Refining Single View Calibration With the Aid of Metric Scene Properties

Manolis I.A. Lourakis and Antonis A. Argyros
Institute of Computer Science, Foundation for Research and Technology - Hellas
Vassiliki Vouton, P.O.Box 1385, GR 71110, Heraklion, Crete, GRe:Tel
{lourakis|argyros}@ics.forth.gr – http://www.ics.forth.gr/cvrl/

Abstract

Intrinsic camera calibration using a single image is possible provided that certain geometric objects such as orthogonal vanishing points and metric homographies can be estimated from the image and give rise to adequate constraints on the sought calibration parameters. In doing so, however, any additional metric information that might be available for the imaged scene is not always straightforward to accommodate. This paper puts forward a method for incorporating into the calibration procedure metric scene information expressed in the form of known segment 3D angles, equal but unknown 3D angles and known 3D length ratios. Assuming the availability of an initial calibration estimate, the proposed method refines the former by numerically minimizing an error term corresponding to the discrepancy between the scene’s known metric properties and the values measured with the aid of the calibration estimate. Sample experimental results demonstrate the improvements in the intrinsic calibration estimates that are achieved by the proposed method.

Keywords: Camera calibration, single view reconstruction, image-based modeling, computer vision.

1 INTRODUCTION

Deriving the geometry and appearance of a scene directly from photographs is an attractive paradigm that has generated strong interest in image-based modeling techniques during recent years. A particular class of such techniques consists of those dealing with Single View Reconstruction (SVR) [LCZ99]. Their aim is to create a 3D graphical model corresponding to a scene for which only a single image is available. Determining the intrinsic calibration parameters (i.e. interior orientation) of the employed camera is an important step towards SVR. This is because intrinsic calibration upgrades a camera to a metric measuring device and, given information related to vanishing points and lines, facilitates computations such as the estimation of line segment angles and length ratios, the metric rectification of planes, the determination of a plane’s orientation relative to the camera, the estimation of the dihedral angle between two planes, etc.

In contrast to standard photogrammetric methods for camera calibration where known 3D points in a world coordinate system are required, single view camera calibration relies on constraints provided by the parallelism and orthogonality of observed 3D lines and planes. For instance, Caprile and Torre [CT90] have shown that three vanishing points corresponding to orthogonal directions allow partial calibration of a camera from a single view. Their method follows a construction showing that the principal point of the camera is at the orthocentre of the triangle having the vanishing points as its vertices and relies on prior knowledge of the aspect ratio plus the assumption that the image skew is zero. Liebowitz and Zisserman [LZ98] studied the properties of a metric planar homography, a transformation that rectifies (i.e. cancels) the projective distortion of an imaged plane. They proved that metric rectification of a plane allows the determination of the images of the circular points and thus partially determines camera calibration by providing a pair of constraints on the intrinsic parameters. Gurdjos and Payrisat [GP00] propose another approach for exploiting the metric structure of a scene plane, according to which the triangle employed in [CT90] can be recovered from a single metric homography combined with additional camera constraints. In more recent work, Liebowitz and Zisserman [LZ99] present a technique for calibrating a camera using a combination of linear constraints arising from vanishing points corresponding to perpendicular directions, from vanishing lines and the vanishing points of their perpendicular directions and from metric rectification homographies. When more than the minimum needed constraints are available, they can be conveniently combined in a least squares fashion. Colombo et al [CBP05] exploit constraints arising from the symmetry of surfaces of revolution to calibrate a camera with known aspect ratio and skew.

Common to all methods briefly reviewed above is the shortcoming that their applicability depends upon whether a minimum number of suitable constraints is available in a certain scene. Often, the number of available constraints is insufficient, forcing the employment of approximate, simplified camera models. An example of such a simplified model is to approximate the principal point by the image center; this assumption, never-
theless, is not always valid [HKU2]. Furthermore, with
the exception of [LZ98, LZ99], none of the aforemen-
tioned methods allows metric scene information other
than orthogonality and parallelism to be explicitly em-
bedded in the calibration process. Even [LZ98, LZ99],
however, require that the line segments whose metric
properties are to be taken into account for estimating
a metric homography are all coplanar and that the un-
derlying plane has been a priori rectified. This paper
presents a novel approach for refining an existing esti-
mate of the calibration parameters. Starting with an
approximate calibration which may have been computed
with any of the aforementioned methods, our approach
refines it to account for metric constraints expressed
in the form of segment angles and length ratios. The rest
of the paper is organized as follows. Section 2 intro-
duces the notation that is used in the remainder of
the paper and reviews some background material. Section
3 presents the proposed technique for single view cali-
bration refinement. Some implementation details are
given in section 4. Experimental results from a proto-
type implementation are presented in section 5 and the
paper concludes with a brief discussion in section 6.

