

Comparing Mouse and MAGIC Pointing for Moving Target Acquisition

Jutta Hild*, Dennis Gill, Jürgen Beyerer*[†]

*Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB), Karlsruhe, Germany

[†]Vision and Fusion Laboratory, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

{jutta.hild; juergen.beyerer}@iosb.fraunhofer.de, dennis.gill@student.kit.edu

Abstract

Moving target acquisition is a challenging and manually stressful task if performed using an all-manual, pointer-based interaction technique like mouse interaction, especially if targets are small, move fast, and are visible on screen only for a limited time. The MAGIC pointing interaction approach combines the precision of manual, pointer-based interaction with the speed and little manual stress of eye pointing. In this contribution, a pilot study with twelve participants on moving target acquisition is presented using an abstract experimental task derived from a video analysis scenario. Mouse input, conservative MAGIC pointing and MAGIC button are compared considering acquisition time, error rate, and user satisfaction. Although none of the participants had used MAGIC pointing before, eight participants voted for MAGIC button being their favorite technique; participants performed with only slightly higher mean acquisition time and error rate than with the familiar mouse input. Conservative MAGIC pointing was preferred by three participants; however, mean acquisition time and error rate were significantly worse than with mouse input.

CR Categories: H.5.2 [Information interfaces and presentation]: User Interfaces — Input devices and strategies.

Keywords: moving target acquisition, input device, eye gaze interaction, MAGIC pointing, mouse, pilot study, video analysis

1 Introduction

Target acquisition on the graphical user interface of a desktop computer requires the user to perform a selection operation. Usually, targets are stationary (e.g., windows, icons). However, various applications also require the selection of moving targets, e.g., computer games, simulations or air traffic control systems [Hasan et al. 2011]. The background of this contribution is expert real-time video analysis. One of the main tasks in this domain is to detect targets or events within the video images and to report them immediately. Video exploitation systems more and more provide automated target recognition algorithms; however, proper target classification often demands the context knowledge of a human operator. Hence, real-time video analysis still requires a human in the loop to monitor the video and algo-

rithm outputs continuously and to select moving targets for subordinate analysis. Due to the visually rich and dynamic input and the constantly required attention, often over hours, it is an exhaustive task.

Today, user interfaces of video exploitation systems often feature double or triple monitor setups displaying the video output with adequate size for detailed analysis, and mouse input for target selection. However, using an all-manual interaction technique to acquire moving targets of a size down to about 30 or 40 pixels is challenging and manually stressful. Moving the pointer on the target can be time consuming and exhausting, especially if the operator is not aware of the current pointer position or if the pointer has to be relocated between monitors. In both cases, operators have to move their visual attention onto the pointer and off the target. Target acquisition might fail if a target moves fast and is visible on screen only for a limited time. Summarized, there is a need for an interaction technique which should support the operator in (1) keeping the visual focus of attention continuously on the video output, (2) reduce target acquisition time, and (3) reduce manual load. Proposed methods to make moving target acquisition using mouse input easier include pausing motion on the display during the target acquisition process, enhancing target or pointer, or using visual feedback [Hasan et al. 2011][Al Hajri et al. 2011]. However, to pause the video is not permitted in real-time video analysis, enhancement or visual feedback of the targets require information about target properties (size, position, speed) which are usually not available, and enhancement of the pointer lacks precision in a dense and dynamically changing field of targets. Besides, these methods address only aspects (2) and (3).

We propose to use eye gaze interaction based on eye tracking technology for moving target acquisition. An eye tracker continuously provides an estimation of the user's gaze position on the screen which can be utilized for pointing at targets. As the gaze pointer is always located at the user's visual focus of attention, gaze pointing requires neither pointer search nor manual pointer repositioning. Hence, the operator can monitor the video output without interruption. Various gaze interaction techniques have been introduced by a large number of contributions, see e.g. [Ware and Mikaelian 1987][Jacob 1991][Zhai et al. 1999]. All employ the user's gaze position for pointing, but differ in the way target acquisition is actuated (the equivalent to the mouse click). Unimodal gaze interaction techniques actuate by a certain fixation dwell-time on a target or by a prolonged eye blink. Multimodal gaze interaction techniques use a different input modality for actuation, e.g. a key press or a mouse. For the dwell-time and key press techniques authors reported reduced acquisition time and manual stress for stationary targets. However, due to eye tracker measurement uncertainty, pointing accuracy is inherently lower as with mouse input. Today's eye

