The omnidirectional acquisition of stereoscopic images of dynamic scenes

by

Luis E. Gurrieri

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the Doctorate degree in Electrical and Computer Engineering

School of Electrical Engineering and Computer Science
Faculty of Engineering
University of Ottawa

Abstract

This thesis analyzes the problem of acquiring stereoscopic images in all gazing directions around a reference viewpoint in space with the purpose of creating stereoscopic panoramas of non-static scenes. The generation of immersive stereoscopic imagery suitable to stimulate human stereopsis requires images from two distinct viewpoints with horizontal parallax in all gazing directions, or to be able to simulate this situation in the generated imagery. The available techniques to produce omnistereoscopic imagery for human viewing are not suitable to capture dynamic scenes stereoscopically. This is a not trivial problem when considering acquiring the entire scene at once while avoiding self-occlusion between multiple cameras.

In this thesis, the term *omnidirectional* refers to all possible gazing directions in azimuth and a limited set of directions in elevation. The acquisition of *dynamic scenes* restricts the problem to those techniques suitable for collecting in one simultaneous exposure all the necessary visual information to recreate stereoscopic imagery in arbitrary gazing directions.

The analysis of the problem starts by defining an *omnistereoscopic viewing model* for the physical magnitude to be measured by a panoramic image sensor intended to produce stereoscopic imagery for human viewing. Based on this model, a novel *acquisition model* is proposed, which is suitable to describe the omnistereoscopic techniques based on horizontal stereo. From this acquisition model, an acquisition method based on multiple cameras combined with the rendering by mosaicking of partially overlapped stereoscopic images is identified as a good candidate to produce omnistereoscopic imagery of dynamic scenes.

Experimental acquisition and rendering tests were performed for different multiple-camera configurations. Furthermore, a mosaicking criterion between partially overlapped stereoscopic images based on the continuity of the perceived depth and the prediction of the location and magnitude of unwanted vertical disparities in the final stereoscopic panorama are two main contributions of this thesis. In addition, two novel omnistereoscopic acquisition and rendering techniques were introduced.

The main contributions to this field are to propose a general model for the acquisition of omnistereoscopic imagery, to devise novel methods to produce omnistereoscopic imagery, and more importantly, to contribute to the awareness of the problem of acquiring dynamic scenes within the scope of omnistereoscopic research.

Acknowledgements

This thesis could not be as it is without the help of Prof. Eric Dubois, whose expertise, understanding and constant support have made this document a reality. Through all these years as my supervisor, he has always found the time to maintain engaging conversations on variety of topics, some of which have helped to shape this thesis.

I would like to thank the participants of my thesis defence: Dr. Pierre Payeur, Dr. Anthony Whitehead, Dr. Jiying Zhao, Dr. Dimitri Adroutsos and Dr. Barry Schneider.

I would like to express my gratitude to my fellow students and especially to Chunxiao (Vicki) Zhang who volunteered to participate in my extensive data collection for this project. I must also give thanks to Prof. Jochen Lang, who gave his time to be interviewed for one of my projects and for his encouraging words during my presentations.

I must also acknowledge the support of Dr. Sima Noghanian who helped me during my graduate program at the University of Manitoba and has not stopped helping from a distance.

I would like to express my gratitude to those who believed in my personal project from the University of Buenos Aires. My deepest gratitude to Ing. Fernando Acero, who encouraged me to keep going ahead from early times, to Ing. Gabriel Venturino, who saw the potential, and to Ing. Atila P. J. Szabo, who opened the doors for us at the very beginning of our careers.

I would like to extend my gratitude to Ing. A. Hanke, Ing. D. Picornel and Ing. A. Zemelman, for keeping the connection alive all these years, and to all those who helped me in one way or another to reach this point.

Finally but not last, my special thanks to Quyen Sy for unconditionally helping me all these years. Without her help, nothing would have been possible.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The advent of panoramic cameras has enabled a variety of image-based applications in recent years. However, the acquisition and rendering of direction-independent stereoscopic imagery, in particular of real-time events is still a problem open to novel contributions. There are numerous technologies capable to produce high-quality omnistereoscopic imagery. However, only a few of them can create direction-independent images capable to provide the correct binocular cues of depth to the human viewer. Furthermore, most of these panoramic technologies are limited to static scenes, whereas a state-of-the-art omnistereoscopic camera should be suitable to capture dynamic scenes as well.

The problem is how to acquire an omnistereoscopic image of a non-static scene, with respect to a single reference viewpoint in space, but representing two viewpoints with horizontal parallax in every gazing direction. This challenging constraint imposed by our binocular visual system requires an acquisition technology capable to sample the whole visual scene at once in order to reconstruct stereoscopic views in arbitrary directions. In this thesis, I investigate techniques suitable to acquire the whole scene stereoscopically around a reference viewpoint in space and in one simultaneous exposure.

In this thesis, the term *omnidirectional* denotes imagery that can be rendered in any direction in azimuth around a reference viewpoint in space. This imagery can have 360° field-of-view (FOV) in the panning (azimuth) direction and a limited FOV in elevation. A wide-angle image or a *panoramic image* can be created from an omnidirectional image by defining a desired FOV in azimuth and elevation around a particular virtual gazing direction in azimuth. Additionally, a panoramic image can have a full (360°) field-of-view in azimuth. In the latter case, an omnidirectional and a panoramic image are the same. The term stereoscopic panorama

will be used interchangeably with omnistereoscopic imagery in this thesis.

My analysis of the problem starts by defining the signal that the omnistereoscopic image sensor needs to measure. This approach has led to an *omnistereoscopic viewing model* which is based on geometrical constraints imposed by our binocular visual system. Based on this model, I derive the physical signal to be measured by an omnistereoscopic image sensor, whether in dynamic or static scenarios.

In addition, based on the omnistereoscopic viewing model, I propose an *acquisition model*, which is used to describe a large variety of omnistereoscopic technologies based on horizontal stereo. By means of changing the geometric parameters of this acquisition model, I propose to summarize the main acquisition technologies of omnistereo into four arrangements of stereoscopic cameras. From these four variations, two model the sequential acquisition while the other two model the simultaneous acquisition of the whole scene. I use this model to contrast the available omnistereoscopic technologies in the context of their capability to acquire dynamic scenes.

One of the contributions of my thesis is to define the acquisition and rendering constraints based on the continuity of the perceived depth, which is applicable to the rendering by mosaicking of partially overlapped stereoscopic images. Another important contribution is the prediction of the location and magnitude of unwanted vertical disparities in the final stereoscopic panorama. Both results are important for rendering stereoscopic views in arbitrary gazing directions based on mosaicking partial stereoscopic views.

I also introduce two novel omnistereoscopic acquisition and rendering techniques. The first technique is based on sequentially acquiring three panoramic images in a co-planar triangular pattern. The second technique is a novel multiple-camera configuration suitable to capture non-static scenes, which is based on the simultaneous acquisition of partially overlapped stereoscopic images using a novel geometric distribution of cameras. Both techniques are supported with extensive experimental results.

In the next sections of this introductory chapter, I present the main problem to be addressed and its relevance in the area of panoramic vision, followed by a discussion of the existent solutions in the literature and the methodology used in this thesis. At the end of this chapter, I describe the organization of this thesis.

1.1 The problem to be addressed

The specific problem to be addressed is to find a suitable strategy to produce direction-independent stereoscopic imagery with horizontal parallax, from a single reference viewpoint in space. A general omnistereoscopic system must be capable to sample the light-field of a scene in order to at least approximate the trichromatic values from each eye's viewpoint and do so for every gazing direction. Additionally, the acquisition system should be able to capture dynamic scenarios. In summary, this thesis is going to study the problem of direction-independent acquisition of stereoscopic imagery in the context of dynamic scenes.

The problem of acquiring stereoscopic snapshots of dynamic scenes depends on the strategy used to sample two distinctive viewpoints, with horizontal parallax, in every gazing direction at once. Stereoscopic panoramic methods proposed in the past are partially inadequate because they are not conceived to acquire dynamic scenes.

A good strategy is to simultaneously acquire a number of partially overlapped stereoscopic snapshots of the scene and then mosaic[1] these samples to create a stereoscopic panorama. This approach can provide the necessary information to generate omnistereoscopic imagery and videos in real-time. However, there are a variety of aspects to take into account during the design of such a camera, e.g., the number of actual cameras necessary to cover the scene omnidirectionally, the field-of-view of each lens, and the geometric specification of the camera arrangement. These design considerations will determine the best rendering strategies, and the inherent limitations to render artifact-free stereoscopic views of near and far scene elements. Moreover, the geometry of the multiple-camera system will affect the rendering strategy. Finally, the consistency in the binocular illusion of depth is connected to the acquisition system and rendering method.

There are no analyses to justify the capabilities of various omnistereoscopic techniques to produce omnidirectional stereoscopic imagery. Therefore, an analytical framework to compare these cameras is needed. Such an analytical model will help to suggest rendering strategies to cope with the horizontal disparity continuity, mosaicking quality and stitching errors, i.e., ghosting due to stitching and blending, undesired sources of visual discomfort such as vertical disparities, among other problems.

Another constraint is the stereoscopic format to display the results. For instance, the rendered stereoscopic image can be a complete panoramic image or interactive stereoscopic views

[1] Mosaic in this context the process by which a set of partially overlapped images of the scene are stitched and blended together to form an image seemingly continuous in the azimuth direction.

of the scene at desired gazing directions. Each display scenario determines a particular model for the stereoscopic viewing. In any case, the acquisition system should provide the information necessary for the real-time or off-line stereoscopic rendering.

Wide-angle stereoscopic formats, i.e., cubic, cylindrical or spherical stereoscopic panoramas, are required in situations where the image is projected in an immersive display. The problem in this case is how to properly render wide-angle stereoscopic images of the scene in order to provide consistent and undistorted binocular cues of depth to the user.

The interactive stereoscopic navigation involves the real-time rendering of stereoscopic images corresponding to left and right eyes for each gazing direction. This scenario is more suitable for interactive applications such as interactive three-dimensional television (3DTV[2]) [1], and image-based stereoscopic virtual environments [2]. In both cases, the experience is individual and supported by feedback information of the user's gazing direction, e.g., a head mounted display equipped with an eye-tracking device.

A multiple-camera system introduces new challenges. For instance, some camera configurations are limited to render stereoscopic views only when the scene is far from the camera. Paradoxically, the human binocular perception of depth uses the horizontal disparity between elements in the foreground of the scene as seen by each eye. In far distant scenes, these binocular cues become secondary. In those cases, other monocular cues of depth are predominant, i.e., perspective, blurring, shading, and even colour contrast. Therefore, the use of certain multiple-camera configurations for omnistereoscopic photography is marginal. More studies are required to understand the limitations and advantages of different multi-cameras configurations.

In this thesis, I am going to focus on suitable strategies to sample the necessary information to produce omnistereoscopic views of dynamic scenarios. The main problem is to define the visual information necessary to sample in order to produce direction-independent stereoscopic imagery intended for human viewers.

1.2 The relevance of this research

The importance of this research is justified by the fact that most real-world scenes are intrinsically dynamic and the large majority of the panoramic cameras and rendering techniques proposed over the last decade are inadequate to capture omnistereoscopic snapshots of such

[2]The complete list of acronyms and abbreviations can be found in Appendix D

events. Recent multiple-camera proposals promise solutions for this problem, but their capabilities and feasibility are not supported by a model. Hence, there is a need for an omnistereoscopic camera configuration suitable to capture real-time events (dynamic scenes) and, moreover, there is a need for a theoretical framework to understand the pros and cons of such cameras.

The commercial development of single-snapshot panoramic cameras has already facilitated the popularization of image-based virtual tours and virtual navigation applications, i.e. Google's StreetView [3], MapJack [4] and EveryScape [5] to mention some of them. These virtual tours require the acquisition of a large number of panoramic snapshots of the scene to facilitate the free navigation in image-based simulations of remote world locations [6]. A camera mounted on a mobile platform, e.g., cameras mounted on cars, bicycles, or robots, are generally used to capture a large number of (monoscopic) panoramic snapshots. Stereoscopic virtual tours can be made in a similar manner, provided the technology to acquire large number of omnistereoscopic samples becomes available.

The addition of stereoscopic visualization will improve the realism of image-based simulations, e.g., image-based backgrounds for stereoscopic video production, training simulations using stereoscopic snapshots of remote locations, educational applications, stereoscopic telepresence, etc. The benefits of omnistereoscopic media can be extended to other applications such as free-viewpoint television [7], which can use an omnistereoscopic camera to provide interactive 3DTV to the viewer, with the additional capability to navigate the scene at will. A further improvement will be to propose an omnistereoscopic camera that can provide the means to adjust the rendered depth of a scene. Although the impact in multimedia systems of this technology is evident, a panoramic camera capable to capture omnistereoscopic snapshots has not been made commercially available yet to my best knowledge.

The virtual navigation of remote-world locations is another example that shows the need for an omnistereoscopic technology to acquire real-time imagery. For instance, in a telepresence system such as the one illustrated in Fig. 1.1, an omnistereoscopic camera is used to acquire a large number of samples of a remote location, which are likely dynamic scenes. The collected visual information can be stored for future access, or processed in real-time depending of the application. For instance, the stereoscopic navigation of this virtual environment will be based on dynamically rendering stereoscopic views in the chosen gazing directions.

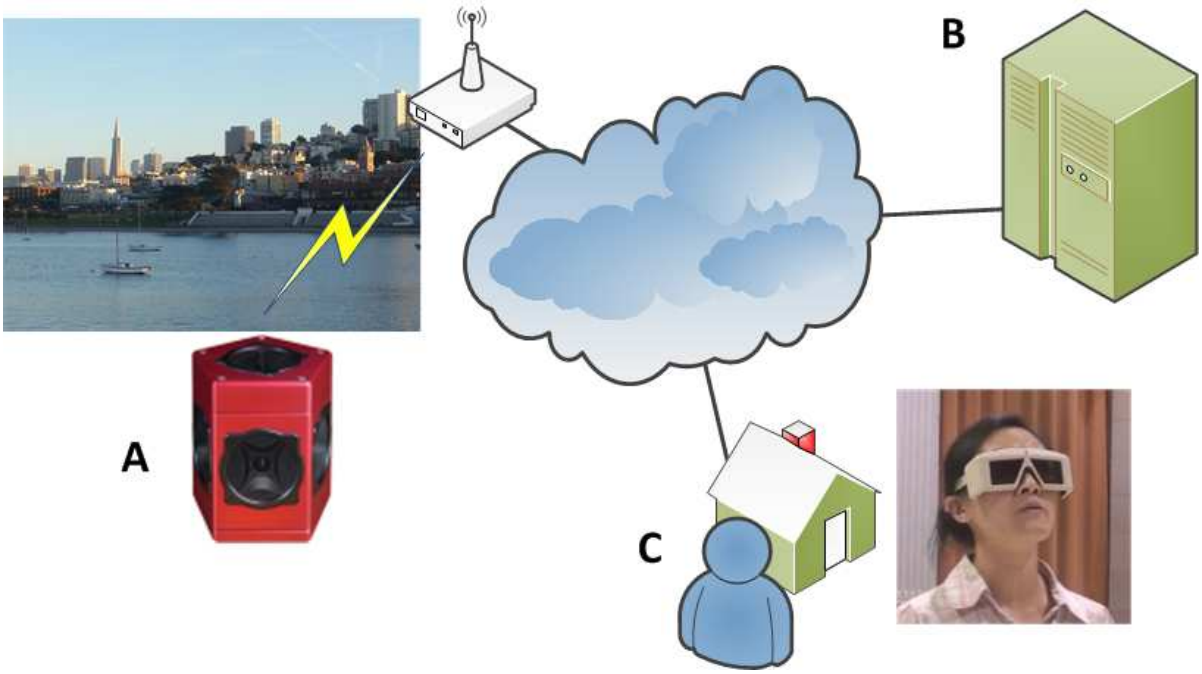


Figure 1.1: Omnistereoscopic telepresence: (A) the omnistereoscopic camera samples the scene at a remote location, sending images to remote server (B) where they are stored for future access by a remote user (C) who explores the virtual scene choosing the gazing direction.

1.3 Solutions in the literature

Over the last decade, problems such as robot navigation and omnidirectional 3D-scene recovery originated strategies for omnistereoscopic depth map acquisition [8, 9, 10]. Although they are attractive for their simplicity, these methods are in general inadequate to simulate human stereopsis directly since they are based on multiple views with vertical parallax. Occlusion handing and other problems arise when trying to adapt these techniques to create omnistereoscopic views of the scene.

Some technologies rely on the sequential acquisition of narrow image columns from two viewpoints. The horizontal parallax between the viewpoints makes this method suitable to emulate binocular visual mechanisms [11, 12, 13, 14, 2]. These techniques can produce high-resolution omnistereoscopic imagery but only for static scenes. Although some of these configurations can be implemented with multiple cameras for simultaneous omnidirectional acquisition, the number of necessary samples makes most of these approaches impractical [15].

In more recent times, the interest in new immersive and interactive media has motivated several proposals suitable to capture real-time omnistereoscopic imagery [16, 17, 18, 19]. Some recent proposals based on multi-camera arrangements, were conceived specifically to satisfy the dynamic scene sampling constraint. These configurations mostly come from patents and technical reports.

In this context, there is a lack of models to reflect the impact of the multiple-camera geometry in problems such as the depth consistency based on binocular cues and other factors that will affect the visual comfort such as ghosting by stitching of adjacent views and vertical disparities. There have been attempts to model the omnidirectional acquisition, but for sequential acquisition only [20, 21]. A model for the omnidirectional viewing signal based on the geometric constraints imposed by the human visual system has not been proposed in the literature.

There have also been some attempts to model the depth distortions in stereoscopic displays based on estimations of the binocular depth resolution in humans [22, 23]. However, these studies were constrained to particular displays and rendering systems. A model applicable to different acquisition configurations has not been proposed.

1.4 The method

The main hypothesis of my thesis is that omnistereoscopic imaging of real-time events can be satisfactorily approximated using a limited number of partially overlapped stereoscopic snapshots of the scene. Given that acquisition strategy, the most attractive rendering method is based on mosaicking the acquired samples. This can be justified by the low computational requirements of this rendering approach. But this particular acquisition and rendering combination needs to provide a satisfactory immersive experience to the user. Therefore, the omnistereoscopic results have to be free of stitching artifacts and have to provide consistent binocular cues of depth in every direction. These are the main problems I am studying in this thesis.

My approach to study the problem is to propose first a model that sets the geometric constraints for the *omnidirectional viewing function* that represents the stereoscopic visual information to be acquired. The development of this model starts from the plenoptic function to propose a multidimensional array of real-valued signals, that, when sampled by a panoramic sensor, provides the necessary information to recreate a binocular image in any desired gazing direction.

Based on the geometric constraints of the *omnistereoscopic viewing model*, I propose a general *acquisition model* applicable to variety of sequential and simultaneous acquisition methods. This acquisition model is based on a pair of pinhole cameras with horizontal parallax between their projection centers.

Using this *acquisition model*, I compare the different omnistereoscopic technologies available in the literature. From this comparison and given the objective of acquiring real-time events, I justify the configurations based on multiple cameras as one of the best candidates for the task. Then, I propose using a rendering approach based on mosaicking since its low complexity enables to produce real-time imagery.

In order to test the proposed acquisition and rendering approach, I discuss the development of the complete mosaicking pipeline. Even though the rendering is not a core element of this thesis, its implementation is necessary to produce omnistereoscopic imagery as a proof-of-concept. Moreover, the rendering helps me to identify important issues to address for each acquisition configuration. Consequently, I emulated each of the four basic acquisition configurations derived from the acquisition model and I tested the rendering blocks on the acquired samples.

During the rendering experiments, I identified two main aspects to model: the problem of depth continuity after mosaicking partially-overlapped images and the modeling of the vertical disparities that affect the visual comfort.

In order to test the usability of the mosaicking approach, I developed a method to define the required number of stereoscopic images to be mosaicked based on the limits of the human depth resolution. My first approach to model the problem relies on two simplified 3D-models of the scene. In a different approach, I use the projective model derived from each variation of the *acquisition model*. Based on the latter, I propose a ray-tracing method to find the relationships between the scene depth and the optimum location of stitching in each image. I used the consistency in the horizontal disparity between stitched stereoscopic images as the criterion to study these relationships.

Using the acquisition model, I estimate the magnitude of undesired vertical disparities as function of the gazing direction. This approach helped me to model the regions where the vertical disparities are negligible and those where they are more severe.

Finally, two novel approaches for the omnistereoscopic acquisition and rendering are presented as part of this thesis.

A note has to be made regarding the thesis scope. The subjective evaluation of the visual

comfort in human is beyond the scope of this thesis. Neither is part of this thesis the rendering algorithms and their implementation. Although the mosaicking implementation has been instrumental to test and to produce a proof-of-concept of the acquisition configurations, their optimization is beyond the scope of this thesis.

1.5 Contributions

The following list summarizes the original contributions of this thesis to field of stereoscopic panoramas.

- I proposed a model for the direction-independent stereoscopic viewing of any scene, which is applicable to static and dynamic scenes. From this model, I defined the physical signal to be measured by an omnidirectional image sensor intended for human stereoscopic vision. This approach for modeling this omnidirectional stereoscopic signal has not been attempted before.

- In addition, I proposed a general model for the omnistereoscopic image acquisition based on projective geometry. This model can describe a large variety of single or multiple-camera configurations and it was used to contrast the available technologies to produce direction-independent stereoscopic imagery intended for human viewing. This is another original contribution.

- Based on the acquisition model, I proposed the mosaicking of a number of partially overlapped images of the scene as a suitable method to acquire and render omnistereoscopic imagery. Hence, I focused on the problems arising when mosaicking these partially overlapped images. This approach has been used recently, but to my best knowledge this is a novel attempt to identify the problems of acquiring real-time imagery. Hence, this is one of the most important contributions of this thesis.

- Although the rendering by mosaicking partial stereoscopic snapshots of the scene is not part of the core of this thesis, the unsuitability of the available commercial panoramic stitching products to produce omnistereoscopic imagery made me propose and implement a suitable rendering pipeline. This stereoscopic mosaicking technique is also novel and therefore is one of the contributions of this thesis.

- I identified and proposed a method to obtain the acquisition system constraints necessary to consistently render depth cues around the reference viewing point. The ray-tracing approach I used to study this problem as well as the acquisition model in which it is supported are both novel contributions.

- I proposed methods to estimate the number of stereoscopic images necessary to acquire for the mosaicking based on the distance between the scene and the camera. This estimation is based on two basic 3D-models of the scene and the development of thresholds for the human depth resolution based on binocular cues. The use of the human depth resolution to determine the usable mosaicking width has not been used before in this problem to my best knowledge.

- In addition, I modeled the vertical disparity based on the general acquisition model. The vertical disparities have been attributed as one of the main sources of visual discomfort in stereoscopic imagery. The approach used in the modeling of this effect is a novel contribution.

- I proposed a novel multiple-camera approach suitable for rendering by mosaicking and for the off-line synthesis of stereoscopic views of the scene in any gazing direction. The stereoscopic conversion (2D-to-3D) is an ongoing research. However, its application to dynamically create omnistereoscopic imagery is a new approach. Furthermore, the acquisition configuration I propose for this purpose is also novel.

- Also, I proposed a novel sequential acquisition method based on clusters of three panoramas that improves similar sequential acquisition approaches (Appendix A). This approach has been presented before for four cameras in a square cluster. The cluster I proposed reduces the number of necessary panoramas to three and it has not been proposed before to my best knowledge.

- Finally, I proposed an uncalibrated method to align pairs of panoramas (Appendix B). Other methods have been proposed to align panoramas in different formats to arbitrary directions in the scene. All these methods are based on assuming a pinhole camera and estimating from it the calibrations parameters to obtain a fundamental matrix and defining a homography between panoramas. The method I proposed is different because it is based on identifying the unique directions of *focal expansion* and *contraction* in a dense disparity map defined between panoramas. The obtained direction in azimuth is unique

and it enables to identify the regions usable for stereoscopic mosaicking in any pair of panoramas with coplanar projection centers and parallel vertical axes. This technique does not assume any camera model and it is uncalibrated. To my best knowledge, this panorama alignment method is another novel contribution of this thesis.

1.6 Thesis organization

A rationale of this thesis organization that has motivated the order of appearance and contents of each chapter is provided next.

Chapter 2: Viewing functions

In this chapter, I first present the derivation of the *omnidirectional viewing model* $(\mathbf{C})$ that represent the multidimensional signal to be measured from a panoramic sensor located in a reference viewpoint in 3D-space. After that, I extend this model to the define the *omnistereoscopic viewing model* $(\mathbf{C_S})$ which encompasses all the constraints necessary to define the signal to be measured by an omnistereoscopic sensor. I called this multidimensional real-valued signal the *omnistereoscopic viewing function*.

Chapter 3: Acquisition model

In this chapter, I use the model for $\mathbf{C_S}$ to propose a *general acquisition model*. This model is based on a pair of pinhole cameras with parallel optical axes and co-planar projective centers. The majority of the omnistereoscopic cameras can be modeled by changing the relative spatial location of the two pinhole cameras on the reference plane where they are located. The four different spatial variations of the general acquisition model are denoted as *configurations*. Finally, I detail the transformations between coordinate systems in the stereoscopic pair as well as the general notation used through this thesis.

Chapter 4: Review of the omnistereoscopic technology

In this chapter, I present a review of the different technologies capable to produce omnistereoscopic imagery based on the *acquisition model* introduced in the previous chapter. In order to classify and compare the different technologies, I define four families based on the strategies

used to acquire the necessary information to produce omnistereoscopic imagery. These families are: catadioptric cameras, sequential methods, methods based on panoramic images, and multiple cameras systems. I evaluate each of the four families in terms of their capabilities to acquire and produce omnistereoscopic imagery of dynamic scenes. From this comparison, I chose the multiple cameras approach as the best candidate in the context of the proposed problem.

Chapter 5: Disparities and rendering approaches

In this chapter, I formally introduce the corresponding coordinate points and their disparities in a stereoscopic image. Moreover, I derive the depth resolution limits based on the human binocular vision system. From this analysis, I propose valid thresholds for the horizontal disparity based on the resolution limits of the human binocular system. In addition, I discuss three related but different rendering approaches in decreasing order of complexity. From this comparison, I conclude that the mosaicking approach has advantages for the rendering of real-time omnistereoscopic imagery using the images acquired by a multiple-camera system. Finally, as a first approximation to calculate the number of stereoscopic images, I apply the proposed thresholds for human depth resolution obtained at the beginning of this chapter.

Chapter 6: Mosaicking of stereoscopic images

In this chapter, I present the mathematical models used for the alignment of left and right image sequences. In addition, I introduce formal definitions for the set of correspondence image coordinates in the overlapping area between consecutive image samples. Finally, I discuss the need for the stereoscopic registration and the approaches used for image stitching and blending.

Chapter 7: Experimental acquisition and rendering

I present in this chapter the result of the acquisition experiments based on the variations of the general acquisition model introduced in Chapter 2. In order to produce stereoscopic panoramas with the acquired image samples, I detail the implementation of the complete rendering pipeline based on mosaicking. At the end of this chapter, I present examples for the four acquisition configurations as a proof-of-concept.

Chapter 8: Omnistereoscopic field-of-view and disparities

In this chapter, I define the *stereoscopic* and *omnistereoscopic* field-of-view. In addition, I present the mathematical formulation and a ray-tracing approach to calculate the minimum distance between the camera and the scene based on the horizontal disparity continuity in all gazing directions. In order to do this, I use the projection on the scene on each camera, which is based on the general *acquisition model*, combined with the threshold for the resolvable horizontal disparity proposed in Chapter 5. Finally, I derive a model for the vertical disparities on the overlapping regions between images.

Chapter 9: Depth consistency and vertical disparities

This chapter is dedicated to compare the performance of the different acquisition configurations. In order to do this, I emulated the different acquisition configurations based on the parameters of three different off-the-shelf cameras and lenses. Actually, I used one of these cameras and lens combination in the experiments presented in Chapter 7. This comparison is based on the ray-tracing method discussed in Chapter 8 to estimate the optimal location of the stitching position based on distance between the camera and the scene. Finally, I present an analysis of the magnitude and value of the expected vertical disparities.

Chapter 10: Conclusions

In this concluding chapter, I discuss the results presented in this thesis as well as the aspects that require further investigation.

Appendix A: Experimental acquisition based on panoramic clusters

In this appendix of my thesis, I present the sequential method based on three panoramas in a triangular-shaped acquisition pattern. This was a initial research on omnistereoscopic image acquisition that I did to test the mosaicking of wide-angle stereoscopic snapshots extracted from panoramic images in a controlled pattern. This work is a valid contribution to research on omnistereoscopic imagery, but it is not suitable to capture dynamic scenes. Therefore, I decided to present it as an appendix of the main thesis.

Appendix B: Alignment of panoramas for best stereoscopic usability

This appendix contains the development of the panoramic alignment used in the Appendix A. The alignment method I present here is one of the novel contributions of this thesis, but since it was developed for the panoramic triad method, I decided to present it as an appendix.

Appendix C: Displaying stereoscopic images

This appendix contains clarifications of the industry definitions for the rendering of stereoscopic imagery. These definitions have been originated in the 3D-movie industry and nowadays have widespread usage.

Appendix D: Acronyms and Abbreviations

This appendix contains an explanation of the typographic convention used in this thesis and a summary of the acronyms and abbreviations.

Chapter 2

Viewing functions

In order to analyze and compare different omnistereoscopic acquisition techniques, first it is necessary to define the physical magnitude to be sampled by the omnidirectional sensor, which must be defined as a signal function. In order to define this function, I am going to model the signal that contains the necessary information to reconstruct an omnistereoscopic view of the scene. The sampling of this signal is the first step towards producing stereoscopic images capable to stimulate the human binocular vision system for all gazing directions around a reference viewpoint.

In this chapter, I am going to derive the *omnidirectional viewing function* to be sampled by an omnistereoscopic sensor starting from the plenoptic model. The omnidirectional viewing function can be used to produce a monoscopic panoramic view of the scene. Using this function, I am going to extend the model to define an *omnistereoscopic viewing function* to describe the color components of the light to be measured from two viewpoints with horizontal parallax, for any gazing direction around a reference point in 3D-space.

This is a novel approach to describe the problem of omnistereoscopic image acquisition which is one of the contributions of this thesis.

The omnistereoscopic viewing function will help to define in the next chapter a general acquisition model that encompasses the geometric constraints for human binocular viewing of the scene.

2.1 The plenoptic model

The plenoptic function $\mathscr{P}$[1] models the light arriving at any viewpoint in 3D-space from any direction [24]. The plenoptic function is defined as

$$\mathscr{P}(\theta, \phi, \lambda, \mathbf{r}, t), \tag{2.1}$$

where $\mathbf{r}$ is a the viewpoint location with respect to $\mathbf{XYZ}$ in $\mathbb{R}^3$, θ (azimuth) and ϕ (elevation) angles define the *gazing direction* $\mathscr{D}$ ($\mathscr{D} \in \mathbb{S}^2$), where $\mathbb{S}^2$ is the unit sphere ($S^2 = \{\mathbf{r} \in \mathbb{R}^3 : \|\mathbf{r}\| = 1\}$), λ is the light wavelength ($\lambda \in \mathbb{R}$) and t is the instant ($t \in \mathbb{R}$). As defined, $\mathscr{P} : \mathbb{S}^2 \times \mathbb{R}^3 \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ models the directional *power spectral density* of light arriving at any point in space.

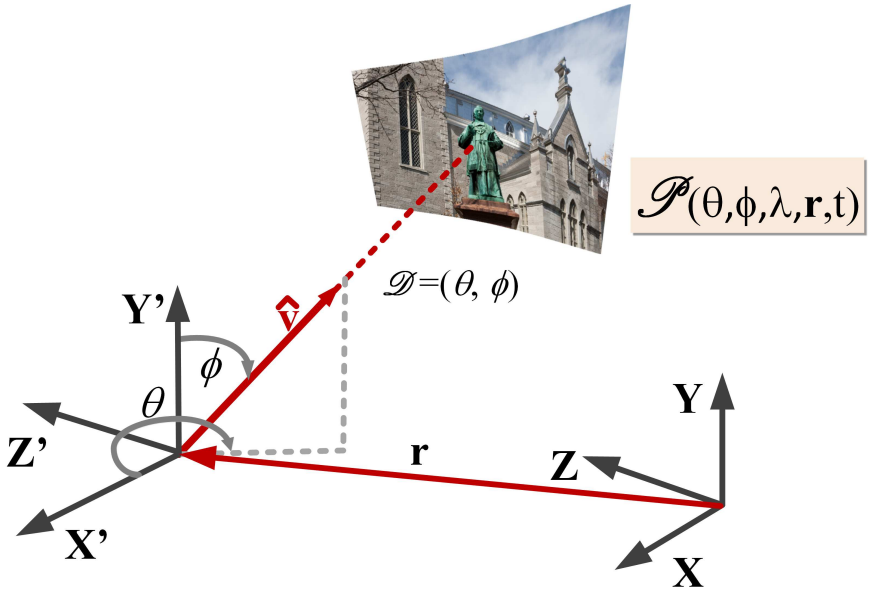


Figure 2.1: The plenoptic function $\mathscr{P}$ models the power spectrum of light arriving at a singular viewpoint in 3D-space from the direction $\mathscr{D}$.

A graphical interpretation of the plenoptic model is illustrated in Fig. 2.1. The gazing direction $\mathscr{D}$ is measured with respect to the viewpoint point located at $\mathbf{r}$. This implies measuring $\mathscr{D}$ using a reference of coordinates $\mathbf{X'Y'Z'}$ centered at $\mathbf{r}$ with respect to a global reference frame $\mathbf{XYZ}$, where the frame of coordinates transformation is a pure translation $\mathbf{T} = \mathbf{r}$. The unit vector $\hat{\mathbf{v}}$, as shown in Fig. 2.1, indicates the gazing direction with respect to $\mathbf{X'Y'Z'}$.

[1]The typographic conventions used in this thesis are explained in Appendix D

The multi-dimensionality characteristic of the plenoptic function is lost when sampled by a regular photographic camera, unless $\mathscr{P}$ is captured by an image sensor specifically designed to prevent that dimensionality reduction [25]. As a matter of fact, this dimensionality reduction is part of the derivation of the omnistereoscopic viewing function as is shown next.

2.2 The viewing function

The first step towards defining the signal to be measured by an omnidirectional sensor is to define the set of signals to be measured in a given color space. Even though the color space is specific to the sensor device used, this approach leads to a general expression of a measurable signal which is applicable to all kinds of image sensor, whether omnidirectional or not.

A given color space $\mathscr{C}$ is defined by a set of color matching functions $\Phi_n(\lambda)$, where $n \in \{1, 2, 3\}$ [26]. For this analysis, I use a trichromatic color space without loss of generality. The functions c_n represent the measurable color information in the sensor sensitivity range, from λ_{min} to λ_{max}, which can be obtained from $\mathscr{P}$ as follows

$$c_n(\theta, \phi, \mathbf{r}, t) = \int_{\lambda_{min}}^{\lambda_{max}} \mathscr{P}(\theta, \phi, \lambda, \mathbf{r}, t) \cdot \Phi_n(\lambda)\, \mathrm{d}\lambda, \tag{2.2}$$

where $c_n : \mathbb{S}^2 \times \mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}$, for the n^{th}chromatic component ($n \in \{1, 2, 3\}$), are functions of $\mathbf{r}$, the gazing direction $\mathscr{D}\ (\theta, \phi)$ and time t.

A trichromatic color basis represents the typical color vision in the human population. However, this biological parallel is not related to the obtention of these tristimulus functions c_n. The color space $\mathscr{C}$ used is strictly dependent on the image sensor technology.

The result of removing the wavelength dependency of $\mathscr{P}$ are the tristimulus functions of the arriving light. Light sensor devices such as digital cameras measure these tristimulus functions over an exposure time Δ_t. For instance, this is done by accumulating voltage proportional to the light intensity in each color component over a given Δ_t in a charged-coupled device (CCD) sensor array. Spatial multiplexing is normally used to sample simultaneously each chromatic component of the image. This is done by using color filter arrays (CFA) such as the widely used Bayer filter and other well known spatial multiplexing techniques.

2.2.1 Omnidirectional viewing function

The *omnidirectional viewing function* $\mathbf{C}$ models the signal to be measured by any omnidirectional image sensor. This function can be defined as an array of tristimulus functions

$$\mathbf{C}(\mathbf{u}) = (c_1(\mathbf{u}), c_2(\mathbf{u}), c_3(\mathbf{u}))^T, \tag{2.3}$$

where $\mathbf{u} = (\theta, \phi, \mathbf{r}, t)$.

An ideal omnidirectional image sensor located at $\mathbf{r}$ integrates the three components of $\mathbf{C}$ over a given exposure time Δ_t, obtaining a snapshot or video frame of the time-varying sequence. The vector function obtained by the sensor is defined as follows

$$\mathscr{I}(\theta, \phi, \mathbf{r}) = \int_{\Delta_t} \mathbf{C}(\theta, \phi, \mathbf{r}, t)\, \mathrm{d}t, \tag{2.4}$$

$$= (\mathscr{I}_1, \mathscr{I}_2, \mathscr{I}_3)^T, \tag{2.5}$$

where $\mathscr{I}_n$ is defined as follows

$$\mathscr{I}_n(\theta, \phi, \mathbf{r}) = \int_{\Delta_t} c_n(\theta, \phi, \mathbf{r}, t)\, \mathrm{d}t, \tag{2.6}$$

where $n \in \{1, 2, 3\}$. Each function $\mathscr{I}_n$ corresponds to the intensity of light of the chromatic component n at $\mathbf{r}$ as a function of the gazing direction $\mathscr{D}$. Therefore, the function array $\mathscr{I} : \mathbb{S}^2 \times \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is a panoramic image of the scene that satisfies the constraints of monocular vision.

2.2.2 Panoramic acquisition example

A digital panoramic camera measures $\mathbf{C}$ to reconstruct a panoramic image $\mathscr{I}$ which can be mapped onto variety of convex surfaces. For example, a cylindrical representation covers $360°$ in azimuth and a partial FOV in elevation. Alternatively, it can be mapped onto any curved surfaces topologically equivalent to a cylinder.

Alternatively, a partial FOV of the scene can be mapped onto a planar screen, allowing the user to change $\mathscr{D}$ while exploring the scene relative to a viewpoint located at $\mathbf{r}$. This is illustrated in Fig. 2.2 where $\mathbf{C}$ is measured for the viewpoint $\mathbf{r}$ by integrating c_n over Δ_t. The acquisition is limited to a range of azimuth and elevation angles around the gazing direction $\mathscr{D}$. The sampling of this section of the spherical space around the viewpoint produces a discrete multidimensional array $\mathbf{I} : \mathbb{Z}^2 \rightarrow \mathbb{R}^3$.

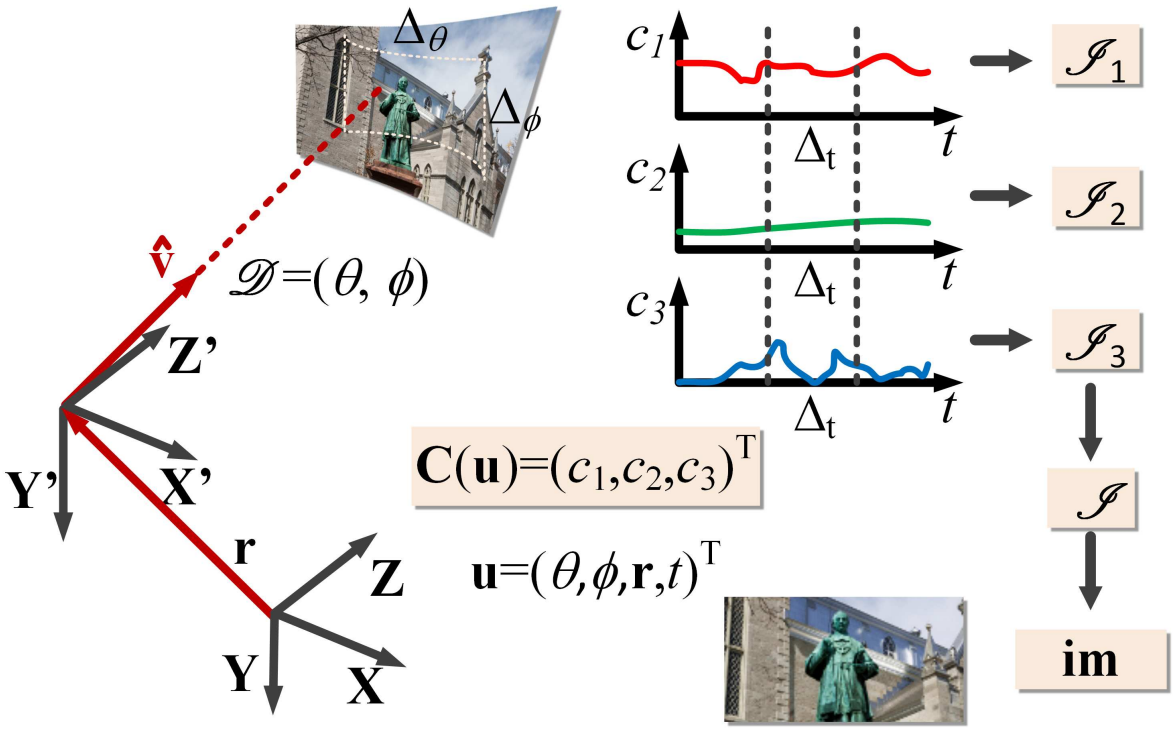


Figure 2.2: Acquisition of the omnidirectional image $\mathscr{I}$ from the viewpoint $\mathbf{r}$: the discrete image $\mathbf{I}$ represent a partial FOV of the scene in a gazing direction $\mathscr{D}$.

The advent of panoramic photography in recent years has enabled the measurement of $\mathbf{C}$. However, the available acquisition techniques for omnidirectional imagery enable only to generate approximations of $\mathbf{I}$ with different degrees of distortion. Many panoramic techniques are based on acquiring multiple images obtained from close but distinct viewpoints instead of a singular viewpoint. Other approaches such as catadioptric based systems acquire the scene omnidirectionally from a singular viewpoint, but at the price of introducing distortions on $\mathbf{I}$.

The omnidirectional viewing function $\mathbf{C}$, although valid only to model the acquisition of omnidirectional monoscopic imagery, can be used to propose a model for the omnistereoscopic viewing function, which is the goal of this chapter.

2.3 Omnistereoscopic viewing function

The *omnistereoscopic viewing function* $\mathbf{C_S}$ models the visual information to be acquired to reconstruct an omnidirectional stereoscopic image of the scene $\mathscr{I}_S$. As a difference with $\mathscr{I}$,

which can be used to render monoscopic images in any $\mathscr{D}$, $\mathscr{I}_S$ must provide the user with binocular visual cues in any gazing direction around the acquisition point.

The function $\mathbf{C_S}$ can be defined as an arrangement of a pair of omnidirectional viewing functions, $\mathbf{C_L}$ and $\mathbf{C_R}$ defined at two viewpoints in 3D-space. These viewpoints have horizontal parallax with respect to a reference viewpoint $\mathbf{r}$. In this model, the location of the left (L) and right (R) viewpoints are denoted $\mathbf{r}_L$ and $\mathbf{r}_R$, respectively. Although the subindexes $\{L, R\}$ respectively refer to the left and right viewpoints, this classification depends on the location of the viewer with respect to the camera coordinate frame. The omnistereoscopic viewing function can be defined as

$$\mathbf{C_S} = (\mathbf{C_L}(\mathbf{u_L}), \mathbf{C_R}(\mathbf{u_R}))^T, \tag{2.7}$$

where $\mathbf{C_L}$ and $\mathbf{C_R}$ are omnidirectional viewing functions, $\mathbf{u_L} = (\theta_L, \phi_L, \mathbf{r}_L, t)$ and $\mathbf{u_R} = (\theta_R, \phi_R, \mathbf{r}_R, t)$ are the input parameters for left and right omnidirectional viewing functions, respectively. A depiction of the omnistereoscopic viewing model is presented in Fig. 2.3.

2.3.1 Gazing vector

The *gazing vector* $\mathbf{v}$ that connects the scene point $\mathbf{P_W}$ with the reference viewpoint $\mathbf{r}$ can be defined as

$$\mathbf{v} = \mathbf{r} - \mathbf{P_W}. \tag{2.8}$$

The magnitude $v = \|\mathbf{v}\|$ defines the fixation distance and it will be used to determine the relative skew ω or *vergence angle* between left and right optical axes.

The unit vector $\hat{\mathbf{v}}$ introduced in Sec. 2.1 can be formally defined as

$$\hat{\mathbf{v}} = \frac{\mathbf{v}}{\|\mathbf{v}\|}. \tag{2.9}$$

This unit vector determines the gazing directions $\mathscr{D}_L = (\theta_L, \phi_L)$ and $\mathscr{D}_R = (\theta_R, \phi_R)$ as shown in Fig. 2.3.

For any $\mathbf{P_W}$ and $\mathbf{r}$, there is a unique 3D localization for $\mathbf{r}_L$ and $\mathbf{r}_R$ given a *baseline distance* b. This is explained next.

2.3.2 Stereoscopic viewing points

The stereoscopic viewpoints $\mathbf{r_L}$ and $\mathbf{r_R}$ are located on a plane parallel to the reference floor $\mathbf{XZ}$-plane which contains $\mathbf{r}$ as shown in Fig. 2.3.

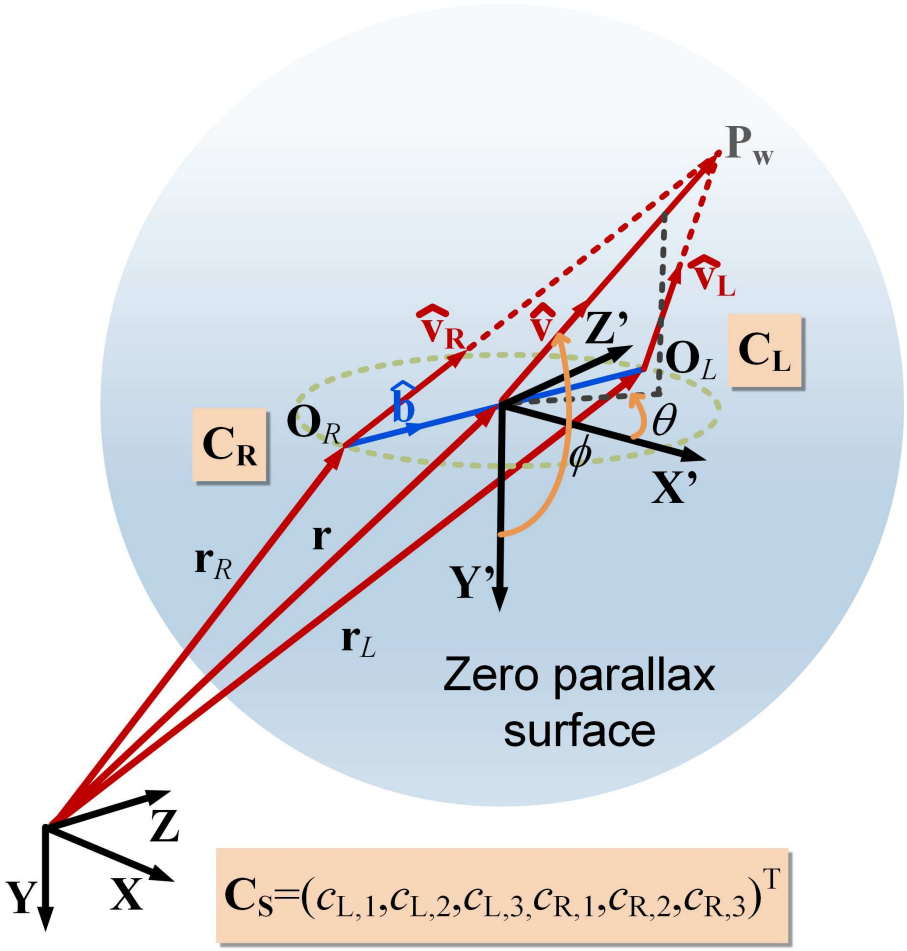


Figure 2.3: Stereoscopic viewing model.

The baseline vector is defined by $\mathbf{b} = \mathbf{r}_L - \mathbf{r}_R$. By definition, the baseline vector $\mathbf{b}$ is on the $\mathbf{X}'\mathbf{Z}'$-plane. The baseline length used in the analysis is denoted $b = \|\mathbf{b}\|$. Finally, the direction between right and left viewpoints is given by the unit vector

$$\hat{\mathbf{b}} = \frac{\mathbf{r}_L - \mathbf{r}_R}{\|\mathbf{r}_L - \mathbf{r}_R\|}. \tag{2.10}$$

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 location of these stereoscopic viewing points is 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 gazing direction.

In Fig. 2.4-(a) is shown the case of a point in the scene $\mathbf{P_W}$ which is located at a distance $\|\mathbf{v}\| = v$ from the reference viewpoint $\mathbf{r}$, which in this figure is at $\mathbf{O}$. 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. 2.4-(b), a different point $\mathbf{P_W}$ is chosen, which is at the same distance from the reference point as before but in another gazing direction in azimuth θ_2. According to the model for $\mathbf{C_S}$, the same vergence $\omega_2 = \omega_1$ is necessary, but different stereoscopic viewpoints $(\mathbf{r_{L,2}}, \mathbf{r_{R,2}})$ are determined by this new gazing direction. The last example shown in Fig. 2.4-(c), corresponds to a point in the scene sufficiently far from the sampling point with respect to the baseline length $(v \gg b)$. In this final example, the gazing direction for left and right viewpoints are parallel, $\hat{\mathbf{v}}_\mathbf{L} \parallel \hat{\mathbf{v}}_\mathbf{R}$, and the vergence angle $\omega = 0°$.

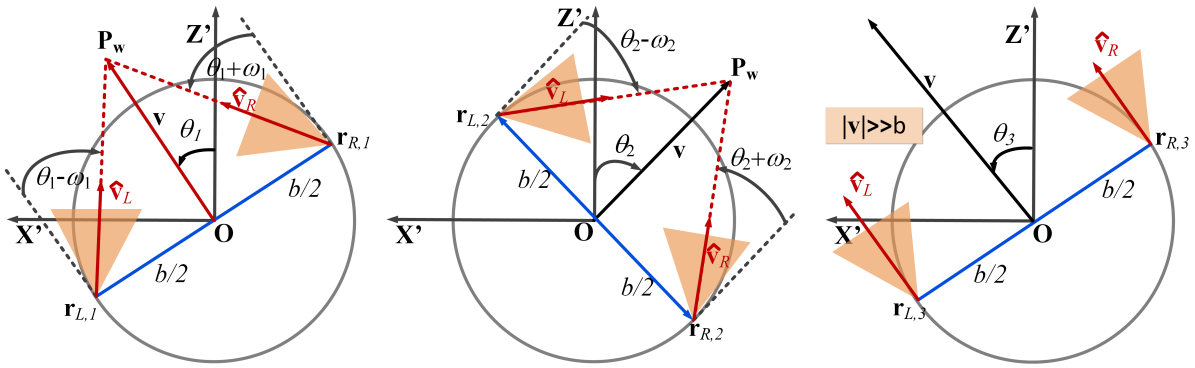


Figure 2.4: The perpendicularity constraint between $\hat{\mathbf{b}}$ and $\hat{\mathbf{v}}$ defines the locations of the stereoscopic viewing points $\mathbf{r_L}$ and $\mathbf{r_R}$: (a) and (b) illustrate the case change in the stereoscopic viewpoints for two scenes points at the same distance from the reference point but in two different gazing directions in azimuth, and the example (c) shows the case of a point in the scene sufficiently far from the reference point.

2.3.3 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

$$\mathbf{r} = \mathbf{r}_L - \frac{b}{2}\hat{\mathbf{b}} \tag{2.11}$$

$$= \mathbf{r}_R + \frac{b}{2}\hat{\mathbf{b}}. \tag{2.12}$$

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* and it is depicted in Fig. 2.3.

If the user's head has freedom of movement in 3D-space, such as 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 $\mathbf{X'Z'}$-plane. In that situation, the set of all possible locations for the stereoscopic viewpoints would fall onto a spherical surface with center in $\mathbf{r}$ and diameter b that would define a *viewing sphere* instead of a *viewing circle*. However, for this thesis, 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 gazing direction $\mathscr{D} = (\theta, \phi)$ and a time variable t. The individual gazing directions defined for $\mathbf{C_L}$ and $\mathbf{C_R}$ can be calculated from these parameters.

2.3.4 Binocular vergence

Any point $\mathbf{P_W}$ in the scene determines a *fixation point*. This can be seen by defining lines containing $\mathbf{v_L}$ and $\mathbf{v_R}$. These lines intersect in a *fixation point* coincident with $\mathbf{P_W}$. The set of all possible fixation points at the same distance v defines a sphere centered at $\mathbf{r}$: the zero parallax sphere. This is the spherical region depicted in Fig. 2.3.

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

$$\hat{\mathbf{v}}_\mathbf{L} = \frac{\mathbf{r}_L - \mathbf{P_W}}{\|\mathbf{r}_L - \mathbf{P_W}\|}, \tag{2.13}$$

and

$$\hat{\mathbf{v}}_\mathbf{R} = \frac{\mathbf{r}_R - \mathbf{P_W}}{\|\mathbf{r}_R - \mathbf{P_W}\|}, \tag{2.14}$$

as depicted in Fig. 2.3. The gazing directions $\mathscr{D}_L$ and $\mathscr{D}_R$ are measured with respect to systems of coordinates $\mathbf{X_L Y_L Z_L}$ and $\mathbf{X_R Y_R Z_R}$, centered at $\mathbf{r}_L$ and $\mathbf{r}_R$, respectively, whose axes are parallel to the axes in the global coordinate system $\mathbf{XYZ}$.

The trigonometric relationship between the panning angles θ_L and θ_R given the vergence

point $\mathbf{P_W}$ is illustrated in Fig. 2.5-(a). The vergence angle ω is defined as follows

$$\omega = \arctan\left(\frac{b}{2v}\right), \tag{2.15}$$

where v is the vergence distance to the scene as shown by the geometric relation ships in Fig. 2.5-(b). When the fixation distance is $v >> b$, $\omega \simeq 0^\circ$ and $\theta_L = \theta_R = \theta$. The latter is equivalent to a stereoscopic camera with parallel optical axes.

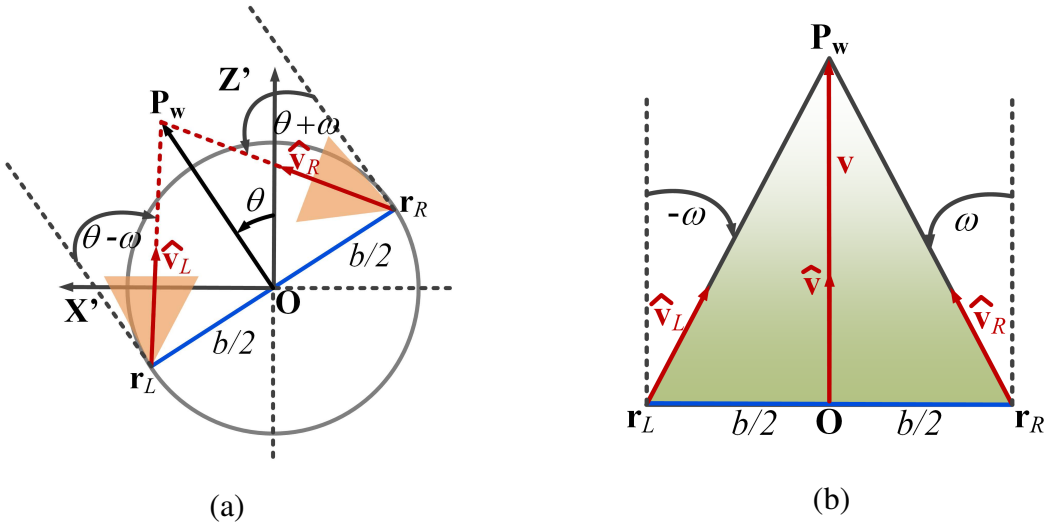


Figure 2.5: Geometric relationship between the gazing direction and location of the stereoscopic viewpoints with respect to $\mathbf{r}$: (a) gazing directions, and (b) stereoscopic viewpoint locations.

2.3.5 Left and right gazing directions

The perpendicularity constraint defined in Section 2.3.2 determines the relationship between $\mathscr{D}_L$, $\mathscr{D}_R$ and the global gazing direction $\mathscr{D} = (\theta, \phi)$. The vergence angle ω given in Eq. 2.15 determines the gazing directions $\mathscr{D}_L$ and $\mathscr{D}_R$ as follows

$$\begin{aligned} \theta_L &= \theta - \omega, \\ \theta_R &= \theta + \omega, \\ \phi_R &= \phi_R = \phi, \end{aligned} \tag{2.16}$$

hence, the $\mathscr{D}$ defines $\mathscr{D}_L$ and $\mathscr{D}_R$

$$\begin{aligned} \mathscr{D}_L &= (\theta - \omega, \phi), \\ \mathscr{D}_R &= (\theta + \omega, \phi). \end{aligned} \tag{2.17}$$

These gazing direction relationship are shown in Fig. 2.5-(b).

2.3.6 Left and right viewing functions

The viewing function defined in Eq. 2.7 can be used to define the viewing functions for the left and right-eye viewpoints

$$\mathbf{C}_j(\mathbf{u}_j) = \left(c_{j,1}(\mathbf{u_j}), c_{j,2}(\mathbf{u_j}), c_{j,3}(\mathbf{u_j})\right)^T, \tag{2.18}$$

where $c_{j,n}$, for $j = \{L, R\}$ and $n \in \{1, 2, 3\}$, are the tristimulus values defined in Sec. 2.2, and $\mathbf{u_j}$ are the input vectors for $\mathbf{C_j}$. These input vectors are

$$\begin{aligned} \mathbf{u_L} &= (\mathscr{D}_L, \mathbf{r}_L, t), \\ &= (\theta - \omega, \phi, \mathbf{r}_L, t), \end{aligned} \tag{2.19}$$

$$\begin{aligned} \mathbf{u_R} &= (\mathscr{D}_R, \mathbf{r}_R, t), \\ &= (\theta + \omega, \phi, \mathbf{r}_R, t). \end{aligned} \tag{2.20}$$

2.3.7 Omnistereoscopic viewing function

The *omnistereoscopic viewing function* $\mathbf{C_S}$ can be derived from the general plenoptic function $\mathscr{P}$ for a given color space $\mathscr{C}$ once the viewpoint in space, the baseline, the gazing direction and vergence are defined. 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 distinct viewpoint with horizontal parallax and can be written as a vector quantity

$$\mathbf{C_S}(\mathbf{u_S}) = \left(c_{L,1}, c_{L,2}, c_{L,3}, c_{R,1}, c_{R,2}, c_{R,3}\right)^T, \tag{2.21}$$

where the input parameters for $\mathbf{C_S}$ are

$$\mathbf{u_S} = (\theta, \phi, \omega, \mathbf{r}, b, t). \tag{2.22}$$

The input parameters $\mathbf{u_L}$ and $\mathbf{u_R}$ can be derived from $\mathbf{u_S}$ using the relationships given by Eq. 2.19.

The function $\mathbf{C_S}$ is defined for $\theta \in [0°, 360°)$ and for $\phi \in (0°, 180°)$. Notice that $\mathbf{C_S}$ is undefined in the *nadir* and *zenith* gazing directions since the omnistereoscopic viewing function is undefined at the poles.

2.4 Omnistereoscopic image acquisition

The acquisition of an omnistereoscopic image of a scene, suitable for binocular human viewing, requires the acquisition of the function $\mathbf{C_S}$. Similarly to the monoscopic case, an omnistereoscopic sensor acquires the color intensity components for the left and right viewpoints by integrating the tristimulus function values over a defined exposure time Δ_t.

The captured information in a reference viewpoint $\mathbf{r}$ can be used to approximate the value of an omnistereoscopic image $\mathscr{I}_S$, which is defined for a set of gazing directions $\mathscr{D}$, e.g., $\Delta_a = 360°$ in azimuth and a limited FOV Δ_e in elevation.

A possible representation of $\mathscr{I}_S$ is as pair of panoramic images $(\mathscr{I}_L, \mathscr{I}_R)$, which are the left and right-eye views of the scene for any $\mathscr{D}$ and a given vergence distance. As in the monoscopic case, $\mathscr{I}_S$ can be mapped onto variety of surfaces and displayed by using diverse stereoscopic technologies.

For a given $\mathbf{r}$, the spatial sampling of $\mathscr{I}_S$ leads to a multidimensional array $\mathbf{I}_S : \mathbb{Z}^2 \rightarrow \mathbb{R}^6$, or equivalently to a pair of arrays $\mathbf{I}_L : \mathbb{Z}^2 \rightarrow \mathbb{R}^3$ and $\mathbf{I}_R : \mathbb{Z}^2 \rightarrow \mathbb{R}^3$.

Once $\mathscr{I}_S$ has been approximated by sampling the C_S function, stereoscopic images with partial FOV can be rendered. This approach enables the interactive navigation of the image-based scene by dynamically generating pairs of left and right images $\mathbf{im_L}$ and $\mathbf{im_R}$, respectively, for each gazing direction $\mathscr{D}$.

2.4.1 The acquisition problem

The sampling of $\mathbf{C_S}$ in order to generate $\mathscr{I}_S$ can be done by using a variety of techniques [27]. Some methods are based on the sequential sampling of $\mathbf{C_S}$ over a defined Δ_t and for a given color space $\mathscr{C}$ (Fig. 2.6-(a)). Other techniques are based on the simultaneous acquisition of partially overlapped stereoscopic snapshots of the scene that are used to reconstruct $\mathscr{I}_S$ according to a given geometric model (Fig. 2.6-(b)). A few proposed techniques rely on light diffractive devices to acquire the scene from two distinct viewpoints in every direction (Fig. 2.6-(c)). The various acquisition techniques mentioned here aim to produce imagery suitable for human

stereopsis in every gazing direction in azimuth. The literature review on this topic is presented in Chapter 4.

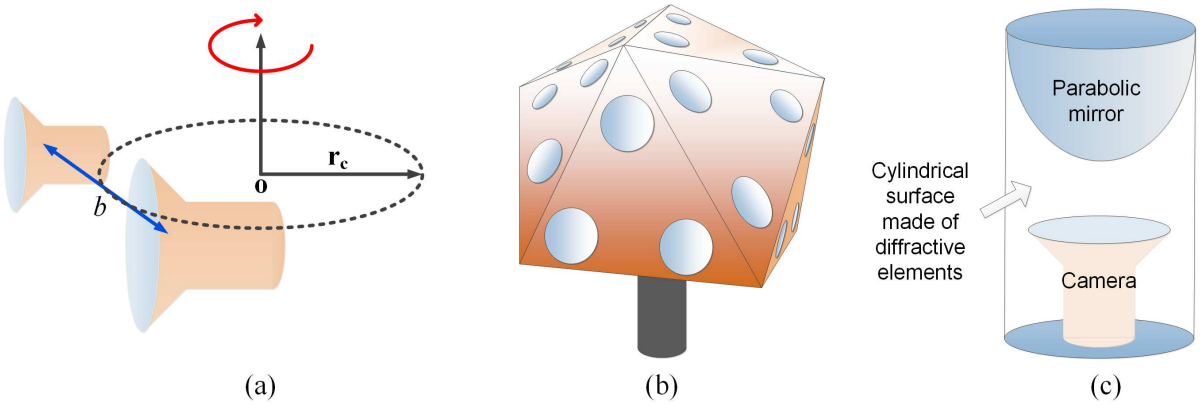


Figure 2.6: Alternatives for the acquisition of $\mathbf{C_S}$: (a) sequential acquisition, (b) acquisition of a reduced number of stereoscopic images, and (c) method based on diffractive lenses.

One problem that is of particular interest for this thesis is the acquisition of omnistereoscopic imagery of dynamic scenes. The majority of the real-world scenes are dynamic and the acquisition of omnistereoscopic snapshots and videos is a problem that has not been addressed by traditional techniques to produce stereoscopic panoramas and videos. An example of this acquisition of $\mathbf{C_S}$, which relies on a rotating pair of pinhole cameras to acquire the scene is shown in Fig. 2.7.

2.5 Summary

In this chapter, I proposed a novel mathematical model for the omnistereoscopic viewing of any scene. This model is applicable to static and dynamic scenes as well and it sets the geometric constraints for the *omnidirectional viewing function* that represents the stereoscopic visual information to be acquired by an omnidirectional image sensor intended for stereo vision.

The model uses the plenoptic function to derive a multidimensional array of real-valued signals, that, when sampled by a panoramic sensor, provides the necessary information to recreate a binocular image at any desired gazing direction.

This approach of modeling the visual information that need to be captured to emulate stereoscopic vision in humans has not been attempted before to my best knowledge. Hence, this model is one of the contributions of this thesis.

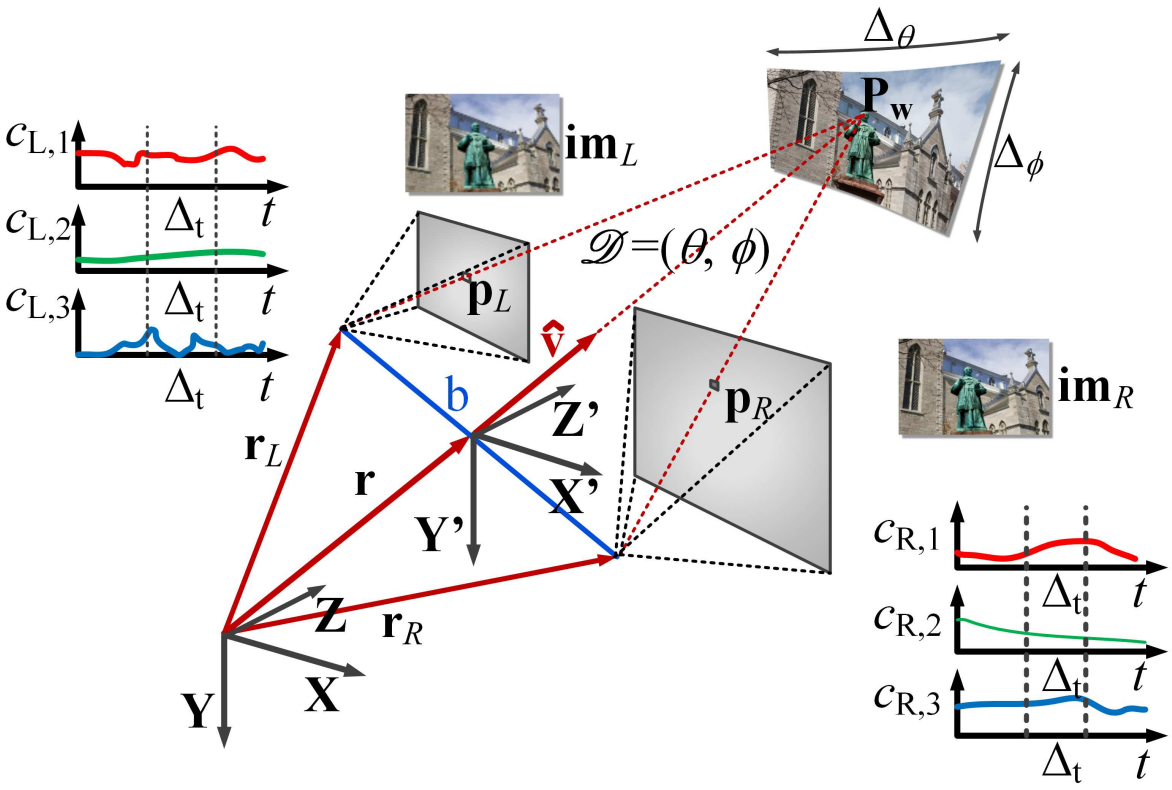


Figure 2.7: Omnistereoscopic viewing function $\mathbf{C_S}$ and the acquisition of a partial stereoscopic view of the scene.

In the next section, I propose a model for omnistereoscopic acquisition of the function $\mathbf{C_S}$. This singular *acquisition model* can model four *acquisition configurations*, each of which is applicable to different omnistereoscopic acquisition methods presented in the literature. Since this acquisition model is constrained to the omnistereoscopic viewing model presented in this chapter, it is a useful mathematical framework to study the problem of $\mathbf{C_S}$ acquisition and $\mathscr{I}_S$ reconstruction in different situations.

Chapter 3

Acquisition model

The main problem in omnistereoscopic imaging, when it is intended for human viewing, is defining a configuration of cameras to acquire the visual field omnidirectionally, from two viewpoints with horizontal parallax, as defined by the function $\mathbf{C_S}$. The problem becomes more constrained when the idea is to acquire dynamic scenes omnistereoscopically. In this scenario, problems such as avoiding camera self-occlusion, depth consistency, vertical disparities and producing a credible reproduction of the scene appealing to the human binocular system are some of the challenges to take into account.

The self-occlusion problem is common to all the conceptualizations of panoramic photography which must be considered when a single or multiple cameras are used to sample the scene omnidirectionally. If the image sampling is sequential, self-occlusion can be avoided. However, the acquisition of dynamic scenes exacerbates the restrictions since all the information to produce omnistereoscopic images has to be acquired at once. The parallax arising from sampling the scene from different viewpoints is another problem common in panoramic photography. The problem gets more complicated when the simultaneous acquisition of stereoscopic images from different viewpoints enters into the equation.

One possible solution to the problem is to acquire multiple stereoscopic snapshots of the scene simultaneously. In this case, the geometric configuration of the multi-sensor device must be carefully designed to avoid self-occlusion which occurs when one camera lies in the field-of-view of another. A solution is using diffractive optics to obtain two views of the scene with horizontal parallax, and doing so omnidirectionally.

