STRATEGIES FOR THE INTERPRETIVE INTEGRATION OF GROUND AND AERIAL VIEWS IN UGV OPERATIONS

Roger. A. Chadwick*, and Douglas. J. Gillan
New Mexico State University
Department of Psychology
Las Cruces, NM 88003

ABSTRACT

Two experiments examined the cognitive process of aerial view target localization. Participants were shown ground-view images with designated targets, and tasked with locating the target in an aerial view. The first study examined photographic image sets in both a qualitative and quantitative manner, including a think-aloud protocol analysis. The second study used manipulated three dimensional model images to isolate effects of color, shape, and other attributes. Results show a strong cue dominance effect for unique colors, sex differences, and minor view angle effects. We discuss a proposed cognitive model for this task and suggest recommendations for assistive unmanned ground vehicle (UGV) interface features.

1. INTRODUCTION

The window of opportunity for precise surgical tactical strikes can be brief. Quickly obtaining and communicating precise targeting information obtained from the imagery of remotely operated ground vehicles (UGVs) is essential to mission success in tactical UGV operations. Yet spatial disorientation in the operation of remote vehicles, especially multiple vehicles, can be a problem (Chadwick, Gillan, Simon, & Pazuchanics, 2004), and operators making targeting decisions based on remote vehicle imagery must be fast and accurate in order to maximize effectiveness because missed opportunities can have lethal consequences. In two studies we examined the cognitive process of identifying a ground viewed target on an aerial map view. Cognitively, this task is an application and conjoinment of the psychological processes and effects of perception, object recognition (i.e. Biederman & Gerhardstein, 1993), navigation (Wickens, 1990), and similarity (Tversky, 1977). Identifying ground viewed objects in an aerial view involves a specific type of mental rotation (Shepard & Metzler, 1971) transform , and the process is affected by many image attributes and object relations. This task is highly relevant to UGV operations where remote ground imagery is transmitted to operators who must comprehend viewed objects in terms of their global spatial relations. The use of informative maps in such cases are essential.

The inherent difficulties with spatial navigation in remote unmanned ground vehicle (UGV) operations motivates the use of map modules detailing position and orientation of the UGVs in their spatial environment. Excellent satellite imagery and GPS technology are available for constructing these maps, but simply placing an icon indicating UGV position and camera orientation does not fully solve the spatial comprehension problem. A tactical question of interest most often involves an object *within* the view of the UGV imagery rather than the position of the camera itself. Therein lies the problem of ground and aerial view integration. Issues regarding map rotation and UGV icon functionality become complex when multiple UGV systems are considered. While track oriented maps that rotate to provide current orientation in the "up" direction may be useful in some circumstances, they are not without their drawbacks (see Wickens, 1990). The operation of multiple ground vehicles is a situation in which such solutions might not be helpful. The consistency of a north-up map should provide the necessary visual momentum (1990) and situational awareness for the maintenance of landmark orientation and global spatial awareness as operators switch from one UGV to another. Constant view switching in multiple robot interfaces will exacerbate spatial disorientation, and a careful analysis of the critical task of mapping ground viewed objects to aerial view maps is addressed in the current studies.

The difficulty of the ground and aerial image integration task became apparent in our simulations of multiple UGV operations where participants were given the task of locating an object in the UGV camera image on a satellite image map (Chadwick, 2006). The task was difficult for many participants, error rates were high and response times often slow. It became clear that assistive interface features that address this issue will be extremely beneficial. Implementation of these features may impact the design of robotic vehicles in that they must provide the necessary information (telemetry). The ultimate goal of the current series of studies at the Human Robotics Interaction Laboratories at New Mexico State University is therefore to propose and test (via computer simulations) ground air view integration enhancement tools for remote vehicle interfaces. In order to maximize effectiveness of design, a deeper understanding of the cognitive demands of this task are required, and these studies address these issues.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01 NOV 2006	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Strategies For The Interpretive Integration Of Ground And Aerial Views In Ugv Operations	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **New Mexico State University Department of Psychology Las Cruces, NM 88003**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM002075., The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**8**	

Complimentary qualitative and quantitative methods were used in a multi-phased examination of this cognitive task. In both studies participants were given the task of finding a designated target, as seen in a ground view, on an aerial image map. In the first experiment actual photographs (e.g., figure 1) of air and ground views were analyzed, and in a follow on experiment three-dimensional (3D) computer generated models of building scenes were manipulated in order to examine specific hypotheses concerning a cognitive model of the task.

