



## City Research Online

### City, University of London Institutional Repository

---

**Citation:** Brillault-O'Mahony, B. (1992). A probabilistic approach to 3D interpretation of monocular images. (Unpublished Doctoral thesis, City, University of London)

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

---

**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/29038/>

**Link to published version:**

**Copyright:** City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

**Reuse:** Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

**A PROBABILISTIC APPROACH TO  
3D INTERPRETATION OF  
MONOCULAR IMAGES**

Béatrice Brillault-O'Mahony

---

Thesis submitted for the obtention of the  
Degree of Doctor of Philosophy  
in Information Engineering

---

CITY UNIVERSITY

Department of Electrical, Electronics  
and Information Engineering

---

March, 1992

---

## TABLE OF CONTENTS

1 INTRODUCTION	21
2 SURVEY	29
2.1 AIMS AND DIFFICULTIES OF A VISION SYSTEM	30
2.2 GENERAL KNOWLEDGE	36
2.3 SCENE REPRESENTATION	41
2.4 EXTRACTION OF FEATURES	46
2.5 METHOD AND DATA REPRESENTATION	48
2.6 CONCLUSION	54
3 OVERVIEW OF THE METHOD	57
3.1 PROBLEM DESCRIPTION	57
3.2 PRINCIPLES OF THE METHOD	59
3.2.1 From low to high level features	61
3.2.2 Dealing with uncertainty	66
3.2.3 Dealing with segmentation noise : Likelihood ratio test and scoring	72
3.2.4 Application to Map construction	80
3.2.5 Conclusion	81
4 PREPROCESSING	83
4.1 OVERVIEW OF THE FEATURE EXTRACTION	83
4.2 EDGE DETECTION	84
4.3 LINE AND ELLIPSE FINDER	107
4.4 STATISTICAL MODEL	112
4.4.1 Modelling of the measurement error	112
4.4.2 Modelling of the segmentation error	115
4.5 CALIBRATION PARAMETERS	118
4.6 CONCLUSION	120
5 DETECTION OF PRINCIPAL DIRECTIONS	123
5.1 INTRODUCTION	123

Table of contents

5.2 VANISHING POINT DETECTION	124
5.2.1 Previous work	124
5.2.2 Accumulator space	126
5.2.3 Comparison with other accumulator space	141
5.2.4 Vanishing point detection	144
5.3 CLASSIFICATION OF THE LINES	152
5.4 MAIN PERPENDICULAR DIRECTIONS	158
5.5 ELLIPTICAL ARC CLASSIFICATION	165
5.6 RESULTS	166
5.7 CONCLUSION	171
6 HIGH-LEVEL 3-DIMENSIONAL CONFIGURATIONS	173
6.1 OVERVIEW OF THE METHOD	173
6.2 CONNECTIVITY	175
6.2.1 Merging lines parallel in the 3D world	176
6.2.2 Grouping lines perpendicular in the 3D world	182
6.3 3D STRUCTURES	185
6.4 RESULTS	197
6.5 DISCUSSION AND CONCLUSION	205
7 TOWARDS A CAD DATABASE	207
7.1 INTRODUCTION	207
7.2 3D MAP CONSTRUCTION	208
7.3 TOWARDS A CAD DATABASE	213
7.4 RESULTS	215
7.5 DISCUSSION AND CONCLUSION	218
8 DISCUSSION AND CONCLUSION	221
8.1 SUMMARY OF THE METHOD	221
8.2 DISCUSSION	222
8.3 FROM MONOCULAR VISION TO A CAD REPRESENTATION	226
REFERENCES	227
APPENDIX 1 <i>Perspective transformation</i>	
APPENDIX 2 <i>The Kalman filter</i>	

APPENDIX 3 *Likelihood ratio*

APPENDIX 4 *Statistics of the parameters of the straight line segments*

APPENDIX 5 *Uncertainty parameters associated with a line (L) in the image*

APPENDIX 6 *Uncertainty parameters associated with the calibration parameters*

APPENDIX 7 *Zero-crossing problem. Application to edge detection. Application to the expected number of false alarms in the accumulator space.*

Table of contents

LIST OF FIGURES

2.1.1	: Object location or viewpoint determination	30
2.1.2	: Application field of various existing systems.	32
2.1.3	: Prior knowledge and image processing	33
2.2.1	: Inferences proposed by Lowe (Lowe, 1985)	38
2.3.1	: Model type for various systems	45
2.4.1	: Vertex pair	47
3.2.1.1	: Perspective view of a cube and associated vanishing points	62
3.2.1.2	: Examples of 3D structures	64
3.2.1.3	: Top view of a cupboard lying against a wall	66
3.2.2.1	: Likelihood test, Kalman filter and data representation in the construction of 3D structures	71
4.2.1	: A real step and a simulated noisy step	89
4.2.2	: A crenellated edge and stair steps	89
4.2.3	: Theoretical comparison of the Canny, Deriche and IEF filters	101
4.2.4	: First test image	103
4.2.5	: Second test image	103
4.2.6	: Comparison of edge detectors on the first test image	104
4.2.7	: Image displayed in figure 4.2.6, thresholded at $\Sigma-1$ .	104
4.2.8	: Comparison of edge detectors on the second test image	105
4.2.9	: Comparison of the Shen detector ( $\alpha=0.3$ ) and the IEF ( $\alpha'=0.2$ )	105
4.2.10	: Initial image of an indoor scene	106
4.2.11	: Gaussian filter $\sigma_f = 2$ .	106
4.2.12	: Deriche's detector $\alpha = 0.85$	106
4.2.13	: IEF, $\alpha' = 0.3$	106
4.3.1	: Principle of Berthod's line finder	109
4.3.2	: Effect of a small perturbation on Berthod's line finder	110
4.3.3	: Berthod's line finder results on a test image	111
4.3.4	: Berthod's line finder	111
4.3.5	: Lowe's line finder	111

## Figures

4.4.1.1	: Model of the endpoint uncertainty	113
4.4.1.2	: Geometric interpretation of $V(\vec{w})$	115
4.4.2.1	: Density function of the lengths of the segments ( $\ell > 15$ )	117
4.4.2.2	: Average value of $1/\ell^2$ in function of d.	117
4.5.1	: Image of the test cube	119
5.2.2.1	: Transformation from the image plane to the accumulator space	127
5.2.2.2	: Line (L) passing near the point P	129
5.2.2.3	: Average value of $1/\ell^2$ in function of d performed over the lines satisfying the constraint 5.2.2.6 with $D_m/D_r = 1.3$	130
5.2.2.4	: $(x'(r), y'(r, \theta))$ accumulation space.	134
5.2.2.5	: Geometric interpretation of the transformation T	136
5.2.2.6	: Comparison between a "whole line" accumulation approach and the "intersection point" accumulation approach	138
5.2.2.7	: Accumulator space corresponding to the sampling 5.2.2.15	141
5.2.3.1	: Projection of an image segment onto the Gaussian sphere	143
5.2.4.1	: Accumulator space corresponding to the accumulation of 2000 lines located at random in the image.	141
5.2.4.2	: Expected value of the accumulator space in function of $x'$ where 2000 lines located at random in the image have been accumulated.	147
5.2.4.3	: Number of false alarms $n_f$ when successively 50, 100, 300 random directions have been accumulated, corresponding to the risk $\tau = 0.01$ .	149
5.2.4.4	: Predicted number versus actual number of false alarms	149
5.2.4.5	: Division of the accumulator space	152
5.4.1	: Interpretation tree of the 3D directions of line segments	165
5.5.1	: Ellipse associated with two vanishing points.	166
5.6.1	: Initial images 1 and 2	167
5.6.2	: Extracted segments of the images 1 and 2	167
5.6.3	: Accumulator spaces with all the lines accumulated	168
5.6.4	: Accumulator spaces : the vertical lines have been removed	168
5.6.5	: Final classification	169
5.6.6	: Extract of the classification with a vanishing point	169
5.6.7	: Result of the search for triplets of perpendicular directions	170
5.6.8	: Initial images	170



5.6.9	: Perpendicular directions	170
6.2.1.1	: Example of close parallel segments, with and without overlapping.	176
6.2.1.2	: Definition of $D_0$ and $D_1$	177
6.2.2.1	: Definition of the distance $D$ and $D'$	183
6.3.1	: Construction of 3D structures from the direction interpretation tree	185
6.3.2	: Example of a subjective linear structure	187
6.3.3	: Example of comb structures	189
6.3.4	: Consistent scaling of a tooth of a comb structure	191
6.3.5	: Example of an edge built from 2 comb structures	192
6.3.6	: Position of the vertex according to the orientations of the edges.	192
6.3.7	: Visualisation of the 3D structures	196
6.4.1	: Initial image	199
6.4.2	: Detected lines.	199
6.4.3	: linear structures	200
6.4.4	: Comb structures	200
6.4.5	: Rectangles	201
6.4.6	: Edges	201
6.4.7	: Vertices	202
6.4.8	: Propagation of depth to adjacent structures	202
6.4.9	: 3D interpretation, same scene, different viewpoint	203
6.4.10	: 3D interpretation. Same scene, different view-point.	203
6.4.11	: Ellipse interpretation	204
6.4.12	: Same structure under 2 different viewpoints	204
6.4.13	: Initial image of another scene	205
6.4.14	: Scene interpretation of the scene displayed in figure 6.4.13.	205
7.2.1	: Orientation of the axes associated with the scene coordinate system	210
7.2.2	: New coordinate system associated with 2 viewpoints.	210
7.3.1	: PDMS/TIMI matching process	214
7.4.1	: 3D map of the scene	216
7.4.2	: Same map in ROBCAD	217

Figures

*To Stéphane,  
Jérémy, Yohan,  
Emilie and Vincent.*

## ACKNOWLEDGEMENTS

First, I would like to thank very much Electricité de France for funding this work, particularly Mr de Montardy and Mr Pavart who have made it possible.

A special thank you is given to the successive supervisors of this work, Dr Geoff West and Dr Tim Ellis, without whom this work would not have been possible. I particularly thank them for their continual technical support, for the time they spent in guiding my English and most of all for their friendship.

Also many thanks to Claude Mersier from Electricité de France for the numerous and very interesting discussions we have had all along this work.

I wish to thank the Machine Vision Group for the nice atmosphere that existed during my stay with them. I also thank very much Dr Pat Samwell, Dr Sanoa Khan and Caroline Butt for their linguistic and very friendly support.

I thank Prof. Castan from the Université Paul Sabatier (Toulouse) for having provided me with their edge detection program.

The last but not the least, very special thanks are given to my Londonian friends, to my parents, to Stéphane and our children, for having always encouraged me.

Acknowledgements

DECLARATION

I grant powers of discretion to the City University Librarian to allow this thesis to be copied in whole or in part without further reference to the author. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement.

Declaration

## ABSTRACT

The work described in this thesis is concerned with the 3D interpretation of monocular images. First, the perspective transformation is interpreted in an image which enables the extraction of 3D information. Then, connectivity in the image is utilized in order to infer 3D groupings in the scene. The result of the process is 3D structures, leading to local 3D maps, the scales of which are unknown. The processing of several images from unknown viewpoints allows the relative scales of the various maps to be known. Thus, an unscaled but consistent 3D map is extracted. This map has a 3D symbolic representation and may be integrated in a CAD database.

A probabilistic approach is used for interpreting the image. First, two types of error are defined : errors due to the measurement uncertainty and errors due to accidents such as the proximity of unrelated features, called segmentation errors. Because of measurement errors, relations are not exactly fulfilled. Accounting for such errors is responsible for segmentation errors, and thereby unreliability of the process. The best trade-off for checking these relations is based on the maximum likelihood test. In order to determine this test, a precise statistical model of the data is defined. Moreover, accounting for measurement uncertainty leads to an original process for detecting the vanishing points in the image in a consistent way over the space.

Another central theme of this work is the 3D representation adopted for the structures extracted from the image. The intrinsic parameters of these structures are viewpoint and scale invariant and the geometric relationships and the degree of freedom of these structures are implicit in such a representation. This considerably eases the construction of the 3D structures, and then of the local and global maps.

The method is illustrated by the processing of images of indoor



Abstract

scenes of a power plant.

## NOTATIONS

$\vec{u}$	Vector $u$ , element of $\mathbb{R}^n$ .
$\vec{u} \cdot \vec{v}$	Scalar product of the vectors $\vec{u}$ and $\vec{v}$ .
$M$	Matrix, element of $\mathbb{R}^n \times \mathbb{R}^p$ (a vector can be written $\vec{u}$ or $u$ ).
$\partial$	Symbol of derivation
$f'(x)$	Derivative of $f$ with respect to $x$ (i.e. $\frac{\partial f}{\partial x}$ ).
$\varepsilon(x)$	$\varepsilon$ tends towards zero like $x$ .
$\log_e(x)$	Neperian logarithm of $x$ .
$\sin \theta$	Sinus of $\theta$
$\cos \theta$	Cosinus of $\theta$
$\tan \theta$	Tangent of $\theta$
$\cotan \theta$	Cotangent of $\theta$
$E(x)$	Expected value of $x$ .
$E(x y)$	Expected value of $x$ knowing $y$ .
$\text{var}(x)$	Variance of $x$ (i.e. $E((x-E(x))^2)$ ).
$\text{erf}(x)$	$\text{erf}(x) = \frac{1}{\sqrt{2\pi}} \int_0^x \exp(-\frac{u^2}{2}) du$
$n!$	Factorial $n$ , i.e. $n! = n \times (n-1) \times \dots \times 2 \times 1$ .
$\binom{n}{p}$	$\binom{n}{p} = \frac{n!}{p!(n-p)!}$
LMS	Least Mean Square
LR	Likelihood Ratio
MD	Mahalanobis Distance
$\mathbb{R}$	Set of real values
$[a, b]$	$\{x \in \mathbb{R} ; a \leq x \leq b\}$
$[a, b[$	$\{x \in \mathbb{R} ; a \leq x < b\}$
$]a, b[$	$\{x \in \mathbb{R} ; a < x < b\}$
EDF	Electricité De France

## Notations

## CHAPTER 1

## INTRODUCTION

*Context*

Guidance of a vehicle in any environment is a key issue in robotics. A number of methods have been developed, ranging from previous training of the vehicle to the use of exteroceptive sensors as elaborate as vision.

Previous training whereby the vehicle is meant to repeat the same displacements, is a method widely used in an undisturbed environment. Although performance is increased by the use of sensors such as an odometer, a range finder, a tactile sensor which provides information about the vehicle orientation and free space or obstacles surrounding the vehicle, the range of mission remains limited.

Vision is a natural step forward to obtain a higher degree of autonomy for a wider range of mission. It would provide a precise description of the scene and thereby the relative position of the vehicle within it. It is a very flexible tool which does not require any conversion of the environment and may be used to perform other tasks (e.g. inspection, repairs).

The principal difficulty of any method involving exteroceptive sensors is the interpretation of the data, as information contained in each datum, qualified "of low level", is extremely poor. The physical process of image formation is well known, but the interpretation process leading to a semantic description of the scene involves very complicated mechanisms not yet fully understood. An elaborated vision system would aim at providing such a "high level" description of the scene, which, up to now, is not possible without much prior knowledge

## Introduction

of the scene.

Existing methods extract information from the image at an "intermediate level", such as edges, straight lines, geometric patterns. If a model of the projections of an object onto the image plane under a number of viewpoints, i.e. a model of the aspects of an object, is provided, then a matching strategy allows the object to be recognized in the image. Although this may be used to locate an object or conversely the camera (and therefore the vehicle in the scene), this does not yet deal with generic objects. Other methods use several images to extract a coarse 3D map of the scene, that is to say a set of points, segments or curves located in the 3D space.

Whatever the method, much remains to do to increase the level of the image interpretation. Monocular vision cannot allow complete 3D interpretation without additional information which may be provided by other sensors such as a laser range finder, multiple images (stereo vision or motion) or a model of the scene (either heuristic or semantic). However, before choosing the most appropriate method it is useful to know the limits of the interpretation of one image. Anthropomorphic considerations suggest that it is possible to extract qualitative 3D information from a single image by using general knowledge of the scene (e.g. it is generally possible to recognize a scene from its photograph). The general knowledge involved depends very much on the type of the scene. In the case of man-made scenes, such as indoor scenes, knowledge about the regularity of expected shapes, occurrence of some geometrical relationships can be used.

### *Aim*

The main purpose of this thesis is the analysis and interpretation of the information contained in a single image in the 3D space, at the highest level possible, by using general knowledge of the scene. The work is concerned with indoor scenes which can often be represented by a limited number of 3D regular geometric primitives (e.g.

parallelepipeds, cylinders) with privileged directions. The boundaries provide numerous straight lines which are likely to be parallel or perpendicular. For instance, a room has three principal directions : one vertical and two horizontal ones defined by the directions of the room walls. In the present work, heuristic information is used in the form of hypotheses tested on the image primitives, such as "three concurrent lines in the image are likely to be parallel in the scene", in a way similar to Lowe's method for perceptual groupings. Hence a local geometric interpretation of the scene is inferred.

The method described in the next chapters consists of two main parts, first the interpretation of the image perspective and second the construction of local 3D configurations.

### *Perspective interpretation*

The image is first segmented and then edges are approximated by straight line segments or elliptical arcs. Finally the scene is hypothesized to contain at least two principal perpendicular directions with which a number of lines in the scene may be associated. These directions are detected through the interpretation of the perspective of the image which is done in three stages

- Vanishing point detection
- Line direction classification
- Perpendicular directions

The detection of the vanishing points is achieved by using an original accumulator space (using the Hough paradigm), with a total consistency whatever their location in the image plane. Each line is then classified with a vanishing point, by using a Bayesian approach. Perpendicularity criterion between directions fixes principal directions.

## Introduction

### *Construction of local 3D configurations*

Only lines associated with a principal direction have been considered in what follows. Since their orientation is known, they can be represented in the 3D space. They are grouped to form significant structures. For instance, close parallel lines form a *linear structure* then perpendicular *linear structures* form *rectangular structures* and *corners*. These structures are represented in the 3D space relative to the camera coordinate system. In fact, at this stage the situation is similar to the 2½ D sketch introduced by Marr (Marr, 1982) as the distance of the structures from the camera is unknown, i.e. the scale of the representation is unknown. Then the adjacent structures are grouped by using a connectivity criterion in order to form local 3D configurations, the substructures of which have consistent relative depths, i.e. only the global scale is unknown (They are called "3D" because of their representation, with no reference to their actual dimension). The structures are hierarchically organized by increasing complexity.

### *Uncertainty and scoring process*

Robustness is clearly a major concern in any system responsible for vehicle guidance and depends on its response to noise. The presence of noise which results in uncertainty of measurement needs to be taken into account throughout the process. Uncertainty plays a key part in the method described in this thesis. Errors of measurement are hypothesized to be normal random variables and accordingly any result of the interpretation process is also a random variable. The hypotheses are tested by applying the likelihood concept, and a Bayesian method is used for the scoring process.

One difficulty of the Bayesian approach is the estimation of *prior* probabilities. This is done by using a statistical model of the features in the image, justified by experiment. It is shown that from this model, it is possible to define a likelihood ratio test for a

given relationship, e.g. parallelism or connectivity. This test is proved to be more reliable than other popular tests, such as the Mahalanobis distance test or the neighbourhood test.

### *3D Representation*

The construction of the "3D configurations" described previously aims at demonstrating that the use of 3D representation from an early stage is powerful. The geometric relationships, such as viewpoint invariant geometric properties, are shown to be implicit in this representation, so is the degree of freedom of the structures extracted.

### *Application*

In order to illustrate the possibilities of the method described, it is applied to the construction of 3D maps of the scene using several images grabbed by the camera during the movement of the vehicle. It is shown that two images allow the determination of the relative scale and location of the structures, without requiring any information about the relative positions of the camera. The merging process of several maps is simplified by the representation of the local configurations in the 3D space. Notice that the general scale of the map remains unknown. The software developed is called TIMI (Three Dimensional Interpretation of Monocular Images). The representation of the scene extracted by TIMI is consistent with some CAD representations (e.g. ROBCAD's representation). Matching with two CAD databases using a different representation is discussed, although the matching process is beyond the scope of this thesis.

### *Industrial application*

The general context of this work is the inspection of nuclear plants of Electricité de France (EDF) by autonomous vehicles. Up to now remote controlled vehicles are used, but this requires skilled staff and limits the set of possible tasks. The environment is known,



## Introduction

but it may be subject to some variations. It is specific : numerous lines and cylinders (e.g. pipes, tanks). In hostile environments, robustness obviously is a key point. A CAD database of these plants may exist. The aim of the project is to match the image grabbed by a camera with a possible viewpoint in the CAD database, in order to locate the vehicle in the database coordinate system. But the CAD database has been designed for purposes other than vehicle guidance by vision. Therefore the database has to be transformed and enhanced to be used for such an application. The work done should demonstrate the possibilities of monocular vision for guidance with respect to a CAD database. It should also determine what type of additional information would be the most useful to improve robustness.

### *Organization of the thesis*

Chapter 2 analyzes the overall characteristics of similar existing systems: type of scene processed, method chosen, type of representation used.

The context of this work, the problem to be solved and the general principles of the method are described in chapter 3. A unified approach is used throughout this work for testing a relationship hypothesized between 2 features by using a maximum likelihood approach. It is also described in this chapter.

Chapter 4 is concerned with the preprocessing of the image, edge detection and line finder algorithms. A section is concerned with the modelling of the measurement uncertainty and the statistical distribution of the feature parameters extracted from the image. Determining the camera parameters is also discussed in this chapter.

Chapter 5 describes the vanishing point detection method and the line classification. This method is tested on a set of images of indoor scenes of power plants.

Chapter 6 describes the construction of the local 3D configuration.

After the method description, results on various images are provided and discussed.

The application of the method to map construction from several views is described in chapter 7. The representation used is compared with the representation of the data of two existing CAD softwares, ROBCAD and PDMS. It is converted into ROBCAD's representation.

The method is discussed in chapter 8. Its major advantages and drawbacks are pointed out.

The method sometimes requires fastidious mathematical developments which are described in the annexes. Some of them are original, others are not but are given for clarity and completeness.



## CHAPTER 2

## SURVEY

The following survey aims at providing an analysis of the difficulties and constraints encountered when designing a vision system, based on previous work achieved in this field. Rather than describing consecutively the existing systems, we had better centre the discussion around principal thema of vision.

For interpreting a photograph, a human extracts meaningful information from the image and compares it to similar information obtained from past experience. What information has been extracted and in which form it has been stored are generally unknown; this is a fundamental problem for artificial vision. The following sections are organized around the nature and the representation of the information used by a vision system. There are various types of knowledge which may be used for interpreting an image, they have been classified as follows : knowledge about the type of scene and the conditions of image acquisition, called general knowledge ; precise knowledge about the objects present in the scene, i.e. knowledge of a model. An image interpretation process may be roughly schematized in the following way :

- general information is represented within the process (geometric reasoning, properties exploited and various approximations).
- model information and data extracted from the image are the input of the process through which they are interpreted and compared.

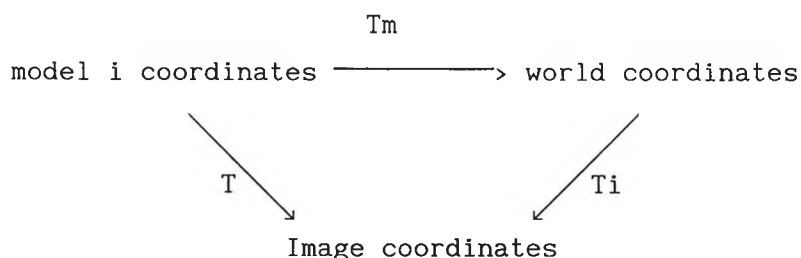
## 2.1 AIMS AND DIFFICULTIES OF A VISION SYSTEM

### *Field of application*

Two linked problems in robotics vision have been the subject of much research. These are :

- To find the location of unattached objects in world coordinates, when the location of the camera is known by a previous calibration.
- To find the location of the camera (i.e. of the robot) by reference to fixed objects.

They may be formalized as in figure 2.1.1 :



$T_m$  : transformation from the model coordinates to the world coordinates

$T$  : transformation from the model coordinates to the image coordinates

$T_i$  : transformation from the world coordinates to the image coordinates

Two types of problem: \*  $T_i$  known,  $T_m$  is searched for

\*  $T_m$  known,  $T_i$  is searched for

Solution: to look for  $T$

Figure 2.1.1 : Object location or viewpoint determination

Although both problems : object location and viewpoint determination, can be formalized in the same way, they present different difficulties, because the types of scenes involved are very different (see section 2.2). It is well known that the conception of a vision system depends very much on the field of application.

Figure 2.1.2 classifies some well-known operational systems with

their main field of application. The same systems have been used later in figure 2.3.1.

Up to now, these systems mainly deal with recognition and location of unattached objects (RAF, Grimson and Lozano-Perez 1984). Albeit potentially more general, ACRONYM (Brook, 1984) and SCERPO (Lowe, 1985) have only been demonstrated on recognition of such objects. The field of application of these systems includes automatic inspection (3DPO (Bolles and Horaud, 1986), HYPER 2D (Ayache, 1985); CAIMAN (Lux, 1985)).

However, over the last few years more general scene understanding systems have been developed, particularly for vehicle guidance. Stereo or motion based vision methods allow the extraction of 3D information and thereby have been quite popular (Ayache, 1988; Pollard et al, 1989; Brown, 1989; Crowley and Stelmaszyk, 1990). Recent research has been to explore the possibilities of monocular vision for scene recognition purpose (Sugihara, 1988; Quan and Mohr, 1988; Coelho et al, 1990).

Another important application for vision is in highly specialized fields such as radiography, astronomy or microscopy, which mostly involves 2D recognition process. The objects in the image are classified according to a number of more or less general criteria (Tsuji and Nakano, 1981; Granger, 1985).

#### *Method*

For every case, the problem consists of extracting relevant information from the image which cannot be done without using prior knowledge of the scene and of the conditions of the image acquisition. The use of this knowledge during image interpretation process is schematized on figure 2.1.3.

	Isolated objects	Environement	Classification
3D/2D systems			
ACRONYM	*		
SCERPO	*	*	
VISIONS			*
SYGAL	*		
FABIUS			
2D/2D systems			
HYPER 2D	*		
3D/3D systems			
HYPER 3D	*		
3DPO	*		
IMAGINE	*		
Other systems			
RAF	*		
STEREO (INRIA)	*	*	

Figure 2.1.2: Application field of various existing systems:

- Identification and/or inspection of loose objects
- Interpretation of natural or artificial scene and/or location of the camera within this environment
- Object classification

General knowledge may be used at any level of the interpretation process, from image segmentation to 3D map construction (dotted lines in figure 2.1.3). The information stored in the model is compared to the features extracted from the image by a matching process at the feature level (arrow lines in figure 2.1.3). The method, which predicts visible features from the model, identifies them in the image and then characterizes them, is called a top down process. The extraction of high level features from the image using general knowledge of the scene and image formation, is called a bottom-up process. Both processes can collaborate (ACRONYM, Brook, 1984; VISIONS, Hanson and Riseman, 1978, 1988). Typically, a specific application, i.e. much prior knowledge, allows a top down process to be efficient (e.g. restriction on the set of possible viewpoints and the set of objects present in the scene allows a

prediction-verification method to be efficient (Ayache, 1983)). Conversely, autonomous guidance of a vehicle should use a minimal amount of information to be of general purpose, which requires an elaborated bottom-up process.

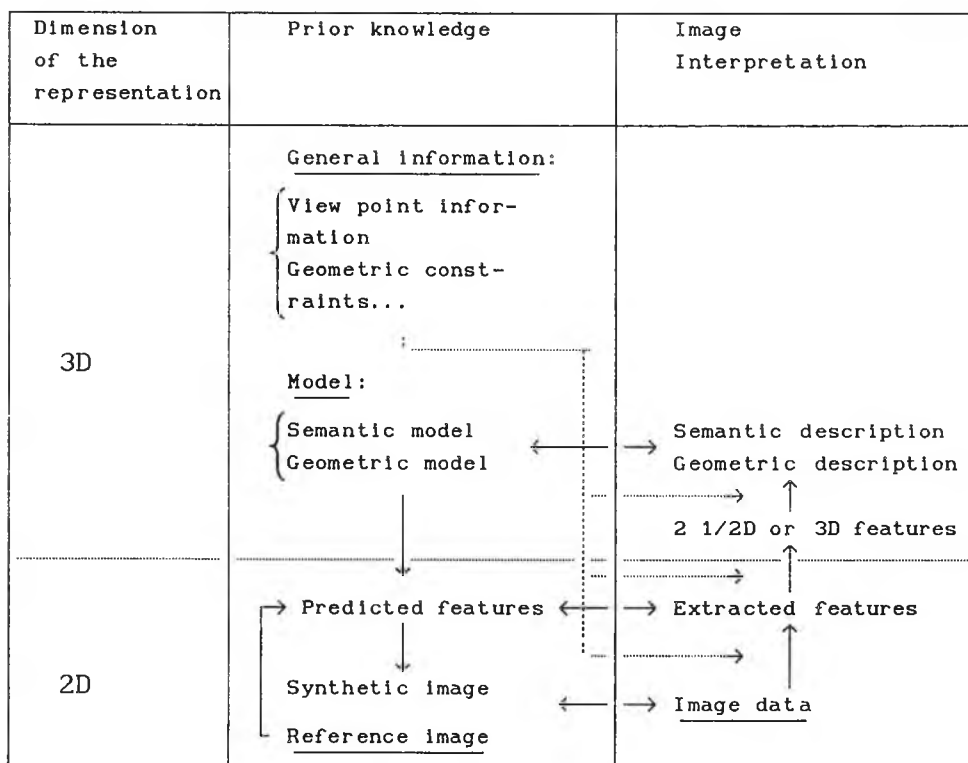


Figure 2.1.3 : Prior knowledge and image processing

Therefore, the matching between the model and the image features can be achieved at different levels :

- At the low level by correlation between the real image and a reference image or a synthetic image deduced from a geometric model (Even and Marse, 1988). The viewpoint or at least a finite range of viewpoints should be known.
- At the intermediate level by matching either geometric 2D features extracted from a geometric model with 2D image



features (Brooks, 1984), or 3D model features and 3D features inferred from 2D image features (Lowe, 1985). As data is represented in a compact form (symbolic form), more viewpoints and more complex models may be processed. Matching at a 3D level has the strong advantage to be viewpoint independent, thereby reducing the combinatorial complexity of matching.

- At a high level, i.e. a semantic level, by taking into account the context of the image features and the relations between them (Granger, 1985; Simoni, 1988; Rosin, 1988). This approach may considerably reduce the combinatorix of matching if a precise semantic description is available in the model, which is in general not the case for an indoor scene (see sub-section 2.3).
- At different levels simultaneously, as information becomes available in a blackboard (VISIONS (Hanson and Riseman, 1988)).

### *Difficulties*

The main difficulties encountered by a vision process are linked to the three following phenomena :

- Noise: information in the image is embedded in noise. This may be due to the presence of shadows and reflections, to the camera distortions, to electronic noise and to digitization. Noise results in substantial uncertainty of the image segmentation process, which affects the significance of the information extracted as well as its accuracy.
- Connectivity : connectivity in the image is poorly related to connectivity in the scene. This is a major problem in the construction of high level primitives, such as a parallelepiped or cylinder, which are successfully used in CAD databases for scene description. Lack of consistency in connectivity is due to occlusion and to noise.
- Combinatorial complexity : Numerous objects may correspond to a single projection in the image. Even the conceptual objects

which are expected to be present in the image are limited in number, the variety of their instantiations and the number of the associated aspects makes the recognition problem extremely complicated. In the case of a precise description of the scene being available, the combinatorix of possible assignments between model features and image features is still of exponential type.

### *Some solutions*

These difficulties are reduced by using much prior information about the application, e.g. the effect of noise may be reduced if small details may be ignored, the connectivity of edges may be enforced if the scene only contains sparse objects, the combinatorix of matching may be considerably reduced if there is a finite range of possible viewpoints. Currently, vision systems are designed according to the type of the scene, the type of knowledge available and the performance required.

Various types of "prior" knowledge of the scene may be used for vision, from general information to a precise geometric model of the scene. They are used in different ways ; general information affects fundamental choices of method, whereas precise models are explicitly part of the data. Section 2.2 mentions various types of general knowledge which are used in image interpretation. This knowledge may be imprecise so the estimation of the reliability of the results is part of the problem and is discussed in the same sub-section. Sub-section 2.3 deals with model representation.

The image data are very numerous and very crude. To extract essential information, the image is segmented into features, having a symbolic representation where possible. This process decreases the number of data to process and increases their significance. From elementary features, e.g. straight line segments or vertices, it is possible to construct more elaborate features, e.g. vertex-pair or

polygons. The higher the level of the features, the less numerous the data to process and the more significant the interpretation. Some of the most popular features are described in section 2.4.

The choice of the method depends on the property which is assumed to prevail in a particular situation, e.g. accumulation of evidence for the viewpoint location is an appealing method if the scene is mainly composed of fixed objects. However the method chosen should have essential mathematical properties to produce sound results, e.g. stability of the results with respect to the data. Data representation is a key point of the formalism used. Essential properties of the method used and data representation (i.e. feature representation) are the subjects of the sub-section 2.5.

Control structure has not been the subject of research in this work. Therefore, albeit it is an important aspect of a vision system, it has been ignored in the following sections.

## 2.2 GENERAL KNOWLEDGE

The *prior* knowledge of the scene may consist of knowledge of the viewpoint and/or the geometry and the complexity of the scene. The following examines some common types of *prior* knowledge exploited by vision systems.

### *General knowledge about the geometry of the scene*

If the scene is a man-made scene, e.g. a room, the shape of the objects are likely to be approximately polyhedral with corners likely to be rectangular. This constraint considerably reduces the number of possible interpretations. Polyhedral objects have been the subject of much attention (Roberts, 1965 ; Nevatia, 1982 ; Sugihara, 1984 ; Nelson and Young, 1985 ; Kanatani, 1989), and among them objects generated by blocks. Huffman and Clowes (huffman, 1971)(Clowes, 1971) give a list of the possible corner aspects of a blocks world scene.

Mackworth (Mackworth, 1973) shows that the graph generated by the projection of a cubic object has a dual graph, a constraint useful to recognize consistent edges. Sugihara (1984) shows that the image of a polyhedral object obeys a number of algebraic constraints which can be used for inferring 3D shape from a single image.

Using the same type of considerations, Lowe (Lowe, 1985) introduces the concept of *perceptual groupings*, i.e. typical configurations of features in the image, for 3D inference. This relies on heuristic knowledge such as: "in a man-made scene parallel lines are numerous so that it is reasonable to suppose that parallel lines in the image are parallel in the scene". A summary of such reasonable inferences is displayed in figure 2.2.1.

Heuristic knowledge also sustains the interpretation of the perspective in the image : "If a number of lines are concurrent in the image, they are likely to be parallel in the scene" (Banard, 1983; Magee and Aggarwal, 1984; Quan and Mohr, 1989; Kanatani, 1989).

The use of heuristic knowledge enables the selection of a number of reasonable solutions among an infinity of them and thereby is an important step forward in the interpretation. However it does not provide a unique solution and the evaluation of the likelihood of each "reasonable" solution is still a problem. Moreover, this often excludes interpretation of natural scenes because such geometric information is rarely available.

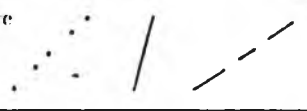
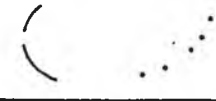


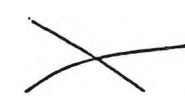

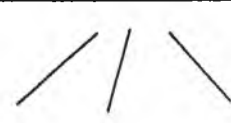
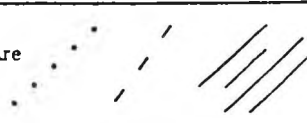
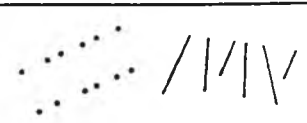
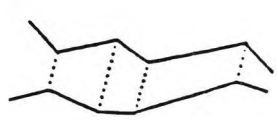
2-D relation	3-D inference	Examples
1. Collinearity of points or line segments	Collinearity in three-space	
2. Curvilinearity of points or arcs	Curvilinearity in three-space	
3. Two or more terminations at a common point	Curves terminate at a common point in three-space	
4. Termination at a continuous curve	Terminating curve is no closer to the camera than the continuous curve	
5. Crossing of continuous curves	Both curves cannot be occluding geometric edges	
6. Parallel curves	Curves are parallel in three-space	
7. Three or more lines converge to a common point	Lines are parallel (seen in perspective) or converge to a common point in three-space	
8. Equal spacing of collinear points or parallel lines	Equal spacing in three-space and parallel lines are coplanar	
9. Relations hold between terminations or virtual lines	Same relation holds between virtual features in three-space	
10. Parallel virtual lines between tangent discontinuities in curves	Curves correspond to geometric edges and their cast shadow boundaries	

Figure 2.2.1: Inferences proposed by Lowe (Lowe, 1985)

Semantic knowledge may concern any type of scenes and increases the level of the interpretation. ANDES (Simoni, 1988) uses semantic knowledge for defining the geometric relationships in the image. It bases the recognition on an expert system, the knowledge representation of which are a set of rules representing the spatial

organization of the scene, e.g. using predicates such as *above*, *inside*. The efficiency of such a description is limited by its simplicity (see sub-section 2.3).

#### *General knowledge about the viewpoint*

The viewpoint may be limited either in range, e.g. a camera may be located in a corner of a room and have a limited number of degrees of freedom, or in number, e.g. the camera is fixed and the object has a finite number of stable positions (Ellis et al, 1988). In the latter case, a 3D problem is reduced to a 2D problem. The camera may also be supposed far from the object so that the perspective transformation may be approximated by an orthographic projection with a scale factor (Thompson and Mundy, 1987). These approximations limit very much the field of possible applications.

#### *Complexity of the scene*

The combinatorial complexity of any matching process is directly related to the complexity of the scene. Strategies for pruning part of the interpretation tree, i.e. tree formed by all possible matches, are of increasing importance as complexity of the scene increases.

If the scene is composed of sparse objects with rather convex shapes and uniform colour on a uniform background, then a region growing algorithm (Rosenfeld and Kak, 1982) or any algorithm enforcing connectivity (Reis, 1991) may be very efficient and so reduces the combinatorial complexity of the matching. However the efficiency of these tools is compromised when the complexity of the scene increases (where it is more needed!).

#### *Measure of evidence*

Whatever the application, various knowledge is used to reduce the difficulties encountered by a vision system, but this also restrains the field of application of the system. A compromise has to be found

between powerful constraints and generality. The use of precise information such as: "the number of possible viewpoints is limited", clearly very much limits the range of applications. Fuzzy heuristic information such as: "numerous lines are *likely* to be parallel" is less limiting but more difficult to handle, and raises the difficult question of reliability of the results.

Fuzzy heuristic information allows hypotheses to be tested on a set of image features but it should provide a *prior* estimate of the probability of success. In practice many systems accumulate evidence for hypotheses until inconsistencies arise. For instance an edge in an image is generally assumed to correspond to an object boundary until evidence for the contrary, e.g. failure in matching with a model feature. Mulgaonkar et al (Mulgaonkar and Shapiro, 1985) test a hypothesis through an inference engine, the rules of which are geometric relationships between 3D configurations of straight lines and their projection onto the image. Such methods give little information about the reliability of the final result.

Incorporating probability reasoning provides a firm basis for interpretation as long as rigorous statistical analysis supports the model. Tsuji & Nakao (1981) achieve an interpretation of a cine-angiogram system. The images vary significantly from patient to patient and do not allow a geometric description. The knowledge is represented as a set of rules, associated with a probability. They obtain good results with a difficult problem. Lowe uses the concept of significance, based on a Bayesian approach, in order to measure the likelihood of the hypothesis tested. The same approach is used by Rosin (Rosin, 1988), in the model-based recognition system FABIUS. Dickson (Dickson, 1990) generates a network where the features are the nodes and the hypotheses the arcs, in a way similar to Pearl's method (1988), and a likelihood measure is propagated through the network. The difficulty of the Bayesian approach used in the previous methods is the estimation of the prior likelihoods, which is usually

arbitrarily performed.

As the interpretation of the scene from one sensor to another sensor may vary substantially, the estimation of the reliability of the results is crucial in multi-sensor fusion. Most research for evidence qualification has been done in this context (Hackett and Shah, 1990). A popular method is based on the Dempster-Shafer theory (Shafer, 1976) according to which a probability measure does not give enough information about the possible occurrence of a hypothesis. Other measures are introduced such as the possibility measure and the plausibility measure, providing an interval of values associated with a set of evidences in a way similar to the uncertainty interval associated with a physical measure.

In spite of the difficulties, handling fuzzy information seems to be essential for the development of a vision system of general purpose. Nevertheless none of the systems described in the previous paragraph has been tested on a large range of applications and cannot be fully evaluated.

General knowledge embedded in the method is an essential component of any vision system. It may be extremely various in nature but is generally the heart of the method, e.g. perceptual groupings of Lowe (1985), algebraic constraints (Sugihara, 1984), affine transform (Thompson and Mundy, 1987). However, its choice remains arbitrary and there is no consensus on the way to represent and process it. Furthermore it may be considered as a set of hypotheses which limits the field of application.

### 2.3 SCENE REPRESENTATION

In this sub-section, various types of models of the scene are described and discussed in contrast with general knowledge.

General knowledge described in the previous sub-section usually is



embedded in the processing itself. In contrast, the use of models for object recognition is explicitly part of the data, e.g. in an expert system approach a model is stored in the fact base. General knowledge is used in a bottom up stage (i.e. to extract symbolic information from the image) whereas the model is usually used in a top down stage (i.e. the image feature corresponding to a particular model feature is looked for). This latter strategy is called a model-driven strategy. Numerous types of models and representations have been investigated by researchers.

The representation will greatly depend upon the type of scene studied and the source of the information. There are various ways of building a model with which the image features are to be compared, from reference image acquisition to a complete symbolic description.

#### *From reference images*

The reference image acquisition is probably the best way to model precise and maybe complicated objects, e.g. cast components. The technique for segmenting the model is the same as the one used for segmenting the image. Ayache (1985) in the system HYPER, Lux (1985), Grimson (1987), model 2D objects randomly placed and partially overlapped, using a polygonal approximation. The acquisition of a number of viewpoints corresponding to the various aspects of the object enables recognition of 3D objects.

This type of modelling gives good results as the features of the model are very likely to be matched with the features of the image since they are extracted by the same process. However, it limits the recognition to precise objects. It requires a limited number of possible viewpoints and does not allow geometric manipulation.

Faugeras and Herbert (1986) achieve 3D object modelling by interpolating 3D points acquired on the object by a set of facets. Fisher (1987) uses quadratic surface interpolation, whereas many researchers prefer bi-cubic surface interpolation which is simpler to

implement and enables accurate modelling of a wide range of objects (Boult, 1985). The problem with both quadratic and bi-cubic interpolation is that a local change in the acquired points induces a complete change in the model. Nevatia and Binford (1977) fit generalized cylinders to real image data but the stability of the results has not been demonstrated.

### *Geometric modelling*

The object to recognize is described as a set of simple geometric forms. This considerably limits the set of objects which can be modelled. The natural environment can not be easily modelled in this form, but many man-made scenes are designed with regular shapes.

CAD tools may be used to achieve geometric modelling. The most common types of CAD representation are constructive solid geometry (CGS) and the boundary representation. The CAD representation of the database used by Electricité de France for modelling power plants and particularly pipes, PDMS (PDMS) is of CGS type. The object is represented by a number of volumetric primitives, e.g. parallelepipeds or cylinders. The boundary representation represents an object by the set of linear primitives, e.g. straight line segments or circular arcs. This latter representation may appear more suitable for vision, but a CAD database may exist independently of the vision task and one may wish to use it as such. Ellis et al (1987) use a CAD representation of widgets to recognize their location and to inspect them.

Binford (1975) introduced generalized cones to represent a larger set of objects. Brooks (1985) used this idea with restriction on the definition of generalized cones. He succeeded to model objects as complicated as a class of aeroplanes. ACRONYM (Brooks, 1985) successfully uses this type of representation to identify particular types of aeroplanes from a range of models.

Geometric modelling is associated with a symbolic description enabling the parameterization and thereby the representation of generic objects. The representation is compact and enables geometric reasoning. However the range of objects which can be modelled in this form is limited.

### *Semantic description*

We define a semantic description as a set of qualitative rather than quantitative descriptors. It includes symbolic form descriptors (circular, lengthened...), color (dark, bright or any color), and predicates such as spatial relations (above, below, included...), order relation (larger, smaller...). It is well adapted to describe a natural environment as it offers a great flexibility for the description.

A symbolic description can be represented as a graph, called a semantic network, where the nodes are objects, classes of objects, descriptors or situations and the arcs are the relations between the nodes. For instance a circular object will be represented by two nodes "object" and "circular" connected by the arc "shape".

A popular knowledge representation is the "frame" representation (Minsky, 1975). A frame is a multi-level representation of an object. With an object (a frame) are associated attributes or slots (sort of, part of, color, texture...) described by facets (value, default...). The frames can be connected to form a semantic network. It enables the modelling of a class of objects from which the objects inherit the properties. A geometric description may also be associated with it.

The frontier between general knowledge and modelling when using semantic knowledge is not as clear as in previous paragraphs. Actually, the distinction is made through the structure of the algorithm, whether prior knowledge is stored in the fact base in (Rosin, 1988; Granger, 1985) or in the rule base (Tsuji and Nakao,

1981; Simoni, 1988).

The semantic representation is very well adapted to highly specialized fields, such as biology or astronomy (SYGAL (Granger, 1985)). The knowledge of the expert, formalized in terms of image, is translated into the database. However, albeit scene interpretation is an elementary task for everybody, it is very difficult to formalize it in terms of image and it is still the subject of extensive research in neuroscience.

Figure 2.3.1 classifies some operational systems according to the type of representation adopted to describe the models.

	Ref. image	Geom. model	Semantic desc.
<b>3D/2D systems</b>			
ACRONYM	*	*	
SCERPO		*	*
VISIONS			*
SYGAL			*
FABIUS	*	*	
<b>2D/2D systems</b>			
HYPER 2D	*		
<b>3D/3D systems</b>			
HYPER 3D	*		
3DPO		*	
IMAGINE	*	*	*
<b>Other syst.</b>			
RAF	*		
STEREO (INRIA)		*	

Figure 2.3.1 : Model type for various systems

- Model from reference image acquisition
- Geometric models as CAD database
- Semantic description

#### *Organization of the database*

The organization of the model database depends on the type of scene. If it consists of unattached objects, the object models are stored in a catalogue, which may be hierarchically organized if objects are formed of sub-objects (Brooks, 1985). In the case of

indoor scenes, the notion of object is ambiguous. For instance the effect of scale can alter its definition. Let us suppose that the database can be structured as a highly interconnected hierarchical tree. The top of the tree corresponds to a coarse description and the bottom to a detailed description of the scene. For example a room may be represented by a parallelepiped or more precisely as the set of six rectangles corresponding to the walls, the floor and the ceiling, the description of which may be refined. It appears that different trees may be associated with the same scene, e.g. whether the frame door is classified as a door element or as a wall element. Besides, in spite of the description refinement it remains probable that most of tree leaves will only be partially visible.

The model database complexity rapidly increases with the complexity of the scene. The problem is further aggravated in an indoor scene because the notion of object may be ambiguous, e.g. due to the connectivity of the different elements.

## 2.4 EXTRACTION OF FEATURES

Raw image data are too numerous and have poor meaning when isolated. To consider only meaningful information, the image is segmented and represented as a set of symbolic features. The quality of the interpretation depends on the complexity level of these features; the highest level is a semantic description of the scene or a geometric description of the CSG type. Features commonly extracted from the image are described in what follows.

Information is extracted from the image in the form of image features or primitives, e.g. edges, points, regions. They are then processed to give more significant primitives such as straight line segments, angles, circular arcs. These primitives may be grouped to provide higher level primitives. Hanson and Riseman (1978; 1988) in the system VISIONS use the primitives "region" and "edge" which may be

grouped into a structure called a token. A set of attributes is associated with each type of token. Thompson and Mundy (1987) have introduced the primitive called the "vertex pair". It consists of a pair of vertices, each formed by the intersection of two edges; the angles between the spine, i.e. imaginary line linking the vertices, and the edges meeting at one vertex are known (see figure 2.4.1).

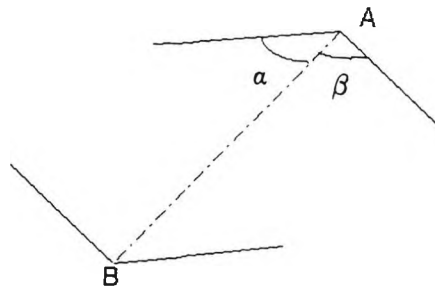


Figure 2.4.1: Vertex pair

Work has been done to increase the level of the primitives extracted by interpreting them as a projection of 3D features. The interpretation of the image perspective gives information on the 3D orientation of the primitives. The primitives used for such purpose are straight line segments (Ballard, 1982 ; Magee and Arggawall, 1984 ; Quan and Mohr, 1988). Coelho et al (1990) recently use such information to construct a 2½ D sketch (Marr, 1982), i.e. a set of planar regions, the orientation of which is known. The level of the interpretation is significantly increased, but the interpretation relies on a number of hypotheses, the likelihood of which varies.

A matching method based on stereo images gives the position of the primitives in the 3D space (Medioni and Nevatia, 1985 ; Ayache, 1988 ; Marapane and Trivedi, 1988). The primitives, typically points, straight line segments or regions, do not usually appear to be connected in the 3D space. Matching 3D segments with a CAD model of the scene is difficult because the boundaries which seem essential to the object description are not necessarily the ones which are easy to

detect in the image. Furthermore segments do not give any information about shape which is necessary when using a CSG representation. Although the combinatorix of the matching is reduced by the 3D representation of the features, grouping them remains a necessity when the complexity of the scene increases.

Increasing the level of features extracted from the image by constructing complex primitives such as vertex-pairs or polygons, requires solution of the connectivity problem. Up to now this is largely solved by setting thresholds on the distance between segments, through experience (Coelho et al, 1990 ; Reis, 1991). As far as the author knows, there is no well formalized solution to this problem.

Characteristic features are extracted from the image and represent the input data of an interpretation process or a recognition process. These features may be grouped or interpreted in the 3D space to increase the level of the representation. The performance and generality of the system depends very much on the achieved level of these primitives. Up to now, among the most significant primitives let us mention: (2D) the vertex-pair (Mundy and Thompson, 1987), (2 $\frac{1}{2}$ D) the oriented planar region (Coelho et al, 1990), (3D) straight line segments located in the 3D space (Ayache, 1988).

## 2.5 METHOD AND DATA REPRESENTATION

The choice of the method depends on the knowledge available and on the properties which are assumed to prevail in the application studied. It also depends of the level of the interpretation one wishes to reach (see figure 2.1) or one can possibly reach. Besides, the method should obey a number of mathematical properties :

- Resolution : the problem should be resolvable
- Stability : a small variation of the data should produce a small variation of the results. This includes stability with a

change of scale or a rotation.

- Efficiency and accuracy : the method should provide the correct solution in most cases, with a minimal uncertainty neighbourhood.

Data representation appears to be a key point of the method. Faugeras (1988) says that a representation should be :

- unique : a unique set of parameters,
- complete : a feature has always at least one representation,
- minimal : minimal number of parameters,
- differentiable : a small variation of the parameters correspond to a small variation of the feature.

To satisfy all the previous criteria is extremely ambitious, furthermore they are not of equal importance and some researchers have given preponderance to some of them. Their relative weight in the choice of the method is demonstrated through a number of examples.

#### *Data*

The data of an interpretation or recognition process are the primitives of various levels extracted from the image (segmentation processes are not investigated here). Whatever this process, it may be formalized as a system of equations with a number of unknown variables.

#### *Resolution of the system*

The system of equations may appear to be complex to solve, from a mathematical point of view, or even not resolvable by algebraic means. A classical way to deal with such a system is to linearize the equations, e.g. Newton-Raphson method for pose determination (Lowe, 1985) or an extended Kalman filter for ellipse fitting (Porill, 1989). Linear algebra is very well known, numerous tools and results are available, thereby linearity becomes a crucial property when the



complexity of the problem increases (e.g. dimension  $\gg 1$ ).

Perspective transformation is inherently not linear in the Euclidean space. The problem may be solved by various approximations. If the viewpoint is approximately known (a small movement since the last positioning) the problem is linearized around this value (Worrall, Baker and Sullivan, 1989). This method may be used in a more general context by using an extended Kalman filter. In that case, the initial guess is associated with an infinite uncertainty (the stability of the process is still a problem). If the object is far from the camera, the perspective transformation may be approximated by an orthographic projection associated with a scale factor (Thompson and Mundy; 1987). However, approximations are not always possible. Naeve and Eklundh (1987) suggest that the right context for solving the problem posed by perspective geometry is projective geometry, as most equations become naturally linear. The elegance of the formalism is very attractive but the associated data representation is not unique, not minimal and not differentiable. For instance, a straight line segment is represented by 6 parameters, 3 of which are used for the detection of the vanishing points (in contrast with 2 using a minimal representation), and the uncertainty of the vanishing points detected is not bounded.

The problem may also be under-determined, e.g. the use of three points for determining 3D object position from a single perspective view (Wolfe, 1988) or the use of three edges for the same purpose (Dhome et al, 1989). In this case all the solutions are found and the ambiguities must then be solved by another process.

#### *Stability*

A small variation of the data should produce a small variation of the results. This obvious property of any reliable system is not trivial in image interpretation.

Differentiability of the system is a necessary condition and applies equally well to data representation. However it is not sufficient, as differentiability is only concerned with small change of the values of the data but not of the data themselves, e.g. missing primitives or segments cut into two segments, the consequences of which are often unpredictable. Methods have been developed to remedy some aspects of the problem. For example, Quan and Mohr (1988) accumulate straight lines by weighting them by their length in order to be insensitive to over segmentation. The use of an edge fragment in the vertex-pair primitive definition (Thompson and Mundy, 1987) aims at the same property. More generally, invariance of the features under a number of transformations is the best way to prevent instability, e.g. the popularity of the vertex primitive is due to its invariance through rotation, change of scale and translation (Whitten, 1988).

The presence of a number of thresholds in a vision process gives little chance to perfect stability of the results. These thresholds are due to the necessity of deciding about the significance of the features or relationships between features throughout the process. This cause for instability of the process is related to the paragraph *measure of evidence* of the sub-section 2.2.

Currently, the use of differentiable and invariant representations is the best way to avoid instability.

#### *Uncertainty and accuracy*

The system may appear to be over-determined when  $n$  features are used to determine  $p$  parameters ( $p < n$ ). For instance, the determination of the transformation between the model coordinate system and the camera coordinate system represented by six parameters is performed by matching  $n$  image features with  $n$  model features ( $n > 6$ ). In fact the equations are fulfilled within a range of uncertainty, so that the over-determination is broken.

A classical way to handle uncertainty is to extract and solve consistent sub-systems, e.g. six non-coplanar points determine a solution for the transformation, then to accumulate the solutions using the Hough paradigm. The solution having the maximum number of votes is selected, i.e. maximum peak in the accumulator space (Thompson and Mundy, 1987). The method does not give a way of estimating the uncertainty of the result.

Another popular way is to minimize an error criterion such as the least mean square criterion (LMS)  $E = \sum_i w_i \varepsilon_i^2$ ,  $\varepsilon_i$  being the distance between the solution searched for and the solution of  $i^{\text{th}}$  sub-system, and  $w_i$  any weight. If the noise responsible for  $\varepsilon_i$  is assumed to be normal, then it is possible to show that the solution which minimizes the global uncertainty, minimizes  $E$  if  $w_i$  is equal to the inverse of the variance of  $\varepsilon_i$  and that it also maximizes the likelihood function. If the system is linear, this solution may be found by using the weighted LMS method, or a Kalman filter (Kalman, 1960) which, in this case, is similar to an iterative LMS method. These methods also provide the uncertainty neighbourhood of the solution. If the noise is not normal with zero mean, they may give very wrong results by emphasizing the importance of a large value of  $\varepsilon_i$ . In this case Weiss (1988) proposes a method based on a prior estimation of the result and segmenting the data by maximizing a likelihood function.

These latter methods are mainly used because they provide the best estimate of the solution in the LMS sense or the likelihood sense, with the associated uncertainty (Ayache, 1988; Porill, 1989; Deriche et al, 1990).

### *Representation*

The linearity of the system of equations depends on the data representation chosen, e.g. the equations giving the coordinates of the intersection point of two straight lines are linear when using a

Cartesian representation but are not when using a polar representation. The use of homogeneous coordinates (Thompson and Mundy, 1987; Ayache, 1988) or reduced coordinates (Kanatani, 1989) in a perspective transformation aim at providing linear systems.

Linear systems are very attractive because they are resolvable with the maximum accuracy in the LMS sense. However the data representation associated with these systems does not necessarily satisfy the criteria set by Faugeras.

Let us focus on a popular primitive, the straight line segment. In the Euclidean space the Cartesian equation of a straight line, i.e.  $ax+by+c = 0$ , is not minimal ; the Cartesian equation normalized by  $c=1$  is minimal but non differentiable at the origin and its polar equation is not unique. There is no representation satisfying the four criteria for this primitive (Ayache, 1988). Ayache proposes to use a set of unique, minimal and differentiable representations, called an atlas, the set being complete . However, the four criteria are not of equal importance : if the representation should obviously be complete, it should not necessarily be minimal or unique. The price to pay for not satisfying both these last criteria is an increase of the dimension or complexity of the problem. The differentiability allows uncertainty to be controlled throughout the process and ensures a better stability (see paragraph on stability, section 2.5).

Cartesian, homogeneous or projective representations have privileged coordinate axes. They are not appropriate for highlighting, for example, the isotropy of some geometrical relationships, e.g. angular properties. Thus, it is sometimes not possible to satisfy the set of fixed criteria or linearity without compromising a fundamental property of the relationship exploited. A trade-off has to be found between good mathematical properties of the representation and the intrinsic qualities of the representation.

