Home Video Editing Made Easy —
Balancing Automation and User Control

Andreas Girgensohn[1], Sara Bly[2], Frank Shipman[3], John Boreczky[1] & Lynn Wilcox[1]

[1]FX Palo Alto Laboratory, Inc., 3400 Hillview Avenue, Bldg. 4, Palo Alto, CA 94304, USA
[2]Sara Bly Consulting, 24511 NW Moreland Road, North Plains, OR 97133, USA
[3]Texas A&M University, Dept. of Computer Science, College Station, TX 77843, USA

{andreasg, johnb, wilcox}@pal.xerox.com, sara_bly@acm.org, shipman@cs.tamu.edu

Abstract: Hitchcock is a system to simplify the process of editing video. Its key features are the use of automatic analysis to find the best quality video clips, an algorithm to cluster those clips into meaningful piles, and an intuitive user interface for combining the desired clips into a final video. We conducted a user study to determine how the automatic clip creation and pile navigation support users in the editing process. The study showed that users liked the ease-of-use afforded by automation, but occasionally had problems navigating and overriding the automated editing decisions. These findings demonstrate the need for a proper balance between automation and user control. Thus, we built a new version of Hitchcock that retains the automatic editing features, but provides additional controls for navigation and for allowing users to modify the system decisions.

Keywords: Video editing, video analysis, video exploration, automatic video clip extraction.

1 Introduction

Video use is becoming increasingly popular among non-professional users. Many people own video cameras and use them to record personal events such as vacations or weddings. Furthermore, Digital Video (DV), a digital camera format, is becoming more common. With DV cameras, people can easily move their video to the computer. Once on the computer, the video can be edited or converted to a streaming format and published on the Web.

However, editing the video is still problematic. With still cameras, it is easy to discard pictures that are of poor quality, that are redundant, or that are simply "boring." With video, the process is more difficult. This is because (1) you have to find the parts of the video that you want and (2) you have to trim the video in such a way that it is visually appealing when viewed. Finding the part of the video you want requires linear search, and finding the cut points for trimming requires looking at individual frames. Although there are systems that automatically select summary segments from a video (Christel et al., 1998; He et al., 1999; Pfeiffer et al. 1996; Smith et al., 1997), these systems are designed to create a video abstract rather than to aid in automatic editing.

There are several commercial systems for non-linear editing of DV such as Adobe Premiere (2000)

and Apple's Final Cut Pro (2000). However, these systems are difficult to learn and to use. They focus on enhancing the look of the output video by providing titling, transitions, and special effects. Simpler editors such as Apple's iMovie (2000) and Ulead's VideoStudio (2000) are easier to use, but still require the user to find a desired portion of video and to determine the in/out points manually. This requires examining the video in real time, on a frame-by-frame basis. Furthermore, these systems treat takes (see Table 1) as the smallest unit even if they are several minutes long and contain many different events.

We began with the hypothesis that video editing could be made simpler by pre-processing and structuring the video for editing. First, we automatically segment long takes into shorter clips. These clips contain a single event, and are trimmed to cut out "unsuitable" video. By unsuitable video, we mean video where the camera is moving too quickly, the camera is unsteady, or the image is too dark. These regions are found by analyzing the video for camera motion and luminance level. Next, we cluster the clips and arrange the keyframes representing each clip into "piles" based on similarity and/or temporal structure. Finally, we provide an intuitive user interface that allows the user to navigate through the piles, arrange selected clips in the desired order in a

Frame	Still image from video
Keyframe	Frame representing video sequence
Take	Video sequence between camera on/off
Clip	Uninterrupted video sequence (subsequence of take)
Pile	Stack of keyframes representing clips

Table 1: Glossary of terms.

timeline, and lengthen and shorten a clip by simply manipulating the size of its keyframe.

In order to test our hypotheses about automating video clip selection, trimming, and navigation, we built Hitchcock, a semi-automatic home video editor. Hitchcock was designed to incorporate the automatic pre-processing of the video with an intuitive user interface to present the video clips to the user. An earlier version of the system and the video analysis algorithms are described in (Girgensohn et al., 2000).

In this paper, we report on a user study conducted with the earlier version of Hitchcock. We then describe the enhancements made to Hitchcock following the study findings. We conclude by discussing the need for a proper balance between automatic determination of video clips and user control.

