

## Chapter 1

# SPACE-TIME MAPPINGS AS DATABASE BROWSING TOOLS

Stephen Mamber

*Department of Film, Television and Digital Media*

*University of California*

*Los Angeles, California, USA*

smamber@ucla.edu

### **Abstract**

One possible value to examining cinematic structures is that once understood, they can serve as visual and conceptual mappings which can then function as an interface to accessing the work in question. Determining such structures is not purely a systematic technique, but some ideas about segmentation can perhaps be useful in developing approaches to accessing visual material.

### **Keywords:**

Visual databases, film, video, browsing, narrative, segmentation, spatio-temporal mapping, Cartesian grid.

### **Introduction: Film and Video as a Database**

It is not immediately evident that a film or video work might be considered as a database. In fact, the impulse to do so is widespread. Based upon some method of segmentation, there are many situations where there is a need to search for one or more specific moments from a much larger amount of moving image material. Also, such databases may be somewhat relational, in that links to tables of other data can be established with segments in the moving image material. How to use film and video, then, as a kind of browseable interface, is one of the principal problems explored in this chapter.

Principal emerging areas where the notion of considering a film as a database have been explored are in academic film studies, and its offshoot, digital media theory. See particularly [1, p. xv-xxxvi and Chapter 5] and also [2]. (A far from negligible related field would be film and video production, which treats all of its footage to be edited as a database from which a finished work emerges.) Because ideas of narrative and segmentation are well advanced in these fields, we can look at the possibilities of drawing upon accepted views of segmentation as applied to a database environment.

While films and video can be thought to be composed of individual images, they are not equivalent to databases of still images. The reason is fairly simple - moving image works have narrative relations. Their order matters, as do sound-image associations. A ten second shot of a flower is not the same as 240 successive images of a flower - the moving image footage has temporal and spatial dimensions which are not inherent in the still images. While some ideas of how to treat databases of still images may carry over, considering moving image media as databases require a respect for their uniqueness. How can we look at a film in ways with which we are familiar, and use those ways to extract desired segments and examine related materials? That is what we will attempt to explore.

## **1. The Need to Segment and The Narrative Map**

A main idea we wish to advance here is of the browseable database composed of thumbnail images representing segments of the work. With such a scheme, a narrative could be “zoomed” in on and out of, in a manner corresponding to a map. A thumbnail representation (each thumbnail corresponding to one contiguous element of segmented narrative) of a film or video is actually a kind of map - a narrative map. But what are the units? There is no simple correspondence of space in a geographic map to time in a narrative. We can explore, though, a number of solutions to this problem, seeing what possibilities we can find in alternative segmentation approaches, and their usefulness in constructing narrative maps.

One goal is to consider thumbnail representations as visualization tools. Which representations enable us to see the contours of the work more clearly? Which lead us to mentally reconstruct where desired segments may lie? Which allow a kind of meaningful zoom function? And how should the presence of additional materials be represented?

It is indeed the case that film and video can be indexed by individual frames, though this is usually too fine a level of granularity to be useful. As film is composed of 24 frames per second, and video is 30 frames per second, one hour of film consists of 86,400 individual frames, and one hour of video, 108,000 frames.

There have been two standard and fairly obvious numerical schemes to mark individual frames - hours/minutes/seconds/frames or simply cumulative frames numbered from one. Each would suggest divisions which could serve to mark segments - such as one minute or one hundred frames constituting a uniform segmentation - establishing something of a timeline mapping.

A limitation is that the thumbnails produced by such segmentation can be both duplicative and misleading. If one creates thumbnails for each ten seconds, a 45 second shot will have four thumbnails to represent it (as would a 40 second shot). This type of indexing, then, would reliably suggest total shot length, at the cost of strings of identical thumbnails. Also, shots which are shorter than the length of the interval risk not being represented by a thumbnail at all. (An eight second shot, for instance, in a database of ten second intervals, could be missed entirely.)

