Visual and IR-Based Target Detection from Unmanned Aerial Vehicle

Patrik Lif(✉), Fredrik Näsström, Gustav Tolt, Johan Hedström, and Jonas Allvar

Swedish Defence Research Agency, Linköping, Sweden
{patrik.lif,fredrik.nasstrom,gustav.tolt,johan.hedstrom,
jonas.allvar}@foi.se

Abstract. In many situations it is important to detect and identify people and vehicles. In this study the purpose was to investigate subject's performance to detect and estimate number of stationary people on the ground. The unmanned aerial vehicle used visual- and infrared sensor, wide and narrow field of view, and ground speed 8 m/s and 12 m/s. Participants watched synthetic video sequences captured from an unmanned aerial vehicle. The results from this study demonstrated that the ability to detect people was affected by type of sensor and field of view. It took significantly longer time to detect targets with the infrared sensor than with the visual sensor, and it took significantly longer time with wider field of view than with narrow field of view. The ability to assess number of targets was affected by type of sensor and speed, the infrared sensor causing more problems than the visual sensor. Also, performance decreased at higher speed.

Keywords: Target detection · Visual sensor · IR sensor · UAV · Human factors

1 Introduction

The information explosion that comes from gathering information with new and better sensors is positive since users can access more information. At the same time it is a challenge to select the vital information in a specific situation. Data overload may be a serious problem, and it is necessary to have an understanding of the whole system. How to help human cognition using a medium (i.e. computer) is fundamental to ensure good user performance in the current situation. To build an effective system in a military setting a number of factors must be taken into account. From an ecological approach [1] and representation design [2] there is a cognitive triad between *domain or environment*, *interface* and *users*. There is a reciprocal coupling between the user and the environment which often is mediated by a user interface. The components of the triad are the cognitive demands from the domain or environment, user resources and limitations, and the interface. The interface effectiveness is determined by the mapping between the environment and interface (correspondence) and the mapping between the user and interface (coherence). To develop an effective and user friendly system all these three parts must be taken into account.

In many applications part of the information that reaches the user has been acquired with some type of sensor system, involving one or several sensors and signal processing,

© Springer International Publishing AG 2017
S. Yamamoto (Ed.): HIMI 2017, Part I, LNCS 10273, pp. 136–144, 2017.
DOI: 10.1007/978-3-319-58521-5_10

acting as a filter between the environment and the interface. In order to be able to understand the complete picture and study sensor-related aspects the model has to be extended.

In our research we aim for an understanding of the whole picture and therefore also sensor type are included since this is an important part of the system we work with (Fig. 1).

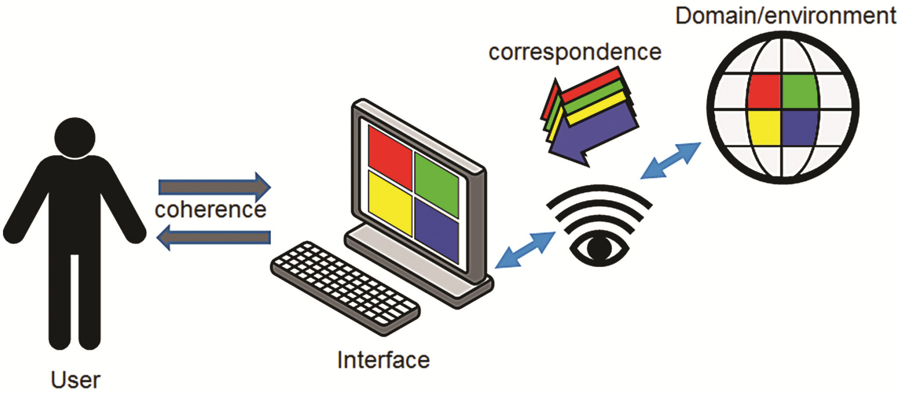

Fig. 1. The relation between user, interface, environment and sensor.

