

Efficient panoramic sampling of real-world environments for image-based stereoscopic telepresence

Luis E. Gurrieri and Eric Dubois

School of Electrical Engineering and Computer Science
University of Ottawa
Ottawa, ON, Canada.

ABSTRACT

A key problem in telepresence systems is how to effectively emulate the subjective experience of *being there* delivered by our visual system. A step toward visual realism can be achieved by using high-quality panoramic snapshots instead of computer-based models of the scene. Furthermore, a better immersive illusion can be created by enabling the free viewpoint stereoscopic navigation of the scene, i.e. using omnistereoscopic imaging. However, commonly found implementation constraints of telepresence systems such as acquisition time, rendering complexity, and storage capacity, make the idea of using stereoscopic panoramas challenging. Having these constraints in mind, we developed a technique for the efficient acquisition and rendering of omnistereoscopic images based on sampling the scene with clusters of three panoramic images arranged in a controlled geometric pattern. Our technique can be implemented with any off-the-shelf panoramic cameras. Furthermore, it does not require neither the acquisition of additional depth information of the scene nor the estimation of camera parameters. The low the computational complexity and reduced data overhead of our rendering process make it attractive for the large scale stereoscopic sampling in a variety of scenarios.

Keywords: stereo vision, stereo image processing, panoramic image acquisition, image-based telepresence, stereoscopic panoramas, omnistereoscopic, clusters of panoramas, stereoscopic virtual environments

1. INTRODUCTION

The ultimate purpose of a telepresence system is to emulate the experience of being virtually immersed in a remote world location, and to enable the realistic, free viewpoint navigation of that virtual environment. The research on various aspects of this idea is the main goal of the NAVIRE (Navigation in Image-Based Representations of Real World Environments) project at the University of Ottawa.

Telepresence has become part of our daily life in the form of online interactive maps such as Google StreetView, MapJack, and EveryScape, and panoramic virtual tours of international museums and art galleries. Even though synthetic virtual environments have been used for years in the training of security and military personnel, the realism provided by image-based virtual scenarios still cannot be surpassed, at least, up to the moment of writing this paper. It is clear that the socioeconomic impact of this technology increases the interest to improve the realism of these immersive simulations.

In terms of improvements, the binocular stereoscopic visualization of these image-based environments is a desirable enhancement. In this regard, a possible method to create the illusion of omnidirectional depth is by sampling a target location with a number of stereoscopic panoramic images. However, the large number of samples necessary to provide a

Further author information: (Send correspondence to Luis E. Gurrieri)

Luis E. Gurrieri: E-mail: Luis.Gurrieri@uOttawa.ca, Website: <http://www.LuisGurrieri.net/>

Eric Dubois: E-mail: edubois@uottawa.ca, Website: <http://www.eecs.uottawa.ca/~edubois/>

Luis E. Gurrieri and Eric Dubois, "Efficient panoramic sampling of real-world environments for image-based stereoscopic telepresence", Andrew J. Woods, Nicolas S. Holliman, Gregg E. Favalora, Editors, Proc. SPIE 8288, 82882D (2012).

Copyright 2012 Society of Photo-Optical Instrumentation Engineers. One print or electronic copy may be made for personal use only. Systematic electronic or print reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited.

<http://dx.doi.org/10.1117/12.908794>

smooth and unrestricted navigation poses a problem. The lengthy acquisition time of existent omnistereoscopic rendering methods was one of the motivations to propose an alternative acquisition and rendering strategy.

In this paper, we present a different approach to the problem of creating omnistereoscopic images. Our goal is to achieve consistent and correct omnidirectional depth rendition without auxiliary depth information of the scene and consequently, reducing acquisition time, data overhead, and rendering complexity. A method that improves the performance in any one of these aspects will greatly simplify the entire logistics of the stereoscopic sampling of a real world scene.

1.1 The Acquisition and Rendering of Stereoscopic Images

The idea of rendering stereoscopic panoramas using rotating cameras (or line sensors) has been widely adopted since the late 90's.^{1,2} This approach is capable of producing high-quality stereoscopic panoramas but at the cost of using long acquisition times. The latter limitation makes the rotating sensor approach impractical for the sampling of a large number of scene viewpoints. However, it is technically feasible to capture left and right parallel eye view with two line sensors separated by a baseline b instead of a single line sensor, and to increase the number of pairs of sensors. This approach would help to capture non-overlapping stereoscopic sections of the scene simultaneously, which can then be mosaicked to create a full omnistereoscopic image in a fraction of time of a single rotating sensor system. Unfortunately, a rotating camera system with such characteristics is still commercially unavailable to the best knowledge of the authors.

The problem of 3-D modeling of a scene gave raise to alternative omnistereoscopic rendering methods.^{3,4} However, these ideas were not conceived for human stereopsis and cannot be directly adapted to create a stereoscopic image-based virtual environment.

In another line of research, robot navigation^{5,6} and teleoperation areas have provided solutions for the stereoscopic visualization of real-world scenes. In these cases, a robot or a probe equipped with binocular cameras transmits stereoscopic images to the user, who controls in real-time the probe operation. In image-based telepresence simulation, the free navigation of the remote location is implemented by using images of the scene in conjunction with telemetric information stored in a database. These image-base environments are generated off-line. Even though a teleoperated mobile platform can be used to sample a scene, the large number of omnistereoscopic images required demands a more efficient method to create images using a limited number of sampled sources.

Many techniques were specifically conceived for the off-line navigation of image-based stereoscopic virtual environments. For instance, Yamaguchi et al.⁷ and, in a more recent follow up, Hori et al.,⁸ proposed a method to generate omnistereoscopic images using panoramic videos. Their approach enables a smooth navigation of the scene, but at the cost of creating large data overhead. Furthermore, without extra depth information of the scene, it is difficult to maintain a consistent depth rendition around and between arbitrary viewpoints. Moreover, the exact position of each panoramic frame must be known in order to find the best panoramic pair to render a binocular image for each user's virtual location and viewing direction, i.e. pan and tilt viewing angles. In addition, it is not feasible to acquire the panorama frames in an equally spaced grid to have control over the stereoscopic baseline. To cope with this problem, Hori et al.⁹ proposed capturing a panoramic video using simultaneously two panoramic cameras mounted on a rig. Unfortunately, this approach cannot provide a full omnistereoscopic rendition of the scene and does not solve the data overhead problem.

An alternative omnistereoscopic acquisition strategy was proposed by Vanijja and Horiguchi,¹⁰ which was specifically tailored for a CAVE type of display. Their idea relies on a limited set of four panoramas acquired in a controlled sampling pattern, which are used to render four wide field-of-view panoramic images which are projected on each one of the four CAVE's walls to produce an omnistereoscopic immersive experience. The CAVE display imposes restrictions in terms of the number panoramas to use, and consequently, the sampling spatial configuration to employ. In addition, at the moment of their proposal, the authors acknowledged limitations of the viable stereoscopic field-of-view at certain viewing angles due to the different location of the panoramic projection centers. Even with these unsolved problems, this method offers advantages in terms of acquisition time and depth consistency between sampled viewpoints, thus making it suitable for a practical stereoscopic telepresence system.

The method we present in this article extends and improves the idea proposed by Vanijja and Horiguchi, reducing the required number of panoramic samples, and introducing image corrections to improve viewing comfort over the whole field-of-view of the scene. One of our goals is to limit the size of the panoramic database and to achieve consistency in the perceived depth without depending on additional scene's depth information. To do so, we propose a sampling strategy based on clusters of three panoramic snapshots distributed in a coplanar and equidistant triangular pattern.

