Applications of Convolutional Neural Networks to Facial Detection and Recognition for Augmented Reality and Wearable Computing

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Engineering

Christopher R. Mitchell

Problem Statement

Facial detection and recognition are among the most heavily researched fields of computer vision and image processing. However, the computation necessary for most facial processing tasks has historically made it unfit for real-time applications. Facial detection within an augmented reality framework has myriad applications, including potential uses for law enforcement, medical personnel, and patients with post-traumatic or degenerative memory loss or visual impairments. Although the hardware is now available, few portable or wearable computing systems exist that can localize and identify individuals for real-time or near-real-time augmented reality.

Abstract

The author presents a system design and implementation that performs facial detection and recognition robust to variations in lighting, pose, and scale. Scouter combines a commodity netbook computer, a high-resolution webcam, and display glasses into a light and powerful wearable computing system platform for real-time augmented reality and near-real-time facial processing. A convolutional neural network performs precise facial localization, a Haar cascade detector is used for facial feature registration, and a Fisherface implementation recognizes size-normalized faces. A novel multiscale voting and overlap removal algorithm is presented to boost face localization accuracy, and a failure-resilient normalization method is introduced.

Background

Facial Recognition Algorithms:
- •PCA and Eigenfaces: Simple, sensitive to lighting and pose
- •Fisherfaces: More robust to changes in lighting and pose
- •Neural Networks: Slow, moderate accuracy
- •Other methods: CNNs, SVMs, HMMs

Facial Detection Algorithms:
- •Eigenfaces: Also sensitive to lighting, accessories, pose
- •Eigenfeatures: More robust
- •Haar Cascades: Fast and accurate
- •Convolutional Neural Networks (CNNs): Fast and very robust

Wearable Computing and Augmented Reality Projects
- •Context-aware system using audio input and output
- •Monocular video capture and display for augmented memory
- •PDA-based augmented reality navigation
- •View augmentation with virtual 3D objects

More Information

Thesis and Defense Presenation
http://www.cemetech.net/projects/ee/scouter

Contact the author
christopher@cemetech.net

Hardware Platform

Computing platform: eeePC 901
- •1.6GHz Intel Atom N270
- •1GB RAM, 20GB SSD
- •Bluetooth, WiFi
- •3-4 hour battery life under Scouter load
- •2.43 pounds
- •Poor floating-point performance

Output: Vuzix VR920 HUD
- •Dual 640x480 LCDs
- •Gyroscope, earbuds, microphone
- •VGA input, USB power

Input: Logitech Quickcam 9000
- •Up to 1600x1200 at 24FPS
- •Used in 640x480 mode

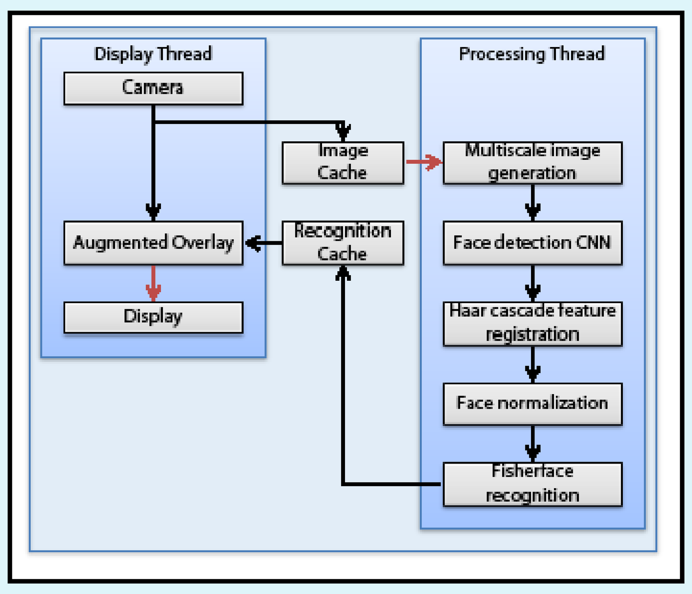
Software Platform

Software Functions
- •Captures real-time video
- •Detects faces, attempts recognition
- •Delivers near-real-time augmented reality