In this chapter, I introduce a model for omnistereoscopic image acquisition to describe the different variations of omnistereoscopic imaging techniques and cameras. The four variations

of this model are used in the next chapter to classify and present the cameras and techniques suitable for omnistereoscopic imaging.

3.1 Omnistereoscopic acquisition model

The *general acquisition model* is composed of two pinhole cameras separated by a baseline distance b with respect to a global reference of coordinates in three-dimensional (3D) space as illustrated in Fig. 3.1. This model is used to derive four camera configurations which can be distinguished by the relative location of the stereoscopic pair of cameras with respect to the reference center. In this thesis, we refer to these four spatial variations of the acquisition model as *Configurations*, and we assign them numbers from one to four.

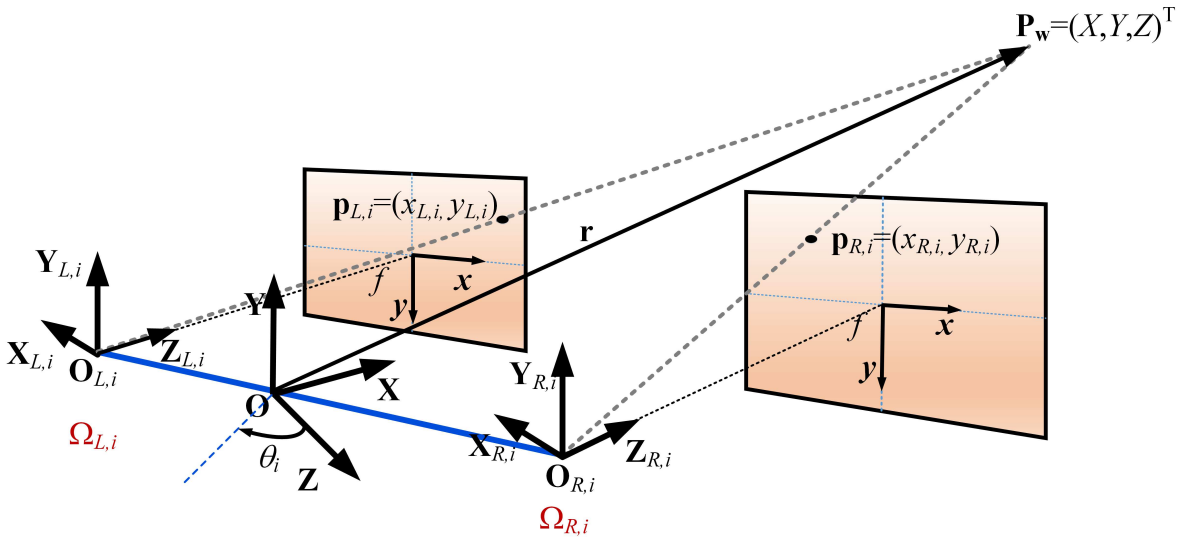


Figure 3.1: The global reference frame versus the camera frames for the stereoscopic camera composed by cameras $\Omega_{L,i}$ and $\Omega_{R,i}$, which is oriented at θ_i in azimuth.

Although this model consists of one pair of cameras, it can model complex multiple camera configurations as well as a single stereoscopic camera rig rotated at different azimuth angles. Also in Chapter 4, I review a large variety of acquisition cameras and methods to produce omnistereoscopic imagery suitable for human viewing that each can be described by one of these four variations of the acquisition model.

The location and orientation of this stereoscopic camera pair in 3D space is restricted by the need to capture two snapshots of the same scene from two viewpoints with horizontal parallax.

One constraint is that all the possible locations for this pair of pinhole cameras are restricted to the horizontal $\mathbf{XZ}$-plane, which is used as the reference horizontal plane. Another constraint is that the optical axes of the stereoscopic pair of 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.

3.1.1 General notation

The subindex j ($j \in \{L, R\}$) refers to the left (L) or right (R) cameras in a stereoscopic camera pair. The subindex i refers to one of N consecutive gazing directions $\mathscr{D} = (\theta_i, 0°)$ ($i \in \{0, ..., N-1\}$) defined as

$$\theta_i = \frac{i \cdot 360°}{N}. \tag{3.1}$$

It follows that $\theta_N = \theta_0$. Unless specifically said, the stereoscopic rig orientation in azimuth will be referred to θ_i. This direction is different of the user's gazing $\mathscr{D}$, which indicates a direction in 3D-space (θ, ϕ) with respect to $\mathbf{XYZ}$.

The subindex (j, i) indicates the camera j ($j \in \{L, R\}$) from a stereoscopic pair of cameras which is oriented at θ_i in azimuth ($i \in \{0, ..., N-1\}$) with respect to the global reference of coordinates. For a complete description of the typographic conventions used in this document, please refer to Appendix D.1.

Each of the four configurations presented in Fig. 3.2 can be modeled by a set of single stereoscopic rigs with different gazing directions or by a singular stereoscopic camera sequentially rotated at different θ_i angles. All configurations can be described by a single stereoscopic camera, after defining the spatial location of left and right cameras and the set of all possible gazing directions for such a virtual stereoscopic camera.

3.1.2 Systems of coordinates

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

Each camera has its own local frame of coordinates, which is denoted as $\mathbf{X}_{j,i}\mathbf{Y}_{j,i}\mathbf{Z}_{j,i}$. Similarly to the world reference of coordinates, a point in the camera coordinates frame is

denoted $\mathbf{P}_{j,i} = (X_{j,i}, Y_{j,i}, Z_{j,i})^T$. The global (reference) and local (cameras) coordinate frames are illustrated in Fig. 3.1 for *configuration 1*.

3.1.3 Cameras and stereoscopic rigs

Each individual camera is denoted $\Omega_{j,i}$, where $j \in \{L, R\}$ and $i \in \{0, ..., N-1\}$. A camera $\Omega_{j,i}$ has a projective center, which is referred in the text using location $\mathbf{O}_{j,i}$ of its projective center. The individual cameras are grouped in stereoscopic camera pairs $(\Omega_{L,i}, \Omega_{R,i})$, where $i \in \{0, ..., N-1\}$ defines its orientation angle in azimuth according to Eq. 3.1 as illustrated in Fig. 3.2 for each configuration.

The distinction between left and right does not necessarily correspond to the relative spatial location of each camera in world coordinates, but it is used here to distinguish each camera in a stereoscopic pair. In order to eliminate ambiguities, the cameras are labeled according to their position in the reference of coordinates $\mathbf{XYZ}$ when $\theta_0 = 0°$, e.g., when the baseline vector $\mathbf{b}$ is parallel to $\mathbf{X}$ axis. Using this convention, the camera $\Omega_{L,i}$ is the one whose projection center $\mathbf{O}_{L,i}$ is centered at $X \leq 0$ and, conversely, the right camera, which is called $\Omega_{R,i}$, is the one whose projection center $\mathbf{O}_{R,i}$ is centered at $X > 0$. This labelling scheme is exemplified in Fig. 3.1. After labelling them for θ_0, each camera retains its left or right for all θ_i in the acquisition sequence.

3.1.4 Location and orientation of each camera

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

Each camera's frame and its projection center are co-located. The $\mathbf{Y}_{j,i}$-axes are parallel to the Y-axis. The optical axis of camera $\Omega_{j,i}$ is aligned with the its corresponding $\mathbf{Z}_{j,i}$-axis. Furthermore, each $\mathbf{Z}_{j,i}$ axis is perpendicular to the vector $\mathbf{b} = \mathbf{O}_{L,i} - \mathbf{O}_{R,i}$, which is defined by each stereoscopic camera. Consequently, the optical axes of each camera pair are parallel. Finally, the image plane is parallel to the $\mathbf{XY}$-plane and it is located at $Z_{j,i} = f$ as shown in Fig. 3.1.

The optical axes of the cameras are parallel and located on the $\mathbf{XZ}$-plane as well as their projection centers $\mathbf{O}_{j,i}$. Parallel optical axes means that the vergence distance for $\mathbf{C_S}$ is fixed at infinity $(\omega = 0°)$ as defined in Section 2.3.4. The latter condition does not affect the generality

of model to describe the sampling of $\mathbf{C_S}$ since the vergence can be defined when reconstructing $\mathscr{I}_S$.

The gazing direction $\mathscr{D}$ is defined by the angle θ_i in azimuth and $\phi_i = 0°$, since optical axes of all cameras are on the $\mathbf{XZ}$-plane.

3.1.5 Stereoscopic image pairs

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

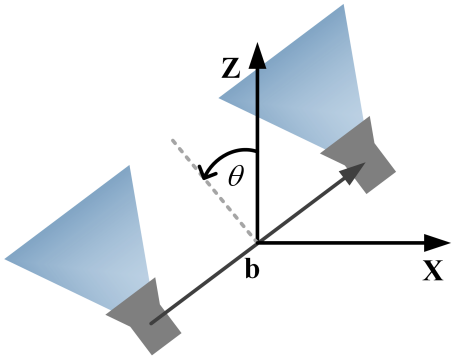
The two-dimensional (2D) reference of coordinates in each image is located at the symmetry center of each image as shown in Fig. 3.1. A coordinate point on each image plane is denoted $\mathbf{p}_{j,i} = (x_{j,i}, y_{j,i})$, where the subindex (j, i) has the meaning given in Section 3.1.1. The image window is a rectangular subregion of the image plane, whose center intersects the camera optical axis $Z_{j,i}$ and whose area is defined by its horizontal width (W_h) and its aspect ratio (a_r).

3.2 Acquisition configurations

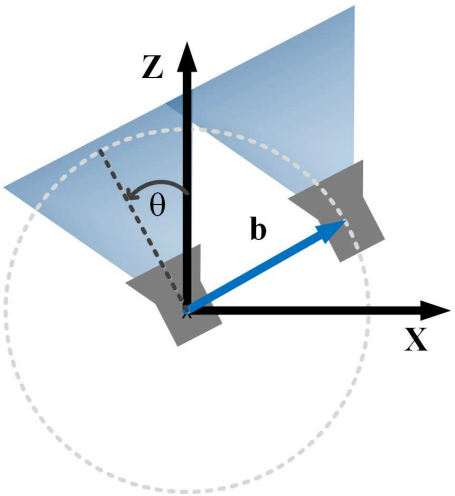
This simple model can describe complex multiple camera configurations for simultaneous capture of the whole scene as well as sequential acquisition techniques. The first two configurations of the acquisition model (*configuration 1* and *2*) are limited to sequential acquisition schemes due to self-occlusion between individual cameras that prevent them from being implemented using multiple sensors. The other two configurations (*configuration 3* and *4*) can also model the simultaneous scene acquisition and they can be implemented using multiple sensors. The simultaneous acquisition case is of particular interest in the context of the problem studied in this thesis.

Acquisition configuration 1

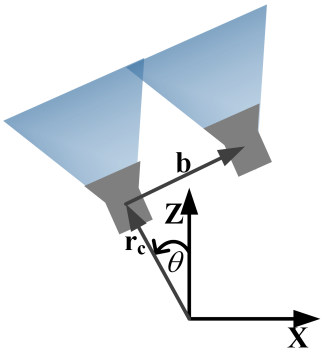
The *central stereoscopic rig* (*configuration 1*) is illustrated in Fig. 3.2-(a). This configuration models the sampling of the scene by means of a rotating stereoscopic camera, which captures partial snapshots at regular angular intervals in azimuth. This method has been the first technique to create omnistereoscopic images with horizontal parallax. This camera configuration consists of two cameras with co-planar projection centers which are separated by a baseline b. The stereoscopic camera rotation is around the Y-axis exclusively (pitch rotation). This configuration describes acquisition methods that have been widely used over the last decade, whether using planar [28, 29] or line sensors [11, 30] for sequential acquisition. This configuration also models a widely used technique based on rotating an off-centered camera at regular angles θ_i, where $i \in \{0, ..., N-1\}$ and N is usually large, producing a set of overlapped stereoscopic images of the scene [12, 31]. Although simple in its conception, *configuration 1* cannot to be used in a parallel acquisition configuration due to self-occlusion between cameras, hence it can only represent sequential omnistereoscopic acquisition.



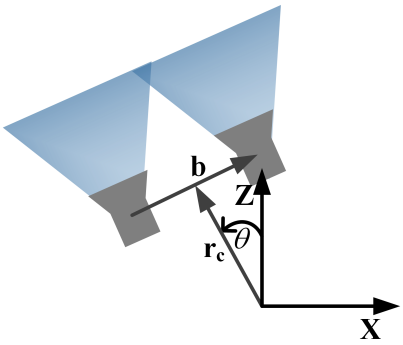
(a) *configuration 1*



(b) *configuration 2*



(c) *configuration 3*



(d) *configuration 4*

Figure 3.2: Variations of the acquisition model: (a) central stereoscopic rig, (b) lateral stereoscopic rig, (c) lateral-radial stereoscopic rig, and (d) off-centered stereoscopic rig.

In this document, the term *parallel* refers to the *simultaneous* acquisition of multiple images, in contrast to *sequential* which in this context refers to the non-simultaneous acquisition of partial FOV stereoscopic images of the scene.

Acquisition configuration 2

The *lateral stereoscopic rig* (*configuration 2*) is shown in Fig. 3.2-(b). This configuration models the sequential acquisition of partial images of the scene by rotating the stereoscopic camera at regular intervals θ_i, where $i \in \{0, ..., N-1\}$. The main difference with *configuration 1* is that the pitch rotation is defined around the nodal point of one of the cameras, i.e., making the rotation axis to coincide with the projection center of the left or right camera.

In this approach, a singular view point panorama is produced by mosaicking the images acquired by the central camera, while a second image with horizontal parallax for each θ_i is acquired by the lateral camera [32, 33]. The lateral camera (stereoscopic counterpart) describes a circle of radius equal to the stereo baseline b. Similar to *configuration 1*, this configuration cannot be used for the simultaneous acquisition of the whole scene due to self-occlusion of the central camera.

Acquisition configuration 3

The *lateral-radial stereoscopic rig* (*configuration 3*) is shown in Fig. 3.2-(c). This configuration models a stereoscopic camera rotated off-center, where one camera is radially displaced from the rotation axis in $\mathbf{O}$ by a distance $\|\mathbf{r_c}\|$, and the second camera is laterally displaced b perpendicularly with respect to the direction of $\|\mathbf{r_c}\|$. This arrangement enables the capture of a second snapshot of the scene with horizontal parallax for each sampling angle θ_i [18].

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

Acquisition configuration 4

The *off-centered stereoscopic rig* (*configuration 4*) is a stereoscopic camera located at a radial distance $\|\mathbf{r_c}\|$ from the geometrical center $\mathbf{O}$ as depicted in Fig. 3.2-(d). This configuration

can be derived from *configuration 1* by locating the pitch axis $\mathbf{O}$ a distance $\|\mathbf{r_c}\|$ behind the stereoscopic camera midpoint.

This configuration models the case where multiple stereoscopic rigs, radially located with respect of the pitch center $\mathbf{O}$, are used to acquire partially overlapped snapshots of the whole scene [34, 17, 35, 36]. In one approach, the successive images acquired by left and right cameras can be mosaicked to create left and right eye panoramas: $(\mathbf{I_L}, \mathbf{I_R})$ [37].

3.3 Coordinate transformations

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

$$\mathbf{P}_{j,i} = \mathbf{R}_i \cdot (\mathbf{P_W} - \mathbf{T}_{j,i}), \tag{3.2}$$

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

The rotation matrix is defined as follows

$$\mathbf{R}_i = \begin{pmatrix} \cos\theta_i & 0 & \sin\theta_i \\ 0 & 1 & 0 \\ -\sin\theta_i & 0 & \cos\theta_i \end{pmatrix}, \tag{3.3}$$

The generic notation to refer to the translation vectors are

$$\mathbf{T}_{L,i} = (t_{L,1}, t_{L,2}, t_{L,3})^T, \tag{3.4}$$

$$\mathbf{T}_{R,i} = (t_{R,1}, t_{R,2}, t_{R,2})^T, \tag{3.5}$$

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

Configuration 1:

$$\mathbf{T}_{L,i} = \left(-\frac{b}{2}\cos\theta_i, 0, -\frac{b}{2}\sin\theta_i\right)^T, \tag{3.6}$$

$$\mathbf{T}_{R,i} = \left(\frac{b}{2}\cos\theta_i, 0, \frac{b}{2}\sin\theta_i\right)^T, \tag{3.7}$$

where $b = \|\mathbf{b}\|$. This is shown in Fig. 3.3-(a).

Configuration 2:

$$\mathbf{T}_{L,i} = (0, 0, 0)^T, \tag{3.8}$$

$$\mathbf{T}_{R,i} = (b\cos\theta_i, 0, b\sin\theta_i)^T. \tag{3.9}$$

This is shown in Fig. 3.3-(b).

Configuration 3:

$$\mathbf{T}_{L,i} = (\|\mathbf{r_c}\|\sin\theta_i, 0, \|\mathbf{r_c}\|\cos\theta_i)^T, \tag{3.10}$$

$$\mathbf{T}_{R,i} = (\|\mathbf{r_o}\|\cos\beta_i, 0, \|\mathbf{r_o}\|\sin\beta_i)^T, \tag{3.11}$$

where $\beta_i = \arctan\left(\frac{\|\mathbf{r_c}\|}{b}\right) + \theta_i$ and $\|\mathbf{r_o}\| = \sqrt{\|\mathbf{r_c}\|^2 + b^2}$. This is shown in Fig. 3.3-(c).

Configuration 4:

$$\mathbf{T}_{L,i} = (-\|\mathbf{r_o}\|\sin\beta_i, 0, \|\mathbf{r_o}\|\cos\beta_i)^T, \tag{3.12}$$

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

$$\mathbf{T}_{R,i} = (\|\mathbf{r_o}\|\sin\alpha_i, 0, \|\mathbf{r_o}\|\cos\alpha_i)^T, \tag{3.13}$$

where $\alpha_i = \arctan\left(\frac{b}{2\|\mathbf{r_c}\|}\right) - \theta_i$. This is shown in Fig. 3.3-(d).

Alternatively, the transformation of coordinates in Eq. 3.2 can be written as,

$$\mathbf{P}_{j,i} = \mathbf{R}_i \cdot \mathbf{P_W} - \mathbf{T}_j, \tag{3.14}$$

where

$$\mathbf{T}_j = \mathbf{R}_i \cdot \mathbf{T}_{j,i}. \tag{3.15}$$

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

3.3.1 Projecting the scene on each camera

The frame of corrdinates transformation to map a scene point $\mathbf{P_W}$ into the frame of coordinates of $\Omega_{j,i}$ is defined by Eq. 3.2. Furthermore, any point $\mathbf{P_W}$ in the FOV of both cameras of a stereoscopic camera can be projected into each image $\mathbf{im}_{j,i}$ as a point with coordinates $\mathbf{p}_{j,i} = (x_{j,i}, y_{j,i})$. The 2D image coordinates in each image are given by the projective equations

$$-(x_{j,i} - o_x)s_x = \frac{X_{j,i}}{Z_{j,i}} = f\frac{\mathbf{R}_{1,i}(\mathbf{P_w} - \mathbf{T}_{j,i})}{\mathbf{R}_{3,i}(\mathbf{P_w} - \mathbf{T}_{j,i})}, \tag{3.16}$$

$$-(y_{j,i} - o_y)s_y = \frac{Y_{j,i}}{Z_{j,i}} = f\frac{\mathbf{R}_{2,i}(\mathbf{P_w} - \mathbf{T}_{j,i})}{\mathbf{R}_{3,i}(\mathbf{P_w} - \mathbf{T}_{j,i})}, \tag{3.17}$$

where the subindex j indicated the left (L) or right (R) image, the orientation of the stereoscopic camera θ_i in the acquisition sequence is denoted by $i \in \{0, ..., N-1\}$ (Eq. 3.1), $\mathbf{R}_{k,i}$ for $k \in \{1, 2, 3\}$ is a *row vector* formed by the k^{th} row of $\mathbf{R}_i$, f is the camera focal length, (s_x, s_y) is the pixel size in each dimension, and (o_x, o_y) is the image center in the image coordinate frame.

It can be assumed without losing generality that $o_x = o_y = 0$ and the pixel size is $s_x = s_y = 1$. Hence, the general expressions for the coordinates in each image are given by

$$x_{j,i} = -f\frac{\mathbf{R}_{1,i}(\mathbf{P_w} - \mathbf{T}_{j,i})}{\mathbf{R}_{3,i}(\mathbf{P_w} - \mathbf{T}_{j,i})}, \tag{3.18}$$

$$y_{j,i} = -f\frac{\mathbf{R}_{2,i}(\mathbf{P_w} - \mathbf{T}_{j,i})}{\mathbf{R}_{3,i}(\mathbf{P_w} - \mathbf{T}_{j,i})}. \tag{3.19}$$

An example that illustrates the mapping of $\mathbf{P_W}$ into $\mathbf{p}_{R,i}$ and $\mathbf{p}_{L,i}$ is shown in Fig. 3.1.

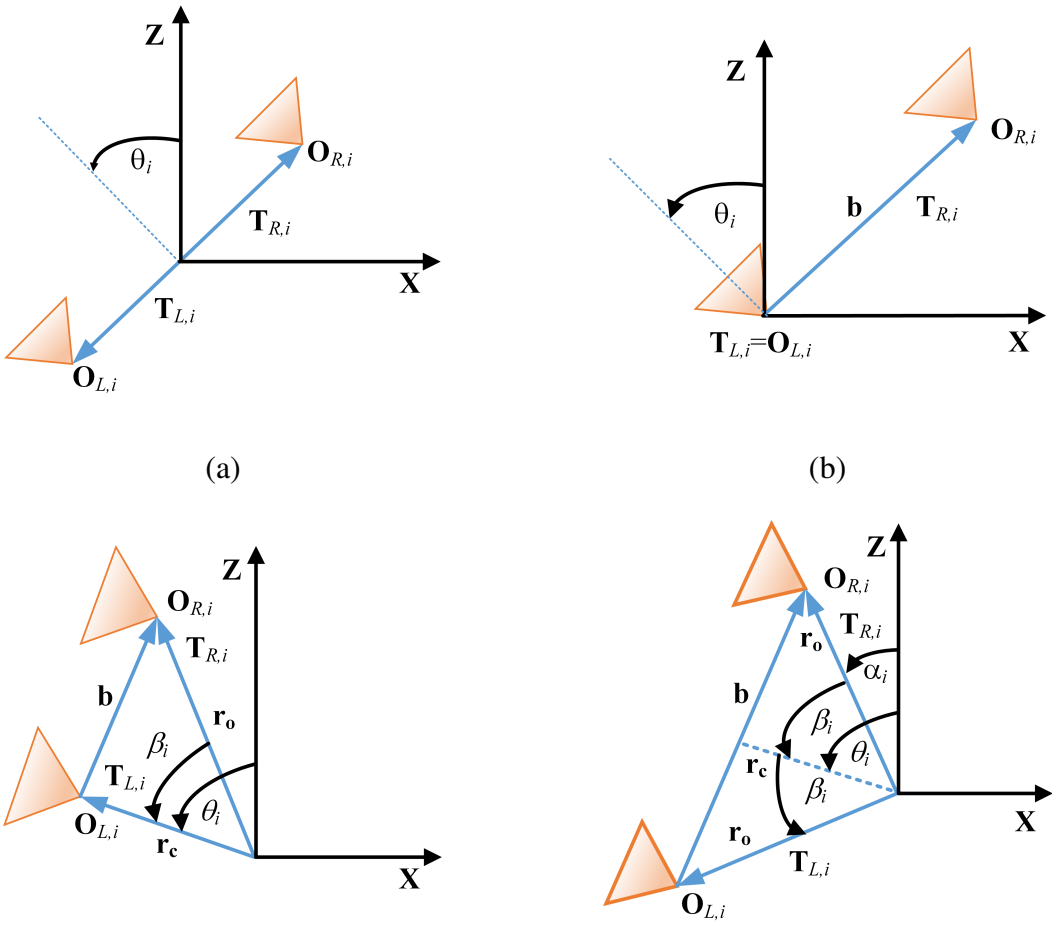


Figure 3.3: The geometric relationships for each configuration.

3.4 Summary

In this chapter, I proposed a general model for the omnistereoscopic image acquisition based on projective geometry. This general *acquisition model* can describe a large variety of omnistereoscopic strategies to acquire samples of $\mathbf{C_S}$. This approach to model the function to be sampled by an omnistereoscopic sensor is an original contribution of this thesis to the research on stereoscopic panoramas.

The next section is dedicated to a literature review of the acquisition technologies suitable to sample the omnistereoscopic viewing function $\mathbf{C_S}$. The acquisition model presented in this chapter will be used in the description of the different methods. Special focus is made on comparing the potential and disadvantages of historical solution to sample the $\mathbf{C_S}$ function in dynamic scenarios.

Chapter 4

Review of omnistereoscopic technology

Most real-world scenarios to be rendered omnistereoscopically are intrinsically dynamic, and hence a practical omnistereoscopic camera should facilitate the reconstruction of $\mathscr{I}_S$ by sampling the $\mathbf{C_S}$ function in different gazing directions $\mathscr{D}$. The reconstructed $\mathscr{I}_S$ or its sample $\mathbf{I_S}$ should be able to stimulate the mechanism of human binocular vision, reproducing a credible and consistent perception of depth in any gazing direction. A few cameras are capable of acquiring $\mathbf{C_S}$ in a single-snapshot. However, not all of them are suitable to produce omnistereoscopic images adequate for human viewing, at least not directly.

In 2001, Zhigang Zhu [38] presented an extensive classification of the different technologies to create omnidirectional stereoscopic imagery. This is an excellent survey of omnistereoscopic methods up to its publication date, which presented a taxonomical classification of camera configurations and methods, comparing their capabilities to produce viewable stereoscopic imagery in any azimuthal gazing direction around the viewer. However, the omnidirectional acquisition of dynamic scenes, which is relevant for today's multimedia applications, was not taken into account in that work.

This chapter presents an up-to-date classification and review of the different technologies to produce omnistereoscopic renditions of real-world scenes in arbitrary gazing directions.

This survey of the different technologies for omnistereoscopic image acquisition has been published in the SPIE Electronic Imaging Journal [27] in July 2013. A companion website [39] for this paper is available online.

4.1 Acquisition of monoscopic panoramas

There are two approaches for the acquisition of the function $\mathbf{C}$ as defined in Chapter 2.2.1: a singular viewpoint (SVP) and a non-singular viewpoint (non-SVP) acquisition, a.k.a. polycentric panoramic model. Any technique available to approximate a monoscopic and omnidirectional image $\mathscr{I}$ from the measuring of $\mathbf{C}$ can be classified into one of these two acquisition alternatives.

In the SVP acquisition, for any gazing direction $\mathscr{D} = (\theta, 0^\circ)$, there is a unique projection center $\mathbf{O}$ which marks a single convergence point for all incident light rays. This model groups the catadioptric cameras used to acquire the whole scene using, for example, a single photosensitive sensor array and a curved mirror. Panoramas created by a rotating camera around its nodal point, or its projection center assuming a pinhole camera model, also satisfy the SVP model. These panoramas are created by acquiring planar images to be mosaicked, or by scanning the scene column-wise, e.g., using line sensor cameras and turning platforms. Examples of SVP acquisition are illustrated in Fig. 4.1-(a) and (b).

In the case of a non-SVP model, the panoramic image is rendered using a centrally symmetric set of N projection centers $\mathbf{O}_k$ ($k \in \{1, ..., N)$, which are not spatially collocated. Cameras based on the non-SVP paradigm are more common than those based on an SVP model because the physical dimension of multiple camera configurations prevents sampling the scene from a single viewpoint. A way around this problem is using planar mirrors to re-position the projection centers closer to each other, approximating an SVP configuration. Examples non-SVP cameras are shown in Fig. 4.1-(c) and (d).

In the context of the problem studied in this thesis, stereoscopic panoramas are by definition non-SVP panoramas since the scene is imaged from two distinct viewpoints (left and right-eye viewpoints) for any $\mathscr{D}$.

4.1.1 Panoramic acquisition: cameras and methods

The omnistereoscopic technologies reviewed herein have been classified into four families based on the strategies used for image acquisition or in their technology. These acquisition families are

- Omnistereo based on catadioptric based cameras
- Sequential techniques to produce stereoscopic panoramas
- Omnistereo based on panoramic snapshots
- Omnistereo based on multiple cameras

This classification into families is independent of the four acquisition models introduced in Section 3.1. Nevertheless, each omnistereoscopic family can be modelled using a specific variation of the acquisition model (*configurations 1-4*). The catadioptric cameras based on vertical parallax are the only exception to this rule, since the acquisition models are based on horizontal stereo.

In the following section, the pros and cons of each family are studied individually, distinguishing those cameras whose characteristics are suitable for acquiring $\mathbf{C_S}$ of dynamic scenes.

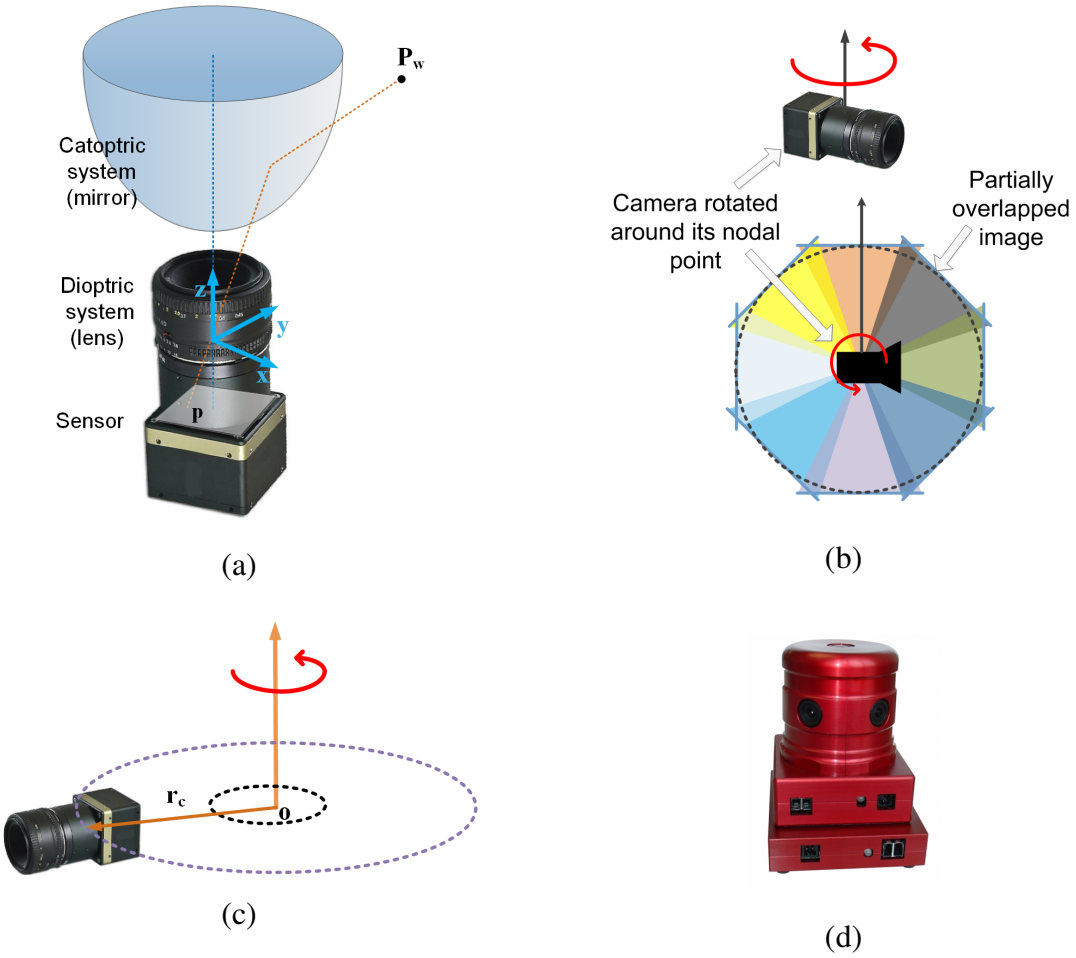


Figure 4.1: Examples of omnidirectional image acquisition: (a) catadioptric cameras based on parabolic or hyperbolic mirrors to produce SVP panoramas, (b) rotating a camera about its nodal point to acquire multiple perspective projections with a common projection center also produces SVP panoramas, while (c) non-SVP panoramas can be created by rotating an off-centered camera to acquire image patches around a point different than its nodal point, as well as (d) multi-sensor cameras such as the Ladybug2 panoramic camera, which also produce non-SVP panoramas.

4.2 Catadioptric cameras

A catadioptric panoramic camera can capture a wide-angle FOV, e.g., 360° in azimuth and a wide-angle range in elevation, at once by combining the use of mirrors (catoptric system) to reflect the light arriving from every direction towards a lens system (dioptric system), which focuses the light over a planar image sensor. This technology enables to measure $\mathbf{C}$ and to estimate the panoramic image $\mathscr{I}$ for a wide FOV. Since a simple sample mapping transformation is between the sampled $\mathbf{C}$ and the panoramic image, this approach favours the fast rendering of $\mathbf{I}$ from $\mathscr{I}$.

In the case of a parabolic profile mirror, light rays emanating from the mirror's focal point are reflected outwards as parallel rays. Conversely, by applying the principle of reversibility of the optical path, ray paths intersecting the mirror's inner focal point are reflected parallel to the mirror's symmetry axis. A dioptric system coaxially located at certain distance from the mirror surface focuses the light over the image sensor. The principle is analogous to a parabolic dish antenna, which collects electromagnetic radiation at its focal point by concentrating incident wavefronts arriving from a source relatively far from the antenna.

In Fig. 4.2, a simplified model of the catadioptric principle is illustrated, showing how a ray of light emanating from the scene point $\mathbf{P_W}$ is reflected by the catoptric system and focused by the dioptric system onto the point $\mathbf{p} = (x, y)^T$ on the image plane. The sampled panoramic image $\mathbf{I}$ is an orthographic projection of the scene, which can then be projected onto a canonical representation used in panoramic imaging, i.e., cylindrical, cubical, spherical, among others, or to extract a partial FOV of the scene in any direction in azimuth.

In a real-world scenario, a parabolic mirror profile reflects light in quasi-parallel fashion, affecting the quality of the orthographic projection. In the case of using a hyperbolic mirror profile, light rays directed toward a focal point (located inside the convex mirror's surface) are reflected towards the other focal point of the hyperbola, where the dioptric system is located.

4.2.1 Catadioptric cameras: antecedents

One of the first panoramic cameras exploiting this idea is attributed to Rees [40], who proposed in 1967 to combine a hyperbolic mirror and a TV camera to provide an omnidirectional view of the scene to the operator of a military armored vehicle. This is an early example of sampling $\mathbf{C}$ function to produce panoramic videos.

Panoramic cameras based on the catadioptric configuration, where light is focused on a

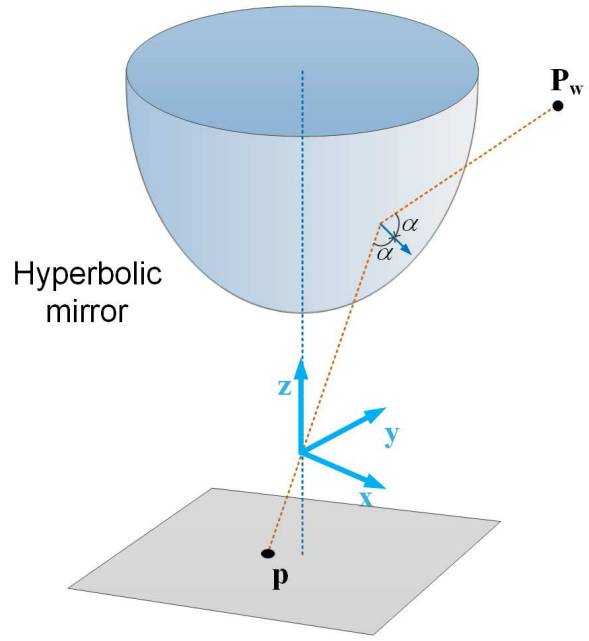


Figure 4.2: SVP catadioptric camera principle using parabolic mirrors.

single projection point, correspond to the SVP model. Following the SVP principle, a full spherical panorama can be approximated by using two coaxial catadioptric cameras back-to-back and mosaicking the semi-hemispherical images originating from each camera. This idea was proposed by Nayar [41] in 1997. At about the same time, S. Baker and Nayar proposed a model for catadioptric image formation [42] from which they concluded that only parabolic and hyperbolic mirror profiles satisfy the SVP criterion.

Other configurations of catadioptric cameras based on different mirror profiles, i.e., semi-spherical or multi-faceted pyramidal mirrors, exhibit multiple focal points, which makes them require multiple cameras. An example of these non-SVP camera is illustrated in Fig. 4.3, where the catoptric system uses planar mirrors. These configurations are being used in commercial panoramic cameras to produce monoscopic panoramas [43, 44].

The camera configurations described so far can only produce monoscopic panoramas when used as a single-snapshot camera. However, catadioptric cameras can be used to produce omnistereoscopic images when used in clusters [14, 2, 45]. The case of omnistereoscopic images based on a number of monoscopic panoramas is studied in Section 4.4. Along with the development of monoscopic catadioptric cameras, there has been a parallel development of catadioptric omnistereoscopic cameras. The family of the catadioptric cameras for omnistereoscopic

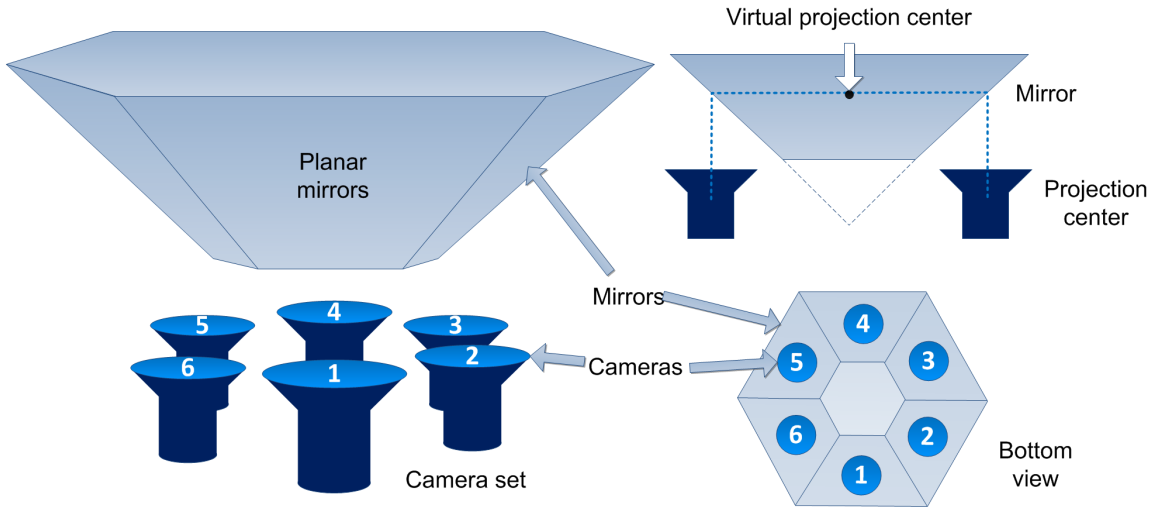


Figure 4.3: Catadioptric camera using planar mirrors instead of hyperbolic or parabolic profiles mirrors.

imagery is described next.

4.2.2 Omnistereo using catadioptric cameras: antecedents

The development of omnistereoscopic catadioptric cameras has paralleled the development of general (monoscopic) catadioptric sensors. It should be mentioned that omnistereoscopic catadioptric cameras were originally intended for the real time estimation of depth maps. In other words, these omnistereoscopic approaches were not intended to produce omnistereoscopic imagery for human viewing, but they were motivated by applications such as robot navigation and 3D scene reconstruction.

One important remark is necessary here: the omnistereoscopic cameras based on a catadioptric configuration with vertical parallax presented in this section are not modelled by any of the acquisition models presented in Section 3.1 since the omnistereoscopic acquisition classification has been constrained to human-viewable omnistereoscopic imagery, i.e., binocular stereo with horizontal parallax.

One of the earlier examples of this technology was an SVP catadioptric camera proposed by Southwell et al. [8] in 1996. This omnistereoscopic catadioptric camera is based on a

coaxial, dual-lobe parabolic mirror, and its main application was to generate omnidirectional depth maps of the terrain. A depiction of this camera appears in Fig. 4.4-(a).

Another configuration was proposed by Gluckman et al. [9] in 1998. Their configuration is based on using two co-axial catadioptric cameras whose vertical baseline helps to acquire an omnistereoscopic pair of images. This camera, illustrated in Fig. 4.4-(b), enables the estimation of the 3D location of a scene point $\mathbf{P_W}$ in space, matching pairs of feature points in the panoramic views from each camera. The larger the vertical baseline b used, the better the accuracy of depth estimation achieved. A theoretical model of the image formation in a dual mirror, axially symmetrical, catadioptric sensor was proposed by Stürzl et al. [46].

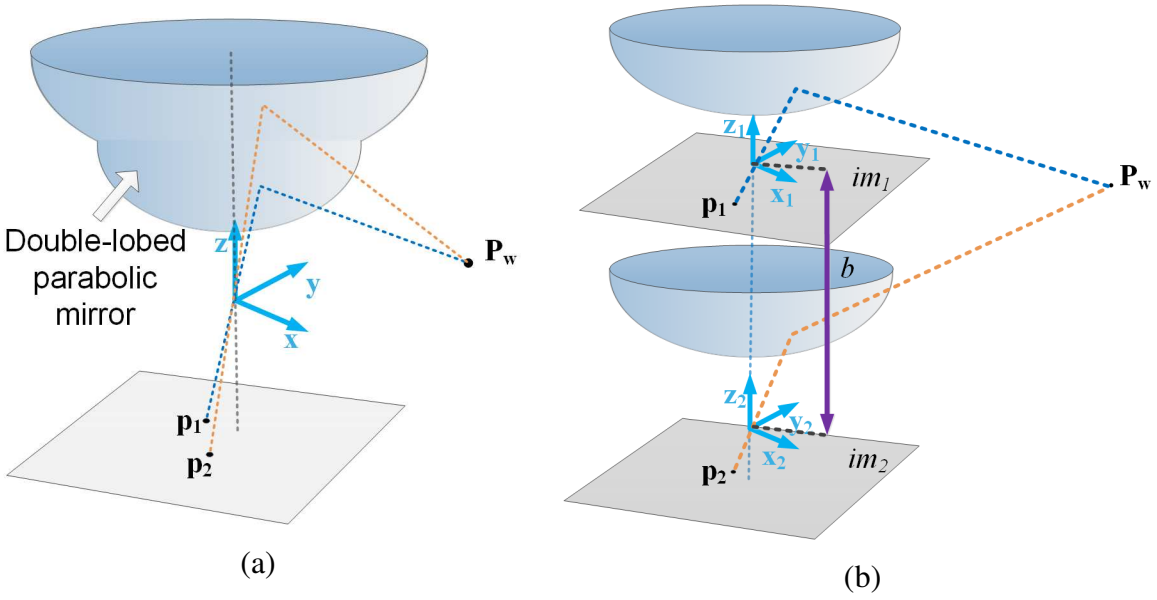


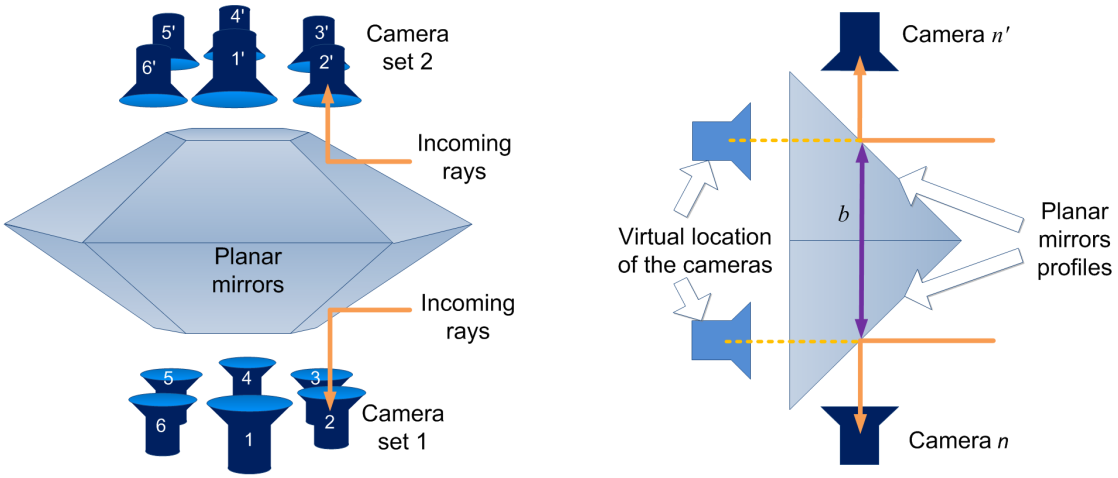
Figure 4.4: Omnistereoscopic catadioptric examples: (a) the camera proposed by Southwell et al. in 1996 using dual-lobe mirror, and (b) an early SVP catadioptric omnistereoscopic camera proposed by Gluckman in 1998: this configuration uses two coaxial catadioptric panoramic cameras with a large vertical baseline to acquire two panoramic views of the scene with vertical parallax b, the 3D scene structure is estimated from the vertical disparity arising between matched feature points in each panorama.

The coaxial catadioptric camera has been a popular method to estimate omnidirectional depth over the past two decades due to its simplicity and hardware economy. Unfortunately, it is not suitable to produce a satisfactory omnistereoscopic rendition of the scene capable to stimulate human stereopsis. The binocular visual process is based on fusing two views of the scene obtained from *horizontally* displaced viewpoints (horizontal parallax). This camera provides two views of the scene in every direction, but based on *vertical* parallax. Although the omnidirectional depth can be estimated using this information, it cannot be used to produce satisfactory horizontal stereoscopic views. For instance, when using a two-dimensional to three-dimensional (2D-to-3D) conversion techniques, the gaps in image coming from occluded areas need to be filled, e.g., using texture synthesis [47]. One of the problems of the vertically coaxial method is that the visual information regarding occluded areas is not acquired.

A similar omnistereoscopic camera based on the catadioptric principle was proposed by Kawanishi et al. [10] in 1998. Their camera consists of two non-SVP catadioptric cameras in vertical co-axial configuration as shown in Fig. 4.5-(a). Each catadioptric camera consists of six planar mirrors, each of which reflects a partial view of the scene over a video camera. This configuration produces twelve video streams covering 360° in azimuth. Each camera in the top of the arrangement is paired with the camera located directly below, i.e., the cameras n and n' form a stereoscopic camera pair, whose vertical baseline is b, as illustrated in Fig. 4.5-(b). Similarly to Gluckman's camera [9], the vertical parallax b between camera pairs enables the panoramic estimation of the scene's depth, but does not provide the means to render viewable stereoscopic images. In a follow up of this design, Shimamura et al. [48] built a working prototype based on the Kawanishi et al. design capable to produce panoramic depth-maps.

Spacek [49] relaxed the non-SVP condition using two conic mirrors, instead of pyramidal mirrors, coaxially aligned with cameras. This configuration was conceived to estimate distances based on vertical disparities. The author reported benefits in using conical mirrors over other profiles in terms of the uniformity of the resolution density. However, this type of profile introduces out-of-focus blurring in some regions of the orthographic image because the optical focus in not uniformly located as in the case of hyperbolic and parabolic mirrors.

A recent interesting development is due to researchers at the Fraunhofer Heinrich Hertz Institute [50], who are currently working on a prototype of an omnistereoscopic high-definition television (HDTV) camera based on a catadioptric design. Conceived for omnistereoscopic 3DTV broadcasting, this set-up uses six stereoscopic camera rigs, each of which is associated with a planar mirror. Each mirror reflects a partial view of the scene on a camera pair,



(a) (b)

Figure 4.5: Omnistereoscopic camera configuration based on coaxial planar mirrors: (a) configuration based on Kawanishi et al. idea, and (b) virtual location of each camera's projection center and vertical baseline b.

for a total of twelve HDTV cameras [51, 16]. These video streams can be mosaicked into a full omnistereoscopic video, or into free-panning 3DTV signal. The concept of this camera is presented in Fig. 4.6. The creators of this camera have reported difficulties when mosaicking partially overlapping stereoscopic frames due to the large parallax between adjacent projection centers [52]. Part of the problem resides, similarly to other star-like configurations, in the excessive parallax introduced by the stereoscopic rigs where both cameras are laterally displaced with respect to each other. The minimum distance to objects in the scene for correct rendering is affected by the large inter-camera parallax, reducing the stereoscopic usability in foreground regions of the scene. This camera corresponds to the *acquisition configuration 4* (Section 3.2), where six stereoscopic camera rigs, equally distributed at a distance $\mathbf{r_c}$ from the geometric center $\mathbf{O}$, capture simultaneously six, partially-overlapped, stereoscopic video signals of the scene.

Peleg et al. [53] proposed a different catadioptric camera configuration capable of acquiring horizontal binocular stereo omnidirectionally and in real time. This catadioptric system uses a complex spiral lenticular lens and an optical prism to acquire a complete omnistereoscopic im-

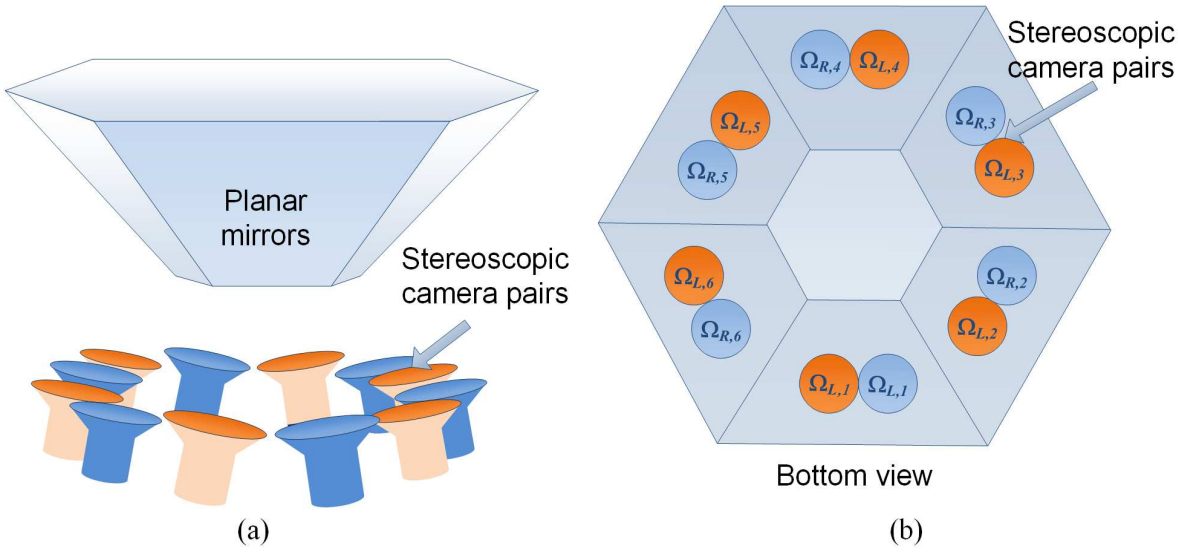
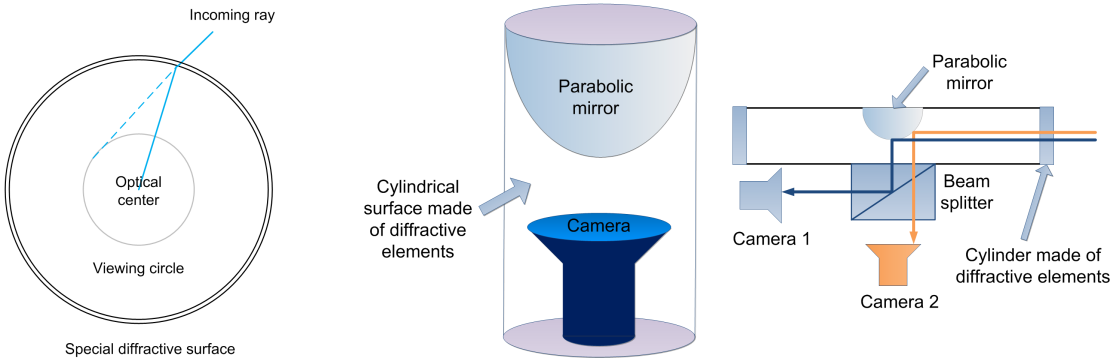


Figure 4.6: Omnistereoscopic video camera developed at the Fraunhofer Heinrich-Hertz Institute: (a) each planar mirror face is associated with a stereoscopic pair of cameras, (b) locations the cameras as seen reflected on the planar mirrors.

age in real time. This configuration deflects incoming light rays as if the scene were acquired simultaneously from multiple perspective points located on a virtual *viewing circle*. This camera can be modeled by the *acquisition configuration 1* illustrated in Fig. 3.2-(a), where a large number of stereoscopic image vertical stripes (central columns of left and right images) are simultaneously sampled and mosaicked to create complete left and right cylindrical panoramas in real time.

Peleg et al.'s lenticular arrangement is shown in Fig. 4.7-(a), which illustrates the acquisition of one eye (left or right) panoramic view. Using a Fresnel-like diffractive system with thousands of lenses to capture both omnistereoscopic views simultaneously, this idea could in theory produce an omnistereoscopic video in real time. The lenticular arrangement can be built around an SVP panoramic camera as shown in Fig. 4.7-(b). The authors have proposed using a beam splitter and the described lenticular system to acquire both viewpoints of a stereo projection simultaneously as illustrated in Fig. 4.7-(c).

This camera could be a solution to the problem of omnistereoscopic image acquisition of dynamic scenes. Furthermore, Peleg, Ben-Ezra and Pritch were granted a patent [54] for this camera in 2004, but no prototype has been built or licensed yet to our best knowledge. This commercialization lag must not be taken as a proof of the inadequacy of the idea; e.g., over



(a) (b) (c)

Figure 4.7: Peleg et al. proposal for a real time omnistereoscopic camera based on a catadioptric principle: (a) a Fresnel-like lenticular lens arrangement diffracts the light over a viewing circle, (b) a catadioptric scheme with a cylindrical diffractive material composed by vertical stripes of the proposed Fresnel lens, to capture one (left or right view) panorama, and (c) using an optical beam splitter such as a prism and combining diffraction lenses for left and right view in the same cylindrical surface, both (left- and right-eye) views can be captured simultaneously.

than 70,000 patents were granted annually in the U.S. by the turn of the century and only very few of them were commercially developed [55]. As a matter of fact, a variation of the lenticular lens, although not an omnistereoscopic application, has been licensed for the production of 3D billboards [56]. Peleg et al. proposed a geometrical model of their elliptic lens [57]; however, there are still aspects of the image formation for this camera that have not been extensively studied. More importantly, the capabilities of such an optical approach to produce high-quality omnistereoscopic imagery are open for further studies.

It is important to remark that a cylindrical stereoscopic panorama produces a correct binocular experience only in the center of each rectangular view extracted from these left and right panoramas. The peripheral image region, outside the image center produces a distorted binocular perception [58, 23]. Although this is a noticeable effect, the user tends to focus in a region of interest at the center of the image, where the binocular perception is correct, reducing the likelihood of uncomfortable effects that could lead to eye strain. Furthermore, if a cylindrical omnistereoscopic image created with this acquisition method were projected in a cylindrical surface around the user (located at the center), a correct binocular depth would be experienced by looking in any direction around the user, as long as the zero parallax distance (eyes ver-

gence) coincide with the distance to the cylindrical screen [59].

Other configurations based on catadioptric cameras have appeared in the last decade mainly focused in the problem of omnidirectional stereo reconstruction [60, 61] following the idea of co-axial catadioptric stereo, which makes them inadequate to produce omnistereoscopic imagery suitable for binocular human viewing.

4.2.3 Pros and cons of the catadioptric omnistereoscopic cameras

The most evident advantage of an SVP catadioptric camera is its simplicity: a single camera and dioptric system can sample the scene's visual field, in addition to estimating the panoramic depth, in a single snapshot. The SVP approach avoids stitching problems and imperfections that arise by parallax between multiple projection centers in non-SVP cameras. However, focusing the light uniformly on a planar image sensor after reflecting light in a non-planar mirror is problematic. Not all the rays are parallel after reflection in a parabolic surface, nor are they perfectly reflected towards the convex focal point in a hyperbolic mirror. In practice, image blurring as a function of the radial distance from the center of the image [42] is difficult to avoid. This problem can be reduced, although not completely, by a careful design of the catoptric system (mirror) and its dioptric (focusing) counterpart. The high resolution CCD sensors available nowadays can help to reduce problems while projecting the acquired orthographic projection of the scene into a canonical panoramic surface. Furthermore, a catadioptric camera has advantages with respect to other mirror-less systems in terms of reducing chromatic aberrations present in most aspheric lenses, e.g., fisheye lenses. Offsetting the advantages of catadioptric cameras, there are still inherent problems in using this camera to render an omnistereoscopic scene for binocular viewing.

Although catadioptric omnistereoscopic configurations using vertically coaxial mirrors are undoubtedly an elegant method for acquiring omnidirectional depth maps in real time, the camera configurations are unable to handle occlusions in the scene. A binocularly correct stereoscopic view can be synthesized with one panoramic image plus a dense horizontal disparity information of the scene. The disparity information can be used to generate a pixel-wise horizontal displacement map. Applying this map to displace horizontally pixels (or regions) in the image can produce a correct illusion of depth. The necessary information to produce this omnistereoscopic image can be acquired using an omnistereoscopic catadioptric configuration: a correct panorama of the scene, and an omnidirectional depth map. Unfortunately, the information to fill image gaps (occluded areas) cannot directly be obtained using a camera

configuration based on only vertical parallax. But, there are still suboptimal solutions coming from the field of 2D-to-3D conversion that can be applied. For instance, pixels can be copied from the adjacent regions of the missing areas to fill these gaps. A much better alternative will be to simultaneously acquire a different view, e.g., a horizontal stereoscopic pair for each gazing direction. The parallax view information can be used to fill-in the occluded areas using a texture inpainting technique [62].

A good candidate to produce directly viewable omnistereoscopic imagery, capable to satisfy the real time acquisition of dynamic events as well, is the optical diffractive approach proposed by Peleg et al. [63, 54] illustrated in Fig. 4.7. However, the design and implementation of such a lenticular arrangement is challenging. Similarly to Peleg's approach, there has been at least one other recent proposal, also based on lenticular arrays to multiplex two horizontal binocular views of the scene, but using non-SVP configurations [64] which will be presented in Section 4.5. Besides the lack of commercial interest for an optical based approach, an alternative solution, based on off-the-shelf lenses and camera sensors, can provide a satisfactory solution to the problem with reduced hardware complexity.

Another good candidate is the panoramic 3DTV camera developed at the Fraunhofer Heinrich-Hertz Institute. This omnistereoscopic camera, which is illustrated in 4.6, uses multiple off-the-shelf high-definition television (HDTV) cameras and mirrors to produce real-time stereoscopic videos with broadcasting quality. The downfall of this camera is its size [51] which makes difficult the mosaicking of individual video streams. Additional improvement may be achieved by using a different camera distribution to enable registering stereoscopic images of scene elements closer to the camera.

4.3 Sequential acquisition methods

A literature review of omnistereoscopic methods and configurations cannot be complete without mentioning the family of sequential acquisition methods. It is necessary to point out that sequential methods are intrinsically inadequate to acquire dynamic scenes omnistereoscopically since they require the scene to be static for correct rendering. However, many multiple camera configurations that will be presented in Section 4.5 can be directly traced back to parallelized (simultaneous acquisition) versions of sequential methods presented in this section. Therefore, these sequential techniques deserve a closer look.

The sequential acquisition of images has been widely used to produce high-quality panoramas. The idea is simple: using a single camera, it is possible to capture samples of $\mathbf{C_L}$ and $\mathbf{C_R}$ functions as defined by the omnistereoscopic viewing model in Section 2.3.7. These sequential sampling leads to an accurate estimation of $\mathscr{I}_S$ accordingly to the omnistereoscopic viewing model of $\mathbf{C_S}$ and, after sampling, the rendering of $\mathscr{I}_S$ for the chosen format of panoramic representation.

An example of sequential sampling of $\mathbf{C_S}$ is using *acquisition configuration 1*. A number of partially overlapped stereoscopic snapshots at discreet gazing directions θ_i, for $i \in \{0, ..., N-1\}$, can be used to reconstruct an approximation of $\mathscr{I}_S$. This camera set-up is illustrated in Fig. 4.13.

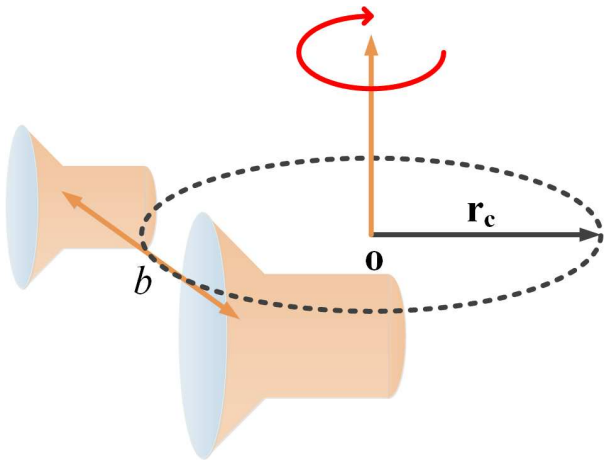


Figure 4.8: Example of sequential acquisition.

One of the first methods to sample $\mathbf{C_S}$ function at a reference viewpoint $\mathbf{r}$ has been to rotate a stereoscopic camera rig around the midpoint between nodal points. This scheme corresponds

to the *configuration 1* (Section 3.2). A valid rendering strategy in this case is mosaicking vertical image stripes from the center of each image pair, to sample $\mathscr{I}_S$ to create $\mathbf{I_L}$ and $\mathbf{I_R}$ panoramic images. The number of stereoscopic images to be acquired determines the incremental panning angle Δ_θ and, consequently, the width of each image to be mosaicked. As was mentioned in Section 4.2.2, the stereoscopic panoramas are correct only at the center of each view, becoming distorted outside the central region. Note that each camera's projection center (nodal point) defines a *viewing circle* of diameter $(b = 2\, r_c)$ during the scene panning.

Alternatively, rotating the stereoscopic rig about one camera's nodal point corresponds to the *configuration 2* (Section. 3.2). Unfortunately, the latter strategy cannot be implemented using multiple cameras for a simultaneous (parallel) acquisition.

The other sequential acquisition strategy corresponds to the *configuration 4* (Section 3.2). In this case, the scene is sampled by acquiring a sequence of stereoscopic snapshots, rotating the stereoscopic rig off-center translation $\mathbf{r_c}$.

A modified version of this sequential method consist of radially aligning the nodal point of one of the cameras with the center $\mathbf{O}$, which corresponds to the *configuration 3* (Section 3.2). Both sequential variants can be parallelized for simultaneous acquisition using multiple cameras.

These sequential techniques, their variations, and other alternatives are individually detailed in the next section.

4.3.1 Omnistereo using sequential acquisition: antecedents

Perhaps one of the most illustrious application of omnistereoscopic sequential acquisition was integrated to the Mars Pathfinder rover [28] in the late 1990s. The camera can be modelled by the *configuration 1* presented in Section 3.2. It was designed to provide a variety of telemetric measurements beyond producing omnistereoscopic imagery; actually, producing stereoscopic images suitable for human stereopsis was not the primary goal of this camera. Two cameras were mounted on a retractable mast and were rotated around the mid-point between each camera's nodal point. The pair of cameras, whose resolution was modest as is expected for an interplanetary probe of that era (512×256), were toed in 0.03 rad, defining a fixation point (zero parallax distance) at 2.5 meters from the rotation center. The stereoscopic panoramas produced by this camera received much attention in the news [65, 66, 67]. Although these omnistereoscopic images are not impressive in terms of quality of the binocular stereo, they constitute an important precedent for the marketing value of realistic immersive imagery to

promote a planetary mission to the layperson [68].

Other authors reported variations on the rotating stereoscopic rig method that are worth mentioning. For instance, Romain et al. [69] proposed a rotating platform with two degrees of freedom, which can rotate in azimuth and elevation to produce spherical omnistereoscopic images. More recently, Ainsworth et al. [70] revisited the method of a rotating stereoscopic rig to minimize (not to eliminate as the authors stated) stitching problems. They reported their method to create stereoscopic panoramas [71] based on mosaicking partially overlapped scene's images, which they tested using an off-the-shelf digital camera (Panasonic Lumix) and a commercial rotating platform (GigaPan EPIC Pro Robotic Controller).

In ref. [11], Ho-Chao Huang and Yi-Ping Hung proposed a rotating camera to acquire stereoscopic images of scenes, which were aligned and mosaicked to render omnistereoscopic images. Their idea, originally published in 1996, is illustrated in Fig. 4.9-(a) showing a central camera which is rotated around its nodal point while a second camera is rotated off-axis providing a parallax view of the scene. Their method corresponds to the *configuration 2* (Section 3.2), which produces non-SVP stereoscopic panoramas, where the stereo budget can be selected by choosing the baseline $r_c = b$ of the rotating stereo rig.

This configuration is interesting because it helps to acquire stereoscopic images where one camera (central) is in the optimum position to minimize stitching errors (nodal point), while the second camera captures a second view of the scene with horizontal parallax. The central camera produces an SVP panorama while the second camera can be used to render a binocular polycentric panorama. Because of camera self-occlusion this idea is difficult to implement using multiple cameras, although a multi-camera configuration using planar mirrors can approximate an SVP for a central camera and acquire a horizontal parallax view simultaneously. There have not been proposals based on this catadioptric design to our best knowledge.

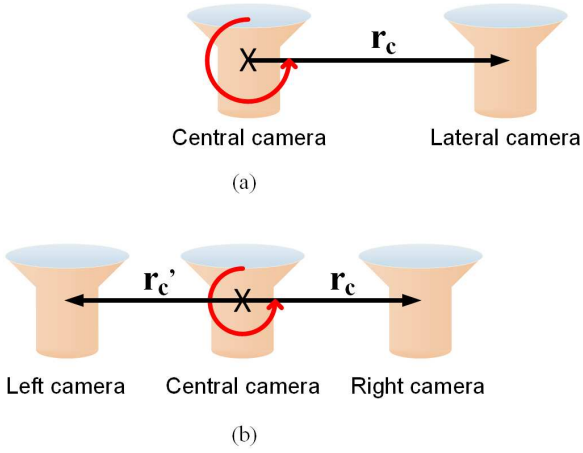


Figure 4.9: Omnistereoscopic methods based on sequential acquisition of partially overlapped images: (a) method proposed by Huang et al. in 1996 to generate a correct panorama (central camera) and accessory information to estimate a horizontally parallax view (lateral camera), and (b) the acquisition strategy proposed by Yamada et al., which is similar to Huang's method, is based on acquiring images to produce an SVP panorama (central camera) and, in this case, estimating the panoramic depth map based on a large baseline stereoscopic pair of images (left and right cameras).

Using a similar approach, Yamada et al. [33, 32, 72] proposed the triple camera system shown in Fig. 4.9-(b). Similarly to Huang et al. method, the central camera produces an SVP panorama while the two satellite cameras help to estimate a panoramic disparity map using a large baseline. This sequential acquisition technique can be seen as a modification of the *configuration 1*, used only to estimate depth by exploiting the larger baseline ($b = 2 \cdot r_c$) between the satellite cameras, and adding a central camera to produce an SVP panorama. This acquisition scheme is a hybrid between *configuration 1* and *2*. Again, self-occlusion between cameras makes it difficult to parallelize Yamada's approach in a simultaneous acquisition scheme. But a catadioptric scheme to acquire an SVP and a multi-sensor scheme to simultaneously acquire a number of partially overlapped stereoscopic snapshots of the scene is an interesting design challenge and a suitable approach for dynamic scene omnistereoscopic acquisition. No camera following the suggested approach has been proposed to our best knowledge.

A variation of the idea of rotating sensors was proposed by Chan and Clarke [73] who devised a camera where a mirror is rotated while a single sensor sequentially captures binocular stereoscopic images of the scene. Their idea was inspired from endoscopic applications

where a single probe has to be inserted in a biological cavity or in an archeological site. The principle behind the idea is similar to other proposals in terms of a sequential acquisition. For instance, carefully selecting the planar mirror and camera location, the *configuration 1* can be implemented with this technique.

Ishiguro et al. [12] proposed in 1992 a method to create omnistereoscopic imagery based on a single rotating camera, but for robot navigation instead of human visualization. This method corresponds to the *configuration 1*. Peleg and Ben-Enzra [31] rediscovered this method in the late 1990's, but tailored the idea with human visualization in mind. This method has became one of the most popular sequential techniques to create high-quality omnistereoscopic imagery given its hardware economy and simplicity.

The principle is based on a single camera rotating around a point behind its projection center, as depicted in Fig. 4.10-(a). In one complete rotation, the camera captures a number of images which are used to extract left ($\mathbf{im_L}$) and right ($\mathbf{im_R}$) columns. These two image columns correspond to the back-projection of the image's vertical stripes defined by intersecting the rays passing through the camera projection center $\mathbf{O}$ and the projection points $\mathbf{O_L}$ and $\mathbf{O_R}$. The ray tracing concept is depicted in Fig. 4.10-(b). These columns are then mosaicked producing left and right-eye panoramic views. The end-result is equivalent to the *configuration 1* (Section 3.2) used to acquire left and right views column-wise. There have been proposals for omnistereoscopic cameras that can be directly traced to parallelization of a single rotating camera [15, 74, 75].

The ubiquity of digital cameras has motivated the continuous research on rendering techniques based on Ishiguro-Peleg's method using a singular off-centered camera. For instance, in [76] the authors presents their approach in solving the alignment between snapshots acquired with a hand-held mobile camera used to manually pan the scene. Disney Research has also being doing research recently in the alignment corrections needed when using Ishiguro-Peleg's acquisition method to produce omnistereoscopic imagery [77, 78].

