Robotica (2011) volume 00, pp. 1–13. © Cambridge University Press 2011
doi:10.1017/S0263574711001068

Catadioptric panoramic stereovision for humanoid robots†

C. Salinas‡, H. Montes‡§*, G. Fernandez¶, P. Gonzalez de Santos‡, and M. Armada‡

‡*Centre for Automation and Robotics – CAR (CSIC-UPM), Robotics Locomotion & Interaction Group, Ctra. de Campo Real. Km 0.200, La Poveda, Arganda del Rey, 28500, Madrid, Spain*
§*Facultad de Ingenieria Electrica, Universidad Tecnológica de Panamá, Republic of Panama*
¶*Departamento de Electronica y Circuitos, Simon Bolivar University, Republic of Venezuela*

(Accepted August 25, 2011)

SUMMARY

This paper proposes a novel design of a reconfigurable humanoid robot head, based on biological likeness of human being so that the humanoid robot could agreeably interact with people in various everyday tasks. The proposed humanoid head has a modular and adaptive structural design and is equipped with three main components: frame, neck motion system and omnidirectional stereovision system modules. The omnidirectional stereovision system module being the last module, a motivating contribution with regard to other computer vision systems implemented in former humanoids, it opens new research possibilities for achieving human-like behaviour. A proposal for a real-time catadioptric stereovision system is presented, including stereo geometry for rectifying the system configuration and depth estimation. The methodology for an initial approach for visual servoing tasks is divided into two phases, first related to the robust detection of moving objects, their depth estimation and position calculation, and second the development of attention-based control strategies. Perception capabilities provided allow the extraction of 3D information from a wide range of visions from uncontrolled dynamic environments, and work results are illustrated through a number of experiments.

KEYWORDS: Catadioptric panoramic cameras; Omnidirectional stereovision; Visual servoing control; Humanoid robot head.

1. Introduction

During last years, several research groups have achieved important advances in humanoid robotics projects.[1–2] At its inception, research efforts were concentrated in the areas of design and construction of biped robots. However, over the years greater emphasis has been given to technological and scientific development, which aims at achieving a higher affinity interaction with human beings, developing humanoid robots with friendly aspects.

The development of humanoid robots that assist human activities in environments such as offices, homes, shops and hospitals is expected. Humanoid robots are called upon to perform these tasks for serving humans because of their anthropomorphic structure, friendly design, locomotion way and so forth. Human society demands the incorporation of new applications for robots to perform various tasks of service, assistance, entertainment[3] and so on. In these tasks, robots will be required to interact with the modifiable environment and surrounded by people.

Most works on humanoid robots, up to date, have concentrated on studying the locomotion problem. In this direction great advances have been achieved and humanoid robots like ASIMO, HRP-2, the QRIO, Johnnie,[1–4] among others, can be mentioned.

The vision system (based on conventional cameras) on these robots in combination with other sensors and appropriate control strategies are used to improve the process of robot movement.[8–11]

By other side, during last two decades interest in panoramic vision systems has grown and their use in the robotics field has gained importance due to technological advances and the increasing need to track and detect objects over large 3D environments. However, the incorporation of panoramic vision into the field of humanoid robots is not yet broadly used. In the case of omnidirectional vision systems, since the Ress' first proposal in US Patent No. 3505465 in 1970 [12] and later on when these systems started to develop again,[13–15] several configurations and theories of catadioptric panoramic system have been presented in order to obtain images of the entire scene.[16–18]

Compared with conventional systems, their greatest advantage is to acquire a wide range of view images, which allow robotic systems to become more suitable for tasks such as navigation, tracking objects and ego-motion detection, since the objects disappear later on the images.

It is common to use rotating, multiple cameras, or catadioptric systems to obtain images of the entire scene. However, the first approach brings in mechanical problems, as the movement of heavy parts, the manufacture costs and the rotation mechanisms are not always suitable for real-time applications, and also to achieve accurate positioning

* Corresponding author. E-mail: hector.montes@csic.es
† This paper was originally submitted under the auspices of the CLAWAR Association. It is an extension of work presented at CLAWAR 2009: The 12th International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines, Istanbul, Turkey.

it requires extra efforts. Multiple cameras present a high computing cost to form a single panoramic image. On the other hand, the catadioptric systems as the result of combination of a refracting (*dioptric*) and reflecting (*catoptrics*) surfaces are considered to be a very interesting configuration. These systems are easily built employing a conventional high-resolution camera as the refracting part and a curved mirror as the reflecting one. In order to acquire a single image containing the information of the whole scene, the camera and mirror must be arranged in a configuration such that the entire system has a *single effective viewpoint*,[18] named as *central catadioptric cameras*.[19] To generate omnidirectional images perfect quadric surfaces are considered as the only candidates for mirror shapes; in this way every incident ray of light that strikes the surface towards the mirror focus is reflected to the second focus. Since the geometry of the system is known, it is possible to compute the ray direction for each pixel and its irradiance value.

Several configurations of stereo system have been presented: The general theory of epipolar geometry for central catadioptric stereo cameras is depicted in ref. [19], a rectified systems is given in ref. [20], where two omnidirectional systems were placed one on top of the other, vertically aligned, the special mirror as a double lobbed in ref. [21], and a series of pairs of distinct curved mirrors with a single camera were proposed by Nene and Nayar.[22] Since our interest lies in high resolution systems, the last two configurations are to be avoided. A rectified system is desirable as it allows simplifying the process of disparity extraction, since the epipolar lines correspond to the radial axis of the omnidirectional image, even though the dimensions of this system for humanoid robot head are too large and not proportioned for a normal size human head.

The Robotics Locomotion and Interaction Group of the Centre for Automation and Robotics of the CSIC-UPM has been conducting research in the field of humanoid robotics. Currently this centre has a humanoid robot prototype called SILO2.[23,24] Also, this group has proposed an initial design of a multi-sensor humanoid head with omnidirectional vision system.[25]

This paper presents the extension and improvement of the first design of the humanoid head. In this extension, the work has been focused on a modular, flexible and adaptable design of the structure of the humanoid head, with which it is possible to experiment a range of actions and mechanisms of motion for the neck, and with various omnidirectional vision systems. Single omnidirectional or omnidirectional stereovision systems can be used, without these changes affecting the harmony of head design. The omnidirectional vision system will provide an extra sense to the robotic system, and at the same time it can bring in a substantial difference with relation to other humanoid robots that presently exist. The omnidirectional stereovision system presented in this work consists of two catadioptric panoramic systems aligned and separated by a constant horizontal distance, and displaced along the vertical axis at a preset distance.

The control system to develop the proposed strategies in this approach has a hierarchical architecture. The hardware consists of an industrial PC Intel Core 2 Duo based processor E4500, 2.2 GHz for the omnidirectional vision system with Windows XP OS, and a Master/Server single board computer Pentium III, 700 MHz+, with QNX 6.4 RTOS, for controlling the installed servomotors in the neck of the humanoid head by means of three PID/slave processors. In addition, the system has a server–client architecture for servoing commands between both PCs.

This paper is divided into six sections. Section 1 introduces the topic of humanoid robotics and the motivation of this work, taking into account the humanoid head design, vision systems in humanoid robots and a brief review of omnidirectional vision systems. In Section 2 design features of the humanoid head are presented. Next, in Section 3 the description of a designed omnidirectional catadioptric system as well as a brief review of hyperbolic surface geometry, the system resolution and the corresponding panoramic transformation of our homemade specially designed mirrors is carried out. Our proposal for a real-time catadioptric stereovision system is presented in Section 4, including stereo geometry for rectifying the system configuration and depth estimation. The experimental stage focused to achieve human-like behaviours (humanoid head attitudes) is divided into two phases, first phase being the depth estimation and position calculation of the moving objects, and the second one being the development of attention-based control strategies. Finally, in last section conclusions and contributions of this work are presented.

2. Design of Humanoid Head

In the design of a humanoid head several aspects must be considered, the anthropometrical aspects being the most important ones, which refer to the study of physical dimensions of human body, and ensure that the humanoid head has similar proportions to its biological simile. Therefore, it is necessary to evaluate the parameters of human dimensions to achieve a humanoid-appropriate configuration.[26] It must also be taking into consideration the characteristics of human movement, in this case the neck movements. It is also important that the structural design of the humanoid head has the ability for sensors and actuators to be carried inside.

Moreover, it is necessary to consider issues related to the sensory system, in this case an omnidirectional stereovision system, allowing the acquisition of a wide range of views of approximate 360°.[14,15]

For the humanoid head of this work, the chosen vision system is formed by catadioptric systems. These vision systems can acquire a range of visions of $\sim 360^\circ$ with the capture of a single image. Since the geometry of the mirror is known, the rectification of the omnidirectional image to panoramic one is possible.

On the other hand, the dimensions of the system could be modified by means of adjustment of the mathematical model of the mirror surface if the profile of the mirror is changed.

2.1. Anthropometric and kinematic considerations

For the integration of robots in human society, it is essential to know the shapes and symmetries of human beings (as far as possible). So, humanoid robots must be designed with

Table I. Range of movement of the human neck.

Vertebrae	Pitch (°)	Yaw (°)	Roll (°)
Upper cervical (C1–C2)	40	77	13
Lower cervical (C2–C7)	82	55	67
Total range	122	132	80

dimensions and characteristics similar to humans. These concepts make it easy for humans "to accept" the humanoid robots in their environment, e.g. offices, homes, shops, exhibition spaces, hospitals. Montes *et al.*[25] describe the most important parts of human skull used as a biological simile for the design of the structure of the proposed humanoid head.