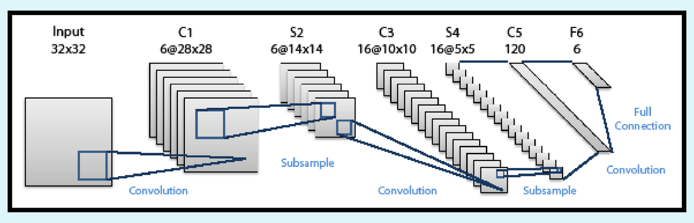
Display Thread
- •**Captures video** at 6FPS to balance augmented reality latency and face recognition delay
- •Overlays **latest face info** on video frames
- •Displays **augmented video** on HUD

Face Processing Thread
- •**Face detection**: Lenet-5 Convolutional Neural Network trained on new training set
- •Novel **multiscale voting** and overlap removal
- •**Feature registration**: Haar Feature Cascade
- •Novel face **normalization**
- •**Face recognition**: Fisherfaces implementation

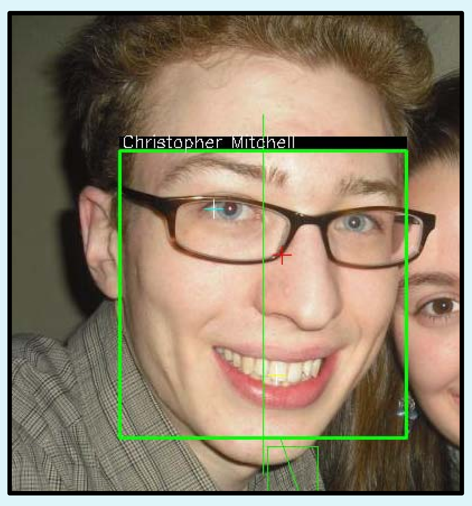
Development and Design
- •CNN Training program written in Lush
- •Testing programs in Lush, C, and Python
- •Final software written in C, using the OpenCV library for matrix manipulation and video functions, and the Intel Performance Primitives (IPP) library for speed optimizations



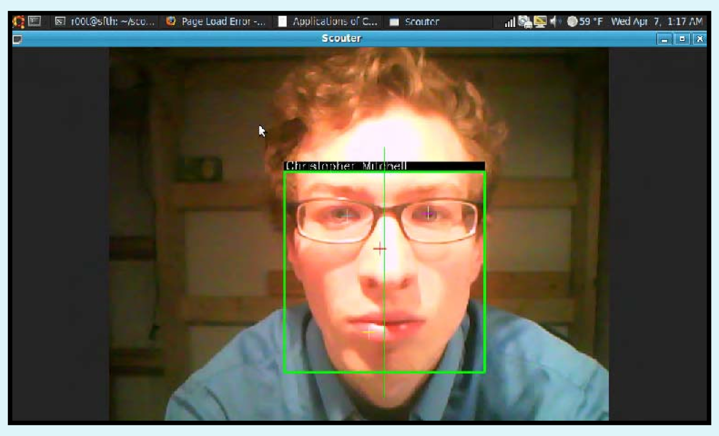
'Scouter' software functionality



Lenet-5 Convolutional Neural Network structure



Successful facial detection, normalization, and recognition on a static image (above) and in real-time (below)



Results and Conclusions

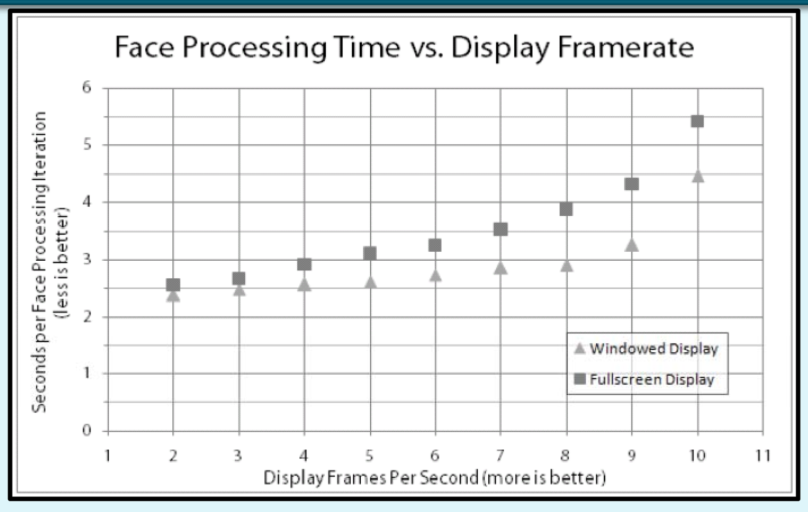
Project Completion
- •Full hardware platform
- •Full face detection and recognition suite

