Metrology from Vertical Objects

Xiaochun Cao and Hassan Foroosh
School of Computer Science
University of Central Florida
Orlando, FL, 32816-3262
{xccao, foroosh}@cs.ucf.edu

Abstract

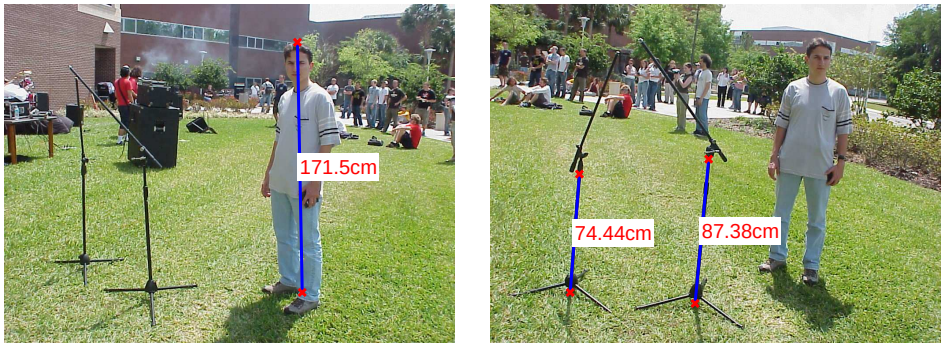
In this paper, we describe how 3D Euclidean measurements can be made in a pair of perspective images, when only minimal geometric information are available in the image planes. This minimal information consists of one line on a reference plane and one vanishing point for a direction perpendicular to the plane. Given these information, we show that the length ratio of two objects perpendicular to the reference plane can be expressed as a function of the camera principal point. Assuming that the camera intrinsic parameters remain invariant between the two views, we recover the principal point and the camera focal length by minimizing the symmetric transfer error of geometric distances. Euclidean metric measurements can then be made directly from the images. To demonstrate the effectiveness of the approach, we present the processing results for synthetic and natural images, including measurements along both parallel and non-parallel lines.

1 Introduction

Metrology from uncalibrated images is becoming of increasing interest for many applications. This problem is trivial if the camera is calibrated meaning that its intrinsic parameters and its position and orientation are known. Camera parameters can be obtained by using standard methods if a calibration object or measurements of enough 3D points in the scene are available [4], or alternatively using self-calibration methods from unstructured scenes [8, 6]. However, such measurements are not always available, and also self-calibration techniques typically require bootstraping from a projective reconstruction, often leading to solving complex non-linear problems that are typically ill-conditioned, and hence are not easily tractable, or may not always converge to the optimum solution.

Vanishing points or parallel lines have proven to be useful features for this task [1, 7, 2, 3, 5, 9]. In a seminal work, Criminisi et al. [3] proposed an approach for single view metrology, and showed that affine scene structure may be recovered from a single image without any prior knowledge of the camera calibration parameters. The limitation of their aproach is that they require that three mutually orthogonal vanishing points to be available simultaneously in the image plane. Also, in order to recover the metric measurements they require three reference distances. Their advantage however is that they need only one image to solve the problem. In contrast, our approach requires only one vanishing point along a vertical to a reference plane. However our method would

BMVC 2004 doi:10.5244/C.18.74



(a) (b)

Figure 1: Measuring height of vertical objects: (a) The standing person has known height, (b) Computed heights of two microphone stands.

require two images to solve the problem with only one reference distance. Examples of images where such scenario may apply are commonly encountered in indoor and outdoor environments, where there is a ground plane and some up-right objects, e.g. humans, street lamps, trees, etc., but not all vanishing points avialable, see for instance Figure 1. Note also that Criminisi et al. can only perform measurements in the reference plane and the planes parallel to it. In our approach, we can directly perform measurements outside the reference plane and along non-parallel lines.

