PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

SCHOOL OF ENGINEERING

COMPARISON OF PANORAMIC STEREOSCOPIC SENSORS BASED ON HYPERBOLOIDAL MIRRORS

FELIPE ANDRÉS BRAVO VALENZUELA

Thesis submitted to the Office of Research and Graduate Studies
in partial fulfillment of the requirements for the degree of
Master of Science in Engineering

Advisor:
MIGUEL TORRES T.

Santiago de Chile, April 2010

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

SCHOOL OF ENGINEERING

COMPARISON OF PANORAMIC STEREOSCOPIC SENSORS BASED ON HYPERBOLOIDAL MIRRORS

FELIPE ANDRÉS BRAVO VALENZUELA

Members of the Committee:

MIGUEL TORRES T.

DOMINGO MERY Q.

PABLO ZEGERS F.

LUCIANO CHIANG S.

Thesis submitted to the Office of Research and Graduate Studies

in partial fulfillment of the requirements for the degree of

Master of Science in Engineering

Santiago de Chile, April 2010

Gratefully to my family

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Panoramic distance maps can provide valuable information for mobile robot navigation. An approach to obtain such distance maps is to implement a stereoscopic system using two hyperboloidal catadioptric sensors. This document analyzes the effects over distance estimation of two typical arrangements of the mirrors. In one of the arrangements the mirrors are placed along a vertical axis, while in the other configuration the sensors are located on plane. Both configurations may yield similar distance estimation errors for an adequately chosen baseline and given measurement range of interest. However, depending on the application, some aspects can make one arrangement preferable over the other. In the vertical configuration, the epipolar lines correspond to radial lines in the omnidirectional image, and hence, the stereoscopic correspondence problem can be solved in a standard and simple way, but a drawback of this arrangement is the space required, which in the case of mobile robot and vehicular application translates into a reduced vertical clearance and poorer aerodynamics. The horizontal arrangement, on the other hand, requires less space if mounted on a mobile robot, a vehicle's roof or its lateral mirrors, but involves complex parabolic, elliptic or hyperbolic epipolar curves that make the solution of the stereo correspondence problem computationally more expensive. Despite the evident implementation advantages and drawbacks of each approach, not enough quantitative information exits in the literature about which of the two configurations provides in general the most accurate 3D reconstruction of the environment.

Keywords: panoramic stereoscopy, catadioptric system, omnidirectional hyperbolic vision sensor.

RESUMEN

Los mapas de distancia panorámicos proveen información valiosa para navegación de robots móviles. Una opción para obtener tal mapa de distancia es implementar un sistema estereoscópico usando dos espejos hiperbólicos. El presente trabajo analiza el efecto sobre la estimación de distancia para dos configuraciones típicas de los espejos. En una de las configuraciones los espejos están ubicados a lo largo del eje vertical, mientras que en la otra, los sensores están dispuestos de forma horizontal. Ambas configuraciones deberían proporcionar errores similares de estimación en la distancia para una adecuada separación de los sensores y una apropiada medida del rango de interés. Aunque, dependiendo de la aplicación, algunos aspectos pueden hacer una configuración más adecuada que la otra. En la configuración vertical, las líneas epipolares corresponden a líneas radiales en la imagen omnidireccional, por lo tanto, el problema de la correspondencia estereoscópica puede ser resuelto de manera simple y tradicional, pero una desventaja de esta configuración es el espacio que requiere, en el caso de robots móviles y aplicaciones en vehículos se traduce en una menor resolución vertical y una aerodinámica más pobre. El arreglo horizontal, por otra parte, requiere menos espacio si se monta en un robot móvil, en el techo de un vehículo o sus espejos laterales, pero implica complejas curvas epipolares parabólicas, elípticas o hiperbólicas que hacen que la solución del problema de la búsqueda de correspondencias en el sistema estéreo tenga un mayor costo computacional. A pesar de las evidentes ventajas y desventajas en la implementación de cada configuración, no suficiente información cuantitativa en la literatura sobre cual de estas dos configuraciones provee, en general, la más precisa reconstrucción 3D.

Palabras Claves: estereoscopía panorámica, sistema catadióptrico, sistema de visión omnidireccional hiperbólico.

1. INTRODUCTION

Visual panoramic information is extremely useful in a wide range of applications from mobile robot navigation to surveillance and teleconferencing (Benosman & Kang, 2001). In particular, the ability to measure distances $360°$ around a moving platform is essential not only for autonomous navigation, but also for driver assistance systems that must anticipate possible collisions to other vehicles on the road. Dense range measurements can be obtained using sensors based on time-of-flight measurements, such as laser scanners, or employing several cameras to solve distance by triangulation (Bertozzi et al., 2002; Bertozzi & Broggi, 1998). Although laser scanners can be extremely accurate and reliable, few scanners exist which can obtain 3D dense panoramic range measurements (e.g. Velodyne and Ibeo) and they can cost between $ 60,000-100,000 USD, thus not being economically viable for passenger car manufacturers. On the other hand, digital cameras can cost less than $ 1,000 USD and therefore can provide an attractive alternative to perform stereoscopic analysis and to obtain distance information. Stereoscopic binocular systems traditionally employ standard perspective cameras mounted so that their optical axes are parallel. With this arrangement the field-of-view is less than the field-of-view of each separate camera (typically around $50°$) and decreases as the baseline increases. Therefore, for a $360°$ coverage it would be necessary to employ several binocular arrangements or a rotating system, unless conic, hyperbolic, parabolic, or spherical mirrors are employed (see fig. 1.1).

The purpose of this thesis is precisely to study the construction of depth maps using two hyperboloidal catadioptric sensors and assess the quality of distance estimation under two typical arrangements of the mirrors. In one of the arrangements the mirrors are placed along a vertical axis, while in the other configuration the sensors are located on a plane. Both configurations may yield similar distance estimation errors for an adequately chosen baseline and given measurement range of interest. However, depending on the application, some aspects can make one arrangement preferable over the other. In the vertical configuration, the epipolar lines correspond to radial lines in the omnidirectional image, and

hence, the stereoscopic correspondence problem can be solved in a standard and simple way, but a drawback of this arrangement is the space required, which in the case of mobile robot and vehicular application translates into a reduced vertical clearance and poorer aerodynamics. The horizontal arrangement, on the other hand, requires less space if mounted on a mobile robot, a vehicle's roof or its lateral mirrors, but involves complex parabolic, elliptic or hyperbolic epipolar curves that make the solution of the stereo correspondence problem computationally more expensive. Despite the evident implementation advantages and drawbacks of each approach, not enough quantitative information exits in the literature about which of the two configurations provides in general the most accurate 3D reconstruction of the environment.

1.1. Problem Definition

Consider a camera with a mirror as shown in figs. 1.1 or 2.1, if two of these cameras are arranged vertically as in fig. 2.2 or horizontally as in fig. 2.3, the first problem is to find the equations that yield the distance d_P from the catadioptric system to a given point $\mathbf{P}$ in space. Once the equations are derived, a practical problem arises, which is to determine where does point $\mathbf{P}$ projects on each image. This problem is the so-called matching problem or correspondence search problem. For an effective solution of the matching problem it is essential to derive algebraically or at least numerically compute the geometry of the epipolar curves determining which part of the space must be searched for correspondences. Summarizing, the problem is threefold:

- finding the projective geometry of the stereoscopic omnidirectional hyperbolic sensors,
- finding the epipolar curves and solving the matching problem, and
- assessing the accuracy and expected error levels occurring under each of the two main possible configurations, the horizontal and the vertical arrangements.

2

1.2. Existing Approaches Using Catadioptric Systems

In the area of autonomous ground robots and vehicles, a wide variety of sensors is often employed, among which vision sensors based on perspective cameras is arguably the most common (Desouza & Kak, 2002). Thus, developing vision sensors with a panoramic field of view has received much interest, particularly since the pioneering works (Nayar, 1988; Yagi & Yachida, 1991; Yamazawa, Yagi, & Yachida, 1993; Yagi, Kawato, & Tsuji, 1994), for an in-depth review see (Benosman & Kang, 2001). Some omnidirectional sensors achieve panoramic views employing an array of detectors on a polyhedron or sphere, others involve rotating sensors, while a third group rely on a combination of mirrors with conventional imaging systems often referred to as *catadioptric systems* (Baker & Nayar, 1999; Benosman & Kang, 2001). The latter has received special interest during recent years as they do not involve additional data bandwidth requirements nor moving parts. Catadioptric systems may be implemented using planar, conical, spherical, ellipsoidal, hyperboloidal or paraboloidal mirrors (see fig. 1.1 for an example). However, only the hyperboloidal and parabolic mirrors produce a panoramic view with a single effective viewpoint, which is a requirement for the generation of true perspective panoramic images (Baker & Nayar, 1999), as well as distance maps when arrangements of two or more catadioptric systems are employed. For this reason, and the lack of results concerning optimal arrangement design, in this document we consider a stereoscopic arrangement of two omnidirectional cameras based on hyperboloidal mirrors. In one of the arrangements the catadioptric cameras are aligned along a vertical axis, while in the other setup the cameras are located on a horizontal plane.

From a computational point of view, the advantage of the vertical arrangement is that the epipolar curves are simple radial lines in the omnidirectional image (Svoboda & Pajdla, 2002). This facilitates the search for correspondences and the solution of the stereoscopic triangulation problem for distance estimation. In the case of the horizontal arrangement, the epipolar curves can be an arc segment of any conic (Svoboda & Pajdla, 2002), depending on the position of the observed point. This fact makes the solution of the stereoscopic

matching problem computationally more difficult and prone to errors whenever the epipolar curve is not calculated accurately. However, the horizontal configuration is often more convenient for mounting on a mobile robot or vehicle (Gandhi & Trivedi, 2005, 2006). Despite the advantages and drawbacks of each arrangement, it is not clear which configuration allows a more accurate distance estimation, which configuration is less sensitive to matching errors and which are the computational costs associated in each case. This investigation provides answers to these fundamental questions, and thus can be valuable to researchers and engineers in the fields of robotics and computer vision.

1.3. Summary of Contributions

The contributions of this work are summarized in:

(i) A detailed derivation of the optical geometry of the omnidirectional hyperboloidal stereoscopic system.

(ii) An analysis of the distance estimation errors under each camera-mirror arrangement.

(iii) The implementation of an algorithm to numerically compute the epipolar curve by back-projection of reference points onto the optical plane of the target system.

(iv) The development of a procedure to select the most adequate arrangement for a given operating distance and level of acceptable error in the distance estimation.

(v) The analytic quantification of the computational cost associated to the solution of the matching problem under both camera-mirror arrangements.

(vi) The construction of a camera rig which can be mounted on a vehicle and allows performing experiments with different camera positions.

(vii) Implementation in C/C++ using OpenCV and Matlab of dewarping and stereoscopic matching algorithms for omnidirectional hyperbolic images.
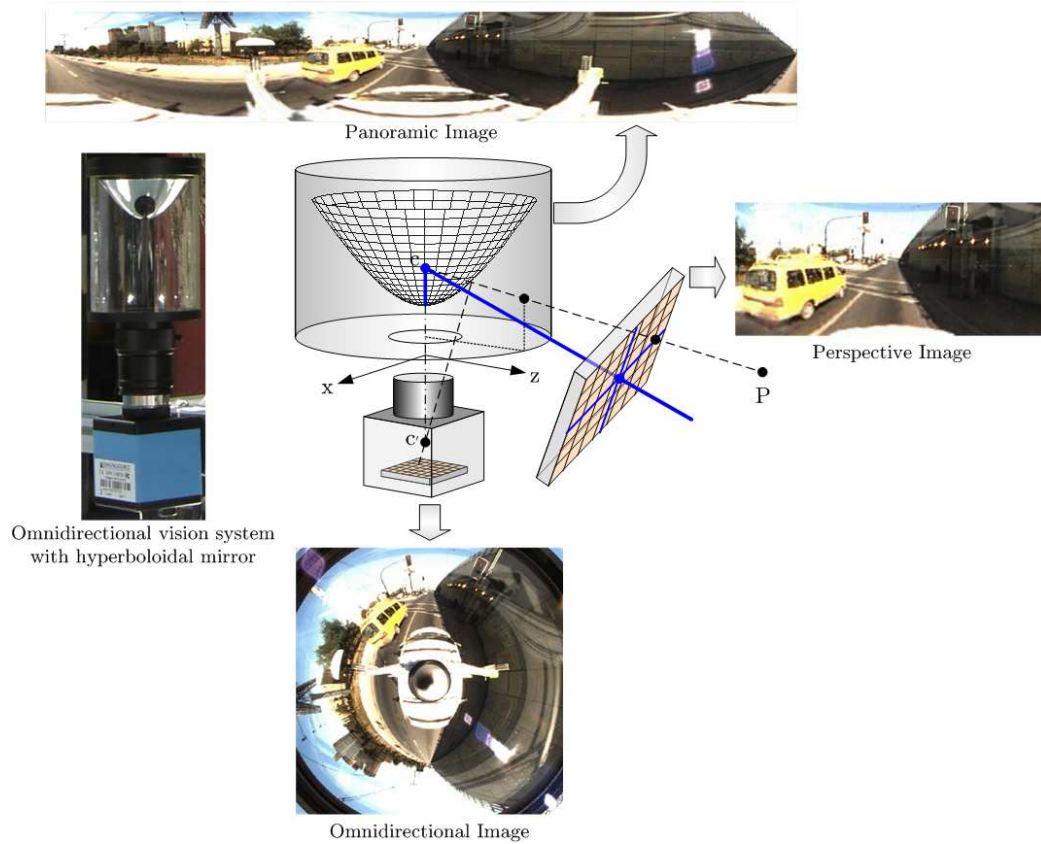
FIGURE 1.1. Omnidirectional hyperboloidal system.

1.4. Thesis Outline

This thesis is organized as follows. Chapter 2 presents the basic ray optics geometry required to formulate the projection equations from points in 3D space to points on the optical plane. The triangulation equations required to compute the distance to a point in the world from a given pair of corresponding points on each omnidirectional image are developed in section 2.3. The distance estimation errors under each configuration and arrangement selection are discussed in chapter 3. The feasibility of each arrangement is validated by computing distances for a synthetic scenario; and the experiments and results obtained with synthetic images and in real urban roads are discussed in chapter 4. Finally, chapter 5 presents the conclusions of this work and discusses some issues concerning ongoing research.

2. BASIC ASSUMPTIONS, FACTS AND PRELIMINARY RESULTS

2.1. Basic Assumptions

The basic assumptions in the analysis and design of stereoscopic omnidirectional system using hyperboloidal mirrors are:

(i) The intrinsic and extrinsic parameters of the system geometry are known.

(ii) The real system is calibrated and the calibration errors are negligible.

(iii) There is a working distance range, i.e. the system will only measure the distance to objects withing the operating distance interval.

(iv) When the system is mounted on a vehicle or moving platform, the ground can be assumed to be a plane and the cameras distance from the plane is fixed, thus the ground plane provides an upper bound for the operating distance interval.

(v) The matching problem can be solved for most points in the scene, i.e. illumination and texture levels are adequate, there are few repetitive patterns and few occlusions.

2.2. Hyperboloidal Catadioptric Sensor Geometry

For simplicity of the derivation, consider a cross section of the hyperbolic mirror, corresponding to a hyperbola that may be written as:

$$\frac{y^2}{b^2} - \frac{x^2}{a^2} = 1$$

with focal points $\mathbf{c} = (0, c)$ and $\mathbf{c}' = (0, -c)$, $c^2 = a^2 + b^2$. For the reader's convenience, a summary of the notation employed in the derivation of the catadioptric system equations is included in table 2.1.

Let α be the angle of the incoming ray measured with respect to an horizontal axis (parallel to $\mathbf{x}$ passing through $\mathbf{c}$, as shown in fig. 2.1). It can be shown that the angle between the tangent to the hyperbola at a point $\mathbf{p} = (x, y)$ and line segment $\overline{\mathbf{cp}}$ is equal to the angle between the tangent at $\mathbf{p}$ and the line segment $\overline{\mathbf{c'p}}$. Therefore, the incoming

SYMBOL	DESCRIPTION
P	point in the world
a, b, c	hyperbolic mirror parameters
c	upper focal point with coordinates $(0, c)$
c′	lower focal point with coordinates $(0, -c)$
f	focal distance between the focal point **c**′ and the CCD
p	point of intersection between the hyperbolic mirror and the incident ray
ϕ	angle of the incident and reflected rays with respect to the normal at **p**
α_i	angle between the line segment $\overline{\mathbf{c}'\mathbf{p}}$ and the horizontal plane through **c**′ containing the **x** axis
β_i	angle between the incoming ray and the horizontal plane through **c** containing the **x** axis
(u_1, v_1)	coordinates in the first (reference) image
(u_2, v_2)	coordinates in the second image
ℓ_i	distance between **P** and **c** of system i
(r_i, θ_i)	polar coordinates of a point in the world with respect to the upper focal point of the mirror system i
(ρ_i, θ_i)	polar coordinates on the CCD for point (r_i, θ_i)
d_i	distance between point **P** and upper focal point of system i
d_P	radial distance between the arrangement's vertical axis and a point in the world
d_c	distance between cameras

TABLE 2.1. Notation for the catadioptric system equations.

ray passing through **p** and **c** is reflected as a ray from **p** to **c**′. This is consistent with the physical law stating that the angle of reflection with respect to the surface's normal is equal to the angle of incidence. Let ϕ denote the angle of the incident and reflected rays with respect to the normal at **p** (see fig. 2.1), then $\gamma \overset{def}{=} \pi/2 - \phi$ is the angle between $\overline{\mathbf{c}'\mathbf{p}}$ and the tangent at **p**, which is, by the property just mentioned, equal to the angle between $\overline{\mathbf{c}'\mathbf{p}}$ and the tangent at **p**. Therefore, if β defines the angle of the reflected ray and an horizontal

axis through $\mathbf{c}'$, then:

$$2\gamma \;=\; \alpha + \beta \tag{2.1}$$

$$2(\pi/2 - \phi) \;=\; \alpha + \beta \tag{2.2}$$

Defining the angle $\delta \overset{def}{=} \pi/2 - \beta$, corresponding to the angle of the incoming ray projected on the optical plane with respect to the camera axis, then

$$\phi \;=\; 180° - (\alpha + \beta) \tag{2.3}$$

$$\;=\; 180° - (\alpha + \pi/2 - \delta) \;=\; \pi/2 + \delta - \alpha \tag{2.4}$$

It will be shown next that α can be computed as a function of δ, and therefore, it is possible to find an expression of ϕ in terms of δ. This is useful for finding a new representation for the omnidirectional image as a type of standard rectangular panoramic image.

Since δ is the orthogonal complement of β, finding α in terms δ, is equivalent to finding α in terms of β. The latter allows for a simpler derivation of the relevant equations. We now proceed as follows. First, define the length of the line segments from the foci $\mathbf{c}$, $\mathbf{c}'$ to point $\mathbf{p} = (x, y)$ on the hyperbola respectively as $\ell_1 = \|\overline{\mathbf{cp}}\|$ a $\ell_2 = \|\overline{\mathbf{c'p}}\|$, then from the geometrical properties of the hyperbola:

$$\ell_2 - \ell_1 \;=\; 2b \tag{2.5}$$

$$\ell_1 \sin(\alpha) + \ell_2 \sin(\beta) \;=\; 2c \tag{2.6}$$

By the Law of Sines:

$$\frac{\ell_1}{\cos(\beta)} \;=\; \frac{\ell_2}{\cos(\alpha)}, \tag{2.7}$$

while by the Law of Cosines:

$$\ell_1^2 \;=\; 4c^2 + \ell_2^2 - 4c\ell_2 \sin(\beta). \tag{2.8}$$

8

From (2.6) and (2.7), respectively:

$$\sin(\alpha) = \frac{2c - \ell_2 \sin(\beta)}{\ell_1}, \tag{2.9}$$

$$\cos(\alpha) = \frac{\ell_2}{\ell_1} \cos(\beta). \tag{2.10}$$

Dividing (2.9) by (2.10) yields

$$\tan(\alpha) = \frac{2c - \ell_2 \sin(\beta)}{\ell_2 \cos(\beta)} \tag{2.11}$$

Replacing ℓ_1 from (2.5) into (2.8):

$$(\ell_2 - 2b)^2 = 4c^2 + \ell_2^2 - 4c\ell_2 \sin(\beta), \tag{2.12}$$

which after rearranging yields

$$\ell_2 = \frac{c^2 - b^2}{c \sin(\beta) - b} \tag{2.13}$$

The last equation replaced into (2.11) yields after simplification the following expression for α in terms of β:

$$\alpha = \arctan\left(\frac{(c^2 + b^2)\sin(\beta) - 2bc}{(c^2 - b^2)\cos(\beta)}\right) \tag{2.14}$$

In a similar way, but employing the Law of Cosines equation equivalent to (2.8) for the angle at corner c of triangle $\triangle \mathbf{cpc'}$, the β angle in terms of α is found to be:

$$\beta = \arctan\left(\frac{(c^2 + b^2)\sin(\alpha) + 2bc}{(c^2 - b^2)\cos(\alpha)}\right) \tag{2.15}$$

If (u, v) is a point in a coordinate system whose origin is located at the center of the optical plane and f is the focal distance of the camera lens, then equation (2.14) together with

$$\beta = \frac{\pi}{2} - \arctan\left(\frac{\sqrt{u^2 + v^2}}{f}\right), \tag{2.16}$$

9
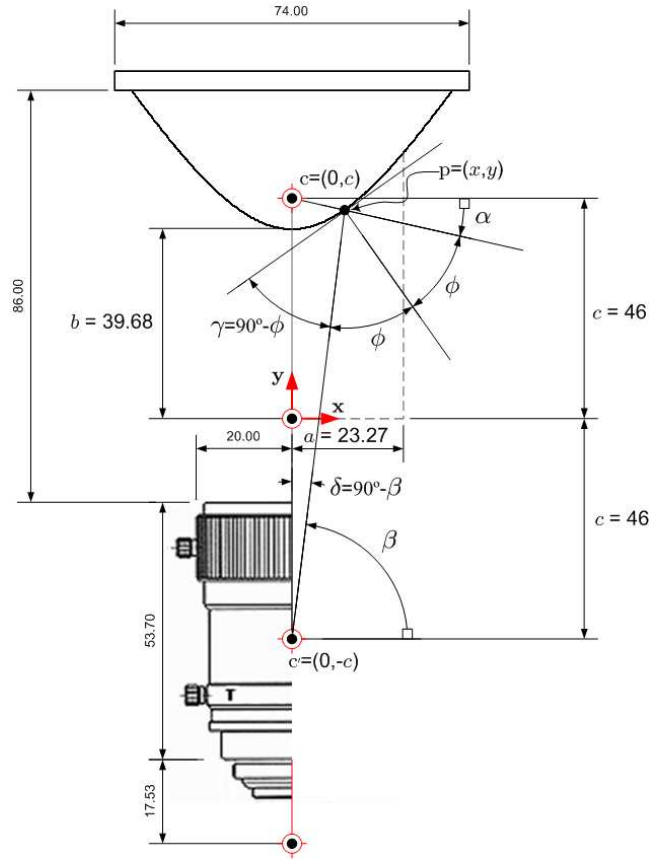
FIGURE 2.1. Hyperbolic mirror.

complete the projection equations relating the angle α of the incident ray with a given optical plane pixel (u, v).

