

# Guiding Attention in Controlled Real-World Environments

Thomas Booth\*  
Rochester Institute of Technology  
Cindy Grimm §  
Oregon State University  
Srinivas Sridharan†  
Rochester Institute of Technology  
Reynold Bailey¶  
Rochester Institute of Technology  
Ann McNamara‡  
Texas A&M University

## Abstract

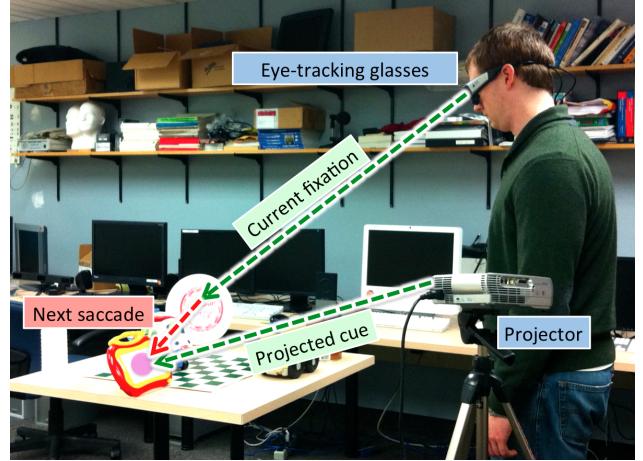
The ability to direct a viewer's attention has important applications in computer graphics, data visualization, image analysis, and training. Existing computer-based gaze manipulation techniques, which direct a viewer's attention about a display, have been shown to be effective for spatial learning, search task completion, and medical training applications. In this work we extend the concept of gaze manipulation beyond digital imagery to include controlled, real-world environments. We address two main challenges in guiding attention to real-world objects: determining what object the viewer is currently paying attention to, and providing (projecting) a visual cue on a different part of the scene in order to draw the viewer's attention there. Our system consists of a pair of eye-tracking glasses to determine the viewer's gaze location, and a projector to create the visual cue in the physical environment. The results of a user study show that we can effectively direct the viewer's gaze in the real-world scene. Our technique has applicability in a wide range of instructional environments, including pilot training and driving simulators.

**CR Categories:** I.3.6 [Computer Graphics]: Methodology and Techniques—interaction techniques I.4.7 [Image Processing and Computer Vision]: Feature Measurements—projections I.4.7 [Image Processing and Computer Vision]: Feature Measurements—feature representation;

**Keywords:** gaze manipulation, eye-tracking, SIFT, training

## 1 Introduction

In this paper, we present a technique for directing a viewer's attention in a controlled, real-world environment. Gaze manipulation approaches have been shown to be effective for computer-based spatial learning [Walther et al. 2005; Bailey et al. 2012], search task completion [McNamara et al. 2009], and medical training applications [Litchfield et al. 2010; Sridharan et al. 2012]. Existing computer-based gaze manipulation techniques, however, are mostly designed to guide a viewer's attention about a static digital display. We focus here on a technique called Subtle Gaze Direction [Bailey et al. 2009] that combines eye tracking with subtle image-space modulations to guide a viewer's gaze in a manner that has minimal impact on the viewing experience. The modulations are only presented to the peripheral regions of the field of view and are termi-



**Figure 1:** Annotated photograph of the real-world gaze manipulation setup. The current fixation is determined using a wearable eye-tracker. A projector is used to create a visual cue on another part of the scene in order to guide the viewer's gaze to that location. This process can be repeated to guide the viewer over time.

nated before the viewer can scrutinize them with high acuity foveal vision. Such approaches are preferred to more overt techniques that require permanent alterations to the scene to highlight areas of interest.

The goal of this work is to develop and test various approaches to extend the concept of gaze manipulation beyond digital imagery to include controlled, real-world environments. There are two main challenges to guiding attention to real-world objects: (1) determining what the viewer is currently paying attention to, and (2) providing (projecting) a visual cue on other parts of the scene to draw the viewer's attention there. In our system (Figure 1), the user wears a pair of eye-tracking glasses that allow for unrestricted head movement. The glasses are equipped with a front facing camera that captures the scene that the viewer is looking at and determines the viewer's gaze position within the scene. We extract the image features (SIFT) from this camera's video feed near to the viewer's current fixation and search for matching features in a corresponding high resolution source image. This allows us to infer the location of the viewer's current fixation on the high resolution image. Once we have established where the viewer is looking, a visual cue (subtle luminance change or a more overt flash) can then be projected onto another part of the scene to attract the viewer's attention. This can be repeated as necessary to guide the viewer's gaze over time.

We evaluated the effectiveness of our system by conducting an experiment in which we guided viewers to look at sequences of objects in a scene. On average, for a given sequence of eight objects, less than one error was reported. Furthermore, only 18 percent of the trials had more than 1 sequence error. We also found that this approach to gaze manipulation was fast as subjects typically attended to the target regions within 0.5 seconds of the onset of the visual cue.

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@acm.org](mailto:permissions@acm.org).

SAP 2013, August 22 – 23, 2013, Dublin, Ireland.  
Copyright © ACM 978-1-4503-2262-1/13/08 \$15.00

Our technique has high applicability in instructional environments such as aircraft cockpits. For example, it can be used to direct novice pilots to pay attention to situation-relevant instrumentation in response to an event such as equipment failure or adverse weather conditions or to help them learn the order of operational protocols such as takeoff and landing procedures. With our system, it is also possible to capture the fixation sequence of veteran pilots and use it to guide novices. Similar training can be done in driving simulators and systems for navigation aid.

The remainder of this paper is organized as follows: background and related work is presented in section 2, the system design is presented in section 3, the user study is described in section 4, results and discussion are presented in section 5, and the paper concludes in section 6 with a summary of the contributions and potential avenues of future research.

