PJIM PARSONS JOURNAL FOR INFORMATION MAPPING

THE PARSONS INSTITUTE
FOR INFORMATION MAPPING
68 Fifth Avenue
New York, NY 10011
212 229 6825
piim.newschool.edu

Gaining Insight Into Films Via Topic Modeling & Visualization

MISHA RABINOVICH, MFA
YOGESH GIRDHAR, PHD

KEYWORDS *Collaboration, computer vision, cultural analytics, economy of abundance, interactive data visualization*

PROJECT DATE 2014

URL http://misharabinovich.com/soyummy.html

ABSTRACT This paper describes an interdisciplinary collaboration that began with creative misuse of artificial intelligence and computer vision algorithms originally developed at McGill's Centre for Intelligent Machines. We began by analyzing image banks and video with software originally designed for surveillance and robotic anomaly detection. We started with basic visual analysis and had only limited success with movie summarization. We moved beyond misuse when the software actually became useful for film analysis with the addition of audio analysis, subtitle analysis, facial recognition, and topic modeling. Using multiple types of visualizations and a back-and-fourth workflow between people and AI we arrived at an approach for cultural analytics that can be used to review and develop film criticism. Finally, we present ways to apply these techniques to Database Cinema and other aspects of film and video creation.

INTRODUCTION

In the summer of 2013, Misha Rabinovich was an artist in residence at McGill's Centre for Intelligent Machines working primarily with Yogesh Girdhar (then earning his doctorate degree). Our goal was to stage ordering interventions in an deluge of cultural information. We began by creatively misusing robotic navigation software which was originally developed for robotic navigation and anomaly detection. This software worked to summarize data by observing a data stream and saving only those observations which would provide an overview of the whole data set. It measured the amount of surprise

FIGURE 1: *Jay Muhlin Edit*

FIGURE 2: *Sonic Lost World Playthrough*

in an observation (compared to previously stored observations) and made sure that the difference between any given observation and the summary of observations was minimized. The surprise score was constantly calculated by quantifying visual features and qualifying them using Girdhar's online (real-time) extremum summary algorithm effective at solving the k-centers problem (which is better at capturing outliers).[1] The summary was chosen so that if the person saw all the samples in the summary he or she would not be surprised by anything else the robot encountered. In the visual domain the program picks the most common and the most rare images from a video stream in order to maximize the amount of surprise between the summary images. One example of an application is to help photog-raphers dig through a plethora of images during the editing process. An automatic edit could be generated which would give a sense of the whole shoot. FIGURE 1 is a sixteen image edit generated from a massive photo bank created by Jay Muhlin on a cross-country trucking road trip.

VIDEO SUMMARIZATION

A carefully chosen series of image thumbnails can communicate the gist of a video. One application of using automatic visual analysis is to give a general overview of environments encountered in a video game by generating a summary of the game playthrough. FIGURE 2 shows a sixteen image summary of a three hour *Sonic Lost World* (2013) playthrough video created by YouTube user Paraxade0. Another way to summarize video content is to use a video montage which is often how trailers summarize films. We decided to summarize popular movies because their broad viewership made it easier for us to find experimental subjects to verify the success of the summaries. We applied the algorithm to create one-minute summary montages of movies by picking the most surprising and most boring clips in the movie. The goal was to have the software automatically pick only those clips that did the best job of revealing the movie's essence. Ideally, after seeing the summary, no part of the movie would be surprising. This software functioned as a sort of automatic trailer generator with the potential to stand in for or even spoil the movie.

For one experiment using thirteen popular movies we asked twenty-five people to choose the better of two summaries of each movie—one generated with our temporal semantic compression algorithm, and the other with with a pseudo-random "drunken walk" (a new pseudo-random summary for each of the human subjects). It turned out that only about a third of the films