2.3. Distance Estimation with the Panoramic Stereoscopic Arrangements

Since the epipolar lines in the vertical arrangement correspond to radial line segments in the omnidirectional image, while in the horizontal configuration the epipolar lines correspond to the projection of conic sections on the hyperbola onto the optical plane (Svoboda & Pajdla, 2002), we consider first, the simpler vertical configuration and then, the horizontal one.

2.3.1. Vertical Arrangement

Consider the vertical arrangement of the mirrors shown in fig. 2.2. The cameras are at a distance d_c, which is equivalent to the so-called *baseline* in a standard parallel perspective camera configuration. Denote by d_1 and d_2 the distance from the foci of the upper and lower mirrors, respectively, to a point $\mathbf{P}$ in the world, and let α_1, α_2 be the angles of the corresponding incident rays. The 3D coordinates of point $\mathbf{P}$ are found as follows.

From the geometry of the arrangement,

$$d_1 \sin(\alpha_1) = d_2 \sin(\alpha_2) + d_c,$$

$$d_1 \cos(\alpha_1) = d_2 \cos(\alpha_2) = d_P.$$

Combining these equations, the value of d_1 can be found as:

$$d_1 = \frac{d_c \cos(\alpha_2)}{\sin(\alpha_1 - \alpha_2)}.$$

Hence, the distance d_P between the symmetry axis and the point $\mathbf{P}$ is given by:

$$d_P = \frac{d_C \cos(\alpha_1) \cos(\alpha_2)}{\sin(\alpha_1 - \alpha_2)} \tag{2.17}$$

The distance d_o from the baseline's mid-point to $\mathbf{P}$ may be found in different ways. By the Law of Cosines, $d_0^2 = (d_c/2 + c)^2 + d_1^2 - 2(d_c/2 + c)d_1 \sin(\alpha_1)$, or by Pythagoras' theorem, $d_0 = \sqrt{d_P^2 + (d_2 \sin(\alpha_2) + (d_c/2 - c))^2}$. Both expressions, together with the previous equations, allow the computation of d_0 in terms of α_1 and α_2. The angle α_0 between an horizontal axis and the line segment joining point $\mathbf{P}$ to the baseline's mid-point can be computed as $\alpha_0 = \arctan((d_1 \sin(\alpha_1) - (d_c/2 + c))/d_P)$ or as $\alpha_0 = \arccos(d_P/d_0)$.

2.3.2. Horizontal Arrangement

Consider now the horizontal arrangement of the mirrors shown in fig. 2.3, in which the cameras are at a distance d_c (baseline), and the point $\mathbf{P}$ in the world has polar coordinates (r_1, θ_1), and (r_2, θ_2) with respect to the origin of the right and left camera coordinate

11
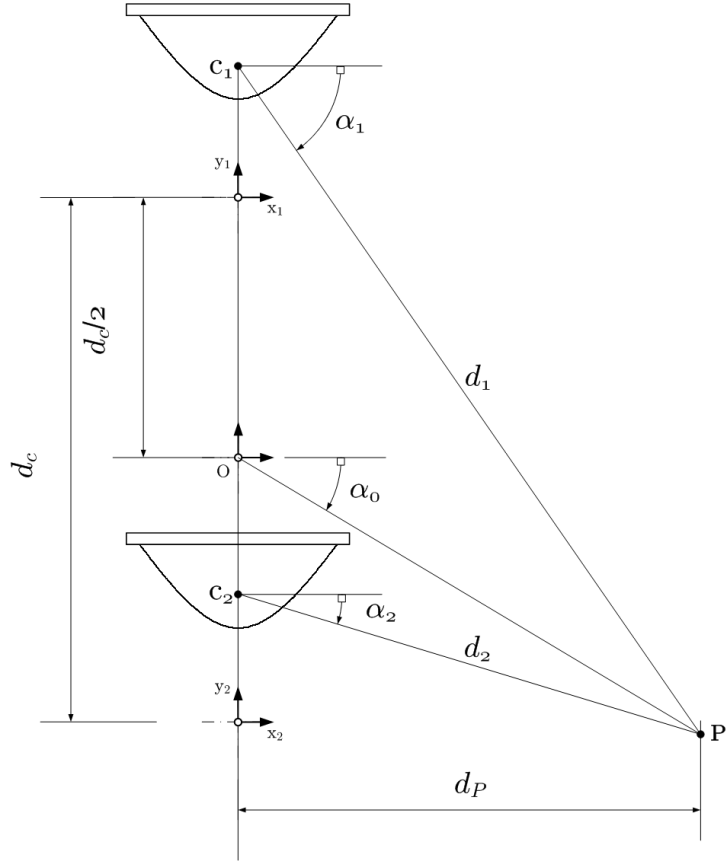
FIGURE 2.2. Lateral view of the vertical arrangement.

systems, respectively. Then from the geometry of the arrangement:

$$r_1 \sin(\theta_1) = r_2 \sin(\theta_2) = d_P \sin(\theta_0) \tag{2.18}$$

$$r_2 \cos(\theta_2) = d_c + r_1 \cos(\theta_1) \tag{2.19}$$

$$d_P \cos(\theta_0) = \frac{d_c}{2} + r_1 \cos(\theta_1) \tag{2.20}$$

Combining (2.18) and (2.19), allows to express r_1 in terms of θ_1 and θ_2 as

$$r_1 = \frac{d_C \sin(\theta_2)}{sin(\theta_1 - \theta_2)} \tag{2.21}$$

12

FIGURE 2.3. Top view of the horizontal arrangement.

Replacing r_1 into (2.18) and (2.20) the expressions for $\sin(\theta_0)$ y $\cos(\theta_0)$:

$$\sin(\theta_0) = \frac{d_c}{d_P} \frac{\sin(\theta_1)\sin(\theta_2)}{\sin(\theta_1 - \theta_2)}, \qquad (2.22)$$

$$\cos(\theta_0) = \frac{d_c}{d_P}\left(\frac{1}{2} + \frac{\sin(\theta_2)cos(\theta_1)}{sin(\theta_1 - \theta_2)}\right), \qquad (2.23)$$

which can be used to calculate $\tan(\theta_0)$, and therefore

$$\theta_0 = \arctan\left(\frac{\cos(\theta_1 - \theta_2) - \cos(\theta_1 + \theta_2)}{\sin(\theta_1 + \theta_2)}\right) \qquad (2.24)$$
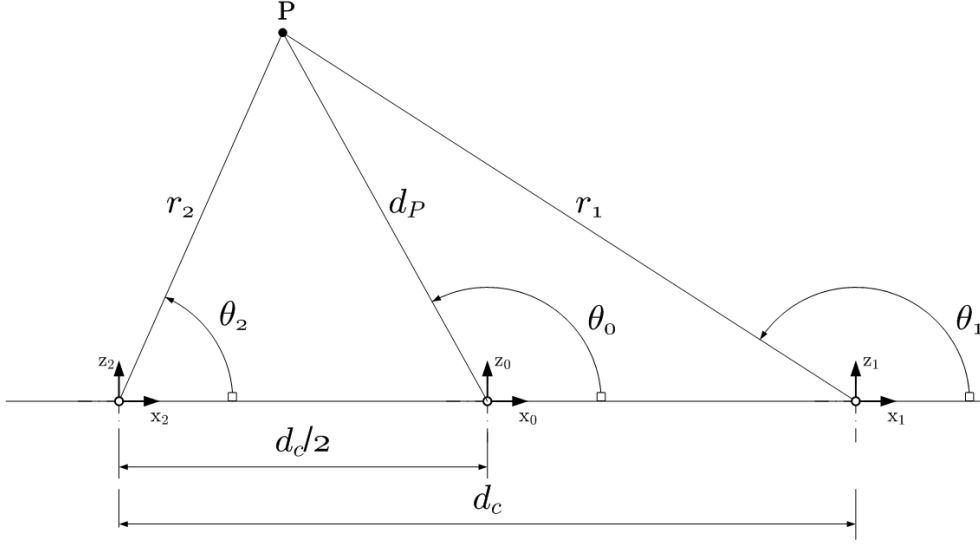
For the computed value of θ_0 from (2.24), the distance to point $\mathbf{P}$ from the baseline's midpoint is calculated employing (2.22) as

$$d_P = \frac{d_c \sin(\theta_1)\sin(\theta_2)}{\sin(\theta_1 - \theta_2)\sin(\theta_0)}, \qquad (2.25)$$

or directly in terms of θ_1 and θ_2 as

$$d_P = \frac{d_c\left(\cos^2(\theta_1 - \theta_2) + 2\cos(\theta_1 - \theta_2)\cos(\theta_1 + \theta_2) + 1\right)}{2\sin(\theta_1 - \theta_2)},$$

by solving for d_P from the addition of the squared equations (2.22) and (2.23).

13

2.3.3. Computation of the Points on the Epipolar Curves

In order to compute points of an epipolar curve on the left camera for a point on the optical plane of the right camera with Cartesian coordinates (u_1, v_1) the direction of the incident ray defined by α_1 is first computed using equations (2.14) and (2.16). To completely establish the orientation of the incident ray, the value of θ_1 is computed as $\theta_1 = \arctan(v_1/x_1)$. The orientation of the incident ray on the left mirror defined by angles α_2 and θ_2 can be computed for a given value of r_1. Thus, for varying values of r_1 the epipolar curve on the left mirror can be found numerically. It is to be noted that the epipolar curve is the locus of the points resulting from the intersection of a plane containing the triangle $\triangle \mathbf{c}_1 \mathbf{P} \mathbf{c}_2$ and the hyperboloid surface of the mirrors. These intersections correspond to a collection of conic sections for different locations of the point $\mathbf{P}$ in the 3D world. Although it is possible to find analytic expressions for the conic sections, the resulting expressions would have to be evaluated at incremental steps along the curve and the points projected onto the optical plane to find the matching point on the other camera. Hence, a numerically simpler approach is to find different values of the incident ray angles α_2, θ_2 for varying values of r_1, and then for the computed α_2, θ_2 find the angle β_2 using (2.15) and the corresponding coordinates (u_2, v_2) on the second optical plane. To this end, θ_2 in terms of r_1 is computed as

$$\theta_2 = \arctan\left(\frac{r_1 \sin(\theta_1)}{d_c + \cos(\theta_1)}\right) \tag{2.26}$$

as can be verified from the geometry of the arrangement shown in fig. 2.3.

To find the location of the corresponding point on the optical plane, first the distance r_2 from the point $\mathbf{P}$ to the origin of the left camera is computed, recalling (2.21) as

$$r_2 = \frac{d_c \sin(\theta_1)}{\sin(\theta_1 - \theta_2)}$$

14

Now from the geometry, $\tan(\alpha_1) = \mathbf{P_Z}/r_1$ and $\tan(\alpha_2) = \mathbf{P_Z}/r_2$, for the point $\mathbf{P}$ with vertical position $\mathbf{P_Z}$. Thus

$$\alpha_2 = \arctan\left(\frac{r_1 \tan(\alpha_1)}{r_2}\right). \tag{2.27}$$

Using equation (2.15) it is possible to obtain β_2 with the value of α_2 just computed. Defining $\rho_2 = \sqrt{u_2^2 + v_2^2}$, by (2.16),

$$\rho_2 = f \tan\left(\beta - \frac{\pi}{2}\right) = \frac{f}{\tan(\beta)}. \tag{2.28}$$

Finally, the Cartesian coordinates (u_2, v_2) are simply given by $(\rho_2 \cos(\theta_2), \rho_2 \sin(\theta_2))$.

2.4. Weighted Correlation

In order to reduce the warping effect produced by the hyperboloidal mirrors, which causes a nonlinear size reduction to objects as they stand farther away from the camera, the correlation is computed using weighting coefficients. The purpose of the coefficients is to reduce the level of importance of those pixels close to the boundaries of the block, while given more attention to the central pixel. By doing this, large and more reliable blocks can be employed without being significantly affected by the fact that one of the sliding blocks may consider regions of the image which are far away, therefore significantly distorted with respect to the central region of the reference block. The computation of the weighted cross-correlation employes the (i, j) weights matrix:

$$\mathbf{M_w} = \begin{bmatrix} w_{1,1} & \cdots & w_{1,q+1} & \cdots & w_{1,2q+1} \\ \vdots & & \vdots & & \vdots \\ w_{q+1,1} & \cdots & w_{q+1,q+1} & \cdots & w_{q+1,2q+1} \\ \vdots & & \vdots & & \vdots \\ w_{2q+1,1} & \cdots & w_{2q+1,q+1} & \cdots & w_{2q+1,2q+1} \end{bmatrix}$$

Which contains the weighting coefficients of each pixel within the correlation window. If the upper-left corner of the correlation block is considered as the origin, then pixel with

coordinates $(row, col) = (i, j)$ is weighted by element $w_{i,j}$ of the matrix M_w. Matrix M_w is constructed by assigning a unitary value to the central element at $(row, col) = (q+1, q+1)$, and decreasing values towards the edges until a minimum value $w_{Min} \in [0, 1]$ is reached at each corner of the matrix M_w. The expression for the computation of $w_{i,j}$ is given by:

$$w_{i,j} = 1 - (1 - w_{Min}) \cdot \frac{\|(i,j) - (q+1, q+1)\|}{q\sqrt{2}} \tag{2.29}$$

In order to obtain a normalized correlation value in the interval $[-1, 1]$, normalized weights $\tilde{w}_{i,j}$ are employed instead of $w_{i,j}$. The normalized weights are computed as:

$$\tilde{w}_{i,j} = \frac{w_{i,j}}{\sum_{i=1}^{2q+1} \sum_{j=1}^{2q+1} w_{i,j}} \tag{2.30}$$

The weighting matrices are illustrated for different values of w_{Min} and block sizes in table 2.2, in which the values of each weight are represented in terms of an intensity level in gray scale ($0 =$black, $1 =$white).

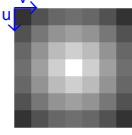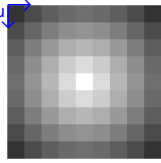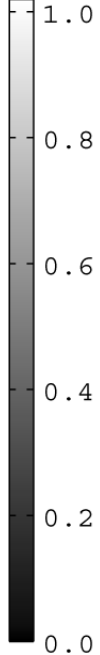The weighted cross-correlation between blocks A and B can then be computed as:
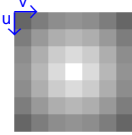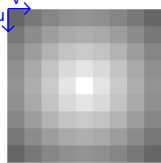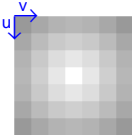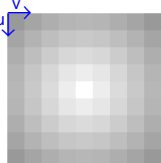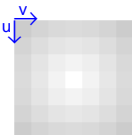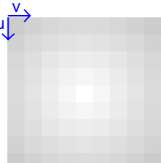
$$Corr_w = \frac{\sum_{i=1}^{2q+1} \sum_{j=1}^{2q+1} \tilde{w}_{i,j} \left(A_{i,j} - \bar{A}\right) \left(B_{i,j} - \bar{B}\right)}{\sqrt{\sum_{i=1}^{2q+1} \sum_{j=1}^{2q+1} \tilde{w}_{i,j} \left(A_{i,j} - \bar{A}\right)^2 \sum_{i=1}^{2q+1} \sum_{j=1}^{2q+1} \tilde{w}_{i,j} \left(B_{i,j} - \bar{B}\right)^2}} \tag{2.31}$$

Where $\bar{A}$ and $\bar{B}$ are the mean intensity values for each block.

2.5. Computational Cost

The computational cost of the algorithm implemented to obtain the distance maps arises from the calculation of angles α_1 and θ_1, together with the cost of fetching an image

TABLE 2.2. Weight matrices.

| w_{Min} | Size of the correlation block $[pixel] \times [pixel]$ | | | Weight |
	5×5	7×7	9×9	color bar
0.2				
0.4				
0.6				
0.8				

block of $(2q + 1) \times (2q + 1)$ pixels for each pixel in the reference image. In addition to the previous cost, the computational effort involved in the calculation of the epipolar curve and the matching process must also be included. In the case of the horizontal arrangement, the epipolar curve is explored for increments of r_1 in the interval $[r_{Min}, r_{Max}]$ using steps proportional to C_P in each iteration. In other words, the value of r_1 at iteration k denoted by $r_1(k)$ is computed as:

$$r_1(k) = C_p r_1(k - 1), k = 1, 2, ..., K$$

with

$$r_1(0) = r_{Min}$$

17

therefore

$$r_1(k) = r_{Min}C_p^k$$

and the value of $k = K$ when $r_1(K) = r_{Max}$ is given by

$$r_{Max} = r_{Min}C_p^K \Rightarrow K = \log_{C_p}\frac{r_{Max}}{r_{Min}}$$

This number K is the number of times that the correlation between the reference and sought block have to be computed. In the case of the vertical arrangement, the cost associated to the search along a radial epipolar line is limited to the calculation of lower and upper bounds for the search radius, the same idea of backprojection implemented in the algorithm of sec. 2.3.3, and the necessary operations to compute the points in the interval incrementally without using the backprojection strategy of the algorithm in sec. 2.3.3. Traversing the epipolar line in the vertical arrangement is done by incrementing ρ_2 in the interval $[\rho_{Min}, \rho_{Max}]$ for each $\theta_2 = \theta_1$ in steps $\Delta\rho_2$ per iteration. The bounds of interval $[\rho_{Min}, \rho_{Max}]$ are obtained from the lower and upper bounds of the working distance values d_{Min} and d_{Max} by solving the backprojetion equations for $d_1 = d_{Min}$ and $d_1 = d_{Max}$ (in the reference image), respectively. In order to obtain a more accurate distance estimation, the peak in the correlation curve is approximated at the subpixel level using bidimensional quadratic interpolation. To compute the parameters of the interpolating paraboloid, the correlation values found before and after the maximum are also stored together with their correponding locations on the epipolar curve.

Finally, the cost of solving the distance equations for the maximum correlation point must also be added to obtain the total computational cost. Thus the cost for an $N \times N$ image can be respectively expressed for the vertical and horizontal arrangement as follows:

$$
\begin{aligned}
C_{Vertical} &= N^2\left(41C_1 + 28C_2 + 2C_3 + 74C_4 + 5C_5 + C_v\right) \\
&+ N^2\left(\frac{\rho_{Max} - \rho_{Min}}{\Delta\rho}\right)\left(3C_1 + 2C_2 + 2C_3 + 4C_4 + C_v + C_c\right)
\end{aligned}
$$

$$
\begin{aligned}
C_{Horizontal} \;=\;\; & N^2 \left(33C_1 + 18C_2 + 2C_3 + 56C_4 + 3C_5 + C_v \right) \\
& + N^2 \log_{C_p} \left(\frac{r_{Max}}{r_{Min}} \right) \left(6C_1 + 7C_2 + 2C_3 + 13C_4 + C_5 + C_v + C_c \right)
\end{aligned}
$$

Where C_1, C_2, C_3, C_4 and C_5 are the number of CPU cycles required to computed sums/subtractions, trascendental functions, rounding operations, multiplications/divisions, and algebraic functions, respectively. The cost C_v is associated to the cost of fetching a block of size $(2q+1) \times (2q+1)$, while C_c is the cost of computing the correlation for a block of the same dimension.

