The Visual Computer
https://doi.org/10.1007/s00371-024-03293-7

ORIGINAL ARTICLE

Generating a full spherical view by modeling the relation between two fisheye images

María Flores[1] · David Valiente[1] · Adrián Peidró[1] · Oscar Reinoso[1] · Luis Payá[1]

Accepted: 20 January 2024

Abstract
Full spherical views provide advantages in many applications that use visual information. Dual back-to-back fisheye cameras are receiving much attention to obtain this type of view. However, obtaining a high-quality full spherical view is very challenging. In this paper, we propose a correction step that models the relation between the pixels of the pair of fisheye images in polar coordinates. This correction is implemented during the mapping from the unit sphere to the fisheye image using the equidistant fisheye projection. The objective is that the projections of the same point in the pair of images have the same position on the unit sphere after the correction. In this way, they will also have the same position on the equirectangular coordinate system. Consequently, the discontinuity between the spherical views for blending is minimized. Throughout the manuscript, we show that the angular polar coordinates of the same scene point in the fisheye images are related by a sine function and the radial distance coordinates by a linear function. Also, we propose employing a polynomial as a geometric transformation between the pair of spherical views during the image alignment since the relationship between the matching points of pairs of spherical views is not linear, especially in the top/bottom regions. Quantitative evaluations demonstrate that using the correction step improves the quality of the full spherical view, i.e. IQ MS-SSIM, up to 7%. Similarly, using a polynomial improves the IQ MS-SSIM up to 6.29% with respect to using an affine matrix.

Keywords Full spherical view · Dual fisheye images · Image stitching · Fisheye projection

1 Introduction

The research line dedicated to the generation of panoramas is receiving much attention. Such panoramas contain a great quantity of information from the environment, which is advantageous in a wide variety of applications, such as mobile robot localization [1, 2], vehicle panoramic view [3] or driving assistance for power wheelchairs [4–6].

A panorama is a single image that contains a wide-angle view of the environment around a vision system. There are multiple panorama formats, which can be classified according to whether only a portion (e.g. cylindrical format [7]) or the whole sphere (e.g. equirectangular or cube map format [8, 9]) is projected.

In terms of acquisition, there are different alternatives to obtain such wide-angle views [10], such as rotating a camera about its optical center, using an array of cameras pointing toward different directions (and subsequently fusing all the images) or combining a camera with lenses or mirrors. However, none of the above acquisition systems cover the whole sphere. Nowadays, the most interesting configuration to obtain an image with a field of view of 360 degrees horizontally and 180 degrees vertically (i.e. a full spherical view) is to arrange with opposite points of view (back to back) two fisheye lenses with a field of view greater than 180 degrees each one and fusing both images. There are some commer-

María Flores, David Valiente, Adrián Peidró, Oscar Reinoso and Luis Payá have contributed equally to this work.

✉ María Flores
m.flores@umh.es

David Valiente
dvaliente@umh.es

Adrián Peidró
apeidro@umh.es

Oscar Reinoso
o.reinoso@umh.es

Luis Payá
lpaya@umh.es

[1] Institute for Engineering Research, Miguel Hernandez University, Avenida de la Universidad, s/, 03202 Elche, Alicante, Spain

Published online: 29 February 2024

cial cameras with this configuration, such as the Samsung Gear 360 [11], the RICOH THETA S [12] or the Garmin VIRB 360 [13]. In the related literature, several works, such as [14–16], use some of these commercial cameras.

The two images captured by the back-to-back dual-fisheye lens cameras can be fused to obtain high-resolution full-view images. In addition to this, these cameras are lightweight, cheap and small. Notwithstanding that, generating a full spherical view from dual fisheye images is challenging owing to the next features. First, the projection centers of the dual fisheye lenses are displaced (parallax). This fact creates mismatches between matching features and typically produces ghost effects in the common area. Second, the common area between the two fisheye images (which is the peripheral area) is strongly distorted and cannot be directly matched using the raw fisheye images. Third, the pair of images has a limited overlapping field of view, meaning that little information can be extracted from the common region; moreover, this region is the most affected by distortion as stated before.

Many researchers are working on solutions to these challenges and thus trying to obtain high-quality full spherical views. In general, the algorithms to generate a full spherical view from dual fisheye images typically start with a transformation of the fisheye images to a spherical format, followed by a subsequent alignment of this pair of spherical views, and a final merging through some blending technique to remove possible inconsistencies in the final full spherical view. Some algorithms have additional stages, such as a calibration process or a photometric compensation.

In this work, we generate a full spherical view from a pair of fisheye images using different algorithms. They differ in: (a) the procedure to project the fisheye images into the unit sphere surface, and (b) the type of geometric transformation used to align the pair of spherical views.

First, to map from the fisheye image to a unit sphere, we use either a calibration-based method or an equidistant fisheye projection-based method. Also, we propose a correction step, which is applied to the equations of the equidistant fisheye projection. This correction models the relation between both fisheye images. During the transformation of the fisheye images to spherical format, it is assumed that the front and back unit spheres have the same center, and the Z-axes are perfectly aligned and opposite. Nevertheless, these assumptions may be erroneous due, for example, to the existence of a slight offset between the centers. In addition, the relationship between the fisheye pixels and the unit sphere projection is considered the same for both cameras when the equidistant projection is applied. Then, the proposed correction step could address the possible errors caused by all these assumptions. The functions that model the correction step are obtained experimentally and consist in a sine function to relate the angular polar coordinates and a linear relationship between the radial polar coordinates.

Thus, the first stage of the algorithm considers that the visual information to be blended has been captured by two fisheye lenses and is jointly analyzed. This is different to most existing algorithms in which the conversion to spherical projection of each of the fisheye images is performed independently. In such algorithms the common visual information is only aligned in the stitching process.

Second, in previous works, the geometric transformation used to align the pair of spherical views in the 2D plane is typically the affine matrix. This transformation is characterized as a linear mapping. However, the use of fisheye lenses may introduce substantial distortions, mainly at the poles of the spherical views, so the relationship between matching points in these regions may not be linear. Hence, while this type of transformation may be a feasible solution when the most textured regions are in the central area, it may not be an adequate solution when much visual information is present in the top and bottom regions of the views. This fact can make the difference between both spherical views more noticeable, appearing undesired effects in the full spherical view. This is the reason why in this paper we propose to use a polynomial to perform this transformation and we evaluate both types of transformations (I) the polynomial and (II) the affine matrix.

In summary, the main contributions of this paper are:

1. A correction step to apply during the transformation of the fisheye images to unit sphere projection. This correction models the relation between the front fisheye and the back fisheye images.
2. Using a polynomial geometric transformation to align the pair of spherical views.
3. A complete evaluation and comparison of full spherical views generated by different variations of the algorithm and the full spherical view provided by the Garmin VIRB 360 camera with the built-in method.

The remainder of this paper is organized as follows. Section 2 presents a review of related works. In Sect. 3, the algorithm to create a full spherical view from dual fisheye images is described. Also, the proposed correction step that models the relationship between the pair of fisheye images is presented. Section 4 describes the vision system, the dataset used in the experiments, and the variations of the algorithm. In addition to this, the full spherical views have been evaluated in this work by means of qualitative and quantitative assessments and the results are shown in Sect. 4.3. The conclusions and future works are presented in Sect. 5.

2 Related works

Nowadays, visual information is frequently used for the resolution of a wide range of tasks. For instance, Zhang et al.

[17] solved the localization problem as a visual odometry and compared the results employing images taken by three vision systems with different field of view. They concluded that the use of cameras with large field of view is advantageous for localization in indoor scenes.

2.1 Panoramic vision systems

A panoramic image can be obtained by employing different vision systems. These systems can be classified depending on the number of conventional cameras that are utilized and whether they are combined with another element (such as fisheye lens or mirror) or not. For vision systems with more than one camera, image stitching must be used to combine all images into a single one. For instance, Zhang and Xiu [18] propose an image stitching method based on the Human Visual System (HVS) and Scale-Invariant Feature Transform (SIFT) algorithm. Also, this method is based on optimal seamline. Lyu et al. [19] present a survey of image stitching techniques to build a panoramic image.

First, a panoramic image can be obtained with a single camera, capturing a sequence of images with some overlapping field of view while the camera performs a full rotation around the vertical axis. The problem is that all images are not captured in a shot, so, this is not suitable for many applications in which the camera has continuous motion, such as mobile robot navigation. Moreover, this fact can cause some difficulties in the stitching process, especially if the environment is dynamic. In [20], some panorama rig systems are described and an automatic panorama imaging rig system is proposed, whose control is made through a smartphone.

Second, vision systems composed of multiple cameras pointing to different directions with overlapping fields of view constitute an alternative that overcomes the disadvantage of rotating cameras. In this case, several cameras are required to obtain a full 360° view, leading to many areas in the image where stitching effects can occur. The number of cameras can be reduced by combining them with fisheye lenses. For instance, Zhang et al. [21] proposed an algorithm based on optical flow to generate a panorama. The experimental data they used are images provided by a Facebook surround360 and a Insta360 PRO camera. The first camera is composed of 17 lenses, 14 wide angle lenses and 3 fisheye lenses, while the second one has six fisheye lenses.

The field of view in a single image can be increased by combining a conventional camera with a reflecting surface (catadioptric vision system) or with a fisheye lens. Such systems are relatively extended in robotic applications. For example, Flores et al. [22] evaluate an Adaptive Probability-Oriented Feature Matching (APOFM) method for visual odometry using images captured by both of these configurations. Also, Cabrera et al. [23] tried to estimate the optimal hyperparameters of a convolutional neural network (CNN), which is employed to address the mobile robot localization problem using images captured by a catadioptric omnidirectional vision system. The previous works are related to the localization problem, but another application in which the large field of view is advantageous is person detection. For instance, Yang et al. [24] propose a network to detect persons in images taken by a top-view fisheye camera. The training carried out is rotation equivariant. Once persons are detected, their physical positions are estimated. Given the pixel location of the human detected, it is expressed in the camera coordinate system using a general fisheye camera model and the known altitude of the fisheye camera.

Although the panoramic image generated by a catadioptric vision system does not suffer from stitching effects, it typically presents a lower resolution. In addition, it captures a field of view smaller than a whole sphere (the top and bottom of the sphere, i.e. the poles, are excluded). By contrast, two fisheye lenses with a field of view greater than 180 degrees each and opposite view directions can be used to achieve the whole sphere with only two cameras.

As regards the fisheye lens, only slightly more than a hemisphere is captured with them. They are usually used for visual surveillance. In this regard, Wang et al. [25] performed a study based on people detection for surveillance using Mask-RCNN in images taken by a top-view fisheye camera.

Among all the vision systems capable of providing a panoramic view of the scenario, dual back-to-back fisheye cameras are gaining an increasing interest due to their many benefits. Nonetheless, as already remarked in Sect. 1, the challenging task with these cameras is to generate a high-quality 360-degree view, that is, without artifacts produced by the stitching process: ghosting, misalignment, structural distortion, geometric error, chromatic aberrations and blur. Tian et al. [26] describe each of these artifacts, their origin and main properties. This vision system is the one chosen for this work. Therefore, the following subsection is focused on the methods that use a dual back-to-back fisheye camera.

In this paper, panorama refers, in general terms, to a single image with wide field of view, while spherical view is a panorama whose field of view is 180 degrees vertically and 360 degrees horizontally (a complete sphere is captured).

2.2 Full spherical view from dual back-to-back fisheye systems

A variety of research works in the state-of-the-art have studied methods to obtain high-quality 360-degree views from dual fisheye lenses. Typically, the main stages of the methods are the fisheye images unwarping or projection, the registration and the stitching or blending. In the next paragraphs the main approaches are described, focusing on the alternatives to address these stages.

2.2.1 Fisheye image unwarping

Fisheye unwrapping is the process of mapping a circular fisheye image to a rectangular panoramic image [27]. It is composed of two mappings: A first one from the fisheye image plane onto a sphere and a second one from the sphere to a flat surface.

The first projection employs a fisheye camera model, which can be based on calibration models or classical lens fisheye projections (stereographic, equidistant, equisolid angle and orthogonal). For the second projection, there are several ways to represent the sphere surface on a plane. Cai et al. [28] present an overview of the projections to obtain a full view of the environment.

As for fisheye image unwarping to generate a full spherical view, the following procedures are examples that have been employed in the literature. Ni et al. [29] propose an algorithm with two main objectives: correcting the distortion introduced by the fisheye lens (f-theta distortion) and eliminating the installation error. To achieve them, first, during the fisheye images unwarping, they use a linear interpolation method to relate the two parameters (r and θ) of the equidistant fisheye projection. Also, the center and the radius of the effective area of the fisheye image are estimated. Lin et al. [30] suggest a method to estimate the effective radius of a fisheye image, as well as the effective field of view. Then, these estimated parameters are used to transform the fisheye images to equirectangular spherical view. Xue et al. [31] use Latitude and Longitude Bilinear Interpolation (LLBI) during the transformation of the fisheye images to rectangular. Lo et al. [32] carry out a dual-fisheye camera calibration. The authors proposed a concentric calibration. The optical centers and the parameters of the projection functions corresponding to the pair of fisheye lenses are jointly estimated.

By contrast, in this paper, we propose a correction step based on a model that relates the pair of matching points between the dual fisheye images. This model is employed during the transformation of the back fisheye image to spherical format. The purpose of this correction step is that the projections of the same points in the fisheye images have the same 3D coordinates on the unit sphere surface.

2.2.2 Image registration and blending

Both image registration and blending are relevant steps. The first one is the process to relate common visual information, whereas the image blending is the process of combining the two images in one. Further details about image stitching can be consulted in the review provided by Szelisk [33].

Concerning the registration and blending process from visual information captured by a dual fisheye back-to-back camera, several solutions are proposed in the literature, some of which are described below.

Ho and Budagavi [34] propose an algorithm in which the pair of fisheye images are expressed in equirectangular format an then registered in two steps. In the first one, given a set of control points selected manually, the affine matrix is estimated in order to minimize the geometric misalignment between both equirectangular images. In the second step, a template matching for objects in the overlapping area is carried out. However, this method only produces a partially accurate alignment. This is because the control points in the central part of the equirectangular spherical view are aligned well, but this does not occur when they are at the top or bottom parts. For this reason, the authors propose an improved method in [35]. In this paper, they suggest transforming by grid interpolation based on rigid Moving Least Squares (MLS), instead of an affine warping matrix like in the previous work. Also, the authors extend the application of this method to video in this paper.

