What Happens in Films?

Andrew Salway, Andrew Vassiliou and Khurshid Ahmad
Department of Computing, University of Surrey, Guildford, United Kingdom
{a.salway, a.vassiliou, k.ahmad}@surrey.ac.uk

Abstract

This paper aims to contribute to the analysis and description of semantic video content by investigating what actions are important in films. We apply a corpus analysis method to identify frequently occurring phrases in texts that describe films - screenplays and audio description. Frequent words and statistically significant collocations of these words are identified in screenplays of 75 films and in audio description of 45 films. Phrases such as 'looks at', 'turns to', 'smiles at' and various collocations of 'door' were found to be common. We argue that these phrases occur frequently because they describe actions that are important story-telling elements for filmed narrative. We discuss how this knowledge helps the development of systems to structure semantic video content.

1. Introduction

Intuitive and innovative forms of video retrieval, browsing and reuse require video data to be made into a structured medium by analysis and description of its content [1]. In the case of feature films, the semantic content is particularly rich. To appreciate a film, a viewer must understand narrative elements like what is depicted, cause-effect relationships, emotional reactions and ideas about the filmmaker's intention. The content is communicated to the viewer by patterns of light and shade, dialogue, sound effects, film editing techniques, music, and the actions of characters. Progress has been made in mapping from audio-visual features to higher-level structures. However, 'mid-level' semantic content has been ignored, i.e. what the characters are doing. On the one hand this is a very hard problem. The sets of entities, actions and events that may be depicted in a film are virtually limitless. But, we wonder, are there any actions that are common in films? If so, then efforts to structure semantic film content could be directed at these actions which could be considered as building blocks of filmed narrative. Descriptions of basic actions might provide a stepping-stone between audio-visual features and high-level narrative structures.

Section 2 reviews research into the analysis and structuring of film content. In order to discover common actions in films, our approach is to identify frequently occurring phrases referring to actions in texts that describe films. Section 3 presents results from the analysis of the screenplays of 75 films and the audio description of 45 films. Frequently occurring phrases and their associated actions are identified and discussed. Section 4 considers how knowledge about common actions might provide a basis for the description of high-level film content.

2. Analyzing and structuring film content

An important step in structuring semantic film content is the segmentation of shots and scenes. This can be achieved by fusing audio and visual boundary detectors [2]; earlier work proposed a data model for navigating films based on shots, scenes and editing effects [3]. Pixel motion and shot length have been extracted and combined, in order to compute a film's changing tempo. Changes in tempo were found to coincide with key points in a film's story [4]. The presence and absence of characters on/off-screen gives a rhythm that may be used for topic segmentation and film classification, though the annotation of character presence is currently a manual task [5]. Color features have been used to automatically classify the moods of scenes and film genres [6, 7].

This review suggests that encouraging progress is being made with various strategies to bridge the semantic gap between audio-visual features and the higher-level structures and meanings of films. However, researchers typically analyze small numbers of films (between 3 and 15). Also, 'mid-level' semantic content has not been addressed, e.g. the entities, actions and events that are depicted in films. This content is important to access films with respect

0-7803-9332-5/05/$20.00 ©2005 IEEE

to their narrative aspects, such as interactions between characters, cause-effect relationships between events, and story threads [8, 9]. MPEG-7 includes descriptors for semantic entities in narrative worlds [10]. Screenplays and audio description have been identified as rich sources of information to structure and to describe video data, complementary to the extraction of audio-visual features from video data. The use of screenplays, and some challenges and solutions for alignment with video data, are discussed in [11]. The opportunities and challenges for using audio description to describe and structure film content with respect to narrative structures are discussed in [12-14].

3. Analysis of texts that describe films

Here we use screenplays and audio description for another purpose. We want to find out what actions are performed frequently by characters in a wide range of films. We think this knowledge will help to describe the narrative structures of film content. Screenplays, and subsequent scripts, are used in the making of films and give information about characters, their actions, settings, dialog and camera directions. Different scripts are made during the production process, ending with the post-production script which most resembles the finished film. Audio description helps visually impaired people to appreciate films in cinemas and on VHS/DVD releases. In between existing dialog, a describer gives important information about on-screen scenes and events, and about characters' actions and appearances. Audio description is scripted with time-codes before it is recorded. Compared with screenplays, audio description is more tightly aligned with the film (by time-codes) and gives a more succinct description of essential visual information.

