Media Streams: An Iconic Visual Language for Video Annotation

Bibliographic Reference:

Marc Davis. "Media Streams: An Iconic Visual Language for Video Annotation." In: *Proceedings of 1993 IEEE Symposium on Visual Languages in Bergen, Norway*, IEEE Computer Society Press, 196-202, 1993.

Media Streams: An Iconic Visual Language for Video Annotation

Marc Davis
MIT Media Laboratory
20 Ames Street, E15-494
Cambridge, MA 02139
mdavis@media.mit.edu

Abstract

In order to enable the search and retrieval of video from large archives, we need a representation of video content. Although some aspects of video can be automatically parsed, a detailed representation requires that video be annotated. We discuss the design criteria for a video annotation language with special attention to the issue of creating a global, reusable video archive. Our prototype system, Media Streams, enables users to create multi-layered, iconic annotations of streams video data. Within Media Streams, the organization and categories of the Director's Workshop allow users to browse and compound over 2200 iconic primitives by means of a cascading hierarchical structure which supports compounding icons across branches of the hierarchy. The problems of creating a representation of action for video are given special attention, as well as describing transitions in video.

1. Introduction

The central problem in manipulating video information lies in representing and visualizing video content. Currently, content providers possess large archives of film and video for which they lack sufficient tools for search and retrieval. For the types of applications that will be developed in the near future (interactive television, personalized news, video on demand, etc.) these archives will remain a largely untapped resource, unless we are able to access their contents.

Given the current state of the art in machine vision and image processing, we cannot now, and probably will not be able to for a long time, have machines "watch" and understand the content of digital video archives for us. We are currently able to automatically analyze scene breaks, pauses in the audio, and camera pans and zooms [23, 24, 10, 20, 21, 12], yet this information alone does not enable the creation of a sufficiently detailed representation of video content. In the near term, it is computer-supported human annotation that will enable video to become a rich, structured data type. Therefore, we need to design tools for creators and users of large video archives, so as to enable their contents to be searched and reused.

2. Criteria for video annotation languages

A language for video annotation needs to support the visualization and browsing of the structure of video content as well as search and retrieval. There has been some excellent work in visualizing and browsing video data [23, 24, 22, 20, 12] with which our work has affinity. The limitations of these systems rest in the question of their scalability and, a related problem, their lack of a developed video annotation language. For as visualization and browsing interfaces must accommodate larger and larger video databases, they need to be able to work with video according to its content as well as its structure, and hence, annotation and retrieval become necessary components of the system.

A video annotation language needs to create representations that are durable and sharable. The knowledge encoded in the annotation language needs to extend in time longer than one person's memory or even a collective memory, and needs to extend in space across continents and cultures. German news teams may shoot footage in Brazil for South Korean television which is then accessed by American documentary filmmakers, perhaps ten years later.

Visual languages may enable the design of an annotation language with which we can create a truly global media resource. Unlike other visual languages that are used internationally (e.g., for traffic signage, operating instructions on machines, etc. [11]) a visual language for video annotation can take advantage of the affordances of the computer medium. We can develop visual languages for video that utilize color, animation, variable resolution, and sound in order to create durable and sharable representations of video content.

3. Representing video

3.1 Streams vs. clips

In designing a visual language for video content we must think about the structure of what is being represented. A video camera produces a temporal stream of image and sound data represented as a sequence of frames played back at a certain rate—normally 30 frames per second. Traditionally, this stream of frames is segmented into units called clips. Current tools for annotating video content used in film production, television production, and multimedia, add descriptors (often keywords) to clips. There is a significant problem with this approach. By taking an incoming video stream, segmenting it into various clips, and then annotating the content of those clips, a *fixed segmentation* is imposed on the content of the video stream. To illustrate this idea, imagine a camera recording a sequence of 100 frames. Traditionally, one or more parts of this stream of frames would be segmented into clips which would then be annotated by attaching descriptors to these clips. Imagine a clip made from this stream of 100 frames which extends from frame 50 to frame 80. This clip is a fixed segmentation of the video stream that separates the clip from its context of origin and encodes a particular chunking of the original data. Annotation of this clip does not produce any new segmentations of the video stream.