Lo et al. [36] suggest using a local warping for the alignment of the image pair. This local mesh warping is based on minimizing a cost function that combines a feature term and a smoothness term through a weighted sum. The first term aims to align each pair of matched feature points as close to their center point as possible, whereas the second term tries to preserve the geometric structure of the mesh. To do it, once the fisheye images have been transformed into equirectangular format, they are divided into a uniform mesh.

Souza et al. [37] propose an adaptive stitching method based on high-texture image regions. Once the dual fisheye images are converted to equirectangular projections, ORB features are extracted from the overlapping regions and clustered into templates. After that, the authors try to minimize the discontinuity through a template matching by using only the templates obtained in the previous step (high texture areas) instead of the whole overlapping regions. Then, the displacement information obtained from the template matching is used to estimate the homography matrix.

For these algorithm steps, other methods proposed in the literature are the estimation of a rotation matrix to relate both fisheye spatial sphere coordinates [29], the implementation of a weighted blending employing a nonlinear function [31] or the application of a mesh-deformation-based local alignment [32].

In this sense, in this paper, we propose the use of a polynomial as 2D geometric transformation for the alignment of the pair of spherical views.

3 Generating a full equirectangular spherical view

In this work, the vision system chosen is Garmin VIRB 360 [13]. This camera is composed of back-to-back dual-fisheye lenses, each with a 201-degree field view. Thus, to get a

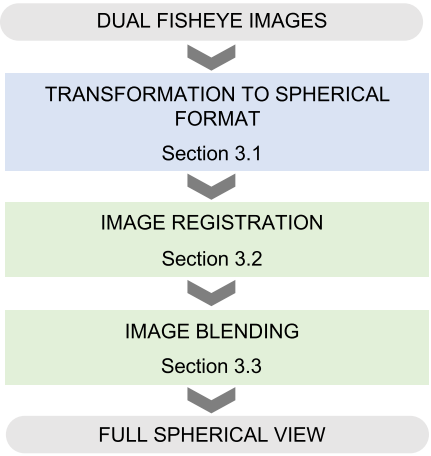


Fig. 1 The algorithm input is a pair of fisheye images. In the initial stage (blue color), both fisheye images are transformed into a spherical format. Finally, this pair of spherical views are merged after carrying out the two steps corresponding to the stitching image process (green color), being the algorithm output a full spherical view

360-degree view, the use of the image stitching technique is required to merge the pair of fisheye images provided by this camera in one shot. Nevertheless, as an initial step, each fisheye image must be projected to a more comprehensible spherical format (e.g. the equirectangular projection) in which the stitching technique can be applied.

The process of combining images with overlapping zones taken by some cameras from different views is mainly composed of two stages: (1) image registration and (2) image blending. Considering this and the above paragraph, the algorithm that generates a full spherical image from dual fisheye images is composed mainly of three stages, including the transformation to spherical format. For that reason, this section is divided into three subsections, one for each stage. In addition, the algorithm is shown by means of a diagram in Fig. 1. In this diagram, the initial stage is highlighted in blue color and the two stages of the image stitching in green color.

3.1 Transformation to spherical format

The initial stage of the algorithm consists in transforming each fisheye image into an equirectangular spherical projection where the pixel relation between the pair of images is more comprehensible. It is carried out by means of two consecutive steps. In the first one, the fisheye image is projected onto a sphere. On the contrary, in the second step, the surface of this sphere is projected to a rectangular plane (i.e. the spherical view). For this last step, the equirectangular projection is employed. In this projection, a sphere is mapped onto a rectangular image whose aspect ratio is 2:1, that is, the width of this rectangular image is twice its height. The north (Zenith) and south (Nadir) poles of the sphere are located at the top and bottom in the rectangular image.

The algorithm used in this stage is based on a backward mapping (Fig. 2), i.e. for each pixel of the equirectangular (output) image, the coordinates on the fisheye (source) image are calculated using the following inverse transformations: (1) mapping from spherical view to a global unit sphere, and, after that, (2) projecting to the fisheye image. The next paragraphs describe these two transformations.

Mapping from 2D (spherical view) to 3D (unit sphere) As shown in Fig. 3, the first mapping consists in calculating the 3D vector (P_G) corresponding to the projection of a pixel from the equirectangular spherical image into the surface of the unit sphere.

According to equirectangular projection, the x-coordinate of each pixel (x_{out}) is proportional to the longitude (θ) and the y-coordinate of the pixel (y_{out}) is proportional to the latitude (α).

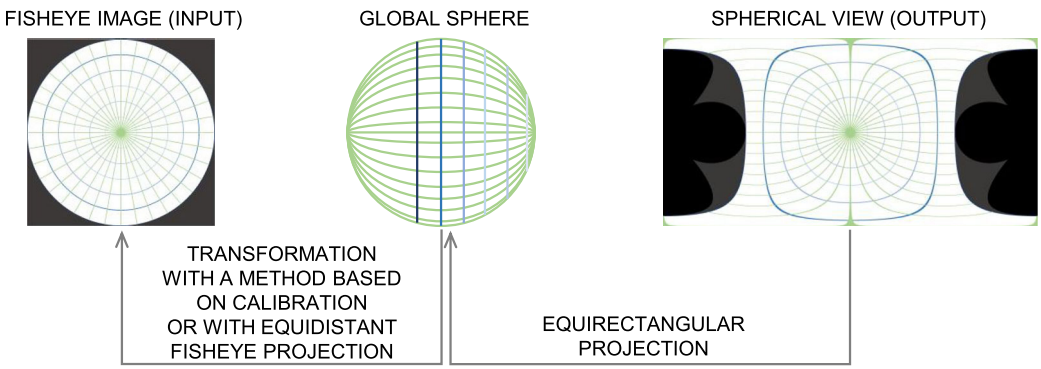


Fig. 2 The fisheye image (input), shown on the left side of the figure, is first projected onto the unit sphere, which is shown in the center of the figure. Finally, the projection of the unit sphere is transformed into a spherical view. The result is shown on the right side of the figure. The direction of the arrows at the bottom of the figure show that both transformations are performed by means of backward mapping

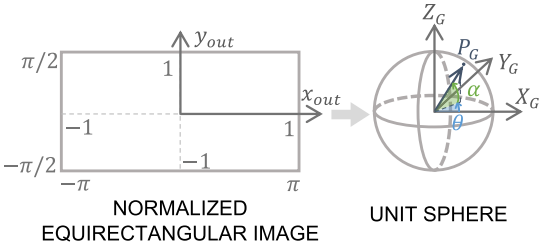


Fig. 3 Mapping of a normalized pixel from the equirectangular spherical image to the unit global sphere. The result is a 3D vector (P_G) defined by azimuth (θ) and elevation (α) angles whose values are proportional to the 2D cartesian coordinates of this normalized pixel

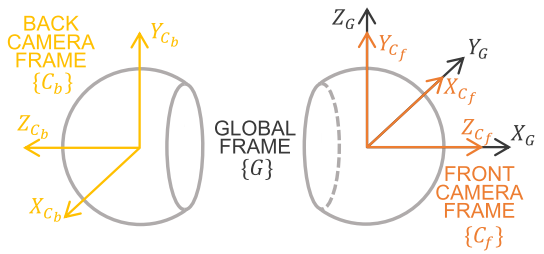


Fig. 4 There are three coordinate systems: $\{C_\mathrm{f}\}$, the front camera frame, $\{C_\mathrm{b}\}$, the back camera frame, and $\{G\}$, the global frame. The transformation from $\{G\}$ to $\{C_\mathrm{f}\}$/$\{C_\mathrm{b}\}$ is given by a rotation matrix

Given the normalized coordinates of a pixel in the spherical image (x_out, y_out), the equirectangular projection is applied. Then, the projection onto the unit sphere (i.e. the 3D vector) is given by:

$$P_G = \begin{bmatrix} X_G \\ Y_G \\ Z_G \end{bmatrix} = \begin{bmatrix} \cos\alpha\cos\theta \\ \cos\alpha\sin\theta \\ \sin\alpha \end{bmatrix} \tag{1}$$

where α is the latitude coordinate that is calculated by $\alpha = \pi/2 \cdot y_\mathrm{out}$, and θ is the longitude coordinate whose value is given by $\theta = \pi \cdot x_\mathrm{out}$.

This 3D vector P_G is expressed in the global frame system ($\{G\}$), where the positive direction of the z-axis (zenith) is perpendicular to the ground plane, as shown in Fig. 4.

However, the z-axis of one camera points right (front camera, $\{C_\mathrm{f}\}$) and the other points left (back camera, $\{C_\mathrm{b}\}$), not up and down. Therefore, before proceeding further, a change of coordinate system must be performed. This frame transformation is composed of a rotation, since we assume that the relative translation among the cameras is zero, and the centers of the three frames are the same. Moreover, this rotation will depend on which fisheye image (front or back) is being processed.

In the case of the front lens ($\{C_\mathrm{f}\}$), this transformation (R_{GC_f}) is composed by a first rotation of 90 degrees around the Z_G-axis and a second rotation of 90 degrees around the X_G-axis. For the back lens ($\{C_\mathrm{b}\}$), this transformation (R_{GC_b}) is defined by a first rotation of -90 degrees around the Z_G-axis and a second rotation of 90 degrees around the X_G-axis.

Mapping from 3D (unit sphere) to 2D (fisheye image) This mapping can be performed using a camera model based on a sphere (e.g. the unified camera model proposed by Scaramuzza et al. [38]) or using a fisheye projection (e.g. equidistant projection). For the first option, the parameters of the camera model must be previously estimated, i.e. a calibration process is required.

As mentioned at the beginning of this paper, this work not only uses a unique method for this second mapping, but also two methods have been implemented in order to compare the correctness of the full spherical views obtained by using each one. Both methods are applied to the 3D vector already expressed in the front or back camera frame ($P_{C_{\mathrm{f}\|\mathrm{b}}}$).

3.1.1 Calibration-based projection method (CPM)

Given $P_M = [X_M, Y_M, Z_M]^T$, the unified camera model proposed by Scaramuzza et al. [38] defines the following relation:

$$\begin{bmatrix} x_{src} \\ y_{src} \\ f(\rho) \end{bmatrix} = \begin{bmatrix} \rho \cdot X_M / \sqrt{X_M^2 + Y_M^2} \\ \rho \cdot Y_M / \sqrt{X_M^2 + Y_M^2} \\ \rho \cdot Z_M / \sqrt{X_M^2 + Y_M^2} \end{bmatrix} \tag{2}$$

where x_{src} and y_{src} are the coordinates of the projection on a hypothetical plane (ideal coordinates) and ρ is the radial distance (i.e. $\rho = \sqrt{x_{src}^2 + y_{src}^2}$). In addition, $f(\rho)$ consists in a Taylor polynomial, whose degree is four and whose coefficients are previously obtained during the calibration process. The coordinates x_{src} and y_{src} can be calculated after estimating the value of ρ through the third element of the vector defined on Eq. (2). This camera model states the ideal coordinates (x_{src}, y_{src}) and their corresponding real coordinates on the image plane (u_{src}, v_{src}) are related by an affine transformation. Figure 5a shows the mapping from the unit sphere (3D) to the fisheye image (2D) using this method.

3.1.2 Equidistant fisheye projection (EFP)

The 3D point ($P_{C_{\mathrm{f}\|\mathrm{b}}}$) projected onto the unit sphere can be expressed as follows:

$$P_{C_{\mathrm{f}\|\mathrm{b}}} = \begin{bmatrix} x_{C_{\mathrm{f}\|\mathrm{b}}} \\ y_{C_{\mathrm{f}\|\mathrm{b}}} \\ z_{C_{\mathrm{f}\|\mathrm{b}}} \end{bmatrix} = \begin{bmatrix} \sin\phi\cos\theta \\ \sin\phi\sin\theta \\ \cos\phi \end{bmatrix} \tag{3}$$

where ϕ is the angle from the camera view direction ($Z_{C_{\mathrm{f}\|\mathrm{b}}}$-axis) to the 3D coordinate vector (called zenith angle) and θ

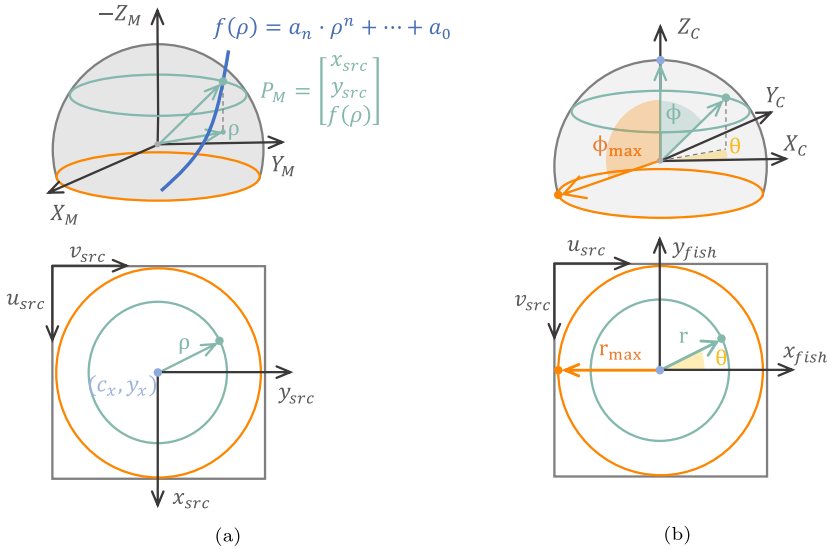


Fig. 5 Mapping from 3D to 2D using: **a** Calibration-based Projection Method (CPM) and **b** Equidistant Fisheye Projection (EFP)

is the angle in the $X_{C_{\text{f}\|\text{b}}}$-$Y_{C_{\text{f}\|\text{b}}}$ plane from the positive $X_{C_{\text{f}\|\text{b}}}$-axis.

$$\theta = \tan^{-1}(y_{C_{\text{f}\|\text{b}}}/x_{C_{\text{f}\|\text{b}}}) \tag{4}$$

This 3D point will be given by a radial (r) and an angular (θ) coordinate in the fisheye image. The second coordinate is the same that appears in Eq. (3), so θ can be calculated using Eq. (4), whereas the former (r) is calculated through the equidistant fisheye projection. This projection explains the linear relationship between r and ϕ:

$$r = a \cdot \phi \tag{5}$$

The maximum value the zenith angle (ϕ_{max}) can take is equal to half of the field of view in radians, and the maximum radial distance (r_{max}) on the fisheye image is equal to one since the coordinates are normalized. Taking this into account, the parameter a can be calculated using these values and the previous equation:

$$a = \frac{r_{\text{max}}}{\phi_{\text{max}}} = \frac{1}{FOV/2} \tag{6}$$

Up to this point, the polar coordinates (r, θ) on the fisheye image of the projected point have been calculated. Then, the final step consists in a first transformation to the cartesian coordinates (x_{src}, y_{src}):

$$(x_{src}, y_{src}) = (r \cdot \cos\theta, r \cdot \sin\theta) \tag{7}$$

and a second one to obtain the denormalized pixel coordinates of the fisheye image (u_{src}, v_{src}) by means of:

$$\begin{bmatrix} u_{src} \\ v_{src} \end{bmatrix} = \begin{bmatrix} a_x & 0 \\ 0 & a_y \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} x_{src} \\ y_{src} \end{bmatrix} + \begin{bmatrix} c_x \\ c_y \end{bmatrix} \tag{8}$$

where a_x is the half of the fisheye image width ($W_{src}/2$) and a_y is the half of the fisheye image height ($H_{src}/2$).

