

ROLE OF SHOT LENGTH IN CHARACTERIZING TEMPO AND DRAMATIC STORY SECTIONS IN MOTION PICTURES

Brett Adams[†], Chitra Dorai[‡], Svetha Venkatesh[†]

Department of Computer Science[†]
Curtin University of Technology
GPO Box U1987, Perth, 6845, W. Australia
{adamsb, svetha}@cs.curtin.edu.au

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

ABSTRACT

Motivated by existing cinematic conventions known as film grammar, we proposed a computational approach to determine *tempo* as a high-level movie content descriptor as well as means for deriving dramatic story sections and events occurring in movies. Movie tempo is extracted from two easily computed aspects in our approach: shot length and motion. Story sections and events are generally associated with changes in tempo, and are thus identified by edges located in the tempo function. In this paper, we analyze our initial founding of the tempo function on the basis that the distribution of both shot length and motion in movies is normal. Given that the distribution of shot length is approximately Weibull as confirmed in our experiments, we examine the impact of modelling and modifying the contributions of shot length to tempo. We derive an appropriate normalization function that faithfully encapsulates the role of shot length in tempo perception, and analyze the changes to the story sections identified in films.

1. INTRODUCTION

Given the explosive growth in multimedia data, effective systems are required for annotation and retrieval. Whilst a great deal of research has focused on video annotation using low level image primitives, it falls far short when applied for the purposes of higher level semantic interpretation in real world user scenarios.

Our project is seeking to build video descriptors from which higher level semantic constructs can be derived, thus establishing a computational framework for film content analysis. We approach the problem uniquely, in first defining high-level semantics associated with the *form* of story narration in films, identify cinematic elements that are associated with these precepts, and then examine mechanisms by which we can compute these cinematic elements.

Tempo/pace is one such descriptor of cinematic elements. It is closely linked to the audience 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” ([11, p. 103]). Encyclopedia Britannica [6] 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 hypothesize that tempo can be initially computed as a function of shot length and motion. The extraction of movie tempo, takes us towards automatic understanding of film and video, enabling content labelling with high level semantics of stories, thus allowing for better video annotation systems.

For the most part in automatic video analysis, film grammar [4] is utilized in an implicit and rudimentary manner. Scene extrac-

tion methods that are based on assumptions like “where there is a dissolve” or “a significant change in colour characteristics over a number of shots” ([8, 10]) are examples of such. While Yoshitaka et al. [13] refer to film grammar explicitly as predicting certain signatures for different dramatic events, their work is directed towards distinguishing from among 3 chosen scene types (where test data is taken from the sample space of these 3 types).

In [3] we sought to make logical deductions about the expressive element of movie tempo, guided by the “tenets” of film grammar. We developed a novel continuous tempo measure and demonstrated its usefulness, thus strengthening the claim that film grammar provides us with a vital body of knowledge for the creation of tools that can extract semantic information from film.

In this paper, we seek to further improve the tempo measure with a more detailed study of the factors that affect shot length, one of the two fundamental components of the proposed tempo function. We examine shot length distribution models and modify the contributions of shot length to tempo using a suitable normalization function.

2. A RECAP. OF THE TEMPO FUNCTION

In [3] we proposed the following function for the purpose of providing a shot by shot measure of the tempo of a movie. A digital movie is automatically segmented into shots using standard techniques published in the literature [1], with average motion (inclusive of camera's) magnitude and shot length computed for each shot. In addition to the per shot data, the mean, μ and the standard deviation, σ of these two features are calculated for the entire film, along with the overall shot length median, med_s .

The proposed tempo/pace function takes the form:

$$\mathbf{P}(\mathbf{n}) = \frac{\alpha(med_s - s(\mathbf{n}))}{\sigma_s} + \frac{\beta(m(\mathbf{n}) - \mu_m)}{\sigma_m}, \quad (1)$$

where s refers to shot length in frames, m to motion magnitude, and \mathbf{n} to shot number. The weights α and β , are given values of 1, effectively assuming that both shot length and motion contribute equally to the perception of pace for a given film.

Our computational framework for tempo includes smoothing $\mathbf{P}(\mathbf{n})$ with a Gaussian filter for the purposes of noise removal, and as a means of reflecting the fact that film tempo is a function of a neighbourhood of shots.