In our representation, the stream of frames is left intact and is annotated by multi-layered annotations with precise time indexes (beginning and ending points in the video stream). Annotations may be made within any of the various categories for media annotation discussed below (e.g., characters, spatial location, camera motion, dialogue, etc.) or contain any data the user may wish. The result is that this representation makes annotation pay off—the richer the annotation, the more numerous the possible segmentations of the video stream. For example, six annotations of the stream of 100 frames would result in 66 possible segmentations. Clips change from being fixed segmentations of the video stream, to being the results of retrieval queries based on annotations of the video stream. In short, in addressing the challenges of representing video for large archives what we need are representations which make clips, not representations of clips.

3.2 Categories for media annotation

A central question in our research is the development of a minimal set of categories for representing video content. One of the principal things that makes video unique is that it is a temporal medium. Any language for annotating the content of video must have a way of talking about temporal events—the actions of humans and objects in space over time. Therefore, we also need a way of talking about the characters and objects involved in actions as well as their mise-en-scene, that is, the spatial location, temporal location, and weather/lighting conditions. The objects and characters involved in actions in particular settings also have significant positions in space relative to one another (beneath, above, inside, outside, etc.).

These categories—*actions*, *characters*, *objects*, *relative positions*, *locations*, *times*, and *weather*—would be nearly sufficient for talking about actions in the world, but video is a *recording* of actions in the world by a camera, and any representation of video content must address further specific properties. First, we need ways of talking about *cinematographic properties*, the movement and framing of the camera recording events in the world. We also need to describe the properties of the *recording medium* itself (film or video, color or black & white, graininess, etc.) Furthermore, in video, viewers see events depicted on screens, and therefore, in addition to relative positions in space, screen objects have a *screen position* in the two-dimensional grid of the frame and in the various layered vertical planes of the screen depth. Finally, video recordings of events can be manipulated as objects and rearranged. We create transitions in video in ways not possible in the real world. Therefore, *cinematic transition*s must also be represented in an annotation language for video content. In working with video archivists from Monitor Television, we found that in their daily practice (in addition to the above mentioned intersubjective categories) video producers would ask for footage according to highly subjective *thoughts* about the video content which relate to the quality of the frame composition, color, and level of activity.

These categories need not be *sufficient* for media annotation (the range of potential things one can say is unbounded), but we believe they are *necessary* categories for media annotation in order to support retrieval and reuse of particular segments of video data from an annotated stream.

These minimal annotation categories attempt to represent information about media content that can function as a substrate:

- on top of which other annotations may be layered
- out of which new annotations may be inferred
- within which the differences between consensual and idiosyncratic annotations may be articulated

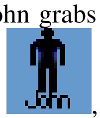
4. Media Streams: An Overview

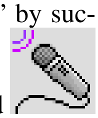
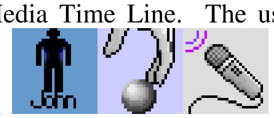
Over the past two years, members of the MIT Media Laboratory's Learning and Common Sense Section (Marc Davis with the assistance of Brian Williams and Golan Levin under the direction of Prof. Kenneth Haase) have been building a prototype for the annotation and retrieval of video information. This system is called *Media Streams*. A paper on an early version of this system was presented at the AAAI-91 Workshop on Intelligent Multimedia Interfaces [10]. Media Streams has developed into a working prototype that is being used by other researchers at the Media Lab in various projects in which content annotated temporal media are required. Media Streams is written in Macintosh Common Lisp and FRAMER [17], a persistent framework for media annotation and description that supports cross-platform knowledge representation and database functionality. Media Streams is being developed on an Apple Macintosh Quadra 950 with three high resolution color displays.

The system has three main interface components: the Director's Workshop (Fig. 1); Icon Palettes (Fig. 2); and Media Time Lines (Fig. 3). The process of annotating video in Media Streams using these components involves a few simple steps:

1) In the Director's Workshop, the user creates iconic descriptors by cascading down hierarchies of icons in order to select or compound iconic primitives.