2 Hitchcock

Hitchcock is a system for editing home video by incorporating automatic analysis of the video to support good editing techniques. The characteristics distinguishing Hitchcock from other video editing tools are its segmentation of takes into clips, the automatic selection of the most suitable portions of the resulting clips, and the clustering and visualization of the clips for the user. By dividing the takes into shorter clips and by selecting the portions of the clips that have the least camera motion as being the most suitable, Hitchcock provides structure for the raw video footage based on expert knowledge of the editing task. The visualization of the resulting clips into clusters offers an organization for the video while still providing users with control over their results. This section provides a brief description of the main features of the version of Hitchcock used in the user study. Enhancements made in response to the study results are described in Section 4.

Hitchcock uses the camera on/off information included in the DV format to determine video takes. These takes are divided into clips based on heuristic notions of video quality. Thus clips are separated by erratic or fast camera motion. The user has access to between 3 and 10 seconds from each of the resulting clips. Whenever the user changes the length of a clip, the best contiguous footage is selected from the original clip. Thus, a 3-second clip and a 7-second clip from the same initial clip may, in fact, consist of non-overlapping segments within that original clip.

Clips are hierarchically clustered using color histogram analysis and the resulting clusters are visualized as piles of *keyframes* (i.e., a single image for each clip). The FotoFile system (Kuchinsky et al., 1999) uses a similar clustering algorithm to present visually similar photographs in a Hyperbolic Tree but that approach does not appear to scale up to a large number of images. In Hitchcock, the piles are shown row-by-row in the time order of the first clip in each pile. We present clips in piles rather than showing them as individual images in a scrollable window because we believe that piles provide an organizational structure that is easier to navigate than to scroll through dozens of screens of images. We offer two different clustering approaches. In the contiguous view, only clips that are adjacent in time may be clustered together so that each pile contains an uninterrupted sequence of clips. In the other (non-contiguous) view, clips from any part of the video may be clustered together based on color similarity. The clips in each pile are still in temporal order and the piles are sorted by the time of the first clip.

The interface consists of two image-based panels (see Figure 1). The selection panel at the top lets users select clips from the raw video. The composition panel lets the users organize the clips to compose an output video sequence and change the lengths of the clips. Users may drag individual clips from the selection panel into the composition panel to include that clip in the resulting presentation.

The timeline at the top of the interface shows the coverage of the clips in the current pile under the mouse (dark area) as well as the start time of the particular clip (triangle) within that pile. Furthermore, the image under the mouse is moved to the top of the pile to reveal its content. By moving the mouse across a pile, users can flips through a sequence of images representing clips within that pile. To navigate through the hierarchy of clips, users click on a pile to open an overlay window displaying the contents at the next level of the hierarchy.

Clips dragged into the composition panel are framed by a purple border in the selection window (similar to a visited link in a Web browser). The user can change the image's size to change the length of the clip. The clips can be reordered by dragging them to different points in the composition panel. We use two designs for placing images. The first design saves horizontal space by staggering the images in two rows so that less scrolling is required to see all images. The images are slightly overlapped to make the time order more obvious. The second design places all images in a single row.

