ANDROGENIC HAIR PATTERN RECOGNITION FOR BIOMETRIC IDENTIFICATION

THESIS

Proposed as One of the Requirement to Finish Electrical Engineering Graduate Program

By:

Regina Lionnie
55415120017

UNIVERSITAS MERCU BUANA
GRADUATE PROGRAM

Abstract

To identify criminal or pedophile in online child pornography images and video is a challenging task when the faces and other distinguish features are not shown. To address these kind of problems, the system of recognition using androgenic hair pattern is being developed. Androgenic hair pattern stated to be the newest biometric trait since 2014. There were 4 different methods that were compared in this thesis, Haar wavelet transform, principal component analysis, hierarchical Gaussian scale space and scale invariant feature transform (SIFT). For the first three methods, 400 images were tested for the recognition system. Hierarchical Gaussian scale space produced 94,23 % precision of recognition using 10-fold cross validation with histogram equalization. The method defeated other two methods, Haar wavelet transform with 83,48 % precision of recognition and principal component analysis with 75.19 % precision of recognition. The next design used smaller version of androgenic hair database, 50 images with extreme condition of only 2 images variation for one respondent. For this latest system design, the SIFT algorithm gave the best performance of 38% precision of recognition and was superior to other previous three methods, Haar wavelet transform, principal component analysis and hierarchical Gaussian scale space that only produced around 30-32 % precision of recognition.

Keywords: androgenic hair pattern; biometric identification; Haar wavelet transform; principal component analysis; scale-invariant feature transform

Abstrak

Untuk dapat melakukan identifikasi pelaku kriminal atau pedofil pada kasus kejahatan seksual anak-anak yang tertangkap kamera atau video digital menjadi tantangan yang sulit ketika wajah atau fitur unik dari kriminal tersebut tidak terlihat. Untuk dapat mengatasi kesulitan ini, telah dibangun sistem pengenalan berbasis pola rambut androgenik. Pola rambut androgenik menjadi ciri biometrik baru sejak tahun 2014. Terdapat 4 metode yang dibandingkan di dalam tesis ini, transformasi wavelet Haar, analisis komponen utama, skala ruang hierarki Gauss dan *scale-invariant feature transform* (SIFT). Untuk ketiga metode pertama, 400 gambar digital diuji untuk sistem pengenalan. Metode skala ruang Hierarki Gauss menghasilkan keakuratan presisi sistem terbaik yaitu 94,23 % menggunakan validasi silang 10-lipat dengan ekualisasi histogram. Metode ini mengalahkan transformasi wavelet Haar yang memberikan presisi 83,48 % dan analisis komponen utama yang menghasilkan 75,19 % presisi keakuratan. Desain sistem selanjutnya menggunakan versi basis data yang lebih kecil, 50 gambar digital dengan kondisi ekstrem hanya 2 variasi gambar yang diambil dari setiap responden. Untuk desain sistem terakhir ini, algoritma SIFT memberikan keakuratan presisi terbaik yaitu 38 % dan mengalahkan ketiga metode sebelumnya transformasi wavelet Haar, analisis komponen utama dan skala ruang Hierarki Gauss yang hanya menghasilkan sekitar 30-32% presisi keakuratan.

Kata kunci: analisis komponen utama; identifikasi biometrik; pola rambut androgenik; *scale-invariant feature transform*; transformasi wavelet Haar

THESIS VALIDATION

Title	:	Androgenic Hair Pattern Recognition for Biometric Identification
Name	:	Regina Lionnie
NIM	:	55415120017
Program	:	Electrical Engineering Graduate Program
Concentration	:	Security in ICT System
Date	:	14 October 2017

Thesis Supervisor

Prof. Dr.-Ing. Mudrik Alaydrus

Verified by:

Graduate Program Director	*Electrical Engineering Graduate Program* Head Department
Prof. Dr. Didik J. Rachbini	Prof. Dr.-Ing. Mudrik Alaydrus

ORIGINALITY STATEMENT

By signing below, I hereby confirm and acknowledge that all writing and statement in this thesis:

Title	:	Androgenic Hair Pattern Recognition for Biometric Identification
Name	:	Regina Lionnie
NIM	:	55415120017
Program	:	Electrical Engineering Graduate Program
Concentration	:	Security in ICT System
Date	:	14 October 2017

Is the results of literature study, research and the creation of my own work with guidance from my thesis supervisor which was appointed with the Letter of Decision from Electrical Engineering Graduate Program Head Department.
This thesis has never been submitted for achieving magister title on this kind of graduate program in another university. The source of all information, data and results from this thesis was declared clearly and can be checked for its truth.