2 BACKGROUND

2.1 Elements of Single View Geometry

In the following, vectors and arrays appear in boldface
and are represented using projective (homogeneous) co-
ordinates [HZ00]. An image point with Euclidean co-
ordinates \((x, y)\) is represented by the homogeneous
3-vector \(\mathbf{x} = (x, y, 1)^T\). Similarly, a line is repre-
sented by a homogeneous 3-vector \(\mathbf{l}\) such that \(\mathbf{P}^T\mathbf{x} = 0\) for all
points \(\mathbf{x}\) lying on it. Assuming a pinhole camera model,
camera calibration is specified by the following five pa-
rameter upper triangular matrix \(\mathbf{K}\) [HZ00]:

\[
\mathbf{K} = \begin{bmatrix}
    f_u & s & u_0 \\
    0 & f_v & v_0 \\
    0 & 0 & 1
\end{bmatrix}.
\]

The parameters \(f_u\) and \(f_v\) correspond to the focal length
expressed in pixel units along the two axes of the image,
s is the skew parameter and \((u_0, v_0)\) are the coor-
dinates of the image principal point in pixels. Parameter s is
related to the angle between the two image axes and
is zero for most cameras. Furthermore, the aspect ratio
\(r = \frac{f_u}{f_v}\) for a certain camera is fixed and equal to one in
most cases. A camera with zero skew and unit aspect
ratio is commonly called a natural camera.

Customarily, single view camera calibration is per-
duced by determining the image of the absolute conic
(IAC). The absolute conic is a special conic lying at the
plane at infinity, having the property that its image pro-
jection depends on the intrinsic parameters only and not
on the camera orientation or position. The IAC is itself
a conic whose equation is defined by a homogeneous
3 x 3 symmetric matrix \(\mathbf{o}\) with five degrees of freedom
given by \(\mathbf{o} = (\mathbf{K}^T)^{-1}\). Working with the IAC is more
convenient than working with \(\mathbf{K}\) since the constraints
involved in single view calibration are linear in the ele-
ments of \(\mathbf{o}\), thus admitting a simple algebraic solution.
The camera calibration matrix \(\mathbf{K}\) can be computed from
\(\mathbf{o}\) via Cholesky decomposition, a factorization that is
unique for a symmetric, positive definite matrix.

A concept that will be repeatedly used in the follow-
ing is that of vanishing points and lines. Assuming an
infinite 3D line that is imaged under perspective, a point
on it that is infinitely far away from the camera projects
to a finite point known as the vanishing point that de-
dends only on the 3D line’s direction and not on its po-
position. Thus, parallel 3D lines share the same vanishing
points. In a similar manner, the vanishing points of a
set of non-parallel, coplanar 3D lines lie on the same
image line, which is known as the vanishing line of the
underlying plane.

2.2 Metric Measurements Using the IAC

Knowing the matrix of the IAC \(\mathbf{o}\) allows similarity (i.e.
scaled Euclidean) invariants such as the size of angles
and the ratio of lengths to be computed directly from
images. Assume two 3D lines whose vanishing points
in an image are denoted by the homogeneous vectors
\(\mathbf{v}_1\) and \(\mathbf{v}_2\). Applying Laguerre’s formula, the acute an-
gle \(\theta\) between the two line directions can be computed from
[HZ00]

\[
\cos(\theta) = \frac{|S(\mathbf{v}_1, \mathbf{v}_2)|}{\sqrt{S(\mathbf{v}_1, \mathbf{v}_1) S(\mathbf{v}_2, \mathbf{v}_2)}},
\]