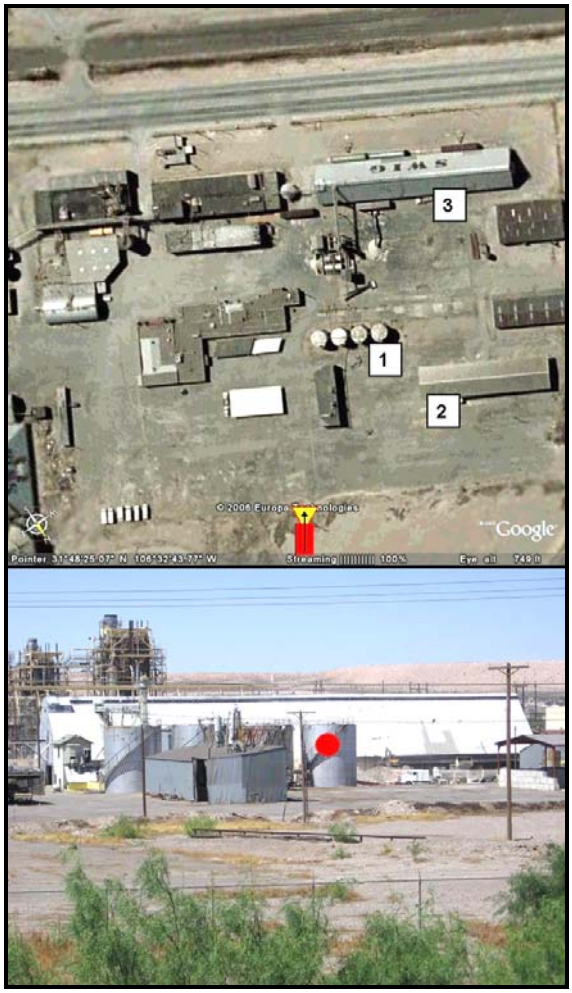

Figure 1. Sample stimuli for study 1 with a canonical aerial view. The filled red circle designates the target. The target designations in the aerial view are added for clarity.

After an initial examination of the photographic stimuli and a review of the commentary of participants performing this task, we proposed an *opportunistic multiple cue abductive model* of the cognitive process. The process is opportunistic in the sense that the solution to any particular set of images or target location depends on a complex set of factors (cues) which are chosen for their appropriateness to the specifics of the image content, using some initial possibly unconscious thought process. As the solution process continues, various hypotheses are tested in a form of abductive reasoning until the target object's location in the aerial image is found and confirmed. Imagining (mental imaging) of the aerial view transformation of ground objects is employed where viable, and in some cases analytic propositional thought is exercised. In this sense the solution is a combination of fast perceptual processes coupled at times with slower analytical thinking. An example of the analytic process would be saying to oneself "I know the building is beyond the railroad tracks, and the railroad tracks are here, so this must be it.", or perhaps counting as in "it's the third one over from the white building, one, two, three, it must be this one".

Our proposed model of the cognitive process of locating an object in an aerial image consists of a) an initial cue selection process which may to a large extent be automatic and unconscious, b) a set of possible perceptual cues, c) imagination of the transformed object or gestalt group shape, d) the use of reference objects, e) analytical propositional thought, f) abductive hypothesis testing, and g) a verification process using a check object. Guiding the process is an affective sense that a satisficing solution has been reached, a meta-sense that the participant has or has not found the object correctly in the aerial view. After a detailed discussion of the methodological details, data from the two experiments are examined in order to evaluate this proposed opportunistic abductive reasoning model.

2. METHOD

2.1 Experiment 1: Photographic image sets.

A small group of *think-aloud* participants performed an aerial localization of a ground viewed target task, and were instructed to express their thoughts verbally subsequent to each trial (see Nielsen, 1993). An independent second group of *response-data* participants performed the target localization task without think-aloud.

Participants. Think aloud verbal protocol was recorded for 7 participants, 4 men and 3 women with a mean age of 26 years. Response data were collected on the same image sets for an additional 27 undergraduate participants consisting of 13 men and 14 women with a mean age of 24 years. Participants gave informed consent and were debriefed upon completion of the session.

