A Viz of Ice and Fire: Exploring Entertainment Video Using Color and Dialogue

Fred Hohman* Sandeep Soni* Ian Stewart* John Stasko†

Georgia Institute of Technology

ABSTRACT

Films and television shows provide a rich source of cultural data and form an integral part of modern life. However, the video medium remains difficult to analyze at scale effectively and its study has generally attracted limited research attention. We propose a method of summarizing the audio and visual aspects of entertainment videos, through the automatic extraction of dominant colors from video frames and textual categories from dialogue. The colors and dialogue are displayed in a visualization that allows the user to explore the video, highlighting both high-level and low-level patterns from the data. Focusing on the hit television series *Game of Thrones*, we show how our visualization supports the detection of scene changes and plot points, providing a new perspective for both scholars and casual viewers.

Index Terms: H.5.m [Information Interfaces and Presentation (e.g., HCI)]: Miscellaneous;

1 INTRODUCTION

With [cinema] you can say so many things, because you've got time and sequences, dialogue, music... you can express a feeling and a thought that can't be conveyed any other way. – David Lynch

Narrative media, including both cinema and television shows, have made an indelible mark on our culture. The impact of narrative media on everyday life has been widespread and lasting, from iconic characters like Darth Vader from *Star Wars* that have inspired many Halloween costumes, to phrases like *yada yada yada* from *Seinfield* that have become part of the modern lexicon. Even with improved availability of narrative media, visualization tools that help digital humanities scholars to discover patterns are lacking or require substantial manual effort to annotate. An interactive visualization could be used to pose high-level stylistic questions, including how the visual style of TV shows change throughout their lifetime, or low-level plot-based questions, such as when a particularly striking event in a series occurs.

Scholars or film critics interested in analyzing narrative media can benefit from visualizations that support both serendipitous and guided discovery. Furthermore, the representation underlying the visualization should be meaningful to a diverse audience with varying film knowledge. In this paper, we focus on the hit series *Game of Thrones*[1] due to its popularity and cultural impact, and propose:

- **An interpretable and computationally efficient representation of unlabelled long-form video**. This representation uses dominant colors in video frames and frequency of words in the dialogue to represent visual and spoken modalities.
- **An interactive visualization that allows users to explore and discover insights about the video**. We demonstrate with two use cases how this visualization can help uncover narrative patterns such as scene changes and plot points.

2 RELATED WORK

Information visualization has permeated many domains and has proved to facilitate cognitive tasks, including video analysis and summarization [4, 7, 18, 25]. Narrative and entertainment videos like films and television shows, however, have received comparably less research attention with most work focused on the content of the video and not on cinematographic elements like color and dialogue [19]. Our primary focus in this work is to visualize cinematographic patterns in a popular show like *Game of Thrones* based on the color extracted from video frames and textual analysis of dialogue.

Color has primarily been used in prior work to visualize high dimensional data, with different color representations to encode different dimensions or range of a dimension [27]. We, on the other hand, use color extracted from video frames as data to visualize, similar to prior research that used colors in videos to track moving objects in a video [22, 26]. A key aspect to visualizing color as data is to decide on the representation that can lead to an interpretable visualization. We find a representative set of colors from each frame, called dominant colors, by partitioning the image to maximize the difference in colors [12]. We describe in Section 3 why we chose to use color dominance and how it leads to visually meaningful color signatures of videos.

In addition to color, we also visualize patterns from the dialogue text obtained from the video subtitles. We employ Linguistic Inquiry and Word Count (LIWC), a popular psycholinguistic lexicon consisting of words and multi-word expressions divided into language, content, and psychological process categories [20, 21]. Although the LIWC lexicon was created to measure the correlation between writing and health, it has since been adapted to a variety of social science research [8, 29]. We choose LIWC over data-driven techniques such as topic modeling [2] because of the simplicity in using a dictionary based method and the interpretability of the thematic categories. The combination of color and text has been exploited to summarize film content [6], and we extend previous work to visualize multiple dimensions of text categories and detailed color palettes.