We analyzed two collections, or corpora, of texts. One corpus comprised screenplays (mainly post-production scripts) for 75 films, totaling 1,971,950 words, and taken from a variety of websites. The other corpus comprised audio description for 45 films, totaling 399,199 words, and supplied by two British organizations that produce audio description. Only two films were duplicated between the two corpora.

3.1 Frequent phrases – common actions?

We followed a method for identifying idiosyncratic language use, or linguistic variance, in text corpora [15] with the *System Quirk* text analysis package [16]. We sorted the words in each corpus by frequency, took the 100 most frequent words in each, and removed grammatical words (*the, of, but* etc.), Table 1. Some words were only frequent in the screenplays, e.g. *day, night* and *ext* (external) which demarcate and introduce scenes. It seems that the audio description has a more tightly restricted vocabulary because more non-grammatical words appear in the top 100. These words seem to refer to actions like *takes, stands, smiles, steps, opens, runs* and common objects like *window, table* and *bed*. We concentrated on words that appeared in the top 100 of both corpora and that seemed to refer to characters' actions, i.e. *look/looks*, *turns*, *door* (assuming characters are opening / closing or walking through). We also included *smiles* (only in the top 100 for audio description) because it can express characters' feelings.

Table 1. Frequent non-grammatical words

Corpus	Non-grammatical words in top 100 most-frequent words
Screenplays	*day, know, ext, see, night, right, will, time, going, blade, cut*
Audio Description	*takes, walk, sits, hands, white, stands, tom, men, open, john, side, pulls, smiles, stares, goes, puts, steps, watches, water, opens, table, black, window, runs, stops, way, woman, bed, go, red*
Common to both	*looks, door, room, man, look, go, head, turns, hand, eyes, face, car*

Single words, especially common ones, can have many meanings or senses which lead to ambiguities and problems for automatic language processing systems. For this reason it is perhaps better to look at phrases which tend to be less ambiguous. One way to identify frequent phrases is by collocation analysis. Collocation is when one word is co-located, i.e. occurs next to, another. We used statistics from [17] to identify statistically significant collocations for *look**, *turn**, *smile** and *door* - where * indicates the endings 0, -ed, -ing, –s. The most statistically significant collocations gave us the phrases *look* at* (especially *looks at*), *turn* to* (especially *turns to*) and *smiles at*, Table 2. There was a variety of collocations of *door*: * *door* * is *open*/close* (the) door* AND *(the) door open*/close**.

Table 2. Occurrences of frequent phrases

Phrase	Average per Film - Screenplay	Average per Film - Audio Description
look at (looks at)*	29 (17)	11 (8)
turn to (turns to)*	10 (8)	8 (7)
smiles at	2	2
** door **	6	3

We believe that these frequent phrases reflect some of the most common actions performed by characters in feature films. These common actions may be important story-telling elements in filmed narrative. The phrase *look* at* was most frequent throughout both corpora and suggests that one character is focusing their attention on an object or another character. Perhaps knowing this helps the audience to understand what the character is thinking about or who the character is talking with. The phrase *turn* to* also suggests the directing of a characters' attention but maybe it emphasizes the changing of attention in reaction to something. The phrase *smiles at* gives some information about a character's feelings, though the character may not always be smiling happily. The collocations of *door* strike us as interesting because they may be informative about characters' entrances and exits, and may also help to convey important spatial information to the audience. The word *room* was also frequent in phrases such as *enters the room* and *leaves the room*. More detailed analysis of all these phrases is ongoing in order to elaborate our interpretation of the actions as story-telling elements.

