Discriminating between camouflaged targets by their time of detection by a human based observer assessment method

G. K. Selj*[a], and M. Søderblom[a]

[a]Norwegian Defence Research Establishment, P.O. Box 25, NO-2027 Kjeller, Norway

ABSTRACT

Detection of a camouflaged object in natural sceneries requires the target to be distinguishable from its local background. The development of any new camouflage pattern therefore has to rely on a well-founded test methodology – which has to be correlated with the final purpose of the pattern – as well as an evaluation procedure, containing the optimal criteria for i) discriminating between the targets and then eventually ii) for a final rank of the targets.

In this study we present results from a recent camouflage assessment trial where human observers were used in a search by photo methodology to assess generic test camouflage patterns. We conducted a study to investigate possible improvements in camouflage patterns for battle dress uniforms. The aim was to do a comparative study of potential, and generic patterns intended for use in arid areas (sparsely vegetated, semi desert).

We developed a test methodology that was intended to be simple, reliable and realistic with respect to the operational benefit of camouflage. Therefore we chose to conduct a human based observer trial founded on imagery of realistic targets in natural backgrounds. Inspired by a recent and similar trial in the UK, we developed new and purpose-based software to be able to conduct the observer trial. Our preferred assessment methodology – the observer trial – was based on target recordings in 12 different, but operational relevant scenes, collected in a dry and sparsely vegetated area (Rhodes). The scenes were chosen with the intention to span as broadly as possible. The targets were human-shaped mannequins and were situated identically in each of the scenes to allow for a relative comparison of camouflage effectiveness in each scene. Test of significance, among the targets' performance, was carried out by non-parametric tests as the corresponding time of detection distributions in overall were found to be difficult to parameterize.

From the trial, containing 12 different scenes from sparsely vegetated areas we collected detection time's distributions for 6 generic targets through visual search by 148 observers. We found that the different targets performed differently, given by their corresponding time of detection distributions, within a single scene. Furthermore, we gained an overall ranking over all the 12 scenes by performing a weighted sum over all scenes, intended to keep as much of the vital information on the targets' signature effectiveness as possible. Our results show that it was possible to measure the targets performance relatively to another also when summing over all scenes.

We also compared our ranking based on our preferred criterion (detection time) with a secondary (probability of detection) to assess the sensitivity of a final ranking based upon the test set-up and evaluation criterion. We found our observer-based approach to be well suited regarding its ability to discriminate between similar targets and to assign numeric values to the observed differences in performance. We believe our approach will be well suited as a tool whenever different aspects of camouflage are to be evaluated and understood further.

.

Keywords: Detection, human observers, camouflage, assessment methodology, vision, search, rank criteria

1. INTRODUCTION

Camouflage and signature suppression is a very vital part of force protection and evaluation of such has recently gained increased attention [1-3]. The primary purpose with camouflage is to blend in with the natural environments. As different natural backgrounds are never alike, camouflage for military purposes has to be adapted reasonably well to a large set of background types. Typically the primary goal in such a context has been to reduce contrast between target and local background as well as reducing the target's conspicuity to an observer [1, 4]. This is a very complex task as one needs to achieve acceptable concealment effectiveness in all/most types of natural backgrounds of operational importance. At the same time, such developments of new camouflage have to be evaluated by objective, reliable methods of high operative

Target and Background Signatures, edited by Karin U. Stein, Ric H. M. A. Schleijpen, Proc. of SPIE Vol. 9653, 965305
© 2015 SPIE · CCC code: 0277-786X/15/$18 · doi: 10.1117/12.2195150

relevance. Of equal importance is the evaluation methodology being capable of ranking the targets under consideration quantitatively – preserving performance differences – and not just some rank by order.

Evaluation of visual concealment effectiveness is very difficult and complex as there is always a risk that the final recommendation (*e.g.* of a camouflage pattern) depends on the test method, the evaluation criteria, or any critical parameter that has to be chosen subjectively by the research team. Ideally, we would want available some broadly applicable, low-cost, unbiased signature evaluation methodology with high credibility. Hence, it is vital in all evaluations of a target's concealment effectiveness to use a proper measure, which is capable of capturing the information that is of relevance and interest to the primary purpose of the test. For evaluation of camouflaged targets that are intended to be used in a broad range of operative scenarios, it is important that the evaluation both spans all kinds of natural (or urban) sceneries as well as being closely related to operative use.

Several methods for camouflage properties assessment have been developed, all aiming to rank the targets under consideration as correct as possible. Photo-simulation by using human observers as an assessment method of concealment effectiveness of a set of targets has been used in various forms in the recent decades [1, 5-7]. Other methodologies have also been tested out, involving video surveillance [8], simulation of human vision [9-12], similarity measures of target-background by image analysis techniques [3,13], assessment by simulation of targets against different kinds of backgrounds [14], as well as image sequences taken by approaching sensors [15,16].

In this paper we present results from a recent camouflage evaluation study in the visual part of the electromagnetic spectrum. Through our observer based trial methodology we present quantitative results for 6 unique targets over 12 different, arid scenes. The results and methodology that is to be presented adds to the current knowledge on concealment effectiveness evaluation with the potential of narrowing the gap between experimental testing and operative use in the field further.

2. METHODS

Six different camouflage patterns, all with colours and pattern intended for use in dry and sparsely vegetated natural backgrounds were fabricated for testing of concealment effectiveness. All patterns were roller printed onto optically opaque cotton textile (225 g/ m^2) by roller print (HolTex, Germany). To allow for camouflage testing in the field of operation, the patterns were sewn into mannequin suits, all consisting of torso and head. The six patterns are from now on referred to as target 1 to target 6, as shown in Figure 1. A styrofoam mannequin was to be dressed up with the six suits, one at the time, and the targets were thereafter recorded in 12 various natural backgrounds (scenes) in Rhodes in August 2013. The scene images were captured with the intention of being suitable for camouflage assessment human observers in a search by photo (computer based) observer trial.

