Journal of Electronic Imaging

SPIEDigitalLibrary.org/jei

Depth consistency and vertical disparities in stereoscopic panoramas

Luis E. Gurrieri
Eric Dubois

Luis E. Gurrieri and Eric Dubois "Depth consistency and vertical disparities in stereoscopic panoramas", J. Electron. Imaging. 23(1), 011004 (Jan 29, 2014).

http://dx.doi.org/10.1117/1.JEI.23.1.011004

Journal of Electronic Imaging 23(1), 011004 (Jan–Mar 2014)

Depth consistency and vertical disparities in stereoscopic panoramas

Luis E. Gurrieri* **and Eric Dubois**
University of Ottawa, School of Electrical Engineering and Computer Science, Ottawa, Ontario K1N 6N5, Canada

Abstract. In recent years, the problem of acquiring omnidirectional stereoscopic imagery of dynamic scenes has gained commercial interest, and consequently, new techniques have been proposed to address this problem. The goal of many of these new panoramic methods is to provide practical solutions for acquiring real-time omnidirectional stereoscopic imagery for human viewing. However, there are problems related to mosaicking partially overlapped stereoscopic snapshots of the scene that need to be addressed. Among these issues are the conditions to provide a consistent depth illusion over the whole scene and the appearance of undesired vertical disparities. We develop an acquisition model capable of describing a variety of omnistereoscopic imaging systems and suitable to study the design constraints of these systems. Based on this acquisition model, we compare different acquisition approaches based on mosaicking partial stereoscopic views of the scene in terms of their depth continuity constraints and the appearance of vertical disparities. This work complements and extends our previous work in omnistereoscopic imaging systems by proposing a mathematical framework to contrast different acquisition strategies to create stereoscopic panoramas using a small number of stereoscopic images. *©2014 SPIE and IS&T* [DOI: 10.1117/1.JEI.23.1.011004]

Keywords: stereoscopic panoramas; omnistereo; omnistereoscopic acquisition; omnistereoscopic cameras; vertical disparities; panoramic depth perception.

Paper 13494SS received Aug. 31, 2013; revised manuscript received Dec. 20, 2013; accepted for publication Jan. 7, 2014; published online Jan. 29, 2014.

1 Introduction

The problem of acquiring stereoscopic panoramas of dynamic scenes has gained relevance in recent years and, consequently, new acquisition methods have been proposed.[1] The goal of many of these novel panoramic methods is to acquire stereoscopic panoramas of real-world scenes and to render stereoscopic views suitable for human viewing.[2–4] In particular, methods based on the acquisition of partially overlapped stereoscopic snapshots of the scene are the most attractive for real-time omnistereoscopic capture.[5,6] However, there is a need to rigorously model these acquisition techniques in order to provide useful design constraints for the corresponding omnidirectional stereoscopic systems.

This paper is about the limitations of omnistereoscopic systems based on acquiring partially overlapped stereoscopic views of the scene from two distinct and coplanar viewpoints with horizontal parallax. In particular, we study the mosaicking problem that has become relevant due to recent developments in stereoscopic panoramic video.[6,7] However, this paper does not address the problems related specifically to panoramic video.

An important problem often ignored in omnistereoscopic acquisition techniques is the continuity in the illusion of depth perceived over all gazing directions. This problem is relevant for omnistereoscopic systems based on acquiring partially overlapping stereoscopic snapshots of the scene to be mosaicked into a complete stereoscopic panorama.[4,5,7,8] One important parameter used to contrast different acquisition techniques is the minimum distance to the scene to provide a continuous illusion of depth in any gazing direction. Another important problem is to characterize the vertical disparities that cause ghosting and visual discomfort at the stitching boundaries between mosaics. In our simulations, we studied the effect of the field of view (FOV) of the lenses, and the pixel size and dimension of the sensor in the design of the system.

In order to study these parameters, we propose a general acquisition model to describe a variety of multiple camera systems and acquisition techniques.[1] Our model is based on a pair of pin-hole cameras with horizontal parallax. By changing the spatial location of the respective projection centers and the panning direction of the stereoscopic pair, this model describes a large variety of omnistereoscopic imaging systems with a horizontal baseline.

First, we detail the general acquisition model. Then, we derive from this generic model four acquisition configurations that describe a variety of omnistereoscopic cameras suitable to produce horizontal stereo for human viewing.[1] Finally, we apply this acquisition model to obtain expressions for the horizontal and vertical disparities observed when mosaicking stereoscopic snapshots of the scene. We obtain the parameters of interest for each configuration using a ray tracing approach. From these simulations, we extract conclusions that can be used in the design of omnistereoscopic cameras for the acquisition of dynamic scenes.

One of the contribution of this paper is to provide a tractable method for analyzing multiple camera configurations intended for omnistereoscopic imaging. Furthermore, we provide methods to study the acquisition constraints necessary to attain a continuous depth perception in all gazing directions in azimuth. Another relevant contribution is to

*Address all correspondence to: Luis E. Gurrieri, E-mail: Luis.Gurrieri@uOttawa.ca

provide a mathematical model for the vertical disparities that would affect the mosaicking process in each configuration. This work complements and extends our previous work in stereoscopic panoramas acquisition[2–4] by proposing a mathematical framework to contrast different omnistereoscopic acquisition strategies.

2 Omnistereoscopic Acquisition Model

The general acquisition model is composed of two pin-hole cameras separated by a baseline distance b with respect to a global reference of coordinates in three-dimensional (3-D) space as illustrated in Fig. 1. This model is used to derive four camera configurations, which can be distinguished by the relative location of the stereoscopic pair of cameras with respect to the reference center. In this paper, we refer to these four spatial variations of the acquisition model as configurations, and we assign them numbers from one to four.

Although this model consists of one pair of cameras, it can model complex multiple camera configurations as well as a single stereoscopic camera rig rotated at different azimuth angles. In the paper,[1] we review a large variety of acquisition cameras and methods to produce omnistereoscopic imagery suitable for human viewing that each can be described by one of these four variations of the acquisition model.

The location and orientation of this stereoscopic camera pair in the 3-D space is restricted by the need to capture two snapshots of the same scene from two viewpoints with horizontal parallax. One constraint is that all the possible locations for this pair of pin-hole cameras are restricted to the horizontal **XZ**-plane, which is used as the reference horizontal plane. Another constraint is that the optical axes of the stereoscopic pair of pin-hole cameras are parallel and lie on the **XZ**-plane. A consequence of these constraints is that the orientation (panning angle) of each virtual stereoscopic rig is described by a pitch rotation around the **Y** axis. The reference point **O** is used to describe the panning direction of the stereoscopic rig. Hence, the locations of the projection centers of each camera can be described by a Euclidean translation in the **XZ**-plane.

2.1 *Acquisition Model: Configurations*

The first configuration of the acquisition model we introduce is the central stereoscopic rig (configuration 1), which is illustrated in Fig. 2(a). This configuration models the sampling of the scene by means of a rotating stereoscopic camera, which captures partial snapshots at regular angular intervals in azimuth. This method has been the first technique

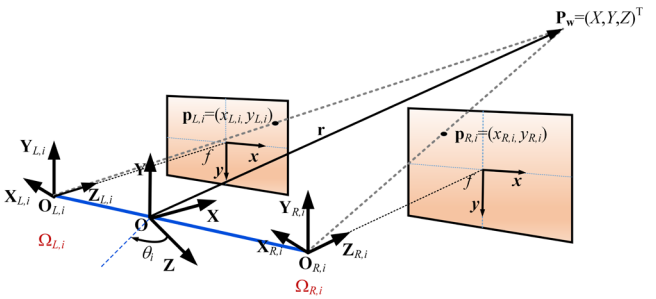

Fig. 1 The global reference frame versus the camera frames for the stereoscopic camera composed by cameras $\Omega_{L,i}$ and $\Omega_{R,i}$, which is oriented at θ_i in azimuth.

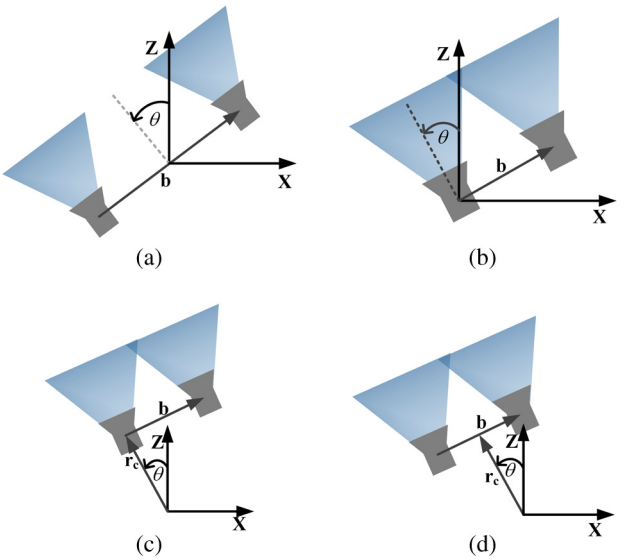

Fig. 2 Variations of the acquisition model: (a) central stereoscopic rig (configuration 1), (b) lateral stereoscopic rig (configuration 2), (c) lateral-radial stereoscopic rig (configuration 3), and (d) off-centered stereoscopic rig (configuration 4).

to create omnistereoscopic images with horizontal parallax. This camera configuration consists of two cameras with co-planar projection centers, which are separated by a baseline b. The stereoscopic camera rotation is around the Y axis exclusively (pitch rotation). This configuration describes acquisition methods that have been widely used over the last decade, whether using planar[9,10] or line sensors[11,12] for sequential acquisition. This configuration also models a widely used technique based on rotating an off-centered camera at regular angles θ_i, where $i \in \{0, \ldots, N-1\}$ and N is usually large, producing a set of overlapped stereoscopic images of the scene.[13,14] Although simple in its conception, configuration 1 cannot to be used in a parallel acquisition configuration due to self-occlusion between cameras; hence it can only represent sequential omnistereoscopic acquisition.

The lateral stereoscopic rig (configuration 2) is shown in Fig. 2(b). This configuration models the sequential acquisition of partial images of the scene by rotating the stereoscopic camera at regular intervals θ_i, where $i \in \{0, \ldots, N-1\}$. The main difference with configuration 1 is that the pitch rotation is defined around the nodal point of one of the cameras, i.e., making the rotation axis to coincide with the projection center of the left or right camera. In this approach, a singular viewpoint panorama is produced by mosaicking the images acquired by the central camera, while a second image with horizontal parallax for each θ_i is acquired by the lateral camera.[15,16] The lateral camera (stereoscopic counterpart) describes a circle of radius equal to the stereo baseline b. Similar to configuration 1, this configuration cannot be used for the simultaneous acquisition of the whole scene due to self-occlusion of the central camera.