Lastly, our work can be considered as an example of artistic information visualization, which Viégas and Wattenburg describe as an emerging form of visualization used by artists to create visually appealing artifacts [30]. They provide the example of the artist Jason Salavon, who visualized the movie *Titanic* by extracting the color from individual frames. When the colors were combined into a mosaic, the visualization revealed frantic scene changes around the climax of the film. Other examples of artistic information visualization use color to explore the visual style of different painters and artistic eras [16], and trace cultural patterns through social media image platforms [13, 17]. Our work belongs to the same category, with the objective of using the dialogue text and the extracted colors from the video frames to explore the aesthetic of TV series such as *Game of Thrones*. Our choice of using *Game of Thrones* is influenced by its profound cultural impact through both the narrative

*{fredhohman, sandeepsoni, istewart6}@gatech.edu
†stasko@cc.gatech.edu

[1] *Game of Thrones* is a popular and critically acclaimed TV series on HBO. The show is based on a series of epic fantasy novels referred by the name *A Song of Ice and Fire*, to which the paper title alludes.

Figure 1: A comparison of the color palettes extracted from one frame of *Game of Thrones* using frequency (left) and dominance (right). In our visualization, we create color palettes based on dominance, to handle frames where the top 10 most frequent colors in an image all have a similar hue. Notice how the color dominance extraction accounts for the single pink tree in the background.

and print media [15]. While there have been diverse studies ranging from estimating the importance of characters in the story by network science methods [1], to tracking public emotion about the show through social media [31] and other online sources [24], we are not aware of any work that jointly analyzes the color and the dialogue for this hit TV series.

3 Data and Methodology

In this section, we describe the data and outline the methodology for pre-processing the color and textual data for the visualization, which includes sampling and slicing video and dialogue data into *time slices*, extracting and sorting dominant colors from video frames, and extracting LIWC text category frequencies from dialogue.

Video Data Our visualization represents the visual component of the video with color and the audio component with text. Since we extract color data from the videos, we ensure each video selected is sufficiently high resolution (1280x720 pixels) in order to have the best color representation possible. For the textual data, we use the dialogue found in the show subtitles that includes both English and text translated to English (e.g., from the fictional Dothraki language). The dialogue includes both spoken words as well as sound effects. While the subtitles leave out information such as which character is speaking at a given time, the extracted dialogue provides a rich signal of plot development in narrative media that may be missing from the video alone.

Sampling and Slicing We sample each video at a fixed rate of 1 frame per second. This sample rate preserves video scene integrity while eliminating redundant information, i.e., nearby frames that are almost identical to the extracted frame. The sampling process converts a video into a large image dataset, where each image is slightly different from the next (when two images exist in the same scene, i.e. similar location and characters), except in the case of scene changes where two consecutive images are completely different.

We then divide each episode's image samples into 60 equal-width *time slices*. For instance, a 60-minute episode would yield 60 1-minute time slices, each with 60 1-second samples. We choose a fixed rate of 60 slices per episode to ensure that the visualization remains consistent across episodes, which preserves the similarity between, for example, a 55-minute and a 65-minute episode. [2] We perform the same slicing procedure with the dialogue data to keep the visual and text data chronologically synchronized.

Dominant Color Extraction A color palette is a finite set of related colors, often used in computer graphics, graphic design, and other creative mediums. One useful method of extracting color palettes is color image quantization, the process of selecting colors to represent the full spectrum of an image [10]. Below we describe our techniques used to prepare the color palette data. We use a technique

[2]Some slices may contain fewer or more frames depending on the length of the time slice: e.g., a slice with 40 seconds would contain 40 frames.

Table 1: The LIWC categories used, with example words.

Categories	Example Words	Categories	Example Words
Anger	distrust, hate, resent	Negative	disaster, idiot, sad
Death	corpse, die, execution	Positive	care, grateful, fun
Family	father, kin, wife	Religion	angel, hell, pious
Home	bedroom, door, furniture	Sexual	breast, lover, prostitute
Humans	adult, baby, society	Swear	ass, bastard, damn

called the modified median cut algorithm whose early version is attributed to Heckbert [12], was later improved by Kruger [14], and has a current state-of-the-art version implemented by Leptonica [3]. The median cut algorithm first successively divides the red, green, and blue (RGB) color space into a set of 3D rectangular boxes, starting with the smallest box that contains all the image pixels and dividing on the median RGB value. However, low frequency colors are often not represented using a median cut. For our work, we use the now popular modified algorithm that weighs the low frequency colors more, to produce a more accurate color palette. In summary, this algorithm takes in an RGB image and outputs a palette of RGB values associated with the image's most representative colors.

