

A model for the omnidirectional acquisition and rendering of stereoscopic images for human viewing

Luis E. Gurrieri* and Eric Dubois

Abstract

Interactive visual media enable the visualization and navigation of remote-world locations in all gaze directions. A large segment of such media is created using pictures from the remote sites thanks to the advance in panoramic cameras. A desirable enhancement is to facilitate the stereoscopic visualization of remote scenes in all gaze directions. In this context, a model for the signal to be acquired by an omnistereoscopic sensor is needed in order to design better acquisition strategies. This *omnistereoscopic viewing model* must take into account the geometric constraints imposed by our binocular vision system since we want to produce stereoscopic imagery capable to induce stereopsis consistently in any gaze direction; in this paper, we present such model. In addition, we discuss different approaches to sample or to approximate this function and we propose a general acquisition model for sampling the omnistereoscopic light signal. From this model, we propose that by acquiring and mosaicking sparse sets of partially overlapped stereoscopic snapshots, a satisfactory illusion of depth can be evoked. Finally, we show an example of the rendering pipeline to create the omnistereoscopic imagery.

Luis E. Gurrieri, "A model for the omnidirectional acquisition and rendering of stereoscopic images for human viewing", 3D Research, Volume 6, Issue 35, December 2015.

Online date of publication: Oct. 1, 2015. The final publication is available at Springer via:
<http://dx.doi.org/10.1007/s13319-015-0069-0>

*Luis E. Gurrieri, University of Ottawa. luis.gurrieri@uottawa.ca

1 Introduction

The goal of interactive visual media is to evoke an immersive experience by means of high-quality images and videos of remote real-world locations. The interactivity is implemented by enabling the user to dynamically choose the gaze direction (\mathcal{D}) and viewpoint location (\mathbf{r}) in the virtual scene. This can be done by using synthetic 3D-models of the remote sites or by using panoramic imagery. A feature that can enhance the realism of such navigable imagery is to incorporate binocular vision.

The development of image-based virtual reality (VR) applications has gained momentum recently with the popularization of head mounted displays (HMD) [1]. Companies such as Google [2] and Samsung [3] have recently proposed omnistereoscopic 3D cameras to produce HMD suitable content. This is done by acquiring the necessary data set of images to produce correct stereoscopic views in all gaze directions; this is a challenging problem in terms of the 3D panoramic camera design.

One of the novelties of this paper resides in the derivation of functions that represent the visual signals to be acquired by an omnidirectional (stereoscopic) sensor. We present in Section 3 the *omnidirectional viewing function* derivation from the plenoptic model discussed in Section 2. This function is used to justify a general model for the acquisition of omnistereoscopic imagery and its different variations or *configurations* [4, 5]. The acquisition model can be applied to most of the panoramic cameras capable to produce stereoscopic imagery suitable for human viewing, e.g., stereoscopic images with horizontal parallax. We describe the acquisition model in Section 4.

In the second part of this paper, we argue that a satisfactory approximation of the omnistereoscopic viewing function can be achieved by using sparse sets of stereoscopic snapshots. This is important since some configurations of our acquisition model are suitable for the simultaneous capture of multiple stereoscopic images, covering the whole scene in one simultaneous snapshot [6] of multiple sensors. These geometric configurations lead to the design of panoramic cameras to acquire stereoscopic snapshots and videos of dynamic scenes omnistereoscopically. Finally, we present in Section 5 our rendering algorithm for creating omnistereoscopic visual media by stitching images from the acquired sparse sets.

This paper's goal is to propose a general model for omnidirectional acquisition of stereoscopic images to be used in analyzing various cameras and techniques to produce stereoscopic panoramas for human viewing. A main conclusion derived from these models is that a satisfactory illusion of depth based on horizontal disparity between the scene's elements can be achieved by using a limited number of stereoscopic samples of the viewing function. We illustrate this claim by producing various omnistereoscopic views of different scenes which can be seen in Section 6 and whose navigable versions can be explored in [7]. We evaluated our acquisition model using parameters such as the amount of vertical disparities and the continuity of the depth illusion after mosaicking partial samples based on horizontal disparity continuity [5]. The subjective evaluation of the sparse set method to create omnistereoscopic imagery is not part of this paper, but its importance is discussed Section 7.

The rendering procedure is given just as a tentative example of using sparse sets of stereoscopic images; hence, we do not evaluate its complexity or contrast its efficiency against other methods. However, in previous papers, we reported on the feasibility and limitations of using sparse sets of stereoscopic images [5] and contrasted this approach against other acquisition-rendering methods [4].

1.1 Background

Panoramic stereo consisting of coaxial cameras and mirrors, a.k.a. catadioptric stereo [4, 8], can acquire images omnidirectionally by using relatively simple hardware; the acquired pairs of panoramas exhibits vertical (instead of horizontal) disparities. This is not an obstacle to produce omnidirectional depth-maps of the scene, which can be used to render synthetic omnistereoscopic imagery by means of region selective image warping guided by the calculated depth information. However, this approach is computationally costly for a real-time scenario and it is prone to depth estimation errors; rendering occluded areas poses another problem. All these limitations have motivated the developing of alternative cameras suitable to produce gaze-direction-independent stereoscopic imagery intended for human viewing. These methods are based on mosaicking partially overlapped stereoscopic views, which are directly acquired with horizontal baselines [2, 3].

Over the last decade, multiple omnistereoscopic techniques have been proposed to produce stereoscopic panoramas [4]; the technique used is determined by the final application. For instance, the virtual navigation of static environments can be implemented using the sequential acquisition methods [9, 10, 11]. However, this approach cannot be used to acquire a large collection of omnistereoscopic imagery, particularly when the scene is time-variant such as in the case of navigation simulations. The latter may require multiple-camera sensors which can acquire omnistereoscopic imagery and videos in one snapshot [12, 13, 14, 15]. All these acquisition scenarios require the sampling of the plenoptic function from two distinct viewpoints in every gaze direction with respect to a reference viewpoint in space.

The omnistereoscopic viewing model represents the visual signal that, once measured and sampled, can be used to produce stereoscopic images. The produced imagery must be suitable to emulate the appropriate stimulus for the binocular visual system. Hence, the model must incorporate the constraints of our visual system to define the function to be acquired for a given gaze direction.

Our previous publications describe our path towards the definition of an omnistereoscopic viewing model. For instance, in [4], we reviewed the different methods and cameras to produce omnistereoscopic imagery for human viewing; we put special attention on techniques to capture the whole scene at once. Then, in [12], we presented a possible solution based on using three panoramic snapshots acquired in a co-planar triangular pattern. We generalized this technique in [16] where we presented a method based on mosaicking a limited number of partially overlapped stereoscopic snapshots; we suggested in this publication what will become our acquisition model and possible rendering techniques. In our last contribution [5], we contrasted various acquisition configurations based on parameters such as vertical disparities and minimum distance to each camera. However, it was still missing a model to describe the omnistereoscopic light to be acquired by an stereo panoramic sensor, so we introduce it in this paper.

The plenoptic model as defined by Adelson and Bergen [17] is a ray-tracing model of the light arriving at a point in space that accounts for the light as a function of time and direction of arrival, time-of-arrival and wavelength. This model has been used to define a model for the light function acquired by an omnidirectional sensor [18] and for rendering of panoramic (monoscopic) images [19]. A typical omnidirectional sensor integrates the plenoptic function over a spectral range to reduce its dimensionality [20]. This time-variant, direction-dependent function contains the color information in a given color space. We denoted this signal as the *viewing function*, which is the omnidirectional signal to be acquired by a panoramic sensor. The integration of each component of the *viewing function* over a time interval (exposure time) produces an omnistereoscopic (monoscopic) image. The spatial sampling of this image produces a discrete valued panoramic image.

The omnistereoscopic viewing model we propose herein is also derived from the plenoptic model, but it considers the light arriving at two distinct viewpoints in space separated by a horizontal

baseline. We define the *omnistereoscopic viewing function* that models the signal to be acquired to produce omnistereoscopic imagery for human viewing. Our model is detailed in Section 3.1.

The geometric constraints of the omnistereoscopic acquisition model have been examined by Huang et al. in [21, 22, 23]. However, this omnistereoscopic model is applicable to line cameras only. Our model in contrast represents the general visual function to be acquired to reproduce a satisfactory depth illusion in arbitrary gaze directions, including line cameras as a special case.

In addition to defining the omnistereoscopic function to be measured, we discuss different approaches to acquire this function. One of the acquisition systems that best approximates the direct capture of the *omnistereoscopic function* is a stereoscopic rig where the user controls \mathcal{D} by selecting the panning direction of the camera arrangement [24]. Alternative acquisition methods use a cluster of stereoscopic panoramas to produce approximations of the binocular views predicted by our model [12]. The imagery produced using these methods is correct in certain gaze directions, but they exhibit disparity distortions in other looking angles. This problem also appears for column-wise image scanning [9, 23], especially when they are displayed on curved surfaces [25, 26]. Other deviations of the model are introduced by using sparse stereoscopic samples of the *omnistereoscopic function* [16, 15, 13].

2 Omnidirectional viewing function

The *viewing function* (\mathbf{C}) contains the necessary visual information to reconstruct omnidirectional images. This function models the color information in a given color space (\mathcal{C}) as a function of the gaze direction \mathcal{D} and the instant t . The sampling of this signal by means of a panoramic sensor is necessary to produce monoscopic, gaze direction-dependent imagery. The *viewing function* is necessary to model another function, the *omnistereoscopic function* (\mathbf{C}_S), which, after being properly measured and spatially-sampled, helps to produce omnistereoscopic imagery for human viewing. The omnistereoscopic model, from where the \mathbf{C}_S is defined, consists a pair of wide-angle images, acquired from two viewpoints ($\mathbf{O}_L, \mathbf{O}_R$) with horizontal parallax with respect to \mathbf{r} , the reference viewpoint in 3D-space where the acquisition takes place.

2.1 Plenoptic model

The plenoptic function (\mathcal{P}) models the rays of light arriving at a reference viewpoint \mathbf{r} and it is a function of \mathcal{D} or the direction of arrival of the ray and the light wavelength (λ). The plenoptic function is denoted

$$\mathcal{P}(\theta, \phi, \lambda, \mathbf{r}, t), \quad (1)$$

where \mathbf{r} is a reference viewpoint with respect to a global frame of coordinates \mathbf{XYZ} in \mathbb{R}^3 , θ (azimuth) and ϕ (elevation) angles define the gaze direction $\mathcal{D} \in \mathbb{S}^2$, where \mathbb{S}^2 is the unit sphere defined by $\mathbb{S}^2 = \{\rho \in \mathbb{R}^3 : \|\rho\| = 1\}$, $\lambda \in \mathbb{R}$ is the light wavelength and $t \in \mathbb{R}$ is the time variable.