If the computation of the distance is not done with subpixel accuracy the main computational effort is that of multiplications/divisions, can be reduce by $1/2$ for the vertical arrangement, while the reduction factor is $1/3$ in the case of the horizontal configuration. In each case, the computational cost without subpixel accuracy is given by:

$$
\begin{aligned}
C_{Vertical} \;=\;\; & N^2 \left(18C_1 + 23C_2 + 2C_3 + 37C_4 + 4C_5 + C_v \right) \\
& + N^2 \left(\frac{\rho_{Max} - \rho_{Min}}{\Delta \rho} \right) \left(3C_1 + 2C_2 + 2C_3 + 4C_4 + C_v + C_c \right)
\end{aligned}
$$

$$
\begin{aligned}
C_{Horizontal} \;=\;\; & N^2 \left(10C_1 + 13C_2 + 2C_3 + 19C_4 + 2C_5 + C_v \right) \\
& + N^2 \log_{C_p} \left(\frac{r_{Max}}{r_{Min}} \right) \left(6C_1 + 7C_2 + 2C_3 + 13C_4 + C_5 + C_v + C_c \right)
\end{aligned}
$$

A detailed flow diagram of the algorithm for panoramic stereoscopic computation has been included in the Appendix A, together with the associated computational cost of each step. Careful study of the process flow and loops allows to obtain the expressions for the computational cost just derived.

3. DISTANCE ESTIMATION ERRORS AND ARRANGEMENT SELECTION

3.1. Distance Estimation Errors

Employing the equations derived in the previous section, it is possible to compute the distance estimation error for (i) errors in the parameters of the system or (ii) errors in the computation of the corresponding pixel coordinate on the other camera for a given point $\mathbf{P}$ in the world. The first type of error can be avoided by an adequate construction of the system and corrected with standard calibration techniques (Micusik & Pajdla, 2004), while mitigating errors of the second kind is more difficult as these do not depend only on the matching technique, but also on the visual properties of the scene (e.g. presence of texture, edges, periodic patterns or illumination changes). For these reasons, we limit our attention to the analysis of distance estimation errors under pixel matching errors.

In the previous section it was shown that for a given point $\mathbf{P}$ in the world with projection coordinates (u_1, v_1) on the optical plane of the first camera, the corresponding angles α_1, β_1 can be found in terms of (u_1, v_1). Also, α_2, β_2, as well as (u_2, v_2), can be found in terms of the matching point (u_1, v_1) for a given value of the distance d_P. If d_P is solved as a function of (u_1, v_1) and the matching (u_2, v_2), i.e. $d_P = f(u_1, v_1, u_2, v_2)$, then for an error of one pixel in finding the location of (u_2, v_2), i.e. $(\tilde{u}_2, \tilde{v}_2) = (u_2 + \Delta u_2, v_2 + \Delta v_2)$, such that $\|(\Delta u_2, \Delta v_2)\| = 1$ and $(\tilde{u}_2, \tilde{v}_2) \in \mathcal{C}^e$, where $\mathcal{C}^e$ denotes the epipolar curve, the average distance estimation error can be computed as:

$$
\begin{aligned}
\Delta d_P = \frac{1}{2} \left(|f(u_1, v_1, u_2, v_2) - f(u_1, v_1, u_2 - \Delta u_2, v_2 - \Delta v_2)| \right. \\
\left. + |f(u_1, v_1, u_2, v_2) - f(u_1, v_1, u_2 + \Delta u_2, v_2 + \Delta v_2)| \right)
\end{aligned}
\tag{3.1}
$$

Here $(-\Delta u_2, -\Delta v_2)$ and $(+\Delta u_2, +\Delta v_2)$ must be understood as the bounding endpoints of the epipolar curve segment around the matching location (u_2, v_2). In other words, the error is computed as the average of the errors obtained when the mismatch is due to an error in locating the corresponding pixel by one pixel in excess or due to falling short by one pixel.
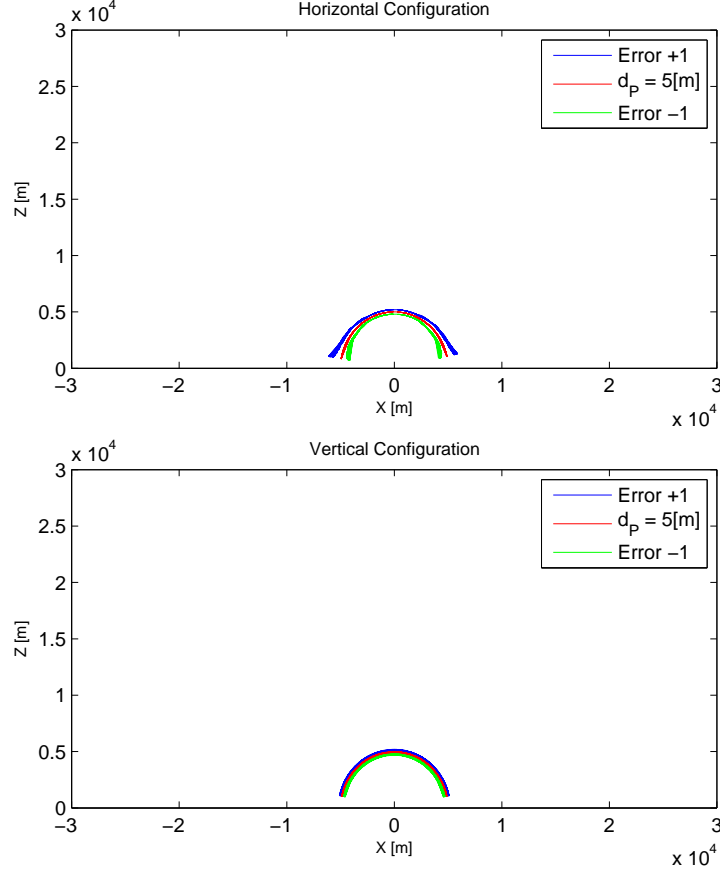
FIGURE 3.1. Working distance circle and estimation error margins for $d_c = 0.5\,m$ and $d_P = 5\,m$

The error intervals for different working distances d_P and looking angles θ_0 appear as semi-circles around the true distance, as shown in figs 3.1– 3.4. It is possible to see that the error interval becomes larger as the working distance increases from $5\,m$ (fig. 3.1), $10\,m$(fig. 3.2), $15\,m$ (fig. 3.3) to $20\,m$ (fig. 3.4), independently of whether the horizontal or vertical configuration is being used. However, for looking angles $\theta_0 = 0$ or $\theta_0 = \pi$, the error under the horizontal configuration becomes larger more rapidly than for the vertical arrangement. This is because equations for the horizontal configuration become singular at looking angles $\theta_0 = 0$ or $\theta_0 = \pi$.
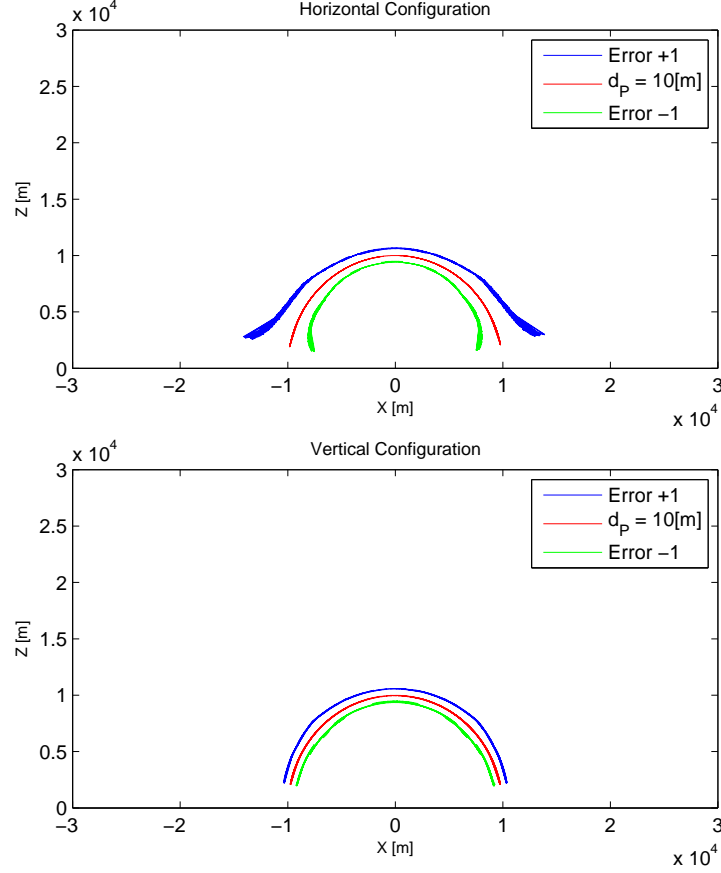
FIGURE 3.2. Working distance circle and estimation error margins for $d_c = 0.5\,m$ and $d_P = 10\,m$

Another way to analyze the distance estimation error under each configuration is by ploting the average error Δd_P, which is approximately half the width of the error band, for different looking directions and baseline values. These error curves are shown in figs. 3.5(a)-3.5(d) for both camera configurations and observation angles $\theta_0 = \pi/6$, $\pi/4$, $\pi/3$, $\pi/2$. Each figure shows different error curves for the vertical configuration (lines with circle markers) and for the horizontal configuration (dotted lines) at different baselines $d_c = 0.5, 1.0, 1.5\,m$. These curves can be obtained for a range of values of the observation angle θ_0, which are not presented here due to space. Comparing the curves for each configuration is possible to conclude that:
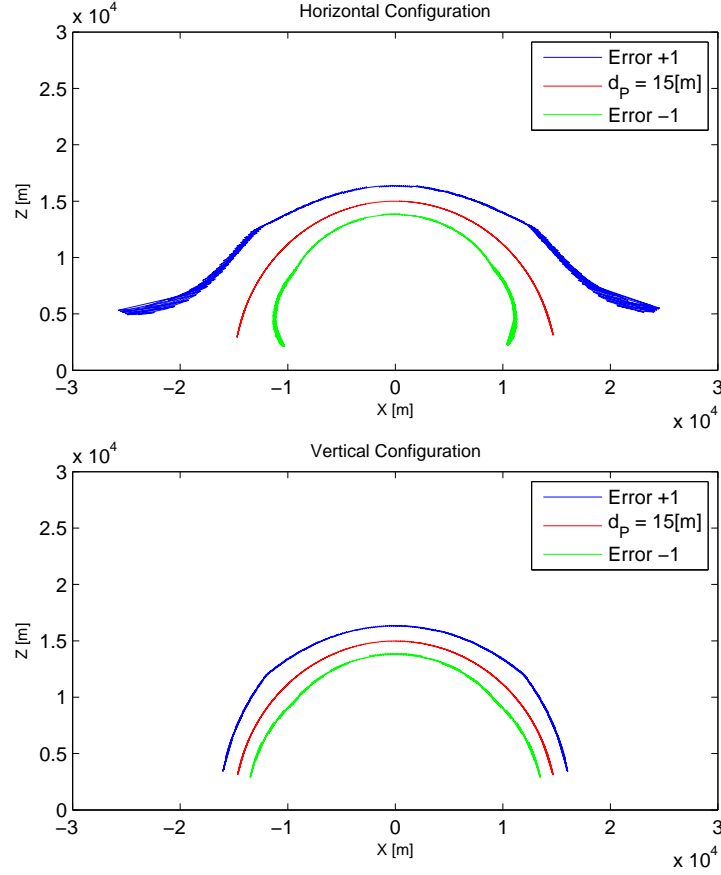
FIGURE 3.3. Working distance circle and estimation error margins for $d_c = 0.5\,m$ and $d_P = 15\,m$

- The distance estimation error increases more rapidly for smaller baselines as in the case of a stereoscopic system with standard perspective cameras.
- For a given distance, there is not much difference between the horizontal or vertical configuration if the observation angle considered is $\theta_0 = \pi/2$ (forward looking) (fig. 3.5(a)). As the observation angle decreases, i.e. looking to the sides, the error in the horizontal configuration rapidly becomes larger than that in the vertical configuration.
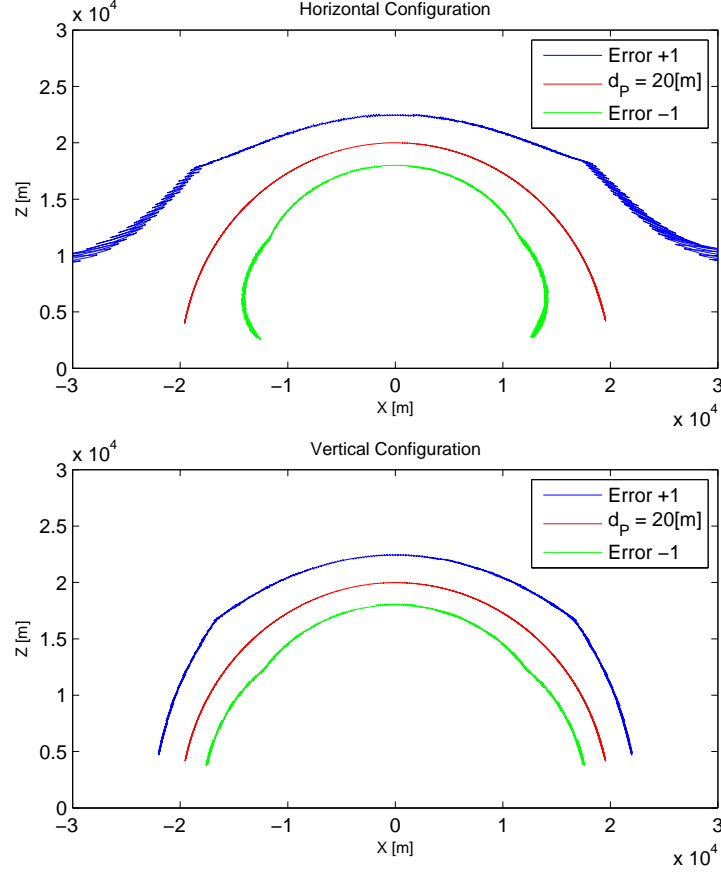
FIGURE 3.4. Working distance circle and estimation error margins for $d_c = 0.5\,m$ and $d_P = 20\,m$

- Considering that smaller distance estimation errors are obtained for larger baselines, the horizontal configuration becomes more convenient for mounting on a mobile platform or vehicle.

3.2. Camera Arrangement and Baseline Selection Procedure

The error curves presented in the previous section give us some insight into the criteria for selecting the most convenient configuration for a given application. Figure 3.6 shows a selection procedure diagram that will allow designers to choose the most convenient arrangement and the appropriate baseline given the acceptable depth estimation error.
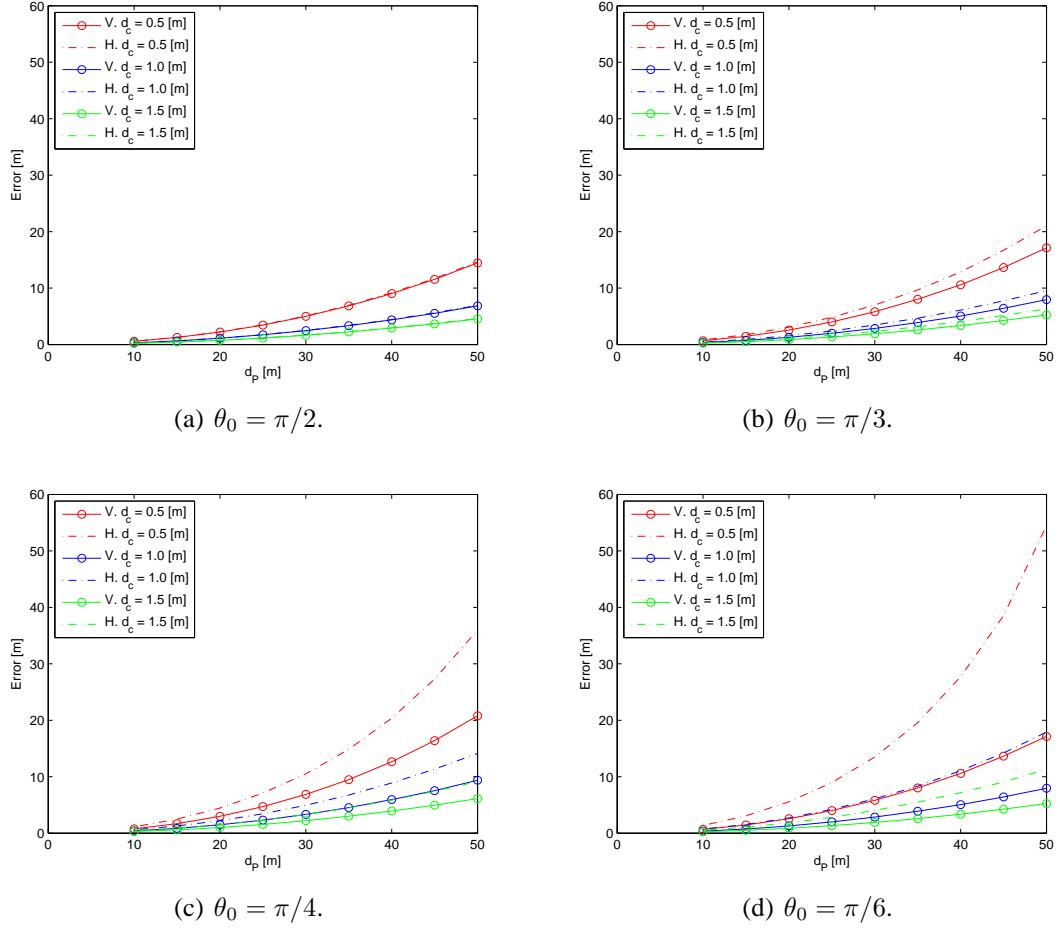
FIGURE 3.5. Distance error at different θ_0 values.

In the first step the user must check if his or her system has similar geometry parameters. If so, then figs. 3.5(a)– 3.5(d) showing the expected error at versus working distance for different baselines, can be used to select the smallest baseline that guarantees an acceptable level of error for the specified working distance. If the system has significantly different intrinsic parameters, differing more than 5% in one or more of the parameters, then the designer must redo the error curves using the projective geometry equations and the epipolar curve computation procedure presented in the previous sections (see sec. 2.2 and 2.3). It is possible that under certain conditions, both arrangements satisfy the acceptable levels of error for a given baseline. If this situation occurs then the designer should prefer the vertical configuration over the horizontal arrangement because the vertical arrangement yields

full $360°$ panoramic distance maps free from singularities. On the other hand, the vertical configuration involves a simpler stereo matching process because the epipolar curves are simple straight lines, and moreover, the estimation error does not depend on the looking direction θ_0. When both arrangement comply with the aceptable error, the only reason for not prefering the vertical configuration would be if there are space constraints. Finally, if the error requirements cannot be archived by neither of the arrangements, the designer must check if the design of the catadioptric system has adequate parameters (i.e. intrinsic parameters, such as camera resolution, focal distance, hyperbola parameters).
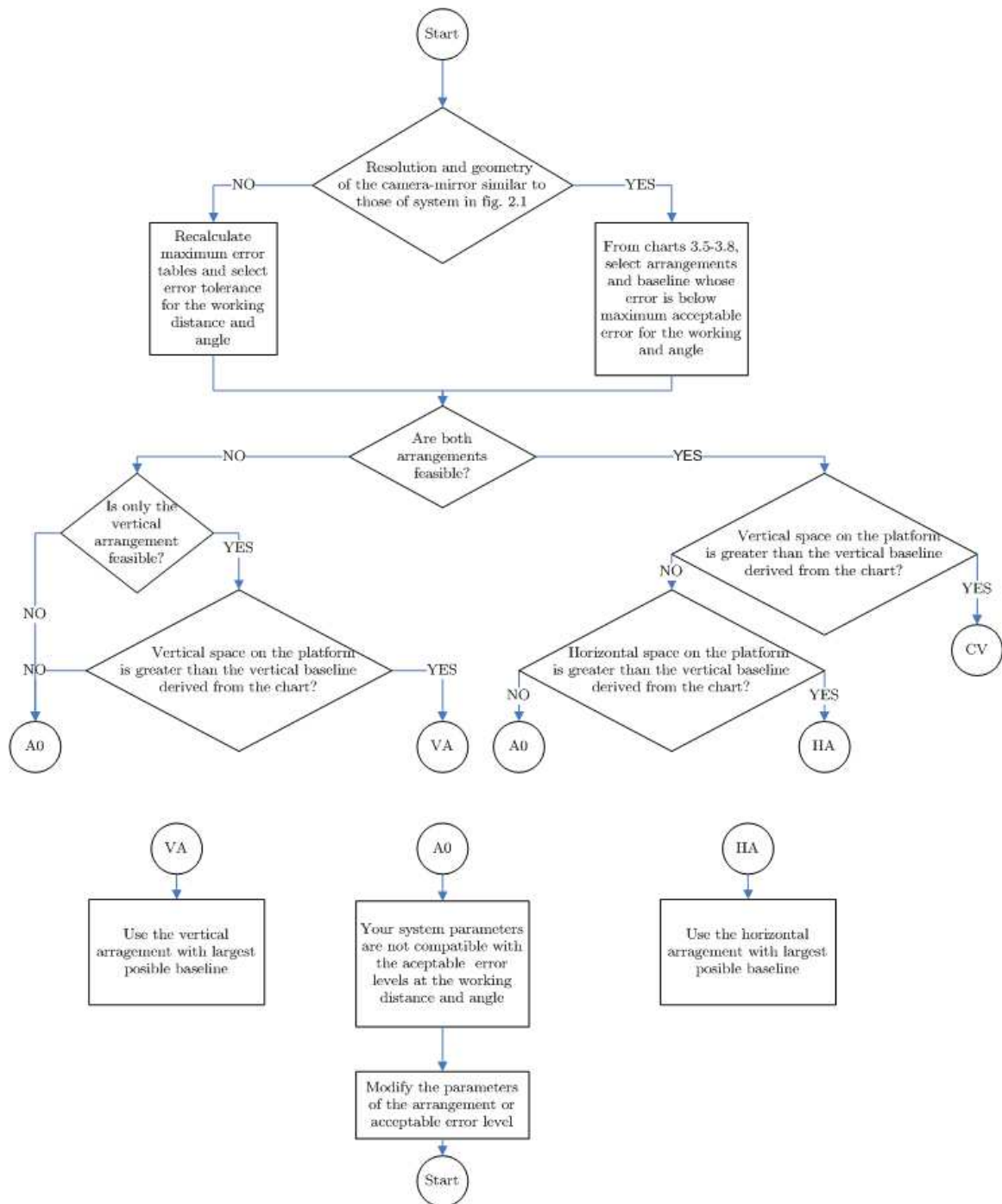
FIGURE 3.6. Camera arrangement and baseline selection procedure

4. EXPERIMENTAL RESULTS

### 4.1. Experimental Validation Using Synthetic Images

A synthetic world is used to evaluate the distance estimation accuracy under each arrangement. The virtual world is a cube (e.g. a room) to which textures have been mapped onto each wall in order to ensure the solvability of the stereoscopic matching problem. Figure 4.1 (a) shows a perspective view of the virtual world. Since our purpose is to evaluate the accuracy of each configuration, rather than the performance of the stereoscopic matching procedure, the actual textures employed are derived from fractal and synthetic random patterns (see fig. 4.9).

