Study of Shot Length and Motion as Contributing Factors to Movie Tempo

Brett Adams
Department of Computer Science
Curtin University of Technology
GPO Box U1987, Perth, 6845, W. Australia
adamsb@cs.curtin.edu.au

Chitra Dorai
IBM T. J. Watson Research Center
P.O. Box 704, Yorktown Heights
New York 10598, USA
dorai@watson.ibm.com

Svetha Venkatesh
Department of Computer Science
Curtin University of Technology
GPO Box U1987, Perth, 6845, W. Australia
svetha@cs.curtin.edu.au

ABSTRACT

This work seeks to lay the framework of film grammar over the video to be analyzed. We use the shot attributes of motion and shot length to produce a novel continuous measure of one of the aesthetic elements of films, namely the movie tempo. We refer to our previous work detailing the study of this construct and its automatic derivation, and also demonstrating its usefulness as an expressive element and as a sound basis for higher semantic descriptions such as dramatic events and story elements. Initial assessment of tempo was performed in our study on the basis that the relative importance of both shot length and motion in formulating the tempo function was the same. In this paper, we analyze their relative contributions to tempo, and demonstrate how these two factors can be manipulated to influence audience perception of movie time.

1. INTRODUCTION

Film "grammar" [3, 5] outlines the cinematic conventions that are employed by directors to manipulate film elements such as the shot, the movement, and the types of editing to convey a narrative structure of a story. Inspired by the existing film grammar, we propose a unique approach to computationally determine the expressive elements of motion pictures conveyed by the manipulation of editing, motion, colour etc. for high level video understanding and annotation. We seek to find computational elements that can map to the "expressiveness" of the medium, and determine thematic movie sections underscored by the expressions.

In our previous work [2, 1], we concentrated on a computational understanding of movie tempo or pace, that manifests in the sense of a story's *experienced* time. Sobchack says that "[tempo] is usually created chiefly by the rhythm of editing and by the pace of motion within the frame" ([5, p. 103]). Encyclopedia Britannica [4] defines tempo as being influenced "in three ways: by the actual speed and rhythm of movement and cuts within the film, by the accompanying music, and by the content of the story". We proceeded from this artistic definition to formulate a computable measure, briefly recapitulated below, of tempo/pace for a given film.

As a new contribution in this paper, we analyze the relative impact of shot length and motion on our tempo function, and demonstrate how the two factors can be manipulated to influence audience's perception of time. Film grammar tells us that different directors and different genres will make use of these complementary techniques with differing measures. We conduct experiments on films of different genres and find that tilting the sensitivity of the tempo function toward one or the other of its constituents causes different "flavours" of movie tempo to be emphasized accordingly.

2. EXTRACTION OF MOVIE TEMPO

We first proposed tempo/pace function of a video sequence in [2] and improved it in [1] as:

$$P(n) = \alpha(W(s(n))) + \frac{\beta(m(n) - \mu_m)}{\sigma_m}, \qquad (1)$$

where s refers to shot length, m to motion magnitude, and n to shot number. The shot length, s, is computed in frames (from source 25 frames/sec) for each shot detected in the video, and m is the absolute value of the sum of pan and tilt values computed for every frame pair in a shot, and averaged for each shot. In addition to the per shot data, the mean, μ and the standard deviation, σ of these features are calculated for the entire film for normalization. $W(s(n))$ is a novel two-part normalization scheme proposed for shot length in [1] where it is defined to be more sensitive near the overall shot length median. For the motivation and derivation of this normalization scheme, see [1]. The weights α and β affect the contributions of shot length and motion to the perception of pace, and are given values of 1. $P(n)$ is smoothed with a bank of Gaussian filters as a response to the neighbourhood nature of pace, and as a means to multi-resolution video analysis.

3. SHOT LENGTH VS. MOTION

While it has been recognized that both motion and shot length contribute to the perception of pace, it is not as simple to make statements about their relative impact.

ACM Multimedia 2000 Los Angeles CA USA

Our research on $P(n)$ has thus far used unit weights for α and β, assuming that motion and shot length contribute equally to the perception of time. It is possible, however, that under certain circumstances one or the other of these two impact more heavily on the audience perception of time. Such circumstances might include sparing use of a technique. For example, a director who makes minimal use of quick cutting techniques might do so only at particularly important junctures in a film's development. Conversely, a director might rely heavily upon one technique to clarify the story. An example of this might be films crafted by montage directors, who make use of shot characteristics (e.g., length) for the purpose of carrying artistic expression to the exclusion of other cinematic factors like motion. This results in one technique being used to influence semantic interpretation, thus relegating other techniques such as motion to a position of secondary importance.

