A Saliency-driven Video Magnifier for People with Low Vision

Ali Selman Aydin†
Department of Computer Science
Stony Brook University
Stony Brook, NY, USA
aaydin@cs.stonybrook.edu

Shirin Feiz†
Department of Computer Science
Stony Brook University
Stony Brook, NY, USA
sfeizdisfani@cs.stonybrook.edu

Vikas Ashok
Department of Computer Science
Old Dominion University
Norfolk, VA, USA
vganjigu@cs.odu.edu

IV Ramakrishnan
Department of Computer Science
Stony Brook University
Stony Brook, NY, USA
ram@cs.stonybrook.edu

ABSTRACT

Consuming video content poses significant challenges for many screen magnifier users, which is the "go to" assistive technology for people with low vision. While screen magnifier software could be used to achieve a zoom factor that would make the content of the video visible to low-vision users, it is oftentimes a major challenge for these users to navigate through videos. Towards making videos more accessible for low-vision users, we have developed the SViM video magnifier system [6]. Specifically, SViM consists of three different magnifier interfaces with easy-to-use means of interactions. All three interfaces are driven by visual saliency as a guided signal, which provides a quantification of interestingness at the pixel-level. Saliency information, which is provided as a heatmap is then processed to obtain distinct regions of interest. These regions of interests are tracked over time and displayed using an easy-to-use interface. We present a description of our overall design and interfaces.

CCS CONCEPTS

• **Human-centered computing ~ Accessibility technologies**

KEYWORDS

Screen magnifier users, video accessibility

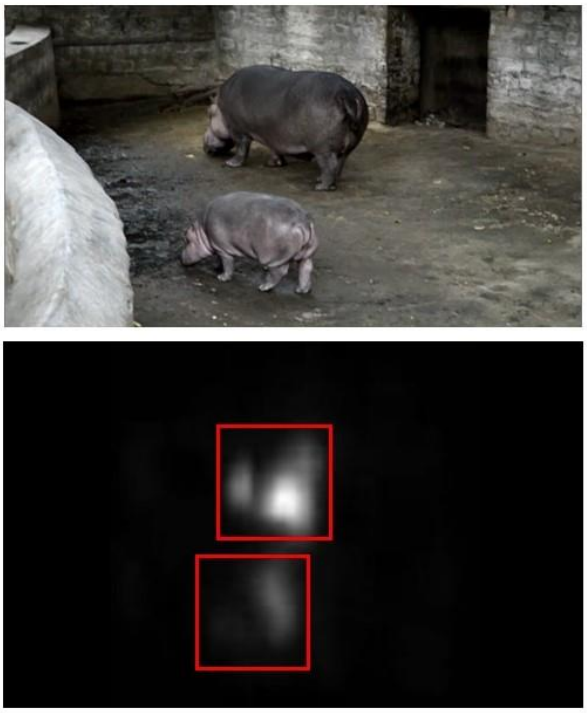

Figure 1: Top: Original image. Bottom: Heatmap and sample ROIs(Photo at the top is derived from video by Freestocks (CC BY 3.0)).

1 INTRODUCTION

Existing screen magnifiers make it easier to interact with desktops [2,3] and mobile devices for people with low vision. Although such magnifier interfaces are useful in many daily tasks, the dynamic and rich content of videos create a significant challenge for screen magnifier users. Keeping track of interesting parts of a video requires manual navigation on the user's part. The users also risk

†Equal contribution

W4A 20, April 20-21, 2020, Taipei, Taiwan

ACM ISBN 978-1-4503-7056-1/20/04.
https://doi.org/10.1145/3371300.3383356

missing important content outside the viewport when they use a screen magnifier to watch a video.

We developed SViM (Saliency-driven Video Magnifier) as part of our previous work. The idea behind SViM system is to predict which regions of a video would be of interest to the users by predicting the video saliency using a state-of-the-art neural network – DeepVS [1]. SViM provides three interfaces with varying functionalities.

2 SYSTEM DESIGN

Our system is composed of several components that determine regions of interest and track these regions of interest over time and ensure smooth transitions between these regions of interest while the video is being watched.

