HDROmni : Extension optique de gamme dynamique en vision panoramique pour la robotique

Julien Jean Ducrocq, Guillaume Caron, El Mustapha Mouaddib

HDROmni : Extension optique de gamme dynamique en vision panoramique pour la robotique

HDROmni: Optical extension of dynamic range for panoramic robot vision

Julien Ducrocq[1] Guillaume Caron[1,2] El Mustapha Mouaddib[1]
[1] Université de Picardie Jules Verne, laboratoire Modélisation, Information et Systèmes
33 rue Saint-Leu 80039 Amiens, France
[2] CNRS-AIST Joint Robotics Laboratory, IRL, Tsukuba, Japon
{julien.ducrocq, guillaume.caron, mouaddib}@u-picardie.fr

Résumé

Cet article présente la méthode de conception qui permet d'étendre optiquement la gamme dynamique d'une caméra panoramique pour la robotique. La large gamme de radiance d'une scène est capturée en temps réel par une combinaison de miroirs paraboliques, une lentille télécentrique et des filtres à densité neutre.

Le premier prototype, nommé HDROmni, est constitué de 4 miroirs et de 3 filtres à densité neutre. Il augmente 100 fois la gamme dynamique de la caméra utilisée. Plusieurs évaluations comparent la nouvelle caméra avec des caméras à dynamique classique dans plusieurs scènes difficiles. Elles montrent qu'en moyenne, la qualité de la gamme dynamique de l'image double. Des évaluations supplémentaires montrent que HDROmni permet aussi d'accroître le nombre d'appariements de points caractéristiques. Enfin, HDROmni engendre aussi un domaine de convergence plus étendu et une meilleure précision à convergence de l'asservissement visuel direct (basé intensités de pixels) d'un robot mobile à l'extérieur.

Mots Clef

Vision omnidirectionnelle; imagerie HDR; asservissement visuel.

Abstract

This paper presents the design method to optically extend the perceptible dynamic range of panoramic vision for robotics. The large range of scene radiance is captured in real-time by a combination of paraboloidal mirrors, a telecentric lens and neutral density filters.

The first prototype, named HDROmni, is made of 4 mirrors and 3 neutral density filters. It increases 100 times the perceptible dynamic range of the used camera. Results in various challenging scenes compare the new camera to low dynamic range ones. They show that the dynamic range quality of images doubles, on average. Supplementary results show significant benefit for feature-based approaches, i.e. much more matched features. Finally, HDROmni also causes a larger convergence domain and a much higher positionning accuracy of direct visual servoing (based on pixel intensities) of a mobile robot outdoors.

Keywords

Omnidirectional vision; HDR imaging; visual servoing.

1 Introduction

Real-time panoramic imaging refers to a one-shot camera of $360^o \times 180^o$ field of view, or more. Such a field of view is obtained with a fisheye lens [13] or with a catadioptric optics including a curved mirror [27].

It decreases the acquisition time to make virtual tours [25] and enhances video-surveillance capabilities while avoiding the use of motorized cameras [5]. In robotics, it improves robustness and accuracy of visual odometry [32] as well as enlarges the convergence domain of direct visual servoing [7].

However, when the illumination of the scene is significantly different in several directions, the image includes under- (Fig. 1b) and over-exposed (Fig. 1c) regions due to the dynamic range (DR) of the camera, that is lower than the scene DR. Indeed, the DR of scene radiance is commonly between 80 and 120 dB [20] whereas few CMOS sensors of linear response, in machine vision cameras, reach 70 dB[1], *eg.* 74.3 dB for Sony's IMX287 [12]. Even though sensors of (pseudo-) logarithmic response overtake 80 dB of DR, *eg.* 100 dB for ON Semicon's MT9V032, there is still a gap to reach 120 dB. So, even when combining a fisheye lens with one of such sensors, there are still risks of non perceived areas,

[1] *Irradiance* DR (iDR), computed with the max/min perceptible irradiances, can be different from *digital* DR (dDR), computed from the sensor bit depth (*eg.* ON Semicon's 12bits MT9P031: dDR = 72 dB; iDR = 60 dB [3]).

thus minimizing the interest of panoramic vision in scenes of challenging lighting.

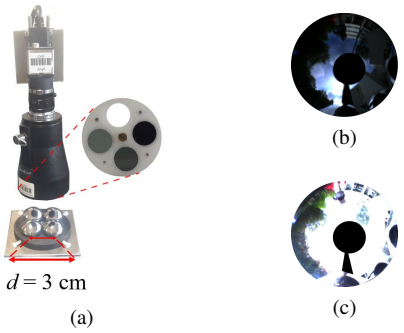


Figure 1: (a) HDROmni (filters zoomed, right) to overcome the dynamic range limits of LDR panoramic images with exposure times 0.25 (b) and 6.5 ms (c).

To handle the latter issue, one increases the DR of a camera either with hardware processing on CMOS chip [31], however not available to consumers yet, or by acquiring several low DR (LDR) images of the same scene with various exposure times, then merging them in a single high DR (HDR) image [10, 11, 15, 21, 30]. The latter methodology samples the range of scene radiance to overcome the DR limit of a camera. However, it implies a static camera in a static scene, which is not suitable for robotics.

The camera motion can be estimated for compensation in images before merging them. However, the DR of images acquired successively [15], or in parallel [4], must overlap significantly for image matching. Alternatively, some design put optical masks, of spatially varying transmittance, directly next to the photosensitive matrix of the camera [24, 28]. Recently, neural networks are able to predict the HDR image from a LDR one [16], [17], [19], to increase the quality of a HDR image resulting of the fusion of a LDR stack [29] or to cancel ghosting artifacts due to camera shakes [18].

All the previous methods can almost directly be applied to panoramic imaging, *e.g.* [14], but not for dynamic robotics. Indeed, [15], the fastest learning-based HDR video generation method we found, would induce a latency of 590 ms, when considering a 20 Hz image acquisition framerate: 150 ms to sequentially acquire three exposures + 440 ms to process 430x430 pixels[2] (same resolution as in the experiments of this article, see Sec. IV and V). Hence, only the generalized assorted pixel camera [28] and the parallel acquisitions of LDR images of different exposures [4] may lead to real time and low latency required by robotics applications, including control, thanks to capturing, not just predicting, a high DR, although at the price of hardware complexity or size.

We propose a new camera design approach between both latter seminal works [4], [28], *i.e.* easy to setup outside the camera box while compact as a single camera. Rigorously chosen neutral density (ND) filters are set between the camera lens and several curved mirrors, to make HDROmni, the real-time panoramic HDR camera (Fig. 1a) featuring:

- Maximum DR by enlarging the range of perceptible irradiances.
- Compactness, as existing catadioptric cameras.
- High precision positioning of mobile robot, outdoors.

