Abstract In this paper we present a deep analysis of the hybrid two-view relations combining images acquired with uncalibrated central catadioptric systems and conventional cameras. We consider both, hybrid fundamental matrices and hybrid planar homographies. These matrices contain useful geometric information. We study three different types of matrices, varying in complexity depending on their capacity to deal with a single or multiple types of central catadioptric systems. The first and simplest one is designed to deal with para-catadioptric systems, the second one and more complex, considers the combination of a perspective camera and any central catadioptric system. The last one is the complete and generic model which is able to deal with any combination of central catadioptric systems. We show that the generic and most complex model sometimes is not the best option when we deal with real images. Simpler models are not as accurate as the complete model in the ideal case, but they provide a better and more accurate behavior in presence of noise, being simpler and requiring less correspondences to be computed. Experiments with simulated data and real images are performed. To show the potential of these approaches we develop two applications. The first is the successful matching between perspective images and hyper-catadioptric images using SIFT descriptors. In the second one, using only the hybrid fundamental matrix and the hybrid planar homography we compute the metric localization of the perspective camera inside the catadioptric view in an indoors environment.

1 Introduction

In recent years the use of omnidirectional catadioptric systems, which combine lenses and mirrors, has increased among the computer vision community. The advantages of such systems are their wide field of view and the central single view point property. The former allows to minimize the possibility of fatal occlusions and partial views, helping the tracking of features. The latter allow us to calculate easily the directions of the light rays coming into the camera [11], helping the computation of 3D information from multiple views. In [1] an analysis of this kind of systems is presented and it describes those systems which have the single view-point property. Among these the most popular are the hyper-catadioptric system, composed of a hyperbolic mirror and a perspective camera, and the para-catadioptric system which is the composition of a parabolic mirror and an orthographic camera.

The combination of central catadioptric views with perspective ones is important since a single catadioptric view contains a more complete description of the scene, and the perspective image gives a more detailed description of the particular area or object we are interested in. Some areas where the combination of these cameras has an important role are localization and recognition, since a database of omnidirectional images would be more
representative with fewer points of view and less data, and perspective images are the simplest query images [15]. In visual surveillance [7] catadioptric views provide coarse information about locations of the targets while perspective cameras provide high resolution images for more precise analysis. Active vision [12] is another area where this mixture is naturally implemented; an omnidirectional camera provides peripheral vision while a controllable perspective camera provides foveal vision. Recently, in [4] a structure from motion approach mixing omnidirectional and conventional cameras is presented. In [13] a stereo system is presented mixing omnidirectional and perspective views by creating virtual perspective cameras which rectify the catadioptric view. In [22] virtual cameras are also used to match catadioptric views.

The two-view relations between uncalibrated perspective cameras has been widely studied, for example, in [5] the fundamental matrix is computed from planar homographies. In this work we are particularly interested in the two-view relations between uncalibrated catadioptric and conventional views working directly in the raw images. In the literature very few approaches are presented to compute hybrid two-view relations mixing uncalibrated catadioptric and conventional cameras. These approaches have in common the use of lifted coordinates to deal with the non-linearities of the catadioptric projection. Most of these approaches are presented theoretically and with simple experiments. In this work we tackle this situation by performing a deep evaluation of such approaches using simulated data and real images, extending [17] and including the analysis of the hybrid homography. We observed that the approaches with less parameters are less sensitive to the presence of noise, giving similar results to the theoretically more complete approaches.

To compute the two-view geometry we require pairs of corresponding points between the views. These correspondences are built from previously detected relevant features. Perhaps the most used extractor is the SIFT [14]. However, if SIFT features extracted in an omnidirectional image are matched to features extracted in a perspective image the results are not good, this is because the SIFT descriptor is scale invariant but not projective invariant. We show that with a simple flip of the omnidirectional image, SIFT matching can still be useful, requiring neither image rectification nor panoramic transformation.

Two applications are developed to show the potential of the hybrid two-view geometry. In the first one the automatic matching between a conventional image with a hyper-catadioptric image is performed. The second one performs the metric localization of a conventional camera inside the catadioptric view [18]. It uses the hybrid fundamental matrix to compute the location of the camera in the omnidirectional image and the hybrid homography to map it to the ground plane.

The rest of the paper is organized as follows. In Section 2 we present related work. In Section 3 we present the projection model and the lifted coordinates used to deal with the central catadioptric systems. In Section 4 we introduce the two-view geometry relations between catadioptric and perspective views. In Section 5 we present experiments with synthetical and real data. In Section 6 we show the two applications which use the hybrid two-view geometric approaches. Finally in Section 7 we present the conclusions of this work.

2 Related Work

The multiview geometry problem for conventional cameras has been studied for a long time [11]. In the case of central catadioptric systems this multi-view geometry has been studied until recent years. Some approaches require the calibration of the systems. Svoboda and Pajdla [21] study the epipolar geometry for central catadioptric systems. Based on the model of image formation they propose the epipolar geometry for catadioptric systems using elliptic, hyperbolic and parabolic mirrors.

Recently some works have been developed considering the hybrid epipolar geometry for different combinations of uncalibrated central catadioptric systems, including the pin-hole model. These use a generic central catadioptric model [10] and lifted coordinates. To a lesser extent homographies have also been studied using uncalibrated catadioptric systems. They establish a relation between the projections on the omnidirectional images of 3D points that lie on a plane.