## **2. The Shortcomings of Common Database Search Practices as They Apply to Moving Image Databases**

One may well ask: why cannot moving image databases be treated like text databases? One reason is likely to be practicality. To make a film conventionally searchable would require manual keywording - assigning topics to selected segments following some established method. Beyond being a time consuming and somewhat arbitrary approach, the shortcoming would be that the assigned keywords would have to correspond to the user's interests in searching the visual material. Those interests are more difficult to predetermine with moving image works - their very utility frequently lies in the complexities and ambiguities which each user seeks to explore. So predetermined categories and text markings, beyond being laborious, might not assist locating what the user may only be starting to discover about the work in question.

Sometimes a search for similar visual characteristics which can be arrived at through visual analysis would be useful, but these methods also ignore narrative segmentation and beg the question of appropriate units to return. Finding all the images with leaves, for instance, while of value in searching a photo database, might return results of a 24th of a second or several minutes, or might bring up instances which are trivial

in terms of their narrative importance. Eventually, searches for visual characteristics might have a role in assisting moving image database browsing by providing a means of establishing narrative segmentation if there is a certain variation in the visual material, though it would strike us at this point as a hit or miss proposition. But, “smart” analysis of streaming video is a promising area of current work. Virage [3] and IBM Almaden Research [4], among others, are doing interesting work in this area.

### **3. The Cartesian Grid as the Spatio-Temporal Mapping for Browsing**

Laying out thumbnails in a grid is a form of spatializing a temporal sequence. While a continual left-to-right stream would seem more in keeping with moving images, the grid looks like what we expect of a database, even if it is just a method to squeeze more thumbnails into the space of a screen. While the points at which a  $32 \times 32$  grid, say, privileges a new row (the 33rd, 65th, etc.), a user can simply learn to “read” a visual database as one reads lines of text - the new line is understood to continue from the previous, just as in reading text.

The grid also serves as an aid to pattern recognition in the overall work, and to understanding “where” in the narrative map one is located. A user can easily sense if one is halfway through the material, which a continuous stream would not provide as succinctly. And moving from left to right and then from top to bottom allows an arrangement which spatializes temporality, since the Cartesian grid is our principal form of representing physical space. So the grid itself is a form of narrative mapping.

#### **3.1 From The Frame to The Shot**

The most useful segmentation of a film for database purposes is very likely the shot, for a number of reasons. The idea of the shot is readily accepted and understood. It corresponds to how films are actually put together, and often how screenplays are designed. In a typical feature film, there are usually from 900-1000 shots [5]. If a 90 minute film has 1000 shots, then they average 5.4 seconds each, so each unit is relatively brief.

It is also important that when viewed as thumbnails, shot representations can still follow remarkably coherent patterns - they can continue to make narrative sense. This is likely because the very mechanisms of film narrative can continue to prevail - alternations of close-ups and establishing shots, reverse angles, fluctuating distances - the fluidity of film

can still be understood by representative thumbnails - so that sequences are “readable”.

The likely candidate for thumbnail status would be the first frame of a new shot. While this is an arbitrary method, it is a consistent one. If dissolves, wipes, or other optical effects overlap the successive shots, the first clear frame of a new shot can be selected. As in any complete segmentation, the only number which would have to be stored (in addition to the shot number) would be of this frame - the end frame would then simply be one frame before the next new shot.

A shortcoming of shot segmentation would be that they give no indication of shot duration. Since a shot can be as brief as a few frames or last as long (in the case of the recent film *Timecode*) as the entire length of the work, there can well be circumstances where a shot segmentation would not yield a functional narrative mapping. But even in amateur video, it is very common to see Hollywood norms followed, and for average shot lengths to be observed.

Thumbnail representations of an entire film on one screen can be accomplished by a  $32 \times 32$  grid. (This would allow 1024 shots to be represented.) On a  $1024 \times 768$  resolution screen, thumbnails of 22 pixel width would be possible, a size which still allows the individual images to be recognizable. This would be a narrative mapping - in that the overall contours of a work can be comprehended all in one glance. (See Figure 1.1 for an approximation in print of seeing a complete film.)

