3D volumetric reconstruction with a catadioptric stereovision sensor

Romain Rossi, Nicolas Ragot, Xavier Savatier, Jean-Yves Ertaud, Bélahcène Mazari

3D Volumetric Reconstruction with a Catadioptric Stereovision Sensor

N. Ragot, R. Rossi, X. Savatier, J-Y. Ertaud, B. Mazari
Institut de Recherche en Systèmes Electroniques EMbarqués
Ecole Supérieure d'Ingénieurs en Génie Electrique
Technopole du Madrillet - Av. Galilée - BP10024
76801 Saint Etienne du Rouvray Cedex - France
Email: ragot,rossi,savatier,ertaud,mazari@esigelec.fr

***Abstract*— A stereoscopic catadioptric sensor and associated algorithms are presented to provide a relevant solution to the problem of $3D$ reconstruction of unknown environments. A novel calibration methodology is detailed, establishing a discrete relation between $3D$ points and the corresponding pixels. The $3D$ reconstruction adopted is an adaptation of the classic volumetric scene reconstruction technique. Complete voxels are projected onto the image planes and consistency measurements are computed on pixel surfaces. To ensure fast processings, $3D/2D$ matchings are processed offline and results are stored in Look-Up Tables (LUT). Calibration results are given and $3D$ reconstruction algorithms are validated on synthetic images. Some results obtained for real scenes are provided.**

I. Introduction

The determination of the $3D$ geometric structure of an unknown environment is a traditional problem in computer vision. Many applications, such as path generation, obstacle avoidance, virtual reality, etc. require a complete reconstruction of a scene. To achieve this, some works deal with camera networks or rotating cameras associated with mosaicing processes [1], [2]. Our attention is focused on *catadioptric* sensors, i.e camera/mirror combination, for their capacity to provide a 360 degrees panoramic image in a single shot [3].

Traditionally, scene reconstruction techniques are based on image matchings using correlation algorithms associated with triangulation and surface fitting. Nevertheless, the matching stage is a very difficult problem because scene path has different shapes and appearances when seen from different viewpoints and at different times [4]. Another approach, consists on a beforehand discretization of a $3D$ space into elementary volumes (i.e. voxels). Computations are based on voxel image consistency and visibility. This method is known as *volumetric scene modeling* and provides a relevant alternative to the drawbacks of stereovision techniques [5].

This paper gives details about an architecture and the associated algorithms for a vision sensor offering a solution to the problem of the $3D$ reconstruction of an environment with no prior knowledge. The main contributions of this paper deal with an innovative calibration methodology based on a $3D/2D$ mesh and the implementation of an adapted $3D$ volumetric reconstruction for a stereoscopic panoramic sensor. The paper is organized as follows: in §II the sensor is described. §III presents the calibration methodology. §IV gives details about the $3D$ reconstruction algorithm. §V presents the results and §VI gives comments. The conclusion is given in §VII.

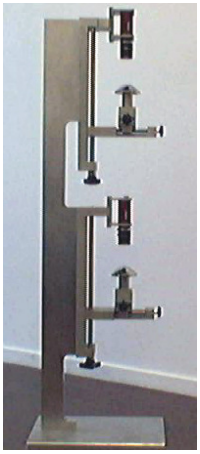
II. Sensor Architecture

A. Catadioptric Sensors

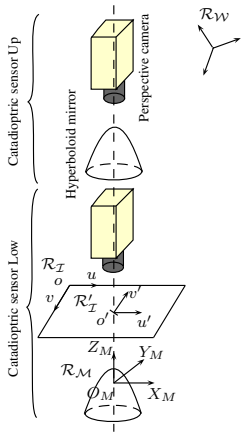
In any imaging system the uniqueness of the center of projection (viewpoint) is highly desirable. For catadioptric sensors, this convergence property is known as the *Single Effective Viewpoint Constraint* (SEVC). It enables the formation of geometrically pure perspective images and allows easier modeling of catadioptric sensors. According to the SEVC, we distinguish on the one hand, the *non-central* catadioptric sensors for which multiple viewpoints are obtained and caustic shapes have to be considered [6], and on the other hand, the *central* ones. [3] derived all the classes of mirrors for which the SEVC is respected. One of these solutions, and the one retained for the development of the sensor, is the association of a hyperboloidal mirror with a perspective camera. Nevertheless, a particular positioning must be accomplished which is difficult to carry out [7].