2.1 Determining ROIs

Our system utilizes DeepVS video saliency network [1] to predict saliency across a video. DeepVS consists of two subnets to determine objectness and motion and a sequential model to learn temporal relationships across the frames. DeepVS outputs a heatmap for each frame in the video that quantifies how salient is each region in the frame. We post-process this frame-wise saliency information to find one or more regions of interest (ROIs). For example, in a video that contains a conversation between two people, we expect to have two regions of interest, each of which correspond to their faces.

Determining exact number of ROIs corresponds to finding grouping (clusters) of salient pixels in an unsupervised manner. To do so, we employ a hierarchical procedure that detect accumulations of salient regions and decide whether to further divide this region or not.

2.2 Tracking ROIs Over Time & Smoothing

ROIs in a frame need to be tracked accurately over time in order to present a smooth experience to the users. Although the movement between successive frames is minimal (i.e., order of few pixels), following the ROIs without stabilizing the viewport may result in jitter, thereby disrupting viewing experience. To mitigate the effect of jitter caused by the movement between the successive frames, we employ a Kalman [4] filter based smoothing mechanism. We also prevent adjusting of viewport across the video unless the accumulated amount of movement is larger than a predefined threshold.

2.3 Display Interfaces

We implemented three interfaces as part of our system: (i) $SViM_B$ (SViM-Basic) which is guided by ROIs and tracks them over time, (ii) $SViM_M$ (SViM-Mixed mode) which provides users with the ability to navigate freely using the mouse in addition to the functionality provided by $SViM_B$, and $SViM_L$ (SViM-Lens Mode) which provides a lens-mode zooming that magnifies ROIs only while keeping the background static.

$SViM_B$ interface starts by fixating on one of the ROIs which is tracked by default. If any other ROIs exist, the user can switch to those ROIs with a mouse click in a round-robin manner. $SViM_M$ offers an additional functionality to the one provided by $SViM_B$ that allows users to explore the video if they demand. This exploration mode is invoked by moving the mouse at an arbitrary direction. Once the exploration mode is invoked the user can navigate the video using mouse movements. The user can exit the navigation mode by clicking an arbitrary location, which will re-invoke the tracking mode. $SViM_L$ stems from the idea that the regions of interest

In addition to ROI-guided functionality described above, we implement an easy to use interface for changing zoom factor using mouse wheel, and commonly used visual transformations(e.g., contrast enhancement, sharpening, brightness adjustment and inversion) that are shown to improve visibility for users with certain conditions.

3 EVALUATION

We conducted a user study with 13 participants to assess effectiveness of SViM system [6], in which we compared the three prototypes we developed to the Windows screen magnifier [2] and VLC Player magnifier [5]. Both subjective results and our measurements have demonstrated that the users had a better experience with SViM interfaces compared to the baseline magnifiers. SViM system works in real-time on a PC once the saliency heatmaps are computed offline. Since the heatmaps are computed only once, it is possible to perform heatmap computation in servers with GPUs, which makes it possible to deploy the system on the web. Our main focus for future work would be to improve the accuracy of ROI detection and tracking algorithms.

ACKNOWLEDGMENTS

This work was supported by NSF Award: 1806076, NEI/NIH Award: R01EY026621, and NIDILRR Award: 90IF0117-01-00.

REFERENCES

[1] Jiang, L., Xu, M., Liu, T., Qiao, M. and Wang, Z., 2018. Deepvs: A deep learning based video saliency prediction approach. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 602-617).

[2] Microsoft Corporation. 1998. *Magnifier.*[Software][Accessed: 2019]

[3] Apple Inc.. *Zoom.*[Software][Accessed: 2019]

[4] Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. Journal of basic Engineering, 82(1), pp.35-45.

[5] VideoLAN organization. 2001. *VLC Media Player.*[Software][Accessed: 2019]

[6] Aydin, A.S., Feiz, S., Ashok, V. and Ramakrishnan, I. "Towards Making Videos Accessible for Low Vision Screen Magnifier Users" in 25th International Conference on Intelligent User Interfaces(IUI), 2020(To appear).