Video Knowledge Bases and Computer-Assisted Video Production: A Forestry Example

Alan J. Thomson and C. Sivertson
Pacific Forestry Centre
506 West Burnside Road
Victoria, B.C., Canada V8Z 1M5

ABSTRACT. A system has been developed to assist in production of customized technical videos in response to user queries, analogous to customized text-generating programs. The domain of the system is forest insect and disease biology, management, and impact. A hypermedia system for query processing provided greater flexibility than could be achieved through database forms. A processed query generated a video edit list used to automate the video assembly process. The system has been used to assist in the production of in-house technical videos. Linking computers and video brings the power of database and AI technologies to the video production process, and also permits video to become an integral part of decision support sytems, with video being the output of a knowledge-based system, or linked to simulation models or GIS. Production of bilingual videos is also facilitated.

The increasing interaction of video and computing technologies is exemplified by the frequent appearance of articles on video in computer magazines such as *BYTE* (Lechner 1992, Yager 1993) and articles on computers in video magazines such as *Videography* (Kohler 1991), with magazines such as *Desktop Video World* and *Video Toaster User* dealing specifically with that interaction. Books such as Luther (1991) provide a good introduction to the field, and integration of video devices into computer networks is discussed by Nadeau and Bailey (1992). Digital video is rapidly becoming simply another software data type, and desktop video systems make video inclusion economically feasible.

A video may be the output of a system similarly to numeric or graphical output. For example, multimedia or hypermedia systems (Rauscher and Host 1990, Rauscher and Johnston 1991, Thomson et al. 1993) could include video as well as text and graphics. We have already linked video laserdiscs to GIS. Video could become a integral part of decision support systems, which could output a video that describes a recommended action. Buhyoff and Fuller (1993) discuss the need for improved explanation of quantitative models; by linking simulation models to a laser videodisc or video tape player, video segments could be assembled based on a knowledge-based system's evaluation of the initial parameters, the management actions simulated, and the output of the model to provide, for example, a walkthrough of a forest during a rotation showing the management actions and their expected results. Predicted pest and disease effects could also be visualized.

Such a procedure could be made fully interactive, permitting the user to select objects displayed in the video segments and interrogate the system about their properties. Parkes and Self (1990) discuss such a process in a description of a video-based intelligent tutoring environment. However, this type of system requires detailed within-frame knowledge of object positions and movement in order that the computer can relate the objects in the video to the part of the screen selected.

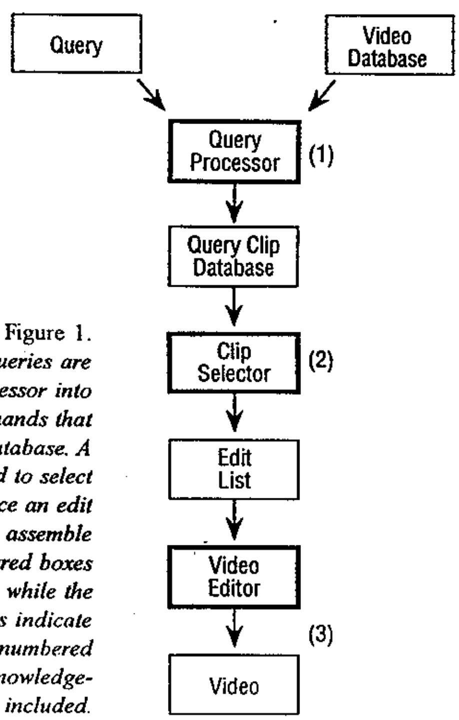

Figure 1.
System overview. User queries are translated by a query processor into video database commands that produce a query clip database. A database form is used to select specific clips and produce an edit list used to automatically assemble the video. The thin-bordered boxes are input/output products, while the thick-bordered boxes indicate software processing. The numbered positions are where knowledge-based systems could be included.

Expert systems that produce customized documents containing text-based knowledge in response to client-specific queries have been described by Gast et al. (1988), Thomson and Taylor (1990), and Thomson and Williamson (1992). These expert systems constructed sentences and paragraphs from fragments of text which represent the basic units of knowledge in the system, including decisions on page layout, grammar, and punctuation. Segments of video may be pieced together in an analogous manner.