The latter, considering HDROmni for photometric visual servoing [7, 8], is one of the most challenging among various real-time applications.

The rest of the paper details the design methodology (Sec. 2) of the optical extension, then the analytical description of environments where it can relevantly be used (Sec. 3). After that, Section 4 presents the first prototype as well as the evaluation of the DR quality of its images. Finally, Section 5 evaluates visual servoing with HDROmni w. r. t. the state-of-the-art, before conclusion (Sec. 6).

2 Multi-view Dynamic Range Extension

The multi-view dynamic range extension performed by HDROmni consists in acquiring images of different exposures in parallel. It is inspired from the ND filters-based multi-camera system for 3D HDR TV [4] which considers several (eight) cameras mounted on a rig, each having the same exposure time for synchronization purpose, but equipped with different ND filters, to optically *simulate* various exposure times. The novelty, in our work, is to consider real-time panoramic vision and a single camera.

2.1 Single Panoramic Camera DR Extension

The same idea as [4] applies to panoramic vision. However, combining a single camera and several mirrors [23, 26] prevents synchronization issues and may keep compactness. Particularly, [23], with four paraboloidal mirrors, makes viewpoints much closer than a rig of several panoramic cameras (Fig. 1a). Such proximity of the four panoramic viewpoints allows the HDR fusion from the basic superimposition of their LDR image regions, under camera-scene distance assumptions (Sec. 3.1).

We emphasize the fact that a multi-omnidirectional camera is generally designed for multi-view perception, *eg.* 3D reconstruction [9, 23]. However, the proposed panoramic HDR camera design targets *complementary* perception to increase the most the DR of the panoramic camera. Indeed, in a multi-camera system [4], optical densities of ND filters must be close to make neighbor image DRs overlap to

[2]The 440 ms are estimated from Table IV of [15] where 2.2 s are required to obtain the HDR image from 3 exposures.

allow image matching. Inversely, our design method is not constrained by such extra-processing. Hence, it selects ND filters for the minimum overlapping of individual DRs so that their combination reaches the largest DR.
Extending the four mirrors system of [23] with ND filters allows acquiring simultaneously four exposures with a single camera. As a first result, it virtually quadruples the bit depth of the acquired digital image, thus adds two bits to the bit depth of any machine vision camera. Such result increases the digital DR of the camera, but the rest of this section expresses optical densities of ND filters to increase the most the irradiance DR of the camera.

2.2 Optical Densities for Complementarity

We consider a panoramic camera made of $M \in \mathbb{N}^*$ mirrors and $M-1$ ND filters. Each filter, in front of a mirror (one mirror is directly seen), must be selected so that they gradually decrease the irradiance to simultaneously capture radiances in the scene, from the lowest to the highest. Assuming the sensor irradiance equals the scene radiance, the exposure $e \in \mathbb{R}_+$ is expressed proportionally to the irradiance $E \in \mathbb{R}_+$ of the image plane, thanks to the exposure time $t_e \in \mathbb{R}_+^*$ [11]:

$$e = t_e E. \tag{1}$$

Thus, the lowest irradiance that can be captured depends on the exposure time, set following the application requirements, *e.g.* 50 ms for 20 images per second. Then, the ND filter free mirror gets the exposure e_0 (of lowest E) and each ND filter $f \in [1, M-1] \subset \mathbb{N}$ achieves optically an exposure reduction to allow the camera sensor getting the highest irradiances E. Consecutive exposures $e_f \in \mathbb{R}_+$ of ND filters f are related by exposure ratios $R_{f,f-1} \in \mathbb{R}_+$, whose expression depends on the camera response function, either linear or non-linear as respectively recalled and introduced hereafter.

Exposure ratio for linear camera responses. In a few words, the camera response function (CRF) relates pixel brightness $B \in [0, 255] \subset \mathbb{N}$ (for an image depth of 8 bits) to the scene radiance, assumed to be equal to E (1). Then, for a camera of linear CRF, exposure ratios to acquire several LDR images, to be fused as an HDR one, follow the rule:

$$R_{f,f-1} = e_f / e_{f-1} = \lambda,\ \lambda \in \mathbb{R}_+ : \lambda \leq 1, \tag{2}$$

even for parallel acquisitions [24][3]. Then, additional exposures e_f are recursively computed as $e_f = e_{f-1} R_{f,f-1}$ [24], which is a geometric suite, *i.e.*:

$$e_f = e_0\ [R_{f,f-1}]^f. \tag{3}$$

If λ (2) tends to 1, the quantization of the perceptible range of scene radiance is the most regular, but the latter range is limited, particularly with a small number M of various exposures, *e.g.* $M = 4$. Inversely, if λ tends to zero, the perceptible range of scene radiance appears wide but at the price of a poor quantization. In the worst case, it leads to a single, partly useful, LDR image among the set of acquired ones, so far from an HDR acquisition. In practice, for cameras of linear response function, λ is set to 0.5 or 0.25 as a compromise [24].
In the next part, we propose a generalization of both the expression of $R_{f,f-1}$ (2), for some cameras of non-linear response function, and e_f (3).

Generalized exposure ratio. The expression of exposure ratios (2) must be extended to fit cameras of non-linear CRF since they have a higher DR potential (Sec. 1). Non-linear CRF can be approximated by polynomial [22] or logarithmic functions [11]. However, in this paper, we focus on the class of cameras whose non-linear CRF can be piece-wise linearly approximated of an order 2 (Fig. 2a, between irradiances of 0 and E_{0max}).
Define E_{0max} as the maximum perceptible irradiance for exposure e_0 and E_{0dis} as the irradiance at which the slope of the CRF changes. For a given exposure time, brightnesses B_{0max} and B_{0dis} are measured for irradiances E_{0max}, respectively E_{0dis}. Then, we define the range of exposure ratios $R_{f,f-1}$ constraining the first slope of the CRF of the second exposure, *i.e* e_1, to lie between slopes of both linear parts of the CRF associated to e_0 as:

$$\frac{B_{0max} - B_{0dis}}{E_{0max} - E_{0dis}} \geq \frac{1}{R_{f,f-1}} \geq E_{0max}/E_{0dis}. \tag{4}$$

Then, setting $E_{md} = E_{0max}/E_{0dis}$, we generalize (2) as:

$$R_{f,f-1} = \lambda / E_{md}. \tag{5}$$

Obviously, if the camera is linear, $E_{md} = 1$ and (5) simplifies as (2). Thus, the latter generalized expression of $R_{f,f-1}$ fits both non-linear and linear CRF. Then, as done in previous works, we can set empirically λ, but *bounded* (4).
After that, in order to get the most complementary exposures, we reformulate the expression of e_f (3), such that e_f decreases faster than following a geometric suite, by making the power of $R_{f,f-1}$ growing itself as a geometric suite:

$$e_f = e_0\ [R_{f,f-1}]^m,\ m = 2^{f-1}. \tag{6}$$

Equation 6 first maximizes the range of perceptible scene radiance. Then, it minimizes the resampling of radiance ranges that were already, and better, sampled by higher exposures. Figure 2 compares, on an example, the impact of considering (3) (Fig. 2b) or (6) (Fig. 2c) on the range of perceptible scene radiance and radiances resampling.