Jakarta, 14 October 2017

REGINA LIONNIE

PREFACE

It has been a wonderful experience and delight opportunity to finish this research and summarize it in this thesis. I would like to express my gratitude to the almighty God above, my son, my family and friends, also to Universitas Mercu Buana Jakarta, Electrical Engineering Graduate Program, especially my supervisor and my MTEL 18 friends.

Researching is a never ending journey. One can find latest hypothesis, proves it and go back to research it again. That is the beauty of the research because we can never have enough of it. I, personally, do research to improve the condition of my country and hope that one day my research can show significant changes to Indonesia.

To all other researchers around the world, especially in Indonesia, I would like to say that if you have ever find a problem in your research or in your life, never give up. The only limitation is yourself. You, are the only one reason, preventing you to achieve your dream. Believe in yourselves and you are doing half the journey already.

Regina Lionnie

to Benji,

my sweet boy,

whose smile can light up the world....

CONTENTS

Title i
Abstract ii
Abstrak iii
Thesis Validation iv
Originality Statement v
Preface vi
Contents viii
List of Table x
List of Figure xi
List of Abbreviation xiii
List of Appendix xiv
Chapter 1. Introduction 1
1.1 Background 1
1.2 Problem Statement 3
1.3 Research Goal 3
1.4 Problem Limitation 3
1.5 Thesis Outline 4
Chapter 2. Literature Review 5
2.1 Previous Work 5
2.2 Pre-Processing 8
2.3 Scale-Invariant Feature Transform 10
2.3.1 Finding Scale Space Extrema 10
2.3.2 Keypoint Location and Filtering 14
2.3.3 Orientation Assignment 16
2.3.4 Creating Descriptor 17
2.4 Classification 17
2.5 Calculating System Performance 18
Chapter 3. Methods 19
3.1 Androgenic Hair Database 19

3.2 Pre-Processing 20
3.3 Feature Extraction and Classification 20
Chapter 4. Results and Analysis 26
4.1 Database of Androgenic Hair 26
4.2 Results and Analysis Using Haar Wavelet Transform 28
4.3 Results and Analysis Using Principal Component Analysis 29
4.4 Results and Analysis Using Hierarchical Gaussian Scale Space 31
4.5 Results and Analysis Using Combination of Pre-Processing and SIFT 37
4.6 Results and Analysis Using Proposed Method, SIFT 44
Chapter 5. Conclusion and Suggestion 52
5.1 Conclusion 52
5.2 Suggestion 53
References 54
Appendix A 58
Appendix B 59

LIST OF TABLES

Table 2.1.	Comparison of Other Previous Research, Author's Previous Research (in bold) and *Thesis* (*in italic and bold*)	7
Table 4.1.	Results of Recognition Performance in [19]	29
Table 4.2.	Performance Results: Average Recognition Precision with Hierarchical Gaussian Scale-Space for Androgenic Hair Pattern	34
Table 4.3.	Performance Results: Average Recognition precision with Hierarchical Gaussian Scale-Space with Histogram Equalization for Androgenic Hair Pattern	34
Table 4.4.	Mean Square Error Results	38
Table 4.5.	Total Matches (M) with SIFT and False Matches (FM) Results	39
Table 4.6.	Results Summary	40
Table 4.7.	Total False Matches (FM) Compared to SIFT	40
Table 4.8.	Results of True Positive (TP) or False Positive (FP)	46
Table 4.9.	Results Using Different Distance Ratio (DR) Number in True Matches/Total Matches Format.	49
Table 4.10.	Results from Haar Wavelet Transform (HWT) 49 Level 1 Decomposition for Subbands Approximation (App), Horizontal (H), Vertical (V) and Diagonal (D).	
Table 4.11.	Results from Principal Component Analysis by Using 10-100 Largest Eigenvalues (ev) and Its Corresponding Eigenvectors.	49
Table 4.12.	Results from Hierarchical Gaussian Scale Space with 4 Octaves (O) and 3 Levels (L).	50
Table 4.12.	Comparison Results by Using Haar Wavelet Transform (HWT), Principal Component Analysis (PCA), Hierarchical Gaussian Scale Space (HGSS) and SIFT Algorithm.	50

