Camera calibration based on arbitrary parallelograms

Jun-Sik Kim\textsuperscript{a,*}, In So Kweon\textsuperscript{b}

\textsuperscript{a}Robotics Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA, USA
\textsuperscript{b}Department of EECS, Korea Advanced Institute of Science and Technology, Kusong-Dong 373-1, Yuseong-Gu, Daejeon, Republic of Korea

\begin{abstract}
Existing algorithms for camera calibration and metric reconstruction are not appropriate for image sets containing geometrically transformed images for which we cannot apply the camera constraints such as square or zero-skewed pixels. In this paper, we propose a framework to use scene constraints in the form of camera constraints. Our approach is based on image warping using images of parallelograms. We show that the warped image using parallelograms constrains the camera both intrinsically and extrinsically. Image warping converts the calibration problems of transformed images into the calibration problem with highly constrained cameras. In addition, it is possible to determine affine projection matrices from the images without explicit projective reconstruction. We introduce camera motion constraints of the warped image and a new parameterization of an infinite homography using the warping matrix. Combining the calibration and the affine reconstruction results in the fully metric reconstruction of scenes with geometrically transformed images. The feasibility of the proposed algorithm is tested with synthetic and real data. Finally, examples of metric reconstructions are shown from the geometrically transformed images obtained from the Internet.
\end{abstract}

\section{Introduction}

The reconstruction of the three-dimensional (3D) objects from images is one of the most important topics in computer vision. After stratified approaches for reconstructing structures were introduced, the meanings of the invariant features in each stratification became much clearer than before\cite{1}. Estimating the camera's intrinsic parameters is a key issue of 3D metric reconstruction for stratified approaches. To estimate the intrinsic parameters of cameras or, equivalently, to recover metric invariants such as the absolute conic or the dual absolute quadric of reconstructed scenes, there are many possible approaches. The approaches are classified into three categories. Classification depends on whether scenes or cameras constraints are used. The former uses only scene constraints of known targets called calibration objects. Faugeras proposed a direct linear transformation-based method using 3D targets\cite{2}. Tsai proposed another camera calibration method using radial alignment constraints\cite{3}. Recently, there have been two-dimensional target-based algorithms; Sturm and Maybank proposed a plane-based calibration algorithm and provided theoretical analysis of the algorithm\cite{4}. Zhang proposed a similar method independently\cite{5}, and Liebowitz showed that the plane-based calibration is equivalent to fitting the image of the absolute conic (IAC) with the imaged circular points (ICP) of the imaged plane, and gave an example of simple calibration objects\cite{6,7}. Because ICPs are always on the images of planar circles, circle-like patterns have been introduced to calibrate cameras. Kim et al. used concentric circles instead of grid patterns and showed that there are significant advantages of the circle-like patterns based on conic dual to the circular points (CDCP) which is dual to the circular points\cite{8}. Recently Gurdjos et al. generalized it to confocal conic cases\cite{9}. The concentric circles are also used as an easily detectable feature to calibrate the cameras\cite{10}. Wu et al. proposed a method to use two parallel circles\cite{11}, and Gurdjos et al. generalized it to the case that has more than two circles\cite{12}. Furthermore, Zhang proposed a method to use one-dimensional objects in restricted motion\cite{13}.

The second approach uses only camera assumptions without calibration targets. This approach is called autocalibration or self-calibration. After Faugeras et al. initially proposed some theories about autocalibration\cite{14,15}, interest in it began to increase. Pollefeys and Van Gool proposed a stratified approach to calibrate cameras and to metric-reconstruct scenes with static camera assumptions and attempted to generalize it to cameras with variable parameters\cite{16}. In some cases, people use camera motion constraints. Armstrong et al. proposed a method that translates cameras with no change of intrinsic parameters\cite{17}. Agapito et al. proposed a linear method for rotating cameras with varying parameters\cite{18}.

Finally, one can choose an approach that is a combination of the first two approaches when there is information available on both the scenes and the camera. Caprile and Torre proposed a camera...
calibration method using three vanishing points from an image [19]. Criminisi et al. showed that it is possible to derive 3D structures from a single image with some measurements [20]. Liebowitz and Zisserman showed that combining camera and scene constraints is possible using the form of the constraints on the image of the absolute conic and it is beneficial for analyzing algorithms [6,7].

