Catadioptric single-shot rangefinder for textured map building in robot navigation

R. Orghidan, E.M. Mouaddib, J. Salvi and J.J. Serrano

Abstract: The paper focuses on combining the advantages of both omnidirectional vision and structured light in order to obtain panoramic depth information for generating a complete 3D visual model of the surroundings of a mobile robot. The model is completed by mapping the imaged texture onto the objects through the central view point. This process is called single-shot scene reconstruction because a visual model of the scene is obtained from a single omnidirectional image. Several such reconstructions are then registered to build a 3D map of the scene while the robot is moving. The authors call this process range scan registration modelling. A mobile platform provided with a catadioptric camera coupled with an omnidirectional structured light projector was used for this purpose. Sensor calibration, image processing and experimental results are included.

1 Introduction

Omnidirectional vision enlarges the field of view (FOV) of traditional cameras by means of special optics and combinations of lenses and mirrors, obtaining a panoramic and complete view (360°) of the surroundings of the camera. Many researchers have explored the advantages of using omnidirectional cameras. If the scene is observed from a single point in space, the sensor has a single projection centre, that is, a single view point (SVP). The SVP is one of the most used classification criteria for the cameras with wide FOV [1, 2].

One of the first groups to use omnidirectional sensors for robot navigation and 3D perception was Yagi and Yachida [3], who used the conical mirror-based sensor COPIS. They obtained the depth by triangulation, using dead-reckoning information. If a 3D scene model is available, the panoramic image can be mapped onto the model to obtain a virtual 3D environment. Researchers from the University of Alberta [4] used a similar method for pipe inspection, using a catadioptric camera with a conical mirror. The omnidirectional image provided by the catadioptric camera is mapped onto a 3D surface defined by a previously available model of the real surface or onto a complex developable surface formed by basic building blocks. Svoboda *et al.* [5] explored the epipolar geometry of panoramic cameras and used this property for extracting point correspondences between images, a key feature for stereo vision. Bunschoten and Kröse [6] used the theory of the epipolar geometry between two panoramic images and performed range estimation from a mobile platform in a synthetic environment. Recently, Fleck *et al.* [7] obtained a 3D model of the environment using a mobile robot that is equipped with a panoramic camera and a laser scanner used for odometry. The scene reconstruction is completed using graph-cut methods, are close to optimal but computationally very intensive. Besides, other authors used catadioptric cameras and time-of-flight laser-range finders for ego-motion estimation and map generation by integrating the depth maps provided by the two systems, see Miura and colleagues [8, 9].

Stereo catadioptric sensors are special structures of mirrors and lenses designed to obtain depth from images with a wide FOV. In order to obtain different points of view of a scene with a single camera, several researchers used structures of convex mirrors [10–12] or planar mirrors [13–15]. For instance, Southwell *et al.* [12] used a non-SVP catadioptric sensor provided with a two-lobbed spherical mirror to obtain two images with separated view points that allow panoramic 3D perception. Later, Fiala and Basu [10] used a similar sensor to obtain a stereo panoramic view in which horizontal and vertical lines were detected using the panoramic Hough transform, which is an extension of the Hough transform adapted for catadioptric images. Depth is retrieved by matching the lines imaged by both spherical mirrors through radial symmetry. Kawanishi *et al.* [15] developed a high-resolution SVP omnidirectional sensor built with six cameras and a hexagonal pyramidal mirror. Stereo views are obtained by connecting two such sensors placed in a back-to-back configuration. Nene and Nayar [11] studied the epipolar geometry of stereo catadioptric images obtained by several SVP devices built using a camera pointed at pairs of planar, elliptic, hyperbolic and parabolic mirrors. Formal mathematical models were obtained and the FOV of each catadioptric combination was deduced. The epipolar geometry was also used by Gluckman *et al.* [16], who obtained stereo panoramic views using either two paracatadioptric cameras aligned vertically [16] or two planar mirrors with a perspective camera [13]. Lin and Bajcksy [17] described an omnidirectional sensor that provides depth information by using two cameras, a beam splitter and a conical mirror. Recently, a similar sensor has been proposed by Spacek [18], which uses an omnidirectional stereo setup built with two coaxial conical mirrors.

doi:10.1049/iet-cvi:20065003

Paper first received 26th December 2006 and in revised form 14th April 2007

R. Orghidan and J. Salvi are with Computer Vision and Robotics Group, Institute of Informatics and Applications, University of Girona, Girona, Spain

E.M. Mouaddib is with Centre of Robotics, Electrotechnics and Automation, University of Picardie Jules Verne, Amiens, France

J.J. Serrano is with the Department of Systems Data Processing and Computers, Polytechnic University of Valencia, Valencia, Spain

E-mail: rorghi@yahoo.com

Stereo techniques are based on the assumption that the feature correspondences between images are found accurately. In the case of catadioptric sensors, the matching process is deteriorated because the resolution of the omnidirectional cameras is usually very poor compared with conventional cameras because of the different FOVs. This shortcoming affects matching since it becomes more difficult to find details in the images even when taking into account the epipolar constraints [5]. This problem can be alleviated by using structured light-based techniques that are a particular case of stereo vision in which one of the cameras is replaced by a pattern projector [19, 20]. Using this method is similar to placing visible landmarks in the scene so that image points can be identified and matched faster and more accurately. For instance, Mei and Rives [21] addressed the SLAM problem using a laser-range finder emitting a laser trace in the scene and an SVP omnidirectional camera. The main problem of this approach is that the laser pattern is projected by rotating a laser spot which deteriorates both the correspondence problem and the real-time behaviour. In our case, we have chosen a circle pattern obtained by diffracting the light emitted by a laser diode. The advantage of such a laser scanner is the high resolution and accuracy obtained compared with other structured light scanners. The energy pattern of a laser beam has a Gaussian profile, which enables it to be detected in the image with sub-pixel accuracy.

2 Sensor design and geometry

Because of their optical properties, catadioptric sensors with SVP were chosen to build the omnidirectional camera of the proposed sensor. Usually, the central projection constraint is fulfilled by using a parabolic mirror in conjunction with an orthographic camera or a hyperbolic mirror with a perspective camera. We finally chose to use the first configuration, which preserves the SVP independently from the translation between the mirror and the camera because of its simplicity in focusing the image.

The third dimension of the scene is perceived by crossing the light rays emitted by the laser with the ones observed by the camera or, in other words, by performing a triangulation. The laser should project a pattern that covers the whole scene and should be easily identifiable. We chose to use the circle as the pattern. The light of the projected laser pattern is reflected by the conical mirror and spread onto the scene along an imaginary conical shape. The parabolic mirror reflects the whole scene into the camera and all the spots illuminated by the laser can be immediately identified (Fig. 1). With the models for both components of the sensor, a precise triangulation can be carried out. In this section, we give an overview of the modelling process and the parameter estimation (calibration) of the sensor. For a more in-depth understanding of the calibration process, the reader can consult our previous publications [22, 23].

2.1 Omnidirectional camera

Assuming that the camera–mirror pair possesses SVP, the omnidirectional camera can be modelled as a projection onto a sphere followed by a projection onto a plane, as stated by Geyer and Daniilidis [24, 25]. Another way of approaching camera calibration is by considering the mirror surface as a known revolution shape and modelling it explicitly, for instance, considering that the reflecting surface is a paraboloid and the camera is orthographic. Both models were tested and the comparative results were reported in [26]. The omni camera used for this work has an SVP but contains two reflecting surfaces, so the first method mentioned (shape independent) was preferred.

The calibration is performed using a set of known 3D points, $P_w = [x_w, y_w, z_w]$, distributed on the four walls of a cube placed around the sensor. Considering $P_s = [x_s, y_s, z_s]$ to be the intersection of the light ray emitted by the point P_w with the sphere of radius R, then the perspective projection of P_s on the image plane from a point $C = [0, \xi]$ produces a point $P_i = [x, y]$ as shown in Fig. 2.