An important parameter in the color extraction algorithm is the number of colors, n, to include in a palette. After experimentation, we choose to use $n = 10$ colors in each of our palettes to balance between under-representing the image with few colors (e.g. $n = 2$) or over-representing the image with too many colors (e.g. $n = 50$). An example of this technique applied to a single image can be seen in Figure 1, where the 10 dominant colors from the frame are extracted and shown as a palette on the right side.

We choose to produce color palettes based on a color's dominance in a single frame: how much a particular color perceptually stands out in an image or "dominates" the image. Another natural approach that we explored was to extract color palettes based on a color's frequency in a frame; however, this produces palettes that are not diverse nor meaningful. For example, in Figure 1, extracting colors based on frequency produces a palette consisting only of light blues, which neglects the other salient colors in the image. Notice how color dominance isolates perceptually meaningful palettes and accounts for unique low-frequency colors, such as the single pink tree in the background.

Sorting Colors After extracting a color palette for a given time slice, we sort the colors to provide consistency across all color palettes. However, since color is a multi-dimensional quantity, there is no way to sort colors linearly such that any color will be similar to its neighbors [5, 11]. There are a number of ways to sort colors in practice, such as RGB component-wise sorting or luminosity sorting. However, we sort colors by hue, the base color which is arranged as a rainbow; saturation, the perceived intensity; and value, the perceived brightness (HSV). Sorting by component in HSV space groups the colors by hue, e.g., dark reds to light reds, followed by dark blues to light blues. This produces a pleasing linear color sort at the expense of losing monotonic change from light to dark.

Episode- and Season-level Color Extraction After extracting colors from the time slices within each episode, we aggregate these colors at the episode and season level. All the color palettes within each episode are concatenated into a single image, and the top 10 dominant colors from this image are extracted into a episode-level color palette, as seen in the middle color block in Figure 2. Next, all the episode-level color palettes are concatenated, and the top 10 dominant colors are extracted a season-level color palette, as seen in the top block in Figure 2. Extracting the dominant colors in this fashion may not seem equivalent to concatenating all images in an episode and extracting the palettes from the combined image. However, it saves memory and runtime while providing a clear connection between the different levels.

LIWC Text Extraction The dialogue data is sliced the same way as the color data to yield 60 time slices of dialogue per episode. We first preprocess the data to remove noise, such as removing En-