*Conclusion on the methods and data representations*

Simplicity and accuracy of the interpretation method depends on the data representation. Many methods aim at providing linear systems as they offer numerous tools including uncertainty calculation mechanisms, such as the Kalman filter, and solutions, such as the best estimate in the LMS or likelihood sense. For instance, in the context of projective geometry, perspective equations become linear (Naeve and Eklundh, 1987). The data representation associated with such systems is not always optimal in the sense defined by Faugeras (1988), nevertheless it is possible to define such a representation, e.g. an atlas (Ayache, 1988).

Approximations, choice of the prior knowledge, choice of the features, choice of the data representation may modify the nature of the initial problem. A trade-off has to be found between complexity, mathematical relevance and accuracy. Up to now no unified formalism exists for vision.

## 2.6 CONCLUSION

The nature of the information used and the way it is used in some existing vision systems have been the subjects of the previous sub-sections. Common difficulties when designing an image interpretation process have been pointed out.

Among them the difficulty of dealing with fuzzy information has not found a satisfactory answer. It may be used in an intuitive way at some stage of the process such that its effect on the result is difficult to quantify. More satisfactory is to explicitly write it in a rule base. Various methods have been developed for quantifying the reliability of the results obtained but no definite consensus seems to have emerged.

Dealing with uncertainty has appeared to be essential in the last few years and a popular answer has been the Kalman filter. However

this approach constrains the linearization of the equations and thereby various approximations which limits the field of application of the system.

Combinatorial explosion may be partly solved by increasing the level of the interpretation which is still relatively low. Errors of measurement, low reliability, inconsistent connectivity between the image and the scene are all parts of the problem.



## CHAPTER 3

## OVERVIEW OF THE METHOD

## 3.1 PROBLEM DESCRIPTION

The aim of this work is to achieve a high level interpretation of the image by using general knowledge of the scene. The method developed is a first step to positioning a camera relative to a CAD representation of the scene, which is a CSG type representation in the case of PDMS, i.e. the application of this work.

The scenes studied are indoor scenes of power plants, typically composed of pipes (usually aligned with the walls), tables, rectangular and circular structures and cylindric tanks. Wall boundaries, door frames, rectangular structures and pipes may be represented by lines mostly perpendicular to three principal directions, the vertical one and two horizontal ones defined by the orientation of the walls. The presence of circular structures, such as joints of pipes or ends of cylindric tanks, correspond to elliptical arcs in the image.

The model provided by PDMS is a set of volumetric primitives such as parallelepipeds, cylinders or toruses. Each primitive is located relative to a unique coordinate system with no information about their relative positions and interactions. PDMS generates a very compact database and is very efficient for pipe runs. PDMS databases at EDF represent a huge amount of information about the power plants which might be used for vehicle guidance.

A first approach for solving this problem consists of deriving a vision-oriented model from the PDMS database and comparing it with the features in the image. A second consists of reconstructing volumetric



Method overall

primitives from the image and comparing it with PDMS primitives. The first approach is computationally very expensive, as numerous models have to be generated corresponding to all possible viewpoints. The complexity of the matching substantially decreases when the interpretation of the image improves, which makes the second approach attractive. Moreover it seems to be closer to how humans interpret photographs. But reconstruction of volumetric shapes from the image is limited because of the indetermination of the problem inherent to the 2D projection, and thereby is not very reliable.

A trade-off is proposed here. On one hand, high level features extracted from the image are represented in the 3D space, which can be called volumetric-oriented primitives. On the other hand, a vision-oriented database is extracted from the CAD database, but kept at a 3D level, thereby not increasing the level of complexity.

The aims of this work are :

- to extract high level features and estimate their reliability
- to demonstrate the feasibility of locating the camera with respect to a CAD database of the scene.

The first point is the main part of this thesis. The second point is demonstrated by constructing a model from the interpretations of several viewpoints which is then included in a CAD database. The matching process involved in the construction is beyond the scope of this work.

The method for extracting high level features consists of the following stages :

- Extraction of the edges of the image and approximation by straight line segments and ellipses.
- Interpretation of the perspective consisting of the detection of the vanishing points and the classification of the lines with the appropriate vanishing point.

- Construction of features represented in the 3D space, called 3D structures, consisting of straight line segments, rectangles, ellipses, vertices, edges, connected structures called local configurations.

Then 3D maps are built in two steps :

- Rotation of the coordinate systems corresponding to two viewpoints, in order to line up the main directions with the coordinate axes.
- Matching of structures (not described here) and determination of the relative scales and, thereby, of the relative positions of the viewpoints.

The scale of the representation remains unknown except if the scale of one matched structure is known, since the matched structures are consistently scaled. The scale indetermination may be broken by the *prior* knowledge of the average depth of the scene.

The next sub-section describes the main principles of the method.

### 3.2 PRINCIPLES OF THE METHOD

The scenes studied allow the following hypotheses to be made :

- geometric constraints : most objects in the scene are parallel to 2 or 3 main directions, e.g. in a room the directions defined by the walls.
- connectivity : connectivity in the scene is inferred from connectivity in the image, which can be done only if there is no occluded part. Therefore, the scene is hypotesized to contain few occluded parts.

The geometric constraints apply to most indoor scenes, such as offices or laboratories but it clearly excludes rooms or corridors with circular shape. The second hypothesis supposes that the scene is not

Method overall

excessively complex and has few unattached objects.

The formulation of the hypotheses is fuzzy in order to be a minimal constraint for the type of application concerned. This approach requires a method for evaluating the reliability of the result.

The main difficulties of a vision system are described in section 2.1.1, which considers the presence of noise in the image, the lack of consistency in the feature connectivity and the combinatorial explosion due to the number of possible aspects of a scene. The lack of consistency is resolved by the second hypothesis. The combinatorix may be decreased by extracting high level features, but this introduces the problem of reliability of these features. The presence of noise implies the need to deal with uncertainty.

The same methodology is used throughout the process. Let  $\mathcal{R}$  be a relationship between 2 features which is to be exploited, the features linked by  $\mathcal{R}$  are selected by a likelihood ratio test and then a higher level feature is determined using a Kalman filter. The part of this process common to several stages is detailed in the following sub-sections, whereas specific processes are detailed in the following chapters.

Sub-section 3.2.1 describes the elementary features selected for the process and the construction of high-level features from them. Two types of error are distinguished : uncertainty of measurement which is assumed to be normal with zero mean, and errors of segmentation (e.g. data selected although in fact they do not satisfy the relationship  $\mathcal{R}$ ), which do not have a zero mean error and thereby corrupts the Kalman filter. Sub-section 3.2.2 deals with the first type of error. It demonstrates how the necessity to deal with uncertainty constrains the choice of the tools and the representation of the variables and the relationships. The likelihood ratio test minimizes the risk of the segmentation error and is the subject of sub-section 3.2.3. The scoring process which provides a measure of the reliability of the results, is

deduced from the result of the likelihood ratio and is described in the same sub-section. Sub-section 3.2.4 gives details on the construction of the 3D maps. The last sub-section concludes with the principal points of the method developed.

### 3.2.1 From low to high level features

Edges are extracted from the image and represent the only information retained from the image. Edges are chosen because they are relatively accurate and seem preponderant in scene interpretation, e.g. the boundary of a closed door in a room is visible because of the shadow area formed, rather than because of the contrast between the door and the wall, which may be of the same colour. The Gaussian filter (Canny, 1986), Deriche's filter (1987) and Shen's filter (1986) are compared. An improved version of Shen's detector is proposed.

Edges are then approximated by straight line segments or ellipses. Berthod's algorithm (1987) has been tested for straight line segment fitting. Rosin and West's algorithm (1988) achieves straight line and circular arc fitting and provides a score with each fit, based on significance. Then, circular arcs are grouped for fitting ellipses (Rosin, et al, 1990). These two steps are shortly discussed (further work on this step is in progress (Ellis et al, 1991)). Let us remark that the quality of these steps plays an essential part in the quality of the final results.

Thus, the data of the process are straight line segments and elliptic arcs. Straight line segments are used for interpreting the perspective of the image. Once the vanishing points are found, elliptic arcs are interpreted as the projection of a circle and straight lines are grouped to form higher level features, called structures, which are represented in the 3D space.

*Interpretation of the perspective*

Parallel straight lines in the scene are projected onto concurrent lines in the image, which meet at a point called a vanishing point. The vanishing point coordinates provide the orientation of the associated lines (appendix 1). Lines lying in a plane in the scene have their associated vanishing points on a straight line in the image. Perpendicularity in the scene corresponds to a relationship between associated vanishing points in the image so that the vanishing points associated with a triplet of perpendicular lines form a triangle, the orthocentre of which is the projection of the optic centre onto the image, and the scale of which is related to the focal length. These well known properties of perspective are exploited for inferring 3D orientation of the features from their projection onto the image (see figure 3.2.1.1).

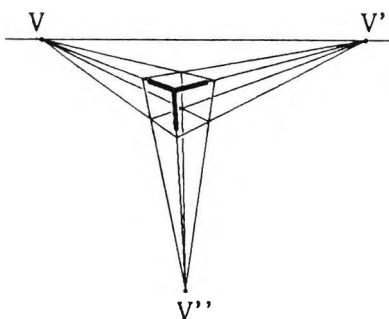


Figure 3.2.1.1: perspective view of a cube and associated vanishing points

The vanishing points are detected by accumulating the straight lines in an accumulator space in a way similar to Barnard (1983), but using a new accumulator space. Concurrent lines in the image are projected onto concurrent curves in this accumulator space which is bounded, isotropic and keeps constant the expected uncertainty of the vanishing point whatever its location. It is shown in chapter 5 that the definition of this accumulator space relies on a probabilistic model of the straight lines in the image and in particular of the expected value of the inverse of the square of their lengths. The model adopted and the

results of the statistics are described in chapter 4. The searched vanishing points correspond to the peaks of the accumulator space. The properties of the accumulator space ensure the same quality of the detection whatever the location of the vanishing points in the image plane.

Once the vanishing points are found, the straight lines are classified with them by using a likelihood test. Then, the perpendicular 3D directions are found by detecting consistent pairs of vanishing points. If possible, the elliptic arcs are classified with a pair of vanishing points, if not they are rejected. Hence, triplets of perpendicular directions are searched for.

At this stage, primitives are straight line segments and circular arcs oriented in the 3D space, i.e. only their distances from the camera is unknown. Actually, circular arcs have been very difficult to extract and have a very small place here; more work on this point is needed.

### *3D Local configurations*

In the following text the prefixes 2D and 3D refer to the dimension of the representation space and not to the dimension of the feature itself, e.g. a 3D straight line segment is a segment of the 3D scene, the exact location of which may be unknown, whereas a 2D straight line segment is its projection into the image.

Connectivity in the image is used for grouping the 3D primitives extracted at the previous stage (see figure 3.2.1.2) in the following way :

- Parallel 3D line segments, the corresponding 2D line segments of which are close in the image, are grouped to form a single line segment, called a linear structure.
- Two perpendicular linear structures, the corresponding 2D projections of which are close in the image, are grouped to form

an L structure.

- Parallel L structures, which share a linear structure and are in the same half-plane with respect to this linear structure, form a Comb structure.
- Two comb structures, which share a linear structure different from the spine, form a rectangular structure.
- Two perpendicular comb structures with the same spine form a 3D edge.
- Two perpendicular L structures, the projections of which have close vertices, form a 3D vertex.

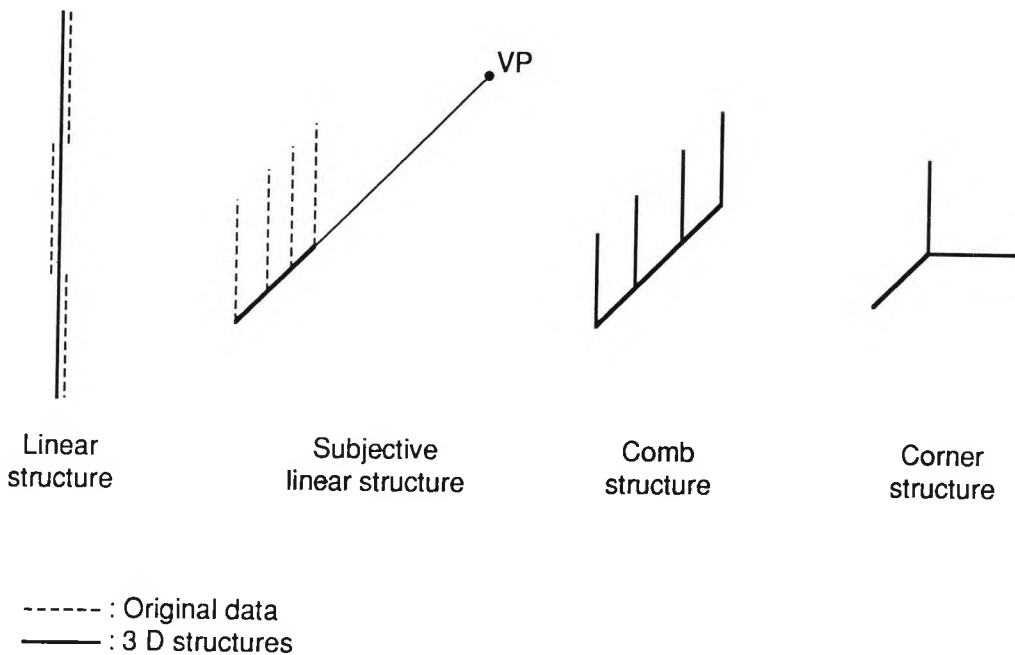


Figure 3.2.1.2: Examples of 3D structures

Most parameters describing a 3D structure are viewpoint invariant because of their representation in the 3D space. This is particularly attractive for a matching process. Most of these structures are scale invariant, e.g. edges and vertices. The size of the comb and L structures has poor significance as only a part of it may be visible, but the size of the rectangles is considered to have some significance. Not all the rectangles have been extracted, only the largest ones, e.g.

only the largest rectangle included within two comb structures facing each other is extracted. The concept of the comb structure is introduced to restrict the construction to a linear complexity, as their number is less than four times the number of linear structures (there are four half-planes perpendicular to a linear structure). Comb structures (an L structure is a particular case of a comb structure) are the basis of all further constructions, i.e. rectangles, edges and vertices.

Two different connectivity criteria are used to form the 3D structures. The first one uses the closeness of the corresponding features in the image, the other one uses the simple heuristic "have a common sub-structure". Both are subject to mistakes; the former because of noise, segmentation errors and hidden parts, the latter because of hidden parts (e.g. an edge may be confounded with its projection onto a wall - see figure 3.2.1.3). To carry on the construction by applying connectivity criterion until the process is stable, may therefore be dangerous. However, it may be used as an accumulation of evidence for the presence of a super-structure. For instance, a number of close parallel planar structures may provide evidence of the presence of a wall, even if actually the substructures are not coplanar because they correspond to, say, a cupboard, a desk and a picture lying on the wall. The iteration of the second connectivity criterion provides structures called 3D local configurations, which should correspond to coarse local maps of the scene.



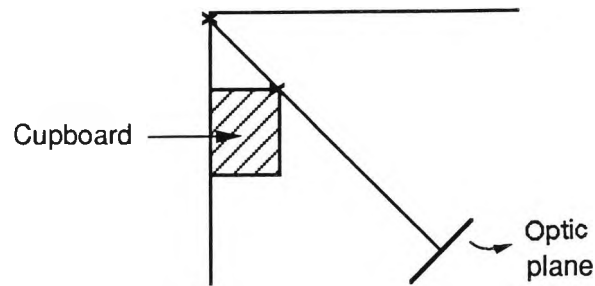


Figure 3.2.1.3 : Top view of a cupboard lying against a wall. The two vertical lines in the scene corresponding to the cross on the top view are confounded in the image.

At this stage the interpretation of the scene is not completed as the depths of the local configurations and the 3D structures are not known, i.e. the scale is unknown. However, it has improved very much as the scene is now symbolically described in terms of 3D rectangles, circular arcs, edges and vertices and more complex but less reliable structures, called local configurations.

These 3D structures may be compared with the 3D structures inferred by Lowe (1985) from the perceptual groupings extracted from the image. However, they are more precisely defined because of the interpretation of the perspective. It will be seen (sub-section 3.2.3) that the measure of the significance is given by a likelihood ratio based on a statistical model of the features in the image, using a unified approach. This approach is based on the existence of a non-negligible uncertainty of the measurements extracted from the image.

### 3.2.2 Dealing with uncertainty

The interpretation of the perspective and the construction method of the 3D structures may be expressed as a set of relationships linking the parameters of the features, i.e. the data, and the unknown variables. As the data are uncertain, the relationships are true within a range of uncertainty. Taking into account uncertainty enables the

best estimate of the solution to be found with its associated uncertainty.

Let us remark that dealing with uncertainty is an absolute necessity for mobile vehicle because data from various sources (e.g. CAD model, multiple images, range-finder, odometer), may be required for achieving even a simple task, and they must be made consistent.

As mentioned in section 2.1.5, Kalman filtering is a very popular tool when dealing with normal random variables. If the relationship between data and unknown variables may be written as a set of linear equations (possibly after linearization) and the noise associated with the data has a zero mean, a Kalman filter (or extended Kalman filter) gives the best estimate, in the LMS sense, of the variable studied and its uncertainty. In addition, the algorithm is very efficient in time and storage because of its iterative implementation. However it requires linear (or linearized in case of an extended Kalman filter) and independent relationships (if 2 relationships are dependent, divergence occurs).

Kalman filtering (appendix 2) is used throughout the method developed in the following pages, whenever using the appropriate representation in the sense defined above (Ayache and Faugeras, 1987). However the tests selecting the input data use a different representation, enabling the exploitation of geometric properties (Figure 3.2.2.1). For example, the detection of vanishing points is achieved through an isotropic but non linear representation, whereas the Kalman filter uses the Cartesian representation, linear but not isotropic ; the use of the Cartesian representation for vanishing point detection would prevent the isotropy of the detection.

The use of a Kalman filter supposes a white noise and therefore a correct segmentation, which does not take into account possible errors at previous stages or simply unreliability of the hypotheses used. This implies that wrong data are used as the input to an iteration of the

Method overall

Kalman filter. The effect is very different from noisy data as no hypothesis can be made upon the statistical distribution of the error, so that the hypotheses required for using the Kalman filter (Kalman, 1960) are not fulfilled. Unfortunately, the larger the error, the greater the influence on the result ; moreover if the uncertainty is low, it substantially reduces the uncertainty of the result, further aggravating the problem. Therefore, a distinction between segmentation error and measurement error should be made so that only reliable data (in the segmentation sense) are used in a Kalman filter. In the method described here, data are segmented by using a likelihood ratio test (section 3.2.3, appendix 3).

Another way to ensure stability of the Kalman filter is to provide a good initial estimate of the unknown variable to the Kalman filter. More flexibility is possible for this first estimation as uncertainty may be over-estimated. Accumulation of data in an appropriate accumulator space is an efficient way of doing this (Weiss, 1988). This approach is used for the detection of the vanishing points. However, as the prior estimate is found by using the same data as the Kalman filter, it is no longer independent of the data. This difficulty is overcome by over-estimating the uncertainty associated with the prior estimate, but not too much, in order to give this estimate some weight in the Kalman filter. It may cause an under-estimation of the uncertainty of the final result but this is in no way comparable to the one resulting from the use of wrong data in the Kalman filter, and it increases the reliability of the results.

The figure 3.2.2.1 demonstrates how the likelihood test and Kalman filter cooperate in the process and how the data representation is related to the process in which the data are involved. In this figure the following notations have been used :

R : result of the likelihood ratio test

(a,b) : a is the slope of a 2D line ( or the inverse of the slope

if the slope is higher than 1), and  $b$  the intercept with the axis  $Oy$  (or  $Ox$  if the slope higher than 1)

$(\vec{u}, G)$  : unit vector and centroid of a 3D segment

$(x, y)$  : Cartesian coordinates in the image

$(\rho, \theta)$  : polar representation of a point in the image

$(X, Y, Z)$  : Cartesian coordinates in the scene

$D(F_i, F_j)$  : Distance between the features  $F_i$  and  $F_j$  (in the image).  $D$  is the Euclidean distance in case of vanishing point detection, and is the longitudinal and transversal distances between two straight line segments (defined in chapter 6) in case of the construction of linear structures and L structures

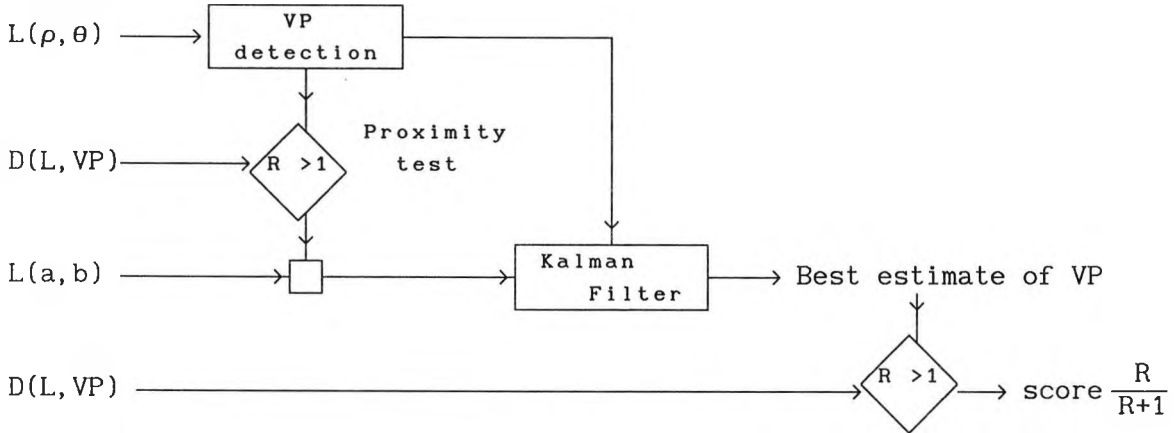
$L(a, b)$  : Straight line segment in the image with endpoints  $a, b$ .

$L(\vec{u}, G, \ell)$  : Straight line segment in the space with orientation  $\vec{u}$ , centroid  $G$  and length  $\ell$ .

Method overall

Vanishing point detection:

Line segments



Pair and Triplet of perpendicular directions:

Vanishing points

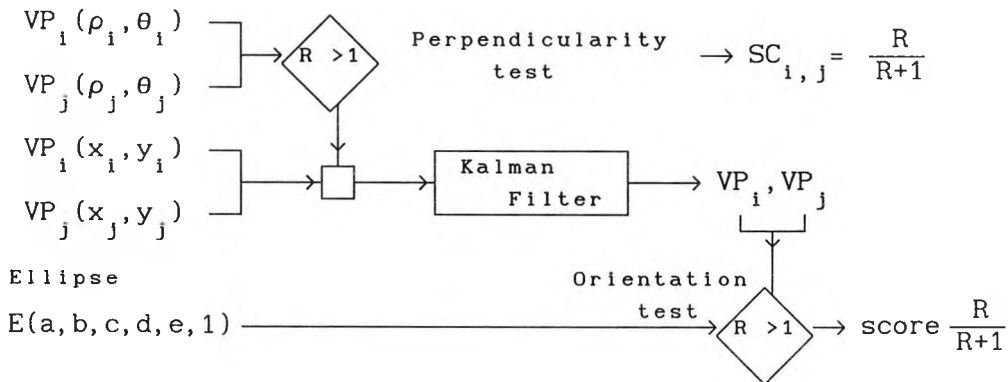
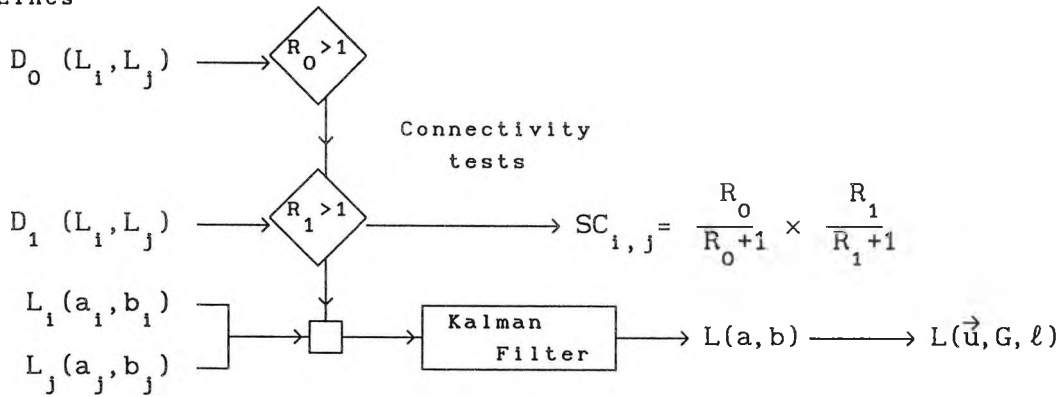


Figure 3.2.2.1.a: Likelihood test, Kalman filter and data representation in the construction of the 3D structures.

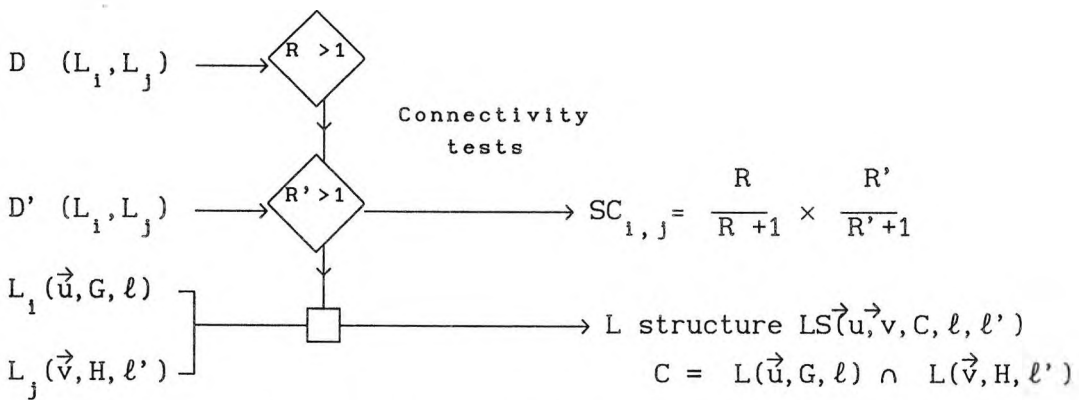
Linear structures:

$L_i, L_j$  associated with the vanishing point  $VP(x,y)$ :

Lines



L structures:



Comb, rectangular and edge structures =  $\cup$  parallel L structures

Corners structures:

From 2 perpendicular L structures  $LS_1(\vec{u}_1, \vec{v}_1, C_1, \ell, \ell_1)$  and  $LS_2(\vec{u}_2, \vec{v}_2, C_2, \ell, \ell_2)$

Vertices

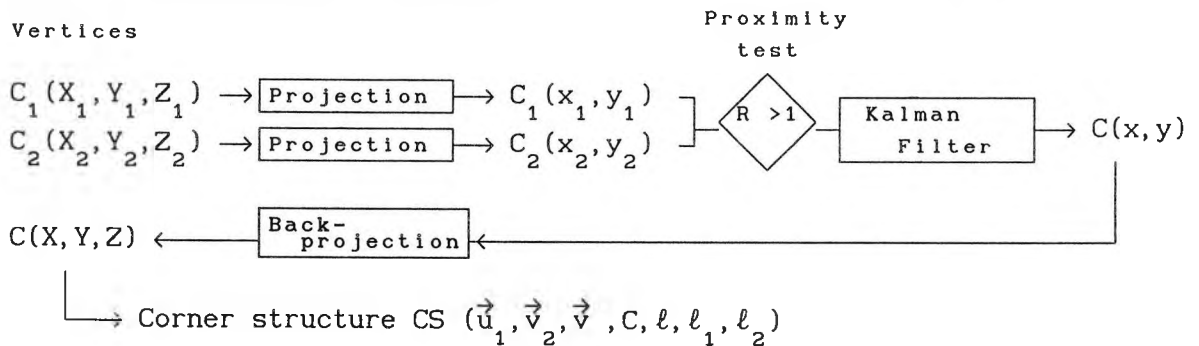


Figure 3.2.2.1: Likelihood test, Kalman filter and data representation in the construction of the 3D structures.

### 3.3 Dealing with segmentation noise : Likelihood test and scoring process

#### *Likelihood ratio test*

The perspective projection of the features in the scene onto the image generates an indetermination relative to the 3D location of these features. However, it is possible to use cues from the image for hypothesizing 3D configurations. For instance, connectivity in the image cannot ensure connectivity in the scene but strongly suggests it. Such particular relationships, e.g. connectivity or parallelism, are very important for interpreting the image. The aim of the likelihood ratio test is to decide whether a 2D relationship between two features in the image is more likely to be due to chance or to be due to a particular relationship between the corresponding 3D features.

As has been explained, the relationships between two features are not perfect because of uncertainty of measurement. But they can also occur by accident, e.g. two lines may be parallel in the image without being parallel in the scene. Hence the segmentation problem : how to determine which data satisfy the relationship of interest. Let us remark that the sources of errors are linked, since the worse the uncertainty of measurement, the higher the risk of erroneous segmentation.

Let us suppose that the relationship of interest,  $\mathcal{R}$ , may be written in the following way : let  $F_1$  and  $F_2$  be two features of the image linked by this relationship, then

$$(\mathcal{R} \text{ true}) \Rightarrow V(F_1, F_2) = 0$$

Because of error of measurement, it should be rewritten

$$(\mathcal{R} \text{ true}) \Rightarrow V(F_1, F_2, e_1, e_2) = V \approx 0$$

where  $e_1$  and  $e_2$  are the errors of measurement with covariance matrices  $C_1$  and  $C_2$ . As the errors are supposed to be normal,  $V$  is a random

variable with a Gaussian distribution with zero mean and covariance  $C_v$ .

Now let us suppose that the features  $F_1$  and  $F_2$  are not linked by the relationship  $\mathcal{R}$ , then  $V(F_1, F_2, e_1, e_2)$  is still a Gaussian random variable no longer centered at 0 but at  $S(F_1, F_2)$ , then

$$V = V(F_1, F_2, e_1, e_2) \approx S(F_1, F_2) \quad (3.3.1)$$

The features  $F_1$  and  $F_2$  are supposed linked by the relationship  $\mathcal{R}$  if  $V$  is not too large.  $V$  is therefore called the decision variable. For example, for classifying a line  $L$  with a vanishing point  $P$  the relationship  $\mathcal{R}$  is : "the point  $P$  is the vanishing point of the line  $L$ ". If  $\mathcal{R}$  is true, the distance  $D$  between the line  $L$  and the point  $P$  should be equal to 0. Because of errors of measurement it is only possible to say that  $D$  is small. But  $D$  small does not mean that  $\mathcal{R}$  is true as the line  $L$  may pass near  $P$  by accident. The problem is to find a criterion applied to  $D$  which minimizes the risk of mis-classification.

Various tests may be used to test the value of  $V$  and decide whether  $\mathcal{R}$  is fulfilled. A coarse test consists of setting a threshold for  $V$  and selecting any pair of features such that  $|V| < V_{\max}$ , i.e. a neighbourhood criterion. Another criterion takes into account the uncertainty with which  $V$  is known by using its Mahalanobis distance, i.e.  $V^t C_v^{-1} V$ , which has to be less than a fixed threshold; this test is called the MD test in what follows. This criterion does not take explicitly into account the risk of error due to accidental situations, i.e. segmentation error. Since the importance of such an error has been mentioned in the previous sub-section, it has been preferred to use the likelihood ratio test (LR test) which does take it explicitly into account.

The likelihood test consists of selecting the more probable hypothesis for the interpretation. Let  $V$  be the result of an experiment testing from which of the hypotheses  $H$  or  $\bar{H}$  is true, i.e. whether two features  $F_1$  and  $F_2$  are linked by the relationship  $\mathcal{R}$  ( $H$ ) or not linked



Method overall

by  $\mathcal{R}(\bar{H})$ . For instance, in the case of the classification of a line  $L$  with a vanishing point  $P$ ,  $V$  is the minimal distance from the line to the point and the hypotheses to test are  $H$  : "the line ( $L$ ) passes through  $P$  'on purpose'" and  $\bar{H}$  : "the point  $P$  and the line  $L$  are not related", i.e. the point  $P$  is close to the line  $L$  by accident. The likelihood ratio tests which hypothesis is more likely, knowing  $V$ . Using the Bayes theorem, it is found that  $H$  is more likely than  $\bar{H}$  knowing  $V = \mathbf{V}$  if :

$$R = \frac{p(H|V=\mathbf{V})}{p(\bar{H}|V=\mathbf{V})} = \frac{p(V=\mathbf{V}|H)}{p(V=\mathbf{V}|\bar{H})} \frac{p(H)}{p(\bar{H})} > 1 \quad (3.3.2)$$

$V|H$  and  $V|\bar{H}$  are two random variables, the former corresponds to the hypothesis  $H$  and the latter to the hypothesis  $\bar{H}$ . If classifying a line with  $H$  by mistake has more serious consequences than classifying a line with  $\bar{H}$  by mistake, this is taken into account by modifying the test (3.3.2) to  $R > R_{\min} > 1$ .

The difficulty with the Bayesian approach is to estimate the distribution associated with  $V|\bar{H}$  and the prior probabilities  $p(H)$  and  $p(\bar{H})$ .  $V|H$  is the random variable corresponding to the error of measurement, the distribution of which is supposed known.  $V|\bar{H}$  depends on the statistical distribution of the features of interest. The ratio  $p(H)/p(\bar{H})$  is often intuitively estimated by the user (e.g. Rosin, 1988; Dickson, 1990) which is dangerous as probability is not a very intuitive concept. In order not to jeopardize the appeal of the likelihood test, the definitions of the random variables  $V|H$ ,  $V|\bar{H}$  and the values of  $p(H)$  and  $p(\bar{H})$  must be precisely defined.

The scene and the conditions of image acquisition are supposed perfectly known so that the exact parameters of the features  $F_i$  and their relationships may also be supposed known. Let  $\Omega$  be the set of pairs of features  $\mathcal{F}_k = (F_i, F_j)$ ,  $\Omega_1$  the set of pairs of features linked by the relationship  $\mathcal{R}$ , and  $\Omega_2$  the set of pairs of independent features.

Then,  $\mathcal{V}|H$ ,  $\mathcal{V}|\bar{H}$ ,  $p(H)$  and  $p(\bar{H})$  are defined by

$$p(H) = \frac{\text{Card}(\Omega_1)}{\text{Card}(\Omega)} = \frac{n_1}{n} \quad \text{and} \quad p(\bar{H}) = \frac{\text{Card}(\Omega_2)}{\text{Card}(\Omega)} = \frac{n_2}{n}$$

$$\mathcal{V}_1 = \mathcal{V}|H = \mathcal{V}|\Omega_1 \quad \text{and} \quad \mathcal{V}_2 = \mathcal{V}|\bar{H} = \mathcal{V}|\Omega_2$$

$$p(\mathcal{V}_1=V) = p_1(V) \quad \text{and} \quad p(\mathcal{V}_2=V) = p_2(V)$$

$p_1$  is a sum of Gaussian functions centred at 0 and scaled by a normalization factor, whereas  $p_2$  is a sum of Gaussian functions centred at points  $S_k = \mathcal{V}(F_i, F_j, 0, 0)$ , where  $(F_i, F_j) = \mathcal{F}_k \in \Omega_2$ . The density  $p_2$  may be written

$$p_2(V) = \sum q_k G_{\sigma_k}(V-S_k)$$

where  $q_k$  is the prior probability associated with  $\mathcal{F}_k \in \Omega_2$  and  $G_{\sigma_k}$  is the Gaussian law associated with the measurement error. In fact, since the scene is unknown, so are  $\Omega_1$  and  $\Omega_2$ . If a statistical model of the distributions of the features  $F_i$  and  $F_j$  is available and defined by the density  $g(P)$ , where  $P$  is the set of parameters corresponding to the pair of features considered, then  $p_2(V)$  may be estimated by replacing  $q_k$  by  $g(P)d(P)$ . Let  $q(S)$  be the density law associated with  $\mathcal{V}|\bar{H} = S$  in the absence of measurement error, then  $\int_{S=\text{cst}} g(P)dP = q(S)dS$  and  $p_2$  becomes

$$p_2(V) = \int G_{\sigma_s}(V-S) q(S)dS .$$

As  $q$  is smooth, it is possible to approximate  $p_2$  by  $q$  (i.e. the convolution of  $q$  by a Gaussian does not change  $q$ ), which means that the effect of the error of measurement is negligible compared with the inaccuracy of the *prior* model.

The ratio  $R$  may be rewritten

$$R = \frac{n_1 p_1(V)}{n_2 p_2(V)}$$

The estimation of the ratio  $n_1/n_2 = p(H)/p(\bar{H})$  may be achieved as follows. Let  $\mathcal{N}$  be a subset of  $\Omega$  such that the probability of " $(F_i, F_j)$  linked by the relation  $\mathcal{R}$  is not in  $\mathcal{N}$ " is small.  $\mathcal{N}$  is the subset of interest.  $\mathcal{N}_1$  is the subset of  $\mathcal{N}$  of features linked by  $\mathcal{R}$  and  $\mathcal{N}_2$  its complement in  $\mathcal{N}$ . Let  $nn$ ,  $nn_1$  and  $nn_2$  be the expected values of the cardinals of  $\mathcal{N}$ ,  $\mathcal{N}_1$  and  $\mathcal{N}_2$ . The probability for a pair of features  $\mathcal{F}_k \in \Omega_i$ , with  $i \in \{1, 2\}$ , to be in  $\mathcal{N}_1$  is  $\int_{\mathcal{N}} dp_1$ . Since  $nn_1 = \sum_{\Omega_1} p(\mathcal{F}_k \in \mathcal{N}_1)$  :

$$nn_1 = n_1 \int_{\mathcal{N}} dp_1 \quad \text{and} \quad nn_2 = n_2 \int_{\mathcal{N}} dp_2$$

As  $\Omega$  and  $\mathcal{N}$  are known,  $n$  and  $nn$  are also known. From  $n_1 + n_2 = n$  and  $nn_1 + nn_2 = nn$ , the ratio  $n_1/n_2$  is deduced :

$$\frac{n_1}{n_2} = \frac{nn - n \int_{\mathcal{N}} dp_2}{n \int_{\mathcal{N}} dp_1 - nn} \quad (3.3.3)$$

Note that theoretically,  $n_1/n_2$  does not depend on the choice of  $\mathcal{N}$ . Practically if  $\mathcal{N} = \Omega$ , this ratio is undetermined. If  $\mathcal{N}$  contains very few features of interest, i.e.  $\int_{\mathcal{N}} dp_1 \approx 0$ , then the numerator and denominator are negative, measuring a lack of features in  $\mathcal{N}$ , a process which is highly unreliable because of the existence of other relationships among the features. If  $\mathcal{N}$  is a minimal subset containing most of the features of interest, i.e.  $\int_{\mathcal{N}} dp_2 \approx 0$  and  $\int_{\mathcal{N}} dp_1 \approx 1$ , the numerator evaluates the amplitude of the local event  $\mathcal{N}_1$ , while the denominator evaluates the remaining noise. This process is much more reliable as it measures an event where it is most likely to occur.

The same definitions may be applied to any subset of  $\Omega$ , say  $\Lambda$ . For example, it is possible to define  $\Lambda$  by fixing one of the features, e.g. the vanishing point in the case of the line classification. The

reliability of  $R$  measurement depends on the choice of the subset  $\Lambda$ . The smaller the initial subset, the more efficient the test, as it takes into account the specificities of the subset. The choice of  $\Lambda$  depends on the relationship studied ( $\mathcal{R}$ ) and on the model used.

This approach contrasts with Lowe's approach (1985). Here, the probability of occurrence of  $\mathcal{R}$  in  $\Omega$  (i.e. in the image) is deduced from the comparison between the appearance of  $\mathcal{R}$  in the image and its accidental appearance in  $\Omega_2$  (i.e. in the prior statistical model of the image), whilst Lowe defines it statistically as the frequency of occurrence of the relationship  $\mathcal{R}$  among a set of typical images. Our approach has the advantage not needing to refer to any specific set of typical images.

A model of the distribution of the straight line segments into the image has been used for determining  $\Lambda$ ,  $p_1$ ,  $p_2$  and  $n_1/n_2$  in the line classification stage (see chapter 5).

*Adaptation of the LR test to the complexity of the image*

The distribution of  $V_2$ , i.e.  $p_2$ , may often be approximated by a uniform distribution around zero, the value of which is represented by  $\tau$ . The likelihood ratio test is then

$$p(V|H) > \tau \frac{p(\bar{H})}{p(H)} = \lambda \quad (3.3.4)$$

where  $\lambda$  is related to the risk of rejecting  $H$  when  $H$  is true, called a type I error.

If  $\Lambda_2$  is very small compared with  $\Lambda_1$ ,  $p(\bar{H})/p(H)$  is very small and the test is not selective, conversely if  $\Lambda_1$  is small compared with  $\Lambda_2$ , then the test is very selective. This means that if the segmentation noise is low, i.e. the relationship  $\mathcal{R}$  is generally fulfilled (e.g. most of the straight lines in the image are parallel in the 3D world), the errors are mainly errors of measurement (which can be dealt with by the

Method overall

Kalman filter) and the test may be tolerant, i.e.  $\lambda$  is small. Conversely if the noise is high (e.g. the lines searched for are embedded in numerous other lines) it is necessary to be very cautious about the line selection, i.e.  $\lambda$  is large.

It is shown in appendix 3 that the LR test is more indulgent than the MD test when the uncertainty of the feature is low and conversely when it is high. The LR test appears to be an intermediate between the neighbourhood test and the MD test. The LR test takes into account uncertainty of measurement but also the risks of segmentation error due to accidental configurations, i.e. the noise, by contrast with the MD test. However it requires the modeling of the distribution of  $V|\bar{H}$  and the estimation of  $p(H)/p(\bar{H})$  (which is not always straightforward!).

#### *Scoring process*

Two features  $F_1$  and  $F_2$  are supposed to be linked by the relationship  $R$  if

$$R > 1.$$

The higher the value of  $R$ , the higher the confidence of the decision. The score associated with this decision is

$$p(H|V=V) = \frac{R}{R+1}.$$

Once all pairs of features have been classified with  $H$  and  $\bar{H}$ , updating  $p(H)$  is possible, e.g. for scoring a class of lines associated with a vanishing point. From the model described previously, it is known that the set

$$\mathcal{V}(\mathcal{N}) = \{V_k(F_i, F_j) = V_k ; (F_i, F_j) \in \mathcal{N}\}.$$

has the density  $p(V) = p(H)p_1(V) + p(\bar{H})p_2(V)$ .  $p(H)$  and  $p(\bar{H})$  has been estimated previously, using a global assumption on the repartition of the noise. Now it is possible to refine these values by taking into account the actual distribution of  $\mathcal{V}(\mathcal{N})$ , where  $\mathcal{N}$  is the neighbourhood

previously described. Let  $\{\alpha\}$  be the set of possible values for  $\text{Card}(N_1)/\text{Card}(N)$ . Then,

$$p(\alpha) = p(nn_2=(1-\alpha)nn),$$

where  $nn$  is the total number of directions in  $N$  and  $nn_2$  is the number of these directions corresponding to noise. Using Bayes' theorem the density of probability  $p$  can be written

$$p(\alpha|V_1, \dots, V_n) = \frac{p(V_1, \dots, V_n|\alpha)p(\alpha)}{p(V_1, \dots, V_n)}. \quad (3.3.5)$$

Using the decomposition of  $p(V_i|\alpha)$  over  $p(H)$ , i.e.  $\alpha$ , and  $p(\bar{H})$ , i.e.  $1-\alpha$ , then

$$p(V_i|\alpha) = p(V_i|\bar{H}) ((O_i-1)\alpha + 1)$$

where  $O_i = p(V_i|H)/p(V_i|\bar{H})$  is the odds of  $V_i$ . The events  $V_i$  being independent,  $p(V_1, \dots, V_n|\alpha)$  is the product of the  $p(V_i|\alpha)$ . The maximization of (3.3.5) is equivalent to the maximization of

$$G(\alpha) = \prod_i ((O_i-1)\alpha + 1)p(\alpha) \quad (3.3.6)$$

The probability  $p(\alpha)$  is maximum for  $\alpha_0 = p(H)$ , the prior probability of  $H$ , and decreases as  $\alpha$  goes away from  $\alpha_0$ . Thus, the maximum of  $G$ ,  $\alpha_m$ , is reached around  $\alpha_0$  and may be found by a simple process such as the parabolic approximation (Press, 1988).

Returning to the definition of  $p(H)$ ,  $p(H|V_1=V_1, \dots, V_n=V_n) = \alpha_m$  represents the expected percentage of pairs of features linked by the relationship  $\mathcal{R}$ , knowing  $V_1, \dots, V_n$ . In the case of the line classification the score of a class represents the expected percentage of lines in the image which are effectively parallel in the scene to the direction associated with the class.

Method overall

### *Conclusion*

A relationship between two features is hypothesized if the decision variable associated with the relationship succeeds the LR test. The quality of the test depends on the accuracy of the modelling of the feature distribution. It has been compared to the neighbourhood test and to the test based on the Mahalanobis distance. Its advantage over the former is to take into account explicitly the uncertainty of the measurement. Its advantage over both is also to take into account the risk of mis-classification. Moreover, the approach enables a natural scoring of the process.

#### **3.2.4 Application to map construction**

Up to now the parameters of the 3D structures have been expressed relative to the camera coordinate system. The best pair or triplet of perpendicular directions is now chosen to be the new coordinate system, the origin of the system being unchanged. The transformation for passing from the old coordinate system to the new one is a rotation around the origin.

Let  $VP_1$  and  $VP_2$  be two viewpoints of the scene ; upside down movement of the camera is discarded so that it is possible to have a point to point correspondence between the vanishing points of both viewpoints. The new 3D coordinate systems associated with  $VP_1$  and  $VP_2$  are therefore parallel.

Matching of a similar rectangular structure in both viewpoints, e.g. identical orientation, same ratio width/length and same range of size, allows the determination of the relative scale of those structures and thereby of the translation of the camera between the viewpoints. All the structures may now be represented in a common coordinate system and matching be propagated to other structures.

The result of this process is a number of structures represented in

a unique coordinate system parallel to the principal directions and with origin, say the optic centre of the first position of the camera. Although the general scale of the map is not known, the elements of the map are now consistently scaled. The scale of the representation can be bounded if the focal distance and the depth of field is known. Scale indetermination is broken once the scale or the depth of one rectangular structure is known.

The model of the scene obtained may be identified to a CAD model. To illustrate this point, it is converted into an object of ROBCAD (ROBCAD is the CAD software used by EDF for robotics simulation). The feasibility of matching such a model with PDMS database is discussed.

The 3D maps constructed from a range of viewpoints using monocular vision give a description of the scene in terms of high level symbolic primitives such as rectangles or vertices, but it is far from being complete, its reliability may be low and the scale remains unknown. However, it illustrates the level of interpretation achieved by the construction of 3D structures and local configurations and demonstrates the possibility of matching with a CAD database which takes advantage of the fact that only principal structures have been extracted.

The 3D representation of the 3D structures eases the change of coordinate systems and matching process, as it is a simple linear transformation, and the intrinsic parameters of the structures are directly comparable, e.g. orientation or ratio length/width.

### 3.2.5 Conclusion

The method briefly described in the previous sections is detailed in the following chapters.

High level structures represented in the 3D space have been constructed by first interpreting the perspective of the image, then by



## Method overall

grouping the connected features. Relationships such as parallelism, perpendicularity and connectivity, are tested by a likelihood ratio test, to be used in a Kalman filter in order to determine higher level features. The Kalman filter maintains the uncertainty of the new feature while the likelihood ratios provide a score reflecting its reliability.

One of the main contributions of this work is concerned with the statistical approach which is used throughout the process, avoiding as much as possible arbitrary choices. This approach has led to the definition of a new accumulator space for the detection of the vanishing points and to a consistent methodology for testing and scoring a relationship between two features.

The high level of the interpretation reached is demonstrated by the construction of 3D maps using a number of viewpoints, without reference to their relative positions.

## CHAPTER 4

## PREPROCESSING

## 4.1 OVERVIEW OF THE FEATURE EXTRACTION

The objects in the scene are represented by variations of grey level in the image. However, a uniform part of the object generally does not correspond to a uniform area in the image because of variations of lighting in the scene. However the boundaries of an object are very likely to be represented by sudden variations of grey level because of change of colour between the object and the background or because of the shadows due to the geometry of the boundaries. Therefore, the edges appear to be very useful features for scene understanding. Various edge detectors are described in section 4.2. They are first studied from a theoretical point of view and then are tested on simulated and real scene images. Then, an improved version of the Shen detector (1986) is proposed.

Edges are still low level information. The shapes of man-made objects are more efficiently described in term of straight line segments or circular arcs, which lead to higher symbolic descriptors, such as parallelepipeds or cylinders. Such shapes are very numerous in the scene studied. Thus, the edges are first approximated by straight line segments, then by elliptical arcs assumed to be the projection of circular arcs in the scene (section 4.3).

Unfortunately edge detection is sensitive to noise. As a result real edges are noisy and many false edges (i.e. corresponding to noise) are detected. The polygonal approximation should be insensitive to the noisy aspect of straight line edges, which is achieved by choosing appropriate parameters. However this results in uncertainty of the endpoints. The response of the edge detector and line finder is a main issue for the evaluation of the reliability of the result. The response to noise of various edge detectors and the Berthod line finder is

studied in sections 4.2 and 4.3. Then the uncertainty of the endpoints is modelled in subsection 4.4.1.

Another source of error, completely different in nature, is due to the superimposition of the 2D projection of the segments of the scene. The effect of this superimposition is unpredictable and is a major difficulty for interpreting the scene. Segments may be adjacent in the image without corresponding to adjacent segments in the scene. This effect is considered as a type of noise, called the segmentation noise. It may be taken into account in the interpretation process by considering a prior distribution of the straight line segments in the image. This distribution is meant to describe the accidental occurrence of various relationships in the image, due to the loss of one dimension. A particular prior statistical model of the feature parameters is described in subsection 4.4.2. This model will be used as a reference in the following chapters.

The interpretation of angles in the image, particularly the angles assumed to be the projection of right angles in the scene, requires the knowledge of the intrinsic calibration parameters. The determination of these parameters is the subject of section 4.5.

## 4.2 EDGE DETECTION

The boundaries of the objects in the scene are represented by discontinuities in intensity in the image. The edge detection should detect these discontinuities, which are localized at maxima of the gradient of the image. Differentiation emphasizes noise, and numerous false maxima may appear in the image gradient. Low-pass filtering is required to decrease their number.

Two main approaches are classically opposed in edge detection : either maxima of the modulus of the gradient are searched for in the direction of the gradient, or the zero-crossings of the Laplacian are detected. The zero-crossings of the Laplacian include not only maxima of the gradient but also saddle points and plateaux of the gradient; however, they allow sub-pixel accuracy. Many researchers (Canny, 1986 ;

Deriche, 1987 ; Shen, 1986 ; de Micheli et al, 1989) have preferred the method of the maxima of the gradient because there are less extraneous edges. This approach has been chosen here as it was not intended to obtain sub-pixel located edges, at least in the first place.

Therefore, the problem is to find the best antisymmetric filter for detecting the edges. First the shape, then the parameter of the filter have to be found. According to de Micheli et al (1989), they are not essential to the goodness of the result for indoor scenes, because of the good signal to noise ratio of such images, even with an average camera. The images studied here include many shadow areas, where the signal to noise ratio is very low, so that the search for the best edge detector is critical to the quality of the interpretation. It will be shown that it is indeed difficult to conclude on the optimal shape of an edge detector on images from real scenes, but that it is possible to appreciate the differences between the shapes of various filters, and still more between various parameters of the same filter, by using a theoretical approach illustrated by the response of the filters on synthesized images. Three different shapes are compared in the following, the Gaussian shape, Deriche's filter (1987) and the exponential shape (Shen, 1986). Then, an improved version of the Shen filter is proposed, having the additional important property of isotropy. For comparing these edge detectors, Canny's schema is first used, then extended by using three additional criteria describing the round-up effect, the sensitivity to thresholding and the sensitivity to multiple edges.

Canny (1986) proposed a convolution by an antisymmetric function followed by the detection of the maxima, called non-maxima suppression. Since  $u * g' = (u * g)'$ , the convolution by an antisymmetric filter  $f$  is equivalent to applying a smoothing filter  $g$  such that  $g' = f$ , and computing the gradient. For choosing the appropriate filter the response of the antisymmetric filters with a finite support to a noisy step is studied. A real edge and a simulated noisy step are displayed in figure 4.2.1. The function optimizing the following product is looked for

$$P = \sum \Gamma,$$

where  $\Sigma$  is the signal to noise ratio and  $\Gamma$  is the localization of the edge. In one dimension the larger the filter, the better the signal to noise ratio  $\Sigma$  and the worse the localization. The optimization of  $\Sigma\Gamma$  leads to a function  $f$  defined by a difference of boxes defined by Herskovitz and Binford (1980). The definition of  $\Sigma$  depends only on the signal and the noise occurring at the very location of the perfect step and not of the behaviour of the signal around this point. Actually, the set of functions defined by a difference of boxes produces a multiple response responsible for another type of noise, the parasite edge, which is not taken into account by  $\Sigma\Gamma$  but which is taken into account by the uniqueness criterion, defined as the inverse of the density of maxima  $\mu_0$  of a Gaussian noise filtered by  $f$ . Canny defines the optimal filter with a finite support  $W$ , as the function optimizing  $P$  under the constraint  $kW = 1/\mu_0$ .

The optimal filter proposed by Canny has the form

$$f(x) = (a_1 \sin(ax) + a_2 \cos(ax)) \exp(\omega x) + (a_3 \sin(ax) + a_4 \cos(ax)) \exp(-\omega x) + c.$$

In order to qualify the uniqueness of the filter, Canny introduces the ratio  $r$ , defined as

$$r = \frac{|f'(0)|}{\Sigma \sqrt{\int_S f''^2(x) dx}} \quad (4.2.2)$$

where  $S$  is the support of  $f$ . The ratio  $r$  is meant to relate the rate of false maxima at the step location and the rate of false maxima far from this step. This ratio is equal to 1 when the two events are equally likely. Canny tried to choose  $r$  as close as possible to 1. In fact, for the optimal filter  $r = 0.57$  (Canny, 1986) (here, the probability of detecting false maxima is higher at the step location than outside).

Shen (1986), then Deriche (1987) have extended Canny's approach to infinite response filters, by using a recursive implementation of the filter. They conclude on filters of different shapes. In the following, the Canny, Deriche and Shen filters are first compared with respect to

3 criteria,  $\Sigma$ ,  $\sigma_t = 1/\Gamma$  and  $\mu$ , where  $\mu$  is the density of false maxima at the step location.  $\mu$  instead of  $\nu$  is used for the uniqueness criterion, because of its clear geometric interpretation. Only 2D case is studied. In 2D case,  $\mu_0$  is given in appendix 7 :

$$\mu_0 = \frac{1}{2\pi} \sqrt{\frac{R^{(4)}(0)}{R''(0)}}, \quad (4.2.1)$$

where  $R(\xi)$  is the correlation function of the convolution product of the noise by the filter, and  $\mu$  is given by eq.A7.13 :

$$\mu = \mu_0 (\sqrt{2\pi} |\zeta| (\operatorname{erf}(\zeta) - 0.5) + \exp(-\frac{\zeta^2}{2})) \quad (4.2.3)$$

where  $\zeta = \nu \Sigma$ . Therefore, the higher  $\nu$ , the better the result. There is no reason for limiting the value of  $\nu$  to 1. Then, it is shown that these criteria are insufficient to characterize an edge detector, and additional criteria will be used : the round-up effect and the sensitivity to multiple edges and to thresholding.

Only the 2D case is considered in the comparison. First, let us shortly explain why the 2D case is really different from the 1D case. For simplicity, the filter is supposed with a separable kernel (defined latter). The image is first filtered in one direction by a symmetric filter (i.e. smoothing filter), then filtered in the orthogonal direction by an antisymmetric filter, which provides with the value of the gradient in this latter direction. This opportunity of filtering along the step, allows the noise but not the step to be smoothed. Actually the central limit theorem shows that if the support of this smoothing filter, assumed positive and symmetric, is included in the interval  $[-n, n]$ , then when  $n$  tends towards infinity,  $\Sigma$  tends towards infinity and  $\mu$  tends towards  $\mu_0$  which tends towards 0. As it will be shown later on, the effect of this filter compensates the effect of the antisymmetric filter on the error of location  $\sigma_t$ . These remarks suggest that, using a single step model, the larger the support of the filter, the better the results for  $\Sigma$ ,  $\mu$  and two effects on  $\sigma_t$  which compensate each other (in fact,  $\sigma_t$  decreases for the IEF and is constant for the Gaussian and Deriche filters). This behaviour is different from the 1D

## Preprocessing

case where the better  $\Sigma$ , the worse the localization. This shows the appeal of Shen's and Deriche's extension to infinite impulse response filters. The limitations of the size of the support are due to the finite size of the image, the presence of a number of edges and the fact that edges are not always straight, e.g. the presence of corners. Canny's criteria are therefore incomplete for describing an edge detector. De Micheli et al (1989) have studied in detail the effect of the Gaussian filter on various types of corners or junctions. Here, for shortness, the study is limited to rectangular corners. The response of the filters is studied on crenellated and stair steps (figure 4.2.2). Then the effect of the thresholding is discussed.

Canny proposed to approximate his optimal filter with support  $W$  maximizing the uniqueness criterion  $\nu$  by the derivative of a Gaussian function :

$$f(x) = -\frac{x}{\sqrt{2\pi} \sigma_f^3} \exp\left(-\frac{x^2}{2\sigma_f^2}\right). \quad (4.2.4)$$

where  $\sigma_f$  depends on  $W$ .

