Extracting Information about Emotions in Films

Andrew Salway and Mike Graham
University of Surrey
Department of Computing
Guildford, UK GU2 7XH
+44 (0)1483 683133
a.salway@surrey.ac.uk

ABSTRACT
We present a method being developed to extract information about characters' emotions in films. It is suggested that this information can help describe higher levels of multimedia semantics relating to narrative structures. Our method extracts information from audio description that is provided for the visually-impaired with an increasing number of films. The method is based on a cognitive theory of emotions that links a character's emotional states to the events in their environment. In this paper the method is described along with some preliminary evaluation and discussions about the kinds of novel video retrieval and browsing applications it may support.

Categories and Subject Descriptors
H.3.1 [**Information Storage and Retrieval**]: Content analysis and indexing

H.5.1 [**Information Interfaces and Presentation**]: Multimedia information systems

Keywords
Semantic video content, video retrieval, video browsing, narrative, film, audio description, emotions.

1. INTRODUCTION
There is a current trend towards dealing with higher-levels of multimedia semantics, which for video data often means dealing with the entities and events depicted in moving images. In the case of films, human understanding of semantic content goes beyond identification of entities and events. Humans seem to construct rich representations of film characters' motives and feelings in order to make sense of and to anticipate the events unfolding on-screen; we are also able to judge 'similar' stories, perhaps based on these representations. The question then is how can we endow our multimedia computing systems with some sense of these narrative structures? We suggest that information about characters' emotions gives a useful foothold in making sense of their goals, actions and the events around them.

MM'03, November 2-8, 2003, Berkeley, California, USA.

An increasing number of television programs and films are being provided with audio description for the visually impaired. As well as being invaluable as a means for improving access to information for the visually impaired, audio description is also a promising source of information for generating descriptions of semantic video content [6]. We present a method to extract information about characters' emotions in films from audio description; this method is grounded in a cognitive theory of emotion. We then discuss how the information about the emotions depicted at different times in a film might be used to generate descriptions of narrative structures, and how these might support novel video retrieval and browsing applications.

2. BACKGROUND

2.1 Audio Description
Audio description is provided for the visually impaired with some television programs and with films in some cinemas and on VHS/DVD releases. In between existing dialogue a describer gives important information about on-screen scenes and events, and about characters' actions and appearances. Legislation and regulation mean that audio description is becoming increasingly available in countries like the UK, US, Canada, Germany and Japan: we currently have English audio description for 500 films.

Audio description is normally scripted with time-codes before it is recorded. The following is an excerpt from the audio description script for *The English Patient* – time-codes indicate when the audio description is to be spoken:

[11:43] Hanna passes Jan some banknotes

[11:55] Laughing, Jan falls back into her seat

[12:01] An explosion on the road ahead

[12:08] The jeep has hit a mine

Audio description refers to what is depicted on-screen at, or near to, the moment it is spoken; note that the describer must work around existing dialogue. The following non-sequential examples include descriptions of a scene, a character's introduction, a character's physical features and clothing, and an action:

Beneath the aircraft the evening sun throws deep shadows amongst the soft rolling sand dunes.

A young French-Canadian nurse, Hana, adjusts her uniform.

She is wearing a simple white dress and her blonde hair is drawn back from her pale face.

She struggles with a soldier who grabs hold of her firmly.

The language of audio description is rich in information concerning the characters and their external appearance, but

information about their cognitive states, including their emotions, is not described directly. However some insight into a character's emotional state is given by audio description when the emotion is being depicted visually in the film. In the following lines audio description provides information that helps a viewer to understand something about a characters' emotions:

She spins slowly on the rope, laughing with exhilaration.

Kim grows alarmed.

2.2 Emotions and Narrative Structures

Narrative is a multi-faceted phenomenon studied by philosophers, literature and film scholars, psychologists, linguists, cognitive scientists and computer scientists. The study of narrative is concerned with how narrative forms of media are created, how different kinds of media can convey narratives, and, how narratives are understood. For some researchers, narrative abilities considered both as a mode of thought and of communication are fundamental to intelligence. In films narrative relates to how audiences are engaged and kept in suspense, as well as how they recognize similar kinds of stories.

Narrative is commonly defined as a sequence of events, organized in space and time, where the events are linked by cause-effect relationships. For films this definition has been elaborated so that the agents of cause-effect relationships are said to be characters [2]. When we watch a film we make sense of and anticipate the unfolding events that are depicted on-screen, based at least in part on what we think about characters' cognitive states, e.g. their goals, beliefs and emotions.

The fact that an explosion took place on-screen can be explained by a physical cause, e.g. somebody pressed the detonator. However, a more detailed understanding for a viewer includes something about why the character wanted to cause the explosion: consider the following chain of events. Maybe the character wanted to kill another character (Goal), because they were angry (Emotion), because they thought that character had cheated them (Belief), because they saw them steal their car (another physical event). The character's 'anger' might turn to 'satisfaction' if the explosion had the desired effect.

