Conspicuity and Identifiability: efficient calibration tools for synthetic imagery

Alexander Toet, Maarten A. Hogervorst, Piet Bijl
TNO Human Factors, Kampweg 5, 3769 DE Soesterberg, The Netherlands

ABSTRACT

We argue that visual conspicuity and identifiability are two efficient task-related measures that can be deployed to calibrate synthetic imagery that is intended to be used for human visual search and detection tasks. The conspicuity of a target is operationally defined as the region around the center of the visual field where the target is capable to attract visual attention. Visual conspicuity predicts human visual search performance in realistic and military relevant complex scenario's. Conspicuity can easily and quickly be measured either in the field (in complex environments) or in the lab. This eliminates the need for costly and time consuming visual search experiments. The agreement between field and lab measurements implies that conspicuity can be used to validate synthetic imagery. Target identifiability is operationally defined as the amount of Gaussian blur that is required to reduce the target signature to its identification threshold. It is an efficient metric that can be used to gain insight into human identification performance without having to resort to elaborate and costly experiments. Identifiability is directly related to PID-performance, and is therefore well suited for comparing synthetic and realistic imagery. We conclude that synthetic imagery can be calibrated for human visual search and detection tasks by setting the conspicuity and identifiability of targets equal to those of their real world counterparts.

Keywords: identifiability, conspicuity, simulation, validation, human perception, synthetic imagery

1. INTRODUCTION

Synthetic imagery is now widely used for a variety of military applications like training and evaluation of camouflage and sensor systems. The generation of synthetic imagery generally involves a range of computational approximations and simplifications of the physical processes involved in the image formation, in order to meet the update rates in real-time systems or simply to achieve reasonable computation times. These approximations may reduce the fidelity of the resulting imagery. However, the degree of fidelity and the level of detail that is actually required in practice depends on the task for which the synthetic imagery has been generated. For instance, when performing a driving task, the color and definition of details like buildings, clouds, trees and grass are irrelevant (these need not necessarily be physically or perceptually correct), as long as the road can be can be seen just as well as in reality. Many military tasks involve human visual search, detection and identification. Synthetic imagery used for training these tasks should ideally result in the same observer performance that is obtained in the field. This implies that the synthetic imagery should be perceptually calibrated with respect to human visual search, detection and identification.

Synthetic image calibration is typically done by performing the same perceptual task both in the lab on the synthetic imagery and in the field, using similar scenes and scenarios in both cases. However, in most cases this results in costly and time consuming experiments, requiring large numbers of observers and resources to obtain statistical significance. In some cases, it may even be too dangerous or impractical to perform real life experiments. Hence there is a need for efficient task performance measures that can be used to quickly assess the fidelity of synthetic imagery with respect to a given observer task.

Targets and Backgrounds X: Characterization and Representation, edited by
Wendell R. Watkins, Dieter Clement, William R. Reynolds, Proceedings of SPIE
Vol. 5431 (SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.539118

We have recently developed two efficient task-related measures that can be deployed to calibrate synthetic imagery that is intended to be used for human visual search and detection tasks: visual conspicuity and visual identifiability.

Visual conspicuity predicts human visual search performance in realistic and military relevant complex scenario's. Conspicuity can easily and quickly be measured either in the field (in complex environments) or in the lab. Only a few observers (typically 2--3) are needed to achieve sufficient accuracy. Also, conspicuity measured in the lab on photographic slides agrees with conspicuity measured in the field. This eliminates the need for costly and time consuming visual search experiments, that typically require in the order or 100 observers and many repetitions to obtain statistically significant results.

Identifiability is defined as is the amount of blur required to reduce the target signature to its identification threshold. We determined the metric for a large set of corresponding synthetic and real FLIR images representing 10 different military vehicles registered at various angles and distances. The results show that the new metric correlates strongly with the probability of identification. Identifiability can quickly be determined using a simple and standard staircase procedure.

The structure of the rest of this paper is as follows. In the next two sections we introduce the concepts of respectively conspicuity and identifiability. For both metrics we discuss the corresponding measurement procedures, and we present some validation results and example measurements. Finally, we argue that both metrics are efficient tools to calibrate simulators that are used to train human visual search and detection tasks.