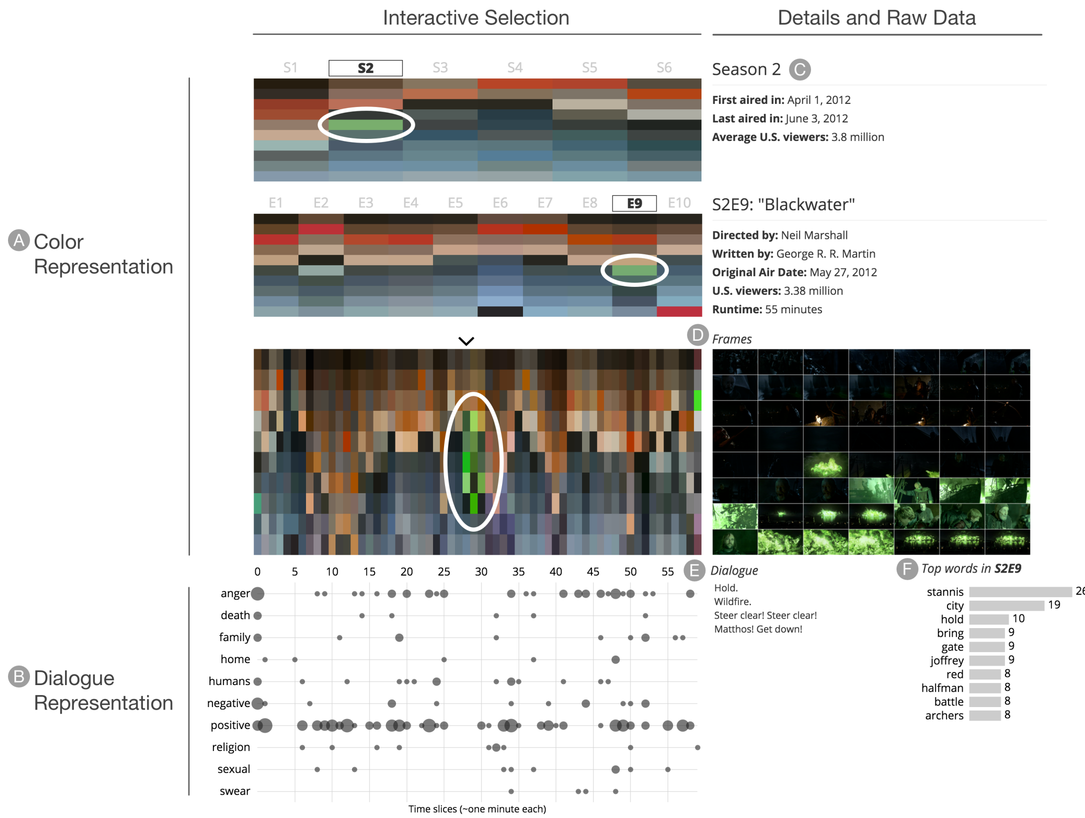

Figure 2: The interactive visualization, displaying the show's Season 2, Episode 9: "Blackwater." The Color Representation plot (A) shows the palettes extracted for the entire series, each season, and each episode as labelled stacked columns, e.g. S2E9 corresponds to Season 2, Episode 9. The dialogue plot (B) shows how much of each textual category is present in the dialogue at a given time. Notice that the bottom color plot in (A) shares the horizontal time axis with the bubbles directly below in (B). Metadata for each season and episode is displayed in (C). The frames view (D) and dialogue view (E) display the video frames and spoken dialogue for the current time slice and update as users interact with the data in (A) and (B). The word frequency histogram (F) displays the top words spoken in the episode.

glish stopwords and lowercasing the text. For each time slice, we extract counts of words from the following 10 LIWC categories: anger, death, family, home, humans, negative, positive, religion, sexual and swears, with examples shown in Table 1. Out of 43 available lexicons[3], we choose the categories that exhibited the most consistent frequencies in *Game of Thrones* determined through preliminary analysis. While not appropriate for all TV shows, the chosen lexicons are well-suited to the show's violent and interpersonal plot [24].

4 VISUALIZATION

To display both the color and textual data, we develop a custom visualization (using HTML, CSS, and JavaScript) that allows a user to observe an overview of an entire entertainment video, while using interaction to provide details on demand. The visualization contains multiple components that are linked together using interaction, and the system's user interface can be seen in Figure 2.

Data Representation The visualization is divided into two rows: *Color Representation* and *Dialogue Representation*. In the Color Representation row (A) three rectangular components display the top 10 dominant colors for each season, episode, and time slice within the selected episode. In the top and middle rectangular components, the labels above each color palette display the selected season or episode, e.g., in Figure 2 *S2E9* corresponds to Season 2, Episode 9. The *Dialogue Representation* row (B) displays the frequencies of the words in the selected LIWC categories across all time slices. If words from one of the categories are spoken in a particular time slice, a transparent circular bubble is placed at the corresponding position, where the radius of the bubble is sized according to the count of unique words in that category that are spoken. The Color Representation and Dialogue Representation row are chronologically linked according to the axis between them (indicated by the slice numbers on the axis), such that the color palette and the text bubbles in the same vertical column represent data from the same time slice.

In addition to the *Interactive Selection* area on the left, the visualization displays *Details and Raw Data* from the current selected episode on the right: metadata (C), the video frames from the selected time slice (D), the dialogue spoken during the selected time slice (E), and the top 10 words spoken during the selected episode, including non-LIWC words such as character names (F).

