Machine Vision
with its Application to
Autonomous Vehicle Navigation

by

Shauna Hui Wang

A Thesis

presented to the University of Manitoba

in partial fulfillment of the

requirements for the degree of

Master of Science

in

Industrial Engineering

Winnipeg, Manitoba

ISBN 0-612-13550-0

Name Shauna Hui WONG

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

Industrial
SUBJECT TERM

0546
SUBJECT CODE

U·M·I

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS
Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION
General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523
Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS
Language
- General 0679
- Ancient 0289
- Linguistics 0290
- Modern 0291

Literature
- General 0401
- Classical 0294
- Comparative 0295
- Medieval 0297
- Modern 0298
- African 0316
- American 0591
- Asian 0305
- Canadian (English) 0352
- Canadian (French) 0355
- English 0593
- Germanic 0311
- Latin American 0312
- Middle Eastern 0315
- Romance 0313
- Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY
Philosophy 0422
Religion
- General 0318
- Biblical Studies 0321
- Clergy 0319
- History of 0320
- Philosophy of 0322

Theology 0469

SOCIAL SCIENCES
American Studies 0323
Anthropology
- Archaeology 0324
- Cultural 0326
- Physical 0327

Business Administration
- General 0310
- Accounting 0272
- Banking 0770
- Management 0454
- Marketing 0338

Canadian Studies 0385
Economics
- General 0501
- Agricultural 0503
- Commerce-Business 0505
- Finance 0508
- History 0509
- Labor 0510
- Theory 0511

Folklore 0358
Geography 0366
Gerontology 0351
History
- General 0578
- Ancient 0579
- Medieval 0581
- Modern 0582
- Black 0328
- African 0331
- Asia, Australia and Oceania 0332
- Canadian 0334
- European 0335
- Latin American 0336
- Middle Eastern 0333
- United States 0337

History of Science 0585
Law 0398
Political Science
- General 0615
- International Law and Relations 0616
- Public Administration 0617

Recreation 0814
Social Work 0452
Sociology
- General 0626
- Criminology and Penology 0627
- Demography 0938
- Ethnic and Racial Studies 0631
- Individual and Family Studies 0628
- Industrial and Labor Relations 0629
- Public and Social Welfare 0630
- Social Structure and Development 0700
- Theory and Methods 0344

Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES
Agriculture
- General 0473
- Agronomy 0285
- Animal Culture and Nutrition 0475
- Animal Pathology 0476
- Food Science and Technology 0359
- Forestry and Wildlife 0478
- Plant Culture 0479
- Plant Pathology 0480
- Plant Physiology 0817
- Range Management 0777
- Wood Technology 0746

Biology
- General 0306
- Anatomy 0287
- Biostatistics 0308
- Botany 0309
- Cell 0379
- Ecology 0329
- Entomology 0353
- Genetics 0369
- Limnology 0793
- Microbiology 0410
- Molecular 0307
- Neuroscience 0317
- Oceanography 0416
- Physiology 0433
- Radiation 0821
- Veterinary Science 0778
- Zoology 0472

Biophysics
- General 0786
- Medical 0760

EARTH SCIENCES
Biogeochemistry 0425
Geochemistry 0996
Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES
Environmental Sciences 0768
Health Sciences
- General 0566
- Audiology 0300
- Chemotherapy 0992
- Dentistry 0567
- Education 0350
- Hospital Management 0769
- Human Development 0758
- Immunology 0982
- Medicine and Surgery 0564
- Mental Health 0347
- Nursing 0569
- Nutrition 0570
- Obstetrics and Gynecology 0380
- Occupational Health and Therapy 0354
- Ophthalmology 0381
- Pathology 0571
- Pharmacology 0419
- Pharmacy 0572
- Physical Therapy 0382
- Public Health 0573
- Radiology 0574
- Recreation 0575
- Speech Pathology 0460
- Toxicology 0383

Home Economics 0386

PHYSICAL SCIENCES
Pure Sciences
Chemistry
- General 0485
- Agricultural 0749
- Analytical 0486
- Biochemistry 0487
- Inorganic 0488
- Nuclear 0738
- Organic 0490
- Pharmaceutical 0491
- Physical 0494
- Polymer 0495
- Radiation 0754

Mathematics 0405
Physics
- General 0605
- Acoustics 0986
- Astronomy and Astrophysics 0606
- Atmospheric Science 0608
- Atomic 0748
- Electronics and Electricity 0607
- Elementary Particles and High Energy 0798
- Fluid and Plasma 0759
- Molecular 0609
- Nuclear 0610
- Optics 0752
- Radiation 0756
- Solid State 0611

Statistics 0463

Applied Sciences
Applied Mechanics 0346
Computer Science 0984
Engineering
- General 0537
- Aerospace 0538
- Agricultural 0539
- Automotive 0540
- Biomedical 0541
- Chemical 0542
- Civil 0543
- Electronics and Electrical 0544
- Heat and Thermodynamics 0348
- Hydraulic 0545
- Industrial 0546
- Marine 0547
- Materials Science 0794
- Mechanical 0548
- Metallurgy 0743
- Mining 0551
- Nuclear 0552
- Packaging 0549
- Petroleum 0765
- Sanitary and Municipal 0554
- System Science 0790

Geotechnology 0428
Operations Research 0796
Plastics Technology 0795
Textile Technology 0994

PSYCHOLOGY
General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

Nom _____

Dissertation Abstracts International est organisé en catégories de sujets. Veuillez s.v.p. choisir le sujet qui décrit le mieux votre thèse et inscrivez le code numérique approprié dans l'espace réservé ci-dessous.

SUJET

CODE DE SUJET

U·M·I

Catégories par sujets

HUMANITÉS ET SCIENCES SOCIALES

COMMUNICATIONS ET LES ARTS

Architecture 0729
Beaux-arts 0357
Bibliothéconomie 0399
Cinéma 0900
Communication verbale 0459
Communications 0708
Danse 0378
Histoire de l'art 0377
Journalisme 0391
Musique 0413
Sciences de l'information 0723
Théâtre 0465

ÉDUCATION

Généralités 515
Administration 0514
Art 0273
Collèges communautaires 0275
Commerce 0688
Économie domestique 0278
Éducation permanente 0516
Éducation préscolaire 0518
Éducation sanitaire 0680
Enseignement agricole 0517
Enseignement bilingue et multiculturel 0282
Enseignement industriel 0521
Enseignement primaire. 0524
Enseignement professionnel 0747
Enseignement religieux 0527
Enseignement secondaire 0533
Enseignement spécial 0529
Enseignement supérieur 0745
Évaluation 0288
Finances 0277
Formation des enseignants 0530
Histoire de l'éducation 0520
Langues et littérature 0279
Lecture 0535
Mathématiques 0280
Musique 0522
Orientation et consultation 0519
Philosophie de l'éducation 0998
Physique 0523
Programmes d'études et enseignement 0727
Psychologie 0525
Sciences 0714
Sciences sociales 0534
Sociologie de l'éducation 0340
Technologie 0710

LANGUE, LITTÉRATURE ET LINGUISTIQUE

Langues
- Généralités 0679
- Anciennes 0289
- Linguistique 0290
- Modernes 0291

Littérature
- Généralités 0401
- Anciennes 0294
- Comparée 0295
- Médiévale 0297
- Moderne 0298
- Africaine 0316
- Américaine 0591
- Anglaise 0593
- Asiatique 0305
- Canadienne (Anglaise) 0352
- Canadienne (Française) 0355
- Germanique 0311
- Latino-américaine 0312
- Moyen-orientale 0315
- Romane 0313
- Slave et est-européenne 0314

PHILOSOPHIE, RELIGION ET THÉOLOGIE

Philosophie 0422
Religion
- Généralités 0318
- Clergé 0319
- Études bibliques 0321
- Histoire des religions 0320
- Philosophie de la religion 0322

Théologie 0469

SCIENCES SOCIALES

Anthropologie
- Archéologie 0324
- Culturelle 0326
- Physique 0327

Droit 0398
Économie
- Généralités 0501
- Commerce-Affaires 0505
- Économie agricole 0503
- Économie du travail 0510
- Finances 0508
- Histoire 0509
- Théorie 0511

Études américaines 0323
Études canadiennes 0385
Études féministes 0453
Folklore 0358
Géographie 0366
Gérontologie 0351
Gestion des affaires
- Généralités 0310
- Administration 0454
- Banques 0770
- Comptabilité 0272
- Marketing 0338

Histoire
- Histoire générale 0578
- Ancienne 0579
- Médiévale 0581
- Moderne 0582
- Histoire des noirs 0328
- Africaine 0331
- Canadienne 0334
- États-Unis 0337
- Européenne 0335
- Moyen-orientale 0333
- Latino-américaine 0336
- Asie, Australie et Océanie 0332

Histoire des sciences 0585
Loisirs 0814
Planification urbaine et régionale 0999
Science politique
- Généralités 0615
- Administration publique 0617
- Droit et relations internationales 0616

Sociologie
- Généralités 0626
- Aide et bien-àtre social 0630
- Criminologie et établissements pénitentiaires 0627
- Démographie 0938
- Études de l' individu et de la famille 0628
- Études des relations interethniques et des relations raciales 0631
- Structure et développement social 0700
- Théorie et méthodes. 0344
- Travail et relations industrielles 0629

Transports 0709
Travail social 0452

SCIENCES ET INGÉNIERIE

SCIENCES BIOLOGIQUES

Agriculture
- Généralités 0473
- Agronomie. 0285
- Alimentation et technologie alimentaire 0359
- Culture 0479
- Élevage et alimentation 0475
- Exploitation des pâturages 0777
- Pathologie animale 0476
- Pathologie végétale 0480
- Physiologie végétale 0817
- Sylviculture et faune 0478
- Technologie du bois 0746

Biologie
- Généralités 0306
- Anatomie 0287
- Biologie (Statistiques) 0308
- Biologie moléculaire 0307
- Botanique 0309
- Cellule 0379
- Écologie 0329
- Entomologie 0353
- Génétique 0369
- Limnologie 0793
- Microbiologie 0410
- Neurologie 0317
- Océanographie 0416
- Physiologie 0433
- Radiation 0821
- Science vétérinaire 0778
- Zoologie 0472

Biophysique
- Généralités 0786
- Medicale 0760

SCIENCES DE LA TERRE

Biogéochimie 0425
Géochimie 0996
Géodésie 0370
Géographie physique 0368
Géologie 0372
Géophysique 0373
Hydrologie 0388
Minéralogie 0411
Océanographie physique 0415
Paléobotanique 0345
Paléoécologie 0426
Paléontologie 0418
Paléozoologie 0985
Palynologie 0427

SCIENCES DE LA SANTÉ ET DE L'ENVIRONNEMENT

Économie domestique 0386
Sciences de l'environnement 0768
Sciences de la santé
- Généralités 0566
- Administration des hipitaux 0769
- Alimentation et nutrition 0570
- Audiologie 0300
- Chimiothérapie 0992
- Dentisterie 0567
- Développement humain 0758
- Enseignement 0350
- Immunologie 0982
- Loisirs 0575
- Médecine du travail et thérapie 0354
- Médecine et chirurgie 0564
- Obstétrique et gynécologie 0380
- Ophtalmologie 0381
- Orthophonie 0460
- Pathologie 0571
- Pharmacie 0572
- Pharmacologie 0419
- Physiothérapie 0382
- Radiologie 0574
- Santé mentale 0347
- Santé publique 0573
- Soins infirmiers 0569
- Toxicologie 0383

SCIENCES PHYSIQUES

Sciences Pures

Chimie
- Généralités 0485
- Biochimie 487
- Chimie agricole 0749
- Chimie analytique 0486
- Chimie minérale 0488
- Chimie nucléaire 0738
- Chimie organique 0490
- Chimie pharmaceutique 0491
- Physique 0494
- PolymÇres 0495
- Radiation 0754

Mathématiques 0405
Physique
- Généralités 0605
- Acoustique 0986
- Astronomie et astrophysique 0606
- Electronique et électricité 0607
- Fluides et plasma 0759
- Météorologie 0608
- Optique 0752
- Particules (Physique nucléaire) 0798
- Physique atomique 0748
- Physique de l'état solide 0611
- Physique moléculaire 0609
- Physique nucléaire 0610
- Radiation 0756

Statistiques 0463

Sciences Appliqués Et Technologie

Informatique 0984
Ingénierie
- Généralités 0537
- Agricole 0539
- Automobile 0540
- Biomédicale 0541
- Chaleur et ther modynamique 0348
- Conditionnement (Emballage) 0549
- Génie aérospatial 0538
- Génie chimique 0542
- Génie civil 0543
- Génie électronique et électrique 0544
- Génie industriel 0546
- Génie mécanique 0548
- Génie nucléaire 0552
- Ingénierie des systämes 0790
- Mécanique navale 0547
- Métallurgie 0743
- Science des matériaux 0794
- Technique du pétrole 0765
- Technique minière 0551
- Techniques sanitaires et municipales 0554
- Technologie hydraulique 0545

Mécanique appliquée 0346
Géotechnologie 0428
Matières plastiques (Technologie) 0795
Recherche opérationnelle 0796
Textiles et tissus (Technologie) 0794

PSYCHOLOGIE

Généralités 0621
Personnalité 0625
Psychobiologie 0349
Psychologie clinique 0622
Psychologie du comportement 0384
Psychologie du développement 0620
Psychologie expérimentale 0623
Psychologie industrielle 0624
Psychologie physiologique 0989
Psychologie sociale 0451
Psychométrie 0632

MACHINE VISION WITH ITS APPLICATION TO AUTONOMOUS VEHICLE NAVIGATION

BY

SHAUNA HUI WANG

A Thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

I hereby declare that I am the sole author of this thesis.

I authorize the University of Manitoba to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Signature

I further authorize the University of Manitoba to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Signature

Abstract

As manufacturing industries move to more flexible manufacturing systems, the demand for a more flexible automatic system increases dramatically. To deal with the particular navigation problems associated with automatic guided vehicles within such a system, a machine vision system has been developed. This vision system, which updates current vehicle position, is based on single camera, a priori knowledge about targets to be viewed, and a map of the related environment. A simple two-dimensional target has been designed to reduce the required complexity of the system for position updating. An algorithm has been developed to complete target recognition and its geometry calculation, and yield the camera position (vehicle) data. This newly developed vision system was then tested in a real industrial environment and the experimental results show that the system is reliable and robust. The accuracy on recovered position data is satisfactory for our purpose (less than five percent error on computing position data). The computing time of this vision system is less than a quarter of a second. Therefore this vision system is deemed adequate for automatic guided vehicles.

Acknowledgments

I would like to express my deep gratitude and sincere appreciation to my advisor, Prof. D. Strong, for his excellent guidance, support, his patience and continuous encouragement throughout this research work and thesis writing.

Special thanks to Craig Muller for providing help when I encountered seemingly endless bugs in programs and other types of technical nuisances. In particular, his collaboration in target design and code optimization as discussed in appendix B and noted in chapter 3.5.

I also wish thank Hassen Zghal for the use of his utility program, Tobie Guerin for editing the english composition, and Al Lohse for taking pictures.

To the students of the Vision Lab, I would like to thank you for providing a helpful working environment.

Finally, but not least, I am much obliged to my husband, Dennis Liu, for his constant support, understanding, encouragement and love. Words do not aptly describe my gratitude.

Table of Contents

List of Figures

CHAPTER 1

Introduction

1.1 Background

Human has the abilities to interact with his environment and to manipulate it in a coherent and stable manner. This interaction is accomplished by an on-going intelligent interplay between perception and motion-control (action). Vision, or visual perception, is one of the fundamentally important processes in this interaction, and provides us with a remarkable amount of information about our surrounding and enables us to interact intelligently with our environment.

In the manufacturing industry, tasks requiring the use of eyesight have in the past always been carried out by human operators, and this has limited the application of automation mainly to situations involving repetitive mechanical tasks. To solve this problem, efforts have been put forward to develop machine vision systems to give the machine an ability to "see", hence to increase the automation level in the manufacturing industry. Machine vision, in certain circumstances, can be better than human eyesight. It will continue to "see" round the clock and will not tire itself out, lose concentration, or have "Friday afternoon" syndrome. It can work in conditions which would be very unpleasant or impossible for a human operator. This is the reason why research and development in the area of machine vision began almost since the time that digital computers first became generally available.

The biological vision system in human beings needs to process a large amount of data in order to derive useful information, and the same requirement is true for the machine vision system. However, machine vision is not backed by the powerful human brain, and the processing task of machine vision is performed by computers which are considered to be the machine's "brain", and usually requires much computation. At the early stage of machine vision due to computation limitations, its application were found only in industries where the visual environment could be controlled, the task of the machine vision could be well defined, and real time response from the vision system was not required. A typical example is the inspection of printed circuited boards in the electronics industry where the preceding limitations are satisfied. The dramatic development in computer technologies in the past two decades reduces computational limitations and enables the system to compute more and more complicated algorithms in a very fast manner for information processing purposes to accomplish sophisticated tasks, resulting in rapidly improving research and development results in the area of machine vision.

Machine vision has been defined by different authors in their books and articles. One broad definition for machine vision is the use of devices for the sensing of X-ray, ultra-violet light, visible light, infrared light to receive and interpret an image of a scene to obtain information. Whatever the application, machine vision always includes the following three elements:

(1) Acquiring a picture image of the object or scene under examination.

(2) Carrying out various processes on the image in order to identify and make a quantitative assessment of the important features.

(3) Taking a decision or reporting on the basis of the information.

At the present time, there are many areas of application for machine vision in the manufacturing industry. Among them, Visual sensing and interpretation of environment for navigation are widely recognized as being an important ability to the control of robots, mobile vehicles. Only with the capability of the visual sensing and interpretation, these mobile machines can behave intelligently and flexibly. The requirements for machine vision in the area of navigation applications are different in many respects from the needs in other areas. However, the basic elements in the process are still the same as mentioned above.

1.2 The motivations of the research

As manufacturing industries move to more Flexible Manufacturing Systems (FMS), the demand for a more flexible automatic system increases dramatically. Automatic guided vehicle (AGV) for material handling is one part of such an integral system. Currently most of the AGV's are guided by a defined path. This is generally accomplished by using wires embedded in the floor or reflective paint on the floor surface. The vehicle treats the path as a close loop and travels along the path in one direction with mandatory stops at predetermined locations, and the sensing system on the vehicle tracks the path and directs the control system to steer the vehicle right or left. In this system the major disadvantage is the lack of flexibility. The drive path is fixed, thus workplaces and production lines of an FMS must be designed and structured to accommodate the drive path. Any modifications on the drive path can be very expensive and time consuming. It also requires the floor to be smooth and even,

in some cases clean and metal free in order to keep a continuous communication between vehicle and guiding system. This is the reason that most drive path guided AGV implementations are in environments where those path requirements can be met, such as the electronics industry.