Apparatus. Aerial (satellite) images of 0.3m resolution were obtained of several distinctive areas around El Paso, Texas; including a milk factory, power plant, downtown area, and cotton processing facility. These areas were chosen because of diverse terrain and object characteristics reflecting a variety of plausible tactical military operational scenarios. We obtained the color images using commercially available internet software

(GoogleEarth) set for an eye-altitude of 750 or 1000 ft., with an image size of 692 x 692 pixels. Color ground photographs of 3.2 Mpixel resolution were taken of the same sites, from various viewpoints at eye-level height. While dynamic environmental features such as vehicles were not consistent between the air and ground images, significant static features were consistent as the aerial images were of recent origin. Aerial images were edited to include a camera viewpoint icon, and rotated at angles (from a canonical ground-view) of 0, 90, and 180 degrees. Targets were designated in the ground images by a filled red circle (red-dot). Three versions of each ground image were created, each with a different target designation. In this manner the differential difficulty of particular target characteristics could be examined while holding image content constant. The ground targets and air view rotations were given to participants in three groups such that each participant responded to a total of 36 image pairs, with angle and target counterbalanced between the three participant groups. The aerial image was displayed in the upper half of the 19" (1024 x 768 resolution) computer display in 512 x 512 pixels, with the ground image centered directly below the air image.

Procedure. Participants were given the task of finding a designated ground-view target and clicking with a mouse pointer on the aerial view at the exact point corresponding to the target red-dot indicator. They were instructed to respond as quickly and accurately as possible and informed that the delay between trials would depend upon their accuracy (maximum delay of 10 s). The intent of this delay and instruction was to minimize guessing. A set of four practice trials were performed with the experimenter's guidance. The correct response was provided for the first two practice trials, which were views taken of the building in which the experiment was performed, ensuring some familiarity. Each of 36 trials began with a full screen view of the ground image coupled with a brief text noun phrase description of the target object (to avoid any possible ambiguity) displayed for 10 seconds. This ground image preview was followed by a pair of ground and aerial images. Image pairs consisting of 36 trials were presented in random order without replacement. Each participant received each image set three times, at three different angles, with each specific target presented only once (at one of the three angles). The mouse cursor was positioned randomly to one side of the vertical center of the screen at the start of each trial. Following each target locating response, the images were removed and participants provided a confidence rating on a seven point scale (1 = *extremely not confident*, to 7 = *extremely confident*). Response time (ms), localization error (pixel offset from actual target location), and confidence ratings (1-7) were recorded.

For the think-aloud participants only, the images were immediately re-displayed after each response while participants were prompted to verbally express their thoughts on finding the target. Participants were encouraged to use the mouse cursor as a pointer to assist in their expression. The computer display video and participant's voice were recorded on VHS tape. Participant comments were later transcribed and organized by specific image sets ranked according to response time and localization error for analysis. After analyzing the think aloud commentary, a set of attributes were defined for the images, and the images in air-ground pairs were rated on these attributes by three independent raters. Mean ratings from the two raters with the highest correlations were then used in a regression analysis of the contribution of these attributes to the response time.

2.2 Experiment 2: Manipulated modeled 3D images.

Photographic images may be ecologically valid, but the image content in terms of terrain, object types, shapes, and arrangements varies along many complex dimensions. In order to control image content and manipulate object attributes in this task, computer generated 3D models were created. Object attributes were manipulated in a systematic fashion in order to examine specific hypotheses regarding the cognitive process of integrating air and ground images.

Participants. A group of nineteen participants consisting of thirteen women and six men, mean age 21 yrs, provided informed consent and participated in this study. All participants were debriefed at the end of their session.

Materials. Scenes consisting of four to six building objects were created using commercially available architectural planning software (GoogleSketchup). Objects included flat and pitched-roof houses, cylinder tanks, and Quonset hut type structures of varying colors, shapes, complexities, and arrangements (e.g., figure 2). The attributes of shape and color uniqueness were varied in four levels. From each scene images were rendered from an aerial (directly overhead) and two separate ground viewpoints. Each ground viewpoint corresponded to a specific target object, which was designated with a filled red circle in the ground image, as in experiment 1. Object color uniqueness (relative to the distractor objects) was manipulated within a specific scene while holding the object shapes constant, and shape and shape uniqueness were manipulated between scenes using differently shaped building types and combinations. Shape similarity is a complex phenomenon consisting of comparisons of both similar and dissimilar features (Tversky, 1977). Shape uniqueness level 1 consisted of a group of same objects, at different orientations, or a group of very similar objects differing only in small features (figure 3). A complete set of 216 air and ground image pairs were rendered, including a series of scenes designed to test separate hypotheses and foils intended to preclude

demand characteristics. Participants in four groups received 66 trials each, with groups representing specific attribute combinations for specific scenes.