For the extension to the 2D case, one should consider that the convolution by  $f(x)$  is equivalent to a convolution by  $g(x)$  followed by the calculation of the gradient. The 2D form of a filter is obtained by considering the smoothing filter  $G(r, \theta) = g(r)$ , where  $g'(r) = f(r)$  and  $(r, \theta)$  are the polar coordinates. First, the image is convolved by  $G$ , then the derivatives in both directions  $x$  and  $y$  enables the modulus of the gradient to be computed. If the kernel is separable in  $x$  and  $y$  where  $(x, y)$  are the Euclidean coordinates, i.e.  $G(r, \theta) = G_x(x)G_y(y)$ , then the gradient in  $y$  direction of the image filtered is given by the convolution of the image by  $f_y(x, y) = G_x(x) * G'_y(y)$ , which is computationally much more efficient. The advantage of the Gaussian filter is to be separable (Canny's 2D optimal filter is not separable). In the following, only the 2D case will be considered.

Remarque : For extending their filter to the 2D case, Deriche and Shen directly used the convolution by  $g(x) * f(y)$ , which in general does not

correspond to an isotropic smoothing filter (see later).

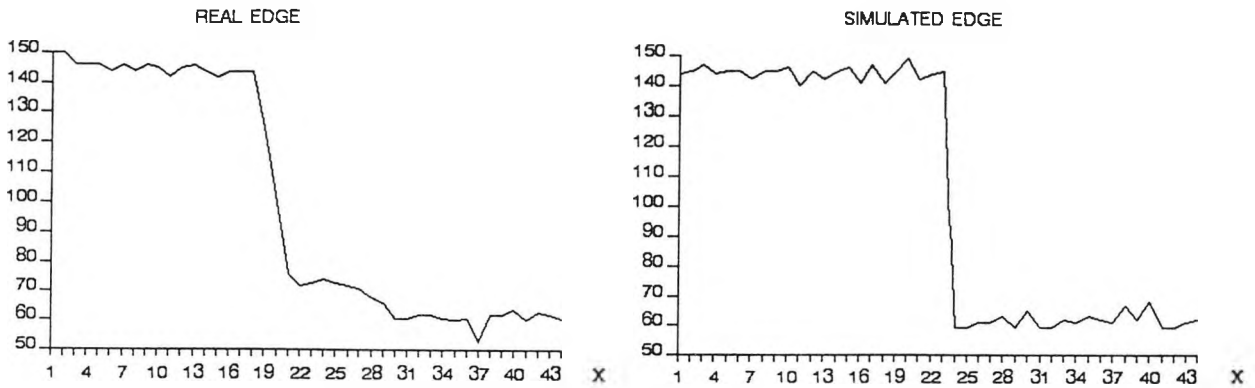


Figure 4.2.1 : A real step and a simulated noisy step

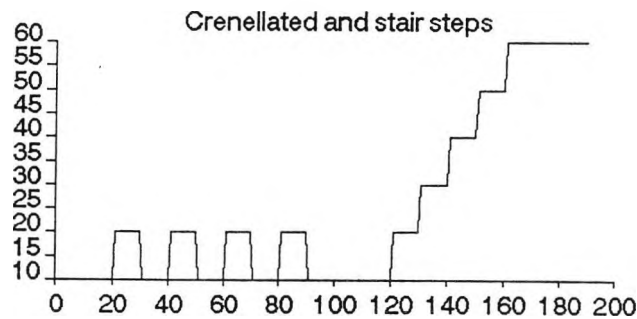


Figure 4.2.2 : A crenellated edge and stair steps

The variance of the noise after filtering is equal to  $R(0)$  ( $R(\xi)$  is the correlation function of the filtered noise and is given in appendix 7 by eq.A7.8), then  $\Sigma$  is deduced

$$R(0) = \frac{\sigma_n^2}{8\pi\sigma_f^4}, \quad \Sigma = 2\sigma_f \Sigma_0 \quad (4.2.5)$$

where  $\sigma_n$  is the standard deviation of the initial noise and  $\Sigma_0$  is the initial signal to noise ratio.

The uncertainty  $\sigma_t$  of the edge location is equal to the inverse of the localization criterion (Canny, 1983),  $\Gamma = \Delta_0 |f'(0)| / \sqrt{-R''(0)}$  (see in appendix 7, A7.14), where  $\Delta_0$  is the initial magnitude of the step. Let  $\sigma_n$  be the standard deviation of the initial noise,  $\sigma_t$  is equal to

$$\sigma_t = \sqrt{\frac{3}{8}} / \Sigma_0 \quad (4.2.6)$$



The density  $\mu$  of parasite maxima around the step is given in appendix 7 by A7.13. It decreases as  $\Sigma$  increases, i.e. as  $\sigma_f$  increases. It is proportional to  $\mu_0$  which is equal to

$$\mu_0 = \frac{1}{2\pi \sigma_f} \sqrt{\frac{5}{2}} \quad (4.2.7)$$

However, the previous analysis considers a perfect infinite length step. The response is in fact different near the corners. At a rectangular corner, say at (0,0), the response to a step with a magnitude equal to  $\Delta_0$  is

$$G(x, y) = \frac{\Delta_0}{\sqrt{2\pi} \sigma_f} \sqrt{F^2(x/\sigma_f) \exp(-y^2/\sigma_f^2) + F^2(y/\sigma_f) \exp(-x^2/\sigma_f^2)}$$

where  $F(x) = 0.5 + \text{erf}(x)$ .

If  $x = y = 0.5 \sigma_f$  then  $G_{\max} = 0.87 \Delta$ ,  $G_{\max}$  is the maximum gradient along the diagonal,

if  $x = 1.3 \sigma_f$  and  $y = 0$  then  $G = 0.95 \Delta$ .

The Gaussian edge detector rounds off the corners and decreases their contrast as  $\sigma_f$  increases. In the following, the value of  $x$  when  $y = 0$ , such that  $G/\Delta = 0.95$  is considered as a longitudinal error  $e$ . The round-up effect is measured by  $x = y = \rho$ , corresponding to the maximum of the gradient along the diagonal.

A multiple edge is defined by  $\Delta_0 \sum_{i=-n}^n \epsilon_i h_i$ , where  $h_i$  is the heaviside function located at  $id$  and  $\epsilon_i$  is equal to  $\pm 1$ , depending on the sense of the step.  $d$  is the distance between two consecutive steps. The signal to noise ratio is

$$\Sigma_d = \sum_{i \geq -n/2}^{n/2} \left( \exp\left(-\frac{id^2}{\sigma_f^2}\right) - \exp\left(-\frac{(2i+1)d^2}{2\sigma_f^2}\right) \right) + \sum_{i \geq n/2}^{n/2} \epsilon_i \left( \exp\left(-\frac{(2i+1)d^2}{2\sigma_f^2}\right) - \exp\left(-\frac{(i+1)d^2}{\sigma_f^2}\right) \right)$$

If  $d > 2\sigma_f$ , then it may be approximated by

$$\Sigma_d \approx (1 + (\epsilon_1 + \epsilon_{-1}) \exp(-\frac{d^2}{\sigma_f^2})) \Sigma \quad (4.2.8)$$

And the uncertainty of the edge point becomes

$$\sigma_{td} = \frac{\sqrt{R''(0)}}{\Delta_0 |f'(0) + (\epsilon_1 + \epsilon_{-1})f'(d) + \dots + (\epsilon_n + \epsilon_{-n})f'(nd)|} \quad (4.2.9)$$

where  $f'(d) = f'(0)(1 - \frac{d^2}{\sigma_f^2}) \exp(-\frac{d^2}{\sigma_f^2})$ ;  $f'(pd)$  tends towards 0 with  $1/p$ . It appears that the response of the edge detector depends very much on  $(\epsilon_1 + \epsilon_{-1})$ . In the case of the Gaussian filter if  $(\epsilon_1 + \epsilon_{-1}) = -2$  (i.e. crenellated edge) then the signal to noise ratio is worse than for the single step model but the localization of the edge is better, and conversely with  $(\epsilon_1 + \epsilon_{-1}) = 2$  (i.e. stair steps). By similarity to Canny's criterion  $\Sigma\Gamma$ , the response to multiple edges for  $d > 2\sigma_f$  may be qualified by the product

$$P_{\pm d} = \Sigma_{\pm d} \Gamma_{\pm d} \approx (1 \pm 2 \int_d^{+\infty} f(x) dx / \int_0^{+\infty} f(x) dx) (1 \pm 2f'(d)/f'(0)) \Sigma\Gamma \quad (4.2.10)$$

In the case of the Gaussian filter and  $d \geq 2\sigma_f$

$$P_{\pm d} \approx (1 \pm 2 \exp(-\frac{d^2}{2\sigma_f^2})) (1 \pm 2(1 - \frac{d^2}{\sigma_f^2}) \exp(-\frac{d^2}{2\sigma_f^2})) 4 \sqrt{\frac{2}{3}} \sigma_f \Sigma_0^2 \quad (4.2.11)$$

Actually, in the case of a crenellated edge, the signal to noise ratio is lower than in the single step model, but the localization is better, and conversely for stair steps. In the case of stair steps then  $P_{+d} < P_0$ , but in the case of crenellated steps if  $d/\sigma_f > 2$ , then  $P_{-d} > P_0$ . Notice that to study the case  $d/\sigma_f < 2$  more terms should be taken in the product (if  $n=1$ , there is an inversion of the sign of the step magnitude for  $d/\sigma_f$  small, which is natural since then the crenellated edge may be seen as a perturbation at the location of an edge with a magnitude  $-\Delta_0$ ).

Once the non-maxima suppression has been performed, many noisy edges

remain in the image. Let  $\Delta$  be the magnitude of the filtered edge, the thresholding is defined by

$$\Delta \geq t\sqrt{R(0)} = T, \quad (4.2.12)$$

where  $t$  is the lowest signal to noise ratio accepted and depends on the probability  $\tau$  that the grey level  $\Delta$  in the filtered image may be produced only by noise. Considering the number of points in an image, it is necessary to give the probability  $\tau$  a very small value, such as 0.0001. The value of  $T$  is linked to  $\tau$  in the following way : since  $\Delta^2/R(0)$  is a distribution function of a  $\kappa^2$  law with two degrees of freedom,  $t$  is equal to  $\sqrt{-2\text{Ln } \tau}$ , which leads to the following value of  $T$  for a normalized filter, i.e. such that  $\Delta = \Delta_0$  so that  $R(0) = \sigma_n^2/(2\sigma_f^2)$ ,

$$T = \frac{\sqrt{-2 \text{Ln } \tau} \sigma_n}{2 \sigma_f} \quad (4.2.13)$$

The effect of the thresholding is to split some edges. For example, if an edge has a magnitude equal to  $\Delta_0$ , corresponding to a magnitude equal to  $\Delta$  after filtering ( $\Delta = \Delta_0$  if the filter is normalised), and if  $T = \Delta$  then the average length of the missing parts is  $\ell_0$  given by (A7.11)

$$\ell_0 = \sqrt{2} \pi \sigma_f \quad (4.2.14)$$

Segments with a length less than  $\ell_0$  are more likely to belong to edges with an initial magnitude inferior to  $\Delta_0$ . Besides, small segments are associated with a high uncertainty on their slope and are of little use for perspective interpretation. Therefore segments with a length inferior to  $\ell_0$  are eliminated.

Thus, in the 2D case of a perfect straight line step, because of the conjugate smoothing effects along and across the step, the localization is independent of the parameter  $\sigma_f$ , and the signal to noise ratio  $\Sigma$ , the average length after thresholding  $\ell_0$  and the density of maxima  $\mu$  are improved if  $\sigma_f$  is larger. Considering only Canny's criteria leads

to the choice of the largest  $\sigma_f$  consistent with the size of the image. However, the round-up effect increases with  $\sigma_f$ , while its response to multiple edges decreases. Therefore, the optimal parameter is the largest one such that the round-up effect and the response to multiple edges with a minimal distance apart  $d$ , depending on the type of the images processed, remains acceptable.

Deriche (1987) extends Canny's approach to infinite impulse response filters. Using the same criteria as Canny, the optimal solution is

$$f(x) = \alpha_1 \sin(wx) \exp(-\alpha|x|) \quad (4.2.15)$$

When  $w$  is very small, the filter becomes  $f(x) = -cx \exp(-\alpha|x|)$ . This filter provides a much better value for  $\Sigma\Gamma$  but a slightly smaller value for  $r$  than the Gaussian filter in the 1D case ( $r$  is defined in appendix 7 and is related to  $\mu$  (A7.13)). However, Deriche shows that it is possible to have an exponential filter for which both values ( $\Sigma\Gamma$  and  $r$ ) are above the values corresponding to the Gaussian filter. For applying the filter to a 2D signal, a smoothing filter is used  $h(x) = k(1+a|x|)\exp(-ax)$ , so that the  $y$  gradient is obtained by convolving the image by  $h(x)$  in the  $x$  direction and by  $f(y)$  in the  $y$  direction.

The same analysis as for the Gaussian filter gives the signal to noise ratio and the transversal uncertainty for edges parallel to  $x$  and  $y$

$$\Sigma = \frac{8}{\sqrt{5} a} \Sigma_0 \quad \sigma_t = \frac{\sqrt{5}}{8} / \Sigma_0, \quad (4.2.16)$$

Actually  $\Sigma$  depends on the orientation of the edge and decreases to  $\sqrt{10} \Sigma_0/a$  when the angle between the nearest axis,  $Ox$  or  $Oy$ , and the edge increases to  $\pi/4$ .

The density of false maxima  $\mu$  around the step (A7.13) is substantially higher than in the case of the Gaussian filter for the same  $\Sigma/\Sigma_0$  because of the value of  $\mu_0$ .

$$\mu_0 = \frac{\alpha \sqrt{5}}{2\pi} \quad (4.2.17)$$

The effect on corners depends on the orientation of the corner because this filter is not isotropic. For a corner, the edges of which are parallel to the x and y axes, the response is

$$G = \Delta \sqrt{K_1^2(x) K_2^2(y) + K_1^2(y) K_2^2(x)}$$

where  $K_1(x) = 1 - (2+\alpha|x|)\exp(-\alpha|x|)/4$  and

$$K_2(x) = (1+\alpha|x|)\exp(-\alpha|x|)$$

If  $x = y = \rho = 0.73/\alpha$  then  $G = 0.8 \Delta$

if  $y = 0$  and  $x = e = 3.2/\alpha$  then  $G = 0.95 \Delta$

The response to multiple edges is given by

$$P_{\pm d} = (1 \pm 2(1+\alpha d) \exp(-\alpha d))(1 \pm 2(1-\alpha d) \exp(-\alpha d)) \frac{64}{5a} \Sigma_0^2 \quad (4.2.18)$$

Actually, as for the Gaussian filter, the signal to noise ratio of the Deriche filter is better but the localization is worse in the case of stair steps than for the single step model ; conversely for the crenellated edge. However, here  $P_{-d} < P_0 < P_{+d}$ , whilst for the Gaussian and the Deriche filters  $P_{-d} P_{+d} < P_0^2$ .

The threshold T associated with  $\tau$  for a normalized filter is

$$T = \sqrt{-2 \ln \tau} \frac{\sqrt{5} a}{8} \sigma_n \quad (4.2.19)$$

The average length of the segments (or missing parts) after thresholding at the step magnitude  $\Delta$  is

$$l_0 = \frac{2\pi}{\alpha} \quad (4.2.20)$$

Also referring to Canny's criteria, Shen (1986, 1990) extends the initial set of functions not only to infinite impulse response but also to discontinuous functions at the origin. The resulting optimal function is the derivative of an exponential

$$f(x) = \text{sign}(x) C \exp(-\alpha|x|) \quad \text{and} \quad h(x) = k \exp(-\alpha x) \quad (4.2.21)$$

The signal to noise ratio after filtering depends on the direction of the edge. Assuming the edge parallel to one of the axes, the signal to noise ratio is equal to

$$\Sigma = \frac{2 \Sigma_0}{\alpha} \quad (4.2.22)$$

For an edge forming an angle  $\theta$  with the nearest axis,  $0 \leq \theta \leq \pi/4$ ,  $\Sigma$  should be replaced by  $\Sigma/(\cos \theta + \sin \theta)$ . Therefore, the signal to noise ratio varies between  $\sqrt{2} \Sigma_0/\alpha$  and  $2\Sigma_0/\alpha$  with respect to the orientation of the edge.

The discontinuity at the origin prevents the use of the Taylor Lagrange development. Two cases should be considered : whether the zero-crossing  $y_1$  is at the step location  $y_0$  or it is delocalized. If  $y_1 \neq y_0$ , then the Taylor-Lagrange development may be used again, but introduces a bias due to the fact that  $f(0^+) \neq 0$ . Thus, three populations of  $\xi = y_1 - y_0$  exist, one population at zero, and two populations centred at  $\pm E(\xi|\xi > 0)$ . The probability of  $\xi$  to be zero is

$$p(0) = 2 \operatorname{erf}\left(\frac{\Sigma_0}{\sqrt{\alpha}}\right) \quad (4.2.23)$$

Thus, the smaller  $\alpha$ , the higher the probability to detect the edge at the proper location. Assuming an edge parallel to one of the axes, the expected value and the variance of  $\xi$  when  $\xi > 0$  are

$$E(\xi > 0) = \frac{1}{\alpha} \quad \text{var}(\xi > 0) = \frac{1}{\alpha \Sigma_0^2} \quad (4.2.24)$$

Therefore, the smaller  $\alpha$ , the larger  $E(\xi > 0)$  and the larger  $E(\xi > 0)/\sqrt{\text{var}(\xi > 0)}$ , which means that if  $\xi \neq 0$  and  $\alpha$  small enough, then  $\xi$  is very likely to be far from the edge and therefore corresponds to a noise maxima which is eliminated by thresholding. Thus, conversely to the Gaussian and Deriche filters which keep the localization constant, here the smaller  $\alpha$ , the better the localization and the smaller the risk of multiple edges at the edge location. Moreover, the localization

## Preprocessing

increases very rapidly with the initial signal to noise ratio, much quicker than for the other filters.

If  $\alpha$  is large, say  $\alpha > \Sigma_0^2$ , then the probability of not detecting the edge at the proper location but at  $\xi \neq 0$  is no longer small. Since  $\xi$  always exists, when  $\xi \neq 0$ ,  $|\xi|$  is likely to be small, corresponding to a signal close to  $\Sigma$ , then the edge is wrongly detected at  $\xi$ . Therefore if  $\alpha$  is large, the variance of the edge location is estimated by

$$\sigma_t^2 = \left( \frac{1}{\alpha \Sigma_0^2} + \frac{1}{\alpha^2} \right) \left( 1 - 2 \operatorname{erf} \left( \frac{\Sigma_0}{\sqrt{\alpha}} \right) \right)$$

This gives  $\sigma_t \approx 0.57/\Sigma_0^4$ , when  $\alpha = \Sigma_0^2$ . Therefore the localization is still good, but the risk of multiple edges at the edge location is high.

The density of noise maxima when  $\alpha$  is small enough is nearly constant and equal to

$$\mu_0 \approx \frac{1}{2\pi}$$

Actually, the discontinuity at the origin is theoretical and in practice  $\mu_0$  decreases slowly with  $\alpha$ . However, for the parameter  $\alpha$  in the range [0.25, 0.5], the above value of  $\mu_0$  has been checked experimentally. This relatively small value of  $\mu_0$  is compensated by the value of  $\zeta$ , equal to

$$\zeta = \sqrt{\frac{2}{\alpha}},$$

so that if  $\alpha$  is small enough, the density  $\mu$  of noisy maxima at the edge location is small (see figures 4.2.3 and 4.2.6).

The response to a rectangular corner with the edges parallel to the axes is

$$\Sigma = \Sigma_0 \sqrt{K_1^2(x) K_2^2(y) + K_1^2(y) K_2^2(x)}$$

where  $K_1(x) = 1 - \exp(-\alpha|x|)/2$  and

$$K_2(x) = \exp(-\alpha|x|)$$

If  $x = y = \rho = 0$  then  $G = 0.5 \Delta$

if  $y = 0$  and  $x = e = 2.3/\alpha$  then  $G = 0.95 \Delta$

Thus, for such corners, there is no round-up effect with the Shen detector but only a loss of contrast.

The response to multiple edges is

$$P_{\pm d} = (1 \pm 2 \exp(-\alpha d) \mp 2 \exp(-2\alpha d) \pm \dots \mp n \exp(-n\alpha d))^2 \frac{2}{\alpha} \Sigma_0 / \sigma_t \quad (4.2.25)$$

If  $\exp(-n\alpha d)$  is small then

$$P_{\pm d} = (1 \pm \frac{2 \exp(-\alpha d)}{1 + \exp(-\alpha d)})^2 \frac{2}{\alpha} \Sigma_0 / \sigma_t \quad (4.2.26)$$

The estimation of  $P_{\pm d}$  is limited by the difficulties for estimating  $\sigma_t$ , however it can be seen that, in the stair case, it tends towards infinity when  $\alpha$  tends towards zero. Actually, in this case both signal to noise ratio and localization are better than in the single step case. In the crenellated step case, when  $\alpha d$  tends towards zero, then

$$P_{-d} \cong \frac{2\alpha d^2 \Sigma_0}{\sigma_t} \quad (4.2.27)$$

As  $\sigma_t$  tends towards zero with  $0.5 - \text{erf}(\Sigma_0 / \sqrt{\alpha})$ ,  $R_{-d}$  tends towards infinity when  $\alpha$  tends towards zero. However it has been seen that  $\alpha$  cannot be too small. Moreover  $P_{-d} = k(\alpha)P_0$ , when  $k(\alpha)$  tends towards zero with  $\alpha$ . Actually, in this latter case both signal to noise ratio and localization are worse than in the single case but still they are better when  $\alpha$  is relatively small. The response of the exponential filter to multiple edges works conversely to the Gaussian or Deriche filter with respect to stair steps or crenellated steps. Actually, if  $\alpha$  is small enough, then the location of the edges using the exponential filter is always good, only the signal to noise ratio of the crenellated edge tends towards zero (and thereby the density of false maxima at the edge location increases). Notice that here  $P_{-d}P_{+d}$  is not upper bounded by  $P_0^2$ , conversely to the Gaussian and Deriche's filters.

The problem of Shen's detector is its strong anisotropy. However,



## Preprocessing

the fact that  $\sigma_t$  and the signal to noise ratio tends toward 0 with  $\alpha$ , is mainly responsible for most other properties to improve when  $\alpha$  decreases, which is very attractive. Therefore, an isotropic exponential filter has been searched for. It is given by the smoothing filter :

$$f(x,y) = K \exp (-\alpha \sqrt{x^2 + y^2}) \quad (4.2.28)$$

which leads to the gradient operators in the x and y directions

$$\begin{aligned} f'_x(x,y) &= - \frac{C x}{\sqrt{x^2 + y^2}} \exp (-\alpha \sqrt{x^2 + y^2}) \\ f'_y(x,y) &= - \frac{C y}{\sqrt{x^2 + y^2}} \exp (-\alpha \sqrt{x^2 + y^2}) \end{aligned} \quad (4.2.29)$$

These 2D masks give excellent results, but they are not separable and hence are computationally very expensive ; moreover a small value of  $\alpha$  substantially reduces the size of the image processed. Part of the appeal of the Shen detector is its recursive implementation, resulting in a fast algorithm, independent of the value of  $\alpha$ . An approximation of  $f'_x$  and  $f'_y$  in the form of separable kernels which could use such an implementation is looked for.

$$f'_x \approx g_x = \varepsilon_x C' \exp (-\alpha' |x| - \alpha'' |y|) \quad (4.2.30)$$

with  $\varepsilon_x = 1$  if  $x \geq 0$  and  $-1$  otherwise. The response of  $g_x$  along an edge with the direction  $\theta$  is given by

$$\Delta = \Delta_0 \sqrt{\Delta_x^2 + \Delta_y^2}$$

$$\text{where } \Delta_x = \Delta_0 \left( 1 - \frac{\alpha'}{\alpha' + \alpha'' \tan \theta} \right)$$

$$\text{and } \Delta_y = \Delta_0 \left( 1 - \frac{\alpha'}{\alpha' + \alpha'' \cotan \theta} \right).$$

$\alpha'' = \alpha'$  is the Shen detector

If  $\alpha'' \leq 2\alpha'$ , then  $\Delta_\theta$  decreases from 1, when  $\theta=0$ , to  $\Delta_{\pi/4}$ , when  $\theta = \pi/4$ , and then increases up to 1, when  $\theta = \pi/2$ . If  $\alpha'' > 2\alpha'$ , three extrema

appear in the range  $]0, \pi/2[$ .

Now, if  $\alpha'' = 2\alpha'$ , then  $\sqrt{\frac{8}{9}} \Delta_0 \leq \Delta_\theta \leq \Delta_0$ . Thus, compared with the Shen detector, the anisotropy has been drastically reduced. The signal to noise ratio of the filtered image is

$$\Sigma \approx \frac{\sqrt{2}}{\alpha'} \Sigma_0 \quad (4.2.31)$$

It gives the same signal to noise ratio as  $f'_x$  for  $\alpha' = \alpha/\sqrt{2}$ . The probability of locating the edge with an error equal to zero is

$$p(0) = 2 \operatorname{erf}\left(\frac{\Sigma_0}{\sqrt{2}\alpha'}\right) \quad (4.2.32)$$

The response on a rectangular corner is

$$\Sigma = \Sigma_0 \sqrt{K_1^2(x) K_2^2(y) + K_1^2(y) K_2^2(x)}$$

where  $K_1(x) = 1 - \exp(-2\alpha' |x|)/2$  and

$$K_2(x) = \exp(-\alpha' |x|)$$

If  $x = y = \rho = 0.088/\alpha'$  then  $\Sigma = 0.61 \Sigma_0$

if  $y = 0$  and  $x = e = 1.15/\alpha'$  then  $\Sigma = 0.95 \Sigma_0$

Thus, a slight round-up effect  $\rho$  is introduced but with a smaller longitudinal error  $e$ .

The threshold associated with the risk  $\tau$  for a normalised filter is

$$T = \sqrt{-2 \operatorname{Ln} \tau} \frac{\alpha'}{\sqrt{2}} \sigma_n \quad (4.2.33)$$

The average length of an edge segment when thresholding at  $\Delta$  is

$$l_0 = \frac{\pi}{2\sqrt{2}\alpha'} \quad (4.2.34)$$

The value of  $l_0$  is very low (see figure 4.2.3). This means that edges such that  $\Sigma_0 < \Sigma_{0\min}$  are very likely not to be detected at all or to have a very small length and thereby will be eliminated (in that case

## Preprocessing

$\ell_{\min} > \ell_0$ , because small segments are associated with too large an uncertainty to be taken into account in the interpretation process). As a result false edges are very unlikely. However, it also means that significant edges may be broken into small segments, which is unacceptable. Let  $T$  be the threshold applied to the filtered image,  $\ell_1$  the average length of the segments detected, and  $\ell_2$  the average length of the gaps between the segments, from eq. A7.10 and  $\delta_T = 2/(\ell_1 + \ell_2)$ , it may be shown that the average length  $\ell_1$  of the segments of an edge such that  $\Delta > T$  is lower bounded

$$\ell_1 > \ell_0 \left( 2 \exp\left(\frac{(\Delta-T)^2}{2R(0)}\right) - 1 \right) \quad (4.2.35)$$

For example if  $\sigma_n = 4$ ,  $\alpha' = 0.2$  and  $\Delta = T+1$ , then  $\ell_1 > 20.6$  (in the case of a Gaussian filter with  $\sigma_f = 2$ , then  $\ell_1 > 20.4$ ). Therefore, a small value of  $\ell_0$  is compensated by a large signal to noise ratio. This also suggests that  $T$  may be smaller than with the Gaussian or Deriche filter, i.e. the risk of false edge detection  $\tau$  may be larger, as the noise is very likely to be eliminated by the further constraint  $\ell \geq \ell_{\min}$ .

Thus, the Shen edge detector has been improved by changing the ratio of the parameter of the smoothing filter to the parameter of the gradient filter. The filter is now isotropic and it keeps all the good properties of the Shen detector for edges parallel to the axes. The implementation is the same as the Shen detector, which means it is fast, independent of the value of  $\alpha$  and it does not reduce the processed part of the image. This allows choice of a very small value for  $\alpha$ , which gives altogether a very good signal to noise ratio, a very good location and a small density of parasite maxima  $\mu$  without substantially increasing the round-up effect or worsening the response to multiple edges (actually it improves the response to stair steps). Therefore, the value of  $\alpha$  is only limited by the acceptable round-up effect and the size of the image (if  $\alpha$  is too small, the response is not homogeneous over the image). Let us remember that the good properties of the isotropic exponential filter (IEF) only holds if  $\alpha$  is small. For large  $\alpha$ , it is worse than the other filters studied with respect to nearly all the criteria adopted. The value of  $\alpha$  from which

IEF is competitive is difficult to make explicit because of the difficulty arising from the discontinuity at zero. But such a difficulty is not a good reason for discarding it as an edge detector.

The Gaussian filter, Deriche filter and IEF have been compared in figure 4.2.3 with respect to a number of parameters : the signal to noise ratio  $\Sigma$ , the standard deviation of the error of localization  $\sigma_t$ , the density of maxima at the edge location  $\mu$ , the round-up effect at rectangular corners  $\rho$ , the expected value of the missing part near a corner estimated by  $e+l_0/2$ , and the average length  $l_0$  after thresholding at  $\Delta$ , and the response to multiple steps (5 units apart) given by  $m = P_5 P_{-5} / P_0^2$ , for an initial signal to noise ratio equal to 1. For proper comparison, the signal to noise ratio is identical for all the filters.

	$\Sigma$	$\sigma_t$	$\mu$	$\rho$	$e+l_0/2$	$l_0$	$m$
Gaussian 2.0	4	0.61	0.009	1	7.4	8.9	0.77
Deriche 0.85	4	0.28	0.03	0.86	6.4	7.4	0.97
IEF 0.35	4	$\approx 0.5$	0.01	0.25	4.2	1.9	0.76
Gaussian 3.5	7	0.61	0.00001	1.75	12.3	15.6	0.39
Deriche 0.5	7	0.28	0.0017	1.46	12.7	12.6	0.74
IEF 0.2	7	$\approx 0.2$	0.0005	0.44	7	2.5	0.5

Figure 4.2.3 : Theoretical comparison of the Gaussian, Deriche and IEF filters

The localization of the Deriche edge detector is better than the localization of the Canny edge detector, but the density of parasite maxima is worse for Deriche's than for Canny's detector. The other criteria have approximately the same magnitude, so it is difficult to conclude for anyone of these filters. The choice of the right parameter has been proved to be much more important than the choice of the filter and depends on the type of the image. Actually, the parameters  $\sigma_f=3.5$  for the Gaussian filter and  $\sigma_f=0.5$  for the Deriche filter are not commonly used for an image  $256 \times 256$ , but have been tested for several reasons. Firstly for showing the variations of the filter behaviour

## Preprocessing

with the parameter and secondly for completeness of the comparison with the IEF filter. For such values,  $\ell_0$  is very large which is a disadvantage as it means that the expected length of a missing part of the segment is large too. To decrease the threshold is not necessarily the answer, as the number of noisy edges with a significant expected length may still be extracted. The IEF filter behaves differently than the other filters because the increase of the signal to noise ratio only produces a small degradation of the corners and of the response to multiple edges, moreover  $\ell_0$  remains low. Conversely to the other filters with the same signal to noise ratio,  $\alpha=0.2$  is meant to be a usual value.

Thus, the Gaussian and Deriche filters with usual parameters (e.g.  $\sigma_f \approx 1$  or  $\alpha \approx 1$ ) give good results for multiple edge detection and preserves corners but have poor  $\Sigma$  and  $\mu$  in presence of noise ; whereas the IEF with  $\alpha'=0.2$  still gives reasonable results for multiple edges and corners and has good  $\Sigma$  and  $\mu$  in presence of noise. This suggests that the IEF is particularly appropriate to noisy images. The recursive implementation of the Deriche filter and the IEF is an important advantage as the corresponding algorithm is fast and does not truncate the image (The IEF is faster than the Deriche filter).

The four filters, Gaussian with  $\sigma_f = 3.5$ , Deriche with  $\alpha = 0.5$ , IEF with  $\alpha' = 0.2$  and IEF using the 2D masks (eq.4.2.29) with  $\alpha = 0.3$  (they all correspond to a signal to noise ratio equal to  $7\Sigma_0$ ), have been tested on a simulated step image. Two rectangles representing step edges in 8 directions have been superimposed with noise so that the initial signal to noise ratio is equal to 1, i.e.  $\Delta_0 = \sigma_n = 16$ , (figure 4.2.4). All the filters have been normalized so that  $\Delta = \Delta_0$ , in order to give comparable results. Figure 4.2.6 shows the response of the edge detectors and figure 4.2.7, the effect of a thresholding to  $\Sigma-1$ . The recursive version of the IEF appears to be an excellent approximation of the 2D IEF. The texture of the noise is remarkably different using the Gaussian filter or the IEF, illustrating the different values of  $\mu_0$ , although the value of  $\mu$  are almost comparable. Considering that the localization is limited by the uncertainty of the edge location due to

the digitization, the experimental results are consistent with the theory.

Another test image (figure 4.2.5) representing crenellated and stair steps distance  $d = 5$  apart, have been used to test the response of the edge detectors to multiple edges. It may be noticed that the localization is very good with the IEF, although the  $\Sigma$  of the crenellated edge is low. As expected, the stair is completely delocalized with the Gaussian and the Deriche filters but is well detected with the IEF. The Shen detector with  $\alpha = 0.3$  and the IEF with  $\alpha' = 0.2$  (same signal to noise ratio) have been compared on the image 4.2.4, the results are displayed in figure 4.2.8.

The three edge detectors, Gaussian 3.5, Deriche 0.5 and IEF 0.2, have been tested on an image of an indoor scene of a power plant (figure 4.2.10). The thresholds have been computed, using the value of  $\sigma_n$  estimated as follows. The value of  $\sigma_n$  has been estimated on the image filtered (before the non maxima suppression has been performed) on large uniform areas, selected by hand. This process has been performed on three different images and on various parts of the same image. The uncertainty of the various areas has been shown to be fairly uniform, in the range [3.0,5.0] and the expected value of  $\sigma_n$  has been fixed to 4.0. The results are displayed in figures 4.2.11 to 4.2.13. The results are roughly comparable ; however it may be noticed that the rectangular texture of the grid at the front is only detected by the IEF and that it is an important feature for determining the direction of the camera in the scene (i.e. the location of the vanishing points).

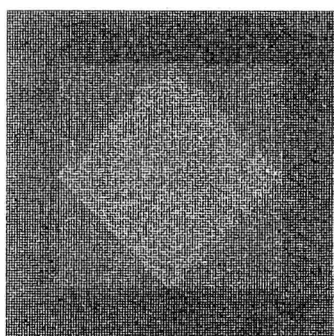


Figure 4.2.4 : First test image  
(Grey levels : 0-3 : bleu; 4-7 : pink; 8-11 : red; 12-16 : violet; >16 black).

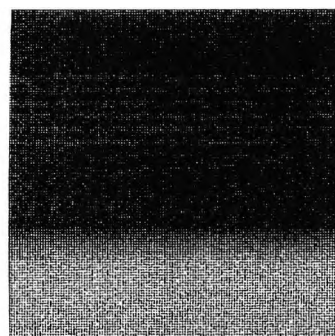


Figure 4.2.5 : Second test image

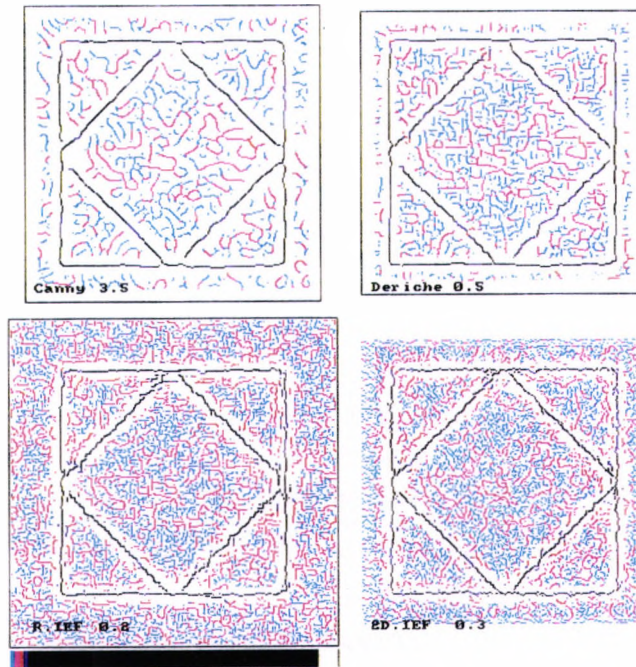


Figure 4.2.6 : Comparison of edge detectors on the first test image

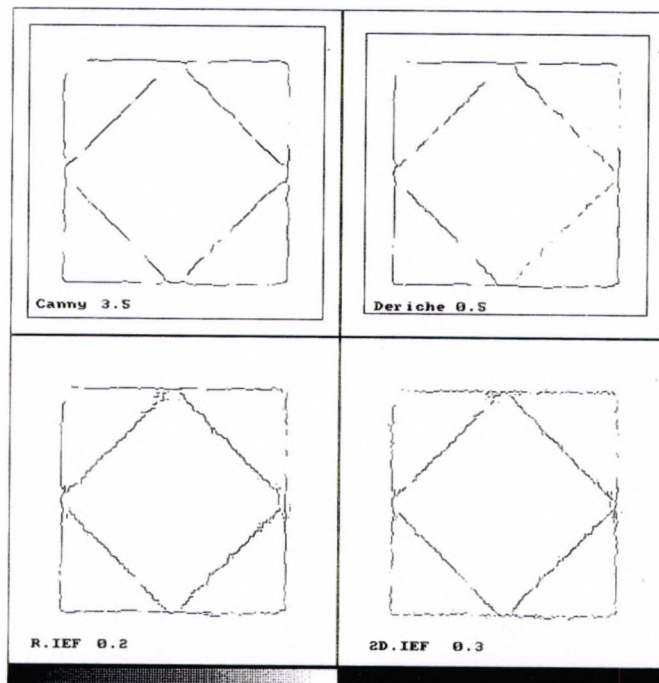


Figure 4.2.7 : Image displayed in figure 4.2.6, thresholded at  $\Sigma=1$ .

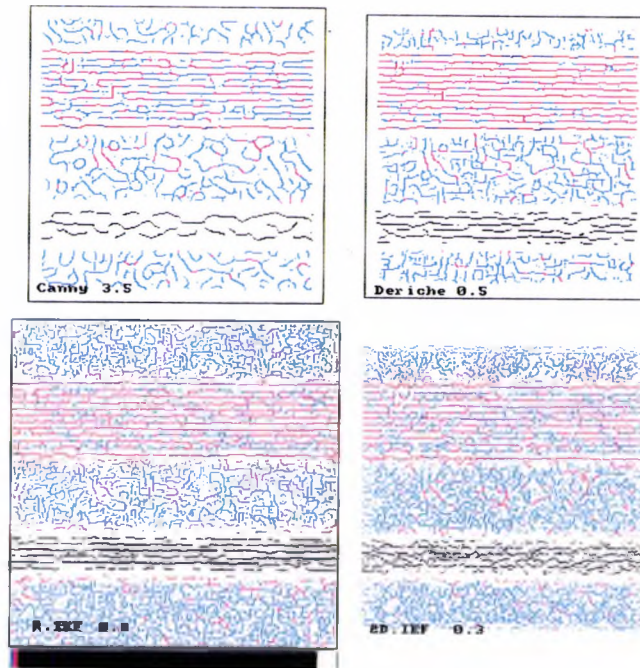


Figure 4.2.8 : Comparison of edge detectors on the second test image (Grey levels identical to figure 4.2.6)

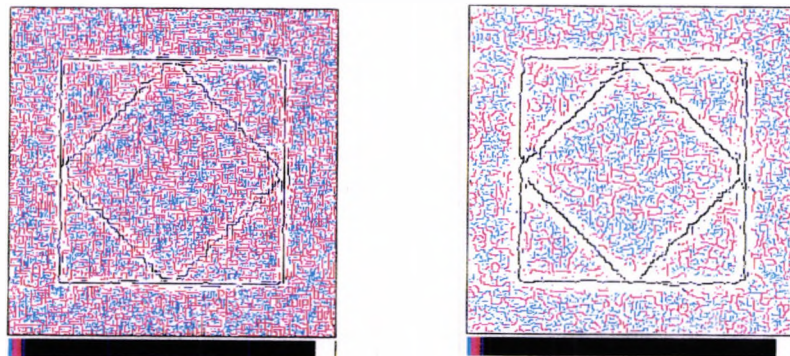


Figure 4.2.9 : Comparison of the Shen detector ( $\alpha=0.3$ ) and the IEF ( $\alpha'=0.2$ ). (Grey levels identical to figure 4.2.6).



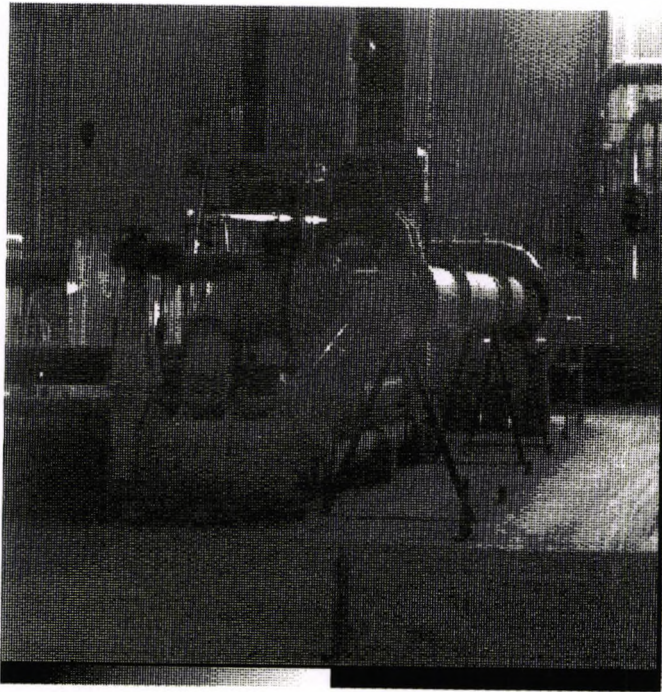


Figure 4.2.10 : Initial image of an indoor scene.

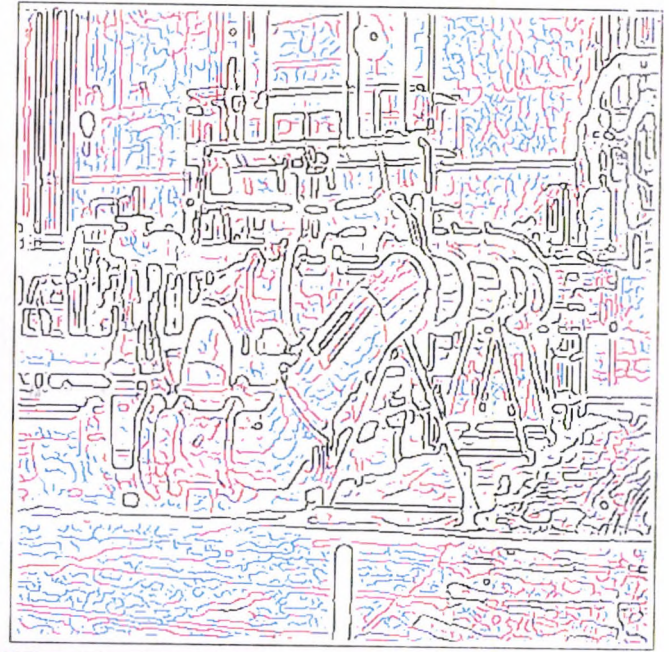


Figure 4.2.11: Gaussian filter  
 $\sigma_f = 2$ .

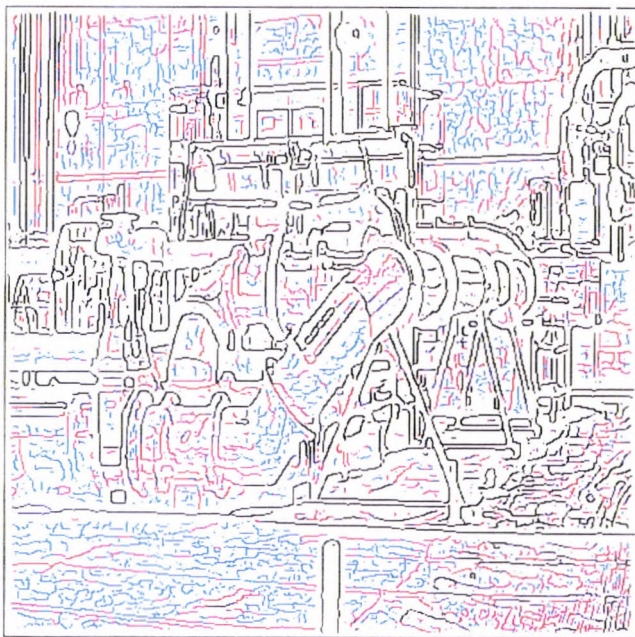


Figure 4.2.12 : Deriche's detector  
 $\alpha = 0.85$   
(Grey levels identical to figure 4.2.6)

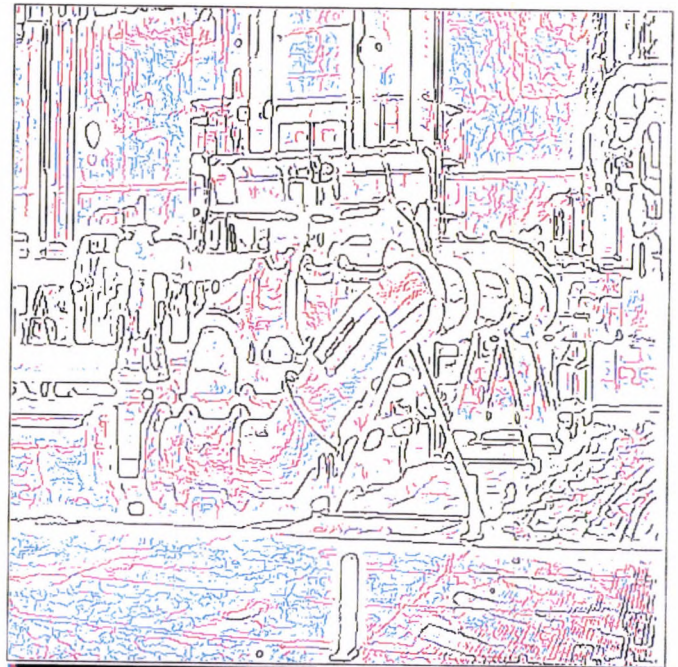


Figure 4.2.13 : IEF,  $\alpha' = 0.3$

Canny's approach has been used for comparing three different shapes of filter for edge detection. It has been shown that Canny's criteria are not sufficient for describing an edge detector and that additional criteria were necessary, such as the round-up effect near a corner, the response to multiple edges and the sensitivity to thresholding. It has been shown that the choice of the parameter has more influence on the chosen criteria than the choice of the shape of the filter.

An improved version of the Shen detector, the IEF has been described. This edge detector has been proved to have important advantages when using a small value of the parameter  $\alpha$  : the signal to noise is high while the uncertainty of the edges is very low and tends rapidly towards zero when the initial signal to noise ratio increases, even for multiple edges ; the density of noisy edges at the edge location is very low, it does not round-off the corners and it produces very few false edges because of the low value of  $\ell_0$ . An additional advantage is its recursive implementation, which gives a fast algorithm which does not truncate the borders of the image. The Shen detector fulfilled all these properties for vertical and horizontal edges, but it is not isotropic, when the IEF is. The IEF has been shown very appropriate to noisy images.

Remark : all computations have been performed in floating point and the non-maxima suppression has been performed using an interpolation between the eight neighbours.

### 4.3 LINE AND ELLIPSE FINDER

The scenes studied are mostly composed of simple geometric shapes, such as rectangles (e.g. the walls) and cylinders (e.g. the tanks). The representation of the scene using such shapes is very efficient, and it is also the representation used for CAD databases. It is necessary to extract their projections in the image. This includes the extraction of straight line segments and elliptical arcs.

A very popular method is the recursive segmentation of the connected edges. A chord links the endpoints of an edge which is then split at

## Preprocessing

the point of maximum deviation, when above a threshold. The difficulties of this method is the instability to noise as a small perturbation may produce a global change, and the fact that the detected segments are not the best approximation in the least mean square (LMS) sense. Once the segmentation is stable, it is possible to compute the LMS approximation, but the determination of the new endpoints is still a problem. Either the connectivity may be jeopardized, or the endpoints may be far from the actual ones, e.g. in the case of adjacent segments with an acute angle.

A similar method is used by Lowe (1985), but there is no thresholding on the deviation. An edge is split into two edges when the significance of the edges increases. Rosin and West have extended the method to circular arcs (1988), which then are used to find elliptical arcs (1990). A sequence of straight line segments is hypothesized to be an arc. A significance measure is associated with the arc, if it is lower than the significance of the corresponding straight lines then the hypothesis fails. The result of the algorithm is the set of straight line segments and selected arcs. These methods are invariant to scale and require no arbitrary thresholds.

Another method inspired by Sklansky and Gonzalez's work (Sklansky, 1980) has been developed by Berthod (Ayache, 1988). Let  $\{M_0, \dots, M_n\}$  be the sequence of points to approximate. The sectors  $C_i$  with summit  $M_0$ , axis  $M_0M_i$  and angle  $\theta_i$  are considered. The intersection of this set of sectors is a sector with an axis  $\mathcal{A}$  closer to  $M_0M_p$  if the sequence  $\{M_0, \dots, M_p\}$  is closer to a straight line. If  $\{M_{p+1}, \dots, M_q\}$  is another straight line, the intersection of the sectors  $C_1, \dots, C_q$  is empty. Let  $D_k$  be the intersection of all  $C_i$ , with  $i \leq k$  and  $\mathcal{A}_k$  its axis, and let  $(M_0, M_j)$  be the axis of the last cone  $C_j$  such that  $C_1 \cap \dots \cap C_j \neq \emptyset$  and  $C_1 \cap \dots \cap C_{j+1} = \emptyset$ . The first minimum  $d(M_k, \mathcal{A}_k)$  of the distance from a point of the sorted set  $(M_j, \dots, M_0)$  to  $\mathcal{A}_k$  determines the splitting point,  $M_k$ .

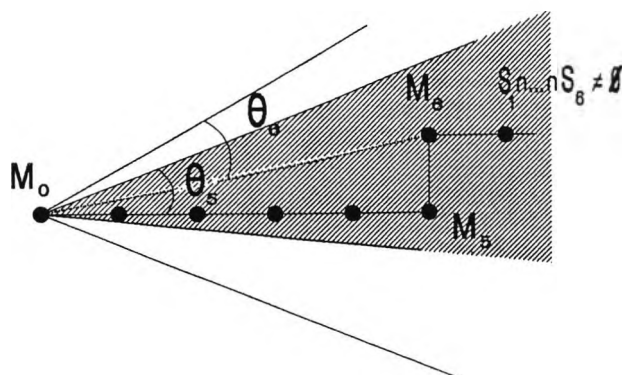


Figure 4.3.1 : Principle of Berthod's algorithm

Unfortunately, too little information is available on the implementation of Berthod's algorithm to make a serious analysis (the algorithm is NOESIS's version (1988) and the source code is not available). An experimental study has been done instead.

On the one hand, the choice of the parameter  $\zeta_0 = i\theta_1$  has to be done in a way consistent with the error  $\sigma_t$  of the edge detector, in order not to produce numerous small segments which would be useless. The maximum deviation between the straight line detected and the edge is upper bounded by  $2\zeta_0$ . The expected error  $\sigma_t$  is given by 4.2.6 and 4.2.4. Therefore  $\zeta_0$  should be much higher than  $\sigma_t$  in order not to break straight line segments. On the other hand, it should not be too high such that it is not responsible for too large an error on the end points.

The main source of error is splitting errors, i.e. the line finder splits the curve at the wrong locations. For example, it may be due to a local perturbation or a rounded corner. Figure 4.3.2 shows how a small local perturbation may affect Berthod's line finder. In that case, the noise cannot be considered as Gaussian noise with a zero mean on the corresponding part of the edge. In this type of configuration a systematic error on the endpoints occurs, which is bounded by  $2\zeta_0$ .



Figure 4.3.2 Effect of a small perturbation on Berthod's line finder

Therefore, if a straight line segment is an approximation of a segment within a straight line edge, e.g. an edge broken because of the thresholding on its magnitude, then the transversal uncertainty of the end points should be  $\sigma_t$ . On the contrary, if a straight line segment approximates the end part of the straight line edge, the transversal uncertainty may be as high as  $2\zeta_0$ . This latter error is more likely with long segments, as long segments are likely to correspond to a high signal to noise ratio and thereby have their connectivity preserved. As long segments play an important part in the following chapter, it is important to keep  $\zeta_0$  relatively small. Thus, a trade-off must be found between the break-up effect produced by too small a value of  $\zeta_0$  and splitting errors bounded by  $\zeta_0$ .

Let us remark that a LMS approximation of the segment after the splitting stage would have substantially reduced the effect of the splitting error. The endpoints are then defined as being the orthogonal projection of the real endpoints onto the straight line segment. But at the time this work began, the loss of the connectivity due to this latter stage seemed a serious drawback and, moreover, this option was not available in our current implementation. Now, as the connectivity of the segments has not been directly used (but only through a proximity criterion), the LMS approximation seems to be a necessary improvement for the future.

Berthod's algorithm with the parameter  $\zeta_0=1.5$  has been tested on the test picture used in section 4.2, using the three edge detectors studied in this previous section.

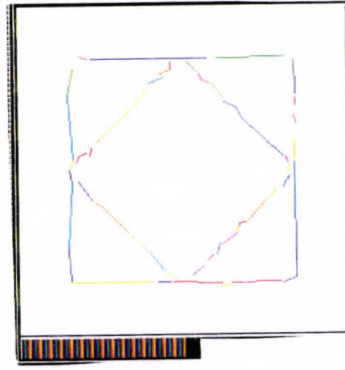


Figure 4.3.3 : Berthod's line finder results on a test image

Berthod's and Lowe's algorithms are compared in figure 4.3.4 and 4.3.5 on a real image.

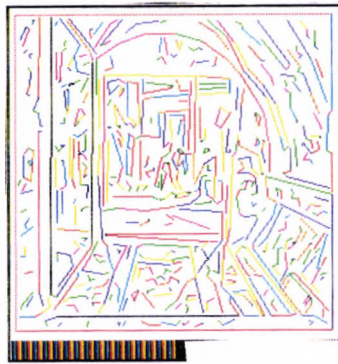


Figure 4.3.4 : Berthod's line finder      Figure 4.3.5 : Lowe's line finder

For practical reasons, Berthod's algorithm has been selected for the detection of lines in most of the images processed here. A more serious analysis of the existing line finders and the research of the appropriate criteria is necessary for improving this stage of the processing. However, the existing detectors could be improved by using a LMS approximation of the detected segments.

The smallest straight line segments extracted by the line finder are discarded, because they are too numerous and the direction is unreliable. Let  $\ell_{\min}$  be the minimal length allowed,  $\ell_{\min}$  is lower or equal to  $\ell_0$ , where  $\ell_0$  is defined in section 4.2.

In Rosin and West (1988) the algorithm has been used when

significant elliptical arcs were expected in the image. Large ellipses have been missed because of various distortions (camera distortion, aggravated by noise). Others have been missed because of the poor connectivity of the edge detected. Current work is being undertaken to estimate the uncertainty of the approximation and to group unconnected edges for hypothesizing an elliptical arc (Ellis et al, 1991).

## 4.4 STATISTICAL MODEL

### 4.4.1 Modelling of the measurement error

The first source of uncertainty is the difference between the real scene and the symbolic representation describing this scene, e.g. CAD representations.

The second source of uncertainty comes from the image acquisition, digitization and preprocessing process. The acquisition of the image is modelled by a pin-hole perspective model. This does not take into account the possible distortions of the image caused by the lens. The Gaussian filter has an accuracy limited by the value of the parameter  $\sigma_f$  of the Gaussian filter (section 4.2). The line finder may introduce non-negligible errors for the slope of a segment.

To clarify, let us assume that the camera calibration parameters are known (though unnecessary for the vanishing point detection) and the ideal representation of the scene is also known, e.g. in the form of a CAD description. The error of an endpoint location is defined as the difference between the observed endpoint location, and the projection of the corresponding ideal endpoint in the scene using a pin-hole model (for which the parameters are known).

The sources of the uncertainty are various and difficult to quantify at times. However it has been seen how to estimate the uncertainty caused by the edge detector and the line finder.

The most obvious source of uncertainty is due to digitization. Let  $y_0$  be the location of the digitized edge on an axis perpendicular to

the edge. The original edge is equally likely on  $[y_0-a, y_0+a]$ , with  $a=0.5$  if the edge is vertical or horizontal and  $a=0.7$  if the edge is along a diagonal. As the directions are assumed equally likely the expected value of the error of an edge location due to digitization is approximately 0.35.

The errors due to the edge detector have various causes. The first cause is the lack of localization of the edge, which is measured by Canny's criterion  $\Gamma$ . The second one is the bad connectivity of the edges due to the threshold on the maximum of the gradient. As a result part of the edge is missing. This is integrated in the modelling in the form of the uncertainty of the endpoint in the direction of the segment. The third cause is the bad response to high curvature features, such as corners. The last cause is the presence of parasite edges.

The errors due to the line finder is break-up, which produces small segments which are then eliminated, and splitting errors, e.g. at corners. As the small segments are eliminated, parts of the edge are missing and this is represented by an uncertainty of the endpoint in the direction of the straight line segment. Splitting errors are responsible for an error in the transverse direction of the segment.

Eventually, the endpoint uncertainty is modelled by a Gaussian law with its main axes parallel and perpendicular to the straight line segment. The transverse uncertainty is  $\sigma_0$  and the longitudinal uncertainty is  $\sigma_1$  (see figure 4.4.1.1).

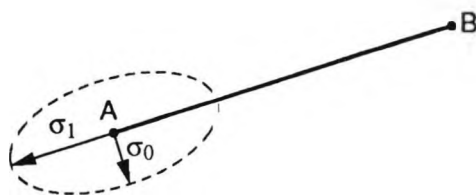


Figure 4.4.1.1 : Model of the endpoint uncertainty



The value of  $\sigma_0$  is lower bounded by the digitization error. It depends on  $\sigma_t$ ,  $\rho$ ,  $\zeta_0$  and errors more difficult to quantify such as the lens distortion. Moreover it should take into account the error due to the difference between the geometric model and the scene.