In a seminal work Sturm [19] proposes two models of hybrid fundamental matrices, a $3 \times 4$ fundamental matrix to relate a para-catadioptric view and a perspective view and a $3 \times 6$ fundamental matrix to relate a perspective view and a general central catadioptric view. He also describes the $3 \times 4$ plane homography which represents the mapping of an image point in a para-catadioptric view to the perspective view. This mapping is unique, unlike the opposite one that maps a point in the perspective image to two points in the para-catadioptric image. He also shows how to use the homographies and fundamental matrices to self-calibrate the para-catadioptric system. All these methods use only lifted coordinates of the points in the omnidirectional image. In [15] Menem et al. propose an algebraic constraint on corresponding image points in a perspective image and a circular panorama. They use a lifting from
3 Projection Model and Lifted Coordinates

According to Geyer and Daniilidis [10], all central catadioptric cameras can be modeled by a unit sphere and a perspective projection, such that the projection of 3D points can be performed in two steps (Fig. 1). First, one projects the point onto the unit sphere, obtaining the intersection of the sphere and the line joining its center and the 3D point. There are two such intersection points, which are represented as $s_\pm$. These points are then projected in the second step, using a perspective projection $P$ resulting in two image points, $q_{\pm}$. One of which is physically true. This model covers all central catadioptric cameras, encoded by $\xi$, which is the distance between the perspective projection center and the center of the sphere, and $\psi$ which is the distance between the center of the sphere and the image plane. We have $\xi = 0$ for perspective, $\xi = 1$ for para-catadioptric and $0 < \xi < 1$ for hyper-catadioptric systems.

For any 3D point $Q$ its projection is defined as follows. The two intersection points of the sphere and the line joining its center and $Q$, are $s_\pm \sim (Q_1, Q_2, Q_3, \pm \sqrt{Q_1^2 + Q_2^2 + Q_3^2})^T$. Their images in the image are

$$q_\pm \sim P s_\pm, \quad \text{where} \quad P \sim K \begin{pmatrix} I & 0 \\ 0 & -\xi \end{pmatrix},$$

(1)

with $I$, a $3 \times 3$ identity matrix. Giving the final definition of

$$q_\pm \sim K \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \pm \xi \sqrt{Q_1^2 + Q_2^2 + Q_3^2} \end{bmatrix},$$

(2)

The theoretical 2 image points $q_{\pm}$ can be represented in a single geometric object, which is the degenerate dual conic generated by the two points. This conic contains all lines incident to either one or both of these 2 points $\Omega \sim q_{+} q_{-}^T + q_{-} q_{+}^T$.

3.1 Lifted coordinates

The derivation of (multi-)linear relations for uncalibrated catadioptric imagery requires the use of lifted coordinates. They allow to generalize the transformations and multiview tensors from conventional perspective images to catadioptric systems, where the projective invariant entities are quadrics instead of lines. The Veronese map $V_{n,d}$ of degree $d$ maps points of $\varphi^m$ into points of an $m$ dimensional projective space $\varphi^n$, with $m = \binom{n+d}{d} - 1$. Consider the second order Veronese map
V_{2,2}, that embeds the projective plane into the 5D projective space, by lifting the coordinates of point \( q = (q_1, q_2, q_3) \) to

\[
\hat{q} = (q_1^2, q_1q_2, q_2^2, q_1q_3, q_2q_3, q_3^2)^	op \tag{3}
\]

This lifting can be performed by the operator \( \Gamma \) which transforms two \( 3 \times 1 \) vector into a \( 6 \times 1 \) vector

\[
\Gamma(q,q) = (q_1, q_1q_2, q_2q_3, q_1q_3, q_2q_3, q_3^2)^	op
\tag{4}
\]

This lifting also preserves homogeneity and it is suitable to deal with quadratic functions because it discriminates the entire set of second order monomials [2]. As we observe, if \( c = (c_1, c_2, c_3, c_4, c_5) \) represents a conic, its equation \( c_1q_1^2 + c_2q_1q_2 + c_3q_1q_3 + c_4q_2q_3 + c_5q_3^2 = 0 \) can be written as \( \hat{q}^Tc = 0 \).

When the conic \( c \) has the particular shape of a circle, we have \( c_2 = 0 \) and \( c_1 = c_3 \). We then use the simplified lifted coordinates of a point \( q = (q_1, q_2, q_3) \) in a 4-vector defined as

\[
\hat{q} = (q_1^2 + q_2^2, q_1q_2, q_2q_3, q_3^2)^	op \tag{5}
\]

Another useful lifting is \( V_{3,2} \) that transforms a 4-vector \( Q = (Q_1, Q_2, Q_3, Q_4) \) into a 10-vector \( \bar{Q} \). It is defined as follows

\[
\bar{Q} = (Q_1^2, Q_1Q_2, Q_1Q_3, Q_1Q_4, Q_2Q_3, Q_2Q_4, Q_3Q_4, Q_4^2)^	op \tag{6}
\]

3.2 Lifted matrices

Similar to lifted coordinates, lifted matrices are also useful in order to generalize the transformations and multiview tensors from conventional perspective images to catadioptric systems. Let us assume the linear transformation \( L \), which maps points \( x \) and \( \bar{x} \) in points \( Lx \) and \( L\bar{x} \). The operator \( \Lambda \), that lifts transformation \( L \) from the projective plane \( \mathbb{P}^2 \) to the embedding space \( \mathbb{P}^5 \), must satisfy

\[
\Gamma(Lx, L\bar{x}) = \Lambda(L) \cdot \Gamma(x, \bar{x}) \tag{7}
\]

Such operator can be derived by algebraic manipulation

\[
\Lambda([v_1v_2v_3]) = [\Gamma_{11} \Gamma_{12} \Gamma_{22} \Gamma_{13} \Gamma_{23} \Gamma_{33}] \bar{D}
\]

where \( \bar{D} = \text{diag}\{1, 2, 1, 2, 2, 1\} \) and \( \Gamma_{ij} = \Gamma(v_i, v_j) \). The operator \( \Lambda \) maps any \( 3 \times 3 \) matrix \( L \) into a \( 6 \times 6 \) matrix \( \bar{L} \).

4 Hybrid Two-view Relations

In this section we explain the epipolar geometry between catadioptric and perspective images. We also explain the hybrid homography induced by a plane observed in these two types of image. For both two-view relations we show and analyze three different models.

