EMPIRICALLY EVALUATING REPRESENTATIONAL AIDS FOR TARGET TRACKING AND SENSOR MANAGEMENT

Taleri Hammack
Department of Human Factors Psychology, Wright State University
Dayton, Ohio
Jared Neely, Terry Stanard, Jason Roll
711th Human Performance Wing, Air Force Research Laboratory
Dayton, Ohio
John Flach
Department of Human Factors Psychology, Wright State University
Dayton, Ohio

Today, security officers at military and civilian installations are often required to track people and vehicles (targets) moving in a remote space using a distributed array of stationary security cameras. A pervasive tracking challenge is maintaining view of the target as it moves through the restricted fields of view of different cameras. The current research explores how different display designs indicating camera fields of view impact the operator's situation awareness of the next best camera to continue viewing a moving target. Three different interface displays (Full North-Up Map, Peripheral Display, and Track-Up Mini-Map) were evaluated over four experimental conditions. While having all display types available was most preferred by participants, the Peripheral Display provided better situation awareness as indicated by a statistically significant increased ability to pick the best camera to continue following the target. This was an encouraging finding since the Peripheral Display was designed to complement the video feed information while preserving spatial relationship information resembling a map-like display.

Between 2012 and 2016, the number of HD CCTV units shipped worldwide increased 140x from 0.2 million to 28 million units (Cropley, 2016). With the expanding deployment of surveillance camera technologies, there is a persistent need for a multi-sensor management interface (MSMI) that will support an operator managing a distributed array of cameras. The MSMI is critical for the success of a wide variety of surveillance scenarios within military (e.g., base perimeter defense) and civilian settings (e.g., train stations, airports, shopping centers).

In high priority target tracking tasks (such as a military base defense scenario) the need to maintain *constant* visual of a moving target is not an uncommon performance standard. Using a network of CCTV cameras placed throughout an urban environment, Roll, Stanard, Ayala, and Bowman (2016) conducted a simulated target tracking task, where the user's objective was to maintain visual of a walking pedestrian (the target) in the video feed of at least one camera at all times. Results of this experiment revealed that participants maintained view of the target 72% of the time on average, falling short of ideal. These results inspired the current research.

The current research focused on a target tracking mission using an array of grounded cameras/sensors distributed throughout a virtual urban environment. Participants were tasked with identifying the next-best camera to maintain visual of a walking pedestrian as he moved beyond the field of view of the current camera. To improve the MSMI user's spatial awareness of the targets movement in the 3D environment, two new displays were developed with information about nearby camera locations and their current and possible fields of view. To evaluate whether use of the two new displays in the interface could improve tracking performance, participants had to retrieve information from the displays to maintain visual of the target, including information to decide the next camera to select and what direction it should be turned.

Experimental Interface Displays

Full North-Up Map Display

All conditions provided the participant with a video feed from one camera (the right-most window seen in Figure 1) and the Full North-Up Map (seen on the left in Figure 1). The Full North-Up Map was a rectangular, north-up oriented map with symbology representing terrain (streets, buildings, parks), all cameras and their current and possible fields of view, and the target's initial starting location as indicated by the yellow dot.

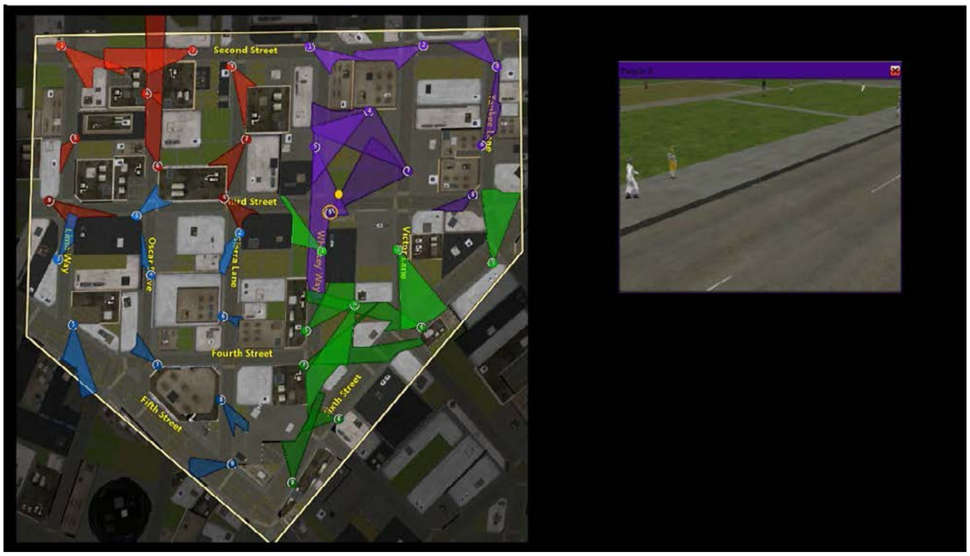

Figure 1. Screen capture of the multi-sensor management interface (MSMI) providing the Full North-Up Map display (left) and the video feed of a single camera (right).

