About conditions for recovering the metric structures of perpendicular planes from the single ground plane to image homography

Pierre Gurdjos, René Payrissat
IRIT-UPS, 118 route de Narbonne
31062 Toulouse Cedex, France
{Pierre.Gurdjos,Rene.Payrissat}@irit.fr

Abstract

The absolute conic is the key for recovering the Euclidean structure in 3-D space. The image of the absolute conic can be parametrized by the set of three vanishing points corresponding to orthogonal directions in space with two independent additional factors. Whereas a single plane to image homography only allows the image to inherit two vanishing points and one of these factors, we describe a method which recovers the metric structures of any perpendicular plane to the ground from the ground to image homography matrix. Indeed, under the assumption of a natural camera, we demonstrate that the third vanishing point lies on a line that we call central line. A direct solution is found with an additional constraint relevant to the application we deal with (automatic analysis of soccer video) : a "pseudo-parallelism" between one of the camera axis and the ground plane. Direct applications on video mpeg-encoded images are presented in the experiment section showing a very satisfactory accuracy.

1. Introduction

Recent papers[1][2][6][7] have described a technical framework for single view metrology using the properties of vanishing points. The absolute conic[3] is the key for recovering metric informations in the 3-D space[8] and is imaged as a 3×3 symmetric matrix ω with 5 degrees of freedom. Matrix ω can be parametrized by the set of three vanishing points corresponding to orthogonal directions and two independent additional factors. In previous works, all the techniques concerned with planar scenes required at least three mutually conjugated vanishing points with respect to ω and two additional constraints (*eg* about the camera). Although a single plane to image homography only allows the image to inherit two vanishing points and one of the factors, we will describe a direct method, relevant to our application, that computes the homography associated with any plane perpendicular to the ground, from the sole ground to image homography. This study has actually been motived by problems encountered in automatic analysis of soccer video, in which the ground plane to image homography can be computed. Frequently, this mapping is not sufficient and metric informations in perpendicular planes are often required (ball trajectory[9], measurements of player heights for tracking[4], augmented reality...) in order to complete the analysis. Unfortunately, most of the time reliable 3-D affine or Euclidean constraints are not available in the image, except for the ground plane. The problem is set under the conditions of a natural camera (zero-skew and known aspect ratio) and a "pseudo" parallelism between of one of the camera axis and the ground. The latter condition comes from the technique used by TV broadcast for shooting soccer images. The direct applications of this method can be measurements of player heights and virtual object insertion. As shown in the application section, it yields a quite good accuracy.

About metrology in the image frame, it is worthy of note that some input data of the algorithms presented in the above-mentioned papers are Euclidean distances or norms computed in the uncalibrated image frame. Consequently, we point out that actually authors *implicitly* assumed that these algorithms dealt with a model of natural camera, *ie* with zero-skew and unit aspect ratio, even if they do not give precise details about it. As a general rule, to implement techniques using Euclidean distances or norms with the most general 5-internal parameter model of camera, is inconsistent since, as soon as the image coordinate system is not orthonormal, the knowledge of the associated scalar product is necessary, *ie* the camera skew and pixel aspect ratio have to be known.

In order to explain how to recover the metric structure of any perpendicular plane to the ground from the ground to image homography matrix, it seems interesting to us to firstly tackle the problem the computation of measurements of person heights. Similar works[5] required the vanishing line associated with a reference (ground) plane, the vanishing point associated with the direction perpendicular to the reference plane and a reference length in the direction of the vanishing point. We will show that the computation of distances orthogonal to the ground is a special case of our algorithm. Our approach differs from similar works because neither reference length or direction are required.

