Fusing of image and inertial sensing for camera calibration *

Jorge Lobo† and Jorge Dias†

Institute of Systems and Robotics
Electrical Engineering Department
University of Coimbra
Coimbra, Portugal.

Abstract

This paper explores the integration of inertial sensor data with vision. A method is proposed for the estimation of camera focal distance based on vanishing points and inertial sensors.

Visual and inertial sensing are two sensory modalities that can be explored to give robust solutions on image segmentation and recovering of 3D structure from images, increasing the capabilities of autonomous vehicles and enlarging the application potential of vision systems.

In this paper we show that using just one vanishing point, obtained from two parallel lines belonging to some levelled plane, and using the cameras attitude taken from the inertial sensors, the unknown scaling factor f in the camera's perspective projection can be estimated. The quality of the estimation of f depends on the quality of the vanishing point used and the noise level in the accelerometer data. Nevertheless it provides a reasonable estimate for a completely uncalibrated camera. The advantage over using two vanishing points is that the best (i.e. more stable) vanishing point can be chosen, and that in indoors environment the vanishing point point can sometimes be obtained from the scene without placing any specific calibration target.

1 Introduction

In the paper we show how fusing information from inertial sensors with image data can be used for camera calibration.

Inertial sensors explore intrinsic properties of body motion. From the principle of generalised relativity of Einstein we known that only the specific force on one point and the angular instantaneous velocity, but no other quantity concerning motion and orientation with respect to the rest of the universe, can be measured from physical experiments inside an isolated closed system. Therefore from inertial measurements one can only determine an estimate for linear acceleration and angular velocity. Linear velocity and position, and angular position, can be obtained by integration. Inertial navigation systems implement this process of obtaining velocity and position information from inertial sensor measurements. Internal sensing using inertial sensors is very useful in autonomous robotic systems since it is not dependent on any external references, except for the gravity field which does provide an external reference.

In human and other mammals the vestibular system in the inner ear gives inertial information essential for navigation, orientation, body posture control and equilibrium. In humans this sensorial system is crucial for several visual tasks and head stabilisation. It is well known that, in humans, the information provided by the vestibular system is used during the execution of visual movements such as gaze holding and tracking, as described by Carpenter [1].

Camera calibration using vanishing points has been widely explored, [2] [3] [4] [5] [6] amongst others. The novelty in this work is using just one vanishing point, and using the inertial sensors to extract camera pose information. Calibration based on vanishing points is limited since a compromise has to be reached on the quality of each point, but since we require just one vanishing point, the best one can be chosen.

The rest of this paper is arranged as follows. Section 2 describes the system's geometric framework. The geometric features extracted from the inertial sensor data, the vertical and the horizon are discussed in section 3. The estimation of camera focal distance based and vanishing point detection is presented in section 4. The experimental setup used and some results are presented in Section 6 and concluding re-

*Project financed by PRAXIS /P/EEI 11218/98

†{jlobo,jorge}@isr.uc.pt

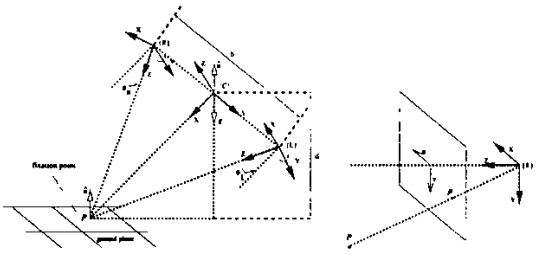

Figure 1: System geometry, and the camera referential.

marks drawn in section 7.

2 Geometric Framework

2.1 System Geometry

While some initial tests were made using a single camera with inertial sensors, we are now using a stereo system to explore other possibilities for the integration of inertial sensor data with vision. The system used has an inertial unit at the middle of the stereo camera baseline, as seen in figure 4. The cameras' pan is controlled so as to have a symmetric verge angle. The system's coordinate frame referential, $\{\mathcal{C}\}$, is defined as having the origin at the centre of the baseline of the stereo cameras, as seen in figure 1. This referential is named *Cyclop*, after the mythological one-eyed giant.