Optical densities of ND filters. The exposure ratio (6) leads to densities of ND filters. We define the attenuation factor k_f of the f-th ND filter as:

$$k_f = [R_{f,f-1}]^{-m},\ m = 2^{f-1} \tag{7}$$

[3] Contrary to [24], we consider e_0 as the closest exposure to $E = 0$.

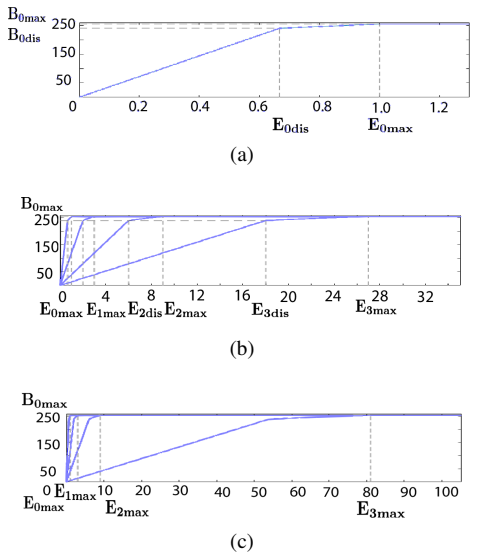


Figure 2: Non-linear CRF of a single camera (a) and 4 exposures e_f increasing as (b) a geometric suite and (c) exponentially ($\lambda = 1/2; E_{md} = 3/2$ for both, to ease comparison).

It simulates exposure e_f (6) while the camera itself keeps exposure time constant. Then, the optical density D_f of ND filter f is expressed as [1]:

$$D_f = log_{10}(k_f). \tag{8}$$

2.3 HDR fusion

The M exposures acquired in parallel lead to M LDR images I_{LDR_j}, one per mirror $j \in [0, M-1] \subset \mathbb{N}$. When the ratio of scene depth over baseline is large enough (see Sec. 3.1), the image disparity falls within a pixel, thus is almost null. Hence, every I_{LDR_j} can directly be fused into a HDR image I_{HDR}. For that, we consider *exposure fusion* [21], the best HDR algorithm according to the perceptual evaluation of Zeng *et al.* [30]. *Exposure fusion* computes three quality measures: well-exposedness (more weight to the pixels of brightness close to the median intensity); saturation (standard deviation in the three color channels) and contrast (Laplacian filter).

Writing ${}^{c_j}\mathbf{u} = ({}^{c_j}u, {}^{c_j}v)^\top \in \mathbb{R}^2$ the coordinates of pixel $I_{LDR_j}({}^{c_j}\mathbf{u})$ of the digital image I_{LDR_j}, the weight $w(I_{LDR_j}, {}^{c_j}\mathbf{u}) \in [0,1]$ is the product of the three quality measures computed for this pixel. Weights are normalized by dividing each $w(I_{LDR_{j_0}}, {}^{c_{j_0}}\mathbf{u}), j_0 \in [0, M-1] \subset \mathbb{N}$, by the sum of weights at coordinates corresponding to ${}^{c_{j_0}}\mathbf{u}$ in all I_{LDR_j}. Thanks to the single orthographic camera with multiple paraboloidal mirrors considered in Section 2.1, assuming mirrors are of the same shape and their summits define a plane parallel to the image plane allows expressing each ${}^{c_j}\mathbf{u}$ from ${}^{c_{j_0}}\mathbf{u}$ by pure translation in the image plane. The latter pure translation is computed from principal points $\mathbf{u}_{0_j}$ and $\mathbf{u}_{0_{j_0}}$ of, respectively, I_{LDR_j} and $I_{LDR_{j_0}}$, leading to ${}^{c_j}\mathbf{u} = {}^{c_{j_0}}\mathbf{u} + \mathbf{u}_{0_j} - \mathbf{u}_{0_{j_0}}$. The LDR images are then blended-in in a multi-scale approach which considers the weights.

Figure 3 shows a 860×860 pixels HDROmni frame and the resulting 430×430 pixels HDR panoramic image.

3 Operational constraints

We model the disparity of the sensor (Sec. 3.1) in order to define for which interval of depths the basic superimposition of LDR views is valid for HDR fusion (Sec. 3.2).

3.1 Disparity modeling

HDROmni is modeled as a rig of M panoramic cameras c_j each modeled by the unified central projection model with intrinsic parameters $\gamma_j = \{\alpha_{u_j}, \alpha_{v_j}, u_{0_j}, v_{0_j}, \xi_j\}$ [2]. Then, considering the rig frame equals the camera frame $\mathcal{F}_{c_0}$, without loss of generality, extrinsic parameters are made of transformation matrices ${}^{c_j}\mathbf{M}_{c_0} \in SE(3)$ from $\mathcal{F}_{c_0}$ to $\mathcal{F}_{c_j}$ [6].

Based on the latter knowledge, one can express the observed disparities $\boldsymbol{\delta}\mathbf{u} = (\delta u, \delta v) \in \mathbb{R}^2$, from two panoramic images, associated to a 3D point $\mathbf{X} = (X, Y, Z)^\top \in \mathbb{R}^3$ of the scene. That link formally defines the relationship between the scene depth and the image resolution to allow the HDR fusion from the basic superimposition of LDR panoramic images. In short, the disparity modeling is here focused on the worst case, *i.e.* leading to the highest disparity: a 3D point belonging to the median plane $\mathcal{P}_M$ (Fig. 4) between two mirrors and parallel to their axes of revolution, supposed exactly parallel.

