

Inferring Artistic Intention in Comic Art through Viewer Gaze

Eakta Jain¹ and Yaser Sheikh¹ and Jessica Hodgins^{1,2*}
¹Carnegie Mellon University ²Disney Research Pittsburgh

(a) Comic art from *Watchmen* (©DC Comics)



First three seconds

Entire duration

(b) A photograph taken automatically by a roaming robot



First three seconds

Entire duration

Figure 1: Eyetracking data from viewers is overlaid on the stimuli images. Each viewer is shown in a different color. The circles are fixation locations and the lines represent saccades. (a) The viewers all attend to the character’s face and the yellow object in his hand. (b) There is less consistency in viewer gaze on robot photographs, compared to artist-created comic art. The green viewer explores the poster board on the left while the purple viewer prefers the poster on the right side of the photograph, for example.

Abstract

Comics are a compelling, though complex, visual storytelling medium. Researchers are interested in the process of comic art creation to be able to automatically tell new stories, and also, summarize videos and catalog large collections of photographs for example. A primary organizing principle used by artists to lay out the components of comic art (panels, word bubbles, objects inside each panel) is to lead the viewer’s attention along a deliberate visual route that reveals the narrative. If artists are successful in leading viewer attention, then their intended visual route would be accessible through recorded viewer attention, i.e., eyetracking data. In this paper, we conduct an experiment to verify if artists are successful in their goal of leading viewer gaze. We eyetrack viewers on images taken from comic books, as well as photographs taken by experts, amateur photographers and a robot. Our data analyses show that there is increased consistency in viewer gaze for comic pictures versus photographs taken by a robot and by amateur photographers, thus confirming that comic artists do indeed direct the flow of viewer attention.

Keywords: Comic art, Eyetracking

1 Introduction

Comic art is the art of telling stories through a sequence of pictures arranged in space [McCloud 1993]. This art form is a compelling and well-loved storytelling medium that is seeing a resurgence of interest as digital devices are making it easier to create,

*e-mails: ejain, yaser, jkh@cs.cmu.edu

distribute and consume comic-based stories [Brothers 2010]. Simultaneously, researchers have been exploring algorithms that aim to create comics automatically, or assist artists in the process. This research aims at not only the telling of new stories via comics, but also summarizing game plays, presenting photo albums and transforming movies into graphic novels.

Comic art is challenging to create because the communication of a visual story depends on the viewer looking at the right place at the right time. The individual pictures (or panels) that comprise a comic are arranged in space rather than time, and audio (sound and dialogue) is replaced with visual symbols: text via word bubbles or effects such as lines that denote speed. Each comic panel is, thus, a carefully crafted moment frozen in time. Perhaps the popularity of the art form is due in part to this characteristic, because it allows the reader-viewer to fill in the gaps with their imagination.

As a result, the creation of a comic (given the story) requires many decisions of the artist. Some decisions, such as whether to have the comic in black and white or color, are an artistic choice. Algorithms can assist artists with other decisions, such as the moments of time to ‘freeze’ in each panel, the arrangement of the various objects in a panel, the lighting in the panel, where to place the word bubbles and whether to embellish the picture with lines of action or words such as “bang”. Computer graphics research in comic book art has focused on automating several of these decisions individually.

How should these individual (automatic or semi-automatic) decisions be combined to computationally create a finished comic book? In addition to communicating the story and the personalities of the characters, a primary goal while putting together the components of a comic book is planning the flow of viewer attention. For example, if the viewer were to see the knife hidden in the jacket of the scheming army chief before she saw the king standing in front of him, the element of surprise would be lost. Because comic panels are arranged spatially rather than presented temporally (as in film), the comic artist must lead the viewer along the storytelling route.

Will Eisner stated that a primary organizing principle in creating comic art is to “...secure control of the readers attention and dictate

