



Recognizing people from dynamic video: Dissecting identity with a fusion approach

A. J. O'Toole¹, S. Weimer¹, J. Dunlop¹, R. Barwick¹, J. Ayyad¹, & P. J. Phillips²

¹The University of Texas at Dallas, ²National Institute of Standards and Technology,
Contact: otoole@utdallas.edu



OBJECTIVE

- dissect identity-specific information in natural videos
- information in static versus dynamic
- face versus body

BACKGROUND

- face recognition & motion (Christie & Bruce, 98; Pike et al., 97; Roark et al., 2006)
 - minimal or no benefit of motion for recognition - why?
- distributed network for face processing (Haxby et al., 2000)
 - identity processing (invariant information)
 - lateral fusiform gyrus - fusiform face gyrus (FFA)
 - social interaction (changeable, expression, gaze, etc) - body motion
 - posterior superior temporal sulcus - pSTS
 - (cf., Downing et al., 2001; extrastriate body area)
- psychological and neural distributed network (O'Toole et al., 2002)
 - Role of motion for recognition**
 - dynamic identity signatures - idiosyncratic face/body movement
 - processed in pSTS
 - "backup" recognition system when viewing conditions are poor
 - recognition - primarily from the ventral temporal face areas
 - body and face motions may contribute, but are secondary

APPROACH

- dissect source of identity information in natural videos of people
 - match identity in pairs of videos or images
 - Experiments
 - 1 - video face and body
 - 2 - video face alone
 - 3 - video body alone
 - 4 - static face and body
 - 5 - static face alone
 - 6 - static body alone

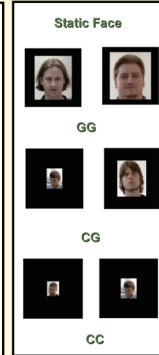
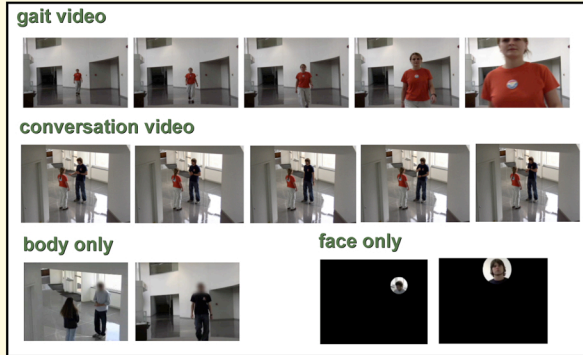
SAMPLE TRIAL



- Response :
 - sure same person
 - think same person
 - don't know
 - think not the same person
 - sure not the same person

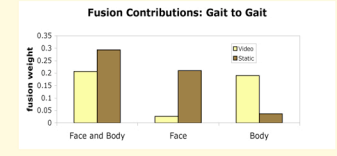
- Face Pairs (n=40)
 - 20 matched identity and 20 unmatched
 - participants (minimum: 30 per experiment)
 - "video" comparison condition
 - GG - gait to gait
 - CG - conversation to gait
 - CC - conversation to conversation

STIMULI & METHODS

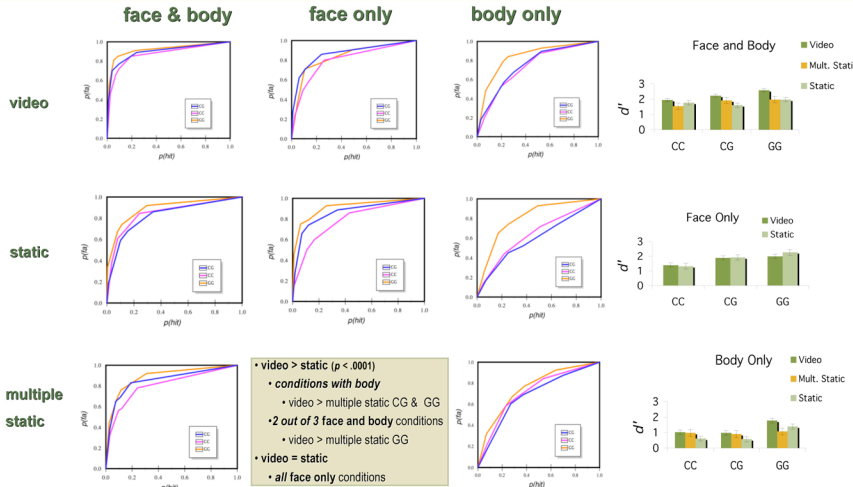


FUSING JUDGMENTS

- Can fusion indicate human use of identity information?
 - Video versus static? Face and body, face only, body only?
- Fusion - combine participant judgments in different conditions
 - algorithm - partial least squares regression (PLS)
 - train predictor network with (n-1) pair judgments (n=40 pairs)
 - predict match status of "left-out" pair - iterate n times
 - tally number of correct predictions
 - PLS yields **weights** for condition predictors
- Results** - 100% correct performance - 3 factor solution



RESULTS



CONCLUSIONS

- Empirical**
 - dynamic > static when body is included
- video advantage**
 - sometimes due to multiple static images
- fusion -> optimal combined information in:**
 - (face and body) and body for **video**
 - (face and body) and face for **static**
- Theoretical**
 - contribution of motion to face for **static**:
 - from body motions
 - role for motion directing attention to face vs body
 - static -> face
 - video -> body

REFERENCES

Christie, F. and V Bruce. 1998. The role of dynamic information in the recognition of unfamiliar faces. *Memory and Cognition* 26:780-790.

Christie, F. and V Bruce. 1998. The role of dynamic information in the recognition of unfamiliar faces. *Memory and Cognition* 26:780-790.

Downing, P. Jiang, Y., Shuman, M. & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, 293, 2402-2403.

Haxby, J. V., E Hoffman, M. I. Gobbini. 2000. The distributed human neural system for face perception. *Trends in Cognitive Sciences* 4:223-233.

O'Toole, A. J., D. Roark, and H. Book. 2002. Recognition of moving faces: A psychological and neural framework. *Trends in Cognitive Sciences* 6:201-206.

Pike G. E. R. I Kemp, N. A. Towell, and K. C. Phillips. 1997. Recognizing moving faces: The relative contribution of motion and perspective view information. *Visual Cognition* 4:409-437.

Acknowledgements - Thanks are due to funding from TSWG/DOD to A. O'Toole for these experiments.