Finding the Beat: An Analysis of the Rhythmic Elements of 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

This paper forms a continuation of our work focused on exploiting film grammar for the task of automated film understanding. We examine film rhythm, a powerful narrative concept used to endow structure and form to the film compositionally and to enhance its lyrical quality experientially. Of the many, often complex, cinematic devices contributing to film rhythm, this paper investigates the rhythmic elements that are present in edited sequences of shots, and presents a novel computational model to detect shot structural rhythm as either metric, accelerated, decelerated, *or* free. *Details of the algorithm for the extraction of these editing rhythm classes are presented, along with experimental results on real movie data. Following this we study the usefulness of combining the rhythmic patterns induced through both motion and editing in film. We show that, whilst detailed content identification via rhythm types alone is not possible by virtue of the fact that film is not codified to this level in terms of rhythmic elements, analysis of the combined motion/shot rhythm can allow us to determine that the content has changed and hypothesize as to why this is so. We present 3 such categories of change and demonstrate their efficacy for capturing useful film elements (e.g., scene change precipitated by plot event), by providing data support from 5 motion pictures.*

1 Introduction

Successful automatic analysis of film requires a careful consideration of the forces that have shaped it. There exists a large body of literature detailing the rules and conventions that apply to film production in general. Termed *film grammar*, it forms the filter through which any automated attempt at film understanding must pass. It is the interpretive grid that guides the analysis and defines the meaning of the structures discovered in the raw data.

In previous work, we have outlined how a systematic use of the body of literature concerning film production from all of its aspects can be exploited when building tools for the automatic understanding of films [3, 2, 1, 4]. This level of careful processing and inspection is vital to treating the raw data of film with integrity, as opposed to bringing a false set of interpretive rules to bear upon it.

This paper forms a continuation to our very recent work on **film rhythm** [4], in which we outlined the first algorithmic study of *motion rhythm*, one induced by the manipulation of movements in film (the interested reader is directed to [4] for a detailed discussion). In this paper, we investigate the second important contributor to film rhythm, namely *editing*. We study rhythmic elements that can be extracted from edited sequences of shot (duration) patterns. Based on film grammar, four categories of shot pattern are identified. We present an algorithm for the automatic extraction of these rhythmic elements and show the results of applying this algorithm to motion pictures. We then investigate the usefulness of combining the rhythmic elements of motion and editing for content description. Although a fine grain identification of content proves to be unresolvable with these rhythm tools (as film content is not coded to a detailed level by these narrative concepts), coarse grain conjectures regarding structure and content *change* and causes for those changes can be made. We hypothesize 3 such categories of motion/shot rhythm configuration and content semantics, and demonstrate their validity by studying the data support across 5 movies.

The novelty of this work lies in the definition and extraction of the shot rhythms, the combined rhythmic element analysis involving both shot and motion rhythms, and the resultant hypotheses about high-level semantics of content. Although rhythm has been mentioned by an earlier research effort in the field of automatic understanding of film (e.g. [9]), attempts to extract it in any way have not been forthcoming. Our work, is the first attempt at a systematic algorithmic study of film editing rhythm and at its connections to scene content and story narration. Mining of such structures is essential to truly automated film understanding, and consequently to content management/retrieval systems based on techniques like those presented in this paper, as it is pertinent to the "trained" and "untrained" audience alike[1].

2 Film Rhythm

Film rhythm forms an integral part of film grammar. In its most generic definition, rhythm is an "organization of time" ([10, p. 90]. Film, being both a construction, and firmly tied to a timetable in at least one sense (i.e., running time, see [12, p. 246] for a list of *times* present in a film), naturally gives rise to a rhythm.

While the term rhythm naturally causes one to think of musical rhythm, generally only the most generic concepts may be transferred to film. Like musical rhythm, film is structured in time. It has flow and discontinuity, "short beats and long" as it were, but unlike musical rhythm, such relationships are not found in strict time. Indeed Mitry goes as far as saying "there is therefore no similarity between film rhythm and musical rhythm involving relationships of proportion and recurring patterns"([10, p. 271]). With film the methods of expression and corresponding order of dimension are much expanded (image composition, sound/music, editing, motion all giving rise to rhythm), and consequently all the more elusive. Often the rhythm of a film is not uppermost in a filmmaker's mind for its duration (to the degree perhaps that tempo is [3]), nevertheless the decisions to place this shot before that, adjust shot lengths in relation to one another, or use this framing type over that, affect what is later perceived to be a film's rhythm by its audience.

