Pre-screen: Assisting Material Screening in Early-stage of Video Editing

Qian Zhu*, Shuai Ma*, Cuixia Ma
HCI Lab, Institute of Software, Chinese Academy of Sciences
School of Computer Science and Technology, University of Chinese Academy of Sciences
{zhuqian172, mashuai171}@mails.ucas.ac.cn cuixia@iscas.ac.cn

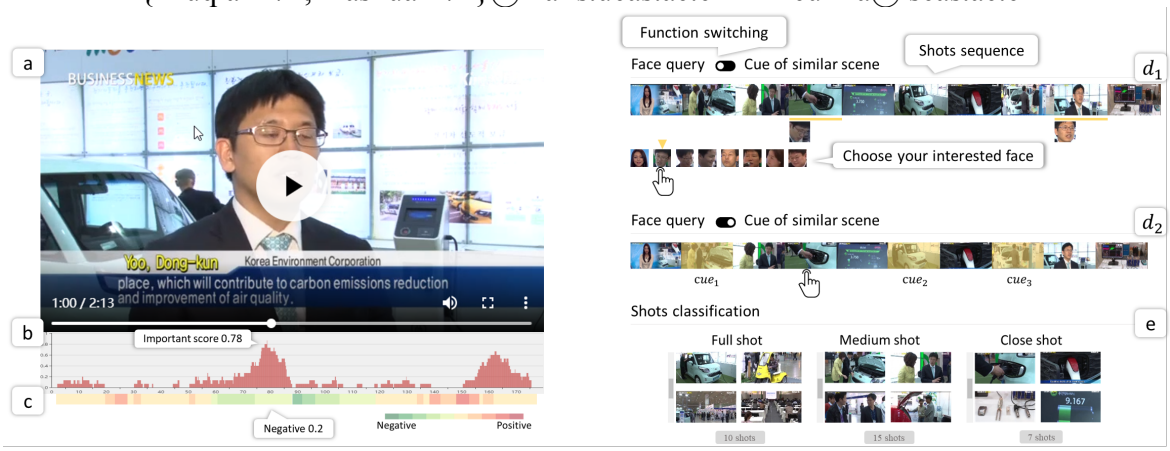

Figure 1. Overview of Pre-screen. Pre-screen adopts several video content analysis methods to help user screen and organize shots in both global view and detailed view, it provides a novel interface with a video player and five support functions: (a) A video player; (b) Importance curve; (c) Emotion analysis; (d_1) Face query; (d_2) Cue of similar scene; (e) Shots classification.

ABSTRACT

Video editing is a difficult task for both professionals and amateur editors. One of the biggest reasons is that screening useful clips from raw material in the early stage of editing is too much time-consuming and laborious. To better understand current difficulties faced by users in editing task, we first conduct a pilot study involving a survey and an interview among 20 participants. Based on the results, we then design Pre-screen, a novel tool to provide users with both global-view and detailed-view video analysis as well as material screening features based on intelligent video processing, analysis and visualization methods. User study shows that Pre-screen can not only effectively help users screen and arrange raw video material to save much more time than a widely used video editing tool in video editing tasks, but also inspire and satisfy users.

Author Keywords

Video content analysis; video editing; material screening.

CCS Concepts

• Human-centered computing~Interactive systems and tools;

INTRODUCTION

Video editing is the manipulation and arrangement of video shots [5]. An essential part of the whole editing process is screening useful shots from all source material. Generally, most editors need to do their manual organization and screening of complex materials, which is time-consuming and laborious. Researchers have proposed various methods and systems to help users create a video/film. One way is to use Artificial Intelligent (AI) video processing technologies to provide video content summaries and generate trailers [9, 10, 14]. But editing is a so creative job that results provided by such automatic video processing methods can't replace the work of editors. Another way is to propose frameworks and interactive systems for assisting the video editing process [7, 16], and there are many mature video editing tools on the market, such as Final Cut Pro [1], Adobe Premiere [2], iMovie [3] and so on. However, most of the tools are oriented to the middle or later stage of the editing process, thus the time-consuming material screening and arrangement work in the early stage are often neglected.