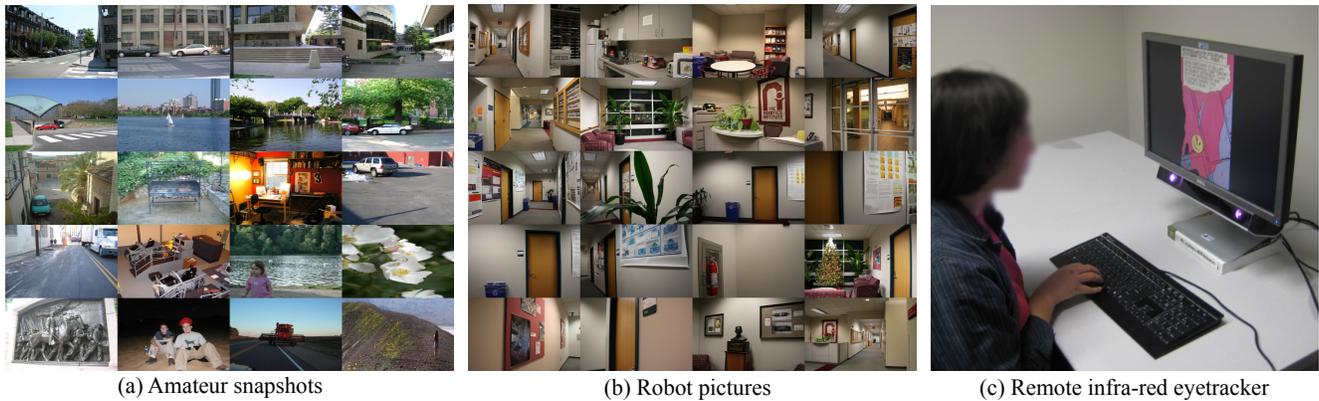


Figure 2: Sample photographs from the set of (a) amateur snapshots and (b) robot pictures. (c) Gaze data was collected with a remote infra-red eyetracker.

the sequence in which the reader will follow the narrative...” [Eisner 2008]. Though a second artist may attempt to guess at the artist-designed visual route in a comic, this guess will be his or her interpretation. It can be expensive or impossible to interview the original comic artist. Our key insight is that eyetracking viewers who read/view the comic art gives us access the original artistic intent. Algorithms that perform computations on comic art could understand and respect this intent.

In this paper, we conduct an experiment to verify if comic book artists are in fact successful in this stated goal of controlling viewer attention. We eyetrack viewers on a variety of images and analyze their recorded gaze data. Because reading conventions are usually implicitly followed in comics with a left to right and top to bottom panel based layout, it is expected that viewer attention from panel to panel will be consistent. Our experiment therefore tests if the artist is successful in securing control of viewer attention inside each individual panel. We find that there is increased consistency in viewer eye movements on comic book stimuli relative to pictures that were not created with the intent of directing viewer attention, for example, pictures taken by amateur Flickr users or by a robot.

The findings indicate that artistic intent is indeed encoded in measurements of viewer gaze and can be inferred by eyetracking multiple viewers. This access to the original artist-designed visual route through a comic picture could enable algorithms to automatically modify existing comic books, for example, by extracting and resizing panels in the correct storytelling order to make a comic more readable on handheld digital displays. For the creation of new comics, algorithms could leverage viewer attention as an energy term to optimally combine the various decisions that go into the making of a comic.

2 Background

Eye movements are driven both by the properties of the presented stimuli (bottom-up cues such as high contrast [Itti and Koch 2001]) and the idiosyncrasies of the viewer (top-down influences such as finding a particular face more attractive or being asked to answer a question about the stimuli [Yarbus 1967]). However, artistic wisdom is that photographers, graphic artists and film directors are able to get viewers to look at certain regions versus others, i.e., manipulate viewer attention. Past research has investigated the difference in the visual perception of artistic stimuli versus ‘natural’ stimuli.

Dorr and colleagues found that the characteristics of eye move-

ments on static images are different from those on continuous videos [Dorr et al. 2010]. In particular, the eye movements of different observers on Hollywood trailers were found to be significantly more coherent than on natural movies (a beach scene, for example). Carmi studied the effect of ‘MTV-style jump cuts’ on viewer gaze compared with continuous video [Carmi and Itti 2006; Carmi 2007]. They found that there is reduced inter-observer variability in eye movements and increased center bias in the MTV-style condition, possibly because bottom-up cues are more dominant in this condition [Carmi 2007]. We contribute to the research in the visual perception of artistic content by experimentally verifying a widely held belief in the comic art community: that the artist is able to purposefully direct the visual attention of readers through the pictorial narrative.

Researchers have made some preliminary explorations into how the comic artist is able to direct viewer attention. Omori and colleagues recorded eye movements on Japanese manga to understand why readers skipped certain panels and paid attention to others [Omori et al. 2004]. They found that changing the locations of dense word bubbles could change whether viewers skipped or attended to the preceding panel.

There is also growing interest in building algorithms to automatically or semi-automatically create the various components of comic art. The first step in creating a comic is selecting the appropriate time points to ‘freeze’ in each panel. Shamir and colleagues [Shamir et al. 2006] and Alves and colleagues [Alves et al. 2008] provide algorithms to perform this selection based on text logs or scripts to summarize an interactive game play in comic book format. Katti and colleagues use gaze information, such as pupil dilation, to infer the portions of a video that are interesting to the viewers to create a storyboard or a comic book summary of a video [Katti et al. 2011]. A recent work by Toyoura and colleagues also uses gaze information to detect scene changes and to insert word bubbles [Toyouira et al. 2012b].