Many authors recognise film rhythm to be a complex entity. Bordwell and Thompson [6] state that "the issue of rhythm in cinema is enormously complex and still not well understood", however, they do go on to say that it roughly involves "a beat or pulse, a pace, and a pattern of accents, or stronger and weaker beats". [6] lists "movement in the mise-en-scene, camera position and movement, the rhythm of sound" as well as *editing*, as contributors to overall rhythm. With our investigation of motion rhythm complete [4], the next candidate from this complex array of contributors that offers itself as a computable entity is that of *editing devices* (i.e., shot patterns).

3 Editing Rhythms

Shot length has a dominant part to play in the organization of a film into time and hence its rhythmic properties. A shot forms an identifiable piece of a film, and the movement from shot to shot has certain demands and effects on the viewer. [6] notes that "the patterning of shot lengths contributes considerably to what we intuitively recognize as a film's rhythm."

[1] Salvaggio [11, p. 137] notes that "[The spectator] may not be able to formulate the rules, but he recognizes an ungrammatical sequence when he sees one." - i.e., You don't have to be a film student to tacitly recognize film structures.

What sort of patterns are we talking about? While shot lengths can obviously be crafted to as many patterns as the human imagination allows, some broad fundamental categories may be defined (upon which more complicated patterns may be built):

1. *Metrical* - a series of shots with very similar shot lengths, providing a constant tempo, or *visual percussion*, often subordinates shot content to the rhythmic expression. (Zettl, p. 292 "A rhythmic structuring device")

2. *Accelerated/Attack* - a series of shots providing a steadily rising tempo, again, this narrative device often subordinates shot content to the rhythmic aims.

3. *Decelerated/Decay* - a slowing shot rate, often used to deaccentuate an event.

4. *Free* - used generally for the largest portion of a movie. It is in this category that the rhythm is perhaps – of the highest degree – secondary. That is, it arises as a function of the placement of shots (and lengths) rather than strictly directing their placement (which is more the case for categories 1-3). See Mitry [10, p. 273] for further thoughts. Free is, by definition, the absence of the other categories.

3.1 Technique to Extract Shot Rhythms

The following process is used to identify these devices (*device* from here on refers to an instance of one of the identified shot length patterns and hence rhythms they give rise to) for a given movie. A full-length movie is digitized and first segmented into shots using WebFlix [8]. A shot transition index is obtained, and from this a series of shot lengths for the movie. The lengths are then smoothed once with a Gaussian ($\sigma = 1.5$). The smoothing is used to avoid complicated classification rules that must deal simultaneously with possibly noisy shot length data *and* the natural tolerances of the rhythm classes. Note that large levels of simple smoothing are not generally justified for shot length however, due to the complicated nature of its underlying distribution, i.e., there is not a simple 1 or 2 order curve that can be rebuilt in the majority of cases (see [1] for a comprehensive discussion of shot length distribution).

A sliding window (of varying sizes from 3 to 11) is then applied to the smoothed lengths to determine the nature of each shot neighbourhood, on a shot by shot basis. The data within the window has the following tests applied to it, and a device classification results:

1. The *metric* rhythmic device generally has the characteristic of a series of (often shorter) shots, designed to create an effect of visual percussion. To have such an effect the length of the shots in question generally has to be close enough to each other for the viewer to perceive them as such. This physical limitation puts some soft upper and lower limits on the problem. The limits have a small degree of flexibility to them due to the fact that audiences will accommodate a small enough difference into the de-

vice to "make it fit" (see [10, p. 271] quoting Paul Fraisse). Added to this is the fact that during the course of such a shot series it often becomes necessary to insert a shot whose length must be dictated by the content. It is possible to do this (within bounds) without destroying the overall rhythmic movement of the series (e.g., the best way of portraying a particular occurrence in a fight scene, say the dropping of a gun, in the circumstances may require the insertion of a longer shot into a flurry of shots elucidating the event. The sequence should still be identified as one metric sequence).

Classification: The lengths in the window under consideration is adjudged a *metrical* series if all of the shot lengths are within 10% of the median of the series. It is a matter of further research as to whether an absolute value of variation from the median obtained from a consideration of human perceptual behaviour would be better here than a proportion. A threshold of 3 seconds is also applied to this series median to prevent series of very long shots of equal length from being classified as a metrical series, as the percussive effect becomes decidedly weaker at these lengths (such a rhythm might still be important for other reasons however). This generally has to do with our decreasing ability to assimilate long events into rhythmic forms perceptually, although intellectually reconstruction may be possible.