Similarly, it is required that both the positions and the transitions between movements of robotic devices are as soft and natural as possible. The study of neck movements have enabled the realisation of the design of mechanisms to execute movements similar to those of human beings.

In the case of human body, the head movement is achieved by the combination of cervical vertebrae and neck muscles. The human neck has three degrees of freedom, which are presented in the movement of pitch, roll and yaw. Table I details the range of neck motion.[27]

2.2. Description of humanoid head

The projected humanoid head is designed as a flexible and adaptable system that consists of three modules. The first module, the frame, is structural and it has the shape of human head. The second module corresponds to the neck motion system (motive module), consisting of mechanical elements and electrical actuators. The actuators are three DC motors with incremental encoders and gearboxes; two of them are assembled around a differential gear to perform pitch and roll movements, and the other actuator supports the humanoid head and it performs the yaw movement. Omnidirectional vision system is located in the third module, which can be reconfigured to "single-omnidirectional" or "stereo-omnidirectional" system. Figure 1(a) shows the perspective view of the humanoid head design and its main parts. Figure 1(b) shows the connecting points between different modules. It is possible to see in Fig. 1(b) that each module can operate independently and can be replaced by other mechanisms (in the case of neck). These changes would not affect the modules of frame and vision.

The external module or frame consists of three parts, two of them form the lower part of the humanoid head and correspond to the set comprising occipital, temporal and maxilla, and the third part consists of the frontal–parietal zone; both parts have been described by Montes *et al.*[25]

The module of the neck motion corresponds to the motion module of the humanoid head. The pieces of this set are joined to the bottom of the vision system module (see Fig. 1(b)). This module includes a mechanism that allows the performance of movements similar to those of human movements (see Fig. 2). In addition, the flexibility of the humanoid head design makes it possible to attach another actuation mechanism.

Figure 1(c) shows the components of the motion module of the neck. This module consists of a differential mechanism,

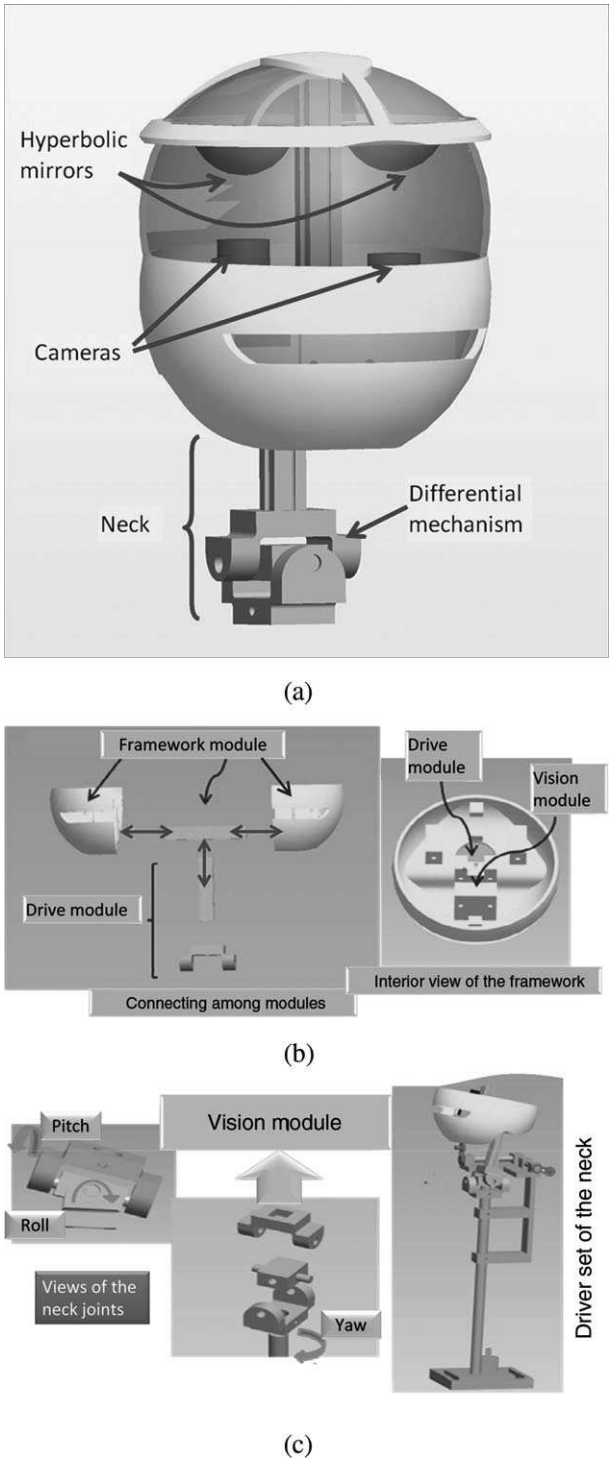

Fig. 1. (Colour online) (a) View of the humanoid head; (b) connection points among the three modules of the humanoid head; (c) motion module of the neck.

whose axes are connected to motors that carry out the respective movements of pitch and roll. The yaw movement is achieved through a vertical shaft attached to the top of the differential mechanism. The motor of this shaft is installed at the base of the neck of the head. The motors connected to the differential mechanism are of same mechanical and electric characteristics so that movements of pitch and roll are carried out properly. These motors are 24 VDC, with similar gearbox and differential encoders of 500 pulses

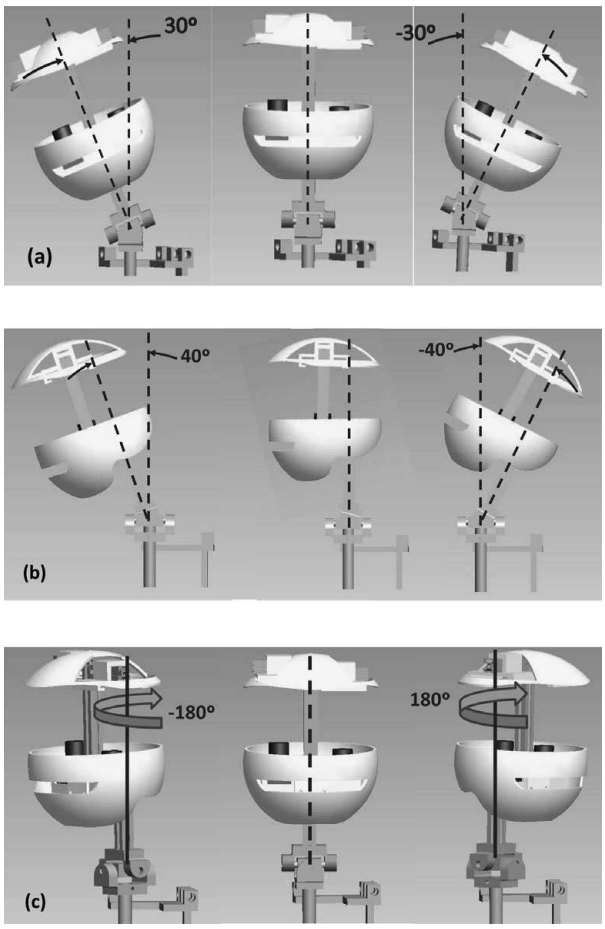

Fig. 2. Simulation of neck movement: (a) roll movement; (b) pitch movement and (c) yaw movement.

per revolution (using the appropriate hardware, it gets an accuracy of about 0.0013°). The motor that performs the yaw movement has similar characteristics as the other two, but has less power and gearing because the gravitational effects are almost insignificant.

In this first prototype of the neck mechanism, the movements of pitch and roll have ranges of angular displacement lower than those of human beings (see Table I in Section 2.1). The range of angular displacement for the movement of pitch is $-40^\circ \leq p1 \leq 40^\circ$ and for the movement of roll it is $-30^\circ \leq r1 \leq 30^\circ$, and the yaw movement is $-90^\circ \leq y1 \leq 90^\circ$. The first two movements are restricted by the mechanical system, and the control algorithm restricts the yaw movement. These movements have been outlined in Fig. 2.

The adaptive design of the humanoid head and neck structure that supports the head has been very useful for carrying out the experiments presented in this work. Since the design of the humanoid head presented in this paper has a modular architecture (as described above), it is possible to perform various experiments with different configurations of the omnidirectional vision. The shape and dimensions of the humanoid head can be changed without affecting the vision system. The interior structure of the head set allows the implementation of different vision systems (different hyperbolic mirrors, cameras, etc) without affecting the overall design of the head. Other adaptive characteristic of the humanoid head is its capability for implementing diverse mechanisms of neck. The entire set of the head and vision system is supported by the mechanism of neck. This set can be coupled with any other possible mechanism used as neck.

3. Designing a Catadioptric Omnidirectional System

The catadioptric panoramic systems, also named as omnidirectional systems because of their enclosed information captured from a scene in all possible directions around an effective viewpoint, are based on a combination of conventional cameras and rotationally symmetric quadric mirrors, where the optical axis and the symmetric mirror axis are aligned. The theory of central perspective projections for a catadioptric image formation has been detailed in refs. [17, 18], where a collection of specific mirror shapes is analysed to achieve a single effective viewpoint that allows the construction of the perspective and panoramic images. Since the image formation is a well-controlled process, it is easy to derive its geometrical properties.