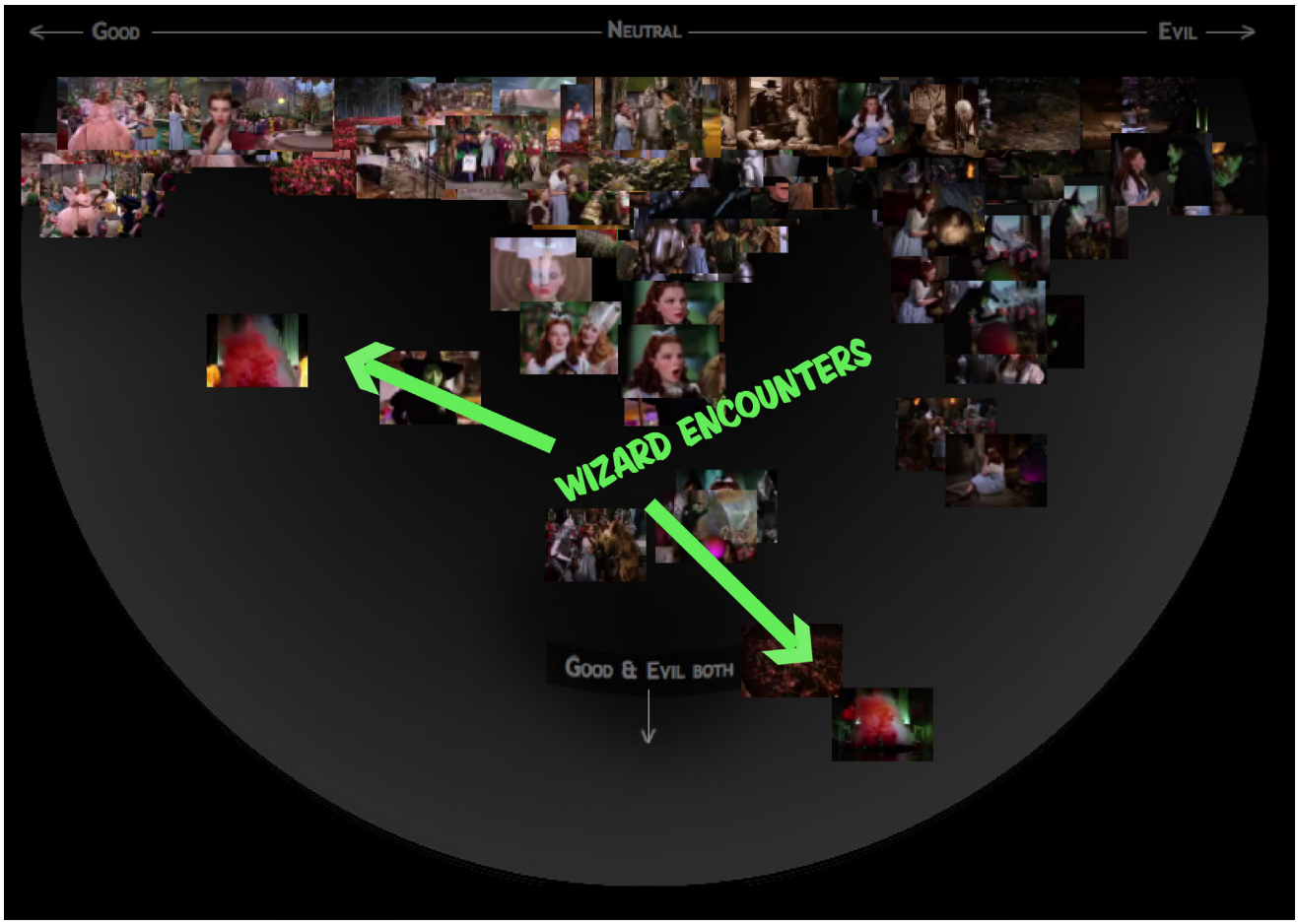

FIGURE 3: *Wizard of Oz Good vs. Evil*

could be better summarized using the AI-generated montage; the AI only worked for films that tightly couple the visual economy with plot development e.g. *Citizen Kane* (1941)[2] and *Fear and Loathing in Las Vegas* (1998).[3]