2) As the user creates iconic descriptors, they accumulate on one or more Icon Palettes. This process effectively groups related iconic descriptors. The user builds up Icon Palettes for various types of default scenes in which iconic descriptors are likely to co-occur, for example, an Icon Palette for "treaty signings" would contain icons for certain dignitaries, a treaty, journalists, the action of writing, a stateroom, etc.

3) By dragging iconic descriptors from Icon Palettes and dropping them onto a Media Time Line, the user annotates the temporal media represented in the Media Time Line. Once dropped onto a Media Time Line, an iconic description extends from its insertion point in the video stream to either a scene break or the end of the video stream. In addition to dropping individual icons onto the Media Time Line, the user can construct compound icon sentences by dropping certain "glommable" icons onto the Media Time Line, which, when completed, are then added to the relevant Icon Palette and may themselves be used as primitives. For example, the user initially builds up the compound icon sentence for "John grabs microphone" by successively dropping the icons 

, , and  onto the Media Time Line. The user then has the compound icon 

on an Icon Palette to use in later annotation. By annotating various aspects of the video stream (time, space, characters, characters' actions, camera motions, etc.), the user constructs a multi-layered, temporally indexed representation of video content.

Media Streams is a large system that attempts to address many questions in video representation. In this paper we focus on Media Streams' iconic visual language for video annotation. It is an iconic visual language that allows composition of iconic primitives in order to form complex expressions. It has a syntax for the composition of iconic sentences and means for extending the visual language.

5. Why icons?

There have been serious efforts to create iconic languages to facilitate global communication [4] and provide international standard symbols for specific domains [11]. We developed Media Streams' iconic visual language in response to trying to meet the needs of annotating video content in large archives. It seeks to enable:

- quick recognition and browsing of content annotations
- visualization of the dense, multi-layered temporal structure of video content
- an accurate and readable time-indexed representation of simultaneous, sequential, overlapping and contained actions (natural languages are not very good at this task)
- articulation of the boundaries between consensual and idiosyncratic annotations (icons can have attached textual annotations and can thus function as the explicit consensual tokens of various idiosyncratic textual descriptions)
- global, international use of annotations
- visual similarities between instances or subclasses of a class (visual resonances in the iconic language)

Media Streams' iconic language encompasses icons which denote both things and actions and thus embodies a distinction analogous to Chang's [8] distinction between object icons and process icons. The difference here is that the objects and processes denoted by the Media Streams' icons are not computational ones, but aspects of the video content which they annotate.

The iconic language gains expressive power and range from the compounding of primitives and has set grammars of combination for various categories of icons. In Korfhage's sense Media Streams is an iconic language as opposed to being merely an iconography [17]. Similar to other syntaxes for iconic sentences [9, 22], icon sentences for actions have the form of subject-action-object or subject-action-direction, while those for relative positions have the form subject-relative position-object. Icon sentences for cinematographic properties are of the form camera-movement-object (as in "the camera-is tracking-Steve" or "the camera-zooms in on-Sally").

6. Director's Workshop

The Director's Workshop is the interface for the selection and compounding of the iconic primitives in Media Streams. To date we have over 2200 iconic primitives. What enables the user to navigate and make use of such a large number of primitives is the way the Director's Workshop organizes these icons into cascading hierarchies. We refer to the iconic primitives in the Director's Workshop as "cascading icons." The Director's Workshop has two significant forms of organization for managing navigational and descriptive complexity:

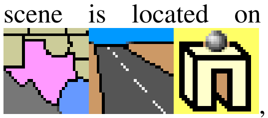
- *Cascading Hierarchy with Increasing Specificity of Primitives on Subordinate Levels*

Cascading icons are organized in hierarchies from levels of generality to increasing levels of specificity. Similar to cascading menus on the Macintosh, when a user cascades down an icon hierarchy by clicking on a cascading icon, its subordinate icons are displayed to the right of the cascading icon. These subordinate icons are arranged *horizon-*

tally and represent an increased level of specificity. Some of the icon hierarchies cascade to as many as 7 or 8 levels deep, yet, similarly to the semantic hierarchies of the CYC Project [18], the design of the categories themselves and their first two or three levels is the hardest and most important representational task.

