

Directing Gaze in Narrative Art

Ann McNamara*
Texas A&M University
Stephen Caffey
Texas A&M University

Thomas Booth
Rochester Institute of Technology
Cindy Grimm
Washington University in St. Louis

Srinivas Sridharan
Rochester Institute of Technology
Reynold Bailey
Rochester Institute of Technology



Figure 1: Fixation maps over an art work that contains separate panels, which, when viewed in the intended order, tell a coherent story. In this work a gaze direction technique is used to visually guide the viewer through the image. As can be seen above the distribution of fixations is very different for static viewing (A), and Gaze Directed viewing (B). The highlighted rectangular regions indicate story panels, and the numbered circles indicate viewing order. In the gaze directed case, both order viewing accuracy and fixations on relevant panels increased. The story of Joseph, by Biagio d'Antonio, reproduced with kind permission of the J. Paul Getty Museum, Los Angeles

Abstract

Narrative art tells a story, either as a moment in an ongoing story or as a sequence of events unfolding over time. In many works of art separate panels within the same frame are used to depict the sequence of events. Often, there is no clear delineation between these panels, or any indication of the optimal viewing order. To improve visual literacy we propose using Subtle Gaze Direction (SGD) to direct the viewers gaze across an image in a manner which reveals the story. SGD uses small image space modulations in the luminance channel to guide a viewer's gaze about an image without disrupting their normal visual experience. Using a simple ordering task we compared performance using no modulation and using subtle modulation with the correct order of narrative episodes as intended by the artist. Results from experiments show improved performance when SGD is employed. This experiment establishes the potential of the method as an aid to visual navigation in images where the viewing order is unclear.

CR Categories: I.3.3 [Computer Graphics]: Three-Dimensional Graphics and Realism—Display Algorithms

Keywords: eye-tracking, gaze direction, art history education

*e-mail: ann@viz.tamu.edu

Copyright © 2012 by the Association for Computing Machinery, Inc. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions Dept, ACM Inc., fax +1 (212) 869-0481 or e-mail permissions@acm.org.

SAP 2012, Los Angeles, CA, August 3 – 4, 2012.
© 2012 ACM 978-1-4503-1431-2/12/0008 \$15.00

1 Introduction

In this paper we describe the use of Subtle Gaze Direction (SGD) to correctly guide observers through pictorial episodes when viewing paintings (Figure 1). To improve visual literacy we propose the use of SGD to direct the viewer's gaze over an image in a manner which reveals the story in the narrative art. Using a simple ordering task we compared performance using no modulation and using subtle modulation, highlighting the center of the panels in the order of the story narrative. Results from experiments show improved performance when SGD is employed. This experiment establishes the potential of the method as an aid to visual navigation in images where the viewing order is unclear.

Narrative art tells a story, either as a moment in an ongoing story or as a sequence of events unfolding over time. A synoptic narrative depicts a single scene in which a character, or characters, are portrayed multiple times within a frame to convey that multiple actions are taking place. This can cause the sequence of events to be unclear within the narrative. Synoptic narratives typically provide visual cues that convey the sequence, but it still might be difficult to decipher for those unfamiliar with the story. The process is best illustrated with an example. Figure 2 (top) shows a synoptic painting titled “The Tribute Money” by Renaissance artist Masaccio. This painting describes a story from the Gospel of Matthew, in which Jesus directs Peter to go to the river and retrieve a coin from the mouth of a fish in order to pay the temple tax. The optimal way to visually navigate this piece is to begin in the center with the tax collector demanding the money. Jesus, surrounded by his disciples, instructs Peter to retrieve the money from the mouth of a fish. By moving their gaze to the *left* of the painting (perhaps counter-intuitive to western civilization who normally read left to right) viewers notice Peter executing Jesus' instruction.

The viewer's eyes next need to travel to the extreme right of the painting to view the third episode in which Peter pays the tax collector. At the time it was painted, audiences were conditioned to recognize repeated elements in a frame and identify panels, thereby intuitively understanding the intended order in which each episode of the painting was to be viewed. However, our ability, as artists and audiences, to correctly "read" these paintings may not be so accurate in present day because our visual literacy is not conditioned to follow the viewing pattern the artist intended. In the 15th century the audience would understand that there are multiple episodes in this painting, and also in which order to view these panels in order to comprehend the story. Web-based solutions do exist which manipulate a digital representation of a painting using strong outlines, or interruptive text over the image to explain where the viewer should direct their gaze (see Figure 2, [Masaccio 1421]). While these represent a promising initial approach, a more elegant solution would not disrupt interrupt the visual experience of the audience. Employing gaze direction techniques would allow the viewer to see the actual painting with areas of interest accentuated in a manner which preserves the visual experience by acknowledging the artist's intent. In this work we investigate the use of Subtle Gaze Direction (SGD) as an aid to navigate narrative art. This goal of this work is to satisfy the need to display information in a manner that minimizes disruption to the viewer, but can accurately direct gaze to certain locations of an image, *in a specific sequence*. In other words, our original SGD technique did not examine how well SGD directs gaze to multiple image locations in a specific sequence.

