3D SURFACE SEGMENTATION USING ACTIVE SENSING

M. J. Aldon, O. Strauss

LAMM - URA CNRS D0370 - Université Montpellier II - Place Bataillon
34095 Montpellier Cedex 5 - FRANCE
Tel. (33) 67 14 37 76 - Fax (33) 67 52 14 96

Abstract - Manufacturing processes involving object inspection or manipulation often require 3D shape acquisition and modeling. Active sensing offers a practical solution to such problems. In this paper, we present a new solution to extract 3D shapes from structured light images. The 2D striped image is segmented into regions corresponding to quadratic surfaces. For that purpose, a fuzzy analysis of the stripe properties is used to locate discontinuities and to caracterise stripe parts. These parts are tracked in consecutive images and matched in order to create regions. We present segmentation results obtained with real scenes consisting of multiple objects of arbitrary shapes.

1. INTRODUCTION

Sensing of 3D shapes is one of the primary capabilities necessary for autonomous robots able to manipulate objects , or to move in more or less structured environments. 3D measurements can be exploited in a number of different applications such as : inspection, bin-picking, assembly or vehicle motion. Many 3D sensing methods are currently under investigation. R.A. Jarvis presents a general overview of these solutions [1]. It appears that visual sensing can be done passively or actively .

- *Passive range finding techniques* using a single intensity image to deduce shape from shading, from texture, or from extracted edges exhibits ambiguities because of the complexity of the reflectance and illumination model.

- *Stereovision and dynamic vision* use at least two images of the same scene taken by two fixed (or one moving) camera. Presently, these methods furnish a polyhedral approximation of the scene, which is sufficient for tasks which only need a rough approximation of objects.

- *Active sensing techniques* which imply the use of an artificial lighting source can be classified according to the fact that they furnish direct or undirect range measurement :

i) Time of flight methods consist in measuring the light propagation time between the emissive source (ultrasonics, lasers,...) and a coaxial detector. Ultrasonics are a simple and cheap solution ; but it cannot be used for accurate sensing, for several reasons : a poor resolution, the existence of confusing multiple reflections, ... Laser range finders are expensive when they require high performance optoelectronic hardware to build up an accurate range map with a reasonable speed .

ii) Active triangulation methods, so-called structured light methods, consist in projecting on the scene a visible light from a source with a known pattern geometry [2]. The scene is viewed with an imaging sensor looking off the emission axis. Knowing the position of an image point on the detector, as well as the lateral distance between the projector and the camera, and the projection angle of the light source, the 3D point location can be computed.

When the light source is a spot, the acquisition time is essentially depending on the scanning time. This is reduced when the source produces a plane of light. Here, the range information manifests itself in the apparent deformation of the projected stripe.

Scanning devices may be avoided by using multistripe structured light (a set of parallel stripes or a rectangular grid). However, stripes does not appear in the image plane in the same sequence as that projected. The first problem to solve is the grid line identification. Different methods are investigated. Boyer and Kak have tested successively a binary-encoded and a color-encoded structured light with parallel stripes [3],[4]. The system which employs color may be restricted to environments in which the color content of the scene is predominently neutral. Le Moigne and Waxman have choosen a grid of horizontal and vertical lines where additional dots are used as landmarks to initiate the labelling process [5]. They address several interesting issues like operating in ambient lighting, grid pattern selection, albedo normalisation and grid extraction. In [6], G. Hu and G. Stockman solve the "grid line identification" like a correspondence problem by using geometric and topological constraints.

Several researchers have attempted to integrate active and passive techniques in order to increase the redundancy of informations for a more reliable world modeling [7], [8], [9].

Our work deals with the development of an inexpensive and accurate structured light vision system which could be used for fine localization and recognition of manufactured objects [10]. We have choosen a stripe scanning technique for several major reasons:

- the system can operate in ambient lighting and does not suffer limitations of passive vision methods;
- presently, dense range images may be built in a few seconds, if individual light stripe images are analyzed at video rates [11]. Acquisition time may be considerably decreased if only sparse light stripes are needed, for instance when measuring planar faces, or when assuming that partial information is a priori known;

- highly accurate range measurements can be expected if the system is correctly calibrated;
- an active 3D imaging system can produce a dense range map which is necessary for recovering curved and complex shapes;
- the 2D light stripe image brings interesting geometrical information which can be used to solve efficiently the segmentation problem without computation of 3D surface points.