The above discussion assumes that the video segments are of high quality, with appropriate sound track. However, production of these video segments, or production of a whole video, from raw video footage is a different problem; segments of video may need to be selected from a wide choice of available footage. In the present study, we describe a project in which a computer system was developed to assist in video segment production in response to specific queries. These segments could be linked to produce a complete, customized, technical video. The domain of the queries was the biology, impact, and management of forest insects and diseases.

The underlying principle of the system (Fig. 1) is a query processed into the format of a database search, where the database contains knowledge of video material and its location. Contents of the database records are then automatically converted into commands for the video-editing software which is used to assemble the final video. At present, the system is designed to assist in the processing of the raw video footage taken by the forest rangers of the Forest Insect and Disease Survey (FIDS) of the Pacific Forestry Centre (PFC). Also included in the system are videotapes of slides from the FIDS slide collection, which incorporates many thousands of images.

Because the video database is central to the system, we start by discussing the nature of video knowledge and its consequences for the design of such databases. We then describe the database and a hypermedia-based database query system. Finally, we describe a database form for final processing of the results of the query, and show how the video assembly is automated.

The Nature of Video Knowledge

In a technical sense, the frame is the basic unit of video. Every frame can be referenced by a time code of the form HH:MM:SS:FF where H is hour, M is minute from 0 to 60, S is second from 0 to 60, and F is frame from 0 to 30. However, the frame is not the smallest unit of meaning because the audio track can direct attention to particular parts of a frame, or can assign many meanings to the same visual material. On the other hand, as frames are normally played at 30 per second in video, a sequential set of frames is necessary to make up a meaningful video segment that can be discerned from adjacent segments when played. Terms such as "shot" and "clip" are used to define such segments.

Many texts have pursued the comparison of film and spoken/written language, suggesting equivalence between the shot and the word, the scene and the sentence, the sequence and the paragraph. However, such analogies break down under close examination, which suggests that video does not consist of discrete units, but is rather a continuum of meaning. "A shot contains as much information as we want to read in it, and whatever units we want to define within the shot are arbitrary" (Monaco 1981). The continuous nature of video knowledge poses a challenge to database design. For the purposes of this study, we do not deal with knowledge below the level of the frame, and we concentrate on visual databases rather than audio databases.

Computer Hardware and Software, and System Use

The project was completed on a Commodore Amiga 2000 computer equipped with Newtek's Video Toaster™, which provides computer-controllable video switching capabilities. The Amiga was selected for its custom video chips, the availability of third-party video processing hardware and software, and the incorporation of the ARexx language as part of the Amiga operating system.

ARexx is an implementation of the REXX language, originally developed for IBM, for multitasking Amiga computers (Zamara and Sullivan 1991). Scripts written in ARexx can control other programs such as databases, spreadsheet packages, graphics, multimedia, and video applications. For example, a value could be automatically extracted from a database or spreadsheet, inserted into video titling software, and overlaid on a particular video segment, using a particular video transition (wipe), all under computer control. The availability of ARexx scripts facilitates development of expert systems for use at several places in the video production process (Fig. 1). ARexx compatibility was an important feature in selecting software products used in the system.

The database was designed in SuperBase Professional 4™, the query interface was developed in Gold Disk's Hyperbook™ (a hypermedia authoring tool), and video editing was carried out using Gold Disk's Video Director™ software. The Video Director software is not ARexx-controllable, but had other features that governed its selection.

Luther (1991) indicates that the list of skills to work with an audio/video/computer system is very long, including audio, video and computer engineering, audio/video production and postproduction, artistic design, graphic art, script writing, creative writing, publishing, programming, and image processing. The aim of the present system was to facilitate video production by an individual, rather than requiring a full production crew. Ultimately, technical video production should be possible by an individual with no video expertise, being simply another form of system output.

Because we did not expect clients to have the necessary equipment for producing videos, we envisaged queries being processed at a central location, with a video being sent to the client in the same manner as text-based output was generated in the study described by Thomson and Taylor (1990).

The Video Database: Video-Related Fields

The basic unit of knowledge in the database is the "clip," defined here as the section of video between two time codes. Fields in the database

(Table 1) fall into two main categories: video-related knowledge and biology-related knowledge. As indicated previously, video does not consist of discrete units, but is rather a continuum of meaning. The evolution of the meaning of "clip" during the project is discussed below.