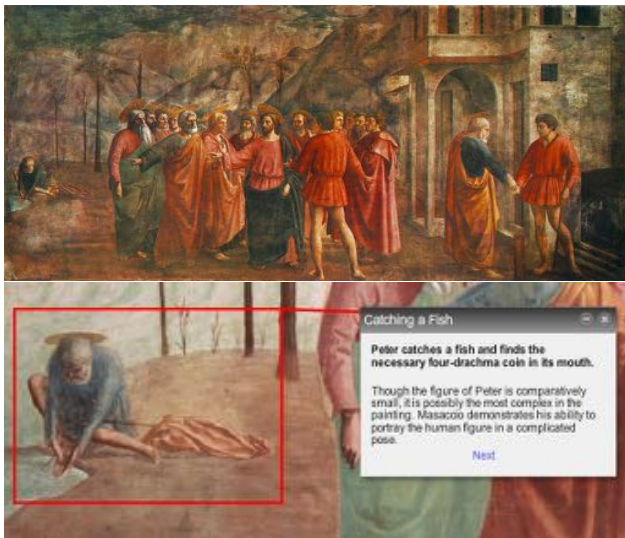


Figure 2: Above: "The Tribute Money", by Masaccio tells the story of Jesus and the tax collector. The piece should be viewed in the following order: center, left, then right. Current web-browser based educational tools use text pop-ups with interruptive rectangular outlines to highlight important information in a visual narrative. This not only distracts the viewer from appreciating the image, but also breaks up the image into smaller pieces so it is not viewed in a holistic manner. The red colored rectangle destroys the visual experience by superimposing a distracting overlay on the original painting.

The **main contributions** of this work is validation of SGD as a technique to influence where viewers look in a scene, both spatially and temporally. This technique will be especially useful in the digital humanities, as it will allow scholars to recuperate various types

of visual literacy specific to a historical moment. We focus on Art History Education as an application but it is easy to extend to any visual task in which viewing order is critical to understanding or task completion.

2 Previous Work

Subtle Gaze Direction (SGD) [Bailey et al. 2009] exploits the well established fact that human peripheral vision processes stimuli faster than foveal vision [Ogden and Miller 1966]. On initial viewing of a scene the low-acuity peripheral vision of the Human Visual System (HVS) locates regions of interest. The foveal vision, which is slower and has higher acuity, is then involuntarily directed to focus on these regions. SGD proceeds by modulating regions of an image that appear only to peripheral vision. In this manner peripheral vision is forced to locate the indicated regions of interest, which are subsequently fixated on. This causes involuntary saccades to move the eye to fixate on the modulated region as it attempts to resolve the detected stimuli. Luminance modulation works well as the HVS is highly sensitive to changes in luminance values [Spillman 1990].

Modulations are constructed by alternately blending small pixel regions with some amount of black, then some amount of white. The rate at which the blend is modulated is 10Hz. A Gaussian falloff with a radius of 32 pixels is used which (in our viewing configuration) corresponds to approximately a 2cm diameter circular screen area.

The marriage of technology and art appreciation is not new — several existing applications have successfully been applied in the Art domain [Gwilt 2009] [Damala et al. 2008] [Andolina et al. 2009] [Bruns et al. 2007] [Choudary et al. 2009] [Chou et al. 2005] [Srinivasan et al. 2009]. To date, however, few have proposed eye-tracking as an added dimension. The novelty of this approach lies in the eye-tracking and in attracting and directing the gaze to the correct region of the artwork in a sequence that will encourage appropriate visual navigation and understanding of the image and strengthen observation skills.