2. Metrology in perpendicular planes

Given the world plane to image homography matrix $\mathbf{H}$ associated with the ground, denoted π, we are concerned with measurements of distances orthogonal to this reference plane π. Let $\mathtt{T}$ be the a 3-D point not in π, imaged as vector $\mathbf{t}$. Let $\mathtt{B}$ be the orthogonal projection of $\mathtt{T}$ onto π, imaged as vector $\mathbf{b}$. Our aim is to compute the distance between $\mathtt{T}$ and $\mathtt{B}$, that is to say the distance of $\mathtt{T}$ orthogonal to π. Let us assume that we have carried out an appropriate coordinate transformation of the 3-D world space such as the origin of the world coordinate system is $\mathtt{B}$ and such as plane π has equation $Z = 0$. Let π^* be the plane perpendicular to π with equation $Y = 0$ and let $\mathbf{H}^*$ denotes the plane to image homography matrix associated with π^*. When homography matrix $\mathbf{H}$ (resp. $\mathbf{H}^*$) is considered as the restriction to plane π (resp. π^*) of the 3-D world space, it follows that the distance between $\mathtt{T}$ and $\mathtt{B}$ is:

$$\| \mathtt{T} - \mathtt{B} \| = Z_{\mathtt{T}} \text{ whereas } \begin{bmatrix} 0 \\ sZ_{\mathtt{T}} \\ s \end{bmatrix} = (\mathbf{H}^*)^{-1} \begin{bmatrix} \mathbf{t} \\ 1 \end{bmatrix} \tag{1}$$

If $\mathbf{H}$ is written in the form of $\mathbf{H} = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{b}]$ then, from the hypothesis about the 3-D world coordinate system, it follows that we have $\mathbf{H}^* = [\mathbf{v}_1, \mathbf{v}_3^*, \mathbf{b}]$, where column vector $\mathbf{v}_3^*$ contains three unknowns quantities. It is easy to show that the columns of matrix $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3^*]$ encode the image homogeneous coordinates of mutually conjugated (vanishing) points with respect to the imaged absolute conic ω. Let the imaged absolute conic matrix be written as:

$$\omega = \mathbf{V}^{-T}\mathbf{V}^{-1} \tag{2}$$

We demonstrate that, under the assumption of a *natural camera model* (*ie* a zero-skew camera with known aspect ratio), the knowledge of the two vectors $\mathbf{v}_1, \mathbf{v}_2$ yields two independent constraints on $\mathbf{v}_3^*$.

The zero-skew and known aspect ratio assumption. If the camera model has a zero-skew and a known aspect ratio κ, it follows that the circular points[8] have image complex projective coordinates $[\kappa, \pm i, 0]^T$. The circular points lie on the imaged absolute conic ω written as:

$$\omega = \begin{bmatrix} q_1 & q_2 & q_3 \\ q_2 & q_4 & q_5 \\ q_3 & q_5 & 1 \end{bmatrix} \text{ and it follows } \begin{cases} q_2 & = 0 \\ q_1 & = \kappa^2 q_4 \end{cases} \tag{3}$$

Let $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3^*]$ be written in the form of:

$$\mathbf{V} = \begin{bmatrix} \alpha u_1 & \beta u_2 & \gamma u_3 \\ \alpha v_1 & \beta v_2 & \gamma v_3 \\ \alpha & \beta & \gamma \end{bmatrix} \tag{4}$$

where (u_i, v_i) are the image coordinates of the vanishing points and where $[\alpha, \beta, \gamma]$ is an homogeneous vector which represents two degrees of freedom. Substituting (4) in (2) we obtain, considering unknowns u_3, v_3, γ, the solution of system (3) as the equation $L_1 u + L_2 v + L_3 = 0$ of a straight line, called *central line* $\mathcal{L}$, which holds for the vanishing point $(u = u_3, v = v_3)$:

$$\begin{aligned} L_1 &= \kappa^2(\alpha^2 + 1)(u_2 - u_1) \\ L_2 &= (\alpha^2 + 1)(v_2 - v_1) \\ L_3 &= -\kappa^2 u_2^2 - v_2^2 + \alpha^2(\kappa^2 u_1^2 + v_1^2) + (1 - \alpha^2)(\kappa^2 u_1 u_2 + v_1 v_2) \end{aligned} \tag{5}$$

It may be easily verified that:

- the principal point of the image plane lies on $\mathcal{L}$,
- the property exhibited by Caprile and Torre in [1], saying that the principal point is the orthocentre of the triangle with vanishing points as vertices, only holds if the image coordinate system is orthonormal, *ie* with a unit aspect ratio κ and a zero skew. Indeed, when $\kappa = 1$, central line $\mathcal{L}$ is orthogonal to the vanishing line of π since, in this case, line $\mathcal{L}$ has normal vector $[v_1 - v_2, u_2 - u_1]^T$ that is orthogonal to $[L_1, L_2]^T$.