Defining the database fields in the sequence shown in Table 1, the first field is a video tape reference (**tape_ref**), after which the time code at the start and end of the clip is specified. The fields **tc_deck_start_** and **tc_deck_end** define the start and end with the standard time code (SMPTE) as used by the video editing deck and some cameras. Other cameras in the study use a SONY™-specific time code, represented by **cam_tc_start** and **cam_tc_end.** The same clip can have both time codes associated with it. A combination of the tape reference and SMPTE start and end codes (**deck_start_end**) provides a unique identifier for use by the database indexing and the video assembling software.

Knowledge about effects (**effects_code**) such as zoom-in, zoom-out, pan-left, pan-right, tilt-up or tilt-down, and their timing (**effect_start** and **effect_end**) was the aspect of the database that evolved most over the course of the project. Initially, "clip" was defined as the section of tape covered by one effect; thus, for example, the exact frame at which a zoom started and ended was recorded. Determining the exact frames of an effect is time consuming, and when every effect is represented by a separate record, it has the effect of multiplying the records in the database while not providing any practical advantage for our study. In the video editing process, material is generally previewed for suitability in relation to the overall script, and, in addition, the duration is tailored to the narration.

"Clip" was therefore redefined as the section of tape between two cuts (where the recording is stopped). Codes for all effects included in the clip were combined in a single index, and approximate rather than exact time codes were used. While this was much more practical than the original definition, some problems remained relating to the acquisition of footage by untrained camera operators. In some cases, a clip would be over five minutes in length, with multiple effects and including a wide range of biological content. In other cases, many brief clips would be taken of the same scene.

The final definition of clip was "a continuous section of tape for which the information content can be defined in a single record using the permitted options for the biology-related fields." This operational definition focuses on the biological content rather than on properties affected by mechanical or operator issues. Its utility depends on the fact that raw footage is previewed and tailored during the video editing and assembly process. As well as the start and end of clips, key frames were recorded. These were frames at arbitrary positions within especially good sections. The preview of material around these key frames facilitated selection of material during the editing process. This approach entails a tradeoff of duration-estimation at preview time against estimation at data entry time.

The duration of a clip in seconds, to one decimal place, is stored in the field **tc_duration**. This process required development of a program for performing time code arithmetic such as subtracting one time code from another and con-

Table 1. *Fields[a] in the video database for raw footage are in two main categories: video-related knowledge and biology-related knowledge. Four fields are database-related.*

Video-related fields	**Biology-related fields**
tape_ref	pest_species
tc_deck_start	pest_stage
tc_deck_end	pest_type
deck_start_end	host_species
cam_tc_start	host_part
cam_tc_end	region
effects_code	comments
effect_start	
effect_end	**Database-related fields**
tc_duration	
platform	videopicture
source	hampicture
date	include_picture
photo_credit	sequence_order
quality	

[a] Field names are indicated in bold type in the body of the text.

verting the results to a decimal in units of seconds or minutes. The program was written in ARexx.

The **platform** field (Table 1) provides information on the camera mounting, with three possible values (air, hand, and stand) at present. Hand-held shots are generally less suitable than tripod-mounted shots because it is almost impossible to keep the camera steady. This is also reflected in the field **quality**. However, if the required length of video is very short, or if it is possible to use a single frame in freeze-frame mode, then segments of such footage may be useful. The entry "air" is for footage from aircraft or helicopters.

"Stand" in the **platform** field is generally a tripod, and for the purposes of our system it makes little difference if other forms of support are used. However, Zeitl (1992) illustrates many types of mounting, each with different properties that might be suitable for making inferences under certain circumstances in knowledge-based systems.

The system is currently evolving in relation to re-evaluation of the way in which discrete fields are used to capture the continuum of meaning discussed earlier; thus, for example, the term "still" is no longer a valid entry for the **platform** field, although it is still indicated in Fig. 2. The knowledge imparted by "still" is now captured in the **source** field. Reconciliation of field contents will be done at a later stage when the system moves from research mode to operational use. **Source** indicates if the footage is live or obtained from a color slide, photograph, map, or diagram. **Date** and **photo_credit** are self-explanatory. Note that the fields used and the possible entries in the fields reflect the previously stated idea that video knowledge is a continuum.