In order to present the geometry of image formation for central catadioptric cameras, the notation of points in the 3D space is chosen to be represented by bold upper case letter, such as $\mathbf{X}$, and its corresponding coordinates by italic upper case letter, such as X. For the representation of points in a 2D space, a bold lower case letter is used, and for its corresponding coordinates are represented by italic lower case letters. The same notation is used for 2D or 3D vectors and planes.

The vision systems designed in our work are based on hyperbolic mirrors. This shape is the solution of quadric surfaces that provides a central perspective projection where one of its two foci is fixed at the pinhole camera, $\mathbf{F^{ii}}$ and the other at the viewpoint, $\mathbf{F^{i}}$. The initial restrictions of our omnidirectional system prototype are imposed by the previously acquired high-resolution cameras. The camera model is the Ueye UI-1485LE-C/M, colour RGB, with resolution of 2560×1920, sensor size equals to $^1/_2$", 6 fps, 2.2 μm pixel pitch.

The geometry of a hyperbolic catadioptric system is described by means of mirror surface ($\mathbf{M} \in \Re^3$), an arbitrary 3D point in the world space $\mathbf{X_W}$ and the intersected point of the light ray of $\mathbf{X_W}$ towards $\mathbf{F^i}$ at the mirror surface $\mathbf{X_M}$. Let the Cartesian coordinates' origin is denoted as $\mathbf{O_W}$, the distance between two foci as c, the distance from $\mathbf{F^{ii}}$ to the images of plane is designated as the focal length f, the projection of the refracted ray that passes through it into the image plane ($\mathbf{I} \in \Re^2$) as $\mathbf{u_i} = (u_i, v_i)$, the azimuthal radius as $\mathbf{r_M} = (X, Y)$ (See fig. 3) if $\mathbf{O_W}$ is placed on the middle of c, then the equations of the hyperbolic system are expressed by the following relations and its two foci are $\mathbf{F^i} = (0, 0, c/2)$ and $\mathbf{F^{ii}} = (0, 0, -c/2)$:

$$\frac{\left(Z - \frac{c}{2}\right)^2}{a^2} - \frac{Y^2 + X^2}{b^2} = 1, \tag{1}$$

$$\|\mathbf{r}\| = \sqrt{X^2 + Y^2}. \tag{2}$$

Since the perspective projection is rotationally symmetric about $\hat{\mathbf{z}}$-axis, the problem can be restricted to the $\hat{\mathbf{z}}\hat{\mathbf{r}}$-plane and the mirror shape as a profile in the 2D plane. The

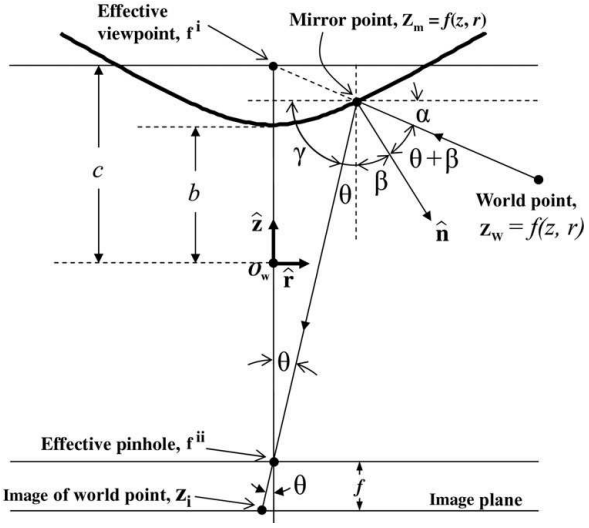

Fig. 3. Hyperbolic catadioptric system geometry.

problem consists in finding an appropriate mirror profile that suits the camera restrictions such as depth of field, working area and, last but not less important, the minimum focus distance. The geometry used to derive the fixed viewpoints as mentioned above has been presented in refs. [17, 18]. As it is well known in the field of perspective cameras' geometry, the relationship between a point in a 3D space ($\vec{\mathbf{X}}_\mathbf{M} = [X_M, Y_M, Z_M, 1]^T$) and its projection onto the image frame ($\vec{\mathbf{u}}_\mathbf{i} = [u_i, v_i, 1]^T$), both expressed in homogeneous coordinates, can be associated according to the equation $\vec{\mathbf{u}}_\mathbf{i} = \mathbf{K\Pi}\vec{\mathbf{x}}_M$, where $\mathbf{K}$ and $\mathbf{\Pi}$ are the intrinsic and extrinsic camera parameter matrices.

Figure 3 illustrates the hyperbolic catadioptric system geometry in the 2D Cartesian coordinates. As in 3D representation, the two foci are aligned along $\hat{\mathbf{z}}$-axis, where $\mathbf{f^i} = (0, c/2)$ and $\mathbf{f^{ii}} = (0, -c/2)$. The mirror profile is a function $z(r)$, where r comes from Eq. (2), the arbitrary world point is $\mathbf{z_w} = (z_w, r_w)$, the intersection of the incoming light ray at the mirror surface is $\mathbf{z_m} = (z_m, r_m)$ and finally the point where the refracting light ray intersects the image plane is denoted by $\mathbf{z_i} = (z_i, r_i)$. The angle θ is the vertical angle of the camera and its complementary angle is γ, and α is the angle between $\hat{\mathbf{r}}$-axis and the incoming light ray from $\mathbf{z_w}$. Subsequently, β is the angle between $\hat{\mathbf{z}}$-axis and the normal $\hat{\mathbf{n}}$ to $\mathbf{z_m}$, therefore the slope at this point is

$$\frac{dz}{dr} = -\tan\beta. \tag{3}$$

Finally, the vertical angle of the catadioptric system is $\phi = 2\beta + \theta$. Next, the following relationships can be deduced:

$$\theta = 90^\circ - \gamma, \tag{4}$$

$$180^\circ = \gamma + 2\theta + 2\beta + \alpha. \tag{5}$$

Substituting Eq. (4) in Eq. (5):

$$2\beta = \gamma - \alpha. \tag{6}$$

Taking the tangent on Eq. (6) and using standard trigonometry relations, such as $\tan(2\beta)$, $\tan(\gamma - \alpha)$ and the slope (3), the resultant equation is

$$4rz\left(\frac{dz}{dr}\right)^2 - \left(4r^2 + c^2 - 8z^2\right)\left(\frac{dz}{dr}\right) - 4rz = 0. \tag{7}$$

The resolution of the catadioptric system must be computed to avoid the degeneration of the geometric relation among the points in a 3D space, the information observed by $\mathbf{f^i}$, and its projection in an open disk. This fact is important to obtain correct panoramic and perspective image transformations. The resolution can be defined as the relation between an infinitesimal area on the image dP and its corresponding solid angle of the world dw; detailed description of the method was presented by Benosman and Kang in ref. [28] . Due to geometrical properties of the hyperbola it is possible to derive the solution by consecutive relations between dP and an infinitesimal area on the mirror surface.

Several simulations have been performed to solve the first-order differential Eq. (7) and consequently to find an appropriate profile that suits the camera parameters, the mean dimensions of a human head and the ability of acquiring a wide vertical field of view. We have performed a simulation of the mirror profile and the resulting hyperbolic mirror designed and manufactured (using CNC micro-mechanisation facilities at the Centre for Automation and Robotics – CSIC-UPM) is presented in Fig. 4. The resulting omnidirectional images acquired by the developed system, and the rectified panoramic as well as its cylindrical representation are shown in Fig. 5.

Other solutions for Eq. (7) can be found in refs [29, 30], where restrictions such as a logarithmic sensor or the proposal for constant resolution cameras, are suggested.

4. Catadioptric Panoramic Stereovision System

The stereovision problem for omnidirectional systems is analogous as for conventional cameras are concerned. The epipolar geometry has been studied thoroughly in ref. [19], which describes the relationship of corresponding 3D points between a pair of images by means of epipolar lines that for the case of catadioptric systems are curved. To present the geometry of our proposed omnidirectional stereo system, the initial step is to consider two catadioptric cameras: *1-omnivision* and *2-omnivision* with an already known geometry (i.e. see Fig. 3). In order to simplify the notation, only the variables related to the mirror and world frame will be retained because their projection onto the image plane can be controlled and *vice versa* from the image plane data onto the mirror frame (please refer to Section 3). The catadioptric systems are positioned in such way that their viewpoints are horizontally aligned, their local $\hat{\mathbf{z}}$-axes are parallel to each other and D_H is the distance between them. The catadioptric system is depicted in Fig. 6, where $\tilde{\mathbf{X}}_\mathbf{W}$ is taken as an arbitrary point in the 3D space and its reflecting light rays at both mirrors' surfaces are $\tilde{\mathbf{X}}^1_\mathbf{M} = (X^1_M, Y^1_M, Z^1_M)$ and $\tilde{\mathbf{X}}^2_\mathbf{M} = (X^2_M, Y^2_M, Z^2_M)$, respectively. Depending on sensors geometry ($c_1 \leq hor c_1 \geq h$), c_1 can take different values.