## 2 Background

In humans, foveal vision (center of gaze) has very high acuity when compared to peripheral vision. The falloff in visual acuity as distance from the fovea increases is directly related to the distribution of the cone photoreceptors in the retina [Osterberg 1935]. The density of cones, and hence the visual acuity, is very high in the fovea and falls off rapidly as the angle increases. The fovea itself has a diameter of 1.5mm and subtends an angle of approximately 2 degrees of the visual field. This means that at any instant, less than 0.05% of our field of view is seen in high resolution. We overcome this limitation by quickly scanning about the scene. These rapid eye movements are called saccades and the brief pauses to focus on objects within the scene are called fixations. Most saccades occur at a level below conscious awareness.

Eye-tracking provides a mechanism for monitoring where our high acuity vision gets focused in a scene or on a display. Eye-tracking systems first emerged in the late 1800s (see [Jacob and Karn 2003] for a review of the history of eye-tracking) and over the years various approaches have been used to track viewer gaze including magnetic coils embedded into contact lenses and electrooculograms (EOG) which attempt to determine eye-position by taking advantage of the voltage difference between the retina and cornea. Most modern eye-tracking systems are video-based — a video feed of the subject’s eye is analyzed to determine the center of the pupil relative to some stable feature in the video (such as the corneal reflection). A brief calibration procedure is usually necessary to establish the mapping of the subject’s eye position to locations within the scene. The accuracy of video-based eye-trackers has improved in recent years with many commercial systems reporting gaze position accuracy  $< 0.5^\circ$ . Eye-tracking systems are primarily used to collect eye movement data during psychophysical experiments. This data is typically analyzed after the completion of the experiments. However during the 1980s, the benefits of real-time analysis of eye movement data were realized as eye-trackers evolved as a channel for human-computer interaction [Levine 1981; Hutchinson et al. 1989].

Wearable eye-trackers began to emerge during the late 1990s [Pelz et al. 2000; Babcock and Pelz 2004]. Today most of the commercial eye-tracking companies offer head-mounted eye-tracking solutions. Head-mounted eye-trackers allow researchers to capture information about the visual behavior and perceptual strategies of people who were engaged in tasks outside of the laboratory. They have already been used in a wide range of settings including driving [Sodhi et al. 2002], sports [Chajka et al. 2006], geology [Evans et al. 2012], and mental health monitoring [Vidal et al. 2012].

There has been extensive research to determine what guides viewer attention. It is well known that the pattern of eye movements de-

pends on the viewer’s intent or the task assigned [Yarbus 1967; Henderson and Hollingworth 1998; Tatler et al. 2010]. Image content also plays a role. For example, it is natural for humans to be immediately drawn to faces or other informative regions of an image [Mackworth and Morandi 1967]. Additionally, research has shown that our gaze is drawn to regions of high local contrast or high edge density [Mannan et al. 1996; Parkhurst and Niebur 2003]. Researchers continue to debate whether it is salient features or contextual information that ultimately drives attention during free viewing of static images [Chen and Zelinsky 2006; Underwood and Foulsham 2006; Torralba et al. 2006; Brockmole and Henderson 2006].

Jonides [Jonides 1981] explored the differences between voluntary and involuntary attention shifts and referred to cues which trigger involuntary eye-movements as *pull cues*. Computer based techniques for providing these pull cues are often overt. These include simulating the depth-of-field effect from traditional photography to bring different areas of an image in or out of focus or directly marking up on the image to highlight areas of interest [DeCarlo and Santella 2002; Grant and Spivey 2003; Thomas and Lleras 2007; Wang and Spelke 2002; Groen and Noyes 2010]. It is also possible to direct attention in a subtle manner. For example, the Subtle Gaze Direction (SGD) technique [Bailey et al. 2009] works by briefly introducing motion cues (image space modulations) to the peripheral regions of the field of view. Since the human visual system is highly sensitive to motion, these brief modulations provide excellent pull cues. To achieve subtlety, modulations are only presented to the peripheral regions of the field of view (determined by real-time eye-tracking) and are terminated before the viewer can scrutinize them with high acuity foveal vision.

Actively guiding viewers’ attention to relevant information has been shown to improve problem solving [Jamat et al. 2008; Grant and Spivey 2003; Groen and Noyes 2010]. It has also been shown to enhance spatial learning by improving the recollection of location, size and shape of objects in images [Walther et al. 2005; Thomas and Lleras 2007; Wang and Spelke 2002; Bailey et al. 2012]. Gaze manipulation strategies have also been used for improving performance on visual search tasks by either guiding attention to previously unattended regions [Qvarfordt et al. 2010] or guiding attention directly to the relevant regions in a scene [McNamara et al. 2009; McNamara et al. 2012].

Gaze manipulation has high applicability in instructional environments. Numerous studies have been conducted to understand experts’ eye movements for specific tasks and to use their fixation sequence to guide novices during training. This has been done in tasks such as aircraft inspection [Sadasivan et al. 2005] and optic disc examinations [O’Neill et al. 2011]. An interesting application of this approach is in the area of medical image analysis to understand the cognitive strategies employed by experts [Vaidyanathan et al. 2011; Li et al. 2012]. Litchfield et al. [2010] [2011] showed that viewing an expert’s eye movements can help to improve identification of pulmonary nodules in chest x-rays. Gaze manipulation has also been used to subtly guide novices along the scanpath of an expert radiologist as they try to identify abnormalities in mammograms [Sridharan et al. 2012]. Results of that study reveal that novices who were guided performed significantly better than the control group (no gaze manipulation). They also continued to perform better once the training was complete and gaze manipulation was disabled.