2. The *accelerated/attack* rhythmic device is often used to lead into a metrical one and is similar in use in that it often subordinates shot content to rhythmic needs. It is also often used as a leading device to accentuate a dramatic occurrence. A rising shot rate demands more from the viewer and correspondingly engages him or her more forcefully.

Classification: This simply requires that a series be monotonically decreasing in length to result in an accelerated rhythm. In addition to this outliers will be accepted up to the specified tolerance.

3. While shot *deceleration* does not seem to be crafted to the extent of the *attack*, it nevertheless retains the ability to perform the opposite rhythmic function to the *acceleration* and will therefore be targeted as such.

Classification: It requires that a series be monotonically increasing in length to result in an decelerated rhythm. In addition, outliers will be accepted up to certain tolerance.

4. A shot series not classified as one of the above is given the label *free*. This component is generally the largest for the majority of films. This rhythm is much more a natural result of the shot organization, than the driving force behind the placement and length of shots. Many factors affect the length of a shot, and result in a wide range of shot durations. Such sections, however, do have a rhythm all of their own, and thus are worthy of classification as free.

An example, including all rhythm categories, can be seen in Figure 1 from "Starwars Episode 1". For this example, a Gaussian smoothing function with $\sigma = 1.5$ was employed, along with a size 9 sliding window and $tolerance = 0$ (see Section 3.2 for explanations of the parameters). The scene studied sees the two Jedis attempting to break through the door to confront the cowering Viceroy. The metrical series coincides with the back and forth cutting showing alternatively Quigon making progress on the door, and the shocked reaction of the Viceroy. The evenly timed shots serve to produce a rhythmically percussive effect. This series gives way to a attack rhythm as the droid "Destroyers" roll onto the scene, with a corresponding conflict escalation. Deciding that it is a standoff, the Jedis flee with a corresponding decay rhythm which serves to deaccentuate the event. The rhythm of the unmarked sections is computed to be free.

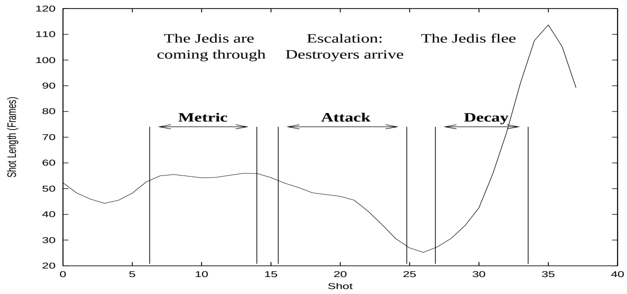

Figure 1. An example of different rhythm types in a scene from Starwars Episode 1.

3.2 Rhythm Classification with Noisy Shot Data

Since there is no perfect automatic shot detection algorithm available to determine precise shot transition index for videos, our algorithm has to deal with missing shots, false positive transitions, inaccurate shot boundaries, etc. Therefore, we experimentally verified the performance of our rhythm classifier on noisy shot data computed from Hollywood movies using WebFlix.

Table 1 reports the results of the experiments with our rhythm classifier using a number of different parameter configurations. The experimental data is drawn from the movies, *The Matrix* and *The Mummy*, with the aim to determine the best configuration for the classifier in the presence of noisy shot indices. Configurations that returned an avalanche of false positive rhythm labels are simply noted as *saturated* (sat.), while those that returned under 5 instances are noted as having *negligible* (neg.) return. The *Precision* is the ratio of correct classification in the three categories, Metrical, Attack, and Decay to the total number of classifications produced for these categories.

The effect of each of the three parameters used by the classifier should be noted.

Window size: The smaller the sliding window size the smaller the devices (rhythmic series) that may be located, resulting in a detailed analysis. However, if too small, the result is a flood of meaningless classification, and too large, the return is too sparse.

Smoothing: This parameter effectively takes the classifier away from the raw data as its value is increased. This is undesirable for our purposes, and therefore this parameter has only been set to *none* or $\sigma = 1.5$ in our experiments. The effect of higher smoothing levels is to target the classifier at larger *trends*, rather than the smaller crafted shot patterns that we desire.

