AI Assisted Video Workflow: Exploring UIs for Human-AI Collaboration

Than Htut Soe
University of Bergen
PO box 7802, 5020, Bergen,
Norway
than.soe@uib.no

The Adjunct Proceedings of ACM TVX2019, Manchester, United Kingdom, June 2019.

Abstract
Video production and distribution have become very affordable and accessible. A large body of research is available in machine learning for audio, visual and language processing and more recently generation of multimedia content. Machine learning provides promising materials for designing innovative video production workflows. However, there is a lack of studies and expertise around how would video editors receive and use machine learning in their work. As a part of ongoing university and industry joint innovation project, this project aims to explore the the challenges of integrating machine learning into video editing workflows. Through setting up and running experiments with AI embedded prototypes on video production workflow, we aim to explore the design space of using AI in video editing interfaces and potential of human-in-the-loop machine learning in creative designs.

Author Keywords
machine learning; video workflows; Intelligent user interfaces; video editing tools

ACM Classification Keywords
H.5.2 [User Interfaces]: Graphical user interfaces (GUI);
H.5.1 [Multimedia Information Systems]: Evaluation/methodology

Introduction

Videos play a very important role in sharing news and stories online. Various studies [5][8] have pointed out the shift in online video consumption behavior from branded news sites to third party online video sharing and social media platforms. This creates two new challenges for journalists and news agencies - to maintain presence in a lot of social media video sharing platforms and to be the first to publish on platforms. With various video sharing platforms to target, there is a need to support various video formats, viewing devices and viewing experiences. For instance, people expect videos on Facebook to be glance-viewable with text-on-video and without audio. The popularity of social media platforms also leads to creating a large number independent video makers with active audience bases.

Machine Learning (ML) technology has become widely adopted in the industry and thus a lot of methods, models, APIs and services are available. In audio and visual data processing, for example, ML has been used for speech recognition, text to speech, object detection, object tracking, scene detection, video captioning, sentiment analysis, video summarization and computational video editing. However, simply plugging ML into an video editing tool does not help user understand and utilize what ML could and could not do. Study of human factors and UI designs for ML is necessary to explore how users understand ML and affordances provided by introducing ML in the process. Dove et al. [2] stated in their study that ML is both underexplored opportunity for HCI researchers and has unknown potential as design material. The authors also pointed out the challenges which are the lack of understanding of ML in UX community, the data dependent nature of ML blackboxes and the difficulty of making interactive prototypes with ML.

With videos easier to produce and distribute than ever before and a lot of video platforms available, there is a need for more streamlined workflows for producing videos; often different versions of the same story for different platforms. In current video editing tools, there is nothing in between consumer level one size fit all tools and professional tools with all low level features like iMovie and Final Cut pro. However, machine learning has the potential to create better video editing tools but not without proper study of decision involved, methods and practices for better interfaces and interactions to connect human and machine learning in a creative work of making a video. For the purose of this project, a simplified video production workflow of is described in Figure 1.

In most parts of this project, a modified version of a video editing tool called Viz Story will be used to create interactive working prototypes. The rationale to use Viz Story is due to practical reasons which includes access to source code, access to internal expertise and chances to evaluate prototypes as a part of a complete workflow. Viz Story [11], is a browser-based complete package of tools for video production and publishing. Viz Story package contains full set of features to support the process of creating video stories and publish multiple versions of them to different platforms.

Related Work

Roughcut [7] is a tool that utilizes machine learning for video editing of dialog driven scenes with a model which can be described roughly as design by description. Given raw video takes and a dialogue, Roughcut enables creative exploration of various editing styles described by editing idioms. PotraitSketch [12] is an interactive drawing system with automated assistance for drawing face sketches. It provides a new way for novices to create face sketches without requiring mastery of drawing techniques. BBC Research

