User Performance for Vehicle Recognition with Visual and Infrared Sensors from an Unmanned Aerial Vehicle

Patrik Lif(✉), Fredrik Näsström, Fredrik Bissmarck, and Jonas Allvar

Swedish Defence Research Agency, Linköping, Sweden
{patrik.lif,fredrik.nasstrom,fredrik.bissmarck,
jonasallvar}@foi.se

Abstract. In many situations it is important to detect and recognize people and vehicles. In this study the purpose was to examine human performance to detect and recognize vehicles on the ground from synthetic video sequences captured from a simulated unmanned aerial vehicle. A visual and an infrared sensor was used on an unmanned aerial vehicle with camera scan rate of the field of view on the ground relative to the ground of either 8 m/s or 12 m/s. The results from this study demonstrated that performance was affected by type of sensor, camera scan rate and type of vehicle. Subjects performed worse with infrared than with visual sensor and increased camera scan rate caused more errors. Also, the results show that recognition performance varied between 67 and 100% depending on type of vehicle. Recognition of specific vehicles was also affected negatively by interference from vehicles of similar appearance. Consequently, a vehicle with unique appearance within the set was easier to recognize.

Keywords: Vehicle recognition · Visual sensor · IR sensor · UAV
Human factors

1 Introduction

Gathering information with new and better sensors is positive since users can access more information, but it is necessary to have an understanding of what is the most vital information in a given situation. To accomplish this, users' need a good understanding of the whole system. Data overload may be a serious problem, and how to help human cognition using e.g. computers is fundamental to ensure good situation awareness and good user performance. Regardless of type of system it is also necessary to have a good understanding of the user and the context. The ecological approach [1] and representation design [2] describes a cognitive triad between *environment*, *interface* and *users*. There is a reciprocal coupling between the user and the environment, which often is mediated by a user 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. Information that reaches the user has often been

M. Kurosu (Ed.): HCI 2018, LNCS 10901, pp. 295–306, 2018.
https://doi.org/10.1007/978-3-319-91238-7_25

acquired with some type of sensor system that involves signal processing, acting as a filter between the environment and the interface. In order to be able to understand the complete picture of study sensor-related aspects the model has to be extended to also include environment, interface and human aspects. Since the sensor is a central part in our research, we add the sensor to the representation visualization (Fig. 1).

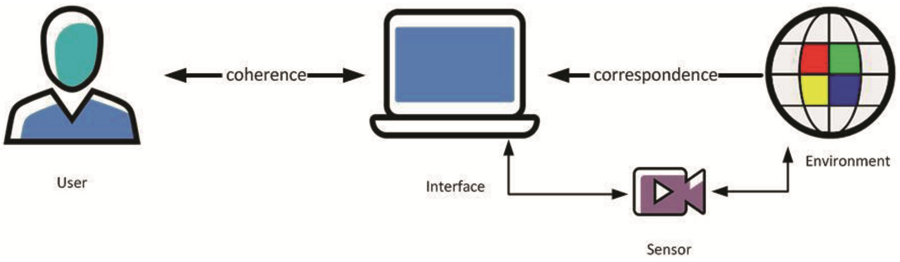

Fig. 1. The relation between user, interface, environment and sensor. Icons were adopted from Iconshock [3].

Even though the whole system always have to be taken into account, the main focus here is on the ability and limitations of the users and their performance to extract correct information from sensor data.

Seeing an object could mean different things, but one way to analyze observers' ability to perform visual tasks is to use the Johnson criteria [4, 5] that distinguish between *detection* (i.e. whether there is something of potential interest), *recognition* (e.g. the difference between a human and car) and *identification* (e.g. whether it is a friend or foe). According to Johnson criteria, possible 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 [6]. However, this should be interpreted as values under best possible conditions. 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 [4]. Also the methods Triangle Orientation Discrimination (TOD), Targeting Task Performance (TTP) and Thermal Range Model (TRM) could be considered. For further description of these methods see Näsström et al. [7], Wittenstein [8], and Vollmerhausen and Jacobs [9].