Therefore, in this paper, we are interested in the problem of affine measurements i.e. the length ratios of parallel and even non-parallel line segments from possibly one or two uncalibrated images of a scene. We then extend the approach to metric measurements by either assuming that the principal point is known, or by minimizing the symmetric transfer errors of geometric distances. It is assumed that the images are obtained by perspective projections with constant intrinsic parameters, and negligible radial distortions, which otherwise can generally be removed [6]. The rest of the paper is organized as follows. In the next section, we present closed-form solutions for metric measurments under two different scenarios, assuming that the principal point is known. We then extend the results in section 3 to metric measurements when the principal point is also unknown and solve the problem by minimizing the symmetric transfer error of geometric distances. Section 4 describes the experimental results. Both computer simulation and real experimentations are used to validate the proposed approach. Finally, in section 5, we present some discussions and concluding remarks.

2 Closed-form Solution

In this section, we consider two different cases, where some objects perpendicular to a reference plane have been observed in two or more uncalibrated images. For instance, we will show that two upright objects standing on the ground plane are sufficient for computing their height ratio, as well as the ratio of the lengthes of other parallel or non-parallel objects in the scene. These affine measurements can be extended to metric measurements

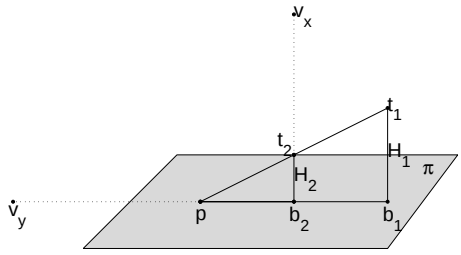


Figure 2: Two objects $\mathbf{b}_1\mathbf{t}_1$ and $\mathbf{b}_2\mathbf{t}_2$ are both perpendicular to the plane π. If the two objects have different height, the line connecting $\mathbf{t}_1$ and $\mathbf{t}_2$ will intersect with π at the point $\mathbf{p}$ which is collinear with b_1 and $\mathbf{b}_2$. The vanishing point $\mathbf{v}_y$ for the direction $\mathbf{b}_1\mathbf{b}_2$ and the three points $\mathbf{p}$, $\mathbf{b}_2$ and $\mathbf{b}_1$ define a cross ratio. The value of the cross ratio determines the ratio of lengths between the two vertical objects; see text.

if we assume that the principal point is known. In section 3 we shall relax this latter assumption, and provide a solution under a more general scenario. The basic geometry is shown in Figure 2, which consists of one line on a reference plane, and one vanishing point for a direction perpendicular to the plane. Let $\mathbf{v}_x$ be the vanishing point along the vertical direction, which is determined by intersecting the two vertical objects in the image plane. The line segment on the reference plane connects two base points $\mathbf{b}_1$ and $\mathbf{b}_2$ of the two vertical objects.

We will first consider in subsection 2.1 the spacial case where the two vertical objects have the same heights. This is equivalent to assuming that two vanishing points are known. One example is shown in figure 4. Although, only the ratio of the two vertical objects is known, we show that it is also possible to do measurements along directions other than perpendicular line segments and outside the reference plane. This may be done given either a single image and two reference distances, or alternatively given only one reference distance if an extra view is available. We then extend this idea in subsection 2.2 to the more general case, where the two vertical objects have different heights, i.e. when only one vanishing point is known.

2.1 Measurement from Two Orthogonal Vanishing Points

Let $\mathbf{v}_x$, $\mathbf{v}_y$, and $\mathbf{v}_z$ denote the three mutually orthogonal vanishing points. As is well known, for a unit aspect ratio and zero skew, the principal point $\mathbf{c}$ is the orthocenter of the triangle with vertices at $\mathbf{v}_x$, $\mathbf{v}_y$, and $\mathbf{v}_z$ [1] (see Figure 3). In the special case of two vertical objects with the same height, the points $\mathbf{p}$ and $\mathbf{v}_y$ in Figure 2 coincide. Therefore, we have two known vanishing points, $\mathbf{v}_x$ and $\mathbf{v}_y$, and one unknown one, i.e. $\mathbf{v}_z$. Metrology on images with three known vanishing points and under a general perspective camera model are described in detail by Criminisi, Reid and Zisserman in [3].