The lateral-radial stereoscopic rig (configuration 3) is shown in Fig. 2(c). This configuration models a stereoscopic camera rotated off-center, where one camera is radially displaced from the rotation axis in **O** by a distance $\|\mathbf{r_c}\|$, and the second camera is laterally displaced b perpendicularly with respect to the direction of $\|\mathbf{r_c}\|$. This arrangement enables the capture of a second snapshot of the scene with horizontal parallax for each sampling angle θ_i.[4] This configuration

can be derived from configuration 2 by radially displacing the stereoscopic camera a distance $\|\mathbf{r_c}\|$ away from the central camera nodal point. This configuration can be used to model a multiple-sensor arrangement since the self-occlusion of the central camera can be avoided.

The off-centered stereoscopic rig (configuration 4) is a stereoscopic camera located at a radial distance $\|\mathbf{r_c}\|$ from the geometrical center $\mathbf{O}$ as depicted in Fig. 2(d). This configuration can be derived from configuration 1 by locating the pitch axis $\mathbf{O}$ a distance $\|\mathbf{r_c}\|$ behind the stereoscopic camera midpoint. This configuration models the case where multiple stereoscopic rigs, radially located with respect to the pitch center $\mathbf{O}$, are used to acquire partially overlapped snapshots of the whole scene.[2,17–19] In one approach, the successive images acquired by left and right cameras can be mosaicked to create left and right eye panoramas: $(\mathbf{I_L}, \mathbf{I_R})$.[20]

2.2 *Practical Acquisition Approach*

Acquiring the scene column-wise using line cameras[21] or through extracting narrow image columns by back-projecting two distinct viewpoints[13,22] can produce stereoscopic panoramas for human viewing. However, their main downfall is their sequential characteristic, which limits them to static scenes. The stereoscopic images rendered by directly mosaicking multiple columns with horizontal parallax are correct only in a limited region of interest,[23] which is located at the center of the image, no matter the gazing direction. This is acceptable since the peripheral vision is not used by the mechanisms of stereo fusion, but adaptation to different display technologies is necessary.

An attractive alternative is to capture a limited number of stereoscopic snapshots of the scene and mosaic them,[4,6,8,24] e.g., five to eight stereo images will be enough to cover the whole scene in azimuth. Partially overlapped snapshots can be acquired sequentially or simultaneously. The latter opens the possibility to acquire omnistereoscopic images and videos for dynamic scenes, bypassing the limitations of line sequential techniques. Mosaicking can produce a continuous and consistent binocular illusion of depth in all gazing directions around the sampling point. However, to achieve this ideal illusion of depth, the camera system has to be carefully designed.

2.3 *Model*

A projective pin-hole camera is a simple yet powerful approach to model each configuration of the acquisition model presented in the previous section. The location of each projection center can be specified for a singular stereoscopic camera rotating around a common vertical axis (configurations 1 and 2) and for multiple stereoscopic pairs with a common symmetry axis (configurations 3 and 4).

In this paper, we use the subindex j ($j \in \{L, R\}$) to refer to the left (L) or right (R) cameras in a stereoscopic camera pair. The subindex i refers to one of N consecutive gazing directions in azimuth ($i \in \{0, \ldots, N-1\}$) defined as

$$\theta_i = \frac{i \cdot 360 \ \text{deg}}{N}. \tag{1}$$

It follows that $\theta_N = \theta_0$. In summary, the subindex (j, i) indicates the camera j ($j \in \{L, R\}$) from a stereoscopic pair of cameras, which is oriented at θ_i in azimuth ($i \in \{0, \ldots, N-1\}$) with respect to the global reference of coordinates.

Note that each of the four configurations presented in Fig. 2 can be modeled by a set of single stereoscopic rigs with different gazing directions or by a singular stereoscopic camera sequentially rotated at different θ_i angles. Using this abstraction, all configurations can be described as a single stereoscopic after defining the spatial location of left and right cameras and the set of all possible gazing directions for such a virtual stereoscopic camera.

2.3.1 *World and camera frame of coordinates*

The reference world frame of coordinates is referred as $\mathbf{XYZ}$ and its spatial location differs for each configuration. A point in the scene is written as $\mathbf{P_W} = (X_W, Y_W, Z_W)^T$, where X_W, Y_W, and Z_W are scalar numbers.

Each camera has its own local frame of coordinates, which we refer as $\mathbf{X}_{j,i}\mathbf{Y}_{j,i}\mathbf{Z}_{j,i}$. Similar to the world reference of coordinates, a point in the camera coordinates frame is denoted as $\mathbf{P}_{j,i} = (X_{j,i}, Y_{j,i}Z_{j,i})^T$. The global (reference) and local (cameras) coordinate frames are illustrated in Fig. 1 for configuration 1.

2.3.2 *Cameras and stereoscopic rigs*

In our acquisition model, we identify each camera with the notation $\Omega_{j,i}$, where $j \in \{L, R\}$ and $i \in \{0, \ldots, N-1\}$. A camera $\Omega_{j,i}$ has a projective center, which we refer to using its spatial location $\mathbf{O}_{j,i}$. The individual cameras are grouped in stereoscopic camera pairs $(\Omega_{L,i}, \Omega_{R,i})$, where $i \in \{0, \ldots, N-1\}$ defines its orientation angle in azimuth according to Eq. (1) as illustrated in Fig. 2 for each configuration.

The distinction between left and right does not necessarily correspond to the relative spatial location of each camera in world coordinates, but it is used here to distinguish each camera in a stereoscopic pair. In order to eliminate ambiguities, we label the cameras according to their position in the reference of coordinates $\mathbf{XYZ}$ when $\theta_0 = 0$ deg, e.g., when the baseline vector $\mathbf{b}$ is parallel to $\mathbf{X}$. Using this convention, the camera $\Omega_{L,i}$ is the one whose projection center $\mathbf{O}_{L,i}$ is centered at $X \leq 0$, and conversely, the right camera, referred as $\Omega_{R,i}$, is the one whose projection center $\mathbf{O}_{R,i}$ is centered at $X > 0$. This labeling scheme is exemplified in Fig. 1. After labeling them for θ_0, each camera retains its left or right for all θ_i in the acquisition sequence.

2.3.3 *Location and orientation of each camera*

Two planes are defined by $\mathbf{X}_{j,i}\mathbf{Z}_{j,i}$ in a stereoscopic camera with orientation θ_i. These two planes are coincident with the $\mathbf{XZ}$-plane. Hence, all projection centers $\mathbf{O}_{j,i}$ are located on the same reference plane.

The camera frame and its projection center are co-located. The $\mathbf{Y}_{j,i}$-axes are parallel to the Y axis. The optical axis of camera $\Omega_{j,i}$ is aligned with the its corresponding $\mathbf{Z}_{j,i}$ axis. Furthermore, each $\mathbf{Z}_{j,i}$ axis is perpendicular to the vector $\mathbf{b} = \mathbf{O}_{L,i} - \mathbf{O}_{R,i}$, which is defined by each stereoscopic camera. Therefore, the optical axes of each camera pair are parallel. Finally, the image plane is parallel to $\mathbf{XY}$-plane and it is located at $Z_{j,i} = f$ as shown in Fig. 1.

2.3.4 *Stereoscopic image pairs*

A stereoscopic pair of images is denoted $(\mathbf{im}_{L,i}, \mathbf{im}_{R,i})$, where $\mathbf{im}_{L,i}$ and $\mathbf{im}_{R,i}$ are images, respectively, acquired by a stereoscopic camera $(\Omega_{L,i}, \Omega_{R,i})$, whose orientation is θ_i with respect to the global reference.

The two-dimensional (2-D) reference of coordinates in each image is located at the symmetry center of each image as shown in Fig. 1. A coordinate point on each image plane is denoted $\mathbf{p}_{j,i} = (x_{j,i}, y_{j,i})^T$, where the subindex (j, i) has the meaning given in Sec. 2.3. The image window is a rectangular subregion of the image plane, whose center intersects the camera optical axis $Z_{j,i}$ and whose area is defined by its horizontal width (W_h) and its aspect ratio (a_r).

2.3.5 *Acquisition and rendering*

The omnistereoscopic strategy is based on acquiring N partially overlapped stereoscopic images. The set of stereoscopic snapshots can be acquired by rotating a single stereoscopic camera in increments of Δ_θ degrees in azimuth (configurations 1 and 2), or by acquiring all the snapshots simultaneously by using multiple stereoscopic cameras (configurations 3 and 4).

The subset of images $\mathbf{im}_{L,i}$ acquired by camera $\Omega_{L,i}$ is mosaicked to render a left-eye view of the scene $\mathbf{I_L}$. The same is done with the set of images $\mathbf{im}_{R,i}$ acquired by camera $\Omega_{R,i}$ to generate the right-eye view $\mathbf{I_R}$. The pair of panoramas $(\mathbf{I_L}, \mathbf{I_R})$ defines an omnistereoscopic image $\mathbf{I_S}$.

2.3.6 *Stitching and blending*

Each image $\mathbf{im}_{j,i}$ will be aligned and stitched with the previous $\mathbf{im}_{j,i-1}$ and the next image $\mathbf{im}_{j,i+1}$ in the sequence. The aligning and stitching is done at $\pm x_b$ from the center in the horizontal dimension as shown in Fig. 3(a).

The camera FOV (Δ) is defined as the angle of view in the main diagonal of the image window. In our acquisition model, we use the camera FOV in the horizontal dimension (Δ_a) since it is better suited to our analysis.

The region of each image defined for $x_{j,i} \in [-x_b, x_b]$ used in the mosaicking is a fraction of the image width (αW_h) as illustrated in Fig. 3. The parameter α is given by

$$\alpha = \frac{2f}{W_h} \tan\left(\frac{\Delta_\theta}{2}\right), \tag{2}$$

where Δ_θ is the sampling angle in azimuth, which is defined as

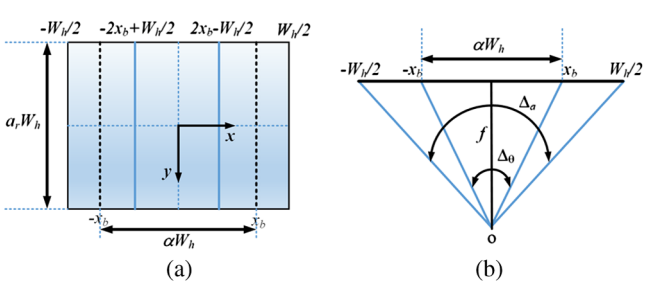

Fig. 3 The parameters for each image: (a) the frame of coordinates location on the image plane and (b) the field-of-view parameters and their relationship with f and W_h.

$$\Delta_\theta = \frac{360 \text{ deg}}{N}. \tag{3}$$

Note that $\Delta_\theta < \Delta_a$ given the partial overlapping requirement between stereoscopic samples i and $i+1$.

