Dynamic Editing Methods for Interactively Adapting Cinematographic Styles

Martin Rougvie
Culture Lab
School of Computing Science
Newcastle University
Newcastle upon Tyne
NE1 7RU
m.g.rougvie@ncl.ac.uk

Patrick Olivier
Culture Lab
School of Computing Science
Newcastle University
Newcastle upon Tyne
NE1 7RU
p.l.olivier@ncl.ac.uk

Abstract
This paper discusses methods for changing the cinematographic style of a film, while the viewer is watching, in order to create a personalized viewing experience. The system achieves this dynamic editing process by choosing the duration of each shot, and which of the available cameras is the most suitable. The choice is based upon information obtained from the viewer, using appropriate heuristics. This paper presents two methods for obtaining the information required for performing editing dynamically. One method requires the viewer to use an interface to inform the system of how they are experiencing the situation. The other method uses an eye-tracker to determine the screen location of the viewer’s gaze. This allows the system to choose a heuristic for the dynamic editing process, which directs the viewer’s attention appropriately.

1. Introduction
Presently, movie experiences are dictated to viewers by directors and editors. In traditional cinematography any piece of film created is intended as a linear sequence of images, defined by the selected sequence of shots comprising each scene. This sequence of images is presented to every viewer in the exact same order. However, recent developments have introduced possible methods to provide a more personal viewing experience in which the displayed images are tailored to fit each individual viewer. By creating and digitally storing simultaneous shots of the same scene, we can dynamically choose the sequence of shots to display to the viewer, according to appropriate heuristics.

2. Motivation
The objective of this project is to enhance and personalize the viewing experience. By personalizing the shots displayed to the viewer according to their individual preferences, they will become more engaged and their experience will be made more meaningful. As the viewing experience changes each time from previous experiences, this increases the motivation of the viewer to do so. Ultimately this approach creates a more viewer-centric environment. A dynamic editing system also allows for a director to influence the personalized viewing experience. By specifying the behaviour of the system under different circumstances, the
director can actively affect the attention of the viewer.

In computer graphics, the process of deciding what should be displayed is typically solved by determining camera parameters and camera movement. Jardillier and Languénou (1998) proposed a constraint-based approach in The Virtual Cameraman, where the camera parameters are constrained by the specified shot properties, such as shot angle or object sizes. Each constraint is considered valid for an interval of values depending on the requirements for the individual shot. The system then uses an interval arithmetic-based solver to find camera placements where all the constraints are satisfied. The ConstraintCam system proposed by Bares et al. (2000) introduces the possibility of relaxing constraints if a solution satisfying all constraints cannot be found. The system maintains the list of constraints in a hierarchical order. When no solution can be found, the system relaxes the constraints with the lowest priority, until it reaches a solution for the remaining constraints.

Constrained-optimisation based approaches combine the use of constraints with shot objectives. Unlike constraints, which are either fulfilled or not, shot objectives have a desired target value. This means that different shots can be compared to each other and the most suitable one can be selected. An example of a shot objective could be that an object should fill 20% of the screen space. Once the shot objectives have been specified, they are used to devise fitness functions that can be optimised, in order to find the most suitable shot for the scene. Drucker and Zeltzer propose the CamDroid system (1995) which specifies camera behaviours in terms of task level goals and constraints. Only camera configurations that satisfy all constraints are considered as possible solutions. The fitness functions constructed from the task level goals can be used to compare possible solutions to each other and decide which is better. The CamDroid system can then use a camera optimiser module to attempt to find the best possible camera configuration.

Tomlinson et al. (2000) introduce a camera control method focused on conveying the emotional state of the scene. This is achieved by monitoring the emotional state of each character, as well as a general emotional state for the scene. The closer the viewpoint is to a character, the more influence that character's emotional state has on how the scene is portrayed. The system changes the shot properties and the lighting intensity and colour to convey a sense of the desired emotion to the viewer.

Christianson et al. (1996) formulate several of the cinematographic principles into a declarative language. Most notably they use film idioms, such as found in Arjio (1976), to increase the storytelling capabilities of 3D computer graphics applications. 16 different idioms are used for controlling the camera configuration. The system decides which of the idioms should be used by analyzing each scene.