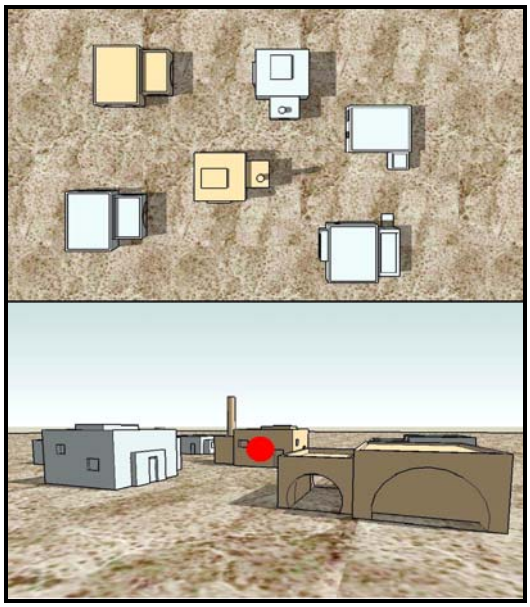

Figure 2. Study 2 3D modeled stimuli. Each scene of objects was rendered from two unique ground viewpoints.

Scenes were created to test a variety of separate hypotheses including the effects of shape and color uniqueness, hidden aerial or ground features, texture, shadows, and adjacent salient reference objects. Hidden aerial features are object features visible in the aerial view, but hidden in the ground view. Hidden ground features are distinctive features, such as the arched doorways in figure 2, that are visible only in the ground view. For the shape-color hypothesis, each participant received one trial in each of 16 combinations of shape and color uniqueness representing a full 4 x 4 within subjects factorial design. Eight scenes consisting of 8 variations each were thus presented to participant groups such that each participant saw each particular ground view and target only once, with different groups receiving different instances of each condition. For the hidden aerial feature hypothesis, trials were arranged in a 2 (hidden aerial features: present vs. absent) x 2 (color uniqueness: all same color vs. target unique color) full factorial design. Trials for the other hypotheses were arranged similarly.

Procedure. Participants were given the task of identifying the location of the target object in the aerial view. After being randomly assigned to one of four image set groups, participants were given a brief training episode consisting of a demonstration video sequence and 4 practice trials. They were instructed to click near the center of the target object in the aerial view. Mouse responses were restricted to the aerial image area of the screen. A click within a radius of 80 pixels from the center point of the target was scored as accurate, the task being to identify the target object rather than any particular point on the object. As motivation against guessing, participants were again instructed that their completion time would be minimized by accurate pointing (maximum inter-trial error proportional delay of 12 seconds). After completing the practice trials each participant responded to 66 randomly presented experimental trials. At the start of each trial a fixation of one second duration (centered on the ground image) was followed by presentation of the images, each air and ground image displayed in 588 x 380 pixels, with the air image directly above the ground image. One of the objects in the ground image was designated as the target by a filled red circle. The images remained displayed until a valid mouse click, after which the images were removed and the participant began the next trial by moving the mouse back to a white circle at the center of the aerial image display area. In this manner the start of each trial was self paced, with the mouse always starting at the center of the aerial image, an approximately equal distance from each of the possible target objects.

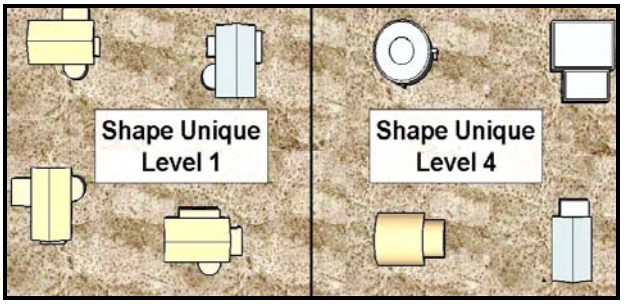

Figure 3. Shape uniqueness varied between all objects of the same or very similar shape (left panel), and all uniquely shaped objects (right panel).

3. RESULTS

3.1 Experiment 1.

Our analysis of the photographic image response is both qualitative and quantitative. Qualitative results from the think aloud study will be discussed first. Quantified effects include effects of image content, gender, angle, and image attribute regression analyses (*alpha* in all analyses = .05). In the case of image content, a picture is worth a thousand words, and while this report is unfortunately too brief to include a complete appendix of images, exemplars are included where beneficial in illustrating some important points.

Think-Aloud. A qualitative summary of the highlights of the think-aloud transcriptions reveal several prominent strategic concepts. It was observed that, without being able to describe the process directly, participants were able to imagine the aerial view transformation of many objects, especially those with distinct and rather simple *geon structural descriptions* (GSD, see Biederman &

Gerhardstein, 1993). A *geon* is a basic three dimensional solid shape, such as a cylinder, cone, or cube. Biederman's recognition by components theory explains object recognition by positing the use of object structural descriptions consisting of combinations of basic geon types and relations. An object is viewpoint invariant if the same GSD is activated by different views. In our study, a cylindrical storage tank or cone shaped tent-top roof are examples of targets for which the imagined aerial transform of the ground view GSD facilitates easy recognition. Thus, viewpoint invariant simple objects that are readily de-composed into geons are more easily recognized in aerial views. The regression analysis on image attributes discussed in a following section supports this observation.

