2020

Towards Making Videos Accessible for Low Vision Screen Magnifier Users

Ali Selman Aydin

Shirin Feiz

Vikas Ashok
Old Dominion University

IV Ramakrishnan

Original Publication Citation

Aydin, A. S., Feiz, S., Ashok, V., & Ramakrishnan, I. V. (2020). Towards making videos accessible for low vision screen magnifier users. IUI '20: International Conference on Intelligent User Interfaces, March 2020, Cagliari, Italy. https://doi.org/10.1145/3377325.3377494

HHS Public Access

Author manuscript

IUI. Author manuscript; available in PMC 2021 February 09.

Published in final edited form as:

IUI. 2020 March ; 2020: 10–21. doi:10.1145/3377325.3377494.

Towards Making Videos Accessible for Low Vision Screen Magnifier Users

Ali Selman Aydin[*],
Stony Brook University

Shirin Feiz[*],
Stony Brook University

Vikas Ashok,
Old Dominion University

IV Ramakrishnan
Stony Brook University

Abstract

People with low vision who use screen magnifiers to interact with computing devices find it very challenging to interact with dynamically changing digital content such as videos, since they do not have the luxury of time to manually move, i.e., pan the magnifier lens to different regions of interest (ROIs) or zoom into these ROIs before the content changes across frames.

In this paper, we present SViM, a first of its kind screen-magnifier interface for such users that leverages advances in computer vision, particularly video saliency models, to identify salient ROIs in videos. SViM's interface allows users to zoom in/out of any point of interest, switch between ROIs via mouse clicks and provides assistive panning with the added flexibility that lets the user explore other regions of the video besides the ROIs identified by SViM.

Subjective and objective evaluation of a user study with 13 low vision screen magnifier users revealed that overall the participants had a better user experience with SViM over extant screen magnifiers, indicative of the former's promise and potential for making videos accessible to low vision screen magnifier users.

Keywords

accessible videos; low vision; screen magnifiers; video magnifiers

aaydin@cs.stonybrook.edu.
[*]Both authors contributed equally to this research.

1 INTRODUCTION

Low vision is characterized as a visual impairment that cannot be fully corrected even with glasses, medication or surgery and is severe enough to interfere with daily functioning but allows for some residual usable vision. It often manifests as loss of sharpness or acuity (ranging from 20/70 to 20/400 [54]), but may be present as decreased field of vision, loss of peripheral or central vision, blurred vision, extreme light sensitivity, tunnel vision, and near-total blindness.

Because low vision encompasses different manifestations of visual impairments as mentioned above, there is no “one size fits all” assistive technology solution for people with low vision. Nevertheless, magnification is shown to be helpful across different low vision conditions, particularly people who have low visual acuity as well as people with blurred vision[15]. Note that magnification may also be accompanied by other visual enhancement techniques such as contrast, edge or color enhancement.

People with low vision who require magnification use screen magnifiers, a special assistive technology, to interact with computing devices (e.g., ZoomText, MaGic on desktops, Zoom, Magnifier on smartphones, and many others [2]). Screen magnifiers primarily enlarge original screen content, and enable users to manually adjust the zoom or pan (i.e., move focus) over the magnified content using predefined keyboard and mouse actions.

Low-vision **s**creen-magnifier **u**sers (LSUs) find it very challenging to watch dynamically changing digital content such as videos of movies, YouTube clips, etc. While there is no time constraint to pan static content such as text and images, in so far as videos are concerned, LSUs need to quickly pan and zoom to keep up with the constantly changing video content. In particular, they have to persistently keep track of the different regions of interest (ROIs) and manually pan the magnifier viewport between these regions quickly - an arduous game-like process that the users have to go through over the entire duration of the videos. This is simply not practical for LSUs.

Accessibility of videos for LSUs remains an understudied research topic. Although well-timed audio descriptions [1] can make videos accessible for people with vision impairments, they primarily cater to people who are blind. Furthermore, creating such descriptions is labor intensive, requiring significant manual effort. Extant research on screen magnifiers [10, 25, 44] are again geared towards only static content.

Therefore, this paper investigates if it is possible to develop intelligent screen magnifiers that can support easy interaction with dynamic content such as videos, and if so, what are the desired interface characteristics of such assistive technologies.

One approach to alleviate the video-watching challenges faced by LSUs is to provide panning assistance. The idea would be to identify the different ROIs automatically, and move the focus of the magnifier to these ROIs on user requests, e.g., in response to the user’s mouse clicks. This is the basis of *SViM*, which embodies our approach to make videos accessible to LSUs. Towards that, SViM leverages advances in Computer Vision, namely, video saliency models [30] to identify the ROIs in a video, techniques to track them

continuously throughout the video, and continuous adjustments to the trajectory of the magnifier lens so that the ROIs can be viewed smoothly, i.e., without any jitters, while transitioning across frames in the video.

We did a pilot study with 11 LSUs to gather user requirements for SViM interface. To this end, using the aforementioned saliency model, we designed a basic SViM prototype that simply identified ROIs automatically and moved the magnifier focus to these ROIs. Users could quickly switch the focus of the magnifier lens between these ROIs in full-screen magnification display mode, with simple mouse clicks. The prototype did not let users manually pan over the video content. Two important takeaways emerged from the pilot study: (1) users wanted a lens magnification display mode where the magnified ROIs overlay the screen, so that they can discern a “rough” overview of the surrounding context even if they are unable to view it clearly; and (2) they desired an interface that enabled both manual and assistive panning (that is automatic positioning of the magnifier focus on the ROIs) thereby providing them with the flexibility to explore other regions close to ROIs.

Informed by the pilot study findings, we refactored SViM to accommodate these user requirements and redesigned the user interface to give users more flexibility to use the mouse to explore other regions of the video besides the ROIs identified by SViM, as well as zoom in/out of any point of interest in the video. We then re-evaluated the refactored SViM in a follow-up user study with 13 LSUs. Our results showed that salient regions displayed by SViM have a high degree of correlation with the regions that users are interested in looking at. In addition, we compared SViM to extant baseline magnifiers, namely the Windows screen magnifier and VLC player magnifier. Our results revealed that compared to these baselines, SViM significantly reduced the amount of panning done by the user while watching the videos (as much as 59%). Furthermore, LSUs found it very easy to dynamically adjust zoom levels with the interface provided by SViM for changing zoom levels. In sum, SViM has taken a step forward towards making videos accessible for people with low vision who rely on magnification for watching videos.

We summarize our contributions as follows:

- SViM - Screen magnifier interfaces that leverage advances in computer vision to enable LSUs to view videos more easily than the status quo with state-of-art extant screen magnifiers.
- Findings of a user study with 13 LSUs, including subjective and objective metrics, demonstrating the promise and potential of SViM in making videos accessible.

2 RELATED WORK

Magnification-based assistive technologies play a vital role in the daily lives of people with low vision as this impairment often manifests itself as low acuity. Magnifiers for low vision can be broadly categorized into two groups: (1) *Magnification Aids* that capture text or physical objects using cameras and magnify them and (2) *Screen Magnifiers* for interacting with computing devices.

2.1 Magnification Aids

Examples of such aids include dedicated desktop magnifiers [28, 58], optical devices for low vision [18], and head-mounted magnifiers with camera attachments [17, 29]. Here we provide a breakdown of representative aids, similar to the one provided by [48]:

Dedicated desktop magnifiers:[19]—Examples of these aids, also known as CCTV include Zoomax [58] and i-See [28]. These aids take as input an image via a camera and display it on a large screen[19, 48]. Besides providing large screens for viewing comfort, these systems are sometimes integrated with OCR for interacting with text content. We note that although these magnifiers are sometimes referred to as “Video Magnifiers”, they only handle videos taken from static objects and do not deal with videos such as movie clips. These dedicated desktop magnifiers are also not portable.

Optical devices for low vision:[20]—These set of devices have both mobile and mounted versions. An example of such device is the smartphone itself, where its camera is used to capture a scene and the magnified view is shown on a smartphone or tablet screen[33]. Another example is explore-5 [18], which acts as a magnifying glass. Note that the user has to manually move these devices for panning across the scene.

Head mounted magnifiers: These tools use a camera to capture the physical world and magnify it on wearable glasses such as virtual and augmented reality glasses. In these head mounted magnifiers, the movement of the camera and the changes in the scene are intrinsically limited in scope as opposed to videos that as a rule are characterized by rapidly changing content from frame to frame. Moreover, in the current prototypes, panning is either manual [16, 17] or is limited to magnifying a few objects selected apriori [47, 48, 56, 57].

2.2 Screen magnifiers

These are special-purpose assistive technology applications that enable people with low vision to interact with computing devices. They primarily enlarge the original screen content of these computing devices; users can manually adjust the zoom level or pan (i.e., move the focus) over the magnified screen content using predefined keyboard and mouse actions. Notable examples include Magnifier in Windows [38] and Zoom in MacOS [4] for desktops, Magnifier in Android [21, 22] and Zoom in Apple devices [3, 5]. In addition, users also have access to several third-party screen magnifiers for desktops (e.g., ZoomText[59], and MAGic [43]) and many others [2].