To overcome the inherent problems associated with the conventional AGV, a few alternative methods have been put forward. All the methods incorporate more intelligence into the AGV system (either the vehicle itself or the environment in which it works). One method is to provide the vehicle with a sensor. The most commonly used sensor for a flexible path vehicles is a laser beam scanner, an active sensor mounted on the vehicle, rotating continuously to scan the work area for bar-coded targets. Information reflected back from a pair of bar codes provides the vehicle position information. These bar codes are placed at convenient intervals throughout the operating area and are accurately surveyed with the information held by the vehicle's on-board computer. This method requires that the bar coded targets be placed at the same level in the plant, which is very restrictive for some manufacturing environments. Furthermore, the light beam of the scanner will interfere with the working environment and workers, and it becomes worse when a number of vehicles are working in the same area. Another method is to provide the environment rather than the vehicles with the intelligence. In this arrangement the intelligence is concentrated in a central position (ceiling mounted camera). Thus, the environment has the means of detecting positions of vehicles by detecting markers on the vehicle as well as providing supervisory capabilities over the latter. Each vehicle is merely provided with the means of driving itself along the given path and the

ability to communicate with the central controller. The vehicle does not have any position sensing ability on board. Because the camera has a limited field of view and is therefore only able to recognize vehicles within that field, this method may be insufficient for many practical applications.

Using machine vision to solve the problem of AGVs is promising. The advantages of machine vision for navigation are mainly (1) greater flexibility, (2) less interference if passive vision is used, and (3) cost effectiveness. There is currently much research on machine vision for autonomous land vehicles, where focus is placed on the navigation of vehicles in world environments (outdoor). Research on machine vision for autonomous vehicles in the manufacturing environment has been around for a few years, but has received less attention than outdoor navigation [2]. Therefore, the objective of this research is to develop an algorithm combined with a machine vision system for the autonomous vehicle navigation in an industrial environment. In the context of this thesis, we will focus on developing a particular technique of machine vision for vehicle navigation and aim at achieving a practical solution.

1.3 Statement of problem

The navigation problem has been perceived as the limiting factor insofar as range and flexibility of autonomous vehicles are concerned. This thesis will explore machine vision as a frame work to solve the navigation problem of autonomous vehicles. We can not, at this stage, build a "universal" vision system. Instead, we are aiming at systems that perform a particular task in a confined environment as

mentioned in the previous section.

The role of our machine vision system as it relates to vehicle nagvigation is to provide position information (distance and orientation) of the camera (vehicle) in a known environment by perceiving objects of interest. The position statement is in the form of distance and orientation specifications since the vehicle movement is limited on a plane. The vision system is not required to provide position statement on a continuous basis. In our particular application, it is very rare that the vehicle loses its way completely. Rather it usually knows where it stood in the immediately preceding step of motion, and how much and in what direction it has moved in the present step. However, problems do arise because motion control systems such as dead reckoning contain errors.

A single camera machine vision system is used here. We decided to use this configuration for several reasons. Using one camera considerably reduces computational work, since only one image is required to be processed by the computer. However, for monocular machine vision, part of the dimensional data is lost due to the three dimensional scene projection onto a two dimensional image. The lost data has to be recovered by machine vision system.

We have chosen a passive target approach, since passive targets create minimum interference with the working environment as opposed to others such as light bulbs and are furthermore easy to maintain.

1.4 Research goals

The model system is based on a single camera, a priori knowledge about the

objects to be viewed, and a map of the related environment. The research goals of this project are (1) to find an efficient and practical method to recover camera position data from a two-dimensional image; and (2) to develop an algorithm combined with machine vision system hardware to recognize the predefined objects in the scene and then calculate the camera position data. This algorithm should be reliable and robust enough to deal with the target environments. The response time from the capture to position determination should be within one second.

1.5 Thesis outline

In the preceding introduction, we presented some of the background knowledge, motivations and objectives of the investigation.

Chapter 2 briefly reviews previous research in machine vision as it relates to range finding and navigation. Different methods, including mark tracking, landmark use, stereo vision and inverse projection are discussed.

Chapter 3 describes our machine vision system, how it functions, its constraints and requirements. Also discussed is our target design for the present vision system, as target shape is crucial to algorithm implementation and vision system speed.

Chapter 4 presents the algorithm required by our vision system. This algorithm implements measurement of the parameters necessary for camera location. Since navigation problems require the system to respond in real time, an efficient search strategy is incorporated into the algorithm scheme.

Chapter 5 discusses experimental test results. All these tests were done in real environments with various conditions to examine the robustness of the algorithm.

Chapter 6 presents conclusions and recommendations.

CHAPTER 2

Related Research

There has been a considerable amount of research work done to develop a machine vision system for navigation. A number of methods or techniques have been put forward in the past few years. In this chapter, we shall briefly review some of the passive techniques related to our current project. These are as follows: mark tracking (the machine vision system derives the position data of the vehicle from the designed marks or objects in the environment), landmark following (the system derives the information from the landmark such as road intersections), stereo vision (the system obtains the information by using depth data in the scene), and reverse projection methods (which try to recover the three-dimensional information or surface orientation from a two-dimensional image).

2.1 Mark tracking for navigation

Most research work on indoor navigation assumes that either the environment is known or it can be modeled. Under these assumptions, we need only to know the relative position of the camera (vehicle) with respect to a particular object in the environment. Some researchers have approached the problem using a designed mark. The purpose of the mark is to provide a reference position in the coordinate system in which the vehicle moves. By observing a single projection of the mark the vehicle is able to determine its position in the coordinate system. The shape of the mark itself is

designed to yield enough information when it is imaged. Generally, a mathematical relationship between the mark, its projection and the camera has to be established to derive the position and orientation data.

2.1.1 Fukui's approach

Fukui [12] has designed a diamond-shape planar marker which is a black square on a white background whose dimensions are known as shown in figure 2.1. The camera lens center (O) and the marker center (O') are assumed to be at the same height with the optical axis of the camera aiming at the center of the marker. After pre-processing the edges of the diamond-shape mark, the length of the projected diagonals are computed. The length of the vertical diagonal in the image plane is inversely proportional to the distance from the camera center to the mark center, the length of the horizontal diagonal, however is related to camera orientation angle as established by the optical axis and the normal to the plane of the mark. By relating

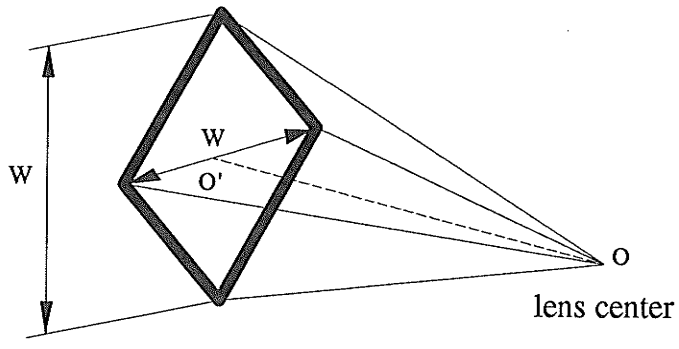


Figure 2.1: Fukui's approach to measure the distance and orientation angle.

the length of the vertical and horizontal diagonals in the image plane and the actual diagonals, the distance and orientation between the mark and camera can be determined. He also showed his experimental results conducted in the laboratory environment where the mark was attached on a white background.

There are two constraints in this approach: (1) the lens center must be kept level with the mark center and (2) the optical axis must point at the mark's center. These constraints greatly limit the flexibility of the system. From a practical point of view, this approach is too restrictive.

2.1.2 Courtney's approach

Courtney *et al.* [6] have extended Fukui's work by relaxing the constraint of the lens center leveled at the mark center. No attempt was made to enhance the

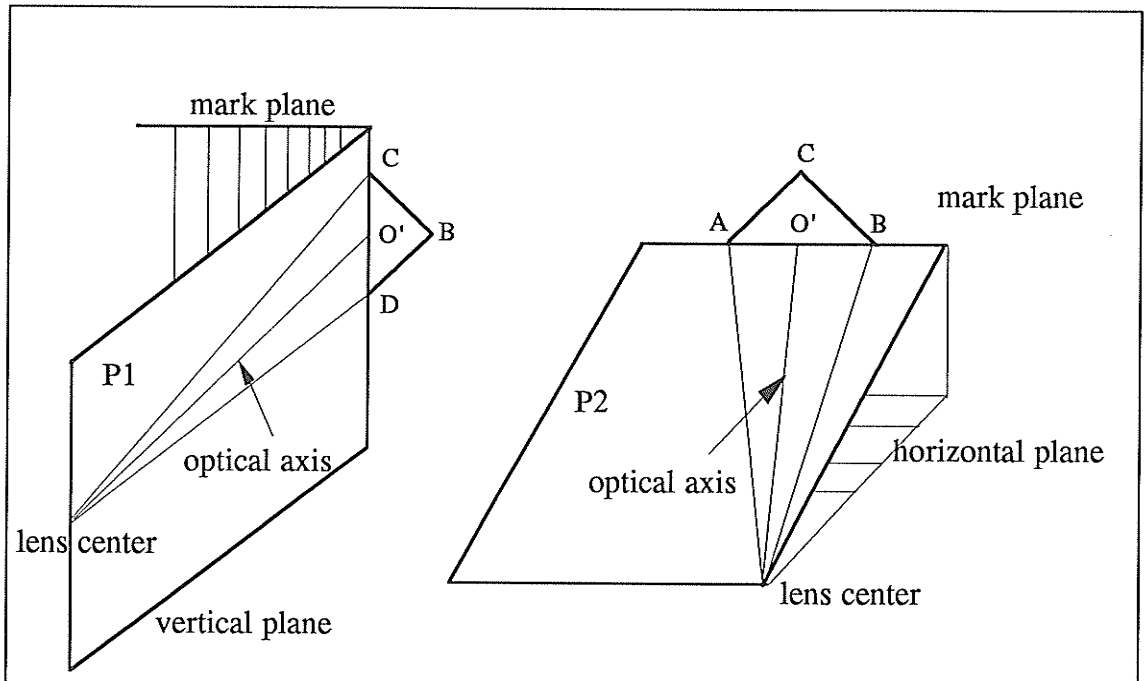


Figure 2.2: Diagram illustrating Courtney *et al.*'s approach.

processing stage of locating the edges of the mark. They have introduced two planes into the problem as shown in figure 2.2. The first plane, P_1, is a vertical plane passing through the camera lens center and points C and D, and the second plane, P2, is a plane passing through the camera lens center and points A and B. The intersection of these two planes is on the optical axis, which is the line joining the mark center and the lens center. By applying the same trigonometric relation in each plane, two equations with three unknowns are yielded. By assuming that the height of camera relative to the mark is known, the system of equations can be solved. This yields three equations with three unknowns. The solution involves a straightforward substitution.

2.1.3 Three-dimensional mark

A sphere has been used as a marker by Magee [17] which is marked with horizontal and vertical great circles for information about orientation purpose as shown in figure 2.3. When the sphere center is viewed along the optical axis, its projection is always a circle whose radius in the image plane can be used to determine the distance. Depending on the viewing angle, the projections of the two great circles in the image would be lines and/or arcs. By computing the closest distance between the sphere's projected center and points on the projected great circles, the distance and orientation of the camera can be found. Magee gave the experimental results conducted over six points.

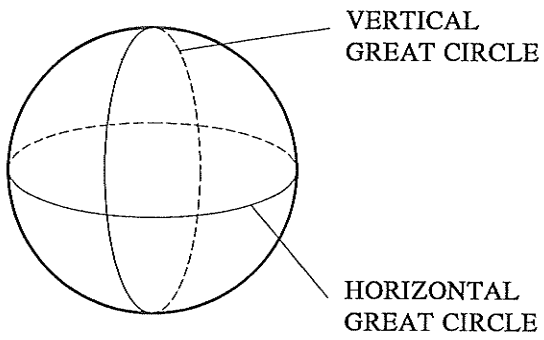


Figure 2.3: A sphere mark used by Magee.

The major drawback of the previously mentioned three mark tracking methods is that the camera optical axis has to be kept pointed at the center of the mark while the camera moves. This suggests that the vision system has to be able to accurately adjust the viewing angle automatically while it moves. These methods also simplify the mark recognition process by placing the mark in a special environment, which is hardly tolerable from a practical point of view.

2.2 Landmark for navigation

2.2.1 Road edge following

Waxman *et al.* [26, 27] have used a monocular image for road following navigation. The linear features such as road curbs are extracted first which are symbolic description of the image domain. They are then passed to a module that tries to establish them into significant groupings called pencils of lines. Each pencil is

given a three dimensional (3D) shape interpretation in the form of parallel lines on a planar surface patch. A reasoning process is used to decide whether the parallels and the corresponding surface patch "make sense" or not. Each successive pencil, interpreted as parallels, is checked for consistency with already accepted parallels, thereby building up a viewer centered 3D description of the road for navigation purposes.

Turk, *et al.* [22] have described a similar road-following technique. Rather than parameterizing the road in terms of centerline and direction, they describe the road boundaries as a sequence of points. The line-of-sight vectors for each of the boundary points are computed by using inverse projection transformation. The intersection of the line-of-sight vectors with the ground plane yields the points in the world between which the vehicle steers to stay on the road. In addition, the system needs to know the predicted position of the road in each image based on the position of the road in the previous image and the motion of the vehicle between the consecutive images.

Wallace *et al.* [25] have employed a color camera for outdoor navigation. The color information was used to segment the images as road region or non-road regions. The detailed map data was supplied to the system which consisted of a topological map and a geometrical map. Road edges were used to construct the road when the vehicle was moving. They used the same local planar surface and continuity of road assumption as Waxman *et al.* when doing the transformation from image points to world coordinates. In order to estimate the vehicle's current position, the system selected crossing lines of intersection of road such as intersections and corresponding lines in the map data, and calculated the transformation.

Indoor road marks can also be used for the navigation of the vehicle when they are available. Frohn and Seelen [9] have used single camera vision to track the road mark during movement of the vehicle in a well-structured environment. The road mark used was a high-contrast stripe on the floor, and the identifying process of vision system for road mark is trivial. The global position of the vehicle was determined by barcode marks integrated into the road mark. It was reported that the vehicle can run at the speed of 1m/s. This approach requires a well-structured environment which can not be satisfied in many manufacturing environments.

2.2.2 Landmark matching

Sugihara [21] has used the objects commonly found in indoor environments as the landmarks necessary to determine the camera positions. He selected the vertical edges of the objects and avoided sophisticated image processing. The method was based on three assumptions: (1) the map where the camera worked was given, (2) the camera's optical axis was parallel to the ground, and (3) the points where vertical edges existed were given. A possible location of the camera was given by a correct correspondence between the edges in the images and those given in the map through an exhaustive search. Krotkov [15] has extended this work by doing experiments and considering errors in landmark direction caused by image processing and image-contained noise. He also analysed the uncertainty of the computed data and determined bounds to the solution errors.

2.3 Stereo vision for navigation

One of best understood aspects of human depth perception is binocular or stereo

disparity, which can be used to interpret how far or how near an object is. Two identical and aligned cameras with parallel optical axes, similar to human eyes, are used to produce stereo pair images from which depth data is derived and then used to interpret the scene for finding spatial relationships.

2.3.1 Feature matching

Triebdl *et al.* [23] and Tsuji *et al.* [24] have used stereo vision to obtain the intrinsic features of an environment confined to buildings. Typical indoor intrinsic features of buildings are doors, ceilings and hallways. By processing and/or matching these features in the images, the vision system can yield information for the purpose of navigation. They have decomposed the navigation task into modeling and planning, and assumed that the environment consists of a flat floor that carries vertical lines, straight walls with hinged doors. This assumption requires that the vision system needs only to look at vertical edges to generate the model. Since all important vertical edges cross a horizontal plane, looking for edges at the horizon is sufficient. After stereo pair images are obtained, edge detection operators are applied to both images. Stereo matching is done by checking similarity and local consistency. The matching result contains uncertainty which is solved by adding potential walls and doors to the model according to the author's criteria. The vision system then builds the model about the local environment. Once the local model is available, the motion planner will generate a path free of obstacles according to the local model, thus accomplishing the navigation task.

2.3.2 Environment knowledge

Onoguchi *et al.* [18] have used stereo vision for creating a multi-information local map for the navigation of the vehicle. They did not tackle the problem as general stereo vision for navigation, but rather, they limited the stereo vision to recognize only circles, lines, ellipses and their combinations in order to obtain the relevant depth information for navigation. They decomposed the navigation task into two operations. The first stage is an off-line teaching operation, carried out by interactively teaching the vehicle key features for localization at check points where the vehicle needs accurate location data. The system is taken around by a human operator to the selected check points, and collects stereoscopic images. The machine vision system then processes those images and creates a multi-information local map with the locations of the objects of interest and processing procedures. Each object of interest has to be unique in the scene to derive correct correspondence. The second stage is the on-line operation. The vehicle moves autonomously according to the multi-information map. At the check point, the vehicle stops and the stereo vision system acquires stereoscopic images, searches for objects of interest according to the extraction procedure in the multi-information local map and calculates the vehicle's position.

One drawback is that this approach needs human interaction each time a new route is introduced. The operator has to determine the image processing parameters and select which objects should be used as targets. Another drawback is that the vehicle has to stop before taking any stereoscopic images, this greatly reducing the

speed of the vehicle. It is reported that the vehicle moves at 300 m/hour [18].

2.4 Inverse projection

Inverse projection tries to obtain 3D information from 2D image features by applying inverse perspective transformation on these image features back to space. This "inverse optics" process of vision (one to many projection), as Poggio [19] and others have pointed out, is an under-constrained problem that requires the introduction of generic constraints to arrive at a unique solution. Since the navigation problem involves the determination of the position and orientation of the vehicle (camera) with respect to objects of fixed positions, research focusing on the recovery of 3D information or surface orientation of commonly found man-made objects relative to the camera position where their two dimensional (2D) image were obtained may be useful for navigation. Generally, researchers assume that the environment is confined and the recognition of objects of interest is complete and object feature correspondence between image and space has been obtained.

There are two approaches in this research field: (1) "shape from" techniques which attempt to recover surface orientation information based on general object independent constraints, and (2) model-based matching techniques which attempt to recover surface orientation information by comparing image features with a known set of 3D object models.

2.4.1 "Shape from"

"Shape from" methods depend on the identification of a few simple general constraints and assumptions which are consistent with the nature of all possible

objects in order to recover a single "best" interpretation from among the many possible for a given image. In other words, it attempts to maximize some general, shape-based evaluation function over the space of a possible inverse projective transformation of a given image feature[3]. The solution is found by searching a possible attitude consistent with these projections.

Horaud *et al.* [14] have analytically solved the equations for finding the position and orientation of a camera with respect to a scene object from corresponding four-point projections. Since objects in space have six degrees of freedom, to determine the values of these six parameters, it is necessary to match at least thee image lines and three model lines. They derived an analytical solution by replacing the four points with three lines and by exploring the geometric constraints available with the perspective camera model. By assuming that the three lines are non-planar and share one of the four points, the problem then becomes that of solving a biquadratic polynomial equation with one unknown. However, the solution does not exclude the case when the four points are coplanar or forming a right vertex. As a matter of fact, the solution is particularly simple for these cases. They reported that the solution is more stable when the points are not coplanar, because it does not depend on the relative orientation of the image plane with respect to the scene plane.