Each camera has its own referential, $\{\mathcal{R}\}$ and $\{\mathcal{L}\}$. Besides being translated from $\{\mathcal{C}\}$ along the baseline (represented by the homogeneous coordinate matrixes $\boldsymbol{T_r}$ and $\boldsymbol{T_l}$), and rotated θ_R and θ_L along the $\{\mathcal{C}\}$ Z-axis ($\boldsymbol{R_r}$ and $\boldsymbol{R_l}$), their axis are also swapped to comply with the typical referential convention used for camera images ($\boldsymbol{S_r}$ and $\boldsymbol{S_l}$) - see figure 1. Notice that in our case we have symmetric vergence, i.e. $\theta = \theta_R = -\theta_L$.

Using these conventions we can express a world point $\boldsymbol{P}$ in the Cyclop referential $\{\mathcal{C}\}$, given its coordinates in the camera referential $\{\mathcal{R}\}$ or $\{\mathcal{L}\}$ by

$$^{\mathcal{C}}\boldsymbol{P} = \boldsymbol{T_r}.\boldsymbol{R_r}.\boldsymbol{S_r}.^{\mathcal{R}}\boldsymbol{P} = ^{\mathcal{C}}\boldsymbol{T_{\mathcal{R}}}.^{\mathcal{R}}\boldsymbol{P} \tag{1}$$

and

$$^{\mathcal{C}}\boldsymbol{P} = \boldsymbol{T_l}.\boldsymbol{R_l}.\boldsymbol{S_l}.^{\mathcal{L}}\boldsymbol{P} = ^{\mathcal{C}}\boldsymbol{T_{\mathcal{L}}}.^{\mathcal{L}}\boldsymbol{P} \tag{2}$$

2.2 Camera Model

The pinhole camera model is used. Assuming that image acquisition maintains square pixel ratio and no skew we have for an image point $\boldsymbol{p_i} = (u, v)$

$$u = f\frac{X}{Z} \qquad v = f\frac{Y}{Z} \tag{3}$$

where u and v are the pixel coordinates with origin at the image centre as show in figure 1, f is the camera effective focal distance (i.e. includes a scale factor) and $\boldsymbol{P} = (X, Y, Z)^{\mathsf{T}}$ is in the camera referential.

2.3 Projection onto Unit Sphere

The previous model derives from the cameras' geometry, but the projection needn't be onto a plane. Consider a unit sphere around the optical centre, with the images being formed on its surface. The image plane can be seen as a plane tangent to a sphere of radius f concentric with the unit sphere. The image plane touches the sphere at the equator, and this point defines, on the image plane, the image centre. Using the unit sphere gives a more general model for central perspective. It also has numerical advantages when considering points at infinity. Consider the unit sphere where every world point $\boldsymbol{P}$ in the world is projected, forming an image on its surface. The image points on this surface can be represented by unit vectors $\boldsymbol{m}$ placed at the sphere's centre, the optical centre of the camera.

$$\boldsymbol{P} \rightarrow \boldsymbol{m} = \frac{\boldsymbol{P}}{||\boldsymbol{P}||} \tag{4}$$

Note that $\boldsymbol{m} = (m_1, m_2, m_3)^{\mathsf{T}}$ is a unit vector and the projection is not defined for $\boldsymbol{P} = (0, 0, 0)^{\mathsf{T}}$ [2]. Projection onto the unit sphere is related to projection onto a plane by

$$(u, v)^{\mathsf{T}} = \left(f\frac{m_1}{m_3}, f\frac{m_2}{m_3}\right)^{\mathsf{T}} \tag{5}$$

Given f, the projection to a sphere can be computed from the projection to a plane and conversely. To avoid ambiguity m_3 is forced to be positive, so that only points on the image side hemisphere are considered.

For a given image point (u, v), its projection onto the unit sphere is given by

$$\boldsymbol{m} = \frac{1}{\sqrt{u^2 + v^2 + f^2}} \begin{bmatrix} u \\ v \\ f \end{bmatrix} \tag{6}$$

Image lines can also be represented in a similar way. Any image line defines a plane with the centre of projection. A vector normal to this plane uniquely defines the image line. A normalized vector, normal to the plane defined by the image line and the optical centre, can be used to represent the line.