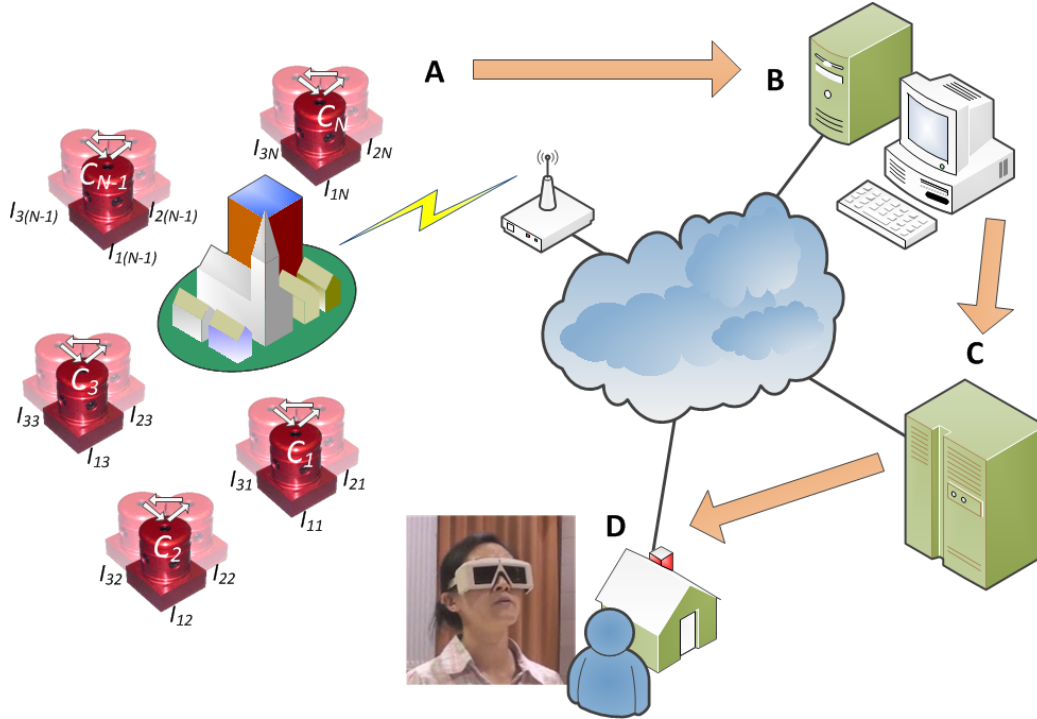


Figure 1. Stereoscopic telepresence system: A) the remote world location is sampled with N panoramic clusters composed by panoramic triads, B) the collected sets of panoramas is stored in a server and processed off-line to generate an omnistereoscopic pair per cluster, C) these omnistereoscopic images are stored in a dedicated server along with their relative location in the scene, D) a user in any location can access this server and navigate the interactive stereoscopic simulation.

2. THE METHOD

Any pair of co-planar cylindrical panoramas exhibits two sections which can be used to compose a stereoscopic image without noticeable depth distortion. This stereoscopic region of interest arises due to the offset between projection centers, which create different viewpoints for any panoramic snapshot. A cluster of co-planar panoramas, taken pairwise, will exhibit these regions of stereoscopic usability in different panning (azimuthal) angles. We found that the idea of Vanijja and Horiguchi of using clusters of panoramas can be improved by reducing the number of panoramas to three instead of four, and by changing the sampling pattern to arrange the panoramic snapshots in an equidistant triangle. A cluster of cylindrical panoramas with such distribution can be used to render a complete omnistereoscopic image by mosaicking the appropriate stereoscopic image sections.

Mosaicking image sections originated in different panoramas within the cluster enables to render two novel omnidirectional images: I_L and I_R , corresponding to the left and right parallel eye views. This is an approximation to panning the scene with two parallel image planes, separated a baseline b , from which only the center image column is mosaicked to create the final omnistereoscopic pair.^{1,2}

We propose to perform the stereoscopic sampling of a scene using N panoramic clusters C_i as illustrated in the part A of Fig. 1. Each C_i consists of three cylindrical panoramas in a triangular coplanar pattern, i.e. $C_i = \{I_{1i}, I_{2i}, I_{3i}\}$, where $i \in \{1, \dots, N\}$. The larger the number of clusters acquired, the easier it is to implement the free viewpoint stereoscopic navigation of the scene, but at the price of increasing the acquisition time and the data overhead. On the other hand, fewer viewpoints implies generating more intermediate images during the navigation. Therefore, the more efficient the view interpolation technique used, the lower the acquisition, transmission, and storage required.

The panoramas and each cluster's absolute location are transferred and stored in a remote workstation (part B of Fig. 1) to be used in the off-line rendering of stereoscopic panoramas (part C of Fig. 1). Alternatively, a real-time rendering can be implemented by generating one stereoscopic image pair according to the user's viewing direction and virtual location in simulated scene. The latter requires the feed back of the user's gaze direction to generate the stereoscopic views (part D

of Fig. 1). If this feed back information is available, it can be used to simulate oculomotor mechanisms such as focus and vergence, improving the binocular experience by reducing the visual fatigue. The stereoscopic image pair of the scene transmitted to the user is generated based on the closest stereoscopic samples to the user’s virtual location.

The user is expected to be immersed in the simulation for extended periods of time, hence, the visual fatigue becomes an important factor. In this paper, we detail our omnistereoscopic rendering technique based on using the triad of panoramic images in section 2.4. Moreover, we propose methods to reduce the visual fatigue and enhance the quality of the final omnistereoscopic image in sections 2.5, 2.6, and 2.7.

2.1 The Format of the Panoramic Images

The 3-D scene is registered on a cylindrical canvas at a distance R with respect to each projection center. However, to process the image, we use the unwrapped 2-D representation of this 3-D cylindrical canvas represented in Fig. 2. This image representation maps the three components of a color sample to a (n, m) location, i.e. $I(n, m) : Z^2 \rightarrow Z^3$, defined for $n \in [1, N]$ and $m \in [1, M]$, where N is the number of rows, and M is the number of columns. Furthermore, this panoramic image representation is periodic in the panning dimension, i.e., $I(n, m) = I(n, m - M)$.

A column in $I(n, m)$ is related to the viewing direction θ in cylindrical coordinates by

$$m = \begin{cases} \lfloor (\theta + 180^\circ) \cdot M/360^\circ \rfloor, & \text{if } 0^\circ \leq \theta \leq 180^\circ \\ \lfloor (\theta - 180^\circ) \cdot M/360^\circ \rfloor, & \text{if } 180^\circ < \theta < 360^\circ \end{cases} \quad (1)$$

where $m \in [1, M]$.

The coordinate system at the center of the cluster, (X_C, Y_C) , which defines the panning angle θ , will be used as reference for the panoramas in the cluster, and to render the columns of the omnistereoscopic image pair, (I_L, I_R) . Each panorama is centered at $(N/2, M/2)$ to exhibit partially overlapped panning regions which are going to be used to create the omnistereoscopic image. The optimal centering does not imply necessarily that all three panoramas will register the same scene element at the center as it will be explained in sections 2.2 and 2.3.

In the following analysis, the panning angle $\theta \in [0^\circ, 360^\circ)$ associated with a column $m \in [1, M]$ by Eq. 1 will be used to indicate the panning direction when rendering the columns of the omnistereoscopic image pair. In order to indicate the column of of the panorama I at the panning angle θ , we are going to use the notation $I(\cdot, \theta)$ to indicate a column vector originated in the panorama $I(n, m)$, where m is related to θ by Eq. 1 and $1 \leq n \leq N$. Conversely, we will use the notation $I(\cdot, m)$ to indicate the column m of the panorama I corresponding to a panning angle θ .

In the rest of the paper, unless stated the contrary, whenever we refer to a *panorama*, we are referring specifically to the *cylindrical* rendering of a panorama.

