DESIGN OF AN AUTOMATED SLIDE CAPTURING SYSTEM WITH A DSP-BASED AUTOMATIC FEATURES EXTRACTION OF THINPREP® IMAGES FOR CERVICAL CANCER

by

NOR RIZUAN MAT NOOR

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

April 2008

ACKNOWLEDGEMENTS

All praise to the almighty Allah, the most merciful and most benevolent for giving me the strength and help in completing this research. Here, I would like to take this opportunity to present my special thanks and appreciation to those who directly and indirectly involved in the accomplishment of this research work.

Firstly, I would like to express my sincere gratitude to my supervisor, Dr. Nor Ashidi Mat Isa and my co-supervisor, Assoc. Prof. Dr. Mohd. Yusoff Mashor for their guideline, support, great ideas and suggestions. With their help, this research work has improved tremendously. They also helped me to recognize my weaknesses and gave numerous advices to help me in completion of this research thesis successfully.

A big thank you goes to pathologist, Prof. Dr. Nor Hayati Othman and other staffs of Pathology Department, Hospital Universiti Sains Malaysia for providing supports, equipments and slides for data collecting stage.

I also want to express my personal thanks to all staffs of School of Electrical & Electronic Engineering, Universiti Sains Malaysia, Engineering Campus and staffs of Engineering Centre, Universiti Malaysia Perlis that gave full cooperation and support for me to successfully finish this study. Last but not least, my deepest thanks to my parents, friends and everyone who was indirectly involved and helped me in completion of this research thesis.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS ii

TABLE OF CONTENTS iii

LIST OF TABLES viii

LIST OF FIGURES x

LIST OF ABBREVIATIONS xvi

ABSTRAK xviii

ABSTRACT xx

CHAPTER 1 - INTRODUCTION

1.1 Introduction 1

1.2 Computer Assisted Systems for Cervical Cancer Screening 2

1.3 Problem Statement 3

1.4 Objectives of Research 4

1.5 Scope of Research 6

1.6 Thesis Outline 8

CHAPTER 2 - LITERATURE REVIEW

2.1 Introduction 11

2.2 Cervical Cancer 11

2.2.1 Cervical Anatomy 12

2.2.2 Cervical Microscopic Anatomy 13

2.2.3 Key Factor: *Human Papilloma Virus* (HPV) 14

2.2.4 Cancer of Cervix; Pre-Cancerous & Cancerous 16

2.2.5 Pap Smear Test and ThinPrep® 20

2.3 Digital Image Processing 26

2.3.1 Digital Image 27

2.3.2 Image Sampling & Quantization 31
2.3.3 Digital Image Processing Techniques 34
2.3.3.1 Segmentation 34
2.3.3.2 Features Extraction 37
2.4 Digital Signal Processor 39
2.5 Image Capturing Applications Using Microscope 40
2.6 DSP Applications to Image & Video Processing 42
2.7 Summary 44

CHAPTER 3 - PC-BASED AUTOMATIC GLASS SLIDE CAPTURING SYSTEM (AutoCapture System)

3.1 Introduction 46
3.2 Overall Diagram of PC-Based Automatic Glass Slide Capturing System (AutoCapture System) 46
3.3 AutoCapture System - Hardware and Interfacing 48
3.3.1 Digital FireWire Camera 48
3.3.1.1 Hardware Features 48
3.3.1.2 Software Features 50
3.3.2 Automated Microscope 52
3.3.2.1 Hardware Features 52
3.3.2.2 Software Features 53
3.4 AutoCapture System - Software 54
3.4.1 Image Capturing 55
3.4.1.1 Single Image Capturing 55
3.4.1.2 Automatic Glass Slide Capturing 57
3.4.2 Image Retrieval/Displaying 60
3.4.2.1 Thumbnail Based on Capture Order 60
3.4.2.2 Thumbnail Based on Most Suspected Images 61
3.4.2.3 100X – 400X Magnification Map 63

3.4.2.4 Inter-Image Cells Display 64
3.5 Summary 65
CHAPTER 4 - AUTOMATIC FEATURES EXTRACTION OF THINPREP® IMAGE
4.1 Introduction 67
4.2 Texas Instrument DSP – TMS320C6416 DSP Starter Kit (DSK) 67
4.3 Features Extraction of ThinPrep® Image 72
4.3.1 Threshold Value Finding 75
4.3.2 Centroid Location Finding 80
4.3.3 Segmentation & Features Extraction 83
4.4 Implementation of Features Extraction of ThinPrep® Image on TMS320C6416 DSP 89
4.4.1 Image Information Transferring 91
4.4.2 Code Optimizations 93
4.4.2.1 Variables and Functions (Local versus Global) 93
4.4.2.2 Level-2 (L2) Cache-Memory Allocation 94
4.4.2.3 Compiler Options 94
4.4.2.4 Data Types, Sizes and Ranges Consideration 96
4.4.2.5 Loop Unrolling 98
4.4.2.6 Intrinsic Operators 99
4.4.3 Code Profiling 100
4.4.3.1 Borland® C++ Builder 101
4.4.3.2 Code Composer Studio 101
4.4.4 Correlation Test 101
4.4.5 Execution Time Comparison 103
4.5 Summary 103
CHAPTER 5 - RESULT & DISCUSSION
5.1 Introduction 106
5.2 AutoCapture System 106

5.2.1 Hardware 107
5.2.2 Software 108
5.2.2.1 Image Capturing 109
5.2.2.1.1 Single Image Capturing 109
5.2.2.1.2 Automatic Glass Slide Capturing 109
5.2.2.2 Image Selection 111
5.2.2.3 Image Retrieval/Displaying 117
5.2.2.3.1 Thumbnail Based on Capture Order 117
5.2.2.3.2 Thumbnail Based on Most Suspected Images 118
5.2.2.3.3 100X – 400X Magnification Map 119
5.2.2.3.4 Inter-image Cells Display 124
5.2.3 Discussion 126
5.3 Automatic Features Extraction 127
5.3.1 Image Selection 128
5.3.2 Image Segmentation & Features Extraction 133
5.3.2.1 Threshold Values, β_{NC}, β_{CB} 133
5.3.2.2 Centroid Location, ... 134
5.3.2.3 Segmentation 138
5.3.2.4 Features Extraction 141
5.3.2.5 Discussion 143
5.3.3 Execution Time 148
5.3.3.1 Effect of Compiler Options Settings (Code Composer Studio) 148
5.3.3.2 Effect of Loop Unrolling 150
5.3.3.3 Effect of Intrinsic Operators 151
5.3.3.4 Execution Time Comparison (With and Without Code Optimization Techniques-DSP System) 153
5.3.3.5 Execution Time Comparison (The PC-based System and the DSP System) 156

5.3.3.6 Discussion 159

5.4 Conclusion 163

CHAPTER 6 - CONCLUSION

6.1 Conclusion 165

6.2 Suggestions for Future Work 168

REFERENCES 170

APPENDIX

APPENDIX A FireWire Digital Camera PixeLink PL-A662 180

APPENDIX B Automated Microscope Leica DM LA 182

APPENDIX C TMS320C6416 Data Sheet 183

APPENDIX D Automatic Features Extraction Results for the 213 Single Cell Images 184

APPENDIX E Execution Time Comparison for the 213 Single Cell Images 189

LIST OF PUBLICATIONS 194

LIST OF RECOGNITIONS 195

LIST OF TABLES

Table 2.1:	The Bethesda System (Spitzer *et al*., 2002)	18
Table 2.2:	Stage of Cervical Cancer (Pihie, 1998)	19
Table 3.1:	Packet Size at Various Clock Frequencies for Resolution of 640 x 480 and 1280 x 1024	49
Table 3.2:	Logical Microscope Components	54
Table 4.1:	Optimization Level of the Optimizing Compiler	96
Table 4.2:	Optimization between Speed of Execution versus Code Size	96
Table 4.3:	Data Types, Sizes and Ranges for Borland® C++ Builder versus TMS320C6416 DSP	97
Table 5.1:	Header File Generated During Automatic Glass Slide Capturing Process	110
Table 5.2:	400X Magnification Images Sorted in Capture Order.	117
Table 5.3:	400X Magnification Image Sorted in Nucleus – Cytoplasm Ratio	119
Table 5.4:	Final Centres and Threshold Values for ThinPrep® Image 1 to 10 for TMS320C6416 DSP System versus PC-Based System	134
Table 5.5:	Correlation Coefficient Result between the DSP System, the PC-Based System and the Clinical Data Set for the 213 Cervical Cell Images.	142
Table 5.6:	Features Extracted from ThinPrep® Image 1 to 10 by Using the DSP System.	142
Table 5.7:	Code Size, CPU Cycles and Time for Each Configuration of Compiler Options Settings	150
Table 5.8:	_add2, _mpy2 and _sub2 Intrinsic Operators Description.	153