On the other hand, the solution of the system (3) simultaneously yields:

$$\gamma = \xi(u_1 - u_3)\alpha\kappa \tag{6}$$

where $\xi = \pm\sqrt{\frac{-c_2}{c_1}}$ $(c_1 \neq 0)$ are the two roots of quadratic equation $c_1\xi^2 + c_2 = 0$ with:

$$\begin{aligned} c_1 &= a^4(v_1 - v_2)^2 + \\ &\quad a^2(2v_2^2 - 2v_1v_2 - 2v_2v3 + 2v_1v_3 - \kappa u_3^2 - \kappa^2 u_1^2 + 2\kappa^2 u_1 u_3) + (v_2 - v_3)^2 \\ c_2 &= -(1 + \alpha^2) < 0 \end{aligned}$$

The case in which there is a real solution for ξ (*ie* $c_1 > 0$) gives a constraint on the domain of value of v_3.

Let us go back to our problem of computations of person heights: we can consider that the selected top and bottom points $\mathbf{t}$ and $\mathbf{b}$ in the image provide a line in direction of the sought-after vanishing point $\mathbf{v}_3^*$. Central line $\mathcal{L}$ intersects this line at coordinates (u_3, v_3) and it follows that $\mathbf{H}^*$ can be computed explicitly. Our method differs from [5], since no reference length and (additional) direction are required. Nevertheless, this method is not of great accuracy because, in soccer games, a player is often in motion and his vertical axis is not the direction of $\mathbf{v}_3^*$. A new constraint is next given. It relies on the fact that, in broadcast TV soccer images, the horizontal U-axis of the image plane is always almost parallel to the ground plane π.

Constraint of parallelism between the camera U-axis and the ground plane. An interesting property of central line $\mathcal{L}$ is that, when $\mathbf{v}_1 = [u_\infty, v_\infty, 0]^T$ represents a point at infinity, $\mathcal{L}$ merges into the vanishing line crossing $\mathbf{v}_2$ and $\mathbf{v}_3^*$. This property holds as soon as the U-axis of the image plane is parallel to the X-axis of the world frame. In practice, it is equivalent to a constraint of parallelism between the U-axis and the ground plane $Z = 0$, since it is always feasible to apply a rotation to the world frame such as the world X-axis becomes parallel to image U-axis. Suppose we happen to be in this configuration, remarking that $\mathcal{L}$ is orthogonal to U-axis (see Figure 1), and let matrix $\mathbf{V}$ be written in the (new) form of:

$$\mathbf{V} = \begin{bmatrix} u_\infty & \beta u_2 & \gamma u_3 \\ v_\infty & \beta v_2 & \gamma v_3 \\ 0 & \beta & \gamma \end{bmatrix} \qquad (7)$$

By solving the system (3), we obtain the new parametrization of the central line $\mathcal{L}$:

$$\begin{aligned} [L_1, L_2, L_3] &= [\kappa^2 u_\infty, v_\infty, -(v_\infty v_2 + \kappa^2 u_\infty u_2)] \\ \gamma &= \xi v_\infty \beta \end{aligned}$$

where $\pm\xi$ are the two roots of the quadratic equation:

$$(\kappa^2 \beta^2 u_3^2 + \kappa^2 \beta^2 u_2^2 - 2\kappa^2 \beta^2 u_2 u_3 - v_\infty^2)\xi^2 - 1 = 0 \quad (8)$$

From this parametrization, it is easy to verify that $\mathcal{L}$ contains both vanishing points $\mathbf{v}_2$ and $\mathbf{v}_3^*$. Unfortunately, in this configuration, there is no real solution for $\mathbf{v}_3^*$. Indeed, since $\mathcal{L}$ is orthogonal to the U-axis, we have $L_2 = v_\infty = 0$ but also $u_2 = u_3$ and equation (8) has no real solution.