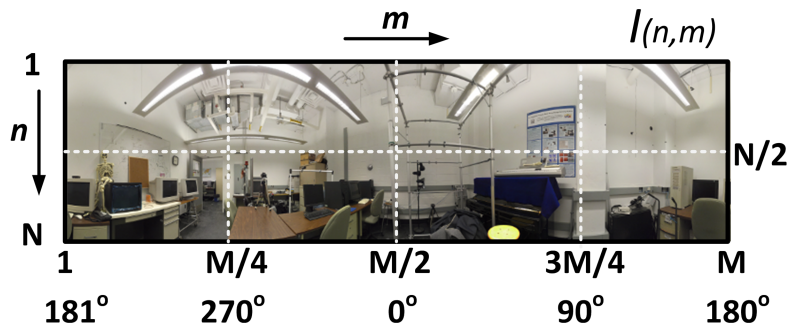
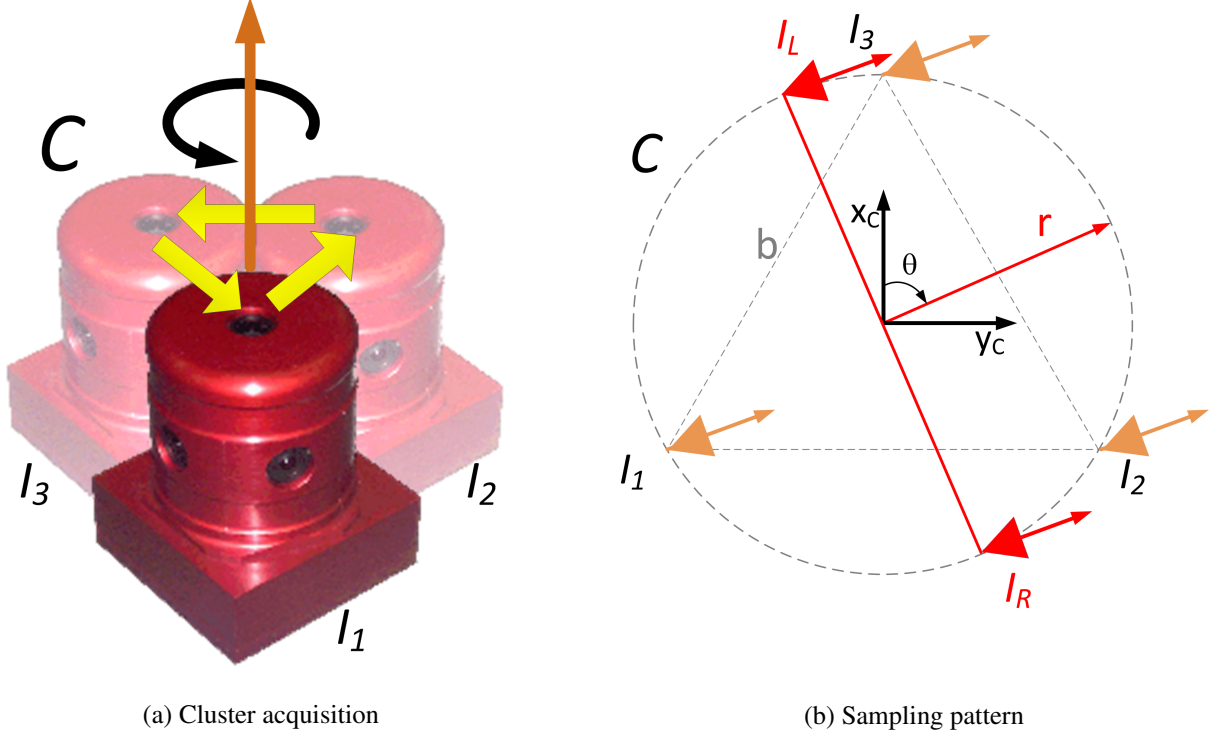


Figure 2. Planar panoramic image $I(n, m)$.



(a) Cluster acquisition

(b) Sampling pattern

Figure 3. Panoramic clusters: (a) example of a cluster acquisition by rotating a panoramic camera located at a distance r of the rotation center, and taken three panoramic snapshots, (b) panoramic cluster scheme where the projection center of each panorama coincide with different vertices of an equilateral triangle of side b (baseline).

2.2 Identifying the Stereoscopic Regions in a Panoramic Cluster

A triad of panoramas C can be obtained by rotating a panoramic camera in a circle of radius r and acquiring three panoramic snapshots at regular intervals as illustrated in Fig. 3-(a). A triangular sampling pattern is defined in this way, where the projection centers of each panorama (I_1, I_2, I_3) coincide with the vertices of an equilateral triangle with side length b as shown in Fig. 3-(b).

When the objects in the scene are at much farther distance than the separation of projection centers, the panning angle θ is approximately the same for all panoramas in C . Hence, any novel column m of the stereoscopic pair (I_L, I_R) will be a function of the columns m of (I_1, I_2, I_3), where the index m will be the same for all panoramas. The distance between the scene elements and the reference center of C is a fundamental constraint for the use of this stereoscopic rendering method since objects too close to the cluster will affect the optimal pairwise alignment between panoramas necessary to find the optimum stereoscopic regions.

The distance from the camera to the scene poses a limitation, particularly in indoor environments. However, this constraint depends on the combination of two factors: the type of panoramic camera used and the rendering algorithm applied to generate the base cylindrical panoramas. Improving the pixel projection algorithm used to render the panoramas in C can effectively reduce the minimum distance between the clusters and the closest objects in the scene.

A key element of our method is to identify the panning angles for optimum stereoscopic rendition between any pair of panoramas. For any pair of panoramas within a cluster there is a distinctive panning direction where the optical motion field exhibits a forward movement pattern called *focal expansion* at θ_{fe} .¹¹ In a cylindrical panoramic pair, the motion field may also exhibit a backward motion pattern called *focal contraction* at $\theta_{fe} - 180^\circ$. We define the panning *direction of alignment* (θ_o) as the panning direction where either a focal expansion (θ_{fe}) or a focal contraction ($\theta_{fe} - 180^\circ$) pattern in the motion field is located (Fig. 4).

Our method exploits the partially overlapped stereoscopic samples of the light field extracted from pairs of panoramas at certain panning angles. In this regard, there are two regions centered around $\theta_o \pm 90^\circ$ which can be used to render a

stereoscopic image (indicated as the shaded areas in Fig. 4-(a) and (b)). We found experimentally that a pair of contiguous panoramic sections, panning $\pm 60^\circ$ around $\theta_o \pm 90^\circ$, captures the scene depth without noticeable distortion. It is possible to compose a complete omnistereoscopic image by mosaicking these wide field-of-view panoramic sections. Experimentally, we also found that the final result reproduces consistently the depth relationships between objects in the scene.