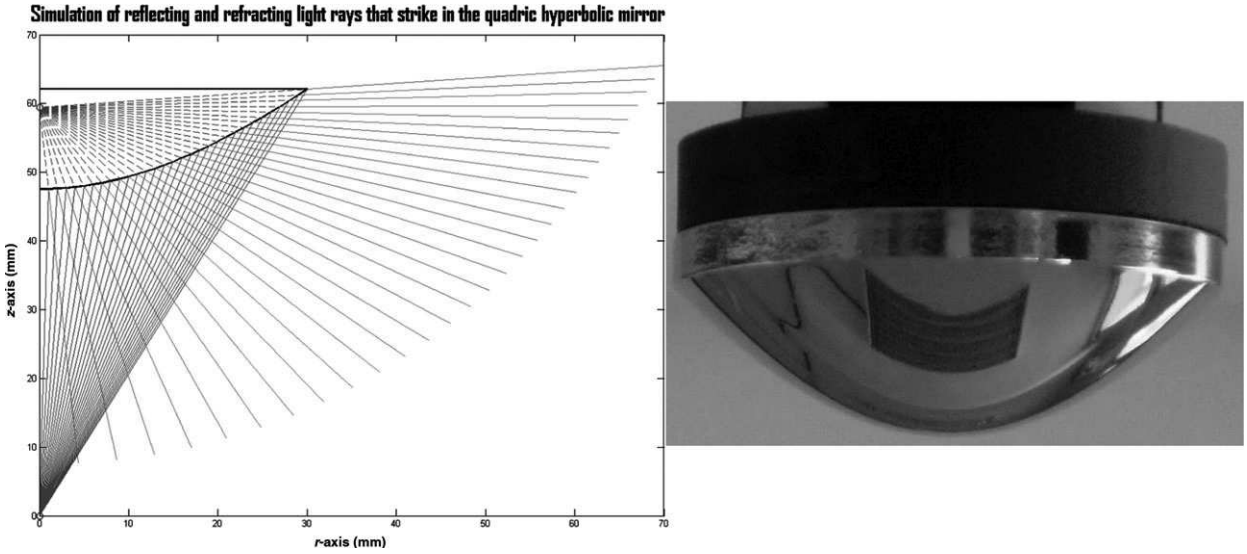

Fig. 4. (Colour online) Hyperbolic mirror. (a) Reflecting ray lights that strike the mirror surface; (b) real image of the specially manufactured mirrors by CNC micro-mechanisation.

Assume that the origin $\tilde{\mathbf{O}}_{\mathbf{W}}$ of the 3D coordinate system is located at $\tilde{\mathbf{F}}_2^{\mathbf{ii}}$ (2-*omnivision*).

Noticing that variables in the world frame $\tilde{\mathbf{O}}_{\mathbf{W}}$ are like "$\tilde{\mathbf{A}}$," the coordinates in the local frame $\mathbf{O}_{\mathbf{W}}$ are transformed to $\tilde{\mathbf{O}}_{\mathbf{W}}$ by $\tilde{\mathbf{A}} = {}^{\tilde{\mathbf{O}}\mathbf{w}}\mathbf{T}_{\mathbf{Ow}}\mathbf{A}$. The plane conformed by $\overline{\tilde{\mathbf{F}}_1^{\mathbf{i}}\tilde{\mathbf{F}}_2^{\mathbf{i}}} \wedge \overline{\tilde{\mathbf{F}}_1^{\mathbf{i}}\tilde{\mathbf{X}}_{\mathbf{W}}}$ *or* $\overline{\tilde{\mathbf{F}}_1^{\mathbf{i}}\tilde{\mathbf{F}}_2^{\mathbf{i}}} \wedge \overline{\tilde{\mathbf{F}}_2^{\mathbf{i}}\tilde{\mathbf{X}}_{\mathbf{W}}} \Rightarrow \hat{\mathbf{n}}_{\Pi}$ is denoted as $\mathbf{\Pi}$ and its normal is denoted as $\hat{\mathbf{n}}_{\Pi}$. Fig. 3 illustrates the geometry of the sensors. The positions of the foci are $\tilde{\mathbf{F}}_1^{\mathbf{i}} = (D_H, 0, h)$, $\tilde{\mathbf{F}}_2^{\mathbf{i}} = (0, 0, h)$, $\tilde{\mathbf{F}}_1^{\mathbf{ii}} = (D_H, 0, h - c_1)$ and $\tilde{\mathbf{F}}_2^{\mathbf{ii}} = (0, 0, 0)$. The baseline $\overline{\tilde{\mathbf{F}}_1^{\mathbf{i}}\tilde{\mathbf{F}}_2^{\mathbf{i}}}$ is parallel to $\hat{\mathbf{x}}$-axis, so equation for $\mathbf{\Pi}$ is derived as follows:

$$\overline{\tilde{\mathbf{F}}_1^{\mathbf{i}}\tilde{\mathbf{X}}_{\mathbf{M}}^{\mathbf{1}}} \Rightarrow \tilde{\mathbf{X}}_{\mathbf{F}}^{\mathbf{M}} = \left(X_M^1 - D_H, Y_M^1, Z_M^1 - h\right), \tag{8}$$

$$\overline{\tilde{\mathbf{F}}_2^{\mathbf{i}}\tilde{\mathbf{F}}_1^{\mathbf{i}}} \Rightarrow \tilde{\mathbf{B}}_2^{\mathbf{1}} = (-D_H, 0, 0), \tag{9}$$

$$\tilde{\mathbf{n}}_{\Pi} = \tilde{\mathbf{X}}_{\mathbf{F}}^{\mathbf{M}} \times \tilde{\mathbf{B}}_2^1 \Rightarrow \Pi : -D_H\left(Z_M^1 - h\right)Y + \left(D_H Y_M^1\right)Z. \tag{10}$$

Let us suppose a third camera 3-*omnivision* is introduced, vertically aligned to 2-*omnivision* (see Fig. 7), their foci belonging to $\hat{\mathbf{z}}$-axis and D_V is the distance between them. The viewpoint of 3-*omnivision* is $\tilde{\mathbf{F}}_3^{\mathbf{i}}$, and the point at the mirror surface coming from $\tilde{\mathbf{X}}_{\mathbf{W}}$ is $\tilde{\mathbf{X}}_{\mathbf{M}}^{\mathbf{3}}$. The geometry between 3–2-*omnivision* is defined by the baseline $\overline{\tilde{\mathbf{F}}_2^{\mathbf{i}}\tilde{\mathbf{F}}_3^{\mathbf{i}}}$, in this particular case the epipolar curves are radial lines. The plane containing $\overline{\tilde{\mathbf{F}}_2^{\mathbf{i}}\tilde{\mathbf{F}}_3^{\mathbf{i}}} \wedge \overline{\tilde{\mathbf{F}}_2^{\mathbf{i}}\tilde{\mathbf{X}}_{\mathbf{W}}}$ *or* $\overline{\tilde{\mathbf{F}}_2^{\mathbf{i}}\tilde{\mathbf{F}}_3^{\mathbf{i}}} \wedge \overline{\tilde{\mathbf{F}}_3^{\mathbf{i}}\tilde{\mathbf{X}}_{\mathbf{W}}} \Rightarrow \hat{\mathbf{n}}_{\Omega}$ is $\mathbf{\Omega}$-plane, $\tilde{\mathbf{F}}_3^{\mathbf{i}} = (0, 0, h + D_V)$ and $\tilde{\mathbf{F}}_3^{\mathbf{ii}} = (0, 0, h + D_V - c_3)$ are the foci of 3-*omnivision* and the equation of $\mathbf{\Omega}$-plane is

$$\overline{\tilde{\mathbf{F}}_3^{\mathbf{i}}\tilde{\mathbf{X}}_{\mathbf{M}}^{\mathbf{3}}} \Rightarrow \tilde{\mathbf{X}}_{\mathbf{F}}^{\mathbf{M_3}} = \left(X_M^3, Y_M^3, Z_M^3 - (h + D_V)\right), \tag{11}$$

$$\overline{\tilde{\mathbf{F}}_3^{\mathbf{i}}\tilde{\mathbf{F}}_2^{\mathbf{i}}} \Rightarrow \tilde{\mathbf{B}}_3^{\mathbf{2}} = (0, 0, -D_V), \tag{12}$$

$$\tilde{\mathbf{n}}_{\Omega} = \tilde{\mathbf{X}}_{\mathbf{F}}^{\mathbf{M_3}} \times \tilde{\mathbf{B}}_3^2 \Rightarrow \Omega : \left(-D_V X_M^2\right)X + \left(D_V X_M^3\right)Y = 0. \tag{13}$$

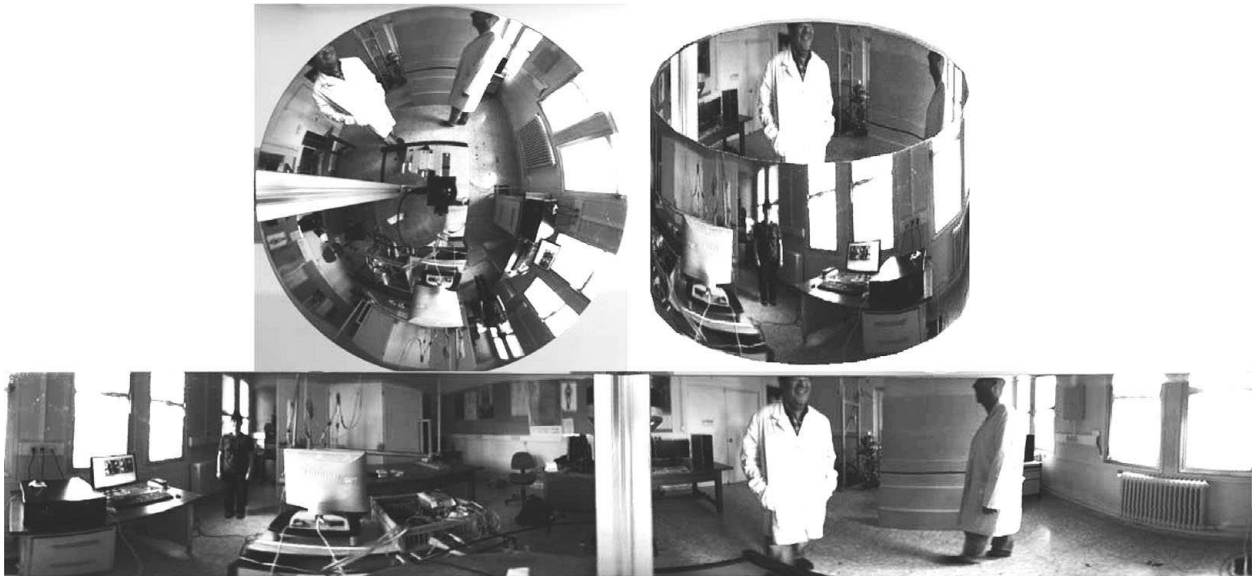

Fig. 5. (Top-left) omnidirectional image; (top-right) cylindrical representation and (bottom) panoramic image.