The Video Database: Biology-Related Fields

The fields **host_species** and **pest_species** are largely self-explanatory, with the common name used for all interface purposes. Scientific names and authorities can easily be included through the relational properties of the database. **Pest_stage** is a description of developmental stage, such as egg or larva for insects, or fruiting body or mycelium for fungi. This field may be blank if no organism is visible, as with trees illustrating symptoms of a problem. This latter possibility is covered by the field **pest_type**, which has possible values such as organism, defoliation, rot, or mortality. The part of the host illustrated (**host_part**) ranges in scale from needle through shoot, branch, and tree to tree-group, stand, and forest, the latter being most appropriate for aerial shots or for shots obtained across lakes or wide valleys.

The geographic **region** field is one of the most important fields for customizing videos for particular clients as this permits selection of footage from landscapes, familiar to the client, that may illustrate features of pest biology not present in other areas. In addition, where a host species may occur over a wide range of regions, there may be regional differences in stem density or undergrowth characteristics that are not covered by the database fields but that can be inferred from the region.

Finally, the **comments** field is used to record any relevant information not included in the other fields. For example, at present, all indications of forest management practices are recorded here.

The Video Database: Database-Related Fields

As indicated in the system overview (Fig. 1), a database form is used to select specific clips. The form includes a representative image from the clip (Fig. 2). A visual representation of the material in the clip contains more information than any verbal description. The digital representation of the clip permits rapid screening of material from a wide range of tapes. The example shown in Figure 2 is the digital representation of a single frame from a videotape of slide material. When the clip is of moving footage, including effects such as pans and zooms, a single frame may be insufficient to adequately represent the clip contents. In such cases, a single digital image may be composited from a number of frames from the video to represent the range of visual material included in the clip.

Due to image palette considerations, a greyscale image is included on the form. The field **videopicture** is the field name (including directory path) of that image. Selecting the button "COLOR PIC" on the form displays a 4096 color image in Amiga hold-and-modify graphics (HAM) format. The field **hampicture** is the file name of the color image. When a specific clip is selected, that information is stored in the field **include_picture**, while information regarding the desired output sequence of clips is stored in the field **sequence_order**.

The Hypermedia Query Interface

At present, system queries are closely tied to the database, with the hypermedia interface permitting selection of specific values from selected fields. The scrolling lists of possible values (Fig. 3) are obtained from the video database. Leaving a choice blank indicates that all values of that field are permissible. For example, selecting "western larch" from the host species list and "larch casebearer" from the pest species lists, then selecting the "Query the Database" button, creates a subset of the video database containing all clips referring to larch casebearer on western larch.

When the power of hypermedia systems (Rauscher and Host 1990, Rauscher and Johnston 1991, Thomson et al. 1993) is linked to the power of ARexx, we found that a hypermedia query system provided greater flexibility for a database query system than could be achieved using database forms; i.e., the hypermedia system generated a database query via ARexx. This facilitates future expansion of the system to provide expert system guidance to clip selection in a manner similar to that described by Thomson et al. (1993).

The Video Clip Selection Form

Although a hypermedia system was best for generating the database query, a database form was found to be the optimal approach to select-

