Mind-reading machines: Inferring emotions from facial expressions

Peter Robinson

University of Cambridge
Computer Laboratory
http://www.cl.cam.ac.uk/~pr/

Introduction

- Computer Science at Cambridge
- Video user interfaces
- Affective interfaces
  - Mind-reading and mental states
  - Computational model of mind-reading
  - The facial affect inference system
  - Evaluation
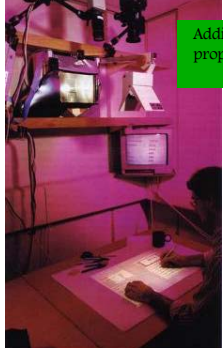 - Applications
- Future work

EDSAC

- Maurice Wilkes (1949)
- Practical computer
  - 650 instructions/s
  - 1k x 17 bits
  - paper tape input
  - teletype output
  - 4m x 3m
  - 3000 valves
  - 12kW

Projectors and cameras

Adding computational properties to everyday objects

DigitalDesk
Pierre Wellner

DoubleDigitalDesk
Steve Freeman
1992

Remote collaboration

Projectors and cameras

Interaction with programs mediated by camera

Animated paper documents
Richard Watts & Dan Sheppard
1997

BrightBoard
Quentin Stafford-Fraser

Projectors and cameras

projector
mirror
mirror
desk
projector

Large display with high resolution fovea

Two-handed interaction

Escritoire
Mark Ashdown
2002

Telepointer traces for collaboration

Projectors and cameras

Arbitrary rotation of documents for territory

High resolution for text

Standard widgets from Java toolkit

Presence for remote collaboration

Distributed Tabletop
Phil Tuddenham
2007

UNIVERSITY OF CAMBRIDGE

Emotional intelligence

- Mind-reading
  - people interp haviour in terms of mental st
  - subtle and el ntial for social interaction
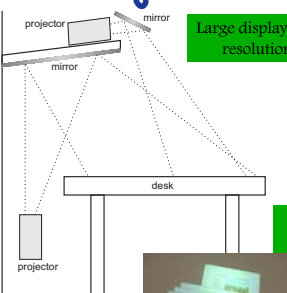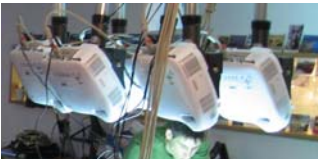- Computers are mindblind
  - most human-computer interaction ignores non-verbal communication cues
  - as a result they fail to:
    - account for what the user knows (and doesn't know)
    - read and respond to the user's intentions
    - understand the reasons behind the user's actions

It looks like you're writing a letter.

Would you like help?

- Get help with writing the letter
- Just type the letter without

☐ Don't show me this tip again

UNIVERSITY OF CAMBRIDGE

The mind in the face

- The face is a spontaneous channel for the communication of social and emotional displays
  - conversation enhancers
  - show empathy
  - acknowledge the actions of other people
- Communicates
  - basic emotions
  - complex (cognitive) mental states

UNIVERSITY OF CAMBRIDGE

Taxonomy of emotions

- Lexicographic analysis (Baron-Cohen, 2003)
  - 1150 words from MS Word thesaurus
  - 412 concepts plus 738 synonyms
  - 24 mutually exclusive groups
- Basic emotions (Ekman, 1971)
  - afraid, angry, disgusted, happy, sad, surprised
  - easily recognised
- Complex (cognitive) mental states
  - bored, bothered, disbelieving, excited, fond, hurt, interested, kind, liked, romantic, sneaky, sorry, sure, thinking, touched, unfriendly, unsure, wanting
  - more important for predicting intentions and actions

UNIVERSITY OF CAMBRIDGE
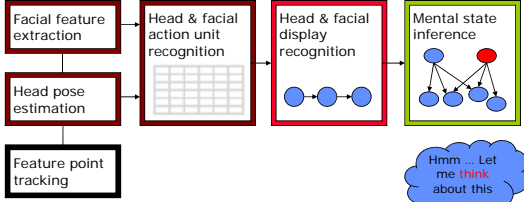
Facial affect analysis

