A VISION SYSTEM FOR A SURGICAL INSTRUMENT-PASSING ROBOT

by

KENNETH LING-MAN CHAN

B. A. Sc., The University of British Columbia, 1983

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Department of Electrical Engineering)

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 1985

© Kenneth Ling-Man Chan, 1985
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Electrical Engineering

The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date 30 August, 1985
ABSTRACT

To help control the high cost of health care delivery, a robotic system is proposed for use in passing surgical instruments in an operating room. The system consists of a vision system, a robotic arm, a speech recognition and synthesis unit, and a microcomputer. A complete vision system has been developed using standard and new techniques to recognize arthroscopic surgical instruments.

Results of the vision system software evaluation gave an overall recognition accuracy of over 99%. Also, error conditions were analysed and found to be consistent with the results of a clinical survey on the proposed instrument-passing robot. As well, a payback and cost benefit analysis using estimated system costs and potential labour savings showed that the instrument-passing robot is economically feasible.

Based on the results of this thesis, it was concluded that the instrument-passing robot would be beneficial for reducing the high cost of health care.
# Table of Contents

ABSTRACT .................................................................................. ii  
LIST OF TABLES ........................................................................ v  
LIST OF FIGURES .......................................................................... vi  
ACKNOWLEDGEMENT ..................................................................... x  

1.0 INTRODUCTION ....................................................................... 1  
1.1. Statement of Problem ............................................................. 1  
1.2. Objectives .............................................................................. 3  

2.0 SYSTEM OVERVIEW ............................................................... 9  
2.1. Block Diagram of System ......................................................... 9  
2.2. Description of Components ..................................................... 11  
2.2.1 Computer ........................................................................... 11  
2.2.2 Vision System ...................................................................... 12  
2.2.3 Speech Recognition and Synthesis System ......................... 14  
2.2.4 Robot Arm ........................................................................... 15  
2.3. Gripper - End Effector ............................................................. 18  
2.4. Surgical Instrument Set ......................................................... 18  
2.5. Structured Lighting and Instrument Tray ............................... 19  
2.6. System Integration ................................................................. 21  
2.6.1 Operation Sequence .......................................................... 21  
2.6.2 Role of Surgeon and Nurse in O.R. ...................................... 25  
2.7. Safety .................................................................................... 28  
2.7.1 Payback Period and Cost Effective Analysis ....................... 29  

3.0 VISION SYSTEM ..................................................................... 31  
3.1. Imaging Components Review ................................................ 31  
3.1.1 High-Resolution Imaging Systems ........................................ 31  
3.1.2 Vision System Design ........................................................ 33  
3.2. Selection and Evaluation of Camera ........................................ 37  
3.2.1 Camera Review .................................................................... 38  
3.2.2 Camera Selection ............................................................... 40  
3.2.3 Camera Evaluation ............................................................. 42  
3.3. Selection and Evaluation of Image Digitizer ............................ 48  
3.3.1 Image Digitizer Review ...................................................... 48  
3.3.2 Digitizer Selection ............................................................. 49  
3.3.3 Digitizer Evaluation ........................................................... 51  

4.0 Recognition Algorithm ........................................................... 56  
4.1. General Recognition Problem ................................................. 56  
4.2. Description of Instruments and Layout ................................. 58  
4.3. Survey of Existing Algorithm ............................................... 59  
4.3.1 Overview ........................................................................... 61  
4.3.2 Criteria for Recognition Algorithm ...................................... 62  
4.3.3 Feature Extraction Algorithms .......................................... 63  
4.3.4 Global Features .................................................................. 63  
4.3.5 Boundary Oriented Features ............................................. 65  
4.3.6 Classification Algorithms ................................................... 69  
4.4. Development of Recognition Algorithms .............................. 70  
4.4.1 Low-Resolution Algorithm ............................................... 71  
4.4.2 High Resolution Algorithm ............................................... 77  
4.4.3 Clinical Requirements ....................................................... 105  

5.0 EVALUATION OF RECOGNITION ALGORITHM ..................... 107
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1. Test Methods</td>
<td>107</td>
</tr>
<tr>
<td>5.1.1 Methods of error estimation methods</td>
<td>107</td>
</tr>
<tr>
<td>5.1.2 Test and Training Data</td>
<td>110</td>
</tr>
<tr>
<td>5.2. Results of Testing Each Algorithm</td>
<td>113</td>
</tr>
<tr>
<td>5.3. Failures of Each Algorithm</td>
<td>117</td>
</tr>
<tr>
<td>5.4. Improvements of Recognition Algorithms</td>
<td>121</td>
</tr>
<tr>
<td>5.5. Implementation and Optimization</td>
<td>123</td>
</tr>
<tr>
<td>5.5.1 Practical Problems</td>
<td>124</td>
</tr>
<tr>
<td>5.5.2 Optimization</td>
<td>131</td>
</tr>
<tr>
<td>5.6. Final Results of the Recognition Algorithm</td>
<td>137</td>
</tr>
<tr>
<td>6.0 CONCLUSIONS AND RECOMMENDATIONS</td>
<td>140</td>
</tr>
<tr>
<td>References</td>
<td>145</td>
</tr>
<tr>
<td>Appendix A Questionnaire on Clinical Requirements</td>
<td>148</td>
</tr>
<tr>
<td>Appendix B Payback Period Calculation</td>
<td>153</td>
</tr>
<tr>
<td>Appendix C Program Pseudo-code</td>
<td>154</td>
</tr>
</tbody>
</table>
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1 Cost of IBM PC.</td>
<td>13</td>
</tr>
<tr>
<td>Table 2.2 List of arthroscopy instruments.</td>
<td>20</td>
</tr>
<tr>
<td>Table 2.3 Proposed system cost and scrub Nurse salary.</td>
<td>30</td>
</tr>
<tr>
<td>Table 3.1 Listing of commercial cameras considered for vision system.</td>
<td>41</td>
</tr>
<tr>
<td>Table 3.2 Listing of commercial image digitizers considered for vision system.</td>
<td>50</td>
</tr>
<tr>
<td>Table 4.1 Typical structural matching results.</td>
<td>103</td>
</tr>
<tr>
<td>Table 5.1 List of test data set.</td>
<td>111</td>
</tr>
<tr>
<td>Table 5.2 Results of tests on Acufex algorithm.</td>
<td>117</td>
</tr>
<tr>
<td>Table 5.3 Confusion matrix of all arthroscopic instruments.</td>
<td>118</td>
</tr>
<tr>
<td>Table 5.4 Types of errors in structural match.</td>
<td>121</td>
</tr>
<tr>
<td>Table 5.5 Time comparison results.</td>
<td>136</td>
</tr>
<tr>
<td>Table 5.6 Arthroscopic instrument usage probability.</td>
<td>139</td>
</tr>
<tr>
<td>Table 5.7 Overall recognition accuracy.</td>
<td>139</td>
</tr>
</tbody>
</table>
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Cartilage knives (top) and Acufex instruments (bottom)</td>
<td>5</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>Obturators (top) and low resolution instruments (bottom)</td>
<td>6</td>
</tr>
<tr>
<td>Figure 1.3</td>
<td>Tips of cartilage knives (top) and 3 Acufex instruments: LGF, LPS, RPS (bottom)</td>
<td>7</td>
</tr>
<tr>
<td>Figure 1.4</td>
<td>Tips of Acufex instruments: L20H, R20H, L60H, R60H (top); LSS, RSS, LBP, RBP (bottom)</td>
<td>8</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Typical robotic system configuration.</td>
<td>10</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>Common geometries for robot arms.</td>
<td>17</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>Drawing of backlight unit.</td>
<td>22</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>Pre-operation initialization flowchart.</td>
<td>24</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>Vision system flowchart.</td>
<td>26</td>
</tr>
<tr>
<td>Figure 2.6</td>
<td>Speech control flowchart for robot arm.</td>
<td>27</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Vision systems using one camera.</td>
<td>35</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>Vision systems using 2 cameras.</td>
<td>36</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>Proposed vision system for instrument-passing robot.</td>
<td>39</td>
</tr>
</tbody>
</table>
Figure 3.4 Resolution test pattern with converging lines

Figure 3.5 RCA P200 test pattern.

Figure 3.6 Test set-up for RCA P200 test.

Figure 3.7 MTF of Javelin and Dage.

Figure 3.8 Digitizer evaluation set-up.

Figure 3.9 Image digitized by IP-512.

Figure 3.10 Image digitized by Oculus 100.

Figure 4.1 Pattern recognition system block diagram

Figure 4.2 Layout of proposed instrument tray.

Figure 4.3 Typical low resolution image.

Figure 4.4 Distribution of low resolution features.

Figure 4.5 Definition of Pose 1 and Pose 2.

Figure 4.6 Determination of handle width.

Figure 4.7 Distribution of lengths of Acufex instruments and cartilage knives.

Figure 4.8 Distribution of handle widths of Acufex instruments and cartilage knives.
Figure 4.9 Determination of pose for Acufex instrument and cartilage knives.

Figure 4.10 Typical high resolution image.

Figure 4.11 Distribution of widths and angles for obturators.

Figure 4.12 Distribution of width of cartilage knives: (top) all knives; (bottom) FCK, HP, SCK, MCK.

Figure 4.13 Curvature of Acufex tip: large grasping forceps.

Figure 4.14 Curvature of Acufex tip: left plain scissors.

Figure 4.15 Curvature of Acufex tip: right plain scissors.

Figure 4.16 Curvature of Acufex tip: left 20 degree hooked scissors.

Figure 4.17 Curvature of Acufex tip: right 20 degree hooked scissors.

Figure 4.18 Curvature of Acufex tip: left 60 degree hooked scissors.

Figure 4.19 Curvature of Acufex tip: right 60 degree hooked scissors.

Figure 4.20 Curvature of Acufex tip: left serrated scissors.

Figure 4.21 Curvature of Acufex tip: right serrated scissors.

Figure 4.22 Curvature of Acufex tip: left basket punch.

Figure 4.23 Curvature of Acufex tip: right basket punch.

Figure 4.24 Noisy high resolution image with intrusions.
Figure 4.25 High resolution structural features.

Figure 4.26 High resolution BMD features

Figure 5.1 Overall recognition flowchart.

Figure 5.2 Histogram of goodness of structural match.

Figure 5.3 Noisy image with histogram of intensity.

Figure 5.4 Binary digitized noisy image: (top) threshold = 55; (bottom) threshold = 65.

Figure 5.5 Binary digitized noisy image: (top) threshold = 45; (bottom) threshold = 35.

Figure 5.6 Noise test of fluids in low resolution.

Figure 5.7 Noise test of fluids in high resolution.

Figure 5.8 Sample instrument tray.

Figure 5.9 Boundary test in low resolution.

Figure 5.10 Boundary test in high resolution.

Figure 5.11 UBC VAX set-up.
ACKNOWLEDGEMENT

I would like to thank Dr. Lawrence and Dr. McEwen for their helpful suggestions throughout the course of this thesis. I would especially like to thank Mr. Neil Cox for much timely advice on completing a thesis, without which this thesis would not have been completed.

I would also like to thank the following people and groups for their kind advice and assistance: Messrs. G. Auchinleck, R. Bohl, S. Chan, J. Clark, J. Ens, B. Fung, H. Garudadri, W. Jager, R. McNeil, C. Osborne, K. Yip, the technical and clerical staff of the Biomedical Engineering Department at Vancouver General Hospital, and the surgical nursing and equipment cleaning staff at Surgical Day Care Centre. Finally, I would like to give a special thanks to my sister, Mabel, for her expert typing and proof-reading of this thesis, and to my family for their support throughout this thesis.

I am very grateful to the B.C. Science Council for their support of my work in the form of GREAT awards, and to Andronic Devices Ltd. for supplying some of the equipment for this project.
CHAPTER 1

INTRODUCTION

1.1. Statement of Problem

The high cost of health care delivery has led to significant reduction in staff and, possibly, reduction in services performed in some hospitals. A large portion, as much as 80%, of the hospital's operating budget is devoted to labour costs. Hence, it is essential for hospitals to control the ever increasing labour cost in order to maintain the quality of health care provided (McEwen, 1984).

One method commonly employed in manufacturing to reduce labour costs and increase productivity is through automation, particularly robotics. A robot, as defined by the Robotics Industry Association, is a "reprogrammable multifunctional manipulator designed to move materials, parts, tools or specialized devices through variable programmed motions for the performance of a variety of tasks." Thus, robots are extremely useful for performing repetitive, well defined tasks, or working in hazardous or hostile environments.

Robotic devices have been used extensively in rehabilitation engineering in efforts to improve the quality of life for disabled people and to reduce their dependence on health care workers (Leifer, 1981). Robotics have not yet been employed extensively in acute care hospitals; however, more widespread usage can be expected in the future not only to control the costs of health care, but also to improve the quality of such care. For example, a few robotic devices are being used in clinical chemistry laboratories to handle hazardous samples or reagents and to reduce
the exposure of infectious material to the laboratory workers (McEwen, 1984).

As another example in the operating room, a robot has been used in stereotactic surgery (Kwok, 1985). The robot is used to align the trajectory of a probe to be inserted into the patient's skull; the surgeon then manually directs the arm to move towards a specified target within the brain while making use of computer tomography (CT) images. The use of the robot allows the probe to move in a more precisely controlled path than a surgeon would have been able to do.

In this thesis, a robotic system with visual sensing capability is proposed to perform non time-critical duties in the operating room (OR). At present, two nurses, a circulating nurse and a scrub nurse, are employed in the OR for each surgical case. The circulating nurse, working in the non-sterile areas of the OR, sets up the surgical equipment prior to surgery, fetches and retrieves material for the other clinical personnel during surgery, and removes the surgical equipment after the conclusion of surgery. The scrub nurse, working in the sterile areas, sets up the sterile equipment prior to surgery, passes the sterile instruments to the surgeon, and removes the surgical equipment after the conclusion of the surgery. Since the circulating nurse is highly mobile during surgery, a robot would not be able to emulate her activities. On the other hand, a scrub nurse spends a considerable amount of her time passing instruments to the surgeon. Hence, these repetitive instrument-passing duties can be performed by an instrument-passing robot, resulting in labour savings. Other advantages are the reduced risk of infection and consistency of robot workers.

Any robotic system for such an application should be able to pass the instruments to the surgeon without compromising the patient's treatment, sterility in the surgical field and safety of the OR personnel. In particular, a vision system is to be developed for recognizing the surgical instruments. The vision system should be able to identify each instrument and determine its location and orientation.
The feasibility of the instrument-passing robot depends on the complexity of the surgical procedure. For cardiac surgery where many surgical instruments are involved, it does not yet seem plausible for a robot to sort through all of these instruments with the required speed. On the other hand, in arthroscopic surgery, only 36 instruments are used and the procedure is less time-critical. Hence, the instrument-passing robot can be used in arthroscopic surgical procedures.

1.2. Objectives

One difficult problem with implementing the instrument-passing robot is the lack of a commercially available vision system that is capable of recognizing the surgical instruments. A survey of the commercial vision systems at the 1984 "Robots 8" exhibition in Detroit revealed that most of the commercial vision system were not suitable for this application. Of the systems examined, the closest to being suitable for this application were the EYE (Analog Devices, Norwood, MA), and the IRRIS–100 (LNK Corp., Silver Spring, MD). However, these system were either too expensive, over $50,000, or did not have the sufficient resolution. Hence, a custom vision system capable of recognizing the surgical instruments will be developed in this thesis.

In any computer vision system, the image pattern recognition problem usually requires extensive development. From Figures 1.1 and 1.2 showing the complete arthroscopic instrument set, it can be seen that some of the instruments are quite similar except for the tips. Figures 1.3 and 1.4 give a close-up view of the tips. The subtle differences in the instruments precludes the use of most of the simple statistical image pattern recognition techniques and require detailed examination of local properties of the instruments.

The objectives of this thesis are:

1. Primary Objectives:
a. to develop and evaluate a real-time, clinically acceptable vision system for recognizing arthroscopic instruments.

2. Secondary Objectives
   a. to investigate the technical problems associated with using such a vision system as part of an instrument-passing robot.
   b. to make recommendations for system components for an instrument-passing robot.
   c. to examine the economic aspects of the instrument-passing robot.

The contributions of this thesis are in developing computer algorithms capable of discriminating the arthroscopic instruments in real-time on a microcomputer, outlining the problems associated with implementing an operating room instrument-passing robot, and examining the economic aspects of implementing a robot in the operating room.
Figure 1.1 Cartilage knives (top) and Acufex instruments (bottom)
Figure 1.2 Obturators (top) and low resolution instruments (bottom)
Figure 1.3 Tips of cartilage knives (top) and 3 Acufex instruments: LGF, LPS, RPS (bottom)
Figure 1.4 Tips of Acufex instruments: L20H, R20H, L60H, R60H (top); LSS, RSS, LBP, RBP (bottom)
CHAPTER 2

SYSTEM OVERVIEW

The primary function of the instrument-passing robot is to replace those functions of a scrub nurse which involve handling the surgical instruments between the sterile instrument area and the surgeon. Although the duties of a human scrub nurse involves considerably more than instrument handling, the present technology limitations in robotics prevent the inclusion of other mobile and decision-making tasks of a scrub nurse. Nevertheless, if a robot is capable of successfully replacing a human operator in a limited capacity, an overall cost saving will result.

The requirement of an instrument-passing robot are considerably more complicated than those of an industrial robot. In addition to having sufficient reach, speed and accuracy, the surgical robot must also be safe, quiet, sterilizable (if necessary), and unobtrusive. This chapter presents the layout of the overall instrument-passing robot.

2.1. Block Diagram of System

The basic components required to perform the instrument passing duties of a scrub nurse are: a computer for overall control and data processing, a speech recognition and synthesis unit for receiving and acknowledging commands from the surgeon, a vision system for recognizing the surgical instruments, and a robot arm to perform the physical "pick and place" operations. A typical system configuration is shown in Figure 2.1
Figure 2.1 Typical robotic system configuration.
To keep the cost low and to limit the number of system components, the computer serves both as the system controller and as the processor for the vision system. The vision system consists of one or more cameras and the appropriate lighting unit. The robot arm consists of a mechanical arm, with 5 or 6 degrees of freedom, and an associated controller. The speech recognition and synthesis units may be one or two units and should be controlled by the computer.

2.2. Description of Components

2.2.1 Computer

The computer for the instrument-passing robot will have the tasks of controlling the robot, responding to the speech recognition and synthesis units, and processing the vision system data. The computer should be fast enough to process the vision system data quickly and should have sufficient memory for image data, vision system program and control program storage. The computer should also have good input/output capabilities for communicating with the robot and the speech recognition/synthesis unit, as well as a good selection of software for program development.

The IBM Personal Computer (PC) (IBM Corp, Armonk, NY) is a popular computer which is suitable for instrument-passing robot. The IBM PC is available in several models, the PC or PC/XT microcomputer or the PC/AT supermicrocomputer. The PC and PC/XT use the Intel 8088 8/16 bit CPU which runs at 4.77 MHz and is capable of addressing 1 Mb of memory while the PC/AT uses the Intel 80286 16 bit CPU which runs at 6 MHz and is capable of addressing 16 Mb of physical memory. Both PC models have 8 expansions slots for communicating with external devices. Other advantages of the IBM PC are extensive graphics capabilities and extensive software availability.
The cost of the IBM PC and PC/AT are listed in Table 2.1. The basic PC and the advanced PC/AT cost approximately $3,500 and $5,000 respectively. Although the PC/AT is slightly more costly than the PC, it can execute programs 4 to 6 times faster than the PC – an execution speed which is very desirable for the vision system.

2.2.2 Vision System

A vision system is used with the instrument-passing robot to identify the different instruments, and to locate the position and orientation of the instruments. The vision system also allows the instruments to be placed in any suitable position within the visual field, keeps track of which instruments are present, and can recalculate the positions of instruments in case they become displaced. Additional safety measures can be incorporated into the vision system to inactivate the robot arm in the event of someone reaching into the visual field to retrieve an instrument or to perform other unforeseen actions.