Assuming a functional browseable database interface such as a shot-segmented set of thumbnails, how would this be utilized? One approach would be to overlay (or switch to, or show on a second window or a monitor) the moving image video when an individual thumbnail is clicked, starting the video at that point. An alternative implementation would be to allow for some sort of dragging, permitting the user to indicate starting and ending segment points, and then playing exactly that corresponding video. The thumbnails, as a starting point, can serve as an interface to viewing the video itself, as the numbers for each shot can mark (in a linked table) where starting and ending points for clips should be. It is also possible to play each successive shot “in place”, directly over the selected thumbnails. In this way, the video moves as each shot is played, representing dramatically the breakdown of the film into shots as the video moves across the grid.

A simple refinement of playing selections from thumbnails would be for the application to store all such selections. For subsequent searches, representations of these selections could be generated on the fly, and the user could either expand or restrict the previous selection. The outcome of such searches would be the segments which the user wished to locate.



*Figure 1.1.* A graphical representation of all the shots in Alfred Hitchcock's *The Birds*, by the author. Each square is one shot of the film.

We will return to other forms of functionality which can be added, principally by considering how to link additional database resources to the moving image material, after we consider additional forms of segmentation beyond the shot.

### 3.2 Self-Generating Segmentations

As films and videos are time-based, many activities captured as moving image material can easily suggest self-generating segmentations. These structures could very naturally determine hierarchical levels of thumbnail representations.

It is the rare sporting event, for instance, which does not have inherent temporal divisions. In most cases, sports are run by the clock, divided into quarters or other fixed periods. With football, it is easy to segment further into plays, or basketball marked by the movement of the ball from one end of the court to the other. Even a sport famously not bounded by time, baseball, could not be more narratively segmented -

halves of innings, outs, batters, pitches - these readily suggest a range of mappings.

Marking individual arrivals and departures, whether of actors or others, can also suggest a structure, assuming those appearances are regular enough to segment the time period. Any regular set of internally generated events can function as a narrative mapping. These might include temporal indicators within the work itself. This illustration of a mapping of the Stanley Kubrick film *The Killing* suggests such an approach. (See Figure 1.2.) The squares represent appearances of certain characters, and the times are provided by the film's narrator. This schematic could then be used to generate thumbnails representing each time period, and thus serve to segment the work.

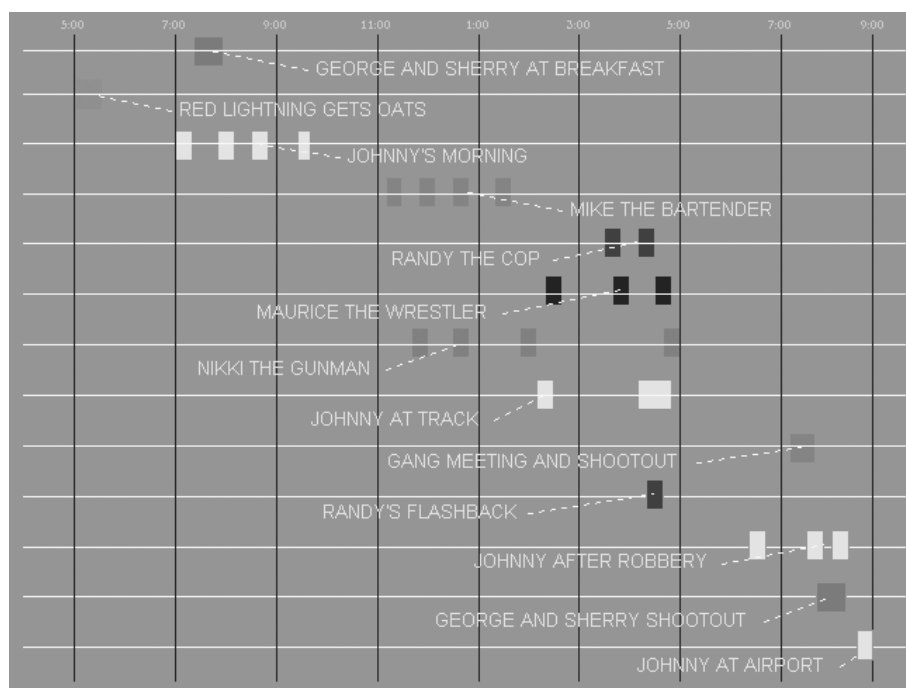
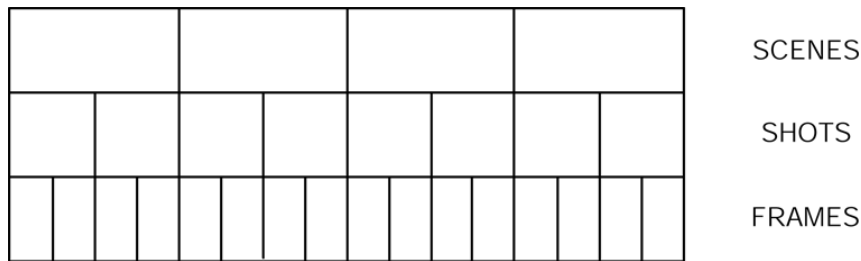