The definition of the transverse component is trivial, but the definition of the longitudinal component  $\sigma_1$  is not. First, the use of this uncertainty has to be defined. As reported several times in the previous sections and chapters, a major difficulty is to extract information from the connectivity of the segments. Through edge detection, thresholding and the line finder, the connectivity of the segments may have been lost. This is not crucial for the interpretation of the perspective, but it is essential to the construction of higher level primitives such as rectangles. Therefore  $\sigma_1$  should represent the standard deviation of the missing parts, i.e. gaps due to thresholding plus eliminated small segments, in order to guide the construction algorithm of the 3D primitives. This error should not take into account the relevance of the connectivity (for example the fact that it corresponds to a 3D connectivity), as this will be done by the likelihood test, but only the fact that two segments might be connected in the image. Several configurations may occur. Let us consider three typical cases. The missing part is a gap due to thresholding and is less than  $l_0$ . Or it is a small segment surrounded by two gaps, the corresponding missing part of which is less than  $2l_0 + l_{\min}$  long. Or it is near a corner, and the missing part may be the truncated corner plus a gap due to thresholding, and the missing part is less than  $e + l_0$ , where  $e$  is the round-up effect defined in section 4.2. The uncertainty  $\sigma_1$  is empirically estimated to  $\sigma_1 = k l_{\min} + l_0/2 + e$ , where  $k$  is the probability of an edge having its length equal to  $l_{\min}$ , e.g.  $k=0.5$  if  $l_{\min} = l_0$ .

The error on each end point location may be represented by an uncertainty ellipse defined by the covariance matrix of the endpoint. Let the matrix  $\mathbf{V}$  be

$$\mathbf{V} = \begin{bmatrix} V_x & V_{xy} \\ V_{xy} & V_y \end{bmatrix}$$

A point M of the uncertainty ellipse is defined by

$$\overrightarrow{AM}^t \mathbf{V}^{-1} \overrightarrow{AM} \leq 1$$

Let  $\vec{u}$  be the unit vector of the line L corresponding to the segment AB, and  $\vec{v}$  be the perpendicular unit vector. The principal axes of the uncertainty ellipse are in the direction  $\vec{u}$  and  $\vec{v}$ , with the major and minor axes  $2\sigma_1$  and  $2\sigma_0$ .

$V(\vec{w})$  is the variance of the error in the direction perpendicular to  $\vec{w}$  (figure 4.4.2) and is equal to

$$V(\vec{w}) = (\vec{w}^t \mathbf{V}^{-1} \vec{w}) \det(\mathbf{V}) \quad (4.4.1)$$

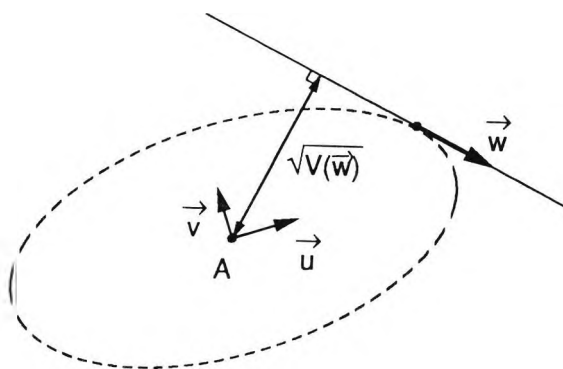


Figure 4.4.1.2 : Geometric interpretation of  $V(\vec{w})$

#### 4.4.2 Modelling of the segmentation error

The aim of this section is to model the noise created by the proximity of unrelated lines to the features of interest. For example, let the straight line segment S be the feature of interest, if the projections of the lines connected to S in the 3D space are looked for, all the other straight line segments connected to S in the image appear to be a type of noise, called segmentation noise. Another example is the search for vanishing points using a whole line accumulation strategy ; a line passing near a vanishing point P far from its own

vanishing point is considered as a segmentation noise for P. This noise is called segmentation noise, because the distinction between parasite features (i.e. features located near the feature of interest by accident) from the good ones is actually a segmentation problem. Since no information is available to distinguish such features in difficult cases, e.g. the parasite line passes very near P, errors necessarily occur whatever the segmentation method. However, it is possible to take them into account during the process, by modelling them. For example, a complex scene is very likely to bring a lot of segmentation noise. A complex scene results in numerous segments in the image, so that the model should take into account the number of segments. Segments may be expected over all the image area or rather in the upper part of the image (i.e. camera looking at the floor). Thus, it is possible to refine the model with respect to the application. The model chosen here pretends to be general. However the statistical parameters used in chapter 5, the average length  $\ell_{av}$  and the expected value of  $1/\ell^2$ , depend on the type of images processed.

In the following chapters, the image is approximated by a circular disk in order to simplify the model and to make as much use as possible of the isotropy property. The lines are assumed to be distributed at random within the image, i.e. the centroid of the segments AB is uniformly distributed on the image. The length of these segments obeys a probability law of density  $f(\ell)$ , which is assumed to be uncorrelated with the centroid location. Figure 4.4.2.1 shows the distribution of the length of the straight line segments. This density may be approximated by the exponential law with the parameter  $1/(\ell_{av} - \ell_0)$ , denoted by  $f(\ell_0)$ . Let us note that, for numerous segments, the location of the endpoints depends on the threshold chosen in the edge detection stage. Thus, the exponential model is justified by the fact that the density of the end-points given in appendix 7 is roughly constant; therefore it may be approximated by a Poisson model.

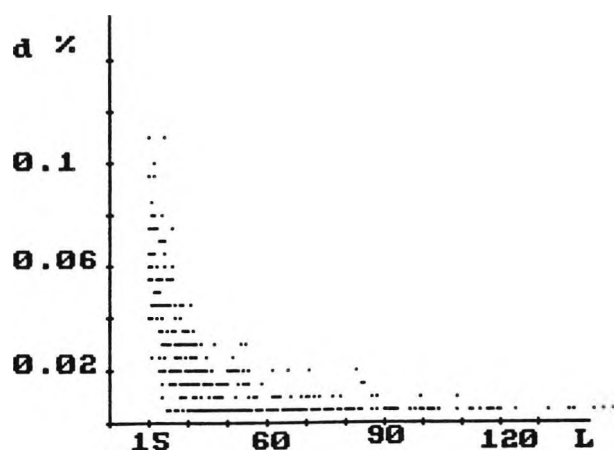


Figure 4.4.2.1 : Density function of the lengths of the segments ( $l > 15$ )

From an image point of view, these assumptions are not exact because of the segments crossing the boundaries of the image. Therefore, for simplicity, the segments are not constrained to lie entirely within the image, except their centroid, so that the assumptions are consistent. It will be seen that  $E(\sigma^2)$  depends on  $E(1/\ell^2)$ . A statistical study of  $1/\ell^2$  on 12 real images demonstrates the validity of the assumption of the independence of  $1/\ell^2$  with the centroid location (figure 4.4.2.2). Besides, in the images studied, the directions are not equiprobable as the vertical lines are always very numerous. In this case, the problem is solved by considering two classes of lines, the vertical lines with their centroid locations and lengths defined as above, and the other lines distributed as described in the previous paragraph.

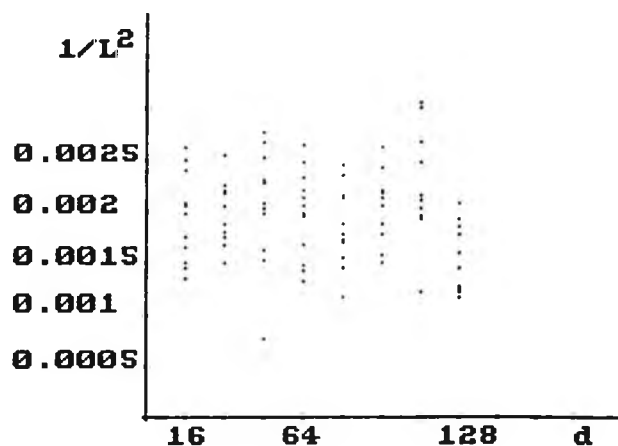


Figure 4.4.2.2 : Average value of  $1/\ell^2$  in function of  $d$ , where  $d$  is the distance from the origine to the line.

Thus a model of the *a priori* distribution of the location of the line segments in the image and of a statistical model of their length is provided to the process. This model is equivalent to a statistical model of the straight line segment parameters. It allows the segmentation error to be taken into account throughout the process.

#### 4.5 CALIBRATION PARAMETERS

The interpretation of the angles between straight line segments requires the knowledge of the intrinsic parameters of the camera; that is to say, the projection  $C$  of the optic centre onto the image, the distance  $f$  between the optic plane and the image plane, and the ratio  $\rho$  of the scales along the  $y$  and  $x$  axes.

Various methods have been developed for calibration (Tsai, 1986; Faugeras, 1986) and have not been investigated thoroughly here, because the method based on the vanishing points, in a similar way to Wei (1988), allowed the method described in chapter 5 to be used.

First, the scale ratio  $\rho$  has been estimated by using the image of a circle parallel to the image plane. The experiment has been done several times without dispersion of the results. The variance of  $\rho$  has been estimated experimentally by the unbiased statistics  $(\sum_1^n s_i)/(n-1)$ , where  $s_i = (\rho_i - \rho)^2$  and  $\rho$  is the average value of  $\rho_i$ .

Once this ratio known, the image of the cube displayed in figure 4.5.1 has been taken. The cube is 1 meter wide, to be consistent with the focusing distance used for the indoor scene images studied. Then, the lines drawn on the cube have been detected and approximated as described in the previous subsections. Eventually, the vanishing points have been extracted with the algorithm described in chapter 5.

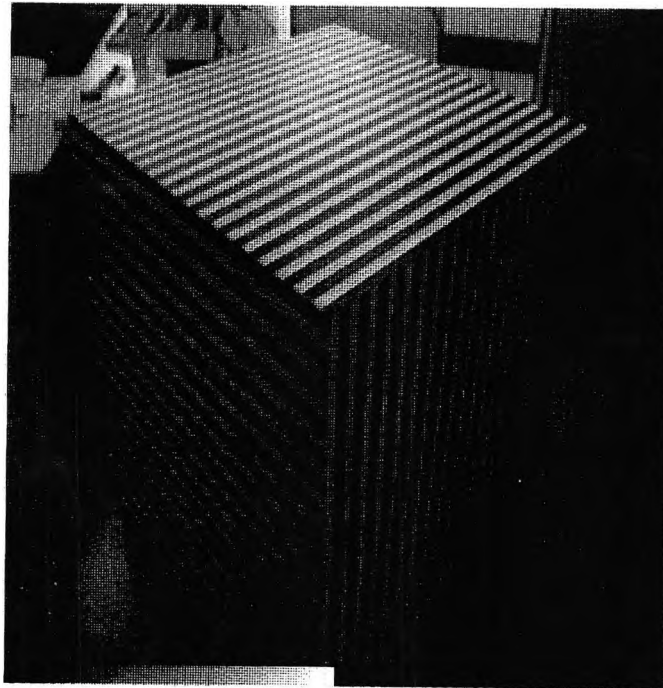


Figure 4.5.1 : Image of the test cube

The vanishing point coordinates are scaled in  $x$  and  $y$  according to the scale ratio  $\rho$ . The projection  $C$  of the optic centre is the orthocentre of the triangle  $(V_1, V_2, V_3)$  and the focal distance is the square root of the scalar product  $\vec{V}_i \vec{C} \cdot \vec{C} \vec{V}_j$ , with  $i \neq j$  (see appendix 1). As  $C$  is defined as the orthocentre of a triangle, this scalar product is constant. The equations are given in appendix 1.

The uncertainty associated with these parameters may be computed by using the covariance matrices associated with  $V_i$ . In order to have the best accuracy, the cube should point a corner towards the camera in order to obtain a nearly equilateral triangle of the vanishing points. If this is not the case, the solution would have a high uncertainty. For example, if  $V_3 = (r_3 \cos \theta_3, r_3 \sin \theta_3)$  is such that  $r_3$  is very large compared with  $r_1$  and  $r_2$ , then  $C(x, y)$  is given by

$$C(x, y) = \begin{cases} x \approx \frac{r_3 \sin \theta_3}{\tan(\theta_{12} - \theta_3)} \\ y \approx -\frac{r_3 \cos \theta_3}{\tan(\theta_{12} - \theta_3)} \end{cases} \quad (4.5.1)$$

## Preprocessing

where  $\theta_{12}$  is the orientation of the line  $V_1V_2$ . Therefore, C is the intersection of line  $(V_1, V_2)$  with the line passing through the origin with the direction  $\theta_3 - \pi/2$ . Actually, the cube is orthogonal, so if  $r_3$  is very large,  $(\theta_{12} - \theta_3) \approx \pi/2$  and the direction  $\theta_3 - \pi/2$  is parallel to the line  $(V_1, V_2)$  and the distance of C from the origin is undetermined. Therefore, the problem is undetermined for  $r_3$  large.

The accuracy of C and f depends not only on the accuracy of the vanishing points but also on the quality of the camera. If some cameras guarantee C with a subpixel accuracy, for the ordinary cameras, C has a position which may vary substantially, depending on the focusing distance. This factor has to be taken into account in the modelling of the parameter uncertainty. Moreover, the main directions in the scene may not be perfectly perpendicular, which may be taken into account by adding a "modelling" uncertainty to the calibration parameters. The cumulation of the sources of error may result in a substantial uncertainty of the parameters ; however, their inaccuracy is not a crucial problem here, as it is taken into account all through the process. On the contrary, it may be considered a source of flexibility of the method.

## 4.6 CONCLUSION

The edges are the features chosen for representing the image because they are assumed to reflect the presence of boundaries in the scene. Most boundaries in the indoor scenes studied are straight line segments, so that most edges in the image are also straight line segments. The scene interpretation process is difficult because of the error of the segment location during the detection, and because of the superposition of all the projected segments on the image. The first error is called measurement uncertainty and may be modelled by a Gaussian law, where the second source of error is called segmentation error and may be modelled by defining a prior statistical model of the features.

First, the error of measurement is minimized by using an optimal

edge detector. The Gaussian, Deriche and Shen edge detectors have been theoretically compared with respect to Canny's criteria and additional criteria, testing the robustness of the detector to corners, multiple edges and thresholding. An improved version of the Shen detector has been proposed. The results have been illustrated on simulated and real images. The Berthod line finder has only been experimentally studied because of lack of time, but necessary improvements need to be made to keep the error of measurement minimal, e.g. a LMS approximation of the straight line segments. This study allows the uncertainty of measurement of the straight line segments to be defined. If a group of straight line segments may be better approximated by an ellipse, then an elliptical arc is considered instead. West and Rosin's(1990) method is used, currently improved by Ellis et al (1991) for taking into account the associated uncertainty.

The segmentation error is due to the fact that features may be connected in the image without being connected in the scene. In order to take it into account in the interpretation process, a prior statistical model of the straight line segments is defined. Thus, the prior probability of two features to be connected by chance can be known.

At this stage, the information available is the set of segments, represented by the list of their endpoints and if required, a set of elliptical arcs represented by the list of their (a,b,c,d,e,f) parameters and their endpoints. The physical parameters of the camera are known with their corresponding uncertainty. Besides, a model of the uncertainty associated with the endpoints is provided, as well as a statistical model of the straight line segment parameters.

The following strategy will be used : a hypothesis on a group of features is tested by using some measure and comparing the real measure with the measure associated with the prior model (i.e. image which represents nothing). If the real measure is comparable to the measure associated with the model, this means that there is no more information in the image than in the prior model ; thereby the hypothesis cannot be



## Preprocessing

validated.

The use of a model for measurement uncertainty and segment parameter distribution allows the choice of optimal (and thereby consistent) parameters throughout the process, avoiding the difficulty of the choice of the right parameters, which increases exponentially with the number of stages of the method. This approach appears to be essential to any high level interpretation process.

## CHAPITRE 5

## DETECTION OF PRINCIPAL DIRECTIONS

## 5.1 INTRODUCTION

The extraction of 3D information from an image is a central problem in computer vision. This chapter is concerned with the interpretation of the perspective in an image. In an indoor scene environment, directions of some features are more likely than others, e.g. for straight lines : the vertical direction and the horizontal directions parallel to the walls. The perspective projection of a set of parallel lines onto an image is a set of lines meeting at a common point, called a vanishing point. The vanishing point coordinates define the direction of the set of lines in the scene relative to the camera coordinate system. The coordinates of the end points of a straight line segment in the image provide the set of possible locations of the end points of the 3D segment. Therefore, once the vanishing point of a 2D straight line segment is known, the corresponding 3D segment is entirely determined except its depth, i.e. except its scale. As mentioned in chapter 3, scale indetermination is inherent to monocular vision.

Section 5.2 describes the detection of potential vanishing points which are the common intersection points of a number of lines. Lines are accumulated in an accumulator space, the peaks of which represent the points looked for. A new accumulator space formulation for the whole line accumulation approach is described which fulfills an important property for robustness: the constancy of the detectability of a vanishing point whatever its location in the image plane. This accumulator space is compact and isotropic. It is compared with the Gaussian sphere accumulator space.

Section 5.3 describes the classification of the lines with each vanishing point candidate and the scoring of the classes obtained. The

## Detection of principal directions

classification is based on a likelihood ratio test (LR test). The vanishing point coordinates are recalculated by a Kalman filter which also provides its uncertainty neighbourhood, assuming the classification is correct. The result of the likelihood ratio is used for scoring the classification.

Because of the complexity of the scene (e.g. many lines are not parallel to any principal direction), because of the relatively high uncertainty of the line parameters after the line finder, and because of the possible ambiguities with corners, many false candidates are found, the scores of which may be good (e.g. for corners). An additional criterion is provided by perpendicularity. The principal directions are supposed to be perpendicular, which again is the case in many indoor scenes. The sets of two or three vanishing point candidates corresponding to perpendicular directions are found by a likelihood process and scored. This additional filter is very powerful as the vertical direction has been proved to always correspond to the highest score, which constrains the vanishing points corresponding to the horizontal directions to lie on a line, i.e. the horizon, within an uncertainty. The detection of the perpendicular directions is the subject of section 5.4.

Section 5.5 describes the information extracted from the image and presents the results obtained from a set of images of indoor scenes of a power plant.

## 5.2 DETECTION OF THE VANISHING POINTS

### 5.2.1 Previous work

The search for vanishing points consists of finding a small neighbourhood in the image plane intersected by a sufficient number of straight lines. In order to reduce the search to a bounded closed set, Barnard (1983) proposed projecting the lines of the image onto a Gaussian sphere centred onto the optic centre. The projection of points

onto a Gaussian sphere is equivalent to a resampling of the image plane, and the number of resampled straight lines crossing each cell is the result of the accumulation (Hough paradigm). Peaks of the accumulator space correspond to potential vanishing points. The form and the size of the cells depend on the sampling used in the accumulator space. Barnard (1983) uses spherical coordinates (elevation, azimuth) ; unfortunately they are irregular and different in the x and y directions. Quan and Mohr (1989) use the same method with a dichotomic approach : the sphere is sampled from a coarse resolution to a fine resolution in order to reduce the number of studied cells. The lines are classified by looking for vanishing points. Once a vanishing point is found with its associated lines, the lines are eliminated and the algorithm is performed again. Dickson (1989) proposed a triangular sampling of the Gaussian sphere which is very attractive because it is isotropic; however the computational efficiency has yet to be proved.

Magee and Aggarwal (1984) accumulate the projection of the intersection points of all pairs of straight lines in the image onto the Gaussian sphere. The accumulation is achieved using the arc distance between two points which leads to an isotropic search.

The isotropy of the search for vanishing points using the accumulator spaces previously described can be ensured by an additional cost of complexity and only guarantees the isotropy in  $\theta$ , not necessarily in  $r$  ( $(r, \theta)$  are the polar coordinates). As a result the probability of detecting a vanishing point depends on its distance from the centre of the image. The reason for this dependence with  $r$  is due to the increasing uncertainty of the vanishing point location with  $r$ .

Kanatani (1989) tests hypotheses on parallelism of three or more straight lines by using the projection onto the Gaussian sphere. The uncertainty of the line parameters is taken into account by using a threshold in a concurrency test which depends on the uncertainty of the intersection point coordinates. We (Brillault, 1989) have previously proposed to solve the problem by resampling the accumulator space such

that the uncertainty remains approximately constant over the space. The variations of the uncertainty of the intersection point of a pair of lines is statistically estimated, but for  $r$  fixed the dispersion of this uncertainty is large, besides which the solution found is not really isotropic in  $\theta$ . It will be seen that a criterion based on distance between intersection points does not seem appropriate to a problem initially defined as the minimization of distances between a point and lines, whatever the representation used.

A number of methods for the detection of the vanishing points have been developed, but for all of them the detectability of a vanishing point depends on its location in the image. The method proposed here is based on the accumulation principle but uses a new mapping that ensures the same detectability of the vanishing point over the space. The accumulator space is therefore isotropic ; moreover it is bounded and so does not increase the complexity of the detection.

### 5.2.2 Accumulator space

Under perspective transformation, parallel lines in a 3-D scene are projected onto concurrent lines in the 2-D image. Ideally, in a man made environment many lines are parallel, e.g. the edges of a wall and a door frame. In practice, this is only approximately true. However, parallelism is a useful concept for representing the scene, e.g. a door frame may be represented as a rectangle.

Let  $P$  be the vanishing point to be determined, which is the common point of the ideal projection of the ideal parallel lines onto the image. The location of  $P$  in the image depends on the viewpoint. The search for  $P$  by accumulation is equivalent to counting the real lines in the image passing through a neighbourhood of  $P$ .

A line ( $L$ ) defined by a line segment in the image is assumed to have its vanishing point in  $P$  if  $D(L,P) < r(P)$ , where  $D(L,P)$  is the distance between the line ( $L$ ) and the point  $P$  and  $r(P)$  is the radius of the chosen neighbourhood  $N$  of  $P$  in the perpendicular direction of ( $L$ ).

The constant quality for the detection of  $P$  irrespective of its location in the image plane increases the robustness and meaning of the detection. The value of the radius  $r(P)$  of the neighbourhood  $N$  defined above should be proportional to the accuracy with which the line  $(L)$  is known in the vicinity of  $P$ , that is to say proportional to the uncertainty of the distance  $D(L,P)$  from  $L$  to  $P$ . The neighbourhood  $N$  has a corresponding neighbourhood  $N'$  in the accumulator space. For practical reasons this neighbourhood  $N'$  should be constant and have a simple shape.

Let  $\sigma$  be the uncertainty of  $D(L,P)$  ; the problem is to find a transformation  $T$  from the image plane to an accumulator space such that the expected value of  $\sigma' = T(\sigma)$  remains constant over the accumulator space. This transformation  $T$  will completely define the new accumulator space (figure 5.2.2.1).

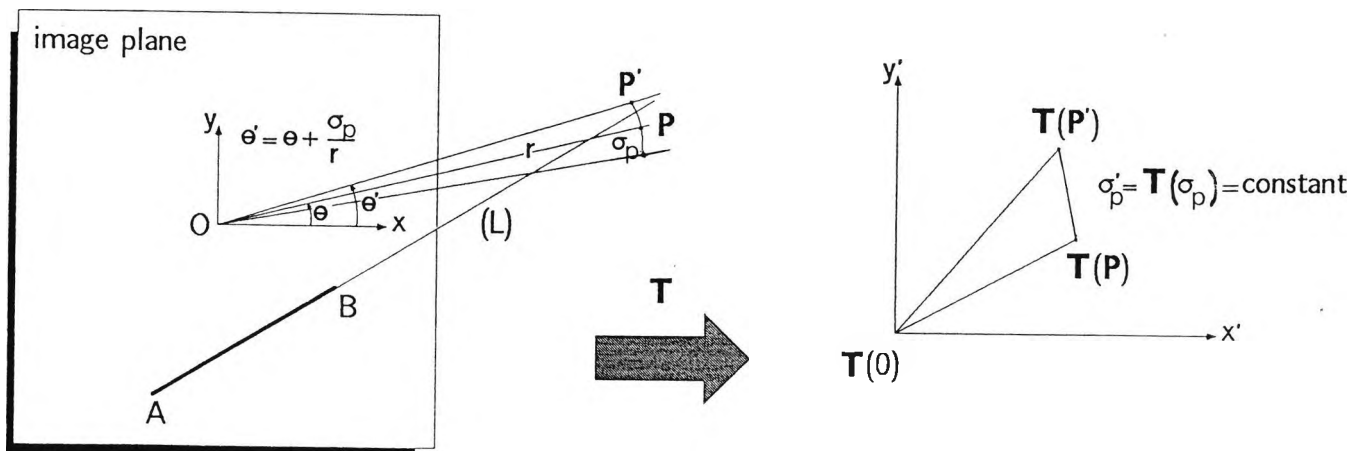


Figure 5.2.2.1: Transformation from the image plane to the accumulator space

*Expected uncertainty of a vanishing point*

The uncertainty of the end points of the segment  $AB$  results in uncertainty  $\sigma$  of its distance to a point  $P$ . Polar coordinates are used, with pole  $O$  and axis  $Ox$ . Let  $O$  be the centre of the image,  $r$  the distance from  $O$  to  $P$ ,  $\theta$  the polar angle of  $OP$ ,  $\alpha$  the polar angle of

(L), and  $d$  its distance from  $O$ .  $Q'$  is the intersection of the line (L) and the circle (C) with centre  $O$  passing through  $P$  and  $\alpha_1$  is the polar angle of  $Q'$  (in the ideal case  $\alpha_1 = \theta$ , see figure 5.2.2.2). Let  $\sigma_p$  be the uncertainty of  $P$  along (C). The variance of  $\sigma_p$  over all the lines having  $P$  for vanishing point is equal to the variance of the error of  $P$  in the tangential direction,  $(r \partial\alpha_1)^2$ . If the line (L) is fixed, the uncertainty  $\sigma_p$  is equal to

$$\sigma_p^2 = r^2 E(\partial\alpha_1^2) \tag{5.2.2.1}$$

where  $E(\partial\alpha_1^2)$  is the variance of the noise on  $\alpha_1$ . Using the polar equation of the line it follows that

$$r \cos(\alpha_1 - \alpha) = d \tag{5.2.2.2}$$

This equation is derived to provide the expression of  $\partial(\alpha_1)$ , when  $\sqrt{r^2 - d^2}$  is not too small,

$$\partial\alpha_1 = \partial\alpha - \frac{\partial d}{r \sin(\alpha_1 - \alpha)} = \partial\alpha - \frac{\partial d}{\sqrt{r^2 - d^2}} \tag{5.2.2.3}$$

Let  $\ell$  be the length of the segment  $AB$  and  $b$  the distance between the centroid  $G$  of the segment and the projection  $O'$  of  $O$  onto (L), when  $\sqrt{r^2 - d^2}$  is large enough, the value of  $\sigma_p^2$  is given by (see appendix 5, eq. A5.4, A5.5 and A5.6)

$$\sigma_p^2 = \left( \frac{2r^2}{\ell^2} + \frac{\frac{1}{2} + \frac{2b^2}{\ell^2}}{1 - \frac{d^2}{r^2}} + 4 \frac{br}{\ell^2 \sqrt{1 - \frac{d^2}{r^2}}} \right) V(u) \tag{5.2.2.4}$$

where  $V(u)$  is defined in section 4.4.1.

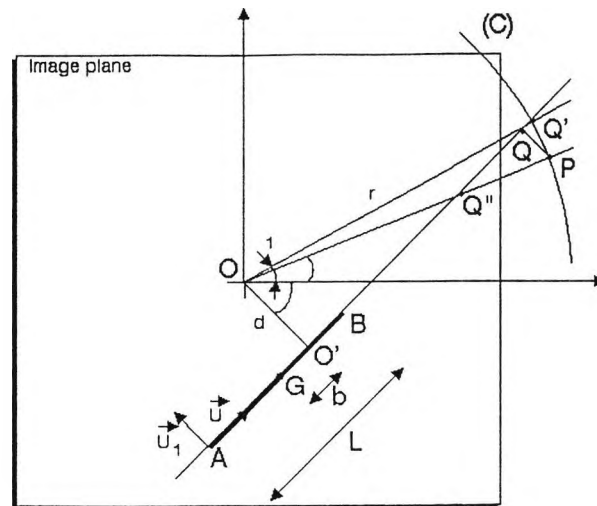


Figure 5.2.2.2: Line (L) passing near the point P

Using the statistical model defined in chapter 4, the expected value of  $\sigma_p$  may be calculated. It has been demonstrated that the length of the segments may be assumed independent of the centroid location. However, if the length  $L$  of a 3D segment  $S$  is assumed fixed, the length  $\ell$  of its projection  $[A,B]$  onto the image depends on the distance of  $S$  from the camera : the further  $S$  from the camera, the closer  $[A,B]$  to the vanishing point and the smaller  $\ell$ . It is shown appendix 1 that the distance between a segment and its vanishing point is constrained by eq. A1.11 :

$$\frac{GQ'}{\ell} > \frac{D_m}{D_f}, \quad (5.2.2.6)$$

where  $D_f$  is the depth of field and  $D_m$  the focussing distance. Figure 5.2.2.3 demonstrates that this constraint has a very small effect on the expected value of  $(1/\ell^2)$ . This is not surprising because the constraint filters the small segments which play a preponderant part in  $E(1/\ell^2)$ . In fact, the distribution of the length is mainly due to the line finder algorithm and little to the viewpoint.



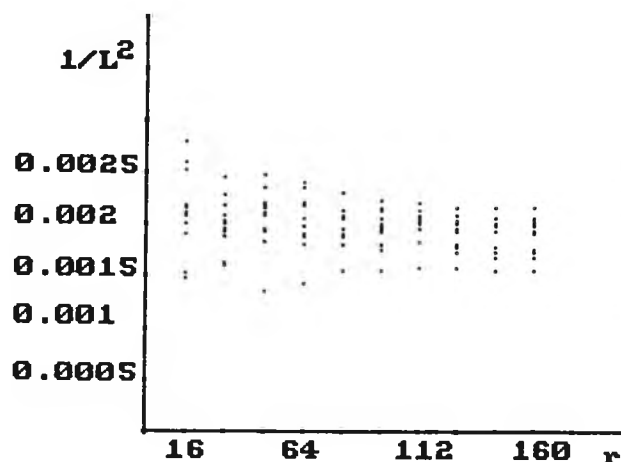


Figure 5.2.2.3 : Average value of  $1/\ell^2$  in function of  $d$  performed over the lines satisfying the constraint 5.2.2.6 with  $D_m/D_r = 1.3$ . (12 images of 6 different scenes have been processed)

Let "a" be  $E(1/\ell^2)$ . Using the independence assumptions between  $1/\ell^2$  and  $(b, d, r)$  and the fact that  $E(b|d)=0$  for symmetry reasons, the expected value of  $\sigma_p^2$  over the set of segments AB at a distance  $d$  from the origin is equal to (from (5.2.2.4))

$$E(\sigma_p^2|d) = (2ar^2 + \frac{1}{2(1 - \frac{d^2}{r^2})} + \frac{2a E(b^2|d)}{1 - \frac{d^2}{r^2}}) V(u). \quad (5.2.2.7)$$

This relation ignores the constraint (5.2.2.6), but still  $E(b^2/\ell^2|d)$  gives more weight to large values of  $b$  and small values of  $\ell$  corresponding to the lines satisfying (5.2.2.6), which justifies (5.2.2.7). Remark: The previous argument only holds when the minimal value allowed for  $\ell$  is small enough.

The distance  $d$  of a line is smaller or equal to  $r$ , within the uncertainty of  $r$ . When  $d$  is smaller than  $r$

$$\frac{1}{(1 - \frac{d^2}{r^2})} = 1 + \epsilon(1/r^2),$$

where  $\epsilon(1/r^2)$  is infinitesimal with  $1/r^2$ .

The centroid  $G$  is located at random on a chord of the image disk (defined in chapter 4), the length of which is equal to  $2R \sqrt{1-d^2/R^2}$ , then (appendix 4),

$$E(b^2|d) = \frac{1}{3} R^2 \left(1 - \frac{d^2}{R^2}\right) \text{ and } E(d^2) = \frac{R^2}{4}$$

Using the model of the measurement uncertainty defined in section 4.4

$$E(\sigma_p^2|d) = \left(2ar^2 + \frac{1}{2} + \frac{2a R^2}{3}\right) \sigma_0^2. \quad (5.2.2.8)$$

A line having  $P$  for its vanishing point satisfies  $d \leq r$ . The eq. (5.2.2.8) has been found using (5.2.2.4) under the assumption that  $\sqrt{r^2 - d^2}$  is not too small.

If  $r > R$ ,  $d$  is always smaller than  $r$ , and (5.2.2.8) holds and is independent of  $d$ . When  $r \leq R$  and  $d = r$ , the definition of  $\sigma_p$  makes no sense, as the line ( $L$ ) may cross once, twice or not at all, the circle ( $C$ ) in the vicinity of  $P$ . When  $P$  is far enough from the origin, a very small proportion of lines are likely to be such that  $d = r$ , and (5.2.2.8) is nearly always valid. If  $r$  is very small, the radial uncertainty of  $P$  makes no sense. The uncertainty neighbourhood of  $P$  is a circle with a constant radius  $\sigma$ , where  $\sigma$  is defined as the root of the expected value of  $PQ^2$ . It is shown in appendix 5 that near the origin

$$\sigma^2 = \left(\frac{1}{2} + \frac{2a R^2}{3}\right) \sigma_0^2. \quad (5.2.2.9)$$

Therefore from (5.2.2.8) and (5.2.2.9) and the previous remarks, the expected value of the variance of  $P$  may be approximated by

## Detection of principal directions

$$\begin{aligned} \text{if } r \text{ is large enough } E(\sigma_p^2) &= (2ar^2 + c) \sigma_0^2, \\ \text{if } r \text{ is small } E(\sigma^2) &= c \sigma_0^2, \end{aligned} \quad (5.2.2.10)$$

where  $a = E(1/\ell^2)$  and  $c = 1/2 + 2a R^2/3$ .

The value of the uncertainty of the line location near P depends on the parameters of the corresponding segment in the image. It has been shown that it is possible to express the expected value of this uncertainty by using statistics of the line segment parameters. The expression for the expected values depends only on the distance of the vanishing point from the image centre. This result allows almost the same detectability to be guaranteed for any vanishing point, by resampling the image plane proportionally to the expected value of the uncertainty.

### *Transformation from the image plane to the accumulator space*

In order to simplify the notation,  $\bar{\sigma}$  represents  $\sqrt{E(\sigma^2)}$  in the following.

If the transformation T exists it is defined by

$$P(r, \theta) \xrightarrow{T} P'(x'(r, \theta), y'(r, \theta)),$$

$x', y'$  being the coordinates of the cell of the accumulator space.

Let  $Q'$  be the point of the circle (C) located at the distance  $\sigma_p$  from P,

then

$$T(Q') = T\left(P\left(r, \theta + \frac{\bar{\sigma}_p}{r}\right)\right) = P'\left(x'\left(r, \theta + \frac{\bar{\sigma}_p}{r}\right), y'\left(r, \theta + \frac{\bar{\sigma}_p}{r}\right)\right)$$

In the accumulator space the uncertainty  $\sigma'_p$  corresponding to the uncertainty  $\sigma_p$  in the image, is equal to the distance of T(P) to T(Q').

Thus

$$\begin{aligned}\bar{\sigma}_p'^2 &= (P'(x'(r, \theta + \frac{\bar{\sigma}_p'}{r}), y'(r, \theta + \frac{\bar{\sigma}_p'}{r})) - P'(x'(r, \theta), y'(r, \theta)))^2 \\ &= (x'(r, \theta + \frac{\bar{\sigma}_p'}{r}) - x'(r, \theta))^2 + (y'(r, \theta + \frac{\bar{\sigma}_p'}{r}) - y'(r, \theta))^2.\end{aligned}$$

After linearization it leads to

$$\bar{\sigma}_p'^2 = \left(\frac{\partial x'}{\partial \theta}\right)^2 \frac{\bar{\sigma}_p'^2}{r^2} + \left(\frac{\partial y'}{\partial \theta}\right)^2 \frac{\bar{\sigma}_p'^2}{r^2}.$$

Thus

$$\left(\frac{\partial x'}{\partial \theta} + \frac{\partial y'}{\partial \theta}\right) = \bar{\sigma}_p'^2 \frac{r^2}{\bar{\sigma}_p'^2}. \quad (5.2.2.11)$$

One solution of (5.2.2.11) is

$$\begin{cases} y' = \bar{\sigma}_p' r \theta / \bar{\sigma}_p', \\ x' = x'(r). \end{cases}$$

$x'(r)$  is independent of  $\theta$  and may be any bijection from  $[0, +\infty]$  to a bounded interval. For simplicity,  $x'$  has been chosen such that

$$y' = \theta \frac{x'}{k},$$

where  $k$  is a scale factor determined by the expected resolution.

Therefore

$$\begin{aligned}x' &= \frac{\bar{\sigma}_p'}{\sigma_0} \frac{k r}{\sqrt{2a r^2 + c}}, \\ y' &= \frac{\bar{\sigma}_p'}{\sigma_0} \frac{r \theta}{\sqrt{2a r^2 + c}}.\end{aligned} \quad (5.2.2.12)$$

The number of straight lines in the image passing near the point  $P(r_0, \theta_0)$  is represented by the number of curves crossing the line

## Detection of principal directions

parallel to  $y'$  axis,  $y' = y'(r_0, \theta_0)$ , in the range  $[P' - \bar{\sigma}'_p, P' + \bar{\sigma}'_p]$  (see figure 5.2.2.4). The counting of these lines is performed by accumulating curves dilated by a vertical kernel with a half-width equal to  $\bar{\sigma}'_p$ . The accumulation of dilated curves prevent the same line from being counted twice.

A line passing near  $P$  such that  $d \in ]r, r + \sigma[$  does not cross the line  $y' = y'(r_0, \theta_0)$ . When  $P$  is far from the origin such a configuration is unlikely, but when  $r$  is small its probability increases. The problem is solved if the  $x'$  resolution around  $P'$  is less than or equal to  $\bar{\sigma}$ . This could be done by choosing the constant  $k$  in eq. 5.2.2.12 such that  $x'(r)$  has a resolution equal to  $\bar{\sigma}$  for small  $r$ , but the resolution of  $x'$  has to be consistent with the resolution of  $y'$  (e.g. 2 curves crossing at  $x' = (2n+1)/2$  must cross the same uncertainty neighbourhood either at  $x' = n$  or at  $x' = n+1$ ). The two constraints on  $k$  are incompatible for large  $r$ . The constant  $k$  is chosen to ensure the consistency with  $y'$  resolution and the constraint for the small value of  $r$  is fulfilled by a pre-accumulation stage in the  $(d, \alpha)$  parameter space (see later).

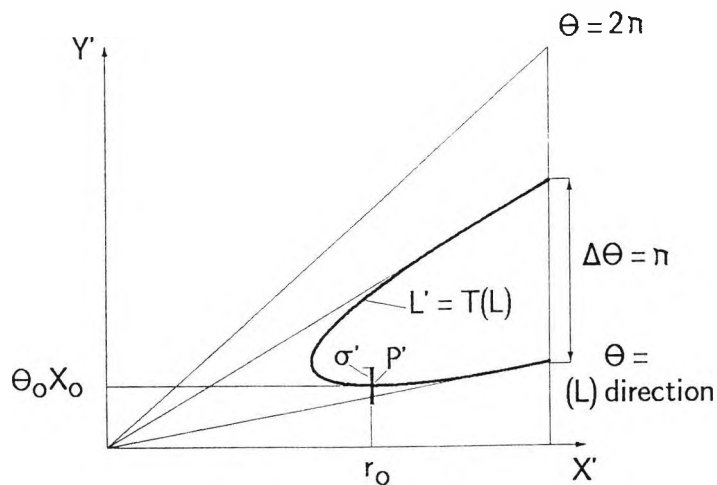


Figure 5.2.2.4 :  $(x'(r), y'(r, \theta))$  accumulation space.

So, a transformation  $T$  has been found such that any point in the image plane is simply transformed to a single point in the accumulator space and the search for vanishing points is reduced to the search for local maxima in the accumulator space. The sampling is isotropic and homogeneous with respect to the uncertainty criterion. Moreover the

accumulator space is bounded.

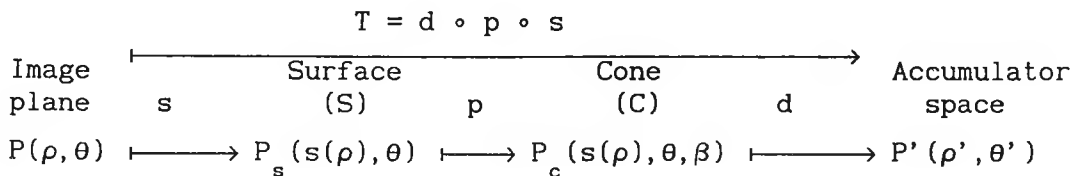
*Geometrical interpretation*

The accumulator space  $(x', y')$  described previously is isotropic in  $\theta$  by definition. In the continuous plane, it may be represented as a sector of a disk. In this sector, a point  $P'$  with the polar coordinates  $(\rho', \theta')$  corresponds to the point  $P(\rho, \theta)$  in the image plane and is defined by using (5.2.2.12)

$$\rho' = x'(\rho) \text{ and } \theta' = \frac{\theta}{k}.$$

In this sector the neighbourhood of uncertainty is a circular arc with arc length  $\bar{\sigma}_p'$ .

Let the origin of the coordinate system be  $O$  and the  $z$  axis be the optical axis. Thus, the transformation defined is equivalent to a projection onto a surface  $(S)$  of revolution with vertex  $O$  and axis  $Oz$ , followed by a projection onto a cone  $(C)$  with vertex  $O$ , axis  $Oz$  and angle  $\beta$  (see figure 5.2.2.5), which is eventually developed onto a plane (to give a sector, i.e. the accumulator space), so that



$P_s$  is the projection of  $P(\rho \cos \theta, \rho \sin \theta, 0)$  on the surface  $(S)$  in a direction parallel with  $Oz$ ;  $P_s$  has the Cartesian coordinates  $(\rho \cos \theta, \rho \sin \theta, s(\rho))$ .  $P_c$  is the projection of  $P_s$  on the cone, in a perpendicular direction with  $Oz$ ;  $P_c$  has the Cartesian coordinates  $(s(\rho) \tan \beta \cos \theta, s(\rho) \tan \beta \sin \theta, s(\rho))$ .  $P'$  is the image of  $P_c$  when the cone is developed;  $P'$  has the polar coordinates  $(\rho', \theta')$  in the accumulator space, such that

$$s(\rho) = \rho' \cos \beta \text{ and } \theta' = \frac{\theta s(\rho) \tan \beta}{\rho'} = \theta \sin \beta$$

## Detection of principal directions

It gives

$$\sin \beta = \frac{1}{k} \text{ and } s(\rho) = \frac{x'(\rho)}{k} \sqrt{k^2 - 1}$$

Remark : the constraint  $k > 1$  which appears above is introduced by the geometric interpretation, but it is possible to show that it is not a real constraint (however, it is verified when the consistency of the resolutions is ensured - see later-).

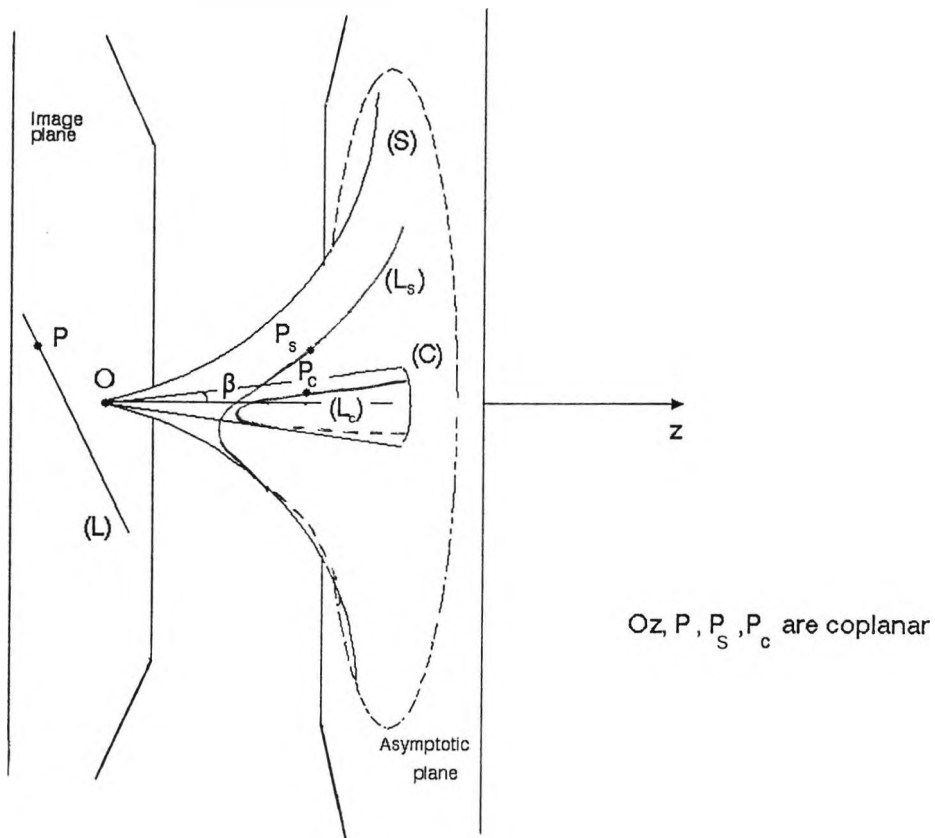


Figure 5.2.2.5 : Geometric interpretation of the transformation T

### *Pre-accumulation stage*

Several straight line segments in the image may be aligned, often not by chance but because they belong to the same structure in the scene, say a window frame. From the point of view of the interpretation of the perspective, only the line is significant. Therefore, before the accumulation stage, aligned segments are grouped to form one line .

The lines are first accumulated with respect to  $d$  and  $\alpha$ , the resolution of which is determined by  $\sqrt{E(\partial d^2)}$  and  $\sqrt{E(\partial \alpha^2)}$  averaged over all possible  $b$  (see appendix 5 (A5.4), (A5.5)), and for each cell  $(d, \alpha)$  the best estimates of the line parameters are computed. Only one line per cell  $(d, \alpha)$  is accumulated in the accumulator space.

When  $r$  is small,  $d$  is also small and the expected value of  $E(\partial d^2)$  over all  $b$  is equal to  $c$ , i.e. to  $\bar{\sigma}^2$ . This means that all the lines such that  $d \in [r, r + \bar{\sigma}[$  and  $\alpha \in \theta$ , are accumulated in the same cell  $(d, \alpha)$ . Therefore, for  $r$  small, the resolution in  $x'$  is equal to  $\sigma$  around  $x'(r)$  and the lines satisfying  $d = r$  crosses the uncertainty neighbourhood around  $P'$  defined above.

The preaccumulation stage avoids a weighting process, e.g. weighting lines by their length (Quan and Mohr, 1989), and thereby increases the significance of the peaks of the accumulator space  $(x', y')$  described above, since the peak value represents the number  $N$  of directions meeting at the same point. For instance, if  $N=3$  the point may be a corner, if  $N>3$  the point is likely to be a vanishing point.

#### *Uncertainty of the vanishing point*

It may be important to have a first evaluation of the uncertainty of the vanishing point location, once detected. Let  $\sigma_r^2$  be the variance of the distance  $Q''P$ , where  $Q''$  is the intersection point of the line  $(L)$  and the line  $(OP)$  (see figure 5.2.2.3); it is equal to

$$\sigma_r^2 = E(Q''P^2) = r^2 \frac{E(QP^2)}{d^2} = \frac{r^2 \sigma^2}{d^2} \quad (5.2.2.13)$$

The line such that  $d = 0$  does not provide any information about  $r$ , and the corresponding uncertainty is infinite, so is the expected value of  $\sigma_r^2$  over all possible  $d$  for  $r$  large enough. But here, the peak in the accumulator space has already been found, which means that the shape of the distribution of  $Q''P^2$  around  $P$  no longer matters. The information about  $r$  is provided by lines passing far from the origin. At least three lines with different directions should cross the neighbourhood of



## Detection of principal directions

P to give it meaning as a possible vanishing point, and therefore at least two lines passing far from the origin, typically at  $\sqrt{E(d^2)}$ . Therefore, for large  $r$ , using the expression of  $E(d^2)$  found in appendix 4

$$\hat{\sigma}_r^2 = \frac{1}{2} \frac{r \bar{\sigma}^2}{E(d^2)} = \frac{2r^2(2ar^2+c)\sigma_0^2}{R^2}, \quad (5.2.2.14)$$

where  $\hat{\sigma}_r$  is an estimate of the radial uncertainty on P.

If  $r$  is small, using the same considerations as for  $\bar{\sigma}$ ,  $\hat{\sigma}_r^2$  can be replaced by the expression of  $\sigma^2$ , say  $c \sigma_0^2$ . The expression (5.2.2.14) for  $\hat{\sigma}_r$  gives a very coarse idea of the vanishing point uncertainty and it should be considered as a temporary information. This approximation is refined by using a Kalman filter applied to each line classified with the point P.

For the resolution to be consistent with this uncertainty value, the derivative of  $r$  relative to  $x'$  must be lower than  $\hat{\sigma}_r$ . The sampling is not regular according to  $\hat{\sigma}_r$ , which does not matter as  $\hat{\sigma}_r$  does not affect the line-counting in the whole line accumulation approach described above. This would not have been the case if only the intersection points had been accumulated (see figure 5.2.2.6).

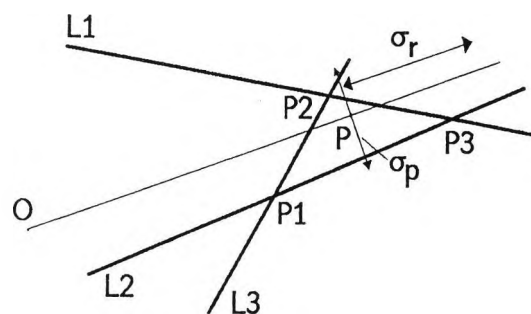


Figure 5.2.2.6 : Comparison between a "whole line" accumulation approach and the "intersection point" accumulation approach

Resampling according to  $\hat{\sigma}_r$

It is possible to choose  $x'(r)$  in order to have a sampling approximately regular according to  $\hat{\sigma}_r$ . Albeit it slightly increases the complexity of the formalism, it has been done to reduce the size of the accumulator space and thereby the number of peaks to process afterwards. However, the consistency between  $x'$  and  $y'$  resolution is not as good as previously, which means a line may pass nearby a point without being accumulated to it, (when the slope of the corresponding accumulated curve is more than  $2\bar{\sigma}'_p$ ). Actually it is unlikely to happen and practically it does not seem to jeopardize the goodness of the results.

Returning to the solution of (5.2.2.11)

$$\begin{cases} y' = \bar{\sigma}'_p r \theta / \bar{\sigma}'_p \\ x' = x'(r), \end{cases}$$

The regularity of the sampling of  $x'$  according to  $\hat{\sigma}_r$  may be written : for small  $r$  (typically  $r < R$ ) :

$$r < R \quad \frac{\partial x'}{\partial r} = \frac{\hat{\sigma}'_0}{\hat{\sigma}'_r} \Rightarrow x'(r) = \frac{\hat{\sigma}'_0}{\sigma_0} \int \frac{dr}{\sqrt{2ar^2+c}},$$

and for large  $r$  :

$$r \geq R \quad \frac{\partial x'}{\partial r} = \frac{\hat{\sigma}'_{x',R}}{\hat{\sigma}'_r} \Rightarrow x'(r) = \frac{\hat{\sigma}'_{x',R}}{2\sigma_0} \int \frac{dr}{r \sqrt{2ar^2+c}}.$$

where  $\hat{\sigma}'_0$  and  $\hat{\sigma}'_{x',R}$  are coefficients defining the resolution in  $x'$ . They are chosen to ensure the continuity of  $x'$  and its first derivative. The solution is given by

$$r < R \quad x'(r) = \frac{\hat{\sigma}'_0}{\sigma_0} \sqrt{\frac{1}{2a}} \operatorname{Argsh} \left( \sqrt{\frac{2a}{c}} r \right),$$

(5.2.2.15)

$$r \geq R \quad x'(r) = \frac{\hat{\sigma}'_{x,R}}{4 \sigma_0 \sqrt{c}} \operatorname{Ln} \left( \frac{\sqrt{2ar^2+c} - \sqrt{c}}{\sqrt{2ar^2+c} + \sqrt{c}} \right) + x'_{\max}$$

The continuity of the derivative is ensured by setting  $\hat{\sigma}'_0 = \hat{\sigma}'_x / 2$ . The value of  $\hat{\sigma}'_0$  fixes the resolution in  $x'$ . A trade-off between the consistency in the resolution of  $x'$  and  $y'$  and the size of the accumulator space should be found. For example a value of  $\hat{\sigma}'_0$  equal to 2.75 leads in the particular case of the images studied here to an accumulator space size equal to  $100 \times 100$ , by contrast with  $256 \times 256$  in the case of the original sampling (5.2.2.12).

The equation giving  $y'$  is clearly unchanged

$$y' = \frac{\bar{\sigma}'_p}{\sigma_0} \frac{r \theta}{\sqrt{2a r^2 + c}}$$

Thus, another sampling has been defined for the accumulator space, which is more regular according to  $\hat{\sigma}'_r$  than the sampling (5.2.2.12), but is less consistent with  $y'$  resolution. The size of the accumulator space has been reduced, without affecting the goodness of the results. The complexity have been slightly increased (see figure 5.2.2.7).

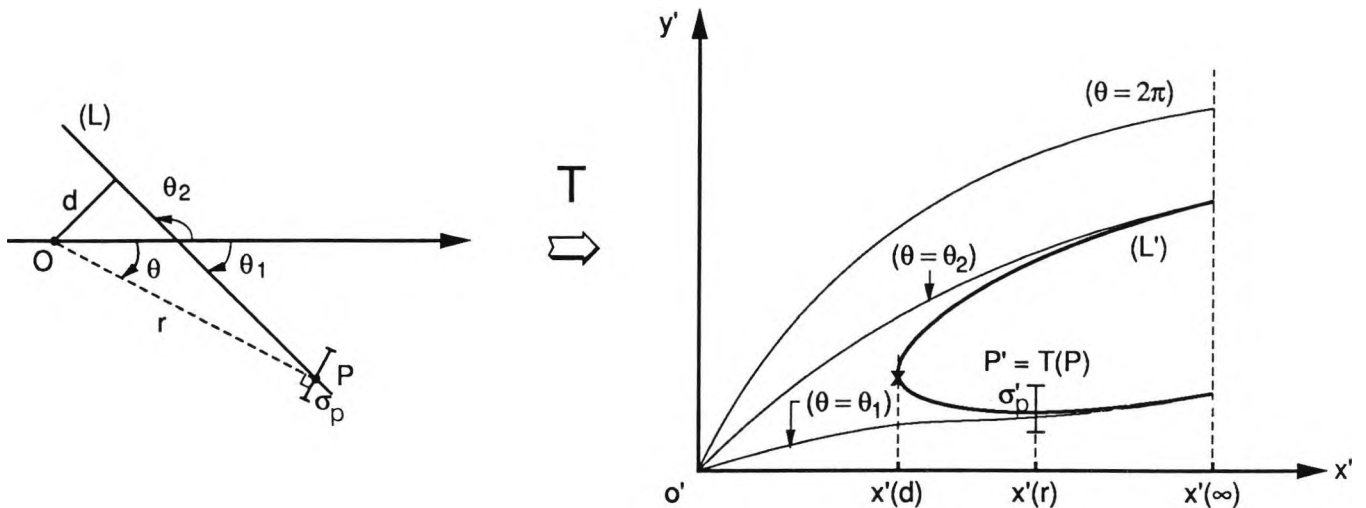


Figure 5.2.2.7 : Accumulator space corresponding to the sampling 5.2.2.15.

### 5.2.3 Comparison with other accumulator spaces

#### *Comparison with intersection points accumulation method*

Figure (5.2.2.5) demonstrates that all the lines around  $Q$  are counted independently of the value of  $\sigma_r$ , but the counting of the corresponding intersection points should take  $\sigma_r$  into account. The equation (5.2.2.13) shows that for  $r$  fixed, the disparity of  $\sigma_r$  is very large, and thereby the disparity of the error of the distance between intersection points is very large too. Whatever the representation used, the disparity of the error of the intersection points is very much larger than the disparity of  $\sigma^2$ , which means that the concurrency test based on the intersection point accumulation approach has a poorer significance than the test used in the whole line accumulation approach.

The best estimate of the intersection point of a set of straight lines in the LMS sense minimises  $\Sigma(d_i/\sigma_i^2)^2$ , where  $d_i$  is the distance from the  $i^{\text{th}}$  line to the intersection point. This criterion may be written in terms of intersection point coordinates and the associated covariance matrix (it is complicated because the intersection points are not independent). Because of the disparity of

## Detection of principal directions

the variance along the  $r$  direction, an accumulator space based on the intersection point accumulation should necessarily be associated with a weighting process using coefficients equal to the inverse of this variance, so that only significant intersection points would be taken into account. If the dependence of the intersection points is ignored, the value of the peaks obtained would approximately represent the inverse of the corresponding intersection point variance along  $r$ . This value is not significant because of the disparity of this value and because it provides no information about the number of lines or directions meeting at this point.

Albeit the set of intersection points and their associated uncertainty matrix theoretically contains sufficient information about the best estimate of the intersection point of a set of straight lines, the corresponding tests have a poor significance. Besides, the information is redundant ( $5n(n+1)/2$  data in contrast with  $5n$  data in the minimal case), i.e. the data are not independent, which is a serious difficulty when reasoning with uncertainty.

### *Comparison with Gaussian sphere method*

If the constraint of the constant detectability is not fulfilled the detection of a vanishing point, i.e. a main direction in the scene, depends on the viewpoint and the orientation of the image. It is important to evaluate the behaviour of the more commonly used accumulator space, the Gaussian sphere, with respect to this constraint. In order to be general, the Gaussian sphere is placed on the optical axis at any distance  $h$  from the principal point (its radius has no importance).

