A General Framework for the Selection of World Coordinate Systems in Perspective and Catadioptric Imaging Applications

JOÃO P. BARRETO AND HELDER ARAUJO
Institute of Systems and Robotics, Department of Electrical and Computer Engineering,
University of Coimbra, Coimbra, Portugal

Abstract. An imaging system with a single effective viewpoint is called a central projection system. The conventional perspective camera is an example of central projection system. A catadioptric realization of omnidirectional vision combines reflective surfaces with lenses. Catadioptric systems with an unique projection center are also examples of central projection systems. Whenever an image is acquired, points in 3D space are mapped into points in the 2D image plane. The image formation process represents a transformation from $\Re^3$ to $\Re^2$, and mathematical models can be used to describe it. This paper discusses the definition of world coordinate systems that simplify the modeling of general central projection imaging. We show that an adequate choice of the world coordinate reference system can be highly advantageous. Such a choice does not imply that new information will be available in the images. Instead the geometric transformations will be represented in a common and more compact framework, while simultaneously enabling newer insights. The first part of the paper focuses on static imaging systems that include both perspective cameras and catadioptric systems. A systematic approach to select the world reference frame is presented. In particular we derive coordinate systems that satisfy two differential constraints (the "compactness" and the "decoupling" constraints). These coordinate systems have several advantages for the representation of the transformations between the 3D world and the image plane. The second part of the paper applies the derived mathematical framework to active tracking of moving targets. In applications of visual control of motion the relationship between motion in the scene and image motion must be established. In the case of active tracking of moving targets these relationships become more complex due to camera motion. Suitable world coordinate reference systems are defined for three distinct situations: perspective camera with planar translation motion, perspective camera with pan and tilt rotation motion, and catadioptric imaging system rotating around an axis going through the effective viewpoint and the camera center. Position and velocity equations relating image motion, camera motion and target 3D motion are derived and discussed. Control laws to perform active tracking of moving targets using visual information are established.

Keywords: sensor modeling, catadioptric, omnidirectional vision, visual servoing, active tracking

1. Introduction

Many applications in computer vision, such as surveillance and model acquisition for virtual reality, require that a large field of view is imaged. Visual control of motion can also benefit from enhanced fields of view. The computation of camera motion from a sequence of images obtained with a traditional camera suffers from the problem that the direction of translation may lie outside the field of view. Panoramic imaging overcomes this problem making the uncertainty of camera motion estimation independent of the motion direction (Gluckman and Nayar, 1998). In position based visual servoing keeping the target in the field of view during motion raises severe difficulties (Malis et al., 1999). With a large field of view this problem no longer exists.

One effective way to enhance the field of view of a camera is to use mirrors (Bogner, 1995; Nalwa, 1996; Yagi and Kawato, 1990; Yamazawa et al., 1993, 1995). The general approach of combining mirrors with conventional imaging systems is referred to as catadioptric image formation (Hecht and Zajac, 1974).

The fixed viewpoint constraint is a requirement ensuring that the visual sensor only measures the intensity of light passing through a single point in 3D space (the projection center). Vision systems verifying the fixed viewpoint constraint are called central projection systems. Central projection systems present interesting geometric properties. A single effective viewpoint is a necessary condition for the generation of geometrically correct perspective images (Baker and Nayar, 1998), and for the existence of epipolar geometry inherent to the moving sensor and independent of the scene structure (Svoboda et al., 1998). It is highly desirable for any vision system to have a single viewpoint. The conventional perspective CCD camera is widely used in computer vision applications. In general it is described by a central projection model with a single effective viewpoint. Central projection cameras are specializations of the general projective camera that can be modeled by a 3×4 matrix with rank 3 (Hartley and Zisserman, 2000). In Baker and Nayar (1998), Baker and Nayar derive the entire class of catadioptric systems with a single effective viewpoint. Systems built using a parabolic mirror with an orthographic camera, or an hyperbolic, elliptical or planar mirror with a perspective camera verify the fixed viewpoint constraint.

In Geyer and Daniilidis (2000), introduce an unifying theory for all central catadioptric systems where conventional perspective imaging appears as a particular case. They show that central panoramic projection is isomorphic to a projective mapping from the sphere to a plane with a projection center on the perpendicular to the plane. A modified version of this unifying model is introduced in the paper (Barreto and Araujo, 2001).

General central projection image formation can be represented by a transformation from $\Re^3$ to $\Re^2$. Whenever an image is acquired, points in 3D space are mapped into points in the 2D image plane. Cartesian coordinate systems are typically used to reference points both in space and in the image plane. The mapping is non-injective and implies loss of information. The relationship between position and velocity in the 3D space and position and velocity in the image are in general complex, difficult and non-linear. This paper shows that

the choice of the coordinate system to reference points in the 3D space is important. The intrinsic nature of image formation process is kept unchanged but the mathematical relationship between the world and the image becomes simpler and more intuitive. This can help not only the understanding of the imaging process but also the development of new algorithms and applications.

The first part of the paper focuses on static imaging systems that include both perspective cameras and catadioptric systems. A general framework to describe the mapping from 3D points to 2D points in the image plane is presented. The mathematical expression of this global mapping depends on the coordinate system used to reference points in the scene. A systematic approach to select the world coordinate system is presented and discussed. Differential constraints are defined to enable the choice of a 3D reference frame. Coordinate transformations satisfying these differential constraints bring advantageous properties when mapping 3D space velocities into 2D image velocities. One such coordinate transformation is described for the case of the perspective camera and then generalized for central catadioptric image formation. Using these coordinate transformations does not imply that new information is available in the images. Instead the geometric transformations are represented in a common and more compact framework, while simultaneously enabling newer insights into the image formation process. Examples and applications that benefit from an adequate choice of the world coordinate system are presented and discussed.

The second part of our article applies the derived mathematical framework to active tracking of moving targets. For this purpose it is assumed that the imaging sensor is mounted on a moving platform. Three different cases are considered: a perspective camera with translational motion in the XY plane, a perspective camera with rotational pan and tilt motion and a parabolic omnidirectional camera with a rotational degree of freedom around the Z axis. The goal of the tracking application is to control the motion of the platform in such a way that the position of the target in the image plane is kept constant.

In the classical eye-in-hand positioning problem the camera is typically attached to the end effector of a 6 d.o.f. manipulator. The platforms considered in this work have less than 3 d.o.f. For the purpose of controlling the constrained 3D motion of these robots it is not necessary to determine the full pose of the target. It is assumed that target motion is characterized by the 3D

position and velocity of the corresponding mass center in an inertial reference frame. It is also assumed that the position of each degree of freedom is known (possibly via an encoder).

In active tracking applications the image motion depends both on target and camera 3D motion. The derived general framework to describe the mapping between 3D points and points in the 2D image plane is extended to central catadioptric imaging systems with rigid motion. The mathematical expression of the global mapping depends on the world coordinates used to reference points in the scene. General criteria to select suitable coordinate systems are discussed. Adequate choices are presented for each type of platform. The derived mathematical framework is used to establish the position and velocity relationships between target 3D motion, camera motion and image motion. The expressions obtained are used to implement image based active visual tracking. Simplifications of the equations obtained (to decouple the degrees of freedom of the pan and tilt vision system) are discussed.

2. Static Imaging Systems

This section focuses on the static central projection vision systems. Examples of such systems are the perspective camera and catadioptric systems that verify the fixed viewpoint constraint (Baker and Nayar, 1998). The image acquisition process maps points from the 3D space into the 2D image plane. Image formation performs a transformation from $\Re^3$ to $\Re^2$ that can be denoted by $\mathbf{F}$. A generic framework to illustrate the transformation $\mathbf{F}$ is proposed. This framework is general to both conventional perspective cameras and central projection catadioptric systems. It is desirable that $\mathbf{F}$ be as simple as possible and as compact as possible. This can be achieved by selecting a specific coordinate systems to reference the world points. General criteria to select the world coordinate system are presented and discussed. Advantages of using different world coordinate systems to change the format of the $\mathbf{F}$ mapping are presented.

2.1. Mapping Points from the 3D Space in the 2D Image Plane

Figure 1 depicts a generic framework to illustrate the transformation $\mathbf{F}$ from $\Re^3$ in $\Re^2$ performed by a central projection vision system. If the vision system has a
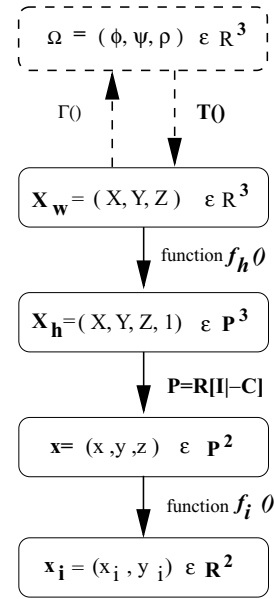

Figure 1. Schematic of the mapping performed by general central projection imaging systems.

unique viewpoint, it preserves central projection and geometrically correct perspective images can always be generated (Gluckman and Nayar, 1998).

$\mathbf{X_w} = (X, Y, Z)^t$ is a vector with the Cartesian 3D coordinates of a point in space. The domain of transformation is the set D of visible points in the world with $D \subset \Re^3$. Function $\mathbf{f_h}$ maps $\Re^3$ into the projective space $\wp^3$. It is a non-injective and surjective function transforming $\mathbf{X_w} = (X, Y, Z)^t$ in $\mathbf{X_h} = (X, Y, Z, 1)^t$ that are the homogeneous world point coordinates. $\mathbf{P}$ is an arbitrary 3×4 homogeneous matrix with rank 3. It represents a general projective transformation performing a linear mapping of $\wp^3$ into the projective plane $\wp^2$ ($\mathbf{x} = \mathbf{PX_h}$). The rank 3 requirement is due to the fact that if the rank is less than 3 then the range of the matrix will be a line or a point and not the whole plane. The rank 3 requirement guarantees that the transformation is surjective. In the case of $\mathbf{P}$ being a camera model it can be written as $\mathbf{P} = \mathbf{KR}[\mathbf{I}| - \tilde{\mathbf{C}}]$ where $\mathbf{I}$ is a 3×3 identity matrix, $\mathbf{K}$ is the intrinsic parameters matrix, $\mathbf{R}$ the rotation matrix between camera and world coordinate systems and $\tilde{\mathbf{C}}$ the projection center in world coordinates (Hartley and Zisserman, 2000). If nothing is stated we will assume $\mathbf{K} = \mathbf{I}$ and standard central projection with $\mathbf{P} = [\mathbf{I} \,|\, \mathbf{0}]$. Function $\mathbf{f_i}$ transforms coordinates in the projective plane $\mathbf{x} = (x, y, z)^t$ into Cartesian coordinates in the image plane $\mathbf{x_i} = (x_i, y_i)^t$. It is a non-injective, surjective

function of $\wp^2$ in $\Re^2$ that maps projective rays in the world into points in the image. For conventional perspective cameras $\mathbf{x_i} = \mathbf{f_i(x)} \Leftrightarrow (x_i, y_i) = (\frac{x}{z}, \frac{y}{z})$. However, as it will be shown later, these relations are more complex for generic catadioptric systems.

The transformation $\mathbf{F}$ maps 3D world points into 2D points in the image. Points in the scene were represented using standard cartesian coordinates. However a different coordinate system can be used to reference points in the 3D world space. Assume that $\mathbf{\Omega} = (\phi, \psi, \rho)^t$ are point coordinates in the new reference frame and that $\mathbf{X_w} = \mathbf{T(\Omega)}$ where $\mathbf{T}$ is a bijective function from $\Re^3$ in $\Re^3$. The transformation $\mathbf{F}$, mapping 3D world points $\mathbf{\Omega}$ in image points $\mathbf{x_i}$ (see Eq. (1)), can be written as the composition of Eq. (2).

$$\mathbf{x_i} = \mathbf{F(\Omega)} \qquad (1)$$

$$\mathbf{F(\Omega)} = \mathbf{f_i(Pf_h(T(\Omega)))} \qquad (2)$$

Equation (3), obtained by differentiating Eq. (1) with respect to time, establishes the relationship between velocity in 3D space $\dot{\mathbf{\Omega}} = (\dot{\phi}, \dot{\psi}, \dot{\rho})^t$ and velocity in image $\dot{\mathbf{x_i}} = (\dot{x_i}, \dot{y_i})^t$. $\dot{\mathbf{x_i}}$ and $\dot{\mathbf{\Omega}}$ are related by the jacobian matrix $\mathbf{J_F}$ of transformation $\mathbf{F}$. Equation (4) shows $\mathbf{J_F}$ as the product of the Jacobians of the transformations that make up $\mathbf{F}$.

$$\dot{\mathbf{x_i}} = \mathbf{J_F}\dot{\mathbf{\Omega}} \qquad (3)$$

$$\mathbf{J_F} = \mathbf{J_{f_i}} \cdot \mathbf{J_P} \cdot \mathbf{J_{f_h}} \cdot \mathbf{J_T} \qquad (4)$$

Function $\mathbf{T}$ represents a change of coordinates. It must be bijective which guarantees that it admits an inverse. Assume that $\mathbf{\Gamma}$ is the inverse function of $\mathbf{T}$ ($\mathbf{\Gamma} = \mathbf{T^{-1}}$). Function $\mathbf{\Gamma}$, from $\Re^3$ into $\Re^3$, transforms cartesian coordinates $\mathbf{X_w}$ in new coordinates $\mathbf{\Omega}$ (Eq. (5)). $\mathbf{J_\Gamma}$ is the jacobian matrix of $\mathbf{\Gamma}$ (Eq. (6)). If $\mathbf{T}$ is injective then the jacobian matrix $\mathbf{J_T}$ is non-singular with inverse $\mathbf{J_\Gamma}$. Replacing $\mathbf{J_T}$ by $\mathbf{J_\Gamma^{-1}}$ in Eq. (4) yields Eq. (7) showing the jacobian matrix of $\mathbf{F}$ expressed in terms of the scalar function of $\mathbf{\Gamma}$ and its partial derivatives.

$$\mathbf{\Gamma(X_w)} = (\phi(X, Y, Z), \psi(X, Y, Z), \rho(X, Y, Z))^t \quad (5)$$

$$\mathbf{J_\Gamma} = \begin{bmatrix} \phi_X & \phi_Y & \phi_Z \\ \psi_X & \psi_Y & \psi_Z \\ \rho_X & \rho_Y & \rho_Z \end{bmatrix} \qquad (6)$$

$$\mathbf{J_F} = \mathbf{J_{f_i}} \cdot \mathbf{J_P} \cdot \mathbf{J_{f_h}} \cdot \mathbf{J_\Gamma^{-1}} \qquad (7)$$

2.2.   *Criteria to Select the World Coordinate System*

Function $\mathbf{F}$ is a transformation from $\Re^3$ (3D world space) into $\Re^2$ (image plane). In Eqs. (8) and (9) $\mathbf{F}$ and $\mathbf{J_F}$ are written in terms of scalar functions and their partial derivatives. The relationship between world and image points can be complex and counter intuitive. The mathematical expression of the mapping function $\mathbf{F}$ depends on the transformation $\mathbf{T}$ (see Eqs. (2), (4) and (7)). The selection of a certain coordinate system to reference points in the scene changes the way $\mathbf{F}$ is written but does not change the intrinsic nature of the mapping. However, with an adequate choice of the world coordinate system, the mathematical relationship between position and velocity in space and position and velocity in the image plane can become simpler, more intuitive or simply more suitable for a specific application. In this section we discuss criteria for the selection of the world coordinate system.

$$\mathbf{F(\Omega)} = (h(\phi, \psi, \rho), g(\phi, \psi, \rho))^t \qquad (8)$$

$$\mathbf{J_F} = \begin{bmatrix} h_\phi & h_\psi & h_\rho \\ g_\phi & g_\psi & g_\rho \end{bmatrix} \qquad (9)$$

2.2.1. *The Compactness Constraint.*   Consider central projection vision systems as mappings of 3D points, expressed in Cartesian coordinates $\mathbf{X_w} = (X, Y, Z)^t$, into the 2D image coordinates $\mathbf{x_i} = (x_i, y_i)$. The transformation is a function from $\Re^3$ into $\Re^2$ with loss of information (depth). In general the two coordinates in the image plane depend on the three coordinates in space. The image gives partial information about each one of the three world coordinates but we are not able to recover any of those parameters without further constraints. The imaging process implies loss of information and there is no additional transformation $\mathbf{T}$ that can change that. However it would be advantageous that image coordinates depend only on two of the 3D parameters. In many situations that can be achieved by means of a change of coordinates $\mathbf{T}$. The coordinate change must be performed in such a way that $\mathbf{F}$ only depends on two of those coordinates. Assuming that $\mathbf{\Omega} = (\phi, \psi, \rho)$ are the new 3D coordinates, $\mathbf{F}$ becomes a function of only ϕ and ψ which means that partial derivatives h_ρ and g_ρ are both equal to zero. Whenever a certain change of coordinates $\mathbf{T}$ leads to a jacobian matrix $\mathbf{J_F}$ with a zero column, it is said that mapping $\mathbf{F}$ is in a compact form and coordinate transformation $\mathbf{T}$ verifies the "compactness constraint".