One of the next steps is determining the shapes and positions of the individual panels to create a page. Hong and colleagues present a system to convert a movie into a comic by extracting the key scenes, laying out the panels in a page and adding word bubbles based on the script and subtitles [Hong et al. 2010], and Toyoura and colleagues map the motion of the camera in a film to panel layouts [Toyouira et al. 2012a]. Myodo and colleagues arrange a collection of photographs in a comic book layout to create a photo album [Myodo et al. 2011]. The system by Tobita and colleagues gives the user guidelines on how the shape of a panel and the place-

ment of word bubbles can be varied to change the ‘tempo’ (pace) of the story [Tobita 2010]. There are also several works that aim to convert printed comics to a format more easily read on a digital display by automatically extracting individual panels from the page layout [Chan et al. 2007; Arai and Tolle 2010; Ho et al. 2011] and changing the panel size based on the corresponding narrative time [Bares 2008]. Our work suggests a direction for making these components more automatic and combining the various individual decisions: leading viewer attention along a deliberate route could be used as an energy term in these optimizations.

3 Eyetracking Viewers on Various Image Categories

The aim of this experiment is to test if comic book artists are successful in controlling viewer attention. Therefore, we pose the following two-part question: if there was no effort put into directing the viewer’s attention, then how much eye movement consistency do we observe across different viewers? If there is expressed effort put into directing the viewer’s attention, then, is the consistency across viewers increased?

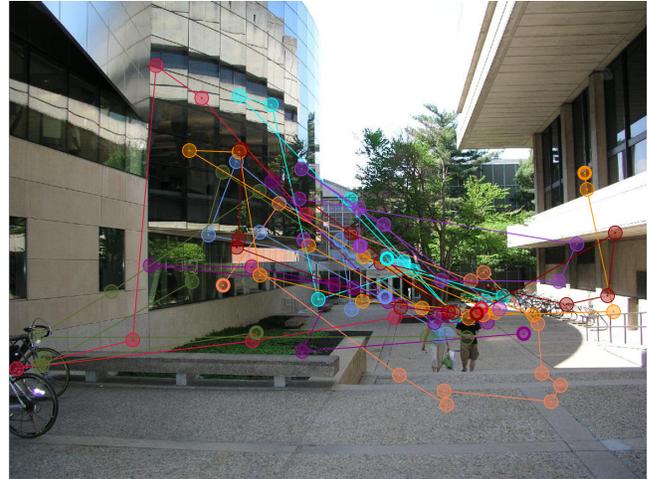
We showed viewers a collection of pictures drawn from five categories: 32 images from the *Watchmen* graphic novel, 35 images from the *Ironman: Extremis* comic book, 18 photographs from photoessays found in online issues of political and nature magazines, 30 images from the collection of ‘robot’ pictures and 30 images from the collection of ‘amateur snapshots’ (sample images are shown in Figure 2 (a) and (b)). Some of the comic book panels contained word bubbles, and these panels were left intact. ‘Robot’ pictures were collected by a robot wandering through the hallways of a campus building while recording images at given intervals [Kang et al. 2009]. The amateur snapshots were randomly chosen from the online dataset made available by Judd and colleagues [Judd et al. 2009]. Judd described the images in the database as ‘amateur snapshots’ (personal communication), which refers to the quality of the photographs as being amateur records of events, rather than carefully composed works of art. The photoessays contained human and non-human (animal) faces, in addition to objects such as cars and trees. In total, every participant observed 145 pictures.

3.1 Method

Ten naïve participants (8 males, 2 females, age ranging from 21 to 44 years) were recruited from a participant pool available via a website, in accordance with the guidelines set by the institute’s review board. They were compensated monetarily for their time. The data for one participant was removed because of large calibration error in the eyetracking device.

Participants were asked to view the stimulus images and told that they would be asked ten questions at random points in the viewing session. The questions were based on comprehension, rather than memorization, for example, ‘What did the man find in the cupboard?’ and ‘Was there a fire extinguisher in the building?’, and the answers were multiple choice. The stimuli images were presented in randomized order. We chose not to advance the stimuli automatically because it would be impossible to set the timing appropriately for every picture and every participant. Too little time might result in the picture advancing too soon, for example, the participant might be reading the text in the word bubbles and would then miss seeing the pictorial portion. On the other hand, it has been observed that consistency in fixation locations decreases as more time is spent on the stimulus, probably because top-down strategies diverge [Tatler et al. 2005], and therefore, too much time could result in individual idiosyncrasies being asserted over artistic direction.