4.1 Hybrid Fundamental Matrix

When we mix perspective and catadioptric images, a point in the perspective image is mapped to its corresponding epipolar conic in the catadioptric image \( c \sim F_p q_p \), while a point in the catadioptric image has to be lifted (the corresponding lifted coordinates are explained later), and it is mapped to its corresponding epipolar line in the perspective image \( l \sim F_q c_q \). In general the relation between catadioptric and perspective images with the fundamental matrix that we call *hybrid fundamental matrix* is established by

\[
q_p^T F_q c_q = 0
\]

whose subscripts \( p \) and \( c \) denote perspective and catadioptric respectively. Using the lifted coordinates explained in section 3.1 we can define different fundamental matrices.

4.1.1 General Catadioptric System

As mentioned before the generic fundamental matrix between two catadioptric images (including perspective ones), is a \( 15 \times 15 \) matrix which uses a double lifting of the coordinates of the points in both the omnidirectional and the perspective image [20]. This lifting represents quartic epipolar curves. Since this matrix is intractable in a practical way we prefer to refer to fundamental matrices easier to compute and that have been successfully applied.

Barreto and Daniilidis [2] propose a \( 6 \times 6 \) fundamental matrix, which is able to deal with different combinations of central catadioptric systems and conventional cameras. This matrix is obtained from the lifted coordinates \( \hat{q}_c \) of points in the omnidirectional and the lifted coordinates \( \hat{q}_p \) of points in the perspective images.

\[
\hat{q}_p^T F_q c_q = 0
\]

This matrix establishes a bilinear relation between the two views, relating a point in the omnidirectional image to a conic in the perspective one \( c_p \sim F_q c_q \). This conic is composed by two lines. These lines are the
forward looking epipolar line and the backward looking epipolar line. These lines can be extracted from an SVD of the epipolar conic. This $6 \times 6$ fundamental matrix is named $F_{66}$ and corresponds to the theoretically correct model.

In [19] Sturm establishes that there exist a simpler formulation to relate a perspective or affine view and a general central catadioptric view. This approach only applies lifted coordinates to the point in the catadioptric image: $\hat{q}_c = (q_1^2, q_1 q_2, q_2^2, q_1 q_3, q_2 q_3, q_3^2)^T$. It is also mentioned that this matrix has the drawback of only working in one direction. In this paper we analyze the behavior of such matrix which is an approximation to the theoretically correct model $F_{66}$. We name this $3 \times 6$ fundamental matrix, $F_{36}$.

### 4.1.2 Para-catadioptric system

The para-catadioptric system is a particular catadioptric system composed by a parabolic mirror and an orthographic camera. In this case the shape of the epipolar conic is a circle and we use the simplified lifting (5) in the coordinates of the points in the omnidirectional image. We name this $3 \times 4$ fundamental matrix, $F_{34}$.

### 4.1.3 Computation of the Hybrid Fundamental Matrix

We use a DLT-like approach [11] to compute the hybrid fundamental matrix. It is explained as follows. Given $n$ pairs of corresponding points $q_c \leftrightarrow q_p$, solve the equations (9) or (10) to find $F_{cp}$. The solution is the least eigenvector of $A^T A$, where $A^T$ is the equation matrix

$$
A^T = \begin{pmatrix}
q_{p1} q_{c1} & \cdots & q_{p1} q_{cm} \\
\vdots & \ddots & \vdots \\
q_{pn} q_{cn1} & \cdots & q_{pn} q_{cnm}
\end{pmatrix}.
$$

The number of pairs of corresponding points $n$ needed to compute the hybrid fundamental matrix depends on the number of elements of the fundamental matrix to be computed. Each pair of corresponding points gives one equation. Therefore 35, 17 and 11 correspondences are required at least to compute the $F_{66}$, $F_{36}$ and $F_{34}$, respectively. It is recommended that $n \gg \text{size}(F_{pc})$.

The number of parameters required to compute these approaches is crucial if we take into account that in wide-baseline image pairs the correspondences are difficult to obtain, and much more in images coming from different sensor types. In this case the $F_{34}$ has a clear advantage over the other two more complex approaches.

### 4.1.4 Rank 2 constraint

The above fundamental matrices are, like for the purely perspective case, of rank 2. If the task we are interested in requires the epipoles, it is mandatory to ensure that the estimated fundamental matrix has rank 2. To impose the rank 2 constraint we have tried two options. One is to enforce this constraint minimizing the Frobenius norm using SVD as explained in [11] which we call direct imposition (DI). The other option is to perform a non-linear re-estimation process minimizing the distance from points in one image to their corresponding epipolar conic or line in the other one, using the Levenberg-Marquardt (LM) algorithm. To guarantee the rank 2 we use a matrix parameterization proposed in [3] which is called the orthonormal representation of the fundamental matrix. This approach was originally applied to $O(3)$ matrices and we adapt it to $F_{34}$, $F_{36}$ and $F_{66}$.

### 4.1.5 Computing the epipoles

The process involved in the computation of the epipoles from the three tested hybrid fundamental matrices, $F_{34}$, $F_{36}$ and $F_{66}$ is based on the computation of their corresponding null-spaces.

The hybrid $F_{34}$ matrix has a one-dimensional left null-space and one right null vector. This one corresponds to the epipole in the perspective image. The two epipoles in the omnidirectional image are extracted from the left null-space. We have to compute the left null-vectors that are valid lifted coordinates. From these, the epipoles are extracted in a closed form. The 4-vectors must fit the following quadratic constraint:

$$
\hat{q}_c \sim \begin{pmatrix} q_1^2 + q_2^2 \\
q_1 q_3 \\
q_2 q_3 \\
q_3^2 
\end{pmatrix} \iff \hat{q}_{c1} \hat{q}_{c4} - \hat{q}_{c2} - \hat{q}_{c3}^2 = 0
$$

In the case of the $F_{36}$ matrix we follow a similar process. The epipole in the perspective image is also given by the right null vector and the two epipoles of the omnidirectional image are extracted from the left null-space of this matrix. In this case the valid lifted coordinates have to satisfy the following equations:

$$
\hat{q}_c \sim \begin{pmatrix} q_1^2 \\
q_1 q_2 \\
q_2^2 \\
q_1 q_3 \\
q_2 q_3 \\
q_3^2 
\end{pmatrix} \iff \hat{q}_{c1} \hat{q}_{c6} - \hat{q}_{c2}^2 = 0
$$
In the case of the F66 matrix we use the same approach as for F36 to extract the epipoles corresponding to the omnidirectional image from the left null-space of the fundamental matrix. For the epipole in the perspective one a different process is required. We extract it from the null-vector of the degenerate epipolar conic \( \Omega_p \sim F_{66} \hat{q}_p \), projected from a point in the omnidirectional image to the perspective image. This conic contains the two points \( q_+ \) and \( q_- \).