Assume that a world coordinate system satisfying the "compactness constraint" is selected. If Eq. (10) is verified then the image coordinates (x_i, y_i) depend only on (ϕ, ψ) and $\mathbf{F}$ becomes a function from $\Re^2$ in $\Re^2$ ($\mathbf{x_i} = \mathbf{F}(\Omega_\mathbf{c})$ with $\Omega_\mathbf{c} = (\phi, \psi)^t$). A function from $\Re^3$ into $\Re^2$ is never invertible, thus putting $\mathbf{F}$ in a compact form is a necessary condition to find out an inverse mapping $\mathbf{F^{-1}}$. If $\mathbf{F^{-1}}$ exists then two of the three 3D parameters of motion can be recovered from image ($\Omega_\mathbf{c} = \mathbf{F^{-1}}(\mathbf{x_i})$) and the jacobian matrix $\mathbf{J_F}$ can be written in term of image coordinates x_i and y_i. By verifying the "compactness constraint" the relationships in position and velocity between the 3D world and the image plane tend to be more compact and intuitive and vision yields all the information about two of the 3D world coordinates and none about the third one.

$$h_\rho = 0 \wedge g_\rho = 0 \qquad (10)$$

2.2.2. The Decoupling Constraint.   Assume that the "compactness constraint" is verified. This means that a coordinate transformation $\mathbf{T}$ is used such that image coordinates (x_i, y_i) depend only on (ϕ, ψ). It would be also advantageous to define a world coordinate system such that x_i depends only of ϕ and y_i depends only of ψ. This is equivalent to say that h_ψ and g_ϕ are both zero. The one to one correspondence is an advantageous feature allowing a better understanding of the imaging process and simplifying subsequent calculations. If a coordinate transformation $\mathbf{T}$ is used such that both Eqs. (10) and (11) are verified then it is said that $\mathbf{F}$ is in a compact and decoupled form and that $\mathbf{T}$ verifies both the "compactness constraint" and the "decoupling constraint".

$$h_\psi = 0 \wedge g_\phi = 0 \qquad (11)$$

In short, given a general central projection mapping, the goal is to select a coordinate transformation $\mathbf{T}$ verifying both:

- the "compactness constraint" (Eq. (10))
- the "decoupling constraint" (Eq. (11))

The coordinate system used to reference points in the scene does not change the intrinsic nature of the mapping nor introduces any additional information. There are situations where it is impossible to find a world coordinates transformation that verifies the "compactness constraint" and/or the "decoupling constraint." Methodologies to find out if it exists such a transformation will be introduced latter.

2.3.   Conventional Perspective Camera

Consider image acquisition performed by a static conventional perspective camera. The image formation process follows the scheme depicted in Fig. 1 where function $\mathbf{f_i}$ is given by Eq. (12). Assume that the matrix of intrinsic parameters is $\mathbf{K} = \mathbf{I}$ and $\mathbf{P} = [\mathbf{I} \,|\, \mathbf{0}]$ (the origin of the cartesian reference frame is coincident with the camera center and the image plane is perpendicular to the Z axis). This section derives a world coordinate system that verifies both the compactness and decoupling constraint. If nothing is stated we will work with the inverse transformation Γ instead of the direct transformation $\mathbf{T}$.

$$\mathbf{f_i}() : (x, y, z) \longrightarrow \left(\frac{x}{z}, \frac{y}{z} \right) \qquad (12)$$

2.3.1. Constraining Γ to Obtain a New World Coordinate System.   Functions $\mathbf{f_i}$, $\mathbf{P}$ and $\mathbf{f_h}$, as well as their jacobian matrices, are defined for the perspective camera case. Replacing $\mathbf{J_\Gamma}$ (Eq. (6)) in Eq. (7) yields $\mathbf{J_F}$ in terms of the partial derivatives of the scalar functions of Γ (the computation is omitted). If $\mathbf{F}$ is in a compact form then the third column of $\mathbf{J_F}$ must be zero (Eq. (10)) which leads to Eqs. (13). A transformation of coordinates Γ that verifies the compactness constraint can be computed by solving the partial differential Eqs. (13) with respect to the scalar functions ϕ, ψ and ρ (Eq. (5)).

$$\begin{cases} Z(\phi_Y \psi_Z - \phi_Z \psi_Y) + X(\phi_Y \psi_X - \phi_X \psi_Y) = 0 \\ Z(\phi_Z \psi_X - \phi_X \psi_Z) + Y(\phi_Y \psi_X - \phi_X \psi_Y) = 0 \end{cases} \qquad (13)$$

The partial differential equations corresponding to the "decoupling constraint" can be derived in a similar way. If the mapping $\mathbf{F}$ is decoupled then both h_ψ and g_ϕ must be zero, which leads to Eq. (14). A world coordinate transformation Γ verifying both the compactness and the decoupling constraint can be computed by solving simultaneously Eqs. (13) and (14). Nevertheless the integration of systems of partial differential equations can be difficult and in general it generates many solutions. Adequate coordinate systems will be derived by geometrical means. Equations (13) and (14) will be used to prove that the selected coordinate transformation verifies the compactness and/or decoupling constraints.

$$\begin{cases} Z(\phi_Z \rho_Y - \phi_Y \rho_Z) + X(\phi_Y \rho_X - \phi_X \rho_Y) = 0 \\ Z(\psi_Z \rho_X - \psi_X \rho_Z) + Y(\psi_Y \rho_X - \psi_X \rho_Y) = 0 \end{cases} \qquad (14)$$
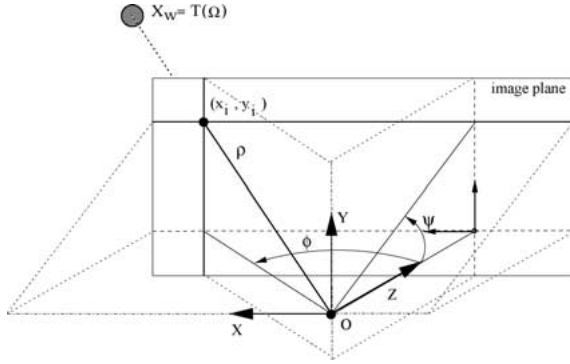
Figure 2.   Conventional perspective camera case.

Table 1.   Using a new coordinate system for perspective camera.

$$\mathbf{\Gamma}(\mathbf{X_w}) = \left(\arctan\left(\frac{X}{Z}\right),\ \arctan\left(-\frac{Y}{Z}\right),\ \sqrt{X^2 + Y^2 + Z^2}\right)^t$$

$$\mathbf{T}(\mathbf{\Omega}) = \frac{\rho}{\sqrt{1 + \tan(\phi)^2 + \tan(\psi)^2}}(\tan(\phi),\ -\tan(\psi),\ 1)^t$$

$$\mathbf{F}(\mathbf{\Omega}) = (\tan(\phi),\ -\tan(\psi))^t$$

$$\mathbf{J_F}(\mathbf{\Omega}) = \begin{bmatrix} \frac{1}{\cos(\phi)^2} & 0 & 0 \\ 0 & -\frac{1}{\cos(\psi)^2} & 0 \end{bmatrix}$$

Figure 2 is a representation of the image formation process for the specified perspective camera situation. A point in the scene with cartesian coordinates (X, Y, Z) is projected in the image plane at position (x_i, y_i). It is assumed that only the points in front of the camera are visible ($Z > 0$). The world points that can be projected in a certain image vertical line, lie in a vertical plane rotated around the camera referential Y axis, containing the center of projection. In a similar manner, the world points projected in an horizontal line of the image belong to an horizontal plane rotated around the X axis. Consider the pencil of the vertical planes where each one is indexed by the corresponding rotation angle ϕ. There is a one to one correspondence between the vertical lines in the image and the ϕ angles. In a similar manner, a pencil of horizontal planes, indexed by the corresponding rotation angle ψ, is defined. Each horizontal line in the image is associated to one ψ angle. Each pair (ϕ, ψ) defines a pair of planes that intersect in a projective ray. If the depth of the world point along the ray is ρ (always positive), the set of values (ϕ, ψ, ρ) gives the position of the world point in an unique way. We have derived a new system of coordinates to represent points in the 3D space. Equation (15) establishes the relationship between (ϕ, ψ, ρ) and the conventional Cartesian coordinates (X, Y, Z).

$$\begin{cases} \phi = \arctan\left(\dfrac{X}{Z}\right) \\[2mm] \psi = \arctan\left(-\dfrac{Y}{Z}\right) \\[2mm] \rho = \sqrt{X^2 + Y^2 + Z^2} \end{cases} \quad (15)$$

Equation (16) gives the jacobian matrix of the derived coordinate transformation $\mathbf{\Gamma}$. The proposed change of coordinates is a solution of the set of differ-

ential Eqs. (13) and (14). $\mathbf{\Gamma}$ satisfies both the compactness and decoupling constraint for the static perspective camera case.

$$\mathbf{J_\Gamma} = \begin{bmatrix} \frac{Z}{X^2+Z^2} & 0 & -\frac{X}{X^2+Z^2} \\[2mm] 0 & -\frac{Z}{Y^2+Z^2} & \frac{Y}{Y^2+Z^2} \\[2mm] \frac{X}{\sqrt{X^2+Y^2+Z^2}} & \frac{Y}{\sqrt{X^2+Y^2+Z^2}} & \frac{Z}{\sqrt{X^2+Y^2+Z^2}} \end{bmatrix} \quad (16)$$

Table 1 summarizes the results obtained in this section. $\mathbf{\Gamma}$ is a world coordinate transformation verifying both the compactness and decoupling constraint. Notice that the new coordinate system is different from the well known spherical coordinates. $\mathbf{T}$ is the inverse function of $\mathbf{\Gamma}$. Replacing $\mathbf{T}$ in Eq. (2) ($\mathbf{f_i}$, $\mathbf{P}$ and $\mathbf{f_h}$ were already defined) leads to the mathematical expression of global mapping $\mathbf{F}$ using the new coordinates. The jacobian matrix $\mathbf{J_F}$ is obtained replacing $\mathbf{J_\Gamma}$ in Eq. (7) by the result of Eq. (16).

2.3.2. Applications.   The global mapping $\mathbf{F}$ and its jacobian matrix establish the relationship between position/velocity in space and position/velocity in image for the conventional camera situation. Assume that both image position $\mathbf{x_i}$ and velocity $\dot{\mathbf{x}}_i$ are known. It was stated that there is a loss of information in the image formation process. Therefore it is not possible to fully recover 3D target motion from images without further information. Nevertheless, using a world coordinate system verifying the compactness and decoupling constraint, it is possible to partially recover the 3D parameters of motion in a straightforward manner.

Table 1 shows the mapping $\mathbf{F}$ and the corresponding jacobian matrix $\mathbf{J_F}$ written in terms of derived system of coordinates. Table 2 gives the mathematical expressions of $\mathbf{F}$ and $\mathbf{J_F}$ using world cartesian coordinates. In the former situation $\mathbf{F}$ is in a compact form and it is possible to invert the mapping to recover position information. One obtains that $\phi = \arctan(x_i)$ and $\psi = -\arctan(y_i)$. When using cartesian coordinates,

Table 2. Using a cartesian coordinate system for perspective camera.

$$\mathbf{F}(\mathbf{X_w}) = \left(\frac{X}{Z}, \frac{Y}{Z}\right)^t$$

$$\mathbf{J_F}(\mathbf{X_w}) = \begin{bmatrix} \frac{1}{Z} & 0 & -\frac{X}{Z^2} \\ 0 & \frac{1}{Z} & -\frac{Y}{Z^2} \end{bmatrix}$$

$\mathbf{F}$ appears as a function of $\Re^3$ in $\Re^2$, and the inversion is not possible.

Equation (17) shows $\mathbf{J_F}$ written in terms of image coordinates. It is derived by replacing in the jacobian matrix of Table 1 (ϕ, ψ) by $(\arctan(x_i), -\arctan(y_i))$. Knowing both position and velocity in the image one obtains $\dot{\phi} = (1 + x_i^2)^{-1}\dot{x}_i$ and $\dot{\psi} = -(1 + y_i^2)^{-1}\dot{y}_i$. Using the derived world coordinate system it is possible to partially recover the position and velocity of the target in the scene.

$$\mathbf{J_F}(\mathbf{x_i}) = \begin{bmatrix} 1 + x_i^2 & 0 & 0 \\ 0 & -(1 + y_i^2) & 0 \end{bmatrix} \quad (17)$$

2.4. Central Catadioptric Imaging System

A catadioptric realization of omnidirectional vision combines reflective surfaces and lenses. Central catadioptric imaging can be highly advantageous for many applications because it combines two important features: a single projection center and a wide field of view. The drawback of this type of sensors is that in general the mapping between points in the 3D world and in the image is highly non-linear. Many times correct perspective images are generated from frames captured by catadioptric sensors, and subsequently processed. We wish to study the advantages of working directly with the catadioptric images without warping them.

Geyer and Daniilidis (2000) introduce an unifying theory for central catadioptric systems. A modified version of this unifying model was presented in Barreto and Araujo (2001). The model is used to derive a general transformation of coordinates that leads to a mapping between points in the world and the in the image that verifies both the compactness and the decoupling constraints. Some applications are presented and discussed.

2.4.1. General Model for Central Projection Systems.
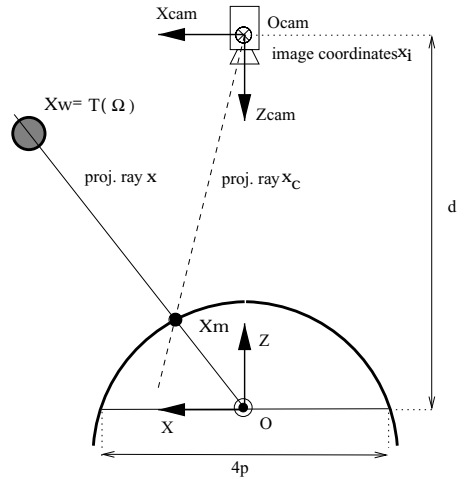In Baker and Nayar (1998) derive the entire class of catadioptric systems with a single effective viewpoint.

Figure 3. Central catadioptric image formation.

Figure 3 is a scheme of the catadioptric system combining an hyperbolic reflective surface with a perspective camera. The hyperbola is placed such that its axis is the Z-axis, its foci are coincident with $\mathbf{O}$ and $\mathbf{O_{cam}}$ (the origin of coordinate systems $\Re$ and $\Re_{\mathbf{cam}}$), its latus rectum is $4p$ and the distance between the foci is d. Light rays incident with $\mathbf{O}$ (the inner focal point) are reflected into rays incident with $\mathbf{O_{cam}}$ (the outer focal point). Assume a perspective camera with projection center in $\mathbf{O_{cam}}$ pointed to the mirror surface. All the captured light rays go originally through the inner focus of the hyperbolic surface. The effective viewpoint of the grabbed image is $\mathbf{O}$ and is unique. Elliptical catadioptric images are obtained combining an elliptical mirror with a perspective camera in a similar way. In the parabolic situation a parabolic mirror is placed such that its axis is the Z-axis, and its unique finite real focus is coincident with $\mathbf{O}$. Light rays incident with $\mathbf{O}$ are reflected into rays parallel with the Z-axis which are captured by an orthographic camera with image plane perpendicular to the Z-axis. The effective viewpoint is in $\mathbf{O}$ and is unique. A catadioptric system made up of a perspective camera steering a planar mirror also verifies the fixed viewpoint constraint. The effective projection center is behind the mirror in the perpendicular line passing through camera center. Its distance to the camera center is twice the distance between planar mirror and the camera.