A 3D point ${}^{c_0}\mathbf{X}$ expressed in $\mathcal{F}_{c_0}$ transforms to $\mathcal{F}_{c_j}$ by composition of its homogeneous representation ${}^{c_0}\tilde{\mathbf{X}} \in \mathbb{P}^3$ to the transformation matrix ${}^{c_j}\mathbf{M}_{c_0}$. Its digital image coordinates ${}^{c_j}\mathbf{u}$ (see Sec. 2.3) are computed as [6]:

$${}^{c_j}\tilde{\mathbf{u}} = pr_{\gamma_j}({}^{c_j}\mathbf{X}) = \mathbf{K} pr_{\xi_j}({}^{c_j}\mathbf{X}), \tag{9}$$

${}^{c_j}\tilde{\mathbf{u}} \in \mathbb{P}^2$ being the homogeneous representation of ${}^{c_j}\mathbf{u}$, $\mathbf{K} \in Aff(2)$ gathering the scale factors (α_{u_j} and α_{v_j}) and

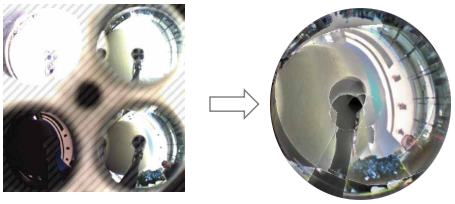


Figure 3: LDR panoramic regions of a HDROmni frame (left) are fused as a HDR image (right). Striped areas are ignored.

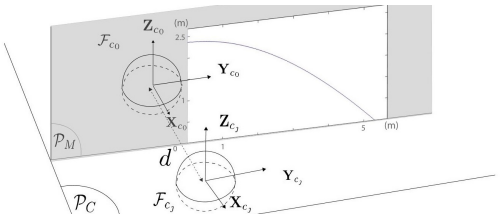


Figure 4: Geometric relationships of two coplanar mirrors.

principal point coordinates (u_{0_j}, v_{0_j}):

$$\mathbf{K} = \begin{bmatrix} \alpha_{u_j} & 0 & u_{0_j} \\ 0 & \alpha_{v_j} & v_{0_j} \\ 0 & 0 & 1 \end{bmatrix}. \tag{10}$$

In (9), $pr_{\xi_j} : \mathbb{R}^3 \longmapsto \mathbb{P}^2$ is the projection function to the normalized image plane, involving ξ_j, the distance between the unit sphere center and the second center of projection of the unified central projection model [2]:

$$\left[\frac{{}^{c_j}X}{{}^{c_j}Z+\xi_j\rho}, \frac{{}^{c_j}Y}{{}^{c_j}Z+\xi_j\rho}, 1\right]^\top = pr_{\xi_j}({}^{c_j}\mathbf{X}) \tag{11}$$

with $\rho = \sqrt{{}^{c_j}X^2 + {}^{c_j}Y^2 + {}^{c_j}Z^2}$.

To simplify the disparity expression, we suppose, in addition to the previously mentioned parallelism of mirror axes of revolution, the coplanarity of mirrors, thus simplifying transformation matrices ${}^{c_j}\mathbf{M}_{c_1}$ to pure translations in the mirrors plane $\mathcal{P}_C$ (Fig. 4), which is parallel to the image plane. Furthermore, by a wise choice of the orientation of $\mathcal{F}_{c_0}$, so that the $\mathbf{X}_{c_0}$ axis of its frame is aligned with the origin of $\mathcal{F}_{c_j}$, ${}^{c_j}\mathbf{M}_{c_0}$ is even simplified to the translation $t_{X_{1,j}} = -d$ along $\mathbf{X}_{c_0}$, $d \in \mathbb{R}_+$ being the distance between origins of $\mathcal{F}_{c_0}$ and $\mathcal{F}_{c_j}$ (see Fig. 4).

Then, since we focus on the maximum disparity theoretically observable, it constrains ${}^{c_0}X = d/2$ (${}^{c_j}X = -d/2$), and since axes of $\mathcal{F}_{c_0}$ and $\mathcal{F}_{c_j}$ are parallel and their axes $\mathbf{X}_{c_0}$ and $\mathbf{X}_{c_j}$ are coplanar, ${}^{c_0}Y = {}^{c_j}Y$ and ${}^{c_0}Z = {}^{c_j}Z$.

Under the last assumption that the mirrors are exactly the same, we have $\alpha_{u_0} = \alpha_{u_j}$, $\alpha_{v_0} = \alpha_{v_j}$ and $\xi_0 = \xi_j$. By setting wisely the orientation of the camera photosensitive matrix with respect to mirrors, the horizontal and vertical axes of the image can be set parallel to both first axes of $\mathcal{F}_{c_0}$, leading to $v_{0_0} = v_{0_j}$. Then, applying (9) to ${}^{c_0}\mathbf{X}$ and ${}^{c_j}\mathbf{X}$ leads to different horizontal coordinates u only, since only ${}^{c_0}X$ and ${}^{c_j}X$, and u_{0_0} and u_{0_j}, are different. Thus, the expression of the disparity $\boldsymbol{\delta}\mathbf{u}$ simplifies to δu only:

$$\delta u = {}^{c_0}u - {}^{c_j}u - \delta u_0, \tag{12}$$

where $\delta u_0 = u_{0_0} - u_{0_j}$ is the transformation in the image plane to match principal points of both panoramic image regions. Then, it leads to:

$$\delta u = \frac{\alpha_{u_0}}{{}^{c_0}Z + \xi_0\rho}({}^{c_0}X - {}^{c_j}X) = \frac{\alpha_{u_0} d}{{}^{c_0}Z + \xi_0\rho} \tag{13}$$

The latter disparity model is used to express constraints on the scene and on the image definition to allow the HDR fusion from the straightforward superimposition of LDR panoramic regions in the next Subsection 3.2.

3.2 Scene depth and image definition constraints

For a given image definition, (13) leads to the minimum camera to 3D points distance ensuring the maximum disparity lies in a pixel:

$$|\delta u| < 1 \Leftrightarrow \frac{\alpha_{u_0} d}{{}^{c_0}Z + \xi_0\rho} - 1 < 0. \tag{14}$$

${}^{c_0}Z$ being non-negative, to be visible by the camera, and the constant intrinsic parameter α_{u_0} denoting implicitly the image definition. Thus, with variables ${}^{c_0}Z$ and ${}^{c_0}Y$, the function expressed as the left part of the above inequality is a quadratic curve in the plane of equation ${}^{c_0}X - d/2 = 0$. That curve (Fig. 4, plotted in blue in plane $\mathcal{P}_M$) is the spatial border beyond which the approximation of unique viewpoint shared by the four mirrors holds. The farthest visible 3D point of that curve corresponds to ${}^{c_0}Z = 0$, leading to:

$$(\alpha_{u_0} d/\xi_0\rho) - 1 < 0 \Leftrightarrow \alpha_{u_0} d/\xi_0 < \rho. \tag{15}$$