Motivated by these observations, the goal of this work is to extend the concept of gaze manipulation beyond digital imagery to include real world environments. The concept of guiding attention (pointing) in physical spaces is well established in art, photography theatre and cinema. In most cases this relies on clever use of contrast and composition, direct manipulation of lights, or manual

post-processing to highlight regions or characters of interest. Several augmented reality systems have also been developed to directly highlight real-world objects [Wellner and Freeman 1993; Raskar et al. 2000; Gurevich et al. 2012] or to highlight objects on a video feed of the real world [Chastine et al. 2008; Gauglitz et al. 2012]. However, none of these systems take the viewer’s gaze into account or attempt to guide attention without using overt cues. One notable exception is the work by Veas et al. [2011] where they pre-process video feeds to increase the saliency of target regions while reducing the saliency of surrounding regions. They do this by adjusting the contrast of the various channels in CIELAB space to create a natural looking change to the image which increases the likelihood of the target regions being attended to. In their work, eye-tracking was only used to test if the subjects actually paid attention to target regions. Furthermore, their technique does not actively try to shift attention between multiple targets. In contrast, we adopt an approach similar to the Subtle Gaze Direction technique, where eye tracking is used to monitor where the viewer is attending to in the scene in real-time. A visual cue is then projected onto another part of the scene to draw the viewer’s attention. This process is repeated to guide the viewer’s gaze about the scene.

### 3 System Design

As discussed earlier, the two main challenges to guiding attention in real world environments are (1) determining what the viewer is currently paying attention to, and (2) projecting visual cues on other parts of the scene to draw the viewer’s attention. Our system combines the use of eye-tracking glasses, a standard projector, and a high-resolution source image of the scene being viewed (taken approximately from the projector’s perspective). The eye-tracking system computes the fixation position within the scene that the viewer is attending to. SIFT descriptors are then extracted in the regions surrounding the fixation point in the scene image. These are compared against pre-computed SIFT descriptors from the source image. The output from this matching algorithm allows us to infer the location of the viewer’s fixation on the source image. This is an important step in our process as it contributes to the subtlety of technique by ensuring that we do not attempt to guide the viewer to objects they are already attending to. Once the current fixation is determined, viewer attention can then be guided by projecting a visual cue onto another part of the scene.

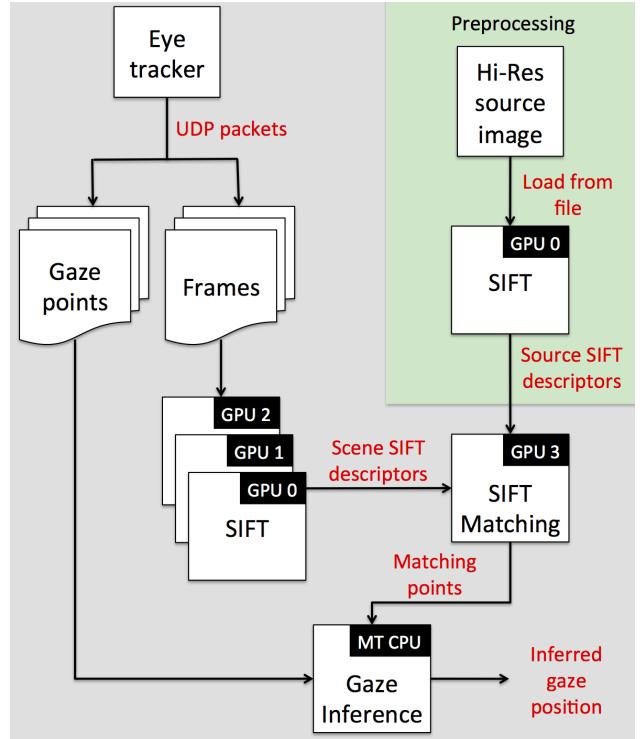
We established the following design goals:

- The system should perform in real-time.
- The system should allow for unrestricted movement of the viewer as long as their fixations remain within the region covered by the source image.
- The system must be robust even under ill-conditioned and ill-posed situations such as track-loss or when the viewer looks away from the scene.

We make the following assumptions:

- The system will be used in a visually stable environment (i.e. unchanging external lighting and no moving objects).
- The objects in the scene are non-transparent and well defined in terms of color, contrast, size, shape and texture.

These assumptions are necessary to ensure a high correlation between the SIFT descriptors from the eye tracker scene camera and the SIFT descriptors from the source image. To accomplish all of this in real-time, our system performs image matching on a frame-by-frame basis using multiple GPUs.



**Figure 2:** System architecture and data flow.

#### 3.1 System Overview

Figure 2 shows the architecture for our real-world gaze manipulation framework. The eye tracker used in our system is a Senso-Motoric Instruments (SMI) Eye Tracking Glasses. The glasses are equipped with a front facing camera (scene camera) that captures the scene that the viewer is looking at as well as two rear facing cameras that capture binocular eye movements (eye cameras). The resolution of the scene camera is 1280 x 960 @ 24p and the field of view is 60° horizontal and 46° vertical. Gaze position within the scene is computed using SMI’s proprietary eye-tracking software at a rate of 30 Hz. The eye tracker is connected to a dedicated laptop which computes the viewer’s fixations on the scene camera frames in real time. The laptop is powered by a dual core processor with 4 GB RAM. The compressed video and fixation data are streamed over the network to a desktop for further processing. The desktop is equipped with 24 GB RAM, dual hexa-core processors and dual GTX 590 graphics cards. A standard RGB projector completes the hardware for our system.

SIFT descriptors of the source image are computed ahead of time using one of the GPU cores. The eye tracker system streams gaze positions and scene camera frames in real time over the network through a UDP socket. These gaze points and the compressed scene frames are added to a FIFO queue. Three GPUs are used to process this information to compute the SIFT descriptors. A dedicated GPU core performs SIFT descriptor matching between the scene and the source image. To reduce computation time and improve the likelihood of solution convergence, only a subset of the SIFT matches are extracted around the current fixation point. The resulting matching points are used by a multi-threaded CPU program to infer where the viewer is attending to in the source image. This distribution of work between the GPUs and CPU allows our system to run in real-time.

Our system was developed in C++ and utilizes a variety of libraries

for specific tasks including SMI’s proprietary API for interfacing with the eye tracking glasses, SiftGPU [Wu 2007], the Eigen library [Guennebaud et al. 2010] for linear algebra and related algorithms, and finally SFML for data visualization. Underlying technologies include CUDA and OpenGL for SiftGPU and OpenMP for the Eigen library.