The method has the advantage of defining a virtual baseline ($b = 2\ r_c$) that can be varied according to the stereoscopic budget desired for the scene by changing the relative distance between left- and right-eye vertical stripes extracted from the successively acquired images [79]. Changing the distance has the effect of changing the viewing circle diameter. More recently, Wang et al. proposed an interactive method to adjust the disparity in panoramas created using the single off-centered rotating camera [80] based on the interactive selection of objects in the scene.

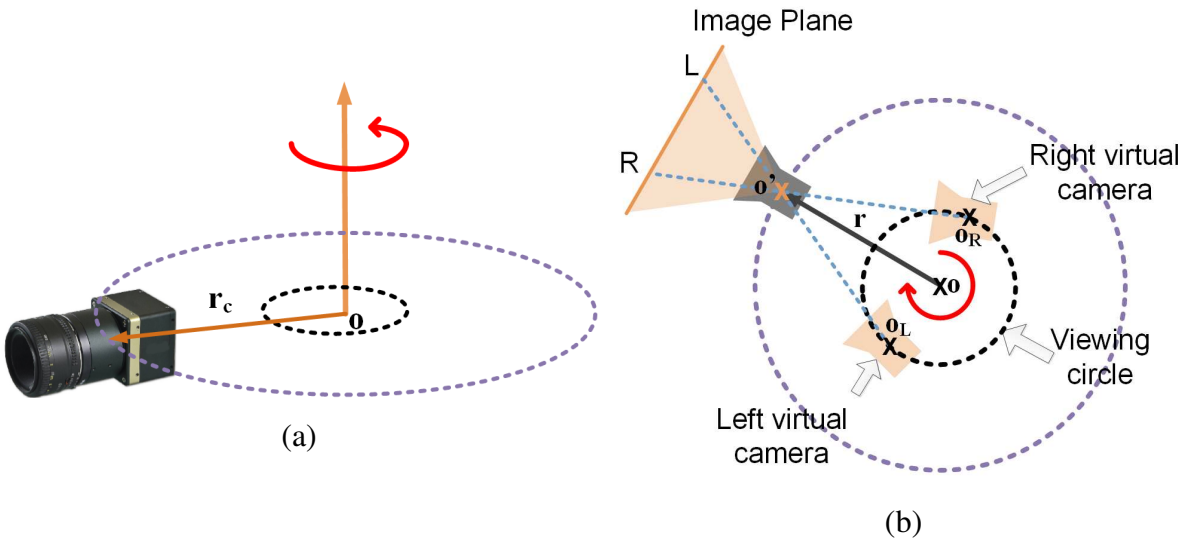
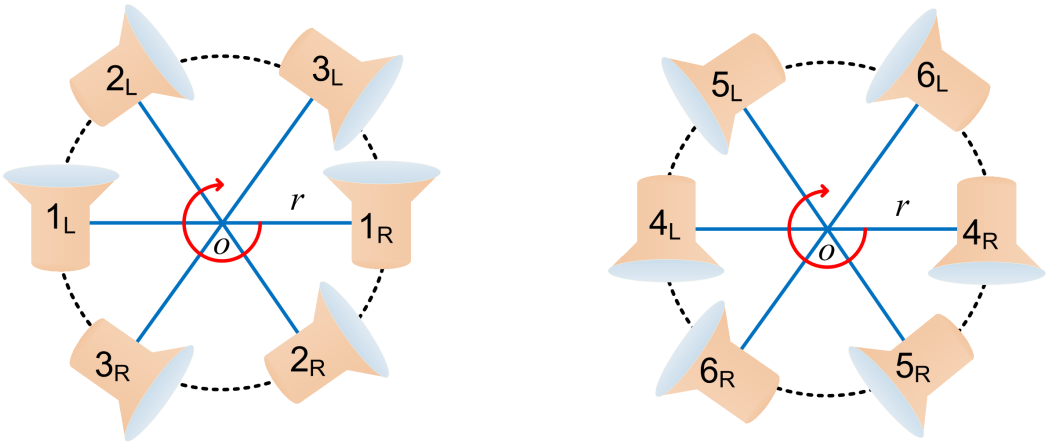


Figure 4.10: Rotating method to produce omnistereoscopic imagery based on a single rotating camera: (a) a single camera is rotated from a off-centered location, and (b) two projections corresponding to left and right-eye projections can be defined intersecting the rays passing through the cameras projection center $\mathbf{O}$ and the points $\mathbf{O_L}$ and $\mathbf{O_R}$ defined over a virtual *viewing circle*.

Examples of 3D (not stereoscopic images) created using the *central stereoscopic rig* method for monoscopic panoramas and omnistereoscopic images can be visited online [81, 82]. Additionally, several patents have been granted to Peleg et al. for acquiring, disparity adjusting, and displaying stereoscopic panoramas using the single camera method [83, 84, 85], one of which has been licensed to HumanEyes [56], not to create omnistereoscopic imagery, but to create multi-view 3D billboards.

A different approach was proposed by Hongfei et al. [29] who proposed using a single digital single-lens reflex (DSLR) camera to successively acquire twelve images of an static scene. Firstly, three stereoscopic image pairs (six images in total) are acquired by placing the camera in the positions labeled (iL, iR), for $i \in \{1, 2, 3\}$, as indicated in Fig. 4.11-(a). After that, the camera is rotated $180°$ and six new stereoscopic images are successively acquired following the pattern illustrated in Fig. 4.11-(b). This acquisition scheme corresponds to the *configuration 1* illustrated in Fig 3.2-(a). Although the Hongfei et al.'s method is not suitable for dynamic scenes acquisition, a parallelized version of the same idea was already proposed by R. G. Baker et al. [86] as a patent application in 2003.



(a) (b)

Figure 4.11: Method based on rotating a single camera to capture six stereoscopic images by positioning the camera in twelve different locations and orientations: (a) first, the stereoscopic pairs (1L,1R),(2L,2R) and (3L,3R) are acquired one-by-one rotating and positioning the camera to the corresponding locations, and (b) finally, the camera is rotated $180°$ around its nodal point and the pairs (4L,4R),(5L,5R) and (6L,6R) are sequentially acquired.

Among the panoramic sequential methods, those based on using line sensor cameras deserve a particular mention. Cylindrical omnistereoscopic imagery produced by this method produce geometrically correct binocular views of the scene at the center of the image, while the depth perception is distorted in the periphery of the image [87]. This viewing paradigm is valid for cylindrical panoramas projected in a cylindrical display, in both monoscopic and stereoscopic panoramas, in part because it does not have to deal with parallax ghosting while blending images. Sequential line scanning is based on mosaicking image columns a single pixel wide and therefore it produces high-quality stereoscopic panoramas. However, this virtue is offset by their lengthy acquisition time, which is common to all sequential scanning methods. The line scanning methods to acquire omnistereoscopic imagery can be modelled by the *configuration 1* (Section 3.2).

High-quality omnistereoscopic images, in cylindrical format, can be produced using line-scanning sensors. This sequential method has been studied over the first decade of the 2000's [21], partially thanks to the commercial availability of line-scanning cameras [88]. As their name indicates, these cameras acquire the scene column-wise. An omnistereoscopic view of the scene can be acquired by rotating the line-scanning sensor off-center at radius $\mathbf{r_c}$ to acquire

column by column left ($\mathbf{I_R}$) and right ($\mathbf{I_L}$) cylindrical panoramas, one at the time. The acquisition model corresponds to the *configuration 1* illustrated in Fig. 3.2-(a), where a line-scanning sensor is in the position of each of the cameras for one complete rotation, acquiring successively each eye viewpoint after two complete rotations. Several authors have contributed to the understanding and modeling of omnistereoscopic imagery using line sensors, covering the line-scanning camera calibration [21, 89], a geometrical model for polycentric panoramas using this acquisition strategy [20], and the omnistereoscopic image acquisition [90, 30, 91, 92]. Although the literature on line sensors is extensive and insightful, this approach cannot be adapted to acquire dynamic scenes.

Hybrid approaches have been proposed that use a laser range finder and rotating sensors to provide a high resolution depth mapping of the scene. For instance, Jiang and Lu [93] proposed a method that combines an off-center CCD camera with a laser range sensor which together produced a monoscopic panorama and its dense depth map. Their approach is shown in Fig. 4.12. Once again, the problem addressed was the reconstruction of a 3D scene and not to produce binocular omnistereoscopic imagery. Beside its sequential conception, this idea cannot be used for an omnistereoscopic 2D-to-3D conversion since occlusion information is not collected during the sequential acquisition. Similar limitations in handling occlusions arise in a recent proposal by Barnes [94] which combines ground based light detection and ranging (LIDAR) and monoscopic panoramas.

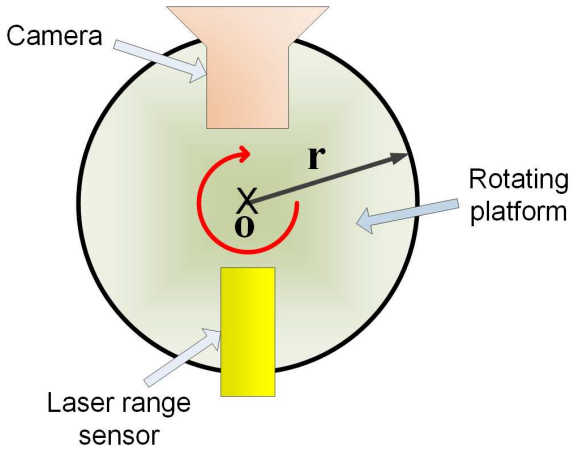


Figure 4.12: Combined camera and laser range sensor in a rotating platform for sequential acquisition: an SVP panorama plus a panoramic depth map can be obtained for static scenes using this acquisition method.

The rotating sensor for omnistereoscopic imagery has appeared with some variations over the last decade [95, 87]. A good summary of methods to create omnistereoscopic images based on rotating sensors was published in 2006 by Burke [96].

4.3.2 Pros and cons of the sequential acquisition

Sequential acquisition strategies are an interesting starting point to devise multi-camera configurations for real time omnistereoscopic acquisition. The configuration proposed by Ho-Chao Huang and Yi-Ping Hung [11] is interesting since it presents a solution to reduce the parallax-induced errors while stitching multiple images. This method can be parallelized to simultaneously acquire a number of image patches and their corresponding stereoscopic counterparts in a non-SVP scheme. The sequential method of Peleg and Ben-Enzra using a single off-centered camera [31] is interesting for reducing the hardware involved, but it is difficult to parallelize since it would require a prohibitively large number of cameras to prevent blending artifacts while mosaicking. In that case, the mosaicked image columns have to be constrained to a few pixels wide, which implies taking hundred of snapshots. A multiple camera configuration that attempts to take this number of simultaneous pictures is impractical, but, as will be shown in the Section 4.5, some multiple camera configurations that do this parallel acquisition have been proposed.

The large overhead when acquiring hundreds of snapshots to just use two narrow slices is partially compensated by line cameras which scan the scene column-wise. Unfortunately, the large acquisition time of line cameras limits them to controlled static environments such as most indoor scenarios.

However, it is possible to devise improvements in acquisition speed of omnistereoscopic images using line sensors. For instance, using multiple stereoscopic line sensors, the acquisition speed can be increased linearly with the number of stereoscopic sets. This approach would enable the simultaneous capture of non-overlapping stereoscopic images of the scene, which can then be mosaicked to create a full omnistereoscopic image in a fraction of time that a single sensor can achieve. Unfortunately, a rotating camera system with such characteristics is still commercially unavailable and, if it were, it would not be suitable to acquire dynamic scenes.

4.4 Omnistereo based on panoramic sources

The panoramic based methods use panoramic snapshots $\mathbf{I}_n$ $(n \in \{1, ..., N\})$ as raw material to synthesize omnistereoscopic images of a scene. This approach relies on sampling the omnidirectional function $\mathbf{C}$ from a number of different viewpoints $\mathbf{r_n}$ $(n \in \{1, ..., N\})$, which provide a set of panoramic images $\mathscr{I}_n$. These panoramic images are sampled to produce the panoramic representations $\mathbf{I}_n$. An omnistereoscopic view of the scene $\mathscr{I}_S$ can be built at any desired viewpoint $\mathbf{r}$ based on the set of $\mathscr{I}_n$, providing their reference viewpoint $\mathbf{r_n}$ is known.

An example illustrating the use of multiple panoramic snapshots to estimate $\mathscr{I}_S$ at different viewpoints [2] is shown in Fig. 4.13. In this example, the $\mathbf{C_n}$ are clusters of three panoramic snapshots acquired in a triangular co-planar pattern. The method uses the three panoramic images $(\mathbf{I_{1n}}, \mathbf{I_{2n}}, \mathbf{I_{3n}}$ to estimate an omnistereoscopic view for the center geometric center of each cluster.

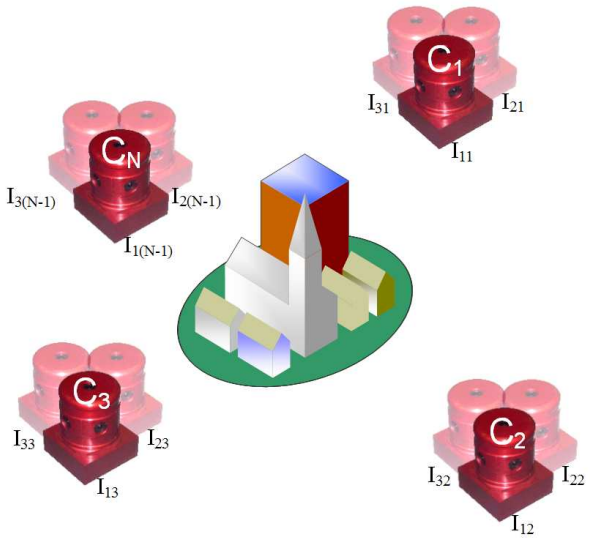


Figure 4.13: Estimation of $\mathscr{I}_S$ at different viewpoints using the information acquired from multiple cylindrical panoramas.

This is a relatively new alternative resulting from the commercial availability of panoramic cameras during the last decade, which made capturing panoramic snapshots of the scene practical.

4.4.1 Omnistereo using panoramic sources: antecedents

The idea of using multiple stereoscopic panoramas of the scene to perform a 3D mapping of a scene has started in the late 1980's [97]. For instance, Ishiguro et al. [12] proposed to render a sequence of omnistereoscopic images, not for human viewing, but to estimate the 3D relationship within objects in a scene. To do this, the authors proposed using a mobile platform (a robot) equipped with a rotating camera mounted on top. The mobile unit moved on a pre-programmed route, stopping at intervals to acquire omnistereoscopic images of the scene using the single rotating sensor. The *configuration 1* was used. This sequence of omnistereoscopic views can be used to estimate the distance to obstacles from the travelling path by matching feature points between stereoscopic images obtained between successive panorama views, e.g., exploiting motion parallax between samples. This method is constrained to static scenes, the panoramas have to be coplanar, the location of each panoramic sample has to be precisely known, and more importantly, the accuracy of the estimation decreases as soon as the viewing direction approximates the direction of motion progression (robot acquisition path). The panoramas can be aligned by determining the cylindrical epipoles, or alternatively, finding the focus of expansion [34] direction using motion analysis between panoramic frames in the sequence (Appendix B). The method is only valid for a limited subset of panning angles around the perpendicular of the planned trajectory. Furthermore, uncertainties in the stereoscopic baseline lead to inconsistencies in the final depth perceived from different viewpoints.

Similarly to Ishiguro, Kang and Szeliski presented a method [98] to reconstruct the 3D structure of a scene using multi-base stereo obtained from randomly located panoramic samples (Fig. 4.14). The idea was to use stereo matching from multiple viewpoints, whose locations were mapped, to estimate an omnidirectional depth map. Fleck also described a method similar to Ishiguro's to reconstruct a 3D model of the scene [99]. These methods were conceived for robot navigation and not to produce omnistereoscopic renderings for binocular human viewing.

Many techniques were specifically conceived for the off-line navigation of image-based stereoscopic virtual environments. For instance, Yamaguchi et al. [100] and, in a more recent follow up, Hori et al. [101], proposed a method to generate stereoscopic images based on a light-field rendering of synthetic novel views using panoramic video frames as sources. This approach enables a smooth navigation of the scene, but creating large data overhead. A key problem with this idea is the exact position of each panoramic frame must be known a priori

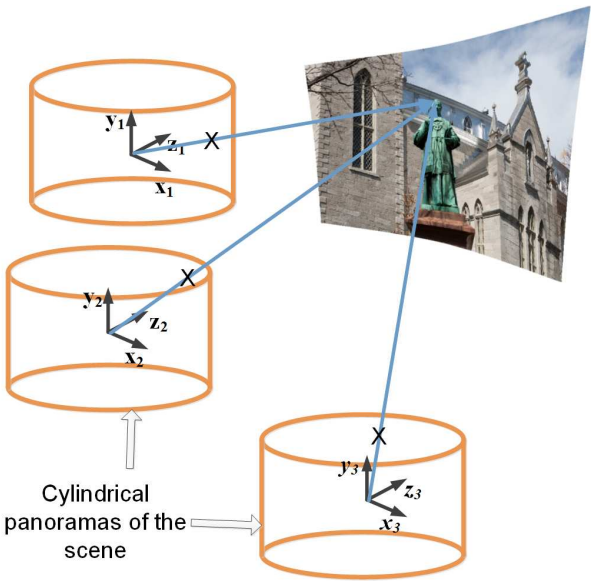


Figure 4.14: Multi-base stereo reconstruction of a scene based on information acquired from multiple cylindrical panoramas.

to find the best panoramic pair to render a binocular image for each user's virtual location and gazing direction. In addition, it is not practically feasible to acquire the panorama frames in an accurate equally spaced grid to have control over the stereoscopic baseline. To cope with this problem, Hori et al. [45] proposed capturing a panoramic video by simultaneously using two panoramic cameras mounted on a rig. Unfortunately, this approach cannot provide a full omnistereoscopic rendition of the scene, but only in directions perpendicular to the travelling path, and does not solve the data overhead problem.

Vanijja and Horiguchi proposed an alternative method [36] specifically tailored for a *cave automatic virtual environment* (CAVE) type of display. Their idea relies on a limited set of four panoramas acquired in a controlled square sampling pattern, as illustrated Fig. 4.15-(a), which were used to render four wide field-of-view panoramic images. These four partial stereoscopic views were projected on each wall of a CAVE [102] to produce an omnistereoscopic immersive experience. The image patches extracted from each panorama of the cluster were arranged according to the camera panning direction in azimuth. This mapping is illustrated in Fig. 4.15-(b). The type of display, in this case a CAVE display, imposed restrictions in terms of the number and the spatial distribution of cylindrical panoramas to use. The authors constrained the viewable stereoscopic range to certain elevation angles because the distinct cylindrical projec-

tion centers of each stereoscopic pair produced undesirable vertical disparities at low and high gazing angles in elevation. Even with these problems, this method offers advantages in terms of acquisition time and depth consistency between sampled viewpoints, making it suitable for a practical stereoscopic telepresence system. Unfortunately, this technique does not satisfy the dynamic scene acquisition constraint since individual panoramas still need to be acquired sequentially. This omnistereoscopic acquisition approach can be modelled using the *configuration 1* where four wide-angle (90° FOV in azimuth) stereoscopic snapshots of the scene are sequentially or simultaneously acquired for $\mathscr{D} = (\theta, 0°)$, where $\theta \in \{0°, 90°, 180°, 270°\}$ in azimuth to produce the image patches $(\mathbf{im}_{L,i}, \mathbf{im}_{R,i})$, for $i \in \{0...7\}$ as indicated in Fig. 4.15.

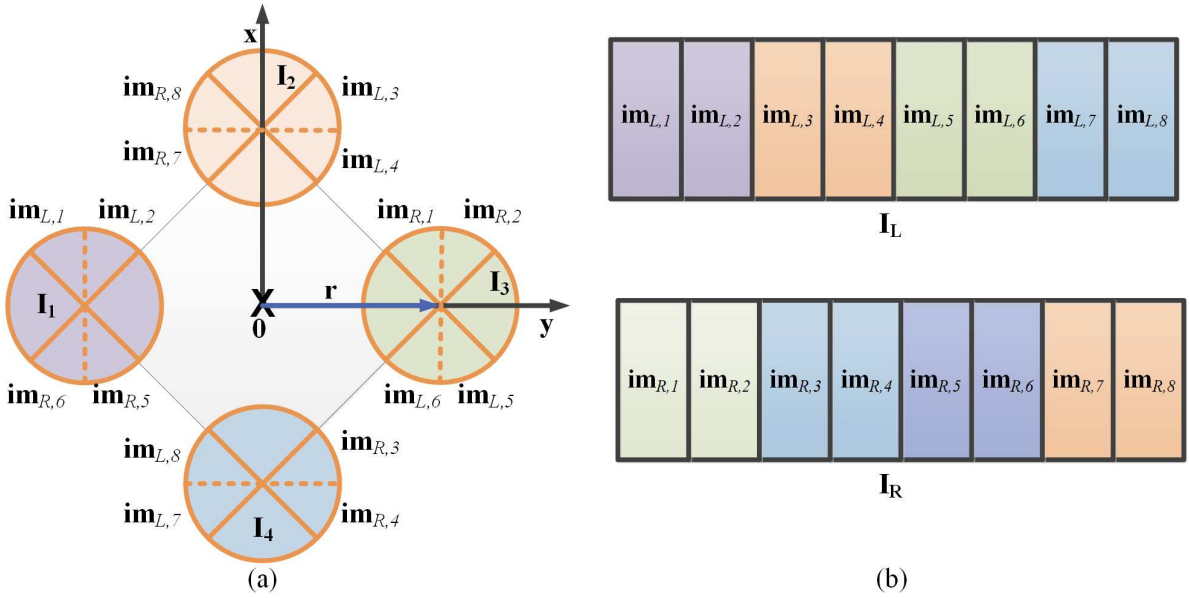
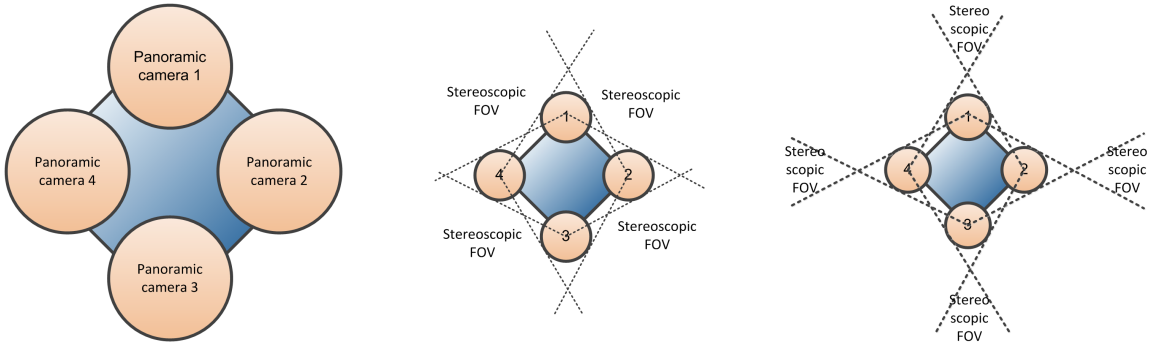


Figure 4.15: Vanijja and Horiguchi proposed in using clusters of four panoramic snapshots of the scene, taken in a square pattern, to extract eight wide-angle stereoscopic images to render a full omnistereoscopic image: (a) the mapping of the eight image sections from panoramas $\{\mathbf{I_1}, \mathbf{I_2}, \mathbf{I_3}, \mathbf{I_4}\}$, and (b) the mosaicking of these sections to create a cylindrical omnistereoscopic pair $(\mathbf{I_L}, \mathbf{I_R})$.

An interesting antecedent of the method of Vanijja et al. can be found in a patent application by Baker et al. [86] from 2003, which was granted in 2007. This omnistereoscopic camera proposal is built around four panoramic cameras in a square pattern. The camera is shown in Fig. 4.16-(a) illustrating four panoramic cameras in a square distribution. The arrangement simultaneously acquires four cylindrical panoramas that are used to compose four wide angle

stereoscopic views of the scene as is illustrated in Fig. 4.16-(b) and (c). Although this is a multi-camera system, it is relevant to include it in this section since it is a parallelization of the sequential method proposed by Vanijja et al. three years later [14]. It should be mentioned that this configuration has a minimum distance to the camera where correct stereoscopic acquisition is possible; however it has the potential to produce omnistereoscopic images of dynamic scenes. Another drawback is its fixed stereoscopic baseline which limits its use to certain scenes. This is a good candidate for the real time omnistereoscopic acquisition of dynamic scenes, but no attempt has been done in formalizing the stereoscopic image formation of this camera.



(a) (b) (c)

Figure 4.16: The omnistereoscopic camera patented by Baker on behalf of Image Systems Inc. proposed a parallel acquisition of a panoramic cluster that precedes Vanijja and Horiguchi's proposal: (a) using four panoramic cameras in a square pattern to simultaneously acquire four overlapping cylindrical panoramic snapshots of the scene, (b) a possible rendering strategy using sections of each cylindrical panorama to render stereoscopic close-ups, and (c) and alternative with a slightly larger baseline, but where the regions for stereoscopic rendering are farther away form the camera.

A different approach, also based on using a cluster of cylindrical panoramas in a controlled pattern, has been developed as part of this thesis research [2]. This technique was designed to reduce the acquisition time and improve the overhead efficiency. The idea improves upon the method proposed by Vanijja et al. [36] by using clusters of three co-planar cylindrical panoramic snapshots of the scene. Each cluster has an equilateral triangular pattern, where each side length is directly related to the desired stereoscopic baseline. The triangular cluster

is shown in Fig. 4.17-(a), where the projection centers of each cylindrical panorama are at the vertices of the triangle. Once the panoramas are aligned [34], it is possible to map and extract stereoscopic images from pairs of panoramas within the cluster as a function of the panning direction, similarly to Vanijja's method. A similar idea based on panoramic triad was suggested by Zhigang Zhu in 2001 [38]. An example of this mapping for a triad of cylindrical images is illustrated in Fig. 4.17-(b), where each image section of panoramas $(\mathbf{I_1}, \mathbf{I_2}, \mathbf{I_3})$ corresponds to a particular camera panning direction in azimuth. Mosaicking these six pairs of images renders a full omnistereoscopic view of the scene in a fraction of the time needed by sequential methods based on acquiring an omnistereoscopic image column-wise. This method can be modeled by the *configuration 1* (Fig. 3.2-(a)), where six stereoscopic pairs $(\mathbf{im}_{L,i}, \mathbf{im}_{R,i})$, for $i \in \{0, ..., 5\}$, are sequentially acquired. This acquisition strategy is illustrated in Fig. 4.17.

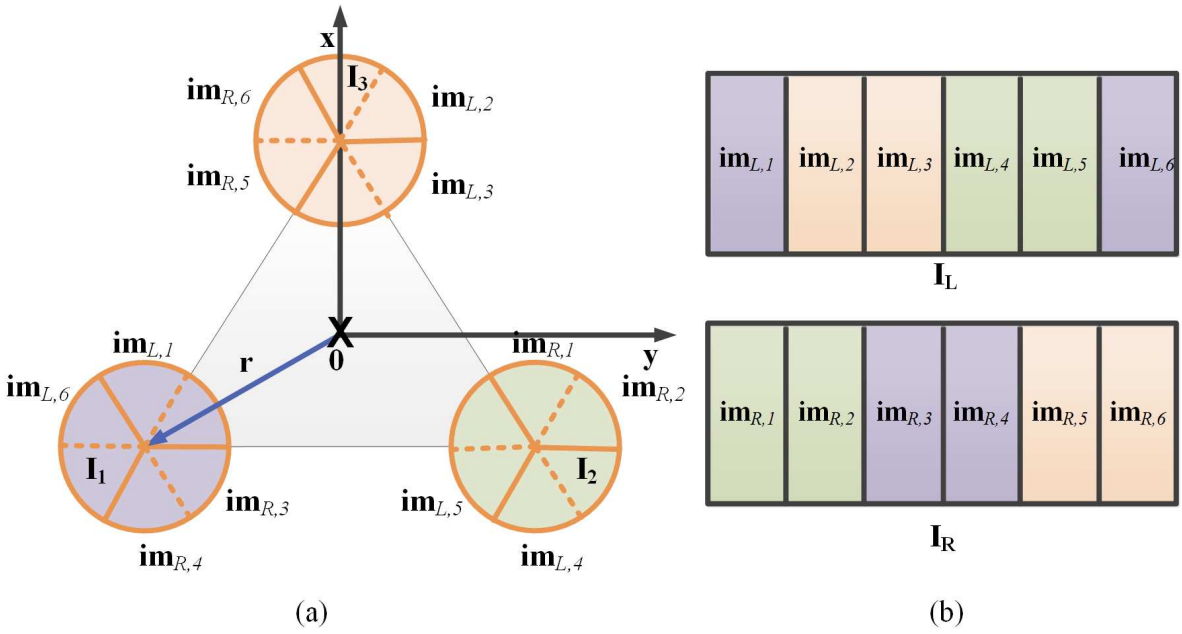


Figure 4.17: Omnistereoscopic images using a cluster of three panoramas: (a) three cylindrical panoramas $(\mathbf{I_1}, \mathbf{I_2}, \mathbf{I_3})$ in a co-planar, triangular pattern can be used to extract six image sections, and then mosaicking them using the sequence illustrated in (b) to create two novel stereoscopic views $(\mathbf{I_L}, \mathbf{I_R})$.

Details about the panoramic triad method and the necessary panoramic alignment can be found in Appendix A and B, respectively.

4.4.2 Pros and cons of using panoramic sources

Although the sequential methods based on panoramic sources are not suitable to acquire dynamic scenes, some of them can be adapted to multi-sensor configurations using wide angle cameras. This is the case of clusters of panoramas where the distance and spatial location between projection centers are precisely known. An example of this parallelization in the acquisition is the camera proposed by Baker et al. [86] using four panoramic sources and the sequential approach proposed by Vanijja et al. [14] using the same square pattern. Similar parallelization can be applied to the panoramic triad cluster [2]. The main cons of using panoramic sources to extract wide angle stereoscopic views of the scene is the natural self occlusion between panoramas in parallel acquisition approach that leads to large data overhead.

4.5 Omnistereo based on multi-camera configurations

This is an interesting approach to sample $\mathbf{C_S}$ in dynamic scenarios since it enable the simultaneous acquisition of visual information for all gazing directions $\mathscr{D}$ at once. The captured set of omnistereoscopic images $(\mathbf{im}_{L,i}, \mathbf{im}_{R,i})$, where $i \in \{0, ..., N-1\}$, can be used to estimate the panoramic image $\mathscr{I}_S$ at the acquisition point $\mathbf{r}$ and for any gazing direction $\mathscr{D} = (\theta, \phi)$ within the camera FOV.

An example of multi-camera acquisition of $\mathbf{C_S}$ can be seen in Fig. 4.18. This camera concept [19] is based on three fish-eye lenses mounted in a triangular pattern, each of which captures a partially overlapped aspherical field-of-view of the scene. This acquisition configuration is similar to the triangular cluster of panoramic cameras [34] presented in Appendix B. A rendering strategy can map $\mathbf{C_L}$ and $\mathbf{C_R}$ samples from individual spherical images to estimate $\mathscr{I}_S$. More interestingly, since this camera capture the whole scene at once, it is suitable for dynamic scenes. More details regarding this acquisition configuration are given at the end of this section.

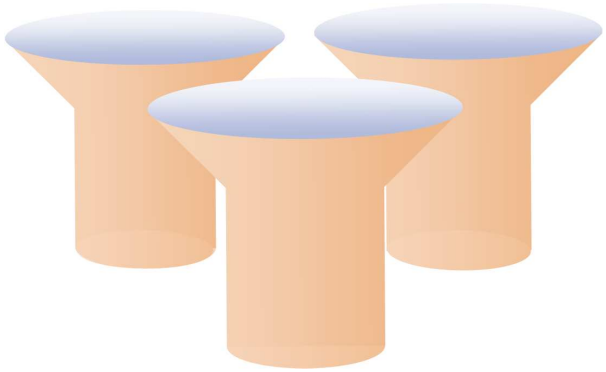


Figure 4.18: Multi-camera configuration to capture partially overlapped stereoscopic snapshots.

4.5.1 Omnistereo using multi-camera configurations: antecedents

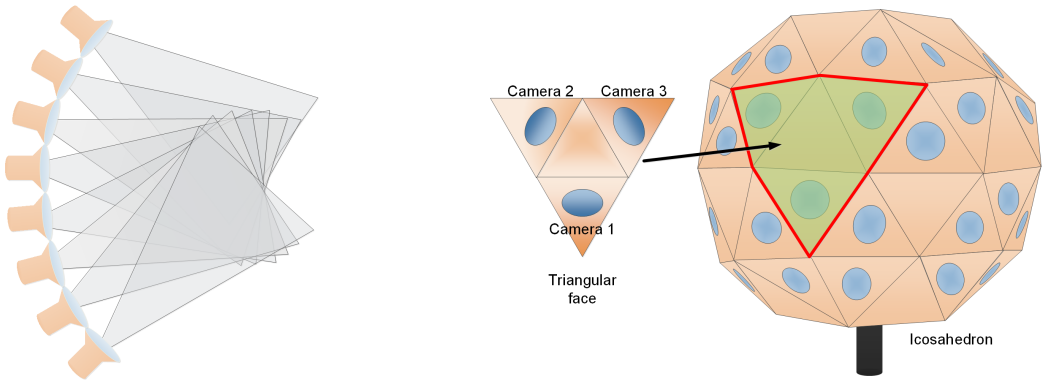
The earlier antecedents of using multiple cameras to compose a wide-angle stereoscopic image date back to 1965, when Clay was granted a patent [37] for a camera designed to capture wide-angle stereoscopic images. Although this camera was not conceived to produce omnistereoscopic imagery, it is one of the earliest antecedents of panoramic (in the wide angle sense)

stereoscopic photography to our best knowledge. The device illustrated in Fig. 4.19-(a) is made of a multitude of cameras. The different optical axes converge in the direction of the object or region-of-interest (ROI) to photograph; each camera captures the image from a different viewpoint. Each camera, paired with its immediate neighbor, constitutes a stereoscopic pair. This can be taken as a precursor of a parallelization of the *configuration 4* presented in Section 3.2.

The downfall of this camera is that it relies on a multiplicity of narrow angle lenses, which creates stereoscopic pairs that can only capture stereoscopic objects when they are located far from the camera. In other words, this camera acquires stereoscopic views of scenes where the distance from the camera is so large that human binocular perception is just marginal. Conversely, the foreground scene can be captured by individual cameras only, hence there is an important *blind* stereoscopic region. In its original conception, this idea used off-the-shelf film cameras, but a similar idea has been reused to capture multiple overlapping sections of the image from different viewpoints using lenticular arrays in front of each individual camera [54, 83, 64].

An approach for constructing a panoramic depth map using multiple cameras distributed over an icosahedric structure (Fig. 4.19-(b)) was proposed by Shimada et al. [103] and by Tanahashi et al. [104, 105]. Each face of the icosahedron houses three cameras, determining an arrangement of 60 cameras, more specifically, 20 sets of three cameras. The authors proposed to use the stereo images from three different stereo pairs per face, e.g., grouping the three cameras in three groups of two, to create disparity maps in three spatial directions. The beauty of this idea is that only one camera per face is used to compose a spherical panorama, while each face's estimated depth map is registered on the final spherical image. The authors of this camera configuration proposed this idea to detect movements in every direction; however, the concept of independently creating a correct panorama and a panoramic depth map can be exploited to create a synthetic omnistereoscopic image [18]. The geometric distribution of cameras makes this configuration attractive to render stereoscopic panoramas in spherical format, unlike the majority of the omnistereoscopic acquisition methods that focus on cylindrical topologies. The problem of acquiring spherical stereoscopic views of dynamic scenes is still open to further research. Although this camera covers a wide-angle FOV in elevation, it can still be modeled by the *configuration 4* where the stereoscopic pairs are at a regular radial distance r_c from the spherical symmetry center.

Firoozfam et al. presented a camera capable to produce omnistereoscopic images by mosaicking six stereoscopic snapshots of the scene [106]. The authors of this work proposed



(a) (b)

Figure 4.19: Examples of omnistereoscopic cameras based on multiple sensors: (a) in an early patent from 1965, Clay exploited the overlapping FOVs between cameras with slightly different viewpoints to produce stereopsis, and (b) Shimada and Tanahashi's multiple camera configuration designed to produce omnidirectional depth maps in real time.

to add omnidirectional depth estimation capabilities to their previous panoramic camera design [107]. To do so, a configuration based on six stereoscopic camera rigs in a star-like hexagonal pattern was used. This camera corresponds to the *configuration 4* in Fig. 3.2-(c), where stereoscopic rigs are located radially in six equality spaced θ angles. An illustration of their camera is shown in Fig. 4.20-(a). A prototype of their omnistereoscopic camera, which was conceived for underwater visualization, was built circa 2002, and even calibration of the stereoscopic camera pairs were reported. Although this camera was proposed for underwater robot navigation, the possibility of real time omnistereoscopic visualization by a remotely located human operator was foreseen.

Baker et al. filed a patent application [17] in 2008 on an omnistereoscopic camera using Firoozfam et al.'s concept. More specifically, their camera was also based on acquiring six partially overlapped stereoscopic images, using a star-like configuration (Fig. 4.20-(a)). This camera can be modeled using the *configuration 4*.

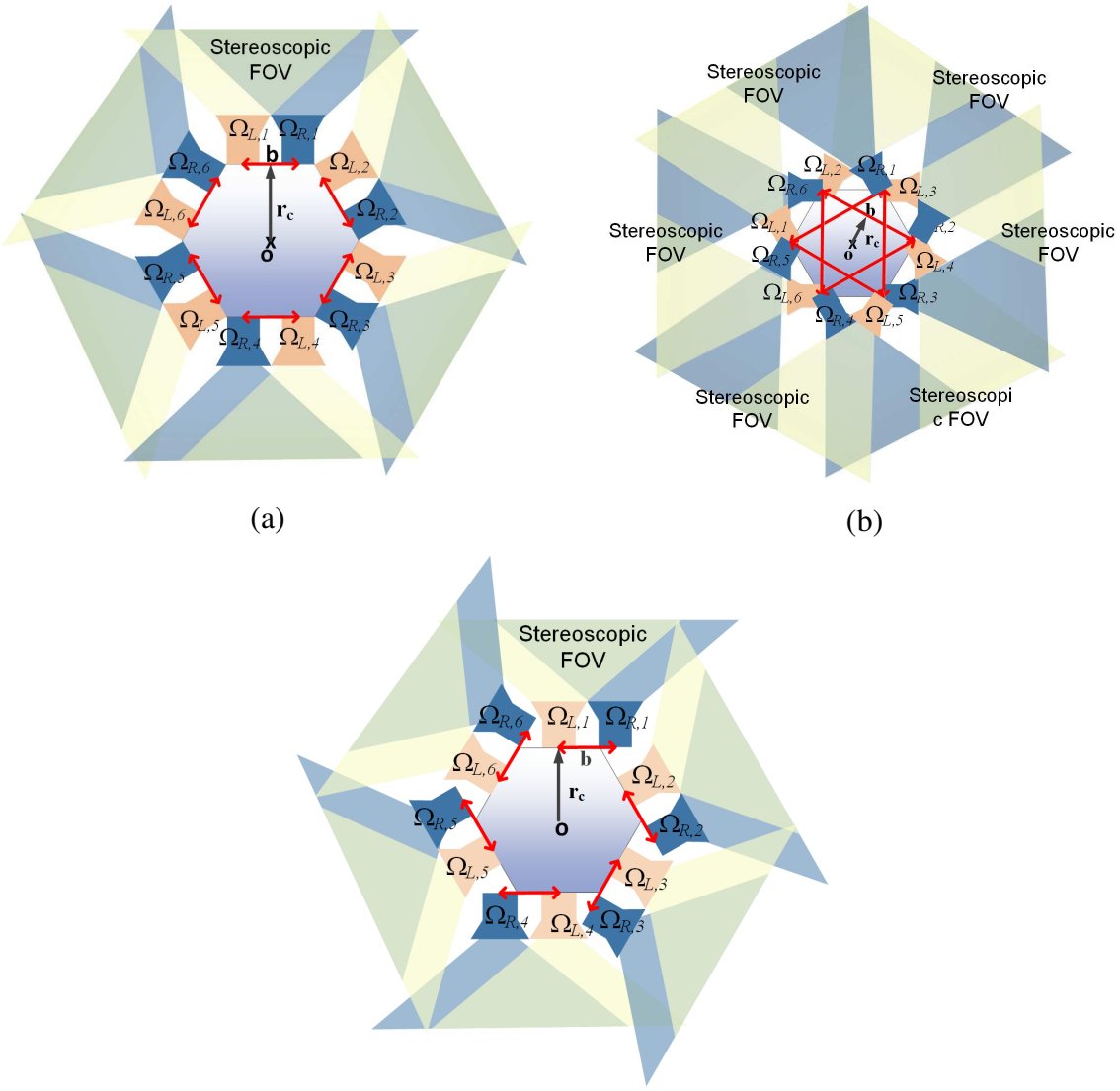


Figure 4.20: Multi-camera examples: (a) panoramic camera configuration using multiple stereoscopic pairs in an hexagonal pattern (Baker et al.), (b) different configuration using narrower FOV lenses and larger number of cameras, and (c) an alternative multi-camera configuration proposed by the authors.

In 2012, H. Baker and Constantin filed a patent application [35] on a different multi-camera configuration. An example of this camera is illustrated in Fig. 4.20-(b) for twelve cameras, although the authors suggested using a larger number of cameras. As in the cases discussed above, the idea is to acquire partially overlapped stereoscopic snapshots of the scene, which can be mosaicked to render a cylindrical omnistereoscopic image. The authors suggested using 16 to 24 individual cameras with 45° to 30° FOV in azimuth, respectively. The distance between projection centers of adjacent cameras $(\Omega_{L,i}, \Omega_{L,i+1})$ and $(\Omega_{R,i}, \Omega_{R,i+1})$, for $i \in \{0, ..., 5\}$ can be kept smaller than in the star-like hexagonal distribution illustrated in Fig. 4.20-(a). However, the price to pay for this increased proximity is using a larger number of cameras with narrower angle lenses to prevent self-occlusion. For example, using 24 cameras, the inter-camera distance is approximately $0.15\ b$, where $b = 65mm$ for a normal interocular distance, and can be smaller for hypo-stereo. The main attraction of this configuration is reducing the parallax between adjacent cameras' projection centers while maintaining a larger baseline than the configuration in Fig. 4.20-(a). However, using narrow angle lenses the minimum distance to objects in the scene is larger for stereoscopic acquisition, e.g., the distance where the scene appears in both cameras FOV of a stereoscopic rig. The same configuration has appeared in another recent patent application [108], but reducing the number of stereoscopic pairs by using wider angle lenses (Fig 4.20-(b)). This camera can be modelled as *configuration 4*.

Also using the *configuration 4*, Jiajun Zhu et al. proposed in 2007 [109] using an array of six pairs of digital single-lens reflex (DSLR) cameras in the arrangement illustrated in Fig. 4.20-(b). Their target application was 3D scene reconstruction instead of producing omnistereoscopic imagery for human viewing. Despite its main application and large baseline, this camera concept can sample the $\mathbf{C_S}$ to synthesize $\mathscr{I}_S$ in certain outdoor scenarios where the distance between the scene an the camera is large. This is mainly due to the excessively large baseline distance b between projection centers.

My contribution to multi-camera configurations [18] is illustrated in Fig. 4.20-(c). This camera addresses the problem of creating a monoscopic panorama with respect to a common cylindrical projection center $\mathbf{O}$, and to acquire accessory information to render an omnistereoscopic counterpart. The stereoscopic information follows the idea behind the approaches of H. C. Huang et al. [11] and Yamada et al. [32], but using a multi-sensor configuration to satisfy the real time omnidirectional constraint of the problem. Unlike the majority of the cameras presented in this section, this multi-sensor configuration can be modelled by the *configuration 3* (Section 3.2), which in this case models acquiring six stereoscopic snapshots separated

in equal panning increments ($\Delta_\theta = 60°$). The use of wide-angle lenses helps to reduce the number of necessary stereoscopic pairs.

The omnistereoscopic rendering based on the images acquired by the camera proposed in Fig. 4.20-(c) can be done by mosaicking stereoscopic snapshots or by synthesizing a stereoscopic view in every direction based on the *central* panorama and the horizontal disparity and occlusion maps extracted from each stereoscopic image. The *central* panorama is always rendered by mosaicking images originating from cameras $\Omega_{L,i}$ ($i \in \{0...5\}$). The mosaicking of the images originating from cameras $\Omega_{R,i}$ produces a right-eye omnidirectional view of the scene, but only when the radial distance r_c and baseline b, which in this case are equivalent ($r_c \simeq b$), are small ($b \leq 3.5$ cm). This is done to prevent excessive ghosting while mosaicking the right-eye panorama. The mosaicking is a low complexity approach suitable for real time acquisition and rendering. The second approach involves using a larger baseline, which also implies a larger $\mathbf{r_c}$. This configuration helps to estimate horizontal depth maps and occlusion maps from each stereoscopic pair. However, the parallax between projection centers is larger and so are the stitching artifacts while mosaicking the *central* monoscopic view, although there are techniques to reduce these artifacts [18]. The second rendering alternative has advantages in terms of controlling the stereoscopic budget of the scene and the visual comfort in postprocessing that make it attractive.

Tzavidas et al. [15] were one of the first in attempting to parallelize the *acquisition method 4* using a multi-camera approach. Their idea is based on using a large number of radially distributed cameras, as is depicted in Fig. 4.21-(a). They proposed to use a minimum of $N = 24$ cameras to capture N snapshots of the scene. Similar to Ishiguro and Peleg's method described in Section 4.3, Tzavidas et al.'s idea is based on extracting a pair of image columns $(\mathbf{im}_{L,i}, \mathbf{im}_{R,i})$ from each image. The concept is illustrated in Fig. 4.10-(b). These image columns are defined by the back-projection of the scene from each camera's own projection center $\mathbf{O}$, onto two new projection centers $(\mathbf{O_L}, \mathbf{O_R})$, for the left- and right-eye viewpoints with horizontal parallax b. The idea is to mosaic the N columns of $\mathbf{im}_{L,i}$ to render a cylindrical left-eye panorama $\mathbf{I_L}$, and following the same procedure, mosaic the N columns of $\mathbf{im_R}$ to generate the right-eye panorama $\mathbf{I_R}$. Since each individual image column must be larger than one pixel, noticeable artifacts appear in the panorama after mosaicking. However, the larger the number of cameras N, the narrower the image column width and the smoother $(\mathbf{I_L}, \mathbf{I_R})$ would be. On the same idea, Peer and Solina [110] estimated that a cylindrical stereoscopic panorama of 1680 pixels wide can be rendered by mosaicking image columns 14 pixels wide extracted

from $N = 120$ cameras in a circular radial configuration. The latter estimation is perhaps too optimistic to produce good results. The large number of cameras that may be necessary to produce an acceptable omnistereoscopic image by simply mosaicking stereoscopic is the main drawback of this approach.

There have been different multi-camera configurations based on distributing a multiplicity of cameras on 3D surfaces as with Shimada and Tanahashi's icosahedric camera. One of them using a semi-spherical surface housing 104 evenly distributed cameras has been built at the École Polytechnique Fédérale de Lausanne (EPFL) [111, 112]. The hemispherical camera, which resembles a multifaceted *insect-eye*, is illustrated in Fig. 4.21-(b). It was conceived to produce polycentric spherical panoramas by mosaicking a large number of snapshots. Furthermore, the developers of this camera propose to estimate the 3D structure of the scene using stereo images extracted from pairs of cameras in selected directions. The fact that the optical axes of adjacent cameras are divergent is not an obstacle to reconstruct a 3D depth map, although it poses a problem to compose stereoscopic images for human viewing. In a recent development [111, 113], this camera was used to reconstruct stereoscopic panoramas in cylindrical format. However, it is necessary an analysis of the omnidirectional depth distortions after mosaicking multiple stereoscopic images captured by cameras pairs with divergent optical axes.

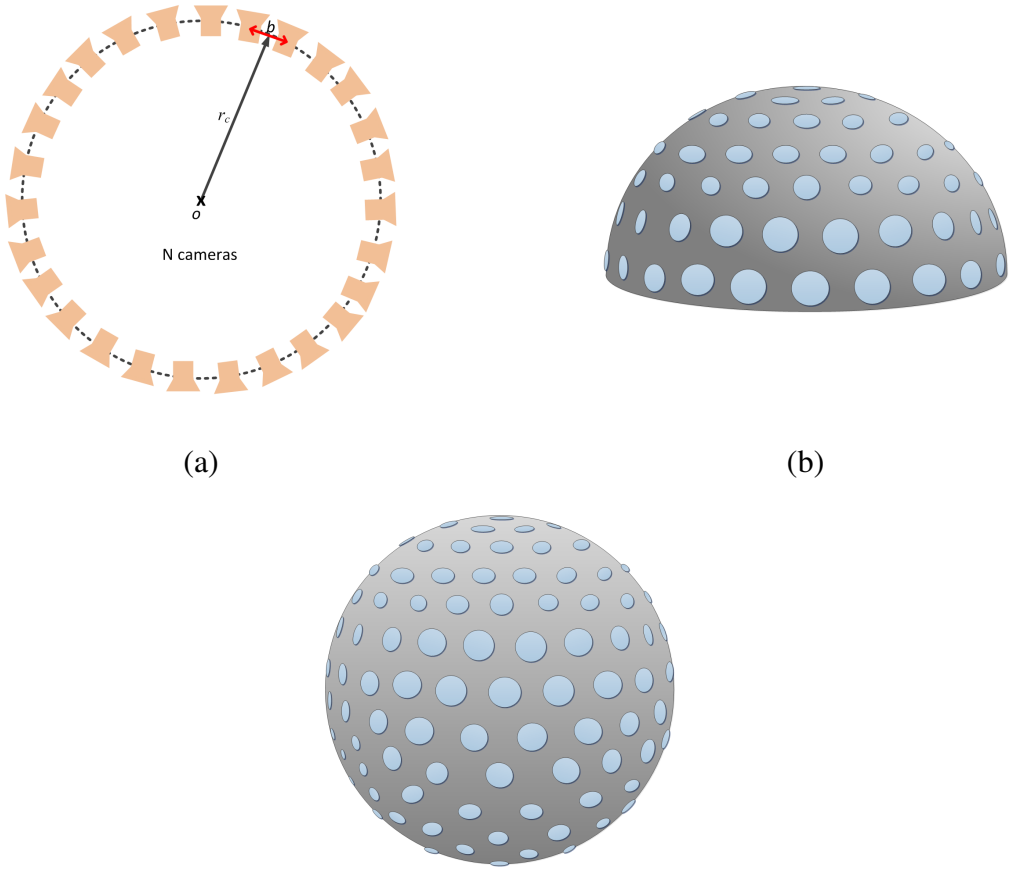


Figure 4.21: Multi-camera configurations: (a) the cameras configuration proposed by Tazvidas et al. in 2002, (b) the omnistereoscopic camera concept developed at the École Polytechnique Fédérale de Lausanne (EPFL), and (c) spherical configuration proposed by Pierce et al. in a patent application.

Similarly to the *insect-eye* camera, Pierce et al. proposed an omnistereoscopic camera based on a spherical surface covered with a large number of narrow angle cameras [114]. The idea is to compose an omnistereoscopic image by mosaicking multiple adjacent images where one camera corresponds to the left-eye view while the immediate horizontal neighbor corresponds to the right-eye view. Following this reasoning, one subset of cameras are mosaicked to form a left-eye spherical panorama while the images originating from the immediate (horizontal) neighbor are mosaicked to form the right-eye panoramic view. The fact that the adjacent cameras taken in pairs have divergent optical axes, instead of parallel or convergent optical axes, seems to be ignored by the authors. There is neither a binocular model of the image formation nor rendering examples to support the authors' claims.

In an attempt to reduce the number of cameras involved to produce omnistereoscopic imagery, Grover patented a multiple camera system using a lenticular array [64], which is based on similar principles to Peleg's spiral lens configuration [54]. Grover's optical approach is based on sampling interleaved image blocks corresponding to left- and right-eye views over a single two-dimensional (2D) planar sensor instead of capturing vertical columns corresponding to left and right-eye views. Using high-resolution sensors and a high-density lenticular array, it would be in theory possible to de-interleave left- and right-eye views from each camera, mosaicking a complete omnistereoscopic image. One of the main problems with this configuration is the complexity of building and calibrating the lenticular arrays. Another major problem is the fixed stereoscopic base-line of this camera, which can lead to undesirable effects in the stereoscopic composition. Furthermore, the number of cameras necessary given a particular lenticular array and the problems of mosaicking these segmented stereoscopic views have not been determined, nor has there been made available omnistereoscopic imagery as proof of concept of this device.

In a recent proposal, Weijia Fen at al. [115], a camera composed of four fisheye lenses in a star-like configuration has been suggested to produce omnistereoscopic imagery. This camera is illustrated in Fig. 4.22 showing four cameras radially distributed around a center $\mathbf{O}$. The designers of this camera propose to use the overlapping areas between adjacent images to extract stereoscopic views of the scene. Each image requires 180° FOV in azimuth or more, so each of them overlaps over at least 90° FOV in azimuth with the adjacent image. The authors of this conceptual camera propose to mosaic these overlapped stereo views, after correcting each lens distortion, to create a full stereoscopic spherical panorama. This is a work in progress at the moment of writing this chapter, hence there are still many issues to be addressed. For instance,

the modeling of stereoscopic views for the poles of the spherical view has not be contemplated. In general, a model for the 3D human binocular viewing is necessary to explain this camera configuration as well as other configurations reviewed herein. Additionally, the disparity distortion due to using fisheye lenses when trying to maintain a consistent and continuous depth perception in every direction has not been addressed.

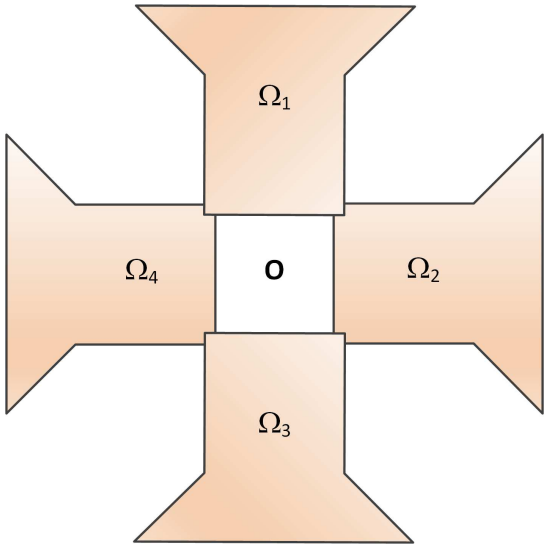


Figure 4.22: Multi-camera configuration to capture partially overlapped stereoscopic snapshots.

A recent development at the Université of Montréal [116] is a novel example of a promising configuration to produce real-time omnistereoscopic imagery [19]. Their *omnipolar* camera consists of three vertically oriented fish-eye lenses to produce a triad of partially overlapped stereoscopic views of the scene, each of them with horizontal parallax (Fig. 4.18). This multiple camera configuration is similar than the panoramic triad method [2] (Section 4.1.1). The similarities reside in the triangular distribution of wide-angle cameras to capture simultaneously three partially overlapped samples of the omnidirectional function $\mathbf{C}$ at three distinct viewpoints. But the similarities ends there. This camera has a major advantage over previous sequential methods [117]: the real-time acquisition of dynamic scenes. However, this acquisition approach requires to define a proper warping strategy to correct lens distortions and a suitable mosaicking algorithm based on an optimal cut stitching [118].

4.5.2 Pros and cons of multi-camera configurations

The parallelization of *configuration 4* [15] is an interesting option, especially given the trend in miniaturization of high-quality cameras. The reason for being skeptical about this approach is simple: the narrower the images's vertical stripes used to extract two parallax views of scene, the better the quality of the mosaicking, but the necessary number of cameras to be able to extract narrow image columns is too large to be practical. However, a more sophisticated rendering approach based on light field rendering [119] can use this samples of $\mathbf{C_L}$ and $\mathbf{C_R}$ at close but continuous gazing directions $\mathscr{D}$ to approximate an artifact-free $\mathbf{I_S}$.

A radial arrangement to capture partially overlapped stereoscopic images using reasonably large baselines implies a careful design of the camera geometry. Multiple camera configurations using this strategy have been proposed recently [17, 86, 64, 35, 18]. However promising, there are still unsolved issues regarding the cameras' parallax effect in the rendering of intermediate views between samples, e.g., stitching panning directions, maintaining the disparity continuity. Moreover, the effect of the geometric distribution of cameras on the minimum distance to objects on the scene for a correct stereoscopic viewing has not been properly studied.

Another point to be considered when mosaicking wide-angle stereoscopic samples is the distortion incurred in creating a cylindrical stereoscopic panorama by mosaicking a limited number of partially overlapped stereoscopic snapshots.

4.6 Summary

This chapter has presented a comprehensive review of methods to acquire the necessary visual information from a scene to render stereoscopic omnidirectional images, most in the form of cylindrical panoramas. The different proposals for omnistereoscopic acquisition have been reviewed with an emphasis on their direct or indirect potential to render stereoscopic snapshots of dynamic scenes in any panning direction. Our review rests on two separate classifications of the proposed methods. The first is a classification into four acquisition models according to the different geometric configurations used to acquire stereoscopic pairs in all directions. The second is a classification into four families based on the camera construction and/or the acquisition strategy. The chapter contains four main sections reviewing proposals from the technical and patent literature belonging to these four families, identifying the acquisition model corresponding to each proposed camera configuration and discussing the relative strengths and weaknesses of each family.

Single-view point catadioptric cameras are the most attractive for their reduced hardware, using a single image sensor, but the omnistereoscopic alternatives based on vertical co-axial configurations are not suitable to produce horizontal stereoscopic views of the scene. The omnistereoscopic catadioptric cameras are suitable to acquire panoramic depth maps, but their incapability to handle occlusions render the information acquired by these cameras incomplete, at least in the context of the problem investigated.

Sequential acquisition methods were reviewed because of their potential to be adapted for a real time omnistereoscopic sampling of dynamic scenes. Following this idea, there were interesting proposals of using multiple cameras to extract left and right-eye views in every direction simultaneously. These methods generally underestimate the number of cameras necessary to produce artifact free omnistereoscopic pairs. A simpler and more efficient approach is sampling the scene with a limited number of partially overlapped stereoscopic snapshots, which is the concept behind multiple camera configurations.

Before discussing the multiple camera methods, another approach has been presented: using multiple panoramic snapshots to render omnistereoscopic images. This approach has also intrinsic problems to render omnistereoscopic views of dynamic scenes since it is based on the sequential acquisition of panoramic snapshots. The idea is to use stereoscopic wide-angle views of the scene obtained from different panoramas to create stereoscopic views in arbitrary gazing directions. When the panoramas are acquired in a known sampling pattern, the consistency in the perceived depth can be guaranteed. Since multiple wide-angle snapshots of

the scene can be strategically acquired to cover all looking directions, the panoramic cluster approach can be replaced by a set of multiple stereoscopic camera pairs strategically oriented to cover the whole scene.

Multiple camera configurations are probably the best candidates in the context of the proposed problem. There are many variants of the same idea, but all of them are based on the simultaneous acquisition of partially overlapped set of stereoscopic snapshots of the scene. The geometric distribution between cameras varies between proposals, and there is no theoretical framework to contrast them. The lack of a formal analysis also extends to the omnistereoscopic rendering methods, which are intrinsically linked to the multiple camera geometry. Furthermore, a model representing the binocular visual field derived from the plenoptic function model will be useful to contrast the different acquisition strategies and rendering results.

The conclusions of this background research in terms of the potential of each panoramic camera to acquire dynamic scenes omnistereoscopically is summarized in Table 4.1.

Table 4.1: Comparison between different omnistereoscopic methods.

Configuration	Pros	Cons
Catadioptric	Simple hardware configuration and, when hyperbolic and parabolic mirrors are used, fits the SVP model, avoiding mosaicking problems due to multiple projection centers. In other non-SVP configurations, it can acquire horizontal stereo in multiple panning directions using lenticular optics [63] or multiple stereoscopic rigs [51].	In vertical coaxial configuration it can only capture depth based on the vertical disparity map; in this mode it cannot handle occlusions. Lens radial distortion may leads to out-of-focus blurring.
Sequential acquisition	Not suitable for stereoscopic acquisition of dynamic events. However, some configurations can be used in parallel acquisition configuration [11, 21, 13].	Their lengthy acquisition time make these methods unsuitable to sample dynamic scenes.
Using panoramic sources	The idea of using wide-angle snapshots of the scene at controlled locations can be used in a multiple camera configuration [2].	Sequential acquisition of panoramas make these methods unsuitable for sampling dynamic scenes stereoscopically and omnidirectionally.
Using multiple sensors	A solution for the problem can be implemented using off-the-shelf cameras in different geometric configurations as long as they acquire partially overlapped stereoscopic snapshots covering the whole scene [17, 35, 103, 64].	Parallax between cameras may introduce artifacts in the images and limits their capability to render objects in stereo to a minimum distance from the camera. The geometric distribution of the camera determines the rendering strategy, which needs further research.

Chapter 5

Disparities and rendering approaches

This chapter introduces the different rendering methods that use the visual information acquired by any of the configurations presented in Section 3.2. Although this thesis does not focus on the rendering of omnistereoscopic imagery, the sampling of $\mathbf{C_S}$ is entangled with the usage of this information to estimate the value of $\mathscr{I}_S$ and to its sampling to produce a discrete omnistereoscopic image $\mathbf{I_S}$.

First, I define the disparity between corresponding coordinates of the same point $\mathbf{P_W}$ after being projected onto images $\mathbf{im}_{j,i}$. This is important in order to define the limits of horizontal disparity resolution in humans and how this limit can be exploited to determine the continuity in the depth perceived after mosaicking stereoscopic images.

The methods to generate $\mathbf{I_S}$ are presented in reverse order of feasibility and complexity in the context of this problem, which is the rendering omnistereoscopic imagery of dynamic scenes. The first method is based on mosaicking a large number of image columns acquired sequentially. The second rendering method is based on capturing the whole scene at once using a reduced number of stereoscopic snapshots and extra information about the scene 3D structure. The last rendering approach is based on the pure mosaicking of a limited number of stereoscopic snapshots to create $\mathbf{I_S}$, without the need for extra depth information of the scene. The acquisition constraints originated in the mosaicking approach for rendering are studied extensively in this thesis.

An example of scene navigation based on projective transformations of a planar stereoscopic image extracted from $\mathbf{I_S}$ is presented and the end of this chapter. Additionally, the estimation of the number of necessary images to produce $\mathbf{I_S}$ by mosaicking is presented at the end of this chapter along with two models for the 3D scene.

5.1 Disparities

A point in the scene $\mathbf{P_W}$, which is in the FOV of cameras $\Omega_{j,i}$, is projected onto $\mathbf{im}_{j,i}$. The coordinates of these corresponding projections are $\mathbf{p}_{j,i} = (x_{j,i}, y_{j,i})^T$. The set of coordinates on each image plane defines a *corresponding pair*, which will be referred to

$$\mathscr{K}_i = (\mathbf{p}_{L,i}, \mathbf{p}_{R,i}), \tag{5.1}$$

where the index i denotes the θ_i and $j = \{L, R\}$. The vector defined by a corresponding pair $\mathscr{K}_i$ is denoted

$$\mathbf{d}_i(\mathscr{K}_i) = \mathbf{p}_{R,i} - \mathbf{p}_{L,i}. \tag{5.2}$$

The components of $\mathbf{d}_i$ in the horizontal and vertical dimensions are defined as the *horizontal disparity* $(d_{h,i})$ and *vertical disparity* $(d_{v,i})$, respectively. The correspondent pair of projections for a point $\mathbf{P_W}$ and θ_i is illustrated in Fig. 3.1.

5.1.1 Disparity functions in a planar stereoscopic image

The disparity can be defined as functions $d_{h,i}, d_{v,i} : \mathbb{R}^2 \rightarrow \mathbb{R}$, which can be defined for $\mathscr{K}_i$ as follows

$$d_{h,i}(\mathbf{K}_i) = x_{R,i} - x_{L,i}, \tag{5.3}$$

$$d_{v,i}(\mathbf{K}_i) = y_{R,i} - y_{L,i}. \tag{5.4}$$

A graphical depiction of the disparities is shown in Fig. 5.1-(a).

A point $\mathbf{P_W}$ in the FOV of a stereoscopic pair of cameras oriented at θ_i defines the corresponding set $\mathscr{K}_i$. Consequently, the set of all $\mathscr{K}_i$ for a given stereoscopic image $\mathbf{im}_{j,i}$ defines a *motion field* [120].

5.1.2 Angular disparity

The projection of $\mathbf{im}_{j,i}$ onto a canvas can be a done onto a planar or warped surface. An alternative definition for the disparities is based on the *angular spread* between corresponding points $\mathscr{K}_i$ with respect to $\mathbf{O}$ reference viewpoint. These angles are defined as follows

$$\alpha_{h,i}(\mathscr{K}_i) = 2 \arctan\left(\frac{d_{h,i}(\mathscr{K}_i)}{2\,f}\right), \tag{5.5}$$

$$\alpha_{v,i}(\mathscr{K}_i) = 2 \arctan\left(\frac{d_{v,i}(\mathscr{K}_i)}{2\,f}\right). \tag{5.6}$$

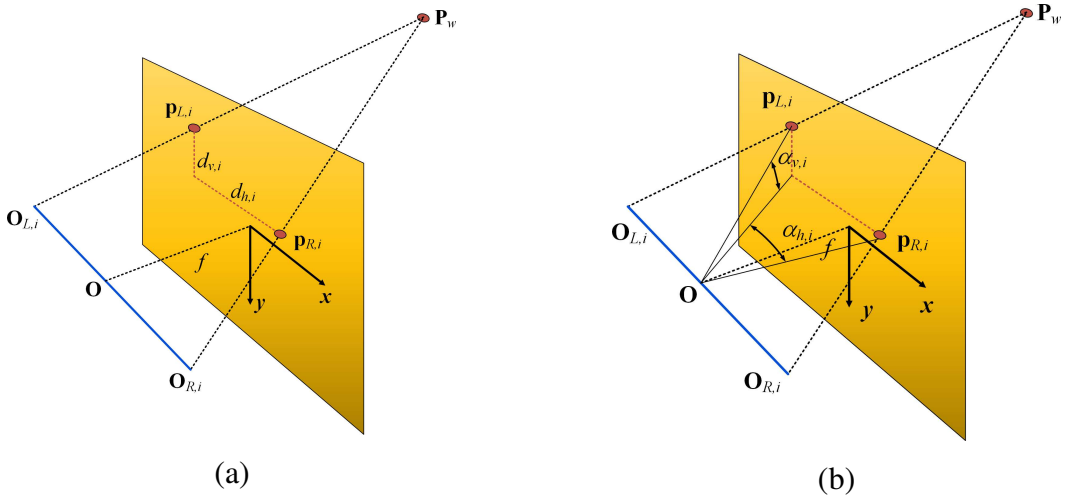


Figure 5.1: Disparities: (a) as the distance between right and left coordinates of $\mathscr{K}_i$ in a planar image and (b) as the angular spread between $\mathscr{K}_i$ with respect to the reference viewpoint $\mathbf{O}$.

The angular disparities are usually expressed in *minutes of arc* (arcmin) or *second of arc* (arcsec) units. The angular disparities in the horizontal or *azimuth* ($\alpha_{h,i}$) and vertical or *elevation* ($\alpha_{v,i}$) directions are illustrated with an example in Fig. 5.1-(b).

The advantage of using this angular representation is to provide a consistent measure of the disparity between corresponding points independent of the display. For instance, after warping and projecting a stereoscopic image on a curved or planar display, the disparity between corresponding points on a planar canvas will vary depending the distance from $\mathbf{O}$, but $\alpha_{h,i}$ and $\alpha_{v,i}$ will remain the same. The idea is illustrated with a simple example in Fig. 5.2 where points $\mathbf{P_W}$ and $\mathbf{P'_W}$, both located at the same distance z_i in $\mathbf{Z_i}$-axis, defines corresponding pairs $\mathscr{K}_i$ and $\mathscr{K}_i'$, respectively. The horizontal disparities in the planar stereoscopic image $\mathbf{im}_{j,i}$ must satisfy that

$$d_{h,i}(\mathscr{K}_i) = d_{h,i}(\mathscr{K}_i').$$

The projection of the same corresponding sets of points on a curved canvas leads to angular disparities

$$\alpha_{h,i}(\mathscr{K}_i) = \alpha'_{h,i}(\mathscr{K}_i'),$$

which conveys the same depth by binocular cues to the user. Therefore, both representations are equivalent.