Dually, for a given minimum distance $\rho = \rho_{min}$ between the camera and the scene 3D point, we can deduce the image definition boundary below which the unique viewpoint approximation holds from (15), *i.e.*:

$$\alpha_{u_0}\, d/\xi_0 < \rho_{min} \Leftrightarrow \alpha_{u_0} < \rho_{min}\, \xi_0/d. \tag{16}$$

However, as mentioned above, α_{u_0} only defines implicitly the image definition, since it includes also optical parameters of the camera. Thus, to deduce the maximum allowed definition of $N_{0_{sup}} \times N_{0_{sup}}$ pixels to satisfy (16), we consider the scale factor $\alpha_{u_{max}}$ associated to the maximum definition $N_{max} \times N_{max}$ pixels of a panoramic region in the image of the used actual camera as the upper boundary for α_{u_0}. Then, for a given ρ_{min} (16), and posing $\alpha_{u_0 sup} = \rho_{min}\frac{\xi_1}{d}$, we define the side definition $N_{0_{sup}}$ as:

$$N_{0_{sup}} = min\left(N_{max}\, \alpha_{u_0 sup}/\alpha_{u_{max}}, N_{max}\right). \tag{17}$$

The constraints provided by (15) and (16) are defined under several simplifying hypotheses about the actual camera. However, since they are obtained considering the worst case of a 3D point belonging to the mirrors plane, those equations are ensured valid for scene points out of that plane.

4 HDROmni prototype evaluation

This section shows and evaluates the prototype of HDR panoramic camera made following the design method of this paper. Even if the method is general with respect to the number of mirrors, we consider the four paraboloidal mirrors case ($M = 4$), leading to select $F = 3$ ND filters.

4.1 Image DR Evaluation Metrics

In order to evaluate the DR quality of a given image I, we exploit the weights $w(I, \mathbf{u})$ (Sec. 2.3), for pixel coordinates $\mathbf{u}$ of I. Such a weight tends to 1 when $I(\mathbf{u})$ is said to represent naturally the scene, or 0, inversely. Then, we compute the mean $\bar{w}(I) \in [0, 1]$ of all $w(I, \mathbf{u})$. The highest $w(I, \mathbf{u})$, the best DR quality.

To take into account practical issues of the camera, its mast reflection on mirrors is ignored in computing $\bar{w}(I)$.

4.2 The actual new camera

Following our optical design method (Sec. 2), we first estimate experimentally the CRF of the considered IDS UI3250-LE camera. It exploits an e2v EV76C570 CMOS sensor (DR of 50 dB, with the used global shutter mode). The choice of the latter camera was mainly made for the evaluation of the optical extension of DR. Indeed, that camera is capable of 10 s exposure times, necessary to estimate the actual CRF of the camera when using filters of high density, for its comparison with expected CRF.

The camera is set in a dark room with a 350 lumen LED light source. Then, pointing the camera to the surface of a cardboard box, assumed Lambertian, illuminated by the light source, images are recorded with exposure times from 1 ms (T_{0min}) to 30 ms (step: 1 ms). As exposure varies proportionally with respect to exposure time or irradiance (1), varying the exposure time as described above simulates proportional variations of irradiance, thus radiance (since the considered object is Lambertian) that our cheap, but easy to reproduce, experimental setup does not allow to quantitatively vary. Since the proportional factor is unknown, the below CRF appear as functions of normalized irradiance.

A CRF is got using the brightness of a unique pixel in the central area of the box for various exposure times. However, due to the experimental setup, the brightness as a function of exposure time is noisy. Thus, brightnesses are slightly smoothed by local regression (sliding window of 30% of the total data) using weighted least squares of a second order degree polynomial model (Matlab Curve Fitting Toolbox). The brightness saturates for $T_{0max} = 21$ ms. The normalized irradiance is then given by $E_{0n} = T_{0n}/T_{0max}$, with $T_{0n} \geq T_{0min}$, $E_{0max} = 1$ being its maximum.

As the CRF of the used camera is not linear (Fig. 5a), we apply the piecewise linear approximation of the generalized exposure ratio modeling (Sec. 2.2) toward the ND filter densities computation. Thus, such CRF approximation leads to $E_{0dis} \approx 2/3$, so that $E_{md} = \frac{1}{2/3} = 3/2$.

Then, to stick to the state-of-the-art [24], we set $\lambda = 1/2$ (Eq. (5)) so that the generalized exposure ratio becomes $R_{f,f-1} = \frac{1/2}{3/2} = 1/3$. Finally, following (6), (7), and (8), we get the ideal optical densities of the three ND filters as $D_1 = 0.47$, $D_2 = 0.96$ and $D_3 = 1.90$.

The closest products available on the market are three absorptive, round-shaped, ND filters of optical densities $\widehat{D}_1 = 0.6$, $\widehat{D}_2 = 0.9$ and $\widehat{D}_3 = 2.0$, which respectively attenuates the exposure by $k_1 = 1/4$, $k_2 = 1/8$ and $k_3 = 1/100$ factors (compared to 1/3, 1/9 and 1/81 for the ideal optical densities). Setting these ND filters between the camera lens and three of the four paraboloidal mirrors (Fig. 1a acquiring images as shown in Fig. 3, left), the HDROmni prototype increases 100 times the DR of the used camera, thus from 50 dB to 90 dB.

Then, to evaluate the actual camera dynamic range with respect to expectations brought by the use of ND filters, the CRF of every panoramic region of the acquired HDROmni image is measured and plotted (Fig. 5b) with respect to normalized irradiance E_{0max}. Due to the most dense ND filter, exposure times of HDROmni ranges from 1 ms to 2500 ms (only to plot the CRF).

Figure 5b shows that the maximum normalized perceptible irradiance is 120 times $E_{0_{max}}$, instead of the expected 100 times. The range and quantization of the perceptible irradiance is improved with respect to a unique image of a single LDR camera. However, obviously, with 4 different exposures only to cover a normalized range of irradiance of $[0, 120]$, the quantization of normalized irradiance $E \in [8, 120]$ is much lower than for $E \in [0, 8[$ (same observation for the range $[0, 100]$). Then, the following subsection evaluates the HDR image quality, after the fusion of LDR panoramic regions of the acquired HDROmni image.