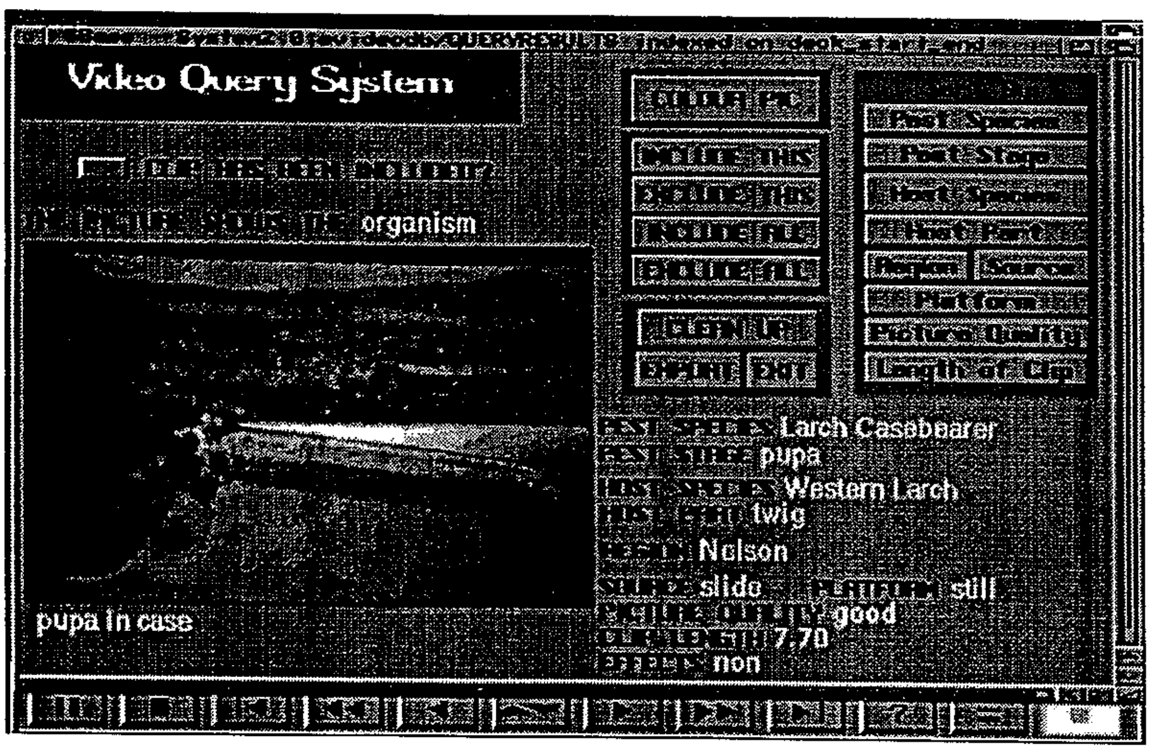

Figure 2. *The database form used to select or exclude specific clips from the subset of records resulting from the original query. Light-colored text is the content of database fields in the current record.*

ing among the records resulting from the query. There are four main sections to the form (Fig. 2):

- Information from current record (including picture)
- Record browsing controls
- Clip selection/exclusion panel
- Clip sorting panel

Only part of the knowledge in a record (Table 1) is displayed on the form. All of the biology-related fields are included, the text "pupa in case" being the contents of the **comments** field. With regard to the database-related fields, the image stored in the file in the **videopicture** field is shown, and the color picture from the **hampicture** field is accessible through the "COLOR PIC" button. The status of the **include_picture** field is indicated in the check box at the top left of the screen, labeled "CLIP HAS BEEN INCLUDED?" Only five of the 15 video-related fields are displayed: **source, platform, quality, tc_duration** (clip length) and **effects_code**.

The VCR-like controls at the base of the screen (Fig. 2) permit the database to be browsed like a video, with fast-forward, pause, and rewind analogies. The three right-most buttons permit Boolean searches of the database and enable/disable the linkage of the image file names to the actual image display.

All clips relevant to the original query are available at this stage, with inclusion/exclusion being possible in two modes. All clips can be initially included using the "INCLUDE ALL" button, then specific clips excluded using the "EXCLUDE THIS" button. Alternatively, all clips can be initially excluded (the default condition), and specific clips included as desired.

The selected clips at this stage are in the order specified by the default indexing mode of the database (**deck_start_end**). Sorting in relation to the theme of the video segment can be accomplished by use of the "Sort By:" buttons (Fig. 2). When clip selection and sorting is com-