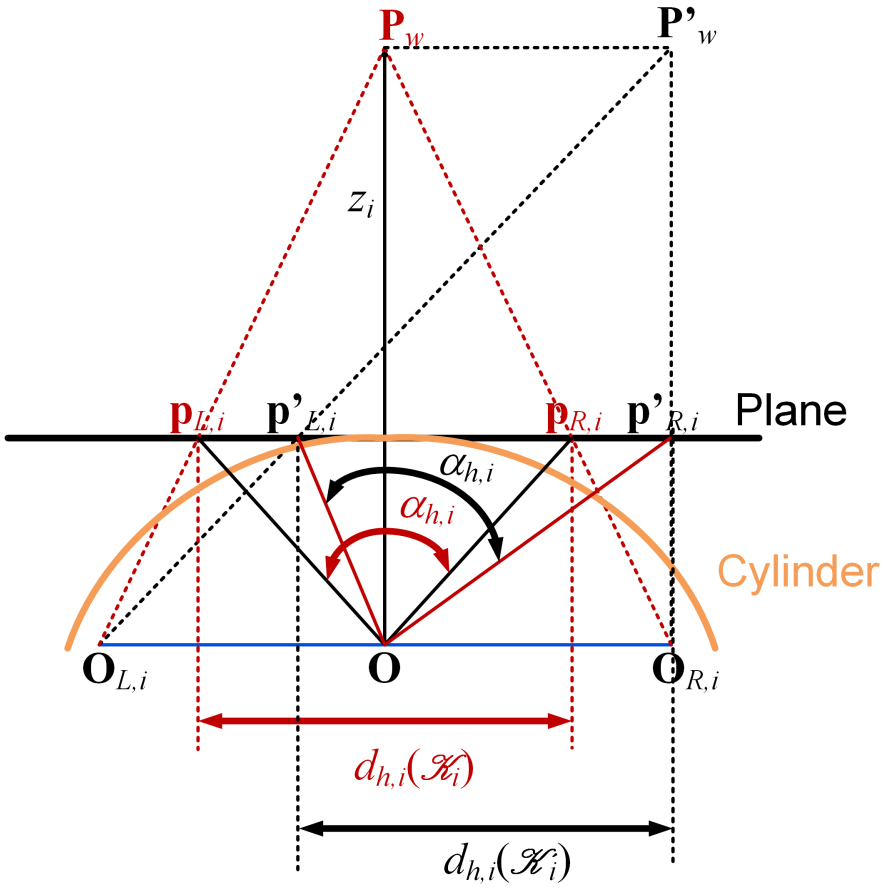


Figure 5.2: Equivalence between disparities of $\mathscr{K}_i$ and $\mathscr{K}'_i$ on a planar and curved canvas.

5.1.3 Depth estimation

The depth of a scene element $\mathbf{P_W}$ is defined by the distance in $\mathbf{Z}$-axis. It can be estimated from $d_{h,i}$ using the geometric constraints for the stereoscopic camera pair defined in Section 3.1.2. For instance, the depth of $\mathbf{P_W}$ is $Z_{L,i}$ for camera $\Omega_{L,i}$ and $Z_{R,i}$ for camera $\Omega_{R,i}$. The parallel optical axes constraint makes $Z_{L,i} = Z_{R_i} = Z_i$, which can be determined by

$$Z_i(\mathscr{K}_i) = \frac{f\ b}{d_{h,i}(\mathscr{K}_i)}, \tag{5.7}$$

where Z_i is the projection of $\mathbf{P_W}$ in an axis perpendicular to $\mathbf{b}$, on the $\mathbf{XZ}$-plane, with origin at $\mathbf{O}$. Hence, the distance with respect to the $\mathbf{XYZ}$ system of coordinates is

$$Z(\mathscr{K}_i) = Z_i(\mathscr{K}_i)\cos\theta_i, \tag{5.8}$$

as illustrated in Fig. 5.3.

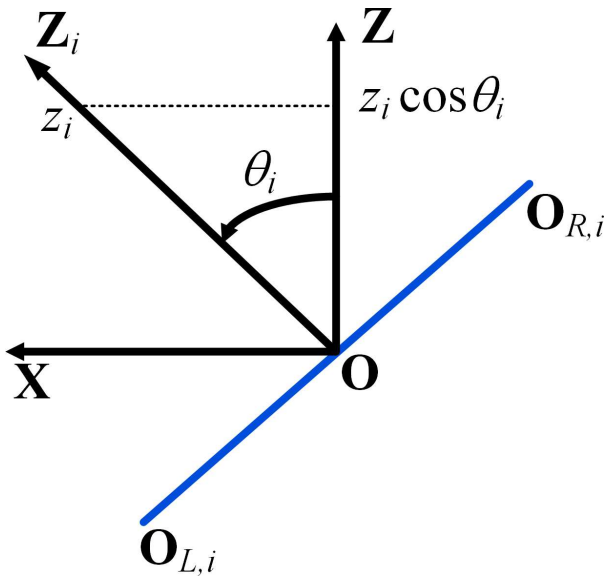


Figure 5.3: Depth of the scene.

The availability of $(d_{h,i}, d_{v,i})$ for all $\mathbf{K}_i$ defined in $\mathbf{I_S}$ enables the definition of an omni-directional depth map $\mathscr{M} : \mathbb{R}^2 \rightarrow \mathbb{R}$, where $\mathscr{M}(\mathbf{p}_{j,i}) = Z(\mathbf{p}_{j,i})$ the depth of the panoramic scene. Conversely, the availability of the scene 3D information contained in $\mathscr{M}$ helps to obtain an accurate estimation of $\mathscr{I}_S$ based on a number of stereoscopic samples of the scene.

5.1.4 Disparity difference function

The disparity difference between any pair of corresponding sets, $(\mathscr{K}_n, \mathscr{K}'_m)$, can be used to determine how the disparity changes over one image or a pair of images. Based on this notation, when $n = m = i$, the comparison is between two corresponding points in $\mathbf{im}_{j,i}$. But, when $n = i$ and $m = i+1$, the comparison is between corresponding sets on neighbour images $\mathbf{im}_{j,i}$ and $\mathbf{im}_{j,i+1}$, respectively. Note that the corresponding sets using in the comparison of disparities does not necessarily are from the same $\mathbf{P_W}$ in the scene.

The *disparity difference functions* can be defined as follows

$$\mathscr{E}_{h,(n,m)}(\mathscr{K}_n, \mathscr{K}'_m) = d_{h,n}(\mathscr{K}_n) - d_{h,m}(\mathscr{K}'_m), \tag{5.9}$$

$$\mathscr{E}_{v,(n,m)}(\mathscr{K}_n, \mathscr{K}'_m) = d_{v,n}(\mathscr{K}_n) - d_{v,m}(\mathscr{K}'_m), \tag{5.10}$$

where $n, m \in \{0, ..., N-1\}$, and h and v indicates horizontal and vertical, respectively. As defined, the disparity difference functions are $\mathscr{E}_{h,(n,m)}, \mathscr{E}_{v,(n,m)} : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$.

The disparity difference can be also expressed as the difference between the angular disparities as follows

$$\mathscr{A}_{h,(n,m)}\left(\mathscr{K}_n, \mathscr{K}'_m\right) = \alpha_{h,n}(\mathscr{K}_n) - \alpha_{h,m}(\mathscr{K}'_m), \tag{5.11}$$

$$\mathscr{A}_{v,(n,m)}\left(\mathscr{K}_n, \mathscr{K}'_m\right) = \alpha_{v,n}(\mathscr{K}_n) - \alpha_{v,m}(\mathscr{K}'_m). \tag{5.12}$$

In the context of this thesis, there are two particular problems where the disparity difference functions can be applied:

- the difference in horizontal disparity within the same stereoscopic image, that is for $n = m = i$, and
- the difference in horizontal disparity between adjacent stereoscopic images, when the projections of the same $\mathbf{P_W}$ in $\mathbf{im}_{j,i}$ is matched (aligned) with its projection in $\mathbf{im}_{j,i+1}$.

The first case is relevant to compare the depth rendition when two different points in the scene, $\mathbf{P_W}$ and $\mathbf{P'_W}$, are projected in the same stereoscopic image $\mathbf{im}_{j,i}$ with corresponding point sets $\mathscr{K}_i$ and $\mathscr{K}_i'$, respectively. In this case, Eq. 5.9 and 5.10 can be denoted as follows

$$\mathscr{E}_{h,i}\left(\mathscr{K}_i, \mathscr{K}_i'\right) = d_{h,i}(\mathscr{K}_i) - d_{h,i}(\mathscr{K'}_i), \tag{5.13}$$

$$\mathscr{E}_{v,i}\left(\mathscr{K}_i, \mathscr{K}_i'\right) = d_{v,i}(\mathscr{K}_i) - d_{v,i}(\mathscr{K'}_i). \tag{5.14}$$

This situation is used in Section 5.2.3 to estimate the usable image width or the number of necessary stereoscopic samples to be acquired based on the practical limits of the human depth resolution.

The second case of interest is when the same point of the scene $\mathbf{P_W}$ is imaged by two neighbour stereoscopic images $\mathbf{im}_{j,i}$ and $\mathbf{im}_{j,i+1}$, originating corresponding points $\mathscr{K}_i$ and $\mathscr{K}'_{i+1}$, respectively. This comparison is especially relevant when one of the points in the $\mathscr{K}_i$ set is mapped (aligned) with another point in the $\mathscr{K}_{i+1}$ set. This is part of the aligning process to stich partially overlapped stereoscopic images, which is detailed in Section 6.1. In this case, the disparity difference become a *disparity error function*, which can be used to study the depth continuity in mosaicked stereoscopic images. The *disparity error functions* can be defined using Eq. 5.9 and 5.10 as follows

$$\mathscr{E}_{h,(i,i+1)}\left(\mathscr{K}_i, \mathscr{K}'_{i+1}\right) = d_{h,i}(\mathscr{K}_i) - d_{h,i+1}(\mathscr{K'}_{i+1}), \tag{5.15}$$

$$\mathscr{E}_{v,(i,i+1)}\left(\mathscr{K}_i, \mathscr{K}'_{i+1}\right) = d_{v,i}(\mathscr{K}_i) - d_{v,i+1}(\mathscr{K'}_{i+1}). \tag{5.16}$$

A graphical depiction of the disparity error functions is presented in Fig. 5.4.

5.1.5 Depth resolution

The binocular vision system in humans is capable to resolve scene elements at different depths. Determining the limits of the human depth resolution is important for the rendering of $\mathscr{I}_S$, especially when the mosaicking of stereoscopic images is involved.

Two scene elements whose respective depths are Z_i and $Z_i + \epsilon_z$ have horizontal disparities given by $d_{h,i}(\mathscr{K}_{j,i})$ and $d_{h,i}(\mathscr{K}'_{j,i})$, where $\mathscr{K}_{j,i}$ and $\mathscr{K}'_{j,i}$ are the corresponding point sets defined in $\mathbf{im}_{j,i}$ for the projections of both points in the scene. This is shown in Fig. 5.4. In this scenario, there is a minimum depth difference ϵ_z that can be resolved by a human viewer using binocular cues only.

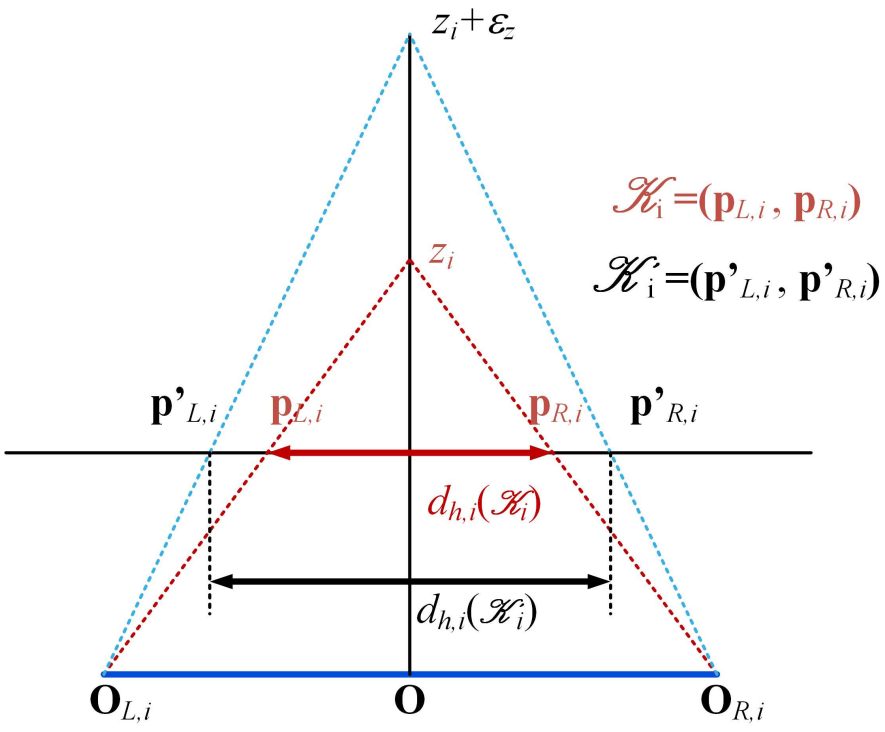


Figure 5.4: Disparity difference.

The horizontal disparity difference is defined by the absolute value of the horizontal dis-

parity function as

$$e_{h,i} = \|\mathscr{E}_{h,i}\|, \tag{5.17}$$

where $\mathscr{E}_{h,i}$ is defined in Eq. 5.9. Equivalently, the horizontal disparity difference can be defined as a function of the angular disparity as follows

$$a_{h,i} = \|\mathscr{A}_{h,i}\|. \tag{5.18}$$

This is valid for two points in the scene projected in the same stereoscopic image $\mathbf{im}_{j,i}$ as shown in Fig. 5.4.

In the case when the same scene is projected in two overlapped stereoscopic images, the horizontal disparity error is defined as follows

$$e_h = \|\mathscr{E}_{h,(i,i+1)}\|, \tag{5.19}$$

or, in terms of the angular disparities, as follows

$$a_h = \|\mathscr{A}_{h,(i,i+1)}\|. \tag{5.20}$$

In summary, e_h is referred to horizontal disparity error between projections of the same point in the scene $\mathbf{P_W}$ onto two stereoscopic images $(\mathbf{im}_{j,i}, \mathbf{im}_{j,i+1})$, while $e_{h,i}$ denotes the horizontal disparity difference between two points in the scene ($\mathbf{P_W}$,$\mathbf{P_W}'$) projected onto one stereoscopic image $\mathbf{im}_{j,i}$.

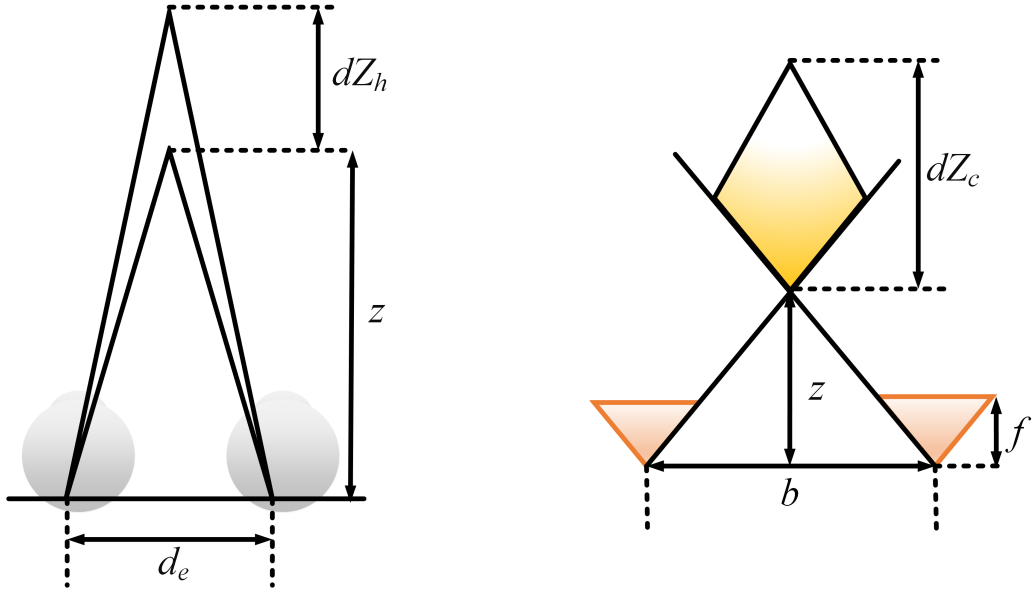
Human depth resolution

The perceptual depth resolution in the average adult population (dZ_h) can be approximated by [121]

$$dZ_h = \frac{{Z_c}^2\ \delta_\theta}{d_e} \tag{5.21}$$

where δ_θ is the *stereo or vergence acuity* in humans (defined in Appendix D), considered here as a dimensionless constant, d_e is the average interocular distance in adults (65 mm), and Z_c is the distance from the reference system defined on the stereoscopic camera. The stereo acuity in humans has been characterized as 20 arcsec by Diner in [122]; however, other authors have attributed a larger figure to this parameters, e.g., 10 arcsec [123, 124]. The stereo acuity may even vary between subjects [125]. As a reference value, I used 20 arcsec as a measure for the human stereo acuity. Note that δ_θ is equivalent to the angular horizontal disparity error $a_{h,i}$

(Eq. 5.18) and $dZ_h = \epsilon_z$ (Fig. 5.4). A diagram that helps to understand these parameters is shown in Fig. 5.5-(a). Note that various visual cues are exploited to estimate depth, but in this analysis only the horizontal disparity between corresponding image elements is considered.



(a) (b)

Figure 5.5: Depth resolution scheme: (a) stereoscopic (pinhole) camera, and (b) simplified binocular vision model.

Camera depth resolution

The depth resolution (dZ_c) of a stereoscopic camera can be approximated by [126]

$$dZ_c = \frac{{Z_c}^2 \, \varepsilon}{f \; b}. \tag{5.22}$$

In Fig. 5.5-(b) is presented an illustration that helps understand dZ_c.

The goal is to have $dZ_c \leq dZ_h$ for a camera to surpass or equal the human depth resolution. The largest threshold is defined by $dZ_c = dZ_h$, hence, from Eq. 5.21 and Eq. 5.22, the disparity threshold is

$$\varepsilon = \frac{f \; b \; \delta_\theta}{d_e}. \tag{5.23}$$

5.1.6 Practical limits of the depth resolution

The model for human depth resolution presented here is an attempt to interpret a series of physiological tests done on human subjects [121]. These tests were designed to obtain measures of the horizontal disparity resolution in subjects. Therefore, these tests attempt to remove other bias that influences the human depth resolution; one of the most important ones are the geometrical cues in the visual scene [127]. Under these circumstances, the parallel between human binocular vision and the stereoscopic pair of pinhole cameras may not be an accurate accurate way to obtain a threshold ε.

As an example, a stereoscopic camera system with the same horizontal FOV as the binocular vision system (120° or $\Delta_a = 60^\circ$), a baseline $b = d_e$ and a horizontal sensor size W_h requires a horizontal disparity resolution ε given by [128]

$$\varepsilon = \frac{W_h \, \delta_\theta}{2 \, \tan \frac{\Delta_a}{2}}. \tag{5.24}$$

The pixel width of a Nikon 800D sensor defines a minimum angular disparity of 65 arcsec, or approximately $3\delta_\theta$ for a 20 arcsec stereo acuity resolution. Hence, the achievable horizontal resolution of a stereoscopic camera using this sensor is below the depth acuity of the human vision.

A good threshold for $e_{h,i}$ is the maximum horizontal resolution of the camera, which is sufficient to guarantee depth consistency in a practical scenario [18, 2]. Hence, a threshold for e_{h_i} in a practical scenario can be defined as

$$\varepsilon = s, \tag{5.25}$$

where s is a pixel width. This threshold is not only a practical compromise due to the available technology, but also has proven to be valid in a rendering scenario based on mosaicking. In that case, it guarantees continuous disparity between seams while preventing visible ghosting after mosaicking and blending.

5.2 Omnistereoscopic rendering

The rendering methods to estimate $\mathscr{I}_S$ and to produce an omnistereoscopic image $\mathbf{I}_S$ depend on the information made available by the acquisition technique. The more information is made available during the acquisition, the better the approximation of $\mathscr{I}_S$ and the better the quality of the depth illusion achieved in any gazing direction.

In this section, I present three cases of rendering based on different acquisition models. The first method is based on the exhaustive sampling of $\mathbf{C_S}$ by sequentially acquiring stereoscopic images of the scene. The second method relies on a limited number of stereoscopic samples to estimate $\mathscr{I}_S$ when the depth of each scene element is available. The third rendering approach does not require the omnidirectional depth information and it is based on stitching and blending a reduced number of stereoscopic images of the scene. The requirements in terms of acquisition constraints to use the mosaicking method are the main focus of this thesis.

The contribution of this thesis is not focused on the rendering of omnistereoscopic imagery, although suitable rendering methods are proposed based on each acquisition technique.

5.2.1 Mosaicking a large number of narrow image columns

A traditional omnistereoscopic acquisition approach has been the exhaustive sampling of $\mathbf{C_S}$ by acquiring a large number of stereoscopic samples using a rotating stereoscopic camera. This technique has been explained in Section 4.3.1. This method uses partial information of images $\mathbf{im}_{j,i}$ to render $\mathbf{I_S}$ when pinhole cameras are used [13]. Line sensors have been alternatively used to capture a pair of narrow image columns that are mosaicked into a full $\mathbf{I_S}$ [11, 21]. An example to illustrate this omnistereoscopic method is presented in Fig. 5.6.

This rendering approach conveys a correct depth illusion to the user located at the center of projection when $\mathbf{I_S}$ is projected in cylindrical displays. However, this binocular correctness is limited to the region of interest at the center of the user's gaze. For instance, based on the model for $\mathbf{C_S}$ (Fig. 2.3.7), in any gazing directions in azimuth, the corresponding points $\mathscr{K}_i$ convey a horizontal angular disparity $\alpha_{h,i}$ that can be directly translated into a depth Z_i when the baseline distance b matches the user's interocular distance.

Line cameras and the omnistereoscopic viewing model

When considering only narrow image columns to approximate $\mathbf{I_S}$, it seems at first a downfall of the omnistereoscopic viewing function model since it does not represent what a user with nor-

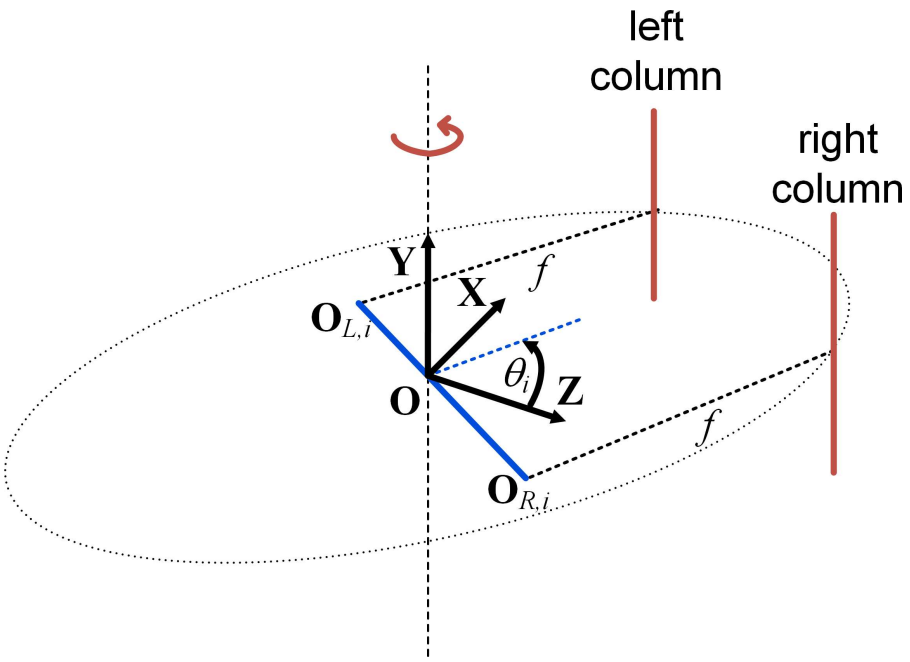


Figure 5.6: Sequential acquisition of image columns for mosaicking into $\mathbf{I_S}$.

mal binocular vision would see when gazing in $\mathscr{D}$ direction. This is true when $\mathbf{I_S}$ is the result of mosaicking image columns obtained from line sensors, but the general acquisition model presented in Chapter 3 consists of two pinhole cameras with a horizontal baseline. Therefore, a wider FOV is acquired in each sampling direction θ_i by sampling $\mathbf{C_L}$ and $\mathbf{C_R}$ for each gazing direction in azimuth θ as illustrated in Fig. 2.4.

Depth distortions

The depth distortion in the periphery of the ROI in cylindrical displays has been estimated by Couture et al. in [22] and by Bourke et al. in connection to dome-shaped displays in [58]. The authors of the former paper concluded that depth distortions are below the human stereoscopic resolution for eccentricities of about 20 arcmin off the central ROI. Although the threshold for human depth resolution the authors of this work used is more conservative that the one calculated in Section 5.1.5, their estimations is good precedent for the analysis of stereoscopic distortions on curved displays.

Although this approach is appropriate for rendering omnistereoscopic image for cylindrical displays, additional processing is necessary to create planar stereoscopic imagery in any gazing directions. More importantly, it is not suitable to render dynamic scenes.

5.2.2 A few stereoscopic samples plus depth information

The alternative to acquiring a large number of column pairs is to capture N stereoscopic samples of the scene. Novel stereoscopic views can be rendered using the projections of $\mathbf{P_W}$ into the corresponding coordinate pair $\mathscr{K}_i$, for a panning angle θ_i ($i \in \{0, ..., N-1\}$). Any of the acquisition configurations presented in Section 3.2 can be used to describe this scheme. For each corresponding pair of coordinates $\mathscr{K}_i = \mathbf{p}_{j,i}$, the depth map $\mathscr{M}$ is necessary to render $\mathbf{I_S}$. An example of this acquisition method is illustrated in Fig. 5.7-(a).

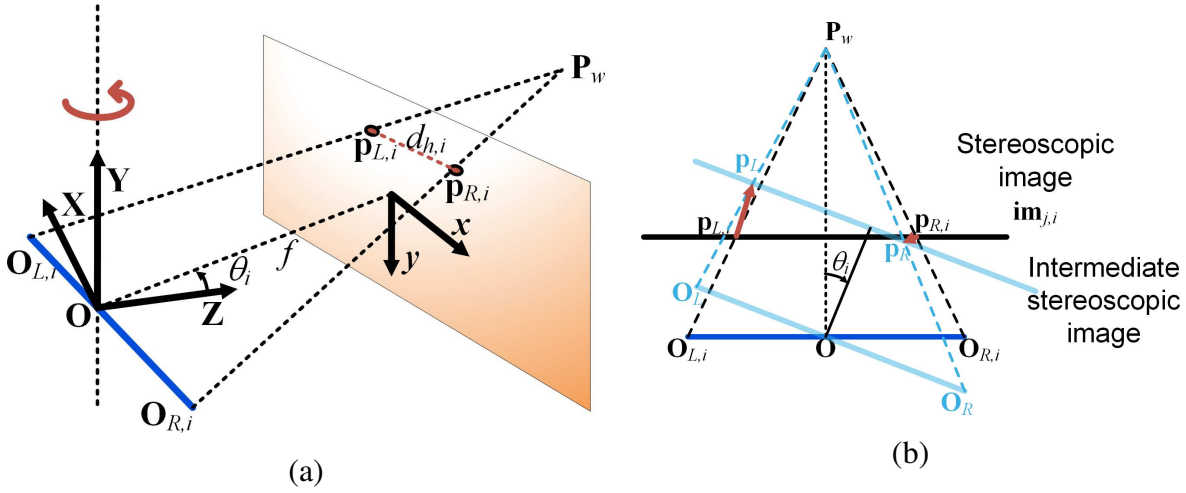


Figure 5.7: Rendering using multiple stereoscopic samples plus depth information: (a) acquisition example based on *configuration 1*, and (b) rendering of an intermediate stereoscopic view using the depth information to enforce the correct horizontal disparity.

This rendering approach can be visualized with an example. A point in the scene $\mathbf{P_W}$, located in the stereoscopic FOV of cameras $\Omega_{j,i}$ in gazing direction θ_i, will be projected into coordinates $\mathbf{p}_{j,i}$ in the corresponding images $\mathbf{im}_{j,i}$. A stereoscopic planar image can be generated using images $\mathbf{im}_{j,i}$ and $\mathbf{im}_{j,i+1}$, for any intermediate gazing direction between θ_i and θ_{i+1}. The image elements with coordinates $(\mathbf{p}_{L,i}, \mathbf{p}_{R,i})$ are mapped into the novel image into coordinates $(\mathbf{p}_L, \mathbf{p}_R)$. The correct horizontal disparity can be enforced by maintaining $dh_{j,i}(\mathbf{p}_{j,i})$ in the novel stereoscopic image. This rendering example is illustrated in Fig. 5.7-(b).

Synthesized stereoscopic views

According to the model for $\mathbf{C_S}$, the real stereoscopic view in intermediate gazing direction $\mathscr{D}$, involves sampling $(\mathbf{C_L}, \mathbf{C_R})$ for a limited FOV around the novel gazing direction and from the projective points $(\mathbf{O}_{L,k}, \mathbf{O}_{R,k})$, where k refers to $\theta_i < \theta_k < \theta_{i+1}$ (Fig. 2.4). However, using this rendering method, a novel stereoscopic image is created using the sampling of the $\mathbf{C_S}$ from $\mathbf{O}_{j,i}$ and $\mathbf{O}_{j,i+1}$. Note that, the novel stereoscopic view is not a mosaic of projective transformations of $\mathbf{im}_{j,i}$ and $\mathbf{im}_{j,i+1}$, since the depth information of the scene is used to correct the horizontal disparity pixel-wise.

The rendered stereoscopic view will be correct only in the acquired gazing directions where $\mathscr{D} = (\theta_i, \phi)$. Any intermediate view rendered for gazing directions between acquired stereoscopic samples is not exactly what the scene would look like in those gazing directions according to $\mathbf{C_S}$ model. However, this method can provide a satisfactory binocular illusion to the user as long as the correct horizontal disparity is enforced and visual artifacts are corrected.

Pros and cons of this method

The main advantage of this approach is that multiple stereoscopic samples can be acquired at once enabling the omnistereoscopic acquisition of dynamic scenes. Additionally, this method enables the synthesis of stereoscopic views incorporating mechanisms of binocular human vision such as vergence and accommodation. This can be done by manipulating the depth-field by using depth-based blurring (accommodation) and exploiting the depth information to manipulate the stereo image pair alignment (vergence). These aspects of the stereoscopic navigation still require further research and are beyond the scope of this thesis.

One challenging aspect of this acquisition-rendering approach is how to acquire high-quality depth maps of the scene omnidirectionally. If $\mathscr{M}$ is estimated from the collected stereoscopic samples, then a large baseline is necessary to improve the depth estimation. However, the larger the baseline the more challenging it is to avoid the occlusion between cameras in a simultaneous acquisition scenario.

Alternative: 2D-to-3D conversion

Another rendering method that uses a number of partial stereoscopic views of the scene plus depth consists in synthesizing a stereoscopic view from a 2D-to-3D conversion [62]. For instance, using a monoscopic view of the scene and the depth information extracted from $\mathscr{M}$

is possible to define a pixel-wise warping to create a novel view, which combined with the source image will compose a stereoscopic pair. In the synthesized stereoscopic pair, the horizontal disparity between corresponding points is enforced based on the desired depth illusion to create in the user [18, 129]. The acquisition model used in *configurations 2* and *3* is ideal for producing a panoramic view with minimum visual artifacts plus an omnidirectional depth map $\mathscr{M}$.

5.2.3 Mosaicking stereoscopic snapshots

A simpler omnistereoscopic rendering approach, which is also relies on a limited number of stereoscopic samples of the scene, is based on the mosaicking [2, 18]. This method only requires the mosaicking and blending of warped stereoscopic samples. The alignment can be done by applying a simple translational transformation [130, 131] and the stitching of stereoscopic pairs can be done using optimal cut algorithms [132, 133, 118].

Pros and cons of the mosaicking method

The low computational complexity of this method makes it ideal for creating video panoramas where a frame-based rendering is necessary [134, 117, 15, 135, 16]. This simultaneous sampling of $\mathbf{C_S}$ for N of gazing directions in azimuth can be done using acquisition *configuration 3* and *4*. However, it can also be implemented using stereoscopic samples obtained from sequential acquisition, i.e., using acquisition *configuration 1* and *2*.

The mosaicking method produces an estimation of the sampling of $\mathbf{I_S}$ that is correct only in the gazing directions that correspond to the acquired stereoscopic samples. As in the previous example, any view in an intermediate panning direction between actual snapshots is the result of mosaicking images from different stereoscopic cameras. The difference with the rendering method explained in Section 5.2.2 is that no depth information is used to correct the horizontal disparity. This is not a problem as long as depth continuity is enforced (Chapter 9). It is also important to correct the vertical disparities that appear within the overlapped regions of neighbour stereoscopic images to stitch. This important point is detailed in Chapter 9.

In this thesis, I focus on the system constraints to facilitate the sampling of $\mathbf{C_S}$ function for mosaicking-based rendering of $\mathbf{I_S}$.

Exploring the scene

The navigation is limited to panning the scene in azimuth unless depth information is used in rendering of intermediate views. The zero vergence point can be set using a horizontal translation in opposite directions to $\mathbf{I_R}$ and $\mathbf{I_L}$ [2] based on where the zero parallax point is located in the panoramic image. The conventions used in stereoscopic displaying, mainly originated in the stereoscopic film industry, are presented in Appendix C.

The omnistereoscopic navigation of the scene in a mosaicked scenario can be done by generating planar images corresponding to the left and right viewpoints with respect to the reference viewpoint of acquisition. This idea of generating stereoscopic images for the panning in azimuth is illustrated in Fig. 5.8-(a).

A projective transformation of any of these planar stereoscopic views can be used to emulate gazing above and below the equatorial plane ($\phi = 90°$) as illustrated in Fig. 5.8-(b) and (c). The projective transformation distorts the horizontal disparity field since the it changes the disparity between any pair of points that are horizontally aligned. However, the altered horizontal disparity cues are masked by other cues of depth such as perspective cues. This is an important effect that can be exploited by omnistereoscopic vision systems to reduce the complexity of the rendering algorithms.

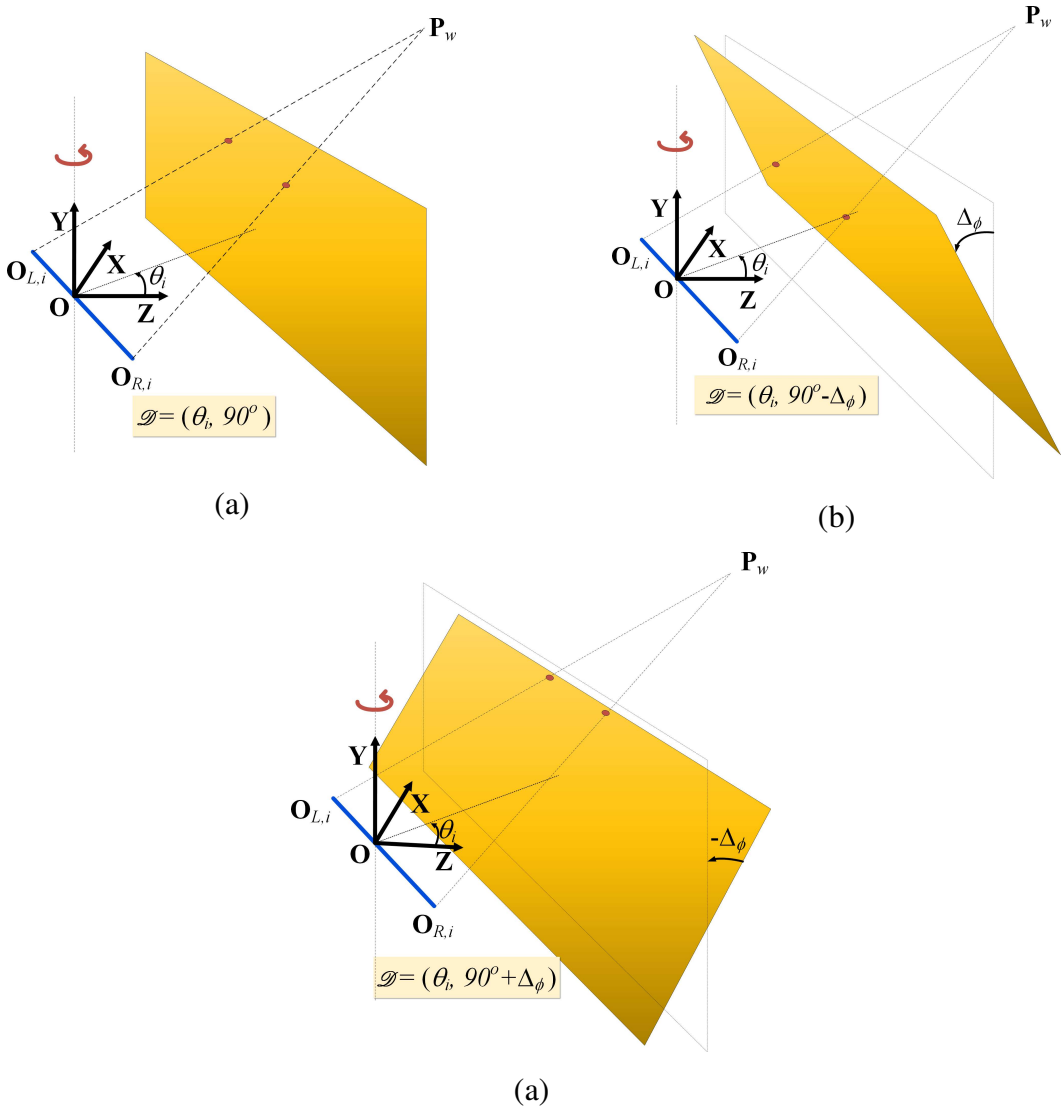
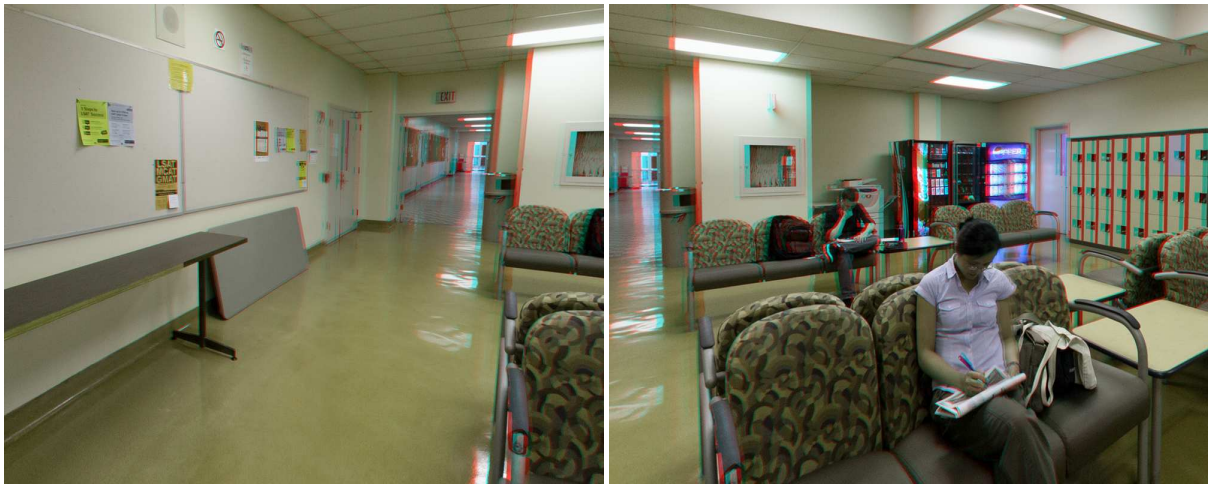
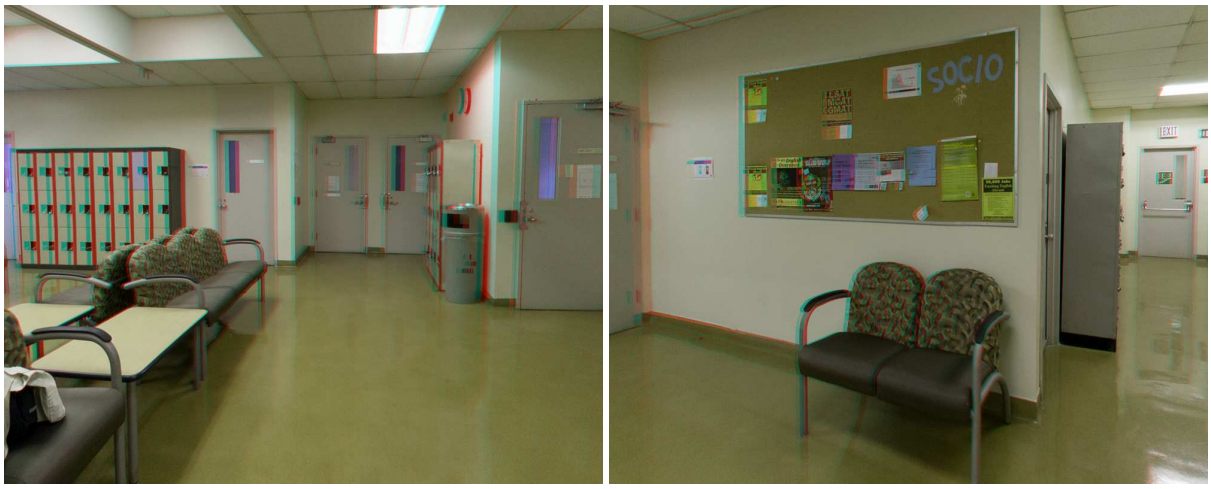


Figure 5.8: Navigation of the omnistereoscopic scene when $\mathbf{I_S}$ was rendered by mosaicking partially overlapped stereoscopic snapshots: (a) panning in azimuth, (b) gazing up, and (c) gazing down.

In Fig. 5.9 there is an example of navigation by generating planar views from $\mathbf{I_S}$ at four different panning angles. The panning angles in this example do not coincide with the sampling angles θ_i. Therefore, these stereoscopic views are the mosaic of partial stereoscopic views of the scene acquired at different but close angles in azimuth. Note that this mosaicking does not affect the continuity in the depth illusion by binocular cues, i.e., those based on the horizontal disparity. This example was created using the stereoscopic images acquired with a camera with the acquisition *configuration 2*, where the baseline is 32.5mm.



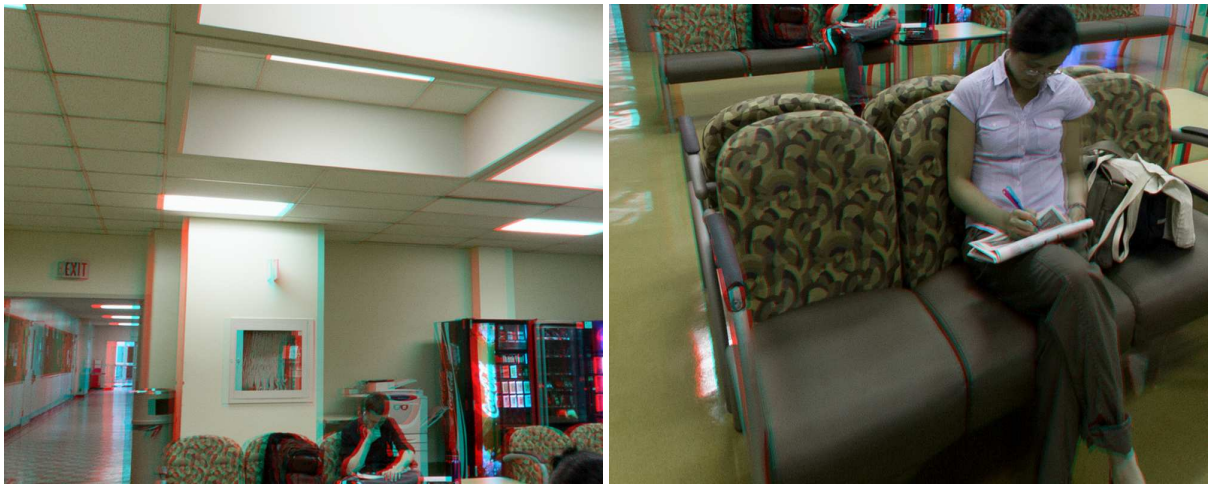
(θ_0) (θ_1)



(θ_2) (θ_3)

Figure 5.9: Example or scene panning: rectangular stereoscopic views (Red-Cyan anaglyph) extracted from $\mathbf{I_S}$, which was rendered by mosaicking stereoscopic images acquired with a camera set-up based on acquisition *configuration 2*.

A limited exploration can be done by changing $\mathscr{D}$ in elevation. This is accomplished by rendering perspective projections of the partial FOV of the scene. Without disparity adjustments, depth distortions are created at different vertical coordinates in the partial FOV images depending of the gazing elevation angle. This is illustrated in Fig. 5.10 for two gazing angles.



$\mathscr{D} = (\theta_1, \phi = 30°)$ $\mathscr{D} = (\theta_1, \phi = 120°)$

Figure 5.10: Two stereoscopic views (Red-Cyan anaglyph) extracted from $\mathbf{I_S}$, which corresponds to θ_1 in Fig. 5.9 and two different elevation angles: the views are created as projective transformations to give the user the illusion of gazing above $(\phi = 30°)$ and below $(\phi = 120°)$ the equatorial plane.

The dynamic generation of stereoscopic views, emulating vergence and accommodation mechanisms requires additional pixel-wise depth information. However, a basic exploration of an omnistereoscopic scene can be done by projective transformations as described in this section without compromising the user's immersive experience.

Mosaicking and depth distortions

The distance between the omnistereoscopic camera and scene affects the depth illusion achieved, in particular, when the depth cues rely on exploiting the binocular mechanism of human vision. The disparity arising from viewing the scene from two viewpoints with horizontal parallax is relevant for the binocular depth perception only when the distance between camera and scene is relatively small, e.g., from a few tens to a few hundreds of baseline units. Beyond that distance, perspective cues are more important than binocular ones for the depth perception.

It will be important to determine how the different acquisition configurations affect the disparity when mosaicking stereoscopic images. In this case, the limit for human depth resolution developed in Section 5.1.5 can be used to assess the limitations of the rendering based on mosaicking.

For stereoscopic cameras with parallel optical axes, when the scene is at several hundreds of times the baseline length, it is valid to assume that the scene is on a plane perpendicular to the optical axes. In this case, each stereoscopic sample $\mathbf{im}_{j,i}$ at θ_i exhibits a constant and small horizontal disparity $d_{h,i}$ over all the image. This 3D model of the scene is shown in Fig. 5.11-(a). The planar model is no longer valid when the camera gets closer to the scene.

Although it is not possible to model all scene distributions, it is valid to assume a particular 3D distribution of the scene around the camera. This abstraction leads to important insights about the mosaicking method. For instance, when the scene and is uniformly distributed around the reference viewpoint, a spherical distribution model as the one illustrated in Fig. 5.11-(b) can be used.

Another useful model for the scene spatial distribution is the cylindrical scene model. The latter can be used when the scene is radially distributed around the viewpoint at a constant distance, but the scene's depth changes with the gazing elevation. The cylindrical model is illustrated in Fig. 5.11-(c).

Both models for the 3D distribution of the scene are mere approximations that can be used to understand what is the problem of mosaicking stereoscopic views when a scene is distributed uniformly around a multiple camera system.

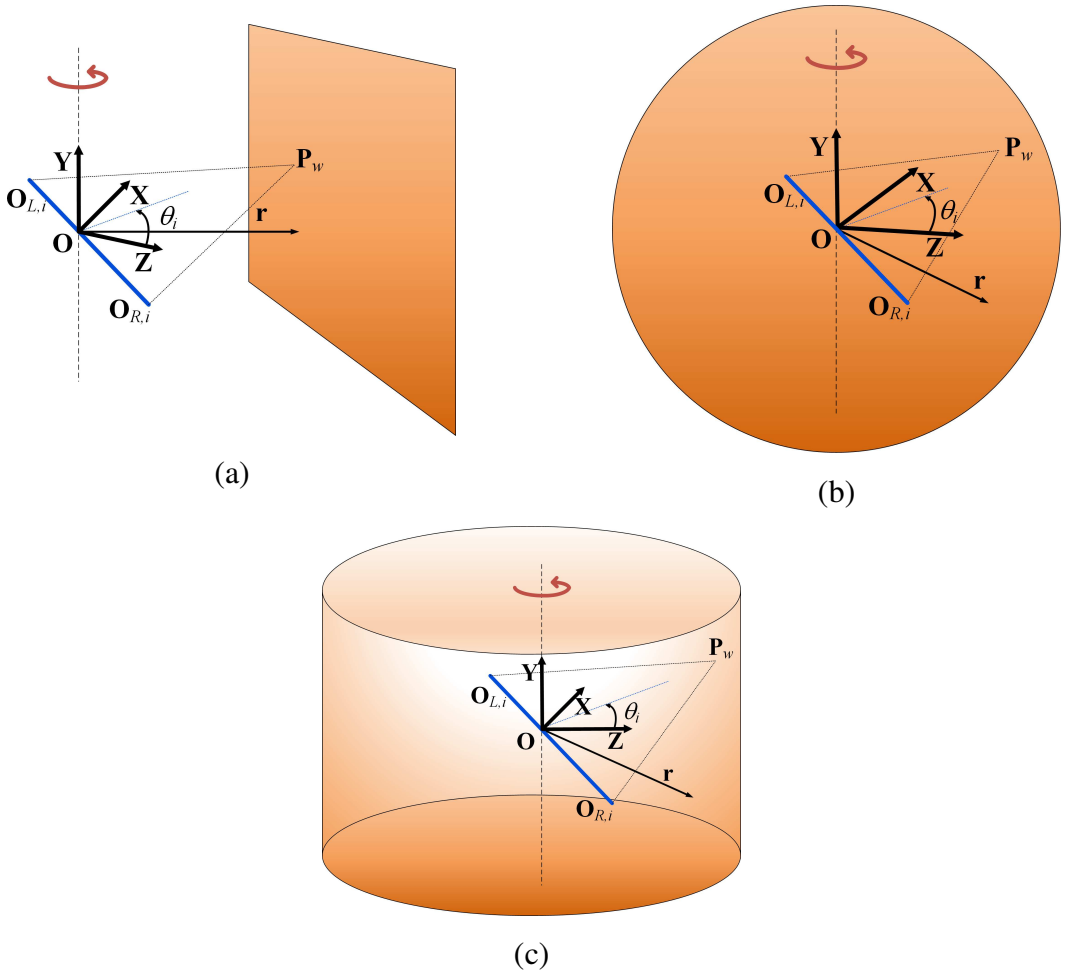


Figure 5.11: Scene depth distribution models: (a) a planar distribution for a large $\|\mathbf{r}\|$, (b) a spherical distribution, and (c) a cylindrical distribution around the camera, both for a closer $\|\mathbf{r}\|$ to the camera.

In both models for the scene, the points on the median plane parallel to the $\mathbf{ZY}$-plane, which is located between $\mathbf{O_L}$ and $\mathbf{O_R}$ in *configuration 1* and *4*, are imaged with the same horizontal disparity $d_{h,i}$ that correspond the scene's depth Z_i. However, the larger is the gazing eccentricity, the closer are the projections $\mathbf{Z'_i}$ from $\mathbf{O}$. Therefore, larger horizontal disparities are observed on the periphery of the $\mathbf{im}_{j,i}$ with respect to its central region. However, in both models of the scene, the same horizontal disparity should be exhibited in all directions given the symmetric distribution of depth around the reference viewpoint.

According to $\mathbf{C_S}$, each stereoscopic image extracted from $\mathscr{I}_S$ has to exhibit the same distribution of horizontal disparities over the whole image given the symmetric distribution of the scene around the reference viewpoint $\mathbf{O}$. The rendering by mosaicking warps and stitches the stereoscopic snapshots. So it is relevant to find out the maximum image width that can be used to make the horizontal disparity continuous.

The region for mosaicking without a noticeable depth discontinuity is based on finding the eccentricity limits where the user will not perceive a noticeable difference in depth due to binocular cues only. The *horizontal disparity difference* defined in Eq. 5.9 can be used here to determine the regions of horizontal disparity continuity for mosaicking.

For example, a point $\mathbf{P_W}$ is in direction $\mathscr{D}$, where $\mathscr{D} = (\theta_i, 90^\circ + \Delta_\phi)$, and it determines the corresponding set of points $\mathscr{K}_i$. Another different point, $\mathbf{P'_W}$, is in direction $\mathscr{D}'$, where $\mathscr{D}' = (\theta_i + \Delta_\theta, 90^\circ + \Delta_\phi)$, and it produces the corresponding set of points $\mathscr{K}_i{}'$. In this example, the angles Δ_θ and Δ_ϕ are the eccentricities in azimuth and elevation, respectively. This example is illustrated graphically in Fig. 5.12. The depth of these two points can be estimated by their respective horizontal disparities $d_{h,i}(\mathscr{K}_i)$ and $d_{h,i}(\mathscr{K}_i{}')$. The discrepancy in the horizontal disparity for each points (Eq. 5.17) will produce a discernible depth for $\mathbf{P_W}$ and $\mathbf{P'_W}$ based on binocular cues only when $e_{h,i}$ is above the threshold of *human depth resolution*.

Evaluation of the mosaicking width

The analysis of the regions in each stereoscopic sample $\mathbf{im}_{j,i}$ that can be used for mosaicking while maintaining horizontal depth continuity provides a first insight on the problem of the distance between the scene and the reference viewpoint. In order to obtain a quantitative evaluation of this problem, I simulated an omnistereoscopic camera in acquisition *configuration 1* for the cylindrical and spherical model of the scene.

In this experiment, each camera has a focal length $f = 9.3$ mm, the image plane width is $W_h = 22.2$ mm and its aspect ratio is $a_r = 1$. These parameters determine a 100° FOV in

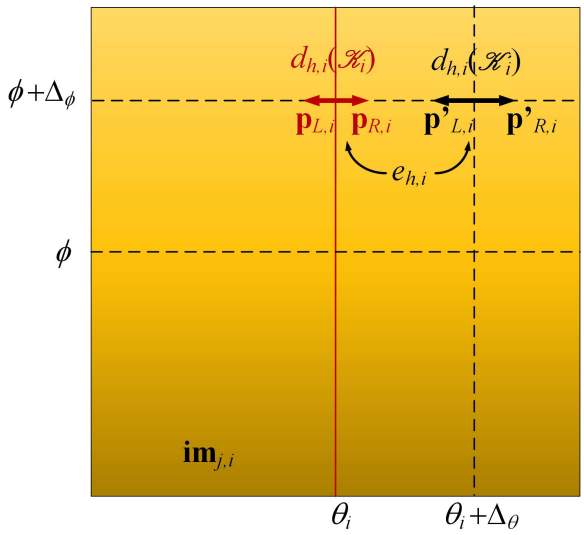


Figure 5.12: The horizontal disparity difference calculated over $\mathbf{im}_{j,i}$.

azimuth and elevation for each pinhole camera. The stereoscopic baseline used in the simulation is $b = 65$ mm which is similar to the interocular distance in the average human population [127], and a narrower baseline of $b = 32.5$ mm. Finally, three distances to the scene $\|\mathbf{r}\|$ are considered: 1 m, 5 m and 10 m. In case of the cylindrical model, $\|\mathbf{r}\|$ is defined in the equatorial reference plane, also defined as the floor plane. The longer the distances from the camera the weaker the binocular cues are, unless a larger baseline is used.

Two different thresholds for $e_{h,i}$ are used: one is derived from the 20 arcsec disparity threshold proposed in [121] and the second is based on the pixel width of the Cannon Rebel Xti 400D sensor which is 126 arcsec. As explained in Section 5.1.6, a pixel width is an attainable and valid threshold for $e_{h,i}$ given today's camera resolutions.

These simulations shows the usable image width as a function of $\|\mathbf{r}\|$. In Fig. 5.13, the usable mosaicking regions for each $\mathbf{im}_{j,i}$ are depicted for a cylindrical scene distribution whereas in Fig. 5.14 the same is illustrated for a spherical scene model. Note that in both cases, the closer the scene is to the viewpoint, the narrower is the usable image width. In both scene models, the equatorial image region ($\theta = 90°$) defines the usable image width for mosaicking.

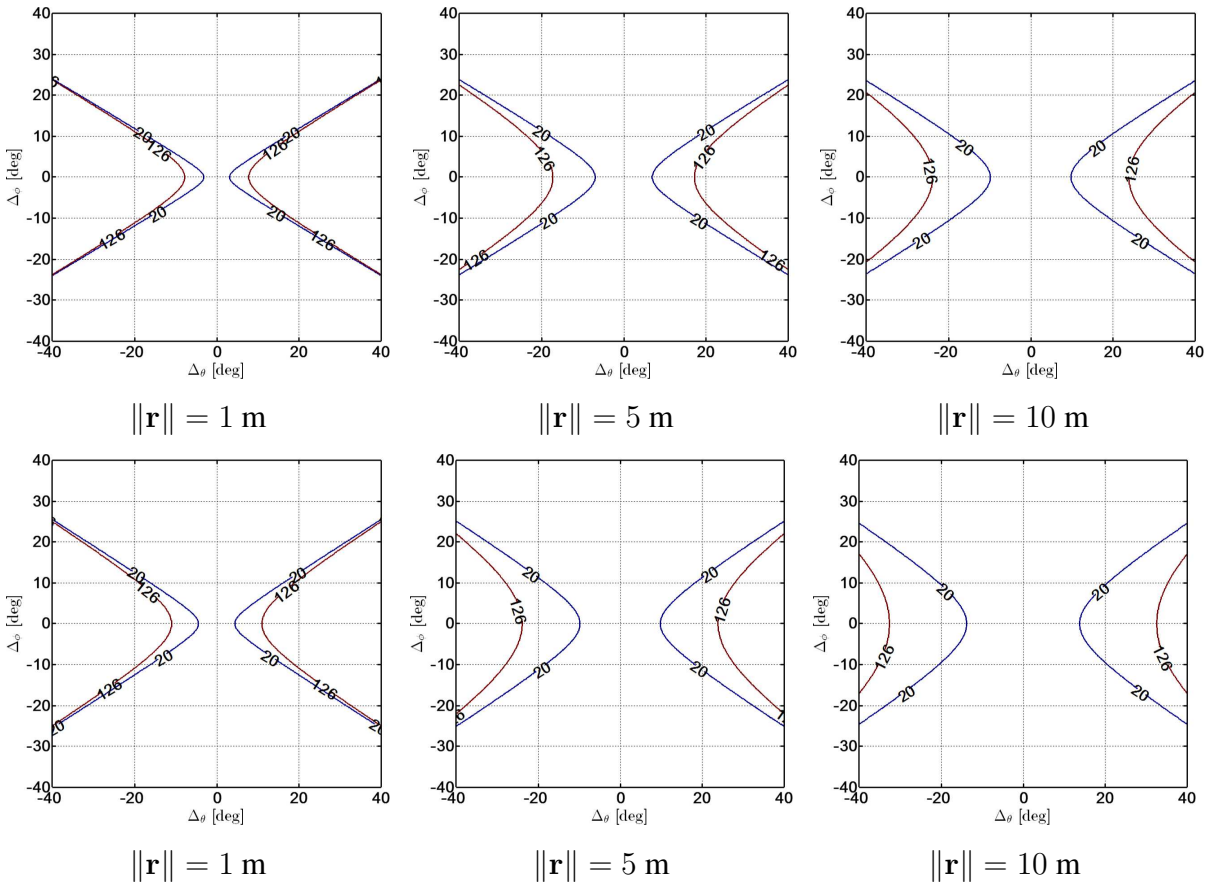


Figure 5.13: Limits of usability of $\mathbf{im}_{j,i}$ based on the horizontal disparity resolution for mosaicking using the cylindrical model of the scene. The region limits for 20 and 126 arcsec are indicated in each figure. The top row corresponds to $b = 65$ mm and the bottom row is for $b = 32.5$ mm.

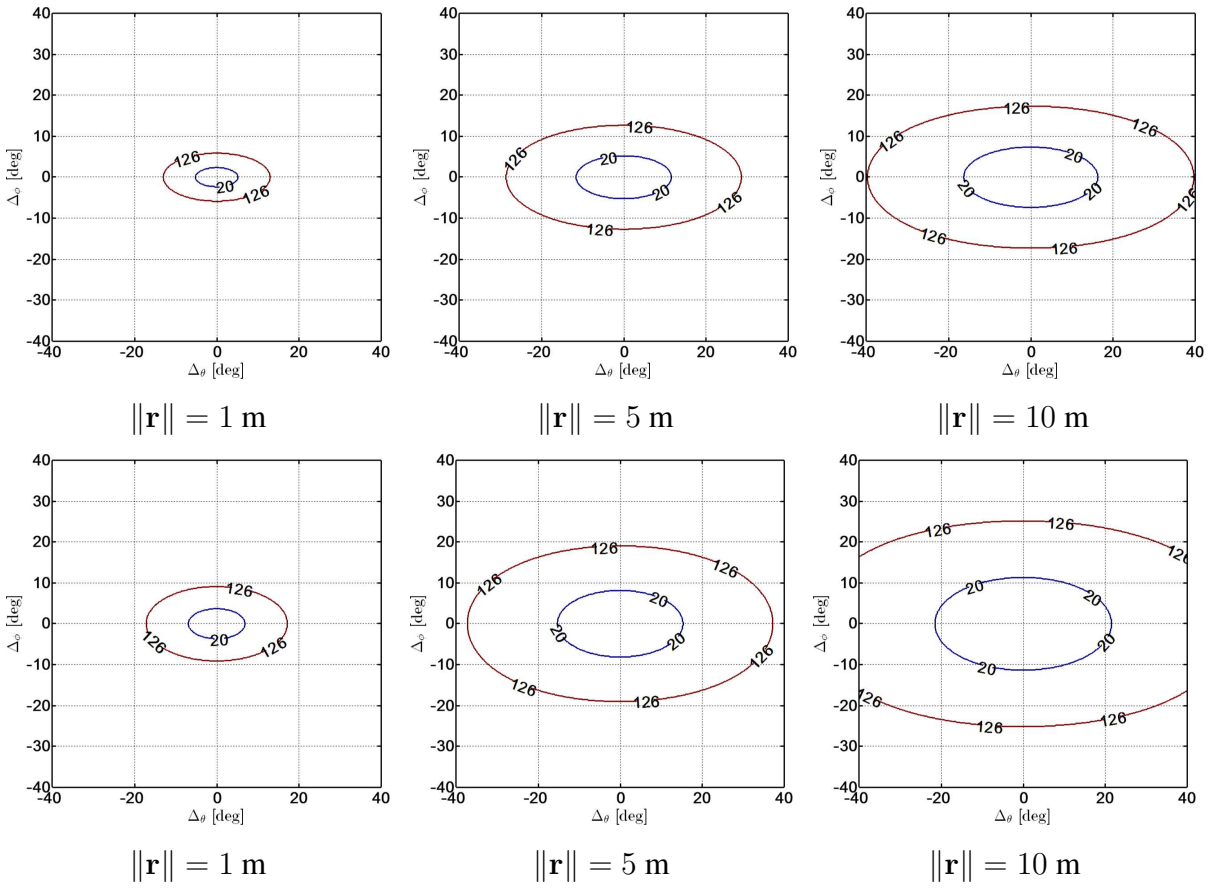


Figure 5.14: Limits of usability of $\mathbf{im}_{j,i}$ based on the horizontal disparity resolution for mosaicking using the spherical model of the scene. The region limits for 20 and 126 arcsec are indicated in each figure. The top row corresponds to $b = 65$ mm and the bottom row is for $b = 32.5$ mm.

A useful information extracted from these simulations is how many stereoscopic samples will be necessary to sample $\mathbf{C_S}$ and to render $\mathbf{I_S}$ by mosaicking these samples. For instance, when the scene can be approximated by the cylindrical model and the distances to the scene are between 6 and 15 m, based on the results in Fig. 5.13 (top right), the FOV usable for mosaicking to guarantee a continuous depth illusion should have between $22.5°$ and $36°$ in eccentricity ($45°$ and $75°$ angular spread azimuth) for a baseline $b = 65$ mm and a threshold $\epsilon = s$. This implies the we must acquire from six to eight stereoscopic samples to cover $360°$ in azimuth. But, when $b = 32.5$ mm (Fig. 5.13 (bottom right)), the usable FOV of each $\mathbf{im}_{j,i}$ can be widen in azimuth. Therefore, the number stereoscopic samples can be reduced for the same threshold for the horizontal disparity resolution.

5.3 Summary

In this chapter, I introduced a formal definition for the disparities that arise when the scene is projected in a stereoscopic pair of cameras from the acquisition model. The formal definition for the disparity functions was necessary to define the problem of human depth threshold. This is important to determine how to efficiently sample $\mathbf{C_S}$, using a reduce number of snapshots, in order to produce a continuous illusion of depth around any viewpoint.

The other important point I presented in this chapter are the rendering methods to create $\mathbf{I_S}$ in the framework of the proposed acquisition model (Chapter 3). In order to do this, I proposed a classification of the rendering methods in three main techniques:

- First, I presented a rendering model based on mosaicking a large number of narrow stereoscopic slides (image columns) acquired by a rotating stereoscopic rig (*configuration 1* and *2*).

- Second, I showed another rendering method that relies on a reduced number of partially stereoscopic images and additional depth information of the scene. The latter comes in the form of an omnidirectional depth map $\mathscr{M}$ which must be extracted from the acquired images or estimated from an additional depth-sensing method. This approach of rendering is applicable to all the acquisition configurations.

- Third, I presented a simplified rendering method that does not require $\mathscr{M}$. This approach relies on mosaicking partially overlapped stereoscopic images of the scene and it is also applicable to all the acquisition configurations.

The second and the third rendering approaches are suitable to produce omnistereoscopic imagery of dynamic scenes. The third rendering method is particularly attractive to produce real-time omnistereoscopic imagery given its reduced computational complexity. However, there are acquisition constraints for the mosaicking approach that need to be researched. In this thesis, I investigate some of the constraints relevant to the mosaicking approach.

Finally, I presented two 3D models of the scene distribution that are simplifications of several real-world scenarios. One of the models is based on an equidistant distribution around the reference viewpoint in all directions (sphere). The second model assumes an equidistant distribution in azimuth only (cylindrical). Applying the thresholds for human depth resolution proposed in this chapter, I obtained the regions on $\mathbf{im}_{j,i}$ where the difference between the real

and the rendered horizontal disparities cannot be perceived by the user. From this results, I presented a first approximation to the usable regions in $\mathbf{im}_{j,i}$ that can be used in the mosaicking.

The 3D models of the scene introduced in this chapter are too simple to represent complex real-world scenes where some elements are closer than other of the reference viewpoint. For that reason, in Chapter 8, I present a ray-tracing method to model the relationship between the stitching position in $\mathbf{im}_{j,i}$ and the acquisition model constraints.

In the next chapter, I present the formulation necessary for the mosaicking of stereoscopic snapshots.

Chapter 6

Mosaicking of stereoscopic images

The mosaicking approach to rendering introduced in the last chapter can be implemented with any of the acquisition configurations defined in Chapter 3. For instance, a set of N partially overlapped stereoscopic images, $(\mathbf{im}_{L,i}, \mathbf{im}_{R,i})$ for $(i \in \{0, ..., N-1\})$, can be acquired by using sequential methods (*configurations 1* or *2*) or by using simultaneous acquisition configurations (*configuration 2* or *3*).

In real-world scenarios, pairwise image rectification between stereoscopic image pairs may be necessary. The rectification enforces the parallel optical axis constraint. This can be done, either by using the known camera parameters, or by defining a set of corresponding feature points when calibration is not available [136, 137].

After stereo rectification, we need to define the coordinate transformation or *stereo correspondence* between corresponding coordinate points in the adjacent images to be mosaicked. This is necessary for the stitching of adjacent images in order to reduce the visual artifacts due to images acquired from distinct projection centers [134, 18]. These visual artifacts are referred herein as *stitching ghosting* [131]. As in the case of stereoscopic rectification, the stereo correspondence between neighbour images can be done by either a calibrated or by an uncalibrated approach. The calibrated method requires to perform a multiple camera calibration [138, 139, 140], while the uncalibrated approach is based on extracting the coordinates of corresponding image features in the overlapping regions to mosaic [131].

6.1 Image alignment

The alignment between left and right images is necessary to reduce the visual artifacts after stitching and blending images to render $\mathbf{I_L}$ and $\mathbf{I_R}$. The alignment requires defining the corresponding points in the overlapping region between neighbor stereoscopic images, e.g., between $(\mathbf{im}_{j,i}, \mathbf{im}_{j,i-1})$ and $(\mathbf{im}_{j,i}, \mathbf{im}_{j,i+1})$, for $j = \{L, R\}$.

In the next section, I extend the definition of corresponding points within a stereoscopic image to a more general case used for the alignment.

6.1.1 Corresponding projections between neighbor images

The *corresponding image coordinates* for a point $\mathbf{P_W}$ have been defined in Section 5.1 for the stereoscopic image $\mathbf{im}_{j,i}$ acquired with a stereoscopic camera oriented at θ_i. This corresponding pair of image coordinates has been denoted as

$$\mathscr{K}_i = (\mathbf{p}_{L,i}, \mathbf{p}_{R,i}),$$

where $\mathbf{p}_{j,i}$ is the coordinates of the left $(j = L)$ and right $(j = R)$ projections of the point $\mathbf{P_W}$ located in the stereoscopic FOV when the camera pair $\Omega_{j,i}$ is oriented in the panning direction θ_i.

A point $\mathbf{P_W}$ located in the FOV of cameras $\Omega_{j,i}$ and $\Omega_{j,i+1}$ is projected on the overlapping region of stereoscopic samples $\mathbf{im}_{j,i}$ and $\mathbf{im}_{j,i+1}$. This point in 3D-space is then located in the *omnistereoscopic FOV*; more about this will be said in Section 8.1. The pairwise alignment of images $(\mathbf{im}_{L,i}, \mathbf{im}_{L,i+1})$ and $(\mathbf{im}_{R,i}, \mathbf{im}_{R,i+1})$ requires the *corresponding coordinate points* between neighbor images. This corresponding set of coordinates is denoted as

$$\mathscr{K}_{j,i} = (\mathbf{p}_{j,i}, \mathbf{p}_{j,i+1}), \tag{6.1}$$

where $\mathbf{p}_{j,i}$ and $\mathbf{p}_{j,i+1}$ are the respective coordinates on $\mathbf{im}_{j,i}$ and on $\mathbf{im}_{j,i+1}$ of the projection of $\mathbf{P_W}$.