Obviously conspicuous objects in a scene (such as a black sheep in a white flock) will draw the viewer's attention first. However, there are more subtle image characteristics that can also draw our gaze. Image properties such as color, size and orientation can be used to control attention [Veas et al. 2011] [Underwood and Foulsham 2006] [Underwood et al. 2009]. In movies, directors use an arsenal of cinematographic "tricks" to lead the audience to look where they want them to look (see [Bordwell 2011]). Taking an automated approach, Itti and Koch [Itti and Koch 2000] [Itti and Koch 2001] developed an algorithm to measure *visual saliency* (how likely people are to look at parts of an image) on the basis of image characteristics such as intensity distribution, color changes, and orientation. Saliency maps could prove to be a good candidate to indicate the initial attention in a painting. Then, by modifying the digital version of the painting to re-distribute saliency, we could build several versions of the painting with the pre-selected interesting regions manipulated to increase saliency. For example, in "The Tribute Money" when it is time to look at Peter retrieving the coin from the mouth of the fish, SGD could boost the saliency in that region and thereby influence the viewer to re-direct their gaze.

This remainder of this paper presents a psychophysical experiment that explores the impact of SGD on performance during a viewing of narrative art works. The results show that this method works well

without introducing noticeable artifacts into images, which might degrade the viewing experience.

3 Experimental Design

The goal of this experiment is to determine to what extent SGD lends itself to aid observers in extracting the intended sequence of events from regions of an episodic image. Participants viewed a sequence of images, each of which contained three or more panels intended as episodes. The intended viewing order of these panels is not always immediately clear. Art history research provides the narrative for each art piece, from which panels are determined [Velli 2007]. Panels (in each image) were manually selected as rectangular regions which enclosed the relevant portion of the image that conveyed an episode of the story. In the non-control group, after viewing the image for a short period of time, relevant panels were highlighted using SGD at the panel center. Participants then indicated the order they perceived to be the correct viewing order by clicking on image sections outlined with boxes. We compare performance using SGD with performance under normal viewing conditions.

3.1 Stimuli

Eleven images served as stimuli for the experiment, two of which were used for observer training, see Figure 7). In each image, episodes or *panels* were identified and served as targets for SGD. The number of episodes (panels) varied from painting to painting, ranging from three to seven.

The size of panels also varied within each image (see Figure 3), initially we were concerned that fixations may be artificially increased in proportion to panel size, but this did not turn out to be the case.

Presentation order was randomized to minimize the introduction of learning effects. Images were presented for a period of time proportional to the number of episodes in the image and were displayed on a 22 inch widescreen monitor, operating at 60 Hz with a resolution of 1680 x 1050. Stimuli transitioned directly from one to the next, however, no action was taken until a finite amount of time had passed. Image size varied as shown in Figure 7. In cases where images size was smaller than the viewing screen, width and height were maximized to fit screen resolution and a black border was added. An example of a single image, with all the regions highlighted for illustration purposes, is shown in Figure 3. In this image there are six panels. The observer would not see the regions highlighted in this obvious manner. This simply illustrates the presence of the panels.

Participants were seated in front of a computer screen in a well-lit room. Using a SensoMotoric Instruments iView X Remote Eye-Tracking Device operating at 250 Hz with gaze position accuracy $< 0.5^\circ$, data pertaining to fixation position and saccades were recorded for each participant. After a brief calibration phase, each observer underwent a short tutorial session to familiarize them with the experimental procedure and user interface. Questions were encouraged during the tutorial session but no data was collected. Participants were then presented with each of the nine art works in a random order. Image complexity varied, as did the number of panels. The two groups are as follows:

- **Group 1: Normal Viewing Conditions:** No actions were applied to the images, in other words images were viewed normally with no modulations. This group served as the control group for the experiment.
- **Group 2: Subtle Modulation:** SGD was employed to highlight the target panel regions in the intended viewing order in



Figure 3: An exemplar image showing all of the panels which contribute to the narrative. Observers first viewed the image without the panels highlighted. Once a certain amount of time had elapsed, participants then clicked on the panels in the order they believed matched the order of the story being told. The modulations for the SGD group (Group 2) were placed at the center of each panel.

an effort to aid in visual navigation. Gaze manipulation was implemented as described in [Bailey et al. 2009]. Modulation was never applied to panels while they were being directly viewed. All modulations were only applied to the peripheral vision. Modulations were stopped as the observers gaze tended toward them i.e. observers never directly viewed modulations. A modulation radius of 0.04° of visual angle was used to ensure that modulations were subtle. The modulations were placed in the center of the panels.