2. OVERVIEW

Different from existing solutions, our method solves the surface identification problem by using a very general constraint that holds in environments consisting of a jumble of manufactured objects. We suppose that surfaces are "regular" enough and that they can be localy approximated by second order patches. We make the assumption that these patches are larger than the stripe spacing given by the scanning system.

Our approach consists of three steps (fig. 1) :

1) Image preprocessing for extracting the stripe skeleton and locating its breaking points and curvature discontinuities which shall correspond to patches boundaries. So, each stripe is segmented into several fragments.

2) Tracking process for grouping 2D stripe parts that present enough common features.

3) Shape recovery via triangulation for computing 3D points and least square minimisation for identifying quadratic patches. 3D construction and refinement the segmented faces are then performed.

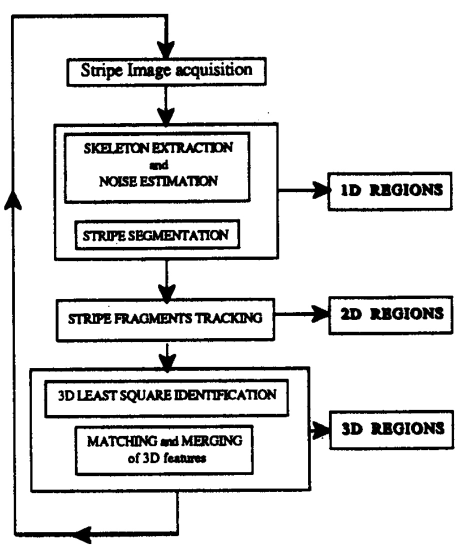

Fig. 1. Processing diagram.

This paper deals only with the algorithms we have developed to solve the two first steps. In the next section, we briefly describe our sensor system and the triangulation principle used to interpret the data in a world coordinate system. Then, we present the data acquisition and the algorithms developed for a preliminary processing of the light stripe image. In section 5, we propose a new approach to the problem of the topographic image segmentation. This 2D image is divided into regions which correspond to surface primitives which do not contain discontinuities in depth or in surface orientation. Section 6 exhibits some experimental results obtained with real scenes including multiple objects of various shapes.

3. RANGE FROM STRIPE SCANNING

3.1 *Sensor system*

Our experimental setup (fig.2) includes a solid camera which provides an image digitization and display module with a standard video signal. Images up to 512*512*8 bits deep are supported in all operations. This camera is equipped with a filter which selects the red light projected on to the scene by a HeNe laser. A cylindrical lens induces a magnification of this light source in one dimension. A galvanometric scanner can address and hold arbitrary angular positions of a rotating mirror which reflects the planar source of light.

3.2 *Sensor model*

The range at illuminated light stripe image points can be calculated by simple triangulation. Therefore, the sensor must be precisely calibrated in order to obtain the parameters of the geometrical transformation between pixel coordinates, and 3D world coordinates.We assume that the camera optical system performs a conical plane projection on a retina whose axis are not necessarily perpendicular. The coordinates (x,y) in mm of a point (u,v) in a pseudo image plane located at a unit distance ($Zc = 1$) from the center Oc may be expressed by :

$$x = l + m\,u + n\,v$$
$$y = p + q\,v$$

where l,m,n,p,q are the intrinsic parameters of the camera. We define a unit vector K normal to the light plane such as:

$$K = [\cos\theta \quad 0 \quad -\sin\theta\,]^T$$

and a fixed frame R_l such as:

. Y_l is attached to the rotation axis,
. Z_l belongs to the laser plane for the position $\theta = 0$,
. O_l belongs to the plane(Xc, Yc).