- *Compounding of Hierarchically Organized Primitives Across Multiple Axes of Description*

In many icon hierarchies on the Director's Workshop, there exists an additional form of organization. When subordinate icons are arranged *vertically*, they represent independent axes of description whose icon hierarchies can be cascaded through separately and whose respective subordinate icons can be compounded together across these axes to form compound iconic descriptors. This form of organization enables a relatively small set of primitives to be compounded into a very large and rich set of descriptors.

To illustrate these forms of organization in our iconic language we can look at how the compound icon for "the scene is located on top of a street in Texas," , was created. Figure 1 shows the cascading icon hierarchy for "space" extended out to the icons for "Texas," "street," and "on top of" which the user compounded to create the icon for "the scene is located on top of a street in Texas" which appears in the Icon Information Editor. The user clicked on the space icon, which cascaded to show its subordinate icons "geographical space," "functional space," and "topological space" *vertically* arranged. Each of these cascading icons has further *horizontally* arranged subordinate icons each of which may go several levels deep. For example, the icons in the path from "geographical space" to "Texas" each represents a distinct level of progressive specification (geographical space->land->continent->North America->United States of America->South-Central United States->Texas). As indicated by the gray square behind the "Texas" icon, it too has further levels of specificity below it which can be displayed by clicking on the icon. In the Director's Workshop, at all but the terminal levels in the hierarchy there exist many icons which themselves have further levels of specification. At any level in the hierarchy, icons can be compounded across the vertical organization to create compound icons. In addition to clicking, cascading icons can be accessed by voice (using the Voice Navigator II™), by typing in text for their names, or by dropping an existing icon onto the Director's Workshop which opens the icon hierarchies up to the terminals of the components of the dropped icon. In all these ways, a vast, structured space of icons can be easily navigated by the user.

It is also important to note that the icon hierarchy of the Director's Workshop is structured not as a tree, but as a graph. The same iconic primitives can often be reached by multiple paths. These paths are important in retrieval because they can guide generalization and specialization of search criteria by functioning as a semantic net of hierarchically organized classes, subclasses, and instances.

6.1 A language for action

The central problem of a descriptive language for temporal media is the representation of dynamic events. For video in particular, the challenge is to come up with techniques for representing and visualizing the complex structure of the actions of characters, objects, and cameras. There exists significant work in the formalization of temporal events in order to support inferencing about their interrelationships [1] and to facilitate the compression and retrieval of image sequences by indexing temporal and spatial changes [2, 3]. Our work creates a representation of cinematic action which these and other techniques could be usefully applied to. For even if we had robust machine vision, temporal and spatial logics would still require a *representation* of the video content because such a representation would determine the units these formalizations would operate on for indexing, compression, retrieval, and inferencing.

A representation of cinematic action for video retrieval and repurposing needs to focus on the granularity, reusability, and semantics of its units. In representing the action of bodies in space, the representation needs to support the hierarchical decomposition of its units both spatially and temporally. Spatial decomposition is supported by a representation that hierarchically orders the bodies and their parts which participate in an action. For example, in a complex action like driving an automobile, the arms, head, eyes, and legs all function independently. Temporal decomposition is enabled by a hierarchical organization of units such that longer sequences of action can be broken down into their temporal subabstractions all the way down to their atomic units. In [18], Lenat points out the need for more than a purely temporal representation of events that would include semantically relevant atomic units organized into various temporal patterns (repeated cycles, scripts, etc.) For example, the atomic unit of "walking" would be "taking a step" which repeats cyclically. An atomic unit of "opening a jar" would be "turning the lid" (which itself could theoretically be broken down into smaller units—but much of the challenge of representing action is knowing what levels of granularity are useful).