Screen magnifiers have been used in assorted applications for low vision. The effect of screen magnification interfaces in terms of user experience is explored in [24]. Magnification of web applications has been a topic of research in [52]. Recent work on web screen magnification appears in [9, 10]. Another application of screen magnification is [42], which magnifies the screen of a smartphone on a head mounted display where panning is controlled by users head movements. Other orthogonal works include Navisio that show how to efficiently magnify PDF documents [8], SeeingVR [55] aids people with low vision to interact with digitally-rendered virtual reality environments and CamBlend [39] is a video collaboration tool that lets users magnify manually selected regions. In addition, video

players such as VLC media player's [37] zoom as well as video editing tools such as Video Studio [49] allow user to magnify videos while playing. However, these magnifiers are not designed for LSUs; users have to specifically locate the pan/zoom controls and use these controls manually to do these operations.

The focus of all the aforementioned screen magnifier applications is either on static content or they require the user to adjust the magnifier manually. In contrast, our SViM system automatically identifies regions of interest in videos and pushes these regions to the user who can switch from one region to another via mouse clicks. SViM works across a wide range of scenarios and works with videos with various profiles (e.g., angle, context, lighting).

2.3 Video saliency

An important idea that SViM uses to identify ROIs is video saliency, a topic that has been extensively studied in computer vision over the past few years [11]. Visual saliency is defined based on users' attention that is determined from human gaze data on the scene. Of late, deep learning models are being trained to predict saliency in images and videos. For instance, [27] and [40] predict saliency of images using deep CNNs that are trained on human fixation maps. Other examples are video saliency models that leverage temporal information to predict saliency [6, 13]. This sequential modeling of video frames led to the use of long short-term memory(LSTM) [26] models. In [7], the authors proposed a combination of 3D convolutional network with LSTM to create a gaussian mixture model for predicting the saliency map. In [23], the authors use a gaze-following model and attention bounce model combined with LSTM cells to predict the saliency. In [30], the authors propose DeepVS, an object-to-motion CNN that contains two subnets, one for modeling objectness and another for modeling motion respectively; the combined outputs of the two subnets is then provided as input to an LSTM model. Our SViM system uses DeepVS to identify salient regions in videos, although this component could be replaced with other contemporary models [32, 34, 53] without affecting the overall functionality of the system.

2.3.1 Video saliency applications.—

Many application require identification of salient regions in videos and images. Thus, a number of applications based on video saliency models have emerged. An early work on video saliency is MixT [14], which uses heuristic-based methods to select and magnify salient UI elements in a tutorial video specifically recorded from the application's screen. More recent applications of video saliency which are inspired by Pano2Vid [51] include [35, 50], where important regions in 360 degree videos are projected onto 2D screens using video saliency networks.

Video saliency has also been used in virtual reality (VR) applications. The authors of [45] study human gaze for VR environments and adapt a saliency prediction model for automatic alignment of VR video cuts, panorama thumbnails, panorama video synopsis, and saliency-based compression.

SViM uses the DeepVS video saliency model [30] for a novel application, namely making videos accessible using screen magnifiers, a problem that had hitherto remained unexplored.

3 SVIM DESIGN AND IMPLEMENTATION

We designed SViM iteratively through a user-centered process. An overview of SViM architecture is shown in Figure 1. It is composed of 3 major components: the *Video Saliency Model* that highlights salient aspects of the video, the *Assistive Panning Model* and the *Magnifier Interface*. The Assistive Panning Model has two sub-components, namely, *Region-of-Interest Detector* that transforms the salient aspects of the video into regions-of-interest (ROIs) and *Region-of-Interest Tracker* that tracks and smooths the trajectory of the ROIs across successive frames.

Operationally, the magnification of videos in SViM is the result of a four-stage process. In the 1^{st} stage the saliency maps are computed from individual frames of the input video using the video saliency model (the grey "blobs" in the Saliency Heatmap box in Figure 2). In the 2^{nd} stage these saliency maps are post-processed to detect ROIs in these maps (the regions enclosed by the red and yellow squares in the ROIs box in Figure 2). To ensure smoother transition of the trajectories of the ROIs across the frames, a tracking procedure is utilized in the 3^{rd} stage (the dashed green and blue trajectories of the ROIs enclosed by green and blue boxes respectively in the Smoothed ROIs box in Figure 2). Once the detection and the tracking steps are completed, the result is passed to the magnification interface in the 4^{th} stage. We describe each of these components next. We point out that our implementation uses well-established methods for each component.

3.1 The Video Saliency Model

Saliency prediction involves detecting where the visual attention is targeted in a given image or a video. Predicting accurate saliency maps, i.e., a map that quantifies saliency of each and every region for each frame is central to the functionality of SViM. To this end, we make use of a state-of-the-art video saliency network, DeepVS [30], that has been trained on human gaze data. DeepVS uses an object-to-motion CNN as well as a convolutional LSTM to produce accurate and coherent predictions across time [30]. The input to DeepVS is a sequence of frames (see Figure 2). Each frame is resized to 448 × 448, which is the input size expected by the network.

Given a sequence of input images, the network produces a saliency map of size 112 × 112 for each frame. The saliency map is a grayscale image with values between 0 and 255, with higher values corresponding to more salient regions. With GPU acceleration, it is possible to obtain real-time performance, which means the saliency maps can be generated at the same rate the video is being played, i.e., 30 frames per second on a high-end GPU [30]. DeepVS was trained on LEDOV dataset [30], which contains 538 videos. These videos were labeled using eye-tracking data compiled from 32 sighted participants under ad hoc conditions. The relatively large number of subjects and the diverse nature of LEDOV dataset allows for the pre-trained network to generalize to unseen videos. We use this publicly available DeepVS network pre-trained with the above LEDOV dataset as the saliency model in SViM (see Figure 1).

3.1.1 ROI Detection.—DeepVS only predicts the saliency of individual pixels in a frame, but not the desired list of regions of interest(ROIs). To transform salient pixels to

ROIs, we first post process the saliency maps using standard image processing binarization and noise removal algorithms. For binarization, we threshold the grayscale values in the saliency map to generate a binary saliency mask (with each pixel value being either 0 or 1). A grayscale value of 30 was set as the experimentally determined threshold value for binarization. For removing noise, we perform an erosion operation on the binarized saliency mask, in which a kernel of size 5 × 5 is passed through the saliency map.

To extract ROIs from the post-processed binary saliency mask, we cluster the salient pixels using the *k*-means algorithm [36]. Since, we are unaware of the number of clusters n beforehand, our algorithm starts with $n = 1$ (i.e., one cluster or ROI) and tests if for a given zoom factor z, at least %60 of the salient pixels in the overall saliency mask is covered by the enclosing rectangle of the ROI. If not, we re-run the k-means clustering with $n = 2$, and check if the total number of salient pixels in the enclosing rectangles of the two ROIs combined is at least %60 of the saliency mask. We repeat this process in a hierarchical manner, until no more splitting occurs. Examples of ROIs are shown as the regions enclosed by the red and yellow squares in Figure 2.

3.1.2 Smooth Temporal Tracking.—

The process of acquiring the center of ROIs is prone to small jitters across frames. Therefore, it becomes necessary to track the ROIs across the frames so that: (a) their trajectories can be smoothed and (b) the predictions are robust even in case of sudden movements. To this end, we use a Kalman filter [31] to account for jitter due to the movement across frames. Furthermore, to keep track of the same ROI, each new centroid is matched to its closest centroid in the previous frame. We acknowledge that the Kalman filter primarily serves as a smoothing filter rather than robust multi-object tracker. In practice, the tracking needs to run in real time which made Kalman filter a good choice. In addition, we observed that Kalman filter is sufficient for tracking videos as long as the observations of ROIs are detected accurately by the detection algorithm over time.

Controlled Movement of Magnifier Viewport.: Although Kalman filtering can ensume smoother auto-panning of magnifier viewport, continuous viewport adjustments are not desirable, especially in case of minor displacements of ROIs between frames, possibly due to minor visual changes such as shadows, color, etc. Therefore, SViM employs an experimentally determined threshold to govern the auto-movement of magnifier viewport. Specifically, the viewport is not adjusted unless the displacement of an ROI between frames exceeds this threshold. If it does, SViM then gradually moves (with a predefined constant speed) the viewport to the new location of that ROI, while ignoring the continuously-generated saliency information during this movement.

3.2 Basic SViM Prototype

The saliency and the assistive panning models together constitute the core of the SViM system. We can implement different magnification interfaces on top of this core. Different interface implementations give rise to different configurations of SViM. The basic SViM configuration, (abbreviated $SViM_B$) is an assistive panning video magnifier that provides users with a magnifier interface that displays ROIs, one at a time, in full magnification mode, while allowing for fixating at other ROIs without much effort. The interface also

provides functionality to change zoom levels without much effort. We use the mouse for implementing both functionality - users switch between different ROIs using mouse clicks, and use the mouse wheel to change zoom levels. It allows users to comfortably follow the videos where multiple ROIs are present.

The workflow in the $SViM_B$ magnifier interface scenario is as follows: The user starts watching a video with this magnifier. The magnifier by default focuses on one of the ROIs. In the case of multiple ROIs, the user can use mouse clicks to check existence of other ROIs and switch to those, with each click resulting in focusing on the next ROI, traversing all the ROIs in a round-robin manner. In this scenario, the only way the user can trigger the movement of the viewport is to click to switch between different ROIs.

Display Customization: The SViM interface provided visual enhancements that are known to be beneficial for LSU's viewing experience, namely sharpening, contrast adjustment, brightness adjustment and inversion, similar to the ones offered by ZoomText [59]. Sharpening and inversion features are binary (i.e., on vs off), while brightness and contrast adjustment features can be adjusted with increments. In our user studies, we give users the ability to customize the magnifier as desired.