Consider a generic scene point, visible by the catadioptric system, with cartesian coordinates $\mathbf{X_w}$ in the world reference frame. The corresponding homogeneous representation is $\mathbf{X_h}$. Visible points in the scene $\mathbf{X_h}$ are mapped into projective rays/points

Table 3. Mapping parameters *l* and *m*.

	l	*m*
Parabolic	1	$1 + 2p$
Hyperbolic	$\frac{d}{\sqrt{d^2+4p^2}}$	$\frac{d+2p}{\sqrt{d^2+4p^2}}$
Elliptical	$\frac{d}{\sqrt{d^2+4p^2}}$	$\frac{d-2p}{\sqrt{d^2+4p^2}}$
Planar	0	1

x in the catadioptric system reference frame centered in the effective viewpoint. The transformation is linear being described by a 3×4 matrix **P** (if nothing is stated it is assumed that $\mathbf{P} = [\mathbf{I} \mid \mathbf{0}]$). To each oriented projective ray/point **x**, corresponds a projective ray/point $\mathbf{x_{cam}}$ in a coordinate system whose origin is in the camera projection center. Notice that **x** and $\mathbf{x_{cam}}$ must intersect in the mirror surface (see Fig. 3). If $\mathbf{x_{cam}} = (x_{cam}, y_{cam}, z_{cam})^t$ and $\mathbf{K} = \mathbf{I}$ (camera intrinsic parameters) then point image coordinates are $(x_i, y_i) = (\frac{x_{cam}}{z_{cam}}, \frac{y_{cam}}{z_{cam}})$. In Barreto and Araujo (2001) it is proved that function $\mathbf{f_i}$, transforming coordinates in the projective plane $\mathbf{x} = (x, y, z)^t$ into cartesian coordinates in the catadioptric image plane $\mathbf{x_i} = (x_i, y_i)^t$, is given by Eq. (18) and that general central catadioptric image formation follows the scheme of Fig. 1. Function $\mathbf{f_i}$ depends on mirror parameters *l* and *m*. Table 3 shows these parameters for the different situations of central catadioptric imaging.

$$\mathbf{f_i}() : (x, y, z)$$
$$\rightarrow \left(\frac{(m-l)x}{z+l\sqrt{x^2+y^2+z^2}}, -\frac{(m-l)y}{z+l\sqrt{x^2+y^2+z^2}} \right) \tag{18}$$

Figure 4 depicts an intuitive "concrete" model for the proposed general central projection mapping. To each visible point in space corresponds an oriented projective ray **x** joining the 3D point with the effective projection center **O**. The projective ray intersects a unit sphere centered in **O** in a unique point **Q**. Consider a point $\mathbf{O_c}$ with coordinates $(0, 0, -l)^t$ in sphere reference frame $\Re$. To each **x** corresponds an oriented projective ray $\mathbf{x_c}$ joining $\mathbf{O_c}$ with the intersection point **Q** in the sphere surface. Assume that the catadioptric image plane is the horizontal plane $Z = m$. The projective ray $\mathbf{x_c}$ intersects the plane at $\mathbf{x_i}$ which are the coordinates of the image point. The scene is projected into the sphere surface and then points on the sphere are re-projected into the catadioptric image plane from
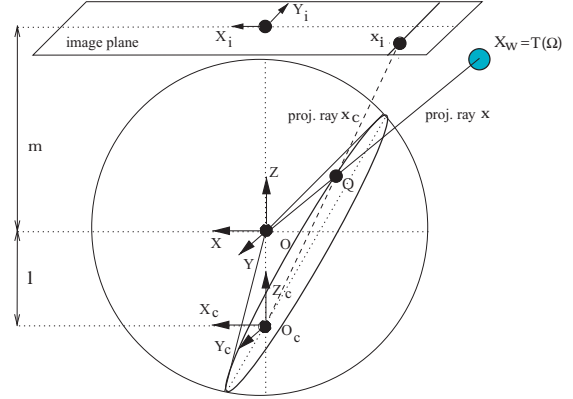

Figure 4. Unit sphere "concrete" model for general central catadioptric image projection.

a novel projection center $\mathbf{O_c}$. Point $\mathbf{O_c} = (0, 0, -l)^t$ only depends on the mirror parameters (see Table 3). For a parabolic mirror $l = 1$ and $\mathbf{O_c}$ belongs to the sphere surface. The re-projection is a stereographic projection. For the hyperbolic and elliptical cases $\mathbf{O_c}$ is inside the sphere in the negative *Z*-axis. The planar mirror is a degenerate case of central catadioptric projection where $l = 0$ and $\mathbf{O_c}$ is coincident with **O**. Function $\mathbf{f_i}$ becomes similar to the one given by Eq. (12) for the perspective camera situation, and the global mapping **F** is equal to the one studied in previous section.

2.4.2. The New World Coordinate System. The image formation process of a general central catadioptric system fits the scheme depicted in Fig. 1. The difference from the previous case of the perspective camera is that function $\mathbf{f_i}$ is given by Eq. (18) instead of Eq. (12) where *l* and *m* depend on the mirror parameters (see Table 3). The goal of this section is to derive a coordinate transformation Γ for which the global mapping **F** between points in the world and in the catadioptric image is in a compact and decoupled form. Differential constraints similar to Eqs. (13) and (14) can be derived in the same manner for this more general situation. As already mentioned, integration of partial differential equations can be a complex task, leading to multiple solutions. Once again the suitable new coordinates are derived geometrically and the differential constraints are used to confirm the results.

In Fig. 4 consider the vertical line in the catadioptric image plane parallel to the *Y* axis. All the points in the world that can be projected in this line lie in a

conic surface, with the vertex coincident with the effective viewpoint **O**. The conic surface intersects the unit sphere on a circumference, passing through the projection point **Q**, which limits a circle containing the point $\mathbf{O_c} = (0, 0, -l)^t$. The axis of the conic surface is always contained in the *XOZ* plane. In a similar way, the world points that can be projected in an horizontal line in the image lie in a conic surface. The difference is that the axis is now in the *YOZ* plane. We will call the first pencil of conic surfaces (with the axis in *XOZ*) the vertical pencil and the second (with the axis in *YOZ*) the horizontal pencil. A vertical and an horizontal conic surface intersect in two oriented projection rays. This ambiguity can be solved assuming that the camera used to acquire the catadioptric images is forward looking and that only points with $Z > -l$ are visible in the image plane. This section defines a coordinate system based in these two pencils of cones. It is shown that the resulting transformation of coordinates $\mathbf{\Gamma}$ verifies both the compactness and decoupling constraint for the general central projection situation.

Consider the dotted conic surface depicted in Fig. 5. Its axis coincides with the referential *X* axis, the vertex coincides with the origin and the aperture angle is α. World points (X, Y, Z) lying on the conic surface verify Eq. (19). Consider now the conic surface rotated around the *Y* axis by an angle ϕ. Equation (20) is obtained multiplying point coordinates by rotation matrix $\mathbf{e}^{\phi\hat{\mathbf{y}}}$ (where $\hat{\mathbf{y}}$ is the unit vector of the *Y* axis) (Murray et al., 1994) and replacing them in Eq. (19). Equation (20) defines the rotated conic surface depicted in Fig 5 (Left). Angle α is the aperture angle and ϕ is the rotation angle around *Y* axis that can be

used to index the cones in the vertical pencil described above.

$$Z^2 + Y^2 = X^2 \tan(\alpha)^2 \qquad (19)$$

$$(X \sin(\phi) + Z \cos(\phi))^2 - (X \cos(\phi) \\ - Z \sin(\phi))^2 \tan(\alpha)^2 + Y^2 = 0 \qquad (20)$$

Observe Fig. 5 (Right) where the rotated conic surface of Eq. (20), the unit sphere and the re-projection center $\mathbf{O_c}$ are depicted. The lower cone must intersect the unit sphere on a circle containing the point $(0, 0, -l)$. World scene points lying in this cone project on a vertical line at the catadioptric image plane (see Fig. 4). Figure 5 (Right) shows that the aperture angle of the conic surface must be $\tan(\alpha) = \frac{c}{e}$ where $e^2 = l^2 \sin^2(\phi)$ and $c^2 = 1 - e^2$ (an unit sphere is assumed). Equation (21) is obtained replacing this result in Eq. (20).

$$l^2 \sin(\phi)^2 (X^2 + Y^2 + Z^2) \\ - (X \cos(\phi) - Z \sin(\phi))^2 = 0 \qquad (21)$$

Solving Eq. (21) with respect to ϕ we are able to compute the vertical conic surface with the desired features that contains a certain world point (X, Y, Z). Notice however that it is a second order equation, thus for each point in 3D space, there are two ϕ solutions. Each solution defines a conic surface containing the point and intersecting the unit sphere in a circle passing through its projection **Q** at the sphere surface. Nevertheless one of these circles contains point $(0, 0, l)^t$, while the other contains point $(0, 0, -l)^t$. The second solution is the one that must be used.
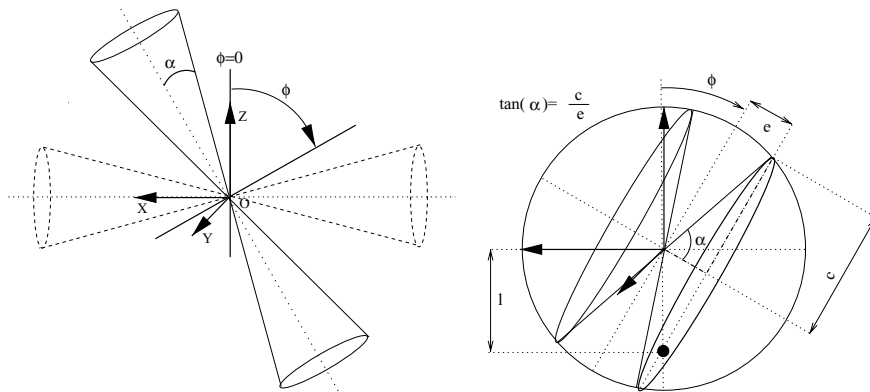

Figure 5. Left: Vertical pencil of conic surfaces indexed by the ϕ rotation angle. Right: Vertical pencil of conic surfaces. Intersection of the lower and upper cone with the unit sphere.

The ψ angle is used in the same manner to represent the horizontal pencil of conic surfaces. The derivation of the relationship between ψ and point world coordinates (X, Y, Z) is similar.

At this point, given a world point represented in Cartesian coordinates (X, Y, Z) we are able to compute a vertical and an horizontal conic surface, respectively referenced in an unique way by ϕ and ψ, that contain the point. In general the two cones intersect in two projective rays. However only one of them is visible in the image. The point lies in this projective ray, and a third coordinate ρ, that is the distance from the center, must be introduced. Equation (22) yields the derived change of coordinates Γ. Both ϕ and ψ are in the range $[-\pi/2, \pi/2]$.

$$
\begin{cases}
\phi = \arctan\left(\dfrac{X}{Z + l\sqrt{X^2 + Y^2 + Z^2})} \right) \\
\psi = \arctan\left(-\dfrac{Y}{Z + l\sqrt{X^2 + Y^2 + Z^2}} \right) \\
\rho = \sqrt{X^2 + Y^2 + Z^2}
\end{cases} \quad (22)
$$

Table 4 summarizes the results obtained so far for general central projection systems. Γ is the derived world coordinate transformation that maps cartesian coordinates $\mathbf{X_w}$ in the new coordinates Ω associated with the conic surfaces reference frame. $\mathbf{T}$ is the inverse function of Γ. Replacing $\mathbf{T}$ in Eq. (2) ($\mathbf{f_i}$, $\mathbf{P}$ and $\mathbf{f_h}$ were already defined) yields the mathematical expression of global mapping $\mathbf{F}$ using the new coordinates. The corre-

Table 4. Using a new coordinate system for general central catadioptric imaging.

$$
\Gamma(\mathbf{X_w}) = \begin{bmatrix} \arctan\left(\frac{X}{Z+l\sqrt{X^2+Y^2+Z^2}}\right) \\ \arctan\left(-\frac{Y}{Z+l\sqrt{X^2+Y^2+Z^2}}\right) \\ \sqrt{X^2+Y^2+Z^2} \end{bmatrix}
$$

$$
\mathbf{T}(\Omega) = \begin{bmatrix} \rho\frac{l+\sqrt{1+(1-l^2)(\tan(\phi)^2+\tan(\psi)^2)}}{1+\tan(\phi)^2+\tan(\psi)^2}\tan(\phi) \\ -\rho\frac{l+\sqrt{1+(1-l^2)(\tan(\phi)^2+\tan(\psi)^2)}}{1+\tan(\phi)^2+\tan(\psi)^2}\tan(\psi) \\ \rho\frac{\sqrt{1+(1-l^2)(\tan(\phi)^2+\tan(\psi)^2)}-l(\tan(\phi)^2+\tan(\psi)^2)}{1+\tan(\phi)^2+\tan(\psi)^2} \end{bmatrix}
$$

$$
\mathbf{F}(\Omega) = (m-l)\begin{bmatrix} \tan(\phi) \\ \tan(\psi) \end{bmatrix}
$$

$$
\mathbf{J_F}(\Omega) = (m-l)\begin{bmatrix} \frac{1}{\cos(\phi)^2} & 0 & 0 \\ 0 & \frac{1}{\cos(\psi)^2} & 0 \end{bmatrix}
$$

sponding jacobian matrix is $\mathbf{J_F}$. Accordingly Eqs. (10) and (11), the derived coordinate transformation Γ verify both the compactness and decoupling constraints. Notice that if $l = 0$ then we have the perspective camera situation studied previously. The aperture of the reference conic surfaces is always $\pi/2$ and the change of coordinates 22 becomes equal to 15. The pencils of conic surfaces degenerate in the pencils of planes derived above. The coordinate transformation proposed for the perspective camera situation is a particular case of this more general solution.

2.4.3. Applications. Applying the derived coordinate transformation Γ does not modify the image formation process nor introduces new information in the problem. Instead the geometric transformations are represented in a way that enables newer insights in the imaging process and is more suitable to develop certain applications. This section shows the advantages in expressing the global mapping function in terms of an adequate system of world coordinates.

Using the derived world coordinates Ω, the global mapping $\mathbf{F}$ is written in a compact and decoupled form (see Table 4). Knowing both target image position $\mathbf{x_i}$ and velocity $\dot{\mathbf{x}}_\mathbf{i}$ the recovery of 3D parameters of motion is straightforward. As a result $\phi = \arctan(x_i)$ and $\psi = \arctan(y_i)$. Replacing these parameters in the jacobian matrix of Table 4 yields the jacobian matrix of the global mapping $\mathbf{F}$ written in terms of image coordinates (Eq. (23)). It follows that $\dot{\phi} = (1 + \frac{x_i^2}{(m-l)^2})^{-1}\frac{\dot{x}_i}{m-l}$ and $\dot{\psi} = -(1 + \frac{y_i^2}{(m-l)^2})^{-1}\frac{\dot{y}_i}{m-l}$. Similarly to the perspective camera case, catadioptric images do not provide any information about the third parameter of motion ρ.

$$
\mathbf{J_F}(\mathbf{x_i}) = (m-l)\begin{bmatrix} 1 + \frac{x_i^2}{(m-l)^2} & 0 & 0 \\ 0 & 1 + \frac{y_i^2}{(m-l)^2} & 0 \end{bmatrix} \quad (23)
$$

It is a well known result that it is always possible to generate a geometrically correct perspective image from an image acquired by a central projection catadioptric system (Baker and Nayar, 1998). The first image of Fig. 6 was acquired by an omnidirectional system that combines a parabolic mirror with an orthographic camera. A generic scene point with cartesian coordinates $\mathbf{X_w}$ is projected at position $\mathbf{x_i}$ in the catadioptric image plane (see Fig. 4). The field of view is nearly 180, thus all points in the world such that $Z > 0$ are visible. Assume a conventional perspective camera with projection center at the effective viewpoint $\mathbf{0}$ and optical axis aligned with the Z-axis. A generic world point with
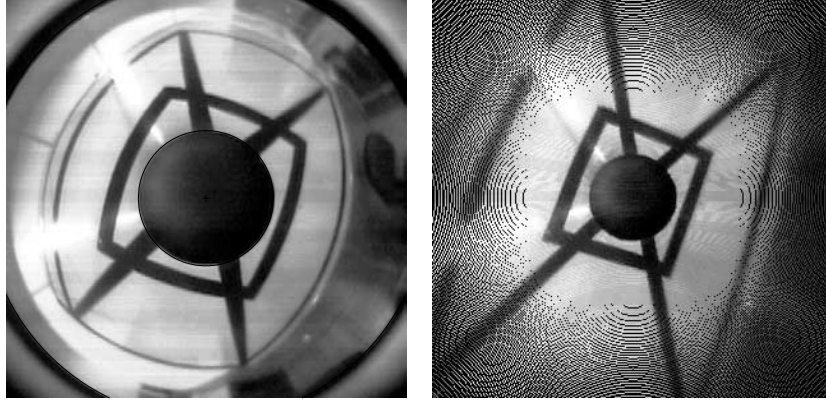
Figure 6. Generating a geometrically correct perspective image.

coordinates $\mathbf{X_w} = (X, Y, Z)^t$ is projected in the perspective image at position $\mathbf{x_p} = (\frac{X}{Z}, \frac{Y}{Z})^t$. The goal is to derive the function $\mathbf{f_p}$ that maps points in the catadioptric image plane into points in the perspective image plane ($\mathbf{x_p} = \mathbf{f_p}(\mathbf{x_i})$). This mapping function can be determined in a straightforward manner using the results of Table 4. As already mentioned $\phi = \arctan(x_i)$ and $\psi = \arctan(y_i)$. Replacing ϕ and ψ in the coordinate transformation $\mathbf{T}$ we are able to compute world coordinates $\mathbf{X_w}$ as a function of catadioptric image coordinates $\mathbf{x_i}$ and ρ. Making $\mathbf{x_p} = (\frac{X}{Z}, \frac{Y}{Z})^t$ the dependence on ρ disappears and $\mathbf{x_p}$ is obtained as a function of $\mathbf{x_i}$. Function $\mathbf{f_p}$ is presented in Eq. (24) for the parabolic system case ($l = 1$). The geometrically correct perspective image generated from the derived mapping function is presented at Fig. 6.

$$f_p() : (x_i, y_i)$$
$$\rightarrow \left(\frac{4px_i}{4p^2 - \left(x_i^2 + y_i^2\right)}, -\frac{4py_i}{4p^2 - \left(x_i^2 + y_i^2\right)} \right) \quad (24)$$