Table 5.9: Average Time per Unit Pixel, t_{av} for the DSP System With and Without the Optimization Techniques for All the 213 Single Cell Images. 155

Table 5.10: Execution Time Comparison for the ThinPrep® Image 1 to 10 for the DSP System With and Without the Optimization Techniques. 156

Table 5.11: Average Time per Unit Pixel, t_{av} for the DSP System and the PC-based System for All the 213 Single Cell Images. 158

Table 5.12: Execution Time Comparison for the ThinPrep® Image 1 to 10 in the DSP System and the PC-based System. 158

LIST OF FIGURES

Figure 2.1: Women Reproductive System 12

Figure 2.2: Ectocervix and Endocervix (Arends, 1999) 12

Figure 2.3: Columnar Epithelium Cells 13

Figure 2.4: Squamous Epithelium Cells 14

Figure 2.5: Cells inside endocervix and ectocervix layers 15

Figure 2.6: Electron Photomicrograph of Human Papilloma Virus 15

Figure 2.7: Changes in Epithelium Cervix 18

Figure 2.8: Stage of Cervical Cancer (Azhar, 1994) 20

Figure 2.9: Obtaining Samples Using Ayre's Spatula - ectocervix 22

Figure 2.10: Obtaining Samples Using Cytobrush – endocervix 22

Figure 2.11: Cervex-Brush® 22

Figure 2.12: Obtaining Samples Using Cervex-Brush® - Both Endocervix & Ectocervix 22

Figure 2.13: Sensitivity and Specificity 23

Figure 2.14: Conventional Pap Smear Slide & ThinPrep® Slide (Cytyc Corporation, 2008) 26

Figure 2.15: Binary Data Representation Using Single Bit Binary 28

Figure 2.16: Grey Scale Data Representation Using 8-bit Binary 28

Figure 2.17: Red-Green-Blue (RGB) Data Representation Using 8-bit Binary 28

Figure 2.18: Pixel Value *f(x, y)*, Coordinate (x, y) & Image Resolution (N x M) in Bitmap Image 29

Figure 2.19: RGB Colour Model (R-red, G-green, B-blue, M-magenta, Y-yellow, C-cyan, W-white and black) (Weeks, 1996) 30

Figure 2.20: Coordinate Representation of 1-bit RGB Colour Model (Ghafar, 2003) 30

Figure 2.21: Coordinate Representation of 8-bit RGB Colour Model 31

Figure 2.22: Image Digitization: Sampling and Quantization (Gonzalez & Woods, 2002) 32

Figure 2.23: Single Imaging Sensor 32

Figure 2.24: An Array of Imaging Sensors 33

Figure 2.25: An Example of Image Histogram; Divided Into 3 Regions, Nucleus, Cytoplasm and Cell Background. 35

Figure 3.2: PixeLINK® PL-A662 FireWire Digital Camera 48

Figure 3.3: Sub-window Area and User-Defined Rectangle 49

Figure 3.4: Host Software Architecture 51

Figure 3.5: PL-A662 Program Flow for Single Frame Capture 52

Figure 3.6: Leica® DM LA Automated Microscope 53

Figure 3.7: AutoCapture Main User-Interface 56

Figure 3.8: Single Image Capture User Interface 57

Figure 3.9: ThinPrep® Glass Slide Diagram 57

Figure 3.10: Direction of Capturing Process on the Glass Slide 58

Figure 3.11: Image Filename Format for AutoCapture System 59

Figure 3.12: Flow Chart of an Automatic Glass Slide Capturing System 59

Figure 3.13: 400X Images Sorted in Capture Order 61

Figure 3.14: 400X Images Sorted in Ratio of Nucleus-Cytoplasm Areas 62

Figure 3.15: 100X – 400X Magnification Map 64

Figure 3.16: Inter-image Cell Display 65

Figure 4.1: TMS320C6416 CPU Core Architecture and Its Functional Block Diagram 69

Figure 4.2: Block Diagram for the TMS320C6416 DSK (Spectrum Digital Inc., 2003) 70

Figure 4.3: Memory Map for TMS320C6416 DSK in Comparison With Generic TMS320C6416 DSP 71

Figure 4.4: Features Extraction Process of ThinPrep® Image Based On Automatic Features Extraction (AFE) Algorithm 75
Figure 4.5: ThinPrep® Image and Its Grey Level Histogram 78
Figure 4.6: The Seed Location at Coordinate *P(0, 0)* 82
Figure 4.7: Growing Techniques, (a) 4-neighbours adjacently, (b) 4-neighbours diagonally and (c) 8-neighbours surround. 84
Figure 4.8: The Seed Location at Coordinate $P(\tilde{x},\tilde{y})$ 85
Figure 4.9: Overall Flow Diagram of Automatic Features Extraction Based on TMS320C6416 DSP System 90
Figure 4.10: Overall Flow for the Code Optimization Process of the C Code 92
Figure 4.11: TMS320C6000 Software Development Flow 95
Figure 4.12: Syntax of Normal Loop Operation on 2-Dimensional Array 99
Figure 4.13: Code in Figure 4.12 with Loop Unrolling 99
Figure 5.1: The AutoCapture System 107
Figure 5.2: Single Image Capture - (a) C1264_1.bmp and (b) C1264_2.bmp for 100X and 400X Magnification Capturing, respectively. 109
Figure 5.3: Automatic Glass Slide Capture - (a) C1253-000-000.bmp and (b) C1253-001-000.bmp for 100X Magnification Capturing 110
Figure 5.4: Automatic Glass Slide Capture - (a) C1253-000-000.bmp and (b) C1253-001-000.bmp for 400X Magnification Capturing 111
Figure 5.5: C1253-000-000.bmp and Its Grey Level Histogram. 111
Figure 5.6: C1253-001-000.bmp and Its Grey Level Histogram 112
Figure 5.7: C1253-002-000.bmp and Its Grey Level Histogram 112
Figure 5.8: C1253-003-000.bmp and Its Grey Level Histogram 112
Figure 5.9: C1253-004-000.bmp and Its Grey Level Histogram 113
Figure 5.10: C1253-005-000.bmp and Its Grey Level Histogram 113
Figure 5.11: C1253-006-000.bmp and Its Grey Level Histogram 113
Figure 5.12: C1253-007-000.bmp and Its Grey Level Histogram 114

Figure 5.13: C1253-007-001.bmp and Its Grey Level Histogram 114

Figure 5.14: C1253-006-001.bmp and Its Grey Level Histogram 114

Figure 5.15: C1253-005-001.bmp and Its Grey Level Histogram 115

Figure 5.16: C1253-004-001.bmp and Its Grey Level Histogram 115

Figure 5.17: C1253-003-001.bmp and Its Grey Level Histogram 115

Figure 5.18: C1253-002-001.bmp and Its Grey Level Histogram 116

Figure 5.19: C1253-001-001.bmp and Its Grey Level Histogram 116

Figure 5.20: C1253-000-001.bmp and Its Grey Level Histogram 116

Figure 5.21: 100X – 400X Navigational Buttons Panel 120

Figure 5.22: C1253-000-000.bmp 100X Magnification Image with Grid and Caption. 120

Figure 5.23: The Corresponding 400X Magnification Image, C1253-000-000.bmp, at Coordinate (0,0) for Figure 5.22 121

Figure 5.24: The Corresponding 400X Magnification Image, C1253-003-000.bmp, at Coordinate (3,0) for Figure 5.22 121

Figure 5.25: 4 100X Magnification Images Stitched Together 122

Figure 5.26: 64 400X Magnification Images Stitched Together, Equivalent to the 4 100X Magnification Images in Figure 5.21 123

Figure 5.27: Initial Image Cached with 8 Neighbours for Inter-Image Cells Display. 124

Figure 5.28: Inter-Image View Panel 125

Figure 5.29: C1253-001-001.bmp Stitched Together with 8 Adjacent Neighbours for Inter-Image Cells Display in (a) Top, (b) Left, (c) Right, and (d) Bottom Directions. 125

Figure 5.30: Images Captured by the AutoCapture System (640x480) at Desktop Resolution of 1280x1024 127

