v_y) \qquad (1)$$

Therefore, a point lies in a line if $f_\rho(\mathbf{r}_i, \mathbf{l}_j) = 0$ holds. Considering Gaussian measurement noise, the likelihood of a point, $\mathbf{r}_i$, belonging to a line j is defined as:

$$p(\mathbf{r}_i|\mathbf{l}_j, \mathbf{v}) = \frac{1}{\sqrt{2\pi}\sigma_{\rho,j}} \exp\left(-\frac{1}{2\sigma_{\rho,j}^2} f_\rho(\mathbf{r}_i, \mathbf{l}_j, \mathbf{v})^2\right) \quad (2)$$

where $\sigma_{\rho,j}$ is the standard deviation of the normal distribution.

For each data sample, the gradient vector $\mathbf{g}_i$ depicts the normal to the direction of the edge at $\mathbf{r}_i$. The difference between the orientation of a line j, which is given by the vector $\mathbf{m}_j = (-b_j, a_j)^\top$, and the one of the data sample is defined as $f_\phi(\mathbf{g}_i, \mathbf{l}_j) = \frac{\mathbf{g}_i^\top \mathbf{m}_j}{\|\mathbf{g}_i\|\|\mathbf{m}_j\|}$.

Considering, as in the case of the position, a normal distribution on the orientation, the likelihood of a vector $\mathbf{g}_i$ to represent the orientation of a line is given by:

$$p(\mathbf{g}_i|\mathbf{l}_j) = \frac{1}{\sqrt{2\pi}\sigma_{\phi,j}} \exp\left(-\frac{1}{2\sigma_{\phi,j}^2} f_\phi(\mathbf{g}_i, \mathbf{l}_j)^2\right) \quad (3)$$

Assuming that the position of data samples and their orientation are independent, the joint likelihood of $\mathbf{x}_i$ to represent a line $\mathbf{l}_j$ is:

$$p(\mathbf{x}_i|\mathbf{l}_j, \mathbf{v}) = p(\mathbf{r}_i|\mathbf{l}_j, \mathbf{v})p(\mathbf{g}_i|\mathbf{l}_j) \quad (4)$$

which is the expression that links each data sample with a hypothesized line oriented in the direction of a vanishing point.

2.2. Mixture model likelihood

If we consider now M lines, the likelihood of a data sample is given by the mixture model:

$$p(\mathbf{x}_i; \Theta, \Omega) = \sum_{j=1}^{M} \omega_j \, p(\mathbf{x}_i|\mathbf{l}_j, \mathbf{v}) \quad (5)$$

where $\Theta = \{\mathbf{v}, \{\mathbf{l}_j\}_{j=1}^M\}$ and $\Omega = \{\omega_j\}_{j=1}^M$, are the parameters of the likelihood distribution. Note that the weighting factors must satisfy $\sum_{j=1}^M \omega_j = 1$, as they can be seen as the a priori probability values for each component of the mixture.

The joint probability density of all the data samples, assumed to be i.i.d., is given by $p(\mathcal{X}|\Theta) = \prod_{i=1}^N p(\mathbf{x}_i; \Theta, \Omega)$. The log-likelihood function transform the product into a summation, expanding the expression of the mixture defined in (5):

$$\log p(\mathcal{X}|\Theta) = \sum_{i=1}^{N} \log \sum_{j=1}^{M} \omega_j \, p_j(\mathbf{x}_i|\mathbf{l}_j, \mathbf{v}) \quad (6)$$

The MLE problem to be solved is to find the parameters Θ that maximize $\log p((\mathcal{X}|\Theta))$. However, partial derivatives of the log-likelihood with respect each parameter yield expressions that depend on the derivative of a natural logarithm of a sum, which is not analytically tractable.

2.3. EM algorithm

The EM strategy offers a solution to this MLE problem [10]. The EM algorithm proceeds iteratively, applying two steps: E-step, that obtains the conditional probability of each data sample to belong to each of the components of the mixture model; and M-step, which searches for the values of the target parameters that maximize the likelihood function.