4.2 Hybrid Homographies

Hybrid homographies relate the projections of points that lie on a plane on different types of images. In particular we analyze the homographies that relate omnidirectional and perspective images. Similarly as before with fundamental matrices, we consider three different models. The general model H66 and two simplified models H36 and H34.

4.2.1 Generic Model

From [20] the projection of a 3D point in any central catadioptric system using lifted coordinates can be described by a \( 6 \times 10 \) projection matrix \( P_{cata} \)

\[
\hat{q} \sim P_{cata} \hat{Q}, \quad P_{cata} = \hat{K}_X \tilde{R} \begin{bmatrix} I_6 & T_{6 \times 4} \end{bmatrix}
\]  

(14)

If we assume that the 3D points lie on a plane \( z = 0 \), \( Q = (Q_1, Q_2, 0, 1)^T \) the non-zero elements of its lifted representation is a \( 6 \)-vector \( \hat{Q}_c = (\hat{Q}_1^T, Q_1^2, Q_2^2, Q_1, Q_2, 1)^T \) and the projection matrix reduces to size \( 6 \times 6 \):

\[
H_{66} = \hat{K}_X \tilde{R} \begin{bmatrix} I_{6 \times 6} & T_{6 \times 6} \end{bmatrix}
\]  

(15)

where \( t_i \) is the \( i \)-th column of the matrix \( T \) and \( H_{66} \) is the \( 6 \times 6 \) homography matrix relating the lifting of the 2D coordinates of the points on the plane to their dual conic representation on the image plane \( \omega \) as explained in Section 3.

This homography can also relate the projection of these 3D points in two different images. And in particular in two images acquired with different sensors, a conventional one \( q_p \) and an omnidirectional one \( q_c \).

\[
\hat{q}_p \sim H_{66} \hat{q}_c
\]  

(16)

To compute this homography we use a DLT-like (Direct Linear Transformation) approach. As in the perspective case we need correspondences \( q_p^c \leftrightarrow q_c^c \) between points lying on the plane in the conventional image \( q_p^c \) and in the omnidirectional one \( q_c^c \). From (16) we obtain

\[
[q_p^c] \times H_{66} \hat{q}_c = 0
\]  

(17)

If the \( j \)-th row of the matrix \( H_{66} \) is denoted by \( h_j^T \) and arranging (17) we have

\[
[q_p^c] \times \hat{q}_c = 0
\]  

(18)

These equations have the form \( A^i h = 0 \), where \( A^i \) is a \( 36 \times 6 \) matrix, and \( h = (h_1^T, h_2^T, h_3^T, h_4^T, h_5^T, h_6^T)^T \) is a \( 36 \)-vector made up of the entries of matrix \( H_{66} \). The matrix \( A^i \) has the form

\[
A^i = \begin{bmatrix}
0 & 0 & q_2^2 \hat{q}_c & \cdots \\
0 & -q_3^2 \hat{q}_c & 0 & \cdots \\
q_3^2 \hat{q}_c & 0 & 0 & \cdots \\
0 & q_3 q_2 \hat{q}_c & -q_3 q_1 \hat{q}_c & \cdots \\
-q_3 q_2 \hat{q}_c & q_3 q_1 \hat{q}_c & 0 & \cdots \\
q_2^2 \hat{q}_c & -2 q_2 q_1 \hat{q}_c & q_1^2 \hat{q}_c & \cdots 
\end{bmatrix}
\]  

(19)

and is rank 3, so each correspondence gives 3 independent equations. Thus we need at least 12 correspondences to compute \( H_{66} \) [9].

4.2.2 Simplified Homographies H34 and H36

We also consider two approximations of the hybrid homography. H34 and H36 are two hybrid homographies that map a lifted vector (3) or (5) corresponding to a point in the omnidirectional image \( \hat{q}_c \) to a point in the corresponding plane \( q_p \) in homogeneous coordinates. The former is related to the theoretical model of a para-catadioptric system and the latter considers any central catadioptric system. Similar to (18) we consider the Kronecker product of \( q_p \) and \( \hat{q}_c \). Both homographies are computed using a DLT approach. Since each correspondence gives two equations we require at least 6 correspondences to compute H34 and 9 correspondences to compute H36.
Hybrid Homographies and Fundamental Matrices Mixing Uncalibrated Omnidirectional and Conventional Cameras.

5 Evaluation of the Hybrid Two-view Models

In this section we analyze the behavior of the three fundamental matrices (F66, F36, F34) and the three homographies (H66, H36, H34). We present some experiments performed with synthetic data and real images.

5.1 Simulated Data

We use a simulator which generates omnidirectional images coming from a hyper-catadioptric system and perspective images from a pin-hole model. The two sensors are placed in a virtual volume of $5 \times 2.5 \times 7$ m. width, height and depth, respectively, where points are located randomly ($n \gg 35$) in the case of the fundamental matrix and in planes in the case of the homographies. The perspective camera has a resolution of $1000 \times 1000$ pixels and is located at the origin of the coordinate system. The omnidirectional camera is located to have a good view of the whole scene. We use the sphere camera model [2] to generate the omnidirectional image. We consider two real hyper-catadioptric systems with mirror parameters of $\xi = 0.9662$ (m₁) and $\xi = 0.7054$ (m₂), from two real hyperbolic mirrors designed by Neovision¹ and Accowle², respectively. As a common practice and because we are using lifted coordinates we apply a normalization to the image coordinates where the origin is the image center and the width and height are 1. Once the points are projected we add Gaussian noise, described by $\sigma$, in both images.

