Distant Seeing TV

Taylor Arnold
tarnold2@richmond.edu
University of Richmond, United States of America

Lauren Tilton
ltilton@richmond.edu
University of Richmond, United States of America

We present work on *Distant Seeing TV*, which applies computational techniques to the study of television series. Our initial set of interest consists of five American sitcoms spanning the 1950's through the 1970's. In order to focus the academic questions that we are interested in exposing, all of the selected shows feature leading women characters and have previously been the topic of academic studies. The poster will outline the kinds of academic questions we hope to answer with our study, the computational methods currently available for answering these questions, our novel extensions of these methods, and some initial results.

The project *Distant Seeing TV* brings the approach of Moretti's "Distant Reading" and Manovich's cultural analytics to the analysis of a large corpus of moving images (Moretti 2013, Manovich 2001). Given that long-running television series broadcast hundreds of episodes and the major networks run dozens of series each season, previous studies, including (Baughman 1993), (Dow 1996), (Spangler 2003), (Morely 2003), have largely had to rely on a close analysis of a small subset of series, episodes, and even scenes. Our project seeks to augment these approaches with computationally-driven analyses that can curate and aggregate the contents of tens of thousands of hours of television programming. To do this, we extract features such as the placement and identities of faces in the shot, as shown in Figure 1, and the time codes for laughter, music, and scene changes as well as the identity of scene locations, as seen in Figure 2. There has been relatively little work done on an aggregate analysis of a corpora of moving images. Noticeable exceptions include *Cinemetrics* (Tsivian and Civjans 2011), which focuses on average shot length in cinema, and the *Arclight Guidebook to Media History and the Digital Humanities* (Acland and Hoyt 2016). Our work extends these to a much wider set of metrics and centers on issues specific to television in contrast to those of film.

Our work attempts to address several types of questions of interest within television studies. These include: Are women characters seen entering a scene more frequently than men? Which characters tend to walk in front of other characters? Does the typical sequence of locations change throughout the run of a show or across different shows? How can we best characterize the narrative arc of an episode? These questions address issues of auteurship, gender, race and narrative in TV.

The initial corpus we are working with includes a diverse set of series with women lead characters that span the 1950s through 1970s. They include a mix of episodes filmed in black and white, in color, with a multiple-camera set-up, with a single camera set up, and from all three major networks:

- I Love Lucy, 1951-1957 (b/w; multiple-camera; CBS)
- The Donna Reed Show, 1958-1966 (b/w; single camera; ABC)
- Bewitched, 1964-1972 (b/w to color; single camera; ABC)
- I Dream of Jeannie, 1965-1970 (b/w to color; single camera; NBC)
- The Mary Tyler Moore Show, 1970-1971 (color; multiple-camera; CBS)

Focusing on a small but diverse set helps in the early stages on this project as we adaptively learn what tools and techniques are most interesting and explore how these methods intersect with current scholarship.

There has been a rapid rate of advancement in modern computer vision techniques over the past decade, which has been particularly driven by the use of deep convolutional neural networks. Given their novelty, there are limited computer vision tools that can be used directly out of the box on a new problem domain. Much of our work on the project has been focused on building a tool set specifically engineered for analyzing television, with a particular focus on parsing black and white images. We will release these tools as an open source package, with wrappers written in python, once it becomes stable enough to be used by other researchers. Techniques that have or will implement include facial detection and recognition (Sun, et al. 2015), scene break detection (Kar 2015; Kumar,

Gupta, and Venkatesh 2015; Pulvar 2015), scene classification (Cheng 2015), dialogue disambiguation (Cervone 2015), speech detection (Sanders, Taubman, and Lee 2015), audience and laugh track detection (Cosentino 2015; Joshi 2016), music segmentation, camera angle detection, and camera selection detection (Nadimi and Bhanu 2004). In all of these cases, training specifically on historical television data has yielded significantly better results that models trained on generic corpora.

The study of television has often been downplayed in favor of textual sources and feature-length film. As described in (Fiske 1978): "Television suffered categorical disadvantage in repute ... its characteristic was oral not literate, whereas 'dominant culture' ... was militantly committed to print-literacy and the values associated with that." From the 1950's onward, however, television has arguably served as the dominant source of mass entertainment in the United States. By 1959, over 83% of households in the US owned their own television set (Baughman 1993). This is not to say that there has been no substantive research in television studies. In fact, there is a large set of prominent examples, including (Baughman 1993), (Dow 1996), (Spangler 2003), (Morely 2003), and many more. Given that long-running television series broadcast hundreds of episodes and the major networks run dozens of series each season, these studies have largely had to rely on a close analysis of a small subset of series, episodes, and even scenes. Our project seeks to augment these approaches with computationally-driven analyses that can curate and aggregate the contents of tens of thousands of hours of television programming.
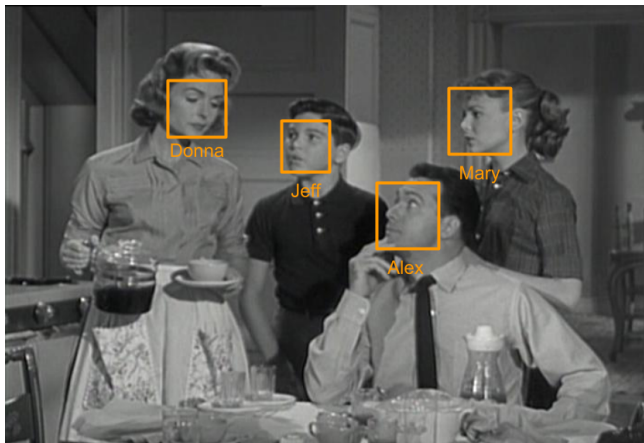