4 PILOT STUDY

We conducted an IRB-approved pilot study to evaluate $SViM_B$. In this pilot study we gathered data on LSUs experience with $SViM_B$ for watching videos. The feedback from the study shaped the design requirements for accessible video magnifiers.

Participants: We recruited low vision participants through a mailing list. During our phone screening, we asked participants if they watched videos visually and we excluded those whose low vision condition prevented them from visual video viewing (e.g., people with low vision who resorted to audio descriptions for watching videos). Our phone screening resulted in the recruitment of 11 participants (6 males, 5 females) for the study. All participants were low vision users, with visual acuity ranging between 20 100 and 20/2000. The average age was 51.6 (median = 47, range = 29–72). Each participant was compensated with an hourly rate of $25. Table 1 presents the participant information.

Apparatus: We used a 15.6" laptop running Windows 10 operating system with a screen resolution of 1920 × 1080, and a wireless mouse with a wheel.

Design: We compare $SViM_B$ with the available magnifier choices LSUs have for video magnification. To this end, we selected (a) WSM_{BL} which is a standard screen magnifier in windows operating systems and (b) VLC_{BL}which is a magnifier interface in the VLC video player (see Figure 3). Note that WSM_{BL} is designed for magnification of static content and as described in the related work, $VLC_{BL}$is not an assistive magnifier for LSUs.

Under this setting, we asked participants to watch a total of 12 videos that varied in the number of ROIs and the degree of movement. Both the order of study conditions (i.e.,

baseline and $SViM_B$ magnifiers) and the tasks (i.e., videos) were counterbalanced in order to minimize the learning effect.

Procedure: We began each session with an informal interview in which the participants were asked about demographic information and their video watching habits (e.g., what technology they used, how often they watched videos, etc.). Following this preamble, we studied three magnification interfaces in randomized order. For each magnifier, the participants were given as much time as they needed to learn how to use each magnifier over a practice video that could be repeatedly run. This was followed by watching four study videos with a think aloud protocol. At the end of each session each participant was asked to answer a questionnaire. This was followed by an open-ended discussion. Each session lasted two hours.

4.1 Pilot Study Evaluation

Our semi-structured interview revealed that the study participants resorted to a number of approaches to watch videos. 8 out of 11 participants mentioned that they hold their phones very close to their eyes or sit close to the TV to watch videos. Three mentioned that they pause the video every few seconds and pan the magnifier to view important content of the video. Below, we summarize some of our other findings based on participants' feedback.

4.1.1 Assistive Panning: Assistive panning (i.e., automatic viewport adjustment to focus on ROIs) was well received by the participants both for its accuracy and for its low-effort usage. All the participants mentioned that they found assistive panning useful in following the videos, compared to the baseline magnifiers. For instance P1 mentioned: "it (assistive panning) minimizes your movements and you focused on the video". Participants specially liked assistive panning on one of the videos showcasing a fast moving subject; for instance, on the exit interview P6 mentioned: "tracking the object across the screen like this is perfect specially because in these cases it is hard to keep up with the object by adjusting the video."

4.1.2 Use of mouse wheel: All the participants mentioned that they liked the use of the mouse wheel for adjusting the zoom factor, specially compared to the baseline magnifiers where they have to operate a separate visual interface for doing zoom-in/zoom-out. Participants also provided suggestions for improving the interface, e.g., 3 participants mentioned that they would like to point to different locations as the desired center of zooming.

4.1.3 Video Context: Although $SViM_B$ magnifier provides an easy-to-follow magnification of the salient objects, users expressed a desire for a flexible interface that would enable them to explore other regions of the video. Participants expressed that exploring other regions, especially the regions around the ROIs gives them contextual information. Our observations revealed that participants would zoom out to see the context although that reduced the visibility of the elements. Thus, a main takeaway from the pilot study was the importance of users being able to explore the surrounding context so that they can get a "rough" overview of the surroundings even if they are unable to view it clearly.

4.1.4 Mouse Cursor: Users expressed a need for larger/more visible cursors, since they were experiencing difficulties in following the position of the cursor in some cases. Although keeping track of the mouse position is not vital for $SViM_B$, it may introduce difficulties in case users want to use mouse cursor.

4.2 Refactoring SViM Design

Informed by the pilot study findings, we refactored SViM as follows. We developed two interface implementations, $SViM_L$, a lens-mode magnifier and $SViM_M$, a mixed-mode magnifier. $SViM_L$ selectively magnifies the ROIs while keeping the background unmagnified. $SViM_M$ is similar to $SViM_B$, but additionally allows users to pan over the video using the mouse. Both manual and assistive panning coexist in $SViM_M$, thereby giving users the flexibility to explore other regions in the proximity of the ROIs.

Figure 4 illustrates the design space for SViM magnifier interface. The three magnifier interfaces of SViM can be readily constructed from this design space. $SViM_B$ is realized by combining Assistive pan, Switch ROI and Full screen; $SViM_L$ is realized by combining Assistive pan, Switch ROI and Lensed; and $SViM_M$ is a combination of Assistive pan, Switch ROI, and Full screen. We also incorporated two additional ways of visualizing the mouse cursor: (i) a high contrast large circular cursor, and (ii) a crosshairs cursor. This is similar to the functionality provided by [59].

5 EVALUATION

Participants.—We conducted an IRB-approved user study with 13 low-vision screen-magnifier users (average age = 49.38, median = 47, Range=29–71). Gender representation was almost equal (6 males, 7 females), and none of the participants had any motor impairments that affected their interactions with SViM. Table 1 presents the participant demographics. Note that 11 out of 13 participants were also part of the pilot study discussed earlier.

Apparatus.—As in pilot study, we used a 15.6” laptop with a screen resolution of 1920 × 1080, and a wireless mouse with a wheel.

Design.—We conducted a within-subject experiment, in which every participant was asked to watch videos over five conditions (magnifiers) namely WSM_{BL} and VLC_{BL} as baseline, and three SViM interfaces: $SViM_B$, $SViM_L$ and $SViM_M$.

Study Videos: Several considerations went into the selection of videos for the user study. Firstly, to minimize the learning effect, all the videos in the user study were unique. We also defined categories of videos, and asked each participant to watch videos of different categories with each magnifier. Specifically, we considered number of ROIs and how these ROIs move on the screen to categorize videos. Each ROI that appears in a video is either static (that is, it does not move on the screen) or the ROI is dynamic (that is, the ROI moves on the screen during the video). Specifically, we focused on the two contrasting cases of static ROIs versus dynamic ROIs that continuously move during the length of the video. We also considered the number of ROIs present in the videos and focused on the two contrasting

cases of one ROI versus multiple ROIs. For the case of multiple ROIs, we ensured that the number of ROIs remained the same across experiments. While selecting videos we noticed that videos with 2+ objects vary a lot in complexity. To limit the confounding in the experiment, we chose videos with 2 ROIs. Specifically, we considered videos with one ROI as our baseline versus videos with two ROIs. For this study, we limited the maximum number of ROIs detected to $n = 2$ in our ROI detection algorithm to avoid undesired inaccuracies. Finally, we considered videos with (a) one static ROI, (b) two static ROIs, (c) one dynamic ROI, and (d) two dynamic ROIs. We collected a dataset comprising of 20 videos (five videos in each category). In case of dynamic ROIs we selected videos where the ROIs' movement had similar speed. These videos covered a variety of context from interviews to discussion panels, group dancing and basketball tutorials. None of these videos were part of the LEDOV[30] dataset.

In the user study we asked each participant to watch four videos (one from each category) with each magnifier. To minimize the learning effect, we counterbalanced the ordering of study conditions (i.e, magnifiers) and the tasks(i.e, videos) [12]. We also counterbalanced the order of videos shown to users.

Procedure.—The experimenter began the study with a semi-structured interview where the participants were requested to provide demographic information as well as discuss their user experience while watching videos.

Configuration and Training: Prior to doing the tasks, the participants were allowed to adjust the sharpness, contrast, brightness and inversion of color as desired with the help of a practice video (different from 20 task videos). Moreover, the participants could also try out and select different types of cursor (e.g., high contrast large cursor or cross-hairs cursor). All user customizations during practice were retained while they did the actual tasks.

After customization, the participants started watching task videos with different magnifiers. Note that there was a practice session prior to using each magnifier, so as to let the participants get comfortable using that magnifier.

Post study interview: At the end of the study, the experimenter engaged in an open-ended discussion to elicit subjective feedback. Each session lasted for two hours.

Data logging: In all sessions, we logged the following data: (a) viewport coordinates(i.e., what part of the screen is being viewed), (b) zoom factor, (c) mouse cursor position for each video frame, (d) button press, (e) mouse clicks, and (f) magnifier customizations.

Saliency Model Accuracy For Study Videos.: During the user study, SViM used DeepVS [30] to identify the ROIs. Since the network accuracy is vital in magnifying what's interesting to the user, we decided to evaluate the performance of the saliency model on the chosen study videos. To do so, we collected ground-truth gaze data from 7 sighted users. Specifically, we asked 7 sighted volunteers to watch the same study videos on a computer screen while wearing commercial gaze tracking glasses [46]. This procedure resulted in (x,y) coordinates for each instant with a sampling rate of 24Hz. The frames in which the

user gaze goes out of the video boundaries were padded with gaze information of the previous frame.

