EUROGRAPHICS 9th Workshop on Cinematography and Editing (2020)
M. Christie, H.-Y. Wu, T.-Y. Li and V. Gandhi (Editors)

GAZED– Gaze-guided Cinematic Editing of Wide-Angle Monocular Video Recordings

K.L.Bhanu Moorthy[1], Moneish Kumar[2], Ramanathan Subramaniam[3] and Vineet Gandhi[1]

[1]CVIT, IIIT Hyderabad, India
[2]Samsung R&D Institude, Bengaluru, India
[3]IIT Ropar, India

Abstract

*We present **GAZED**– eye GAZ-guided EDiting for videos captured by a solitary, static, wide-angle and high-resolution camera. Eye-gaze has been effectively employed in computational applications as a cue to capture interesting scene content; we employ gaze as a proxy to select shots for inclusion in the edited video. Given the original video, scene content and user eye-gaze tracks are combined to generate an edited video comprising of cinematically valid actor shots and shot transitions to generate an aesthetic and vivid representation of the original narrative. We model cinematic video editing as an energy minimization problem over shot selection, whose constraints capture cinematographic editing conventions. Gazed scene locations primarily determine the shots constituting the edited video. Effectiveness of GAZED against multiple competing methods is demonstrated via a psychophysical study involving 12 users and twelve performance videos.*

Professional video recordings of stage performances are typically created by employing skilled camera operators, who record the performance from multiple viewpoints. These multi-camera feeds, termed rushes, are then edited together to portray an eloquent story intended to maximize viewer engagement. Generating professional edits of stage performances is both difficult and challenging. Firstly, maneuvering cameras during a live performance is difficult even for experts as there is no option of retake upon error, and camera viewpoints are limited as the use of large supporting equipment (trolley, crane .) is infeasible. Secondly, manual video editing is an extremely slow and tedious process and leverages the experience of skilled editors. Overall, the need for (i) a professional camera crew, (ii) multiple cameras and supporting equipment, and (iii) expert editors escalates the process complexity and costs.

Consequently, most production houses employ a large field-of-view static camera, placed far enough to capture the entire stage. This approach is widespread as it is simple to implement, and also captures the entire scene. Such static visualizations are apt for archival purposes; however, they are often unsuccessful at captivating attention when presented to the target audience. While conveying the overall context, the distant camera feed fails to bring out vivid scene details like close-up faces, character emotions and actions, and ensuing interactions which are critical for cinematic storytelling.

*GAZED denotes an end-to-end pipeline to generate an aesthetically edited video from a single static, wide-angle stage recording. This is inspired by prior work [GRG14], which describes how a plural camera crew can be replaced by a single high-resolution static camera, and multiple virtual camera shots or rushes generated by simulating several virtual pan/tilt/zoom cameras to focus on actors and actions within the original recording. In this work, we demonstrate that the multiple rushes can be automatically edited by leveraging user eye gaze information, by modeling (virtual) shot selection as a discrete optimization problem. Eye-gaze represents an inherent guiding factor for video editing, as eyes are sensitive to interesting scene events [RKH*09, SSSM14] that need to be vividly presented in the edited video.*

The objective critical for video editing and the key contribution of our work is to decide which shot (or rush) needs to be selected to describe each frame of the edited video. The shot selection problem is modeled as an optimization, which incorporates gaze information along with other cost terms that model cinematic editing principles. Gazed scene locations are utilized to define gaze potentials, which measure the importance of the different shots to choose from. Gaze potentials are then combined with other terms that model cinematic principles like avoiding jump cuts (which produce jarring shot transitions), rhythm (pace of shot transitioning), avoiding transient shots . The optimization is solved using dynamic programming. [MKSG20] refers to the detailed full article.

CCS Concepts

• ***Information systems*** → ***Multimedia content creation;*** • ***Mathematics of computing*** → *Combinatorial optimization;* • ***Computing methodologies*** → *Computational photography;* • ***Human-centered computing*** → *User studies;*

Keywords: Eye gaze, Cinematic video editing, Stage performance, Static wide-angle recording, Gaze potential, Shot selection, Dynamic programming

DOI: 10.2312/wiced.20201130

References

[GRG14] GANDHI V., RONFARD R., GLEICHER M.: Multi-clip video editing from a single viewpoint. In *Proceedings of the 11th European Conference on Visual Media Production* (2014), ACM, p. 9. 1

[MKSG20] MOORTHY K. L. B., KUMAR M., SUBRAMANIAN R., GANDHI V.: Gazed gaze-guided cinematic editing of wide-angle monocular video recordings. Association for Computing Machinery. 1

[RKH*09] RAMANATHAN S., KATTI H., HUANG R., CHUA T.-S., KANKANHALLI M.: Automated localization of affective objects and actions in images via caption text-cum-eye gaze analysis. In *ACM International Conference on Multimedia* (2009), p. 729–732. 1

[SSSM14] SUBRAMANIAN R., SHANKAR D., SEBE N., MELCHER D.: Emotion modulates eye movement patterns and subsequent memory for the gist and details of movie scenes. *Journal of vision 14*, 3 (2014), 1–18. 1