LIST OF FIGURES

Figure 2.1. Blood Vessel Pattern Recognition System 5

Figure 2.2. Example of Gaussian filter 8

Figure 2.3. Before and After Histogram Equalization of Two Different Histogram of Images (left) and Finding a Point of Operation f_{eq} to Shift Histogram Line and Create Cumulative Histogram H_{eq} is approximates linear (right) 9

Figure 2.4. Hierarchical Gaussian Scale-Space 10

Figure 2.5. Flowchart of Hierarchical Gaussian Scale-Space for U octaves and V levels 12

Figure 2.6. How to Obtain DoG from Scale-Space 13

Figure 2.7. Detecting Local Extrema with x marks the interest pixel that needs to be compared (left image) [6] and red blocks in (right image) 14

Figure 2.8. Key point Orientation Assignment: magnitude and gradient of 4x4 window (left) and assigned orientation into 8 bin (right) 16

Figure 2.9. Key point descriptor of 8x8 neighbourhood into 2x2 histogram of oriented gradient 17

Figure 3.1. The Proposed Androgenic Hair Pattern Recognition System 19

Figure 3.2. Flowchart of Proposed Method in [19] 21

Figure 3.3. Flowchart of Proposed System in [20] 22

Figure 3.4. Flowchart of Proposed Research Method in [21] 23

Figure 3.5. Flowchart of Proposed Method in [37] 25

Figure 3.6. Thesis Proposed Method Flowchart 25

Figure 4.1. The Results from LAB Colour Space 26

Figure 4.2. Wide Range of LAB Colour Space to Detect Skin Colour (1) 27

Figure 4.3. Wide Range of LAB Colour Space to Detect Skin Colour (2) 27

Figure 4.4. Wide Range of LAB Colour Space to Detect Skin Colour (3) 27

Figure 4.5. Example of Extraction Skin Colour Result 28

Figure 4.6. Results from [20] Average Precision of Androgenic Hair Recognition by Using Principal Component Analysis with and without Histogram Equalization 30
Figure 4.7. The Graph of Average Precision of Androgenic Hair Recognition for Every System Setting 35
Figure 4.8. Performance Comparison for Androgenic Hair Pattern Researches: 2-fold cross validation with Haar wavelet (2-fold Haar) [19], 10-fold cross validation with Haar wavelet (10-fold Haar) [19], Principal Component Analysis method for feature extraction (PCA) [20], PCA with histogram equalization (PCA with heq) [20] and the performance of the system in [21]: 2-fold cross validation and 10-fold cross validation with (-heq) or without histogram equalization using hierarchical Gaussian scale-space (HGSS) 36
Figure 4.9. Examples of Original and Tampered Images from The MICC-F220 Database with 3 group of images with original images (left) and its tampered forgery images (right) 37
Figure 4.10. Total Number of Keypoints Found for Every Methods (3 groups of image of total 30 images) 38
Figure 4.11. Histogram plot of an image in group 1 (left), group 2 (middle) and group 3 (right) respectively with their cumulative distribution function 41
Figure 4.12. Variation of SIFT's Small Database 44
Figure 4.13. The Histogram of Dominant Condition (left) and Non Dominant Condition (right) 48

LIST OF ABBREVIATION

GOH	Gabor orientation histogram
HWT	Haar wavelet transform
PCA	principal component analysis
HGSS	hierarchical Gaussian scale-space
SIFT	scale-invariant feature transform
IA	image alignment
FP	hair follicle position
HE	histogram equalization
SF	sharpening filter
SmF	smoothing filter
SIFT	scale-invariant feature transform
LoG	laplacian of Gaussian
DoG	difference of Gaussian
TP	true positive
FP	false positive
P	percentage
RGB	red green blue
SVD	singular value decomposition
PDF	probability distribution function
dom	dominant
NM	no match
DR	distance ratio
App	approximation subband
H	horizaontal subband
V	vertical subband
D	diagonal subband
ev	eigenvalues
O	octave
L	level

LIST OF APPENDIX

Appendix A Examples Of Androgenic Hair Database and Its Variations 58
Appendix B Proof of Concept Dot Product and Euclidean Distance 5