### 3.1.1 SIFT descriptor extraction and matching

Sparse SIFT descriptors are computed for both the scene and source images using SiftGPU. The thresholding values are set manually to obtain optimum results for the given scene. This only has to be done once since we are dealing with controlled environments where the thresholds will not change from frame-to-frame.

Once descriptors for the current frame have been identified, they are matched with the precomputed descriptors from the source image as shown in Figure 3. To filter the matching results, we perform a radial search about the fixation point until 5% of the total number of descriptors in the source image have been found. We limit this as it provides sufficient information to infer the fixation position while not incurring any additional computational costs. To eliminate incorrectly matched points from this subset of descriptors we use Cook’s Distance measure [Cook 1977] which performs dimensionless weighting of all values in a set, with respect to the mean of that set. In our case, we expect that the positional average of the subset of descriptors obtained in the scene image will be close to the fixation point (in the Cartesian sense). This means that the matched points with higher Cook’s Distance values are more likely to be outliers. Once the outliers have been eliminated, we use a least squares solver to infer the fixation on the source image.

### 3.1.2 Stability Considerations

Occasionally, situations may arise that lead to inaccurate results or an inability to compute the fixation position on the source image. The most common is a track-loss (i.e. the eye-tracker is not able to detect the pupil due to blinks or extreme pupil position). In this case the scene video frames will have no accompanying gaze information making fixation inference impossible. When this occurs we simply discard the frames and resume once gaze positions are again available. Also, if the viewer happens to look away from the scene, then the inferred gaze position will naturally be incorrect. We detect when this is happening by keeping track of the number of matched descriptors from frame-to-frame. If the number changes drastically (indicating dissimilar scenes) then we ignore the results from the SIFT descriptor matching algorithm. Finally, it is possible (though highly improbable) for the least squares solver not to converge. This can only happen in cases where there are fewer than two descriptors present (i.e. fixation occurs in a large uniform region of the scene) or if all descriptors are co-linear. In these cases we simply reuse the previous frame’s result. By addressing all of these ill-posed and ill-conditioned situations, we ensure that the system is both accurate and stable.

## 4 User Study

The goal of the user study was to test the effectiveness of our real-world gaze manipulation technique. Participants viewed a simple scene consisting of eight objects. The intended viewing order was not prescribed. After viewing the scene for a short period of time, we then attempted to guide their attention through six sequences of objects in the scene. The relevant objects were highlighted by projecting a brief luminance modulation using the projector.

The modulations were constructed by alternately blending some amount of black, then some amount of white. The rate at which

the blend is modulated is 10Hz. A Gaussian falloff is used to soften the edges of the modulated regions. In our viewing configuration, the modulation diameter on the physical objects ranged from approximately 2 centimeters to 4 centimeters (depending on distance from the projector).

Figure 3 (top left) shows the source image of the scene used in our user study. The objects in the scene were non-transparent and well defined in terms of color, contrast, size, shape and texture. The objects were highlighted in a randomized order to minimize the introduction of learning effects.

Twenty participants (16 males, 4 females) between the ages of 18 and 29 volunteered for the user study. They all had normal, or corrected-to-normal, vision and were naive to the purpose of the experiment. They were simply instructed to look at the scene. Each participant underwent a brief calibration procedure, away from the scene, to ensure proper eye-tracking. Six randomly generated viewing sequences were presented to each participant with a 10 second gap between sequences. Each sequence consisted of 8 objects. Between sequences the participants were told to reposition themselves however they saw fit but to keep the scene visible. This ensured that the data collected across all subjects covered a wide range of vantage points. The entire experiment (including calibration) for each participant was less than 10 minutes. We recorded the viewer’s gaze positions within the scene as well as the individual scene camera frames during the experiment. Data from one participant was excluded due to an extended period of track loss.