However, all of the above methods, which are based on the camera constraints, are not applicable to image sets that have geometrically transformed images such as pictures of pictures, scanned images of printed pictures, and images modified with image-processing tools. These various types of images are easily found on the Internet. Recently, a method has been proposed to use many Internet images to reconstruct and navigate the forest of images [21], but it deals only with images captured by digital cameras whose intrinsic parameters are highly constrained, not with pictures of pictures. To solve this problem, the constraints on scene structures should be investigated. Wilczkowiak et al. [22] showed that combining the knowledge of imaged parallelepipeds and camera constraints is beneficial to metric reconstruction. They introduced a duality of parameters of a parallelepiped with a constrained camera, and show how the calibration is achieved. Although their work explains various scene constraints well, it requires one camera that sees two identical parallelepipeds or two identical cameras that see one parallelepiped to calibrate intrinsic parameters of the camera [22]. In order to use geometrically transformed images in calibrating cameras or metric reconstruction of scenes with the previous methods, it is needed to find a way to transform the scene constraints in the image into another image without camera constraints and this requires additional processes such as estimating epipolar geometry [16,11] which are not necessary in using the cameras constrained intrinsically [6,7]. Thus, it is hard to find the unified way to deal with both the scene constraints and the camera constraints with a set of images containing geometrically transformed images.

In this paper, we propose a unified framework to deal with geometrically transformed images, as well as geometric images from cameras, based on parallelograms rather than parallelepipeds. Because the parallelograms are two-dimensional features, we can warp images containing parallelograms into a pre-defined standard form, and imagine virtual cameras which capture the warped images. The virtual cameras are highly constrained both intrinsically and extrinsically. The image warping converts the problems with scene constraints of parallelograms into the well-known autocalibration problems of intrinsically constrained cameras, even when the images include geometrically transformed images.

In addition, we propose a method to estimate infinite homographies between views using the parallelograms. It enables us to calibrate cameras by transferring constraints between views, even if there are not sufficient number of independent constraints in a single image. Moreover, we show how to reconstruct the physical cameras in affine space by using the newly introduced constraints of the warped images without knowledge about the cameras or projective reconstruction. Metric reconstruction of a scene is possible with image sets including pictures of pictures or scanned images of the printed pictures in a unified framework.

The rest of the paper is organized as follows: Section 2 deals with the virtual cameras of the warped images, and their intrinsic parameters. In Section 3, it is shown that the calibration using scene constraints is converted into the well-known autocalibration methods by warping images, and how the infinite homography is estimated using the warped images. In Section 4, some experimental results using synthetic and real data are discussed, followed by conclusion in Section 5.
ence plane have simple forms. The ICDCP and the ICP in the warped image are expressed with the parameters of the features [23].

**Theorem 1.** In the warped image by \( H_{pp} \), the ICDCP of the reference plane is given as diag(\( R_m^2, R_{pp}^2 \)), where \( R_m \) is an aspect ratio of the real rectangle and \( R_{pp} \) is an aspect ratio of the warped rectangle.

**Proof.** Assume that a projection process of a plane is described with \( H_{pp} \).

The CDCP is a dual circle whose radius is infinity [8] and we assume that there is a circle whose radius is \( r \) in the model plane (Euclidean world). The conic dual to the projected circle in the projected image is

\[
A' = H_{pp}\text{diag}(1, 1, -1/r^2)H_{pp}^T.
\]

The dual conic is transferred to the warped image as

\[
A_{pp} = H_{pp} H_{pp}\text{diag}(1, 1, -1/r^2)H_{pp}^T = \text{diag}(1/a^2, 1/b^2, 1/c^2/r^2).
\]

Meanwhile, a point \((X,Y,1)^T \) on the reference plane is warped to the point at

\[
(X', Y', 1)^T = H_{pp} H_{pp}(X, Y, 1)^T \sim \left( \frac{cX}{a}, \frac{cY}{b}, 1 \right)^T.
\]

\( R_m \equiv Y/X \) and \( R_{pp} \equiv Y_{pp}/X_{pp} \) gives

\[
a : b = R_{pp} : R_m.
\]

As the radius goes to infinity in Eq. (4), the CDCP of the warped plane is expressed as

\[
\text{diag}(1/a^2, 1/b^2, 0) \sim \text{diag}(R_m^0, R_{pp}^0).
\]

Because the ICDCP in the warped image is expressed as \( \text{diag}(R_m^2, R_{pp}^2, 0) \), the ICPs \( I_{pp} \) and \( J_{pp} \), which are their dual features in the warped image, are simply expressed as

\[
I_{pp} = [R_m, i R_{pp}, 0]^T \quad \text{and} \quad J_{pp} = [R_m, -i R_{pp}, 0]^T.
\]

2.2. **Intrinsic parameters of fronto-parallel cameras**

We can imagine that there is a camera, called a fronto-parallel (FP) camera, that may capture the warped image. To find the intrinsic parameters of a FP camera, we derive its IAC.