Dhome *et al.* [7] have given a method to find the analytical solution to the problem of the determination of the 3D object attitude in space from a single perspective image. This method is based on the interpretation of a triplet of image lines representing the perspective projection of linear ridges of objects such as polyhedral or cube, which can be coplanar, intersecting or arbitrary in space (not

necessary having common point as oppose to the Horaud *et al.* method [14]), and on the search of a possible attitude consistent with these projections. The solutions are given by the roots of a polynomial of order eight in one unknown. By using some logical rules, the number of solutions can be reduced to a few.

2.4.2 Model-based methods

In some applications, such as low-level aerial image understanding for nagivation, there exists known models of the objects being viewed (i.e. an airport run way). Using knowledge obtained from known object model is simpler than "shape from". Model-based methods require two phases: (1) determining the correspondence between image features and model features and (2) recovering 3D information or surface orientation by determining the transformation based on an assumed imaging geometry from the 2D image features to the 3D model.

The problem of determining surface orientation of an image contour from a given planar model assuming a known correspondence between image and model features has been well studied. Fischler and Bolles [8] have solved camera location with respect to a set of control points by using a perspective image. The control points coordinates in space and in image and their correspondences are assumed known. They established the relations between the camera position and control points. The camera location is the solutions of system of three polynomial equations for the case of three control points. There are four possible solutions in general. When four control points lie in a common plane, a unique solution can be found. They used the "random sample consensus paradigm" to solve the equation which relies on an initial

correlation procedure to determine the correspondence between model and images points. Barnard [4] determines surface orientation by means of back projection of corner features. In his solutions, he considers all possibilities and selects the solutions that correspond to the correct corner angle magnitudes given by the object model.

Horaud [13] has developed an approach which consists of a model-based interpretation of a single perspective image. Lines intersecting in the image are assumed intersecting in space, and the values of angle of lines (non-parallel or three non-coplanar lines correspond to the edges of objects) are assumed known. He used the angle between two intersecting image lines and the junction formed by the intersection of three image lines to derive the spatial constraints. Then he proposed a constructive method to find the orientation solution while avoiding solving nonlinear equation. A geometric model is used for selecting possible solutions. There are no restrictions on the angles between model edges, but orthogonality of these image lines in space results in a simple solution. The location and orientation of the object is obtained based on the possible geometric model and interpretation of several lines.

Augusteijn and Dyer [3] assumed object models can be specified as either planar polygonal contours (defined as a list of line segments) or arbitrary planar point sets and have recovered the surface orientation under orthographic projection geometry. They have developed an iterative procedure that simultaneously determines the correct correspondence between model and image features and the correct surface orientation of the pattern relative to the image plane. They shown results of when object model consists of two non-parallel lines with known correspondence, and when object model is a rectangular with known ratio between its sides but no initial

corresponence. There are no restrictions on the types of patterns or contours as long as angles between these pattern are known.

The difficulty with these methods is that it is hard to get these linear features or point patterns which satisfy the assumed conditions. Another difficulty with these methods is that the correspondence problem may be hard to solve, especially when the image contains a large number of points because all the possible correspondences have to be tried in order to get a correct one.

CHAPTER 3

Machine Vision System

In the previous chapter, we discussed some of the navigation related issues in machine vision. In particular, we reviewed some of the approaches proposed by various other researchers. However, these methods proved to be not well-suited to our application, as we are looking for a fast, economical and reliable machine vision system and the previous approaches appeared unable to meet some of these demands to some extent. In the following section, we will explain the requirements our machine vision system has to meet as well as its internal workings.

3.1 Demands and constraints on the vision system

The function of the machine vision system for an autonomous vehicle is to determine where the vehicle is within its working environment based on the geometrical relationship between the camera and the objects of interest being viewed. Though any object with a fixed and known position in the environment could be used by the vision system as a reference point, using a specifically designed target as the object of interest will greatly simplify the task of target recognition and interpretation. Therefore, in our research we have concentrated on target design as well as on quick target recognition techniques.

Target design is constrained by a couple of factors concerning the planned application. A planar target of reasonable size and no fine details is preferred, since a

planar target takes less space and its installation is much easier than that of a three-dimensional target, and a target with no fine details can be clearly seen at a great distance by the machine vision system. Lastly target design should allow recognition under a wide variety of lighting conditions.

The faster the machine vision system is able to find the target and obtain the necessary information, the higher the speed at which the vehicle can travel, thereby increasing efficiency. This also means incidentally that position updating can be made more frequently. However, since in manufacturing environments there are dozens of objects which would cause the machine vision system to confuse targets with those objects, it is compulsory for the system to use uniqueness of target effectively and find the targets and reject background. The vision system also has to satisfy centain accuracy requirements. Lastly to keep costs down, it is proposed that system be made of general off-the-shelf components. This design aspect as well as previously mentioned ones will be addressed in later sections of this work.

3.2 System overview

Given the task and the constraints that the vision system is going to fulfill, we have chosen a system with monocular vision and a special designed target.

Figure 3.1 shows a block diagram of the system's configuration. The system consists of a target, a video camera, an image capture and storage board and a computer complete with a vision software. The target is black and white and is placed in the working environment to be picked up by the camera and used as a landmark for determining the camera's position and orientation. Upon receiving an image

acquisition command from the computer, the camera picks up the image within its field of view and sends the image signal to the capture and storage board. The Image capture and storage board functions as an interface between the camera and the computer. It stores the image signals from the camera in its memory as a two dimensional array for the computer to access. The computer driven by the vision software is the central part of this vision system. It executes the programmed vision task, processes the image data and finds the target to obtain the information of the camera location.

Image plane processing can be described in three steps. The first step is the target recognition. In a captured image there are some other objects and background in addition to the target, but the focus of the vision system is only on the latter. To extract the target from the image, a horizontal line scanning procedure has been developed that searches for the unique features of the target i.e. three vertical segments of equal width.

The measurement process follows target recognition. Given the location of the

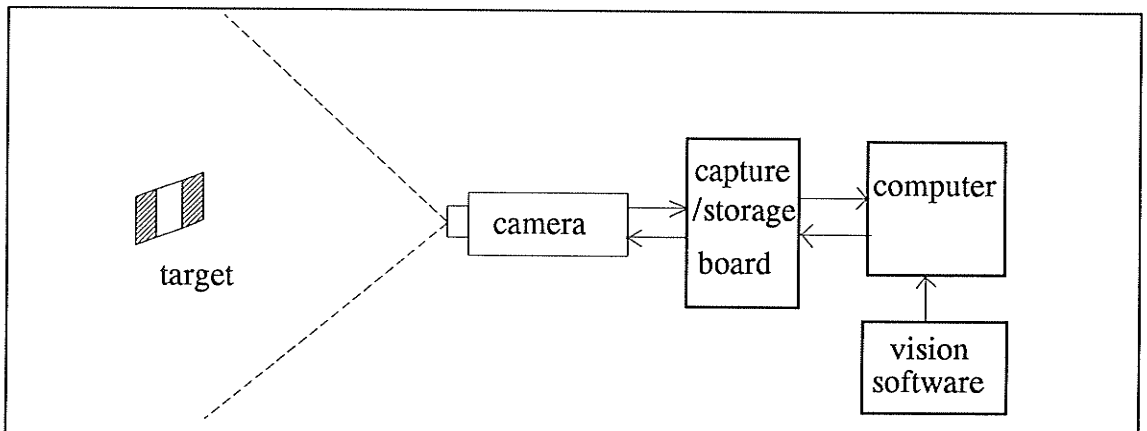


Figure 3.1: Block diagram of the system's configuration

group of pixels corresponding to the target, measurements are subsequently made to identify the target's shape characteristics which will be used by the position and orientation determination process. Those measurements are height and width of the target image.

In the final step, the position and orientation determination, a relationship between the target being viewed and the camera is established from the measurement of target image's shape and the actual geometry of the target. This relationship is based on a perspective projection of the three dimensional world including the target onto the two dimensional image plane.

The system achieves its desired goals. It is able to work with camera-to-target distance ranging from 1m to 10m with 16mm lens and 225mmx150mm target (or any single order of magnitude) in less than five percent error. Because it uses monocular vision and a uniquely designed target, the vision system is able to complete position data in a quarter of a second. It also has little effect on its working environment due to its passive approach and the types of target used.

3.3 System hardware

3.3.1 Camera

A general purpose charge coupled device (CCD) camera (Pulnix TM-450) is employed to transform the three-dimensional world scene into a two-dimensional digital image. Its imaging device is a pixel array of 8.8mm x 6.6mm with a 512x480 (horizontal x vertical) resolution. The camera lens mount is of a C type which can be coupled with commonly available focal length lenses. The image size together with

the focal length determines angles of coverage. For a 16mm lens, the angle of coverage of the camera is 30.8° x 23.3° (horizontal x vertical). Image device size, image resolution and focal length are part of the database held by an onboard computer. The aperture and focus of the camera are adjusted to an appropriate setting to provide the most suitable dynamic brightness range and an adequate depth of the field in the image. The camera records at a rate of 30 frames/sec as per the NTSC video standard used by the image capture and storage board.

3.3.2 Image capture/storage board and the computer

An image processing board (Matrox MVP-AT) interfaced with the computer is used to store image data obtained from the CCD camera. A frame buffer composed of 512x480 bytes is used to store one frame of video data in 256 grey levels. The frame buffer has several other storage and processing abilities that are not used in the research.

A general purpose 33MH 80486 computer with an 8 MB memory and a 320 MB hard disk is used. With this capacity, the image processing can be performed within the computer without any attached hardware.

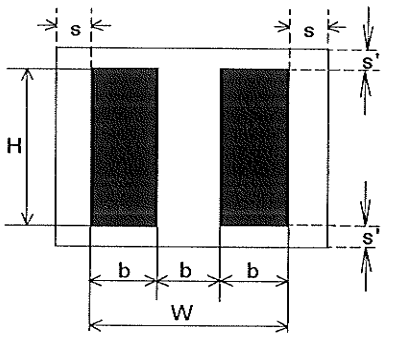
3.4 Target design and geometry

Geometry of the target is very important to the vision system, since it relates directly to the recognition process (algorithm) and thus affects the speed of the overall vision system. Target design constraints have been presented in earlier paragraphs. Evolution of the target have been documented in appendix B. Within the allowed category, we have proposed two targets as shown in figure 3.2 (a) and (b). These are

simple and unique. The simplicity of these targets facilitates the recognition process for the vision system and their uniqueness prevents them from being confused with other objects in a scene. Hence the target can be easily identified during the image processing phase.

The targets in figure 3.2 (a) and (b) are respectively called positive and negative targets. They have identical geometrical shapes but inverse black and white configurations. In the following description and discussion, we will focus on the positive target. This target consists of two identical vertical bars placed seperately with the same width on a 20cm x 28cm (8'' x 11'') plane. The middle area and the plane surrounding the bars are white, while the left and right bars are black and the horizontal grouping of the bars is located in the center of the target. The black bars are as mat and black as practically possible. Since the middle area is the same color as the plane, there is no clear mark defining its top and bottom.

The target geometry is as follows. The active target height H, the width b of the bars is 75mm. The active target width, which is 225mm equals the cumulative width



(a) (b)

Fig. 3.2: Positive (a) and negative (b) targets

of the three bars. This gives its dimensions being single multiplier of one dimension with $H=2b$, $s=b/2$, $s=s'$. This data is stored in the database as part of the a priori knowledge of the vision system. With reference to figure3.2 (a), the stripes "s" and "s' " beside the two vertical black bars and above and below the three bars do not constitute the active area of the target for image plane measurement purposes, they are significant in the recognition process. In the working environment, these two stripes prevent the target features from merging with the background, especially if the target background colors are similar to the black bars of the target. In our case, the stripe "s" width are about an half width of the vertical bar at 37mm.

The designed target has the following characteristics:

- Uniqueness, which is provided by the combination of the three bars. There is hardly any other object in the environment which possesses the same attributes.
- Simplicity. Simple geometry shapes, i.e. three identical bars are used to make up the target which makes identification easy.
- High contrast. The target black and white colors provide a maximum contrast and make the target recognition less subject to environment lighting variances. As a result, recognition can be more reliable.

3.5 Recognition and measurement

Recognition processes usually take several processing stages and can be quite complicated. However, when the object to be recognized is simple in shape and represented by simple features in the image domain, the recognition process can be

quite straight forward. The following assumptions are made in relation to the target and the camera orientation:

(1) the camera's optical axis and the horizontal axis of its image plane are kept parallel to the ground and the vehicle's motion is limited to planar one.

(2) the target plane is perpendicular with the ground, and the vertical bars are normal to the ground.

(3) the range difference between the closest point and the farthest point of the target plane as measured along the optical axis is much less than the range between the camera and the target center.

3.5.1 Thresholding

A digital image is a two-dimensional array. The elements of the array are called pixels, each of which has a grey-level values ranging from 0 to 255. Indices for each dimension define pixel position in the image domain. In our recognition process, only two grey levels i.e. "black" and "white" are useful. To obtain this statement at each pixel, a threshold is used to define the pixel as either "black" or "white" in the image.

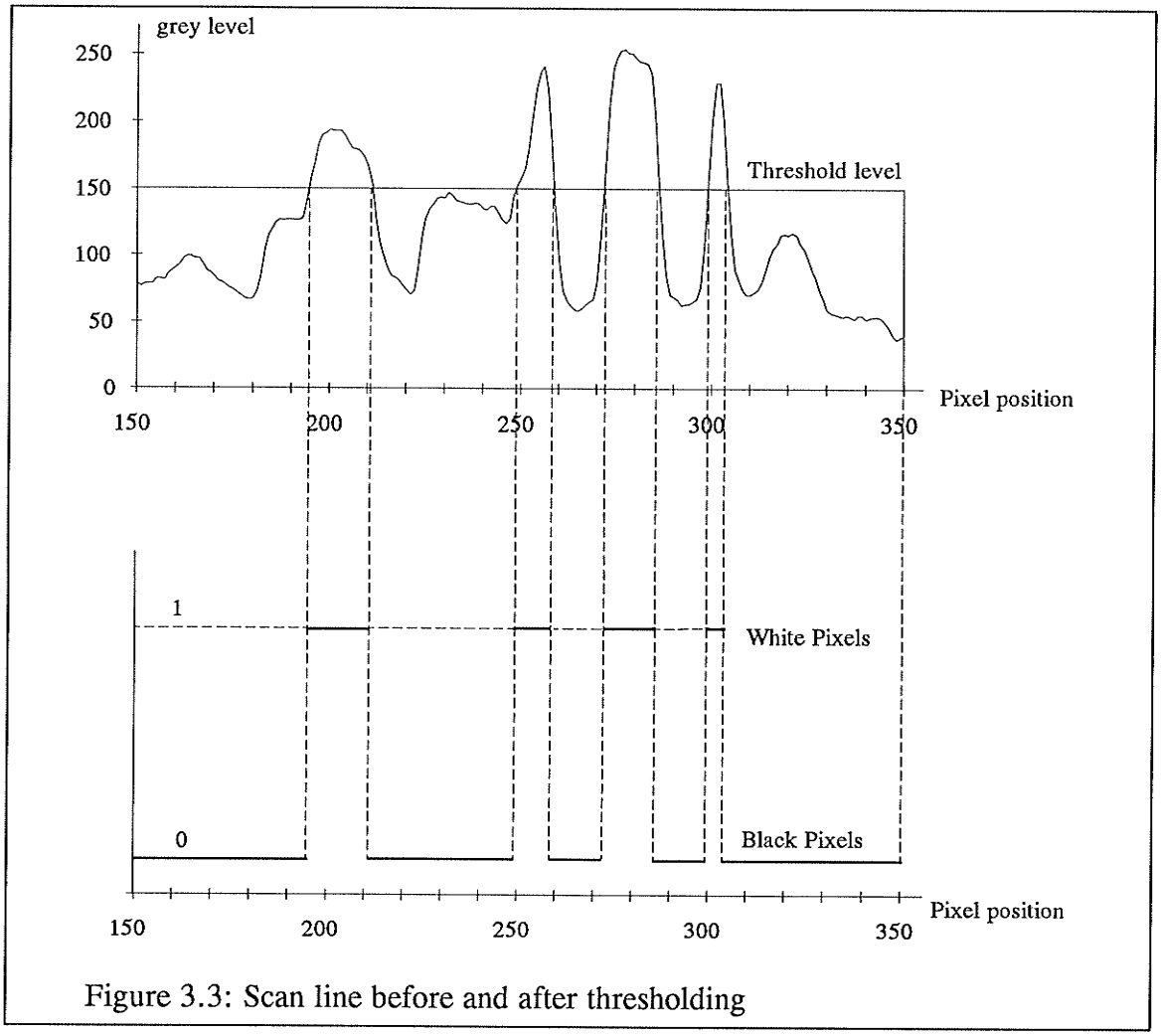
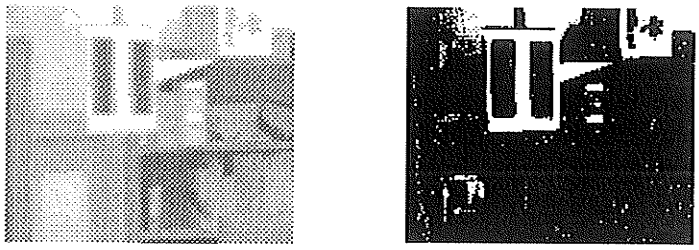


Figure 3.3: Scan line before and after thresholding

Figure 3.3 shows a grey level representation of the pixels along a section of horizontal scan line and the thresholded result. The threshold level used in this case is 150. If a pixel grey level is greater than or equal to the threshold level (or threshold for short), it is considered "white", otherwise it's "black". This way the grey level image data is compressed into binary data resulting in a large reduction in image data. The horizontal scan after thresholding consists of an alternating of the "black" pixel groups and the "white" pixel groups.

The thresholding operation keeps the high contrast features in the image while eliminating details of lower contrast such as texture and background. As for the



(a) (b)

Figure 3.4: Grey level mage (a) and its corresponding binary image (b)

target, features and geometric properties such as size and position are preserved in the thresholded data and can be used in the recognition process as representations of the target. Figure 3.4 (a) and (b) show a grey level image and its corresponding binary image respectively.

The threshold level determines the size of "black" and "white" groups within grey level image data. But it is difficult to select the best threshold suitable for the wide range of light conditions encountered in a manufacturing environment. If one threshold works fine for the targets in a fair illumination condition, it becomes probably too high or too low in an under or over-illuminated environment for thresholding to work properly. A more elegant technique based on the above method combined with spatial differentiation is described in chapter 4.4.