In the video feed shown in Figure 1, if the pedestrian wearing a white robe continued walking up the sidewalk (towards the upper right corner), he would eventually move outside the possible field of view of the current camera. To maintain sight of this target, the user must determine what other camera to select and which direction to turn it to bring the target pedestrian back into view. To make these decisions the user must mentally project the 3D view of the environment seen in the camera feed into the top-down 2D map. These spatial transformations are both cognitively effortful and time consuming, since there is a mental reset time for the viewer to establish the context for the new scene in each display. This difficulty integrating data across successive displays is indicative of a MSMI with low "visual momentum" (Woods, 1984). To increase visual momentum and the ease of making camera selection and turning decisions, a "peripheral display" was designed.

Peripheral Display

Two conditions included the Peripheral Display (see Figure 2), which wrapped icons of the nearby cameras peripherally around the camera video feed. The location of each camera icon around the video feed corresponded to the approximate direction each camera would be located in the environment, relative to the scenery in the video feed. For example, the camera icon labeled "4," seen located in the upper left-hand corner around the video feed, indicates that this camera is located forward and leftward of the current camera view. Also surrounding the video feed was a colored border that indicated what part of the current camera view the nearby cameras could also see. A dashed line connected each camera icon to a border section, and overlapping border sections were darker colored (see Figure 2). This possible field of view was also indicated on the camera icon by the white arc. The yellow dash on the camera icon referred to the relative direction the camera was pointing. The colored border contained information somewhat redundant to the camera icons, providing an alternative representation of the fields of view of nearby cameras with respect to the current field of view in the video feed.

Track-Up Mini-Map Display

In addition to the Full North-Up Map and a camera video feed, two conditions also provided participants with the Track-Up Mini-Map (see Figure 2). The Track-Up Mini-Map was a smaller, circular map oriented so that the current field of view of the selected camera (providing the video feed) was centered and pointed upward. The Track-Up Mini-Map contained the same symbology for streets, buildings, nearby cameras and their current and possible fields of view, and the initial starting location of the target (indicated by the yellow dot). Surrounding the Track-Up Mini-Map were up to four icons of nearby cameras which were outside the boundary of the mini-map. These camera icons did not include the conical shapes indicating the current field of view. These camera icons surrounding the Track-Up Mini-Map were included to give the user knowledge about additional nearby cameras while minimizing interface clutter.

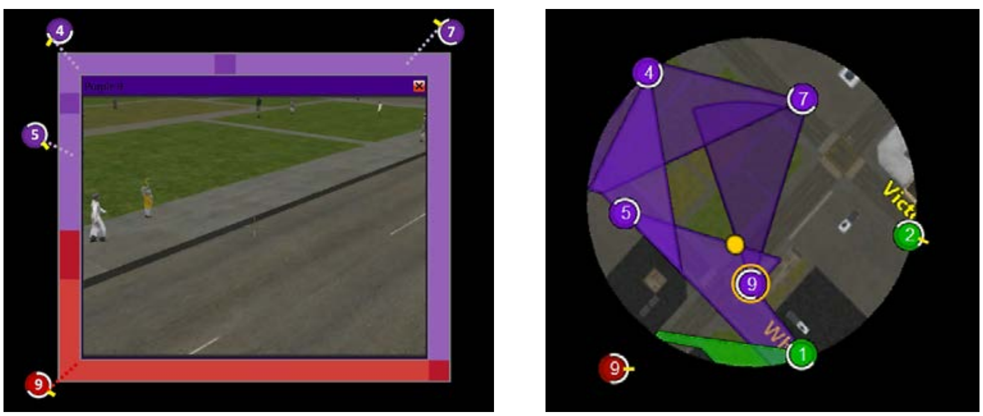

Figure 2. Screen capture of the Peripheral Display (left) and a screen capture of the Track-Up Mini-Map display (right). In the actual interface the Track-Up Mini-Map display was about two-thirds of the size of the Peripheral Display.

Method

Participants

A total of thirteen volunteer Wright-Patterson Air Force Base employees (8 males, 5 females) between the ages of 22 - 47 (M = 27.38, SD = 6.59) participated in this study. All participants reported normal/normal to corrected vision and normal color vision.