2.1. Tempo Based Dramatic Story Sections and Events

Noting the fact that significant pace changes often occur across story sections and are also precipitated by local dramatic events,

we extract the edges of $P(n)$ that capture significant changes in pace to demarcate these events and sections as useful features. Different story sections and events are generally associated with changes in tempo, and are thus identified by the edges located in the tempo function. Deriche’s [7] recursive filtering algorithm is used to extract edges of different slope and scale indicated by various Σ and τ combinations employed in the edge detection process.

2.2. Experimental Results

Results from one of the four movies analyzed, *Titanic*, *Colour Purple*, *Lethal Weapon 2*, and *The Lost World: Jurassic Park 2* are presented here.

Lethal Weapon 2 (LW2) is a typical action film starring Mel Gibson (Riggs) and Danny Glover (Murtaugh), who are joined by Joe Pesci (Leo Getz). Detectives Riggs and Murtaugh stumble across a South African diplomatic official who is running drugs, and attempt to stop him and his band of “crooks”. Figure 1 shows the pace plot of a section of LW2 with located edges indicated for each of the 4 Σ/τ combinations used, and Table 1 matches each automatically discovered edge to a brief description of the *story section* bounded by, or the dramatic *event* coinciding with the discovered edges. The zero axis in Figure 1 may be roughly considered as the average pace mark for the film.

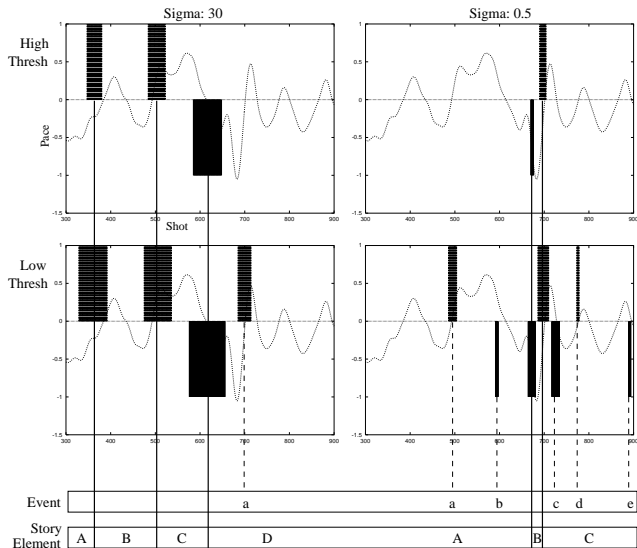


Figure 1: Results of edge detection on tempo/pace flow and corresponding story sections and events from Lethal Weapon 2.

Consider Table 1. A large gradual pace change occurs at the transition between the story elements, *B* and *C*, labelled as “Meet Leo Getz” and “At Crooks House” respectively. The change in tempo occurs after the revelation that Leo knows the location of the Crooks, the team head out to investigate and upon arrival a gunfight and chase ensues. This rise in tempo is faithfully captured by our algorithm. On a finer scale note edge *d*, labelled as “Riggs Calls Bomb Squad”. This sharp rise in pace is precipitated when, upon discovering Murtaugh in a precarious position, Riggs calls the bomb squad resulting in hordes of officious squad members and other interested onlookers descending upon Murtaugh.

Even from this brief summary, it should be apparent that the developed tempo function has a number of desirable qualities. In this paper, we seek to further improve the tempo function with a detailed study of shot length characteristics.

Gradual Edge, Sigma: 30		Sharp Edge, Sigma: 0.5
<i>Story Element</i>		
A	R and M Home, and Office	With Crooks
B	Meet Leo Getz	Next Day’s Fallout
C	At Crooks House	Crooks Hit Back
D	Meet Boss and Next Day	
<i>Event</i>		
a	Riggs Haunts Boss	Arrive at Crooks House
b		End of Car Chase
c		Riggs Leaves Consulate
d		Riggs Calls Bomb Squad
e		Riggs Slips Past Diversion

Table 1: Labelled story sections and events identified from tempo changes in Lethal Weapon 2.

3. ENHANCEMENTS TO SHOT LENGTH NORMALIZATION

$P(n)$ is a composite of two shot features, motion and shot length. Each feature is normalized such that a given value makes an intuitive contribution to the overall calculation of pace for a shot. The normalization process is undergirded by assumptions that should be noted here.

First, motion. With no impetus to proceed otherwise, it is assumed that motion is normally distributed and therefore the form of normalization seen in Equation 1 for motion is adopted. The second component of the pace function is shot length. Currently, it too is treated as being drawn from a normal distribution (see Equation 1). However, an analysis of the raw data, and of the processes leading to the formation of shot length, suggest that a better model for normalizing shot length is required.