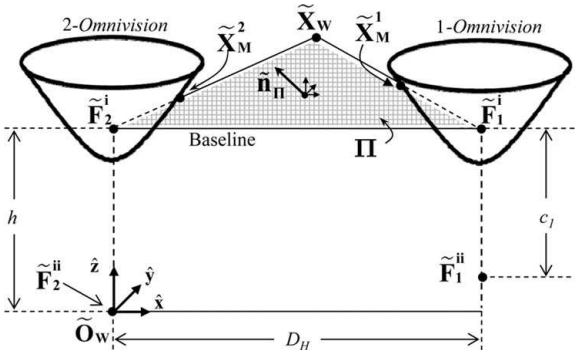

Fig. 6. Stage # 1 for the epipolar geometry of two parallel omnidirectional vision systems with hyperbolic mirrors.

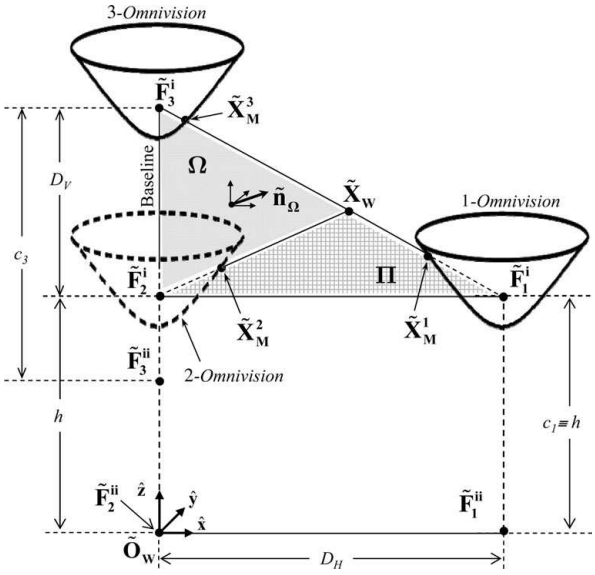

Fig. 7. Complete model for the epipolar geometry of three omnidirectional vision systems, two parallel and one aligned vertically.

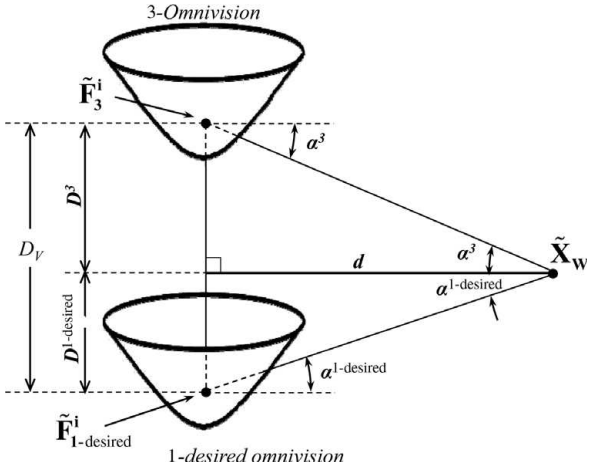

Fig. 8. Triangulation and depth computation of two catadioptric systems aligned vertically.

Nevertheless, since we are interested in getting a high resolution compact stereovision system with only two cameras, we propose to consider 2-*omnivision* as a virtual camera equivalent to the desired position of 1-*omnivision*, leading to $\widetilde{\mathbf{X}}^{\mathbf{2}}_{\mathbf{M}} \Rightarrow \widetilde{\mathbf{X}}^{\mathbf{1-desired}}_{\mathbf{M}}$ (see Fig. 7). If we compute $\widetilde{\mathbf{X}}^{\mathbf{1-desired}}_{\mathbf{M}}$, it will be possible to get a vertically aligned rectified configuration with two decoupled high-resolution catadioptric systems. The theory and properties of rectified images were presented by Hartley in ref. [31]. For catadioptric systems, the rectification process[20,32] provides epipolar radial lines onto the image plane and when they are projected onto the panoramic perspective, they become parallel lines to the vertical axis. Consequently, the depth is isotropic in all directions.

As the points $\widetilde{\mathbf{X}}^{\mathbf{1}}_{\mathbf{M}}$ and $\widetilde{\mathbf{X}}^{\mathbf{3}}_{\mathbf{M}}$ are known positions, the rectification process becomes a problem for solving the equation system of the intersection between **Ω**-plane (Eq. (13)), **Π**-plane (Eq. (10)) and the quadric equation of the mirror of 2-*omnivision* (named $\mathbf{M_{1-desired}}$, Eq. (1)). The system has two possible solutions, which are the intersections of the planes (**Ω**- and **Π**-planes) with the mirror surface $\mathbf{M}_{1-\mathbf{desired}}$.

However, since the azimuthal angle of the $\widetilde{\mathbf{X}}^{\mathbf{3}}_{\mathbf{M}}$ has been computed and it must be same for $\widetilde{\mathbf{X}}^{\mathbf{1-desired}}_{\mathbf{M}}$, it is then possible to select the correct solution for $\widetilde{\mathbf{X}}^{\mathbf{1-desired}}_{\mathbf{M}}$.

Once $\widetilde{\mathbf{X}}^{\mathbf{1-desired}}_{\mathbf{M}}$ has been calculated, the problem for depth estimation from a pair of catadioptric systems is reduced to

Table II. Pseudo-code for objects detection and depth estimation.

Imaging procedure

1. Image acquisition:
   - Two images (img 1–1 and img 1–2) acquisition by 1-*omnivision*.
   - Two images (img 3–1 and img 3–2) acquisition by 3-*omnivision*.
2. Motion detection (segmentation):
   - Image segmentation (imgSeg1): between img 1–1 and img 1–2.
   - Image segmentation (imgSeg3): between img 3–1 and img 3–2.
3. Feature extraction:
   - $F_{\text{omni 1}} = \{\text{motion, colour}\} \Rightarrow \text{ROI}_{\text{omni 1}}$
   - $F_{\text{omni 3}} = \{\text{motion, colour}\} \Rightarrow \text{ROI}_{\text{omni 3}}$
4. Rectification: transformation 1-*omnivision* to 2-*omnivision* (1-desired omnivision), Eqs. (13), (10) and (1).
5. Disparity map and depth computation between $\text{imgSeg}_{\mathbf{1-desired}}$ and imgSeg_3 by Eq. (14).
6. Closest object localisation strategies.

a simple triangulation, similar to what happens when using conventional cameras. Figure 8 illustrates the triangulation and depth computation process where α_3 and $\alpha_{1-\textbf{desired}}$ can be obtained from Figs. 3 and 7 and Eqs. (4) and (5), so depth d is obtained using the following equation:

$$\left.\begin{aligned} D_V &= D^{1-\text{desired}} + D^3 \\ \tan(\alpha^{1-\text{desired}}) &= \frac{D^{1-\text{desired}}}{d} \\ \tan(\alpha^3) &= \frac{D^3}{d} \end{aligned}\right\} \Rightarrow d$$

$$= \frac{D_H}{\tan(\alpha^{1-\text{desired}}) + \tan(\alpha^3)}. \tag{14}$$

It is common that when rectification of images is been doing some region onto the image plane will present singularities or will be occluded in one of the images. Hence, these regions must me identified and avoided. In the case of our system, the regions close to two epipoles and the centre of the images are where these singularities or occlusions are created.

5. Experimentation

Our goal for the experimental stage is to present an initial approach for human-like behaviour by means of the attitude of our proposed humanoid head. The visual servoing control task is focused in the attention strategies where the system reacts to any movements in its surroundings, in this case to the nearest moving object.

The first stage involves image processing for the omnidirectional vision system. Promising results in matching correspondences depend on previous segmentation procedures. Commonly, robust techniques, such as the well-known methods of Mean-shift, CAMshift[33] and so on are applied. However, since our purpose is to introduce robots in dynamic and changing environments for real-time tasks, we have used a robust algorithm for motion segmentation based on robust affine regression,[35,36] and also proposed in previous work,[34] where several experiments in hard dynamic outdoor scenes have been successfully carried out. In order to obtain robust feature for interest regions, we also include the variance and mean of the colour RGB of each segmented region. Then the rectification of the system is applied to the segmented regions (solving the equation system of Eqs. (10) and (13)) and the disparity map is computed (see Eq. (14)). The pseudo code used for this stage is summarised in Table II.