The advantage of a vision system is the unconstrained handling of the instruments. Without a vision system, the instruments must be palletized, and the exact positions and orientations of the instruments must be known; otherwise, the robot arm would not be able to grasp the instruments. As well, a robotic system without visual sensing cannot keep track of the instruments easily if more than one instrument is removed from the field. Hence, the additional cost and complexity of using a vision system can be justified by the improved flexibility, reliability and safety.

To decrease the complexity of the system, two constraints are applied to the vision system: the images are digitized to binary (1 bit) levels and the instruments are non-touching and non-overlapping. Good binary images can be generated by using appropriately structured lighting. To prevent the instruments from touching and
Table 2.1 Cost of IBM PC.

<table>
<thead>
<tr>
<th>Description</th>
<th>IBM PC</th>
<th>IBM PC/AT</th>
</tr>
</thead>
<tbody>
<tr>
<td>computer with 2 disk drives, 256Kb</td>
<td>$2843</td>
<td>$4414</td>
</tr>
<tr>
<td>serial and parallel ports</td>
<td>$200</td>
<td>$200</td>
</tr>
<tr>
<td>monochrome monitor</td>
<td>$200</td>
<td>$150</td>
</tr>
<tr>
<td>monochrome display card</td>
<td>$150</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>$3393</strong></td>
<td><strong>$4964</strong></td>
</tr>
</tbody>
</table>
overlapping, a compartmentalized tray was developed to hold the instruments. The structured lighting and the instrument tray will be discussed in more detail in a later section.

A major concern with the vision system is the preservation of a sterile field around the sterile surgical instruments. Even though the vision system will not be in direct contact with the sterile instruments, there may be concern that any overhead lighting or camera equipment may deposit dirt onto the sterile field. The appropriate action to prevent dirt from entering the sterile area, after consultation with surgical nursing personnel, is to clean the overhead equipment prior to the beginning of the surgical procedure.

2.2.3 Speech Recognition and Synthesis System

The speech recognition and synthesis system is used by the computer to communicate with the surgeon. The speech recognition system receives the commands from the surgeon and the speech synthesis system echoes the received commands for verification purposes. It is important to have a high performance speech recognition system for fast and accurate response to the surgeon’s request. A speech recognition system which cannot recognize or mis-recognizes commands will prolong the duration of the surgical procedure and annoy the OR staff. The only criterion for the speech synthesis system is that it produces clear, easily recognizable speech sounds.

The different types of speech recognition systems may be classified as: discrete or connect speech recognizers; and speaker-dependent or speaker-independent recognizers. The majority of the speech recognition systems available are for discrete speaker-dependent speech. There are connected speech recognition systems available but they are usually very expensive, e.g. the NEC DP-200 which costs over $10,000. Although some systems claim to be speaker independent, the vocabulary for recognition
may be limited only to a few distinct sounds; the general topic of speaker-independent speech recognition is still very much in the research stage (Flanagan, 1982).

Of the speech recognition systems available, the NEC SR100 seems to have the best performance for the price range, $2,000. It has specifications of over 99% recognition accuracy, even in the presence of background noise. The recognition algorithm employed by the SR100 uses the powerful dynamic programming technique to align the input speech signals to match the reference speech signals. Tests of the SR100 were performed at the Electrical Engineering Department of UBC (Frenette, 1985). The best recognition accuracy obtained was 98.5% for subjects with little training and a noiseless background. Another factor which can affect the recognition accuracy is the vocabulary of the application.

For speech synthesis, the NEC AR100 can be used as a companion to the SR100 since both units have similar input/output protocols. The AR100 uses ADPCM techniques to store the speech patterns and costs $2,000.

2.2.4 Robot Arm

A robot arm is used to perform the instrument-passing duties of a scrub nurse. The arm should have a sufficient payload to handle all the surgical instruments with ease, have sufficient reach and accuracy to access all the surgical instruments, and have sufficient speed to perform the passing, or 'pick and place', operations quickly and reliably. Other requirements of the robot arm are that it should have: a convenient power source for the operation room, proper geometry for the type of movements required, cleanable or sterilizable external components, and low cost.

The most important factor in selecting the robot arm for this application is perhaps cost. At present, there exist industrial robot arms that can satisfy all of the
above requirements; however, these industrial robot arms cost between $50,000 and $100,000. The lower-cost robot arms are generally smaller, less accurate and have a lighter payload. Of all the common geometries shown in Figure 2.2, the revolute robot seems to be the most appropriate, since it can move over or around obstacles easily in a smooth path. The cost convenient power source for the operating robot is electric, although a 50 psi pneumatic line is available in most operating rooms. The payload of the robot arm is not a major concern in this application since all the instruments weigh less than 150g and most robot arms have a payload greater than 500g.

At present, there are at least two educational-commercial robot arms with specifications that are suited to the instrument-passing robot. They are the $12,000 Robotic Systems International RT3 and the $14,000 Mitsubishi RM501. Both robot arms use electric power and have the revolute geometry. The RT3 has 6 degrees of freedom (DOF) and a 60 cm reach while the RM501 has 5 dof and a 38 cm reach. Although the RT3 has an extra DOF and a longer reach over the RM501, it is less stable, less accurate, and has a higher drift. Experiments performed by Andronic Devices Ltd. (Andronic), a company performing research into medical and surgical robotics at the Vancouver General Hospital (VGH) and UBC Health Science Center Hospital (HSCH), have shown that the reach of the RM501 can be extended slightly by adding a fixed link between the gripper and the robot wrist without decreasing the performance of the robot arm.

Since the robot arm will be working near to the sterile field, sterility is an important consideration. Proper 'draping' of the robot include covering all surfaces that are adjacent to sterile areas and covering the actual arm since it passes over sterile areas. Sterile 'draping' techniques for the RM501 have been developed by Andronic at VGH and successfully tested in an operating room at the HSCH.
Figure 2.2 Common geometries for robot arms.
2.3. Gripper - End Effector

The gripper, or end effector, is an important part of the instrument-passing robot. The gripper must be able to firmly grip the instrument, while stationary and in motion, and release the instrument on the command of the surgeon, either via sensors on the gripper or electrical signals from the computer and the speech recognition unit. In addition, the gripper should be detachable from the robot arm and be sterilizable, and it should be safe to touch without compromising sterility, such as puncturing the surgical glove.

An important design consideration for the gripper is the exchange in handing off the instrument to the surgeon. Ideally, the gripper should release the instrument after the surgeon has obtained a firm grasp; it is important to avoid dropping the instrument since it would cause a delay as the instrument is replaced, as well as damaging the instrument.

Two successful designs have been tested by Andronic at VGH. One design uses an electrically driven gripper with a membrane switch to activate the release action while the other design is pneumatically driven with a micro switch release activation (Fengler, 1984).

2.4. Surgical Instrument Set

The arthroscopic instruments used to develop the vision system are listed in Table 2.2. They are part of the standard arthroscopy set and the Acufex set used for arthroscopic procedure at the Surgical Day Care Centre (SDCC) at VGH. Other instruments that are used in arthroscopy but not included are the arthroscope and the Dyonics shaver (Dyonics Surgical Ltd., Andover, MA). These two instruments are normally attached to a long cord and cannot be easily handled by the robot arm. Since they are not passed back and forth between the scrub nurse and the surgeon,
they are usually placed within reach of the surgeon at the beginning of the surgical procedure.

The 32 different instruments to be recognized by the vision system consist of knives, knife holder, scissors, needle holder, towel clips, trocar sleeves, pyramidal trocars, blunt obturators, cartilage knives, hook probe, and Acufex scissors, punches and graspers. The Acufex scissors and punches contain pairs that differ only in the rotation direction of the cutting edge, clockwise (CW) and counter-clockwise (CCW). The trocar sleeves, pyramidal trocars and blunt obturators are available in 2 sizes, 5.0 mm and 3.8 mm.

In diagnostic arthroscopy, the scalpel is used to make initial incision in the patient's knee, after which only the cartilage knives, arthroscope and one set of the trocar sleeve, pyramidal trocar, and blunt obturator are used extensively in most procedures. The scissors are used at the end of the procedure to cut away the disposable drape used on the patient's knee. In surgical arthroscopies, the Acufex instruments are increasingly being used to resurface the patient's knee and hence are used extensively in the middle portion of many arthroscopic procedures. If a large amount of resurfacing is required, the Dyonics shaver is often used instead of the Acufex instruments.

2.5. Structured Lighting and Instrument Tray

Structured lighting techniques are used with the vision system to produce high-quality binary images. The structured lighting involves generating backlighting for the instruments to increase the contrast between the background and the instruments. The backlighting technique is the most common for generating binary images, although other methods exist (Schroeder, 1984). In most backlit binary vision systems, the work environment is usually darkened for best image contrast; however, since the lighting in
Table 2.2 List of arthroscopy instruments.

<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol</th>
<th>Width</th>
<th>Length</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large Towel Clip</td>
<td>LTC</td>
<td>7.9 cm</td>
<td>13.5 cm</td>
<td>Low Res</td>
</tr>
<tr>
<td>Small Towel Clip</td>
<td>STC</td>
<td>6.2 cm</td>
<td>8.8 cm</td>
<td>Low Res</td>
</tr>
<tr>
<td>Trocar Sleeve 5.0mm</td>
<td>TS50</td>
<td>5.5 cm</td>
<td>18.5 cm</td>
<td>Low Res</td>
</tr>
<tr>
<td>Trocar Sleeve 3.8mm</td>
<td>TS38</td>
<td>5.5 cm</td>
<td>19.3 cm</td>
<td>Low Res</td>
</tr>
<tr>
<td>Drainage Cannula</td>
<td>TS28</td>
<td>2.8 cm</td>
<td>10.7 cm</td>
<td>Low Res</td>
</tr>
<tr>
<td>Cannula Trocar</td>
<td>PT28</td>
<td>1.0 cm</td>
<td>11.7 cm</td>
<td>Low Res</td>
</tr>
<tr>
<td>Needle Holder</td>
<td>NH</td>
<td>7.5 cm</td>
<td>15.5 cm</td>
<td>Low Res</td>
</tr>
<tr>
<td>Dissecting Scissors</td>
<td>SCI</td>
<td>5.5 cm</td>
<td>17.0 cm</td>
<td>Low Res</td>
</tr>
<tr>
<td>Tweezers</td>
<td>TW</td>
<td>1.0 cm</td>
<td>15.0 cm</td>
<td>Low Res</td>
</tr>
<tr>
<td>Scalpel</td>
<td>KF</td>
<td>1.3 cm</td>
<td>14.2 cm</td>
<td>Low Res</td>
</tr>
<tr>
<td>Long Knife</td>
<td>LK</td>
<td>2.0 cm</td>
<td>20.3 cm</td>
<td>Low Res</td>
</tr>
<tr>
<td>Long Knife Handle</td>
<td>RH</td>
<td>2.5 cm</td>
<td>23.0 cm</td>
<td>Low Res</td>
</tr>
<tr>
<td>Blunt Obturator 3.8mm</td>
<td>BLOB3</td>
<td>1.8 cm</td>
<td>20.8 cm</td>
<td>Obturator</td>
</tr>
<tr>
<td>Pyramidal Trocar 3.8mm</td>
<td>PYTR3</td>
<td>2.5 cm</td>
<td>20.8 cm</td>
<td>Obturator</td>
</tr>
<tr>
<td>Blunt Obturator 5.0mm</td>
<td>BLOB5</td>
<td>2.7 cm</td>
<td>19.5 cm</td>
<td>Obturator</td>
</tr>
<tr>
<td>Pyramidal Trocar 5.0mm</td>
<td>PYTR5</td>
<td>2.8 cm</td>
<td>20.5 cm</td>
<td>Obturator</td>
</tr>
<tr>
<td>Meniscotome</td>
<td>FCK</td>
<td>1.9 cm</td>
<td>24.5 cm</td>
<td>Cart Knife</td>
</tr>
<tr>
<td>Hook Probe</td>
<td>HP</td>
<td>1.9 cm</td>
<td>23.5 cm</td>
<td>Cart Knife</td>
</tr>
<tr>
<td>Small Meniscotome</td>
<td>SCK</td>
<td>2.0 cm</td>
<td>23.5 cm</td>
<td>Cart Knife</td>
</tr>
<tr>
<td>Medium Meniscotome</td>
<td>MCK</td>
<td>2.3 cm</td>
<td>24.2 cm</td>
<td>Cart Knife</td>
</tr>
<tr>
<td>Large Meniscotome</td>
<td>LCK</td>
<td>2.0 cm</td>
<td>23.7 cm</td>
<td>Cart Knife</td>
</tr>
<tr>
<td>Left Basket Punch</td>
<td>LBP</td>
<td>2.3 cm</td>
<td>22.0 cm</td>
<td>Reg Acufex</td>
</tr>
<tr>
<td>Left Serrated Scissors</td>
<td>LSS</td>
<td>2.5 cm</td>
<td>22.5 cm</td>
<td>Reg Acufex</td>
</tr>
<tr>
<td>Left 20 Deg Hooked Sci</td>
<td>L2OH</td>
<td>2.7 cm</td>
<td>21.8 cm</td>
<td>Reg Acufex</td>
</tr>
<tr>
<td>Left 60 Deg Hooked Sci</td>
<td>L6OH</td>
<td>2.5 cm</td>
<td>21.5 cm</td>
<td>Reg Acufex</td>
</tr>
<tr>
<td>Right Basket Punch</td>
<td>RBP</td>
<td>2.3 cm</td>
<td>22.0 cm</td>
<td>Reg Acufex</td>
</tr>
<tr>
<td>Right Serrated Scissors</td>
<td>RSS</td>
<td>2.5 cm</td>
<td>22.5 cm</td>
<td>Reg Acufex</td>
</tr>
<tr>
<td>Right 20 Deg Hooked Sci</td>
<td>R2OH</td>
<td>2.7 cm</td>
<td>21.8 cm</td>
<td>Reg Acufex</td>
</tr>
<tr>
<td>Right 60 Deg Hooked Sci</td>
<td>R6OH</td>
<td>2.5 cm</td>
<td>21.5 cm</td>
<td>Reg Acufex</td>
</tr>
<tr>
<td>Left Plain Scissors</td>
<td>LPS</td>
<td>3.3 cm</td>
<td>24.5 cm</td>
<td>Long Acufex</td>
</tr>
<tr>
<td>Left Grasping Forceps</td>
<td>LGF</td>
<td>3.7 cm</td>
<td>25.5 cm</td>
<td>Long Acufex</td>
</tr>
<tr>
<td>Right Plain Scissors</td>
<td>RPS</td>
<td>3.3 cm</td>
<td>24.5 cm</td>
<td>Long Acufex</td>
</tr>
</tbody>
</table>
the operating room could not be dimmed, the backlighting must be of sufficient intensity to provide adequate contrast. Also, it is important to have uniform backlight intensity for binary images.

For the purpose of obtaining test images for the vision system, a simple backlight unit was built (see Figure 2.3). Four 24-inch fluorescent tubes were used to provide the lighting. Several sheets of frosted mylar were used to diffuse the light to obtain even intensity, with the amount of diffusion controlled by the number of mylar sheets used.

To ensure that the instruments do not touch or overlap, as required by the vision system, an instrument tray with compartments to separate the instruments is required. The instrument tray should be transparent to allow the backlighting to pass through and should be made of a sterilizable material which does not lose its clarity with use.

Each compartment should be designed to be piecewise linear and should be slightly larger than the length and width of the instrument. Whenever possible, the compartments should be suitable for several different instrument of the same type, allowing for different arrangements of the instruments.

2.6. System Integration

2.6.1 Operation Sequence

In designing robotic systems, it is helpful to know the operational sequence required for the process to help in defining the specifications of components. This section discusses several proposed operational sequences for the instrument-passing robot. Although implementation of the operational sequences were considered beyond the scope of the thesis, an understanding of their operations would be useful in defining
Figure 2.3 Drawing of backlight unit.
performance specifications for the instrument-passing robot.

There are two main operational sequences involved in the instrument-passing robot: initialization and control. Initialization sequence is performed during set-up to ensure that the proper programs and data files are present and that the instruments match the data file specified. The control sequence then takes over and operates the vision system, the speech recognition and synthesis units, and the robot arm.

In initialization, the vision system checks the compartments for the correct instruments. After checking all the instruments, the vision system displays a list of instruments according to the compartments and waits for verification by the circulating nurse. If the vision system is unable to recognize an instrument or has misclassified an instrument, the circulating nurse enters the correct instrument from the keyboard; therefore, when initialization is done, the vision system should have a correct map of the instruments in the visual field.

If time permits, the initialization program will request a voice training with the surgeon, which requires 2 to 3 minutes; otherwise, the data file of the voice of the particular surgeon is retrieved and used. A flow chart of the initialization sequence is given in Figure 2.4.

In the main control sequence, the vision system is continually executed to update the position and orientation information on the instruments. To ensure that nothing is altered while the vision system programs are being executed, an interrupt-driven image boundary-testing subroutine is executed. If a string of non-background pixels are detected at the boundary, the vision system generates an alarm until the boundary is clear again; otherwise, the vision system programs are continued.
Figure 2.4 Pre-operation initialization flowchart.
In the event that the vision program changes the classification of an instrument without any intrusion by the robot or an alarm due to the image boundary test, then the new classification is rejected and the old classification is retained. This permits the programs to continue to operate even in the presence of artifacts. Assuming that the map of the instruments is correct at initialization, then the system will function normally if the instruments which changed classification are not requested, since the initial, correct classification is retained. If, however, an instrument which changed classification has been requested and returned, then the vision system would be unable to classify or misclassify the instrument; as a result, the vision system would request assistance from the circulating nurse. Alternatively, since the vision system can keep a record of which instruments are being used currently, the returned instrument may be classified at a reduced confidence level, only if the result is consistent with the accounting of all instruments. The vision program flow chart is shown in Figure 2.5.

Operation of the robot arm is only required after an input from the speech recognition unit or a signal for instrument retrieval. The speech recognition unit, on receiving a command, interrupts the processing and stores the command in a command list. The command processor then interprets and executes the commands in a sequential manner; special commands requiring immediate action may be stored in another list for urgent commands and executed quickly. When all commands have been executed, the control is returned to the vision programs. The proposed speech control flow charts for the robot arm are shown in Figure 2.6.

2.6.2 Role of Surgeon and Nurse in O.R.

In order to effectively implement the instrument-passing robot in an operating room, the interactions between the robot arm and the OR staff must be clearly defined to avoid any accidents. To find out the necessary interactions between the robot and the OR staff, a questionnaire (see appendix 1) was circulated to experienced
Figure 2.5 Vision system flowchart.
Figure 2.6 Speech control flowchart for robot arm.
surgical nursing staff to inquire of the type of assistance available at different stages of arthroscopic procedures.

The questionnaire results indicate that the circulating nurse and scrub nurse would be available to set up the robotic system prior to the surgical procedure. The circulating nurse, when she is not required elsewhere, is available to assist the robotic system during the procedure and to shut down the robotic system at the end of the procedure. However, the surgeon is generally not available to assist the robotic system. Therefore, the circulating nurses should be trained to respond to any alarms generated by the robotic system.

2.7. Safety

Safety is one of the most important criteria in the successful implementation of a robot. Any robot should be programmed to avoid causing any harm to any human being through its actions and inactions (Thring, 1983). Safety is especially important in an instrument-passing robot since it is designed to work closely with the surgeon and to handle some potentially hazardous instruments, such as scalpels. Special attention must be given to how the robot arm picks up an instrument, what path is selected in transporting the instrument, and which end of the instrument is presented to the surgeon during handoff.