The cylinder in the middle of the cube in fig. 4.1 (a) represents the omnidirectional vision sensor. In the vertical configuration both cameras are located along a vertical axis through the center of the cube. In the horizontal configuration the cameras are on a plane parallel to the ground plane and offset to the right and left of the vertical central axis along a horizontal axis, which intersects the vertical axis and runs parallel to two opposite lateral faces of the cube. The omnidirectional image as projected onto the optical plane of the camera is shown in fig. 4.1 (b). The latter translates into a panoramic image as shown in fig. 4.1 (c). The corresponding distance ground truth measured from the center of the omnidirectional vision sensor is shown as a panoramic distance map in fig. 4.1 (d).

Figures 4.9 (a) and (b) show the right and left omnidirectional images for the horizontal configuration. The regions shown as blue disc segments in fig. 4.9 (a) correspond to the reference pixels considered for matching, while the regions in fig. 4.9 (b) correspond to the matched pixels found. Not all pixels where successfully matched and thus there are some holes in the regions of fig. 4.9 (b). It is worth noting that unlike the case of standard parallel stereoscopic system with perspective cameras, the epipolar constraint for the horizontal configuration with hyperboloiodal mirrors imposes a corresponding region that does not preserve shape similarity with respect to the original reference region, and hence, the difficulty in finding correspondences in the horizontal configuration.
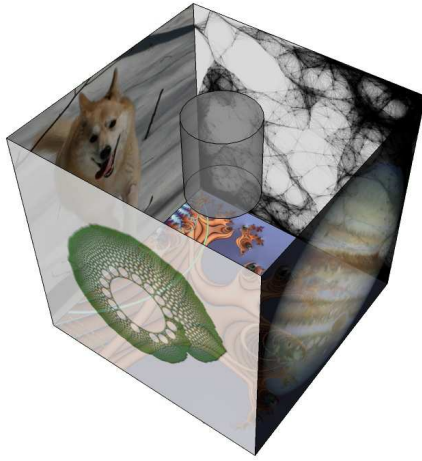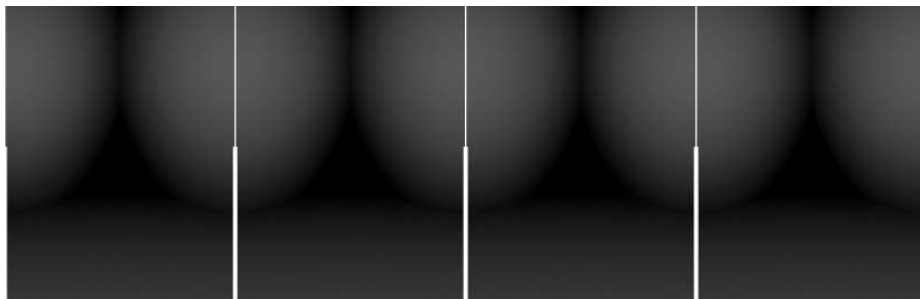
(a) Perspective view.
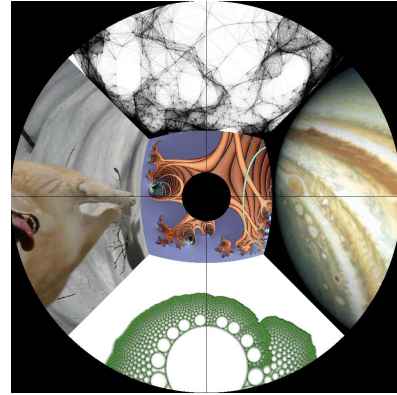
(b) Omnidirectional image.
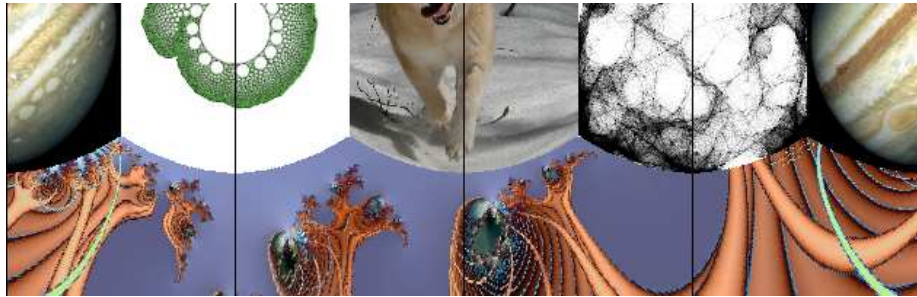
(c) Panoramic image.


(d) Panoramic distance map.

FIGURE 4.1. The test cube and its projections as an omnidirectional and panoramic images.

Given the complexity of the epipolar curves arising in the horizontal configuration, it is worth verifying the validity of the equations derived in section 2.3.3. To this end, four points are selected in the synthetic omnidirectional test image (left camera view), as shown in fig. 4.2(a). Then the corresponding epipolar curves on the second image (right camera view) are computed using the ecuations of section 2.3.3 and plotted, as shown in the fig. 4.2(b). It can be observed in fig. 4.2(b) that the epipolar curves in the target image pass through points that were selected in the reference image, thus confirming the validity of the equations. The same can be observed from the epipolar curves traced on the panoramic view of the target (secondary) image.

In order to compare the accuracy of the distance estimation under each configuration in a "controlled environment", each face of the synthetic test cube is covered using a chessboard pattern image (see fig. 4.4) in which each square corresponds to a $100 \times 100\,mm^2$ square in real world units. The colors of the chessboard pattern correspond to randomly chosen intensity levels of the grayscale palette, with the exception of the central squares and those at the corners of the chessboard. This facilitates the matching procedure for the points at the corners of each square within the chessboard. Figures 4.6 and 4.5 show the set of omnidirectional images obtained for the horizontal and vertical arrangement with $d_c = 120\,mm$, respectively. The analysis of the depth estimation accuracy considers only corner points in each view of the chessboard that satisfy that:

- Their projection is visible in each view of the left-right or top-bottom pair of the horizontal or vertical arrangement respectively,
- the angle $\theta_0 \in [\pi/15, 14\pi/15]$ or interval $\theta_0 \in [16\pi/15, 29\pi/15]$.

The second constraint is employed to make the results from both configurations comparable by avoiding the singularities of the horizontal arrangement occurring at $\theta_0 = 0$ and $\theta_0 = \pi$. Fig. 4.7(a) shows points in the right omnidirectional view that satisfy the above constraints, while fig. 4.7(b) shows the matched points in the left omnidirectional image. Similarly, fig. 4.8(a) shows the reference points in the top omnidirectional view.
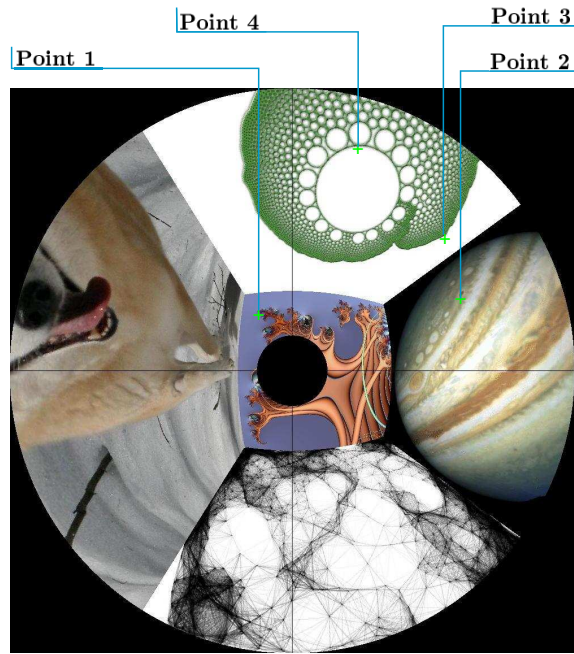
(a) Test points in the left omnidirectional image (reference)
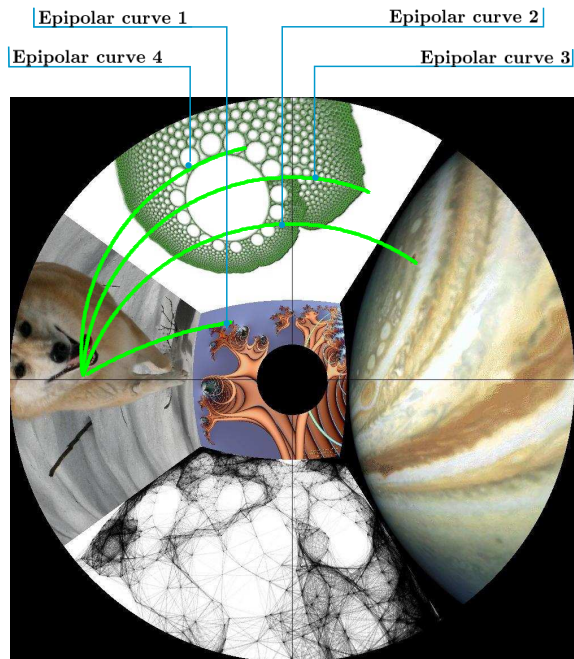
(b) Epipolar curves in the right camera omnidirectional image (target)

FIGURE 4.2. Epipolar curves in the right camera omnidirectional image (target).
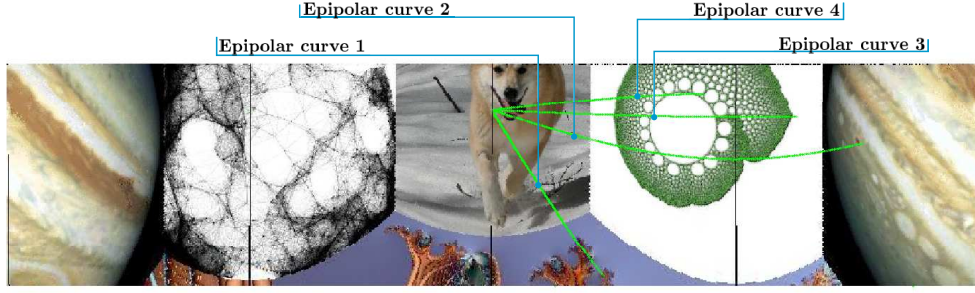
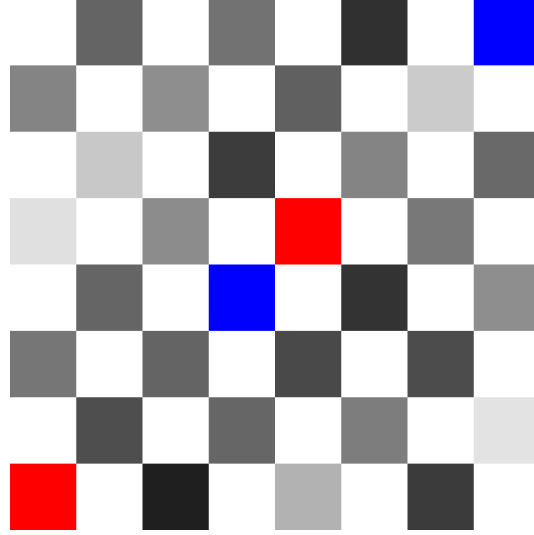FIGURE 4.3. Test points and epipolar curves using synthetic images.

FIGURE 4.4. Chessboard pattern.

The corresponding matched points found in the bottom omnidirectional view are shown in fig. 4.8(b).

If d_{0i} denotes the real distance between the origin of the system and the control point $\mathbf{P}_i$, and $\hat{d}_{0i}$ denotes the distance found by solving the matching problem, then the absolute value of the distance error is given by $e_i = \left| \hat{d}_{0i} - d_{0i} \right|$, and the average absolute error is $\bar{e} = \frac{1}{M} \sum_{i=1}^{M} e_i = \frac{1}{N} \sum_{i=1}^{N} \left| \hat{d}_{0i} - d_{0i} \right|$, where M is the number of selected reference points. Computing $\bar{e}$ for different values of the block size $(2q+1) \times (2q+1)$, w_{Min} and baseline d_c
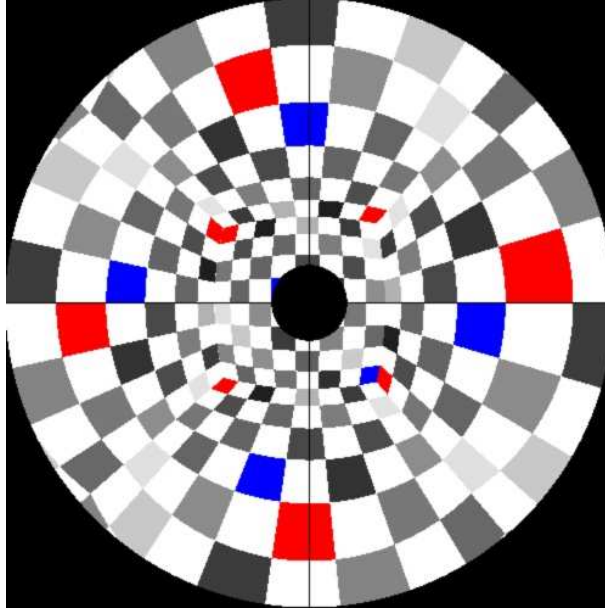
(a) Right camera view
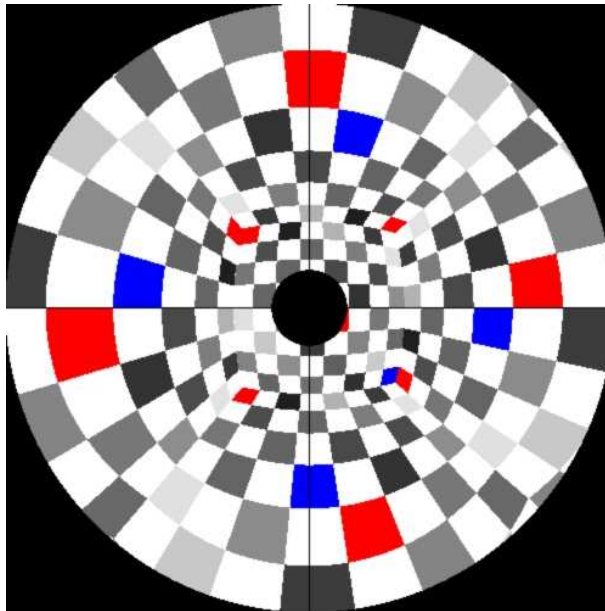
(b) Left camera view

FIGURE 4.5. Synthetic omnidirectional stereoscopic image pair of the cube with chessboard pattern for the horizontal arrangement with $d_c = 120\,mm$.

(a) Top camera view
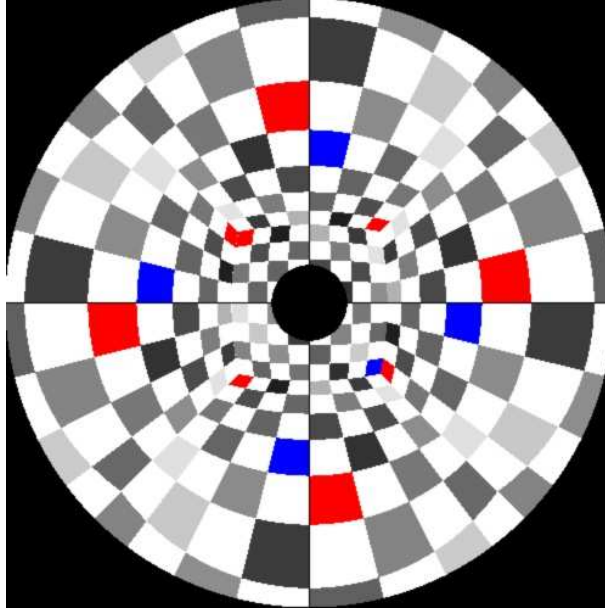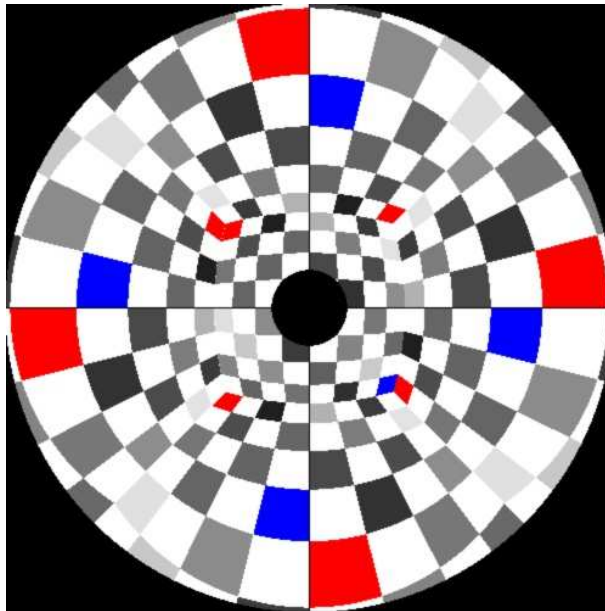

(b) Bottom camera view

FIGURE 4.6. Synthetic omnidirectional stereoscopic image pair of the cube with chessboard pattern for the vertical arrangement with $d_c = 120\,mm$.
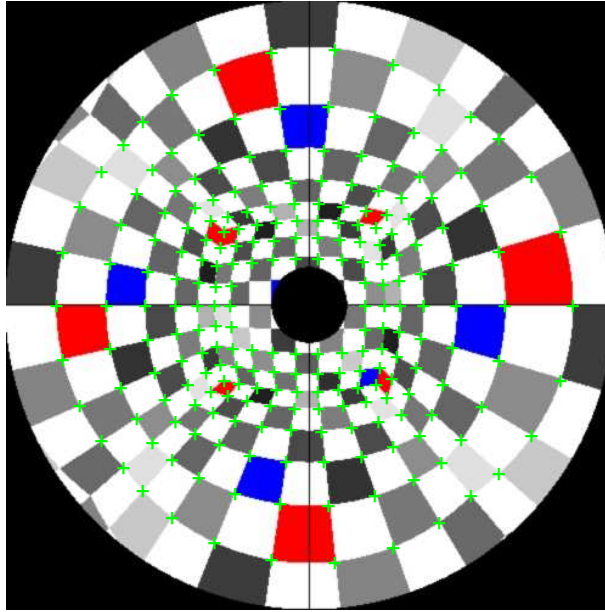
(a) Reference points in the right camera view
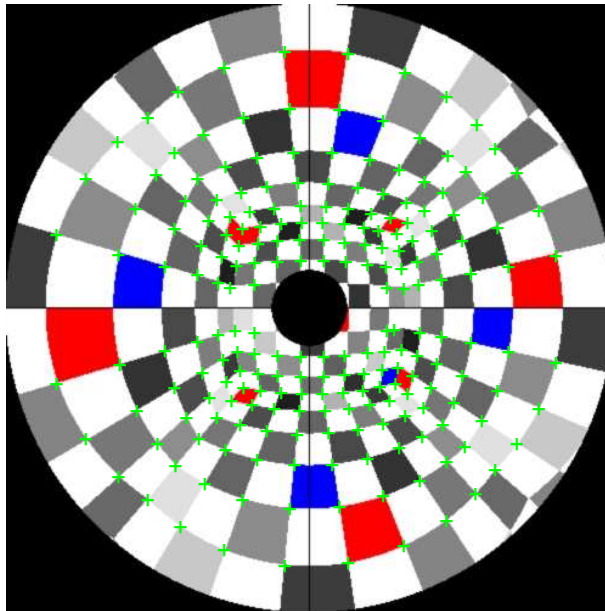
(b) Matched points in the left camera view

