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 and body

BACKGROUND
• face recognition & motion (Christie & Bruce, 1991; Phillips et al., 1997; Ross et al., 2005)
• minimal or no benefit of motion for recognition - why?
• distributed network for face processing (Hadjy et al., 2000)
• identity processing (invariant information)
• lateral fusiform gyrus - fusiform face area (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 - diosyncratic face/face 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

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 (o=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

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

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
    • body and video
    • (face and body) and body for static
  • Theoretical
    • contribution of motion to identification:
      • from body motions
      • role in motion directing attention to face vs body
      • static -> face
      • video -> body

REFERENCES

Acknowledgements. Thanks are due to funding from TSWG/DoD to A. O’Toole for these experiments.