Top: only first three seconds



Bottom: all eyetracking data

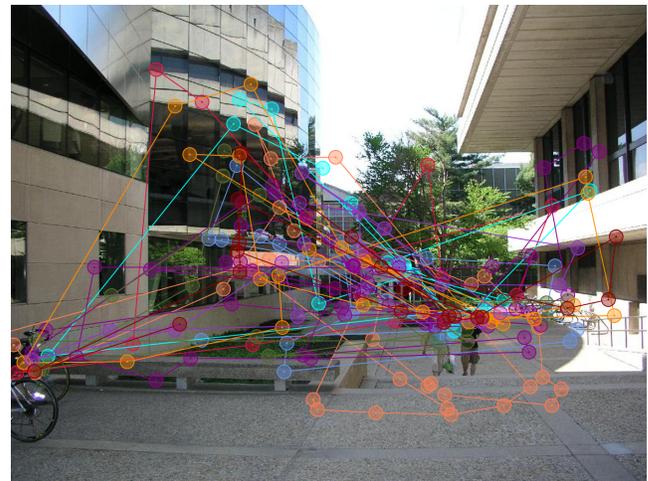


Figure 3: Eyetracking data from nine participants is overlaid on a stimulus image from the amateur snapshots category. The circles represent fixations and the lines represent saccades. The stimulus image is made as large as possible on a 1680×1050 screen.

To balance these constraints, we paced the study so that participants could advance through the images by pressing a button, but were required to spend a certain minimal amount of time on each stimuli image. The minimal time was set to be the average time spent by the pretest participants (four seconds).

Participants viewed the stimuli on a 22 inch monitor, at pixel resolution 1680×1050 . All images were resized to be as large as possible, while keeping their aspect ratio intact. They were letter-boxed with a gray matte (RGB=(127, 127, 127)) to fit the stimuli presentation screen. The participants sat approximately 18-24 inches away from the screen. A visual angle of 1 degree was approximately 30 pixels at these settings. Figure 4 illustrates the dimensions of various objects in two images as a comparison.

Participants were asked to adjust the chair to a height and distance comfortable to them; then the system was calibrated. After calibration, they could move their head freely. This setup allowed for a natural setting in which to observe the presented materials, unlike systems that require a chin rest or a head-mounted eyetracker.

Left: only first three seconds for each participant.

Right: all eyetracking data for each participant.

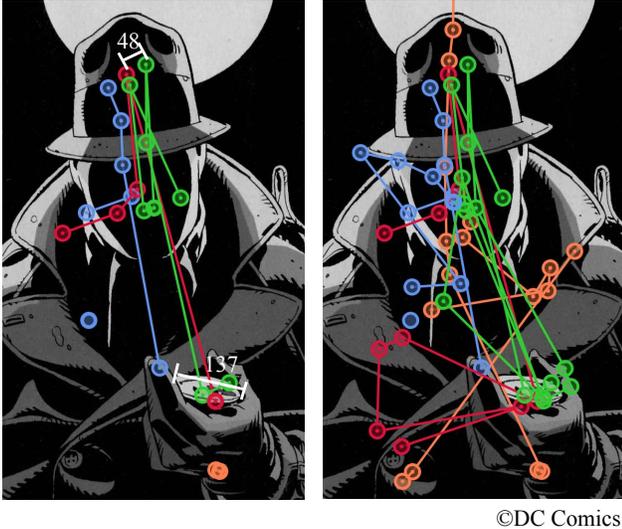


Figure 4: We show eyetracking data from four of the nine participants for clarity. The plots are overlaid on a grayed-out stimulus image. The circles represent fixations and the lines represent saccades. The stimulus image is resized to fit to a 1680×1050 screen. The diameter of the ‘smiley’ button is about 137 pixels. Calibration error is 30 – 40 pixels. Left: eyetracking data for the first three seconds. Right: eyetracking data for the full duration.

The participants all had normal vision or wore contact lenses to correct to normal vision. For eyetracking, we used SensoMotoric Instruments’ RED eyetracker, running the iViewX software (Figure 2(c)). Points of regard, that is, the two-dimensional gaze locations on the calibrated screen, were recorded at 60 Hz. Raw data was converted into saccades (rapid eye movements between objects of interest) and fixation points (the eye is positioned so that the object of interest is in the foveal region) inside the proprietary SMI analysis software before being exported as a spreadsheet.

3.2 Analysis

Our goal is to discover numerically if the viewers followed a more similar visual route to each other while viewing stimuli belonging to the comic art categories, compared with the robot or amateur pictures. We analyzed similarity in the unordered and ordered fixation locations of observers across categories. Because we are interested in understanding the spread of visual attention through the scene, we compute fixation locations and fixation order, but not fixation durations, in our analysis. We also normalize for the difference in stimuli sizes by rescaling all the stimuli images to the same area (while keeping aspect ratio intact) before computing the similarity metrics (to ensure that we do not achieve a smaller score on an image just because it is smaller in size).