A point M (Xc,Yc,Zc) of the laser stripe obeys the equation:

$$K \;[L]\; [\,Xc \;\; Yc \;\; Zc \;\; 1]^T = 0$$

where [L] is the identified homogeneous transformation matrix between R_c and R_l.Let Lij be the (i,j) element of the [L] matrix. It can be shown that :

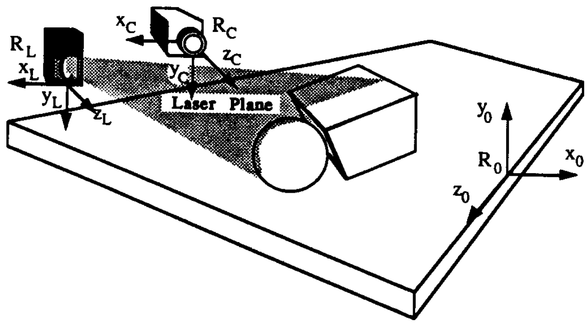

Fig. 2. Light stripe range sensor.

$$Zc = \frac{s\theta\, L34 - c\theta\, L14}{(c\theta\, L11 - s\theta\, L31)x + (c\theta\, L12 - s\theta\, L32)y + (c\theta\, L13 - s\theta\, L33)}$$

with : $s\theta = \sin\theta$ and $c\theta = \cos\theta$,
and that : $Xc = x\ Zc$, $Yc = y\ Zc$.

Then, we apply the identified transformation between R_C and R_O in order to compute the 3D coordinates of the point in the reference frame

4. STRIPE IMAGE PREPROCESSING

By increasing the angle through which the mirror rotates, the light is scanned across the scene. At each scanning step, we obtain a grey-level image showing the scene points illuminated by the light plane. In robotics applications, the variable range between sensor and objects causes defocussed images for the projected stripe. Roughness and reflectivity variations produce large variation in peak intensities of this stripe [2]. Moreover, ambient illumination and surface state of the objects cause lightened pixels on the background.

In order to apply the triangulation method, we must extract the skeleton of the projected pattern which represents the intersection between a perfect plane and the 3D scene. A simple thresholding of the video data cannot be used to isolate the line. We have developed a more accurate solution which uses a statistical analysis of the intensity arrangement in the image. Simultaneously, breaking points which correspond to occluding boundaries between faces are detected and the stripe is divided into several parts. In a second step, a more complete analysis allows us to subdivide stripe parts including angular points and other curvature discontinuities like bending points.

4.1 *Extracting the stripe skeleton*

Sensor's geometry causes the light stripe to be oriented along the Y_C axis of the camera frame (fig. 2). So, a monodimensional processing can be used to select one pixel on each image line.

Firstly, in order to eliminate the background noise, we determine a minimum intensity threshold by applying a statistical method to a set of randomly chosen image points.

On each line of the image, the intensity can be modeled by a discrete function I(x), and the most representative pixel is the one which receives the maximum of energy. In order to identify and to locate accurately this unique maximum, we convolute the I(x) function with a symmetric exponential filter F [12] (see Appendix). Then a statistical computation is used to reduce a possible bias on the maxima location and to estimate the standard deviation of the signal I(x).

4.2 *Locating stripe curvature discontinuities*

Our aim is to approximate object faces by quadratic models. It can be proved that the intersection of a quadric by a plane is a conical curve whose homographic projection is also a conic [13]. Consequently, we propose a segmentation method that allows us to approximate the laser stripe by a set of adjacent second order curves. To do that, it is necessary to locate the stripe discontinuities such as :

- breaking points (zero order discontinuities) which have been detected during the skeleton extraction,
- angular points (first order discontinuities) which result from the junction of two planar faces,
- retrogression points (second order discontinuities) located at the edge between two curved faces,
- bending points which correspond to zero crossings of the curve second derivative.

