Automated video editing: A five-point ratio scale for comparing levels of automation in video editing tools

Marco Braune[1]

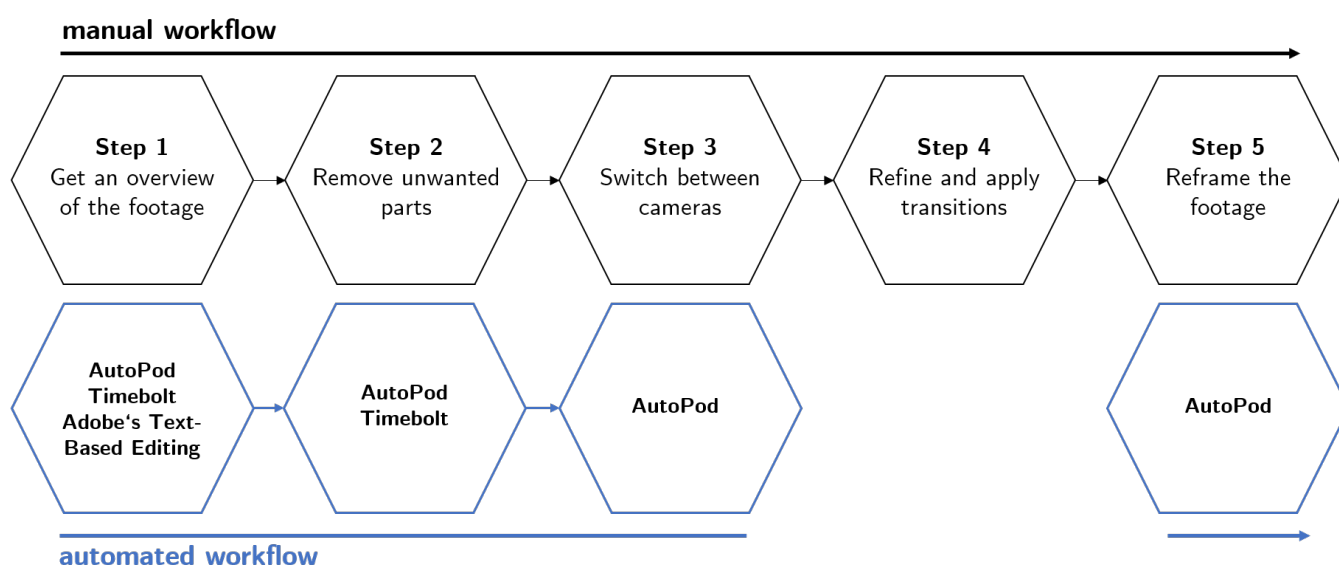[1]Film University Babelsberg *KONRAD WOLF*, Germany

Figure 1: *Comparison of the manual video editing workflow with a workflow that integrates automation tools*

Abstract

More and more software tools become available for video editors that expand the functionalities of video editing programs by automating processes. For a lot of editors, it is important to know how much work a tool still leaves for them, so they can bring in their own ideas. We propose a five-point ratio scale to compare the dependence of automation tools on the human editor. Each point stands for the number of steps a tool completes independently. This scale is based on a minimal workflow for editing a multi-camera video: watching the footage, removing unwanted parts, switching between cameras, refining transitions and reframing the shot. We compare the tools AutoPod, Timebolt and Adobe's Text-Based Editing. While AutoPod is almost an independent tool by scoring four points, Timebolt is more of a supportive tool with two points. To be able to also analyse the Text-Based Editing, which uses a different workflow than the one proposed above, we extend our method.

1. Introduction

Video editing software offers a lot of options to expand the creative possibilities for video creators. Nowadays, a lot of those options include automation of processes to assist the creators even further. On the one hand, this raises the question on how much creativity is left for the human video editor. On the other hand, it can take away a lot of time-consuming work and give the editor more time for other tasks, which could increase the creativity. But it is difficult to choose a tool, that is best suited for the editor and their current project. A categorisation of automation tools would be helpful to be able to differentiate on the different types of tools. But even in the general research of AI it is a problem, that the goals of each AI are different but not clearly distinguished [Wan08]. Because of

that, we propose a five-point ratio scale based on a basic editing workflow. We then conduct a case-study on three automation tools: AutoPod, Timebolt and Adobe's Text-Based Editing. These tools will be ranked on the proposed scale, to exemplarily show how automation tools can be compared. How the workflow with automation tools can differ from a manual workflow is shown in Figure 1.