Even though theoretically calculated values (e.g. Johnson criteria) could be of some value to get an indication of what objects that can be detected, recognized or identified, experiments with users should be conducted to get a better understanding of a real situation. There is an obvious risk of confusion regarding the interpretation of concepts, since the concepts are used by researchers in different context without a standardized definition. It is absolutely necessary to clarify and define the concepts used.

Identification of friend or foe is different from actual identification of a face from memory or a database. In many situations one must be absolutely certain about the identity of a person or vehicle to make a decision whether to use military force. Also, it is necessary to have a good understanding of roles of engagement (ROE), when military force can and cannot be used. Friendly fire, where a soldier accidentally opens fire on his own troops, is a well-known phenomenon that must be avoided. In other cases, such as intelligence, is it important to describe what is seen according to a predetermined classification scheme and not just describe what users think they see.

In military contexts, it is sometimes important to find a particular type of vehicle among other similar military vehicles, and it is also important to distinguish between military and civilian vehicles. To increase knowledge about this, our work involves assessing actual sensor performance but also investigating how operators use and interpret sensor information. Even though the interest from a human factors perspective is mainly on user performance to detect and recognize people and vehicles, we also conduct technology driven sensor studies [10] and thorough investigation of the real setting [11]. There are many interesting studies focusing on detection, recognition and identification. Colomina and Molina [12] discuss the evolution and use of unmanned aerial systems in photogrammetry and remote sensing that can be used in both military and civilian operations, e.g. search and rescue missions. Other research with unmanned aerial vehicle (UAV) and target detection focus has a more technical approach, e.g. develop algorithms for autonomous target detection [13] or autonomous UAVs for search and rescue [14]. There are also interesting studies using multiple cooperative vehicles [15] or a swarm of unmanned vehicles [16] which shows that multiple vehicles can improve performance. Other research has a clearer connection to human factors issues and user performance. Hixson et al. [17] used soldiers to investigate the relation between performance in the laboratory and in the field for tasks including detection, recognition and identification. The results shows that perception laboratory performance using real or simulated imagery relates well to imagery performance in the field.

The research question in the first experiment was to investigate how fast and to what degree of correctness can users detect and recognize one selected military vehicle among other similar vehicles and how is performance affected by type of sensor, camera scan rate of the field of view on the ground (hereafter referred to as scan rate) and distance? The research question in the second experiment was to investigate to what degree of correctness can users recognize eight military vehicles with an infrared sensor, at camera scan rate of 8 m/s at a distance of 400 meters?

It is important to investigate and understand the sensors' pros and cons in different situations. Only the infrared sensor can be used at night while both the visual and infrared sensor can be used during daytime. However, it is not obvious which sensor is preferred during daytime in different situations and it is therefore important to investigate this. In some situations it is certainly better to use the visual sensor, but sometimes the vehicle can be partly hidden under e.g. branches or trees and then it is advantageously to use the infrared sensor also during daytime. From a tactical perspective it may be advantageous to fly the unmanned aerial vehicle at night, but then only the infrared sensor can be used. Also, at night there are significantly fewer civilian vehicles in motion and less vehicles that gives heat signatures which facilitates detection and recognition of military

vehicles. If performance decreases with one of the sensors quantification would be important. It is preferable to use high camera scan rate since larger geographic areas can be covered, but if it results in decreased performance it may be necessary to use a lower speed. Even though user performance is expected to decline at increased camera scan rate it is important to objectively quantify performance decrease. If the unmanned aerial vehicle fly at high altitude there are tactical advantages such as lower risk for the UAV being detected, but if it results in decreased performance it is not recommended.