We present briefly our analysis method which leads to a reliable location of these discontinuities. Let $x = \mathcal{L}(y)$ be the curve resulting from the previous processing (fig. 3a). Noise on this curve is issued from :

- the image sampling,
- the errors involved by the skeleton extraction.

a) *Computing the second derivative of* $L\ (y)$:

Discontinuities correspond to maxima or zero crossing of the second derivative of $\mathcal{L}(y)$. Classical differential operators being sensitive to noise, a preliminary smoothing is necessary. However, the smoothing filter must preserve the accuracy of the discontinuities location . In order to satisfy this constraint, we apply again the symmetric exponential filter described in the appendix. This low-pass filter reduces the truncature noise and provides a good estimate L"(y) of the second derivative (fig.3b). We have choosen the filtering parameters in order to preserve a 6 pixels long motif, i.e. a period of about 4π for $\mathcal{L}(y)$.

b) *Estimating the L" residual noise* :

The estimated second derivative L" includes a residual noise whose mean must be evaluated. This step is necessary to

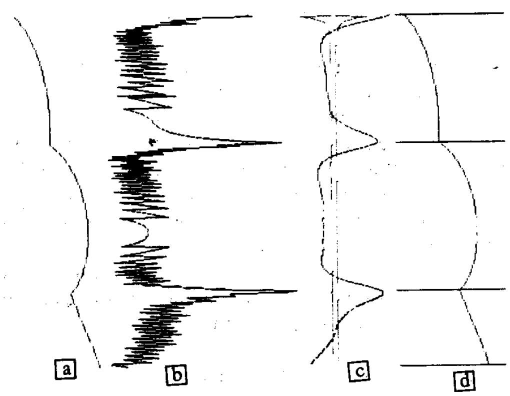

Fig. 3. Laser stripe processing :
3a) Stripe profile $\mathcal{L}(y)$ - **3b)** L"(y) computed by the symmetric filter - **3c)** L"(y) after smoothing, and its confidence interval - **3d)** Discontinuity points on $\mathcal{L}(y)$ -

determine accurately the number of smoothing which must be applied to L" in order to locate exactly the discontinuity points. We thought of applying a Kalman filter to estimate a reference for the zero crossing detection. However, the classical use of this filtering method assumes a better knowledge on the noise than on the signal itself. So, we have defined and applied a qualitative method which controls the filter. This controler uses the Student test results for updating the predicted errors variance. With such a control, the Kalman filter furnishes a good estimate of the noise (fig. 3c).The results obtained show that the estimated confidence interval will be a good criterium for the segmentation. Moreover, the computed mean error is consistent with the noise amplitude.

c) *Analyzing the L" discontinuities* :

The analysis of the smoothed L" signal will consist in locating its zero crossing and its sign change, in order to subdivide $\mathcal{L}(y)$ into second order approximable curves (fig. 3d). We have to establish a decision-making function using fuzzy information such as :
- the smallness of L"(y)
- the straightness of a set of adjacent points of $\mathcal{L}(y)$.

Such information cannot be correctly analyzed by a boolean process. Fuzzy algebra seems suitable to make a good decision in the sense that it preserves the analysis acuteness. So, our curve analysis uses a fuzzy logical operation based on the comparison between the current value of L" and those of the previous points. This method is specially efficient on noisy curve areas. Experiment shows that it then produces a more complete segmentation than the one given by a simple boolean decision.

5. STRIPED IMAGE SEGMENTATION

5.1. *The segmentation problem*

Segmentation of a scene image into regions corresponding to single shapes is one of the hardest problem in 3D vision. Classical methods which are applied to range images can be divided into two groups: region based methods and boundary based methods. *Region based methods* look for similarities to group 3D points or elementary patches into surface regions. For instance, surface normals may be considered as a common local feature. Partitioning them involves thresholding using a histogram analysis [14]. *Boundary based methods* try to find significant changes that separate regions by isolating discontinuities in both depth and surface orientation. Differential geometry is often used to find region boundaries [15]. All these segmentation techniques use dense 3D sensed data and generally need a large computation time.

We are also doing 3D sensing via structured light images. However, we show that an efficient 3D shape segmentation can be done without computing any 3D surface point. Our method is based on the direct analysis of the parameters of the projected stripes in the image frame and of their connectivity relations. We show how it is possible to track stripe parts in the consecutive images and to match them in order to create regions.