FIGURE 4.7. Stereoscopic matching for the vertical arrangement using the synthetic test cube with chessboard pattern.
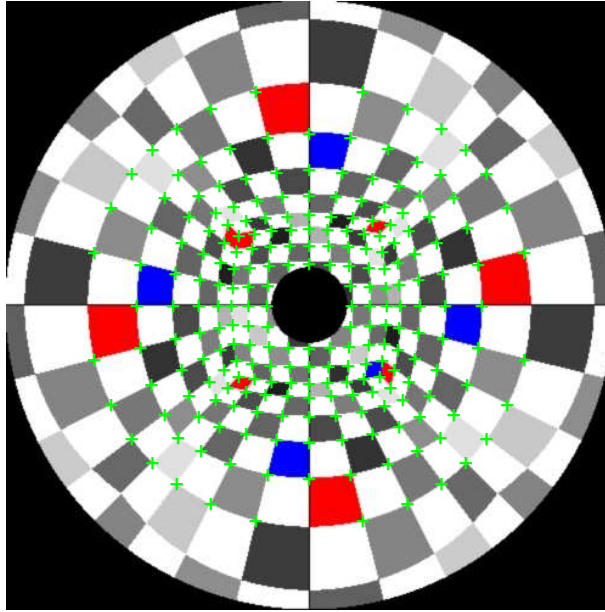
(a) Reference points in the top camera view
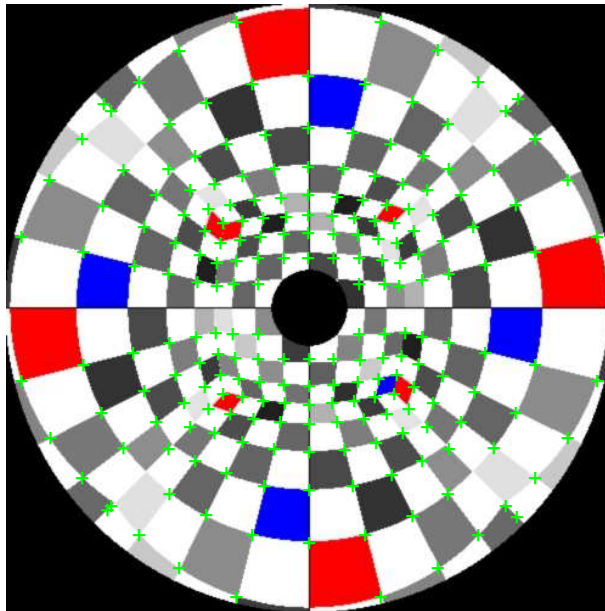
(b) Matched points in the bottom camera view

FIGURE 4.8. Stereoscopic matching for the horizontal arrangement using the synthetic test cube with chessboard pattern.

allows to determine the optimal block size. The result of the experiments for the horizontal and vertical configurations are summarized in tables 4.1 and 4.2, respectively.

TABLE 4.1. Simulation results for the horizontal configuration.

| d_c	$\bar{e}$	$\sigma(e)$	e_{iMax}	$(2q+1) \times (2q+1)$	w_{Min}
$[mm]$	$[mm]$	$[mm]$	$[mm]$	$[pixel] \times [pixel]$	$[\%]$
20	65.8	52.1	350.3	13×13	90
40	55.0	41.5	271.2	13×13	20
60	36.7	28.5	136.8	17×17	0
80	27.0	22.9	105.6	9×9	10
100	23.9	19.5	80.1	9×9	90
120	21.2	16.8	86.1	13×13	70
140	16.9	12.9	58.0	13×13	50
160	15.2	12.4	58.4	13×13	30
180	15.0	12.3	73.8	13×13	30

The estimation errors were computed for different values of the baseline d_c in increments of $20\,mm$, with $d_c \in [20, 180]\,mm$. As predicted from the error curves presented in sec. 3.1, the average error decreases when the separation between cameras increases. Equivalent trends may be observed for the error's standard deviation and the maximum error. Comparing tables 4.1 and 4.2 for both arrangements, it is possible to conclude that for a given baseline the vertical arrangement has on average a slightly large average estimation error. This is due to the reduced effective resolution along the radial direction of the camera-mirror system. As far the optimal block size is concerned, it is possible to observe that under the vertical configuration, larger baseline requires larger block sizes, while the optimal window size for the horizontal configuration remains about the same regardless of the distance between the cameras. Dense distance maps computed for another pair of synthetic omnidirectional images (see figs. 4.9(a) and 4.9(b)) using standard normalized cross-correlation to solve the stereoscopic correspondence problem are shown in figs. 4.9(c) and 4.9(d) for the horizontal and vertical configurations, respectively. In both cases, the

TABLE 4.2. Simulation results for the vertical configuration.

d_c [mm]	$\bar{e}$ [mm]	$\sigma(e)$ [mm]	e_{iMax} [mm]	$(2q+1) \times (2q+1)$ [pixel] $\times$ [pixel]	w_{Min} [%]
20	72.3	63.8	248.8	9×9	50
40	60.3	60.3	279.5	5×5	100
60	42.3	48.2	257.8	13×13	20
80	37.8	48.8	262.3	9×9	60
100	33.3	46.6	266.9	13×13	20
120	31.5	46.9	271.4	17×17	10
140	27.5	39.5	191.4	21×21	50
160	26.8	45.5	197.5	21×21	70
180	25.1	38.8	197.5	21×21	70

arrangement was centered about the origin of the global (world) coordinate frame. The baseline distance between focal points of each hyperboloidal mirror was $20\,mm$. The disparity search range for the horizontal configuration corresponds to that defined by a point $200\,mm$ away from the mirror on ℓ_1 the line $\overline{\mathbf{cP}}$ in $10\,mm$ steps until no significant change in the corresponding point projection is achieved for 10 iterations in a row. The vertical configuration uses a disparity search range which spans the current reference point position up to 90% the radius of the circle inscribed in the CCD. The experiments consider mirror parameters shown in fig. 2.1 and a cube of $800\,mm$ by side.

Considering only the correctly matched points, the relative 3D reconstruction (RMS/true distance) is 5.32% and 7.89% for the horizontal and vertical configurations, respectively. However, for observation angles $\theta_0 \approx 0$ or $\theta_0 \approx \pi$, the distance estimation error in the horizontal configuration is appreciably larger.
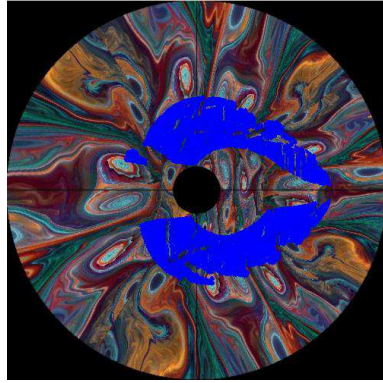
(a) Reference area in the right omni-directional image
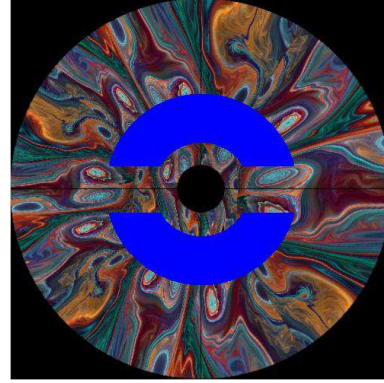
(b) Matched area in the left omnidirectional image
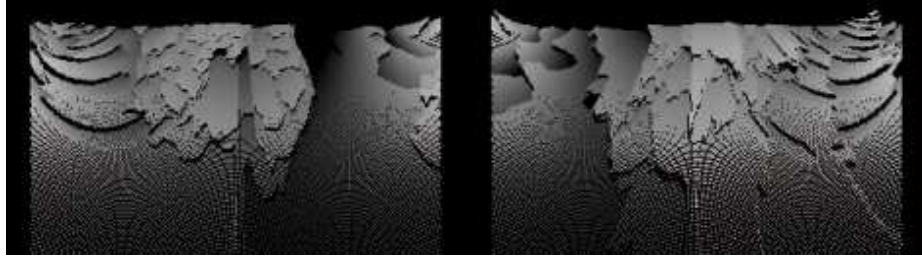
(c) Distance maps for the horizontal configuration.
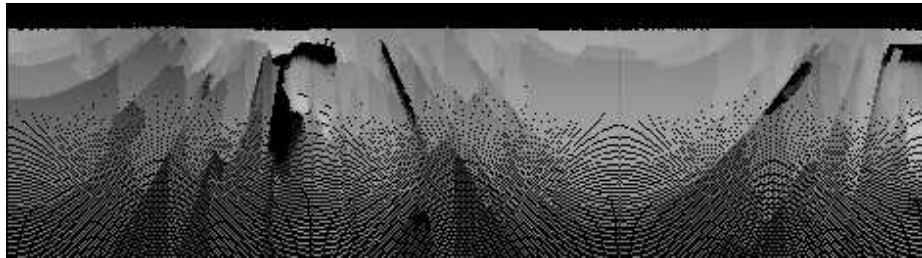
(d) Distance maps for the vertical configuration.

FIGURE 4.9. Dense distance maps for the synthetic test cube with random pattern.

4.2. Experiments with a Real Vehicle

A panoramic stereoscopic system can be very useful in the implementation of collision alert useful in the implementation of a collision alert system for autonomous mobile robots or standard human driven vehicles. Therefore it is important to assess the performance of proposed catadioptric sensor arrangements in real world scenes with varying illumination conditions and different levels of texture in order to determine the practical feasibility of

the system. This section explains the experiments carried out with the sensing arrangement mounted on a standard car driven in urban areas and presents the distance estimation results obtained.

4.2.1. Defining the Working Distance

Defining the operational range of the omnidirectional stereoscopic arrangement to be mounted on a vehicle must consider safe driving distances to other vehicles. A safe driving distance depends on several aspects, such as the driving speed, visibility, road material and conditions, vehicle type, driver skills, reaction time and fatigue, to name a few. Due to the diversity of aspects which can affect the driver's reaction time, the breaking time (the time elapsed between the driver's action and the moment the vehicle comes to a full stop) and the breaking distance, here we will rely on criteria for determining safe driving distances provided by (*The Official Highway Code For Northern Ireland*, 2008). Basically, the criteria establishes from experimental studies that a standard passenger vehicle in a urban road at $50\,km/h$ requires a total breaking distance of $23\,m$, of which $9\,m$ are due to the driver's reaction time and $14\,m$ are the typical breaking distance for a passenger car. If in addition a safety distance of $8\,m$ corresponding to the length of two vehicles is also considered, then the working distance amount is $31\,m$.

4.2.2. Mounting the Cameras on the Vehicle and Selecting the Arrangement

Once the operating distance has been determined the level of acceptable error must be defined in order to employ the arrangement selection procedure. Considering that at the operating distance of $31\,m$, an acceptable error would be $\pm2.5\,m$ (the length of one car), i.e. less than 10% of the length of the operating range, then with this level of tolerable distance estimation error fig. 3.5(a) indicates that at $31\,m$ both the horizontal and vertical arrangements with baselines $d_c = 1\,m$ or $d_c = 1.5\,m$, satisfy the application requirements. Using the selection procedure of fig. 3.6 the vertical option would be preferable unless there are space constraints. Although it would be possible to mount the vertical arrangement on

the roof of a vehicle extending $1.5\,m$ above its standard height (see fig. 4.10) practical considerations make the horizontal arrangement (see fig. 4.11) more convenient. In fact, prior work by (Bertozzi et al., 2002) consider that one of the basic requirements of any driver assistance system is that it should be compact in size and affect as less as possible the vehicle structure. Some authors as (Bensrhair et al., 2002) have suggested mounting the cameras inside the vehicle, while others like (McCall & Trivedi, 2006) position the sensing arrangement on top of the roof with system mounted at lower heights, recommending to mount the cameras on the lateral rear-view mirrors. The latter work mentions as additional advantages of mounting the cameras on the lateral rear-view mirrors: the possibility of studying the driver's behavior, the larger baseline and the compactness of the system. A disadvantage, however, is that the covered area is smaller than the area visible using the system mounted on the roof, which offers a better view of the road. The cameras mounted on the rear-view mirrors mostly include the view of the vehicle on which they are mounted, instead of the surrounding road. Due to these reasons, including the in depth study by (Gandhi & Trivedi, 2005, 2006), and the possibility of mounting the cameras with variable baselines, it was decided that the most convenient location for experimental evaluation purpose would be to mount the cameras on the roof of the vehicle as shown in figs. 4.10 and 4.11 show the rig mounted on hatchback and sedan versions of the Toyota car Yaris model. Figures 4.12(b) and 4.12(b) show the right and left camera omnidirectional view for the horizontal arrangement configuration. The cameras were mounted $82\,cm$ apart, i.e. $d_c = 82\,cm$.
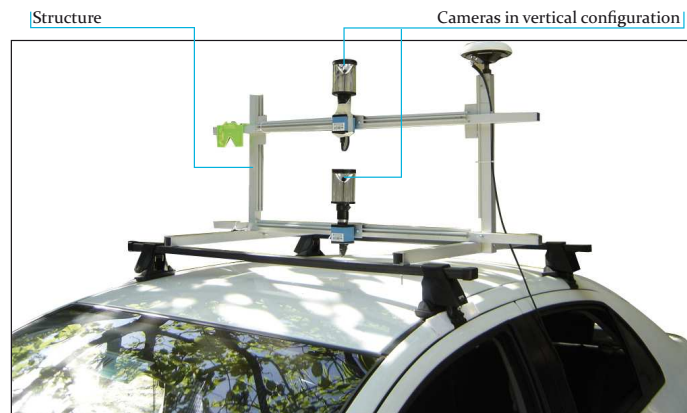

FIGURE 4.10. Structure and vertical cameras arrangement in the vehicle

FIGURE 4.11. Structure and horizontal cameras arrangement in the vehicle

(a) Right camera view

(b) Left camera view

FIGURE 4.12. Street view using the omnidirectional stereoscopic pair in the horizontal arrangement with $d_c = 82\,cm$

(a) Right camera view

(b) Left camera view

FIGURE 4.13. Panoramic view obtained from the omnidirectional stereoscopic pair of fig. 4.12.

4.2.3. System Calibration

System calibration was done manually using a calibration pattern. The first step consisted in aligning the cameras with respect to the pattern until crosshairs matched the alignment marks. By doing so it is possible to reduce the source of error in the posterior estimation of intrinsic and extrinsic parameters. Since manual alignment can practically eliminate axial misalignment, and relative rotation of one camera with respect to the other. Only a few parameters need to be estimated. These parameters are the focal length f, the baseline (d_c), and the hyperbola constants a,b. To this end, marks of another calibration pattern similar to cylinder employed by (Matuszyk, Zelinsky, Nilsson, & Rilbe, 2004) are used to align the symmetric axis for the camera and the mirror. Then, another pattern (see figs. 4.14 and 4.15) are matched and the values of the parameters are adjusted until their projections on both cameras match their actual locations in the image, i.e. the norm of the error between the projected points and the points on the image is minimized. This simple calibration approach relying on calibration tools similar to those of (Matuszyk et al., 2004) proved to be sufficient for our purposes. More automated calibration approaches exist in the literature; the reader is referred to (Hartley & Zisserman, 2000) for an in depth discussion on calibration techniques. Another interesting work is the one by (Scaramuzza, Martinelli,

& Siegwart, 2006), the author shows in his website an omnidirectional camera calibration toolbox for Matlab called *OCamCalib*.

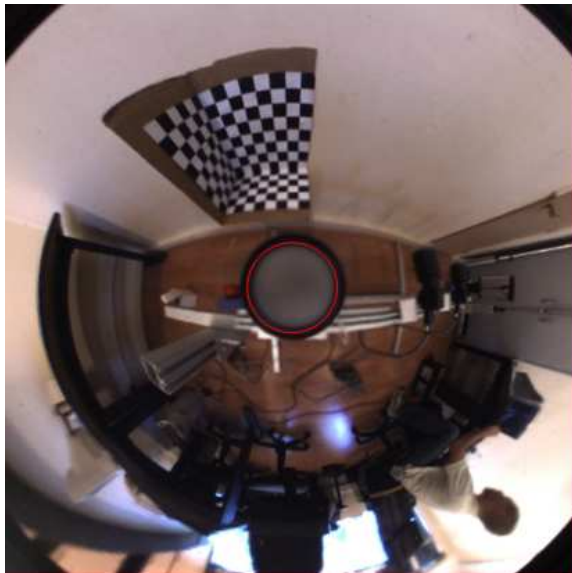FIGURE 4.14. Cameras and calibration pattern.

FIGURE 4.15. Camera view of calibration pattern.

The correctness of the alignment and calibration steps can also be verified manually by simple procedures. First, if the cylindrical pattern is used correctly aligned cameras should provide panoramic images that show perfectly horizontal lines when the omnidireccional image is dewarped (mapped onto cylinder much like a Mercator projection). Secondly,

when tracing the epipolar curves corresponding to the reference points in the primary images, the curves in the target image should contain the projection of associated points. Fig. 4.19(a) shows four manually selected points and fig. 4.16(b) shows the corresponding epipolar curves traced on the target image. It is possible to observe that the corresponding points in the target images are contained by the respective epipolar curves, and therefore the system parameters have been estimated with a sufficiently good accuracy. Fig. 4.17 shows a panoramic view corresponding to the omnidirectional image 4.16(b) in which it is also possible to see that the reference points are contained by the epipolar curves.
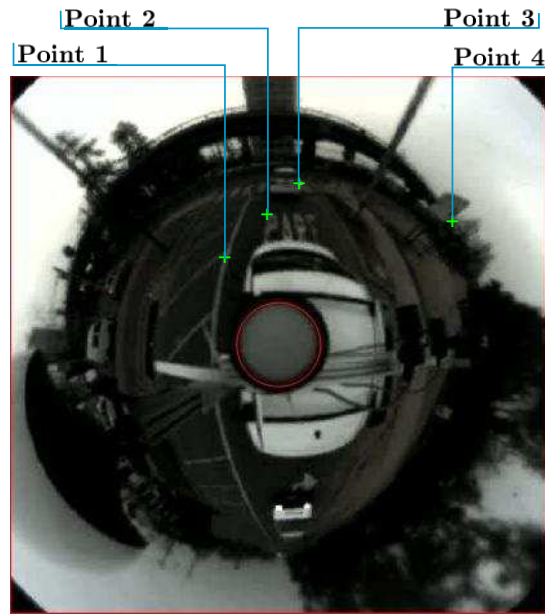
(a) Reference points in the right omnidirectional camera view
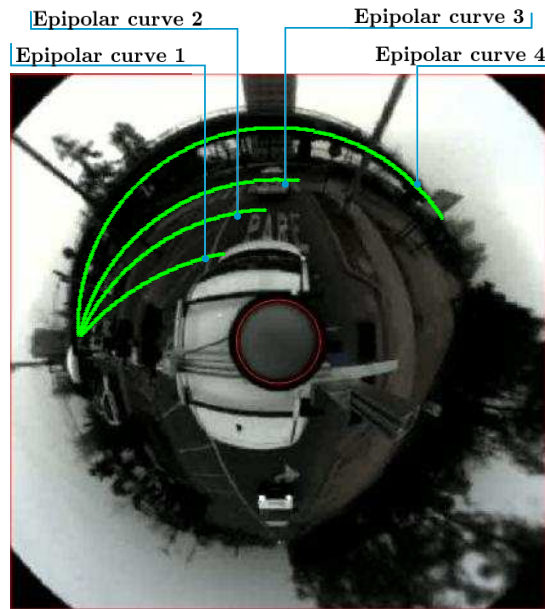
(b) Epipolar curves for the reference points in the left omnidirectional camera view

FIGURE 4.16. Epipolar curves for the omnidirectional view of the road using the horizontal arrangement.
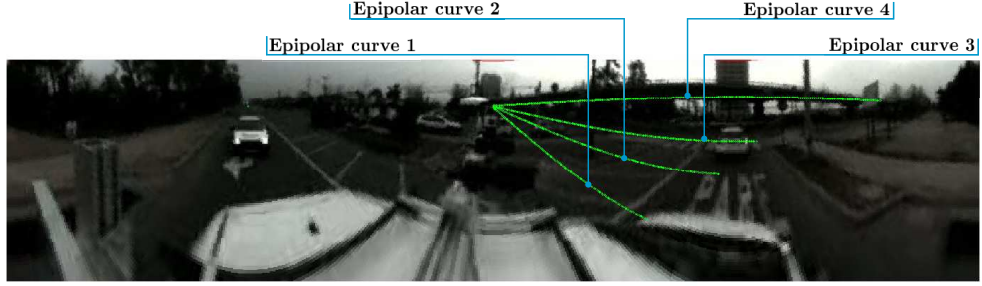
47

FIGURE 4.17. Epipolar curves on the panoramic view of the road.

