GLOBAL JOURNAL OF PURE AND APPLIED SCIENCES VOL. 10, NO. 2, MARCH 2004: 335 - 341
 ISSN 1118- 0579

CORRESPONDENCE OF LINE SEGMENTS BETWEEN TWO IMAGES: COMPARISION BETWEEN EPIPOLAR, BAYESIAN AND NEURONAL APPROACHES

C. N'ZI, H. LOAIZA, J. TRIBOULET, S.LELANDAIS and C. MONTAGNE

(Received 24 July 2003; Revision accepted 6 February 2003)

ABSTRACT

In order to permit the localization and the navigation of a mobile robot within an interior environment, we have built a stereoscopic sensor and implemented all the algorithms which allow to obtain 3D coordinates of real objects from data images. Sensor uses two mini cameras with vertical disposition. Processing on the images leads us to have segments on the two images. To match segments, we have to answer at the question: "Is there two segments, in a pair from top and bottom images segments, are two views of the same part of the real world of the robot ?". For choosing the best matching algorithm, we compare three approaches: epipolar , Bayesian and neuronal techniques. In the epipolar method, we use only geometrical features. Two images of a single scene/object are related by the epipolar geometry, which can be describe by a 3x3 singular matrix called the fundamental matrix. It captures all geometric information contained in two images, and its determination is very important in many applications such as scene modeling and navigation of a mobile robot. With the two others, we associate to each segment, a set of 16 parameters which includes geometrical, gray level, textural and neighborhood features. For the comparison of these approaches, we have built a database of over 2500 segments. Results about this work are presented in our paper. Results with the two techniques from the data classification, are very good. In order to increase the quality of the segments matching, it is possible to find a combination rule for these two approaches. But, even if for the epipolar method 22% segments are not recognized, we think it is interesting to introduce it with the aim to obtain better results in some difficult configurations, by example when gray levels and textures are not very different in the scene, and also in order to initialize the classification algorithm.

KEY WORDS : Stereoscopic vision – Epipolar matching – Bayes classifier – Neural classifier – Mobile robotic.

INTRODUCTION

In this paper we describe the comparison of three algorithms for matching stereoscopic segments. In a first part we describe briefly the stereoscopic head characteristics and the processing performs on images. In second part we present the characteristics that we compute on each segments. Then we describe the three methods for matching segments that we have compared. First is the epipolar algorithm that we have implemented, then we describe Bayesian and neuronal techniques. For these two last methods we compute some features on each segment of an image. Then a pair of two segments, one from the top image and the other from the bottom image, is characterized by 16 parameters obtained from these features. In a third part we present the results obtained by these three approaches on 2500 data. Finally, conclusions and ideas for future work are explained.

HOW WE OBTAIN SEGMENTS FROM STEREOSCOPIC SENSOR?

The segments are provided by a stereoscopic sensor. The stereo head is composed by two mini cameras. The two cameras rig has a vertical disposition. The stereo head is mounted on top of a stepper motor embarked on a mobile robot (on figure 1 the prototype of sensor is mounted on the mobile robot Real MagiCol (LOAIZA, SUAREZ, GONZALEZ, LELANDAIS and MORENO, 1998, P.475)). The field of view of the bottom camera is oriented toward the floor, the top camera is directed toward the horizontal. So, the two vision-fields are partially overlapped.