The structuring of the work area of the robot with the work area of the surgeon should be done such that they do not overlap. Additional devices should be placed around the robot arm such that anyone entering the work area of the robot would inactivate the robot arm and raise an audible alarm. However, these safety alarms should not impede the movement of the surgical staff in the operating room, or interfere with the operation of the other clinical equipment. Hence, a careful study of the equipment layout and clinical staff work areas must be done before the
instrument-passing robot may be introduced into the operating room.

2.7.1 Payback Period and Cost Effective Analysis

In order to assess the cost-effectiveness of an instrument passing robot, it is necessary to know the cost of the total robot system, the salary of a scrub nurse, and the time savings involved. The total cost of the system, as outlined in the previous sections, and the salary of a scrub nurse are given in Table 2.3.

In industrial robotics, it is generally accepted that feasible applications for robotics should have a 1-3 years payback period, i.e. savings arising from the introduction of a robot should equal the capital cost of the robot within two years. A similar criterion might be employed in health care.

To determine the labour savings involved, consider a typical diagnostic arthroscopic procedure of 20-minute duration. Initial set up requires approximately 10 minutes and clean up requires approximately 5 minutes. Thus, the total time that a scrub nurse spends in the OR is approximately 35 minutes. By introducing an instrument-passing robot, only the 20 minutes during the procedure would the robot be useful; hence the cost savings is calculated as 57% (see Appendix B). Using the system cost and the scrub nurse's salary, the payback time is approximately 1.6 years, if the robotic system could completely replace all of the functions of a scrub nurse. Since this may not be true, the payback period should be pro-rated accordingly.
Table 2.3 Proposed system cost and scrub Nurse salary.

System Cost for Instrument Passing Robot

<table>
<thead>
<tr>
<th>Item</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>computer IBM PC/AT</td>
<td>$5,000</td>
</tr>
<tr>
<td>Robot Mitsubishi RM501</td>
<td>$14,000</td>
</tr>
<tr>
<td>Speech Recognition and Synthesis System</td>
<td>$4,000</td>
</tr>
<tr>
<td>NEC SR100/AR100</td>
<td></td>
</tr>
<tr>
<td>Vision System</td>
<td></td>
</tr>
<tr>
<td>Camera Javelin Je2062</td>
<td>$4,000</td>
</tr>
<tr>
<td>Digitizer Oculus 100</td>
<td>$1,000</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>$1,000</td>
</tr>
<tr>
<td><strong>Total System Cost</strong></td>
<td>$29,000</td>
</tr>
</tbody>
</table>

Labour Cost

Salary of a Scrub Nurse (at VGH, August 1985)

<table>
<thead>
<tr>
<th>Description</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Payrate per hour</td>
<td>$12.84 - $14.85</td>
</tr>
<tr>
<td>Including 18% benefit</td>
<td>$15.15 - $17.52</td>
</tr>
</tbody>
</table>

Average total salary including benefits $16.34

Work week = 37.5 hours per week
CHAPTER 3
VISION SYSTEM

3.1. Imaging Components Review

3.1.1 High-Resolution Imaging Systems

Before embarking on a discussion of high-resolution imaging systems, it is appropriate to first explain what is meant by "resolution." The resolution of an imaging system refers to the system's ability to separate fine details and is usually described by the system modulation transfer function - the spatial frequency equivalent of the time domain frequency transfer function. It is commonly measured using a standard test chart consisting of converging black and white lines. The "limiting resolution" is defined as the resolution at which the response, as measured by the test chart, has fallen to 3% of the maximum, which corresponds to the minimum contrast resolvable by the human eye. The unit for resolution is number of lines per picture height.

It is unclear how to relate system specifications to the resolution requirements. The "limiting resolution" may not be suitable since a computer may not be able to detect an object if the response is only 3% of the maximum. Nevertheless, the resolution figures may be used to compare vision systems.

The limitations in resolution for a particular imaging device is usually due to the image sensor. There are two general types of image sensors: camera tubes and solid state sensors. The term "camera tube" commonly refers to a vacuum tube device with a photoconductive target which can convert an image into electrical signals, an
example being the vidicon. The resolution limitations of camera tubes have been discussed in detail by Cope et al. (Cope et al, 1971). The camera tube resolution generally depends on the construction of the tube and the photosensitive target used. Solid state image sensors consist of discrete photosensitive sites and the resolution is determined by the number of these discrete sites. A review on image sensors has recently been published (Flory, 1985).

At present, the highest image resolution achievable is approximately 10,000 lines, using a return-beam vidicon (RBV) (Cantella, 1971). Alternatively, a high-resolution system marketed by Eikonix Inc. can scan at variable resolution up to 4096 x 5200 pixels. It uses a solid state linear imaging sensor to sense one line of the image; the sensor is then mechanically scanned perpendicular to the line image to produce a full two-dimensional image. However, the high-resolution systems noted above are generally slow and cannot be used for real-time applications (i.e. 30 frames/s).

Real-time high-resolution imaging research has intensified recently in anticipation of the introduction of "High Definition Television" (HDTV) system. One result of these research is the development of the saticon high-resolution camera tube (Isozaki, 1978; Isozaki et al, 1981). The saticon tube uses a selenium–tellurium–arsenic photoconductor target, which has very good resolving-power characteristics. The limiting resolution of the saticon is reported to be over 1600 lines for a 1-inch diameter target. In comparison, the limiting resolutions of 1-inch antimony trisulfide vidicons and 1-inch lead–oxide plumbicons are approximately 1100 and 900 lines, respectively. For camera tube sensors, the limiting resolution generally increases with the diameter of the tube.

The resolution of solid state image sensors depends on the number of discrete photosensitive sites. Currently, the highest density sensor available is an 800 x 800 CCD area array imager, at a cost of $20,000 (Frank, 1985). Recent improvements in
solid state device fabrication techniques will enable even larger array areas to be built; several research laboratories have reported development of 1024 x 1024 pixels resolution devices (Smith, 1985; Frank, 1985).

At present, commercial high-resolution imaging technology is limited to be at 512 x 512 pixels resolution, since high quality, inexpensive cameras and digitizers are readily available. The cost for 1024 x 1024 pixel imaging systems are quite high, $20,000 or over. However, with the anticipated introduction of HDTV in the future high-resolution cameras, capable of 1024 x 1024 pixel resolution, will soon be available at reasonably low cost, using either saticon or solid state sensors. As well, high-speed digitizing and computing equipment is becoming available to make 1024 x 1024 imaging economically feasible in the near future.

3.1.2 Vision System Design

There are many different approaches to assembling a robotic vision system, depending on the number of cameras used and the resolution requirement. Most commercial systems use only one camera to cover the entire field of view (FOV). The resolution of such a camera system would depend on the resolution of the camera. Special accessories, such as variable zoom lenses, and pan-and-tilt units, allow the one-camera system to extend beyond the resolution limit of the camera. The single or multi-camera approaches are described in this section and a special configuration is proposed for the instrument-passing robot.

Several one camera system configurations are shown in Figure 3.1. The standard fixed zoom camera system is shown in Figure 3.1(a). This system is simple but has the disadvantage that the resolution is limited by the resolution of the camera. The standard configuration can be enhanced by adding variable zoom and/or pan and tilt capabilities, Figures 3.1(b) and (c). Variable zoom allows the camera to zoom in on a
small area for detailed examination. The zoom unit may be continuously variable, which requires complicated feedback circuitry, or it may be a simple two-position zoom. By adding pan and tilt capability to the camera, the scanning field is significantly increased beyond the FOV of a fixed camera. If both the variable zoom and pan and tilt capabilities are combined on one camera, a high-resolution system may be assembled using low-cost, low-resolution components. The main disadvantage of the zoom is the complicated software required to make full use of the zoom feature. The pan and tilt system has the major disadvantage of tangent error for extreme angles and complicated position-tracking hardware and software.

Another approach to increasing the apparent resolution of a one-camera system is to mount the camera on the robotic arm, or on another arm or track, Figure 3.1(d) and (e). By keeping the camera sensor perpendicular to the field of view, i.e. there is no tilting of the camera axis, tangent errors can be eliminated. However, the disadvantages are the additional cost of either an arm or a motorized track, and slightly more complicated software. If the camera is mounted on the robot arm, the payload of the arm is decreased as well.

Two-camera or multi-camera systems use the basic concepts of the one-camera system; various approaches using two cameras are shown in Figure 3.2. A common two-camera approach is the dual-resolution approach in Figure 3.2(a). A fixed camera is used to scan the entire FOV at low-resolution while a camera mounted on the robot arm can examine details at high-resolution. The primary disadvantages of this system are: decreased payload of the arm, and slightly more complicated software to deal with the dual resolution of the two cameras. Alternately, two cameras can work side by side to double the resolution achievable with only one camera. However, image alignment at the common boundary of the two cameras could be a difficult problem.
Figure 3.1 Vision systems using one camera.
(a) one camera fixed, one camera on arm

(b) two cameras, side by side

Figure 3.2 Vision systems using 2 cameras.
For the vision system of the instrument-passing robot, the camera resolution would be limited to 512 x 512 pixels since 1024 x 1024 pixels resolution systems were beyond the budget of this project. As well, tests on simulated 1024 x 1024 pixel images, i.e. using 512 x 512 pixels images of one quarter of the desired field of view, did not produce sufficient details on several of the instruments for reliable recognition. From the photographs of the instruments, Figures 1.1 to 1.4, it is evident that a high-resolution image will be required to classify the Acufex scissors and Acufex graspers, and cartilage knives since the only significant differences are at the tip. For this reason, a dual-resolution approach was chosen for the vision system.

The configuration used for this project is similar to Figure 3.2(a). Instead of mounting the high-resolution camera on the robot arm, the high-resolution camera is fixed over a special region. The tips of the instruments requiring high resolution vision are placed in the special region in the FOV of the high-resolution camera; The other instruments are placed in the remaining area in the FOV of the low-resolution camera, as shown in Figure 3.3.

The advantage of this approach is that the payload of the arm is unaffected such that a low-cost arm with a light but sufficient payload could be used. Also, the system software would only be slightly more complicated than a one-camera system since the high-resolution camera is fixed in space and cannot zoom. The limitations of this dual-resolution approach is that the details requiring high-resolution imaging must be able to fit within the FOV of the high-resolution camera. If the FOV of the high-resolution camera is insufficient, then the camera must be mounted on the robot arm to extend the area covered by the high-resolution camera.

3.2. Selection and Evaluation of Camera

3.2.1 Camera Review
Two cameras are required for the dual resolution approach for this project. Although implementation of the camera related hardware was not done as part of this thesis, the planned configuration is as follows. The cameras will be mounted vertically, looking down on the backlit translucent tray of arthroscopic surgical instruments; the camera outputs will be fed into the image digitizer. The desirable characteristics for the cameras are high-resolution, low image retention and lag, low noise, stable electronics and low distortion. Also, the cost of the cameras should be within the given budget of $10,000 for the imaging system.

As mentioned in the previous section, the two types of image sensors are camera tubes and solid state image sensors. Tube-type cameras, for the higher-priced models, generally have superior resolution to the solid state cameras. They may also have the option of selectable scan rates, which is impossible in solid state cameras. However, tube-type cameras are susceptible to stability and distortion problems, even for higher-priced models. Another problem associated with tube-type cameras is shading, the variation in output signal due to the non-uniform response across the photoconductive surface of the camera tube. Shading can usually be compensated by applying a linear or parabolic bias to the output signal to counteract the non-uniformity. The other characteristics such as image retention, lag, and noise depend on the type of photoconductive surface of the tube. The common photoconductive surfaces for camera tubes are: the antimony trisulfide vidicon, newvicon, saticon, and plumbicon. For high background light applications, the newvicon and saticon photoconductors are the most suitable as they are less susceptible to burn-in problems due to prolonged exposure to high-intensity light sources.

There are many advantages to solid state cameras, such as having low image retention and lag, and no shading, as well as being distortion free, sensitive, stable, and light weight. The solid state fabrication process ensures that there is no noticeable
Figure 3.3 Proposed vision system for instrument-passing robot.
distortion since the sensor pattern is etched into silicon and ensures that there is no shading due to non-uniformity of response between pixels. Stability of solid state cameras is much better since solid state circuits use only low voltage, and solid state cameras do not have heating problems like camera tubes. The major disadvantage of solid state cameras is the resolution. Most solid state cameras costing about $2,000 currently have resolutions of less than 300 lines, which is much worse than 600 lines or better resolution of comparably priced tube cameras.

3.2.2 Camera Selection

Table 3.1 list various tube type and solid state cameras, as well as several special industrial cameras. As there are far too many cameras available to be considered here, only the ones which are potentially useful in this project are presented.

Of the cameras listed in Table 3.1, the best resolution available in a real-time camera is the Sierra Scientific 2601 (Sierra Scientific, Mountain View, CA), which uses a 1.5 inch plumbicon sensor and is capable of digitizing to 1024 x 1024 pixels resolution; however, the cost of $26,190 precludes the use of 1024 resolution in this project. Two other cameras, the Cohu 8000 (Cohu Inc., San Diego, CA) and the Ikegami ITC-82 (Ikegami Electronics, Maywood, NJ), have variable line rates which can scan over 1000 lines and limiting resolutions of 1000 lines or more. The cost of these cameras are slightly less than $10,000. Although the specified resolution is over 1000 lines, a demonstration of the Cohu 8000 showed a resolution of only 650 lines at a scan rate of 1125 lines. Therefore, it does not appear possible to purchase a 1024 x 1024 pixels resolution imaging system for under $10,000 at the present time.

For 512 x 512 pixels resolution systems, the most suitable tube type camera is the Dage 68 (Dage-MTI Inc., Michigan City, IN). This camera has 800 lines
Table 3.1 Listing of commercial cameras considered for vision system.

<table>
<thead>
<tr>
<th>Model and Manufacturer</th>
<th>cost</th>
<th>resolution H</th>
<th>resolution V</th>
<th>BW</th>
<th>Signal</th>
<th>Sensor</th>
<th>SNR</th>
<th>line rt.</th>
<th>frame rt.</th>
<th>distort</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAV-16 Scierra Sci.</td>
<td>11,000</td>
<td>512</td>
<td>512</td>
<td>5.0</td>
<td>RS170,330,CCIR</td>
<td>1 inch</td>
<td>50</td>
<td>559</td>
<td>30</td>
<td>1.0% max</td>
</tr>
<tr>
<td>DAV2P-16 Scierra Sci.</td>
<td>15,300</td>
<td>512</td>
<td>512</td>
<td>5.0</td>
<td>RS170,330,CCIR</td>
<td>1 inch</td>
<td>60</td>
<td>559</td>
<td>30</td>
<td>1.0% max</td>
</tr>
<tr>
<td>2601 Scierra Sci.</td>
<td>26,190</td>
<td>1024</td>
<td>1024</td>
<td>5.0</td>
<td>RS170,330,CCIR</td>
<td>1.5 inch</td>
<td>66</td>
<td>1118</td>
<td>7.5</td>
<td>1.0%</td>
</tr>
<tr>
<td>8000 Cohu</td>
<td>9,500</td>
<td>1100</td>
<td>775</td>
<td>32.0</td>
<td>RS170,330,CCIR</td>
<td>1 inch</td>
<td>30</td>
<td>1125</td>
<td>30</td>
<td>1.5%</td>
</tr>
<tr>
<td>ITC-82 Ikegami</td>
<td>12,000</td>
<td>1000</td>
<td>715</td>
<td>30.0</td>
<td>RS170,330</td>
<td>1 inch</td>
<td>40</td>
<td>1023</td>
<td>30</td>
<td>1.0%</td>
</tr>
<tr>
<td>68 Dage</td>
<td>3,564</td>
<td>800</td>
<td>10.0</td>
<td>RS330</td>
<td></td>
<td>1 inch</td>
<td>N/S</td>
<td>525</td>
<td>30</td>
<td>0.5%</td>
</tr>
<tr>
<td>68 Dage</td>
<td>5,056</td>
<td>800</td>
<td>18.0</td>
<td>RS330</td>
<td></td>
<td>1 inch</td>
<td>52</td>
<td>525</td>
<td>30</td>
<td>0.5%</td>
</tr>
</tbody>
</table>

SOLID STATE CAMERAS

<table>
<thead>
<tr>
<th>Model and Manufacturer</th>
<th>cost</th>
<th>resolution H</th>
<th>resolution V</th>
<th>BW</th>
<th>Signal</th>
<th>Sensor</th>
<th>SNR</th>
<th>line rt.</th>
<th>frame rt.</th>
<th>distort</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDR 460 Video Logic</td>
<td>280</td>
<td>350(491)</td>
<td>N/S</td>
<td>RS170</td>
<td></td>
<td>CCD</td>
<td>46</td>
<td>525</td>
<td>30</td>
<td>nil</td>
</tr>
<tr>
<td>AVC-O1 Sony</td>
<td>2,250</td>
<td>350(491)</td>
<td>N/S</td>
<td>RS170</td>
<td></td>
<td>CCD</td>
<td>43</td>
<td>525</td>
<td>30</td>
<td>nil</td>
</tr>
<tr>
<td>JE2062 Javelin</td>
<td>2,000</td>
<td>450(384)</td>
<td>450(485)</td>
<td>N/S</td>
<td>RS170</td>
<td>MOS</td>
<td>43</td>
<td>525</td>
<td>30</td>
<td>nil</td>
</tr>
<tr>
<td>4TN2500A1 GE</td>
<td>4,050</td>
<td>(248)</td>
<td>(244)</td>
<td>N/S</td>
<td>RS170</td>
<td>CID</td>
<td>43</td>
<td>525</td>
<td>30</td>
<td>nil</td>
</tr>
</tbody>
</table>

SPECIAL CAMERAS

<table>
<thead>
<tr>
<th>Model and Manufacturer</th>
<th>cost</th>
<th>resolution H</th>
<th>resolution V</th>
<th>BW</th>
<th>Signal</th>
<th>Sensor</th>
<th>SNR</th>
<th>line rt.</th>
<th>frame rt.</th>
<th>distort</th>
</tr>
</thead>
<tbody>
<tr>
<td>4TN2200A1 GE</td>
<td>1,340</td>
<td>(128)</td>
<td>(128)</td>
<td>N/S</td>
<td></td>
<td>CID</td>
<td>48</td>
<td>N/A</td>
<td>N/A</td>
<td>nil</td>
</tr>
<tr>
<td>MC9256 Reticon</td>
<td>3,848</td>
<td>(256)</td>
<td>(256)</td>
<td>N/S</td>
<td></td>
<td>Ph. diode</td>
<td>60</td>
<td>N/A</td>
<td>N/A</td>
<td>nil</td>
</tr>
<tr>
<td>850 Eikonix</td>
<td>4096</td>
<td>(5200)*</td>
<td>(5200)*</td>
<td>N/S</td>
<td></td>
<td>CCD</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>1 pix max</td>
</tr>
<tr>
<td>EC 78/99 Eikonix</td>
<td>27,000</td>
<td>(1728)</td>
<td>(2048)*</td>
<td>N/S</td>
<td></td>
<td>Ph. diode</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>1 pix max</td>
</tr>
<tr>
<td>610 Datacopy</td>
<td>10,800</td>
<td>(1728)</td>
<td>(2846)*</td>
<td>N/S</td>
<td></td>
<td>CCD</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>1 pix max</td>
</tr>
<tr>
<td>300 Datacopy</td>
<td>10,800</td>
<td>(1720)</td>
<td>(2592)*</td>
<td>N/S</td>
<td></td>
<td>CCD</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>1 pix max</td>
</tr>
</tbody>
</table>

Notes: N/S = not specified, N/A = not applicable, * = mechanically scanned in one direction, (I) = number of elements
resolution and has excellent distortion specifications of 0.5% geometric and 0.25% linearity distortion. However, at approximately $5000, the Dage 68 is too expensive to be used in this project. The Dage 66, which has slightly higher distortion specifications than the Dage 68, cost $3564.