2. Related Work

In previous work, the focus is either on automation processes and AI tools in video editing or on defining categories for AI in general. We analyse the results of both topics and combine them to define a method that can categorise automation tools, which are specifically used for video editing. Automation tools can be applied to a lot of different steps in the video editing process. Soe [Soe21a] found out, that the two most common tasks are segmentation and composition of a video. Because of that, we first take a look at automation tools that are handling tasks in this area of video editing. But it is important to keep in mind, that there are a lot of use cases even beyond the editing process. Automation tools can also be used to analyse already existing films and videos, on which AI can be trained and then used to anticipate the next shot for a sequence [AHL*22]. This could help AI tools, to assist the human editor even further. Another important aspect is, that not all video types are edited in the same way [ZLHW21]. Therefore, the ways the tools are working may also differ.

As mentioned above, there are different approaches for categorising AI. Already in 1995 Russel and Norvig [RN95] defined four categories: human-like systems and rational systems. Both further split into two sub-categories: Thinking and acting. In contrast, Wang [Wan08] defined categories based on their research goals, and defined five different categories. Each one of the categories contributes to different studies. A more abstract way for categorising AI is proposed by Sowa [Sow95]. In Sowa's work, AI categories are described by utilising ontological methods. This leads to a very complex set of categories. One of them - the Cyc - contains more than 10000 concepts. All these different approaches make it difficult, to find a common base for comparing automation tools. A more common base already exists for categorising methods for video editing by differentiating between levels of the editing process [ZLHW21], and for the workflow of video editing [Soe21b]. Soe [Soe21b] describes it as an essential part to understand the fundamental workflow to be able to analyse automation processes in video editing. This leads us to base our method on a common workflow and selecting tools that are applied on the same video editing level.

3. Comparison of automation tools

To be able to differentiate between automation tools in regard to the number of tasks they take over from the human editor, we first define a minimal workflow that a human editor follows when editing a multi-camera video. This workflow leads to a five-point ratio scale on which three exemplary tools are ranked.

3.1. Manual workflow for editing a multi-camera video

Every editor has to follow basic steps for editing a multi-camera video. The differences between editors lie in the way how each editor performs the steps and which decision they take within those steps. Soe [Soe21b] proposes a general workflow for video editing: It starts at searching for assets in the editor's library. Then it goes over adding the media files to the timeline, adding cuts, titles, audio, effects and reframing the shot by panning and adjusting the aspect ratio. Finally, the video can be published. Adobe [Ado23b] suggests a basic editing workflow, too, that applies to both novice and advanced editors. Going into a bit more details for each step, this results in the following workflow: First, a new project file has to be created. Secondly, the media files can be imported and then added to the timeline. Multiple clips have to be synchronised to be able to edit them together. To edit the timeline media clips, the editor scrolls through the timeline and watches the footage to find the wanted and unwanted parts. For a multi-camera setup, the clips for each speaker must be switched between enabled and disabled to show the different speakers. Once all the clips are cut, the transitions between clips can be refined by adding effects. After that, the editor may want to reframe the footage to fit it to the required aspect ratio. For reframing, the aspect ration needs to be set and the visible section of the potentially cropped video needs to be adjusted by repositioning and scaling the clip. Finally, the editor can continue by adding titles and additional visual effects, applying colour correction and grading, mixing the audio and as a last step exporting the file.