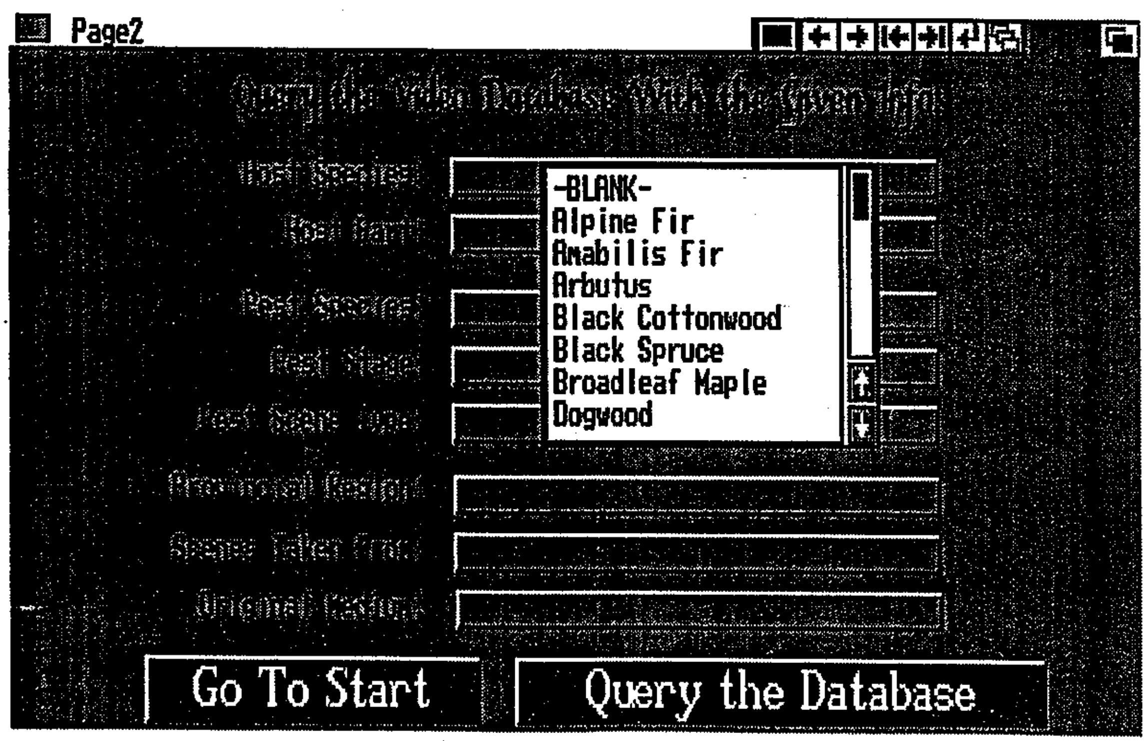

Figure 3. *The hypermedia query interface illustrating the host species selection list.*

plete, the "EXPORT" button is used to process the database knowledge into a form suitable for the video-editing software.

Video Editing and Assembly

Selecting the button "EXPORT" on the clip selection form (Fig. 2) prepares the database knowledge for use by the video-editing software, writing it as an ASCII file, and starts up the program Video Director™ (Fig. 4), which has three main sections. Control and calibration of the source (player) and destination (recorder) of the video material are carried out using the top right panel. Editing of the start and end time codes of particular clips is carried out using the top left panel. This enables clips to be timed to a script, as described below. The bottom panel (Fig. 4) shows (in white) the data exported from the database and imported into the video editing software. The column of black numbers to the right of the panel is the cumulative duration of all clips to that point.

Apart from the time code start and end, the imported material is a unique clip comment combining the tape identifier with an abbreviated description of the clip contents. For example, the first record indicates larch casebearer (LCB) pupae (pup) on western larch (WL) twigs (tw), using a zoom-in (zi) effect. The four tool icons on the left of this panel permit cutting, copying, and pasting of clips. If the clips had been sorted by stage at the clip-selection screen as described earlier, the order would have reflected this choice. When the desired arrangements and timings of clips have been defined, selection of the "ASSEM" button assembles the final videotape segment from the raw footage on the player onto the recording device.

The "EXPORT" button (Fig. 2) described above results in creation of an ASCII file which, at present, must be imported manually using the menu bar's load file option. The final set of clips and timings actually used can in turn be exported