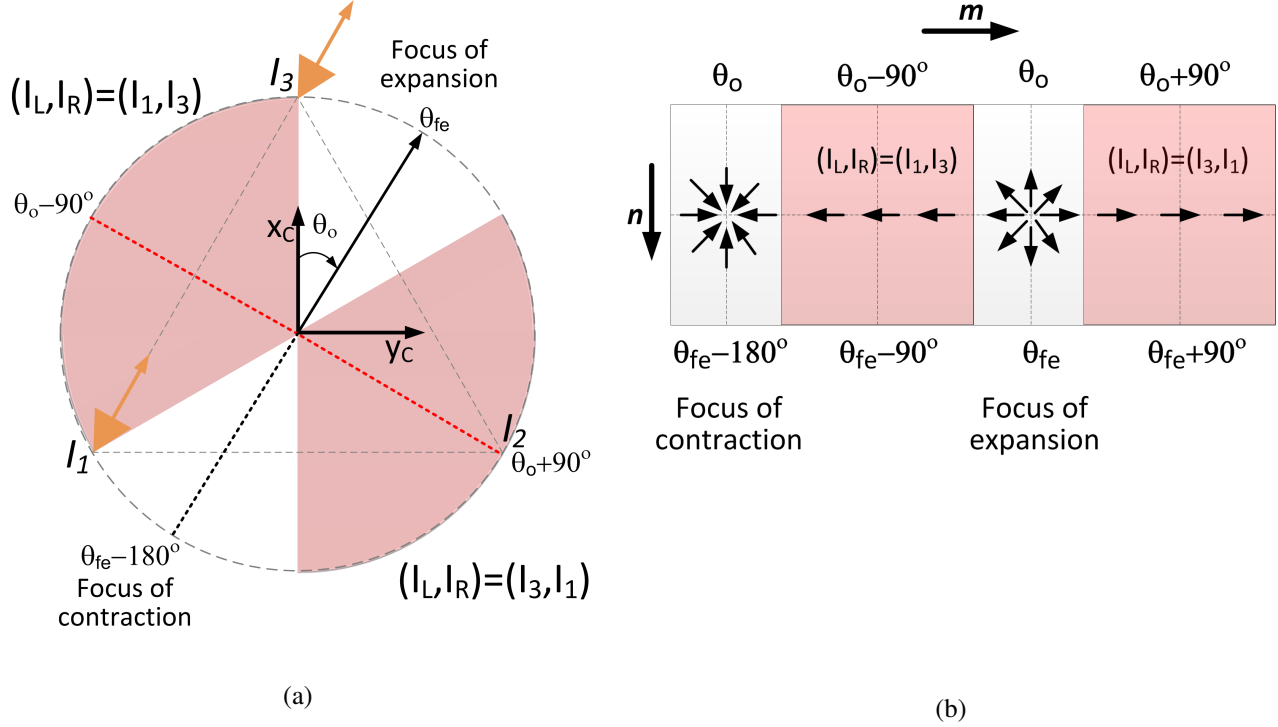


Figure 4. This example illustrates the location of the optimum alignment for I_1 and I_3 : (a) the motion field’s focal expansion and contraction panning directions are located at θ_{fe} and $\theta_{fe} - 180^\circ$, respectively, determining the directions of alignment (θ_o) and two panoramic sections usable for stereoscopic rendering around $\theta_o \pm 90^\circ$, (b) the 2-D optical flow for (I_1, I_3) indicating the areas usable for stereoscopic rendering in color, the alignment directions, and the optical motion field directions.

We can now explain with an example how we propose to create I_L and I_R . Based on the example in Fig. 3-(b), in the panning direction θ or m^* , the columns $(I_1(\cdot, m^*), I_3(\cdot, m^*))$ will be used to render the column $I_L(\cdot, m^*)$, and the columns $(I_1(\cdot, m^*), I_2(\cdot, m^*))$ will be used to render the column $I_R(\cdot, m^*)$. Since our rendering method is based on mosaicking panoramic sections from the closest panorama in C with respect to the projection centers (I_L, I_R) for any given θ , in the previous example, $I_L(\cdot, m^*) = I_3(\cdot, m^*)$ and $I_R(\cdot, m^*) = I_2(\cdot, m^*)$. The whole rendering process is detailed in section 2.4.

Notice that, since the projection center of (I_L, I_R) is moving in a circular trajectory while rendering novel columns according to the panning angle θ , there is no unique projection center for each panorama of the stereoscopic pair. Hence, the omnistereoscopic rendering will not produce a pair of conventional cylindrical panoramas according to the definition of a cylindrical panorama. Nevertheless, we will refer to (I_L, I_R) as panoramic images in this paper.

The distance b between pairs of panoramas in Fig. 3 determines the stereoscopic baseline. In this regard, the triangular sampling pattern determines a circle of radius $r = b/2 \cdot \sec(30^\circ) \approx 0.58 b$. In our experiments, we found that a stereoscopic baseline $b = 0,0065 \text{ cm}$, equivalent to the average interocular distance in the adult human population, is appropriate for indoor scenes where the scene elements are far from the cluster center with respect to b . The parameter b determines a rotation radius $r = 0.0038 \text{ m}$, which gives a clear idea of the magnitude of the camera rotation needed. A smaller baseline (hypo-stereo), e.g. $b = 0.0025 \text{ m}$, is a better option in indoor locations where objects are closer to the panoramic cluster and conversely, a larger baseline (hyper-stereo), e.g. $b = 0.009 \text{ m}$ or larger, can be used for outdoor locations, depending the illusion of depth we want to achieve in the final binocular image.

2.3 Finding the Optimum Center for the Panoramas within a Cluster

The three panoramas must be optimally centered as a necessary condition to identify the panoramic sections to mosaic. This optimum direction of alignment is not arbitrary. In fact, there is a single alignment direction between each pair of panoramas for the co-planar triangular configuration in Fig. 3-(b). As an example, using the cluster reference system represented in Fig. 3-(b), when panoramas I_1 and I_2 are optimally centered, the focal of expansion (θ_{fe}) and contraction ($\theta_{fe} - 180^\circ$) are exhibited at $\theta = 90^\circ$ and $\theta = 270^\circ$, respectively. The panoramic triad is optimally centered when the three optimal alignment directions, defined by the focal expansion and contraction panning angles, correspond to the directions given in Table 1.

The panoramic triad can be optimally centered by controlling the acquisition process, i.e. knowing the orientation of the camera head for each snapshot. However, further pairwise alignment may still be necessary in some cases. First, we align the three panoramas to the same scene element imaged at the center of I_3 . This can be done by adapting the zero phase representation proposed by Pajdla and Hlaváč¹² or by finding the horizontal shifts to apply on I_1 and I_2 to compensate the horizontal disparity* with respect to I_3 at the center of the panorama, $(N/2, M/2)$. Once the three panoramas' centers are aligned to the same scene element, the optimum centering can be optimized by finding the shifts (m_1, m_2, m_3) , such that $(I_1(n, m - m_1), I_2(n, m - m_2), I_3(n, m - m_3))$ would exhibit the focal expansion and contraction pattern in the panning direction of alignment (m_o) given by Table 1. In order to optimize the panoramic alignment in this context, we proposed an uncalibrated method based on finding the image flow pattern characteristic of the focal expansion and contraction patterns.¹³

Table 1. Panning angles of alignment (θ_o) and the corresponding panoramic columns of alignment (m_o).

Panoramic Pair	θ_o	m_o
(I_1, I_2)	$(90^\circ, 270^\circ)$	$(3M/4, M/4)$
(I_2, I_3)	$(150^\circ, 330^\circ)$	$(11M/12, 5M/12)$
(I_3, I_1)	$(30^\circ, 210^\circ)$	$(7M/12, M/12)$

The regions of interest after alignment for the case panoramas I_1 and I_3 are illustrated in Fig. 4-(a). This example shows the panning angles usable for stereoscopic rendering and the alignment direction. The image flow map after aligning calculated over (I_1, I_3) shows the motion field patterns on the directions of focal expansion (θ_{fe}) and contraction ($\theta_{fe} - 180^\circ$), and the areas usable for stereoscopic rendering at $\theta_{fe} \pm 90^\circ$ (Fig. 4-(b)).

2.4 The Rendering of the Omnistereoscopic Pair: I_L and I_R

There are different approaches to synthesize novel columns ($I_L(\cdot, \theta), I_R(\cdot, \theta)$) for $\theta \in [0^\circ, 360^\circ)$ based on the triangular configuration between panoramic projection centers illustrated in Fig. 3-(b). For instance, the plenoptic modeling proposed by McMillan and Bishop¹⁴ is a valid approach for the view interpolation of panoramic columns. In this case, the plenoptic function at arbitrary viewing angles can be estimated by interpolating the image flow between the closest pair of panoramas to the novel projection centers of I_L or I_R for each θ . Interpretatively, the novel columns of (I_L, I_R) can be generated by morphing the views from the nearest panoramic sources in the cluster, adapting the technique of view morphing proposed by Seitz and Dyer.¹⁵ The latter method requires applying the appropriate transformations to the source image pair to satisfy the parallel cameras precondition.

Conveniently, the geometry of the problem suggests a simpler approach for the rendering of the stereoscopic pair. As we have already anticipated, the panning direction that marks the alignment between any pair of panoramas in C determines two panoramic sections where the horizontal disparity around the central region of the panoramic pair encodes the depth of the scene (section 2.3). These sections overlap in a way that by choosing the correct panoramic pair for each given θ , it is possible to create binocular views of the scene. Hence, we proposed to generate each novel column of the omnistereoscopic pair using the closest panoramas in C to the projection centers of (I_L, I_R) for any given θ .