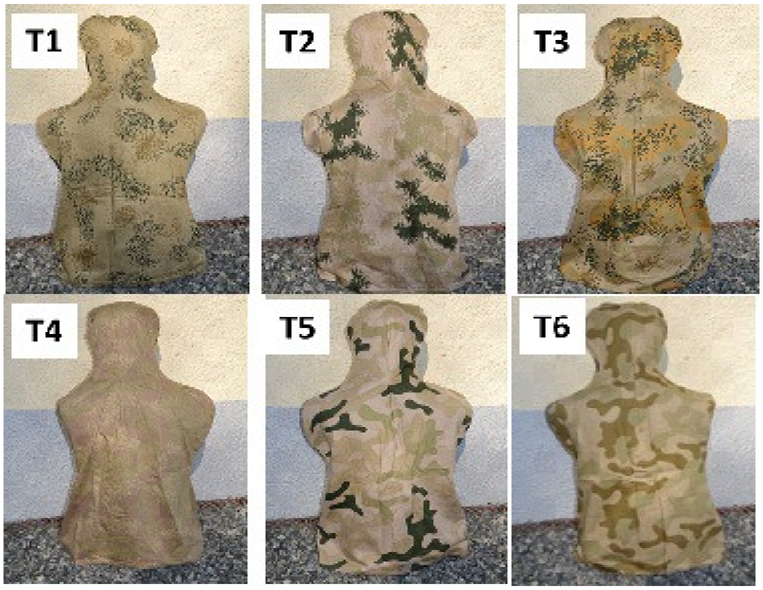


Figure1. Close-up images of the 6 different targets that were used in the observer based search by photo assessment trial.

Scene image capture

The scenes were chosen so that they contained different types of local (Mediterranean) backgrounds in the proximity of the target. We recorded the targets in as identical conditions as we were able to, considering, target orientation, position and area exposed. Identical positioning amongst the targets in a scene was ensured by mounting the mannequin onto a spear that was fixed to the ground. Furthermore, we were trying to record the scene images under as stable illumination conditions as possible, as the latter is very important to the targets' colour representation to the observer [17]. Our aim was to assure that the two targets' camouflage patterns were assessed solely based on their relative camouflage effectiveness. To achieve that, we carried out a near-continuous (within minutes) recording of the targets in each scene. This was done by a digital camera (Nikon D5200). Furthermore, only one target was recorded per image to avoid confusion about what is actually to be assessed by the human observer during the trial.

To ensure variation amongst the scenes, the targets were placed randomly in the image frame from one scene to the next in order to avoid observers' expectations on where to start to search. The targets were recorded so that they would not always appear centered in the image frame as it has been reported that observers tend to start searching from a central point on screen [6, 18]. The physical distance to the targets in the field was varied between 9 m and 70 m in the 12 different scenes. We recorded the scene images with the intention that the target actually was possible to be detected, whenever the observer's eye focus was at the target's spot in the image frame. Hence a detection of a target by a human observer was to reflect camouflage effectiveness and not to be based on observer's making guesses about too far objects. In 5 of the in total 12 scenes the target was recorded with both torso and head (ref Fig 1), whereas in the 7 of the remaining scenes only the target's torso (i.e. no head section) was used for recording. A collage of images of the 12 different scenes (with one target located in it) is shown in Fig. 2.

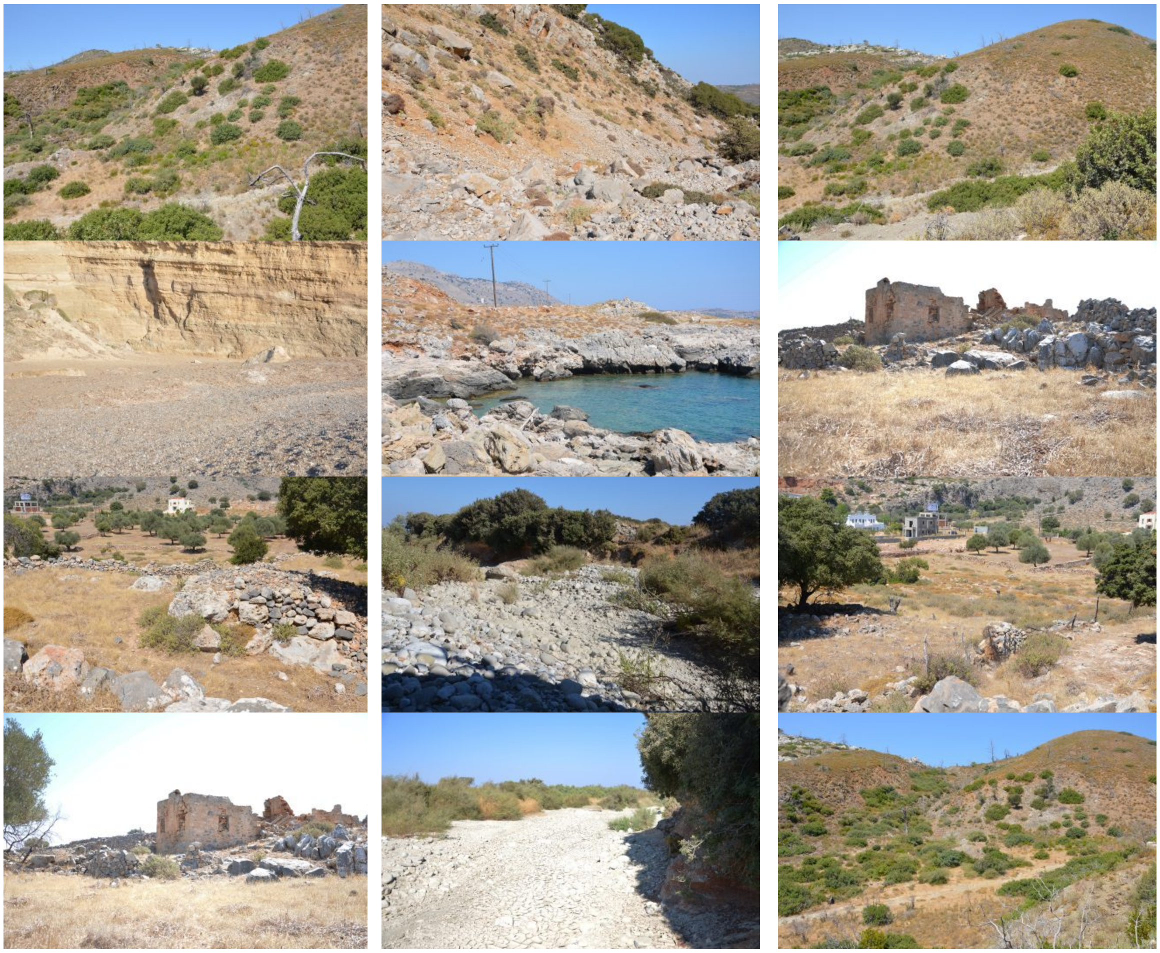


Figure 2. Overview of the 12 distinct scene images (Rhodes, Mediterranean) with different local background next to the target. The scenes were used in the observer trial.