2. CONSPICUITY

2.1. Introduction

A target will be easier to detect when it stands out more from its background. It is therefore a priori likely that a measure that captures a target's visual distinctness as perceived by a human observer should correlate with human visual search performance.

TNO Human Factors developed a psychophysical procedure to quantify the visual conspicuity of a target in a complex scene[4-6]. In this approach, target conspicuity is operationally defined as the maximal lateral distance between target and eye-fixation at which the target can be distinguished[7]. This conspicuity measure can quickly be determined in situ, and can be used with full prior knowledge of the target and its location in the scene. It characterizes the extent to which a target stands out from its immediate surroundings.

This section introduces the concept of visual conspicuity and the associated measurement procedure.

2.2. Target conspicuity

The conspicuity area of a target is defined as the region around the center of the visual field where the target is capable to attract visual attention, because it is perceived as significantly distinct from its local background. Figure 1 illustrates the concept of the conspicuity area. This figure shows two subjects in a wooded environment. Both subjects have different conspicuity areas. The subject on the left is relatively conspicuous because she wears a bright retro-reflective jacket that has a high luminance and color contrast with the local background. The subject on the right is less conspicuous because his sweater has minimal luminance and color contrast with the local background. The bright areas represent the conspicuity areas of the subjects in the center. A target can be distinguished for any eye-fixation inside its corresponding conspicuity area. The observer (lower left) is using the patented TNO conspicuity meter to measure the conspicuity of the subjects.

Figure 1 Illustration of the concept of conspicuity area. The scene shows two subjects (indicated by the arrows) with different conspicuity areas (indicated by the bright areas). The subject on the left is conspicuous because she wears a bright retro-reflective jacket that has a high luminance and color contrast with the local background. The subject on the right is less conspicuous because his sweater has minimal luminance and color contrast with the local background. The brighter areas represent the conspicuity area of the subjects in the center. The observer (lower left) is using the patented TNO conspicuity meter to measure the conspicuity of the subjects.

2.3. Measurement procedure

To determine the visual conspicuity of an object in a background one needs to quantify the degree to which it attracts attention. The conspicuity measurement procedure introduced here is as follows. First, the observer visually inspects (foveates) the target. This is done to ensure that the observer knows to which object in the scene he should direct his attention. This is especially relevant for complex scenes where many details compete for the observer’s attention. Next, the observer fixates a point in the scene that is both (a) at a large angular distance from the target location, and (b) at a viewing distance that is comparable to the viewing distance of the target. When the observer fixates this point, the target should be positioned far in his peripheral visual field, such that it can not be distinguished. The observer then successively fixates locations in the scene that are progressively closer to the target location, until he can perceive (distinguish) the target in his peripheral field of view. The successive fixation points are along a line through the initial fixation point and the center of the target. The angular distance between the fixation location at which the target is first noted and the center of the target is then recorded. The measurement is repeated at least three times. Subjects are usually able to make a setting within one minute. The mean of the distances thus obtained is adopted as the characteristic spatial extent of the conspicuity area of the target, in the direction of the initial fixation point.

We define two different criteria to determine whether the target can be perceived or not. The first criterion is whether there is anything at the location of the target with a visual signature that differs from the local background. This criterion results in a detection conspicuity distance. The second criterion that can be used is whether the spatial structure at the location of the target indeed resembles the target, or can actually be identified as the target. This criterion yields an identification conspicuity distance.

2.4. Examples

Conspicuity measurements involve a free viewing procedure. This section describes experimental techniques that can be used in laboratory and field situations, and presents examples of conspicuity measurements in both conditions.

The stimuli used in the laboratory experiments reported in the rest of this section are photographic slides[3]. The slides were taken during the DISSTAF (Distributed Interactive Simulation, Search and Target Acquisition Fidelity) field test, that was designed and organised by NVESD (Night Vision & Electronic Sensors Directorate, Ft. Belvoir, VA, USA) and that was held in May and June 1995 in Fort Hunter Liggett, Califonia, USA. Each scene represents a military vehicle in a complex rural background. The conspicuity of the vehicles was also measured in the field during the DISSTAF trials. This enables a direct comparison of the lab and field results.