Here, two experiments were conducted that is part of a larger study where the overall goal is to investigate how different sensors should be used in unmanned aerial vehicles to gather information. The purpose with these two experiments were to investigate subjects' performance of vehicle detection and recognition from a simulated unmanned aerial vehicle. In the first experiment, detection and recognition of one selected vehicle among a total of eight vehicles was investigated at two different camera scan rate (seen from the UAV) with visual- and IR-sensor. In the second experiment recognition of all eight vehicles was investigated at a camera scan rate of 8 m/s with an IR-sensor. Although the results here are only presented and analyzed strictly linked to these experiments, later it can be analyzed and compared to other experiments. Also, this information can be used to better understand how information from different sensors can be aggregated to increase performance. However, this is not the focus here and is therefore not presented in this paper.

2 Experiment 1 – Detection and Recognition of Selected Vehicle

In the first experiment, detection and recognition of one selected vehicle among a total of eight vehicles was investigated.

2.1 Method

Participants watched synthetic video sequences captured from an UAV. All video sequences were generated by a sensor simulation system [10]. The task was to detect and recognize a selected vehicle among other vehicles. A within-group design with *two visualizations (*visual and IR) × *two distances (400 and 520 meters)* × *two camera* scan rate (8 and 12 m/s) was used.

Subjects

Twelve subjects (5 women and 7 men) between 25 and 48 years participated in the experiment. Half of the participants' had military background and the other participants' were well acquainted with military activities through their civilian jobs. However, none of the participants were experts on the vehicles presented in these experiments and therefore trained prior to the experiments started. All had adequate vision with or without correction.

Apparatus

The video sequences were presented on a Dell Latitude 7240 with 12.5 inch display with a resolution of 1366 × 768 pixels. The computer had 4th generation Intel® Core i5 and i7. A self-developed software was used to present stimuli and record participant's response time.

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 infrared sensor (Fig. 2). The overall mission was similar to a real UAV flying along a predefined path with vehicles stationary on the ground.

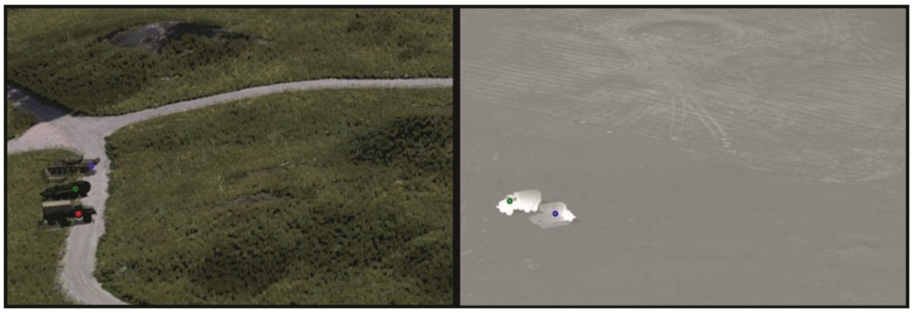

Fig. 2. Still images from the visual sensor (left) and the IR sensor (right).

The task was to detect and recognize one selected vehicle among a total of eight vehicles. The eight vehicles were BMP-3, BTR-80, MT-LB, SA-19, T-72, TOS-1, Ural 4320 Ammunition truck, and Ural 4320 fuel truck (Fig. 3).

Fig. 3. The eight vehicles used.

Four scenarios were generated with the visual- and infrared sensor respectively. Each scenario had 18 areas with different target positions. The same areas and positions were used for the visual- and infrared scenarios. A total of eight videos were generated according to the aforementioned design. The visual- and infrared scenarios were presented in a balanced order between subjects', and within each sensor the four scenarios were presented in a randomized order.

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 infrared image visualizations and received training, which consisted of two three minutes scenarios, one for visual- and one for infrared stimuli. The participants watched the videos and answered by first pressing the space bar whereby the response time (RT) was recorded and then used the left mouse button to annotate in the image to indicate the selected vehicle position. The annotation was later used to calculate number of correct answers. 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.