Evaluation of camouflage effectiveness of the targets

Human observers were used to evaluate the visual camouflage effectiveness of the 6 targets. Based on the scene images, a photo-simulation observer trial set-up was designed [7], allowing for an evaluation of images from the field of operation under controlled and reproducible conditions indoor.

Preparing the trial

In order to prepare the scene images to be used in the trial, software developed for the purpose was used to mark the target digitally with a certain tolerance in each image as shown in figure 3. The images were also scaled and adopted to fit on a high definition 2560x1600 screen used in the observer trial. The scenes that finally were picked out for the observer trial were chosen with the intention of a reasonable detection time of targets and to assure that the targets were detectable whenever the eyes' focus was on the target spot in an image.

A total number of 148 observers (mainly conscripts) were used for the trial. This was to ensure the number of observers per target in each scene was sufficient to smoothen out inevitable differences in performance among the observers. The number of observers per target varied from 11(scene 8) to 19 (scene 1) due to a variable number of different targets being tested in each scene as the results to be presented in the paper was part of a larger study with more than 6 targets.. Furthermore, the group of observers was chosen to be as homogenous as possible with regard to their training and experience, specifically their training in observation techniques. They all gave information that could affect their ability to search for targets by filling out a demographic questionnaire according to recommendations in [19].

Much attention was paid by the research team to minimize errors and spread in performance due to individual human factors. Each observer was given an individual training session in an identical manner before the trial. After this brief, the observer was given a short training session to eliminate possible misunderstandings. During the training session the observers did not see the targets up close. This was avoided so that the observers did not familiarize with the specific targets in detail prior to the trial [20].

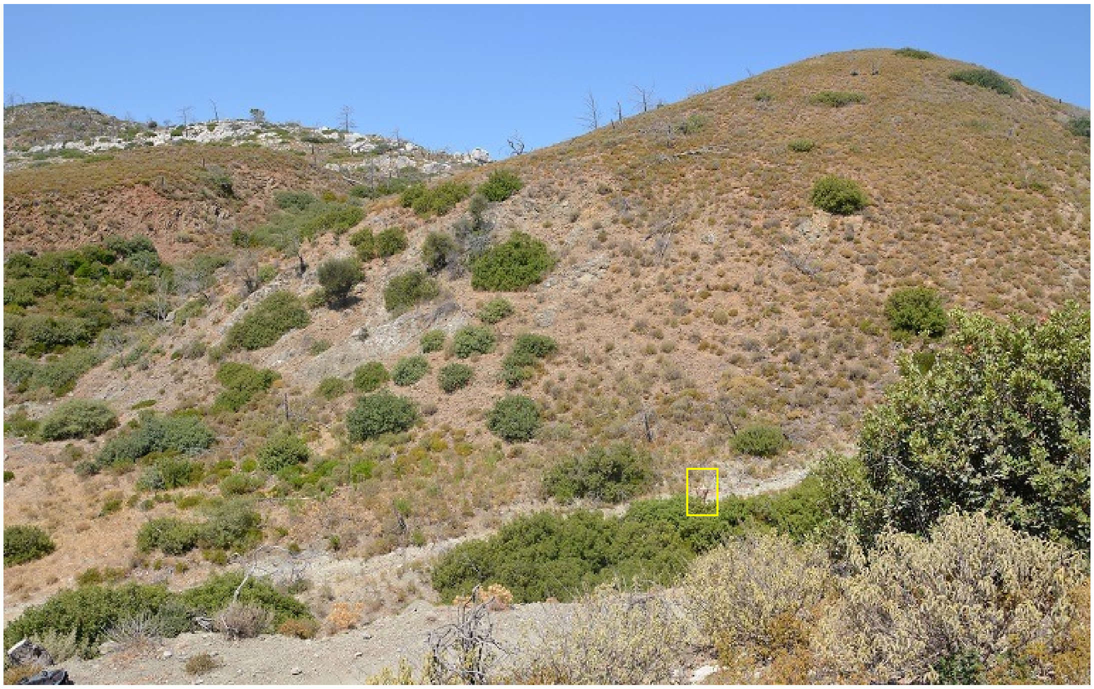


Figure 3. Example of a scene image with a target. Each target was recorded at the identical spot in the terrain. The allowed tolerance for the observer to designate the target is indicated by the yellow rectangle. For obvious reasons, the tolerance frame was not visible to the observers during the trial

Observer trial

The observer trial set up was implemented taking advantage of several recommendations from [19] and from DSTL implementation of the CASE method. The images were presented to the observers one by one on a high definition PC-screen. Each observer was positioned in a dimly lit room with a controlled, short distance to the screen to mimic the eyes natural field of view as shown in Figure 4.

Each observer was then presented to a randomized sequence of the 12 scene images, all with one or no target. The images were organized so that each observer was presented to each scene only once, but the observer was allowed to see the same target be used in another scene. The target was designated by pointing the pc-mouse and click. The observer was allowed a maximum search time of 60 seconds for each image, the observers were asked to focus on a specific point over the screen in the 5 second intervals between each image. This was to ensure equal search conditions in each scene. The detection time and location (coordinates of mouse click) were both automatically logged. Also possible non-detections due to the observer running out of time or miss detections were logged. This was done by purpose written software.

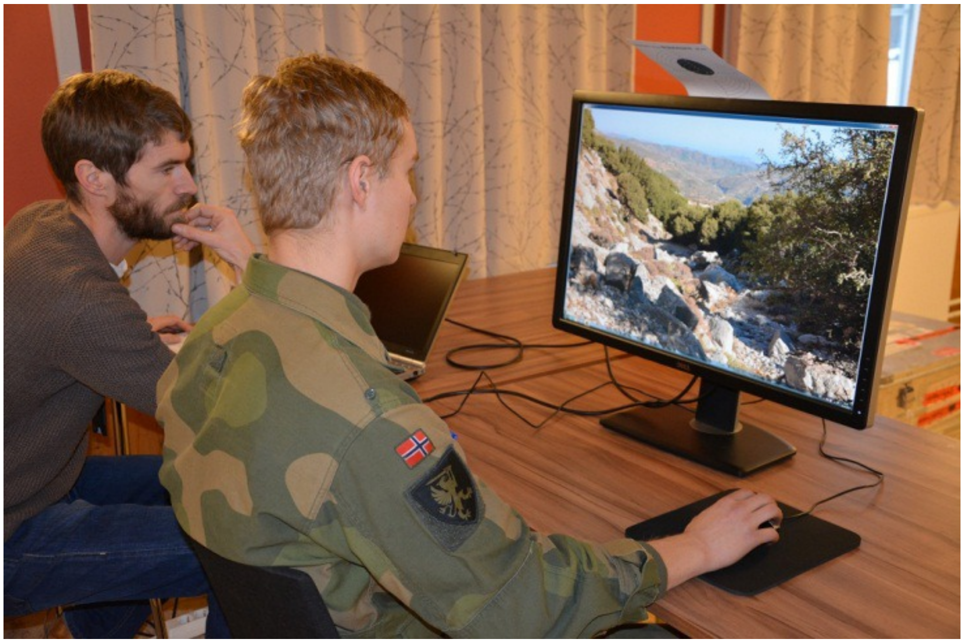


Figure 4. During the observer trial one and one observer were presented to one image (with or without target) from each of the 12 scenes. The supervisor did not assist the observer in any way during the trial. The figure is reprinted from Selj et al. [7] with permission.