Vascelos et al. [12] make the point that shot length appears to be adequately modelled by a member of the Weibull distribution family. Our experimental results also confirm this, from the distributions of shot length data for movies from various genres such as *Titanic*, *Lethal Weapon 2*, *Lost World (Jurassic Park 2)*, *Colour Purple*, and *The French Connection*. These distributions are roughly Weibull. Figures 2(a) to 2(c) show several examples.

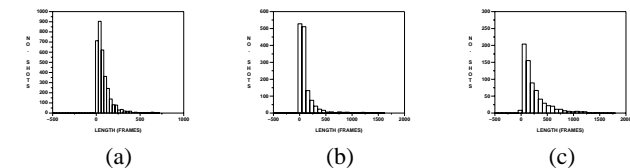


Figure 2: Shot length distributions: (a) *Titanic*; (b) *Lethal Weapon 2*; (c) *Jurassic Park 2*.

This apparent distribution is also somewhat predictable from a consideration of human abilities and the movie making process as explained in the following section.

3.1. Shot Length Distribution Characteristics

For most situations there is a practical lower limit placed on shot length by the ability of a human viewer to adjust to and process the information of a new shot. It is a testimony to the art of directing that these transitions, for the most part, go unnoticed. The fact is that a movie is made up of a series of disjoint shots that would appear very confusing were it not for that fact that we are “trained” to interpret them ([9, p. 121]). [5] refers to studies that indicate that it can take between 0.5 to 3 seconds for a viewer to adjust to a

new shot, let alone absorb the contents of that shot. While this can sometimes be used as a creative tool, ([5] notes the sequence in Patriot Games where a series of very short shots is used to create tension as Jack Ryan watches the Terrorists killed on the other side of the world via satellite. It should be noted also, that there is very little new information to be assimilated in each of the consecutive shots), usually it is desired that the information in each shot be taken in and added to the growing story.

The upper limit to shot length is a lot more amorphous. Whereas the lower limit has a large degree of inflexibility to it that is derived from its physical nature, the upper limit generally tends to be dependent on more subjective factors like audience interest levels and the degree of complexity of story to be captured. For example, many very long shots may lose audience interest, or be unable to aptly convey the desired information intended by the story teller.

The logistics of movie making also impacts on the possible shot makeup of a film. The more intricate the pattern of shots, the more the cost in time and money (although new editing and recording technology will affect this to a degree). Often a director has to resort to using significant pieces of the Master (or Cover) shot when planned interleaved shots are deemed unsatisfactory [9]. In other words, the final shot makeup of a film is not always, if ever, the creative ideal envisioned by the director/editor and hence will, to a degree, reflect these underlying processes in its manifest shot length distribution.

Given that shot length data exhibit something like a Weibull distribution, we address the following question: How do we then formulate the pace equation such that it makes comparable and intuitive contributions from the shot length to the output?

3.2. Modeling an Appropriate Normalization Function

In one sense the shape of the distribution for shot length does not tell us anything about the connection of a certain shot length to a certain perceived time. For this we need to consider perception of time, goals (from the director's point of view), and reception (from the audience's point of view).

Zettl [14] offers some clues as to what a director is trying to achieve with different shot length ranges. Small shot lengths are manifest during the latter part of an increasing metric montage, or during a fast paced metric montage. The director's goal here is to intensify the event by an increasing, or high event density, achieved by means of rapid shot changes. The result being a fast paced section. Shots whose lengths lie closer to the overall median value are possibly the result of a great many factors. Rhythmic montage, medium paced metric montage, and many narrative restraints all result in shot lengths at or near the median. In one sense, this range of shots is the default range for maintaining audience interest levels. As such these shots assume the role of "mid-point" in a tempo estimation. Above the median, typically shot length can range up to a considerable maximum length. Shots of longer length have less of a role in decreasing event density as event clarification can be carried by other methods. That is, the role of shot length in relation to influencing audience perception of time becomes subordinated by the in-shot elements such as primary/secondary (object/camera) motion, story development, and dialogue etc. From an indication based purely on shot length these shots indicate a below normal tempo. Proportionally however, their influence should be considered to be decreasing based on the assumption that the in-shot parts may be being made the vehicle of tempo.