where \(S(\mathbf{a}, \mathbf{b}) = \mathbf{a}^T \mathbf{o} \mathbf{b}\) for vectors \(\mathbf{a}\) and \(\mathbf{b}\). In the case
of right angles, Eq. (2) simplifies to \(S(\mathbf{v}_1, \mathbf{v}_2) = 0\),
which is linear in the elements of \(\mathbf{o}\) and hence con-
stitutes the most commonly used type of constraint for
single view calibration. Note also that for parallel lines,
Eq. (2) reduces to the trivial identity \(1 = 1\). The fact
that angles can be computed from \(\mathbf{o}\) using Eq. (2), fa-
cilitating the computation of length ratios. More spe-
cifically, consider four non collinear points \(A, B, C\) and
\(D\) and assume that the length ratio of segments \(AB\) and
\(CD\) should be computed. Applying the law of sines to
the triangles \(ABC\) and \(BCD\), allows the sought ratio to be
specified as a ratio of sine ratios. Using the equality
\(\sin(\theta) = \sqrt{1 - \cos^2(\theta)}\) to express sines as cosines
and applying Eq. (2) to compute cosines, finally leads to
the following expression for the length ratio:

\[
\frac{AB}{CD} = \frac{\sqrt{S(\mathbf{v}_{ab}, \mathbf{v}_{ab}) S(\mathbf{v}_{ac}, \mathbf{v}_{ac}) S(\mathbf{v}_{bc}, \mathbf{v}_{bc}) S(\mathbf{v}_{ac}, \mathbf{v}_{bc}) S(\mathbf{v}_{bd}, \mathbf{v}_{bd}) S(\mathbf{v}_{cd}, \mathbf{v}_{cd})}}{\sqrt{S(\mathbf{v}_{bd}, \mathbf{v}_{bd}) S(\mathbf{v}_{cd}, \mathbf{v}_{cd}) S(\mathbf{v}_{bd}, \mathbf{v}_{bd}) S(\mathbf{v}_{cd}, \mathbf{v}_{cd})}}
\]

where \(\mathbf{v}_{ab}, \mathbf{v}_{ac}, \mathbf{v}_{bc}, \mathbf{v}_{bd},\) and \(\mathbf{v}_{cd}\) denote the vanishing
points of line segments \(AB, AC, BC, BD\) and \(CD\) respec-
tively.
3 REFINING AN INTRINSIC CALIBRATION ESTIMATE

It is clear that through Eqs. (2) and (3), known scene properties such as angles and length ratios impose high-order polynomial constraints on the elements of $\omega$. However, the difficulties associated with solving (possibly overdetermined) systems of nonlinear polynomial equations, hinder the direct employment of such constraints for estimating $\omega$. The key observation upon which the proposed method is based is the fact that those constraints can be used to refine an existing estimate of $\omega$ so that they are satisfied approximately. Specifically, assuming that an initial estimate of $\omega$ is available, this can be used to define an error term which corresponds to the discrepancy between the values measured from it with Eqs. (2)-(3) and the known scene properties. Minimizing this error term over $\omega$ with numerical optimization techniques will refine the calibration estimate. This procedure is explained in more detail in the remainder of this section.

Assume that a set of line segments has been extracted from an image. Assume further that these line segments form a set $A$ of pairs with known 3D angles, a set $E$ of quadruples each defining two unknown but equal 3D angles and a set $R$ of pairs with known 3D length ratios. For a pair of segments $(s, r) \in A$ with a priori known angle $\phi$, denote by $\alpha(s, r; \omega)$ the cosine of their angle as estimated from Eq. (2) with the current estimate of $\omega$. Then, the term $\alpha(s, r; \omega) - \cos(\phi)$ amounts to their cosine difference error. Similarly, for a quadruple of segments $(s, r, p, q) \in E$, the term $\alpha(s, r; \omega) - \alpha(p, q; \omega)$ amounts to their cosine estimates difference error. For a pair of segments $(s, r) \in R$ with a priori known ratio $\lambda$, denote by $\rho(s, r; \omega)$ the ratio of their lengths as estimated with Eq. (3). The term $\rho(s, r; \omega) - \lambda$ represents their length ratios difference error. Thus, a cumulative error term can be constructed by summing up the squares of all available cosine and length difference errors, as follows:

$$
e(A, E, R; \omega) = \sum_{(s, r) \in A} [\alpha(s, r; \omega) - \cos(\phi)]^2 + \sum_{(s, r, p, q) \in E} [\alpha(s, r; \omega) - \alpha(p, q; \omega)]^2 + \sum_{(s, r) \in R} [\rho(s, r; \omega) - \lambda]^2 . \tag{4}
$$

Note that the above definition does not require that all line segments involved in it lie on the same plane, concerns quantities directly measurable from the image and can incorporate as many constraints as desired.