4.3 Dynamic range image quality evaluation

As defined in Section 3 some rules about image resolution and scene depth ρ_j must be followed to permit the HDR fusion. The four mirror centers of the HDROmni prototype make a square of $d = 0.03$ m side. The maximum resolution of each panoramic LDR image in the used camera is $N_{max} = 430$. The HDROmni prototype was calibrated [6] with checkerboards. Considering the means $\overline{\xi} = 0.89$ and $\overline{\alpha} = 157$ of intrinsic parameters ξ_j and, respectively and together α_{u_j}, α_{v_j}, for $j \in [0, 3] \subset \mathbb{N}$, we

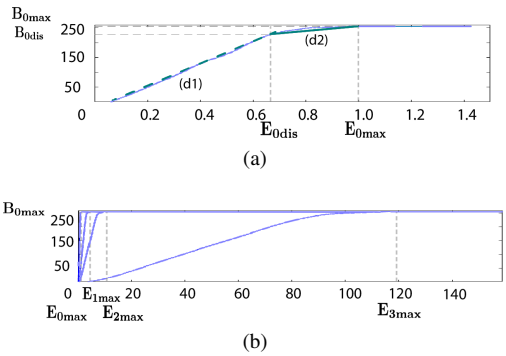


Figure 5: (a) Measured CRF of UI3250-LE, piecewise linearly approximated ((d_1) and (d_2)), (b) measured CRF for the 4 panoramic regions of the acquired HDROmni image.

found, thanks to (15), that beyond depth $\rho_j = 5.30\ m$ the disparity falls within a pixel of the maximum 430×430 pixels resolution per LDR image.

Then, the dynamic range of HDROmni is evaluated outdoor in 9 challenging scenes of clear, bright sky with a few clouds, presented in the supplementary video. We kept the maximum resolution $N_{max} = 430$ as working outdoor, where the distance constraint is always met. Hence, (16) leads to $N_{0_{sup}} = N_{max} = 430$ pixels. To gain space, only 3 of them are presented below. These scenes are mainly made of white flat and curvy concrete walls, windows and both bright sun highlight and dark shadowed areas (Fig. 6). The second scene (Fig. 6b) also features colorful small walls. The other scenes (rest of Fig. 6) are parking lots.

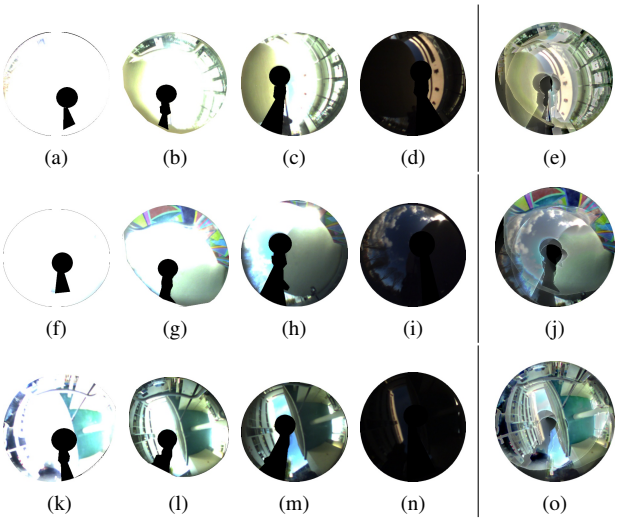
HDROmni images are acquired in each scene. Their LDR panoramic regions (first 4 columns of Fig. 7) are fused into an HDR image (last column of Fig. 7).

We consider the DR image quality criterion of Section 4.1 to evaluate quantitatively the perceptible DR extension provided by the HDROmni prototype with respect to common omnidirectional camera. Writing $\bar{w}(I_{HDR})$ the DR image quality of the HDR fusion and $\bar{w}(I_{LDR_j})$ the ones of the four LDR panoramic regions, their ratio $\bar{w}(I_{HDR})/\bar{w}(I_{LDR_j}) \in \mathbb{R}_+$ is greater than 1 if the DR image quality of I_{HDR} is better than the other one.

I_{HDR} always leads to a higher DR quality than any I_{LDR_j}. Indeed, the mean of the ratios between the 9 HDR images and the best LDR ones in each dataset is 2.0. The computed ratios for three challenging scenes of Figure 6a, 6b, 6c are displayed in Table 1. The 6 others are shown in the accompanying video. The DR increase varies from case to case. Indeed, the highest increase ratio is 17542 in scene 2 in Table 1 (I_{HDR} of Fig. 7j with respect to I_{LDR_0} of Fig. 7f), where it is obvious that only decreasing a bit the exposure time would improve the image content. It is confirmed by the ratio $\bar{w}(I_{HDR})/\bar{w}(I_{LDR_1}) = 3.87$ in scene 2 (Tab. 1). Visual comparison between Figures 7g and 7j qualitatively confirms the quantitative comparison. Despite the lower exposure of I_{LDR_1} with respect to I_{LDR_0}, a large part of the image (sky and top parts of a building) is still over-exposed, contrary to I_{HDR}. It is also interesting to note that decreasing the exposure, while solving the overexposure of the sky, underexposes the rest of the field-of-view and decreases the DR of I_{LDR_2} and I_{LDR_3} compared to I_{LDR_1} (Tab. 1, Scene 2). The latter

(a) (b) (c) (d) (e)

Figure 6: Scenes for evaluating HDROmni ((c), (d): same scene, different weather).



(a) (b) (c) (d) (e)
(f) (g) (h) (i) (j)
(k) (l) (m) (n) (o)

Figure 7: Panoramic LDR images of an acquired HDROmni image (shown split in the 4 first columns) and the fused HDR image (right). A scene per row.

observation is nothing but another illustration of the LDR of the camera that the fusion of the four I_{LDR_j} in I_{HDR} solves.

The analysis of Scene 1 and Scene 3 leads to similar conclusions than for Scene 2, obviously with shifts in the LDR image number the closest to the I_{HDR} of the scene, in term of DR, since scenes as well as illumination are different. Furthermore, the default 50 ms exposure time is set for the first and second scenes, whereas the exposure time in the third scene was manually decreased to 43 ms to allow one of the four I_{LDR_j} to get the image of widest DR the camera could acquire (Fig. 7l). Despite that change, the DR quality of I_{HDR} is still twice the one of I_{LDR_1} in Scene 3.