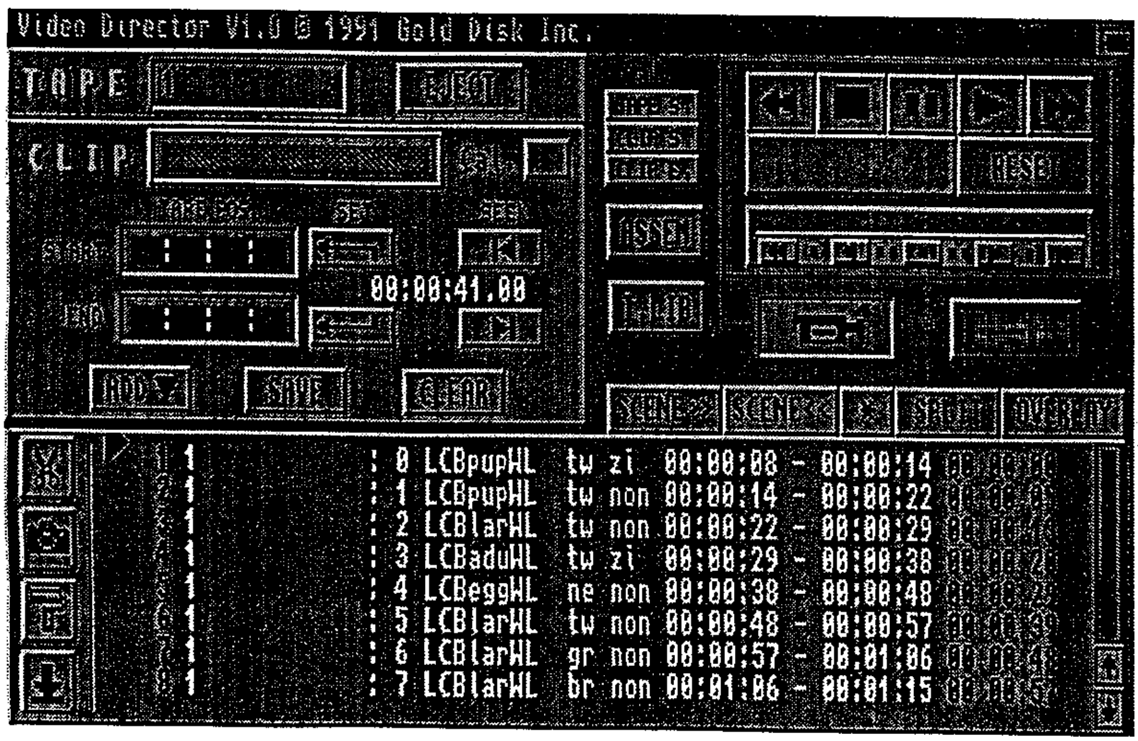

Figure 4. *The video-editing screen illustrating data imported from a processed query regarding larch casebearer on western larch.*

as an ASCII file by the Video Director™ software. Appending these ASCII files in different orders can be used to generate customized videos. The sequence in which they are appended could be determined by a knowledge-based system.

Relating Clips to the Script

Technical videos closely follow a prepared script. The script can therefore be narrated and the sound track laid down prior to assembling the related video footage. The time codes associated with natural breaks in the text, such as commas, periods, or paragraphs, can be determined because these indicate potential changes in requirements for visual material. The start and end time codes for selected clips (Fig. 4) are then tailored to fit the video to these breaks.

Discussion

A database of video knowledge was central to our concept of providing computer assistance to the video production process. However, video knowledge follows a continuum of meaning, posing challenges for database design. Our original design was based on an assumed close analogy between video and text, with the video clip being analogous to a phrase. This design rapidly proved unwieldy for working with raw footage, leading to evolution of the database design guided increasingly by operational considerations. Expert knowledge in the biological domain was required to design the set of fields and define the set of possible values for each field, as well as to determine the values used to define each clip in the knowledge base, including the time code start and end.

The problem of defining where one clip ends and a new one starts is a feature that distinguishes the general video footage, such as was obtained in the present study by forest rangers, from footage obtained specifically in relation to a script. Data entry for general video is facilitated if the camera operator gives a continuous commentary describing the subject material, especially if the commentary uses the permitted values of the database fields. A new sound track based on the script will ultimately replace the commentary in the final video.

Computer assistance in producing video segments is mediated by a query system, where the queries are largely based on biological considerations (in the domain of forest insect and disease biology, impact, and management in this case), with the query format being closely linked, at present, with the database field contents. Many of the database fields relate to the cataloguing of the position of the clips in different time code formats. These are hidden from the system user, but are used for the automation of the video production. The visual nature of the material is captured in digital form by both greyscale and color images included in the record browsing, clip selection interface. These images capture more information than any verbal description.

At present, the query format is closely related to the fields of the video database, and for raw footage this is probably most appropriate. However, if clips were pre-edited to ensure consistent high quality and to have a close relationship to a restricted domain, such as diagnosis and management of a particular pest, then knowledge-based systems at positions (1) and (2) in Figure 1 could be used to produce a brief video giving an analysis of the situation and a case-specific treatment in a manner analogous to the text output of current systems (Gast et al. 1988, Thomson and Taylor 1990, Thomson and Williamson 1992). The video produced could be enhanced by additional knowledge-based systems at position (3) to determine the most appropriate visual effects to use to combine video segments using devices such as Newtek's Video Toaster™ video switcher. Knowledge-based systems could also be used to assist in obtaining the original footage, for example, to advise on lighting, lenses, and filters. Other possible applications described in the present study include output of a video by a knowledge-based system, or linking video to simulation models and GIS.