To illustrate this idea, we use again the scheme in Fig. 3-(b). In that example, for that particular panning θ , the column $I_L(\cdot, \theta)$ will be generated using the columns of the closest panoramic pair for that panning angle, in this case

*In this article, we use disparity (horizontal or vertical) to indicate the relative distance in pixel units between corresponding feature points of the scene projected onto different 2-D panoramic images.

$(I_1(\cdot, \theta), I_3(\cdot, \theta))$. Similarly, $I_R(\cdot, \theta_C)$ will be synthesized from $(I_1(\cdot, \theta), I_2(\cdot, \theta))$. The proximity of (I_L, I_R) projection centers to (I_1, I_2, I_3) will determine what panoramas in C will contribute to each novel column rendered. For the particular θ in the example, $I_L(\cdot, \theta) = I_3(\cdot, \theta)$ and $I_R(\cdot, \theta) = I_2(\cdot, \theta)$.

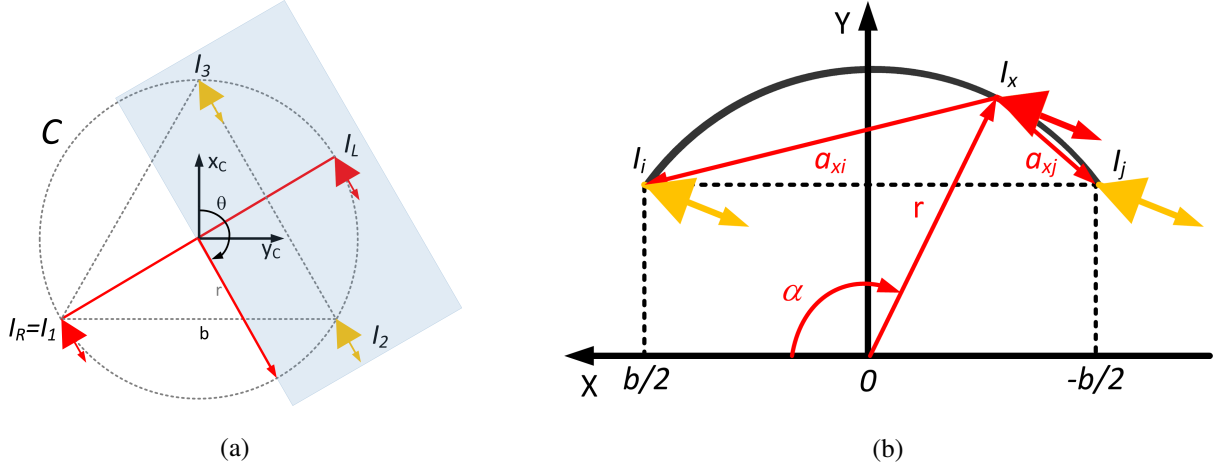


Figure 5. The rendering is modeled by studying the proximity between the target panorama to synthesize and the closest pair of panoramas in C : (a) in the example, the panoramas in the boxed area are used to model the proximity relationships, (b) the distances a_{xi} and a_{xj} between I_x (I_L or I_R), and the two closest panoramas I_i and I_j ($i, j \in \{1, 2, 3\}$), respectively.

It is useful to model the relationship between the rendered virtual projection center of (I_L, I_R) and the closest two panoramas in the cluster. In Fig. 5-(a), the boxed area illustrates a projection center distribution that appears repeatedly while panning all possible angles, i.e. the target panorama to render I_x , which could be I_L or I_R , and the closet panoramas in C , I_i and I_j , where $i, j \in \{1, 2, 3\}$. In this model, illustrated in Fig. 5-(b), the angle $\alpha \in [30^\circ, 150^\circ]$ marks the location of the projection centers. The panning angle when rendering columns progressing clockwise is $\theta = \alpha + 90^\circ$. The distances between the projection centers of (I_x, I_i) and (I_x, I_j) are a_{xi} and a_{xj} , respectively.

A simple inspection shows that for $30^\circ \leq \alpha < 90^\circ$, I_x is closer to I_i , and for $90^\circ < \alpha \leq 150^\circ$, I_x is closer to I_j . When $\alpha = 90^\circ$, I_x and I_j are optically aligned as defined in section 2.3, and this angle marks the transition between rendering columns of I_x using I_i to using I_j . Hence, we can express the column rendering as

$$I_x(\cdot, \theta) = w(\alpha) \cdot I_i(\cdot, \theta) + (1 - w(\alpha)) \cdot I_j(\cdot, \theta), \quad (2)$$

where the transfer function $w(\alpha)$ is defined as

$$w(\alpha) = \begin{cases} 1, & \text{if } 30^\circ \leq \alpha < (90^\circ - \tau) \\ f(\alpha), & \text{if } (90^\circ - \tau) \leq \alpha < 90^\circ \\ 0, & \text{if } 90^\circ \leq \alpha \leq 150^\circ \end{cases} \quad (3)$$

where $f(\alpha) : R \rightarrow R$ is a monotonically decreasing function of the angle α , and $0 \leq f(\alpha) \leq 1$ for $(90^\circ - \tau) \leq \alpha < 90^\circ$. The blending between panoramic sections is defined in the interval τ . Experimentally, we found that $\tau \simeq 10^\circ$, or equivalently $\Delta_m = \lfloor M/36 \rfloor$ columns before the transition column (m_o), was appropriate to blend panoramic sections when using the ghosting reduction process defined in section 2.5. The two image sections where the blending takes place can be defined as: $im_i = I_i(n, m)$ and $im_j = I_j(n, m)$, where $1 \leq n < N$, $(m_o - \Delta_m) \leq m < m_o$, and where m_o is the column corresponding to the alignment angle θ_o defined in Table 1. Note that θ_o corresponds to panning direction of where alternatively the focal expansion θ_{fe} or contraction $(\theta_{fe} - 180^\circ)$ patterns are exhibited (Table 2). The defined two image sections (im_i, im_j) will be the target of the ghosting reduction as it will be explained in section 2.5.

Modeled in this way, the rendering of columns is based on the proximity of the projection center of I_x to the projection centers of I_i and I_j , using only columns of the closest source panorama and blending only around the aligning angle ($\alpha = 90^\circ$). From Eq. 2 and 3, $I_x(\cdot, \theta) = I_i(\cdot, \theta)$ for $0^\circ < \alpha < (90^\circ - \tau)$, and $I_x(\cdot, \theta) = I_j(\cdot, \theta)$ for $90^\circ \leq \alpha \leq 150^\circ$. The

novel column $I_i(\cdot, \theta)$ is rendered based on the weighted combination of columns originated in $I_i(\cdot, \theta)$ and $I_j(\cdot, \theta)$ only for $(90^\circ - \tau) \leq \alpha \leq (90^\circ)$.

The previous model can be used to describe the generation of I_L and I_R over the whole range of θ . In order to do this, we need to define the correspondence between α and θ , and which panoramas correspond to (I_i, I_j) for each θ . All the necessary information to render the omnistereoscopic pair using a optimally centered panoramic triad is provided in Tables 1 and 2. Finally, the panoramic sequences to mosaic in order to render (I_L, I_R) as a function of column number is summarized in Fig. 6.