4.4 Evaluation on feature matching

In three scenes (Fig. 6c-6e), 30 images were captured while the robot moved straight for 2 m (datasets 1, 2, 3). SURF features (OpenCV 4.1.0 implementation, with default parameters) were detected in every acquired LDR images and the resulting HDR image. Then, features were matched between consecutive images in each sequence. False matches were manually counted (2.74 for HDR images and 1.17, 1.01, 1.07 and 0.17 for LDR ones, on

	$j=0$	$j=1$	$j=2$	$j=3$
Scene 1	1448.47	5.66	2.47	6.96
Scene 2	17542.76	3.87	4.58	12.35
Scene 3	4.71	2.09	3.92	25.80

Table 1: I_{HDR} dynamic range quality increase relatively to the four I_{LDR_j}.

average). As results, HDROmni allows detecting, at worst, 36% more features than a LDR camera. Furthermore, there are, at least, 26% more matches in successive HDR images than in LDR ones (on average, 433 matches with HDR, but 221, 306, 257 and 19 for LDR). SURF detection and matching results are shown in the supplementary video.

5 HDR panoramic visual servoing

The real-time property of HDROmni allows its use in robot control. We consider the panoramic photometric visual servoing (pPVS) of a Pioneer 3AT mobile robot [7]. Shortly, photometric visual servoing is the robot control law that automatically drives the robot from an initial pose to the desired pose, only using pixel brightnesses of images: the difference between the desired image, acquired at the desired pose, and the current image is minimized over time. [7] shows that pPVS's convergence domain (up to 1.2 m) is much wider than using a conventional camera [8] (up to 0.4 m, in ideal conditions) while both are very precise at convergence. In our experiments, we observe a decrease of precision outdoors, particularly in scenes of high dynamic range. Inversely, HDROmni leads to constantly precise performance of pPVS, even in scenes of challenging illumination, only thanks to HDROmni, since the control law is not changed.

To distinguish between the original pPVS and the one using HDROmni, we name the former LDR-pPVS and the latter HDR-pPVS. We also consider a third pPVS, namely LDR-4-pPVS, that considers the HDROmni camera without the ND-filters. In this case, since filters are not used, the 4 LDR images have the same exposure and their *exposure fusion* is approximately their mean. Comparing LDR-pPVS, LDR-4-PVS and HDR-pPVS makes clear the impact of using several mirrors, on the one hand, and ND-filters, on the other hand, with respect to the state-of-the-art pPVS [7].

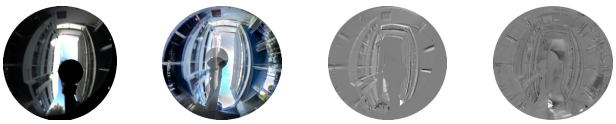
A total of 32 experiments done in the same three scenes as in Section 4.4 is reported for quantitative evaluation. 8 of them are shown in the supplementary video. In each scene, a single desired pose is considered for four initial poses at 500, 1000, 1500 and 2000 mm backward the desired pose. Then, LDR-pPVS, LDR-4-PVS and HDR-pPVS are run from these initial poses. The distance between the final pose of the robot, *i.e.* where the control law no longer makes the robot moving, and the desired pose is reported in Table 2.

Experiments of LDR-pPVS and LDR-4-pPVS in low DR scenes show that the fusion of four mirrors prevents divergence of pPVS (see results for initial errors of 1500 and 2000 mm in the first low DR scene parking lot 1, Tab. 2). In parking lot 1, LDR-pPVS is in line with [7], showing a possible border of the convergence domain of pPVS between 1 and 1.5 m. Interestingly, our implementation of LDR-pPVS could converge even with an initial error of 2 m in parking lot 2, showing more favorable for pPVS.

Still in low DR scenes, HDR-pPVS always increases the positioning precision (see mean final errors of Tab. 2 for low DR scenes), particularly in the parking lot 2 where, even if the sky was full of clouds, the DR of that scene was higher than at parking lot 1. Thus, as using four mirrors without ND-filters does not contribute as much as when combined with ND-filters to form the HDROmni camera, LDR-4-pPVS is not considered in further experiments.

Experiments of LDR-pPVS and HDR-pPVS in the scene of high DR show HDR-pPVS outperforms LDR-pPVS by a factor of almost 10, regarding the mean final error. Interestingly, the final error of LDR-pPVS is almost constant, for the four initial poses, highlighting it converges to a local minimum. Indeed, in LDR images considered by LDR-pPVS, some parts of the field-of-view are under- or over-exposed (Fig. 8a). So the inherent lack of details prevents the control law to reach the true desired pose, contrary to HDR-pPVS for which HDR images are mostly well-exposed (Fig. 8b) leading, at convergence, to an almost null difference between current and desired images (Fig. 8d).

Furthermore, for the three scenes, the mean final errors are almost the same for HDR-pPVS (31.3 to 36.5 mm), unlike the ones computed for LDR-4-pPVS (36.3 and 87.5 mm) and LDR-pPVS (64.5 to 585.0 mm).



(a) (b) (c) (d)

Figure 8: Parking lot 1 of high scene DR. Desired images for pPVS: (a) LDR (b) HDR. Differences between the image at convergence and the desired image for (c) LDR- and (d) HDR-pPVS.

	Scene		**Initial error**				**Mean**
	l.	DR	500	1000	1500	2000	error
L.	1	low	65	75	600	1600	585.0
L4	1	low	5	40	60	40	36.3
H.	1	low	30	45	35	25	**33.8**
L.	2	low	10	8	100	140	64.5
L4	2	low	10	20	140	180	87.5
H.	2	low	15	20	20	70	**31.3**
L.	1	high	330	310	330	353	323.3
H.	1	high	11	22	51	62	**36.5**

Table 2: Positioning errors for 3 different pPVS: LDR (L.), LDR-4 (L4) and HDR (H., ours) (unit: mm) in various outdoor scenes. Bold values indicate the lowest mean final errors.

6 Conclusion and future works

We have introduced a design method for extending optically the dynamic range of panoramic vision with a single camera in real-time. The first HDROmni prototype extends more than 100 times the dynamic range of the used machine vision camera. In scenes of challenging illumination, it doubles, on average, the dynamic range image quality. Such contribution allows detecting and matching more features, for future benefit to visual odometry, for instance. It also significantly improves the positioning precision of direct visual servoing in scenes of challenging illumination.
Future works will improve the prototype and combine it with neural networks-based HDR rendering approaches for better visual appearance of HDR images for the human eye.

7 ACKNOWLEDGMENT

This work is carried out as part of the Interreg VA FCE ADAPT project "Assistive Devices for empowering disAbled People through robotic Technologies" (adapt-project.com). The Interreg FCE Programme is a European Territorial Cooperation programme that aims to fund high quality cooperation projects in the Channel border region between France and England. The Programme is funded by the European Regional Development Fund (ERDF).

References

[1] Manoj Aggarwal and Narendra Ahuja. Split aperture imaging for high dynamic range. *Int. Journal of Computer Vision*, 58(1):7–17, 2004.