Spherical coordinates are broadly used in computer vision and robotics. For some applications it can be useful to reference world points using spherical coordinates. Assume, as an example, that we intend to use our omnidirectional camera in a surveillance application to obtain position information of a set of targets. The imaging system is fixed at the ceiling as depicted in Fig. 7. The target position can be referenced in a simple way using spherical coordinates $\mathbf{\Omega_s} = (\phi_s, \psi_s, \rho_s)^t$. The goal is to derive the function $\mathbf{f_s}$ which transforms catadioptric image coordinates $\mathbf{x_i}$, in the two world spherical coordinates (ϕ_s, ψ_s) that can be fully recovered ($(\phi_s, \psi_s) = \mathbf{f_s}(\mathbf{x_i})$) without further restrictions. $\mathbf{X_w} = (X, Y, Z)^t$ can be computed
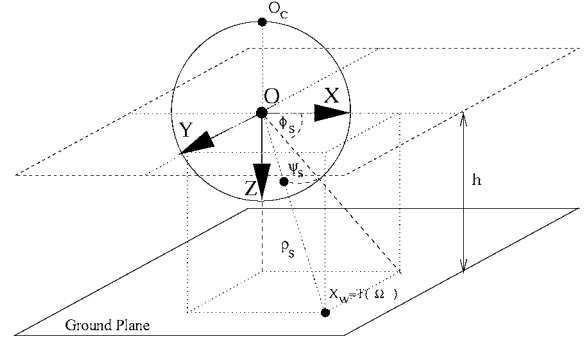

Figure 7. Surveillance application using central panoramic imaging.

in terms of $\mathbf{x_i}$ and ρ. Since $\phi = \arctan(x_i)$ and $\psi = \arctan(y_i)$, then $\mathbf{X_w} = \mathbf{T}(\arctan(x_i), \arctan(y_i), \rho)$ (see Table 4). Replacing (X, Y, Z) in $\phi_s = \arctan(\frac{X}{Z})$ and $\psi_s = \arctan(\frac{Y}{\sqrt{X^2+Z^2}})$ by the obtained result, the dependence on ρ is eliminated and function $\mathbf{f_s}$ is obtained (Eq. (25)). In addition, if the height h of the room is known and if the imaged point is on the floor one obtains $\rho_s = h\sqrt{(4p^2 + x_i^2 + y_i^2)^2/(4p^2 - x_i^2 - y_i^2)^2}$.

$$f_s() : (x_i, y_i) \rightarrow \left(\arctan\left(\frac{4px_i}{4p^2 - \left(x_i^2 + y_i^2\right)} \right), \right.$$
$$\left. \arctan\left(\frac{-4py_i}{\sqrt{\left(4p^2 + x_i^2 - y_i^2\right)^2 - 4x_i^2 y_i^2}} \right) \right) \quad (25)$$

3. Imaging Systems with Motion

Section 2 focuses on static central projection imaging systems. The global mapping $\mathbf{F}$ from points in the 3D

world in points in the 2D image plane is derived. It is shown that the mathematical expression of **F** depends on the coordinate system used to reference points in the scene and that an adequate choice of coordinates presents several advantages. In this situation the imaging system does not move, thus motion in the image plane only depends on 3D motion in the scene.

This section focuses on active tracking of moving targets where the goal is to use visual information to control camera motion such that target image is kept constant. The mathematical framework derived in Section 2 is extended to central projection imaging systems with rigid motion. Global mapping between 3D world points and 2D image points is derived. Criteria to select adequate world coordinates are discussed. The results obtained are applied to active visual tracking using three different platforms and/or imaging sensors.

As mentioned in Section 1 the platforms considered have less than 3 d.o.f. To control such a constrained motion it is enough to work with 3D target position and velocity. If nothing is stated it is assumed that $\mathbf{X_w}$ is a vector with target mass center position in cartesian coordinates, $\mathbf{\Omega}$ represents the target position in the established alternative system of coordinates and $\mathbf{x_i}$ is the vector with the target image coordinates. Vector $\mathbf{\Theta}$ is introduced to represent camera/platform position in an inertial world reference frame.

3.1. Mapping Points from the 3D World into the 2D Image Plane

Consider the schematic of Fig. 1 which depicts the general mapping **F** performed by a static central projection imaging system. Consider $\mathbf{P} = \mathbf{R}[\mathbf{I} \,|\, -\mathbf{C}]$, with **R** the rotation matrix between camera and world and **C** the projection center coordinates. Matrix **P** depends on the pose of the imaging system in world coordinates. If the imaging system is static then the matrix **P** is constant. On the other hand if the camera moves, its pose changes, and matrix **P** is no longer constant. Assume the imaging system is mounted on a moving platform. The pose of the camera depends on the position $\mathbf{\Theta}$ of the platform. Since matrix **P** depends on the pose of the camera then **P** is a function of $\mathbf{\Theta}$.

The mapping between the scene and the image plane depends on the target 3D coordinates $\mathbf{\Omega}$ and on the camera pose parameterized by $\mathbf{\Theta}$. The target image coordinates $\mathbf{x_i}$ are given by Eq. (26) where transformation **F** can be written as the composition of Eq. (27). The difference between Eqs. (2) and (27) is that matrix **P** is

no longer constant and appears as a function of $\mathbf{\Theta}$.

$$\mathbf{x_i} = \mathbf{F}(\mathbf{\Omega}, \mathbf{\Theta}) \qquad (26)$$

$$\mathbf{F}(\mathbf{\Omega}, \mathbf{\Theta}) = \mathbf{f_i}(\mathbf{P}(\mathbf{\Theta}) \cdot \mathbf{f_h}(\mathbf{T}(\mathbf{\Omega}))) \qquad (27)$$

The target velocity in the image $\mathbf{x_i}$ is computed in (28). Equation (28) is obtained by differentiating Eq. (26) with respect to time. $\mathbf{J_F} = [\mathbf{J_F^\Omega} \,|\, \mathbf{J_F^\Theta}]$ is the jacobian matrix of function **F**, $\dot{\mathbf{\Omega}}$ is the target 3D velocity, and $\dot{\mathbf{\Theta}}$ represents the camera/platform velocity. $\mathbf{J_F^\Omega}$ is given in (29) with $\mathbf{J_\Gamma}$ the jacobian matrix of the inverse coordinate transformation $\mathbf{\Gamma}$. $\mathbf{J_F^\Theta}$ is computed in (30) and does not depend on the jacobian matrix of the world coordinate transformation $\mathbf{J_\Gamma}$. The image velocity $\dot{\mathbf{x_i}}$ depends both on target velocity $\dot{\mathbf{\Omega}}$ and on the camera/platform velocity $\dot{\mathbf{\Theta}}$. The second term in Eq. (28) is known in the literature by egomotion and represents the image motion induced by camera/platform motion.

$$\dot{\mathbf{x_i}} = \mathbf{J_F^\Omega} \dot{\mathbf{\Omega}} + \mathbf{J_F^\Theta} \dot{\mathbf{\Theta}} \qquad (28)$$

$$\mathbf{J_F^\Omega} = \mathbf{J_{f_i}} \cdot \mathbf{J_P^{X_h}} \cdot \mathbf{J_{f_h}} \cdot \mathbf{J_\Gamma^{-1}} \qquad (29)$$

$$\mathbf{J_F^\Theta} = \mathbf{J_{f_i}} \cdot \mathbf{J_P^\Theta}. \qquad (30)$$

The target image depends on the relative position between camera and target. Describing target motion in a coordinate frame attached to the imaging device simplifies the position/velocity relationships of Eqs. (26) to (30). The egomotion term disappears and the position/velocity in the image depends only on the position/velocity of the target in the camera system of coordinates. The platform position, provided, for example, by encoder readings, and joint control commands are usually defined in an inertial base coordinate frame. If the control input is defined in the camera coordinate system then the transformations between the two reference systems can not be avoided. In the position/velocity relationships of Eqs. (26) to (30), both camera and target motion are described in a common inertial system of coordinates. Errors in the image plane can be directly related with the control inputs commands in the task space, thus there is no need of additional coordinate transformations. Multiple cameras can be integrated in a natural way by describing the target motion in a common coordinate frame. The explicit computation of an egomotion term can be used for image segmentation (Sharkey et al., 1993; Batista et al., 1997; Barreto et al., 1999).

The mathematical expression of the global mapping $\mathbf{F}$ depends on system of coordinates used to reference target position in the scene (Eq. (27)). The intrinsic nature of the mapping does not depend on the selected coordinate frame. However, as seen in Section 2 for static central catadioptric imaging systems, an adequate choice of the coordinate system can be highly advantageous. Criteria to select suitable coordinate transformations $\mathbf{\Gamma}$ for tracking applications are discussed in the present section.

3.2.1. Analytical Solution for Visual Control of Motion in Tracking Applications.   Consider the imaging system mounted on a moving platform. If nothing is stated assume a platform with 2 d.o.f ($\mathbf{\Theta}$ is a 2×1 vector). To perform active tracking of a moving target, the platform motion must be controlled such that the target projection is kept in constant position/velocity in the image plane. Assume that $\mathbf{x_d}$ and $\dot{\mathbf{x}}_\mathbf{d}$ are the desired target position and velocity in the image. Consider that $\mathbf{\Theta_c}$ and $\dot{\mathbf{\Theta}}_\mathbf{c}$ are the position and velocity commands sent to the platform actuators which have an unitary transfer function. Thus, $\mathbf{\Theta_c}$ and $\dot{\mathbf{\Theta}}_\mathbf{c}$ must verify both $\mathbf{x_d} = \mathbf{F}(\mathbf{\Omega}, \mathbf{\Theta_c})$ and $\dot{\mathbf{x}}_\mathbf{d} = \mathbf{J}_\mathbf{F}^\Omega \cdot \dot{\mathbf{\Omega}} + \mathbf{J}_\mathbf{F}^\Theta \cdot \dot{\mathbf{\Theta}}_\mathbf{c}$ with $\mathbf{\Omega}$ and $\dot{\mathbf{\Omega}}$ target 3D position and velocity (see Eqs. (26) to (30)).

If nothing is stated assume in the sequel that at each frame time instant both target position $\mathbf{x_i}$ and velocity $\dot{\mathbf{x}}_\mathbf{i}$ are measured in the image plane and that the platform position $\mathbf{\Theta}$ and velocity $\dot{\mathbf{\Theta}}$ are estimated using the encoder readings. The goal is to determine the position and velocity commands $\mathbf{\Theta_c}$ and $\dot{\mathbf{\Theta}}_\mathbf{c}$ knowing $\mathbf{x_i}$, $\dot{\mathbf{x}}_\mathbf{i}$, $\mathbf{\Theta}$ and $\dot{\mathbf{\Theta}}$.

Proposition 1.   *Assume that it is possible to compute target 3D position $\mathbf{\Omega}$ from target position in image $\mathbf{x_i}$ and camera pose $\mathbf{\Theta}$. In a similar way consider that camera position $\mathbf{\Theta}$ can be uniquely calculated given target position in image $\mathbf{x_i}$ and in the scene $\mathbf{\Omega}$. If these two conditions hold then, given target position in image $\mathbf{x_i}$ and camera pose $\mathbf{\Theta}$, it is possible to compute camera position $\mathbf{\Theta_c}$ such that target is projected in a pre-defined position $\mathbf{x_d}$ in the image plane. A similar statement can be made for velocity relationships.*

Proof of Proposition 1:   Function $\mathbf{F}$ computes target position in image $\mathbf{x_i}$ given target 3D position $\mathbf{\Omega}$ and camera pose $\mathbf{\Theta}$ (Eq. (26)). Assume that it exists a function ω which enables the computation of $\mathbf{\Omega}$ given both

$\mathbf{x_i}$ and $\mathbf{\Theta}$ (Eq. (31)). In a similar way, using function θ, $\mathbf{\Theta}$ can be calculated knowing both $\mathbf{x_i}$ and $\mathbf{\Omega}$ (Eq. (32)).

$$\mathbf{\Omega} = \omega(\mathbf{x_i}, \mathbf{\Theta}) \tag{31}$$

$$\mathbf{\Theta} = \theta(\mathbf{x_i}, \mathbf{\Omega}) \tag{32}$$

Assume that the target image position $\mathbf{x_i}$ is measured and camera pose $\mathbf{\Theta}$ is estimated from encoder readings. The world target 3D position is $\omega(\mathbf{x_i}, \mathbf{\Theta})$. The camera position $\mathbf{\Theta_c}$ such that target image is projected in position $\mathbf{x_d}$ is given in Eq. (33).

$$\mathbf{\Theta_c} = \theta(\mathbf{x_d}, \omega(\mathbf{x_i}, \mathbf{\Theta})) \tag{33}$$

Equations (34) and (35) are obtained differentiating Eqs. (31) and (32) with respect to time. $\mathbf{J}_\omega = [\mathbf{J}_\omega^\mathbf{i} \mid \mathbf{J}_\omega^\Theta]$ and $\mathbf{J}_\theta = [\mathbf{J}_\theta^\mathbf{i} \mid \mathbf{J}_\theta^\Theta]$ are the jacobian matrices of functions ω and θ.

$$\dot{\mathbf{\Omega}} = \mathbf{J}_\omega^\mathbf{i} \dot{\mathbf{x}}_\mathbf{i} + \mathbf{J}_\omega^\Theta \dot{\mathbf{\Theta}} \tag{34}$$

$$\dot{\mathbf{\Theta}} = \mathbf{J}_\theta^\mathbf{i} \dot{\mathbf{x}}_\mathbf{i} + \mathbf{J}_\theta^\Omega \dot{\mathbf{\Omega}} \tag{35}$$

Knowing both the target velocity in the image $\dot{\mathbf{x}}_\mathbf{i}$ and the camera velocity $\dot{\mathbf{\Theta}}$ at the frame acquisition time instant, it is possible to compute the camera velocity $\dot{\mathbf{\Theta}}_\mathbf{c}$ such that target is projected in the image plane with desired velocity $\dot{\mathbf{x}}_\mathbf{d}$. This results is shown in Eq. (36) derived by differentiating (33) with respect to time.

$$\dot{\mathbf{\Theta}}_\mathbf{c} = \mathbf{J}_\theta^\mathbf{i} \dot{\mathbf{x}}_\mathbf{d} + \mathbf{J}_\theta^\Omega \mathbf{J}_\omega^\mathbf{i} \dot{\mathbf{x}}_\mathbf{i} + \mathbf{J}_\theta^\Omega \mathbf{J}_\omega^\Theta \dot{\mathbf{\Theta}} \tag{36}$$

$\square$

Proposition 2.   *Consider the global mapping function $\mathbf{F}$ and the corresponding jacobian matrix $\mathbf{J_F} = [\mathbf{J}_\mathbf{F}^\Omega \mid \mathbf{J}_\mathbf{F}^\Theta]$. If it exists a function ω that enables the computation of the target 3D position from target image and camera pose, then $\mathbf{J}_\mathbf{F}^\Omega$ must be invertible. In a similar way if it exists a function θ that allows the calculation of the camera position from the target position in the image and in the world, then matrix $\mathbf{J}_\mathbf{F}^\Theta$ must have inverse*

Proof of Proposition 2:   Assume that the function ω exists (Eq. (31)). The mathematical relationship of Eq. (37) is obtained replacing $\dot{\mathbf{\Omega}}$ in Eq. (28) by the result of Eq. (34). Equation (38) is derived replacing in (34) $\dot{\mathbf{x}}_\mathbf{i}$ by the result presented in (28).

$$\dot{\mathbf{x}}_\mathbf{i} = \mathbf{J}_\mathbf{F}^\Omega \mathbf{J}_\omega^\mathbf{i} \dot{\mathbf{x}}_\mathbf{i} + (\mathbf{J}_\mathbf{F}^\Omega \mathbf{J}_\omega^\Theta + \mathbf{J}_\mathbf{F}^\Theta) \dot{\mathbf{\Theta}} \tag{37}$$

$$\dot{\mathbf{\Omega}} = \mathbf{J}_\omega^\mathbf{i} \mathbf{J}_\mathbf{F}^\Omega \dot{\mathbf{\Omega}} + (\mathbf{J}_\omega^\mathbf{i} \mathbf{J}_\mathbf{F}^\Theta + \mathbf{J}_\omega^\theta) \dot{\mathbf{\Theta}} \tag{38}$$

If function ω exists then equalities (37) and (38) must hold. From these equalities come $\mathbf{J}_\omega^\mathbf{i}\mathbf{J}_\mathbf{F}^\Omega = \mathbf{J}_\mathbf{F}^\Omega\mathbf{J}_\omega^\mathbf{i} = \mathbf{I}$. Thus matrix $\mathbf{J}_\mathbf{F}^\Omega$ must have inverse which is $\mathbf{J}_\omega^\mathbf{i}$. Considering that $\mathbf{J}_\mathbf{F}^\Omega\mathbf{J}_\omega^\Theta + \mathbf{J}_\mathbf{F}^\Theta = \mathbf{J}_\omega^\mathbf{i}\mathbf{J}_\mathbf{F}^\Theta + \mathbf{J}_\omega^\Theta = 0$ and making $\mathbf{J}_\omega^\mathbf{i} = (\mathbf{J}_\mathbf{F}^\Omega)^{-1}$ comes the results of Eq. (39).

$$\begin{cases} \mathbf{J}_\omega^\mathbf{i} = (\mathbf{J}_\mathbf{F}^\Omega)^{-1} \\ \mathbf{J}_\omega^\Theta = -(\mathbf{J}_\mathbf{F}^\Omega)^{-1}\mathbf{J}_\mathbf{F}^\Theta \end{cases} \tag{39}$$

Consider now that function θ exists. In a similar way it can be proved that $\mathbf{J}_\mathbf{F}^\Theta$ must be invertible and that $\mathbf{J}_\theta^\mathbf{i}$ and $\mathbf{J}_\theta^\Omega$ are given by Eq. (40).

$$\begin{cases} \mathbf{J}_\theta^\mathbf{i} = (\mathbf{J}_\mathbf{F}^\Theta)^{-1} \\ \mathbf{J}_\theta^\Omega = -(\mathbf{J}_\mathbf{F}^\Theta)^{-1}\mathbf{J}_\mathbf{F}^\Omega \end{cases} \tag{40}$$

Equation (36) is rewritten in Eq. (41) using the results of Eqs. (39) and (40)

$$\dot{\Theta}_\mathbf{c} = \dot{\Theta} + (\mathbf{J}_\mathbf{F}^\Theta)^{-1}(\dot{\mathbf{x}}_\mathbf{d} - \dot{\mathbf{x}}_\mathbf{i}) \tag{41}$$

$\square$

3.2.2. Coordinate Transformation Constraints.
In the previous section Proposition 2 establishes the necessary conditions for the existence of functions ω and θ and Proposition 1 establishes the sufficient conditions for the existence of an analytical solution for the tracking problem.