When comparing Hitchcock to commercial systems, Apple's iMovie (2000) has many similarities. It is aimed at the same target group of home video users and has a user interface similar to Hitchcock's. In both interfaces, clips are represented by keyframes
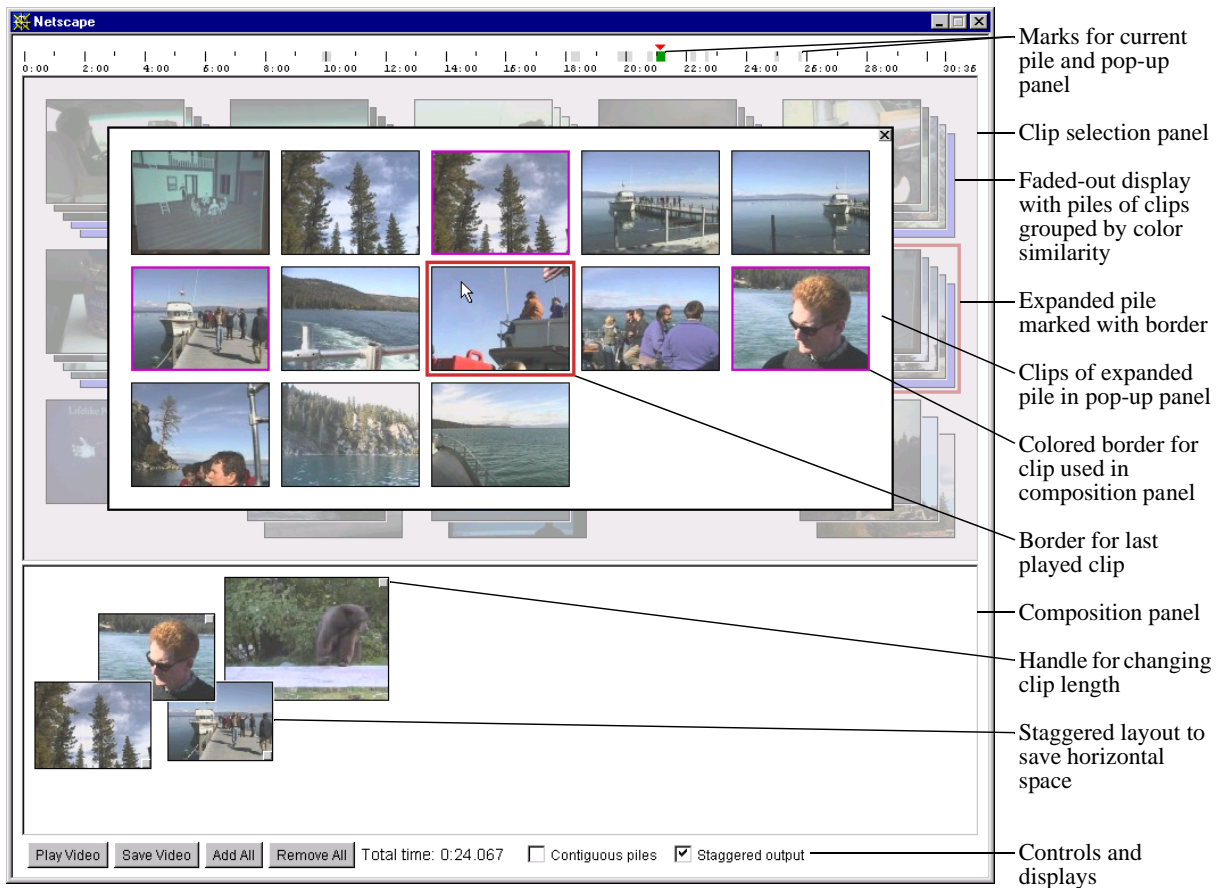
Figure 1: Video editing interface with selection and composition panels.

that can be placed in either a 2D area or in a scrollable output timeline. However, the initial set of clips in iMovie consists of the takes from the DV. iMovie's shelf can only contain individual clips and offers no means to organize clips into piles. iMovie also does not provide an automatic breakup of takes into smaller clips or a selection of a good portion of a clip. Clips can be split or cropped manually, but only by using the player and examining the frames.

3 User Study

Because Hitchcock is based on a substantially different user model than other home video editing systems, we felt it was important to get input at an early stage of the prototype. We were not as interested in details of the user interface as much as in getting feedback on the overall approach of using clips and piles. To get realistic use of the prototype, we felt it was necessary that home video users try it with footage they had shot themselves. Our user study focussed on the segmentation of takes into clips, the clustering of clips into piles, and the selection of suitable footage by trimming.

3.1 Method

We observed nine participants individually using Hitchcock to edit their own video footage. Each participant was given a DV camera and asked to shoot at least 30 minutes of video with the suggestion that they take no more than 5 minutes of video each time they turned on the camera. We began and ended the 90-minute sessions with questions and discussion. Participant P7 only had time for ten minutes of editing and is not included in the data that follows.

The remaining participants consisted of four women and four men. Four participants record video at least once a month and three more at least a few times a year. No one records it less than once or twice a year. All participants want to edit their own video and all but two want to be able to edit video someone else has taken. Only one participant has been able to edit any of his own home video.

