Modeling Videos as Works

James M. Turner, Université de Montréal
Abby A. Goodrum, Syracuse University

Abstract

Defining works is complex, but defining video works is extremely complex because of the large number of instantiations available and because of the intricate relationships among them. Sorting these out is critical to ensuring adequate management of the information they contain, and in order to do so it is necessary to gain a clear understanding of all these components and their various roles in the overall information retrieval picture. The terrorist attacks on the World Trade Center on September 11, 2001 and other important news events offer rich examples of the complexity of the problem and help us gain an understanding of how managing video news material can be modeled. The need for a taxonomy of some of the instantiations of material that need to be managed in the context of a news video library is identified.

Introduction

Videos convey information that may be similar to that contained in texts; however, videos do not communicate in the same fashion as texts. Previous work on the nature of moving image works has revealed the multiplicity of entities that can occur in the bibliographic universe among moving image materials and has linked these to variability among exemplars of a work that are distinct from manifestations of written works. The focus of much of this work has been on exploring the limits of moving image manifestations (Yee, 1994a). Whenever a version or edition of a moving image is released that is significantly different from the original, the new version is considered to be a separate manifestation. Whether through editing, alteration or supplementation, manifestations represent new intellectual or artistic content and are treated as separate entities by catalogers. Thus, when a film is released for television, it may be edited, among other reasons, "to remove profanity, sex, violence and product identifications (if Lucky Strike was a sponsor, Humphrey Bogart couldn't be seen smoking Camels), and then footage was either added or removed to enable them to fit into standard time slots between commercials" (Yee 1994a, 246), constituting a new manifestation of the work. Another example is versions of films shown onboard aircraft. These are "edited for sex, violence (especially airplane crashes), language, and length" (Yee 1994a, 246). In addition, there are sometimes regional versions of works, edited to accommodate local laws and sensibilities (Yee 1994a, 246-247). On the other hand, releasing a work in a new physical format, such as a film released on video or a video released on DVD, does not consitute a new manifestation. Catalogers treat these as near-equivalents, retaining the original cataloging record to describe the minor changes, although guidance as to what constitutes major and minor changes varies among cataloging codes (Yee 1994b, 358-359). Yee offers specific recommendations for treatment of various types of manifestations and near-equivalents (Yee 1994b, 368-369).

Cataloging practice is rooted within a conceptual framework that is inherently bibliographic. Despite the vast number of problems that Yee exposes for consideration, the transfer of cataloging principles to moving image products is relatively straightforward in the case of completed films. However, the level of granularity of the description involves fundamental differences between considering moving image products in a bibliographic framework and considering them in a visual archives framework. Books and completed moving image products represent discrete entities to be catalogued. They are the "item level" under consideration. While it is true that users may seek finer grained access to the contents of books (e.g. through browsing, tables of contents, or indexes), catalogers address themselves to the book as intellectual object. For moving image archivists and visual resource managers, however, the line can not be drawn at the tape box or the DVD case. Moving images in this context are seen as fluid, nomadic entities capable of embodying the intellectual contribution of a wide array of creators. The use made of such documentation is

often more than mere consultation. Researchers “mine” such products for material that can be used in the production of other audiovisual works. Scientists study the occurrences of some detail across numerous works. Artists make links among shot-level events from a number of productions.

In addition, studies of image seeking behavior and use point to behaviors which are distinct from those employed in textual information seeking. (Goodrum 2001). These studies indicate that keyframes and other visual extracts of a video function as visual equivalents for the parent video in ways that textual extracts such as abstracts or titles do not. Visual resource managers and moving image archivists realize that the value of the material they manage resides not only in the moving image as a whole, but also within the shot and even the frame as these are often sought as surrogates or even replacements for the moving image. This discussion attempts to provide a framework for considering moving image works at the shot and scene levels, in addition to their manifestation as completed products.

Access at the scene, shot, and frame levels

Providing access to moving images at the shot and scene levels has been a goal for many years. The need for this type of access is so compelling that the Text Retrieval Conference (TREC) established Video Trec in 2001 with the explicit goal of providing better shot-level access. Large national institutions such as the National Film Board of Canada are considering providing scene- and shot-level access to the users of their databases (Bidd, pers. com. 2002). Providing frame-level access is useful in special contexts.

For the purposes of cataloging, indexing, storage and retrieval of moving images, a scene can be described as a group of shots related to a common theme or that take place around a specific event or theme of a movie or television show. The shots play one after another in the edit sequence and can be considered as an intellectual unit used to advance the plot of a movie or to provide a unit of information in a television program.