Statistical analysis of data

The detection time data from the observer trials, using humans, were inspected further in order to look for significant differences between the two target's camouflage effectiveness in each of the 12 scenes. We first carried out a Jarque-Bera test for each target per scene, testing whether the corresponding distribution of detection times was normally distributed or not. Whenever at least one of the target's detection times, in a particular scene, failed to fulfill normality we carried out a Wilkoxon on rank test (Mann-Whitney U-test) or Kruskal Wallis as such a non-parametric test has shown to be more trustworthy than the parametric counterpart (such as ANOVA) in such cases [21,22]. Also, the Wilcoxon's rank test or Kruskal Wallis, being non-parametric, has the ability to account for non-detections (i.e. detection times larger than the search time limit set to 60 s) that appeared for some of the targets in some scenes during the observer trial.

In our data analysis of the experimental data we also calculated the median values of all detection times (as the median also is rank based, as is the statistical tests we used). The median value is simply the middle value when all detection times, for a single target in a specific scene, were ranked chronologically. By assigning the non-detections a numeric value of 60 seconds (corresponding to the search time limit) the median also captured non-detections (albeit with a conservative estimate) and not just the detections within the search time limit. Consequently, if the number of non-detections outnumbered the number of actual detections ($t < 60$ s) then the median was set to 60 s. In the rare occasions where the distribution of detection times contained an equal number of detections ($t < 60$ s) and non-detections, the median was assigned the average value of the sum of the largest detection time and 60 s (which was the assigned values of the non-detections when medians were calculated).

3. RESULTS

We present the results from 4 individual scenes in detail in this section as an illustration of the experimental data that were collected by the observer trial methodology. An overall ranking of the six targets over all 12 scenes is given in Table 1 towards the end of this section.

Single scene results

Figure 5 shows the distribution of detection times amongst the 6 different targets in scene 1. The green squares represent the median detection time for each target and is a measure of the characteristic detection time per target per scene. The numbers that are framed in the rectangle on top of the detection time distributions show the number of non-detections of the corresponding target in scene 1. The (star marked) hatches that connect some of the targets indicate the targets being significantly different ($p < 0.05$) with respect to their detection time distributions. An inspection of the results in Figure 3 shows that target 5 was performing significantly worse than target 1, 2, 3, and 4, but not when compared to target 6. This finding corresponds reasonably well with the visual impression of the individual performances in Figure 5.

Correspondingly, Figure 6, 7 and 8 show the detection times and non-detections for each of the 6 targets in scene 2, 5 and 11. We observe that there was large differences in relative performance (as evaluated by the median detection time) between the targets, particularly in scene 2 and 11 (Figure 6 and 8) as well as in (typical) median detection times amongst the scene were scene 5 had the lowest detection times (Figure 7). In scene 2 (Figure 6) we found significant differences in performance among the following combinations of targets (T1-T6, T2-T4, T3-T6, T4-T5, and T5-T6). Furthermore, we note that the median of target 6 was un-defined as half of the observations were non-detections, being some un-defined value above 60 s. As for scene 5 (Figure 7) no significant difference amongst the 6 targets was found ($p > 0.061$, which was T3-T4) despite the visual impression from Figure 7 that some targets (T3) seemed to perform poorer than other targets (T1). Finally, in scene 11 (Figure 8) significant differences were found amongst the following combination of targets (T1-T2, T2-T3, T2-T3, T2-T4, T2-T5, T2-T6, T3-T5, and T3-T6). As the majority of observations assigned with target 2 were non-detections, the corresponding median value was undefined, above 60 s.

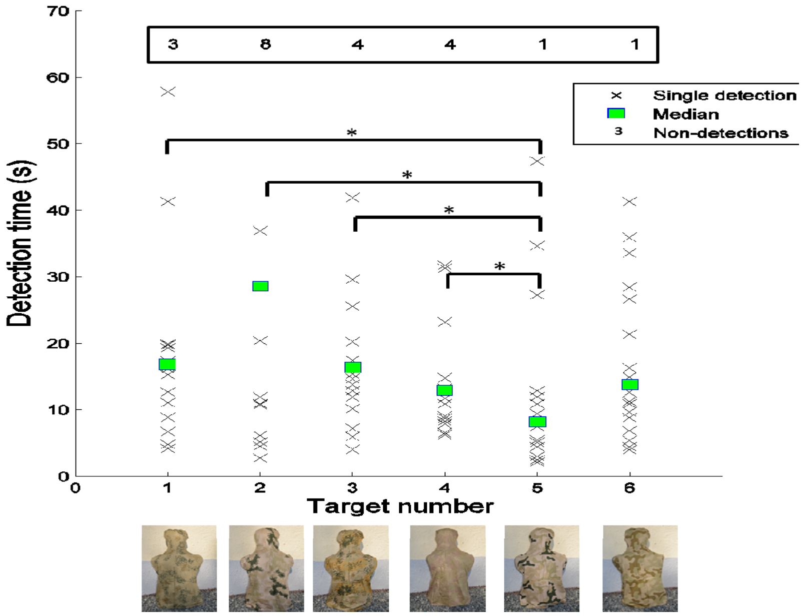


Figure 5. Detection times distribution for the 6 targets in scene 1. The numbers in the rectangle on top of each distribution shows the number of non-detections for the corresponding target and the star marked hatches connecting pairs of targets indicate significant findings ($p < 0.05$). The number of observers per target was 18. The targets were assessed with torso and head section in this scene.