Thirty-six participants were assigned randomly to one of the two groups. Participants were volunteers from a group of undergraduates. All had normal, or corrected-to-normal, vision and were naive to the purpose of the experiment. Viewing time for each image varied in direct proportion to the number of panels present. Each image was presented for t seconds before the user was allowed to respond. For the control group, Group 1, t was chosen to be equal to the number of regions in the image. In Group 2, t is the time taken to guide the viewers exactly once through the correct sequence of regions. Previous studies [Bailey et al. 2009], [McNamara et al. 2008] revealed that SGD modulations typically attracted gaze within 0.5 seconds. To ensure that we had comparable viewing times between both groups a 0.5 second delay was added between successive modulations.

After t seconds, the relevant regions were highlighted with rectangles and the mouse activated to allow the users to respond. Both groups of participants were instructed to click on the highlighted regions in the order they believed the story unfolds. Each participant reported an order which they believed matched the intended sequence of the story in the art work.

3.2 Analysis of Data

In addition to recording eye-movements for each participants, each participant reported an order for each image, based on their understanding of panel sequence within that image. We needed a robust mechanism to compare accuracy of performance between the two groups. Levenshtein distance [Navarro 2001] [Levenshtein 1965] [Levenshtein 1966] is a string metric, developed in the field of information theory and computer science to compute *differences between sequences*. Levenshtein distance provides an appropriate measure to compare distances between ordered sequences, such as those recorded during our experiment. To accurately compare sequences using Levenshtein distance the correct (intended) viewing order of each image is converted into a string sequence. All responses from each participant are also converted to an appropriate string sequence in order to facilitate comparison to the correct sequence. Since the number of relevant regions varies across the images we normalize the distance measure computed for each image by dividing by the number of panels. The normalized Levenshtein distance L between the correct sequence $S_{correct}$ and user sequence S_{user} is as follows:

$$L = \frac{\text{Levenshtein Distance}(S_{correct}, S_{user})}{\# \text{ of Panels}} * 100 \quad (1)$$

For example, let the correct panel order be [ABCDE] and let [ACBDE] denote the participant's response. The number of *panels* in this image is five. Using Equation 1, we obtain a normalized Levenshtein distance value of 40. A distance of 0 would indicate no difference, whereas a distance of 100 would indicate maximal distance.

4 Results and Discussion

The predicted sequence of panels reported by each participant for each image was recorded. Normalized distances for each image were compared to the actual intended sequence for that image using the distance metric expressed in Equation 1. The calculated normalized Levenshtein distance measures between conditions showed differences across the two groups with a mean distance measures of 57.32 and 34.79 for groups 1 and 2 respectively, as illustrated in Figure 4. These values were calculated by averaging the normalized Levenshtein distance (L) for all the participants in a group over all the images used in the experiment. This implies that participants from Group 2 (guided by SGD) consistently proved to be more accurate at predicting the intended sequence of panels contained in narrative art when compared to Group 1, the control group (no SGD). This measure indicates that, for example, when viewing a narrative art image containing ten panels, the static viewing group will incorrectly predict the order of approximately 5 – 6 panels, while the gaze directed group, Group 2, will return a prediction with only 3 – 4 panels out of the sequence. An independent-samples paired t – test suggests that this was a significant effect:

$$t(316) = 1.9675; p < 0.05 \quad (2)$$

Figure 5 shows the average L value for each image across all the participants in each group. This analysis gives some intuition on the influence of the number of panels over the accuracy in detecting the correct sequences in the narrative art. The images in the graph are arranged in the increasing order of number of panels. In all nine images, the average L value indicates that the participants in Group 2, the gaze directed group, predict panel order more accurately than the participants from Group 1, the static viewing group. This also shows that the gaze directed group performed better than the static

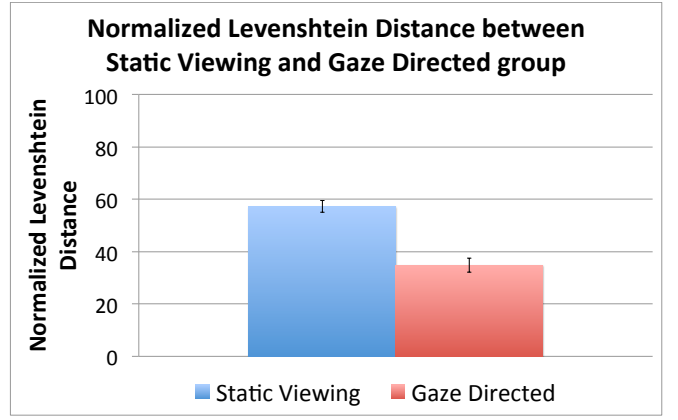


Figure 4: Normalized Levenshtein distance measure between Group 1 (static viewing group) and Group 2 (gaze directed group). Error bars represent one standard error. The graph shows that Group 2 participants, who viewed SGD images were able to predict the intended viewing order of panels more accurately than those in Group 1 that did not have the benefit of SGD as a gaze direction aid.