Of the solid state cameras listed in Table 3.1, the Javelin JE2062 (Javelin Elec. Inc., Torrance, CA) has the best resolution, at approximately 450 lines. This camera uses a 384 x 485 MOS image sensor and incorporates some signal processing techniques to improve the resolution (Flory, 1985). This camera was tested against a Dage model 65, which is very similar to the Dage model 66, for resolution. Both cameras showed a limiting resolution of approximately 400 lines. The Javelin JE2062 has all the advantages of a solid state sensor, as well as a resolution comparable to a high quality tube type camera such as the Dage 65, accordingly, two of them were purchased, at $2000 each, for this project.

3.2.3 Camera Evaluation

The major factor in selecting the Javelin JE2062 instead of other solid state cameras is the higher resolution specified; therefore, resolution will be the major test in evaluating the camera. Other tests of the camera will be actual operational tests in the digitization of images.

There are two common methods to test the resolution of a camera. As mentioned before, one method is to use a wedge pattern of converging black and white lines (see Figure 3.4). With the test pattern properly adjusted in the FOV of the camera, the "limiting resolution" is obtained by observing the point when the black and white lines in the wedge pattern cannot be distinguished any more. Normally, two sets of the wedge pattern, oriented horizontally and vertically, are used to determine the horizontal and vertical resolution of the camera. This test method is
simple to perform but gives only the approximate limiting resolution of the camera.

A more detailed resolution test method involves using the RCA P200 test chart (Neuhauser, 1979) (see Figure 3.5). The P200 test chart consists of blocks of slanted parallel lines at different resolution; the angle of the parallel lines depends on the line resolution of the particular block.

The advantage of the P200 test chart is that by decreasing the slant of the parallel lines with increasing resolution, the frequency of black/white transitions during a horizontal line scan can be kept constant at 1.45 MHz. This results in amplitude measurements that are independent of the frequency response of the camera's video amplifier; also, this allows the use of a low bandwidth video amplifier which results in significantly less noise in the video signal. Using an oscilloscope, the camera's response, measured horizontally in a scan line, for each block is recorded. A plot of the response versus the line resolution yields the MTF of the camera, which is much more useful in comparing different cameras.

The test setup required for the P200 test chart is shown in Figure 3.6. Amplitude response measurements were taken for line resolutions up to 500 TV lines (TVL). The MTF of the Javelin JE2062 is shown in Figure 3.7. A Dage 65 vidicon camera was also tested and the MTF was plotted with the Javelin JE2062. Comparing the MTF's of both cameras, it can be seen that the Dage 65 has a higher response than the JE2062 up to approximately 300 TVL. Beyond 300 TVL, the response of the two cameras are approximately the same. Thus, Javelin JE2062 has a lower response than the Dage 65 - a mid-to-high priced vidicon camera.

Although the Javelin JE2062 did not perform as well in the resolution test, it had advantages over the Dage 65 in stability and burn resistance. In actual image digitization sessions performed at different times, the Dage 65 had heat problems
Figure 3.4 Resolution test pattern with converging lines
Figure 3.5 RCA P200 test pattern.
RCA P200
Test Chart

Camera

Low Pass Filter
0 - 3 KHz.
(Wavetek 452)

Ch. 1 B output
Tektronics 468
oscilloscope
Ch. 2
delayed trigger output

Video
Monitor

Figure 3.6 Test set-up for RCA P200 test.
Resolution Test Results

Figure 3.7 MTF of Javelin and Dage.
which caused spurious signals to appear on the screen and affected the aspect ratio; also, after prolonged exposure to the backlighting unit, shadows of the previous image could be seen on the current image as a result of sensor burn. Even though other photoconductors are available for camera tubes which are more resistant to sensor burn, exposure to bright light sources can still decrease the life of the tube; on the other hand, the Javelin JE2062 solid state camera is not plagued by heat or sensor burn problems.

In a subjective test, images taken with the Javelin JE2062 contained the same information when compared to images taken with the Dage 65. In addition, features extracted on images taken with the two cameras did not show any significant difference. Hence, the Javelin JE2062 solid state camera is suitable for this application.

3.3. Selection and Evaluation of Image Digitizer

3.3.1 Image Digitizer Review

An image digitizer converts an input video signal into a digital signal by passing the video signal through an analog to digital (A/D) converter and storing the digital output in memory. Image digitizers vary according to the image resolution, the A/D converter resolution, speed of digitization, and other minor differences.

The common resolutions for digitizers are 512 x 512 pixels and 256 x 256 pixels. Although 512 lines, or 256 in the latter case, are processed by the digitizer, only 480 lines, or 240, correspond to the number of active lines in most cameras, and therefore contain actual image information.

The A/D converter in the digitizer determines the number of gray levels in the digitized image. Common gray scale digitizers have resolutions of 4 to 8 bits. For a binary digitizer, a voltage comparator is used instead of an A/D converter, and the
binary voltage threshold can usually be set to one of 256 levels.

The speed of digitization usually depends on the speed of the A/D converter used. Real-time image digitizers with fast A/D converters can digitize a complete frame of the video signal in \( \frac{1}{30} \) sec.; slow scan image digitizers may take several frames to digitize one image. The advantage of real-time digitizers is the reduced noise due to motion artifact since the integration time of the image is just one frame, \( \frac{1}{30} \) sec.

Other important considerations for an image digitizer are how to access the image data stored in the digitizer's memory and the availability of look up tables. The simplest method to access the image data is to have the image, or portions of the image, mapped into the memory addressing space of the computer such that accessing a pixel is done by a simple memory read or write. Other access methods use the I/O ports and may involve forming a pixel pipeline at the digitizer board to speed up the data transfer process. Look up tables (LUT) are extremely useful in performing gray-scale image transformation, such as binarization, reverse video, and histogram equalization. LUT's may be located at the video input stage, which causes the transformed image to be stored, or at the output video stage, which leaves the digitized image intact, while displaying the transformed image.

3.3.2 Digitizer Selection

A digitizer capable of producing binary images is required for the imaging system. The digitizer should have 512 x 512 pixels resolution, have at least two camera inputs and should capture the images in real-time. Table 3.2 lists most of the image digitizers available for the IBM PC. As with the camera selection, the total cost of the imaging system should not exceed $10,000.
Table 3.2 Listing of commercial image digitizers considered for vision system.

<table>
<thead>
<tr>
<th>Model and Manufac.</th>
<th>Price</th>
<th>Img Res</th>
<th>Dig Res</th>
<th>Disp Res</th>
<th>Fr Rate</th>
<th>In Sig</th>
<th>In LUT</th>
<th>Out LUT</th>
<th>Computer</th>
<th>Img Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCVision-1 Img. Tech.</td>
<td>4043</td>
<td>512x512</td>
<td>6</td>
<td>8</td>
<td>30</td>
<td>RS170</td>
<td>0</td>
<td>4</td>
<td>PC,XT,AT</td>
<td>Memory</td>
</tr>
<tr>
<td>PCVision-2 Img. Tech.</td>
<td>4043</td>
<td>512x512</td>
<td>8</td>
<td>8</td>
<td>30</td>
<td>RS170</td>
<td>opt.</td>
<td>4</td>
<td>PC,XT,AT</td>
<td>Memory</td>
</tr>
<tr>
<td>IVG-128 Datacube</td>
<td>4050</td>
<td>384x485</td>
<td>6 or 8</td>
<td>8</td>
<td>30</td>
<td>RS170</td>
<td>2</td>
<td>6</td>
<td>PC,XT</td>
<td>1/O</td>
</tr>
<tr>
<td>Silicon Video Epix</td>
<td>3368</td>
<td>752x480</td>
<td>8</td>
<td>8</td>
<td>30</td>
<td>RS170</td>
<td>0</td>
<td>opt.</td>
<td>PC,XT</td>
<td>Memory</td>
</tr>
<tr>
<td>Oculus 200 Coreco</td>
<td>2850</td>
<td>512x512</td>
<td>7</td>
<td>8</td>
<td>30</td>
<td>RS170</td>
<td>2</td>
<td>0</td>
<td>PC,XT</td>
<td>1/O</td>
</tr>
<tr>
<td>Oculus 100 Coreco</td>
<td>985</td>
<td>512x512</td>
<td>1</td>
<td>N/A</td>
<td>30</td>
<td>RS170</td>
<td>N/A</td>
<td>N/A</td>
<td>PC,XT,AT</td>
<td>1/O</td>
</tr>
<tr>
<td>Oculus 150 Coreco</td>
<td>unk</td>
<td>512x512</td>
<td>1</td>
<td>1</td>
<td>30</td>
<td>RS170</td>
<td>N/A</td>
<td>N/A</td>
<td>PC,XT,AT</td>
<td>1/O</td>
</tr>
<tr>
<td>DT2801 Data Trans.</td>
<td>unk</td>
<td>256x256</td>
<td>6</td>
<td>8</td>
<td>30</td>
<td>RS170</td>
<td>8</td>
<td>4</td>
<td>PC,XT</td>
<td>Memory</td>
</tr>
<tr>
<td>PC-Eye Chorus</td>
<td>668</td>
<td>640x512</td>
<td>N/A</td>
<td>10 max</td>
<td>RS170</td>
<td>N/S</td>
<td>N/S</td>
<td>PC,XT</td>
<td>N/S</td>
<td></td>
</tr>
</tbody>
</table>

notes: unk = unknown, N/A = no applicable, N/S = not specified, opt = optional
Both gray-scale and binary digitizers are capable of producing binary images, except gray-scale digitizers use 8 bits per pixel while binary digitizers use 1 bit per pixel. There is no need to use 8 bits per pixel unless extensive mathematical computations are performed on each pixel; hence, the less expensive binary digitizers are more appropriate.

The Coreco Oculus 100 and Oculus 150 (Coreco Inc., Longueuil, P.Q.) are binary digitizers with very similar features. Both digitizers have 512 x 512 resolution and the binary threshold can be set at one of 256 levels. It is suitable for the dual resolution approach as up to 4 cameras can be used simultaneously. Additionally, the Oculus 150 digitizer has a video output for displaying binary images and some hardwired OR, AND, XOR and NOT logic for comparing images which are not available with the Oculus 100. The output display feature is particularly useful for viewing images since the graphic capability of the IBM PC is limited. However, as the Oculus 150 was not yet available at time of selection, the Oculus 100 was chosen. The cost of the Oculus 100 is $985, which is well within the overall budget for the imaging system.

3.3.3 Digitizer Evaluation

The Oculus 100 image digitizer was evaluated in its ability to produce a clean, usable image for subsequent feature extraction. As well, it was compared to a $9000 Imaging Technology IP-512 (Imaging Technology Inc., Woburn, MA) image digitizer at the Electrical Engineering Department at UBC.

In the evaluation setup, the digitizers were tested on a low-resolution image only since this is the worst case with the smallest details; the set up is shown in Figure 3.8. The IP-512 was first used to obtain a binary low-resolution image, see Figure 3.9. The Oculus 100 was then tested and compared to the IP-512 digitized
image. It should be noted that the comparison is not exact due to the limitations in software for the Oculus 100 and hardware for the IBM PC. The IP–512 can display and print a full 512 x 512 pixels image, using programs developed by Mr. J. Clark and K. Chan, whereas the Oculus 100 can only display and print a 512 x 200 pixels image of one field, using vendor supplied programs.

In evaluating the Oculus 100, it was discovered that the digitized field of the Oculus 100 is slightly smaller than the IP–512, by approximately 15% in width. This will cause the aspect ratio of the image to change as well, to approximately 1.03.

After adjusting the digitized field to cover the entire width of the tray, 512 x 200 pixels images were displayed on the IBM PC’s screen. Significant noise problems were seen in the digitized images and could not be eliminated by adjusting the binary threshold or the aperture of the lens. These problems were probably due to the video amplification or conditioning circuitry in the Oculus 100 since the IP–512 did not show excessive noise. The noise problems were finally overcame by increasing the backlight diffusion in certain parts of the tray. This indicates that perhaps the DC or gain level may be the cause of the noise problem. The final 512 x 200 pixels Oculus 100 image is shown in Figure 3.10.

From Figure 3.10, it can be seen that the Oculus 100 can produce a sufficiently clean image for feature extraction. However, the object edges in the image were quite coarse as compared to the edges in the IP–512 image; the coarse edges may be partially due to the 200 line resolution printed. The tests performed in this evaluation shows that the Oculus 100 is capable of producing clean images for feature extraction, but it has drawbacks in its video circuitry and its smaller digitized field.
Figure 3.8 Digitizer evaluation set-up.
Figure 3.9 Image digitized by IP-512.
Figure 3.10 Image digitized by Oculus 100.
CHAPTER 4

Recognition Algorithm

4.1. General Recognition Problem

The task of recognizing objects, or more generally patterns, in a visual field comes under the category of image pattern recognition. The recognition task can be divided into several subtasks as shown in Figure 4.1.

Figure 4.1 Pattern recognition system block diagram

In general, an image pattern recognition system obtains an image as input and produces descriptive results about the input image. The subtasks involved are: image sensing, where the image is converted to a format suitable for processing, image preprocessing, where an input image is modified to enhance certain characteristics of the image, feature extraction, where qualitative and quantitative measurements are obtained from the enhanced image, and classification, where the features are used to obtain the results of the recognition system.

The image sensor may produce a 2 dimensional (2D) or 3 dimensional (3D) representation of the input scene. Two dimensional sensors are more common since the
system complexity, both in hardware and software, is much less than 3D systems and most objects are relatively flat such that a 2D planar representation is sufficient. Only when depth information becomes significant are 3D vision systems used.

Image preprocessing usually performs one or more transformations on the image to facilitate the feature extraction task. For 2D images, the transformation may be geometric, such as rotations and translations, and/or scalar, such as intensity modification. Generally, scalar transformations are extremely easy to perform using hardware look-up tables and are often very useful in highlighting desired features.

The most common scalar transform is binary thresholding, which is used in most first generation in computer vision systems. Although some new systems use gray scale image pattern recognition techniques, binary vision systems are still very much in use. Some of the differences between binary and gray scale vision have been discussed by Kelley (Kelley, 1983). The main advantage of gray scale vision is the additional intensity information in the image which permits the use of powerful pattern recognition algorithms to solve complicated scenes, such as touching or overlapping parts. However, if the objects to be recognized can be adequately represented in binary, and if suitable lighting conditions exist, such as backlighting in this application, then binary vision would be preferable since it is much faster, less complicated and less expensive than gray scale vision.

Image feature extraction and classification are subjects of intense research in both pattern recognition and robotic vision. Feature extraction is perhaps the most difficult problem in image pattern recognition. Once the features have been selected, the classification scheme is usually limited to a few choices. However, the following quotes from Levine’s survey (Levine, 1969) summarizes the underlying difficulties of feature extraction:

"No general theory exists to allow us to choose what features are
relevant for a particular problem ... Design of feature extractors is empirical and uses many ad hoc strategies. At present the only way the machine can get an adequate set of features is from a human programmer. The effectiveness of any particular set can be demonstrated only by experiment."

Thus, clever selection of features will normally lead to simple solutions. Unfortunately, there are no general features which work with a large variety of patterns but many individual features which are developed for specific patterns or specific classes of patterns. Also, as stated above, the only method of evaluating a feature is by experimentation; hence the feature selection process may be very time consuming.

Important criteria in feature selection are system constraints, noise sensitivity, reliability, consistency and resolution of the images. Features should be selected to satisfy as many, if not all, of the criteria. System constraints and noise sensitivity criteria may sometimes limit the choice of the features to a small set. There is, in general, no way of determining the resolution necessary for a particular feature other than by experimentation; studies have been performed to determine the minimum resolution for recognizing military targets (RCA, 1974; Hafemeister et al., 1985) but the results only pertain to human observers.

4.2. Description of Instruments and Layout

In the dual resolution approach, the 36 instruments are divided into 2 groups, the 20 high-resolution instruments and the 16 low-resolution instruments. The high-resolution instruments are the instruments which are in the high-resolution image as well as the low-resolution image; the low-resolution instruments are the instruments which can be seen in the low-resolution image only. The layout and grouping of these instruments will be discussed in this section.
The area on the tray with the 20 high-resolution instruments is designated the high-resolution area. The instruments in this area are arranged in a semi-circular pattern with their tips pointing towards the centre. The tips, which contain the fine details used to separate the individual instrument are positioned such that they are within the field of view (FOV) of the camera for the high-resolution area. The high-resolution instruments are further divided into three groups: the Acufex instruments, the obturators and the cartilage knives. The Acufex instruments refer to the graspers, scissors, and punches manufactured by Acufex Microsurgical Inc.; the obturators refer to the obturators and the trocars; the cartilage knives refer to the 4 cartilage knives and the hooked probe.

In arranging the high-resolution instruments, the compartments are designed and positioned so as to achieve the highest resolution possible, and to minimize the possible confusion between groups while retaining the flexibility that any instrument of the same group can fit in any of the compartments for that group. In doing so, the Acufex instruments are further subdivided into 3 long and 8 regular instruments. Three extra long compartments are reserved for the long Acufex instruments so as to minimize the area used by the other compartments.

There are no special grouping requirements for the low-resolution instruments. The only constraints for the low-resolution area are that it should be as small as possible to achieve the highest resolution for the low-resolution image, and also to arrange the instruments so that they are all within reach of a 5-degrees of freedom robotic arm. The dimensions of the tray should also be similar to the aspect ratio of the camera to make full use of the FOV. The final layout of the instruments are shown in Figure 4.2.

4.3. Survey of Existing Algorithm
Figure 4.2 Layout of proposed instrument tray.
4.3.1 Overview

Algorithms for image pattern recognition generally consist of 2 parts, a feature extractor and a classifier. There are many ways to extract features from an input pattern and the choice of the features greatly influences the type of classifier used. As stated in a previous section, there is no general theory for feature selection and appropriate selection of features can simplify the recognition problem considerably.

In binary vision, the common features for pattern recognition are geometric and topologic properties such as area, length, width, perimeter compactness, number of holes, hole area, minimum and maximum radii, and invariant moments. Another useful technique is matched filtering. These features obtain information using the whole input pattern and are referred to as global features.

For input patterns in which the global features have difficulty discriminating, the boundary is often a source for generating more detailed features; hence, these new features are call boundary-oriented features. For most binary vision applications, the boundary is sufficient to completely describe the objects to be recognized. Thus, boundary-oriented feature techniques are very powerful for binary-image pattern recognition.

The features generated by the feature extractor are used by the classifier to make a decision. There are generally two approaches to pattern classification: the decision-theoretic method and the structural or syntactic method. Decision-theoretic classifiers apply some form of discriminant function to the input features and are extremely easy to use. However, the input features to a decision-theoretic classifier must be organized in the form of a feature vector.

In some cases, due to the type of feature selected, it may not be possible to organize the feature into a feature vector. For such features, the structural or syntactic
classifiers may be more suitable. These classifiers use powerful graph-matching techniques to find the amount of similarity between the input pattern and the references. The disadvantages of the structural or syntactic classifiers are that they are fairly complicated and are somewhat difficult to use.

Most of the techniques mentioned have been employed in commercial vision systems. In the early vision systems, which were mostly binary, global features were used with decision-theoretic classifiers to recognize industrial objects (Gleason & Wilson, 1981; Carlisle et al, 1981). More recently, with the advances in computer hardware and software, many systems have been developed which use boundary-oriented features and a structural classifier. These systems have been successful in recognizing touching and overlapping objects (Rummel and Beutel, 1984).

In the following sections, the criteria for selecting pattern recognition algorithms will be given. More details will be presented on different global and boundary oriented features, and on different classifiers.

4.3.2 Criteria for Recognition Algorithm

The first step in designing a vision system is to examine the conditions of the working environment. The application constraints must be well understood before algorithm development can begin. This section outlines the conditions of the application environment and the constraints of the system, and discusses the desirable features for the recognition algorithm.

