Video Analysis for Sociology

Neva Cherniavsky, Ivan Laptev,
Josef Sivic, Andrew Zisserman

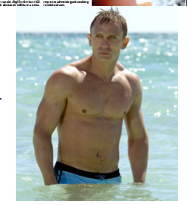
Describing Changes in Human Appearance Over Time

Charlie's Angels: 1976 and 2000

Miami Vice: 1984 and 2006

Dukes of Hazzard: 1979 and 2005

Describing Changes in Human Appearance Over Time



Sociology Research

The Incidence and Context of Tobacco Use in Popular Movies from 1988 to 1997

Madeline A. Dalton, Ph.D.,*,+ Jennifer J. Tickle, B.A.,† James D. Sargent, M.D.,*
Michael L. Beach, M.D., Ph.D.,‡ M. Bridget Ahrens, M.P.H.,* and Todd F. Heatherton, Ph.D.†

- Typical data sets: 250 movies

Preventative Medicine Vol 34, 2002

Sociology Research

The Incidence and Context of Tobacco Use in Popular Movies from 1988 to 1997

Gender Displaying Television Commercials: A Comparative Study of Television Commercials in the 1950s and 1980s[1]

Kenneth Allan[2]
University of North Carolina at Greensboro
Scott Coltrane
University of California, Riverside

- Typical data sets: 250 movies, 617 commercials

Sex Roles Vol 35 Nos 3/4, 1996

Sociology Research

The Incidence and Context of Tobacco Use in Popular Movies from 1988 to 1997

Gender Displaying Television Commercials: A Comparative Study of Television Commercials in the 1950s and 1980s[1]

Kenneth Allan[2]
University of North Carolina at Greensboro
Scott Coltrane
University of California, Riverside

Volume 51, 1990 > Issue 5: September 1990

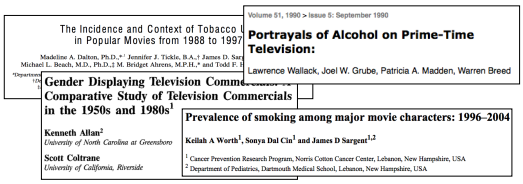
Portrayals of Alcohol on Prime-Time Television:

Lawrence Wallack, Joel W. Grube, Patricia A. Madden, Warren Breed

- Typical data sets: 250 movies, 617 commercials, 195 television episodes

Journal of Alcohol Studies Vol 51 No 5, 1990

Sociology Research



- Typical data sets: 250 movies, 617 commercials, 195 television episodes, 900 movies

Tobacco Control Vol 15, 2006

Sociology Research

- Typical data sets: 250 movies, 617 commercials, 195 television episodes, 900 movies
- Raters (usually students) view video in entirety twice and view each incidence multiple times; usually 10% overlap for inter-rater reliability

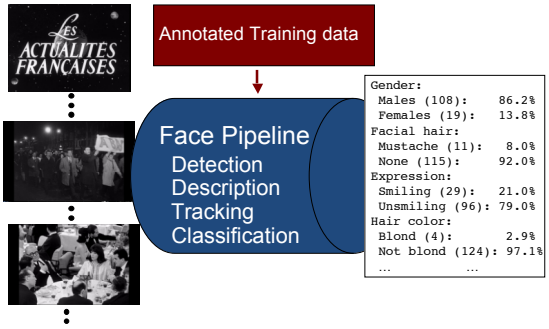
Goal: Video to Statistics

- Automatically find attributes, and number of occurrences, in video data
- Minimize supervision (many different possible attributes)

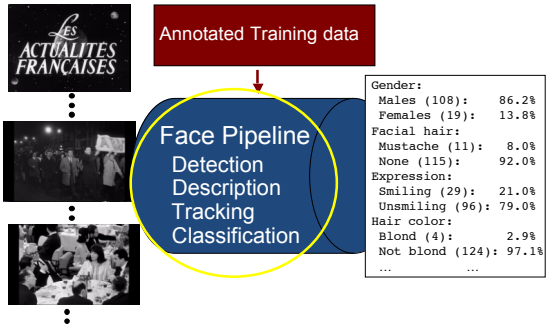
Data

- Hollywood movies from different time periods
  - The Graduate, Roman Holiday, When Harry Met Sally, Love, Actually
- Institut National de l'Audiovisuel
  - R&D: L. Laborelli and D. Teruggi
  - 1.5 Mhours of annotated audiovisual archives, 50 years of TV