In order to solve these problems, we first conducted a preliminary survey to investigate users' needs for clips and their attitudes towards material screening. Based on the results, we propose Pre-screen with the help of video auto-processing algorithms and visualization. We constructed intuitive video content expression from global and detailed views to help editors quickly search, arrange and screen video materials, and provide a basis for story editing and conceiving in the next step. User study shows that our

* Both authors contributed equally to this work.

UIST '19 Adjunct, October 20–23, 2019, New Orleans, LA, USA.

ACM ISBN 978-1-4503-6817-9/19/10.
https://doi.org/10.1145/3332167.3357112

system can greatly reduce the pre-editing time and improve the efficiency of editing work. At the same time, users are satisfied with Pre-screen and enjoy interacting with it.

SYSTEM DESIGN

Preliminary survey

We invited 20 participants (10M, 10F) to fill an online questionnaire, 8 were professional editors and 12 were amateurs but often participated in editing work in universities. And we interviewed 3 editors, two engaged in film and television later-stage editing and one was a video advertising editor and director. The 20 participants had over 2 years of editing experience on average. We've got a lot of useful results, but briefly describe them to save space. 80% of the participants thought that the material screening stage accounted for a large proportion of the whole editing work (more than 50%). Notably, 100% of the participants believe that material selection is a very important step which determines the content and quality of the entire edited work. More importantly, we got some design requirements which contribute to the system design shown as below.

Video content analysis from two views

Firstly, we divided the original video material into shots for getting a shot sequence using automatic frame-cutting toolkit [4]. Then we provided two views of video material analysis and visualization assistance to help users with material screening in the pre-editing stage.

Global view: This part contains two support functions. Figure 1 (b) shows the *Importance curve* which provides users with an importance score curve for different parts of a video. We used LSTM neural network-based video summarization technology [12] to get the importance score of different parts in a video with the content expression of diversity and representativeness considered. Users can observe the importance of all video materials from a global perspective and explore different shots based on it. Figure 1 (c) shows the *Emotion analysis* function which provides a linear-bar of emotional change according to the emotional positivity contained in different shots. Through text emotion analysis [15] and speech emotion classification [8] in video content, we get a continuous emotional change conveyed by a video, which is represented by a visual linear heatmap. Based on it, users can directly locate a part of the video, which will help control and balance the proportion of various emotional material from an emotional perspective.

Detailed view: This part includes three support functions. The first one is *Face query* in shots shown in figure 1 (d_1). According to the results of preliminary study, editors usually need to find clips specified with a certain person to edit. To satisfy this need, we used face recognition [6] to detect and recognize all faces in every shot in advance and get a face list in which faces appeared. Users can select a specific person and get cues of all the clips that contain this person. The second one is *Cue of similar scene*. Editors usually adjust the position of video clips in the clipping time track according to the scene. As shown in figure 3 (d_2), we evaluated the scene similarity between shots according to the similarity of the visual features extracted from different shots [13]. When a user chooses a shot, the similar shots will be highlighted with tags, which can help users find the video clips they need more quickly. Note that, the above two functions can be switched alternatively by a button. The third one is *Shots classification* shown in figure 1 (e). From the survey results, editors need to pay attention to such factors as lens switching or depth of field changing. So, we used an object detection method [11] based on depth neural network to calculate the area ratio of the detected object and the whole scene from which we can infer the depth of field. Then we classified shots in video segments and got three kinds of shot, full shot, medium shot and close shot, which makes material search more effective.