One of the most important consideration in designing a recognition algorithm is the sensitivity to noise. In the operating room environment, the noise sources which may cause difficulties for the vision system are overhead lighting, small tissue fragments on the instruments, and blood or other opaque fluid on the instrument tray. Other sources of noise unrelated to the operating-room environment are noise due to
the camera and noise due to the instrument tray.

To eliminate or minimize the effects of noise on the system, the features chosen must not depend on characteristics which may be corrupted by the noise. For overhead lighting, the noises are usually bright reflections along the boundary or in the middle of the instrument. Similarly, tissue fragments may lead to distortions in the boundary. Therefore, the features should not depend on the absolute shape of the boundary or use the interior points in an input object.

For the problem of opaque material, such as blood, being left on the instrument tray, there may not exist a software solution since the blood may hide an essential portion of the instrument from the camera, making recognition impossible.

4.3.3 Feature Extraction Algorithms

This section presents the details of various feature extraction algorithms, divided into subsections of global features and boundary-oriented features. The advantages and disadvantages of each feature will be discussed in the context of this project. As it is impossible to discuss the myriad of feature-extraction algorithms that have been developed over the years, only those that can potentially be useful in this project will be presented. Complete discussions on the topic of feature extraction may be found elsewhere, e.g. (Levine, 1969; Rosenfeld, 1981; Gonzalez and Safabachsh, 1982).

4.3.4 Global Features

Global features for binary vision systems are generally fairly easy to compute. The most common of the global features are geometric and topologic properties. The geometric features are area, perimeter, maximum/minimum radii, compactness, elongatedness ... etc. The topological properties are the number of holes, branches...etc in the object. Other global feature which are used often are centroidal or invariant
moments, and matched filtering.

The geometric features can be computed easily by first tracing the boundary of the object. The tracing process generates a series of direction changes along the boundary which may be represented by the Freeman chain code (Freeman, 1961). Starting from a given pixel and tracing the boundary in either the clockwise (CW) or counter-clockwise (CCW) direction, the next pixel may be in any one of 8 neighbours to the given pixel. The direction vectors for the 8 neighbours to a pixel 'P' are shown below,

\[
\begin{array}{cccc}
3 & 4 & 5 \\
2 & P & 6 \\
1 & 0 & 7 \\
\end{array}
\]

Given the chain code of an object, Freeman has shown that the area, perimeter, centroid, and centroidal moments can be calculated easily (Freeman, 1961). Other methods have been proposed for obtaining area and perimeter faster, based on information from a raster scan, or more accurately, based on the different sampling grids used (Grant and Reid, 1981; Agrawala and Kulkarni, 1977; Capson, 1984).

The advantage of area, perimeter and centroidal moments is that they are translation invariant. Area and perimeter are also rotation invariant. To achieve rotation invariance for the centroidal moments, a series of mathematical transformations may be performed to obtain invariant moments (Wong and Hall, 1978). The disadvantage of these features is that they are sensitive to noise which corrupts the object, such as reflections causing portions of the binary object to become the background. Nevertheless, in a well planned environment, these features have been successful in recognizing many industrial objects.

From the area and perimeter, other geometric properties, such as the compactness and the eccentricity or elongation can be computed (Ballard and Brown,
The compactness $C$ is defined as

$$C = - \frac{P^2}{A}$$  \hspace{1cm} \text{(4.1)}$$

where $P$ is the perimeter and $A$ is the area of the object. The eccentricity is defined as the ratio of the principal axes of inertia. These features, although simplistic, provide additional information on the object to be recognized.

Topological properties such as the number of holes, branches in the object are often useful features. Associated parameters such as the hole area, perimeter, moments and position with respect to the object centroid are also very valuable as features. A straightforward method to obtain the number and type of branches in the object is thinning. Thinning shrinks the body of an object symmetrically by sequentially deleting pixels from the object boundary, leaving only a thin single pixel wide connected line skeleton. The problem with thinning is that it is not a unique representation; two objects differing only in a scalar factor can produce the same skeleton. As well, thinning require connectivity to be preserved, which may not be possible in the noisy low-resolution images.

Matched filtering, or template matching, is a straightforward method of comparing an unknown image to a reference image. However, in its simplest form, it is not scale or rotation invariant. Although transformations can be used to align the templates prior to matching, it would be computationally intensive unless implemented in hardware. Another problem with matched filtering is that it is only good if gross differences are present and is unreliable for small details.

4.3.5 Boundary Oriented Features

The first step to extract features from the boundary of an object is to obtain a representation of the boundary. One method, the Freeman chain code, has already
been presented in the last section. Other representations are Fourier descriptors (Persoon and Fu, 1977), Walsh descriptors (Sarvarayudu and Sethi, 1983), shape numbers (Bribiesca and Guzman, 1980) and \( \psi - S \) curve (Ballard and Brown, 1982). A survey of the above and other boundary oriented techniques may be found in Pavlidis' paper (Pavlidis, 1980).

To calculate the Fourier descriptors, the boundary pixels are rewritten as a complex function as follow:

\[
b(t) = b_x(t) + jb_y(t) \tag{4.2}
\]

where \( b(t) \) is the boundary contour and \( b_x(t) \) and \( b_y(t) \) are pixel \( x, y \) positions along the contour. The Fourier descriptors \( T_n \) are given by

\[
T_n = \frac{1}{L} \int_0^L b(t) e^{-j(2\pi n/L)} \, dt \tag{4.3}
\]

Alternatively, the Fourier descriptors may be computed from the \( \psi - S \) curve representation, to be discussed later. The advantages of Fourier descriptors are that the theory and computation of Fourier transforms are well developed, are information preserving, and are invariant to rotation and translation. The disadvantages of Fourier descriptors are that they require a large number of coefficients for good boundary representation, they have difficulties discriminating opposite symmetries (Pavlidis, 1977) and they are sensitive to noises which corrupt portions of the boundary.

Walsh descriptors are similar to Fourier descriptors except the Walsh transform is used instead of the Fourier transform. The advantage of Walsh descriptors is that they require less computation than Fourier descriptors. However, the Walsh descriptors have the same disadvantages as the Fourier descriptors, in addition to being sensitive to the starting point where boundary tracing begin.
The shape number of an object is a sequence of numbers representing different types of lines or corners which make up the object boundary. The algorithm begins by finding the principal axes of the object and assigns a rectangular grid binding the object. The resolution of the grid is allowed to vary until the number of boundary pixels is equal to a number specified by the user - this number is the order of the shape number. The numbers in the shape number describing the boundary is assigned as follows: convex corner is 1, straight line is 2, and concave corner is 3. After the object has been labelled, the sequence of numbers is circularly shifted until the sequence represents the smallest number - the shape number. This method has a sound theory and seems reasonably simple. However, it requires the ability to assign the sampling direction and the sampling resolution, which makes it impractical.

The $\psi$-$S$ curve is a one dimensional continuous curvature representation of the boundary. $\psi$ is defined as the tangent angle to the boundary at the point $S$, the arc length along the boundary. In the discrete case where $\psi$ is the angle between adjacent points, the $\psi$-$S$ curve becomes the angular equivalent of the chain code. For angles measured several pixels apart along $S$, the $\psi$-$S$ curve provides a smoother representation of the curvature than the chain code.

Having presented various methods to represent object boundaries, the next step is to extract features from them. For the Fourier and Walsh descriptors and the shape number, further processing is unnecessary as they already form a complete feature set. The chain code and the $\psi$-$S$ curve, however, contain local features which can be extracted. Duda and Hart (Duda and Hart, 1977) have suggested that points of high curvature along a boundary can be used to describe the boundary. This is echoed by Pavlidis (Pavlidis, 1977) who stated that curvature maxima and corners are important in shape perception, as discovered in the early theories of vision. Also, from the work of Attneave, Bennett and MacDonald (Bennett and MacDonald, 1975) stated that "the
information sufficient for the recognition of familiar shapes, is contained in a knowledge of the points of maximum absolute curvature on the boundaries of those shapes and their relative location and connectivity."

Several methods have been proposed to calculate or approximate the curvature \( K(s) \), which is defined as

\[
K(s) = \frac{d\theta(s)}{ds}
\]

where \( s \) is the arc length along the boundary and \( \theta \) is the tangent angle along \( s \). Let a forward vector be a vector from the current pixels to the pixel \( n \) places forward along the arc length, and the backward pixel as \( n \) places back. Geisler used the angle difference between a forward vector and a backward vector to the pixel of interest, with each vector being at least 10 pixel distances in length (Geisler, 1982).

Dessimoz defined curvature \( K(s_k) \) as

\[
K(s_k) = \frac{\Delta \theta_k}{\Delta S_k} = \left[ \theta_k - \theta_{k-1} \right] / \left[ (S_{k+1} - S_{k-1}) / 2 \right]
\]

which is consistent with the previous definition of \( K(s) \) (Dessimoz, 1978). However, his angle measurements seemed somewhat inconsistent and led to a different angle change from the mathematical definition.

Rosenfeld and Weszka proposed a method of angle detection by computing the \( k \)-cosine, the angle between the forward and backward vector, for several vector lengths (Rosenfeld and Weszka, 1975). The \( k \)-cosine is defined as

\[
\cos C^{(k)} = \left[ S_k \ t_k \right] / \left[ |S_k| \ t_k | \right]
\]

where \( S_k \) and \( t_k \) are the forward and backward vectors of length \( k \). The angle of the
particular point is chosen to be the minimum of the angles calculated.

So far, the curvature or angle techniques presented have been based on angle measurements between pixel points. An ad hoc technique have been developed by Hung and Kasvand to select the significant corners from a binary line in Chinese characters (Hung and Kasvand, 1983). Hung's method uses the chain code, and calculates the difference between successive chain codes and the sum of any non-zero pairs of these differences. Based on a set of seven rules, corner pixels are labelled as critical points. The advantage of this method is speed, since it does not require any intensive calculations.

4.3.6 Classification Algorithms

The choice of classifiers in a pattern recognition system depends on the features chosen. Two common techniques for classifiers are decision-theoretic and structural. Decision-theoretic classifiers use quantitative features and make a decision based on some partitions in the feature space. Structural, or syntactic, classifiers use qualitative and quantitative features and make a decision based on a hierarchical process. This section discusses the advantages and disadvantages of various methods using these two techniques.

There are many approaches to partitioning the feature space in decision-theoretic classifiers. The nonparametric approaches are linear or polynomial discriminant functions, minimum distance classifier and nearest neighbour classifier; a parametric approach is the Bayes Classifier. These classifiers are discussed in detail elsewhere (Fu, 1982; Duda and Hart, 1977). The primary advantages of decision-theoretic classifiers are that they are fast, simple to use and train. In fact, given an appropriate feature set for the reference objects, it may be possible to perform unsupervised training. The primary disadvantage of this classification approach
is that it lacks flexibility in using features since the input feature set must be in the form of a feature vector.

Structural classifiers, on the other hand, are extremely flexible in dealing with features. Most structural classifiers implement some form of graph or string matching to assign a class to the input object. As there is no restriction on the data type at the nodes of the graph, any characterizable properties of the object may be used as features. Brief discussions of various graph matching techniques may be found elsewhere (Ballard and Brown, 1982). The matching process in a structural classifier can be controlled by the program designer, who can specify the degree of match desired for the application. Thus, it is possible to obtain a result based on only a partial match between the references and the input object. This powerful feature of the structural classifier makes it suitable for classifying overlapping or touching objects, where a complete match is not usually possible. The drawbacks of a structural classifier are complicated programming, slow execution and, possibly, complicated training procedures.

4.4. Development of Recognition Algorithms

The recognition of arthroscopic instruments has been divided into two subproblems, that of recognizing a subset of the instruments at a lower resolution, over the entire instrument tray, and that of recognizing the remaining instruments at a higher resolution. The algorithms developed for the low-resolution image discriminate the low-resolution instruments and also obtain any gross features for the high-resolution instruments to assist the algorithms for the high-resolution images. The algorithms for the high-resolution image concentrate on the fine details of the instruments under the high-resolution view.
This section presents the algorithms developed for the low-resolution and high-resolution images. The results of a survey leading to the formulation of a set of clinical performance requirements are also discussed.

4.4.1 Low-Resolution Algorithm

The low-resolution image is a 512 x 476 pixels image of all 36 instruments to be recognized. The overall field of view (FOV) of the camera is 68 cm x 52.6 cm, which is slightly larger than the 68 cm x 50 cm tray. To achieve the maximum resolution in this large FOV, the width of the tray is aligned with the picture width; the picture height is allowed to extend beyond the tray area by a small amount. A typical low-resolution image with all 36 instruments is shown in Figure 4.3

In digitizing the low-resolution image, the binary threshold was initially set to a low value, which caused a large portion of the image to be filled with noise. As the threshold was increased, the noise decreased until it could only be seen in the corners of the image. If the threshold was increased further, the noise would completely disappear from the screen; however, the increased threshold would also cause portions of the instruments to disappear, creating gaps in the instruments. The final threshold level chosen for the low-resolution image was the level at which the noise was relatively removed from the instrument compartments. For consistency in the features, the binary threshold of the test and training data was determined from the first image and kept constant throughout the digitizing session.

After the binary digitization process, each image was subjected to feature extraction. At this stage, all background noise had been eliminated from the instrument compartments; however, numerous small gaps could be seen in the image. These small gaps occurred at places where the instruments were extremely narrow, such as the tips of the large and small towel clips, and the tip of the drainage cannula. Gaps were
Figure 4.3 Typical low resolution image.
also found along the finger holes of the small towel clip. These gaps were due to the backlighting reflecting off the finger holes and onto the camera, for instruments near the centre of the tray; this effect could be observed in the thickness of the finger holes, which increases with the distance from the tray centre.

From the low-resolution image, Figure 4.3, it can be seen that there is very little information present for discriminating the individual high-resolution instruments. Therefore, it was considered more appropriate to recognize these instruments under high-resolution. The remaining low-resolution instruments were found to exhibit various gross differences which could be characterized by simple global features, provided that such features were insensitive to the small gaps that are present in some instruments.

Of the global features reviewed in a previous section, the methods that are insensitive to small gaps are length, width and area. Since the presence of small gaps changes the topology of the objects, topological properties cannot be used as feature. As well, perimeter cannot be used as errors would occur around the gaps. Maximum/minimum radii and centroidal moments would not be effective in this case since the gaps may cause portions of the instrument to become disconnected, creating difficulties when calculating the centroid.

To increase execution speed, only the length and width were chosen as features for the low-resolution instruments. Although these features are not rotation invariant, it does not affect this application as the instruments are limited in their rotation by the compartment boundaries. The length and width are obtained by scanning each compartment line by line, for background-to-object transitions, or vice versa. The first and last scanned object lines, and the maximum width are recorded in the scanning process. From the first and last scanned lines, the extremities of the instrument are located to obtain the length of the instrument. The width is simply given by the
maximum width of the scanned lines.

The speed of the algorithm can be increased by scanning every few lines instead of every line. By backtracking and forward tracking at the first and last scanned lines respectively, a close approximation of the length and width can be obtained. If the area is to be calculated, the scan line frequency must be increased before a good approximation to the area can be obtained. For the low-resolution images, the scanning frequency was set to be every four lines.

A graph of the distribution of the instruments in the length and width space is given in Figure 4.4. The separation of different instruments in the graph indicates that the length and width are sufficient to discriminate the low-resolution instruments. For a classifier, the BMDP discriminant function is used. This is a linear discriminant function which is extremely easy to use. Training is done by entering training data into the BMDP stepwise discriminant analysis program which generates the coefficients. As this test is intended for the low-resolution instruments, it will be referred to as the low-resolution test.

Since the FOV of the high-resolution image is restricted to the tips of the high-resolution instruments, this suggested that the coarse features for the high-resolution instruments should be extracted in the low-resolution image. There are three coarse features which were found to be useful in discriminating the instruments: length, width of the handle, and pose of the instrument. The pose of a high-resolution instrument (either "pose 1" or "pose 2") defines whether the instrument is lying on one side or the other. Figure 4.5 illustrates the difference between pose 1 and pose 2. As will be shown in the next section, knowledge of the pose greatly simplified the recognition of Acufex instruments in high-resolution. The width of the handle was found to be useful in separating the cartilage knives and the Acufex instruments while the length of the instruments was useful in separating the long and
Figure 4.4 Distribution of low resolution features.
the regular Acufex instruments.

The algorithms to obtain the length, handle width and pose of high-resolution instruments were called the Acufex length test, the handle width test and the Acufex pose test, respectively. The data to calculate the three features are generated by scanning the high-resolution compartments every 4 lines, as in the low-resolution test. During each scan, the positions of the edges of the instruments are recorded.

From the scanning data of a compartment, the length is calculated using the method of the low-resolution test. To calculate the handle width, the angles along the two edges of the handle are calculated. Using simple trigonometry, the handle width is approximated, as shown in Figure 4.6. The distribution of the length and the width are shown in Figures 4.7 and 4.8. Since these 2 tests have only one feature value to separate 2 classes, a simple threshold classifier suffices. To obtain a suitable threshold, the training data for each test are entered in a single-variable BMDP stepwise discriminant analysis. The threshold is then assigned as the value for which the classification probabilities are equal.

To calculate the pose of the instrument, the angle of the long narrow tip of the Acufex instruments or cartilage knives are measured. This angle is compared to the two edge angles calculated previously for the handle width test, as shown in Figure 4.9. A minimum distance classifier is used to determine pose 1 and pose 2, based on the difference between the tip angle and the edge angles.

It should be noted that the Acufex length, Acufex pose, and handle width tests are only applied to the Acufex instruments and the cartilage knives. The obturators are not tested since two of the tests, handle width and pose, are not meaningful and the best test to separate them is performed in high-resolution.

4.4.2 High Resolution Algorithm
The pose of an instrument is arbitrarily defined as:

**pose 1**

**pose 2**

Figure 4.5 Definition of Pose 1 and Pose 2.
Figure 4.6 Determination of handle width.
Figure 4.7 Distribution of lengths of Acufex instruments and cartilage knives.
Figure 4.8 Distribution of handle widths of Acufex instruments and cartilage knives.
if $|\theta_1 - \theta_T| > |\theta_2 - \theta_T|$
then pose is 1
else pose is 2

Figure 4.9 Determination of pose for Acufex instrument and cartilage knives.
The high-resolution image is a 512 x 476 pixel image covering the tips of the Acufex instruments, the obturators and the cartilage knives. The tray area covered by the camera is a 13.6 cm x 10.6 cm region, oriented 90 degrees to the low-resolution image, located at the bottom of the image and displaced slightly to the left of centre. The change in the orientation of the high-resolution image camera is to take advantage of the aspect ratio of the camera to achieve the maximum possible resolution. A typical high-resolution image is shown in Figure 4.10.

In the high-resolution image, only the tips of all 20 high-resolution instruments are within view. Thus, intuitively, it did not appear feasible to apply global feature extraction techniques to these instruments. For the obturators, the structural differences between them are pyramidal or blunt tip and 3.8 mm or 5.0 mm diameter. From these clues, a simple obturator test was designed to measure the width of the obturator shaft and the angle at the tip. The width was measured with a line scanning procedure for every fourth line, as in the low-resolution test. To measure the angle at the tip, the boundary of the obturator tip was traced and recorded as a chain code. The angles along the boundary were then calculated as the angle between the forward vector, 5 pixels forward, and the backward vector, 5 pixels back. The minimum of these angles usually denotes the tip angle. The distribution of the data in the width-angle space is shown in Figure 4.11. The classifier used was the BMDP discriminant function.

The remaining Acufex instruments and the cartilage knives showed curvature differences at the tip, which made them ideal for boundary-oriented curvature techniques. The cartilage knives also showed some variation in the width near the tip. Using the same method as the handle width test on the low-resolution image, the width of the cartilage knives were measured and graphed in Figure 4.12. The results showed sufficient separation between the different classes that no additional feature was
Figure 4.10 Typical high resolution image.
Figure 4.11: Distribution of widths and angles for obturators.
tried. Again, for a single variable feature, a simple multi-threshold classifier was used. The different thresholds were calculated from the classification function generated by the BMDP stepwise discriminant analysis program.