4.2.4. Correspondence Search Range

Adjusting the length of the epipolar curve on which correspondence are sought is particularly important for real world applications because the chances of erroneous match increase when an excessively long search range is considered likely due to lack of texture or the presence of periodic patterns. As mentioned in section 4.1, adjusting the length of the epipolar curve is done considering the lower and upper bounds of the working distance interval. Some characteristics of the application considered here allow us to obtain tighter constraints on the correspondence search range. More specifically, by assuming that the vehicle moves on a plane, then the points of the epipolar curve must correspond to points laying on the ground plane or before, i.e. the epipolar curves will be limited by points corresponding to the projection of points in space belonging to the ground plane as depicted in fig. 4.18. For all points below the horizon, i.e. $\alpha_1 > 0$ using our convention (see fig. 4.18) there is an r_{Max} associated to the orthogonal projection of the intersection between the ray from the origins of the coordinate system and the ground plane onto the XZ-plane. Distances r_1 from the coordinate origin to any object in the scene will be at most r_{Max}, hence $r_1 \leq r_{Max}$. By simple geometry:

$$r_{Max} = \frac{h}{\tan(\alpha_1)}$$

48

where h is the distance from the origin of the primary system to the ground plane. Comparing figs. 4.16(b) and 4.19(b) it is possible to see that the epipolar curves are significantly shortened in the latter and that the shorter curves still contain the points corresponding to the reference points shown in fig. 4.19(a). The projection of the epipolar curves considering the ground plane constraint onto the panoramic image are shown in fig. 4.20. Once gain the reader may verify that the shorter epipolar curves contain the reference points of interest.
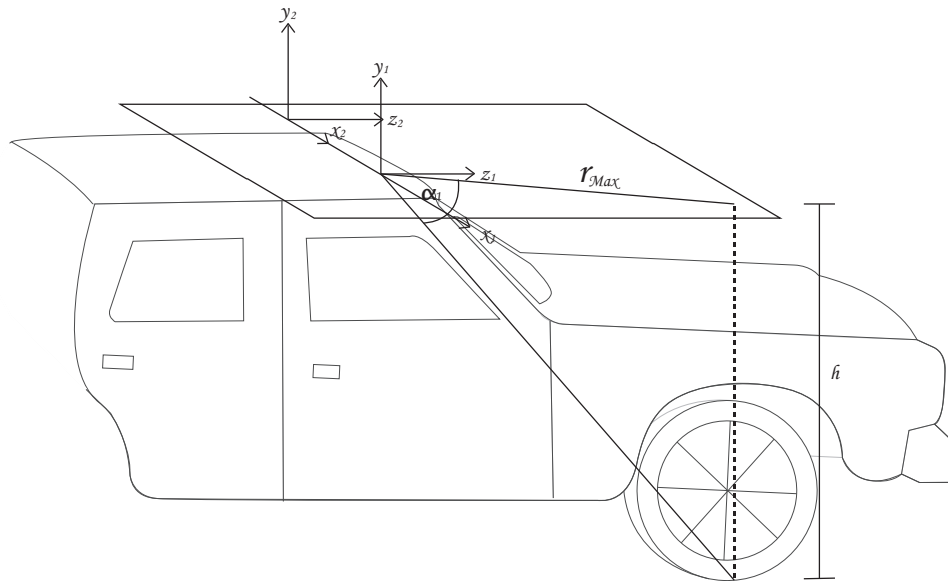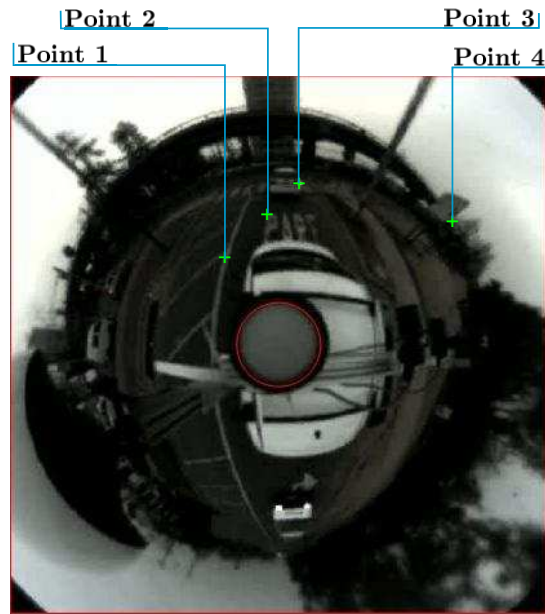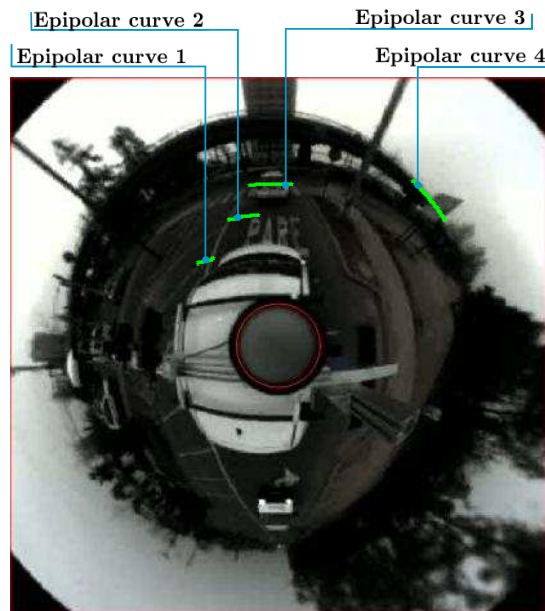
FIGURE 4.18. Maximum search distance for the system mounted on a vehicle.

(a) Reference points in the right omnidirectional camera view

(b) Epipolar curves for the reference points in the left omnidirectional camera view

FIGURE 4.19. Constrained epipolar curves for the omnidirectional view of the road using the horizontal arrangement.
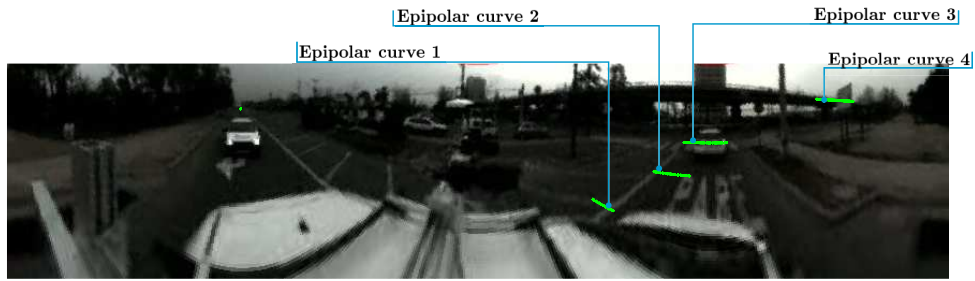
50

FIGURE 4.20. Constrained epipolar curves on the panoramic view of the road.

4.2.5. Matching and Dense Distance Map Results

Before computing a dense panoramic distance map, the correctness of the matching process is verified using a selection of reference points shown on the right image of the omnidirectional stereoscopic pair of fig. 4.21. With the exception of a few points on areas without much texture, the majority of the points are matched correctly to points in the left image, as may be verified by following the connecting lines in fig. 4.21.
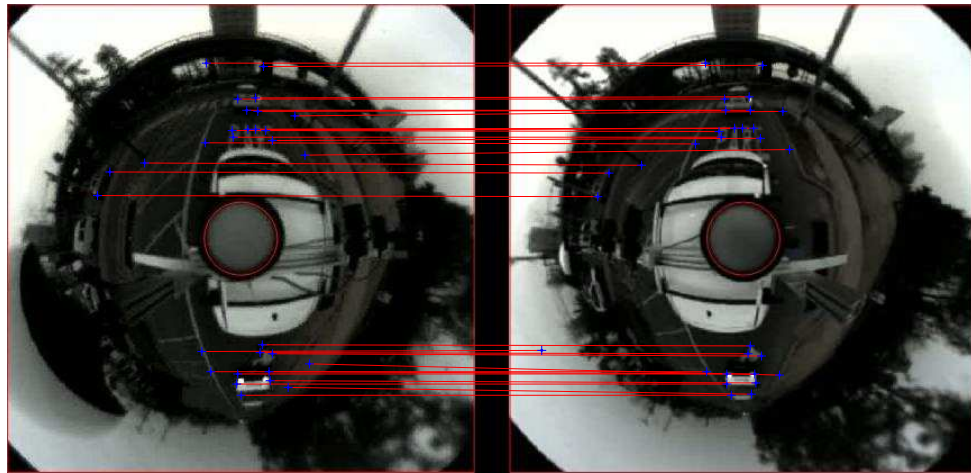

FIGURE 4.21. Matched points in the omnidirectional images.

Repeating the correspondence search for every point in the reference image and employing the distance equations derived in sec. 2.3 is possible to obtain a panoramic dense

distance map. Fig. 4.22(b) shows the reference region employed for the matching, which considers a $160°$ field-of-view towards front and $160°$ field-of-view towards the rear side of the vehicle. The resulting matched points on the target image are shown in fig. 4.22(a). It is possible to observe that the matching for the front region is more dense because of the richer texture. Figures 4.23(a) and 4.23(b) show the panoramic view of the front regions as seen by the right and left cameras, respectively. The dense distance map for this pair of views is presented in fig. 4.23, which shows the presence of the car ahead at $3.8\,m$. In a darker intensity of gray, the columms of the metropolitan railway platform can also be seen. Regions in the distance map closer to the vehicle, such as the road pavement can be seen in lighter shades of gray right at the lower middle area of the image. However, due to the lack of texture on the right side, wrongly matched points appear as very closer points to the vehicle. This problem unfortunately is frequent when attempting dense distance maps using stereoscopic matching techniques applied to scenes with little texture or repetitive patterns (Zitnick & Kanade, 2000). A possible way to eliminate these regions is to use confidence measures as proposed by (Torres & Guesalaga, 1999).
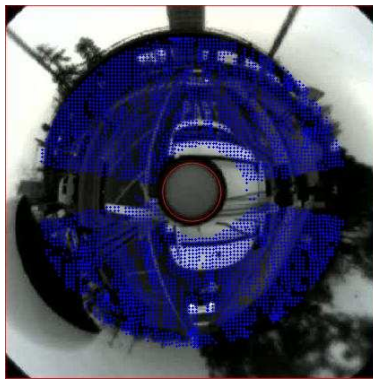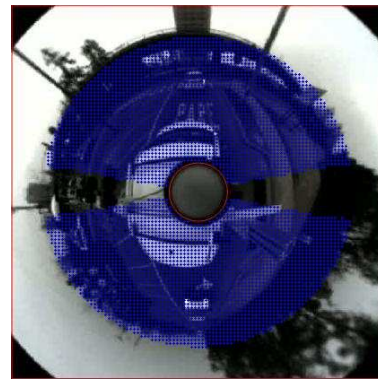
(a) Matched points on the target image   (b) Reference points

FIGURE 4.22. Matched areas for the reference regions of interest (front and rear $160°$ field-of-view).

(a) Right camera view

(b) Left camera view

(c) Distance map

FIGURE 4.23. Frontal panoramic road view and distance map.

5. CONCLUSION AND FUTURE RESEARCH

An analysis of two omnidirectional stereo configurations using hyperboloidal mirrors was presented. In one of the system configurations the mirrors were arranged vertically, while on the other, the mirrors where arranged horizontally. Both configurations may yield similar distance estimation errors for an adequately chosen baseline and given measurement range of interest. However, depending on the application, space constraints can make one arrangement preferable over the other.

Considering the numerical results for synthetic images, the horizontal configuration provides on average of about a 2% more accurate reconstruction if the strictly lateral view under which there is sensor occlusion and singularity are excluded. The solution to the matching problem is more difficult, but space constraints on mobile robots is another strong reason for preferring the horizontal arrangement. Application of the matching algorithm to real world urban road scenes did not yield results as good as those obtained for synthetic images, mainly due to lack of textures and oclusions. Another practical aspect that limits the use of panoramic distance maps obtained using catadioptric systems is the loss of actual resolution, and given the lowering price of cameras, a more reliable solution would be to use several pair of perspective cameras, especially if the system is intended for navgation applications requiring distance measurements laser range scanners still remain the most reliable solution. However, the price of $3D$ scanners is above \$50.000 USD, therefore it is important to continue dedicating efforts to the improvement of computer vision and image understanding techniques.

To summarize, the main contributions of this investigation are the:

  (i) detailed derivation of the optical geometry of the omnidirectional hyperboloidal stereoscopic system

  (ii) analysis of the distance estimation errors under each configuration

  (iii) implementation of an algorithm to numerically compute the epipolar curve by back-projection of reference points onto the optical plane of the target system

  (iv) generation of procedure to select an adequate arrangement

(v) analytic quantification of the computational cost associated to the solution of the matching problem under both arrangements.

The derivation of the equations in a clear manner has a tutorial value to a wide audience of researchers in the fields of robotics and computer vision considering the application of omnidirectional vision sensors to environment perception, pedestrian detection and navigation problems, to name a few. The analysis of the estimation errors provides general guidelines for system design and implementation. The procedure exploits the geometrical and epipolar constraints to numerically compute the points of the epipolar line, without requiring the analytic solution of the intersection problem between the epipolar plane and the hyperbolic mirrors, which would yield an elliptic curve that would then have to be reprojected onto the optical planes of each camera. This procedure proved to be computationally effective. Ongoing research is concerned with automating the system calibration procedure and introducing confidence measures to discard incorrectly matched points due to occlusions or lack of textures.

REFERENCES

Baker, S., & Nayar, S. (1999). A theory of single-viewpoint catadioptric image formation. *International Journal of Computer Vision*, *35*(2), 175–196.

Benosman, R., & Kang, S. (2001). *Panoramic vision: sensors, theory, and applications*. Springer Verlag.

Bensrhair, A., Bertozzi, M., Broggi, A., Fascioli, A., Mousset, S., & Toulminet, G. (2002). Stereo vision-based feature extraction for vehicle detection. In *IEEE Intelligent Vehicles Symp., Versailles, France*.

Bertozzi, M., & Broggi, A. (1998). Gold: A parallel real-time stereo vision system for generic obstacle and lane detection. *IEEE Transactions on Image Processing*, *7*(1), 62–81.

Bertozzi, M., Broggi, A., Cellario, M., Fascioli, A., Lombardi, P., & Porta, M. (2002). Artificial vision in road vehicles. *Proceedings of the IEEE*, *90*(7), 1258–1271.

Desouza, G., & Kak, A. (2002, Feb). Vision for mobile robot navigation: a survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *24*(2), 237-267.

Gandhi, T., & Trivedi, M. (2005, Sept.). Vehicle mounted wide fov stereo for traffic and pedestrian detection. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on* (Vol. 2, p. II-121-4).

Gandhi, T., & Trivedi, M. (2006, Sept.). Vehicle surround capture: Survey of techniques and a novel omni-video-based approach for dynamic panoramic surround maps. *Intelligent Transportation Systems, IEEE Transactions on*, *7*(3), 293-308.

Hartley, R., & Zisserman, A. (2000). *Multiple view geometry*. Cambridge University Press.

Matuszyk, L., Zelinsky, A., Nilsson, L., & Rilbe, M. (2004). Stereo panoramic vision for monitoring vehicle blind-spots. In *2004 IEEE Intelligent Vehicles Symposium* (pp. 31–36).

McCall, J., & Trivedi, M. (2006, March). Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation. *Intelligent Transportation Systems, IEEE Transactions on*, *7*(1), 20-37.

Micusik, B., & Pajdla, T. (2004, June-2 July). Autocalibration & 3d reconstruction with non-central catadioptric cameras. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on* (Vol. 1, p. I-58-I-65 Vol.1).

Nayar, S. K. (1988, Nov). Sphereo: recovering depth using a single camera and two specular spheres. In *Proceedings of SPIE: Optics, Illumination, and Image Sensing for Machine Vision II.*

The official highway code for northern ireland (Tech. Rep.). (2008). Department of the Environment.

Scaramuzza, D., Martinelli, A., & Siegwart, R. (2006, Jan.). A flexible technique for accurate omnidirectional camera calibration and structure from motion. In *Computer Vision Systems, 2006 ICVS '06. IEEE International Conference on* (p. 45-45).

Svoboda, T., & Pajdla, T. (2002). Epipolar geometry for central catadioptric cameras. *Int. J. Comput. Vision*, *49*(1), 23–37.

Torres, M., & Guesalaga, A. (1999). Confidence factor and adaptive disparity range for reliable 3d reconstruction. In *International Conference on Computational Intelligence for Modelling, Control and Automation, Vienna, Austria.*

Yagi, Y., Kawato, S., & Tsuji, S. (1994, Feb). Real-time omnidirectional image sensor (copis) for vision-guided navigation. *Robotics and Automation, IEEE Transactions on*, *10*(1), 11-22.

Yagi, Y., & Yachida, M. (1991). Real-time generation of environmental map and obstacle avoidanceusing omnidirectional image sensor with conic mirror. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91.* (pp. 160–165).

Yamazawa, K., Yagi, Y., & Yachida, M. (1993, Jul). Omnidirectional imaging with hyperboloidal projection. In *Intelligent Robots and Systems '93, IROS '93. Proceedings of the 1993 IEEE/RSJ International Conference on* (Vol. 2, p. 1029-1034 vol.2).

Zitnick, C., & Kanade, T. (2000, Jul). A cooperative algorithm for stereo matching and occlusion detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *22*(7), 675-684.