3.1.3 The correction step proposed

During the registration process, the aim is to align the images from matches of points and thus reduce the discontinuity between them before fusion. Taking into account that initial pair of images are fisheye, we propose an additional step in the algorithm. When this correction step is applied during the transformation to spherical, the difference between the pair of images output at this stage will be reduced.

The algorithm described in the previous subsection assumes, concerning the pair of fisheye images, the following facts:

- Both unit spheres have the same center, which means that there is no displacement between the front and back camera.
- The Z-axes of both coordinate systems are perfectly aligned and opposite.
- In the case of EFP, the relationship between the coordinates of a 3D scene point and its projection on the image is the same for both cameras.

However, it is possible that these assumptions are not fulfilled in reality. Accordingly, we propose the correction step composed of two functions that relate the polar coordinates of the same 3D point projected on the pair of fisheye images.

In order to estimate the functions describing a 2D relation, the coordinates of both projections must be expressed in the same planar coordinate system. As Fig. 6 shows, given a pair of fisheye images, the first step is to find matching features between them. The objective is to obtain a set of N feature points detected in the front fisheye image ($p^{C_{\text{front}}} = [u_p^{C_{\text{front}}}, v_p^{C_{\text{front}}}]$) and their matching feature points

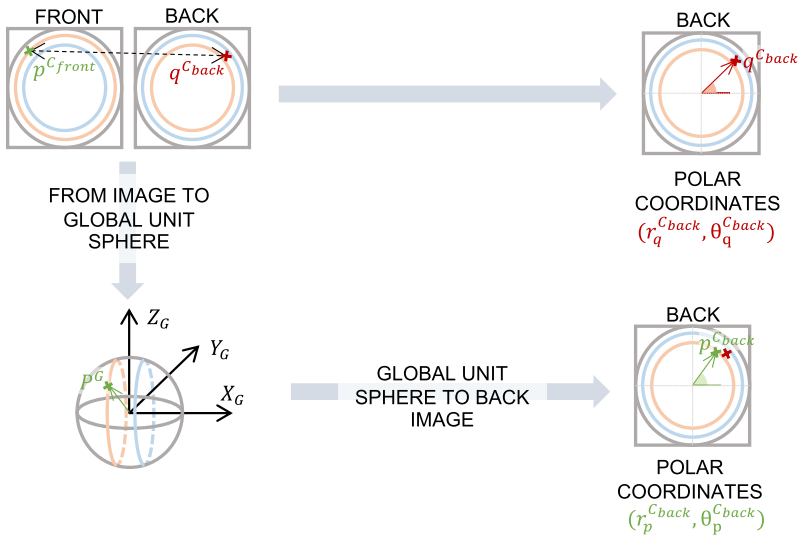


Fig. 6 Process to obtain the relation between pixel pairs of the dual fisheye images. Considering a pair of matching points, the point corresponding to the front image is expressed in the same coordinate system as the point of the back image. This is achieved by a projection onto the surface of the global unit sphere followed by a projection of this 3D point to the back image. This is performed by means of the transformations between the coordinate systems of both cameras and the unit global sphere (Fig. 4), and the equations for 3D to 2D forward/backward mapping (Sect. 3.1.2)

detected in the back fisheye image ($q^{C_{\text{back}}} = [u_q^{C_{\text{back}}}, v_q^{C_{\text{back}}}]$). The second step involves projecting $p^{C_{\text{front}}}$ onto the unit sphere, whose z-axis points upwards (global sphere). Then, they are projected on the back fisheye image plane ($p^{C_{\text{back}}}$). These coordinates are the ones expected in the back image, in ideal conditions, in such a way that the corresponding pair would have the same position on the sphere and thus on the spherical view. Both $p^{C_{\text{back}}}$ (expected coordinates) and $q^{C_{\text{back}}}$ (real coordinates) are expressed in polar coordinates: ($r_p^{C_{\text{back}}}$, $\theta_p^{C_{\text{back}}}$) and ($r_q^{C_{\text{back}}}$, $\theta_q^{C_{\text{back}}}$).

After that, we define the functions of the correction step. On the one hand, the relation between angle coordinates ($\theta_p^{C_{\text{back}}}$ and $\theta_q^{C_{\text{back}}}$) can be modeled by f_θ as follows:

$$\theta_q^{C_{\text{back}}} = f_\theta(\theta_p^{C_{\text{back}}}) \tag{9}$$

On the other hand, the function, f_r corrects the radial distance coordinate:

$$r_q^{C_{\text{back}}} = f_r(r_p^{C_{\text{back}}}) \tag{10}$$

In Sect. 4.2.2, these functions and their corresponding parameters are obtained by an experimental analysis.

For each pair of fisheye images, the parameters of both functions are estimated. Then, these functions are used during the transformation of the back fisheye image to spherical view, concretely they are introduced in the 3D to 2D mapping equations.

In this case, once the projection on the unit sphere is known, Eq. (3), the expected polar coordinates $r_p^{C_{\text{back}}}$ and $\theta_p^{C_{\text{back}}}$ can be calculated by means of Eqs. (4) and (5), respectively. After that, the real polar coordinates of the back fisheye image ($r_q^{C_{\text{back}}}$ and $\theta_q^{C_{\text{back}}}$) are estimated by the functions of the correction step, Eqs. (9) and (10) . To finish the mapping from the unit sphere to the back fisheye image, the cartesian coordinates (x_{src}, y_{src}) are calculated using Eq. (7) with the real polar coordinates. After that, the pixel coordinates are obtained by Eq. (8).

Taking into account the information included in Sect. 3.1, the transformation to spherical format has several variations in this work. They differ in the method employed to relate the 3D point on the sphere and the pixel in the fisheye image. This mapping can be carried out by (a) a Calibration-based Projection Method (CPM) or (b) an Equidistant Fisheye Projection (EFP). Also, the latter can be combined with the correction step proposed in the present paper. Thus, an additional case is obtained: (c) an Equidistant Fisheye Projection (EFP) + correction. The correction can be carried out in a polar coordinate, θ, (c.1) or both polar coordinates, r and θ, (c.2). It can be visualized in Fig. 7, where the contribution of this paper related to this stage (i.e. correction step) is highlighted in orange color.

3.2 Image registration

After converting each fisheye image to a spherical view by the equirectangular projection, the next stage to generate a full spherical view consists in aligning the pair of spherical views using an image registration method. For this purpose, the technique employed is a feature-based image registration, which means that the geometrical transformation is estimated by the pairs of matching feature points. In this work, the method chosen to detect and describe local features is ORB [39].

The main steps of the image registration are: feature point extraction and description, feature matching, estimation of the geometric transformation and performing the transformation.

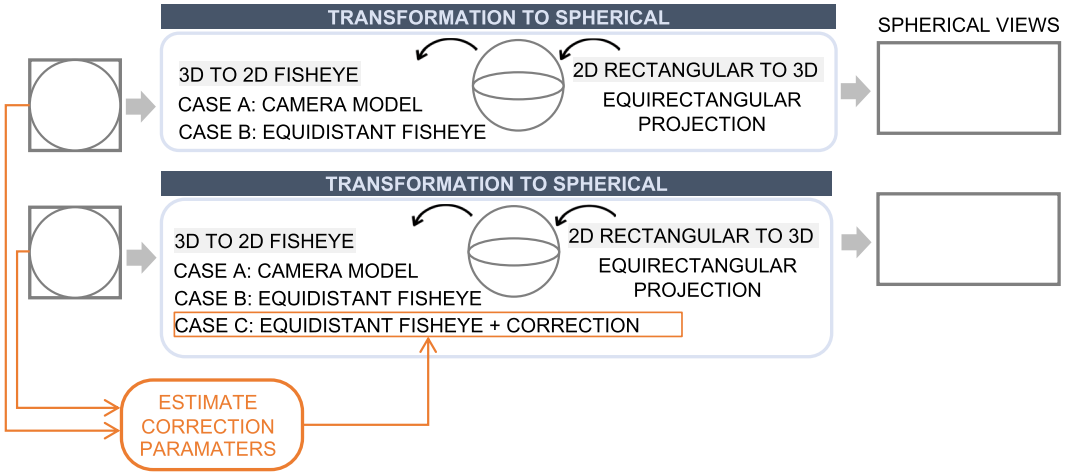


Fig. 7 Block diagram of the first stage of the algorithm: fisheye images transformation into spherical views (Sect. 3.1). Three different cases will be analyzed in this stage of the algorithm, depending on the method to map from 3D to fisheye image: CPM (Sect. 3.1.1), EFP (Sect. 3.1.2) or EFP with correction (Sect. 3.1.3). The correction step proposed in this paper is highlighted with orange color

The most common type of transformation used for this kind of image is the 2D affine. This geometric transformation is a linear mapping method that preserves the points, straight lines and planes, but not angles and length (even though ratios) [40]. For instance, with this transformation, a rectangle becomes a parallelogram. An affine transformation combines linear transformations (rotation, scale, shear and reflection) and translations.

However, a spherical view presents huge distortion at the poles. Thus, there is a nonlinear difference between the pairs of correspondences in this representation. It occurs predominantly in the most distorted parts of the view. This statement will be supported by experimental data in Sect. 4.2.1.

As a consequence, we propose using a 2D polynomial geometric transformation. For this type of transformation, a second-degree polynomial was selected. The inverse 2D polynomial transformation is given by:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} a_5 & a_4 & a_3 & a_2 & a_1 & a_0 \\ b_5 & b_4 & b_3 & b_2 & b_1 & b_0 \end{bmatrix} \cdot \begin{bmatrix} y^2 \\ x^2 \\ xy \\ y \\ x \\ 1 \end{bmatrix} \tag{11}$$

where a_5, a_4, a_3, a_2, a_1, a_0 are the polynomial coefficients for estimating the coordinate u and b_5, b_4, b_3, b_2, b_1, b_0 are the polynomial coefficients for estimating the coordinate v.

In this work, the two types of geometric transformation described in this section have been implemented. In this way, the full spherical views obtained using each of them can be compared to assess whether the quality is improved by using the proposed polynomial transformation. The results of this study are shown in Sect. 4.3.3.

Considering the information described in Sect. 3.2, the image registration stage of the algorithm has two variations in this paper. The difference is in the type of geometric transformation to align the pair of spherical views. The block diagram of this stage and its variations are shown in Fig. 8. The alignment of the spherical view pair can be carried out employing (a) an affine matrix or (b) a polynomial. The contribution of this paper related to this stage (i.e. the use of the polynomial) is highlighted in orange color.

3.3 Image blending

The image blending attempts to create a unique image without visible seams, that is, minimizing discontinuities in the global appearance of the final image caused by geometrical or/and photometrical misalignment. Two main approaches can be found to perform image blending: optimal seam finding and transition smoothing.

The first approach tries to estimate the optimal seam location in the overlapping zone to minimize the differences between both sides of this seam line. Since these algorithms consider the visual information of the scene in the overlapping region, optimal seam finding is a good solution for dealing with the artifacts due to parallax or dynamic scenes (moving objects), but not with those caused by exposure differences or illumination variations in the scene between the images. In contrast, the approaches based on transition smoothing fuse the image information of the overlapping region. These methods can address discontinuities caused by photometric misalignment, such as those which are not caused by small registration errors or moving objects. In [41,

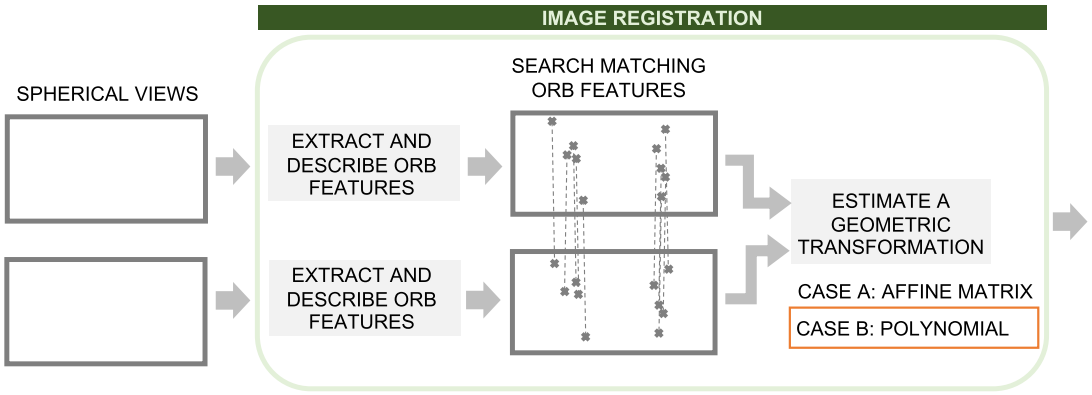


Fig. 8 Block diagram of the second stage of the image alignment. There are two variations depending on the type of geometric transformation used to align the pair of spherical views (Sect. 3.2): affine matrix or polynomial (Eq. (11)). The kind of geometric transformation proposed in this paper is highlighted with orange color

42], more information about algorithms for image blending can be found.

In the present paper, the method employed to blend the pair of spherical views is the described in [34].Then, the blending is carried out by means of a ramp function in the overlapping zones.

4 Experiments

In this section, the results of the experiments are presented. First, the vision system is described. Second, the pixel-based relationship between the pair of fisheye images and the pair of spherical views is studied in Sect. 4.2. On the one hand, the first study is performed with the purpose of confirming the nonlinearity between the spherical views pair (Sect. 4.2.1). On the other hand, the objective of the second study is to support the proposed correction step and identify the types of functions (Sect. 4.2.2). After that, the quality of the full spherical views is evaluated in Sect. 4.3.