Figure 5.31: ThinPrep® Image 1 and Grey Level Histogram 129

Figure 5.32: ThinPrep® Image 2 and Grey Level Histogram 130

Figure 5.33: ThinPrep® Image 3 and Grey Level Histogram 130
Figure 5.34: ThinPrep® Image 4 and Grey Level Histogram 130
Figure 5.35: ThinPrep® Image 5 and Grey Level Histogram 131
Figure 5.36: ThinPrep® Image 6 and Grey Level Histogram 131
Figure 5.37: ThinPrep® Image 7 and Grey Level Histogram 131
Figure 5.38: ThinPrep® Image 8 and Grey Level Histogram 132
Figure 5.39: ThinPrep® Image 9 and Grey Level Histogram 132
Figure 5.40: ThinPrep® Image 10 and Grey Level Histogram 132
Figure 5.41: Centroid Location for ThinPrep® Image 1 135
Figure 5.42: Centroid Location for ThinPrep® Image 2 135
Figure 5.43: Centroid Location for ThinPrep® Image 3 136
Figure 5.44: Centroid Location for ThinPrep® Image 4 136
Figure 5.45: Centroid Location for ThinPrep® Image 5 136
Figure 5.46: Centroid Location for ThinPrep® Image 6 137
Figure 5.47: Centroid Location for ThinPrep® Image 7 137
Figure 5.48: Centroid Location for ThinPrep® Image 8 137
Figure 5.49: Centroid Location for ThinPrep® Image 9 137
Figure 5.50: Centroid Location for ThinPrep® Image 10 138
Figure 5.51: Segmentation Result for ThinPrep® Image 1 138
Figure 5.52: Segmentation Result for ThinPrep® Image 2 139
Figure 5.53: Segmentation Result for ThinPrep® Image 3 139
Figure 5.54: Segmentation Result for ThinPrep® Image 4 139
Figure 5.55: Segmentation Result for ThinPrep® Image 5 140
Figure 5.56: Segmentation Result for ThinPrep® Image 6 140
Figure 5.57: Segmentation Result for ThinPrep® Image 7 140
Figure 5.58: Segmentation Result for ThinPrep® Image 8 140
Figure 5.59: Segmentation Result for ThinPrep® Image 9 141
Figure 5.60: Segmentation Result for ThinPrep® Image 10 141

Figure 5.61: Code Snippet for Effect of Compiler Options Settings 149
Figure 5.62: Code Snippet for Effect of Loop Unrolling 151
Figure 5.63: Code Snippet in Figure 5.63 after Loop Unrolling 151
Figure 5.64: Code Snippet for Effect of Intrinsic Operators (CPU Cycle: 271, 746) 152
Figure 5.65: Code Snippet in Figure 5.65 after Using Intrinsic Operators (CPU Cycle: 260, 604) 152
Figure 5.66: Graph of Relationship between Execution Time and Cells' Size for the Code without Code Optimization Techniques Implementation. 154
Figure 5.67: Graph of Relationship between Execution Time and Cells' Size for the Code with Code Optimization Techniques Implementation. 154
Figure 5.68: Graph of Relationship between Execution Time and Cells' Size (The DSP System) 157
Figure 5.69: Graph of Relationship between Execution Time and Cells' Size (The PC-based System) 157

LIST OF ABBREVIATIONS

AFE	-	Automatic Features Extraction
ANSI	-	American National Standards Institute
API	-	Application Programming Interface
CCD	-	Charge Coupled Device
CCS	-	Code Composer Studio
CIN	-	Cervical intra-epithelial neoplasia
CMOS	-	Complementary Metal-Oxide Semiconductor
COFF	-	Common Object File Format
CPU	-	Central Processing Unit
DIP	-	Dual In-Line Packaging
DLL	-	Dynamic Link Library
DPI	-	Dot per Inch
DSK	-	DSP Starter Kit
DSP	-	Digital Signal Processor
EDMA	-	Enhanced Direct Memory Access
EMIFA	-	External Memory Interface type A
EMIFB	-	External Memory Interface type B
EMIFs	-	External Memory Interfaces
FCM	-	Fuzzy *c*-means
FDA	-	Food and Drug Administration
FIGO	-	The International Federation of Gynaecology and Obstetrics
FOV	-	Field of View
FPS	-	Frame per Second
H^2MLP	-	Hierarchical hybrid multilayered perceptron

HPI	-	Host Port Interface
HPV	-	Human Papilloma Virus
HSIL	-	High grade squamous intraepithelial lesion
I/O	-	Input/Output
IDE	-	Integrated Development Environment
KM	-	*K*-means
LED	-	Light Emitting Diodes
LSIL	-	Low grade squamous intraepithelial lesion
MADU	-	Minimum Addressable Data Unit
MFE	-	Manual feature extraction
MIPS	-	Million Instructions per Cycle
MKM	-	Moving *k*-means
MRI	-	Magnetic Resonance Image
OHCI	-	Open Host Controller Interface
PPM	-	Page per Minute
RGB	-	Red-Green-Blue
RGBFE	-	Region growing based feature extraction
ROI	-	Region of Interest
RTDX	-	Real-Time Data Exchange
SBRG	-	Seed-based region growing
SDK	-	Software Developers' Kit
SDRAM	-	Synchronous Dynamic Random Access Memory
SIMD	-	Single Instruction Multiple Data
SVM	-	Support Vector Machines
VLIW	-	Very Long Instruction Word

REKABENTUK SISTEM PERAKAMAN SLAID AUTOMATIK DENGAN SISTEM PENGEKSTRAKAN CIRI BERASASKAN PEMPROSES ISYARAT BERDIGIT (DSP) UNTUK IMEJ THINPREP® BAGI KANSER PANGKAL RAHIM

ABSTRAK

Kaedah konvensional untuk mengesan kanser pangkal rahim melibatkan pakar patologi dan sitologis memeriksa slaid palitan ThinPrep® di bawah mikroskop cahaya di bawah pembesaran 100X dan 400X. ThinPrep® merupakan teknologi daripada Cytyc Corporation yang telah digunakan secara meluas sebagai kaedah untuk menghasilkan slaid yang ekalapisan dan bersih. Penyelidikan ini tertumpu kepada pembinaan sistem perakaman slaid ThinPrep® automatik dan sistem pengekstrakan ciri berasaskan pemproses isyarat berdigit (DSP) untuk imej ThinPrep®. Sistem perakaman slaid automatik tersebut terdiri daripada mikroskop automatik, kamera digital dan komputer peribadi yang berhubung di antara satu sama lain bagi proses perakaman tersebut. Proses tersebut dilaksanakan di bawah pembesaran 100X dan 400X seperti yang disarankan oleh ahli patologi dan sitologis. Kemudahan pemaparan imej turut dilengkapkan ke dalam sistem tersebut yang mampu memaparkan imej-imej terakam melalui pelbagai kaedah. Imej-imej tersebut boleh dipaparkan mengikut turutan perakaman, imej yang paling disyaki, peta pembesaran 100X-400X dan juga pemaparan sel yang terpapar sebahagiannya di dalam satu imej. Sistem pengekstrakan ciri berasaskan DSP berupaya mengekstrak ciri-ciri tertentu daripada imej ThinPrep® yang dirakam. Sistem tersebut berasaskan DSP daripada Texas Instrument iaitu TMS320C6416 yang merupakan DSP dari kategori titik tetap. Ciri-ciri yang perlu diekstrak adalah purata skala kelabu, saiz dan perimeter untuk kedua-dua nukleus dan sitoplasma bagi sel serviks. Ciri-ciri tersebut diperlukan oleh ahli patologi

dan sitologis untuk menentukan tahap pra-kanser sel-sel terbabit. Proses pengekstrakan ciri dimulakan dengan teknik pengelompokan yang menggunakan algoritma purata-*k* boleh gerak (moving-*k* means, MKM) untuk menentukan nilai ambang bagi membezakan kawasan nukleus, sitoplasma dan latar belakang bagi sel serviks. Kaedah momen nalar digunakan untuk mencari titik sentroid bagi sel terbabit. Daripada titik sentroid tersebut, algoritma pertumbuhan kawasan secara titik benih (seed-based region growing, SBRG) digunakan untuk meruas dan pada masa yang sama mengekstrak ciri-ciri daripada nukleus dan sitoplasma sel serviks tersebut. Kesemua langkah-langkah ini adalah daripada algoritma pengekstrakan ciri automatik (automatic features extraction, AFE) yang telah digunakan untuk mengekstrak ciri-ciri sel serviks di dalam komputer peribadi. Di dalam penyelidikan ini, algoritma AFE akan diimplemen di dalam DSP dan beberapa teknik pengoptimuman akan digunakan untuk memaksimumkan penggunaan DSP yang terhad jika dibandingkan dengan komputer peribadi. Algoritma AFE di dalam DSP dan komputer peribadi, masing-masingnya mampu memproses imej ThinPrep® dalam masa 0.4640 dan 0.1048 milisaat per unit piksel. Kedua-dua sistem perakaman slaid automatik dan sistem pengekstrakan ciri automatik berasaskan DSP mencadangkan kaedah baru untuk memeriksa dan menyaring slaid ThinPrep®.

