ELSEVIER

Signal Processing: *Image Communication* 8 (1996) 269–280

SIGNAL PROCESSING: IMAGE COMMUNICATION

Filmic space–time diagrams for video structure representation

Sean Butler*, Alan P. Parkes

Computing Department, SECAMS, Lancaster University, Lancaster, LA1 4YR, UK

Abstract

In this paper, we propose a novel approach to computer representation for interactive film editing. We describe a prototype system developed at Lancaster which uses this approach, and show how such a system supports film editing. Our work involves the isolation of several principles of filmicity and the combination of these with an event structure and film grammar to produce a system which can intelligently edit a film. Filmic principles are rules which can be used when editing a film to produce a sequence which guarantees a smooth perception of the film in its entirety. The visual structure representation method discussed herein is a tool used in our research.

Keywords: Film; Video; Structure editor; Visual representation; Space–time diagrams

1. Introduction

The system described in this paper was created because of the necessity of supporting film editing whilst producing a visual representation of the film structure. We found that during our research into film theory we continually resorted to hand drawn sketches to represent the structure of a scene being discussed. These sketches became more consistent and evolved into *filmic space–time diagrams.*

Several visual representations exist in computer systems for representing film structure [1, 14, 15]. However, none of these is completely satisfactory. Following a description of these systems we discuss several factors influencing filmic space and time, and go on to introduce our own visual representation, highlighting its advantages, how it supports a particular model of video editing, and presents to the editor a visual representation of the structure of a film by displaying the path of the film through space and time.

2. Visual representations of a film

2.1. Traditional desktop video editing

With the advent of Microsoft™ [1] Video for Windows [9] and QuickTime™ [2] on the Macintosh™ [3], most researchers will be aware of the linear editing model used in these and other similar products. VidEdit is a typical product, in that it does not attempt to represent the structure of the

* Corresponding author. Tel.: +44(0)1524 65201. E-mail: sean@comp.lancs.ac.uk.

[1,2,3] All trademarks acknowledged.

SSDI 0923-5965(95)00052-6

video being edited and is essentially a list editor, supporting cut and paste between video files.

2.2. A magnifier tool for video data

Mills et al. [10] describe a magnifier tool for video data that adequately serves as a tool for exploring a hierarchical database of video but does not address the problems involved in editing/re-editing and arranging video fragments into alternative orderings. The magnifier tool, rather than parsing the video to detect the shot boundaries and using those detected boundaries to specify the current magnification scope, arbitrarily sets the current magnification scope as a function of screen width. However, the magnifier tool does overcome problems that are caused by the cramped workspace on most modern workstations, by varying the ratio of presented frames to selected frames, of a given sequence.

A video browsing tool based on Mills' magnifier tool has been developed which parses the video to detect shot boundaries and thus the magnification level is dictated by the currently viewed shot length [17].

2.3. MediaBENCH

Tonomura and Abe [14] implemented MediaBENCH (Fig. 1) as a basic multimedia database system. The structure visually represented by the system is the structure of presentation of the video rather than the video content. The user can place video icons on a time axis and on two spatial axes. These spatial axes do not represent the filmic space, but rather the space of the presentation display.

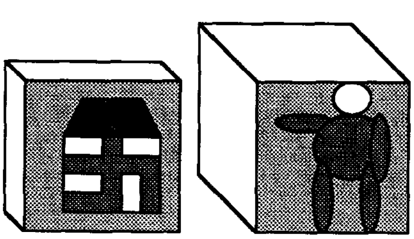

Fig. 1. MediaBENCH.

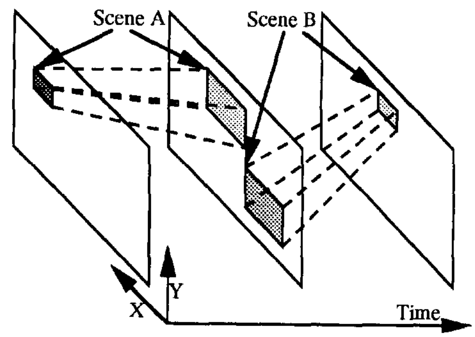

Fig. 2. IMPACT.

2.4. IMPACT

With their IMPACT system, Ueda et al. [15, 16] differ in their approach from others because they aim to visually represent the structural information they can extract from the film using automatic analysis. They use moving icons (micons) to represent a shot with a depth indicator to show the length of the shot. For example, in Fig. 2 the shot of the man is three times as long as the shot of the house. This is indicated by the increased depth (Z-axis) of the micon.

IMPACT also allows dragging of shots arranged in time order into another window where they can be rearranged and played in an arbitrary order.