In this context, the plenoptic function is used to derive the *viewing function* which models the omnidirectional signal to be captured by a single-viewpoint panoramic sensor. A graphical interpretation of the plenoptic model is illustrated in Fig. 1. The gaze direction \mathcal{D} is measured with respect to the singular viewpoint located at \mathbf{r} . The reference of coordinates $\mathbf{X}'\mathbf{Y}'\mathbf{Z}'$ is centered at \mathbf{r} , which is measured with respect to a global reference frame \mathbf{XYZ} . Hence, the coordinate transformation between $\mathbf{X}'\mathbf{Y}'\mathbf{Z}'$ and \mathbf{XYZ} is a pure translation $\mathbf{T} = \mathbf{r}$. The unit vector $\hat{\mathbf{v}}$, as shown in Fig. 1, indicates the gaze direction with respect to $\mathbf{X}'\mathbf{Y}'\mathbf{Z}'$.

Some dimensions of \mathcal{P} are lost when acquired by a panoramic sensor, unless it is captured by an image sensor specifically designed to preserve all plenoptic characteristic of the arriving light [27].

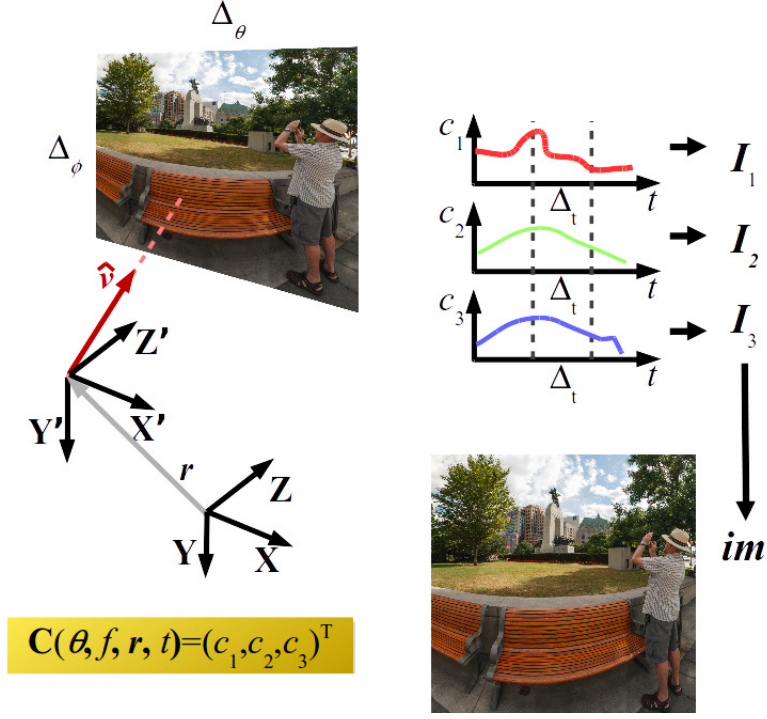


Figure 1: The image \mathbf{im} is defined as a subset of \mathbf{I} for a partial field-of-view (FOV) $(\Delta_\theta, \Delta_\phi)$, a gaze direction \mathcal{D} and a location \mathbf{r} .

2.2 Color function

The function to be measured is a direction-dependent signal for a given color space \mathcal{C} , which is determined by the sensor technology. The color signal is defined by integrating the product of the *color matching functions* $(\Phi_n(\lambda))$ and \mathcal{D} over a wavelength range. This approach leads to define a measurable signal \mathbf{C} which is measured by all image sensors. The color matching functions of the sensor are usually designed to approximate those of the human visual color space [28]. In our model, we use trichromatic color spaces without loss of generality.

The measurable color information in the sensor sensitivity range is denoted by c_n , also known as the *tristimulus values*. A limited spectral range between λ_{min} and λ_{max} is used to produce imagery matching the response in the human visual system. However, most sensors are capable of wider spectral ranges. The tristimulus values are given by

$$c_n(\theta, \phi, \mathbf{r}, t) = \int_{\lambda_{min}}^{\lambda_{max}} \mathcal{D}(\theta, \phi, \lambda, \mathbf{r}, t) \cdot \Phi_n(\lambda) d\lambda, \quad (2)$$

where $c_n : \mathbb{S}^2 \times \mathbb{R}^3 \times \mathbb{R} \rightarrow \mathbb{R}$, for each chromatic component $n \in \{1, 2, 3\}$, are functions of \mathbf{r} , the gaze direction \mathcal{D} and the instant t .

2.3 Monoscopic viewing function

The *viewing function* $\mathbf{C} : \mathbb{S}^2 \times \mathbb{R}^3 \times \mathbb{R} \rightarrow \mathbb{R}^3$ is defined by

$$\mathbf{C}(\mathbf{u}) = (c_1(\mathbf{u}), c_2(\mathbf{u}), c_3(\mathbf{u}))^T, \quad (3)$$

where $\mathbf{u} = (\mathcal{D}, \mathbf{r}, t)$. In other words, for a given point $\mathbf{r} \in \mathbb{R}^3$, a gaze direction $\mathcal{D} \in \mathbb{S}$, and an instant $t \in \mathbb{R}$, the signal to be measured in a color space \mathcal{C} is given by the array of three real-valued functions $(c_1, c_2, c_3)^T$. Note that we can choose to work with any non-singular transformation of \mathbf{C} , corresponding to different choices for the basis of the color space \mathcal{C} .

2.4 Exposure integration

Image sensors measure each c_n by accumulating electric charges over a time interval Δ_t . The voltage produced is proportional to the light intensity for each color component of \mathcal{C} . This technology is used in charged-coupled devices (CCD) and the most widespread active pixel sensors (APS); the latter technology is commonly found in mobile cameras, webcams, security and digital single lens reflex (DSLR) cameras.

The vector function obtained by the exposure integration is the *omnidirectional image* (\mathcal{I}), which can be defined as

$$\mathcal{I}(\theta, \phi, \mathbf{r}) = \int_{\Delta_t} \mathbf{C}(\theta, \phi, \mathbf{r}, t) dt, \quad (4)$$

$$= (\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3), \quad (5)$$

where

$$\mathcal{I}_n(\theta, \phi, \mathbf{r}) = \int_{\Delta_t} c_n(\theta, \phi, \mathbf{r}, t) dt. \quad (6)$$

Each color component of the omnidirectional image is $\mathcal{I}_n : \mathbb{S}^2 \times \mathbb{R}^3 \rightarrow \mathbb{R}$, which models the intensity of the n^{th} chromatic component as a function of \mathcal{D} and \mathbf{r} . Once the acquisition point \mathbf{r} is defined, \mathcal{I}_n is dependent on \mathcal{D} only. In that case, $\mathcal{I}(\theta, \phi)$ defines a panoramic (omnidirectional) image that satisfies the constraints of monocular vision.

2.5 Omnidirectional (monoscopic) image

Spatial multiplexing is normally used to sample each chromatic component \mathcal{I}_n of \mathcal{I} to produce an omnidirectional (monoscopic) image (\mathbf{I}). This is done by using color filter arrays (CFA) such as the widely used Bayer filter and other well known spatial multiplexing techniques. The sampling of each chromatic component of \mathcal{I} for a given \mathbf{r} leads to a panoramic image \mathbf{I} , which can be represented in any of the canonical panoramic formats; e.g., cylindrical, cubical.

Moreover, a partial FOV of the scene can be described by the same model. For instance, an image \mathbf{im} is obtained by acquiring \mathbf{C} at \mathbf{r} , for \mathcal{D} restricted to $(\Delta_\theta, \Delta_\phi)$ around \mathcal{D}_o . This image is $\mathbf{im} = (\mathbf{im}_1, \mathbf{im}_2, \mathbf{im}_3)$, where $\mathbf{im}_n : \mathbb{Z}^2 \rightarrow \mathbb{Z}$ after sampling and quantizing a subset of \mathcal{I} . Each array \mathbf{im}_n contains a two-dimensional array of color samples of the n^{th} color component. This is illustrated in Fig. 1 showing the generation of an image \mathbf{im} by acquiring the function \mathbf{C} by a sensor with a color space \mathcal{C} , for a partial FOV $(\Delta_\theta, \Delta_\phi)$ around a gaze direction \mathcal{D}_o and for an exposure time Δ_t .

The advent of panoramic photography in recent years has enabled the direct acquisition of \mathbf{C} , or at least its approximation. A direct acquisition of \mathbf{C} can be done by catadioptric cameras by means

of which panoramic images \mathbf{I} can be acquired in a single snapshot. Alternatively, multiple images from distinct viewpoints can be used to compose an approximation of \mathbf{I} [4].

A lot of attention has been devoted to correcting visual artifacts caused by stitching and blending partially overlapped images, attributing to these types of visual artifacts the source of most (if not all) the visual discomfort. However, the amount of geometric distortion in a monoscopic panorama that is actually perceived, which hinders the quality of the visual experience, is still undetermined.

3 Omnistereoscopic model

The *omnistereoscopic function* (\mathbf{C}_S) models the visual signal to be acquired to reconstruct a continuous omnistereoscopic image (\mathcal{I}_S) and, after being spatially sampled, a discrete stereoscopic panorama (\mathbf{I}_S) of the scene. A depiction of the omnistereoscopic model is presented in Fig. 2.

Differently than the acquisition of \mathbf{C}_S , an image \mathbf{I}_S must provide binocular visual cues in every \mathcal{D} around the acquisition point. The same requirement must be satisfied by any partial FOV image $\mathbf{im}_S = (\mathbf{im}_L, \mathbf{im}_R)$.

The *omnistereoscopic function* can be defined by a pair of *viewing functions*, \mathbf{C}_L and \mathbf{C}_R , which are respectively defined for $\mathbf{r}_L = \mathbf{O}_L$ and $\mathbf{r}_R = \mathbf{O}_R$, respectively. This function is

$$\mathbf{C}_S = (\mathbf{C}_L(\mathbf{u}_L), \mathbf{C}_R(\mathbf{u}_R))^T, \quad (7)$$

where \mathbf{C}_L and \mathbf{C}_R are respectively left and right *viewing functions* whose input parameters are $\mathbf{u}_j = (\theta_j, \phi_j, \mathbf{r}_j, t)$ with $j \in \{L, R\}$.

The viewpoints \mathbf{r}_L and \mathbf{r}_R exhibit horizontal parallax with respect to a reference \mathbf{r} . Even though the subindex $j \in \{L, R\}$ refers to the left or right viewpoints, this classification depends on the viewer locations with respect to the camera coordinate frame.