2.2 Results

The results include statistical analysis of time to detect targets and recognition of the selected vehicle. The data were analyzed with a three-way ANOVA [18] with type of visualization (visual and infrared), camera scan rate (8 and 12 m/s), and distance (400 and 520 meters). Tukey HSD was used for post hoc testing [19].

Detection
The ability to detect targets was measured by response time (RT) and analysis was performed by ANOVA repeated measures. The results showed no significant main effects of response time ($p > .05$).

Recognition of one selected vehicle
The ability to recognize one selected vehicle was analyzed by ANOVA repeated measurement, where mean values for each condition was used for each participant. The results showed a main effect for type of sensor $F(1, 11) = 9.02$, $p < .05$, where participant's performance were lower with the infrared sensor than with the visual sensor (Fig. 4).

There was also a significant main effect of camera scan rate $F(1, 11) = 8.75$, $p < .05$, where higher camera scan rate caused more errors (Fig. 5). There was no significant main effect of distance, and no significant interaction effects $p > .05$.

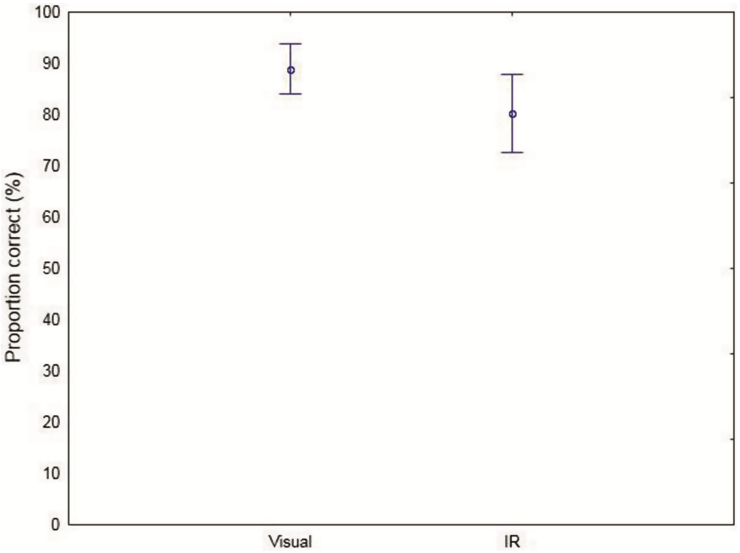

Fig. 4. Mean and standard error of mean for proportion correct answers for visual- and infrared (IR) sensor.

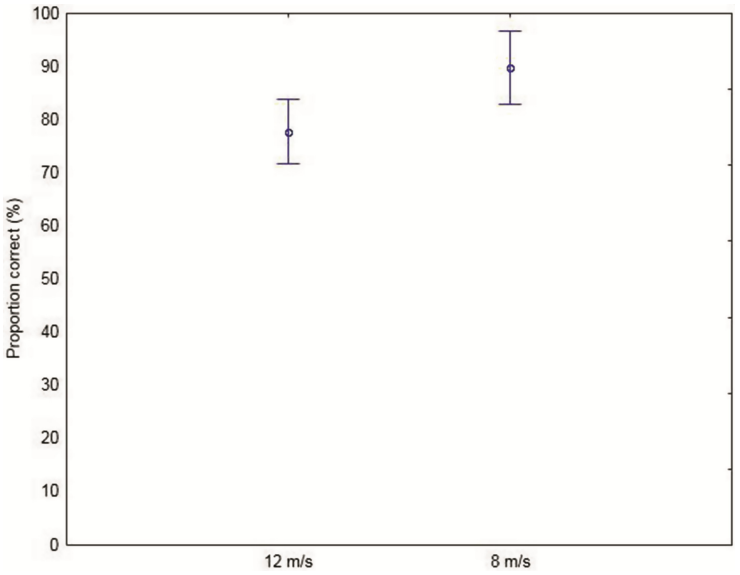

Fig. 5. Mean and standard error of mean for proportion correct answers for 12 m/s and 8 m/s.