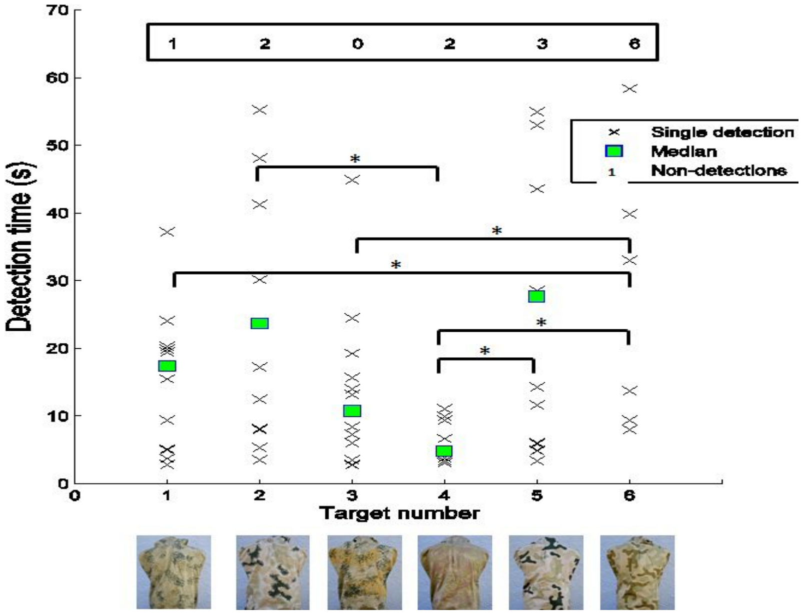


Figure 6. Detection times distribution for the 6 targets in scene 2. The numbers in the rectangle on top of each distribution shows the number of non-detections for the corresponding target and the star marked hatches connecting pairs of targets indicate significant findings ($p < 0.05$). Note that the median was un-defined for target 6 as the median exceeded the search time limit of 60 seconds. The number of observers per target was between 12 and 15 (target 5). The targets were assessed without the head section in this scene.

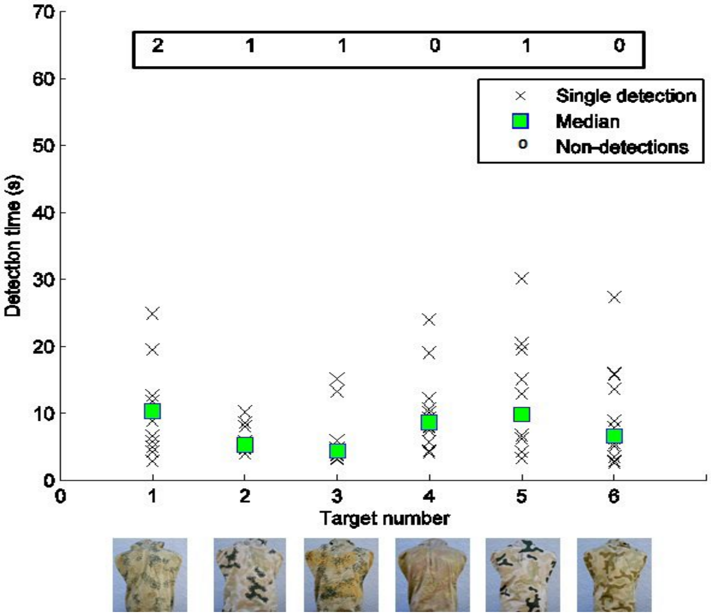


Figure 7. Detection times distribution for the 6 targets in scene 5. The numbers in the rectangle on top of each distribution shows the number of non-detections for the corresponding. The number of observers per target was between 12 and 13 (Target 6). The targets were assessed without the head section in this scene.

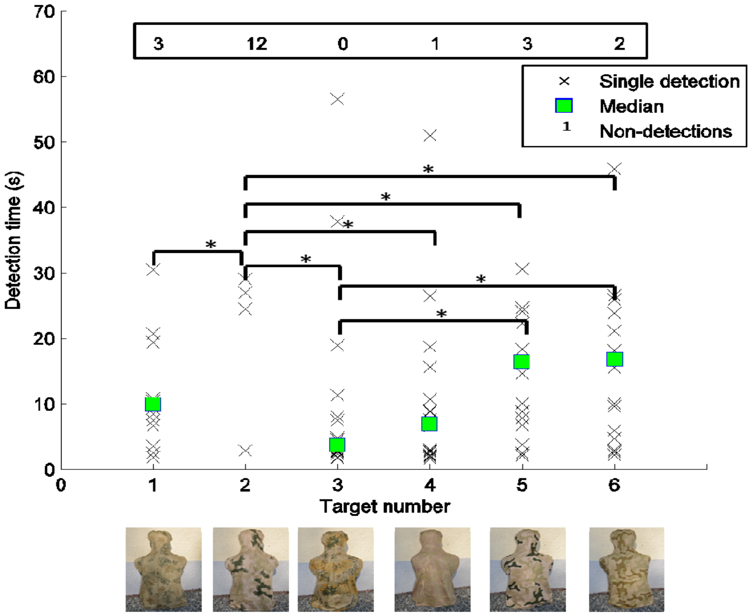


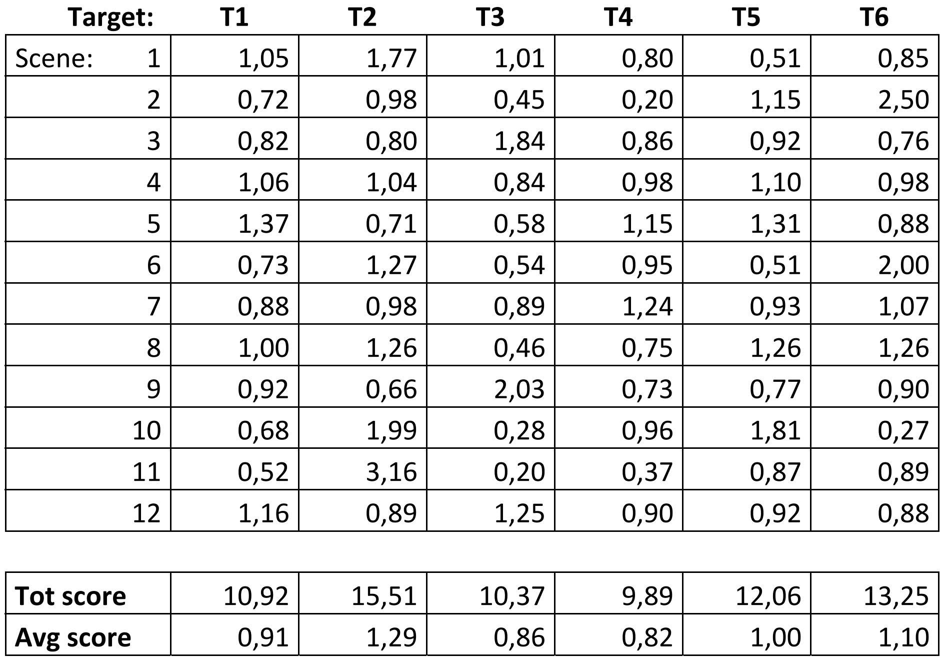
Figure 8. Detection times distribution for the 6 targets in scene 11. The numbers in the rectangle on top of each distribution shows the number of non-detections for the corresponding target and the star marked hatches connecting pairs of targets indicate significant findings ($p < 0.05$). Note that the median was un-defined for target 2 as the median exceeded the search time limit of 60 seconds. The number of observers per target was between 16 and 19 (target 4). The targets were assessed with torso and head section in this scene.