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.
ETRA 2014, March 26 – 28, 2014, Safety Harbor, Florida, USA.
Copyright © ACM 978-1-4503-2751-0/14/03 \$15.00

tracking devices provide measurement uncertainty not better than 0.5° of visual angle. At a typical distance of 60 cm between user and monitor this corresponds to a radius of uncertainty of about 1 cm for the estimated gaze position on screen. Isokoski et al. [2009] investigated gaze-based moving target acquisition in computer games and report that the combination of gaze pointing (crosshair control) and mouse click (triggering weapon fire) could not beat mouse input for a first-person-shooter game, as gaze pointing provided too low accuracy for fine targeting. For a chicken-shoot game, “after four to five trials, most subjects outperformed the mouse and keyboard control condition in the final score”, but no information is provided about target size, target hits or misses. Hild et al. [2013] investigated moving target acquisition using a simulation of targets (visible size = 24 pixels, selectable size = 4° visual angle) moving over the screen. Mean target acquisition time was significantly lower for gaze+key press than for mouse input; mean acquisition error rate was similar for the two.

Zhai et al. [1999] introduced the multimodal gaze-enhanced interaction “MAGIC pointing” aiming to combine the high speed of gaze pointing and the high accuracy of manual target acquisition. The technique makes use of the estimated gaze position only for coarse pointing while fine pointing and actuation of target acquisition are done by manual input using, e.g., a track point, or a mouse. Pointer position is dynamically redefined based on the current gaze position. The liberal MAGIC variant displays the pointer at any new object the user looks at, *new* being defined by sufficient distance, e.g., 120 pixels, from the current pointer position. The conservative variant warps the pointer next to the estimated gaze position only if the user is actuating the manual device. Comparing liberal and conservative MAGIC with mouse input for stationary target acquisition, the authors report rather similar completion times, being shortest for liberal MAGIC, followed by mouse input, and being the longest for conservative MAGIC. Acquisition error rates are similar with the three techniques. Drewes and Schmidt [2009] introduced MAGIC touch, using a touch-sensitive left mouse button to actuate pointer warping. They report speed improvements for pointing tasks on complex backgrounds. On plain backgrounds, performances with mouse and MAGIC touch were similar, “but users perceived the gaze-sensitive interaction of the touch-sensitive mouse as being faster and more convenient”. Fares et al. [2013] introduced Animated MAGIC, a variant similar to conservative MAGIC, but the pointer warps pixel by pixel to a new gaze position. In a Fitts’ Law experiment, their technique outperformed mouse input in terms of throughput. Hild et al. [2013] investigated in their moving target acquisition task liberal MAGIC pointing, too. Mean target acquisition time was significantly higher than with gaze + key press, mean acquisition error rate was significantly worse for liberal MAGIC pointing compared to both mouse and gaze+key press. The authors attribute this to its complexity which would require longer practice time. But they state, that at least some participants were able to perform fast and accurate using this technique.

Given the size of the targets from about 1° visual angle in the shooting games (roughly estimated from the figures in Isokoski et al. [2009]) to 4° in Hild et al. [2013], it is doubtful that target acquisition using gaze for fine pointing will provide the required accuracy to reliably select targets of 30 to 40 pixels in a video. On the other hand, MAGIC pointing has been reported several times to outperform mouse input for stationary target acquisition. We therefore conducted a pilot study to elaborate whether one of the MAGIC pointing variants would provide a promising alternative to mouse input for moving target acquisition.

2 Methodology

2.1 Interaction Techniques

In the pilot study, three interaction techniques were compared: mouse input, conservative MAGIC pointing, and the self-designed MAGIC button. Mouse input worked in the traditional way of pointing and left-button click. As the standard pointing and selection device in state of the art video exploitation is the computer mouse, we decided to use a mouse as manual pointing device also for the MAGIC variants.