3.5.2 Target recognition

The target's features can be found in a thresholded scan data. Figure 3.5 shows an ideal case of the target projection after thresholding. Under the perspective projection, the rectangular target plane changes its shape when one side of the target is closer to the camera than the other. The three equally wide vertical bars now have

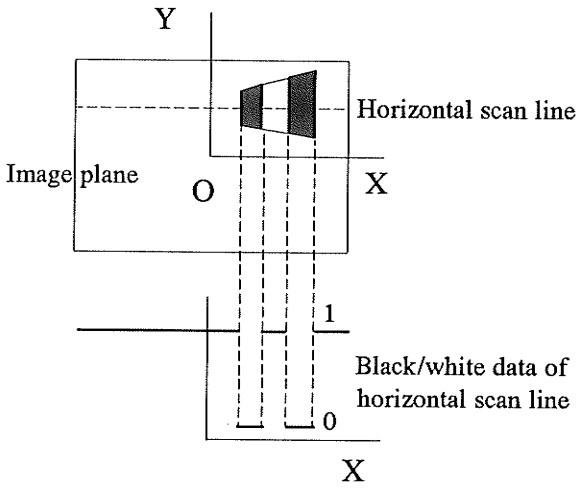


Figure 3.5: A target image under projection.

slightly different widths, while remaining vertical. The top and bottom edges of the bars closer to the camera are more separated than the bars further away, giving a slant to the upper and bottom edges of the target.

Considering the example of scan line in figure 3.5, if one examines the scan line's thresholded data, one finds a pattern consisting of three consecutive, equal (or approximately equal) segments of pixels alternately black and white. If this horizontal pattern is repeated at the same vertical position for a number of horizontal lines, we can say that the region formed by all these horizontal segments is a potential target. Because of inaccuracy factors such as noise, tolerances are allowed when measuring for equality of width. Since the pattern statement is simple, some objects in the background may also be registered as potential targets. Certain tests such as uniform grey level in each group, high contrast (grey level difference in black and white pixel) and continuity tests are carried out on these potential targets to eliminate false patterns and ensure a correct recognition.

3.5.3 Image plane measurement

The measurement process computes the geometrical properties of the recognized target based on measured parameters. The following measurements are of interest to the vision system regarding the target geometry in the image. The first of these parameters is the positions of the four vertical edges of the bars, from which the width of the target can be calculated, the next one is the heights of the two bars in its respective center from which the height of the target at its center can be obtained. Figure 3.6 illustrates a magnified target clipped from an image and the corresponding measurement parameters. It can be seen that the target edges are blurred as a result of sampling. Since the measurement depends on the target's outline or contour, well-defined edge points are crucial for accurate measurement results. However, since the target outline points resulting from the recognition process are based on thresholded data, inevitably many of these points turn out not to be edge point.

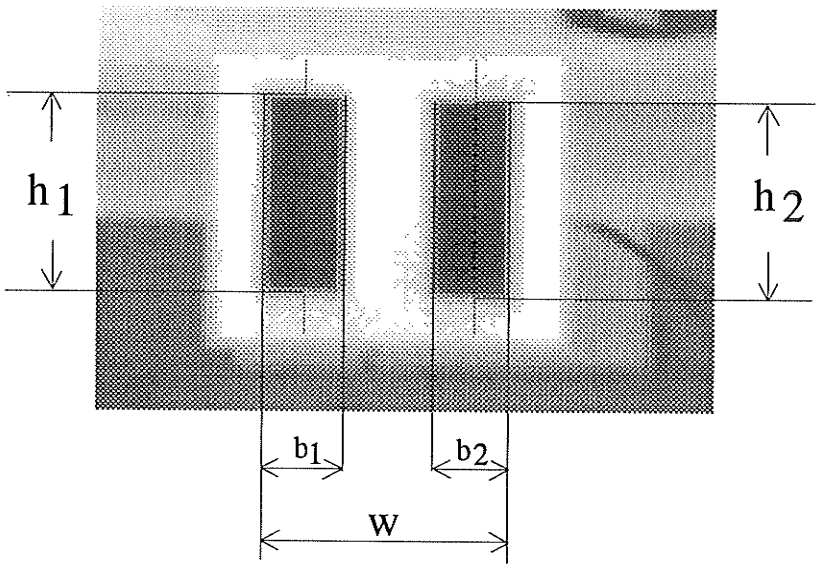


Figure 3.6: Image plane measurement

To get an accurate measurement, a two step procedure is under taken. In order to find the target's true vertical edge points, a vertical edge detector is used to spot edge points from the neighborhood of approximate edge point locations obtained from the recognized target. The obtained vertical edge points then are averaged to derive the target vertical edge positions in the image plane as shown in the figure 3.6. Since there is no clear mark in the target's center, the target's height at center furthermore is found by averaging the heights of the two black bars as measured at their respective centers. To get a measurement of h_1 and h_2, a horizontal edge detector is used around the top and the bottom of the bars along the dotted lines as shown in the figure. Horizontal edge points are found and the respective distances between top and bottom edge points, h_1 and h_2 are measured. Target width w in the image is the distance between the two outmost vertical edges.

3.6 Determination of position and orientation

3.6.1 Geometrical structure

The image formation process provides a relation between points in the three dimensional world and the image plane. A model of the image formation process is shown in figure 3.7. For the convenience of the description, we introduce a Cartesian coordinate system whose origin sits at the camera's lens center which is modeled as a pin hole. The Z-axis coincides with the optical axis and points towards the scene; the X-axis is chosen horizontally, pointing to the right away from the lens center; and the Y-axis would then be vertical, since the camera is not tilted. The image plane is parallel to the XY plane and its center is at the Z axis. The distance between these

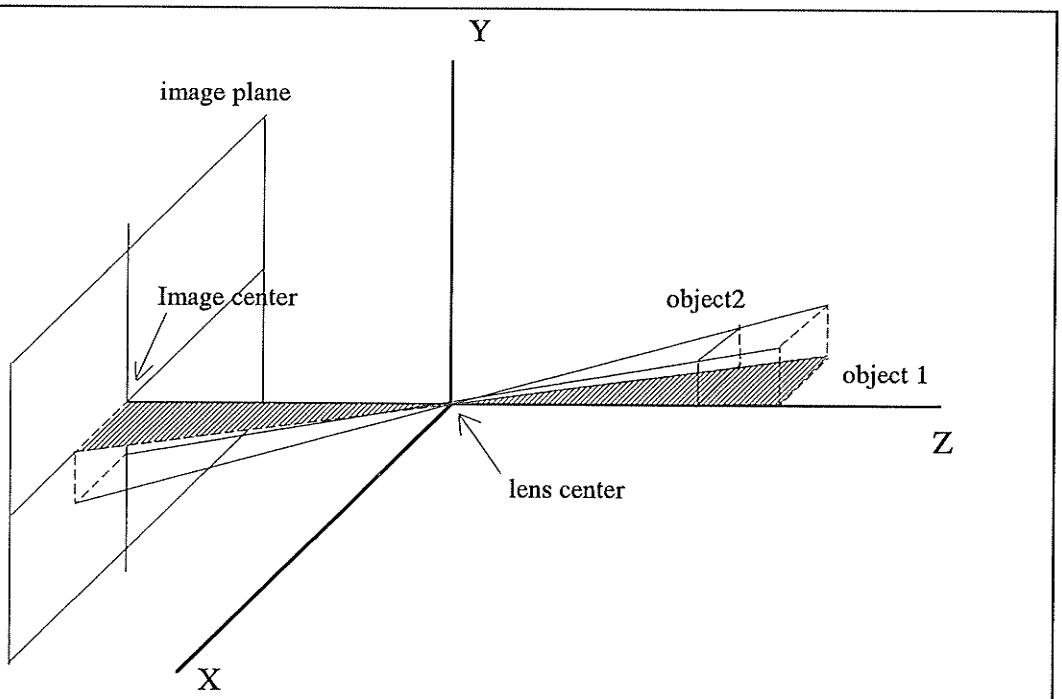


Figure 3.7: Perspective image formation model

two planes is the image distance.

Since light travels along straight lines, the image of any point on a object is found by the ray from this point through the lens center intersecting with the image plane. It can be seen from this model that projecting a three dimensional scene onto the image plane is a many to one transformation. For example, the objects 1 and 2 are different in size but of identical shape and are placed in such way that their respective object corners and lens center lie on one straight line. As a result the projected images coincide. If the size of the objects are known, the Z axis position of the objects can be found by using simple geometry.

3.6.2 Range recovering

The image obtained by perspective projection is actually inverted. For convenience of observation, the displaying device inverts the image so that it is

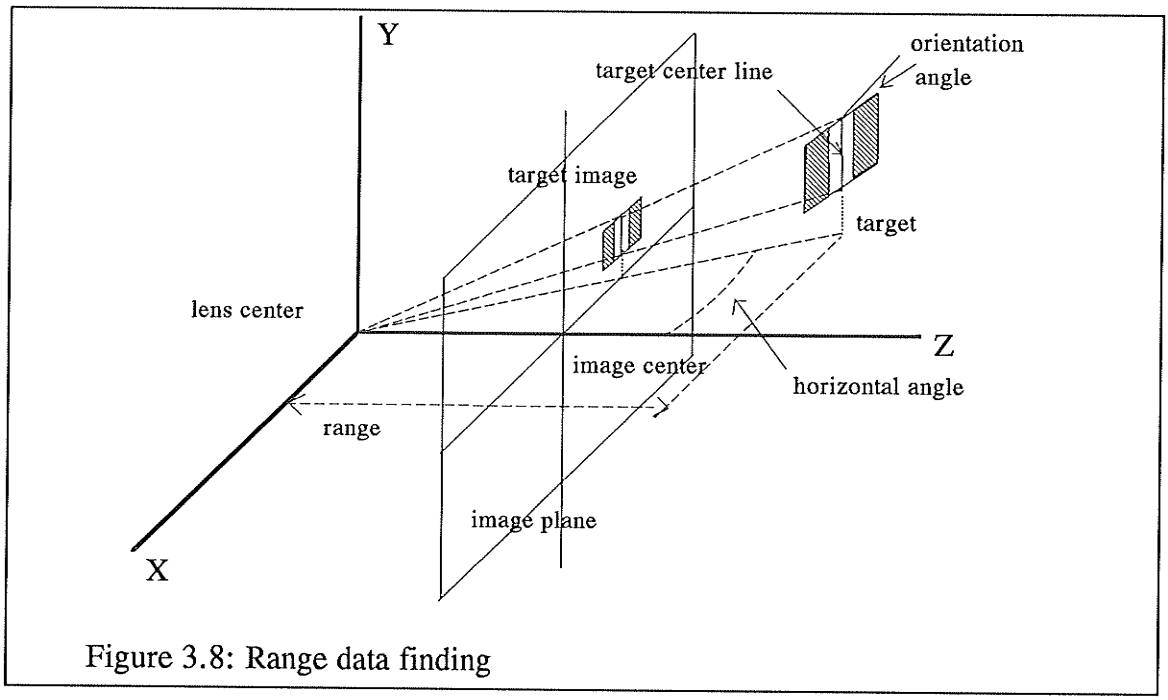


Figure 3.8: Range data finding

equivalent to placing the image plane in front of the lens center with an equal distance from the latter to the image plane. However, the relationships described earlier are still true except for the fact that the object and image are now located in the same quadrant. In the following sections, both configurations will be used as applicable.

Recovering the range data is a prerequisite for determining the location and orientation of the target. As in figure 3.8, when the target plane is perpendicular to the ground, the spatial relation between the camera and the target can be defined by one translation and one rotation. The translation can be determined by the range and the horizontal angle between the camera's optical axis and the target's center, while the rotation is determined by the orientation of the target relative to the image plane.

In the configuration of figure 3.8, one more similar triangle relation can be obtained by considering the ratio between the target's vertical center line height and its image height. This relation, however, can be solved to arrive at target center's

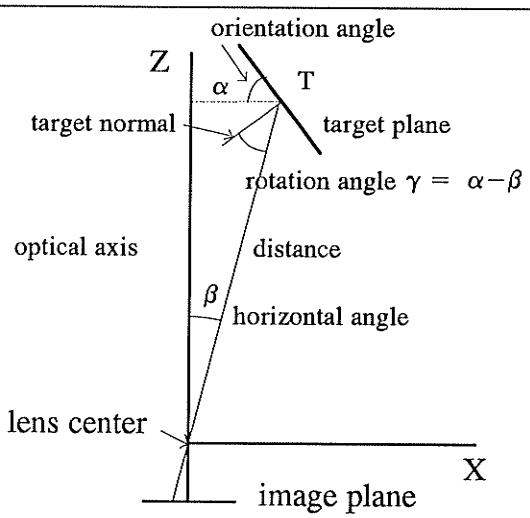


Figure 3.9: A top view of the camera and the target relationship

range. The tangent of the horizontal angle is determined by the image distance (focal length) and the horizontal coordinate of the target center in the image plane. The sign of the horizontal angle β indicates the position of the target's center relative to the optical axis. With this information, the location of the target's center on the XZ plane can be determined.

When referring to the camera's position with respect to the target, it is convenient to use polar coordinate parameters of distance and the rotation angle between the target normal and the line joining the lens center and the target's center as shown in figure 3.9. Since the distance can be obtained with the help of the range and the horizontal angle, the problem remaining is that of finding the rotation angle. Figure 3.9 shows that this angle can be obtained from the orientation angle. When these variables have been established, the camera's orientation and position can be precisely determined.

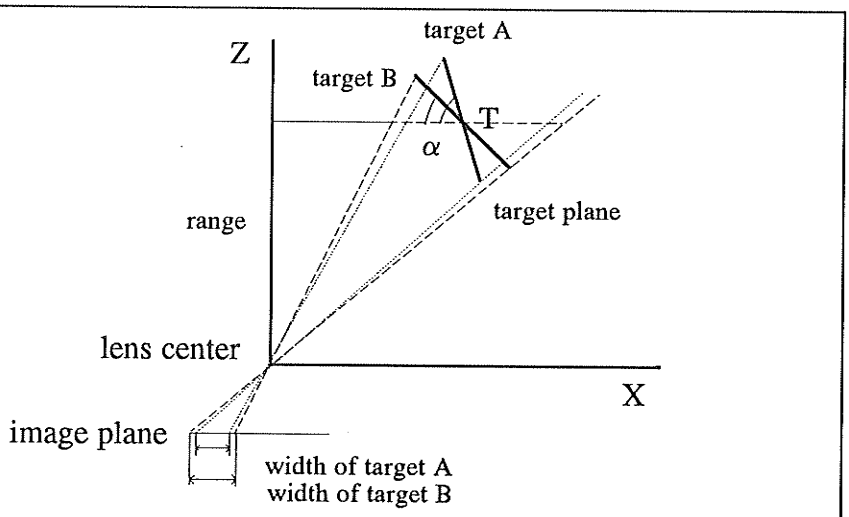


Figure 3.10: Orientation angle and the target width

3.6.3 Orientation angle recovering

The orientation angle determines the target plane's orientation with respect to the image plane and vice versa. In the previous section, we have shown how the ratio of the target's height to its image height used to determine range and horizontal angle values. Here it will be shown that the orientation angle can be found by considering the target height-width ratio in space and on the image plane with the aid of the obtained range data and of a priori knowledge of the target.

Figure 3.10 is a top view of two targets having a same center position but different orientation angles and images. It illustrates the relation between the target's orientation angle, center position and its image width. For a fixed target center position, one can see that the bigger the orientation angle is, the shorter the width of the target image is. The orientation angle being fixed, the farther the target is away from the camera, the shorter the target image width becomes, since target image width is inversely proportional to the range. The ratio of the target's height to its

width in space over the ratio of the height of the target's image to its width in the image plane relates to the target's orientation angle. From this information, however, the orientation angle can be determined.

CHAPTER 4

Geometry and Algorithms

In chapter three, we described the implemented machine vision system and its method of operation. In this chapter, we elaborate on the techniques used for target finding, accurate measurement of the target image and determination of the target position and orientation as described in section 3.5 and 3.6. We first establish the geometric relationship between the camera and the target being viewed, then develop an algorithm of the image domain processing for target recognition which leads to the determination of the camera position and orientation.

4.1 Imaging geometry

In the previous chapter, a camera centered coordinate system was established. Within this coordinate system, with reference to figure 4.1, space point *P(X,Y,Z)* and its projection *p'(x,y,f)* on the image plane are related by similar triangles. It follows that:

$$\frac{x}{X} = \frac{f}{Z} \tag{4.1}$$

$$\frac{y}{Y} = \frac{f}{Z} \tag{4.2}$$

where f is the distance from image plane to the lens center. For this discussion, the camera is assumed to focus at infinity, so image distance *f* will equal to the focal length of the camera. Note that no negative sign appears in these two equations since

an inverted image model is adopted. Equations (4-1) and (4-2) give a line that passes through the image point *(x,y,f)* and the origin *(0,0,0)*. To determine the *X*, *Y* and *Z* coordinates of the vertical target from the image, one first has to obtain the *Z* coordinate. Suppose that SP is the target's vertical centerline which equals the target height *H*, it has a projected image denoted by the *s'p'*, one more relationship can be obtained by considering the similar triangles *ΔOSP~ΔOs'p'* and *ΔORQ~ΔOr'o'*:

$$\frac{H}{s'p'} = \frac{Z}{f} \tag{4.3}$$

Z can be found from the known target height *H*, the focal length and target's vertical centerline projection *s'p'*.

The target's vertical center is offset from the optical axis by an angle β as in figure 4.1 and is determined by the formula:

$$\tan\beta = \frac{x}{f} \tag{4.4}$$

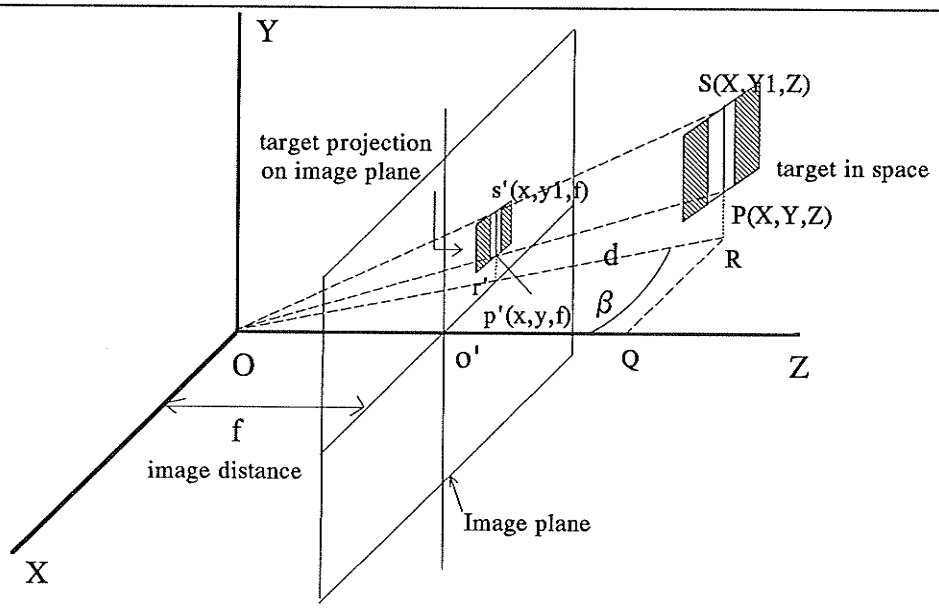


Figure 4.1: Imaging geometry.