C. N'ZI, INP_HB, Département Génie Electrique BP 1093 Yamoussoukro, Côte d'Ivoire.
H. LOAIZA, School of Electrical and Electronic Engineering, University of Valle, Santiago de Cali, Colombia
J. TRIBOULET, Université d'Evry Val d'Essonne-CEMIF Systémes Complexes, 40 Rue du Pelvoux-91020 Evry Cedex, France
S. LELANDAIS, Université d'Evry Val d'Essonne-CEMIF Systémes Complexes, 40 Rue du Pelvoux-91020 Evry Cedex, France
C. MONTAGNE, Université d'Evry Val d'Essonne-CEMIF Systémes Complexes, 40 Rue du Pelvoux-91020 Evry Cedex, France

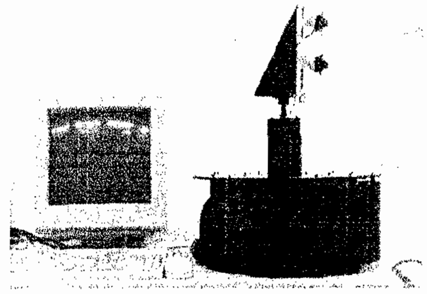


Figure 1: Mobile robot and stereoscopic system.

Calibration is the first step to characterize the cameras and to estimate intrinsic and extrinsic parameters of each one. Two types of models are considered. The first model is a pinhole camera model that neglects all optical distortion (AYACHE and FAUGERAS, 1989). The second model takes into account the radial distortion (LENZ and TSAI, 1988).

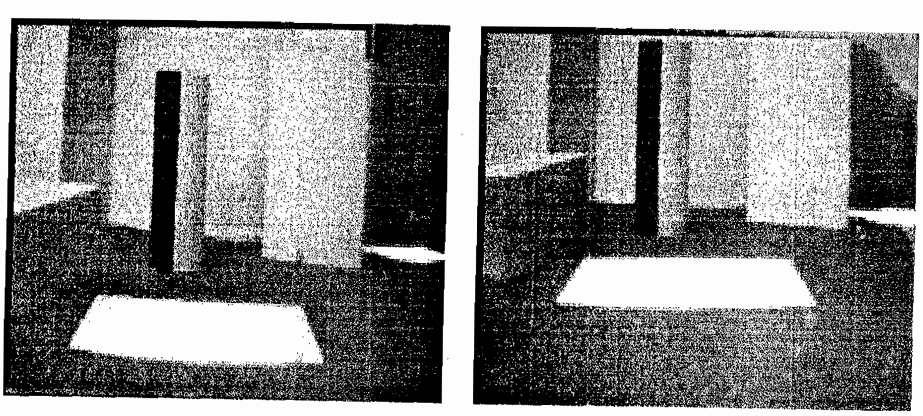
In order to specify the cameras and to define the image quality provided by the sensor, a comparison between one camera of the stereoscopic head and a standard camera was achieved (LOAIZA, 1999). The images of the same scene taken in identical conditions by the two kind of cameras were compared. One could remark a barrel distortion and weak definition of the gray levels in the image obtained from our camera. The image processing which has been implemented has to correct cameras distortions and to extract segments from the images. We have retained straight-line segments as primitives, because they are easily detected. It is also possible to compute many descriptors associated to this kind of primitive and to obtain a function which is very sensitive to the descriptors variations (FAUGERAS, 1996). The different steps of image processing are described in the following lines.

Image enhancement (GONZALEZ and WOODS, 1993) : In this part we correct the radial distortion by using a bidimensional look up table and we give a high contrast to the image by using an histogram specification method. *Optimal filtering* (DERICHE, 1987, P.167) : The method of the gradient to obtain edge points was used. The gradient components are computed by separable and optimal filtered as Deriche defined it. *Local maximum :* The aim of this step is to obtain thin edges. So we must extract edge points having a gradient magnitude value greater than their neighbors. *Threshold computation* (LOAIZA, 1999) : We implement an original algorithm which allows us to automatically obtain two threshold values. *Hysteresis segmentation* (CANNY, 1986, P.679): In this part, we compute both the follow-up of edge points and the building of the segments.

The result of image processing and segment extraction is presented on two images from the stereoscopic head on figures 2 and 3.