Assume that the three vanishing points that are in orthogonal directions are \( v_1, v_2 \) and \( v_3 \). The IAC is expressed as [6]

\[
\omega = \alpha^2 I_1 I_1^T + \beta^2 I_2 I_2^T + \gamma^2 I_3 I_3^T,
\]

where \( \alpha, \beta \) and \( \gamma \) are proper scale factors and \( I_1, I_2 \) and \( I_3 \) are vanishing lines given by

\[
I_1 = v_1 + v_2, \quad I_2 = v_2 + v_3, \quad \text{and} \quad I_3 = v_3 + v_1.
\]

In the warped image, setting the vanishing points \( v_1, v_2 \) and \( v_3 \) as \( v_1 = [1 \; 0 \; 0]^T, \; v_2 = [0 \; 1 \; 0]^T, \) and \( v_3 = [m_1 \; m_2 \; m_3]^T \)

\[
\text{gives the IAC \( \omega_{pp} \) of the FP camera}
\]

\[
\omega_{pp} = \begin{bmatrix}
\gamma^2 m_2^2 & 0 & -\beta m_1 m_3 \\
0 & \beta^2 m_1^2 & -\gamma m_2 m_3 \\
-\beta m_1 m_3 & -\gamma m_2 m_3 & \alpha^2 + \beta^2 m_1^2 + \gamma^2 m_2^2
\end{bmatrix}.
\]

Because the ICPs of a plane are on the IAC,

\[
I_{pp} \omega_{pp} I_{pp} = 0, \quad J_{pp} \omega_{pp} J_{pp} = 0,
\]

and we can find the relation:

\[
\frac{R_{pp}^2}{R_m^2} = \frac{\gamma^2}{\beta^2},
\]

which means that the ratio of \( \beta \) and \( \gamma \) is the same as that of \( R_{pp} \) and \( R_m \).

By decomposing (8), the fronto-parallel camera matrix is:

\[
K_{pp} = \begin{bmatrix}
1/R_{pp} & 0 & m_1 \\
1/R_m & m_2 & m_3
\end{bmatrix}
\]

defined up to scale. Note that \( m_1, m_2 \) and \( m_3 \) are elements of \( v_3 \), which is an orthogonal vanishing point with respect to the reference plane.

Consequently, the camera matrix \( K_{pp} \) expresses a camera with zero-skewed pixels and whose pixel aspect ratio is equal to the ratio between an aspect ratio of the reference rectangle and that of the corresponding warped rectangle. The principal point of the camera is expressed with the scaled orthogonal vanishing point \( v_3 \) and the scale plays a role of the focal length. Note that the FP camera matrix is determined with scene information.

Assuming that there is a reference parallelogram rather than a rectangle, the FP camera matrix would be generalized to

\[
K_{pp} = \begin{bmatrix}
1/R_{pp} \cot \theta/R_m & m_1 \\
1/R_m & m_2 & m_3
\end{bmatrix}
\]

where \( \theta \) is an angle between two parallel line sets.

2.3. **Motion constraints of fronto-parallel cameras**

We can find the projection matrix of the FP camera directly from (2), (3) and (11) as

\[
P_{pp} = H_{pp} K [r_1, r_2, r_3, t]
\]

\[
= (K [r_1, r_2, t] \text{diag}(a, b, c))^{-1} K [r_1, r_2, r_3, t]
\]

\[
= \text{diag}(1/a, 1/b, 1/c) [e_1, e_2, e_3] [r_1, r_2, t]^{-1} [r_1, r_2, e_3],
\]

where \( e_1, e_2 \) and \( e_3 \) are \([1 \; 0 \; 0]^T, [0 \; 1 \; 0]^T, \) and \([0 \; 0 \; 1]^T, \) respectively.