In order to explore this further, we carried out two experiments on the movies, Colour Purple (CP) and Lethal Weapon 2 (LW2). For each movie, edges were automatically detected in its $P(n)$ using Deriche's multi-scale edge detection algorithm and the resulting story sections manually labeled, since significant pace changes as indicated by the edges often accompany dramatic events, or are precipitated by story/scene boundaries. The results of our equal weights-based pace scheme for LW2 can be seen in [1] as an example of this process. These two movies have been chosen as good representatives of a slow thoughtful movie and an action movie respectively. Having computed their reference pace functions with the equal weights-based pace scheme, experiments reported in this paper consist of tilting the weights α and β, in effect adjusting the sensitivity of $P(n)$ to motion or shot length respectively.

3.1 Increase Sensitivity to Less Used Technique

For the first experiment the weight corresponding to the "least" used technique is increased. For CP, this is motion (α), with an overall average of only 0.58 indicating that it is not a very visually dynamic movie. Conversely, for LW2 it is shot length (β), due to its large motion average of 0.83 (as expected for an action movie). In each case, the amplified mode receives a weighting of 1.5, while the other weighting is dropped to 0.5. Figures 1(a) and 1(b) show pace plots for the two movies, using the normal and amplified weighting schemes. See Table 1 for edges of $P(n)$ from new weightings.

For LW2 the most obvious effect is a small swing towards less "action" oriented edges. That is, edges more to do with character/plot development and background than with the dynamic action sequences that are the hallmark and raison d'etre of this genre of film. This is to be expected as the technique of crafting shot length is generally much more dominant in such scene types as dialogue, character development etc. (to well made scenes of these types at least). As such, these events often serve as breathers to the main action of the film, or provide vital pieces of information to the dramatic development of the plot that serve to clarify and justify its existence.

As an example of this, consider edge "B" from Table 1, labelled "Murtaugh's daughter on commercial". The scene takes place at Murtaugh's home. Family, friends and Riggs have gathered to watch Murtaugh's daughter on her small screen debut. The result is a mixed reaction of disbelief, anger, and humour as it turns out to be quite an embarrassing situation for her poor father. The event is an occasion to see the family bonds that exist, the friendship between Riggs and Murtaugh, and also serves as a breather from the hectic pace of the surrounding scenes. Detection of this event would be desirable as it provides a useful index to the overarching story, but in this case we see that it would be lost in the mass of action oriented edges that arise when no special consideration is given to edges of this type.

The results from CP are not as amenable to clear conclusions. The reason for this becomes clear after a consideration of the movie. Unlike LW2, the story told by CP is by nature "amotion", i.e., motion not used to such a degree, with the centrepiece of the story being the relationships that exist between the characters. This is the essence of the story. Rather than forcing motion on the story to alter time perception, the more readily avaliable and natural method open to the director is shot rate change. Thus highlighting motion is missing the point somewhat. In other words, of the methods available to the director to manipulate the perception of time for CP, shot length is much more appropriate and fits seamlessly with the story. Therefore amplifying the effect of any incidental motion is unjustified and results in meaningless edges. No resulting clear trend exists and fits with the fact that we are amplifying *semantic noise*.

3.2 Increase Weight of Primary Technique

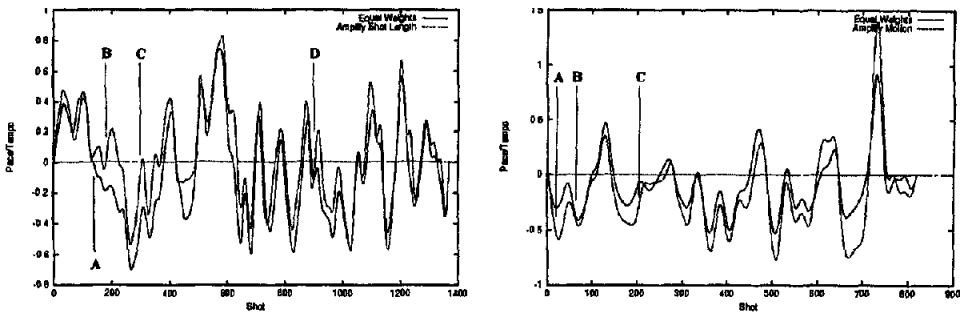
For the second experiment the weight corresponding to the most used mode is increased. This results in an emphasis on the opposite modes as compared to the previous experiment. For LW2 motion is emphasized, for CP it is shot length. Once again, the amplified mode is given a weighting of 1.5, with the remaining mode receiving a weighting of 0.5. This second experiment in effect says "we recognize that this mode is predominant, and therefore we're going to pay special attention to it as the director seems to be placing much upon it". Refer to Figures 2(a) and 2(b) for comparative pace plots, and Table 2 for edge descriptions.

Table 1: Experiment 1 - Resulting significant edge changes in $P(n)$ (cf. Figures 1(a), 1(b)).