3 Experiment 2 – Recognition of Eight Vehicles

In the second experiment, detection and recognition of a total of eight vehicles were investigated.

3.1 Method

From Experiment 1, the scenario with an infrared sensor, distance of 400 meters, and camera scan rate 8 m/s was selected. For this setting recognition of eight different vehicles was investigated. In this experiment the focus was on proportion correct recognized vehicles only. The subjects' watched the video sequences for five seconds and then reported their answers, no response time was measured.

Subjects
Twelve subjects (4 women and 8 men) participated in the experiment. Five of the participants' had military background and the other participants' were well acquainted with military activities through their civilian jobs. However, none of the participants were experts on the vehicles presented in these experiments and therefore trained prior to the experiments started. All had adequate vision with or without correction.

Apparatus
See experiment one for technical description. Superlab [20] was used to present the video sequences and to record proportion correct answers.

Stimuli
Two video sequences (640 × 480 pixels) with a total of nine stops, where the subject's task was to recognize vehicles from a total of eight vehicles. The same eight vehicles were used as in experiment 1, but in this experiment the task was to identify all eight vehicles, not only one selected vehicle.

Procedure
Overall the procedure was as in experiment 1. However, there were some differences due to another design. In experiment 2, the video sequences was paused at predefined occasions, and one vehicle was indicated by a circle. The subject's answered by pressing the number 1–8 on the keyboard, and then the next stimuli was indicated by a circle. The procedure was repeated 1–4 times at each scenario stop depending on the number of vehicles in that particular stop.

3.2 Results

The results include statistical analysis of recognition of the eight vehicles, which are analyzed with one-way ANOVA [10]. Tukey HSD was used for post hoc testing [11]. The results showed a significant effect of vehicle type $F(7, 77) = 5.54$, $p < .001$ (Fig. 6).

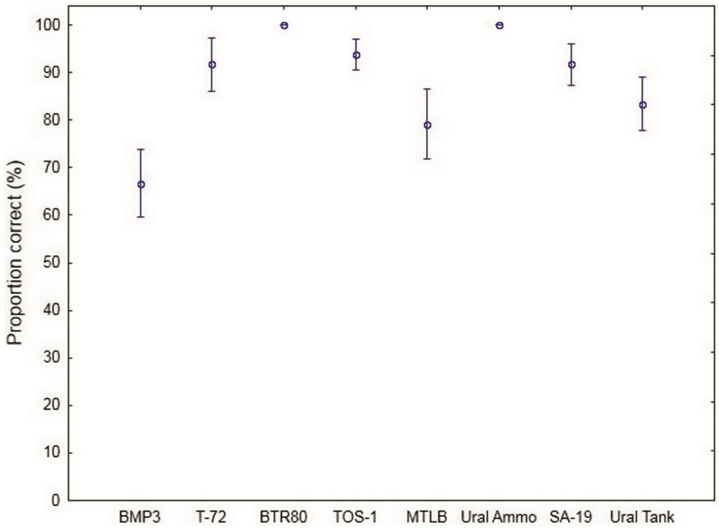

Fig. 6. Mean and standard error of mean for proportion correct answers of the eight vehicles.

Tukey HSD showed that BMP-3 was significantly harder to recognize than most of the other vehicles (except the MT-LB). Vehicles T-72, BTR-80, TOS-1, Ural 4320 ammunition truck, and SA-19 were recognized in 90% of cases or more, while BMP-3, MT-LB and Ural 4320 fuel truck were more difficult. BMP-3 was mainly confused with BTR-80 and MT-LB. MT-LB was mainly confused with BTR-80, and Ural 4320 fuel truck was mainly confused with Ural 4320 ammunition truck. For a more detailed description see the confusion matrix (Table 1). The first column show the vehicle name and the second column show percentage correct recognition. Column three to eight show which other vehicles the vehicle (in first column) was confused with.