4.1 Vision system

The vision system used in this work is a Garmin VIRB 360 camera [13], whose main features are shown in Table 1.

The Garmin VIRB 360 camera is composed of two back-to-back fisheye lenses and two backside-illuminated CMOS sensors (1/2.3"). The field of view of each lens is 201.8 degrees, therefore, a full spherical view can be constructed using the two images captured with both cameras.

This camera can provide different types of images in ".JPEG" format. The type of the image captured depends on the setting of the lens mode (Table 1): *360*, *front only*, *rear only* or *RAW*. In this paper, we only worked with images captured using *360* (Fig. 9c) and *RAW* (Fig. 9b and a) lens mode.

4.2 Experimental evaluation of the difference between feature matchings

The objective of this section is to experimentally support the theoretical aspects on which the two contributions of this paper are based. For this purpose, two studies have been carried out in order to obtain the pixel-based relationship between the image pair before applying any transformation. The first one is related to the pair of spherical views in rectangular coordinates (Sect. 4.2.1) with the aim of confirming the nonlinearity. The second one is concerning to the pair of fisheye images in polar coordinates (Sect. 4.2.2) to model the function of the proposed correction step.

4.2.1 Spherical views: rectangular coordinates

In the case of a pair of spherical views, the procedure consists in a feature matching search between pairs of spherical views that are the outputs of the first stage (see Sect. 3.1), i.e. without any transformation applied. With the purpose of having the lowest number of false positives, this search is based on Aruco Markers. The pairs of feature matches are the corners of an Aruco marker with a specified identifier captured in both images. To accomplish this, the camera was positioned so that several Aruco markers appeared in the overlapping area. The camera was then rotated around the z-axis of one of the fisheye lenses to capture these marks over the whole overlapping area. After multiple rotations, a set of several image pairs were acquired from which a total of 1628 feature matches were obtained.

Table 1 Main technical features of the Garmin VIRB 360 camera

Physical	Weight	160 g (with battery)
	Size	39.0 mm(H)×59.3 mm(W)×69.8 mm(D)
Optics	Sensor	1/2.3" Backside-Illuminated CMOS (2 sensors)
	Lens count	2
	FOV	201.8 degree per Lens
	Effective focal length	1.036 mm
Lens Mode	*360*	This mode outputs stitched fully spherical photos in equirectangular format. The resolution is 5640×2820 (15 MP)
	front only	This mode outputs a perspective image calculated from the photo captured by the front lens. The resolution is 1920×1440 (3 MP)
	rear only	This mode outputs a perspective image calculated from the photo captured by the rear lens. The resolution is 1920×1440 (3MP)
	RAW	This mode captures one raw picture for each lens (2 files). The resolution is 3008×3000 for file (2×9MP)

(a) (b) (c)

Fig. 9 Example of the images output by the Garmin VIRB 360 camera: **a** and **b** show, respectively the images captured in RAW mode with the front and the rear fisheye cameras and **c** shows the full spherical view output in 360 mode

Figure 10 shows the results of the x-coordinate study, where each overlapping zone is analyzed independently: left (Fig. 10a and c) and right (Fig. 10b and d).

First, Fig. 10a and b shows the error in the x-coordinates of each pair of matching points for the left and right overlapping region, respectively. These values are represented (plot y-axis) versus the x-coordinate of the back spherical view (plot x-axis). Also, the y-coordinate can be visualized with a color that depends on its value. After analyzing both figures, we can observe that this difference is higher for points located on the upper and lower areas (highest and lowest values for y-coordinate) and also far from the center of each overlapping zone, i.e. $x = 1410$ (-90 degrees of longitude) and $x = 4230$ (90 degrees of longitude), respectively.

Second, Fig. 10c and d shows the x-coordinate of the front spherical view versus the x-coordinate of the back spherical view.

Figure 11 shows the same results than Fig. 10, but for the y-coordinate. For this coordinate, we observe a more straightforward relationship. After studying Fig. 11a and b, we can observe that the error in y-coordinates is practically linear in the middle region of the image, but this linearity is lost in the upper and lower regions of the image (i.e. in the poles of the global sphere, where there is more distortion).

Making a comparison between both overlapping zones, the behavior of the y-coordinate is inverse, as it is clearly illustrated in Fig. 11c and d. In regards to the left overlapping region (Fig. 11c), the error is maximum and positive in the top part of the view (lower y-coordinate values), whereas it is minimum and negative in the bottom part (higher y-coordinate values) of the view. On the contrary, in the right overlapping region (Fig. 11d), the error is minimum and negative in the top part and maximum and positive in the bottom part.

This first study has allowed us to observe the nonlinear relationship between pairs of matched features in a pair of spherical views. This nonlinearity is more noticeable at the top and bottom parts of the spherical view. An affine matrix

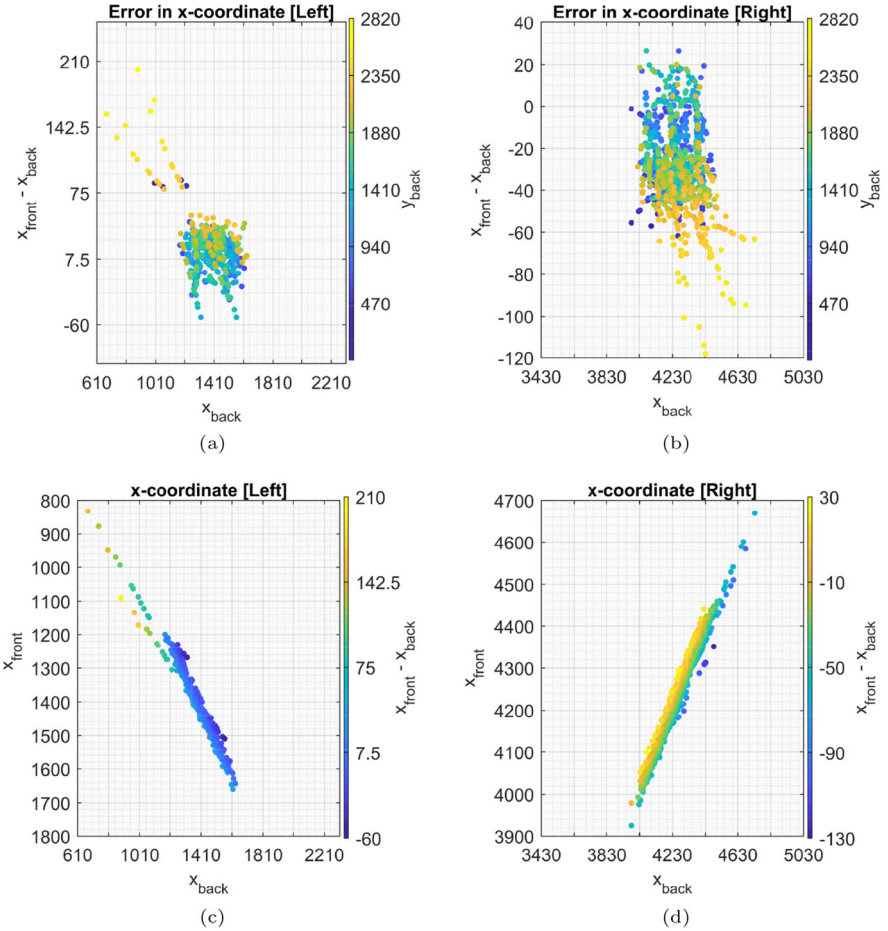


Fig. 10 Results of the x-coordinate study, where **a** and **c** are related to the left overlapping region, whereas **b** and **d** are related to the right overlapping region

can correctly register the middle part of one spherical view with respect to the other, where certain linearity exits, but not the top and bottom parts. As a result, we propose the use of a polynomial geometric transformation during the image registration process.

4.2.2 Fisheye images: polar coordinates

In the case of the fisheye image pair, the procedure is more elaborated due to the fact that the points must be expressed in the same image frame to calculate the distance. The algorithm is shown in Fig. 6. The last step is calculating the difference between each pair of these coordinates. The results are shown in Fig. 12.

As can be seen in Fig. 12b, the difference between $\theta_p^{C_{\text{back}}}$ and $\theta_q^{C_{\text{back}}}$ can be modeled by a sine function. As for the radial distance coordinate (Fig. 12c), even though the relationship is not so direct and common for most cases, we have noted that an α factor models the relationship between the pair of radial distance coordinates.

After analyzing the results of this experimental study, the functions of the correction step are identified. The function f_θ can be defined as a sine function as follows:

$$\theta_q^{C_{\text{back}}} = f_\theta(\theta_p^{C_{\text{back}}}) = \theta_p^{C_{\text{back}}} - a \cdot \sin\left(b \cdot \theta_p^{C_{\text{back}}} + c\right) \quad (12)$$

As for the correction of the radial distance coordinate, both $r_p^{C_{\text{back}}}$ and $r_q^{C_{\text{back}}}$ are related by a proportional factor denominated α.

$$r_q^{C_{\text{back}}} = f_r(r_p^{C_{\text{back}}}) = \alpha \cdot r_p^{C_{\text{back}}} = \alpha \cdot a \cdot \phi \quad (13)$$

The factor α is estimated as the mean of the ratios between the pairs of radial distance coordinates:

$$\alpha = \sum_{j=1}^{N} \frac{r_{q_j}^{C_{\text{back}}}}{r_{p_j}^{C_{\text{back}}}} \quad (14)$$

The parameters a, b, c and α are estimated before the transformation to spherical format. Then, these functions are applied during the transformation of the back fisheye image to spherical format.

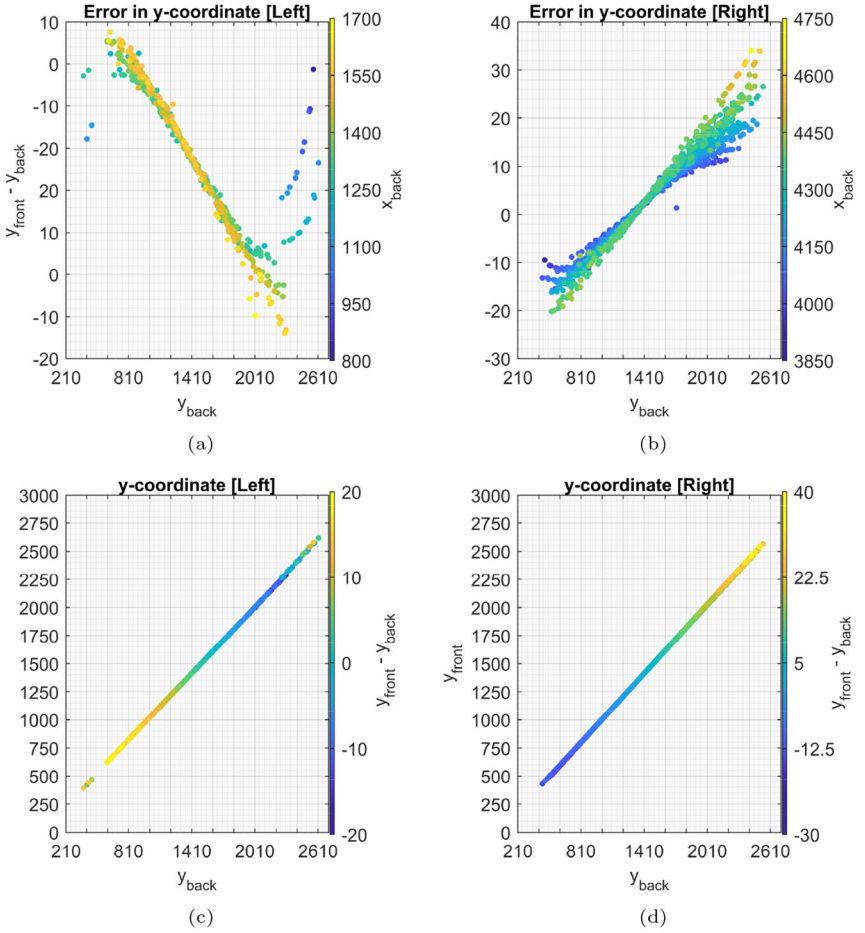


Fig. 11 Results of the y-coordinate study, where **a** and **c** are related to the left overlapping region, whereas **b** and **d** are related to the right overlapping region

4.3 Quality evaluation

In this section, the ability of the algorithm to output a correct full spherical view is evaluated, i.e. with a global appearance as homogeneous as possible and without visible stitching artifacts. The full spherical view provided by the Garmin VIRB 360 is the only visual information provided by this camera. That is to say, the firmware of the camera is not public, so that any previous composition nor internal data are available for evaluation purposes. Nonetheless, to perform this evaluation, two approaches are considered to obtain quality measurements.

The first approach is a no-reference quality method (see Sect. 4.3.2), which will be employed to make a comparative evaluation among the four types of full spherical views generated according to the projection method (see Table 2) and the one provided by the Garmin VIRB 360 in 360 lens mode (VIRB). In this first experiment, the full spherical views generated are the result of applying an affine matrix as geometric transformation between the pairs of spherical views.

The second approach is a full-reference quality approach (see Sect. 4.3.3), which will be applied only to the final full spherical views generated (see Table 2) using both types of geometric transformation (i.e. affine and polynomial).

4.3.1 Dataset garmin VIRB 360

About the visual dataset, to perform the experiments, a set of images has been captured in a variety of scenarios. These scenarios present different levels of visual information and challenging features, to obtain a complete evaluation. More concisely, the dataset is composed of images captured at a total of 50 positions in four kinds of scenarios: *Office*, *Laboratory*, *Meeting room* and *Hall*. In Table 3, a brief description of each scene is shown. A pair of fisheye images (RAW mode) and a full spherical view (360 mode) have been captured at each position. The RAW fisheye images are the input of the proposed algorithm, while the full spherical view will be used as reference to check the performance of the proposal (among other parameters). Figure 9 displays an example of the images corresponding to a single position in the database (*Hall*).