APPENDIX A. COMPUTATIONAL COST ESTIMATION

- $N \times N$: Image size in pixels
- $(2q + 1) \times (2q + 1)$: Correlation block size in pixels
- (u_1, v_1): Coordinates in the first (reference) image
- (u_2, v_2): Coordinates in the second image
- (u_{2Bef}, v_{2Bef}): Coordinates of the corresponding point in the second image computed in the previus cycle
- α_i: Angle between the incoming ray and the horizontal plane to coordinate system i
- (r_i, θ_i): Polar coordinates of a point in the world with respect to the upper focal point of the mirror system i
- (ρ_i, θ_i): Polar coordinates on the CCD for point (r_i, θ_i)
- d_P: Radial distance between the arrangement's vertical axis and a point in the world
- $corr_{Act}$: Correlation between blocks 1 and 2
- $corr_{Max}$: Maximum correlation value between blocks
- $corr_{MaxBef}$: Correlation value one iteration before finding the maximum correlation
- $corr_{MaxAft}$: Correlation value one iteration after finding the maximum correlation
- C_1: Additions and sustractions cost
- C_2: Trascendental functions cost
- C_3: Rounding cost
- C_4: Multiplication and divisions cost
- C_5: Algebraic functions cost
- C_v: $(2q + 1) \times (2q + 1)$ block fetching cost
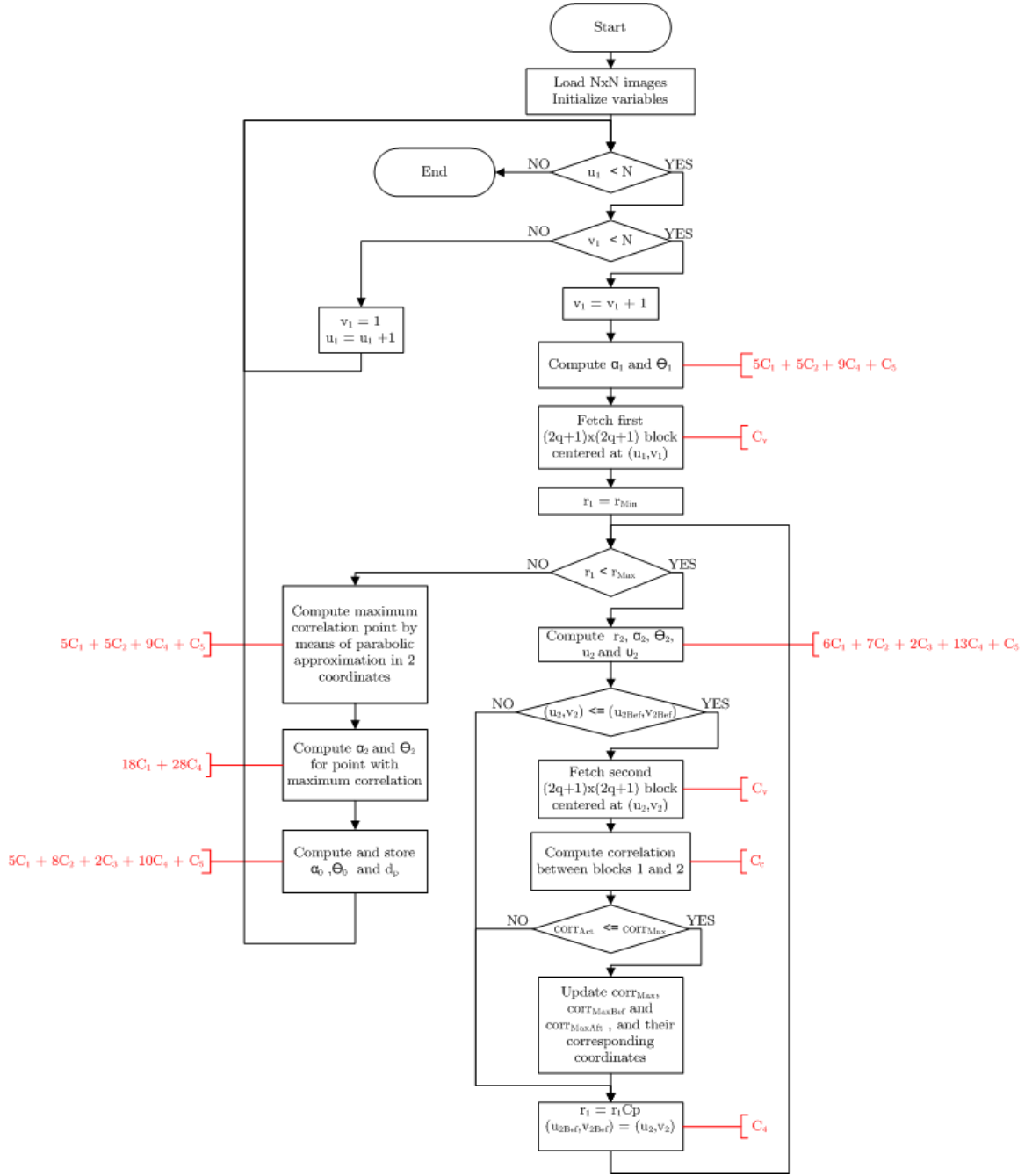- C_c: $(2q + 1) \times (2q + 1)$ block correlation cost

FIGURE A.1. Flow diagram and computational cost for the distance estimation process under the horizontal arrangement
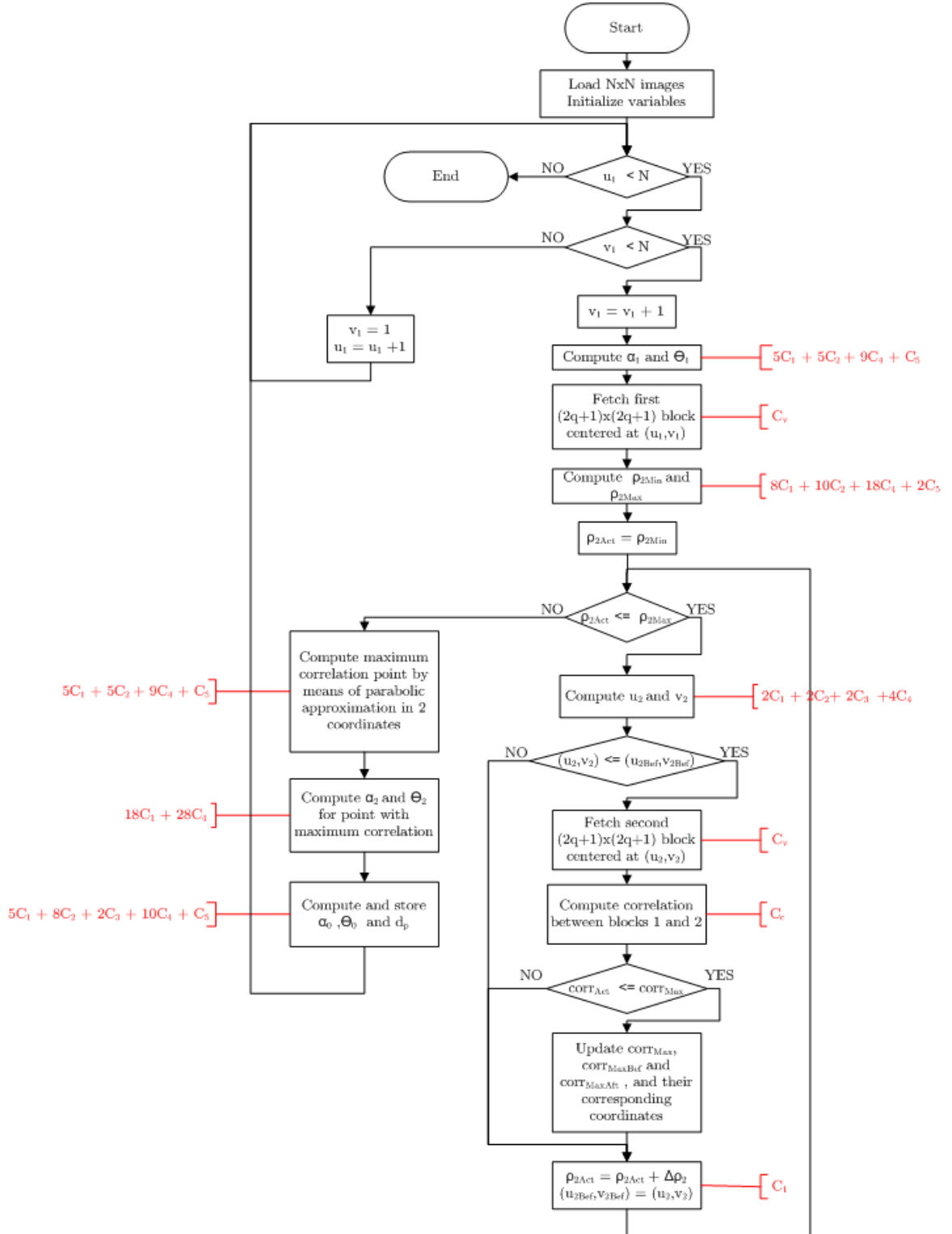
FIGURE A.2. Flow diagram and computational cost for the distance estimation process under the vertical arrangement
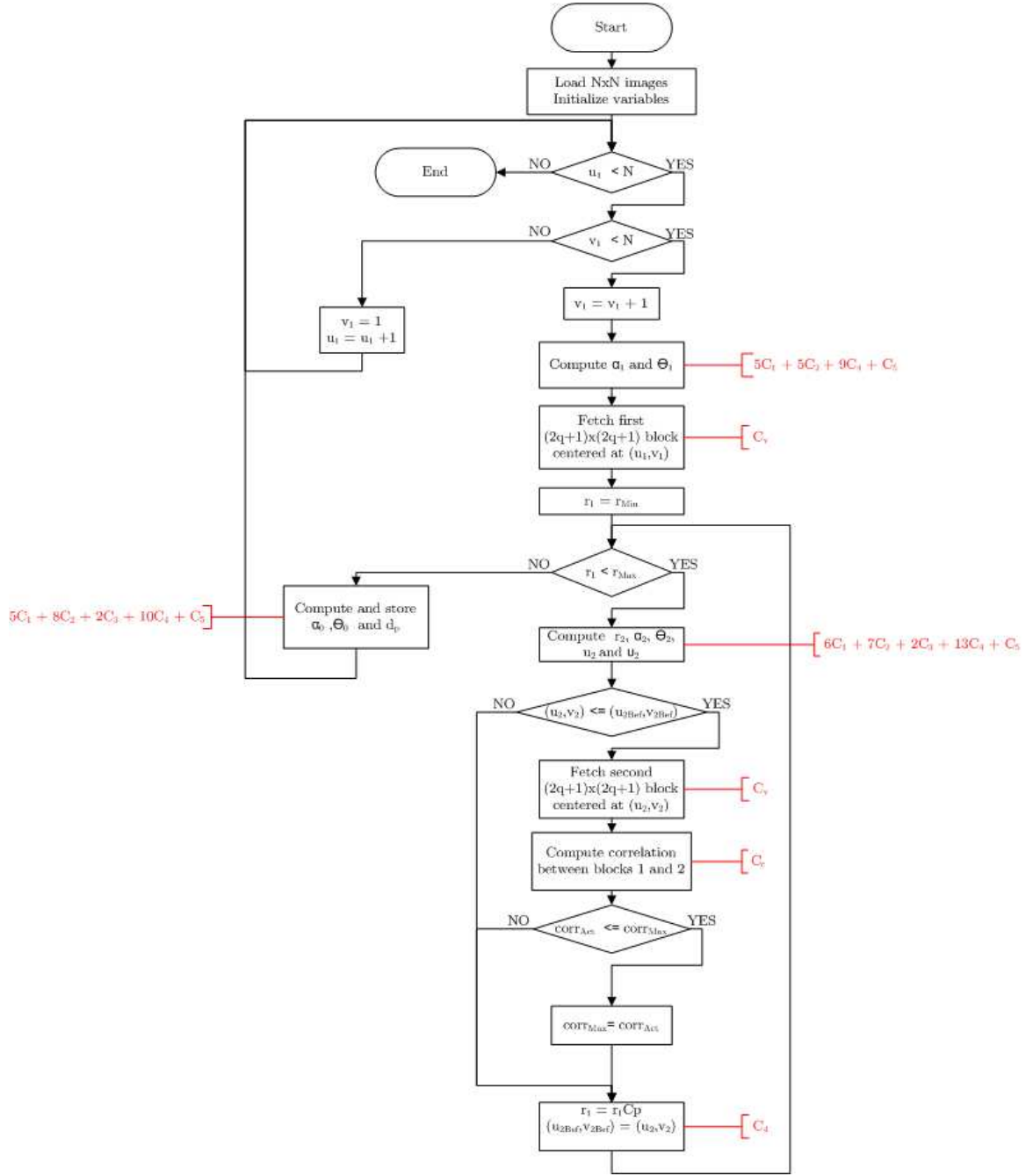
Start

Load NxN images
Initialize variables

End ← NO $u_1 < N$ YES

NO $v_1 < N$ YES

$v_1 = 1$
$u_1 = u_1 + 1$

$v_1 = v_1 + 1$

Compute α_1 and Θ_1 — $5C_1 + 5C_2 + 9C_4 + C_5$

Fetch first
$(2q+1)\times(2q+1)$ block
centered at (u_1,v_1) — C_v

$r_1 = r_{Min}$

NO $r_1 < r_{Max}$ YES

$5C_1 + 8C_2 + 2C_3 + 10C_4 + C_5$ —
Compute and store
α_0, Θ_0 and d_p

Compute r_2, α_2, Θ_2,
u_2 and v_2 — $6C_1 + 7C_2 + 2C_3 + 13C_4 + C_5$

NO $(u_2,v_2) <= (u_{2Bef},v_{2Bef})$ YES

Fetch second
$(2q+1)\times(2q+1)$ block
centered at (u_2,v_2) — C_v

Compute correlation
between blocks 1 and 2 — C_v

NO $corr_{Act} <= corr_{Max}$ YES

$corr_{Max} = corr_{Act}$

$r_1 = r_1 Cp$
$(u_{2Bef},v_{2Bef}) = (u_2,v_2)$ — C_4

FIGURE A.3. Flow diagram and computational cost for the distance estimation process under the horizontal arrangement
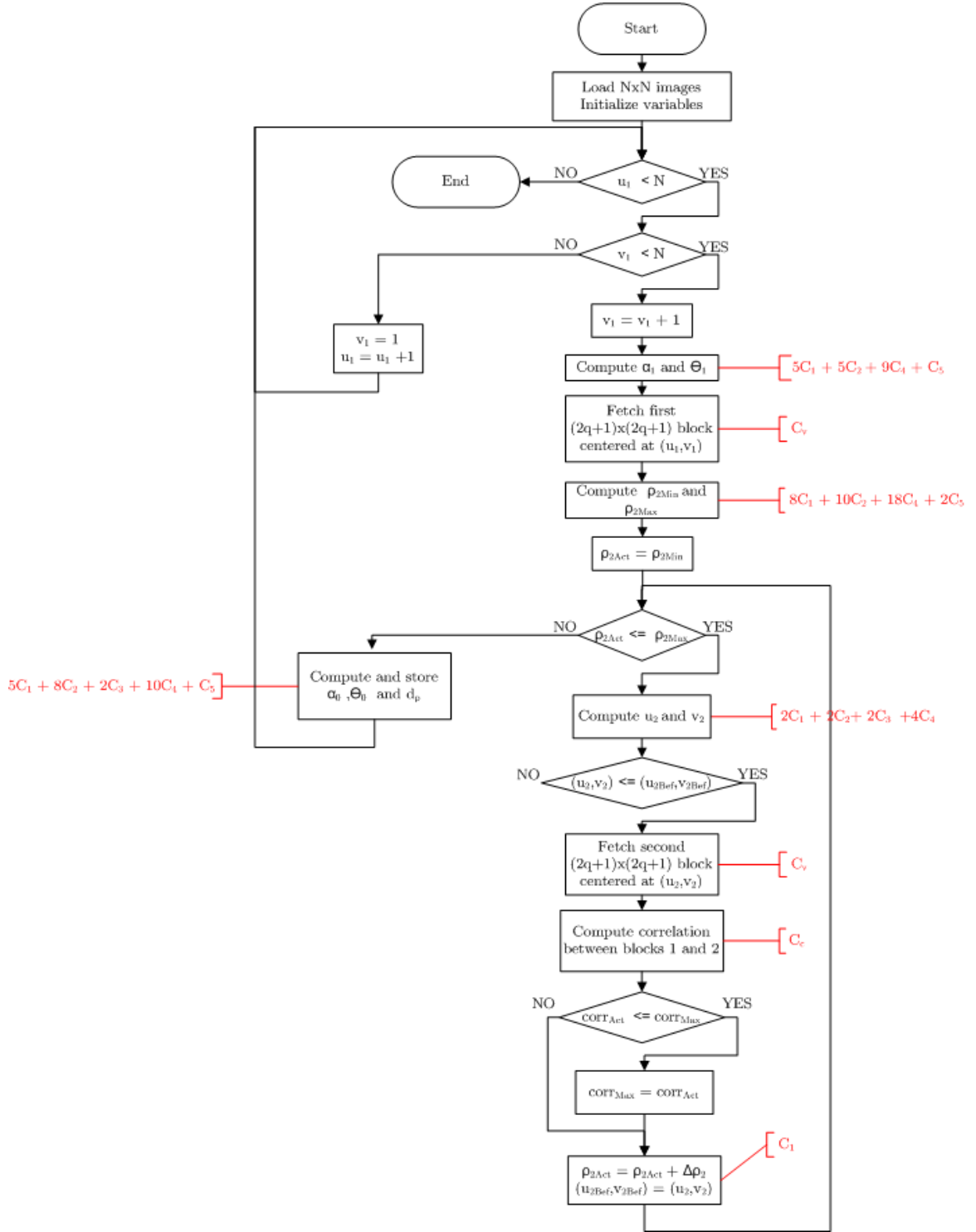
FIGURE A.4. Flow diagram and computational cost for the distance estimation process under the vertical arrangement