Assuming a platform with 2 d.o.f, then $\mathbf{J}_\mathbf{F}^\Theta$ is a 2 × 2 matrix. This matrix is invertible or not depending on the features of the system, in particular the kinematics of the robotic platform and the type of image sensor. The coordinate frame where the platform position Θ is defined is considered as a problem specification that can not be changed.

In general, the world target position is referenced by $\Omega = (\phi, \psi, \rho)^t$, which is a 3 × 1 vector. The corresponding jacobian matrix $\mathbf{J}_\mathbf{F}^\Omega$ is a 2 × 3 matrix similar to the one shown in Eq. (9). A 2 × 3 matrix is never invertible and, accordingly to Proposition 2, it is not possible to define a function ω to recover target 3D position information. As mentioned in Section 2, the image formation process transforms points in $\Re^3$ into points in $\Re^2$. There is a loss of information and world target position can not be fully recovered from a single image. Nevertheless, as mentioned in 2.2, if $\mathbf{F}$ is written in a compact form then Ω can be partially recovered in a straightforward manner. If the recovered target 3D parameters are enough to derive function θ

then an analytical solution for the tracking problem can be reached (Proposition 1).

If it exists a transformation of coordinates Γ verifying the compactness constraint then it is possible to write the global mapping $\mathbf{F}$ in a compact form (Section 2.2). This means that $\mathbf{F}$ does not depend on ρ and that the third column of $\mathbf{J}_\mathbf{F}^\Theta$ is zero. In practice, if the compactness constraint is verified, we can replace in Eqs. (26) to (41) $\Omega = (\phi, \psi, \rho)^t$ by $\hat{\Omega} = (\phi, \psi)^t$ and assume that $\mathbf{J}_\mathbf{F}^\Omega$ is a 2 × 2 matrix by discarding the null column. The verification of the compactness constraint is a necessary conditions for $\mathbf{J}_\mathbf{F}^\Omega$ being a square matrix with inverse. Considering the statement of Proposition 2, the verification of the compactness constraint is a necessary condition to exist an ω function. If both ω and θ exist then there is an analytical solution for the active tracking problem (Proposition 1).

From the statements above we can conclude that it is desirable to select a transformation of coordinates Γ verifying the compactness constraint (Eq. (10)). If Γ also verifies the decoupling constraint (Eq. (11)) then $\mathbf{F}$ is written in a decoupled way which simplifies the calculations to obtain the tracking control laws (Eqs. (33) and (36)). Another useful guideline is to select transformation Γ such that target 3D position is referenced in a coordinate frame of the same type as the one where camera position Θ is defined. This can also lead to several simplifications in the calculation as well as a deeper understanding of the tracking task.

In the sequel the derived mathematical framework is applied to active tracking of a moving target using three different robotic platforms. These examples help the understanding of the exposed ideas and illustrate the usefulness of a judicious selection of the world coordinate system.

3.3. Active Tracking Using a Perspective Camera with Translation Motion in the XY Plane

Figure 8 depicts a conventional perspective camera with translation motion in the *XOY* plane. A schematic of the framework to derive global mapping $\mathbf{F}$ is shown in Fig. 1. The imaging sensor is a perspective camera with intrinsic parameters $\mathbf{K} = \mathbf{I}$, therefore function $\mathbf{f}_\mathbf{i}$ is the one shown in Eq. (12). The camera position in the *XOY* plane in world cartesian coordinates is $\Theta = (t_X, t_Y)^t$. For this particular case $\mathbf{P}(\Theta)$ is given by Eq. (42). The goal of the application is to control camera position and velocity such that target position and velocity in image are zero ($\mathbf{x}_\mathbf{d} = \dot{\mathbf{x}}_\mathbf{d} = 0$). Therefore
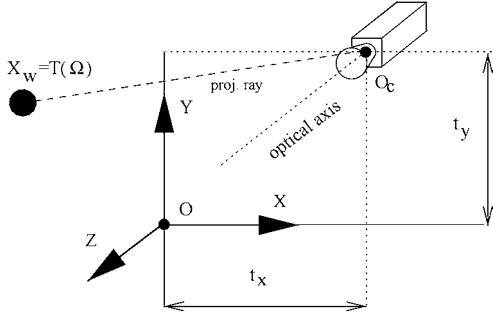
Figure 8. Active tracking using a perspective camera with translation motion.

the camera motion must be controlled in such a way as to keep the optical axis aligned with target mass center.

$$\mathbf{P}(\Theta) = \begin{bmatrix} 1 & 0 & 0 & -t_X \\ 0 & 1 & 0 & -t_Y \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (42)$$

3.3.1. Selecting Coordinate Transformation Γ.
In a similar way to what was done in Section 2 we intend to select a suitable coordinate transformation Γ (Eq. (5)) for this specific application. Functions $\mathbf{f_i}$, $\mathbf{P}(\Theta)$ and $\mathbf{f_h}$ have been defined already. The jacobian matrix $\mathbf{J_F^\Omega}$ can be written in terms of the partial derivatives of Γ by replacing $\mathbf{J_\Gamma}$ in Eq. (29) by the result of Eq. (6).

If we intend to select a coordinate transformation Γ verifying the compactness constraint then the third column of matrix $\mathbf{J_F^\Omega}$ must be zero (Eq. (10)). This yields the partial differential equations shown in Eq. (43).

$$\begin{cases} Z(\phi_Y\psi_Z - \phi_Z\psi_Y) + (X - t_X)(\phi_Y\psi_X - \phi_X\psi_Y) = 0 \\ Z(\phi_Z\psi_X - \phi_X\psi_Z) + (Y - t_Y)(\phi_Y\psi_X - \phi_X\psi_Y) = 0 \end{cases}$$
$$(43)$$

The system of Eqs. (43) can be solved with respect to the partial derivatives of Γ. $\mathbf{J_\Gamma}$ is a 3×3 matrix which yields 9 unknowns for 2 equations. The problem is under determined and multiple solutions can be found. Some of the solutions obtained can be discarded by considering the additional constraint $\det(\mathbf{J_\Gamma}) \neq 0$ (transformation Γ must be bijective). Following this procedure the result of Eq. (44) is derived. If the coordinate transformation Γ verifies the compactness constraint then the structure of the corresponding jacobian matrix $\mathbf{J_\Gamma}$ must be the one shown in Eq. (44). Notice that the provided jacobian matrix depends both

on $\mathbf{X_w} = (X, Y, Z)^t$ and $\Theta = (t_x, t_y)^t$. Any function Γ verifying Eq. (43) must depend, not only on $\mathbf{X_w}$, but also on Θ. Γ is no longer a transformation of inertial world cartesian coordinates and we can conclude that it is not possible to find a transformation of coordinates verifying the compactness constraint.

$$\mathbf{J_\Gamma} = \begin{bmatrix} \phi_X & \phi_Y & -\frac{(X-t_X)\phi_X + (Y-t_Y)\phi_Y}{Z} \\ \psi_X & \psi_Y & -\frac{(X-t_X)\psi_X + (Y-t_Y)\psi_Y}{Z} \\ \rho_X & \rho_Y & \rho_Z \end{bmatrix} \quad (44)$$

The conclusions drawn from the discussion above were expected. If the camera has translation motion there is no way to suppress the dependence on the third coordinate and it is not possible to perform the required motion control using only visual information. Some authors overcome the problem by assuming additional constraints such the target moving in a plane in the world (Corke, 1996; Krautgartner and Vincze, 1998). A systematic approach to determine if it exists any coordinate transformation verifying the constraints specified in 2.2 has been presented. The proposed procedure establishes necessary conditions. The fulfillment of the conditions does not guarantee the existence of a desired coordinate transformation Γ. This systematic approach will be repeated in the next two cases.

3.4. Active Tracking Using a Perspective Camera with Pan and Tilt Rotation Motion

A perspective camera is mounted on a pan and tilt unit such that both rotation axes go through the optical center $\mathbf{O_c}$. The camera first rotates in pan and then in tilt (Fick Model) as depicted in Fig. 9. The camera position
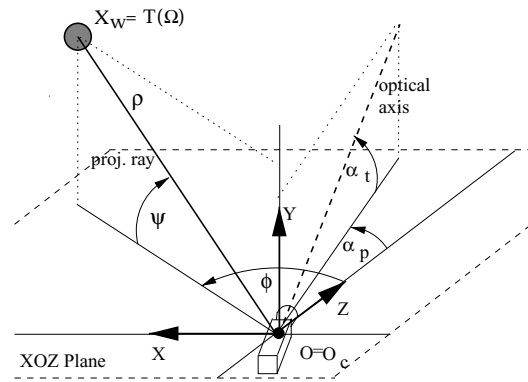

Figure 9. Active tracking using a perspective camera mounted on a pan and tilt unit.

vector is $\Theta = (\alpha_p, \alpha_t)^t$ with α_p the pan angle and α_t the tilt angle.

Consider the global mapping scheme depicted in Fig. 1. For this particular case function $\mathbf{f_i}$ is provided in Eq. (12) (perspective camera with $\mathbf{K} = \mathbf{I}$), Eq. (45) shows $\mathbf{P}(\Theta)$ with $e^{-\alpha_p \hat{\mathbf{y}}_c}$ and $e^{-\alpha_t \hat{\mathbf{x}}_c}$ the pan and tilt rotation matrices (Murray et al., 1994). Unit vectors $\hat{\mathbf{x}}_c$ and $\hat{\mathbf{y}}_c$ are associated with the X_c and Y_c axes of the cartesian coordinate frame attached to the camera (not depicted in Fig. 9). The goal of the tracking application is to control camera rotation such that both target position and velocity in image is zero ($\mathbf{x_d} = \dot{\mathbf{x}}_\mathbf{d} = 0$). To achieve this the camera optical axis must be aligned with target in 3D space.

$$\mathbf{P}(\Theta) = e^{-\alpha_t \hat{\mathbf{x}}_c} e^{-\alpha_p \hat{\mathbf{y}}_c} [\mathbf{I} \,|\, \mathbf{0}] \qquad (45)$$

3.4.1. Selecting Coordinate Transformation Γ.

Functions $\mathbf{f_i}$, $\mathbf{P}(\Theta)$ and $\mathbf{f_h}$, as well as the corresponding jacobian matrices, are defined. Assume $\mathbf{J_\Gamma}$ given by Eq. (6). According to Eq. (29), matrix $\mathbf{J_F^\Omega}$ can be written in terms of the partial derivatives of coordinate transformation Γ. The procedure described in Section 3.3.1 is repeated. If Γ verifies the compactness constraint then its jacobian matrix must have the structure shown in Eq. (46). For the pan and tilt tracking situation there is a solution $\mathbf{J_\Gamma}$ that only depends on $\mathbf{X_w}$. Thus the existence of a cartesian coordinate transformation Γ verifying the compactness constraint is not excluded. Nevertheless, repeating the procedure to achieve a decoupled $\mathbf{F}$ function, allows us to conclude that it is not possible to find a transformation Γ which verifies the decoupling constraint.

$$\mathbf{J_\Gamma} = \begin{bmatrix} \phi_X & -\frac{X\phi_X + Z\phi_Z}{Y} & \phi_Z \\ -\frac{Y\psi_Y + Z\psi_Z}{X} & \psi_Y & \psi_Z \\ \rho_X & \rho_Y & \rho_Z \end{bmatrix} \qquad (46)$$

One of the suggested guidelines for the selection of transformation Γ is to reference the target 3D position in a coordinate frame of the same type as the one where the camera position Θ is defined.

Consider the transformation from cartesian coordinates into spherical coordinates proposed in Eq. (47). The angles ϕ and ψ, used to reference the target 3D position, are similar to the camera pan and tilt angles (see Fig. 9). Moreover the jacobian matrix of the coordinate transformation Eq. (47) follows the structure presented in Eq. (46) which assures that the compact-

ness constraint is verified.

$$\begin{cases} \phi = \arctan\left(\dfrac{X}{Z}\right) \\ \psi = \arctan\left(-\dfrac{Y}{\sqrt{X^2 + Z^2}}\right) \\ \rho = \sqrt{X^2 + Y^2 + Z^2} \end{cases} \qquad (47)$$

Table 5 summarizes the results obtained using the coordinate transformation of Eq. (47). Both Γ and its inverse $\mathbf{T}$ are presented. Equation (48) defines an angular error vector Δ. The goal of the tracking application is to align the camera optical axis with the target mass center. This is verified whenever Δ is zero. Notice that by using coordinates of the same type to reference both the target and the camera position, the dependence of the global mapping $\mathbf{F}$ on the angular tracking errors δ_p and δ_t is explicit. The third column of $\mathbf{J_F^\Omega}$ is zero which means that Γ verifies the compactness constraint. Due to lack of space the polynomials Υ_1, Υ_2 and Υ_3 are presented at the bottom of the table.

$$\Delta = (\delta_p, \delta_t)^t = (\phi - \alpha_p, \psi - \alpha_t)^t \qquad (48)$$

3.4.2. Active Tracking Control Law.

Assume that Θ_c and $\dot{\Theta}_c$ are the position and velocity commands that must be sent to platform actuators to accomplish a specified task (the actuators transfer function is unitary). The goal is to track a moving object such that its projection is kept in the image center. The position command Θ_c can be obtained by making $\mathbf{x_d} = \mathbf{0}$ in Eq. (33) ($\Theta_c = \theta(0, \omega(x_i, \Theta))$). Notice that the specified tracking task is accomplished by controlling the camera motion such that the optical axis becomes aligned with target mass center. Therefore the position command must be $\Theta_c = \hat{\Omega}$ with $\hat{\Omega} = (\phi, \psi)^t$. Considering that $\hat{\Omega} = \Theta + \Delta$ (Eq. (48)) it results $\Theta_c = \Theta + \Delta$. The explicit computation of ω and θ functions can be avoided by determining angular error Δ as a function of target position in image $\mathbf{x_i}$ and camera pose Θ

In Table 5 the global mapping $\mathbf{F}$ is written in a compact form such that $\mathbf{x_i}$ depends on $\hat{\Omega}$ and Θ. The 3D position parameters ϕ and ψ can be recovered from the target position in the image (x_i, y_i) and the camera pose (α_p, α_t) (Eq. (49)). The derivation of the angular error vector Δ from Eq. (47) as a function of target position in the image and the camera pose is straightforward. The position command Θ_c is obtained by making

Table 5. Using a spherical coordinate system for active tracking with a pan and tilt perspective camera.