When the third vanishing point is unknown and only one view is available, we need either two reference distances to measure up to only a rigid ambiguity, or the ratio of the object height to $\mathbf{b}_1\mathbf{b}_2$ in Figure 2, leaving a scale ambiguity. Either information can be used to determine the position of the two vertical objects, and hence the 3D coordinates of the four end points. For this purpose, note that the 3×3 planar homography $\mathbf{H}$, which

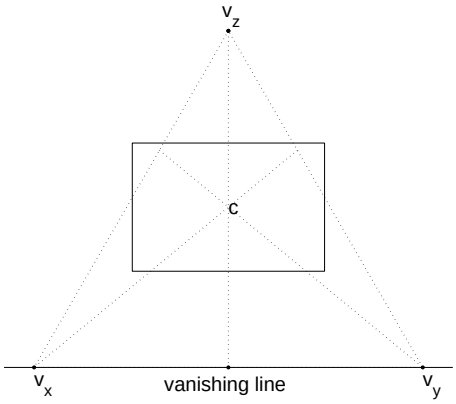


Figure 3: Principal point **c** is the orthocenter of the three vanishing points of mutually orthogonal directions.

maps the world plane passing through the two vertical objects to the image plane, can be computed as described below. Assuming a zero skew and unit aspect ratio,

$$\mathbf{H} \sim \begin{bmatrix} fr_{11}+u_0r_{31} & fr_{12}+u_0r_{32} & ft_x+u_0t_z \\ fr_{21}+v_0r_{31} & fr_{22}+v_0r_{32} & ft_y+v_0t_z \\ r_{31} & r_{32} & t_z \end{bmatrix} \tag{1}$$

where f is the camera focal length, $(u_0 \;\; v_0)$ is the principal point, t_x, t_y, and t_z are the components of the translation vector, and r_{ij} are the components of the rotation matrix $\mathbf{R} = \mathbf{R}_z\mathbf{R}_y\mathbf{R}_x$ given by

$$r_{11} = \cos(\theta_y)\cos(\theta_z) \tag{2}$$
$$r_{21} = \cos(\theta_y)\sin(\theta_z) \tag{3}$$
$$r_{31} = -\sin(\theta_y) \tag{4}$$
$$r_{12} = \sin(\theta_x)\sin(\theta_y)\cos(\theta_z) - \cos(\theta_x)\sin(\theta_z) \tag{5}$$
$$r_{22} = \sin(\theta_x)\sin(\theta_y)\sin(\theta_z) + \cos(\theta_x)\cos(\theta_z) \tag{6}$$
$$r_{32} = \cos(\theta_y)\sin(\theta_x) \tag{7}$$

where the pan angle θ_x, tilt angle θ_y, and yaw angle θ_z describe the rotation between the world coordinate system and the camera coordinate system. Note that both coordinate systems are up to a metric ambiguity, which in general is of no significant concern in measurement and reconstruction.

Since the principal point **c** is the orthocenter of the three vanishing points [1], we have

$$(\mathbf{v}_x - \mathbf{c})^T(\mathbf{v}_y - \mathbf{c}) + f^2 = 0 \tag{8}$$

In other words, the focal length depends only on the principal point **c**. As a result, the three rotation angles θ_x, θ_y and θ_z can also be expressed as a function of principal point. Taking the ratio of (2) and (3), we get

$$\theta_z = \tan^{-1}\left(\frac{v_{xy}-v_0}{v_{xx}-u_0}\right) \tag{9}$$

where (v_{xx}, v_{xy}) are the coordinates of the vanishing point $\mathbf{v}_x$ and (v_{yx}, v_{yy}) are those of the vanishing point $\mathbf{v}_y$. Taking the ratio of (3) and (4), we get

$$\theta_y = -\tan^{-1}\left(\frac{f\sin(\theta_z)}{v_{xy} - v_0}\right) \tag{10}$$