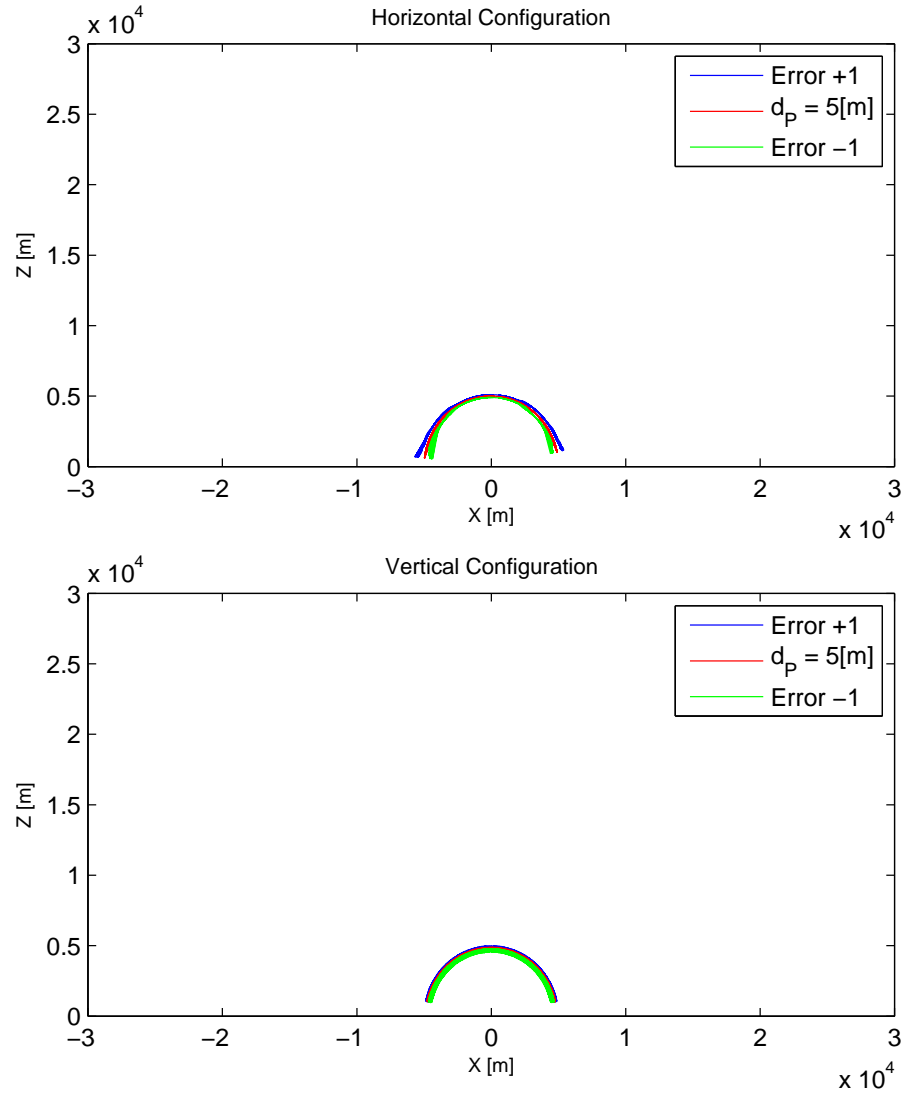# APPENDIX B. WORKING DISTANCE CIRCLE AND STIMATION ERROR MARGINS

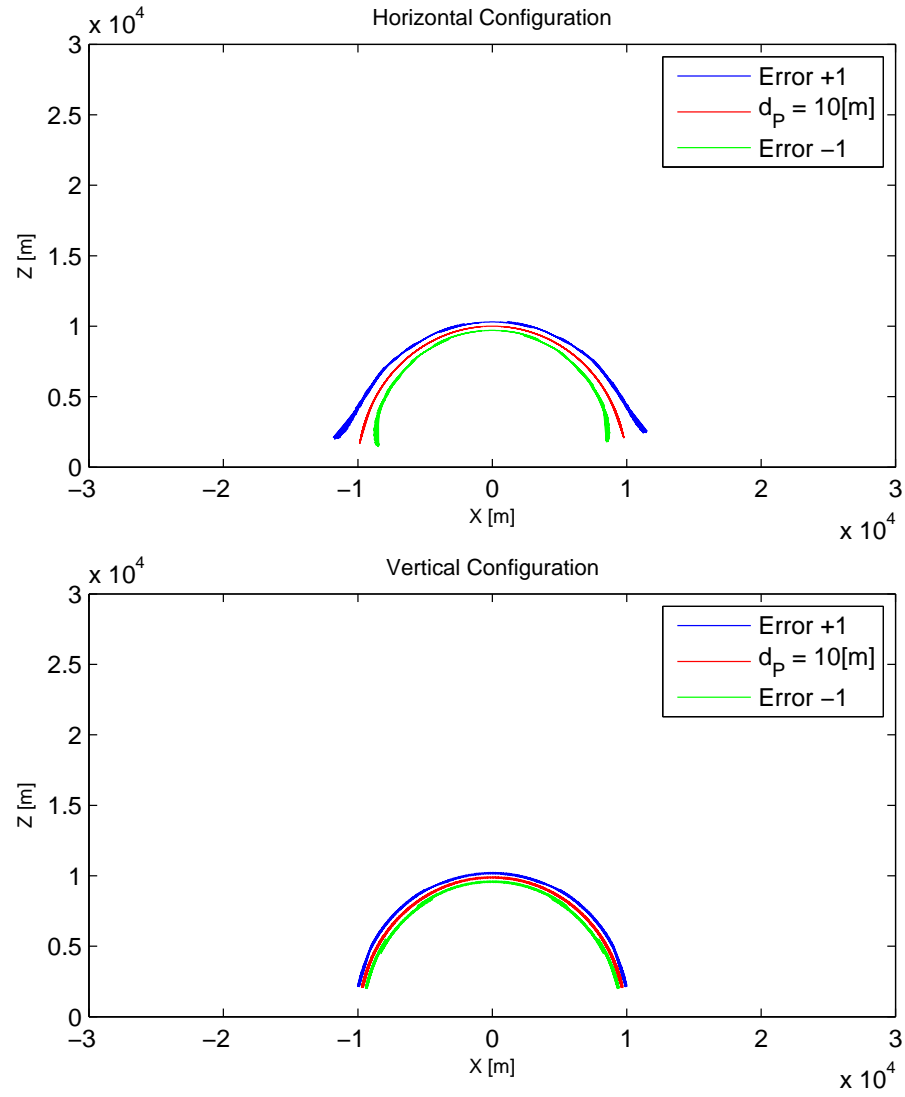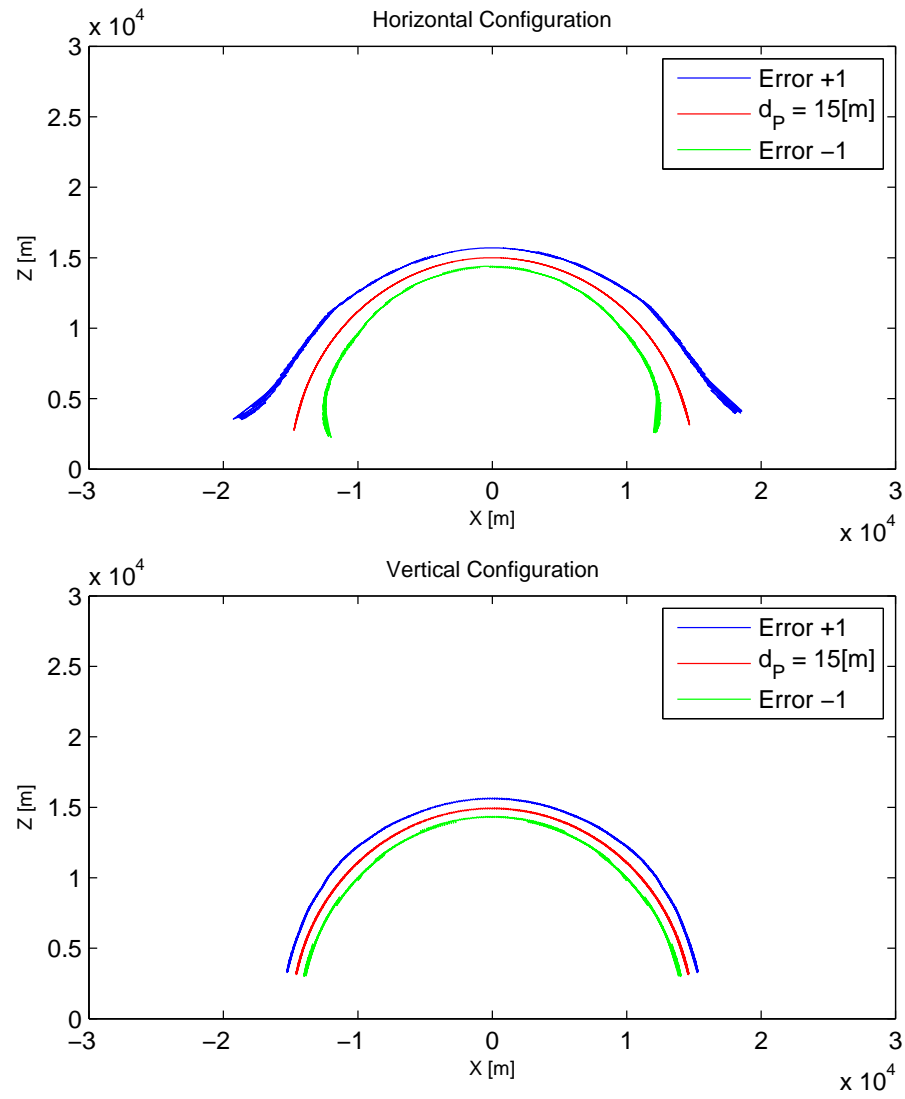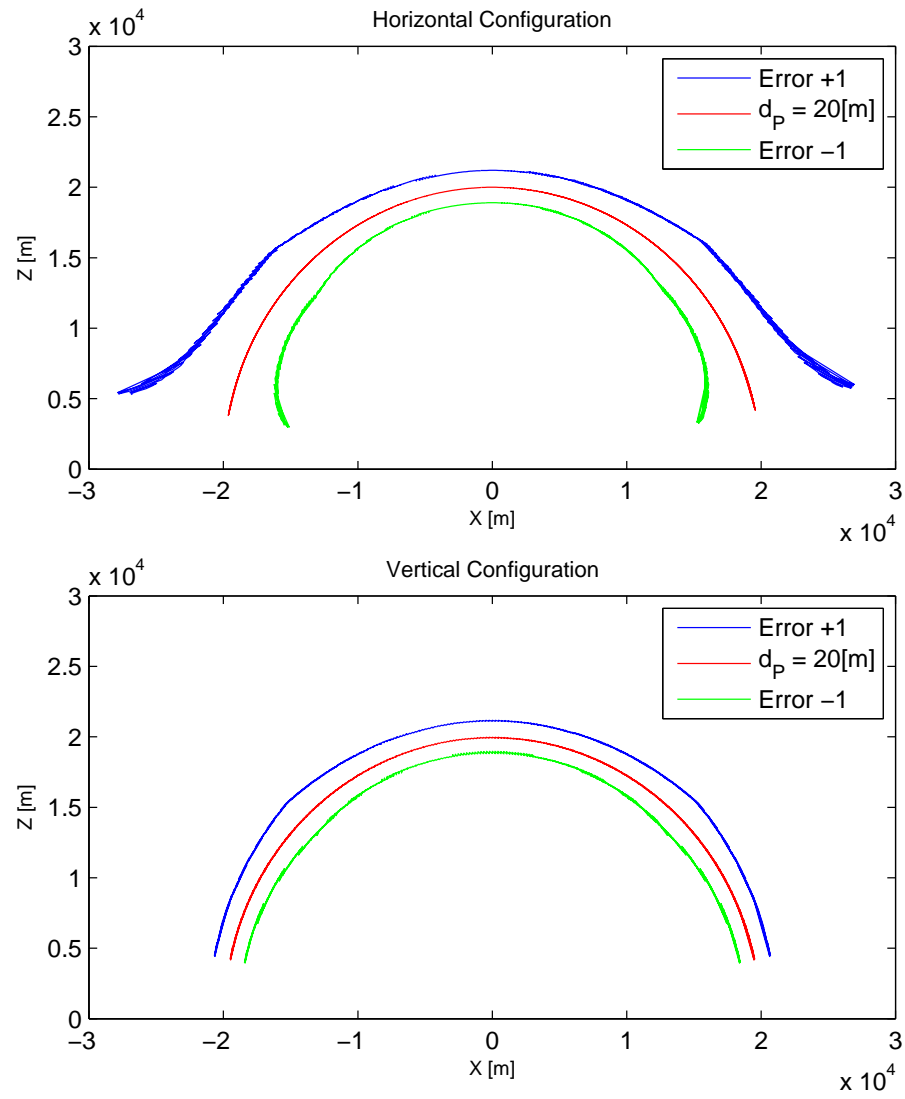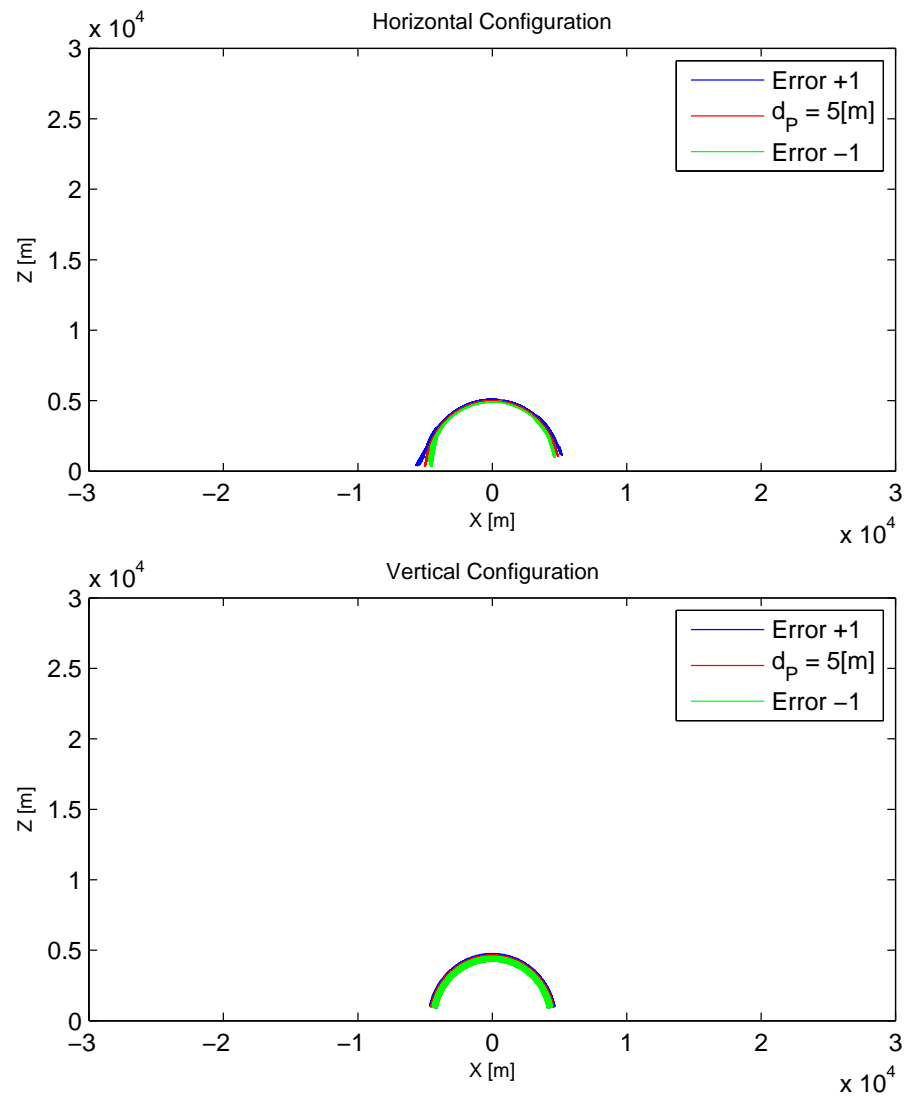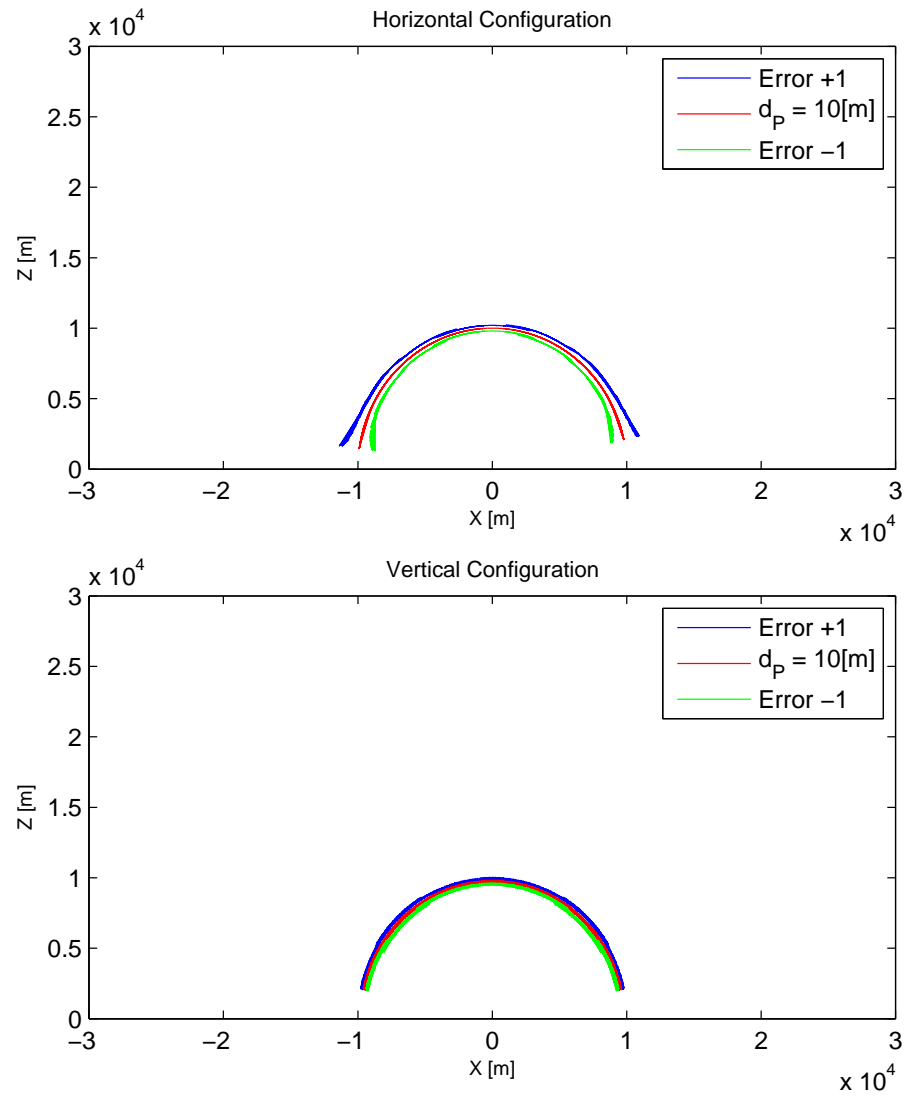FIGURE B.1. Working distance circle and estimation error margins for $d_c = 1\,m$ and $d_P = 5\,m$

FIGURE B.2. Working distance circle and estimation error margins for $d_c = 1\,m$ and $d_P = 10\,m$
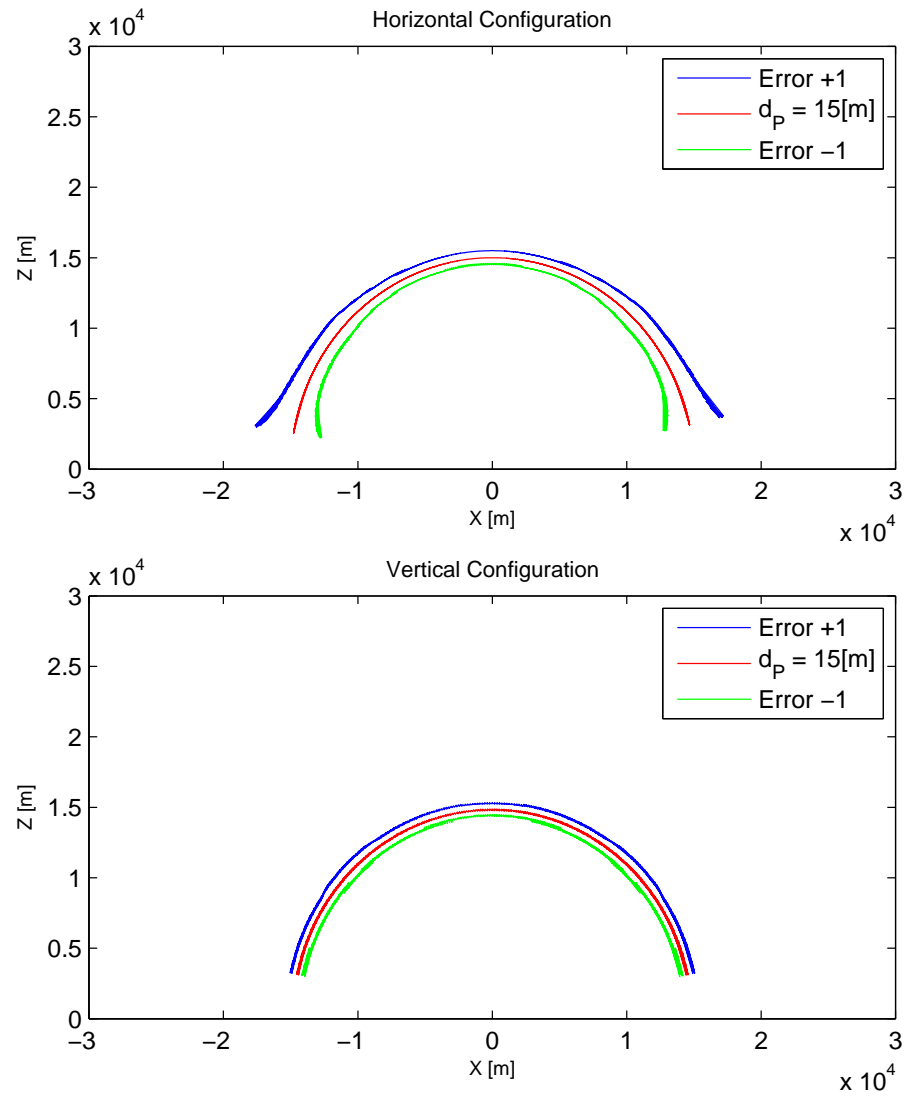
FIGURE B.3. Working distance circle and estimation error margins for $d_c = 1\,m$ and $d_P = 15\,m$

FIGURE B.4. Working distance circle and estimation error margins for $d_c = 1\,m$ and $d_P = 20\,m$

FIGURE B.5. Working distance circle and estimation error margins for $d_c = 1.5\,m$ and $d_P = 5\,m$

FIGURE B.6. Working distance circle and estimation error margins for $d_c = 1.5\,m$ and $d_P = 10\,m$

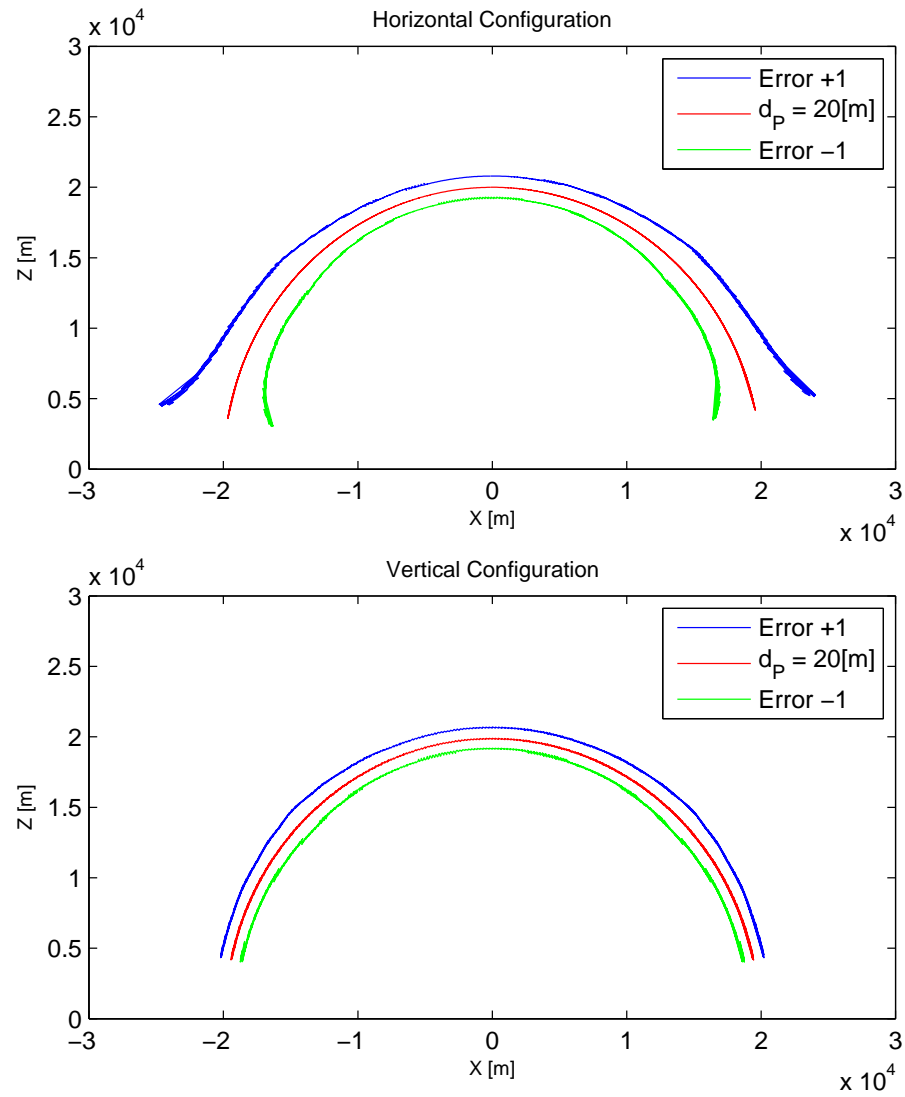FIGURE B.7. Working distance circle and estimation error margins for $d_c = 1.5\,m$ and $d_P = 15\,m$

FIGURE B.8. Working distance circle and estimation error margins for $d_c = 1.5\,m$ and $d_P = 20\,m$