After gathering ground-truth data with the aforementioned procedure, we compared this data with the saliency model's predicted output, using the Pearson correlation metric. In order to convert gaze data into saliency maps for comparison, we followed the procedure proposed by Jiang et al [30], which includes placing a Gaussian mask with a fixed standard deviation, centered at the gaze coordinates. We computed this saliency correlation for each frame of each video used in the study. We found out that the average correlation between the network-predicted saliency maps and ground-truth saliency maps was 0.428. This degree of correlation indicates a high conformance between the network predictions and the ground-truth, given the fact that the fixed size of the Gaussian used to create ground-truth saliency maps can introduce inaccuracies with varying size of salient regions. We also computed Normalized Scan-path Saliency [41] between the ground-truth and predicted saliency maps as 1.584, which is comparable to some of the benchmark dataset results reported in [30].

5.1 Results

In our study panning can either be initiated by the user (manual) or by SViM (assistive). The baseline magnifiers (WSM_{BL} and VLC_{BL}) only support manual panning while the SViM had $SViM_B$ and $SViM_L$ magnifier interfaces which only support assistive panning, and finally $SViM_M$ which is a combination of both panning options. In either case, to measure the amount of panning, we computed magnifier-viewport displacement per frame which is the Euclidean distance between the centers of the viewports of two consecutive frames.

5.1.1 Panning Effort.—To focus on the manual panning, we excluded the $SViM_B$ and $SViM_L$ magnifiers since they do not include manual panning. Panning statistics for the videos that participants watched using the remaining 3 magnifiers are shown in Figure 5. A Kruskal-Wallis test between these 3 magnifiers showed that there was a significant difference in the amount of panning between the 3 magnifiers ($p = .0002$, $H = 25.24$). Pairwise two-tailed Mann-Whitney U test showed that the average user-panning amount with $SViM_M$ magnifier is significantly lesser than that with the WSM_{BL} ($p = 0.0001456, U = 1104.5, U = 351.5$) and VLC_{BL} ($p < 0.001, U = 1867, U = 577$) magnifiers. However, we did not observe a significant difference between the baseline WSM_{BL} and VLC_{BL} magnifiers. ($p = 0.53, U = 715.5, U = 600.5$). These results suggest that the proposed $SViM_M$ magnifier can significantly reduce the amount of manual panning effort for the user while watching videos.

5.1.2 Assistive Panning.—Figure 5 shows the average panning assistance for the 3 SViM magnifiers. Note that assistive panning is not available in baseline magnifiers, and hence we do not consider them in our analysis of assistive panning.

Although a Kruskal-Wallis test did not reveal any statistical significance in the auto-panning amount between the 3 magnifiers ($p = .056$, $H = 5.76$), we observed that the median value of $SViM_M$is slightly lower compared to the other two. This can be attributed to the fact that the participants occasionally take over the control by manually panning the video. Assistive panning amount is also dependent on user-induced triggers, especially when watching multi-ROI videos. Therefore, we examined the number of times the participants switched focus

between the ROIs in the study videos. We observed that there was no statistical significance in the average number of switches while using any of the 3 SViM magnifiers (Kruskal-Wallis test, $p = .95$, $H = 0.083$).

As part of our analysis regarding the zooming behavior, we computed the number of times the participants adjusted the zoom factor while watching the task videos. Average number of zoom-factor adjustments per minute are shown in Figure 6. We found that there was a significant difference in the number of zoom adjustments made with the different magnifiers (Kruskal-Wallis test, $p - value \approx 0$, $H = 94.31$,).

Pairwise Mann-Whitney U test (see Table 2) showed that while the zoom adjustments using the baseline magnifiers (average zoom adjustment in WSM_{BL} magnifier was 3.68 per minute and for VLC_{BL} was 6.98 per minute) are not significantly different from each other, they were still significantly lesser than that using all 3 SViM magnifiers. Post-study interviews suggest that this was due to the participants facing difficulties interacting with the interfaces of baseline magnifiers, given their low usability (See Figure 6). On the contrary, the wheel-based zooming available in SViM magnifiers was easy to use, hence the greater number of zoom adjustments.

Among the SViM magnifiers, only the $SViM_L$ zooming count was significantly different from the other two. Since the difference between $SViM_L$ with other SViM magnifiers is in the display mode, the zoom count difference suggests that zooming behavior was affected by the display mode rather than panning interface.

We started our analysis by looking into why the variance of zoom count is higher in $SViM_L$ magnifier. We noticed that in the user study the $SViM_L$ magnifier received mixed feedback from participants. Participants who had better vision liked it especially because it let them see the context while people with limited vision, struggled to see content in the $SViM_L$ display mode. We note that the display mode in $SViM_L$ is only an overlay and is smaller than the full screen.

5.1.3 Subjective Feedback. Ease of Use.—

Participants were asked a single ease question (SEQ) following each task completion (i.e., watching each video). The question was “Overall, this task was…” and the participants selected a number between (1–7), with 1 being very difficult and 7 being very easy. The difference between the answers that the participants provided for different magnifiers was found to be statistically significant (Kruskal-Wallis test, $H = 22.81, p = 0.00014$). Pairwise comparisons using two-tailed Mann-Whitney U Test (see Table 3) showed that: (1) the two baseline magnifiers (WSM_{BL} and VLC_{BL} magnifier) are not significantly different in terms of ease of use ($p = 0.20, U = 1546, U = 1158$); and (2) there was a significant difference in the difficulty ratings between $SViM_B$ and $SViM_M$ magnifier ($p = 0.047$). The test also showed a significant difference between baseline magnifiers and the SViM magnifiers ($p < 0.05$), thereby indicating that the SViM magnifiers were significantly easier to use. The statistics SEQ scores given to the magnifiers by the participants are shown in the Figure 7. This result indicates that the participants found the $SViM_M$ and $SViM_L$ magnifiers easy to use.

Further analysis showed that watching videos with a different number of ROI with the same magnifier did not significantly affect the ease of use for that magnifier (two-tailed Mann-Whitney U test for single-ROI versus two-ROI videos of WSM_{BL}: $p = 0.92, U_1 = 332.5, U_2 = 343.5$, VLC_{BL}:$p = 0.91, U_1 = 345, U_2 = 331$, $SViM_B$: $p = 0.47, U - 1 = 301, U_2 = 375$,, $SViM_L$:$p = 0.37, U - 1 = 291, U_2 = 385$ and $SViM_M$: $p = 0.62, U_1 = 312.5, U_2 = 363.5$).

Comparison of Magnifiers.: At the end of each session, we asked each participant to rank the 5 magnifiers from best to worst(1(best) to 5(worst)). The Friedman rank sum test on these data suggests that there was significant variation in the magnifier preferences ($p \approx 0$). The post-hoc pairwise comparison using Conover p-values, further adjusted by the Benjaminyi-Hochberg FDR method, suggests that the $SViM_M$ magnifier, compared to the rest of the magnifiers, was rated significantly better($p \approx 0$). On the other hand, the WSM_{BL} magnifier was rated significantly worse than all the other magnifiers($p \approx 0$). However, the pair-wise ranking of $SViM_L$, VLC_{BL} and $SViM_B$ magnifiers were not affected by the magnifier type significantly($p > 0.05$).

5.1.4 User Feedback.—

In the interviews, participants mentioned that assistive panning in SViM relieved them of having to do any panning and hence they could simply focus on watching the video. P2 described the experience: "*you don't need to think about magnifiers anymore, you are immersed in the video watching experience.*" Participants also expressed that assistive panning helped them in easily finding important regions, e.g., P1 stated: "*it zooms in the important areas so you don't miss what's going on*".

Also, the participants very much appreciated the control given to the user in the $SViM_M$ magnifier, with 4 participants expressing that they would like to readily use the $SViM_M$ magnifier in their daily activities. Participants specifically appreciated the flexibility and sense of control given to them, for instance P5 mentioned: "*you can control things and move around, it's more broader!*".

We also noticed that given the chance to dynamically adjust the magnifier, users tend to move the magnifier so as to test the look and feel of the hybrid feature rather than feeling the necessity to move the magnifier per every adjustment. For instance P9 explained: "I was looking at her face but then I wanted to see if with this [magnifier] I can go and look at her hand".

Visual Distraction.: In the study, the participants mentioned that with the baseline magnifiers, they have to look at a specific part of the screen in order to control the panning and zooming (in WSM_{BL} zooming-in/zooming-out and in VLC_{BL} for both zooming and panning menus are visually located on the screen). They stated that this distracts them from watching the video, and therefore results in missing important content while they are trying to steer the magnifier. Moreover, they stated that because the size of those menus are not adjustable per user they find it difficult to see. P10 mentioned: "*can't see it, it is too visual for me*", P12 compared controlling the magnifiers via the menus as playing a game and called it "*tedious and distracting*". In contrast, the SViM interface supports a 'distraction-free' magnifier interaction while watching videos.

Multi-ROI switching.: While state-of-the-art magnifiers can help the user capture the general semantics of the content in videos (e.g., “there are two people playing in this video”), we observed that with SViM, they could capture the fine-grained time-sensitive semantics of the video, thanks largely to the focus-switching feature that lets them switch between ROIs quickly (such as when the first person asked a question, the other person laughed). This feature was well received by the participants. For instance P6 explained why she liked the SViM magnifiers:”*…to find the other one you just click, it is so easy!*”.

6 DISCUSSION

Making videos accessible for people with low vision has remained largely unexplored. This paper has taken a first step in this direction with the SViM system. User studies of SViM with the target population, i.e., low-vision people, show its promise in fulfilling a long-standing accessibility need in the context of video magnification. The design and implementation of SViM opens up a number of interesting questions. Addressing them will pave the way for ushering in many-fold improvements in accessible video technology. We touch upon a few of them here.