2.5. Field measurements

When measuring the conspicuity of objects in real world scenes the observer should first fixate the target and note its characteristic details. This may help to distinguish the target from its surround during the actual conspicuity measurement. Then, the observer should make a rough estimate of the minimal threshold angle in his peripheral visual field at which the target can no longer be perceived. He should then fixate a detail in the scene that is roughly at the same distance as the target. The visual angle between this fixation spot and the target should exceed the threshold visual angle. Then, the observer should slowly move his fixation in the direction of the target, until he first notices the target. The angle between the target and the fixation point at which the target is first noticed determines the target conspicuity at the given viewing distance. The threshold viewing angle and the viewing distance determine the extent of the conspicuity area. This extent is defined as the distance between the target and the fixation point in the frontoparallel plane through the target at which the target is first noticed. It has been shown that this extent is invariant for variations in viewing distance (for viewing distances that are not so small that the target covers a large area of the visual field or so large that the target can no longer be distinguished in the foveal visual field). Hence, the peripheral angle at which the target can be noticed will generally decrease with increasing viewing distance.

When using the above mentioned free viewing procedure in the field, target distances and visual angles can for instance be measured quickly using a pair of binoculars with a built in laser range finder and an electromagnetic compass (e.g. a Leica Vector 1500 DAES).

The conspicuity of the targets that were deployed in the DISSTAF field trials was low and the viewing distances were large (ranging up to 8 km). Therefore, the conspicuity was measured in the field using a pair of binoculars with a magnification of 10 times.

2.6. Laboratory measurements

The new procedure to measure conspicuity (introduced in Section 2.3) requires little effort and time, and can therefore be performed in situ easily, even in dynamic and complex environments. However, in some cases it may be more practical to measure conspicuity from photographs, slides, or video material. In that case, one can also help the observer to maintain fixation during measurements by introducing a fixation point into the scene. This can for instance be done by projecting a laser dot over the scene area.

The conspicuity of the DISSTAF targets was measured in the laboratory, on projections of slides taken from the field scenerios. The combination of viewing distance and the extent of the projected images was chosen such that the scenes were perceived with a magnification of 10 times, similar to the magnification of the binoculars that were used in the actual field trials.

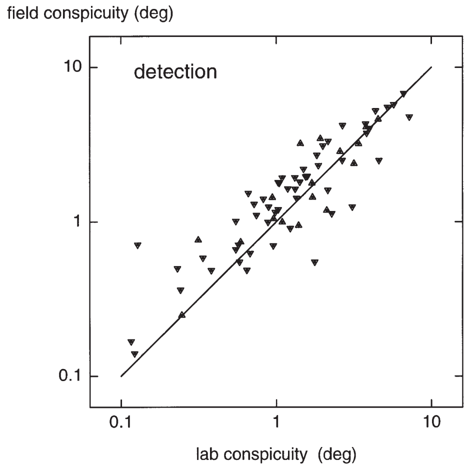


Figure 2 The characteristic spatial extent of the detection conspicuity area (in degrees), measured respectively in the field and in the lab on projections of photographic slides representing the field scenes. Data are shown for two observers (subject PB and FK, denoted respectively by upward and downward pointing triangles). The drawn line connects points of equal lab and field conspicuity, and closely fits the data points (r=0.89). The error is about 17% of the value (in both directions).

2.7. Relation between field and laboratory measurements

Field and laboratory measurements of the conspicuity of a target may differ because the photographic reproduction process influences the visual signature of a target. To assess the relation between both types of measurements we determined the conspicuity of the above mentioned military target vehicles both in the field and in the lab using slides taken from the field scenes.

