Real-time Vision-based Hand and Face Tracking and Recognition of Gesture

A PhD dissertation submitted in partial fulfillment of the requirement for the degree of

Doctor of Philosophy (Ph.D.)

in

Computer Science

by

Farhad Dadgostar

Institute of Information and Mathematical Sciences

College of Science

Massey University

December 2006

Real-time Vision-Based Hand and Face Tracking and Recognition of Gesture

Farhad Dadgostar

Dedication

With love, to the one who always encouraged me and provided me indefinite unconditional support…

To my wife Nasim

Acknowledgements

I would like to thank Dr. Abdolhossein Sarrafzadeh, my supervisor, for all of his support, his patience, and his guidance and endless encouragement throughout my PhD study. Without his support, I would have never gained such an interest in research and nor learned as much as I have. I have been lucky and am proud of being a student of him. Working with him was a great pleasure.

I would like to thank to Dr. Scott P. Overmyer and Dr. Liyanage De Silva for their valuable suggestions and comments on my work.

I also have enjoyed working with a number of fellow PhD colleagues over the past years. In particular Andre Barczak and Chao Fan with whom sharing the authorship of several papers and reports reflect the closeness of our cooperation.

The financial support provided by the Massey University and the Institute of Information and Mathematical Sciences made my research and presenting the results in international conferences possible. In particular, I would like to thank Professor Robert McKibbin, the head of the institute for his support during my study.

This achievement would not have been possible without endless support provided to me by my family. I would like to thank my parents who have always been supportive and encouraging through all the time of my study. I would like to thank my precious daughter Kiana for her company and patience when I was working on weekends. Finally, I would like to express my deep thanks to my lovely wife, Nasim, for all her love, sacrifice, support and encouragement which could be found in every word, in this work.

List of Publications

Journal Articles

- **Farhad Dadgostar**, and Abdolhossein Sarrafzadeh, *"An adaptive real-time skin detector based on Hue thresholding: A comparison on two motion tracking methods"*, In Pattern Recognition Letters, Vol. 27, Issue 12, pp. 1342-1352, March 2006, Elsevier.
- Abdolhossein Sarrafzadeh, Samuel Alexander, **Farhad Dadgostar**, Chao Fan, and Abbas Bigdeli, *"How do you know that I don't understand? A look at the future of intelligent tutoring systems"*, Accepted for publication in Journal of Computers in Human Behavior, 2006.

Book Chapters

- **Farhad Dadgostar**, and Abdolhossein Sarrafzadeh, *"A Fast Skin Detection Algorithm for Video Sequences"*, In Mohamed Kamel, Aurelio Campilho (Ed.), Lecture Notes in Computer Science, ICIAR 2005, Vol. 3656, pp. 804-811, 2005, ISBN 3-540-29069-9, Springer-Verlag, Berlin, Heidelberg.
- **Farhad Dadgostar**, Abdolhossein Sarrafzadeh, and Scott P. Overmyer, *"Face Tracking Using Mean-Shift Algorithm: A Fuzzy Approach for Boundary Detection"*, In J. Tao, T. Tan and R. W. Picard (Ed.), Lecture Notes in Computer Science: Affective Computing and Intelligent User Interfaces, Vol. 3784, pp. 56-63, 2005, ISBN 3-540-29621-2, Springer-Verlag, Berlin, Heidelberg.
- **Farhad Dadgostar**, Hokyoung Ryu, Abdolhossein Sarrafzadeh, and Scott P. Overmyer, *"Making sense of student use of nonverbal cues for intelligent tutoring systems"*, In the ACM International Conference Proceeding Series: 19th conference of the computer-human interaction special interest group (CHISIG), Vol. 122, pp 1-4, 2005, ISBN 1-59593-222-4, Canberra, Australia.

International Refereed Conferences

- **Farhad Dadgostar**, Abdolhossein Sarrafzadeh, Chao Fan, Liyanage De Silva, and Chris Messom, *"Modeling and Recognition of Gesture Signals in 2D Space: A comparison of NN and SVM approaches"*, The 18th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2006, Washington D.C., USA.
- **Farhad Dadgostar**, Abdolhossein Sarrafzadeh, Scott P. Overmyer, and Liyanage De Silva, *"Is the Hand really quicker than the Eye? Variances of the Mean-Shift algorithm for real-time hand and face tracking"*, IEEE International Conference on

Computational Intelligence for Modelling, Control and Automation, CIMCA 2006, Sydney, Australia.