AUDIO, SUBTITLE, AND FACE ANALYSIS

To generate better summaries we decided to access other modalities beyond the visual. For our next experiments we incorporated analyses of audio using MFCC features.[4] We also employed subtitle analysis and face detection. We focused on the stem words within the subtitles using Lancaster Stemmer, and face extraction using Haar Feature-based Cascade Classifiers mapped onto 140 different face types. These face types were learned from the LFW database using Fisher-faces algorithm. This face database consists of public domain images of celebrities.[5] This gave us access to many types of features, and so we needed a way to intelligently group them together. Instead of only counting discrete features and comparing feature histograms, we used another level of abstraction by including high-level topic modeling using ROST resulting in a sort of semantic sensing.[6] Topics represent high-level themes that are discovered automatically by finding commonly occurring low-level features. The number of topics and their specificity is tunable by input parameters. An example of a topic overlap is Bogart's face appearing in half the topics detected in *Casablanca*, while the few scenes of airplanes were concentrated into just one.[7] We used topic histogram text files to store the strength of each of the topics at regular time intervals of a movie and visualized this information with browser-based interactive visualizations.

We used an iterative process to glean insights into films. We generated a set of twenty topic montages per film, and carefully watched each montage (of duration between a few seconds to a few minutes). We then labeled each montage with the semantic idea it represented. Some of the topics were easy to name, such as "*Wizard of Oz* Song and Dance" (every clip of the characters singing and dancing the theme song) and "Dorothy" (Dorothy's face being the key fea-

ture) while others seemed more abstract, and a few didn't express anything salient at all.[8] One of the most salient topics
combined clips of the good witch, the colorful and abundant Munchkins, and Dorothy being welcomed to Oz. We felt that this topic montage represented Goodness within the schema of the film. Glinda the good witch appeared in other topics but clips of her appeared more often in this particular montage. Another idea was based on the bad witch and lots of dark scenes in the forest (including the kidnapping of the characters), and seemed to encapsulate Evil. Lev Manovich and Nadav Hochman have written about cultural analytics and visualization of photographs based on hue and saturation.[9] Going beyond formal analysis of the image we picked two of the most salient topics and plotted thumbnails from the movie based on how much of the "Good" and "Evil" topics those moments contained (FIGURE 3).

Curiously, scenes of encounters with the Wizard appeared to contain high levels of both good and evil topics. This analysis led us to understand the Wizard as a liminal character who at first appears to be terrifying (and inciting our heroes to murder), but, when the curtain is pulled back, he is revealed to be a bumbling impostor who nonetheless tries to help. Salmon Rushdie's critique of the *Wizard of Oz* points out that the women characters have all the power in this picture.[10] The female witches are strong oppositional characters but the Wizard is a two-faced flip-flopper. This kind of existing film criticism gave us a ground truth against which to test the performance of the visualizations. We also used metadata such as plot keywords from the *Internet Movie Database* (IMDB) to understand the film's conceptual schema. A fruitful dichotomy was found in *Casablanca* (1942) between the theme of imprisoned independence (drinking, Ilsa, singing Germans) vs. collaboration, hope, and freedom (letters of transit, Ricky). These topics are anchored more by specific faces, sound frequencies, and key words than dramatic shifts of the visual domain (since most scenes are claustrophobically set at Rick's club). Kendall R. Phillips, a critic of horror films, discussed the bird motif running throughout the film *Psycho*.[11, 12] One of the montages generated by the software condenses bird references to create a foreboding montage foreshadowing the machinations of the main character's psychosis. The Norman Bates character says that he loves stuffing birds and that taxidermy is more than a hobby for him. The montage immediately cuts to him telling the Marion Crane character (another bird reference) that she eats like a bird. The insinuation of the impending violence and the bird references were automatically distilled into a menacing two-minute bird topic montage