In each image, the stitching point is symmetrically located a distance x_b with respect to the image center in the horizontal dimension. The distance x_b is given by

$$x_b = f \tan\left(\frac{\Delta_\theta}{2}\right), \tag{4}$$

where f is the focal length of the camera, which is defined as

$$f = \frac{W_h}{2 \tan\left(\frac{\Delta_a}{2}\right)}. \tag{5}$$

The stitching positions $x_{j,i} = \pm x_b$ are in the horizontal middle of the overlapping regions for any Δ_a and Δ_θ. The geometric depiction of these magnitudes is presented in Fig. 3(b).

2.3.7 *Overlapping regions*

The overlapping regions between any two neighbor images in a set can be defined by the image regions where the same scene is simultaneously projected in both images. In our acquisition model, two overlapping regions can be defined in any image. These regions are located in

$$-\frac{W_h}{2} \le x_{j,i} \le -2x_b + \frac{W_h}{2}, \tag{6}$$

$$2x_b - \frac{W_h}{2} \le x_{j,i} \le \frac{W_h}{2}. \tag{7}$$

The overlapping regions spread $W_h - 2x_b$ from each horizontal edge of the image toward its center. However, not all this region is used for blending: only a region defined by few pixels width around x_b is used.

2.3.8 *Blending regions*

The horizontal stitching coordinate may be different from a fixed x_b when using any optimal cut algorithm.[25,26] For instance, in a recently proposed method based on using graph cuts to find the optimal stitching position,[27] the authors claim to produce consistent mosaics of stereoscopic images. Their method is based on minimizing an energy function specifically designed to account for the depth continuity followed by a warping transformation to smooth disparity transitions between mosaics. Consequently, depending on the stitching technique used, x_b provides a reference for the stitching position search. The actual stitching position can be a point $\mathbf{p}_s = (x_s, y)$, where x_s has different values for each image row.

However, we use a constant stitching coordinate x_b and a blending region of size Δ_b around the stitching coordinate. In other words, the blending region is defined by $x_{j,i} \in [-x_b - (\Delta_b/2), -x_b + (\Delta_b/2)]$ and $x_{j,i} \in [x_b - (\Delta_b/2), x_b + (\Delta_b/2)]$, for all image rows $y_{j,i}$. This is sufficient to model the design parameters for the different configurations.

2.4 *From Global to Camera Coordinates*

The transformation to represent a point $\mathbf{P_W}$ in each camera frame of coordinates can be defined as

$$\mathbf{P}_{j,i} = \mathbf{R}_i \cdot (\mathbf{P_W} - \mathbf{T}_{j,i}), \tag{8}$$

where $\mathbf{R}_i$ is the rotation matrix defined by θ_i and $\mathbf{T}_{j,i}$ defines the location of $\mathbf{O}_{j,i}$ in the global frame.

The rotation matrix is defined by

$$\mathbf{R}_i = \begin{pmatrix} \cos\theta_i & 0 & \sin\theta_i \\ 0 & 1 & 0 \\ -\sin\theta_i & 0 & \cos\theta_i \end{pmatrix}. \tag{9}$$

The generic notations to refer to the translation vectors are

$$\mathbf{T}_{L,i} = (t_{L,1}, t_{L,2}, t_{L,3})^T, \tag{10}$$

$$\mathbf{T}_{R,i} = (t_{R,1}, t_{R,2}, t_{R,3})^T, \tag{11}$$

where $t_{j,k}$, for $j \in \{L, R\}$ and $k \in \{1, 2, 3\}$, are the translation components in the global frame of coordinates for $\mathbf{O}_{j,i}$. The translation vectors for each configuration are defined in Secs. 2.4.1 to 2.4.4.

2.4.1 *Configuration 1*

$$\mathbf{T}_{L,i} = \left(-\frac{b}{2}\cos\theta_i, 0, -\frac{b}{2}\sin\theta_i\right)^T, \tag{12}$$

$$\mathbf{T}_{R,i} = \left(\frac{b}{2}\cos\theta_i, 0, \frac{b}{2}\sin\theta_i\right)^T, \tag{13}$$

where $b = \|\mathbf{b}\|$. This is shown in Fig. 4(a).

2.4.2 *Configuration 2*

$$\mathbf{T}_{L,i} = (0, 0, 0)^T, \tag{14}$$

$$\mathbf{T}_{R,i} = (b\cos\theta_i, 0, b\sin\theta_i)^T. \tag{15}$$

This is shown in Fig. 4(b).

2.4.3 *Configuration 3*

$$\mathbf{T}_{L,i} = (-\|\mathbf{r_c}\|\sin\theta_i, 0, \|\mathbf{r_c}\|\cos\theta_i)^T, \tag{16}$$

$$\mathbf{T}_{R,i} = (\|\mathbf{r_o}\|\cos\beta_i, 0, \|\mathbf{r_o}\|\sin\beta_i)^T, \tag{17}$$

where $\beta_i = \arctan(\frac{\|\mathbf{r_c}\|}{b}) + \theta_i$ and $\|\mathbf{r_o}\| = \sqrt{\|\mathbf{r_c}\|^2 + b^2}$. This is shown in Fig. 4(c).

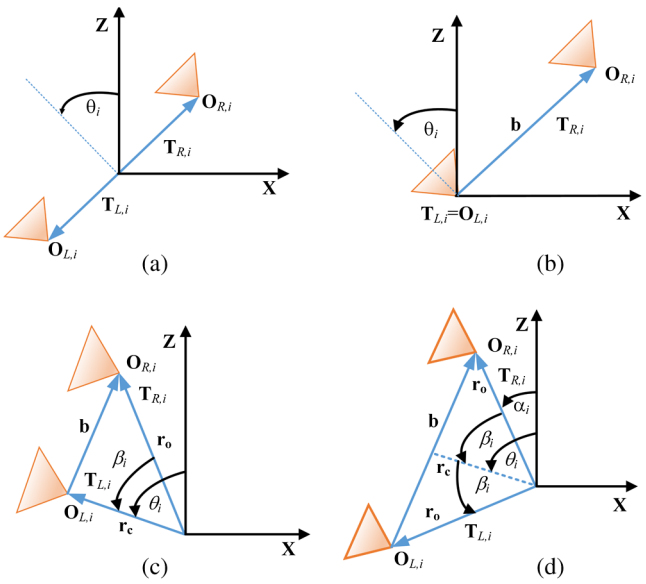

Fig. 4 The geometric relationships for each configuration.

2.4.4 *Configuration 4*

$$\mathbf{T}_{L,i} = (-\|\mathbf{r_o}\|\sin\beta_i, 0, \|\mathbf{r_o}\|\cos\beta_i)^T, \tag{18}$$

where $\beta_i = \theta_i + \arctan\left(\frac{b}{2\|\mathbf{r_c}\|}\right)$ and $\|\mathbf{r_o}\| = \sqrt{\|\mathbf{r_c}\|^2 + (\frac{b}{2})^2}$.

$$\mathbf{T}_{R,i} = (\|\mathbf{r_o}\|\sin\alpha_i, 0, \|\mathbf{r_o}\|\cos\alpha_i)^T, \tag{19}$$

where $\alpha_i = \arctan\left(\frac{b}{2\|\mathbf{r_c}\|}\right) - \theta_i$. This is shown in Fig. 4(d).

Alternatively, the transformation of coordinates in Eq. (8) can be written as

$$\mathbf{P}_{j,i} = \mathbf{R}_i \cdot \mathbf{P_W} - \mathbf{T}_j, \tag{20}$$

where

$$\mathbf{T}_j = \mathbf{R}_i \cdot \mathbf{T}_{j,i}. \tag{21}$$

Here $\mathbf{T}_j$ is the camera translation with respect to $\mathbf{O}$ of the left or right camera frame of coordinates. By this formulation, $\mathbf{T}_R = \mathbf{T}_L + (b, 0, 0)^T$ for all configurations.

2.5 *Projecting the Scene on Each Camera*

The perspective transformation to map a scene point $\mathbf{P_W}$ into the frame of coordinates of each camera $\Omega_{j,i}$ was defined by Eq. (8). Furthermore, any point $\mathbf{P_W}$ in the FOV of both cameras of a stereoscopic camera can be projected into each image as a point with coordinates $\mathbf{p}_{j,i} = (x_{j,i}, y_{j,i})^T$. The 2-D image coordinates in each image are given by the projective equations [28]

$$x_{j,i} = -f\frac{\mathbf{R}_{1,i}(\mathbf{P_w} - \mathbf{T}_{j,i})}{\mathbf{R}_{3,i}(\mathbf{P_w} - \mathbf{T}_{j,i})}, \tag{22}$$

$$y_{j,i} = -f\frac{\mathbf{R}_{2,i}(\mathbf{P_w} - \mathbf{T}_{j,i})}{\mathbf{R}_{3,i}(\mathbf{P_w} - \mathbf{T}_{j,i})}, \tag{23}$$

where $\mathbf{R}_{k,i}$ for $k \in \{1,2,3\}$ is a row vector formed by the k'th row of $\mathbf{R}_i$. Without losing generality, we assume zero bias in image center and a unit pixel size.

3 Acquisition Constraints

One important requirement arising in omnistereoscopic imaging is how to provide a continuous and consistent depth illusion in all directions. For instance, the acquisition of multiple stereoscopic snapshots of the scene for mosaicking will be correct only at the center of each mosaic, which coincides with the θ_i orientation of the stereoscopic camera at the moment of acquisition. The reproduced stereoscopic view in any intermediate gazing directions between θ_i and θ_{i+1} is subject to distortion.[6,23]

3.1 *Omnistereoscopic FOV*

The FOV of a single camera $\Omega_{j,i}$ is defined by all possible rays of light that pass through the projection center $\mathbf{O}_{i,j}$ which simultaneously intersect its image window (Sec. 2.3.4). The FOV defines the region of space in front of camera that can be acquired. This definition is valid for the geometric approach used here. However, the minimum distance to the scene can be several times f in front of the camera using real lenses. In Fig. 5, we present the different visibility scenarios using various locations in the **XZ**-plane. The point $\mathbf{P_1}$ in this example is in the FOV of camera $\Omega_{R,i}$ only.

Also in Fig. 5, the point $\mathbf{P_2}$ is in the stereoscopic FOV of the stereoscopic camera $(\Omega_{L,i}, \Omega_{R,i})$. This stereoscopic FOV is defined by the intersection of the FOVs of each camera in the stereoscopic pair. The same point $\mathbf{P_2}$ is in the FOV of camera $\Omega_{R,i+1}$, but not in the stereoscopic FOV of $(\Omega_{L,i+1}, \Omega_{R,i+1})$. Conversely, a third point $\mathbf{P_3}$, which is at the same distance from $\mathbf{O}$ as point $\mathbf{P_2}$, is located in the stereoscopic FOV of $(\Omega_{L,i+1}, \Omega_{R,i+1})$, but is not in the stereoscopic FOV of $(\Omega_{L,i}, \Omega_{R,i})$.

