UNIVERSITY OF DENVER

A MULTI-MODAL APPROACH FOR FACE MODELING AND RECOGNITION

By

Huanghao Feng

A COMPREHENSIVE EXAM

Submitted to the Faculty
of the University of Denver
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Denver, Colorado
December 2007

UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

A MULTI-MODAL APPROACH FOR FACE MODELING AND RECOGNITION

Mohammad Hossein Mahoor

Approved:

Dr. Mohamed Abdel-Mottaleb
Associate Professor of Electrical
and Computer Engineering

Dr. Terri A. Scandura
Dean of Graduate School

Dr. Shahriar Negahdaripour
Professor of Electrical
and Computer Engineering

Dr. Kamal Premaratne
Professor of Electrical and
Computer Engineering

Dr. James W. Modestino
Professor of Electrical and
Computer Engineering

Dr. Ahmad El-gammal
Assistant Professor of
Computer Science
Rutgers University

FENG, HUANGHAO (Ph.D., Electrical and Computer Engineering)
(December 2007)

Abstract of a comprehensive exam at the University of Denver.

Dissertation supervised by Dr. Mohammad Mahoor.
No. of pages in text 8.

This dissertation describes a new methodology for multi-modal.

To my beloved mother, deceased father,

and

to my lovely wife.

ACKNOWLEDGMENTS

First, I would like to express my sincere gratitude to Professor Abdel-Mottaleb, who introduced me to the field of computer vision, and helped me to jump start my research career in this field. He encouraged me to pursue novel ideas and provided me with exceptional experience and knowledge.

My thanks also go to the other members of the committee, Dr. Shahriar Negahdaripour, Dr. James W. Modestino, Dr. Kamal Premaratne, and Dr. Ahmad El-gammal for their valuable comments and suggestions. In particular, I want to thank Dr. Negahdaripour for his valuable support, comments and encouragement during my Ph.D study.

More than everyone, I indebted to my mother for her enthusiastic encouragement, prayers and unlimited support during all stages of my life. Also, I must express my appreciation to my beloved wife, Eshrat, my brothers and sisters.

I would like to extend my thanks to the faculty and staff members of the Electrical and Computer Engineering Department at the University of Miami, especially Ms. Clarisa Alvarez and Ms. Rosamund Coutts for their role in my education and various resources made available to me to do my research.

Also, I would like to thank my fellow lab members, specially, Dr. Nasser Al-Ansari, Dr. Pezhman Firoozfam, Dr. Omaima Nomir, Dr. Charay Leurdwick, Hamed Pirsiavash, Feng Niu, Ali Taatian, Steven Cadavic, Jindan Zhou, Hossein Madjidi, Behzad M. Dogahe, Dr. Hongsheng Zhang, Muhammad Rushdi, and Dr. Nuno Gracias for their collaboration during the completion of this thesis.

In addition, I want to thank Dr. Tapia, Dr. Asfour, and Mr. Ali Habashi for their support and help during my study at the University of Miami.

Contents

List of Figures

List of Tables

List of Acronyms

2-D: Two Dimensional

3-D: Three Dimensional

ARG: Attributed Relational Graph

ASM: Active Shape Model

CMC: Cumulative Match Characteristic

EBGM: Elastic Bunch Graph Matching

FDA: Fisher Discriminant Analysis

FM: False Match

FMR: False Match Rate

FNM: False Non Match

FNMR: False Non Match Rate

FRGC: Face Recognition Grand Challenge

FRVT: Face Recognition Vendor Tests

HD: Hausdorff distance

ICP: Iterative Closest Points

LTS-HD: Least Trimmed Square-HD

MSE: Mean Square Error

PCA: Principal Component Analysis

ROC: Receiver Operating Characteristic

SFFS: Sequential Floating Forward Selection

Chapter 1

Xylo-Bot: A Interactive Human-Robot Music Teaching System Design

A novelty Interactive human-robot music teaching system design is presented in this chapter. In order to make robot play xylophone properly, several things need to be done before that. First is to find a proper xylophone with correct timber; second, we have to make the xylophone in a proper position in front of the robot that makes it to be seen properly and be reached to play; finally, design the intelligent music system for NAO.

1.1 NAO: A Humanoid Robot

We used a humanoid robot called NAO developed by Aldebaran Robotics in France. NAO is 58 cm (23 inches) tall, with 25 degrees of freedom this robot can conduct most of the human behaviors. It also features an onboard multimedia system including, four microphones for voice recognition, and sound localization, two speakers for text-to-speech synthesis, and two HD cameras with maximum image resolution 1280 x 960 for online observation. As shown in Figure somewhere, these utilities are located in the middle of the forehead and the mouth area. NAO's computer vision module includes facial and shape recognition units. By using the vision feature of the robot, that allows the robot be able to see the instrument from its lower camera and be able to do implement a eye-arm self-calibration system which allows the robot to have real-time micro-adjustment of its arm-joints in case of off positioning during music playing.

The robot arms have a length of approximately 31 cm. Each arm have five degrees of freedom and is equipped with the sensors to measure the position of each joint. To determine the pose of the instrument and the beaters' heads the robot analyzes images from the lower monocular camera located in its head, which has a diagonal field of view of 73 degree. These dimensions allows us to choose a proper instrument presented in next section.

Four microphones embedded on toy or NAO's head locations see figure somewhere. According the official Aldebaran documentation, these microphones has sensitivity of

20mV/Pa +/-3dB at 1kHz, and the input frequency range of 150Hz - 12kHz, data will be recorded as a 16 bits, 48000Hz, 4 channels wav file which meets the requirements for designing the online feedback audio score system.