Taking the ratio of (6) and (7), we get

$$\theta_x = \tan^{-1}\left(\frac{f\cos(\theta_z)}{(v_{yy} - v_0)\cos(\theta_y) - f\sin(\theta_z)\sin(\theta_y)}\right) \tag{11}$$

Therefore, the first two columns of $\mathbf{H}$ depend only on the principal point. Because the principal points of recent CCD cameras are very close to the center of the image, a closed-form solution may be obtained by assuming that the principal point $\mathbf{c}$ is at the center of the image. This assumption will be relaxed later in the more general case using non-linear estimation discussed in section 3. The last column of equation (1) denotes the homogeneous image point corresponding to the projection of origin of the world coordinate system [4], and thus can be assumed at $\mathbf{b}_1$ in figure 2. Since world origin is visible in images, t_z in our cases cannot be close to zero. Without loss of generality, assume $\hat{t}_z$ in one view equals to 1. Therefore the estimated homography is of the form:

$$\hat{\mathbf{H}} = \bar{\mathbf{H}}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/t_z \end{bmatrix} \tag{12}$$

In other words, an inhomogeneous world point $\bar{\mathbf{M}}$ and its estimated inhomogeneous world point $\hat{\mathbf{M}}$ are related via

$$\hat{\mathbf{M}} = \frac{1}{t_z}\bar{\mathbf{M}} \tag{13}$$

Therefore, with the principal point assumed to be known, we get a closed-form solution for the planar homography. Given this homography, the measurements inside the plane passing through the two vertical objects becomes straightforward.

¿From the above derivations, one can not directly solve for all the camera parameters without additional information, since the homography is defined up to a global scalar. One possible solution is to use a second image with a different pose, but with the same internal parameters. Given this new image, neither the two reference distances nor the ratio of an object height to $\mathbf{b}_1\mathbf{b}_2$ would be anymore required. Also note that the closed-form solution is affected by the error due to the assumption that the principal point is at the center of the image. This problem may also be mitigated by using the additional image. Using the same approach as discussed above, the extra image is also computed up to a column-wise scalar t'_z. The scale can be determined by forcing the image point $\mathbf{m}'$ and the corresponding image point $\mathbf{m}$ in the first image (for which t_z is assumed known) to be both projected to the same 3D world point $\mathbf{M}$. We can therefore determine all the internal parameters, the three rotation angles, and the third column of the 3×4 camera projection matrix. Traditional stereo or multiple view techniques [6] can then be applied to measure distances along directions other than the vertical or outside the refernce plane. Camera position also can be computed up to scale factor.

2.2 Measurements From One Vanishing Point

A more general and frequently occuring scenario would be when the two vertical objects are not of the same height, or are not known *a priori* to have the exact same height, e.g. two pedestrians, a street lamp and a tree, etc.; an example is shown in Figure 2. In this case the point $\mathbf{p}$ does not coincide with $\mathbf{v}_y$ anymore, and only one vanishing point i.e. $\mathbf{v}_x$ is known. The only information we have about $\mathbf{v}_y$ is that it is along the line $\mathbf{b}_1\mathbf{b}_2$, i.e.

$$\mathbf{v}_y{}^T [l_1 \;\; l_2 \;\; 1]^T = 0 \tag{14}$$

where $(l_1, l_2, 1)$ is the line $\mathbf{b}_1 \times \mathbf{b}_2$. In addition, equation (8) can be written as

$$\mathbf{v}_y{}^T \begin{bmatrix} v_{xx} - u_0 \\ v_{xy} - v_0 \\ f^2 - u_0 v_{xx} + u_0^2 - v_0 v_{xy} + v_0^2 \end{bmatrix} = 0 \tag{15}$$

Therefore by combining equations (14) and (15) and some simplification, we can show that $\mathbf{v}_y$ is of the form

$$\mathbf{v}_y = \begin{bmatrix} v_{xy} - v_0 - (f^2 - u_0 v_{xx} + u_0^2 - v_0 v_{xy} + v_0^2) l_2 \\ -v_{xx} + u_0 + (f^2 - u_0 v_{xx} + u_0^2 - v_0 v_{xy} + v_0^2) l_1 \\ (v_{xx} - u_0) l_2 - (v_{xy} - v_0) l_1 \end{bmatrix} \tag{16}$$