- **Farhad Dadgostar**, Abdolhossein Sarrafzadeh, and Scott P. Overmyer, *"Genetic Algorithms and Long-Haar features: A Method for Object Detection"*, The 3rd International Conference on Cybernetics and Information Technologies, Systems and Applications, CITSA 2006, Orlando, Florida, USA.
- Abdolhossein Sarrafzadeh, Samuel Alexander, **Farhad Dadgostar**, Chao Fan, and Abbas Bigdeli, *"See Me, Teach Me: Facial Expression and Gesture Recognition for Intelligent Tutoring Systems"* , IEEE International Conference on Innovation in Information Technology, IIT 2006, Dubai, U.A.E.
- **Farhad Dadgostar**, Abdolhossein Sarrafzadeh, Hokyoung Ryu, *"A Macro model of Human Emotional-Response for Intelligent Agents Applications"*, 1st Korea-New Zealand Joint Workshop on Advances of Computational Intelligent Methods and Applications, Feb 2006, Auckland, New Zealand.
- **Farhad Dadgostar**, Abdolhossein Sarrafzadeh, *"A Component-based Architecture for Vision-based Gesture Recognition"*, In the proceedings of the International Image and Vision Computing Conference, IVCNZ 2005, pp. 322-328, Dunedin, New Zealand.
- Chao Fan, Abdolhossein Sarrafzadeh, **Farhad Dadgostar**, and Hamid Gholamhosseini, *"Facial Expression Analysis by Support Vector Regression"*, In the proceedings of the International Image and Vision Computing Conference, IVCNZ 2005, pp. 311-317, Dunedin, New Zealand.
- **Farhad Dadgostar**, Chao Fan, and Abdolhossein Sarrafzadeh, *"A Hybrid Approach for Robust Real-time Face Tracking in Video Sequences"*, In the proceedings of the Institute of Information and Mathematical Sciences Postgraduate Conference, pp. 35-42, 2005, Auckland, New Zealand.
- Chao Fan, Abdolhossein Sarrafzadeh, **Farhad Dadgostar**, and Hamid Gholamhosseini, *"Face and Eye Detection Using Support Vector Machines"*, The 3rd International Conference on Computational Intelligence, Robotics and Autonomous Systems, CIRAS 2005, Singapore.
- **Farhad Dadgostar**, Abdolhossein Sarrafzadeh, and Martin Johnson, *"An Adaptive Skin Detector for Video Sequences Based on Optical Flow Motion Features"* , In the proceedings of the 7th International IEEE Conference on Image and Signal Processing, IASTED-SIP 2005, In M.W. Marcellin (Ed.), Hawaii, USA.
- Andre L. C. Barczak, **Farhad Dadgostar**, and Martin Johnson, *"Real-time Hand Tracking Using the Viola and Jones Method"* , In the proceedings of the 7th International IEEE Conference on Image and Signal Processing, IASTED-SIP 2005, In M.W. Marcellin (Ed.), Hawaii, USA.
- **Farhad Dadgostar**, Abdolhossein Sarrafzadeh, and Scott P. Overmyer, *"An Adaptive Real-time Skin Detector for Video Sequences"*, In the Proceedings of the 2005 International Conference on Computer Vision, Vision 2005, H. R. Arabnia (Ed.), CSREA Press 2005, ISBN 1-932415-65-3, pp. 65-70, Las Vegas, Nevada, USA.
- Andre L. C. Barczak, **Farhad Dadgostar**, and Chris Messom, *"Real-time Hand Tracking Based on Non-invariant Features"*, Proceedings of the International IEEE Conference on Measurement and Instrumentation, IMTC 2005, Vol. 3, pp 2192-2192, Ottawa, Canada.

- Chao Fan, **Farhad Dadgostar**, Abdolhossein Sarrafzadeh, Hamid Gholamhosseini, and Martin Johnson, *"Facial Expression Reconstruction Using Polygon Approximation"*, In the proceedings of the 7th International IEEE Conference on Image and Signal Processing, IASTED-SIP 2005, In M.W. Marcellin (Ed.), Hawaii, USA.
- Abdolhossein Sarrafzadeh, Chao Fan, **Farhad Dadgostar**, Sam Alexander, and Chris Messom, *"Frown Gives Game Away: Affect Sensitive Tutoring Systems for Elementary Mathematics"*, International IEEE Conference on Systems, Man and Cybernetics, SMC 2004, The Hague, The Netherlands.

Other Publications

- **Farhad Dadgostar**, Abdolhossein Sarrafzadeh, *Gesture recognition through angle space*, Research Letters in the Information and Mathematical Sciences, 2006, Vol. 9, pp 112-119.
- **Farhad Dadgostar**, Abdolhossein Sarrafzadeh, *A formal model of emotional-response, inspired from human cognition and emotion systems*, Research Letters in the Information and Mathematical Sciences, 2006, Vol. 9, pp 89-97.
- **Farhad Dadgostar**, Andre L. C. Barczak, Abdolhossein Sarrafzadeh, *A Color Hand Gesture Database for Evaluating and Improving Algorithms on Hand Gesture and Posture Recognition*, Research Letters in the Information and Mathematical Sciences, ISSN 1175-2777, 2005, Vol. 7, pp 127-134.
- Andre L. C. Barczak, **Farhad Dadgostar**, *Real-time Hand Tracking Using a Set of Cooperative Classifiers and Haar-Like Features*, Research Letters in the Information and Mathematical Sciences, ISSN 1175-2777, 2005, Vol. 7, pp 29-42.