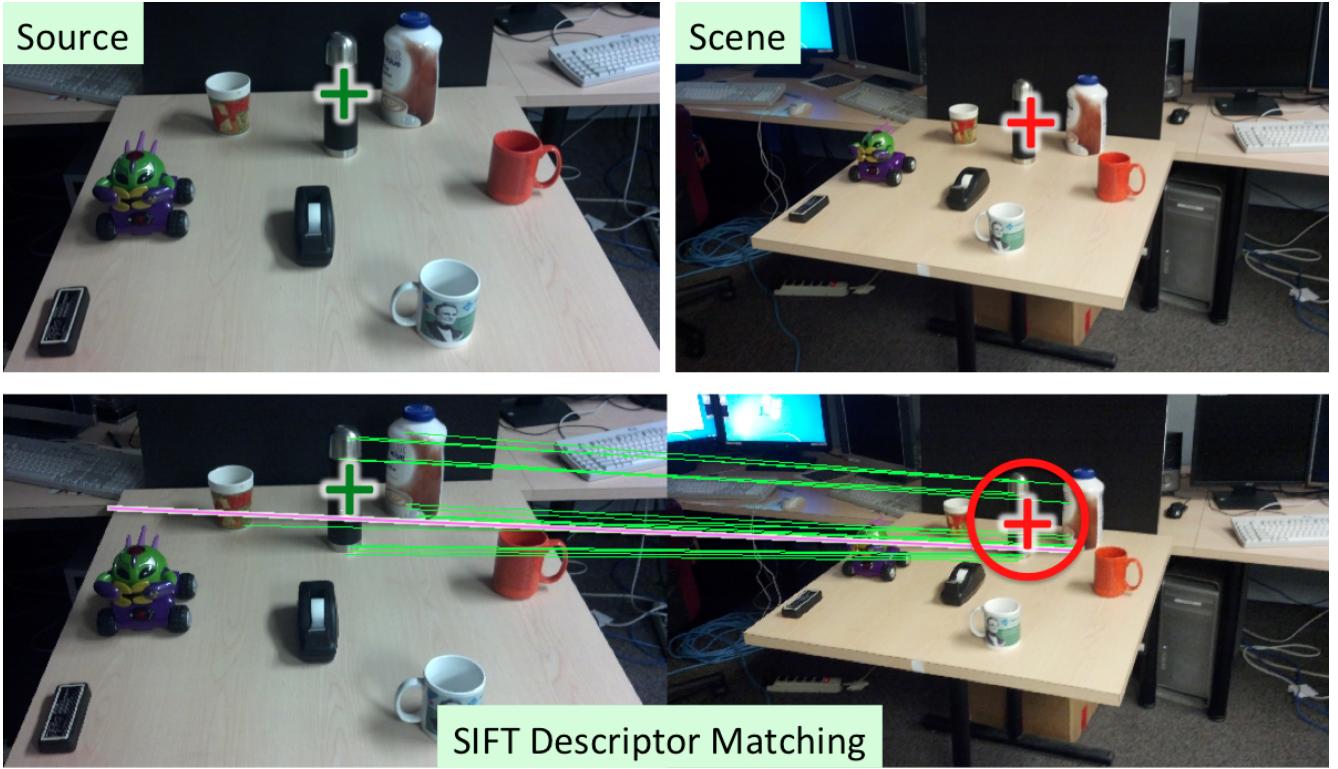
## 5 Results

We needed a robust mechanism to compare the intended viewing sequence with the actual viewing sequence of each participant. The actual viewing sequence is extracted from the eye-tracking data by identifying the first fixation that occurs after the onset of the visual cue. Levenshtein distance [Levenshtein 1965] provides an appropriate measure to compare distances between ordered sequences, such as those recorded during our experiment. We assign labels A through H to the eight objects in the scene as shown in Figure 6 (inset) and compare the intended viewing sequence with the actual viewing sequence. Suppose for the eight objects in the scene that the correct viewing order is [ABCDEFGH]. A Levenshtein distance of 0 would indicate no difference, whereas a distance of 8 would indicate the maximum difference.

Figure 4 shows the average and standard deviation of the Levenshtein distance for the six sequences for all participants. The similar averages and standard deviation values indicate that participant performance remains consistent over time. The average of the Levenshtein distance across all sequences for all participants is 0.85 (recall that the Levenshtein distance measure for this study ranges from 0 to 8). This means that we can typically expect less than one error per sequence. The histogram in Figure 5 shows the distribution of Levenshtein distance for all sequences. Notice that it is skewed to the left indicating that a large number distance measures are close to zero. Only 18% of the trials had more than one sequence error and the error rate falls off rapidly.

We also observed that the response time between the onset of the modulation and the subsequent fixation on the object was approximately 0.5 seconds. This is consistent with what was observed in the Subtle Gaze Direction paper for digital images.

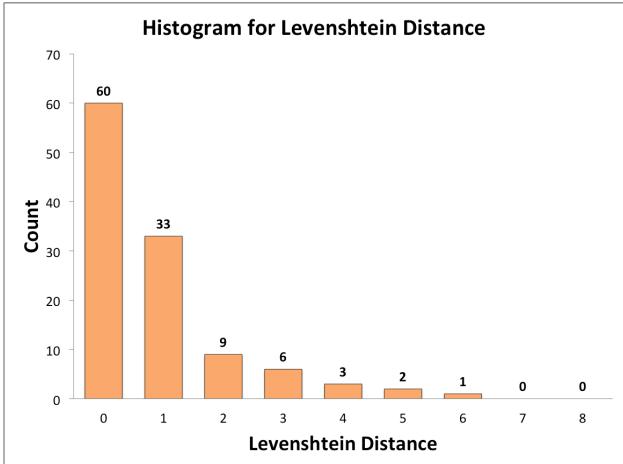
Overall these results indicate that our technique is indeed effective at guiding attention in simple controlled real-world environments. Figure 6 shows a representative frame from the scene camera video that was captured from one subject for one of the sequences. It is overlaid with the subject’s scanpath. The numbered red circles indi-



**Figure 3:** SIFT descriptor extraction and matching and fixation inference. Source image (top left), scene image (top right), and 5% of the resulting descriptor matches (bottom). Correct matches are shown in green and an incorrect match (outlier) is highlighted in magenta; current fixation is shown as a red cross; inferred fixation is shown as a green cross; the red circle shows the area in which the SIFT descriptors are matched.

	Sequence 1	Sequence 2	Sequence 3	Sequence 4	Sequence 5	Sequence 6	All Sequences
Average	0.58	0.53	0.89	1.11	1.05	1.05	<b>0.85</b>
Standard deviation	1.12	0.90	1.15	1.70	1.51	1.08	<b>1.24</b>