Knowledge-based systems could also be used to piece together digitized speech segments. For example, when a number is obtained from a database or spreadsheet, fuzzy logic could be used to relate it to an appropriate segment of sound: 1327 ha would then become "more than 1000 hectares," while 1879 ha would become "almost 2000 hectares." Alternatively, phrases such as "has increased significantly since last year" could be used. It is the audio track that provides the interpretation of the scene in terms of the system's domain. Thomson (1993) discusses the manner in which environmental values can change rapidly. Such changes would be reflected in different audio tracks being associated with the same scene.

Linking computers and video brings the power of database and AI technologies to the video production process, and also permits video to become an integral part of decision support systems. The automation of many aspects of technical video production permits customized videos to be produced in a cost-effective manner to meet the needs of specific clients. This includes assistance in production of bilingual videos, where the duration of each phrase in the script varies between languages, but the sequence of clips would remain the same, with the start and end timings adjusted appropriately.

Acknowledgment

This study was supported by the Canada/British Columbia Forest Resource Development Agreement study FC-ETT 17.

References

Buhyoff, G.J., and L.G. Fuller. 1993. Explanation of quantitative models: A need and an example. AI Applications 7(1): 37-43.

Gast, S.J., M. Stock, and L. Livingston. 1988. Customized document generation for forestry extension. AI Applications 2(1): 53-56.

Kohler, C. 1991. Amiga: The video natural. Videography, April 1991 (Supplement): 3-6.

Lechner, B.J. 1992. HDTV is coming to desktop. Byte 17(7): 189-192.

Luther, A.C. 1991. Digital video in the PC environment. McGraw-Hill Book Company, New York.

Monaco, J. 1981. How to read a film. The art, technology, language, history, and theory of film and media. Oxford University Press, New York.

Nadeau, D.R., and M.J. Bailey. 1992. Network video device control. Pages 299-306 in: Proceedings, Visualization '92. IEEE Computer Society Press, Los Alamos, California.

Parkes, A.P., and J.A. Self. 1990. Towards "interactive video": A video-based intelligent tutoring environment. Pages 56-82 in: Intelligent tutoring systems: At the crossroad of artificial intelligence and education, C. Frasson and G. Gauthier, editors. Ablex Publishing Corporation, Norwood, New Jersey.

Rauscher, H.M., and G.E. Host. 1990. Hypertext and AI: a complementary combination for knowledge management. AI Applications 4 (3): 58-61.

Rauscher, H.M., and S. Johnson. 1991. Authoring hypertext documents using Hyperwriter!™ AI Applications 5(2): 116-119.

Thomson, A.J. 1993. Paradigm green: AI approaches to evaluating the economic consequences of changing environmental viewpoints. AI Applications 7(4): 61-68.

Thomson, A.J., and C.M.A. Taylor. 1990. An expert system for diagnosis and treatment of nutrient deficiencies of sitka spruce in Great Britain. AI Applications 4 (1): 44-52.

Thomson, A.J., and D.R. Williamson. 1992. Formation and use of intermediate inferences in advisory systems: a herbicide example. AI Applications 6(4) 29-37.

Thomson, A.J., J.R. Sutherland and C. Carpenter. 1993. Computer-assisted diagnosis using expert system-guided hypermedia. AI Applications 7(1): 17-27.

Yager, T. 1993. Inside video for Windows. Pages 57-60 in: Byte's Essential Guide to Windows, Spring 1993.

Zamara, C., and N. Sullivan. 1991. Using ARexx on the Amiga. Abacus, Grand Rapids.

Zeitl, H. 1992. Television production handbook (5th edition). Wadsworth Inc., Belmont, California.

Alan J. Thomson graduated in 1968 with a B.Sc. (Hons.) in Zoology (entomology) from Glasgow University. He received a Ph.D. in ecology and parasitology from McMaster University, Hamilton, Ontario, in 1972. Following four years of postdoctoral research (1972-1976) into simulation of nutritional and physiological systems at the Institute of Resource Ecology, University of British Columbia, Vancouver, B.C., he joined the Canadian

Forestry Service as a Biological Systems Analyst, involved in development of pest management and forest management systems. From 1988-1989 he was on leave at the Department of Forestry and Natural Resources of Edinburgh University, developing expert systems for natural resource management, and is currently developing such systems in a forestry setting. He is presently a member of the Advanced Forest Technologies program at the Pacific Forestry Centre.

Cheryl Sivertson is currently in her graduating year at the University of Victoria where she is majoring in computer science.