It is worth noting that the rotation matrix is an identity matrix in the FP camera, which means that the coordinate axes are aligned with those of the reference plane. Intuitively, we can present two motion constraints of multiple FP cameras that are derived with an identical scene. From the fact that the FP camera has an identity rotation matrix, which means that the camera axis is aligned with that of the reference plane, it is straightforward to show the following two motion constraints [24]:

**Result 1.** Assume that there is a parallelogram in 3D space and two views \( i \) and \( j \) that see the parallellogram. The FP cameras \( P_{pp} \) and \( P_{pp} \), which are derived from an identical parallelogram in 3D, are pure translating cameras.

**Result 2.** Assume that there are two static parallelograms \( I \) and \( J \) in general position in 3D space that are seen in two view \( i \) and \( j \). The relative rotation between two FP cameras of the view \( i \) is identical to that between two FP cameras of the view \( j \).

3. **Camera calibration by transferring scene constraints**

A FP camera matrix is defined from the parameters of scene constraints. Because the warped image is obtained by a simple planar homography, the IAC of a physical camera and that of the derived FP camera are transformed to each other with the planar
homography. If we have sufficiently many scene constraints from a single image (or equivalently images from a static camera), all the FP cameras from the physical camera are regarded as rotating cameras. Therefore, estimating the intrinsic parameters of the physical camera is equivalent to estimating the IAC of the physical camera from the constraints of FP cameras, like the autocalibration of rotating cameras [18]. This means that camera calibration using geometric constraints of rectangles is equivalent to the autocalibration of rotating cameras whose intrinsic parameters are varying.

However, it is difficult to find sufficiently many scene constraints from a single image. To deal with more general cases, we generalize our framework to IAC-based autocalibration. The IACs of cameras are transformed with the infinite homographies between cameras [1].

Fig. 2 sketches the generalization of the proposed framework to freely moving cameras whose intrinsic parameters are not constrained. Because all the FP cameras of one physical camera are rotating with known image transformations, this group of cameras has only five degrees of freedom (DOF). If there is another group of rotating cameras, the DOF of the whole system increases by five, unless the static camera assumption is used. However, if we can find just one transformation between the two groups, as in Fig. 2, the DOF of the whole system remains at five. The transformation should be an infinite homography between two cameras.

The problem can be regarded as IAC-based autocalibration with known infinite homographies. Recall that the autocalibration of rotating cameras is a special case of IAC-based autocalibration. The scene constraints in the original views can be converted into those of rotating cameras and the newly derived rotating cameras are transferred to the first camera with an infinite homography. We can make a linear system on the first camera as

\[ A \omega = 0, \]

where \( A \) is the coefficient matrix of the camera constraint equations, and \( \omega \) is a vectorized IAC of the first camera. To solve this problem, five or more independent camera constraints are needed, including ones derived from scene geometry, and ones from the original cameras, if exist. In the proposed algorithm, there is no difference between constraints from the scene and ones from the cameras.

The remaining problem is to estimate an infinite homography between views, which is one of the most difficult parts in stratified autocalibration without prior knowledge about cameras [16,1]. In order to estimate the infinite homography using scene constraints, we need vanishing points of four different directions, which are very hard to get, or we should have three corresponding vanishing points with a fundamental matrix between views.

We will show that it is much easier to estimate the infinite homography between FP cameras than between physical cameras directly, without prior estimation of the epipolar geometry between views. In addition, we can easily get the fundamental matrix from the estimated infinite homography by using the properties of the FP cameras.

### 3.1. Parameterization of the infinite homography

Assume that there are two views \( i \) and \( j \) that see an unknown parallelogram. FP camera matrices \( K_{\text{FP}_i} \) and \( K_{\text{FP}_j} \) are expressed as

\[
K_{\text{FP}_i} = \begin{bmatrix} 1/R_{\text{FP}_i} & \cot \theta / R_{\text{FP}_i} & m_i^1 \\ 1/R_{\text{FP}_i} & m_i^2 \\ m_i^3 \end{bmatrix}, \quad K_{\text{FP}_j} = \begin{bmatrix} 1/R_{\text{FP}_j} & \cot \theta / R_{\text{FP}_j} & m_j^1 \\ 1/R_{\text{FP}_j} & m_j^2 \\ m_j^3 \end{bmatrix},
\]

where \( \mathbf{v}_i = [m_i^1 \ m_i^2 \ m_i^3]^{\top} \) and \( \mathbf{v}_j = [m_j^1 \ m_j^2 \ m_j^3]^{\top} \) represent scaled orthogonal vanishing points in the warped images of view \( i \) and \( j \), respectively. Note that we can set \( R_{\text{FP}_i} \) as one, and \( R_{\text{FP}_j} \) is the same in \( K_{\text{FP}_i} \) and \( K_{\text{FP}_j} \) because the parallelogram is identical.