The evaluation consisted of four phases: a training session, a written questionnaire, video editing, and a summary discussion.

- In the training session, the facilitator explained the pre-processing features of Hitchcock, and a videotape gave an overview of the Hitchcock user interface. Participants played with the demo

	Original footage	Takes	Clips	Output clips	Output	Edit time
P1	60:08	38	354	38	4:21	40 min.
P2	21:32	47	129	16	1:59	36 min.
P3	55:52	57	231	17	1:16	45 min.
P4	50:06	50	296	67	8:21	40 min.
P5	54:50	117	317	65	5:17	45 min.
P6	45:25	166	265	75	7:47	45 min.
P8	30:31	158	189	25	2:39	45 min.
P9	89:57	72	477	172	21:28	45 min.

Table 2: Video footage for each participant.

	#takes used in output	#takes w/ some clips in output	#takes w/ all clips in output
P1	17	16	1
P2	11	9	2
P3	8	6	2
P4	24	19	5
P5	48	12	36
P6	56	19	37
P8	21	5	16
P9	29	16	13

Table 3: Use of takes in output.

video footage and were asked to perform several tasks to indicate their understanding and to elicit further questions of Hitchcock features.

- The questionnaire had 18 questions about their individual video editing habits and their anticipated wishes for editing, about their initial impressions of Hitchcock, and about the video footage they shot for the study session.
- Next, participants had 45 minutes in which to edit their own video footage using Hitchcock. The participants were left alone to edit, having been asked to talk-aloud as they worked.
- After the editing session, the participants were asked a series of open-ended questions to elicit their perceptions and feedback on Hitchcock.

3.2 Data

Data consisted of videotapes, questionnaires, and logs of user interface events for each session. Participants varied widely in the amount of video they shot and in the final output they created. The eight participants took between 21 and 90 minutes of raw digital footage. Five participants said deleting unwanted footage was one of their goals in editing and six wanted to put together random sequences of interesting bits. Only three participants mentioned wanting a chronological sequence as a goal in editing.

The takes from each participant's footage were divided into clips. Table 2 shows the original length of the raw footage, the number of takes, the total number of clips, the number of clips used in the output, and the final output length.

All participants worked consistently throughout their editing session and were able to create output films. Participants said that they liked using Hitchcock, particularly the ability to play and to drag and drop clips easily. P3 said she particularly liked the process, noting: *"I like snapshots over the whole time. Working with video is so difficult and I really like being able to look at them in this piecemeal fashion and then being able to go in and explore."* P5 summed it up by saying that it's a quick and easy way to get interesting parts out of the raw footage and to arrange them for viewing. Seven of the partic-

ipants said that they would like to use Hitchcock at home. The eighth remarked that it would be conditional on having some of the changes he suggested.

Many participants did have trouble finding a few particular clips in the piles, often never locating the footage at all. Occasionally a participant would also "lose" a clip in moving between views or in accidentally closing a pile expansion. Summary discussions in conjunction with comments made during the editing sessions suggest that people were particularly concerned with identifying and/or finding clips, the lack of control in clip length (and the resulting choppy nature of contiguous clips), and the confusion about color-based similarity.

Clips. Participants varied widely in their original footage (i.e., in number of takes, number of clips/ take, etc.) but all participants had occasions when they used fewer clips than were available in a given take. Table 3 shows the number of takes in which only some clips from that take were included in the output. Four of the participants only occasionally used all of the clips in a take. P5, P6, and P8 had many takes in the output that did not segment into more than one clip (30, 31, and 16, respectively).

Even when all the clips were used from a take, the clips were not necessarily used sequentially. In particular, none of P6's clips in a take with all clips were in sequential order. There were three instances of repeating a clip in the output (P4, P6, P9).

Piles. The participants indicated no problems with the overall notion of piles and appeared comfortable with having all their footage easily represented on a single screen. However, participants were confused about the visual similarity as a grouping mechanism (i.e., the colors did not appear similar to the participants). Nevertheless, three of the eight participants used the non-contiguous piles for almost half of their editing time. P8 commented that different views offer different perspectives on the footage and thus help in identifying desired footage. In manipulating piles, participants were often surprised to find piles expanding many levels deep. A few participants also were disoriented when working at deep nesting levels.