For a given image line $au + bv + c = 0$, the normalized vector is given by

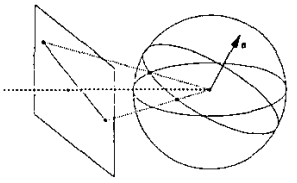

Figure 2: Line projection onto Unit Sphere.

$$\boldsymbol{n} = \frac{1}{\sqrt{a^2 + b^2 + (c/f)^2}} \begin{bmatrix} a \\ b \\ c/f \end{bmatrix} \quad (7)$$

A given set of image plane points are collinear if their corresponding unit sphere vectors, $\boldsymbol{m_i}$, all lie within the same plane, touching the unit sphere along a great circle. All $\boldsymbol{m_i}$ are orthogonal to the unit vector $\boldsymbol{n}$ that defines the line passing through them all, i.e.

$$\boldsymbol{n}.\boldsymbol{m_i} = 0 \quad (8)$$

From the duality of points and lines it follows that a given set of image lines are concurrent if their vectors, $\boldsymbol{n_i}$, all lie within the same plane, all $\boldsymbol{n_i}$ are orthogonal to the unit vector $\boldsymbol{m}$ that defines their common point, i.e.

$$\boldsymbol{m}.\boldsymbol{n_i} = 0 \quad (9)$$

For both cases, the mentioned plane is defined by the unit sphere centre (centre of projection) and the normal.

Image points $\boldsymbol{m}$ and $\boldsymbol{m'}$ are said to be conjugate to each other if

$$\boldsymbol{m}.\boldsymbol{m'} = 0 \quad (10)$$

and image lines $\boldsymbol{n}$ and $\boldsymbol{n'}$ are conjugate to each other if

$$\boldsymbol{n}.\boldsymbol{n'} = 0 \quad (11)$$

In image coordinates we have that image points $(u, v)^\mathsf{T}$ and $(u', v')^\mathsf{T}$ are conjugate to each other if

$$uu' + vv' + f^2 = 0 \quad (12)$$

and the projective lines passing through each point and the center of projection are orthogonal.

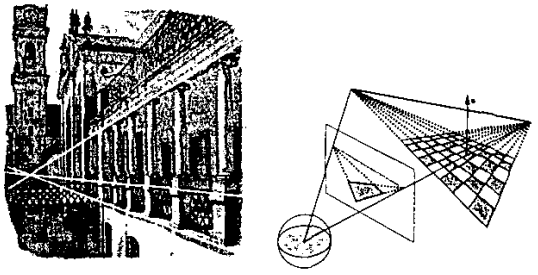

Figure 3: *Via Latina* of Coimbra University with vanishing point, and vanishing line of planar surface.

2.4 Vanishing points and vanishing lines

Since the perspective projection maps a 3D world onto a plane or planar surface, phenomena that only occurs *at infinity* will project to very finite locations in the image. Parallel lines only meet at infinity, but as seen in figure 3, the point where they meet can be quite visible and is called the *vanishing point* of that set of parallel lines.

A space line with the orientation of an unit vector $\boldsymbol{m}$ has, when projected, a *vanishing point* with unit sphere vector $\pm\boldsymbol{m}$. Since the vanishing point is only determined by the 3D orientation of the space line, projections of parallel space lines intersect at a common vanishing point.

A planar surface with a unit normal vector $\boldsymbol{n}$, not parallel to the image plane has, when projected, a *vanishing line* with unit sphere vector $\pm\boldsymbol{n}$. Since the vanishing line is determined alone by the orientation of the planar surface, then the projections of planar surfaces parallel in the scene define a common vanishing line. A vanishing line is a set of all vanishing points corresponding to the lines that belong to the set of parallel planes defining the vanishing line.

In an image the horizon can be found by having two distinct vanishing points as seen in figure 3. With a suitable calibration target (e.g. a levelled square with well defined edges) the horizon can be determined.