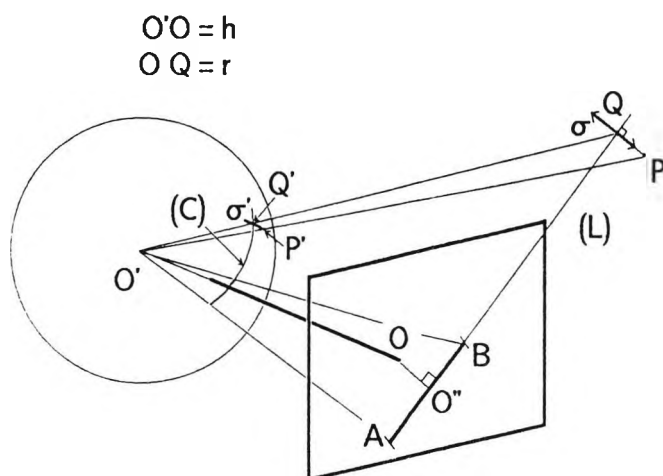


Figure 5.2.3.1: Projection of an image segment onto the Gaussian sphere

Let  $r$  be the distance from  $Q$  to  $O$ . The projection process transforms  $L$  into a big circle  $C$ ,  $P$  into  $P'$  and  $Q$  into  $Q'$ .  $P$  is assumed to be near  $Q$ , so that the distance  $D'$  from  $C$  to  $P'$  can be approximated to  $D' = D(P', Q')$ , and its uncertainty  $\sigma'$  approximated by the projection of the uncertainty  $\sigma$ . Let  $h$  be the distance between  $O$  and the centre of the Gaussian sphere, then

$$\frac{\sigma'}{\sigma} \approx \frac{O'Q'}{O'Q} = \frac{\rho}{\sqrt{r^2 + h^2}},$$

where  $\rho$  is the radius of the sphere. The constraint  $\sigma' = \text{constant}$  implies that  $(r^2 + h^2)$  is proportional to  $\sigma^2$ . It is shown in appendix 6 that it is possible to approximate  $\sigma^2$  for any  $r$  by

$$\sigma^2 = (2a r^2 + c) \sigma_0^2,$$

which leads to, when applied to the Gaussian sphere

$$h = \frac{c}{2a}. \quad (5.2.3.1)$$

This relation means that there exists a value for  $h$  such that the uncertainty is the same at the origin and at infinity. A large value of  $h$  favours the detection of vanishing points located near the centre of the image, whereas a small value of  $h$  favours the detection of

## Detection of principal directions

vanishing points located at infinity. Here, the neighbourhood used for the counting of the lines is a circular spherical cap with a constant radius  $\sigma'$ , which requires an isotropic sampling of the sphere.

The Gaussian sphere method appears to be a correct accumulator sphere for the constraint considered here in the context of the whole line accumulation, provided a correct value has been chosen for  $h$  and an isotropic sampling is used. This result does not hold when only intersection points are accumulated. But an isotropic sampling of the sphere in the whole line accumulation approach implies an involved process, whereas this problem does not exist for the accumulator space which is described previously.

### 5.2.4. Vanishing point detection

#### *"Noise" of the accumulator space*

The vanishing points looked for are peaks of the accumulator space; however a number of peaks are not vanishing points. Classically, the peaks are selected if they are above a fixed threshold (Quan and Mohr, 1989; Barnard, 1983). This approach is not very satisfactory because the threshold should obviously depend on a number of factors such as the number of lines or, in a case of lines weighted by their length, the average length of the straight line segments in the image. Usually, this threshold is fixed experimentally, which is arbitrary. In the following, a method for finding the optimal threshold is described.

It is important to estimate the significance of the peaks, in order to select a minimum number of false candidates and a maximum of good ones. A peak is selected if it is very unlikely to happen by accident. It will be seen that a peak actually generates a chapelet of peaks, therefore further precautions are needed (only the highest peaks over a large area of the accumulator spaces are selected). First the probability of a peak happening by accident is determined and the criterion to select a peak is defined ; then additional precautions are

discussed.

A number of lines located at random in the image cross each other at any point which do not necessarily correspond to a particular point such as their vanishing point - they just have to cross somewhere!-. The more lines in the image, the more likely several lines pass through the same neighbourhood in the image plane, i.e. in the same cell of the accumulator space. Using the statistical model defined in chapter 4, applied to the number of lines accumulated, it is possible to determine the probability of a peak occurring by accident.

A line is characterised by its distance  $d$  from the origin of the coordinate system, i.e. the centre of the image, and its direction  $\theta$ . The lines are supposed distributed at random in the image, according to the statistical model defined in chapter 4. In the accumulator space, a line is represented by a curve which crosses each vertical  $x'$  constant twice, in the interval  $]x'(d), x'_{\max}]$ , the vertical  $x' = x'(d)$  once, and never the vertical  $x' < x'(d)$ . The expected number of lines crossing the vertical  $x' = x'(d)$  is equal to  $n'(d)$  (see its expression in appendix 4). As all the directions are equiprobable the distribution of the lines along a vertical  $x'$  constant is uniform in the interval  $[0, y'_{\max}]$  (illustrated in figure 5.2.4.1). Therefore the density of probability at the point  $(x', y')$  corresponds to a binomial law corresponding to the parameters  $(n'(d), p = 2(2\bar{\sigma}'_p / y'_{\max}))$ , which is

$$p(N(x', y') = k) = \binom{n'}{k} p^k (1-p)^{n'-k} \quad (5.2.4.1)$$

Therefore, the expected value at a point of the accumulator space is

$$E(N(x', y')) = n'p = \frac{4 \bar{\sigma}'_p n'(d(x'))}{y'_{\max}(x')} = f(x'), \quad (5.2.4.2)$$

and its variance is



## Detection of principal directions

$$V(N(x', y')) = n'p (1-p) = f(x') \left(1 - \frac{4 \bar{\sigma}'_p}{y'_{\max}(x')}\right) = \psi(x'). \quad (5.2.4.3)$$

Remark : The value of  $y'_{\max}(x')$  depends on the sampling chosen for the accumulator space.

The function  $f(x')$  for the sampling (5.2.2.15) is displayed in figure 5.2.4.2. For illustrating the type of distribution obtained, 2000 lines have been synthesized according to the statistical model defined in chapter 3 and have been accumulated, the result is displayed figure 5.2.4.1.

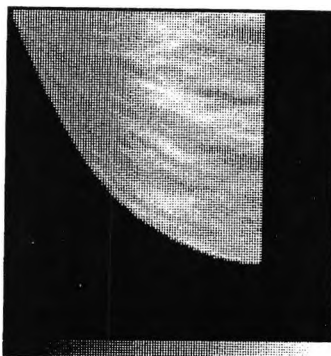


Figure 5.2.4.1 : Accumulator space corresponding to the accumulation of 2000 lines located at random in the image.

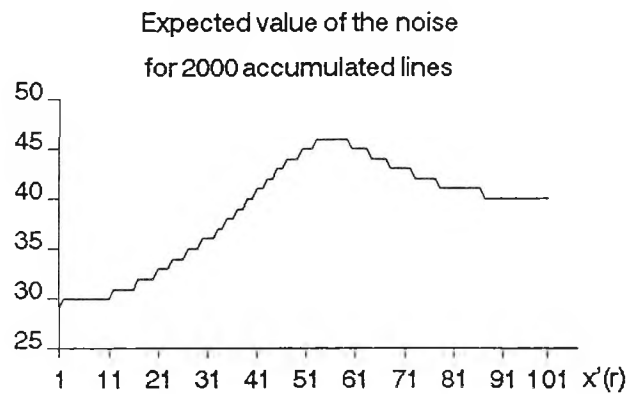


Figure 5.2.4.2 : Expected value of the accumulator space in function of  $x'$  where 2000 lines located at random in the image are accumulated.

*Detection of the significant peaks of the accumulator space*

A peak of the accumulator space is selected when the probability that it happens by accident is less than the admitted risk  $\tau$ , say 0.01.

$$(x', y') \text{ selected if } N(x', y') \geq k(x') \quad (5.2.4.4)$$

$$\text{with } k(x') \text{ such that } p(N(x', y') \geq k(x')) = \sum_{j \geq k} \binom{n'}{j} p^j (1-p)^{n'-j} \approx \tau$$

For large  $n'$ , say  $n' > 10$ , the Moivre-Laplace theorem (Papoulis, 1965) allows the binomial law to be approximated by a Gaussian law with the parameters  $(n'p, n'p(p-1))$ .

The expected number of points in the accumulator space with a value higher than  $k(x')$  is equal to  $\tau A$ , where  $A$  is the area of the accumulator space. Actually, only local maxima with a value higher than  $k(x')$  are selected as potential candidates. So, the expected number of false candidates depends on  $\tau$  and the density of peaks in the accumulator space. Using the Gaussian approximation for  $p(N=k)$ , it is possible to consider that

$$w_x(y') = N(x', y') - f(x')$$

behaves approximately as a normal process. Furthermore, if  $\tau$  is small it is possible to ignore the probability of having two local maxima for  $w_{x'}(y')$  between two successive passages by  $k(x')$ . Therefore, the expected number of local maxima above  $k(x')$  along a vertical,  $x'$  constant, is equal to half the expected number of passages by  $k(x')$ . Appendix 7, eq. A7.16, shows that it is equal to

$$n_{\tau}(x') = \frac{y'_{\max}(x') \sqrt{\varepsilon}}{2\pi \sqrt{\sigma'_p}} \exp\left(-\frac{h(x')^2}{2 \omega(x')}\right) \quad (5.2.4.5)$$

where  $h(x') = k(x') - \ell(x')$ , and  $\varepsilon$  is the height of a cell of the accumulator space, that is to say  $\varepsilon$  is equal to 1 (it is maintained in the equation for homogeneity). Using the Gaussian approximation

$$h(x')^2 = [\text{erf}^{-1}(0.5-\tau)]^2 \omega(x')$$

where  $\omega(x')$  is given by eq. 5.2.4.3. Therefore

$$n_{\tau}(x') = \frac{y'_{\max}(x') \sqrt{\varepsilon}}{2\pi \sqrt{\sigma'_p}} \exp\left(-\frac{[\text{erf}^{-1}(0.5-\tau)]^2}{2}\right). \quad (5.2.4.6)$$

However, the local maxima along a vertical are not necessarily local maxima in the 2D accumulator space (though the converse is true). Therefore  $d_{\tau} = n_{\tau}(x')/y'_{\max}(x')$  does not represent the density of local maxima in the accumulator space. The correlation function along  $x'$  is not easy to find, so that it is difficult to be conclusive in the 2D case. It is only possible to give an estimation, by using the fact that the correlation is obviously high (because of the continuity of the lines into the accumulator space) and the additional following condition :

- if two candidates P and P' are found such that the distance between P and P',  $d(P,P')$ , is inferior to its corresponding uncertainty  $\sigma_d$ , then they are merged into one.

Thus, an intuitive reasoning indicates that the maxima along the vertical  $x'$  being likely connected to the maxima of the vertical  $x'-1$ , the expected number of false candidates is of the same order of magnitude as  $d_{\tau} A/\xi$ , where  $1/\xi$  is equal to the average value of  $1/\hat{\sigma}_{x'}$ . For instance, in the case of the accumulator space defined by (5.2.2.15),  $A = 6950$ ,  $(1/\hat{\sigma}_{x'})_{av} = 1/3.82$  and  $\bar{\sigma}_p' = 1.5$  which gives, for  $\tau = 0.01$ , an expected number of false alarms of approximately 15, which is confirmed by the experience achieved on the accumulation of successively 50, 100, 300 lines, randomly chosen (see figures 5.2.4.3 and 5.2.4.4).

Nb of direct.	50		100		300	
Trials	$\tau'$	nf	$\tau'$	nf	$\tau'$	nf
1	0.009	14	0.01	15	0.015	17
2	0.01	19	0.011	16	0.0082	12
3	0.015	13	0.009	12	0.007	10

Figure 5.2.4.3 : Number of false alarms nf when successively 50, 100, 300 random directions have been accumulated, corresponding to the risk  $\tau = 0.01$ .  $\tau'$  corresponds to the real percentage of points above  $k(x')$ .

Predicted risk $\tau$	actual $\tau'$	predicted nf	actual nf
0.001	0.	1.9	0
	0.00015		1
	0.00015		1
0.005	0.0048	8.5	8
	0.0041		8
	0.006		10
0.01	0.01	15.5	15
	0.011		16
	0.009		12
0.02	0.036	28	35
	0.035		36
	0.013		21
0.04	0.063	51	27
	0.044		37
	0.062		37

Figure 5.2.4.4 : Predicted number versus actual number of false alarms when 100 random directions been accumulated, for different risks  $\tau$ .

Thus the detection of the vanishing point is consistent over the image plane not only with respect to the expected uncertainty, but also with respect to the expected level of noise. Furthermore, it is remarkable that the expected number of false alarms is independent of the number of lines accumulated, i.e. of the complexity of the scene.

Actually the statistical model does not take into account the lines associated with another vanishing point, although several main directions are supposed to exist. The presence of a real vanishing point near a peak produces noise in the neighbourhood of this peak which is not properly described by the probabilistic law described above. For example, the image of a set of pipes may produce several local maxima corresponding to the intersection of this set with various lines. This problem is due to the lack of accuracy of the model used and is clearly worse when no model is used and the peaks are selected when above a fixed threshold. Some of the other methods (Magee and Aggarwall, 1984), (Quan and Mohr, 1989) use the following strategy : once a vanishing point has been found, all the lines classified with it are taken away from the list and the algorithm is performed on the rest of the lines. However, a number of lines in the image may pass through several vanishing points and such a method may ignore an important direction in the image (usually the less represented horizontal direction), by having eliminated too many lines.

Here, this problem is empirically solved by the following considerations. First, the vertical lines are nearly always very numerous and are represented in the accumulator space by a very high peak ; second, the main directions are supposed to be perpendicular and thereby cannot be located in the same area of the accumulator space.

The risk of removing important lines is limited by the following strategy. The lines are removed from the list after classification with a vanishing point candidate, only if the corresponding peak is very significant, e.g the probability that it happens by accident is less than 0.0005 ( $\tau=0.0005$  for the sampling (5.2.2.15) corresponds to a

number of false alarms  $\ll 1$ ). This means that a class of lines is discarded if the number of these lines is so high that it would be responsible for numerous significant peaks in the accumulator space. It is always the case for the class of vertical lines in the images studied.

The classification of the lines with a vanishing point candidate is achieved by a likelihood ratio test (sections 3.3.3 and 5.4). The more numerous the class, the more tolerant the test. This means that the number of lines missed by this test, which could be responsible for further noise, is approximately constant whatever the size of the class, by contrast with a MD test where the *percentage* of lines missed is constant. It is seen in the results (section 5.5) that the efficiency of the LR test for eliminating the lines associated with a preponderant class is more efficient than the MD test.

Once the classification corresponding to numerous classes is achieved, the accumulation is then performed again on the remaining lines. The search for all significant peaks is expensive and too many peaks may still appear. A first method consists of selecting only vanishing points perpendicular to the directions already detected, by using a MD test. This method is interesting but is computationally expensive as the MD test depends on each pair of points tested ; moreover it would definitely ignore directions not perpendicular to the first directions found. An alternative consists of dividing the accumulator space in several areas, in each of which no more than one vanishing point is expected (see figure 5.2.4.5). The maximum peak (or peaks if several peaks have the same value) of each area is found and tested by eq. 5.2.4.4). Let us remark that the test inside an area should be approximately constant, i.e.  $k$  approximately constant in the area, in order to have comparable peaks. This method is quick and efficient, but produces more candidates than the first method.

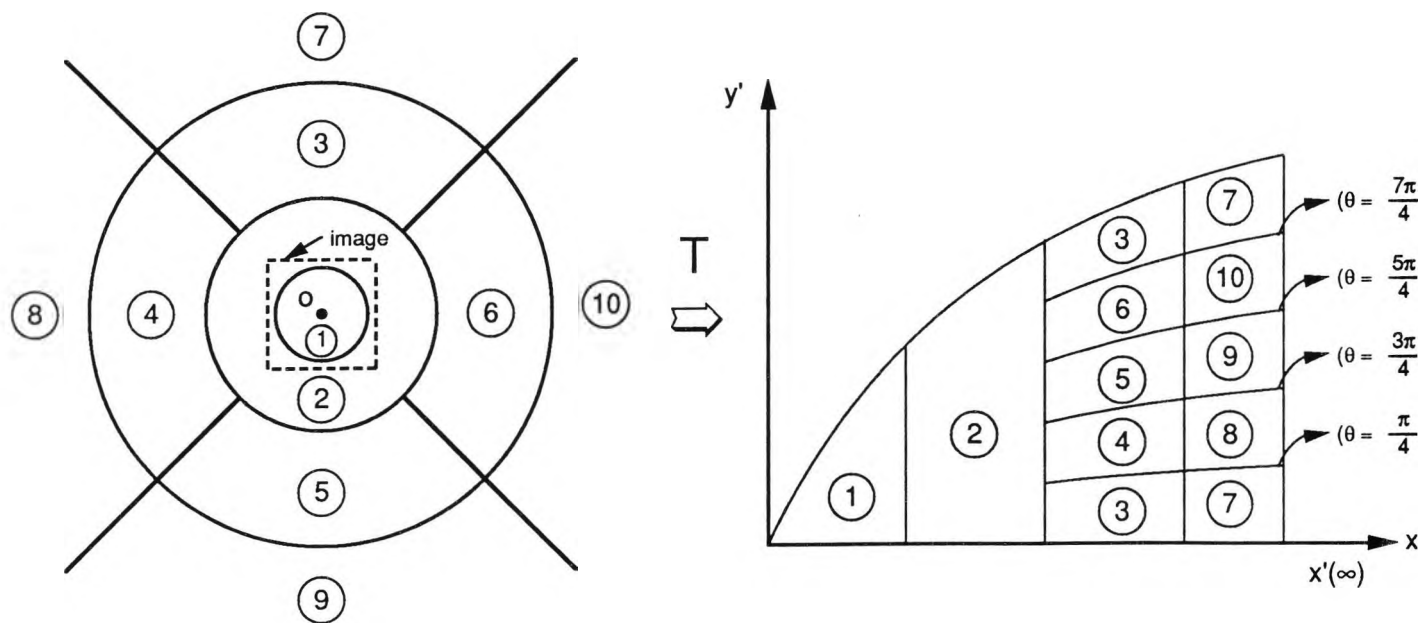


Figure 5.2.4.5 : Division of the accumulator space in several areas in each of which no more than one vanishing point is expected. The triangle corresponds to three perpendicular directions (see appendix 1).

The method described in the previous paragraphs minimizes the probability of false alarms in a consistent way over the image, furthermore it allows the selection of vanishing points corresponding to a few straight lines by minimizing the risk of loss of these lines in the process of elimination of larger classes. The method does not depend on the complexity of the scene.

### 5.3 CLASSIFICATION OF THE LINES

The lines are individually classified with the vanishing point candidates in order to take into account the value of  $\sigma$  corresponding to each line (appendix 5, eq. A5.9) and to perform the test 5.2.2.6 .

For each vanishing point candidate and for each line a LR test is performed. If the line succeeds the test, it is used by a Kalman filter to improve the determination of the coordinates of the vanishing point, and the line is classified with the vanishing point.

*Definition of the LR test*

A particular vanishing point  $P$  is considered in the following. The set  $\Omega$  is the set of the pairs  $(P, L_1)$ , where  $L_1$  is any straight line segment extracted from the image. A sub-set  $\Lambda$  is the set of pairs such that the centroid of the segment  $L_1$  is at a fixed distance  $r'$  from the point  $P$  with a constant length  $\ell$ .

The hypothesis  $H$  assumes that the line  $L_1$  should pass through  $P$ , e.g. because  $P$  is its vanishing point (let us remark that this reasoning also holds for a corner). The hypothesis  $\bar{H}$  assumes that the line  $L_1$  passes near  $P$  by chance, i.e. the corresponding 3D line is not related to the 3D meaning of  $P$  in the scene. The decision variable is the distance  $D_1$  from the line  $L_1$  to the point  $P$  when the distance  $r$  from  $P$  to the origin is less than  $r_s$ . Otherwise, it is the distance  $\Delta_1$  from the curve  $C_1 = T(L_1)$  to the point  $P'(x', y') = T(P)$  along a vertical of the accumulator space (for large  $r$ , the accumulator space is the appropriate space because it is bounded).

The methodology described in chapter 3 may now be applied. Under the hypothesis  $H$ , the distance  $D_1$  is assumed to be the error of measurement and is modelled by a normal law with zero mean and a variance equal to  $\sigma_1^2$  defined in appendix 5, eq. A5.9.

Under the hypothesis  $\bar{H}$ , the line  $L_1 \in \Lambda$  may have any direction  $\alpha$  in the image with a uniform probability. From

$$D_1 = d |\sin(\alpha)| \text{ and } d |\cos(\alpha)| = \sqrt{r'^2 - D_1^2},$$

the probability law of  $D_1$  is deduced

$$p_2(D_1) = p_2(\alpha(D_1)) \frac{\partial \alpha}{\partial D_1} = \frac{1}{\pi \sqrt{r'^2 - D_1^2}}.$$

Therefore if  $r < r_s$ , the LR test is



Detection of principal directions

$$\frac{p(H)}{p(\bar{H})} \sqrt{\frac{\pi}{2}} \frac{\sqrt{r'^2 - D_1^2}}{\sigma_1} \exp\left(-\frac{D_1^2}{2\sigma_1^2}\right) > 1, \quad (5.3.1)$$

In the case  $r > r_s$ , the calculation is performed in the accumulator space.

In the accumulator space, the distribution of the curves along a vertical under the hypothesis  $\bar{H}$  is uniform with the density  $2/y'_{\max}$  (the curves cross a vertical twice).

The uncertainty of the location of the curve  $C_1$  along the vertical passing through  $P'(x', y')$ , using the geometric interpretation of the accumulator space (section 5.2) and the hypothesis  $r'$  large, is equal to

$$\sigma_{\Delta_1} = \frac{\sigma_1}{\sigma_p} \frac{\bar{\sigma}_p}{\sigma_p} \approx \frac{\sigma_1}{2\pi r'} y'_{\max}(r'),$$

Let  $\sigma_{\theta_1}$  be the expected value of the error of the angle  $\theta_1$  of the line ( $L_1$ ), then if  $r'$  is large  $\sigma_{\theta_1} \approx \sigma_1/r'$ . The LR test is therefore

$$\frac{p(H)}{p(\bar{H})} \sqrt{\frac{\pi}{2}} \frac{1}{\sigma_{\theta_1}} \exp\left(-\frac{\Delta_1^2}{2\sigma_{\Delta_1}^2}\right) > 1. \quad (5.3.2)$$

When  $r'$  large

$$\Delta_1 \approx \frac{D_1 y'_{\max}}{r' 2\pi}$$

Thus when  $r'$  is large, the tests (5.3.1) and (5.3.2) are equivalent. Therefore, the value of  $r_s$  may be set to any large but finite value, e.g.  $4R$ ,  $R$  being the radius of the image.

The value of  $P(H)/P(\bar{H})$  is assumed independent of  $\Lambda$ , i.e. the distance  $r'$  and the length of the segment  $l$  does not give any

information on  $p(H)/p(\bar{H})$ . This hypothesis is not contradictory with the fact that a segment should necessarily be located at some distance from the vanishing point, as  $P$  could as well be a corner. No reference is made to its physical meaning, but only to the fact that the line does or does not pass through  $P$  by accident.

Applying eq. 3.3.3 to the set  $\mathcal{N}$  of lines crossing  $[y' - \bar{\sigma}'_p j, y' + \bar{\sigma}'_p j]$ , where  $y'$  is the ordinate of  $P' = \sigma(P)$ , it comes

$$\frac{p(H)}{p(\bar{H})} = \frac{nn - n' \int_{\mathcal{N}_2} p_2}{n' \int_{\mathcal{N}_1} p_1 - nn},$$

where  $n'(x')$  is the expected number of directions accumulated along the vertical  $x'$  (see eq. 5.2.4.1),  $nn$  is the value of the peak  $P'$  in the accumulator space,

$$\int_{\mathcal{N}_1} p_1 = 1 - 2 \operatorname{erf} \left( \frac{\bar{\sigma}'_p}{\sigma_1} \right),$$

and  $n' \int_{\mathcal{N}_2} p_2$  is equal to  $\ell(x')$  (see eq. 5.2.4.2).

Remark : if  $r'$  is large the LR test is equivalent to

$$\frac{\Delta_1^2}{\sigma_{\Delta_1}^2} < 2 \operatorname{Log}_e \left( \frac{\sqrt{\pi} p(H) \ell}{2 p(\bar{H}) \sigma_0} \right) = T_0 + 2 \operatorname{Log}_e(\ell). \quad (5.3.3)$$

Thus, it corresponds to a MD test with a threshold depending on  $p(H)/p(\bar{H})$  and the logarithm of the length  $\ell$ . If the length of the segment is large, i.e. its uncertainty is low, the test is more tolerant. However, the tolerance grows less rapidly than the uncertainty decreases, thereby the LR test may be described as intermediate between a MD test and a neighbourhood test.

The equation 5.3.3 may be used for approximating the LR test where

## Detection of principal directions

$p(H)/p(\bar{H})$  is intuitively estimated by the user. Often, it is possible to select a reasonable threshold for the MD test in a particular case (either an average case or an extreme case). Let  $T_s$  be a reasonable threshold in the MD test for the straight line segments with a length  $l_s$ , then

$$\frac{\Delta_i^2}{\sigma_{\Delta_i}^2} < T_s + 2 \text{Log}_e(l/l_s) .$$

### Scoring of a line

The line  $L_i$  is associated with the vanishing point  $P$  if  $R > 1$ . The higher the value of  $R$ , the higher the confidence of the classification. The score of the classification of the line  $L_i$  with the point  $P$  is

$$p(H|D_i) = \frac{R_i}{R_i + 1} . \quad (5.3.4)$$

### Scoring a class of lines associated with a vanishing point candidate

Once all the lines have been evaluated for classification with the point  $P$ , it is possible to update the probability of the hypothesis  $H$ , i.e. the probable ratio of the number of directions in the image having  $P$  for vanishing point, to the total number of directions considered. It is more interesting to limit this definition to a neighbourhood of  $P$ , say  $\mathcal{N}$ , where  $\mathcal{N}$  is the set of straight lines at a distance inferior to  $\bar{\sigma}_p$  from  $P$ . Let  $\alpha_0$  be the prior probability of  $H$ , knowing that the line crosses  $\mathcal{N}$  (i.e. using the notations of section 3.3  $\alpha_0 = p(H|\mathcal{N}) = (n' \int_{\mathcal{N}} p_1) p(H)$ ,  $n'$  being the total number of lines expected along the vertical of the accumulator space considered). Let  $nn$  be the number of directions in  $\mathcal{N}$ ,  $\alpha_0 nn$  is the expected number of directions crossing at  $P$  "on purpose", and  $(1-\alpha_0) nn$  the expected number of directions passing near  $P$  by chance (corresponding to noise).

The methodology defined in chapter 3 is used with  $p(\alpha)$  defined as follows. As seen previously, the noise of the accumulator space may be

modelled by a binomial law with parameters  $(p, n_2)$ , where  $p$  is the number of cells considered along a vertical and  $n_2$  the expected number of lines corresponding to noise crossing these cells. The expected number of lines crossing the cell  $P$  by accident is  $n_2 p = nn(1-\alpha_0)$ . Hence,  $p(\alpha)$  is equal to

$$p(\alpha) = p(n_2 = (1-\alpha)nn | nn) = \frac{\binom{n_2}{(1-\alpha)nn} p^{(1-\alpha)nn} (1-p)^{n_2 - (1-\alpha)nn}}{\sum_{k \leq nn} \binom{n_2}{k} p^k (1-p)^{n_2 - k}} \quad (5.3.5)$$

The denominator is a normalization factor. When  $nn$  is large enough, the Gaussian approximation may be used and

$$p(\alpha) = \frac{\exp\left(-\frac{(\alpha - \alpha_0)^2 nn^2}{2\upsilon}\right)}{\sqrt{2\pi \upsilon} (0.5 + \operatorname{erf}(\alpha_0 nn / \sqrt{\upsilon}))}$$

where  $\operatorname{erf}(\alpha) = \frac{1}{\sqrt{2\pi}} \int_0^\alpha \exp(-\frac{\alpha^2}{2}) d\alpha$ , and  $\upsilon$  is the variance of the noise at the point  $P$  of the accumulator space,  $\upsilon = (1-\alpha_0)(1-p)nn$ .

To compute the function  $G(\alpha)$  (eq. 3.3.6), the odds  $O_1$  of a direction is defined as the maximum value of the odds corresponding to the straight line segments grouped with this direction, because it corresponds to the most likely segment which may have  $P$  for a vanishing point. Since the function  $G(\alpha)$  is proportional to  $p(\alpha)$  and  $p(\alpha)$  is very small when  $\alpha$  is far from  $\alpha_0$ , the maximum of  $G(\alpha)$  should be close to  $\alpha_0$ . This property ensures the stability of the process.

The updating process measures the degree of convergence of the lines around  $P$ . For instance, if a set of lines crosses the neighbourhood of  $Q$  located at  $d$  from  $P$ , such that  $Q$  and  $P$  are considered as two possible candidates (frequent case),  $\alpha_0$  is supposed to be the same in both cases (if  $d$  small enough), but  $\alpha_Q$  is inferior to  $\alpha_P$ , because the lack of convergence in  $Q$  "flattens" the Gaussian distribution, favouring the uniform distribution, i.e.  $1-\alpha_Q$ .

*Updating the risk of false alarms*

As has been seen sub-section 5.2.4, it is possible to estimate the expected number of false alarms corresponding to the risk  $\tau$  which defines the thresholding process. Conversely it is possible to associate a risk  $\tau_p$  to a peak  $P$  with a value  $nn$  by chance in the following way

$$\tau_p = 1 - \sum_{k=0}^{k=nn-1} \binom{n'}{k} p^k (1-p)^{n'-k}, \quad (5.3.6)$$

where  $(p, n')$  are the parameters of the binomial law corresponding to the noise, defined in subsection (5.3.2).

The expected number of directions associated with a vanishing point and the risk of false alarms are complementary information, albeit related. The former indicates the significance of the interpretation of  $P$  as a vanishing point and the latter gives information about the confidence of such an interpretation.

#### 5.4 MAIN PERPENDICULAR DIRECTIONS

In the following, the coordinate system of the image is  $(Ox, Oy/\rho)$  in order to correct the distortion between  $Ox$  and  $Oy$  scales, so that the reference coordinate system  $(Ox, Oy)$  is Euclidean.

If two vanishing points  $P_1$  and  $P_2$  correspond to perpendicular directions in the scene, then (appendix 1)

$$\vec{OP}_1 \cdot \vec{OP}_2 + f^2 = 0. \quad (5.4.1)$$

Because of uncertainty of measurement  $V(P_1, P_2) = \vec{OP}_1 \cdot \vec{OP}_2 + f^2$ , is a normal variable with zero mean and variance  $\sigma_v^2$  deduced from the covariance matrices associated with  $O, P_1, P_2$  and  $f$ . Let  $V$  be the value taken by  $V(P_1, P_2)$ .

Conversely, if two vanishing points obey the relation (5.4.1) they

correspond to perpendicular directions in the scene. Therefore there is no notion of such a relation happening by accident. If  $P_1$  and  $P_2$  are two real vanishing points and if the corresponding  $V^2/\sigma_v^2$  is less than a threshold, the corresponding directions in the scene are approximately perpendicular. Thus, a test based on the Mahalanobis distance seems appropriate. The threshold may be chosen reasonably high as most lines in the scene have been supposed parallel to three main perpendicular directions, therefore only three vanishing points are expected in the image, obeying the relation (5.4.1) in pairs. However, there are usually more than three candidates and  $P_1$  or  $P_2$  may be a false alarm. As the uncertainty of  $V$  is high (due to the uncertainty of the principal point and to  $\sigma_r$  eq. 5.2.2.14), a MD test is not selective enough when using a "common sense" threshold (above 2 for selecting more than 95% of the good candidates). The choice of a lower threshold is arbitrary if no explicit reference to the segmentation noise is made. Actually the perpendicularity test is not aimed at checking the perpendicularity of the main directions in the scene, but rather to provide an additional filter against false alarms. It has been seen in chapter 3 that the advantage of a likelihood ratio test is that it is dependent on the level of segmentation noise. If the noise is high the test is more selective than if it is low. Moreover, the test fails when the uncertainty of  $V$  is very high, independently of  $V$ , which seems appropriate as  $V$  would have no significance.

In order to define a likelihood ratio test, two complementary hypotheses have to be found. The set  $\Gamma$  of the vanishing point candidates is supposed to have the following characteristics :

- it contains at most three real vanishing points corresponding to three perpendicular directions (it may also contain duplicates of some of them corresponding to large  $\sigma_r$  and slightly different classifications).
- the remaining candidates are false alarms due to the noise of the accumulator space.

## Detection of principal directions

Let  $\Omega$  be the set of the pairs of the elements of  $\Gamma$ , and  $\Lambda$  be a subset of  $\Omega$ , such that  $P_1$  is at the distance  $r_1$  from  $O$ , and  $P_2$  at the distance  $r_2$  from  $O$ . Let  $H$  be the hypothesis that  $P_1$  and  $P_2$  are two real vanishing points. By hypothesis, they correspond to perpendicular directions in the scene and must obey the relation (5.4.1). The hypothesis  $\bar{H}$  is : either  $P_1$  or  $P_2$  is a false alarm.

The calibration parameters are assumed to be exactly known, and the principal point is assumed to correspond to the centre  $O$  of the image. The decision variable  $V$  may be rewritten

$$V = r_1 r_2 \cos\theta + f^2,$$

where  $\theta$  is the angle  $(OP_1, OP_2)$ . Then

$$\frac{\partial V}{\partial \theta} = -r_1 r_2 \sin\theta.$$

Thus, the density of probability  $p_2$  in the case of  $\bar{H}$  is equal to

$$p_2(V) = p(\theta) \frac{\partial \theta}{\partial V} = \frac{1}{2\pi r_1 r_2 |\sin\theta|} = \frac{1}{2\pi \sqrt{r_1^2 r_2^2 - (V-f^2)^2}}$$

and the variance of  $V$  in the case of  $H$  is

$$\sigma_v^2 = r_1^2 r_2^2 \sin^2\theta \sigma_\theta^2.$$

The variance  $\sigma_\theta^2$  of the angle  $\theta$  depends on the covariance matrix of  $P_1$  and  $P_2$  and is given in appendix 6, eq. A6.2. The density of probability  $p_1$  is a Gaussian with zero mean and a variance  $\sigma_v^2$ .

The likelihood ratio test may be written

$$\frac{p(H)}{p(\bar{H})} \frac{\sqrt{2\pi}}{\sigma_\theta} \exp\left(-\frac{V^2}{2r_1^2 r_2^2 \sin^2\theta \sigma_\theta^2}\right) > 1, \quad (5.4.2)$$

Unfortunately, the uncertainty of the calibration parameters often is far too important to be negligible and should be taken into account

in the likelihood test. The error of the calibration parameters is assumed to be normal with zero mean and covariance matrix  $C_c$ . Let  $\vec{C}$  be the vector associated with the calibration parameters

$$\vec{C} = (x_c, y_c, f, \rho)^t,$$

$V$  may be written at the first order

$$V \approx V_0 + \frac{\partial V}{\partial \theta} d\theta + \vec{\nabla}_c(V) \cdot d\vec{C},$$

where  $\vec{\nabla}_c$  is the gradient of  $\vec{C}$ . Thus, the law of  $V$  is now the convolution product of the law  $p_1$  (in case of  $H$ ) or  $p_2$  (in case of  $\bar{H}$ ) previously found, with the Gaussian law  $p_c$  with zero mean and variance  $\sigma_c^2$  (eq. A6.5) describing the variations of  $\vec{\nabla}_c(V) \cdot d\vec{C}$ . The LR test becomes

$$\frac{p(H)}{p(\bar{H})} \frac{p_1 * p_c(V)}{p_2 * p_c(V)} > 1$$

The convolution of  $p_1$  with  $p_c$  is a Gaussian law with zero mean and variance equal to  $\sigma_v^2 + \sigma_c^2$ . The convolution of  $p_2$  with  $p_c$  is computed by numerical means (Press, 1988).

The last problem is the determination of  $p(H)/p(\bar{H})$ . It is defined as the ratio of the expected number of pairs of real vanishing points in  $\Lambda$  to the expected number of false alarms. Prior knowledge of the viewpoint may be introduced here. A model of a vertical upright camera is used in the following.

In the application studied here, all the pictures have been taken with the camera approximately vertical. The model of the prior knowledge is therefore chosen as follows. One of the horizontal directions, say the left side, may be any direction in the horizontal plane forming an angle  $\varphi$  between 0 and  $\pi/2$  with the projection of the optic axis on the horizontal plane.

The prior knowledge of the vertical direction may be described in



## Detection of principal directions

the accumulator space by a Gaussian law centred on the nearest point  $V'_1$  of the points  $V'_1$ ,  $V'_2$  and  $V'_3$  with abscisse  $x'_{\max}$  and ordinates  $0$ ,  $y'_{\max}/2$  and  $y'_{\max}$ , with a covariance matrix  $C_V$ . Thus, if  $P$  is assumed to correspond to a vertical direction, then the expected density of vanishing point in the accumulator space at  $P$  is

$$\lambda_P = \exp(\overrightarrow{PV}_1^t C_V^{-1} \overrightarrow{PV}_1 / 2) / (2\pi \sqrt{\det(C_V)}).$$

Now, let  $L$  be the line  $(OV)$  in the image, where  $V$  is the vanishing point corresponding to the vertical direction,  $H_0$  be the horizon and  $O'$  be the intersection of  $L$  with  $H_0$ . Let  $r'$  be the distance between  $O'$  and the vanishing point  $P$  corresponding to an horizontal direction. From the relationship between the direction and its corresponding vanishing point (see appendix 1), the density of probability  $p_h(r')$  may be deduced

$$p_h(r') = p(\varphi) \frac{\partial \varphi}{\partial r'} = \frac{2}{\pi f' \left(1 + \frac{r'^2}{f'^2}\right)},$$

where  $f'^2 = f^2 + OO'^2$ .  $OO'$  is described by a Gaussian law centred at zero with variance  $\sigma_h^2$ ,  $\sigma_h$  is related to  $C_V$  and to the covariance matrix of the calibration parameters (see appendix 6). Then, from  $p_h(x') = p_h(r') \partial r' / \partial x'$ ,  $p_h(x')$  in the accumulator space is deduced. Thus, if  $P(x, y)$  may correspond to an horizontal direction then  $r' = x$ ,  $OO' = y$ ,  $r = (x^2 + y^2)^{1/2}$ ,  $x' = x(r)$ . The expected density of a vanishing point in the accumulator space at  $P$  is

$$\lambda_P = p_h(x') \cdot \exp(-y^2 / 2\sigma_h^2) / (\sqrt{2\pi} \sigma_h).$$

Then  $p(H)/p(\overline{H})$  is defined as

$$\frac{p(H)}{p(\overline{H})} = \frac{\lambda_{P1} \lambda_{P2}}{\kappa_{P1} \lambda_{P2} + \kappa_{P2} \lambda_{P1} + \kappa_{P1} \kappa_{P2}}, \quad (5.4.3)$$

where  $\kappa_{P1} = \overline{\sigma}_p n'_{\tau_{P1}}(x'_1) / y_{1\max}$  and  $\kappa_{P2} = \overline{\sigma}_p n'_{\tau_{P2}}(x'_2) / y_{2\max}$ ,  $n'_{\tau_1}(x')$  is the

expected density of false alarms along the vertical passing through  $P'_1$  and is given by eq. 5.2.4.5, where the updated expected number of directions associated with  $P_1$ ,  $h(x) = nn_1 \cdot x$ ,  $\tau_{P_1}$  is given by eq. 5.3.6.

If  $\kappa_{P_1}$  is negligible, which is the case of peaks corresponding to numerous classes, e.g.  $\tau < 0.0005$ , then (5.4.3) becomes

$$\frac{p(H)}{p(\bar{H})} = \frac{\lambda_{P_2}}{\kappa_{P_2}},$$

which is the ratio of the expected number of main vanishing points to the expected number of false alarms at  $P_2$ . This happens when one of the two points corresponds to the vertical direction. If both classes are very numerous, then the LR ratio is always very large even if they do not correspond to proper perpendicular directions. If both classes are relatively small and therefore unreliable, then the test is very selective : only real perpendicular directions succeed the test. Thus, this LR test behaves exactly as required : it does not provide a measure for the real perpendicularity of the main directions looked for, but it provides an additional filter against false alarms by using the prior knowledge that such directions are likely to be perpendicular. Lower and upper limits for  $p(H)/p(\bar{H})$  may be fixed, in order to give any candidate a chance of succes (i.e. the corresponding threshold of the MD test should always be positive) and to ensure a minimal significance to the perpendicularity relationship.

#### *Score of a pair or a triplet of directions*

The score of a pair of directions, i.e. a pair of vanishing points  $P_1, P_2$ , is equal to  $\frac{R}{R+1}$  (see section 3.3.3). It corresponds to a measure of likelihood of having two real vanishing points using available prior information on the feature extraction and on the viewpoint. Let us assume that it may be written  $s = s_1 s_2$ ,  $s_1$  corresponding to the score of  $P_1$  and  $s_2$  corresponding to the score of  $P_2$  (which is done for convenience but cannot be the case as their likelihoods are no longer independent). Then it is natural to define

## Detection of principal directions

the score of a triplet of perpendicular directions as being equal to the product  $s_1 s_2 s_3$  equal to

$$\sqrt{\frac{R_1 R_2 R_3}{(R_1+1)(R_2+1)(R_3+1)}}$$

This score takes into account the confidence associated with each point of the triplet and the quality of the perpendicularity. The larger the confidence, the less important the perpendicularity.

A hierarchical tree may be built : first, the straight line segments are accumulated to form straight line directions which are grouped into classes corresponding to parallel lines in the 3D scene, then the classes are grouped by pairs, each pair defining approximately perpendicular directions in the scene ; eventually the consistent pairs are merged to form triplets of perpendicular directions hypothesised to be the main directions of the scene (see figure 5.4.1). With each node is associated a score.

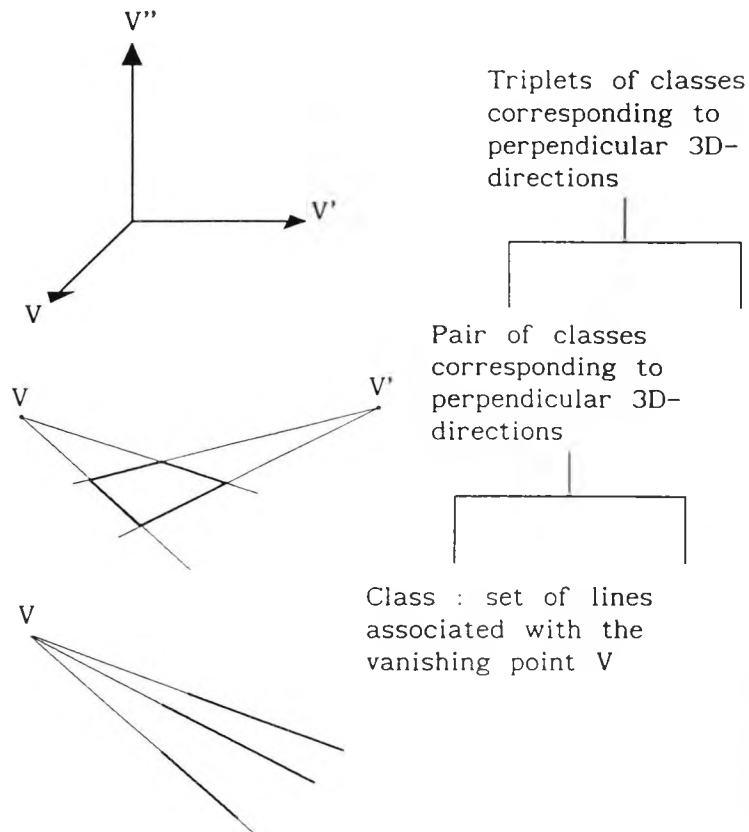


Figure 5.4.1 : Interpretation tree of the 3D directions of line segments from an image.

The triplets and then the pairs of perpendicular directions are sorted according to their scores. If no pair has been found, the interpretation has failed (this usually means that only the vertical direction has been found, which is a very poor performance for the effort involved!).

## 5.5 ELLIPTICAL ARC CLASSIFICATION

The elliptical arcs are assumed to be the projection of circular arcs in the scene. If an arc lies in a plane associated with two main directions, then the image of the arc may be associated with the two corresponding vanishing points in the following way.

Let  $V$  and  $V'$  be the two vanishing points associated with the perpendicular directions considered, let the points  $A$  and  $B$

## Detection of principal directions

(respectively  $A'$  and  $B'$ ) be defined by the tangents to an ellipse passing through the point  $V$  (respectively  $V'$ ). The ellipse may be associated with two vanishing points  $V$  and  $V'$  if and only if (figure 5.5.1) the line joining the points  $A$  and  $B$  (respectively  $A'$  and  $B'$ ) of the ellipse passes through the vanishing point  $V'$  (respectively  $V$ ). This ellipse is the projection of a circle if the 3D axis lengths are equal, that is to say :

$$AB \frac{\sqrt{(f^2 + OV)^2}}{f MV} = A'B' \frac{\sqrt{(f^2 + OV')^2}}{f MV'} \quad (5.5.1)$$

The two lines  $(AB)$  and  $(A'B')$ , each of them corresponding to a vanishing point, intersect at the projection  $q(x_g, y_g)$  of the circle centre onto the image.

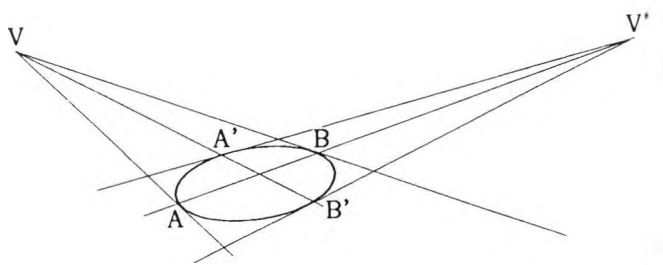


Figure 5.5.1 : Ellipse associated with two vanishing points.

## 5.6 RESULTS

The accumulator space depends on the statistic of  $1/\ell^2$  which depends on the complexity of the image and on the preprocessing stages. The size of the image used here is 256 and only lines with a length greater than 15 have been accumulated (lines with a length below 15 are very numerous and do not provide meaningful information as their uncertainty is very high and they often correspond to noise). Thus the pre-accumulation in  $(d, \alpha)$  results in a narrow distribution of  $1/\ell^2$  around  $a = 0.0019$ , determined experimentally over a number of images (see figure 4.4.2.1).

The method has been applied to various types of indoor scene. Figure

5.6.1 shows the initial images. Their respective accumulator spaces are displayed with all the lines accumulated, figure 5.6.3, and with a number of lines removed, figure 5.6.4. The final result of the classification is displayed in figure 5.6.5. All the main directions represented by a sufficient number of segments have been found. False candidates have been found, corresponding generally to corners. They are filtered out by using the constraint 5.2.2.6 which is usually not satisfied by the corners. It can be noticed that vanishing points have been found corresponding to a small value of  $r$  ( $r = 18$ ), to a average value of  $r$  ( $r = 135$ ) and to an infinite value of  $r$ , with peaks of similar shape along  $y'$ .

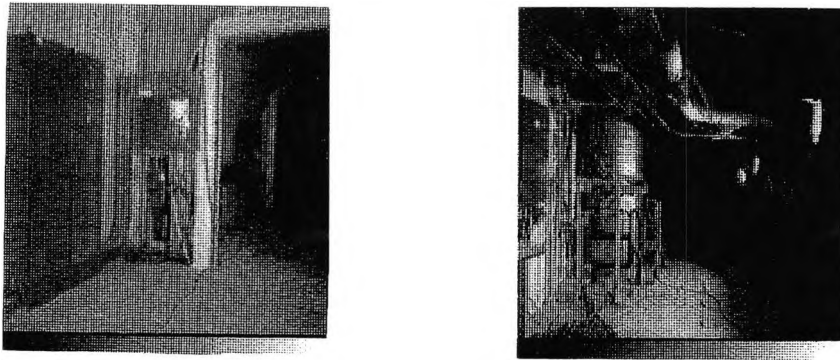


Figure 5.6.1: Initial images 1 and 2

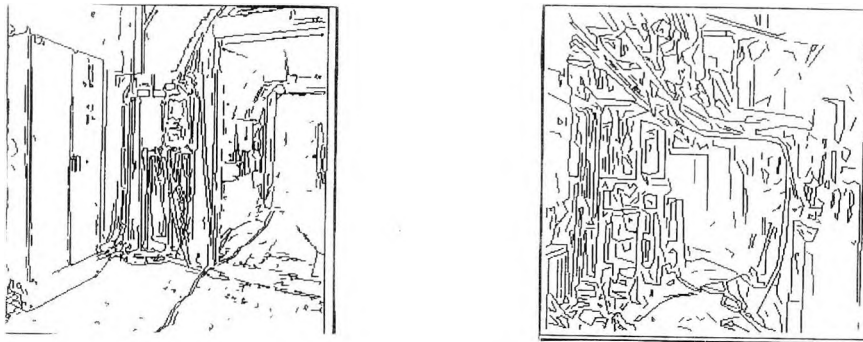


Figure 5.6.2 : Extracted segments of the images 1 and 2.

## Detection of principal directions



Figure 5.6.3 : Accumulator spaces with all the lines accumulated (For clarity the image has been scaled from 0 to 255). The axis  $y=0$  corresponds to the vertical lines (For convenience, the  $\theta$  reference has been shifted by  $\pi/2$ ).



Figure 5.6.4 : Accumulator space : the vertical lines have been removed (For clarity the images have been scaled from 0 to 255).

An extract of the result of the line classification is displayed in figure 5.6.6, and the result of the search for perpendicular triplets of directions in image 1 is displayed in figure 5.6.7. The best scored triplet corresponds to the triplet displayed in figure 5.6.5. The other triplet is composed of the vertical and the horizontal directions nearly parallel to the image plane, the other class has been falsely detected and corresponds to a corner. In image 2, the score of the triplet is 0.5 which is bad because the perpendicularity of the two corresponding horizontal directions is very bad ; it is compensated by the very good scores of each of these directions.

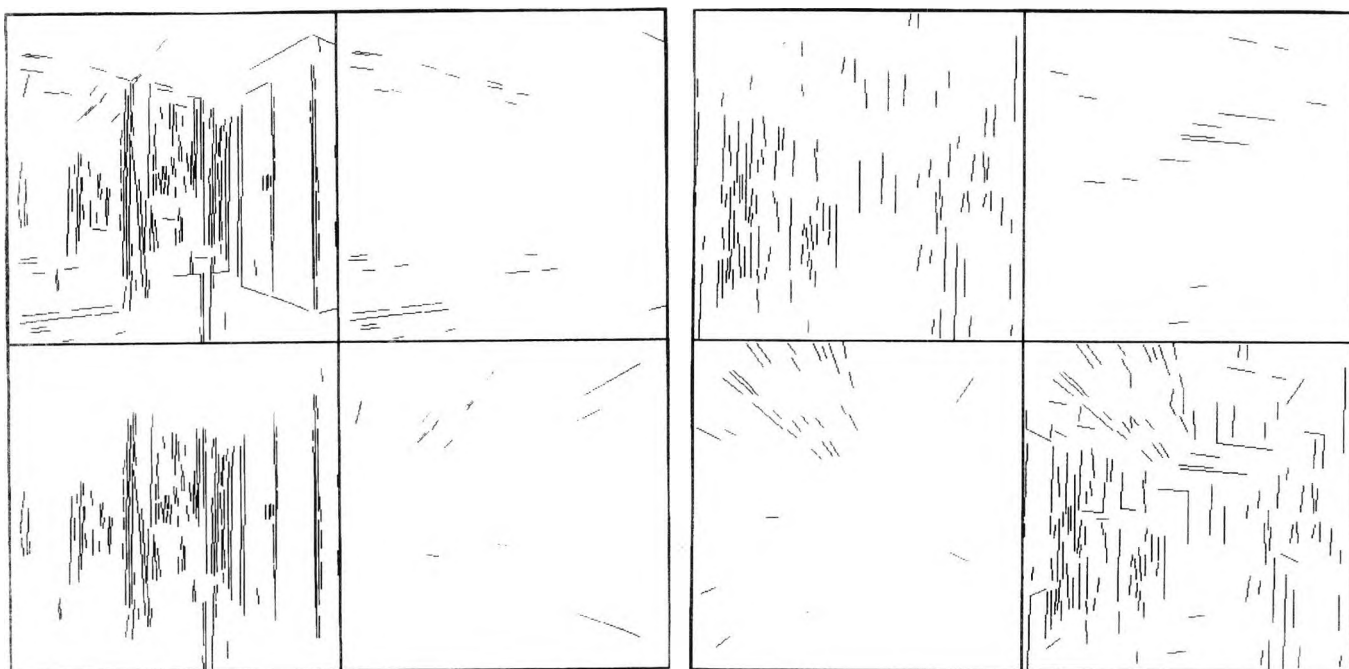


Figure 5.6.5 : Final classification: the top and the bottom left images correspond to the main directions detected ; the bottom right images show all the lines classified with a main direction.

```

object number : 1
number of lines : 31
  expected number of directions : 6
  score : 0.78
  perp. classes and associated scores :
    5  0.71
    6  0.70
   11  0.73
   13  0.75
  Vx, Vy and covariance matrix :
    -442.04   89.44 11839.77   197.37   367.64
  element number : 1
    xa, ya, xb, yb :
      214 21 248 17
  map, a, b and covariance matrix :
    0 -0.118 46.24 0.0045 240.91 -1.04
  score value : 0.69

```

Figure 5.6.6 : Extract of the classification of a line with a vanishing point (image 1,  $V_h$ ).



Detection of principal directions

```
perp. triplets and scores :  
1 6 13    0.608  
1 5 13    0.605
```

Figure 5.6.7 : Result of the search for a triplet of perpendicular directions (image 1).

Images of various indoor scenes with various viewpoints have been processed, some examples of the results obtained are given in the next figures. It has not been always possible to extract a triplet of perpendicular directions, but it has nearly always been possible to extract two of them.

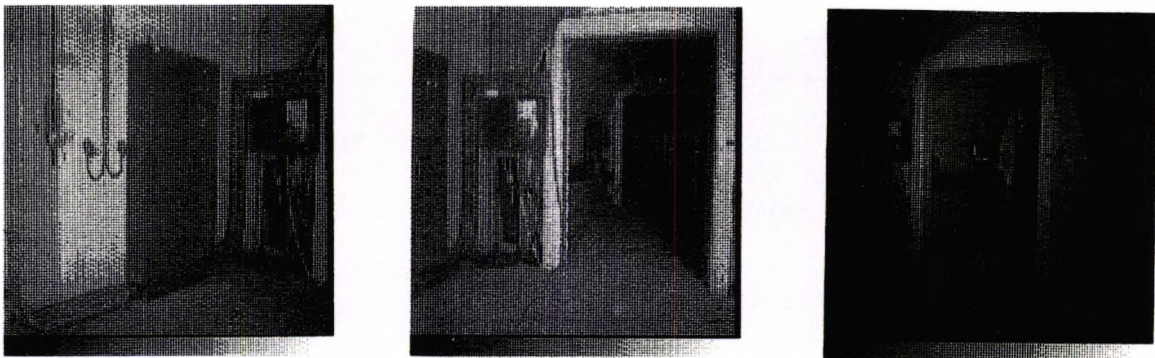


Figure 5.6.8 : Initial images

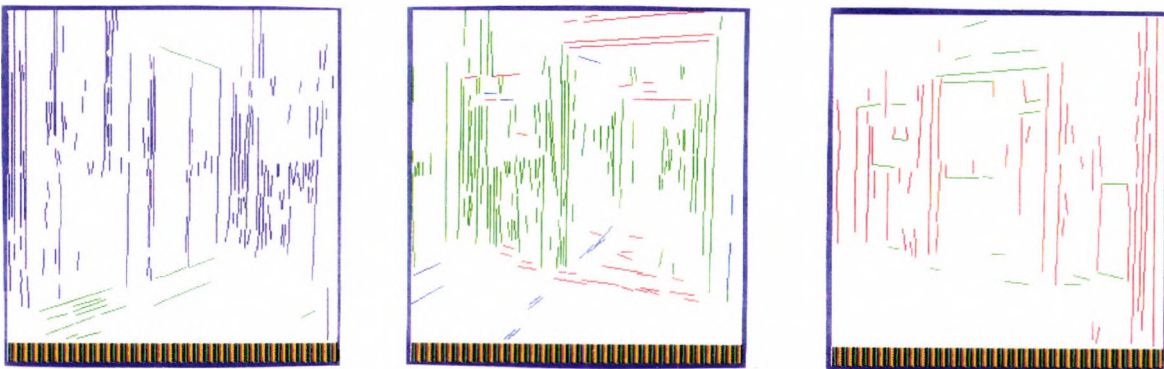


Figure 5.6.9 : Perpendicular directions

## 5.7 CONCLUSION

A new accumulator space for the detection of the vanishing points has been presented. This accumulator space is isotropic and bounded ; moreover it ensures a constant quality of the detection of the vanishing points over the space. In addition it is easily implemented. It has been compared with other methods, using the Gaussian sphere as an accumulator space.

Only significant peaks are detected, taking into account the noise produced by the accidental presence of lines near the peaks. The line segments are then classified with each of the vanishing point candidates, i.e. significant peaks of the accumulator, by using a likelihood ratio test which takes into account the parameter uncertainties associated with each segment. A scoring process provides the expected number of directions which do not cross the neighbourhood of the selected peaks by chance, and a measure of the reliability of these peaks.

An additional filter is provided by taking into account prior knowledge about the perpendicularity of the main directions. This process compensates for the unreliability of some vanishing point candidates, when associated with few directions, by the quality of the perpendicularity to more reliable vanishing points. It does not filter out reliable candidates even if their perpendicularity is poor.