Overall performance of the targets

In Table 1 the normalized medians over all 12 scenes are shown together. In each scene the median of each of the 6 targets was divided with a common factor which was the average median detection time in that scene. Consequently, values larger than 1.0 in Table 1 show that the corresponding target performed above average in the given scene, and values below 1.0 show a performance below average. We note the large variation in (normalized) median values amongst the targets, showing that there were large individual differences between the targets regarding their corresponding detection times in most of the 12 scenes.

The normalization of the median detection time values, allow for a summation of characteristic (median) values over the 12 unique scenes. This sum, representing a target's overall performance in the entire observer trial is shown for each of the targets in the lowermost two rows in Table 1. The two rows mentioned show the total (performance) score of each target over all scenes, both in absolute values as well as in values relative to the number of scenes (the lowermost row in Table 1). We see that target T2 came out with the overall highest score, whereas T3 achieved the poorest overall score over the 12 scenes.

Table 1. Normalized medians of detection time for each target in each scene. The bottom row presents the average normalized medians for each target, a parameter indicating the total, relative performance of a target over all scenes.



Target:	T1	T2	T3	T4	T5	T6
Scene: 1	1,05	1,77	1,01	0,80	0,51	0,85
2	0,72	0,98	0,45	0,20	1,15	2,50
3	0,82	0,80	1,84	0,86	0,92	0,76
4	1,06	1,04	0,84	0,98	1,10	0,98
5	1,37	0,71	0,58	1,15	1,31	0,88
6	0,73	1,27	0,54	0,95	0,51	2,00
7	0,88	0,98	0,89	1,24	0,93	1,07
8	1,00	1,26	0,46	0,75	1,26	1,26
9	0,92	0,66	2,03	0,73	0,77	0,90
10	0,68	1,99	0,28	0,96	1,81	0,27
11	0,52	3,16	0,20	0,37	0,87	0,89
12	1,16	0,89	1,25	0,90	0,92	0,88
Tot score	10,92	15,51	10,37	9,89	12,06	13,25
Avg score	0,91	1,29	0,86	0,82	1,00	1,10

Combining measures of performance

Figure 9 shows the overall performance of the 6 targets when evaluated by two different evaluation criteria, detection time and probability of detection. For each of the targets the overall (over all 12 scenes), normalized median time as well as the average probability of being non-detected during the search were calculated. The corresponding pairs of coordinates are shown in Figure 9. The upper right corner in the figure corresponds to high performance in both measures whereas the lower left corner corresponds to poor performances as when evaluated by both time and probability of detection. We note the striking correspondence between the two measures of performance regarding the overall ranking of the 6 targets.

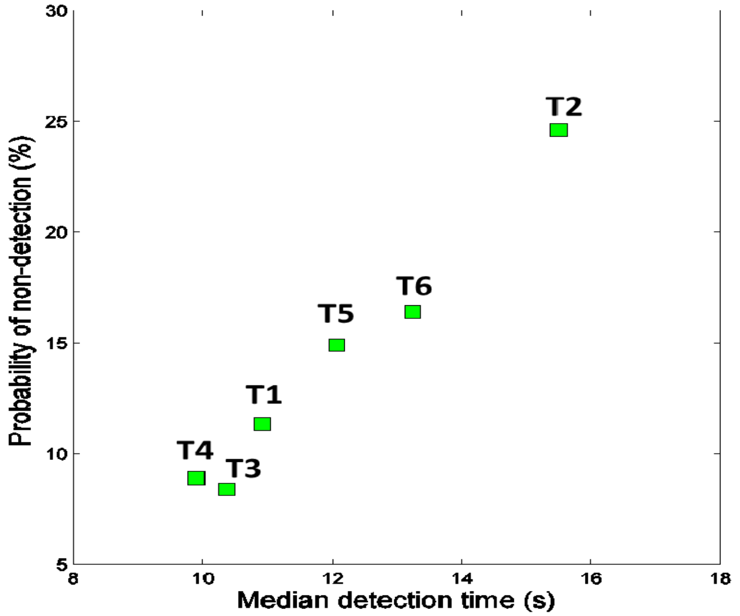


Figure 9. Median detection time of the 6 targets plotted against the corresponding probability of remaining non-detected during the 60 seconds of search time. The median times of each target are the sum of the normalized median times over all the 12 scenes. The probability of non-detection is the average probability of being non-detected (throughout the search time) over all 12 scenes.

4. DISCUSSION

In this study we have evaluated the visual concealment effectiveness of 6 distinct camouflage patterns, individually in each separate scene, and in addition achieved an overall, quantitative ranking of the targets in 12 unique scenes. The evaluation and ranking has been carried out by two different measures of performance, time of detection (our primary evaluation criterion) and probability of detection (secondary criterion), yielding quantitative performance data, of high operative relevance, on the performance of each of the targets.

General findings

From the four scenes that were presented in detail in the Results section in Figure 5-8, there are several aspects of interest. We note the (obvious) spread in the detection times' distributions amongst the 6 targets in all of these scenes. Similarly, we also found that the detection times' distribution for a single target (within a single scene) contained numeric values that varied substantially. Consequently, it is not obvious what kind of numeric time value that is the most representative measure of a target in a scene. From our results we found most of the distributions to be difficult to parameterize mathematically, and they were typically not normally distributed (by Jarque-Bera). In addition, there were in most cases one or more non-detections (given as integer numbers in the rectangles above each time distributions in Fig 5-8).