B. Mechanical Configuration

As the main objective of the study is the $3D$ reconstruction of a scene, we have developed a stereoscopic panoramic bench made of two hyperboloidal catadioptric sensors (cf. Fig. 1). There are various mechanical configurations for mirror/camera combinations. [8] uses multiple mirrors observed by a single camera. Compactness and the absence of camera synchronization problems are the main advantages of this sensor. Nevertheless, the smallness of the baseline does not allow the reconstruction of a large zone. We favor a mechanical structure similar to the one proposed by [9]. The two catadioptric sensors are combined in a co-axial configuration. In such a configuration the epipolar geometry of the sensor is simplified: the epipolar curves are radial lines in the omnidirectional image and columns in the panoramic image (obtained by a cylindrical projection). Moreover, the baseline can be adjusted, which means that the reconstructed area can be user-controlled according to the targeted applications.



(a)



(b)

Fig. 1. Illustration of the panoramic stereovision sensor. (a): Picture of the system - (b): Schema of the stereoscopic device. Details about catadioptric sensors are given: It consists on a classic camera observing a revolution hyperboloidal mirror. The different coordinate systems used in this paper are given (i.e. $\mathcal{R}_\mathcal{W}$, $\mathcal{R}_\mathcal{M}$, $\mathcal{R}_\mathcal{I}$, $\mathcal{R}'_\mathcal{I}$)

III. Panoramic Stereovision Sensor Calibration

A. Brief Review of Catadioptric Sensor Calibration Methods

Calibration of catadioptric sensors is the most studied aspect in omnidirectional vision. For more details, the reader can refer to [10]. We distinguish *intrinsic* calibration which provides the intrinsic parameters of the camera [7], *weak* calibration available for stereovision sensors [11] and *strong* calibration. This last category uses an external calibration pattern from which features are extracted and matched with their corresponding $2D$ points. Traditionally this method is based on a model (generic or ad-hoc) whose parameters are estimated using optimization algorithms.

B. Calibration Pattern

To calibrate the sensor, a cylindrical calibration tool is used. The calibration tool is made of luminous markers (LEDs) whose $3D$ coordinates are known in the frame of the calibration tool (i.e. $\mathcal{R}_\mathcal{W}$ located at the bottom center of the cylinder). A hypothesis is that the revolution axis of the cylinder and the optical axis of the sensor are in common. The markers are distributed along eight vertical bars describing the direction circles of the cylinder entirely. According to the calibration methodology adopted (i.e. $3D/2D$ mesh) and the mirror shapes of the catadioptric sensors, the position of the markers and their quantity are critical parameters. A beforehand prototyping stage is provided to determine a relevant vertical distribution of the markers onto the cylinder. Moreover, a technique to densify the number of vertices of the mesh belonging to direction circles of the cylinder is presented.

1) Horizontal distribution: To maximize the number of nodal points constituting the mesh (i.e. to increase the precision of the calibration), artificial vertices are created. Calibration images are taken so that the markers defining the direction circles of the cylinder are illuminated simultaneously. Theoretically, direction circle projections define circles on the image planes. Nevertheless, due to the effect of noise, theirs projections are not perfectly circles and conic curves must be considered. Let $\mathbf{p}_i = [u_i, v_i]$ be the projections of the $3D$ markers belonging to a direction circle of the cylinder. On the image plane, $\mathbf{p}_i$ lie on a conic $\mathbf{\Omega} = [A, B, C, D, E, F]^T$. The problem of conic fitting can be written as (eq. 1), and a solution is given by the normalized eigenvector of $\mathbf{A}^T\mathbf{A}$ corresponding to the smallest eigenvalue [12]. Once $\mathbf{\Omega}$ is identified, the conic curve is sampled following an angular parameter. Artificial $3D$ points corresponding to the $2D$ conic sampling points are created and the mesh is horizontally densified.

$$\underbrace{\begin{bmatrix} u_1^2 & 2u_1v_1 & v_1^2 & 2u_1 & 2v_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_i^2 & 2u_iv_i & v_i^2 & 2u_i & 2v_i & 1 \end{bmatrix}}_{\mathbf{A}} \mathbf{\Omega} \neq 0 \qquad (1)$$

2) Vertical distribution: The purpose is to determine a relevant distribution of the markers onto the calibration pattern. First of all a *High Density Mesh* (HDM) is created so that consecutive markers are vertically distributed with a constant and small spacing. Due to the mirror shape, spacings between marker projections are not constant. To validate the mesh, an average euclidean projection error (ϵ_{2D} eq. 2) is computed between projections ($\mathbf{p}_i = [u_i, v_i]$) of a sample of $3D$ points (S) randomly located in a scene and projections obtained using the *interpolation-based* calibration method ($\widehat{\mathbf{p}}_i = [\widehat{u}_i, \widehat{v}_i]$).

$$\epsilon_{2D} = \frac{1}{S}\sum_{i=1}^{S}\sqrt{(u_i - \widehat{u}_i)^2 + (v_i - \widehat{v}_i)^2} \qquad (2)$$

Secondly, based on the HDM calibration data, an appropriate distribution of markers onto the calibration pattern is defined. A sample of $2D$ points regularly spaced on a radius of a omnidirectional image is back-projected onto the calibration pattern. Several simulations are carried out in order for ϵ_{2D} and the number of markers (N) to be minimal.