All tests are based on the principle of the maximum of likelihood, which does not introduce arbitrary thresholds. The only parameters to determine are the covariance matrix elements associated with the data and the parameter  $a=E(1/\ell^2)$ , which may be determined by statistical means. Because of their physical meanings, the order of magnitude of these quantities is generally known. In other words, the thresholding parameters, usually present in any process, have been replaced by physical parameters, such as covariance matrix elements, which are included in statistical reasonings. The consistency of the various reasonings is ensured by the reference to the same probabilistic model

## Detection of principal directions

of the locations of the straight line segments in the image (see section 4.4). This results in a more robust and predictable method that works on a large number of different images of indoor scenes, independently of their complexity.

## CHAPTER 6

## HIGH-LEVEL 3-D CONFIGURATIONS

## 6.1 OVERVIEW OF THE METHOD

For matching a 3D model of an object with the features extracted from the image it is necessary to transform the representations of the model and/or of the features to obtain comparable representations. Two approaches may be proposed : either the model is projected onto the image and the matching is performed between 2D features, or the image features are back-projected into the model space and the matching is done at a 3D level. The advantage of the latter approach is that the back-projection of the image features into the model space makes explicit the geometrical constraints linked to the consistency of the 3D objects, e.g. the equality of the opposed edges of a rectangle.

A method for building a 3D representation of the image features through back-projection is described in this chapter. This approach is fundamentally equivalent to the former approach but it allows a much more powerful representation of the information contained in the image.

*Previous work*

A number of systems propose to match a 3D model with a single view (Brook, 1984), (Lowe, 1985). The common approach is to project the model onto the image over a range of viewpoints (e.g. to obtain an aspect of the object) and to perform the matching between the projected model features and the image features. The representation of the different aspects of a 3-D model has been the subject of much research (Minsky, 75; Brooks, 1984). Brooks (1984) does not predict all the instances of an object but rather quasi-invariant features. Lowe (Lowe, 1985) detects significant groupings, called perceptual groupings, which are matched with similar groupings in the model using a prediction-verification method. Mohan et al. (Mohan et al, 1989) also use the idea of perceptual

## High-level 3D configurations

organisation to extract 2D high level structures e.g. rectangles, for stereovision. Knowledge representation is a major concern for all these systems. The data are structured in the form of frames, schema or graphs corresponding to a particular view-point. Perceptual groupings are performed on the image to obtain a high level of representation and reduce the combinatorix at the matching stage. In spite of finding viewpoint invariant features, the viewpoint dependency is inherent to the type of representation, which remains 2D.

Ballard (1982), Quan et al. (1989), Kanakani (1989) and other researchers interpret perspective transformation for constraining the set of viewpoints. This is a very important constraint which relies heavily on statistical inference, i.e. lines in the image are unlikely to meet at the same point by accident. The matching is still performed at a 2D representation but exploiting additional viewpoint invariance properties such as the bi-ratio (Quan et al, 1989) or the angle relationships (Shakunaga, 89).

Kanade (1981) used parallelism as well as symmetry considerations to recover the 3D shape of objects. Nelson et al (1985) proposed a least-slant-angle heuristic to predict the orientation of an object face. The construction of an object using such techniques is based on connectivity criteria. For a simple scene, it may be inferred from connectivity in the image using a simple criterion, but for a complex scene errors are likely to occur and it is important to minimize them.

### *Problem definition*

Image interpretation is concerned here with complex indoor scenes. For such scenes, the notion of an object, e.g. a wall or a window, is ambiguous. For example the window frame may be either part of the wall or part of the window. The scale choice may also modify the object definition. Most of the time only part of the object is visible. To allow a large degree of flexibility at the matching stage, a high level representation of the image knowledge is required.

Triplets of perpendicular directions have been found in the image. The

best scored triplet is assumed to correspond to the main directions of the scene, e.g. the wall limits. All lines classified with these directions have a known orientation in the space ; only their depth is unknown. Proximal lines are grouped in order to form rectangles, vertices or edges, which are further grouped to form still higher-level configurations. This construction relies very much on connectivity : if two segments are close in the image they are likely to be connected in the scene. Because of numerous hidden faces and thereby numerous accidental proximity relationships in the image, many errors may occur. These errors are minimized by defining the connectivity criterion by a likelihood ratio test.

The construction of these 3D high-level structures is organised in a hierarchical way. If a grouping is false, sub-groupings may still be correct and therefore must be remembered. The representation used is (as much as possible) viewpoint and scale invariant so that matching is straightforward. The more complex the 3D structures, the more constraints on the matching. The only unknown parameter of these structures is depth (i.e. scale), which will be deduced from matching in a latter stage. The viewpoint invariance of parameters and relations is inherent to the type of representation and thereby is fully exploited in a natural way.

Section 6.2 describes the connectivity criterion used for the construction. Then, section 6.3 describes the set of 3D structures built. The method has been tested on indoor scenes and the results are discussed in section 6.4.

## 6.2 CONNECTIVITY

First, close lines of the same class, i.e. lines which are parallel in the 3D world, are merged, in order to restore the connectivity of long edges and to simplify the representation. A set of 3D parallel lines merged together is called a linear structure, the representation of which will be detailed in section 6.3. Then, perpendicular linear structures, assumed to be connected in the scene, are grouped in order

to form higher level structures. The connectivity criterion associated with each case is defined in this section, followed by a description of the merging process.

### 6.2.1 Merging lines parallel in the 3D world

Merging close lines which are parallel in the 3D world aims at restoring the connectivity of the edges which may have been broken by thresholding of the maxima of the gradient, or by eliminating small segments after having extracted the lines (see chapter 4). This merging process also aims at simplifying the final representation. Indeed, close parallel lines may be hypothesized to belong to the same 3D structure and would better be represented by one line. For example, the mouldings of a frame door may generate numerous close parallel lines in the image, which are useless for the semantic interpretation of the scene and unlikely to be present in the model of the scene. When restoring connectivity, the segments cannot be overlapped, but when simplifying the representation they may overlap (figure 6.2.1.1).

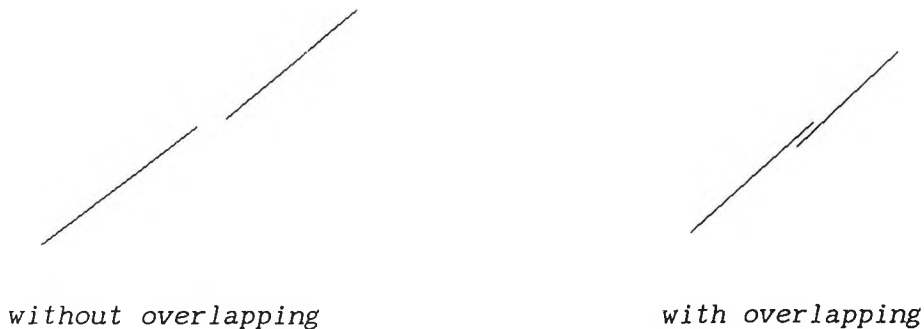


Figure 6.2.1.1 : Example of close parallel segments, with and without overlapping.

The restoration and simplification processes are complementary. The criteria for restoring or simplifying are based on two different decision variables,  $D_0$  and  $D_1$ , where  $D_0$  is the transverse distance between the segments candidates and  $D_1$  is the longitudinal distance between these two segments (see figure 6.2.1.2). The straight line segments are candidates for the edge restoration if  $D_1 \neq 0$ , otherwise they are candidates for the simplification process.

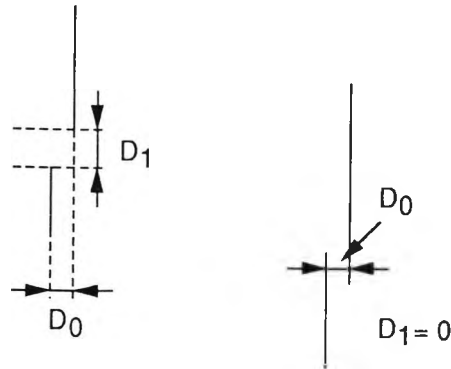


Figure 6.2.1.2 : Definition of  $D_0$  and  $D_1$

Two straight line segments  $S_1$  and  $S_2$  are merged if they satisfy both tests for the restoration (if  $D_1 \neq 0$ ) and the simplification.

*Definition of the simplification test*

The simplification process is arbitrary by nature, since it is not associated with an uncertainty but with details unnecessary to the interpretation. Close parallel lines are assumed to belong to the same 3D structure, but little information is available to quantifying "close". In order to control the process, an upper boundary,  $k\sigma_d$ , of the distance between such "close" lines is fixed, where  $\sigma_d$  is the equivalent of an uncertainty and is defined in the following. As the simplification test is performed with colinear segments assumed to belong to the same edge, it should also deal with the transverse uncertainty. Thus, by similitude with the distribution of the transverse uncertainty, the distribution of the distance between two lines belonging to the same 3D structure is assumed to be a Gaussian law, with the uncertainty  $\sigma_d$ . Therefore  $\sigma_d$  is assumed to be at least equal to  $\sqrt{2} \cdot \sigma_0$  (as the variance of the transverse error is  $2\sigma_0^2$ ), where  $\sigma_0$  is the transverse uncertainty defined in section 4.4.1. The likelihood test may now be defined.

Let the straight line segments  $S$  and  $S'$  correspond to  $(D_0, D_1=0)$ . Then the hypothesis  $H$  is defined by : the segments  $S$  and  $S'$  are close parallel straight line segments in the scene. If the hypothesis  $H$  is verified, then, as been seen previously,  $D_0$  is assumed to obey a



## High-level 3D configurations

Gaussian law with variance  $\sigma_d^2$ . Otherwise,  $D_0$  is assumed to correspond to a uniform law. Thus the densities of probability  $p_1$  and  $p_2$  are

$$p_1(D_0) = \frac{1}{\operatorname{erf}(k) \sqrt{2\pi} \sigma_d} \exp\left(-\frac{D_0^2}{2\sigma_d^2}\right) \quad \text{and} \quad p_2(D_0) = \frac{1}{k \sigma_d} \quad (6.2.1.1)$$

To estimate the ratio  $p(H)/p(\bar{H})$ ,  $\bar{H}$  has to be made explicit. The segments  $S$  and  $S'$  verify  $\bar{H}$  if, either

- one of the segments or both have been misclassified,
- or the two segments are close in the image by chance.

Let  $H_0$  be the hypothesis that the straight line segment  $S$  has been correctly classified (respectively  $H'_0$  for  $S'$ ), and let  $H_1$  be the hypothesis that  $S$  and  $S'$  are close in the 3D space. Then,  $p(H)/p(\bar{H})$  is equal to

$$\frac{p(H)}{p(\bar{H})} = \frac{p(H_0)p(H'_0)p(H_1)}{p(H_1)(p(\bar{H}_0)p(H'_0)+p(H_0)p(\bar{H}_0)+p(\bar{H}_0)p(\bar{H}'_0)) + p(\bar{H}_1)} \quad (6.2.1.2)$$

The probabilities  $p(H_0)$  and  $p(H'_0)$  are given by the scores associated with the classification of  $S$  and  $S'$  with the class of 3D parallel lines studied (see section 5.3). The probability of  $H_1$  is deduced from the probability of  $\bar{H}_1$ . To compute the probability of  $\bar{H}_1$ , two cases  $\mathcal{C}_1$  and  $\mathcal{C}_2$  are considered whether it is (case  $\mathcal{C}_1$ ) or it is not possible (case  $\mathcal{C}_2$ ) to have two segments  $S'$  and  $S''$  lying on the same side of  $S$  and at the same distance  $D_0$  from  $S$ . According to the statistical model defined in section 4.2, the centroid of the segments are assumed to have a uniform distribution in the image. Let  $P$  be the centroid of  $S$ ,  $B$  be the band of the image parallel to  $S$  with centre  $P$  and width  $k\sigma_d$ , and  $n'$  be the expected number of straight line segments in the band  $B$  of the image. The probability of having a segment  $S'$  in  $B$  such that  $D_1 = 0$  is

$$p(\bar{H}_1) = 1 - \left(1 - \frac{n'(\ell + \ell')}{n h}\right)^n \quad (6.2.1.3)$$

where  $h$  is the height of the band  $B$ . As  $n'$  is proportional to the surface of the band  $B$ ,  $n'k/h = n\sigma_d/A$ , where  $A$  is the area of the image.

In the case  $\mathcal{C}_2$ , this probability is equal to

$$p(\bar{H}_1) = n k \frac{\sigma_d (\ell + \ell')}{A} \quad (6.2.1.4)$$

The real case is intermediate between cases  $\mathcal{C}_1$  and  $\mathcal{C}_2$ . If  $\sigma_d = \sqrt{2}\sigma_0$ , then the case to consider is  $\mathcal{C}_2$ , if  $\sigma_d$  is very large then it is  $\mathcal{C}_1$ . However, if  $k\sigma_d(\ell + \ell')/A$  is small, then eq. 6.2.1.3 and 6.2.1.4 are equivalent. Then when  $p(\bar{H}_1)$  is small enough, eq. 6.2.1.4 is valide, which is supposed to be the case here. Thus, the LR test is completely defined.

From eq. 6.2.1.2 and 6.2.1.4, the likelihood ratio test defined is equivalent to

$$\frac{D_0^2}{\sigma_0^2} \leq T(n, \ell, \ell', p(H_0), p(H'_0)) \quad (6.2.1.5)$$

where  $T$  decreases when  $n$ ,  $\ell$  and  $\ell'$  increase, and increases when  $p(H_0)$  and  $p(H'_0)$  increases. The more numerous the class, the more selective the test. This allows the relative importance of each class in the image to be kept. As  $p(H_0)$  and  $p(H'_0)$  also depends on  $\ell$  and  $\ell'$ ,  $T$  decreases when  $n$ ,  $\ell$  and  $\ell'$  are large enough and increases when  $V$  and  $V'$  decrease,  $V$  and  $V'$  being the decision variables associated with  $S$  and  $S'$  for classifying them with the vanishing point considered. Therefore if  $V$  and  $V'$  are small, the quality of the parallelism of the corresponding 3D lines is likely to be very good, then the lines are likely to belong to the same structure and it is normal to have a tolerant test for  $D_0$ . If the segments are long, then they should be very close in order to be merged. This is very good, since if the segments are long, they usually are very significant and to merge such parallel segments is dangerous because the risk that they belong to different structures is relatively high (i.e.  $p(\bar{H}_1)$  is high) and merging them would result in a substantial error of localisation for one or the other structure. It is more natural to group the small segments around a large one, than to group large segments together. Therefore, the test defined behaves in a very satisfactory way.

*Definition of the restoration test*

Let the straight line segments  $S, S'$  correspond to  $(D_0, D_1 \neq 0)$ . The hypothesis  $H$  tested is : the straight line segments  $S$  and  $S'$  are colinear and connected in the 3D world. If  $H$  is verified then  $D_1$  is assumed to correspond to an exponential law with expected value  $2\sigma_1$ ,  $\sigma_1$  being defined in section 4.4.1, but if they are close by chance, then  $D_1$  is assumed to correspond to a uniform law. For avoiding the difficulties due to the image bound, the set  $\Lambda$  of pairs of segments considered here is such that  $(D_0 < k\sigma_d, D_1 < 2k\sigma_1)$ , the segment  $S$  and the endpoint considered being fixed. This has the additional advantage of fixing an upper boundary to the possible values for  $D_1$ , useful when the segmentation error risk is very low. The value of  $k$  is such that if  $\Lambda$  is known to contain one pair, the probability that it contains an additional pair is nearly zero. Typically  $k$  is equal to 2, corresponding to the risk of missing a pair of segments connected in the 3D space equal to 5%. Thus,  $p_1$  and  $p_2$  corresponding to  $H$  are

$$p_1(D_1) = \frac{1-e^{-k \frac{D_1}{\sigma_1}}}{\sigma_1} \exp\left(-\frac{D_1}{\sigma_1}\right) \quad \text{and} \quad p_2(D_1) = \frac{1}{2 k \sigma_1} \quad (6.2.1.6)$$

The ratio  $p(H)/p(\bar{H})$  is given by eq. 6.2.1.2, where  $H_0$  and  $H'_0$  are defined as previously and  $H_1$  is the probability of  $S$  and  $S'$  belonging to colinear structures. To compute the probability of  $\bar{H}_1$ , two cases  $\mathcal{C}_1$  and  $\mathcal{C}_2$  are considered whether it is ( $\mathcal{C}_1$ ) or it is not ( $\mathcal{C}_2$ ) possible to have two segments  $S'$  and  $S''$  lying on the same side of  $S$  and at the same distance  $D_1$  from  $S$ . According to the statistical model defined in section 4.2, the centroids of the segments are assumed to have a uniform distribution in the image, so are the set of the first endpoints and the set of the second endpoints, where the first endpoint is defined as the endpoint closer to the vanishing point. Let  $a$  be the area of the rectangle with centre  $P$ , length  $k\sqrt{2}\sigma_1$  and width  $k\sigma_d$ . In case  $\mathcal{C}_1$ , if  $a$  is assumed to contain one first endpoint, then the probability for  $a$  to contain one second endpoint is

$$p(\bar{H}_1) = 1 - \left(1 - \frac{a}{A}\right)^n \quad (6.2.1.7)$$

where  $A$  is the area of the image. In case  $\mathcal{C}_2$ , this probability is equal to

$$p(\bar{H}_1) = n \frac{a}{A} \quad (6.2.1.8)$$

The real case is intermediate between cases  $\mathcal{C}_1$  and  $\mathcal{C}_2$ . If  $\sigma_d = \sqrt{2}\sigma_0$ , then the case to consider is  $\mathcal{C}_2$ , if  $\sigma_d$  is very large then it is  $\mathcal{C}_1$ . However, if  $a/S$  is small, then eq. 6.2.1.7 and 6.2.1.8 are equivalent. Now, by hypothesis  $a/S$  is small, so that eq. 6.2.1.8 is valid and the likelihood ratio test is completely defined.

Thus, the likelihood ratio test defined is equivalent to

$$\frac{D_1^2}{\sigma_1^2} \leq T(n, p(H_0), p(H'_0)) \quad (6.2.1.9)$$

where  $T(n, p_0, p'_0)$  decreases when  $n$  increases, and increases with  $p(H_0)$  and  $p(H'_0)$ . As previously, the more numerous the class, the more selective the test. The less reliable the straight line segments, the more selective the threshold. It has been seen that the small segments are associated with a small value for  $p(H_0)$  in section 5.3, therefore merging two small segments requires a small distance  $D_1$ , which is very satisfactory as the evidence for a long linear structure is low in such a case. If segments are long but associated with a low score, either they have been misclassified, or they are unlikely to be exactly parallel in the 3D scene and thereby to belong to the same object. In this case the LR test is more selective, which is what was expected. On the contrary, segments with a high score are necessarily long and with a good parallelism, and the LR test is tolerant for  $D_0$ , i.e. the length of the gap between them. Once again the LR test behaves in the expected way.

### *Discussion*

The definitions of the density of probability  $p_1$  for the simplification stage are somewhat arbitrary. Actually, the distribution of close lines, corresponding to various details of the same structure

## High-level 3D configurations

is unknown. The Gaussian model describes the simplification process, rather than a physical reality. However, it is able to deal properly with the uncertainty defined by  $\sigma_0$  in chapter 4. The exponential model associated with the length of the gaps between two line segments of the same edge is justified in a similar way to the distribution of the segment lengths (see section 4.4.2).

In the above definitions of the tests, the convergence of the segments of the same class due to perspective is assumed negligible when these segments are close, i.e.  $D_0$  is assumed constant. This assumption is largely justified compare with the other assumptions already made, such as the distribution of  $D_0$  for the hypothesis ( $H_1$ ).

The separation of the tests allows emphasis of the importance of the endpoints in a connectivity test. It is noticeable that the longitudinal merging is easier if the segments are long and that the converse is true for a transverse merging, and that it is a satisfactory behaviour in both cases.

Thus, because of the simplifications made, the model chosen is not claimed to represent the full complexity of the problem of connectivity, but is rather a reasonable guide for finding adaptative tests which behave in a satisfactory way.

### *Merging process*

Before performing the LR test, the segments are sorted by scores. Once two straight line segments have succeeded the LR tests defined above, then they are merged by using the Kalman filter, in a way similar to (Ayache, 1988). The Kalman filter provides the uncertainty of the line parameters and the longitudinal uncertainty of the endpoints is still  $\sigma_1$ .

### 6.2.2 Grouping lines perpendicular in the 3D world

The purpose of this section is to define a test for selecting perpendicular lines, likely to belong to the same planar structure in

the scene, e.g. a rectangle. The straight line segments have been merged into linear structures, as described in section 6.2.1. Therefore the connectivity criterion described here applies to perpendicular linear structures.

A similar reasoning to that previously is applied to the distances  $D$  and  $D'$ , between the segment  $S$  and  $S'$  (see figure 6.2.2.1). For computing the ratio  $p(H)/p(\bar{H})$ , it is considered that two parallel linear structures cannot lie in the neighbourhood of interest by construction (due to the simplification process, see section 6.2.1). Actually, it is not strictly true, since the LR test is not a neighbourhood test, but it is sufficient to consider only case  $\mathcal{C}_2$  described in section 6.2.1. Again, for defining the LR test, two cases are considered, whether  $D$  is equal to zero or not (figure 6.2.2.1).

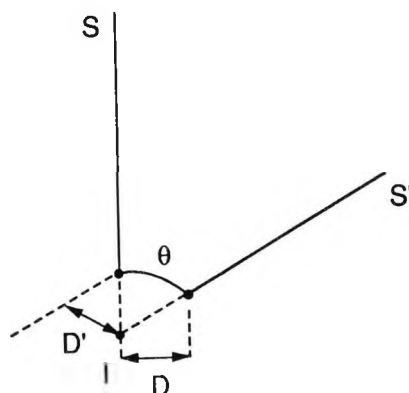


Figure 6.2.2.1 : Definition of the distance  $D$  and  $D'$  associated with the images of 2 perpendicular segments

#### *Intersection point lying on S*

If the intersection point of the straight line segments  $S$  and  $S'$  lies on  $S$ , then  $D=0$ . Either it also lies on  $S'$  and  $D'=0$ , in which case the segments are assumed connected in the 3D space. Or it does not lie on  $S'$ , i.e.  $D' \neq 0$ . Similarly to previously,  $D'$  is bounded by  $k\sqrt{u}$ , where  $u$  is the variance of  $D'$ . The density of probability  $p_1$  of  $D'$  is a Gaussian law with variance  $\sigma_{D'}^2 = (\sin\theta)^2\sigma_1^2 + \sigma_t^2$ , where  $\theta$  is the angle between  $S$  and  $S'$  and  $\sigma_t$  is the transverse uncertainty of  $S$ . The density  $p_2$  is uniform

## High-level 3D configurations

$$p_1(D') = \frac{2 \operatorname{erf}(k)}{\sqrt{2\pi} \sigma_{D'}} \exp\left(-\frac{D'^2}{2 \sigma_{D'}^2}\right) \quad p_2 = \frac{1}{k \sigma_{D'}} \quad (6.2.2.1)$$

The ratio  $p(H)/p(\bar{H})$  is given by eq. 6.2.1.2, where  $p(H_0)$  and  $p(\bar{H}_0)$  are defined as in section 6.2.1 and  $p(\bar{H}_1)$  is

$$p(\bar{H}_1) = n \frac{k \sigma_{D'} \ell}{A} \quad (6.2.2.2)$$

where  $n$  is the number of linear structures parallel to  $S$  and  $A$  the area of the image. Thus, the LR test is completely defined

### *Intersection point outside of the segments $S$ and $S'$*

Supposing the intersection point of the straight line segments lies outside of both segments  $S$  and  $S'$ , i.e.  $D \neq 0$  and  $D' \neq 0$ , let  $V$  be  $(D, D')$  and  $C_V$  the covariance matrix associated with  $V$ , equal to the sum of the variances associated with  $D$  and  $D'$ . The densities of probability are defined as eq. (6.2.2.1)

$$p_1(V) = \frac{2}{\operatorname{erf}(k) 2\pi \sqrt{\det(C_V)}} \exp\left(-\frac{V^t C_V^{-1} V}{2}\right) \quad p_2(V) = \frac{1}{k \sqrt{\det(C_V)}} \quad (6.2.2.3)$$

The probability  $p(\bar{H}_1)$  is now equal to

$$p(\bar{H}_1) = n n' \frac{k^2 \det(C_V)}{A} \quad (6.2.2.4)$$

where  $n$  is the number of linear structures parallel to  $S$  and  $n'$  is the number of linear structures parallel to  $S'$ .

### *Discussion*

Again, the test is tolerant when there are few structures. The test for connecting the linear structures through the endpoints is more tolerant than the test for connecting the structures along one of the segments. This is due to the probability of false connection which is higher in the latter case, which is easy to admit.

The model adopted for  $p_1$  is Gaussian, because it allows the 2D

round-up effect to be more easily dealt with than by using an exponential model (because of the easy generalisation of the Gaussian law to the 2D case). Practically, the difference of results due to the choice of this law is negligible compared with the difference of the results when using various values  $\sigma_1$ . This result must be compared to similar results for the relative importance of the shape of the filter and the choice of the parameter, described in chapter 4.

Thus, a LR test has been defined for grouping structures assumed perpendicular and connected in the 3D world. This test will be used to construct higher level primitives, such as rectangles or vertices.

### 6.3 3D STRUCTURES

At each level of the direction interpretation tree (3D direction pair of perpendicular directions, triplet of perpendicular directions), it is possible to associate 3D structures (3D lines, 3D rectangles and ellipses, vertices and edges) by using proximity criteria (figure 6.3.1).

Parallel and perpendicular groupings	3-dimensional structures
Triplets of perpendicular directions and associated lines	Vertices and adjacent structures Edges and adjacent structures
Pairs of perpendicular directions and associated lines and ellipses	Rectangles "Comb" structures (U,L,T...) Circular arcs
Parallel lines	Linear structures

Figure 6.3.1: Construction of 3-D structures from the direction interpretation tree.

The perpendicularity of the principal directions are clearly not perfect. As the representation looked for aimed at being symbolic,



## High-level 3D configurations

including forms such as rectangles and rectangular corners, it is necessary to correct for the lack of perpendicularity of the directions involved. This correction is achieved by a LMS method applied to the directions found, using the covariance matrices associated with them (they are deduced from the vanishing points covariance matrices). Thus, the parallelism or perpendicularity involved in the structures described in the following is quasi perfect.

In the following, the structures are described in the 3D scene with respect to the camera coordinate system. The only reference to the image is concerned with the connectivity test.

### *Linear structures*

Main directions of the scene are supposed to be known from the method described in chapter 5 and image lines are associated with each of them. A merging process of these lines allows 3D linear structures to be built. A linear structure is assumed to represent a 3D straight line segment present in the scene.

Two types of linear structures are created, the linear structures obtained from a set of close parallel segments in the image and the linear structures obtained from an alignment of points with a vanishing point. The latter structures are called subjective structures.

The former type of linear structure is obtained by merging close segments which are parallel in 3D space, by using a merging process similar to the process described in (Ayache,1988). Two segments are parallel in the 3D space if they are classified with the same vanishing point and they are close if they succeed the likelihood test defined in sub-section 6.2.

When numerous and close parallel lines end on the same perpendicular line, the round-up effect of the edge detector prevents the perpendicular line from being extracted. To recover such lines which may have an important part in the scene (they often are the limit of a wall or a main structure) a linear structure has been created called a

subjective linear structure. A subjective linear structure is obtained by accumulating the direction of the segments defined by the end points of an image segment and a vanishing point. Only the image segments classified with a direction perpendicular to the direction defined by the vanishing point are considered (figure 6.3.2).

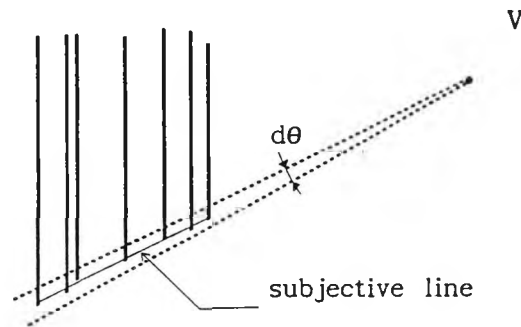


Figure 6.3.2 : Example of a subjective linear structure

A linear structure is defined by its orientation, its centroid position and its length within the camera coordinate system. These parameter values are deduced from the location of the corresponding segment in the image and the vanishing point coordinates. Only the scale of this representation, i.e. the depth of the structure, is not known. Let  $(C, \vec{I}, \vec{J}, \vec{K})$  be the coordinate system associated with the camera, where  $C$  is the optic centre,  $(C, \vec{K})$  the optic axis and  $(C, \vec{I})$  and  $(C, \vec{J})$  the axes parallel to the image coordinate system. Furthermore, let  $f$  be the algebraic distance from  $C$  to the image plane and  $O$  the principal point (i.e.  $\vec{CO} = f \vec{K}$ ). A 3D linear structure is represented in this coordinate system by :

Linear structure : associated vanishing point  $V$  ;  $\vec{CV} = \vec{OV} + f \vec{K}$

Orientation	$\vec{U} = \frac{\vec{CV}}{ \vec{CV} }$
Centroid/depth	$\frac{\vec{CG}}{d} = \frac{\vec{CM}}{f} + \frac{ \vec{AB} ^2}{4f \vec{MV} ^2} \vec{MV}$
Length/depth	$\frac{L}{d} = \frac{ \vec{AB}   \vec{CV} }{f  \vec{MV} }$
Depth	$d = \text{unknown}$

## High-level 3D configurations

where A and B are the end points of the image segment corresponding to the linear structure, and where M is the mid-point of the segment AB (M is not the projection of G). The centroid location and the length are represented by the "reduced" coordinates, i.e. the ratios of the centroid coordinates to the depth, and the "reduced" length, i.e. the ratio of the length to the depth. By definition the depth is the third centroid coordinate ; it defines the scale of the above representation.

Therefore, a set of close segments in the image, supposed parallel in the 3D scene, has a 3D representation parameterized by depth. This representation enables the correction of the distortions due to the perspective, e.g. recovering of the real centroid of the segment.

### *Circular structures*

An ellipse in the image is assumed to be the projection of a circle in the scene. This circle is assumed to be lying in one main plane, that is a plane associated with two principal perpendicular directions (see section 5.5).

A circular structure is a 3D arc characterised by two directions, its centroid location and its radius. These parameters are deduced from the location of the ellipse in the image and the associated vanishing point coordinates.

Circular structure associated with the vanishing points  $V_1$  and  $V_2$

Orientations	$\begin{cases} \vec{U}_1 \\ \vec{U}_2 \end{cases}$
Centroid/depth	$\frac{\vec{CG}}{d} = \frac{\vec{Og} + f \vec{K}}{f}$
Radius/depth	$\frac{R}{d} = \frac{ AB   CV_1 }{2 f  MV_1 }$
Depth	unknown

where A, B, M and g are in the image and are defined in section 5.5 and the orientations are defined as above.

### Rectangular structures

There are 2 types of rectangular structures : the "comb" structures including U,L and T shape structures and the rectangular structures.

The "comb" structures are formed from a linear structure, called the principal structure, and the perpendicular linear structures that are in close proximity, called the "teeth". Two linear structures are close if they succeed the LR test defined in section 6.2. There is no notion of a "cross" structure, therefore a "comb" structure is the set of the principal linear structure and the associated perpendicular structures (the teeth of the comb) lying in a half plane defined by the principal structure. Two "comb" structures are created in the case of a cross configuration (see figure 6.3.3).

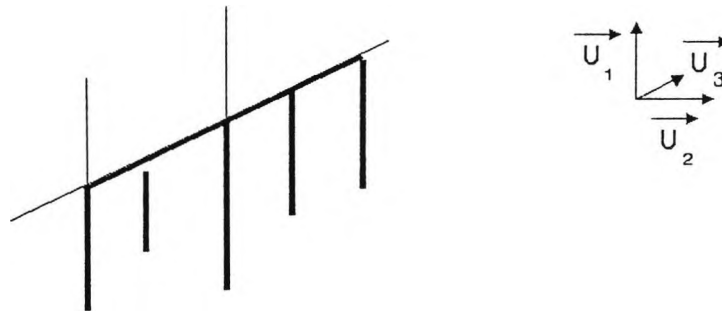


Figure 6.3.3 : Example of comb structures

The "comb" structures are much less numerous than the set of specific structures such as L,T,U and X shape structures ( $O(np)$  for L, T and X shapes or  $O(np^2)$  for U shapes against at most  $2n$  for the comb structures, where  $n$  is the number of linear structures associated with the direction considered and  $p$  is the number of linear structures associated with a perpendicular direction). They are more significant as they usually represent parts of a plane and not any specific structure, the detail of which would often be omitted in the model.

The comb structure is the basis of the following construction, i.e. any structure described in the following is a set of comb structures which have a common linear structure in a way which will be described later. Therefore, the further reasoning is exclusively in 3D space.

## High-level 3D configurations

A rectangular structure is defined by two comb structures such that the principal linear structures associated with them are parallel and they have at least one tooth in common. If they have several teeth in common, only the largest rectangle is generated.

A comb or rectangular structure is characterised by its two directions, its centroid location and its length and width. That is to say :

Comb or rect structure associated with the vanishing points  $V_1$  and  $V_2$

orientations	$\vec{U}_1$ $\vec{U}_2$
centroid/depth	$\frac{\vec{CG}}{d}$
half plane	+1 or -1 (in case of comb structure)
{ $d_i$ /depth}	$d_1 \dots d_n$ (in case of comb structure)
length1/length2	$\lambda = L_1/L_2$
area/depth <sup>2</sup>	$\alpha = L_1 L_2 / d^2$
depth	$d = \text{unknown}$

where the variables are :

- comb structure :  $\{d_i\}$  are the algebraic distances between the perpendicular structures and the centroid of the principal structure (in the 3D space),  $L_1$  is equal to  $d_n - d_1$  and  $L_2$  is the maximum length of the teeth, where the teeth are consistently scaled (figure 6.3.4). Here the centre is the centroid of the principal linear structure.
- rectangle :  $L_1$  is the length of one edge and  $L_2$  is the length of the other edge, with the "comb" structures consistently scaled. The point G is defined from the centroid  $G_1$  of one edge (corresponding to the direction  $\vec{U}_1$  and the length  $L_1$ ) such that  $\vec{GG}_1$  is parallel to  $\vec{U}_2$  with a length equal to  $L_2/2$ . It is calculated using the 3D representation associated with the comb structures.

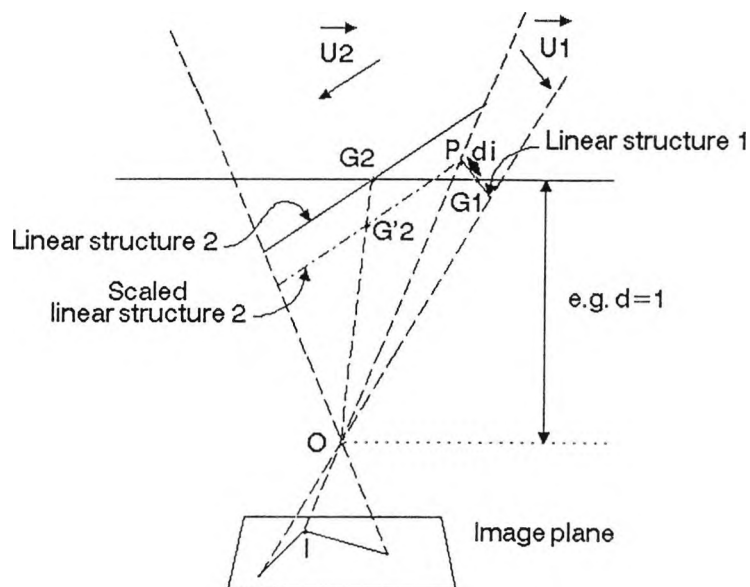


Figure 6.3.4 : Consistent scaling of a tooth of a comb structure

#### *Edges and vertices*

An edge is formed by two perpendicular comb structures which have the same principal linear structure. It is characterised by this linear structure, the two other perpendicular directions and its concavity. It is possible to hypothesize the concavity of the edge by considering the relative positions of the linear structures and the associated vanishing point in the image (see figure 6.3.6). When two directions are parallel to the image, the concavity is undetermined, case rare since the vanishing point associated with the third direction should be exactly at the principal point location (let us remember that, now, the perpendicularity of the directions is quasi perfect). An edge is not necessarily bounded by one or two corners as no connectivity criterion is applied to the respective ends of the comb structures.

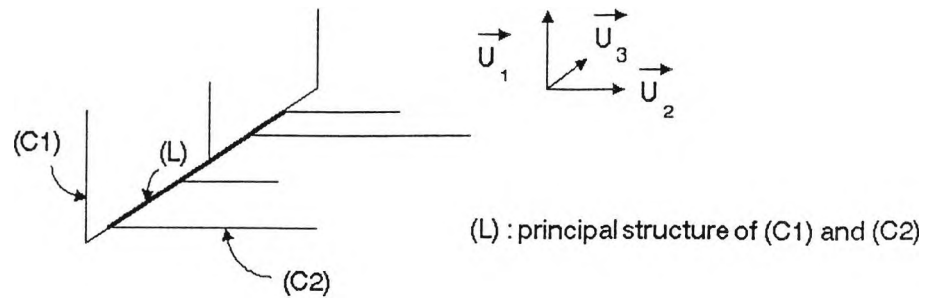


Figure 6.3.5 : Example of an edge built from 2 comb structures

A vertex is defined by a triplet of consistent perpendicular comb structures, i.e. each principal linear structure is a tooth for the 2 other comb structures. It is possible to hypothesize the concavity of the vertex and the existence of a hidden face by considering the relative positions of the linear structures, the vertex and the associated vanishing point in the image (see figure 6.3.6).

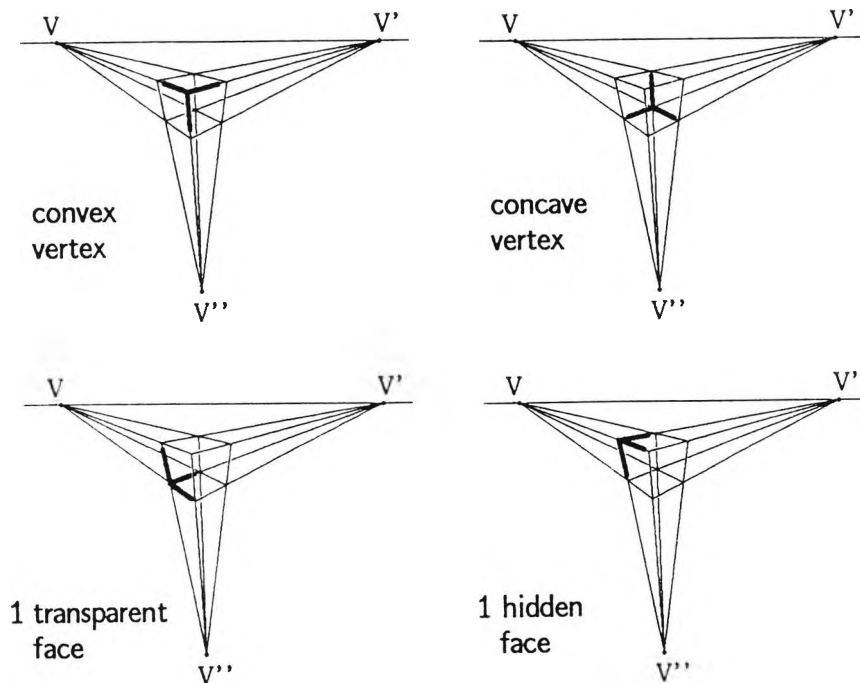


Figure 6.3.6 : Position of the vertex according to the orientations of the edges.

A vertex is characterised by :

Vertex structure associated with the vanishing points  $V_1$ ,  $V_2$  and  $V_3$

orientations	$\vec{U}_1$
	$\vec{U}_2$
	$\vec{U}_3$
Concavity	<i>concave, convex, convex with 1 hidden face</i>
corner/depth	$\frac{\vec{CG}}{d}$
lengths/depth	$L_1, L_2, L_3$
depth	$d = \text{unknown}$

where  $G$  is the intersection of the 3 principal linear structures, which are first consistently scaled. Because of uncertainty of measurement, the linear structures do not intersect exactly, and  $G$  is found by a LMS method.

### 3D configurations

The hierarchical links are recorded with each structure. For example, /a rectangle is made of two comb structures, the identity of which is recorded with the rectangle. It is therefore possible to create a 3D configuration by propagating the depth information down to the linear structures involved, then up to all structures concerned with these linear structures (i.e. structures adjacent to the initial structure).

The 3D configurations may be built from any structure. However, it would result in numerous redundant 3D configurations. We have chosen to build them only from significant features, such as the rectangular corners and the rectangles. The propagation of the depth of a structure, say  $S_1$  with centroid  $G_1$ , to adjacent structures, for example a structure  $S_2$  with centroid  $G_2$ , is a simple geometric problem, which may be decomposed into a set of elementary problems of two types :



## High-level 3D configurations

- either the point  $G_2$  is constrained to lie on the line  $(P, \vec{U}_2)$ , where  $P$  and  $\vec{U}_2$  are known (see figure 6.3.4) ;
- or the point  $G_2$  is constrained to lie on the plane  $(G_1, \vec{U}_1, \vec{U}_2)$ , where  $G_1$ ,  $\vec{U}_1$  and  $\vec{U}_2$  are known.

In the first case, the point  $G'_2$  is the intersection of the lines  $(P, \vec{U}_2)$  and  $(O, G_2)$ . In the second case,  $G'_2$  is the intersection between the plane  $(G_1, \vec{U}_1, \vec{U}_2)$  and the line  $(O, G_2)$ . In spite of the connectivity in the image, no intersection point may exist because of the correction of the perpendicularity. Therefore a LMS method must be performed.

The third coordinate of  $G'_2$  fixes the depth and hence the scale of the structure  $S'_2$ .

The propagation to the adjacent structures may be stopped or carried on until the process is stable. The result of this propagation is unreliable when there are numerous hidden faces. However, even if erroneous at times, the 3D configurations gives information about the general organization of the scene. For example, information such as "a long vertical structure (e.g. a wall) have been detected on the left" is valuable, even if the relative depths of the different objects along this structure are wrong. When no hidden face exists on the part of the scene studied, then a 3D configuration is a local 3D map, only the scale of which is unknown.

Thus, the 3D structures built are hierarchically organized, from the linear structure to the 3D configurations, which gives a local 3D interpretation of the image in terms of a set of connected symbolic forms, such as rectangles and corners.

### *3D representation*

The 3D structures have a symbolic representation defined by a number of parameters independent of the depth (the depth is the single unknown variable). This representation allows the intrinsic geometric properties of the 3D form to be implicitly used. For example,

- The use of a 3D representation associated with the linear structures has very much eased the construction of higher level primitives.
- For a "comb" structure the ratio  $d_i/d_j$  is obviously viewpoint invariant; this property is a natural translation of the viewpoint invariance of the bi-ratio used to reduce the combinatorix by some researchers (Quan et al, 1989). Moreover, only 3 points are required here against 4 in the bi-ratio.
- For a rectangular structure, the ratio of the lengths is independent of the viewpoint.
- The variations of the area parameter are bounded by the extrema of the depth, i.e. by a relation depending on the ratio of the focus distance to the depth of field of the camera.

It has been seen that the lack of perpendicularity of the triplet or pair of directions found by the vanishing point detection process must be corrected. In order to minimize the effects of a possible error during the detection of the vanishing points, this correction may only involve the directions concerned with the structure studied. For instance, no correction is made for the direction of a linear structure, and for the two directions of a rectangle a LMS solution is computed only involving these two directions (so that the plane of the rectangle is unchanged). In practice, the orientations associated with the structures are recorded using labels associated with the triplet of directions studied, so that the degree of correction of perpendicularity may depend on the results of a further process. For instance, if numerous constructed structures may be successfully matched with model structures, then the corresponding triplet of perpendicular directions is valid, and the correction is made once for all, using the three directions. But if there is not enough evidence for a triplet of directions, it is better not to propagate the possible errors. In any case, the consistency of the representation built is ensured by the use of labels.

A 3D structure is visualized by setting the depth to any arbitrary

## High-level 3D configurations

value. For convenience, this value has been chosen equal to the focal length. The lack of perpendicularity between some structures, such as the linear structures, comes from the absence of correction of their associated directions (see the previous paragraph), but a rectangle or a rectangular corner is always perfect. As an example, a 3D rectangle is visualised in figure 6.3.7.

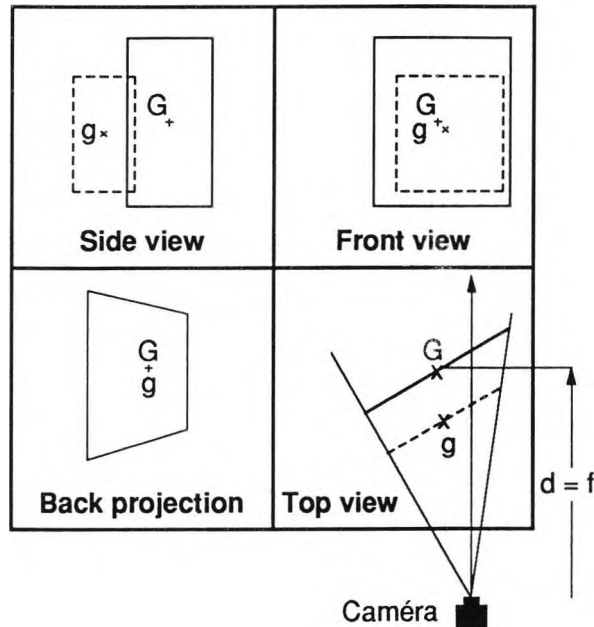


Figure 6.3.7 : Visualisation of the 3D structures

Now, the matching may be performed between two 3D representations, which avoids the extraction of the visible features from the CAD data-base and the projection of the model onto the image for a range of viewpoints and therefore it reduces the combinatorix.

It has been shown that it is possible to extract a natural 3D representation of the features extracted from the image by using a set of assumptions. The first assumption is concerned with the type of scene processed and relies on the validity of the vanishing point detection. The second assumption is concerned with the connectivity criteria. These criteria are used once for the construction of the linear structures and the comb structures. Then, the construction of higher level primitives is only concerned with the comb structures, using reasoning in the 3D

space. This construction exploits the hierarchical links between the structures, e.g. structures sharing a linear structure. At the top of this hierarchical construction are the 3D configurations which are similar to local 3D maps. The 3D geometrical properties of the features extracted are intrinsic to the 3D symbolic representation used.

## 6.4 RESULTS

The approach previously described has been applied to a set of indoor scenes of a power plant. The results are qualitative because matching with a model has not been performed. Only the most significant triplet of perpendicular directions has been taken into account. It has been possible to extract numerous significant structures from each image such as the door frame, the cupboard frame, parts of the walls and floor and corners.

The 3D structures are displayed by using orthographic projections in the camera coordinate system. They are back-projected onto the image to evaluate the consistency of the 3D location of the structure and its projection onto the image.

The entrance to an air-lock is displayed in figure 6.4.1. 107 segments have been extracted and 116 linear structures, 66 rectangular structures and 3 vertices and 4 edges have been constructed. The segments are displayed in figure 6.4.2. The linear structures are displayed in figure 6.4.3, then the U structures in figure 6.4.4, the rectangular structures in figure 6.4.5 and the corners structures in figures 6.4.6 and 6.4.7. The depth has been propagated over the first adjacent structures in figure 6.4.8. Four connected components have been found, the first component represents the air-lock entrance, the second a part of the wall and the remaining two the top and the bottom of the cupboard. This local interpretation of the scene is the highest level of the representation described. It is possible to see how the interpretation of the scene has been improved from figure 6.4.3 to figure 6.4.8. The back-projection is very satisfactory. It demonstrates that the vanishing point has been accurately detected. The lack of perpendicularity between some structures is due to the fact that it has

## High-level 3D configurations

only been corrected when necessary, e.g. for computing a rectangular structure but not for computing the linear structures.

Two different views of the same scene have been processed ; the results are displayed in figures 6.4.9 and 6.4.10. In figure 6.4.9, the shadows of the cupboard and a part of the wall behind it have been aligned with its doors. This illustrates the risk of error associated with the propagation of depth to adjacent structures. An ellipse has been found on the second one and has been interpreted as the projection of a circle onto a plane parallel to the air-lock entrance. The two representations of the cupboard front, associated with different view points (figure 6.4.12), show that the ratio of the lengths and the ratio  $d_{min}/d_{max}$  are almost identical in accordance with the viewpoint invariance of both ratios. The areas and the distances  $d_{min}$  and  $d_{max}$  are similar because the camera was approximately at the same distance (along the Z axis) from the cupboard.

A different and more complex scene is displayed in figure 6.4.13. The interpretation does not clearly show the main components of the scene such as the cylindrical tank, the blackboard or the table. Nevertheless it shows pretty well the organisation of the space : a wall on both sides with an obstacle in front of the tank and some elements on the right wall as well as volumetric elements on the ceiling (pipe structures). Some linear structures may have either horizontal direction. Both alternatives remain in the interpretation. The two parts of the horizontal pipe on the ceiling are not really perpendicular. In this case, the correction of the lack of perpendicular does not correspond to a physical reality and therefore may be subject to discussion. This is responsible for the bad quality of the back-projection. It could have been decided to correct the lack of perpendicularity only when it is low.

These examples show the limits of the interpretation. Shadows, hidden parts are subject to mistakes ; however, shadows and hidden parts depends on the viewpoint and their mis-interpretation should not persist throughout the analysis of a sequence of images. The use of

perpendicularity may introduce inaccuracy and must correspond to a reliable prior knowledge of the scene. In spite of some mistakes or inaccuracy, the interpretation of the monocular images studied gives a good idea of the spacial organization of the scene.

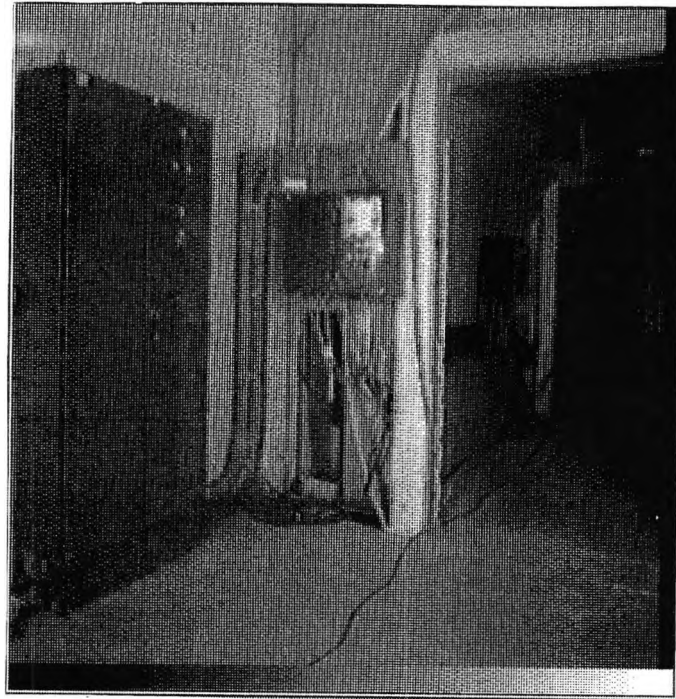


Figure 6.4.1 : Initial image



Figure 6.4.2 : Detected lines.

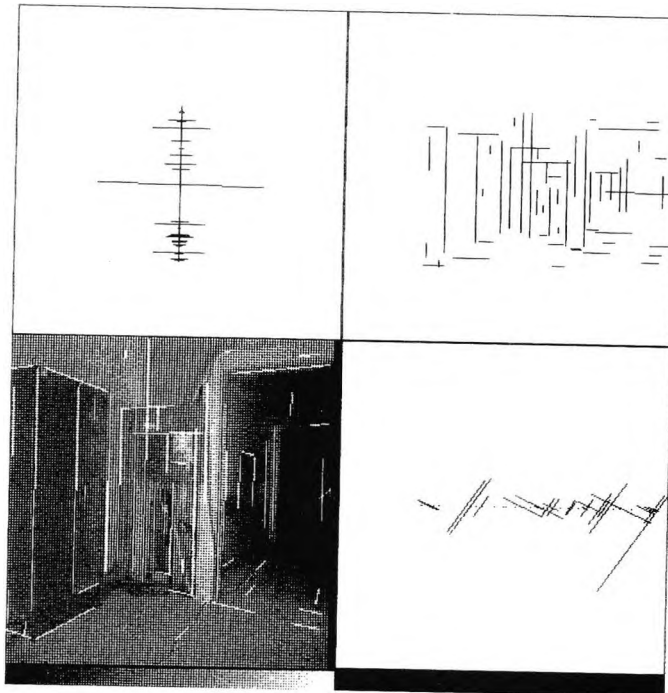


Figure 6.4.3 : linear structures

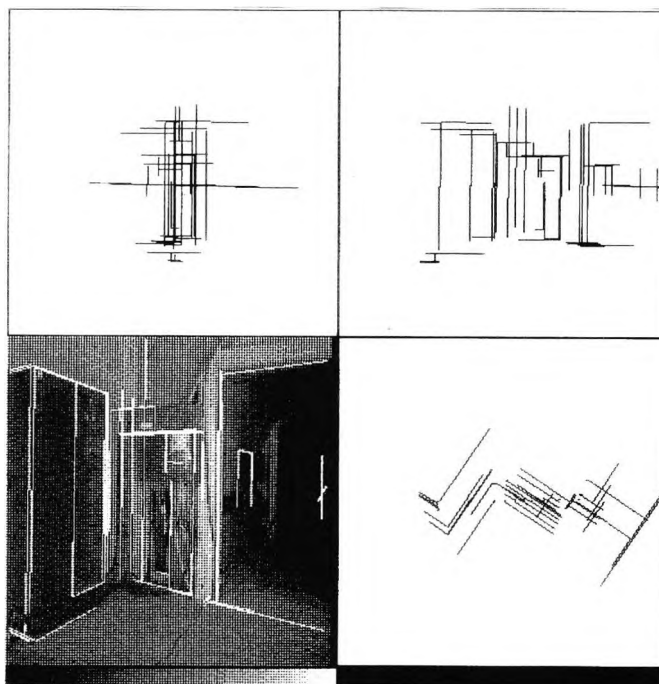


Figure 6.4.4 : Comb structures

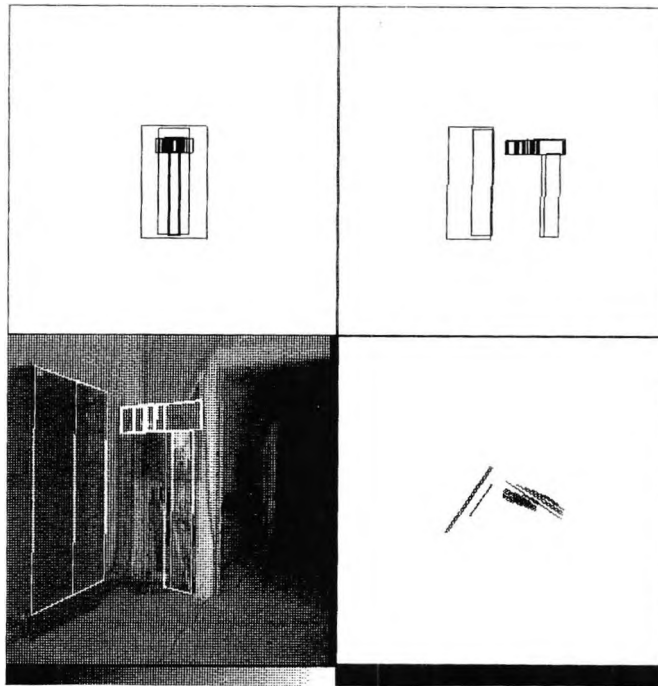


Figure 6.4.5 : Rectangles

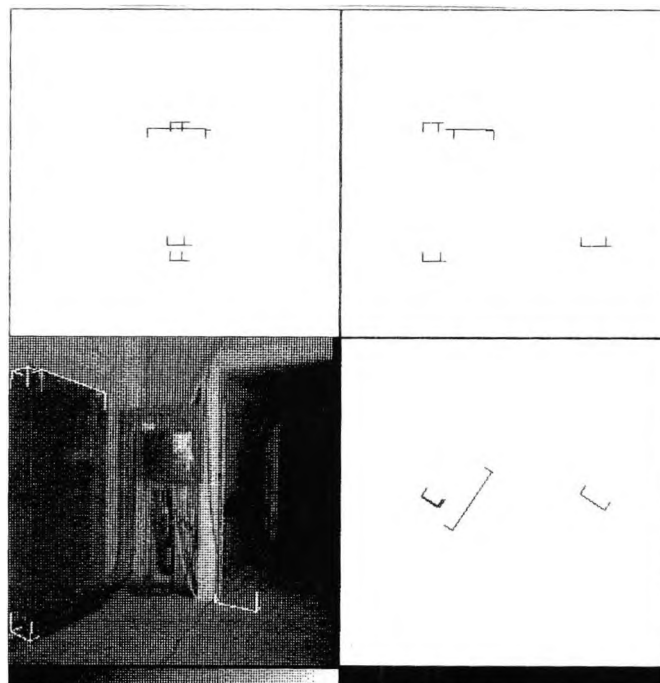


Figure 6.4.6 : Edges



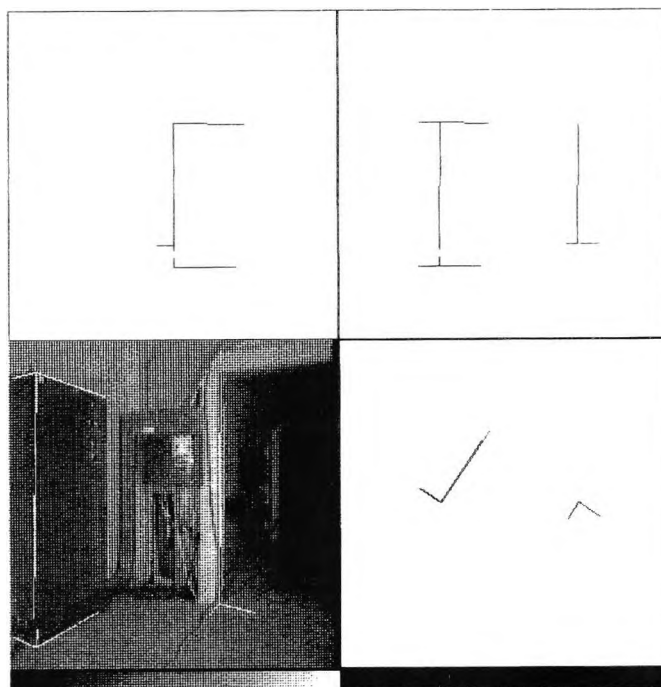


Figure 6.4.7 : Vertices

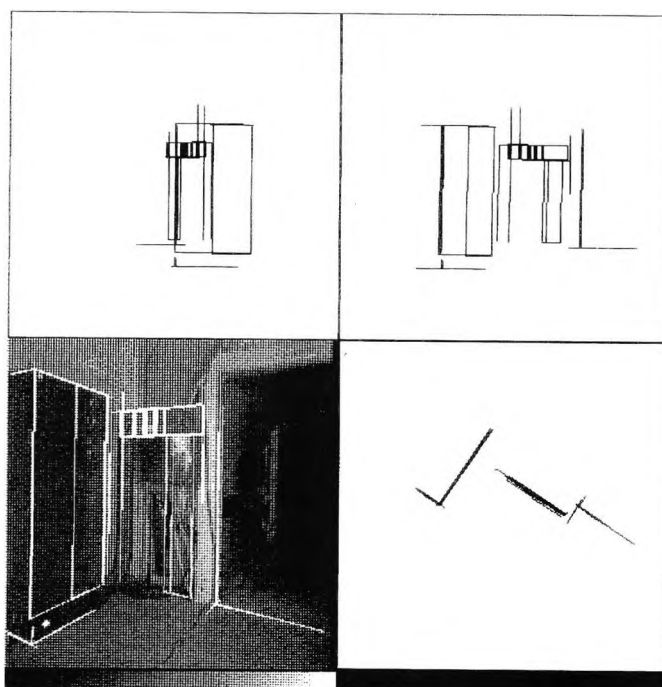


Figure 6.4.8 : Propagation of depth to adjacent structures

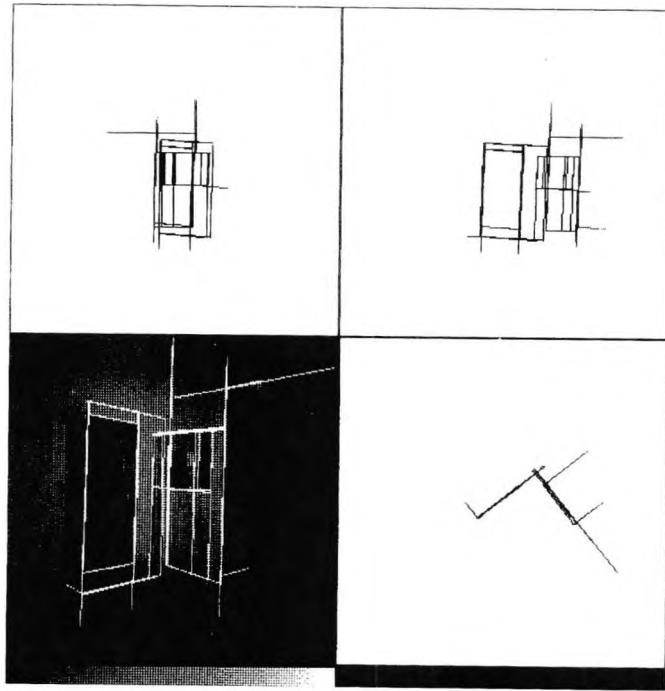


Figure 6.4.9 : 3D interpretation, same scene, different viewpoint

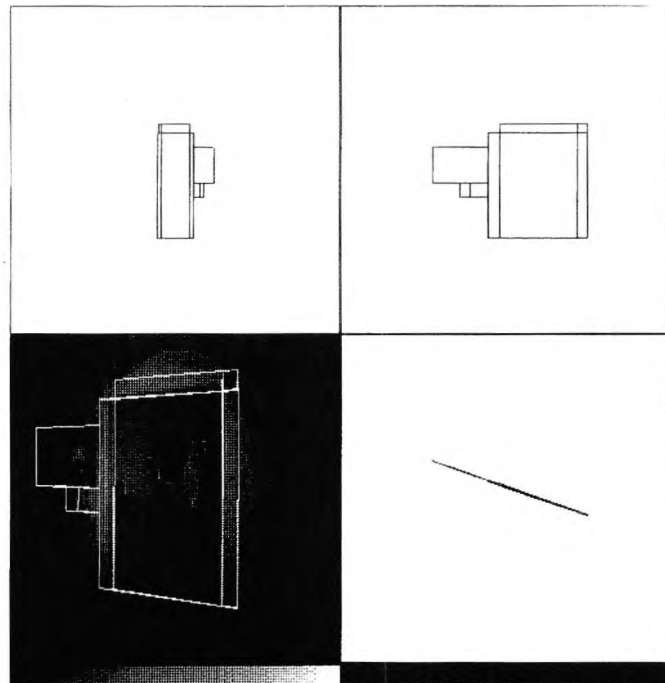


Figure 6.4.10 : 3D interpretation. Same scene, different view-point.