Tolerance: The tolerance parameter is particularly useful in the unsmoothed case. A tolerance of one shot, for example, allows a shot with a non-conforming rhythm label to be part of the larger series such that its presence does not destroy the effect of the device detected for the series.

Table 1. Results of shot rhythm (Metric, Attack, Decay) classification on movie data.

Movie	Wind. Size	Smooth. Level	Toler. (shots)	Precision
Matrix	3	-	0	sat.
	5	-	0	35/44 (79%)
	5	-	1	sat.
	7	-	0	neg.
	7	-	1	22/28 (79%)
	7	-	2	sat.
	7	$\sigma = 1.5$	0	sat.
	9	-	1	neg.
	9	$\sigma = 1.5$	0	68/93 (73%)
	11	$\sigma = 1.5$	0	29/42 (69%)
Mummy	3	-	0	sat.
	5	-	0	28/32 (87%)
	5	-	1	sat.
	7	-	0	neg.
	7	-	1	22/25 (88%)
	7	-	2	sat.
	7	$\sigma = 1.5$	0	sat.
	9	-	1	neg.
	9	$\sigma = 1.5$	0	54/74 (73%)
	11	$\sigma = 1.5$	0	30/42 (71%)

Table 1 can be summarized as follows: A window size of 9, $\sigma = 1.5$ and $tolerance = 0$, returns an acceptable precision (73% with both movies). The number of shot devices detected is reasonable in detail (93 for Matrix, 74 for Mummy). Increasing the window size decreases the total number of shot devices found (to 42 for both movies). Decreasing the window size while keeping $\sigma = 1.5$ and $tolerance = 0$, results in saturation (too many false positives) in both cases. Decreasing the window size and reducing the level of smoothing does result in acceptable precision, but the number of devices detected decreases dramatically (with $windowsize = 7$, to 28 in Matrix and 25 in Mummy). The rhythm miss rates to complete Table 1 will be reported when available.

There are instances for each of the classifier configurations listed in Table 1 where one configuration is able to locate devices that another misses. This suggests that there may be advantages to a multi-mode analysis followed by consolidation (removal of duplicates etc.). For the remainder of this paper, however, we will confine the rhythm classifier to a size 9 window, with Gaussian ($\sigma = 1.5$) smoothing and zero tolerance, as this is seen to yield acceptable precision with reasonable number of detected rhythms. While offering slightly lower resolution than the unsmoothed configurations, it was found to salvage many of the smaller devices and provide robustness against a noisy shot index in many instances.

3.3 Discussion

For different parts of a movie, typically, a director/editor is more or less concerned with crafting a shot to a very specific length. More often than not the first two classes, *metrical* and *attack* (and the third, *decay* to a lesser degree) receive much more attention in this respect. This is due to the fact that usually their employment is precisely due to their rhythmic effects, i.e., rhythm is the primary consideration. For the *free* class however, the content dictates the shot length. Shot rhythm in this situation is, as noted, in a sense secondary, arising from the placement of shots and shot lengths. As a result this category shows the greatest degree of variation, and is correspondingly the hardest to draw conclusions from regarding possible content.

For classes *metrical, attack, and decay*, identifying where these devices are employed is useful in that it they are clear sign posts that the director is creating or reinforcing something by their use. For class *free*, at the very least, sections of this type say something by the absence of the other classes. Also, if different content gives rise to different shot patterns (whatever they may be) then sections of free shot rhythm with markedly different characteristics may indicate a change of content, and be a guide as to where to apply more computationally rigorous tests to determine the nature of that content (e.g., an establishing shot may indicate an overarching colour scheme/possibilities).

The simple rhythm classifier might be improved by additional criteria. Outliers could have constraints placed on them such as a maximum deviation from the ramp under consideration. For example, an outlier that exceeds the maximum length in a ramping series (and therefore probably destroying the rhythmic shot device) could be designated a miss. Another criterion might be a minimum drop or increase in the length of consecutive shots in a ramping series. A prime candidate here might be the minimum difference that a human viewer will notice. Both of these possible criteria would have been of use in the analysis and improvement of Table 1.

4 Visual Motion Rhythms

A full discussion of the extraction of rhythmic patterns induced through a perception of motion can be found in [4]. A brief review is necessary here for understanding Section 5, the combined shot and motion rhythm analysis.