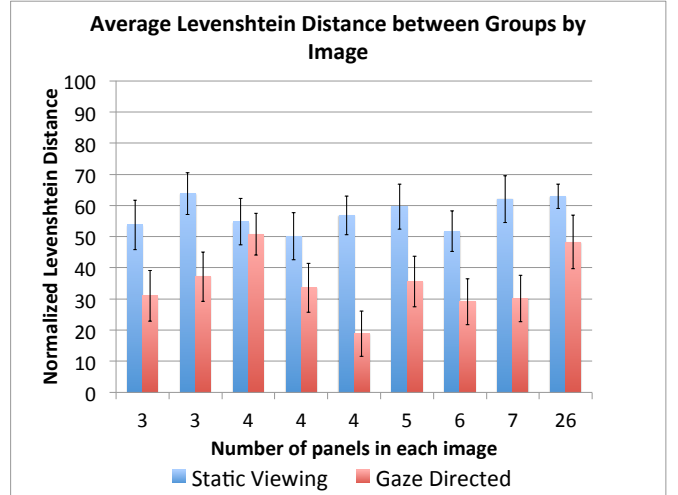


Figure 5: Normalized Levenshtein distance measure between static viewing group and gaze directed group for each image. The x-axis indicates the number of panels in each image. The error bars represent one standard error.

group for images having relevant regions varying from 3 to 26. In eight of the nine images used in the study, this result was shown to be significant. Independent-samples t-tests reveal that this effect was significant and not due to chance, the t-test results for images with relevant regions 3,4,5,6,7 and 26 are shown in Table 1.

The results did reveal a single anomaly where the t-test did not show a significant difference between groups. Image “C”, shown in Figure 7, revealed no significant difference in performance between the two groups. Further inspection showed that in this image, the artist has gradually decreased the luminance of the narrative art over the story. This analysis was possible as the same characters appear over multiple regions in the image. We reason that this luminance change in itself would provide a strong enough visual cue to enable the participants in Group 1 to correctly navigate the story. This also suggest that luminance changes could serve to guide direct gaze in