2.5. Stratification

Aguierre Smith and Davenport [1] describe a system based on keywords used as annotations of intervals of the video. Combinations of multiple keywords produce a layered data structure. The visual representation occurs as a histogram with the frame numbers of the video along the x axis and each position on the y axis containing a keyword. See Fig. 3. This results in a series of horizontal bars representing the span of relevance for each keyword. This system has useful features because it allows the video manipulator to see the context that a frame has. Not only can we see the keywords pertaining to the fragment of interest, but also those which precede and follow it. Thus, the system provides an effective visual contextual index.

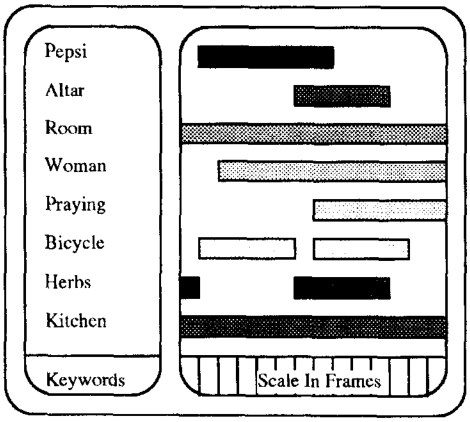

Fig. 3. Stratification system.

3. Cinema theory and its relevance to film structure

Cinema theory is an established field of knowledge which is often ignored by contemporary researchers. The authors propose that to correctly and visually represent the structure of film of any type, one has to investigate cinema theory. Film editors use, even if subconsciously, the principles and theories developed since the birth of film which have been formalised in cinema theory. None of the approaches described above attempt to support film or video using cinema theory, and hence are unlikely to provide complete support for the process of film editing.

3.1. Filmic space

In the nineteen-twenties, D. Kuleshov conducted an experiment (described in [13]) which showed quite clearly the creation of a non-existent space which is realised only on film and in the spectators' mind. Visualise five clips of a film ...

1. A young man walks from left to right (near the G.U.M. Building, Moscow).
2. A woman walks from right to left (near Gogol's monument, Moscow).
3. They meet and shake hands. The young man points (near the Bolshoi Teatr, Moscow).
4. A large white building is shown, with a broad flight of steps. (The White House, Washington D.C.)
5. The two ascend some steps (St. Saviour's Cathedral, Moscow).

Even though the filming had been done in separate locations the audience who were shown the film perceived it as a whole. Kuleshov called the effect demonstrated here "creative geography". Shots representing various locations can be juxtaposed, with the only constraint being the viewers' real world knowledge. For example, we could not have the Tower Of London in Paris, or snow in the Sahara (except in circumstances where we wish to break the rule to achieve a comic effect, for example).

3.2. Filmic time

Filmic time is created by the director when an event is filmed. It is determined by the lengths of the pieces of film edited together. Thus, the same event can be filmed, and the film edited in such ways as to portray the event over two or more distinct durations. There are two types of filmic time: *story time* and *film time*. *Film time* is the time spent by the audience watching the film. Story time is the time through which the film time weaves backward and forward. Using the *event termination rule* and *action off shot rule* the time can be stretched.

The *Event Termination Rule* states ...

"If an event or action has a terminating destination and a terminating state, then we can infer the occurrence of the entire event by portraying only the terminating state" [3].

An example of this in use is in a film of a bomb falling. We see the bomb bay doors open and the bomb leave the plane. There is then a cut to the bomb's destination, showing the location being obliterated. The viewer can infer that the bomb has fallen from the plane, because of the portrayal of the terminating state, the explosion.

The *action off shot rule* states ...

"When using the *event termination rule* we can portray the amount of time taken for the entire event by portraying the location of the terminating state prior to the state being entered into, for the duration of the event" [3].

To portray an actor entering an office, then walking from one side to the other, we would show the actor entering the office to just inside the door, cut to the location where they were going to walk (the location of the terminating state), and hold that location for the duration of the walk across the room (the duration of the event). Finally, we would show them entering the shot in this location (the terminating state). We hope that the reader can see that the duration of the walk is the crucial element and can be lengthened or shortened.

It should be noted that use of the above two rules has a side effect when the event chosen involves movement in filmic space. The side effect is to alter the length or speed of that movement. From these considerations, we can form the principle of *conservation of filmic space-time.*

4. Principle of conservation of filmic space-time

Viewers bring common-sense knowledge with them and use it to construct a revisable fiction [13, 6] as they watch a film. One piece of this knowledge roughly corresponds to *speed = distance/time.*

The principle of *conservation of filmic space-time* is used if we show an event portraying two of the values from the above equation, i.e., distance and speed, or time and speed, or distance and time, then the viewer will use common-sense knowledge to empirically estimate the missing value. If we show only one of the values then the viewer will invent a missing value (from their own experience) and do the same estimation for the third. *Time* is usually the value which is supplied. This is especially true if there are no cuts or edits in the sequence.