In the preceding, admittedly rather ad-hoc, example an Emotion features prominently as a connection between two physical events. A cognitive theory of emotion, proposed by Ortony and colleagues, treats emotions as agents' appraisals of actions, events and objects in their environment – where their appraisals are made with respect their important goals [4]. In these terms, if someone is 'delighted' it is because something unexpected has happened that helps towards one of his or her important goals; for example, you long for a sports car and then win one. We use Ortony's cognitive theory of emotion as the basis for our work: it links 22 types of emotions to beliefs, goals and behaviors.

2.3 Related work

We are not aware of any other work attempting to extract information about emotions from films, nor of any other work that uses audio description as a source of information for semantic video content. However we feel that our work is complementary to other research that has dealt with film data in multimedia systems. A system was developed to browse between intervals in a film via a semantic network that included cause-effect relationships [5]. Other work explicitly introduced theories related to narrative and story threads in a system to browse multimedia stories, including films [1]. In terms of extracting information directly from digital film data, interesting approaches have been described for recovering information about film narrative from visual and aural patterns in video data [3].

3. METHOD

Our method is to scan for possible descriptions of visibly manifested emotions in time-coded audio description and to classify them: here we use the 22 types of emotion proposed by [4] for classification. A list of emotion tokens (keywords) was produced for each emotion type, such that when used in audio description each keyword could be indicative of an emotion being depicted. For example, 'exhilaration' could be indicative of a character experiencing JOY, and 'alarmed' could be indicative of a character experiencing FEAR. A sample of the lists generated is given in Table 1: in total 679 emotion tokens were selected for the 22 emotion types. Occurrences of emotion tokens were then automatically plotted against the time-code of the audio description for 30 films. Manual inspection of the graphs suggests that the distribution of emotion tokens may reflect some aspect of films' narrative structures.

3.1 Creating lists of emotion tokens

The aim was to create lists of semantically-related keywords clustered around the words used in [4] to refer to the 22 types of emotions, e.g. FEAR, JOY, HOPE, etc. For each of these words we added its grammatical variants, e.g. FEAR → fear, fearful, afraid. We then entered each grammatical variant into WordNet [8] to automatically retrieve its synonyms and hyponyms, e.g. afraid → apprehensive, petrified, unnerved.

Table 1. Some of the 627 emotion tokens selected for 22 emotion types.

Emotion Type	Total tokens	Example emotion tokens
JOY	47	euphoria, elation, happy, jolly, pleased
DISTRESS	50	distraught, anguished, miserable, depressed
LIKE	31	love, passionately, adoration, fondness
DISLIKE	33	hatred, loathing, disgust, aversion, distaste
HOPE	31	anticipation, excited, expectant, optimistic,
FEAR	115	terrified, panicked, worried, concerned

Of course expanding our keywords through a thesaurus-like structure meant that some inappropriate words were introduced, however one-off manual intervention was required for only about 30 of the words retrieved from WordNet. This involved: (i) removing duplications, so that a keyword appeared in the list for only one emotion type; (ii) reclassifying keywords where WordNet was clearly at odds with Ortony's more specialist definitions; and, (iii) removing words which have more common non-emotional meanings, e.g. 'like' - using a part-of-speech tagger could alleviate this kind of problem.

3.2 Output

In the audio description for the film *Captain Corelli's Mandolin* there were 52 tokens of 8 emotion types, as shown in Figure 1.

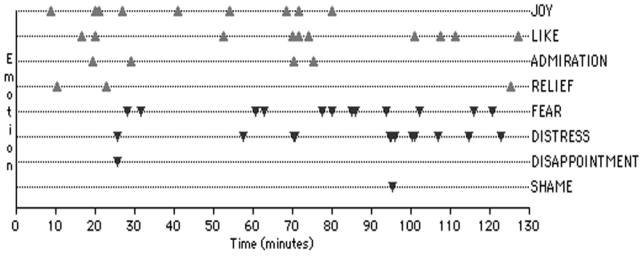

Figure 1. A plot of emotion tokens found in audio description for *Captain Corelli's Mandolin*; emotions only included if > 0.

The story of this film concerns a love triangle between an Italian officer (Corelli), a Greek woman (Pelagia), and a Greek partisan (Madras) on the occupied Greek island of Cephallonia during World War II. A high density of positive emotion tokens appear 15-20 minutes into the film, e.g. JOY and LIKE, corresponding to Pelagia's betrothal to Madras. The negative emotion tokens which immediately follow are associated with the invasion of the island. The cluster of positive emotions between 68-74 minutes occurs during scenes in which the growing relationship between Pelagia and Corelli becomes explicit. The group of FEAR, DISTRESS and SELF-REPROACH tokens between 92-95 minutes maps to a scene in which German soldiers are disarming their former Italian allies, during which a number of Italians are gunned down. The clusters of emotion tokens appear to identify many of the dramatically important sequences in the film.