3.1 Omnistereoscopic viewing function

The function $\mathbf{C}_S : \mathbb{S}^2 \times \mathbb{S} \times \mathbb{R}^3 \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}^6$, models the three tristimulus values from two stereoscopic viewpoints with horizontal parallax as a vector quantity

$$\mathbf{C}_S(\mathbf{u}_S) = (c_{L,1}, c_{L,2}, c_{L,3}, c_{R,1}, c_{R,2}, c_{R,3})^T, \quad (8)$$

where the input parameters for \mathbf{C}_S are

$$\mathbf{u}_S = (\theta, \phi, \omega, \mathbf{r}, b, t). \quad (9)$$

The input parameters \mathbf{u}_L and \mathbf{u}_R can be derived from \mathbf{u}_S based on the geometric constraints of the model.

The function \mathbf{C}_S is defined for $\theta \in [0^\circ, 360^\circ)$ and for $\phi \in (0^\circ, 180^\circ)$ and it is undefined in the *nadir* and *zenith* gaze directions.

3.2 Stereoscopic viewpoints

The stereoscopic viewpoints \mathbf{r}_L and \mathbf{r}_R and the viewpoint \mathbf{r} are located in the $\mathbf{X}'\mathbf{Z}'$ -plane, which is parallel to the reference $\mathbf{X}\mathbf{Z}$ -plane (floor-plane), as illustrated in Fig. 2.

A baseline vector is defined by $\mathbf{b} = \mathbf{r}_L - \mathbf{r}_R$. By definition, \mathbf{b} is on the $\mathbf{X}'\mathbf{Z}'$ -plane. The baseline, defined by the Euclidean distance between stereoscopic viewpoints, is denoted $b = \|\mathbf{b}\|$. Finally, the

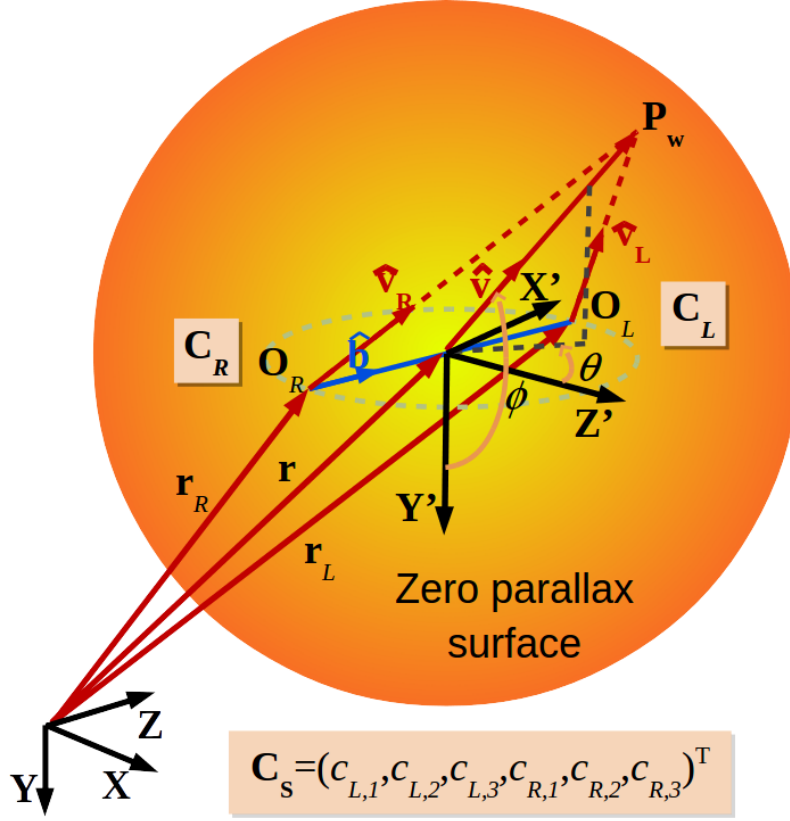


Figure 2: Omnistereoscopic model.

direction between right and left viewpoints is given by the unit vector

$$\hat{\mathbf{b}} = \frac{\mathbf{r}_L - \mathbf{r}_R}{\|\mathbf{r}_L - \mathbf{r}_R\|}. \quad (10)$$

In this model, the direction of $\hat{\mathbf{b}}$ is always perpendicular to $\hat{\mathbf{v}}$ for any panning angle. This constraint determines that, for each point \mathbf{P}_w , a different location for \mathbf{r}_L and \mathbf{r}_R is uniquely defined. Furthermore, the locations of these stereoscopic viewing points are defined uniquely by the panning direction in azimuth.

The perpendicularity between $\hat{\mathbf{b}}$ and $\hat{\mathbf{v}}$ guarantees that the scene will be acquired from two distinct viewpoints with a constant (horizontal) parallax b independently of the gaze direction.

3.3 Gaze direction

The *gaze vector* \mathbf{v} that connects the scene point \mathbf{P}_w with the reference viewpoint \mathbf{r} is defined by

$$\mathbf{v} = \mathbf{P}_w - \mathbf{r}. \quad (11)$$

The magnitude $v = \|\mathbf{v}\|$ defines the fixation distance. This parameter will be used to determine the relative skew ω or *vergence angle* between the left and right light rays. The unit vector is defined as

$$\hat{\mathbf{v}} = \frac{\mathbf{v}}{\|\mathbf{v}\|}, \quad (12)$$

determines the gaze direction of the left $\mathcal{D}_L = (\theta_L, \phi_L)$ and right $\mathcal{D}_R = (\theta_R, \phi_R)$ viewpoints (Fig. 2). For any \mathbf{P}_W and \mathbf{r} , there is a unique location for \mathbf{r}_L and \mathbf{r}_R which is determined by the parallax between viewpoints, or more specifically, the baseline distance b .

3.4 Fixation point

Any point \mathbf{P}_W in the scene determines a *fixation point*. This point is defined by the intersection between rays between \mathbf{P}_W and \mathbf{r}_L , which determines \mathbf{v}_L , and between \mathbf{P}_W and \mathbf{r}_R , which defines \mathbf{v}_R . The set of all possible fixation points at the same distance v defines a sphere centered at \mathbf{r} referred to as the zero parallax sphere. This is the spherical region depicted in Fig. 2.

The fixation distance v determines two dependent gaze directions $\mathcal{D}_L = (\theta_L, \phi_L)$ and $\mathcal{D}_R = (\theta_R, \phi_R)$ defined by

$$\hat{\mathbf{v}}_L = \frac{\mathbf{r}_L - \mathbf{P}_W}{\|\mathbf{r}_L - \mathbf{P}_W\|}, \quad (13)$$

and

$$\hat{\mathbf{v}}_R = \frac{\mathbf{r}_R - \mathbf{P}_W}{\|\mathbf{r}_R - \mathbf{P}_W\|}. \quad (14)$$

The gaze directions \mathcal{D}_L and \mathcal{D}_R are measured with respect to systems of coordinates $\mathbf{X}_L\mathbf{Y}_L\mathbf{Z}_L$ and $\mathbf{X}_R\mathbf{Y}_R\mathbf{Z}_R$, centered at \mathbf{r}_L and \mathbf{r}_R , respectively. These local references of coordinates can be obtained from \mathbf{XYZ} by applying a coordinate transformation consisting of a translation $\pm\mathbf{b}/2$ and rotations defined by each viewpoint gaze direction.

3.5 Vergence

In this model, the vergence is defined by the difference in azimuthal gaze directions of each viewing function. The perpendicularity constraint determines the relationship between \mathcal{D}_L , \mathcal{D}_R and the global gaze direction \mathcal{D} . The vergence angle ω determines the gaze directions \mathcal{D}_L and \mathcal{D}_R as follows

$$\begin{aligned} \mathcal{D}_L &= (\theta - \omega, \phi), \\ \mathcal{D}_R &= (\theta + \omega, \phi). \end{aligned} \quad (15)$$

The trigonometric relationship between the panning angles θ_L and θ_R given the vergence point \mathbf{P}_W is illustrated in Fig. 3-(a). The vergence angle is $\omega = \arctan\left(\frac{b}{2v}\right)$, where v is the vergence distance to the scene as shown by the geometric relationships in Fig. 3-(b). When the fixation distance is $v \gg b$, $\omega \simeq 0^\circ$ and $\theta_L = \theta_R = \theta$. The latter is equivalent to a stereoscopic camera with parallel optical axes.

The case of a point in the scene \mathbf{P}_W located at a distance $\|\mathbf{v}\| = v$ from the reference viewpoint \mathbf{r} is illustrated in Fig. 4-(a). According to the omnistereoscopic model, this situation determines a vergence angle ω_1 and the location of the stereoscopic viewpoints $(\mathbf{r}_{L,1}, \mathbf{r}_{R,1})$. In Fig. 4-(b), a different point \mathbf{P}_W is chosen, which is at the same distance from the reference point as before but in another gaze direction in azimuth θ_2 . The omnistereoscopic model determines that $\omega_2 = \omega_1$, but different stereoscopic viewpoints $(\mathbf{r}_{L,2}, \mathbf{r}_{R,2})$ are determined by this new gaze direction. Finally, in Fig. 4-(c) is shown the case of a point sufficiently far from the sampling point ($v \gg b$). In this final

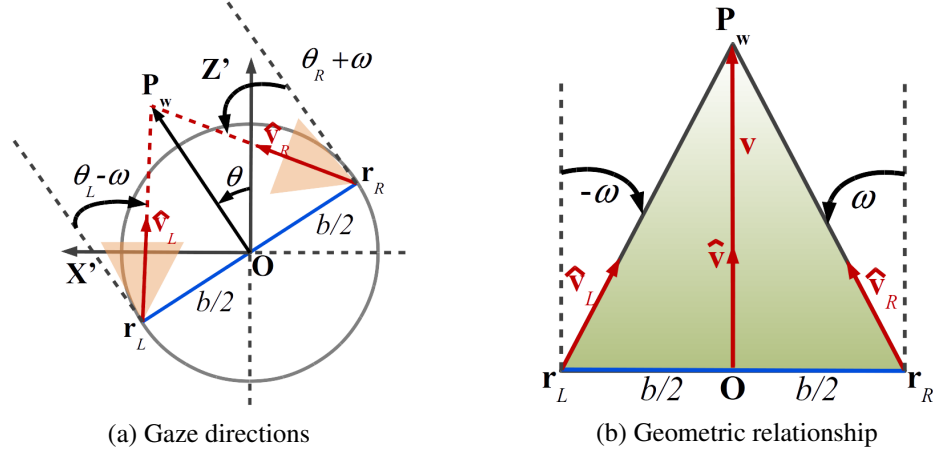


Figure 3: Geometric relationship between the gaze direction and location of the stereoscopic view-points.

example, the gaze directions for left and right viewpoints are parallel, $\hat{\mathbf{v}}_L \parallel \hat{\mathbf{v}}_R$, and the vergence angle $\omega = 0^\circ$.