Figure 2 shows that conspicuity measured in the lab on projections of photographic slides agrees with conspicuity measured in the field (r=0.89), for conspicuity values larger than about 1 degree. For targets that have conspicuity values smaller than about 1 degree in the field, the lab results consistently underestimate the target conspicuity. This is probably a result of the fact that the resolution of these targets is limited by the grain size of the photographic material, because these images were taken at large viewing distances. This implies that conspicuity is to a large degree robust for overall changes in the luminance of the scene, and for changes in the color balance. This finding is important, since it implies that the results obtained with photosimulation studies or with visual scene generators relate directly to the real life situations they represent. Hence, it is for instance in principle possible to interactively optimize the design and placement of traffic signs by measuring conspicuity from simulated imagery.

2.8. Relation between conspicuity and search time

We investigated the relation between target conspicuity and human visual search performance in the realistic and military relevant complex DISSTAF scenario's.

Figure 3 and Figure 4 show the relation between the mean search time and respectively the detection and identification conspicuity. These figures show that target conspicuity and mean search time are indeed strongly related.

A linear least squares fit to the logarithm of the data points yields the following relations between detection conspicuity C_d (in degrees) and mean search time ST:

$$\log ST = 0.81 - 0.67 \log C_d$$

and between identification conspicuity C_i (in degrees) and mean search time ST:

$$\log ST = 0.61 - 0.49 \log C_i \,.$$

The correlation between the logarithmic values of the identification conspicuity and mean search time is somewhat higher (r=0.89) than the correlation between the logarithmic values of detection conspicuity and mean search time (r=0.84). This is probably because a detail in the scene is more likely to attract fixation when it resembles a potential target. A detail that is highly visible (has a large detection conspicuity) need not attract attention when a single peripheral fixation is sufficient to notice that it does not resemble (cannot be) the target.

2.9. Discussion

The results presented in this section convincingly show that target conspicuity predicts visual search performance. This is an important result, since the conspicuity method provides an enormous reduction in time and number of observers required to obtain reliable search time estimates, compared to conventional methods. The agreement between field and lab measurements implies that conspicuity van be used to validate simulated imagery.

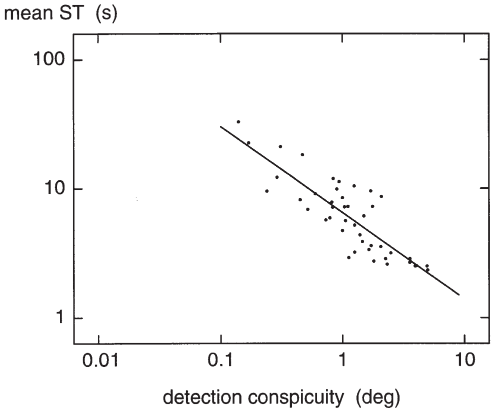


Figure 3 The relation between mean search time and detection conspicuity. The drawn line represents the result of a linear least squares fit to the logarithm of the data points (r=0.84). The error in the detection conspicuity estimates is about 17% of the value, and the error in the mean search time is about 14% of the value.

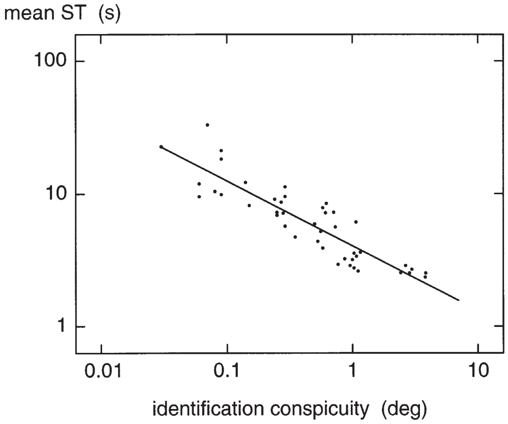


Figure 4 The relation between mean search time and identification conspicuity. The drawn line represents the result of a linear least squares fit to the logarithm of the data points (r=0.89). The error in the identification conspicuity estimates is about 23% of the value, and the error in the mean search time is about 14% of the value.