Table 2. Panorama selection for the stereoscopic pair rendering

I_x	Panning angle: θ	Column: m	α	Aligning: $\theta_o(m_o)$	I_i	I_j
I_L	$330^\circ \leq \theta < 90^\circ$	$5M/12 \leq m < 3M/4$	$\theta + 60^\circ$	$30^\circ (7M/12)$	I_1	I_3
I_L	$90^\circ \leq \theta < 210^\circ$	$3M/4 \leq m < M - 1, 1 \leq m < M/12$	$\theta - 60^\circ$	$150^\circ (11M/12)$	I_3	I_2
I_L	$210^\circ \leq \theta < 330^\circ$	$M/12 \leq m < 5M/12$	$\theta - 180^\circ$	$270^\circ (M/4)$	I_2	I_1
I_R	$270^\circ \leq \theta < 30^\circ$	$M/4 \leq m < 7M/12$	$\theta + 120^\circ$	$330^\circ (5M/12)$	I_3	I_2
I_R	$30^\circ \leq \theta < 150^\circ$	$7M/12 \leq m < 11M/12$	θ	$90^\circ (3M/4)$	I_2	I_1
I_R	$150^\circ \leq \theta < 270^\circ$	$11M/12 \leq m < M - 1, 1 \leq m < M/4$	$\theta - 120^\circ$	$210^\circ (M/12)$	I_1	I_3

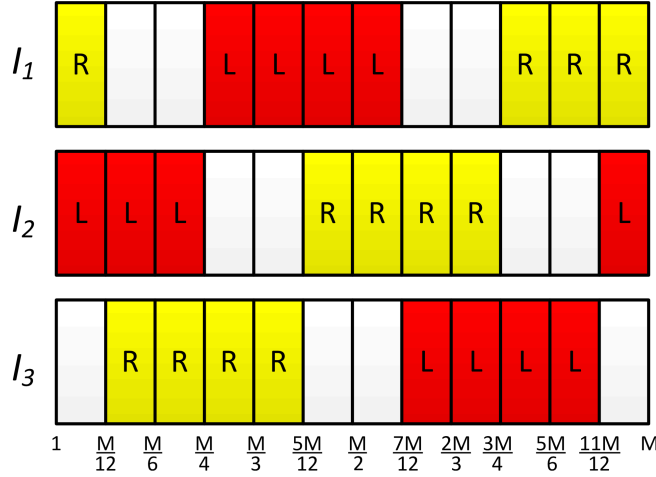


Figure 6. Panorama sections using in the mosaicking.

The rendering based on panoramic columns is possible due to the relative distance between the scene and the center of projection of each panorama. For an ideal coplanar panoramic triad scenario, scene elements projected at the center of each panoramic pair exhibit only horizontal disparities in the target zones for stereoscopic rendering due to the baseline between panoramic projection centers. When the scene is close to the camera, e.g. floors and ceilings, undesirable vertical disparities arise in distinctive viewing directions. In section 2.6, we propose a method to improve the visual comfort by applying a warping transformation to the omnistereoscopic image to compensate for the excessive vertical disparities. In addition, the vertical disparities between image sections create ghosting while blending (im_i, im_j) before the alignment angle θ_o . In the next section, we describe a method to reduce the ghosting by applying a warping transformation over the localized regions to blend.

2.5 Reducing the Ghosting when Blending Panoramic Sections

The rendering method described in the previous section is based on mosaicking of sections from different panoramic images as a function of the panning angle. Since these sections originate in cylindrical panoramas whose projection centers are not collocated, vertical disparities among correspondent features appear in the stereoscopic pair in different tilt directions. This effect is illustrated in Fig. 7-(a), where the projection center of I_j is translated $(b, 0, 0)$ with respect to the projection

center of I_i causing a vertical disparity δ between a correspondent scene element projected onto the planar images im_i and im_j as shown in Fig. 7-(b).

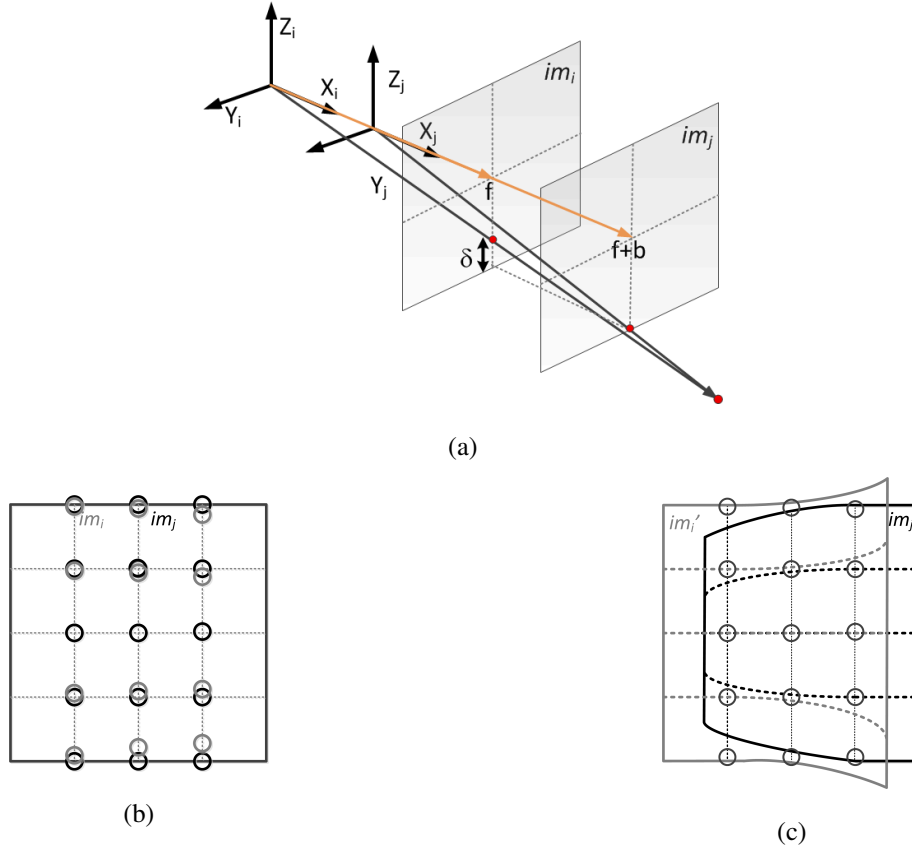


Figure 7. Image warping transformation to prevent ghosting: (a) a translation b in the direction of alignment between projection centers explain the vertical disparity δ which can be measured between corresponding features, (b) this vertical disparity creates ghosting after blending (im_i, im_j) if not corrected, (c) a proper warping transformation, based on compensating the vertical disparities reduces the ghosting likelihood.

For each column m_τ (Table 2), we can extract two cylindrical image sections im_i and im_j , as defined in section 2.4. Using the corresponding features between (im_i, im_j), we can define a pair of warping maps (W_{iv}, W_{jv}) using the criterion of maximum a posterior similarity between images. Note that a warping map is defined in this article as an array $W(n, m)$, with the same size of the image im , such as $W(n, m)$ contains the displacement to apply to the pixel $im(n, m)$ in the vertical or horizontal dimension. In this case, (W_{iv}, W_{jv}) re-map pixels vertically only, hence the horizontal disparity map is not affected. A better mapping between corresponding features is achieved after applying the warping map using the pixel retargeting transformation¹⁶ $T: im'_i = T(W_{iv}, im_i)$ and $im'_j = T(W_{jv}, im_j)$. The idea is illustrated in Fig. 7-(b) and (c).

An iterative method to obtain (W_{iv}, W_{jv}) can be adapted from the registration technique proposed by Shum and Szeliski.¹⁷ We used a similar approach, defining an energy function based on the mean squared difference between pixel intensities as a measure of the similarities between images. The recursive energy minimization defined under this constraint helps to obtain the desired warping maps.

The described method provides a solution to reduce the ghosting. However, there are residual vertical disparities on the stereoscopic image pair after mosaicking that need to be corrected.