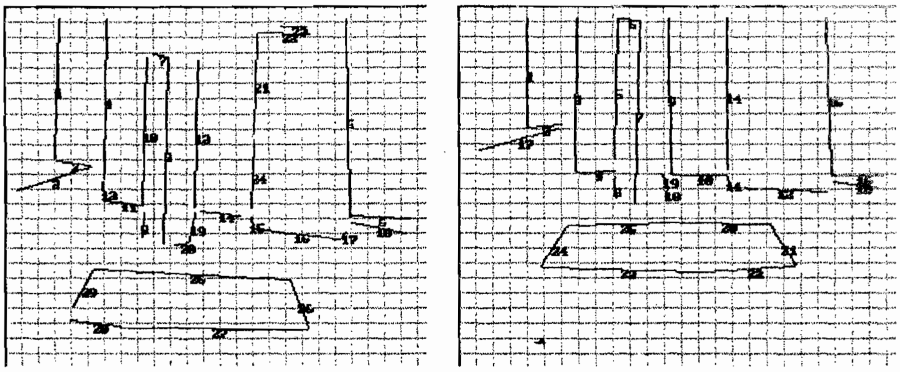
SEGMENTS CHARACTERISTICS

After we have obtained the segments, we calculate descriptors' values on each one. Four kinds of descriptors have been chosen : geometrical, gray level, textural and neighborhood features. During image processing and segments building, we compute geometrical parameters and mean gradient value. For the other gray level features and for the textural features, we define a pixels area on the right and left of the segments as figure 4 shows it.



(a) (b)

Figure 2 : Images from camera, a)upper b) lower.



(a) (b)

Figure 3 : Images after segmentation, a) upper, b) lower

We compute on these areas the mean of gray levels, the internal contrast and the gray levels differences on four directions (0°, 45°, 90° and 135°) which represent a local variation. These textural parameters come from simplified cooccurrence matrix (HARALICK, SHANMUGAM and DINSTEIN, 1973, P.610) in order to obtain a result quickly. So each segment is described by 16 parameters(see table1). In a second time, it appears that 16 variables is a too big number in order to realize a real time application. It is also possible that variables are correlated. So we decide to reduce the size of variables space. In this way, three approaches are tested: Principal components analysis, Factorial discriminant analysis and a combinatory method which uses the trace criterion (LOAIZA, 1999). With all these three methods, we find that the best number of variables is 8. It is interesting to note that these variables come from the three kinds of features: geometrical, gray level and textural features.

The goal of this work is to compare a couple of segments, one from the top and the other from the bottom images, and to decide if they are the representations of the same part of the world scene. In order to do that, we create two classes. Class number 1, or 2, members are pairs of segments which are good, or bad, matched respectively. A 8 components vector is associated to each member of one class. These 8 variables (x^i) are computed as the differences between the parameters of each segments of the pair to classify.

STEREOSCOPIC MATCHING

Introduction

Now we have segments, we have to match them in pair, one from the top camera and the other from the bottom camera. After the matching, we perform the 3D reconstruction. In this step, precision is important, but also we do not want wrong matching. With a false matching, the robot can make a mistake about its position. So, we have to find the best matching algorithm. We compare a classical method, as epipolar geometry, with two others like Bayesian and neuronal classification.

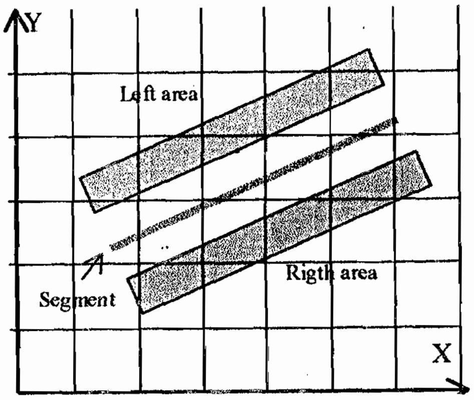


Figure 4 . Definition of the areas on the left and on the right of a segment.