3.2. Tool selection

For the comparison we take a look at three different automation tools:

- AutoPod
- Timebolt
- Adobe's Text-Based Editing

Each of those tools can take over different tasks. AutoPod focuses on switching between multiple clips from different cameras, according to the current speaker in the video, to output a finished video podcast [Aut23a]. Timebolt analyses a single video clip for pauses and offers the option to delete all the pauses at once [Tim23]. By deleting all pauses automatically, it saves the editor time in making a video more fast-paced and makes it easier to skip from one take to another to find the best one. Adobe's Text-Based Editing transcribes the audio from the video and allows the user to edit the video based on the text instead of watching the video [Ado23c]. This way, the editor can search for specific text passages and is able to delete them instead of watching and cutting the video file directly. All of those tools affect the timeline and supports the editor in the cutting process. Additionally, all those tools can be used directly with *Adobe Premiere Pro*. This choice was made, because according to a survey from Soe [Soe21a] *Premiere Pro* is one of the most used software among video editors. These two aspects make it an exemplary selection of automation tools that follow the defined workflow above.

3.3. Comparison scale

All those three automation tools have an impact on only one segment of steps in the editing workflow. The first and last segments are still left for a human editor or different kind of tools. Extracting only the affected steps from the workflow defined above results in the following shortened workflow:

1. Scroll through timeline and watch the footage
2. Remove unwanted parts by adding cuts
3. Select which speaker is shown
4. Refine transitions between clips by adding effects
5. Reframe the footage to the required aspect ratio

Based on this minimal editing workflow, we an derive a five-point ratio scale, shown in Figure 2. For each step that an automation tool performs instead of a human editor, the tool scores a point. This score gives an impression on how dependent the tool is on the human editor. Zero points would define a workflow, that is done completely manual by a human editor. As described in the tool selection, each tool is able to take over at least one step. This leads to excluding zero from the scale. Five points define a completely independent automation tool. A score between zero and five indicates on how much support a tool offers.
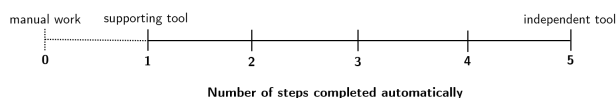

Figure 2: *Five-point ratio scale for ranking automation tools*

3.4. Comparison execution

All of the three tools are tested on a video podcast project with multiple cameras. The following subsections describe the way the tools are used and which steps they perform. The first steps before being able to watch and edit the footage on the timeline, as described in section 3.1, must be fulfilled for using these tools.

3.4.1. AutoPod

The *AutoPod Multi-Camera Editor* can be started as an extension of *Adobe Premiere Pro*. The tool then requires input for selecting the amount of speakers and cameras, speaker names and optional settings for the editing style, regarding the cutting method and the multi-shot frequency, as shown in Figure 3 Once, everything is set up, the software automatically cuts the video clips in realtime. It sets cuts and switches to the current speaker or another fitting clip - either a reaction shot from another speaker or a wide shot with multiple speakers - by disabling the other clips. To change the aspect ratio, AutoPod offers another extension called *AutoPod Social Clip Creator*. By selecting the desired aspect ratio, the *Social Clip Creator* crops the video accordingly and positions the video in a way, that the main action of a shot, often times the speaker, is in the centre of the video. For an option to automatically remove pauses inside a video, AutoPod comes with another feature: *AutoPod Jump Cut Editor*. According to the manually set silence dB cutoff, it analyses the video for silences and cuts them out. AutoPod does not have an option for automatically inserting transitions.

Figure 3: *Settings in AutoPod (Source: [Aut23b])*

3.4.2. Timebolt

First, the master audio track must be selected and added to Timebolt. The master audio track is the one, that contains audio from all cameras. This track can also be connected to a video file. A preview will then show, which parts of the audio will be removed based on the settings for the dB level, silence duration and padding, which are shown in Figure 4. These cuts can then be applied to all the clips. This outputs a timeline with cut video clips, but all of them are enabled. Based on those cuts, the editor can then jump from one segment to the next one, to detect unwanted takes or to disable a video clip to make another speaker visible. There are also workarounds to use the pauses from one video to find out which video clip is the one with the active speaker. But because this is a workaround and not completely automated, we do not consider this as a feature of Timebolt. There is also a feature that allows the user to select a transition effect that should be applied between all cuts. But it does not automatically decide which transition fits. Timebolt does not offer an option for reframing the video.