Table 1. Confusion matrix that shows which other vehicles the eight vehicles were confused with. All numbers are presented in percent, and each row add up to 100%

Vehicle	% correct	BMP-3	T-72	BTR-80	TOS-1	MT-LB	Ural 4320 ammunition	SA-19	Ural 4320 fuel truck
BMP-3	66,6		4,2	12,5		12,5		4,2	
T-72	91,7		8,3						
BTR-80	100								
TOS-1	93,8						4,2	2,0	
MT-LB	79,1	4,2		12,5					4,2
Ural 4320 Ammunition	100								
SA-19	91,7	2,8			5,5				
Ural 4320 fuel	83,3						16,7		

4 Discussion

The experiments presented here is part of a larger study where user performance to detect and recognize people and vehicles is investigated, technology driven sensor studies are performed [10], and thorough investigation of the real setting [11] conducted. The combination of improved technical knowledge, understanding of the real environment, and user performance is seen as a good interdisciplinary combination to better understand how a final system can make a difference in real settings.

The purpose with these two experiments were to investigate subjects' performance of vehicle detection and recognition from a simulated unmanned aerial vehicle. The results shows that the ability to recognize vehicles is affected by type of sensor, camera scan rate, and type of vehicle that is to be recognized. User performance to recognize the selected vehicle among a total of eight vehicles was significant lower with the infrared- than with the visual sensor, and significant lower at camera scan rate 12 m/s than at 8 m/s. Also, the results show that recognition performance varied between 67% and 100% depending on type of vehicle. The results from the second experiment clearly shows that vehicle recognition with the infrared sensor is problematic, even though short distance (400 meters) and slow camera scan rate (8 m/s). The results also show that certain types of vehicles are particularly difficult to recognize which is an important operational information in military contexts.

In situations where the vehicles are placed in open terrain as in these experiments, it is advantageous to use the visual sensor. However, the infrared sensor allows detection of vehicles in situations where there is no clear view, in situations with low visibility and at night. In these situations the heat signature from the infrared sensor can be used to detect and recognize vehicles. Another possibility is to combine information, either by switching between the two sensor images or by fusing the sensor images into one image. However, this was not investigated in these experiments and therefore not reported here.

From a scientific perspective it is important to understand perceptual and cognitive possibilities and limitations. As a part of this we investigated how the type of information presented (visual and infrared), camera scan rate and distance affected user performance. Although there are a number of other factors that affect performance, this contributes to knowledge about vehicle detection and recognition in this military context. For practical and economic reasons it is not always possible to conduct field studies and therefore laboratory studies can be used as an important compliment. The results presented here can also be correlated with results from similar field studies (not yet performed). Results from laboratory experiments are especially valuable if they can predict performance in real environments, which for us is a future challenge.

It is also important to use systematic methods for data collection and result analysis, which gives the possibility to compare and analyze the results relative to other scientific results. The results from our experiments can later on be analyzed and correlated with calculated values from e.g. the Johnson criteria or relative to sensors' technical performance to get an understanding of correlation between human performance and technical performance. In this study, no specific sensors have been presented, but the results from

this work can be used for evaluating the existing sensors that the simulation here were based on. This work remains to be done and is therefore not presented here.

Even though these experiments and prior experiments [21] gives a good understanding of user performance to detect and recognize people and vehicles from an unmanned aerial vehicles there are some limitations. In these studies we used a predefined flying path, as often is the case in real settings, but it would be interesting to let the users manually control the sensor direction and give them the possibility to zoom-in to targets. Also, in these experiments it was daytime and strong sunshine, which gave clear shadows of vehicles. 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. In this study, vehicles were placed on open surfaces in the terrain, but it would be interesting to see how different camouflage (such as nets or trees) would affect the ability to recognize vehicles especially with visual camera.

Detection and recognition was investigated in this study, but it would be interesting to also investigate identification of people and vehicles in a similar setting as the experiments presented here. One limitation of this study is that although the visual and infrared 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 [10]. In the future, this procedure need to be improved with standardized objective measures.