Augmentation &Processing Speed
- •Augmented video feed at 6FPS
- •Maximum 3.25 second latency for face processing

Resource Utilization
- •95% of processing spent running face detection CNN
- •71% total processing spent in C5 layer of the CNN
- •Future work: optimize CNN structure to reduce face processing lag

Face Detection Performance
- •Lenet-5 CNN as trained is robust against: extreme or inconsistent lighting, facial accessories and occlusion, small variations in pitch and yaw, scale and translation.
- •Reliably finds faces with low false positive rate
- •Provides accurate roll estimate on [-25,25] degrees
- •Overlap removal algorithm improves performance

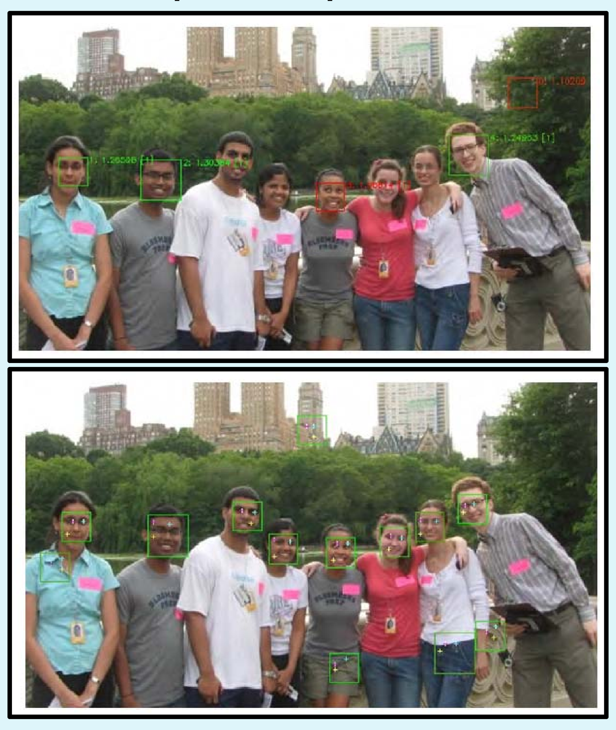
Face Recognition Performance
- •Accurate when feature registration is correct
- •Inaccurate when feature registration dictates incorrect scale or roll corrections
- •Future improvement: tighter integration of CNN pose estimation into face normalizer



Experimental performance balancing

Section	Time (sec)	Pct.	Subsection	Time (sec)	Pct.
Setup	0.0504	1.8%	Multiscale image population	0.0504	1.8%
CNN	2.5917	94.9%	CNN C1/S2	0.2024	7.4%
			CNN C3/S4	0.2589	9.5%
			CNN C5	1.9369	70.9%
			CNN F6	0.0467	1.7%
			CNN RBF	0.1469	5.4%
Overlap Removal	0.00007	0.003%	Multiscale voting	0.00005	0.002%
			Overlap pruning	0.00002	0.001%
Haar Features	0.0693	2.5%	Feature registration	0.0693	2.5 %
Normalization & Recognition	0.0195	0.7%	Face normalization & recognition	0.0195	0.7%

Processing resource utilization by section



Face detection performance before (top) and after parameter tuning (bottom)

(a) Final Overlap Removal Algorithm, Stage 1 (b) Final Overlap Removal Algorithm, Stage 2
(c) Final Overlap Removal Algorithm, Stage 3 (d) Overlap Removal Algorithm Comparison

Face detection performance after overlap removal with multiscale voting (c) and without (d).

CEMETECH

S*ProCom[2] Center for Signal Processing Communications and Computer Engineering

A Research Center at The Cooper Union

150 The Cooper Union Years