Adding the intrinsic camera parameters (α_u, α_v, u_0, v_0), the pixel coordinates of the image points (u, v) are computed by (1)

$$u = \frac{\alpha_u(\xi + \varphi)x_w}{\xi\sqrt{x_w^2 + y_w^2 + z_w^2} - z_w} + u_0$$
$$v = \frac{\alpha_v(\xi + \varphi)y_w}{\xi\sqrt{x_w^2 + y_w^2 + z_w^2} - z_w} + v_0 \quad (1)$$

The 12 parameters of the model are ξ, which controls the eccentricity; φ, which is a function of both the eccentricity

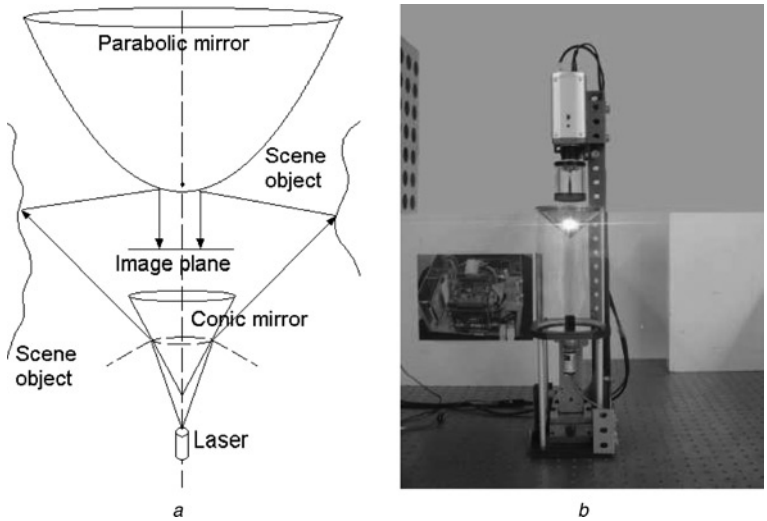

Fig. 1 *Proposed sensor*

a Catadioptric omnidirectional camera with an embedded structured light projector
b Laboratory prototype

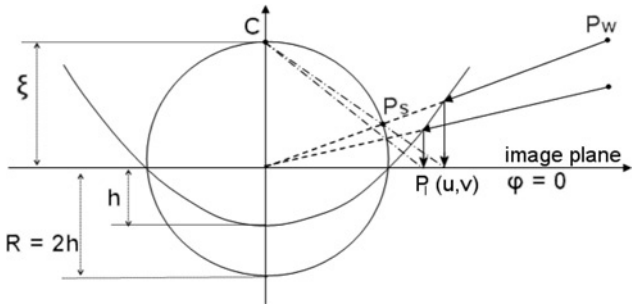

Fig. 2 *Image formation using the projective equivalence of an SVP catadioptric projection, with the projection on the sphere*

and the scale; α_u, α_v, u_0, v_0, the intrinsic camera parameters; $r_X(\alpha)$, $r_Y(\beta)$, $r_Z(\gamma)$, the orientation angles and t_x, t_y, t_z, the translation between the world and the camera coordinate systems [27].

2.2 Omnidirectional laser projector

The omnidirectional light projector is formed by a laser that emits a circle and is pointed at a conical mirror so that the projected light covers the entire FOV of the catadioptric camera. The proposed projector can be considered a reversed omni-camera, where the light flows in the opposite direction. Therefore the projector takes the advantage of the attributes revealed by previous studies of catadioptric cameras based on the conical mirror shape. Lin and Bajcsy [28] pointed out that the conical mirror can be used for building true SVP configurations with the advantage that it preserves image point brightness better than other mirrors since it does not distort the image in longitudinal directions. Yagi [2] highlighted the fact that the conical mirror on vertical section behaves like a planar mirror and consequently provides much better resolution than any other omni-mirror shape. Baker and Nayar [29] proved that curved mirrors (such as parabolic, hyperbolic, etc.) increase defocus blur because of their bend. Consequently, the cone is the ideal mirror shape for building the structured light projector.

The bright spots on the scene are observed by the previously calibrated omni-directional camera, which has a single projection centre. This property allows the direction of the light source to be calculated for each image point. Since the location of the calibration planes is known, the 3D coordinates of the laser stripe lying on those planes can be determined. A set of such points can be used for calibrating the laser-conic mirror pair. Ideally, when the laser is perfectly aligned with the conic mirror, the 3D shape formed by the reflected laser pattern can be imagined as a circular cone. Unfortunately, the precision to obtain the coordinates of the bright spots is limited by the catadioptric camera calibration accuracy and by its resolution. Moreover, a perfect alignment of the laser and the conic mirror is difficult to guarantee, so a more general shape than the circular cone should be considered. The general quadratic surface was chosen for modelling the laser projection. The general quadratic shape is expressed by the parametric equation (2), in which x, y and z are the coordinates of a point in the 3D space.

$$\begin{bmatrix} x & y & z \end{bmatrix} \cdot \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} + f = 0 \quad (2)$$

The parameters of the equation are a_{ij}, with $i = \overline{1,3}$, $j = \overline{1,3}$, b_i, with $i = \overline{1,3}$, and the coefficient f. The shape

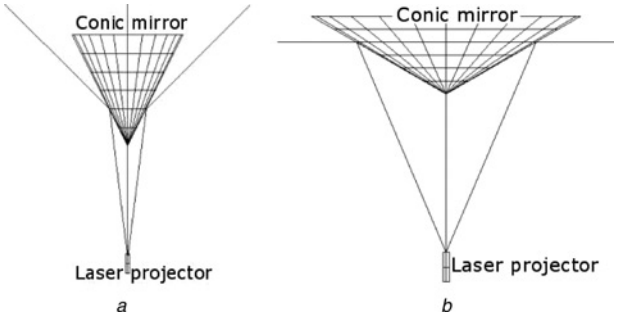

Fig. 3 *Omnidirectional-structured light projector*

a Projection of a conical shaped pattern
b Projection of a planar pattern

representing the projection of the circular laser pattern into the conic mirror was fitted to the 3D points and the parameters were found by a minimisation algorithm.

In a general case, the light hitting a conic mirror (Fig. 3*a*) produces a conical shape. If there is a special configuration between the vertical angle of the mirror plane and the incident light angle, the projected pattern becomes a plane as illustrated in Fig. 3*b*.

3 Image processing and map building

Generally, a 3D scene model can be represented either as a cloud of points or as a parametric model [30]. The first technique is generally used for object visualisation by means of polygonal mesh surfaces generated from the point cloud. The resulting models have limited accuracy for describing the object and require a large amount of computing time and memory space. The alternative to the point cloud model is the parametric model that provides useful information such as the size, shape, position and rotation of the objects. Moreover, the parametric model describes the object in a compact form alleviating the computation and storage issues. Because of the specific data acquisition of the proposed sensor, the parametric model representation is an appropriate choice when using the sensor in structured human-made environments. For these reasons, the parametric model was selected for reconstructing the scene.

The image processing flowchart from image capture to building the map is presented in Fig. 4.