Our approach tries to address these issues in multiple ways with special attention paid to the problems of representing human action as it appears in video. It is important to note in this regard—and this holds true for all aspects of representing the content of video—that unlike the project of traditional knowledge representation which seeks to represent the world, our project is *to represent a representation of the world.* This distinction has significant consequences for the representation of human action in video. In video, actions and their units do not have a fixed semantics because their meaning can shift as the video is recut and inserted into new sequences [19, 16]. For example, a shot of two people shaking hands, if positioned at the beginning of a sequence depicting a business meeting, could represent "greeting," if positioned at the end, the same shot could represent "agreeing." Video brings to our attention the effects of context and order on the meaning of represented action. In addition, the prospect of annotating video for a global media archive brings forward an issue which traditional knowledge representation has largely ignored: cultural variance. The shot of two people shaking hands may signify greeting or agreeing in some cultures, but in others it does not. How are we to annotate shots of people bowing, shaking hands, waving hello and good-bye? The list goes on. In order to address the representational challenges of action in video we do not explicitly annotate actions according to their particular semantics in a given video stream (a shot of two people shaking hands is not annotated as "greeting" or alternately as "agreeing"), but rather according to the motion of objects and people in space. We annotate using physically-based description in order to support the reuse of annotated video in different contexts—be they cinematic or cultural ones. We create analogy mappings between these physically-based annotations in their concrete contexts in order to represent their contextual synonymy or lack thereof.

We represent actions for characters and objects separately in the Director's Workshop because of the unique actions afforded by the human form. Our icons for action are *animated* which takes advantage of the affordances of iconography in the computer medium as opposed to those of traditional graphic arts.

We *horizontally* subdivide character actions into full body actions, head actions, arm actions, and leg actions. Under each of these categories of human action (and their own subdivisions) action is represented in two ways: *conventionalized* physical motions and *abstract* physical motions.

We built into our ontology many commonly occurring, complex patterns of human motion which seem to have cross-cultural importance (e.g. walking, sitting, eating, talking, etc.). We also provide a hierarchical decomposition of the possible motions of the human body according to articulations and rotations of joints. Since Media Streams enables multi-layered annotation, any pattern of human motion can be described with precision by layering temporally indexed descriptions of the motion of various human body parts.

6.2 Transitions

Transitions between shots are both the tools editors use to construct scenes and sequences out of a series of shots, and the gaps in a video stream of recorded space-time which are bridged by the viewer's inferential activity [6].

For example, if a viewer sees the two shot sequence of a person entering an elevator and then exiting an elevator, the viewer infers that a certain amount of time has passed and that a certain type of spatial translation has occurred. Noël Burch has developed a systematic categorization of spatio-temporal transitions between shots in cinema [7]. He divides temporal transitions into: continuous; forward ellipses in time of a determinate length; forward ellipses of an indeterminate length; and the corresponding transitions in which there is a temporal reversal. Spatial transitions are divided into: continuous; transitions in which spatial proximity is determinate; and transitions in which spatial proximity is indeterminate. Burch's categorization scheme was used by Gilles Bloch in his groundbreaking work in the automatic construction of cinematic narratives [5]. We adopt and extend Burch's categorization of shot transitions by adding "temporal overlaps" as a type of temporal transition and the category of "visual transitions" for describing transition effects which unlike traditional cuts, can themselves have a duration (icons for transition effects which have durations are animated icons).

When a transition icon is dropped on a Media Time Line, Media Streams creates a compound icon in which the first icon is an icon-sized (32 x 32 pixels, 24 bits deep) QuickTime Movie containing the first shot, the second icon is the transition icon, and the third icon is an icon-sized QuickTime Movie containing the shot after the transition.

A search using transition icons would enable the user to find a "matching" shot in the following way: The user could begin with a shot of a person getting into an automobile and use one or more of the transition icons as analogical search guides in order to retrieve a shot of the person exiting the automobile in a nearby location. The query would have expressed the idea of "find me a Shot B which has a similar relation to Shot A as Shot D has to Shot C."