In order to present the results of the image procedure, we have selected a representative pairs of images from large image sequences acquired with the panoramic stereo system (top 1-*omnivision* and bottom 3-*omnivision*), displayed in Fig. 9. It is possible to observe three interesting situations, the first an easiest scenario when a single object is moving in the surroundings of the system (see Fig. 9(a)). In the second pair, another object enters the scene; therefore, the current problem involves multiple moving objects' detection (see Fig. 9(b)). The third pair shows several objects that move in random trajectories around the system (see Fig. 9(c)). The sequences were acquired in dynamic changing scenarios with

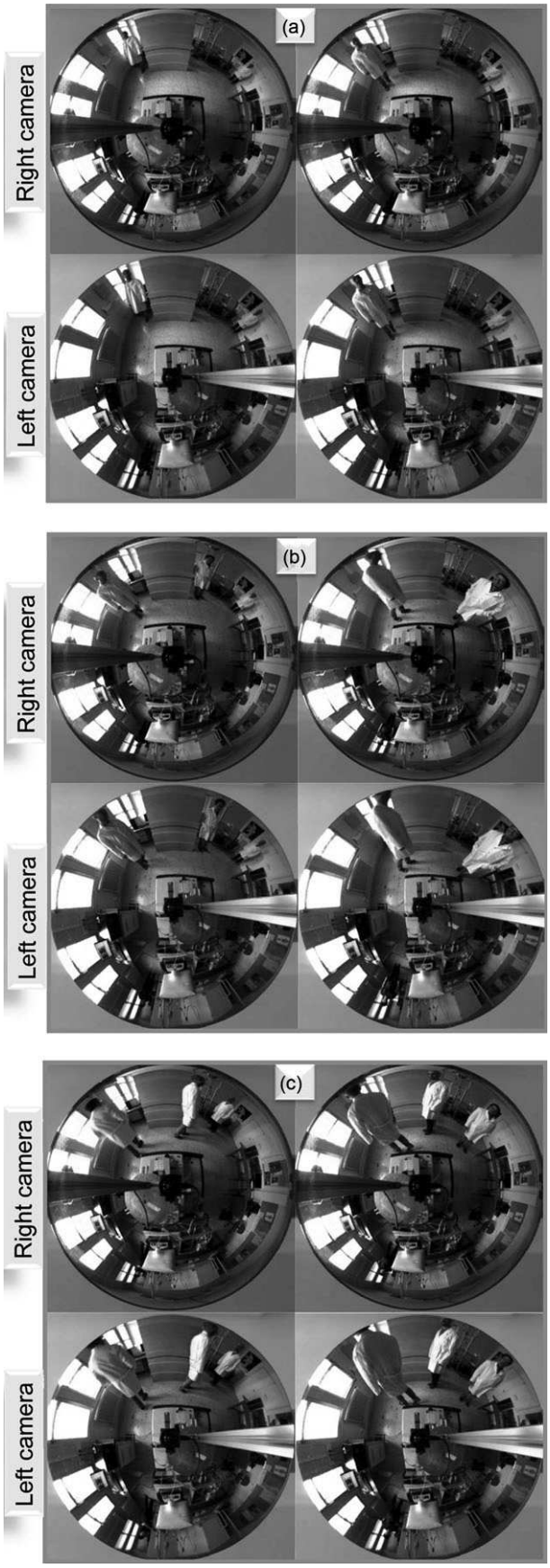

Fig. 9. (Colour online) Omnidirectional images sequences: three pairs from (top) 1-*omnivison* and (bottom) 3-*omnivision*.

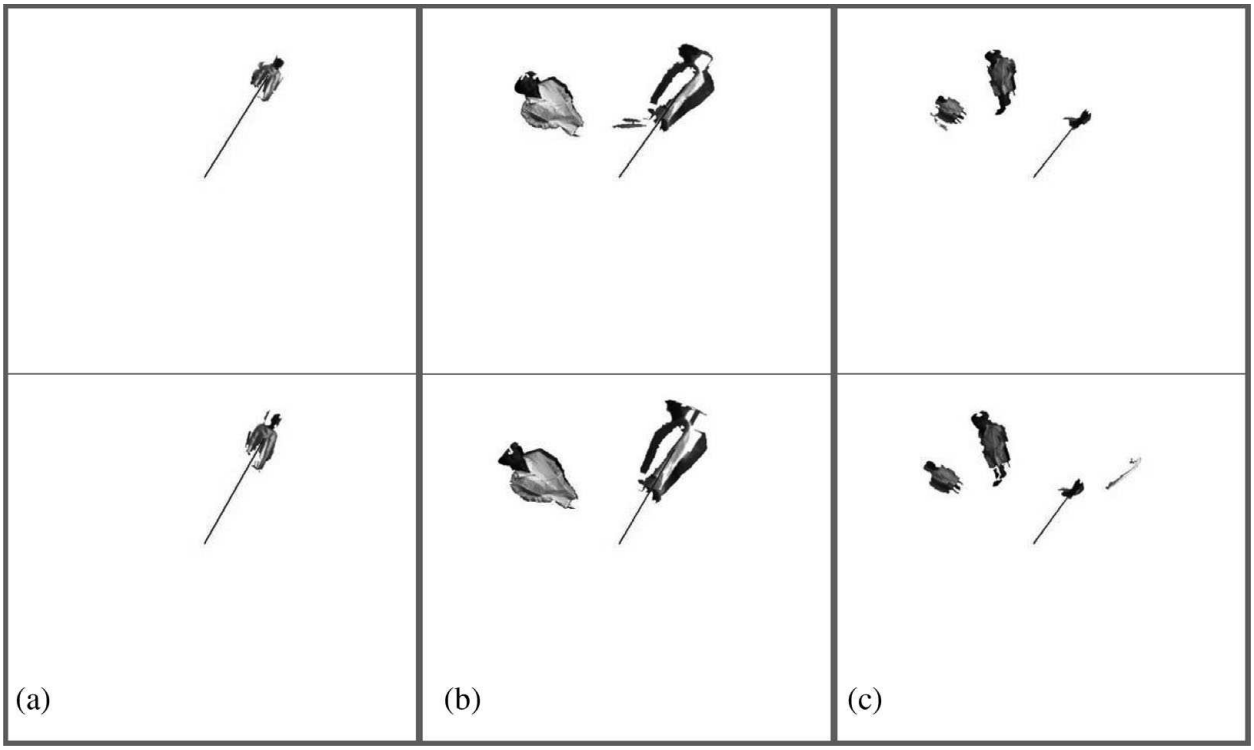

Fig. 10. The robust motion segmentation of image pairs and feature extraction for ROI.

uncontrolled light conditions; it is possible to observe the light coming through the window.

The extraction of robust features of the regions (objects – possible targets) in movement is the main goal of the image-processing stage (see Table II). The combination of the robust algorithm for motion detection and region clustering by means of the gradient of colour variances allows us to compensate the changes in light conditions. Figure 10 shows the motion detection and feature extraction of three pairs of images. These three pairs describe similar situations as the images presented in Figs. 9(a), (b) and (c), respectively.

Once the regions of interest (ROI) are identified on both segmented images, we transform 1-*omnivision* to 2-*omnivision* (1-*desired omnivision)*. When the images are vertically aligned, the disparity can be computed (inversely proportional to depth) for each ROI, therefore the nearest object can be defined. In order to perform the process of matching correspondences, the translation of the ROI needs a transformation due to catadioptric resolution. Figure 11 exemplifies a disparity map and depth representation where the darker objects represent the nearest regions of interest. In addition, as a result of the disparity calculation, in Fig. 10 the nearest object of each pair of images, captured by 1-*omnivision* and 3-*omnivision*, is selected by a radial line from the image centre.

Several experiments were performed in order to test the visual attention control strategy, the results generated for the vision system for tracking the nearest object are presented as a function of the angular position (the elevation **β** and the azimuthal **α** angles), distances (*disp.*) and magnitude of the movement (**M**). The decision stage weighs the disparity and magnitude of the movement to select the nearest object; in this way the system will let alone near static objects and will

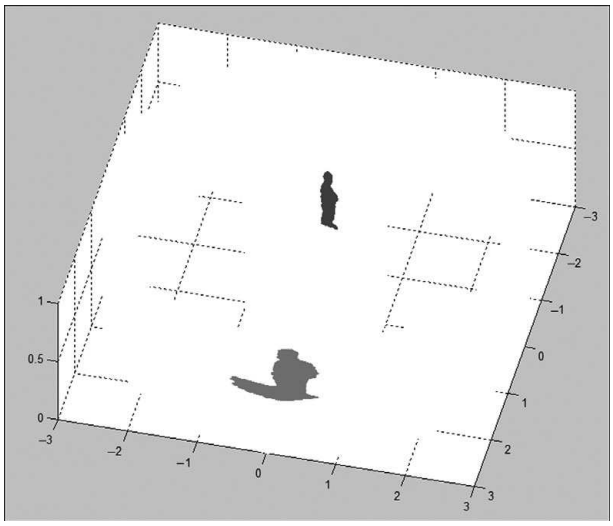

Fig. 11. Disparity map of the objects in motion, dark grey is the closest.

attend the next nearest object with the largest movement. Under this rule our humanoid head will move in a "curious" human-like manner.