3.1 Automatic laser stripe detection with sub-pixel accuracy

Once the sensor has been calibrated, the image acquisition process is initiated. In each shot, the laser stripe in the image is automatically detected. In the omnidirectional image, the laser stripe is affected by the superposition of a certain amount of undesired lighting peaks. This fact is because of noise produced by four main sources: electrical noise, quantisation noise, speckle and image blur. The first two noise sources are associated with the image sensor whereas the speckle is related to the reduced wavelength of light compared with the surface roughness and the monochromatism of the laser light. The image blur is inherent in catadioptric cameras because of the mirror curvature. An accurate reconstruction needs precise laser detection, therefore we used the peak detection method described by Forest *et al.* [31] in order to automatically extract the peak of the laser with sub-pixel accuracy. In the original paper, Forest used perspective images containing vertical laser lines. In our case, the omnidirectional image provided by the sensor

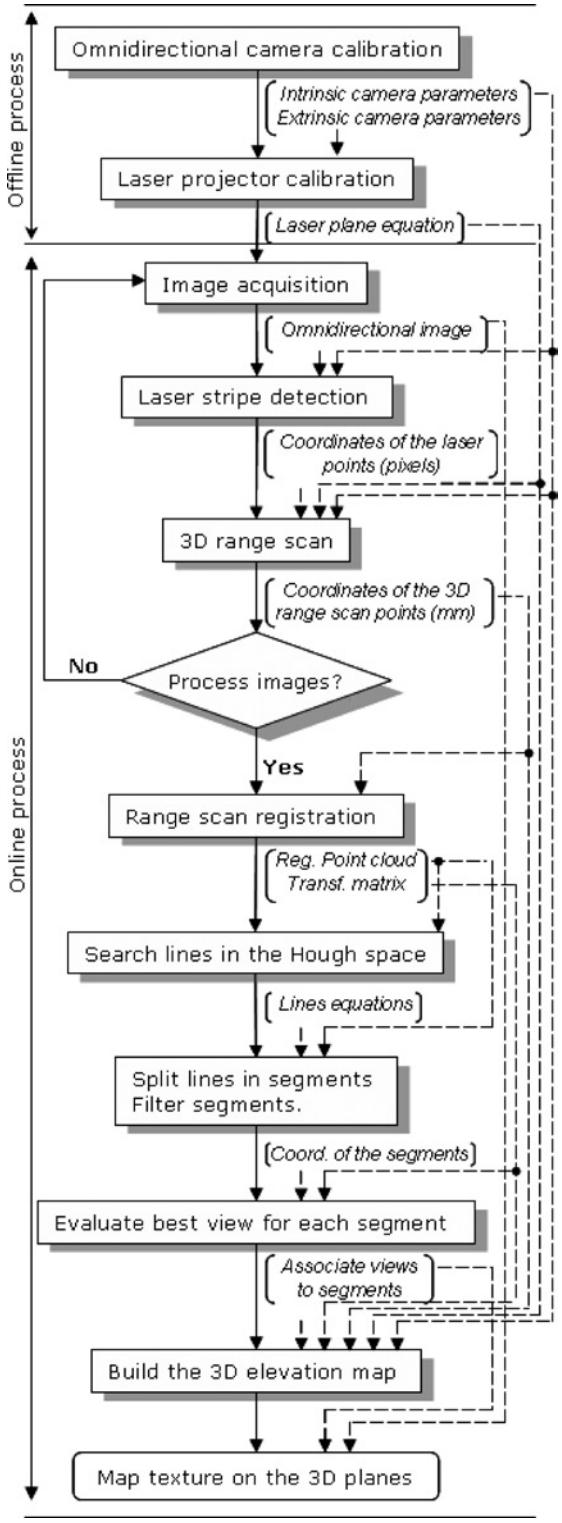

Fig. 4 *Image-processing flowchart*

is scanned radially in order to find the light peaks along the circular laser profile projected on the scene (Fig. 5).

3.2 Range scan computation and registration

The detected laser peaks are back-projected through the camera model and the 3D range scan is computed by triangulation. Fig. 6 shows four range scans calculated at different positions in a scene. It can be found that each range scan contains different details and even the surfaces that are seen in more than one range scan are detected with different accuracy.

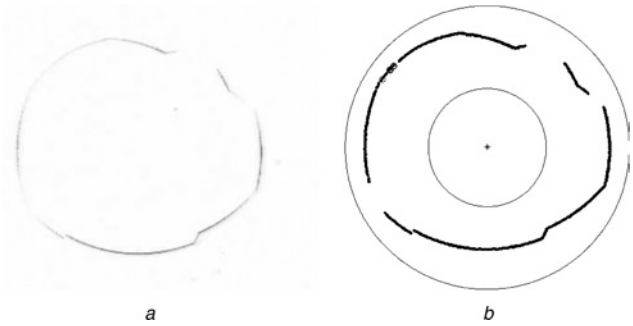

Fig. 5 *Peak detection of a laser stripe in the omnidirectional image*

a Laser profile corresponding to a real scene. In the original image, the bright laser stripe contrasts with the dark background. For printing convenience, the image is presented in negative
b Detecting the laser profile. Laser peaks are marked with dots and the outliers are marked with small circles. The image centre is shown with a star. The two concentric circles are the bounds of the stereo FOV and therefore the limits of the automatic search of the laser in the image

The single-shot scene model provides a rich representation but has two inherent drawbacks: there are occluded parts of the scene that are obviously not modelled and the lower resolution of objects in the distance. These shortcomings can be alleviated if several points of view of the scene are used, that is by reconstructing a larger scene using different range scans with overlapping regions. In addition, using a wide FOV is very useful for finding correspondences between parts of different images because the points are in the sensor's sight for longer. The process of aligning several lines or surfaces in the same coordinate system is called registration. The most common method for registration is the iterative closest point (ICP) algorithm, introduced by Besl and McKay [32]. This algorithm finds the motion between two clouds of points by minimising the distance between them. Consider that the coordinates of the N_p pairs of points that belong to the two clouds are stored in the $\boldsymbol{m}_i$ and $\boldsymbol{p}_i$ vectors. The distance between the pairs of points in the two clouds can be expressed by the function f, described by (3). The rotation R and the translation t, known as the motion parameters, between the two clouds of points, are found by minimising this function.

$$f = \frac{1}{N_p}\sum_{i=1}^{N_p} \|\boldsymbol{m}_i - R\boldsymbol{p}_i - \boldsymbol{t}\|^2 \tag{3}$$

Although good results are obtained with free-form shapes, the main shortcoming of this algorithm is that it can only be used in surface-to-model registrations. However, in real applications, surfaces usually contain non-overlapping regions. A different version of the ICP algorithm was also proposed by Chen and Medioni [33]. The main difference between the two algorithms is the method used for finding the correspondences: Chen and Medioni used the point-to-plane distance while Besl and McKay used the point-to-point distance. Generally, the point-to-plane distance solves the problem of local minima better since it overcomes the difficulty of finding correspondence between the points of the two clouds. These techniques, known as fine registration techniques, converge fast enough but require a good initial solution. When the initial solution is not available, the coarse registration techniques can be used to estimate it. We refer the reader to a more in-depth comparison on surface registration recently published in [34]. In our