The information presented from this study is important since user performance and technical knowledge can be aggregated and used to understand operational performance and limitations. Issues such as camera scan rate, type of sensor, flight altitude, weather conditions, and time of day, are important and can be put in a broader context to understand how a task best can be solved.

References

1. Flach, J.M., Hancock, P.A.: An ecological approach to human-machine systems. In: Proceedings of the Human Factors Society Annual Meeting, vol. 36, no. 14, pp. 1056–1058 (1992)
2. Woods, D.: Toward 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 Ecologixal Approach to Human Machine System, Erlbaum, Hillsdale, New Jersey, pp. 157–188 (1995)
3. Iconshock: "Icon man," Smashingmagazine (2017). SmashingMagazine.com
4. Donohue, J.: Introductionary review of target discrimination criteria. Wilmington, MA (1991)
5. Sjaardema, T.A., Smith, C.S., Birch, G.C.: History and Evolution of the Johnson Criteria. Albuquerque, New Mexico (2015)
6. Kopeika, N.S.: A System Engineering Approach To Imaging. SPIE Optical Engineering Press (1998)
7. Näsström, F., Bergström, D., Bissmarck, F., Grahn, P., Gustafsson, D., Karlholm, J.: Prestationsmått för sensorsystem, Linköping, FOI-R–4139–SE (2015). (In swedish)

8. Wittenstein, W.: Thermal range model TRM3. In: Proceedings Volume 3436, Infrared Technology and Applications XXIV, vol. 3436, p. 413 (1998)
9. Vollmerhausen, R.H., Jacobs, E.: The Targeting Task Performance (TTP) Metric A New Model for Predicting Target Acquisition Performance. Fort Belvoir, VA, Technical Report AMSEL-NV-TR-230, 2004
10. Näsström, F., Allvar, J., Deleskog, V.: Simulation framework for research of 'intelligent' reconnaisance systems (2017)
11. Näsström, F., et al.: Värdering av sensorsystem, Linköping, FOI-R–4474–SE (2017). (In swedish)
12. Colomina, I., Molina, P.: Unmanned aerial systems for photogrammetry and remote sensing: a review. ISPRS J. Photogramm. Remote Sens. **92**, 79–97 (2014)
13. Hinas, A., Roberts, J.M., Gonzalez, F.: Vision-based target finding and inspection of a ground target using a multirotor UAV system. Sensors **17**(12), 2929 (2017)
14. Sun, J., Li, B., Jiang, Y., Wen, C.: A camera-based target detection and positioning UAV system for search and rescue (SAR) purposes. Sensors **16**(11), 1778 (2016)
15. York, G., Pack, D.J., York, G., Pack, D.J.: Ground target detection using cooperative unmanned aerial systems. J. Intell. Robot. Syst. **65**, 473–478 (2012)
16. Altshuler, Y., Pentland, A., Bruckstein, A.M.: The cooperative hunters – efficient and scalable drones swarm for multiple targets detection. Swarms and Network Intelligence in Search. SCI, vol. 729, pp. 187–205. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-63604-7_7
17. Hixson, J.G., Teaney, B.P., May, C., Maurer, T., Nelson, M.B., Pham, J.R.: Virtual DRI dataset development. In: Proceedings of Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXVIII, vol. 10178 (2017)
18. Hays, W.: Statistics, 5th edn. Harcourt Brace College Publishers, New York (1994)
19. Greene, J., D'Oliveira, M.: Learning To Use Statistical Tests In Psychology. Open University Press, Philadelphia (1982)
20. Cedrus: "SuperLab 5," Cedrus (2017). https://www.cedrus.com/superlab/. Accessed 20 Nov 2017
21. Lif, P., Näsström, F., Tolt, G., Hedström, J., Allvar, J.: Visual and IR-based target detection from unmanned aerial vehicle. In: Yamamoto, S. (ed.) HIMI 2017. LNCS, vol. 10273, pp. 136–144. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58521-5_10