Figure 4: *Settings in Timebolt (Source: [Tim23])*

3.4.3. Adobe's Text-Based Editing

The Text-Based Editing, which comes as a native feature in *Adobe Premiere Pro*, lets the user transcribe the video by selecting the video file on the timeline and opening the text panel. The transcription is then divided into sections with timestamps attached. Dots symbolise pauses in the video. These pauses or parts of the text can be deleted, which directly affects the clips in the timeline by cutting out the segments that contain the selected phrases. An exemplary screenshot of the interface is shown in Figure 5 Beyond transcribing the video, it offers options to create subtitles from the generated text. It does not offer options for switching between speakers, adding transitions or reframing the footage. This shows, that it does not automate a lot of processes, but adds a lot of new ways of editing, that saves the editor time.
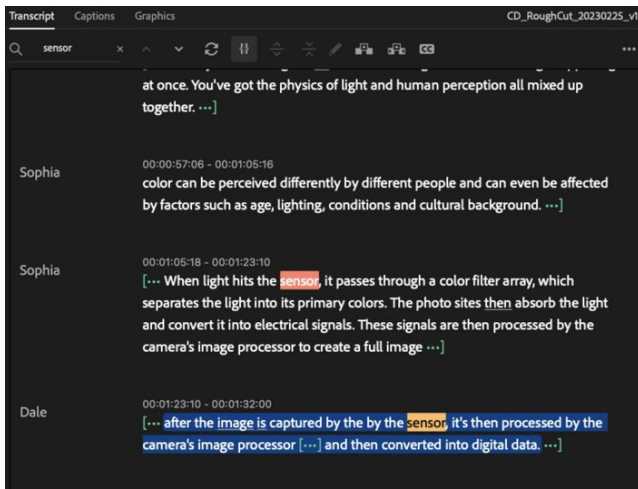

Figure 5: *Interface of Adobe's Text-Based Editing (Source: [Ado23c])*

3.5. Results

Using the tools AutoPod, Timebolt and Adobe's Text-Based Editing, have lead to an overview of what tasks these tools can handle. By mapping the abilities of the tools to the five steps described in section 3.3 we get the following results, shown in Table 1.

	AutoPod	Timebolt	Adobe's Text-Based Editing
Step 1	x	x	x
Step 2	x	x	-
Step 3	x	-	-
Step 4	-	-	-
Step 5	x	-	-

Table 1: *Overview of steps that can be automatically performed by the tools AutoPod, Timebolt and Adobe's Text-Based Editing*

An *x* indicates that a tool can perform this task, while - indicates that the tool cannot perform this task. We can now apply the five points ratio scale to assign a score to each tool. The final results of the scores are shown in Table 2.

Tool	Score
AutoPod	4
Timebolt	2
Adobe's Text-Based Editing	1

Table 2: *Scores for each tool based on a five-point ratio scale*

3.6. Evaluation

AutoPod can be described as the most independent automation tool in comparison to the other two. But it still is not completely independent from a human editor, as it cannot automatically apply transitions. But it automatically analyses the video, so the human editor does not need to watch the whole video anymore and it sets cuts to remove pauses and selects a fitting video clip according to the current speaker. Finally, the human editor does not need to reposition the shot anymore for a cropped video format, as the tool automatically detects the most important part in a video and positions it in the middle of the frame. This leads to a score of four point, meaning the human editor can have the most impact by adding custom transition effects after using the automation tool. But human editors also still have to make decisions when setting up the tool before starting it and they still have the option to make adjustments to the cuts and repositioning that were made by AutoPod. Even though the editor can still make changes, it is important to note, that a lot of editors feel more free in their decision making, when there is not already a suggestion on how to edit a video. Therefore, using a tool that scores a high number of points, is meant for editors, that do not only want support for tasks to save time but to also get help for making decisions.

Timebolt scores only two points. With that, the editor does not need to manually scroll through the whole timeline and cut out pauses. But again, there is the option for the editor to refine the work done by Timebolt. In comparison to the jump cut editor from AutoPod, it also offers more options for detecting pauses. This means, that even though they both score a point for performing step 2 automatically, it does not mean the same amount of work is necessarily taken away from the editor. Of course, the editor can still resume with the standard settings. This justifies the scoring of a point for this step, as it indicates the same possible independence of both tools. The editor can also use Timebolt to add transitions, but this can also be done directly in *Adobe Premiere Pro*. Thus, offering more ways to the editor for completing this step, but not taking it away from the editor. Because of no automation for adding transitions, there is no time saving and no independence from the editor in this step. Therefore, it does not score a point for the fourth step. By only scoring two points, it indicates that this tool can be used as a supporting tool for the editor to take over time consuming tasks. But it does not take away the option for creative input from the editor.