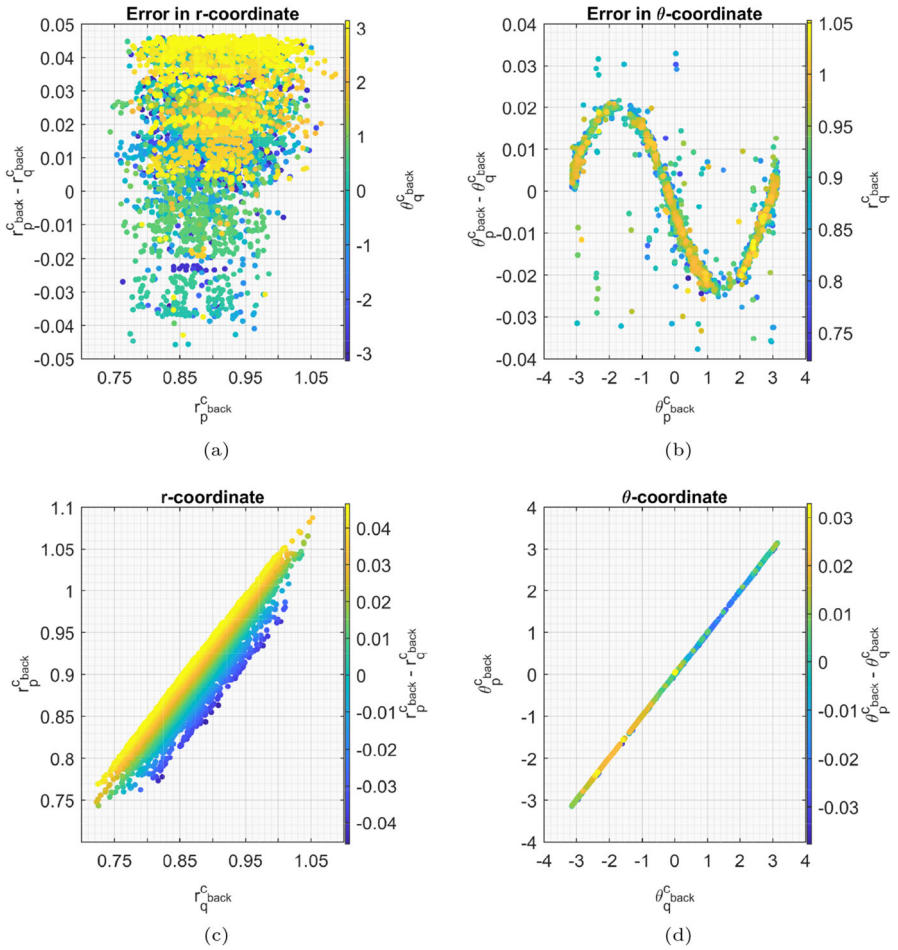


Fig. 12 Study of the polar coordinates, where **a** and **c** are related to the radial distance coordinate (r), whereas **b** and **d** are related to the angular coordinate (θ)

Table 2 Main information about each variation related to the transformation of a fisheye image into a spherical view

Case	Abbr	Method: Projection from sphere to fisheye image	Equations
a	CPM	Calibration-based projection method (Sect. 3.1.1)	[38]
b	EFP	Equidistant Fisheye Projection (Sect. 3.1.2)	Eqs. (3), (4) and (5)
c.1	EFP+θ	Equidistant Fisheye Projection with the correction of the θ polar coordinate (Sect. 3.1.2 and Sect. 3.1.3)	Eqs. (3), (12) and (5)
c.2	EFP+θ+r	Equidistant Fisheye Projection with the correction of both polar coordinates (r and θ) (Sect. 3.1.2 and Sect. 3.1.3)	Eqs. (3), (12) and (13)

4.3.2 No-reference quality assessment

In this first experiment, the evaluation focuses on the sharpness of the final image, using a no-reference image quality assessment method. The sharpness of an image is related to the presence of high-frequency components. Therefore, this approach studies the image in the frequency domain.

Firstly, the discrete 2D Fourier transform of the overlapping area is computed using a Fast Fourier Transform (FFT) algorithm. Then, after shifting the zero-frequency components to the center, a high filter pass is applied, removing the low frequencies. Once the zero frequency components are moved back to their original location ($\mathcal{F}_i$), the mean of the magnitude spectrum, which is the Image Quality (IQ) score in this evaluation, is calculated by means of Eq. (15). The

Table 3 Brief description of the scenarios in which the images that compose this dataset were captured

	Description	Number of positions	Area
Office	In these images, the objects that typically appear are computers, desks, cupboards, coat stands, posters and a whiteboard. Besides, this scene has some Aruco markers, providing more visual information	24 positions	27 m^2
Laboratory	This scene is a laboratory. This room is the largest one of this dataset. There considerable amount of and variety of objects, thus being a space rich in visual information. However, owing to the dimensions of the scenario, this information usually appears in the middle rows of the spherical view, what challenges the registration step	6 positions	117 m^2
Meeting room	The predominating structures in this scenario are bookcases with books, chairs and desks, among other objects. While the space is rich in visual information, this scenario is specially challenging because of the repetitivity and symmetry of the visual appearance	11 positions	55 m^2
Hall	This scene is less rich in detail, since it is primarily constituted by walls, large windows, and doors, though there exists some scarce visual information due to the emergency exit signs or informative posters, for instance	9 positions	69 m^2

higher this value, the sharper the image is.

$$IQ_{\text{sharpness}}\ \text{score} = \frac{\sum_{i=1}^{M}(1 + |\mathcal{F}_i(u, v)|)}{M} \tag{15}$$

where M is the total number of pixels in the overlapping region. Figure 13 shows the mean for each kind of scenario of this sharpness IQ score using bar graphs. Also, the standard error of the mean is represented with orange color and the deviation with respect to the best result of mean $IQ_{\text{sharpness}}$ score (the highest mean value) is indicated as a percentage.

Regarding the results shown in this figure, the highest average values have been reached in the *Office* (Fig. 13a and b) and *Laboratory* (Fig. 13c and d) scenarios. Additionally, the *Meeting room* (Fig. 13e and f) and *Hall* (Fig. 13g and h) scenarios have the lowest average values. This result is in line with the fact that the *Office* and *Laboratory* scenarios are rich in visual information, as noted in Table 3.

Analyzing in depth the results obtained, we can observe that the views generated by the Calibration-based Projection Method (CPM) present relatively good results for all the scenes (CPM has the highest values of $IQ_{\text{sharpness}}$ score). On the contrary, the results with the lowest quality values are from the Garmin VIRB camera. In scenes with poor visual information, the percentage deviation is lower.

In regards to the results using the variations of equidistant fisheye projection described in this paper, the difference between the three cases (i.e. EFP, EFP+θ and EFP+θ+r) is minimal, and although they do not have the best $IQ_{\text{sharpness}}$ scores, they are very close to the Calibration-based Projection Method (CPM), being the percentage deviation around one percent.

It is worth noting that a relatively small percentage deviation may not imply higher quality but instead, that more visual information appears in the overlapping region (more or less visual information can appear in this region depending on the projection type). However, when the percentage deviation is significant, we can consider that it is due to blur effect.

In an overall analysis, the $IQ_{\text{sharpness}}$ scores are more similar among them in scenarios with poor visual information. This fact is expected since the transformation between the pair of spherical views is estimated using matching features. In this respect, a more significant number of matching features will only be found if there is distinctive visual information in the common area. Particularly, in the case of the

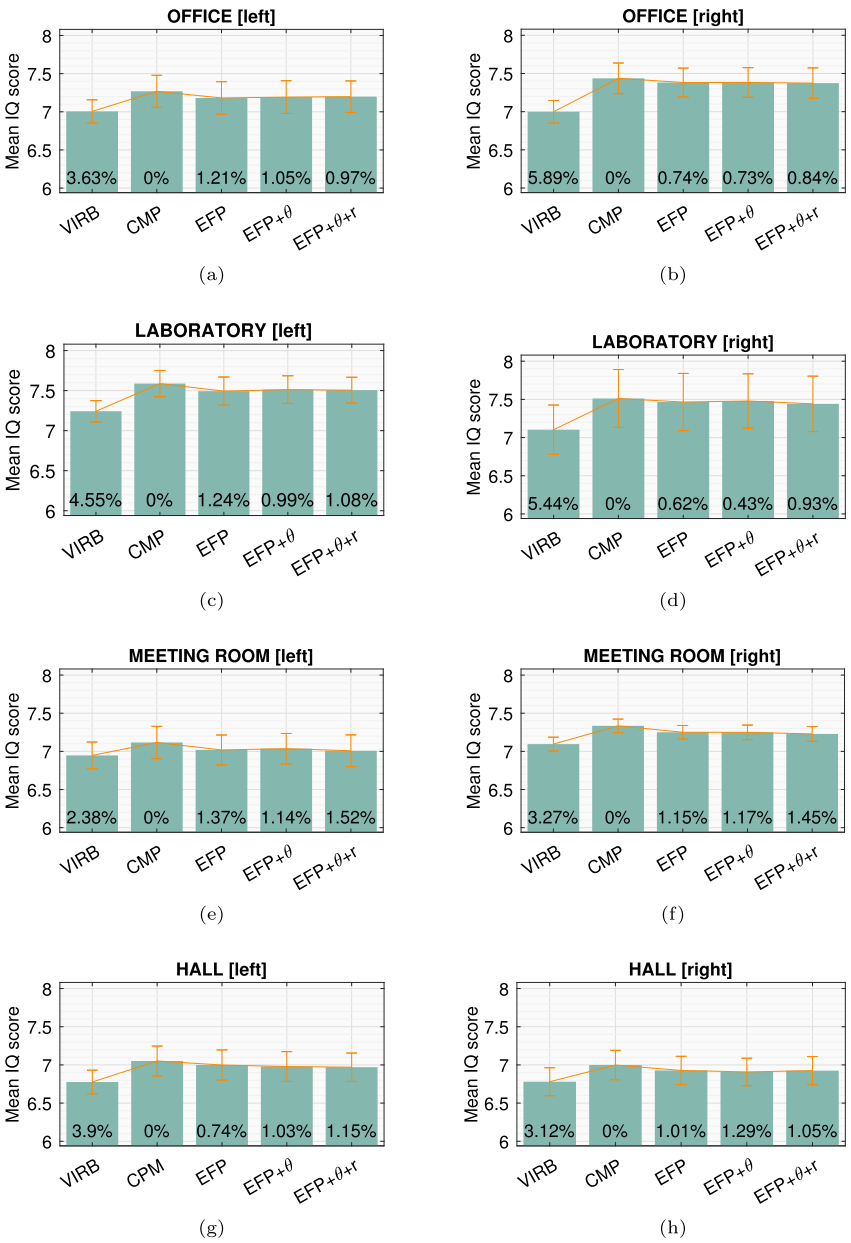


Fig. 13 Evaluation of the full spherical views based on sharpness (no-reference metric). Comparison between the views obtained with the different configurations of the algorithm (CPM, EFP, EFP+θ and EFP+θ+r) and the one provided by the Garmin VIRB 360 (VIRB) in 360 mode. This $IQ_{sharpness}$ score was calculated for the left (**a**, **c**, **e**, **g**) and right (**b**, **d**, **f**, **h**) overlapping region of each full spherical view

algorithms based on the proposed correction model, this drawback is more remarkable due to the fact that not only the image registration but also the correction model is based on matching features.

4.3.3 Full-reference quality assessment

In this section, the evaluation is realized by a full reference image quality approach. MS-SSIM is the full-reference quality method chosen for this section.

A full-reference approach calculates the quality score of a test image as a result of comparing it with a reference image. The most used full-reference image quality methods are mean squared error (MSE) and peak signal-to-noise ratio (PSNR), which is a variation of the former. The score of these methods is calculated by means of a pixel-to-pixel comparison between the test and reference image. However, the result of these methods might not be correlated with the human perception of quality unlike Structured Similarity Indexing Method (SSIM) [43] since it is based on the structural information from the scene.

In the literature, there are also full-reference image quality evaluation alternatives for omnidirectional images. In this line, Li et al. [44] provide a cross-reference omnidirectional image dataset. It contains stitched images as well as dual-fisheye images. The main feature of this dataset is that it is composed of four images captured from the same position but with different orientations (0°, 90°, 180° and 270°).

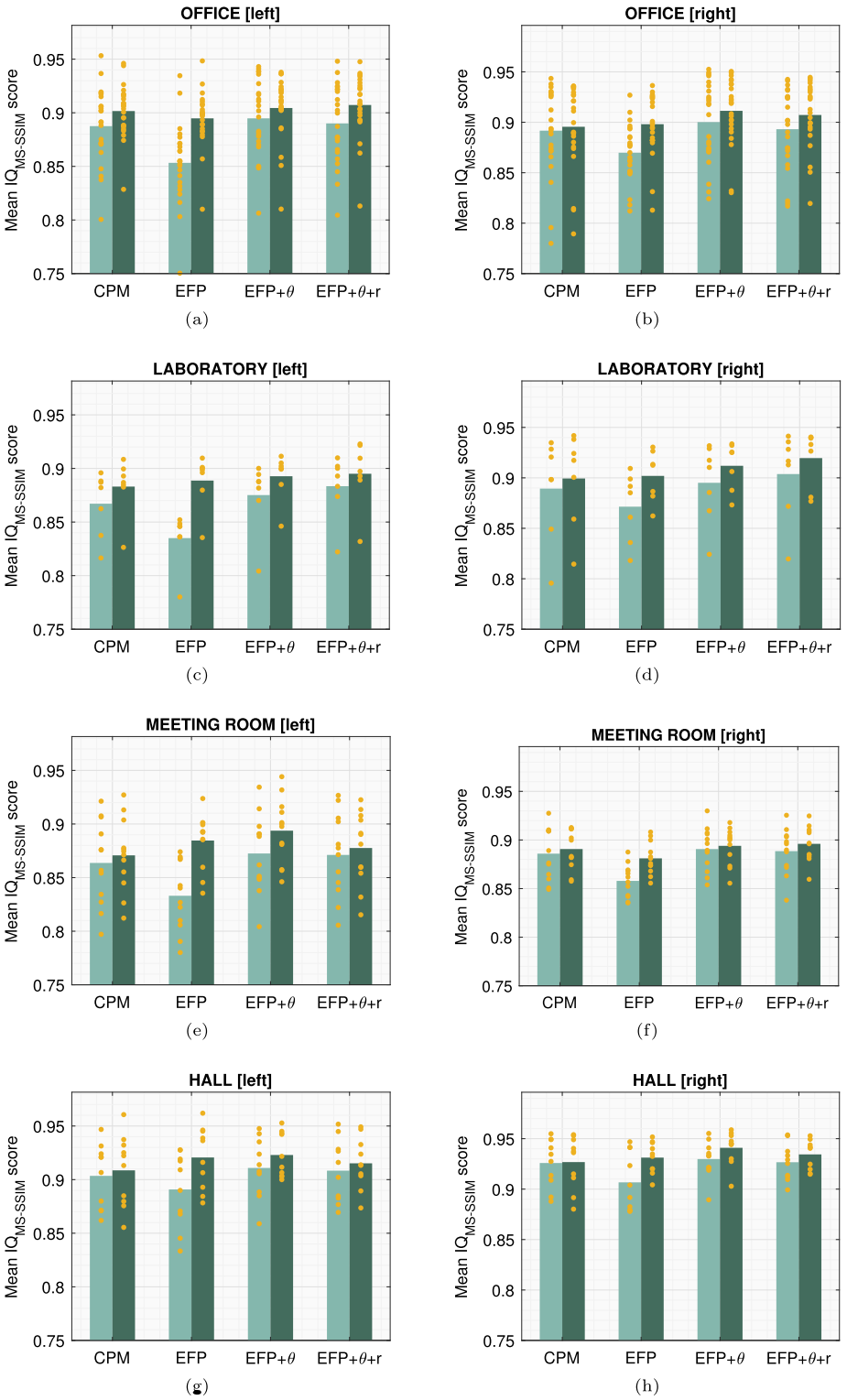


Fig. 14 Evaluation based on MS-SSIM of the full spherical views generated using an affine matrix light green color or a polynomial dark green color as geometric transformation. The scores for each case are represented by Yellow color. This $IQ_{MS-SSIM}$ score was calculated for the left (**a**, **c**, **e**, **g**) and right (**b**, **d**, **f**, **h**) overlapping region of each full spherical view