Experimental Design

The three display types were evaluated over four experimental conditions, and trials were blocked by interface configuration. Conditions 1-4 provided the video feed of a single camera showing a walking pedestrian (the target) and the Full North-Up Map. The Full North-Up Map was the only display provided in Condition 1. In Condition 2, the Peripheral Display was provided to participants in addition to the Full North-Up Map. For Condition 3 however, the Track-Up Mini-Map display was provided instead of the Peripheral Display. In Condition 4, all displays were made available (the Full North-Up Map, the

Peripheral Display, and the Track-Up Mini-Map). All participants completed the block of trials with each interface configuration (Conditions 1-4), with block order counterbalanced across participants. Each block contained 28 experimental trials, with the same randomized order of trials used for each participant. Each participant answered a total of 308 experimental questions.

Experimental Task

In each trial participants were presented an 8 - 18 second (M = 13.81, SD = 2.59) video clip of a walking pedestrian (the target) as viewed by one camera. Each trial required participants to answer two different multiple choice questions by retrieving information from the display(s) provided, and a third question based on the participants individualized use of the different available displays in that condition. The questions were: (1) "In order to maintain view of the target, what would be the next best camera to switch to?" (2) "In order to maintain view of the target, which way would the next best camera you selected need to be turned?" and (3) "Which display(s) did you use the *most* to answer the two previous questions?" The questions were presented sequentially so that once Question #1 was answered Question #2 would appear, and then once Question #2 was answered Question #3 would appear. Question #3 was not given to participants in the Condition 1 block of trials since there was only the Full North-Up Map display available.

Once all three questions were answered participants were presented a screen asking them if they would like to continue to the next trial; this was done to ensure participants were not rushed into the proceeding trial before they were ready. The target's physical appearance (gender, body type, clothing) did not change throughout the experiment. Feedback was not provided during the experimental trials, but a hard copy of the rules for selecting the next best camera to maintain view of the target (Question #1) was available to all participants throughout the experiment. The general rule was to select the closest camera the target would next approach if he continued walking in the same direction.

Procedure

Upon arrival, participants read and signed the informed consent document, filled out a short demographics questionnaire, and were given an overview of the study. In the overview, participants were presented with introductory training slides specifying the goals of the research, the nature of the task, and the requirements to successfully complete the upcoming trials, including the rules for selecting the "next best camera" (Question #1). Participants were then trained on the individual displays available in the interface configuration block they were going to receive next. After this training, participants were given 8 practice trails that they could repeat until they felt confident in their ability to retrieve the necessary information from the given interface display(s).

Participants were given a Post-Block Questionnaire after each condition, specific to the display(s) they just experienced. Questions included their perceived speed and accuracy, their ability to retrieve the necessary information, and their thoughts about possible display modifications. After all four blocks were completed (and the respective questionnaires) a Post-Experiment Questionnaire was administered. The Post-Experiment Questionnaire had participants compare the different display types, indicate their preferences, and provide any additional feedback or recommendations. Total session time for each participant was approximately 1.5 hours.

Results & Discussion

Data was collapsed across participants and analyzed for each condition. Performance data (response accuracy and time) and questionnaire responses were analyzed with a repeated measures Analysis of Variance (ANOVA) model with a Greenhouse-Geisser correction.

Accuracy

The results showed a significant difference in accuracy on Question #1 ("In order to maintain view of the target, what would be the next best camera to switch to?") across conditions ($F(2.13, 25.64) = 4.29$, $p < .05$), but not a significant difference for Question #2 accuracy ("In order to maintain view of the target, which way would the next best camera you selected need to be turned?"). Post hoc Bonferroni t-test results indicate that accuracy on Question #1 was significantly higher when the Peripheral Display was present with the Full North-Up Map than when only the Full North-Up Map was provided ($p = .013$). Accuracy on Question #1 was also significantly higher when the Peripheral Display + Full North-Up Map were provided than when the Track-Up Mini Map + Full North-Up Map were provided ($p = .028$). Finally, when all display types were made available (Full North-Up Map + Peripheral Display + Track-Up Mini-Map), Question #1 accuracy was marginally significantly higher than when only the Full North-Up Map was available ($p = .064$).

Response Time

In order to better reflect the time required to retrieve information to answer the questions correctly, response time was calculated from the time each question was presented until the participant selected their response. This enabled response times for both Question #1 and Question #2 to be recorded separately. There was not a significant difference in average response time on Question #1 across conditions, but there was a significant difference in average response time for Question #2 ($F(2.02, 24.32) = 3.34$, $p = .051$). Post hoc Bonferroni t-test results indicated that response times were significantly faster on Question #2 with the Full North-Up Map (only) than with the Peripheral Display + Full North-Up Map ($p = .042$). Average response time on Question #2 was also marginally faster with the Full North-Up Map (only) than with the Track-Up Mini-Map + Full North-Up Map ($p = .062$). Interestingly, average response time on Question #2 was not significantly different between the Full North-Up Map (only) and when all three displays were provided.