C. Interpolation-Based Method

1) Calibration principle: A catadioptric sensor provides a circular image of the environment. As illustrated in Fig. 2, a pixel $\mathbf{p}' = [u', v']$ in $\mathcal{R}'_\mathcal{I}$ is derived from $\mathbf{p} = [u, v]$ expressed in $\mathcal{R}_\mathcal{I}$, so that: $u' = u - u_0$, $v' = v_0 - v$ with $\mathbf{o}' = [u_0, v_0]$ the image center coordinates. Also, $\mathbf{p}'$ can be defined in polar coordinates, with θ $(-\pi \leq \theta \leq \pi)$ as its azimuth angle and $r = ||\mathbf{o}'\mathbf{p}'||$ as its radius. The only assumption made is the parallelism between the plane $(X_MO_MY_M)$ and the retinal plane of the camera. In the mirror frame $\mathcal{R}_\mathcal{M}$, θ defined the position of a $3D$ point $\mathbf{P} = [X, Y, Z]$ in the plane $(X_MO_MY_M)$. The orthogonal projection of $\mathbf{P}$ in the plane $(X_MO_MY_M)$ is denoted as $\mathbf{P}'$ and θ is defined by (eq. 3):

$$\begin{cases} \theta = \arccos\left(\frac{X}{\sqrt{X^2+Y^2}}\right) & \text{if} \quad Y \geq 0 \\ \theta = -\arccos\left(\frac{X}{\sqrt{X^2+Y^2}}\right) & \text{else} \end{cases} \qquad (3)$$

Moreover, r is linked to the vertical position (Z) of the $3D$ point. There is a close relationship between r and φ, with φ ($-\frac{\pi}{2} \leq \varphi \leq \frac{\pi}{2}$) the angle of the position vector of a $3D$ point in the plane $(P'O_M P)$. From the geometry of the vision sensor, φ and the cartesian coordinates of the $3D$ point are linked by a simple trigonometric relationship (eq. 4):

$$\varphi = \arctan\left(\frac{Z}{\sqrt{X^2+Y^2}}\right) \tag{4}$$

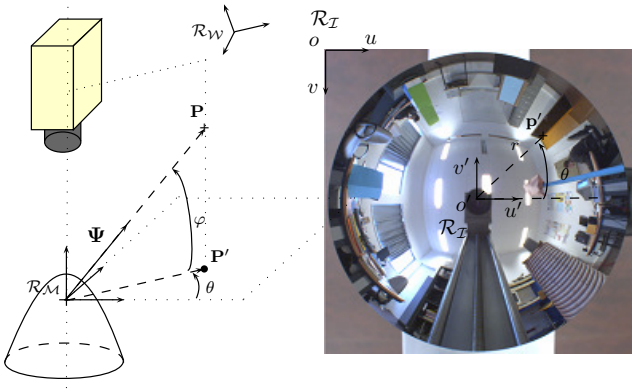


Fig. 2. Illustrations of : - the relation linking a $3D$ point $\mathbf{P}$ to its matching pixel $\mathbf{p}'$. - the relation linking a pixel $\mathbf{p}'$ to its corresponding ray of light $\mathbf{\Psi}$

Thus, the $3D/2D$ matchings depend on the variables (θ, φ, r). A projection function $\mathcal{F}$ can be defined, linking $2D$ image points to their corresponding rays of light. Nevertheless, the calibration procedure gives only partial knowledge of the $3D/2D$ matchings at the vertices of the mesh and $\widetilde{\mathcal{F}}$, the discrete function associated to $\mathcal{F}$, is defined by (eq. 5):

$$\begin{cases} 2D = \mathcal{F}(3D) & \Rightarrow r = \widetilde{\mathcal{F}}(\theta, \varphi) \\ 3D \cong \mathcal{F}^{-1}(2D) & \Rightarrow \varphi \cong \widetilde{\mathcal{F}}^{-1}(\theta, r) \end{cases} \tag{5}$$

To establish all of the entire $3D/2D$ and $2D/3D$ matchings, a local interpolation is provided in the vicinity of the nodal points defining the mesh.