Using the principle of *conservation of filmic space-time* we can either confuse the viewer, or cause them to recalculate in the light of new evidence, by supplying a missing value during a later part of the film. This principle is usually adhered to, except in the case of unnecessary information e.g. a tedious journey from one location to another with no significant details. In Naked Gun 2 1/2: The Smell Of Fear (Zucker, 1991, US) we see the two main characters fall in love and do many *falling in love* things all in the space of a day. This scene is humorous not only because of the absurdity of the actions themselves, but also because the principle of *conservation of filmic space–time* is broken when the characters manage to carry them out *all in one day.*

5. Why cinema theory is important to film structure visualisation

The scene structure is determined by the editor, who is influenced by the story, by the director's instructions, by the footage available, and by his/her own knowledge of cinema theory. The editor's goal is to tell the story, and to present it in an understandable way. He/she knows that the phenomena of audience perception of film can be described using cinema theory. So to realise the script as film he/she *must* take these theories into account (at least unconsciously) if the audience is to perceive the script as intended.

The structure of a film is determined by the position of the cuts within it. The position of the cuts are specified by the movement of the viewpoint through space and time (as well as other artistic influences). This movement is dictated by the narrative/story. The editor has to realise this narrative given a set of shots which may or may not map directly onto the narrative itself. The footage made available to him/her creates a filmic, space-time through which the camera moves. The editor can alter the movement of the audience's viewpoint by cutting and rearranging film fragments. In order to visualise the structure of the film we can simply portray the movement of the viewpoint through the filmic space–time. This will in turn allow an editor to perform manipulations of narrative relations as well as film units.

6. Space–time diagrams for filmic representation

We can plot film time along the x-axis of a graph and spatial position of the camera along the y-axis to obtain a visual representation of the cutting in a piece of film. The solid lines in Figs. 4–11 represent the actual footage. The dotted lines join the fragments of film in their correct order, they are used to represent a *cut* and therefore show the path to be taken through the graph by a film when it is played.

We now demonstrate that each of the major types of cut has a unique visual appearance by this representation method.

6.1. Spatial deletion

Spatial deletion [5] occurs either when we move the camera from one position to another without showing the intervening locations on film or when space is not portrayed because it is outside the frame of the camera. I wish to restrict the definition of spatial deletion to the former, and introduce a new term for the latter *inherent deletion.*

When spatial deletion occurs, movement of the camera is instantaneous. The dotted line 90° to the time axis (in Fig. 4) shows a movement in the spatial position of the camera, without any change in the time.

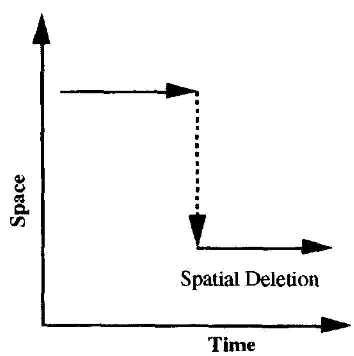

Fig. 4. Spatial deletion.

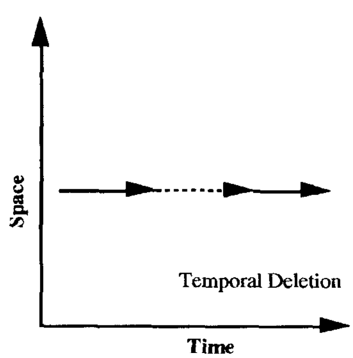

Fig. 5. Temporal deletion.

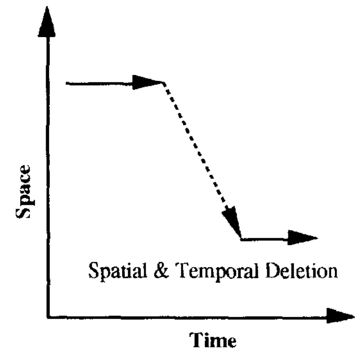

Fig. 6. Spatial and temporal deletion.

6.2. Temporal (or linear) deletion

Temporal (or linear) deletion, is the removal of time from the film [5]. For example, in Naked Gun 33 1/3 (Zucker, 1993, US) we see Priscilla Presley enter a room, change her clothes and leave the room all in the space of a second. The changing has been temporally deleted. The horizontal line in Fig. 5 represents the fact that no movement in the position of the camera has occurred.

6.3. Spatial and temporal deletion

The previous two types of cut are usually combined to produce spatial and temporal deletion (Fig. 6). This is the most used type of cut, and almost every cut we see in a film is of this type. It is the combination of this and the above two types of cut in various orientations which give us the structures most prevalent in films.

7. Flashback and flashforward

Flashback (Fig. 7) can be thought of as a "digression" [2] and is a temporal aside or recount of earlier events.

A flashforward (Fig. 8) is the same but to the future. Note that the orientation of the lines on the graph is reversed.