Another observed strategy employed in the solution of this task is the use of salient, easily recognizable reference objects. A *reference object* is a non-target object with proximity to the target itself, which can be used to identify the target when the GSD of the target is, in itself, insufficient for a rapid or easy recognition. Gestalt groups of similar objects were often used as reference objects, such as "the four round tanks here", or in the identification of targets that were part of such a group, a case where *counting* was employed. Another observation is the ease of recognition and orientation made possible by building objects containing lettering or symbols (such as the rooftop "SWIG" text and cotton symbol in images of the cotton processing facility). As one participant said, "what stuck out was the cotton [symbol] underneath the SWIG lettering... any lettering gives away location". For the images which contained text or symbols on building rooftops, these were mentioned by most participants many times. Distinct colored targets were identified in a similar manner, this effect somewhat attenuated by the somewhat washed-out colors in the aerial photographs, miss-match between ground and aerial colors, and inability to see roof top colors in ground images. At times, an imaginary camera view line was mentioned by participants as they tried to imagine relations of objects along the camera line of sight. Upon hypothesizing the identification of a target participants often reported a secondary check-process, testing their finding against other object relations. Finally, shadows were used, especially in the case of tall objects such as smoke-stacks, but also surprisingly in cases of verifying some detailed features. It was also clear that distortions in depth perception (objects appearing closer together in the ground view) made identifications more difficult, as did any unimaginable (unexpected or hidden) details of target gestalt group appearance in the aerial view. Analytical thinking such as "the target is a guard shack, that should be near the entrance to the facility" was also evident, augmenting the perceptual process.

From analysis of the think aloud protocol several basic constituents of aerial object identification were extracted including: a) distinct color, texture, or markings; b) shadows, c) reference objects, c) gestalt groupings, e) imaginary camera line of sight, f) viewpoint invariant GSDs, g) imagined aerial views, h) analytical thought (including counting), and i) hypothesis testing (including a checking process).

Image content. The range of response times and overall localization errors for various images is noteworthy. Indicative of motivated responding, 50 % of responses were accurate within 8 pixels of actual target location, and 75% were accurate within 30 pixels. Response times and localization errors were positively correlated (r = .221, p < .01), indicating that inaccurate responses took longer, and there was no speed-accuracy trade off. While the mean response time for any target was 14.5 seconds (*SD* = 14.5 sec, *Mdn* = 9.7 sec), there was a great deal of variation across targets. The range of means for specific targets was from 5.5 to 30.1 seconds. With response time as a criteria, the easiest (fastest) image scene was that of a large mall building with targets of a doorway, roof-top tent structure, and roof-top dome (*M* = 9.6 s, 5.5 s, and 6.0 s for targets 1, 2, and 3 respectively). The cotton processing facility scene depicted in figure 1 was one of the most difficult scenes (M = 11.3 s, 30.1 s, and 15.8 s for targets 1, 2, and 3 respectively), with the most difficult target (2) being a shed which was only partially visible in the ground image, and for which the elongated roof structure seen in the aerial view was probably unexpected and unimaginable. Target 3, the long white building in the background of the ground view is also somewhat difficult, perhaps due to the color inconsistency between air and ground views, and the effect of relative size distortions and depth compression.

Confidence. Participants rated the confidence in their judgments after each trial on a 7-point scale. Confidence ratings accurately reflected performance, with significant correlations with both response time (r = -.439, p < .01) and localization error (r = -.458, p < .01). Longer response times and greater localization error resulted in lower confidence ratings. This implies that participants were meta-cognitively aware of their errors, and were less confident of more difficult judgments that took longer.

Gender effects. These data reveal a strong gender effect, with men (*M* = 11.2 sec, *SE* = 0.45) faster than women (*M* = 19.1 sec, *SE* = 0.76), a result consistent with many studies in spatial reasoning (Astur, Ortiz, & Sutherland, 1998; Voyer, Voyer & Bryden, 1995) and an effect especially strong in mental rotation and spatial perception tasks (1995). In a repeated measures general linear model (GLM) analysis with image content (12 image sets) and targets (3 different targets per image set) as variables, the effect was significant with $F(1,24) = 7.4$, $p < .02$. While there was no significant interaction between gender and image, the difference was present in all 12 images to a

greater or lesser degree. There were differences across specific targets which, on a case by case basis, can be revealing, although a complete analysis of all images and targets is beyond the scope of this paper.