5.2 *Stripe parts model*

Grouping similar stripe parts in homogeneous 2D regions involve modeling these curves, and defining matching rules.We identify each part with a simple parabolic model which can be written :

$$x = a\,y^2 + by + c$$

So, a curve fragment is described by the parameters a,b,c, and their variance/covariance. Moreover, we preserve a complete description of its two end points, i.e.:
- their image coordinates,
- a fuzzy value attached to their type (breaking, bending, retrogression or angular point),

We also consider another fuzzy value which caracterizes the straightness of the curve.

5.3 *Matching and fusion process*

The problem we address here is to "track" each curve fragment in the sequence of images acquired when scanning the 3D scene. Presently, we assume that the curve motion D can be modelled by a translation and a very small rotation in the image plane. This small rotation is first order approximated, which allows us to keep the consistency of the parabolic model. For each fragment of a new image, we find which fragment might correspond to it by calculating the distance between each parameter of this fragment, and the corresponding parameters of the set of fragments which have not been matched. We use the so called Battacharyya distance [16] which takes into account the variance/covariance A_i of the parameters. In practice, we have defined an extended distance $\mathcal{B}$ between two parabolic fragments F1 and F2 which are supposed to be locally linear :

$$\mathcal{B} = \frac{1}{4}\left(f(P_1,P_2,D)^t . \mathcal{V}^{-1} f(P_1,P_2,D) \right) + \frac{1}{2} \mathrm{Log} \frac{|\Delta_1 + \Delta_2|}{2\sqrt{|\Delta_1 . \Delta_2|}}$$

with :

$$\Delta_i = \left(\widehat{\frac{\partial f}{\partial P_i}} \Lambda_i \widehat{\frac{\partial f}{\partial P_i}}^t \right) \qquad (i=1,2)$$

$$\mathcal{V} = \Delta_1 + \Delta_2$$

$$P_i = \begin{bmatrix} a_i \\ b_i \\ c_i \end{bmatrix} \qquad D = \begin{bmatrix} d_x \\ d_y \\ \varphi \end{bmatrix}$$

$$f(P_1,P_2,D) = \begin{cases} a_2 - \dfrac{a_1}{1+\alpha\varphi b_1} \\ b_2 - b_1 + 2\alpha d_x . a_1 + \dfrac{b_1 + \alpha\varphi}{1+\alpha\varphi b_1} \\ c_2 - c_1 - (\alpha d_x)^2 . a_1 + \alpha d_x . b_1 - \alpha d_y \end{cases}$$

The translation components αd_x and αd_y, and the rotation angle $\alpha\varphi$ represent the motion from F1 to F2 in the image plane.

Other criteria are considered in this matching process , such as:

- the overlapping of the fragment projections on the Y_c axis which allows us to avoid the matching of similar curves belonging to different regions,
- the likeness between the corresponding fragment end points which is generally stronger in case of curves issued from the same object face.

The decision process is controlled by a fuzzy logical operation on the fuzzy values that we have associated to these distances and criteria.

6. EXPERIMENTAL RESULTS

The algorithms proposed in this study have been implemented on a PC computer and tested on indoor scenes including planar and curved objects. Figure 4a show the striped image of a scene superimposed on the corresponding grey-level image. Figures 4b,...4d present some regions extracted with the segmentation process. We have choosen examples where the stripes exhibit many connected curve parts in order to demonstrate the performances of our algorithms for analyzing complex scenes. Figure 4d shows an example where an efficient tracking can be achieved on a cylindrical face by taking into account a rotation component for describing the fragment motion.