**Figure 4:** Summary of Levenshtein distances across all sequences for all participants.



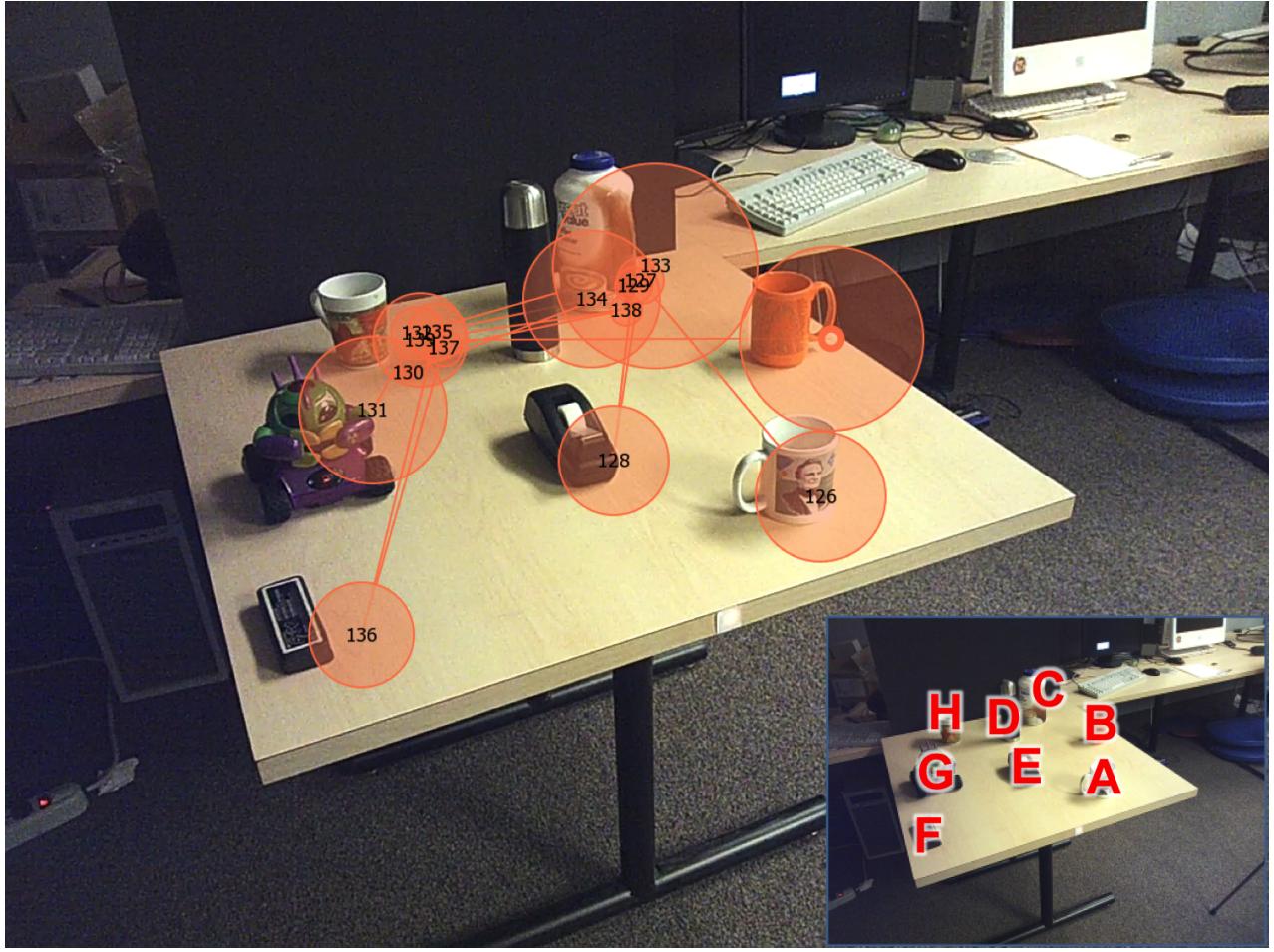
**Figure 5:** Histogram of Levenshtein distances across all sequences for all participants.

cate the order of the fixations that occurred during the presentation of the sequence and the size of the circle indicates the duration of the fixation. In this particular example there was a perfect match between the target sequence and the actual sequence.

To demonstrate the usefulness of real-world gaze manipulation, we used our framework in a parts retrieval experiment. Parts from a toy building set were arranged on a table in piles (see Figure 7) and we asked six subjects to retrieve 10 parts. Three subjects were given a sheet of paper with photos of the 10 parts (control) and three subjects were told that we were going to attempt to guide them to the correct parts. The results of this experiment are shown in Figure 7. As expected, subjects in the control group had to shift their attention back and forth between the printed sheet and the parts on the table. They also had to develop a strategy for keeping track of which parts were already found. This resulted in longer completion times and greater likelihood of error. On the other hand, the gaze guided group performed much better in terms of completion time and error rate due to reduced cognitive load.

## 6 Conclusions & Future Work

The work presented in this paper extends the Subtle Gaze Direction technique to guide viewer attention in controlled real-world environments. By projecting subtle luminance modulations into the physical world, our system is capable of drawing attention to a target region very quickly (response time  $\approx 0.5$  seconds) and the process can be repeated on other parts of the scene to guide the viewer attention over time. Results of a user study reveal that our approach effectively guides viewers through sequences of objects with less than one error per sequence consisting of eight objects.



**Figure 6:** Representative images showing the real-world gaze-guided scanpath of a single subject. The numbered red circles indicate the order of the fixations that occurred during the presentation of the sequence and the size of the circle indicates the duration of the fixation. Inset show the labels assigned to the various objects in the scene.

	Control			Gaze Guided		
	Subject 1	Subject 2	Subject 3	Subject 1	Subject 2	Subject 3
Completion Time (min:sec)	1:04	0:37	2:54	0:30	0:24	0:25
Errors	1	0	2	0	0	0

**Figure 7:** Performance of subjects on a parts retrieval task. Photograph shows initial arrangement of parts on the table.

Furthermore, the likelihood of more than one error is only 18%.

The main limitation of our setup is restricted mobility due to the dependence on a fixed projector. It would be better to present the visual cues on a head-mounted display, however, commercial eye-trackers with head-mounted displays for augmented reality applications are not yet commonplace. Note that other researchers have attempted to track the eyes from below/above the frame of head-mounted displays with limited success. Transitioning to an integrated system would be simple and will greatly improve mobility and also be less distracting to bystanders.

An obvious next-step for our work is to experiment with more

complex and dynamic environments. We will be able to leverage recently published work from the eye-tracking community which documents ‘best practices’ for collecting eye-tracking data in outdoor environments [Pelz et al. 2011; Evans et al. 2012] as well as work from computer vision and augmented-reality on object detection and tracking [Wagner et al. 2009; Klein and Drummond 2003].