In summary, the visualization reads top to bottom, displaying color on the top and the dialogue on the bottom, and left to right, displaying the aggregate data on the left and details on the right.

Interactions The Color Representation area supports several user interactions. The user can click on a season or episode block to update the visualization's adjacent views, including the episode-level colors, text data, and metadata. In addition, the user can click on a color slice to enlarge it for closer inspection. Our visualization also provides details on demand: hovering over a color slice will update the image frames from the given slice displayed on the right. This allows the user to connect the dominant colors in a given time slice with that slice's actual video frames.

Similarly, the Dialogue Representation area responds to user interaction: hovering over a text bubble reveals all raw dialogue spoken during the associated time slice in the *Dialogue* area. Similar

[3]Obtained through official LIWC source at `http://liwc.wpengine.com/`. Accessed March 14, 2017.

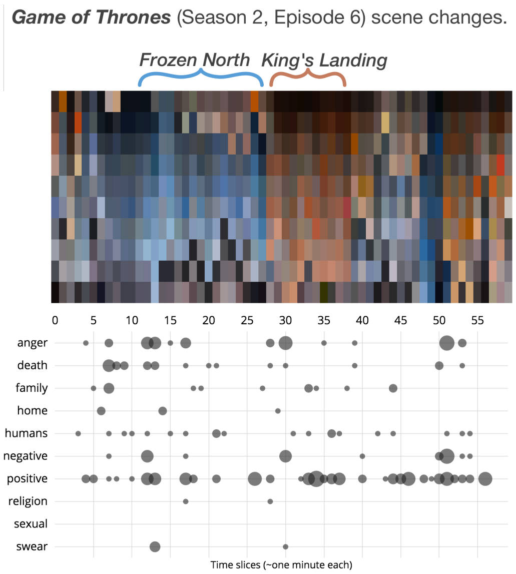

Figure 3: A scene change in *Game of Thrones* Season 2, Episode 6. Note the contrast between the blue-dominant Frozen North scene in the beginning of the episode (time slices 10-28) and the brown-dominant King's Landing scene in the middle (time slices 28-36).

to the frames provided for the color slices, the raw dialogue helps the user to interpret the text categories. In addition, the text plot interactions are synchronized with the color slice chart. Both plots update simultaneously when the user clicks a new episode or season, and hovering over a color slice also highlights the text bubbles in that slice to differentiate nearby bubbles. The time slice is marked by an arrow above the color slice plot. This allows the user to keep track of the color and text data in parallel.

5 Use Cases

Our visualization provides both high-level and low-level insights, and we explore two use cases that illustrate these capabilities: detection of scene changes and revealing plot points.

Use Case I: Detecting Scene Changes *Game of Thrones* is a complex narrative involving multiple storylines, which result in frequent scene changes. A film critic may wish to understand the show's storytelling techniques by studying these changes. Previous work has shown that color can help segment scenes automatically [28] without considering dialogue. Since specific scenes are often defined by their background colors and dialogue patterns, we show how both colors and LIWC categories can signal that a scene has changed. Two such scene changes happen in *Game of Thrones* Season 2, Episode 6, shown in Figure 3: the change from the blue-dominant scene to the brown-dominant scene at time slice 28, and the change from the browns back to the blues at time slice 46. These color changes signal a switch from the blue ice and snow of the Frozen North to the brown buildings of King's Landing, which are key locations for the plot that correspond to two separate main characters. The difference between the scenes is underscored by the dialogue plot, where the switch to browns coincides with a burst of anger and negative words, followed by a stream of positive words. A film critic may further explore whether the same blue-to-brown scene change occurs consistently in other episodes in the same season, to explore the interaction between storylines. Detecting scene changes can also differentiate episodes that remain in one location from episodes that jump across multiple locations, and each of these styles of episode may contribute differently to the overall storytelling.

For a show such as *Game of Thrones* with highly differentiated environments, the combination of color and dialogue visualization can highlight scene changes.