DESIGN OF AN AUTOMATED SLIDE CAPTURING SYSTEM WITH A DSP-BASED AUTOMATIC FEATURES EXTRACTION OF THINPREP® IMAGES FOR CERVICAL CANCER

ABSTRACT

Conventional method of cervical cancer screening involves of pathologist or cytologist examining the ThinPrep® cervical smear slide under normal light microscope with 100X and 400X magnification. ThinPrep® is a technology from Cytyc Corporation that has been used widely for producing mono-layer and clean cervical smear slide. This study focuses on developing an automated ThinPrep® slide capturing system and a DSP-based automatic features extraction system for the ThinPrep® images. The automated slide capturing system consists of automated microscope, digital camera and personal computer that work together to handle the slide in order to acquire it into digital form. The capturing process is carried out under 100X and 400X magnification as preferred by pathologist and cytologist. Image retrieval or displaying facilities are embedded inside the system that capable to display the captured images in various ways. The images could be displayed in capturing order, most suspected images, 100X-400X magnification map and also capable to display the cell that partially displayed in one image. The DSP-based automatic features extraction system on the other hand capable to extract certain features from the ThinPrep® images. The system is based on digital signal processor (DSP) from Texas Instruments, TMS320C6416, a fixed-point DSP. The features that need to be extracted are average grey level, size and perimeter for both of nucleus and cytoplasm of the cervical cell. The features are needed by the pathologist and cytologist to determine the pre-cancerous level of the cells. The extraction process starts with clustering technique that uses moving *k*-mean (MKM) algorithm to find the threshold value to differentiate nucleus, cytoplasm and

background of the cervical cell. Invariant moment technique is used to find the centroid location of the cell's nucleus. Starting from the centroid location, seed-based region growing (SBRG) algorithm is used to segment and at the same time extracts all the features from both of nucleus and cytoplasm of the cell. All of these steps are from automatic features extraction (AFE) algorithm that have been used for extracting features from cervical cell image on a personal computer. In the current study, the AFE algorithm will be implemented inside the DSP and some code optimization techniques will be used to fully utilize all local resources inside the DSP that are limited as compared to the personal computer. The AFE algorithm inside the DSP and personal computer could process the ThinPrep® image in 0.4640 and 0.1048 milliseconds per unit pixel, respectively. Both of the automated slide capturing system and the DSP-based automatic features extraction system proposed the new ways of examining and screening the ThinPrep® slide.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Technological innovation has yielded truly remarkable advances in health care during the last three decades. In just the last several years, breakthroughs in biotechnology, biomaterials, surgical techniques, and computer technology have helped to improve health care delivery and patient outcomes. The instruments that have been used in medical fields have helped experts to solve medical problems. Technological developments have given a new inspiration to researchers who have been struggling to obtain new knowledge to apply into new medical invention.

One of the areas in health care that requires new technologies to help maintain healthy life is cancer screening program. Generally, cancer remains the number one concern for the human population worldwide (Othman, 2002). In the year 2005, a total of 59,777 cancerous cases have been reported in Malaysian government hospitals (Dewan Rakyat Board Meeting, 2007). In the year 2003, a total of 21,464 reported cancer cases were diagnosed among Malaysians in Peninsular Malaysia comprising of 9,400 men and 12,064 women (Lim & Halimah, 2003). From the report, lung cancer was identified as the most common cancer among men (13.8% out of the 9,400 reported male cancer cases). As for women, breast cancer was the main problem (31.0% or 3,738 cases of total female cancers) making it the most commonly diagnosed cancer in women. The second most common problem among women was the cancer of the cervix that constituted 12.9% or 1,557 of reported total female cancers.

The cancer of the cervix could be prevented since the precursor cancer cells can be detected by regular screening test called Pap smear test (Othman, 2002). Peoples, especially women should be given early exposure on information and risk of the cancer. The Malaysian government and non-government organizations (NGO) such as Majlis Kanser Nasional (MAKNA) and Obstetrical & Gynaecological Society of Malaysia (OGSM) have made enormous efforts to encourage women to do the Pap smear test as a prevention step for cervical cancer. The test remains the most effective strategy for the detection of pre-cancerous stage and consequent control of cervical cancer (Ministry of Health, Malaysia, 2003).

1.2 Computer Assisted Systems for Cervical Cancer Screening

The AutoPap® Primary Screening System is one of the computer assisted systems for cervical screening that uses high-speed video microscope and image interpretation software. The system is capable to classify the input slides for normal and abnormal cases. The software algorithms rank each slide based on the likelihood that an abnormality is present. The process requires computer evaluation of visual patterns and localisation of significant objects and groupings, followed by systematic comparison of these findings against thousands of reference images. The algorithms detect both squamous and glandular lesions and also evaluate the adequacy of the specimen (Medical Services Advisory Committee, 2003). This system has been approved by Food and Drug Administration (FDA) for screening Pap smear slide that produced from PrepStain™ technique (Palmer, 2005).

PAPNET® Testing System uses automated microscope with an attached video camera for the detection of abnormal cells. Based on neural network technology, the program recognises suspicious cells and cell clusters and presents them to the cytologist for further assessment. On each slide, 128 images consisting of 64 cells and 64 cell clusters that most likely to be abnormal are identified. The images are then

stored in a digital tape and send to the pathologist for diagnose. The images could also be referred to other pathologists for exchange of opinion through the internet facilities (Medical Services Advisory Committee, 2003).

AutoCyte SCREEN is a computerised image analysis system for primary screening. This system was developed mainly for slides that have been produced via the liquid-based method that is PrepStain™ technique. Similar to PAPNET® Testing System, the AutoCyte SCREEN only selects slides that have suspicious cells and cell clusters for further examination by cytologists (Medical Services Advisory Committee, 2003). The results from the cytologists that differ with the results from the system are subject to further cytological review (Bishop *et al.*, 2000).

NeuralPap System was developed by Mat-Isa (2003) to classify image from conventional Pap smear slide into 3 stages of pre-cancerous level: normal, Low-grade Squamous Intra-epithelial Lesion (LSIL) and High-grade Squamous Intra-epithelial Lesion (HSIL). The system used moving *k*-mean (MKM) algorithm for clustering the cell image into nucleus, cytoplasm and background regions. Region Growing-based Features Extraction (RGBFE) algorithm was introduced from Seed-based Region Growing (SBRG) algorithm for segmentation and features extraction of the cervical cell. The features extracted were size, perimeter and grey level of nucleus and cytoplasm, respectively. The same features were extracted by pathologist and compared with the features that were extracted by the system. The features were then became inputs to Hierarchical Hybrid Multi-layered Perceptron (H^2MLP) neural network to train the system for classification processes.

1.3 Problem Statement

Conventional method of diagnosing cervical cancer cases requires cytologist or pathologists to look at tissue samples through a microscope and to visually interpret

the complex microscope images. The magnifications involved are 100X and 400X that makes the process is time-consuming and too subjective. The shortage of pathologist worldwide further compounded the problem. Moreover, pathologists or cytologists are also susceptible to fatigueness and this will affect the diagnostic process. It is estimated that only 20 to 30 slides that can be processed by a pathologist in one day. In United States of America, the government has prevented cytologist and pathologist from reviewing more than 100 slides per day and the more stringent Australian guidelines allow only 70 slides per day (Medical Services Advisory Committee, 2003).

For reviewing slides, it is highly depends on the effectiveness of archiving system at laboratories or hospitals. In most of the cases, the slide retrieving is difficult and nearly impossible when there are certain cases that needs to be reviewed. Moreover, the slides tend to loss some of information such as color from staining process if kept in archive beyond the suggested time period.

Digital camera is the best solution to replace human eyes. Unlike a human, it can be used 24 hours a day and 7 days a week and will never be affected by fatigueness. Virtual archiving system that involves digitized information of the original slide is seems to have a very good potential to support modern medical facilities. The archive can be referred at any time and can be stored for a great period of times.