Displaying the pointer constantly at any new object the user looks at could be distracting. Hence, the liberal MAGIC approach is not appropriate for real-time video analysis lasting over several hours. Instead, we implemented the conservative MAGIC approach, where pointer warping happens only on manual demand of the user, modifying the original by Zhai et al. [1999] in two ways. We did not implement the *intelligent offset* introduced to minimize the directional uncertainty after the pointer appears, but place the pointer at the estimated gaze position, i.e., as near to the target as possible. In addition, we implemented two thresholds to limit over-active pointer behavior which can occur if the user accidentally moves the mouse. Firstly, the pointer is only relocated if the last mouse movement exceeds a time-threshold of 25ms. Secondly, a gaze position is only considered being *new* if the distance to the preceding gaze position exceeds a predefined distance-threshold of 150 pixels; by this means, pointer warping every 25 ms is suppressed, if users are not able to move the mouse without any interruption. In case the last detected eye movement was a saccade, the distance-threshold is set to 20 pixels, as by high chance now a small eye movement will follow to actually land the gaze position on the wanted target. Thus, the system repositions the pointer near the target once more, quite close to the estimated gaze position. In case the last detected eye movement represents a fixation, the chance of a saccade following is higher, and the distance-threshold is set to 150 pixels again.

As a new approach, we designed the MAGIC button technique, inspired by MAGIC touch. Pointer replacement is now initiated by a right mouse button click. As users sometimes might move the mouse involuntarily by habit, conservative MAGIC could produce unwanted pointer warping. In contrast, to actuate warping, MAGIC button requires a distinct action which is not part of the internalized point-and-left-button-click-procedure of mouse input, and allows the users to exactly control the point in time of pointer warping. Thus, no more thresholds are required.

2.2 Apparatus

Figure 1 (left) shows the experimental setup. The gaze data required for the MAGIC pointing techniques was recorded using a Tobii X60 remote eye tracker placed beneath a 24” monitor (1920 x 1200 resolution) using a Tobii Monitor Mount. According to the manufacturer, the Tobii X60 features an accuracy of 0.5° , and allows a freedom of head movement of 44 x 22 x 30 cm (W x H x D). Gaze position is sampled at 60 Hz. The participants sat at a distance of 65 cm from the monitor, resulting in a region of uncertainty of the estimated gaze position of 1.13 cm (corresponding to 43 pixels on the used monitor). The eye tracker raw data was filtered using the I-VT fixation detection algorithm (velocity threshold = $50^\circ/\text{s}$). For mouse input a Microsoft Comfort-Mouse 6000 for Business 3 (1000 dpi) was used. Mouse pointer speed was set to 8/11 using Windows7.

2.3 Experimental Task

The experimental task was similar to the one in Hild et al. [2013]. However, the colors and target size (30 pixels, 0.71° visual angle) were redesigned simulating the visual properties of real-time aerial video images. Figure 1 (right) shows a screen shot of the visual stimulus used in the experimental task. 294 targets in light grey, light yellow and light blue move from left to right over the screen like persons or vehicles would do. They move with constant, but different speeds (115, 122, 128, 135 pixels/s). The participants were told to select the 91 targets being highlighted in red as quickly as possible. Successful acquisition was indicated by a color change to green. To induce some of the time pressure and stress video analysts face, the number of targets visible simultaneously on the screen steadily grows during the task. Task duration is 3 minutes.



Figure 1: Experimental setup and visual stimulus of exp. task.

2.4 Procedure

12 participants (11 male, 1 female) between 21 and 27 years (average age = 23.5 years) volunteered in the study. All had normal or corrected to normal sight. 3 participants wore glasses, 2 used contact lenses. All were experienced mouse users (more than five years of daily use). One had used an eye tracker before.

Each participant performed the task with all three interaction techniques. The pilot study used a complete, counterbalanced within-subjects design. The participants completed a 9-point-eye tracker calibration before performing the first MAGIC pointing technique. For practice, they performed three multidirectional tasks with increased difficulty and a short version of the experimental task. The multidirectional practice tasks were based on the circular design proposed by the ISO/TS 9241-411 standard on “Evaluation methods for the design of physical input devices” (Figure 2). On clicking the red square, one of the 16 targets was red marked by random for selection. Targets were stationary (practice task 1, distance to square = 500 pixels), moved clockwise (practice task 2, distance = 500 pixels), and moved radially (practice task 3, distance to square between 250 and 500 pixels).

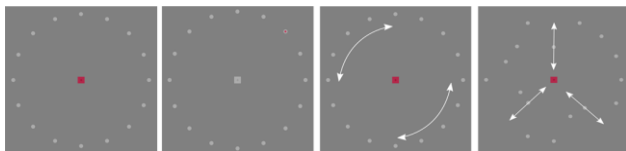


Figure 2: Visual stimuli of the multidirectional practice tasks.