Even though the whole system must always be taken into account, the research presented here focuses on the ability and limitations of the users to extract the correct information about the environment based on data from sensors observing it. The purpose of this study was to investigate and compare performance of manual visual and IR-based detection of static targets seen from a simulated unmanned aerial vehicle (UAV).

It is not always certain what seeing an object means. One way to analyze observers' ability to perform visual tasks is to use the Johnson criteria [3, 4], often used by scientists who study the capability of sensors, e.g., infrared systems. A differentiation is made on *detection* (i.e. whether there is something of potential interest), *orientation* (which sometimes are excluded), *recognition* (e.g. the difference between a human and car) and *identification* (e.g. whether it is a friend or foe). The Johnson criteria proposes in detail how many pixels (originally line-pairs) an object needs to contain to make the classification possible. According to Johnson criteria, the detection distance is calculated based on how many pixels an object must contain. In order to detect static objects it requires 2 × 2 pixels, orientation 8 × 2.8 pixels, 8 × 8 pixels for recognition and identification required 12.8 × 12.8 pixels [5]. These figures should not be interpreted as guidelines but rather as values under the best possible conditions. It is not entirely clear in various descriptions but the described values probably mean that the task can be solved in 50% of the cases, which in most cases is unacceptably low. There is also a variety of factors that must be considered, including the contrast between objects and background, atmospheric disturbances, the number of objects in the picture, light, contextual clues, color and type of optics. Moreover, performance is affected by the type of task, the experience of the participants and their level of training for the specific task, motivation, and the relative importance between quick decisions and correct results [3]. There are several methods that can be regarded as further development of Johnson criteria, e.g. TOD (Triangle Orientation Discrimination), TTP (Targeting Task Performance) and TRM (Thermal Range Model). For further description of these methods see Näsström et al. [6], Wittenstein [7], and Vollmerhausen and Jacobs [8].

In order to assess and evaluate a system, experiments with users should be conducted. Theoretically calculated values could be of some value to get an indication of what objects that can be detected or identified (e.g. Johnson criteria) but is never enough. Identification of friend or foe is different from being able to identify who the person is and it is therefore important to be clear about exactly what you mean by different concepts. There is an obvious risk for confusion regarding the interpretation of concepts, since the concepts are used by researcher in different context without a standardized definition. In many situations one must be absolute sure about the identity of a person to make a decision whether to use military force. Friendly fire, where a soldier accidentally opens fire on his own troops, is a well-known phenomenon. In other cases, such as intelligence, is it important to describe what is seen according to a predetermined classification scheme and not just described what they think they see.

In this study the purpose was to investigate subjects' performance for visual and IR-based detection of targets seen from a simulated unmanned aerial vehicle (UAV).

2 Method

Participants watched synthetic video sequences captured from an UAV. All video sequences were generated by a sensor simulation system. The task was to detect and count stationary humans (referred to as targets in this paper). The participants also estimated how confident they were in their answers on a scale from 0 to 100%. A within-group design with *two visualizations (*visual and IR) × *two field of view* × *two UAV speeds* (12 m/s and 8 m/s) was used. The field of view (FOV)) used were 2.3 × 1.7° and 1.4 × 1.1°.

2.1 Subjects

Eight subjects (four women and four men) participated in the experiment. All had adequate vision with or without correction.

2.2 Apparatus

The video sequences were presented on a 21.5 inch full HD widescreen display with a resolution of 1920 × 1080 pixels. A PC with Windows 7, Intel® Core™ 2 Duo processor with 3 GHz and 4 GB of RAM memory was used. The software Matlab [9] was used on a separate computer to register response time when the participants clicked with the left mouse button. The response tool was designed so participants could always look at the screen where the stimuli material (video sequences) were presented.

2.3 Stimuli

A total of eight videos (640 × 480 pixels) were generated during a clear sunny day with shadows from targets on the ground to depict sensor information from a visual and IR sensor (Fig. 2). The overall mission was similar to a real UAV flying along a predefined path (Fig. 3) with humans (targets) standing on the ground.