Scene-level access is widespread in large news and television archives databases in North America and Europe. DVD products often provide access at this level as well, or alternately, at the broader level of the main sections of a movie. In using audio description text for indexing moving images, the need for describing some of the material at the scene level rather than the shot level was observed (Turner 1998). In computer science research on the automatic indexing of moving images, scene-level access is described in various models as one of the access levels needed (Mohan 1998, Ide et al. 2000).

Shot-level access has long been seen as desirable, but the costs of providing it are so high that it is usually found only in television news archives, where it is an absolute necessity, and in stockshot libraries, where the potential revenues from the sale of rights to the material make the practice cost-effective. In information science, the research in this area focuses on automating the process by repurposing text creating during the pre-production, production, and post-production processes.

Just as the the various manifestations of a work invite much discussion and sorting out, so do the various manifestations of a shot. Context needs to be provided in order to discuss the question, and three contexts are given here: the camera, the information system, and the finished product. In the camera, a shot is what is recorded between the time the camera starts rolling and the time it stops rolling. In collections that use shot-level description, such as stockshot libraries or television archives, there may be an initial edit of the raw material, for example to discard technically imperfect or unusable parts of a shot. As a result, the shot found in the information system may be different from the one that came from the camera. To paraphrase Yee, manifestations of a shot and near-equivalents of a shot require treatment for the creation of records in information systems that catalog at this level. Five cases are given for what consitutes a shot for purposes of storage and retrieval from information systems that index shot by shot (Turner 1996). When a shot appears in a finished product such as a film or a television program, it can be defined as what occurs between two cuts. However, in light of the editing that takes place in the contexts pointed out by Yee (1994a, 246-247), we need to consider even different cuts or “manifestations” of the same shot.

Frame-level access is exceptional but in some contexts it is very useful. As animators and other filmmakers know, what happens between the frames is more important than what is on the frames (Haynes 1999, McLaughlin 2001). The Zapruder film recording the assassination of John F. Kennedy has been analyzed frame by frame, over and over again. Forensic work involving analysis of information recorded on security cameras makes use of frame-by-frame analysis on a daily basis. Insurance companies are likewise interested in such analysis for the critical information it can provide in some cases. All these players, as well as historians, researchers and other analysts will be making extensive use of the available material from the historic events of September 11, 2001 by studying it frame by frame, notably the footage of the planes hitting the towers of the World Trade Center and that of the buildings collapsing. Goodrum (2001) and others have shown the importance of individual frames in indexing shots. Computer science researchers in the area of automatic analysis of moving images also use the frame as a unit in some of the information models they describe. Models for cataloguing and indexing individual frames have yet to be described in the context of building information systems for storage and retrieval of moving images, but there is some precedent and guidance available from research in the areas of the management of still images and the use of keyframes for indexing moving images.

In addition to forensic, security, and historical study, access at these levels of granularity is desirable for entertainment, including dramas, situation comedies, variety shows and quiz shows, as well as for the documentation of experiments, lectures, and other real-world phenomena. Perhaps most in need of shot-level access, however, is the domain of television news, because of its immediacy.

Television news material as works

News library stock footage sales is a billion dollar industry worldwide and the last decade has seen a significant rise in the number of media asset management and digital content management systems for sale. According to market research firm Gistics Inc., (www.gistics.com) every year companies jeopardize $300 billion worth of their brand-related digital content such as video, packaging, print, and audio because it is either lost, misplaced or duplicated. Of primary concern is how to access not only original raw footage of an event, but all subsequent uses of that footage. A first step in gaining some videographic control over these elements would be to construct a of taxonomy of elements within broadcast news footage that represent derivative works. We need to model relationships among works of news footage having a common progenitor or common ideational content.

First we need to eliminate the erroneous distinction made between “deliberate” works of creativity and news footage. The tacit assumption has been that motion pictures and other “crafted” or edited works are deliberate acts of creativity while news footage is not. We do not generally conceive of newspapers or the *Six O’clock News* on television as deliberate works, but of course they are. They are the object of editorial meetings and decisions, and are crafted and edited, even timed to the second in the case of television news. Much more than for writing, painting, or other visual arts, it is safe to say that in moving image works, nothing is there by chance. In addition, these are serial works, presumably without pre-ordained end. And they are works which often include reports or eye-witness accounts that are themselves serial works, such as a series of investigative reports (Smiraglia, pers. com. 2001). Sometimes these are later compiled into books, documentaries or motion pictures such as the recent film *Black Hawk down,* which is an adaptation of a serialized report that appeared in the Philadelphia *Inquirer* (newspaper) in print and online (Philadelphia 2001).