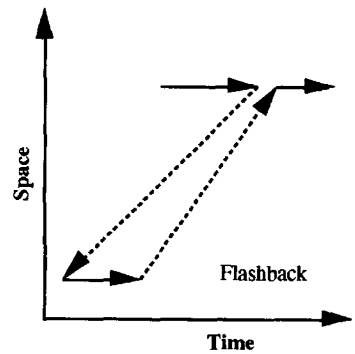

Fig. 7. Flashback.

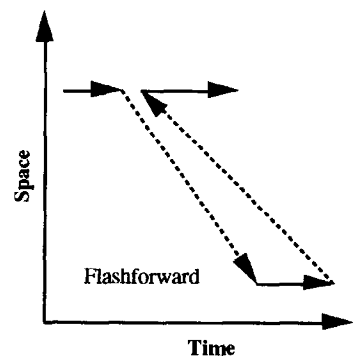

Fig. 8. Flashforward.

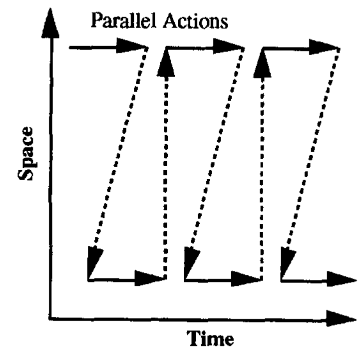

Fig. 9. Parallel actions.

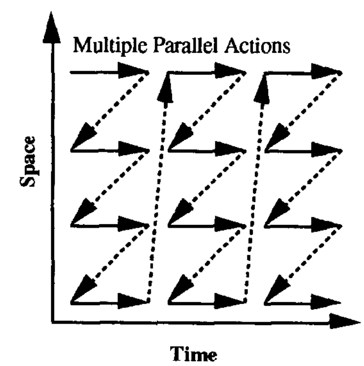

Fig. 10. Multiple parallel actions.

7.1. *Parallelism*

A more complex example is parallelism, which consists of multiple repeated spatial and temporal deletion followed by spatial deletion. *The Birth Of A Nation* (D.W. Griffith, 1915, US) was an early and influential film using this technique. *The French Connection* (William Friedkin, 1971, US) is more recently renowned for one of the finest car-chase scenes ever made. This type of scene relies on parallel action (Fig. 9).

8. Combinations of space–time diagrams

As can be seen from Fig. 10, space–time diagrams can be combined to produce more complex versions. Here we see multiple parallel actions. Four alternate threads of plot/film are being interwoven.

In Fig. 11 spatial and temporal deletions are nested within a flashback. This is a more likely structure of a flashback.

8.1. *Film editing model*

Cinema theory in isolation is not enough. Any system to be used in a practical environment must take into account the actual working practices of the users. Nack and Parkes [8] propose a model of film editing based on observation of film editors at work. This model is summarised in Fig. 12.

From a stated goal a *story board* of potential scenes is created. Then the editors (often working in pairs to facilitate creative discussion) browse through the available material grouping shots into *heaps* each representing a potential scene. Each heap is ordered and annotated with every shot's

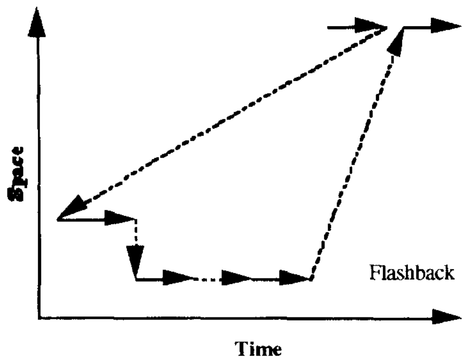

Fig. 11. Flashback.