The alignment between neighbor stereoscopic images requires the knowledge on $\mathscr{K}_{j,i}$ for all $i \in \{0, ..., N-1\}$.

6.1.2 Matrix form of the projection equations

The transformations from world to camera frame of coordinates can be expressed as a simple product [136] by using the homogeneous equations. Given a point in the scene $\mathbf{P_W} =$

$(X_W, Y_W, Z_W)^T$ in the world reference of coordinates, the same point can be expressed on the camera $\Omega_{j,i}$ frame of coordinates as $\mathbf{P_{j,i}} = (X_{j,i}, Y_{j,i}, Z_{j,i})^T$ as follows

$$\begin{aligned}\mathbf{P_{j,i}} &= \mathbf{E_{j,i}} \cdot (\mathbf{P_w}^T, 1)^T \\ &= \mathbf{E_{j,i}} \cdot (X_w, Y_w, Z_w, 1)^T,\end{aligned} \tag{6.2}$$

where $\mathbf{E_{j,i}}$ is the extrinsic matrix for camera $\Omega_{j,i}$. These matrices contains the spatial information of each camera frame with respect to the global reference frame and are defined as

$$\mathbf{E_{j,i}} = \begin{pmatrix} \cos\theta_i & 0 & \sin\theta_i & -\mathbf{R_{1,i}}\mathbf{T}_{j,i} \\ 0 & 1 & 0 & -\mathbf{R_{2,i}}\mathbf{T}_{j,i} \\ -\sin\theta_i & 0 & \cos\theta_i & -\mathbf{R_{3,i}}\mathbf{T}_{j,i} \end{pmatrix}, \tag{6.3}$$

where $\mathbf{R}_{k,i}$ for $k \in \{1, 2, 3\}$ was defined in the previous section as a row vector of $\mathbf{R}_i$, which was defined along the translation vectors $\mathbf{T}_{j,i}$ in Section 3.3.

The matrix form of the projective transformation is given by

$$\mathbf{p}_{j,i} = \mathbf{D} \cdot \mathbf{P}_{j,i}, \tag{6.4}$$

where the projection or intrinsic matrix $\mathbf{D}$ is the same for both cameras since projective parameters are assumed the same for all cameras and is given by

$$\mathbf{D} = \begin{pmatrix} -f & 0 & 0 \\ 0 & -f & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{6.5}$$

where f is the cameras' focal length (Eq. 6.21). From Eq. 6.4 and Eq. 6.2, it follows that

$$\begin{aligned}\mathscr{P}_{j,i} &= \mathbf{D} \cdot \mathbf{E_{j,i}} \cdot (\mathbf{P_w}^T, 1)^T \\ &= \mathbf{M_{j,i}} \cdot (\mathbf{P_w}^T, 1)^T,\end{aligned} \tag{6.6}$$

where $\mathscr{P}_{j,i} = (\mathscr{X}_{j,i}, \mathscr{Y}_{j,i}, \mathscr{Z}_{j,i})^T$ and $\mathbf{M_{j,i}}$ is the projective transformation given by

$$\mathbf{M_{j,i}} = \mathbf{D} \cdot \mathbf{E_{j,i}}. \tag{6.7}$$

The matrix $\mathbf{M_{j,i}}$ enables mapping a point $\mathbf{P_W}$ in the world coordinate frame into a 2D coordinate point $\mathbf{p}_{j,i}$ in the image frame of coordinates. Replacing Eq. 6.5 and Eq. 6.3 into Eq. 6.7, the closed form of the projective matrix for the camera $\Omega_{j,i}$ is given by

$$\mathbf{M_{j,i}} = \begin{pmatrix} -f\cos\theta_i & 0 & -f\sin\theta_i & f\mathbf{R_{1,i}}\mathbf{T}_{j,i} \\ 0 & -f & 0 & f\mathbf{R_{2,i}}\mathbf{T}_{j,i} \\ -\sin\theta_i & 0 & \cos\theta_i & -\mathbf{R_{3,i}}\mathbf{T}_{j,i} \end{pmatrix}. \tag{6.8}$$

The image coordinates $\mathbf{p}_{j,i} = (x_{j,i}, y_{j,i})$ for the image plane j and the panning angle θ_i are given by

$$x_{j,i} = \frac{\mathscr{X}_{j,i}}{\mathscr{Z}_{j,i}} = -f\frac{X_{j,i}}{Z_{j,i}}, \tag{6.9}$$

$$y_{j,i} = \frac{\mathscr{Y}_{j,i}}{\mathscr{Z}_{j,i}} = -f\frac{X_{j,i}}{Z_{j,i}}. \tag{6.10}$$

The mapping between $\mathbf{P_W}$ and its coordinates in $\mathbf{im}_{j,i}$ after projection is relevant to define a set of $\mathscr{K}_{j,i}$ on the overlapping regions of consecutive image samples. This set of $\mathscr{K}_{j,i}$ can be used to align consecutively acquired planar images prior to mosaicking them.

6.2 Stereo rectification

A point $\mathbf{P_W}$ located in the *stereoscopic FOV* of a stereoscopic camera $\Omega_{j,i}$ can be projected onto the corresponding image planes by the transformations

$$\mathbf{P}_{L,i} = \mathbf{M_{L,i}} \cdot (\mathbf{P_w}^T, 1)^T \tag{6.11}$$

$$\mathbf{P}_{R,i} = \mathbf{M_{R,i}} \cdot (\mathbf{P_w}^T, 1)^T. \tag{6.12}$$

The constraints of *epipolar geometry* help to obtain a correspondence transformation between coordinates in any stereoscopic image of the acquired set. The calibrated approach to stereo rectification requires the pairwise camera calibration of the whole multiple camera system to obtain the *intrinsic* and *extrinsic* parameters of each camera. The prior system calibration enables the definition of transformations that makes the conjugate epipolar lines collinear and parallel to the $\mathbf{X}_j$-axe [136].

Since, this calibrated approach was not used in the experimental part of this thesis, the deduction of the closed form of these transformation matrices will not be presented here. The uncalibrated method used for the stereo rectification will be presented in Section 7.7.

6.2.1 Stereo correspondence for mosaicking

The same point $\mathbf{P_W}$ is projected in the overlapping region of two consecutive stereoscopic samples in $\mathbf{p}_{j,i}$ into $\mathbf{im}_{j,i}$ and $\mathbf{p}_{j,i+1}$ into $\mathbf{im_{j,i+1}}$. The overlapping regions are defined for

$j = L$ and $j = R$. The projections of the point $\mathbf{P_W}$ on each of the consecutive images in the acquired sequence are

$$\mathbf{P}_{j,i} = \mathbf{M_{j,i}} \cdot (\mathbf{P_w}^T, 1)^T \tag{6.13}$$

$$\mathbf{P}_{j,i+1} = \mathbf{M_{j,i+1}} \cdot (\mathbf{P_w}^T, 1)^T. \tag{6.14}$$

This is valid for a point $\mathbf{P_W}$ visible simultaneously in cameras $\Omega_{j,i}$.

The *epipolar geometry*, defined for any pair of pinhole cameras of the acquisition model, helps to define a matching transformation between corresponding points [136, 137]. In this context, the aligning transformation is denoted as

$$\mathbf{P}_{j,i+1} = \mathbf{H_{j,(i,i+1)}} \cdot \mathbf{P}_{j,i}, \tag{6.15}$$

where the matrix $\mathbf{H_{j,(i,i+1)}}$ defines the a transformation between frames of coordinates of camera $\Omega_{j,i}$ and camera $\Omega_{j,i+1}$.

The stereo correspondence matrix defined in Eq. 6.15 makes possible the alignment of neighbour images for a rendering based on mosaicking. However, this method requires to know the *intrinsic* and *extrinsic* camera parameters for $(\Omega_{j,i}, \Omega_{j,i+1})$, for $i \in \{0, ..., N-1\}$ and for the left $(j = L)$ and right $(j = R)$ camera sets.

An uncalibrated approach that relies on a large set of $\mathscr{K}_{j,i}$ to define a transformation between planar projections to match corresponding points is presented in Section 7.9.

In this contest, the alignment operation that enables matching image coordinates $\mathbf{p}_{j,i}$ to the correspondent coordinate $\mathbf{p}_{j,i}$, defining a corresponding set $\mathscr{K}_{j,i}$, is denoted as

$$\mathbf{p}_{j,i+1} = \mathscr{A}_{j,(i,i+1)}\mathbf{p}_{j,i}, \tag{6.16}$$

where $\mathscr{A}_{j,(i,i+1)}$ can be derived from the stereo correspondence procedure used for the alignment.

6.3 Stitching and blending

Each stereoscopic pair $\mathbf{im}_{j,i}$ must be stitched with the next image $\mathbf{im}_{j,i+1}$ in the acquisition sequence for all $i \in \{0, ..., N-1\}$. The alignment and stitching is done for coordinate points $\mathbf{p}_{j,i} = (x_b, 0)$ and $\mathbf{p}_{j,i+1} = (x'_b, 0)$. This example assumes clockwise acquisition of stereoscopic images. The point of alignment in the image $\mathbf{im}_{j,i+1}$ is given by

$$(x'_b, 0) = \mathscr{A}_{j,(i,i+1)}(x_b, 0). \tag{6.17}$$

In the case of stitching images acquired from a camera rotated at regular Δ_θ intervals around its nodal point, the alignment point will be symmetrically located in the image, e.g., $x'_b = -x_b$ and the coordinate transformation for aligning both images will be a pure horizontal translation. This case is illustrated in Fig. 6.1-(a).

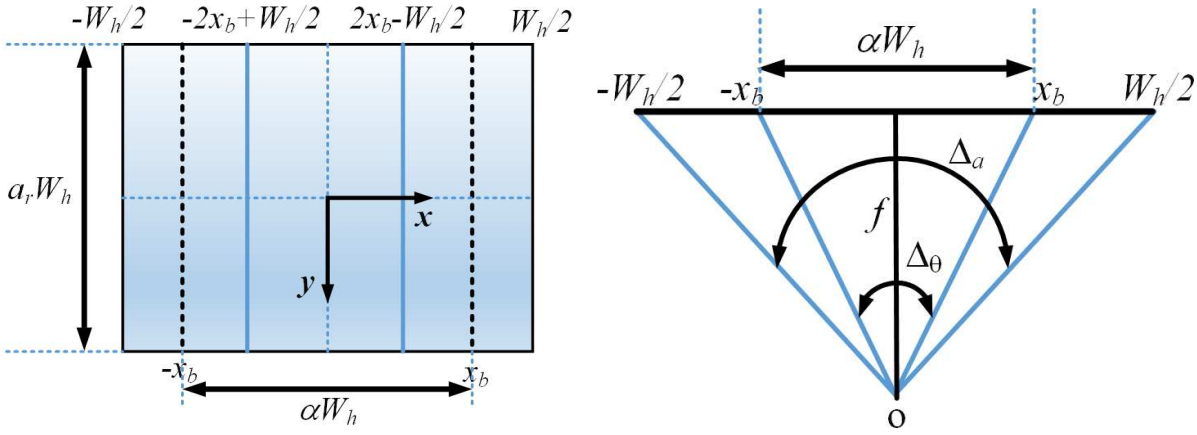


Figure 6.1: The parameters for each image: (a) the frame of coordinates location on the image plane and (b) the field-of-view parameters and their relationship with f and W_h.

Acquisition system parameters

The case of a pinhole camera rotated around its projection center to capture images at regular intervals is useful to make a first estimation of the number of samples N for a given focal length f. So, I use this model to determine these parameters.

The camera FOV (Δ) is defined as the angle-of-view in the main diagonal of the image window. In our acquisition model, the camera FOV in the horizontal dimension (Δ_a) is used.

The region of each image defined for $x_{j,i} \in [-x_b, x_b]$ used in the mosaicking is a fraction of the image width (αW_h) as illustrated in Fig. 6.1. The parameter α is given by

$$\alpha = \frac{2\,f}{W_h} \tan\left(\frac{\Delta_\theta}{2}\right), \tag{6.18}$$

where Δ_θ is the sampling angle in azimuth defined as

$$\Delta_\theta = \frac{360^\circ}{N}. \tag{6.19}$$

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

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

$$x_b = f \tan\left(\frac{\Delta_\theta}{2}\right), \tag{6.20}$$

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

$$f = \frac{W_h}{2\tan\left(\frac{\Delta_a}{2}\right)}. \tag{6.21}$$

The stitching positions $x_{j,i} = \pm x_b$ are in the horizontal middle of the overlapping regions for any Δ_a and Δ_θ. The geometric depiction of these magnitudes is presented in Fig. 6.1-(b).

6.3.1 Overlapping regions

The overlapping regions between any two neighbour images in a set can be defined by the image regions where the same scene is simultaneously projected in both images. In our acquisition model, two overlapping regions can be defined in any image $\mathbf{im}_{j,i}$. Following the same simplified model than in the previous section, these regions are located in

$$-\frac{W_h}{2} \leq x_{j,i} \leq -2x_b + \frac{W_h}{2}, \tag{6.22}$$

$$2x_b - \frac{W_h}{2} \leq x_{j,i} \leq \frac{W_h}{2}. \tag{6.23}$$

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

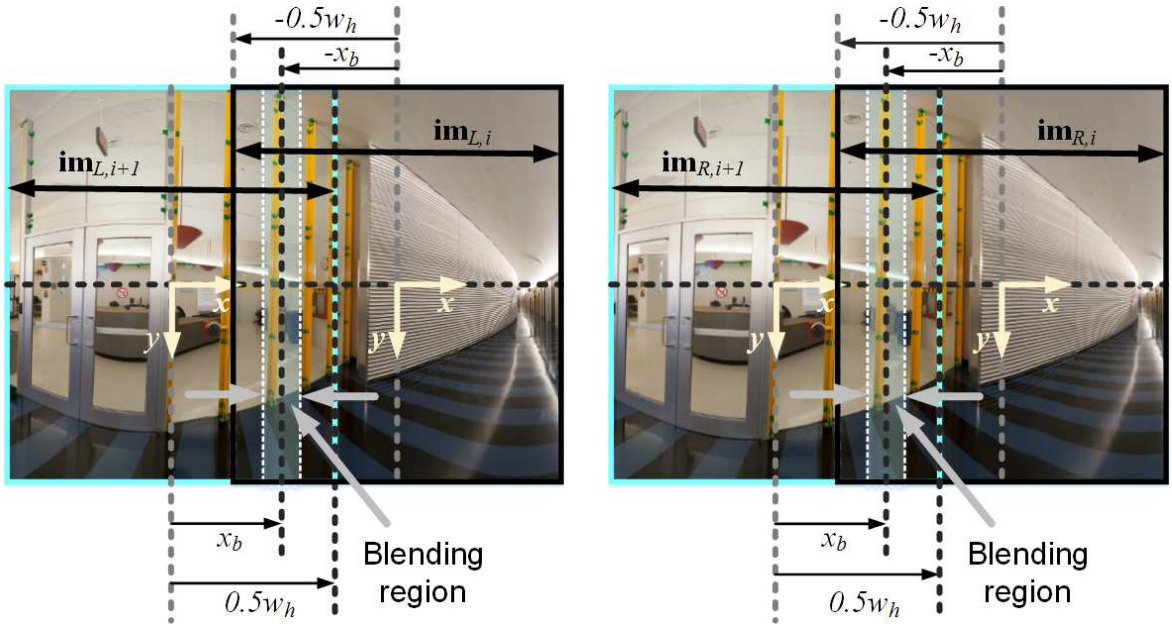
This is just an estimation based on the simplified model represented in Fig. 6.1. The location of these overlapping regions varies slightly for each acquisition model. However, for the purpose of its localization in the image coordinate, this estimation is enough.

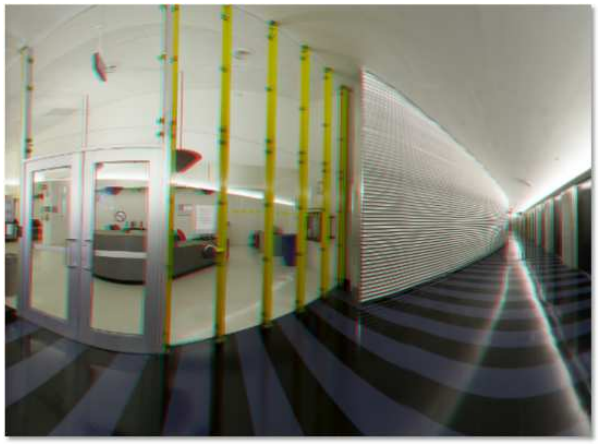
6.3.2 Blending regions

The horizontal stitching coordinate may be different from a fixed x_b when using any optimal cut algorithm [132, 133]. For instance, in a recently proposed method based on using graph cuts to find the optimal stitching position [141], the authors claim to produce consistent mosaics

of stereoscopic images. Their method is based on minimizing an energy function specifically designed to account for the depth continuity followed by a warping transformation to *smooth* disparity transitions between mosaics. Consequently, depending on the stitching technique used, x_b provides a reference for the stitching position search. The actual stitching position can be a point $\mathbf{p}_s = (x_s, y)$, where x_s has different values for each image row.

However, the analysis is done considering a constant stitching coordinate x_b and a blending region of size Δ_b around the stitching coordinate. In other words, the blending region is defined by $x_{j,i} \in [-x_b - \frac{\Delta_b}{2}, -x_b + \frac{\Delta_b}{2}]$ and $x_{j,i} \in [x_b - \frac{\Delta_b}{2}, x_b + \frac{\Delta_b}{2}]$, for all image rows $y_{j,i}$. This is sufficient to model the design parameters for the different configurations. The overlapping and blending regions in the image are shown in Fig. 6.2.





(c)

Figure 6.2: The overlapping regions of (a) the left eye sequence $(\mathbf{im}_{L,i}, \mathbf{im}_{L,i+1})$, and (b) the right eye sequence $(\mathbf{im}_{R,i}, \mathbf{im}_{R,i+1})$, and (c) stereo anaglyph (red-cyan) of the two images mosaicked.

6.4 Summary

In this chapter, I presented the formal method necessary to implement the omnistereoscopic rendering by mosaicking partially stereoscopic images of the scene. Among these formal methods, I described the calibrated techniques to match corresponding point coordinates from neighbour images. This is necessary to formalize the alignment of left and right sequences of images previous their stitching into the panoramic images $\mathbf{I_L}$ and $\mathbf{I_R}$. I also presented the calibrated approach to stereo registration applicable to each $\mathbf{im}_{j,i}$. I presented both calibrated methods to illustrate a robust approach applicable to multiple camera system. However, neither of these calibrated methods is suitable for the experimental set-up I used to emulate the four acquisition configurations.

One of my goals it to provide an analytical framework for the problem of omnistereoscopic acquisition that will help to develop a multiple camera system in the suggested guidelines. The robust methods for stereoscopic registration and aligning are a necessary part of this thesis since they are the most suitable alternatives for a multiple camera set-up.

I also defined the stitching and blending regions on each image for a given set of acquisition parameters.

In the next chapter, I am presenting the experiments in omnistereoscopic image acquisition and the building blocks of the mosaicking processing pipeline. The main purpose of implementing the mosaicking algorithms is to test the capabilities of each acquisition configuration in similar scenarios and under comparable system constraints.

Chapter 7

Experimental acquisition and rendering

In this section, I present an example from each acquisition configuration to illustrate the creation of $\mathbf{I}_S$ based on mosaicking. In these experiments, I used the same stereoscopic baseline to compare different configurations. In the case of *configuration 3*, I also included additional results from a different set-up, which consisted on a larger baseline and a where I applied different rendering method.

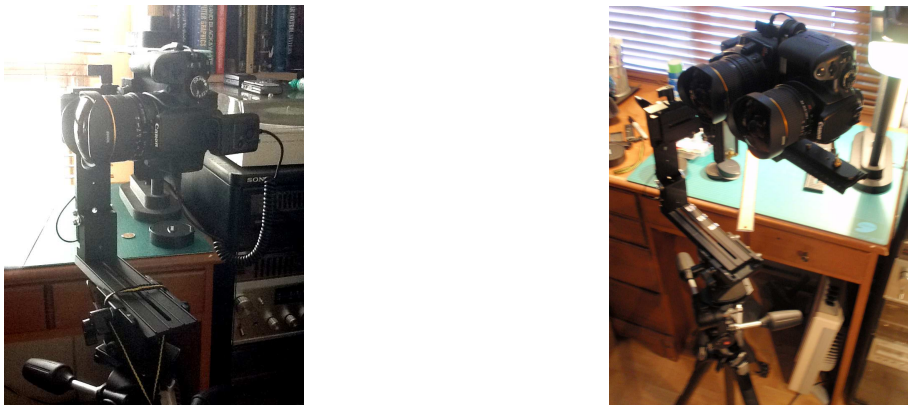
Partial results have been presented at the SPIE/IS&T Electronic Imaging conference in 2013 [18] and a companion website of this conference paper is available online [142].

7.1 Acquisition platform

Each acquisition configuration was emulated using a single or dual DSLR camera (Canon Rebel XTi 400D), which was mounted in portrait orientation on a panoramic head (Manfrotto 303SPH). In these experiments, I used a *fish-eye* lens Bower/Samyang model SLY-358C. This acquisition set-up is illustrated in Fig. 7.1 for the single and dual camera.

Each acquisition configuration determines a different spatial distribution of projective centers $(\mathbf{O}_{\mathbf{L},\mathbf{i}}, \mathbf{O}_{\mathbf{R},\mathbf{i}})$ as defined in Section 3.2. All configurations were emulated by manually positioning the nodal point of the single camera (Fig. 7.1-(a)) at predetermined locations.

Alternatively, when using the stereoscopic camera rig (Fig. 7.1-(b)), the stereoscopic image pair $\mathbf{im}_{j,i}$ was acquired in one snapshot for each orientation in azimuth (θ_i). Using a pair of cameras helped to acquire dynamic events, at least on each stereoscopic sample, but the physical dimensions of the cameras used imposed a large baseline which affected the mosaicking processes. The latter mode was used to emulate *configuration 3* using a larger baseline. An



(a) (b)

Figure 7.1: Test set-up used to emulate the proposed camera: (a) sequential acquisition using a single camera, and (b) rotating stereoscopic rig.

example of the manual acquisition to emulate *configuration 1* is illustrated in Fig. 7.2.

A constant number of six stereoscopic samples were acquired in each experiment. However, based on the 3D models of the scene presented in Section 5.2.3, the chosen value of N was too small for some indoor scenarios where the scene and the reference viewpoint were too close to each other, at least, in some panning directions in azimuth. This apparent disadvantage was offset by the fact that scenes were at larger distances than required to produce a consistent depth rendition in most of $\mathscr{D}$ around the reference viewpoints.

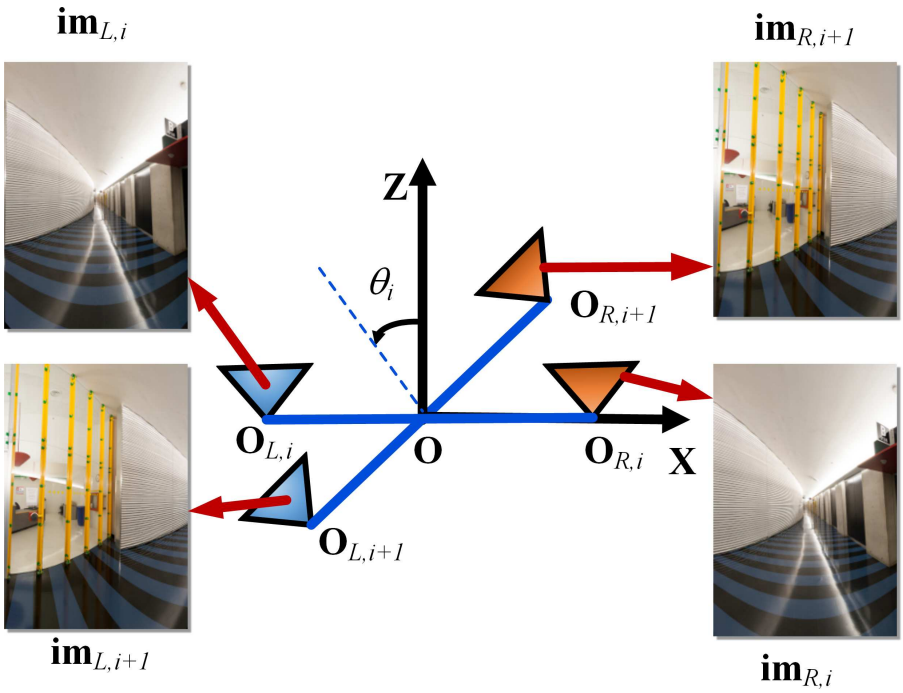


Figure 7.2: Example of acquisition of two adjacent stereoscopic snapshots of the scene using *configuration 1*: the acquisition of the stereoscopic pair of images $\mathbf{im}_{j,i}$ for θ_i and $\mathbf{im}_{j,i+1}$ where the angular increment between stereoscopic samples is $\Delta_\theta = \theta_i$.

7.2 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 the 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 mosaics. 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 interocular distance within the adult human population. A reduced baseline is convenient for mosaicking since less parallax between projective centers is introduced. In outdoor scenes, where an omnidirectional depth map $\mathscr{M}$ need to estimated from the stereoscopic samples, a larger baseline may be necessary. Unfortunately, a larger baseline implies more parallax between projection centers and more complex rendering techniques are needed to reduce the visible ghosting after stitching neighbour samples. The acquisition parameters used in all the experiments presented in this section are detailed in Table 7.1.

For large baselines, the acquisition *configuration 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 radial distance r_c.

Table 7.1: Acquisition parameters: Canon Rebel XTi 400D, Bower fish-eye lens (SLY-358C).

Parameters	**Value**
W_h	22.2 mm
a_r	1.5
f	9.3 mm
s	5.6 μm (3888×2592 pixels)
Δ_a	100°
b	32.5 mm
r_c	32.5 mm

7.3 Rendering implementation

The mosaicking of $\mathbf{im}_{j,i}$, although is not the main focus of this thesis, was necessary to test the viability of different acquisition configurations. For this reason, the computational efficiency of the implemented algorithms was not a critical factor, and so I used commercially available tools whenever relevant to implement part of the rendering pipeline. In addition, I developed essential rendering blocks using Matlab, which is an environment that facilitates the testing of the necessary algorithms, which was advantageous in terms of reducing the debugging time.

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 lens distortion and color correction were both done using *Adobe Lightroom 4.4* [143]. After these initial corrections, the image sets were stored in *tagged image file format* (TIFF) for mosaicking. In the case 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* [144]. The rest of the mosaicking blocks and additional processing were implemented using *Matlab R2012a* [145] using the *Computer Vision System* and *Image processing* toolboxes. Note that, when not explicitly mentioned, the algorithms described in the following sections were implemented by me in *Matlab*, whether completely or partially using specific

toolboxes included in the release or implemented by third parties. In the latter case, the links to the available source code online are provided.

The anaglyph version of the stereoscopic images presented in this document have been created using the Dubois anaglyph technique [146] using the *StereoPhotoMaker* authoring tool [147].

7.4 Mosaicking rendering pipeline

The block diagram in Fig. 7.3 illustrates the complete acquisition and rendering pipeline implemented. The implementation of each of these blocks is presented in the following sections.

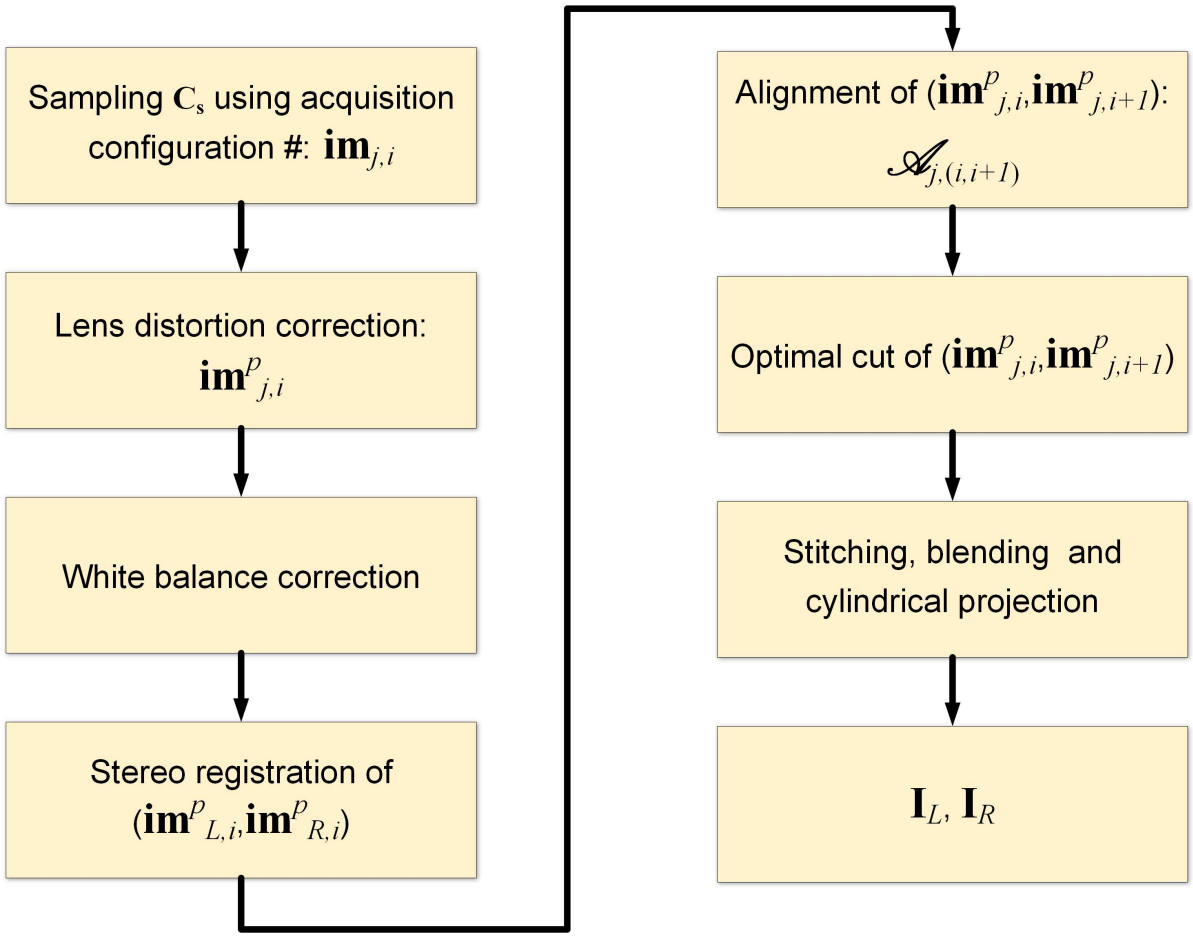
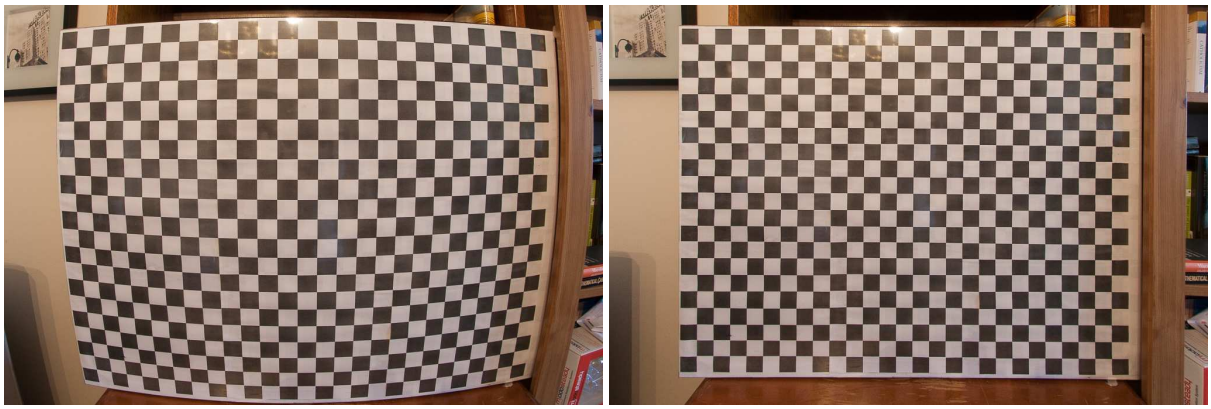


Figure 7.3: Omnistereoscopic rendering based on mosaicking partially overlapped stereoscopic snapshots of the scene.

7.5 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. Hence, the first step in the processing pipeline was to re-sample $\mathbf{im}_{j,i}$ into planar images $\mathbf{im}^{p}_{j,i}$. This correction requires the optical system characterization to estimate the image center and the distortion parameters according to a given radial distortion model [148].

Initially, I developed correction routines in *Matlab* using a radially spherical distortion model and calibration packages [149]. Fortunately, I found an alternative solution that led to better results. The *Adobe Lightroom 4.4* has a lens correction add-on that surpassed my own implementation in terms of quality of the corrected images. The calibration information for Samyang SLY-358C can be downloaded and added to the *Adobe Lightroom*'s lens corrector toolbox. As a test I corrected the distortion of the calibration patterns I was using to estimate the distortion parameters. The planar image after correcting the calibration patterns' distortion is shown in Fig. 7.4. The distortion correction results seemed appropriate for the intended application, so I applied the lens correction to each set of images.



Before distortion correction After distortion correction

Figure 7.4: Lens correction on the calibration pattern.

I saved a great deal of development time and I achieved greater accuracy to generate the planar set $\mathbf{im}^{p}_{j,i}$ by using this tool. An example showing the acquired raw image and the image after correction is shown in Fig. 7.5.

Acquired image | After correction | After cropping

Figure 7.5: Lens correction.

7.6 White balance correction

The next step is to correct white balance of the planar set of images $\mathbf{im}^{p}_{j,i}$. This processing was done manually using *Adobe Lightroom 4.4* [143] and the *Color Checker Passport* [150]. The latter provides a color calibration pattern to be used as color reference. An example is shown in Fig. 7.6. This color pattern was photographed at the beginning of each image acquisition for each viewpoint location.

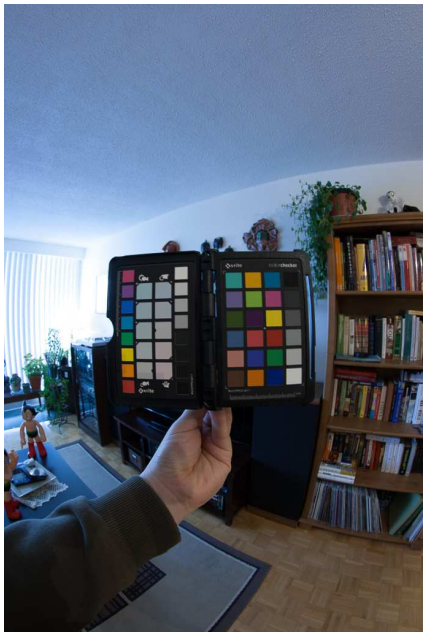


Figure 7.6: Color pattern acquired alongside each image set.

This color reference image provides a sample of the light on each scene to correct the white balance during post-processing as shown in Fig. 7.7. 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.

After lens distortion and white balance corrections, the images were cropped and stored as TIFF images to be mosaicked.

Acquired image

After white balancing and distortion correction

Figure 7.7: Color correction.

7.7 Stereo rectification

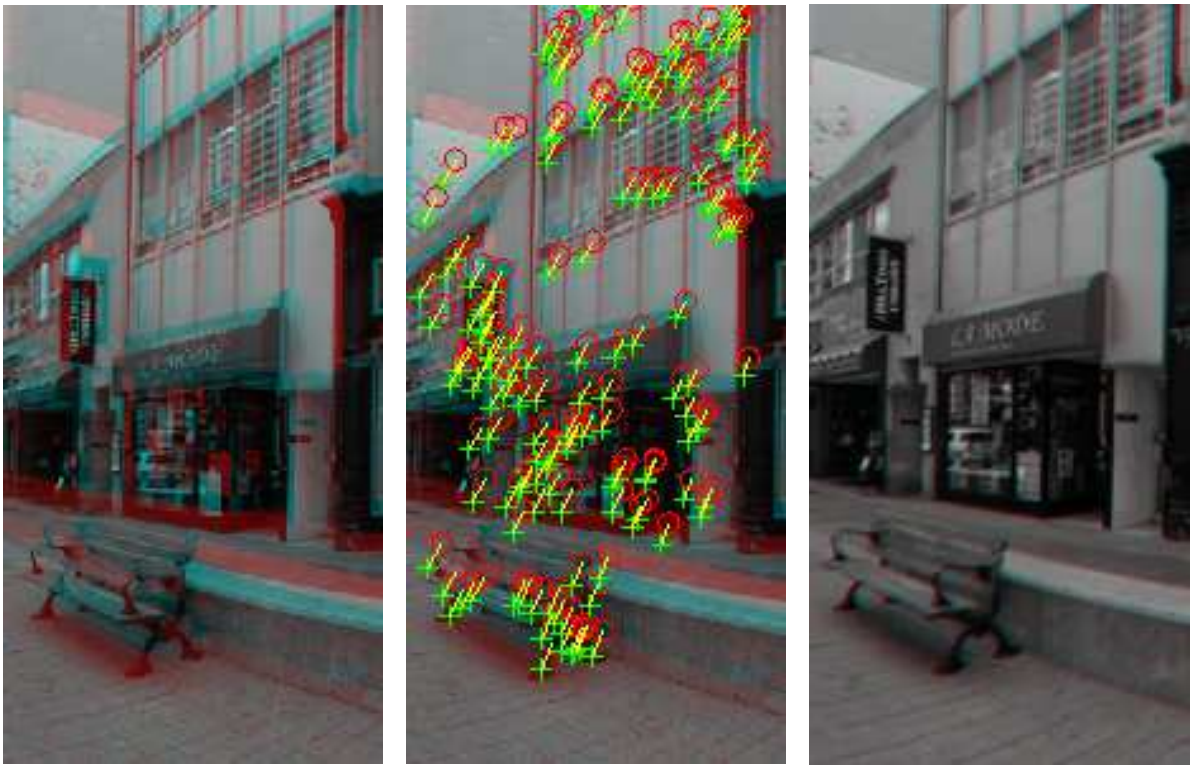
The stereo rectification of the pair of undistorted images $\mathbf{im}^{p}_{j,i}$ was implemented to correct any deviation from the ideal camera model with parallel optical axes. The image rectification method projects both images $\mathbf{im}^{p}_{j,i}$ onto the same image plane so that the vertical disparities $d_{v,i}$ defined over the corresponding coordinate points $\mathscr{K}_i$ (Section 5.1) are minimized. This is a method to enforce the ideal case of parallel optical axes between cameras by reducing a 2D correspondence problem to a 1D problem as explained in Section 6.2. In other words, the corresponding points will exhibit only horizontal disparities after the stereo rectification correction [136, 137].

The calibrated approach for stereo rectification requires the *intrinsic* and *extrinsic* camera parameters for cameras $\mathbf{\Omega_{j,i}}$, which is not possible in these experiments since I used a single camera or a camera rig to emulate each acquisition configuration. Hence, I implemented a feature-based method to estimate the necessary transformations to apply to each $\mathbf{im}_{j,i}$.

The uncalibrated approach for the stereoscopic rectification involves the estimation of a set of corresponding points $\mathscr{K}_i$ on $\mathbf{im}^{p}_{j,i}$ 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. Since the acquisition set-up was emulated using a single camera, the uncalibrated approach for the stereoscopic registration was implemented on *Matlab R2012a* using the routines from the *Computer Vision System* toolbox [151].

The implementation uses the *speeded up robust features* (SURF) algorithm [152, 153] to identify a set of corresponding point features $\mathscr{K}_i = \mathbf{p}_{j,i}$ for each stereoscopic pair. The recursive process eliminates feature points from the set whose sum of absolute distances are above a pre-determined distance threshold. On each recursion of the algorithm, the point sets $\mathscr{K}_i$ must satisfy the epipolar constraints, e.g., any point $\mathbf{p}_{L,i}$ must be on the epipolar line determined by $\mathbf{p}_{R,i}$. I used the *Matlab* implementation of the *random sample consensus* (RANSAC) algorithm [154, 155] to eliminate outliers candidate points using the epipolar constraint. On each iteration, a new projective matrix is estimated. After all outliers have been eliminated from the set $\mathscr{K}_i$, 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. 7.8-(a) shows the stereoscopic input $\mathbf{im}^{p}_{j,i}$ before registration. The corresponding point set $\mathscr{K}_i$ after a several recursions is shown in Fig. 7.8-(b). Finally, the result of the stereoscopic rectification is presented in Fig. 7.8-(c) after the projection of $\mathbf{im}_{L,i}$ and $\mathbf{im}_{R,i}$ on a common plane to enforce the parallel optical axes constraint.



(a) (b) (c)

Figure 7.8: Stereoscopic registration (red-cyan anaglyphs): (a) the overlapping area between images $\mathbf{im}_{L,0}$ and $\mathbf{im}_{R,0}$, (b) feature points calculated in the overlapping areas, and (c) red-cyan anaglyph of the overlapping area after alignment.

For the cases where not enough corresponding points were automatically detected by this SURF/RANSAC recursive method, I implemented an alternative stereoscopic rectification algorithm that relies on a reduced number of corresponding points manually selected. The reasons for not detecting enough samples with the automatic methods can be explained by areas that were insufficiently illuminated or where the lack and contrast in the images prevents to detect enough corresponding features. In these cases, the manual selection of a limited set of corresponding points was enough to implement a satisfactory rectification.

7.8 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. In these acquisition experiments, exposure and f-stop were manually selected at the beginning of the acquisition, therefore, color and luminance adjustments were not necessary in general. However, some experiments were done adjusting the photographic parameters in each gazing direction with the purpose of testing the developed color and luminance blending routines. Note that these routines were implemented in *Matlab*.

There have been recent applications developed for luminance blending in mobile devices. The attractiveness of these proposals resides in their low computational complexity, which can be translated into real-time applications in other platforms. Among them, the method proposed by Xiong and Pulli [156] is a simple but effective approach to achieve exposure equalization for mosaicking multiple overlapped images. Their method is based on calculating local compensation coefficients over the overlapped areas between $\mathbf{im}^p_{j,i}$ and $\mathbf{im}^p_{j,i+1}$. 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 neighbour images are close.

I implemented using *Matlab* the luminance homogenization technique of Xiong and Pulli. In order to test it, I acquired a set of images changing the exposure parameters. The results were satisfactory, so I incorporated this processing block in the rendering pipeline. An example of the mosaicking before and after the exposure compensation is shown in Fig. 7.9.

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.

Mosaicking result before luminance compensation

Mosaicking result after luminance compensation

Figure 7.9: Effect of the local and global color correction using Xiong and Pulli low complexity algorithm for color blending.

7.9 Mosaic alignment

The rendering by mosaicking relies on mapping of a point coordinate $\mathbf{p}_{j,i}$ from $\mathbf{im}^{p}_{j,i}$ 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 corresponding transformation between corresponding points $\mathscr{K}_{j,i}$ as defined in Section 6.1.1. This process is referred herein to as alignment.

The calibrated approach to alignment can be done by obtaining $\mathbf{H}_{j,(i,i+1)}$ that transforms the coordinate point $\mathbf{P}_{j,i}$ in $\Omega_{j,i}$ frame into $\mathbf{P}_{j,i+1}$ in $\Omega_{j,i+1}$ frame, and doing so for $j \in \{L, R\}$. This requires the pairwise calibration between cameras $\Omega_{j,i}$ and $\Omega_{j,i+1}$, for $j \in \{L, R\}$ and $i \in \{0, ..., N-1\}$. This correspondence transformation has been detailed in Section 6.2.1.

As in the case of stereoscopic rectification, the approach I used for the alignment of mosaics was uncalibrated. The corresponding point set $\mathscr{K}_{j,i}$ on the overlapped regions between neighbor images were estimated using SURF and RANSAC algorithms following the implementation described in Section 7.7. The $\mathscr{K}_{j,i}$ set 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 [151] 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 [131]. This coordinate transformation to align coordinates is referred herein by the operation $\mathscr{A}_{j,(i,i+1)}$ defined by Eq. 6.16.

In some scenarios, the overlapped regions between certain images were not suitable to produce a large number of corresponding points using the automatic feature detection mentioned above. Hence, I implemented an alternative method that relies on a reduced set of manually selected corresponding points. This reduced set is used to estimate the alignment operation $\mathscr{A}_{j,(i,i+1)}$.

7.10 Cylindrical projection

The central symmetry of all the acquisition configurations suggests to project $\mathbf{im}^p_{j,i}$ 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 [157]. Therefore, I implemented a mapping function to project points from different $\mathbf{im}^p_{j,i}$ onto a cylindrical surface.

The cylindrical canvas, which is illustrated in Fig. 7.10, has parameters Δ_θ for the panning azimuth and h for the gazing height. The 2D mapping of a pixel with coordinates $\mathbf{p}_{j,i} = (x_{j,i}, y_{j,i})$ in $\mathbf{im}^p_{j,i}$ onto cylindrical coordinates $\mathbf{p}^c_{j,i} = (x^c_{j,i}, y^c_{j,i})$ is given by

$$x^c_{j,i} = f\ \Delta_\theta = f\ \arctan\frac{x_{j,i}}{f}, \tag{7.1}$$

$$y^c_{j,i} = f\ h = f\ \frac{y_{j,i}}{\sqrt{x^2_{j,i} + f^2}}, \tag{7.2}$$

where the radius of the cylindrical surface was chosen to be f [158].

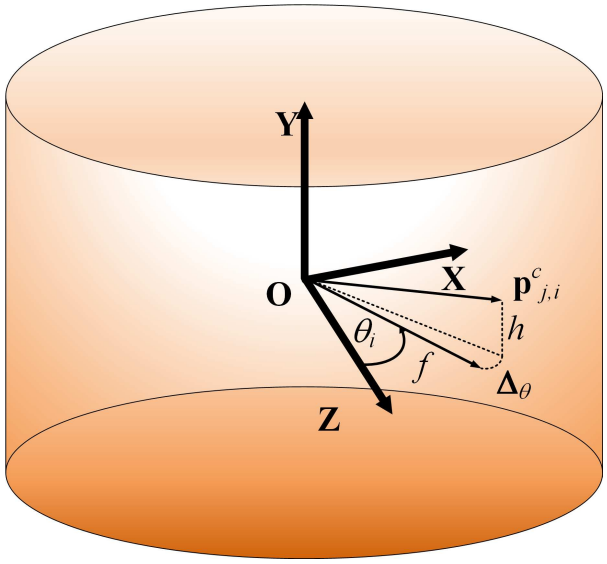


Figure 7.10: Cylindrical canvas.

Note that the cylindrical coordinates $\mathbf{p}^c_{j,i} = (x^c_{j,i}, y^c_{j,i})$ are defined for the projection of the planar image $\mathbf{im}^p_{j,i}$ 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 $\mathbf{p}^c_{j,i} = (0, 0)$, which corresponds to cylindrical coordinate $\theta = \theta_i$ and $h = 0$. This transformation is illustrated in Fig. 7.11.

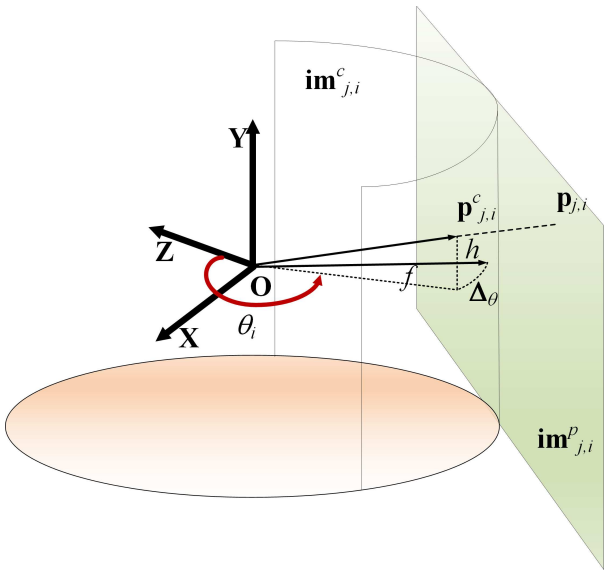


Figure 7.11: Mapping from a planar to a cylindrical surface.

All images $\mathbf{im}^p_{j,i}$ can be mapped onto the same cylindrical surface with center $\mathbf{O}$ using the transformation defined in Eq. 7.3 for any θ_i and Δ_θ deviation with respect to panning angle θ_i in azimuth.

The mapping from a planar to cylindrical coordinates can be expressed in a matrix form as

$$\mathbf{p}^c_{j,i} = \mathscr{C}_y \cdot \mathbf{p}^p_{j,i}, \tag{7.3}$$

where cylindrical projection is given by

$$\mathscr{C}_y = f \begin{pmatrix} \arctan \frac{x_{j,i}}{f} & 0 \\ 0 & \frac{y_{j,i}}{\sqrt{x^2_{j,i}+f^2}} \end{pmatrix}. \tag{7.4}$$

Conversely, the mapping from a cylindrical coordinates onto a planar surface can be done by

$$x_{j,i} = f \ \tan \Delta_\theta = f \ \frac{x^c_{j,i}}{f}, \tag{7.5}$$

$$y_{j,i} = y^c_{j,i} \sec \frac{x^c_{j,i}}{f}. \tag{7.6}$$

7.11 Stitching

Any coordinate point in the planar image that is not in the overlapping region between two neighbour samples is mapped into a cylindrical canvas using the coordinate transformation defined by Eq. 7.3.

A point with the stitching coordinate $\mathbf{p}^p_{j,i} = (x_b, y_{j,i})$ from $\mathbf{im}^p_{j,i}$ will be mapped into $\mathbf{p}_{j,i+1} = (x'_b, y_{j,i})$ in $\mathbf{im}^p_{j,i+1}$ using the correspondence transformation defined in Eq. 6.16. At that point, the corresponding point coordinate in $\mathbf{im}^p_{j,i+1}$ is projected onto the cylindrical canvas. The corresponding stitching coordinates in $\mathbf{im}^p_{j,i+1}$, which is $(x'_b, y_{j,i})$, can obtained using Eq. 6.16.

The alignment enables to continue the rendering by mapping $\mathbf{im}^p_{j,i+1}$ into the cylinder following the same cylindrical projection defined in Section 7.10, but this time 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 continue until the last image in the sequence $(\mathbf{im}^p_{j,N-1})$ is stitched with the first image $(\mathbf{im}^p_{j,0})$ 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 $j \in \{L, R\}$.

The stitching method defined in this section defines the mapping of all $\mathbf{im}_{j,i}$ into a single cylindrical canvas using the aligning process to match neighbor images. The method I implemented to smooth the visible transition between mosaicked images is explained next.

7.12 Blending

The correct blending between image pairs is necessary to eliminate visible ghosting in the final mosaicked panorama. The simplest approach is using a *linear blending* method where 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 [159].

In this context, the linear blending can be defined as function of the Euclidean distance between the point in the nearby the stitch coordinate x_b in $\mathbf{im}^p_{j,i}$ and its corresponding coordinate point x'_b in $\mathbf{im}^p_{j,i+1}$. The mix of the luminance values for the three color channels $\mathbf{im}^p_{j,i}$ and $\mathbf{im}^p_{j,i+1}$ is the one used on the cylindrical projection of a point in the neighbor of the stitching coordinate. This region around x_b defines the *blending region* between neighbor images. This section is about defining the luminance value (pixel) in the blending region. The coordinate mapping of these coordinate points in the blending region onto a cylindrical surface was explained in the previous section.

A weighting function to mix the luminance values of two neighbour planar images is defined as follows

$$\mathscr{Q}(\Delta_x) = \frac{2\,\Delta_x + W}{2\,W}, \tag{7.7}$$

where $\Delta_x \in [-W/2, W/2]$ is the horizontal deviation from the stitching position (x_b in $\mathbf{im}^p_{j,i}$ and x'_b in $\mathbf{im}^p_{j,i+1}$) and W is the blending width. The pixel value to map onto the cylindrical surface is then multiplied by this weighting function so

$$\mathbf{I}(x) = \mathscr{Q}(\Delta_x) \cdot \mathbf{im}^p_{j,i}(x, y_{j,i}) + (1 - \mathscr{Q}(\Delta_x)) \cdot \mathbf{im}^p_{j,i+1}(x', y'_{j,i+1}), \tag{7.8}$$

where $x = x_b + \Delta_x$, and $x'\ y'_{j,i+1}$ are defined as follows

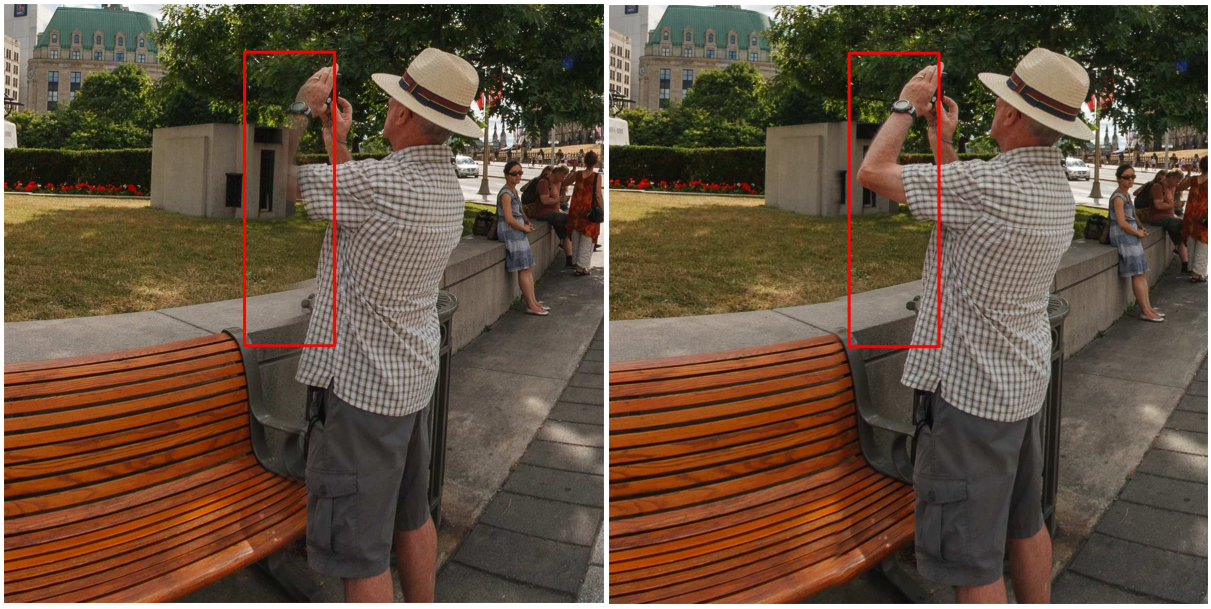
$$(x', y'_{j,i}) = \mathscr{A}_{j,(i,i+1)} \cdot (x_b + \Delta_x, y_{j,i}), \tag{7.9}$$

An alternative approach is using a multi-band decomposition, such as a Gaussian pyramidal decomposition [160]. The decomposition produces $\mathscr{N}$ increasingly blurred images, which are blended linearly. The reverse decomposition combines the new $\mathscr{N}$ blended images to produce the final blended image. This approach 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. The latter case is relevant for the acquisition of dynamic scenes and led to search for an optimal cut selection of the stitching point.

7.13 Optimal cut

The optimal cut method relies on estimate $\mathscr{A}_{j,(i,i+1)}$ 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.

The stitching and blending method I implemented relies on the *linear blending* applied line by line for a coordinate x_b defined by the optimal cut algorithm. This processing is applied independently to left and right images. In order to do this, I implemented the algorithm proposed by Ha et al. [133] 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 a real-time applications, i.e., video mosaics. An example of the image blending is presented in Fig. 7.12 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 has changed between the neighbor samples.



(a) (b)

Figure 7.12: 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.

The mosaicking implementation I developed is based on estimating the optimal cut coordinate x_b per image line $y_{j,i}$ and blending linearly the aligned neighbor images over a distance of five pixel widths ($W = 5\ s$).

7.14 Mosaicking examples

In this section, I present some of the results of testing the rendering pipeline described in this chapter over the images acquired by emulating the four configurations presented in Section 3.2. An explanation of the zero parallax adjustments done on each $\mathbf{I_S}$ is given in Appendix C.

The interactive version of the rendering examples presented next can be found in the companion website of this thesis [161].

7.14.1 Acquisition configuration 1

The first experimental result presented was obtained emulating the acquisition *configuration 1* with a single camera set-up illustrated in Fig. 7.1-(a). The baseline length and camera settings are in Table 7.1.

The original six images $\mathbf{im}_{j,i}$ are presented in Fig. 7.13. The first row of images corresponds to $j = L$ and the row below belongs to $j = R$.

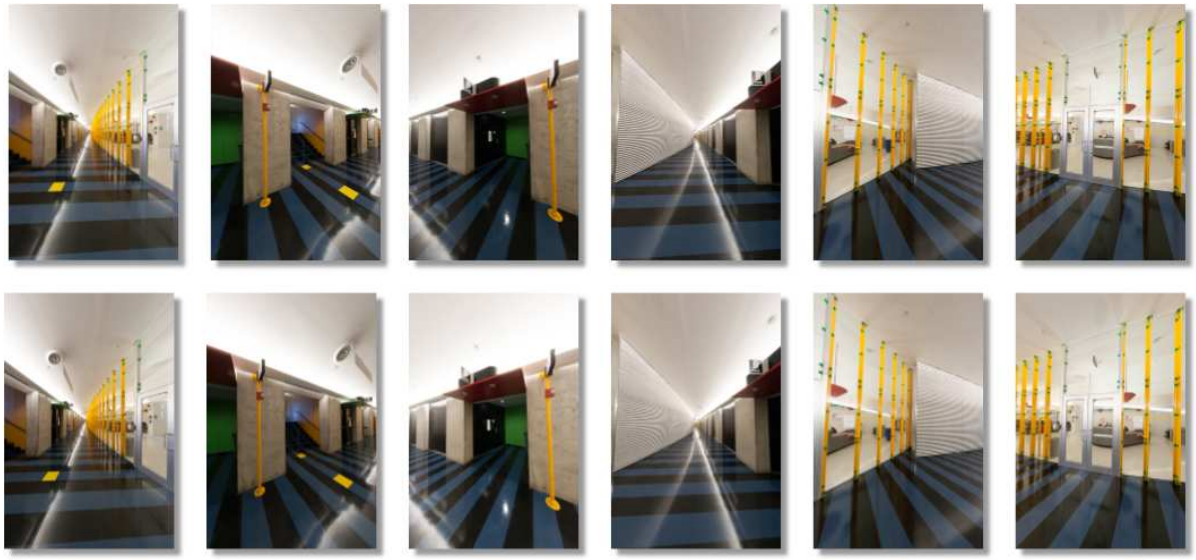


Figure 7.13: Source images $\mathbf{im}_{j,i}$, where $j = L$ (top row) and $j = R$ (bottom row) and $i \in \{0, ..., N-1\}$ (from left to right).

The result of the mosaicking is presented in Fig. 7.14 where the left-eye ($\mathbf{I_L}$) and right-eye ($\mathbf{I_R}$) cylindrical panoramas are shown. In Fig. 7.15, a red-cyan anaglyph of the omnistereoscopic image $\mathbf{I_S}$ is presented. Since the cylindrical projection does not convey the immersive experience provided by an interactive navigation based on producing stereoscopic planar projection in desired gazing directions, an interactive version of the same example is available online [161].

The interactive projections presented as supplementary information of this thesis were created with *Autopano Tour* [162], which is a software tour to display panoramic images. The application create the planar view with a limited FOV by projecting a panoramic image into a planar canvas based on the gazing direction. In this case, the anaglyph version of the cylindrical omnistereoscopic images were used as input for the navigable on-line version. This navigation principle has been explained in Section 5.2.3.

$\mathbf{I_L}$

$\mathbf{I_R}$

Figure 7.14: Example of mosaicking using the source images acquired from *configuration 1*.

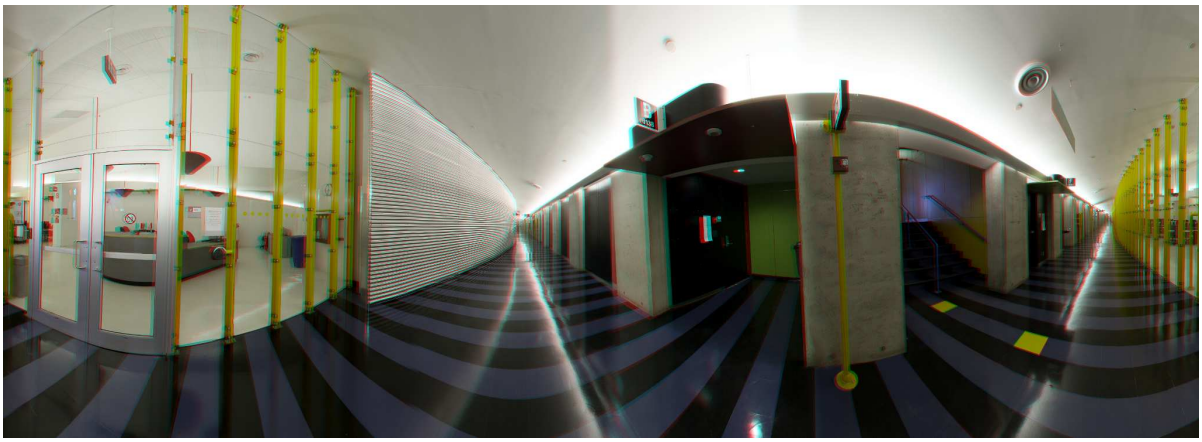


Figure 7.15: Red-Cyan anaglyph of $\mathbf{I_S}$ (*configuration 1*).

This example would fit the cylindrical scene model (Section 5.2.3) if all the 3D distribution were uniformly distributed in space around the camera. But in this case, the scene was closer than 2 m in two distinct directions, while it was much farther away in two other directions in azimuth. In this case, the chosen number of six stereoscopic samples was satisfactory to create a continuous horizontal disparity transition between mosaics. A better approach for the optimal cut based on the horizontal disparity in the blending regions between mosaics in presented in the next chapter.

Another point to remark about this experiment is that it requires manual selection of corresponding features in half of the stitching directions. This is because the alignment and stereo rectification implementation relies on the automatic identification of features, but in this case, the poor illumination and lack of distinguishable textures made the automatic identification of features difficult. A proper calibration of each pair in a multiple camera system will help to implement a rendering strategy based on the homography, which can make the rendering independent of the scene captured. This was not possible in this experiment since a single camera was manually located in different spatial locations, which prevented the use of a calibrated approach.

7.14.2 Acquisition configuration 2

This example refers to the *configuration 2* (Section 3.2, which was emulated using a single camera. In this case, the left camera was rotated around its nodal point while the right camera was shifted off-center to the right $b = 32.5$ mm. Since a short baseline was used in this experiment, it was possible to mosaic left and right views.

The mosaic was better for the images acquired with the central camera (Fig. 7.16-(a)) than for the off-centered camera (Fig. 7.16-(b)), even though their parallax was small. This can be explained by the fact that the images acquired by the centered camera ($\mathbf{\Omega_L}$) always share the same projective center for all panning angles. The latter facilitates the stitching in ideal conditions and therefore the ghosting can be eliminated in this set of images [131]. However, the set of images acquired by the off-centered cameras will always present some degree of parallax, even for small baselines, and this translates into some degree of ghosting after mosaicking them.

In order to illustrate this difference in the mosaicking, no optimal-cut was used in the stitching.

An example of the final omnistereoscopic image for this configuration is presented in

$\mathbf{I_L}$

$\mathbf{I_R}$

Figure 7.16: Example of mosaicking using the source images acquired from *configuration 2*.

Fig. 7.17. An interactive version of the same example can be found in the companion website of this thesis [161].

Figure 7.17: Red-Cyan anaglyph of $\mathbf{I_S}$ (*configuration 2*).

7.14.3 Acquisition configuration 3

The acquisition *configuration 3* (Section 3.2) accounts for the physical limitation to collocate spatially the projection centers (nodal points) of multiple cameras. In this case, the problem is related to the centrally located camera that in the acquisition *configuration 2* is used to produce the SVP panoramic image. In this example, the camera $\Omega_{L,i}$ was radially displaced a distance $r_c = 32.5$ mm, while the right camera $\Omega_{L,i}$ was displaced laterally $b = 32.5$ mm with respect to the left camera. Note that smaller radial distances are also possible, but the lenses FOV and number of samples have to be accordingly changed. An example of a multiple camera system using a larger N and narrower FOV can be seen in [35]. A stereoscopic pair of cylindrical panoramas created in an indoor environment using this acquisition configuration is presented in Fig. 7.18.

The final rendered $\mathbf{I_S}$ is illustrated in Fig. 7.19 and an interactive version of the same example can be visited online [161].

$\mathbf{I_L}$

$\mathbf{I_R}$

Figure 7.18: Example of mosaicking using the source images acquired from *configuration 3*.

Figure 7.19: Red-Cyan anaglyph of $\mathbf{I_S}$ (*configuration 3*).

In order to illustrate the effect of the parallax between projection centers, a simple stitching at a constant position x_b was used instead of the optimal-cut stitching correction. Hence, in certain directions, when the scene was too close to the camera, a visible ghosting appeared in the right panorama due to the parallax between neighbour images to mosaic. This is shown in Fig. 7.20-(a). However, for another direction, the stitching does not exhibit visible ghosting since the scene was further away from the camera (Fig. 7.20-(b)). The same stitching directions do not exhibit visible ghosting in $\mathbf{I_L}$, which have been rendered using images acquired by the radially displaced camera as seen in Fig. 7.20-(c) and (d).

The problems arising in stitching these images should not be interpreted as an indication of an irremediable flaw in this acquisition configuration. It only shows that the parallax between left and right cameras makes somewhat more challenging a ghost-free mosaicking, at least for scene close-ups. However, a different rendering than mosaicking may be suitable, if not desirable, for this particular configuration.

The results of this acquisition configuration and the mosaicked rendering proposed in this section have been presented in the Stereoscopic Displays and Applications conference at the SPIE Electronic Imaging 2013 [18]. More interactive examples of omnistereoscopic panoramas created with this combined method can be visited online [142].

This particular configuration enables a different approach to rendering: the dynamic 2D-to-3D synthesis of stereoscopic views of the scene.

Rendering based on synthetic stereo

The acquisition *configuration 3* can be used to generate a monoscopic SVP panorama $\mathbf{I_L}$ using the images collected by the radially central camera, in this experiment $\Omega_{L,i}$, and to estimate an omnidirectional depth map $\mathscr{M}$, based on the images acquired by cameras $\Omega_{R,i}$.

This rendering approach aims to create $\mathbf{I_L}$ by mosaicking images acquired by the subset of cameras $\Omega_{L,i}$. This can be done using the mosaicking pipeline described in Section 7.4. The stereoscopic counterpart $\mathbf{I_R}$ are rendered dynamically for each $\mathscr{D}$ by warping the left-eye image according to the information extracted from the depth map $\mathscr{M}$. A possible method to deal with the occluded regions after the pixel displacement is using texture inpainting or other hole-filling techniques[163, 62].

Unlike the mosaicking method, which in this experiment relies on the *microstereopsis* effect arising from using narrow baselines, a better estimation of $\mathscr{M}$ requires wider baselines to improve the resolution of the depth estimation. Hence, the proposed mosaicking pipeline has

to account for larger baselines to produce artifact-free stitching. This rendering can be done exploiting $\mathscr{M}$ as proposed in Section 5.2.2.

The rendering method has been implemented as a proof of concept only.

Experimental set-up

The camera set-up used in this experiment was modified to include a second identical camera and lens. This set-up is illustrated in Fig. 7.1-(b). The baseline was set to $b = 90$ mm and the radial displacement $r_c = b$. As in the previous experiments, the camera rig was manually positioned at different panning angles θ_i, for $i \in \{0, ..., 5\}$, to acquire six partially overlapped stereoscopic samples $\mathbf{im}_{j,i}$.

Optical flow

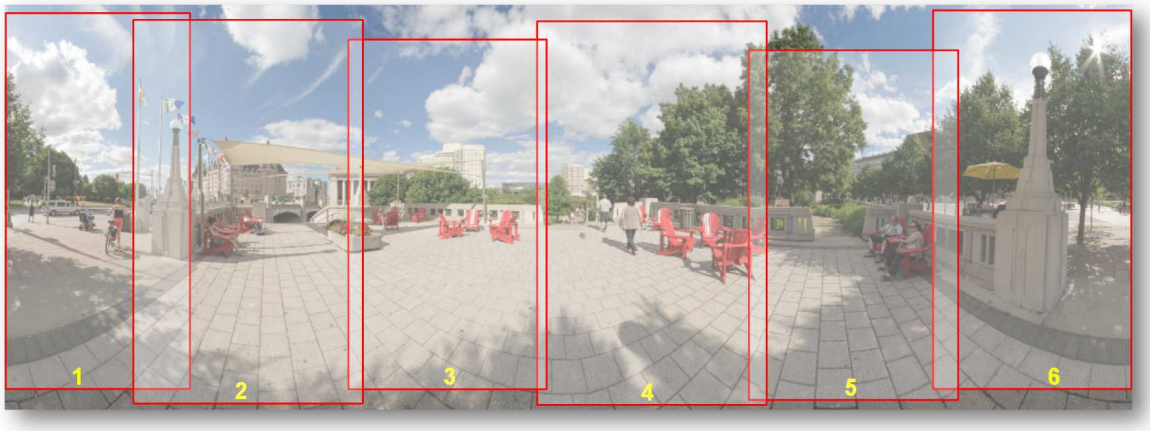
After the stereoscopic registration of each planar image (Section 7.7), the dense optical flow was estimated for each image pair $\mathbf{im}^p_{j,i}$. In order to estimate the dense optical flow, I used the *total variation L1 regularization* (TV-L1) algorithm for convex optimization proposed by Chambolle and Pock in [164]. The Matlab implementation available online [165] was used in this rendering experiment.