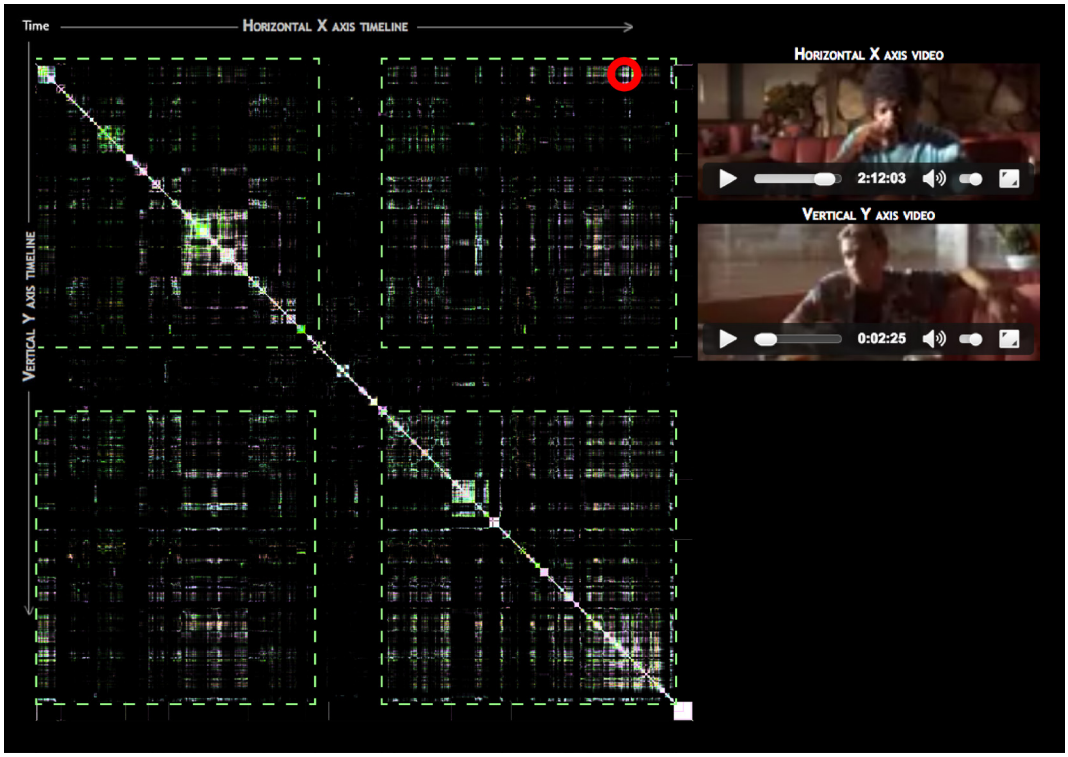

FIGURE 4:
Pulp Fiction Similarity Browser

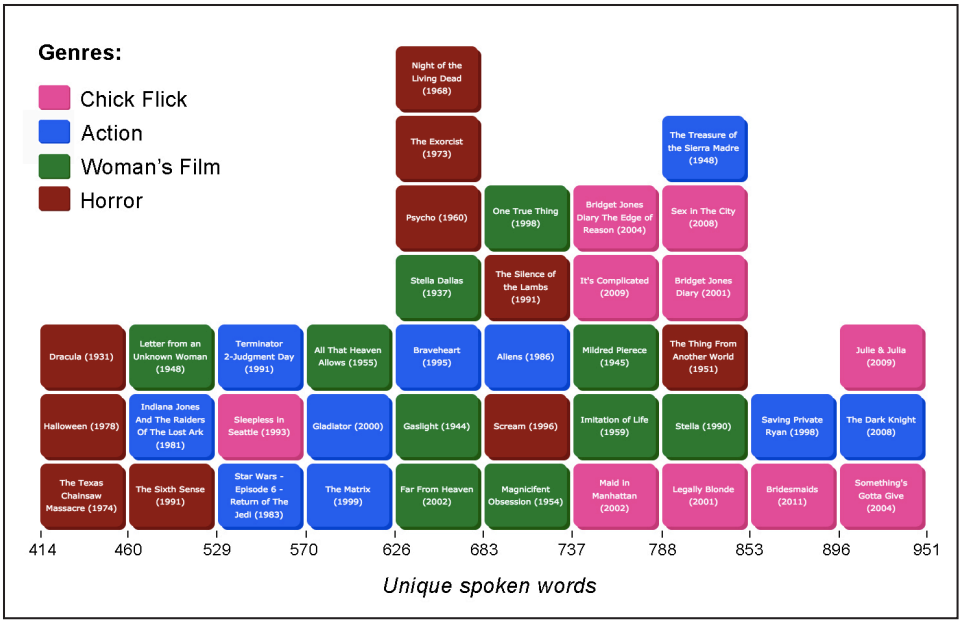

FIGURE 5: *Silence of the Lambs Similarity*

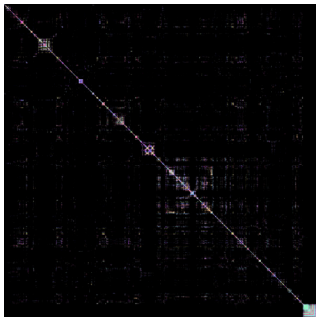

FIGURE 6: *Genre Vocab Plot*

which could either spoil the film—Spoiler Alert!—or help a critic catch the connections faster than watching the film multiple times. This bird topic montage is a video syllogism that says, "Norman kills and preserves birds. Women are like birds. Therefore Norman kills and stuffs women."

A useful and fun interactive summary visualization is our similarity browser. Biologist and educator, Gabriel A. Harp has suggested that working metaphors of cinema can be applied to biology and vice versa.[13] Similarity graphs are used to find patterns in DNA sequences. Patterns hiding in the data are easier to spot because of the symmetry of the graph. Our similarity graph compares the movie to itself. Similarity between moments was calculated by contrasting the strengths of each of the twenty topics at each moment to find the differences between every moment and every other moment in the movie. The pixel brightness corresponds to the similarity between the two moments and the pixel color shows the dominant topic. *Pulp Fiction* (1994)[14] is a self-reflexive film with a circular narrative, and the similarity graph (FIGURE 4) reveals four big squares of pattens because the film starts and ends at the same restaurant. The four main squares are separated by a black plus sign in the center where non-repeating scenarios occurred (the gold watch flashback, Butch's boxing match and subsequent getaway). We turned the similarity graph into an