Classifying Motion Behaviour of a Shot: Bordwell and Thompson [6] mention that "frame mobility involves time as well as space, and filmmakers have realized that our sense of duration and *rhythm* is affected by the mobile frame". Prompted by [6] who note that "a camera movement can be fluid, staccato, hesitant, and so forth", we undertook to automatically categorize the motion of each shot in a given film as either non-existent, fluid or staccato.

Qualitatively the classes may be defined in the following manner: *No Motion*: No appreciable camera or object movement; *Fluid*: Constant or smoothly changing speed; *Staccato*: Containing abrupt speed changes.

A shot classifier was then implemented to label shots using smoothed shot motion values, with a two stage process. First a given shot is deemed to either contain motion or not. This is achieved by thresholding the average motion magnitude for the entire shot. A shot with motion falling below the threshold is deemed *No Motion*.

The second stage involves classifying the remaining shots as either *Fluid* or *Staccato*. The essential characteristic employed is the average of the first derivative of motion magnitude, calculated for the sections of the given shot that contain motion and this is normalised for both shot length and motion duration. The more abrupt the motion changes within the shot, the higher this value will be, and vice versa. The final step is to threshold this value, with values below the threshold producing a *Fluid* classification, and those above *Staccato*.

In order to test this classifier, ground truth was constructed for sections from a number of movies. While this process demonstrated the difficulties inherent to the classification scheme, and the somewhat spectral nature of the fluid/staccato division, we found that each shot could be manually classified with a good level of confidence. The automatic shot motion classifier performed well against the ground truth, with correct classification of 85% plus of *No Motion* shots, and 75-85% of the motion classes.

Detecting Motion Rhythms from Shot Arrangements: Since rhythm is formed from the arrangement of a number of shots, the next step was to classify a *neighbourhood* of shots. We chose to classify the (sliding) neighbourhood centered on a shot by the percentage makeup of shot motion behaviour types. The following classes were defined:

1. Predominantly No Motion (NM)
2. Predominantly Fluid (F)
3. Predominantly Staccato (S)
4. Mixed, No Motion and Fluid (NM/F)
5. Mixed, No Motion and Staccato (NM/S)
6. Mixed, Fluid and Staccato (F/S)
7. Equally mixed, No Motion, Fluid and Staccato (*All*)

We have applied this process to over 20 movies so far in their entirety, and the resulting rhythm classification has been observed as being useful for finding sequence changes, particularly if the neighbourhood size used is large (i.e., targeted at large trends). The size of the sliding window was set to 41 shots and it was found to be about the order of the "clumps" of shots of interest, i.e., scenes. This value merely affects the level of detail for analysis.

5 Content Semantics and Combined Rhythm Types

Given the classification of the two rhythmic elements, editing and shot motion, are there any insights that can be gained from considering them together? Do they interact with one another to add a new rhythmic dimension, or are they essentially different modes of the same?

One crucial question is whether any new light is shed upon the content of a piece of film when the two rhythmic modes are considered together. For example, does a certain motion behaviour rhythm combined with a certain shot rhythm device indicate a flashback sequence? Our study indicates that in the general case this degree of content identification is not possible but that important contributions are still produced by considering both rhythmic modes together.

Our research suggests that while some inferences about the content of sections of different rhythmic characteristics (in the terms described in this paper) can be made in general, what can be determined is that the content has *changed*, rather than what that content is. If a director employs an accelerated montage then he is reinforcing or creating *something*, but in general we don't know what sort of an event it is. However, the ability to detect changes in the rhythm of a film is, in and of itself, a useful ability [5, p. 233]. Our work highlights where these changes or surprises occur and how they are interesting in and of themselves, without necessarily requiring that the content at the time of change be identified. Absolute classification of such a complex entity as rhythm in simple labels and semantic terms is a daunting, if not impossible task, but detecting changes to the elements that constitute it has been successfully demonstrated in this paper in part.

Crittenden notes that "No shot is neutral, both its content and its form already contain the elements which contribute to the innate rhythm that reacts upon any juxtaposition we create" [7, p. 76]. Therefore, different content often gives rise to different rhythmic characteristics (e.g., dialogue vs. action scene) and finding where rhythmic elements change is equivalent to finding where some part or all nature of the depicted content has changed. Locating content change, and additionally the nature or cause of that change, would be a valuable addition to any simple or composite index in the domain of content management or retrieval.

As a demonstration of our proposed dual mode analysis of motion and editing rhythms, consider Figure 2.