5.1.1 Analysis of the Fundamental Matrices

The fundamental matrices are computed using a Levenberg-Marquardt³ non-linear minimization of the geometric distance from image points to epipolar lines and conics using the point to conic distance proposed in [5]. For every $\sigma$ representing the amount of noise we repeat the experiment 10 times to avoid particular cases due to random noise. We show the mean of these iterations. Fig. 2 shows the distances from points to their corresponding epipolar conics and lines as a function of image noise.

---

¹ http://www.neovision.cz
² http://www.accowle.com
³ lsqnonlin function provided by Matlab
Table 1 Epipoles estimated by the three fundamental matrices. DI = direct imposition. LM = Levenberg-Marquardt.

<table>
<thead>
<tr>
<th>Value</th>
<th>DI</th>
<th>LM</th>
<th>DI</th>
<th>LM</th>
<th>DI</th>
<th>LM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_1$</td>
<td>(500,303.07)</td>
<td>(499.24,302.48)</td>
<td>(499.60,303.42)</td>
<td>(500.23,303.73)</td>
<td>(500.07,303.41)</td>
<td>(499.66,303.98)</td>
</tr>
<tr>
<td>$e_2$</td>
<td>(500,200)</td>
<td>(503.31,199.17)</td>
<td>(501.55,201.08)</td>
<td>(501.03,201.27)</td>
<td>(501.52,202.03)</td>
<td>(500.53,201.79)</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0</td>
<td>18.04</td>
<td>0.76</td>
<td>1.16</td>
<td>1.01</td>
<td>0.85</td>
</tr>
</tbody>
</table>

Fig. 4 Behavior of the three fundamental matrices as a function of the mirror parameter ($\xi$): Mean distances from points to epipolar conics in (a) omnidirectional image, and (b) perspective image.

From Fig. 2 we can observe that when there is no noise present in the image the $F_{66}$ shows the best performance, which is expected since $F_{66}$ is the theoretically correct model. This changes when noise increases. In this case the $F_{34}$ and $F_{36}$ show a better performance, being consistent with the noise present in the images. The $F_{34}$ shows a better performance with the mirror $m_1$ since this one is closer to a parabolic mirror, the one the matrix $F_{34}$ was designed to deal with. The residuals of $F_{36}$ are slightly larger than the ones from $F_{34}$. We observe that $F_{66}$ is instable when noise is present in the images. This behavior can be caused by the over-parameterization of the model, the more the parameters the higher the sensitivity to noise; that can also explain the difference between the $F_{36}$ and $F_{34}$.

We also estimate the epipoles from the 3 hybrid fundamental matrices, using the $m_1$ hyper-catadioptric system. In this experiment we add $\sigma = 1$ pixel Gaussian noise to both images. We test the two approaches DI and LM to get a rank 2 matrix (cf. Section 4.1.4). We evaluate the performance of these approaches by the accuracy of the estimated epipoles and by the residual, which is the RMSE of the distances from the points used to compute the fundamental matrix to their corresponding epipolar lines and conics. In Fig. 3 we show the residuals for the three approaches imposing the rank 2 constraint by the direct imposition and by using the LM algorithm with orthonormal representation.

In Fig. 3(a) we can observe that $F_{66}$ is very sensitive to the direct imposition of the rank 2 property with maximum errors of 8 and 12 pixels. This occurs because we are transforming a good solution that passes through the points in the perspective and omnidirectional images into a new matrix of rank 2 that contains the epipoles and makes all epipolar lines and conics to pass through them but far from the points in the corresponding images. This does not occur with the LM algorithm which uses the orthonormal representation because it imposes the rank 2 property and at the same time minimizes the distance between points and epipolar lines and conics, having a maximum error of 3 pixels.

Table 1 shows the epipoles from these two approaches. We can see from it that the three approaches give similar results in computing the epipole but we also observe an increment in the distance from points to conics and the minimization obtained with the LM algorithm, all this as expected. Once more $F_{34}$ shows an interesting behavior giving a small distance to conics even with the DI approach. This adds another advantage to $F_{34}$.

As observed from the previous experiments $F_{34}$ shows a good performance dealing with images coming from a hyper-catadioptric system. In order to test this behavior we designed the following experiment. We modify the mirror parameter $\xi$ from the hyperbolic case ($0 < \xi < 1$) to the parabolic case ($\xi = 1$) [2]. We add $\sigma = 1$ pixels Gaussian noise in both images and repeat the experiment 10 times to avoid bias since we are using random noise. In Fig. 4 we observe that $F_{34}$ can deal better with hyper-catadioptric images when the mirror shape is close to a parabola ($\xi = 1$) and not as good as $F_{66}$ and $F_{36}$ models which are designed to deal with
Hybrid Homographies and Fundamental Matrices Mixing Uncalibrated Omnidirectional and Conventional Cameras.