Use Case II: Revealing Plot Points Our visualization also helps to reveal a story's key plot points: significant events that can lead to unforeseen consequences [9]. A film critic may judge the storytelling of an episode based on its execution of a particular plot point: what is the signature of the plot point, and when did it occur? In *Game of Thrones*, a plot point occurs in the battle of Blackwater (Season 2, Episode 9) with an explosion of green wildfire that the Lannisters (a ruling family under siege) use to halt the advance of Stannis Baratheon (enemy to the Lannisters). This explosion shocks the characters and tilts the battle in favor of the Lannisters. The scene is iconic[4] for its visual depiction of the explosion, whose bright green color even a non-fan would identify as interesting.

Although the explosion happens quickly, it can be recognized at the highest level of the visualization in the season-level blocks at the top of Figure 2. Not knowing where the green comes from, a curious viewer may follow the green by first clicking on the Season 2 block, then the Episode 9 block. The viewer could then notice the bright green dominating the middle of the episode, from time slice 28 to 30, and note the explosion happening in the video frames. Looking at the associated dialogue in the Dialogue Representation, the viewer may also note the silence that occurs during the explosion. Although surprising in the context of this dialogue-packed episode, the lack of dialogue here makes sense: the suddenness of the explosion catches characters off-guard, and the episode's director may intend to leverage the spectacular explosion imagery to have maximum standalone impact on the viewer. The film critic may look for similar periods of sparse dialogue, or imbalanced dialogue (e.g. excessive negative language), in other episodes to help pick out plot points that may lack corresponding distinctive colors. Both the striking color and the unusual dialogue pattern mark the wildfire explosion as a plot point that stands out even at the season-level representation.

6 Conclusion

Our work lays the foundation for future studies to improve the representation and evaluation of narrative media visualization.

First, representing video frames with their dominant colors and LIWC words only captures surface-level trends and could miss the underlying semantics of a scene. For instance, it is difficult to identify scenes involving a specific character, such as *Jon Snow*, using solely colors. More advanced image and text processing techniques, such as facial recognition and topic modeling, could provide better context for the data visualized. Second, future work should address user evaluation, because we have provided only qualitative results about the utility of our method for the use cases we described. It is difficult to evaluate quantitatively the value of the proposed visualization tool, because of its focus on exploratory analysis. One way to approach evaluation is to elicit user feedback such as self-reported enjoyment or satisfaction ratings, similar to other causal visualization systems [23], to assess how this system contributed to their overall viewing experience.

By highlighting scene changes and plot points, we show how the visualization can allow a viewer, such as a film critic, to explore the storytelling techniques of narrative media. Digital humanities research about films would greatly benefit from such storytelling analysis done at scale without being limited by expensive manual annotation effort. Automated analysis can help scholars make a preliminary analysis of new media such as *Game of Thrones* without waiting for annotations.

Acknowledgments

The authors thank the members of the GT Visualization Lab and the anonymous reviewers for their constructive feedback. We do not own any of the images or text data associated with *Game of Thrones*. All rights belong to HBO.

[4] `https://www.wired.com/2012/05/game-of-thrones-blackwater/`. Accessed March 14, 2017.

References

[1] A. Beveridge and J. Shan. Network of thrones. *Math Horizons*, 23(4):18–22, 2016.

[2] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan):993–1022, 2003.

[3] D. Bloomberg. Color quantization using modified median cut. *Leptonica*, 2008.

[4] R. Borgo, M. Chen, B. Daubney, E. Grundy, G. Heidemann, B. Höferlin, M. Höferlin, H. Leitte, D. Weiskopf, and X. Xie. State of the art report on video-based graphics and video visualization. In *Computer Graphics Forum*, vol. 31, pp. 2450–2477. Wiley Online Library, 2012.

[5] C. A. Brewer. Guidelines for use of the perceptual dimensions of color for mapping and visualization. In *IS&T/SPIE 1994 International Symposium on Electronic Imaging: Science and Technology*, pp. 54–63. International Society for Optics and Photonics, 1994.

[6] M. Burghardt, M. Kao, and C. Wolff. Beyond shot lengths–using language data and color information as additional parameters for quantitative movie analysis. In *Digital Humanities*, 2016.

[7] M. Chen, R. Botchen, R. Hashim, D. Weiskopf, T. Ertl, and I. Thornton. Visual signatures in video visualization. *IEEE Transactions on Visualization and Computer Graphics*, 12(5), 2006.