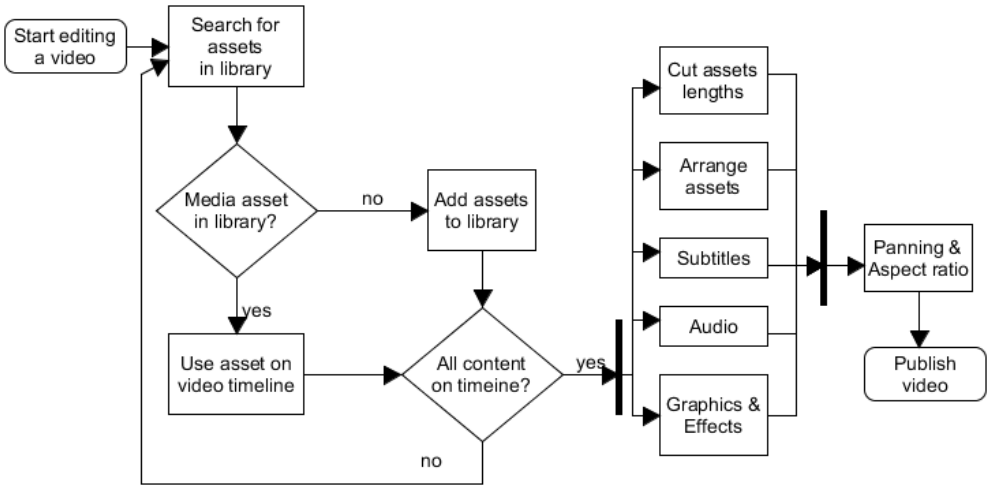

Figure 1: Tasks involved in a video worlflow.

tool named Audiogram [6] allows repurposing of radio content into vidoes with animated wave-forms and subtitles. AutoEdit [9] is a open source tools which uses automated speech recognition to transcribe the audio and allows editors to make selections on the video using text.

Another field where machines can help humans perform creative tasks is Writing with machines in the loop. Clark et.al. [1] performed an experiment with two machine-in-the-loop systems for story writing and slogan writing tasks and the participants enjoyed collaborating with machine even though third-party evaluations rated of stories written with machine-generated suggestions are not as good as stories written by humans alone. Visual story telling models generates descriptions of a series of pictures that describes an event. Hus et.al [4] analyzed how humans edits those machine-generated text. Explainable Artificial Intelligence (XAI) is an emerging field in machine learning to come up with techniques that are more explainable to human users[3].

Objectives

The main purpose of the research is to answer the questions how human and AI perform together in collaboration as human-in-the-loop for video workflow tasks. To answer the question, it is necessary to identify what are the interaction design challenges and how do we measure efficiency for video workflow tasks.

Working at the intersection of HCI and AI research in video workflow context, the aim of this project are as follows

- To reveal interaction design challenges for human-in-the-loop tasks in video workflow.
- To measure performance of improvements, if any, by using AI and human-in-the loop.

- To provide insights on people opinions into AI being pushed into their workflow and how it influences their experiences and creativity.

Research Methods

As this research project is situated at the intersection of HCI, machine learning and video workflow practices, it requires research methodology from all three fields. Research through design [13] approach serves as a method of inquiry through designer practices in designing for machine learning in video production context. Empirical research approach with experiments on interactive prototypes will be used to observe and collect the data about video workflows tasks with human-in-the-loop.

Current Work

Automated Subtitling using ML Subtitling has been identified as a time consuming task for video production [10]. To explore the challenges of integrating ML based subtitling in the video workflow, a prototype on Viz Story has been implemented. Upon uploading video materials to edit, subtitles are generated using online ML based speech recognition APIs. Those generated subtitles are then included as starting subtitles for the video creators to begin their subtitling task. First pilot test has been conducted internally with two persons on this prototype. Each of the participants is assigned to provide English subtitles for two short video clips. One of the clip has no subtitles and another has ML generated subtitles. During the pilot test, we captured the screen, recorded time taken and saved the subtitles. After that, two participants provided open feedbacks.

The ML generated subtitles has around 10% word error rate. From observing the pilot test, it is still time consuming for the users to manually fix errors found in automated subtitles. With the UI lacking features to point out ML inaccuracies, the participants questioned the quality of whole ML generated subtitles. Pilot testing on this first prototype highlights the need to design proper UI around pro and cons of ML based speech recognition. It revealed some challenges of integrating ML services into the UI and it is the first step to understand how ML will impact the workflow. The second prototype has been iteratively designed with consideration of having automated speech recognition in the loop. An experiment will be conducted on both prototypes to measure empirical data.