The role of shot length in affecting perceived time is most pro-

nounced from the minimum shot length to somewhere above the median (but well short of the maximum shot length). Our shot normalization scheme should thus be most sensitive in these areas, and less so as shot length proceeds beyond this area. Added to this are some others factors that influence human perception of pace.

3.3. The Proposed Shot Length Normalization Scheme

We choose to separate the shot length data computed from a movie into two sections, shot lengths below and above the overall median.

The median is chosen as the "zero" point of the contribution of shot length to pace. Half of the shots have durations above this point, and half below (by definition). The median provides a more robust estimate of the average in the presence of outliers.

The sample space below the median is well contained (by the minimum shot length below, and the median above) and as such can be weighted with a simple linear model. The slope of the curve is chosen such that the minimum shot length coincides with a unit weighting (for symmetry).

Above the median is a different story. Shot length is theoretically unbounded, and in practice includes outliers far removed from the preponderance of data. A linear model would, in general, under weight the majority of data.

A better scheme would be to use the hazard function of an underlying Weibull model. Experiments with distribution fitting software [2] reports a beta of the order of 1.5, which results in a hazard function of the form $xpow(1/2)$. Fitting this curve to reach -1 (for symmetry) at the 95% mark of shot lengths, for robustness, results in the following overall two part weighting scheme. This function has the property of being more sensitive near the median, but slows in gradient as shot length increases into the "longer" range. This is a desirable attribute given the above discussion of shot length ranges and their contribution to pace. Figure 3 shows a plot of the new shot length normalization scheme, and the tempo function assumes the form,

$$P(n) = \alpha(W(s(n))) + \frac{\beta(m(n) - \mu_m)}{\sigma_m}. \quad (2)$$

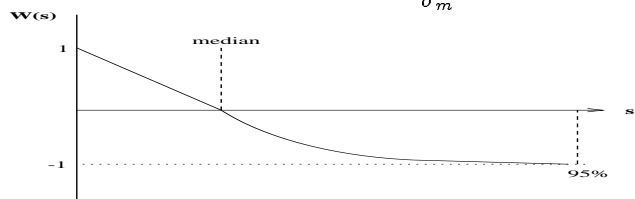


Figure 3: New shot length normalization scheme for tempo computation.

4. RESULTS WITH THE NEW P(N)

Figure 4 shows what the new shot normalization scheme does to the pace and to the detected events of LW2. It shows that the new scheme fills in the large tempo drops that occur at shots 700 and 1200. These scenes ("Boss talks to girl/Riggs, Murtaugh and Leo in car", and "the Calm before Riggs pulls the house down" respectively) involve sequences of very long shots, which are exacerbated by degradation of shot detection due to poor lighting. In both cases the diminishing contribution of shot length to pace has been reflected by the new normalization scheme, resulting in a more intuitive drop in the pace level for each scene. Analysis of the differences caused by the new scheme is looked at in further detail below.

Figure 5 and Table 2 show new results from LW2 for shots 450 through 850. Overall, the new scheme resulted in a number of

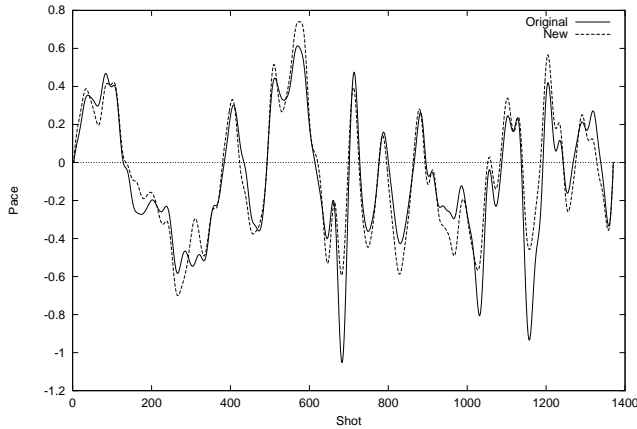


Figure 4: Comparison of shot normalization schemes with LW2.

useful edges emerging (or, in some cases, being made more pronounced). As an example, it can be seen that the section between the end of the fight at the crooks house (edge A) and the start of the ensuing car chase (edge B) has been more accurately resolved in the bottom plot. This is a result of the increased sensitivity to the rise in shot length just above the median. Previously the large amount of camera motion in that area caused the tempo of the linking shots to be smoothed, resulting in no clear edges, as seen in the top plot of Figure 5.