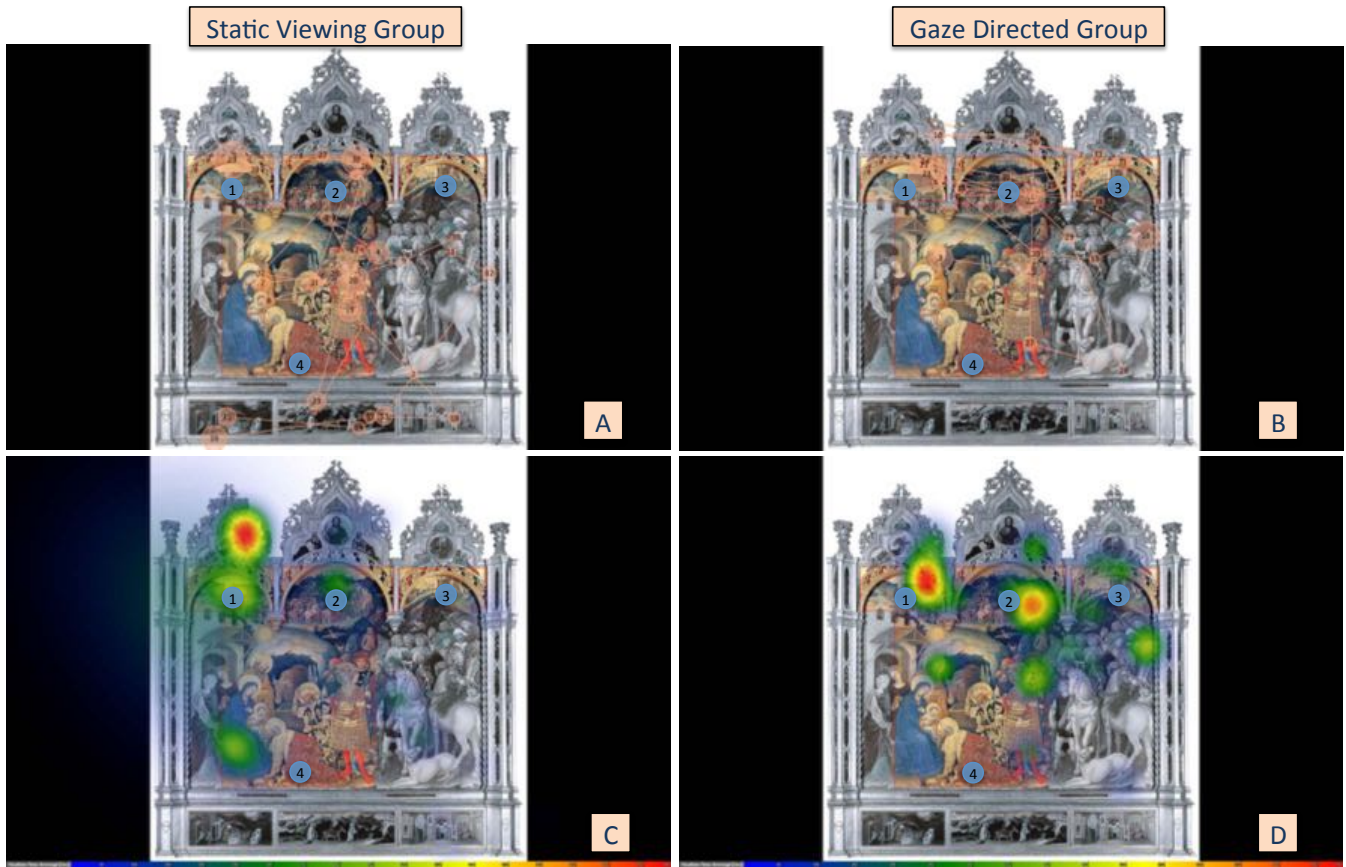


Figure 6: This Figure shows the scan paths and heat-maps for one participant from each of the two groups (no modulation and SGD), Image A & C represent data collected from Group 1, while images B & D were collected from Group 2. Rectangular highlighted regions denote panels. Blue numbered circles indicate the correct viewing order of panels within the image. As can be seen from the images, gaze distribution and fixations are more accurately aligned with panel (modulated) regions in the SGD condition.

Number of panels	Independent t-test (Group 1 V Group 2)
3	$t(33) = 2.0322$
4	$t(33) = 2.0346$
5	$t(33) = 2.0345$
6	$t(33) = 2.0340$
7	$t(33) = 2.028$
26	$t(33) = 2.0364$

Table 1: Independent *t*-tests indicate significant differences in the ability to correctly predict intended panel sequences for images with vary in numbers of panels. In each case, $p < 0.05$.

imagery. This phenomenon is a topic for future research.

To further illustrate the success of SGD we present a single example. Figure 6 shows the images for the static viewing and the gaze directed group respectively. Images are placed side-by-side for comparison.

Image A depicting the scan path of the viewer's gaze over the static images shows that the viewer's gaze does not coincide with all of the relevant story panels. Contrast this with image B, which shows the scan path over the SGD enhanced image. This image reveals a more coherent scan path in terms of visitation to each relevant panel.

Comparing heat maps reveals a similar story. The heat maps represent the amount of time spent fixating in each image region. Figure 6, Image C reveals that most fixations fall to the left of the image, and the distribution does not encompass the story panels. Conversely, examination of the heat map for image D (SGD) indicates that viewer fixations are distributed over the story panels.

Examining the L value measures (as described in Section 3.2) for this single image, the Group 1 (no SGD) participant's value is 71.45 compared to 28.57 for the participant from Group 2 (the gaze directed group). Thus the heat map and scan path analysis not only reflect the increase in the gaze coverage and attention to all relevant regions of the image for the gaze directed group over the static viewing group, but also correspond well with the L-value metric chosen to compare performance.

It is important to note that Figure 6 serves as a representative example of a consistent trend across all nine images viewed. This analysis reveals that, *without SGD*, not only did participants fail to view all of the story panels, but they failed to fixate on all the relevant story panels. The exact opposite is true for those images presented with SGD applied to the story panels, giving a high level of confidence in the success of applying SGD to subtly reveal an intended viewing sequence.