Adobe's Text-Based Editing depends the most on the human editor by scoring only one point. By using this tool, editors can get a quicker overview of the content of a video without watching it. Not only are pauses easily detectable, editors can also quickly search for fill words, duplicates which may come from multiple takes and sentenced or topics that may need to be deleted. As noted in section

3.4.3, it does not automate even the first step completely. Instead, it offers a new way for the editor to get an overview of the video content. The editor does not need to watch the video anymore, but can search for specific words or skim the text, which can be quicker than watching and therefore takes away the first step from the editor. For this reason, it still scores a point for the first step. Although, the cuts do not have to be done on the timeline but by editing the text, editors still have to find the unwanted parts, make decisions on which video clips should be shown, adding transition effects and reframe the footage in case another aspect ratio is required. It offers new ways of editing videos and can support the editor in steps beyond the defined five-steps workflow. But based on the five-point ratio scale it shows that this tool does not take over a lot of tasks, and therefore still leaving a lot of work and decisions to the editor.

4. Discussion

By making a comparison of automation tools, it shows that each tool follows their own workflow. This does not only apply to tools but also to human editors. For human editors we can follow official guidelines from software developers. But automation tools are developed to revolutionise the editing process. This makes it difficult to make an assumption of how many tasks a tool can perform, because it still depends on how it is used in the editing process. Adobe's Text-Based Editing stands out the most. It only scores one point, even though it changes the editing workflow more than the other two automation tools. To be able to also evaluate the number of new possibilities, to see what is the state-of-the-art in automation tools, we would need to extend our method. For that, it is not enough to only compare it to the established steps a human editor takes, but it must be compared to all current possible ways of editing. But with each tool that is developed, the possible ways are expanded, thus making it difficult to make this comparison. Our method does not include to give an overview on how advanced a tool is, but it can be applied as long as video editing software is not fundamentally changed. Therefore, an editor can get a quick impression on the influence of an automation tool before using it.

Another aspect when using the scale is to check which steps are performed by a tool. It is defined as a ratio scale, to make it easier for an editor to compare which tools takes over more tasks than another. It shows how high or low the dependence on the human editor is, but it does not show, how much the creative freedom may be limited by using one of those tools. There is no direct solution to it, as it depends on the editor and on the project in which steps and to which capacity the tool should assist the editor [Soe21a].

Furthermore, we selected three tools that are all working on the same part in the editing process. But there are a lot of more automation tools available for editors, for example Adobe offers AI tools for automatic colour matching, volume adjustments and even more [Ado23a]. All of the steps that are taken after the cutting process are already described in section 3.1. But to define a scale for each section or even a scale that includes the complete workflow, these steps must be analysed in more detail and broken down into more specific sub-tasks.

5. Future work

As the number of automation tools is increasing, it is also important to adjust the scale accordingly. There are already a lot of different tools available that focus on the cutting process. For now, they all mainly focus on the same steps, for example Gling [Gli23] can perform the same steps as Timebolt and AutoPod's Jump Cut Editor. But as Adobe's Text-Based Editing tool shows, there are ways to change the workflow and therefore offering possibilities for more variation in tools. The scale needs to be constantly updated to be able to include all the new steps that can be taken by editors. Then, it would be interesting to compare more than three tools. By comparing more tools, it could hint to possible adjustments for the scale. Additionally, it would be useful to look into automation tools that perform steps that were excluded in the five-step workflow. This could lead to finding a new method for comparing different kinds of automation tools. Finally, a survey conducted among video editors on how much the scores reflect on their personal impression of the dependence of automation tools would make sense to align the theoretical method to practical experiences.