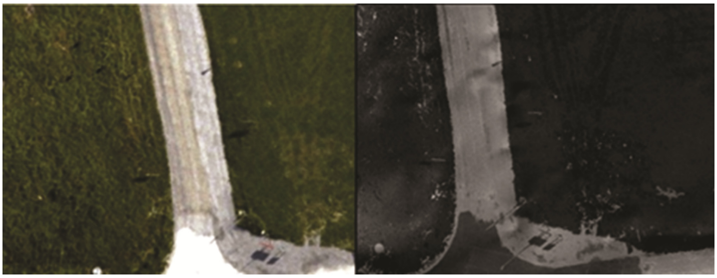

Fig. 2. Still images from the visual sensor (left) and the IR sensor (right).

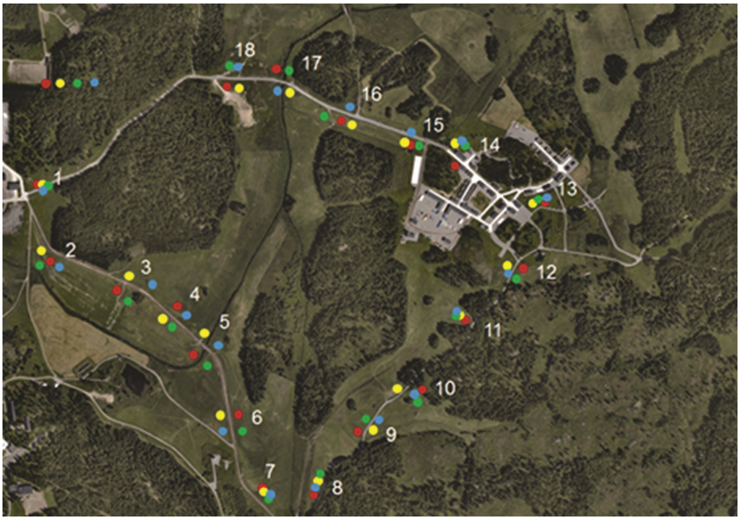

Fig. 3. Location of subjects on the ground at eighteen areas with different position for each of the four scenarios.

According to the mission, four scenarios were generated with the visual- and IR-sensor respectively. Each scenario had eighteen areas with different target positions. The same areas and positions were used for the visual- and IR scenarios. A total of eight videos were generated according to the aforementioned design. The visual and IR scenarios were presented in a balanced order between subjects', and within each sensor the four scenarios were presented in a randomized order.

2.4 Procedure

After welcoming the participants individually and briefing them about the experiment purpose and procedure they received written information and had the opportunity to ask questions to the experiment leader. Then an introduction was given to make sure that the participants were familiar with the situation and test material. They were introduced with both visual and IR image visualizations and received about ten minutes training. The participants watched the videos and answered by first pressing the left mouse button whereby the response time (RT) was recorded, then continued looking at the video and verbally answered how many targets they spotted and how confident they were (Fig. 4). The participants were instructed to always focus on the screen with the stimuli. Because the task was mentally demanding it was divided into eight separate videos with the possibility to rest before continuing with the next one.

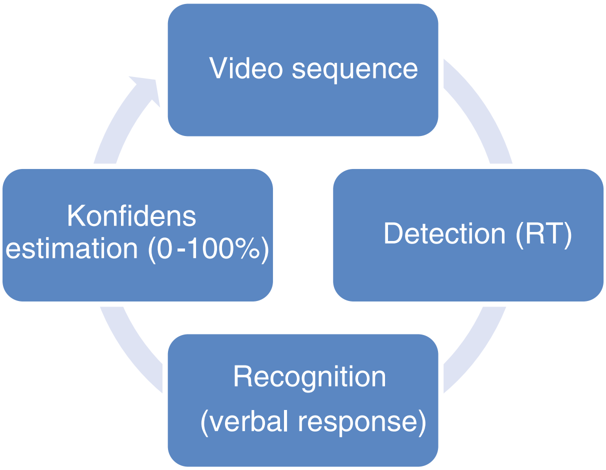

Fig. 4. Process for answering with video sequence, detection, recognition and confidence estimation.