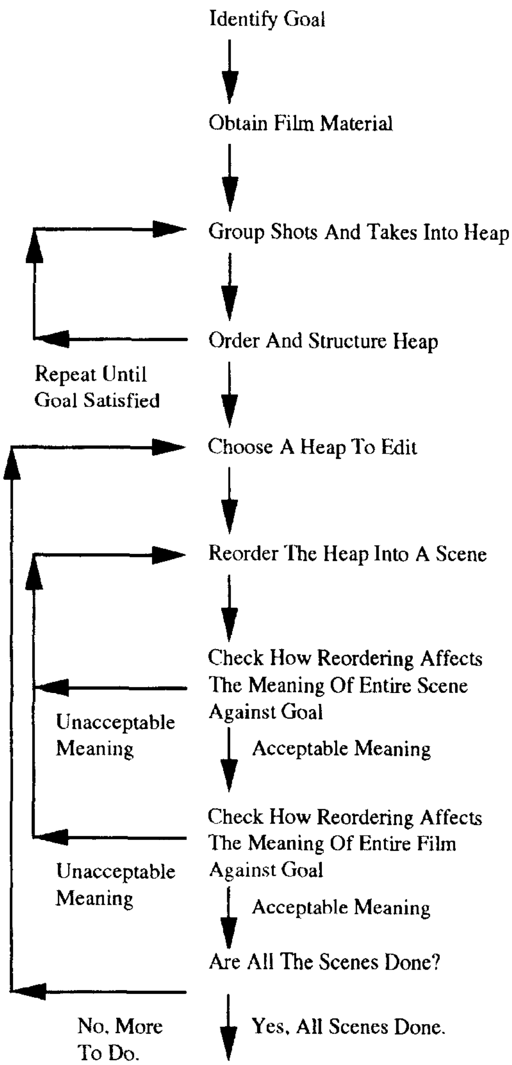

Fig. 12. Nack's film editing model.

attributes, such as length, location, etc. Choosing one of the heaps as a starting point, the editors reorder the scene based on the effect they wish to achieve or the idea they wish to communicate, i.e. they *edit the scene to tell the story*. The next two stages are comparisons to check that the current reordering is consistent with the ultimate goal. This whole sequence is repeated for each heap in turn.

8.2. Prototype user interface

The space–time graphs are being used as the main interaction method in a prototype video editing system under development here at Lancaster University. This system was written in GNU C on a Sun Sparcstation using VideoPix, the X-Windows System, and Motif widget set. The prototype (see Figs 13–15) supports the stages of Nack's model from "obtain film material" to "check how reordering affects the meaning of entire scene against goal" with the functions listed here:

- The capture and annotation of shots.
- The grouping of shots into heaps.
- The ordering and structuring of each heap occurs automatically based on the annotations entered at capture time.
- The choice of a heap to edit is made using a motif file selection dialog.
- The reordering of a heap and thus creation of the space-time diagram is achieved using the edit window, by unlinking and linking the fragments of video. See the menu on Fig. 14.
- A partially edited scene can be played from the current edit point, or from the beginning to check that the desired effect is being achieved.
- Any given fragment may be trimmed or shaved by a few frames at any point.
- Movement from one section of the system to another has been implemented to allow the feedback loops of the model to occur.

It can be seen from Figs. 13–15 how the interface incorporates the space–time diagrams outlined earlier. The horizontal arrows representing fragments of film are replaced by horizontal bars with a label naming the film fragment, and the dotted arrows representing the cuts and movements in space–time without film are replaced by white lines.