Table1: Definition of the 16 components of each vector

Variables	Computed as the difference between:
x1	vertical coordinates of segment centers
X2	horizontal coordinate of segment centers
X3	gradient intensities along the segment
X4	segment orientations(in degrees)
X5	inside contrast in the left area
X6	inside contrast in the right area
X7	means of gray levels in the right area
X8	means of gray levels in the left area
X9	gray level variation at 0° in the rigth area
X10	gray level variation at 45° in the rigth area
X11	gray level variation at 90° in the rigth area
X12	gray level variation at 135° in the rigth area
X13	gray level variation at 0° in the left area
X14	gray level variation at 45° in the left area
X15	gray level variation at 90° in the left area
X16	gray level variation at 135° in the left area

Table 2 : 2D points used to estimate a fundamental matrix

Image UP		Image Down	
V	*U*	*V'*	*U'*
155	103	154	104
116	114	114	119
183	39	184	47
246	19	247	23
206	202	209	211
126	200	127	210
222	77	224	83
165	179	167	191
227	125	229	130
202	170	203	181
135	107	133	107
236	195	241	201

Table 3 : Classification results with 8 variables

Method	Bayesian %	Neural %
Class 1	93,75	96,43
Class 2	92,73	92,73
Mean	93,24	94,58

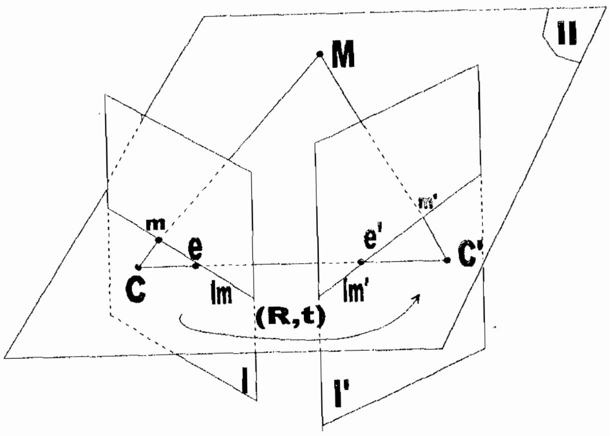


Figure 5: Epipolar geometry

Epipolar method

Epipolar geometry describes the relation between two images of the same scene using a 3x3 singular matrix. This matrix is known as the essential matrix if intrinsic parameters of the images are known. If those parameters are not known the matrix is called fundamental matrix.
This matrix contains all of the geometric information about the two images. So its estimation is really important for scene modelisation in case of an autonomous mobile robot system.
In image processing solving stereoscopic problem, this approach is well known. Ayache (AYACHE, 1989) calculates analytically the equations of the epipolar lines using global parameters of both cameras. Zhang (ZHANG, 1996) works on this problem taking the fundamental matrix into account.
Let m be a point from image ***I*** and m' the corresponding point in the second image ***I'*** is on the epipolar line (l'm) as we see on figure 5. This line is the intersection between the image plane ***I'*** and the epipolar plane passing through points m, **C** and **C'**. This plane contains also m'. All of the points on line **CM** have the same epipolar line. The set of epipolar lines has a common point e' called the epipole point. It is also the intersection between the line defined by **CC'** and the image plane ***I'***.

By the same way it is possible to determine (lm') the epipolar line from m' and epipole point e in the image plane ***I***. (lm') and (l'm) are called conjugated epipolar lines.
A point m from the upper image and its corresponding point m' in the lower image are known using the fundamental matrix $\mathbf{F}_{3x3}$:

$$\tilde{m}'^T F\tilde{m} = 0 \qquad \text{where} \quad \tilde{m} = \begin{pmatrix} m^T \\ 1 \end{pmatrix} \qquad (1)$$