A depth-based warping transformation $\mathscr{W} : \mathbb{R}^2 \rightarrow \mathbb{R}$ was calculated based on the horizontal disparity desired between left and right view. The undesired vertical disparities, which pose a visual comfort problem for the binocular visualization, are not a problem using this method since the novel view is created by shifting horizontally the pixels of the source image. The occluded areas were filled using a pixel duplication technique. The warping transformation was implemented in *Matlab* while the depth-based warping based on $\mathscr{W}$ was done using the image warping tool of *GIMP* [144] to facilitate the processing. An example of this rendering approach is presented in Fig. 7.22.

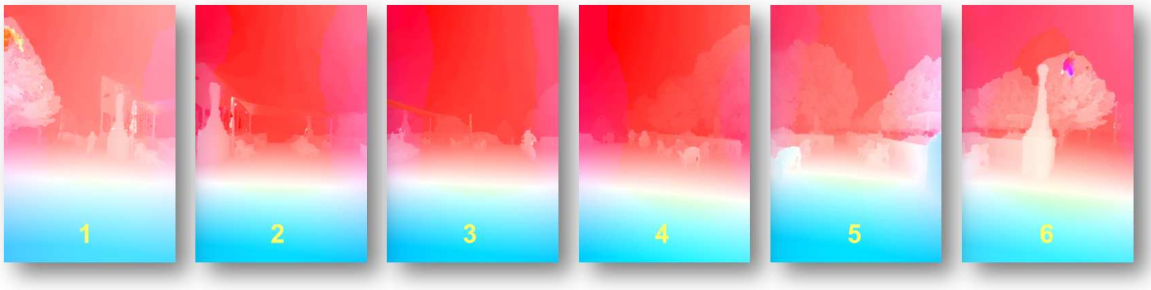
(a) (b)

(c) (d)

Figure 7.20: Mosaicking errors: (a)-(b) on stitching region extracted from $\mathbf{I_R}$ and (c)-(d) on the same stitching direction but extracted from $\mathbf{I_L}$ *configuration 3*.



Targeted views of the omnistereoscopic



Colour coded depth-based segmentation of the above regions

Figure 7.21: An example of the rendering of stereoscopic views of the scene at different $\mathscr{D}$ based on a 2d-to-3d conversion of the monoscopic views extracted from $\mathbf{I_S}$ (*configuration 3*).

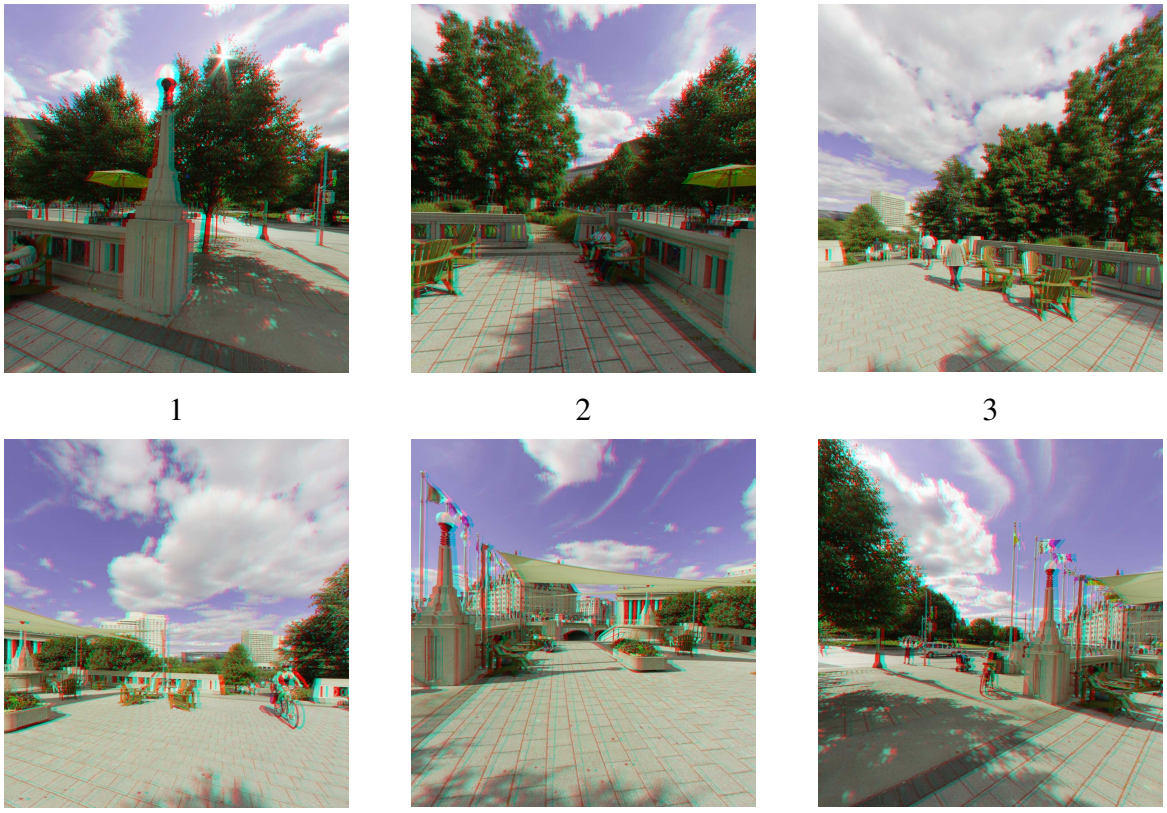


Figure 7.22: Red-cyan anaglyphs of the stereoscopic views synthesized for each target region.

An example of applying the horizontal displacement defined by $\mathscr{W}$ on different target regions of the image is presented in Fig. 7.23.

Figure 7.23: The result of applying the transformation $\mathscr{W}$ to different image regions: the left image in each pair is the original image extracted from $\mathbf{I_S}$ and their right side counterpart is the synthesized right-eye image.

This implementation was done with the sole purpose of showing the potential of the 2D-to-3D conversion on the images acquired with the *configuration 3*. However, the same approach can be applied to any of the other camera configurations using a larger baseline.

Despite the potential of this approach, there are still many aspects to be improved. First, a technique to fuse the depth map estimated from individual image pairs must be implemented. After this data fusion, a complete omnidirectional depth map $\mathbf{M}$ can be estimated. Another aspect to improve is the optical flow (motion field) estimation, which is an ongoing research topic. The incorrect segmentation of the images extracted from $\mathbf{I_S}$ is the source of visual artifact after applying $\mathscr{W}$. This warping error can be seen in the *upper-left* and *bottom-right* image pairs in Fig. 7.23. Despite these limitations, this acquisition and rendering technique is a promising alternative to mosaicking that can facilitate the stereoscopic exploration of the scene [2].

7.14.4 Acquisition configuration 4

The last example is the result of mosaicking six partially overlapped stereoscopic images, which have been acquired using the acquisition *configuration 4*. To keep consistency with the other experiments, the baseline and radial displacements were $b = 32.5$ mm and $r_c = b$. The left and right cylindrical panoramas corresponding to $(\mathbf{I_L}, \mathbf{I_R})$ are presented in Fig. 7.24 while

a red-cyan anaglyph of the omnistereoscopic image is presented in Fig. 7.25. An interactive version of this example can be seen online [161].

$\mathbf{I_L}$

$\mathbf{I_R}$

Figure 7.24: Example of mosaicking using the source images acquired from *configuration 4*.

Figure 7.25: Red-Cyan anaglyph of $\mathbf{I_S}$ (*configuration 4*).

7.15 Summary

In this chapter, I presented the experimental set-up I used to emulate the four acquisition configurations introduced in Chapter 3. Furthermore, I also presented a rendering implementation and I described the rationale behind each building block.

Even though the rendering by mosaicking is not a core element of this thesis, implementing the rendering by mosaicking is necessary to test the capability of different acquisition configuration to render $\mathbf{I_S}$ by the mosaicking method. Furthermore, the commercially available applications for stitching images were not conceived for mosaicking stereoscopic images, which requires a coordinated processing of left and right source images. For this reason, the rendering pipeline I propose in this chapter can be considered a novel contribution of this thesis.

The fact that these examples were created by emulating the four configurations of the acquisition model using a single camera does not disqualify them in terms of acquiring dynamic scenes. These examples show that the rendering by mosaicking stereoscopic snapshots of the scene can produce a satisfactory and continuous illusion of depth around the acquisition point. As a matter of fact, in Section 7.14.3, a dual camera set-up shows examples of piecewise dynamic scenes. It is immediate the extrapolation of these results to a full multiple camera set-up capable to acquire the sparse set of stereoscopic images simultaneously. Hence, any of these multiple camera configurations combined with a rendering pipeline as the exemplified in this chapter is suitable to produce direction-independent stereoscopic imagery of dynamic scenes.

One of the contributions presented in this chapter is the test of acquisition *configuration 3* (Section 7.14.3). This camera configuration is novel. As explained in Section 3.2, this configuration is particularly interesting for the synthesis of stereoscopic imagery, which is another novel idea for omnistereoscopic rendering problem that has intrinsic advantages for the navigation and visual comfort.

In the next chapter, I elaborate on the problem of depth continuity for rendering by mosaicking. In Chapter 5, I introduced this problem with a novel approach that relies on determining the usable eccentricity of each $\mathbf{im}_{j,i}$ based on the limits of the human depth resolution. The framework to compare the different configurations of the acquisition model is presented next.

Chapter 8

Omnistereoscopic field-of-view and disparities

One important requirement for any omnistereoscopic imaging system is to provide a continuous and consistent depth illusion around the reference viewpoint in all directions. For instance, acquiring multiple stereoscopic snapshots of the scene for rendering by mosaicking will produce a correct depth illusion only at the center of each mosaicked stereoscopic image, which coincides with the θ_i orientation of the stereoscopic camera at the moment of acquisition. The reproduced stereoscopic view in any intermediate gazing direction between θ_i and θ_{i+1} is subject to a depth distortion [52, 58].

In Chapter 5, I propose a method to estimate the eccentricity of a stereoscopic image that is usable for direct mosaicking based on the human resolution thresholds. I made this estimation with the help of two simplified 3D-models of the scene. In addition, I develop the projection equations to evaluate the disparity that occurs when mosaicking partially overlapped snapshots. Moreover, I propose a method that compares the difference in horizontal disparities in the overlapped areas between neighbor stereoscopic samples to constrain the minimum distance between the camera and the scene to produce a continuous and artifact-free illusion of depth around any viewpoint. Finally, the analysis is extended to model the vertical disparities in the blending regions.

8.1 The omnistereoscopic FOV

The FOV of a single camera $\Omega_{j,i}$ is defined by the all possible rays of light that pass through the projection centers $\mathbf{O}_{i,j}$ which simultaneously intersect its image plane at a distance $Z_{j,i} = f$. The FOV defines the region of the space in front of camera that can be acquired. This definition is valid for the projective approach used in our acquisition model. However, the minimum distance to the scene can be several times f in front of the camera using real lenses. In Fig. 8.1, I present an example of the different visibility scenarios using various locations in the **XZ**-plane. The point $\mathbf{P_1}$ in this example is in the FOV of camera $\Omega_{R,i}$ only.

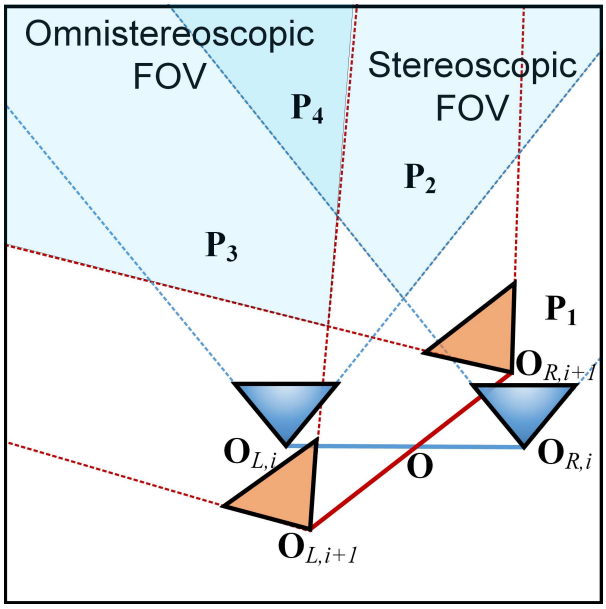


Figure 8.1: The continuous depth perception for all gazing directions.

Also in Fig 8.1, the point $\mathbf{P_2}$ is in the *stereoscopic FOV* of the stereoscopic camera $(\Omega_{L,i}, \Omega_{R,i})$. This region of space is defined by the intersection of the FOV of each camera in the stereoscopic pair. The same point $\mathbf{P_2}$ is in the FOV of camera $\Omega_{R,i+1}$, but not in the *stereoscopic FOV* of $(\Omega_{L,i+1}, \Omega_{R,i+1})$. Conversely, a third point $\mathbf{P_3}$, which is at the same distance from $\mathbf{O}$ than $\mathbf{P_2}$, is located in the *stereoscopic FOV* of $(\Omega_{L,i+1}, \Omega_{R,i+1})$, but is not in the *stereoscopic FOV* of $(\Omega_{L,i}, \Omega_{R,i})$.

The distance from the camera to the intersection of the *stereoscopic FOV* of two neighbour stereoscopic cameras (i and $i+1$) defines the *omnistereoscopic FOV*. Any point in the scene at this distance from the omnistereoscopic camera can potentially be imaged in $\mathbf{I_S}$. In this

example, $\mathbf{P_4}$ belongs to the intersection of two neighbour stereoscopic FOVs, which defines the *overlapping region* in space. A point in this overlapping region will be imaged simultaneously by cameras $(\Omega_{L,i}, \Omega_{R,i})$ and $(\Omega_{L,i+1}, \Omega_{R,i+1})$. Any point is the overlapping region can be used for stitching neighbour stereoscopic images.

The stereoscopic FOV of the panoramic camera does not define the minimum distance between the omnistereoscopic camera and the scene. In some camera configurations, a point in the overlapping of *stereoscopic FOV*s may be registered with different horizontal disparities in neighbour stereoscopic samples. This is especially true in asymmetrical acquisition *configurations 2* and *3*. This affects the perceived depth illusion after mosaicking the partial images.

However, a continuous depth illusion can be maintained as long as the difference between horizontal disparities is below the human threshold for depth perception (stereos acuity) around the stitching position, or at least, below the sub-pixel threshold $\epsilon = s$ (Section 5.1.6).

The perceived depth is directly related to the horizontal disparities in each stereoscopic sample. Since the difference between the horizontal disparities of the same scene in different samples decreases with the distance between the camera and the scene, there is a minimum tolerable distance for each acquisition model after which a difference in the registered depth is below a desired perceptual threshold.

The *omnistereoscopic FOV* is the spherical surface centered at $\mathbf{O}$, with a radius r_{min}, that marks the minimum distance from the omnistereoscopic camera after which a perceptually continuous depth illusion around the acquisition point can be reproduced. Any point $\mathbf{P_W}$ located at a distance $\|\mathbf{P_W}\| \geq r_{min}$ can be imaged whether by two (around θ_i) or by four cameras (around $\theta = \theta_i + \Delta_\theta/2$), maintaining a consistent illusion of depth between stereoscopic samples. The *stereoscopic FOV* of the panoramic camera and the *omnistereoscopic FOV* limits in the $\mathbf{XZ}$-plane are shown in Fig. 8.1.

8.2 The minimum distance to the scene

The minimum distance from the camera to guarantee a consistent illusion of depth requires defining the closed form of the horizontal disparity equations. The projective equations developed in Section 3.3.1 can be used to obtain the horizontal disparity error, which is especially important in the overlapping regions between neighbor stereoscopic images.

The vertical disparities are also of interest in the mosaicking. Any amount of vertical disparities will produce eye stress during the binocular visualization. In order to explain the appearance of vertical disparities in different regions of the omnistereoscopic image after mosaicking, we need to obtain the closed form of the vertical disparity error.

8.2.1 Horizontal disparity error

The horizontal disparity in the vicinity of the horizontal stitching coordinate $x_{j,i} = x_b$ remains below a tolerable error. In order to measure the horizontal disparity difference in the blending regions, the horizontal disparity error e_h defined in Sections 5.1.4-5.1.5 will be used. The case we are interested is when $\mathbf{P_W}$ is imaged onto $\mathbf{im}_{j,i}$ and $\mathbf{im}_{j,i+1}$, producing the pairs of corresponding points $\mathscr{K}_i$ and $\mathscr{K}'_{i+1}$, respectively.

In the case in which we are interested, at least one of the points in the set $\mathscr{K}_i$ is mapped (aligned) with another point in the set $\mathscr{K}'_{i+1}$. In this context, since we are referring to projections of the same scene into different stereoscopic images and with different horizontal disparities, the function e_h (Eq. 5.19) represents the *horizontal disparity error*.

The region of stitching and blending is critical since it is there where the artifacts caused by the parallax between individual cameras $(\Omega_{j,i}, \Omega_{j,i+1})$ will appear. It is also the most critical area in terms of defining the *omnistereoscopic FOV*. Hence, it is relevant to estimate what is the closest distance between camera and scene to guarantee the continuity of the horizontal disparity among mosaicked stereoscopic snapshots.

8.2.2 Horizontal disparity equations

The coordinates of the scene point$\mathbf{P_W}$ after applying the projective transformation defined in Eq.6.9 6.10 are

$$x_{L,i} = -f\frac{\cos\theta_i\,(X_w - t_1) + \sin\theta_i\,(Z_w - t_3)}{-\sin\theta_i\,(X_w - t_1) + \cos\theta_i\,(Z_w - t_3)} = -f\frac{X_{L,i}}{Z_{L,i}}, \tag{8.1}$$

$$x_{R,i} = -f\frac{\cos\theta_i\,(X_w - t_4) + \sin\theta_i\,(Z_w - t_6)}{-\sin\theta_i\,(X_w - t_4) + \cos\theta_i\,(Z_w - t_6)} = -f\frac{X_{R,i}}{Z_{R,i}}, \tag{8.2}$$

where $(X_{j,i}, Y_{j,i}, Z_{j,i})^T$ are the coordinates of a point $\mathbf{P_W}$ in each camera frame ($j \in \{L, R\}$ and $i \in \{0, ..., N-1\}$), and t_k ($k \in \{1, 3, 4, 6\}$) are components of the translation vectors defined in Eq. 3.6-3.12.

The image plane coordinates ($x_{L,i}$ and $x_{R,i}$) can be used to expand $d_{h,i}$ and $d_{h,i+1}$ (Eq. 5.3) as

$$d_{h,i} = -f\left(\frac{X_{R,i}\,Z_{L,i} - X_{L,i}\,Z_{R,i}}{Z_{R,i}\,Z_{L,i}}\right), \tag{8.3}$$

$$d_{h,i+1} = -f\left(\frac{X_{R,i+1}\,Z_{L,i+1} - X_{L,i+1}\,Z_{R,i+1}}{Z_{R,i+1}\,Z_{L,i+1}}\right). \tag{8.4}$$

Note that, since $Z_{R,i} = Z_{L,i} = Z_i$ and $X_{R,i} = X_{L,i} + b$ holds for all i, the disparities in Eq. 8.3 can be simplified to

$$d_{h,i} = -f\,b\left(\frac{1}{Z_i}\right), \tag{8.5}$$

$$d_{h,i+1} = -f\,b\left(\frac{1}{Z_{i+1}}\right). \tag{8.6}$$

8.2.3 Closed form of the horizontal disparity error

The closed form of e_h is

$$\begin{aligned} e_h &= f\left|\frac{X_{R,i+1}\,Z_{L,i+1} - X_{L,i+1}\,Z_{R,i+1}}{Z_{R,i+1}\,Z_{L,i+1}} - \frac{X_{R,i}\,Z_{L,i} - X_{L,i}\,Z_{R,i}}{Z_{R,i}\,Z_{L,i}}\right|, \\ &= f\,b\left|\frac{1}{Z_{i+1}} - \frac{1}{Z_i}\right|, \end{aligned} \tag{8.7}$$

As defined, e_h is a function of $\mathbf{P_W}$ after being transformed into each camera frame of coordinates.

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

$$Z_i = Z_w - t_{R,3}, \tag{8.8}$$

$$Z_{i+1} = -\sin\Delta_\theta\,(X_w - t_{R,1}) + \cos\Delta_\theta\,(Z_w - t_{R,3}), \tag{8.9}$$

where $t_{R,1}$ and $t_{R,3}$ can be obtained from Eq. 3.6-3.12 by replacing θ_{i+1} by Δ_θ.

The farther the scene is from the omnistereoscopic camera reference $\mathbf{O}$, the smaller are $d_{h,i}$ and $d_{h,i+1}$ since they are both monotonically decreasing functions of the relative distance to each stereoscopic camera for distances larger than r_{min}. It follows that there is a distance from the omnistereoscopic camera after which $e_h \leq \varepsilon$ consistently for all $\mathscr{D}$.

8.2.4 Finding the minimum distance to the scene

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

This approach requires only two consecutively acquired stereoscopic images, e.g, the samples $i = 0$ $(\theta_0 = 0^\circ)$ and $i = 1$ $(\theta_1 = \Delta_\theta)$. The search can be restricted to the $\mathbf{XZ}$-plane for simplicity and without losing generality. Hence, I am going to assume that the point $\mathbf{P_W}$ is located on the $\mathbf{XZ}$-plane. This is illustrated in Fig. 8.2.

First, I propose to define a ray between the point $\mathbf{P_W}$ and $\mathbf{O}_{R,0}$. This ray $(\mathbf{L}_{R,0})$ intersects the image plane of camera $\Omega_{R,0}$ in the coordinates $\mathbf{p}_{R,0} = (-x_b, 0)^T$ as shown in Fig. 8.2. The ray $\mathbf{L}_{R,0}$ is modeled as

$$\mathbf{L}_{R,0} = \mathbf{a_0} + t_r\,(\mathbf{a_1} - \mathbf{a_0}), \tag{8.10}$$

where $\mathbf{a_0}$ is the location of $\mathbf{O}_{R,0}$, $\mathbf{a_1}$ is the intersection of the ray with the image plane of camera $\Omega_{R,0}$, both expressed in term of the global reference frame $\mathbf{XYZ}$, and $t_r \in \mathbb{R}$, where $t_r \geq 1$. The parameters for $\mathbf{L_{R,i}}$ are given in Table 8.1 for each camera model.

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

The analysis starts by defining a point $\mathbf{P_0} \in \rho\mathbf{L}_{R,0}$ located inside the intersection between *stereoscopic FOV* of cameras $(\Omega_{L,0}, \Omega R, 0)$ and $(\Omega_{L,1}, \Omega R, 1)$, i.e., the *stereoscopic FOV* of

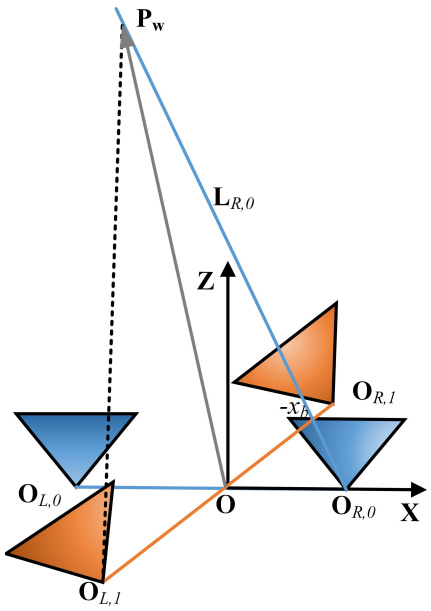


Figure 8.2: Ray tracing method to find the minimum distance r_{min} (*configuration 1*): the ray passing through $\mathbf{O_{R,0}}$ and x_b defines all the points on the $\mathbf{XZ}$-plane to find the minimum distance after which $e_h \leq \varepsilon$.

the panoramic camera (Section 8.1). This can be granted by calculating the projection of $\mathbf{P_0}$ in the other three image planes. This projections must satisfy: $x_{j,i} \in [-W_h/2, W_h/2]$.

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

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

A comparison between acquisition models in terms of the achievable minimum distance to the scene for a consistent depth rendition is presented in Chapter 9.

Table 8.1: Parameters to define $\mathbf{L}_{R,0}$.

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

8.3 The vertical disparities

There are two situations where vertical disparities need to be modeled. The first case involves the vertical disparities exhibited when a point $\mathbf{P_W}$ in the *stereoscopic FOV* of a stereoscopic camera pair produces a pair of corresponding points $\mathbf{K}_i$ in $\mathbf{im}_{j,i}$. The other situation happens when the point $\mathbf{P_W}$ in the *omnistereoscopic FOV* between two consecutive stereoscopic images, produces corresponding points sets $\mathbf{K}_i$ and $\mathbf{K'}_{i+1}$ in $\mathbf{im}_{j,i}$ and $\mathbf{im}_{j,i+1}$, respectively.

For points located in the *stereoscopic FOV*, the vertical disparities within $\mathbf{im}_{j,i}$ are negligible for any configuration of the acquisition model. However, in a real-world scenario, the optical axes of any pairs of cameras would not be parallel. But, even in that case, the vertical disparities can be eliminated using stereoscopic registration techniques such as the one described in Section 7.7.

For points in the *omnistereoscopic FOV*, the parallax between cameras will unavoidably produce vertical disparities between corresponding points in $\mathbf{im}_{j,i}$ and $\mathbf{im}_{j,i+1}$. This situation leads to stitching artifacts after mosaicking these stereoscopic images. The magnitude of these vertical disparities depends of the elevation component of $\mathscr{D}$ as I will show in this section.

8.3.1 Vertical disparity equations

The coordinates of a point $\mathbf{P_W}$ after applying the projective transformation defined in Eq. 3.18-3.19 are

$$y_{L,i} = -f\frac{Y_w}{-\sin\theta_i\,(X_w - t_1) + \cos\theta_i\,(Z_w - t_3)} = -f\frac{Y_{L,i}}{Z_{L,i}}, \tag{8.11}$$

$$y_{R,i} = -f\frac{Y_w}{-\sin\theta_i\,(X_w - t_4) + \cos\theta_i\,(Z_w - t_6)} = -f\frac{Y_{R,i}}{Z_{R,i}}. \tag{8.12}$$

where $(X_{j,i}, Y_{j,i}, Z_{j,i})^T$ are the coordinates of a point $\mathbf{P_W}$ in each camera frame and t_k ($k \in \{1, 3, 4, 6\}$) are components of the translation vectors defined in Eq. 3.6-3.12. Since the $\mathbf{XZ}$-plane contains all the projection centers and optical axes, it is valid to say that $Y_{j,i} = Y_W$ for all the cameras (Eq. 8.11-8.12).

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

$$d_{v,i} = y_{R,i} - y_{L,i} = -f\ Y_W \left(\frac{Z_{L,i} - Z_{R,i}}{Z_{R,i}\ Z_{L,i}} \right). \tag{8.13}$$

8.3.2 Vertical disparities outside the blending region

Any point $\mathbf{P_W}$ in the *stereoscopic FOV* of $(\mathbf{im}_{L,i}, \mathbf{im}_{R,i})$ will have projections $Z_{L,i} = Z_{R,i}$. Therefore, $d_{v,i} = 0$ for all Y_W (Eq. 8.13). This is valid outside the overlapping regions of every stereoscopic image to mosaic (Section 6.3.1).

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

8.3.3 Vertical disparities within the blending region

The parallax between projection centers produces unwanted vertical disparities in the overlapping regions. The vertical disparities between neighbor images $(\mathbf{im}_{j,i}, \mathbf{im}_{j,i+1})$, which is defined as the vertical component of the disparity between a coordinate point $\mathscr{K}_{j,i} = (\mathbf{p}_{j,i}, \mathbf{p}_{j,i+1})$, for $j = L$ or $j = R$ (Eq. 6.1.1), is given by

$$\begin{aligned} dv_{j,i}(\mathscr{K}_{j,i}) &= y_{j,i+1} - y_{j,i} \\ &= -f\ Y_w \left(\frac{Z_{j,i} - Z_{j,i+1}}{Z_{j,i}\ Z_{j,i+1}} \right) \\ &= -f\ Y_w \vartheta_j, \end{aligned} \tag{8.14}$$

where

$$\vartheta_j = \left(\frac{Z_{j,i} - Z_{j,i+1}}{Z_{j,i}\ Z_{j,i+1}} \right), \tag{8.15}$$

for $j \in \{L, R\}$ and $i \in \{0, ..., N_1\}$. Note that this disparity is different than $d_{v,(i,i+1)}$ defined in Eq. 5.10 in that it is not the vertical disparity difference between $\mathbf{im}_{j,i}$ and $\mathbf{im}_{j,i+1}$. This disparity measures the vertical component error between the projections of $\mathbf{P_W}$ in $(\mathbf{im}_{L,i}, \mathbf{im}_{L,i+1})$

or in $(\mathbf{im}_{R,i}, \mathbf{im}_{R,i+1})$. Furthermore, the coefficient $\vartheta_L = \vartheta_R = \vartheta$ given that $Z_{L,i} = Z_{R,i}$ and $Z_{L,i+1} = Z_{R,i+1}$. Hence, the factor that affects the vertical disparities on the overlapping regions is the same for left and right sequences of images.

The vertical disparities in the overlapping regions between images are null for $Y_W = 0$. However, $dv_{j,i} \neq 0$ for gazing directions above and below the horizontal reference plane. Therefore, as can be seen from Eq. 8.14, the vertical disparities increase with the parameters ϑ in the overlapping region. The $\vartheta \neq 0$ explain the appearance of vertical disparities while mosaicking stereoscopic snapshots originating in cameras with distinct projection centers [2, 14]. This occurs because a point $\mathbf{P_W}$ has different projections on each camera, i.e., $Z_{j,i} \neq Z_{j,i+1}$. This unwanted effect diminishes with the distance between the scene and the omnistereoscopic camera reference center.

The vertical disparities are manifested as ghosting after the blending process. This ghosting will affect the stereoscopic result only when it is not corrected before stitching and blending, e.g., by local registration and warping images [2], and it will be restricted to the overlapping regions.

A comparison between the different configurations in terms of vertical disparities is presented at the end of Chapter 9.

8.4 Summary

In this chapter, I presented a classification of the regions of the interest in the scene based on their visibility from one or two stereoscopic cameras. From this classification, I defined the regions of *stereoscopic* and *omnistereoscopic* FOV in the scene. The latter is the most important for mosaicking partially overlapped stereoscopic images since it defines the stitching and blending region in neighbor images (Section 6.3).

In addition, I developed closed forms for the projective equations to express a point in the scene into the frame of coordinates of each camera and acquisition configuration based on the equations developed in Section 3.3. Using these equations, I developed a closed form for the *horizontal disparity error*, which I introduced in Chapter 5. From the closed form of e_h, I showed that the horizontal disparity error is a function of the projections of $\mathbf{P_W}$ on the $\mathbf{Z}$-axis of each stereoscopic camera pair. Finally, I introduced a ray-tracing method to calculate the minimum distance between a point $\mathbf{P_W}$ in the *omnistereoscopic FOV* and the reference viewpoint using as a criterion the human threshold for depth perception developed in Section 5.1.5.

In terms of the vertical disparities, using the same projective model, I proposed a closed form for the *vertical disparity error* between corresponding projections of $\mathbf{P_W}$. From this projective model, I proved that, when this point is in the *stereoscopic FOV* region of a camera pair, it does not produce vertical disparities after being imaged into $\mathbf{im}_{j,i}$, at least, under ideal conditions. However, when $\mathbf{P_W}$ is in the *omnistereoscopic FOV* of a camera pair, it produces vertical disparities that depends on its distance from the camera and its gazing direction.

In the next chapter, I present the results of applying the ray tracing method proposed in this chapter to compare the four acquisition models.

Chapter 9

Depth consistency and vertical disparities

A first approach to determine the usable FOV of each stereoscopic image acquired for mosaicking was done in Section 5.2.3. That analysis was made relying on two simplified 3D-models of the scene and showed that, by using the limits in the binocular depth resolution in humans, and when certain spatial distribution around the camera are assumed valid, the number of stereoscopic samples necessary for rendering $\mathbf{I_S}$ using mosaicking can be dramatically reduced in comparison with sequential acquisition methods. Although, considering the depth perception limits in humans was the right approach, the models used in the first estimation do not necessarily represent real-world scenarios where elements in scene can be at diverse distances from the camera in diverse gazing directions.

For that reason, a ray-tracing method that takes into account the depth consistency while mosaicking stereoscopic snapshots has been proposed in Section 8.2.4. This method is applied in this chapter on the camera used to produce the experimental results on acquisition and rendering by mosaicking presented in Chapter 6. Furthermore, two other real-world camera models and various lenses were used in these simulations in order to extract relevant insights for the design of omnistereoscopic acquisition systems.

At the end of this chapter, an analysis of the vertical disparities that appears in the overlapping region between stereoscopic images is presented for all the acquisition configurations.

Parts of the results presented in this chapter have been published in SPIE Electrotonic Imaging journal on Feb 2014 [166]. A companion website is available online [167].

9.1 Simulation parameters

The goal of our simulations is to contrast the various acquisition configurations in order to identify characteristics useful to improve the design of the acquisition system. As the horizontal disparity threshold I used the native horizontal resolution of each camera given by the pixel size, i.e., $\varepsilon = s$ (Section 5.1.6). Furthermore, I used real values extracted from the specifications of three off-the-shelf cameras: one *Advanced Photo System type-C* (APS-C) sensor (Canon 400D) and two full-frame sensors (Nikon 800D and Canon EOS6D). Specifications of each camera are presented in Table 9.1. In terms of the lenses, I used three generic lenses with horizontal FOV $\Delta_a \in \{122^\circ, 100^\circ, 76^\circ\}$. I chose this particular set of lenses to illustrate the effect of changing the focal length (Eq. 6.21) in the estimated distance to the scene.

Table 9.1: Image sensor specifications: sensor width W_h, aspect ratio a_r, pixel width s.

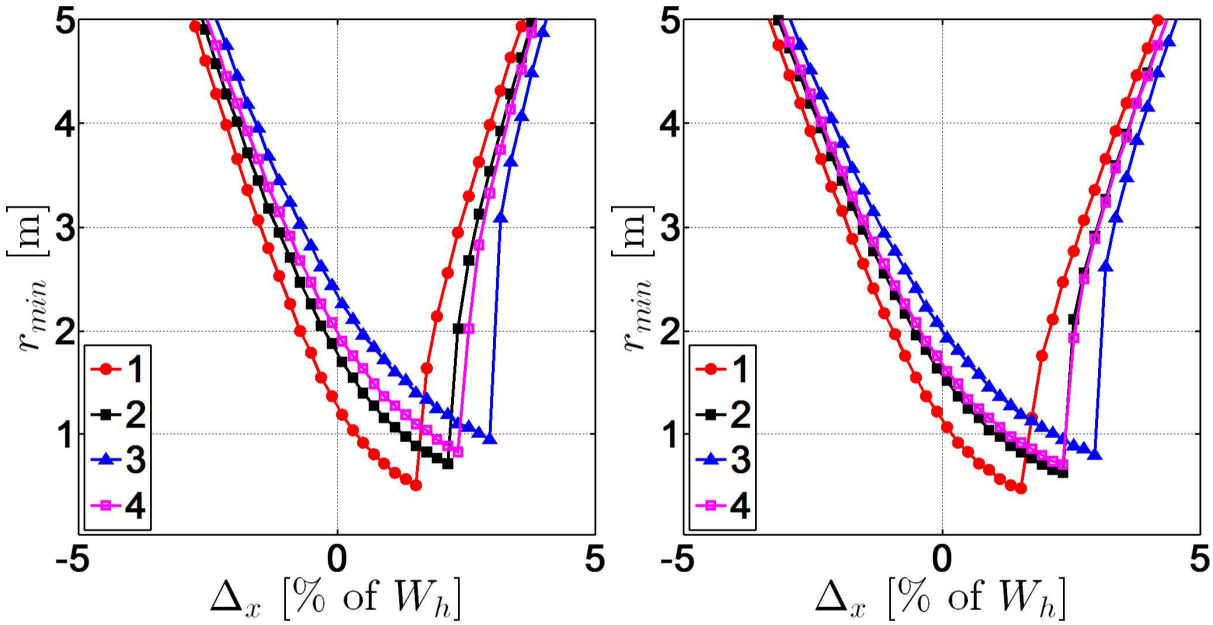
Camera model	W_h	a_r	s
Canon 400D	$22.2\ 10^{-3}$ m	1.5	$5.71\ 10^{-6}$ m
Nikon 800D	$35.9\ 10^{-3}$ m	1.5	$4.88\ 10^{-6}$ m
Canon EOS6D	$35.8\ 10^{-3}$ m	1.5	$6.55\ 10^{-6}$ m

The parameters of interest are r_{min}, which depends on e_h (Section 5.1.4), and the vertical disparities within the blending regions, which depends on ϑ_j ($j = \{L, R\}$) (Section 8.3.3).

9.2 Horizontal disparity continuity

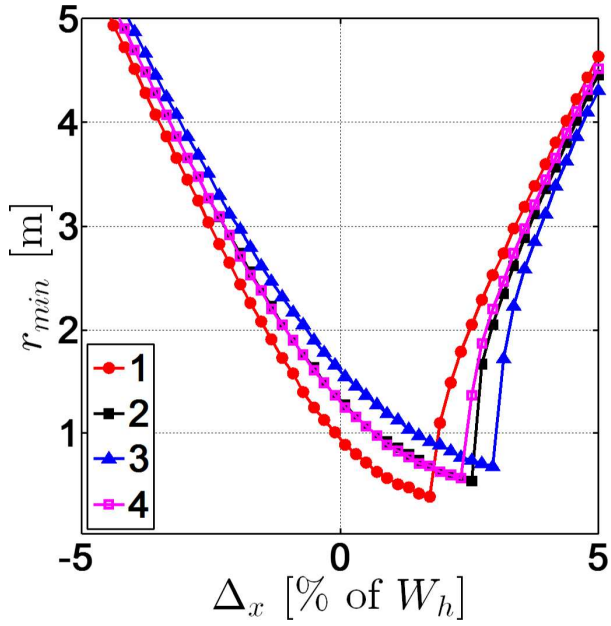
In order to find the distance r_{min}, I calculated first e_h between two neighbour stereoscopic samples to mosaic. This was done following the procedure described in Section 8.2.4. In brief, I defined a ray $\mathbf{L}_{R,0}$ on the $\mathbf{XZ}$-plane, using the stitching position with horizontal coordinates $x_{R,0} = -x_b$ in $\mathbf{im}_{R,0}$ and the projection center $\mathbf{O}_{R,0}$. Then, I computed e_h for the points $\mathbf{P_W} \in \mathbf{L}_{R,0}$, starting at a minimum distance from the camera $\|\mathbf{r_0}\| = 0.3$ m. In order to obtain e_h, I used Eq. 5.19 for all $x_{j,0}$ and $x_{j,1}$ within the blending region (Section 6.3.2). The maximum horizontal width of the blending region was set to 10% of W_h, i.e., the blending area evaluated is between $\pm 5\%$ of W_h around the stitching point x_b. The variable Δ_x is the deviation from x_b in the horizontal coordinate.

In the first simulation, I compared r_{min} for the four configurations using different numbers of stereoscopic samples N (different Δ_θ), for $b = 35$ mm (*configuration 1* and *2*) and the same radial distance $\|\mathbf{r_c}\| = b$ (*configuration 3* and *4*). I used the same $f = 9.3$ mm for three cases, which corresponds approximately to $\Delta_a = 100^\circ$ and W_h from the Canon 400D (APS-C sensor) in portrait orientation. This simulation gives information about r_{min} as a function of the horizontal blending coordinate $X_{R,1} = x_b + \Delta_x$. The results of this simulation are presented in Fig. 9.1.



(a) $\Delta_\theta = 72°$, $N = 5$

(b) $\Delta_\theta = 60°$, $N = 6$



(c) $\Delta_\theta = 45°$, $N = 8$

Figure 9.1: The minimum distance for stereoscopic rendering r_{min} as a function of the horizontal bias Δ_x from stitching coordinate x_b ($X_{R,1} = x_b + \Delta_x$), for Configurations 1-4, and for different number of stereoscopic samples N (Canon 400D, $f = 9.3$ mm ($\Delta_a = 100°$), $b = 35\ mm$ and $\|\mathbf{r_c}\| = b$).

The *critical minimum distance* (ϱ_{min}) is defined by the largest value of r_{min} within blending width Δ_b, which is a few pixels width around x_b (Section 6.3.2). In other words, ϱ_{min} is the practical minimum distance between the omnistereoscopic camera and the scene for a given set of acquisition parameters. Based on the simulation results shown in Fig. 9.1, the values of ϱ_{min} for each configuration are presented in Table 9.2 for x_b ($\Delta_x = 0$) and a given blending region $\Delta_b = 10 * s$. This table provides a numerical example to show how the minimum achievable distance is reduced by increasing the number of stereoscopic samples for a given f.

Table 9.2: Example of the effective minimum distance ϱ_{min} for x_b ($\Delta_x = 0$, $\Delta_b = 10 * s$): ϱ_{min} (Canon 400D, $f = 9.3$ mm $\Delta_a = 100^\circ$ and $b = 35\ mm$).

configuration	$N = 5$	$N = 6$	$N = 8$
1	1.20 m	1.10 m	0.90 m
2	1.70 m	1.50 m	1.30 m
3	2.30 m	2.0 m	1.60 m
4	1.90 m	1.60 m	1.35 m

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

A small ϱ_{min} is necessary to render omnistereoscopic images of indoor scenes where objects may be closer to the camera than in outdoor scenes. This can be achieved by increasing the number of samples N, but that option implies adding more cameras in *configuration 3* or *4*, or taking more samples in *configuration 1* and *2*, which leads to increasing the complexity and cost of the omnistereoscopic system. Another option is changing the focal length and the number of samples to get a compromise between cost and efficiency. The results of doing this is simulated in Section 9.5.

Another useful conclusion from this simulation is that the selection of stitching position x_b, as defined by Eq. 6.20, can be optimized for each configuration. For instance, a stitching point shifted $\Delta_x = -0.05\ W_h$ with respect to x_b leads to an effective minimum distance in the range

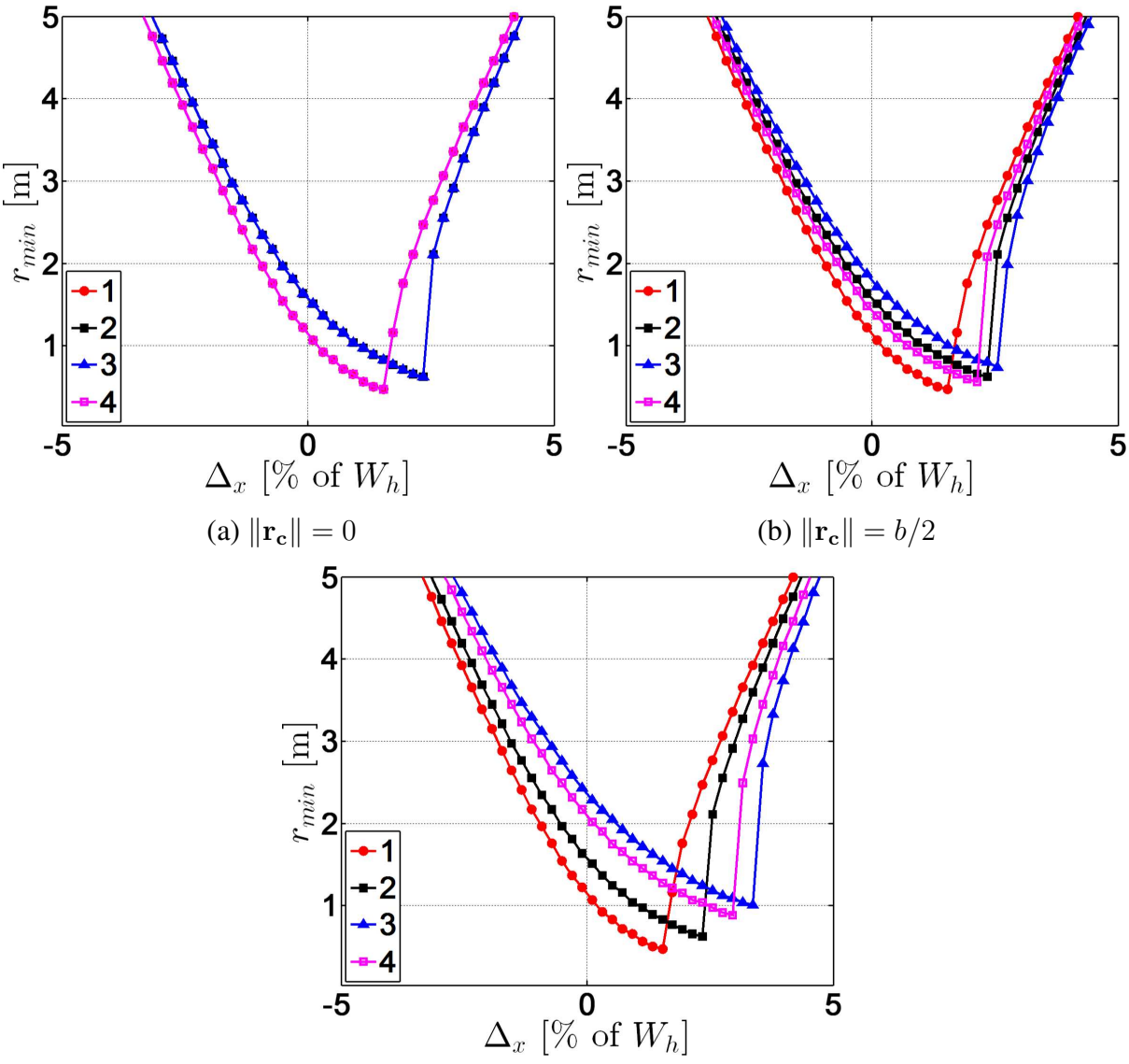
of $4.5-5.2$ m instead of $1.9-1.2$ m as it would be stitching at x_b. Hence, the acquisition model shows that the stitching position and blending width has a larger impact than the configuration used.

Probably the most important conclusion is that there is not much difference between different configurations in terms of the parameter ϱ_{min}. For instance, it can be seen from Table 9.2 that ϱ_{min} varies a few tens of centimeters between configurations, which means that all variations of the acquisition model performs similarly in real acquisition scenarios.

9.3 Effect of radial displacement

The radial displacement $\|\mathbf{r_c}\|$ is relevant for *configuration 3* and *4* to model the physical limitations in a multiple camera system to co-locate the projection centers. The smaller $\|\mathbf{r_c}\|$, the more similar are *configurations 3* and *2*, and similarly, the more similar *configuration 4* is to *configuration 1*. In order to see this effect, I recalculate r_{min} for the case presented in Fig. 9.1-(b) for $\|\mathbf{r_c}\| = 0$, $\|\mathbf{r_c}\| = b/2$ and $\|\mathbf{r_c}\| = 2\ b$. The results are presented in Fig. 9.2. Notice that *configuration 1* coincides with *configuration 4* and *configuration 2* coincides with *configuration 3* for $\|\mathbf{r_c}\| = 0$ (Fig. 9.2-(a)).

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



(a) $\|\mathbf{r_c}\| = 0$ (b) $\|\mathbf{r_c}\| = b/2$

(c) $\|\mathbf{r_c}\| = 2\,b$

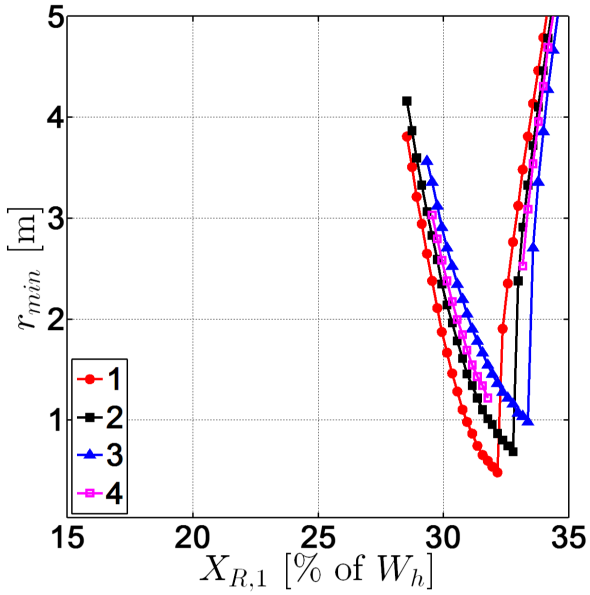
Figure 9.2: Comparison of r_{min} as a function of the stitching bias Δ_x, for all Configurations, after changing the radial distance $\|\mathbf{r_c}\|$ in *configuration 3* and *4* (Canon 400D, $N = 6$, $f = 9.3$ mm ($\Delta_a = 100°$) and $b = 35\ mm$).

9.4 Optimum stitching

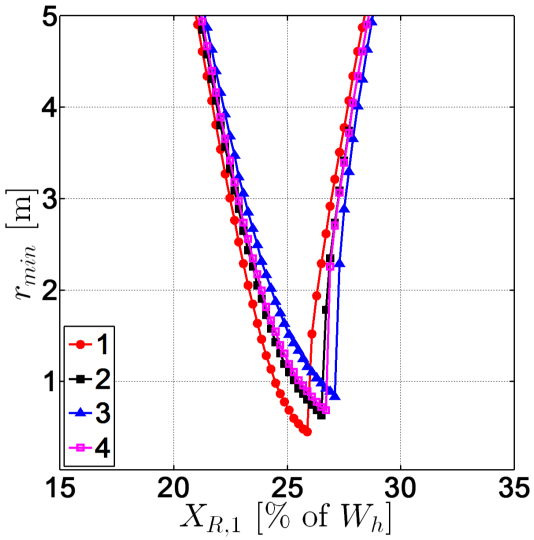
The parameter Δ_θ represents the relative azimuthal angle between two consecutive stereoscopic samples. In addition, each Δ_θ determines a shift in the horizontal coordinates for the optimum stitching position x_b. The optimality of this new stitching position is determined by the closest r_{min} to the camera as shown in Fig. 9.3.

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

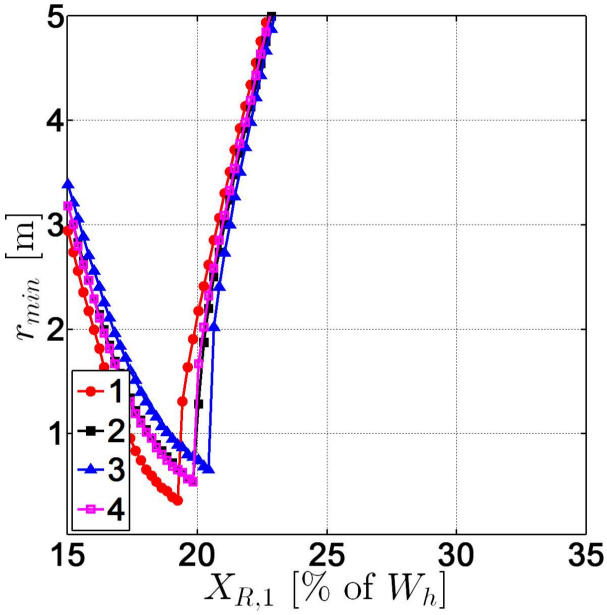
The results for r_{min} shown in Fig. 9.3 are presented as a function of coordinate $x_{R,1}$ instead of around a determined stitching coordinate x_b as were presented in Fig. 9.1. This helps to illustrate the different locations, relative to the image center, of the optimum stitching point when changing Δ_θ.



(a) $\Delta_\theta = 72°$ $(N = 5)$



(b) $\Delta_\theta = 60°$ $(N = 6)$



(c) $\Delta_\theta = 45°$ $(N = 8)$

Figure 9.3: Comparison of r_{min} as a function of the stitching bias Δ_x, for *configuration 1* to *4*, and for different N (Canon 400D, $f = 9.3$ mm $(\Delta_a = 100°)$ $b = 35\ mm$ and $\|\mathbf{r_c}\| = b$).

9.4.1 The optimum stitching location

As can be seen in Fig. 9.1, the curves for the minimum distance r_{min} present a minimum value at different absolute stitching positions x. This optimum stitching position in the images, where the distance between the camera and the scene can be minimized, differs for each acquisition configuration. The shape of the r_{min} vs. x curves and the location of the optimum stitching coordinate is connected to how similar is the imaged scene in two consecutive stereoscopic samples.

The maximum similarity between dh_0 and dh_1, which is measured from the images captured by cameras $(\mathbf{\Omega}_{L,0}, \mathbf{\Omega}_{R,0})$ and $(\mathbf{\Omega}_{L,1}, \mathbf{\Omega}_{R,1})$, respectively, is given by the projections of the scene on the $\mathbf{Z}_j$-axis of each camera pair. In other words, if a point $\mathbf{P_W}$ has projections z_0 on $\mathbf{Z}_0$ and z_1 on $\mathbf{Z}_1$, then e_h will have a minimum when z_0 is close to z_1 or, equivalently, when the scene depth is similar in a certain direction on the overlapping area.

The boundary where the scene depth presents minimum error is defined by the plane that contains the camera reference center $\mathbf{O}$ and all the closest points to the camera center defined by the omnistereoscopic FOV. The intersection of that plane with the horizontal reference plane $\mathbf{XZ}$ defines a line $\mathbf{L_S}$. The maximum similarity boundary line $\mathbf{L_S}$ is illustrated in Fig. 9.4 for the configuration 1.

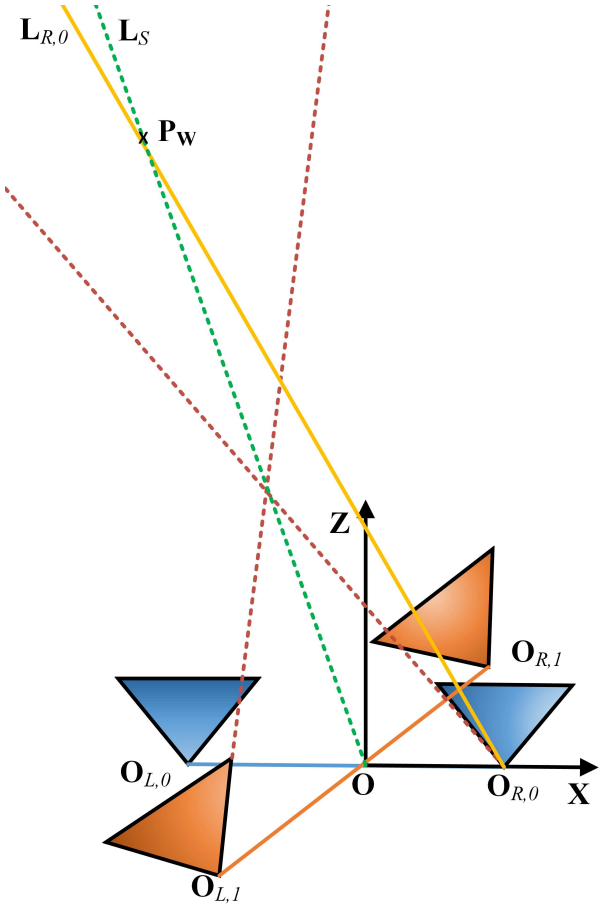


Figure 9.4: The similarity boundary $\mathbf{L}_S$ in the $\mathbf{XZ}$-plane for configuration 1: the ray $\mathbf{L}_{R,0}$, which is defined for a stitching coordinate x, intersects $\mathbf{L}_S$ in $\mathbf{P_W}$. A minimum in e_h is observed for a distance $\|\mathbf{r}\| = \|\mathbf{P_W}\|$.

The maximum similarity boundary line $\mathbf{L}_S$ divides the omnistereoscopic FOV region in two symmetrical volumes. A point $\mathbf{P_W} \in \mathbf{L}_S$ will exhibit the same relative distance to cameras $(\mathbf{\Omega}_{L,0}, \mathbf{\Omega}_{R,0})$ and $(\mathbf{\Omega}_{L,1}, \mathbf{\Omega}_{R,1})$. In the ray tracing approach used for obtaining the minimum distance to the scene presented in Fig. 9.1, the location of the stitching point x determines the intersection point between $\mathbf{L}_{R,0}$ and the boundary $\mathbf{L_S}$. This can be analyzed by presenting the curves of e_h as function of the absolute distance to the camera $\|\mathbf{r}\|$ for a stitching point x.

The crossing point of e_h curves with the threshold $\epsilon = s$ (1 pixel) defines the minimum distance r_{min}. This is shown in Fig. 9.5-(a). However, for some x, the curve may have more than one intersection with $\epsilon = s$ level. In those situations, r_{min} is such that e_h is monotonically decreasing for $\|\mathbf{r}\| \geq r_{min}$ as shown in Fig. 9.5-(b). After establishing this constraint in the selection of r_{min}, I analyzed the peculiar shape of the r_{min} vs x shown in Fig. 9.1.

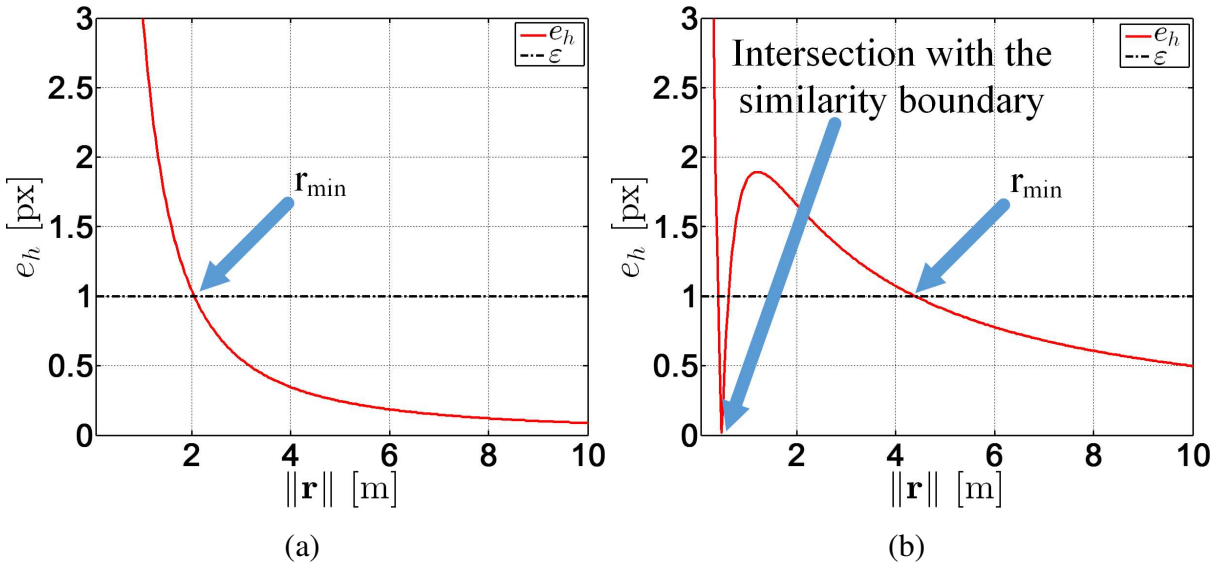


Figure 9.5: e_h as a function of $\|\mathbf{r}\|$.

The camera configuration illustrated in Fig. 9.4 can be used to illustrate the r_{min} obtained by scanning the scene on the $\mathbf{XZ}$-plane with different $\mathbf{L}_{R,0}$ rays (Fig. 9.6-(a)).

The first case occurs when the crossing point between $\mathbf{L}_{R,0}$ and the boundary $\mathbf{L_S}$ is relatively far from $\mathbf{O}$. This is illustrated in Fig. 9.6-(a) for the ray $\mathbf{A}$ from which the e_h illustrated in Fig. 9.6-(b) is obtained. The crossing point between e_h and the threshold s indicates the value of r_{min}. for this horizontal coordinate x of stitching point. It can be observed that e_h is monotonically decreasing for all $\|\mathbf{r}\| \geq 0$. In addition, the crossing point between $\mathbf{L}_{R,0}$ and $\mathbf{L_S}$ is father than r_{min}. The value of r_{min} will decrease as the stitching point x shifts closer to the

left edge of $\mathbf{im}_{R,1}$, illustrated by the decreasing part of the curve shown in Fig. 9.6-(a).

The second example corresponds to x shifted towards the edge of $\mathbf{im}_{R,0}$ with respect to the first example (or shifted towards the center of $\mathbf{im}_{R,1}$). This is shown in Fig. 9.6-(a) for the ray $\mathbf{B}$. In this example, the crossing point between $\mathbf{L}_{R,0}$ and $\mathbf{L_S}$ is closer to the camera. Therefore, e_h exhibits a global minimum for $\|r\| < r_{min}$ where the scene shows maximum similarity between cameras in the overlapping region. This minimum distance r_{min} occurs for the optimum x distance illustrated in Fig. 9.6-(b).

The third and last case, which is illustrated in Fig. 9.6-(a) for the ray $\mathbf{C}$, illustrates what happens for x values larger than the optimum stitching position. For distances $\|\mathbf{r}\|$ larger than the crossing between $\mathbf{L}_{R,0}$ and $\mathbf{L_S}$, which defines e_h the global minimum, the scene shows different depths for each camera pair. Hence, e_h increases for $\|\mathbf{r}\|$ larger than the crossing point between $\mathbf{L}_{R,0}$ and $\mathbf{L_S}$. This is shown in Fig. 9.6-(c). This effect is eventually compensated by the increasing distance between $\mathbf{O}$ and a point $\mathbf{P_W} \in \mathbf{L}_{R,0}$, which will temper the value of e_h. In other words, e_h will decrease for increasing distances until crossing the threshold $\epsilon = s$ again. This last crossing point defines r_{min} for that x. The value where e_h crosses the threshold will be larger for stitching coordinates x that are closer to the $\mathbf{im}_{R,1}$ center. This explains the increasing part of the curve shown in Fig. 9.6-(a), which can be observed after the global minimum.

The location of the similarity boundary $\mathbf{L_S}$ that marks the maximum similarity between the imaged scene in neighbour images (Fig. 9.6-(a)) differs between configurations due to the difference in the relative locations between cameras. Hence, the location of the optimum stitching position will also be different for each configuration. It is interesting to mention that, if the threshold ϵ were larger than s, e_h as function of $\|\mathbf{r}\|$ would no exhibit that shape and not optimum position of stitching can be defined.

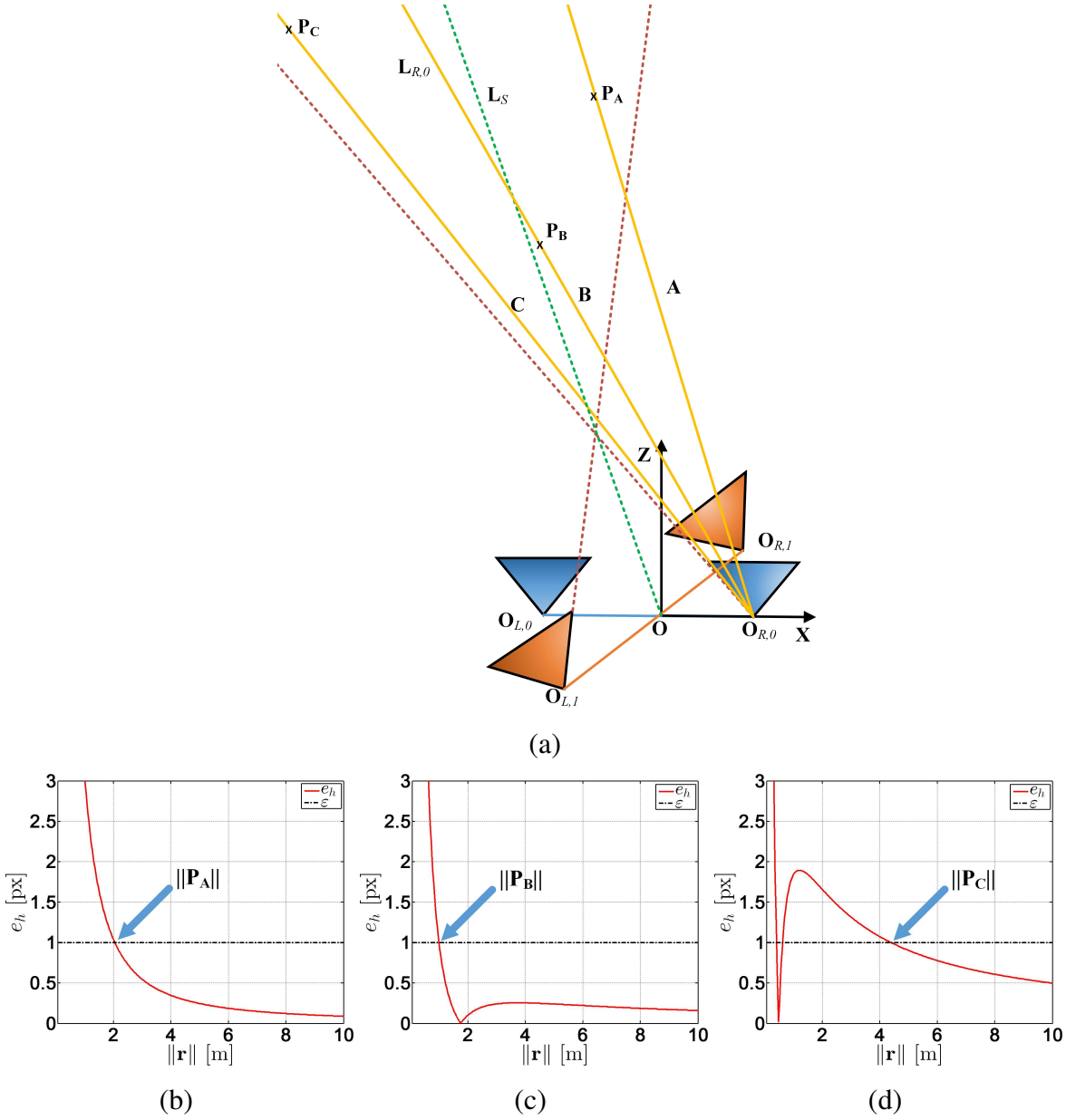
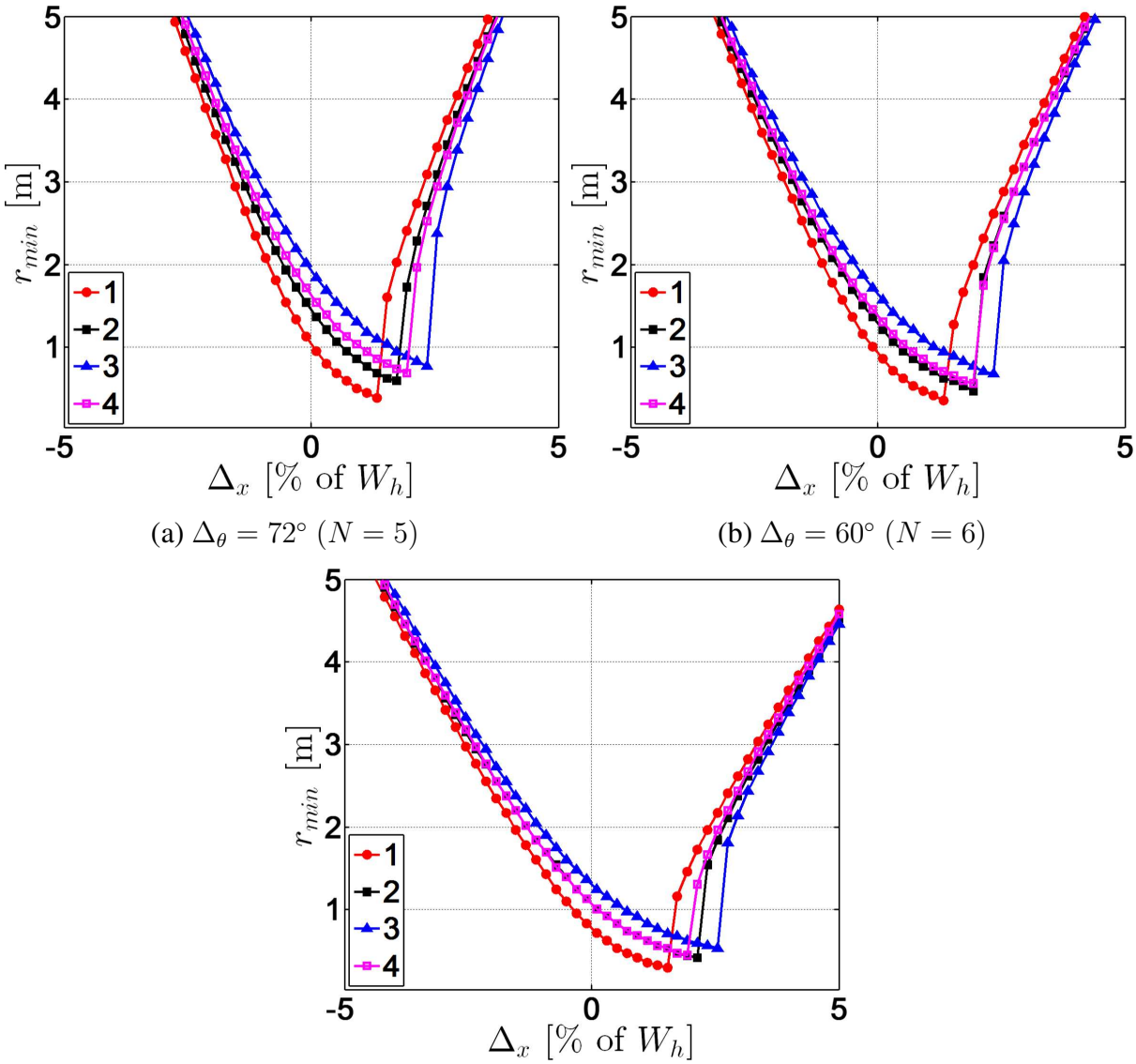


Figure 9.6: e_h as a function of $\|\mathbf{r}\|$: (a) Three rays $\mathbf{L}_{R,0}$ defined by three different stitching positions, (b) for the ray $\mathbf{A}$, the minimum in e_h is closer than the intersection with $\mathbf{L_S}$, (a) for the ray $\mathbf{B}$, the intersection with $\mathbf{L_S}$ happen at a closer distance from r_{min}, and (c) the intersection with $\mathbf{L_S}$ is closer than r_{min}.