In practice, in configurations where $\mathbf{v}_1$ "tends towards" $[u_\infty, v_\infty, 0]^T$, $\mathbf{v}_3^*$ can be computed as the intersection between central line $\mathcal{L}$, imaged from the "old" world frame, and line $u = u_2'$, imaged from the "new" world frame. The matrix $\mathbf{H}^*$ is actually recovered up to an infinitesimal 3-D rotation around the point which has coordinates $\mathbf{A}^{-1}[u_3, v_3, 1]^T$ in the camera normalized frame, where $\mathbf{A}$ is the camera internal parameter matrix. This is is illustrated in Figure 1. This gives a simple algorithm, given the homography matrix $\mathbf{H} = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{b}]$:

1. compute $[L_1, L_2, L_3]^T = \mathcal{L}(\mathbf{v}_1, \mathbf{v}_2)$ as explained in (5),
2. compute $\mathbf{H}' \leftarrow \mathbf{H}\mathbf{R}'$, where $\mathbf{R}'$ is the rotation around Z in the world space such that the image of the new X'-axis is parallel to U-axis in the image frame,
3. compute the "new" u_2 from $\mathbf{H}'$ and denote it u_2',
4. compute $(u_3 = u_2', v_3 = \frac{-(L_1 u_2' + L_3)}{L_2})$ and γ as described in (6),
5. compute matrix $\mathbf{H}^* = [\mathbf{v}_1, \mathbf{v}_3^*, \mathbf{b}]$,
6. compute the world space to image projection matrix $\mathbf{P} = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3^*, \mathbf{b}]$.

The implementation of this algorithm turned out to be highly satisfactory, in particular with TV soccer images.

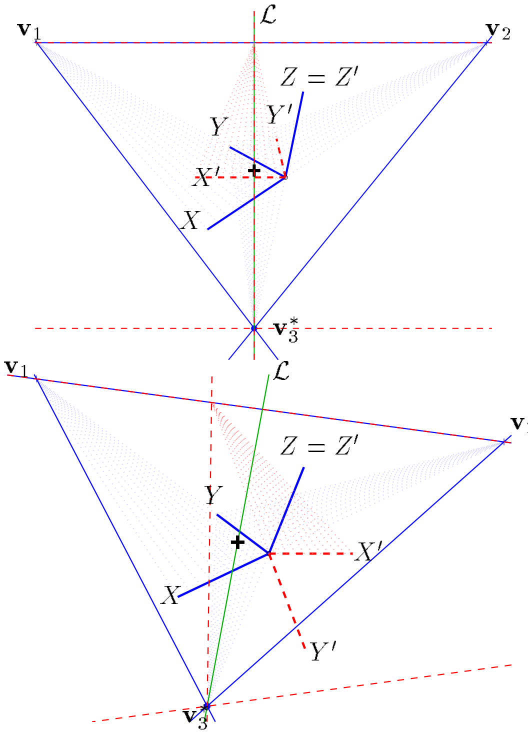

Figure 1. Both figures represent the images of the world frame and the associated vanishing points $\mathbf{v}_i$ $(i = 1, 2, 3)$. In the former, the U-axis of the image plane is parallel to the ground plane $Z = 0$. There exists an appropriate rotation of the world frame such as the new triangle ,with (dotted) vanishing lines as sides, is degenerated ; central line $\mathcal{L}$ merging into the vertical side. In the latter figure, when the U-axis of the image plane is not strictly parallel to the ground plane, the vanishing point $\mathbf{v}_3^*$ is the intersection between line $\mathcal{L}$ and line defined by the vertical side.