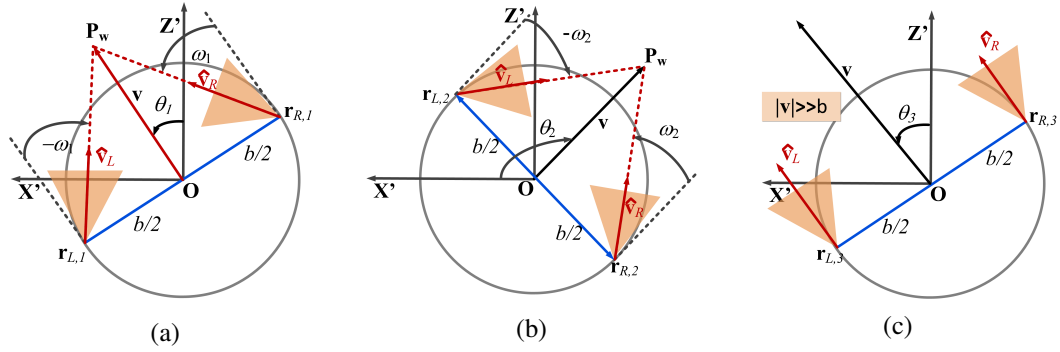


Figure 4: The perpendicularity constraint between $\hat{\mathbf{b}}$ and $\hat{\mathbf{v}}$ defines the locations of the stereoscopic viewing points \mathbf{r}_L and \mathbf{r}_R : (a) and (b) illustrate the case of change in the stereoscopic viewpoints for two scene points at the same distance from the reference point but in two different gaze directions in azimuth, and the example (c) shows the case of a point in the scene sufficiently far from the reference point.

3.6 Left and right viewing functions

The *viewing functions* for the left and right-eye viewpoints are given by

$$\mathbf{C}_j(\mathbf{u}_j) = (c_{j,1}(\mathbf{u}_j), c_{j,2}(\mathbf{u}_j), c_{j,3}(\mathbf{u}_j))^T, \quad (16)$$

where $c_{j,n}$, for $j = \{L, R\}$ and $n \in \{1, 2, 3\}$, are the tristimulus values. The input vectors are defined by

$$\mathbf{u}_L = (\mathcal{D}_L, \mathbf{r}_L, t) \quad (17)$$

$$= (\boldsymbol{\theta} - \boldsymbol{\omega}, \phi, \mathbf{r}_L, t), \quad (18)$$

$$\mathbf{u}_R = (\mathcal{D}_R, \mathbf{r}_R, t) \quad (19)$$

$$= (\boldsymbol{\theta} + \boldsymbol{\omega}, \phi, \mathbf{r}_R, t). \quad (20)$$

3.7 Viewing circle

The reference viewpoint \mathbf{r} is defined in this model by the midpoint between \mathbf{r}_L and \mathbf{r}_R , in vector expression

$$\mathbf{r} = \mathbf{r}_L - \frac{b}{2} \hat{\mathbf{b}} \quad (21)$$

$$= \mathbf{r}_R + \frac{b}{2} \hat{\mathbf{b}}. \quad (22)$$

The possible location of all the stereoscopic viewing points defines a circle centered at \mathbf{r} , whose diameter is the stereoscopic baseline b . This circle is the *viewing circle* which is depicted in Fig. 2.

In the case when the user's head has total freedom of movement, e.g., in the case of a zero-gravity environment, a more general omnistereoscopic viewing model would eliminate the restriction of having $(\mathbf{r}_L, \mathbf{r}_R)$ on the $\mathbf{X}'\mathbf{Z}'$ -plane. In that situation, the set of all possible locations for the stereoscopic viewpoints would fall on a spherical surface with center at \mathbf{r} and diameter b that would define a *viewing sphere* instead of a *viewing circle*. However, for this paper, limiting the possible locations of $(\mathbf{r}_L, \mathbf{r}_R)$ to a reference plane whose normal is the vertical axis defined by the user's head orientation (perpendicular to the floor plane) is a reasonable assumption.

The geometric constraints introduced so far reduce the necessary information to characterize \mathbf{C}_S to only a few parameters: a reference viewpoint \mathbf{r} , a gaze direction \mathcal{D} , a time variable t a baseline b and a vergence $\boldsymbol{\omega}$. The gaze directions defined for \mathbf{C}_L and \mathbf{C}_R can be calculated from these parameters.

3.8 Omnistereoscopic image

As in the monoscopic case, an omnistereoscopic image \mathcal{I}_S is created by acquiring the function \mathbf{C}_S in a point \mathbf{r} , for a given a color space \mathcal{C} and a defined exposure Δ_t . For a given \mathbf{r} , $\mathcal{I}_S : \mathbb{S}^2 \rightarrow \mathbb{R}^6$ is defined by

$$\mathcal{I}_S(\boldsymbol{\theta}, \phi, \mathbf{r}) = \left(\int_{\Delta_t} \mathbf{C}_L(\boldsymbol{\theta}, \phi, \mathbf{r}, t) dt, \int_{\Delta_t} \mathbf{C}_R(\boldsymbol{\theta}, \phi, \mathbf{r}, t) dt \right), \quad (23)$$

$$= (\mathcal{I}_{L,1}, \mathcal{I}_{L,2}, \mathcal{I}_{L,3}, \mathcal{I}_{R,1}, \mathcal{I}_{R,2}, \mathcal{I}_{R,3}), \quad (24)$$

where $\mathcal{I}_{j,n} : \mathbb{S}^2 \rightarrow \mathbb{R}$ is the n^{th} color intensity function for $j \in \{L, R\}$ and for the acquisition point \mathbf{r} . This direction dependent color function is

$$\mathcal{I}_{j,n}(\boldsymbol{\theta}, \phi, \mathbf{r}) = \int_{\Delta_t} c_n(\boldsymbol{\theta}, \phi, \mathbf{r}_j, t) dt. \quad (25)$$

Furthermore, for a defined acquisition point $\mathbf{r} = \mathbf{r}_o$, the omnistereoscopic image is a function of the gaze direction only. In this case, each color component of the omnidirectional image is $\mathcal{I}_{j,n} : \mathbb{S}^2 \rightarrow \mathbb{R}^3$.

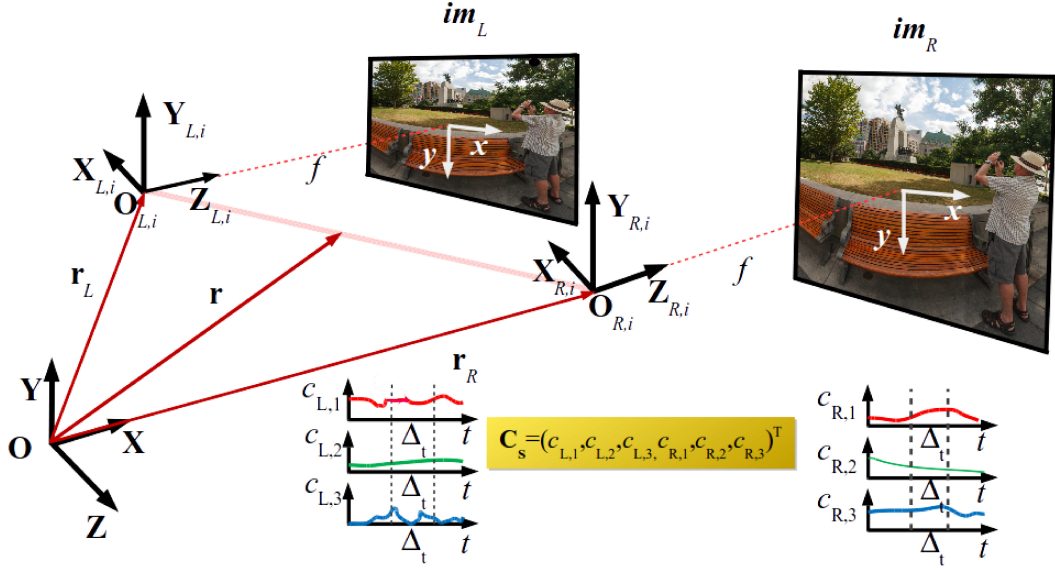


Figure 5: \mathbf{C}_S and the acquisition of a partial stereoscopic view of the scene.

The spatial sampling of \mathcal{S}_S for 360° in azimuth and a wide but restricted elevation FOV determines a discrete omnistereoscopic image $\mathbf{I}_S = (\mathbf{I}_L, \mathbf{I}_R)$, where each panoramic image is defined by $\mathbf{I}_j = (\mathbf{I}_{j,1}, \mathbf{I}_{j,2}, \mathbf{I}_{j,3})$. These images are topologically equivalent to a cylinder.

Instead of acquiring the complete omnistereoscopic function at once for an acquisition point \mathbf{r} , we may be interested in acquiring a wide-angle stereoscopic image in the direction \mathcal{D}_o and for a limited FOV $(\Delta_\theta, \Delta_\phi)$. In that case, after sampling and quantizing a subset of \mathcal{S}_S , the stereoscopic image $\mathbf{im}_S = (\mathbf{im}_L, \mathbf{im}_R)$, where $\mathbf{im}_j : \mathbb{Z}^2 \rightarrow \mathbb{Z}$, can be produced. This is illustrated in Fig. 5.

4 Acquisition

The function \mathbf{C}_S can be acquired by capturing sets of monoscopic or stereoscopic images covering the whole scene in azimuth. One of the earlier approaches relies on the simultaneous acquisition of two cylindrical panoramas using a pair of catadioptric cameras co-axially aligned [29, 30]. This method produces two single-viewpoint panoramas where the vertical disparities provides depth information of the scene. However, this approach is not suitable to acquire \mathbf{C}_S since it does not correspond to the geometric constraints given by our omnistereoscopic model. Fortunately, other omnistereoscopic techniques are well suited to acquire, or at least, to approximate the value of \mathbf{C}_S according to the constraints of our model [4].

Next, we present an acquisition model that can represent the different omnistereoscopic techniques to produce stereoscopic panoramas based on horizontal parallax.

4.1 Acquisition model

An acquisition strategy based on two pinhole cameras with horizontal parallax is suitable to sample \mathbf{C}_S around a reference viewpoint \mathbf{O} for arbitrary gaze directions. Furthermore, this acquisition

model [5] can produce direction-dependent stereoscopic imagery that conveys correct binocular cues of depth at the center and peripheral region of each image.

The acquisition 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. The location and orientation of the camera pair is restricted by the need to capture two snapshots from a pair of viewpoints with horizontal parallax. One constraint is that all the possible locations for this pair of pinhole cameras are restricted to the horizontal \mathbf{XZ} -plane, which we denoted as the reference *horizontal* plane. Another constraint is that the optical axes of both pinhole cameras are parallel and lay on the \mathbf{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 \mathbf{Y} -axis. The reference point \mathbf{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 \mathbf{XZ} -plane.