Several metrics have been proposed to compute the similarity between eye movements of different observers. One class of metrics constructs a fixation map from recorded gaze data by convolving the fixation locations with Gaussian kernels [Pomplun et al. 1996; Wooding 2002]. The fixation map for one observer is thresholded to create a binary map, which serves as a classifier for the fixation locations of all other observers [Tatler et al. 2005; Judd et al. 2009]. The percentage of fixations that fall inside the map are the inliers and the percentage of total area inside the binary map is the salient

percentage for a given threshold value. An ROC curve is plotted by varying the value of the threshold. The mean curves for each stimuli category, by leaving out each observer in turn, are shown in Figure 5 (a).

The area under the ROC curve measures how similar the fixation locations of a given observer are to other observers. Curves closer to the diagonal indicate lesser similarity compared with curves that rise sharply. Because comic books typically contain text inside word bubbles, while photographs do not, we plot ROC curves for the comic book categories with text (solid lines with cross markers) and without text by color segmenting out the word bubbles. The curves for *Watchmen* are nearly identical with and without text because only four of the 32 images contain word bubbles, with only one word each.

The analysis of the ROC areas (Figure 5 (b)) shows that fixation locations are more consistent across observers for the comic book categories, *Ironman* and *Watchmen* (red and magenta curves, respectively), compared to the robot pictures and amateur snapshots (green, dark blue curves). The difference in ROC areas is significant (one-way ANOVA, $F(4, 140) = 27.62, p < 0.05$, followed by Tukey-Kramer for post-hoc analysis, $p < 0.05$).

We also verified this analysis with another measure: Normalized Scanpath Saliency (NSS) [RJ Peters and Koch 2005; Dorr et al. 2010]. The fixation map is normalized to zero mean and unit standard deviation. For each observer, the NSS score is computed by looking up the values of the fixation map (for all other observers) at the fixation locations of this observer and taking the average. If this observer had fixated on locations similar to other observers, the average value will be larger than if this observer had fixated on different locations. In Figure 5 (c), we show the NSS score for each category. For our analysis, we used a Gaussian kernel with $\sigma = 50$, which is about 1° - 1.5° visual angle.

We computed the NSS score with each participant in turn as the test observer (leave-one-out) and averaged this score to obtain the NSS measure of inter-observer gaze variability for each stimulus image. Gaze data from different observers was significantly more consistent on *Watchmen* and *Ironman*, compared with robot pictures, amateur snapshots and photoessays (one-way ANOVA, $F(4, 140) = 13.79, p < 0.05$, followed by Tukey-Kramer for post-hoc analysis, $p < 0.05$).

These two metrics measure the spatial similarity in fixation locations, irrespective of the order in which the fixations occurred. If an observer looked at the top left corner and then the bottom right corner, and another observer looked at the same two locations but in the opposite order, they would still have a high similarity score. Because artists aim to deliberately direct the viewer’s attention through their picture, we expect that observers will not only look at similar locations, but also do so in the same order.

The second class of metrics to measure similarity in eye movements considers both spatial and temporal similarity, that is, fixation locations and their order. As an example, the stimulus images may be preprocessed to mark out regions of interest (ROI’s) which are assigned unique alphabet identifiers. Eye movements for different observers are compared by computing the string-edit (Levenshtein) distance between strings obtained by concatenating ROI identifiers. Because this metric requires us to know the regions of interest a priori, it is suitable for a study where there is a well-known structure to the stimuli (e.g., the presence of a face).

Instead, we follow the method by Jarodska and colleagues, who expressed recorded gaze data as vectors of positions and durations and presented a suite of vector distance metrics to compute inter-observer consistency [Jarodska and Holmqvist 2010]. We denote