9.5 Effect of focal length

Using wide-angle lenses, i.e., larger Δ_a (shorter f), reduces the number of stereoscopic samples. However, a larger Δ_a introduces distortions at the edges of the stereoscopic images, especially for large baselines. Despite this disadvantage, a slight improvement in the effective minimum distance can be achieved by reducing f. I repeated the calculation r_{min} presented in Section 9.2, but this time, I used a smaller focal length ($f = 6.1$ mm, $\Delta_a = 122^\circ$). The results presented in Fig. 9.7 show a small but measurable reduction of ϱ_{min} with respect to using $f = 9.3$ mm ($\Delta_a = 100^\circ$) on the same sensor.

Conversely, it is expected that a larger ϱ_{min} will result when using larger f. This is shown Fig. 9.8 for the same camera parameters as the previous example but for $f = 14.2$ mm ($\Delta_a = 76^\circ$).

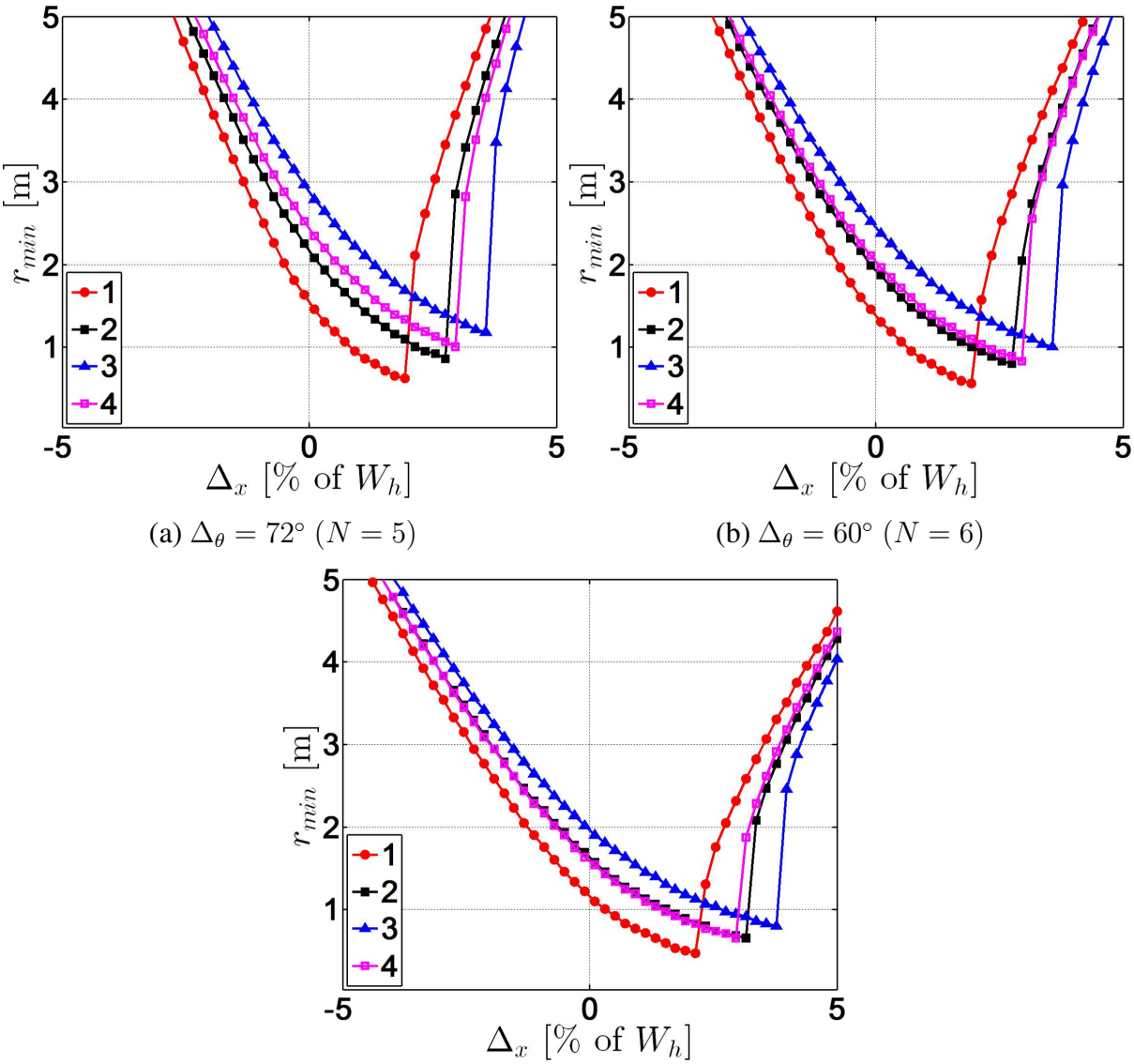


(a) $\Delta_\theta = 72°$ $(N = 5)$

(b) $\Delta_\theta = 60°$ $(N = 6)$

(c) $\Delta_\theta = 45°$ $(N = 8)$

Figure 9.7: Comparison of r_{min} as a function of the stitching bias Δ_x, for Configurations 1 to 4, and for different N, when reducing the focal length to $f = 6.1$ mm $(\Delta_a = 122°)$ (Canon 400D, $b = 35\ mm$ and $\|\mathbf{r_c}\| = b$).



(a) $\Delta_\theta = 72°$ $(N = 5)$

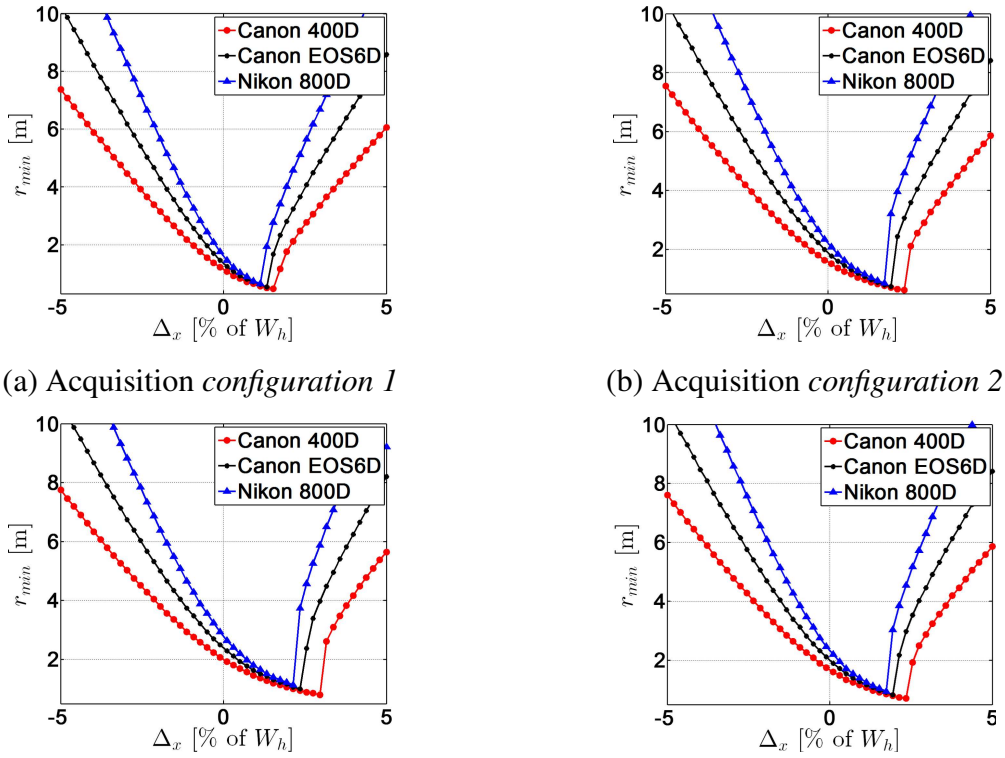
(b) $\Delta_\theta = 60°$ $(N = 6)$

(c) $\Delta_\theta = 45°$ $(N = 8)$

Figure 9.8: Comparison of r_{min} as a function of the stitching bias Δ_x, for *configurations 1* to *4*, and for different N, when increasing the focal length to $f = 14.2$ mm $(\Delta_a = 76°)$ (Canon 400D, $b = 35\ mm$ and $\|\mathbf{r_c}\| = b$).

9.6 Effect of the sensor size

The sensor size affects f for a given Δ_a (Eq. 6.21). For instance, in this configuration, a Canon 400D (APS-C sensor) has $f = 9.3$ mm for $\Delta_a = 100°$, while $f = 15.1$ mm for a full-size sensor ($W_h = 35$ mm) (Table 9.1). Compared to the pixel size of the Canon 400D sensor, the pixel size of the Canon EOS6D is approximately 15% wider while the pixel width of the Nikon 800D is 15% smaller. The results for finding r_{min} for the four configurations are presented in Fig. 9.9.



(a) Acquisition *configuration 1*

(b) Acquisition *configuration 2*

(c) Acquisition *configuration 3*

(d) Acquisition *configuration 4*

Figure 9.9: Comparison of r_{min} as a function of Δ_x, for Configurations 1 to 4, and for different sensors while maintaining a given FOV Δ_a.

For a comparable sensor size and the same focal length, the Nikon 800D sensor performance is worse than Canon EOS6D sensor in terms of ϱ_{min} because the pixel size of the former is smaller, which makes the threshold for e_h smaller and pushes the minimum tolerable distance away from the camera. The Canon 400D sensor is smaller and its f is also smaller, so ϱ_{min} for the same sampling angle Δ_θ is closer to the omnistereoscopic camera.

9.7 Vertical disparities

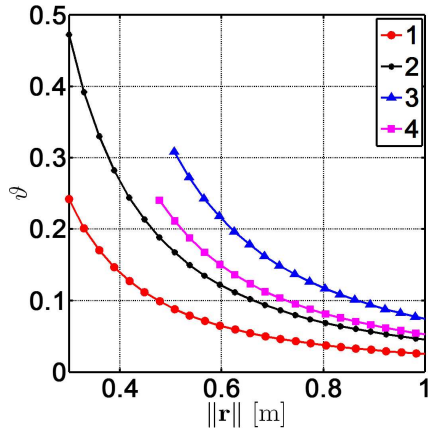
An important side effect of the spatial distribution between cameras in each acquisition configuration is the manifestation of vertical disparities in the overlapping region between neighbour stereoscopic images (Section 8.3.3). In this section, I show the result of comparing the four acquisition configurations in terms of the coefficients ϑ, which determine the magnitude of the vertical disparities in the blending areas.

As seen in Eq. 8.14, the undesired vertical disparities are proportional to $\mathbf{Y_W}$. But, this error also depends on the distance between the camera and $\mathbf{P_W}$ on the $\mathbf{XZ}$-plane. The results of calculating the value of ϑ as a function of the distance to the camera $\|\mathbf{r}\| = \|\mathbf{P_W}\|$ for the four configurations are presented in Fig. 9.10. This results are valid for a Canon 400D camera, $b = 65$ mm, $\|\mathbf{r_c}\| = 65$ mm, $N = 6$ and $\Delta_a = 100°$.

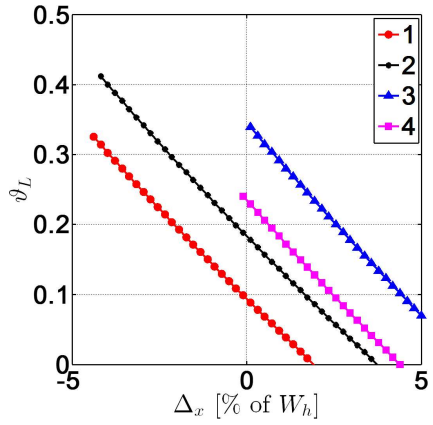
A non-zero ϑ leads to a *vertical disparity error* (Section 8.3.3) in the blending areas, hence, it is desirable to take into account this effect when positioning the omnistereoscopic camera, especially for acquiring scenarios that are relatively close in certain elevation angles. This relationship between vertical disparities, the proximity to the camera and the gazing direction in elevation explains the vertical disparities that appear at the top and bottom parts of the mosaicked images, which are both at the shorter distance $\|\mathbf{r}\|$ on the $\mathbf{XZ}$-plane and, at the same time, have large $\mathbf{Y_W}$ components. Fortunately, both coefficients converge to zero at a relatively short distance from the camera for all the configurations of the acquisition model.

The magnitude of this error is also dependent on the particular stitching position in the image pair x_b. In Fig. 9.10-(c)-(f), I present the value of ϑ over the blending region, for $\|\mathbf{r}\| = 50$ cm of the camera and $\|\mathbf{r}\| = 100$ cm. These results show that the vertical disparities over each blending region decrease with the distance between the scene projection on the $\mathbf{XZ}$-plane and the camera.

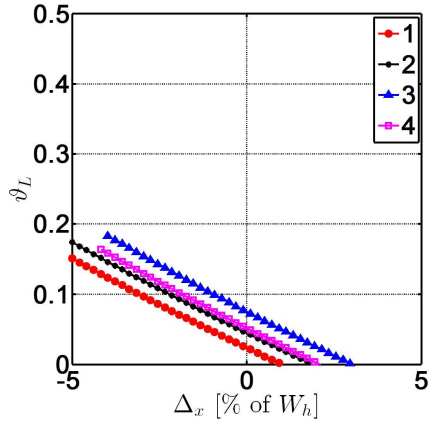
The magnitude of this error is also dependent of the particular stitching position in the image pair x_b. In Fig. 9.10-(c)-(f), I present the value of ϑ over the blending region, for $\|\mathbf{r}\| = 50$ cm of the camera and $\|\mathbf{r}\| = 100$ cm. These results show that the vertical disparities over each blending region decrease with the distance between the scene projection on $\mathbf{XZ}$-plane and the camera.



(a)



(b)



(c)

Figure 9.10: Vertical disparity coefficients ϑ as a function of the distance to the camera $\|\mathbf{r}\|$, for all the acquisition configurations, and for $x_{R,1} = x_b$ ($\Delta_x = 0$): (a)ϑ as a function of the distance r and the variation of this coefficient around x_b as function Δ_x, when the distance to the scene is (b) $\|\mathbf{r}\| = 50$ cm and (c) $\|\mathbf{r}\| = 100$ cm.

9.8 Summary

In this chapter, I presented a comparison of the depth consistency for the different acquisition configurations. This comparison is based on the ray tracing method to obtain the minimum distance to the scene to achieve a consistent binocular experience of depth (Section 8.2.4). In this chapter, I analyzed the performance of the system emulated in Chapter 5. In addition, I contrasted these results against omnistereoscopic systems based on two other DSLR cameras.

The main result is that there is no substantial difference between camera geometries in terms of the minimum distance between camera and the scene. However, I found differences in the optimal location for stitching stereoscopic samples, which depends of the geometry of the omnistereoscopic system. This result can be used to improve the rendering process by implementing an optimization method to find x_b based on the distance between the camera and the scene. The development of such an optimal cut algorithm is beyond the scope of this thesis. Additionally, based on these simulation, I suggested strategies affecting the design of simultaneous acquisition systems that can increase the efficiency in the omnistereoscopic acquisition.

In Section 8.3, I presented the closed-forms expressions of the vertical disparity equations, which are important in the overlapping areas between stereoscopic samples. Based on these equations, in this chapter I studied the correlation of these vertical disparities with the elevation component of $\mathscr{D}$ and the distance between the camera and the scene.

In the next chapter, I am presenting the conclusion for this thesis.

Chapter 10

Conclusions

The main hypothesis of this thesis is that omnistereoscopic imaging of real-time events can be satisfactorily approximated using a limited number of partially overlapped stereoscopic snapshots of the scene.

The research method described in Chapter 1 led to propose an acquisition strategy based on multiple cameras to collect partially overlapped stereoscopic snapshots of the scene, followed by a rendering approach based on mosaicking the collected images. This solution has been founded on the practicality of acquiring a sparse set of stereoscopic samples of $\mathbf{C_S}$ and the low-complexity of the rendering approach to produce omnistereoscopic imagery of real-time events. Both advantages are attractive for the real-time omnidirectional acquisition of scenes.

The experimental part of this thesis was supported by the sequential sampling of $\mathbf{C_S}$, which prevents me to demonstrate the suitability of my proposed approach in truly dynamic environments. However, I provided sufficient evidence to support the idea that arrangements of stereoscopic camera rigs, as long as occlusion between cameras is avoided, can be used to acquire a sparse set of partially overlapped stereoscopic snapshots. This sparse set of stereoscopic samples can be mosaicked into satisfactory omnistereoscopic images of the scene in any panning direction. I have supported this hypothesis with extensive experimental examples.

Having established the viability of the multiple camera approach, I focused on the development of the analytical foundations of the proposed omnistereoscopic acquisition technique. Based on this analysis, I studied some of the main issues that arise from this omnistereoscopic acquisition technique: depth consistency, vertical disparities and design parameters for a multiple camera system. This analysis and their results are main contributions of my thesis.

In order to provide a satisfactory immersive experience to the user, this particular acqui-

sition and rendering technique needs to provide the means to create imagery free of stitching artifacts. Not only that, the omnistereoscopic imagery rendered by this method has to provide consistent binocular cues of depth in every gazing direction. These two main problems have been examined in this thesis and the results are summarized next.

10.1 The viewing models

The thesis started with the definition of an *omnidirectional viewing function* (Chapter 2). This function is defined from the plenoptic function $\mathscr{P}$ and models the physical signal to be measured by any panoramic sensor. The measuring of $\mathbf{C}$ leads to a real-valued multi-dimensional signal $\mathscr{I}$, which after being sampled generates the panoramic image $\mathbf{I}$. This approach is not novel and has been used before to model the multi-dimensional signal measured by a generic camera to produce an image, which can be omnidirectional (panoramic), a wide-angle angle image (photographic camera), a narrow FOV image (line sensor) or to obtain a sample of the light field.

The first contribution presented in this thesis has been to propose an *omnistereoscopic viewing model*, which abstracts all the geometric constraints of the binocular visual system. The model is based on two omnidirectional viewing functions, which are denoted $(\mathbf{C_L}, \mathbf{C_R})$, to model the light captured at two distinct viewpoints with horizontal parallax for a given FOV. The constraints imposed by the model in the gazing direction $\mathscr{D}$ enables us to model the vergence of the human eyes towards a region of interest in the scene.

The *omnistereoscopic viewing model* defines a multi-dimensional array of real-valued signals that represents the *omnidirectional viewing function*, denoted $\mathbf{C_S}$, which encompasses all the necessary information to be acquired by any omnistereoscopic sensor that intends to recreate direction-independent stereoscopic views of a scene for human viewing. This model of the physical magnitude to be measured by an omnistereoscopic device has not been proposed before based on my own research on the subject up to this day. Hence, it is one of the contributions of this thesis.

As defined, the *omnistereoscopic viewing model* has an indetermination in proximity of the poles (*zenith* and *nadir*) where multiple horizontal baselines can be defined for the same $\mathscr{D}$. This indetermination does not prevent to acquire and render stereoscopic images in these extreme directions. However, this problem has not been studied in this thesis.

An interesting point for future research is modifying the *omnistereoscopic viewing model*

to eliminate the restriction of having a single reference plane. More specifically, the model proposed herein is restricted to one horizontal reference plane which is convenient to model acquisition cameras with horizontal parallax, but it is interesting for people in the panoramic community to render omnistereoscopic images where the stereoscopic viewpoints are not restricted to a single plane. In this extended model, the possible locations for the left and right viewpoints define a spherical region around the center $\mathbf{O}$ instead of a *viewing circle* on the horizontal plane.

10.2 The acquisition model

In this thesis, the *omnistereoscopic viewing model* has been used to define the geometric constraints for a generic omnistereoscopic acquisition system, which becomes the general *acquisition model* introduced in Chapter 3. This model can describe the variety of omnistereoscopic cameras to produce stereoscopic images with horizontal parallax.

In order to sample $(\mathbf{C_L}, \mathbf{C_R})$, the model includes two pinhole cameras with parallel optical axes which are located in place of the omnidirectional projection centers. The horizontal parallax distance and coplanarity defined by the *omnistereoscopic viewing model* are followed by the *acquisition model*. Since the pinhole cameras have a limited FOV, only partial samples of the $\mathbf{C_S}$ are acquired in any gazing direction $\mathscr{D}$.

The relative location of left and right pinhole cameras can be manipulated while maintaining the same baseline and panning restrictions. Moreover, the different spatial distribution of projection centers defines a set of acquisition *configurations*, which can model the omnistereoscopic cameras and acquisition techniques proposed in the literature (Chapter 4).

The *acquisition model* enables a tractable method to study a variety of effects such as the consistency of the depth over the whole panorama when mosaicking stereoscopic snapshots, the minimum acceptable distance between the camera and scene and the vertical disparities.

In addition, the *acquisition model* helps to devise novel strategies to acquire samples of $\mathbf{C_S}$ in order to estimate $\mathscr{I}_S$. For instance, a brute-force acquisition approach to sample $\mathbf{C_S}$ used to acquire narrow vertical samples of $(\mathbf{C_L}, \mathbf{C_R})$ can be modeled by a stereoscopic camera in any of the four *configurations* of this model. However, without post-processing, the binocular image created in this way is not what we would see in any gazing direction $\mathscr{D}$. Alternatively, a pair of wide-angle snapshots can be acquired for any given gazing direction. The latter acquisition approach represents more closely the stereoscopic qualia of the scene that we would

experience when looking in $\mathscr{D}$ according to the *omnistereoscopic viewing model*. Moreover, the latter approach can be emulated by acquiring a few stereoscopic images at constant panning increments to mosaic them into an approximation of $\mathscr{I}_S$.

The *acquisition model* in any of its four basic *configurations* can describe most of the omnistereoscopic acquisition techniques available today. However, it cannot model omnistereoscopic methods based on vertical parallax since these configurations are not contemplated in the *omnistereoscopic viewing model* presented herein.

It is worth mentioning that, even though the *omnistereoscopic viewing model* contemplates the vergence of the left and right gazing directions, parallel optical axes between cameras in a stereoscopic pair are used in the experiments on acquisition to facilitate the rendering by mosaicking. The vergence for all panning directions is emulated in rendering time by defining the zero parallax point in $\mathbf{I_S}$ (Appendix C), although the acquisition model contemplates the emulation of vergence mechanisms. It will be interesting to evaluate the effect of vergence in the visual comfort of the user, but such a study is beyond the scope of this thesis.

The general *acquisition model* is another of the original contributions of this thesis to the area of research on omnistereoscopic image formation.

10.3 The limits of human depth resolution

A threshold for human depth resolution was proposed in Chapter 5. In this analysis, the threshold was developed by comparing a stereoscopic camera model with a projective model developed for human depth resolution. The stereoscopic camera was composed of two pinhole cameras with parallel optical axes and horizontal parallax equivalent to the interocular distance in the average human population. Based on an estimation of the FOV in humans, the minimum perceived error between two distances in front of the eyes was used to estimate a threshold for the horizontal disparity resolution in humans. The estimated threshold for the horizontal disparity is several times lower than one pixel width in the average off-the-shelf DSLR cameras available at the moment of writing this thesis.

The horizontal disparity threshold was used before in the context of measuring the expected distortions when projecting an omnistereoscopic image in a cylindrical display. However, in that case the threshold of depth resolution used was several times higher than the one estimated in this thesis.

This discrepancy in the depth resolution can be attributed to the effect of different uncer-

tainties in the model. First, the human depth resolution does not depend on binocular cues only. The effect of textures, illumination, color and perspective of the scene affect the thresholds of depth perception. In both experimental cases, key parameters in the calculation, e.g., the human depth-of-field resolution, were taken from the results of visual tests performed on humans.

Since this uncertainty was found, a more practical threshold for the horizontal disparity resolution was used in this thesis: a horizontal disparity error in the order of one pixel width. This threshold is convenient in the context of rendering by mosaicking. Additionally, it guarantees no visible discontinuities between mosaicked stereoscopic images.

However, further research should be made in determining the real thresholds for depth resolution in stereoscopic displays.

10.4 The mosaic width approximation

Although determining the human minimum resolvable error in horizontal disparity using a model based on a pair of pinhole cameras is not a novel contribution, the use of such a threshold to determine the usable width of a stereoscopic mosaic is new.

In the test experiment described in Chapter 5, a stereoscopic pinhole camera acquired two images of a particular scene. This scene obeys two simplified 3D-models: in one model the scene is located at constant distance in every direction $\mathscr{D}$ from the reference center $\mathbf{O}$ (sphere), in the other model the scene is distributed at the same distance in the panning direction (cylinder). The baseline of the stereoscopic rig and the distance between the scene and the camera for each model determines the variation in the horizontal disparity along the stereoscopic image. This is valid for any configuration of the acquisition model since the baseline is the determining parameter.

The idea of the test was to determine the regions in the stereoscopic image that can be distinguished as having a different depth by applying the threshold for horizontal disparity proposed in Chapter 5. In other words, if $\mathscr{D}$ in the center of region of the image conveys certain binocular cues of depth, the usable region to mosaic extends from that region to the periphery until crossing the threshold of depth resolution. A larger eccentricity will be perceived as being at different depth.

This simple approach enables us to determine the usable width of a stereoscopic sample and consequently the number of samples necessary to acquire. The problem with this method is

that normally scenes cannot be modeled completely with such simple models. In some experimental examples, i.e., the example in Section 7.14.1, the scene presents minimum distances to the camera that vary for different panning angles. Using the number of samples determined by this method, based on the closest points to the camera, a larger number of samples than necessary may be necessary. However, the knowledge of the average distance to the scene and its type of spatial distribution can be used to obtain an educated guess based on this model.

An interesting point for future research is to extend this approach to study the perceived distortion in depth when projecting stereoscopic images onto different planar or warped canvases.

10.5 Experiments in acquisition

In Chapter 7, the experimental set-up used in the experiments was presented. This acquisition strategy was based on using an off-the-shelf DSLR camera equipped with a wide-angle lens to consecutively capture each of the $2N$ images a multiple camera system in any of the four acquisition configurations would acquire. This situation is suitable in static scenarios and it was used to develop and test different strategies of rendering. In order to test dynamic scenes, a similar strategy was used, but this time, a stereoscopic rig composed of two identical DSLR cameras was used. However, only partial dynamic scenes in the omnistereoscopic FOV of each panning angle of the stereoscopic rig were acquired. The building of a multiple camera set-up that can simultaneously acquire the entire scene is beyond the scope of this thesis.

These experimental set-ups allowed me to acquire piece-wise dynamic scenes (dual camera set-up) or static scenes (singular camera set-up). Despite this limitation, the suitability of a multiple camera configuration to acquire a complete set of stereoscopic images simultaneously as a solution for the dynamic scene acquisition problem has been sufficiently demonstrated in this thesis. Therefore, my experiments focused on the acquisition and reconstruction problems.

Efficient stitching and rendering algorithms need to be proposed to support multiple camera set-ups. These algorithms should be optimized for the real-time acquisition of omnistereoscopic imagery. In that sense, the *acquisition model* helps in the design of such multiple camera arrangement. The cost and size of high-resolution cameras and the necessary wide-angle lenses makes it attractive to build a full omnistereoscopic acquisition system. The design and implementation of such a prototype has attached a variety of technical challenges such as the synchronization of multiple cameras, the data transmission and storage, the real-time stitching

and blending, etc.

10.6 Novel acquisition strategies

One of the original contributions of this thesis is the omnistereoscopic camera based on acquisition *configuration 3*. Some experimental results using this camera are shown in Section 7.14.3.

This camera concept consists of a set of cameras, whose projection centers are radially displaced with respect to the reference viewpoint, and a second set of cameras laterally shifted with respect to each of the radially displaced cameras (Fig. 4.20-(c)). The first set of cameras captures a set of partially overlapped images that can be easily mosaicked to render one panoramic image, i.e., $\mathbf{I_L}$ or $\mathbf{I_R}$. The second set of cameras can be either mosaicked when the baseline is narrow (Section 7.14.3) or used to extract a panoramic depth map of the scene (Section 7.14.3). The former rendering approach exploits the microstereopsis exhibited in stereoscopic images acquired with narrow baselines. The latter rendering method is more versatile since it synthesizes a second stereoscopic image using 2D-to-3D conversion. The latter approach has advantages in terms of emulating mechanisms of human vision such as *vergence* and *accommodation* in post-processing time. However, the emulation of these mechanisms is not part of this thesis. This novel configuration has been emulated by using a single camera for narrow baselines and a stereoscopic rig for large baselines. Even though the ideal experimental set-up is to build a multiple camera based on this configuration, this approach is beyond the scope of this thesis as I mentioned before.

One interesting aspect that has not been tested is the effect on the visual comfort of combining one artifact-free image and one mosaicked image in a stereoscopic pair. This is applicable to the mosaicking of stereoscopic images acquired with this camera configuration. In that case, stereoscopic views in intermediate $\mathscr{D}$ between actual sampling directions may be composed of one stitching-free image and another with visual artifacts. An approach for this test would involve human trials and therefore it is beyond the scope of this thesis. Another method is to devise a metric for the stereoscopic quality that takes into account the amount of *asymmetrical distortion* that is tolerable, e.g., the structural distortion of one image only in a stereoscopic image.

The 2D-to-3D technique implemented is a proof-of-concept of the dynamic generation of stereoscopic pairs. More research needs to be done in that direction. For instance, the fusing of the estimated depth maps from partial views of the scene into an omnidirectional $\mathscr{M}$ has not

been developed as part of this thesis.

Another original acquisition method proposed in this thesis is the use of three panoramic images acquired in a triangular co-planar pattern. The panoramic cameras must have coplanar projection centers and parallel vertical axes. This approach is based on extracting six partially overlapped stereoscopic snapshots, each of them covering $60°$, from the set of three panoramas. This method can be modeled as a variation of the acquisition *configuration 3*. A similar approach has been tested before but using four panoramas to mosaic eight partially overlapped stereoscopic images. Unfortunately, the panoramic triad method cannot be adapted for simultaneous acquisition using multiple panoramic cameras. Therefore, it is limited to acquire static scenes only. An extensive explanation of the panoramic triad method has been included in Appendix A.

The panoramic alignment proposed alongside the panorama triad method is another original contribution of this thesis. Other methods have been proposed to align panoramas in different formats to arbitrary directions in the scene. All these methods are based on assuming a pinhole camera and estimated from it the calibration parameters to estimate a fundamental matrix and defining a correspondence transformation between panoramas. The alignment technique proposed as part of this thesis is based on identified the unique directions of *focal expansion* and *contraction* in a dense disparity map defined between panoramas. The alignment direction in azimuth is unique and it enables us to identify the regions usable for stereoscopic mosaicking in any pair of panoramas from the panoramic cluster method proposed in this thesis. This technique does not assume any camera model and it is uncalibrated. The pattern detection implemented is based on the dense disparity map analysis. Better results can be obtained by using a more robust approach of correspondence feature detection. The panoramic alignment technique has been included in Appendix B.

10.7 Depth consistency and vertical disparities

In Chapter 8, the projective equations derived from each configuration were used to study the parameters of interest for the rendering approach based on mosaicking. In order to model the depth continuity, the theoretical and practical thresholds for the depth resolution proposed in Chapter 5 were used. One of the constraints studied was the minimum distance between the camera and the scene to reproduce a continuous depth illusion in all panning directions. The effect of the focal length, number of stereoscopic samples and radial distance was corre-

lated to the minimum distance from the camera after extensive ray tracing simulations. These comparison are presented in Chapter 9.

The main conclusion from these simulations is that there is no substantial difference between the four acquisition *configurations* in terms of the minimum distance between camera and the scene. However, differences were found in the optimal location for stitching stereoscopic samples, which can be attributed to the multiple camera geometry. This effect can be used to improve the rendering process. Also from our simulation results, strategies affecting the design of simultaneous acquisition systems were proposed.

Finally, the *acquisition model* was used to model the vertical disparities. The derived model for the vertical disparities enables to calculate their dependency of the gazing angle in elevation and the distance to the scene.

10.8 Future work

An interesting follow up to this research will be to study the adaptive generation of stereoscopic views around and between different viewpoints. This should be done considering the binocular mechanisms of human viewing. This will enhance the stereoscopic visualization of image-based virtual environments by providing a consistent and more natural illusion of depth in all directions around the user. This problem involves novel methods to generate stereoscopic images dynamically based on the users' gazing direction and the distance to the observed region-of-interest in the scene.

The omnistereoscopic imagery of dynamic scenes based on the mosaicking of a limited number of partially overlapped stereoscopic images is a feasible and computationally inexpensive approach. This claim has been supported by this thesis research. However, without a post-processing step, the stereoscopic views produced by this method have a fixed vergence point at infinity and maximum depth-of-field so all the scene appears in focus to the user. A more natural visualization of an image-based stereoscopic simulation must render binocular views by taking into account the perceptual mechanisms of the human binocular vision.

An important problem to be addressed is the perceived distortion introduced by mosaicking stereoscopic images acquired from multi-camera configurations with different geometries. Research has been done to propose a stereoscopic image quality benchmark, but a measure of quality applicable to mosaicked stereoscopic images that considers binocular vision mechanism requires further research. Some steps in that direction based on an omnistereoscopic viewing model are suggested in my thesis, but further research needs to be done.

The *vergence* of the eyes towards a region-of-interest implies the alignment of stereoscopic images before producing a limited field-of-view. This aligning transformation depends on the distance between the scene and the reference point in the scene. A warping transformation to align the stereoscopic image pair introduces undesired effects on the peripheral regions of the images that affect the visual comfort. The second important mechanism of binocular vision is the *accommodation* of the eyes, which in this context implies to emulate a depth-of-field that is consistent with the distance between the reference viewpoint in space and the region-of-interest in the scene. It has been demonstrated experimentally that stereopsis occurs only in regions of the image that are focused, which is a consequence of the eyes' accommodation on the region-of-interest. This phenomenon can be exploited to improve the stereoscopic visual comfort.

The synergy between vergence and accommodation mechanisms should guide the adaptive

generation of stereoscopic pairs while the user explores the virtual environment around the users viewpoint in space. Hence, the rendering pipeline must exploit the feedback information of the gazing direction and the distance to the virtual scene, which can be determined by means of eye and head tracking devices.

A feasible rendering approach using stereoscopic image pairs is to define an image warping transformation based on disparity maps to emulate *vergence*. After the alignment and correction of the image pairs, a region-based blurring based on the estimated dense motion field can be used to emulate *accommodation*. Alternatively, the dynamic generation can be based on a single panoramic image and its dense omnidirectional depth-map. In this case, stereoscopic images can be created by extracting the view for one eye from the panoramic image, whereas the view for the other eye can be synthesized using 2D-to-3D conversion techniques. Even though this idea has been used in automatic stereoscopic conversion, further research needs to be done in the context of an omnistereoscopic interactive environment.

Appendix A

Experimental acquisition based on panoramic clusters

In this appendix of my thesis, I am presenting an alternative technique to acquire a set of partially overlapped stereoscopic images to produce $\mathbf{I_S}$ using the mosaicking technique explained in Section 5.2.3. This approach to acquisition is sequential. Consequently, it cannot be used to acquire dynamic scenes. Despite this disadvantage, this is an efficient method to acquire a set of wide-angle stereoscopic snapshots $\mathbf{im}_{j,i}$ of a static scene.

The acquisition time, rendering complexity, storage capacity, and camera self-occlusion are some of the constraints that make the idea of using panoramic snapshots to sample $\mathbf{C_S}$ challenging. Offsetting these disadvantages, cameras to produce high-resolution monoscopic panoramas are easily available nowadays. In addition, they provide an economic alternative to acquire wide-angle images.

For all these reasons, my research on omnistereoscopic techniques to sample $\mathbf{C_S}$ started with the development of a technique based on sampling the scene with clusters of three panoramic images arranged in a controlled geometric pattern. This technique, which can be modeled using the acquisition *configuration 4*, can be implemented with any off-the-shelf panoramic cameras. The low computational complexity and reduced data overhead of the rendering process based on mosaicking make this method attractive for the large scale acquisition of omnistereoscopic images in a variety of scenarios.

Part of this method has been presented in the SPIE/IS&T Electronic Imaging conference in 2012 [2]. The companion website containing interactive omnistereoscopic images created

with this method is available online [168].

A.1 Antecedents of the idea

The use of panoramic sources to sample the omnistereoscopic viewing function has been covered in Section 4.4. Many of these omnistereoscopic methods were specifically conceived for the off-line navigation of image-based stereoscopic virtual environments. For instance, Yamaguchi et al. [100] and, in a more recent follow up, Hori et al. [101], 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.

A different approach based on using four panoramic cameras in a square co-planar pattern has been proposed in a patent application from 2003 by Baker et al. [86]. This camera aims to acquire four stereoscopic snapshots to sample $\mathbf{C_S}$ in four orthogonal panning directions. An example of this camera is in Fig. 4.16. The authors propose to exploit the *stereoscopic FOV* that arises between pairs of panoramas, However, as seen in Section 5.2.3, a rendering based on mosaicking only cannot produce a continuous horizontal disparity between mosaics, at least, if the scene is close to the camera. Therefore, this system requires a more complex rendering approach, e.g., relying on the omnidirectional depth map $\mathscr{M}$ to describe the 3D-model of the scene. This patent was granted in 2007.

Vanijja and Horiguchi [14] proposed an acquisition method 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 walls of a CAVE display 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.

The omnistereoscopic acquisition method I am presenting extends and improves the idea proposed by Vanijja and Horiguchi, reducing the required number of panoramic samples. 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 depth information of the scene. In order to do this, I propose a sampling strategy based on clusters of three panoramic snapshots distributed in a coplanar and equidistant triangular pattern.

A.2 The proposed 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, exhibits these regions of stereoscopic usability in different panning angles in azimuth when considering the panoramas in pairs.

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.

This method of stereoscopic sampling can be modeled using the acquisition *configuration 4*, where a virtual stereoscopic rig is located off-center a distance r_c with respect to the reference viewpoint $\mathbf{O}$ (Section 3.2). The stereoscopic rig is displaced behind $\mathbf{O}$ to model the acquisition of $\mathbf{im}_{j,i}$, where $j \in \{L, R\}$ and $i \in \{0, ..., N-1\}$. This is illustrated in Fig. A.1-(a). The next stereoscopic sample, $\mathbf{im}_{j,i+1}$ is acquired with a virtual stereoscopic rig displaced r_c in front of $\mathbf{O}$ and rotated Δ_θ in azimuth with respect to the previous sample as illustrated in Fig. A.1-(b). Following this acquisition strategy, a set of N stereoscopic images can be extracted from the panoramas.

The centers of projection for each cylindrical panorama $\mathbf{I}_{q,k}$ ($q \in \{1, 2, 3\}$) are denoted $\mathbf{O}_{q,k}$ for the cluster C_k, where k indicates the cluster number in a set of $\mathscr{N}_s$ clusters. The three $\mathbf{O}_{q,k}$ are on the $\mathbf{XZ}$-plane to satisfy the co-planarity constraint (Fig. A.1).

This variation of the acquisition model was used to acquire six ($N = 6$) partially overlapped stereoscopic images that can be mosaicked to produce $\mathbf{I_L}$ and $\mathbf{I_R}$ with respect to the viewpoint located in $\mathbf{O}$ using the techniques proposed in Section 5.2.3. The selection of the number of images is determined by the geometry of the cluster and reduces the usable FOV in azimuth to 30° by interleaving images extracted from different panoramas. This is illustrated in Fig. 4.13.

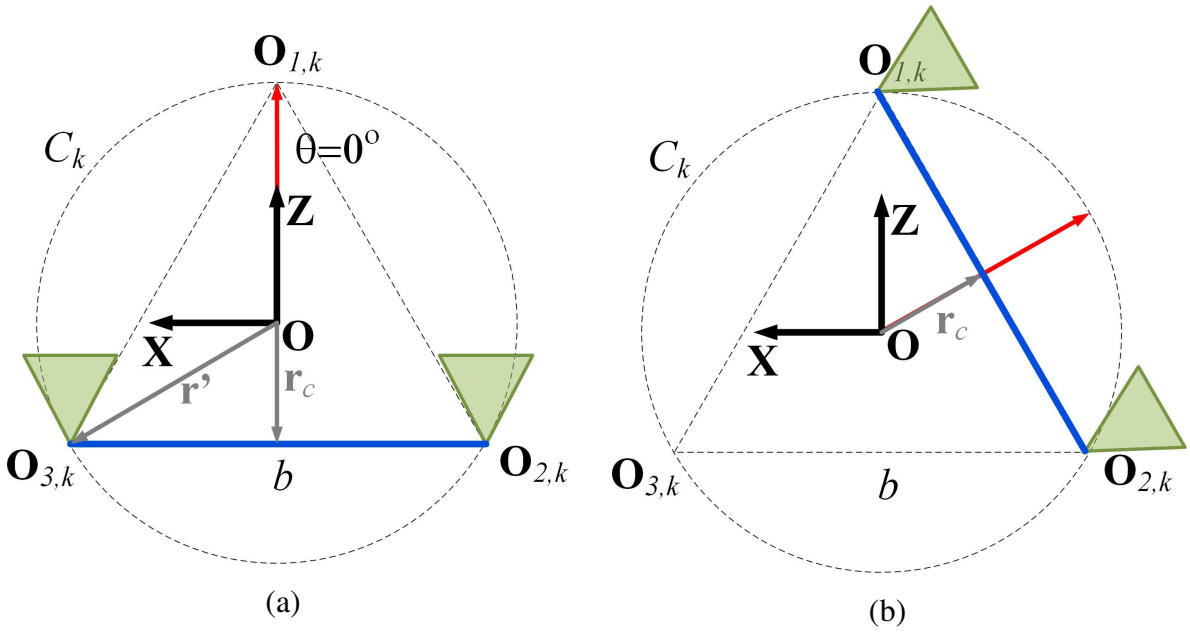


Figure A.1: Acquisition *configuration 4* to model the virtual stereoscopic camera rig that acquires $\mathbf{im}_{j,i}$, for $j = \{L, R\}$ and $i = \{0, ..., N-1\}$

A.2.1 The acquisition

The suggested acquisition scheme is depicted in Fig. A.2. A triangular sampling pattern is defined in this way, where the projection centers of $\mathbf{I}_{q,k}$ ($q \in \{1, 2, 3\}$) coincide with the vertices of an equilateral triangle with side length b as illustrated in Fig. A.1-(a).

The assumption is that a pair of panoramic images has their projection centers on the $\mathbf{XZ}$-plane and the camera vertical axis is parallel to the $\mathbf{Y}$-axis in the global reference of coordinates.

A.2.2 Telepresence application

The free-viewpoint stereoscopic navigation of a virtual environment is a good example of an application where it is necessary to sample the environment with a large number of omnistereoscopic images. The omnistereoscopic sampling of a scene using $\mathcal{N}_s$ panoramic clusters C_i as illustrated in the part **A** of Fig. A.3.

The larger $\mathcal{N}_s$, the easier it is to implement a free-viewpoint stereoscopic navigation of the scene, but at the price of increasing the acquisition time and the data overhead. On the other

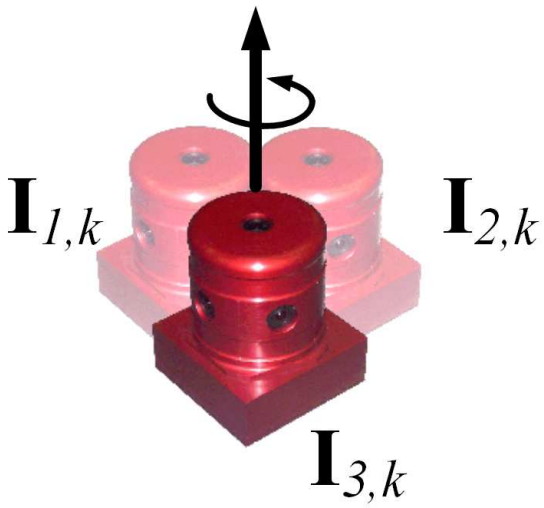


Figure A.2: Acquisition of a panoramic cluster C_k: method based on rotating a panoramic camera located at a distance r' of **O**, and taken three panoramic snapshots.

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 location are transferred and stored in a remote workstation (part **B** of Fig. A.3) to be used in the off-line rendering of stereoscopic panoramas (part **C** of Fig. A.3). 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 feedback of the gazing $\mathscr{D}$ of the user to generate the stereoscopic views (part **D** of Fig. A.3). If this feedback 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 (Section 5.2.2). The stereoscopic image pair of the scene transmitted to the user is generated based on the closest stereoscopic samples from the user's virtual location.

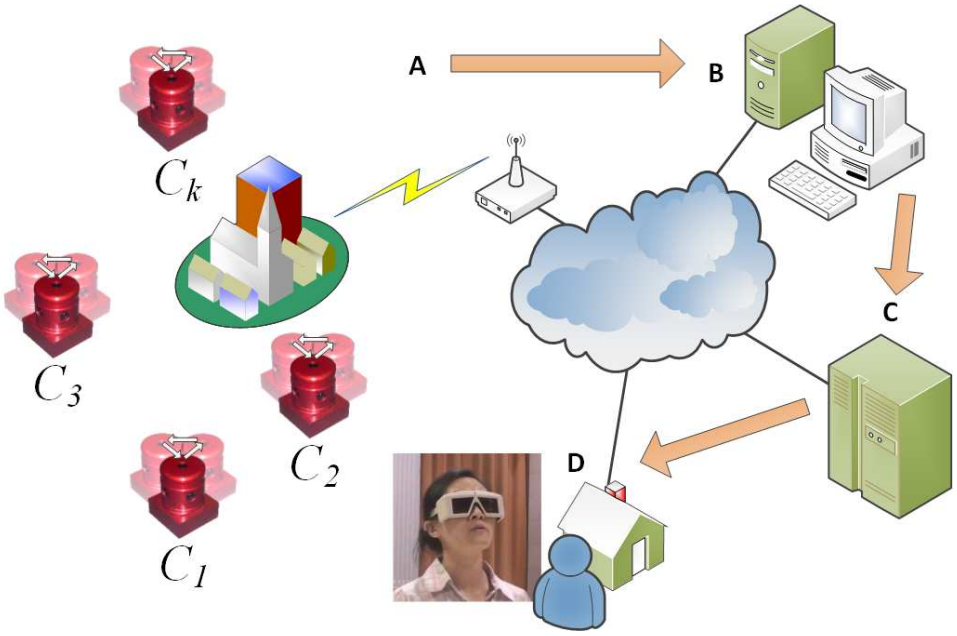


Figure A.3: Stereoscopic telepresence system: A) the remote world location is sampled with N panoramic clusters composed by panoramic triads, B) the collected sets of panoramas is stored in a server and processed off-line to generate an omnistereoscopic pair per cluster, C) these omnistereoscopic images 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.

A.2.3 The stereoscopic images

The $\mathbf{I}_{q,k}$ ($q \in \{1,2,3\}$) must be aligned to locate the stereoscopic regions in each pair of panoramas. This part has been done using the panorama alignment method described in Appendix B. The alignment process set a common reference at the center of each cylindrical image $\mathbf{I}_{q,k}$, which coincides with the direction of the $\mathbf{Z}$-axis.

After aligning the panoramas with respect to $\mathbf{XYZ}$, six stereoscopic images $\mathbf{im}_{j,i}$ can be defined. At this point, the acquisition *configuration 4* can be used to visualize these stereoscopic image pairs (Section 3.2). For instance, a virtual pair of pinhole cameras with horizontal parallax b, can be used to model the acquisition process. In Fig. A.4, the panning of this virtual stereoscopic rig helps to visualize the origin of the $\mathbf{im}_{j,i}$ to be extracted from $\mathbf{I}_{q,k}$.

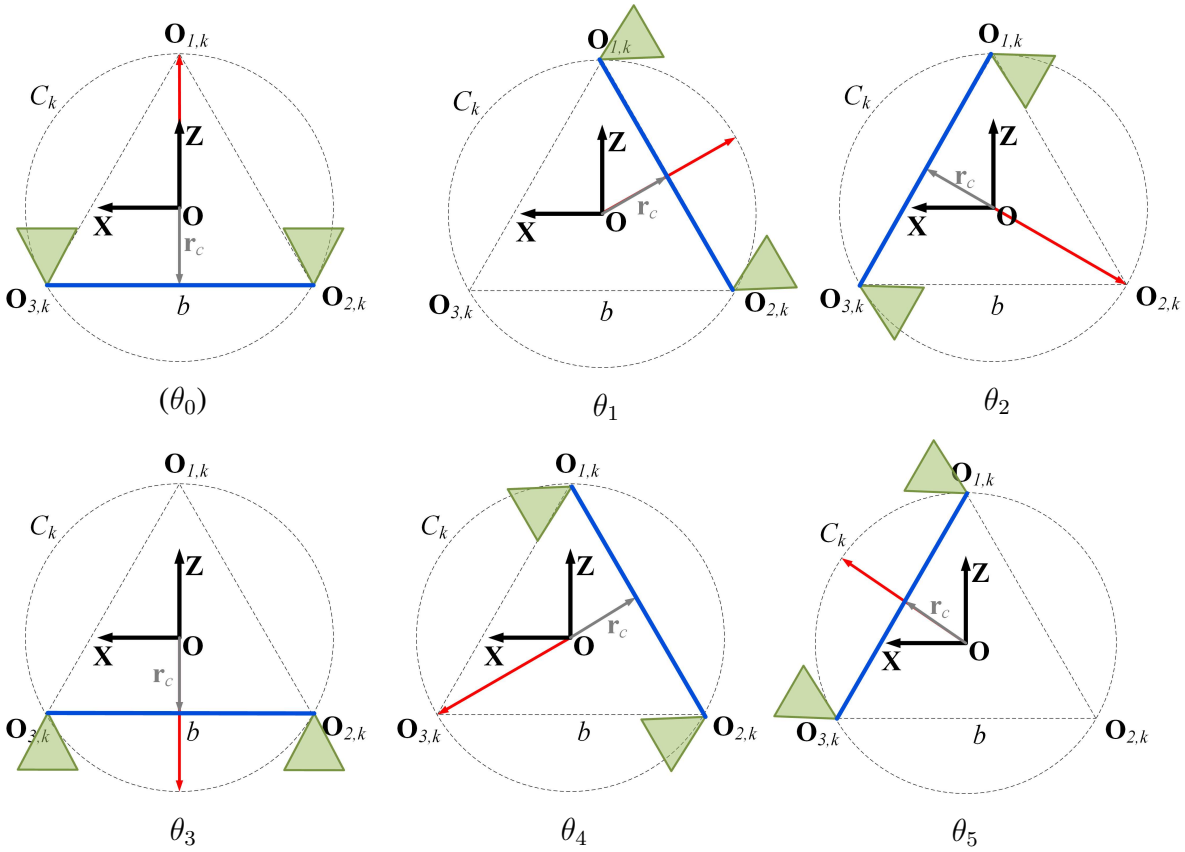


Figure A.4: The acquisition model applied to the panoramic triad method.

A.2.4 Format of the panoramic images

In order to extract each $\mathbf{im}_{j,i}$ from the cylindrical images $\mathbf{I}_{q,k}$, it is helpful to define the 2D representation of the 3D cylindrical canvas as shown in Fig. A.6. This image representation is a function $\mathbf{I}_{q,k} : Z^2 \rightarrow Z^3$. Furthermore, the panoramic images are periodic in the horizontal dimension, in other words,

$$\mathbf{I}_{q,k}(x, y) = \mathbf{I}_{q,k}(x - \mathscr{W}_h, y), \tag{A.1}$$

where in this context $\mathscr{W}_h$ is the image width in the horizontal dimension.

The panning angle θ' is defined in the reference of coordinates with center in cylindrical projection center $\mathbf{O}_{q,k}$. Since the distance between the scene and the cluster center $\mathbf{O}$ is several times the distance r' or r_c, we can consider that $\theta' = \theta$ without loss of generality. This is illustrated in Fig. A.5.

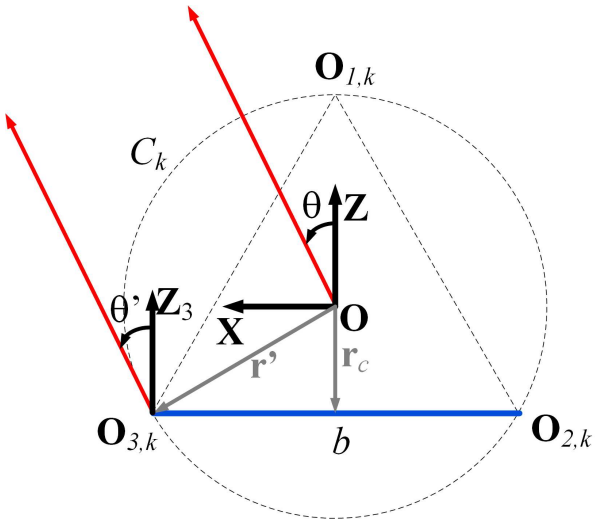


Figure A.5: Panoramic center $\mathbf{O}_{q,k}$ and cluster center $\mathbf{O}$.

A column in $\mathbf{I}_{q,k}$ is related to the viewing direction θ in cylindrical coordinates by

$$\theta = x \cdot 180/\mathscr{W}_h, \tag{A.2}$$

where $\theta \in (-180°, 180°]$ and $x \in (-\mathscr{W}_h/2, \mathscr{W}_h/2]$. This mapping is illustrated in Fig. A.6.

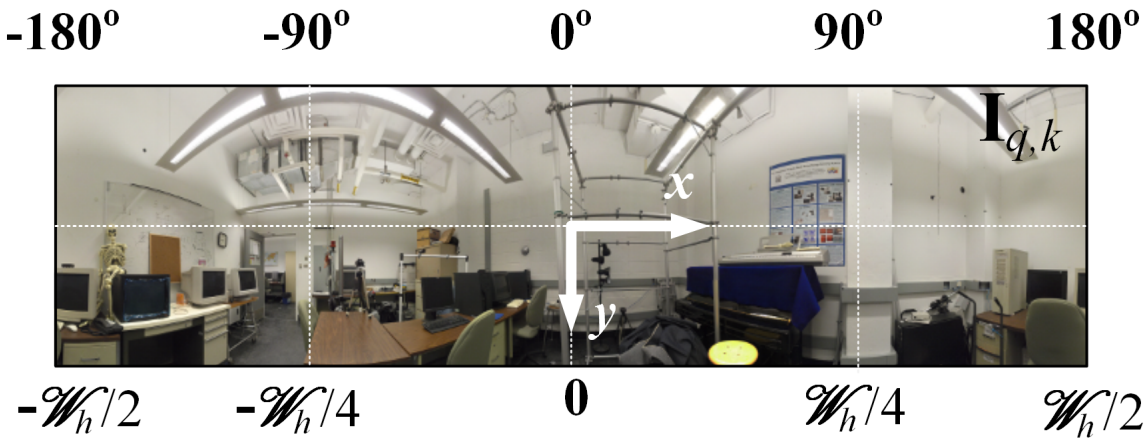


Figure A.6: Planar panoramic image $\mathbf{I}_{q,k}$.

Using this convention, an image column of the cylindrical panorama will be denoted $\mathbf{I}_{q,k}(\theta, .)$. This is equivalent to $\mathbf{I}_{q,k}(x, .)$, where x can be calculated from a panning direction θ using Eq. A.2 and the $\cdot$ represent all the vertical coordinates y. Hence, a panoramic region that spreads Δ_a in azimuth around the direction θ will be denoted $\mathbf{I}_{q,k}(\theta \pm \Delta_a/2, .)$.

A.2.5 Extracting the stereoscopic images

The mapping of $\mathbf{im}_{j,i}$ into regions of the panoramic images $\mathbf{I}_{q,k}$ can be done for each θ_i by defining the angular spread in azimuth Δ_a. In this case, six stereoscopic images should be extracted from the panoramic images, each of them has a FOV in azimuth given by $\Delta_a = 60°$. All the panoramas in C_k have a common reference center after aligning them. Therefore, the mapping of each stereoscopic image $\mathbf{im}_{j,i}$ to a region in different panoramas $\mathbf{I}_{q,k}(\theta \pm \Delta_a/2, .)$ is defined in Table A.1. This correspondence is based on the panorama labeling and clockwise progression of θ illustrated in Fig. A.4.

Table A.1: Panorama selection for the stereoscopic pair rendering

i	θ_o	$\mathbf{im}_{L,i}$	$\mathbf{im}_{R,i}$
0	0°	$\mathbf{I}_{3,k}$	$\mathbf{I}_{2,k}$
1	60°	$\mathbf{I}_{1,k}$	$\mathbf{I}_{2,k}$
2	90°	$\mathbf{I}_{1,k}$	$\mathbf{I}_{3,k}$
3	120°	$\mathbf{I}_{2,k}$	$\mathbf{I}_{3,k}$
4	180°	$\mathbf{I}_{2,k}$	$\mathbf{I}_{3,k}$
5	−120°	$\mathbf{I}_{2,k}$	$\mathbf{I}_{1,k}$
6	−60°	$\mathbf{I}_{3,k}$	$\mathbf{I}_{1,k}$

The left and right images to mosaic are defined by

$$\mathbf{im}_{L,i} = \mathbf{I}_{q_1,k}(\theta_o \pm \Delta_a/2, .), \tag{A.3}$$

$$\mathbf{im}_{R,i} = \mathbf{I}_{q_2,k}(\theta_o \pm \Delta_a/2, .), \tag{A.4}$$

where $\Delta_a = 60°$ and the indexes (q_1, q_2) indicates the order number in the cluster as defined in Table A.1.

The partially overlapped $\mathbf{im}_{j,i}$ can be mosaicked into a cylindrical panorama $\mathbf{I_S}$ using the rendering method proposed in Chapter 6 and 7.

A.3 Experimental results

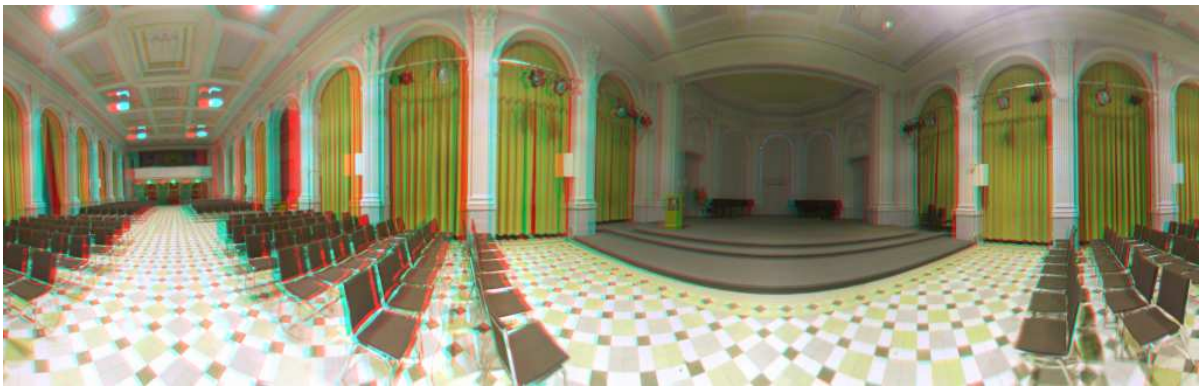
In order to test the omnistereoscopic method proposed herein, I performed several tests in a variety of indoor scenarios. To do this, we used a Ladybug2 [169] panoramic camera mounted on a tripod at 1.70m above the floor level. Considering the vertical FOV of this camera, a safe zone of ~ 2m around the cameras was required to prevent rendering problems.

The acquisition was done manually, using a triangular sampling pattern of side $b = 65$mm. In some trials, we mounted the camera off-center a distance $r' = 38$mm and captured three panoramic snapshots at regular intervals while rotating the camera. The panoramic camera software renders cylindrical panoramas of 4096 horizontal by 1304 vertical pixels by mapping the images into a cylinder of radius $R = 5$m.

The radial distortions and the parallax created by off-set sensors in the panoramic camera create distortions in each $\mathbf{I}_{q,k}$ in different directions. This distortion is commonly neglected in monoscopic panoramic snapshots, but becomes visible when comparing the scene registered in pairs of panoramas within a cluster. In order to correct this issue, I performed a fine alignment for each panoramic pair independently in order to minimize errors in the depth rendition. This procedure is detailed in Appendix B. An example of $\mathbf{I_S}$ created using this method is presented in Fig A.7.

(a) $\mathbf{I_L}$

(b) $\mathbf{I_R}$



(c) $\mathbf{I_S}$ (red-cyan anaglyph)

Figure A.7: Stereoscopic panoramas in cylindrical format (red-cyan anaglyph).

A.4 Summary

The omnistereoscopic acquisition technique presented in this appendix is an alternative for the fast acquisition and rendering of stereoscopic panoramas using a cluster of three panoramic snapshots of the scene. This method can reproduce the illusion of depth 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.

This method reduces the required data acquisition time, transmission bandwidth and storage capacity by using a limited number of panoramic samples. Furthermore, this technique is attractive for the large scale stereoscopic survey of a scene for telepresence applications.

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, preventing its usage to sample $\mathbf{C_S}$ of dynamic scenes.

Appendix B

Alignment of panoramas for best stereoscopic usability

One of the methods to acquire samples of the $\mathbf{C_S}$ function involves using clusters of panoramas in a known spatial arrangement. However, misalignments between panoramas introduced by single-shot panoramic cameras must be corrected in order to capture the depth of the scene consistently in every direction. In this regard, I propose a novel alignment method applicable to cylindrical panoramas. This is an uncalibrated method based on the identification motion flow patterns.

This technique was successfully tested in the rendering of omnistereoscopic images in different scenarios (Appendix A) and it is applicable to any off-the-shelf panoramic cameras. This technique relies on the dense optical flow between panoramic images to identify the regions of best binocular rendering.

This research has been done in the context the panoramic triad method experiments. Part of this method has been presented in the workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras (OMNIVIS 2011), part of ICCV 2011 conference [34]. The companion website can be visited online [129].

B.1 The problem

The absolute location and relative orientation of the panoramas need to be known to exploit the parallax between panoramic snapshots. The prior information relative to spatial orientation

between panoramas in a cluster must be known to identify the usable stereoscopic regions of interest as well as to keep a consistent depth rendering in every direction.

A pair of cylindrical panoramas exhibits two zones of optimum stereoscopic depth registration located in perpendicular directions in azimuth to their direction of alignment. In this context, the alignment occurs in a unique direction where, comparing the views, a focal *divergence* or *convergence* pattern is exhibited. This pattern can be identified using the optical flow between the central region in a pair of panoramas.

In a cluster of co-planar cylindrical panoramas, which are symmetrically distributed around a reference viewpoint **O**, the area of stereoscopic usability can be found in different directions in azimuth for different panoramic pairs. The correct geometric distribution of samples within a cluster enables to use these panoramic sections to compose two novel panoramic views corresponding to the left and right eyes. The key problem is to find the optimal alignment between any pair in order to maintain a consistent depth rendering in every direction.

Panoramic cameras based on multiple sensors have been increasingly used in image-based navigation applications [170]. This panoramic technology is based on mapping several overlapping wide-angle images on a 3D surface, i.e., cylinders, semi-spheres or cubes. However, the radial distortions in each image in combination with inherent parallax between sensors introduce distortions in the panoramic image in distinctive directions, which are more noticeable the closer the objects are to the camera [171]. This distortion is commonly neglected in single panoramic snapshots but becomes evident when comparing the scene registered in pairs of panoramas within a cluster.

The acquisition method presented in Appendix A uses images extracted from cylindrical panoramas to compose an omnistereoscopic image. In that acquisition method, the orientation of the panoramic camera is controlled. Hence, the alignment of the panoramic triad to a particular direction, in this case the reference **Z**-axis, can be controlled knowing the orientation of the triangular cluster with respect to the scene.

A more general problem arises when panoramic images within a cluster are acquired controlling the location of the camera in the global reference system but not its orientation. In those cases, a fine alignment correction must be done independently for each panoramic pair to minimize rendering errors in the omnistereoscopic composition.

In recent years, methods for aligning cubic panoramas have been proposed, sharing the same goal of facilitating the navigation between viewpoints in image-based virtual environments [172][173][174]. These proposals are based on epipolar geometry, which requires the

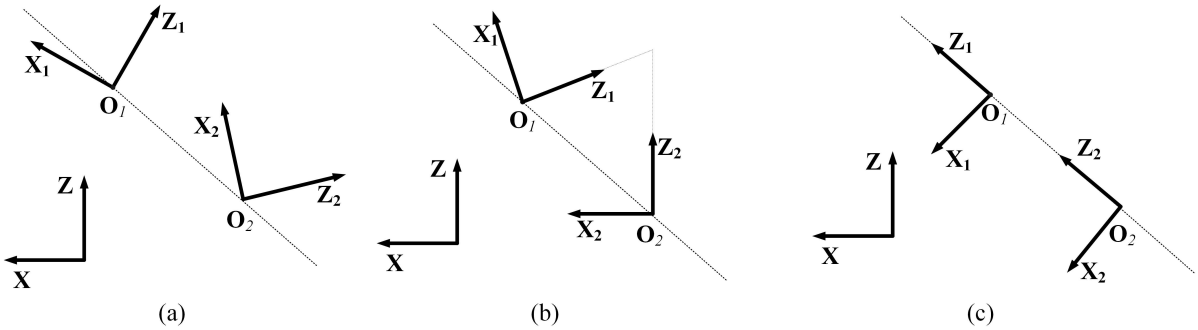


Figure B.1: (a) Misaligned panoramic pair, (b) panoramic pair aligned to an arbitrary scene element, and (c) optimum panoramic alignment for stereoscopic mosaicking; the large circles indicate the panorama locations.

estimation of calibration parameters from the rendered panoramic image assuming a pinhole camera model. These techniques were focused on equalizing the viewing direction between largely spaced panoramic samples. Although these calibrated methods provide robust solutions to the registration problem, the alignment necessary to localize the stereoscopic regions of interest is a different problem. This is illustrated with the example in Fig. B.1.

In this appendix, I present a panoramic alignment method applicable to any off-the-shelf panoramic camera in the context of the omnistereoscopic image composition. This method relies on the localization of specific disparity patterns by analyzing the dense image flow between pairs of panoramas.

B.2 The method

The alignment method is based on identifying the *focus of divergence* and *convergence* patterns in the central region of a panoramic pair of images. This solution is valid when a pair of panoramic images have their projection centers on the $\mathbf{XZ}$-plane and the camera vertical axis is parallel to the $\mathbf{Y}$-axis in the global reference of coordinates. The same constraints are valid for the panoramic cluster method presented in Appendix A.

B.2.1 The region of interest

A pair of cylindrical panoramas can be represented as $(\mathbf{I}_i, \mathbf{I}_j)$, where $\mathbf{I}_q : Z^2 \rightarrow Z^3$. The target patterns in the image flow are located in the central region-of-interest in the panoramic pair,

which is denoted as $\mathbf{R}$ in Fig. B.2. Each panoramic image exhibits periodicity in the horizontal dimension, so $\mathbf{I}_q(x, y) = \mathbf{I}_q, (x - \mathscr{W}_h, y)$, where $\mathscr{W}_h$ is the horizontal image width. This is a useful property that will be used in the development of the aligning algorithm.

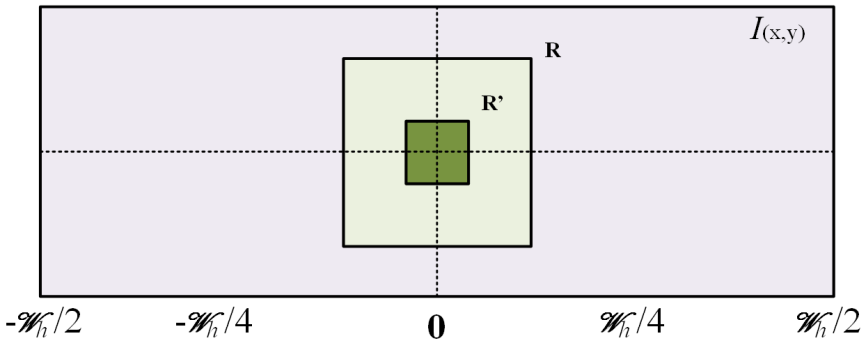


Figure B.2: Region of interest in a panoramic image pair.

It is important to note that there is a unique subset of contiguous horizontal coordinates, x or θ (Eq. A.2), where the pairwise image flow around the central region $\mathbf{R}$ exhibits the motion pattern of interest. The optimum alignment consists in finding a direction of alignment (x_a or θ_a) where a maximal image correlation is observed in the region $\mathbf{R}$. This maximal image correlation is exhibited at the horizontal coordinate x_a and at $x_a + \mathscr{W}_h/2$. After this panoramic alignment, the regions for best stereoscopic rendering will be located at $x_a \pm \mathscr{W}_h/4$, in other words, in perpendicular directions in azimuth of the *focus of convergence* and *divergence* directions in azimuth found at x_a and at $x_a + \mathscr{W}_h/2$.

The idea is illustrated in Fig. B.3 for a panoramic pair $(\mathbf{I}_i, \mathbf{I}_j)$. In this example, the panoramas are aligned at $x_a = 0$ which is the direction of $\mathbf{Z}$-axis, and the panoramic projection centers are $\mathbf{O}_i$ and $\mathbf{O}_j$. The regions $\mathbf{A}$ indicates where the patterns of *convergence* and *divergence* are exhibited. These image regions are centered at $x = 0$ and $x = \mathscr{W}_h/2$ in the 2D image representation (Fig. B.3-(right)). Consequently, the regions labeled $\mathbf{B}$, located at $x = x_a \pm \mathscr{W}_h/4$ in the 2D image representation (Fig. B.3-(right)) are the regions of optimum stereoscopic usability.