4.2 Acquisition configurations

The relative location of the projection centers for left \mathbf{O}_L and right \mathbf{O}_R cameras with respect to the symmetry center \mathbf{O} determines the four basic acquisition configurations. These basic configurations can describe all the panoramic cameras capable of producing omnidirectional stereo with horizontal parallax. We have demonstrated this point in our review paper on omnidirectional stereo [6] where we showed with examples that cameras and methods to produce stereoscopic panoramas fall in one or another of these four basic acquisition configurations.

The simplest arrangement of cameras is *configuration 1* (Fig. 6-(a)), which can be used in sequential acquisition of stereoscopic images with horizontal parallax. It cannot be used for the simultaneous acquisition of multiple stereoscopic samples of \mathbf{C}_S due to self-occlusion between cameras.

The *configuration 2* (Fig. 6-(b)) is a variation of *configuration 1*, which consists in making \mathbf{O} coincident with either \mathbf{O}_L or \mathbf{O}_R projection centers. This configuration enables the sampling of one *viewing function* from a singular viewpoint (\mathbf{O}), while a stereoscopic view for each \mathcal{D} is acquired by the off-centered image sensor. The result is an approximation of \mathbf{C}_S from closely located viewpoints [10]. This model is suitable only for sequential acquisition due to the limitation of spatially collocating multiple cameras at the same projection center. This constructive problem is addressed in *configurations 3* and *4*.

Finally, *configuration 3* (Fig. 6-(c)) is a variation of the *configuration 2* where the stereoscopic pair of cameras is radially displaced from \mathbf{O} by a distance $|\mathbf{r}_c|$, but where their projection centers \mathbf{O}_j remain in the \mathbf{XZ} -plane.

The *configuration 4* (Fig. 6-(d)) can be derived by defining a radial displacement $|\mathbf{r}_c|$ to the stereoscopic rig in *configuration 1* (Fig. 6). The radial displacement addresses the physical limitation that the cameras in a multiple-camera configuration cannot have spatially collocated projection centers. Then, *configurations 3* and *4* can be implemented with multiple cameras in a simultaneous acquisition scheme. This configuration models the camera rig presented by Google [2] which uses sixteen cameras synchronously controlled to produce eight, partially overlapped, stereoscopic video streams; Samsung also proposed a camera that can be modeled by the same configuration [3].

4.3 Dense sampling

A narrow-width column at the center of each \mathbf{im}_j can be acquired and mosaicked to produce \mathbf{I}_S . However, large number of samples must be acquired sequentially by rotating a stereoscopic rig in small increments of θ . This acquisition method models a technique based on line cameras, which is based on sequentially sampling \mathbf{C}_S , collecting stereoscopic pairs of image columns, sometimes, one

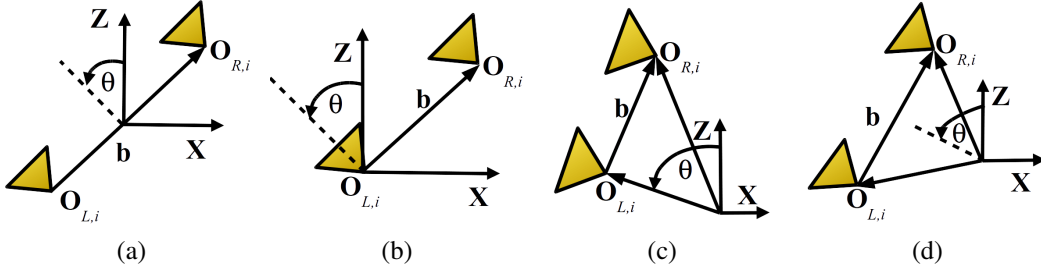


Figure 6: Variations of the acquisition model: configuration 1 (a), 2 (b), 3 (c), and 4 (d).

view (left or right-eye) at a time, by changing \mathcal{D} in small increments of azimuth. For instance, the acquisition model presented by Huang et al. [23] can be modeled by *configuration 1*, when a single central column of \mathbf{im}_j is collected for a given vergence (camera toe-in) and baseline.

Alternatively, *configuration 1* can model the acquisition approach presented by Peleg et al. [9] where a pair of image columns is obtained from a single camera rotated off-center. These image columns are back-projected according to two virtual projection centers defined on a *viewing circle* centered in \mathbf{O} and whose diameter is b . The narrow width of the extracted image columns implies the acquisition of a dense set to be able to reconstruct \mathcal{I}_S . Additionally, the scene must be static.

In the previous two acquisition examples, the produced \mathbf{I}_S is a distorted approximation of the stereoscopic image produced by acquiring \mathbf{im}_S , in any direction \mathcal{D} , by means of wide-angle FOV pair of cameras. This is particularly noticeable when \mathbf{I}_S is projected on curved displays where the binocular depth cues in any gaze direction \mathcal{D} would be correct, but the perceived depth when the gaze eccentricity increases is distorted [25].

An example of dense acquisition that more closely approximates the acquisition model in *configuration 1* is the rotating stereoscopic rig used in the Mars Pathfinder probe [24]. This setup can be used in an interactive scene exploration. However, the off-line stereoscopic navigation would require the storage of an unnecessarily large set of stereoscopic images, each of which would be a stereoscopic snapshot corresponding to a slightly different gaze directions in azimuth. Fortunately, a sparser sampling of \mathbf{C}_S can also produce a pleasant and consistent depth illusion.

Any of the four configuration presented in Section 4.2 can be used for the acquisition of a dense set of samples of the \mathbf{C}_S function.

4.4 Sparse sampling

A limited number of partially overlapped stereoscopic snapshots can be used to collect a sparse set of samples of \mathbf{C}_S . The mosaicking of these stereoscopic samples can be used to produce a satisfactory illusion of depth based on horizontal disparities around the reference viewpoint [16, 12]. The horizontal disparity between mosaicked images must be consistent and the vertical disparities over the blending regions between mosaicked images must be corrected to avoid artifacts and visual discomfort [5].

The sparse collection of stereoscopic samples of \mathbf{C}_S can be done with any of the acquisition configurations presented before [10, 12]. In particular, *configurations 3* [16] and *4* [13, 15, 14] enable the simultaneous acquisition of a sparse set of samples of the \mathbf{C}_S function. This set is sufficient to reconstruct an approximation of the ideal \mathbf{I}_S . The number of stereoscopic samples necessary is related to the distance between the reference viewpoint and the scene, which also affects the magnitude and direction of vertical disparities to correct [5]. The simultaneous acquisition using sparse

stereoscopic samples is attractive since it enables the acquisition of dynamic scenes which was not possible by using sequential acquisition methods.

The acquisition of a reduced number N of stereoscopic images, which corresponds to incremental values of $\theta \in \{\theta_0, \dots, \theta_{N-1}\}$, raises the question of how many stereoscopic images are really necessary to correctly convey a consistent illusion of depth around a reference viewpoint. The produced \mathbf{I}_S , or more specifically, the stereoscopic images \mathbf{im}_S generated are faithful renditions of the scene for a limited number of directions in azimuth, i.e., when $\theta = \theta_i$ ($i \in \{0, \dots, N-1\}$). In a mosaicking scenario, any intermediate view is a mosaic of two stereoscopic samples in the acquired set, but it is not a sample of \mathbf{C}_S in that particular gaze direction. Fortunately, if the horizontal disparities are kept continuous and consistent in mosaic boundaries, a satisfactory and compelling illusion of depth can be delivered by this acquisition and rendering approach [5].

5 Experimental acquisition and rendering

In this section, we illustrate the use of sparse sets of stereoscopic images to approximate the sampling of \mathbf{C}_S to produce an omnistereoscopic image \mathbf{I}_S . The examples provided in this section have the sole purpose of illustrating the sparse set concept (Section 1). The curious reader may find interesting the navigable versions of these examples available in [7] as red-cyan anaglyphs.

In our experiments, we wanted to test the sparse acquisition and rendering approach using the different configurations introduced in Section 4.2. Our experimental set-up consisted in a single or dual DSLR camera (Canon Rebel XTi 400D), which was mounted in portrait orientation on a panoramic head (Manfrotto 303SPH). We used a *fish-eye* lens Bower/Samyang model SLY-358C. As the reader may notice, we did not attempt to create a functional camera for the real-time production of omnistereoscopic imagery; hence, we do not report comparative results with other methods, which is beyond the scope of this paper. Similarly, we do not make any claim in terms of computational complexity of our rendering technique since it is simply illustrative of how we rendered each \mathbf{I}_S .

Each acquisition configuration determines a different spatial distribution of projective centers ($\mathbf{O}_{L,i}, \mathbf{O}_{R,i}$). All acquisition configurations were emulated by manually positioning the nodal point of the camera at predetermined locations defined by each variation of the acquisition model.

The omnistereoscopic viewing function \mathbf{C}_S was sparsely sampled by acquiring $N = 6$ stereoscopic images ($\mathbf{im}_{L,i}, \mathbf{im}_{R,i}$), with $i = \{1, \dots, 6\}$, used to emulate the four *configurations* illustrated in Fig. 6. An example of this approach is depicted in Fig. 7 and the six stereoscopic snapshots acquired are presented in Fig. 8.

5.1 The acquisition system parameters

The camera exposure and aperture were controlled manually based on the requirements of each scenario. For instance, the *f-stop*, which indicates the size of the diaphragm aperture, was adjusted to provide *depth-of-field* long enough to focus all the elements of the scene. Similarly, the exposure was manually selected according to the chosen f-stop to provide a balanced illumination on the stereoscopic image set. The challenge was to keep a balanced exposure between poorly illuminated and over-exposed directions of the scene.

