



Hand Gesture Spotting and Recognition Based on Generative and Discriminative Models in Stereo Color Image Sequences

Thesis

**Submitted for the partial fulfillment of the requirements
of the PhD degree in computer science**

Submitted by

Fatma Shabaan Abd Elbary Ebrahim

Supervised by

Ass. Prof.

Fayed Fayek Ghaleb
Department of Mathematics,
Faculty of Science,
Ain Shams University.

Prof.

Ebrahim Abd Allah Youness
Department of Mathematics,
Faculty of Science,
Tanta University.

Lecturer

Mahmoud Othman Elmezain
Department of Mathematics,
Faculty of Science,
Tanta University.

Submitted to

**Department of Mathematics,
Faculty of Science,
Ain Shams University,
Cairo-Egypt,
2018**

Curriculum Vitae

Name: Fatma Shaban Abd El_Bary Ebrahim.

Date of Birth: 5-2-1978, Tanta, Egypt

Nationality: Egyptian

Status: Married, three Children

E-mail: fatma1sh2008@yahoo.com

Education:

1994 – 1997 Higher Secondary Certificate, Tanta, Egypt.

1997 - 2001 B.Sc. in Pure Mathematics . Faculty of Education,
Kafer-Eshakg University, Egypt.

2001 - 2003 B. Sc. In Statistics and Computer Science. Faculty
of Science, Tanta University, Egypt.

2003-2005 Postgraduate courses for M.Sc. Faculty of Science,
Ain Shams University, Egypt.

2006 – 2009 M.Sc. in Computer Science, Faculty of Science,
Ain Shams University, Egypt.

2011 - Now Works towards Ph.D. degree Faculty of Science ,
Ain Shams University, Egypt.

Work Experience:

2003 - 2009 Demonstrator in Dept. of Statistics and Computer Science, Faculty of Science, Tanta University, Egypt.

2009 - 2016 Assistant lecturers in Dept. of Statistics and Computer Science, Faculty of Science, Tanta University, Egypt.

جامعة عين شمس
كلية العلوم

اسم الطالب : فاطمة شعبان عبد الباري ابراهيم
عنوان الرسالة: اكتشاف حركة اليد والتعرف عليها باستخدام النماذج التوليدية
والتمييزه في سلسله من الصور الملونة
اسم الدرجة: دكتوراه في علوم الحاسب

لجنة الإشراف

(١) د/ فايد فائق محمد غالب

أستاذ مساعد متفرغ بقسم الرياضيات - كلية العلوم - جامعة عين شمس.

(٢) أ.د/ ابراهيم عبدالله يونس

أستاذ بقسم الرياضيات- كلية العلوم - جامع طنطا

(٣) د/ محمود عثمان المزين

مدرس علوم الحاسب بقسم الرياضيات-كلية العلوم- جامعة طنطا

لجنة التحكيم :

(١) أ.د/ طه إبراهيم العريف

استاذ علوم الحاسب - كلية الحاسبات والمعلومات - جامعة عين شمس.

(٢) أ.د/ هالة حلمي محمد زايد

استاذ علوم الحاسب - كلية الحاسبات والمعلومات - جامعة بنها.

(٣) د / فايد فائق محمد غالب

أستاذ مساعد متفرغ بقسم الرياضيات - كلية العلوم - جامعة عين شمس.

الدراسات العليا :

أجيزت الرسالة بتاريخ / / ٢٠١٨

ختم الإجازة

موافقة مجلس الجامعة

/ / ٢٠١٨

موافقة مجلس الكلية

/ / ٢٠١٨

List of publication during PH.D study:

The presented thesis is based on the following international

Reviewed journal:

- [1] Fayed .M. Ghaleb, E. A. youness, M. Elmezain and F. Sh. Dewdar, " SVM and HMM-Based Hand Gesture Spotting and Recognition in Stereo Color Image Sequences", Journal of INFORMATION, vol. 18, No. 8, August, pp. 3641-3654, 2015.

- [2] Fayed .M. Ghaleb, E. A. youness, M. Elmezain and F. Sh. Dewdar, "Vision-Based Hand Gesture Spotting and Recognition using CRF and SVM", Journal of Software Engineering and Applications (JSEA): vol. 8 NO. 7, 2015.

- [3] Fayed .M. Ghaleb, E. A. youness, M. Elmezain and F. Sh. Dewdar, " Generative and Discriminative Models for Hand Gesture Spotting and Recognition in Stereo Color Image Sequences", INFORMATION, vol. 17 , Number 6(A) , PP. 2185- 2205, 2014 .

- [4] Fayed .M. Ghaleb, E. A. youness, M. Elmezain and F. Sh. Dewdar, "Hand gesture Spotting and Recognition in Stereo Color Image Sequences Based on Generative Models" , International Journal of Engineering Science and Innovative Technology (IJESIT) , vol. 3 , Issue 1, ISSN: 2319-5967, ISO 9001: 2008 Certified, 2014.