An infinite homography \( H_{\text{FP}}^{ij} \) between the two FP cameras \( K_{\text{FP}_i} \) and \( K_{\text{FP}_j} \) is

\[
H_{\text{FP}}^{ij} = K_{\text{FP}_j} R_{\text{FP}_j} K_{\text{FP}_i}^{-1} = \begin{bmatrix} 1 & 0 & h_x \\ 0 & 1 & h_y \\ 0 & 0 & 1 \end{bmatrix}, \quad (14)
\]

which has three parameters \( h_x, h_y \), and \( h_0 \), because the two FP cameras \( K_{\text{FP}_i} \) and \( K_{\text{FP}_j} \) are purely translating to each other. One can see that the form of (14) is preserved even when \( \theta \neq 90\degree \), which means that the cameras see a parallelogram, not a rectangle. Thus, all the following derivation can be applied to the parallelogram case.

Applying the infinite homography \( H_{\text{FP}}^{ij} \) between two FP cameras \( i \) and \( j \) gives us

\[
\omega_{\text{FP}_i} = H_{\text{FP}}^{ij} \omega_{\text{FP}_j},
\]

where \( \omega_{\text{FP}_i} \) and \( \omega_{\text{FP}_j} \) are the IACs of FP cameras of view \( i \) and \( j \), respectively. Through conic transformation, we have

\[
\mathbf{a}_i = H_{\text{FP}}^{ij} H_{\text{FP}}^{ji} \mathbf{a}_j,
\]

and the infinite homography \( H_{\text{FP}}^{ij} \) from view \( i \) and view \( j \) is

\[
H_{\text{FP}}^{ij} = H_{\text{FP}}^{ij} H_{\text{FP}}^{ji} H_{\text{FP}}^{ji} H_{\text{FP}}^{ij}.
\] (15)

This result shows that the infinite homography is expressed with fronto-parallel warping matrices \( H_{\text{FP}}^{ij} \) and \( H_{\text{FP}}^{ji} \), and an infinite homography \( H_{\text{FP}}^{ij} \) between FP cameras. Note that there are no assumptions about cameras such as a static camera or zero-skewed pixels.

### 3.2. Linear estimation of infinite homography from two arbitrary rectangles

If a captured scene contains two arbitrary parallelograms whose aspect ratios are unknown, the infinite homography is estimated linearly using a parameterization in (15). The unknowns are three parameters of \( H_{\text{FP}}^{ij} \).

Assume that there are two views \( i \) and \( j \) that see two unknown rectangles \( I \) and \( J \) in general positions. Two infinite homographies with respect to the two rectangles are given as:

\[
H_{\text{FP}}^{ij} = H_{\text{FP}}^{ij} H_{\text{FP}}^{ji} H_{\text{FP}}^{ji} H_{\text{FP}}^{ij} H_{\text{FP}}^{ji},
\]

where \( H_{\text{FP}} \) means a fronto-parallel warping matrix of the view \( i \) w.r.t. the rectangle \( I \). Because these two should be identical, the constraint equation of
\[
\rho H_{i1}^{-1} H_{i2}^{-1} H_{i3} = H_{j1}^{-1} H_{j2}^{-1} H_{j3},
\]

should hold, where \( \rho \) is a proper scale factor.

The unknowns are the parameters of \( H_{i1}, H_{i2}, \) and \( H_{i3} \), and a scale factor \( \rho \). Each infinite homography has three parameters as in (14), so we have nine equations with seven unknowns, which gives a closed-form linear solution. Note that we do not use any metric measurements such as lengths and aspect ratios of scene parallelograms.

3.3. Affine reconstruction of cameras from the estimated infinite homography

The estimated infinite homography has information about the epipolar geometry between two views. We can extract affine projection matrices of the two cameras directly from the parameterization of the infinite homography.