Further, there appears to be an intuitively appealing sequence of emotion types over the entire course of the film. For the first 75 minutes or so there is a clear sequence of JOY tokens, punctuated by a handful of DISTRESS and FEAR tokens. From this time onwards the JOY tokens disappear, and DISTRESS and FEAR tokens increase in frequency as violence descends on Cephallonia. Towards the end of the film, the last DISTRESS token appears, followed by a RELIEF token when Pelagia discovers her father has survived an earthquake, and a LIKE token as Pelagia and Corelli are reunited. There is some correspondence here with notions of story structure and conventional character behavior.

The graph in Figure 1 can be compared to a graph produced from the audio description for *The Postman*, Figure 2, which contains fewer LIKE and JOY tokens, and contains PRIDE and HOPE tokens that were not found in *Captain Corelli's Mandolin*. This science-fiction film portrays events in a post-apocalyptic America when a nameless man, dressed in a postman's uniform and posing as an emissary of the Restored United States of America, inspires an uprising against a tyrannical warlord which eventually leads to a true restoration of government. Like the previous graph, FEAR tokens are a recurrent feature of this audio description, but perhaps the lower number of LIKE and JOY tokens is due to the less prominent love interest in this film. Similarly the greater number of PRIDE, ADMIRATION and HOPE tokens might be related to the patriotic nature of the story.

We acknowledge that in the preceding discussion our interpretation of the graphs is subjective, and informed by our prior knowledge of the films. However we are encouraged by these, and other results, to think that we might be taking a step towards extracting something about films' narrative structures.

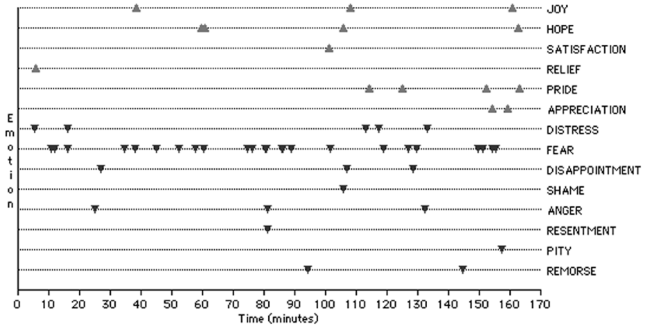

Figure 2. A plot of emotion tokens found in audio description for *The Postman*; emotions only included if >0 tokens found.

3.3 Validation Test

Our preliminary evaluation to date consists of a small-scale test to see how the emotions extracted by our method correspond with those observed by human subjects.

Subjects were asked to watch short excerpts from recent films and then report up to three emotions that they thought were depicted in the sequence. Ten sequences (1-3 minutes in length) from four films were selected according to our judgment to cover a wide range of emotion types. The sequences were played to 10 subjects who were asked to choose from Ortony's 22 emotion types when giving their responses. Note our automated method extracted 23 occurrences of emotions in the 10 sequences.

If an emotion was selected by a majority of subjects for a sequence then that emotion was considered to be *important* for the subject group. This gave 19 important emotions across the 10 film sequences, normally only one or two per sequence: 12 out of these 19 were identified by our method (63%). Of the remaining 11 suggested by our method, 7 were observed by one or more subject – leaving 4 'false positives'.

Such false positives may be due to the fact that our method ignores the strength of the emotion tokens. For example, 'concerned' and 'terrified' are both recognized as tokens of FEAR, whereas human subjects may not consider 'concerned' to be important amongst other stronger emotions. With regards to 'false negatives' it seems that all the prominent emotions that our method missed were being conveyed non-visually, e.g. through the intonation of the dialogue.

3.4 Sources of Further Information

Dialogue, music and lighting can all be informative about characters' emotions. Audio description only refers to emotions that are visibly manifested, such as those conveyed by facial expressions. It would be interesting to fuse information from audio description with other media streams. There are also other textual information sources to consider like closed captions, screenplays and plot summaries. Combining information from these different sources might help to provide more evidence for the occurrence of emotional states. It might also be possible to extract more information about each occurrence, like who is experiencing the emotion, how strongly, and why.

4. NARRATIVE STRUCTURES?

If we can extract information about what emotions are being depicted on-screen at certain times then it is trivial to retrieve video intervals depicting particular emotions. More interestingly

perhaps we can consider how the distribution of emotions, linked as they are to characters' goals and the events taking place around them, might reflect some aspects of films' narrative structures. Film theorists have proposed structures relating to tension, inciting incidents and dramatic climaxes, genres, and character motivation.