3. Application

Our algorithm was applied to very low-quality mpeg images (with dimensions around 350×220) taken from three different TV broadcast soccer matches (Figure 2). The model of camera was supposed to be natural, with zero-skew and unit aspect ratio. The field of play is the reference ground plane with equation $Z = 0$. We have preliminary computed the matrix $\mathbf{H}$ from the lines dotted in black. The goal posts and crossbar, even if present, have not been used as input. They should be considered as verification data and remained unknown to our algorithm. Firstly, we have computed the distance from the lower edge of the goal crossbar to the ground, indicated by a black arrow. The algorithm provided, respectively, 2.48m, 2.38m and 2.49m. Considering the low-quality of images, the results seem quite accurate since the actual height is 2.44m according to the FIFA's law of the game. In the second place, we have recovered the projection matrix $\mathbf{P}$ from $\mathbf{H}^*$ and $\mathbf{H}$. It was henceforth possible to insert virtual 3-D objects in the image, like the (re-drawn) posts and crossbar. In addition, a "fake" advertising board with "IRIT" has also been inserted.

4. Conclusion

We have developed a method which recovers the metric structure of a perpendicular plane π^* to the ground from the ground to image homography matrix $\mathbf{H}$. The pre-conditions of our algorithm are the two related to the common natural camera model (zero-skew and known aspect ratio) and a "pseudo" parallelism between of one of the camera axis and the ground. This condition comes from the applications we deal with (automatic soccer analysis) regarding the technique used by TV broadcast for shooting soccer images. Once the π^*-plane to image homography matrix $\mathbf{H}^*$ recovered, we are able to construct the world space to image matrix $\mathbf{P}$. Direct applications have been tested on video images in order to compute distances orthogonal to the ground and to insert virtual 3D objects in the image. We have noticed a surprising accuracy for such low-quality mpeg images.

References

[1] B. Caprile and V. Torre. Using vanishing points for camera calibration. *IJCV*, 4(2):127–140, March 1990.

[2] A. Criminisi, I. Reid, and A. Zisserman. Single view metrology. In *Proc. 7th International Conference on Computer Vision, Kerkyra, Greece*, pages 434–442, September 1999.

[3] O. Faugeras. *Three-dimensional Computer Vision : a geometric viewpoint*. MIT Press, Boston, USA, 1993.

[4] P. Gurdjos, P. Dalle, and S. Castan. Tracking 3d coplanar points in the invariant perspective coordinates plane. In *Proceedings of the 13th International Conference on Pattern Recognition (ICPR'96)*, pages 493–497, Vienna, Austria, August 1996.

[5] D. Liebowitz, A. Criminisi, and A. Zisserman. Creating architectural models from images. In *Proc. EuroGraphics*, volume 18, pages 39–50, September 1999.

[6] D. Liebowitz and A. Zisserman. Metric rectification for perspective images of planes. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 482–488, June 1998.

[7] D. Liebowitz and A. Zisserman. Combining scene and auto-calibration constraints. In *Proc. 7th International Conference on Computer Vision, Kerkyra, Greece*, September 1999.

[8] J. Semple and G. Kneebone. *Algebraic Projective Geometry*. Oxford University Press, 1979.

[9] K. Taeone, A. Yongduek, and K. Hong. Physics-based 3d position analysis of a soccer ball from monocular image sequences. In *Proceedings of the International Conference on Computer Vision. Bombay, India*, pages 721–721, January 1998.

Figure 2. Given the ground to image homography, our algorithm carries out the computation of distances orthogonal to the ground and the insertion of virtual 3D objects. It is applied to images from three different TV broadcast soccer matches. Regarding the former task, the height of the upright post, indicated by a black arrow, is measured. It yields, respectively 2.48m, 2.38m and 2.49m whereas the actual distance from the lower edge of the goal crossbar to the ground is 2.44m. Regarding the the latter task, we "re-paint" in yellow the goal crossbar and posts, and insert a "fake" advertising board "IRIT".