In Fig. 12 we present the results of the elevation (φ) and the azimuthal (α) angles, obtained by the omnidirectional stereovision system – a long and large image sequence, the image processing of this initial approach takes $\sim$ 300 ms.

Three interesting cases are represented in Fig. 12. In case 1, the system is tracking the trajectory of an object in movement (*Object_1*), and unexpectedly another object moves (*Object_2*), a swiftly movement (to kick, to drop something, etc), since the position of *Object_2* with respect to the humanoid head is closer than *Object_1*, the system attends this action.

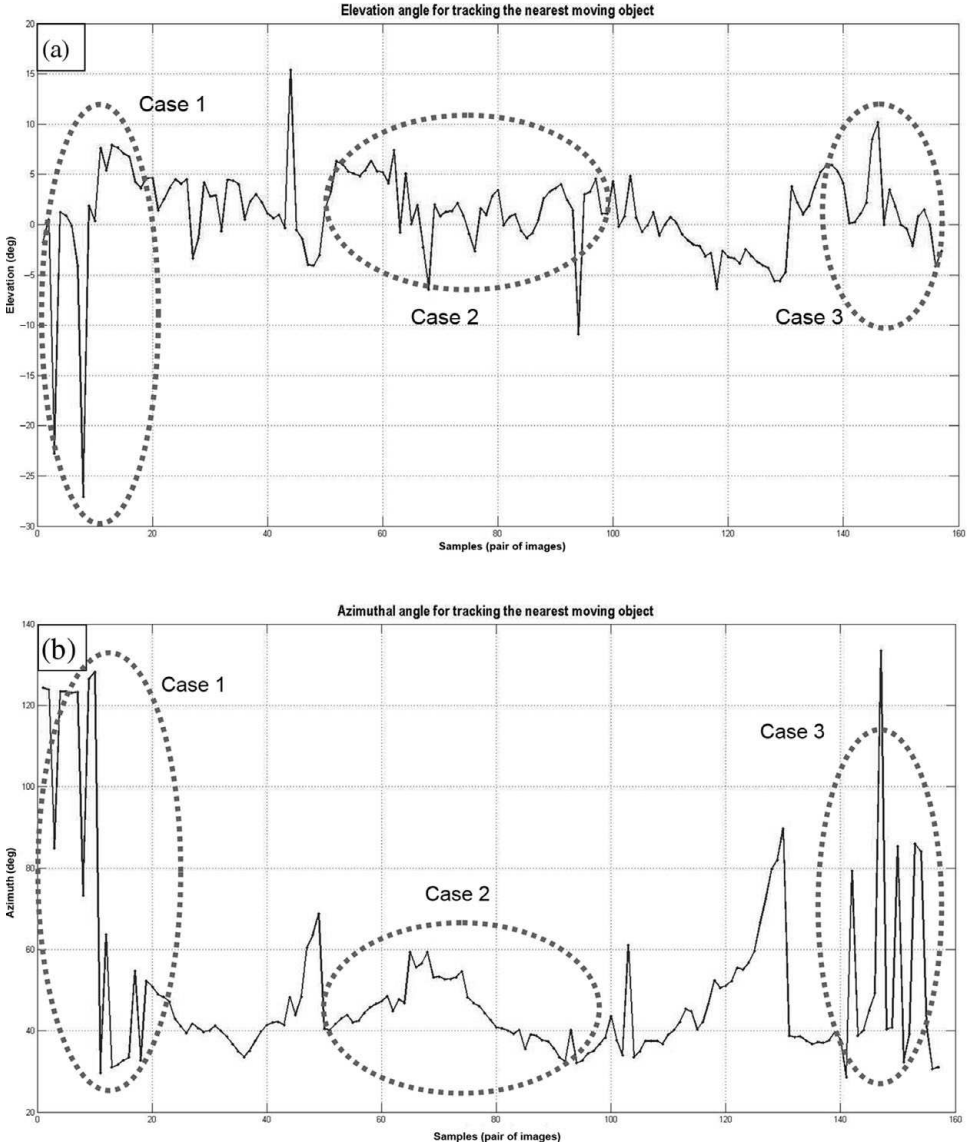

Fig. 12. Experimental results obtained by omnidirectional stereovision system: (a) elevation angle; (b) azimuthal angle.

This is represented as outliers in the natural tracking of *Object_1* (this scenario is represented in Fig. 10(c)).

The second case (case 2) shows the natural tracking of the nearest object in movement (*Object_1*), where none of the objects is nearer than *Object_1*, this situation is presented in Fig. 10(a). And finally (case 3), a scenario where two people are interacting around the human head at equivalent distances; in this case both people are selected intermittently as a target, for example when people are talking, moving their hands and so on (see Fig. 10(b)).

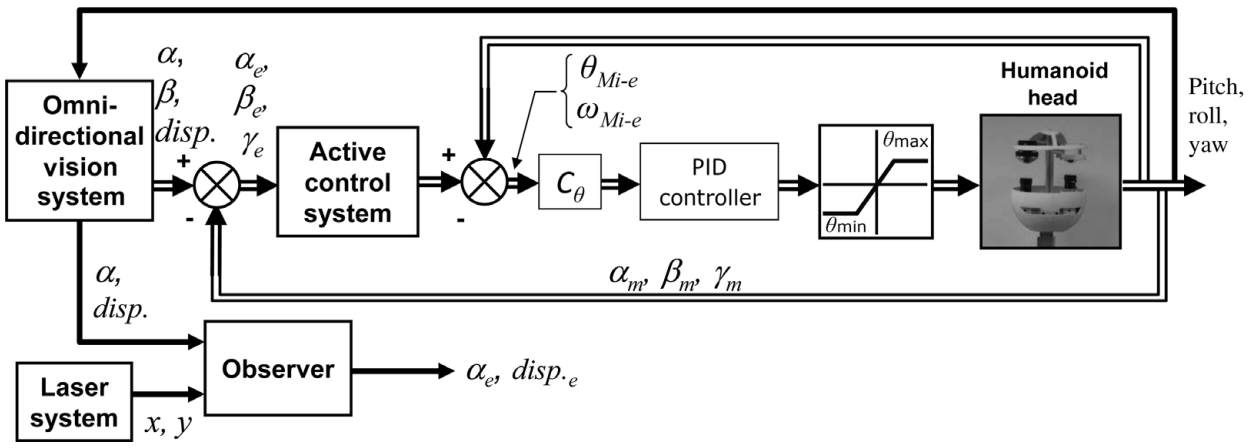

Fig. 13. Block Diagram of active control system for humanoid head using omnidirectional stereovision system.

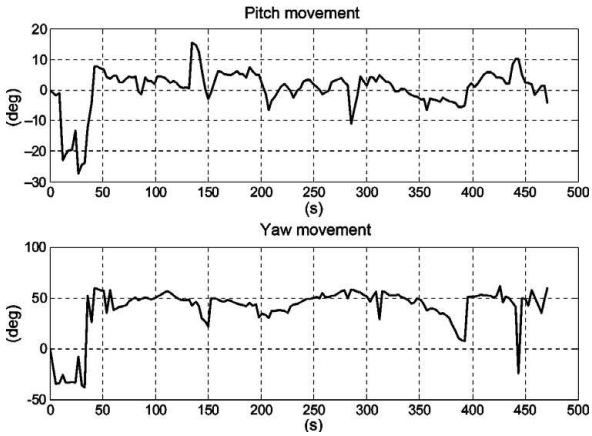

Fig. 14. Experimental results obtained by active control system for tracking the nearest object.

Figure 13 shows a block diagram of the active control system of the humanoid head using omnidirectional stereovision system. The omnidirectional stereovision system supplies the control system input, the azimuth angle (α), the elevation angle (β) and the disparity of the images (*disp*.). Both the azimuth angle and the disparity of the images provided by the omnidirectional stereovision system are validated with data from the laser system, Sick LMS-291, verifying the effectiveness of the vision system and algorithm described in Table II. Angles α, β and *disp*. values supplied by the vision system are the inputs for active control system of the humanoid head, so the head can "seek" the objects in movements. The actions carried out by the humanoid head are the pitch (β), roll (*disp*., β) and yaw (α). The angular displacements of the neck are limited by the control system to perform natural tracking tasks.

Figure 14 shows the pitch and yaw movements of the humanoid head according to the results presented in Fig. 12. The results of the pitch and yaw movement (Fig. 14) are equivalent to the results of the elevation and azimuthal angles given in Fig. 12. There is a difference between these, because as a normal human behaviour the attention of the humanoid head changes if the most interesting condition appears on scenario. The commands to the active control system are consecutively refreshed when an object is detected by the omnidirectional vision system.

Several experiments were realised in order to test the basic visual servoing control algorithm. In this case (Fig. 15) the omnidirectional stereovision system detects three people in movement, and the vision system is able to localise the closest one. On the other hand, the system can detect the most distant object if it is moving and the nearest objects are in static position. Subsequently, the humanoid head turns its "attention" (central point of the catadioptric system) to the

Fig. 15. Experimental testing of the implemented visual servoing algorithm to detect moving objects using omnidirectional stereovision system.

central point of the nearest object (or moving object in the area detected by the vision system).

When two or more moving objects interact with each other at similar distances with respect to the humanoid head, it performs oscillatory movements paying attention to all them.

The behaviour of the humanoid head, briefly described above, can be noted in the photographic sequences shown in Fig. 15. Sequences depicted in Fig. 15 should be read from left to right and top to bottom. In each picture there are arrows that indicate the attitude of the humanoid head (pitch and yaw angles) and the area towards the humanoid head turns his attention.