2) $3D/2D$ and $2D/3D$ matching recovery:

- $3D \rightarrow 2D$: Starting with a known $3D$ point ($\mathbf{P} = [X, Y, Z]$) in $\mathcal{R}_\mathcal{M}$, the purpose is to determine its corresponding pixel $\mathbf{p}' = [\theta, r]$ in $\mathcal{R}'_\mathcal{I}$. According to $\widetilde{\mathcal{F}}$, $\mathbf{P}$ is also defined by: $\mathbf{P} = [\theta, \varphi, r]$. θ and φ are derived from (3) and (4). The problem consists in determining the coordinate r. We apply a Delaunay triangulation in the plane $(\theta O \varphi)$ onto the vertices $\mathbf{V}_i$. This enables the selection of three nodes for which $\mathbf{P}$ is inside the triangle. The three selected vertices define a plane (Δ) whose equation is given by (eq. 6):

$$\Delta \ : \ r = a\theta + b\varphi + c \tag{6}$$

By inserting the coordinates of the three vertices into (eq. 6) and solving the linear system, the plane equation parameters (a, b, c) are determined. Once the plane parameters are identified, r is easily computed and so, the pixel $\mathbf{p}'$ in $\mathcal{R}'_\mathcal{I}$ is entirely defined (cf. Fig. 3).

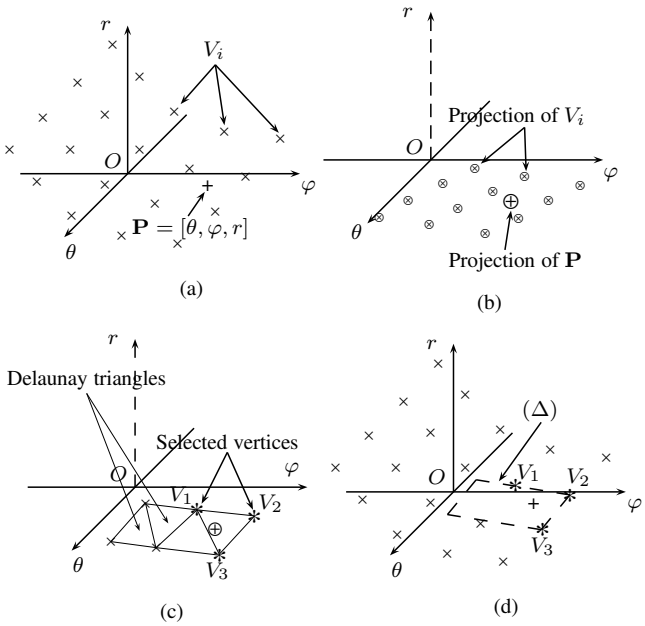


Fig. 3. Illustration of the *interpolation-based* calibration method and the recovery of $3D \rightarrow 2D$ correspondences. (a): Illustration of the projection function $\widetilde{\mathcal{F}}$. $\times$ symbolize the vertices of the mesh ($\mathbf{V}_i$). $\mathbf{P} = [\theta, \varphi, r]$ symbolized by $+$ belongs to $\widetilde{\mathcal{F}}$ - (b): Projection of the vertices onto the plane $(\theta O \varphi)$ represented by $\otimes$. Projection of $\mathbf{P}$ is represented by $\oplus$ - (c): Delaunay triangulation applied onto the projected vertices. Three nodes for which $\mathbf{P}$ is inside a triangle are selected $(*)$. (d): Plane (Δ) defined by the three vertices $(V_1 V_2 V_3)$

- $2D \rightarrow 3D$: The problem is similar to the one detailed previously. Starting with a known pixel $\mathbf{p}' = [\theta, r]$ in $\mathcal{R}'_\mathcal{I}$, the aim is to determine the direction vector $\mathbf{\Psi}$ of the corresponding ray of light. Considering $\mathbf{\Psi}$ as a unit direction vector, its coordinates are defined by (eq. 7), and so, the objective is φ component computation.

$$\mathbf{\Psi} = \begin{pmatrix} \sin\varphi\cos\theta \\ \sin\varphi\sin\theta \\ \cos\varphi \end{pmatrix} \tag{7}$$

According to $\widetilde{\mathcal{F}}^{-1}$, the Delaunay triangulation is established in the plane $(\theta O r)$. Three vertices $\mathbf{V}_i$ are selected and a plane is defined ($\Delta' \ : \ \varphi = a'\theta + b'r + c'$). Once the plane parameters (a', b', c') are computed, φ is easily determined. Thus, $\mathbf{\Psi}$ expressed in $\mathcal{R}_\mathcal{M}$ is defined.