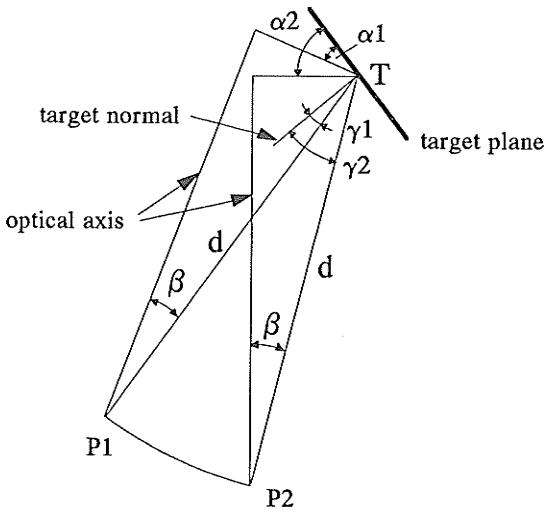


Figure 4.2: The target position and orientation relative to the camera

where x is the horizontal coordinate of $s'p'$ in the image plane. With range Z and angle β, the target's center on the XZ plane is determined.

It is convenient to use polar coordinates of distance and angle when referring to the camera's position relative to the target. The horizontal distance between the camera center and the target center can be obtained by:

$$d = Z\sqrt{1 + \tan^2 \beta} \qquad (4.5).$$

The camera's position with respect to the target's center can not be determined solely from d and β as can be seen in the figure 4.2, where $P1$ and $P2$ represent two different positions on a circle of radius d with center T and the same horizontal offset angle β. If rotation angle γ defined as the angle between the target normal and the line joining the camera and the target center is known, the position and orientation of the camera can be obtained. The rotation angle γ relates to orientation α which is defined as the angle between the target plane and image plane as in:

$$\gamma = \alpha - \beta \tag{4.6}.$$

4.2 Orientation determination

In the previous section it was mentioned that the angle between the two planes is used for the determination of orientation, but there was no mention of a relation leading to the technique for computating this angle. In this section, we are going to explore the geometry dealing with orientation. An equation for solving the angle is obtained by modifying the formula given by Courtney *et al* [6], which is based on the assumption that the positions of the target's two outermost vertical edges and the vertical centerline are known and that as a result their corresponding horizontal angles with the optical axis can be obtained (see appendix for details). The difficulty with this equation is that the geometrical center of the target in the perspective image can not be obtained easily, since there is no clear mark indicating the center of target in the image. To obtain the orientation angle, we estimated the target center as described in the following for the calculation of orientation. By introducing some approximations to the analytical calculation of orientation angle in the appendix, the solution shows an identical form as the one described in the following.

4.2.1 Geometry of orientation angle

Figure 4.3 shows a top view of the spatial relationship between the camera and the target being viewed. In order to obtain the orientation angle α, we have to make some approximations. Line AB represents the target plane with its length being equal to the width of the target W; line ab is the image of the AB in the image plane. T denotes the target's vertical center which divides the target plane into two equal parts.

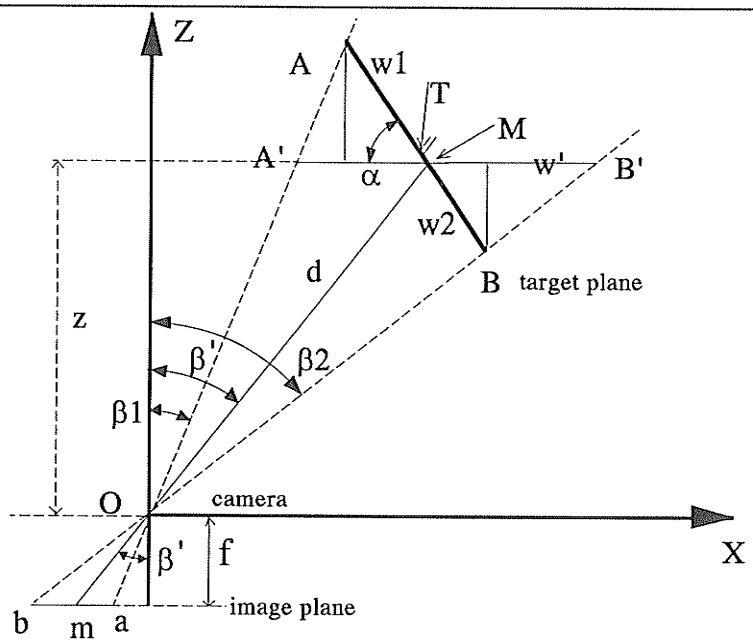


Figure 4.3: Top view of the relationship between the camera and the target.

The middle point named M is an approximation of the target vertical center. $M's$ image is the arithmetic center of ab and is denoted by m. M is found by extending the line mO to intersect with the target plane, which also divides the target plane into two parts but with different widths denoted by w_1 and w_2 as shown in figure 4.3. With reference to the figure, the following equations are obtained:

$$\begin{aligned} w' &= w_1 \cos\alpha + w_1 \sin\alpha \tan\beta_1 + w_2 \cos\alpha + w_2 \sin\alpha \tan\beta_2 \\ &= W\cos\alpha + \sin\alpha\,(w_1 \tan\beta_1 + w_2 \tan\beta_2) \end{aligned} \tag{4.7}$$

where w' denotes a plane perpendicular to the camera's optical axis, intersecting with point M, and bounded by the lines OA' and OB'. Angles $\beta 1$ and $\beta 2$ are the horizontal angles between the optical axis and lines OA and OB respectively. Substituting the values of $tan\beta_1$ and $tan\beta_2$ into the equation (4.7) and noting that the summation of w_1 and w_2 gives target width W, the equation (4.7) can be rewritten as:

$$w' = W\cos\alpha + \sin\alpha\,(w_1\frac{M_x - w'/2}{z} + w_2\frac{M_x + w'/2}{z})$$
$$= W\cos\alpha + \sin\alpha\,(W\frac{M_x}{z} + \frac{w'}{2}\frac{(w_2 - w_1)}{z}) \quad (4.8)$$

where M_x is x coordinate of the point M , z is its range.

Because of $(w_2\text{-}w_1) << z$ and $tan\beta' = M_x/z$, the second term inside the outer brackets in the above equation can be neglected, so that this equation can be further simplified as:

$$w' \approx W\cos\alpha + W\sin\alpha\ \tan\beta' \quad (4.9)$$

where β' is the horizontal angle between the optical axis and the line OM, which is also an approximate to the target center horizontal angle β.

Noting that $\Delta OA'B' \sim \Delta Oab$, we derive the following equation:

$$\frac{W'}{\Delta x} = \frac{z}{f} \quad (4.10)$$

where Δx is the width of the target image i.e. the line ab; f is the camera's focal length, and w' and z are as before. If we consider three dimensional geometry, another similar trianglar relation can be shown to exist between the target's vertical centerline and its image, as mentioned in the previous section, and expressed by:

$$\frac{H}{\Delta y} = \frac{z}{f} \quad (4.11)$$

where Δy is the height of the target image at middle point m and H corresponds to the height of the target at the point M, i.e., the target's height. Substituting equations (4.9) and (4.11) into equation (4.10), we obtain:

$$\frac{W(\cos\alpha + \sin\alpha \tan\beta')}{H} = \frac{\Delta x}{\Delta y}$$

$$\cos\alpha + \sin\alpha \tan\beta' = \frac{H/W}{\Delta y/\Delta x} \tag{4.12}$$

where H/W is the ratio of target height to target width, $\Delta y/\Delta x$ is the ratio of the target's height at the middle point to the target's width in the image plane. So α can be obtained by solving equation (4.12), since H/W is known in advance; $\Delta x/\Delta y$ can be measured from the image plane; and the angle β' can be calculated from the measurement of the middle point m and image plane distance f.

4.2.2 Solving for angle α

Solving the equation (4.12) for angle α, we have the solution

$$\alpha = \beta' \pm \cos^{-1}\left[\left(\frac{H}{W} * \frac{\Delta x}{\Delta y}\right)\cos\beta'\right] \tag{4.13}$$

When the target center is on the right side of the optical axis (Z axis), β' takes a positive value, otherwise it is negative. To determine $\alpha's$ sign from image measurement, we have to first define it in relation to image plane. With reference to figure 4.4, we define the angle α as positive if the target plane rotates counter clockwise, the amount of angle $\alpha(<90°)$ being established in relation with the image plane. If the rotation is clockwise it is negative as in figure 4.4(b). The $\alpha's$ sign can be determined from the measurement of image plane, i.e. from the relative height of the left and right bars of the target image. For example, if the left bar is shorter than the right bar, it means that the left bar is further away from the camera's center along the optical axis. So the target has two possible orientations as shown in the

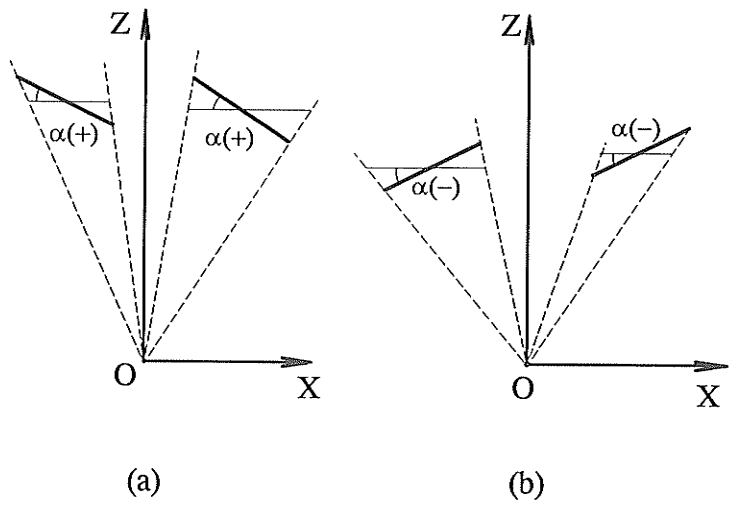


Figure 4.4: Possible orientations of the target

figure 4.4(a). By knowing β' sign, one can determine which is the proper one.

Within these sign definitions, the solution for orientation angle α (4.13) does not lose its generality. If we rewrite solution (4.13) as two separate ones, we get:

$$\alpha_1 = \beta' + \cos^{-1}\left[\left(\frac{H}{W} * \frac{\Delta x}{\Delta y}\right)\cos\beta'\right] \tag{4.14}$$

$$\alpha_2 = \beta' - \cos^{-1}\left[\left(\frac{H}{W} * \frac{\Delta x}{\Delta y}\right)\cos\beta'\right] \tag{4.15}$$

There are four different sign combination cases for α and β': namely when α and β' have opposite signs (two counts), and when α and φ have the same signs (two counts). We will discuss each category separately. When the α and β' are opposite in sign, i.e. α is positive and β' negative or vice versa, there is only one legitimate solution, i.e. when α is positive, α_1 that conforms with the signs of angles is the solution; α_2 in (4.15) is the legitimate solution in the case of α being negative. When the α and β' are both negative or positive, the things become a little bit more

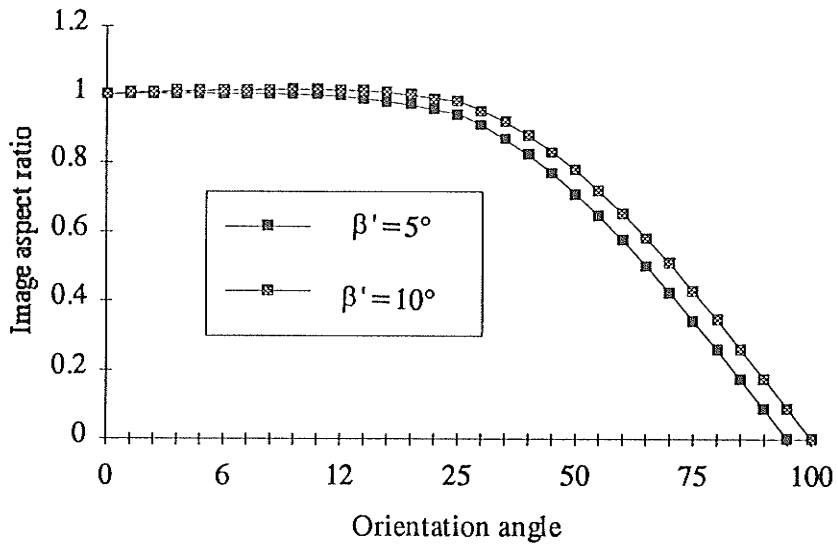


Figure 4.5: (H/W)(Δx/Δy) as a function of angle α for two values of β'.

complicated. We will only discuss the case where both α and β' are positive, the other case where the angles are both negative being similar.

When the second term in these two solutions yields an angle which is greater than angle β' in these equations, only α_1 in (4.14) yields a positive value which conforms with the premises on the angles' signs while α_2 in (4.15) becomes an illegimate solution. In the case when the second term yields an angle which is smaller than or equal to β', both α_1 and α_2 produce legitimate solutions for the orientation angle.

If we consider solutions in terms of range of α ($0<=\alpha<=90°$), we can say that for φ being less or equal to β' (φ represents the resulting angle of the second term), two valid solutions, $\alpha_1=\beta'+\varphi$ and $\alpha_2=\beta'-\varphi$, result in two orientation angles α in the interval $0 \le \alpha \le 2\beta'$,while for φ being greater than ', one valid solution $\alpha_1=\beta'+\varphi$ results in the orientation angle α in the interval $\alpha>2\beta'$. This causes uncertainty in a small interval when solving for orientation angle α.

When the target's center is fixed in a particular position, angle β' becomes a constant. So different orientations are reflected by different values for $(H/W)(\Delta x/\Delta y)$. Figure 4.5 illustrates $(H/W)(\Delta x/\Delta y)$ as a function of α for two different values of β' (5° and 10°). Some observations can be made from the figure: (1) for a different angle of β', the curves have a similar shape, (2) when α is small and between the interval of $0\sim2\beta'$, the value of $(H/W)(\Delta x/\Delta y)$ shows a lack of sensitity to the changes of orientation angle and $(H/W)(\Delta x/\Delta y)$ becomes a non-monotonic function of angle α, since two values of α gives same ratio of $(H/W)(\Delta x/\Delta y)$ and (3) it also can be deduced that a smaller β' can yield a larger monotonic range. In a practical situation, the maximum β' is confined by the camera's focal length. For a 16mm lens, this angle is about 15°. So in the worst case ($\beta'=15°$), the monotonic interval for $(H/W)(\Delta x/\Delta y)$ range from 15° to 90°. To get a greater monotonic interval, one has to reduce angle β'.

4.3 Flowchart of the algorithm

The algorithm is to process the image and obtain the parameters necessity to determine the position and orientation of the target. Its processing procedures can be described by the block diagram as shown in figure 4.6. It consists of several major parts (1) searching for the target features in the image, (2) finding potential targets, (3) verifying the potential targets against tests, (4) edge detection and measurement, (5) based on the measurement and a priori knowledge determining the target position and orientation. These tasks are fused into the following steps.

The first three steps corresponds to the first rectangular block in figure 4.6. The

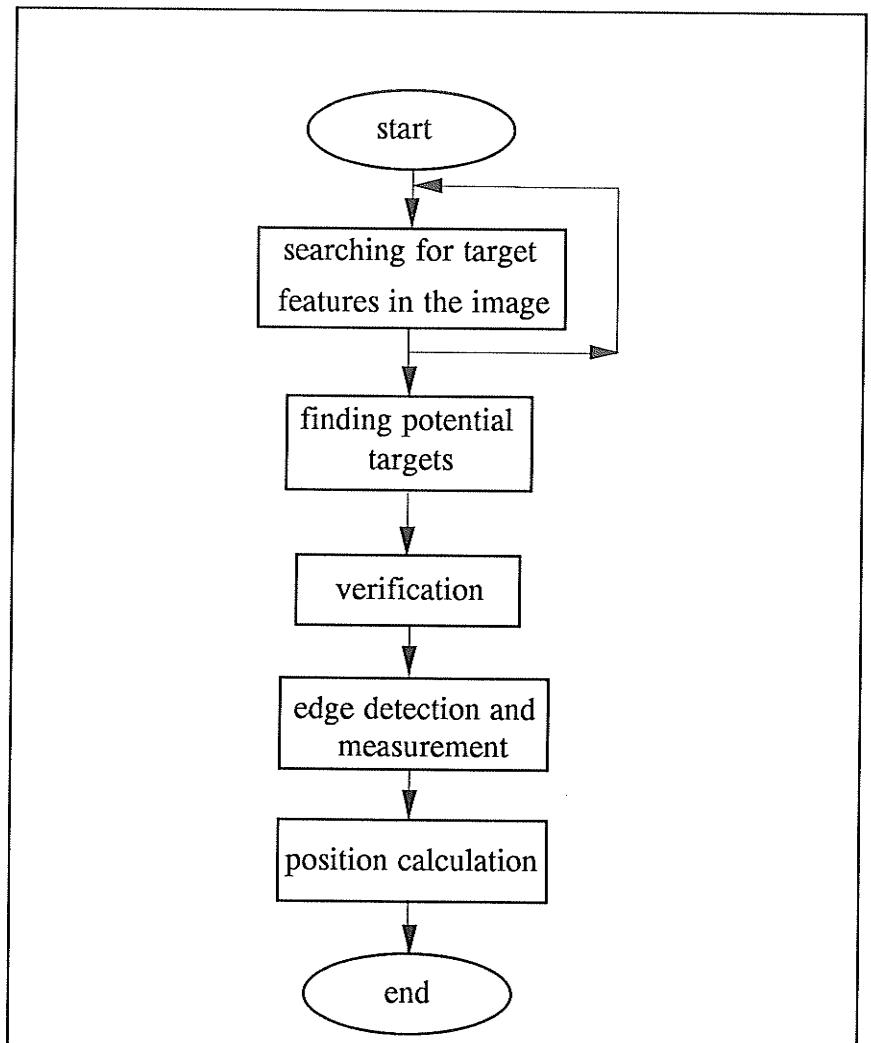


Figure 4.6: Block diagram of processing flow.

first step is thresholding. According to an initial search scheme (coarse search), the image is scanned from top left to bottom right in every 10th line. A threshold value is used to reduce the gray scales of the image data to two levels on one scanned line at a time. This order can be interrupted by a finer search if the preceding processing results indicate to do so.

After the thresholding process, the image data has been classified as either black or white. As a result, the data of scanned line then can be represented by much more

compact parameters. A modified one dimension run length coding method is employed for this purpose.

In the target extraction step, the target can be extracted by detecting the target features in the coded data, since we define the feature representation in a runlength code. If a target feature is matched in the data, the search scheme changes to a finer search if it has not been so, and the next thresholding process is continued in the top and bottom neighborhood of this scanned line until no matches are found. Then the initial search scheme resumes.

Following target extraction, the found potential targets are verified against the random noise and false pattern by using the target characteristics and constraints, such as high contrast and uniform target region.