The conditional probability of the samples to belong to a specific component j is given by:

$$p(j|\mathbf{x}_i, \Theta^*) = \frac{\omega_j^* p_j(\mathbf{x}_i|\mathbf{l}_j, \mathbf{v})}{\sum_{k=1}^{M} \omega_k^* p_k(\mathbf{x}_i|\mathbf{l}_k, \mathbf{v})} \quad (7)$$

where Θ^* and ω_j^* are, respectively, the set of parameters and the a priori probabilities for component j at the previous iteration. From here on, $p(j|\mathbf{x}_i, \Theta^*)$ will be denoted as γ_{ij}.

Regarding the M-step, the estimation of the a priori probabilities is given by $\hat{\omega}_j = \frac{1}{N} \sum_{i=1}^N \gamma_{ij}$, which is, as expected, independent of the rest of parameters of the model. For the maximization related to the line parameters, Lagrange multipliers are required to add the restriction $a^2 + b^2 = 1$ which leads to a linear system of equations:

$$C\mathbf{n}_j = \lambda \mathbf{n}_j \quad (8)$$

where $\mathbf{n}_j = (a_j, b_j)^\top$, and $C = \hat{\sigma}_\rho^{-2} A + \hat{\sigma}_\phi^{-2} B$. Matrices A and B contain, respectively, the information given by the weighted samples of position and orientation. For the sake of simplicity in the notation, we define the weighted sample mean of a generic data set $\{z_i\}_{i=1}^N$, as $\xi(z) = \sum_{i=1}^N z_i \gamma_{ij} (\sum_{i=1}^N \gamma_{ij})^{-1}$, yielding the following matrices:

$$A = \begin{pmatrix} \xi(\tilde{x}^2) - \xi^2(\tilde{x}) & \xi(\tilde{x}\tilde{y}) - \xi(\tilde{x})\xi(\tilde{y}) \\ \xi(\tilde{x}\tilde{y}) - \xi(\tilde{x})\xi(\tilde{y}) & \xi(\tilde{y}^2) - \xi^2(\tilde{y}) \end{pmatrix} \quad (9)$$

where random variables, $\tilde{x}$ and $\tilde{y}$ include the vanishing point components: $\tilde{x}_i = x_i - v_x$ and $\tilde{y}_i = y_i - v_y$, and:

$$B = \begin{pmatrix} \xi(g_y^2) & -\xi(g_x g_y) \\ -\xi(g_x g_y) & \xi(g_x^2) \end{pmatrix} \quad (10)$$

The maximization of the vanishing point is obtained as:

$$\hat{\mathbf{v}} = \left(\sum_{j=1}^{M}\sum_{i=1}^{N} N_j \gamma_{ij}\right)^{-1} \sum_{j=1}^{M}\sum_{i=1}^{N} \mathbf{r}_i^\top N_j \gamma_{ij} \quad (11)$$

where $N_j = \mathbf{n}_j \mathbf{n}_j^\top$ is a 2×2 matrix. As expected, the maximum likelihood estimation of a_j and b_j depend on $\mathbf{v}$ as in (9), and viceversa, as shown in (11). Therefore, the maximization process must be done in a two-steps process, which is known as gradual M-step [7]. That way it is necessary to proceed fixing one of this parameters and updating the other, so that in two steps, both parameters are updated.

2.4. Outliers control

In our EM scheme proposal, the presence of outliers is handled with the inclusion of an additional component, $p_{M'}(\mathbf{x}_i|\mathbf{v})$, for the mixture model. That way, M components are assumed by the EM to represent the data samples drawn from the M most significant lines in the scene. The rest of the data samples are absorbed by the outlier distribution. Hence, in our approach, the complete mixture model has $M' = M + 1$ components.

The mixture model (5) is rewritten as:

$$p(\mathbf{x}_i; \Theta, \Omega) = \omega_{M'} \, p_{M'}(\mathbf{r}_i|\mathbf{v}) + \sum_{j=1}^{M} \omega_j \, p(\mathbf{x}_i|\mathbf{l}_j, \mathbf{v}) \quad (12)$$

where the components $j = 1, 2, \ldots, M$ are the ones already described, and $p_{M'}(\mathbf{x}_i|\mathbf{v})$, is the outlier distribution. We propose to define this distribution as the sum of a uniform distribution of value $\rho = (\pi W H)^{-1}$ that integrates to one within the data range, and a normal distribution on the distance of the data samples position, $\mathbf{r}_i$, to the vanishing point, $d^2(\mathbf{r}_i, \mathbf{v}) = (x_i - v_x)^2 + (y_i - v_y)^2$:

$$p_{M'}(\mathbf{r}_i|\mathbf{v}) = \rho + \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2} d^2(\mathbf{r}_i, \mathbf{v})\right) \quad (13)$$

Table 1. Mean number of iterations for pEM and gpEM (100 tests).