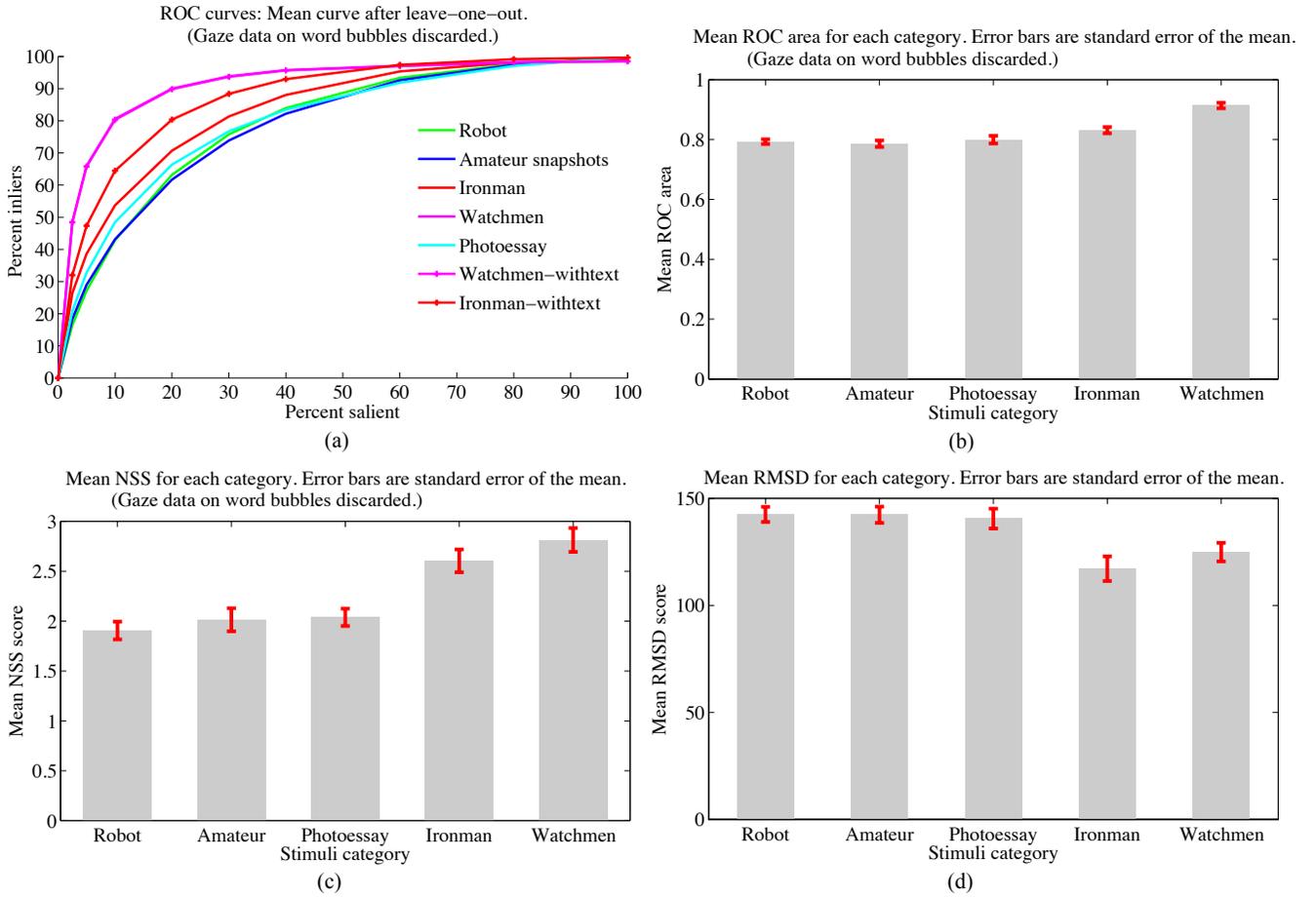


Figure 5: Viewer gaze data is more consistent for the comic book panels compared with the robot pictures, amateur snapshots and photoessays. (a) ROC curves for the different categories. (b) Mean ROC area for each category. (c) Mean Normalized Scanpath Saliency (NSS) score for each category. (d) RMS distance between ordered fixation locations.

the ordered fixation locations for observer s on picture p by \mathbf{X}^{ps} . Then, $\mathbf{X}^{ps} = [\mathbf{X}_1^{ps}, \mathbf{X}_2^{ps}, \dots, \mathbf{X}_{\gamma_{ps}}^{ps}]^T$, where \mathbf{X}_l^{ps} is the l -th fixation location, and γ_{ps} is the total number of fixations on this picture. The value γ_{ps} could be different because two observers might have different reading speeds, or because one of them fixated longer at each location, while the other jumped back and forth between the locations, for example. In addition, the same observer might attend to a different number of regions on different stimuli, or, change their speed of perusal.

We warp the vectors \mathbf{X}^{ps} to be the same length via the Dynamic Time Warp algorithm [Sakoe and Chiba 1990; Ellis 2003]. Let $\tilde{\mathbf{X}}^{ps}$ be the warped gaze path for observer s on the picture p , and $\tilde{\gamma}_{ps}$ be the length of this vector. The distance $\Delta_{i,j}^p$ between the gaze paths of observers i and j on the picture p is the root mean squared Euclidean distance in image space between the warped fixation location vectors,

$$\Delta_{ij}^p = \sqrt{\frac{\|\tilde{\mathbf{X}}^{pi} - \tilde{\mathbf{X}}^{pj}\|_2}{\tilde{\gamma}_{pi}}} \quad i, j = 1, 2, \dots, S, \quad (1)$$

where S is the total number of observers in the experiment. This measure captures spatio-temporal variability and is similar to the vector comparison metric presented by Jarodska and colleagues when the vector consists of fixation locations only [Jarodska and Holmqvist 2010].