- [5] Fayed .M. Ghaleb, E. A. youness, M. Elmezain and F. Sh. Dewdar, "Hand gesture Spotting and Recognition Based on Discriminative Models in Stereo Color Image Sequences" , Journal of Software Engineering and Applications (JSEA), Accepted, 2014.

Acknowledgement

IN THE NAME OF ALLAH

MOST GRACEFUL AND MERCIFUL

"BESM ELLAH ERRAHMAN ERRAHEEM"

I humbly acknowledge the blessings of almighty Allah who has enabled me to complete this thesis. May Allah pray on Mohamed "Peace Be upon Him" the Prophet and the messenger of Allah.

Many thanks are dedicated to *Dr. Fayed Fayek Ghaleb* Associate professor, Department of Mathematics, Faculty of Science, Ain Shams University and Dr. **Ebrahim Abd Allah Youness**, professor, Department of Mathematics, Faculty of Science, Tanta University, they have brainstormed the principle ideas and concepts of my research and directed me to tackle them. All my sincere prayers and gratitude go to **Dr. Mahmoud Othman Elmezain**, Lecturer of computer science, Department of Mathematics, Faculty of Science, Tanta University; for her kindness, continuous interest to help, useful advices and guidance to focus on the main ideas and link between them.

I thank my parents for their constant encouragement and motivation in my pursuit for higher studies and for supporting me at every step in life. Dr. Noor (my brother), thank you for being by my side as steadfastly as ever and for all the helpful ideas and valuable suggestions. I am also grateful to all my friends for their company and support. I would like to express my special thanks

to my husband. And finally, I thank my brothers whose love and affection has seen me through all the difficult times in my life.

All appropriations for the academic staff of Computer Science of Department of Mathematics, Faculty of Science, Ain Shams University. Many thanks also go to the staff members of Department of Mathematics, Faculty of Science, Tanta University.

Fatma Shabaan Abd El bary Ebrahim.

Department of Mathematics,

Faculty of Science,

Tanta University.

Summary

The main objective of this thesis is to study the Human Computer Interaction (HCI). Mainly, we shall study the system of gesture spotting and recognition by using hand. This faced by several problems: Extracting (spotting) key gesture from the continuous hand motions sequence is the first problem which arises in the recognition of hand gesture. The second problem is that there is quite a bit of variability in trajectory, shape, and duration of the same gesture even for the same person.

Throughout literature, to discover the start points of gestures they firstly detect the end points of gestures and then track back through their optimal paths, this technique is called the backward spotting. The recognizer receives the point's trajectory between the detection of start and end points for classification. Thus, there is time delay among the spotting process of key (meaningful) gestures and the recognition process. It is being noted that the time delay is not acceptable to online implementation and applications.

In this work, we are handled meaningful gesture spotting and recognition according to two different classification techniques: generative model such as Hidden Markov Models (HMMs) with coupling of Support Vector Machine (SVM) and neural network (NNW) separately as well as discriminative models like

Conditional Random Fields (CRFs) in conjunction with SVM and NNW to specify the superior in terms of recognition outcomes.

For designing the model of non-gesture a stochastic method is proposed to spot meaningful gestures accurately with HMMs in conjunction with a SVM and NNW versus CRFs in conjunction with a SVM and NNW with no training data. For HMM, the non-gesture model is used to select the starting and ending points of key gestures as an adaptive threshold which gives a trust measure which is embedded in input videos. Firstly, the hand's segmentation is done via $YCbCr$ color space and $3D$ depth map. We captured depth map by kind of stereo camera (bumblebee) where complex background sense is neutralized. The depth map which based on passive stereo measuring has been obtained using calibration data of the camera in addition to mean absolute difference. To track a hand motion a set of hand postures is extracted by mean-shift technique coupling of depth map to correctly achieve accurate and robust hand trajectory. Secondly, in addition to the extracted features of three angles we derived also the features of Zernike moments and elliptic Fourier of dynamic affine-invariants from $3D$ spatio-temporal hand volume. Finally, the generative model of Hidden Markov Model and discriminative models of Conditional Random Fields perform the processes of spotting as well as recognition using the set of angles features. In addition, we employed the technique of forward

spotting in coupling with circular buffer mechanism which carries out hand gesture spotting and recognition simultaneously with no time delay. After that, a Support Vector Machine is employed to verify the hand shape in between the start and the end point of key/meaningful gesture, that subject powerful view invariant process.