6. Conclusions

A systematic study of the proposed omnidirectional vision system was carried out in order to present a reconfigurable stereovision system for an adaptive humanoid head. The selected approach consists of panoramic stereovision system, composed by two hyperbolic catadioptric systems, resulting finally in a compact high resolution configuration easy to reproduce and feasible for real-time applications. A specially designed hyperbolic mirror was also micro-manufactured in the Centre for Automation and Robotics (CSIC-UPM). In addition, a simplified method for depth estimation was presented.

Initial experimental results have been introduced. As an illustration of good performance of the proposed system, a wide range of typical human-like scenarios was used for these experiments. We have validated the stereo omnidirectional vision system (depth estimation) and the target position (azimuthal angle) by using as reference system such as LIDAR range sensor.

The problem of singularities and lateral occlusions decreases the range of view; nevertheless this problem can be solved by using the neck movement. Since the objects position could be computed and tracked, the system could gather the situation when an object is approaching to singularities or occlusions.

Acknowledgments

This research was partially funded by Consejería de Educación de la Comunidad de Madrid under grant RoboCity2030 S-2009/DPI-1559, and Agencia Española de Cooperación Internacional para el Desarrollo (AECID) under grant FORTUNA D/030531/10. Dr. H. Montes acknowledges the support received from Universidad Tecnológica de Panamá and CSIC under JAE-Doc Programme.

References

1. Y. Sakagami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki and K. Fujimura, "The Intelligent ASIMO: System Overview and Integration," **In**: *Proceedings of the IEEE/RSJ, International Conference on Intelligent Robots and Systems*, EPFL, Lausanne, Switzerland (Sep. 30–Oct. 4, 2002) pp. 2478–2483.
2. K. Kaneko, F. Kanehiro, S. Kajita, H. Hirukawa, T. Kawasaki, M. Hirata, K. Akachi and T. Isozumi, "Humanoid Robot HRP-2," **In**: *Proceedings of the IEEE International Conference on Robotics and Automation*, New Orleans, LA (Apr. 26–May 1, 2004) pp. 1083–1090.
3. F. Tanaka and H. Suzuki, "Dance Interaction with QRIO: A Case Study for Nonboring Interaction by Using an Entrainment Ensemble Model," **In**: *Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication*, ROMAN (Sep. 20–22, 2004) pp. 419–424.
4. S. Lohmeier, K. Loffler, M. Gienger, H. Ulbrich and F. Pfeiffer, "Computer System and Control of Biped 'Johnnie'," **In**: *Proceedings of the IEEE International Conference on Robotics and Automation*, New Orleans, LA, Vol. 4 (Apr. 26–May 1, 2004) pp. 4222–4227.
5. C. L. Breazeal, Sociable Machines: Expressive Social Exchange between Humans and Robots *Ph.D. Dissertation* (Massachusetts Institute of Technology, Cambridge, MA, USA, 2000).
6. R. Brooks, C. Breazeal, M. Marjanović, B. Scassellati and M. Williamson, *The Cog Project: Building a Humanoid Robot*, Lecture Notes in Computer Science (LNCS). (Springer-Verlag, Heidelberg, Germany, 1999) pp. 52–87.
7. J. Hirth, N. Schmitz and K. Berns, "Emotional Architecture for the Humanoid Robot Head ROMAN," **In**: *Proceedings of the IEEE International Conference on Robotics and Automation*, Roma, Italy (Apr. 10–14, 2007) pp. 2150–2155.
8. E. Yoshida, J-P. Laumond, C. Esteves, O. Kanoun, A. Mallet, T. Sakaguchi and K. Yokoi, "Motion autonomy for humanoids: experiments on HRP-2 No. 14," *Comput. Animat. Virtual Worlds* **20**, 511–522 (2009).
9. O. Stasse, B. Verrelst, B. Vanderborght and K. Yokoi, "Strategies for humanoid robots to dynamically walk over large obstacles," *IEEE Trans. Robot.* **25**(4), 960–967 (2009).
10. J. Chestnutt, P. Michel, J. Kuffner and T. Kanade, "Locomotion Among Dynamic Obstacles for the Honda ASIMO," *Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, San Diego, CA, USA (Oct 29–Nov 2, 2009).
11. F. Pfeiffer, "The TUM walking machines," *Phil. Trans. R. Soc.* **365**(1850), 109–131 (2007).
12. D. W. Rees, "Panoramic Television Viewing System," US Patent No. 3505465 (1970).
13. J. Hong, "Image Based Homing," **In**: *Proceedings of the International Conference on Robotics and Automation*, Sacramento, USA (1991) pp. 620–625.
14. K. Yamazawa, Y. Yagi, M. Yachida, "Omnidirectional Imaging with Hyperboloidal Projection," **In**: *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Yokohama, Japan (Jul. 26–30, 1993) pp. 1029–1034.
15. Y. Yagi, Y. Nishizawa and M. Yachida, "Map-based navigation for a mobile robot with omnidirectional image sensor COPIS," *IEEE Trans. Robot. Autom.* **11**(5), 634–648 (1995).
16. C. Geyer and K. Daniilidis, "Catadioptric projective geometry," *Int. J. Comput. Vision* **45**(3), 223–243 (2001).
17. T. Svodoba, "Central Panoramic Cameras Design, Geometry, Egomotion," *Ph.D. Thesis* (Center for Machine Perception, Czech Technical University, Prague, Czech Republic, 1999).
18. S. Baker and S. K. Nayar, "A theory of single-viewpoint catadioptric image formation," *Int. J. Comput. Vis.* **35**(2), 1–22 (1999).
19. T. Svodoba, T. Padjdla and V. Hlavac, "Epipolar Geometry for Panoramic Cameras," **In**: *Proceedings of the European Conference on Computer Vision*, Bombay, India (Jan. 1998) pp. 218–232.
20. J. Gluckman, S. K. Nayar and K. J. Thoresz, "Real-Time Omnidirectional and Panoramic Stereo," **In**: *Proceedings of DARPA Image Understanding Workshop* (Nov. 1998) pp. 299–303.
21. E. L. Cabral, J. C., Junior and M.C. Hunold, "Omnidirectional Stereo Vision with a Hyperbolic Double Lobed Mirror," *Proceedings of the Pattern Recognition, 17th International Conference*, Vol. 1 (IEEE CS Press, Washington, DC, 2004).

22. S. A. Nene and S. K. Nayar, "Stereo with Mirrors," **In**: *Proceedings of International Conference on Computer Vision*, Bombay, India (Jan. 1998) pp. 1087–1094.
23. M. Armada, R. Caballero, T. Akinfiev, H. Montes, C. Manzano, L. Pedraza and P. González de Santos, "Design of SILO2 Humanoid Robot," **In**: *Proceedings of IARP Workshop on Humanoid and Friendly Robotics*, Tsukuba, Japan (Dec. 11–12, 2002) pp. 37–42.
24. H. Montes, "Análisis, diseño y evaluación de estrategias de control de fuerza en robots caminantes," *Ph.D. Thesis* (U. Complutense, Spain, 2005).
25. H. Montes, C. Salinas, G. Fernandez, P. Clarembaux, P. Gonzalez de Santos and M. Armada, "Omnidirectional Stereo Vision Head for Humand Robots," **In**: *Proceedings of CLAWAR'09*, Istabul, Turkey (Sep. 9–11, 2009) pp. 909–918.
26. D. A. Winter, *Biomechanics and Motor Control of Human Movement* (John Wiley, Hoboken, NJ, 1990).
27. A. Vasavada, L. Siping and S. Delp, "Influence of muscle morphometry and moment arms on the moment-generating capacity of human neck muscles," *SPINE* **23**(4), 412–422 (1998).
28. R. Benosman and S. B. Kang, *Panoramic Vision: Theory System and Applications* (Springer-Verlag, New York, 2001).
29. T. Padjdla, "Localization Using SVAVISCA Panoramic Image of Agam Fiducials – Limits of Performance," *Technical Report* (Center for Machine Perception, Czech Technical University, Prague, Czech, 2001).
30. J. Gaspar, C. Deccó, J. Okamoto and J. Santos-Victor, "Constast Resolution Omnidirectional Cameras," *Proceedings of Workshop on Omni-Directional Vision*, Copenhagen, Denmark (2002).
31. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision* (Cambridge University Press, Cambridge, UK, 2004).
32. Z. Zhu, "Omnidirectional Stereo Vision," *Workshop on Omnidirectional Vision, Proceedings of the 10th IEEE ICAR*, Budapest, Hungary (2001).
33. D. Comaniciu and P. Meer. "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(5), 603–619 (2002).
34. C. Salinas and M. Armada, "Analysing Human-Robot Interaction Using Omnidirectional Vision and Structure from Motion," *Proceedings of CLAWAR'08*, Coimbra, Portugal (2008).
35. M. Black, "The robust estimation of multiple motions: Parametric and piecewise smooth flow fields," *Comput. Vis. Image Underst.* **63**(1), 75–104 (1996).
36. A. Bab-Hadiashar and D. Suter, "Robust optical flow computation," *IJCV* **29**(1), 59–77 (1998).
37. E. R. Davies, *Machine Vision, Third Edition: Theory, Algorithms, Practicalities (Signal Processing and its Applications)* (Morgan Kaufmann, Massachusetts, 2005).