Informal reporting, post experiment, showed that a small subset of participants (approximately 15%) reported noticing the modulations, but (as designed) the modulation disappeared before they

could inspect it. A single participant reported trying to follow the modulations, again performing exactly the action SGD is designed for. In these cases we reason that certain individuals may have heightened peripheral vision e.g. it has been proposed by several researchers, for example, that strong peripheral vision may give basketball players, and other athletes a distinct advantage during game play [Vickers 2007]. In general, however the premise of subtly holds well for SGD. Even in cases where individual participants noticed SGD task performance was not impeded.

In summary our results show that gains can be made in task performance when modulation is employed to direct gaze to target a sequence of panels in a specific order. Even if the modulations are noticed to some degree, there is still an increased accuracy in task performance. This seems to hold true over a range of images, and over a range of panel numbers. For applications that require a specific viewing order for understanding or performance, SGD can serve as a subtle aid to boost accuracy of performance on a sub-image ordering task.

5 Conclusions and Future Work

We presented an experiment to compare task performance in digital images across two groups of stimuli. In one group no image alterations were used (Group 1), in the second group small modulations were applied to image panels in an effort to direct the viewers gaze (Group 2). The participant's task was to specify the order of panels (contained in episodic art pieces) which revealed the intended story. The results indicate that using a subtle gaze direction technique, which modulates the appropriate panel in the intended sequence, does indeed improve the precision of panel ordering. The difference between performance between the two groups was shown to be significant.

For this study we chose to modulate the luminance channel within a small radius of pixels. All modulations took the same shape and were modulated using the same oscillation strength. Given the variation of the subset of images we included in this study, and the variation in the number of panels, it may be that variation of modulation shape and strength would yield further improvements in task performance. For example, when the number of relevant panels is large, perhaps a stronger modulation would attract gaze faster. For future work it may be useful to implement modulations characteristics based on image content rather than take a "one-size-fits-all" approach.

Also, in this chosen domain, all of the imagery is static. We are interested in applying SGD to dynamic environments, which may pose increased difficulty due to the attentional draw of objects in motion. We anticipate that stronger modulations will be required to successfully use SGD effectively in dynamic environments, but the payoff could be beneficial in areas such as simulation training and educational gaming.

We have shown that SGD can improve performance on a within-image panel ordering set without noticeably disrupting the visual experience of the image. This technique can be applied to help guide gaze in complex applications where viewing order is critical to understanding, such as in story telling, or training task performance where sequence of operations is important, for example, construction instructions.

Acknowledgements

This material is based in part on work supported by the National Science Foundation under Award No. IIS-0952631. Any opinions, findings, and conclusions or recommendations expressed in

this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

- ANDOLINA, S., SANTANGELO, A., CANNELLA, M., GENTILE, A., AGNELLO, F., AND VILLA, B. 2009. Multimodal virtual navigation of a cultural heritage site: the medieval ceiling of steri in palermo. In *Proceedings of the 2nd conference on Human System Interactions*, IEEE Press, Piscataway, NJ, USA, HSI'09, 559–564.
- BAILEY, R., MCNAMARA, A., SUDARSANAM, N., AND GRIMM, C. 2009. Subtle gaze direction. *ACM Trans. Graph.* 28 (September), 100:1–100:14.
- BORDWELL, D., 2011. <http://www.davidbordwell.net/blog/2011/02/14/watching-you-watch-there-will-be-blood/>.
- BOSCOTRECASE, IBC. Landscape with perseus and andromeda , <http://www.metmuseum.org/toah/works-of-art/20.192.16>.
- BRUNS, E., BROMBACH, B., ZEIDLER, T., AND BIMBER, O. 2007. Enabling mobile phones to support large-scale museum guidance. *IEEE MultiMedia* 14 (April), 16–25.
- CHOU, S.-C., HSIEH, W.-T., GANDON, F. L., AND SADEH, N. M. 2005. Semantic web technologies for context-aware museum tour guide applications. In *Proceedings of the 19th International Conference on Advanced Information Networking and Applications - Volume 2*, IEEE Computer Society, Washington, DC, USA, AINA '05, 709–714.
- CHOUDARY, O., CHARVILLAT, V., GRIGORAS, R., AND GURDJOS, P. 2009. March: mobile augmented reality for cultural heritage. In *Proceedings of the 17th ACM international conference on Multimedia*, ACM, New York, NY, USA, MM '09, 1023–1024.
- DA FABRIANO, G., 1423. Adoration of the magi ,<http://arthistoryblogger.blogspot.com/2011/11/continuous-narrative-in-art.html>.
- DAMALA, A., CUBAUD, P., BATIONO, A., HOULIER, P., AND MARCHAL, I. 2008. Bridging the gap between the digital and the physical: design and evaluation of a mobile augmented reality guide for the museum visit. In *Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts*, ACM, New York, NY, USA, DIMEA '08, 120–127.
- DI BUONINSEGNA, D., 1311. Christ taken prisoner. <http://www.wga.hu/frames-e.html?/html/d/duccio/maesta/verso.1/verso07.html>.
- DI BUONINSEGNA, D., 1311. Maesta alterpiece, <http://www.casasantapia.com/art/duccio/maestabackpanels.htm>.
- DANTONIO, B., 1485. The story of joseph. http://www.getty.edu/education/teachers/classroom_resources/curricula/stories_in_art/downloads/sia_story_joseph.pdf.
- GWILT, I. 2009. Augmented reality and mobile art. In *Handbook of Multimedia For Digital Entertainment and Arts*, B. Furht, Ed. Springer US, 593–599.
- ITTI, L., AND KOCH, C. 2000. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research* 40, 10-12 (May), 1489–1506.
- ITTI, L., AND KOCH, C. 2001. Computational modelling of visual attention. *Nature Reviews Neuroscience* 2, 3 (Mar), 194–203.