For two views that have images of a planar parallelogram with an unknown aspect ratio, they can be transformed into the commonly warped image. Assume that the fronto-parallel projection matrices are given as
\[
P_{FP} = [K_{FP} \ e_1] \quad \text{and} \quad P_{FP} = [K_{FP} \ e_3]
\]
as in (13).

From (12), the camera centers \( C_{FP} \), and \( C_{FP} \), are given by
\[
C_{FP} \sim [R_{FP}(m_1 - m_2 \cot \theta) \ R_{m_1} m_1^2 - 1 \ m_1^T]^T
\]
and
\[
C_{FP} \sim [R_{FP}(m_1 - m_2 \cot \theta) \ R_{m_1} m_1^2 - 1 \ m_1^T]^T,
\]
respectively.

The epipoles \( e_{FP} \) and \( e_{FP} \) are
\[
e_{FP} = P_{FP} C_{FP} \sim m_i - m_i \quad \text{and} \quad e_{FP} = P_{FP} C_{FP} \sim m_i - m_i \sim e_{FP}
\]
Consequently, the two epipoles defined in the corresponding warped images are identical defined up to scale. Fig. 3 shows that the epipole between two warped views is on the line through the scaled orthogonal vanishing points \( m_i \) and \( m_n \).

A fundamental matrix induced by a planar homography \( H \) is given as
\[
F = [e', \ e, \ H],
\]
where \([\ ]\) is a skew-symmetric matrix [1]. In this case, the planar homography can be chosen as \( I_{3 \times 3} \), because the two images are warped with respect to the same reference plane. Because the epipole is simply \( m_i - m_i \), the resulting fundamental matrix is
\[
F_{FP} = [m_i - m_i] = [m_i] - [m_i].
\]
The above derivation is equivalent to that of a fundamental matrix in plane + parallax approaches [25,26]. The difference is that we describe the epipolar constraint of two views with in-image features, which are the scaled orthogonal vanishing points, by adopting a reference parallelogram.

The common epipole \( e_{FP} \) can be directly estimated from the infinite homography between two FP cameras. (14) can be rewritten as
\[
\begin{bmatrix}
1/R_{FP} \cot \theta / R_{m} m_1^T \\
1/R_{m} m_2^T \\
0 0 1
\end{bmatrix}
\]
and we can find the equation of
\[
m_j = [m_1 + m_2 h_x \ m_1^2 + m_2 h_y \ m_2^2 h_z ]^T
\]
\[
= [m_1 + m_2 h_x \ m_1^2 + m_2 h_y \ m_2^2 h_z - m_2 ]^T
\]
(17)

Because the common epipole is \( m_i - m_i \) in the commonly warped image,
\[
e_{FP} = [h_x \ h_y \ h_z - 1]^T
\]
defined up to scale.

Once \( H_{i1} \) is estimated between two FP cameras, we can obtain the resulting affine projection matrices of the physical cameras \( i \) and \( j \) using transformation matrices as
\[
P_i = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad P_j = [H_{i1} H_{FP}, H_{FP}, 1] e_{FP}
\]
without explicit projective reconstruction and estimation of a fundamental matrix. Affine reconstruction of scene is possible using simple triangulation between two views, because the affine projection matrices are determined when estimating the infinite homography between two FP cameras.

4. Experiments

In this section, we show various experiments for verifying the feasibility of the proposed method with simulated and real data.

4.1. Accuracy analysis by simulation

The performance of the proposed framework depends on the accuracy of the infinite homography estimation between views, because it transfers the scene constraints from one to the others using the infinite homography. The performance of the algorithm was analyzed to estimate the infinite homography in various situations. Three views were generated having two arbitrary rectangles in general poses and Gaussian noise whose standard deviation is 0.5 pixels was added on the corner of the rectangles. It is assumed that the images see two real rectangles whose aspect ratio are not known for the proposed method. For comparison, the camera was calibrated using the plane-based calibration method using ICPs [7] with the real aspect ratio of the rectangles. Note that the plane-based methods need more information than the proposed method because the metric structures of the scene planes are needed to estimate the ICPs of the planes. In Fig. 4, RMS errors of focal length estimation are depicted, with 500 iterations.

First, the effect of the angle between the two model planes was tested. Poles of cameras are selected from those of the real multicamera system. The average area of the projected rectangles is about 30% of the image size. Fig. 4(a) shows the performance of the proposed algorithm in rotating the plane in 3D along the X axis.