Subjective Data

The final Post-Experiment Questionnaire asked participants to rank the four different display configurations (i.e., experimental conditions) on: Ability to identify the next best camera, ability to identify the direction to turn the next best camera, predicted ability if they were tracking a target in real-time, and predicted ability if they were target tracking in real-time *and* had to track multiple targets. After collapsing the data across these four dimensions, the results showed a significant difference in condition preference ($F(1.88, 94.03) = 31.58$, $p < .01$). Post hoc Bonferroni t-test results indicate that the highest ranked, and thus most preferred option was when all display types were available (Condition 4). The next most preferred option was when the Peripheral Display was available (Condition 2), followed by having the Track-Up Mini-Map available (Condition 3), and then least preferred by participants was only having the Full North-Up Map (Condition 1). The only pairwise comparison that was not statistically significant at $p < .05$ was the preference for displays in Condition 3 over Condition 1. When averaging across the four ranked dimensions on the Post-Experiment Questionnaire, 9 of the 13 participants most preferred having all displays available.

Results from analyzing Question #3 responses revealed that people did not refer to the Full North-Up Map even half as often when they were given the Track-Up Mini-Map as they did when they were given the Peripheral Display (21.16% vs 55.53% respectively). This is a particularly interesting finding because accuracy performance was significantly higher when people were given the Peripheral Display compared to the Track-Up Mini-Map. These results suggest that the Peripheral Display, although useful, did not have all the necessary information to answer Questions #1 and #2. The need to include more information in the Peripheral Display, specifically information regarding the proximity of nearby cameras, was reiterated in several of the Post-Experiment Questionnaire comments made by participants regarding possible display improvements.

One limitation to the current study was the use of strict rules for choosing the one correct “next best camera” to continue viewing the target (Question #1). In the real world, the rules for selecting another camera are dynamic and depend on the context and goals of the operator doing the tracking. For example, the operator may seek a camera providing a close-in view of the target so details are visible, or the operator may instead seek a camera providing the longest view time of a moving target. Although the rules chosen for this experiment were found to be used by operators to track targets in a previous study (Roll, Stanard, Ayala, & Bowman, 2016), they are not necessarily the only criteria used by operators for selection of a camera to switch to.

Conclusion

Results from the subjective data revealed that participants most preferred having all display types available. This is not an unexpected finding since the layout of this interface allows participants to use each of the three displays independently or in combination, so having all displays available to them allows for the largest range of resources to answer each question. An encouraging finding was that situation awareness was increased when the Peripheral Display was made available, as indicated by a statistically significant increased ability to pick the best camera to continue following the target.

Future testing of the Peripheral Display should be explored in a more dynamic environment, with the operator tracking a target in real-time for extended periods of time. Furthermore, operators could be tasked with tracking multiple targets simultaneously, since this is a realistic scenario in real world applications such as in military base defense events. The utility of including camera proximity information in the Peripheral Display (e.g., icon size changes with camera distance, with further cameras having smaller icons) should be explored in future iterations of design. Including this information in the Peripheral Display could greatly reduce response times by reducing the need to consult a second display, namely a map. Furthermore, the camera icons wrapping around the Peripheral Display would support direct manipulation if operators could directly preview and/or switch to the desired camera video feed just by clicking on the icon. These design implementations tested in a real-time tracking task, would help verify that the Peripheral Display supports situation awareness by enabling faster, more accurate decision making when operators switch camera perspectives in order to maintain visual of a moving target.

Acknowledgments

The authors sincerely thank Elizabeth Frost and Antonio Ayala for their assistance in the experiment development process and to Mike Bowman for software development.

References

Cropley, J. (2016, February). *Top video surveillance trends for 2016*. Retrieved from IHS Markit Technology: https://technology.ihs.com/520142/top-video-surveillance-trends-for-2016

Roll, J., Stanard, T., Ayala, A., & Bowman, M. (2016). Enabling human-automation collaborative target tracking using an array of sensors. *Ground/Air Multi-Sensor Interoperability, Integration, and Networking for Persistent ISR VII, Proceedings of the SPIE, 9831. Baltimore, MD: SPIE.*

Woods, D. (1984). Visual momentum: A concept to improve the cognitive coupling of person and computer. *International Journal of Man-Machine Studies*, *21*(3), 229-244.