Future Work

Future work In the short term, different functional prototypes for automated subtitling and automated object tracking and automated content adaptation will be iteratively designed. Throughout the design process we hope to discover the issues and challenges of designing ML assisted video editing interfaces. Developed prototypes will then be used to perform a study. The purpose of the studies are to reveal design considerations and issues that exists in designing ML assisted editing interfaces and impact of ML in the video workflow process.

Another interesting question is "will automation of tedious and repetitive tasks leads to more time spent on creative tasks?". After the first study is completed, we are considering an ML system for generating various styles of video stories using something called smart video templates. As video makers usually follow their own style as well as that of the organization, can we apply machine learning techniques in this exploratory task of video editing. The project hopes to explore the area of applied machine learning in the video production.

Acknowledgements

This research is supported by the Research Council of Norway through the User-driven Research based Innovation (BIA) programme. Special thanks to Vizrt UX team at Bergen Norway for their valuable input.

REFERENCES

1. Elizabeth Clark, Anne Spencer Ross, Chenhao Tan, Yangfeng Ji, and Noah A. Smith. 2018. Creative Writing with a Machine in the Loop: Case Studies on Slogans and Stories. In *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval - IUI '18*. ACM Press, Tokyo, Japan, 329–340. DOI: http://dx.doi.org/10.1145/3172944.3172983
2. Graham Dove, Kim Halskov, Jodi Forlizzi, and John Zimmerman. 2017. UX Design Innovation: Challenges for Working with Machine Learning as a Design Material. ACM Press, 278–288. DOI: http://dx.doi.org/10.1145/3025453.3025739
3. David Gunning. 2017. Explainable Artificial Intelligence (XAI). (Nov. 2017), 38.
4. Ting-Yao Hsu, Yen-Chia Hsu, and Ting-Hao 'Kenneth' Huang. 2019. On How Users Edit Computer-Generated Visual Stories. *arXiv:1902.08327 [cs]* (Feb. 2019). http://arxiv.org/abs/1902.08327 arXiv: 1902.08327.
5. Antonis Kalogeropoulos. 2018. Online News Video Consumption. *Digital Journalism* 6, 5 (May 2018), 651–665. DOI: http://dx.doi.org/10.1080/21670811.2017.1320197
6. BBC News Lab. 2019. Audiogram Generator. (2019). /projects/Audiograms/
7. Mackenzie Leake, Abe Davis, Anh Truong, and Maneesh Agrawala. 2017. Computational video editing for dialogue-driven scenes. *ACM Transactions on Graphics* 36, 4 (July 2017), 1–14. DOI: http://dx.doi.org/10.1145/3072959.3073653
8. Nic Newman, David A. L. Levy, and Rasmus Kleis Nielsen. 2015. Reuters Institute Digital News Report 2015. *SSRN Electronic Journal* (2015). DOI: http://dx.doi.org/10.2139/ssrn.2619576
9. Pietro Passarelli. 2019. autoEdit Fast Text Based Video Editing. (2019). http://www.autoedit.io/
10. Hassan Sawaf. 2012. Automatic speech recognition and hybrid machine translation for high-quality closed-captioning and subtitling for video broadcast. *Proceedings of Association for Machine Translation in the AmericasâĂŞAMTA* (2012), 5.
11. Vizrt. 2019. Viz Story. (2019). https://www.vizrt.com/products/viz-story
12. Jun Xie, Aaron Hertzmann, Wilmot Li, and Holger WinnemÃűller. 2014. PortraitSketch: face sketching assistance for novices. In *Proceedings of the 27th annual ACM symposium on User interface software and technology - UIST '14*. ACM Press, Honolulu, Hawaii, USA, 407–417. DOI: http://dx.doi.org/10.1145/2642918.2647399
13. John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07*. ACM Press, San Jose, California, USA, 493. DOI: http://dx.doi.org/10.1145/1240624.1240704