5.1.2 Analysis of Homographies

We perform experiments computing the hybrid homographies relating a plane in the ground and its projection in an omnidirectional image as well as the projection of a planar scene in both omnidirectional and perspective images. We use the same simulator as in the fundamental matrix case also considering the two different hyperbolic mirrors \( m_1 \) and \( m_2 \). The 3D points are distributed in a planar pattern. This pattern is composed of a squared plane with \( 11 \times 11 \) points and a distance between points of 40cm. The goal of the first experiment is to know the behavior of the three homography approaches in presence of noise. We add different amounts of Gaussian noise described by its standard deviation \( \sigma \) to the coordinates of the points in the omnidirectional image. The DLT algorithm followed by a non-linear step using Levenberg-Marquardt minimizing errors in the image are used to compute the homographies. For every \( \sigma \) we repeat the experiment 10 times in order to avoid particular cases due to random noise. The error of the projected points in the ground plane is shown in Fig. 5. We observe that the three approaches have a similar behavior. When the amount of noise is low the best performance is given by \( H_{66} \), in fact it is the only one that has a zero error when we use noiseless data. When the amount of noise increases the performance of \( H_{66} \) decreases and \( H_{34} \) and \( H_{36} \) remain with smaller errors. This result shows that \( H_{66} \) is more sensitive to noise than the other two approaches. The difference between the errors using the different mirrors is explained because the area occupied by the plane using \( m_2 \) is bigger than the area covered using \( m_1 \). With the \( m_1 \) mirror we have errors of 5.2 cm with the \( H_{66} \) but with the \( m_2 \) mirror this error decreases to 3 mm in both cases with \( \sigma = 1 \) pixel.

The next experiment maps a point from a plane in a perspective image to its projection in the omnidirectional image. In this case we added Gaussian noise to both perspective and omnidirectional image coordinates. We project a point from the omnidirectional to the perspective image, where the map is direct. In Fig. 6 we can observe the experiment using different amounts of Gaussian noise \( \sigma \). Again \( H_{34} \) and \( H_{36} \) give better results than \( H_{66} \) except for the case with a very small amount of noise.

Fig. 5 Comparison between the three approaches to compute the hybrid homography. Using mirrors (a) \( m_1 \), (b) \( m_2 \).

Fig. 6 Noise sensitivity of the hybrid homographies between omnidirectional and perspective images. (a) Using \( m_1 \) and (b) using \( m_2 \).
In the opposite direction, the homography maps a point in the perspective image to a conic in the omnidirectional one. Since the extraction of the corresponding point from this conic is difficult, a way to overcome this problem is to compute a different homography, which maps lifted coordinates in the perspective image to a single point in the omnidirectional one.

From the simulations we observe that the hybrid fundamental matrices and the hybrid homographies with less parameters, \( F_{34} \) and \( H_{34} \) have a good performance even dealing with hyperbolic mirrors. Also they are less sensitive to noise than the theoretically correct and more general models \( F_{66} \) and \( H_{66} \). Note also that simpler models \( F_{34} \) and \( H_{34} \) require fewer point correspondences to be computed and therefore they have advantages in practice.

### 5.2 Experiments with Real Images

We also performed experiments with real images coming from a hyper-catadioptric system and from a conventional camera. We compare the accuracy of the three methods to compute both the hybrid fundamental matrix and the hybrid homography.

#### 5.2.1 Hybrid Fundamental Matrix

In this case we use 70 manually selected pairs of corresponding points to compute the three approaches (\( F_{34} \), \( F_{36} \), \( F_{66} \)). In order to measure the performance of \( F \) we calculate the root mean square error of the geometric distance from each correspondence to its corresponding epipolar conic or line. Table 2 shows these distances for the estimated \( F \) without imposing rank 2 and for the two ways to obtain the rank 2 fundamental matrix. We can observe that when we impose the rank 2 the error increases in particular with \( F_{66} \). With the orthogonal normalization using the LM algorithm \( F_{66} \) gives the best result but with very few difference with alternate models \( F_{34} \) and \( F_{36} \). When we impose the rank 2 constraint we eliminate a few degrees of freedom of the matrix that better adjusts to the data so, the residual error must be worse actually. From Fig. 7 we can observe the epipolar lines and conics from the three approaches. We also observed that a great number of correspondences, larger than the minimum is required to have a reasonable accuracy. Using \( F_{36} \) we obtain good results with 50 (three times the minimum) correspondences. This gives a good reason to use the \( F_{34} \) for further applications.

#### 5.2.2 Hybrid Homographies

In this experiment we select 55 correspondences manually. From these correspondences we use 35 to compute the hybrid homographies \( H_{34} \), \( H_{36} \) and \( H_{66} \). We use the rest as test points. If we want to map points in the opposite direction, i.e., from the perspective image to the omnidirectional one we require the inverse mapping
of matrices $H_{34}$ and $H_{36}$. Since these matrices are not square their computation is not possible. In this order we compute separate homographies to map points in this direction. With this computation we also avoid the extraction of the points from the corresponding conics. With respect to $H_{66}$ it was shown in [9] that two homographies have to be computed, since the inverse matrix does not correspond to the opposite mapping. In Fig. 8 we show the images used to compute the hybrid homographies. In Fig. 9(a) we show the error corresponding to the Euclidean distance between the estimated and the test points in the omnidirectional image. Fig. 9(b) shows the error in the perspective image. We observe that $H_{34}$ and $H_{36}$ have a similar behavior. In both images we also show the corresponding means of the error. $H_{34}$ has the best performance in the perspective image while $H_{36}$ has it in the omnidirectional one. The worst behavior in both images corresponds to $H_{66}$. All the approaches show a considerable error, which can be caused by the small area occupied by the points in the omnidirectional image.

6 Applications using hybrid two-view relations

In this section we present two applications developed using the hybrid two-view relations. The first one consists of a robust automatic matching between uncalibrated hyper-catadioptric images and perspective images. The second one is a localization application of a mobile perspective camera in an indoors environment. It requires the perspective camera, which is constrained to planar motion, to have a part of its field of view in common with previously recorded omnidirectional images.