3) Calibration validation: To validate the calibration methodology, the complete stereoscopic panoramic sensor is evaluated. Back-projection error ϵ_{3D} (eq. 8) is computed between real $3D$ coordinates of $3D$ points ($\mathbf{P}_i = [X_i, Y_i, Z_i]$) randomly distributed in a scene and the coordinates of the $3D$ points ($\widehat{\mathbf{P}}_i = [\widehat{X}_i, \widehat{Y}_i, \widehat{Z}_i]$) obtained by the inverse projection function $\widetilde{\mathcal{F}}^{-1}$ associated with a *mid-point triangulation* method.

$$\epsilon_{3D_i} = \frac{\sqrt{\left(X_i - \widehat{X}_i\right)^2 + \left(Y_i - \widehat{Y}_i\right)^2 + \left(Z_i - \widehat{Z}_i\right)^2}}{\sqrt{X_i^2 + Y_i^2 + Z_i^2}} \tag{8}$$

IV. $3D$ RECONSTRUCTION

Using an omnidirectional sensor, we can observe the surrounding environment and build a $3D$ model of the scene from a few pictures. In our application, it is assumed that the scene is unknown (i.e. scene model objects are not available). In such conditions, a generic reconstruction method is preferred. Voxel coloring algorithms are known to be useful in such situations [5].

A. Volumetric reconstruction framework

Many applications use voxel coloring to obtain photo-realistic reconstructions [13]. But the simplicity of this algorithm makes it suitable to obtain a rough geometric approximation of the scene objects. Voxel coloring algorithm uses a discretized volume as a representation of the $3D$ scene. To process the reconstruction, each voxel of the discretized scene is virtually projected onto the image planes of the cameras. This projection is made according to the calibration data. A voxel can be projected onto the images with various shapes and sizes, depending on its distance from the sensor and its viewing angle. Three different approaches are defined to project the voxels [14]. The first considers only the center of gravity of the voxel. The second approximates the projection as a simple shape (ex. a rectangle). The third processes the exact voxel projection. Once the voxel is projected onto each image, its projections are compared to determine if the projected voxel is consistent across all its projections. Many different algorithms can be used to make this comparison [15]. Our attention is focused on a photoconsistency-test similar as the one developed by [13]. The algorithm processes each voxel and consistent ones are marked *full*. When all the voxels of the discretized scene have been processed, the $3D$ reconstruction is obtained with the set of filled voxels.

There are two major key points with this algorithm. The first is the consistency test which determines the reconstruction quality. The second is the $3D/2D$ projection which is usually time-consuming. Our contribution focuses on the adaptation of the voxel coloring algorithm for an omnidirectional stereoscopic sensor. We developed algorithms to provide fast $3D/2D$ projections and consistency-tests for decreasing reconstruction times.

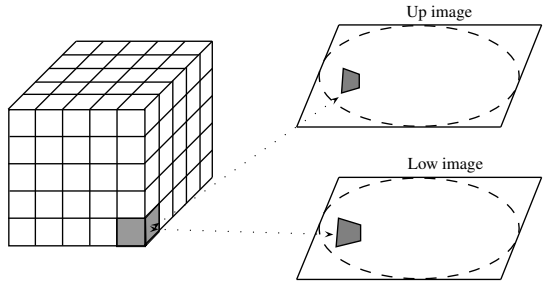


Fig. 4. Illustration of the *scene discretization* and $3D \rightarrow 2D$ projection

B. Methodology

Our reconstruction method is based on an adapted voxel coloring algorithm. Some variations are introduced due to the specific sensor and calibration method developed. Our approach focused on obtaining a fast $3D$ reconstruction with limited computation power to target embedded applications. To achieve this, we used some additional techniques to reduce reconstruction time.

1) Scene description: The sensor is placed in an unknown scene, and a bounded reconstruction volume is defined in this scene. A discretization resolution is chosen in accordance with application goals. The scene illumination is approximated by a Lambertian model. The scene is supposed static and the sensor is optionally moved to different places in the scene; a pair of pictures is taken at each position. This leads to two types of reconstruction, a *local* one by processing only a pair of pictures, and a *global* one by taking into account multiples views of the same static scene. All positions of the sensor are known before processing the reconstruction. Both reconstruction types share the same algorithm. The *global* one contains an additional step to merge multiple *local* reconstructions.

```
* FOR EACH voxel IN scene volume
  * FOR EACH image IN {Up, Low}
    * PROJECT voxel ONTO image
    * Calculate Average Color
      of the voxel projection
  * IF ||Avg_Up - Avg_Low|| < Avg_TH
    * Mark voxel as full
```

Fig. 5. Pseudo algorithm of the *local* $3D$ reconstruction. Avg_TH is the threshold on the Average Color to set a voxel as consistent or not.