USER STUDY

To verify the usefulness of Pre-screen, we deployed our system on a website and invited the same participants in pre-survey to use it. We divided participants into two groups and ensured that the two groups have the same editing experience. Then, we let participants do two video editing tasks from raw video material each with a total length of 2 hours. 10 participants screened the first video manually and edited it with Adobe Pr (we name the total process Baseline). After that, they screened the second video with Pre-screen and edited it with Adobe Pr (we name the total process Pre-screen). The other group is the opposite. We instructed participants on how to use the two systems and encouraged them to play with all the functions before starting this task. And we recorded the time spent on this task. In the end, participants were required to fill a post-task questionnaire on a 5-point Likert scale.

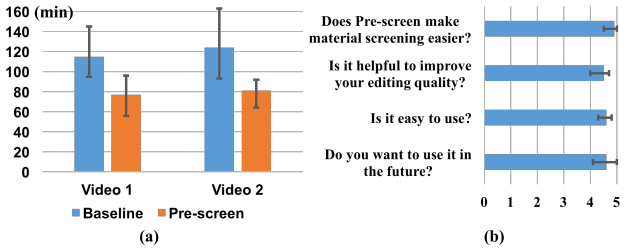

Figure 2. Results of user study. (a) shows the time spent with two systems; (b) shows the result in the questionnaire.

Result

Figure 2 (a) shows the time in the video editing tasks. It can be seen that Pre-screen outperforms Baseline by a large margin both in the two tasks. Based on pair-wise t-tests we found that the difference is significant ($p<0.001$). Figure 2 (b) shows that users were overall satisfied with Pre-screen.

DISCUSSION AND FUTURE WORK

The video content analysis methods adopted in our work are not perfect, which may affect users' experience. In the future, we will further explore the interaction between intelligent automation algorithms and editors. How to let the editors and artificial intelligence algorithm cooperate to create an excellent video is a direction worth studying.

REFERENCES

[1] https://www.apple.com/final-cut-pro/

[2] https://www.adobe.com/products/premiere.html

[3] https://www.apple.com/imovie/

[4] https://pyscenedetect.readthedocs.io/

[5] The Art Of Film And Video Editing. *Video University.* 2012. https://www.videouniversity.com/articles/.

[6] Schroff, F., Kalenichenko, D., & Philbin, J. 2015. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 815-823.

[7] Leiva, G., & Beaudouin-Lafon, M. 2018. Montage: A Video Prototyping System to Reduce Re-Shooting and Increase Re-Usability. In *The 31st Annual ACM Symposium on User Interface Software and Technology*, 675-682.

[8] Trigeorgis, G., Ringeval, F., Brueckner, R., Marchi, E., Nicolaou, M. A., Schuller, B., & Zafeiriou, S. 2016. Adieu features? end-to-end speech emotion recognition using a deep convolutional recurrent network. In *2016 IEEE international conference on acoustics, speech and signal processing (ICASSP),* 5200-5204.

[9] Xu, H., Zhen, Y., & Zha, H. 2015. Trailer generation via a point process-based visual attractiveness model. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*.

[10] Smith, J. R., Joshi, D., Huet, B., Hsu, W., & Cota, J. 2017. Harnessing ai for augmenting creativity: Application to movie trailer creation. In *Proceedings of the 25th ACM international conference on Multimedia*, 1799-1808.

[11] He, K., Zhang, X., Ren, S., & Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778.

[12] Zhang, K., Chao, W. L., Sha, F., & Grauman, K. 2016. Video summarization with long short-term memory. In *European conference on computer vision*, 766-782.

[13] Tapaswi, M., Bauml, M., & Stiefelhagen, R. 2014. Storygraphs: visualizing character interactions as a timeline. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 827-834.

[14] Merler, M., Mac, K. N. C., Joshi, D., Nguyen, Q. B., Hammer, S., Kent, J., & Feris, R. S. 2018. Automatic curation of sports highlights using multimodal excitement features. *IEEE Transactions on Multimedia*, *21*(5), 1147-1160.

[15] Loria, S. 2018. textblob Documentation, 1-73. *Technical report.*

[16] Gandhi, V., & Ronfard, R. 2015. A computational framework for vertical video editing.