Next let us review the definition of a work and examine relationships among different types of works. The unit of analysis for cataloging purposes is the “bibliographic entity.” An “entity” (bibliographic or otherwise), is generally defined as a unique instance of recorded knowledge having both physical and intellectual properties. In general terms, a “work” is defined as a collection of ideas set into a document or other signifying object used for communication. The document is simply a container and embodies the physical aspects of a bibliographic entity. The intellectual content of the entity is the collection of ideas embodied in the work (Smiraglia 2001). From the perspective of a library of news footage, the bibliographic entity could conceivably be the original raw footage taken of a specific event. The physical property of this entity is the tape itself. The intellectual component is the content of the tape. In journalistic

parlance, it is the who, what, where, when, why, and how of the tape. Further, this content is embodied not only in the native events recorded by the camera, but also in how the camera was used to record these events. The same event recorded by different persons will yield distinctly different images. Differences in perspective, camera movement (panning, zooming, tilting) color, light and shadow manifest themselves differently and constitute related but distinct works of creativity and communication. This is the equivalent of offering differing eye-witness accounts of the same event.

From a progenitor work may spring subsequent works. These derivations contain some of the ideational content of the progenitor work but present a different expression of that content. With respect to textual works, derivative works are generally translations, extractions, amplifications, or successive editions of a progenitor work. In terms of video news footage, derivative works occur whenever any part of the original footage (video or audio) is rebroadcast, edited, repackaged or extracted in some way. Taken together these instantiations of video works form the families that we seek to cluster for information storage and retrieval and for media asset management.

The goal is to organize these instantiations such that we are able to distinguish between individual progenitors, and at the same time to reveal relationships among the various derivations from each work. The record for an individual work should point to the known derivative versions of that work, arranged so that manifestations that are most similar (e.g., subsequent broadcasts of the entire work) appear close together while manifestations that are most different (e.g., audio extractions) appear farther down the list from the progenitor. The records for each derivative instantiation must also be capable of linking back to the original work.

Capturing these relationships within the current cataloging rules is challenging. Cataloging rules for video and for serialized works such as news reports pose a number of important problems in the practice of video news librarians, moving image archivists and media asset managers. AACR2 defines two "types of publication": the "monograph" and the "serial." A monograph is defined as being complete or intended to be complete in a finite number of parts. A serial is defined as being issued in a succession of numbered parts that are intended to continue indefinitely. These distinctions between monograph and serial are problematic for video news cataloguers because different rules apply to each. Moreover, AACR2 does not accommodate serials in any of the special category chapters, such as maps, sound and video recordings, and computer files.

To complicate matters, most local and regional television news archivists have not yet agreed upon a "standard" method of cataloguing television news. Because of the need to access derivative materials at varying levels of granularity, it may be impossible to give each instantiation full AACR2/MARC cataloging treatment. As a result, some news archives are only cataloging material at the collection level. Others have opted out of cataloging entirely in favor of providing only inventory and indexing information for each item.

Elements for a taxonomy

At this stage of the sytematization of methods for organizing moving image collections at the shot level, perhaps the most useful contribution to the problems faced by video news librarians, moving image archivists and media asset managers would be a taxonomy of the types of materials they might find in the collections they manage. Such a taxonomy would no doubt help improve description and cataloging methods available, but elaborating such a tool is not a simple matter. This has largely to do with trying to sort out the various instantiations, versions, manifestations and near-equivalents of a work.

To get an idea of the extent of the problem, following are a number of situations in which a simple piece of material might be found or used. We take as a main example the planes crashing into the twin towers of the World Trade Center, images familiar to everyone, and make references as well as to other familiar news events. The same material could be found all the following situations in moving image databases.

Raw footage. An obvious division in news archives is professional and amateur material. Professional material is created by a news crew on location for the event, or a freelance crew hired for the occasion and working in an institutional setting. In cases such as the arrival of some visiting dignitary, the shoot is planned and discussed, and decisions are made about what images to capture. In cases where there is urgency because of the nature of the event, such as the attacks on the World Trade Center, there may be some direction or none at all, with crews relying on their experience and know-how to capture what they deem to be important or whatever they are able to get when shooting circumstances are difficult or impossible. Where there are more than one camera, such material involves multiple perspectives, multiple views of the same event.

Amateur news footage usually involves bystanders who happen to capture the event on video because it takes place very near to something else they were shooting. They hear a noise, and focus the camera on the news event. Alternately, they capture an event because they happen to be already shooting something that becomes the object of news, such as a stunt at an air show that turns into a plane crash. In our example, a tourist already has a camera on the towers of the World Trade Center because of the smoke emanating from the first crash when the second plane smashes into the second tower. Other examples of these situations include the Zapruder film of John F Kennedy's assassination and the amateur video footage of the Concorde accident at Charles de Gaulle airport.