interactive similarity pattern browser by adding two video players and placing a draggable circle on the graph. The horizontal position of the circle maps to the scrollbar of the first player and the vertical position maps to the scrollbar of the second player. This allows scrubbing through video clips to see and hear their content by moving the circle across the patterns (and vice versa). A similarity graph for *The Silence of the Lambs* (1991)[15] reveals a linear narrative that runs forward in pursuit of new evidence and uncovers new clues without returning to the familiar (FIGURE 5).

GENRE ANALYSIS

Film criticism often focuses on genre-based analysis so we decided to topic model genres. We looked at four genres including Action as defined by IMDB, Horror as defined by Kendall Phillips, and "Chick Flicks" and Woman's Films as defined by Karen Hollinger.[16] In her book *Feminist Film Studies* (2012), Hollinger describes the Woman's Film as a sort of meta genre consisting of films comprised of various core genres such as motherhood melodramas and female-centered action-adventure films. Hollinger investigates whether the contemporary Chick Flick—a genre with a romantic comedy core and incorporating other conventions i.e. the female friendship film—grew out of the Woman's Film, or if it is a distinct genre of its own. One key difference is that the Woman's Films focus on the loss and pain experienced by women under patriarchy, while the Chick Flick either celebrates female triumph over patriarchy or takes female equality for granted and presents a shallow female desire linked to sex and shopping. Despite overlap of the two genres, Hollinger nevertheless derives two