3 Results

The results include statistical analysis of time to detect targets and estimation of number of targets with confidence estimations. The data were analyzed with a three-way ANOVA [10] with type of visualization (visual and IR), speed (12 m/s and 8 m/s), and field of view (1.4 × 1.1°). Tukey HSD were used for post hoc testing [11].

3.1 Detection

The ability to detect targets were measured by response time (RT) and analysis was performed by ANOVA repeated measures. The results showed a main effect for type of sensor F (1, 7) = 33.62, p < .001, where the response time of the visual sensor was shorter than for the IR sensor (Fig. 5).

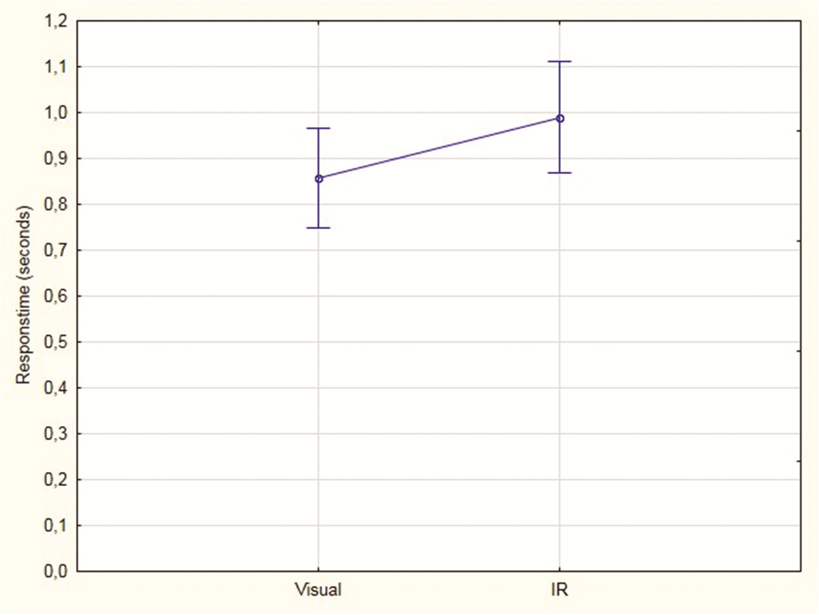

Fig. 5. Mean and standard error of mean for response time for visual and IR sensor.

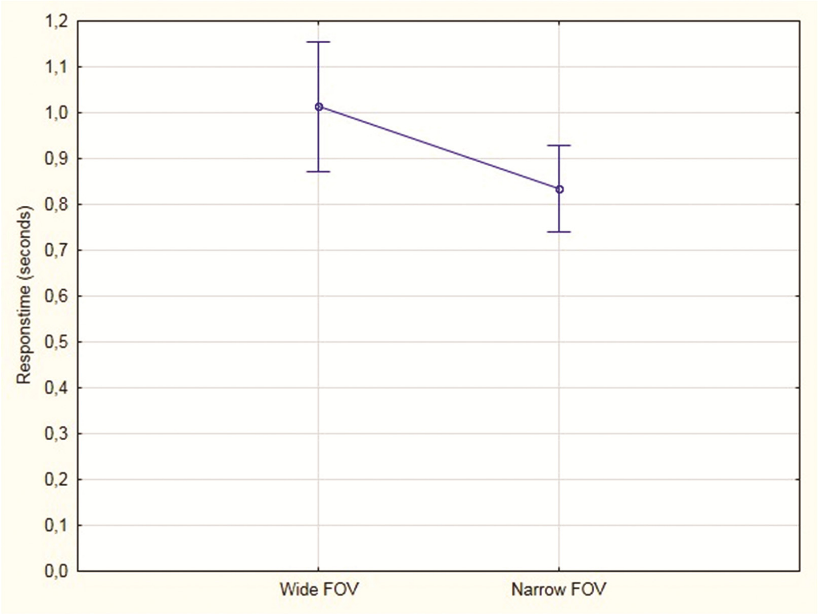

Fig. 6. Mean and standard error of mean for response time for wide and narrow FOV.