Sometimes there are "hybrid" situations, in which amateur footage is added to professional footage because it is taken from an interesting point of view or reveals details not captured in the official material. Another such situation is exemplified by the memorial documentary film about the events of the World Trade Center entitled *9/11* that was broadcast on CBS March 10, 2002, six months almost to the day after the event. French filmmakers Jules et Gédéon Naudet were already working on a documentary about the New York Fire Department. On the morning of September 11, 2001, Jules Naudet was with a crew of firefighters responding to a report of a gas leak at the corner of Lispenard and Church streets. As they were packing up to return to the station, Naudet heard the roar of a low-flying jet and turned his camera up in time to record the American Airlines jet flying into the first tower. Moments later, the firefighting crew had rushed to the tower and Jules Naudet captured footage inside the lobby. Gédéon Naudet got there in time to capture footage of the action in the street. In this case, a professional crew found itself in an unexpected situation and immediately turned its attention elsewhere. Editorial decisions were made on the spot about refusing to film sensational footage of people on fire, people jumping from the buildings, and so on.

Satellite feeds. Crews on the scene shoot material of newsworthy events and upload it to satellites. News agencies are among the brokers of this material, which subscribers can edit into news and other programs. Users typically also have feeds from other networks as well as agreements with them for use of their material in exchange for a fee. Sometimes re-feeds which include natural sound not included in a first upload are sent, creating another version of the same material.

Edited packages. This category represents several layers and several generations of material. From all the material available, compilations of the most interesting material are made available to broadcasters. Domestic and international editions reflect narrower or broader interest. In using this material, both domestic and international users affix their logos as a kind of watermark as the material is edited into news items which include their own commentary. Thus a large number of variations are created, which need to be accounted for when they are included in information systems.

Audio. In situations where the image is not available, the audio is sometimes accompanied by still images. An example of this situation occurs in the case of the historic concert given by Marian Anderson on the steps of the Lincoln Memorial in Washington on Easter Sunday, April 9, 1939. In ongoing work at the UCLA Film and Television Archive on the *Treasures of American Film Archives* film entitled *Marian Anderson's Lincoln Memorial Concert,* images and audio are being added to complete missing moving image information. The film crew from Hearst Metrotone News stopped shooting the concert once they had their news story. However, the NBC radio broadcast of the concert was preserved, and the UCLA Film and Television Archive are adding still images to the radio broadcast sound in order to provide a more complete document about the event.

Keyframes. These are usually selected automatically as a clue to the content of the moving image. In systems that use automatic detection of scene changes, the first image of a new scene is usually the one generated automatically as a keyframe. Some systems allow users to select another, more representative frame if this is deemed to be preferable. A screen full of still images can then be browsed and the user can click on a key frame to see the moving image. As computers are able to process more and more information, we can hope to see all those key frames replaced by moving clips.

In the examples we have already given, the assassination of John F. Kennedy, the terrorist attacks on the World Trade Center, as well as other important events such as the volcanic eruption of Mount Saint Helens or the explosion of the space shuttle Challenger, it is easy to see the usefulness of access to individual frames. Newspapers and magazines may wish to publish all the frames as a sequence of still images. Other publication venues include the Web, news digests, books, journal articles, presentations, and still image databases, as well as the classroom. Depending on the context of the material and how it is presented, each situation can involve new works or instantiations of existing works.

Icons, graphics, appropriation. Some particularly striking images take on an iconic value, and because of the widespread use of them that follows, it becomes next to impossible to keep track of all the uses and versions created. Examples of this include the still photo of the Vietnamese peasant girl whose body is covered with napalm, running naked and with a look of anguish on her face that seems to capture all the horror of war, or the moving (in more ways than one) image of Chinese student defiantly facing the army tanks in Tien An Minh square.

Sometimes such an iconic image becomes part of a piece of still or video art, becomes an object of study in other contexts, or is reduced to a graphic version and becomes the logo of some organization.

Appropriation is a venerable practice in the art world. All kinds of borrowing of images or ideas expressed in them is responsible for the creation of new images, new versions, new instantiations. The drawn depiction of the firefighter carrying the little girl from the Oklahoma City bombing became an icon in its own right. A clip of Fred Astaire dancing in a movie appears in a tv commercial for some product, with current video editing technology allowing new, composite images to be created.