As the non-detections also contain important information about a target's concealment effectiveness, we wanted to include also the non-detections in any representative measure of performance of each target. The median time, which is capable of just that, is, in our opinion therefore the most reliable parameter to give a representative measure on individual targets being evaluated in an observer based search by photo trial [7] or similar. The median is also not so much affected by outliers. Consequently, the median is not overly shifted in those cases where single (high valued) detection times occur. From our data we see this in Figure 7 where both T2 and T3 had one single non-detection, each deviating much from the remaining, low valued (and clustered) detection times. The corresponding mean time detection times, which have been used as measure of concealment effectiveness in other studies [1], would, at least in our data have been more affected towards higher values, as the mean is in general more affected by outliers than the median [19].

It is interesting to compare T2 and T5 regarding their time of detection in the 4 scenes shown in Figure 5-8. In Fig 5 and 8 all targets were assessed with a head section on top of the torso (ref Fig. 1), whereas in Figure 6 and 7 all targets were assessed with no head section. We note that T2 and T5, which were near-identical camouflage patterns, performed almost identical in the scenes with no head section (scene 2, 4, 5, 7, 8, 10, and 12 in Table 1), but that T2 performed much better than T5 in scene 1, 6, and 11 where all targets were assessed with a head section. We address this difference to the specific pattern of the head section of T2 versus T5, where T2 potentially had a so called disruptive preference compared to T5 [4, 23, 24] yielding enhanced detection times under certain conditions [24]. Finally, we remark that as our targets were both assessed with and without a head section, different aspects of concealment effectiveness were assessed as body parts such as a head section is thought to be a salient feature of a target [25].

Measures of performance

As it is shown in the Figures 5-8 it seems that our primary chosen measure, time of detection, was able to distinguish much between the individual targets regarding their concealment effectiveness in a scene. Of particular interest is that our methodology allows for quantitative measures of the targets, providing numeric values of the performance differences in each scene. This can be done by extracting a characteristic measure (the median) from the corresponding detection time distribution of a target. An example is shown in Figure 6 where T4 achieved a median of roughly 5 s and T5 achieved a median detection time of about 28 s. Additionally, the time distributions associated with the two targets were significantly different ($p < 0.05$). Hence, there was a factor of more than 5 between these two targets and this particular factor is, as it describes differences in characteristic detection time, also thought to be directly linked to an expected operative performance when such two camouflage patterns are in use.

There is no gold standard regarding the best measure of concealment effectiveness in natural settings. It will in general depend on the primary purpose of the targets under consideration. Consequently, different measures, ranging from detection time [1,7], detectability [10], detection distance [15], probability [1], to subjective measure [2] and psychophysical measures [26] have been used in order to evaluate camouflage in earlier studies. One important feature of detection time is that it is a continuous variable, and consequently allows for a detailed (and un-biased) data harvest of signature effectiveness of targets. As we see it, this may have the potential that more of the vital target information is captured during a trial compared to similar evaluations where the primary measure is either binary, as for detection probability [1], or quantized in some other way, as in [15, 16] where detection distances were recorded in steps rather than continuous. However, we would like to underline that we do not state that detection time always is the best measure on a target's visual concealment effectiveness as proper rank criteria are inextricably intertwined with how an evaluation test is set up.

Scenes and methodology discussion

As the primary purpose in our study was to assess camouflage patterns intended for use in combat suits (or similar) the methodology was developed to capture a target's concealment effectiveness in a variety of natural backgrounds at close range (ca 10- 70 m). The choice of time as the primary rank criteria, including also non-detections, allowed for a location of targets at close range and still achieve significant differences in performance amongst the targets under consideration. We see this from our results in Figure 8 where all targets were only 17 m apart from the recording camera, and hence the observer during the trials. Still, there were significant (and partly surprisingly large) differences in performance, given by the targets' time of detection. We believe our methodology therefore opens for evaluation of targets also at close range, which is potentially of operative high importance. Through our methodology we were also capable to assess the camouflaged targets in a very broad span of natural backgrounds considered to be relevant for the final purpose and use of the camouflage patterns. In general there were few restrictions on the type of scene in this study, other than the targets having free line of sight to the recording camera as well as being detectable (except reasonable doubt) whenever the observer's eye focus was at the target spot in the image frame.

Compared to other camouflage assessment trials, such as many based on detectability or detection range, our methodology may potentially allow for a much broader selection of scenes (and hence closing the gap between test and operative use further). If targets must have free line of sight to the sensor or observer over a large span of distances, such as is in use in methodologies based on detection distances [15, 16], then the overall concealment effectiveness evaluation may be much hampered by a narrow selection of natural, local backgrounds, as open fields or similar is often required. This eventually limits the number of distinct background that targets are tested in and thereby also the operative relevance of the test.

Such close range concealment evaluations, as we report in this study, can generally not be easily carried out and at the same time generate high credibility results whenever measures such as detectability distance (which are much longer) or probability of hiding (which has the risk of being hampered by either a low number of non-detections or biased by the search time limit) are used. However, the approach suggested by Toet et al, where target distinctness was quantified through visual conspicuity [26] may also provide such information. It would therefore be very interesting to compare a rank by our observer methodology (which we believe is partially conspicuity driven) by a corresponding rank by any suitable conspicuity measure as a correlation in this respect may optimize an observer based photo simulation assessment methodology further.

The methodology that has been used in this study is based on real suits applied to mannequins in real operational settings, permitting a high degree of realism. Effects from silhouette, shadowing and partly solar illumination which all could affect the camouflage effectiveness, are therefore integrated in the study. Other methods based on digital simulation of the target, background or both [10-12] could well miss the influence of these effects. However, careful attention to preparation and conduction of the observer trial as described in this paper is essential to avoid methodic errors that could lead to an erroneous overall ranking. The spread in detection time for all targets in each scene highlight the importance of a sufficient number of observers. Furthermore, the selection of scenes is important. We find there should be several scenes for each relevant terrain category to be able to capture different immediate backgrounds, illuminations and distances.

Overall performance of targets

The overall ranking of the targets over all the 12 scenes was carried out by adding normalized performance values, derived in each scene individually. Such an approach preserves the relative performance differences amongst the targets in an overall summation. We chose to use the average of the target median detection times in each scene as the normalization factor. Consequently, all scene performance values (medians) were normalized by a factor that was characteristic for the detection times achieved in each scene. Hence values above 1 in Table 1 were found in targets performing above average in a scene and values below 1 were found in targets with concealment effectiveness below average.