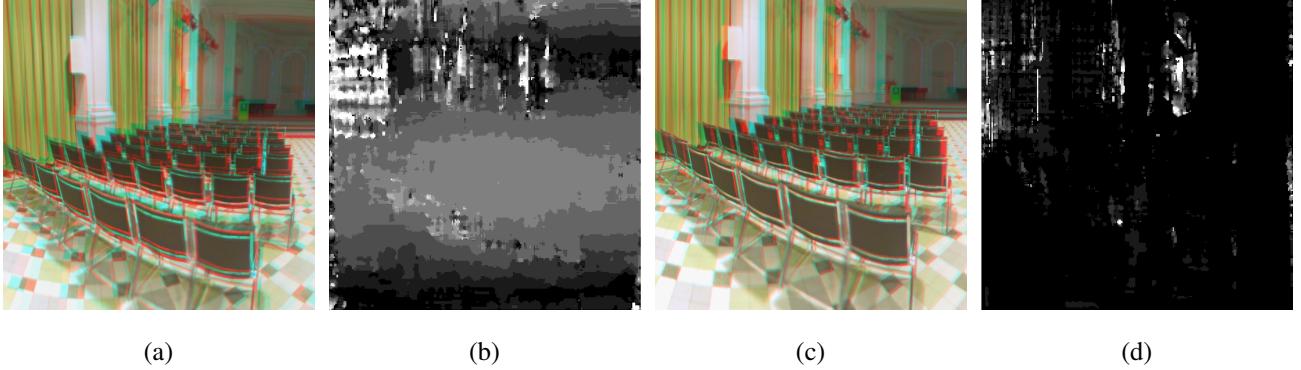


Figure 8. Vertical disparity correction: (a) a section of an omnistereoscopic image where the excessive vertical disparities need to be corrected (red-cyan anaglyph), (b) vertical disparity map extracted from the optical flow estimation where the darkest areas represent the lower absolute value of vertical disparity, (c) the same section of the omnistereoscopic image after applying the correction (red-cyan anaglyph), (d) vertical disparity map after the vertical corrections was applied; the majority of the post-warping disparity map exhibits less non-zero vertical disparities.

2.6 Reducing the Excessive Vertical Disparities

The excessive vertical disparities at certain angles in the omnistereoscopic pair is one of the causes of binocular discomfort and must be corrected. As in the previous case, excessive vertical disparities are more noticeable at the bottom part of the panorama ($n > 2N/3$) since floors are the closer scene’s elements registered. In addition, the undesired tilting in the camera vertical axis between snapshots may create noticeable vertical disparities distributed over the whole omnistereoscopic pair.

Using a similar solution as in the case of the ghosting reduction problem, a pair of vertical warping maps (W_{lv}, W_{rv}) can be used to re-map pixels in the vertical direction only in order to reduce the vertical disparities over the complete or partial omnistereoscopic pair. Like in the ghosting reduction method, horizontal disparities are not going to be affected by this transformation. Despite the similarities of this problem with the ghosting reduction problem, we used a different approach to find (W_{lv}, W_{rv}).

First, we estimated the dense image flow¹⁸ over (I_L, I_R) . The result of this process consists of three $(N \times M)$ signals: D_v and D_h corresponding to the vertical and horizontal disparity maps, respectively, and O corresponding to the pixel-wise occlusion map. The dense disparity maps (D_v, D_h) , as defined in this article, provide estimations of the displacement between corresponding pixels in both images. The occlusion map O contains a boolean estimation of whether a pixel is occluded or not, i.e. $O(n, m) = 1$ if the pixel at (n, m) is occluded, and $O(n, m) = 0$ if not. Since all these estimations are subject to noise, they cannot be used directly. Therefore, we divided the mapping into non-overlapping blocks of size $N_B \times M_B$ in order to estimate the disparities for each block using a majority vote method.¹³ Experimentally, we found that $\lfloor N/N_B \rfloor = 15$ and $\lfloor M/M_B \rfloor = 40$ enable to characterize the disparities in typical indoor scenes. Before estimating the vertical disparity for each block, we have to eliminate from the voting the occluded pixels of D_v using the information provided by O .

This majority vote approach helps to extract a $(\lfloor N/N_B \rfloor \times \lfloor M/M_B \rfloor)$ data set containing the block-wise vertical disparity estimation of the omnistereoscopic pair. Next, dividing the vertical displacement compensation between the two images and using a bilinear interpolation method to up-sample the obtained data-set, we created two $(N \times M)$ smooth warping maps (W_{lv}, W_{rv}) to re-map the image pair pixel-wise in the vertical dimension. As in the previous section, we used a retargeting transformation to apply the warping maps: $I'_L = T(W_{lv}, I_L(n, m))$ and $I'_R = T(W_{rv}, I_R(n, m))$. The new omnistereoscopic pair (I'_L, I'_R) exhibits less vertical undesired disparities preserving the original horizontal disparity map. An example of this vertical disparity correction technique is illustrated in Fig. 8.

2.7 Setting the Zero Parallax and Reducing the Excessive Horizontal Disparities

In some cases, it may also be necessary to compensate for excessive horizontal disparities. This is mostly due to using an incorrect baseline b , or to adapt the omnistereoscopic image to a certain display size and viewing distance from the screen, or to maximize the stereoscopic visual comfort.

The simpler approach to solve this problem is shifting both panoramas in different directions, i.e. $I_L(n, m - m_x)$ and $I_R(n, m + m_x)$, where m_x is determined by the horizontal disparity of the scene’s element chosen to have zero parallax. Considering that we are dealing with cylindrical panoramas, this strategy is equivalent to subtract $2 \cdot m_o$ from the whole horizontal disparity map, reducing the depth perception in all directions. This can be used as a quick automatic method to set the zero parallax point in the closer scene element. However, caution should be taken to prevent creating regions of negative parallax in the image, and possible window violation problems. After setting the zero parallax reference, the rest of the panorama will display positive parallax, creating the illusion that the scene depth progresses behind the display.

An elegant method to correct the horizontal disparities is using a non-linear disparity retargeting transformation as the one proposed by Lang et al.¹⁶ This correction implies estimating a pair of image warping functions (see section 2.6), but based on the horizontal disparity map D_h over the vertical disparity compensated pair (I'_L, I'_R) . A non-linear transformation applied to D_h to limit only those disparities above a maximum threshold value, can effectively reduce the binocular discomfort while maintaining the depth illusion on most of the registered scene. This threshold depends on many perceptual factors involving the visual comfort watching the binocular image in planar displays: distance from the display, size of the screen, interocular distance, etc.¹⁹ After this non-linear transformation over D_h , two new warping maps (W_{lh}, W_{rh}) are defined to re-map pixels of each image in the horizontal direction only. This is an effective method to improve the stereoscopic visual comfort, but at the price of distorting the depth relationships between scene elements.

3. EXPERIMENTAL RESULTS

In order to test the omnistereoscopic method proposed herein, we performed several tests in a variety of indoor scenarios. To do this, we used a Ladybug2²⁰ panoramic camera mounted on a tripod at 1.70 m above the floor level. Considering the vertical field-of-view of this camera, we estimated and then we found experimentally a safe zone of $\sim 2 m$ around the camera to prevent rendering problems. Notice that the minimum distance limitation of this camera is a problem in indoor scenarios, but such limitation is less likely to cause problems for rendering stereoscopic panoramas of outdoor scenes. The acquisition was done manually, using a triangular sampling pattern of side $b = 0.065 m$. In some trials, we mounted the camera off-center a distance $r = 0.038 m$ and captured three panoramic snapshots at regular intervals while rotating the camera. We used the camera calibration settings to project the multiple wide angle images acquired by the camera onto cylinders of radius $R = 5 m$. Each 3-D cylinder is represented by 2-D images of 4096 horizontal by 1304 vertical pixels.

The camera we used relies on overlapping wide-angle images on a 3-D surface to create a panoramic image, e.g. cylinders, spheres or cubes. In these cameras, the image radial distortions and the parallax created by off-set sensors introduce distortions in the panoramic image in different directions. This distortion is commonly neglected in single panoramic snapshots but becomes evident when comparing the scene registered in pairs of panoramas within a cluster. In our experiments, we found that each panorama exhibited distortions in different directions as result of the camera and the method used to capture the panoramic cluster. Therefore, we performed a fine alignment correction for each panoramic pair independently to minimize errors in the depth rendition as detailed in section 2.3.

Once we estimated the individual shifts to optimally center the panoramic triad according to the direction of alignment defined in Table 1, we mosaicked the panoramic sections as detailed in section 2.4. The ghosting in the boundaries between mosaicked sections was reduced by finding the appropriate warping transformation as described in section 2.5. The mosaicking result in an omnistereoscopic pair (I_L, I_R) for each cluster.