The distance from the camera that marks the intersection of the stereoscopic FOV of two neighbor stereoscopic cameras (i and $i+1$) defines the stereoscopic FOV of the panoramic camera. Any point in the scene at this distance from the omnistereoscopic camera can potentially be imaged in $\mathbf{I_S}$. In this example, $\mathbf{P_4}$ belongs to the intersection of two neighbor stereoscopic FOVs, which defines the overlapping region in space. A point in this overlapping region will be imaged simultaneously by cameras $(\Omega_{L,i}, \Omega_{R,i})$ and $(\Omega_{L,i+1}, \Omega_{R,i+1})$. Any point in the overlapping region can be used for stitching neighbor stereoscopic images.

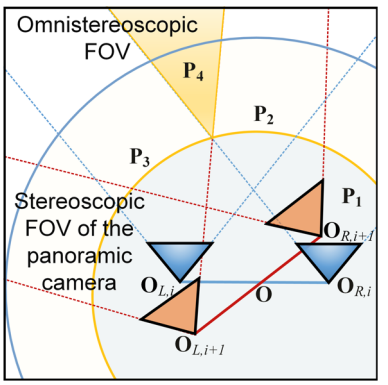

Fig. 5 The continuous depth perception for all gazing directions.

The stereoscopic FOV of the panoramic camera does not define the minimum distance between the omnistereoscopic camera and the scene. In some configurations, a point in the overlapping of stereoscopic FOVs may be registered with different horizontal disparities in neighbor stereoscopic samples. This is especially true in configurations with asymmetrical camera locations with respect to the virtual rotation axis, such as configurations 2 and 3. This affects the perceived depth illusion after mosaicking the partial images.

However, a continuous depth illusion can be maintained as long as the difference between horizontal disparities is below the human threshold for depth perception around the stitching position. The perceived depth is directly related to the horizontal disparities in each stereoscopic sample. Since the difference between the horizontal disparities of the same scene in different samples decreases with the distance between the camera and the scene, there is a minimum tolerable distance for each configuration after which a difference in the registered depth is below the human perceptual threshold.

Hence, we define the omnistereoscopic FOV as the spherical surface centered at O, with a radius $r_{\min}$, that marks the minimum distance from the omnistereoscopic camera after which a perceptually continuous depth illusion around the acquisition point can be reproduced. Any point $\mathbf{P_W}$ located at a distance $\|\mathbf{P_W}\| \geq r_{\min}$ can be imaged by either two or four cameras, maintaining a consistent illusion of depth between stereoscopic samples. The stereoscopic FOV of the panoramic camera and the omnistereoscopic FOV are shown in Fig. 5.

3.2 *Disparities*

A point in the scene $\mathbf{P_W}$ is projected by the stereoscopic camera $(\Omega_{L,i}, \Omega_{R,i})$ into the image coordinates $\mathbf{p}_{L,i} = (x_{L,i}, y_{L,i})^T$ and $\mathbf{p}_{R,i} = (x_{R,i}, y_{R,i})^T$ in $\mathbf{im}_{L,i}$ and $\mathbf{im}_{R,i}$, respectively. In this context, we define the horizontal (dh_i) and vertical (dv_i) disparities for the stereoscopic image pair i as

$$dh_i = x_{R,i} - x_{L,i}, \tag{24}$$

$$dv_i = y_{R,i} - y_{L,i}. \tag{25}$$

3.3 *Horizontal Disparity Equations*

The coordinates of point in the scene $\mathbf{P_W}$ after applying the projective transformation defined in Eqs. (22) and (23) are

$$\begin{aligned} x_{L,i} &= -f \frac{\cos\theta_i(X_w - t_{L,1}) + \sin\theta_i(Z_w - t_{L,3})}{-\sin\theta_i(X_w - t_{L,1}) + \cos\theta_i(Z_w - t_{L,3})} \\ &= -f \frac{X_{L,i}}{Z_{L,i}}, \end{aligned} \tag{26}$$

$$\begin{aligned} x_{R,i} &= -f \frac{\cos\theta_i(X_w - t_{R,1}) + \sin\theta_i(Z_w - t_{R,3})}{-\sin\theta_i(X_w - t_{R,1}) + \cos\theta_i(Z_w - t_{R,3})} \\ &= -f \frac{X_{R,i}}{Z_{R,i}}, \end{aligned} \tag{27}$$

where $(X_{j,i}, Y_{j,i}, Z_{j,i})^T$ are the coordinates of a point $\mathbf{P_W}$ in each camera frame, and $t_{j,k}$ ($k \in \{1,2,3\}$) are components of the translation vectors defined in Eqs. (10) to (19).

The image plane coordinates ($x_{L,i}$ and $x_{R,i}$) can be used to expand dh_i and dh_{i+1} [Eq. (24)] as

$$dh_i = -f\left(\frac{X_{R,i}Z_{L,i} - X_{L,i}Z_{R,i}}{Z_{R,i}Z_{L,i}}\right), \tag{28}$$

$$dh_{i+1} = -f\left(\frac{X_{R,i+1}Z_{L,i+1} - X_{L,i+1}Z_{R,i+1}}{Z_{R,i+1}Z_{L,i+1}}\right). \tag{29}$$

Note that since $Z_{R,i} = Z_{L,i} = Z_i$ and $X_{R,i} = X_{L,i} + b$ holds for all i, the disparities in Eq. (28) can be simplified to

$$dh_i = -f\,b\left(\frac{1}{Z_i}\right), \tag{30}$$

$$dh_{i+1} = -f\,b\left(\frac{1}{Z_{i+1}}\right). \tag{31}$$

3.4 *Horizontal Disparity Error*

The difference in the depth estimation from the perspective of two neighbor stereoscopic samples is the parameter we define to study the depth continuity. This depth estimation is the distance of $\mathbf{P_W}$ in the $\mathbf{Z}_i$ axis when imaged by stereoscopic cameras at θ_i and θ_{i+1}. Such depth estimation $(\mathbf{Z}_i, \mathbf{Z}_{i+1})$ is related to the horizontal disparities (dh_i, dh_{i+1}).

The mosaicking of overlapped stereoscopic images requires the registration and alignment of neighbor images. However, this prerequisite is independent of the calculation $(\mathbf{Z}_i, \mathbf{Z}_{i+1})$. Furthermore, warping the stereoscopic images onto a cylinder (or a topologically equivalent surface) centered at $\mathbf{O}$, the horizontal angular disparity defined over the curved surface has to be consistent with d_i and d_{i+1} to convey the same illusion of depth captured from each stereoscopic camera.

After warping the stereoscopic images onto a curved surface centered at $\mathbf{O}$, the angular disparity in azimuth direction defined by corresponding left and right projections of $\mathbf{P_W}$ over any display has to produce the same illusion of depth to the viewer than the originally captured stereoscopic images, which is determined by dh_i and dh_{i+1} in a planar image. Hence, consideration of the similarity between Z_i and Z_{i+1} for each acquisition configuration is a valid approach to study the depth continuity.

The horizontal disparity in the vicinity of the horizontal stitching coordinate $x_{j,i} = \pm x_b$ should remain below a tolerable error. In order to quantize the difference between registered horizontal disparities in neighbor stereoscopic samples, we define the horizontal disparity error (e_h) as

$$e_h = |dh_{i+1} - dh_i|. \tag{32}$$

The region of stitching and blending is critical since it is the region where the artifacts caused by the parallax between individual cameras will appear. It is also the most critical area in terms of defining the omnistereoscopic FOV. Hence, it is important to know what is the closest distance to the scene to guarantee the continuity of the horizontal disparity among mosaicked stereoscopic snapshots.

The closed form of e_h is

$$\begin{aligned} e_h &= f\left|\frac{X_{R,i+1}Z_{L,i+1} - X_{L,i+1}Z_{R,i+1}}{Z_{R,i+1}Z_{L,i+1}} - \frac{X_{R,i}Z_{L,i} - X_{L,i}Z_{R,i}}{Z_{R,i}Z_{L,i}}\right| \\ &= f\,b\left|\frac{1}{Z_{i+1}} - \frac{1}{Z_i}\right|. \end{aligned} \tag{33}$$

As defined, e_h is a function of $\mathbf{P_W}$ after being transformed into each camera frame of coordinates. Hence, in the analysis, we are using $e_h(\mathbf{P_W})$ to express the dependency on the horizontal disparity error of the depth of the point in scene.

The depth of $\mathbf{P_W}$ can be written in terms of cameras $(\Omega_{L,0}, \Omega_{R,0})$ and $(\Omega_{L,\Delta_\theta}, \Omega_{R,\Delta_\theta})$ as follows:

$$Z_i = Z_w - t_{R,3}, \tag{34}$$

$$Z_{i+1} = -\sin\,\Delta_\theta(X_w - t_{R,1}) + \cos\,\Delta_\theta(Z_w - t_{R,3}), \tag{35}$$

where $t_{R,1}$ and $t_{R,3}$ can be obtained from Eqs. (12) to (19) by replacing θ_{i+1} by Δ_θ.

3.5 *Depth Resolution*

The threshold for e_h is a distance ε, which is related to the depth resolution among humans. The threshold ε defines the tolerable horizontal difference between horizontal disparities of the same point $\mathbf{P_W}$ projected in two consecutively acquired stereoscopic images. Each configuration must satisfy

$$e_h \le \varepsilon \tag{36}$$

in the region of overlapping (Sec. 3.1). The value of ε can be estimated by assuming that the depth resolution of each stereoscopic camera is at least equal to the average depth resolution in humans.

The perceptual depth resolution in the average adult population (dZ_h) can be approximated by[29]

$$dZ_h = \frac{z^2\delta_\theta}{d_e}, \tag{37}$$

where δ_θ is the vergence acuity in humans, d_e (20 arcsec) is the average interocular distance in adults (65 mm), and Z_c is the distance from the reference system defined on the stereoscopic camera. A diagram that helps to understand these parameters is shown in Fig. 6(a).

The depth resolution (dZ_c) of a stereoscopic camera can be approximated by[30]

$$dZ_c = \frac{z^2\varepsilon}{f\,b}. \tag{38}$$

In Fig. 6(b), we present an illustration that helps understand dZ_c.

The goal is to have $dZ_c \le dZ_h$ for a camera to surpass or equal the human depth resolution. The largest threshold is defined by $dZ_c = dZ_h$; hence, from Eqs. (37) and (38), the disparity threshold is

$$\varepsilon = f\, b \frac{\delta_\theta}{d_e}. \tag{39}$$

The larger the product $f\,b$, the larger the horizontal disparity threshold. The latter expression can be written as

$$\varepsilon = \left(\frac{W_h b}{2\,\tan\frac{\Delta_a}{2}}\right)\frac{\delta_\theta}{d_e}. \tag{40}$$

The depth accuracy of human vision leads to a very stringent restriction for a stereoscopic camera system.[31] As an example, the sensor width of Nikon 800D is $W_h = 35.8\,10^{-3}$ m, which requires a horizontal disparity resolution of $\varepsilon = 0.19$ μm in order to be comparable with the human depth resolution. However, the pixel size of this camera is 4.8 μm, which is approximately 25 times larger than ε. Hence, the achievable horizontal resolution of a stereoscopic camera using this sensor is well below the depth acuity of the human vision.