Generalization in the Saliency model: As mentioned earlier, the DeepVS model was trained on a particular dataset. Just like any other human-labeled dataset, the bias due to samples in the dataset and the subjects who participated in collecting labels are limiting factors for the trained model’s generalization power. SViM_{M} magnifier interface reduces the impact of inaccurate predictions by allowing assistive and user panning to co-exist. Nevertheless, larger and more diverse video saliency datasets will be needed to improve the generalization power of the saliency model.

Robust Tracking: Since our goal was to explore the feasibility of accessible video magnifier that users can interact with in real time, the ROI detection and tracking mechanisms in SViM were kept rather simple and thus have some limitations. Firstly, small-sized ROIs can give rise to false negatives since the noise removal step removes small salient regions in the saliency maps. Secondly, the movement of ROIs across consecutive frames is assumed to be minor thus can be easily tracked with our algorithms. Third, ROI detection and tracking are less accurate in videos that have a large or changing number of ROIs across the frames.

Relaxing these assumptions will expand SViM’s reach to many more arbitrary videos, but will require more complex detection and tracking methods and compute power.

Incorporating Audio: SViM system does not differ in existing magnifiers in the way it utilizes or handles audio. In our user studies the effect of audio was not analyzed, and all videos were played on mute. However, one common feedback from low-vision users was the importance of audio in following the videos. We will investigate the possibility of incorporating audio as a second modality for deciding which part of the video to magnify. For example, if current audio of the video is classified as speech, this would create further motivation to magnify the face(s).

Portability: We noted that it is possible to run DeepVS in real time on high-end GPUs. Computing saliency maps on stock CPUs will go a long way towards porting SViM on to other platforms, in particular mobile devices. Towards that, one idea is to reduce the saliency map inference rate and use deep networks with lower inference time but at the expense of a drop in accuracy in predicting and tracking salient regions. A principled understanding of this trade off will inform the implementation of SViM on other platforms such as mobiles and wearables.

7 CONCLUSION

People with low-vision who need magnification have a difficult time watching videos with screen magnifiers. It essentially stems from having to zoom and pan a video continually, which can become a tiring experience for people with low-vision. SViM offers a solution to alleviate these difficulties and enhance the user experience. In the age of the ubiquitous online video platforms such as YouTube, a system like SViM can dramatically and beneficially impact the lives of people with low vision. Successfully addressing the research questions raised in the Discussion section will go a long way towards establishing SViM as the “go to” accessible video magnifier for low-vision people.

ACKNOWLEDGMENTS

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

REFERENCES

[1]. Federal Communication Act. 2019 Audio Description. https://www.fcc.gov/consumers/guides/video-description

[2]. AFB. [n.d.]. Screen Magnification Systems https://www.afb.org/node/16207/screen-magnification-systems.

[3]. Apple. [n.d.]. Apple iOS Accessibility Features. https://www.apple.com/accessibility/iphone/vision/.

[4]. Apple. [n.d.]. Vision Accessibility - Mac - Apple. https://www.apple.com/accessibility/mac/vision/.

[5]. Apple. 2019 Use Magnifier with your iPhone or iPad. https://support.apple.com/en-us/HT209517.

[6]. Bak Cagdas, Kocak Aysun, Erdem Erkut, and Erdem Aykut. 2018 Spatio-temporal saliency networks for dynamic saliency prediction. IEEE Transactions on Multimedia 20, 7 (2018), 1688–1698.

[7]. Bazzani Loris, Larochelle Hugo, and Torresani Lorenzo. 2016 Recurrent mixture density network for spatiotemporal visual attention. arXiv preprint arXiv:1603.08199 (2016).

[8]. Bernard Jean-Baptiste, Tlapale Emilien, Faure Geraldine, Castet Eric, and Kornprobst Pierre. 2008. Navisio: Towards an integrated reading aid system for low vision patients.

[9]. Jeffrey P Bigham. 2014 Making the web easier to see with opportunistic accessibility improvement. InProceedings of the 27th annual ACM symposium on User interface software and technology. ACM, 117–122.

[10]. Syed Masum Billah Vikas Ashok, Donald E Porter, and Ramakrishnan IV. 2018. SteeringWheel: A Locality-Preserving Magnification Interface for Low Vision Web Browsing. InProceedings of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, 20.

[11]. Borji Ali. 2018 Saliency prediction in the deep learning era: An empirical investigation. arXiv preprint arXiv:1810.03716 (2018).

[12]. James V Bradley. 1958 Complete counterbalancing of immediate sequential effects in a Latin square design. J. Amer. Statist. Assoc 53, 282 (1958), 525–528.

[13]. Chaabouni Souad, Benois-Pineau Jenny, and Chokri Ben Amar. 2016 Transfer learning with deep networks for saliency prediction in natural video. In 2016 IEEE International Conference on Image Processing (ICIP). IEEE, 1604–1608.

[14]. Chi Pei-Yu, Ahn Sally, Ren Amanda, Dontcheva Mira, Li Wilmot, and Hartmann Björn. 2012 MixT: automatic generation of step-by-step mixed media tutorials. In Proceedings of the 25th annual ACM symposium on User interface software and technology. ACM, 93–102.

[15]. Christen Michael and Abegg Mathias. 2017 The effect of magnification and contrast on reading performance in different types of simulated low vision. Journal of Eye Movement Research JEMR 10, 2 (2017).

[16]. Ashley D Deemer, Bonnielin K Swenor, Kyoko Fujiwara, James T Deremeik, Nicole C Ross, Danielle M Natale, Chris K Bradley, Frank S Werblin, and Robert W Massof. 2019 Preliminary Evaluation of Two Digital Image Processing Strategies for Head-Mounted Magnification for Low Vision Patients. Translational vision science & technology 8, 1 (2019), 23–23.

[17]. eSight. [n.d.]. Electronic Glasses for Blind People | From 20/200 Vision to 20/20! | eSight. https://www.esighteyewear.com.

[18]. explore 5. [n.d.]. https://store.humanware.com/hus/explore-5-handheld-electronic-magnifier.html.

[19]. American Foundation for the Blind. [n.d.]. CCTVs/VideoMagnifiers. https://www.afb.org/blindness-and-low-vision/using-technology/assistive-technology-products/video-magnifiers.

[20]. American Foundation for the Blind. [n.d.]. Low Vision Optical Devices. https://www.afb.org/node/16207/low-vision-optical-devices.

[21]. Google. [n.d.]. https://support.google.com/accessibility/android/answer/6006949?hl=en&ref_topic=9079043.

[22]. Google. [n.d.]. Google Android Accessibility Features. https://support.google.com/accessibility/android/answer/6006949.

[23]. Gorji Siavash and James J Clark. 2018 Going from image to video saliency: Augmenting image salience with dynamic attentional push. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 7501–7511.

[24]. Elyse C Hallett Wayne Dick, Jewett Tom, and Kim-Phuong L Vu. 2017. How Screen Magnification with and without Word-Wrapping Affects the User Experience of Adults with Low Vision In International Conference on Applied Human Factors and Ergonomics. Springer, 665–674.

[25]. Makoto J Hirayama. 2018 A book reading magnifier for low vision persons on smartphones and tablets. In Advanced Image Technology (IWAIT), 2018 International Workshop on. IEEE, 1–4.

[26]. Hochreiter Sepp and Schmidhuber Jürgen. 1997 Long short-term memory. Neural computation 9, 8 (1997), 1735–1780. [PubMed: 9377276]

[27]. Huang Xun, Shen Chengyao, Boix Xavier, and Zhao Qi. 2015 Salicon: Reducing the semantic gap in saliency prediction by adapting deep neural networks. In Proceedings of the IEEE International Conference on Computer Vision. 262–270.

[28]. i See. [n.d.]. https://irie-at.com/product/i-see-19/.

[29]. IrisVision. [n.d.]. IrisVision. http://www.irisvision.com/.

[30]. Jiang Lai, Xu Mai, Liu Tie, Qiao Minglang, and Wang Zulin. 2018 Deepvs: A deep learning based video saliency prediction approach. In Proceedings of the European Conference on Computer Vision (ECCV). 602–617.

[31]. Rudolph Emil Kalman. 1960 A new approach to linear filtering and prediction problems. Journal of basic Engineering 82, 1 (1960), 35–45.

[32]. Koutras Petros and Maragos Petros. 2019 SUSiNet: See, Understand and Summarize it. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 0–0.

[33]. Raja S Kushalnagar, Stephanie A Ludie, and Poorna Kushalnagar. 2011. Multi-view platform: an accessible live classroom viewing approach for low vision students. In The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility. ACM, 267–268.

[34]. Lai Qiuxia, Wang Wenguan, Sun Hanqiu, and Shen Jianbing. 2019 Video Saliency Prediction using Spatiotemporal Residual Attentive Networks. IEEE Transactions on Image Processing (2019).

[35]. Lai Wei-Sheng, Huang Yujia, Joshi Neel, Buehler Christopher, Yang Ming-Hsuan, and Kang Sing Bing 2017 Semantic-driven generation of hyperlapse from 360 degree video. IEEE transactions on visualization and computer graphics 24, 9 (2017), 2610–2621. [PubMed: 28910772]

[36]. Lloyd Stuart. 1982 Least squares quantization in PCM. IEEE transactions on information theory 28, 2 (1982), 129–137.

[37]. VideoLAN media player. [n.d.].

[38]. Microsoft. 2019 Use Magnifier to make things on the screen easier to see - Windows Help. https://support.microsoft.com/en-us/help/11542/windows-use-magnifier.