If the vanishing points, $(u, v)^\mathsf{T}$ and $(u', v')^\mathsf{T}$, correspond to orthogonal sets of parallel lines, they are conjugate to each other and from (12) we have

$$f = \sqrt{-uu' - vv'} \quad (13)$$

3 Gravity vector gives image horizon, vertical and ground plane

3.1 Gravity vector

The measurements $\boldsymbol{a}$ taken by the inertial unit's accelerometers include the sensed gravity vector $\boldsymbol{g}$

summed with the body's acceleration a_b:

$$a = g + a_b \tag{14}$$

Assuming the system is motionless, then $a_b = 0$ and the measured acceleration $a = g$ gives the gravity vector in the system's referential. So, with a_x,a_y and a_z being the accelerometer filtered measurements along each axis, the vertical unit vector will be given by

$$\hat{n} = -\frac{g}{\|g\|} = \frac{1}{\sqrt{a_x^2+a_y^2+a_z^2}} \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} = \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} \tag{15}$$

and

$${}^{c}\hat{n} = \begin{bmatrix} n_x & n_y & n_z & 1 \end{bmatrix}^{\mathsf{T}} \tag{16}$$

Notice that if our assumption of the system being motionless or subject to constant speed is correct, than in the above equation

$$\sqrt{a_x^2 + a_y^2 + a_z^2} \cong 9.8ms^{-2} \tag{17}$$

and this condition can be tested and monitored by the system.

3.2 Vertical

In equation (16) the vertical unit vector is given in the Cyclop referential. The vertical for each of the cameras is given by

$${}^{r}\hat{n} = \boldsymbol{S}_r^{-1}.\boldsymbol{R}_r^{-1}.{}^{c}\hat{n} \quad \text{and} \quad {}^{l}\hat{\mathrm{n}} = \mathrm{S}_l^{-1}.\mathrm{R}_l^{-1}.{}^{c}\hat{\mathrm{n}} \tag{18}$$

This vertical corresponds to the vanishing point of all vertical lines.

3.3 Horizon

In the previous section we saw how the horizon can be found by having two distinct vanishing points. Knowing the vertical in the cameras referential and the focal distance, an artificial horizon also can be traced. A planar surface with a unit normal vector $\hat{n}$, not parallel to the image plane has, when projected, a *vanishing line* given by

$$n_x u + n_y v + n_z f = 0 \tag{19}$$

where f is the focal distance, u and v image coordinates and $\hat{n} = (n_x, n_y, n_z)^{\mathsf{T}}$

Since the vanishing line is determined alone by the orientation of the planar surface, then the projections of planar surfaces parallel in the scene define a common vanishing line. The horizon is the vanishing line of all levelled planes, parallel to the ground plane.

4 Calibration of f

With one vanishing point $\boldsymbol{p}_v = (u, v)^{\mathsf{T}}$, obtained from two parallel lines belonging to some levelled plane, and from equation (19) the unknown scaling factor f in equation (3) can be estimated as

$$f = -\frac{n_x u + n_y v}{n_z} \tag{20}$$

where $\hat{n} = (n_x, n_y, n_z)^{\mathsf{T}}$ is taken from (18)

4.1 Determining the Vanishing Points

Using an image with dominant ground plane parallel lines, the lines have to be detected so that the vanishing points can be found.

The edges in the image are found with an optimized Sobel filter. The filter estimates the gradient $\mathcal{D}$ as

$$\mathcal{D} = \begin{bmatrix} \mathcal{D}_x \\ \mathcal{D}_y \end{bmatrix} \tag{21}$$

where $\mathcal{D}_x$ and $\mathcal{D}_x$ are pixel mask operators given by

$$\mathcal{D}_x \approx \frac{1}{32}\begin{bmatrix} 3 & 0 & -3 \\ 10 & 0 & -10 \\ 3 & 0 & -3 \end{bmatrix} \quad \mathcal{D}_y \approx \frac{1}{32}\begin{bmatrix} 3 & 10 & 3 \\ 0 & 0 & 0 \\ -3 & -10 & -3 \end{bmatrix} \tag{22}$$

The optimized Sobel filter has a lower angle error than the standard Sobel filter [7]. By choosing an appropriate threshold for the gradient magnitude, the potential edge lines can be identified.

$$\mathcal{D} = \sqrt{\mathcal{D}_x^2 + \mathcal{D}_y^2} > threshold \tag{23}$$

The Hough transform is used to group the edge points into lines. Since the Sobel filter provides a local edge orientation estimate, we can use an orientation-based fast Hough transform that avoids the high computational effort of the parameter space tranform [7].