Being a sum of squares, the above error term can be iteratively minimized over the parameters of $\omega$ using a nonlinear least squares algorithm such as Levenberg-Marquardt. The initial estimate of $\omega$ serves as the starting point for the numerical minimization. A suitable parametrization for $\omega$ can be chosen by recalling that the former is a homogeneous matrix, i.e. it can be multiplied by an arbitrary scalar without any effect on the angles and ratios computed by Eqs. (2) and (3). Since modern cameras have rectangular pixels, the skew parameter is assumed to be fixed to zero. Substituting $K$ from Eq. (1) with $s = 0$ into the definition of $\omega$ and multiplying by $f^2$, yields the following expression for $f^2 \omega$:

$$
f^2 \omega = \begin{bmatrix}
  r^2 & 0 & -r^2 u_0 \\
  0 & 1 & -v_0 \\
  -r^2 u_0 & v_0 & r^2 f_0^2 + v_0^2 + r^2 u_0^2
\end{bmatrix} . \tag{5}
$$

As it is clear from the above expression, $\omega$ can be parametrized with four parameters $o_i$, $i = 1 \ldots 4$, directly corresponding to the parameters of $K$: $o_1 = f_x$, $o_2 = u_0$, $o_3 = v_0$, $o_4 = r$. When more information is known regarding the camera to be calibrated, less than four parameters suffice for parametrizing $\omega$. For instance, a natural camera has $r = 1$, thus only the first three of the above parameters need to be estimated.

4 IMPLEMENTATION DETAILS

The line segments that are required by the proposed method were defined manually. Alternatively, they could have been obtained via edge detection, edge linking and orthogonal line fitting to edge lists obtained by segmentation at points of high curvature. The nonlinear technique suggested in [LZ98] was employed to compute maximum likelihood estimates (MLE) of the vanishing points corresponding to parallel line segments. Vanishing lines of planes were estimated from pairs of vanishing points corresponding to two sets of parallel, coplanar lines. Given the vanishing line $l$ of a plane, the vanishing point of an arbitrary line segment $s$ on the plane is their point of intersection, simply computed from the cross product $l \times s$. Initial estimates for the calibration matrix were obtained by combining linear constraints arising from orthogonal vanishing points and metric rectification homographies, as described in [LZ98, LZ99]. The minimization of Eq. (4) was achieved using the Levenberg-Marquardt algorithm as implemented by the freely available levmar library [Lout04]. It necessary, convergence can be improved by employing the box-constrained L-M variant from levmar, which allows the imposition on the minimization parameters of constraints regarding their minimum and maximum permissible values. The Jacobian of Eq. (4) with respect to the calibration parameters that is necessary for the non-linear minimization has been computed analytically with the aid of MAPLE’s symbolic differentiation facilities. Finally, it is worth mentioning that the analysis of section 3 has assumed that the radial lens distortion in the image is negligible. If this is not the case, the effects of distortion can be corrected by applying a technique such as [DF95].

Journal of WSCG     ISSN 1213-6972     131     ISBN 978-80-86943-00-8
frames sequence imaging the calibration object from different viewpoints. Knowledge of the calibration object's shape, allowed the calibration matrix to be estimated from all 27 frames using Bouguet's MATLAB calibration toolkit [Bou04] as

$$\begin{bmatrix}
1565.7 & 0 & 800.9 \\
0 & 1565.5 & 642.44 \\
0 & 0 & 1
\end{bmatrix}.$$  

Fig. 1(a) has plenty of lines for computing vanishing points and metric rectification homographies. The vertical and horizontal lines shown in Fig. 1(a) were employed for estimating two orthogonal vanishing points. Assuming a natural camera and using the three constraints provided by the pair of vanishing points and the metric homography corresponding to the left checkerboard plane, the initial calibration estimate from a single view was computed as

$$\begin{bmatrix}
1552.0 & 0 & 816 \\
0 & 1552.0 & 612 \\
0 & 0 & 1
\end{bmatrix}.$$  