3.3 Observations

Participants were quickly able to manipulate their clips and to create output sequences. The ease with which participants were able to work successfully with Hitchcock and their enthusiasm about its process suggest the value of the clip and pile model. The drag and drop interface for output clip arrangement also worked well. However, participants had problems with the clip identification and trimming capabilities. The study indicates that many of the problems centered around features of clip and pile use that were also considered helpful in the editing process.

Segmenting takes into clips. Clip selection problems were primarily caused by the fact that not all of the original footage was included in the Hitchcock piles. If a participant wanted footage that had erratic or fast camera motion, it was most likely not available. Participants understood that the clips represented "good" footage but wanted all of their original footage in clips. Making all footage available will most likely solve many of the problems encountered.

Clustering clips into piles. Although participants expressed some concerns with the visual similarity of piles, the pile mechanism itself was well received. Participants were easily able to flip through images, often getting a sense of the clips without having to watch the video itself. It also appears that users liked having multiple views onto their footage beyond just that of being contiguous in time. Redesign plans should continue to build on the pile mechanism but explore other ways of grouping clips into piles.

Clip trimming features. Clip trimming always kept the most suitable footage within a clip and never allowed a clip to go beyond ten seconds in length. Thus, participants were unable to stretch a clip to include all possible footage. Even if a participant only wanted to capture a particular portion of the clip, they were unable to determine whether the trimming would occur at the front or back of the clip and how to cut off the clip at a particular location. These features were often problematic for users who wanted to override the system decisions.

Even when all video footage is included in the clips, users still need control over the point from which a clip is lengthened or shortened. Several suggestions were made by participants for ways to trim clips: use the timeline to lengthen or shorten the desired end(s); use two handles on the clip image; use the video playback window and corresponding timeline to indicate a total possible clip length as well as the currently selected clip endpoints.

4 Hitchcock Enhancements

Overall, the study results offered positive feedback on the approach of using clips and piles. For the next
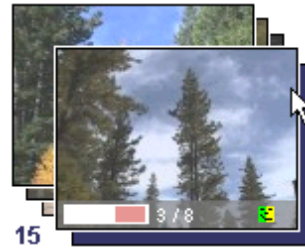

Figure 2: Pile with information bar.

iteration in the design, we planned to address the shortcomings with the selection and manipulation of video clips without giving up the value of these features. Experiences with the use of visual workspaces led to several solutions: (1) providing the user with more information about the clips, (2) improving navigation among clips and their visualization in piles, and (3) giving users more control over the clips. The sections below describe these extensions.

4.1 Providing More Clip Information

Much of the Hitchcock user's effort is devoted to the selection of clips. One difficulty that users have when selecting clips to view is not having any information beyond the keyframe and its position in the overall video (identified in the timeline at the top of the selection panel). Given that Hitchcock has more information about the clips, our first extension aimed at providing users with this additional information while not increasing the space required to present a clip in the visualization. We added an information bar to display basic information about clips and tooltips to display more detailed information.

The information bar for clips, shown in Figure 2, is a translucent overlay for the keyframe. On the left side of the information bar is a gauge showing the position and length of the default segment. A suitability icon using a combination of icon color and shape appears near the right end of the bar. Faces ranging from smiling to frowning expressions with background colors from green to yellow to red indicate five levels of highly to poorly suitable default segments (see Figure 3).

The information bar presents basic information about the clip. The addition of tooltips to the Hitchcock interface facilitates the communication of more specific information to the user. A tooltip with general information about the clip, including the take and clip numbers, appears when the mouse lingers over the keyframe. When the pointer pauses over the gauge in the information bar, an explanation of the information in the gauge is displayed. If the pointer pauses over the suitability icon, a tooltip provides

Figure 3: Faces to indicate suitability.

Figure 4: Tooltip providing detailed information.

more detailed information about the suitability of the clip (see Figure 4).

4.2 Navigating Clips and Piles

Besides providing more information about the clips, the study indicated the need for an improved piling strategy and navigation aids to find clips within the piles. The current visualization creates piles based on a color histogram analysis. There are two visualizations, one where the elements in a pile are required to be temporally contiguous and the other where they are not. Although subjects in the study indicated confusion about the analysis, several used both views, switching back and forth between the views regularly. This implies users are willing to switch between visualizations and might benefit from a set of user-selectable, and perhaps even user-modifiable, clusterings. For example, similar to the Intel home video abstracting system (Lienhart, 1999) we could cluster clips by recording time so that clips recorded on different days would be more distant than those recorded during the same hour.