Consequently, in a practical scenario, the maximum horizontal resolution of the camera, e.g., one pixel width (s) defines the ε threshold. This threshold has been proven to be sufficient in our rendering experiments.[3,4]

4 Minimum Distance to the Scene

The problem of finding the minimum distance to the scene ($r_{\min}$) that defines the omnistereoscopic FOV requires us to find the points in the scene located in the intersection of stereoscopic FOVs of cameras $(\Omega_{L,i}, \Omega_{R,i})$ and $(\Omega_{L,i+1}, \Omega_{R,i+1})$, constrained to $e_h \leq \varepsilon$. This is done for the neighborhood of the stitching coordinate as explained in Sec. 3.5.

This approach requires only two consecutively acquired stereoscopic images, e.g., the samples $i = 0$ ($\theta_0 = 0$ deg) and $i = 1$ ($\theta_1 = \Delta_\theta$). The search can be restricted to the **XZ**-plane for simplicity and without losing generality. Hence, we are going to assume that the point $\mathbf{P_W}$ is located on the **XZ**-plane. This is illustrated in Fig. 7.

First, we propose to define a ray between the point $\mathbf{P_W}$ and $\mathbf{O}_{R,0}$. This ray ($\mathbf{L}_{R,0}$) intersects the image window of camera $\Omega_{R,0}$ in the coordinates $\mathbf{p}_{R,0} = (-x_b, 0)^T$ as shown in Fig. 7. The ray $\mathbf{L}_{R,0}$ is modeled as

$$\mathbf{L}_{R,0} = \mathbf{a_0} + t_r(\mathbf{a_1} - \mathbf{a_0}), \tag{41}$$

where $\mathbf{a_0}$ is the location of $\mathbf{O}_{R,0}$, $\mathbf{a_1}$ is the intersection of the ray with the image window of camera $\Omega_{R,0}$, both expressed in terms of the global reference frame **XYZ**, and $t_r \in \mathbb{R}$, where $t_r \geq 1$. The parameters for $\mathbf{L_{R,i}}$ are given in Table 1 for all configurations.

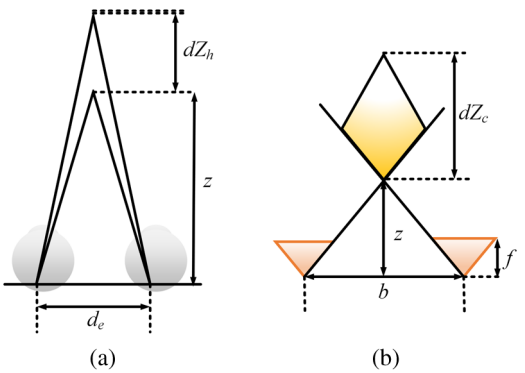

Fig. 6 Depth resolution scheme: (a) human eyes and (b) stereoscopic camera.

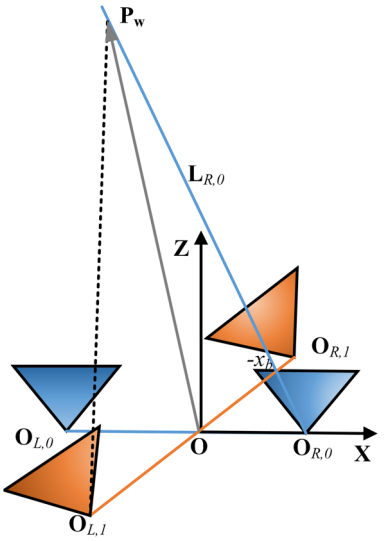

Fig. 7 Ray tracing method to find the minimum distance $r_{\min}$: the ray passing through $\mathbf{O_{R,0}}$ and x_b defines the points on the **XZ**-plane to find the minimum distance for which $e_h \leq \varepsilon$.

The points in the scene $\mathbf{P_W}$, which are in the line $\mathbf{L}_{R,0}$, are projected into the same coordinate point $x_{R,0} = -x_b$ and therefore it is used as reference. The projections of $\mathbf{P_W}$ in the other three image planes are $x_{L,0}$, $x_{R,1}$, and $x_{L,1}$, since $y_{j,0} = y_{j,1} = 0$. These projection points are used to calculate the horizontal disparities dh_0 and dh_1 and e_h for each point $\mathbf{P_W} \in \mathbf{L}_{R,0}$.

The analysis starts by defining a point $\mathbf{P_0} \in \rho\mathbf{L}_{R,0}$ located inside the intersection between stereoscopic FOV of cameras $(\Omega_{L,0}, \Omega R, 0)$ and $(\Omega_{L,1}, \Omega R, 1)$, i.e., the stereoscopic FOV of the panoramic camera (Sec. 3.1). This can be granted by calculating the projection of $\mathbf{P_0}$ in the other three image planes. This projection must satisfy $x_{j,i} \in [-W_h/2, W_h/2]$.

If $e_h(\mathbf{P_0}) > \varepsilon$, then the points $\mathbf{P_W} \in \mathbf{L}_{R,0}$, such that $\|\mathbf{P_W}\| > \|\mathbf{P_0}\|$ must be evaluated. The point $\mathbf{P_1} \in \mathbf{L}_{R,0}$ that $e_h(\mathbf{P_0}) \leq \varepsilon$ defines the minimum distance for omnistereoscopic rendering as $r_{\min} = \|\mathbf{P_1}\|$.

The error in the horizontal disparity e_h is a monotonically decreasing function of the distance between the scene and the camera ($\|\mathbf{P_W}\|$). Therefore, the search for a point $\mathbf{P_W} \in \mathbf{L}_{R,0}$, such that $\|\mathbf{P_W}\| > \|\mathbf{P_0}\|$ will converge to $r_{\min}$. Any point in the scene whose distance from the camera is larger than $r_{\min}$ will be projected in both stereoscopic

Table 1 Parameters to define $L_{R,0}$.

Configuration	$\mathbf{a_0}$	$\mathbf{a_1}$
1	$(b/2, 0, 0)$	$(-x_b + b/2, 0, f)$
2	$(b, 0, 0)$	$(-x_b + b, 0, f)$
3	$(b, 0, \|\mathbf{r_c}\|)$	$(-x_b + b, 0, f + \|\mathbf{r_c}\|)$
4	$(b/2, 0, \|\mathbf{r_c}\|)$	$(-x_b + b/2, 0, f + \|\mathbf{r_c}\|)$

cameras with an error in the horizontal disparity below the perceptual threshold.

This approach to calculate $r_{\min}$, which defines the omnistereoscopic FOV, can be applied to all the acquisition models.

A comparison between acquisition models in terms of the achievable minimum distance to the scene for a consistent depth rendition is presented in Sec. 6.

5 Vertical Disparities

There are two situations where vertical disparities need to be modeled: the first involves the vertical disparities that appear within stereoscopic images $(\mathbf{im}_{L,i}, \mathbf{im}_{R,i})$, and the other case involves the vertical disparities between left and right images to be stitched $[(\mathbf{im}_{L,i}, \mathbf{im}_{L,i+1})$ and $(\mathbf{im}_{R,i}, \mathbf{im}_{R,i+1})]$.

The vertical disparities within stereoscopic samples are negligible unless there are misalignments between stereoscopic pairs of cameras, e.g., when the optical axes of the cameras are not parallel. Even in that case, the vertical disparities can be eliminated using stereoscopic registration techniques.

The vertical disparities between consecutive samples appear within the overlapping regions between neighbor images because of the parallax between contiguous cameras. These disparities may lead to image artifacts, such as ghosting after mosaicking, if not corrected. The amount and variation of these vertical disparities varies with the gazing angle (ϕ) in elevation.

5.1 *Vertical Disparity Equations*

The coordinates of a point $\mathbf{P_W}$ after applying the projective transformation defined in Eqs. (22) and (23) are

$$y_{L,i} = -f\frac{Y_w}{-\sin\theta_i(X_w - t_{L,1}) + \cos\theta_i(Z_w - t_{L,3})} = -f\frac{Y_{L,i}}{Z_{L,i}}, \tag{42}$$

$$y_{R,i} = -f\frac{Y_w}{-\sin\theta_i(X_w - t_{R,1}) + \cos\theta_i(Z_w - t_{R,3})} = -f\frac{Y_{R,i}}{Z_{R,i}}, \tag{43}$$

where $(X_{j,i}, Y_{j,i}, Z_{j,i})^T$ are the coordinates of a point $\mathbf{P_W}$ in each camera frame and t_k $(k \in \{1,2,3\})$ are components of the translation vectors defined in Eqs. (10) to (19). Since the **XZ**-plane contains all the projection centers and optical axes, we can say that $Y_{j,i} = Y_W$ for all the cameras [Eqs. (42) and (43)].

The vertical disparity expression within stereoscopic images $(\mathbf{im}_{L,i}, \mathbf{im}_{R,i})$ is given by

$$dv_i = y_{R,i} - y_{L,i} = -fY_W\left(\frac{Z_{L,i} - Z_{R,i}}{Z_{R,i}Z_{L,i}}\right). \tag{44}$$

5.2 *Vertical Disparities Outside the Overlapping Region*

Any point $\mathbf{P_W}$ in the stereoscopic FOV of $(\mathbf{im}_{L,i}, \mathbf{im}_{R,i})$ will have projections $Z_{L,i} = Z_{R,i}$. Therefore, $dv_i = 0$ for all Y_W [Eq. (44)]. This is valid outside the overlapping regions of every stereoscopic image to mosaic (Sec. 2.3.7).

In a real scenario, cameras will exhibit slight variations from the ideal case, e.g., the optical axes will not be perfectly parallel and their projection center might not lie on the reference plane. Hence, $Z_{L,i} \neq Z_{R,i}$ and $dv_i \neq 0$ for $Y_W \neq 0$. In these cases, pairwise camera calibration followed by stereoscopic image registration will help to reduce or to eliminate undesired vertical disparities before mosaicking.