Figure 2 depicts two of the three proposed combinations

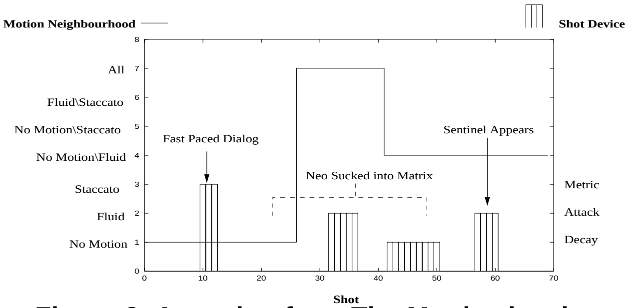

Figure 2. A section from The Matrix showing editing rhythms and motion rhythms for combined analysis.

of editing and motion rhythms. As an example of what situations give rise to the different categories:

Category 1: Shot device found, without Motion Neighbourhood (motion rhythm) change, in the vicinity of shots 10, Metrical and 60, Attack. All of these devices arise within the same film setting as a result of an event. The Metrical device coincides with a fast paced dialog following Neo's decision to take the red pill. The Attack comes as the Matrix drone discovers Neo, and the ensuing struggle.

Incidentally, the metric montage combined with motion neighbourhood class 1 (No Motion) is often the result of back and forth dialog, as is the case at shot 10.

Category 2: The other two shot devices found in the figure coincide with a change in Motion Neighbourhood Class (actually are within a small neighbourhood of that change). The Attack rhythm at shot 31 and the corresponding Decay at shot 41 occur across as Neo is about to be sucked into the Matrix, and the change of scene respectively. Such a configuration often occurs where the filmmaker is trying to reinforce the change, or where the scene change is precipitated by an event.

Category 3: This category, not shown in the figure, is Motion Neighbourhood Class change without Shot Device and it often happens with the change to a different locale/situation abruptly. The motion rhythm change is due to the new scene/situation characteristics, and the shot profile changes so abruptly also that no ramping device is employed (this is sometimes used to disorient the viewer).

The fourth possible category where there is neither a motion rhythm change nor a planned shot device is not considered here. Table 2 summarizes the above observations enabling a dual mode analysis.

In order to investigate the validity and coverage of these combined rhythm modes we conducted the following experiment. Five mainstream movies, namely The Colour Purple, The Matrix, The Mummy, Starwars Episode 1 and Titanic, were used as data. With each movie and its shot data the following procedure was performed:

1. Shot rhythm is classified as Attack/Decay/Metric using a window size of 9, $\sigma = 1.5$, and tolerance=0.

Table 2. Shot device and motion rhythm configurations with typical causes in the proposed combined analysis.

Cat.	Shot Device Found	Motion Rhythm Change	Probable Cause
1	Attack, Decay, Metric	No	Same film setting, as a result of event
2	Attack, Decay, Metric	Yes	(i) Reinforce change, (ii) Scene change precipitated by event
3	Free	Yes	Change to different locale, scene, sequence abruptly

2. Motion rhythm is classified using the automatic process defined in Section 4, window size = 31. Locations where this rhythm changes are noted.

3. For each shot device found by step 1, a neighbourhood (called *search scope*, = 5 shots) around the shot device is searched for motion rhythm changes (located in step 2). If a motion rhythm change is found within the search scope the shot receives a classification of *category 2*, if not the classification given is *category 1*.

4. A second pass is performed that gives a classification of *category 3* to those motion rhythm changes that have not already resulted in a classification of category 2 in step 3.

5. For each classification automatically obtained by the procedure so far, a ground truth classification is manually obtained from the movie itself and is recorded. Where the actual movie content does not accord with any of the 3 possible categories (i.e., a film occurrence not defined in the 3 combined categories), it is marked as *Other*.

6. A comparison is made between the classification produced by this automatic procedure, and the manual classification. The results can be seen in Table 3. The figures given are the instances of correct classification over the number of classifications produced for each category.

Table 3. Combined rhythm analysis: Classification results of categories.

Movie	Cat. 1	Cat. 2	Cat. 3	Overall (%)
C. Purple	13/19	13/16	9/9	35/44 (80%)
Matrix	53/57	25/31	17/32	95/120 (79%)
Mummy	48/51	18/20	15/19	81/90 (90%)
Starwars	54/59	17/29	16/28	87/116 (75%)
Titanic	63/74	23/37	12/22	98/133 (74%)
Totals	231/260 (89%)	96/133 (72%)	69/110 (63%)	396/503 (79%)

The results reported focused only on the accuracy of our combination classifier and its miss rates will be forthcoming. Let us examine the results for each of the 3 categories. Although, in the following discussion, a single problem is focused on for each category, all categories are generally affected by these problems to one degree or another.