$$\mathbf{\Gamma(X_w)} = \begin{bmatrix} \arctan\left(\frac{X}{Z}\right) \\ \arctan\left(-\frac{Y}{\sqrt{X^2+Z^2}}\right) \\ \sqrt{X^2+Y^2+Z^2} \end{bmatrix}$$

$$\mathbf{T(\Omega)} = \begin{bmatrix} \rho \sin(\phi)\cos(\psi) \\ -\rho \sin(\psi) \\ \rho \cos(\phi)\cos(\psi) \end{bmatrix}$$

$$\mathbf{F(\Omega,\Theta)} = \begin{bmatrix} \frac{C(\psi)S(\phi-\alpha_p)}{C(\psi)C(\alpha_t)C(\phi-\alpha_p)+S(\psi)S(\alpha_t)} \\ -\frac{S(\psi)C(\alpha_t)-C(\psi)S(\alpha_t)C(\phi-\alpha_p)}{C(\psi)C(\alpha_t)C(\phi-\alpha_p)+S(\psi)S(\alpha_t)} \end{bmatrix}$$

$$\mathbf{J_F^\Omega(\Omega,\Theta)} = \begin{bmatrix} \frac{C(\psi)(S(\psi)S(\alpha_t)C(\phi-\alpha_p)+C(\psi)C(\alpha_t))}{\Upsilon_1} & -\frac{S(\alpha_t)S(\phi-\alpha_p)}{\Upsilon_1} & 0 \\ -\frac{C(\psi)S(\psi)S(\phi-\alpha_p)}{\Upsilon_1} & -\frac{C(\phi-\alpha_p)}{\Upsilon_1} & 0 \end{bmatrix}$$

$$\mathbf{J_F^\Theta(\Omega,\Theta)} = \begin{bmatrix} -\frac{C(\psi)(S(\psi)S(\alpha_t)C(\phi-\alpha_p)+C(\psi)C(\alpha_t))}{\Upsilon_1} & \frac{C(\psi)\Upsilon_2}{\Upsilon_1} \\ \frac{C(\psi)S(\psi)S(\phi-\alpha_p)}{\Upsilon_1} & \frac{\Upsilon_3}{\Upsilon_1} \end{bmatrix}$$

$$\Upsilon_1 = C(\psi)^2 C(\alpha_t)^2 S(\psi)^2 (S(\phi)^2 + S(\alpha_p)^2 + 2C(\phi)C(\alpha_p)C(\phi-\alpha_p))$$
$$- C(\alpha_t)^2 + 2S(\alpha_t)C(\alpha_t)S(\psi)C(\psi)C(\phi-\alpha_p)$$

$$\Upsilon_2 = C(\psi)S(\alpha_t)(S(\alpha_p)C(\alpha_p) + 2C(\alpha_p)C(\phi)S(\phi-\alpha_p)$$
$$- S(\phi)C(\phi)) - S(\psi)C(\alpha_t)S(\phi-\alpha_p)$$

$$\Upsilon_3 = 1 - C(\psi)^2(C(\phi)^2 + C(\alpha_p)^2 - 2C(\alpha_p)C(\phi)C(\phi-\alpha_p))$$

Table 6. Position and velocity command for active tracking with a pan and tilt perspective camera.

$$\mathbf{\Theta_c} = \mathbf{\Theta} + \begin{bmatrix} \arctan\left(\frac{x_i}{y_i S(\alpha_t)+C(\alpha_t)}\right) \\ \arctan\left(\frac{\frac{y_i-\tan(\alpha_t)}{1+y_i\tan(\alpha_t)} + \tan(\alpha_t)\sqrt{1+\frac{x_i^2}{(y_i S(\alpha_t)+C(\alpha_t))^2}}}{\frac{y_i-\tan(\alpha_t)}{1+y_i\tan(\alpha_t)}\tan(\alpha_t) - \sqrt{1+\frac{x_i^2}{(y_i S(\alpha_t)+C(\alpha_t))^2}}}\right) \end{bmatrix}$$

$$\mathbf{\dot\Theta_c} = \mathbf{\dot\Theta} + \begin{bmatrix} \frac{1+y_i^2}{(1+x_i^2+y_i^2)(y_i S(\alpha_t)+C(\alpha_t))} & -\frac{x_i y_i}{(1+x_i^2+y_i^2)(y_i S(\alpha_t)+C(\alpha_t))} \\ \frac{x_i(y_i C(\alpha_t)-S(\alpha_t))}{(1+x_i^2+y_i^2)(y_i S(\alpha_t)+C(\alpha_t))} & -\frac{(y_i S(\alpha_t)+(1+x_i^2)C(\alpha_t))}{(1+x_i^2+y_i^2)(y_i S(\alpha_t)+C(\alpha_t))} \end{bmatrix} \mathbf{\dot x_i}$$

$\mathbf{\Theta_c} = \mathbf{\Theta} + \mathbf{\Delta}$ (Table 6).

$$\begin{cases} \phi = \arctan\left(\frac{(y_i\sin(\alpha_t)+\cos(\alpha_t))\sin(\alpha_p)+x_i\cos(\alpha_p)}{(y_i\sin(\alpha_t)+\cos(\alpha_t))\cos(\alpha_p)-x_i\sin(\alpha_p)}\right) \\ \psi = -\arctan\left(\frac{y_i\cos(\alpha_t)-\sin(\alpha_t)}{\sqrt{x_i^2+(y_i\sin(\alpha_t)+\cos(\alpha_t))^2}}\right) \end{cases}$$
$$(49)$$

Consider the jacobian matrices of Table 5. Replacing angles ϕ and ψ by the result of Eq. (49), both $\mathbf{J_F^\Omega}$ and $\mathbf{J_F^\Theta}$ can be written in terms of target position in the image and the camera pose. Equation (50) shows $\mathbf{\hat J_F^\Omega}$ such that $\mathbf{J_F^\Omega} = [\mathbf{\hat J_F^\Omega} \mid \mathbf{0}]$, and Eq. (51) gives $\mathbf{J_F^\Theta}$. Making $\mathbf{\dot x_d} = 0$ in Eq. (41) the velocity command $\mathbf{\dot\Theta}$ is obtained (Table 6).

$$\mathbf{\hat J_F^\Omega}$$
$$= \begin{bmatrix} y_i S(\alpha_t) + (1+x_i^2)C(\alpha_t) & -\frac{x_i S(\alpha_t)(1+x_i^2+y_i^2)}{\sqrt{x_i^2+(y_i S(\alpha_t)+C(\alpha_t))^2}} \\ (y_i C(\alpha_t) - S(\alpha_t))x_i & -\frac{(y_i S(\alpha_t)+C(\alpha_t))(1+x_i^2+y_i^2)}{\sqrt{x_i^2+(y_i S(\alpha_t)+C(\alpha_t))^2}} \end{bmatrix}$$
$$(50)$$

$$\mathbf{J_F^\Theta} = \begin{bmatrix} -(y_i S(\alpha_t)+(1+x_i^2)C(\alpha_t)) & x_i y_i \\ -(y_i C(\alpha_t)-S(\alpha_t))x_i & 1+y_i^2 \end{bmatrix} \quad (51)$$

If $y_i = -\cot(\alpha_t)$ then $y_i\sin(\alpha_t)+\cos(\alpha_t) = 0$ and matrix $\mathbf{J_F^\Theta}$ becomes non-invertible and a singularity occurs in the derived expressions of $\mathbf{\Theta_c}$ and $\mathbf{\dot\Theta_c}$. This happens whenever target image lays in an horizontal line that contains the intersection point of pan rotation axis with image plane. For this case $\delta_p = \pm\pi/2$ and $\tan(\delta_p) = \pm\infty$.
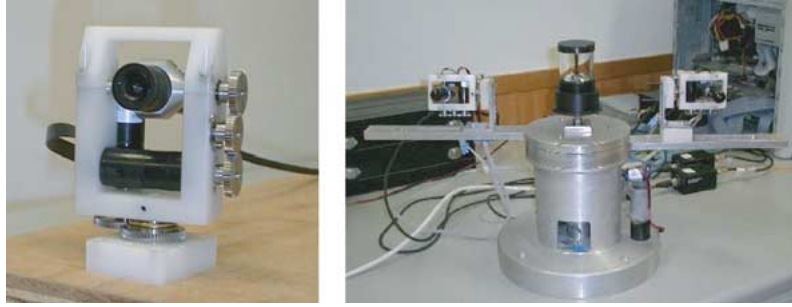
Figure 10. The modular vision system (MVS). Left: The MVS robotic eye. Right: The MVS platform with two robotic eyes and a parabolic camera.

3.5. Tracking Applications Using the MDOF and the MVS Robotic Platforms

In our laboratory two robotic platforms have been developed for active tracking: the MDOF robot head (Batista et al., 1995) and the MVS modular platform. The MDOF head has been mainly used in monocular and binocular tracking of a single target (Batista et al., 1997; Barreto et al., 1999a, 1999b). The MVS head has been recently built with the purpose of working on simultaneous tracking of multiple targets (Barreto et al., 1999c). Figure 10 (Left) depicts one of the MVS robotic eyes. The camera has two degrees of freedom: pan and tilt. The rotation axes go through the optical center and the camera undergoes pure rotation motion. This section describes the application of the equations derived in Section 3.4 for monocular tracking. Two different behaviors must be considered: the saccadic motion and the smooth pursuit (Batista et al., 1995, 1997).

The active vision system is in standby until the target appears in the field of view. Initially the object is projected somewhere in the image periphery. The goal of the saccadic motion is to position the target image in the foveal region. The performance of the saccadic control is highly dependent on the accuracy of the angular error estimation. The importance of using the exact position command of Table 6 to accomplish the task is shown. After a successful saccadic motion the object is projected nearby the image center. The smooth pursuit behavior adjusts the system position such that the target is kept in the image center. We show that for the smooth pursuit the global mapping $\mathbf{F}$ can be approximated by function $\bar{\mathbf{F}}$ which yields simpler mathematical expressions and decoupled control of pan and tilt d.o.f. Function $\bar{\mathbf{F}}$ is established assuming that the target is projected near the image

center during tracking. The approximation errors are studied.

3.5.1. Saccadic Motion. As mentioned, the angular error $\mathbf{\Delta} = (\delta_p, \delta_t)^t$ (Eq. (48)) can be written as a function of the target position in the image and of the camera pose (Table 6). Solving the system of equations with respect to x_i and y_i yields the result of Eq. (52).

$$\begin{cases} x_i = (y_i S(\alpha_t) + C(\alpha_t)) \tan(\delta_p) \\ y_i = -\dfrac{(S(\alpha_t)^2 C(\delta_p) + C(\alpha_t)^2) \tan(\delta_t) + S(\alpha_t) C(\alpha_t)(1 - C(\delta_p))}{(S(\alpha_t)^2 + C(\alpha_t)^2 C(\delta_p)) + S(\alpha_t) C(\alpha_t)(1 - C(\delta_p)) \tan(\delta_t)} \end{cases}$$
$$(52)$$

The saccadic motion consists in rotating the camera such that target image jumps from the periphery to the center of the retina. Table 6 provides the exact position command $\mathbf{\Theta_c}$ to accomplish the task. If command $\mathbf{\Theta_c}$ is sent to pan and tilt motors, then the angular error vector $\mathbf{\Delta}$ becomes null. The saccadic motion is perfect because target projection moves from its initial position on the image periphery to the center (replace δ_p, δ_t by zero in Eq. (52)).

The results presented on Table 6 are novel. In Chroust et al. (2000) the equations to track an object moving in a plane with known depth are derived. The control equations proposed on Table 6 generalize this result since they assume unconstrained target motion. In Sharkey et al. (1993) and Batista et al. (1997) the angular error $\mathbf{\Delta}$ is approximated by $\bar{\mathbf{\Delta}} = (\arctan(x_i), -\arctan(y_i))^t$. This approximation is rather intuitive and simplifies the derivation of the control law. However the performance of the saccadic control tends to be poor as shown in Fig. 11. Consider that the position command is $\bar{\mathbf{\Theta}}_\mathbf{c} = \mathbf{\Theta} + \bar{\mathbf{\Delta}}$. After the saccadic motion, target image moves towards the foveal region since angles δ_p and δ_t decrease. However the
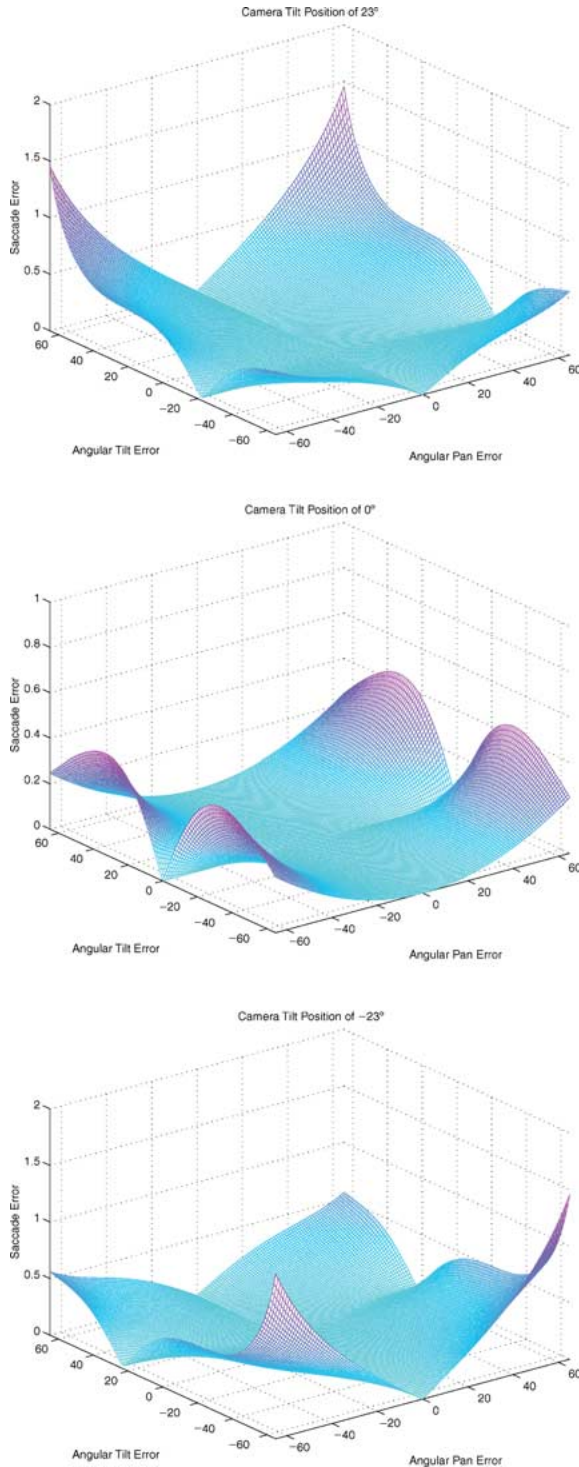
Figure 11. Distance to the image center after the saccadic motion. The position command is $\bar{\Theta}_c = \Theta + \bar{\Delta}$ with $\bar{\Delta} = (\arctan(x_i), -\arctan(y_i))^t$. Figures correspond to camera tilt position of $23°$ (top), $0°$ (middle) and $-23°$ (bottom).

error angles do not become necessarily zero, nor the target is projected in the image center. Figure 11 shows the distance to the image center after the saccadic motion for three different camera tilt positions. The X and Y axes correspond to the initial pan and tilt errors (δ_p, δ_t) which take values in the range $[-60°, 60°]$. The Z axis is the normalized distance to the image center after the saccadic motion. To convert into image pixels the value must be multiplied $k \cdot f$ with k the number of pixels per metric unit and f the camera focal length (in our case $k \cdot f = 225$ for both the MDOF and MVS robot head).

Observing Fig. 11 comes that the performance of the saccadic motion decreases when the angular errors increase. Since the target starts by appearing in the image periphery, both $\boldsymbol{\delta_p}$ and $\boldsymbol{\delta_t}$ take in general high values. One concludes that the performance of the saccadic control using command $\bar{\boldsymbol{\Theta}}_{\mathbf{c}}$ tends to be poor. The approximation $\bar{\boldsymbol{\Delta}} = (\arctan(x_i), -\arctan(y_i))^t$ is only valid for small angular errors and camera tilt angles. The command equation of Table 6 must be used to control the saccadic motion.

3.5.2. Function $\bar{\mathbf{F}}$ for Smooth Pursuit. In general the active tracking process is initialized by the saccadic motion. Camera pose changes abruptly such that target projection jumps from the image periphery to the foveal region. After the saccadic motion the camera orientation is smoothly adjusted to keep the target image in the center of the retina. This stage is called the smooth pursuit control.

The global mapping function $\mathbf{F}$ is shown in Table 5. In the smooth pursuit it is reasonable to assume that most of the time the target projection is near the image center. This assumption is used to derive an approximate mapping $\bar{\mathbf{F}}$. Equation (53) results from making $y_i = 0$ in the first equation of (52) and $\delta_p = 0$ in the second one. Function $\bar{\mathbf{F}}$ is an approximation of the global mapping $\mathbf{F}$ with simpler mathematical expressions and the advantage of decoupling the pan and tilt control.

$$\bar{\mathbf{F}}(\Omega, \Theta) = \begin{bmatrix} \cos(\alpha_t)\tan(\delta_p) \\ -\tan(\delta_t) \end{bmatrix} \quad (53)$$

Consider the approximation error vector $\mathbf{E} = \mathbf{x_i} - \bar{\mathbf{x}_i}$ (Eq. (54)) measured in the image plane. The exact image target position is $\mathbf{x_i} = \mathbf{F}(\Omega, \Theta)$, and $\bar{\mathbf{x}_i} = \bar{\mathbf{F}}(\Omega, \Theta)$ is the approximate one. Error $\mathbf{E}$ depends on the camera

tilt position α_t and on the angular tracking error $\mathbf{\Delta}$.

$$\mathbf{E}(\alpha_t, \mathbf{\Delta}) = (E_x(\alpha_t, \delta_p, \delta_t), E_y(\alpha_t, \delta_p, \delta_t))^t \quad (54)$$

Figure 12 helps to understand the approximation function $\mathbf{\bar{F}}$. Each figure represents the image plane for three different camera tilt positions. If α_t is known then target position in the image depends on the angular tracking error $\mathbf{\Delta}$. Both the exact ($\mathbf{x_i}$) and approximate ($\mathbf{\bar{x}_i}$) target position in the image are calculated for a set of values of $\mathbf{\Delta}$. The values computed of $\mathbf{x_i}$ and $\mathbf{\bar{x}_i}$ are represented in the image plane. The exact positions on the solid grid are approximated by the positions of the dashed grid. Conclusions about $\mathbf{E} = (E_x, E_y)^t$ can be drawn by observing Fig. 12.