Equation (16) defines the vanishing point $\mathbf{v}_y$ for the direction $\mathbf{b}_1\mathbf{b}_2$ as a function of the focal length f and the principal point (u_0, v_0), which can be assumed to be at the center of the image as in the last subsection in order to obtain a closed-form solution 2.1. This assumption will be relaxed later. The four points $\mathbf{v}_y$, $\mathbf{p}$, $\mathbf{b}_2$ and $\mathbf{b}_1$ define a cross ratio. The value of the cross ratio determines a ratio of the lengths of two vertical objects.

$$\frac{d(\mathbf{b}_2, \mathbf{t}_2)}{d(\mathbf{b}_1, \mathbf{t}_1)} = \frac{d(\mathbf{p}, \mathbf{b}_2) d(\mathbf{v}_y, \mathbf{b}_1)}{d(\mathbf{v}_y, \mathbf{b}_2) d(\mathbf{p}, \mathbf{b}_2)} \tag{17}$$

where $d(,)$ denotes the distance between two points in the image plane. Substituting equation 16 into equation 17, and using the fact that cross ratios remain invariant under projection transformations between two images, we can solve for the focal length f. One such equality provides four relations for f, two of which can be eliminated by verifying that the points are in front of the cameras. For removing the remaining ambiguity one can either rely on a third image to obtain a closed-form solution, or avoid the third image by minimizing the symmetric transfer error of the geometric distances as shown in the next section.

Once the focal length is computed, $\mathbf{v}_y$ can be obtained from equation (16), and hence one can get the length ratio from equation (17). Rotation angles can be computed from equations (9)-(11). Translations are obtained from the last column of the projection matrix by fixing the scale for the first camera and following the approach discussed in subsection 2.1. Given all external and internal parameters one can then perform metric measurements.

3 Nonlinear Solution

As described above, when no reference lengths or distance ratios are known, the problem can be solved given one vanishing point and a known principal point, if two images

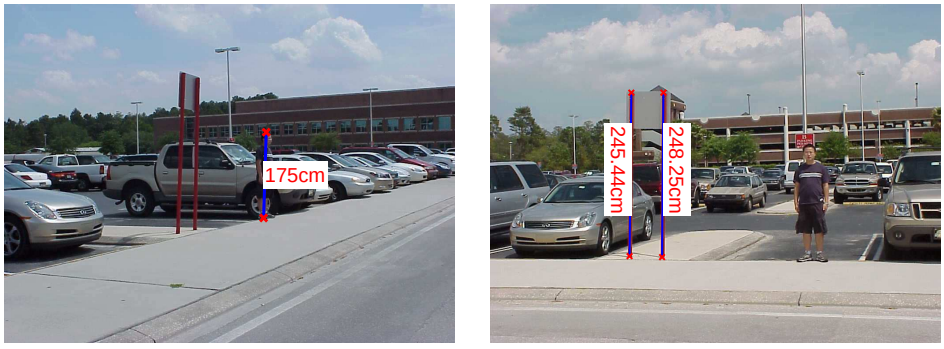
are available. If however the principal point is unknown, or if the errors due to taking the principal point as the image center are not negligible, then the closed-form solution would fail. In this case an accurate solution can still be obtained by minimizing the symmetric transfer error of the geometric distances. In which case, the closed form solution described in section 2 can be used to initialize the minimization step described below.