3. IDENTIFIABILITY

3.1. Introduction

A target will be easier to identify when its details are well resolved and clearly visible. It is therefore a priori likely that a measure that captures a target's visual articulateness should correlate with human visual identification performance. This suggests that the identifiablility of a target in a complex background can be determined by quantifying the visual articulateness of the target. Here we operationally define target identifiability as the amount of Gaussian blur that is required to reduce the target signature to its identification threshold. The rationale for the choice of a low-pass signature degradation filter is the fact that all spatial frequencies contribute to target identification[1]. The Gaussian blurring proces is easy to implement. A simple adjustment procedure can be applied to quickly determine the threshold extent of the Gaussian filter. This makes the target identifiability metric an attractive alternative for intricate identification experiments.

In the next sections we describe an experiment that was performed to validate the new target identifiability metric. The results of this experiment will show that identifiability indeed determines identification performance, and that it therefore provides an efficient alternative for measuring identification scores.

3.2. Validation experiment

Jacobs et al.[2] recently performed an experiment in which they compared identification performance on nominally identical sets of real and synthetic thermal imagery. Real imagery representing notional first generation and advanced scanning sensor systems was obtained. Parameters derived from the sensor data were used to generate synthetic imagery. Both image sets were then used in a target identification experiment with trained human observers. An example of these images is shown in Figure 5. For details about the image registration and generation procedures we refer to Jacobs et al.[2] In this study we use a subset of their imagery and the corresponding identification data to validate our new identifiability metric.

3.3. Experimental setup

Two observers participated in the experiment. The experiment was run on a PC controlled by Matlab software. The images were shown on a 22 inch PC-monitor, displaying 1280 x 1024 pixels in a screen area of 29.7 x 40.5 cm^2. The observer was seated at a regular distance from a PC-monitor (around 50 cm) and was free to adjust his/her distance to the screen. The experiments were performed in a dimly lit room.

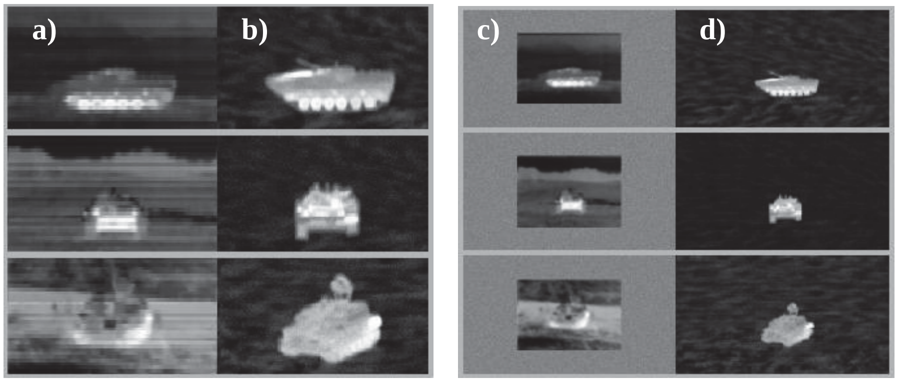


Figure 5 Example reference images (a and c) and synthetic (uncorrected) versions (b and d) for first generation FLIR (left: a and b) and advanced scanning FLIR (right: c and d). For details see Jacobs et al.[2]

2.1 Imagery

The reference image set represented 5 different military vehicles (2S3; BMP-1; T-62; T-72 ; ZSU) in natural terrain backgrounds, registered at 5 different viewing distances, and for two different aspects[2]. The 50 reference images were registered with an advanced scanning Agema LWIR camera. A second set of 50 reference images were generated by degrading the original 50 reference images to simulate a first generation sensor. The resulting 100 reference images were also simulated twice, using two different parameter settings of the simulation program. In their validation experiment Jacobs et al. [2] found that the set of parameters that was initially used to generate the synthetic images was not entirely complete, and additional parameters were needed. They therefore generated another set of corrected synthetic images. In the rest of this paper we will refer to the original and corrected synthetic image sets as the “synthetic” and “corrected synthetic” images. The entire image set therefore consisted of 100 nominally identical triplets of reference, synthetic and corrected synthetic thermal images.