Edge detection gives a clearly defined boundary points of the target. This detection is only applied to designated area in the target image so that the processing time is reduced. The boundary points obtained by the edge detection process are used by the measurement procedure to obtain the information of the target image geometry.

The final step uses the obtained target geometry in image plane and a priori knowledge to calculate the camera position and orientation. The relationship discussed in previous sections is used in this step.

4.4 Thresholding and image data coding

This section is the continuation of section 3.5.1 but in more detail. Digital representations of images usually require a very large number of bits. Reducing the

representation to a small number of components while still carrying enough discriminating information is important to the succeeding processes such as feature detection. Certain coding technique can compact the image data even further, and the results can be used directly for target feature detection.

4.4.1 Search procedures

Since one does not know where the targets will appear, in order to extract the target feature, it is necessary to process the data of the whole image. However, this is a rather inefficient way of processing. As the target features most likely repeat themselves over several scan lines, It is not necessary to carry out the searching line by line.

For a whole image, the scan procedure is started with every N scan lines (N=10, it is the number that target image will still produce a minimum of two scan patterns at the maximum range), defined as an initial search pattern (coarse search). It begins with the first scan line (the first processing line), and proceeds to the next processing line which is the next *Nth* line down of current line and then proceeds to next *2*Nth* line down of current line, and so on. The initial search pattern is interrupted when the feature detection result of current processing indicates that there is a matched target feature in the current processing line. Then the second search pattern (finer search) is invoked, which processes the neighborhood area of the current processing line. For example, the current processing line is m (say 30) and there is a matched target feature, the next processing will first proceed from the current processing line (30) upward line by line (maximum number is *N-1*, since this is the first encounter) until

no match is found, and then the searching process will go from the first encounter line (30) downward line by line until no match is found. Then the initial search pattern is resumed. The procedure repeats itself until it reaches the last scan line of the image. A scan line is called processing line in the context of the following paragraph if image data on that line is undergoing the processing.

4.4.2 Thresholding technique

Grey-level thresholding is a technique which classifies the pixels into different classes according to their gray levels. Usually one knows that pixels in each class represent a meaningful region in the sense of his interest. In simple cases, a threshold operation is based exclusively on the grey-level of a test point, irrespective of its position in the image or any local context. It may be viewed as a test involving a function T of the grey-level at a point:

$$T = T(g(x, y)) \tag{4.16}$$

where $g(x, y)$ is the grey-level at the point (x, y). If $g(x, y) > T$ then the point (x, y) is labeled an object point, otherwise it is labeled a background point, or conversely.

When an object exhibits a uniform grey-level and rests against a background of a different grey-level, such thresholding can result a good segmentation of the object and the background. But the real problem in practice is that same object exhibits a different grey level range under different illuminations. For example, pixels of a dark object have grey level range of 0-40 under a fair lighting, but possibly have 20-80 under a high illumination. If a threshold is determined based on the first situation, it probably will not be suitable for the second case.

Currently the implemented thresholding process is a combination of global and local thresholding. Global threshold levels are determined empirically beforehand. The local threshold level is determined from the processing line image data. To be specific, the contrast in the area of five pixel is computed across the scan line. If the maximum of the contrast exceeds a certain value, the grey level midway bewteen the maximum and minimum brightness values will be used as the threshold in the scan line. This process sometimes creates a different threshold but it will not affect the thresholding of the target, since the target image consists of high contrast uniform regions, the maximum contrast in the lines containing the target are almost same.

According to processing procedure, a particular image line is scanned. The image data on this line is read into a buffer where the thresholding process is used. The original image data, however is maintained for later use. By comparing each pixel's grey level in the line buffer with the threshold T, a bilevel image data, "g_0" and "g_1" is obtained. Group "g_0" is a string of pixels which are "black" continuously. Group "g_1" is a string of pixels which are "white" continuously. Thus a typical line consists of combinations of these two groups. For example, g_0-g_1-$g_1$$\cdots$ or g_0-g_1-$g_0$$\cdots$. In this type of processing lines, "black" and "white" groups alternate.

Figure 4.7 shows a typical scan line before and after the thresholding process. The threshold used in this figure is chosen arbitrarily to be 120, the g_0 and g_1 are assigned the value 10 and 180 respectively. It can be seen from the figure that the selected threshold level determines each group size. When the threshold level increases, the number of "black" pixels of group g_0 increases while the number of "white" pixels of g_2 decreases.

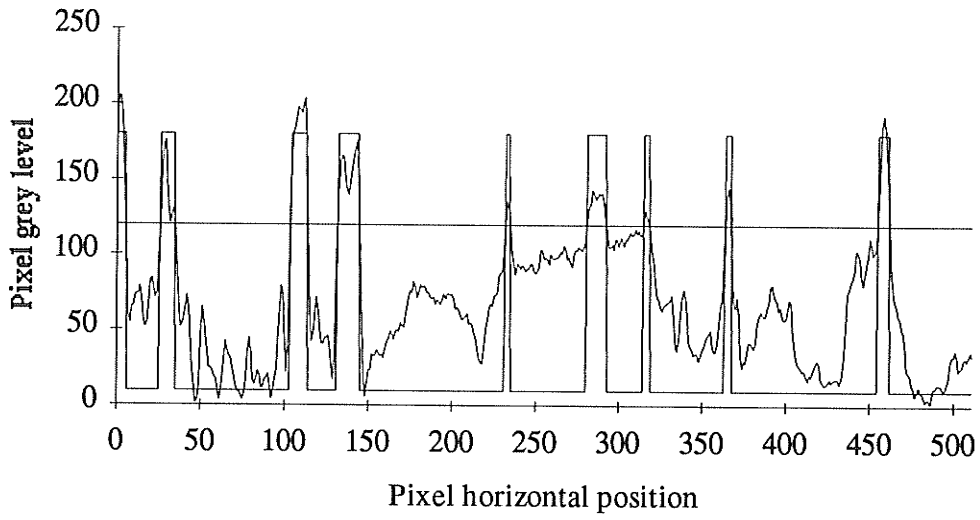


Figure 4. 7: Before and after thresholding on a scan line

4.4.3 Binary data coding

After the thresholding process, the scanned line image data is reduced into two grey levels, or binary data which can be compressed even further by using some coding techniques. Coded data represents the pixels in a much more compact form which is more useful for the target feature detection. The important consideration here is the choice of a proper coding technique. It should be able to reduce the data, keep the information we are interested in, and the most importantly, that the feature reresentation is of simple form after the coding.

We have implemented a modified one-dimensional run-length coding (RLC). Run-length coding exploits the fact that along any particular thresholded scan line there will usually be long *runs* of zeros or ones. It maps a sequence of integers(binary data) into a sequence of integer pairs (g_i, l_j), where g_i denotes gray levels ("1" or

"0") and l_j denotes run-lengths which is the number of pixels having the same gray level g_k. For example, if image data is as shown in the following, the run-length coding result of the image data is a sequence of integer pairs, ("0", 2), ("1", 4), ("0",

0	0	1	1	1	1	0	0	0	1	1	0	0	0	0

3), ("1", 2), ("0", 4), and called records, where the actual value of "1" and "0" are not important, since it functions only as a label. This process starts at the leftmost pixel of processing line till it comes to the end of the processing line, this procedure is repeated for the next processing line data.

In the modified coding process, the positions of some important pixels such as the vertical positions of the first pixels in each integer pairs are also included in the coded data. Thus retrieving the pixels' positions can be made from single coding record rather than decoding whole scan line data to reduce computing time. So the actual coding process maps the binary data of a scan line into a sequence of integers (g_i, l_j, y, x_k), g_i and l_j as mentioned before, y is the position of the processing line (for this line RLC, y is unchanged), x_k is the vertical position of the first pixel in each runs.

4.5. Target recognition

4.5.1 Target feature detection

The target feature detection is to identify the target features within the modified run-length coding data. Since the target's black-white-black features are represented by a number of stacked horizontal line segments of black-white-black in the image,

and a representation of the feature can be found in some scanned lines. Target feature detection is performed by examining the coding data to see if any data sequences fit into the target feature pattern, i.e. three consecutive sequences of integers having an approximately equal number of pixels. If found, and the segments are a first encounter, a new target group is created, otherwise the segment will be put into an existing and corresponding target group. If more than one data sequence is found to satisfy the test of equal number of pixels in one processing line, the same operation is applied to each of these segments. After feature detection, the rest or the whole of the coded data of this processing line is discarded and the processing for this line ends. The program returns to the thresholding process again.

4.5.2 Pattern verification

At this point of the process, potential targets are obtained which are represented by a stack of horizontal line segments. Due to simple matching of the target feature, the obtained potential targets need to be tested on the original image data against some assumptions and constraints to yield the true ones, which are:

(1) Minimum height assumption (expressed in terms of number of scan lines, 15 is our empirical number): this number represents the minimum number of horizontal lines that a target has to have, based on the fact that the target has certain size. Aspect ratio of each potential target is also checked to see if it satisfies the target aspect ratio. This condition is very effective for filtering out some small features in the background which show a similar target feature after thresholding.

(2) High contrast and uniform grey-level: the three line segments should satisfy

this condition. Within each segment, the grey-levels of pixels should be uniform in an ideal case. Practically, there are certain variations in each segment. Grey-level difference of non-boundary pixels within one segment should be less than 20 to satisfy uniform grey-level constraint. Grey-level difference in two adjacent segments should be more than 100 to satisfy the high contrast constraint.

(3) Connectivity: since each target is represented by a stack of horizontal line segments. In the image domain, these line segments are vertically adjacent to each other to form solid regions as one sees in the visual world. The line number consisting in an entity of target should be made up of a continuous sequence of integers (say, 28, 29, 30, 31,...), i.e. exhibit connectivity. If there is no connectivity breaks, i.e. the line numbers are no longer sequential, we will treat the stack as two separate parts which will be verified separately.

A stack of horizontal line segments is accepted as a target entity when it satisfies all three conditions, otherwise it is abandoned.

4.6 Edge detection

The most useful information in an image is contained in the region where a change of gray level or intensity occurs, i.e. at the edge location, since edges define the boundaries between regions with different properties. Up to this stage, the targets are found and their "rough" positions in the image are known. We use "rough" here, since the accurate boundary points have not been obtained. To get an accurate measurement of the target, clearly defined boundary points are necessary. Figure 4.8 is a typical gray level profile along a horizontal scan line which contains vertical

sharp edges in the visual world. On a close look of the "seemingly" vertical lines in the figure 4.8, which indicate the presence of the vertical edges, one can see that for a sharp edge in the direction normal to the edge, the transition from the black to white pixels or white to black usually takes several pixels. For our image device, the sharp edge transition from the white to black region or vice versa takes five pixels.

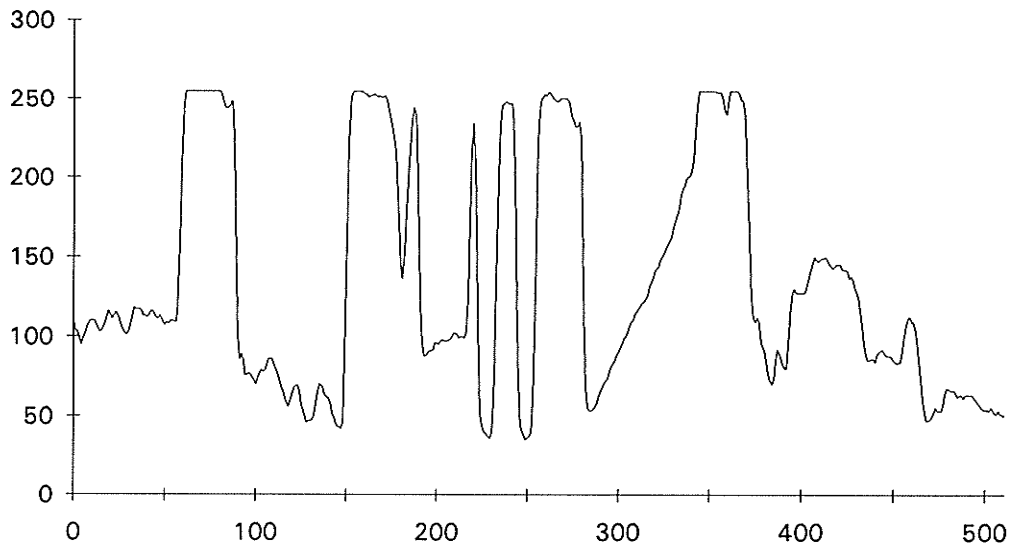


Figure 4.8: Gray-level profile of a scan line.

4.6.1 Basic formula for edge detection

An edge point can be detected from local edge area where gray level is changing rapidly by using an edge detector which is a mathematical operator (or its computational equivalent). Figure 4.9 shows an image of a simple light (dark) object on a dark (light) background, the gray-level profile along a horizontal scan line of the image, and the first derivative of the profile. An edge is modeled as a ramp rather than an abrupt change of grey level due to the fact that edges in digital images are blurred as a result of sampling [9].

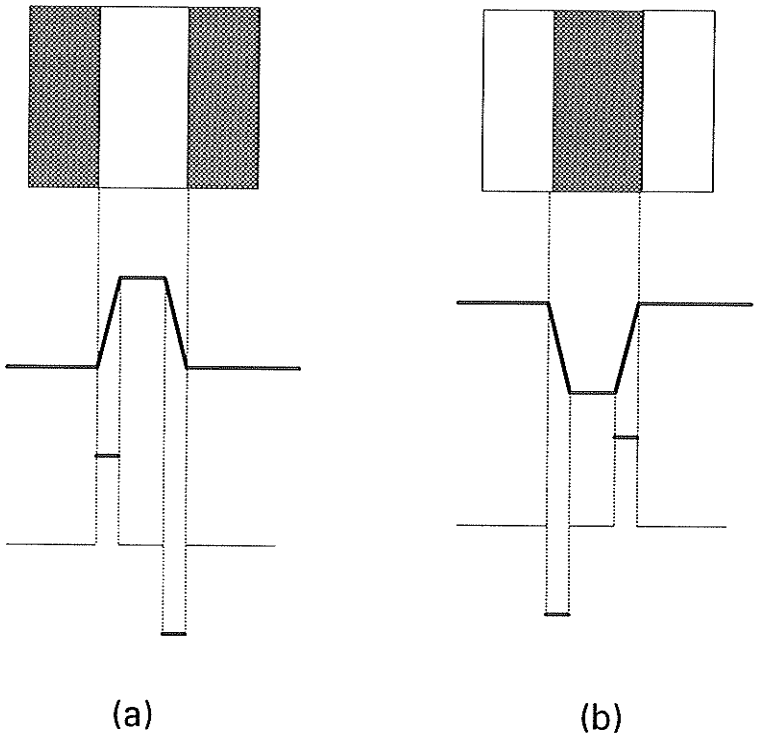


Figure 4.9: Elements of edge detection by derivative operators (a) light object on a dark background (b) dark object on a light background.

The first derivative of an edge modeled in this manner is zero in all region of constant gray level, and assumes a constant value during a gray level transition, which is very similar to actual case. Based on these remarks as illustrated in figure 4.9, the magnitude of the first derivative can be used to detect the presence of an edge. The most commonly used method of obtaining first derivative at any point in an image is the gradient at that point. Given a function $f(x,y)$ (assuming $f(x,y)$ is continuous), the gradient of $f(x,y)$ at coordinates (x,y) is defined as the vector

$$\mathbf{G}[f(x,y)] = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} G_x \\ G_y \end{bmatrix} \tag{4-17}$$

The vector ***G[f(x,y)]*** points in the direction of the maximum rate of increase of the function *f(x,y)*. The magnitude of ***G[f(x,y)]*** gives the strength measure of *f(x,y)* in the direction of ***G***.

It is a common practice to approximate the gradient magnitude by absolute value:

$$G[f(x,y)] \cong |G_x| + |G_y|. \tag{4-18}$$

For discrete image data, the computation of equation (4-18) can be approximated by differences. One typical approximation is given by

$$G[f(x,y)] \cong |f(x,y)-f(x+1,y)| + |f(x,y) - f(x,y+1)|. \tag{4-19}$$

The equation requires 2x2 pixel region as illustrated in figure 4.10 for the implementation. The value of the gradient is proportional to the difference in gray levels between two adjacent pixels. Thus, as expected, the gradient will be relatively large in the regions which have prominent edges, relatively small in the regions which are fairly smooth, and zero only in the regions which have a constant gray level.

It is noted that the gradient magnitude at any point is composed of two parts. The

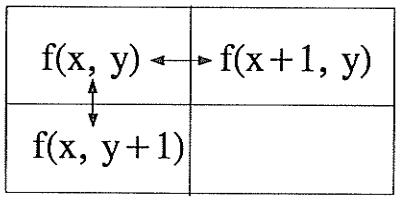


Figure 4.10: A 2x2 pixel region

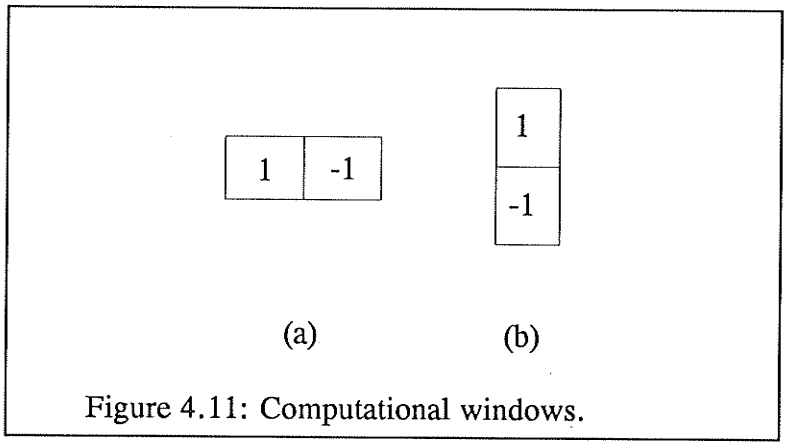


Figure 4.11: Computational windows.

first part is the difference in gray level between the pixel being considered and its next right pixel, and the second part is the difference in gray level between the pixel being considered and its right below pixel as shown in figure4.10. For a sharp vertical edge in the image (such as the target vertical bar in the image), the gradient magnitude on the edge points will depend largely on the first part, while the second part has little contribution to the magnitude of gradient. For a horizontal sharp edge, gradient magnitude on the edge points will be dominated by the second part of equation (4-19). For these two cases (edges are either vertical or horizontal as in our case), the computation of gradient magnitude can be simplified further without affecting much of accuracy in the position determination of edge points. For extracting vertical edges, the following simplified equation has be used:

$$G[f(x,y)] \approx |f(x,y) - f(x+1,y)|; \qquad (4\text{-}20a)$$

while for extracting horizontal edge points:

$$G[f(x,y)] \approx |f(x,y) - f(x,y+1)|. \qquad (4\text{-}20b)$$

4.6.2 Target edge point detection procedures

The computational windows equivalent to the equation (4.20) is shown in figure 4.11. To detect the vertical edge points of the found targets, a 1x2 horizontal window as shown in figure 4.11(a) is used on the horizontal lines defined by the found targets. Since approximate positions of vertical edge points of the target are known, the window moves from left to right horizontally on the local area around the target edge points. We define the local area being nine pixel wide, since we want that the local area covers the sharp edge transaction area to save computing time. At each window position, the two pixels masked by the window are multiplied with the corresponding window element and the products are summed up.

