Distant Seeing TV

Taylor Arnold\textsuperscript{1}  Lauren Tilton\textsuperscript{2}

07 August 2017

\textsuperscript{1}Assistant Professor of Statistics
University of Richmond
@statsmaths

\textsuperscript{2}Assistant Professor of Digital Humanities
University of Richmond
@nolauren
Objectives
Distant Viewing TV applies computational methods to the study of television series, utilizing and developing cutting-edge techniques in computer vision to analyze moving image culture on a large scale.

The project analyzes how visual space is used by characters over a set of fourteen sitcoms from the Network Era of American Television (1952-1985), modeling a new mode of cultural analysis within TV studies. Given that long-running television series broadcast hundreds of episodes, and the major networks run dozens of series each season, previous studies of network television have had to rely on a close analysis of a subset of series, episodes, and scenes.
<table>
<thead>
<tr>
<th>Show</th>
<th>Years Active</th>
<th>Format</th>
<th>Camera Setup</th>
<th>Media Type</th>
<th>Episode Count</th>
<th>Total Hours</th>
<th>Selected Locations</th>
<th>Starring Characters</th>
</tr>
</thead>
<tbody>
<tr>
<td>I Love Lucy</td>
<td>1951-1957</td>
<td>b/w</td>
<td>multi</td>
<td>film</td>
<td>181</td>
<td>72</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Father Knows Best</td>
<td>1954-1960</td>
<td>b/w</td>
<td>single</td>
<td>film</td>
<td>203</td>
<td>88</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>The Andy Griffith Show</td>
<td>1960-1968</td>
<td>b/w</td>
<td>single</td>
<td>film</td>
<td>249</td>
<td>104</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>The Dick Van Dyke Show</td>
<td>1961-1966</td>
<td>b/w</td>
<td>single</td>
<td>film</td>
<td>158</td>
<td>61</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>The Beverly Hillbillies</td>
<td>1962-1971</td>
<td>b/w to color</td>
<td>single</td>
<td>film</td>
<td>274</td>
<td>114</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Bewitched</td>
<td>1964-1972</td>
<td>b/w &amp; color</td>
<td>single</td>
<td>film</td>
<td>254</td>
<td>106</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Gomer Pyle, U.S.M.C.</td>
<td>1964-1969</td>
<td>b/w &amp; color</td>
<td>single</td>
<td>film</td>
<td>150</td>
<td>58</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>The Mary Tyler Moore Show</td>
<td>1970-1977</td>
<td>color</td>
<td>multi</td>
<td>film</td>
<td>168</td>
<td>71</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>All in the Family</td>
<td>1971-1979</td>
<td>color</td>
<td>multi</td>
<td>video</td>
<td>210</td>
<td>89</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Sanford and Son</td>
<td>1972-1977</td>
<td>color</td>
<td>multi</td>
<td>film</td>
<td>136</td>
<td>53</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Happy Days</td>
<td>1974-1984</td>
<td>color</td>
<td>multi</td>
<td>film</td>
<td>255</td>
<td>106</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>The Jeffersons</td>
<td>1975-1985</td>
<td>color</td>
<td>multi</td>
<td>video</td>
<td>253</td>
<td>97</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Laverne &amp; Shirley</td>
<td>1976-1983</td>
<td>color</td>
<td>multi</td>
<td>film</td>
<td>178</td>
<td>68</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Three's Company</td>
<td>1977-1984</td>
<td>color</td>
<td>multi</td>
<td>video</td>
<td>172</td>
<td>72</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td><strong>TOTALS:</strong></td>
<td><strong>2310</strong></td>
<td><strong>947</strong></td>
<td><strong>61</strong></td>
<td><strong>36</strong></td>
<td><strong>41</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Methodology
we want a general approach for studying moving images

computer vision: approximate *seeing* the images by extracting features from the images