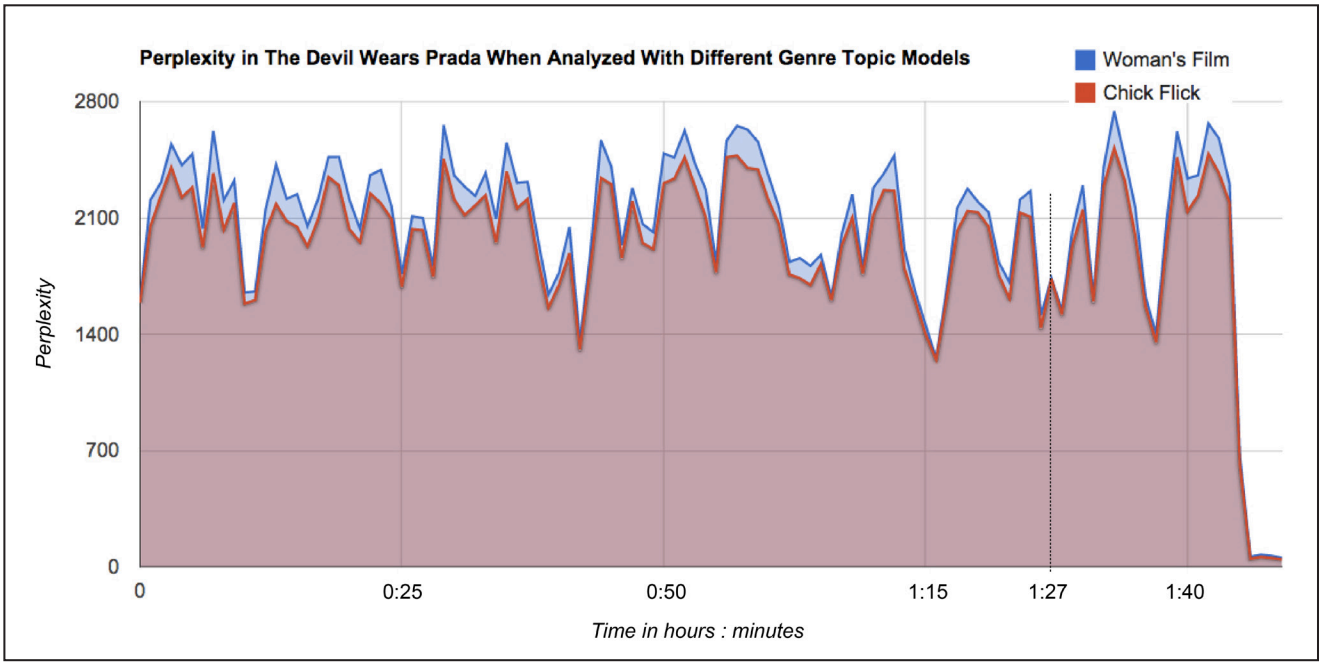

FIGURE 7: *Perplexity Using Different Genre Models*

distinct lists of film titles for each of them. We counted the amount of unique stem words that were spoken in forty films using ten films per genre. We then visualized this information in FIGURE 6 using a unique color for each genre. What this graphic shows is that Chick Flicks tend to have higher vocabularies, while Woman's Films tend toward more mid-range counts which are not too dissimilar from Action movies. Chick Flick are typically very dialogue heavy and hence have a higher chance of having a wide-ranging vocabulary of spoken words. It seems that Woman's Films employ a wider range of techniques to activate their audience, suggesting a more complex and deeper genre. Horror films rely on music and pacing to achieve their affect and hence have the smallest vocabularies.

Hollinger analyzed the pop movie The Devil Wears Prada to see if it is a mere formulaic Chick Flick or if it goes beyond and qualifies as a Woman's Film. She finds that the movie follows a Chick Flick formula but goes beyond (e.g. the female protagonist is young but intelligent and the elder career woman antagonist is not altogether evil). We wanted to see if our software would confirm her analyses. Up until this point we worked with topic models generated for each movie. For genre analysis we generated topic models for whole genres by merging the topic models of the 10 constituent movies into a single model of 200 topics per genre. We analyzed *The Devil Wears Prada* (2006) using the topic model for Chick Flicks and Woman's Films. The software generated perplexity readouts at each minute of the movie which indicated how perplexed or confused the software was when looking at the movie while using each genre model. The less perplexity experienced by the software while using a certain genre model, the better that genre was at describing the movie in question. FIGURE 7 shows that the software was more surprised by the movie when using the Woman's Film topic model suggesting that we are dealing with a Chick Flick. At one hour and twenty-seven minutes, the software thought of the film as more of a Woman's Film and this happens to be where the female protagonist defends the female antagonist against accusations by a male character that the woman is a sadist.

CONCLUSION

Cultural analytics tools can serve as mind augmentation prostheses for gaining insights into film and video faster. Algorithmic genre analysis can be used to understand how a film might be received by the audience and by critics and to affect this reception by making a film or video fit into, or confound, genre-based expectations. Beyond helping photographers and producers dig for diamonds in the rough of media libraries, this software can help film and video makers to develop different perspectives in their work. We found that software, originally used to help robots surveil and find their way around

an environment, can be a useful tool for gaining insight into film and video when the other modalities beyond the visual are accessed and higher order topic modeling is leveraged. Multi-modal interactive visualizations used in concert are better for understanding the gist of a film than a single temporally compressed summary montage. As opposed to the Netflix-style exhaustive upfront micro-tagging of films by people—as described by Alexis C. Madrigal in an article in *The Atlantic* on January 2, 2014 —an iterative back and fourth approach between people and AI software could save time and be more enjoyable.