5.3 *Vertical Disparities Within the Overlapping Region*

The parallax between projection centers produces unwanted vertical disparities in the overlapping regions. The measure of the vertical disparities between neighbor images $(\mathbf{im}_{j,i}, \mathbf{im}_{j,i+1})$ is given by

$$dv_{j,i} = y_{j,i+1} - y_{j,i} = -fY_w\left(\frac{Z_{j,i} - Z_{j,i+1}}{Z_{j,i}Z_{j,i+1}}\right) = -fY_w\vartheta_j, \tag{45}$$

where

$$\vartheta_j = \left(\frac{Z_{j,i} - Z_{j,i+1}}{Z_{j,i}Z_{j,i+1}}\right)$$

for $j \in \{L,R\}$ and $i \in \{0,\ldots,N_1\}$. As defined, this disparity measures the vertical component error between the projections of $\mathbf{P_W}$ in $(\mathbf{im}_{L,i}, \mathbf{im}_{L,i+1})$ or in $(\mathbf{im}_{R,i}, \mathbf{im}_{R,i+1})$.

The vertical disparities in the overlapping regions between images are null for $Y_W = 0$. However, $dv_{j,i} \neq 0$ for gazing directions above and below the horizontal reference plane. Therefore, as can be seen from Eq. (45), the vertical disparities increase with the parameters ϑ_j in the overlapping region.

The $\vartheta_j \neq 0$ explains the appearance of vertical disparities while mosaicking stereoscopic snapshots originating in cameras with distinct projection centers.[3,8] This occurs because a point $\mathbf{P_W}$ has different projections on each camera, i.e., $Z_{j,i} \neq Z_{j,i+1}$. This unwanted effect diminishes with the distance between the scene and the omnistereoscopic camera reference center.

The vertical disparities affect the stitching of neighbor images and are manifested as ghosting after the blending process. This ghosting will affect the stereoscopic result only when it is not corrected before stitching and blending, e.g., by local registration and warping images,[3] and it will be restricted to the overlapping regions.

A comparison between the different configurations in terms of vertical disparities is presented in Sec. 6.6.

6 Results

The goal of our simulations is to contrast the four configurations in order to identify the characteristics useful to improve the design of the acquisition system. We used the native horizontal resolution of each camera given by the pixel size as the horizontal disparity threshold, i.e., $\varepsilon = s$. We used real values extracted from the specifications of three off-the-shelf cameras: one APS-C sensor (Canon 400D) and two full-frame sensors (Nikon 800D and Canon EOS6D). Specifications of each camera are presented in Table 2. In terms of the lenses, we used three generic lenses

Table 2 Image sensor specifications: sensor width W_h, aspect ratio a_r, pixel width s.

Camera	W_h	a_r	s
Canon 400D	$22.2\,10^{-3}$ m	1.5	$5.71\,10^{-6}$ m
Nikon 800D	$35.9\,10^{-3}$ m	1.5	$4.88\,10^{-6}$ m
Canon EOS6D	$35.8\,10^{-3}$ m	1.5	$6.55\,10^{-6}$ m

with horizontal FOV $\Delta_a \in \{122 \text{ deg}, 100 \text{ deg}, 76 \text{ deg}\}$. We chose this particular set of lenses only to illustrate the effect of changing the focal length [Eq. (5)] in the minimum distance to the scene.

The parameters of interest are $r_{\min}$, which depends on e_h (Sec. 3.4), and the vertical disparities within the overlapping regions, which depends on ϑ_j ($j = \{L, R\}$) (Sec. 5.3).

6.1 *Horizontal Disparity Continuity*

In order to find the distance $r_{\min}$, we first calculated e_h between two neighbor stereoscopic samples to mosaic. This was done following the procedure described in Sec. 4. In brief, we defined a ray $\mathbf{L}_{R,0}$ on the $\mathbf{XZ}$-plane, using the stitching position with horizontal coordinates $x_{R,0} = -x_b$ in $\mathbf{im}_{R,0}$ and the projection center $\mathbf{O}_{R,0}$. Then, we computed e_h for the points $\mathbf{P_W} \in \mathbf{L}_{R,0}$, starting at a minimum distance from the camera $\|\mathbf{r_0}\| = 0.3$ m. We calculated e_h using Eq. (32) for all $x_{j,0}$ and $x_{j,1}$ within the overlapping region (Sec. 2.3.8). The depth continuity and vertical disparities were calculated for the overlapping region extending $\pm 0.05 W_h$ around x_b in the horizontal dimension. The variable Δ_x is the deviation from x_b in the horizontal coordinate.

In the first simulation, we compared $r_{\min}$ for the four configurations using different numbers of stereoscopic samples N (different Δ_θ), for $b = 35$ mm (configurations 1 and 2) and the same radial distance $\|\mathbf{r_c}\| = b$ (configurations 3 and 4). We used the same $f = 9.3$ mm for three cases, which corresponds approximately to $\Delta_a = 100$ deg. We used the sensor size of a Canon 400D (APS-C sensor) in landscape orientation. This simulation gives information about $r_{\min}$ as a function of the horizontal stitching coordinate $X_{R,1} = x_b + \Delta_x$. The results of this simulation are presented in Fig. 8.

The critical minimum distance ($\varrho_{\min}$) is defined by the largest value of $r_{\min}$ within blending width Δ_b, which is a few pixels width around x_b (Sec. 2.3.8). In other words, $\varrho_{\min}$ is the practical minimum distance between the omnistereoscopic camera and the scene for a given set of acquisition parameters. Based on the simulation results shown in Fig. 8, the values of $\varrho_{\min}$ for each configuration are presented in Table 3 for x_b ($\Delta_x = 0$) and a given blending region $\Delta_b = 10 \cdot \text{s}$. This table provides a numerical example to show how the minimum achievable distance is reduced by increasing the number of stereoscopic samples for a given f.

The best performance in terms of $\varrho_{\min}$ is achieved by configuration 1 followed by configuration 2, both modeling acquisition methods that are not suitable for acquiring dynamic scenes. The $\varrho_{\min}$ increases for configurations 3 and 4, with configuration 3 being the worst in terms of allowed minimum distance to the camera. Hence, the price to pay for having $\|\mathbf{r_c}\| \neq 0$ is to increase the achievable minimum distance to the camera.

A small $\varrho_{\min}$ is necessary to render omnistereoscopic images of indoor scenes where objects may be closer to the camera than in outdoor scenes. This can be achieved by increasing the number of samples N, but that option implies adding more cameras in configuration 3 or 4, or taking more samples in configurations 1 and 2, which leads to increasing the complexity and cost of the omnistereoscopic system. Another option is changing the focal length and the number of samples to get a compromise between cost and efficiency. The results of doing this is simulated in Sec. 6.4.

Another useful conclusion from this simulation is that the selection of stitching position x_b, as defined by Eq. (4), can be optimized for each configuration. For instance, a stitching point shifted $\Delta_x = -0.05 W_h$ with respect to x_b leads to an effective minimum distance in the range of 4.5 to 5.2 m instead of 1.9 to 1.2 m as it would be stitching at x_b. Hence, the acquisition model shows that the stitching position x_b has a larger impact than the acquisition model used.

Probably, the most important conclusion is that there is not much difference between different configurations in terms of the parameter $\varrho_{\min}$. For instance, it can be seen from Table 3 that $\varrho_{\min}$ varies a few tens of centimeters between configurations, which means that all variations of the acquisition model performs similarly in real acquisition scenarios.

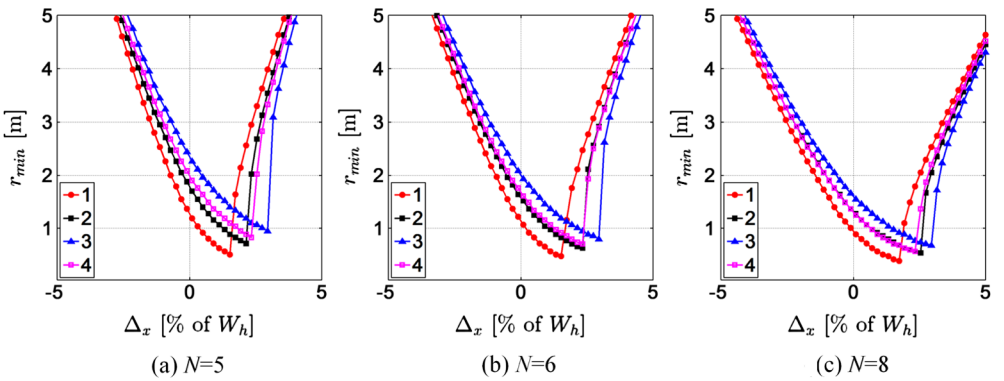

Fig. 8 The minimum distance for stereoscopic rendering $r_{\min}$ as a function of the horizontal bias Δ_x from stitching coordinate x_b ($X_{R,1} = x_b + \Delta_x$), for configurations 1 to 4 and for different number of stereoscopic samples N [Canon 400D, $f = 9.3$ mm ($\Delta_a = 100$ deg), $b = 35$ mm, and $\|\mathbf{r_c}\| = b$].

Table 3 Example of the effective minimum distance $\varrho_{\min}$ for x_b ($\Delta_x = 0$, $\Delta_b = 10 \cdot s$): $\varrho_{\min}$ (Canon 400D, $f = 9.3$ mm $\Delta_a = 100$ deg, and $b = 35$ mm).

Configuration	$N = 5$	$N = 6$	$N = 8$
1	1.2 m	1.1 m	0.9 m
2	1.7 m	1.5 m	1.3 m
3	2.3 m	2.0 m	1.6 m
4	1.90 m	1.60 m	1.35 m

6.2 *Effect of Radial Displacement*

The radial displacement $\|\mathbf{r_c}\|$ is relevant for configurations 3 and 4 to model the physical limitations in a multiple-camera system to co-locate the projection centers. The smaller $\|\mathbf{r_c}\|$, the more similar are configurations 3 and 2, and similarly, the more similar configuration 4 is to configuration 1. In order to see this effect, we recalculate $r_{\min}$ for the case presented in Fig. 8(b) for $\|\mathbf{r_c}\| = 0$, $\|\mathbf{r_c}\| = b/2$, and $\|\mathbf{r_c}\| = 2b$. The results are presented in Fig. 9. Notice that configuration 1 coincides with configuration 4 and configuration 2 coincides with configuration 3 for $\|\mathbf{r_c}\| = 0$ [Fig. 9(a)].

The larger $\|\mathbf{r_c}\|$, the larger is $\varrho_{\min}$; in other words, the larger is the distance between the camera and the scene. A possible acquisition system is using configuration 4 for $\|\mathbf{r_c}\| \leq b$, but interleaving a number of stereoscopic rigs in order to reduce $\varrho_{\min}$. Another method to reduce $\|\mathbf{r_c}\|$ is using mirrors to relocate each camera projection center closer to $\mathbf{O}$. However, the latter may require a smaller Δ_a (larger f) and more camera pairs, all of which increase the complexity and cost of the acquisition system. A better solution to reduce $\varrho_{\min}$ is selecting a bias in the stitching point, in this case, $|x_b'| > |x_b|$.