[39]. Norris James, Schnädelbach Holger, and Guoping Qiu. 2012 CamBlend: an object focused collaboration tool. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 627–636.

[40]. Pan Junting, Sayrol Elisa, Xavier Giro-i Nieto Kevin McGuinness, and Noel E O'Connor. 2016 Shallow and deep convolutional networks for saliency prediction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 598–606.

[41]. Robert J Peters Asha Iyer, Itti Laurent, and Koch Christof. 2005 Components of bottom-up gaze allocation in natural images. Vision research 45, 18 (2005), 2397–2416. [PubMed: 15935435]

[42]. Pundlik Shrinivas, Yi Huaqi, Liu Rui, Peli Eli, and Luo Gang. 2016 Magnifying smartphone screen using google glass for low-vision users. IEEE Transactions on Neural Systems and Rehabilitation Engineering 25, 1 (2016), 52–61. [PubMed: 28113862]

[43]. Scientific Freedom. [n.d.]. MAGic® - Freedom Scientific https://www.freedomscientific.com/products/software/magic/.

[44]. Seakins Paul J, Jonathan D Cartwright, David J Haughey, N Lovegrove David, and Darryl J Best. 2009. Image magnifier for the visually impaired. US Patent App 11/578,486.

[45]. Sitzmann Vincent, Serrano Ana, Pavel Amy, Agrawala Maneesh, Gutierrez Diego, Masia Belen, and Wetzstein Gordon. 2018 Saliency in VR: How do people explore virtual environments? IEEE transactions on visualization and computer graphics 24, 4 (2018), 1633–1642. [PubMed: 29553930]

[46]. SMI. [n.d.]. SMI Eye Tracking Glasses 2 Wireless.

[47]. Stearns Lee, Victor DeSouza Jessica Yin, Findlater Leah, and Jon E Froehlich. 2017 Augmented reality magnification for low vision users with the microsoft hololens and a finger-worn camera. In Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility. ACM, 361–362.

[48]. Stearns Lee, Findlater Leah, and Jon E Froehlich. 2018 Design of an Augmented Reality Magnification Aid for Low Vision Users. In Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility. ACM, 28–39.

[49]. Video Studio. [n.d.]. https://www.videostudiopro.com/en/tips/basics/zoom-in-on-video/.

[50]. Su Yu-Chuan and Grauman Kristen. 2017 Making 360 video watchable in 2d: Learning videography for click free viewing. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 1368–1376.

[51]. Su Yu-Chuan, Jayaraman Dinesh, and Grauman Kristen. 2016 Pano2Vid: Automatic Cinematography for Watching 360° Videos. In Proceedings of the Asian Conference on Computer Vision (ACCV).

[52]. Mary Frances Theofanos and Janice Ginny Redish. 2005 Helping low-vision and other users with web sites that meet their needs: Is one site for all feasible? Technical communication 52, 1 (2005), 9–20.

[53]. Wang Wenguan, Shen Jianbing, Xie Jianwen, Cheng Ming-Ming, Ling Haibin, and Borji Ali. 2019 Revisiting video saliency prediction in the deep learning era. IEEE transactions on pattern analysis and machine intelligence (2019).

[54]. WHO. [n.d.] Low Vision Characterization. https://www.who.int/blindness/Change%20the%20Definition%20of%20Blindness.pdf

[55]. Zhao Yuhang, Cutrell Edward, Holz Christian, Meredith Ringel Morris Eyal Ofek, and Andrew D Wilson. 2019 SeeingVR: A Set of Tools to Make Virtual Reality More Accessible to People with Low Vision. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM, 111.

[56]. Zhao Yuhang, Szpiro Sarit, and Azenkot Shiri. 2015 Foresee: A customizable head-mounted vision enhancement system for people with low vision. In Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility. ACM, 239–249.

[57]. Zhao Yuhang, Szpiro Sarit, Knighten Jonathan, and Azenkot Shiri. 2016 CueSee: exploring visual cues for people with low vision to facilitate a visual search task. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, 73–84.

[58]. Zoomax. [n.d.]. https://www.zoomax.com/low-vision-products/easy-to-use-desktop-video-magnifier-Panda.html.

[59]. Zoomtext. [n.d.] Zoom Text Magnifier/Reader https://www.zoomtext.com/products/zoomtext-magnifierreader/.

CCS CONCEPTS

- **Human-centered computing → Human computer interaction (HCI)**; **Accessibility systems and tools**; User studies.

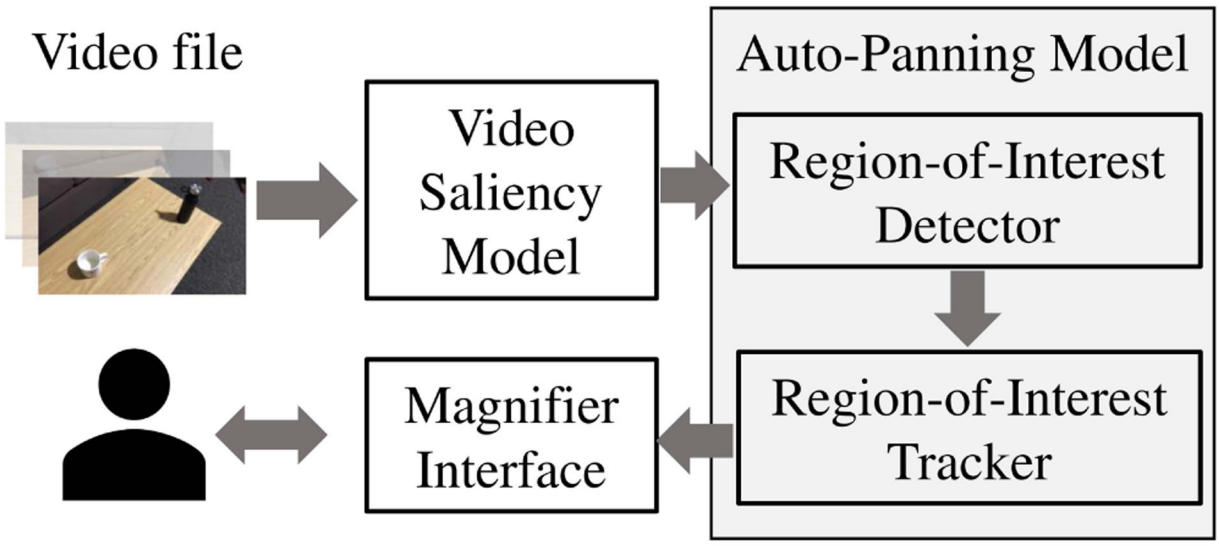

Figure 1:
SViM architectural overview.

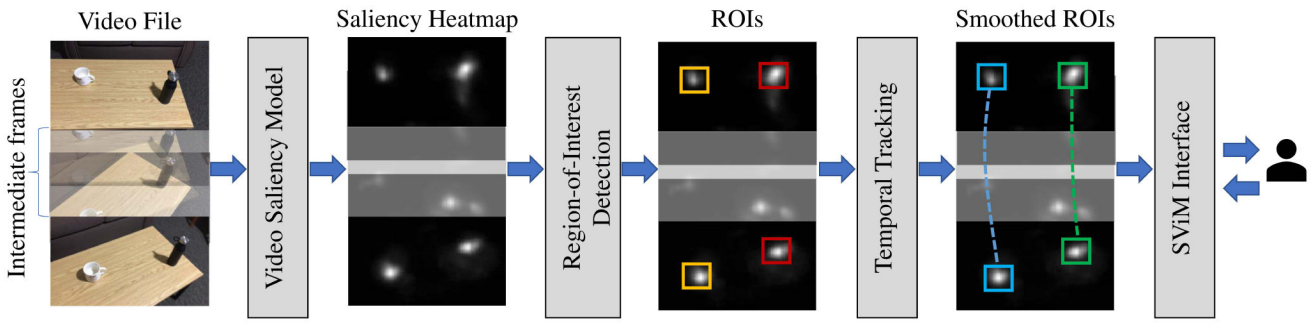

Figure 2:

The SViM video magnification pipeline. The process begins with a sequence of RGB frames. Video saliency model predicts saliency in the form of a heatmap. Next, the ROI detector produces a set of bounding boxes identifying ROIs. This is followed by temporal tracking, which smooths the trajectory of ROIs across the frames. The result of this process is then passed to the SViM interface, which displays the magnified ROIs to the user in the magnifier viewport.