There was also a significant main effect of FOV, F (1, 7) = 19.36, p < .005 (Fig. 6), where the response time for the narrow FOV (more zoomed) gave faster response time than the wider FOV.

3.2 Estimation of Number of Objects with Confidence Estimation

The ability to assess number of targets was analyzed by ANOVA repeated measurement. Mean values for each condition was used for each participant and multiplied by the participants' confidence estimation. The results showed a main effect for type of sensor F(1, 7) = 13.73, p < .01, where participants assessed the number of targets more correct with visual than with IR sensor (Fig. 7).

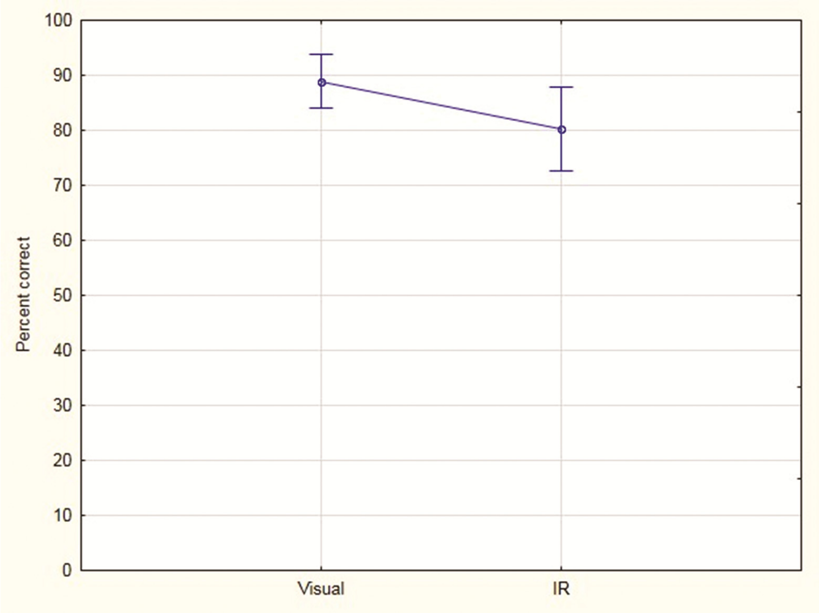

Fig. 7. Mean and standard error of mean for percent correct answers for visual- and IR-sensor.

There was also a significant main effect of UAV speed F(1, 7) = 52.53, p < .001, where participants assessed the number of targets more correct at low than at high speed (Fig. 8). There were no significant effect of FOV, and no interaction effects.

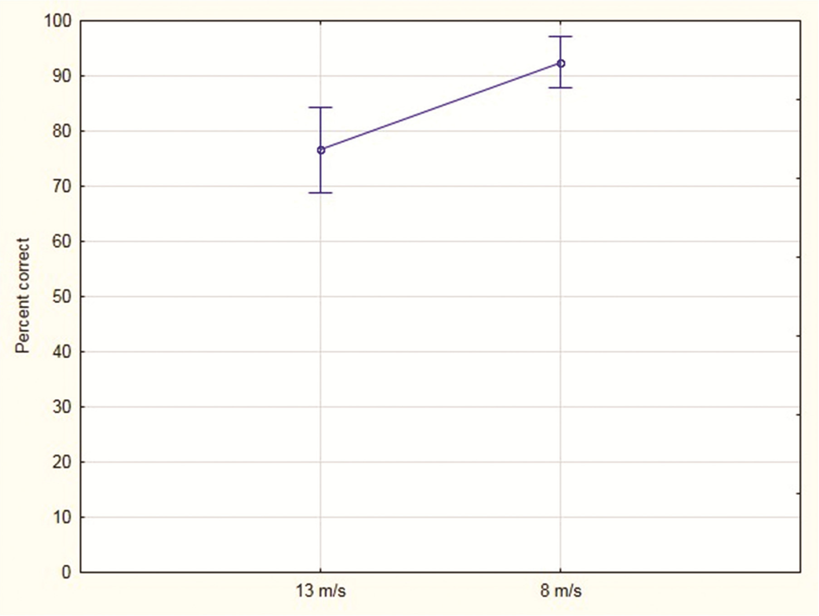

Fig. 8. Mean and standard error of mean for percent correct answers for 13 m/s and 8 m/s.