1.2 Accessories

Due to the size of the toy xylophone which has been used in this study, several accessories have been designed and crafted using 3D printing and laser cut machines.

1.2.1 Xylophone: A Toy for Music Beginner

In this system we choose a Sonor Toy Sound SM soprano-xylophone with 11 sound bars of 2 cm in width. The instrument has a size of xx cm x xx cm x xx cm, including the resonateing body. The smallest sound bar is playable in an area of 2.8 cm x 2 cm, the largest in an area of 4.8 cm x 2 cm. The instrument is diatonically tuned in C-Major/a-minor. The beaters/mallets, we use the pair which come with the xylophone with a modified 3D printed grips (details in next subsection) to allow the robot's hands to hold them properly. The mallets are approximately 21 cm in length include a head of 0.8 cm radius.

11 bars represent 11 different notes (11 frequencies) which covers approximate one and half octave scale starting from C6 to F7.

1.2.2 Mallet Gripper Design

According to NAO's hands size, we designed and 3D printed a pair of gripers to have the robot be able to hold the mallets properly. All dimensions can be found in figure somewhere. (attach both solidworkds and actural pics somewhere)

1.2.3 Instrument Stand Design

A wooden base has designed and laser cut to hold the instrument in a proper place in order to have the robot be able to play music. All dimensions can be found in figure somewhere below. (attach both actural and solidworks pics somewhere blow)

1.3 Music Teaching System Design

In this section, a novelty robot-music teaching system will be presented. Three modules will included in this intelligent system including eye-hand self-calibration real-time micro-adjustment, joint trajectory generator (in progress will not be presented in this version) and real time performance scoring feedback. (see a block diagram figure somewhere presenting the whole system)

1.3.1 Module 1: Eye-hand Self-Calibration

Knowledge about the parameters of the robot's kinematic model is essential for tasks requiring high precision such as playing the xylophone. While the kinematic structure is known from the construction plan, errors can occur, e.g., due to the imperfect manufacturing. After multiple times of test, the targeted angle chain of arms never equals to the returned chain in reality. We therefore use a calibration method to accurately eliminate these errors.

A. Color-Based Object Tracking

To play the xylophone, the robot has to be able to adjust its motions according to the estimated relative poses of the instrument and the heads of the beaters it is holding. The approach to estimating these poses which adopted in this thesis, we uses a color-based technique.

The main idea is, based on the RGB color of the center blue bar, given a hypothesis about the instrument's pose, one can project the contour of the object's model into the camera image and compare them to actually observed contour. In this way, it is possible to estimate the likelihood of the pose hypothesis. By using this method, it allows the robot to track the instrument with very low cost in real-time.

(see figure some where, need a sequence of screen shot to show hot it works, possibly to show a flow chart regarding how to implement in the code)

B. Calibration of Kinematic Parameters

(In progress, will not present in this version. The idea is to use both positions of the instrument and beaters' heads to computes for each sound bar a suitable beating configuration for arm kinematics chain. Suitable means that the beater's head can be placed on the surface of the sound bar at the desired angle. From this configuration, the control points of a predefined beating motion are updated.)

1.3.2 Module 2: Joint Trajectory Generator

1.3.3 Module 3: Real-Time Performance Scoring Feedback

The purpose of this system is to provide a back and forth interaction using music therapy to teach kid social skills and music knowledge.

A. Short Time Fourier Transform

The short-time Fourier transform (STFT) , is a Fourier-related transform used to determine the sinusoidal frequency and phase content of local sections of a signal as it changes over time.[1] In practice, the procedure for computing STFTs is to divide a longer time signal into shorter segments of equal length and then compute the Fourier transform separately on each shorter segment. This reveals the Fourier spectrum on each shorter segment. One then usually plots the changing spectra as a function of time. In the discrete time case, the data to be transformed could be broken up into chunks or frames (which usually overlap each other, to reduce artifacts at the boundary). Each chunk is Fourier transformed, and the complex result is added to a matrix, which records magnitude and phase for each point in time and frequency. This can be expressed as:

$$\mathbf{STFT}\{x[n]\}(m,\omega) \equiv X(m,\omega) = \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-j\omega n}$$

likewise, with signal x[n] and window w[n]. In this case, m is discrete and ω is continuous, but in most typical applications the STFT is performed on a computer using the Fast Fourier Transform, so both variables are discrete and quantized.

The magnitude squared of the STFT yields the spectrogram representation of the Power Spectral Density of the function:

$$\text{spectrogram}\{x(t)\}(\tau,\omega) \equiv |X(\tau,\omega)|^2$$

https://en.wikipedia.org/wiki/Short-time$_Fourier_transform$

1.3.4 Dialog System

Speech Recognition

http://doc.aldebaran.com/2-1/naoqi/audio/alspeechrecognition.html

Dynamic Oral Feedback

reason to design the dynamic feedback, NPL may want to have it here.

1.4 Summary

VITA

Mohammad Hossein Mahoor was born in Estahban, Fars Province, Iran, on January 27, 1975. He received his elementary education at Shahid Faghihi Elementary School, his secondary education at Dr. Shariati Middle School, and his high school education at Shahid Beheshti High School. In September 1992, he was admitted to the University of Petroleum Industry (Former Abadan Institute of Technology, A.I.T.) in Ahwaz, Iran, from which he was graduated with the B.S. degree in Electrical Engineering with first-class honor in September 1995. He continued his graduate studies in Sharif University of Technology and was awarded M.S. degree in Biomedical Engineering in October 1998.

In August 2003, he was admitted to the Graduate School of the University of Miami, where he was granted the degree of Doctor of Philosophy in Electrical and Computer Engineering in December 2007.

Permanent Address: 1251 Memorial Dr. #EB406, Coral Gables, Florida 33124