Edges Gained	
A	End of fight, crooks house
B	Start of car chase
C	Boss arrives
D	Police leave crooks house
E	Riggs left with Murtaugh on bomb

Table 2: LW2 - New edges and corresponding story sections.

5. CONCLUSION

Taking the conventions of film grammar as the key to unlocking the semantic door of film, we have focused on extracting the acknowledged expressive element of tempo or pace. Using the fundamental components of tempo, shot length and motion, we have constructed a continuous measure that has been demonstrated as a useful tool in its own right, and a promising component of more complex semantically meaningful constructs. In addition to this, a new shot normalization scheme was developed and tested. This was shown to offer improvements to the original tempo measure and serves to highlight the value of considering the processes that lead to film creation in greater detail.

6. REFERENCES

[1] Mediaware solutions webflix pro v1.5.3. <http://www.mediaware.com.au/webflix.html>, 1999.

[2] NIST dataplot software, 1999.

[3] 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, NY, USA, July 2000.

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

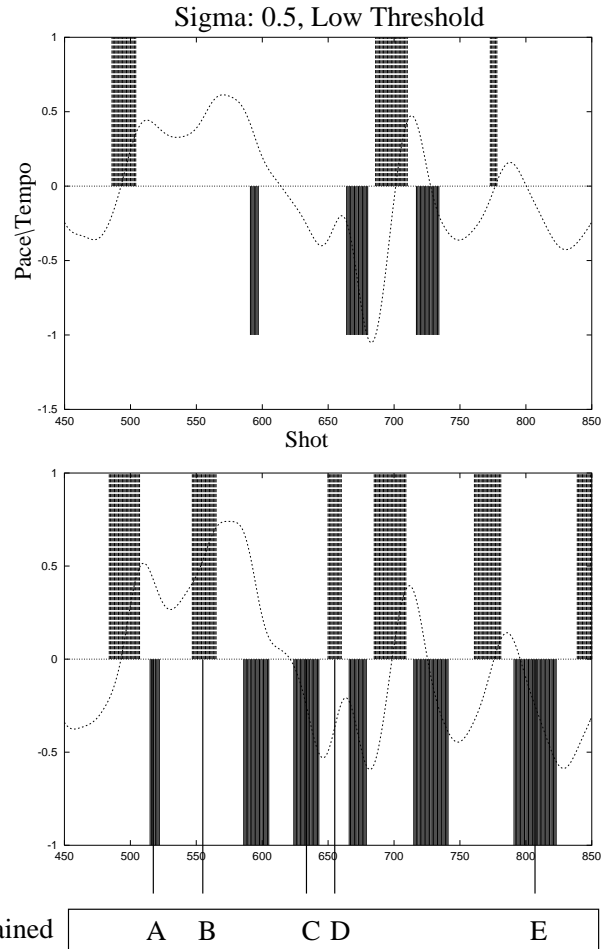


Figure 5: LW2 - Top: original pace plot. Bottom: pace plot and edges gained with the new shot normalization scheme.

[5] M. Brandt. Traditional film editing vs. electronic nonlinear film editing: A comparison of feature films. <http://www.nonlinear3.com/brandt.htm>, 1998.

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

[7] R. Deriche. Recursively implementing the Gaussian and its derivatives. In *ICIP'92, Proc. 2nd Singapore Int. Conf. on Image Processing*, pages 263–267, 1992.

[8] W. Mahdi, L. Chen, and D. Fontaine. Improving the spatial-temporal clue based segmentation by the use of rhythm. In *Second European Conference, ECDL '98*, 1998.

[9] J. Monaco. *How to read a film: The Art, Technology, Language, History and Theory of Film and Media*. Oxford University Press, 1981.

[10] R. L. S. Pfeiffer and W. Effelsberg. Video abstracting. *Communications of the ACM*, 40(12):54–63, 1997.

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

[12] N. Vasconcelos and A. Lippman. A Bayesian video modeling framework for shot segmentation and content characterization. In *CVPR'97, San Juan, Puerto Rico*, 1997.

[13] A. Yoshitaka, T. Ishii, M. Hirakawa, and T. Ichikawa. Content-based retrieval of video data by the grammar of film. In *IEEE Symposium on Visual Languages, Capri, Italy*, 1997.

[14] H. Zettl. *Sight, Sound, Motion: Applied Media Aesthetics*. Wadsworth Pub Co., 1973.