Parameters noise	# it. pEM	# it. gpEM
$\{\sigma_{a,b} = 0.10, \sigma_{v_{x,y}} = 10\}$	8.4	4.7
$\{\sigma_{a,b} = 0.15, \sigma_{v_{x,y}} = 15\}$	9.2	5.1
$\{\sigma_{a,b} = 0.20, \sigma_{v_{x,y}} = 20\}$	12.7	6.5
$\{\sigma_{a,b} = 0.25, \sigma_{v_{x,y}} = 25\}$	16.4	7.8

The uniform component models the general uncertainty about outliers, while the second term models that data samples near the vanishing point are more likely outliers (as they are images of very remote elements of the scene). The equations of the EM algorithm are still valid by just updating the mixture model with the outlier distribution.

3. HOMOGRAPHY COMPUTATION

The road plane rectification can be done by means of a projective transform, known as homography, that links, up to scale, points in the plane coordinate system and their image in the image plane. Typically the homography is computed using the DLT algorithm [11], using at least four correspondences between points of the image plane and road plane, or in an stratified manner, knowing two vanishing points and other information about the scene, such as known length or angle ratios[3].

In this paper we propose to use the DLT method, for which we need to find four image points that correspond to points in the road that form a rectangle. This way we can remove both the projective and the affine distortion, which is enough for the purposes of the application (the only remaining distortion is an anisotropic scaling and rotation). In order to do so, we use the information of the lines obtained with the EM vanishing point tracker, that we assume to be equally spaced in the plane. To compute the line at the infinity, without need to compute two orthogonal vanishing points, we follow the method of Schaffalitzky [4]:

$$\mathbf{l}_\infty = \det[\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}]\mathbf{l}_3 - \det[\mathbf{l}, \mathbf{l}_2, \mathbf{l}_3]\mathbf{l}_1 \tag{14}$$

where $\mathbf{l}$ is any 3-vector for which the determinants are nonzero.

Knowing the camera calibration matrix, K, we can translate both the line at the infinity and the known vanishing point, $\mathbf{v}_1$, to the camera coordinate system, as $\mathbf{l}'_\infty = K^\top \mathbf{l}_\infty$ and $\mathbf{v}'_1 = K^{-1}\mathbf{v}_1$ and compute the position of the orthogonal vanishing point as $\mathbf{v}'_2 = \mathbf{l}'_\infty \times \mathbf{v}'_1$. This way we can trace two lines converging in $\mathbf{v}_2$ that intersect with two of the lines obtained with the EM. The intersection of these lines are the four required points.

4. RESULTS

In this section we first compare the proposed EM procedure with the one of Minagawa [7]. For fair comparisons, we add the outlier distribution to this approach and test the benefits of using the orientation information to the models. We will denote this method as point-EM (or pEM), and ours as gradient-point-EM (gpEM). In second place, we present some results of the plane rectification achieved through the computation of the vanishing point and the line at the infinity as described in previous sections.

4.1. Synthetic data tests

For the experiments, let's consider three lines converging into a common vanishing point, as shown in Fig. 1 (a), for which a set of N data
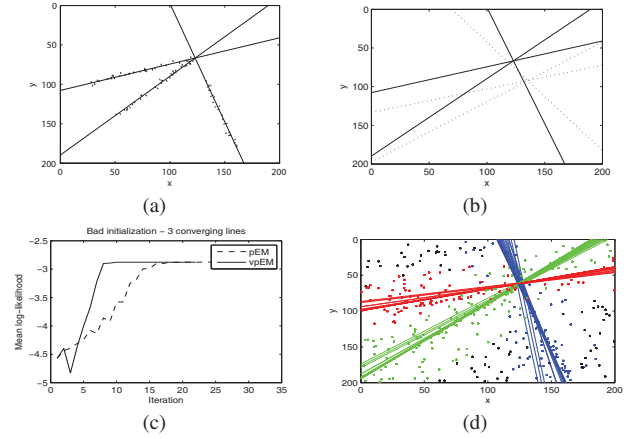
Fig. 1. Synthetic data test: (a) the ground truth model with N = 100 data samples drawn with Gaussian noise; (b) initialization, in dotted lines, and final estimation in solid lines; (c)likelihood value at each iteration in solid lines for the proposed method, in dotted lines for the method without using the orientation information; (d) adding 300 outliers, the estimation is still correct.