3. Proposed System

3.1 Dynamic Editing

In order to perform any type of dynamic editing it is necessary to capture the same scene from various different perspectives simultaneously. Different cinematographic styles and effects require different types of shots. The more variations of shot types available, the more freedom the system has to perform the dynamic editing process. However, in order to perform dynamic editing, it must be possible to evaluate the shot properties for each individual shot, regarding its appropriateness for the situation. When determining criteria for the evaluation of a shot, much can be learnt from traditional accounts of cinematography, such as Arjio (1976), Maselli (1965), and Katz (1991). From these accounts, important targets for the various shot properties, such as camera positions, cutting heights, and screen composition, can be determined. Arjio also provides models for editing a scene in the form of idioms, which describe different ways of capturing a scene and which effects are created by the different approaches. The final target however, does not depend solely on these theoretical ideals, but is largely dependant on the intention of the director. Also, the target will be influenced by the input received from the viewer interaction.
Once the required shot objectives have been decided upon, the system needs a method that allows it to determine the relevant shot properties of each individual camera. A good way of obtaining the relevant shot properties would be to equip the cameras, actors and important scene elements with sensors. This would allow us to abstract a model of the environment, and would enable the system to identify the exact contents of each camera. In absence of such sensor information, we need to devise a suitable form of annotation. The annotation must contain all criteria on which the system bases the camera choice, such as:

- Shot type (e.g. establishing shot, external reverse etc.)
- Actors in shot (in order of importance) and their screen location
- Cutting height
- Camera angle
- Camera movement
- Action description

3.2 Direct Dynamic Editing

We propose directly controlling the dynamic editing process using physical blocks, equipped with RFID tags, and multiple RFID readers embedded in a flat surface that the viewer can interact with. This allows us to determine where on the board the viewer has placed the RFID blocks. The viewer can influence the dynamic editing, by positioning the blocks on different locations of the board, or by removing blocks from the surface entirely. One of the main challenges is devising a setup for the RFID board that is intuitive for the viewer to interact with. This is important, because if a viewer is being allowed direct control over the dynamic editing process, they need to be able to anticipate the effect that their actions will have. The effect a block has is both dependent on its absolute location, i.e. the RFID reader it is placed on (if any), as well as its location relative to other blocks. For example, blocks may represent individual characters. A character block that is located close to the viewer might indicate that the character it represents should be considered more important than other characters. Similarly, the blocks may influence various shot properties directly, such as camera angle or cutting height. So placing a camera block close to a character block may lead to the use of more close shots. On the other hand, if the viewer wants to see more wide angles, this can be achieved by placing the blocks further apart.

3.3 Implicit Dynamic Editing

A more subtle method for influencing the dynamic editing process relies on the data provided by an eye-tracking system. The eye-tracker allows the system to determine what screen location the attention of the viewer is focused on. The system controls the dynamic editing process based upon the gaze of the viewer. Therefore it requires shots with multiple areas of interest, so that the user has a choice of different screen locations to look at. If there is only a single area of interest, such as a typical close shot, the viewer will always look at the same location of the screen, and the system will gain no information about the viewers attention. The majority of shots recommended by typical cinematographic idioms only contain one area of screen space that a viewer would normally focus on. Shots that contain separate screen areas that compete for the viewers’ attention are usually used sparingly. Therefore it will be necessary to develop an individual set of rules for presenting a scene to a viewer that uses more shots, which compete for the viewers attention. This allows the system to gather more information from the gaze of the viewer, which can be used to influence the dynamic editing process accordingly. This should be considered during the principal photography of the scene, as well as during the editing process.

This system also provides opportunities for the director to influence the dynamic editing process directly. The director has the opportunity to create their own rules, to dictate how the system changes the style of editing in specific situations. For instance, if a viewer is showing a large amount of interest in a certain character, then the director may choose to reinforce the viewers’ attention, by showing more shots of the relevant actor. Or if the director wants to subvert the attention of the viewer towards the other characters, fewer shots of that actor will be shown.
4. Future Work

The next step is to construct appropriate editing heuristics for the system, and to determine the exact way in which the viewer interaction should change the editing process. The viewer interaction can affect the content being displayed by, for example, focusing the attention to a different character. It can also influence the cinematographic style and change the way the viewer perceives the action. Possible criteria to consider include the average shot duration, the camera angles and cutting heights used. The heuristics for the implicit dynamic editing method using the eye-tracker, must use enough shots that have multiple areas of attention. This ensures that enough data is being supplied for the editing process to run effectively.

Ultimately, a system will be designed and implemented, providing effective methods for performing dynamic editing. The system should provide the necessary functionality combined with an interface that is simple and intuitive to use so as not to impede the viewer from watching the film. Instead it will enhance the viewing experience by personalizing it according to their individual preferences.

In order to evaluate the performance of the system, several empirical user studies will be performed. In one user study, viewers will be asked to evaluate a scene, once created using the RFID-interface, and once created using the eye-tracker, based upon certain criteria. They will also be asked to evaluate a version of the scene created using a default editing approach. The viewers may be asked to comment on criteria such as level of involvement they felt in the scene, or how easy they felt it was to follow the scene.

An interesting evaluation approach for the eye-tracker system may be to split the subjects into two separate groups. One group will be told that the eye-tracker is controlling the editing process, the other will not. The two sets of results will then be compared to each other, in order to determine, how the information affects the gaze behaviour and the evaluation of the viewer.

5. References