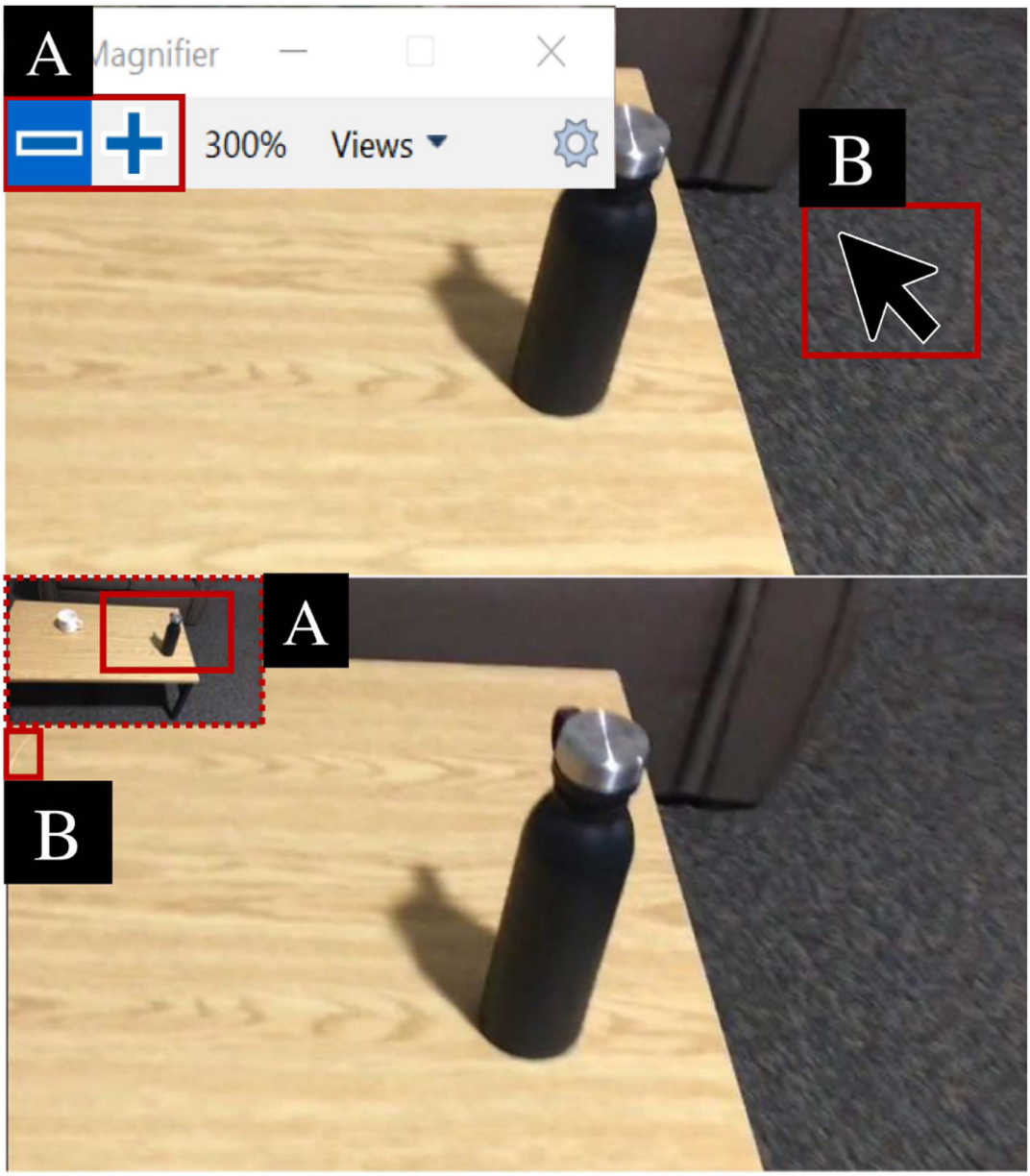

Figure 3:
The baseline magnifiers. Top: Windows screen magnifier. Mouse cursor (B) is used for user panning. + and − buttons (A) are used for adjusting the zoom factor. Bottom: VLC magnifier. A thumbnail of the frame on the upper left corner is used for user panning (A). A vertical bar (B) is used for adjusting the zoom factor.

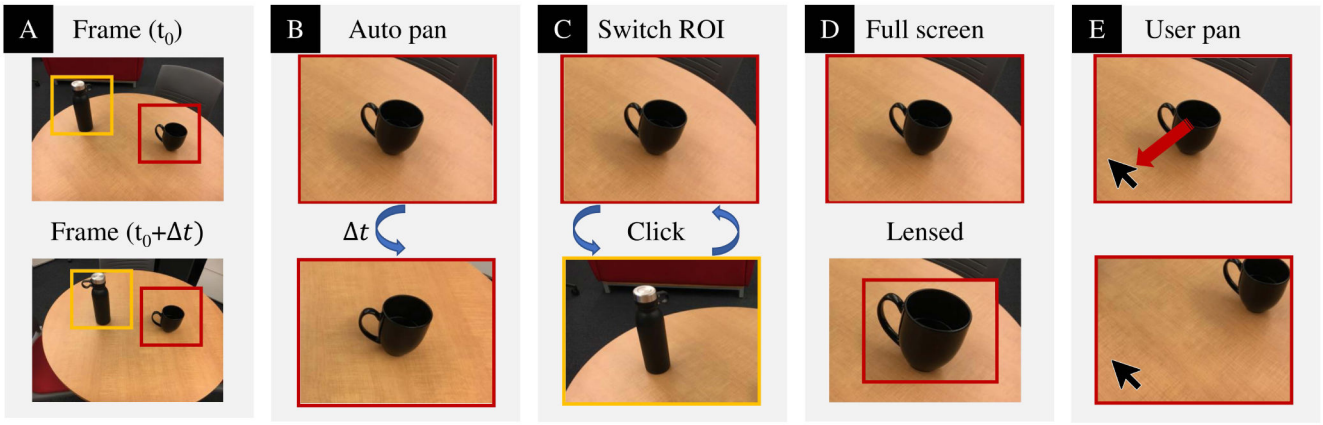

Figure 4:

Interface design space for SViM magnifier. (A) shows two frames of a video with two salient ROIs enclosed by red and yellow squares; the top and bottom frames have time stamps t_0 $t_0 + \Delta t$ respectively. (B) shows assistive panning following the ROIs through time by automatically panning the viewport to the ROIs. Specifically, observe that assistive panning has tracked the ROI enclosed in red square through time ranging from t_0 through $t_0 + \Delta t$ and panned the viewport to this ROI at $t_0 + \Delta t$. (C) shows switching between ROIs with mouse clicks - the ROI in the red square is switched to the ROI in the yellow square. (D) shows the display modes, the full screen mode (on top) populates the entire screen to show magnified content whereas in the lensed mode (bottom) the ROI is magnified (such as the one in the red square) overlaying the rest of the content unmagnified. (E) shows that the user moves the mouse to pan to any location.

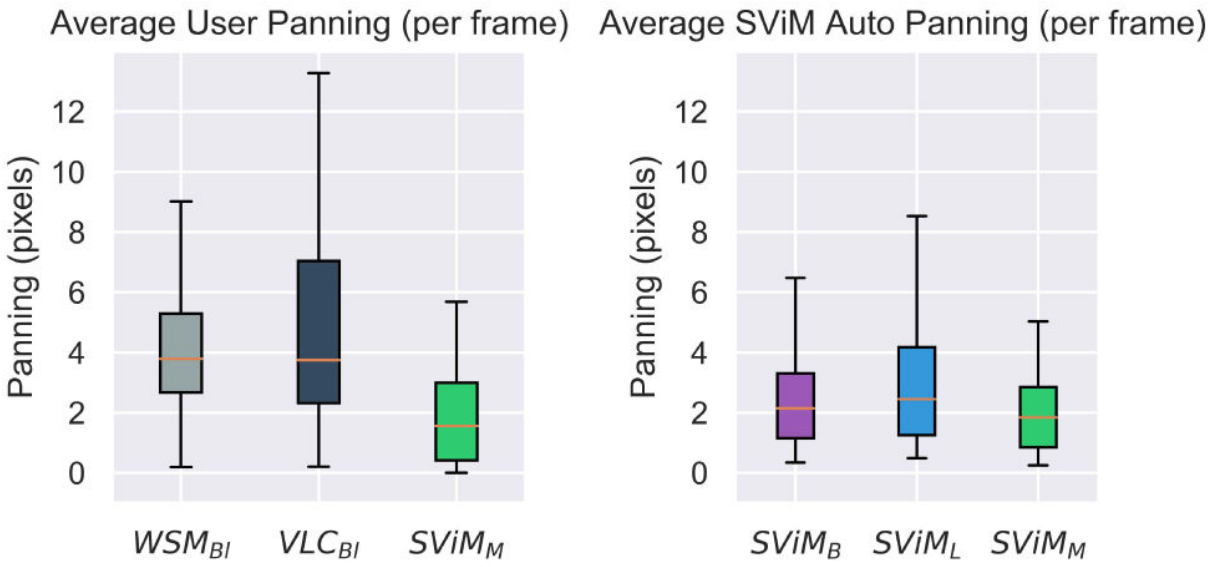

Figure 5:

Left: Average user panning per frame for each magnifier. Right: Average auto panning per frame for each magnifier.