1.4 Objectives of Research

The main objectives of this research are:

1. To develop an automatic glass slide capturing system for ThinPrep® cervical slide
2. To explore possibility of implementing automatic features extraction of the cervical cell image on digital signal processor (DSP)

Sub-objectives for objective 1 are:

a) **Providing automatic glass slide capturing system that can be easily operated by technologist.**

   A system that consists of automated microscope, digital camera and personal computer will be developed to automatically capture glass slide or specifically ThinPrep® slide. The system will store the digital images in the personal computer memory storage. A simple user interface will be used to make the system much easier for use even by technologist after a short training period.

b) **Imitating the actual pathologist or cytologist screening procedures of the smear slide by providing 100X and 400X magnification image previewing.**

   The pathologist and cytologist screening procedures involve the reviewing of the slide in 100X magnification, looking for abnormal cell or abnormal cell cluster. The 400X magnification is used to see the detail characteristics of the abnormal cell or abnormal cell cluster itself. The system will be able to display the captured images in both of 100X and 400X magnification. The target user will start the screening process by previewing the 100X magnification image for scanning and screening for groups of abnormal cells and then could easily switch to the respective 400X magnification image when required to look at the detail characteristic of the cells.

c) **Providing access to the live image of the slide on the screen, different to most of slide scanners in the market.**

   The screening of cervical cancer will become much comfortable when the user could just scan the slide on the computer screen rather than seeing through the microscope's eyepiece that tends to affect the eyes. User may freely move the slide by using joystick and at the same time could watch the live preview of the slide on the screen. It also differs from most of the slide scanners that prohibit

user from moving the slide during the scanning process and the only image that can be observed is the captured image.

d) **Lessening the time taken to review and screen the slide, especially for tasks that require exploring and/or analyzing complete slides.**

The software system provides graphical user interface that will guide the user to browse the captured images through 100X and 400X magnification. Browsing will be available between 100X and 400X magnification images, between 100X magnification images and also between 400X magnification images when required. Thus, the slide reviewing process could be done easily even when the ThinPrep® slide is not available. The software system also provides facility to sort the captured images in thumbnail panel by displaying the most suspected images on the top position, suggesting for the image to be diagnosed first. The facility will present an overall prognosis and will lessen the time taken for the screening process.

The second objective of the current study is to explore the possibilities of implementing the automatic features extraction process of the ThinPrep® image on a DSP. The features that need to be extracted are size, perimeter and grey level of nucleus and cytoplasm of the cervical cell, respectively. The algorithms are capable to segment the image into 3 main regions: nucleus, cytoplasm and background of the cell. This will be followed by features extraction process that will examine every single pixel inside the image. The entire process will execute automatically and give the results right after the features extraction process.

1.5 Scope of Research

The scope of research of the current study is mainly on image capturing system and digital signal processing that is digital image processing. The research is based on

medical application that is cervical cancer screening that involves ThinPrep® slides and images. A ThinPrep® slide contain thin-layered cervical cells taken from a woman's cervix by using special methods and equipments. The slide is also clean and contains appropriate amount of cells for the diagnosis process. In the medical field, the ThinPrep® slide has been used widely to detect the existence of cancer in women's cervix.

For the first main objective, the automatic glass slide capturing system will be built by using automated microscope from Leica® from DM LA model and digital camera from PixeLink® that is PL-A662. The system uses the Software Developers' Kit (SDK) from both of the manufacturers extensively that contain special instructions to control both of the devices. The examples of control are such as moving the microscope's stage, lenses, diaphragm aperture and light intensity control for the automated microscope. As for the digital camera, the control involves of video previewing, gamma correction, colour adjustment and image capturing. The personal computer on the other hand should be able to communicate with the digital camera through FireWire and serial port for the automated microscope.

The automated microscope should use eye-piece with 10X magnifications and lenses with 10X and 40X magnifications. Thus, the total magnifications for the capturing process are 100X and 400X magnifications as usually used by pathologists and cytologists. The lenses supposed to be high quality and from the type of Plan Apochromat 'dry' lenses that do not require any immersion oil to see through the ThinPrep® slides specimen. All the images captured by the system are only available in 640x480 resolution for maximum viewing in monitor resolution of 1280x1024. This is also for fast displaying inside the system that uses thumbnail method very extensively.

For the second main objective, the code that will be developed is purely for DSP from Texas Instruments from TMS320C6416 model. The code uses C language and some special instruction known as intrinsic operators that specifically for this model only. For the automatic extraction process, the maximum resolution of the input image is fixed to 640x480 or less. This is due to limited memory resources and the extraction process uses the memory for storing and processing each pixel of the image as a fixed-memory known as arrays. The arrays' size is fixed to 640 in width and 480 in height from the type of char (8-bit data) and at the same time limits the size of the input image

The multi-cell ThinPrep® image that has been captured should be cropped so that only one cervical cell appeared in one image. The algorithm uses inside the system only extract features from single cell and incapable of extracting features from multi-cell ThinPrep® image. The image information such as pixels value are in grey level format and needs to be transferred into the DSP's memory by using Code Composer Studio along with the image resolution.

1.6 Thesis Outline

Overall, this thesis comprises 6 chapters. It starts with Chapter 1 that gives concise explanation about the research in the current study. The explanation starts with an introduction and is followed by discussion on a few computer assisted systems for cervical cancer screening. Besides, objectives and scope of the research, along with thesis guidelines, are also presented in this chapter.

Chapter 2 serves as a literature review of this research. It provides in-depth discussion on digital signal processor (DSP) from Texas Instrument – TMS320C6000 family, including its basic architecture and Code Composer Studio™ as an integrated development environment (IDE) of the DSP. Next, it will proceed into the description of

cancer of the cervix from anatomy of the cervix to the Pap smear test and ThinPrep® technology. Digital image processing will be carried out by discussion on image sampling and quantization for conversion of analog image to the digital image. Basic knowledge of segmentation and features extraction will be discussed and followed by review of existing studies on image capturing applications by using microscope and DSP applications for image and video processing.

Chapter 3 will provide in-depth discussion on the AutoCapture System. AutoCapture System is the proposed system of automatic glass slide capturing system. The presentation of the chapter starts with an overall diagram of the system with its entire component. Next, the hardware and interfacing of the system will be presented including the software part for communicating with personal computer. Finally, the features inside the AutoCapture System that are image capturing and image displaying will be discussed in detail.

Chapter 4 will discuss on the implementation of automatic features extraction of cervical cell image from ThinPrep® slide on TMS320C6416 DSP from Texas Instrument. The architecture of the DSP will be explained including its hardware interconnections. Next, for features extraction of the cervical cell, discussion on threshold value finding, centroid location finding, segmentation and features extraction will be presented. The implementation of the processes on the DSP will be discussed from image information transferring, code optimization techniques and code profiling. The comparison between the DSP-based system and personal computer (PC)–based system will be made by correlation test and execution time comparison.

Results from both of the systems in Chapter 3 and 4 will be presented in Chapter 5. The presentation starts with results from the AutoCapture System. The hardware and software features of the system will be shown through a few images that

have been captured by the system. Next, features that have been extracted from the DSP System and PC-based System will be presented and compared to each other. The execution time comparison will be carried out on how the code optimization process on the DSP System could shorten the time taken to extract all the features.

The last chapter, Chapter 6 will provide conclusion on the AutoCapture System and the DSP System. It includes how the AutoCapture System could be used as a pre-screening tool for cervical cancer and the conclusion about the possibility of implementing the automatic features extraction algorithm on the DSP. Chapter 6 will also discuss on the future improvements that can be implemented on both of the AutoCapture System and the DSP System.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter discusses on 3 main topics that are cancer of cervix, digital image processing and digital signal processor (DSP). In Section 2.2, an overview of cancer of cervix will be reviewed in detail. The discussion begins with the description of the cervical anatomy and the cervical microscopic anatomy. The key factor of cervical cancer-Human Papilloma Virus, an overview of cervical cancer including Pap smear test, conventional slide preparation technique and Thin Prep® will follow next. Section 2.3 will deal with digital image processing fundamentals, such as digital image, image sampling and quantization, and certain image processing techniques. Next in Section 2.4, a brief overview of DSP will be discussed, with special focus on Texas Instruments DSP, TMS320C6000 family including the Code Composer Studio, and the DSP/BIOS Kernel, which makes real-time digital signal processing possible. Section 2.5 will focus on image capturing application by using microscope. Finally, in the last section, a few applications regarding DSP in image and video processing will be reviewed.

2.2 Cervical Cancer

Cancer refers to a malfunction of certain body cells, which would divide very fast and produce too much tissue that in turn forms a tumour. Cervical cancer is a cancer that attacks the cervix; the lower part of women's uterus (womb) (Pihie, 1998). 1,557 new cases of cervical cancer were found in Malaysia for the year 2003 (Abdul Jalil, 2005; Lim & Halimah, 2003). Malaysian government has made a lot of efforts to control, detect and treatment of the cancer disease in general through National Cancer Control Program since 1960s.

2.2.1 Cervical Anatomy

The word 'cervix' (from Latin "neck") is the lower, narrow portion of the uterus where it joins with the top end of vagina with uterus. Internal os connects cervix and uterus whereas external os connects cervix and vagina (Gray & McKee, 2002). The anatomy of women reproductive system is represented in Figure 2.1.