Therefore, two problems need to be addressed in this section. Firstly, we relax the assumption that the principal point is known. Secondly, the ambiguity in the solution of the focal length will be removed. Recall that the projection matrices and hence the inter-image homography $\mathbf{H}$ depend on the position of the principal point up to an ambiguity caused by the quadratic equation in (17) giving two solutions for f^2. This ambiguity is resolved immediately if one of the solutions for f^2 leads to a complex value for f. Otherwise, the correct value of the principal point as well as the principal point should minimize the symmetric transfer error of the geometric image distances between pairs of corresponding points $(\mathbf{x}_i, \mathbf{x}'_i)$.

$$(u_0, v_0, f) = \arg\min_{W} \sum_i d(\mathbf{x}_i, \mathbf{H}^{-1}\mathbf{x}'_i)^2 + d(\mathbf{x}'_i, \mathbf{H}\mathbf{x}_i)^2 \tag{18}$$

where $d(\cdot,\cdot)$ is the Euclidean distance between the points, W is the search window, and the inter-image homography is given by $\mathbf{H} = \hat{\mathbf{H}}'\hat{\mathbf{H}}^{-1}$. Note that this minimization process is similarity-invariant because only image distances are minimized and the points $\mathbf{x}_i$ and $\mathbf{x}'_i$, which are the projections of 3D points $\mathbf{X}_i$, do not depend on the scale in which $\mathbf{X}_i$ are defined, i.e. different scaled points will project to the same points [6].

Using the closed form solution described in the previous section we can find an initial estimate for f. Also assuming that the principal point is in some neighborhood of the image center, we reduce the search space for the minimization problem in (18) to a window around the image center and the initial estimate of f. We therefore found the solution by discretizing the search space and used an exhaustive search to find the solution to avoid converging to a local minimum.



(a) (b)

Figure 4: Measuring height of vertical objects: (a) The standing person has known height, (b) Computed heights of two sign board posts.

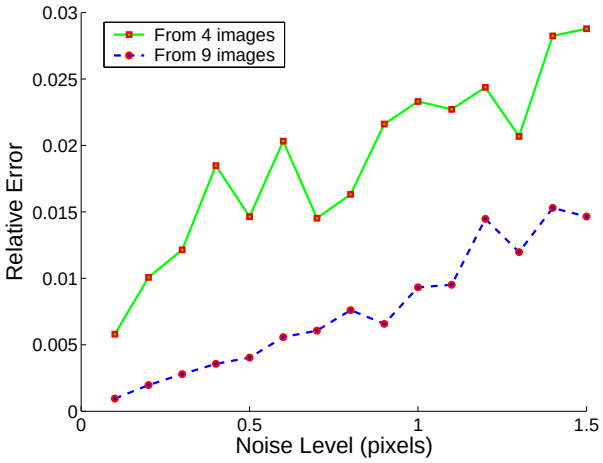


Figure 5: Performance versus noise (in pixels) using four views and nine views separately. The results shown here are averaged over 50 independent trials.

4 Experimental Results

4.1 Computer Simulation

The simulated camera has a focal length of $f = 1000$, unit aspect ratio, zero skew, and the principal point close to the center of the image. The image resolution is 720×360. In the experiments presented here, we observed the two vertical objects with height 100 and 50 units separately in nine positions. The nine observations are all close to the plane $Z = 300$ units, and with around 25 unit distances along x-axes and 50 unit distances along y-axes.

First we evaluate the performance versus noise level. In principle, two images suffice to solve the calibration, but in this experimentation we used four image pairs observed by cameras in most distant of the nine positions in order to improve the quality of the results. Gaussian noise with zero mean and a standard deviation of $\sigma \leq 1.5$ was added to the projected image points. The estimated ratio is then compared with the ground truth and shown in figure 5. The relative error of estimated ratio is 1.46% for a typical noise level $\sigma = 0.5$, and increased to 2.88% when the added noise was $\sigma = 1.5$ which is larger than the typical noise in practice.

We also examined the performance with respect to the number of viewpoints (i.e. the number of image pairs). We show the results using nine views also in figure 5. With nine views, the relative error of estimated ratio are not beyond 1% until much noise ($\sigma \geq 1.2$) is added. For all the noise level, the more viewpoints we have, the more accurate camera calibration will be in practice, since data redundancy compensates for the noise in the data.