samples are drawn, corrupted by Gaussian noise. For these experiments we define $\sigma_\rho = 5$ (pixels) and $\sigma_\phi = 0.1$ (rads), which is in line to the expected noise present in a real image.

An example initialization and final estimation are shown in Fig. 1 (b). As previously mentioned, the algorithm iterates until convergence is achieved. We define the convergence related to the mean log-likelihood value, which is given by:

$$\bar{\mathcal{L}}(\Theta|\mathcal{X}) = \frac{1}{N} \sum_{i=1}^{N} \log \sum_{j=1}^{M} \omega_j \, p(\mathbf{x}_i|\mathbf{l}_j, \mathbf{v}) \tag{15}$$

At each iteration, indexed as k, the improvement of mean log-likelihood is checked as $\Delta\bar{\mathcal{L}}_k = \frac{\bar{\mathcal{L}}_k - \bar{\mathcal{L}}_{k-1}}{\bar{\mathcal{L}}_{k-1}}$. When $\Delta\bar{\mathcal{L}}_k$ is below a certain threshold, ϵ, the process is considered to have converged. Typically it is accurately enough to use $\epsilon = 10^{-5}$. In Fig. 1 (c), a comparison of the convergence of the gpEM and the pEM is shown. It is clear that our method achieves faster convergence for this example, showing just a small oscillation at the firsts iterations due to the gradual M-step.

This experiment was repeated several times, using different initializations of the parameter set, with similar results. For convenience, the parameters were initialized randomly, using zero mean Gaussian noise and a defined standard deviation for each parameter. Table 1 shows the results, in terms of the mean number of required iterations to converge, for 100 executions of the proposed method, and the method we compare with. The contents of the table reveal that the required number of iterations of the pEM increases dramatically as the initialization get worse, while for the gpEM, the number of iterations is approximately delimited and in all cases much more lower than for pEM. The conclusion is that the utilization of gradient information makes the system much faster and less dependent on the deviation of the initialization.

Another test is done to demonstrate the need for the outlier distribution. Its addition to the mixture model aims to absorb the mismatching between the data set and the model. The performance of our method against outliers randomly distributed among the image is

shown in Fig. 1 (d). As shown, the data set of the example in (a) contains a large number of outliers. The outlier distribution makes that most part of outliers are discarded for the estimation process of the hypothesized lines, so that the result is correct. This method shows excellent results in all cases and only begins to suffer significant errors when the proportion of outliers reaches 75%. The color code is the following: blue, red and green lines are fitted using the corresponding colored data samples, which are automatically selected. The black data samples are those marked as outliers.

4.2. Plane rectification results

We have applied our method to different road sequences, with a forward looking camera installed inside a vehicle. Fig. 2 shows some example images of the computed sequences.

The EM works as a tracker, where each instantaneous estimation is used as initialization for the next time instant, since the motion between two consecutive frames is low. For this reason, initializations are quite close to the optimal solution and the EM algorithm finds it with just very few iterations, typically between two and three.

As shown in the examples of Fig. 2 (a), where outliers are depicted as black dots, the method is very robust, leading to the correct estimation of the vanishing point, since outliers are absorbed by the outlier distribution and thus do not corrupt neither the estimation of the vanishing point nor the parameters of the lines. This is particularly significant in the example of the last row of Fig. 2. This way, the estimation of the target parameters is very steady along time, and thus the proposed method allows to compute a road plane rectification adapted to the motion of the vehicle. Therefore, this proposal overcomes problems of other approaches that also compute road rectification but without updating the information of the homography [2].

This strategy can also be applied to static surveillance scenarios, to avoid the computation of the second vanishing point, for which typically there is not enough information, or its hard to obtain [1].