Following this, known scene properties were used to refine the single view calibration estimate. More precisely, the refinement was based on the known, equal to 45\(^\circ\), angles between line segment pairs 0-4, 2-4, 0-5 and 2-5 that are illustrated in Fig. 1(b) as well as the unit length ratios corresponding to segments 0-1, 2-3, and 0-2. The refined calibration estimate using the proposed method was

$$\begin{bmatrix}
1553.52 & 0 & 805.293 \\
0 & 1553.52 & 638.787 \\
0 & 0 & 1
\end{bmatrix}.$$  

Since the initial calibration estimate was close to that obtained via grid-based calibration, the refinement did not change the calibration matrix much. Nevertheless, it has improved slightly the estimate of the principal point.

The second experiment employed one frame from the well-known Valbonne church sequence, shown in Fig. 2(a). This sequence consists of several 512 \(\times\) 768 images depicting the church from different viewpoints. True camera calibration parameters for the Valbonne sequence are unavailable. However, calibration of the employed camera has been possible using self-calibration techniques applied to the whole sequence. As reported in [LD00], assuming unit aspect ratio and a principal point coinciding with the image center, the intrinsic calibration matrix was estimated as

$$\begin{bmatrix}
682.84 & 0 & 256 \\
0 & 682.84 & 384 \\
0 & 0 & 1
\end{bmatrix}.$$  

Single view calibration was performed assuming a natural camera with its principal point fixed at the image center. The two orthogonal vanishing points corresponding to vertical and horizontal directions from the church’s front wall and bell-tower (see Fig. 2(a)) provided one constraint on the IAC. Notice that these van-
ishing points are located far outside the image, therefore their accurate localization is difficult. Two more constraints were provided by the metric rectification homography of the front wall that was estimated from the known length ratios of segments 10-11 and 8-9 shown in Fig. 2(b). The calibration matrix was estimated as

\[
\begin{bmatrix}
529.6 & 0 & 256 \\
0 & 529.6 & 384 \\
0 & 0 & 1
\end{bmatrix}.
\]

Clearly, the discrepancy between the above estimate and that obtained via self-calibration is considerable. The proposed method was then employed to refine this single view calibration estimate using a few line segments with equal 3D angles and lengths. Specifically, the a priori information that was employed concerned the equality of the angles formed by segments 0-1 and 2-3 shown in Fig. 2(b) and the equality of the lengths of segment pairs 4-5, 6-7, 8-9 and 10-11. The refined calibration estimate was

\[
\begin{bmatrix}
672.0 & 0 & 256 \\
0 & 672.0 & 384 \\
0 & 0 & 1
\end{bmatrix},
\]

which can be verified to be very similar to that obtained with multiple image self-calibration.

The third experiment refers to the first frame of another well-known sequence, shown in Fig. 3(a). This sequence consists of 11 512 × 512 frames acquired as the camera approached the scene. According to http://www.robots.ox.ac.uk/~vgg/data.html, the intrinsic calibration matrix computed by Oxford’s Visual Geometry Group (VGG) structure and motion recovery system [FZ98] for this sequence is

\[
\begin{bmatrix}
496.9 & 0 & 273.5 \\
0 & 496.9 & 280.0 \\
0 & 0 & 1
\end{bmatrix}.
\]

Single view calibration was carried out by employing the two orthogonal vanishing points corresponding to the vertical and horizontal directions illustrated in Fig. 3(a) and the metric homography corresponding to the floor plane that was estimated using the line segments of Fig. 3(b). Single view calibration with a variable principal point failed due to the estimated \(\omega\) not being positive definite. Therefore, calibration was performed assuming a natural camera with the principal point fixed on the image center. The calibration matrix estimated in this case was

\[
\begin{bmatrix}
312.1 & 0 & 256.0 \\
0.0 & 312.1 & 256.0 \\
0.0 & 0.0 & 1.0
\end{bmatrix},
\]

which differs significantly from that estimated as a byproduct of multiple view reconstruction. The single view estimate was next refined with the proposed method, assuming a natural camera and using the three pairs of equal in length line segments 0-1, 2-3 and 4-5 that are shown in Fig. 3(b). The refined estimate was

\[
\begin{bmatrix}
527.5 & 0 & 254.8 \\
0 & 527.5 & 303.7 \\
0 & 0 & 1
\end{bmatrix}.
\]
REFERENCES