Figure 1. An example of face detection and disambiguation from a still image taken from The Donna Reed Show. Tracking these over the course of a scene and episode

reveal characteristics of how characters interact and describe the narrative flow of the series.

Figure 2. An example of scene classification from Bewitched. Each still image from the episode is tagged with the description of the sets on the sound stage. Following the progress of these over the course of the episode can serve as a way to compactly describe and aggregate the narrative arc of an episode. Comparing across episodes, seasons and shows reveals similarities and differences across the various series of interest.

Bibliography

Acland, C. R. and Hoyt, E. (2016) Editors. *The Arclight Guidebook to Media History and the Digital Humanities.* REFRAME Books.

Baraldi, L., Grana, C., and Cucchiara, R. (2015). "Measuring Scene Detection Performance." *Iberian Conference on Pattern Recognition and Image Analysis.* Springer International Publishing.

Baughman, J. L. (1993) "Television Comes To America, 1947-1957'." *Illinois History* 46.3.

Baughman, J. L. (2006) *The Republic of Mass Culture: Journalism, Filmmaking, and Broadcasting in America Since 1941.* JHU Press, 2006.

Cervone, A., et al. (2015) "Towards automatic detection of reported speech in dialogue using prosodic cues." *Sixteenth Annual Conference of the International Speech Communication Association.*

Cheng, G., et al. (2015) "Auto-encoder-based shared mid-level visual dictionary learning for scene classification using very high resolution remote sensing images." *IET Computer Vision* 9.5: 639-647.

Cosentino, S., et al. (2015) "Automatic discrimination of laughter using distributed sEMG." *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on*. IEEE.

Dow, B. J. (1996) *Prime-Time Feminism: Television, Media Culture, and the Women's Movement Since 1970*. University of Pennsylvania Press.

Fiske, J. (1978). *Reading Television*. Routledge.

Joshi, A., et al. (2016) "Harnessing Sequence Labeling for Sarcasm Detection in Dialogue from TV Series 'Friends'." *CoNLL 2016*: 146.

Kar, T., and Kanungo, P. (2015). "A texture based method for scene change detection." *2015 IEEE Power, Communication and Information Technology Conference (PCITC)*. IEEE.

Kumar, R., Gupta, S., and Venkatesh, K. S. (2015) "Cut scene change detection using spatio temporal video frame." *2015 Third International Conference on Image Information Processing (ICIIP)*. IEEE.

Pulver, A., Chang, M-C., and Lyu, S. (2015) "Shot Segmentation and Grouping for PTZ Camera Videos." *10th Annual Symposium on Information Assurance (ASIA'15)*.

Manovich, L. (2001). *The Language of New Media*. MIT press.

Moore, B., Bensman, M. R., and Van Dyke, J. (2006). *Prime-Time Television: A Concise History*. Greenwood Publishing Group.

Morley, D. (2003). *Television, Audiences and Cultural Studies*. Routledge,

Moretti, F. (2013). *Distant reading*. Verso Books.

Nadimi, S., and Bhanu, B. (2004) "Physical models for moving shadow and object detection in video." *IEEE transactions on pattern analysis and machine intelligence* 26.8 (2004): 1079-1087.

Sanders, J., Taubman, G., and Lee, J. J. (2015). "Background audio identification for speech disambiguation." U.S. Patent No. 9,123,338. 1 Sep. 2015.

Silverstone, R. (1994) *Television and everyday life*. Routledge, 1994.

Spangler, L. C (2003). *Television Women from Lucy to Friends: Fifty Years of Sitcoms and Feminism*. Greenwood Publishing Group, 2003.

Sun, Y., et al. (2015) "Deepid3: Face recognition with very deep neural networks." *arXiv preprint arXiv:1502.00873* (2015).

Tsivian, Y., and Civjans, G. (2011) "Cinemetrics: Movie Measurement and Study Tool Database." (2011).

Xu, P., et al (2001). "Algorithms And System For Segmentation And Structure Analysis In Soccer Video." *ICME*. Vol. 1. 2001.