Fundamental matrix estimation

Several methods were developed as presented in (LUONG, DERICHE, FAUGERAS and PAPADOPOULOS, 1993). The method presented here was implemented by Zhang (ZHANG, 1996).
Equation (1) can be rewritten :

$$\mathbf{u}_i^T\mathbf{f} = 0 \qquad (2)$$

where :u_i = [$u_i'u_i$, $u_i'v_i$, u_i', $v_i'u_i$, $v_i'v_i$, v_i', u_i, v_i, 1] andf = [F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33}]

Then, for n couples of points [(u,v), (u',v')], equation (2) leads to:

$$\mathbf{U}_n\mathbf{f} = 0 \quad \text{where } U_n = [u_1, \ldots, u_n] \qquad (3)$$

Least square method is about finding solutions of equation (3) minimizing:

$$\mathbf{min}_f\|\mathbf{U}_n\mathbf{f}\|^2 \qquad (4).$$

Numerical results can give bad results in case of zero divide. To avoid that problem, the solutions can be obtained using Lagrange method. It is an optimization method with constraint which can be written :

$$\boldsymbol{f}(\mathrm{f},\lambda) = \|\mathrm{U_n f}\|^2 + \lambda(1 - \|\mathrm{f}\|^2) \qquad \text{where } \|\mathrm{f}\|=1 \qquad (5)$$

3. Bayesian approach

A p dimensional features vector is denoted by $x = [x_1,..,x_p]^T \in R^p$ and its associated class by $k \in \{1,..,K\}$. A classifier can be regarded as a function $c: R^p \to \{1,..,K\}$ that classifies a given features vector x to the class $c(k)$ (RIPPLEY, 1996). The optimal Bayes rule $c(x)$ for 2 classes which minimizes the expected error rate is:

$$c(x) = \begin{cases} 1 \text{ if } h(x) > 1 \\ 2 \text{ if } h(x) < 1 \end{cases} \quad \text{with:} \quad h(x) = \frac{C_{1/2}\, p_1 f_1(x)}{C_{2/1}\, p_2 f_2(x)} \qquad (6)$$

p_k denotes the prior probability of class k, $f_k(x)$ is the density probability of class k, C_{k_1/k_2} represents the classification cost of a vector of class k_1 into class k_2 and $h(x) = 1$ is the discriminating surface.

In the following we assume that pattern vectors from class *k* are normally distributed with mean vector μ_k and covariance matrix Σ_k. The detection method quality is defined by a rates: *Dk* indicates the success percentage for class k.

4. Neural approach

The neural network used is a multilayer perception with the gradient back-propagation algorithm. It is composed of one hidden layer with a non linear sigmoid, 16 inputs and 2 outputs with linear activation function The best architecture is chosen after comparing the generalization error for different initializations and hidden layer sizes. An algorithm of cross-validation allows the maximizing of the generalization capabilities (DROR, ZAGAESKI and MOSS, 1995, P.149). Data set are divided in three equals parts a learning data set, a validation data set and a generalization data set (BISHOP, 1995). A softmax function (BRIDLE, 1995, P.227) evaluates the a posteriori probability of neural classification.

RESULTS OF MATCHING

1. Epipolar results

To calculate the fundamental matrix, we use the stereo images presented on figure 2. The couples of coordinates [(u,v), (u',v')] are given by the same 3D point. To help the algorithm, a rough match is done before. It is performed taking the segment slope into account. So two segments can be matched only if the slope difference is less than a threshold (20° is used here). Then, epipolar geometry is used to improve the matching. On figure 3, a manual matching is made as a reference (18 couples). The epipolar algorithm presented here gives 78% of correct matching, 4% of incorrect matching and 22% segments are not recognized. Other tests must be done using experimental images with more than 500 segments.