Representation of hierarchy. The initial prototype for representing clips in a pile used one or two blue frames to indicate the depth of the hierarchy in that pile. Users were unsure of the exact relationship between the number of clips in a pile and the number of blue frames. To provide more explicit information regarding the size of the piles, Hitchcock now displays a single blue frame with a number indicating the height of the pile for piles containing more than five clips (see Figure 2).

Bookmarks. In the study of the initial Hitchcock prototype, users were occasionally frustrated by not being able to find a clip they had seen before. To improve the location of clips, we added a bookmark panel and enabled navigation from clips in the composition and bookmark panels. Figure 5 shows the new overall interface. When the "Find Clip" operation for a clip is chosen in the composition or bookmark panel, the selection panel will display the level of the hierarchy where that clip appears by itself (not as part of a pile). In this way, users can navigate to the context in which they remember other potentially useful clips. Figure 6 shows the results of finding the bookmarked clip with the bear.

Stable clip focus. In the study, users attempted to find clips recorded at the same time as the selected clip by switching from the non-contiguous to the contiguous pile view. To simplify that operation, a user can now click on the timeline to display the level of hierarchy corresponding to the clip at that time. In switching between different views, Hitchcock always initialized the visualization to the top level of the hierarchy. A number of times, users had found clips in the non-contiguous view and wanted to locate the clips contiguous in time. Now when switching views, the selected clip will be navigated to and selected in the new visualization.

4.3 Controlling Output

The most frustrating experience for many of the participants in the study of the initial Hitchcock interface was their lack of control over what video they could add to the presentation they were authoring. The system, with its heuristics for identifying and choosing suitable video clips, removed the need and ability for users to specify the in/out points for video clips. Users found the choices made by the system reasonable for much of their work but expressed frustration when they wanted to override the automatic choices and make their own changes.

The most emphatic lesson learned by the study was that no video should be removed. The maximum length of 10 seconds for clips in the output meant that a majority of the video taken was not visible through the Hitchcock interface. Clips now include all video and users may specify any length ranging from 0.7 seconds[1] to the length of the clip. Thus it is now possible to view and use all of the available video. However, as indicated earlier, the system still offers the user information about the best footage within a clip.

Due to the large range for output clip lengths, the keyframe size to clip length function had to be altered. Before, size was linearly related to clip length but now the size is linearly related up to a point (around 7 seconds in clip length) and then the size is relative to the logarithm of the clip length. This allows for the manipulation of very long clips. However, the new approach requires more vertical space because we chose to keep 7-second clips at the normal size (120x90). A comparison of Figures 1 and 5 shows how much more space is required.

Figure 7 shows the new interface for resizing clips in the authored presentation. A new timeline at the bottom of the output panel shows detailed information about the length of the clip and the portion of the clip currently being used. As the user resizes the keyframe beyond the 3- and 10-second boundaries,

---

[1] We choose 0.7 seconds because we consider the corresponding image width of 40 pixels to be the minimum desirable size.