This manner, the stitched image from the pair of fisheye images taken from 0° and 180° can be used as perfect reference groundtruth when evaluating the stitched image from the pair of fisheye images taken from 90° and 270°, and vice versa. As for quality score, Li et al. [45] present the Attentive Quality Assessment (AQA). This evaluation is carried out using different metrics and also human subjective evaluations. The aim of the latter is to supervise their linear classifier. With the classifier, the metric is consistent with human subjective assessment. About metrics, the authors propose two local quality assessment metrics (sparse reconstruction and appearance similarity) and two global quality assessment metrics (color chromatism and blind zone). Duan et al. [46] suggest the use of an Attentive Multi-channel IQA Neural Network for designing an objective IQA metric. They propose both full reference and no reference quality assessment algorithms and are based on the subjective ratings that the authors obtained. The method consists of a first transformation of the omnidirectional images to cubic images in which data refinement and data augmentation methods are applied. As for the deep convolution neural networks, ResNet is the backbone and the authors present a sub-network for spatial attention to extract the features associated with the stitching distortions.

In the present work, an IQ score based on SSIM is used in this subsection. This IQ score compares the test image (I) with a reference image (I_r) and is based on three features: luminance, contrast and structure. Then, the SSIM score is the combination of three comparison functions related to these features. In addition, an extended version called Multi-Scale Structural Similarity Index Method (MS-SSIM) [47] evaluates the structural similarity of both images at different image scales.

MS-SSIM combines the luminance comparison at the highest scale M, $l_M(I, I_\mathrm{r})$, with the structure, $s_j(I, I_\mathrm{r})$, and contrast, $c_j(I, I_\mathrm{r})$, comparison calculated at different scales. The multiple scales are obtained by applying a low-pass filter and down-sampling the image by a factor of two $M-1$ times, corresponding the scale 1 to the original resolution image and the scale M to the lowest resolution. The MS-SSIM quality score is calculated as:

$$\mathrm{IQ}_{\mathrm{MS-SSIM}}\ \mathrm{score}(I, I_\mathrm{r}) = l_M(I, I_\mathrm{ref})]^{\alpha M} \prod_{j=1}^{M} [c_j(I, I_\mathrm{r}]^{\beta_j} [s_j(I, I_\mathrm{r})]^{\gamma_j}) \tag{16}$$

where the exponent of each term is used to adjust its relative importance.

As stated previously, MS-SSIM requires a reference for calculating the score. In this evaluation, the reference image is the overlapping region of the back spherical view before the blending step. Besides, the test image is the overlapping region of the full spherical view (i.e. after the blending step). In other words, the test image is the result of blending the reference image and the overlapping region of the front spherical view.

Figure 14 shows the mean scores of the four full spherical views generated using either an affine matrix (light green bars) or a polynomial (dark green bars) as geometric transformation. Like in the previous evaluation, the results of each scene are studied separately.

In general terms, this figure shows that the results obtained with a polynomial transformation have higher quality than applying an affine matrix. This difference is less noteworthy for the Calibration-based Projection Method (CPM) or for the two possible proposed methods (EFP+θ and EFP+θ+r) than for equidistant fisheye projection (EFP). In this last case, the use of a polynomial greatly improves the quality based on this measure with respect to the affine In general terms, this figure shows the results obtained with a polynomial transformation.

Concerning the projection method, we can see that the equidistant fisheye projection (EFP) has the worst quality based on MS-SSIM using affine matrix. However, the results for the proposed methods (EFP+θ and EFP+θ+r) are good. In fact, they are more similar to the Calibration-based Projection Method.

4.3.4 Study of the misalignment error and computation time

Another approach to evaluate the stitching process, concretely the image registration, is to calculate the error after applying the geometrical transformation. This error can be calculated as the Euclidean distance between matching feature pairs found in a pair of spherical views.

In Fig. 15, we can visualize the error in pixels for all spherical view pairs: without transformation (Fig. 15a) and applying an affine (Fig. 15b) or a polynomial (Fig. 15c) as geometric transformation in one of the spherical views. A boxplot represents the error, and the number of matching features is indicated below.

About the results without transformation (Fig. 15a), the error is higher with the Calibration-based Projection Method (CPM). On the contrary, the proposed method with two corrections (EFP+θ+r) has the best result regarding the misalignment error, followed by the proposed method with one correction (EFP+θ). The Equidistant Fisheye Projection (EFP) has worse results than the Calibration-based Projection Method and the number of matching features is lower.

In relation to the results of affine transformation (Fig. 15b), the mean error value is lower, getting the best results for Calibration-based Projection Method (CPM) and the proposed method with two corrections (EFP+θ+r).

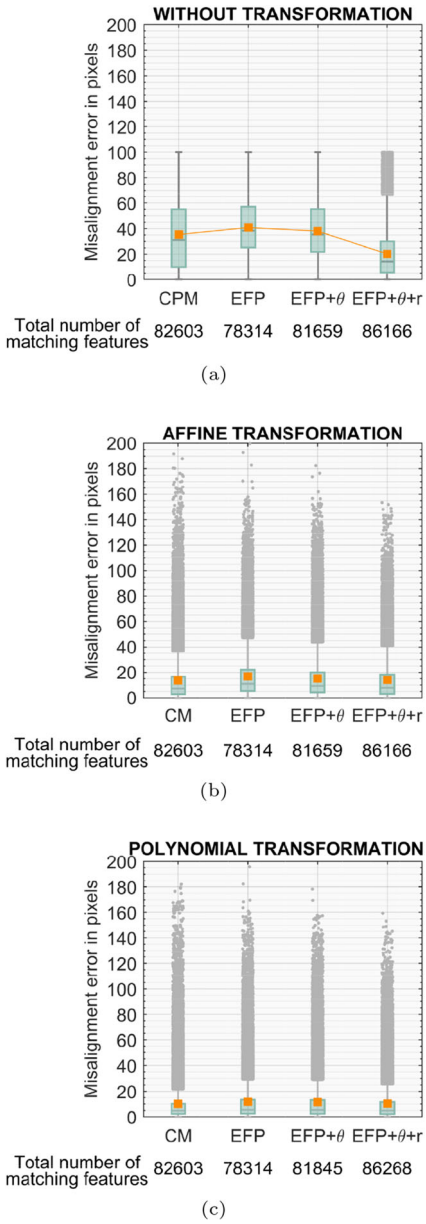


Fig. 15 Misalignment error in pixels considering each pair of spherical views: **a** without transformation and after estimating and applying **b** an affine geometric transformation or **c** a polynomial geometric transformation. The mean error for each case is represented by Orange color

Finally, concerning the results of polynomial transformation (Fig. 15c), the misalignment error is lower than in the two previous cases.

For some applications, the computation time is crucial. As a consequence, we also studied this factor. For each pair of fisheye images, the time spent to create a full spherical view (run the complete algorithm) was calculated. Table 4 shows the mean time of the 50 fisheye image pairs of the dataset for each possible combination between the two stages of the spherical view generation: the method used to project on the sphere (MP, EFP, EFP+θ or EFP+θ+r) and the method to estimate the geometric transformation (affine or polynomial) for the image registration.

Table 4 Average computation time of each possible combination

	CPM (s)	EFP (s)	EFP+θ (s)	EFP+θ+r (s)
Affine	368.88	36.02	36.87	37.59
Polynomial	375.48	36.77	37.70	38.35

After assessing these values, we can confirm that the fastest way to obtain a full spherical view is by using the Equidistant Fisheye Projection (EFP) and the affine matrix. The use of the proposed correction slightly increases the computation time, which is higher if the two polar coordinates are corrected (EFP+θ+r). This fact was expected since, first, estimating the correction parameters requires some additional time and, as we have observed in Fig. 15. Second, more matched features are detected when applying this correction, which implies a higher time during the image registration process. Despite these facts, the difference in computation time between the Equidistant Fisheye Projection (EFP) without and with correction is less than one second, whereas the computation time using the Calibration-based Projection Method (CPM) is approximately ten times higher than using the equidistant fisheye projection.

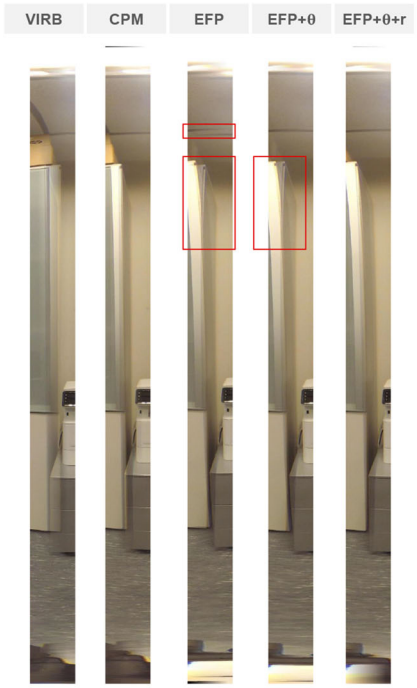
Regarding the geometric transformation, the polynomial takes more computation time than the affine matrix, but the difference is slight (less than one second in all cases).

4.3.5 Visual qualitative assessment

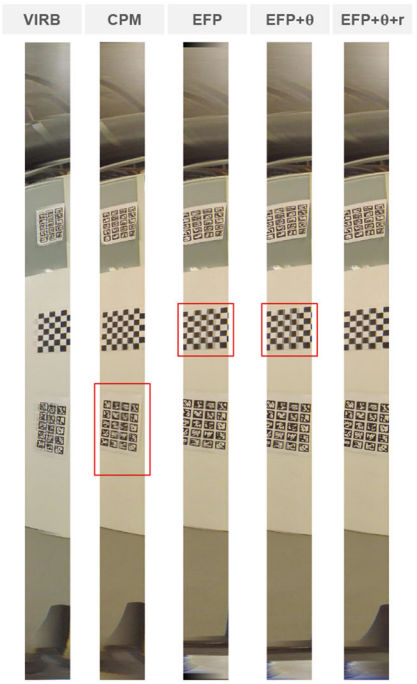
In the previous sections, a quantitative evaluation has been addressed. However, sometimes, the image quality metric may not agree with the quality perception appreciated by a human or not take into account all possible stitching artifacts. Considering, we also propose a qualitative evaluation. This section is divided into two parts. In the first one, the qualitative evaluation is performed with the camera Garmin VIRB 360. In contrast, the second part shows the qualitative evaluation of the proposed correction step for another dual fisheye camera, the Samsung Gear 360.

Garmin VIRB 360 The quantitative evaluation results establish that the correction step proposed and the polynomial provide a spherical view with good quality, as the quantitative assessment has shown. Now, these views are evaluated qualitatively. Figure 16 shows an overlapping region for each scenario. In each example, we can see and compare the four full spherical views calculated with the algorithm and the ones provided by the Garmin VIRB 360.