[2] Joao Pedro Barreto and Helder Araujo. Issues on the geometry of central catadioptric image formation. In *IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, 2001.

[3] Basler. EMVA data overview, t-explorer.co.jp/pdf/170331.pdf, 2017.

[4] Jennifer Bonnard, Céline Loscos, Gilles Valette, Jean-Michel Nourrit, and Laurent Lucas. High-dynamic range video acquisition with a multiview camera. In *Optics, Photonics, and Digital Tech. for Multimedia App. II*, volume 8436, 2012.

[5] T.E. Boult, X. Gao, R. Micheals, and M. Eckmann. Omni-directional visual surveillance. *Image and Vision Computing*, 22(7):515 – 534, 2004.

[6] Guillaume Caron and Damien Eynard. Multiple camera types simultaneous stereo calibration. In *IEEE Int. Conf. on Robotics and Automation*, pages 2933–2938, 2011.

[7] Guillaume Caron, Eric Marchand, and El Mouaddib. Photometric visual servoing for omnidirectional cameras. *Autonomous Robots*, 35(2-3):177–193, October 2013.

[8] C. Collewet and E. Marchand. Modeling complex luminance variations for target tracking. In *IEEE Conf. on Computer Vision and Pattern Recognition*, Anchorage, Alaska, June 2008.

[9] Diana-Margarita Córdova-Esparza, Juan Terven, Julio-Alejandro Romero-González, and Alfonso Ramírez-Pedraza. Three-dimensional reconstruction of indoor and outdoor environments using a stereo catadioptric system. *Applied Sciences*, 10(24):8851, 2020.

[10] Arnaud Darmont. *High Dynamic Range Imaging: Sensors and Architectures, 2nd Ed.* SPIE, 2019.

[11] PE Debevec. Recovering high dynamic range radiance maps from photographs. *SIGGRAPH'97*, 1997.

[12] Flir. Sensor review: mono cameras (EMVA standard compliant), https://www.flir.com/globalassets/iis/guidebooks/2019-machine-vision-emva1288-sensor-review.pdf, 2019.

[13] R. Hill. A lens for whole sky photographs. *Quarterly Journal of the Royal Meteorological Society*, 50(211):227–235, July 1924.

[14] Fan Jinsong, Fan Yanbing, and Pei Jigang. HDR spherical panoramic image technology and its applications in ancient building heritage protection. In *IEEE Int. Conf. on Computer-Aided Industrial Design & Conceptual Design*, pages 1549–1553, 2009.

[15] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep hdr video from sequences with alternating exposures. In *Computer Graphics Forum*, volume 38, pages 193–205. Wiley Online Library, 2019.

[16] Soo Ye Kim, Jihyong Oh, and Munchurl Kim. Deep SR-ITM: Joint learning of super-resolution and inverse tone-mapping for 4K UHD HDR applications. In *IEEE/CVF Int. Conf. on Computer Vision*, October 2019.

[17] Siyeong Lee, Gwon Hwan An, and Suk-Ju Kang. Deep chain hdri: Reconstructing a high dynamic range image from a single low dynamic range image. *IEEE Access*, 6:49913–49924, 2018.

[18] Yan Liu, Bingxue Lv, Wei Huang, Baohua Jin, and Canlin Li. Anti-shake HDR imaging using raw image data. *Information*, 11(4):213, 2020.

[19] Yu-Lun Liu, Wei-Sheng Lai, Yu-Sheng Chen, Yi-Lung Kao, Ming-Hsuan Yang, Yung-Yu Chuang, and Jia-Bin Huang. Single-Image HDR Reconstruction by Learning to Reverse the Camera Pipeline. In *IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, pages 1651–1660, 2020.

[20] Rafał Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. High dynamic range imaging. In *Wiley Encyclopedia of Electrical and Electronics Engineering*, pages 1–42. Wiley, 2015.

[21] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion: A simple and practical alternative to high dynamic range photography. In *Computer graphics forum*, volume 28, pages 161–171, 2009.

[22] Tomoo Mitsunaga and Shree K Nayar. Radiometric self calibration. In *IEEE Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 374–380, 1999.

[23] E. Mouaddib, R. Sagawa, T. Echigo, and Y. Yagi. Stereo Vision with a Single Camera and Multiple Mirrors. In *IEEE Int. Conf. on Robotics and Automation*, pages 812–817, Barcelona, Spain, 2005.

[24] Shree K Nayar and Tomoo Mitsunaga. High dynamic range imaging: Spatially varying pixel exposures. In *IEEE Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 472–479, 2000.

[25] Yeonju Oh, Ramviyas Parasuraman, Tim McGraw, and Byung-Cheol Min. 360 VR based robot teleoperation interface for virtual tour. In *Proc. of the 1st Int. Workshop on Virtual, Augmented, and Mixed Reality for HRI*, 2018.

[26] Yuichi Taguchi, Amit Agrawal, Ashok Veeraraghavan, Srikumar Ramalingam, and Ramesh Raskar. Axial-cones: Modeling spherical catadioptric cameras for wide-angle light field rendering. *ACM Trans. Graph.*, 29(6):172, 2010.

[27] Y. Yagi, S. Kawato, and S. Tsuji. Real-time omnidirectional image sensor (COPIS) for vision-guided navigation. *IEEE Trans. on Robotics and Automation*, 10(1):11–22, Feb 1994.

[28] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar. Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum. *IEEE Trans. on Image Processing*, 19(9):2241–2253, Sep. 2010.

[29] Jia-Li Yin, Bo-Hao Chen, Yan-Tsung Peng, and Chung-Chi Tsai. Deep prior guided network for high-quality image fusion. In *IEEE Int. Conf. on Multimedia and Expo*, pages 1–6, 2020.

[30] Kai Zeng, Kede Ma, Rania Hassen, and Zhou Wang. Perceptual evaluation of multi-exposure image fusion algorithms. In *Int. Workshop on Quality of Multimedia Experience*, pages 7–12, 2014.

[31] Jie Zhang, Jonathan P Newman, Xiao Wang, Chetan Singh Thakur, John Rattray, Ralph Etienne-Cummings, and Matthew A Wilson. A closed-loop, all-electronic pixel-wise adaptive imaging system for high dynamic range videography. *IEEE Trans. on Circuits and Systems*, 67(6):1803–1814, 2020.

[32] Z. Zhang, H. Rebecq, C. Forster, and D. Scaramuzza. Benefit of large field-of-view cameras for visual odometry. In *IEEE Int. Conf. on Robotics and Automation*, pages 801–808, May 2016.