4.2 Real Data

The proposed method was also tested on real data sets, some of which are shown below and throughout the paper. For demonstrated results shown in figure 1, 4, 6, we all use the

(a) (b)

Figure 6: Measurements which might be difficult in practice: (a) The height of the pine is around 264cm, and the distance from the head of the standing people to the top of the pine is 154cm, (b)The distance from the head of the standing people to the top of the stick is around 140cm.

computed standing person as the reference. In figure 4, the computed heights of sign board posts are similar which coincide with the fact in real life. In order to test the accuracy of our algorithm, we also compared the computed results with ground truth measurements. For instance in figure 6, the estimated stick's height is 100.75cm, while the ground truth is 99.4cm. The distance between the bottom of the standing person and bottom of the stick is 116cm, the estimated one is 119.47cm. The approach can be used to measure heights of the objects that are not accessible for direct measurement too. For instance, we estimated the tree's height as 263.62cm as shown in Figure 6. Other estimated distances which might be difficult to measure in practice are also shown.

5 Conclusion

We have explored new solutions for metrology from uncalibrated images that require minimal geometric information. This is achieved by making some simplifying assumptions about the camera intrinsic parameters or by using additional images. This work therefore extends the work of Criminisi et al. [3], whereby external geometric constraints are relaxed by trading off the intrinsic constraints. The results show the high accuracy and the effectiveness of the approach as compared to the ground truth. The approach can be made further robust by using additional feature points or extra images, in which case one can use bundle adjustment to improve the accuracy.

Appendix: Feature Extraction

Features can be extracted either manually or automatically (e.g. using an edge detector or Harris corner detector). The features are mostly the image locations of the top and base points **t**, **b**. These features however are subject to errors, and hence similar to [3], it is possible to use a maximum likelihood estimation method, with the uncertainty in the

top and base points modelled by the covariance matrices Λ_b and Λ_t. Since in the error-free case, these points must be aligned with the vertical vanishing point $\mathbf{v}_x$ as in figure 2, we can determine maximum likelihood estimates of their true locations ($\hat{\mathbf{t}}$ and $\hat{\mathbf{b}}$) by minimizing the sum of the Mahalanobis distances between the input points $\mathbf{t}$ and $\mathbf{b}$ and their MLE estimates $\hat{\mathbf{t}}$ and $\hat{\mathbf{b}}$

$$\arg\min_{\hat{\mathbf{b}},\hat{\mathbf{t}}}[(\mathbf{b}-\hat{\mathbf{b}})^T\Lambda_b^{-1}(\mathbf{b}-\hat{\mathbf{b}})+(\mathbf{t}-\hat{\mathbf{t}})^T\Lambda_t^{-1}(\mathbf{t}-\hat{\mathbf{t}})] \tag{19}$$

subject to the alignment constraint

$$\mathbf{v}_x^T(\hat{\mathbf{b}}\times\hat{\mathbf{t}}) \tag{20}$$

References

[1] B. Caprile and V. Torre. Using vanishing points for camera calibration. *International Journal of Computer Vision*, 4(2):127–140, 1990.

[2] R. Cipolla, T. Drummond, and D. Robertson. Camera calibration from vanishing points in images of architectural scenes.

[3] A. Criminisi, I. Reid, and A. Zisserman. Single view metrology. *International Journal of Computer Vision*, 2001.

[4] O.D. Faugeras. *Computer Vision: a Geometric Viewpoint*. MIT Press, 1993.

[5] Tao Zhao Fengjun Lv and Ram Nevatia. Self-calibration of a camera from video of a walking human. In *Proc. of the 16th International Conference on Pattern Recognition*, pages 562–567, 2002.

[6] R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[7] David Liebowitz and Andrew Zisserman. Combining scene and auto-calibration constraints. In *Proc. International Conference on Computer Vision*, pages 293–300, 1999.

[8] S.J. Maybank and O.D. Faugeras. A theory of self-calibration of a moving camera. *International Journal of Computer Vision*, 8(2):123–152, 1992.

[9] K.-Y. Wong, R.S.P. Mendonca, and R. Cipolla. Camera calibration from surfaces of revolution. *IEEE Trans. Pattern Analysis & Machine Intelligence*, 2003.