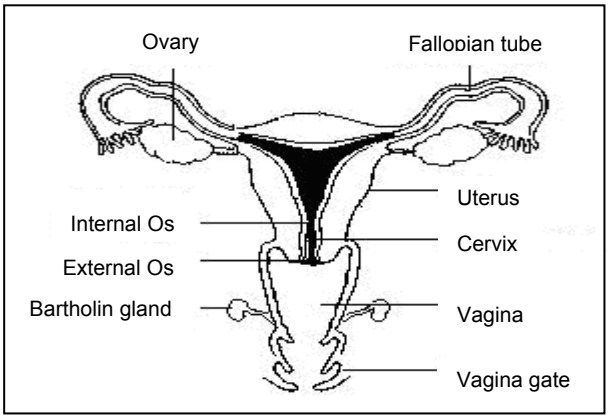


Figure 2.1: Women Reproductive System

Women cervix is divided into two sections; endocervix and ectocervix as depicted in Figure 2.2. The ectocervix is the portion of cervix that is exterior to the external os which is readily visible during speculum examination, whereas the endocervix is the portion above the external os (Sankaranarayanan & Wesley, 2003).

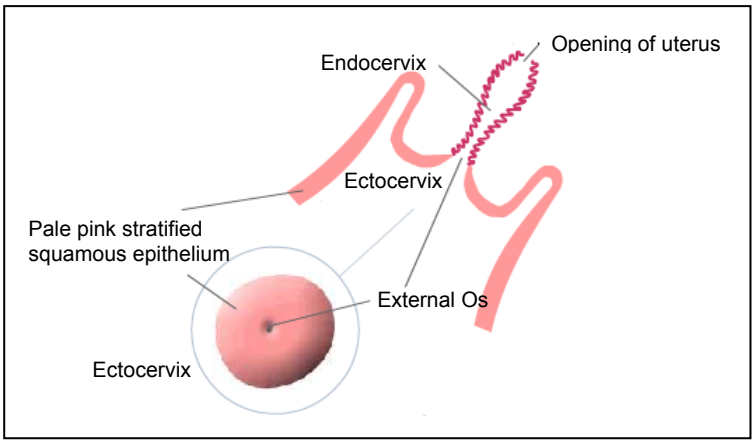


Figure 2.2: Ectocervix and Endocervix (Arends, 1999)

2.2.2 Cervical Microscopic Anatomy

In humans, epithelium is one of the four primary body tissues beside connective, muscle and nervous tissues. Epithelium is a tissue composed of a layer of cells that functions as protection, sensation detection, absorption, and selective permeability. In a woman cervix, two types of epithelium cells are common; squamous epithelium cells and columnar epithelium cells.

Columnar Epithelium Cells

Columnar epithelium cells usually can be found in the endocervix section as shown in Figure 2.3. This type of tissue cells is also known as endocervix cells. It consists of single layer of tall cells with dark-staining nucleus. It appears as a grainy, oval shape, with the function of secreting the mucus that lubricates the cervix and vagina (Sankaranarayanan & Wesley, 2003).

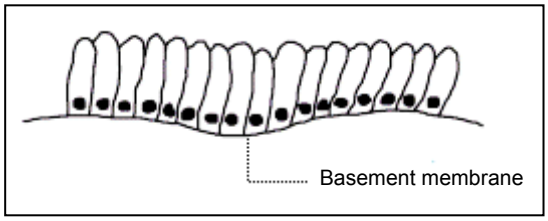


Figure 2.3: Columnar Epithelium Cells

Squamous Epithelium Cells

Large areas of ectocervix are covered by squamous epithelium cells as shown in Figure 2.4. It has multiple layers of cells (i.e. 15 to 20 layers) and appears pale pink in colour on visual examination. It consists of a single layer of round basal cells with a large dark-staining nucleus and little cytoplasm at the basement membrane. The basal cells divide and differentiate to form the parabasal, intermediate and superficial layers. From the basal to the superficial layer, the cells undergo an increase in cytoplasm and a reduction of nucleus size (Gray & McKee, 2002; Sankaranarayanan & Wesley, 2003).

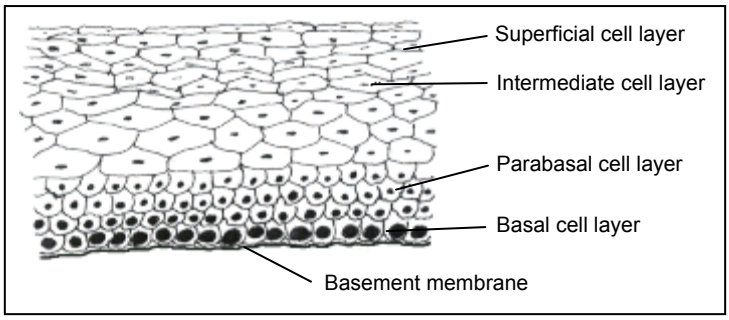


Figure 2.4: Squamous Epithelium Cells

Squamous Epithelium Metaplastic Cells

Changes from columnar epithelium cells to squamous epithelium cells always occur and are affected by frequent sexual intercourse (Othman, 2002). This is perceived as a form of trauma to the fragile columnar epithelium cells that tends to change to another kind of epithelium which is stronger; squamous epithelium cells (Sankaranarayanan & Wesley, 2003).

Cells that are in the process of transformation are called squamous epithelium metaplastic cells and the region is called transformation zone. This zone is also called squamocolumnar junction and located between the region where columnar epithelium cells and squamous epithelium cells meet (Pihie, 1998). This transformation phenomenon is called squamous metaplasia and reflected by age and sexual activity (Othman, 2002). Metaplasia refers to the change or replacement of one type of epithelium by another (Sankaranarayanan & Wesley, 2003). Figure 2.5 visualizes the cells inside squamocolumnar junction, endocervix and ectocervix layers.

2.2.3 Key Factor: *Human Papilloma Virus* (HPV)

Human Papilloma Virus (HPV) is one of the most common causes of sexually transmitted infection (STI) in the world, including cancer of cervix (Othman, 2002; Walboomers *et al.*, 1999). HPV is from DNA-based viruses that infect skin and mucous membranes of human.

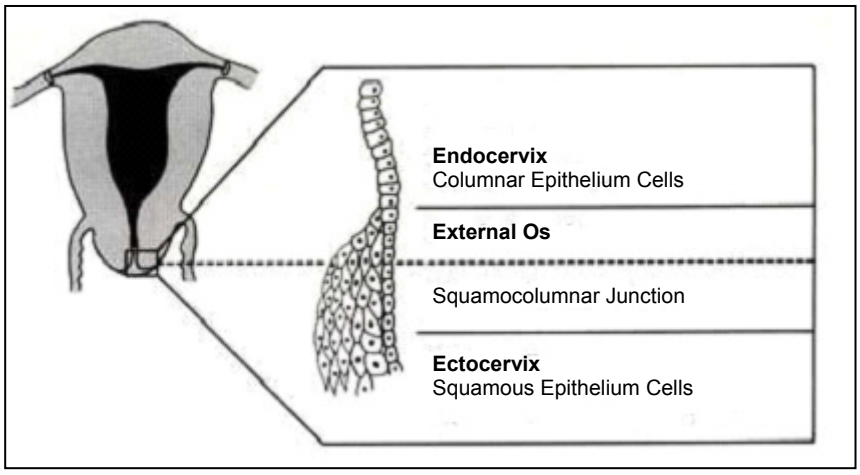


Figure 2.5: Cells inside endocervix and ectocervix layers

More than 130 different types of HPV have been characterized, and about 70 subtypes infect human, and out of these, 35 subtypes of HPV infect genital tracts. These 35 subtypes are then classified to low risk and high risk HPV (LR HPV and HR HPV) groups according to the risk of development of cervical cancer (Othman, 2002). The LR HPV subtypes are 6, 11, 42, 43, 44, and the HR HPV subtypes are 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 69 and possibly a few others. Figure 2.6 shows HPV image taken from electron photomicrograph.