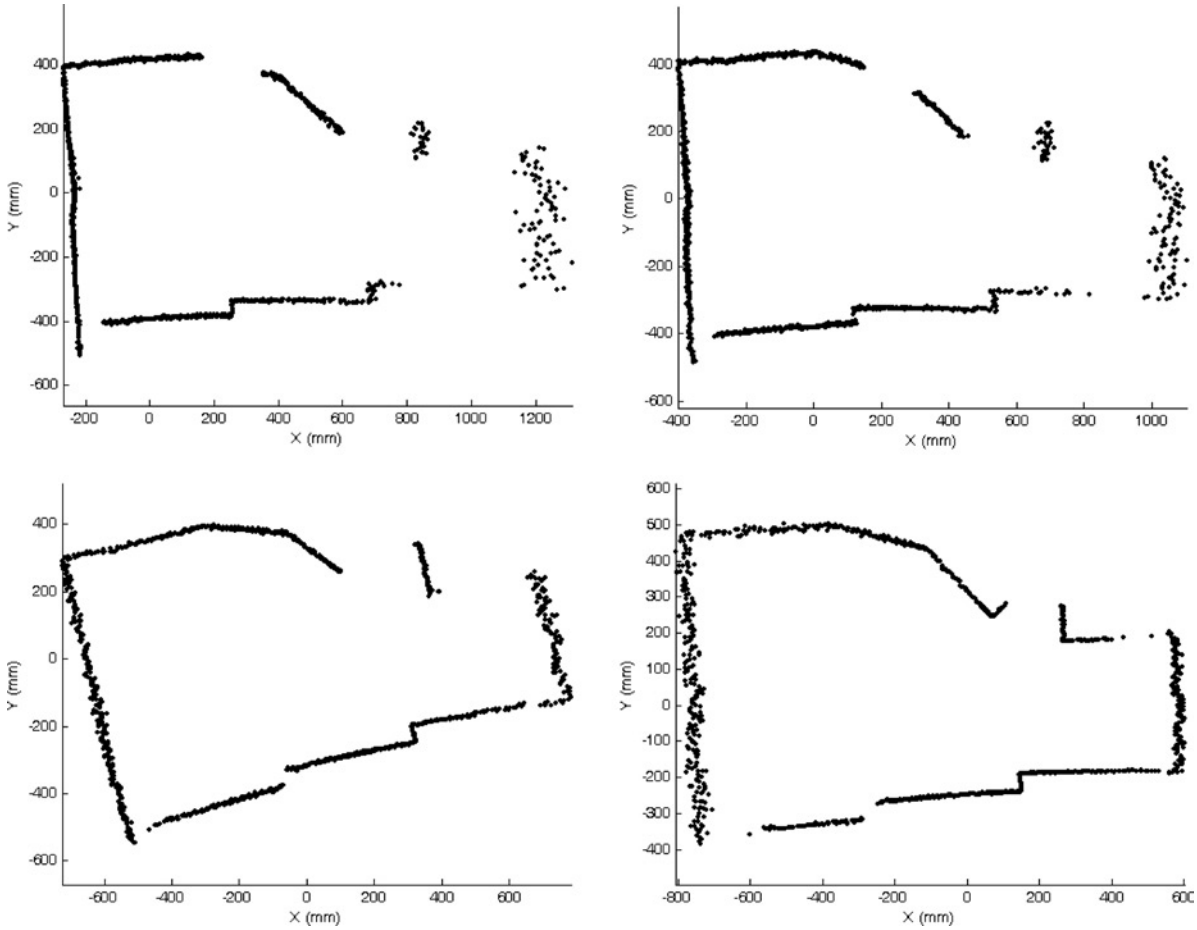

Fig. 6 *Four range scans calculated at different positions in a scene*

case, the motion between two consecutive positions is small and the wide FOV offers large overlapping areas which help the ICP algorithm converge even when the initial solution is the identity matrix. Therefore we chose the ICP algorithm proposed by Chen and Medioni.

A preliminary step for registration is transforming the coordinates of the 3D points of each range scan from the sensor to the laser plane reference system. Therefore the third coordinate can be ignored and the registration is performed in the 2D space of the laser plane. Consider the laser plane described by the equation $ax + by + cz = d$ and a range scan point in the sensor reference coordinate system $P_c(x_c, y_c, z_c)$. The transformation of the point P_c into the point $P_L(x_L, y_L, z_L)$ expressed in the laser plane reference frame is carried out using (4). The transformation matrix $\boldsymbol{R}_L$ and the translation vector $\boldsymbol{t}_L$ are described in (5).

$$\begin{aligned} P_L &= \boldsymbol{R}_L P_c + \boldsymbol{t}_L \\ \boldsymbol{R}_L &= \begin{bmatrix} X & Y & Z \end{bmatrix} \end{aligned} \tag{4}$$

where

$$\begin{aligned} X &= \begin{bmatrix} 1-0 \\ 1-0 \\ \dfrac{d-a-b}{c} - \dfrac{d}{c} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -\dfrac{a+b}{c} \end{bmatrix} \\ Z' &= \begin{bmatrix} a & b & c \end{bmatrix} \\ Y &= Z \times X \\ \boldsymbol{t}'_L &= \begin{bmatrix} 0 & 0 & z_L \end{bmatrix} = \begin{bmatrix} 0 & 0 & -\dfrac{d}{c} \end{bmatrix} \end{aligned} \tag{5}$$

Subsequently, the range scan points transformed in the 2D space are filtered through a normal space sampling algorithm [35]. The algorithm consists in transforming the (x_L, y_L) space into a grid of $(n_x \times n_y)$ cells and it selects only one element in each cell. Therefore the new space is formed by the sampled points $P_s(n_{xi}, n_{yi})$ and thus the weight of the points becomes inversely proportional to their density.

The distances from the sampled points of the current view i to the planes (or lines in the 2D space) obtained from the unsampled points of the previous view $i-1$ are calculated and the point-to-plane correspondences are established for all the points of the view i. Considering a point in the current view, its pair in the previous view is the nearest point in the corresponding plane.

Finally, the 2D registration is performed by minimising the distance between the two clouds of points using (3). The result of the registration of the range scans presented in Fig. 6 is shown in Fig. 7. The base line can be seen in Fig. 7*b*, in which the distance between the camera focal point and the laser plane is visible.

3.3 Line tracking and segment detection

Most of the indoor scenes are formed by human-made shapes, and the majority of these shapes contain vertical planes. Therefore a mobile robot that carries the sensor and inspects an indoor scene should take advantage of the inherent particularity of this environment. A line-tracking algorithm is used to extract the planes in the cloud of points obtained from the range scan registration. Then the extracted lines are divided into segments representing the projections of the vertical planes onto the laser plane.

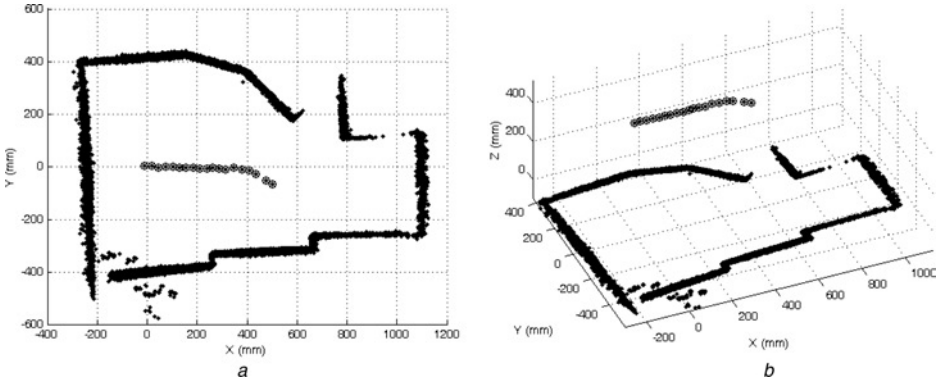

Fig. 7 *Cloud of points obtained by the registration of several partial range scans obtained by placing the sensor at different locations in the scene*

Camera view point is represented by the black sphere
a Projection on the *XY* plane
b 3D view

One or several registered panoramic range scans provided by the sensor can be transformed into a 2D binary matrix, which is analysed using a recursive Hough algorithm with a variable threshold adapted for the detection of the maximum number of lines.

All the lines identified in the image are bounded and the segments corresponding to planar surfaces in the scene are detected. The incorrect segments are filtered and removed. The 2D map obtained from the registered cloud of points is shown in Fig. 8.

3.4 Best pose evaluation

During the registration process, the transformation between the range scans taken at different locations is evaluated. Since the points of the range scans obtained in each shot are represented in the sensor coordinate system, the different poses of the sensor relative to each segment can be calculated. Therefore the best view pose of each segment is determined as being the sensor position from which the scene plane corresponding to the segment is observed at the most favourable angle. The image taken from the best pose is subsequently used for mapping the texture onto the scene objects.