Bookmark panel

Better layout for expanding piles

Overlay display under the mouse for clip length, coverage, and suitability

Logarithmic image size for long clips

Overlay display for all clips

Border to indicate timeline clip
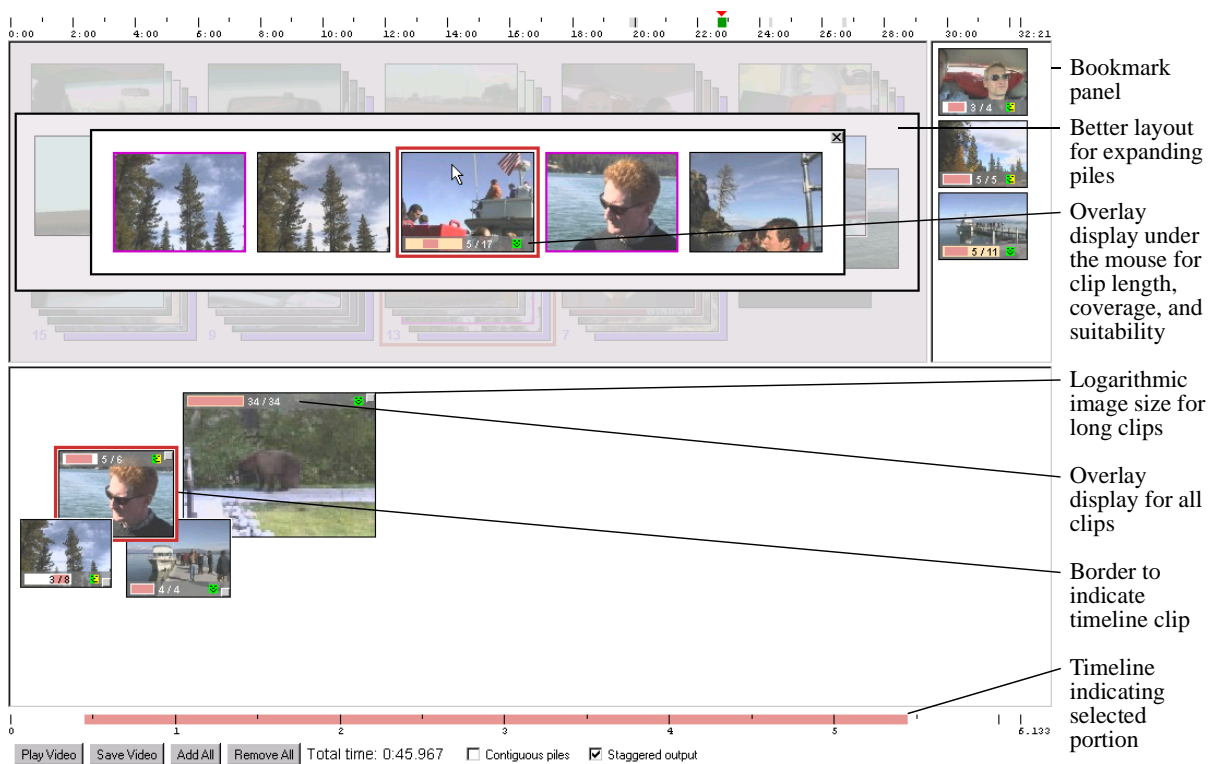
Timeline indicating selected portion

Figure 5: New Interface with bookmark panel and composition clip timeline.

the bounding box showing the new size turns from black to red indicating the user is going beyond the suggested clip length. Another change is to the algorithm for selecting segments within a clip. We now select a "core segment" for the clip and all expansions and contractions extend or reduce that core so a longer clip is always a superset of a shorter one.

There are several additional features that were indicated as being necessary, or at least useful, by the study. First, users need to have more control over clip in/out points when they want to override Hitchcock's selections. Second, audio is an important factor in the automatic selection of in/out points to avoid clips starting and ending in mid-sentence. Finally, users should be able to play only portions of the authored presentation or collapse parts of the output presentation they considered done. We are currently working on addressing these needs.

Figure 6: Results of using "Find Clip" for a bookmark.

5 Discussion

Our initial semi-automatic approach provided both strong positive and strong negative feedback. Users clearly benefited from the clip level of control and from the pile clusters. They quickly and easily flipped through clips and selected clips for their output. However, the system did not always provide enough information about its actions. Also, when users wanted a particular segment of video, they were frustrated if they were not able to locate the clip or change its length to include the desired footage.

Segmenting takes into clips. The data indicates that clips are an extremely useful tool for the users. Users took advantage of the clips to reduce quickly the amount of video shown from any single take and to rearrange clips, often within a take, to create their desired output. The changes to Hitchcock keep the clip segmentation but allow the user to extend any
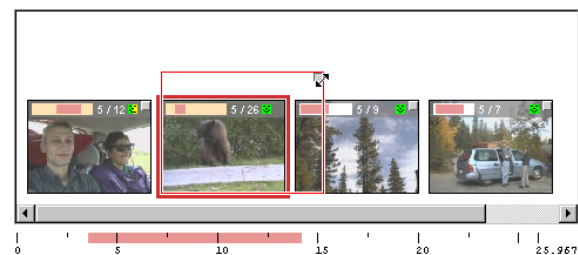

Figure 7: Resizing the keyframe to change the clip length.

particular clip to include as much of the original footage as they wish.