After the experimental task, the participants rated the interaction technique using the ISO/TS 9241-411 standard questionnaire on a 7-point scale (7: most favorable response, 1: least favorable response) and reported their favorite interaction technique.

3 Results and Discussion

Given the small number of participants and the short test task, we consider the results of the pilot study to be preliminary. Acquisition time was measured by the difference between target highlighting and clicking at the target. Only successfully selected targets were used in the calculation. Figure 3 (left) shows that the result was best with mouse input, followed by MAGIC button, and was worst with conservative MAGIC. A repeated-measures ANOVA showed that participants’ performance varied significantly with techniques ($F(2,32) = 26$; $p < 0.001$). A post-hoc analysis with Bonferroni correction revealed a highly significant difference between mouse input and conservative MAGIC ($p < 0.001$), and a significant difference between MAGIC button and conservative MAGIC ($p < 0.05$); the difference between mouse input and MAGIC button was not significant. Acquisition error rate was calculated by the proportion of not successfully selected targets. Figure 3 (right) shows that, again, the result was best with mouse input, closely followed by MAGIC button, and worst using conservative MAGIC. A repeated-measures analysis of variance (ANOVA) revealed a highly significant difference between techniques for acquisition error rate ($F(2,32) = 11.96$; $p < 0.001$). A post-hoc analysis with Bonferroni correction revealed highly significant differences between mouse input and conservative MAGIC as well as between MAGIC button and conservative MAGIC; the difference between mouse input and MAGIC button was not significant. Comparing the results with Hild et al. [2013], acquisition time was now higher (mean = 1744 ms vs. mean 1422 ms), presumably because of the much smaller selectable size of the targets (0.71° vs. 4°) and because of the larger distances between targets due to the larger monitor ($24''$ vs. $17''$). However, the difference between means of mouse and MAGIC button (about 330 ms) is similar to the difference between mouse and liberal MAGIC (about 260 ms). The larger monitor might also contribute to the less mouse acquisition error rates (mean 1.1% vs. mean 3.3%). Using MAGIC button, participants yielded now a much better error rate than using liberal MAGIC (2,4% vs. 8,3%).

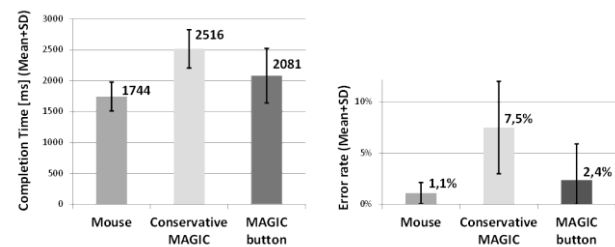


Figure 3: Completion time as a function of interaction technique and error rate as a function of interaction technique.

Figure 4 shows the results of the subjective ratings. As all participants were novices using MAGIC pointing, we divided the aspect of required effort into physical and mental effort. As two gaze-based techniques were tested, we added the question for eye fatigue. Asked for their favorite interaction technique for moving target acquisition, eight out of twelve participants voted for MAGIC button, three voted for conservative MAGIC, and one voted for mouse input. The good mouse input ratings for mental effort and overall input are presumably related to the great experience all participants had with this technique. The fact that mouse input was rated worse for physical effort confirms that moving target acquisition is a stressful task even after performing for about 15 minutes (including the practice tasks).