Successfully directing an observer’s gaze in the real world has many applications including:

- **Training:** Beyond the cockpit and automobile console examples described earlier, directing gaze in the real world could benefit training in a multitude of ways. For example, gaze manipulation can be used to assist first responders in learning the spatial layout of real buildings including the location of fire exits and AED stations.
- **Tourism:** Tour guides are employed at attractions, parks, museums and galleries around the world. Their main function is to point out interesting artifacts or landmarks in their specific environment. In situations where guides are not available, real-world gaze manipulation could be used to highlight interesting objects on a self-guided tour.
- **Advertising** An obvious domain is advertising. Product placement is an elegant science with companies competing

for limited shelf space for their products. Employing real-world gaze manipulation techniques could prove valuable to companies hoping to emphasize their products.

In summary we explored the question of guiding a viewers' attention in a real world setting. Our technique extends the Subtle Gaze Direction paradigm to guide attention to physical objects. Our system provides an exciting foundation for future work.

## Acknowledgments

This material is based 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

- BABCOCK, J. S., AND PELZ, J. B. 2004. Building a lightweight eyetracking headgear. In *Proceedings of the 2004 Symposium on Eye Tracking Research and Applications*, ACM, New York, NY, USA, ETRA '04, 109–114.
- BAILEY, R., McNAMARA, A., SUDARSANAM, N., AND GRIMM, C. 2009. Subtle gaze direction. *ACM Trans. Graph.* 28 (September), 100:1–100:14.
- BAILEY, R., McNAMARA, A., COSTELLO, A., SRIDHARAN, S., AND GRIMM, C. 2012. Impact of subtle gaze direction on short-term spatial information recall. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ACM, New York, NY, USA, ETRA '12, 67–74.
- BROCKMOLE, J. R., AND HENDERSON, J. M. 2006. Recognition and attention guidance during contextual cueing in real-world scenes: evidence from eye movements. *Q J Exp Psychol (Hove)* 59, 7 (Jul), 1177–87.
- CHAJKA, K., HAYHOE, M., SULLIVAN, B., PELZ, J., MENNIE, N., AND DROLL, J. 2006. Predictive eye movements in squash. *Journal of Vision* 6, 6, 481.
- CHASTINE, J., NAGEL, K., ZHU, Y., AND HUDAČEK-BUSWELL, M. 2008. Studies on the effectiveness of virtual pointers in collaborative augmented reality. In *Proceedings of the 2008 IEEE Symposium on 3D User Interfaces*, IEEE Computer Society, Washington, DC, USA, 3DUI '08, 117–124.
- CHEN, X., AND ZELINSKY, G. J. 2006. Real-world visual search is dominated by top-down guidance. *Vision research* 46, 24 (Nov.), 4118–4133.
- COOK, R. D. 1977. Detection of Influential Observation in Linear Regression. *Technometrics* 19, 1, 15–18.
- DECARLO, D., AND SANTELLA, A. 2002. Stylization and abstraction of photographs. In *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, ACM Press, New York, NY, USA, 769–776.
- EVANS, K., JACOBS, R., TARDUNO, J., AND PELZ, J. 2012. Collecting and analyzing mobile eye-tracking data in outdoor environments. *Journal of Eye Movement Research* 5(2):6, 1–19.
- GAUGLITZ, S., LEE, C., TURK, M., AND HÖLLERER, T. 2012. Integrating the physical environment into mobile remote collaboration. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, ACM, New York, NY, USA, MobileHCI '12, 241–250.
- GRANT, E., AND SPIVEY, M. J. 2003. Eye movements and problem solving: guiding attention guides thought. *Psychological Science* 14, 5, 462–466.
- GROEN, M., AND NOYES, J. 2010. Solving problems: How can guidance concerning task-relevancy be provided? *Comput. Hum. Behav.* 26 (November), 1318–1326.
- GUENNEBAUD, G., JACOB, B., ET AL., 2010. Eigen v3. <http://eigen.tuxfamily.org>.
- GUREVICH, P., LANIR, J., COHEN, B., AND STONE, R. 2012. Teleadvisor: a versatile augmented reality tool for remote assistance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, New York, NY, USA, CHI '12, 619–622.
- HENDERSON, J. M., AND HOLLINGWORTH, A. 1998. Eye movements during scene viewing: An overview. In *Eye Guidance in Reading and Scene Perception*, G. Underwood, Ed. Oxford: Elsevier, 269–293.
- HUTCHINSON, T. E., WHITE, K. P., MARTIN, W. N., REICHERT, K. C., AND FREY, L. A. 1989. Human-computer interaction using eye-gaze input. *Systems, Man and Cybernetics, IEEE Transactions on* 19, 6, 1527–1534.
- JACOB, R. J. K., AND KARN, K. S. 2003. *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*. Elsevier Science, Amsterdam, ch. Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises (Section Commentary), 573–605.
- JAMET, E., GAVOTA, M., AND QUAIREAU, C. 2008. Attention guiding in multimedia learning. *Learning and Instruction* 18, 2, 135 – 145.
- JONIDES, J. 1981. *Voluntary versus automatic control over the mind? s eye's movement*, vol. 9. Erlbaum, 187–203.
- KLEIN, G., AND DRUMMOND, T. 2003. Robust visual tracking for non-instrumented augmented reality. In *Proc. Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'03)*, 113–122.
- LEVENSHTEIN, V. 1965. Binary codes capable of correcting spurious insertions and deletions of ones. *Problems of Information Transmission* 1, 8–17.
- LEVINE, J. L. 1981. An eye-controlled computer. Research Report RC-8857, IBM Thomas J. Watson Research Center, Yorktown Heights, N.Y.
- LI, R., PELZ, J., SHI, P., ALM, C. O., AND HAAKE, A. R. 2012. Learning eye movement patterns for characterization of perceptual expertise. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ACM, New York, NY, USA, ETRA '12, 393–396.
- LITCHFIELD, D., AND BALL, L. J. 2011. Using another's gaze as an explicit aid to insight problem solving. *Q J Exp Psychol (Hove)* 64 (Apr), 649–656.
- LITCHFIELD, D., BALL, L. J., DONOVAN, T., MANNING, D. J., AND CRAWFORD, T. 2010. Viewing another person's eye movements improves identification of pulmonary nodules in chest x-ray inspection. *J Exp Psychol Appl* 16 (Sep), 251–262.
- MACKWORTH, N. H., AND MORANDI, A. J. 1967. The gaze selects informative details within pictures. *Perception and Psychophysics* 2, 547–552.