The Hough transform maps image points $\boldsymbol{p}_i = (u_i, v_i)^{\mathsf{T}}$ to parameter space $= (d, \theta)^{\mathsf{T}}$ where

$$u_i \cos\theta + v_i \sin\theta = d \tag{24}$$

The fast Hough transform uses the edge orientation given by the Sobel filter. Since the local edge orientation is noisy, a small neighbourhood bell shaped mask is used when voting in the transform space. The magnitude of the gradient is also used so that stronger edges have a higher weighting factor in the accumulator parameter space.

The parameter space is than parsed to find the highest peaks, corresponding to the image dominant lines. Having identified the line, a neighbourhood of this peak is zeroed so that the second line can be found, and the process is iterated to find all four lines

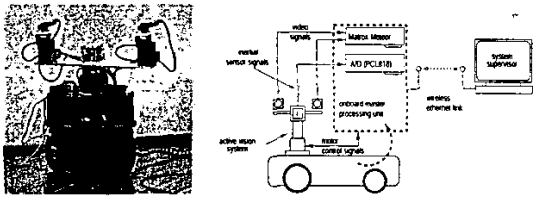

Figure 4: System on Mobile Robot, and System Architecture.

that must be dominant in the image. The lines are than sorted by angle, so that they can be paired and the vanishing points found. We used two sets of lines to compare our method with others, but only one vanishing point is needed. Having multiple vanishing points, the best one can be chosen.

Our emphasis is on fast methods, and methods such as presented by Palmer [8], Lutton [9] and Tuytelaars [10] where not tested since the added accuracy might not improve the results due to the noise level in the accelerometers.

5 Results

5.1 Experimental Setup

For this work we needed a pair of cameras with a stereo rig capable of controlling camera vergence, and inertial sensors to measure the systems attitude.

An inertial system prototype built at our lab was used. The system is based on low-cost inertial sensors and is intended for robotic applications. The sensors used in the prototype system include a three-axial accelerometer, three gyroscopes and a dual-axis inclinometer. For this work we are only extracting the system's attitude from the accelerometer data when it is motionless, by keeping track of the gravity vector g. See [11] for complete details of the experimental setup.

To study the integration of inertial information and vision in artificial autonomous mobile systems, the system was mounted onto a mobile robot platform. Figure 4 shows the complete system.

5.2 Calibration of f

With this setup, the method for the estimation of f was tested. The calibration target shown in figure 5 was rotated so that , and 10 samples taken at each position.

As seen in figure 6 and table 1 using just one vanishing point and $\hat{n}$ provides the estimate with lower error.

Figure 7 and table 2 show the estimate results for 100 samples of a single position, favourable for the

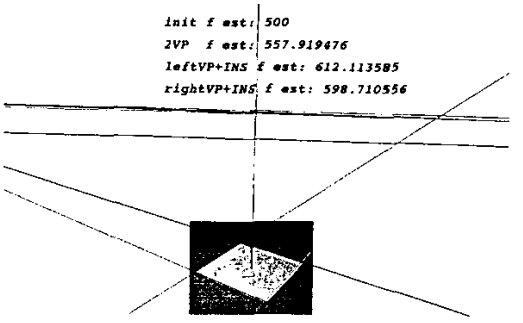

Figure 5: Target image and estimated values of f. The vanishing lines are shown, as well as the nearer vanishing point. The lower horizon is based on an initial estimate of f and n, and the others are based on the the left and right vanishing points and n.

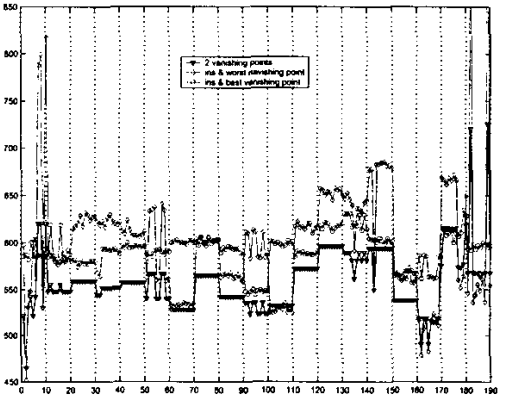

Figure 6: Estimation of f using vanishing points and $\hat{n}$, the target was rotated through 19 positions, with 10 samples taken at each position.