For the 11 Acufex instruments, as there did not appear to be any clever scheme which would allow simple features to be used, a boundary-oriented approach was taken. The boundary of each instrument was traced and recorded as a chain code and the angles for the boundary were calculated as in the obturator test. A peak detector was then applied to the boundary angles to extract the significant corners. The results for all 11 instruments are illustrated in Figures 4.13 to 4.23. Figure 4.24 shows a high-resolution image in a "noisy" environment generated by overhead lighting. This image shows the types of degradation that could occur to the objects—"intrusions" appearing in the object silhouettes. To overcome this type of noise problems, a structural technique is proposed.

Drawing on the work of Pavlidis in matching island contours using syntactic pattern recognition techniques (Pavlidis, 1979), the tips on the Acufex instruments can be modeled as a sequence of sharp corners. In Pavlidis' algorithm, the corners represent a cyclic sequence around a contour with features such as size, type and orientation. Size is described by small, medium, large and huge; type is described by sharp protrusion or intrusion, convex or concave corner and convex or concave arc; orientation is described by directions such as East, North-East, North ..., etc. Pavlidis then sequentially matches the unknown corners to the reference corners according to the similarities of the features; a strength is used to measure the amount of similarity of each match feature pair. After different references have been matched, the unknown is classified as the reference that has the highest total strength.

For the Acufex instruments, the tips are always traced in the counter-clockwise direction, and knowing that the boundary waveforms are not cyclic, there are distinctive
Figure 4.12 Distribution of width of cartilage knives: (top) all knives; (bottom) FCK, HP, SCK, MCK.
LEFT GRASPING FORCEPS

Figure 4.1.3: Curvature of Acufex tip: large grasping forceps.
Figure 4.14: Curvature of Acuti, tip: left plain scissors.

**LEFT PLAIN SCISSORS**

![Graph of arc length vs angle](image)

<table>
<thead>
<tr>
<th>Angle</th>
<th>ARC LENGTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>-160</td>
<td>0</td>
</tr>
<tr>
<td>-140</td>
<td>40</td>
</tr>
<tr>
<td>-120</td>
<td>80</td>
</tr>
<tr>
<td>-100</td>
<td>120</td>
</tr>
<tr>
<td>-80</td>
<td>160</td>
</tr>
<tr>
<td>-60</td>
<td>200</td>
</tr>
<tr>
<td>-40</td>
<td>240</td>
</tr>
<tr>
<td>-20</td>
<td>280</td>
</tr>
<tr>
<td>0</td>
<td>320</td>
</tr>
</tbody>
</table>

- **f**: first peak
- **l**: last peak
- **/**: peaks detected
Figure 4.15 Curvature of Acufex tip: right plain scissors.
LEFT 20 DEG HOOKED SCISSORS

Figure 4.16 Curvature of Acute: up, left 20 degree hooked scissors.

f: first peak
l: last peak
Æ: peaks detected

ANGLE

ARC LENGTH
Figure 4.17: Curvature of Acuflex up-right 20 degree hooked scissors.

RIGHT 20 DEG HOOKED SCISSORS

ANGLE

ARC LENGTH

f: first peak
l: last peak
/ : peaks detected
Figure 4.18 Curvature of AcuFlex up: left 60 degree hooked scissors.
Figure 4.19: Curvature of Acufex up; right 60 degree hooked scissors.

- $f$: first peak
- $l$: last peak
- Peaks detected

**Graph:**
- X-axis: Arc Length
- Y-axis: Angle
- Plot shows peaks at specific arc lengths.
LEFT SERRATED SCISSORS

Figure 4.20: Curvature of Acufax tip: left serrated scissors.
Figure 4.2.1 Curvature of Acufex tip: right serrated scissors.

The diagram illustrates the curvature of the Acufex tip for right serrated scissors. It shows a plot of angle versus arc length, with peaks labeled as follows:

- f: first peak
- l: last peak
- Peaks detected

The graph indicates variations in curvature along the arc length, with notable peaks and troughs.
LEFT BASKET PUNCH

Figure 4.22 Curvature of Acufex tip: left basket punch.
Figure 4.23: Curvature of Acu-Text up: right basket punch.
Figure 4.24 Noisy high resolution image with intrusions.
first and last corners in the boundary. Instead of qualitative features, absolute quantities
such as distances and angles are used as features. Denoting the absolute angle of the
instrument shaft as the reference shaft angle $\theta_{sa}$, then all other corners between the
first and last corners are characterized by the distances $d_f$ and $d_l$ to the first and last
corners and by the angles $\theta_{fs}$ and $\theta_{ls}$ between the reference shaft angles and angles
to the first and last corners, as shown in Figure 4.25.

The classifier for this structural model is a sequential matching algorithm. The
matching algorithm starts with the corners in the reference models and tries to match
them to the object corners. Since the corners must appear in a specific order, if one
reference corner is matched, the algorithm then begins matching the next reference
corner at the next object corner - never allowing overlapping of matched corners. For
each reference model tried, a score is given as:

$$\text{Score} = \frac{(\text{number of matches})}{(\text{number of corners in reference})} \quad ...(4.7)$$

The classifier assigns the instrument to the class with the highest score. If the
matching algorithm is unable to match any corner to the references, then the result is
a no match condition.

The choice of first and last corners are crucial in this algorithm; if in the
presence of noise, the wrong first or last corner was chosen, then this method would
not work. To eliminate this drawback, an improvement was implemented whereby if
the algorithm could not match any of the references to the object, it would
automatically choose a new first or last corner and restart the classification process. In
actual tests of the algorithms, this extension was able to recover when the first or last
corner was incorrect. In the case of an actual no-match condition, this extension did
not produce a false match.
Figure 4.25 High resolution structural features.
To obtain the reference models for the structural classifier, all significant corners for each instrument in the training data are first generated. For each instrument, the corners from different data are then compared; the corners that correspond to the same corner in the image are grouped together. For each corner, the mean and standard deviation for each distance and angle feature are calculated. If the standard deviation is large, then the particular corner may be discarded or the mean and standard deviation may be recalculated without the outliers. The mean values of the feature of the significant corners are used to construct the reference models. In most cases, only the significant corners that are detected in a large proportion of the training data are used in the reference model.

In an attempt to improve the speed and accuracy, six features associated with the first and last corners were defined. They were:

1. the angle between the shaft angle $\theta_{sa}$ and the angle $\theta_{l}$ from first to last corner;
2. the distance $d_{fl}$ between the first and last corners;
3. the angle $\theta_{f}$ of the first corner;
4. the angle $\theta_{l}$ at the last corner;
5. the difference between the angles formed by the first corner and a pixel 10 positions before, and the angle formed by the last corner and a pixel 10 positions past ($\theta_{10}$);
6. the same test as above but with the pixel 20 positions away ($\theta_{20}$);

An illustration of the features is given in Figure 4.26. These features were used with the BMDP stepwise discriminant analysis to generate a classification function. Depending on the highest scores from the classification function results, between 3 to 5 of the highest scores are considered as "possible matches" for the input. This limits the set to be used with the structural matching algorithm and should speed up the algorithm.
In addition to the BMDP feature set, the pose and length information from the low-resolution image also help to limit the set to be matched. Given that the same instrument can look different in pose 1 from pose 2, they are treated as two different instruments, with different BMDP coefficients and reference models. Therefore, the knowledge of instrument pose reduces the "possible matches" by one-half. Similarly, knowing the length of the Acufex instrument, the sequential matching algorithm will ignore the "possible matches" which are not consistent with the instrument lengths.

To summarize the algorithm for the Acufex instruments, referred to as the Acufex test, the algorithm consists of 2 stages. In the first stage, using the results of the instrument pose, the algorithm performs a BMDP classification and identifies 3 to 5 instruments as a "possible match" for the second stage. The second stage, after eliminating the "possible matches" that are inconsistent according to the instrument length, performs a sequential match on the remaining "possible matches". The instrument with the highest score will have the object assigned to its class; if the sequential matching algorithm cannot find a match, then the result is a no-match condition.

A sample corner-matching sequence of the L60H scissors is shown in Table 4.1. The BMD classification function gave 3 possible matches: L60H, R20H and LPS. The length and pose of the unknown were 190 and pose 2 respectively. From the length of the instrument, it was determined that the unknown cannot be LPS; hence, it was unnecessary to match the unknown to the LPS reference model. After extracting the peaks in the unknown, the first and last corners were found to be corners 2 and 5, leaving only two corners to be matched. In matching against R20H, only one corner could be matched out of 3 corners in the R20H reference, giving a goodness of match of 0.333. In matching against L60H, both corners were matched to the two
Table 4.1 Typical Structural matching results.

Unknown Instrument = Left 60 Deg. Hooked Scissors (L60H)

BMD Discriminant Function Scores:

<table>
<thead>
<tr>
<th></th>
<th>L60H</th>
<th>R20H</th>
<th>LPS</th>
<th>LSS</th>
<th>RPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>714.55</td>
<td>682.89</td>
<td>681.20</td>
<td>678.57</td>
<td>677.39</td>
</tr>
</tbody>
</table>

Significant Corners Detected and Features Calculated:

<table>
<thead>
<tr>
<th></th>
<th>ANG</th>
<th>FD</th>
<th>FA</th>
<th>LD</th>
<th>LA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corner</td>
<td>1.</td>
<td>-143</td>
<td>2.</td>
<td>59</td>
<td>first corner</td>
</tr>
<tr>
<td></td>
<td>3.</td>
<td>-77</td>
<td>19</td>
<td>69</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>4.</td>
<td>98</td>
<td>19</td>
<td>84</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>5.</td>
<td>76</td>
<td>19</td>
<td>69</td>
<td>10</td>
</tr>
</tbody>
</table>

Structural Matching:

<table>
<thead>
<tr>
<th>Unknown</th>
<th>Reference</th>
<th>R20H</th>
</tr>
</thead>
<tbody>
<tr>
<td>-77 19 69 10 119</td>
<td>-78.4</td>
<td>24.8</td>
</tr>
<tr>
<td>98 19 84 8 91</td>
<td>-151.5</td>
<td>21.4</td>
</tr>
</tbody>
</table>

match score = 0.333

<table>
<thead>
<tr>
<th>Unknown</th>
<th>Reference</th>
<th>L60H</th>
</tr>
</thead>
<tbody>
<tr>
<td>-77 19 69 10 119</td>
<td>-74.8</td>
<td>18.6</td>
</tr>
<tr>
<td>98 19 84 8 91</td>
<td>98.5</td>
<td>19.7</td>
</tr>
</tbody>
</table>

match score = 1.000
Figure 4.26 High resolution BMD features

\[ \theta_{10} = \theta_{10f} - \theta_{10L} \]

\[ \theta_{20} = \theta_{20f} - \theta_{10L} \]
reference corners, giving a goodness of match of 1.000. Therefore, the structural classifier assigned the unknown to class L60H. Since the structural classifier result agreed with the BMD classifier result, the clinically acceptable match was also L60H.

4.4.3 Clinical Requirements

In designing equipment for use in the health care field, special considerations must be given to safety. The equipment must not fail in a manner which may bring harm to a patient. As well, the mode of operation must be acceptable to the clinical users, such as nurses and surgeons. This section addresses these problems and outlines some basic criteria for a clinically acceptable implementation of an image pattern recognition system.

In order to obtain pertinent data for this application, a questionnaire was drafted to solicit comments on the clinical issues related to the vision system. The questionnaire attempted to find the acceptable level of performance for a vision system, the type of assistance available for the vision system in the operating room, and other questions relevant to the instrument-passing robot. In determining the acceptable level of performance, it was desired to know what type of error and what error rate would not be tolerated. It was also considered to be useful, for establishing system performance requirements, to know (a) what assistance would be available to monitor and, possibly, to correct system alarms, and (b) the tolerable frequency of assistance. A copy of the questionnaire is shown in Appendix A.

Responses to the questionnaires were taken from 3 nurses, each with over 9 years of experience at Vancouver General Hospital and UBC Health Sciences Centre Hospital. The maximum number of errors rated as tolerable varied from 0 to 2 errors per procedure, which is approximately the same level of performance as human scrub nurses. This number, however, depends on the patience of the surgeon. It was also
learned from the questionnaire that the circulating nurse in the operating room is available to assist the vision system, by taking minor corrective actions, if she is not engaged in other activities. The nurses indicated that the frequency of alarms from the vision system should not exceed once every 15 minutes. They stated that, in general, the surgeon is not available to assist in the operation of the pattern recognition system as the surgeon should not lose visual contact with the surgical site. Complete results of the questionnaire survey are given in Appendix A.

The primary implications of the questionnaire results with respect to the recognition algorithm design are that there should be as few errors as possible, and no more than one error per procedure of 30 minutes. Since assistance is available during the procedure, it is possible for the vision system to reduce the number of errors by responding with a no-match condition when the recognition confidence is low. The no-match condition can be solved by a request for intervention by the circulating nurse, either immediately or when the unmatched instrument has been called for.

These results were used as performance criteria in the development of the pattern recognition algorithms to be used in the operating room.
CHAPTER 5

EVALUATION OF RECOGNITION ALGORITHM

The usual question that is asked of any pattern recognition system is; "How good is the system?" In this chapter, the algorithms developed are evaluated based on standard error-estimation methods. The failures of the algorithms are analysed and possible improvements to the algorithms are presented. Practical problems associated with implementation, such as noise and computation time are presented. Finally, the error estimates for each algorithm are combined, with other considerations, to provide a performance estimate for the overall system.

5.1. Test Methods

5.1.1 Methods of Error Estimation

Various methods exist to estimate the probability of error, or misclassification, in a pattern recognition system. These methods are described and reviewed in detail elsewhere, e.g. (Toussaint, 1974; Toussaint and Sharpe, 1975; Glick, 1978; Kanal, 1974). The purpose of these error estimators is to provide an accurate error estimate given a small data set. The differences between the methods are the manner in which the data set is used, the bias of the resulting estimate, and the variance of the estimate.

Resubstitution Method

One pertinent method of error estimate is the resubstitution (R) method. The procedures for the R method is as follows,

1. train the classifier on all of the available data set N;
2. test the classifier on the entire data set N.
This method makes efficient use of the data set in that the whole set is used to train the classifier. However, the results of the R method is "overly optimistic" since the classifier has encountered all the test data during training (Toussaint, 1974).

**Cross-validation (π) method**

Alternatively, several methods that are known generically as "cross-validation" may be used for error estimation. The most general of these cross-validation methods is the rotation π method. The procedure for the π method is:

1. divide the data set N into P mutually exclusive sets
2. N/P is an integer and P/N <= 0.5;
3. For i = 1, ..., P
   a. train on set (j), (j = 1, ..., P & j ≠ i)
   b. test on set (i)
4. average the results of the P training and testing sessions.

The efficiency of using the available data set in the π method depends on the ratio N/P; the efficiency increases as N/P decreases toward 2. The error estimate of the π method has been shown to be a pessimistic estimate, e.g. (Toussaint, 1974).

There are two special cases of the π method, when P = N/2 and P = 1. For P = N/2, the π method is also known as the "double cross-validation" method. The data usage is least efficient in this case since only half of the data set is used for training, and the resulting error estimate is the most pessimistic. For P = 1, the π method is also known as the "leave one out" (U) method. Of all the π methods, this method makes the most efficient use of the data set and has the least (pessimistic) bias in the resulting error estimate.
Although the U method has the least bias in its error estimate, Glick has shown that this estimate has a larger variance than "any other well known estimator of success rate (Glick, 1978)." Another disadvantage of the U method is the large number of training sessions required. For a data size of \( N \), the U method requires \( N \) training sessions while the R, double cross-validation, and \( \pi \) methods require 1, 2, and \( P \) training sessions, respectively.

Toussaint's method

An alternative for obtaining an unbiased estimate was proposed by Toussaint (Toussaint and Sharpe, 1975). Since the R method provides an optimistic error estimate while the \( \pi \) method provides a pessimistic estimate, an unbiased estimate may be obtained using the following.

\[
P_e = \frac{N}{P} \sum_{i=1}^{N/P} P_e[\pi_i] + \left(1 - W(N,P,\lambda)\right) P_e[R] \quad ...(5.1)
\]

where \( W \) is a weighting function which depends on the dimensionality \( \lambda \) and the parameters \( N \) and \( P \) in the \( \pi \) method. Experiments by Toussaint, with \( W = 1/2 \), have shown that the estimate from equation (5.1) approaches that of U method.

In designing an evaluation protocol for the pattern recognition algorithms, it is important to provide a least biased but reliable estimate of the error. The U method is not suitable since it has the largest variance of all the methods. Therefore, the \( \pi \) method was chosen to provide the error estimate. Another advantage of the \( \pi \) method is that its estimates are pessimistic, which implies that the final estimates would be a worst case (lower bound) performance estimate. Similarly, the R method may be used to obtain a best case (upper bound) performance estimate.

Ideally, the data set size \( N \) and the number of set partitions \( K \) should be large to minimize the bias in the result. However, there is a cost in user time and computing time associated with obtaining, training and testing a large data set.
Therefore, two versions of \( \pi \) method error estimator are used: \( P=4 \) and \( P=2 \). The \( P=2 \) \( \pi \) method will be used on the algorithms which are expected to perform well while the less biased \( P=4 \) \( \pi \) method will be used on the more complicated algorithms.

5.1.2 Test and Training Data

Besides choosing the appropriate error estimation method to give a reliable estimate of performance, it is also important to have appropriate samples in the data set to represent the various conditions which may arise in the practical implementation of the pattern recognition system.

An ideal data set may contain a large number of noise-free images for training and testing, and a reasonable number of images representing all possible noise sources in the system. As well, images of all possible error conditions should be included in the ideal data set.

As with the protocol for performance evaluation, there are similar costs associated with obtaining and testing of the data set which may make it necessary to use a less than ideal data set. The data set for this evaluation consists of noise free and noisy images of instruments within the group compartment. Images of error conditions for wrong instrument groups in compartments are not included in the data set. A listing of the data set is given in Table 5.1.

The following are the keyword descriptors for different test images.

1. Noisy images: images taken with backlighting and direct overhead lighting.
2. Noise-free images: images taken with backlighting and no overhead lighting.
3. Low-resolution images: images of entire tray area
4. High-resolution images: images of the lower central region of tray containing the tips of the Acufex, obturators and cartilage knives
Table 5.1 List of test data set.

**Low Resolution images:**
- 4 pose 1 noise free images (LT101...104)
- 4 pose 2 noise free images (LT201...204)
- 4 mixed pose noise free images (LTM01...04)
- 1 pose 1 noisy image (LN01)
- 1 pose 2 noisy image (LN02)
- 1 mixed pose noisy image (LN03)

**High Resolution images:**
- 16 pose 1 random comp. noise free images (HT101...HT116)
- 16 pose 2 random comp. noise free images (HT201...HT216)
- 16 pose 1 fixed comp. noise free images (HT120...HT135)
- 16 pose 2 fixed comp. noise free images (HT220...HT235)
- 4 pose 1 random comp. noisy images (HN101...HN104)
- 4 pose 2 random comp. noisy images (HN201...HN204)
- 4 pose 1 fixed comp. noisy images (HN105...HN108)
- 4 pose 2 fixed comp. noisy images (HN205...HN208)
- 6 pose 1 noise images (HT140...HT145)
- 6 pose 2 noise images (HT240...HT245)
- 2 pose 1 noise (HN110)
- 2 pose 2 noisy images

**Conditions:**
- **Position** - the positions of all instruments were moved slightly from image to image to randomize the image data.
- **Pose** - the pose of some low-resolution images were changed to randomize the image data. The changes include right side up & upside down, front to end and end to front...etc.
- **Threshold** - the binary threshold was kept constant for noise free images but allowed to vary for noisy images.
5. Pose-1 images: low or high-resolution images with the Acufex & cartilage knives lying in pose 1 (as defined in chapter 4)

6. Pose-2 images: low or high-resolution images with the Acufex & cartilage knives lying in pose 2 (as defined in chapter 4)

7. Mixed-pose images: low or high-resolution images with the Acufex & cartilage knives alternatively lying in pose 1 and 2.