7. Extensibility of the icon language

Currently, we have two ways of extending the iconic visual language of Media Streams beyond the composition of iconic primitives. Icons and the components of compound icons can be titled. This enables the user to attain a level of specificity of representation while still making use of the generality and abstraction of icons. For example, if I were to annotate the video of an automobile with the descriptor "XJ7," this description may be very opaque. If, however, I title a car icon XJ7, in addition to the computer learning that XJ7 is a type of car, which is what the computer can then learn, a human reading this annotation, can see simply and quickly the similarity between an XJ7 and other types of automobiles. A form of system maintenance would be to periodically find titles for which there are many occurrences and create an icon for them.

Users can also create new icons for character and object actions by means of an animated icon editor (Fig. 4). This editor allows users to define new icons as subsets or mixtures of existing animated icons. This is very useful in conjunction of our complete body model, because a very wide range of possible human motions can be described as subsets or mixtures of existing icons.

Applying the results of work on automatic icon incorporation would be a fruitful path of exploration [13]. Already in our icon language, there are many iconic descriptors which we designed using the principle of incorporation (by which individual iconic elements are combined to form new icons). Creating tools to allow users to automatically extend the language in this way is a logical extension of our work in this area.

8. Conclusions and future work

Media Streams is about to be subjected to some rigorous real-world tests. In addition to several internal projects at the MIT Media Laboratory which will be building other systems on top of Media Streams, external projects involving large archives of news footage will be exploring using Media Streams for video annotation and retrieval. Clearly these internal and external projects will teach us much about the claim made in this paper: that an iconic visual language for video annotation and retrieval can support the creation of a stream-based, reusable, global archive of digital video. We believe that this goal articulates an important challenge and opportunity for visual languages in the 1990's [14]. Our next step is to use our system to create an archive of annotated digital video (approx. 5 Gig) in order to explore mechanisms for computational video storytelling which was the goal that originally inspired the creation of Media Streams out of the necessity of having a representation of video content.

Acknowledgments

The research discussed above was conducted at the MIT Media Laboratory. The support of the Laboratory and its sponsors is gratefully acknowledged. I want to thank Brian Williams and Golan Levin for their continually creative and Herculean efforts and my advisor, Prof. Kenneth Haase, for his insight, inspiration, and support. Thanks also to Warren Sack, Axil Comras, and Wendy Buffett for editorial and moral support.