- Infer cognitive mental state groups
  - agreement, concentration, disagreement, interest, thought, unsure
- Characteristics
  - multiple asynchronous information sources
  - multi-level abstractions
    - Action unit: per frame
    - Facial and head displays: 6th of a second
    - Mental state: 2-4 seconds

UNIVERSITY OF CAMBRIDGE

Computational model

Facial feature extraction

Head pose estimation

Feature point tracking

Head & facial action unit recognition

Head & facial display recognition

Mental state inference

Hmm ... Let me think about this

Video input (observed)

UNIVERSITY OF CAMBRIDGE

Feature point tracking

- FaceTracker (Neven Vision, 2002)
- 320×240 video stream at 30 fps
- Tracks 22 points in real-time
  - anchor point interpolated and two further nodes extrapolated
- Self-starting
- Rotation up to 4°/frame and translation to 20 pixels/frame
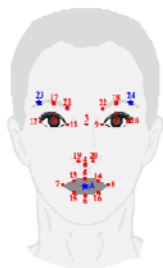
Head pose estimation

- Use expression-invariant feature points to estimate pitch (50°), yaw (50°) and roll (30°)
- Output is head action unit
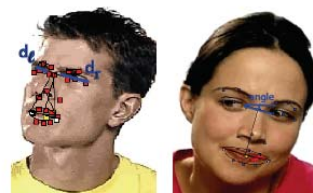
(a) Estimation of head yaw using the ratio of left d_l to right d_r eye widths

(b) Estimation of head roll using the angle between the two inner eye corners

Facial feature extraction

- Colour, shape and motion analysis
- Output is facial action units (+ intensity)

Facial Action	Feature	Description
Lip Pull	Anchor point Polar distance	$\frac{(d1_r + d2_r) - (d1_0 + d2_0)}{(d1_0 + d2_0)} \geq k$
Lip Puckered	Anchor point Polar distance	$\frac{(d1_r + d2_r) - (d1_0 + d2_0)}{(d1_0 + d2_0)} \leq k$

Aperture   Teeth
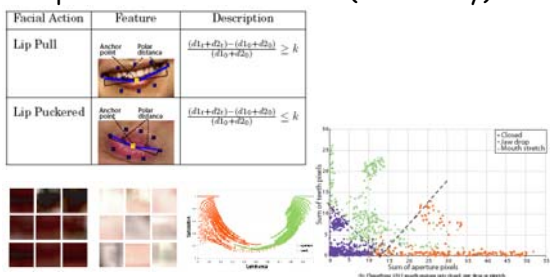
Action units

- Facial Action Coding System (Ekman, 1978)
- Consider 15 of 44 Action Units
  - pitch up & down
  - yaw left & right
  - roll left& right
  - lip pull, pucker, depress & stretch
  - mouth stretch, jaw drop, lips part
  - eyebrow raise inner & outer
- Coded empirically

Head and facial displays

- New classification for additional level in model
- Consider 9 displays (2 periodic & 7 episodic)
  - head nod & shake
  - head tilt & turn
  - lip corner pull & pucker
  - mouth open
  - teeth visible
  - eyebrow flash

Mental states

- Groups of emotion concepts from the Baron-Cohen taxonomy
- Consider 6 groups encompassing 29 concepts
  - Agreement (sure): assertive, committed, convinced, knowing, persuaded, sure
  - Concentration (interested): absorbed, concentrating, vigilant
  - Disagreement (unfriendly): contradictory, disapproving, discouraging, disinclined
  - Interest (interested): asking, curious, fascinated, impressed, interested
  - Thought (thinking): brooding, choosing, fantasizing, judging, thinking, thoughtful
  - Unsure (unsure): baffled, confused, puzzled, undecided, unsure

Combined inference

DBN Level
(model per mental state)

HMM Level
(head+facial displays)

Feature
Extraction
Level

frame 45   64   68   94   99   107   128

Temporal abstraction

The Mindreading DVD

- Comprehensive labelled corpus
  - 412 mental states (or emotions)
  - 6 videos per mental state
  - 24 groups
    - meta-groups
    - fine shades of
      the same mental state
  - posed using 30 actors
  - classified by panel of 10

Mindreading DVD (Jessica Kingsley Publishers), courtesy of the Autism Research Centre,
University of Cambridge www.human-emotions.com