By far the most common emotion identified in the 30 films we have analyzed is FEAR. Maybe this reflects an integral ingredient of films, which require elements of forward-looking suspense to sustain audience interest. FEAR is a so-called *prospect-based* emotion because it involves characters expectations about a future, undesirable, event. Many of the films we have analyzed end with positive emotion tokens, like RELIEF and LIKE which could suggest the resolution of tension.

We might assume that characters' emotional states are most intense at the moments of highest drama. If so then the occurrence of relatively dense groups of emotions would point to dramatically important sequences within a film, such as an inciting incident or a dramatic climax.

Further, if we were able to attribute emotions to characters automatically then perhaps we could start to build up chains of reasoning to explain some of the main events of a film. For example, in between a character's FEAR and their subsequent DISTRESS we might expect to find the event causing these emotions. We have used Ortony's cognitive theory of emotion in our work with the long-term aim of linking representations of emotional states to events, actions and objects in a film.

As well as looking into the dramatic elements of a film, the relative frequencies and distribution of emotions may characterize a film as a whole. For example, a love story might contain a higher proportion of LIKE, HOPE and ADMIRATION than a horror film dominated by FEAR and DISTRESS. Our informal analyses suggest that this is indeed the case. We have also observed 'dark' films with few or no positive emotions, and comedies with a distinct balance of FEAR and ANGER; in comedies the FEAR emotions tend to of weaker kinds than in horror, e.g. 'nervous' rather than 'terrified'. We expect that more extensive, statistically-grounded analyses will lead to interesting ways of classifying films by 'emotion histograms'. We are also interested in trying to characterize films by the prominent emotion in the first, second and final third of the film.

The ideas discussed here are being investigated at the University of Surrey as part of the TIWO project [7]: this project aims to develop video retrieval and browsing applications, like:

- Video Retrieval by Story Similarity – based on films' 'emotion histograms'
- Video Summarization – based on identifying important sequences from dense groupings of emotions
- Video Browsing via Cause-Effect Links – based on attributing emotions to characters, and associating them with physical events

The work presented here is a promising step towards generating richer representations of semantic video content. Information about characters' emotions in films seems to give insight into narrative structures – at least to the human eye. The challenges being addressed now in the TIWO project relate to: (i) automating the analysis of the 'basic' emotion information – e.g. similarity metrics for comparing film clips and reasoning about sequences of emotions and associated events; and, (ii) supplementing the basic emotion information by combining cues from other sources like music, color and intonation, and by further language processing to attribute emotions to characters and to determine relative strengths of emotions. This work will be informed by relevant film theory and cognitive theories of emotions.

5. ACKNOWLEDGMENTS

This research is supported by EPSRC grant GR/R67194/01 – TIWO: Television in Words. We thank the members of the TIWO Round Table (BBC, ITFC, Royal National Institute of the Blind and Softel) for sharing their knowledge of audio description. Thanks also to Elia Tomadaki who assisted in the preparation of data for this research, and thanks to three anonymous referees for their helpful comments.

6. REFERENCES

[1] Allen, R. B. and Acheson, J. Browsing the Structure of Multimedia Stories. In Procs. 5th ACM Conference on Digital Libraries, 11-18. ACM Press, New York, 2000.

[2] Bordwell, D. and Thompson, K. Film Art: An Introduction. McGraw-Hill 5th Edition, New York, 1997.

[3] Dorai, C. and Venkatesh, S. Computational Media Aesthetics: Finding Meaning Beautiful! IEEE Multimedia 8(4), 10-12, (Oct.-Dec. 2001).

[4] Ortony A., Clore G. L. and Collins A. The Cognitive Structure of Emotions. Cambridge University Press, 1988.

[5] Roth, V. Content-based retrieval from digital video. Image and Vision Computing 17, 531-540 (1999).

[6] Salway, A., Graham, M., Tomadaki, E. and Xu, Y. Integrating Video and Text via Representations of Narrative. In Procs. AAAI Spring Symposium 2003, Intelligent Multimedia Knowledge Management.

[7] TIWO – Television in Words, EPSRC research project, www.computing.surrey.ac.uk/ckm/tiwo_project/Index.html

[8] WordNet – online lexical reference system, http://www.cogsci.princeton.edu/~wn/

Films

Ford Coppola, F. (1972) Apocalypse Now. Zoetrope Studios, USA.

Madden, J. (2001) Captain Corelli's Mandolin. Universal Pictures, USA.

Minghella, A. (1996) The English Patient. Miramax Films, USA.

Raimi, S. (2002) Spiderman. Columbia Pictures, USA.

Soderbergh, S. (2001) Ocean's Eleven. Warner Brothers, US