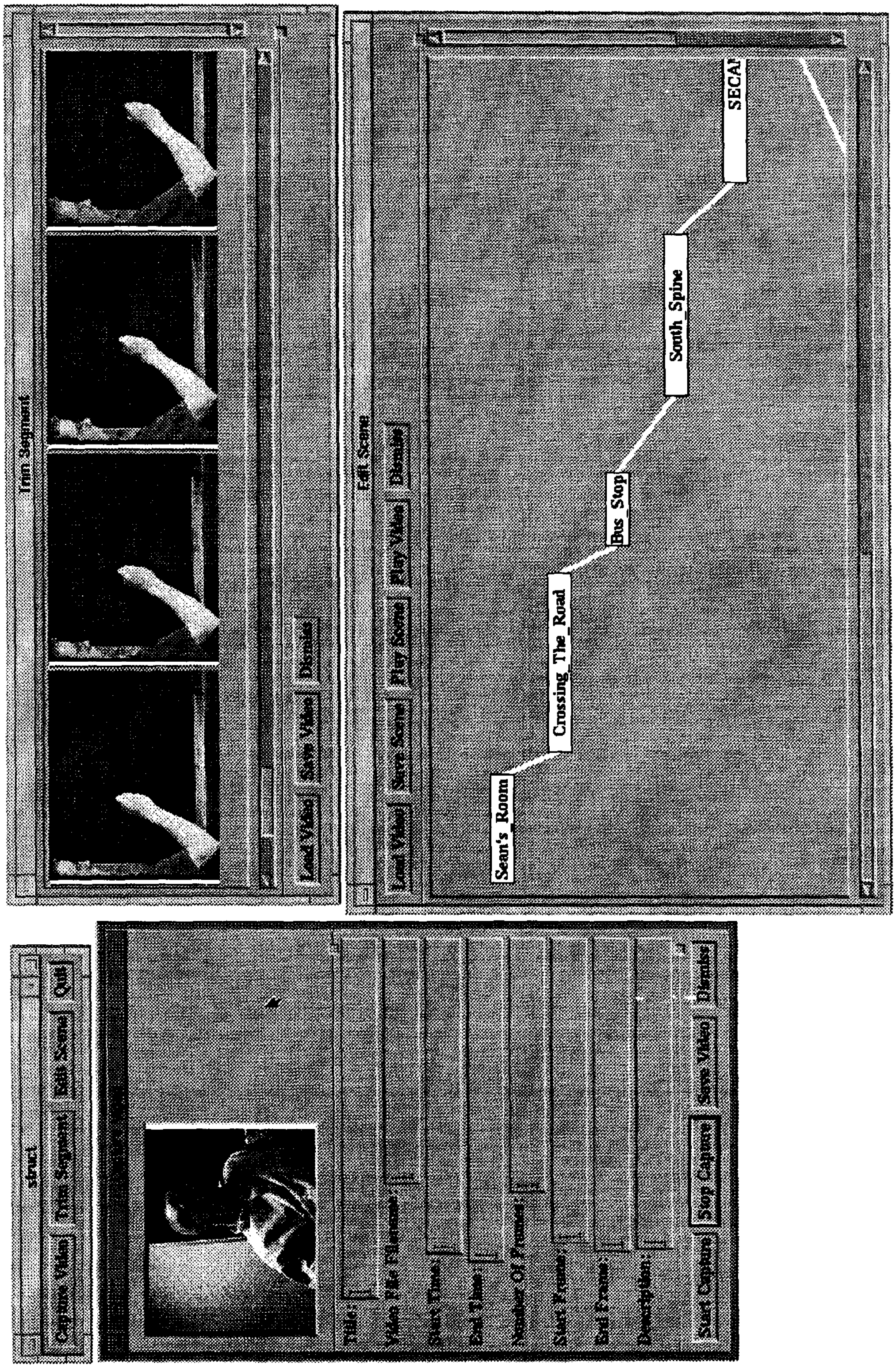

Fig. 13. Structured editor showing temporal and linear deletion.