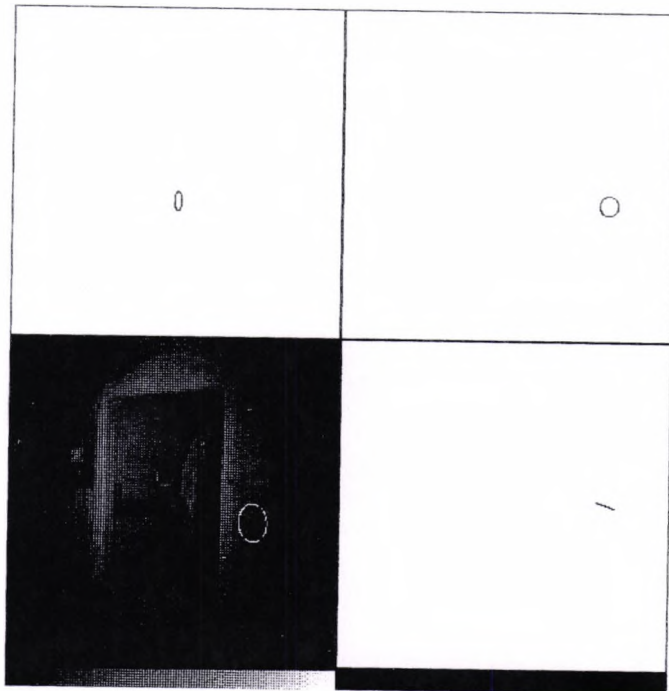


Figure 6.4.11 : Ellipse interpretation

edt12d.str

```
type: Comb structure
orientations: 0.559 0.041 0.828
               0.028 -0.998 0.030
G : -31.76 -44.22 250.00
half plane: +1
d1, d2: 41.6 -42.4
λ : 0.73
α f2: 9648.0
```

edt11d.str

```
type: Comb structure
orientations: 0.773 0.036 0.633
               0.034 -0.999 0.015
G : 60.55 -42.23 250.00
half plane: +1
d1, d2: 40.5 -44.4
λ : 0.76
α f2: 9454.6
```

Figure 6.4.12 : Same model structure under 2 different viewpoints, with the depths set to  $f = 250$ .



Figure 6.4.13 : Initial image of another scene

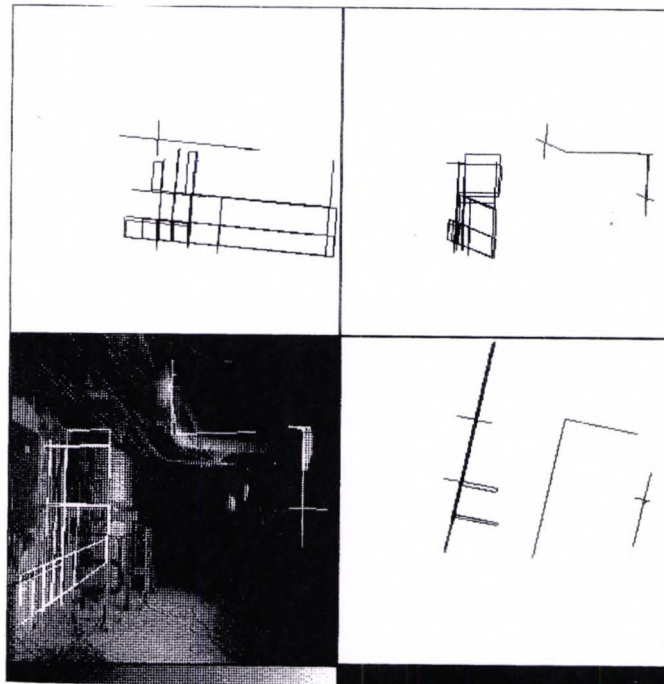


Figure 6.4.14 : Scene interpretation of the scene displayed in figure 6.4.13.

## 6.5 DISCUSSION AND CONCLUSION

A method for constructing 3D structures from straight lines and elliptical arcs extracted from an image has been described in this chapter. Two or three principal perpendicular directions in the scene are supposed known, and the features considered parallel to them in the

## High-level 3D configurations

3D space (it is the subject of chapter 5). The construction of the 3D structures relies on the connectivity of the features in the 3D space. Since connectivity in the image is not sufficient to ensure connectivity in the scene, a likelihood ratio test has been defined in order to optimize the probability of a correct decision. As there is redundancy of the features, linked to the unnecessary details of the structures considered, e.g. the mouldings of a frame, close structures have been grouped in order to simplify the final representation. Albeit the LR test is based on statistical models which are somewhat arbitrary (e.g. the distance between "close" structures is defined by a Gaussian law) it behaves very well, as it simplifies the representation around the longest, and therefore the more reliable, structures.

It has been shown that the representation of the knowledge contained in the image in the form of a hierarchical tree of 3D structures is very powerful for exploiting viewpoint invariant properties and geometrical relationships between the image features in a 3D image interpretation process. It relies heavily on the robustness of the vanishing point detection and the classification of lines.

The 3D representation of the information extracted from the image allows the matching process to be performed between similar representations. The combinatorix is limited by the high level representation of the complex structures. The matching strategy may be based on a prediction-verification paradigm and its feasibility is fundamentally the same as Lowe's method because the information available is the same ; only its representation differs. The representation described in this chapter is a powerful and homogeneous data structure which implicitly contains Brooks' quasi-invariant features, Lowe's perceptual groupings and Quan's bi-ratio property.

## CHAPTER 7

### TOWARDS A CAD DATABASE

#### 7.1 INTRODUCTION

This chapter aims at illustrating the efficiency of the 3D representation described in chapter 6, on a particular problem : 3D map construction. Only the aspects relevant to the work described in the previous chapters are fully studied in the following sections ; the matching process required for the construction is only examined with respect to its feasibility. Then, the constructed map is transferred to the database of ROBCAD, a CAD software for robotic simulation which uses a b-rep representation. Matching with PDMS (the CAD database used by EDF for storing a representation of a nuclear plant) is demonstrated. The construction of a 3D map has been tested on an indoor scene of a power plant (where the matching process has been done by hand).

Several views of the same scene are used to infer the relative depth of the structures, in order to build a symbolic 3D map of the scene, the scale of which remains unknown. The relative positions of the cameras for the different views are not known a priori, but are given by the process with respect to the constructed map. Thus, the same camera in motion may be used for the different views (which may be recorded on a video tape). This makes the acquisition procedure very simple.

The strategy is the following. The coordinate system of the 3D scene is defined such that its origin is located at the origin of the first viewpoint and its axes are parallel to the main directions of the scene. The first view is called the principal view and a second view of the scene is used for the determination of the relative depths of the structures of the principal view. From this process, the location of the second view in the scene coordinate system is known. The second view then becomes the principal view, the structures of which are located by using a third view, and so on. Up to now, there is no possible

contradiction between the different views, during the construction of the map. This is the object of further work. At the end of the process, a 3D map of the scene, in terms of rectangles, corners, vertices and so on, has been built with respect to the scene coordinate system previously defined. The scale of the whole representation remains unknown, but the different elements of the map are consistently scaled, by contrast with the 3D configurations described in chapter 6.

Matching 3D structures from different views is required. It is demonstrated that the 3D representation adopted should make it easier, but no precise matching process is described, as it will be the subject of further work.

## 7.2 3D MAP CONSTRUCTION

### *General principle*

Classically (e.g. (Ayache, 1988; Marapane et al, 1989)), the depth of a structure (e.g. edge, region) is determined by triangulation of the images of this structure, when using different viewpoints. This requires matching the images corresponding to the same structure in the 3D world. Here, the 3D world has already been interpreted with respect to the coordinate systems associated with the viewpoints. The relative depth of the structures is computed in two steps :

- determination of the second viewpoint location,
- determination of the depths of the structures of the principal view matched with structures of the second view,
- propagation of the depth to the 3D configuration extracted from the principal view.

Since the three main directions of the scene are supposed detected by the process described in chapter 5 for each viewpoint (if only two directions have been detected, the third one is automatically deduced for completing the orthogonal triplet), and the camera is supposed approximately up-right, the 3D structures extracted of each view may be represented in a coordinate system parallel to the main directions of

the scene, by simply performing a rotation. The origin of one system with respect to the other remains to be determined. This may be achieved by matching a pair of connected points or a rectangle.

### *Rotation*

First, let us remark that any linear geometrical transformation (e.g. a rotation, a translation or a change of scale) is straightforward when using the 3D representation described in chapter 6.

The rotation matrix is given by the main directions of the scene,  $\vec{U}_1$ , which are the unit vectors of the new coordinate system. This system is oriented as follows :  $(\vec{U}_1, \vec{U}_2, \vec{U}_3)$ , where  $\vec{U}_2$  is the vertical direction oriented towards the bottom and  $\vec{U}_3$  is an horizontal direction oriented towards infinity (opposite to the camera) (figure 7.2.1). The choice of the horizontal direction corresponding to  $\vec{U}_3$  is made with the first view, e.g. the one corresponding to the vanishing point the closest to 0 or such that  $x_v > 0$ , if there is an ambiguity. The correspondence of the horizontal directions between two views follows the same principle. Whether there is a vanishing point close to the principal point of the image and the other one is at infinity then the correspondence is straightforward ; or both vanishing points are at similar distances from the principal point and they appear in the same order on both views. Thus, the 3D structures corresponding to both viewpoints studied are now represented in coordinate systems  $\mathcal{P}_1$  and  $\mathcal{P}_2$  parallel to the main directions of the scene and parallel between themselves (figure 7.2.2)



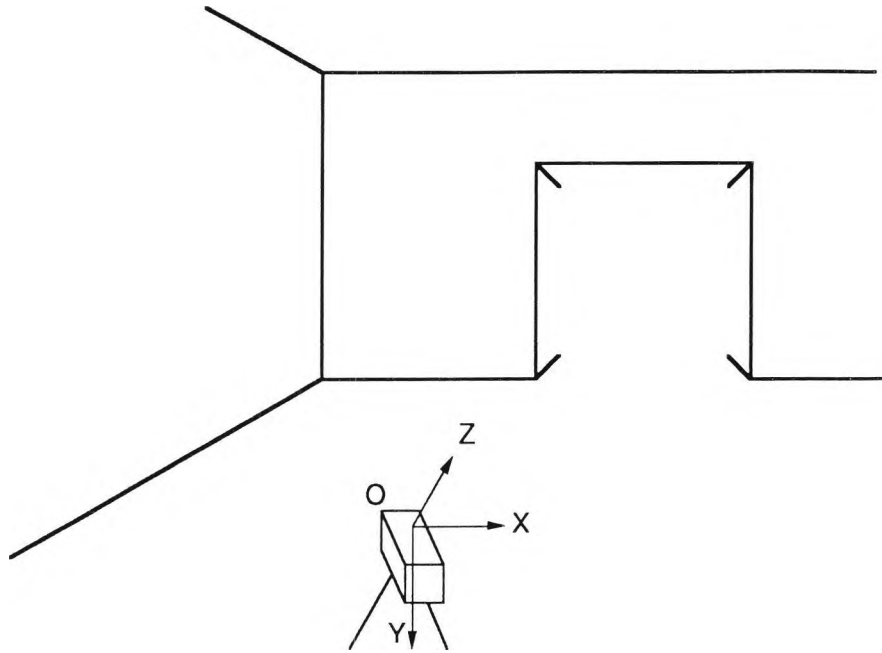


Figure 7.2.1 : Orientation of the axes associated with the scene coordinate system

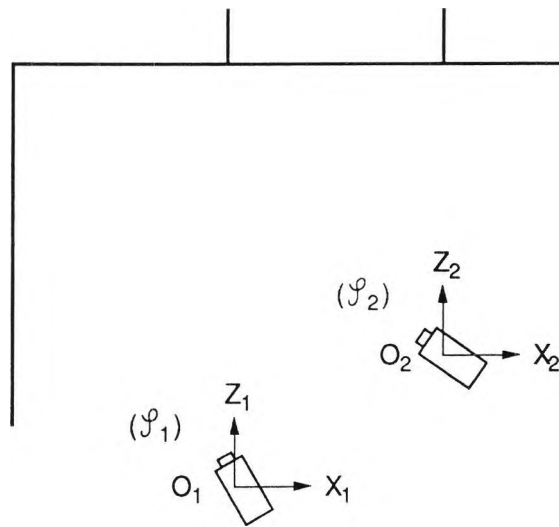


Figure 7.2.2 : New coordinate system associated with 2 viewpoints. (top view).

*Translation*

Let  $O_1$  and  $O_2$  the origins of  $\mathcal{P}_1$  and  $\mathcal{P}_2$ , i.e.  $O_1$  and  $O_2$  are the optic centre of the first and second camera. The origin  $O_2$  of the second viewpoint coordinate system with respect to the coordinate system  $\mathcal{P}_1$  must be determined. The coordinates of  $O_2$  are determined by matching two

pairs of connected points or two rectangles extracted from each view. Two pairs of connected points (typically two rectangular corners connected by a linear structure) or two rectangles may be matched if they have same directions and relative sizes and locations consistent with the set of possible displacements (the displacement is limited as both views must be comparable). As the scale cannot be known without additional information on the displacement, let the depth of this structure be fixed for the principal viewpoint. From the equality of the lengths in the previous matching, the scale and therefore the depth of this structure with respect to the second viewpoint is known. Its representation in  $\mathcal{P}_2$  is then updated. Let  $G_1$  (respectively  $G_2$ ) be the centroid of the structure with respect to  $\mathcal{P}_1$  (respectively  $\mathcal{P}_2$ ), the origin  $O_2$  is given by

$$\overrightarrow{O_1 O_2} = \overrightarrow{G_1 G_2} \quad (7.2.1)$$

For example, the rectangle  $\mathcal{R}_1$ , extracted from the first view, is matched with the rectangle  $\mathcal{R}_2$ , extracted from the second view. This means that their orientations must be identical, with the same ratio  $\lambda$  of the edge lengths (i.e. parameter  $L_1/L_2$ ); moreover the sizes and locations must be consistent (same magnitude). The depth of  $\mathcal{R}_1$  is fixed to  $d_1=d$ . From the parameter  $\alpha = \text{area}/\text{depth}^2$ , the relative scale of  $\mathcal{R}_2$  is found and therefore the depth of  $\mathcal{R}_2$

$$d_2 = \sqrt{\frac{\alpha_1}{\alpha_2}} d$$

The centroid  $G_i$ , for  $i=1$  and  $2$ , is updated

$$G_i \longrightarrow G'_i ; \overrightarrow{O_1 G'_i} = d_i \overrightarrow{O_1 G_i} \quad (7.2.2)$$

Since by hypothesis  $G'_2 \equiv G'_1$ , the origin  $O_2$  is given by

$$\overrightarrow{O_1 O_2} = \overrightarrow{O_2 G'_2} - \overrightarrow{O_1 G'_1} \quad (7.2.3)$$

Thus, the location of the second viewpoint is known with respect to the first one (the previous example shows how simple this determination is when using a 3D representation of the structure). All the structures

extracted from the second viewpoint are now represented in the scene coordinate system  $\mathcal{S}_1$ , centred at  $O_1$ .

*Map construction*

The structures of both views are represented with respect to a unique coordinate system, the scene coordinate system. All types of structures are considered. The 3D structures of the principal view have one degree of freedom, i.e. the depth from  $O_1$ , which is eliminated by matching with the corresponding structure extracted from the second view. This matching is highly constrained ; indeed, the structures should still have the same directions, consistent parameters and locations. Moreover  $\overrightarrow{O_{11}G_1}$ ,  $\overrightarrow{O_{12}G_2}$  and  $\overrightarrow{O_{12}O_1}$  should be coplanar, that is to say the associated determinant should be zero (this is the epipolarity constraint). The intersection point of the lines  $(O_1, \overrightarrow{O_{11}G_1})$  and  $(O_2, \overrightarrow{O_{22}G_2})$  is the centroid of the structure.

$$d_2 \overrightarrow{O_2G_2} - d_1 \overrightarrow{O_1G_1} = \overrightarrow{O_1O_2} \tag{7.2.4}$$

The system is overdetermined (two unknown variables,  $d_1$  and  $d_2$ , and three equations), which is dealt with by using a LMS method, which also allows the uncertainty of the input data to be taken into account.

*Propagation*

The locations of the matched structures with respect to the scene coordinate system are known. The depths of the located structures are now propagated to the 3D configuration to which the structure belongs. This process is only applied to the structures corresponding to the principal view in order to avoid the propagation of errors, as it has been seen that 3D connectivity from a single view is unreliable. At the moment, there is no process for contradiction, but it may be noticed that this process is very important for determining the reliability of the constructed map. Indeed, if the location of a structure is confirmed in the process of several views, its reliability rapidly increases.

### Conclusion

An unscaled 3D map of the scene has been built, using a number of views without reference to the relative positions of the viewpoints. Typically, the acquisition of such a sequence is done by a camera in motion and may be recorded on a video tape for off line processing. It has been seen how elementary the problem becomes when using the 3D representation described in chapter 6. Furthermore, the map constructed is of a high level, since it is in terms of sets of 2D or 3D geometrical forms, hierarchically organized and using a symbolic representation. During the construction of the 3D representation, the forms extracted have been rectified in order to correspond to an *ideal* world, such as represented in a CAD database, where the geometrical relationships, e.g. perpendicularity or parallelism, are supposed to be perfect. As a consequence, the map is not accurate, but is meant to describe the general organisation of the space in a qualitative way.

### 7.3 TOWARDS A CAD DATABASE

The processing for extracting the 3D maps described in the previous sections is referred as TIMI (Three-dimensional Interpretation of Monocular Images). The representation of the scene extracted by TIMI is compared with the representation used by two CAD software packages, ROBCAD (Technomatics) (the CAD software used by EDF for robotic simulation) and PDMS (CAD Centre) (the CAD software used by EDF for storing a representation of the nuclear plants).

The representation extracted by TIMI is surfacic, each facet being described in a symbolic way. The database of ROBCAD allows the representation of surfaces, described in a symbolic way, and the representation of volumes in terms of polyhedrons, the facets of which are represented by their corners. PDMS describes only volumetric elements, by using a symbolic representation.

The level of the representation extracted by TIMI is "intermediate" between the levels of ROBCAD's representation and PDMS's representation.

Therefore, at this stage of TIMI's development, it is possible to convert the maps extracted by TIMI into ROBCAD objects, and it is possible to extract a TIMI representation from a PDMS database. The conversion TIMI→ROBCAD is trivial. The conversion PDMS→TIMI requires the extraction of the plane faces from the volumetric primitives. There are eleven primitives, symbolically represented, so that this extraction is quite simple. Matching between similar primitives (i.e. rectangles, edges, vertices, circular arcs) is now possible (figure 7.3.1). If model primitives remain unmatched, then they may be matched directly with comb structures, by using the relation "is a part of" (e.g. "may an extracted comb structure be a part of a model rectangle?"). The conversion of a PDMS model to a TIMI model should be limited to the "extractable" primitives, e.g. primitives parallel to the three principal directions of the room, if known. As in section 7.2, matching a pair of points or a rectangle enables the determination of the transformation between the model coordinate system and the TIMI map coordinate system. The depth computed by the process described in section 7.2 is used for matching purposes, but is updated by the model in case of success. If one image is used, matching can still be done in the same way as in section 7.2.

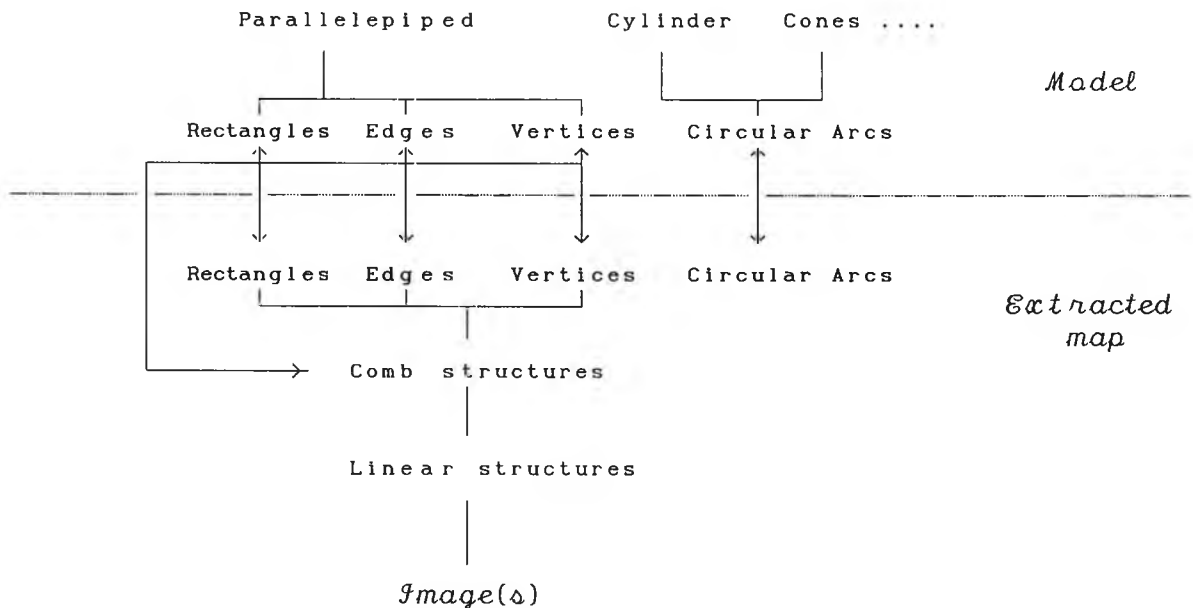


Figure 7.3.1 : PDMS/TIMI matching process

The map constructed is very incomplete. However, this is not a drawback when matching with a CAD database, as it reduces the combinatorix. Actually, the ideal case is to have a particular configuration of a minimum number of primitives, and in that respect circular arcs may be very good cues.

It has been seen that the map constructed may be interfaced with the CAD databases used by ROBCAD and PDMS. The feasibility of a matching process with PDMS' database has been discussed.

## 7.4 RESULTS

Three views of the same scene have been acquired using a camera in motion and have been recorded on a video tape. The processing of these views have been studied in chapter 6 (figures 6.4.1, 6.4.9 and 6.4.10) and the results have been displayed in figures 6.4.8 to 6.4.10. The 3D map extracted is displayed in figure 7.4.1. The matchings of the rectangles corresponding to the cupboard and the pair of points corresponding to the left edge of the door have been performed as follows. The rectangle and the pair of points have been chosen by hand on the first view and possible candidates on the other views have been searched for by filtering on the parameters. For each structure, only one candidate (or no candidate) has been found in the other views. The image at the bottom right represents the perspective projection of the constructed map from a viewpoint parallel to the scene coordinate system. The viewpoints are displayed as the projection of a 3D rectangle centred at the optic centre. The scene is correctly represented : a cupboard on the left wall lying at some distance from the front wall, the door and the beginning of the corridor on this wall. Details have not been properly interpreted, such as the meter lying on the front wall nearby the door. The end of the face wall is aligned with the wall behind the cupboard and the base line of the wall lies in the same plane of the base line of the cupboard, although they belong to unconnected configurations in all views. This demonstrates the fairly good quality of the reconstruction.

The ROBCAD version of the extracted map is displayed in figure 7.4.2.

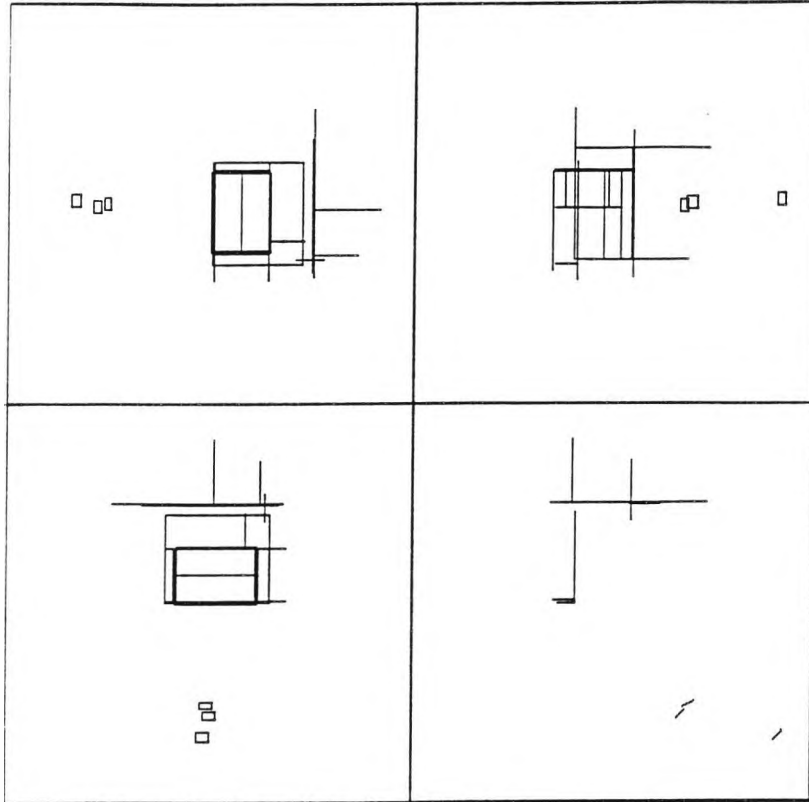


Figure 7.4.1 : 3D map of the scene (a camera is represented in the 3D space by a small square parallel to the image plane with the centre located at the optic centre)

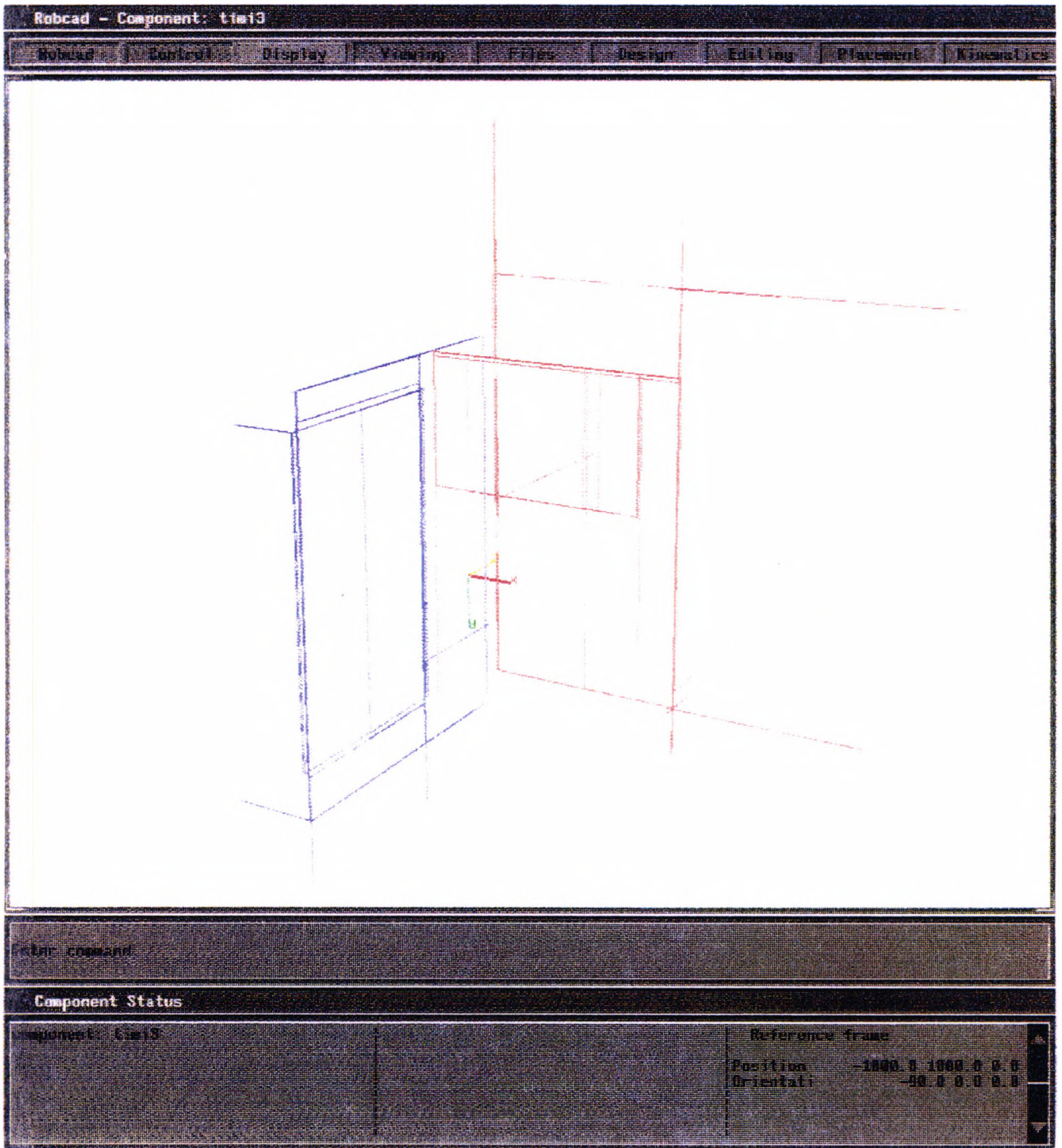


Figure 7.4.2 : same map in ROBCAD



## 7.5 DISCUSSION AND CONCLUSION

It has been seen how to use the 3D configurations described in chapter 6 for the construction of a 3D map. The process appears to be extremely simple, once matching of 3D structures from the various views has been performed. The matching process has not been fully implemented, but it has been seen that it is highly constrained. This is due to the high level of the interpretation of the scene for each view. As a consequence, the combinatorix of the matching is very much simplified.

The epipolarity constraint is exploited by checking the value of a determinant, the bi-ratio property is exploited by checking the equality of the parameter  $L_1/L_2$ , the consistency of the directions of the structures matched is trivial to check, the consistency of locations and sizes is checked by giving upper and lower bounds' on the relative variations of the parameters centroid/depth and size/depth (deduced from the admissible displacements of the camera between two views of the same scene). The simplicity of these comparisons results from the fact that the geometrical properties linked to the 3D consistency of the scene are intrinsic to the representation used. Moreover, 3D representation eases any geometric reasoning, such as merging similar structures, rotation, translation or change of scale.

The construction does not require the prior knowledge of the viewpoint location. Therefore, the calibration process is limited to the acquisition of the intrinsic parameters of the camera, which may be done once only (sometimes they are given by the constructor). However, the scale of the representation cannot be known. This does not seem a serious drawback as the knowledge of the size of one structure allows its determination.

The representation described is intermediate between ROBCAD's representation and PDMS' representation. Therefore, it has been possible to interface it with ROBCAD, and it has been seen how PDMS database could be used for matching.

The accuracy and the degree of detail are clearly limited by the type of features extracted (only in three perpendicular directions and mostly straight line features), by the performance of the matching process and by the corrections made for ending up with an *ideal* schema of the scene. However, it has been seen on an example that the quality of the reconstruction is good when the cubic model is appropriate. Moreover this *ideal* representation is an advantage when matching with a geometric model, as it reduces the combinatorix and simplifies the calculations. Thus, this chapter illustrates the possibilities of a high-level 3D representation of structures from a single view but does not validate an accurate system of 3D map construction.

The 3D interpretation of an indoor scene has been achieved from monocular vision with unknown viewpoints, and its result is a CAD database object.



## CHAPTER 8

## DISCUSSION AND CONCLUSION

## 8.1 SUMMARY OF THE METHOD

A process for 3D interpretation of a scene from a single image has been described. The interpretation has relied upon geometrical properties of the scene, e.g. parallelism and perpendicularity of straight line borders, and the knowledge of the perspective transformation.

The perspective is interpreted, first by searching for the vanishing points, then by classifying the straight line segments and the elliptical arcs extracted from the image with these vanishing points. The straight lines of the scene have been assumed to be parallel to three principal perpendicular directions. Pairs and triplets of vanishing points corresponding to perpendicular directions have been selected in order to reduce the number of false vanishing point candidates. Thus, the 3D orientation of segments and circular arcs of the scene has been determined.

From this stage, the interpretation has been carried on in the 3D space, with respect to the camera coordinate system ; the only reference to the image has been for testing the proximity of the segments. Proximal straight line segments have been merged into a linear structure ; proximal perpendicular linear structures have been grouped to form a comb structure. The comb structures are the basis of the construction of higher level structures, e.g. corners and vertices. It has been chosen for restraining the process to a linear complexity. The result of the interpretation is unscaled 3D local maps, called 3D configurations.

The utilisation of two or more images has allowed the relative scales of these maps to be known, resulting in a consistently scaled 3D map of the scene, which is of a CAD database type.

## 8.2 DISCUSSION

### *Extraction of the primitives from the image*

First, the interpretation of the image is based on the primitive "edge". Three edge detectors have been compared with respect to the uncertainty of the edge detected in various configurations. An improved version (IEF) of the Shen detector (Shen, 1986) has been developed. The edges have been approximated by straight lines and elliptical arcs, which are the data of the interpretation process.

This stage is primordial to the quality of the further process. The smaller the error of localisation, the more selective the various tests involved in the process for selecting the features of interest. The quality of the detection is also very important, as the process relies on accumulation of evidence. The accuracy of the straight line segments could easily be improved by using a LMS method, but the detection of elliptical arcs has been proved more difficult and is the object of further research.

### *Probabilistic approach*

The reliability of the process has been adressed by using a probabilistic approach. Two types of error in this process have been defined : uncertainty of measurement (small errors), e.g. error of the localisation of the edge, and errors of segmentation (gross errors), e.g. choice of the wrong edge due to an accidental connectivity in the image. In order to determine probabilistic laws corresponding to the second type of error, a *prior* statistical model of the image data has been defined. These models are the basis of the method described in the previous chapters :

- A new accumulator space for the detection of vanishing points has been defined by taking into account such models for ensuring a constant reliability of the detection over the accumulator space.
- Segmentation is achieved through a likelihood ratio test based

on these models, which therefore takes into account the two types of errors mentioned. It is found to be equivalent to an adaptative MD test. This method has been used whenever segmentation has been involved in the process, i.e. for classifying the straight line segments and the circular arcs with the vanishing point candidates, for selecting vanishing points corresponding to perpendicular directions and for testing the connectivity of the straight line segments.

The probabilistic approach raises some difficulties. The choice of appropriate decision variables is not trivial and at times the associated probabilistic law is hard to justify other than by its correct behaviour on the boundaries and by its nice mathematical properties (e.g. choice of the exponential law for  $D_1$  in section 6.2.1). If the measurement uncertainty is too high ( $\sigma_0$  too high), the likelihood tests have little significance and the interpretation process no value.

#### *Detection of the vanishing points*

The effort has been concentrated on the consistency of the detection of the vanishing points, whatever their location on the image plane. The method has been proved to be superior to other methods, such as the accumulation of the projections of the lines onto the Gaussian sphere or the accumulation of the intersection points. A first accumulation stage of the lines according to their uncertainty allows the complexity of the accumulation stage to be bounded.

In spite of these efforts to optimize the quality of the detection, the lack of accuracy of the parallelism and the straight line segments results in the presence of numerous false candidates when a good sensitivity of the process is required. The perpendicularity test provides a good filter for these false candidates, however this is not really satisfactory as it restrains the field of application. We believe that higher accuracy for straight line segments is required, which includes a high resolution image and the correction of the lens distortions. Unfortunately, this would result in a larger accumulator

## Discussion and Conclusion

space.

The detection of the vanishing points allows the orientation of the 3D structures to be known. Thus, it appears from this study that information about 3D orientation of the surface (from which the orientation of any line lying on this surface may be deduced) is essential to the interpretation of the image. Indeed, the knowledge of the orientation of the structures enables their 3D representation, which is the basis of this interpretation. The detection of the vanishing points is reliable in a simple environment, e.g. a corridor, but no longer in a complex environment where the lines parallel to principal directions are not in majority. In such environments, this stage could be associated with or replaced by another type of system allowing the determination of surface orientation (range finder), the remaining part of the interpretation being unchanged.

### *3D representation*

The 3D structures constructed have been chosen to avoid combinatorial complexity (comb structures), for completeness of the representation (subjective linear structures) and for their flexibility (a comb structure is part of a higher level structure). Since the construction of high-level structures relies only on the comb structures, the hierarchical structure is very simple. The depth, which remains unknown, parameterises the 3D representation of the structures.

The appeal of this parameterised 3D representation of the structures relies on the fact that the geometrical properties are implicit and therefore are fully exploited by further processing. It has been seen how simple any geometrical transformation is, and how constrained a matching process becomes, when using such a representation.

### *3D map construction*

A coarse but high level 3D map has been constructed from a sequence of images. This map uses an *ideal world* type representation, which eases matching or interfacing with a CAD database. As an example, the map

has been converted into a ROBCAD object.

Relative depth of the structures is deduced from the interpretation of a sequence of images. Depth may also be deduced from matching with a CAD database. In contrast with orientation, depth has not appeared to be a significant parameter for the interpretation, which has largely been carried out without it. This is an interesting consideration, since it is difficult to estimate depth with good accuracy (e.g. using active or passive stereovision).

The fact that the 3D representation extracted is coarse, eases matching with representations from different viewpoints or/and with a CAD database. Once the matching is satisfactory, a top down process should allow the representation to be completed. Thus, the approach proposed for interpreting a scene from monocular vision consists of identifying the principal elements of the scene, e.g. the walls, relative to which the other elements will be located. The feasibility of the first stage has been demonstrated throughout this thesis.

Such a method for 3D interpretation does not require any extrinsic calibration of the camera (e.g. relative to another camera) in contrast with stereo-vision or active ranging, or a precise geometric model in contrast with model driven strategy. It only uses general knowledge, such as the occurrence of geometrical relationships, e.g. parallelism. As a result, both *visible* parts of the process, that is to say the input and the output, are extremely convenient ; the former is a simple sequence of images and the latter is a CAD type symbolic representation of the scene.

#### *Further improvement*

The approach described seems very appealing, however further work is needed to generalize it and to increase the quality of the results.

- The accuracy of the straight line segments and the detection of elliptical arcs should be improved.
- Too many erroneous peaks are detected in the accumulator space.



## Discussion and Conclusion

To reduce their number, the significance of the test should still be increased.

- The likelihood ratio tests have been deduced by hand. They could be generated by using mathematical tools.
- The interpretation should be extended to other type of primitives (e.g. regions).
- Complete scoring of the 3D structures and development of a matching strategy remains to be done.
- The interpretation could be completed by using a top down process in a second stage.

### 8.3 FROM MONOCULAR VISION TO A CAD REPRESENTATION

A 3D representation of the scene has been extracted from monocular image(s) with unknown viewpoints. It has been based on the detection of the vanishing points. Then, connectivity in the image has enabled the construction of 3D structures, such as rectangular structures or vertices, the parameters of which are completely determined, except the depth. It has been proved that this 3D representation of structures extracted from the image is very powerful. This point has been illustrated by the construction of a 3D map from a sequence of monocular images with unknown viewpoints. Then, the map has been converted into an object of a ROBCAD database and the feasibility of a matching process with a PDMS database has been discussed. Therefore, from a (set of) monocular image(s), a very convenient 3D representation of the scene has been extracted. The explicit definitions of models for the measurement error and the segmentation error have enabled an optimal and consistent definition for the parameters involved in the process ; this results in an adaptative method for a wide range of scenes.

## REFERENCES

AYACHE, N. (1988). Construction et Fusion de Representation Visuelle 3D Application à la robotique mobile, These de doctorat d'Etat, Université Paris-Sud, Orsay, France.

AYACHE, N. (1983). Un système de vision bidimensionnelle en robotique industrielle. Thèse de docteur ingénieur, Université Paris-Sud, Orsay, France.

AYACHE, N. and FAUGERAS, O.D. (1987). Maintening representations of the environment of a mobile robot. *International Symposium on Robotics Research, Santa-Cruz, California. Proceedings.*

BALLARD, D. H. and C.M. BROWN. (1982). Computer Vision. (New Jersey : Prentice-Hall, Englewood Cliffs).

BARNARD, S.T. (1983). Interpreting perspective images, *Artificial Intelligence 21* (3). 435-462.

BERTHOD, M. (1987). Un nouvel algorithm d'approximation polygonale. Unpublished manuscript.

BOLLES, R.C. and HORAUD, P. (1986). 3DPO : A Three-Dimensional Part Orientation System. *The International Journal of Robotics Research 5* (3). 3-26.

BOULT, T.E. (1985). A survey of some three-dimensional viszion systems. *SIGART Newsletter* (92).

BRILLAULT, B. (1989). Parallel and Perpendicular Line Grouping in a 3D Scene from a Single View. *5<sup>th</sup> Alvey Vision Conference, Reading UK, 1989, Proceedings.* 257-262.

BRILLAULT, B. (1991). New Method for Vanishing Point Detection. *CGVIP : Image Understanding, July (1991).*

## References

- BROOKS, R.A. (1984). Model-Based Computer Vision. (UMI research Press, Michigan, 1984).
- BROWN, C. (1989). Predictive Gaze Control. *Fifth Alvey Vision Conference, Reading, UK. Proceedings.* 103-108
- CANNY, J. (1986). A computational approach to edge detection. *IEEE transactions on PAMI* 8 (6). 679-698.
- CAD CENTRE. PDMS. CAD software distributed by CAD Centre, London, England.
- CASTAN, S., ZHAO, J. and SHEN, J. (1990). Optimal Filter for Edge Detection Methods and Results. *First European Conference, Antibes, France, 1990. Proceedings.* 13-17.
- CLOWES, M.B. (1971). On Seeing Things". *Artificial Intelligence*, 2 (1). 79-116.
- COELHO, C., STRAFORINI, M., CAMPANI, M. (1990). Using geometrical rules and a priori knowledge for the understanding of indoor scenes. *1<sup>st</sup> British Machine Vision Conference, Oxford UK, 1990, Proceedings.* 229-234.
- CROWLEY, J.L., BOBET, P. and SARACHIK, K. (1990). Dynamic world modeling using vertical line stereo. *In: First European Conference on Computer Vision, Antibes, France, 1990. Proceedings.* 241-248.
- CROWLEY, J.L., and STELMASZYK, P. (1990). Measurement and intergration of 3-D structures by tracking edge lines. *First European Conference on Computer Vision, Antibes, France, 1990. Proceedings.* 269-280.
- DERICHE, R. (1987) Using Canny's criteria to derive an optimal edge detector recursively implemented. *The International Journal of Computer Vision* 1(2). 167-187.
- DERICHE, R. and FAUGERAS, O.D. (1990). Tracking Line Segments. *First*

- European Conference on Computer Vision, Antibes, France, 1990.* 259-268.
- DICKSON, W. (1989). Personal communication.
- DICKSON, W. (1990). Feature Grouping in a Hierarchical Probabilistic Network. *1<sup>st</sup> British Machine Vision Conference, Oxford, UK. Proceedings.* 13-18.
- DU LI, SULLIVAN, G.D. and BAKER, K.D. (1989). Edge Detection at Junctions. *Fifth Alvey Vision Conference, Reading, UK. Proceedings.* 121-126.
- ELLIS, T.J., MOUKAS, P. and WEST, G.A.W. (1987). Complete Object Inspection using CAD Models and Robotic Manipulation. *Alvey Vision Conference, Cambridge, UK. Proceedings.*
- ELLIS, T.J., ABOOD, A., BRILLAULT, B. (1991). Ellipse Detection and Matching with Uncertainty. *British Machine Vision Conference 1991. Proceedings.* 136-144.
- EVEN, P. and MARCE, L. (1988). Manned geometric modelling for computer aided teleoperation. *International Symposium of Teleoperation and Control. Proceedings.* 113-122.
- FAUGERAS, O.D. and HERBERT, M. (1986). The Representation, Recognition, and Locating of 3-D Objects. *The International Journal of Robotics Research* 5 (3). 27-52.
- FAUGERAS, O.D. (1988). *Vision Artificielle pour la Robotique.* (Cours d'été EDF-CEA-INRIA).
- FISHER, R.B. (1987). Model invocation for three dimensional scene understanding. *IJCAI, Milan, ITALY.*
- GRANGER, C. (1985). *Reconnaissance d'objets par mise en correspondance en vision par ordinateur.* Thèse d'Université, Nice, France.
- GRIMSON, W.E.L. (1987). Recognition of object families using

## References

parameterized models. *IEEE, Object Recognition*. 93-117.

GRIMSON, W.E.L. and LOZANO-PEREZ, T. (1984). Model-based recognition and Localization from Sparse range and tactile data. *The International Journal of Robotics Reseach*, 3 (3). 3-35.

HACKETT, J.K. and SHAH, M. (1990). Multi-Sensor Fusion : A Perspective. *IEEE*. 1324-1330.

HANSON, A.R. and RISEMAN, E.M. (1978). VISIONS : a computer system for interpreting scenes. *Computer Vision System*. 303-333.

HANSON, A.R. and RISEMAN, E.M. (1988). Advances in computer vision. (Prentice Hall, editor C. Brown, 1988).

HORAUD, R. and VEILLON, F. (1990). Finding geometric and relational structures in an image. *First European Conference in Computer Vision, Antibes France, 1990. Proceedings*. 374-384.

HOUGH, P.V.C. (1962). Method and Means for Recognizing Complex Patterns. (U.S. patent 30 69 654).

HUFFMAN, D.A. (1971). Impossible Objects as Nonsense Sentences. In *Machine Intelligence 6*, (Edinburgh University Press, editors B. Meltzer and D. Michie).

KALMAN, R.E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, march 1960. 35-45.

KANADE, T. (1981). Recovery of the Three-dimensional Shape of an Object from a Single View. *Artificial Intelligence (17-1)*. 409.

KANATANI, K. (1989). Reconstruction of Consistent Shape from Inconsistent Data : Optimization of 2 $\frac{1}{2}$ D Sketches. *International Journal of Computer Vision (3)*. 261-292.

LOWE, D. G. (1985). Perceptual Organization and Visual Recognition. (Kluwer Academic Publishers).

- LUX, A. (1985). Algorithmique et contrôle par ordinateur. Thèse de doctorat d'Etat, Université de Grenoble, France.
- MACKWORTH, A.K. (1973). Interpreting Pictures of Polyhedral Scenes. *Artificial Intelligence* 4. 121-137.
- MAGEE, M.J. and AGGARWAL, J.K. (1984). Determining Vanishing Points from Perspective Images. *Computer Vision, Graphics, and Image Processing* (26). 256-267.
- MARAPANE, S.B. and TRIVEDI, M.M. (1989). Region-Based Stereo Analysis for Robotic Applications. *IEEE Transactions on, Systems, Man, and Cybernetics*, 29 (6). 1447-1463.
- MARR, D. (1982). Vision. (W.H. Freeman and Co).
- MINSKY, M. (1975). A Framework for Representing Knowledge. in *Psychology of Computer Vision*, (Mac-Graw Hill, editor P. Winston). 211-277.
- de MICHELI, E., Caprile B., Ottonello P. and Torre V. (1989). Localisation and Noise in Edge Detection. *IEEE Transaction on PAMI* (11-10), 1989. 1-11.
- MOHAN, R. and NEVATIA, R. (1989). Using perceptual Organisation to extract 3D Structures. *IEEE PAMI* (11-11).
- MULGAONONKAR, P.G. and SHAPIRO, L.G. (1985). Hypothesis-Based Geometric Reasoning about Perspective Images. *IEEE, 3<sup>rd</sup> workshop of computer vision*. 11-18.
- NAEVE, A. and EKLUNDH, J.O. (1987). On Projective Geometry and the Recovery of 3D Structure. *First International Conference on Computer Vision, Londres, 1987. Proceedings*. 128-135.
- NELSON, R.N. and YOUNG, T.Y. (1986). Determining Three-dimensional Object Shape and Orientation from a Single Perspective View. *Optical Engineering* (25-3).

## References

NEVATIA, R. and BINFORD, T. (1977). Description and Recognition of curved Objects. *Artificial Intelligence* 8 (1).

NEVATIA, R. (1982). Machine Perception. (Prentice-Hall, Inc., Englewood Cliffs, New-Jersey).

NOESIS. (1988). Image processing software Visilog ( 27 rue Hoche, 78000 Versailles, France).

PDMS. CAD software distributed by CAD Centre, London, UK.

PEARL, J. (1988). Probabilistic reasoning in intelligent system. (Morgan Kaufmann Publishers).

POLLARD, S.B., PORILL, J. and MAYHEW, J.E.W. (1989). Predictive Feed Forward Stereo Processing. *Fifth Alvey Vision Conference, Reading, UK. Proceedings.* 97-102.

PORILL, J. (1989). Fitting Ellipses and Predicting Confidence Envelopes using a Bias Corrected Kalman Filter. *Fifth Alvey Vision Conference, Reading, UK. Proceedings.* 175-180.

PRESS, W. H. (1988). Numerical C recipes. (Cambridge University Press, 1988).

QUAN and MOHR, R. (1988). Matching perspective images using geometric constraints and perceptual grouping. *ICCV, Florida. Proceedings.* 679-683.

QUAN and MOHR, R. (1989). Determining perspective structures using hierarchical Hough transform, *Pattern Recognition Letters* (9). 279-286.

ROBERTS, L. (1965). Machine perception of 3D solids. (Optical and Electro-optical Information Processing, editors J.T. Tippett et al, Eds MIT press, Cambridge)

ROSENFELD, a. and KAK, A.C. (1982). Digital Picture Processing. Volume 2. (Academic Press, Orlando).

- ROSIN, P. and WEST, G. (1988). Circular Arc Detection in Images. *4th. Alvey Vision Conference, Manchester, England, Sept 1988. Proceedings.*
- ROSIN, P.L. and WEST, G.A.W. (1990). Segmenting curves into elliptic arcs and straight lines. *ICCV 3, OSAKA, JAPAN, 1990. Proceedings.*
- SERRA, J. (1982). *Image Analysis and Mathematical Morphology.* (London: Academic Press).
- SHAFER, G. (1976). *A Mathematical Theory of Evidence.* (Princeton University Press).
- SHEN, J. and CASTAN, S. (1986). An optimal Linear Operator for Edge Detection, . *CVPR, Miami, 1986. Proceedings.*
- SILVEY, S.D. (1979). *Statistical Inference.* (Chapman and Hall, Monographs on Statistics and Applied Probability).
- SIMONI, P.O. (1988). An overview of ANDES : A Knowledge-based Scene Analysis System. *IEEE.*
- STRAFORINI, M., COELHO, C., CAMPANI, M. (1990). A fast and precise method to extract vanishing points. *SPIE Close-Range Photogrammetry Meets Machine (1395): 266-274.*
- SUGIHARA, K. (1988). Some Location Problems for Robot Navigation Using a Single Camera. *Computer Vision Graphics and Image Processings (42).* 112-119.
- TECHNOMATICS. ROBCAD. CAD software for robotics simulation distributed by Technomatics, ISRAEL.
- THOMPSON, D.W. and MUNDY, J.L. (1987). Three dimensional Model Matching From an Unconstrained Viewpoint. *IEEE International Conference on Robotics and Automation, (1987).* 208-220.
- TSUJI, S. and NAKANO, H. (1981). Knowledge-based identification of artery branches in cine-angiograms. *IJCAI 7.* 710-715.



## References

WEI, G.Q. and HE, Z.Y. (1988). Determining Vanishing Point and Camera Parameters : New Approaches. 9<sup>th</sup> *International Conference on Pattern Recognition, Italy, 1988*.

WEISS, I. (1988). Straight Line Fitting in a Noisy Image. *IEEE : Conference on Computer Vision and Pattern Recognition* : 647-652.

WHITTEN, G. (1988). Vertex space and its Application to Model Based Object Recognition. *IEEE*. 847-857.

WORRALL, A.D., BAKER, K.D. and SULLIVAN, G.D. (1989). Model-Based Perspective Inversion. *Image and Vision Computing (7-1)*. 17-23.

## APPENDIX 1

## PERSPECTIVE TRANSFORMATION

Let  $(C_0, X, Y, Z)$  be the world coordinates, where  $C_0$  is the optic centre,  $C_0Z$  the optic axis and  $C_0X$ ,  $C_0Y$  are parallel to the image axes  $Ox$  and  $Oy$ . First, the projection  $C$  of  $C_0$  onto the image, i.e. the principal point  $C$ , is chosen to be the origin  $O$ . In the  $(C_0, X, Y, Z)$  coordinate system a straight line in the 3D world may be expressed in the form

$$(L) \quad \begin{cases} X = X_0 + \mu U_x \\ Y = Y_0 + \mu U_y \\ Z = Z_0 + \mu U_z \end{cases} \quad (A1.1)$$

where  $P_0 = (X_0, Y_0, 0)$  is the intersection between the optic plane and the straight line (L) and  $\vec{U}$  is the unit vector of (L).  $\mu$  represents the distance from the point  $P = (X, Y, Z)$  to the point  $P_0$  (see figure A1.1).