build complex features by combining lower level ones
# Feature Taxonomy

<table>
<thead>
<tr>
<th></th>
<th>Text Examples</th>
<th>Moving Image Examples</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outputs</strong></td>
<td>topic models</td>
<td>?</td>
<td>summary plots</td>
</tr>
<tr>
<td></td>
<td>authorship detection</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Abstract Features</strong></td>
<td>sentiment</td>
<td>emotion</td>
<td>application specific, subjective elements</td>
</tr>
<tr>
<td></td>
<td>readability</td>
<td>genre</td>
<td></td>
</tr>
<tr>
<td><strong>Mid-level Features</strong></td>
<td>coreferences</td>
<td>faces</td>
<td>basic building blocks, independent of the medium</td>
</tr>
<tr>
<td></td>
<td>dependencies</td>
<td>objects</td>
<td></td>
</tr>
<tr>
<td></td>
<td>parts of speech</td>
<td>texture</td>
<td></td>
</tr>
<tr>
<td></td>
<td>lexemes</td>
<td>color</td>
<td></td>
</tr>
<tr>
<td><strong>Low-level Features</strong></td>
<td>paragraphs</td>
<td>shapes</td>
<td>medium-specific features</td>
</tr>
<tr>
<td></td>
<td>sentences</td>
<td>edges</td>
<td></td>
</tr>
<tr>
<td></td>
<td>tokens</td>
<td>bounding boxes</td>
<td></td>
</tr>
<tr>
<td><strong>Inputs</strong></td>
<td>letters</td>
<td>pixels</td>
<td>aspects particular to digital storage</td>
</tr>
<tr>
<td></td>
<td>encodings</td>
<td>audio</td>
<td></td>
</tr>
<tr>
<td></td>
<td>png, jpeg, bmp, ...</td>
<td>mp4, MOV, VOB, ...</td>
<td></td>
</tr>
</tbody>
</table>
Implementation

▶ several great libraries for working with **low-level features**:
  ▶ dlib
  ▶ cvv
  ▶ OpenCV

▶ many papers and algorithms that attempt to identify mid-level features, however:
  ▶ code often non-existent or not publicly available
  ▶ prototype ≠ library
  ▶ not generalizable: black & white, low-definition, out of sample
  ▶ no interoperability

▶ abstract features are fairly new territory
Implementation

▶ several great libraries for working with **low-level features**:
  ▶ dlib
  ▶ cvv
  ▶ OpenCV

▶ **many** papers and algorithms that attempt to identify **mid-level** features, however:
  ▶ code often non-existent or not publicly available
  ▶ prototype \(\neq\) library
  ▶ not generalizable: black & white, low-definition, out of sample
  ▶ no interoperability
Implementation

- several great libraries for working with low-level features:
  - dlib
  - cvv
  - OpenCV

- many papers and algorithms that attempt to identify mid-level features, however:
  - code often non-existent or not publicly available
  - prototype \(\neq\) library
  - not generalizable: black & white, low-definition, out of sample
  - no interoperability

- abstract-features are fairly new territory
Face Detection
Face Recognition
Face Recognition
Bewitched, Season 5, Episode 18: Samantha the Bard
Pilot Study: I Dream of Jeannie vs. Bewitched
Character Co-occurrence

Bewitched

Neither Lead
Only Male Lead
Only Female Lead
Both Present

Percentage

I Dream of Jeannie

Neither Lead
Only Male Lead
Only Female Lead
Both Present

Percentage
Next Steps
Next Steps

Phase I

- scene segmentation and classification
- identify minor and one-off characters
- detect laugh track and music

Phase II

- build detailed training set from several different shows in the corpus
- validate methods

Phase III

- run annotators on entire corpus
- exploratory analysis
- document and publish the DTV-toolkit
Thank you!
Team and Advisory Board

Core Team

▶ Nathaniel Ayers, University of Richmond
▶ Annie Berke, Hollins University
▶ Claudia Calhoun, New York University

Advisory Board

▶ Paul Achter, University of Richmond
▶ Joel Burges, University of Rochester
▶ Jeremy Butler, University of Alabama
▶ Lev Manovich, The Graduate Center at CUNY
▶ Trevor Muñoz, University of Maryland
▶ Miriam Posner, UCLA
▶ Joshua Romphf, University of Rochester
▶ Holly Rushmeier, Yale University
▶ Mark Williams, Dartmouth College
Check out the current website: distanttv.org