In such a summation over all scenes the individual scenes were assigned a weight factor. In principle this weight factor may deviate from 1, whenever the scene is considered to be of particular high or low importance. However, as we did not have any rationale for assigning different weights to the scenes (as the latter is also a subjective act which will influence in the final ranking of the targets) all scenes were weighted equally. Rather than weighting the scenes by their assumed importance to the overall purpose of the test, we strived for a high number of distinct scenes, all of high operative importance, which in total covered the background types that were assumed to be, at least, the most important for arid camouflage.

In this respect, it is therefore very interesting to compare the ranking of the 6 targets by their time of detection and by the corresponding probability of detection. This was done in Figure 9 and we see the striking correlation between the evaluation and rank by the two measures. It is not obvious that the overall rank of the targets was to be unaffected (apart from the individual rank of the two poorest targets, T3 and T4) by switching between two measures of performance that are, by nature, different. Of importance in this manner is that the number of non-detections was typically low (< 3) for the majority of targets in most of the 12 scenes in our study. Consequently a rank based on the probability of detection will, as long as the number of non-detections is sufficiently large, will be much sensitive to the performance of each individual observers as adding one non-detection potentially will influence the result much.

On the other hand, the correlation between the two measures strengthens our overall rank of the targets (Table 1) as our secondary measure (probability) now adds support to our primary, and most trusted, measure (time). In our case the overall rank of the targets based on Figure 9 is obvious as the two performance measures showed a high degree of (visual) correlation. However, whenever there are discrepancies between the measures of performance (i.e. one target comes out best by the time of detection, whereas another target comes out best by the probability of detection, as in Toet et al. [1]), the overall ranking of the targets will depend on the weight associated with each measure of performance.

5. CONCLUSION

In this study we have shown that it is possible to achieve a quantitative rank with high operative relevance of 6 different camouflage patterns in a series of different, arid natural backgrounds. The method we used ensures results that are of high reliability and reproducibility and that allows for an evaluation of a broad variety of concealment effects, preferably in the visual range. Hence, our study is a step towards closing the gap between experiments on camouflage effectiveness and the final, operative use of the targets to be considered.

REFERENCES

[1] Toet, A. and Hogervorst, M. A., "Urban camouflage assessment through visual search and computational saliency," Opt Eng 52 (2013).

[2] Baumbach, J., "Color and pattern composition to blend objects into a natural environment," Association Internationale de la Couleur, (2008).

[3] Chang, C. C., Lee, Y. H. and Lin, C. J., "Visual assessment of camouflaged targets with different background similarities," Percept Motor Skills 114, 527-541 (2012).

[4] Endler, J. A., "Disruptive and cryptic coloration," Proc. R. Soc. B. 273, 2425-2426 (2006).

[5] Gretzmacher, F. M., Ruppert, G. S. and Nyberg, S., "Camouflage assessment considering human perception data ," Proc. SPIE 3375 (1998).

[6] Toet, A., Bijl, P. and Valeton, J. M., "Image dataset for testing search and detection models," Opt Eng 40(9), 1760-1767 (2001).

[7] Selj, G. K. and Heinrich, D., "Search by photo methodology for signature properties assessment by human observers,“ Proc. SPIE 9474, 947411 (2015).

[8] Boult, T. E., Micheals, R. J., Gao, X. and Eckmann, M., "Into the woods: Visual surveillance of noncooperative and camouflaged targets in complex outdoor settings," Proc. IEEE 89(10), 1382-1402 (2001).

[9] Tamura, H., Mori, S. and Yamawaki, T., "Textural features corresponding to visual perception," IEEE Trans Syst Man and Cybern, 8, 460-473 (1978).

[10] Hecker, R., "CHAMELEON-CAMOUFLAGE ASSESSMENT BY EVALUATION OF LOCAL ENERGY, SPATIAL-FREQUENCY AND ORIENTATION," Proc. SPIE 1687, 342-349 (1992).

[11] Kilian, J. C. and Hepfinger, L., "Computer based evaluation of camouflage," Proc. SPIE 1687, 359-369 (1992)

[12] Birkemark, C. M., "CAMEVA, a methodology for computerised evaluation of camouflage effectiveness and estimation of target detectability," Proc. SPIE 3699, 229-238 (1999).

[13] Nyberg, S. and Bohman, L., "Characterizing low signature targets in background using spatial and spectral features," Proc. SPIE 5152, 139-149 (2003).

[14] Houlbrook, A. W., Moorhead, I. R., Filbee, D., Stroud, C., Hutchings, G. and Kirk, A., "Scene simulation for camouflage assessment," Proc. SPIE 4029, 247-255 (2000).

[15] Schoene, R., Meidow, J. and Mauer, E.,"Feature evaluation for target/background discrimination in image sequences taken by approaching sensors ," Proc. SPIE 7697 (2010).

[16] Mauer, E. and Koffler, P., " A new method for observer-based evaluation of object detectability using image sequences by approaching sensors" Proc. SPIE 7113, 711317 (2008)

[17] Ohta, O. and Robertson, A.R., [Colorimetry], John Wiley & Sons Ltd, West Sussex, 92-93 (2005).

[18] Mannan, S. K., Ruddock, K. H. and Wooding, D. S., "The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images," Spatial Vision 10(3), 165-188 (1996).

[19] Peak, J et al., "Guidelines for camouflage assessment using observers," AG-SCI Rapport 095 (2006).

[20] O'Kane, B. L., Bonzo, D. and Hoffman, J. E., "Perception studies," Opt. Eng. 40, 1768-1775 (2001).

[21] Bickel, P. J. and Doksum, K. A., [Mathematical statistics - Basic ideas and selected topics], Holden-Day, Oakland, USA, 344-390 (1977).

[22] Sawilowski, S., S. "Misconceptions leading to choosing the t test over the Wilcoxon Mann-Whitney test for shift in location parameter," J Mod Appl Statist Method 4, 598-600 (2005).

[23] Selj, G. K., "Disruptive coloration tricks the human eye – a study of detection times of two near-similar targets in natural backgrounds,“ Proc. SPIE 9653, 965330 (submitted) (2015).

[24] Heinrich, D. H. and Selj, G. K., "The effect of contrast in camouflage patterns on detectability by human observers and CAMAELEON," Proc. SPIE 9476, 947604 (2015).

[25] Marr, D. and Hildreth, E., "Theory of edge detection," Proc. R. Soc. B. 207, 187-207 (1980).

[26] Toet, A., Bijl, P., Kooi, F. L. and Valeton, J. M., "Quantifying target distinctness through visual conspicuity," Proc. SPIE 3375, 152-163 (1998).