The registration process provides the transformation matrix between the different range scans and thus the robot's position can be calculated with respect to any coordinate system. Fig. 8 shows the locations through which the mobile platform moved and that were retrieved by registration. The transformation between two locations is computed by (6), in which $\boldsymbol{R}_{\mathrm{L}}$ and $\boldsymbol{t}_{\mathrm{L}}$ are the rotation and the translation matrices, respectively, between the sensor coordinate system and the laser plane coordinate system, and $t_{n(3x1)}$ is the translation between the nth sensor position and the reference one.

$$\begin{bmatrix} x_{\mathrm{pos}} \\ y_{\mathrm{pos}} \\ z_{\mathrm{pos}} \\ 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{R}'_{\mathrm{L}(3x3)} & (-\boldsymbol{R}'_{\mathrm{L}}\boldsymbol{t}_{\mathrm{L}})(3x1) \\ 0_{(3x3)} & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{t}_{n(3x1)} \\ 1 \end{bmatrix} \quad (6)$$

Therefore the sensor location and the segments detected in the scene map can be represented with respect to the same reference. The optimal position for observing the texture of a given segment (i.e. planar object in the real scene) is the corner of an equilateral triangle containing the segment in one of its sides. The nearest robot location to the optimal position is chosen as the best pose (Fig. 8).

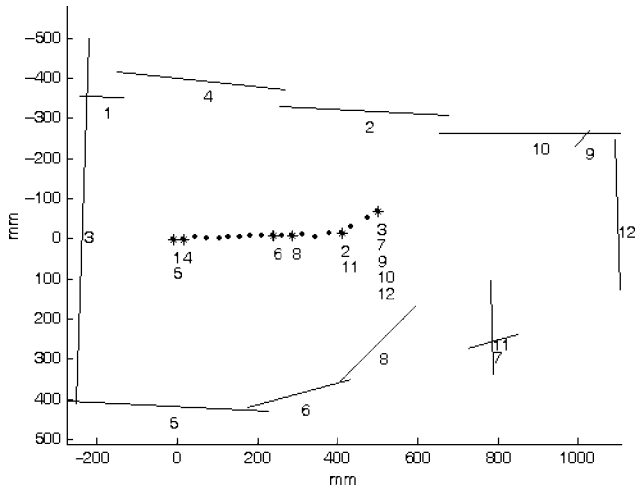

Fig. 8 *Best view for each segment in the scene*

Segments in the scene are labelled with numbers from 1 to 12
Positions of the robot are marked with dots, and the positions selected for mapping the texture on objects are marked with stars
Under each selected position, the labels of the segments that were observed best from that position are displayed

3.5 Texture mapping

Assuming that the analysed scene is composed mostly of vertical planar shapes placed on a flat floor surface, modelling the 3D scene is performed by generating the elevation map. The elevation map is obtained directly from the positions of the 2D segments by considering the vertical FOV of the omnidirectional camera (Fig. 9). In the following, the term 'maximum vertical FOV' will be used for describing the maximum height, hence on the vertical direction, visible by the catadioptric device. Although the lower bound of the scene objects is clearly given by the floor, the upper bound depends on the vertical FOV and is a function of the position of the robot with respect to the object. However, it is worth to remark that in practice, because of the low resolution of the catadioptric cameras, the robot has to get close to the target plane in order to obtain an accurate depth measurement and a clear texture image. Therefore the upper bound of the planes is implicitly limited.

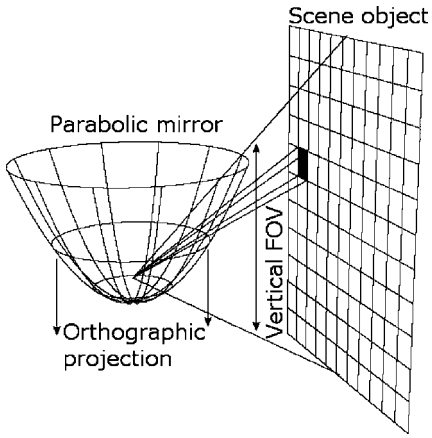

Fig. 9 *3D reconstruction of a plane in the scene*

Height of the reconstructed plane depends on the distance from the sensor and on the vertical FOV

The easiest way to build the scene is to use a single reference point and the maximum vertical FOV for each plane in the scene, as shown in Fig. 10*a*. Naturally, in the resulting 3D model, the planes nearby the sensor will be lower than the far planes. This model ensures an intuitive spatial view of the scene since high planes are immediately identified as being far from the robot. However, the large FOV of the parabolic mirror may create high planes artificially. A natural idea is to set the lower bound of the walls (Fig. 10*b*) since the robot is assumed to move on a flat surface.

Even though a single reference point can be used for the map elevation, the texture of the resulting models cannot be fully retrieved since there are planes that are not visible from the reference position. Fig. 10*c* shows the reconstruction of a scene from different points of view using the maximum vertical FOV at each position. This configuration ensures the best available texture for each object in the scene and an intuitive scene geometry. The final choice was to bound the vertical FOV only for the planes that are

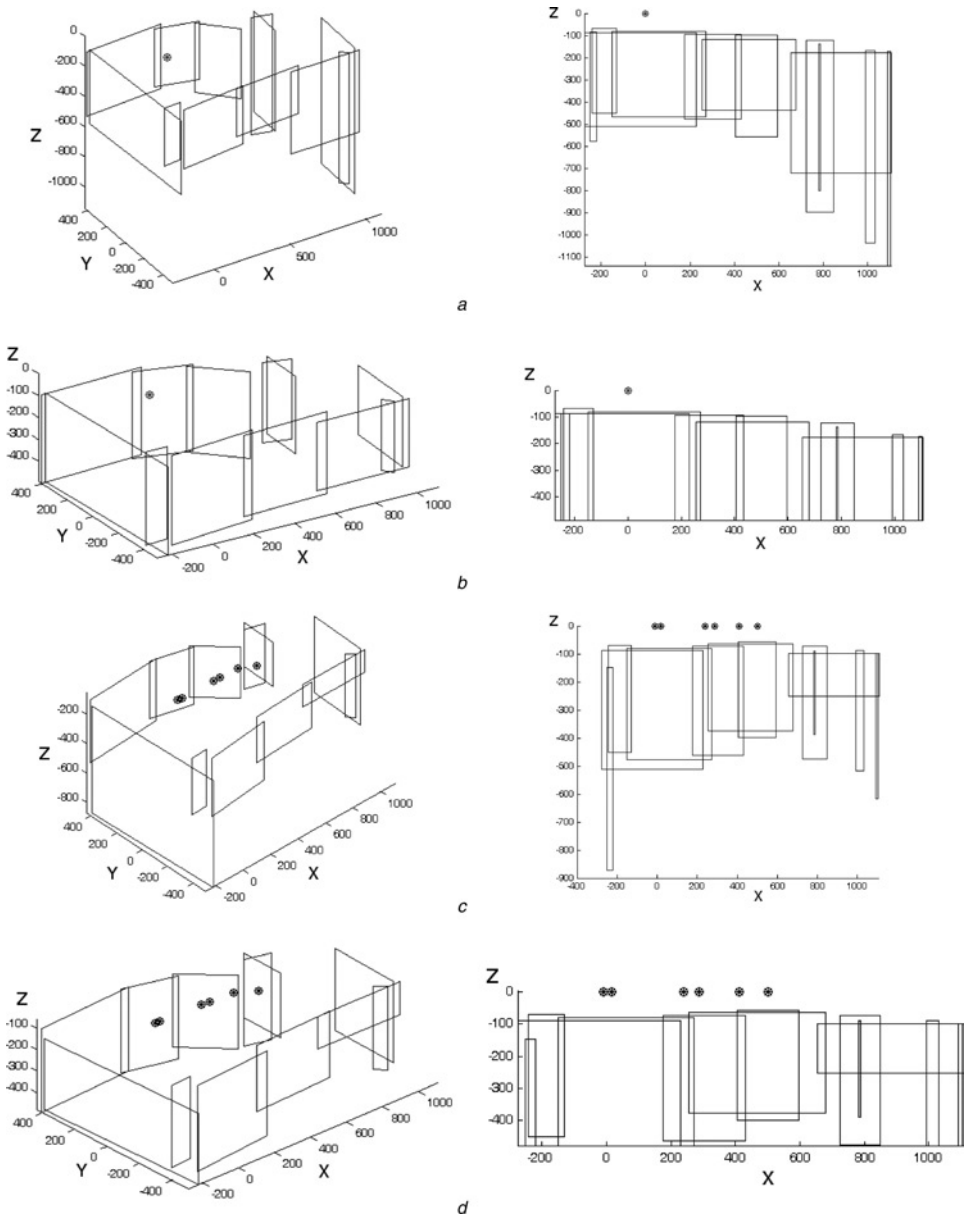

Fig. 10 *Different scene models*

a Scene from an SVP with maximum vertical FOV
b Scene from an SVP with constant vertical FOV
c Scene from multiple view points with maximum vertical FOV
d Scene from multiple view points with thresholded vertical FOV

'crossing the floor'. The resulting model is presented in Fig. 10*d*.