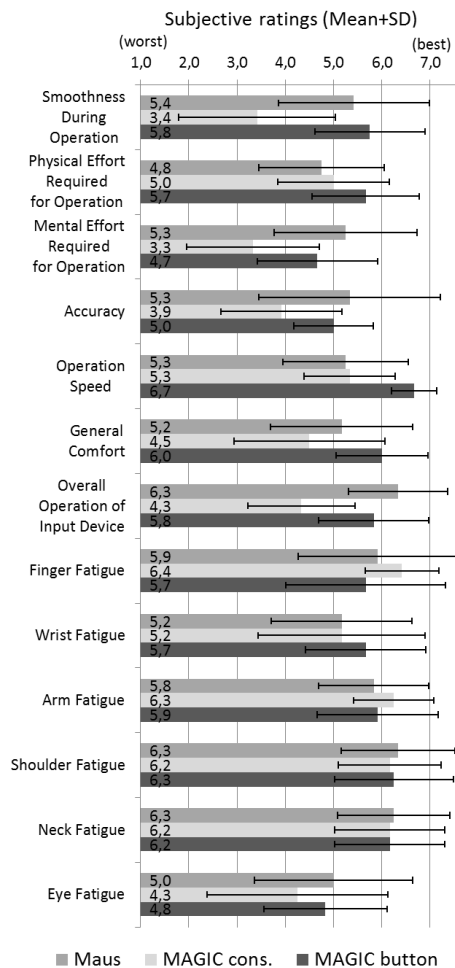


Figure 4: Subjective ratings as a function of interaction technique characteristics proposed by ISO/TS 9241-411.

Participants preferring conservative MAGIC described it as very intuitive and liked the little manual action required, resulting in the good ratings for finger and arm fatigue. However, considering the overall results for conservative MAGIC indicate that it is a complex technique. Concerning MAGIC button, participants very much liked the control over the pointer, even if the worse rating for finger fatigue results from the required two buttons “right-left-clicking” which confused a few participants. The fact that operation speed was rated best, even if the acquisition time results do not confirm this, has also been reported by Zhai et al. [1999] and Drewes and Schmidt [2009]; the fact that participants felt positive performing with this technique clearly contributed to good vote. The good ratings for physical effort and general comfort might indicate that this technique reduces manual load.

4 Conclusion and Future Work

The reduced requirement of mouse pointer search and replacement of the MAGIC pointing variants did not result in less missed targets or faster target acquisition. However, MAGIC button was subjectively rated best by the participants. All participants were novices using MAGIC pointing and thus had no internalized strategies for best use. The fact that mouse results for both acquisition time and error rate were statistically not

significantly better than the results for MAGIC button shows that MAGIC button could be a promising mouse alternative, especially in scenarios using more than one monitor and pointer search and manual replacement probably would take a longer time. Currently, we are investigating the use of a 5-button mouse using the thumb for warping control. It would be interesting to conduct a long-term study to see how the learning curves will evolve with respect to acquisition time and error rate.

Acknowledgements

The underlying projects to this article are funded by the WTD 81 of the German Federal Ministry of Defense. The authors are responsible for the content of this article.

References

- AL HAJRI, A., FELS, S., MILLER, G., AND ILICH, M. 2011. Moving target selection in 2D graphical user interfaces. In *Proceedings of Human-Computer Interaction--INTERACT 2011*, Springer, 141–161.
- DREWES, H., AND SCHMIDT, M. 2009. The MAGIC touch: Combining MAGIC-pointing with a touch-sensitive mouse. In *Proceedings of Human-Computer Interaction--INTERACT 2009*, Springer, 415–428.
- Ergonomics of human-system interaction - Part 411: Evaluation methods for the design of physical input devices, International Standard, International Organization for Standardization (2012) ISO, ISO/TS 9241-411
- FARES, R., FANG, S., AND KOMOGORTSEV, O. 2013. Can We Beat the Mouse with MAGIC? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 1387–1390.
- HASAN, K., GROSSMAN, T., AND IRANI, P. 2011. Comet and Target Ghost: Techniques for Selecting Moving Targets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 839–848.
- HILD, J., MÜLLER, E., KLAUS, E., PEINSIPP-BYMA, E., AND BEYERER, J. 2013. Evaluating Multi-Modal Eye Gaze Interaction for Moving Object Selection. In *Proceedings of the The Sixth International Conference on Advances in Computer-Human Interactions*, IARIA, 454–459.
- ISOKOSKI, P., JOOS, M., SPAKOV, O., AND MARTIN, B. 2009. Gaze controlled games. *Universal Access in the Information Society* 8, 4, 323–337.
- JACOB, R. J. K. 1991. The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Transactions on Information Systems (TOIS)* 9, 2, 152–169.
- WARE, C., AND MIKAELIAN, H. 1987. An evaluation of an eye tracker as a device for computer input2. *ACM SIGCHI Bulletin* 17, SI, 183–187.
- ZHAI, S., MORIMOTO, C., AND IHDE, S. 1999. Manual and gaze input cascaded (MAGIC) pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 246–253.