3.2 Variation between film genres?

Tables 3a and 3b show the occurrence of the phrases in some of the genres we analyzed; percentage values are included to remove the effects of film and text length. Our initial analysis suggests no significant variation, so it seems that the phrases, and their respective actions, are common across film genres. There is a hint of *looks at* being more important in romantic/period drama films. This is too simplistic an analysis of film genres, with small samples, to make any firm conclusions.

Table 3a. Phrases in screenplays

Phrase	Action / Thriller (44 Films)	Romantic / Period (7 Films)
look at*	29.6 (60%)	47.7 (64%)
turn to*	11.6 (24%)	15.3 (21%)
smil at*	2.0 (4%)	3.7 (5%)
** door **	5.8 (12%)	7.6 (10%)

Table 3b. Phrases in audio description

Phrase	Action / Thriller (9 Films)	Romantic / Period (6 Films)	Children's Films (10 Films)
look at*	9.1 (33%)	17.3 (50%)	8.7 (52%)
turn to*	12.7 (45%)	10.2 (29%)	3.3 (20%)
smil at*	2.0 (7%)	4.2 (12%)	1.5 (9%)
** door **	4.2 (15%)	3.2 (9%)	3.2 (19%)

3.3 Occurrence of phrases over time

Time is an integral and complex component of filmed narrative. Given previous work about tempo, rhythm and the distribution of characters' emotions in films [4,5,14], we are interested to see if any patterns emerge when the occurrences of the frequent phrases are plotted over time, Figures 1a and 1b. Time-codes in audio description makes plotting quite precise, but for the screenplays we have approximated by assuming that the line number of the screenplay correlates with film time. The phrases appear to cluster, which might indicate dramatically important sequences. We are currently working on extracting information about who and what is involved in looking / turning / smiling, etc.

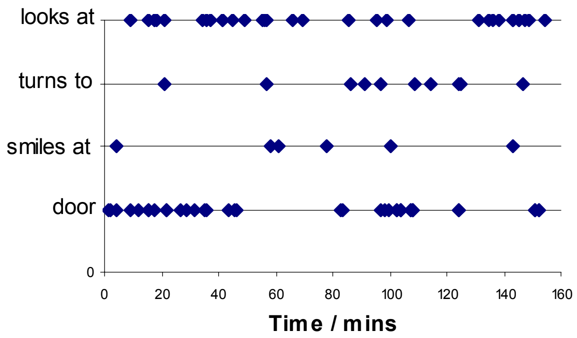

Figure 1a. Plot of phrases occurring in the audio description of *The Horse Whisperer* (1998, Robert Redford, Touchstone).

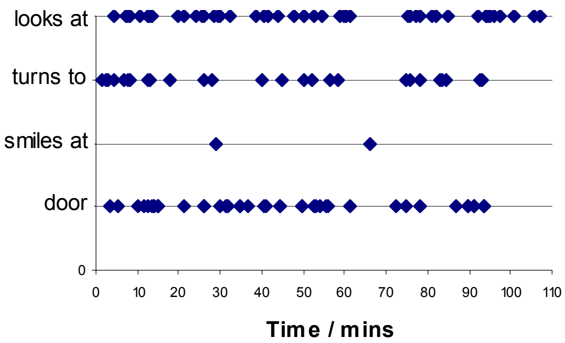

Figure 1b. Plot of phrases occurring in a script of *Wild Things* (1998, John McNoughton, Mandalay).

4. Discussion

The application of a corpus analysis method was successful in giving insights into semantic film content from texts that describe films. A statistical analysis found that phrases with the words *looks*, *turns*, *smiles* and *door* were especially frequent in screenplays and audio description. We argue that the occurrence of these phrases is frequent and widespread because the actions that they describe are important story-telling elements for filmed narrative. If so, then this is a valuable finding for the analysis and description of semantic film content. The extraction of information about these actions, whether from video data or from text, may provide a stepping-stone for bridging the semantic gap. We envisage representations of basic actions like '`looks at`', '`turns to`', '`opens door`' and '`smiles at`' as being empirically-based building blocks for higher-level representations of film content. These representations would address important narrative ideas like characters' entrances and exits, focus of attention and character interactions. We are continuing our corpus analysis to identify patterns that could be used to extract descriptions of film content automatically. Our approach may also be applicable to identifying important elements of semantic content for other kinds of video data that is described by texts.