Demonstration

- Classify the video
  - Agreement
  - Concentration
  - Disagreement
  - Interest
  - Thinking
  - Unsure

Source: Cambridge Autism Research Centre

Result

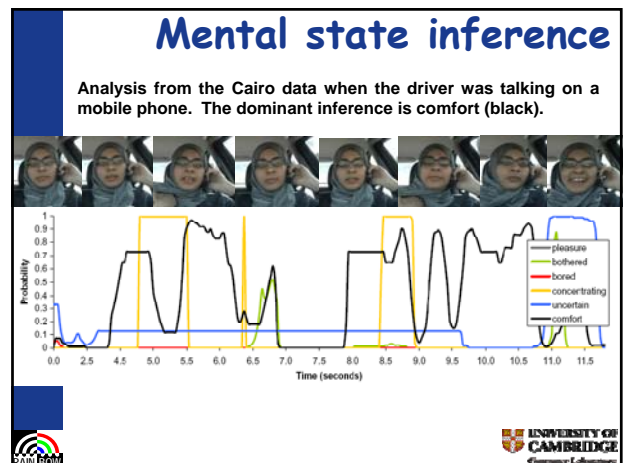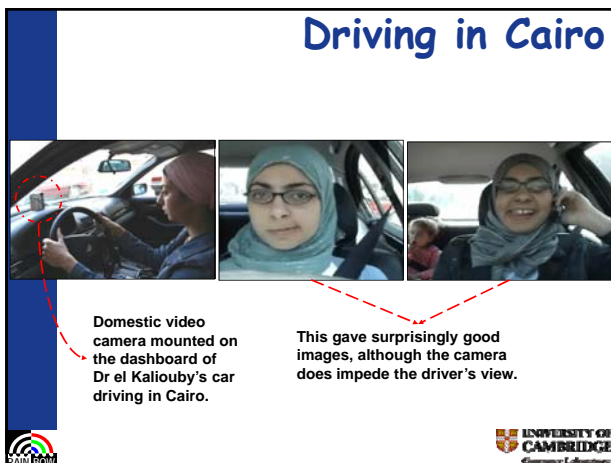- This was a video of *brooding* from the *thinking* group

Training

- Six groups of complex mental states
- 174 videos from Mindreading DVD
  - FaceTracker failed on 10
  - 164 videos featuring 30 actors over 855 seconds
- 164 runs using leave-one-out
  - each actor appears at most once for any concept but possibly more than once in a group
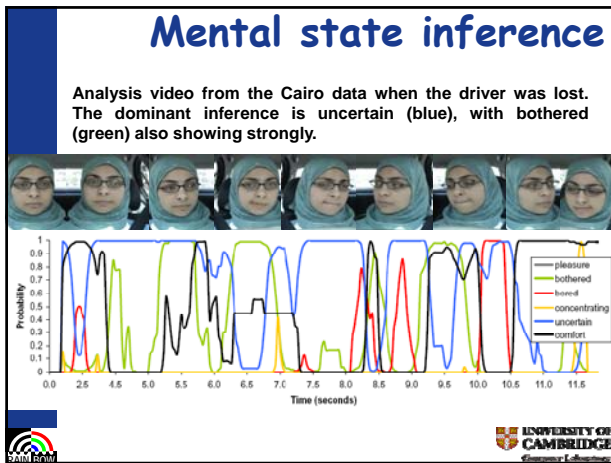- 77% accuracy
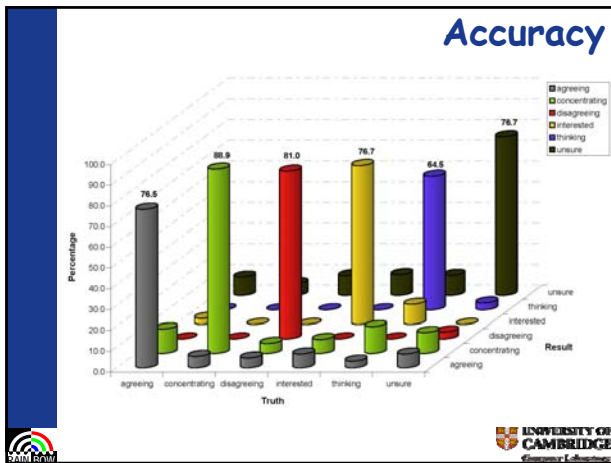
Accuracy