Category 1 performed well for all movies, with the exception of Colour Purple (68%). For most cases the shot rhythm was recorded correctly. In one case, the problem was that the device was used to transition to a new scene, and in a second the device occurred during a sequence that spanned a number of locations (i.e., in violation of the stated category 1 hypothesis). In the first case, The Colour Purple, being a fairly static movie in general, did not show pronounced motion neighbourhood change and the new scene had very similar motion rhythm. In the second case, this might rightly be called an event within a *sequence*, as *location* is overly precise.

For category 2 the principal problem was the resolution of the motion neighbourhood window. While a size 31 window targets large trends successfully, the resolution regarding exactly where motion rhythms change is accordingly poor. This resulted in many shot devices falling just outside the search scope, when in actual fact a more precisely located motion neighbourhood edge (change) would have caused them to be correctly associated with that edge. A motion neighbourhood finding algorithm with improved edge precision will greatly improve the results for this category. In other words it is a problem with the underlying motion rhythm precision, not the hypothesis in question.

A common cause of misclassification for category 3 was small shot devices. A number of shot devices were no more than 6 or 7 shots long, and as such were missed by the size 9 classifier. A multi-scale analysis, as mentioned in Section 3.2 might help alleviate this problem by providing these short shot devices. Again, this is a problem with the underlying shot device misses, not the hypothesis in question.

In addition to these causes, the “problem” of sheer cinematic creativity is not to be neglected. Often a filmmaker will use a non-conventional set of techniques to convey an idea. If this were the norm across all techniques and conventions the problem of automatic analysis would indeed be completely impossible. As it is, the tenets of film grammar exist because this is not the usual case, and our task remains possible, if not easy.

6 Conclusion

Starting with the film literature concerning rhythm in film, we noted its complexity when taken as a whole and sought to focus our analysis to that of the rhythmic elements expressed via editing (shot length) and motion. The first identified rhythmic element that was extracted was that of shot patterning due to editing. Four categories of shot pattern were derived from the film literature and defined, including Metric, Attack, Decay and Free. Computational tools for their extraction were developed and demonstrated.

The second rhythmic element discussed was that of shot motion. Tools for classifying a given shot as either without Motion, Fluid, or Staccato were developed. A natural extension to this work was the classification of shot neighbourhood arrangements, and the resulting neighbourhood motion classes proved to be useful in that they are reflective of certain film content categories.

A final investigation into the usefulness of combining the results of both analyses was conducted. The outcome proved to be that more detailed codification of content does not become available with such a dual mode analysis. However, the rhythmic elements do combine to reinforce each other and offer pointers to the nature of the rhythm change (e.g. due to a localised film event or not). Given the ability to find sections of likely content, the tools developed here could be applied to the task of identifying narrative structures and dramatic progression.

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 2000*, pages 54–57, Sydney, Australia, December 2000.

[2] B. Adams, C. Dorai, and S. Venkatesh. Study of shot length and motion as contributing factors to movie tempo. In *8th ACM International Conference on Multimedia*, pages 353–355, Los Angeles, California, November 2000.

[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*, volume II, pages 641–645, New York City, USA, July 2000.

[4] B. Adams, C. Dorai, and S. Venkatesh. Automated film rhythm extraction for scene analysis. In *IEEE International Conference on Multimedia and Expo*, Tokyo, Japan, August 2001.

[5] R. A. Block. *Cognitive Models of Psychological Time*. Lawrence Erlbaum Associates, Publishers, 1990.

[6] D. Bordwell and K. Thompson. *Film Art, 5th Ed.* McGraw-Hill, 1997.

[7] R. Crittenden. *The Thomas and Hudson Manual of Film Editing*. Thomas and Hudson, 1981.

[8] M. S. P. Ltd. Webflix Pro. http://www.mediaware.com.au/webflix.html, 1998, 1999.

[9] 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.

[10] J. Mitry. *The Aesthetics and Psychology of the Cinema*. The Athlone Press, London, 1998.

[11] J. L. Salvaggio. *A Theory of Film Language*. Arno Press, 1980.

[12] H. Zettl. *Sight Sound Motion Applied Media Aesthetics*. Wadsworth Pub Co., 1998.