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 13.

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 and X-Elophone

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.

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 Module-Based Acoustic Music Interactive System Design

In this section, a novelty module-based robot-music teaching system will be presented. Three modules have built in this intelligent system including module 1: eye-hand self-calibration micro-adjustment; module 2: joint trajectory generator; and module 3: real time performance scoring feedback. (see a block diagram figure somewhere presenting the whole system)

1.3.1 Module 1: Eye-hand Self-Calibration Micro-Adjustment

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.

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

Our system parses a list of numerical numbers (from 1 to 11) to obtain the sequence of notes to play. It converts the notes into a joint trajectory using the beating configurations obtained from inverse kinematic as control points. The timestamps for the control points will be defined by user in order to meet the experiment requirement. The trajectory is then computed using Bezier interpolation in joint space by the manufacturer-provided API and send to the robot controller for execution. In this way, the robot plays in-time with the song.

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. This module creates the

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$_Fourier_transform$

1.4 X-Elophone: A Revolution of Xylophone

reason why need this design. Due to the limitation of keys. This provides more possibility for different timber and major minor keys. That allows this system to play more customized song which kids love.

1.4.1 Components Selection

Piezo Vibration Sensor

The LDT0-028K is a flexible component comprising a 28 m thick piezoelectric PVDF polymer film with screen-printed Ag-ink electrodes, laminated to a 0.125 mm polyester substrate, and fitted with two crimped contacts. As the piezo film is displaced from the mechanical neutral axis, bending creates very high strain within the piezopolymer and therefore high voltages are generated. When the assembly is deflected by direct contact, the device acts as a flexible "switch", and the generated output is sufficient to trigger MOSFET or CMOS stages directly. If the assembly is supported by its contacts and left to vibrate "in free space" (with the inertia of the clamped/free beam creating bending stress), the device will behave as an accelerometer or vibration sensor. Adding mass, or altering the free length of the element by clamping, can change the resonant frequency and sensitivity of the sensor to suit specific applications. Multi-axis response can be achieved by positioning the mass off center. The LDTM-028K is a vibration sensor where the sensing element comprises a cantilever beam loaded by an additional mass to offer high sensitivity at low frequencies. https://cdn.sparkfun.com/datasheets/Sensors/ForceFlex/LDT$_{S}eries.pdf$

Also have to show the circuit, how to design this and attach the figure from here https://www.sparkfun.com/datasheets/Sensors/Flex/MSI-techman.pdf page 39

Op-Amp

An operational amplifier (often op-amp or opamp) is a DC-coupled high-gain electronic voltage amplifier with a differential input and, usually, a single-ended output.[1] In this configuration, an op-amp produces an output potential (relative to circuit ground) that is typically hundreds of thousands of times larger than the potential difference between its input terminals. Operational amplifiers had their origins in analog computers, where they were used to perform mathematical operations in many linear, non-linear, and frequency-dependent circuits. The popularity of the op-amp as a building block in analog circuits is due to its versatility. By using negative feedback, the characteristics of an op-amp circuit, its gain, input and output impedance, bandwidth etc. are determined by external components and have little dependence on temperature coefficients or engineering tolerance in the op-amp itself. Op-amps are among the most widely used electronic devices today, being used in a vast array of consumer, industrial, and scientific devices. Many standard IC op-amps cost only a few cents in moderate production volume; however, some integrated or hybrid operational amplifiers with special performance specifications may cost over US 100 in small quantities.[2] Op-amps may be packaged as components or used as elements of more complex integrated circuits. The op-amp is one type of differential amplifier. Other types of differential amplifier include the fully differential amplifier (similar to the op-amp, but with two outputs), the instrumentation amplifier (usually built from three op-amps), the isolation amplifier (similar to the instrumentation amplifier, but with tolerance to common-mode voltages that would destroy an ordinary op-amp), and negative-feedback amplifier (usually built from one or more op-amps and a resistive feedback network). https://en.wikipedia.org/wiki/Operational$_{a}mplifierhttps$: $//ww1.microchip.com/downloads/en/DeviceDoc/21733j.pdf$