Abstract

In this dissertation, we present the research pathway to the design and implementation of a real-time vision-based gesture recognition system. This system was built based on three components, representing three layers of abstraction: i) detection of skin and localization of hand and face, ii) tracking multiple skin blobs in video sequences and finally iii) recognition of gesture movement trajectories.

The adaptive skin detection, the first component, was implemented based on our novel *adaptive skin detection algorithm* for video sequences. This algorithm has two main sub-components: i) the static skin detector, which is a skin detection method based on the hue factor of the skin color, and ii) the adaptive skin detector which retrains itself based on new data gathered from movement of the user. The results of our experiments show that the algorithm improves the quality of skin detection within the video sequences.

For tracking, a new approach for boundary detection in blob tracking based on the Mean-shift algorithm was proposed. Our approach is based on continuous sampling of the boundaries of the kernel and changing the size of the kernel using our novel Fuzzy-based algorithm. We compared our approach to the kernel density-based approach, which is known as the CAM-Shift algorithm, in a set of different noise levels and conditions. The results show that the proposed approach is superior in stability against white noise, and also provides correct boundary detection for arbitrary hand postures, which is not achievable by the CAM-Shift algorithm.

Finally we presented a novel approach for gesture recognition. This approach includes two main parts: i) gesture modeling, and ii) gesture recognition. The gesture modeling technique is based on sampling the gradient of the gesture movement trajectory and presenting the gesture trajectory as a sequence of numbers. This technique has some important features for gesture recognition including robustness against slight rotation, a small number of required samples, invariance to the start position and device independence. For gesture recognition, we used a multi-layer feed-forward neural-network. The results of our experiments show that this approach provides 98.71% accuracy for gesture recognition, and provides a higher accuracy rate than other methods introduced in the literature.

These components form the required framework for vision-based real-time gesture recognition and hand and face tracking. The components, individually or as a framework, can be applied in scientific and commercial extensions of either vision-based or hybrid gesture recognition systems.

Table of Contents

List of Figures

List of Tables

List of Acronyms

ANN	Artificial Neural Network
AP	Attention Point
ASD	Adaptive Skin Detector
ASL	American Sign Language
BBN	Bayesian Belief Network
CAM-Shift	Continuously-Adaptive Mean-Shift
CCD	Charged Coupled Device
CIELAB,CIEXYZ	Both color spaces, refer to perceptually linear color spaces known as CIE-L-A-B and CIE-X-Y-Z
CP	Color Predicate
CPU	Central processing Unit
CRCNN	Hyper Rectangular Composite Neural Network
DOG	Difference of Gaussian
DTW	Dynamic Time Warping
GHz	Giga Hertz
GMM	Gaussian Mixture Model
GSD	Global Skin Detector
HCE	Histogram Error
HCI	Human-Computer Interaction
HI	Histogram Intersection
HIS	Hue-Saturation-Intensity (Color model)
HMM	Hidden Markov Model
HS	Hue-Saturation (color model)
HSV	Hue-Saturation-Value (Color model)
ICrCb	Intensity-Chrominance red-Chrominance blue (Color model)
ICT	Information-Communication Technology
IUV	Intensity-Luminance-Chrominance (Color model)
LCS	Localized Contour Sequence
MSEPF	Mean-Shift Embedded Particle Filter
NN	Neural Network
PC	Personal Computer

PCA	Principal Component Analysis
PDA	Personal Digital Assistant
PDF	Probability Density Function
PUI	Perceptual User Interface
RBF	Radial Basis Function
RGB	Red-Green-Blue (Color model)
SASOM	Structure Adaptive Self-Organized Map
SMA	Specialized Mapping Architecture
SOM	Self-Organized Map
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine
tMHI	Time-Motion History Image
TMS	Transcranial Magnetic Stimulation
UAV	Unmanned Aerial Vehicle
UI	User Interface
UR	Unreliability Rate
USB	Universal Serial Bus
YCrCb	YCrCb is an encoded nonlinear RGB signal, commonly used by European television studios and for image compression.
YIQ	Color model inspired from human vision system, formerly used in NTSC television broadcasting
YUV	Luminance-Chrominance (Color model, mainly is used in PAL analog television broadcasting)