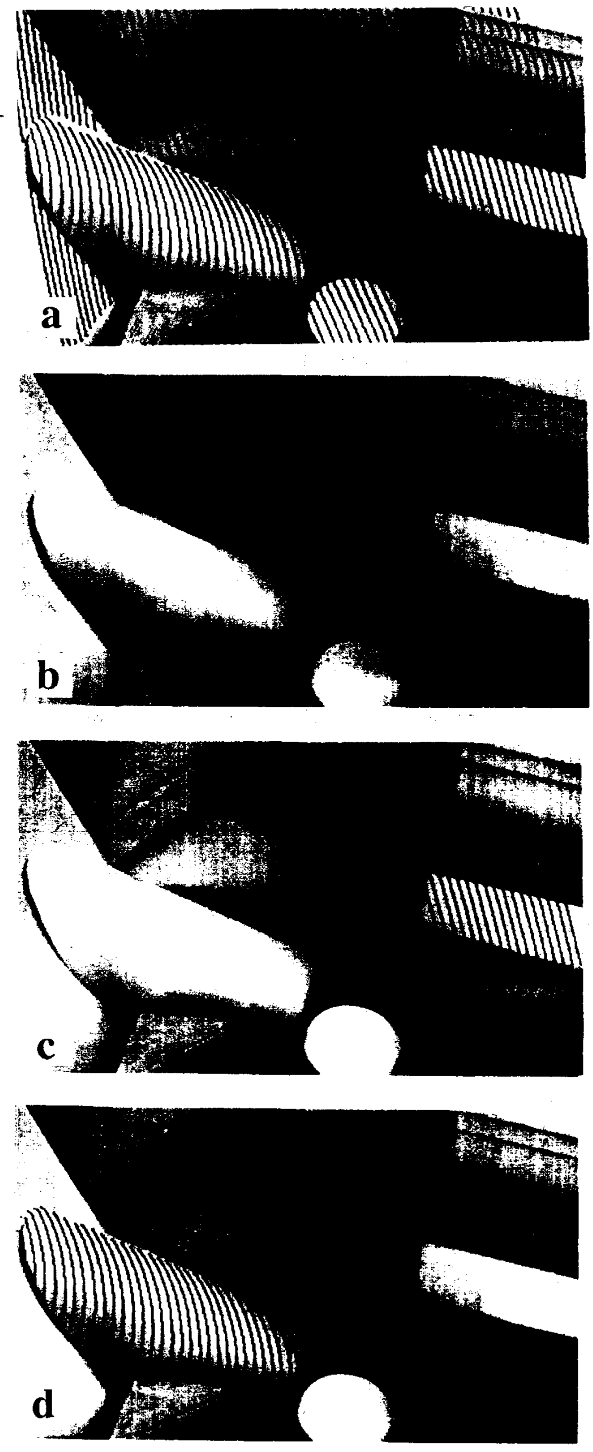

Fig. 4. Segmentation results

One of the major limitation of our method is the computation time needed by the stripe preprocessing (40 sec). However, we think that this time can be significantly decreased by using a dedicated hardware.

7. CONCLUSION

In this paper we have presented algorithms for solving the segmentation problem when sensing 3D scenes with structured light vision. The proposed method has been successfully tested in few constrained environments, with fixed sensors.

Presently, we are improving our shape recovery method which uses an analytic representation. A recursive least square identification is applied to sparse 3D measurements issued from the matched stripes, in order to obtain a quadratic model for each created patch [10]. We are also developping algorithms for modeling the face boundaries.

Future extensions of this work are directed at :

- the implementation of hardware solutions for reducing the stripe preprocessing computation time (skeleton extraction),
- the study of cases where sensors are able to move in the scene, in order to avoid shadow effects and to obtain a more complete modeling. Then, the estimation of sensors motion, and the interpretation of measurements in the same reference frame involve using multi-sensory fusion methods for reducing incertainty, and ensuring a coherent global model of the world.

APPENDIX : The symmetric exponential filter

This low-pass filter has been proposed by S. Castan and al. [12] for an optimal edge detection in 2D images. It allows to reduce the truncature noise and to obtain an accurate estimation of the second derivative of a function.