View-angle. Aerial views were presented either canonically (0 degrees), or rotated 90 or 180 degrees with respect to the viewpoint of the ground image. There was no effect of view-angle on response time, except for some specific targets. Target identification localization error increased linearly with increased angle discrepancy (M = 26.3, 32.0, and 40.2 pixels; SE = 3.1, 3.2, 3.9 pixels, Beta = .102, $p < .01$), although angle accounted for only a very small percentage of the variance ($R^2 = .01$). This linear effect of view angle discrepancy is not consistent with related studies in identifying cardinal direction of targets relative to an object, where there is generally a significant rotation effect with a reduced effect at the "upside down" 180 degree orientation (e.g., Gugerty & Brooks, 2004), explained by the switching from a mental rotation strategy to an analytic "reversal" judgment. The effect of view-angle was not consistent across image sets (figure 4), or all of the targets within an image set, but represents an average effect of minor consequence that was statistically significant in only 3 of 36 target cases analyzed separately.

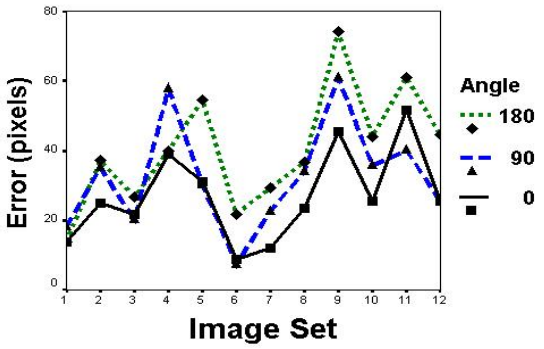

Figure 4. The effect of aerial viewpoint angle on target localization error varied across the image sets and targets.

A close look at a specific image set for which angle was a significant factor in accuracy is revealing. As an example, the effect of view-angle is especially strong for the 180 degree angle in image set 5, target 3 shown in figure 5 (means for 0, 90, and 180 degrees = 60.2, 71.6, 132.2 pixels, respectively, F (2,18) = 4.28, p < .05). Looking at the images we can see that the difficult target (3) for 180 degree detection in this image set is the "power pole". The 180 degree reversal of object left-right relations did not significantly affect identification of the clearly distinguishable targets (cylindrical storage tanks), but drastically impacted the localization of the difficult target (power pole), which is embedded in clutter, not de-composable into geons, difficult to distinguish from background and similar objects, and in general small and hard to see. The difficulty of target 3, the power pole, is also seen in the response time measure, with mean response times for this target doubling from 14 seconds for 0 degrees to 28 seconds for 180 degrees.

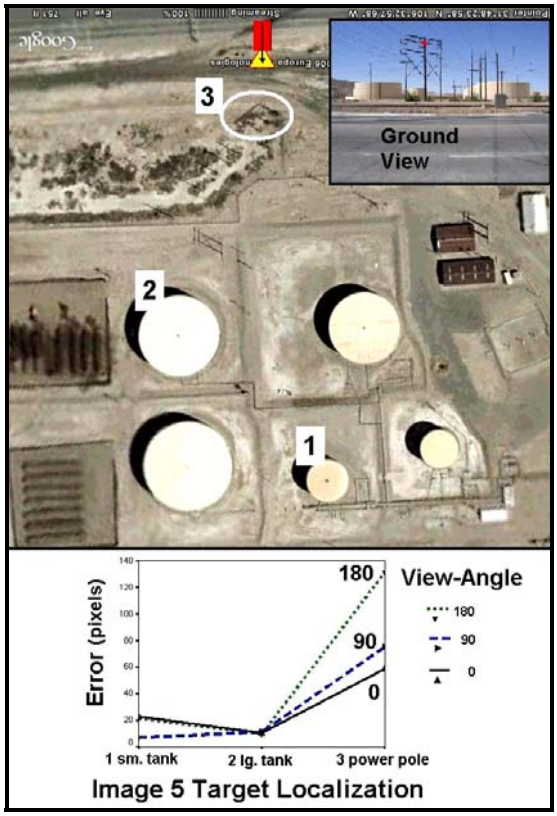

Figure 5. View angle effects were often greatest for the difficult targets, both in terms of error (shown here) and response times.

