IMCE: INTEGRATED MEDIA CREATION ENVIRONMENT

Brett Adams[†], *Svetha Venkatesh*[†], *Ramesh Jain*[‡]

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

School of Electrical and Computer Engineering[‡]
and College of Computing
Georgia Tech, Atlanta, GA 30332-0250
jain@ece.gatech.edu

ABSTRACT

We discuss the design and implementation of an integrated media creation environment, and demonstrate its efficacy in the generation of two simple home movies. The significance for the average user seeking to create home movies lies in the flexible and automatic application of film principles to the task, removal of tedious low-level editing by means of wellformed media transformations in terms of high-level film constructs (e.g. tempo), and content repurposing powered by those same transformations added to the rich semantic information maintained at each phase of the process.

1. INTRODUCTION

We will outline a computationally-assisted, that is semi-automatic, media creation framework, aimed at helping the average user to build media artifacts that faithfully communicate their intent whilst harnessing the full expressive powers of the medium chosen.

We address the chief obstacles facing the collection, generation and presentation of quality media by the average user, namely lack of willingness to invest time in learning the expressive properties of the chosen media and applying that knowledge in polishing the artifact under construction, and the informal nature of the context the amateur often works within. This lends our media creation framework the following significance:

- automated application of domain specific knowledge,
- flexibility to user knowledge and context,
- inherent editing and transformation of media, and
- an accordingly wider scope for media reuse/repurposing.

We will discuss the design motivations for the components of our architecture and demonstrate with two small movies generated by the implemented system.

2. MOTIVATION

We are interested in helping people create better media more easily. Typical scenarios that bring about the creation of new home movies[1] might be a proud Dad seeking to capture something of his daughter's birthday party, or a vacationing couple desiring to compile a manifest of their trip. Usually there is little thought put into what will be filmed and how, the person attempting the capture has only a moderate (largely intuitive) grasp of the film making craft, and the final result is often far from what it could be due to the time and effort required to edit the captured footage.

Motivated by common experience and also from the explicit recommendations of film-makers themselves, [1, p. 16][2, p. 60], we use narrative principles to prime and direct our video collection environment. *Story* provides the *why* that drives the *what* and *how* to capture.

There exists a well known and simpler parallel to the 'video collection' environment we are seeking here: integrated development environments for software creation. IDEs support the software creation process by providing automated indenting, context sensitive help (such as method expansions) and even rule checking. Analogously, our video collection environment enables the user to appraise and identify what they have collected and where to put new material, alerts the user to capture options and their communicative impact on the presentation and viewer in turn, and provides checking for well-formedness, which is defined in the terms of cinematic convention and practice (i.e. film grammar).

For related work, see [3, 4, 5].

3. MEDIA CREATION FRAMEWORK

If the two key factors identified above as being detrimental to home video quality, namely lack of coherency in the content, and lack of adequate expression and affective (and effective) reinforcement of that content, were routinely addressed, the process would look like Figure 1. The labels at the top of each arrow can be thought of as *roles*, with corresponding rules for transforming the media, while the objects below could be termed *deliverables*.

There is nothing new in this view of movie creation – if we were talking about a Hollywood feature film that is.

[1] Here the term *home movie* embraces more than just 'family' content, having more to do with the amateur nature of the capture.

0-7803-8603-5/04/$20.00 ©2004 IEEE

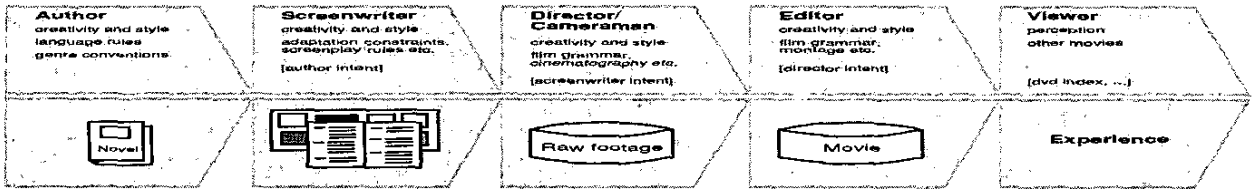

Fig. 1. Professional workflow for media collection.

What *is* new, is the application of these concepts to home video, which normally looks like Figure 2:

Fig. 2. Current amateur workflow for media collection.