We implemented the gestures numbers from 0 to 9 for gesture-based interaction as an application to explain the coactions of indicated contents and the validation of gesture spotting and recognition system. We trained the HMMs models by Baum-Welch (BW) algorithm while we trained CRFs by using gradient ascent along Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimization technique. By Bumblebee stereo camera system of 6mm focal point we captured the input images at 15FPS with 240×320 pixels frame resolve, application by Matlab. Founded form dataset, we get the classification results which contain 700 video piece for isolated gestures (i.e. 70 video piece for each gesture number from 0 to 9). For training NNW, HMMs, CRFs and SVMs each isolated number from 0 to 9 have 42 videos. But for testing, the data set includes 280 videos of hand gestures. Experiments explain that the suggested system can effectively spot and recognize hand gesture from continuous hand trajectories with 94.64%, 91.07, 88.93% and with 86.08, 92.50 % recognition rate for HMMs in coupling with SVM, HMMs in coupling with

NNW, NNW separately and CRFs in conjunction with NNW, CRFs in conjunction with SVM respectively.

Contents

Acknowledgement	v
Summary	vii
List of publication during PH.D study	xi
Contents	xii
List of Figures	xvi
List of Tables	xix
List of Abbreviations	xxi
Chapter 1	1
1. Introduction	1
1.1 Introduction to gesture spotting	1
1.2 Problem definition	3
1.3 Application	5
1.4 Contributions	7
1.5 Thesis Structure	10
Chapter 2	13
2. Literature Review	13
2.1 Introduction to Gesture Recognition	13
2.2 Related Work	15
2.2.1 Gesture Recognition	16
2.2.2 Gesture Spotting	18
2.2.3 Sign Language Recognition	19
2.3 Approaches of Gesture Recognition	23
2.3.1 Gesture Recognition Using Dynamic Time Warping	25
2.3.2 Gesture Recognition Using Neural Network	27
2.3.3 Gesture Recognition Using Hidden Markov Models	30
2.3.4 Gesture Recognition Using Conditional Random Fields	33
Chapter 3	36
3. Basic Techniques and Fundamental tools	36
3.1 Color spaces for skin modeling	36
3.1.1 RGB	36
3.1.2 YCrCb	37
3.2 Stereo Camera	38
3.3 K- Means Algorithm	42
3.4 Skin Color Detection	44

3.4.1 Gaussians Mixture Model	46
3.4.2 Support Vector Machines	49
3.4.3 Artificial Neural Network	52
3.4.4 Hidden Markov Model	57
3.4.4.1 Elements of HMMs	57
3.4.4.2 The main key issues	58
3.4.4.3 Topologies of HMMs	67
3.4.5 Conditional Random Fields	70
3.4.5.1 Learning CRF Parameter	73
3.4.5.2 CRFs Inference	74
3.4.6 Zernike Moment	75
3.4.7 Fourier descriptors	77
3.4.8 Relative Entropy	79
Chapter 4	82
Hand Gesture Recognition using Neural Networks	82
4.1 Proposed Approach	84
4.1.1 Preprocessing	85
4.1.2 Tracking and Feature Extraction	86
4.1.2.1 NNW-based orientation features	86
4.1.2.2 Affine-invariants features NNW-based dynamic	87
4.1.2.2.1 Elliptic Fourier descriptors	88
4.1.2.2.2 Zernike Moments	88
4.1.3 Recognition	89
4.1.3.1 Artificial Neural Network	89
4.1.3.2 Creating the NNW architecture	90
4.1.3.3 Spotting and Recognition	94
4.1.3.4 NNWs-Based Hand Gesture Verification	94

4.2	Experimental Results	95
	Chapter 5	103
	Hand Gesture Spotting and Recognition using CRF and SVM	103
	5.1 Spotting and Recognition Approach	105
	5.1.1 Preprocessing	105
	5.1.2 Tracking and Feature Extraction	108
	5.1.2.1 CRF-based orientation features	108
	5.1.2.2 Affine-invariants features SVM-based dynamic	110
	5.1.2.2.1 Elliptic Fourier descriptors	110
	5.1.2.2.2 Zernike Moments	111
	5.1.3 Spotting and Recognition	112
	5.1.3.1 Spotting with CRFs	112
	5.1.3.2 Gestures and Non-gesture Model	114
	5.1.3.3 The Parameters N-CRFs Model	115
	5.1.3.4 Forward Spotting by CRF with Circular Buffer	118
	5.1.3.5 Verification SVMs-Based hand Gesture	120
	5.2 Experimental Results	121
	5.2.1 Spotting and Recognition via CRFs	121
	Chapter 6	128
	SVM and HMM-Based Hand Gesture Spotting and Recognition in Stereo Color Image Sequences	128
	6.1 Hand Gesture Spotting and Recognition Approach	128
	6.1.1 Preprocessing	128
	6.1.2 Tracking and Feature Extraction	129
	6.2 Spotting and Recognition	129

6.2.1 HMM Forward Spotting	130
6.2.2 SVMs-Based Gesture Verification	134
6.3 Experimental Results	135
Chapter 7	144
Conclusions and Future Work	144
7.1 Conclusions	144
7.2 Future Work	148
Appendices:	150
Appendix A	150
Appendix B	188
References	191
الملخص العربي	203