References

1. Allen, J.F., "Maintaining Knowledge about Temporal Intervals," in *Readings In Knowledge Representation,* R.J. Brachman and H.J. Levesque, Editor. Morgan Kaufmann Publish-ers, Inc.: San Mateo, California. p. 510-521. 1985.
2. Arndt, T. and S.-K. Chang. "Image Sequence Compression by Iconic Indexing." In: *Proceedings of 1989 IEEE Workshop on Visual Languages*. Rome, Italy: IEEE Computer Society Press. p. 177-182. 1989.
3. Bimbo, A.D., E. Vicario, and D. Zingoni. "A Spatio-Temporal Logic for Image Sequence Coding and Retrieval." In: *Proceedings of 1992 IEEE Workshop on Visual Languages*. Seattle, Washington: IEEE Computer Society Press. p. 228-230. 1992.
4. Bliss, C.K., *Semantography-Blissymbolics*. 3rd ed. Sydney, N.S.W., Australia: Semantogra-phy-Blissymbolics Publications. 1978.
5. Bloch, G.R., *From Concepts to Film Sequences*. Yale University Department of Computer Science: 1987.
6. Bordwell, D., *Narration in the Fiction Film*. Madison: University of Wisconsin Press. 1985.
7. Burch, N., *Theory of Film Practice*. Princeton: Princeton University Press. 1969.
8. Chang, S.-K., "Visual Languages and Iconic Languages," in *Visual Languages,* S.-K. Chang, T. Ichikawa, and P.A. Ligomenides, Editor. Plenum Press: New York. p. 1-7. 1986.
9. Chang, S.-K., *et al*. "A Methodology for Iconic Language Design with Application to Augmentative Commmunication." In: *Proceedings of 1992 IEEE Workshop on Visual Languages*. Seattle, Washington: IEEE Computer Society Press. p. 110-116. 1992.
10. Davis, M. "Director's Workshop: Semantic Video Logging with Intelligent Icons." In: *Proceedings of AAAI-91 Workshop on Intelligent Multimedia Interfaces*. Anaheim, California: p. 122-132. 1991.
11. Dreyfuss, H., *Symbol Sourcebook: An Authoritative Guide to International Graphic Symbols*. New York: McGraw-Hill. 1972.
12. Elliott, E.L. "WATCH • GRAB • ARRANGE • SEE: Thinking with Motion Images via Streams and Collages." M.S.V.S., Massachusetts Institute of Technology Media Laboratory. 1993.
13. Fuji, H. and R.R. Korfhage. "Features and a Model for Icon Morphological Transformation." In: *Proceedings of 1991 IEEE Workshop on Visual Languages*. Kobe, Japan: IEEE Computer Society Press. p. 240-245. 1991.
14. Glinert, E., M.M. Blattner, and C.J. Frerking. "Visual Tool and Languages: Directions for the 90's." In: *Proceedings of 1991 IEEE Workshop on Visual Languages*. Kobe, Japan: IEEE Computer Society Press. p. 89-95. 1991.
15. Haase, K. and W. Sack, *FRAMER Manual*. MIT Media Laboratory: Cambridge, Massachusetts. 1993.
16. Isenhour, J.P., "The Effects of Context and Order in Film Editing." *AV Communications Review*, 23(1): p. 69-80. 1975.
17. Korfhage, R.R. and M.A. Korfhage, "Criteria for Iconic Languages," in *Visual Languages,* S.-K. Chang, T. Ichikawa, and P.A. Ligomenides, Editor. Plenum Press: New York. p. 207-231. 1986.
18. Lenat, D.B. and R.V. Guha, *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*. Reading, Massachusetts: Addison-Wesley Publishing Company, Inc. 1990.
19. Levaco, R., ed. "Kuleshov on Film: Writings by Lev Kuleshov." University of California Press: Berkeley. 1974.
20. MacNeil, R. "Generating Multimedia Presentations Automatically Using TYRO: the Constraint, Case-Based Designer's Apprentice." In: *Proceedings of 1991 IEEE Workshop on Visual Languages*. Kobe, Japan: IEEE Computer Society Press. p. 74-79. 1991.
21. Mills, M., J. Cohen, and Y.Y. Wong. "A Magnifier Tool for Video Data." In: *Proceedings of CHI'92*. Monterey, California: p. 93-98. 1992.

22. Tanimoto, S.L. and M.S. Runyan, “PLAY: An Iconic Programming System for Children,” in *Visual Languages,* S.-K. Chang, T. Ichikawa, and P.A. Ligomenides, Editor. Plenum Press: New York. p. 191-205. 1986.

23. Tonomura, Y. and S. Abe. “Content Oriented Visual Interface Using Video Icons for Visual Database Systems.” In: *Proceedings of 1989 IEEE Workshop on Visual Languages*. Rome, Italy: IEEE Computer Society Press. p. 68-73. 1989.

24. Ueda, H., T. Miayatake, and S. Yoshizawa. “IMPACT: An Interactive Natural-Motion-Picture Dedicated Multimedia Authoring System.” In: *Proceedings of CHI '91*. New Orleans, Louisiana: ACM Press. p. 343-350. 1991.