And in fact due to the complex and expert nature of the edit stage even this is sometimes avoided, and thus the raw footage is promoted directly to "movie" status, a title it rarely deserves to hold. Ofcourse the reason that home movie creation normally looks like the latter is because of the effort and expertise required to do the former. The amateur movie maker is required to take on each of the roles indicated: author, screenwriter, director, and editor. But it is precisely that effort and expertise which we are seeking to inject by computational means. The process for home movie creation we're after, then, looks more like Figure 3.

What do we need to achieve our twin goals?:

1. Creation of *better* media, where media is potentially movie, still images, audio, presentation, novel, or other combinations thereof, etc.,
2. Creation of better media by the *average user*, who has a variable degree of familiarity with the medium in question, and possibly less opportunity or incentive to learn, i.e. we are targeted at the non-expert,

Let us explore the requirements that derive from each of these abstract goals in turn, and offer some concrete examples relevant to our video collection framework.

3.1. Better media

The goal of "better media" requires: 1. A *representation* for expressing the desired media (i.e. a directive telling the user to capture media "like this"), which will naturally be expressed in media specific structures, manipulations etc., and 2. A way of *transforming* a story and creative intent into statements in that media dependent representation, and further transformations from media to media, such that they are imbued with desirable properties.

What properties should the media specific *representation* have?

1-a It must be as domain specific as possible, because without precise 'terminology', appropriate representations and structures, we can only make imprecise generative 'statements' about the media. E.g., for the case of the movie domain, representations such as shots, scenes, and elements of cinematography, such as framing type and motion type, are appropriate, as they form the vocabulary of film production, theory and criticism.

1-b It will ideally provide a sound basis for inferring other, higher order, properties of the media, at differing resolutions, dependent upon whether more or less detail can be called for or reliably known. E.g., representations like shot and motion allow us to infer about the property 'film tempo,' of which they are constituents. Certain types of inference will require a degree of orderliness to the representations used, such as that imparted by means of hierarchical systems of classification among members - taxo/partonomies, ontologies. E.g. a partonomy of 3-act narrative structure allows us to infer that any dramatic events **within** the first act are likely to carry the dramatic function of setup, i.e. some aspect of the characters or situation is being introduced.

What properties should the media *transformers* have?

2-a Incorporate domain knowledge. The transformers must observe known domain rules. E.g. It may be possible to cut two pieces of footage together, in terms of the raw media manipulations required, but if in doing so a movie domain rule like "don't cross the eyeline" is violated, the transformation should be deemed illegal.

2-b The goal of a transformation is to imbue the media with a *desirable property*. The movie literature tells us that there should be variation in a movie's tempo, so there should be a transformer whose clearly defined scope and objective is to manipulate tempo to effect this result in the final movie.

2-c There is potentially much variation in those desirable properties that we wish to imbue the media with and hence the transformers should be *explicitly parameterizable*, effectively providing the ability to generate multiple 'views' of the collected media automatically. Continuing with the tempo example, if the desirable property is variation, that still leaves the question of how much variation, and of what sort? These should be able to be dictated to the process by means of parameters such as mean or standard deviation.

2-d Separate roles and make them explicit, so their progressive influence upon the media can be traced. Where possible, the effects of the transformers should be mutually exclusive; the tempo transformer should be the only agent manipulating elements of the shot directives (instructions to the

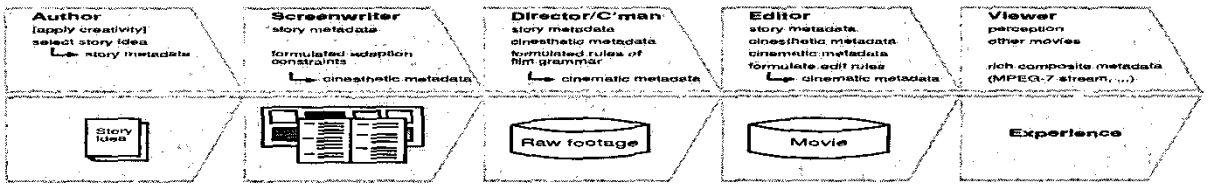

Fig. 3. New amateur workflow for media collection.

user about how to capture a desired shot) that effect tempo.