An example of a fundamental matrix is given(7). This matrix is obtained with 12 couples of 2D coordinates points(table 2).

$$F = \begin{bmatrix} 0.000020 & 0.004014 & -0.737703 \\ -0.004081 & 0.000128 & 0.924646 \\ 0.746738 & -1.000000 & 8.783325 \end{bmatrix} \qquad (7)$$

2. Data classification results

For testing the quality of these variables we have built four data bases, each one of 660 members. The bases 1 and 2 use segments from very textured images. The bases 3 and 4 use segments from poorly textured images. Bases 1 and 3 elements are from the class 1 (good matching). Bases 2 and 4 elements are from the class 2 (bad matching). The two classification methods have been used to classify all the pairs in the data bases. Table 3 shows the classification performances of Bayesian and MLP methods.

CONCLUSION

We have developed a vertical sensor for stereoscopic vision. This stereoscopic sensor is used in autonomous mobile robot domestic applications. With the segments which are extracted from the two images, it is possible to obtain 3D information. In order to have the best 3D information, we want to match the segments from the top and bottom images with the better algorithm as possible. We test three methods: epipolar, Bayesian and neural methods. Results with the two techniques from the data classification, are very good. In order to increase the quality of the segments matching, it is possible to find a combination rule for these two approaches. But, even if for the epipolar method 22% segments are not recognized, we think it is interesting to introduce it with the aim to obtain better results in some difficult configurations, by example when gray levels and textures are not very different in the scene, and also in order to initialize the classification algorithm.
In other way we work now with color cameras. We have to test our methods on color images. May be the combination of these two kinds of approaches: geometrical and data classification, will be useful on more complex images.

REFERENCES

Ayache, N. and Faugeras O. D., 1989. Maintaining Representation of the Environment of a Mobile Robot. Proceedings IEEE of the International Symposium and Exposition on Robots.

Ayache, N., paris 1989. *Vision Stéréoscopique et Perception Multisensorielle : Application à la Robotique*. InterEditions.

Bishop, C., 1995. Neural networks for pattern recognition. Clarendon Press.Oxford, 1995.

Bridle, J. S., 1990. Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. Neuro-computing: Algorithms, Architectures and Applications, Berlin: Springer, Pp. 227-236.

Canny, J. F., 1986. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8: 679-698.

Deriche, R., 1987. Using Canny's criteria to derive an optimal edge detector recursively implemented. The International Journal of Computer Vision, Pp. 167-187.

Dror, I.E., Zagaeski, M. and Moss, C. F., 1995. Three-dimensional target recognition via sonar: a neural network model. Neural Networks, 8(1): 149-160.

Faugeras, O., 1996. Three-Dimensional Computer Vision. MIT Press, Massachusetts.

Gonzalez, R. and Woods, R., 1993. Digital image processing. Addison-Wesley Publishing Co. Massachusetts.

Haralick, R. M., Shanmugam, K. and Dinstein, I., 1973. Textural features for image classification. IEEE Trans. on Systems, Man, and Cybernetics. vol. smc-3(6): 610-621.

Lenz, R. K. and Tsai r. Y., 1988. Techniques for calibration of the scale factor and image center for high accuracy 3-d machine vision metrology. IEEE Transactions on PAMI, vol. 10, No. 5.

Loaiza, H., Suarez, A., Gonzalez, E., Lelandais, S. and Moreno, C., Paris, 1998. Real MagiCol: Complex Behavior through simpler Behavior Oriented Commands. Proceedings of the second Robocop Workshop. Robocup-98: 475-482.

Loaiza H., 1999. *Capteur de vision stéréoscopique multiconfiguration pour la localisation d'un robot mobile.* Thesis on Robotic, University of Evry, France.

Luong, Q. T., Deriche, R., Faugeras, O., Papadopoulos, T., 1993. On Determining the Fundamental matrix : Analysis of Different Methods and Experimental Results, INRIA Sophia Antipolis, Research report.

Rippley, B. D., 1996. Pattern Recognition and Neural Networks. Cambridge University Press.

Zhang, Z., 1996. Determining the Epipolar Geometry and its Unvertaintyj : A Review, INRIA, Research report.