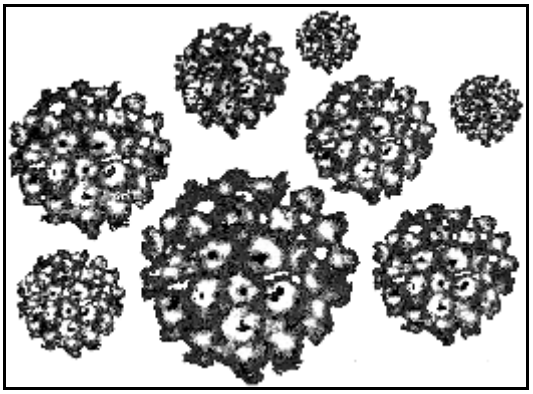


Figure 2.6: Electron Photomicrograph of Human Papilloma Virus

Some types of HPV may cause warts to appear on or around the genitals or anus. Genital warts (technically known as condylomata acuminatum) are most commonly associated with two LR HPV types, HPV–6 and HPV–11. Warts may appear within several weeks after sexual contact with a person, who is infected with this LR HPV, or they may take months or years to appear. These HPVs may also cause flat,

abnormal growths in the genital area and on the cervix. (National Cancer Institute, 2008). However, unlike the LR HPV, HR HPV infections usually do not cause any symptoms, meaning that, there are no signs of its occurrence and cannot be seen with naked eye (Cadman, 2006).

2.2.4 Cancer of Cervix; Pre-Cancerous & Cancerous

The worldwide HPV prevalence in cancer of cervix is 99.7% (Walboomers *et al.*, 1999). Certain types of genes from HPV could attack cervical cells that make the cells divide (mitosis process) uncontrollably than usual, erode and destroy normal tissues (Ghafar, 2003). The uncontrolled growth of unwanted cells will become abnormal mass of tissues that is called tumour (National Cancer Institute, 2008; Mat-Isa, 2003).

There are two (2) types of tumours namely: benign and malignant tumours (Pihie, 1998). Benign tumours are not dangerous and can sometimes be ignored, or removed entirely via surgery. It does not have the capabilities to spread among the neighbour cells or tissues. Malignant tumours are cancerous and sometimes referred to malignant growth or malignancy. It has the ability to spread among the cells and tissues, via lymph and blood circulation system in human body through metastasize process (National Cancer Institute, 2008). Normally for the cancerous tumours, a few options including radiation, chemotherapy and surgery can be taken to stop the abnormal cells activities.

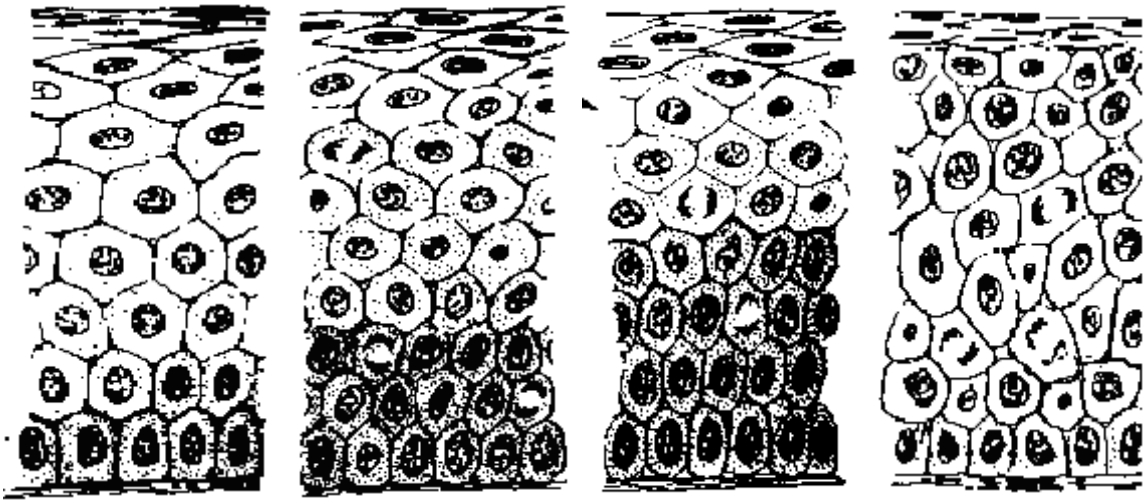
Changes in cells level are always used as a way to detect transformation of cancerous cells in the early stages. The cells will have an increase in nucleus-cytoplasm ratio as a result from mitosis process that occur only in nucleus but not in cytoplasm, and exhibit a loss of polarity (Pihie, 1998). The abnormal changes can easily be monitored under normal light microscope.

Cancer of cervix could originate from any type of tissues in women cervix. Nearly 90% of cancer of cervix cases start from squamous epithelium cells and is called squamous cell carcinoma whereas for the cases that started from columnar epithelium cells, it is called adenocarcinoma (Kuie, 1996). After infecting epithelium cells, the abnormalities could be invasive and attack the muscle tissues nearby. The phenomenon of infected cells among the epithelium cells is called Cervical Intra-Epithelial Neoplasia (CIN). The word Neoplasia here is from Greek meaning new growth and is classified as pre-cancerous case (Wikipedia Foundation Inc., 2008).

CIN is categorized into three namely CIN I, CIN II and CIN III according to the percentage of abnormal cells inside the epithelium cells layers. CIN I exists if there are less than ⅓ of abnormal cells from the epithelium layers. CIN II case may occur if the abnormal cells amount is greater than ⅓ but less than ⅔ from the epithelium layers. For more than ⅔ of abnormal cells in the epithelium layers, this will be classified as CIN III or carcinoma in-situ (Meisels & Morin, 1991; Othman, 2002).

All the 3 categories of CIN are pre-cancerous cases and are not yet cancerous. The possibility to become cancerous is high for CIN III compared to CIN I. The affected cell tissues also might self-cured or remain unchanged from time to time (Cronjé, 2004). Figure 2.7 shows the changes from normal to CIN III epithelium cells.

The concept of cervical intra-epithelial neoplasia or CIN was introduced by Richart, R. M. in 1973. It was a major breakthrough in encompassing all of the pre-cancerous stages of the epithelium cells. The concept however did not necessarily drift down to the primary care physicians or alleviate the need for a uniform reporting system of cervical smears (Koss, 1990).



(a) Normal (b) CIN I (c) CIN II (d) CIN III

Figure 2.7: Changes in Epithelium Cervix

In 1988, a group of experts lead by Bethesda, M. D. were met in an attempt to redefine the reporting system of the cervical smears. The outcome from this meeting, The Bethesda System, was since adopted by many practitioners. Inside the system, there are still co-relationship between the concept of CIN and the system itself plus details from infection to patient history. The classification of squamous epithelial cell abnormalities and its relationship with CIN concept are presented in Table 2.1.

Table 2.1: The Bethesda System (Spitzer *et al*., 2002)

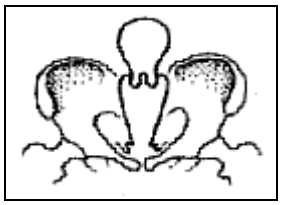
Classification	Dysplasia	CIN
Negative	Normal	Normal
LSIL	Least Dysplasia	CIN I
HSIL	More and most Dysplasia	CIN II and CIN III
Cancerous Stage	Cancerous Stage	Cancerous Stage

The term “dysplasia” refers to abnormality in the appearance of cells indicative of an early step towards transformation into a neoplasia which is cancerous stage (Sankaranarayanan & Wesley, 2003). The new terms added to the Bethesda System are Low-grade Squamous Intra-epithelial Lesion (LSIL) and High-grade Squamous Intra-epithelial Lesion (HSIL).

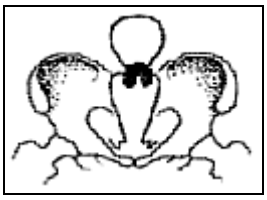
Pre-cancerous or CIN changes is not classified as cancer of cervix yet since the cells might self-cure or remain unchanged from time to time. The risks to become cancerous still exist and could be decreased by numerous ways of surgery and medication to stop the abnormal cells activities. As pre-cancerous stage, cancerous stage have been systemized and classified based on the system from The International Federation of Gynaecology and Obstetrics (FIGO). The stages are shown in Table 2.2 (Pihie, 1998) and visualize in Figure 2.8 (Azhar, 1994).

Table 2.2: Stage of Cervical Cancer (Pihie, 1998)

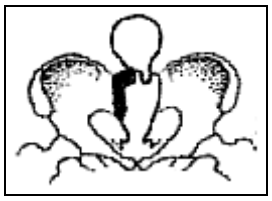
Stages		Cancer of Cervix
0	Carcinoma in-situ	Carcinoma in-situ; CIN III. The abnormal cells are in the surface layer of the cervix and have not invaded deeper tissues.
1a	Cancer in the cervix only	Micro invasive cervical cancer. Can only be seen with a microscope.
1b		Macro invasive cervical cancer. Can be seen without a microscope.
2a	Cancer extends beyond the cervix, but not as far as the pelvic wall	Extends to upper part of the vagina, but not to the surrounding tissues.
2b		Extends to the surrounding tissues, but not to the pelvic wall.
3a	Extended beyond Stage 2, but not beyond the pelvic area.	The cancer has spread to the lower third of the vagina, but nowhere else.
3b		The cancer has spread to the pelvic wall, or has blocked a ureter.
4a	Cancer has spread to the bladder, rectum, or outside the pelvis	Spread to the rectum or bladder.
4b		Spread to distant organs such as lungs or liver.