Finally, the texture is mapped on each plane of the 3D model using the information provided by the omnidirectional camera. The image taken from the location that ensures the best view of a plane is selected from among all the available images and is used for extracting the texture of the object. The texture patch is then stuck on the 3D model of the map.

4 Experimental results

The sensor was built using off-the-shelf components. The optics and the mirror used for the omnidirectional camera were provided by remote reality. The camera used was a Sony SSC-DC198P with a 1/3 in. CCD and a resolution of 768 × 576 pixels. The laser and its optics are produced by Lasiris, the diode power is 3 mW and produces red light with a wavelength of 635 nm.

The calibrated parameters of the camera model are listed in Table 1.

The height of the conic mirror used to build the laboratory prototype is $h = 4.4$ cm and the cone aperture angle is $\beta = 52^\circ$. The laser projects a circular cone with a fan angle $\alpha = 11.4^\circ$. Since the relation between the two angles is $\beta \simeq 0.5(\alpha + \pi/2)$, the laser is reflected along a very flat surface that can be approximated to a plane: $ax + by + cz + d = 0$. The laser parameters obtained by calibration are $a = -0.0212$, $b = 0.0019$, $c = 1$ and $d = 91.7291$.

Depth is perceived because of the disparity generated by the two components of the sensor placed along a baseline, as explained in the previous sections. At the design stage, the baseline was implicitly chosen because the laser stripe had to be visible in the central part of the omnidirectional image. After calibration, the length of the baseline was calculated to be 95.26 mm. Note that, because of the SVP property of the catadioptric camera, the baseline is constant and does not depend on the elevation of the observed scene point, which is not the case for non-SVP stereo catadioptric systems.

The calibrated sensor was located on a mobile platform and placed in a man-made scene. The structured light projector produces the light stripe in the scene, which is automatically detected with sub-pixel accuracy. A detail of the laser stripe is shown in Fig. 11*a*, in which the effects of the noise on the laser stripe are visible. The peaks detected are represented by dots in Fig. 11*b*, which are subsequently approximated to an ellipse in order to improve the process of detecting straight lines at the calibration stage.

In order to evaluate the sensor accuracy, we placed a planar surface at different known positions and its location was calculated (Fig. 12). The plane was placed at different distances from the sensor within a range from 40.64 cm (16 in.) to 78.74 cm (31 in.) and the error was calculated for each position.

Table 2 presents the sensor accuracy extracted from the ground truth evaluation.

The range scan provided by the sensor was analysed using a recursive Hough algorithm and the vertical planes in the scene were detected. Since the FOV of the camera is known, the elevation map illustrated in Fig. 13*a* can be generated. The maximum vertical FOV was used because of the fact that all the scene objects were close enough. The parts of the image corresponding to each plane in the scene have been unwrapped and two examples of these patches are shown in Fig. 13*b*.

Different points of view are obtained by changing the location of the sensor. The images captured by the omnidirectional camera at each location are shown in Fig. 14*a*, whereas the range scan obtained is depicted in Fig. 14*b*. Screen shots of the Virtual Reality Modeling Language (VRML) scene models are shown in Fig. 14*c*.

A more complex scene has been reconstructed using photographic and registered range scan data from 19 different points of view which permitted reconstructing the planes that are not visible from all the view points. Two screen shots of the VRML model are presented in Fig. 15.

5 Conclusions and future work

The aim of this work is to build the full 3D model of an indoor environment by means of a sensor that combines omnidirectional vision and structured light. The two omnidirectional systems that compose the sensor are calibrated and the resulting model is used to measure distances in a real scene. Finally, the scene model is reconstructed by moving the sensor in a scene containing planar surfaces and collecting data at different locations. The virtual scene is built using two informational streams: the range-finder scan data and the texture of the scene objects. Both data sources were used to reconstruct the scene, as they offer complementary hints about the environment. The position and orientation of the robot are also estimated in each step. The experimental results show that the shapes are properly retrieved and the textures are correctly mapped on the objects that surround the sensor. Mapping the texture on the corresponding objects enables the human operators to detect discontinuities that are not spotted by the laser pattern.

The single camera catadioptric system offers a wide FOV but has lower resolution than common cameras. Therefore such an omnidirectional device can be used for gathering information from the surrounding scene but is not suitable for very accurate detail detection. Following this logic, we conclude that the proposed sensor is appropriate for applications that need a large FOV and fairly low accuracy such as human-made environment modelling that is useful for robot navigation and remote inspections.

An inherent disadvantage of the proposed sensor is that it can recover only one line of 3D dots at a given position. The first prototype presented here can be improved by the use of a more complex-structured light pattern. For instance, a set of concentric circles or colour-coded light can increase the vertical FOV. Thus, full cylindrical surfaces of 3D dots could be recovered. Another shortcoming of the proposed sensor is its range. The laser used for the prototype

Table 1: Calibrated parameters for the omnidirectional camera

ξ	φ	α_u	α_v	u_0	v_0	α	β	γ	t_x	t_y	t_z
1.07	−6.12	−55.10	56.84	394.23	292.20	−3.08	−3.12	−3.13	10.22	20.94	−444.29

Measurement units: mirror parameters (ξ and φ) – mm; intrinsic parameters (α_u, α_v, u_0 and v_0) – pixels; rotation angles (α, β and γ) – radians; translations (t_x, t_y and t_z) – mm

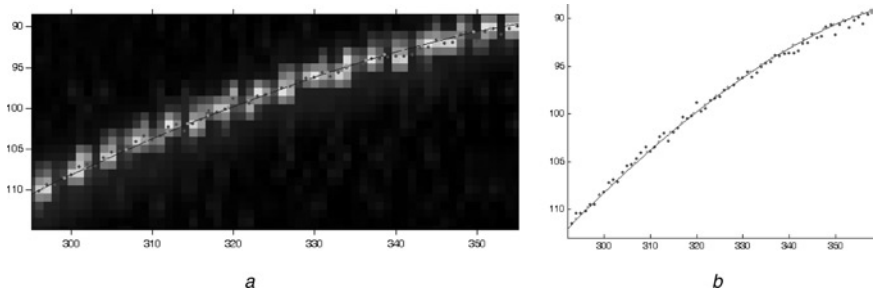

Fig. 11 *Peak detection of a laser stripe in the omnidirectional image*

a Detail of the real image
b Real and approximated points in the omnidirectional image

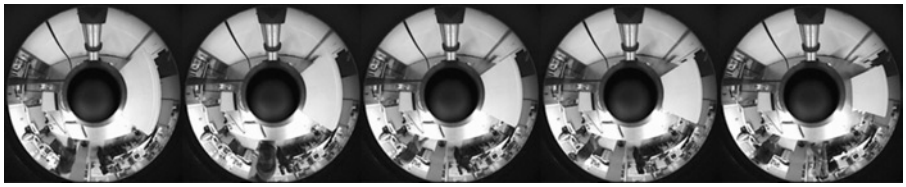

Fig. 12 *Planar surface, placed at different locations, imaged by the omnidirectional camera*