2) $3D \rightarrow 2D$ projection: In order to process the voxel projections quickly at run-time, the projections of all voxels are computed during the calibration process and results are stored into a LUT. This LUT contains correspondences between $3D$ points in $\mathcal{R}_{\mathcal{W}}$ and $2D$ points in $\mathcal{R}_{\mathcal{I}}$ for both cameras. This LUT is then used at run-time (during the actual reconstruction) for a fast voxel-to-surface projection. When projecting a voxel, the convex hull of its projection is approximated as an axis-aligned rectangular area. This approximation of the projection allows fast reading of the projected surface.

3) Photo-consistency algorithm: The consistency test performed between two different projections of the same voxel on different images is a key part of the reconstruction. Our approach tries to obtain a rough $3D$ reconstruction of the surrounding environment with limited processing performance. Thus we use a photo-consistency algorithm that calculates for each voxel projection the average color of this surface. Color averages of all projections are then compared to determine the consistency of the projected voxel (made by a euclidean distance computation in the RGB color space).

V. Results

A. Calibration

This section presents the results obtained for the calibration stage. The TABLE I specifies the number of markers (N) constituting the calibration pattern, the corresponding average projection error ϵ and the standard deviation σ.

N	$\epsilon \pm \sigma$
HDM	0.26253± 0.15038
40	0.16085 ± 0.087807
30	0.16569 ± 0.09538
20	0.31549 ± 0.18551
15	0.39709 ± 0.1974
10	0.68764 ± 0.32136
5	2.4666 ± 1.313

TABLE I

MEAN PROJECTION ERRORS WITH ASSOCIATED STANDARD DEVIATIONS ($\epsilon \pm \sigma$) FOR DIFFERENT QUANTITIES OF MARKERS (N) ALONG A SINGLE VERTICAL. THE MEAN PROJECTION ERROR IS EXPRESSED IN PIXELS.

The TABLE II summarizes the results obtained for the $3D$ metric recovery. A sample of known $3D$ points $\mathbf{P}_i = [X_i, Y_i, Z_i]$ and corresponding triangulated $3D$ points are detailed. The back-projection percentage errors ($\epsilon_{3D}\%$) are given.

X	Y	Z	$\hat{X}$	$\hat{Y}$	$\hat{Z}$	$\epsilon_{3D}\%$
-2650	-500	550	-2648.1	-500.6	553.9	0.16
500	500	1000	497.1	497.1	998.8	0.40
1500	-700	800	1499.2	-698.9	802.5	0.15
-850	-1150	2200	-848.9	-1146.7	2200.7	0.13
-1500	3000	1800	-1501.9	3004.2	1806.4	0.20
-1500	500	450	-1498.0	499.3	452.2	0.18
-750	1500	550	-748.7	1496.3	553.5	0.30
-450	650	900	-448.2	647.7	901.3	0.26
-750	-800	1000	-748.5	-798.2	1001.1	0.17

TABLE II

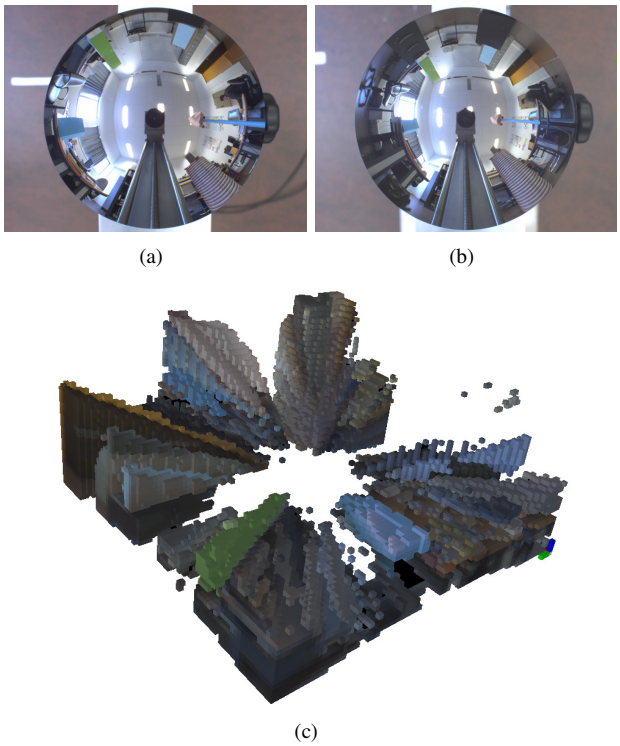
STEREOSCOPIC PANORAMIC SENSOR EVALUATION BY MEASUREMENTS OF A $3D$ METRIC ERROR. ALL DIMENSIONS ARE EXPRESSED IN MILLIMETERS

To corroborate these results, a previous study has been made which consisted in a comparison between a classic parametric calibration technique and the *interpolation-based* one. In terms of projection and $3D$ metric errors, similar results have been founded [16].