B.2.2 The optical flow patterns

In order to classify the motion flow on the central region $\mathbf{R}$, I divided it into nine non-overlapping blocks. There are four main motion flow patterns that can be detected: a perfect

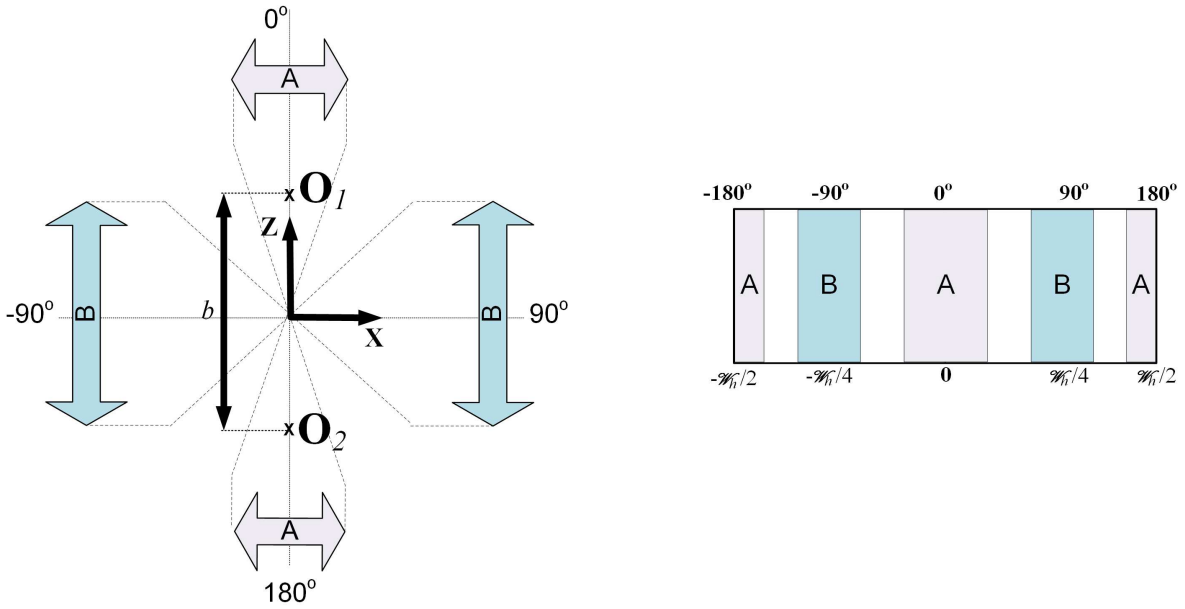


Figure B.3: A pair of optimally aligned panoramic images. The projection centers of the acquired panoramas are located at $\mathbf{O_1}$ and $\mathbf{O_2}$, aligned with the $\mathbf{Z}$-axis in the $\mathbf{XZ}$-plane. The directions of *focal convergence* and/or *divergence* ($\mathbf{A}$) and the stereoscopic regions of interest ($\mathbf{B}$) are indicated with respect to $\mathbf{XYZ}$ (left) and in 2D coordinates of $\mathbf{I_i}$ (right).

alignment (Fig. B.4-(a)), a *focal convergence*, which is indicated by the convergence of the image disparity vectors towards the center of reference (Fig. B.4-(b)), a *focal divergence*, which happen when the disparity vectors diverge from the center of the image (Fig. B.4-(c)), and *near alignment* between both images, when the pattern cannot be determined (Fig. B.4-(d)).

This pattern classification method was applied to the set of motion vectors estimated for each image block in $\mathbf{R}$. The result was effective in identified the desired patterns in the target image regions.

B.2.3 Disparity map

The dense optical flow calculated over $\mathbf{R}$ provides the dense disparity map necessary for the motion pattern identification. In this case, I used the implementation in *Matlab* of the optical flow algorithm proposed by Ogale et al. [120] to obtain the horizontal and vertical components of the disparity map. The image block size of the $\mathbf{R}$ can be defined as a percentage of the image height. In my implementation, I used a squared image block of 50% of the panoramic

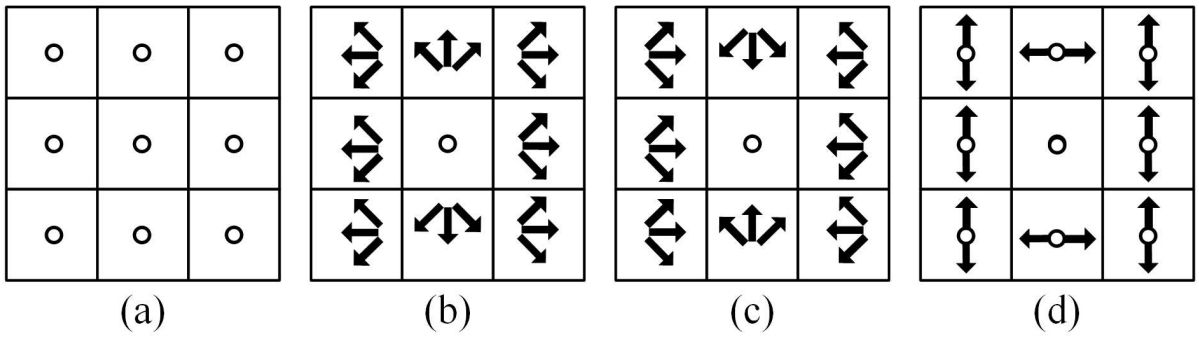


Figure B.4: Pattern classification: the circle represent a null motion vector and the arrows marks the direction of the motion flow for that block. The cases presented here are: (a) maximum image correlation on the nine blocks defined on $\mathbf{R}$, (b) focal convergence, (c) focal divergence, and (d) near alignment.

image height. The disparity map is returned in two matrices, $\mathbf{D_x}$ and $\mathbf{D_y}$, each of the same size as the region $\mathbf{R}$.

In order to determine the characteristic motion vector, I defined a disparity operator function $\Delta(\mathbf{D_x}, \mathbf{D_y})_{\mathbf{R}'}$ to calculate the frequency of occurrence of each disparity pattern in $\mathbf{R}'$. I used a voting system that used this frequency of occurrence to define a motion vector (d_x, d_y) for each block $\mathbf{R}' \in \mathbf{R}$. This motion vector was then used to classify the motion pattern on the region $\mathbf{R}$ according it proximity to any of the patterns defined in Fig. B.4.

This method to identify the characteristic motion can be improved using the feature detection method implemented in Section 7.7.

B.2.4 The panorama alignment algorithm

The algorithm to search for the direction of alignment between a pair of co-planar panoramas is presented in the block diagram of Fig. B.5. Details of the implementation can be found in [34].

B.3 Experimental results

The method was applied to a set of panoramic snapshot pairs, that I acquired using a *Ladybug 2* panoramic camera [169] to test the acquisition method proposed in Appendix A. From among these tests, two examples are presented in Fig. B.6. The example shows the horizontal

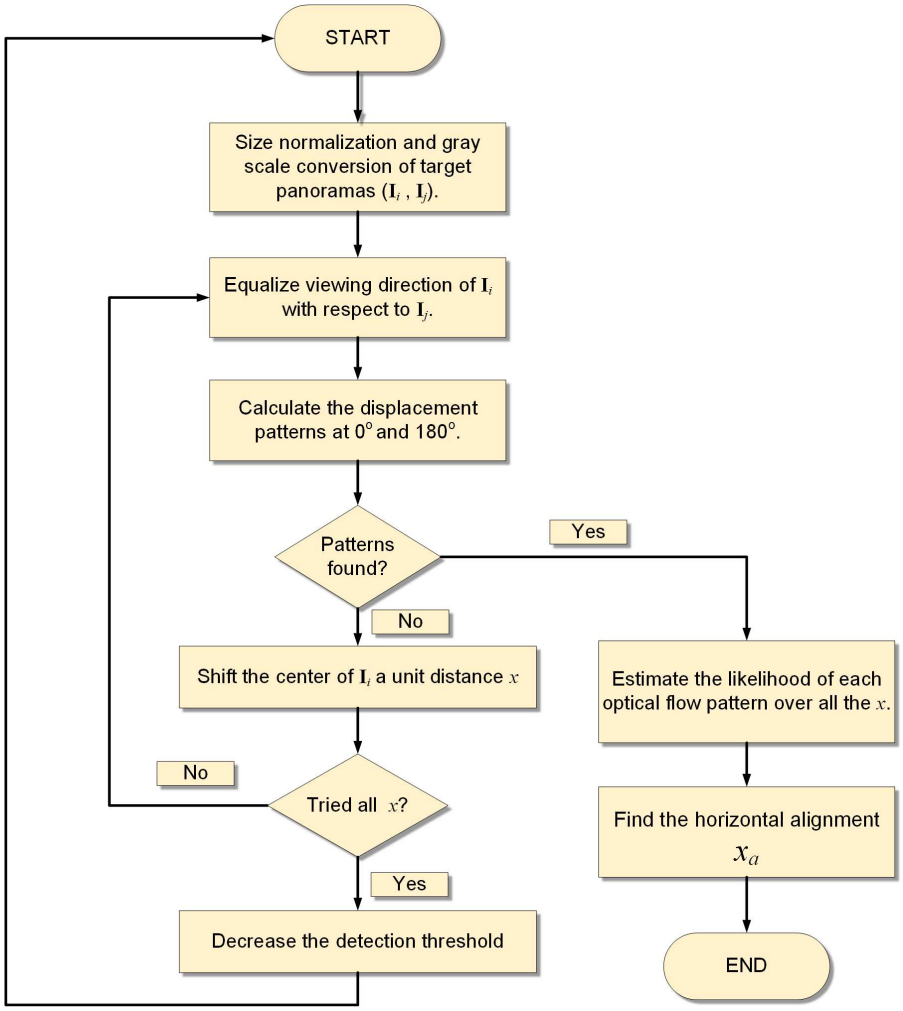
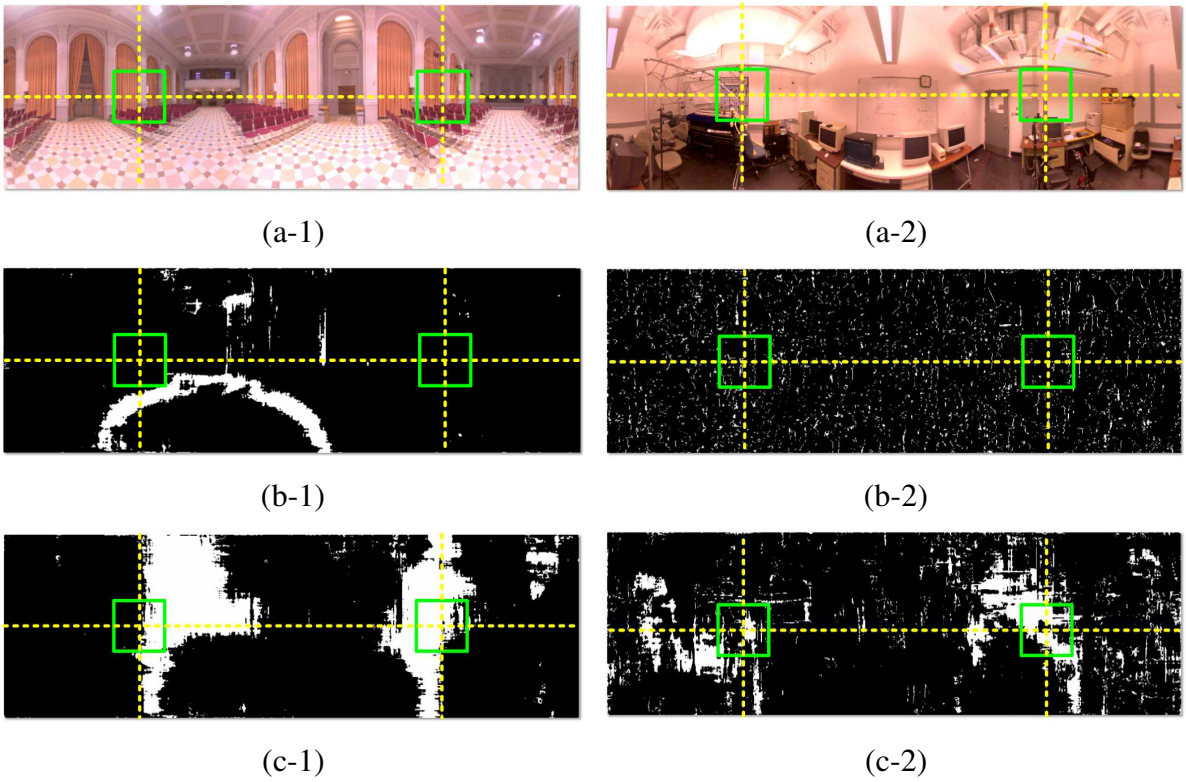


Figure B.5: Optimal alignment algorithm.

disparity maps calculated on ($\mathbf{I_i}$, $\mathbf{I_j}$) when the panoramas are aligned to an arbitrary direction in azimuth and when their are optimally aligned, exhibiting the desired image flow patterns centered around $x = \pm\mathscr{W}_h/4$. Even though three panoramic pairs per triangular cluster were aligned independently, a single pair of panoramas per cluster is enough to illustrate the effectiveness of the proposed method.

More results can be found on the companion website at:
`http://www.luisgurrieri.net/publications/cpia`



(a-1) (a-2)

(b-1) (b-2)

(c-1) (c-2)

Figure B.6: The alignment results: (a)-(b) one of the panoramas of the pair ($\mathbf{I_i}$, $\mathbf{I_j}$) to be aligned, (c)-(d) are their respective horizontal disparity maps, where the zero disparity areas are represented in white, and (e)-(f) the respective horizontal disparity map for each pair of panoramas after aligning them showing the focal convergence and divergence region at $x = \pm\mathscr{W}_h/4$.

B.4 Summary

The panorama alignment method I presented was developed as an alternative to image alignment techniques that rely on calibrated methods. The main problem of using a calibrated approach when the images are extracted from panoramic imagery is that the assumed projective model is not necessary valid. For instance, a panorama can be rendered by projecting and mosaicking a multiplicity of partially overlapped images onto a warped surface. In the process of rendering, several distortions may have been introduced into the final panorama in order to reduce or eliminate visual artifact. Hence, the camera intrinsic and extrinsic parameters estimated based on a planar image extracted from the panorama, emulating a pinhole camera model, may not be valid for some panoramic cameras.

In that context, I proposed novel technique for the optimal alignment of panoramas based on the information that can be extracted from the panoramic imagery. I conceived this method for the identification of the region of best stereoscopic rendering (Appendix A). The main advantage of this approach resides in its simplicity, making it suitable to the processing of the large numbers of panoramas needed in massive omnistereoscopic surveys of scenes intended for navigation.

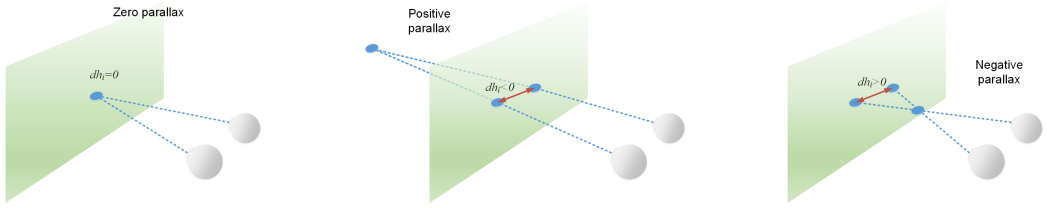
However, the alignment technique can be improved by using more robust methods to estimate the characteristic motion vector of a block in the image pair. For instance, this can be improved by finding a set of corresponding coordinates $\mathscr{K}_i$ using the robust feature identification techniques implemented in Section 6.1. Another disadvantage of this method is its dependency of the image quality. For instance, poorly illuminated regions in certain directions $\mathscr{D}$ or uniformly textured regions are challenging to classify.

Appendix C

Displaying stereoscopic images

The existing convention for displaying stereoscopic images was originated in the stereoscopic cinema industry [175]. In this convention, a scene element is said to have *zero parallax* when left and right images perfectly overlap in the display, i.e., they exhibit zero disparity (Fig. C.1-(a)). This effect conveys the illusion that the object is at the same distance as the projection canvas surface (screen), since the eyes need the same vergence angle to converge on the virtual scene than to converge on the screen. In other words: the scene is at the same distance than the screen is from the user point of view. Using the zero parallax vergence as a reference, a scene element that requires less vergence of the eyes is said to have *positive parallax* (Fig. C.1-(b)), conveying the illusion of *being* behind the screen or farther way than the screen is from the user reference point. Conversely, an object is said to have *negative parallax* when the eyes need more inward eccentricity inwards (towards the median plane defined between the eyes and parallel to the head axis) than in the zero parallax case (Fig. C.1-(c)), conveying the illusion that the object is closer than the screen surface with respect to the user reference point.

The acquisition model relies on stereoscopic rigs with parallel axes to acquire partial stereoscopic views of the scene to produce an omnistereoscopic image $\mathbf{I_S}$. If displayed as captured, the zero parallax will correspond to the farther elements in the scene, while the rest of the scene would exhibit negative parallax. In other words, the scene will appear mostly in negative space (in front of the screen) in all directions. However, the desired illusion is the opposite: the scene *begins* at a screen distance and *unveils* behind the screen. In other words, the display screen should be the window frame that marks where the 3D scene begins with respect to the user reference viewpoint.



(a) (b) (c)

Figure C.1: Stereoscopic displaying convention versus dh_i after applying the cross-eye inversion: (a) zero parallax $(dh_i = 0)$, (b) positive parallax $(dh_i < 0)$, and (c) negative parallax $(dh_i > 0)$.

Some elements in the scene may remain be rendered in negative space, so they will convey the illusion of being closer than the display screen. This effect is known as *pop-up stereo*. Although this may be attractive for some applications, this effect should be avoided because it would impose unnecessary stress to the binocular vision system [176]. This is particularly true in immersive stereoscopic scenarios where the users are supposed to interact with a virtual environment for extended periods of time [177].

In the context of stereoscopic panoramas, I choose a determined point in the image to be displayed with zero disparity (zero parallax), which is usually the closest scene element imaged by the omnistereoscopic camera. A *horizontal translation* applied to $\mathbf{I_R}$ or $\mathbf{I_L}$ equal to the maximum negative disparity observed in $\mathbf{I_S}$ is enough in this application to set the zero disparity. Note this image transformation applied to the $\mathbf{I_R}$ or $\mathbf{I_L}$ is a circular shifting of the images given that they are omnidirectional views of the scene in azimuth. Finally, I must add that, adapting $\mathbf{I_S}$ for different displaying scenarios, a process also known as *image retargeting*, requires to adapt the stereoscopic image by using image warping techniques [178].

Appendix D

Acronyms and Abbreviations

D.1 Typographic conventions

- Scalar variables, indexes and functions are represented in italic font, i.e., a, b, c_i.
- Vector variables or functions are represented in bold font, i.e., $\mathbf{C_S}$, $\mathbf{T}_{j,i}$, $\mathbf{R}_i$. These magnitudes can be row or column vectors and also matrices.
- The subindex $j \in \{L, R\}$ denotes the left (L) or right (R) member of a stereoscopic pair, i.e., $\mathbf{r}_j$ represents the spatial location of the j camera in a stereoscopic pair of cameras as a vector in 3D space.
- The subindex $i \in \{0, ..., N-1\}$ denotes the position in an ordinal set of images, i.e., the angle θ_i is the i-angle value in set of N given by the series $i \in 0, ...i, i+1, ..., N-1$.
- The image $\mathbf{im}_{j,i}$ is the j image (left of right) corresponding to the gazing direction θ_i in azimuth.
- In situations where the subindex does not denote the order in a stereoscopic pair and/or the order in set of images, bold or regular fonts are used for vector or scalar variables or functions, respectively. For instance, $\mathbf{P_W}$, c_n.
- The symbolic representation of elements in a pair uses italic fonts, i.e., the camera $\Omega_{j,i}$.
- The Euclidean distance defined by a vector norm is indicated with the same symbol used to name the vector, but in italic fonts, e.g., $r_c = \|\mathbf{r_c}\|$.

- The style used for the acronyms that denoted two-dimensional and three-dimensional are 2D and 3D, respectively, omitting the hyphen [179]. Their usage in compound nouns may or not maintain the hyphen, i.e., 3D-space, 2D-to-3D, but not 3DTV.

D.2 List of acronyms and definitions

2D Two-dimensional.

3D Three-dimensional.

2D-to-3D Two-dimensional to three-dimensional conversion.

3DTV Three-dimensional television.

AoA Angle-of-Arrival (See gazing direction $\mathscr{D}$).

APS-C Advanced Photo System type-C (APS-C) image sensor.

arcsec Second of arc, defined as $\frac{1}{3600}$ of a degree, or $\frac{\pi}{648000}$ of a radian, e.g., $9.6963e-05$ rad.

aspherical A type of lens whose FOV varies slightly from being spherical. For instance, commercially available *fish-eye* lenses have aspherical FOV.

azimuth Horizontal direction expressed as the angular distance between the direction of a fixed point (as the observer's heading) and the direction of the object [180].

DOF Depth-of-field.

DSLR Digital single-lens reflex camera.

elevation The angle formed by the line of sight and the horizontal plane for an object above the horizontal [180].

Exposure time the exposure time or shutter speed is the time length an image sensor is exposed to light from the scene.

fish-eye Type of lens characterized for a wide-angle FOV, which typically is in the order of $180°$.

FOV Field-of-view $\mathscr{F}$ of a lens measured in the longer diagonal of the image sensor following the convention of photographic equipment. Unless specifically indicated in the text, the FOV of a lens in this context is the angle-of-vision in azimuth.

f-stop In photography the f-stop of a camera is f/D, where D is the diameter of the diaphragm that controls the amount of light that is allowed to illuminate the photographic sensor.

FVTV Free-viewpoint television.

HDTV High definition television.

LIDAR Light detection and ranging.

non-SVP Non-singular viewpoint.

RANSAC Random sample consensus algorithm.

ROI Region-of-interest.

SURF Speeded up robust features algorithm.

Stereo acuity See vergence acuity.

SVP Singular viewpoint.

TV Television.

TIFF Tagged image file format.

Vergence acuity The ability to resolve to scene elements at different depths based on minimum difference between vergence angles of the eyes. In other words, if two scene points $\mathbf{P_W}$ and $\mathbf{P_W}'$ are separated by the minimum distance resolvable by our binocular vision, then they are imaged in the eyes central foveal area by vergence angles ω and ω', respectively, which defines a the vergence acuity given by $|\omega - \omega'|$.

D.3 List of symbols

$\alpha_{h,i}$ Horizontal angular disparity between a corresponding point set $\mathscr{K}_i$.

$\alpha_{v,i}$ Vertical angular disparity between a corresponding point set $\mathscr{K}_i$.

$\mathscr{A}_{j,(i,i+1)}$ Alignment operation that enables matching corresponding coordinates from the set $\mathscr{K}_{j,i}$.

a_r Aspect ratio.

a_0, a_1 Parameters to define the line $\mathbf{L}_{j,i}$

$\mathscr{A}_{h,i}$ Horizontal disparity error (angular) between any two picture elements with coordinates $\mathbf{p}_{j,i}$ and $\mathbf{p'}_{j,i}$.

$\mathscr{A}_{v,i}$ Vertical disparity error (angular) between any two picture elements with coordinates $\mathbf{p}_{j,i}$ and $\mathbf{p'}_{j,i}$.

$a_{h,i}$ Horizontal disparity error (angular) defined by the absolute difference between $\alpha_{h,i}(\mathbf{p}_{j,i})$ and $\alpha_{h,i}(\mathbf{p'}_{j,i})$.

b Length of the stereoscopic baseline.

$\mathbf{b}$ Baseline vector defined by $\mathbf{r}_L - \mathbf{r}_R$.

$\hat{\mathbf{b}}$ Unit vector that defines the baseline direction on the $\mathbf{XZ}$-plane.

$\mathscr{C}$ Color space.

$\mathbf{C}$ Omnidirectional viewing function.

$\mathscr{C}_y$ Transformation from planar to cylindrical coordinates in 2D space.

C_i The i-th cluster in a set of $\mathscr{N}_s$ clusters.

c_n Given a color space, power of light arriving at a point r from a given direction in space as a function of time, for the n color space component.

$c_{L,n}$ **and** $c_{R,n}$ The n tristimulus value $n = \{1, 2, 3\}$, output of the viewing function $\mathbf{C_L}$ or $\mathbf{C_R}$, respectively.

$\mathbf{C_S}$ Stereoscopic viewing function.

$\mathbf{C}_j$ The viewing function corresponding to the j viewpoint, where $j \in \{L, R\}$.

δ_θ Vergence resolution or stereo acuity in humans which is below 20 arcsec ($4.8481e-06$ rad) in the majority of the adult population (a widely used value is $\delta_\theta = 20$ arcsec).

Δ Field-of-view in the diagonal of the image sensor.

Δ_a Field-of-view in azimuth [deg].

Δ_b Width of the blending region in the horizontal dimension.

Δ_e Field-of-view in elevation [deg].

Δ_t Time interval, i.e., exposure time in this context.

Δ_x Deviation from x_b in the horizontal dimension.

$\Delta(\mathbf{D_x}, \mathbf{D_y})_{\mathbf{R}}$ Disparity operator over a central region $\mathbf{R}$.

$\mathbf{D_x}$ Horizontal disparity map.

$\mathbf{D_y}$ Vertical disparity map.

$\mathscr{D}$ Gazing direction defined by the azimuth (θ) and elevation (ϕ) angles.

$\mathscr{D}_L$ Gazing direction with respect to the left viewpoint $\mathbf{e_L}$.

$\mathscr{D}_R$ Gazing direction of the right viewpoint.

d_e Interocular distance in the average adult population (65 mm).

$d_{h,i}$ Horizontal (Euclidean) disparity between a corresponding point set $\mathscr{K}_i$.

$d_{v,i}$ Vertical (Euclidean) disparity between a corresponding point set $\mathscr{K}_i$.

$dv_{j,i}$ The vertical disparities between neighbor images $(\mathbf{im}_{j,i}, \mathbf{im}_{j,i+1})$ defined between a coordinate point $\mathbf{p}_{j,i}$ and $\mathbf{p}_{j,i+1}$, for $j = L$ or $j = R$.

ε Tolerable horizontal disparity error.

$\mathscr{E}_{h,i}$ Horizontal disparity error (Euclidean) between any two picture elements with coordinates $\mathbf{p}_{j,i}$ and $\mathbf{p'}_{j,i}$.

$\mathscr{E}_{v,i}$ Vertical disparity error (Euclidean) between any two picture elements with coordinates $\mathbf{p}_{j,i}$ and $\mathbf{p'}_{j,i}$.

$e_{h,i}$ Horizontal disparity error (Euclidean) defined by the absolute difference between $d_{h,i}(\mathbf{p}_{j,i})$ and $d_{h,i}(\mathbf{p'}_{j,i})$.

ϕ Elevation angle (panning) in spherical or cylindrical coordinates.

Φ_n The n-th color matching function in a color space $\mathscr{C}$.

$\mathbf{H}_{j,(i,i+1)}$ Homography between planar coordinates defined on $\mathbf{im}^p_{j,i}$ and $\mathbf{im}_{j,i+1}$

$\mathbf{I}$ Image luminance component to use in blending.

$\mathscr{I}_S$ Omnistereoscopic image, $\mathscr{I}_S : (R)^2 \rightarrow (R)^6$.

$\mathbf{I_S}$ Discrete valued omnistereoscopic image, $\mathbf{I_S} = (\mathbf{I_L}, \mathbf{I_R})$, , $\mathbf{I_R} : (Z)^2 \rightarrow (Z)^6$.

$\mathbf{I_L}$ Discrete valued omnidirectional image image pair corresponding to the left-eye viewpoint, $\mathbf{I_L} : (Z)^2 \rightarrow (Z)^3$.

$\mathbf{I_R}$ Discrete valued omnidirectional image image pair corresponding to the right-eye viewpoint, $\mathbf{I_R} : (Z)^2 \rightarrow (Z)^3$.

$\mathbf{I}_{q,k}$ The q cylindrical panorama in the k cluster of three panoramas, where $q \in \{1, 2, 3\}$ and $k \in \{1, ..., \mathscr{N}_s$.

$\mathbf{im}_{j,i}$ Stereoscopic image pair where $j \in L, R$ and $i \in \{0, ..., N-1\}$.

$\mathbf{im}^p_{j,i}$ Planar projection of the stereoscopic image pair after correcting the lens distortion.

j Subindex that indicates the camera order in an stereoscopic pair, e.g., $j = L$ for the left side camera and $j = R$ for the right side camera.

$\mathscr{K}_i$ Corresponding coordinate points $\mathbf{p}_{j,i}$ arising when the same point $\mathbf{P_W}$, in the stereoscopic FOV of cameras $\Omega_{j,i}$, is projected onto the images $\mathbf{im}_{j,i}$.

$\mathscr{K}_{j,i}$ The corresponding coordinate points $\mathbf{p}_{j,i}$ and $\mathbf{p}_{j,i+1}$ of the projection of $\mathbf{P_W}$ on $\mathbf{im}_{j,i}$ and $\mathbf{im}_{j,i+1}$, respectively.

λ Light wavelength.

λ_{min} The shorter light wavelength in the visible spectrum (400nm).

λ_{max} The longest light wavelength in the visible spectrum (700nm).

$\mathbf{L}_{j,i}$ A ray of light modeled as a line between a point in the scene $\mathbf{P_W}$ and a camera projection center $\mathbf{O_{j,i}}$.

$\mathscr{M}$ The dense omnidirectional depth map of the scene ($\mathscr{M} : \mathbb{R}^2 \rightarrow \mathbb{R}$).

$\mathscr{N}$ Number of band in a Gaussian pyramidal decomposition.

$\mathbf{N}_s$ Number of omnistereoscopic samples using the panoramic triad method.

nadir Opposite direction of *zenith* defined by $\phi = 180°$.

$\Omega_{j,i}$ The j camera in a stereoscopic rig whose azimuth gazing direction is θ_i with respect to the omnistereoscopic camera frame $\mathbf{XYZ}$.

$\mathbf{O}_j$ The j side reference viewpoint in an stereoscopic pair, where $j \in \{L, R\}$.

$\mathbf{O}_{j,i}$ The location of j reference viewpoint in an stereoscopic pair oriented to the panning angle θ_i, where $j \in \{L, R\}$ and $i \in \{0, ..., N-1\}$.

$\mathscr{P}_{j,i}$ A vector defined as $(\mathscr{X}_{j,i}, \mathscr{Y}_{j,i}, \mathscr{Z}_{j,i})^T$

$\mathscr{P}$ Plenoptic function.

$\mathbf{P}_{j,i}$ The coordinates of the scene point $\mathbf{P_W}$ expressed in the camera Ω_{j_i} frame of coordinates.

$\mathbf{p}_{j,i}$ The projection of $\mathbf{P_{j_i}}$ into the image $im_{j,i}$ as a point with coordinates $(x_{j,i}, y_{j,i})$ in the image reference of coordinates.

$\mathbf{p}^c_{j,i}$ Coordinates in the 2D-representation of a 3D-cylindrical canvas of a coordinate point $\mathbf{p_{j_i}}$ on the image plane.

θ Azimuthal angle (panning) in spherical coordinates.

θ_i The i-th sampling angle in azimuth from an ordered set $\{0, ..., N-1\}$, measured with respect to a reference frame $\mathbf{XYZ}$.

$\mathscr{Q}$ Linear blending function.

ϱ_{min} Effective minimum distance to the scene for omnistereoscopic rendering for a given x_b and a width Δ_x.

$\mathscr{R}$ Central region of the cylindrical image.

$\mathbf{r}$ The vector viewpoint location with respect to a $\mathbf{XYZ}$ reference of coordinates in a viewing model. Also the vector reference point with respect to a $\mathbf{XYZ}$ reference of coordinates in a stereoscopic viewing function.

$\mathbf{r}_L$ The vector location of the left viewpoint with respect to a $\mathbf{XYZ}$ reference of coordinates in a stereoscopic viewing model.

$\mathbf{r}_R$ The vector location of the left viewpoint with respect to a $\mathbf{XYZ}$ reference of coordinates in a stereoscopic viewing model.

$\mathbb{S}^2$ The unit sphere in 3D-space $\mathbb{S}^2 = \{\mathbf{r} \in \mathbb{R}^3 : \|\mathbf{r}\| = 1\}$. It can be expressed as a 3D-vector vector $\mathbf{r} = (x, y, z)$ in cartesian coordinates or as a 2D-vector in polar coordinates $\mathbf{r} = (\theta, \phi)$, where $\|\mathbf{r}\| = 1$.

s Pixel width in the horizontal dimension.

t Time variable.

$\mathbf{u}$ Input vector containing the parameters necessary to calculate $\mathbf{C}$.

v Distance between the reference viewing point $\mathbf{r}$ and the scene $\mathbf{P_W}$.

$\mathbf{u}_j$ Vector containing the input parameters of $\mathbf{C}_j$, where $j \in \{L, R\}$.

$\mathbf{u_S}$ Input vector containing the parameters necessary to calculate $\mathbf{C_S}$.

$\hat{\mathbf{v}}$ Unit vector indicating the viewing direction $\mathscr{D}$.

$\mathbf{v}_{j,i}$ Input vector of $\mathbf{C}_j$, which is defined as $\mathbf{v}_{j,i} = (\theta_i, \phi, \mathbf{r}_j, t)^T$.

$\mathscr{W}$ Warping transformation used to generate a synthetic stereoscopic view based on an image and its dense depth map. The transformation is used to displace each pixel a defined horizontal amount.

$\mathscr{W}_h$ The width in the horizontal dimension of the panoramic image $\mathbf{I}_{q,k}$.

W Blending width in the horizontal dimension.

W_h Sensor width [m].

W_v Sensor height [m].

x_b Stitching coordinate in the horizontal dimension defined on $\mathbf{im}_{j,i}$ for counterclockwise mosaicking.

$\mathscr{X}_{j,i}$ A magnitude defined as $-f\ X_{j,i}$.

$x_{j,i}$ The x image coordinate in $\mathbf{im}_{j,i}$.

$\mathbf{XYZ}$ Global reference of coordinates.

$\mathbf{X'Y'Z'}$ Reference of coordinate at r, whose axes are parallel to the axes of $\mathbf{XYZ}$.

$\mathbf{X}_{j,i}\mathbf{Y}_{j,i}\mathbf{Z}_{j,i}$ Reference of coordinate centered at the projection center of camera $\Omega_{j,i}$, whose axes are parallel to $\mathbf{XYZ}$ axes, i.e., the $\mathbf{X}_{j,i}\mathbf{Z}_{j,i}$-plane and the $\mathbf{XZ}$-plane are the same and $\mathbf{Y}_{j,i}$ is parallel to $\mathbf{Y}$ and has the same orientation.

XZ-plane Reference floor plane defined on $\mathbf{XYZ}$.

X'Z'-plane Reference floor plane defined on $\mathbf{X'Y'Z'}$, parallel to the $\mathbf{XZ}$-plane.

nadir Direction defined by $\phi = 180°$.

$\mathscr{Y}_{j,i}$ A magnitude defined as $-f\ Y_{j,i}$.

$y_{j,i}$ The y image coordinate in $\mathbf{im}_{j,i}$.

$\mathscr{Z}_{j,i}$ A magnitude defined as $Z_{j,i}$.

zenith Direction defined by $\phi = 0°$.

Bibliography

[1] A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen, and C. Zhang, "Multiview imaging and 3DTV," *Signal Processing Magazine, IEEE*, vol. 24, no. 6, pp. 10–21, 2007.

[2] L. E. Gurrieri and E. Dubois, "Efficient panoramic sampling of real-world environments for image-based stereoscopic telepresence," in *Proc. SPIE 8288, Stereoscopic Displays and Applications XXIII*, Feb. 2012, p. 82882D.

[3] Google Inc., "Street view: Explore the world at street level," Website, 2012. [Online]. Available: http://maps.google.ca/intl/en/help/maps/streetview/

[4] MapJack, "Mapjack: image-based virtual navigation of streets and cities," Website, 2012. [Online]. Available: http://www.mapjack.com/

[5] EveryScape, "Everyscape: virtual tours of indoor environments," Website, 2012. [Online]. Available: http://www.everyscape.com/

[6] M. Uyttendaele, A. Criminisi, S. B. Kang, S. Winder, R. Szeliski, and R. Hartley, "Image-based interactive exploration of real-world environments," *IEEE Computer Graphics and Applications*, pp. 52–63, 2004.

[7] M. Tanimoto, "FTV (free viewpoint television) creating ray-based image engineering," in *IEEE International Conference on Image Processing*, vol. 2, Sep. 2005, pp. 25–32.

[8] D. Southwell, A. Basu, M. Fiala, and J. Reyda, "Panoramic stereo," in *Proc. IEEE Int. Conf. Pattern Recognition*, vol. 1, 1996, pp. 378–382.

[9] J. Gluckman, S. Nayar, and K. Thoresz, "Real-time omnidirectional and panoramic stereo," in *Proc. of DARPA Image Understanding Workshop*, vol. 1, 1998, pp. 299–303.

[10] T. Kawanishi, K. Yamazawa, H. Iwasa, H. Takemura, and N. Yokoya, "Generation of high-resolution stereo panoramic images by omnidirectional imaging sensor using hexagonal pyramidal mirrors," in *Proc. of the 14th International Conference on Pattern Recognition*, 1998, pp. 485–489.

[11] H.-C. Huang and Y.-P. Hung, "Panoramic stereo imaging system with automatic disparity warping and seaming," *Graphical Models and Image Processing*, vol. 60, no. 3, pp. 196–208, 1998.

[12] H. Ishiguro, M. Yamamoto, and S. Tsuji, "Omni-directional stereo," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no. 2, pp. 257–262, 1992.

[13] S. Peleg and M. Ben-Ezra, "Stereo panorama with a single camera," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, June 1999, pp. 395–401.

[14] V. Vanijja and S. Horiguchi, "A stereoscopic image-based approach to virtual environment navigation," *The Computer, the Internet and Management*, vol. 14, no. 2, pp. 68–81, Aug. 2006.

[15] S. Tzavidas and A. Katsaggelos, "Multicamera setup for generating stereo panoramic video," in *Proc. SPIE Three-Dimensional Image Capture and Applications V*, vol. 4661, 2002, pp. 47–58.

[16] Fraunhofer Heinrich Herzt Institute (HHI), "2D&3D OmniCam: end-to-end system for the real-time acquisition of video panoramas," Website, 2012. [Online]. Available: http://www.hhi.fraunhofer.de/en/departments/image-processing/applications/

[17] R. G. Baker, F. A. Baker, and J. A. Conellan, "Panoramic stereoscopic camera," U.S. Patent 2008/0 298 674 A1, 2008.

[18] L. E. Gurrieri and E. Dubois, "Stereoscopic cameras for the real-time acquisition of panoramic 3D images and videos," in *Proc. SPIE 8648, Stereoscopic Displays and Applications XXIV*, Feb. 2013, p. 86481W.

[19] V. Couture and S. Roy, "The omnipolar camera: A new approach to stereo immersive capture," in *IEEE Int. Conf. on Computational Photography*, 2013, pp. 1–9.

[20] F. Huang, S. Wei, and R. Klette, "Geometrical fundamentals of polycentric panoramas," in *Proc. 8th IEEE International Conference on Computer Vision*, vol. 1. Los Alamitos, CA, USA: IEEE Comput. Soc., July 2001, pp. 560–565.

[21] F. Huang, R. Klette, and K. Scheibe, *Panoramic Imaging: Sensor-Line Cameras and Laser Range-Finders*. Hoboken, NJ: John Wiley & Sons Inc., 2008.

[22] V. Couture, M. S. Langer, and S. Roy, "Analysis of disparity distortions in omnistereoscopic displays," *ACM Transactions on Applied Perception (TAP)*, vol. 7, pp. 25:1–25:13, July 2010, ACM ID: 1823743.

[23] P. Bourke, "Capturing omni-directional stereoscopic spherical projections with a single camera," in *IEEE 16th International Conference on Virtual Systems and Multimedia (VSMM)*, Seoul, Korea, Oct. 2010, pp. 179–183.

[24] E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision," *Computational models of visual processing*, vol. 1, pp. 3–20, 1991.

[25] G. Wetzstein, I. Ihrke, D. Lanman, and W. Heidrich, "Computational plenoptic imaging," *Computer Graphics Forum*, vol. 30, no. 8, pp. 2397–2426, Dec. 2011.

[26] E. Dubois, *The structure and properties of color spaces and the representation of color images*, ser. Synthesis Lectures on Image, Video, and Multimedia Processing. Morgan & Claypool, 2009, vol. 4, no. 1.

[27] L. E. Gurrieri and E. Dubois, "Acquisition of omnidirectional stereoscopic images and videos of dynamic scenes: a review," *SPIE Journal of Electronic Imaging*, vol. 22, no. 3, pp. 030 902–030 902, July 2013.

[28] R. O. Reynolds, "Design of a stereo multispectral CCD camera for Mars Pathfinder," in *Proceedings of SPIE*, vol. 2542, San Diego, CA, USA, 1995, pp. 197–206.

[29] F. Hongfei, J. Jinyuan, W. Hongkai, and T. Luchen, "Immersive roaming of stereoscopic panorama," in *Proc. Int. Conf. on Cyberworlds*, ser. 2008 International Conference on Cyberworlds. Hangzhou, China: IEEE, 2008, pp. 377–82.

[30] F. Huang and Z.-H. Lin, "Stereo panorama imaging and display for 3D VR system," in *IEEE Congress on Image and Signal Processing*, vol. 3, 2008, pp. 796–800.

[31] S. Peleg and M. Ben-Ezra, "Stereo panorama with a single camera," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, vol. 1, June 1999, pp. 395–401.

[32] K. Yamada, T. Ichikawa, T. Naemura, K. Aizawa, and T. Saito, "Generation of high-quality stereo panoramas using a three-camera panorama capturing system," *Kyokai Joho Imeji Zasshi/Journal of the Institute of Image Information and Television Engineers*, vol. 55, no. 1, pp. 151–158, 2001.

[33] ——, "Structure analysis of natural scenes using census transform and region competition," in *Proceedings of SPIE*, vol. 4310, San Jose, CA, USA, 2000, pp. 228–237.

[34] L. E. Gurrieri and E. Dubois, "Optimum alignment of panoramic images for stereoscopic navigation in image-based telepresence systems," in *Proc. of the 11th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras (OMNIVIS 2011)*, vol. 11. IEEE, Nov. 2011, pp. 351–358.

[35] H. H. Baker and P. Constantin, "Panoramic stereoscopic camera," U.S. Patent 2012/0 105 574, 2010.

[36] V. Vanijja and S. Horiguchi, "Omni-Directional stereoscopic images from one omnidirectional camera," *The Journal of VLSI Signal Processing*, vol. 42, no. 1, pp. 91–101, Jan. 2006.

[37] W. A. Clay, "Methods of stereoscopic reproduction of images," U.S. Patent 3 225 651, Dec., 1965.

[38] Z. Zhu, "Omnidirectional stereo vision," in *Proceedings of the Workshop on Omnidirectional Vision, Budapest, Hungary*, 2001.

[39] L. E. Gurrieri, "The acquisition of omnidirectional stereoscopic images and videos of dynamic scenes: a review," website, 2013. [Online]. Available: http://luisgurrier.net/pub/background/

[40] D. W. Rees, "Panoramic television viewing system," U.S. Patent 3 505 465, Apr., 1970.

[41] S. Nayar, "Catadioptric omnidirectional camera," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1997, pp. 482–488.

[42] S. Baker and S. Nayar, "A theory of catadioptric image formation," in *Proc. of the IEEE 6th International Conference on Computer Vision*, vol. 1, Jan. 1998, pp. 35–42.

[43] V. S. Nalwa, "A true omnidirectional viewer," technical report, Bell Laboratories, Tech. Rep., 1996.

[44] ——, "Stereo panoramic viewing system," U.S. Patent 6 141 145, Oct., 2000.

[45] M. Hori, M. Kanbara, and N. Yokoya, "Arbitrary stereoscopic view generation using multiple omnidirectional image sequences," in *Proc. IEEE Int. Conf. Pattern Recognition*, 2010, pp. 286–289.

[46] W. Stürzl, H. Dahmen, and H. A. Mallot, "The quality of catadioptric imaging application to omnidirectional stereo," in *Computer Vision (ECCV)*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2004, vol. 3021, pp. 614–627.

[47] H. Igehy and L. Pereira, "Image replacement through texture synthesis," in *Proc. of International Conference on Image Processing*, vol. 3, Oct. 1997, pp. 186–189.

[48] J. Shimamura, N. Yokoya, H. Takemura, and K. Yamazawa, "Construction of an immersive mixed environment using an omnidirectional stereo image sensor," in *Proc. IEEE Workshop Omnidirectional Vision*, 2000, pp. 62–69.

[49] L. Spacek, "Coaxial omnidirectional stereopsis," in *Computer Vision - ECCV 2004*. Springer Berlin / Heidelberg, 2004, vol. 3024, pp. 354–365.

[50] Fraunhofer Heinrich Hertz Institute (HHI), "Multiview generation for 3D digital signage," Website, 2012. [Online]. Available: http://www.hhi.fraunhofer.de/fields-of-competence/image-processing/applications/multiview-generati

[51] O. Scheer, P. Kauff, C. Weissig, and J.-C. Rosenthal, "Geometrical design concept for panoramic 3D video acquisition," in *20th European Signal Processing Conf.*, Bucarest, Romania, Aug. 2012, pp. 2757–2761.

[52] C. Weissig, O. Scherr, P. Eisert, and P. Kauff, "Advances in multimedia modeling," in *The Ultimate Immersive Experience: Panoramic 3D Video Acquisition*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2012, vol. 7131, pp. 671–681.

[53] S. Peleg, Y. Pritch, and M. Ben-Ezra, "Cameras for stereo panoramic imaging," in *Proc. IEEE Computer Vision and Pattern Recognition*, vol. 1, 2000, pp. 208–214.

[54] S. Peleg, M. Ben-Ezra, and Y. Pritch, "Stereo panoramic camera arrangements for recording panoramic images useful in stereo panoramic image pair," U.S. Patent 6 795 109, Sep., 2004.

[55] J. M. Diamond, *Guns, Germs, and Steel: The Fates of Human Societies*. New York: W. W. Norton & Company, Apr. 1999.

[56] HumanEyes Technologies Ltd., "Lenticular 3D billboards," Website, 2012. [Online]. Available: http://www.humaneyes.com/3d-gallery-products/

[57] Y. Pritch, M. Ben-Ezra, and S. Peleg, *Foundations of Image Understanding*, L. S. Davis, Ed. Dordrecht, the Netherlands: Kluwer Academic, 2001.

[58] P. Bourke, "Omni-directional stereoscopic fisheye images for immersive hemispherical dome environments," *Int. Conference on Computer Games, Multimedia and Allied Technology*, pp. 136–143, 2009.

[59] M. S. Banks, "Achieving near-correct focus cues in a 3D display using multiple image planes," *SID Symposium Digest of Technical Papers*, vol. 37, no. 1, pp. 77–80, June 2006.

[60] M. Fiala and A. Basu, "Panoramic stereo reconstruction using non-SVP optics," *Computer Vision and Image Understanding*, vol. 98, no. 3, pp. 363–397, June 2005.

[61] L. Bagnato, P. Frossard, and P. Vandergheynst, "Optical flow and depth from motion for omnidirectional images using a TV-L1 variational framework on graphs," in *Proc. of 16th IEEE Int. Conf. on Image Processing (ICIP)*, 2009, pp. 1469–1472.

[62] L. Zhang, C. Vazquez, and S. Knorr, "3D-TV content creation: Automatic 2D-to-3D video conversion," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 372–383, June 2011.

[63] S. Peleg, M. Ben-Ezra, and Y. Pritch, "Omnistereo: Panoramic stereo imaging," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 3, pp. 279–290, 2001.

[64] T. Grover, "Multi-dimensional imaging," U.S. Patent 7 796 152B2, Sep., 2010.

[65] NASA Mars Team Online, "Mars Pathfinder Mission - Stereo panorama of the Sagan Memorial Station (QuickTime VR)," Website, 1997. [Online]. Available: http://science.ksc.NASA.gov/mars/vrml/qtvr_stereo.html

[66] NASA, "Mars Pathfinder Mission: Stereo Monster Panorama," Website, 1998. [Online]. Available: http://mars.jpl.NASA.gov/MPF/ops/stereopan.html

[67] Mars Pathfinder CD-ROM Directory, "Jet Propulsion Laboratory - NASA, Mars Pathfinder," Website, 1998. [Online]. Available: http://mars.jpl.NASA.gov/MPF/sitemap/anaglyph.html

[68] National Geographic Magazine, "Return to Mars @ nationalgeographic.com," Website, 1998. [Online]. Available: http://www.nationalgeographic.com/features/98/mars/

[69] O. Romain, T. Ea, C. Gastaud, and P. Garda, "An omnidirectional stereoscopic sensor: spherical color image acquisition," in *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 2, 2002, pp. II–209–II–212.

[70] R. A. Ainsworth, D. J. Sandin, J. P. Schulze, A. Prudhomme, T. A. DeFanti, and M. Srinivasan, "Acquisition of stereo panoramas for display in VR environments," in *SPIE*, vol. 7864, 2011, pp. 786 416–1–15.

[71] R. A. Ainsworth, "Environmental imaging: Photographic capture of visual environments for virtual reality systems," Website, 2010. [Online]. Available: http://www.qwerty.com/Environmental_Imaging/Index.html

[72] K. Yamada, "Virtual view generation of natural panorama scenes by setting representation," in *Proceedings of SPIE*, vol. 4660, San Jose, CA, USA, 2002, pp. 300–309.

[73] S. Chan and A. Clark, "Periscopic stereo for virtual world creation," in *Proc. 6th Int. Conf on Image Processing and Its Applications*, vol. 1, 1997, pp. 419–422.

[74] P. Peer and F. Solina, "Panoramic depth imaging: Single standard camera approach," *International Journal of Computer Vision*, vol. 47, no. 1, pp. 149–160, 2002.

[75] ——, *Real time panoramic depth imaging from multiperspective panoramas using standard cameras*, ser. Robot vision: new research, T. Matsuda, Ed. New York: Nova Science, 2009.

[76] J. Cho, J. H. Cha, Y. M. Tai, Y.-S. Moon, and S. Lee, "Stereo panoramic image stitching with a single camera," in *IEEE Int. Conf. on Consumer Electronics*, 2013, pp. 256–257.

[77] C. Richardt, Y. Pritch, H. Zimmer, and A. Sorkine-Hornung, "Megastereo: constructing high resolution stereo panoramas," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 1256–1263.

[78] ——, "Megastereo: constructing high resolution stereo panoramas," Website, 2013. [Online]. Available: http://www.disneyresearch.com/project/megastereo/

[79] Y. Pritch, M. Ben-Ezra, and S. Peleg, "Automatic disparity control in stereo panoramas (OmniStereo)," in *Proc. IEEE Workshop on Omnidirectional Vision*, 2000, pp. 54–61.

[80] C. Wang, C.-Y. Chang, and A. A. Sawchuk, "Object-based disparity adjusting tool for stereo panoramas," in *Stereoscopic Displays and Virtual Reality Systems XIV*, vol. 6490. SPIE, 2007, pp. 64 900D–64 900D–12.

[81] M. Ben-Ezra, Y. Pritch, and P. S., "Omnistereo: 3D stereo panorama," Website, 2001. [Online]. Available: http://www.vision.huji.ac.il/stereo/

[82] M. Ben-Ezra, "Omnidirectional stereo imaging," Website, 2001. [Online]. Available: http://www.ben-ezra.org/omnistereo/omni.html

[83] S. Peleg, M. Ben-Ezra, and Y. Pritch, "System and method for capturing and viewing stereoscopic panoramic images," U.S. Patent 7 477 284B2, Jan., 2009.

[84] ——, "System and method for facilitating the adjustment of disparity in a stereoscopic panoramic image pair," U.S. Patent 6 831 677, Dec., 2004.

[85] ——, "System and method for generating and displaying panoramic images and movies," U.S. Patent 6 665 003, Dec., 2003.

[86] R. G. Baker, "Immersive imaging system," U.S. Patent 7 224 382 B2, May, 2007.

[87] P. Bourke, "Capturing omni-directional stereoscopic spherical projections with a single camera," in *16th International Conference on Virtual Systems and Multimedia (VSMM)*, 2010, pp. 179–183.

[88] Panoscan, "Panoscan Mark III panoramic camera," Website, 2012. [Online]. Available: http://www.panoscan.com/MK3/index.html

[89] S. Wei, F. Huang, and R. Klette, "Determination of geometric parameters for stereoscopic panorama cameras," *Machine Graphics & Vision*, vol. 10, no. 3, pp. 399–427, 2001.

[90] S.-K. Wei, F. Huang, and R. Klette, "The design of a stereo panorama camera for scenes of dynamic range," in *Proc. 16th Int. Conf. on Pattern Recognition*, vol. 3, 2002, pp. 635–638.

[91] F. Huang and R. Klette, "Stereo panorama acquisition and automatic image disparity adjustment for stereoscopic visualization," *Multimedia Tools and Applications*, vol. 47, pp. 353–377, 2010, 10.1007/s11042-009-0328-2.

[92] F. Huang, A. Torii, and R. Klette, "Geometries of panoramic images and 3D vision," *Machine Graphics and Vision*, vol. 19, no. 4, pp. 463–477, 2010.

[93] W. Jiang and J. Lu, "Panoramic 3D reconstruction by fusing color intensity and laser range data," in *2006 IEEE International Conference on Robotics and Biomimetics*, Kunming, China, Dec. 2006, pp. 947–953.

[94] C. Barnes, "Omnistereo images from ground based LIDAR," in *ACM SIGGRAPH 2012 Posters*, ser. SIGGRAPH '12. Los Angeles, California: ACM, 2012.

[95] W. Jiang, S. Sugimoto, and M. Okutomi, "Panoramic 3D reconstruction using rotating camera with planar mirrors," *Proceedings of Omnivis '05*, vol. 1, pp. 123–130, 2005.

[96] P. Bourke, "Synthetic stereoscopic panoramic images," *Interactive Technologies and Sociotechnical Systems*, pp. 147–155, 2006.

[97] K. Sarachik, "Characterising an indoor environment with a mobile robot and uncalibrated stereo," in *IEEE Int. Conf. on Robotics and Automation*, May 1989, pp. 984–989.

[98] S. B. Kang and R. Szeliski, "3-D scene data recovery using omnidirectional multibaseline stereo," *International Journal of Computer Vision*, vol. 25, no. 2, pp. 167–183, 1997.

[99] S. Fleck, F. Busch, P. Biber, W. Straßer, and H. Andreasson, "Omnidirectional 3D modeling on a mobile robot using graph cuts," in *Proc. of the IEEE International Conf. on Robotics and Automation (ICRA 2005)*, 2005, pp. 1748–1754.

[100] K. Yamaguchi, H. Takemura, K. Yamazawa, and N. Yokoya, "Real-time generation and presentation of view-dependent binocular stereo images using a sequence of omnidirectional images," in *Proc. IEEE Int. Conf. Pattern Recognition*, vol. 4, 2000, pp. 589–593.

[101] M. Hori, M. Kanbara, and N. Yokoya, "Novel stereoscopic view generation by image-based rendering coordinated with depth information," in *15th Scandinavian Conference on Image Analysis*, 2007, pp. 193–202.

[102] C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon, and J. C. Hart, "The CAVE: audio visual experience automatic virtual environment," *Communications of the ACM*, vol. 35, pp. 64–72, June 1992.

[103] D. Shimada, H. Tanahashi, K. Kato, and K. Yamamoto, "Extract and display moving object in all direction by using stereo omnidirectional system (SOS)," in *Proc. the 3rd Int. Conf. on 3-D Digital Imaging and Modeling*, 2001, pp. 42 –47.

[104] H. Tanahashi, K. Yamamoto, C. Wang, and Y. Niwa, "Development of a stereo omnidirectional imaging system (SOS)," in *Proc. of the 26th Conf. of the IEEE Industrial Electronics Society (IECON)*, vol. 1, 2000, pp. 289–294.

[105] H. Tanhashi, D. Shimada, K. Yamamoto, and Y. Niwa, "Acquisition of three-dimensional information in a real environment by using the stereo omni-directional system (SOS)," in *Proc. of 3-D Digital Imaging and Modeling*, 2001, pp. 365–371.

[106] P. Firoozfam, S. Negahdaripour, and C. Barufaldi, "A conical panoramic stereo imaging system for 3-D scene reconstruction," in *Proc. of OCEANS 2003*, vol. 4, 2003, pp. 2303–2308.

[107] S. Negahdaripour, H. Zhang, P. Firoozfam, and J. Oles, "Utilizing panoramic views for visually guided tasks in underwater robotics applications," in *MTS/IEEE Conference and Exhibition Oceans*.

[108] L. P. Steuart III, "Digital 3D/360 degree camera system," U.S. Patent US20 120 327 185 A1, 2012.

[109] J. Zhu, G. Humphreys, D. Koller, S. Steuart, and R. Wang, "Fast omnidirectional 3D scene acquisition with an array of stereo cameras," in *Sixth International Conference on 3-D Digital Imaging and Modeling, 2007. 3DIM '07*, Aug. 2007, pp. 217–224.

[110] P. Peer and F. Solina, "Towards a real time panoramic depth sensor," in *Computer Analysis of Images and Patterns*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2003, vol. 2756, pp. 107–115.

[111] H. Afshari, L. Jacques, L. Bagnato, A. Schmid, P. Vandergheynst, and Y. Leblebici, "Hardware implementation of an omnidirectional camerawith real-time 3D imaging capability," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, May 2011, pp. 1–4.

[112] EPFL News, "Revolutionary 360 degrees 3D camera," website, 2007. [Online]. Available: http://youtu.be/KFsERnHu0Cc

[113] H. Afshari, V. Popovic, T. Tasci, A. Schmid, and Y. Leblebici, "A spherical multi-camera system with real-time omnidirectional video acquisition capability," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 4, pp. 1110–1118, Nov. 2012.

[114] D. Pierce, S. M. Herrnstadt, T. Grover, and F. Good, "Stereoscopic panoramic image capture device," U.S. Patent 6 947 059, Sep., 2005.

[115] W. Feng, B. Zhang, J. Röning, X. Zong, and T. Yi, "Panoramic stereo sphere vision," in *Proc. SPIE 8662, Intelligent Robots and Computer Vision XXX: Algorithms and Techniques*, Feb. 2013, p. 866206.

[116] Université de Montréal, "Lab Vision3D: omnistereo," Website, 2013. [Online]. Available: http://vision3d.iro.umontreal.ca/en/projects/omnistereo/

[117] V. Couture, M. S. Langer, and S. Roy, "Panoramic stereo video textures," Nov. 2011, pp. 1251–1258.

[118] V. C. Couture, M. S. Langer, and S. Roy, "Omnistereo video textures without ghosting," in *IEEE Int. Conference on 3DTV*, 2013, pp. 64–70.

[119] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. of the 23rd Conf. on Computer Graphics and Interactive Techniques, SIGGRAPH'96*, 1996, pp. 31–42.

[120] A. S. Ogale and Y. Aloimonos, "A roadmap to the integration of early visual modules," in *Proc. IEEE Int. Conf. Computer Vision: Special Issue on Early Cognitive Vision*, vol. 72, no. 1, 2007, pp. 9–25.

[121] J. M. Harris, "Monocular zones in stereoscopic scenes: A useful source of information for human binocular vision?" in *SPIE*, vol. 7524, 2010, pp. 151–162.

[122] D. B. Diner, *Human engineering in stereoscopic viewing devices*, ser. Advances in computer vision and machine intelligence. New York: Plenum Press, 1993.

[123] C. Ware, "Dynamic stereo displays." ACM Press, 1995, pp. 310–316.

[124] A. Çöltekin, "Foveation for 3D visualization and stereo imaging," PhD Thesis, Aalto University. [Online]. Available: http://lib.tkk.fi/Diss/2006/isbn9512280175/

[125] J. Dakin and R. G. W. Brown, Eds., *Handbook of optoelectronics*. New York: Taylor & Francis, 2006.

[126] C. Chang and S. Chatterjee, "Quantization error analysis in stereo vision," in *IEEE Asilomar Conference on Signals, Systems and Computers*, vol. 2, 1992, pp. 1037–1041.

[127] S. Steinman, B. Steinman, and R. Garzia, *Foundations of Binocular Vision: A Clinical Perspective*, 1st ed. New York, NY: McGraw-Hill Medical, June 2000.

[128] M. Kytö, M. Nuutinen, and P. Oittinen, "Method for measuring stereo camera depth accuracy based on stereoscopic vision," in *SPIE*, vol. 7864, 2011, pp. I1–I9.

[129] L. E. Gurrieri, "Alignment of panoramic images for the creation of stereoscopic image-based virtual environments," website, 2011. [Online]. Available: http://luisgurrieri.net/pub/cpia/

[130] R. Szeliski, "Image mosaicing for tele-reality applications," in *Proc. of the Second IEEE Workshop on Applications of Computer Vision*, 1994, pp. 44–53.

[131] ——, "Image alignment and stitching: A tutorial," *Foundations and Trends in Computer Graphics and Vision*, vol. 2, no. 1, pp. 1–10, 2006.

[132] V. Kwatra, A. Schdl, I. Essa, G. Turk, and A. Bobick, "Graphcut textures: image and video synthesis using graph cuts," in *Proc. of ACM SIGGRAPH 2003*, vol. 22, July 2003, pp. 277–286.

[133] S. J. Ha, H. Koo, S. H. Lee, N. I. Cho, and S. K. Kim, "Panorama mosaic optimization for mobile camera systems," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 4, pp. 1217–1225, Nov. 2007.

[134] R. Szeliski, "Video mosaics for virtual environments," *IEEE Computer Graphics and Applications*, vol. 16, no. 2, pp. 22–30, 1996.

[135] M. T. Uyttendaele and R. S. Szeliski, "Panoramic video," U.S. Patent 6 788 333, Sep., 2004.

[136] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*. Englewood Cliffs, New Jersey: Prentice Hall, Mar. 1998.

[137] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.

[138] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[139] S. Ikeda, T. Sato, and N. Yokoya, "A calibration method for an omnidirectional multi-camera system," in *Stereoscopic Displays and Virtual Reality Systems X, 21-24 Jan. 2003*, ser. Proc. SPIE - Int. Soc. Opt. Eng. (USA), vol. 5006. SPIE-Int. Soc. Opt. Eng, 2003, pp. 499–507.

[140] J. Heikkila and O. Silven, "A four-step camera calibration procedure with implicit image correction," in *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 1997, pp. 1106–1112.

[141] T. Yan, Z. Huang, R. W. Lau, and Y. Xu, "Seamless stitching of stereo images for generating infinite panoramas," in *Proc. of the* 19^{th} *ACM Symposium on Virtual Reality Software and Technology*, 2013, pp. 251–258.

[142] L. E. Gurrieri, "Stereoscopic cameras for the real-time acquisition of panoramic 3D images and videos," website, 2013. [Online]. Available: http://luisgurrier.net/pub/spcamera/

[143] Adobe, "Photoshop Lightroom," Website, 2013. [Online]. Available: http://www.adobe.com/ca/products/photoshop-lightroom.html

[144] GIMP, "GIMP: The GNU image manipulation program," Website, 2013. [Online]. Available: http://www.gimp.org/

[145] Mathworks, "Matlab," Website, 2013. [Online]. Available: http://www.mathworks.com/

[146] E. Dubois, "A projection method to generate anaglyph stereo images," in *Proc. IEEE Int. Conf. Acoustics Speech Signal Processing*, vol. 3, May 2001, pp. 1661–1664.

[147] M. Sutu and D. Sykes, "StereoPhotoMaker: stereo photography authoring tool," Website, 2013. [Online]. Available: http://stereo.jpn.org/eng/stphmkr/

[148] C. Hughes, M. Glavin, E. Jones, and P. Denny, "Review of geometric distortion compensation in fish-eye cameras," in *IET Irish Signals and Systems Conference*, 2008, pp. 162–167.

[149] J.-Y. Bouguet, "Camera Calibration Toolbox for Matlab," Website, 2010. [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc/

[150] Xrite, "Color Checker Passport," Website, 2013. [Online]. Available: http://xritephoto.com/ph_product_overview.aspx?id=1257&catid=28&action=overview

[151] Mathworks, "Stereo Image Rectification," Website, 2012. [Online]. Available: http://www.mathworks.com/help/vision/examples/stereo-image-rectification.html

[152] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: speeded up robust features," in *Computer Vision ECCV 2006*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2006, vol. 3951, pp. 404–417.

[153] Mathworks: Computer Vision Toolbox, "Matlab's implementation of SURF:detectSURFFeatures," Website, 2012. [Online]. Available: http://www.mathworks.com/help/vision/ref/detectsurffeatures.html

[154] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. of the ACM*, vol. 24, no. 6, pp. 381–395, June 1981.

[155] Mathworks: Computer Vision Toolbox, "Estimate fundamental matrix from corresponding points in stereo images: estimateFundamentalMatrix," Website, 2012. [Online]. Available: http://www.mathworks.com/help/vision/ref/estimatefundamentalmatrix.html

[156] Y. Xiong and K. Pulli, "Color correction for mobile panorama imaging," in *Proc. of the 1st Int. Conf. on Internet Multimedia Computing and Service*, 2009, pp. 219–226.

[157] H.-Y. Shum and R. Szeliski, "Construction of panoramic image mosaics with global and local alignment," in *Panoramic Vision*, R. Benosman and S. B. Kang, Eds. Springer Verlag, 2001, ch. 13, pp. 227–268.

[158] R. Szeliski, *Computer vision algorithms and applications*. London; New York: Springer, 2011.

[159] R. Daz Hernndez, "Rendering stereoscopic 3D images in cylindrical spaces," Ph.D. dissertation, Universidad Politecnica de Valencia, Nov. 2011.

[160] P. Burt and E. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 532–540, 1983.

[161] L. E. Gurrieri, "The omnidirectional acquisition of stereoscopic images of dynamic scenes," website, 2013. [Online]. Available: http://luisgurrieri.net/omnistereo/

[162] Kolor, "Autopano Giga: panoramic tours authoring tool," website, 2010. [Online]. Available: http://www.kolor.com/panotour-software-virtual-tours-360-home.html

[163] L. Zhang and W. Tam, "Stereoscopic image generation based on depth images for 3D TV," *IEEE Transactions on Broadcasting*, vol. 51, no. 2, pp. 191–199, June 2005.

[164] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.

[165] ——, "A matlab toolbox that solves several image processing tasks using the primal-dual algorithm," Website, May 2009, Season 10, Episode 13. [Online]. Available: http://gpu4vision.icg.tugraz.at/index.php?content=downloads.php

[166] L. E. Gurrieri and E. Dubois, "Depth consistency and vertical disparities in stereoscopic panoramas," *Journal of Electronic Imaging*, vol. 23, no. 1, p. 011004, Jan. 2014.

[167] L. E. Gurrieri, "Depth consistency and vertical disparities in stereoscopic panoramas," website, 2014. [Online]. Available: http://luisgurrieri.net/pub/drdv/

[168] ——, "Efficient panoramic sampling of real world environments for image-based stereoscopic telepresence," website, 2012. [Online]. Available: http://luisgurrieri.net/pub/tspi/

[169] Point Grey Research Inc., "Ladybug2," Website, 2011. [Online]. Available: http://www.ptgrey.com/products/ladybug2/

[170] M. Uyttendaele, A. Criminisi, S. B. Kang, S. Winder, R. Szeliski, and R. Hartley, "Image-based interactive exploration of real-world environments," *IEEE Computer Graphics and Applications*, pp. 52–63, 2004.

[171] S. B. Kang and R. Weiss, "Characterization of errors in composing cylindrical panoramic images," *Panoramic vision: sensors, theory and applications, Ch. 12*, pp. 205–226, 2001.

[172] M. Fiala and G. Roth, "Automatic alignment and graph map building of panoramas," in *IEEE Int. Workshop on Haptic Audio Visual Environments and their Applications*, 2005, pp. 103–108.

[173] F. Kangni and R. Laganiere, "Epipolar geometry for the rectification of cubic panoramas," in *Proc. of the 3rd Canadian Conf.on Computer and Robot Vision*, 2006, pp. 70–77.

[174] S. Salehi and E. Dubois, "Alignment of cubic-panorama image datasets using epipolar geometry," in *Proc. IEEE Int. Conf. Acoustics Speech Signal Processing*, Prague, Czech Republic, May 2011, pp. 1545–1548.

[175] L. Lipton, *Foundations of the Stereoscopic Cinema: A Study in Depth*. New York: Van Nostrand Reinhold, 1982.

[176] M. Siegel, Y. Tobinaga, T. Akiya, J. O. Merritt, M. T. Bolas, and S. S. Fisher, "Kinder, gentler stereo," in *Proc. of SPIE*, vol. 3639. San Jose, CA, USA: SPIE, May 1999, pp. 18–27.

[177] M. Siegel and S. Nagata, "Just enough reality: comfortable 3-D viewing via microstereopsis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 3, pp. 387 –396, Apr. 2000.

[178] M. Lang, A. Hornung, O. Wang, A. Poulakos, A. Smolic, and M. Gross, "Nonlinear disparity mapping for stereoscopic 3D," *ACM Transactions on Graphics*, vol. 29, no. 4, pp. 75–85, 2010.

[179] A. J. Woods, "3d or 3-d: a study of terminology, usage and style." *European Science Editing*, vol. 39, no. 3, 2013.

[180] Merriam-Webster, *Merriam-Webster's Collegiate Dictionary, 11^{th} Edition*, 11th ed. Merriam-Webster, Inc., Apr. 2008.

Index