Both exposure and f-stop were adjusted for each new viewpoint and kept unchanged during the acquisition of N stereoscopic samples, making the rendering by mosaicking easier since less complex exposure and color equalization were needed. Note that, in a real multiple-camera implementation, the exposure and aperture are automatically adjusted for each individual camera, which requires the implementation of blending algorithms to smooth the transitions between mosaicked

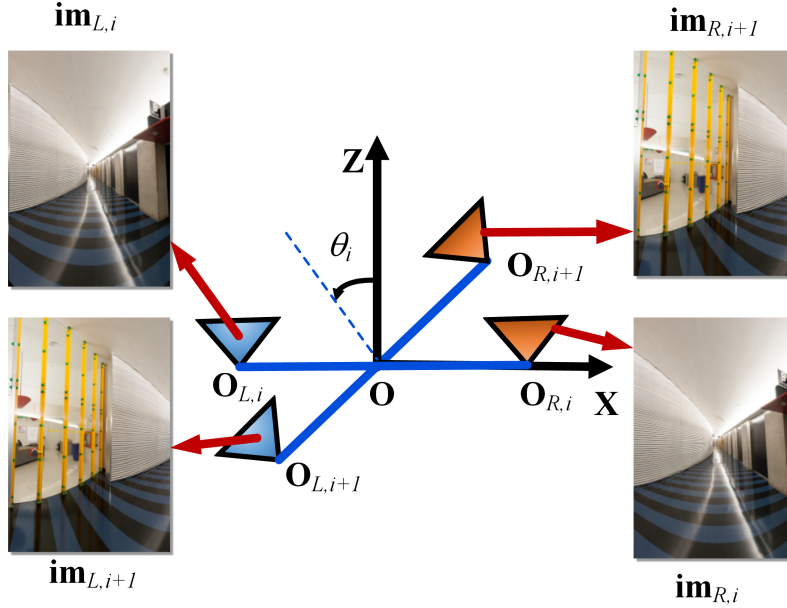


Figure 7: Example of acquisition of two adjacent stereoscopic snapshots of the scene using *configuration 1*: the acquisition of the stereoscopic pair of images $\mathbf{im}_{j,i}$ for θ_i and $\mathbf{im}_{j,i+1}$ where the angular increment between stereoscopic samples is $\Delta\theta = \theta_{i+1} - \theta_i$.

views. The exposure smoothing and color correction algorithms were implemented and tested in later experiments.

The baseline used in these experiments was $b = 32.5$ mm or half the mean inter-ocular distance within the adult human population. A reduced baseline is convenient for mosaicking the images since less parallax between projective centers is introduced. A larger baseline implies more parallax between projection centers and more complex rendering techniques are needed to reduce the visible ghosting after stitching neighbor samples. The acquisition parameters used in all the experiments presented in this section are detailed in Table 1.

For large baselines, the acquisition *configurations 2* and *3* have advantage over other configurations. For instance, in *configuration 2* one camera is rotated around its nodal point, facilitating a seamless mosaicking of a monoscopic panorama while a second camera exploits a large baseline to improve a depth map estimation. In acquisition *configuration 3*, the same idea is implemented maintaining an off-center configuration r_c .

5.2 Rendering implementation

The sets of stereoscopic images were initially stored in the camera's native raw file format in a portable memory unit. A posteriori, these sets of images were manually transferred to a computer hard drive along with additional information to help the necessary post-processing, such as additional images for the color correction.

The goal was to test the viability of the proposed acquisition configurations, so we used commercially available tools when suitable to implements parts of the rendering process. For instance, the lens distortion and color correction were both done using *Adobe Lightroom 4.4* [31]. In the case



Figure 8: Source images $\mathbf{im}_{j,i}$, where $j = L$ (top row) and $j = R$ (bottom row) and $i \in \{0, \dots, 5\}$ (from left to right).

Table 1: Acquisition parameters: Canon Rebel XTi 400D, Bower fish-eye lens (SLY-358C).

Parameters	Value
Sensor width (W_h)	22.2 mm
Aspect ratio (a_r)	1.5
focal length (f)	9.3 mm
pixel width (s)	$5.6 \mu\text{m}$ (3888×2592 pixels)
Horizontal FOV Δ_a	100°
Baseline (b)	32.5 mm
Radial off-set (r_c)	32.5 mm
Vergence (ω)	0°

of testing the 2D-to-3D image synthesis, part of the image warping based on the depth-map was implemented using the toolbox for image warping effects of *GIMP* [32]. The rest of the mosaicking blocks and additional processing were implemented using *Matlab* [33], with the *Computer Vision System* and *Image processing* toolboxes. Finally, the anaglyph version of the stereoscopic images presented in this paper were done using the *StereoPhotoMaker* authoring tool [34].

The omnistereoscopic acquisition and rendering can be summarized in the following sequential blocks:

1. Lens distortion and white balance correction is applied to each $\mathbf{im}_{j,i}$, independently.
2. Stereoscopic registration is applied to each $\mathbf{im}_{j,i}$ to reduce or eliminate undesired vertical disparities.
3. Cylindrical projection and image alignment between $\mathbf{im}_{j,i}$ and $\mathbf{im}_{j,i+1}$ is applied to the left and right image sets.

4. The optimal cut between consecutive image samples is calculated.
5. The stitching and blending of is applied to $\mathbf{im}_{j,i}$ obtain cylindrical panoramas for the left (\mathbf{I}_L) and the right (\mathbf{I}_R) views.

Each of this rendering blocks from our implementation are detailed next.

5.3 Lens distortion correction

The acquired images, denoted as $\mathbf{im}_{j,i}$, were radially distorted as a result of using aspherical lenses in the acquisition. The correction was implemented using *Adobe Lightroom 4.4's* lens correction tool. The calibration information for Samyang SLY-358C is available on-line. An example showing the acquired raw image and the image after correction is shown in Fig. 9.



Figure 9: Lens correction.

5.4 White balance correction

The next step is to correct white balance of the planar set of images $\mathbf{im}_{j,i}^p$. This processing was done manually using *Adobe Lightroom 4.4* [31] and the *Color Checker Passport* [35]. The latter provides a color calibration pattern to be used as color reference. In a multiple camera system, white balancing is set at the beginning of each acquisition for all the cameras to ensure the a posteriori correct color of the omnidirectional scene.

5.5 Stereo rectification

The stereo rectification of the pair of undistorted images $\mathbf{im}_{j,i}^p$ is necessary to correct any deviation from the ideal camera model with parallel optical axes. The uncalibrated stereo rectification used is based on the estimation of a set of corresponding points \mathcal{X}_i on $\mathbf{im}_{j,i}^p$ to define an affine transformation that, after being applied to the stereoscopic pair, eliminates the vertical disparities and make the

conjugate epipolar lines parallel to the horizontal image axis. This uncalibrated stereoscopic registration method was implemented on *Matlab* using the routines from the *Computer Vision System* toolbox [36].

The implementation uses the *speeded up robust features* (SURF) algorithm [37, 38] to identify a set of corresponding point features for each stereoscopic pair. The recursive process eliminates feature points from the set whose sum of absolute distances are above the chosen distance threshold. On each recursion of the algorithm, the feature points \mathcal{K}_i must satisfy the epipolar constraints. We used the *Matlab*'s implementation of the *random sample consensus* (RANSAC) algorithm [39, 40] to eliminate outlier candidate points using the epipolar constraint. On each iteration, a new projective matrix is estimated. After all outliers have been eliminated from the set \mathcal{K}_i , and if enough corresponding points have been identified, the affine transformation that minimize the vertical component of the disparity between corresponding points is calculated. The example presented in Fig. 10 shows the result of the stereoscopic rectification in the alignment before and after stitching and blending a pair of contiguously acquired images.

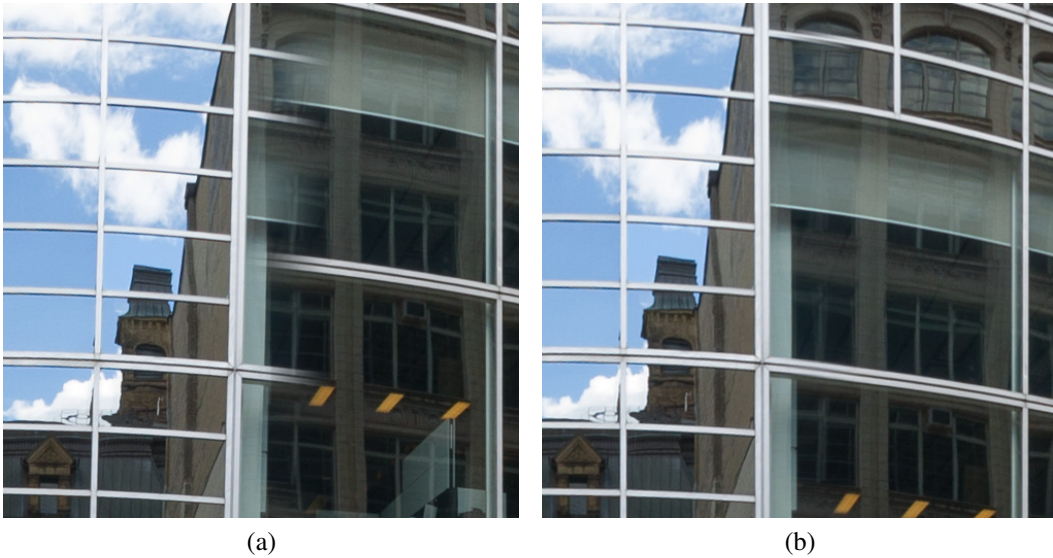


Figure 10: Stereoscopic registration: (a) before alignment and (b) after alignment.

In a few stereoscopic images of a given set, the automatic rectification could not be applied since not enough samples could be detected. The reasons for this is the insufficient illumination or the lack of contrast in the images prevent the automatic detection of enough corresponding features. In these few cases, the manual selection of a limited set of corresponding points was enough to implement a satisfactory rectification.

5.6 Color and luminance blending

A homogenization of the color and luminance of the $\mathbf{im}_{j,i}$ set needs to be done before blending to prevent visible transitions between mosaicked images. This is particularly necessary when each camera adjusts automatically the exposure time and aperture to optimize each shot. The method proposed by Xiong and Pulli [41] was used in these tests. Their method is based on calculating local compensation coefficients over the overlapped areas between $\mathbf{im}_{j,i}^p$ and $\mathbf{im}_{j,i+1}^p$. Local coefficients

are estimated using the average luminance value for each color channel in RGB color space after removing the gamma correction. The method also estimates a global compensation coefficient to reduce accumulative errors. This algorithm can be used when the luminance between neighbor images are close.

Once the individual images to mosaic have been processed to compensate for exposure discrepancies, a global histogram equalization was implemented and tested to minimize the exposure difference between the final \mathbf{I}_L and \mathbf{I}_R . This final equalization was applied at the end of the mosaicking process in cases where the left and right image sets were purposely acquired with different photographic parameters.

5.7 Cylindrical projection

In order to facilitate the alignment by applying a horizontal translation, the images were first projected onto a cylindrical canvas. The central symmetry of all the acquisition configurations suggests to project $\mathbf{im}_{j,i}^p$ onto a warped surface with central symmetry centered at \mathbf{O} . Since \mathbf{C}_S is sampled at different θ_i and for a limited FOV in elevation, a cylindrical canvas or any topologically equivalent surface is adequate to map the acquired samples of the omnistereoscopic viewing function [42]. Therefore, we implemented a mapping algorithm to project points from different $\mathbf{im}_{j,i}^p$ onto a cylindrical surface.