8. Prespecified-compartment images: high-resolution images with Acufex instruments in any prespecified compartments


In obtaining the image data, certain conditions were defined for the position and pose of the instruments, and the threshold level for the binary images. To create a degree of randomness in the images, the positions of the instruments were changed from image to image. The variations in position included moving the instruments to the extremes of the compartment and to the extreme amount of rotation permitted by the boundaries of the compartment. Since an actual tray with compartments was not available, the effect of instruments touching the compartment boundaries was not tested here. This problem will be considered in a later section with tests on a small sample tray.

The poses of all instruments, except the Acufex and the cartilage knives, were changed in some of the images to include this effect in the data testing. The types of pose changes involved were right-side-up and upside-down; and front-to-end and end-to-front. The poses of the Acufex instruments and cartilage knives were not allowed to change freely since controlled data were required for proper evaluation of the Acufex pose test.
For noise-free images used for training, the binary threshold level was kept constant throughout the digitization session. The constant threshold is essential to obtain accurate data for the features. For noisy images used for testing only, the threshold was readjusted to compensate for the noise and was allowed to vary slightly for different noisy images.

5.2. Results of Testing Each Algorithm

All of the algorithms which were developed, and which were subsequently subjected to performance-estimation tests, are outlined in Figure 5.1. For each algorithm developed, the $\pi$ test will be used to estimate its performance. Since the $\pi$ test gives pessimistic results, the resulting performance estimate is a lower bound for the actual performance of the system. In addition, noisy images were tested with each algorithm trained on noise-free images to observe the effects of lighting noise on that algorithm.

Low-resolution Test

The "low-resolution test" classifies the 12 different low-resolution instruments using their length and width as features. Classification is done by a classification function generated by the BMD Stepwise Discriminant Analysis (SDA). A $\pi$ test, with $P=2$, was used to evaluate the performance on 12 noise free images. The result of the $\pi$ test was 100% correct classification. Using the same 12 noise-free images to train the algorithm, it was then tested on 3 noisy images. The result from the noisy images was 100% correct.

Obturator Test

The "obturator test" classifies the obturators and the trocars in the high-resolution image using the shaft width and minimum angle as features.
Figure 5.1 Overall Recognition Flowchart

Low resolution image
- Low resolution instruments
- High resolution instruments

High resolution image
- Acufex instr.
- Obturators & Trocars
- Cartilage knives

FEATURE EXTRACTOR
- Width
- Length
- Width
- Length
- Pose
- Tip curvature—Peak extractor
- Width
- Tip curvature—Min. Angle
- Width

CLASSIFIER
- Discriminant function
- Structural matching
- Discriminant function
- Threshold
- Low resolution instruments
- Acufex instr.
- Obturators & Trocars
- Cartilage knives

Acufex instr.
Obturators & Trocars
Cartilage knives
Classification is again done by a classification function generated by the BMD SDA. Using a $\pi$ test, with $P=2$, on 12 noise-free images, the result was 100% correct classification. Similarly, with the algorithm trained using the 12 noise-free images, tests done on 4 noisy images gave a 100% correct classification result.

**Tip Width Test**

The "tip width test" is done only if the "handle width test" shows that the instrument is in the cartilage knife group. It classifies the 4 cartilage knives and the hooked probe by the width of their tip. Classification is done by comparing the width with thresholds calculated by BMD SDA. A $\pi$ test, with $P=2$, was used on 12 noise-free images to evaluate the algorithm. The result was 100% correct classification. Testing the algorithm on 4 noisy images, with training done on the 12 noise-free images, also gave 100% correct classification.

**Acufex Test**

The "Acufex test" classifies the 11 Acufex instruments. It uses results from 3 other tests performed on the low-resolution image, the handle width test, the Acufex pose test and the Acufex length test. The handle width test separates the Acufex and the cartilage knives by the width of the handle. Classification is done by a threshold calculated using the BMD SDA. A $\pi$ test, with $P=2$, was done using 12 noise-free images. The result was 100% correct classification. With the same training data as in the $\pi$ test, the algorithm was tested on 3 noisy images; the result was also 100% correct classification.

**Acufex Pose Test**

The "Acufex pose test" determines the pose of the Acufex instruments by comparing the edge angles at the handle with the tip angles. This test is empirical
and does not require any training. Using 12 noise–free images, the following results were obtained: 131(99.2%) correct classifications; 1 no match condition; and 0 errors. Using 3 noisy images, the result was 100% correct classification. This test is also used to find the pose of the cartilage knives. With the same 12 noise–free images and 3 noisy images, the result was 100% correct classification.

Acufex Length Test

The "Acufex length test" classifies the long and regular Acufex instruments by their length. The classifier uses a threshold calculated by the BMD SDA. A $\pi$ test with $P=2$ on 12 noise–free images gave the following results: 130(98.5%) correct classifications; 0 no match conditions; and 2 errors. The algorithm was then tested on 3 noisy images, using the training data from 12 noise–free images, and the result was 100% correct classification.

Finally, using the results of the 3 previous tests, the Acufex test classifies the 11 Acufex instruments using a combination of a BMD SDA–generated classification function and a structural matching algorithm. Two results are available from each Acufex test: a best match which tries to maximize the number of correct classifications or a clinically acceptable match which tries to minimize the number of misclassifications.

Two performance estimation methods were used for the Acufex test, the $\pi$ test with $P=4$ and the resubstitution (R) test. Using $P=4$ instead of $P=2$ in the $\pi$ test results in 4 times as many training sessions but would give a less pessimistic (lower–bound) performance estimate. Since the Acufex test had the poorest performance of all tests, the R test was done to give an optimistic (upper–bound) performance estimate. The data set for both estimation methods consisted of 32 noise–free images. In addition, using the training data from the R test, 4 noisy prespecified compartment
images and 4 noise-free random compartment images were tested on the Acufex test to observe the effects of lighting noise and random compartments independently. The results of all the tests applied to the Acufex test are listed in Table 5.2.

Table 5.2 Results of tests on Acufex algorithm.

<table>
<thead>
<tr>
<th>Test Type</th>
<th>Max Corr</th>
<th>Clin Accept</th>
<th>Misclass</th>
<th>No Match</th>
<th>Corr Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resubstitution</td>
<td>2.8%</td>
<td>0.3%</td>
<td>2.8%</td>
<td>94.9%</td>
<td>90.6%</td>
</tr>
<tr>
<td>Noisy Images</td>
<td>4.5%</td>
<td>0.9%</td>
<td>4.5%</td>
<td>89.8%</td>
<td>85.2%</td>
</tr>
<tr>
<td>Free compartment</td>
<td>6.8%</td>
<td>3.4%</td>
<td>8.0%</td>
<td>85.2%</td>
<td>75.0%</td>
</tr>
</tbody>
</table>

A confusion matrix containing the results of the low resolution test, the obturator test, the cartilage knife test and the Acufex test is given in Table 5.3.

5.3. Failures of Each Algorithm

From the results of tests of each algorithm presented in the last section, it is apparent that the algorithm that performed worst was the Acufex test. The other algorithms with less than perfect performance are the pose test and the Acufex length test. In this section, the results of each algorithm will be examined for their significance and an analysis will be given for the algorithms with less than 100%
Table 5.3 Confusion matrix of all arthroscopic instruments.

<table>
<thead>
<tr>
<th>Confusion Matrix - Rotation test</th>
</tr>
</thead>
<tbody>
<tr>
<td>L K T K P T S L S T T N R L L L L R R R R F H S M L B P B P</td>
</tr>
<tr>
<td>2 2 1 C C 5 3 S F S 0 0 S P 0 0 S P K K K K K O T O T</td>
</tr>
<tr>
<td>8 8 0 8 H H H H 1 1 1 1 1 1 1 1 1 1 1 1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SEE NOTE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
</tr>
<tr>
<td>32</td>
</tr>
<tr>
<td>25</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>31</td>
</tr>
<tr>
<td>25</td>
</tr>
<tr>
<td>32</td>
</tr>
<tr>
<td>26</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>SEE NOTE 4</td>
</tr>
<tr>
<td>SEE NOTE 3</td>
</tr>
<tr>
<td>SEE NOTE 2</td>
</tr>
<tr>
<td>SEE NOTE 1</td>
</tr>
<tr>
<td>SEE NOTE 1</td>
</tr>
<tr>
<td>SEE NOTE 1</td>
</tr>
</tbody>
</table>

| 12                |
| 12                |
| 12                |
| 12                |
| 12                |
| 12                |
| 12                |
| 12                |
| 12                |
| 12                |
| 12                |
| 12                |
| 12                |
| 12                |

NOTES: 1. Confusion is zero due to physical incompatibility of instruments and compartments
2. Confusion is assumed to be zero with program error checking
3. Handle width test - confusion is tested to be zero in these areas
4. Only results of Acufex test given in this sub-matrix
correct classification.

Comparing the results and the feature-space plots in Chapter 4 for the algorithms with 100% correct classification (the obturator test, the handle width test and the low-resolution test) it is evident that the features are well separated, and the algorithms performed as expected. For the cartilage knife test, the algorithm was able to separate two of the groups which were clustered closely in the feature space. However, the results were based on only 12 samples, and more samples may be required to properly test the efficacy of the algorithm in separating the two groups.

In the Acufex length test, the data seemed to be well separated except for two samples of the regular Acufex instruments which clustered near the long Acufex instruments (see Figure 4.12). The length of the regular Acufex instruments ranged from 184 to 203 pixels while the length of the long Acufex instruments ranged from 213 to 228 pixels. In the misclassified cases, 2 of the regular Acufex instruments have lengths of 209 and 211 pixels. In both cases, the instruments, LSS and LBP, were in the same compartment immediately on the right of the obturators. The average lengths of the two instruments over 12 training images were 197.8 and 196.2 pixels. It is not known why the lengths were significantly longer in those two cases. However, these two errors should not cause a misclassification in future if error checks are implemented to ensure that long and regular Acufex instruments are in their correct compartment.

In the Acufex pose test, the results (99.2% correct classification, no error and 1 no match condition) show that this test is acceptably reliable. In the case of the no-match condition, the edge angles on both sides of the handle were found to be the same, such that no decision was made. This error was probably due to the instrument handle being partially closed, for some unknown reason. The range of angle differences in all other handle measurements ranged from 2.83 to 8.67 degrees. If the
wrong instrument is in the compartment, or the handle is partially closed such that the angle difference is less than 2.5 degrees, the algorithm reports an error; similarly, an error is reported if the angle difference is greater than 10.0 degrees. These error-checking measures should prevent confusion with some of the other instruments.

Of the algorithms developed, the Acufex test is the most complicated, and has the most difficult job of differentiating among a number of similar Acufex instruments; thus, it is not surprising that it has the poorest performance. The errors in this algorithm can be described as fundamental errors and selection errors. Fundamental errors are errors that the algorithm cannot handle and usually results in no matches or errors in the best match case. A no match can be caused either (1) by the BMD features being incorrect such that the correct instrument is not selected for structural matching, or (2) by some of the features for the structural matching process being sufficiently different to prevent a match. In most of the no matches tested, the latter was the major cause. Errors in the best match case were primarily caused by similar instruments matching with the structural features of the other instruments. These errors occurred most frequently with the pairs of LPS/RPS and LBP/RBP where the structural differences are very minor. The errors due to structural similarities may be difficult to avoid since even human subjects have trouble separating these similar instruments when examining an image of the instruments. It should be noted that the algorithm was able to recover from the fundamental error of detecting a wrong first or last corner in most of the tests.

Selection errors are considered to be differences which cause the clinically acceptable match to differ from the best match. The selection errors may be due to an error in the BMD classification function output or in the structural pattern recognition algorithm. A listing of the causes of these errors is given in Table 5.4. It can be seen that a significant portion of the no matches in the clinically acceptable
results is due to incorrect BMD results with correct structural match. A smaller but still significant portion of the no matches is due to the inverse of the above – correct BMD results with incorrect structural match. Again, a number of the incorrect structural matches are due to structurally similar instrument pairs, as mentioned before.

Table 5.4 Types of errors in structural match.

**Resubstitution Test:**
- no match and BMD correct: 6
- no match and BMD incorrect: 3
- no match – BMD and struct. match wrong: 1
- BMD wrong and struct. match correct: 14
- BMD correct and struct. match incorrect: 7

**π Test:**
- no match and BMD correct: 6
- no match and BMD incorrect: 4
- BMD wrong and struct. match correct: 26
- BMD correct and struct. match wrong: 12
- error – BMD and struct. match wrong: 3
- no match – BMD and struct. match wrong: 1

An analysis of the performance of the structural matching algorithm was done by examining the goodness of match (the percentage of corners matched) in the cases of the R and π methods. The results are shown in Figure 5.2. It can be seen that the performance of the structural matching algorithm did not degrade significantly from the resubstitution test to the π test. Hence, the decrease in performance in the π test may be due to the BMD features.

5.4. Improvements of Recognition Algorithms
Figure 5.2 Histogram of goodness of structural match.
In discussing the failures of the algorithms, it was noted that two of the groups in the cartilage knife test were clustered closely in the feature space and that the separation of the two groups may not be as reliable even though test results had indicated a 100% estimated recognition accuracy. In order to improve the recognition reliability of the cartilage knife test, the distinctive tips of the cartilage knives may be used as features for discrimination. Since the knife tips consist of several significant corners, the algorithms for the Acufex test may be used.

Although the cartilage knives, with the exception of the hook probe, have the same general shape at the tip, there appear to be sufficient differences in the size and angle of the tip for use with the Acufex test algorithms. In fact, it may be sufficient to apply only the BMD features to improve the recognition reliability. The structural matching algorithm can also be used but it will increase the algorithm processing time considerably.

Another algorithm that may be improved is the structural matching algorithm. The drawback of using quantitative features for each significant corner, instead of Pavlidis's qualitative features, is that the matching algorithm may be sensitive to large variations in the features. To reduce this sensitivity, the structural matching process should be examined in detail to determine the conditions which cause correct matches to fail. One possible improvement may be to assign a weighting or scoring scheme to the feature matching procedure such that one bad feature data does not produce an immediate rejection in the match.

5.5. Implementation and Optimization

Beside the recognition algorithms, other practical problems exist in the vision system which may result in degraded performance of the system. This section discusses these practical problems as well as methods to optimize the system towards
implementation in the operating room.

5.5.1 Practical Problems

Some of the practical problems in the implementation of the vision system are noise due to lighting, noise due to blood, tissue fragments, or other material from the surgical site, noise due to boundary effects, inconsistencies due to manufacturing variations, and disturbances due to geometric effects.

There are two sources of noise due to lighting, room light and backlighting. Room lighting had been investigated in the noisy images where overhead lights were left on during digitization of the images. The resulting images showed cavities in the silhouettes of the instruments due to reflections. The amount and location of the reflections will generally vary according to the direction of the overhead light source. The overhead lighting noise can be eliminated in most cases by adjusting the binary threshold of the images. Figure 5.3 shows a gray-level image and its intensity histogram. A suitable threshold for the image is 55. If the threshold is too high, the image would be filled with noise, as shown in Figure 5.4. If the threshold is too high the silhouettes of the instruments would begin to thin out, see Figure 5.5. At present, the threshold is selected manually to obtain images that are free of noise cavities. Automatic threshold selection techniques exist (Weszka, 1978) but evaluation and implementation of such techniques was beyond the scope of the thesis.

The effect of noise due to non-uniform backlighting has not been tested since it has not presented any major problems. Nevertheless, backlighting is an important part of binary vision since inconsistencies in backlighting would seriously affect the binary threshold selection and create noises in the image. This type of noise was encountered in the evaluation of the Oculus 100 digitizer.
Figure 5.3 Noisy Image with histogram of intensity.
Figure 5.4 Binary digitized noisy image: (top) threshold = 55; (bottom) threshold = 65.
Figure 5.5 Binary digitized noisy image: (top) threshold = 45; (bottom) threshold = 35.
The type and level of noise due to tissue or other material remaining on the instruments following use in the surgical site have been tested using actual operating room samples. Samples of two types of fluids were collected and tested on a small plexiglass tray. The fluids were: (1) the residue from instruments that had been inserted into the knee, and (2) a purple-coloured skin preparation solution, tincture of chlororhexadene, used at the start of the surgical procedure. There were no blood or small tissue fragment found in the fluid from the knee, i.e. it was primarily saline solution. The tray of fluids was then dried and examined in low resolution and high resolution.

In low resolution, the fluid tray was placed near the centre of the instrument tray with the trocar sleeves placed across the residue of the fluid, as shown in Figure 5.6. Both types of fluids were found to be invisible in the low-resolution image and the shape of the trocar sleeves were not distorted in any way.

In high resolution, the fluid tray was again placed near the centre of the FOV with the tips of several instruments lying across the fluid residue as shown in Figure 5.7. The saline from the knee was invisible while the skin-preparation solution caused numerous spots to appear along the edge of the compartment and also near the instruments. The problem may not be a major concern since the skin preparation solution is not used with the high-resolution instruments, but this should be verified in actual clinical trials.

The effect of the boundaries of the instrument compartments were tested using the small plexiglass tray mentioned above (see Figure 5.8). In low resolution, see Figure 5.9, the entire tray was invisible and did not cause any distortion of the instruments placed in the tray. In high resolution, however, the boundaries created a dark band along certain portions, while being invisible in others, see Figure 5.10. There are two probable causes for the dark band, uneven lighting and extreme
Figure 5.6 Noise test of fluids in low resolution.
Figure 5.7 Noise test of fluids in high resolution.
curvature. Referring to Figure 5.9, the boundaries were visible on the right side of the image which had a slightly lower backlight intensity, due to the proximity to the edge of the backlight projector; on the other hand, the boundaries were invisible on the left-hand side where the backlighting is slightly stronger. The effects of the tray curvature can be seen in the divider which separates the two compartments in the plexiglass tray. In Figure 5.10, the upper part of the divider with the lesser curvature was invisible while the lower part with the higher curvature was visible.

Differences in instruments due to manufacturing variations are difficult to predict; hence, it would be difficult to test the algorithm's sensitivity to such variations. The least sensitive algorithm is the obturator algorithm since it measures two very distinct shapes. The sensitivity of the low-resolution and cartilage knife algorithms can be tested easily by extracting the features and comparing with the other instruments in the feature space. For the Acufex test, manufacturing variations could lead to instruments that look entirely different and it is not certain that these instruments could be recognized by the structural matching algorithm. Nevertheless, if the instruments satisfy the basic criterion for the structural matching algorithm, which is the presence of distinct corners, then there is a very good chance these instrument would be recognized.

5.5.2 Optimization

The algorithms presented in this thesis have been developed and tested on the UBC Electrical Engineering VAX-11/750 Superminicomputer (see Figure 5.11). Since it would not be economical to incorporate such an expensive computer into the vision system, an IBM PC microcomputer, as described in chapter 2, was chosen for the system. Some of the algorithms on the VAX, written in VAX FORTRAN, were transferred to an IBM PC and modified to compile using MS-Fortran compiler to obtain a time performance estimate of the algorithms using the IBM PC.
Figure 5.8 Sample instrument tray.
Figure 5.9 Boundary test in low resolution.
Figure 5.10 Boundary test in high resolution
Figure 5.11 UBC VAX set-up.
In modifying the programs to be compatible with MS-Fortran, an assembly-language function was written to increase the image pixel access speed and decrease the image memory requirements. The binary images on the VAX-11/750 use 2 bytes per pixel while the IBM PC uses 1 bit per pixel.