Figure 5 (d) illustrates the mean distances for each of the five categories. The mean distances for the comic books *Ironman* and *Watchmen* are statistically different from the robot dataset and the amateur snapshots (one-way ANOVA $F(4, 140) = 6.8, p < 0.05$, followed by Tukey Kramer test, $p < 0.05$). We find greater inter-observer consistency for *Ironman* compared to *Watchmen* (though not a statistically significant difference), most likely because the pictures in the *Ironman* set have word bubbles with a fair amount of text.

In all the analyses, eye movements across observers are less consistent for the photoessays compared with comic artwork. This difference could be because photographs, however carefully composed, are still tied to the physical world. Photographers have limited control over many aspects of the scene they are capturing, for example, it may be difficult to avoid a distracting billboard in the background. The painterly style of comic art, involving strong contrast, sharp outlines and bold colors, is also different from the photographs. It is likely that graphic artists manipulate the painterly effects to direct the viewer's attention, something that photographers have limited control over. Finally, photographers are limited to a physical camera, so they cannot exaggerate too much, for example, by enlarging a person's fist more than their lens will allow (without using photo-editing software).

4 Discussion

In this paper, we conducted an experiment to verify if comic book artists are successful in controlling viewer attention. We eyetracked viewers on stimuli images drawn from five categories and analyzed the recorded gaze data with various metrics to measure inter-observer gaze consistency. Our results confirm that there is increased consistency among viewers, both in the locations of fixations and the order in which they occur, when looking at pictures from the comic book categories, relative to robot pictures and amateur snapshots (i.e., pictures that were not created by artists with the goal of leading the viewer). The findings suggest that artistic intent is encoded in the eye movements of viewers looking at comic art, and that it can be inferred from the recorded gaze data.

In comic art, the communication of the story depends on the reader-viewer looking at the right place at the right time. Though film is also a visual medium, comics offer different challenges because the pictures are arranged in space rather than time, and audio cues are replaced with visual symbols. As a result, the onus is on the artist to lead the viewer's eye along the storytelling route. The artist must be able to direct the viewer by appropriately leveraging pictorial devices such as color, lighting, lines of action and word bubbles. Perhaps this planning of the flow of viewer attention is formalized as the concept of 'visual composition'. An interesting direction of future research could be to test if pictures that are rated better composed are ones where viewer gaze is more consistent.

Future studies could also investigate the impact of the various artistic devices in influencing viewer gaze, for example, if the same panel was rendered in black and white versus color. We could compare photographs with controlled artistic renderings of the same scene to understand the factors that lead to better direction of viewer gaze. For comics, in particular, intra-panel factors (such as color contrast) interact with inter-panel factors (such as page layout) to lead the viewer. As shown by Yarbus [1967], eye movements are dependent on higher mental processes. We could expect gaze consistency to increase when comic panels are viewed in order, because viewers would be following the story across the panels.

A finished comic book is the product of the interactions of many individual decisions, from the layout of panels in a page and the choice of viewpoint and lighting, to the shape, color and position of word bubbles. Though recent algorithms can assist artists or automatically make several of these decisions, there is much progress to be made before we computationally understand the process of creating comic art. Our findings suggest that directing viewer attention could be used as an energy term to combine the various individual decisions. For example, in addition to placing the word bubble so that it does not occlude the speaker, the algorithm should arrange the panel so that the word bubble leads the viewer from the speaker to the listener. Or, image saliency algorithms could be trained for comic book pictures using eyetracking data and then, when putting together the contents of a panel, the algorithm could compute the saliency map for the panel and optimize the computed flow of attention. For example, it might move the various components of a panel to generate a more 'peaky' saliency map, which indicates that viewer attention is being focused. As a further sophistication, the algorithm could learn the types of visual routes designed by artists and then fit these shapes to the peaks of the saliency map for a new panel. It might also ensure that the shapes in the preceding panel point to the shapes in the following panel, thus leading the viewer through the entire page. Algorithms could also modify existing comic books, for example, by extracting individual panels and presenting them in the correct storytelling order to make a given comic more readable on a small handheld display.

Acknowledgments

We thank DC Comics for permission to use the comic images.