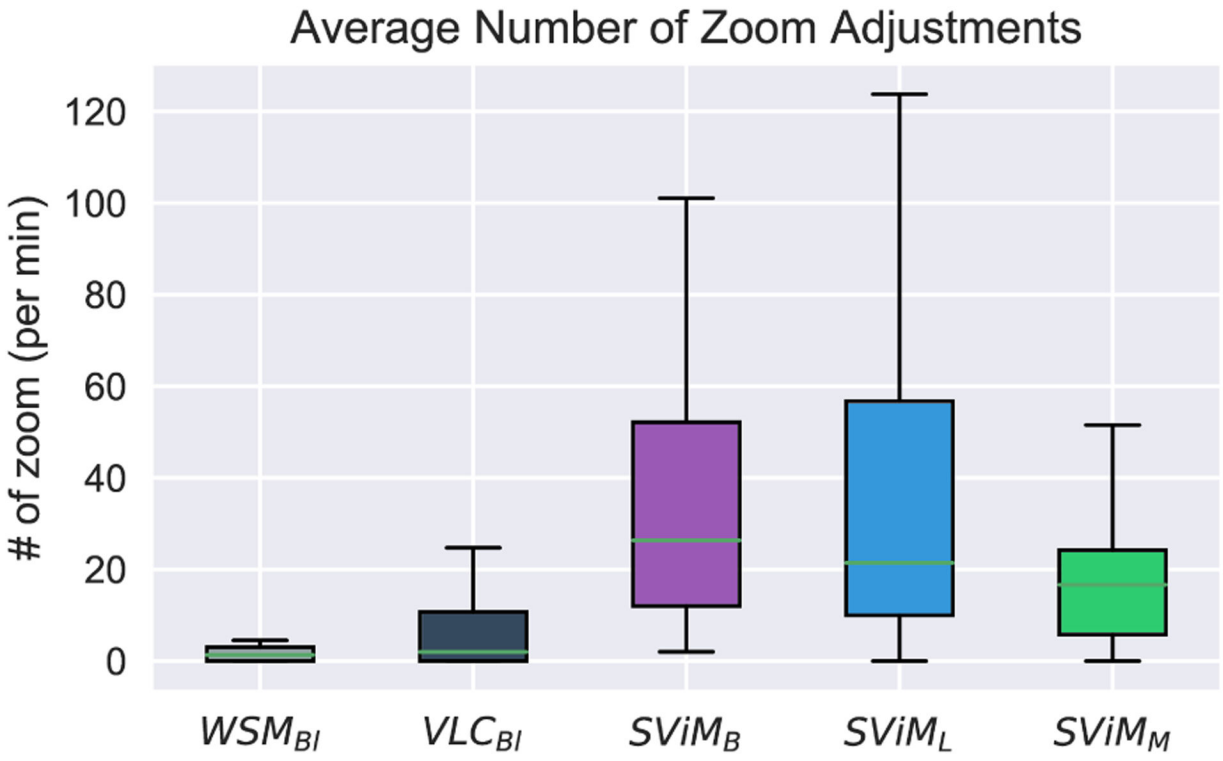

Figure 6:
Average number of zoom adjustments per minute.

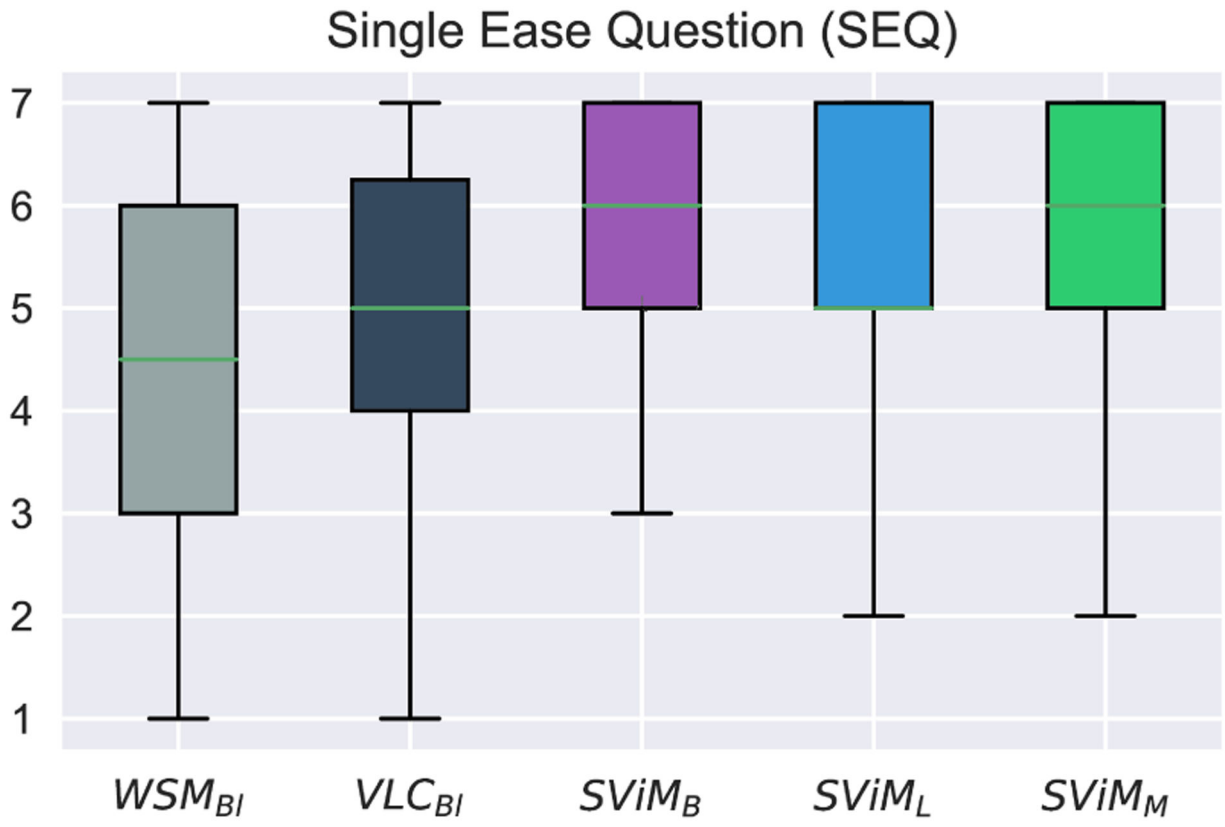

Figure 7:
Average score given in the single ease question (SEQ), with 1 being *very difficult* and 7 being *very easy*.

Table 1:

Participant demographic information.

ID	Pilot/Study	Age/Sex	Diagnosis (C: Congenital, A: Adventitious)	Visual Acuity (L: Left, R:Right)	Video Watching Habits	Screen Mag Experience
P1	Both	35/M	Optic atrophy, Retinitis pigmentosa (A)	L: 20/700; R: 0	iPhone 6s: everyday, TV(14”): occasionally	Expert
P2	Both	34/M	Albinism (C)	L: 20/200; R: 20/100	iPhone 4: everyday, TV(60”): everyday	Expert
P3	Both	71/F	Diabetic retinopathy (A)	Unknown, but good	Android (Galaxy S7): everyday, TV(40”): everyday	Intermediate
P4	Both	33/M	Optic atrophy(A)	Unknown	iPhone: everyday, TV(32”): everyday	Expert
P5	Both	45/F	Leber’s Congenital Amaurosis (C)	L: 20/200; R: 20/400	Android(HTC): everyday, TV(46”): everyday	Intermediate
P6	Study	68/F	Glaucoma (A)	L: 0; R: Unknown	Android: everyday, TV(46”): everyday	Intermediate
P7	Study	71/F	Glaucoma (A)	L: 20/200; R: 0	iPhone 8: occasionally, TV(65”): everyday	Intermediate
P8	Both	47/M	Astigmatism (C)	L: 20/200; R: 20/400	iPhone: everyday, TV(55”): everyday	Expert
P9	Study	66/F	Glaucoma (A)	L: 20/400; R: 0	Android: No , TV(55”): occasionally	Beginner
P10	Both	29/F	Glaucoma (C)	L: 0; R: 20/400	iPhone 10: occasionally, TV: No	Expert
P11	Both	62/F	Congenital retinal scar (C)	L: 20/400; R: 20/400	iPhone 6s plus: No, TV(34”): occasionally	Beginner
P12	Study	47/M	leber’s optic neuritis (A)	L: 20/200; R: Unknown	iPhone: everyday , TV(55”): everyday	Intermediate
P13	Study	34/F	Cancer (C)	L: 20/200; R: 0	Android: occasionally , TV(32”): occasionally	Intermediate
P14	Pilot	72/F	Retinitis pigmentosa	L: 20/2000; R: 20/1000	Occasionally, hardware unknown	Unknown
P15	Pilot	65/M	Cataracts(A)	20/500 (Eye unknown)	TV: everyday	Beginner
P16	Pilot	44/M	Retinitis pigmentosa (C)	Unknown	Everyday, hardware unknown	Unknown

Table 2:

Pair-wise comparisons between the different magnifiers for number of zoom adjustments.

	WSM_{BL}	**VLC_{BL}**	**$SViM_{B}$**	**$SViM_{L}$**
VLC_{BL}	p = 0.075 U1 = 1614.5 U2 = 1089.5			
$SViM_{B}$	p <0.001 U1 = 2524 U2 = 180	p <0.001 U1 = 2326.5 U2 = 377.5		
$SViM_{L}$	p <0.001 U1 = 2306.5 U2 = 397.5	p <0.001 U1 = 2147 U2 = 557	p = 0.631 U1 = 1277.5 U2 = 1426.5	
$SViM_{M}$	p <0.001 U1 = 2293 U2 = 411	p <0.001 U1 = 2057.5 U2 = 646.5	p = 0.005 U1 = 915 U2 = 1789	p = 0.069 U1 = 1072.5 U2 = 1631.5

Table 3:

Pairwise comparisons between magnifiers for SEQ questionnaire ratings.

	WSM_{BL}	**VLC_{BL}**	**$SViM_{B}$**	**$SViM_{L}$**
VLC_{BL}	p = 0.207 U1 =1546 U2 =1158			
$SViM_{B}$	p <0.001 U1 = 1892.5 U2 = 811.5	p = 0.011 U1 = 1726.5 U2 = 977.5		
$SViM_{L}$	p = 0.010 U1 = 1746.5 U2 = 957.5	p = 0.267 U1 = 1518.5 U2 = 1185.5	p = 0.093 U1 = 1103.5 U2 = 1600.5	
$SViM_{M}$	p <0.001 U1 =1956 U2 = 748	p = 0.004 U1 = 1775.5 U2 = 928.5	p = 0.792 U1 = 1390 U2 = 1314	p = 0.047 U1 = 1644.5 U2 = 1059.5