[8] M. De Choudhury, S. Counts, and E. Horvitz. Predicting postpartum changes in emotion and behavior via social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3267–3276. ACM, 2013.

[9] S. Field. *Screenplay: The foundations of screenwriting*. Delta, New York, 2007.

[10] S. Fredrick. Quantization of color images using the modified median cut algorithm. *Virginia Polytechnic Institute and State University*, 1992.

[11] T. Gray. A Visit to Disney's Magic Kingdom. *Wolfram Alpha Blog*, 2013.

[12] P. Heckbert. Color image quantization for frame buffer display. *SIGGRAPH Computing Graphics*, 16(3):297–307, July 1982.

[13] N. Hochman and R. Schwartz. Visualizing Instagram: Tracing Cultural Visual Rhythms. In *Proceedings of ICWSM*, pp. 6–9, 2012.

[14] A. Kruger. Median-cut color quantization. *Dr. Dobbs Journal*, pp. 46–54, 91–92, 1994.

[15] J. Lanchester. When did you get hooked? *London Review of Books*, 35(7):20–22, 2013.

[16] S. Lynch, J. Haber, and S. Carpendale. Colourvis: Exploring colour in digital images. *Computers & Graphics*, 36(6):696–707, 2012.

[17] L. Manovich and J. Chow. Taipei Phototime. *Software Studies Initiative*, 2014.

[18] A. H. Meghdadi and P. Irani. Interactive exploration of surveillance video through action shot summarization and trajectory visualization. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2119–2128, 2013.

[19] A. Nandakumar and J. Murray. Companion apps for long arc TV series: supporting new viewers in complex storyworlds with tightly synchronized context-sensitive annotations. In *Proceedings of the 2014 ACM international conference on Interactive experiences for TV and online video*, pp. 3–10. ACM, 2014.

[20] J. W. Pennebaker, M. E. Francis, and R. J. Booth. Linguistic inquiry and word count: LIWC 2001. 71, 2001.

[21] J. W. Pennebaker, M. R. Mehl, and K. G. Niederhoffer. Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54(1):547–577, 2003.

[22] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet. Color-based probabilistic tracking. In *European Conference on Computer Vision*, pp. 661–675. Springer, 2002.

[23] Z. Pousman, J. Stasko, and M. Mateas. Casual information visualization: Depictions of data in everyday life. *IEEE Transactions on Visualization and Computer Graphics*, 13(6), 2007.

[24] A. Scharl, A. Hubmann-Haidvogel, A. Jones, D. Fischl, R. Kamolov, A. Weichselbraun, and W. Rafelsberger. Analyzing the public discourse on works of fiction–Detection and visualization of emotion in online coverage about HBOs Game of Thrones. *Information Processing & Management*, 52(1):129–138, 2016.

[25] K. Schoeffmann, M. Taschwer, and L. Boeszoermenyi. The video explorer: a tool for navigation and searching within a single video based on fast content analysis. In *Proceedings of the first annual ACM SIGMM conference on Multimedia systems*, pp. 247–258. ACM, 2010.

[26] Y. Seo, S. Choi, H. Kim, and K.-S. Hong. Where are the ball and players? Soccer game analysis with color-based tracking and image mosaick. In *International Conference on Image Analysis and Processing*, pp. 196–203. Springer, 1997.

[27] S. Silva, B. S. Santos, and J. Madeira. Using color in visualization: A survey. *Computers & Graphics*, 35(2):320–333, 2011.

[28] B. T. Truong, S. Venkatesh, and C. Dorai. Scene extraction in motion pictures. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(1):5–15, 2003.

[29] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe. Predicting elections with Twitter: What 140 characters reveal about political sentiment. *ICWSM*, 10(1):178–185, 2010.

[30] F. Viégas and M. Wattenberg. Artistic data visualization: Beyond visual analytics. *Online Communities and Social Computing*, pp. 182–191, 2007.

[31] K. Wongsuphasawat. How every #gameofthrones episode has been discussed on twitter. https://interactive.twitter.com/game-of-thrones/, 2016. Accessed 2017-03-14.