Fig. 2 (c) shows examples of the obtained plane rectification. The two first rows depict a typical road, with well painted lane markings and straight lanes. The third one is more complex as there is less information about lane markings and there is also curvature in the road, which is much more simple to estimate using the rectified images.

5. CONCLUSIONS

In this paper we have introduced a new approach towards plane rectification, using the EM algorithm for the task of vanishing point tracking in sequences of images. Our strategy allows to compute simultaneously the dominant vanishing point of the scene as well as the most significant lines in that orientation. The proposed method is very robust against large amounts of outliers, thanks to the integration of an outlier distribution into the EM framework.

The plane rectification is achieved by means of the computation of the line at the infinity, avoiding to compute a second vanishing point, which is not typically feasible in traffic scenarios. The results show that the proposed method is able to remove the projective and the affine distortion of the images, which results on a stable generation of fronto-parallel views of the road plane.

6. REFERENCES

[1] C. Maduro, K. Batista, P. Peixoto, and J. Batista, "Estimation of vehicle velocity and traffic intensity using rectified images.,"
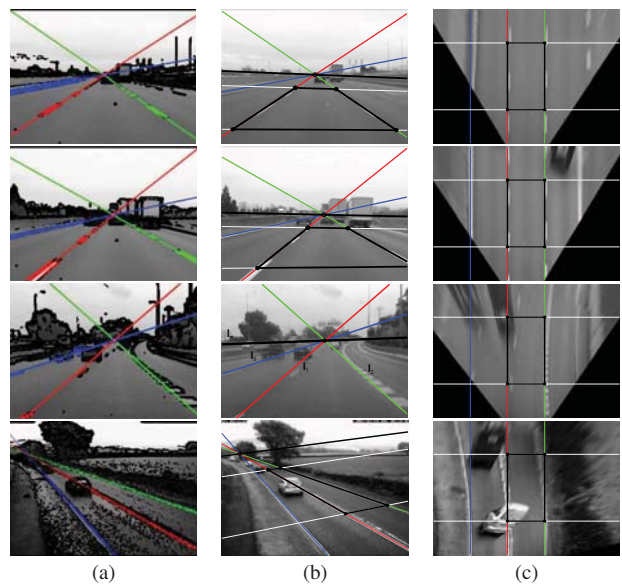
Fig. 2. Examples of the application of the proposed EM algorithm for plane rectification in onboard and static traffic sequences: (a) EM-based classification of the data; (b) computation of l_∞ and v_2 and the four control points; and (c) the virtual fronto-parallel view of the road plane, with the projected control points and detected lines.

in *IEEE Int. Conf. on Image Processing, ICIP 2008*, 2008, pp. 777–780.

[2] M. Bertozzi and A. Broggi, "Gold: A parallel real-time stereo vision system for generic obstacle and lane detection," *IEEE Trans. on Image Processing*, vol. 7, No. 1, pp. 62–81, 1998.

[3] D. Liebowitz and A. Zisserman, "Metric rectification for perspective images of planes," *IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 0, pp. 482, 1998.

[4] F. Schaffalitzky and A. Zisserman, "Planar grouping for automatic detection of vanishing lines and points," *Image and Vision Computing*, vol. 18, pp. 647–658, 2000.

[5] T. Tuytelaars, M. Proesmans, and L. Van Gool, "The cascaded hough transform," in *in IEEE Proc. International Conference on Image Processing 1998*, 1998, pp. 736–739.

[6] V. Cantoni, L. Lombardi, M. Porta, and N. Sicard, "Vanishing point detection: representation analysis and new approaches," in *in Proc. International Conference on Image Analysis and Processing (ICIAP 2001)*, 2001, pp. 26–28.

[7] A. Minagawa, N. Tagawa, T. Moriya, and T. Gotoh, "Vanishing point and vanishing line estimation with line clustering," *IEICE Trans. Inf. & Syst.*, vol. E83-D, no. 7, July 2000.

[8] J. Kosecka and W. Zhang, "Video compass," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2002.

[9] R. Pflugfelder, *Self-calibrating cameras in video surveillance*, Graz University of Technology, May 2008.

[10] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[11] R. Hartley and A. Zisserman, *Multiple View Geometry in computer vision*, Cambridge University Press, 2001.