Additionally, we are looking forward to using the data generated by the human-machine hybrid process to create story visualizers for Database Cinema. Instead of using 3D animation, a fable could be constituted by stitching together clips of existing video and film. Instead of executing cuts in a film based solely on human-generated tags such as in the provocative piece whiteonwhite:algorithmicnoir (2011) by Rufus Corporation, algorithmic movie transitions could be based on deeper formal and semantic ideas derived through collaborative human/machine introspection. The resulting montages could serve as remix stories of pleasant surprise.

BIOGRAPHY

Misha Rabinovich is an artist and educator working collaboratively with artists and scientists on both the east and wests coasts of North America.

Yogesh Girdhar is a data scientist and the postdoctoral scholar of applied ocean physics & engineering at *Woods Hole Oceanographic Institution.*

NOTES

1 Yogesh Girdhar and Gregory Dudek, "Efficient on-line data summarization using extremum summaries" (paper presented at the annual meeting for the IEEE International Conference on Robotics and Automation Saint Paul, Minnesota, USA May 14-18, 2012)

2 *Citizen Kane*, DVD, directed by Orson Welles (1941; Burbank, CA: Warner Home Video, 2001).

3 *Fear and Loathing in Las Vegas*, DVD, directed by Terry Gilliam (1998; New York, NY: Criterion Collection, 2004)

4 Arnold Kalmbach, Yogesh Girdhar, et al., "Unsupervised Environment Recognition and Modeling using Sound Sensing" (paper presented at the annual meeting for the International Conference on Robotics and Automation Karlsruhe, Germany May 6-10, 2013)

5 Gary B. Huang, Marwan Mattar et al., "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments" (paper presented at the annual meeting for the European Conference on Computer Vision Marseille, France October 12-18, 2008)

6 Yogesh Girdhar, Philippe Giguere, et al., "Autonomous Adaptive Exploration using Realtime Online Spatiotemporal Topic Modeling," *International Journal of Robotics Research* (2013), doi: 10.1177/0278364913507325

7 *Casablanca,* DVD, directed by Michael Curtitz (1942; Burbank, CA: Warner Home Video, 2000)

8 *The Wizard of Oz*, DVD, directed by Victor Fleming (1939; Burbank, CA: Warner Home Video, 1999)

9 Nadav Hochman and Lev Manovich, "Zooming into an Instagram City: Reading the local through social media," First Monday 18 (2013), http://firstmonday.org/ojs/index.php/fm/article/view/4711/3698

10 Salmon Rushdie, *The Wizard of Oz (BFI Film Classics)* (London, UK.: British Film Institute, 2008)

11 Kendall R. Phillips, *Projected Fears: Horror Films and American Culture* (Westport, CT: Praeger, 2005), 77-80.

12 *Psycho*, DVD, directed by Alfred Hitchcock (1960; Universal City, CA: Universal Studios Home Entertainment, 2010)

13 Gabriel A. Harp, "Deconstructing the Genome with Cinema", *Leonardo* 40, No 4, (2009): 376-381.

14 *Pulp Fiction*, DVD, directed by Quentin Tarantino (1994; New York City, NY: Miramax Home Entertainment, 2002)

15 *The Silence of the Lambs*, DVD, directed by Jonathan Demme (1991; New York, NY: Criterion Collection, 1998)

16 Karen Hollinger, *Feminist Film Studies* (New York: Routledge, 2012), 35–66.

17 *The Devil Wears Prada*, DVD, directed by David Frankel (2006; Century City, CA: 20TH Century Fox Home Entertainment, 2006)

18 Oleksandr Kolomiyets, Marie-Francine Moens, "Towards Animated Visualization of Actors and Actions in a Learning Environment" (paper presented at the International Workshop on Evidence Based and User Centered Technology Enhanced Learning Trento, Italy, September 16, 2013)