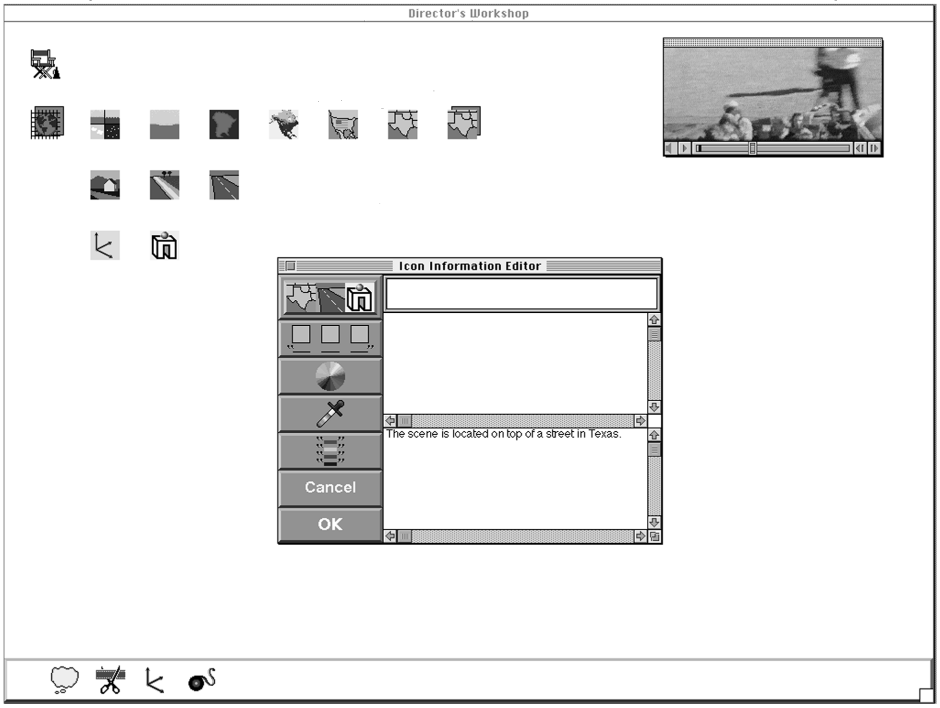


Figure 1: The Director's Workshop

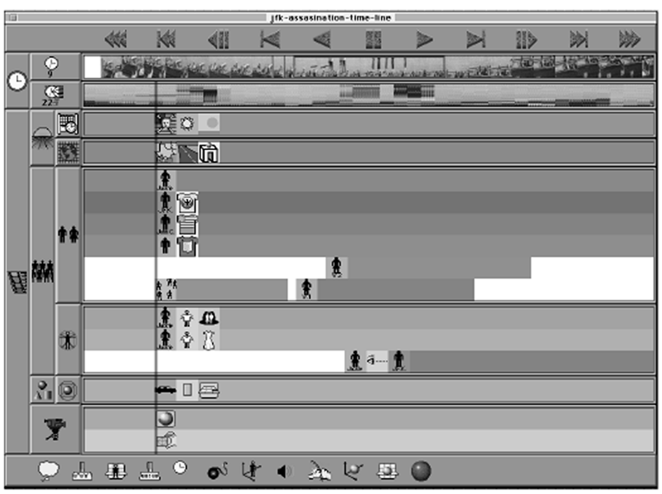


Figure 2: JFK Assasination Media Time Line

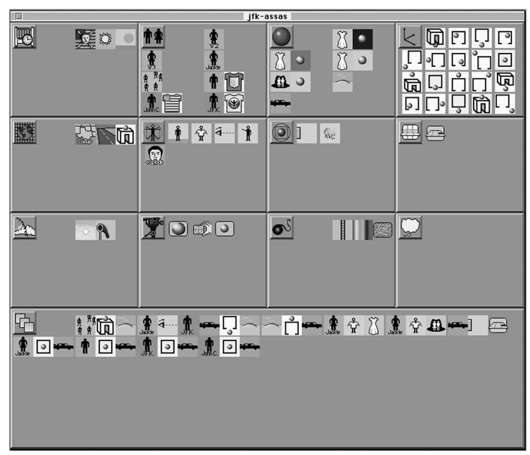


Figure 3: JFK Assasination Icon Palette

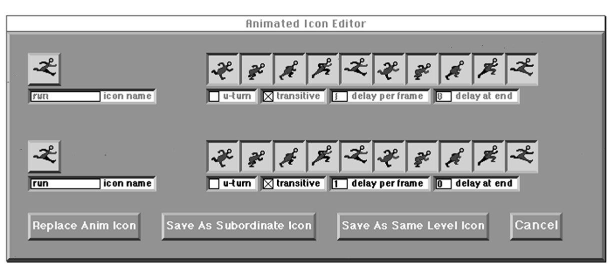


Figure 4: Animated Icon Editor for "run"