Repackaging and reuse. Repurposing existing products is big business because it provides new sources of revenue. There are a number of manifestations of this kind of activity. “Year-enders”, compilations of the important news events of the year for broadcast in the last week of December and the first week in January, are a ready example. Historical documentaries borrow footage from many sources, sometimes using filmed events to represent other historical events (“The Office Cat”, an enlightening column which appears in the *AMIA Newsletter,* documents falsified or inappropriate uses). Home video and accompanying texts are a common source of material appearing in documentaries. Educational video and accompanying texts likewise make use of material created in other contexts. DVDs illustrate the recycling of peripheral shooting materials and other material as added value to accompany the film being purchased. Multimedia products on CD-ROM and other supports make broad use of repurposed images. Compilations of material previously distributed alone or in other compilations are commonplace. Finally, extracts from existing products sometimes get used in other products, sometimes taking on an iconic aura because of widespread circulation.

Our enumeration of some of the elements that would need to be considered in building a taxonomy is no doubt incomplete. However, we hope it gives some idea of the extent of the problem faced by those responsible for managing all these components of their collections, as well as of the theoretical complexities involved in constructing a taxonomy for these elements.

Conclusions

In a study conducted by Leazer & Smiraglia (1999), the popularity and age of a progenitor work was shown to be an important factor in the generation of bibliographic families. Anecdotal evidence suggests that some pieces of raw news footage are selected for re-use more frequently than others. Similarly, some

videographers find their work disproportionately distributed in the edited broadcast offerings. It is not unusual for editors to select from among several sources of footage depicting the same event by asking for the work of a particular videographer whose work is especially appreciated. Storage space in any video repository is at a premium and as new videos enter the archive, older, less used items must be weeded. In this environment, footage that is of continuous use is kept in primary storage while less-used tapes are relegated to offsite storage. Over time, certain highly used works of news footage take on an almost canonical caste. How like citation distributions this is. In light of this, perhaps the exploration of derivative bibliographic relationships suggests an additional tool that can be added to the arsenal of bibliometric measurement.

References

AMIA Newsletter: the newsletter of the Association of Moving Image Archivists. Beverly Hills, CA: Association of Moving Image Archivists.

Bidd, Donald. 2002. Interview at the Université de Montréal, 2002 01 31.

Goodrum, A. 2001. Multidimensional scaling of video surrogates. *Journal of the American Society for Information Science* 52, no. 2, 174-183.

Haynes, M.L. 1999. Animation aerobics: an interview with Ed Counts. *Frame by Frame: a quarterly publication of ASIFA/Central* (spring), 3.

Ide, I., R. Hamada, S. Sakai, and H. Tanaka. 2000. Scene analysis in news video by character region segmentation. In *Proceedings of the ACM Multimedia 2000 Workshop, Marina del Rey CA, USA* (November), 195-200.

Leazer, G. & Smiraglia, R. 1999. Bibliographic families in the library catalog: a qualitative analysis and grounded theory. *Library Resources and Technical Services (LRTS)* 43, 191-212.

McLaughlin, Dan. 2001. Film/TV 181A Animation design in film and television Available at http://animation.filmtv.ucla.edu/program/curric/181A.html (page consulted 2002 03 18).

R. Mohan. 1998. Video sequence matching. In *Proceedings of International Conference on Speech, Accoustics and Signal Processing* 6, 3697--3700.

Phildelphia Inquirer: Black Hawk Down (website containing the entire story that appeared in serialized form in the print publication. Includes additional audio, video and textual materials). Available at http://inquirer.philly.com/packages/somalia/nov16/default16.asp.

Smiraglia, Richard K. 2001. E-mail correspondance with Abby A. Goodrum. 2001 11 27.

Smiraglia, R. 2001. *The nature of "a work": implications for the organization of knowledge.* Lanham, Maryland: Scarecrow Press.

TREC Video Retrieval Track. Available at http://www-nlpir.nist.gov/projects/trecvid/ (document consulted 2002 03 18).

Turner, James M. 1996. Issues in shot-level indexing of moving images: what constitutes a shot? *The SIG VIS News* 1, no. 2 (spring). Available at http://www.asis.org/SIG/SIGVIS/j_turner.html (document consulted 2002 03 09).

Turner, James M. 1998. Some characteristics of audio description and the corresponding moving image. *Information access in the global information economy: proceedings of the 61st ASIS Annual Meeting, Pittsburgh, Pennsylvania, October 24–29 1998,* vol. 35. Medford, NJ: Information Today, 108–117.

Yee, Martha M. 1994a. Manifestations and near-equivalents: theory, with special attention to moving-image materials. *Library Resources and Technical Services (LRTS)* 38, no. 3, 227-255.

Yee, Martha M. 1994b. Manifestations and near-equivalents of moving image works: a research project. *Library Resources and Technical Services (LRTS)* 38, no. 4, 355-372.