2.2 Procedure

Prior to participating in the test, each observer was trained in target identification using the Recognition of Combat Vehicles (ROC-V) training software (see http://www.peostri.army.mil/PRODUCTS/ROCV/). ROC-V is a Windows-based thermal sight training program developed by CECOM NVESD. It helps soldiers learn to identify the thermal signatures of combat vehicles through the use of an interactive curriculum that teaches the unique patterns and shapes of vehicle 'hotspots,' and overall vehicle shapes.

At the start of each trial a randomly selected image from the test set was presented and the observer had to indicate the type of the vehicle that was displayed. After the observer had identified the target by pressing the appropriate response button on a keyboard, written feedback was given by displaying the name of the actual vehicle type on the screen. After a second button press the name disappeared and the observer started the actual identifiablilty measurement procedure. This process involved the adjustment of the width of a Gaussian blur kernel. By pressing the up and down arrow keys on the keyboard the observer could increase or decrease the amount of blur of the displayed image. The amount of blur could be altered in steps of 15% or 3.5% (for fine tuning), amounting to a factor 2 in 5 respectively 20 steps. In a typical measurement the observer initially increases the amount of blur until he can no longer identify the target. He then decreases the blur until he feels certain he can identify the target again. This process of blurring and deblurring is repeated until the subject feels confident that he has reached the threshold.

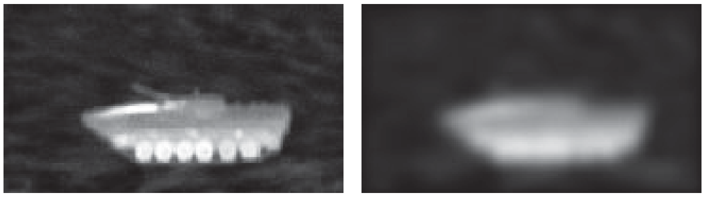
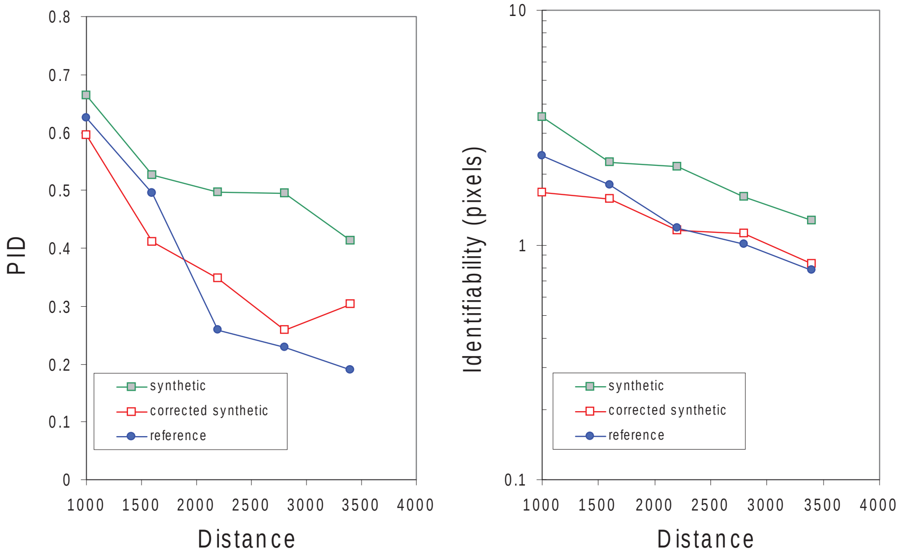


Figure 6 Example of an original image (left) and its blurred version (right). The task for the observer was to set the amount of blur to a value for which the vehicle could just be identified. This threshold value is referred to as the *identifiability* measure.

In case the observer could not identify the unblurred image the threshold was set to a value of 0.76 pixels, corresponding to 2 steps below a blur of 1 pixels, i.e. $\sigma = 2^{-2/5}$) (Note that, in principle, the image should be deblurred in this case to obtain a threshold image). By artificially fixing the value we assured that the average over all images corresponding to a given distance tended towards a low value, i.e. the average is influenced by the number of images that could not be identified. We obtained identifiability measures for each of the 300 images. The experiment was run in 8 sessions that lasted approximately 7 min. per session per observer (amounting to a total of 56 minutes per observer).