![Fig. 3. Epipolar geometry in multiple fronto-parallel warped images. Geometry in warped image is basically the same as the plane + parallax approaches. By determining the orthogonal axis of the reference plane, the epipole and the fundamental matrix are estimated from the vanishing points \( m \) and \( n \) (see text).](image-url)
while one of the plane is fixed. The solid line is for the proposed method, and the dashed line for the plane-based method. For all interim angles, the accuracy of the calibration is under 4% error with respect to the true value. In $0^\circ$ and $180^\circ$, the proposed algorithm does not work, because the two scene planes are parallel and the infinite homography cannot be estimated. This is one of the degenerate cases. At $70^\circ$, the performance of the calibration degrades in both methods because the rotating plane is almost orthogonal to the image plane of one camera, and all the features lie within a narrow band. This is another degenerate case, and the calibration is not significantly degraded in the other cases. The proposed method shows a worse performance comparing with the plane-based methods using all the six scene planes. This is understandable as it is assumed that the real aspect ratios of the scene rectangles are unknown. In this situation, the plane-based method would not work.

Second, the effects of planar rotation of the model plane were analyzed. In this experiment, the camera and the scene planes are fixed while the rectangles on the planes are rotating in the plane. Fig. 4(b) shows the effects of the planar rotation of the world plane. In this case, the performance is tested as a function of the differences of the directions of the orthogonal axis of two rectangles. It is concluded that the directions of the model axis do not affect the performance of the algorithm.

Third, the effects of area of the projected rectangles in input images were tested. In this case, the poses of cameras remain the same with those of the first experiment and the angle between two planes are set to $90^\circ$. Fig. 4(c) shows the performance as a function of the area of the used rectangles in images. Naturally, when the imaged rectangle is larger, the performance of the estimation algorithm is better. The performance is degraded exponentially as the area decreases. The proposed algorithm works robustly if we have projected rectangles larger than 10% of the whole images averagely, as shown in Fig. 4(c). It is interesting that the proposed algorithm outperforms the plane-based method when the average area of the projected rectangles became larger.

In the fourth experiment, we used parallelograms rather than rectangles to test the performance of the proposed algorithm under various included angles of parallelograms. The parallelograms used in this experiment are skewed squares by the included angle, and all the sides have the same lengths. As in Fig. 4(d), the proposed algorithm gives more accurate results when the included angle goes to $90^\circ$. We can expect this because the areas of the projected parallelograms are maximized if the parallelograms are rectangles. However, one can notice that the performance is not degraded so much even with the parallelograms whose included angles are $50^\circ$ in Fig. 4(d). The proposed algorithm works without any information about the length ratios or the included angles of the parallelograms. The only information is that the cameras see parallelograms.

4.2. Camera calibration using static cameras

The framework was applied to real images. Fig. 5 shows input images containing two arbitrary rectangles, captured with a SONY DSC-F717 camera in 640 × 480 resolution. The white solid lines show the manually selected projected rectangles. The selected rectangles have different aspect ratios and the metric properties of each are unknown. We need only three images captured from different positions. Note that there are small projec-
tive distortions on some projected rectangles. The estimated camera matrix is

\[
K_{\text{estimated}} = \begin{bmatrix}
728.5874 & 28.4393 & 352.8952 \\
0 & 718.3721 & 285.1658 \\
0 & 0 & 1.0000 
\end{bmatrix}
\]

and a result from Zhang’s calibration method[5] with six metric planes is

\[
K_{\text{Zhang}} = \begin{bmatrix}
721.3052 & 2.7013 & 335.3498 \\
0 & 724.9379 & 247.3248 \\
0 & 0 & 1.0000 
\end{bmatrix}
\]

This shows that calibration using the proposed method can be applicable to real cameras by just tracking two arbitrary rectangles in general poses.

4.3. Reconstruction using image sets including pictures of pictures

Using the proposed framework, metric reconstruction of scenes is possible from image sets containing pictures of pictures because the proposed framework does not need to have constraints on all cameras. Fig. 6 shows three input images that are captured by different cameras with unknown intrinsic parameters. Note that the second and the third images are pictures of pictures, so we cannot make any assumptions about the capturing devices of the original pictures. Additionally, as we use different cameras at different zoom levels with autofocusimg, there is no available camera information. We only assume that the first camera has square pixels and provides two linear constraints. We do not constrain the parameters of the second and third cameras.