2-e Explicitly represent motivation for the transformation as part of evolving metadata. In the case where it is unavoidable that two transformers wish to manipulate the same shot directive elements toward their own goals (in contrast to the desired mutual exclusivity mentioned above), this provides necessary information for the resolution of conflict. In order to decide whether, say, a tempo transformer or a motion rhythm transformer should have precedence over the motion components of a shot directive, it is helpful to know *why* each has called for differing motion (pan vs. static). It may be that the tempo transformer is responding to local goals, whereas rhythm has more global goals in mind.

2-f Flexibility to cope with uncertainty in aquisition of media. Home movie footage capture is an inherently noisy process, and thus the process should allow for multiple attempts at a given shot directive, and selection of the best footage from among that captured for the shot directive.

3.2. Better media for the *average user*

What requirements does our focus on the non-expert media creator produce?

1. By default, it must put only a low burden on the user. In particular, the parts of this new process that are missing from the current amateur workflow, such as the story idea generation, must not require much effort: E.g. story templates may be selected from a library. Consequent stages of the workflow must have defaults that require no input, all the way to finished product.
2. The inputs and decisions in the workflow must be transparent. All cinematic directives must be traceable to the story and creative purpose chosen. The user should be able to select a shot directive and discover the hoped for impact of its parameters on, say, the tempo profile of the final presentation, and the reason for that target tempo signal.
3. 'Average' skill-level is variable, and therefore we need the ability to grade the media directives to an appropriate level for the current user. That is, the level of direction given to the user can be thresholdable. E.g. a novice might only want to know what to shoot and whether to use camera motion or not, whereas an expert or someone interested in improving their skills might want to know about the reasons for that tempo target mentioned above.
4. In addition to skill level, users exhibit varying levels of *desire* to input to the process. Thus, although there are default actions provided at all stages, there should remain the ability for the user to shape the work at any time. E.g., if the user thinks that a particular shot directive will be hard to capture in footage, he should be able to alter it.

4. ARCHITECTURE

We refer the reader to Figure 4 for an overview of the media creation framework for the domain of home movies. The familiar roles are listed at the top of the figure, and below them, the process has been further split into a number of stages (i.e. functional transformations). The boxes in the 'Directives' row are Storyboard-cum-Shootingscripts. The large circles indicate narrative events, while the smaller circles are shot directives belonging to their respective events (scenes). These concepts are further explained below.

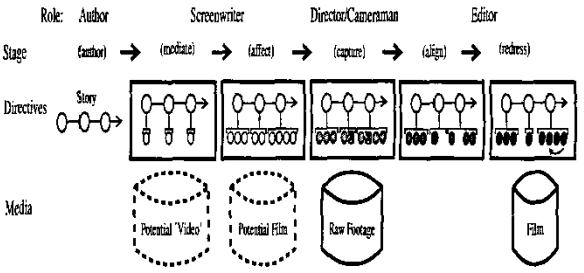

Fig. 4. Media creation framework for home movie domain.

Author: The purpose of the first stage on Figure 4, labelled "author," is to create the abstract, media non-specific story for the occasion (wedding, party, anything...) that is to be the object of the media creation project. That is to say, the given occasion separated into parts, selected and ordered, and thus made to form the content of a narrative. It culminates in a "narrative template," which is the deliverable passed to the next stage. In its simplest form, a narrative template may be represented as a sequence of plot events, where an event is an incident drawn from the raw source of the occasion by the author. The selection and placement of said events enables the author to highlight some incidents and ignore others, in short, to tell a story.

Mediate: The purpose of the Mediate stage is to "apply" or "specialize" the narrative template obtained in the Author stage to a specific media and domain. This stage encapsulates the knowledge required to manifest the abstract events of the plot in a concrete 'surface' manifestation – in

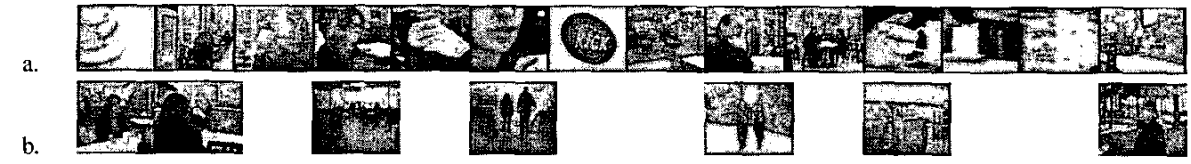

Fig. 5. "Morning Coffee" capture comparison: a. emotive/intense feel vs. b. intellectual/clarity

this case the pixels and sound waves of video.