Image attribute regression analysis. Based on the information gathered in the think aloud session, we defined a group of potentially predictive attributes of the targets and images, and independent judges blind to the response results rated the images along these dimensions. The data and think-aloud protocol suggested target object attributes of distinct color, color match between images, geon decomposability, camera proximity, gestalt group membership, and viewpoint invariance (imaginability of the aerial shape). Of these attributes, a linear regression model ($p < .001$) accounting for 12.5% of the variance in response time was derived from participant gender (Beta = -.269), target viewpoint invariance (Beta = -.208), target gestalt group membership (Beta = .151), and ground-air view target color matching (Beta = .086), $p < .05$ for all predictors. Note that gestalt membership is positively correlated with increased response time.

3.2 Experiment 2.

A GLM repeated measures analysis using within-subject variables of shape uniqueness (4 levels), color uniqueness (4 levels), and a methodological between-subjects group variable (particular image set, four groups)

was used. The group factor was not statistically significant. There was a main effect for color uniqueness with fastest responses for target unique color (F (3,54) = 20.1, $p < .001$) and a main effect for shape uniqueness (F (3,54) = 6.03, $p < .01$) with fastest response times for unique target shape. The interaction between color and shape uniqueness was also significant (F(9,162) = 2.26, $p < .05$), revealing that the effect of shape uniqueness is greatly reduced as color uniqueness increases, such that for a uniquely colored target there is no effect of shape uniqueness with a very fast response (figure 6). Mean response times for color unique levels 1 through 4 were 14.6, 10.2, 8.6, and 2.8 seconds, respectively (SE = 2.2 s, 2.3 s, 1.5 s, 0.40 s). This is a rather large effect (partial eta squared, $\eta^2 = .67$), with the response time for a uniquely colored target 520% faster than for a target the same color as all the distractors. Mean response times for shape uniqueness levels 1 through 4 are 12.2 , 11.1 , 8.1 , and 4.8 seconds (SE = 2.9 s, 2.2 s, 0.98 s, and 0.39 s), respectively, (partial eta squared, $\eta^2 = .51$).

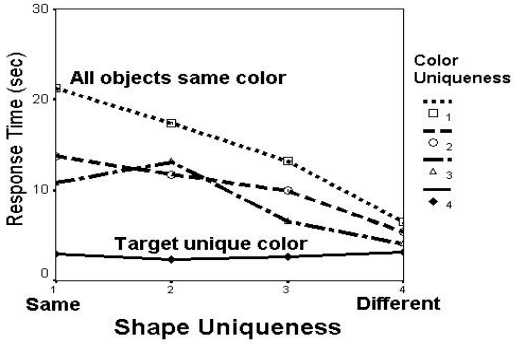

Figure 6. Target color uniqueness is a dominant cue in the aerial view identification task .

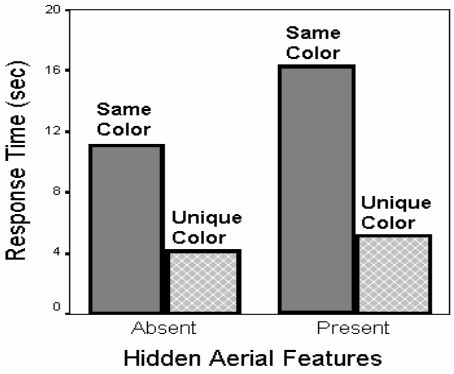

Figure 7. Localization is delayed by "hidden" aerial view features that are *unimaginable* from the ground view .

The presence of adjacent salient reference objects reduced response time by 4.9 seconds (F(1,18) = 15.6, $p < .01$). Other variables examined in preliminary analyses indicate that there was a trend towards effects of hidden aerial features (figure 7), and textures, All of these effects interacting with color uniqueness, which is an overriding effect. Gender (men = faster) and target complexity (number of features, simple = faster) are also significant variables. A linear regression on gender, complexity, shape uniqueness, and color uniqueness across all hypotheses resulted in significant standardized Beta coefficients for each factor, of -.118, .074, -.149, and -.361, respectively, with $R^2 = .17$. There was no effect for hidden ground features, and the effect of shadows is inconclusive at this point, (insufficient statistical power).

4. DISCUSSION

Many aspects of our proposed cognitive model were corroborated by these studies. The regression analysis on photographic image attributes converges with the 3D model results on several points. Color uniqueness is an excellent cue, which can in itself lead to a fast identification in cases where it is present. Our proposition that aerial transforms are imagined will be supported by the finding that hidden aerial features slow identification, a result consistent with both the 3D model data and photo-image analysis. Hidden ground features do not impact the process because they play no part in imagined aerial views. Consistent with the proposal regarding simple geon structures, less complex targets are identified faster, more easily. The finding on adjacent reference objects supports the view that gestalt groups of objects are used. We have not yet examined the abductive reasoning and analytical reasoning aspects of the process.