Assume that the angular pan error δ_p is constant along time. The target is positioned somewhere in a vertical plane in the 3D world, going through the origin O of the inertial coordinate frame. This plane is projected in a line in the image. If $\alpha_t = 0$, the line is vertical, if $\alpha_t \neq 0$ the line has a slope whose module is inversely proportional to module of camera tilt angle. Using the approximation of Eq. (53) the mentioned plane, containing the target, is projected in a vertical line in the image. As depicted in Fig. 12, $\bar{x}_i = x_i$ whenever $\alpha_t = 0$ or $y = 0$. Some properties of error function $E_x(\alpha_t, \delta_p, \delta_t)$ can be observed:

- $E_x(0, \delta_p, \delta_t) = 0$
- $E_x(\alpha_t, \delta_p, \delta_t) = -E_x(\alpha_t, -\delta_p, \delta_t)$
- $E_x(\alpha_t, \delta_p, \delta_t) = E_x(-\alpha_t, \delta_p, -\delta_t)$

Consider that the angular tilt error δ_t constant. In 3D space the target is somewhere in a cone with vertex in the origin O of the inertial coordinate frame. The image projection of these surfaces are the hyperbolic lines shown in Fig. 12. Whenever $\delta_t = -\alpha_t$ the conic surface degenerates in the OXZ plane. The plane projection is an horizontal line also observable in the figure. The approximation of Eq. (53) generates horizontal lines in the image. Notice that $\bar{y}_i = y_i$ whenever $\delta_t = -\alpha_t$ or $\delta_p = 0$. The following properties of error function $E_y(\alpha_t, \delta_p, \delta_t)$ can be observed:

- $E_y(\alpha_t, 0, \delta_t) = 0$
- $E_y(\alpha_t, \delta_p, -\alpha_t) = 0$
- $E_y(\alpha_t, \delta_p, \delta_t) = E_y(\alpha_t, -\delta_p, \delta_t)$
- $E_y(\alpha_t, \delta_p, \delta_t) = -E_y(-\alpha_t, -\delta_p, -\delta_t)$

A statistical characterization of the angular tracking error $\mathbf{\Delta} = (\delta_p, \delta_t)^t$ has been done experimentally.
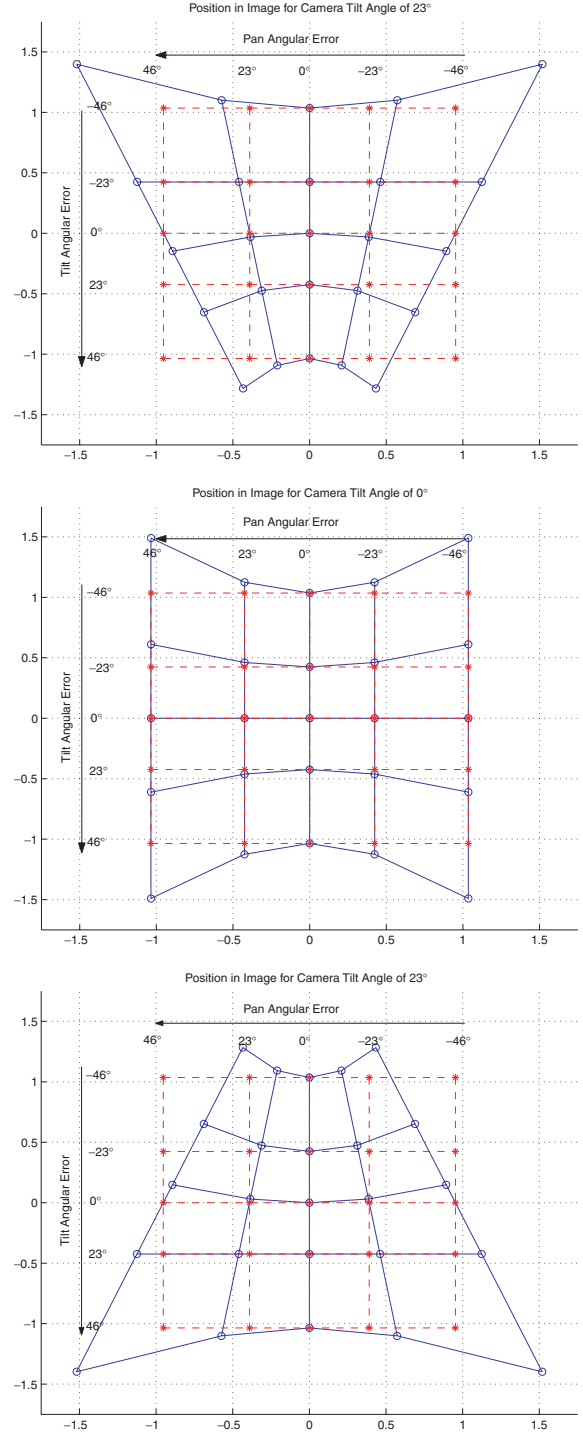

Figure 12. Approximating the global mapping function. Exact (o) and approximate (*) target position in the image are calculated for the set of pairs of values (δ_p, δ_t) with $\delta_p, \delta_t \in \{-46°, -23°, 0°, 23°, 46°\}$. Figures correspond to camera tilt position of 23° (top), 0° (middle) and −23° (bottom).
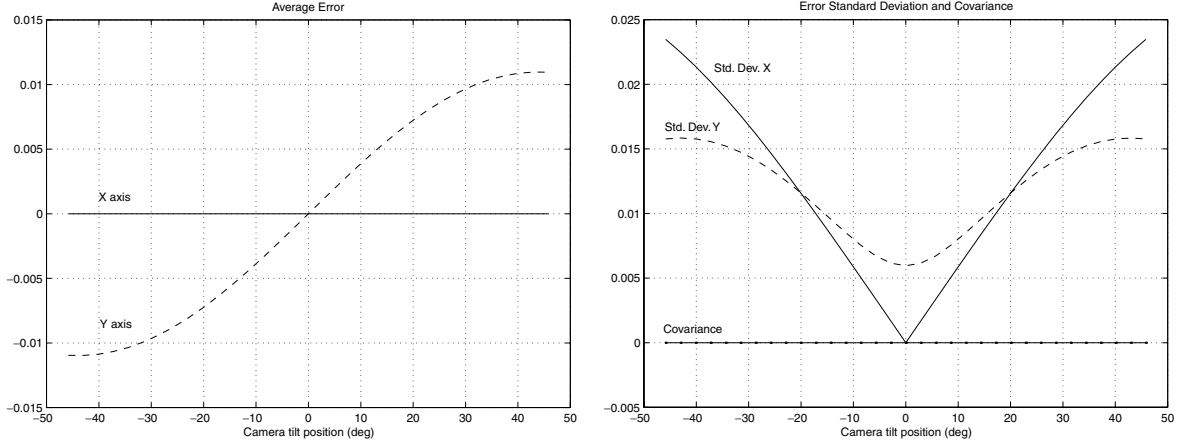
Figure 13. Statistical analysis of approximation error function **E**. Both δ_p and δ_t are assumed to have an independent gaussian probability distribution of average $0°$. The standard deviation of the pan and tilt angular tracking errors are $12°$ and $8°$ respectively. Left: Averages of the approximation error E_x (-) and E_y (- -). Right: The standard deviation of the approximation error E_x (-) and E_y (- -).

Assume that both δ_p and δ_t have a gaussian probability distribution of average $0°$ (the target is almost always near the center of the image). The standard deviation of δ_p is $12°$ and the standard deviation of δ_t is $8°$. The pan and tilt errors are statistically independent thus the covariance is zero. The approximation errors $\mathbf{E_x}$ and $\mathbf{E_y}$ (Eq. (54)) depend on α_t and $\mathbf{\Delta}$. Given the camera tilt position α_t and the statistical characterization of the angular tracking error $\mathbf{\Delta}$, the statistical properties of $\mathbf{E_x}$ and $\mathbf{E_y}$ can be derived. Figure 13 shows the averages μ_x and μ_y, the standard deviations σ_x and σ_y and the covariance σ_{xy} as a function of camera tilt angle α_t. μ_x and σ_{xy} are zero because E_x is an odd function of δ_p. μ_y is an odd function of α_t because $E_y(\alpha_t, \delta_p, \delta_t) = -E_y(-\alpha_t, -\delta_p, -\delta_t)$. The figure presents normalized values for the statistical parameters. To convert into image pixels the value must be multiplied $k \cdot f$ with k the number of pixels per metric unit and f the camera focal length (in our case $k \cdot f = 225$ for both the MDOF and MVS robot head).

The approximation error $\mathbf{E}$ tends to increase with the tracking angular errors $\mathbf{\Delta}$. $\mathbf{E}$ also increases with the absolute value of the camera tilt angle. If $|\alpha_t|$ is high then significant approximation errors arise even when the target is projected near the image center. If we intend to replace the global mapping $\mathbf{F}$ by function $\bar{\mathbf{F}}$ then both $\mathbf{\Delta}$ and α_t must be small, which means that the target image must be near the center and the operating range of tilt d.o.f. can not be large.

3.5.3. Velocity Relationships.
From Eq. (28) it results that the target velocity in the image is the sum $\dot{\mathbf{x}}_i = \dot{\mathbf{x}}_{ind} + \dot{\mathbf{x}}_{ego}$ with $\dot{\mathbf{x}}_{ind} = \mathbf{J}_{\mathbf{F}}^{\Omega}\dot{\Omega}$ the velocity induced by target motion, and $\dot{\mathbf{x}}_{ego} = \mathbf{J}_{\mathbf{F}}^{\Theta}\dot{\Theta}$ the velocity induced by camera motion (egomotion). The first two rows of Fig. 14 depict the image velocity fields $\dot{\mathbf{x}}_{ego}$ when the camera moves in pan and in tilt. Different camera tilt angles are used.

The goal of our tracking application is to keep a zero target image velocity $\mathbf{x_i} = 0$. Considering that $\dot{\mathbf{x}}_i = \dot{\mathbf{x}}_{ind} + \dot{\mathbf{x}}_{ego}$, the camera velocity must be $\dot{\mathbf{\Theta}}_\mathbf{c}$ such that the egomotion compensates for the image velocity component induced by target motion in the scene ($\dot{\mathbf{x}}_{ego} = -\dot{\mathbf{x}}_{ind}$). The velocity command $\dot{\mathbf{\Theta}}_\mathbf{c}$ is written in Eq. (55) as a function of the target velocity in the world (the compactness constraint is verified, therefore $\dot{\rho}$ does not play any role). Notice that if $\dot{\psi} = 0$ then perfect tracking is achieved for $\dot{\mathbf{\Theta}}_\mathbf{c} = (\dot{\phi}, 0)^t$. The image velocity induced by the target motion is compensated for by the camera pan rotation. In general the decoupling between pan and tilt control does not hold as can be seen in the last row of Fig. 14.

$$\dot{\Theta}_\mathbf{c} = -\left(\mathbf{J}_{\mathbf{F}}^{\Theta}\right)^{-1}\hat{\mathbf{J}}_{\mathbf{F}}^{\Omega}\begin{bmatrix} \dot{\phi} \\ \dot{\psi} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \dfrac{x_i(y_i C(\alpha_t) - S(\alpha_t)}{(y_i S(\alpha_t) + C(\alpha_t))\sqrt{x_i^2 + (y_i S(\alpha_t) + C(\alpha_t)^2}} \\ 0 & \dfrac{\sqrt{x_i^2 + (y_i S(\alpha_t) + C(\alpha_t))^2}}{y_i S(\alpha_t) + C(\alpha_t)} \end{bmatrix}\begin{bmatrix} \dot{\phi} \\ \dot{\psi} \end{bmatrix}$$

(55)

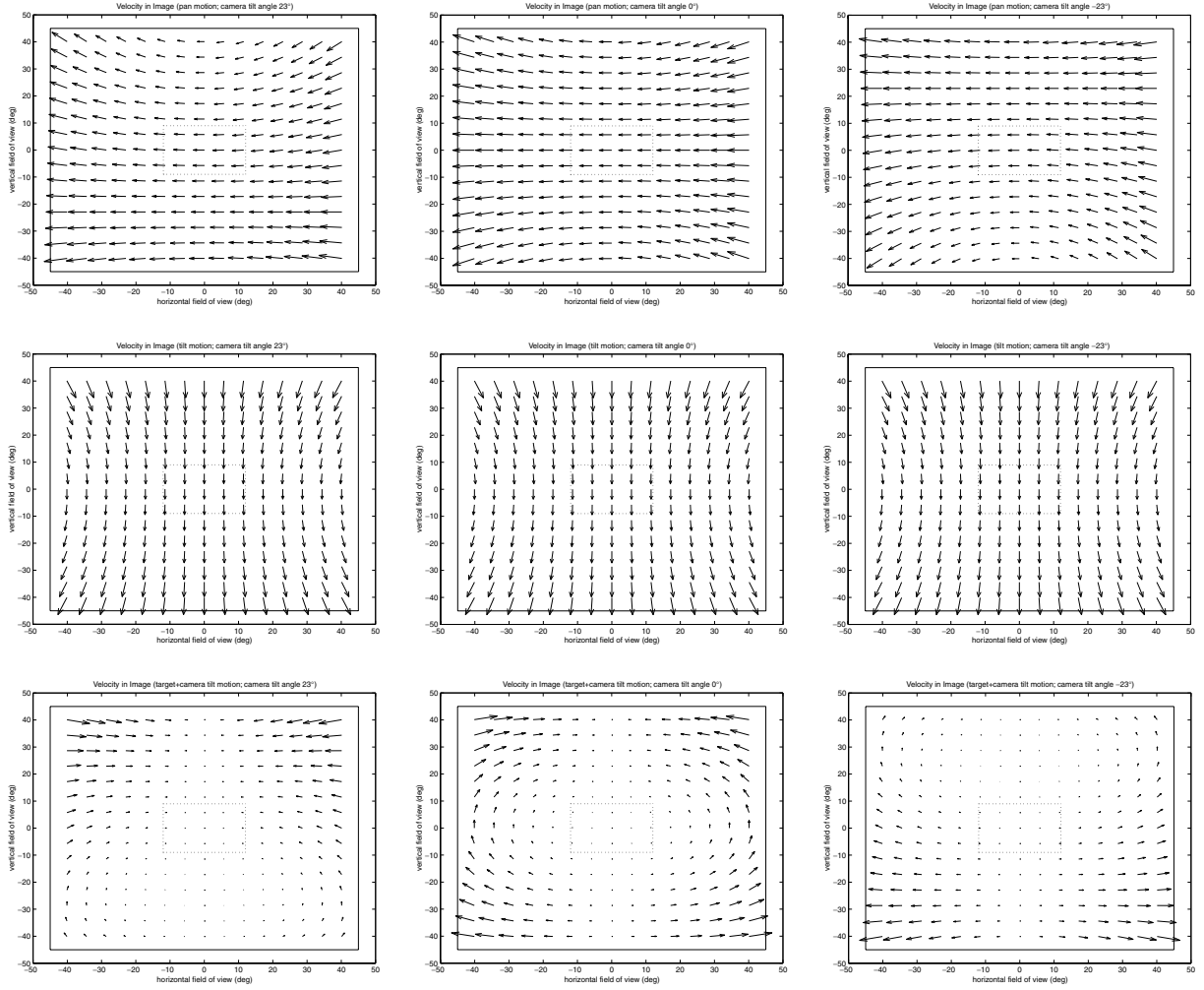*Figure 14.* Image velocity field. First column $\alpha_t = -23°$, second column $\alpha_t = 0°$ and third column $\alpha_t = 23°$. First row: $\dot{\Theta} = (1, 0)^t$; $\dot{\Omega} = (0, 0)^t$ (rad/s). Second row: $\dot{\Theta} = (0, 1)^t$; $\dot{\Omega} = (0, 0)^t$ (rad/s). Third row: $\dot{\Theta} = (0, 1)^t$; $\dot{\Omega} = (0, 1)^t$ (rad/s). Large rectangle (-): image with a field of view of $86° \times 86°$. Small rectangle (:): image with a field of view of $24° \times 18°$.