6.3 *Optimum Stitching Position*

The parameter Δ_θ represents the relative azimuthal angle between two consecutive stereoscopic samples. In addition, each Δ_θ determines a shift in the horizontal coordinates for the optimum stitching position x_b. The optimality of this new stitching position is determined by the closest $r_{\min}$ to the camera as shown in Fig. 8.

The larger the number of stereoscopic samples to mosaic, the smaller is Δ_θ, and the optimal stitching coordinate moves closer to the center on each image to mosaic. Hence, the position x_b can be corrected in rendering time to avoid discontinuities in horizontal disparities in gazing directions where the scene is too close to the omnistereoscopic camera. Hence, there is no need to use the same relative position x_b for the stitching; a different distance to the scene can determine a different stitching position for each stereoscopic pair defined for each configuration.

The results for $r_{\min}$ shown in Fig. 8 are presented as a function of coordinate $x_{R,1}$ instead of around a determined stitching coordinate x_b as were presented in Fig. 8. This helps to illustrate the different locations, relative to the image center, of the optimum stitching point when changing Δ_θ.

6.4 *Effect of Focal Length*

Using wide-angle lenses, i.e., larger Δ_a (shorter f), reduces the number of stereoscopic samples. However, a larger Δ_a introduces distortions at the edges of the stereoscopic images, especially for large baselines. Despite this disadvantage, a slight improvement in the effective minimum distance can be achieved by reducing f. We repeated the calculation $r_{\min}$ presented in Sec. 6.1, but this time we used a smaller focal length ($f = 6.1$ mm, $\Delta_a = 122$ deg). The results presented in Fig. 10 show a small but measurable reduction of $\varrho_{\min}$ with respect to using $f = 9.3$ mm ($\Delta_a = 100$ deg) on the same sensor.

Conversely, it is expected that a larger $\varrho_{\min}$ will result when using larger f. This is shown in Fig. 11 for the same camera parameters as the previous example but for $f = 14.2$ mm ($\Delta_a = 76$ deg).

6.5 *Effect of the Sensor Size*

The sensor size affects f for a given Δ_a [Eq. (5)]. For instance, in this configuration, a Canon 400D (APS-C sensor) has $f = 9.3$ mm for $\Delta_a = 100$ deg, while $f = 15.1$ mm for a full-size sensor ($W_h = 35$ mm) (Table 2). Compared to the pixel size of the Canon 400D sensor, the pixel size of the Canon EOS6D is ~15% wider, while the pixel width of the Nikon 800D is 15% smaller. The results for finding $r_{\min}$ for the four configurations are presented in Fig. 12.

For a comparable sensor size and the same focal length, the Nikon 800D sensor performance is worse than Canon EOS6D sensor in terms of $\varrho_{\min}$ because the pixel size of the former is smaller, which makes the threshold for e_h smaller and pushes the minimum tolerable distance away

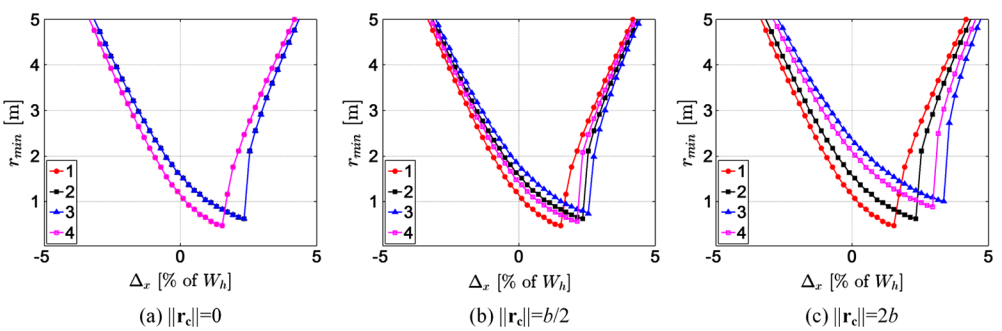

Fig. 9 Comparison of $r_{\min}$ as a function of the stitching bias Δ_x for all configurations after changing the radial distance $\|\mathbf{r_c}\|$ in configurations 3 and 4 [Canon 400D, $N = 6$, $f = 9.3$ mm ($\Delta_a = 100$ deg), and $b = 35$ mm].

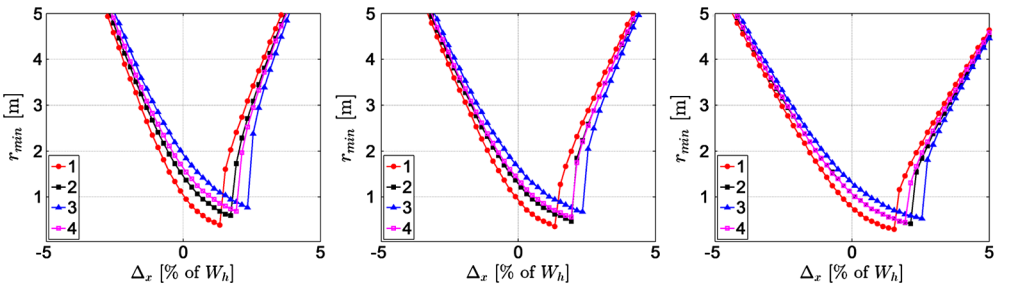

Fig. 10 Comparison of r_{min} as a function of the stitching bias Δ_x for configurations 1 to 4 and for different N when reducing the focal length to $f = 6.1$ mm ($\Delta_a = 122$ deg) (Canon 400D, $b = 35$ mm and $\|\mathbf{r_c}\| = b$).

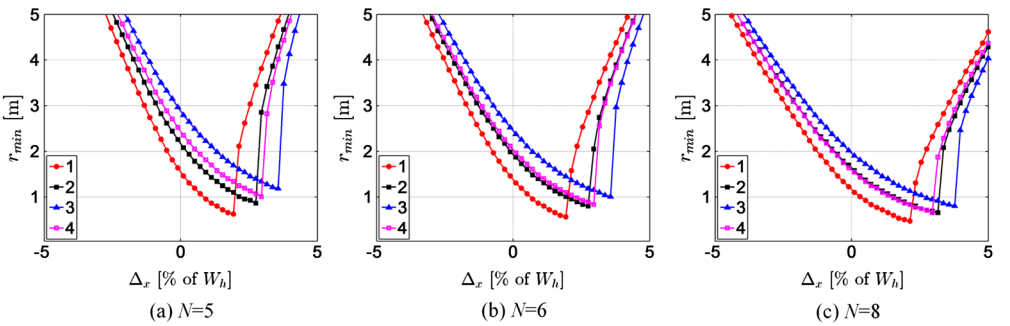

Fig. 11 Comparison of r_{min} as a function of the stitching bias Δ_x for configurations 1 to 4 and for different N when increasing the focal length to $f = 14.2$ mm ($\Delta_a = 76$ deg) (Canon 400D, $b = 35$ mm and $\|\mathbf{r_c}\| = b$).

from the camera. The Canon 400D sensor is smaller and its f is also smaller, so ϱ_{min} for the same sampling angle Δ_θ is closer to the omnistereoscopic camera.

6.6 *Vertical Disparities*

An important side effect of the spatial distribution between cameras in each acquisition configuration is the manifestation of vertical disparities in the overlapping region between neighbor stereoscopic images (Sec. 5.3). We compared the four configurations in terms of the coefficients ϑ_j, which determine the magnitude of the vertical disparities in the overlapping areas. As seen in Eq. (45), the undesired vertical disparities are proportional to $\mathbf{Y_W}$. But, this error also depends on the distance between the camera and $\mathbf{P_W}$ on the $\mathbf{XZ}$-plane. The results of calculating the value of ϑ_j as a function of the distance to the camera $\|\mathbf{r}\| = \|\mathbf{P_W}\|$ for the four configurations are presented in Fig. 13. These results are valid for a camera Canon 400D, $b = 65$ mm, $\|\mathbf{r_c}\| = 65$ mm, $N = 6$, and $\Delta_a = 100$ deg.

A nonzero ϑ_j leads to a vertical disparity error (Sec. 5.3) in the overlapping regions; hence, it is desirable to take into account this effect when positioning the omnistereoscopic camera, especially for acquiring scenarios that are relatively close in certain elevation angles. This relationship among vertical disparities, the proximity to the camera, and the gazing direction in elevation explains the vertical disparities that appear at the top and bottom parts of the mosaicked images, which are both at the shorter distance $\|\mathbf{r}\|$ on the $\mathbf{XZ}$-plane and, at the same time, have large $\mathbf{Y_W}$ components. Fortunately, both coefficients converge to zero at a relatively short distance from the camera for all the configurations of the acquisition model.

The magnitude of this error is also dependent on the particular stitching position in the image pair x_b. In Figs. 13(c) to 13(f), we present the value of ϑ_j over the overlapping region, for $\|\mathbf{r}\| = 50$ cm of the camera and $\|\mathbf{r}\| = 100$ cm.

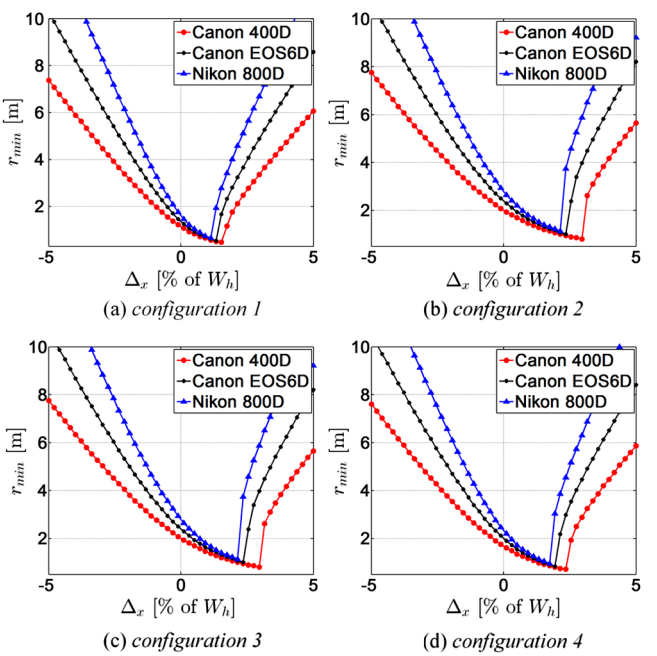

Fig. 12 Comparison of r_{min} as a function of Δ_x for configurations 1 to 4 and for different sensors while maintaining a given field of view Δ_a.