Figure 4.12 illustrates the procedures of vertical edge point detection. The result of the operation is, in fact, the magnitude of gradient $G[f(x,y)]$, of the point masked by the window's left element as described. Since the computation window used is discrete, we assume that when the maximum gradient of the local area occurs, the actual edge point is lying half way between the two pixels masked by the window. Thus the maximum error in the edge point position is one half pixel. This step is repeated on all the horizontal lines of the found targets. A mathematical average operation is then applied on corresponding vertical edge points. If X (which can take a fraction) defines the horizontal coordinate in pixel of a vertical edge, it can be obtained by $X = \Sigma x_i/n$, where x_i is an edge point of this vertical edge, and n is the total number of the edge points consisting of this vertical edge.

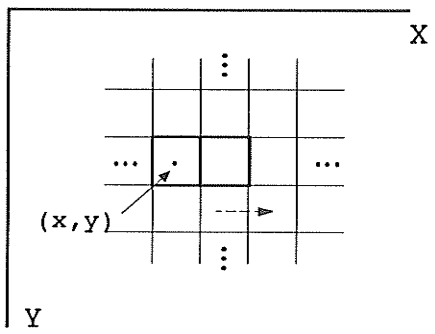


Figure 4.12: Vertical edge point detection

To detect the bar heights, the vertical centerline of each bar is first calculated by using the obtained vertical edge information from the previous step and rounded to the cloest integer and than a 2x1 vertical window as shown in Figure 4.11 (b) is used. It moves up and down in the local area defined by the top-most lines and then on the the bottom-most lines. Similarly, the horizontal edge points in relation to the vertical centerline of each bar are obtained.

4.7 Target geometry parameters

After obtaining the edge positions, we are able to determine the target geometry in the image as shown in figure 3.6. The target width w in pixel would be $w = X_1 - X_2$, where X_1 and X_2 are two outmost vertical edge coordinates. The center heights of the two vertical bars h_1 and h_2 , is equal to the distance between the two corresponding horizontal edge points respectively. The center height h is the average of h_1 and h_2, $h = (h_1 + h_2)/2$, and its horizontal coordinate is the arithmetic center of the two vertical centerlines of the two bars. However, since these parameters are expressed in pixels in the image domain, we have to convert them into physical units

to calculate the camera position and orientation with respect to the individual target.

Pixel and physical units are related according to the image device used. In our system, the dimensions of the image are is 8.8mm x 6.6mm (H x V) corresponding to 512 x 480 pixels. The centre of imaging device is located at pixel (256, 240) with respect to the left-top corner (0,0). To convert the height and width along the target centerline, the following equations are used:

$$h_m = h*(6.6mm/480)$$

$$w_m = w*(8.8mm/512)$$

where h_m and w_m are height and width of the target in millimeters respectively. To convert the position of the target center into physical units, the following equations are used:

$$x_m = (x-256)*(8.8/512)$$

$$y_m = (y-240)*(6.6/480)$$

where the x_m and y_m are the coordinates of the point in millimeters. These parameters are substituted into the obtained equations of (4.4), (4.11) and (4.13) for the camera position determination.

CHAPTER 5

Experimental Results

In the previous chapter, we have discussed the relationship between the target and its image and derived equations necessary to determine the camera's position and orientation with respect to the target. We also developed an algorithm to obtain the necessary parameters for this purpose. Here we present the experimental results conducted on different situations to test the robustness of the algorithm.

5.1 Experiments

5.1.1 Test environment

To demonstrate the validity of our algorithm for target recognition and position determination purposes, the vision system was exposed to a real manufacturing environment, E.H. Price. Figure 5.1 is an image of an industrial manufacturing environment taken at E. H. Price. The image reflects a typical scenario on the manufacturing floor. There are many walls on which our targets can be easily fastened, this reiterates the advantage of planar targets as discussed in chapter three. There are various objects with different shapes or edges in the environment. This fact suggests that unique target design as mentioned in chapter three is essential to a reliable target recognition algorithm. Lighting, surrounding, etc. vary from place to place. All these factors indicate that many considerations must be taken into account

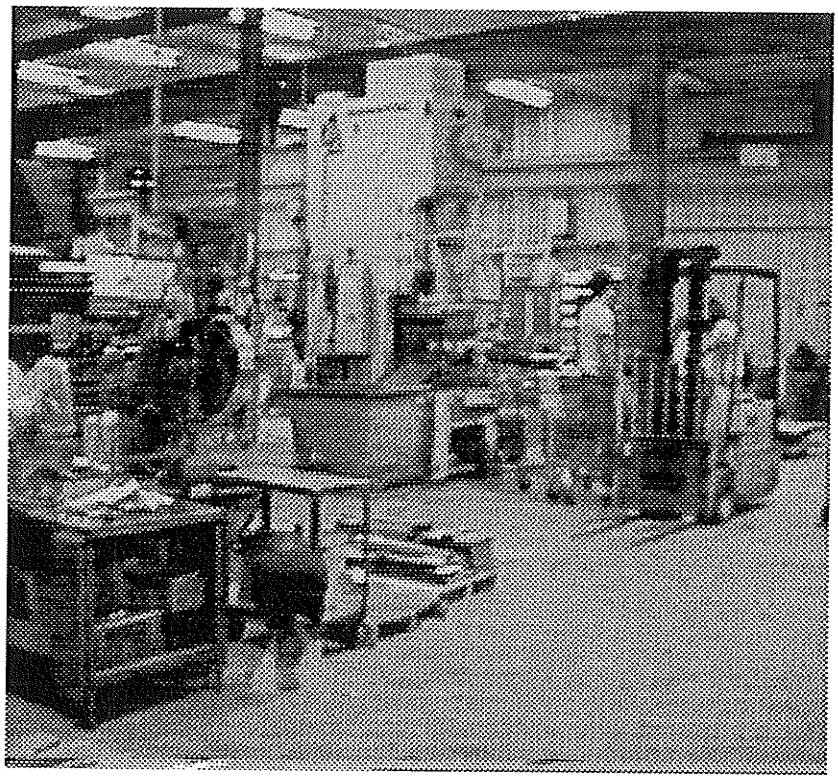


Figure 5.1: Typical scene of an industrial environment from E.H. Price.

in the system's development.

During the algorithm development phase, tests were conducted on different sites in the Engineering building at University of Manitoba, made to resemble as much as possible the real manufacturing environments as shown in figure 5.2 with the result in next section. After having developed the algorithm, we tested our vision system in a real industrial environment (E.H. Price manufacturing floor) using a modified code.

5.1.2 Test methods

The targets of size 225mmx150mm (WxH) were used in the experiments, and set at different positions and orientations under various lighting and surrounding conditions to test robustness of the algorithm. The vision system mounted on a cart

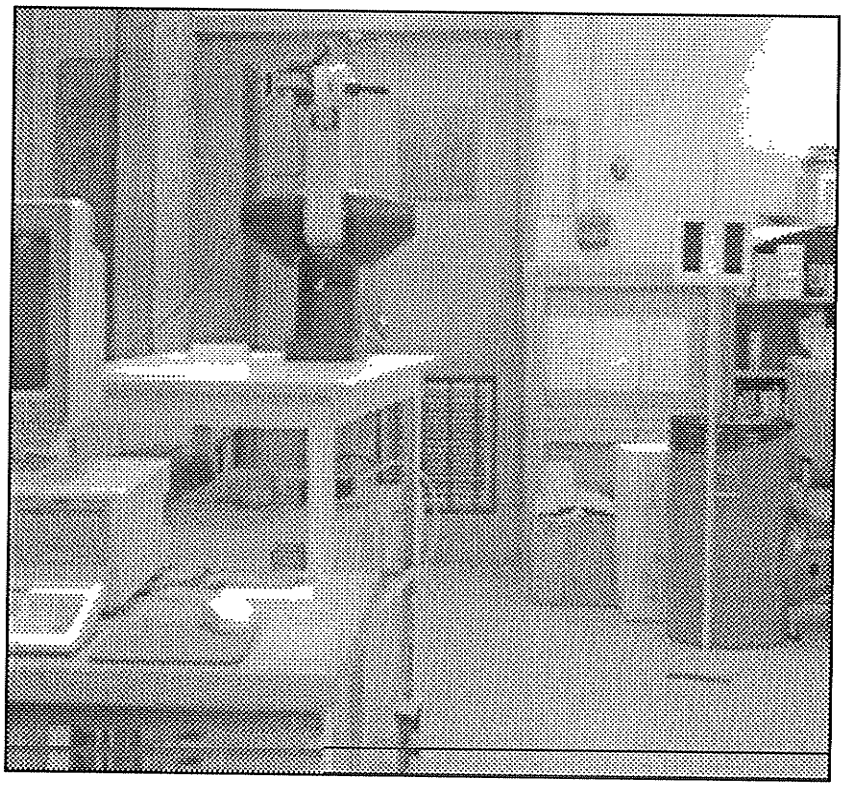


Figure 5.2: Typical scene inside the engineering building.

was moved freely by an operator within the test environment, and stopped at test points which were selected according to different combinations of distance and orientation with respect to the target. The camera focus and aperture were preset. Two lenses with different focal lengths were used for comparison. In the algorithm, the range z , horizontal angle β' , and orientation α were initially obtained from the image plane. Then the horizontal distance d between the camera and the target was obtained from z and β' by using equation (4.5). To check the accuracy of the algorithm, the horizontal distance d and orientation α were also hand-measured. The robustness of the algorithm has been checked over a wide range of conditions as shown in the following experimental results.

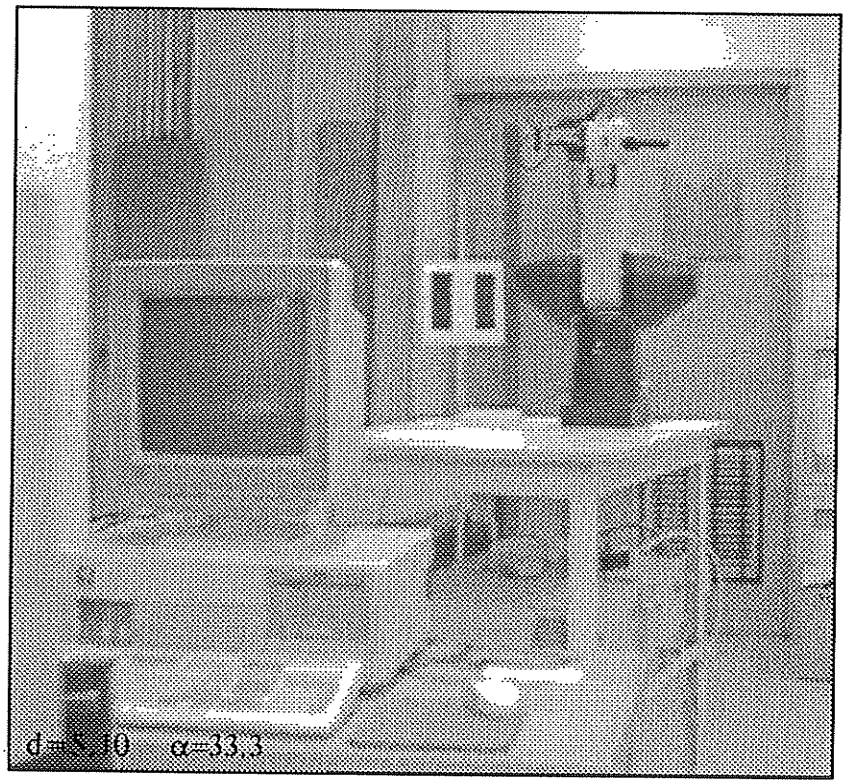


Figure 5.3: An image containing a found target.

5.2 Results and discussion

Figure 5.3 shows the image of a target in typical conditions and the results of position determination. The number on the image are d and α calculated from the algorithm. The algorithm can successfully recognize the target and calculate the orientation angle α and the horizontal distance d, which are calculated from range and angle β'. Because of efficiency consideration within the algorithm, total computing time for this kind of calculation turns out to be less than a quarter of a second, which is very short compared to other algorithms [5] [11], suggesting that the system is suitable for real time applications. This being a typical result from a test point, the same test was conducted over 18 other points.

To facilitate the results analysis, Table 5.1 and Table 5.2 are arranged according to actual distance in an ascending order. They show the results of actual measurement data, calculated data and accuracy on each test point in a wide range of conditions. The camera focal length for the experiments shown in Table 5.1 and Table 5.2 are 16 mm and 50 mm, respectively. The distances are 1410 mm to 5720 mm in Table 5.1 and 4910 mm to 9520 mm in Table 5.2. It is clear that the vision system successfully computed a wide range of distance data. Comparing the measured distance and calculated distance, one can find that the average relative error is less than 4.3 percent and 2 percent in table 5.1 and table 5.2 respectively, which is satisfactory for our particular application. Since the vision system will be asked to search the target and update the position data from time to time, the error from each point will not be accumulated in the vision system. Causes of errors are mainly considered to be following factors:

Distance d (mm)		Relative error	Orientation angle α (deg.)	
Measured	Computed	(distance)	Measured	Computed
1410	1362	3.4%	0.0	4.4
1990	1905	4.3%	43.0	45.8
2030	1975	2.7%	5.0	13.0
3240	3164	2.3%	-23.0	-20.0
3710	3584	3.4%	-50.0	-46.4
4060	3913	3.6%	51.0	48.1
4380	4109	6.2%	42.0	40.0
4400	4126	6.2%	30.0	28.3
5720	5341	6.6%	70.0	71.8

Table 5.1 Experiment results arranged by distance (focal length =16mm)

Distance d (mm)		Relative error	Orientation angle α (deg.)	
Measured	Computed	(distance)	Measured	Computed
4910	4881	0.6%	37.0	38.8
5910	5714	3.3%	-40.0	-37.6
5940	5929	0.2%	10.0	14.5
6330	6103	3.6%	65.0	61.7
6370	6281	1.4%	46.0	43.0
6540	6318	3.4%	57.0	53.1
7330	7410	-1.1%	60.0	57.3
8150	8074	0.9%	-20.0	-17.2
9520	9334	2.0%	0.0	6.8

Table 5.2: Experiment results arranged by distance (focal length=50mm)

(1) Target resolution in the image: the further the camera is far away from the target, the smaller the target image on the image plane becomes. In other words, the target image has a lower level of resolution. As a result, different focal lengths will result in different levels of accuracy. In this respect, the higher the focal length used, the lower the error margins is. (comparing point 8 in Table 5.1 with point 1 in Table 5.2, which have an approximate distance). By using a zooming optical system, the accuracy can be improved.

(2) Camera focus: When the camera is out of focus, the edges of the target image become blurred, which causes some errors in the computation. By using a auto-focusing camera, the accuracy of the system can be improved.

Table 5.3 and Table 5.4 show the same results arranged according to α in an ascending order. The average accuracy error is less than 5°(except for two cases, which will be explained later). Since computing α depends on two independent

Orientation angle α (deg.)		Absolute error	Distance d (mm)	
Measured	Computed	(degrees)	Measured	Computed
0.0	4.4	-4.4	1410	1362
5.0	13.0	-8.0	2030	1975
-23.0	-20.0	-3.0	3240	3164
30.0	28.3	1.7	4400	4126
42.0	40.0	2.0	4380	4109
43.0	45.8	-2.8	1990	1905
-50.0	-46.4	-3.6	3710	3584
51.0	48.1	2.9	4060	3913
70.0	71.8	-1.8	5720	5341

Table 5.3: Experiment results arranged by orientation angle (focal=16mm)

parameters in the image plane, the target image height and width, the errors for angle α are larger than that in the case of distance computing. In practice, when α is larger than a certain angle, the target image is so distorted that the target image fails to meet criteria of the algorithm. The test results indicate that the upper limit for the α angle is about *70* degrees. Both Table 5.3 and Table 5.4 show no data obtained after 70 degrees. Examining the first row of the tables, one notes that poor accuracy results when angle α is smaller than *20* degrees. The major reasons are that (1) α is not monotonic in the interval of $0\sim2\beta'$, and (2) the change of target image width due to the change of angle α is very insensitive in the interval of $0\sim2\beta'$ degrees. For example, the target image width has only changed slightly from 100% to 94% when α changed from *0* to *20* degrees.

Our system can process multi-targets in one image and relate the camera's position to each target. Figure 5.4 shows two targets appearing in one frame of image. This type of situation may be encountered by the vision system in the practical applications. Since the algorithm treats each target as an individual entity, the search procedure is identical to that of a single target case. The parameters obtained from one target are not affected by those originating from otherss. When one target fails to give the required information, another target can be used to calculate the camera's position. The computing time for processing a multi-target image is lengthened only by a fraction of time required for a single target image.

Orientation angle α (deg.)		Absolute error	Distance d (mm)	
Measured	Computed	(degrees)	Measured	Computed
0.0	6.8	-6.8	9520	9334
10.0	14.5	-4.5	5940	5929
-20.0	-17.2	-2.8	8150	8074
37.0	38.8	-1.8	4910	4881
-40.0	-37.6	-2.4	5910	5714
46.0	43.0	3.0	6370	6281
57.0	53.1	3.9	6540	6318
60.0	57.3	2.7	7330	7410
65.0	61.7	3.3	6330	6103

Table 5.4: Experiment results arranged by angle (focal length=50mm)

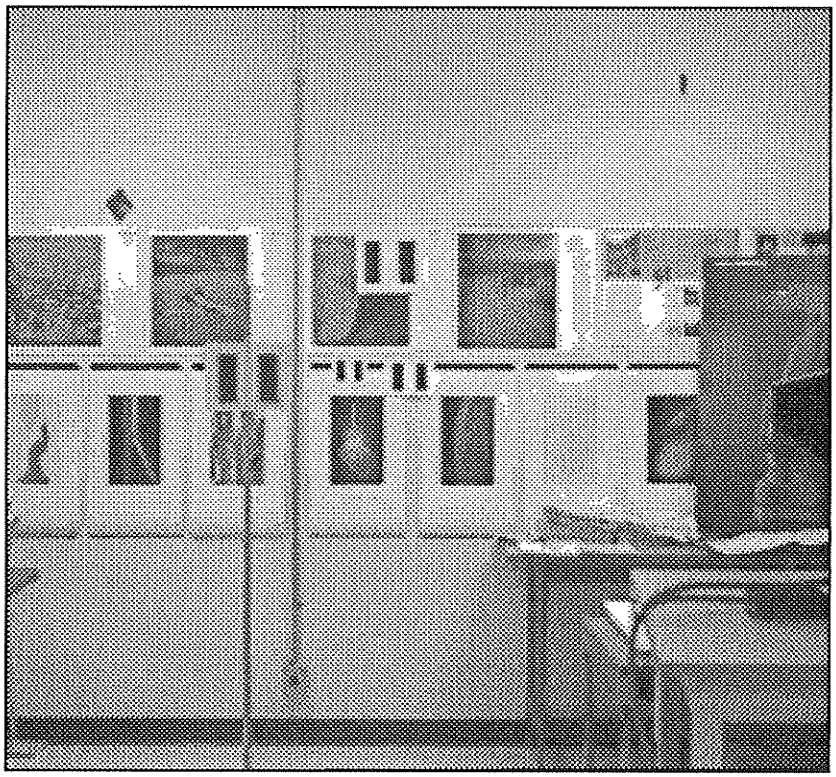


Figure 5.4: Multi-targets appearing in one frame of image

Figure 5.5 shows an image in which the target is exposed to high illumination conditions under which tests were conducted. As the image under the high illumination becomes saturated, contrast is reduced. In this case, the described threshold detection was used to detect target features. The results show that the selected threshold was able to accommodate the change of illumination. The algorithm can also deal with the negative target equally well.