Figure 1.2. A mapping of Stanley Kubrick's film *The Killing*. Vertical lines represent hours, and horizontal lines represent scenes of individual characters.

### 3.3 Beyond Shots

Conventional notions of scenes can also provide a narrative mapping. Most commercial DVD's include such an interface to access parts of the

digital video. Scene breaks in films are usually marked by space and time separations (another location at a time subsequent to the scene which preceded it). See [6] for a good discussion of spatial and temporal articulations constituting the basic language of cinema. In general, either temporal divisions or shifts of location could each in themselves serve as forms of segmentation. That they often accompany each other only serves to underscore the ease with which multiple forms of segmentation can be complementary - either hierarchically or as an alternative at the same level. Various such structures suggest themselves. For instance, one hierarchical model would be the following (See Figure 1.3):



*Figure 1.3.* A hierarchical model of progressively finer mappings.

Christian Metz, a leading narrative theorist, offered a branching model of all possible narrative forms - branching itself a form of establishing a hierarchy [7].

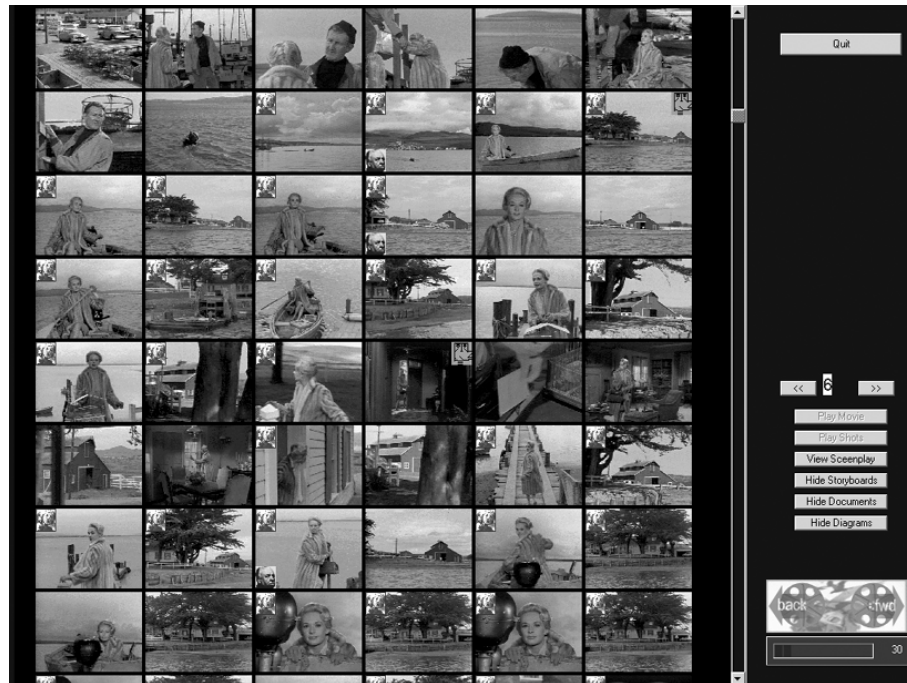
Overlays function by allowing for another perspective-shifting activity: the zoom. A valuable aspect of thumbnails is that they are resizable on the fly. Browsing a narrative mapping of a moving image database can be greatly enhanced by forms of moving in and out - in thumbnail terms this means changing the size of each image. The trade-off, of course, is that zooming in means you are no longer seeing the whole picture; but that is what navigating is about. This kind of searching is visual and requires a grasp of overall structure. Moving “in” can be represented by both greater detail and larger representations. A 32x32 grid can give way to an 8x8 but still fill the screen.