6.1 Automatic Matching

The first step of a matching process between wide-baseline images is to obtain an initial or putative set of pairs of corresponding features. Reasonable matching of two omnidirectional images using well-known features like SIFT, SURF or MSER has been reported [16, 22]. As it is known catadioptric systems use a reflective surface, which corresponds to the mirror. This mirror produces an inverted image of the scene. It should be noted that a mirror does reverse the forward/backward axis, and we define left and right relative to front and
back. Flipping front/back and left/right is equivalent to a rotation of 180 degrees. When we try to match directly conventional images with catadioptric ones using SIFT features, this effect makes that task impossible to succeed. This is because the SIFT descriptor is based on a spatial discretization of the image patch 

histograms of gradients and these are not invariant to the mirror effect. In an earlier work [17], we observed an improving in the matching. We assumed this improving was caused by the unwarping of the catadioptric image. In a subsequent work [18] we realized that this improvement was mainly caused by the simple vertical flip of the catadioptric image and not by the whole unwarping process. Since this flip undoes the mirror effect, and both images the conventional and the catadioptric one observe the scene in a similar way. Further experiments show that the type of flip performed over the omnidirectional image could be either vertical or horizontal, since SIFT descriptor is invariant to rotation and these two flips are equivalent to 180 degrees. From this situation we conclude that the SIFT descriptor is designed to be scale and rotation invariant and even camera invariant if we consider that flipping a catadioptric image can produce good matches with a conventional image. However the SIFT descriptor is not projective invariant, since the projective mirror effect is responsible for the majority of matching failures.

To evaluate that, we show in Fig. 10 the direct matching between SIFT points from a normal omnidirectional image and a perspective image. In this work we use the SIFT implementation by Vedaldi [23]. The inliers and outliers obtained were counted manually. Table 3 shows that near by all matches are wrong if the omnidirectional image is directly used. Using the unwarped and the flipped transformation of the omnidirectional image we repeat the experiment. In these cases we observe an important increment on the number of correct matches showing both similar results. More results are shown in Table 4.

Note that this initial matching between the perspective and the flipped omnidirectional image has a considerable amount of inliers but also many outliers. This scenario requires a robust estimation technique and a geometric model to detect the inliers and reject the outliers. Depending on the situation, either the hy-

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image1.png}
\caption{Matching directly the SIFT points in the omnidirectional and perspective images. (a) using the unwarped image. (b) using the flipped omnidirectional image. The matches with the normal omnidirectional image are not shown since near by all are outliers.}
\end{figure}
Fig. 11 Some of the images used to test the automatic matching using the fundamental matrix.

![Image](image1.png)

![Image](image2.png)

![Image](image3.png)

![Image](image4.png)

![Image](image5.png)

![Image](image6.png)

Table 4 Numerical results of the hybrid matching using the set of images.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Unwarped Omni SIFT</th>
<th>Flipped Omni SIFT</th>
<th>Persp SIFT</th>
<th>Initial Matches (inliers/outliers)</th>
<th>Robust Epipolar Geometry matches (inliers/outliers)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Unwarped/Flipped</td>
<td>Unwarped/Flipped</td>
</tr>
<tr>
<td>Experiment 1</td>
<td>3251</td>
<td>2867</td>
<td>1735</td>
<td>68/111 76/131</td>
<td>40/8 57/4</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>4168</td>
<td>4172</td>
<td>1528</td>
<td>18/68 21/71</td>
<td>16/7 17/7</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>3280</td>
<td>2967</td>
<td>1682</td>
<td>41/101 33/112</td>
<td>27/9 20/9</td>
</tr>
<tr>
<td>Experiment 4</td>
<td>2275</td>
<td>2208</td>
<td>15658</td>
<td>125/322 164/360</td>
<td>80/5 129/22</td>
</tr>
</tbody>
</table>

Table 3 Output from the SIFT matching using the original, unwarped and flipped omnidirectional image.

<table>
<thead>
<tr>
<th></th>
<th>SIFT points</th>
<th>Matches/Inliers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Omnidiirectional</td>
<td>2877</td>
<td>137/9</td>
</tr>
<tr>
<td>Unwarped image</td>
<td>4182</td>
<td>179/68</td>
</tr>
<tr>
<td>Flipped image</td>
<td>2867</td>
<td>207/76</td>
</tr>
</tbody>
</table>

6.1.1 Hybrid Fundamental Matrix

In a general case where the points are in any part of a 3D scene the fundamental matrix is used. The automatic process to perform the matching between an omnidirectional image and a perspective one, using the hybrid fundamental matrix as geometric constraint is as follows:

1. **Initial Matching.** Scale invariant features (SIFT) are extracted from perspective and flipped omnidirectional images and matched based on their intensity neighborhood.

2. **RANSAC robust estimation.** Repeat for $r$ samples, where $r$ is determined adaptively:
   (a) Select a random sample of $k$ corresponding points, where $k$ depends on what model we are using (if $F_{34}$, $k = 11$, if $F_{36}$, $k = 17$ or if $F_{66}$ $k = 35$). Compute the hybrid fundamental matrix $F_{cp}$ as mentioned before.
   (b) Compute the distance $d$ for each putative correspondence, $d$ is the geometric distance from a point to its corresponding epipolar conic.
   (c) Compute the number of inliers consistent with $F_{cp}$ by the number of correspondences for which $d < t$ pixels, $t$ being a defined threshold.
   Choose the $F_{cp}$ with the largest number of inliers.

3. **Non-linear re-estimation.** Re-estimate $F_{cp}$ from all correspondences classified as inliers by minimizing the distance in both images to epipolar cones and epipolar lines, using a non-linear optimization process.

For the next experiments we have selected the $F_{34}$ model since its performance is similar to the other models and the number of correspondences required to be computed is the smallest. In a RANSAC approach the
number of parameters to estimate is important since it
determines the number of iterations required. In prac-
tice, there is an agreement between the computational
cost of the search in the space of solutions, and the
probability of failure \((1 − p)\). A random selection of \(r\)
samples of \(k\) matches ends up with a good solution if all
the matches are correct in at least one of the subsets.
Assuming a ratio \(ε\) of outliers, the number of samples to
explore is \(r = \frac{\log(1−p)}{\log(1−(1−ε)p)}\). For example using a prob-
ability \(p = 99\%\) of not failing in the random search and
30\% of outliers \((ε)\), 231 iterations are needed to get a
result using the F34. If we use the F36, 1978 iterations
are needed for the same level of confidence. In the case
of the F66 the number of iterations increases to \(1.2 \times 10^6\)
and becomes prohibitive for some applications.