- LEVENSHTIN, V. 1965. Binary codes capable of correcting spurious insertions and deletions of ones. *Problems of Information Transmission* 1, 8–17.
- LEVENSHTIN, V. I. 1966. Binary codes capable of correcting deletions, insertions, and reversals. Tech. Rep. 8.
- MASACCIO, A., 1421. The tribute money, http://en.wikipedia.org/wiki/the_tribute_money.
- MCNAMARA, A., BAILEY, R., AND GRIMM, C. 2008. Improving search task performance using subtle gaze direction. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization*, ACM, New York, NY, USA, APGV '08, 51–56.
- NAVARRO, G. 2001. A guided tour to approximate string matching. *ACM Comput. Surv.* 33, 1 (Mar.), 31–88.
- OGDEN, T. E., AND MILLER, R. F. 1966. Studies of the optic nerve of the rhesus monkey: Nerve fiber spectrum and physiological properties. *Vision Research* 6, 1, 485–506.
- SPILLMAN, L. 1990. *Visual Perception: The Neurophysiological Foundations*. Academic Press, San Diego.
- SRINIVASAN, R., BOAST, R., FURNER, J., AND BECVAR, K. M. 2009. Digital museums and diverse cultural knowledges: Moving past the traditional catalog. *The Information Society* 25 (July), 265–278.
- UNDERWOOD, G., AND FOULSHAM, T. 2006. Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *Q J Exp Psychol (Colchester)* 59, 11 (Nov.), 1931–1949.
- UNDERWOOD, J., TEMPLEMAN, E., AND UNDERWOOD, G. 2009. Attention in cognitive systems. Springer-Verlag, Berlin, Heidelberg, ch. Conspicuity and Congruity in Change Detection, 85–97.
- VEAS, E. E., MENDEZ, E., FEINER, S. K., AND SCHMALSTIEG, D. 2011. Directing attention and influencing memory with visual saliency modulation. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, ACM, New York, NY, USA, CHI '11, 1471–1480.
- VELLI, S. T. 2007. *Le immagini e il tempo: Narrazione visiva, storia e allegoria tra Cinque e Seicento*. Edizioni Della Normale, Italy.
- VICKERS, J. 2007. *Perception, Cognition, and Decision Training: The Quiet Eye in Action*. Human Kinetics, New York, NY.
- WITZ, K., 1434. The miraculous draught of the fishes, <http://mydailyartdisplay.wordpress.com/2011/07/18/the-miraculous-draught-of-the-fishes-by-konrad-witz-2/>.



Figure 7: Narrative artwork images used in this study, from top left to right, *The Miraculous Draught of the Fishes* (Konrad Witz) [Witz 1434], *Adoration of the Magi* (Gentile da Fabriano) [da Fabriano 1423], *The Tribute Money* (Masaccio), [Masaccio 1421], *Christ Taken Prisoner* (Duccio di Buoninsegna) [di Buoninsegna 1311a], *The Story of Joseph* (Biagio d'Antonio), [d'Antonio 1485] reproduced with kind permission of The J. Paul Getty Museum, Los Angeles, *Landscape with Perseus and Andromeda*, [Boscotrecase 1BC], *Maesta Altarpiece* (Duccio di Buoninsegna), [di Buoninsegna 1311b].