### 3.4 Embedded Linkages and Taggability

Another advantage to thumbnails is that they allow for further economical overlaying: they can provide the means to indicate the availability of related database materials. To use a feature film as an example, if there are any sets of linked materials, they can be checked in real time



(say by the user clicking one of a set of buttons, each bringing up an additional set of materials) and if there is anything available corresponding to elements of the current segmentation, small icons along the edges or corners of each thumbnail can be used to indicate their availability. (See Figure 1.4.) These icons can then be “live”, in that clicking on them brings up that linked item from the database.



*Figure 1.4.* A shot database as an interface, from an application by the author. Notice the small icons in the corners of the shots, which are live links to database material.

This kind of searching and accessing is also fully visual - no keywords were typed and the user did not have to know what was there prior to asking for a new set of linkages. All that was needed was a desire to know what was available in a given database which corresponds to the current segmentation.

Some examples of how this could operate would be to see any text materials related to a current moment in the film or video, such as a transcript or a version of a script. Audio commentaries could be similarly accessed, as could drawings, sketches, statistics, and the like. The idea is that each item bears a relation to the specific segment, so that the

user clicking on any icon brings up data of that type. The narrative mapping is basically taggable. Like geographic maps, this might be thought of as the equivalent of adding (and removing) at will symbols for schools, hospitals, and other places of interest. Essentially another series of overlays, they still employ the underlying structure of the narrative segmentation, so that the additional resources are locatable through their relation to the work as a whole.

#### 4. Alternatives to The Shot

We will now review some potential alternatives to shot breakdowns, all suggested by the notion on narrative mapping - how to represent visually a moving image work.

*Timelines of Significant Actions* - While “significant” is certainly a loaded term, a work can be marked by its principal events and the times they occur. Simply charting when such actions take place can lead to a useful segmentation. This may be especially efficient when the material contains a lot of “dead” time, periods of little action, such as surveillance video.

Here is an illustration of such a timeline, taken from a web site of the author’s devoted to hidden camera footage. “The Center for Hidden Camera Research” is at [www.cinema.ucla.edu/Mamber2](http://www.cinema.ucla.edu/Mamber2). Other examples of narrative mappings are to be found in the “Gallery” section. (See Figure 1.5.) When segments are pulled out in this manner, indications of relative duration can also aid in contextualizing the positions of each segment. (In this example, when video is selected, explanatory notes may also appear because each segment has been annotated, but this is not essential.

These incidents are used as reference markers, so that they signpost the time surrounding them, making it possible hopefully to locate a broader range of events than just those highlighted. Mapping generally functions to allow for situating small locales in relation to larger, easier to locate “landmarks”, so this would be such a scheme to facilitate that process.

*Appearances of People or Objects* - Such strategies can singly, or in concert, provide useful narrative mappings. The illustration from *The Killing* (see again Figure 2) can be viewed as an example of such an approach, although in this case, the film is very precise in dividing scenes based upon the appearance of individual characters. Here, one new character appearance equals one new scene, so a mapping by scenes is essentially equivalent to a mapping by characters. This will rarely



Figure 1.5. A timeline mapping, from “Center for Hidden Camera Research” web project. Clicking on small picture squares changes both video and text commentary above.

work so neatly, so charting people or objects, regardless of scene or shot divisions, can be a potential alternative.

Regarding the use of objects as a mappable component, this will lead to possible structures if the objects in question are suitably varied yet consistent. With each mapping structure, if there are regular appearances of any element, they can serve as forms of segmentation.

*Stylistic Segmentation* - While this may seem more cinematically specific or to require greater familiarity with artistic nuances, there can be a pragmatic dimension to using elements of the medium as a mapping approach. Camera movements, for example, might be easily marked, or alternations of light or color. Visual materials can have regular fluctuations in stylistic attributes, which would serve to provide a useful mapping.