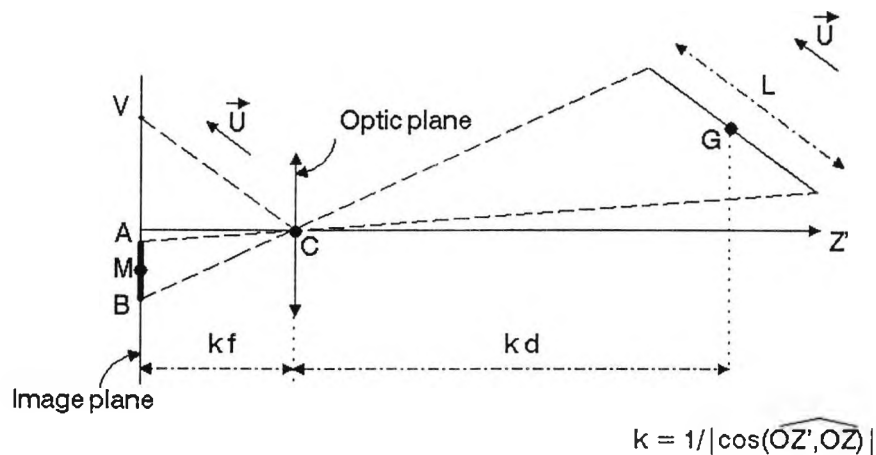


Figure A1.1 : Perspective transformation

L is assumed not to be parallel to the image plane ( $U_z \neq 0$ ). The projection (D) of the line (L) onto the image is represented by

$$(D) \quad \begin{cases} x = f \frac{U_x}{U_z} + \frac{1}{\mu} \left( f \frac{X_0}{U_z} \right) = x_v + \tau u_x \\ y = f \frac{U_y}{U_z} + \frac{1}{\mu} \left( f \frac{Y_0}{U_z} \right) = y_v + \tau u_y \end{cases} \quad \text{with } \tau > 0 \quad (A1.2)$$

where  $f$  is the distance between the optic plane and the image plane. The point  $V$  with the coordinates  $(x_v, y_v)$  is the vanishing point of the line (D) and depends only on the direction of the line (L).  $\tau$  represents the distance from the point  $Q(x, y)$ , image of  $P$ , to the vanishing point  $V$ .

*Parallelism and perpendicularity*

As the vanishing point of a 2D line only depends of the direction of the corresponding 3D line, if two lines are parallel in the 3D world, they have a common vanishing point in the image. The coordinates of the vanishing point allows the common direction in the 3D world to be known.

Two lines (L) and (L') are perpendicular in the 3D world if

$$U_x U'_x + U_y U'_y + U_z U'_z = 0. \quad (A1.3)$$

If  $U_z$  and  $U'_z$  are different from 0, then the corresponding 2D lines (D) and (D') are related in the following way (from A1.2 and A1.3)

$$x_v x'_v + y_v y'_v = \vec{OV} \cdot \vec{OV}' = -f^2. \quad (A1.4)$$

If  $U_z = 0$  then  $\vec{u} \cdot \vec{OV}' = 0$ , and if  $U_z = U'_z = 0$  then  $\vec{u} \cdot \vec{u}' = 0$ .

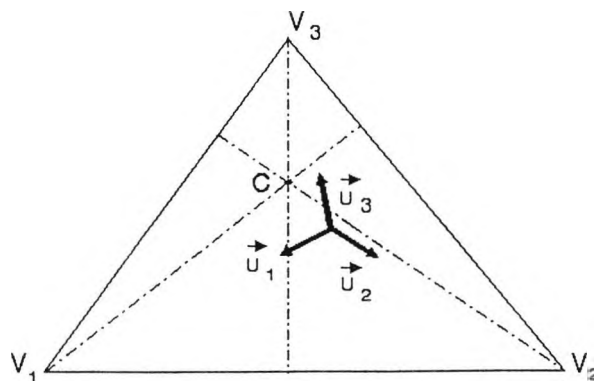


Figure A1.2 : Triangle formed by the vanishing points associated with three orthogonal directions in the scene.

*Calibration parameters*

In order to relate the vanishing point coordinates with the 3D direction of the line, the projection C of the optic centre (the origin of the image so far) and the distance between the optic plane and the image plane should be known. The origin of the image is now any arbitrary point O, and C is looked for.

Let  $T = (M_1, M_2, M_3)$  be a triangle, and  $\Omega$  its orthocentre (the heights dropped from the vertices meet at the same point, called the orthocentre of the triangle). If  $\{S_i\}$  is the set of the feet of the heights of the triangle, then :

$$\vec{\Omega S}_i \cdot \vec{\Omega S}_j = \text{constant } K. \quad (\text{A1.5})$$

$K < 0$  means that the orthocentre is inside the triangle, while it is outside if  $K > 0$ .

If T is a triangle formed by three vanishing points corresponding to perpendicular directions, then the orthocentre  $\Omega$  of the triangle is equal to C, and the constant K is equal to  $-f^2$ .

The coordinates of C are given by solving the following system

$$\begin{cases} \vec{CV}_1 \cdot \vec{V}_2 \vec{V}_3 = 0 \\ \vec{CV}_2 \cdot \vec{V}_3 \vec{V}_1 = 0 \end{cases} \quad (\text{A1.6})$$

If the cartesian coordinates of  $V_i$  are  $(x_i, y_i)$ , the cartesian coordinates  $(x_c, y_c)$  of C are given by

$$\begin{cases} x_c = \frac{\sum_i x_i y_i (x_k - x_h) - \prod_{j>i} (y_i - y_j)}{\sum_{i \neq j} \epsilon_{ij} x_i y_j} \\ y_c = \frac{\sum_i x_i y_i (y_k - y_h) - \prod_{j>i} (x_i - x_j)}{\sum_{i \neq j} \epsilon_{ij} x_i y_j} \end{cases}, \quad (\text{A1.7})$$

with  $h \equiv i+1 \pmod{3}$  and  $k \equiv i+2 \pmod{3}$ ;  $\epsilon_{ij} = 1$  if  $(j-i) \equiv 1 \pmod{3}$

and -1 otherwise . If one of the point is at infinity, then the system is undetermined.

Then the value of f is deduced

$$f = \sqrt{(x_1 - x_c)(x_c - x_2) + (y_1 - y_c)(y_c - y_2)} \quad (A1.8)$$

Let us remark that the vanishing point coordinates should first be consistently scaled by dividing the ordinate in the image by  $\rho$ , where  $\rho$  is the scale ratio  $y/x$ .

### *Filtering criterion*

Let the straight line segment  $[a,b]$  be the projection of the straight line segment  $[A,B]$  onto the image, then

$$\frac{\tau_a}{\tau_b} = \frac{\mu_A}{\mu_B} = \frac{Z_A}{Z_B} \quad (A1.9)$$

Let  $a$  be the further end point from the vanishing point. As the focussing distance and the depth of field should be approximately known, it is possible to bound the ratio  $Z_a/Z_b$  and therefore  $\tau_a/\tau_b$

$$1 < \frac{\tau_a}{\tau_b} \leq k \quad (A1.10)$$

where  $k = Z_{\max}/Z_{\min} > 1$

This constraint may be rewritten in the following form

$$\frac{d}{\ell} \geq \frac{D}{F} \quad (A1.11)$$

where  $d$  is the distance from the centroid of  $[a,b]$  to the vanishing point,  $\ell$  the length of  $[a,b]$ ,  $D$  the focussing distance and  $F$  the depth of field.

A point  $V$  cannot be the vanishing point of the segment  $[a,b]$  if the inequality A1.11 is not fulfilled. An important number of impossible vanishing points are filtered in this way.

*Bi-ratio property*

The bi-ratio  $(P_1, P_2, P_3, P_4)$  of four aligned points is invariant by perspective transformation (see eq. A1.1 and A1.2). If  $(p_1, p_2, p_3, p_4)$  are the projections of these points onto the image, then :

$$\begin{aligned} (P_1, P_2, P_3, P_4) &= \frac{p_1 p_3}{p_1 p_4} \times \frac{p_2 p_4}{p_2 p_3} = \frac{\tau_3 - \tau_1}{\tau_4 - \tau_1} \times \frac{\tau_4 - \tau_2}{\tau_3 - \tau_2} = \frac{\mu_3 - \mu_1}{\mu_4 - \mu_1} \times \frac{\mu_4 - \mu_2}{\mu_3 - \mu_2} \\ &= \frac{P_1 P_3}{P_1 P_4} \times \frac{P_2 P_4}{P_2 P_3} \end{aligned} \quad (\text{A1.12})$$

This property may be used as a constraint in a matching process (Quan, 1988).



## APPENDIX 2

## THE KALMAN FILTER

Let  $a$  be a vector to optimize, and  $x_i$  the vector measured. In absence of noise, the vectors  $a$  and  $x_i$  are supposed related by the equation

$$f_i(x_i, a) = 0 \quad ; \quad f_i \in \mathbb{R}^{p_i}, \quad x_i \in \mathbb{R}^{m_i}, \quad a \in \mathbb{R}^n \quad (\text{A2.1})$$

Because of noise the vector measured is in fact

$$x'_i = x_i + e_i$$

where  $e_i$  is noise with zero mean and covariance matrix  $C_i$ . The vectors  $e_i$  are assumed independent.

If  $f_i$  is not a linear relation of  $x_i$  and  $a$ ,  $f_i$  is linearized around  $(x'_i, a_{i-1})$  by Taylor's expansion

$$f_i(x_i, a) \approx f_i(x'_i, a_{i-1}) + \frac{\partial f_i}{\partial x_i} (x_i - x'_i) + \frac{\partial f_i}{\partial a} (a - a_{i-1}) \quad (\text{A2.2})$$

It is possible to rewrite the measure equation (A2.1), using (A2.2) in the form of

$$y_i = M_i a + w_i \quad (\text{A2.3})$$

where  $y_i$ ,  $w_i$  are vectors with  $p_i$  lines and  $M_i$  is a matrix ( $p_i \times n$ ), such that :

$$\left\{ \begin{array}{l} y_i = -f_i(x'_i, a_{i-1}) + \frac{\partial f_i}{\partial a} a_{i-1} \\ M_i = \frac{\partial f_i}{\partial a} \\ w_i = \frac{\partial f_i}{\partial x_i} (x_i - x'_i) \end{array} \right. /$$



Appendix 2

The noise  $w_i$  has zero mean and covariance  $W_i$

$$W_i = \frac{\partial \ell_i}{\partial x_i} C_i \frac{\partial \ell_i}{\partial x_i}^t$$

The estimate  $a_i$  of  $a$  after  $i$  iterations satisfies

$$a_i = a + \delta_i$$

where  $\delta_i$  is noise with zero mean and covariance  $S_i$ .

The Kalman filter minimizes the following least square criterion

$$C = \sum (y_i - M_i a)^t W_i^{-1} (y_i - M_i a) + (a_0 - a)^t S_0^{-1} (a_0 - a)$$

The minimization of  $C$  has a solution

$$a_k = (S_0^{-1} + \sum_1^k M_i^t W_i^{-1} M_i)^{-1} (S_0^{-1} a_0 + \sum_1^k M_i^t W_i^{-1} y_i) \quad (A2.4)$$

The size of the matrix to invert is  $n \times n$ . Moreover, incorporating one extra measure requires the complete recomputation of  $a$ . The Kalman filter provides a recursive solution to this problem of minimization.

The recursive equations of the Kalman filter gives the estimate of  $a$  knowing the set of data  $y_i$

$$\begin{cases} a_i = a_{i-1} + K_i (y_i - M_i a_{i-1}) \\ K_i = S_{i-1} M_i^t (W_i + M_i S_{i-1} M_i^t)^{-1} \\ S_i = (I - K_i M_i) S_{i-1} \end{cases} \quad (A2.5)$$

$K_i$  is the Kalman gain. The matrix to invert is now  $p_i \times p_i$  (typically  $p_i$  is equal to 1 or 2).

If the noise vectors  $e_i$  and  $\delta_i$  are Gaussian, the estimate of  $a$  provided by the Kalman filter is the expected value of  $a$  knowing the set

of data  $\psi_1$ .



## APPENDIX 3

## LIKELIHOOD RATIO

A traditional approach in image processing when a decision has to be taken concerning an event  $H$  is : is the observation  $V$  likely when the event  $H$  occurs? For instance, the classical Mahalanobis distance test is based on this approach. Often, such an approach does not take into account all the information available on  $H$ , e.g. the prior probability of  $H$  to occur.

The aim of the likelihood ratio test is to compare the probability of two mutually exclusive events, called  $H$  and  $\bar{H}$ . In other words, the problem is changed to : according to the observation  $V$ , is the event  $H$  more likely than the event  $\bar{H}$ ? In the work presented here, the hypothesis  $H$  corresponds to a theoretical value  $V_0$  of the observation  $V$  which actually varies around  $V_0$  because of inaccuracy of measurement, whilst  $\bar{H}$  is assumed to represent the absolute disorder, i.e. the absence of such a value. In terms of image, it means that the hypothesis that two features are linked by a specific relationship  $\mathcal{R}$  (due to related 3D features) is opposed to the situation where these features are completely independent (unrelated 3D features). The hypothesis  $\bar{H}$  is represented by a statistical model corresponding to a completely unstructured image  $I_n$ , that is to say a set of straight line segments randomly placed, with a random direction (see chapter 3). Of course it is an extreme case and intermediate events between  $H$  and  $\bar{H}$  are possible. However, most often these intermediate events are not individually observable and are considered events of  $\bar{H}$ . For instance, in a room the location of an edge of the window is related to the location of an edge of the door, so should be their images ; in practice, the information about the connection is extremely diffused in the picture and may be ignored in a first place. This is no longer true for two consecutive edges of the door, as their images are always connected. Thus, the

problem may be rewritten : is the observation in the image significant of a well known 3D relationship or could it occur by chance?

*Model of probability*

Let  $\mathcal{V}$  be the variable under observation and  $V$  its value. The hypothesis  $H$  may be modelled by a Gaussian law with zero mean and covariance matrix  $C_v$  representing errors of measurement (if the mean is not zero but  $V_m$ ,  $\mathcal{V}-V_m$  is considered). The hypothesis  $\bar{H}$  is modelled by the distribution of  $\mathcal{V}$ , corresponding to the unstructured image  $I_n$ . Using the decomposition over the two exclusive events  $H$  and  $\bar{H}$

$$\begin{aligned} p(\mathcal{V}=V) &= p(\mathcal{V}=V|H)p(H) + p(\mathcal{V}=V|\bar{H})p(\bar{H}) \\ &= p(H) p_1(V) + p(\bar{H}) p_2(V) \end{aligned} \tag{A3.1}$$

where

$$p_1(V) = \frac{1}{\sqrt{2\pi}^d \sqrt{\det C_v}} \exp\left(-\frac{1}{2} V^t C_v^{-1} V\right) \tag{A3.2}$$

where  $d$  is the dimension of  $\mathcal{V}$  and  $p_2(V)$  depends on  $\mathcal{V}$ . Usually,  $p_2(V)$  is nearly constant around 0, so that the density of probability of  $\mathcal{V}$  has a general form as displayed in figure A3.1.

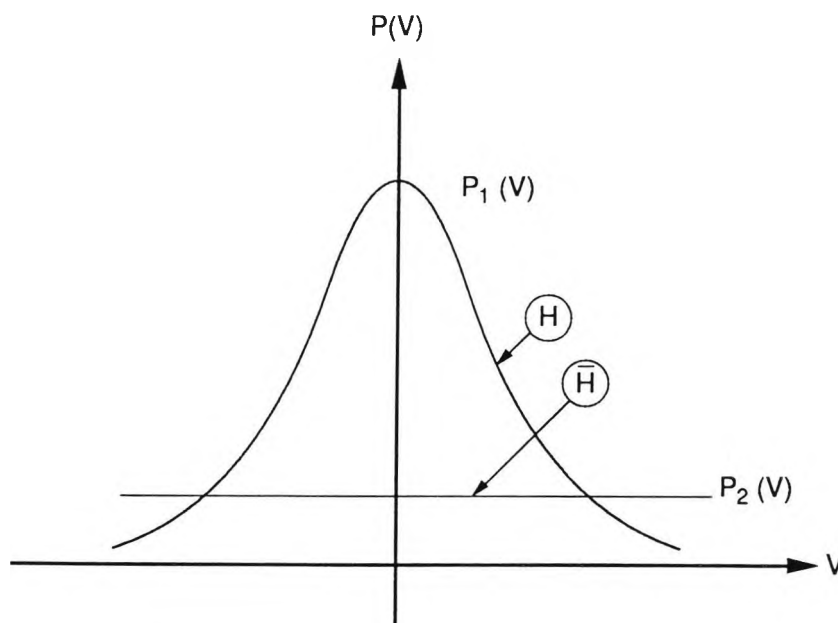


Figure A3.1 : Model of probability for  $V$  for hypothesis  $H$  and hypothesis  $\bar{H}$ .

The determination of  $p_1$  is a mere calculation of covariance matrix but the determination of  $p_2$  depends on  $V$  and  $I_n$ , and may be complex to determine.

In order to estimate  $p(H)$  and  $p(\bar{H})$ , the initial hypotheses should be completed. Let  $\mathcal{R}$  be the relation studied between the features  $(F_i, F_j)$  of the image, and  $V$  the decision variable. Let  $\Omega$  be the set of pairs of features  $(F_i, F_j)$ ,  $\Omega_1$  the set of pairs of features linked by the relationship  $\mathcal{R}$ , and  $\Omega_2$  the set of pairs of independent features. Then,

$$p(H) = \frac{\text{Card}(\Omega_1)}{\text{Card}(\Omega)} \quad \text{and} \quad p(\bar{H}) = \frac{\text{Card}(\Omega_2)}{\text{Card}(\Omega)} \quad (\text{A3.3})$$

#### *Definition of the LR test*

The likelihood ratio test consists in determining whether a pair of features is more likely to be element of  $\Omega_1$  or element of  $\Omega_2$ . Therefore the likelihood ratio is defined as

$$R = \frac{p(H|V)}{p(\bar{H}|V)}$$

and the test (i.e. LR test) succeeds if  $R \geq 1$ . Using Bayes' theorem, the LR test may be rewritten

$$R = \frac{p(H) p_1(V)}{p(\bar{H}) p_2(V)} \geq 1 \quad (\text{A3.4})$$

*Comparison of the LR test with the MD test and the neighbourhood test*

In order to compare the likelihood ratio test with the MD test previously defined, the dimension of  $V$  is supposed to be 1 for simplicity. Let  $\sigma_v^2$  be the variance of  $V|H$  and  $\sigma_{v0}^2$  the expected value of  $\sigma_v^2$  on  $\Omega_1$ . The quantity  $p_2(V) p(\bar{H})/p(H)$  is assumed to be approximately equal to a constant  $p_0$  around zero, which is independent of  $\sigma_v$  as  $p_2$  does not depend on error of measurement by hypothesis. Therefore, the LR test eq. A3.4 may be rewritten

$$\frac{V^2}{\sigma_v^2} < \nu_0 - 2 \text{Log}_e \frac{\sigma_v}{\sigma_{v0}} = \nu(\sigma_v) \quad (\text{A3.5})$$

with  $\nu_0 = -2 \text{Log}_e (\sqrt{2\pi} \sigma_{v0} p_0)$

This relation can be easily generalized to any dimension  $d$  of the decision variable  $V$

$$V^t C_v^{-1} V < \nu_0 - 2 \text{Log}_e \frac{\det C_v}{\det C_{v0}} = \nu(C_v) \quad (\text{A3.6})$$

with  $\nu_0 = -2 \text{Log}_e (\sqrt{2\pi}^d \sqrt{\det C_{v0}} p_0)$

The threshold  $\nu(C_v)$  on the Mahalanobis distance depends on the covariance matrix of measurement  $C_v$ .

In order to compare the test with the neighbourhood test, the

variations of  $\sigma_v^2 \nu(\sigma_v)$  are studied for  $d=1$ . The value of  $\sigma_v^2 \nu(\sigma_v)$  increases with  $\sigma_v$ , then reaches its maximum value for

$$\sigma_m = \sigma_0 \exp((\nu_0 - 1)/2),$$

and decreases to 0, reached for

$$\sigma_M = \sigma_0 \exp(\nu_0/2).$$

If  $\sigma_v$  is higher than  $\sigma_M$ , the value of  $\sigma_v^2 \nu(\sigma_v)$  is negative and the test fails whatever the value of  $V$ . This is comforting as it means that the uncertainty of  $V$  is so big that  $V$  has no significance at all.

When the uncertainty is small, the LR test is less constraining than the MD test whatever the fixed threshold. When the uncertainty is very large, the LR test always failed. When the uncertainty is around its expected value then the LR test is equivalent to a MD test with a threshold equal to  $\nu_0$ . The behaviour of the LR test is therefore intermediate between a neighbourhood test, i.e.  $V > V_{\max}$ , and the MD test. It is more satisfactory than this latter test because it eliminates very uncertain, and thereby not significant features whereas the risk to miss a very significant feature is smaller. The risk  $\beta$  of selecting a feature by mistake, i.e. the type II error, is bounded by  $\beta_{\max}$ :

$$\beta_{\max} = 2 p_2 \sigma_m \quad (\text{A3.7})$$

where  $p_2 = p_0 p(H)/p(\bar{H})$ .

*Definition of  $p(x)$ , where  $x=p(H)$*

The prior probability  $p(H)$  has been found by comparing the value of the peak  $P$ , to the expected value of this point when the lines are supposed randomly crossing the vertical of the accumulator space passing by  $P$ . Let  $\alpha_0$  be  $p(H)$ . It is possible to refine this value by taking into



account not only the value of the peak P of the accumulator space, but also the real distribution of the V corresponding to the directions elements of this peak. If this distribution is the ideal representation of the ideal case, it should be proportional to  $\alpha_0 p_1(V) + (1-\alpha_0) p_2(V)$  (cf eq. A3.1). However, it may be better represented by  $\alpha p_1(V) + (1-\alpha) p_2(V)$ ,  $\alpha \neq \alpha_0$ . In order to find the best value of  $\alpha$ , the set of all  $\alpha \in [0,1]$  are considered,  $nn \alpha$  being the expected number of directions meeting at P "on purpose",  $nn (1-\alpha)$  being the expected lines passing near P by chance (considered as noise). The probability of having such a number of lines passing near P by chance determines the probability of the hypothesis  $\alpha$ . Using the model of the noise described in section 5.2.3, it comes

$$p(\alpha) = p(nn_2 = (1-\alpha)nn | nn) = \frac{\binom{n'}{(1-\alpha)nn} p^{(1-\alpha)nn} (1-p)^{n'-(1-\alpha)nn}}{\sum_{k \leq nn} \binom{n'}{k} p^k (1-p)^{n'-k}} \quad (A3.8)$$

where  $p = 4\bar{\sigma}_p / y'_{\max}$  (see eq. (5.2.4.1)) and  $n'$  is the number of directions associated to noise crossing the vertical  $x'$ ; using eq. (5.2.4.2) and substituting  $f(\alpha)$  by  $nn (1-\alpha_0)$ ,  $n'$  is equal to

$$n' = \frac{nn (1-\alpha_0)}{p} \quad (A3.9)$$

Using a Gaussian approximation of the binomial law,  $p(\alpha)$  may be rewritten :

$$p(\alpha) \approx \frac{\exp\left(-\frac{(\alpha-\alpha_0)^2 nn^2}{2\omega}\right)}{\sqrt{2\pi \omega} (0.5 + \operatorname{erf}(\alpha_0 nn / \sqrt{\omega}))}, \quad (A3.10)$$

where  $\omega = (1-\alpha_0)(1-p)nn$  and  $(0.5 + \operatorname{erf}(\alpha_0 nn / \sqrt{\omega}))$  is a normalization factor corresponding to the constraint  $\alpha \leq 1$ .

## APPENDIX 4

## STATISTICS OF THE PARAMETERS OF THE SEGMENTS

Using the statistical model described in the text, the segment (AB) may be generated by any point G located on the string (S) of the image disk, such that (AB) lies on (S) and has its centre is G. The string (S) is at the distance d from the centre of the image. G is assumed to have a uniform density of probability on (S), therefore the density of probability of d is proportional to the length of the string (S)

$$f(d) = 2 K R \sqrt{1-d^2/R^2} .$$

The constant K is found by integration of f(d) over d ( $-R \leq d \leq R$ )

$$K = 1/(\pi R^2).$$

The expected value of  $d^2$  is equal to

$$E(d^2) = (2/(\pi R)) \int_{-R}^R d^2 \sqrt{1-d^2/R^2} \partial d = R^2/4 \quad (\text{A4.1})$$

Let b be the distance between the centre G of (AB) and the centre of the string (S). The expected value of  $b^2$ , knowing d, is equal to

$$E(b^2|d) = \frac{1}{2\sqrt{R^2-d^2}} \int_{-\sqrt{R^2-d^2}}^{\sqrt{R^2-d^2}} b^2 \partial b = \frac{R^2-d^2}{3} \quad (\text{A4.2})$$

*Probabilistic model for the distribution of the directions relative to d*

It has been seen above that the density of probability of d, distance from a straight line to the origin O is equal to

$$f(d) = \frac{2}{\pi R} \sqrt{1-d^2/R^2}. \quad (\text{A4.3})$$

The density of probability of  $d$  for a direction should be corrected by the accumulation process which is performed over the lines before accumulation. The resolution in  $d$  and  $\theta$  of the accumulation process is given by (see eq. A5.15 and A5.4), using the assumption  $V(u)=\sigma_0^2$  for clarity,

$$\sqrt{E(\partial d^2)} = \sqrt{c-2a \cdot d^2/3} \sigma_0 \quad \text{and} \quad \sqrt{E(\partial \theta^2)} = \sqrt{2a} \sigma_0$$

Therefore the probability for a line to be accumulated in a cell corresponding to the direction  $(d, \theta)$  is

$$g(d) = \frac{2\sqrt{2a} \sigma_0^2}{\pi^2 R} \sqrt{1-d^2/R^2} \cdot \sqrt{c-2a \cdot d^2/3}. \quad (\text{A4.4})$$

The probability for the direction  $(d, \theta)$  to have been accumulated at least once when  $n$  lines have been accumulated is  $1 - (1-g(d))^n$ , therefore the expected number of directions with a distance from the origin inferior to  $d$  is

$$n'(d) = \frac{\pi}{\sqrt{2a} \sigma_0} \sum_{d_i < d} (1 - (1-g(d_i))^n). \quad (\text{A4.5})$$

## APPENDIX 5

## UNCERTAINTY PARAMETERS ASSOCIATED WITH A LINE (L) IN THE IMAGE

Let  $\alpha$  be the polar angle of the line (L),  $d$  its distance to the origin,  $\vec{u}$  its unit vector and  $\vec{u}_1$  the perpendicular unit vector. Let  $G$  be the centroid of the segment  $AB$ ,  $L$  its length,  $O'$  the projection of the origin  $O$  onto (L) and  $b$  the algebraic distance between  $O'$  and  $G$  (Figure 5.2.2.2).

The line (L) is known within the errors of measurement  $\partial A$  and  $\partial B$  of the end points  $A$  and  $B$ . These errors result in errors on  $\alpha$ ,  $d$ ,  $\vec{u}$  and  $\vec{u}_1$ .

$$\partial\alpha = (\vec{\partial B} \cdot \vec{u}_1 - \vec{\partial A} \cdot \vec{u}_1) / \ell \quad (\text{A5.1})$$

And

$$\vec{\partial u} = \partial\alpha \vec{u}_1 ; \partial\vec{u}_1 = -\partial\alpha \vec{u} \quad (\text{A5.2})$$

Now  $d = \vec{OA} \cdot \vec{u}_1$ , then  $\partial d = \vec{\partial A} \cdot \vec{u}_1 + \vec{OA} \cdot \partial\vec{u}_1 = \vec{\partial A} \cdot \vec{u}_1 - \vec{OA} \cdot \vec{u} \partial\alpha$

Substituting  $\vec{OA} \cdot \vec{u}$  by  $(b-L/2)$  and  $\partial\alpha$  by (A5.1), it is found that

$$\partial d = (b/\ell + 1/2) \vec{\partial A} \cdot \vec{u}_1 - (b/\ell - 1/2) \vec{\partial B} \cdot \vec{u}_1 \quad (\text{A5.3})$$

The variance of  $(\vec{\partial A} \cdot \vec{u}_1)$  and  $(\vec{\partial B} \cdot \vec{u}_1)$  is  $V(\vec{u})$  (see eq. 4.4.1 and fig. 4.4.1.2 in the text), therefore the second moments of  $\partial\alpha$  and  $\partial d$  are

$$E(\partial\alpha^2) = 2 V(\vec{u}) / \ell^2 \quad (\text{A5.4})$$

$$E(\partial d^2) = (2b^2/\ell^2 + 1/2) V(\vec{u}) \quad (\text{A5.5})$$

$$E(\partial\alpha \partial d) = -2b V(\vec{u}) / \ell^2 \quad (\text{A5.6})$$

Let  $D$  be the distance from a fixed point  $P$  to the line (L),  $D$  is

equal to

$$D = AP \cdot \vec{u}_1 \quad (A5.7)$$

The derivation of (A5.7) provides  $\partial D$

$$\partial D = -\vec{\partial A} \cdot \vec{u}_1 + \vec{AP} \cdot \partial \vec{u}_1$$

Let Q be the projection of P onto (L), using (A5.2) and (A5.3) and then substituting  $\vec{AP} \cdot \vec{u}_1$  by AQ and (AQ-L) by BQ, it is found that

$$\partial D = (BQ/\ell) \vec{\partial A} \cdot \vec{u}_1 - (AQ/\ell) \vec{\partial B} \cdot \vec{u}_1 \quad (A5.8)$$

$$E(\partial D^2) = (AQ^2 + BQ^2) V(\vec{u}) / \ell^2 \quad (A5.9)$$

Let  $r'$  be the distance from  $O'$  to Q, (A5.9) may be rewritten

$$E(\partial D^2) = (2(r'+b)^2/\ell^2 + 1/2) V(\vec{u}) \quad (A5.10)$$

$O'Q' = \sqrt{r'^2 - d^2}$  and  $OQ = r'$ . The vector  $\vec{QQ'}$  depends only upon the error on the end points, therefore the substitution of Q by  $Q'$  in (A5.8) only adds terms of degree higher than 1 in  $\vec{\partial A} \cdot \vec{u}_1$  and  $\vec{\partial B} \cdot \vec{u}_1$ . Thus the following approximation can be made

$$r' = \sqrt{r^2 - d^2}.$$

Let  $\sigma^2$  be the expected value of  $\partial D^2$ , square of the distance from a given line (L) and a fixed point P. The variations of  $\partial D$  are due to the errors of measurement of the end points. Developing (A5.9) by substituting  $r'$  by  $\sqrt{r^2 - d^2}$ , it gives

$$\sigma^2 = (2r^2/\ell^2 + 4br(1-d^2/r^2)/\ell^2 + 2b^2/\ell^2 + 1/2 - 2d^2/\ell^2) V(\vec{u})$$

If  $d$  is fixed and  $\sigma^2$  is averaged over all the segments AB producing the line (L), then by using (A4.2) and the fact that  $E(b) = 0$  for symmetry reasons, it is found that

$$E(\sigma^2|d) = (2a r^2 - 8a d^2/3 + 2a R^2/3 + 1/2) V(\vec{u})$$

where  $a = E(1/\ell^2)$  (for the independence between  $a$  and the other parameters, see section 4.4.2). If the line (L) passes through P then  $|d| \leq r$ . Therefore, to compute the expected value of  $\sigma^2$  over the set of lines (L) crossing at the point P, 2 cases should be considered whether  $r$  is larger or smaller than  $R$ . If  $r$  is larger than  $R$ ,  $E(d^2) = R^2/4$  (see (A4.1)) and using the assumption  $V(\vec{u}) = \sigma_0^2$  for clarity, the relation (A6.1) may be rewritten

$$E(\sigma^2) = (2a r^2 + 1/2) \sigma_0^2 \quad (\text{A5.11})$$

If  $r$  is smaller than  $R$  then (Integration by parts)

$$E(d^2) = R^2/4 (1 - 2r(1-r^2/R^2)^{3/2}/(R \text{Arcsin}(r/R) + (R^2-r^2)^{1/2}))$$

When  $r$  is small

$$E(d^2) = r^2/3$$

and substituting in (A5.11)

$$E(\sigma^2) = (10a r^2/9 + 2a R^2/3 + 1/2) \sigma_0^2 \quad (\text{A5.13})$$

It is more convenient to use an approximation valid for any  $r$ . Besides the comparison with the Gaussian sphere method requires that such an approximation can be made. Using (A5.12) and (A5.13),  $E(\sigma^2)$  is approximated for any  $r$  by

$$E(\sigma^2) = (2a r^2 + c) \sigma_0^2, \quad (\text{A5.14})$$

where  $c = 2a R^2/3 + 1/2$ .

*Determination of the resolution in  $d$  of the preaccumulation stage.*

The set of lines at a distance  $d$  from 0 is considered. As  $b$  is assumed independent of  $1/\ell^2$  (see section 5.2), the expression (A5.5) becomes

$$E(\partial d^2 | d) = (2E(b^2)E(1/\ell^2) + 1/2) V(\vec{u})$$

and using eq. A4.2

$$E(\partial d^2 | d) = (c - 2a d^2/3) V(\vec{u}) \quad (\text{A5.15})$$

with  $a = E(1/\ell^2)$  and  $c = 2a R^2/3 + 1/2$ . This value fixes the resolution of the preaccumulation stage for this  $d$ .

## APPENDIX 6

## UNCERTAINTY PARAMETERS IN THE EUCLIDEAN IMAGE PLANE

*Calculation of the variance of the angle  $(OV_1, OV_2)$ .*

The reasoning in chapter 6 is performed using a Euclidean system with the origin at the principal point location, whilst the covariance matrix associated with a vanishing point candidate  $V$  of the image is expressed in the image coordinate system, which is not Euclidean. Let  $(x, y)$  be the Cartesian coordinates of  $V$  in the Euclidean system and  $(x', y')$  be the Cartesian coordinates of  $V$  in the image system, then

$$(x, y) = (x' + x'_0, (y' + y'_0)/\rho)$$

The covariance matrix of  $V$  in the Euclidean system is

$$\begin{cases} \sigma_x^2 = \sigma_{x'}^2 + \sigma_{x'_0}^2 \\ \sigma_y^2 = (\sigma_{y'}^2 + \sigma_{y'_0}^2)/\rho^2 + y^2 \sigma_\rho^2/\rho^2 \\ \sigma_{xy} = \sigma_{x'y'}/\rho \end{cases} \quad (\text{A6.1})$$

The coordinates of the principal point are assumed uncorrelated (for simplicity as there is no technical difficulty in taking their correlation into account).

Let  $\theta$  be the polar angle of  $V$  in the Euclidean system, then

$$\sigma_\theta^2 = \frac{x^2 \sigma_y^2 + y^2 \sigma_x^2 - 2xy \sigma_{xy}}{x^2 + y^2}$$

Now, let  $\theta$  be equal to  $\theta_1 - \theta_2$ , where  $\theta_1$  and  $\theta_2$  are the polar angles of two vanishing point candidates  $V_1$  and  $V_2$ , then  $\sigma_\theta^2$  is deduced from the previous equation applied to  $V_1$  and  $V_2$



$$\sigma_{\theta}^2 = \sigma_{\theta_1}^2 + \sigma_{\theta_2}^2 \quad (\text{A6.2})$$

Calculation of  $\sigma_c^2$ , variance of  $V$  when the calibration parameters vary.

The decision variable associated with the perpendicularity test  $V$  is equal to

$$V = \overrightarrow{OV}_1 \cdot \overrightarrow{OV}_2 + f^2$$

in the Euclidean coordinate system  $(Ox, Oy)$ , where  $O$  is equal to the principal point,  $f$  the distance of the image plane from the optic centre. In this system, the decision variable  $V$  associated with the perpendicularity test of  $V_1$  and  $V_2$  may be written

$$V = \overrightarrow{OV}_1 \cdot \overrightarrow{OV}_2 + f^2 \quad (\text{A6.3})$$

Then

$$\partial V = 2f\partial f + x_2 \partial x_1 + x_1 \partial x_2 + y_2 \partial y_1 + y_1 \partial y_2 \quad (\text{A6.4})$$

And (A6.1) is used for computing  $\sigma_c^2 = E((\partial V)^2)$  from

$$\begin{aligned} \sigma_c^2 = & 4f^2 \sigma_f^2 + x_2^2 \sigma_{x1}^2 + x_1^2 \sigma_{x2}^2 + 2 x_2 y_2 \sigma_{x1y1} \\ & + y_2^2 \sigma_{y1}^2 + y_1^2 \sigma_{y2}^2 + 2 x_1 y_1 \sigma_{x2y2} \end{aligned} \quad (\text{A6.5})$$

The distance  $f$ , the scale ratio  $\rho$  and the principal point coordinates have been assumed uncorrelated.

*Determination of  $\sigma_h$  associated with a vanishing point corresponding to a horizontal direction.*

Let  $V'$  be the vanishing point corresponding to the horizontal direction studied and  $V$  the vanishing point corresponding to the perpendicular vertical direction. Then if  $O'$  is the projection of  $V'$  onto the line  $OV$ ,  $O$  being the principal point, then

$$OO' \cdot OV = -f^2$$

Therefore, after differentiation it comes

$$\text{var}(O'_y) = (1 - 2 OO'/r) \text{var}(O_y) + 4 \text{var}(f)/r^2 + OO'^2 \text{var}(r)/r^2$$

where  $r = OV$ .

Only  $\text{var}(r)$  is not known. Let  $C_v$  be the covariance matrix associated with the vertical direction in the accumulator space (defined in chapter 5).

$$C_v = \begin{pmatrix} \sigma_{vx'}^2 & 0 \\ 0 & \sigma_{vy'}^2 \end{pmatrix}$$

Now, using the sampling 5.2.2.15 from the image system to the accumulator space and  $dr' = (\partial r'/\partial x') dx'$  where  $r'$  is the polar distance in the image coordinate system,

$$\text{var}(r') = \frac{4 \sigma_0^2 r'^2 (2ar'^2 + c)}{\hat{\sigma}_{x'}^2 R^2} \sigma_{vx'}^2$$

Since the camera is assumed to be approximately vertical,  $r$  is very large and  $r \approx r'/\rho$ . Therefore, the variance of  $O'$  along the vertical is equal to

$$\text{var}(O'_y) \approx \text{var}(O_y) + \frac{8f^4 \sigma_0^2 a \rho^2}{\hat{\sigma}_{x'}^2 R^2} \sigma_{vx'}^2 \quad (\text{A6.6})$$



## APPENDIX 7

## ZERO-CROSSING PROBLEM

APPLICATION TO EDGE DETECTION  
AND TO THE NUMBER OF FALSE ALARMS IN THE ACCUMULATOR SPACE*Zero-crossing problem*

To extract meaningful information from a noisy signal, e.g. the image or the accumulator space, it is important to know the significance of the value of the signal at a particular point, that is to say the probability of this value to happen by chance or for a particular reason. The problem of determining the statistical properties of the crossings of the signal by a level  $a$  is known as the  $a$ -crossing problem, which is in fact a generalisation of the well known "zero-crossing problem".

Details about the zero-crossing problem are given in (Rice, 1944, 45) and (Papoulis, 1965). A short presentation of this problem is given in the first part of this appendix. The demonstrations are not complete for the sake of shortness, but the main hints are given. The results are then used to describe the response of an edge detector to a noisy step. The Gaussian filter is used as an example, but the approach is similar for any other filter. Then, they are applied to a very different problem, the expected number of false alarms in the accumulator space.

Let  $s(x)$  be a stationary normal process with zero mean and an autocorrelation function  $R(\xi)$  defined by

$$R(\xi) = E(s(x+\xi)s(x)).$$

If  $s(x+\xi)s(x) < 0$ , then the number of zero-crossings between  $x$  and  $x+\xi$  is odd. Let  $p(\xi)$  the probability of having an odd number of zero-crossings between  $x$  and  $x+\xi$ .

Let  $y$  be the random variable  $s(x)$  and  $z$  the random variable  $s(x+\xi)$ , then the random variable  $u = y/z$  has a Cauchy density, the distribution function of which is (Papoulis, 1965)

$$F(u) = \frac{1}{2} + \frac{1}{\pi} \arctan \frac{u - r}{\sqrt{1-r^2}},$$

where  $r = R(\xi)/R(0)$ . As  $p(\xi)$  is equal to  $F(0)$ , then

$$p(\xi) = \frac{1}{2} + \frac{1}{\pi} \arctan \frac{r}{\sqrt{1-r^2}} = \frac{\text{Arccos } r}{\pi} \quad (\text{A7.1})$$

If  $\xi$  is small, then  $p(\xi)$  is also small and the equation (A7.1) may be developed to the first order

$$r \approx 1 - \frac{\pi^2 p^2(\xi)}{2}$$

and  $p(\xi)$  is deduced

$$p(\xi) \approx \frac{1}{\pi} \sqrt{\frac{2(R(0)-R(\xi))}{R(0)}} \quad (\text{A7.2})$$

and using the development of  $R(\xi)$  at the first order, it comes that

$$p(\xi) \approx \frac{1}{\pi} \sqrt{-\frac{2R'(0^+)\xi}{R(0)}}, \quad (\text{A7.3a})$$

if the derivative of  $R$  is discontinuous at the origin and

$$p(\xi) \approx \frac{1}{\pi} \sqrt{-\frac{R''(0)}{R(0)} \xi}, \quad (\text{A7.3b})$$

otherwise.

Let  $u$  be  $s(x)$ , and  $v$  be  $s(x+\xi)$ , it is possible to show that if  $R(\xi) \approx R(0)$  (i.e.  $\xi \approx 0$  and  $R$  smooth enough around zero),

$$p((u-a)>0 \ \&\& \ (v-a)<0) \approx p(u>0 \ \&\& \ v<0) \exp\left(-\frac{a^2}{2R(0)}\right) \quad (\text{A7.4})$$

Notes :

1) Joint density function of  $u=s(x)$  and  $v=s(x+\xi)$

$$\frac{1}{2\pi \sigma^2 \sqrt{1-r^2}} \exp\left(-\frac{1}{2\sigma^2(1-r^2)}(u^2+v^2-2ruv)\right).$$

2) A7.4 may be shown by integrating this density function over  $[a, +\infty][-\infty, a]$  with  $u'=u-a$ ,  $v'=v-a$ , then by using the fact that  $r \approx 1$ .

The same relation as (A7.4) is true on the opposite quadrant so that

$$p((u-a)(v-a) < 0) \approx p(uv < 0) \exp\left(-\frac{a^2}{2R(0)}\right) \quad (\text{A7.5})$$

If  $\xi$  is small then  $R(\xi) \approx R(0)$ , i.e.  $r \approx 1$ , and the probability of having an odd number of zero-crossings or crossings by a level "a" is nearly equal to the probability of having only one crossing by this level between  $x$  and  $x+\xi$ . Let  $p_a(\xi)$  be the probability of having one crossing by the level "a" between  $x$  and  $x+\xi$ , using A7.2 or A7.3 and A7.5, it may be approximated by

$$\begin{aligned} \text{if } R'(0^+) \neq 0 \quad \text{then} \quad p_a(\xi) &= \frac{1}{\pi} \sqrt{-\frac{2R'(0^+)\xi}{R(0)}} \exp\left(-\frac{a^2}{2R(0)}\right) \\ \text{if } R'(0) = 0 \quad \text{then} \quad p_a(\xi) &= \frac{1}{\pi} \sqrt{-\frac{R''(0)}{R(0)}} \exp\left(-\frac{a^2}{2R(0)}\right) \xi \end{aligned} \quad (\text{A7.6})$$

This represents the expected density of crossings by a level "a" at any point. Now, the expected density of crossings may be different at particular points, such as the crossing points. In the following the density of zero-crossings (given by A7.7) around a zero-crossing is studied in the regular case, i.e.  $R'(0) = 0$ . Let  $P_0(\xi)$  be the probability  $P(s'(x)s(x+\xi) < 0 | s(x) = 0)$ . Using the joint density of  $s'(x)$  and  $s(x+\xi)$  and the relation A7.2, it comes

$$P_0(\xi) \approx \frac{1}{\pi} \sqrt{2(1-r')} = \frac{-R'(\xi)}{\sqrt{-R''(0)(R(0)-R^2(\xi)/R(0))}}$$

where  $r'$  is the correlation coefficient of the random variables  $s'(x)|s(x)$  and  $s(x+\xi)|s(x)$ . Let  $P_1(\xi, d\xi)$  the probability that there is an odd number of zero crossings between  $\xi$  and  $\xi+d\xi$ , knowing that  $s(0)=0$ , when  $\xi$  is small. It may be shown by using a Lagrange development around  $(\xi, \zeta) = (0, 0)$  that

$$P_1(\xi, d\xi) \approx \frac{1}{2\pi} \sqrt{\frac{R''(0)}{R(0)} - \frac{R^{(4)}(0)}{R''(0)}} d\xi,$$

Therefore the density of zero-crossings around a zero-crossing is equal to

$$\lambda_0 = \frac{1}{2\pi} \sqrt{\frac{R''(0)}{R(0)} - \frac{R^{(4)}(0)}{R''(0)}} \quad (A7.7)$$

*Application to edge detection*

The Gaussian filter is used to illustrate the reasoning, but it could have been any other filter shape. The Gaussian filter may be written

$$J = \sqrt{(I * G_y * G'_x)^2 + (I * G_x * G'_y)^2},$$

where  $I$  is the initial image,  $J$  the filtered image,  $G_x$  and  $G_y$  are Gaussian functions with zero mean and variance  $\sigma_f^2$ . Then a non-maxima suppression algorithm is performed.

The edge detector is applied to a perfect step in the  $y$  direction on which has been added white noise  $w$  with variance  $\sigma_n^2$

$$\begin{aligned} I(x, y) &= w(x, y) && \text{if } y < y_0 \\ I(x, y) &= \Delta_0 + w(x, y) && \text{if } y \geq y_0 \end{aligned} \quad (A7.8)$$

Let us consider  $J_y = I * G_x * G'_y$ . The convolution of  $I$  by the Gaussian filter may be divided into two parts, the convolution of the signal and the convolution of the noise. The correlation functions along  $x$  and  $y$

corresponding to noise are

$$R_x(\xi) = \frac{\sigma_n^2}{8 \pi \sigma_f^4} \exp\left(-\frac{\xi^2}{4\sigma_f^2}\right)$$

$$R_y(\xi) = \frac{\sigma_n^2}{8 \pi \sigma_f^4} \left(1 - \frac{\xi^2}{2\sigma_f^2}\right) \exp\left(-\frac{\xi^2}{4\sigma_f^2}\right)$$
(A7.9)

and  $\Delta$  the magnitude of the edge found by the filter is

$$\Delta = \frac{\Delta_0}{\sqrt{2\pi} \sigma_f}$$

Once the edge detector is performed, a non-maxima suppression algorithm is performed and the image output is thresholded at the value  $T$ . First the effect of the thresholding is studied. It affects mainly the  $x$  direction. If  $T=\Delta$ , the density of endpoints should be equal to the density of zero-crossings given by A7.6, where  $R(\xi)$  is equal to  $R_x(\xi)$ . So

$$\delta_0 = \frac{1}{\sqrt{2} \pi \sigma_f}$$

The density of crossings by a level " $T$ " is

$$\delta_T = \frac{1}{\sqrt{2} \pi \sigma_f} \exp\left(-\frac{(\Delta - T)^2}{2 R(0)}\right).$$
(A7.10)

The average length  $\ell_0$  of the segments when thresholding at  $\Delta$  is equal to the average length of the holes for reasons of symmetry, thereby it may be defined as the limit of  $L/n$  when  $n$  tends towards infinity, where  $L$  is the total length of the edge, and  $n$  the number of endpoints. As  $n/L$  may be considered as the average value of the random variable  $\nu(x)$ , equal to 1 if  $x$  is an endpoint and 0 otherwise, Borel's theorem may be applied and  $n/L$  tends towards  $\delta_0$  with the probability 1. Therefore



$$\ell_0 = \frac{1}{\delta_0} \quad (\text{A7.11})$$

This means that if a threshold  $\Delta$  is applied to a Gaussian edge image, then the average length of the segments of an edge having a magnitude equal to  $\Delta_0$  is  $\ell_0$  (the boundary segments are not included). Consequently a segment with a length smaller than  $\ell_0$  is more likely to correspond to an edge with a magnitude less than  $\Delta_0$  than to an edge with a magnitude more than  $\Delta_0$ .

Using the above expression of  $\lambda$  in the case of an edge of magnitude  $\Delta_0$  and thresholding at  $\Delta$ , the density of endpoints around an endpoint is equal to

$$\lambda = \frac{1}{2\pi\sigma_f} \quad (\text{A7.12})$$

This means that knowing that the point P is an endpoint, the probability of having another endpoint nearby is lower than the one for any neighbourhood. This means that the endpoints are more equally distributed along the edge than it would be for a random distribution with the same density  $\delta_0$ , i.e. a Poisson process.

Now the effect of the non-maxima suppression algorithm is studied, which affects almost only the y direction. The points  $J(x,y)$  such that  $J'(x,y-dy) > 0$  &&  $J'(x,y+dy) < 0$  are selected, the other points are suppressed. Canny's uniqueness criterion only considers the maxima due to noise ; thus applying A7.3b to  $w * G_x * G_y''$  and the fact that only the maxima are selected (and not the minima) it found that the density of maxima due to noise at the point  $(x,y+y_0)$  is

$$\mu_0 = \frac{1}{2\pi} \sqrt{-\frac{R^{(4)}(0)}{R''(0)}} ,$$

where  $R(\xi) = R_y(\xi) = \frac{\sigma_n^2}{8\pi\sigma_f^4} \left(1 - \frac{\xi^2}{2\sigma_f^2}\right) \exp\left(-\frac{\xi^2}{4\sigma_f^2}\right)$ ,  $\mu_0$  is equal to

$$\mu_0 = \frac{1}{2\pi\sigma_f} \sqrt{\frac{5}{2}}.$$

In fact the density of maxima is different at the edge location  $y=y_0$ , as it should take into account the variations of the signal. The number of zero-crossings of  $s(\xi) = J'_y(y_0+\xi)$  between  $-Y/2$  and  $Y/2$  is given by (Blanc-Lapierre, 1963, pp34)

$$N(Y) = \int_{-Y/2}^{Y/2} |s'(\xi)| \delta(s(\xi)) \partial\xi,$$

where  $\delta(s)$  is the dirac measure at  $s$ . At  $\xi$  fixed the expected value of the integrant is

$$E(|s'(\xi)| \delta(s(\xi))) = \int_{-\infty}^{+\infty} |s'| p(s', 0) ds'$$

where  $p(s', s)$  is a Gaussian density centred at  $(E(s'(\xi)), E(s(\xi)))$  with variances  $R^{(4)}(0)$ ,  $R''(0)$  and a covariance equal to 0. Let  $\zeta$  be

$$\zeta = s'(0) / \sqrt{R^{(4)}(0)}$$

$\zeta$  is the signal to noise ratio of  $J''_y(y_0)$ . Then, if  $E(s(0)) = 0$ , the density of zero-crossings of  $s$  at  $y_0$  is

$$E(|s'(0)| \delta(s(0))) = \frac{1}{\pi} \sqrt{\frac{R^{(4)}(0)}{R''(0)}} (\sqrt{2\pi} \zeta \operatorname{erf}(\zeta) + \exp(-\frac{\zeta^2}{2}))$$

If  $\zeta$  is large, i.e.  $\sigma$  small,  $E(|s'(0)| \delta(s(0))) \cong \mu_0 \sqrt{2\pi} |\zeta|$ . If  $\zeta$  tends towards infinity then the density tends towards infinity, i.e. towards  $|s'(0)| \delta(s(0))$ , representing the fact that the signal surely crosses 0 at  $\xi=0$ . It is possible to show that  $\mu_0 \sqrt{2\pi} |\zeta|$  is the density corresponding to the maxima looked for (the corresponding integral over  $[-\infty, +\infty]$  is equal to 1), the density corresponding to other maxima, i.e. parasite maxima, is

$$\mu = \mu_0 (\sqrt{2\pi} (|\zeta| (\operatorname{erf}(\zeta) - 0.5) + \exp(-\frac{\zeta^2}{2}))). \quad (\text{A7.13})$$

(the division by 2 is due to the fact that only maxima are considered).  $\mu$  is decreasing from  $\mu_0$  to 0 when  $|\zeta|$  is increasing from 0 to  $+\infty$ . The ratio  $r$  introduced by Canny (1986) as a uniqueness response criterion, verifies

$$\zeta = r J_y(y_0)/\sqrt{R(0)}$$

where  $J_y(y_0)/\sqrt{R(0)}$  is the signal to noise ratio of  $J_y$  at  $y_0$ . The largest  $r$ , the smaller the risk of multiple response. However, it is not understood why  $r$  should be as near as 1 as possible (Canny, 1986). In the case of the Gaussian filter  $r = 2/\sqrt{15}$ . In order to compare different filters, the value of  $\mu$  seems a more appropriate uniqueness response criterion than  $r$ , as it really gives the probability of multiple response at the edge location (what Canny actually meant) by taking into account not only  $r$  but also  $\mu_0$ .

*Uncertainty associated with the zero-crossings of the edge detectors*

Let  $J_0$  be the response of the filter to the step  $\Delta_0$ , and  $J_n$  the response of the filter to noise  $w$ . The maxima of  $J$  corresponds to the zero-crossing of  $J'$  which occurs at  $y_1$  which may be different from  $y_0$  because of the noise. Assuming that it is close enough to  $y_0$  for using the Taylor-Lagrange development of  $J'_0$  around  $y_0$ , then (Canny, 1986)

$$J'_0(y_1) \approx J'_0(y_0) + J''_0(y_0) (y_1 - y_0).$$

As by definition  $J'(y_1) = J'_n(y_1) + J'_0(y_1) = 0$  and  $J'_0(y_0) = 0$ , then

$$(y_0 - y_1) \approx J'_n(y_1)/J''_0(y_0)$$

Therefore, substituting  $J''_0(y_0)$  by  $\Delta_0 f''(0)$ , where  $f$  is the filter considered, the uncertainty of the edge location is

$$\sqrt{E((y_0 - y_1)^2)} \approx \frac{1}{\Delta_0} \sqrt{\frac{R''_y(0)}{f''(0)}}. \quad (\text{A7.14})$$

It is not possible to use the Taylor-Lagrange development for the Shen edge detector (1985), because of the discontinuity of the filter at the origin. In the following only the 1D case is considered for simplicity. The zero-crossing equation is

$$\Delta_0 f(\xi) + J'_n(y_0 + \xi) = 0$$

The function  $\Delta_0 f(\xi) + J'_n(y_0 + \xi)$  changes its sign at 0 if

$$\Delta_0 f(0^-) \leq J'_n(y_0) \leq \Delta_0 f(0^+)$$

Because of the antisymmetry of  $f$ ,  $f(0^-) = -f(0^+)$ . The probability of this event is

$$p(y_1=y_0) = 2 \operatorname{erf}\left(\frac{\Delta_0 f(0^+)}{\sqrt{-R''(0)}}\right) \quad (\text{A7.15})$$

Application to the antisymmetric exponential filter  $f(x) = \pm c \exp(-\alpha|x|)$ . The derivative is

$$f'(x) = -\alpha c \exp(-\alpha|x|) + 2c \delta(0)$$

where  $\delta$  is the Dirac measure at zero.

$$p(y_1=y_0) = 2 \operatorname{erf}\left(\frac{\Delta_0}{2\sigma_n}\right)$$

which is different from 1. So the localisation is different from  $\infty$ , unlike Shen's claim. In the 2D case  $\Sigma_0 = \Delta_0/\sigma_n$  should be replaced by  $2\Sigma_0/\sqrt{\alpha}$ . Unlike the Canny edge detector, the better the signal to noise ratio (i.e. the smaller  $\alpha$ ), the better the localisation.

*Application of the zero-crossing problem to the expected number of false alarms in the accumulator space.*

The notations are the same as the notations of chapter 5, subsection 5.2.4.

Using the model of the lines described in chapter 4, the noise of the

accumulator space at the point  $P(x', y')$  may be described as a random process  $w_{x'}(y')$ , obtained as the number of lines crossing the vertical  $y'$  of  $P$  within the neighbourhood  $[y' - \bar{\sigma}'_p, y' + \bar{\sigma}'_p]$

$$w_{x'}(y') = \sum_{|i| < \bar{\sigma}'_p} n_{x'}(y' - i),$$

where  $w_{x'}(y')$  is the number of lines crossing the cell  $(x', y')$  of the accumulator space.  $n_{x'}(y')$  is a random process, the covariance of which is zero when  $\xi \neq 0$ , and equal to  $\sigma_0^2$  when  $\xi = 0$ , which is independent of  $y'$ .  $x'$  being fixed, for simplicity let  $s(y')$  be  $w_{x'}(y') - f(x')$ ,  $f(x')$  being the expected value of  $w_{x'}(y')$  at  $P$ . The correlation function of  $s(y')$  is

$$R(\xi) = E(s(y')s(y'+\xi)) = \sigma_0^2 \sup(2\bar{\sigma}'_p - |\xi|, 0) = \frac{R(0)}{2\bar{\sigma}'_p} \sup(2\bar{\sigma}'_p - |\xi|, 0)$$

The expected number of maxima of  $s(y')$  with a value above  $k(x')$  is looked for. As  $R(\xi)$  is not derivable at 0, it is not possible to apply the equation A7.3 to  $s'(y')$ . However, as  $k(x')$  is assumed large enough, it is possible to ignore the probability of having two local maxima for  $s(y')$  between two successive crossings of  $s$  by the level  $k(x')$ . Therefore it is possible to approximate the expected number of maxima above  $k(x')$  by half the number of expected crossings by the level  $k(x')$ . The formulae A7.3a may be applied to  $s(y')$ . Substituting  $R(0^+)$  by  $-R(0)/(2\bar{\sigma}'_p)$ ,  $R(0)$  by  $\omega(x')$  and "a" by  $k(x') - f(x')$ , the expected number of local maxima above  $k(x')$  is

$$n_{\tau}(x') = \frac{y'_{\max}}{2\pi} \sqrt{\frac{\varepsilon}{\bar{\sigma}'_p}} \exp\left(-\frac{(k(x') - f(x'))^2}{2\omega(x')}\right), \quad (A7.16)$$

where  $\varepsilon=1$  and  $\omega(x')$  is given by eq. 5.2.4.3. Using the definition of  $\tau$  given eq. 5.2.4.3,  $(k(x') - f(x'))^2/\omega(x')$  may be replaced by  $(\text{erf}^{-1}(0.5 - \tau))^2$ .