Such a filter can be written :

$$
\begin{aligned}
f_L(x) &= C\, a_0 (1-a_0)^{|x|} \\
&= f_1(x) * f_2(x) \\
&= C\,(f_1(x)+f_2(x)-a_0\,\delta(x))
\end{aligned}
$$

where :

$*$ means the convolution

$$C=\frac{1}{2-a_0}$$

$$a_0=e^{-\alpha}$$

$$f_1(x)=\begin{cases} a_0\,(1-a_0)^{x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$$f_2(x)=\begin{cases} 0, & x > 0 \\ a_0\,(1-a_0)^{-x}, & x \leq 0 \end{cases}$$

The first and second derivatives of $f_L(x)$ are computed by using the correspondent derivatives of f_1 and f_2.

For instance, an estimated second derivative of $\mathcal{L}(x)$ is given by :

$$L''(x) = C''\,(L_1(x)+L_2(x) - 2\,L(x))$$

where : $L_1 = \mathcal{L} * f_1$ and $L_2 = \mathcal{L} * f_2$

A recursive method is used to compute the symmetric functions L_1 and L_2.

REFERENCES

[1] R. A. Jarvis , " A perspective on range finding techniques for computer vision ", IEEE Trans. on Pattern analysis and Machine Intelligence, Vol.5, n° 2, p. 122-139, Mars 1983.

[2] J. A. Jalkio, R.C. Kim, S.K. Case, "Three dimensional inspection using multistripe structured light," Optical engineering, Vol. 24, n° 6, p. 966-974, Nov. Dec. 1985

[3] H. S. Yang, K. L. Boyer, A. C. Kak, "Range data extraction and interpretation by structured light", 1st IEEE Conf. on Artificial Intelligence Applications,p. 199-205, Dec. 1984.

[4] K.L. Boyer, A.C. Kak, " Color encoded structured light for rapid active ranging", IEEE Trans. on Pattern analysis and Machine Intelligence, Vol.9, n° 1, p. 14-28, January 1987.

[5] J.J. Le Moigne, A.M. Waxman, " Structured light patterns for robot mobility," IEEE Journal of Robotics and Automation, Vol. 4, n° 5, October 1988, P. 541-548.

[6] G. Hu, G. Stockman,"3D surface solution using structured light and constraint propagation," IEEE Trans. on Pattern analysis and Machine Intelligence, Vol.11, n° 4, p. 390-402, April 1989.

[7] G. Hu, G. Stockman, "3D scene analysis via fusion of light striped image and intensity image", Proc. of AAAI Workshop on Spatial reasoning and multi-sensor fusion, St Charles, IL, p. 138-147, October 1987.

[8] M. J. Aldon, H. Huot-Soudain, O. Strauss, "Structured light for 3D robot vision", Int. Workshop SIFIR'89, p. 149-154, Zaragosa, November 1989.

[9] Y. F. Wang, J. K. Aggarwal, "Integration of active and passive sensing techniques for representing 3D objects", IEEE Trans. on Robotics and Automation, Vol. 5, n° 4, P. 460-471, August 1989.

[10] M. J. Aldon, O. Strauss "Structured light and 2D vision for 3D scene analysis", Int. Conference IMAGE'COM, Bordeaux, November 1990.

[11] G.T. Reid, R.C. Rixon, H. Stewart , "Stripe scanning for engineering". Sensor Review, 8 (2), p. 67-71, April 1988.

[12] S. Castan, J. Zhao, J. Shen, "Optimal filter for edge detection methods and results" . E.C.C.V. 90, Antibes, P. 13-17, April 1990.

[13] N. Efimov, "Formes quadratiques et matrices", Ed. Mir-Moscou, 1976.

[14] D. Laurendeau, D. Poussart," Model building based on the segmentation of 3D range images using the hemispheric histogram", Int. Conference AFCET-IASTED on Identification and Pattern Recognition, p. 729-740, Toulouse, Juin 1986.

[15] N. Yokoda, M.D. Levine, "Range image segmentation based on differential geometry".IEEE Trans. on Pattern analysis and Machine Intelligence, Vol.11, n° 6, p. 643-649, June 1989.

[16] M. Basseville, "Distance measures for signal processing and pattern recognition", Research Report n° 422, INRIA, September 1988.