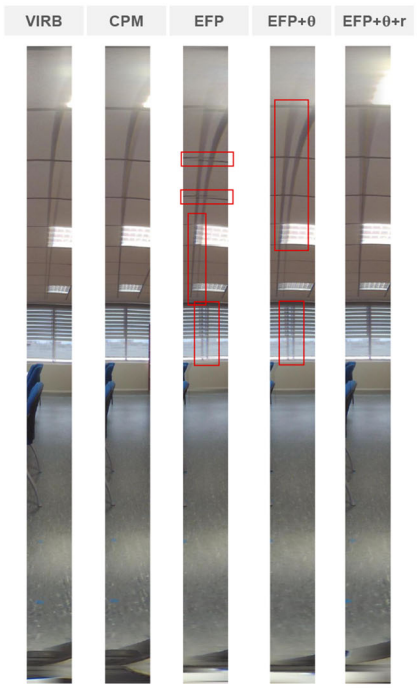
Analyzing Fig. 16, we can state that the vertical error after applying the Equidistant Fisheye Projection (EFP) is reduced with the correction of the θ coordinate (EFP+θ). On



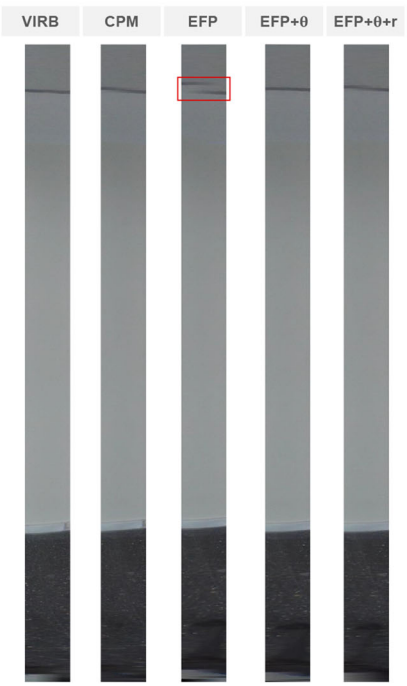
(a) Office



(b) Laboratory

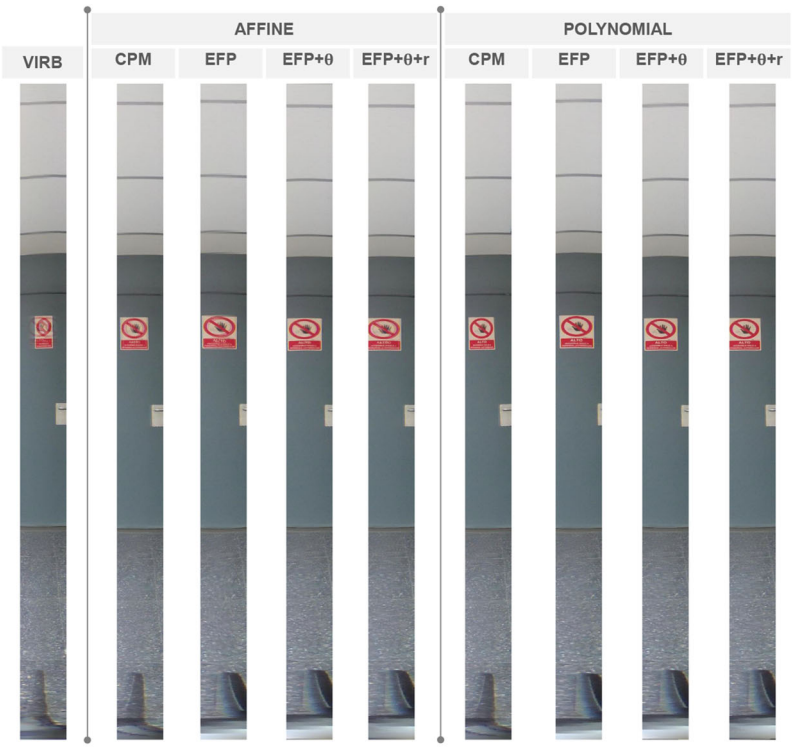


(c) Meeting room

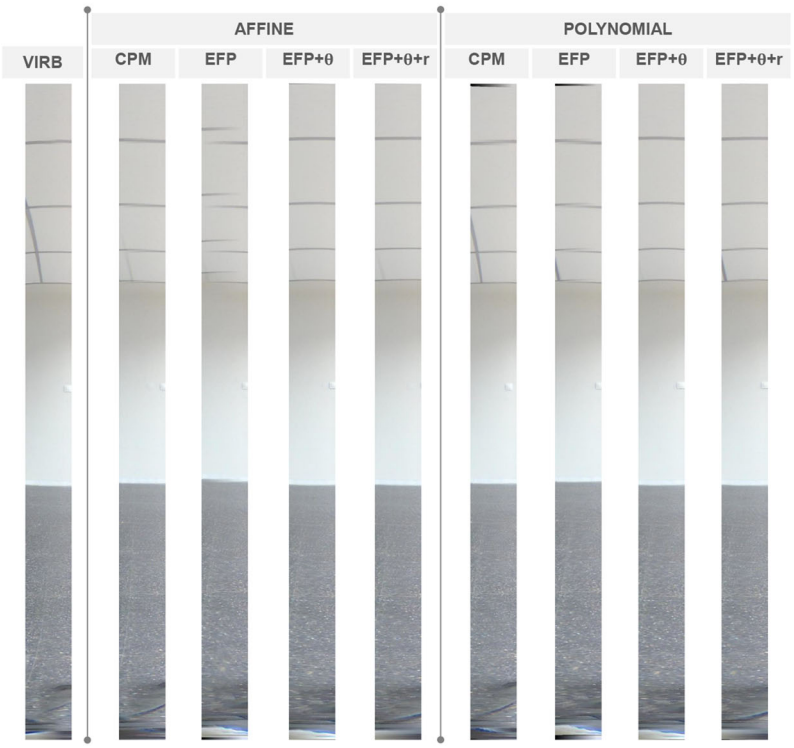


(d) Hall

Fig. 16 Overlapping zones of the final images generated with CPM, EFP, EFP+θ and EFP+θ+r (using affine matrix) and of the image provided by the Garmin VIRB 360



(a) Left overlapping region



(b) Right overlapping region

Fig. 17 Left (**a**) and right (**b**) overlapping regions of the different full spherical views at the same position

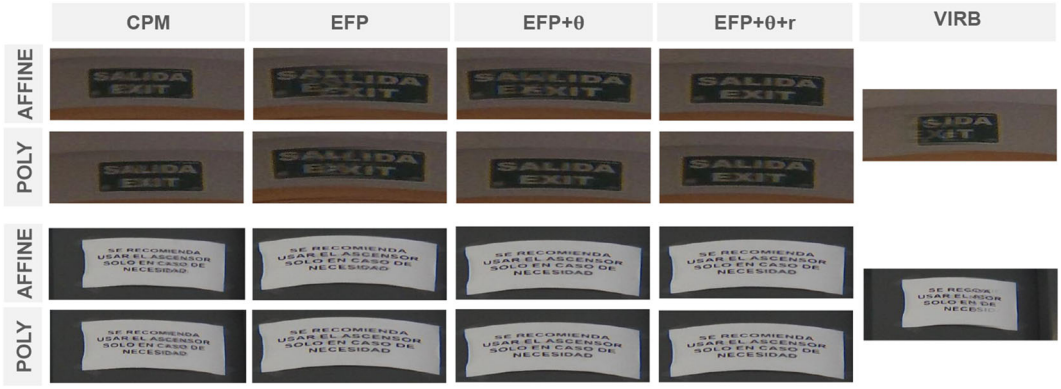


Fig. 18 Examples of signs captured in the overlapping region

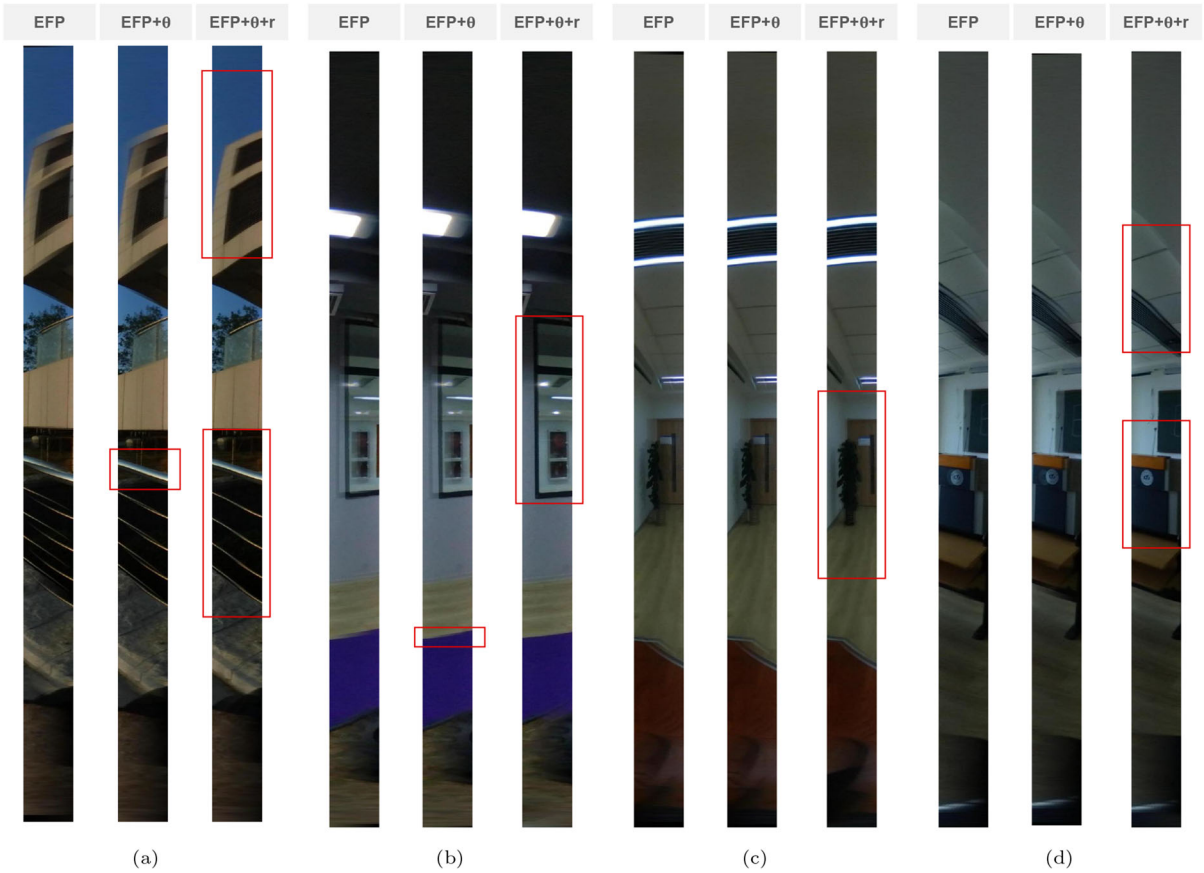


Fig. 19 Samsung Gear 360: overlapping zones of the final images generated with EFP, EFP+θ and EFP+θ+r (using affine matrix)

the contrary, the horizontal error is lower with the additional correction of the r coordinate (EFP+θ+r).

As previously described, we have not only applied an affine matrix to align both spherical views, but we have also proposed to use a polynomial. In this regard, Fig. 17 shows the same overlapping regions belonging to full spherical views arising from the combination of the different projections from sphere to fisheye image and the geometric transformation types. These overlapping regions are part of the same full spherical view. Figure 17a presents the left overlapping region and Fig. 17b the right overlapping region.

By visually comparing them, we can say that the quality of the full spherical view is higher using a polynomial than an affine matrix, as happened in the evaluation based on MS-SSIM. This fact can be clearly seen in those image regions which are rich in texture.

After all these studies, it can be confirmed that the proposed method produces a powerful solution for applications

such as objects, text of face recognition, jointly considering the qualitative and quantitative quality results and the computing time. As an example of this, in Fig. 18, the signs or informative posters can be easily recognized and readable.

Samsung Gear 360 The functions of the proposed correction step have been estimated using experimental data provided by images captured by a Garmin VIRB 360 camera. Therefore, this section evaluates the proposed correction step qualitatively using another dual fisheye commercial camera. For this evaluation, we have chosen a set of fisheye image pairs of the publicly available Cross-reference dataset [44]. For this dataset, the authors use a set of Samsung Gear 360 cameras.

The main difference of the Samsung Gear 360 camera with respect to the Garmin VIRB 360 camera is that the fisheye lenses of the first one have a shorter field of view. Each fisheye lens of the Samsung Gear 360 camera has 195 degrees of field of view.

Figure 19 shows the overlapping zones of four dual fisheye images of the set. For each dual fisheye image, the algorithm has been run three times: (1) using the equidistant fisheye projection, (2) the equidistant fisheye projection and correcting the angular polar coordinate and (3) the equidistant fisheye projection and correcting both polar coordinates. The affine matrix is the geometric transformation to align the pair of spherical views.

Analyzing visually the overlapping zones, we can confirm that the quality of the spherical views generated by employing the correction step is improved.

5 Conclusions

The purpose of this paper is to propose some algorithms to achieve a high-quality full spherical view from dual fisheye images. A Garmin VIRB 360 camera is used to obtain the experimental datasets.

The algorithm implemented to generate the full spherical view from dual fisheye images has different variations associated to: (a) the equations to project from the sphere to the fisheye image during the spherical format transformation stage, and (b) the transformation used to align the pair of spherical views. Regarding the projection, the options are: a Calibration-based Projection Method (CPM), the Equidistant Fisheye Projection (EFP), and, also, this latter combined with a correction step which is one of the main contributions of this paper (EFP+θ or EFP+θ+r). As for the alignment of the spherical views pair, the options of transformation are: an affine matrix or a polynomial.

To determine the performance of each configuration, a variety of evaluations has been carried out. From the results obtained using a no-reference method based on sharpness and visual qualitative assessment, we can conclude that the full spherical view provided by the proposed algorithm is a good solution that substantially improves the one provided by the Garmin VIRB 360 camera.

Concerning to the generated full spherical views, the conclusion is that the correction step proposed and the alignment based on polynomial improve the quality of the view, being very similar to the generated using a Calibration-based Projection Method (CPM). This fact is expected since more effectiveness is achieved with the calibration but requires a previous process, whereas it is not necessary with the proposed correction. However, the drawback we have observed is that both contributions depend on local feature points. The proposed correction step and the registration process are based on feature matches between dual fisheye images and spherical views, respectively. Besides, estimating the parameters of the correction step and the polynomial requires a certain number of relevant pairs (e.g. at least six for the polynomial). If the requirement is not reached, the estimation of these parameters may not be adequate and cause lower image quality.

The results show the validity of the proposed correction step, specially in the image areas with more texture. Then, as future work, we will evaluate the utility of the resulting full spherical views in some high-level tasks, such as people or object detection and recognition (specially when they are in the overlapping areas), mapping or localization of mobile robots.

Acknowledgements This work is part of the project TED2021-130901B-I00 funded by MCIN/AEI/10.13039/501100011033 and by the European Union "NextGenerationEU"/PRTR, of the project PROMETEO/2021/075 funded by Generalitat Valenciana, and of the grant ACIF/2020/141 funded by Generalitat Valenciana and Fondo Social Europeo (FSE).