The Acufex and low-resolution algorithms, representative of the structural matching, boundary tracking and the scanning type of data processing used in the algorithms developed, were tested on the IBM PC/XT microcomputer and the IBM PC/AT supermicrocomputer. As well, tests of two simple operations, such as multiplication and memory reference performance, were conducted. The results of the comparative speed tests for different computers are listed in Table 5.5.

Table 5.5 Time comparison results.

<table>
<thead>
<tr>
<th></th>
<th>VAX</th>
<th>PC/AT</th>
<th>PC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real*8 multiplications</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10,000</td>
<td>0.12</td>
<td>1.32</td>
<td>14.74 sec.</td>
</tr>
<tr>
<td>50,000</td>
<td>0.60</td>
<td>5.49</td>
<td>73.48 sec.</td>
</tr>
<tr>
<td>Image Read - 10 full scans of 512 x 512 pixels</td>
<td>48.32</td>
<td>132.17</td>
<td>410.63 sec.</td>
</tr>
<tr>
<td>Acufex Test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>test on single image</td>
<td>2.65</td>
<td>18.92</td>
<td>53.79 sec.</td>
</tr>
<tr>
<td>range for 5 images</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>max</td>
<td>2.65</td>
<td>18.92</td>
<td>53.79</td>
</tr>
<tr>
<td>min</td>
<td>1.82</td>
<td>13.04</td>
<td>38.34</td>
</tr>
<tr>
<td>average time</td>
<td>2.23</td>
<td>15.51</td>
<td>45.17</td>
</tr>
<tr>
<td>Low Resolution Test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>test on single image</td>
<td>0.76</td>
<td>2.15</td>
<td>6.66 sec.</td>
</tr>
<tr>
<td>range for 5 images</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>max</td>
<td>0.81</td>
<td>2.20</td>
<td>6.88</td>
</tr>
<tr>
<td>min</td>
<td>0.50</td>
<td>2.15</td>
<td>6.66</td>
</tr>
<tr>
<td>average</td>
<td>0.67</td>
<td>2.16</td>
<td>6.80</td>
</tr>
</tbody>
</table>
In numerical calculations such as REAL*8 multiplications, the VAX-11/750 was over 10 times faster than the IBM PC/AT and over 100 times faster than the IBM PC/XT. In memory reference operations such as reading the image pixels, the VAX-11/750, using 512 KB of image memory, was 2.7 times faster than the IBM PC/AT and 8.5 times faster than the IBM PC/XT.

The Acufex algorithm required both memory reference for boundary tracing and numerical calculation for structural matching. Tests of the Acufex algorithm showed that the VAX 11/750 was approximately 7 times faster than the IBM PC/AT and 20 times faster than the IBM PC/XT. The average CPU times were 2.23s, 15.51s, and 45.17s for the VAX, PC/AT and PC/XT respectively. The low-resolution test required mostly image pixel references and test results showed that the VAX-11/750 was approximately 3.0 times faster than the IBM PC/AT and 10 times faster than the IBM PC/XT. The average CPU times were 0.67s, 2.16s and 6.8s for the VAX, PC/AT and PC/XT respectively.

Although the VAX-11/750 is much faster than the IBM PC, the speed of the IBM PC/AT may be sufficient for this project. The complicated Acufex algorithm required, on the average, 15 seconds to execute while the line scan type low-resolution algorithm required only 2 seconds using an IBM PC/AT. Hence, an estimate of the time required for the overall vision system would be approximately 20 to 25 seconds.

5.6. Final Results of the Recognition Algorithm

Evaluation of the algorithms yielded recognition accuracies of 100% for the low-resolution test, the obturator test and the cartilage knife test. The upper- and lower-bound accuracy estimates for the Acufex instruments were 94.9% and 92.6% for the best-case match, and 90.6% and 85.2% for the clinically acceptable match.
To obtain an overall classification accuracy for the vision system, the individual recognition accuracies were combined with the probability of usage for each group. From observing four videotaped arthroscopic cases performed by one surgeon, with case durations ranging from 12 to 18 min., the approximate instrument usage by group is shown in Table 5.6. Weighting the accuracies according to Table 5.6 yields an overall recognition accuracy of 99.1% correct, 0.69% no match and 0.16% misclassification. Table 5.7 shows the recognition accuracies for different weighting percentages of the Acufex instruments. Even at 40% Acufex usage, the correct classification is estimated to be 93.3%.

To obtain an estimate of the frequency of the assistance required from the nursing personnel, it was assumed that an Acufex request occurred every 2 min. With that assumption, there would be 7.5 Acufex requests in a 15 minute period. The worst case recognition results for the 15-min. period are 6.38 correct classifications, 1.04 no-match conditions, and 0.06 misclassifications. The above results indicate that there will be approximately 1 instrument which could not be recognized within the 15-min. period and hence would require the assistance of the surgical nursing personnel. Since this result is within the design criteria given in the clinical requirements, the estimated accuracy of the vision system is considered to be suitable for clinical use.
Table 5.6 Arthroscopic instrument usage probability.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cartilage Knives</td>
<td>23.7 %</td>
</tr>
<tr>
<td>Obturator</td>
<td>44.1 %</td>
</tr>
<tr>
<td>Low Resolution Instruments</td>
<td>27.1 %</td>
</tr>
<tr>
<td>Acufex Instruments</td>
<td>5.1 %</td>
</tr>
</tbody>
</table>

Table 5.7 Overall recognition accuracy.

<table>
<thead>
<tr>
<th>Acufex Usage</th>
<th>correct</th>
<th>no match</th>
<th>misclass</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 %</td>
<td>99.1 %</td>
<td>0.7 %</td>
<td>0.2 %</td>
</tr>
<tr>
<td>10.0 %</td>
<td>98.3 %</td>
<td>1.4 %</td>
<td>0.3 %</td>
</tr>
<tr>
<td>20.0 %</td>
<td>96.7 %</td>
<td>2.7 %</td>
<td>0.6 %</td>
</tr>
<tr>
<td>30.0 %</td>
<td>95.0 %</td>
<td>4.1 %</td>
<td>0.9 %</td>
</tr>
<tr>
<td>40.0 %</td>
<td>93.3 %</td>
<td>5.4 %</td>
<td>1.2 %</td>
</tr>
</tbody>
</table>
CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

The development and evaluation of a vision system for recognizing arthroscopic instruments has been presented in this thesis. Also, a robotic system which uses the vision system for performing surgical instrument passing duties in an operating room has been discussed.

A two camera, dual-resolution approach was employed by the vision system to recognize the arthroscopic instruments. A camera, digitized to 512 x 512 pixels, was used to view all the instruments at a low resolution while another camera, also digitized to 512 x 512 pixels, was used to focus on a small area containing the fine details necessary to discriminate the instruments which are extremely similar.

Four algorithms were developed to recognized the four different groups in the set of 36 arthroscopic instruments. For three of the algorithms; the low-resolution algorithm, the obturator algorithm and the cartilage-knife algorithm, simple boundary scanning techniques were adequate to obtain features such as length, width and minimum angle of the instruments. These simple features were able to discriminate the three groups of instruments, with an estimated recognition accuracy of 100 percent for each algorithm.

A combination of statistical and structural features were found to be necessary to recognize the fourth group of Acufex instruments. The statistical features consisted of 6 angles and distances taken from the first and last significant corners found after boundary tracing; the statistical features were then matched using a BMD-generated discriminant function. The best scores resulting from the discriminant function were
subsequently matched using the structural-matching algorithm. The significant corners between the first and the last corner formed the structural features and were sequentially matched to the reference corners. The Acufex algorithm provided two possible matches: one which maximized the number of correct matches, and another which minimized the number of incorrect matches. The latter matching method was used to generate clinically acceptable recognition results. Using a resubstitution test to obtain an upper bound and a $\pi$ test to obtain a lower bound on the recognition accuracy, the results were 94.9% and 92.6% for the maximally correct matching method, and 90.6% and 85.2% for the clinically acceptable matching method.

An overall estimate of the recognition accuracy, incorporating results of all four recognition algorithms, and weighted according to assumptions concerning instrument usage, gave a recognition accuracy of 99.1% correct classification, 0.69% no match condition, and 0.16% misclassification. As well, simple calculations showed that the frequency of requests for assistance is below the acceptable limit defined by operating-room personnel in a clinical questionnaire.

Based on the above results, it can be concluded that the primary objective of the thesis, the development and evaluation of a clinically acceptable vision system for recognition of arthroscopic surgical instruments, has been satisfied.

When practical vision system tests were performed involving overhead lighting, a sample plexiglass tray, and contamination or degradation of the surgical instruments, it was found that overhead lighting produced bright reflections on the instruments, but the reflections were compensated by increasing the binary threshold to give continuous boundaries. Tests using the sample plexiglass tray showed that the tray was invisible in low-resolution. In high-resolution however, portions of the compartment boundaries with high curvature created dark lines in the image. Samples of fluid from the surgical site were collected in the tray and examined under high and low-resolution. In
low-resolution, the saline and the dark skin-preparation solution were transparent. In high-resolution, the saline was transparent but the dark skin-preparation solution generated some dark spot in the image. This may not be a problem since the solution is not used with the instruments in the high-resolution field.

Computer time performance comparisons were done for the VAX-11/750, IBM PC/AT and IBM PC/XT using the vision programs. The VAX-11/750 was 3 to 7 times faster than the IBM PC/AT and 10 to 20 times faster that the IBM PC/XT. The average execution times for the vision programs tested were 15.51s and 2.16s using the IBM PC/AT.

Based on these performance results, the vision system developed in this thesis was capable of recognizing the arthroscopic surgical instruments within a clinically acceptable recognition accuracy. It was predicted that the algorithms developed will execute in real-time with an anticipated cycle time of 20–25 seconds for the overall vision system.

A complete system outline was given for an instrument-passing robot for use with the vision system. Recommendations for individual system components were given and evaluations were done for the two primary vision system components: the solid state camera and the binary digitizer. Although the solid state camera did not perform as well as a high-quality vidicon camera in the resolution test, it did produce comparable high- and low-resolution images that were used in the evaluation of the vision algorithms.

A very simple payback analysis was done based on the total system cost and the average salary of a scrub nurse. The payback time was calculated to be 1.6 years, assuming a 20 minutes labour savings in 35 minutes of nursing time. This payback period is within the guidelines for introduction of industrial robotics; hence, assuming
similar guidelines for the health care industry, the instrument-passing robot appears to be economically feasible for use in hospitals.

Several of the major problems associated with implementing a real-time vision system for clinical use have been discussed. Equipment recommendations for the instrument-passing robot and a simple payback analysis were presented. Based on these results, it can be concluded that the three secondary objectives of this thesis have been adequately met.

Recommendations for future work

The results of this thesis indicate that it is technically feasible to implement an surgical instrument-passing robot for arthroscopies. However, much work remains to be done before actual clinical implementation is possible. An extensive study should be done on operating-room (OR) activities during such surgical procedures to ensure that the presence of a robot and a vision system would not hinder the effectiveness of the operating room staff. Also, issues concerning safety should also be investigated thoroughly, especially concerning the transfer of instruments between the robot and the surgeon.

One area of improvement in the vision system is the structural matching algorithm. The possibility of weighting the individual features of each structural feature should be studied to improve the matching accuracy. At present, the structural feature selection process is performed manually; however, the selection can be simplified or partially automated by a computer program which looks for similar feature in reference inputs. Such a program would be desirable if the vision system is to be used commercially. As well, the structural matching algorithm should be tested on the group of cartilage knives to reinforce the decision from the single width feature. Finally, sample instrument trays using different moulding methods should be tested to find a
pattern that can be used in the high-resolution field.

In the evaluation of the Oculus 100 digitizer, it was found that the digitizer did not produce images with comparable sharpness as the IP-512 gray-level digitizer. However, a more extensive software and hardware evaluation should be done to determine its actual performance.

By improving the reliability of the vision system and the overall robotic system, it may be possible to introduce a cost effective instrument-passing robot into an operating room in the near future to reduce the cost of health care delivery.
References


Appendix A Questionnaire on Clinical Requirements

Biomedical Engineering Department
Computer Vision Survey

Your initials (optional) ____________________________

What is your experience as a circulating nurse? (number of procedures)
X - 26 years x 280 days x 0.5 = 35,100; Y = 10 years x 280 days x 0.5 = 23,000

What is your experience as a scrub nurse? (procedures) ____________________________

Does your surgical nursing experience cover a wide variety of procedures (Y/N) ________

If the response is 'No' above, what specific procedures are you experienced in

1. On the average, what percentage of the scrub nurse's activities in the OR is spent where her primary task is to pass surgical instruments to the surgeon (percent) ________

2. If an instrument-passing robot is to be used in the OR, would assistance from the surgical nursing personnel be available always available during setup: to drape robot (Y/N) ________

   to unpack and/or to lay out instruments (Y/N) ________

   to attach sterile components to robot (Y/N) ________

   to select activities from a menu on a computer display (Y/N) ________

   to perform other minor activities (Y/N) ________

   to unpack and/or to lay out instruments (Y/N) ________

   to attach sterile components to robot (Y/N) ________

   to select activities from a menu on a computer display (Y/N) ________

   to perform other minor activities (Y/N) ________

   during surgery: to press selected keys once in a while (if necessary) (Y/N) ________

   to remove loose tissue fragments from instruments or instrument tray if requested by computer system (Y/N) ________

   to take other remedial action requested by computer system (Y/N) ________

   post surgery: to initiate shut down of computer system (Y/N) ________

   to remove instruments (Y/N) ________

   to undrape robot (Y/N) ________

3. If minor problems arose during surgery, would it, in your opinion, be possible for the surgeon to take minor corrective actions requested by computer system (Y/N) ________

   Generally NO! Do not want surgeon to lose visual contact with surgical site.

148
Questions on expected level of performance from robot:

1. Approx. what percentage of time does the scrub nurse pass the wrong instrument to the surgeon, including incorrect requests from the surgeon 1 or 2 per case (less than 1%)

   What percentage of the above is due to surgeon's error (less than 1%) __________

2. In view of the above, what is an acceptable percentage of time for a robot to pass the wrong instrument to the surgeon, including incorrect requests from the surgeon

   without nursing assistance __________

   with nursing assistance 1 or 2 per case __________

3. What, in your opinion, is the maximum allowable number of occurrences in which the robot passes an incorrect instrument to the surgeon before surgeon becomes irritated (0 or number) 1 or 2 per case (not 3 or 4) depending on surgeon

   Is this rate of error acceptable to most surgeons (please explain) No (but 1 or 2 errors may be tolerated in arthroscopy)

   (average of 20 - 30 passes per arthroscopy)

   Would greater errors be tolerated? If yes, under what circumstances?

4. During a procedure, is there usually a circulating nurse available with sufficient time and skill to respond to an alarm by entering a few key strokes to correct any mistakes. Yes circulating nurse only.

5. If the computer becomes confused and requires assistance in verifying one or two instruments from the circulating nurse during surgery, how often would this be acceptable (never or no more than once every 7 min.)

   Once every 20-30 minutes.

   If an acceptable level has been given above, at what level would it become intolerable or unacceptable Once every 5 to 10 minutes.
6. Would it take more, same or less time or skill for a scrub nurse to lay out the instruments according to a computer generated map?

More time initially, but may be less as nurse gains experience.

to randomly place the instruments in suitably sized compartments

Less time.

to position preloaded trays

Less time.

7. In the event that nursing assistance is requested, would it require more, same or less effort if the instruments have been placed in an order specified by the nursing personnel?

Same

Same

Same
Some general questions on arthroscopy instrument usage:

1. What is your experience in arthroscopy procedures (approx.):
   - 0 to 10
   - 11 to 30
   - over 30

2. Does the surgeon normally use all or a percentage of the instruments (range):
   - 25%

3. Approximately how many instruments does the surgeon use at one time (range):
   - 2 to 3 instruments

4. Approx. how many instruments does the surgeon keep near him/her during the procedure (range):
   - 1 or 2 instruments

5. Approx. what percentage of time can the scrub nurse anticipate the needs of the surgeon (percent range):
   - 75% with TV, 25% without TV

6. Are there any instrument(s) (or set of instruments) that are difficult to identify (Y/N, give names if possible):
   - Acufex, 3.8 mm or 5.0 mm trocars

   Are the difficulties due to unfamiliarity with instruments, similarity to other instruments, or other reasons:
   - Unfamiliarity and similarity

7. Assuming that the scrub nurse could not anticipate the surgeon's needs, what is the average time from the surgeon's request to placing the requested instrument in the surgeon's hand:
   - for easily identifiable instruments (seconds)
   - 5
   - for the more difficult instruments (seconds)
   - 30 to 40
   - if assembly of parts is required

8. Does the scrub nurse do anything differently to help identify the difficult instruments during setup or strategically place instruments on the stand:
   - Visual and operational check
9. What is the maximum delay in passing the instrument without jeopardizing the patient's treatment or irritating the surgeon (seconds) 

   not significant in terms of patient's treatment; 2-30s depending on surgeon

10. How often does the surgeon ask for multiple instruments (percent) 

    Usually asks scrub nurse to give or have ready

11. Do you know of any time performance studies or related work for instrument passing between scrub nurse and the surgeon (journal and approx. date, or name of text) RIL

Any other comments or suggestions

   In arthroscopies at SDCC, 25% of cases may have a resident in addition to
   surgeon (up to 50% for other areas at VGR).
robotic system cost is $29,000

present salary of scrub nurse $16.34 per hr.

assume 20 minutes per case time of 35 minutes may be replaced by the robot.

labour savings = 20 / 35 = 57%

for a 37.5 hours work week, the savings per week is

savings per week = $16.34 \times 57\% \times 37.5 = $349.27

payback period = system cost / savings per year

= $29,000 / ($349.27 \times 52)

= 1.597 years = 1.6 years
Appendix C Program Pseudo-code

Cartilage Knife Test

Program Knife
read image data
read reference data
for each knife compartment
begin
get start position and scan direction
scan every 4th line until end
record edge locations
obtain edge locations for
  1. 7th line from end
  2. 12th line from end
use edge locations to calculate width
classify knife
end
stop

Obturator Test

Program Obturator
read image data
read reference data
for each obturator compartment
begin
get start position and scan direction
scan every 4th line until end
record edge locations
obtain edge locations for
  1. 7th line from end
  2. 12th line from end
use edge location to calculate width
obtain edge location for
  1. 8th line from end
trace around boundary starting at 8th line
calculate angles between 5 pixels forward &
  5 pixels backward
find minimum angle
classify obturator
end
stop
Acufex Pose and Length Tests

Program Acufex Pose
read image
for each compartment
  scan line and record all edge locations
  repeat every 4th line until tip end
    if tip is less than 3 pixels wide
      - then tip found
  obtain edge locations for
    1. 7th line from 1st scanned line
    2. 12th line from 1st scanned line
  calculate handle angles on either side
  obtain edge location for
    1. 7th tip location
    2. 12th tip location
  calculate angle at tip
  calculate pose from angles
  obtain first and last scanned lines
  calculate length of Acufex instrument
end
stop
Low Resolution Test

Program LRTEST
read image
read references
for each compartment
begin
scan line and record all edge locations
repeat every 4th line until end
calculate length
calculate width
classify instrument
end
stop

Acufex Width Test

Program Acufex width
read image data
for each Acufex compartment
begin
get start position and scan direction
scan every 4th line until 30 lines scanned
calculate width using trigonometric approx.
end
stop
Acufex Instruments Test

Program Acufex test
read image data
read reference data
for each Acufex compartment
begin
  get start position & scan directions
  trace around boundary
  calculate angles between 5 pixels forward and 5
  pixels backwards
  detect peaks in angles
  calculate BMD features
  test BMD discriminant function
  check length of Acufex instrument
  for each possible match
  begin
    match reference features to object features
    until all reference features are matched
    or all object features have been tried.
    calculate score
  end
  if no match, then move first or last peak
  try match again starting with BMD features
end
end
stop