One problem with taking full advantage of the color cue dominance in real images is that, at least with the images available for this study, satellite images appear to contain desaturated color (colors are washed out), and lighting conditions that vary between the acquisition of such images and the real time ground imaging can reduce color matching (e.g., the sun reflecting on the grey tin roof of the SWIG building in figure 1). One suggestion therefore, is to acknowledge the benefit of color matching and produce satellite image maps that reflect ground image coloration parameters to the extent this is possible.

The accuracy of imagined spatial relations of gestalt groups of objects seems critical. Depth perception in ground images is often poor and this contributes to the difficulty. Consider the image pair shown in figure 8. The imagined spatial relations of the buildings and large generator unit in the background differs significantly from the actual spatial relations seen in the aerial view. Furthermore, the presence of the ponds, distinctive "hidden" aerial view features below ground level, destroys the gestalt and contributes to the difficulty.

Operators of unmanned robotic ground vehicles are at a significant disadvantage in comparison to human scouts when it comes to the spatial comprehension of tactical areas of interest. Human scouts have the benefits of a

single vantage point, wide field of view, depth perception, and continuity in navigation. UGV operators on the other hand suffer from narrow fields of view, discontinuities in navigational attention, poor depth perception, and constant viewpoint changes as a result of switching between cameras from multiple vehicles. In our photographic study, despite the presence of a camera position and orientation icon, individual target identifications in the aerial view took anywhere from just a few seconds to several minutes. In a time critical situation, a minute is an excessive lag. We suggest the development of interface tools specifically designed to facilitate this critical task. Proposed interface enhancements including view-lines drawn on situational awareness maps based on ground image target designation, using UGV position, orientation, and camera pointing information should be examined. UGV designers must include necessary information for such features in the telemetry of their vehicles. Also, depth perception enhancement techniques should be examined.

Figure 8. Depth in ground image is greatly foreshortened, presenting a problem in object gestalt viewpoint invariance.

Future directions. Because each image set, despite complexities of content, were analyzed for three separate targets by two groups (genders) of individuals presumed to differ in strategies (Rahman, Andersson, & Govier, 2005), we can analyze various contributors to difficulty in the comparison both between and within image sets, across various viewpoint angles. A complete appendix consisting of an analysis of each image set along these lines will be forthcoming and revealing. Manipulation of more sophisticated 3D models simulating actual scenes can be used to further validate our model.

ACKNOWLEDGMENTS

Prepared through participation in the Advanced Decision Architectures collaborative Technology Alliance sponsored by the U.S. Army Research Laboratory under Cooperative Agreement DAAD19-01-2-0009. The views and conclusions contained in this document are those of the authors, and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory, or the U.S. Government. Author contact: rchadwic@nmsu.edu.

REFERENCES

Astur, R. S., Ortiz, M. L., & Sutherland, R. J. (1998). A characterization of performance by men and women in a virtual Morris water task: A large and reliable sex difference. *Behavioral Brain Research, 93*, 185-190.

Biederman, I., & Gerhardstein, C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology, 19*, 1162-1182.

Chadwick, R. (2005). Multiple robots and display views: An urban search and rescue simulation. *Proceedings of the Human Factors and Ergonomics Society*, 2005, FL: Orlando .

Chadwick, R. (2006). Operating multiple semi-autonomous robots: Monitoring, responding, detecting. *Proceedings of the Human Factors and Ergonomics Society*, 2006, San Francisco, CA.

Chadwick, R.A., Gillan, D. J., Simon, D., Pazuchanics, S. (2004). Cognitive analysis methods for control of multiple robots: robotics on $5 a day. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*. (pp. 688-692). Santa Monica, CA: HFES.

Gugerty, L., & Brooks, J. (2004). Reference frame misalignment and cardinal direction judgments: Group differences and strategies. *Journal of Experimental Psychology, Applied, 10*, 75-88.

Nielsen, J. (1993). Thinking aloud. *Usability Engineering* (pp. 195-200). Academic Press, CA: San Diego.

Rahman, Q., Andersson, D., & Govier, E. (2005). A specific sexual orientation-related difference in navigation strategy. Behavioral Neuroscience, 119, 311-316.

Shepard, R.. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701-703.

Tversky, A. (1977). Features of similarity. Psychological Review, 84, 327-352.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. Psychological Bulletin, 117, 250-270.

Wickens, C. D. (1990). Navigation ergonomics. In E. J. Lovesey (Ed.), Contemporary Ergonomics 1990: Proceedings of the Ergonomics Society's 1990 Annual Conference (pp. 14-29). London: Taylor & Francis.