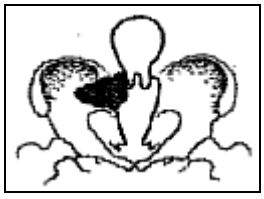
Stage 1a - Micro invasive cervical cancer



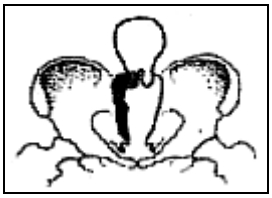
Stage 1b - Macro invasive cervical cancer



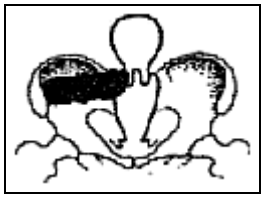
Stage 2a - Extends to upper part of the vagina, but not to the surrounding tissues



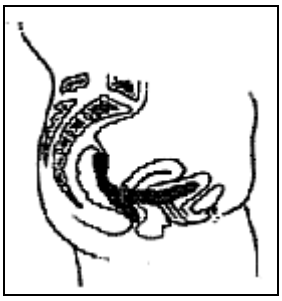
Stage 2b - Extends to the surrounding tissues, but not to the pelvic wall



Stage 3a - The cancer has spread to the lower third of the vagina, but nowhere else



Stage 3b - The cancer has spread to the pelvic wall, or has blocked a ureter



Stage 4a - Spread to the rectum or bladder

Stage 4b - Spread to distant organs such as lungs or liver

Figure 2.8: Stage of Cervical Cancer (Azhar, 1994)

2.2.5 Pap Smear Test and ThinPrep®

In gynaecology, Pap Smear Test is a medical screening method for detecting infectious, pre-malignant or malignant changes in the ectocervix and endocervix, which is originally invented by Georgios Papanikolaou (1883-1962) in 1940s (Spitzer *et al*., 2002).

Infection with certain type of HPV can cause changes in cervical cells that can be detected through Pap Smear Test (Othman, 2002). The Pap Smear Test can detect if a women have an infection, abnormal (unhealthy) cervical cells, or cancerous cells. The test can find the earliest signs of cervical cancer or pre-cancerous stages from the cell level that does not show any external symptoms as in cancerous stage. If detected early, the chances of curing cervical cancer is very high and it can prevent most cases of cervical cancer from developing (Crandall, 2005). An increased frequency of screening of the cervix can result in a greater reduction in the cumulative rate of invasive cervical cancer (Spitzer *et al*., 2002).

The best time for screening is between 10 and 20 days after the first day of woman menstrual period. For about 2 days before testing, a woman should avoid douching or using spermicidal foams, creams, or jellies or vaginal medicines (except as directed by a physician). These agents may wash away or hide any abnormal cervical cells (Crandall, 2005; Spitzer *et al*., 2002).

A speculum is inserted into the vaginal area (the birth canal). The speculum is an instrument that allows the vagina and the cervix to be viewed and examined. A cotton swab is sometimes used to clear away mucus that might interfere with an optimal sample (Crandall, 2005).

There are two ways of collecting samples from women cervix, by using Cytobrush - Ayre's Spatula or by using broom-like device or Cervex-Brush® (Cytyc Corporation, 2008). In the first method, Ayre's Spatula is used to obtain an adequate sampling from the ectocervix, as depicted in Figure 2.9. Cytobrush then, is used to obtain samples from the endocervix. The brush is needed to be inserted into the cervix until only the bottommost fibres are exposed as shown in Figure 2.10 (Cytyc

Corporation, 2008). The sample from both of Ayre's Spatula and Cytobrush are then smeared onto a glass slide.

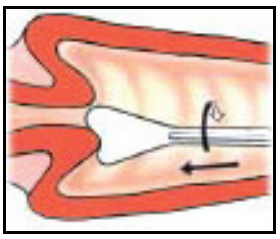


Figure 2.9: Obtaining Samples Using Ayre's Spatula - ectocervix

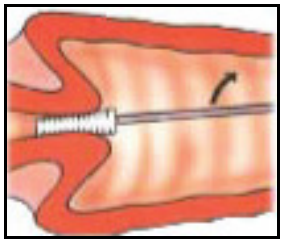


Figure 2.10: Obtaining Samples Using Cytobrush – endocervix

For the second method, a broom-like device or Cervex-Brush® (Figure 2.11) is used. The central bristles of the brush are inserted into the endocervix canal deep enough, to allow the shorter bristles to fully contact the ectocervix. The brush is then pushed gently, and the broom is rotated in a clockwise direction for five times. The brush is then smeared onto a glass slide. Figure 2.12 gives an illustration on how to get adequate samples by using the Cervex-Brush®.

Figure 2.11: Cervex-Brush®

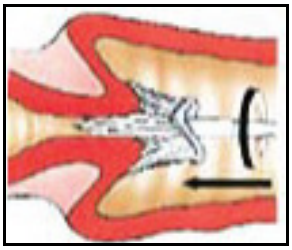


Figure 2.12: Obtaining Samples Using Cervex-Brush® - Both Endocervix & Ectocervix

After smearing the samples in both ways mentioned above, the glass is then sprayed with or immersed in a fixative solution to preserve the cells morphology

(Karnon *et al*., 2004). Coloring process is needed to increase the visibility of the samples (Tay, 1996). Cytologist will then examine the smear under a microscope to detect the existence of any abnormal cell. After that, pathologist will check on the smear once again to make the confirmation (Mat-Isa, 2003).

These methods of preparing smear slide have been practiced since 50 years ago (Othman, 2002). The problems that are always compounding these conventional methods are sample contamination by blood, mucus, air-drying artifact and inflammation (Grace *et al*., 2001). The cells collected are always clumped together and make the screening process becomes difficult. In most cases, the samples that have been taken from women are not totally smeared on the slide (Cytyc Corporation, 2008). Some of the cells are still on the sampling devices, and as a result, the preservation of all the cells is nearly impossible.

The conventional methods are also poor in sensitivity and specificity, and high in false negative. Sensitivity is a statistical measure of how well the screening methods give the result on true cases whereas specificity tell us on how well the screening give result in classifying the cases that not in the true cases. Figure 2.13, Equation 2.1 and 2.2 give explanations on how the sensitivity and specificity are calculated.

		Condition as determined by "Gold" standard	
		True	False
Test outcome	Positive	True Positive	False Positive
	Negative	False Negative	True Negative
		↓ Sensitivity	↓ Specificity

Figure 2.13: Sensitivity and Specificity

$$\text{Sensitivity} = \frac{\text{Number Of True Positives}}{\text{Number Of True Positives} + \text{Number Of False Negatives}} \quad (2.1)$$

$$\text{Specificity} = \frac{\text{Number Of True Negatives}}{\text{Number Of True Negatives} + \text{Number Of False Positives}} \quad (2.2)$$

According to Stein (2003) from his study, conventional Pap screening program has a sensitivity of only 51% and a false negative rate of 5 to 10% which is quite high. Noorani *et al.* (2003) has found that the sensitivity for this method was only 30% and can increase up to 87% in rare cases, with false negative from 13% to 70%. These limitations can significantly be decreased if the screening program was done repeatedly at the expense of examination costs that usually become a burden to patients.

To overcome all the problems, ThinPrep®, the liquid-based cytology method was introduced and approved by the U.S. Food and Drug Administration (FDA) in May 1996 as a replacement for the conventional Pap smear preparation technique (Noorani *et al*., 2003). It has been credited by many studies demonstrating a wide range of clinical benefits of the ThinPrep® Pap test such as increased disease detection, improved specimen adequacy and cost effectiveness (Cytyc Corporation, 2008).

The method of collecting samples out of woman cervix in ThinPrep® is the same with the conventional technique. Cervex-Brush or sometimes Ayre's spatula and Cytobrush are used exactly as the conventional method as sampling devices. Instead of smearing the cells on a slide, the sampling device is rinsed into a ThinPrep® vial transport medium that is PreservCyt®, capped, labelled and sent to the laboratory for slide preparation. (Karnon *et al.*, 2004)

Slide preparation is handled by automation systems; ThinPrep® 2000 Processor (T2000) or ThinPrep® 3000 Processor (T3000), semi-automated and fully automated respectively, both are from Cytyc Corporation (Cytyc Corporation, 2008, Noorani *et al*.,