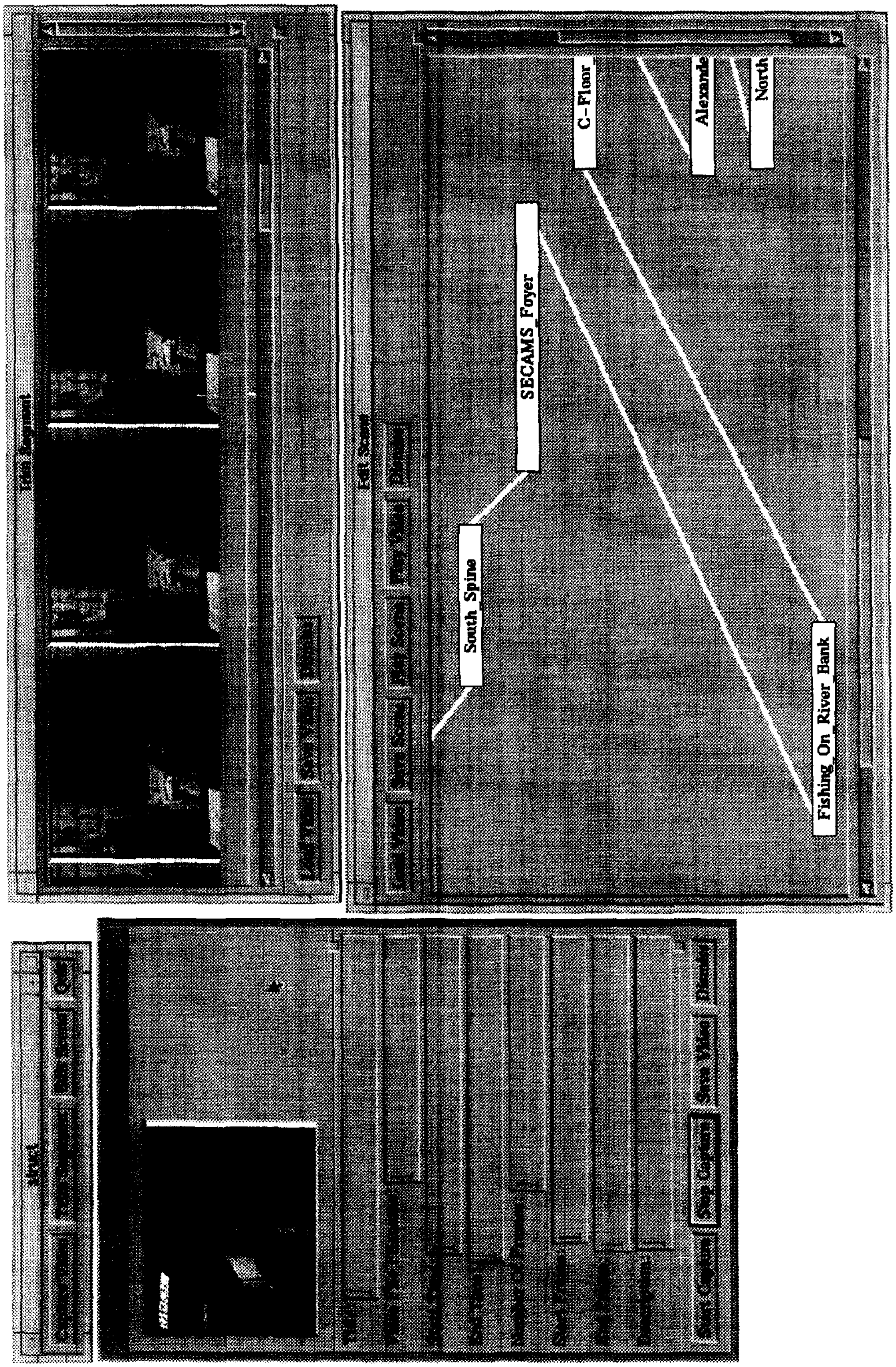

Fig. 14. Structured editor showing a flashback.

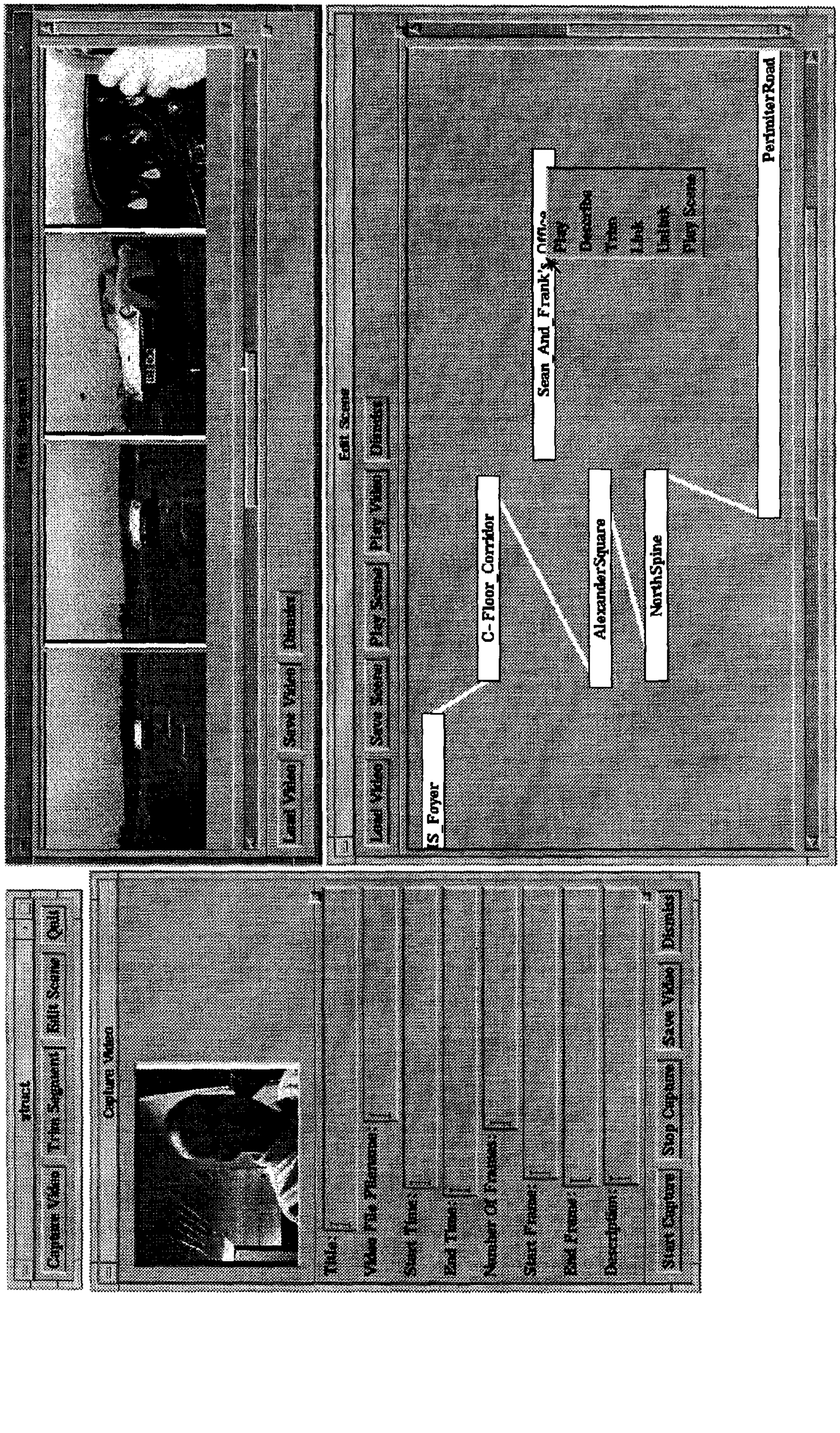

Fig. 15. Structured editor showing parallel actions.