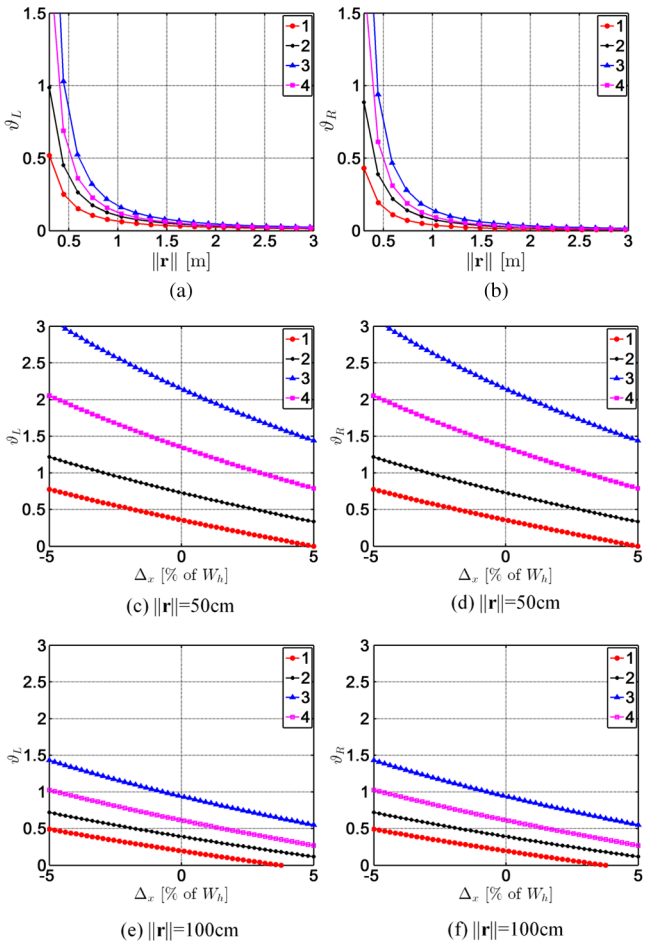

Fig. 13 Vertical disparity coefficients ϑ_j as a function of the distance to the camera $\|\mathbf{r}\|$ for configurations 1 to 4 and for $x_{R,1} = x_b$ ($\Delta_x = 0$): (a) ϑ_L and (b) ϑ_R, and the variation of these coefficients around x_b as function Δ_x when the distance to the scene is [(c) and (d)] $\|\mathbf{r}\| = 50$ cm and [(d) and (e)] $\|\mathbf{r}\| = 100$ cm.

These results show that the vertical disparities over each overlapping region decrease with the distance between the scene projection on the **XZ**-plane and the camera.

7 Conclusions

In this paper, we presented an acquisition model with four configurations suitable to describe the variety of imaging systems to acquire and render stereoscopic panoramas with horizontal stereo. In other words, we focused on the acquisition techniques to capture the visual scene from two distinct and coplanar viewpoints with horizontal parallax. From these methods, we focused on the mosaicking of a reduced number of partially overlapped stereoscopic images given its benefits to acquire dynamic scenes. We studied two parameters relevant for the design of omnistereoscopic acquisition systems based on mosaicking: one involves the constraints to reproduce a continuous depth illusion around the acquisition point, and the other is the appearance of undesired vertical disparities.

We proposed a projective approach based on pin-hole cameras to model each configuration, which can be adapted easily to all the four variations of the acquisition model to study the parameters of interest for the mosaicking approach. In order to model the depth continuity, we introduced theoretical and practical thresholds for the depth resolution. Based on our projective approach, we defined a parameter to study: the minimum distance between the camera and the scene to reproduce a continuous depth illusion around the acquisition point. Furthermore, we introduced the concept of a safe distance around the omnistereoscopic camera to acquire the scene stereoscopically, which we called the omnistereoscopic FOV, and we proposed a ray traced method to determine its location. Based on extensive ray tracing simulations, we compared the effect of the focal length, number of stereoscopic samples, and radial distance on the minimum distance from the camera.

The main conclusion from our simulations is that there is no substantial difference between camera geometries in terms of the minimum distance between camera and the scene. However, we found differences in the optimal location for stitching stereoscopic samples, which can be attributed to the geometry of the omnistereoscopic system, and that can be used to improve the rendering process. Also from our simulation results, we proposed strategies affecting the design of simultaneous acquisition systems that may lead to increasing efficiency in the omnistereoscopic acquisition and rendering.

Finally, we used our acquisition model to find a closed form for the vertical disparity equations. These equations help us to model the relationship between the gazing angle in elevation and the projection center parallax in the introduction of undesired vertical disparities. We simulated the effects of distance to the scene and the elevation angle in the appearance of vertical disparities around the stitching coordinates.

We did not cover in this paper the perceptual distortions introduced by mosaicking a limited number of images. This is an interesting problem related to the stereoscopic quality of the omnistereoscopic image formation that requires further research.

Acknowledgments

This work was supported by the Ontario Graduate Scholarship fund.

References

1. L. E. Gurrieri and E. Dubois, "Acquisition of omnidirectional stereoscopic images and videos of dynamic scenes: a review," *J. Electron. Imaging* **22**(3), 030902 (2013).
2. L. E. Gurrieri and E. Dubois, "Optimum alignment of panoramic images for stereoscopic navigation in image-based telepresence systems," in *Proc. of the 11th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras*, Vol. 11, pp. 351–358, IEEE (2011).
3. L. E. Gurrieri and E. Dubois, "Efficient panoramic sampling of real-world environments for image-based stereoscopic telepresence," *Proc. SPIE* **8288**, 82882D (2012).
4. L. E. Gurrieri and E. Dubois, "Stereoscopic cameras for the real-time acquisition of panoramic 3D images and videos," *Proc. SPIE* **8648**, 86481W (2013).
5. H. Baker et al., "Capture considerations for multiview panoramic cameras," in *2012 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition Workshops*, pp. 37–44, IEEE (2012).
6. C. Weissig et al., "The ultimate immersive experience: panoramic 3D video acquisition," *Lec. Notes Comput. Sci.* **7131**, 671–681 (2012).
7. V. Couture, M. S. Langer, and S. Roy, "Panoramic stereo video textures," in *IEEE Int. Conf. on Computer Vision*, pp. 1251–1258, IEEE (2011).
8. V. Vanijja and S. Horiguchi, "A stereoscopic image-based approach to virtual environment navigation," *Comput. Internet Manage.* **14**(2), 68–81 (2006).
9. R. O. Reynolds, "Design of a stereo multispectral CCD camera for Mars pathfinder," *Proc. SPIE* **2542**, 197–206 (1995).
10. F. Hongfei et al., "Immersive roaming of stereoscopic panorama," in *2008 Int. Conf. on Cyberworlds*, pp. 377–382, IEEE (2008).

11. H.-C. Huang and Y.-P. Hung, "Panoramic stereo imaging system with automatic disparity warping and seaming," *Graph. Models Image Process.* **60**(3), 196–208 (1998).
12. F. Huang and Z.-H. Lin, "Stereo panorama imaging and display for 3D VR system," in *IEEE Congress on Image and Signal Processing*, Vol. 3, pp. 796–800, IEEE (2008).
13. H. Ishiguro, M. Yamamoto, and S. Tsuji, "Omni-directional stereo," *IEEE Trans. Pattern Anal. Machine Intell.* **14**(2), 257–262 (1992).
14. S. Peleg and M. Ben-Ezra, "Stereo panorama with a single camera," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 1, pp. 395–401, IEEE (1999).
15. K. Yamada et al., "Generation of high-quality stereo panoramas using a three-camera panorama capturing system," *J. Inst. Image Inf. Telev. Eng.* **55**(1), 151–158 (2001).
16. K. Yamada et al., "Structure analysis of natural scenes using census transform and region competition," *Proc. SPIE* **4310**, 228–237 (2000).
17. R. G. Baker, F. A. Baker, and J. A. Conellan, "Panoramic stereoscopic camera," U.S. Patent 2008/0298674 A1 application (2008).
18. H. H. Baker and P. Constantin, "Panoramic stereoscopic camera," U.S. Patent 2012/0105574 application (2010).
19. V. Vanijja and S. Horiguchi, "Omni-directional stereoscopic images from one omni-directional camera," *J. VLSI Signal Process.* **42**(1), 91–101 (2006).
20. W. A. Clay, "Methods of stereoscopic reproduction of images," U. S. Patent 3225651 granted (1965).
21. F. Huang, R. Klette, and K. Scheibe, *Panoramic Imaging: Sensor-Line Cameras and Laser Range-Finders*, John Wiley & Sons Inc., Hoboken, NJ (2008).
22. S. Peleg and M. Ben-Ezra, "Stereo panorama with a single camera," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 395–401, IEEE (1999).
23. P. Bourke, "Omni-directional stereoscopic fisheye images for immersive hemispherical dome environments," in *Proc. of the Computer Games and Allied Technology*, pp. 136–143, World Academy of Science, Engineering and Technology (2009).
24. R. Szeliski, "Video mosaics for virtual environments," *IEEE Comput. Graph. Appl.* **16**(2), 22–30 (1996).
25. V. Kwatra et al., "Graphcut textures: image and video synthesis using graph cuts," *Proc. ACM SIGGRAPH* **22**(3), 277–286 (2003).
26. S. J. Ha et al., "Panorama mosaic optimization for mobile camera systems," *IEEE Trans. Consum. Electron.* **53**(4), 1217–1225 (2007).
27. T. Yan et al., "Seamless stitching of stereo images for generating infinite panoramas," in *Proc. of the 19th ACM Symp. on Virtual Reality Software and Technology*, pp. 251–258, Association for Computing Machinery (2013).
28. E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*, Prentice Hall, Englewood Cliffs, New Jersey (1998).
29. J. M. Harris, "Monocular zones in stereoscopic scenes: a useful source of information for human binocular vision?," *Proc. SPIE* **7524**, 151–162 (2010).
30. C. Chang and S. Chatterjee, "Quantization error analysis in stereo vision," in *IEEE Asilomar Conf. on Signals, Systems and Computers*, Vol. 2, pp. 1037–1041, IEEE (1992).
31. M. Kytö, M. Nuutinen, and P. Oittinen, "Method for measuring stereo camera depth accuracy based on stereoscopic vision," *Proc. SPIE* **7864**, I1–I9 (2011).

Luis E. Gurrieri received his BEng in electronic engineering from the University of Buenos Aires in 1998 and MSc in electrical engineering from the University of Manitoba in 2006. From 1998 to 2005, he worked in IT and Telecommunication companies, including Ericsson and AT&T. From 2005 to 2009, he was a research engineer at the Communications Research Center in Ottawa, Canada. He is currently working toward his PhD degree in electrical and computer engineering at the University of Ottawa, where his main research area is stereoscopic vision for image-based telepresence.

Eric Dubois is a professor at the School of Electrical Engineering and Computer Science, University of Ottawa, Canada. His research has centered on the compression and processing of still and moving images, and on multidimensional digital signal processing theory. His current research is focused on stereoscopic and multiview imaging, image sampling theory, image-based virtual environments, and color signal processing. He is a fellow of IEEE, the Canadian Academy of Engineering, and the Engineering Institute of Canada, and is a recipient of the 2013 George S. Glinski Award for Excellence in Research of the Faculty of Engineering.