References

- ALVES, T., SIMÕES, A., FIGUEIREDO, R., VALA, M., PAIVA, A., AND AYLETT, R. 2008. So tell me what happened: turning agent-based interactive drama into comics. In *International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 3*, International Foundation for Autonomous Agents and Multiagent Systems, 1269–1272.
- ARAI, K., AND TOLLE, H. 2010. Automatic e-comic content adaptation. *International Journal of Ubiquitous Computing 1*.
- BARES, W. 2008. Panel beat: Layout and timing of comic panels. In *Smart Graphics*, vol. 5166 of *Lecture Notes in Computer Science*. 273–276.
- BROTHERS, D., 2010. The State of Digital Comics: Censorship, Price, Distribution, and More. www.comicsalliance.com/2010/11/22/state-of-digital-comics-censorship-price-distribution/.
- CARMI, R., AND ITTI, L. 2006. The role of memory in guiding attention during natural vision. *Journal of Vision 6*.
- CARMI, R. 2007. *Attention, Movie Cuts, and Natural Vision: A Functional Perspective*. PhD thesis, University of Southern California.
- CHAN, C., LEUNG, H., AND KOMURA, T. 2007. Automatic panel extraction of color comic images. In *Advances in Multimedia Information Processing PCM 2007*, vol. 4810 of *Lecture Notes in Computer Science*. 775–784.
- DORR, M., MARTINETZ, T., GEGENFURTNER, K., AND BARTH, E. 2010. Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision 10*.
- EISNER, W. 2008. *Comics and Sequential Art*. W.W. Norton & Company.
- ELLIS, D., 2003. Dynamic time warp (DTW) in Matlab. www.ee.columbia.edu/~dpwe/resources/matlab/dtw/.
- HO, A. K. N., BURIE, J.-C., AND OGIER, J.-M. 2011. Comic page structure analysis based on automatic panel extraction. In *Ninth International Workshop on Graphics Recognition*.
- HONG, R., YUAN, X.-T., XU, M., WANG, M., YAN, S., AND CHUA, T.-S. 2010. Movie2comics: a feast of multimedia artwork. In *ACM International Conference on Multimedia*, 611–614.
- ITTI, L., AND KOCH, C. 2001. Computational modelling of visual attention. *Nature Neuroscience*.
- JARODSKA, H., AND HOLMQVIST, K. 2010. A vector-based, multidimensional scanpath similarity measure. In *Eye Tracking Research and Applications (ETRA)*.
- JUDD, T., EHINGER, K., DURAND, F., AND TORRALBA, A. 2009. Learning to predict where humans look. In *IEEE International Conference on Computer Vision*.
- KANG, H., EFROS, A. A., HEBERT, M., AND KANADE, T. 2009. Image matching in large scale indoor environment. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Workshop on Egocentric Vision*.

- KATTI, H., YADATI, K., KANKANHALLI, M., AND TAT-SENG, C. 2011. Affective video summarization and story board generation using pupillary dilation and eye gaze. In *IEEE International Symposium on Multimedia (ISM)*, 319–326.
- MCCLLOUD, S. 1993. *Understanding Comics*. HarperPerennial.
- MYODO, E., UENO, S., TAKAGI, K., AND SAKAZAWA, S. 2011. Automatic comic-like image layout system preserving image order and important regions. In *ACM International Conference on Multimedia*, 795–796.
- OMORI, T., IGAKI, T., ISHII, T., KURATA, K., AND MASUDA, N. 2004. Eye catchers in comics: Controlling eye movements in reading pictorial and textual media. Tech. rep., Keio University.
- POMPLUN, M., RITTER, H., AND VELICHKOVSKY, B. 1996. Disambiguating complex visual information: Towards communication of personal views of a scene. *Perception* 25, 931–948.
- RJ PETERS, A IYER, L. I., AND KOCH, C. 2005. Components of bottom-up gaze allocation in natural images. *Vision Research*.
- SAKOE, H., AND CHIBA, S. 1990. Dynamic programming algorithm optimization for spoken word recognition. *Readings in speech recognition*, 159–165.
- SHAMIR, A., RUBINSTEIN, M., AND LEVINBOIM, T. 2006. Generating comics from 3d interactive computer graphics. *IEEE Computer Graphics & Applications* 26, 3, 30–38.
- TATLER, B. W., BADDELEY, R. J., AND GILCHRIST, I. D. 2005. Visual correlates of fixation selection: effects of scale and time. *Vision Research* 45, 643–659.
- TOBITA, H. 2010. Comic engine: interactive system for creating and browsing comic books with attention cuing. In *Proceedings of the International Conference on Advanced Visual Interfaces*, 281–288.
- TOYOURA, M., KUNIHIRO, M., AND MAO, X. 2012. Film comic reflecting camera-works. In *Advances in Multimedia Modeling*, K. Schoeffmann, B. Merialdo, A. Hauptmann, C.-W. Ngo, Y. Andreopoulos, and C. Breiteneder, Eds., vol. 7131 of *Lecture Notes in Computer Science*. 406–417.
- TOYOURA, M., SAWADA, T., KUNIHIRO, M., AND MAO, X. 2012. Using eye-tracking data for automatic film comic creation. In *Eye Tracking Research and Applications (ETRA)*.
- WOODING, D. S. 2002. Eye movements of large populations II: Deriving regions of interest, coverage, and similarity using fixation maps. *Behaviour Research Methods*.
- YARBUS, A. L. 1967. *Eye Movements and Vision*. Plenum Press.