Multiplexer

In electronics, a multiplexer (or mux) is a device that selects between several analog or digital input signals and forwards it to a single output line.[1] A multiplexer of $2^n 2^n$ inputs has n n select lines, which are used to select which input line to send to the output.[2] Multiplexers are mainly used to increase the amount of data that can be sent over the network within a certain amount of time and bandwidth.[1] A multiplexer is also called a data selector. Multiplexers can also be used to implement Boolean functions of multiple variables. An electronic multiplexer makes it possible for several signals to share one device or resource, for example, one A/D converter or one communication line, instead of having one device per input signal. Conversely, a demultiplexer (or demux) is a device taking a single input and selecting signals of the output of the compatible mux, which is connected to the single input, and a shared selection line. A multiplexer is often used with a complementary demultiplexer on the receiving end.[1] An electronic multiplexer can be considered as a multiple-input, single-output switch, and a demultiplexer as a single-input, multiple-output switch.[3] The schematic symbol for a multiplexer is an isosceles trapezoid with the longer parallel side containing the input pins and the short parallel side containing the output pin.[4] The schematic on the right shows a 2-to-1 multiplexer on the left and an equivalent switch on the right. The *sel* sel wire connects the desired input to the output. The 74HC4051; 74HCT4051 is a single-pole octal-throw analog switch (SP8T) suitable for use in analog or digital 8:1 multiplexer/demultiplexer applications. The switch features three digital select inputs (S0, S1 and S2), eight independent inputs/outputs (Yn), a common input/output (Z) and a digital enable input (E). When E is HIGH, the switches are turned off. Inputs include clamp diodes. This enables the use of current limiting resistors to interface inputs to voltages in excess of VCC. https://en.wikipedia.org/wiki/Multiplexer https://cdn.sparkfun.com/assets/learn$_t utorials$/5/5/3/74$HC_H CT$4051.pdf

Arduino UNO

The Arduino Uno is an open-source microcontroller board based on the Microchip ATmega328P microcontroller and developed by Arduino.cc.[2][3] The board is equipped with sets of digital and analog input/output (I/O) pins that may be interfaced to various expansion boards (shields) and other circuits.[1] The board has 14 Digital pins, 6 Analog pins, and programmable with the Arduino IDE (Integrated Development Environment) via a type B USB cable.[4] It can be powered by the USB cable or by an external 9-volt battery, though it accepts voltages between 7 and 20 volts. It is also similar to the Arduino Nano and Leonardo.[5][6] The hardware reference design is distributed under a Creative Commons Attribution Share-Alike 2.5 license and is available on the Arduino website. Layout and production files for some versions of the hardware are also available.

The word "uno" means "one" in Italian and was chosen to mark the initial release of the Arduino Software.[1] The Uno board is the first in a series of USB-based Arduino boards,[3] and it and version 1.0 of the Arduino IDE were the reference versions of Arduino, now evolved to newer releases.[4] The ATmega328 on the board comes preprogrammed with a bootloader that allows uploading new code to it without the use of an external hardware programmer.[3]

While the Uno communicates using the original STK500 protocol,[1] it differs from all preceding boards in that it does not use the FTDI USB-to-serial driver chip. Instead, it uses the Atmega16U2 (Atmega8U2 up to version R2) programmed as a USB-to-serial converter.[7] https://en.wikipedia.org/wiki/Arduino$_{U}noshowblockdiagramofthecode:$

1.4.2 ChucK: A On-the-fly Audio Programming Language

https://www.researchgate.net/profile/Ge$_{W}$ang9/publication/259326122$_{T}$he$_{C}$hucK$_{P}$rogramming$_{L}$a
Timed$_{O}$n−the−fly$_{E}$nvironmentality/links/0c96052b02acb79c2c000000.pdfbrieflydescribethelang
butonlytotheextentthatweareabletoexpresstothecomputerwhattodo, andhowtodoit.Tothisend, thepr
specificȷlanguagescanbringadditionalexpressiveness, conciseness, andperhapsevendifferentway
purposeprogramminglanguagetailoredforcomputermusic.Thegoalistocreatealanguagethatisexpre
basedconcurrentprogrammingmodelthatallowsprogrammerstoflexiblyandpreciselycontroltheflo
timedȷ), andfacilitiestodevelopprogramson−the−fly−astheyrun.AChucKianapproachtolivecodin
the−flyprogramming, tovisualizeandmonitorChucKprogramsinreal−time, andtoprovideaplatfo
mediatedensembles.Additionalapplicationsarealsodescribed, includingclassrooms, livecodingaren
basedprogrammingmechanism(bothlanguageandunderlyingimplementation)forultrapreciseaudi
timeaudioanalysis.2)Anonpreemptive, time/event−basedconcurrentprogrammingmodelthatprov
the−fly.Thisrapidprototypingmentalityhaspotentiallywideramificationsinthewaywethinkabout
timeaudio), aswellasnewparadigmsandpracticesincomputer−mediatedliveperformance.4)TheAu
andhowthesetwodisciplinescanreinforceeachother.

1.5 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