Currently: focus on facial attributes

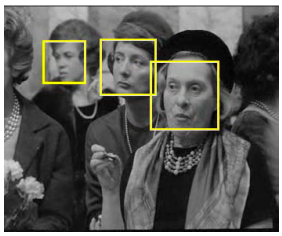


Currently: focus on facial attributes



Face Pipeline: Detection

- Run face detection on each frame (Viola-Jones)



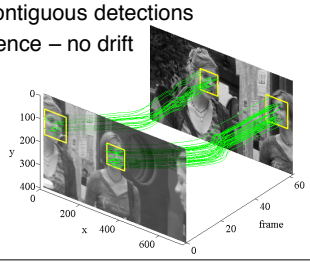
Face Pipeline: Description

- Face representation - local image descriptors at facial feature points
- Extended pictorial structure model

[Everingham, Sivic, Zisserman, 2006]

Face Pipeline: Tracking

- Measure "connectedness" of a pair of faces by point tracks intersecting both
- Doesn't require contiguous detections
- Independent evidence – no drift
- Faces into tracks

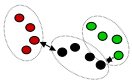


[Everingham et al. 2006]

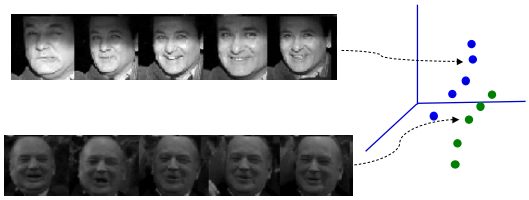
Face Pipeline: Classification

- Classify tracks using SVM
- Distance between tracks is the minimum distance between facial features (not a kernel):

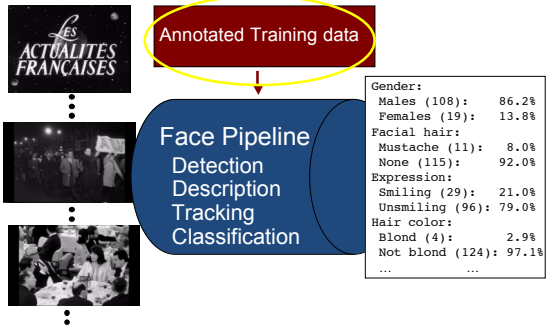
$D(T_i, T_j) = min(d(x,y) \mid x \in T_i, y \in T_j)$



Classification: Matching face sets



Training data



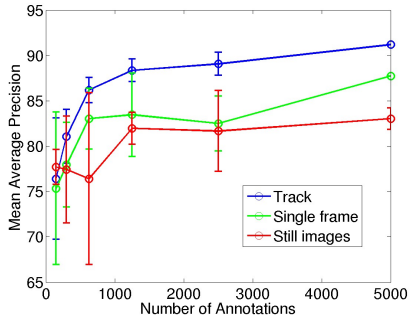
Training data

- Need annotated training data
- Ideally we would train on a large number of attributes with limited supervision
- Looked at two sources: video or still images
- Mechanical Turk (Amazon)
  - Large scale coordination of manual tasks
  - Turks label one frame of the track or a single still image

Training from still images vs video

- Still images:
  - + Variation across people
  - + Potentially labeled data from web for free
  - + Higher quality (resolution, no motion blur)
  - – Not much variation in expression
- Videos:
  - + Variation across viewpoint/expression
  - + Same domain as the testing set
  - – Not much variation in people

Current results: gender



Automatically tagged video

Current work

- Preliminary conclusions: Better to train on videos
- Ongoing work: Study how to combine still images and videos to improve attribute labeling
- More attributes:
  - Race, age, hair color, eye wear
  - Use upper body detection to capture clothing, hairstyles
  - Dynamic attributes: smoking, drinking, smiling
- Video to Statistics
  - Understand where we fail so even when we miss faces, we can report statistics