The cylindrical canvas, which is illustrated in Fig. 11, has parameters $\Delta\theta$ for the panning azimuth and h for the gaze height. The 2D mapping of a pixel with coordinates $\mathbf{p}_{j,i} = (x_{j,i}, y_{j,i})$ in $\mathbf{im}_{j,i}^p$ onto cylindrical coordinates $\mathbf{p}_{j,i}^c = (x_{j,i}^c, y_{j,i}^c)$ is given by

$$x_{j,i}^c = f \Delta\theta = f \arctan \frac{x_{j,i}}{f}, \quad (26)$$

$$y_{j,i}^c = f h = f \frac{y_{j,i}}{\sqrt{x_{j,i}^2 + f^2}}, \quad (27)$$

where the cylindrical surface radius was chosen to be f [43].

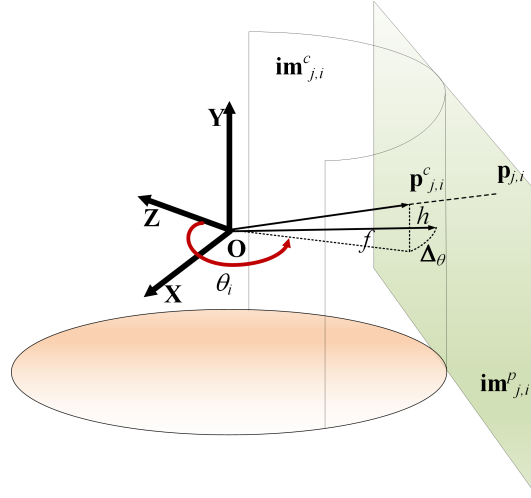


Figure 11: Mapping from a planar to a cylindrical surface.

Note that the cylindrical coordinates $\mathbf{p}_{j,i}^c = (x_{j,i}^c, y_{j,i}^c)$ are defined for the projection of a planar image $\mathbf{im}_{j,i}^p$ for θ_i . The angle in azimuth Δ_θ is defined as an angular deviation in azimuth from θ_i . For example, the cylindrical coordinate $\mathbf{p}_{j,i}^c = (0,0)$, which corresponds to cylindrical coordinate $\theta = \theta_i$ and $h = 0$.

5.8 Mosaic alignment

The rendering by mosaicking relies on mapping of a point coordinate $\mathbf{p}_{j,i}$ from $\mathbf{im}_{j,i}^p$ onto \mathbf{I}_j , for $j \in \{L, R\}$. In order to map samples of \mathbf{C}_S obtained from different viewpoints into an omnistereoscopic image pair $(\mathbf{I}_L, \mathbf{I}_R)$, it is necessary to define the transformation between corresponding points.

The corresponding point set detected on the overlapped regions between neighbor images were estimated using SURF and RANSAC algorithms. This set of points was used to define an affine transformation to align the points in the defined stitching coordinates. The criterion used to define this transformation was the sum of absolute differences [36] between disparities. For the ideal case where the projection centers are collocated and the panning is limited to the \mathbf{XZ} -plane, this aligning transformation is reduced to a horizontal translation [44].

5.9 Optimal cut

The optimal cut method relies on estimate $\mathcal{A}_{j,(i,i+1)}$ and obtaining the best stitching coordinate $(x_b, y_{j,i})$ over the region that would minimize the luminance difference between images. This strategy minimizes visual artifacts such as ghosting. This processing is applied independently to left and right images.

In order to do this, we implemented the algorithm proposed by Ha et al. [45] which relies on obtaining the locations of minimum intensity gradient between two images to blend. A scoring system enables to find the optimal cut line by line. This algorithm is attractive for its simplicity and can be easily parallelized for real-time applications such as video mosaics. An example of the image blending is presented in Fig. 12 to illustrate the effect of pyramidal blending using a fixed stitching coordinate x_b versus the optimal-cut plus linear blending method where the optimal x_b is dynamically selected line by line. This example in particular shows the effect of mosaicking when the scene changes between the consecutive snapshots.

5.10 Stitching

A point with the stitching coordinate $\mathbf{p}_{j,i}^p = (x_b, y_{j,i})$ from $\mathbf{im}_{j,i}^p$ is mapped into $\mathbf{p}_{j,i+1} = (x'_b, y_{j,i})$ in $\mathbf{im}_{j,i+1}^p$. The alignment enables to continue the rendering by mapping $\mathbf{im}_{j,i+1}^p$ into the cylinder following the same cylindrical projection, but for an angular deviation Δ_θ defined with respect to neighbor image, which corresponds to the sampling angle θ_{i+1} .

The projection onto a cylindrical canvas and stitching on neighbor images continues until the last image in the sequence ($\mathbf{im}_{j,N-1}^p$) is stitched with the first image ($\mathbf{im}_{j,0}^p$) and projected onto a cylinder. This process is done independently for the left and right images in the acquired sequence. The result of this process is a pair of cylindrical panoramas I_j , for $j \in \{L, R\}$. This stitching method maps of all $\mathbf{im}_{j,i}$ into a single cylindrical canvas using the aligning process to match neighbor images.

5.11 Blending

The correct blending between image pairs is necessary to eliminate visible ghosting in the final mosaicked panorama. The simplest approach and the one used in these experiments is *linear blending*

in which each color component is multiplied by a coefficient that weights its luminance contribution to the blend depending on its Euclidean distance to the stitching coordinate point x_b in the horizontal dimension [46]. Another possible approach is Gaussian pyramidal decomposition [47]. The latter method improves the blending by adding more processing, but it does not reduce ghosting if the stitching coordinate point x_b is not properly matched, e.g., the images may not be exactly aligned or the scene may have changed from sample to sample.

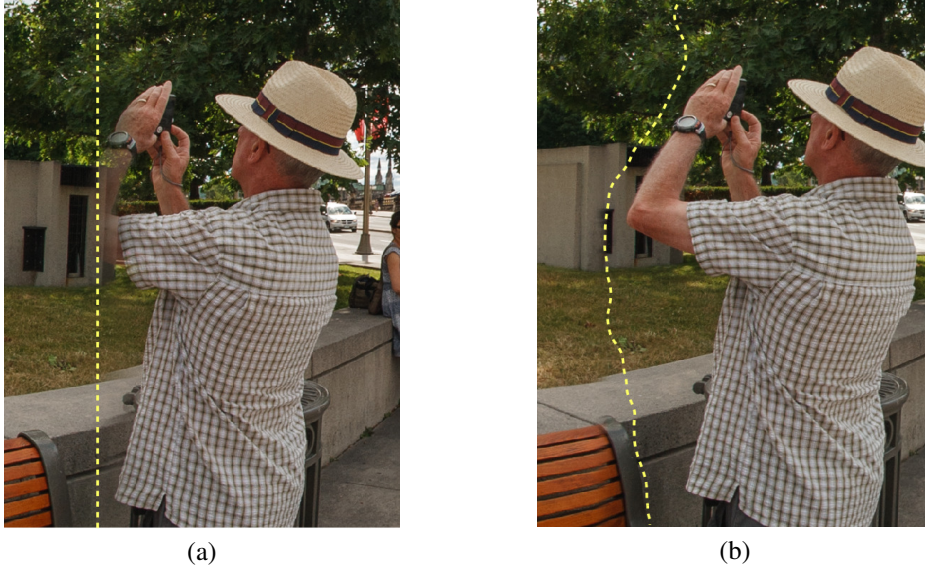


Figure 12: Image stitching and blending: (a) multi-band (pyramidal) blending with fixed x_b and (b) linear blending plus optimal-cut selection of x_b for each vertical coordinate.

6 Results

A sample of the \mathbf{I}_S produced by mosaicking the acquired six stereoscopic images is presented in Fig. 13 for *configurations 1-4*. In order to appreciate the effect of interactively exploring the scene changing the gaze direction around the acquisition point, we prepared a companion website [7] with interactive stereoscopic version of these four examples of \mathbf{I}_S using sparse sample sets.

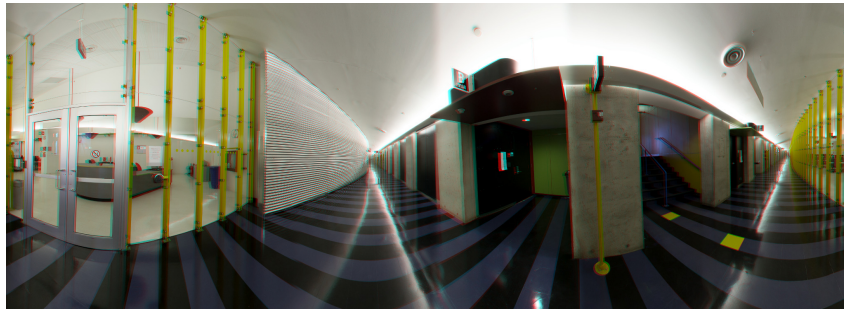
7 Discussion

By examining closer the presented \mathbf{I}_S and their interactive versions [7], the reader will find that the horizontal disparity continuity is maintained for all gaze directions in azimuth. Furthermore, vertical disparities can also be controlled avoiding visual strain. Both advantages can be derived from the geometrical constraints of the acquisition model in its various configurations [5]. However, by using a limited number of wide-angle snapshots to approximate \mathbf{C}_S , we are introducing horizontal disparity distortions on the edge between mosaicked images; the wider the individual sensor's FOV, the larger the distortion [26]. This effect can also be derived by the acquisition model presented

in Section 4 by noticing that the effective stereoscopic baseline \mathbf{b} is reduced with the Δ_θ angular shift from each sampling angle θ_i . But, in most cases this is not apparent since our perception of depth relies on other depth cues beyond binocular disparity; i.e., geometry of the scene, color and shadows. A more accurate sampling \mathbf{C}_S , as derived from our acquisition model, is to acquire narrow image columns where Δ_θ is minimized; this sampling method models other techniques based on line cameras [9, 23]. The number of snapshots necessary to evoke a satisfactory 3D experience for any gaze direction requires the study of viewers’ response to stimuli presented in \mathbf{I}_S ; we left the subjective evaluations open for future contributions. However, we presented an objective analysis of this problem in [5], which was based on cues of stereoscopic quality such as vertical disparities, horizontal disparity continuity between image seams and minimum distance to the camera.