B. $3D$ Reconstruction

$3D$ reconstruction results are presented. Pictures of the reconstructed scene are given associated with their original pictures. The reconstruction performance is also measured in terms of reconstruction speed.

1) Reconstruction results: Two reconstructions are presented. Fig. 6 is a *local* reconstruction of a real environment. The resolution is rough with a $5 \times 5 \times 5cm$ voxel size. The roof and walls are manually erased from the final reconstruction in order to be able to see objects inside the room. Fig. 7 show a *global* reconstruction made by merging eight *local* reconstructions of the same scene from eight different sensor positions. This simulation is done in a synthetic environment made with PovRay.



(a) (b) (c)

Fig. 6. Omnidirectionnal pictures of our lab. (a): Up camera - (b): Low camera - (c): Corresponding 3D reconstruction. This is a *local* reconstruction built from only one pair of pictures. The background (roof and walls) is not displayed

2) Performance measurement: By using the proposed method for projecting the voxels onto pictures by means of LUTs, we achieved a fast reconstruction, even for high resolution reconstructions. TABLE III shows reconstruction time in seconds for different scene voxel resolutions (i.e. different quantity of voxels).

Number of vxls	Reconstruction time (s)
$64 \times 64 \times 48 = 196'608$	10
$80 \times 80 \times 40 = 256'000$	20
$100 \times 100 \times 45 = 450'000$	28
$200 \times 200 \times 100 = 4'000'000$	131

TABLE III

3D RECONSTRUCTION PERFORMANCES. MEASURES HAVE BEEN DONE ON A STANDARD (LOW END) DESKTOP PC (AMD 2GHZ, 1GB RAM) RUNNING UNDER WINDOWS XP SP2 32BITS. DURATIONS ARE ROUNDED TO THE NEAREST SECOND AND INCLUDES TIME FOR LOADING FILES AS WELL AS SAVING RECONSTRUCTION RESULTS ON THE HARD DRIVE.

VI. Discussion

A. Calibration

From the results presented in the TABLE I, more the number of markers is important more the projection error is minimal.

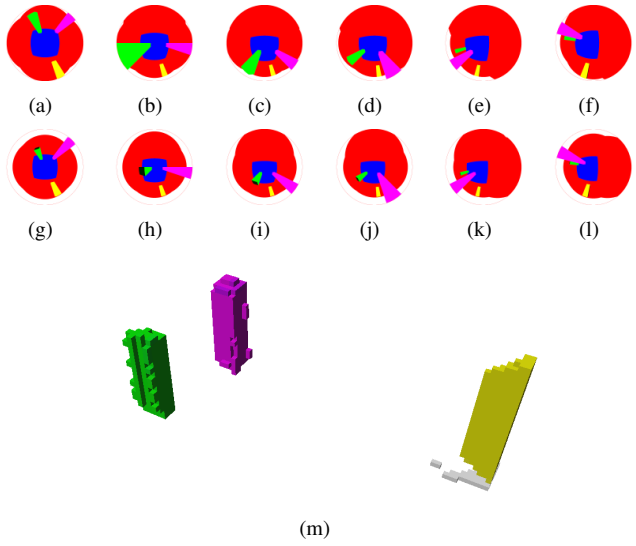


Fig. 7. Synthetic omnidirectionnal pictures of an empty room with 3 colored columns. (a-f): Up camera - (g-l): Low camera - (m): Corresponding 3D reconstruction. This reconstruction was made by merging 8 *local* reconstructions built from 8 different places in the room. Blue roof and red walls are not displayed in the reconstruction

Nevertheless, to build the calibration tool, a compromise has to be made between the projection error and the quantity of markers. A quantity of 15 markers for a single bar introduces an error less than $\frac{1}{2}$ pixel: this seems to be a good compromise. The results given in the TABLE II are made with a 15×8 markers constituting the calibration tool (i.e. quantity of markers for a single bar times the number of bars). The back-projection error percentages are strictly inferior to 1% which means that the calibration performed with the *Interpolation-based* method, according to the new distribution of the markers on the cylinder, is relevant.