Virtually any noticeable stylistic attribute might suffice as a marking point, so some perhaps can be quite obvious. Beyond movement and light variations (of which there are many possible kinds), a few others

might be camera angle, variations of distance (from close to medium to far), shifts of point of view, or clarity of image (such as use of deep focus lenses or rapid shifts of focus).

The idea here is that such visual qualities may structure a work as extensively as shot or scene segmentations. These mappings might be even more functional in terms of allowing a user to grasp overall patterns and design.

*Geographic Mappings* - On one level, this could simply mean a segmentation based upon shifts of location. As such, if once again there are regular and structured changes, this would serve as a useful approach. Another common variation is shifts between exterior and interior locations. These also might correspond roughly to scene breaks, but they might be themselves more indicative of segments within the work than would scene breakdowns.

A good number of films are structured around traveling from place to place, so direct mappings of locations can serve as interesting and useful reference points. Well known and typical examples of such films would include Alfred Hitchcock's *North by Northwest*, or virtually any so-called road movie, which use travel as their basis. Once more, even collections of home movies might suggest that such mappings could be quite functional. When locations are clearly identifiable and serve to divide a work, they provide a clear opportunity for mapping in this manner.

*Dialog-Based Mappings* - Links between transcripts or screenplays and digital video are a most useful form of linkage and segmentation. A script is itself a kind of mapping, one which has added utility because it frequently exists already and does not have to be specially created. A simple mapping would be to link pages of a script to their corresponding points in the work, so that "pages" become the structural equivalent of "shots" in terms of segmentation. The level of specificity (in fact, an issue like zooming in a shot representation of a database) can be as one wishes - sentences, paragraphs, speeches, etc. The desirable result of free movement and locatability between one data set (or script) and another (the digital video) is easily obtainable through a visual representation of a set of materials.

Many kinds of film and video also have accompanying text databases, either in the form of closed captioning or subtitles (sometimes in multiple languages). The process of breaking these into locatable segments is akin to the more visually oriented mappings discussed here, and in many cases may be the most suitable form of creating a mapping of the work.

*Sound-Image Relationships* - While they require the digital video to be playing in order to be heard, sound is an equally significant element for

potential narrative mappings, even beyond what might be spoken. The use of recognizable music, variations between silence and sound, or any such aural element can serve as marking points. Tracking when certain people speak, regardless of what they say, can also establish patterning in the work.

## **5. Conclusion - Generalizing the Notion of Segmentation**

While a number of forms of segmentation have now been discussed, the overriding idea is to imagine these possibilities as freely co-existing, even with other potential mappings which might arise out of the specifics of a particular set of visual materials. There is no single method which need be standard, and the general reliance so far on shots as a structuring device may be more a matter of convenience than anything else.

The goal, then, has been to consider how moving images might be well served by turning the idea of a database upside down. By using the work itself as the browsing tool, we can develop frameworks for accessing segments and for linking to collections of related material. Beyond other efficiencies, this approach has the elegance of being a visual method to explore visual materials. As desktop computers have come to handle digital video much more easily and the quality greatly improved, as well as the advent of DVD as a popular storage and playback medium for such works, the need to browse visual databases has greatly increased. Explorations of approaches to segment meaningfully such material will not only lead to increased ease of access to parts of these works, but will hopefully also encourage greater understanding of the structures of these materials themselves, of how they are put together, how they entertain and enlighten us, and how we can more greatly appreciate their contents.



## References

- [1] Manovich, L., *The Language of New Media*, MIT Press, 2001.
- [2] Mamber, S., “Hitchcock: The conceptual and the pre-digital,” *Stanford Humanities Review*, vol. 7.2, pp. 128–136, March 1999.
- [3] Virage, “Better business through smarter video<sup>TM</sup>,” <http://www.virage.com>.
- [4] CueVideo, “Visual media management,” <http://www.almaden.ibm.com/cs/cuevideo/>, 1997.
- [5] Salt, B., *Film Style & Technology: History and Analysis*, London: Starword, 1983.
- [6] Burch, N., *Theory of Film Practice*, Princeton University Press, 1981.
- [7] Metz, C., *Film Language: A Semiotics of the Cinema*, University of Chicago Press, 1991.