Table 2: Sensor accuracy within different ranges

Range	Minimum error	Maximum error	Mean error	Standard deviation
400–450	4.7	24	6.2	3.1
450–500	5.1	21	9.1	2.2
500–550	6.5	18	11.5	3.9
550–600	8.2	22.5	14.1	5.6
600–650	9.1	26.8	15.6	7.3
650–700	11	32.5	17.3	6.9
700–750	14.2	39.8	19.9	9.5
750–800	16.3	47.1	22.5	11.3

The presented error is the average distance between the 3D points reconstructed by the sensor and the reference plane. The values are given in millimetres

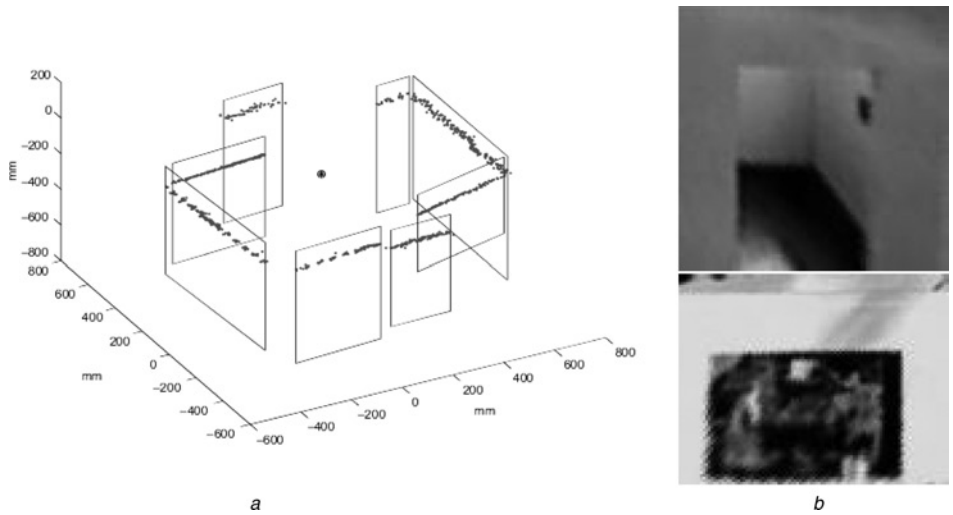

Fig. 13 *Elevation map and texture unwrapping*

a Elevation map of the experimental scene. The position of the camera is marked by a black sphere
b Examples of unwrapped image parts

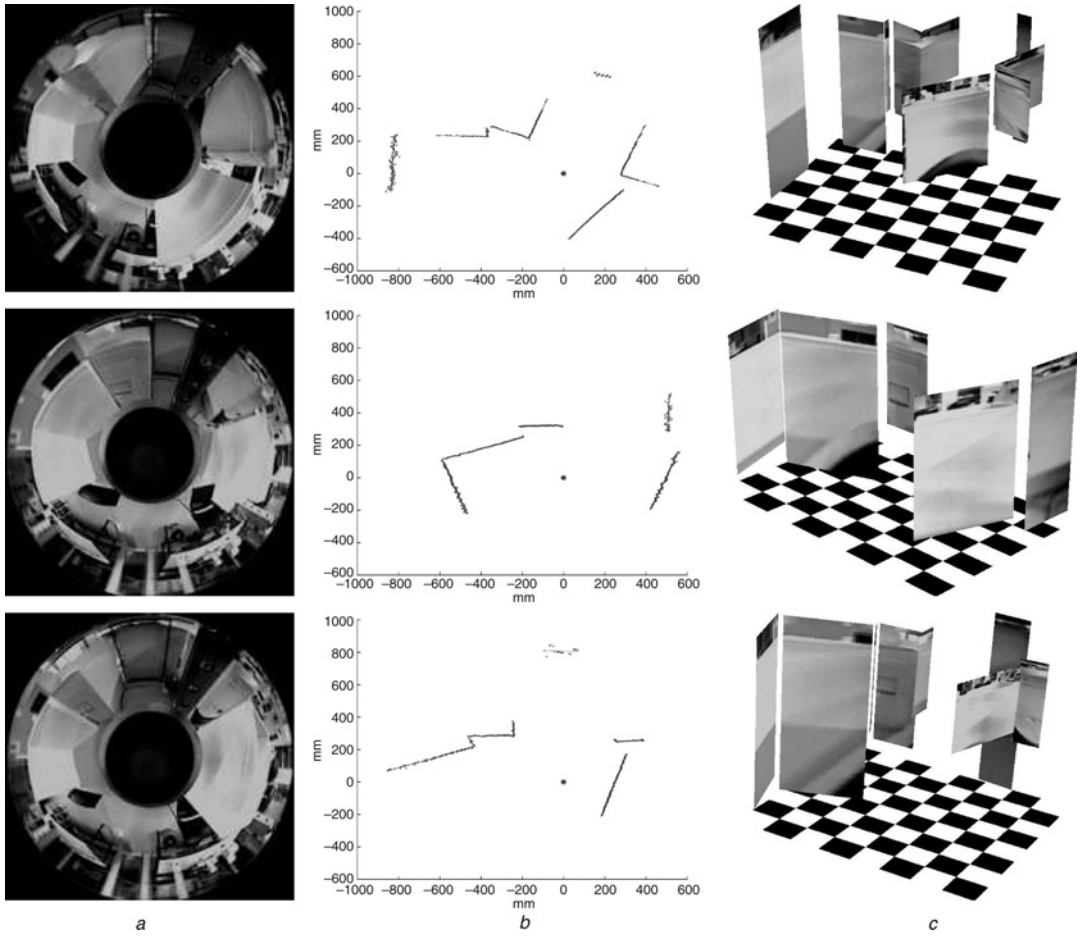

Fig. 14 *Scene modelling with the sensor placed at different locations*

a Images captured by the omnidirectional camera
b Range scans provided by the sensor at each position. The position of the sensor is marked by a black sphere
c Texture mapped on the surrounding planes

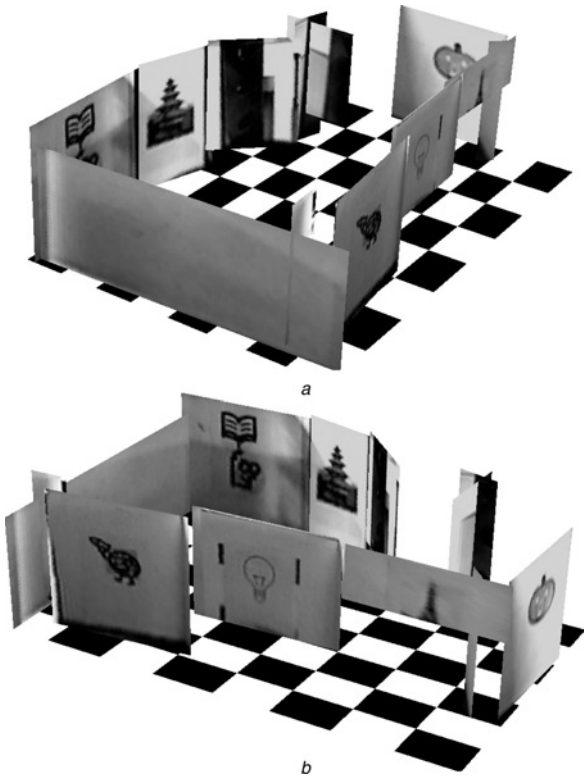

Fig. 15 *3D reconstruction of a scene from 19 omnidirectional views*

presented here is not visible at distances greater than 80 cm. A stronger laser would overcome this problem. The only concern about increasing the laser strength is the eye safety restrictions that would have to be applied in the environment where the sensor was used.