The jacobian matrices $\hat{\mathbf{J}}_{\mathbf{F}}^{\Omega}$ and $\mathbf{J}_{\mathbf{F}}^{\Theta}$ can be approximated by $\bar{\mathbf{J}}_{\mathbf{F}}^{\Omega}$ and $\bar{\mathbf{J}}_{\mathbf{F}}^{\Theta}$. The result of Eq. (56) is derived assuming in Eqs. (50) and 51 that the target is projected in the center of the image ($\mathbf{x_i} = 0$). The approximated jacobian matrices yield simpler mathematical expressions and decoupled pan and tilt control. The egomotion term is approximated by $\dot{\bar{\mathbf{x}}}_{\mathbf{ego}} = (\cos(\alpha_t)\dot{\alpha}_p, \dot{\alpha}_t)^t$ which is independent of the image position. As can be observed in Fig. 14 the egomotion is only constant around the image center. The approximated velocity command $\dot{\bar{\Theta}}_{\mathbf{c}}$ is derived making $\mathbf{x_i} = 0$ in Eq. (55). As a result $\dot{\bar{\Theta}}_{\mathbf{c}} = (\dot{\bar{\phi}}, \dot{\bar{\psi}})^t$ with the target world velocity approximated by $(\dot{\bar{\phi}}, \dot{\bar{\psi}})^t = (\dot{x}_i / \cos(\alpha_t), \dot{y}_i)^t$. As can be seen in Fig. 14 the approximation yielding decoupled

pan and tilt control is only valid for the central area of the image.

$$\bar{\mathbf{J}}_{\mathbf{F}}^{\Omega} = -\bar{\mathbf{J}}_{\mathbf{F}}^{\Theta} = \begin{bmatrix} \cos(\alpha_t) & 0 \\ 0 & -1 \end{bmatrix} \qquad (56)$$

3.6. *Active Tracking Using an Omnidirectional Camera with a Rotational Degree of Freedom Around the Z Axis*

Figure 15 shows a picture of the MVS vision system. The robotic eyes of Fig. 10 (Left) are mounted on the tips of a rotative platform. We intend to use the omnidirectional camera to control the platform rotation.
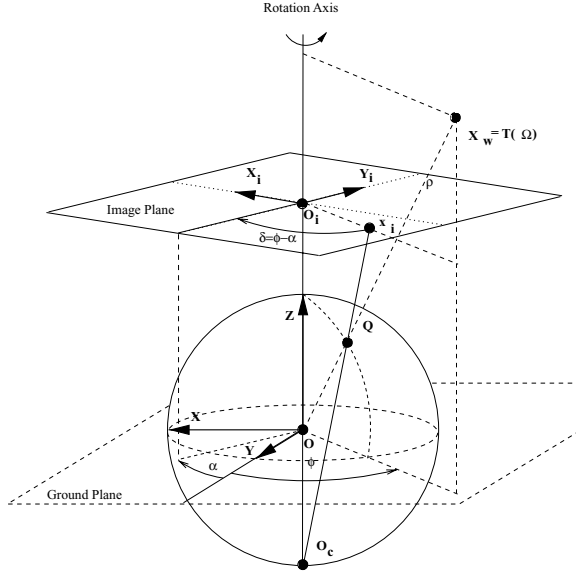
Figure 15. Active tracking using an omnidirectional camera.

The catadioptric system is mounted in the center of the robotic platform (Fig. 10 (Right)). The platform rotation axis goes through the catadioptric effective viewpoint **O** and the re-projection center **O_c**. The goal of the tracking applications is to control the rotation angle α, such that the target image lays in the Y axis of the catadioptric image plane. Figure 15 represents the setup.

Function **f_i** on the global mapping scheme of Fig. 1 is provided in Eq. (18) (our catadioptric system is parabolic thus $l = 1$). **P(Θ)** is given in Eq. (57) where $e^{-\alpha\hat{z}}$ is the matrix rotation around the Z axis (Fig. 15) (Murray et al., 1994). In this case vector **Θ** has 1×1 dimension (**Θ** $= \alpha$). In the sequel vectors **Θ** and **Θ_c** will be replaced by α and α_c to reference the camera pose and command. The goal of the tracking application is to control the camera rotation such that the X target image position and velocity are zero. The desired image target position is $\mathbf{x_d} = (0, y_d)$ and the velocity is $\dot{\mathbf{x}}_d = (0, \dot{y}_d)$ with y_d and $\dot{y}_d$ arbitrary values.

$$\mathbf{P}(\Theta) = e^{-\alpha\hat{\mathbf{Z}}}[\mathbf{I} \mid \mathbf{0}] \tag{57}$$

3.6.1. Selecting Coordinate Transformation Γ.
The procedure explained in Sections 3.3.1 is repeated to select a suitable world coordinate frame to reference the target position.

For this particular case it is possible to choose a coordinate transformation Γ verifying the compactness

constraint. Nevertheless it is not possible to write global mapping **F** in a decoupled form.

The proposed change of coordinates is given in Eq. (58). It is similar to the coordinate transformation used in static catadioptric imaging (Eq. (22)). The only difference is in the ϕ coordinate which is defined according to the camera position parameter α. The projective ray constraining the target is defined by the intersection of a vertical plane, referenced by ϕ, with an horizontal conic surface, indexed by ψ. The proposed coordinate system can not be used to reference points lying on the *XOZ* plane (Fig. 15). We assume that the target is always in front of the MVS system ($Y > 0$). In this case the transformation of Eq. (58) is bijective.

$$\begin{cases} \phi = \arctan\left(-\dfrac{X}{Y}\right) \\[2ex] \psi = \arctan\left(-\dfrac{Y}{Z + l\sqrt{X^2 + Y^2 + Z^2}}\right) \\[2ex] \rho = \sqrt{X^2 + Y^2 + Z^2} \end{cases} \tag{58}$$

Table 7 summarizes the results obtained. The coordinate transformation Γ and its inverse **T** are presented. The global mapping **F** is written using the new world coordinates $\mathbf{\Omega} = (\phi, \psi, \rho)^t$. The corresponding jacobian matrices are presented as well. Notice that the third column of $\mathbf{J_F^{\Omega}}$ is zero, which means that transformation Γ verifies the compactness constraint.

Table 7. New coordinate system for active tracking with an omnidirectional camera.

$$\mathbf{\Gamma}(\mathbf{X_w}) = \begin{bmatrix} \arctan\left(-\frac{X}{Y}\right) \\ \arctan\left(-\frac{Y}{Z+l\sqrt{X^2+Y^2+Z^2}}\right) \\ \sqrt{X^2+Y^2+Z^2} \end{bmatrix}$$

$$\mathbf{T}(\mathbf{\Omega}) = \begin{bmatrix} \frac{\tan(\phi)\tan(\psi)(l+\sqrt{1+(1-l^2)\tan(\psi)^2(1+\tan(\phi)^2)})}{1+\tan(\psi)^2(1+\tan(\phi)^2)} \\ -\frac{(\tan(\psi)(l+\sqrt{1+(1-l^2)\tan(\psi)^2(1+\tan(\phi)^2)}))}{1+\tan(\psi)^2(1+\tan(\phi)^2)} \\ \frac{l(1-\tan(\psi)^2(1+\tan(\phi)^2))+\sqrt{1+(1-l^2)\tan(\psi)^2(1+\tan(\phi)^2)}}{1+\tan(\psi)^2(1+\tan(\phi)^2)} \end{bmatrix}$$

$$\mathbf{F}(\mathbf{\Omega}, \alpha) = (m-l)\cos(\alpha)\tan(\psi)\begin{bmatrix} \tan(\phi) - \tan(\alpha) \\ 1 + \tan(\phi)\tan(\alpha) \end{bmatrix}$$

$$\mathbf{J_F^{\Omega}}(\mathbf{\Omega}, \alpha) = (m-l)\cos(\alpha)\begin{bmatrix} \frac{\tan(\psi)}{\cos(\phi)^2} & \frac{\tan(\phi)-\tan(\alpha)}{\cos(\psi)^2} & 0 \\ \frac{\tan(\psi)\tan(\alpha)}{\cos(\phi)^2} & \frac{1+\tan(\phi)\tan(\alpha)}{\cos(\psi)^2} & 0 \end{bmatrix}$$

$$\mathbf{J_F^{\alpha}}(\mathbf{\Omega}, \alpha) = -(m-l)\cos(\alpha)\tan(\psi)\begin{bmatrix} 1 + \tan(\phi)\tan(\psi) \\ -(\tan(\phi) - \tan(\alpha)) \end{bmatrix}$$

3.6.2. Active Tracking Control Law. Consider the angular tracking error $\delta = \phi - \alpha$. The result of Eq. (59) is derived in a straightforward manner and provides the angular position error δ as a function of target image coordinates.

$$\delta = \arctan\left(\frac{x_i}{y_i}\right) \tag{59}$$

The global mapping function $\mathbf{F}$ is in a compact form, thus it does not depend on ρ. Solving the equation $\mathbf{x_i} = \mathbf{F}(\mathbf{\Omega}, \mathbf{\alpha})$ with respect to $\mathbf{\Omega}$, both ϕ and ψ are obtained as a function of target position in the image and catadioptric system pose. Replacing the result in the two first columns of $\mathbf{J_F^\Omega}$ the compact jacobian $\mathbf{\hat{J}_F^\Omega}$ is obtained as a function of image coordinates and camera pose (Eq. (60)). Matrix $\mathbf{J_F^\alpha}$ of Eq. (62) is derived in the same way. The singularity for $\frac{x_i}{y_i} = \cot(\alpha)$ is similar to the one that appears for the pan and tilt tracking situation.

$$\mathbf{\hat{J}_F^\Omega} = \begin{bmatrix} \frac{x_i^2 + y_i^2}{y_i - x_i \tan(\alpha)} & \frac{((m-l)^2(1+\tan(\alpha)^2) + (y_i - x_i \tan(\alpha))^2)x_i \cos(\alpha)}{(m-l)(y_i - x_i \tan(\alpha))} \\ \frac{\tan(\alpha)(x_i^2 + y_i^2)}{y_i - x_i \tan(\alpha)} & \frac{((m-l)^2(1+\tan(\alpha)^2) + (y_i - x_i \tan(\alpha))^2)y_i \cos(\alpha)}{(m-l)(y_i - x_i \tan(\alpha))} \end{bmatrix} \tag{60}$$

$$\mathbf{J_F^\alpha} = (-y_i, x_i)^t \tag{61}$$

The X coordinate of the target position in image is zero whenever the angular error δ is zero. The position command such that x_i becomes null is $\alpha_c = \alpha + \delta$ with δ provided by Eq. (59).

Consider the target velocity vector $\mathbf{\dot{\hat{\Omega}}} = (\dot{\phi}, \dot{\psi})^t$. Since matrix $\mathbf{\hat{J}_F^\Omega}$ is non-singular, comes that $\mathbf{\dot{\hat{\Omega}}} = (\mathbf{\hat{J}_F^\Omega})^{-1}(\mathbf{\dot{x}_i} - \mathbf{J_F^\alpha}\dot{\alpha})$ (Eqs. (34) and (39)). Assume that $\mathbf{\bar{J}_F^\Omega}$ and $\mathbf{\bar{J}_F^\alpha}$ are the first rows of $\mathbf{\hat{J}_F^\Omega}$ and $\mathbf{J_F^\alpha}$. The goal is to determine the velocity command $\dot{\alpha}_c$ such that the image velocity $\dot{x}_i$, along the X direction, is zero. The desired velocity is $\dot{x}_d = \mathbf{\bar{J}_F^\Omega}\mathbf{\dot{\hat{\Omega}}} + \mathbf{\bar{J}_F^{bm\alpha}}\dot{\alpha}_c$ (Eq. (28)). Making $\dot{x}_d = 0$ and replacing $\mathbf{\dot{\hat{\Omega}}}$ comes that $\dot{\alpha}_c = -(\mathbf{\bar{J}_F^\alpha})^{-1}\mathbf{\bar{J}_F^\Omega}(\mathbf{\hat{J}_F^\Omega})^{-1}(\mathbf{\dot{x}_i} - \mathbf{J_F^\alpha}\dot{\alpha})$. The result of Eq. (62) is achieved taking into account that $\mathbf{\bar{J}_F^\Omega}(\mathbf{\hat{J}_F^\Omega})^{-1} = (1, 0)$.

$$\begin{cases} \alpha_c = \alpha + \arctan\left(\frac{x_i}{y_i}\right) \\ \dot{\alpha}_c = \dot{\alpha} - \frac{\dot{x}_i}{y_i} \end{cases} \tag{62}$$

4.   Conclusions

In this paper we show that the selection of coordinate systems is important in vision applications. In addition we introduce two constraints, the compactness constraint and the decoupling constraint. A coordinate system that verifies one or both of the constraints has several advantages in what concerns the relationships between the position and velocity variables. Coordinate systems defined according to these criteria are also presented for several applications.

References

Baker, S. and Nayar, S. 1998. A theory of catadioptric image formation. In *Proc. of IEEE International Conference on Computer Vision*, Bombay, pp. 35–42.

Barreto, J.P. and Araujo, H. 2001. Issues on the geometry of central catadioptric imaging. In *Proc. of the IEEE Int. Conf. on Computer Vision and Pattern Recognition*, Kauai, Haway, USA.

Barreto, J.P., Batista, J. and Araujo, H. 1999a. Tracking multiple targets in 3d. In *IROS99—IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, Kyongju, Korea, pp. 233–246.

Barreto, J.P., Peixoto, P., Batista, J., and Araujo, H. 1999b. Evaluation of the robustness of visual behaviors through performance characterization. In *Robust Vision for Vision-Based Control of Motion*, Markus Vincze and Gregory D. Hager (Eds.), IEEE Press.

Barreto, J.P., Peixoto, P., Batista, J., and Araujo, H. 1999c. Improving 3d active visual tracking. In *Computer Vision Systems, Lectures Notes in Computer Science*, vol. 1542, Henrik I. Christensen (Ed.), Springer.

Barreto, J.P., Peixoto, P., Batista, J., and Araujo, H. 2000. Mathematical analysis for visual tracking assuming perspective projection. In *SIRS2000—8th Int. Symposium on Intelligent Robotic Systems*, Reading, England, pp. 233–246.

Batista, J., Dias, J., Araujo, H., and Almeida, A. 1995. The isr multi-degree-of-freedom active vision robot head: Design and calibration. In *Proc. of M2VIP'95—Second International Conference on Mechatronics and Machine Vision in Practice*, Hong-Kong.

Batista, J., Peixoto, P., and Araujo, H. 1997. Real-time visual behaviors with a binocular vision system. In *ICRA97—IEEE International Conference on Robotics and Automation*, New Mexico.

Bogner, S. 1995. Introduction to panoramic imaging. In *Proc. of IEEE SMC Conference*, pp. 3100–3106.

Chroust, S., Vincze, M., Traxl, R., and Krautgartner, P. 2000. Evaluation of processing architecture and control law on the performance of vision-based control systems. In *AMC00–IEEE/RSJ Int. Workshop on Advanced Motion Control*, Japan, pp. 19–24.

Corke, P.I. 1996. *Visual Control of Robots: High-Peformance Visual Servoing*. Mechatronics, John Wiley.

Courant, R. 1998. *Differential and Integral Calculus*. Wiley Classic Library.

Geyer, C. and Daniilidis, K. 2000. An unifying theory for central panoramic systems and pratical implications. In *Proc. of European Conference on Computer Vision*, Dublin, pp. 445–461.

Gluckman, J. and Nayar, S. 1998. Egomotion and omnidirectional cameras. In *Proc. of IEEE International Conference on Computer Vision*, Bombay, pp. 999–1005.

Hartley, R. and Zisserman, A. 2000. *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Hecht, E. and Zajac, A. 1974. *Optics*. Addison-Wesley.

Krautgartner, P. and Vincze, M. 1998. Performance evaluation of vision-based control tasks. In *IEEE International Confernce on Robotics and Automation*, Leuven, Belgium.

Laveau, S. and Faugeras, O. 1996. Oriented projective geometry for computer vision. In *Proc. of European Conference on Computer Vision*.

Malis, E., Chaumette, F., and Boudet, S. 1999. 2 1/2 d visual servoing. *IEEE Trans. on Robotics and Automation*, 15(2):238–250.

Murray, R.M., Li, Z., and Sastry, S. 1994. *A Mathematical Introduction to Robotic Manipulation*. CRC Press.

Nalwa, V.S. 1996. A true omnidirectional viewer. Technical Report, Bell Laboratories, Holmdel, NJ, USA.

Sharkey, P., Murray, D., Vandevelde, S., Reid, I., and Mclauchlan, P. 1993. A modular head/eye platform for real-time reactive vision. *Mechatronics*, 3(4):517–535.

Stolfi, J. 1991. *Oriented Projective Geometry*. Academic Press Inc.

Svoboda, T., Padjla, T., and Hlavac, V. 1998. Epipolar geometry for panoramic cameras. In *Proc. of European Conference on Computer Vision*, Freiburg, Germany, pp. 218–332.

Yagi, Y. and Kawato, S. 1990. Panoramic scene analysis with conic projection. In *Proc. of the International Conference on Robots and Systems*.

Yamazawa, K. Yagi, Y., and Yachida, M. 1993. Omnidirectional imaging with hyperboloidal projection. In *Proc. of the International Conference on Robots and Systems*.

Yamazawa, K. and Yagi, Y., and Yachida, M. 1995 Obstacle avoidance with omnidirectional image sensor hyperomni vision. In *Proc. of the IEEE International Conference on Robotics and Automation*, pp. 1062–1067.