Author Contributions Conceptualization: DV, LP, OR; Methodology: MF, DV, LP; Software: MF, DV, AP; Formal analysis and investigation: MF, AP; Writing—original draft preparation: MF, AP; Writing—review and editing: MF, DV, LP, OR, AP; Supervision: DV, OR, LP; Funding Acquisition and Project administration: OR, LP.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Data availability To clearly show the performance of the proposal, a selection of results has been included in [48]. This link includes one folder with the results obtained in the *Meeting room* scenario, and one folder with the results of *Hall* (these are two of the scenarios described in Table 3). Each folder contains results for a number of positions of the GARMIN VIRB 360 on the ground plane (11 positions in the *Meeting room* and 9 in the *Hall*). For every position, the folder includes the initial dual fisheye images, the full spherical view provided by the camera and

the full spherical views generated with the eight possible variations of the algorithms described in this paper.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Cebollada, S., Payá, L., Flores, M., Román, V., Peidró, A., Reinoso, O.: A Localization Approach Based on Omnidirectional Vision and Deep Learning. In: Gusikhin, O., Madani, K., Zaytoon, J. (eds.) Informatics in Control, Automation and Robotics, pp. 226–246. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-92442-3_13
2. Román, V., Payá, L., Peidró, A., Ballesta, M., Reinoso, O.: The role of global appearance of omnidirectional images in relative distance and orientation retrieval. Sensors **21**(10), 3327 (2021). https://doi.org/10.3390/s21103327
3. Zhang, J., Yin, X., Luan, J., Liu, T.: An improved vehicle panoramic image generation algorithm. Multimed. Tools Appl. **78**(19), 27663–27682 (2019). https://doi.org/10.1007/s11042-019-07890-w
4. Delmas, S., Morbidi, F., Caron, G., Albrand, J., Jeanne-Rose, M., Devigne, L., Babel, M.: SpheriCol: A Driving Assistance System for Power Wheelchairs Based on Spherical Vision and Range Measurements. In: 2021 IEEE/SICE International Symposium on System Integration (SII), pp. 505–510. IEEE, Iwaki, Fukushima, Japan (2021). https://doi.org/10.1109/IEEECONF49454.2021.9382766
5. Ha, V.K.L., Chai, R., Nguyen, H.T.: A telepresence wheelchair with 360-Degree vision using WebRTC. Appl. Sci. **10**(1), 369 (2020). https://doi.org/10.3390/app10010369
6. Morbidi, F., Devigne, L., Teodorescu, C.S., Fraudet, B., Leblong, E., Carlson, T., Babel, M., Caron, G., Delmas, S., Pasteau, F., Vailland, G., Gouranton, V., Guegan, S., Le Breton, R., Ragot, N.: Assistive Robotic Technologies for Next-Generation Smart Wheelchairs: Codesign and Modularity to Improve Users' Quality of Life. IEEE Robotics & Automation Magazine, 2–14 (2022). https://doi.org/10.1109/MRA.2022.3178965
7. Cebollada, S., Payá, L., Jiang, X., Reinoso, O.: Development and use of a convolutional neural network for hierarchical appearance-based localization. Artif. Intell. Rev. **55**(4), 2847–2874 (2022). https://doi.org/10.1007/s10462-021-10076-2
8. Rana, A., Ozcinar, C., Smolic, A.: Towards generating ambisonics using audio-visual cue for virtual reality. In: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2012–2016. IEEE, Brighton, United Kingdom (2019). https://doi.org/10.1109/ICASSP.2019.8683318
9. Saura-Herreros, M., Lopez, A., Ribelles, J.: Spherical panorama compositing through depth estimation. Vis. Comput. **37**(9), 2809–2821 (2021). https://doi.org/10.1007/s00371-021-02239-7
10. Gledhill, D., Tian, G.Y., Taylor, D., Clarke, D.: Panoramic imaging-a review. Comput. Graph. **27**(3), 435–445 (2003). https://doi.org/10.1016/S0097-8493(03)00038-4
11. Samsung: Gear 360 (2017) | Samsung Soporte España. https://www.samsung.com/es/support/model/SM-R210NZWAPHE/ Accessed 2022-11-18
12. Ricoh: Producto | RICOH THETA S. https://theta360.com/es/about/theta/s.html Accessed 2022-11-18
13. Garmin: VIRB 360. https://www.garmin.com/es-ES/p/562010 Accessed 2022-11-18
14. Colonnese, S., Cuomo, F., Ferranti, L., Melodia, T.: Efficient video streaming of 360° cameras in unmanned aerial vehicles: an analysis of real video sources. In: 2018 7th European Workshop on Visual Information Processing (EUVIP), pp. 1–6 (2018). https://doi.org/10.1109/EUVIP.2018.8611639
15. Benseddik, H.-E., Morbidi, F., Caron, G.: PanoraMIS: an ultra-wide field of view image dataset for vision-based robot-motion estimation. Int. J. Robot. Res. **39**(9), 1037–1051 (2020). https://doi.org/10.1177/0278364920915248
16. Zhang, Y., Huang, F.: Panoramic visual slam technology for spherical images. Sensors **21**(3), 705 (2021). https://doi.org/10.3390/s21030705
17. Zhang, Z., Rebecq, H., Forster, C., Scaramuzza, D.: Benefit of large field-of-view cameras for visual odometry. In: 2016 IEEE International Conference on Robotics and Automation (ICRA), pp. 801–808 (2016). https://doi.org/10.1109/ICRA.2016.7487210
18. Zhang, J., Xiu, Y.: Image stitching based on human visual system and SIFT algorithm. Vis. Comput. (2023). https://doi.org/10.1007/s00371-023-02791-4
19. Lyu, W., Zhou, Z., Chen, L., Zhou, Y.: A survey on image and video stitching. Virtual Real. Intell. Hardw. **1**(1), 55–83 (2019). https://doi.org/10.3724/SP.J.2096-5796.2018.0008
20. Lee, S.-H., Lee, S.-J.: Development of remote automatic panorama VR imaging rig systems using smartphones. Clust. Comput. **21**(1), 1175–1185 (2018). https://doi.org/10.1007/s10586-017-0930-4
21. Zhang, W., Wang, Y., Liu, Y.: Generating high-quality panorama by view synthesis based on optical flow estimation. Sensors **22**(2), 470 (2022). https://doi.org/10.3390/s22020470
22. Flores, M., Valiente, D., Gil, A., Reinoso, O., Payá, L.: Efficient probability-oriented feature matching using wide field-of-view imaging. Eng. Appl. Artif. Intell. **107**, 104539 (2022). https://doi.org/10.1016/j.engappai.2021.104539
23. Cabrera, J.J., Cebollada, S., Flores, M., Reinoso, O., Payá, L.: Training, optimization and validation of a CNN for room retrieval and description of omnidirectional images. SN Comput. Sci. **3**(4), 271 (2022). https://doi.org/10.1007/s42979-022-01127-8
24. Yang, L., Li, L., Xin, X., Sun, Y., Song, Q., Wang, W.: Large-Scale Person Detection and Localization using Overhead Fisheye Cameras (2023). https://doi.org/10.48550/ARXIV.2307.08252
25. Wang, T., Hsieh, Y.-Y., Wong, F.-W., Chen, Y.-F.: Mask-RCNN Based People Detection Using A Top-View Fisheye Camera. In: 2019 International Conference on Technologies and Applications of Artificial Intelligence (TAAI), pp. 1–4. IEEE, Kaohsiung, Taiwan (2019). https://doi.org/10.1109/TAAI48200.2019.8959887
26. Tian, C., Chai, X., Shao, F.: Stitched image quality assessment based on local measurement errors and global statistical properties. J. Vis. Commun. Image Represent. **81**, 103324 (2021). https://doi.org/10.1016/j.jvcir.2021.103324
27. Krams, O., Kiryati, N.: People detection in top-view fisheye imaging. In: 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1–6. IEEE, Lecce, Italy (2017). https://doi.org/10.1109/AVSS.2017.8078535. http://ieeexplore.ieee.org/document/8078535/

28. Cai, Y., Li, X., Wang, Y., Wang, R.: An overview of panoramic video projection schemes in the IEEE 1857.9 standard for immersive visual content coding. IEEE Trans. Circuits Syst. Video Technol. **32**(9), 6400–6413 (2022). https://doi.org/10.1109/TCSVT.2022.3165878
29. Ni, G., Chen, X., Zhu, Y., He, L.: Dual-fisheye lens stitching and error correction. In: 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), pp. 1–6. IEEE, Shanghai (2017). https://doi.org/10.1109/CISP-BMEI.2017.8302053
30. Lin, B.-H., Cheng, H.-Z., Li, Y.-T., Guo, J.-I.: 360 Degree Fish Eye Optical Construction For Equirectangular Projection of Panoramic Images. In: 2020 International Conference on Pervasive Artificial Intelligence (ICPAI), pp. 194–198. IEEE, Taipei, Taiwan (2020). https://doi.org/10.1109/ICPAI51961.2020.00043
31. Xue, L., Zhu, J., Zhang, H., Liu, R.: A high-quality stitching algorithm based on fisheye images. Optik **238**, 166520 (2021). https://doi.org/10.1016/j.ijleo.2021.166520
32. Lo, I.-C., Shih, K.-T., Chen, H.H.: Efficient and accurate stitching for 360° dual-fisheye images and videos. IEEE Trans. Image Process. **31**, 251–262 (2022). https://doi.org/10.1109/TIP.2021.3130531
33. Szeliski, R.: Image Alignment and Stitching: A Tutorial. Foundations and trends in computer graphics and vision. now publishers Inc, Hanover (2006)
34. Ho, T., Budagavi, M.: Dual-fisheye lens stitching for 360-degree imaging. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2172–2176. IEEE, New Orleans, LA (2017). https://doi.org/10.1109/ICASSP.2017.7952541
35. Ho, T., Schizas, I.D., Rao, K.R., Budagavi, M.: 360-degree video stitching for dual-fisheye lens cameras based on rigid moving least squares. In: 2017 IEEE International Conference on Image Processing (ICIP), pp. 51–55. IEEE, Beijing (2017). https://doi.org/10.1109/ICIP.2017.8296241
36. Lo, I.-C., Shih, K.-T., Chen, H.H.: Image Stitching for Dual Fisheye Cameras. In: 2018 25th IEEE International Conference on Image Processing (ICIP), pp. 3164–3168. IEEE, Athens (2018). https://doi.org/10.1109/ICIP.2018.8451333
37. Souza, T., Roberto, R., Silva do Monte Lima, J.P., Teichrieb, V., Quintino, J.P., da Silva, F.Q.B., Santos, A.L.M., Pinho, H.: 360 Stitching from Dual-Fisheye Cameras Based on Feature Cluster Matching. In: 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), pp. 313–320. IEEE, Parana (2018). https://doi.org/10.1109/SIBGRAPI.2018.00047
38. Scaramuzza, D., Martinelli, A., Siegwart, R.: A Toolbox for Easily Calibrating Omnidirectional Cameras. In: 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 5695–5701 (2006). https://doi.org/10.1109/IROS.2006.282372. ISSN: 2153-0866
39. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: An efficient alternative to SIFT or SURF. In: 2011 International Conference on Computer Vision, pp. 2564–2571 (2011). https://doi.org/10.1109/ICCV.2011.6126544. ISSN: 2380-7504
40. Anand, S., Priya, L.: A Guide for Machine Vision in Quality Control, 1st edn. CRC Press, Boca Raton (2019)
41. Prados, R., Garcia, R., Neumann, L.: State of the Art in Image Blending Techniques. In: Image Blending Techniques and Their Application in Underwater Mosaicing. SpringerBriefs in Computer Science, pp. 35–60. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-05558-9_3
42. Ghosh, D., Kaabouch, N.: A survey on image mosaicing techniques. J. Vis. Commun. Image Represent. **34**, 1–11 (2016). https://doi.org/10.1016/j.jvcir.2015.10.014
43. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004). https://doi.org/10.1109/TIP.2003.819861
44. Li, J., Yu, K., Zhao, Y., Zhang, Y., Xu, L.: Cross-Reference Stitching Quality Assessment for 360° Omnidirectional Images. In: Proceedings of the 27th ACM International Conference on Multimedia, pp. 2360–2368. ACM, Nice France (2019). https://doi.org/10.1145/3343031.3350973. https://dl.acm.org/doi/10.1145/3343031.3350973
45. Li, J., Zhao, Y., Ye, W., Yu, K., Ge, S.: Attentive deep stitching and quality assessment for 360° omnidirectional images. IEEE J. Select. Top. Signal Process. **14**(1), 209–221 (2020). https://doi.org/10.1109/JSTSP.2019.2953950
46. Duan, H., Min, X., Sun, W., Zhu, Y., Zhang, X.-P., Zhai, G.: Attentive deep image quality assessment for omnidirectional stitching. IEEE J. Select. Top. Signal Process. (2023). https://doi.org/10.1109/JSTSP.2023.3250956
47. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In: The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, pp. 1398–1402. IEEE, Pacific Grove, CA, USA (2003). https://doi.org/10.1109/ACSSC.2003.1292216
48. ARVC: Laboratorio de Automatización Robótica y Visión por Computador (ARVC) - UMH. https://arvc.umh.es/db/360views/. Online; accessed 16 February 2023

María Flores holds a Bachelor's degree in Electronic and Industrial Automation Engineering from Miguel Hernández University (UMH) in 2017 and a Master's degree in Robotics from Miguel Hernández University (UMH) in 2018. Since 2019, she is at the Miguel Hernández University as Ph.D. Candidate student. The topic of research is focused on wide-FOV vision, visual localization with feature points and visual mapping. Since 2020 she has a Ph.D.-Candidate scholarship supported by Valencian Government (ACIF/2020/141).

David Valiente received the M.Eng. degree in Telecommunications Engineering in 2009, and the Ph.D. degree in Industrial and Telecommunications Technologies in 2016, both with honours. Since 2009, he works as a researcher at the Systems Engineering and Automation Department of the Miguel Hernández University. His teaching experience lies in subjects of robotics perception, general electronics and electronic measurement and instrumentation. His research interests comprise visual localization, visual mapping, feature matching and wide-FOV vision, like omnidirectional and fisheye images.

Adrián Peidró holds a M.Eng. degree in Mechanical Engineering (2013) and a Ph.D. in Industrial Technologies (2018), both at Miguel Hernandez University (UMH) of Elche, Spain. He has authored 17 scientific papers published in JCR-indexed journals and 50 conference papers. His research interests are focused on parallel robots, climbing robots, artificial intelligence, robot simulation, and robots in education

Oscar Reinoso received the industrial engineer and Ph.D. degrees from Polytechnic University of Madrid (UPM) in 1991 and 1996 respectively. From 1994 to 1997 he works in the Research Development department of Protos Desarrollo in a visual inspection system. Since 1997, he has been at the Miguel Hernández University, as professor in control, robotics and computer vision. His research interests include robotics, teleoperated robots, climbing robots, visual servoing, visual inspection systems. He is author of several books, papers and communications in the cited topics. Prof. Reinoso is a member of the CEA-IFAC and senior member IEEE.

Luis Payá holds a M. Eng. in Industrial Engineering (Spain, 2002) and a Ph.D. in Industrial Technologies (Spain, 2014). Currently he works as associate professor at Miguel Hernández University in Spain (Department of Systems Engineering and Automation). He teaches some subjects related to the fields of automatic control, electronics and robotics. His current research interests include omnidirectional vision and global appearance algorithms; topological map building and localization of mobile robots; and also implementation and testing of remote laboratories. He is author of several books, papers and communications in the cited topics.