Finally, we would like to highlight some of the advantages of the proposed solution for omnidirectional scene modelling: it is not intrusive since there are no modifications of the scene; unlike traditional stereo, the camera synchronisation is not an issue since only a single camera is used; the sensor is more robust and easier to calibrate since it is compact and does not have mobile elements and, last but not least, it is cheaper than other similar structures since we use only two mirrors, one camera and a simple laser projector.

In conclusion, in this paper, we have presented a new single-shot 3D range-acquisition sensor which provides a 360° panoramic image of the scene surrounding the robot, which is useful for a wide variety of applications such as robot navigation, pipe inspections and so on.

6 Acknowledgment

This work was partially supported by the Spanish project CICYT TIC 2003-08106-C02-02 and by the AIRE mobility grant provided by the Government of Catalonia, which allowed the first author to have a 4 month stay at the University of Picardie.

7 References

1 Svoboda, T., and Pajdla, T.: 'Panoramic cameras for 3d computation'. Proc. Czech Pattern Recognition Workshop, 2000, pp. 63–70

2 Yagi, Y.: 'Omnidirectional sensing and its applications', *IEICE Trans. Inf. Syst.*, 1999, **E82-D**, pp. 568–578

3 Yagi, Y., and Yachida, M.: 'Real-time generation of environmental map and obstacle avoidance using omnidirectional image sensor with conic mirror', *Comput. Vis. Pattern Recognit.*, 1991, pp. 160–165

4 Southwell, D., Vandegriend, B., and Basu, A.: 'A conical mirror pipeline inspection system'. IEEE Int. Conf. on Robotics and Automation, 1996, vol. 4, pp. 3253–3258

5 Svoboda, T., Pajdla, T., and Hlavac, V.: 'Epipolar geometry for panoramic cameras'. Fifth European Conf. on Computer Vision, 1998, pp. 218–232

6 Bunschoten, R., and Kröse, B.: 'Robust scene reconstruction from an omnidirectional vision system', *IEEE Trans. Robot. Autom.*, 2003, **19**, (2), pp. 351–357

7 Fleck, S., Busch, F., Biber, P., Straber, W., and Andreasson, H.: 'Omnidirectional 3d modeling on a mobile robot using graph cuts'. Int. Conf. on Robotics and Automation, 2005, pp. 1760–1766

8 Shirai, Y., Miura, J., and Negishi, Y.: 'Mobile robot map generation by integration omnidirectional stereo and laser range finder'. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, 2002, pp. 250–255

9 Negishi, Y., Miura, J., and Shirai, Y.: 'Mobile robot navigation in unknown environments using omnidirectional stereo and laser range finder'. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, 2004, pp. 2737–2742

10 Fiala, M., and Basu, A.: 'Panoramic stereo reconstruction using non-svp optics'. 16th Int. Conf. on Pattern Recognition, 2002, vol. 4, pp. 27–30

11 Nene, S.A., and Nayar, S.K.: 'Stereo with mirrors'. Sixth Int. Conf. on Computer Vision, 1998, pp. 1087–1094

12 Southwell, D., Basu, A., Fiala, M., and Reyda, J.: 'Panoramic stereo'. 13th Int. Conf. on Pattern Recognition, 1996, vol. 1, pp. 378–382

13 Gluckman, J., and Nayar, S.K.: 'Planar catadioptric stereo: geometry and calibration'. IEEE Computer Vision and Pattern Recognition, 1999, vol. 1, pp. 22–28

14 Gluckman, J., and Nayar, S.K.: 'Rectified catadioptric stereo sensors', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, (2), pp. 224–236

15 Kawanishi, T., Yamazawa, K., Iwasa, H., Takemura, H., and Yokoya, N.: 'Generation of high-resolution stereo panoramic images by omnidirectional imaging sensor using hexagonal pyramidal mirrors'. Fourteenth Int. Conf. on Pattern Recognition, 1998, vol. 1, pp. 485–489

16 Gluckman, J., Nayar, S.K., and Thoresz, K.J.: 'Real-time omnidirectional and panoramic stereo'. DARPA98, 1998, pp. 299–303

17 Lin, S.S., and Bajcsy, R.: 'High resolution catadioptric omni-directional stereo sensor for robot vision'. IEEE Int. Conf. on Robotics and Automation, 2003, pp. 1694–1699

18 Spacek, L.: 'A catadioptric sensor with multiple viewpoints', *Robot. Auton. Syst.*, 2005, **51**, pp. 3–15

19 Pagès, J., Salvi, J., and Forest, J.: 'Optimized de Bruijn patterns for one-shot shape acquisition', *Image Vis. Comput.*, 2005, **23**, (8), pp. 707–720

20 Salvi, J., Pagès, J., and Batlle, J.: 'Pattern codification strategies in structured light systems', *Pattern Recognit.*, 2004, **37**, (4), pp. 827–849

21 Mei, C., and Rives, P.: 'Calibration between a central catadioptric camera and a laser range finder for robotic applications'. IEEE Int. Conf. on Robotics and Automation, 2006, pp. 532–537

22 Orghidan, R., Mouaddib, E., and Salvi, J.: 'A computer vision sensor for panoramic depth perception', *Lect. Notes Comput. Sci.*, 2005, **3522**, (I), pp. 153–160

23 Orghidan, R., Salvi, J., and Mouaddib, E.: 'Modelling and accuracy estimation of a new omnidirectional depth computation sensor', *Pattern Recognit. Lett.*, 2006, **27**, pp. 843–853

24 Daniilidis, K., and Geyer, C.: 'A unifying theory for central panoramic systems and practical applications'. IEEE Sixth European Conf. on Computer Vision, 2000, pp. 445–461

25 Geyer, C., and Daniilidis, K.: 'Equivalence of catadioptric projections and mappings of the sphere'. First Workshop of Omnidirectional Vision, 2000, pp. 91–96

26 Orghidan, R., Mouaddib, E.M., and Salvi, J.: 'Omnidirectional depth computation from a single image'. IEEE Int. Conf. on Robotics and Automation, 2005, pp. 1234–1239

27 Salvi, J., Armangue, X., and Batlle, J.: 'A comparative review of camera calibrating methods with accuracy evaluation', *Pattern Recognit.*, 2002, **35**, (7), pp. 1617–1635

28 Lin, S.S., and Bajcsy, R.: 'The true single view point (svp) configuration for omni-directional view catadioptric system using cone mirror'. 'Technical Report' ms-cis-00-24, Computer and Information Science Department, University of Pennsylvania, Philadelphia, USA, 2001

29 Baker, S., and Nayar, S.K.: 'A theory of catadioptric image formation'. IEEE Sixth Int. Conf. on Computer Vision, 1998, pp. 35–42

30 Ahn, S.J.: 'Least-squares orthogonal distance fitting of curves and surfaces in space, (Springer-Verlag GmbH, 2004), vol. 3151

31 Forest, J., Salvi, J., Cabruja, E., and Pous, C.: 'Laser stripe peak detector for 3d scanners. a fir filter approach'. Int. Conf. on Pattern Recognition, 2004, vol. 3, pp. 646–649

32 Besl, P., and McKay, N.: 'A method for registration of 3-d shapes', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1992, **14**, pp. 239–256

33 Chen, G., and Medioni, Y.: 'Object modeling by registration of multiple range images'. IEEE Int. Conf. on Robotics and Automation, 1991, pp. 2724–2729

34 Salvi, J., Matabosch, C., Fofi, D., and Forest, J.: 'A review of recent range image registration methods with accuracy evaluation', *Image Vis. Comput.*, 2007, **25**, pp. 578–596

35 Rusinkiewicz, S., and Levoy, M.: 'Efficient variant of the icp algorithm'. 3rd Int. Conf. on 3-D Digital Imaging and Modeling, 2001, pp. 145–152