Generalization

- CVPR 2004 corpus
  - 16 conference delegates (*not* actors)
  - each acting six mental state groups
- 96 videos tested on panel of 18 people
  - average accuracy 54%
  - 85% consensus only achieved on 11% of videos
- 88 videos tested on automatic system
  - 3 too short and FaceTracker failed on 5
  - average accuracy 64%
  - 80% accuracy for videos achieving consensus
- System as good as best 6% of panel

Accuracy

Panel of 18 people

Automatic system
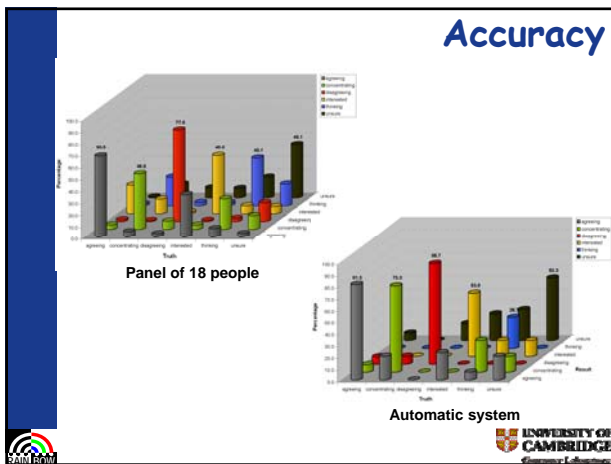
Driver monitoring

- Pleasure
  - enjoyable driving
- Comfort
  - easy driving
- Concentrating
  - neutral
- Bored
  - distracted thinking about something else
- Uncertain
  - distracted finding route
- Bothered
  - discomforted stuck in heavy traffic

Driving in Cairo

Domestic video camera mounted on the dashboard of Dr el Kaliouby's car driving in Cairo.

This gave surprisingly good images, although the camera does impede the driver's view.

Mental state inference

Analysis from the Cairo data when the driver was talking on a mobile phone. The dominant inference is comfort (black).

Mental state inference

Analysis video from the Cairo data when the driver was lost. The dominant inference is uncertain (blue), with bothered (green) also showing strongly.


Simulator at TRL

The UK's Transport Research Laboratory in Crowthorne has several simulators. Video samples were obtained from a camera monitoring the driver in this simulator.

The camera is in a similar location to the Cambridge car and the video also includes the forward view from the simulator. The driver-cam is slightly further away from the driver which makes the view more direct so the tracker coped better.

Mental state inference

**Analysis from the TRL data.
The dominant inference is comfort (black).**
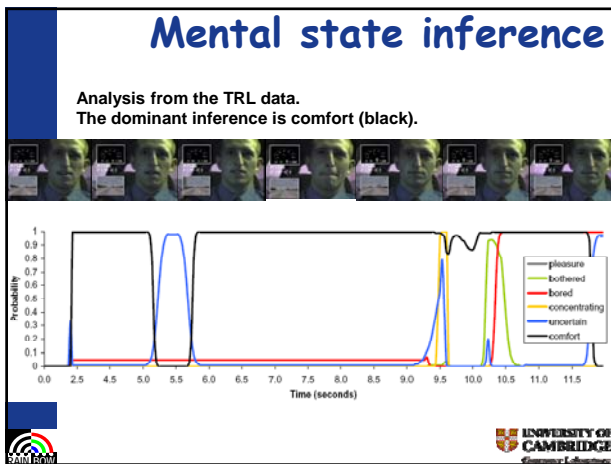
Other applications

- Facial affect instant messaging
  - plug-in for MS Messenger
- Emotional hearing aid
  - assistive tool
- Empathic avatars
  - synthesis as well as analysis
- Usability
- On-line learning
- Sales assistant
- Indexing video

Multi-modal inference

- Face, voice, posture, gesture, physiology

Future work

- Further groups of mental states
- Multi-modal sensors
  - voice, skin conductivity, posture, gesture
- Extended model
  - additional sensors, context
- Generalisation
- Naturally-evoked emotions