Information about clips. Users had no problems using keyframes as representatives for clips. However, they did not have any additional information about the clips so that they frequently played the clips. We now provide additional information about clips without complicating the user interface through a combination of image overlays and tooltips.

Clustering clips into piles. No specific changes were made to the clustering algorithms used in the initial Hitchcock prototype. Although users found the groups non-intuitive, several of them nevertheless found the non-chronological perspective helpful. We plan to offer more clustering options including the grouping of clips of the same take. While we initially expected that an abstract representation for the number of clips in a pile would be helpful, it caused confusion so that we now show the exact number of clips for large piles.

Locating clips. While users could quickly locate clips that were visible in the top-level piles by flipping through the clips in those piles, they had problems locating the remaining clips. In response to this problem, we added several navigation aids such as a bookmark panel and access to clips through the timeline that should make locating clips easier.

Clip trimming features. Although users often wanted to include additional footage, they liked the ease with which they could lengthen or shorten a clip. The new features of Hitchcock give them the advantages of the automatic processing to define suitable footage while offering a simple mechanism for overriding the system decision.

6 Conclusions

In this paper, we have described our initial design of the Hitchcock system and user reactions in using Hitchcock to edit their own digital video footage. Overall, the study showed that our approach is very promising. The participants liked the system and had no major problems using it. However, the study results also pointed to problems and we have made several changes to the system to address some of those shortcomings. The changes to the Hitchcock interface combine automatic analysis techniques with better information and user control. This seems to be the right approach to make home video editing easier, faster, and more enjoyable.

A lesson from this study for system designers in general is that balance is required when introducing automation into an application. Automating tasks will frequently achieve the goal of reducing the effort required of the user, making acceptable tasks that were previously too much work to perform. Equally importantly, automating tasks can frequently process data in a way that goes beyond the user's ability or expertise. But however intelligent or helpful the automation, it also tends to reduce the users' control, producing results that are unacceptable. In either case, the system is not successful. The introduction of automation in ways that still enable user control, e.g., as defaults or suggestions, can both reduce the effort of the user and enable acceptable results.

References

Adobe (2000), Premiere. http://www.adobe.com/products/premiere/

Apple (2000), iMovie. http://www.apple.com/imovie/

Apple (2000), Final Cut Pro. http://www.apple.com/finalcutpro/

Christel, M., Smith, M., Taylor, C. and Winkler, D. (1998), Evolving Video Skims into Useful Multimedia Abstractions, in *Human Factors in Computing Systems, CHI 98 Conference Proceedings*, ACM Press, pp. 171-178.

Girgensohn, A., Boreczky, J., Chiu, P., Doherty, J., Foote., J., Golovchinsky, G., Uchihashi, S., and Wilcox, L. (2000), A Semi-Automatic Approach to Home Video Editing, in *UIST '00 Proceedings*, ACM Press, pp. 81-89.

He, L., Sanocki, E., Gupta, A., and Grudin, J. (1999), Auto-Summarization of Audio-Video Presentations, in *ACM Multimedia 99*, ACM Press, pp. 489-498.

Kuchinsky, A., Pering, C., Creech, M.L., Freeze, D. Serra, B., and Gwizdka, J. (1999), FotoFile: A Consumer Multimedia Organization and Retrieval System, in *CHI 99 Conference Proceedings,* ACM Press, pp. 496-503.

Lienhart, R. (1999), Abstracting Home Video Automatically, in *ACM Multimedia 99 Proceedings* (Part 2), ACM Press, pp. 37-40.

Pfeiffer, S., Lienhart, R., Fischer, S. and Effelsberg, W. (1996), Abstracting digital movies automatically, in *Journal of Visual Communication and Image Representation*, 7(4), pp. 345-353.

Smith, M. and Kanade, T. (1997), Video Skimming and Characterization through the Combination of Image and Language Understanding Techniques, in *Proceedings of Computer Vision and Pattern Recognition*, pp. 775-781.

Ulead (2000), VideoStudio. http://www.ulead.com/vs/