6. Conclusion

The comparison of the automation tools AutoPod, Timebolt and Adobe's Text-Based Editing based on a five-point ratio scale shows, how different the dependence on the human editor is for each tool. AutoPod is an almost independent tool, as it automatically performs four out of five steps from the minimal workflow for cutting a video: it automatically analyzes the footage, so watching the footage is not necessary, it removes pauses, switches between cameras to show the most fitting video clip and it can reframe the shot by adjusting the position, when the format is cropped. The only step left for the human editor is refining and adding transitions. In contrast, Timebolt scores only two points by only performing the first two steps automatically. The most innovative tool, by adding new ways to the established workflow is Adobe's Text-Based Editing. But regarding the independence on the human editor, it only scores one point. For now, this five-point ratio scale is limited to the cutting process only. Nevertheless, for tools that assist in the cutting process, it offers video editors a quick impression on how much work is taken away from them. On the one hand, this can be time saving, on the other hand editors may feel limited in their creative freedom if too much work is taken away from them. After comparing the scores, it may become easier for the editor to decide which tool is the right one for their project.

References

[Ado23a] ADOBE: Adobe Sensei powers Creative Cloud. https://www.adobe.com/sensei/creative-cloud-artificial-intelligence.html, 2023. Viewed September 02, 2023. 5

[Ado23b] ADOBE: Get started with Adobe Premiere Pro. hhttps://helpx.adobe.com/premiere-pro/using/basic-workflow.html, 2023. Viewed August 22, 2023. 2

[Ado23c] ADOBE: Text-Based Editing in Premiere Pro. https://helpx.adobe.com/premiere-pro/using/text-based-editing.html, 2023. Viewed August 28, 2023. 2, 4

[AHL*22] ARGAW D. M., HEILBRON F. C., LEE J., WOODSON M., KWEON I. S.: The Anatomy of Video Editing: A Dataset and Benchmark Suite for AI-Assisted Video Editing. In *Computer Vision - ECCV 2022 - 17th European Conference, Tel Aviv, Israel, October 23-27, 2022, Proceedings, Part VIII* (2022), vol. 13668 of *Lecture Notes in Computer Science*, Springer, pp. 201–218. `doi:10.1007/978-3-031-20074-8\_12`. 2

[Aut23a] AUTOPOD: Automatic editing for video podcasts and shows. `https://www.autopod.fm/`, 2023. Viewed September 03, 2023. 2

[Aut23b] AUTOPOD: AutoPod Multi-Camera Editor. `https://www.youtube.com/watch?v=YtIBG5PjXIU`, 2023. Timestamp: 00:29, Viewed October 20, 2023. 3

[Gli23] GLING: Creator, get your time back. `https://www.gling.ai/`, 2023. Viewed September 02, 2023. 5

[RN95] RUSSEL S. J., NORVIG P.: *Artifical Intelligence - A Modern Approach*. Prentice Hall, 1995. 2

[Soe21a] SOE T. H.: AI video editing tools. What editors want and how far is AI from delivering? *CoRR abs/2109.07809* (2021). `doi:10.48550/arXiv.2109.07809`. 2, 5

[Soe21b] SOE T. H.: Automation in Video Editing: Assisted Workflows in Video Editing. In *Proceedings of the Workshop on Automation Experience at the Workplace, AutomationXP 2021, co-located with the ACM Conference on Human Factors in Computing Systems (CHI 2021), Online Workshop (originally Yokohama, Japan), May 7, 2021* (2021), vol. 2905 of *CEUR Workshop Proceedings*, CEUR-WS.org. 2

[Sow95] SOWA J. F.: Top-level ontological categories. *International Journal of Human-Computer Studies 43* (1995), 669–685. 2

[Tim23] TIMEBOLT LLC: Jump cut video app | Automatic Remove Silence. `https://www.timebolt.io/`, 2023. Viewed August 30, 2023. 2, 3

[Wan08] WANG P.: What Do You Mean by "AI"? *Frontiers in Artificial Intelligence and Applications 171* (2008), 362–373. 1, 2

[ZLHW21] ZHANG X., LI Y., HAN Y., WEN J.: AI Video Editing: a Survey. MDPI AG. `doi:10.20944/preprints202201.0016.v1`. 2