4 Discussion and Summary

The results from this study demonstrated that the ability to detect targets was affected by type of sensor and FOV. It took significantly longer time to detect targets with the IR sensor than with the visual sensor and it took significantly longer time with wider FOV than narrow FOV. The ability to assess the number targets was affected by type of sensor and speed, IR sensor causing more problems than the visual sensor. The ability to assess the number of targets was overall very high, at 8 m/s (93–99%) and relatively high at 12 m/s (79–91%).

This study also shows a method that can be used to investigate the users' performance related to detection of ground targets from an UAV. The task was to detect and count number of humans on the ground and make confidence estimations. The task can be modified to identify different type of targets, e.g. to investigate possible differences between target types or classify objects as friend or foe.

One limitation of this study is that although the visual and IR sensor data are realistic, no scientific verification has been made to confirm the similarity between the used stimuli material and real data from sensors. However, one researcher compared the simulated videos with real sensor information and confirmed that the material looked similar. In the future, this procedure need to be improved with objective measures.

This experiment was the first experiment in a series of planned experiments, where different platform speeds and field of view angles were studied. Another possibility for further studies is to use other missions, e.g. manually control the UGV instead of using predefined paths. In this experiment it was daytime in strong sunshine, which gave clear shadows of people and objects. It would be interesting to compare the results achieved in this study with results from a daytime scenario with cloudy weather without clear and sharp shadows visible. Also, night time scenarios' would be interesting to investigate. Another possibility is not to only examine human performance of detection, but also examine recognition and identification. The follow-up experiment (now in progress) investigates recognition of various vehicles in a similar setting as the experiment presented here.

References

1. Flach, J., Hancock, P.: An ecological approach to human-machine systems. In: Proceedings of the Human Factors and Ergonomics 36th annual meeting, Santa Monica, CA, USA, pp. 1056–1058 (1992)
2. Woods, D.D.: Towards a theoretical base for representation design in the computer medium: ecological perception and aiding human cognition. In: Flach, J., Hancock, P., Caird, K., Vicente, K. (eds.) An Ecological Approach to Human Machine Systems I: A Global Perspective, pp. 157–188. Erlbaum, Hillsdale (1995)
3. Donohue, J.: Introductory review of target discrimination criteria. Philips laboratory Air force system command, Wilmington, MA, US (1991)
4. Johnson, J.: Analysis of image forming systems. Technical report, U.S. Army Engineer Research and Development Laboratories, Fort Belvoir, Virginia, US (1958)
5. Kopeika, S.: Contrast-limited resolution and target acquisition. In: A System Engineering Approach to Imaging, vol. PM38. SPIE (1998)
6. Näsström, F., Bergström, D., Bissmarck, F., Grahn, P., Gustafsson, D., Karlholm, J.: Prestandamått för sensorsystem (FOI-R–4139–SE). Linköping: Sensor och TK-system (2015). (In swedish)
7. Wittenstein, W.: Thermal range model TRM3. Paper presented at the SPIE Conference in Infrared Technology and Application XXIV, San Diego (1998)
8. Vollmerhausen, R., Jacobs, E.: The targeting task performance (TTP) metric. Technical report AMSEL-NV-TR-230, Modeling and Simulation Division, Fort Belvoir, VA (2004)
9. MATLAB (2017). https://se.mathworks.com/products/matlab.html
10. Hays, W.L.: Statistics. Harcourt Brace College Publishers, Fort Worth (1994)
11. Greene, J., D'Oliveira, M.: Learning to Use Statistical Tests in Psychology. Open University Press, Milton Keynes (1982)