As we expected, these omnistereoscopic images exhibited different amount of vertical disparities depending on their distance to objects in the scene. Since the closer scene elements were located at the bottom of the panorama ($n > 2N/3$), we applied the vertical disparity correction method described in section 2.6 only to the bottom part of the omnistereoscopic pair. In some individual cases, a slight tilt in the camera axis during the acquisition created vertical disparities distributed over the whole panorama. In those cases, we applied the vertical compensation to the complete image pair.

In terms of the horizontal disparity correction, we restricted the adjustment to a horizontal shift in opposite directions of one or both panoramic in the omnistereoscopic pair to set the zero parallax reference at the bottom of the image. This was done automatically by finding the horizontal disparity of the non-occluded scene element registered at the bottom of the stereoscopic pair $(2 \cdot m_x)$ and shifting $I_L(n, m - m_x)$ and $I_R(n, m + m_x)$ as explained in section 2.7.



(a) http://www.luisgurrieri.net/stereo/tspi/flash-tabaret_05.html



(b) http://www.luisgurrieri.net/stereo/tspi/flash-tabaret_10.html



(c) http://www.luisgurrieri.net/stereo/tspi/flash-tabaret_06.html

Figure 9. Stereoscopic panoramas (red-cyan anaglyph) with links to flash 3-D navigable views.

Some representative results of our rendering method are presented in Fig 9. In addition, we have posted the results of an experimental stereoscopic sampling of an indoor scene in the companion website[†] of this article, along with a number of navigable omnistereoscopic viewpoints from this sampling campaign.

4. CONCLUSIONS AND FUTURE RESEARCH

The technique we have presented is an attractive alternative for the fast acquisition and rendering of stereoscopic panoramas using a cluster of three panoramic snapshots of the scene. Our approach can effectively reproduce the illusion of depth

[†]The companion website can be visited at: <http://www.luisgurrieri.net/publications/tspi/>

of a real-world scenario by using partially overlapped wide field-of-view stereoscopic samples of the scene obtained by controlling the panoramic sampling pattern. In addition, using a limited number of panoramic samples, we reduced the required data acquisition time, transmission bandwidth and storage capacity. The low computational complexity algorithms we proposed for aligning and rendering the omnistereoscopic images make this technique attractive for the large scale stereoscopic survey of a scene. Furthermore, we proposed methods to reduce ghosting while mosaicking sections, and to correct excessive vertical and horizontal disparities, which improve the overall visual comfort of viewing the final omnistereoscopic image.

The commercially available panoramic cameras enable to capture a sequence of panoramas at high speed. However, it is still necessary to move the camera head to obtain different panoramic viewpoints, which has to be done at a limited speed to prevent motion blur. When sampling outdoor scenes, having to stop the mobile platform to acquire one panoramic cluster at the time may be inconvenient. Moreover, in certain scenarios where the scene is not static, the sequential acquisition of a number of panoramas may be problematic. Alternative methods to enable acquiring omnistereoscopic images of dynamic scenarios in one single snapshot is an ongoing research.

ACKNOWLEDGMENTS

The authors would like to thank Quyen Sy and Chunxiao Zhang for their invaluable assistance in the acquisition of panoramic samples. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

- [1] Peleg, S. and Ben-Ezra, M., “Stereo panorama with a single camera,” *Proc. IEEE Conf. Computer Vision Pattern Recognition*, **1**, 395–401 (1999). [2](#), [3](#)
- [2] Huang, F. and Klette, R., “Stereo panorama acquisition and automatic image disparity adjustment for stereoscopic visualization,” *Multimedia Tools and Applications*, **47**(3), 353–377 (2009). [2](#), [3](#)
- [3] Ishiguro, H., Yamamoto, M., and Tsuji, S., “Omni-directional stereo,” *IEEE Trans. Pattern Anal. Machine Intell.*, **14**(2), 257–262 (1992). [2](#)
- [4] Southwell, D., Basu, A., Fiala, M., and Reyda, J., “Panoramic stereo,” *Proc. IEEE Int. Conf. Pattern Recognition*, **1**, 378–382 (1996). [2](#)
- [5] Fleck, S., Busch, F., Biber, P., Straßer, W., and Andreasson, H., “Omnidirectional 3D modeling on a mobile robot using graph cuts,” *Proc. of the IEEE International Conf. on Robotics and Automation (ICRA 2005)*, 1748–1754 (2005). [2](#)
- [6] Shevlin, F., McCullagh, B., Eadie, D., Navas-Herreros, M., and Rabaud, C., “Stereoscopic 3-D acquisition, processing, and display for telerobotic applications,” *Advances in telerobotics*, **31**, 91–105 (2007). [2](#)
- [7] Yamaguchi, K., Takemura, H., Yamazawa, K., and Yokoya, N., “Real-time generation and presentation of view-dependent binocular stereo images using a sequence of omnidirectional images,” *Proc. IEEE Int. Conf. Pattern Recognition*, **4**, 589–593 (2000). [2](#)
- [8] Hori, M., Kanbara, M., and Yokoya, N., “Novel stereoscopic view generation by image-based rendering coordinated with depth information,” *Proceedings of the 15th Scandinavian Conference on Image Analysis*, 193–202 (2007). [2](#)
- [9] Hori, M., Kanbara, M., and Yokoya, N., “Arbitrary stereoscopic view generation using multiple omnidirectional image sequences,” *Proc. IEEE Int. Conf. Pattern Recognition*, 286–289 (2010). [2](#)
- [10] Vanijja, V. and Horiguchi, S., “A stereoscopic image-based approach to virtual environment navigation,” *The Computer, the Internet and Management*, **14**, 68–81 (2006). [2](#)
- [11] Thompson, W., Fleming, R., Creem-Regehr, S., and Stefanucci, J. K., “Motion of viewer and objects,” in [*Visual Perception from a Computer Graphics Perspective*], ch. 11, 282–283 (2011). [5](#)
- [12] Pajdla, T. and Hlaváč, V., “Zero phase representation of panoramic images for image based localization,” *Computer Analysis of Images and Patterns*, 838–838 (1999). [7](#)
- [13] Gurrieri, L. E. and Dubois, E., “Optimum alignment of panoramic images for stereoscopic navigation in image-based telepresence systems,” *Proc. 11th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras (OMNIVIS 2011)*, **11**, 351–358 (2011). [7](#), [11](#)

- [14] McMillan, L. and Bishop, G., “Plenoptic modeling,” *Proc. of the 22nd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH’95)*, 39–46 (1995). 7
- [15] Seitz, S. and Dyer, C., “View morphing: Uniquely predicting scene appearance from basis images,” *Proc. Image Understanding Workshop*, 881–887 (1997). 7
- [16] Lang, M., Hornung, A., Wang, O., Poulakos, A., Smolic, A., and Gross, M., “Nonlinear disparity mapping for stereoscopic 3D,” *ACM Transactions on Graphics (TOG)*, **29**(4), 75–85 (2010). 10, 12
- [17] Shum, H.-Y. and Szeliski, R., “Construction of panoramic image mosaics with global and local alignment,” in [*Panoramic Vision*], Benosman, R. and Kang, S., eds., ch. 13, 227–268 (2001). 10
- [18] Ogale, A. S. and Aloimonos, Y., “A roadmap to the integration of early visual modules,” *Proc. IEEE Int. Conf. Computer Vision: Special Issue on Early Cognitive Vision*, **72**(1), 9–25 (2007). 11
- [19] Lambooij, M. T. M., IJsselsteijn, W. A., Fortuin, M., and Heynderickx, I., “Visual discomfort in stereoscopic displays: a review,” *Journal of Imaging Science and Technology*, **53**(3), 030201–14 (2009). 12
- [20] Point Grey Research Inc., “Ladybug2 panoramic camera.” Website (2011). <http://www.ptgrey.com/products/ladybug2/>. 12