3.4. Results

The results of the validation experiment are shown in Figure 7. Figure 7a reproduces the results from Jacobs *et al*[2]. This figure show the proportion of correctly identified targets (PID) as a function of the viewing distance. Results for realistic ("reference") imagery is compared with results for original synthetic imagery ("synthetic"), and synthetic imagery for which the parameter setting has been corrected ("corrected synthetic") such that PID performance for these images matches PID performance for realistic images. The results clearly show that PID for the original synthetic imagery was much higher than the PID for realistic imagery. After correction PID performance for the synthetic images matches PID performance for realistic images.



(a) **(b)**

Figure 7 **(a)** The PID measured by Jacobs *et al.*[2] as a function of distance, for a first generation FLIR. The curves respectively correspond to realistic imagery ("reference", filled circles), synthetic (filled squares) and corrected synthetic (open squares) imagery. **(b)** Identifiability (blur threshold) in pixels for the same images, where the judgments of two observers have been averaged.

Figure 7b show identifiability expressed in pixels (averaged over the two observers) for the same images. The results show a similar pattern as the PID-performance results. In both cases the thresholds corresponding to the original synthetic data are higher than the reference data, but the corrected synthetic data matches the reference data.

Figure 8 shows the relationship between PID (measured by Jacobs *et al.*[2]) and identifiability (in pixels). PID-score is plotted against identifiability, averaged over the two observers. The data are fitted by a logarithmic line. The correlation between PID and identifiability is high, with an average correlation coeffient (*R*) of 0.91 (for the individual observers the correlation coefficients are respectively: 0.92 for observer MH, and 0.89 for observer PB)[1].

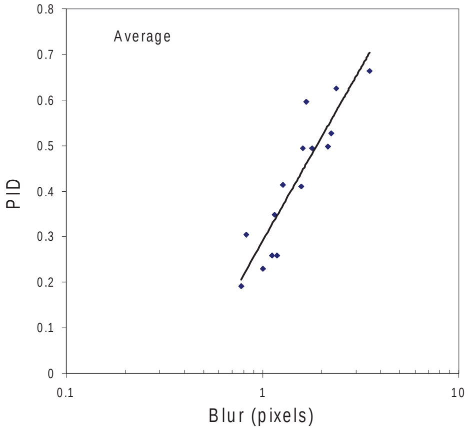


Figure 8 PID-score, as measured by Jacobs *et al.*[2], as a function of identifiability (in pixels) for the average over the two observers. Identifiability measures are averaged over all images registered at the same viewing distance (averaged over sensor type, vehicle type and orientation). The drawn line represents a logarithmic fit to the data.

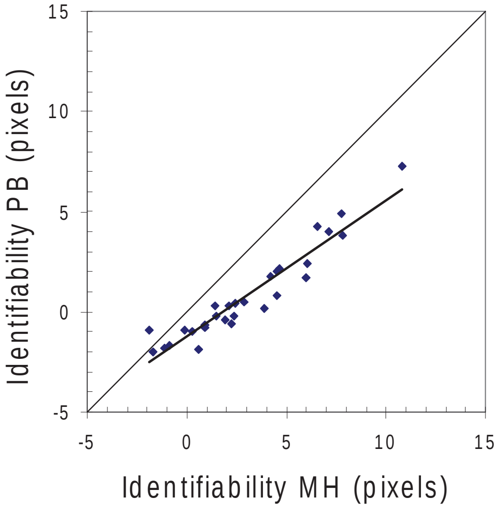


Figure 9 Correlation between the identifiability judgments of two observers (MH and PB). The drawn line corresponds to the linear fit. The dashed line corresponds to the condition in which both observers are in complete agreement.

---

[1] *R* represents the linear correlation correlation coefficient between PID and log(identifiability).

Figure 9 shows the correlation between the identifiabilty judgments of two observers. The data points correspond to averages over all images registered at the same viewing distance and with the same sensor type. The judgments of both observers are highly correlated, with a linear correlation coefficient of $R = 0.97$. The judgments of both observers show a linear relationship, in which the estimates of observer MH are about twice as high as the judgments of observer PB. The results show that observers agree well on the order of how well targets can be identified, but may differ in the absolute value of their judgments. It is therefore advisable to use the same observers when comparing synthetic with real imagery.