When the user first encounters the edit window there are no linking lines. Only the horizontal bars exist. By using the trim, link and unlink menu options editing becomes a process of drawing the space–time diagram desired. Fig. 13 shows examples of spatial and temporal deletion. A flashback can be seen visualised in Fig. 14, and Fig. 15 shows parallel action as it appears on the system.

9. Results

The system described is under continued development, however, it is used by the authors and other research students at Lancaster to create short films for presentation during research group seminars. These short films are created to highlight a particular filmic principle for discussion. Space–time diagrams have been most useful in the iterative re-editing that often occurs when creating films and in the communication of a film's structure between two editors.

10. Further work

Further development is underway to flexibly support the entire editing model. The aim is to remove the ordering constraint of the model and allow experienced editors to move quickly back and forth between creating heaps and reordering them into scenes, whilst still constraining novices to a structured approach.

Other main direction of research takes the form of determining a grammar for film and embedding this with an event structure, to support various applications of automatic intelligent film/video editing [4, 7]. However, several researchers have commented on the difficulty of determining a terminal symbol for such a grammar, and the mistaken belief that comparisons can be drawn between natural language and film language [11]. Parkes investigated this problem and defined the *setting* as a minimal working unit for film manipulation [12]. It is this work upon which we base our current research. Our ultimate goal is to produce a system which can automatically and intelligently edit film in response to a user request taking into consideration the film medium and the semantic effects that change in the juxtaposition of film fragments will have on the sequence as a whole.[4]

Acknowledgements

This work is supported by the EPSRC and Lancaster University.

References

[1] T.G. Aguierre Smith and G. Davenport, The stratification system: A design environment for random access video, MIT Media Lab Technical Report, 1992.
[2] R. Arnheim, *Film as Art*, Faber & Faber, London, 1958.
[3] S. Butler, Filmic principles for automatic intelligent video editing, Applied AI Group Technical Report: 106. Lancaster University, 1994.
[4] S. Butler and A. Parkes, "Intelligent strategies for the presentation and interpretation of video in intelligent tutoring systems", *Proc. ED-MEDIA '94*, Vancouver, Canada, 25–30 July 1994.
[5] J.M. Carroll, *Toward a Structural Psychology of Cinema*, Mouton, The Hague, 1980.
[6] S.M. Eisenstein, *Selected Works: Writings 1922–1934*, BFI Publishing, London, 1988.
[7] F. Nack and A. Parkes, "Automated film editing for educational applications? Don't make me laugh!", in: *Proc. ED-MEDIA* '95, 17–21 June 1995, Graz, Austria, pp. 477–482.
[8] F. Nack and A. Parkes, "AUTEUR: The creation of humorous scenes using automated video editing", in: *Proc. IJCAI-95 Workshop on AI and Entertainment and AI/Alife*, Montreal, Canada, 20–25 August 1995, pp. 82–89.
[9] Microsoft, *Video for Windows™ User's Guide*, Microsoft Corporation, USA, 1992.
[10] M. Mills et al., "A magnifier tool for video data", *Proc. CHI'92*, Monterey, CA, 1992, pp. 93–98.
[11] J. Manaco, *How to Read A Film*, Oxford University Press, New York, 1981.
[12] A.P. Parkes, "Settings and the settings structure: The description and automated propagation of networks for persuing videodisk image states", in: N. J. Belkin and C. J. van Rijsbergen, eds., *Proc. Twelfth Annual Internat.*

[4] Published Papers and Group Technical Reports by the Authors and other members of the Lancaster Applied AI/AI-ED Group are available from the URL http://www.comp.lancs.ac.uk/ using Mosaic, LYNX or similar utility.

ACMSIGIR Conf. on Research and Development in Information Retrieval, Cambridge, MA, 1989, pp. 229–238.

[13] V.I. Pudovkin, *Film Technique*, Vision Mayflower, London, 1958.

[14] Y. Tonomura and S. Abe, "Content oriented visual interface using video icons for visual database systems", *Proc. IEEE Workshop on Visual Languages*, IEEE Computer Society Press, Rome, 1989, pp. 68–73.

[15] H. Ueda et al., "IMPACT: An interactive natural-motion-picture dedicated multimedia authoring system", in: *Proc. ACM CHI'91 Conference on Human Factors In Computing Systems*, 1991, pp. 343–450.

[16] H. Ueda et al., "Automatic structure visualization for video editing", *Proc. ACM & IFIP INTERCHI'93*, 1993, pp. 137–141.

[17] H. Zhang, S.W. Smoliar and J.H. Wu, "Content-based video browsing tools", in: *Proc. IS&T/SPIE'95 Multimedia Computing and Networking*, San Jose, California, 1995, pp. 389–398.