B. $3D$ Reconstruction

Our photo-consistency algorithm introduces errors when processing empty voxels placed in front of an object with a uniform color. These errors leads to a fulfilled cone in front of objects (cf. Fig. 6). Discarding these *false* voxels needs more reliable consistency-test. We are working on improving the existing algorithm without introducing additional complexity. A solution to improve the reconstruction consists in merging $3D$ *local* reconstructions obtained from more views at wider spaced positions. This approach has been tested on synthetic scene and cheering results have been obtained (cf. Fig. 7).

VII. Conclusion

In this paper, we have presented a panoramic architecture and algorithms for a stereoscopic panoramic bench providing a relevant solution to the problem of the $3D$ reconstruction of an unknown scene. First of all, we have defined the sensor architecture, made of two catadioptric sensors, and justifications of such a design have been detailed. Secondly, we have elaborated an innovative and generic calibration methodology. This method is available for different kinds of catadioptric sensors and can be used to calibrate catadioptric sensors made of low-cost mirrors. The concept is based on a mesh which establishes $3D/2D$ matchings. These knowledges being established only at the vertices of the mesh, the projection function is a discrete representation of the $3D/2D$ matchings. Thus, $3D/2D$ recovery is done by a local interpolation at the vicinity of the vertices. Finally, we have presented a $3D$ volumetric reconstruction adapted to a panoramic sensor. Each complete voxel is projected onto the two image planes and a photo-consistency test is computed on an approximation of the surfaces. The implementation of the algorithms, i.e. $3D/2D$ correspondences made offline and stored in LUTs, provides fast reconstructions even for high resolution reconstructions.

Acknowledgment

This study took place in an *OmniViss* European InterregIIIA program (Embedded Omnidirectional Vision System for Environment Reconstruction and Analysis). Moreover, this study has received financial support from the council of Haute-Normandie.

References

[1] L. Smadja, E. Bigorgne, R. Benosman, and J. Devars, "Génération de cartes de disparité denses à partir d'images panoramiques cylindriques haute définition," in *Colloque GRETSI*, 2003.

[2] H. Tanahashi, K. Yamamoto, C. Wang, and Y. Niwa, "Development of a stereo omnidirectional imaging system (sos)," in *IEEE IECON*, 2000, pp. 289–294.

[3] S. Baker and S. Nayar, "A theory of single-viewpoint catadioptric image formation," *IJCV*, vol. 35, no. 2, pp. 175–196, 1999.

[4] J. Pons, R. Keriven, and O. Faugeras, "Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score." *IJCV*, vol. 72, no. 2, pp. 179–193, 2007.

[5] C. Dyer, *Foundations of Image Understanding*. Kluwer, 2001, ch. 16 : Volumetric Scene Reconstruction from Multiple Views, pp. 469–489.

[6] S. Ieng and R. Benosman, "Les surfaces caustiques par la géométrie - applications aux capteurs catadioptriques," *Traitement du Signal*, vol. 22, no. 5, pp. 433–442, 2005.

[7] J. Fabrizio, J. Tarel, and R. Benosman, "Calibration of panoramic catadioptric sensors made easier," in *IEEE Omnivis Workshop*, 2002, pp. 45–52.

[8] M. El-Mouaddib, R. Sagawa, T. Echigo, and Y. Yagi, "Stereovision with a single camera and multiple mirrors," in *IEEE ICRA*, 2005, pp. 800–805.

[9] J.-J. Gonzalez-Barbosa, "Vision panoramique pour la robotique mobile : Stéréovision et localisation par indexation d'images," Ph.D. dissertation, Université Toulouse III, Janvier 2004.

[10] M. El-Mouaddib, "Introduction à la vision panoramique," *Traitement du Signal*, vol. 22, no. 5, pp. 409–417, 2005.

[11] T. Svoboda and T. Pajdla, "Epipolar geometry for central catadioptric cameras," *IJCV*, vol. 49, no. 1, pp. 23–37, 2002.

[12] J. P. Barreto and H. Araujo, "Paracatadioptric camera calibration using lines," in *IEEE ICCV*, 2003, pp. 1359–1365.

[13] S. Seitz and C. Dyer, "Photorealistic scene reconstruction by voxel coloring," in *IEEE CVPR*, 1997, pp. 1067–1073.

[14] E. Steinbach, B. Girod, P. Eisert, and A. Betz, "3-d reconstruction of real-world objects using extended voxels," in *ICIP*, 2000, pp. 569–572.

[15] O. Ozun, U. Yilmaz, and V. Atalay, "Comparison of photoconsistency measures used in voxel coloring," in *ISPRS Workshop BenCOS*, 2005, pp. 51–56.

[16] N. Ragot, J. Ertaud, X. Savatier, and B. Mazari, "Calibration of a panoramic stereovision sensor : Analytical vs interpolation-based methods," in *IEEE IECON*, 2006, pp. 4130–4135.