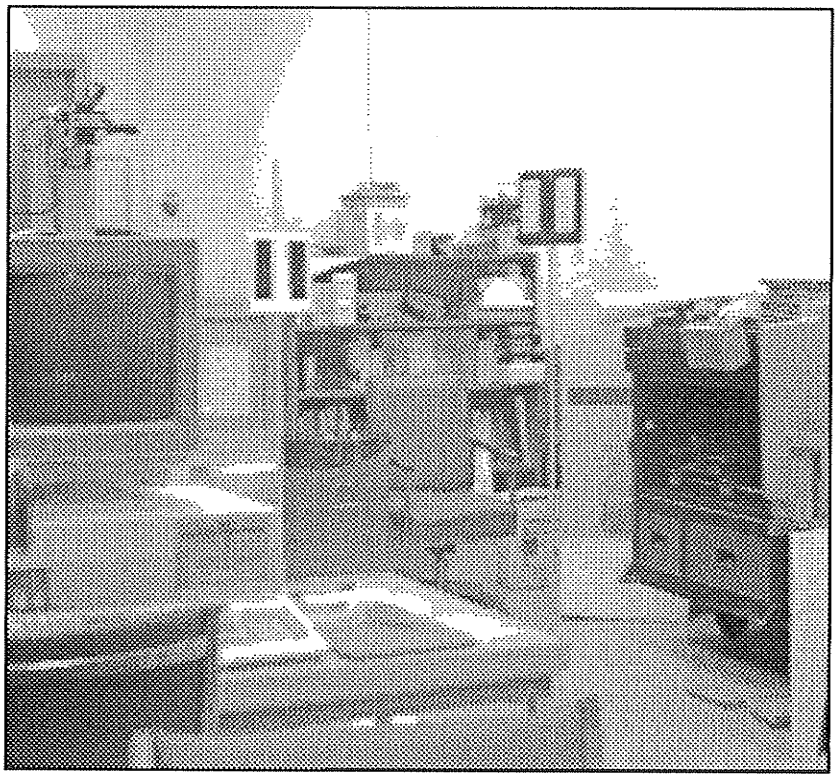


Figure 5.5: Image of a target under conditions of high illumination

CHAPTER 6

Conclusions and Recommendations

In this thesis, a machine vision system has been proposed for the navigation of autonomous guided vehicles. The features of the system and experimental results can be summarized as follows:

(1) The system is based on a single camera, a priori knowledge about the objects to be viewed, and a map of the related environment for updating the current position and orientation of the vehicle while avoiding general scene analysis and interpretation. Neither three-dimensional reconstruction nor solution of the feature correspondence problem are necessary with this vision system. It has been proved that this vision system provides an effective way for updating vehicle position.

(2) A two-dimensional target was designed to be used for with vision system. This target is uniquely designed so as to distinguish itself from the surroundings, simple so as to reduce computing time, and easy to install. We have used the target image middle point in an approximating approach to derive a set of equations which efficiently compute the position and orientation of the vehicle with respect to the target. Using this method to obtain orientation angles is simpler than dealing with theoretical solutions.

(3) An algorithm was developed to complete target recognition and image plane measurement, as well as, subsequently, to yield the camera's position data. The computing time of this algorithm was shown to be less than a quarter of a second.

(4) The developed vision system has been tested in a real industrial environment. The experimental results prove that the system is reliable, robust and that the accuracy of the recovered position data is satisfactory for our purposes. The vision system is adequate for an autonomous guided vehicle.

Presently, in our research, we are concentrating on developing system features which would provide a statement of the camera's (vehicle) position within a map-given environment. Further work would be desirable to expand the vision system to be able to cope with unforeseen changes in the environment, such as the appearance of new obstacles in the path of the vehicle. This would therefore enhance the capabilities of the vehicle.

References

[1] F. Andresen, L. Davis, R.D. Eastman and S. Kambbampati, *"Visual algorithm for autonomous navigation"*, Proceeding, IEEE International Conference on Robotics and Automation, pp.656-661, (1985).

[2] R. Arkin and R. Murphy, *"Autonomous Navigation in a manufacturing environment"*, IEEE Transaction on Robotics and Automation, Vol.6, No.5, pp.445-454,.(1990).

[3] M. Augusteijn and C. Dyer, *"Recognition and recovery of the three dimensional orientation of planar point patterns"*, Computer Vision, Graphics, and Image Processing, Vol.36, pp.76-99, (1986).

[4] S. T. Barnard, *"Interpreting perspective images"*, Artificial Intellegence 21, pp.435-462, (1983).

[5] D. Brzakovic and L. Hong, *"Road edge detection for mobile robot navigation"*, Proceeding, IEEE International Conference on Robotics and Automation, pp.1153-, (1989).

[6] J. Courtney and J. k. Aggarwal, *"Robot guidance using computer vision"*, Pattern Recognition, Vol.17, No.6, pp.585-592, (1984).

[7] M. Dhome, M. Richetin, J. Lapreste and G. Rives, *"Determination of the attitude of 3D objects from a single perspective view"*, IEEE transaction on pattern analysis and machine intelligence, Vol.11, No.12, pp.1265-1278, (1989).

[8] M. Fischler and R. Bolles, *"Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartograpgy"*,

Communications of ACM, Vol 24, pp.381-395, (1981).

[9] H. Frohn and W. Seelen, *"VISOCAR: An autonomous industrial transport vehicle guided by visual navigation"*, Proceeding, IEEE International Conference on Robotics and Automation, pp.1155-1159, (1989).

[10] R. C. Gonzalez and P. Wintz, *"Digital Image Processing"*, Second Edition, Addison Wesley,(1987).

[11] Y. Goto and A. Stentz *"The CMU system for mobile robot navigation"*, Proceeding, IEEE International Conference on Robotics and Automation, pp.99-105, (1987).

[12] I. Fukui, *"TV image processing to determine the position of a robot vehicle"*, Pattern Recognition, Vol.15, nos.1-6, pp.101-109, (1981).

[13] R. Horaud, *"New methods for matching 3-D objects with single perspective views"*, IEEE transaction on pattern analysis and machine intelligence, Vol pami-9, No.3, (1987).

[14] R. Horaud, B. Conio, O. Leboulleux and B. Lacolle, *"An analytic solution for the perspective 4-point problem"*, Computer vision, Graphics, and Image Processing, Vol.47, pp.33-44, (1989).

[15] E. Krotkov, *"Mobile robot localization using a single image"*, Proceeding, IEEE International Conference on Robotics and Automation, pp.978-983, (1989).

[16] M. D. Livine, *"Computer determination of depth maps"*, Computer Graphics and Image Processing, Vol.2, pp.131-150, (1973).

[17] M. J. Magee, *"Determining the position of a robot using a single calibration object"*, Proceeding, IEEE International Conference on Robotics and

Automation, pp.140-149, (1984).

[18] K. Onofuchi and M. Watanabe, *"A visual navigation system using a multi-information local map"*, Proceeding, IEEE International Conference on Robotics and Automation, pp.767-775, (1990).

[19] T.Poggio, *"Early vision: From comptutational structure to algorithm and parallel hardware"*, Computer vision, Graphics, and Image Processing, Vol.31, pp.139-155, (1985).

[20] K. Sarachik, *"Characterizing an indoor environment with a mobile robot and uncalibrated stereo"*, Proceeding, IEEE International Conference on Robotics and Automation, pp.985-989, (1989).

[21] K.Sugihara, *"Some location problem for robot navigation using a single camera"*, Computer Vision, Graphics, and Image Processing, Vol.42, pp.112-129, (1988)

[22] M. Turk and D.Morgenthaler, *"Video road-following for the autonomous land vehicle"*, Proceeding, IEEE International Conference on Robotics and Automation, pp.273-280, (1987).

[23] E. Triendl and David J Kriegman, *"Stereo vision and navigation within buildings"*, Proceeding, IEEE International Conference on Robotics and Automation, pp.1725-1730, (1987).

[24] S.Tsuji, J. Zheng and M. Asada, *"Stereo vision of a mobile robot: world constraints for image matching and interpretation"*, Proceeding, IEEE International Conference on Robotics and Automation, pp.1594-1599, (1986).

[25] R. Wallace, K. Matsuzaki, Y. Goto, J.Crisman, J.Webb and T. Kanade

"Progress in robot road following", Proceeding, IEEE International Conference on Robotics and Automation, pp.1615-1621, (1986).

[26] A. Waxman, J.Moigne and B. Srinivasan, *"Visual navigation of roadways"*, Proceeding, IEEE International Conference on Robotics and Automation, pp.662-667, (1985).

[27] A. Waxman, J. LeMoigne, L. Davis E. Liang and T. Siddalingaiah, *"A visual navigation system"*, Proceeding, IEEE International Conference on Robotics and Automation, pp.1600-1606. (1986).

Appendix A

Analytical solution for orientation angle α

The following shows how to calculate orientation angle α assuming the three horizontal angles are known. Figure A.1 shows the spatial relation between the target and the camera, we are going to express the angle α in term of three horizontal angles, width of *AB* and horizontal distance *r*.

Since the sum of the area of ΔAOT and the area of ΔBOT is equal to the area ΔABT, we have following equation:

$$r_1 r_2 \sin(\beta_2 - \beta_1) = r r_2 \sin(\beta - \beta_1) + r r_2 \sin(\beta_2 - \beta) \quad (A.1)$$

where r_1, r_2, and r is the horizontal distance from the lens center to the target left boundary, right boundary, and vertical center, respectively, β_1, β_2, and β are the correspondent angles between optical axis (*YO*) and r_1, r_2, and r, respectively, *T* is the center of the line *AB*. Two more equations can be derived by applying the law of sine to ΔAOT and ΔBOT. For ΔAOT, we have:

$$\frac{\sin(\beta - \beta_1)}{W/2} = \frac{\sin(\alpha + (90^0 - \beta))}{r_1} \quad (A.2)$$

and for ΔBOT, we have:

$$\frac{\sin(\beta_2 - \beta)}{W/2} = \frac{\sin(\alpha + (90^0 - \beta_1))}{r_2}. \quad (A.3)$$

The equations (A.2) and (A.3) can be rewritten as

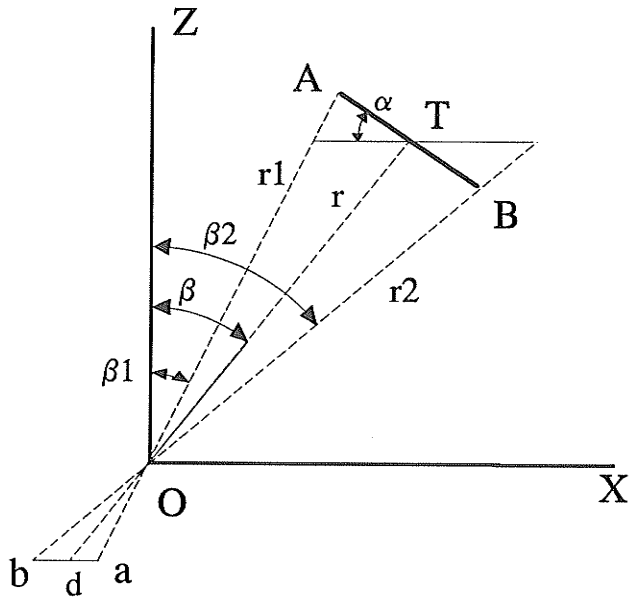


Figure A.1: Spatial relation between target and camera.

$$r_1 \sin(\beta - \beta_1) = \frac{W}{2} \sin(\alpha + 90^0 - \beta) \tag{A.4}$$

and

$$r_2 \sin(\beta_2 - \beta) = \frac{W}{2} \sin(\alpha + 90^0 - \beta), \tag{A.5}$$

where W is the target width. These equations for $r_1 sin(\beta-\beta_1)$ and $r_2 sin(\beta_2-\beta)$ can now be substituted into the right hand side of the equation (A.1) to yield

$$\begin{aligned} r_1 r_2 \sin(\beta_2 - \beta_1) &= r \frac{w}{2} \sin(\alpha + 90^0 - \beta) + r \frac{w}{2} \sin(\alpha + 90^0 - \beta) \\ &= rw \sin(\alpha + 90^0 - \beta) \end{aligned} \tag{A.6}$$

Since we know r ($r=Z/cos\beta$) and W, we are going to express the r_1 and r_2 of the above equation in terms of r and W. Applying the law of cosines to the two triangles ΔAOD and ΔBOD respectively. It gives

$$r_1^2 = r^2 + (W/2)^2 - 2r\frac{W}{2}\cos\left(\alpha + 90^0 - \beta\right) \tag{A.7}$$

and

$$r_2^2 = r^2 + (W/2)^2 - 2r\frac{W}{2}\cos\left(\alpha + 90^0 - \beta\right). \tag{A.8}$$

These last two equations combine to give

$$r_1^2 + r_2^2 = 2r^2 + 2(W/2)^2. \tag{A.9}$$

When the law of cosines is applied to the large triangle ΔABC, a second relation for

$$W^2 = r_1^2 + r_2^2 - 2r_1 r_2 \cos(\beta_2 - \beta_1) \tag{A.10}$$

The formula for $r_1^2 + r_2^2$ in equation (A.9) and the formula for $r_1 r_2$ of equation (A.6) can be substituted into (A.10) to obtain

$$\begin{aligned} W^2 &= 2r^2 + 2(W/2)^2 - 2r_1 r_2 \cos(\beta_2 - \beta_1) \\ &= 2\left(r^2 + \left(\frac{W}{2}\right)^2\right) + 2\frac{r W \sin\left(\alpha + 90^0 - \beta\right)}{\sin(\beta_2 - \beta_1)}\cos(\beta_2 - \beta) \end{aligned} \tag{A.11}$$

The equation (A.11) can be rewritten as

$$r^2 - (W/2)^2 = r\, w\frac{\cos(\alpha - \beta)}{\tan(\beta_2 - \beta_1)} \tag{A.12}$$

i.e.

$$\cos(\alpha - \beta) = \frac{r^2 - (W/2)^2}{r\, W}\tan(\beta_2 - \beta_1), \tag{A.13}$$

and i.e.

$$\alpha = \beta + \cos^{-1}\left[\frac{r^2 - (W/2)^2}{r\, W}\tan(\beta_2 - \beta_1)\right] \tag{A.14}.$$

If we introduce some approximations in the above equation, we can get an identical solution as the one in equation (4.13).

Considering $\beta_1$and $\beta_2$are small, $tan(\beta_1 - \beta_2) \approx \beta_1 - \beta_2$ and $\beta_1 \approx \beta_2 \approx \beta$, we ontain:

$$\alpha \approx \beta + \cos^{-1}[\frac{r}{W}(\beta_2 - \beta_1)]$$

$$= \beta + \cos^{-1}[\frac{r}{W}\frac{\Delta x}{f}\frac{z}{z}] \qquad \text{(A.15)}$$

$$= \beta + \cos^{-1}[\frac{\Delta x}{W}\frac{z}{f}\cos\beta] \approx \beta' + \cos^{-1}[(\frac{\Delta x}{W})(\frac{H}{\Delta y})\cos\beta']$$

Equation (A.15) has same form as equation (4.13).

Appendix B

Target shape evolution

Target shape plays an important role in the vision system as mentioned in chapter 3. In search of the optimum shape, there were a few alternatives being considered during this study. The first one, butterfly shape, is shown in figure B.1. The slant bottom and top edges were chosen and thought to be useful for determining its orientation. However, studies show that the target width in the image plane contains

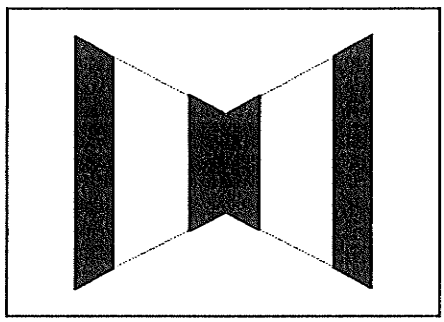


Figure B.1: Butterfly shape target.

the orientation information, and it is not necessary to have slants in top and bottom edges.

The second one, rectangular shape with three black bars, is shown in figure B.2. The space between middle black bar and outside black bar is equal to the width of middle black bar. The features used for the target recognition were width of the spaces and the middle bar which are three segments of equal width alternated in black and white. The two outside narrow black bars served only as a separator between the

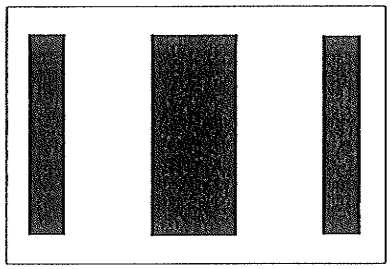


Figure B.2 : Rectangular shape with three black bars.

target features and the background. The experiments as documented in chapter five Table 5.1 and Table 5.2 used mainly this target.

This was later replaced by the current targets as shown in figure B.3. In the current one, there are two black bars on a white plane (or two white bars on black plane). The space between bars is equal to the width of the bars, which also provides three segments of equal width alternated in black and white, so that this modification will not affect the recognition process. In both targets, the area on the left and right side of the two bars are used as feature separator, and the top and bottom of the center bar are not required in the recognition and measurement process.

This last design gives the most usable information in least area. It has been used in the phase of real manufacturing environment test along with the collaborated code .

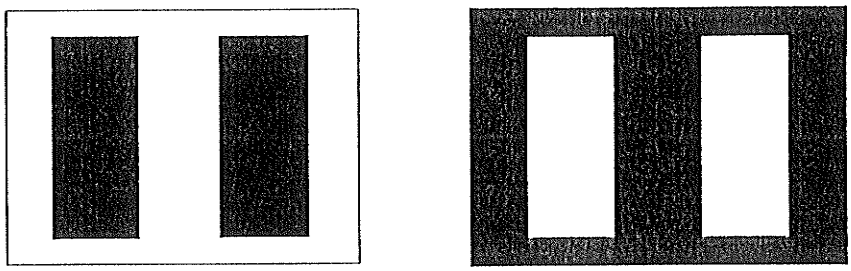


Figure B.3: Target with two black or white bars.