Previous work in mapping audio-visual features to high-level film content has benefited by being based in the theory of film making and the conventions of film grammar, e.g. [2, 4]. The conventions result in regularities which can be exploited by audio-visual feature detectors. We are interested to explore whether the depiction of basic actions follows patterns in line with film-making and story-telling conventions. A recent theory of narrative describes how actions, time, space and focalization (who can see what) all combine in story telling [18]. Perhaps our kind of analysis can provide empirical data to test theories about film-making and story-telling. In the longer term we would like to synthesize information from texts with audio-visual features.

5. Acknowledgements

EPSRC grant GR/R67194/01 – TIWO: Television in Words. We also thank the members of the TIWO Round Table - BBC, ITFC, Royal National Institute of the Blind and Softel.

6. References

[1] N. Dimitrova et al, "Applications of Video-Content Analysis and Retrieval", *IEEE Multimedia* July-Sept. 2002, pp. 42-55.

[2] H. Sundaram, and S.-F. Chang, "Computable Scenes and Structures in Films", *IEEE Trans. Multimedia* 4 (4), 2002, pp. 482-491.

[3] J. Corridoni et al, "Multi-perspective Navigation of Movies", *Journal of Visual Languages and Computing* 7, 1996, pp. 445-466.

[4] B. Adams, C. Dorai, and S. Venkatesh, "Towards Automatic Extraction of Expressive Elements for Motion Pictures: Tempo", *IEEE Trans. Multimedia* 4 (4), 2002, pp. 472-481.

[5] K. Shirahama, K. Iwamoto, and K. Uehara, "Video Data Mining: Rhythms in a Movie", *Procs. IEEE Int. Conf. Multimedia and Expo*, ICME 2004.

[6] C-Y. Wei, N. Dimitrova, and S.-F. Chang, "Color-Mood Analysis of Films Based on Syntactic and Psychological Models", *Procs. IEEE Int. Conf. Multimedia and Expo,* ICME 2004.

[7] H.-B. Kang, "Affective Content Detection using HMMs", *Procs. ACM Multimedia 2003*, pp. 259-262.

[8] V. Roth, "Content-based retrieval from digital video," *Image and Vision Computing* 17, 1999, pp. 531-540.

[9] R. Allen, and J. Acheson, "Browsing the Structure of Multimedia Stories", *Procs 5th ACM Int. Conf. Digital Libraries*, 2000, pp. 11-18.

[10] www.chiariglione.org/mpeg/

[11] R. Turetsky, and N. Dimitrova, "Screenplay Alignment for Closed-System Speaker Identification and Analysis of Feature Films", *Procs. IEEE Int. Conf. Multimedia and Expo*, ICME 2004.

[12] A. Salway, and E. Tomadaki, 'Temporal Information in Collateral Texts for Indexing Moving Images', *LREC 2002 Workshop on Annotation Standards for Temporal Information in Natural Language*.

[13] A. Salway, M. Graham, E. Tomadaki, and Y. Xu, 'Linking Video and Text via Representations of Narrative', *AAAI Spring Symposium on Intelligent Multimedia Knowledge Management*, Palo Alto, 2003, pp. 104-112.

[14] A. Salway, and M. Graham, 'Extracting Information about Emotions in Films', *Procs. ACM Multimedia 2003*, pp. 299-302.

[15] K. Ahmad, "The Role of Specialist Terminology in Artificial Intelligence and Knowledge Acquisition", in S.-E. Wright and G. Budin (eds.), *Handbook of Terminology Management* (Volume 2), John Benjamins, 2001.

[16] www.computing.surrey.ac.uk/SystemQ.

[17] F. Smadja, "Retrieving Collocations from Text", in Armstrong-Warwick, S., *Using Large Corpora*, pp. 143-177, MIT Press, 1994.

[18] Herman, D., *Story Logic: problems and possibilities of narrative*, Uni. of Nebraska Press, 2002.