8 Conclusions

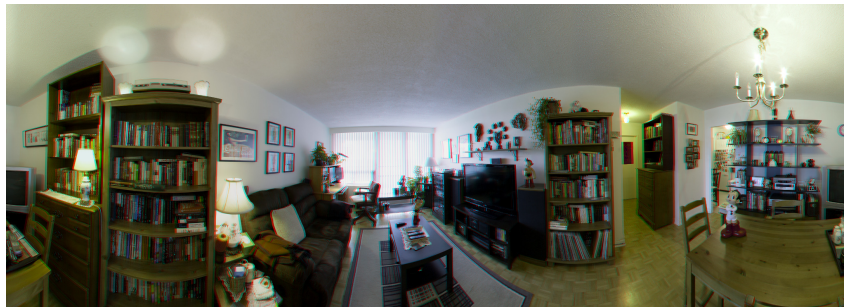
In this paper, we presented a novel contribution to the omnidirectional acquisition of stereoscopic imagery for immersive interactive media. Our model describes the visual function to be acquired by any omnistereoscopic sensor or a panoramic technique designed to produce direction-dependent stereoscopic imagery capable to stimulate binocular cues of depth in the viewer. Firstly, we modeled the viewing function that represents the signal to be acquired by a panoramic (monoscopic) sensor. This ray-tracing model of the omnidirectional light enables us to propose a more general model for the omnistereoscopic function, which we denote \mathbf{C}_S . Our omnistereoscopic model is based on the geometric constraints imposed by the human binocular system and represents the stereoscopic signal to be captured in any gaze direction around a reference viewpoint in space. In addition, we introduced an acquisition model versatile enough to describe a wide range of omnistereoscopic cameras and techniques to produce stereoscopic panoramas for human viewing. Using this acquisition model, we discuss the pros and cons of acquiring dense versus sparse samples of the omnistereoscopic function. We argue that acquisition of a few partially-overlapped stereoscopic samples of \mathbf{C}_S can produce satisfactory approximations of the ideal omnistereoscopic image defined by our model. Then, we presented examples with the sole purpose of illustrating that omnidirectional stereoscopic imagery can produce a continuous depth illusion in all gaze directions. Additionally, we described a representative rendering method based on mosaicking six stereoscopic samples. We believe that our models and rendering algorithm open the possibility to new practical strategies for omnidirectional 3D content creation needed in immersive visual media.



Configuration 1



Configuration 2



Configuration 3



Configuration 4

Figure 13: Cylindrical projections of \mathbf{I}_S (Red-Cyan anaglyphs) for the different acquisition configurations. Interactive versions are available in the companion website [7]

References

- [1] Oculus VR, LLC, “Oculus Rift VR,” Accessed 28 Jun 2015
<https://www.oculus.com/>. (2015).
- [2] Google, I., “Google Jump: camera rig,” Accessed 28 Jun 2015
<http://www.google.com/get/cardboard/jump/>. (2015).
- [3] Team, T. T., “Project Beyond,” Accessed 28 Jun 2015
<http://www.google.com/get/cardboard/jump/>. (2015).
- [4] Gurrieri, L. E. and Dubois, E., “Acquisition of omnidirectional stereoscopic images and videos of dynamic scenes: a review,” *SPIE Journal of Electronic Imaging* **22**, 030902–030902 (Jul. 2013).
- [5] Gurrieri, L. E. and Dubois, E., “Depth consistency and vertical disparities in stereoscopic panoramas,” *Journal of Electronic Imaging* **23**, 011004 (Jan. 2014).
- [6] Gurrieri, L. E., “The acquisition of omnidirectional stereoscopic images and videos of dynamic scenes: a review,” website
<http://luisgurrieri.net/pub/background/>. (2013).
- [7] Gurrieri, L. and Dubois, E., “A model for the omnidirectional acquisition of stereoscopic images,” Website
<http://luisgurrieri.net/pub/model/>. (2014).
- [8] Adelson, E. and Wang, J., “Single lens stereo with a plenoptic camera,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **14**(2), 99–106 (1992).
- [9] Peleg, S. and Ben-Ezra, M., “Stereo panorama with a single camera,” in *Proc. IEEE Conf. Computer Vision Pattern Recognition* 395–401 (Jun. 1999).
- [10] Huang, H.-C. and Hung, Y.-P., “Panoramic stereo imaging system with automatic disparity warping and seaming,” *Graphical Models and Image Processing* **60**(3), 196–208 (1998).
- [11] Huang, F. and Klette, R., “Stereo panorama acquisition and automatic image disparity adjustment for stereoscopic visualization,” *Multimedia Tools and Applications* **47**, 353–377 (2010). 10.1007/s11042-009-0328-2.
- [12] Gurrieri, L. E. and Dubois, E., “Efficient panoramic sampling of real-world environments for image-based stereoscopic telepresence,” in *Proc. SPIE 8288, Stereoscopic Displays and Applications XXIII* 82882D (Feb. 2012).
- [13] Baker, H. H. and Constantin, P., “Panoramic stereoscopic camera,” U.S. Patent 2012/0105574 application (2010).
- [14] Couture, V. and Roy, S., “The omnipolar camera: A new approach to stereo immersive capture,” in *IEEE Int. Conf. on Computational Photography* 1–9 (2013).
- [15] Scheer, O., Kauff, P., Weissig, C., and Rosenthal, J.-C., “Geometrical design concept for panoramic 3D video acquisition,” in *20th European Signal Processing Conf.* 2757–2761 (Aug. 2012).

- [16] Gurrieri, L. E. and Dubois, E., “Stereoscopic cameras for the real-time acquisition of panoramic 3D images and videos,” in *Proc. SPIE 8648, Stereoscopic Displays and Applications XXIV* 86481W (Feb. 2013).
- [17] Adelson, E. H. and Bergen, J. R., “The plenoptic function and the elements of early vision,” *Computational models of visual processing* **1**, 3–20 (1991).
- [18] Nayar, S., “Catadioptric omnidirectional camera,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 482–488 (1997).
- [19] McMillan, L. and Bishop, G., “Plenoptic modeling: an image-based rendering system,” in *Proc. of the 22nd annual conference on Computer graphics and interactive techniques SIGGRAPH '95*, 39–46, ACM, New York, USA (1995).
- [20] Wetzstein, G., Ihrke, I., and Heidrich, W., “On plenoptic multiplexing and reconstruction,” *International Journal of Computer Vision* , 1–17 (2013).
- [21] Huang, F., Wei, S., and Klette, R., “Geometrical fundamentals of polycentric panoramas,” in *Proc. 8th IEEE International Conference on Computer Vision* **1**, 560–565, IEEE Comput. Soc., Los Alamitos, CA, USA (Jul. 2001).
- [22] Wei, S., Huang, F., and Klette, R., “Determination of geometric parameters for stereoscopic panorama cameras,” *Machine Graphics & Vision* **10**(3), 399–427 (2001).
- [23] Huang, F., Klette, R., and Scheibe, K., [Panoramic Imaging: Sensor-Line Cameras and Laser Range-Finders], John Wiley & Sons Inc., Hoboken, NJ (2008).
- [24] Reynolds, R. O., “Design of a stereo multispectral CCD camera for Mars Pathfinder,” in *Proceedings of SPIE* **2542**, 197–206 (1995).
- [25] Couture, V., Langer, M. S., and Roy, S., “Analysis of disparity distortions in omnistereoscopic displays,” *ACM Transactions on Applied Perception (TAP)* **7**, 25:1–25:13 (Jul. 2010). ACM ID: 1823743.
- [26] Bourke, P., “Capturing omni-directional stereoscopic spherical projections with a single camera,” in *16th International Conference on Virtual Systems and Multimedia (VSMM)* 179–183 (2010).
- [27] Wetzstein, G., Ihrke, I., Lanman, D., and Heidrich, W., “Computational plenoptic imaging,” *Computer Graphics Forum* **30**, 2397–2426 (Dec. 2011).
- [28] Dubois, E., [The structure and properties of color spaces and the representation of color images], vol. **4** of *Synthesis Lectures on Image, Video, and Multimedia Processing*, Morgan & Claypool (2009).
- [29] Southwell, D., Basu, A., Fiala, M., and Reyda, J., “Panoramic stereo,” in *Proc. IEEE Int. Conf. Pattern Recognition* **1**, 378–382 (1996).
- [30] Spacek, L., “Coaxial omnidirectional stereopsis,” in [*Computer Vision - ECCV 2004*] **3024**, 354–365, Springer Berlin / Heidelberg (2004).
- [31] Adobe, “Photoshop Lightroom,” Website <http://www.adobe.com/ca/products/photoshop-lightroom.html>. (2013).

- [32] GIMP, “GIMP: The GNU image manipulation program,” Website <http://www.gimp.org/>. (2013).
- [33] Mathworks, “Matlab,” Website <http://www.mathworks.com/>. (2013).
- [34] Sutu, M. and Sykes, D., “StereoPhotoMaker: stereo photography authoring tool,” Website <http://stereo.jp.org/eng/stphmkr/>. (2013).
- [35] Xrite, “Color Checker Passport,” Website http://xritephoto.com/ph_product_overview.aspx?id=1257&catid=28&action=overview. (2013).
- [36] Mathworks, “Stereo Image Rectification,” Website <http://www.mathworks.com/help/vision/examples/stereo-image-rectification.html>. (2012).
- [37] Bay, H., Tuytelaars, T., and Van Gool, L., “SURF: speeded up robust features,” in [*Computer Vision ECCV 2006*] *Lecture Notes in Computer Science* **3951**, 404–417, Springer Berlin / Heidelberg (2006).
- [38] Mathworks: Computer Vision Toolbox, “Matlab’s implementation of SURF:detectSURFFeatures,” Website <http://www.mathworks.com/help/vision/ref/detectsurffeatures.html>. (2012).
- [39] Fischler, M. A. and Bolles, R. C., “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Commun. of the ACM* **24**, 381–395 (Jun. 1981).
- [40] Mathworks: Computer Vision Toolbox, “Estimate fundamental matrix from corresponding points in stereo images: estimateFundamentalMatrix,” Website <http://www.mathworks.com/help/vision/ref/estimatefundamentalmatrix.html>. (2012).
- [41] Xiong, Y. and Pulli, K., “Color correction for mobile panorama imaging,” in *Proc. of the 1st Int. Conf. on Internet Multimedia Computing and Service* 219–226 (2009).
- [42] Shum, H.-Y. and Szeliski, R., “Construction of panoramic image mosaics with global and local alignment,” in [*Panoramic Vision*] Benosman, R. and Kang, S. B., eds., ch. 13, 227–268, Springer Verlag (2001).
- [43] Szeliski, R., [Computer vision algorithms and applications], Springer, London; New York (2011).
- [44] Szeliski, R., “Image alignment and stitching: A tutorial,” *Foundations and Trends in Computer Graphics and Vision* **2**(1), 1–10 (2006).
- [45] Ha, S. J., Koo, H., Lee, S. H., Cho, N. I., and Kim, S. K., “Panorama mosaic optimization for mobile camera systems,” *IEEE Transactions on Consumer Electronics* **53**, 1217–1225 (Nov. 2007).

- [46] Daz Hernandez, R., Rendering stereoscopic 3D images in cylindrical spaces, PhD thesis, Universidad Politecnica de Valencia (Nov. 2011).
- [47] Burt, P. and Adelson, E., "The laplacian pyramid as a compact image code," *IEEE Transactions on Communications* **31**(4), 532–540 (1983).