3.5. Discussion

Identifiability gives reliable estimates with a small number of observers and limited measuring time. We have shown that identifiability is directly related to PID-performance. Identifiability is therefore well suited for comparing synthetic and realistic imagery.

The general method is to degrade an image until perceived identifiability is at a threshold level. We chose to use blur as the degradation method. In principle other degradation method can be deployed just as well, such as the addition of (white) noise. However, blur is one of the major factors to determine identification distance with and without a sensor, while noise determines identification distance to a lesser extent. Therefore it makes sense to use blur to estimate ID-performance.

To obtain reliable estimates of identifiability it is best to use trained observers, since they are capable to judge for what amount of blur an image of a vehicle can still be discriminated from images of other vehicles.

There are several advantages to using identifiability over performing an experiment in which PID-scores are recorded. Apart from the fact that it can be obtained easy, fast and with a limited number of observers, identifiability makes it possible to estimate PID performance of single images. Synthetic imagery/models can be evaluated and adjusted (in a continuous loop) until identifiability matches that of realistic imagery.

4. CONCLUDING REMARKS

We presented visual conspicuity and identifiability as two efficient task-related measures that can be deployed to calibrate synthetic imagery that is intended to be used for human visual search and detection tasks.

Visual conspicuity predicts human visual search performance in realistic and military relevant complex scenario's. Conspicuity can easily and quickly be measured either in the field (in complex environments) or in the lab. This eliminates the need for costly and time consuming visual search experiments. The agreement between field and lab measurements implies that conspicuity can be used to validate synthetic imagery.

Identifiability is an efficient metric that can be used to gain insight into human identification performance without having to resort to elaborate and costly experiments. Target identifiability is operationally defined as the amount of Gaussian blur that is required to reduce the target signature to its identification threshold. Identifiability is directly related to PID-performance, and is therefore well suited for comparing synthetic and realistic imagery.

Summarising, synthetic imagery can be calibrated for human visual search and detection tasks by setting the conspicuity and identifiability of targets equal to those of their real world counterparts.

REFERENCES

1. Driggers, R.G., Vollmerhausen, R.H. and Krapels, K., Target identification performance as a function of low spatial frequency image content, *Optical Engineering*, 39*(9)*, pp. 2458-2462, 2000.
2. Jacobs, E., Edwards, T., Miller, B. and Hodgkin, V., Comparison of ID performance using real and synthetic imagery, In: R. Appleby, G.C. Holst & D.A. Wikner (Ed.), *Infrared and Passive Millimeter-wave Imaging Systems: Design, Analysis, Modeling, and Testing*, SPIE-4719, pp. 34-41, The International Society for Optical Engineering, Bellingham, WA, 2002.
3. Toet, A. and Bijl, P., Search experiment example, In: R.G. Driggers (Ed.), *Encyclopedia of optical engineering*, pp. 2566-2576, Marcel Dekker Inc., New York, USA, 2003.
4. Toet, A. and Bijl, P., Visual conspicuity, In: R.G. Driggers (Ed.), *Encyclopedia of optical engineering*, pp. 2929-2935, Marcel Dekker Inc., New York, USA, 2003.
5. Toet, A. and Kooi, F.L., Conspicuity: an efficient alternative for search time, In: A.G. Gale, I.D. Brown, C.M. Haslegrave & S.P. Taylor (Ed.), *Vision in Vehicles*, VII, pp. 451-462, Elsevier Science Ltd, Oxford, UK, 1999.
6. Toet, A., Kooi, F.L., Bijl, P. and Valeton, J.M., Visual conspicuity determines human target acquisition performance, *Optical Engineering*, 37*(7)*, pp. 1969-1975, 1998.
7. Wertheim, A.H. *A quantitative conspicuity index; theoretical foundation and experimental validation of a measurement procedure*, (Report C-20), TNO Human Factors, Soesterberg, The Netherlands, (1989).