Affect: The Affect stage transforms the initial media specific directives produced by the Mediate stage into directives that maintain correct or wellformed use of the given medium, in this case good technical and cinematic convention, better utilize the particular expressive properties of that medium, and this with reference to user goals for the final media. It does this with a user-set parameterization, *affect parameters*. We detail one of the parameters: Desired viewer *response*, which dictates whether we desire an emotive/intense response from our audience, or we wish to engage the intellect by means of clarity in the presentation? All parameters are mapped to a user-friendly genre label, enabling the user to dictate the creative purpose or style for the desired movie. E.g. do we want our holiday movie to feel more like a documentary or a Hollywood action flick?

Capture: The Capture stage "realizes" all media directives in actual media. For our domain of home movies this means capturing all the shot directives as film footage. A realized shot directive is simply a shot directive that has been (potentially) "instantiated" or augmented with actual media.

Align: The purpose of the Align stage is to check (where possible) whether the footage captured for a given shot directive has indeed been captured according to the directive, and in the case where more than the required duration has been captured, select the "optimal" footage according to the shot directive. Thus a section of footage of the desired duration is obtained that maximizes the match between shot directive and the actual footage. For any shot directive properties not satisfied in the final footage selection, their properties in the shot directive are altered to match what has been realized. The resulting realized shot directive thus *actually* accords with what has been captured, not what was originally desired (i.e. shot directives given by Affect the stage).

Redress: After the Align stage, we may have affect goals gone awry due to failures in actual capture. This stage attempts to achieve, or get closer to, the original affect goals with the realized shot directives. This is achieved by means of a distance function in affect space between the original target shot directives, and the realized shot sequence as transformed by an edit function which manipulates a shot sequence by means of insertions or deletions, and also manipulations of the realized shot directives themselves.

If you allow that the *affect parameters*, and consequent shot directives, be altered at this stage, you have the ability to explicitly generate multiple views of the generated media: Would you like the Hollywood or documentary version?

5. DEMONSTRATION

In order to demonstrate the implemented framework from beginning to end, we have used it to create two small films, two movie versions of the same narrative, about the universal ritual of "The Morning Coffee." We first constructed a simple narrative template around the idea of the two caffeine addicts heeding the call to obtain their morning fix.

We present the narrative template to the system twice, with differing *affect parameters*, one with an emphasis on an emotive/intense response from the audience and the other seeking to maintain a higher level of clarity. Figure 5 presents about 40 seconds of footage from both of the automatically compiled versions of Morning Coffee.

Two aspects of the resulting movies are able to be observed from this representation. The first row, consisting of thumbnails from the version of Morning Coffee with an emotive or intense response sought from the audience, shows a higher *tempo*. I.e. in the 40 second section shown, there is a higher number of shots when compared to the version below, which was created with a more clarity preserving feel aimed at allowing the viewer room to think about the content. The second observable aspect is the automatic choice of framing type (i.e. distance from subject) chosen. The first version contains shots that are on average much closer to the subjects (extreme closeups, closeups). This forces the viewer to do more work to understand and integrate the content and thus supports the emotive/intense goal. Conversely, the version emphasizing clarity has more shots captured at a greater distance (long shots, medium shots), which allow the user more *mise-en-scene* context by which they may understand the movement of the story and its constituents.

6. REFERENCES

[1] E. Schultz and D. Schultz, *How to make exciting home movies and stop boring your friends and relatives*, Robert Hale, London, 1972.

[2] J.D. Beal, *Cine craft*, Focal Press, London, NY, 1974.

[3] B. Barry and G. Davenport, "Documenting life: Videography and common sense," in *ICME*, July 2003.

[4] M. Davis, "Editing out editing," in *IEEE Multimedia Magazine, spec. ed. Computational Media Aesthetics*. IEEE Computer Society, April-June 2003.

[5] X.-S. Hua, L. Lu, and H.-J. Zhang, "AVE - Automated home video editing," in *Proc. ACM MM*, Nov. 2003.