Table 1: Estimation of f using vanishing points and $\hat{n}$ with rotating target.

	mean	σ
2 vanishing points	558.94	32.11
ins & worst vanishing point	601.23	66.33
ins & best vanishing point	592.38	18.17

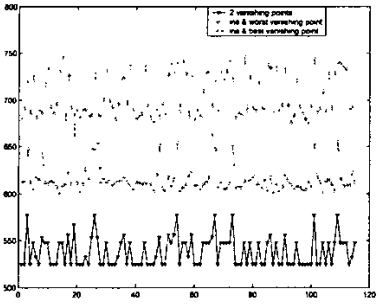

Figure 7: Estimation of f using vanishing points and $\hat{n}$, with 100 samples taken with fixed target.

Table 2: Estimation of f using vanishing points and $\hat{n}$ with fixed target.

	mean	σ
2 vanishing points	538.30	15.68
ins & worst vanishing point	704.50	22.46
ins & best vanishing point	615.80	14.90

two vanishing points. Although the vanishing points are both near the image, one is nearer and using this vanishing point and $\hat{n}$ still provides the estimate with lower error.

6 Conclusions

This article presented our recent results on the integration of inertial sensor data with vision, namely in the estimation of camera focal distance.

With just one vanishing point, obtained from two parallel lines belonging to some levelled plane, and using the cameras attitude taken from the inertial sensors, the unknown scaling factor f in the camera's perspective projection can be estimated. The quality of the estimation of f depends on the quality of the vanishing point used and the noise level in the accelerometer data. Nevertheless it provides a reasonable estimate for a completely uncalibrated camera. The advantage over using two vanishing points is that the best (i.e. more stable) vanishing point can be chosen, and that in indoors environment the vanishing point point can sometimes be obtained from the scene without placing any specific calibration target.

Future work involves improving the vanishing point detection, exploring the methods used by Palmer [8], Lutton [9] and Tuytelaars [10], so that our single vanishing point approach can benefit. Ongoing work is being done in statistical error models and sensitivity analysis, as well as tests with diverse indoor scenes.

References

[1] H. Carpenter. *Movements of the Eyes.* London Pion Limited, 2nd edition, 1988. ISBN 0-85086-109-8.

[2] Kenichi Kanatani. *Geometric Computation for Machine Vision.* Oxford University Press, 1993. ISBN 0-19-856385-X.

[3] Ling-Ling Wang and Wen-Hsiang Tsai. Camera Calibration by Vanishing Lines for 3-D Computer Vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4):370–376, April 1991.

[4] B. Caprile and V. Torre. Using Vanishing Points for Camera Calibration. *International Journal of Computer Vision*, 4(2):127–140, 1990.

[5] B. Brillault and O'Mahony. New Method for Vanishing Point Detection. *CVGIP: Image Understanding*, 54(2):289–300, 1991.

[6] M. Li. Camera calibration of the kth head-eye system. In *ECCV94*, pages A:543–554, 1994.

[7] Bernd Jahne. *Digital Image Processing.* Springer-Verlag, 1997. ISBN 3-540-62724-3.

[8] P. Palmer, J. Kittler, and M. Petrou. Accurate Line Parameters from an Optimising Hough Transform for Vanishing Point Detection. In *Proceedings of the Fourth Inetrnational Conference on Computer Vision*, pages 529–533, Berlin, Germany, May 1993.

[9] E. Lutton, H. Ma itre, and J. Lopez-Krahe. Contribution to the Determination of Vanishing Points Using Hough Transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(4):430–438, April 1994.

[10] T. Tuytelaars, M. Proesmans, and L. van Gool. The Cascaded Hough Transform. In *ICIP Proceedings*, pages 736–739, 1998.

[11] Jorge Lobo and Jorge Dias. Integration of Inertial Information with Vision towards Robot Autonomy. In *Proceedings of the 1997 IEEE International Symposium on Industrial Electronics*, pages 825–830, Guimaraes, Portugal, July 1997.