To estimate the infinite homographies between views, we apply the algorithm in Section 2.3 with the corresponding two rectangles depicted with dashed lines. Note that any kind of algorithm to estimate infinite homographies can be applied. The solid rectangles are used to supply additional scene constraints. The only information we use is the zero-skew constraint of each rectangle. Therefore, we use five independent camera constraints in all sequences. We select the features and boundaries of the building manually.

Affine reconstruction of cameras is done by using the method in Section 3.3, directly from the estimation of the infinite homography. After recovering the metric projection matrices of the cameras, the structure of the scene is computed by simple

---

**Fig. 5.** Input images for calibration using the proposed method.

**Fig. 6.** Input images for verification of the proposed flexible algorithm. The first image is captured with a commercial digital camera, but the second and the third images are pictures of pictures. Dashed rectangles are used to estimate infinite homographies between views, and solid rectangles are used as scene constraints. Note that the solid rectangles do not correspond in every view.
triangulation. The correspondences between views are found manually. Fig. 7 shows a reconstructed VRML model built using the proposed framework from various viewpoints. The orthogonality and parallelism between features are well preserved. The angle between two major walls of the building was estimated to be 90.69° using the proposed method, and the true angle measured 90° using SICK laser range finder as shown in Fig. 7 as blue lines. Note that we do not constrain the right angle condition at all.

Additionally, note that the scene constraints depicted in solid lines are not reconstructed since their correspondences do not exist in the other views. This means that the scene constraints can be used freely, although correspondences of the features cannot be

Fig. 8. Input images of the pagoda Sokkat'ap downloaded from the Internet.
found because the proposed framework is defined only in the image space.

This calibration algorithm can be applied to fully unknown intrinsic parameter cases, such as scanned photographs and pictures of pictures.

Another example includes images downloaded from the Internet. The Internet contains raw images from digital cameras, images processed with various image-processing tools, pictures of pictures, scanned images of books or printed materials, etc. Fig. 8 shows the input images used for this example. The first one is taken with a commercial digital camera, assuming a zero-skew camera whose principal point is at the center of the image. The second one is an image projectively warped with an image-processing tool. The third one is a scanned image from a book. We cannot assume anything about the cameras used to capture the second and the third images, because there is no information about the sensor used or the transformations. The fourth and fifth images are pictures from digital cameras, but are not used to calibrate the cameras; they are only used to reconstruct the rear of the pagoda after recovering the motion of the cameras in metric space. Note that it is nontrivial to estimate three vanishing points in the images because some parallel lines are quite parallel in the images, which gives difficulties in estimating vanishing points in images.

Fig. 9 shows a resulting 3D model from the input images and the motions of the cameras. To calibrate the cameras, we use two rectangles within the pagoda. Note that the orthogonality between planes is well preserved, although we do not constrain that condition explicitly. The position and viewing directions of the cameras are also estimated well.

5. Conclusion

Conventional camera calibration methods based on the camera constraints are not applicable to images containing transformed images, such as pictures of pictures or scanned images from printed pictures. In this paper, a method is presented that provides effective calibration and metric reconstruction using geometrically warped images.

Because there is no information about the cameras, we propose a method to convert scene constraints in the form of parallelograms into camera constraints by warping images. The calibration using the scene constraints are equivalent to autocalibration of cameras whose parameters are constrained by the scene parameters. Practically, planar features such as arbitrary rectangles or parallelograms are used to warp the image to constrain the cameras of the warped images. The imaginary cameras, called fronto-parallel cameras, are always parallel to the scene planes containing the planar features. The intrinsic parameters of the fronto-parallel cameras are constrained by the parameters of the used planar features.

A linear method is also proposed to estimate an infinite homography linearly using fronto-parallel cameras without any assumptions about cameras. The method is based on the motion constraints of the fronto-parallel cameras. Once we have an infinite homography between views, it is very straightforward to transfer camera constraints derived from scene constraints to another camera, and we can utilize all the scene constraints in calibrating one camera in the camera system. In addition, we show that the affine reconstruction of cameras is possible directly from the infinite homography derived from images of parallelograms.
The feasibility of the proposed framework has been shown by calibrating cameras and metric reconstructing the scenes with synthetic and real images containing pictures of pictures and images downloaded from the Internet.

References