Several omnidirectional and perspective image-pairs
are used to perform the experiment of automatic match-
ing (Fig. 11). We avoid the rank 2 constraint since we
are just concerned about the matching problem. Ta-
ble 4 summarizes the results giving the number of SIFT
features extracted in the two valid versions of omnidi-
rectional images tested (unwarped and flipped), and in
the perspective one. It also shows the quantity of inliers
and outliers in the initial (SIFT matching) and the ro-
 bust matching (hybrid fundamental matrix), using both
the unwarped and the flipped transformations. In Ex-
periment 1 we use images Fig. 11(a) and Fig. 11(e).
The number of SIFT points extracted from the flipped
and the unwarped images are similar. We observe that
the initial matches are similar using the unwarped and
flipped versions of the omnidirectional image, with a
small advantage for the flipped one. These results con-
firm that SIFT is not projective invariant but it works
well with such a distortion of catadioptric cameras. De-
spite the use of either of the transformed omnidirec-
tional images the automatic matching using the hybrid
epipolar geometry is able to filter most of the outliers.
In Experiment 4 the increment in the SIFT features
of the perspective image is caused by the resolution of
the image\((1280 \times 960)\). Fig. 12 shows two examples of
the matching between omnidirectional and perspective
images. The results show that the hybrid epipolar con-
straint eliminates most of the outliers.

6.1.2 Hybrid Homography

Analogous to the matching process using the hybrid
fundamental matrix we can use the hybrid homography
when most of the scene points lie in a plane. An example of the automatic matching using the hybrid homography as a geometrical constraint can be observed in Fig. 13. Fig. 13(a) shows the putative correspondences given by the SIFT matching and Fig. 13(b) after applying the robust matching process, computing a $3 \times 4$ homography.

6.2 A Localization Application

Now we show an application which performs indoors localization of a perspective camera mounted on a mobile platform using a set of previously recorded omnidirectional images (visual memory) taken from fixed positions. The main steps of this application are the two-view geometrical approaches we have studied in this work. These steps are described as follows:

1. **Matching.** A perspective image is acquired with the on-board camera, which is matched with an omnidirectional one contained in the visual memory using the algorithm from section 6.1.1.

2. **Location of the camera in the scene.** From the fundamental matrix we extract the epipole that corresponds to the current position of the on-board perspective camera in the catadioptric view.

3. **Mapping the position in the image to the ground plane.** Using a planar homography ($H_{34}$) previously computed we map the position of the on-board camera (epipole) to the actual ground plane. At this step we have the relative position of the on-board camera in meters with respect to the origin of the ground plane.

Since the homography transformation produces a non-homogeneous uncertainty distribution we studied the error propagation from omnidirectional images to the ground plane. We analyze in particular the hybrid homography $H_{34}$. We consider the Jacobian to translate the error in the coordinates of the omnidirectional points into their corresponding lifted coordinates. Then the error is propagated from the points in the image to the points in the plane. In Fig. 14 we observe the error propagation of two Gaussian distributions, centered on two points in the omnidirectional image with the same variance ($\sigma = 1 \text{pixel}$), located at different distances from the image center.

We observed that the error at the periphery of the image, with a low resolution, propagates a bigger error than a point located close to the center of the image, where the resolution is higher (Fig. 14). We also tested under simulations the importance of the common field of view (CFOV) of the two cameras. The bigger the CFOV the better the estimation of the epipole that corresponds to the position of the on-board camera.

In Fig. 15(a-c) we can see the two main phases of the whole approach, camera location on the omnidirectional image (matching and epipole computing) and the mapping from the image to the ground plane through the planar homography. Fig. 15(d) shows the uncertainty estimation of the position of the on-board camera. We observe that the estimation of the position is better when the camera is in the central area of the omnidirectional image. When the camera is in the periphery the location uncertainty increases as expected.

7 Conclusions

In this work we have presented a deep analysis of the two-view geometry combining a central catadioptric system and a conventional camera. In particular we studied the hybrid epipolar geometry and the hybrid planar homography. We use lifted coordinates to generalize the two-view constraints, well known for perspective image pairs. We selected three approaches to compute the hybrid fundamental matrix $F_{34}$, $F_{36}$ and $F_{66}$ and three approaches to compute the hybrid homography $H_{34}$, $H_{36}$ and $H_{66}$. We performed several experiments comparing the different approaches from the more complex and complete ($F_{66}$, $H_{66}$) to a more particular and simplified one ($F_{34}$, $H_{34}$), that in principle only can deal with a certain type of central catadioptric systems. From the simulation and real data experiments these simplified models obtained better results in presence of noise. We observed that the complete models can deal with any catadioptric system under ideal conditions but these approaches are more sensitive to the presence of noise. We successfully introduce the geometrical constraints in a robust matching process with initial putative matches given by SIFT points computed in the perspective image and the flipped version of the catadioptric one. Moreover, combining the two simplified approaches, the hybrid fundamental matrix to localize the conventional camera inside the field of view of the catadioptric image and the hybrid homography to map points from the catadioptric image to the ground plane, we are able to localize the on-board perspective camera using a visual memory composed of omnidirectional images.

References

Fig. 13 Matching between omnidirectional and perspective image using, (a) putative matches, (b) matches after the robust estimation using the hybrid homography.

Fig. 14 Propagation error in the plane using $H_{34}$. (a) Omnidirectional image with Gaussian distributions in the central area (green) and in the periphery (blue). Theoretical ellipses of uncertainty corresponding to the points close to the center (b) and the points in the periphery (c).

Hybrid Homographies and Fundamental Matrices Mixing Uncalibrated Omnidirectional and Conventional Cameras.

Fig. 15 (a) Matching between omnidirectional and perspective images. (b) Epipolar conics corresponding to the matched points. (c) Epipole trajectory superimposed in one single omnidirectional image. (d) Uncertainties of the epipoles in the ground plane, units are in meters.