- MANNAN, S. K., RUDDOCK, K. H., AND WOODING, D. S. 1996. The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision* 10, 165–188.
- MCNAMARA, A., BAILEY, R., AND GRIMM, C. 2009. Search task performance using subtle gaze direction with the presence of distractions. *ACM Trans. Appl. Percept.* 6 (September), 17:1–17:19.
- MCNAMARA, A., BOOTH, T., SRIDHARAN, S., CAFFEY, S., GRIMM, C., AND BAILEY, R. 2012. Directing gaze in narrative art. In *Proceedings of the ACM Symposium on Applied Perception*, ACM, New York, NY, USA, SAP '12, 63–70.
- O'NEILL, E. C., KONG, Y. X. G., CONNELL, P. P., ONG, D. N., HAYMES, S. A., COOTE, M. A., AND CROWSTON, J. G. 2011. Gaze behavior among experts and trainees during optic disc examination: does how we look affect what we see? *Invest Ophthalmol Vis Sci* 52, 7 (Jun), 3976–83.
- OSTERBERG, G. 1935. Topography of the layer of rods and cones in the human retina. *Acta Ophthalm. suppl.* 6, 11–97.
- PARKHURST, D., AND NIEBUR, E. 2003. Scene content selected by active vision. *Spatial Vision* 16, 125–154.
- PELZ, J., CANOSA, R., KUCHARCZYK, D., BABCOCK, J., SILVER, A., AND KONNO, D. 2000. Portable eyetracking: A study of natural eye movements. In *Human Vision and Electronic Imaging V, SPIE Proceedings*, 3659.
- PELZ, J., KINSMAN, T., AND EVANS, K. 2011. Analyzing complex gaze behavior in the natural world. *SPIE-IS&T Human Vision and Electronic Imaging XVI*, 1–11.
- QVARFORDT, P., BIEHL, J. T., GOLOVCHINSKY, G., AND DUNNINGAN, T. 2010. Understanding the benefits of gaze enhanced visual search. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ACM, New York, NY, USA, ETRA '10, 283–290.
- RASKAR, R., LOW, K., AND WELCH, G. 2000. Shader lamps: Animating real objects with image-based illumination. Tech. rep., Chapel Hill, NC, USA.
- SADASIVAN, S., GREENSTEIN, J. S., GRAMOPADHYE, A. K., AND DUCHOWSKI, A. T. 2005. Use of eye movements as feed-forward training for a synthetic aircraft inspection task. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, New York, NY, USA, CHI '05, 141–149.
- SODHI, M., REIMER, B., COHEN, J. L., VASTENBURG, E., KAARS, R., AND KIRSCHENBAUM, S. 2002. On-road driver eye movement tracking using head-mounted devices. In *Proceedings of the 2002 symposium on Eye tracking research and applications*, ACM, New York, NY, USA, ETRA '02, 61–68.
- SRIDHARAN, S., BAILEY, R., MCNAMARA, A., AND GRIMM, C. 2012. Subtle gaze manipulation for improved mammography training. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ACM, New York, NY, USA, ETRA '12, 75–82.
- TATLER, B. W., WADE, N. J., KWAN, H., FINDLAY, J. M., AND VELICHKOVSKY, B. M. 2010. Yarbus, eye movements, and vision. *i-Perception* 1.
- THOMAS, L., AND LLERAS, A. 2007. Moving eyes and moving thought: on the spatial compatibility between eye movements and cognition. *Psychonomic bulletin and review* 14, 4, 663–668.
- TORRALBA, A., OLIVA, A., CASTELHANO, M. S., AND HENDERSON, J. M. 2006. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological review* 113, 4 (Oct.), 766–786.
- UNDERWOOD, G., AND FOULSHAM, T. 2006. Visual saliency and semantic incongruity influence eye movements when inspecting pictures. *Q J Exp Psychol (Hove)* 59, 11 (Nov), 1931–49.
- VAIDYANATHAN, P., PELZ, J., LI, R., MULPURU, S., WANG, D., SHI, P., CALVELLI, C., AND HAAKE, A. 2011. Using human experts' gaze data to evaluate image processing algorithms. In *IVMSP Workshop, 2011 IEEE 10th*, 129–134.
- 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 SIGCHI Conference on Human Factors in Computing Systems*, ACM, New York, NY, USA, CHI '11, 1471–1480.
- VIDAL, M., TURNER, J., BULLING, A., AND GELLERSEN, H. 2012. Wearable eye tracking for mental health monitoring. *Comput. Commun.* 35, 11 (June), 1306–1311.
- WAGNER, D., SCHMALSTIEG, D., AND BISCHOF, H. 2009. Multiple target detection and tracking with guaranteed framerates on mobile phones. In *Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality*, IEEE Computer Society, Washington, DC, USA, ISMAR '09, 57–64.
- WALTHER, D., RUTISHAUSER, U., KOCH, C., AND PERONA, P. 2005. Selective visual attention enables learning and recognition of multiple objects in cluttered scenes. *Comput. Vis. Image Underst.* 100, 1–2 (Oct.), 41–63.
- WANG, R. F., AND SPELKE, E. S. 2002. Human spatial representation: insights from animals. *Trends in Cognitive Sciences* 6, 9, 376 – 382.
- WELLNER, P., AND FREEMAN, S. 1993. The doubledigitaldesk: Shared editing of paper documents. Tech. Rep. Tech. Rep. EPC-93-108, Xerox EuroPARC.
- WU, C. 2007. SiftGPU: A GPU implementation of scale invariant feature transform (SIFT). <http://cs.unc.edu/~ccwu/siftgpu>.
- YARBUS, A. L. 1967. *Eye Movements and Vision*. Plenum. New York.