Gained and Lost Edges			
Lethal Weapon 2 Edges			
A	-	End of car chase	action
B	+	Murtaugh's daughter on commercial	tense/humour
C	+	Riggs talks about his wife's death	tense/sad
D	+	Riggs meets the baddies in room	tense
Colour Purple Edges			
A	+	Mr asks to marry Nettie	tense
B	-	Mr slaps Celie	tense
C	+	Harpo's wedding	lively

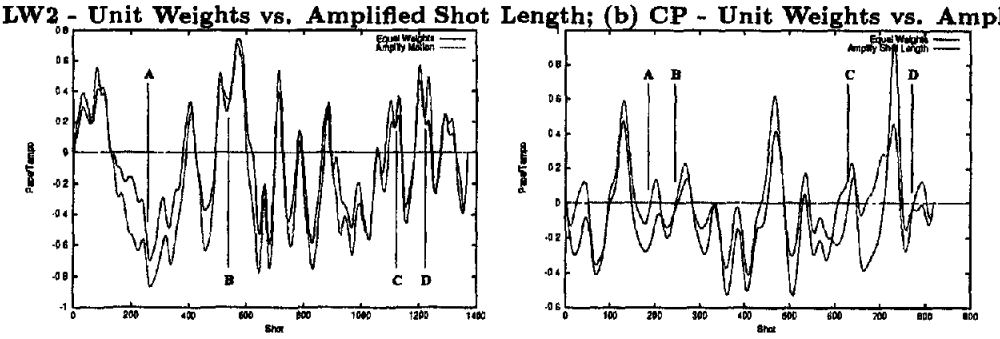
For LW2 we observe an effect opposite to that found in the first experiment. There is a shift towards action based events, those noted above as being the trademarks of this kind of film. We also find better resolution of different events within larger action sequences. Edge "D" is an example of this; a scene entailing the destruction of the crooks house.

A second observation has to do with the fact that, especially with this kind of intense film, both motion and shot length techniques are often used to compound their effects to cause a greater impact on the viewer. Both techniques are often used together to overwhelm the audience visually. Given this, where one set of data is subject to error (par-



(a) (b)

Figure 1: (a) LW2 - Unit Weights vs. Amplified Shot Length; (b) CP - Unit Weights vs. Amplified Motion.



(a) (b)

Figure 2: (a) LW2 - Unit Weights vs. Amplified Motion; (b) CP - Unit Weights vs. Amplified Shot Length.

Table 2: Experiment 2 - Resulting significant edge changes in $P(n)$ (cf. Figures 2(a), 2(b)).

Gained and Lost Edges			
Lethal Weapon 2 Edges			
A	-	End of Rian's TV debut	tense/humour
B	-	After gun fight at crooks house	action breather
C	+	Riggs fights back (caravan scene)	action
D	+	Crooks house begin to shake	action
Colour Purple Edges			
A	+	Sophia talks to Mr about Harpo	tense/dialogue
B	+	Shug Avery arrives	rising tension
C	→	Celie about to kill Mr	rising tension
D	+	Shug and others join Church	climax

ticularly the shot index), the other technique may be used to locate edges that would otherwise be missed. This may be viewed as taking advantage of a form of inherent redundancy. Edge "C" offers an example of this observation. The scene involves the gunship attack on the caravan in which Riggs lives. Originally only the edge marking the start of the attack was detected. There is however a dip in shot rate just before Riggs retaliates, that is not captured in the shot index due to the fact that it is a darkly lit scene. There is also a motion pause in concert with this shot rate drop which is now detected with amplified motion.

The converse is true for CP. There is a swing towards dramatic events. Given the discussion above of the relatively minor role of motion in the fashioning of this story, it is to be expected that the result of desensitizing the pace function to motion is to more faithfully bring out the ebb and flow of the story. Edge "C" is "shifted" to the point where Spielberg brings his increasing metric montage to a climax. In addition to this improved accuracy, edges "A", "B", and "D" refer to dramatic events that have been detected in this new configuration, with the lessening of the motion *noise* (i.e., not indicative of events) in the pace measure.

4. CONCLUSIONS

We have shown how the film grammar relating to the production of motion pictures can be exploited for the purpose of creating tools that can extract semantic information from motion pictures. A function for the calculation of the expressive element, pace or tempo has been briefly outlined and demonstrated. Further, we have experimented with the relative contribution of each of the two fundamental components of pace, namely shot length and motion. The results indicate that it is possible to sensitize the pace function to different flavours of dramatic event. They also serve as further impetus to sift the body of film grammar for aid in the task of extracting useful information from the film medium.

5. REFERENCES

[1] B. Adams, C. Dorai, and S. Venkatesh. Role of shot length in characterizing tempo and dramatic story sections in motion pictures. In *IEEE Pacific-Rim Conference on Multimedia*, Sydney, Australia, December 2000.

[2] B. Adams, C. Dorai, and S. Venkatesh. Towards automatic extraction of expressive elements from motion pictures: Tempo. In *IEEE International Conference on Multimedia and Expo*, New York City, USA, July 2000.

[3] D. Arijon. *Grammar of the film language.* Silman-James Press, 1976.

[4] E. Britannica. Encyclopedia Britannica Online, 1999.

[5] T. Sobchack and V. Sobchack. *An introduction to film.* Scot, Foresman and Company, 1987.