Camouflage performance of winter uniforms – Photosimulations in the visual spectrum

A. Mikkelsen*[a], G. K. Selj[a], N. H. Nielsen[a]
[a]Norwegian Defence Research Establishment, Instituttveien 20, 2007 Kjeller, Norway

ABSTRACT

Minimizing electro-optical signatures of soldiers against modern sensors is a challenging task, but a task with high importance and benefits for operative soldiers that need to stay undetected. Optimizing camouflage uniforms for winter conditions efficiently reduces soldier signatures in winter scenes, especially in the visual spectrum. Snow usually dominates winter scenes and is difficult to mimic because the spectral properties of snow change with several parameters such as grain size, structure, and wetness. Developing efficient winter camouflage thus requires knowledge and data on the spectral properties of snow. This paper presents spectral data on common snow types in Norway and evaluates the camouflage performance of several winter uniforms of different colors and patterns. We assessed and ranked the camouflage performance of the uniforms quantitatively in the visible spectrum using an observer-based photosimulation where many soldiers searched for targets in various Norwegian winter scenes. By collecting a large number of detection times, indicating how difficult it was for an observer to detect each camouflage in each of the unique winter scenes, it was possible to rank the camouflage targets quantitatively. The results show how each camouflage performed (given by time of detection or as a percentage) compared to all the other camouflages in the test for each scene. The photosimulation method is time-consuming, but it gives a realistic estimation of camouflage performance over the different scenes. We discuss the performance of the various winter camouflages with their pattern and similarity to snow (color coordinates).

Keywords: Snow, uniforms, soldiers, signature, winter, camouflage

1 INTRODUCTION

Signature management is vital for soldiers who seek to accomplish their mission undetected by modern sensors. By reducing their signatures, soldiers can undertake more advanced operations without risking themselves or their mission. The Arctic regions are particularly significant for military operations, especially in light of recent developments in Europe [1-3]. To achieve enhanced protection against detection in winter conditions, soldiers use winter camouflage uniforms and similar equipment to reduce their signatures. It is thus essential to understand which types of camouflage are most effective against specific sensor threats and in various conditions.

Camouflage is crucial in effective concealment in the animal kingdom and warfare. Its primary purpose is to blend in with the environment and attract as little attention as possible from any specific observer or sensor. Effective camouflage can be achieved through spectral reflectance matching and spatial fitting to the background characteristics, reducing the contrast between the camouflage and the local background [4-18]. Camouflage and nature must reflect light similarly over the entire wavelength region where electro-optical sensor threats operate [19-22]. Spatial fitting of camouflage to natural backgrounds can be achieved through frequency and orientation matching to resemble the spatial "pattern"-fingerprints of natural backgrounds, without making exact copies of a particular microhabitat [17, 23-25]. Utilizing these methods, FFI has developed different camouflage patterns that resemble the statistical "fingerprint patterns" or "noise" found in nature.

A thorough evaluation is necessary to ensure that new camouflage (with or without patterns) offers optimal protective properties. The three main approaches for camouflage evaluation are physical measurements (e.g., reflectance, emissivity, color, gloss, and polarization), computer-based image analysis (search models, target signature metrics, saliency, clutter metrics), and observer experiments (e.g., search and detection, range estimation and conspicuity). Reference [26] presents an overview of the various approaches and (dis-)advantages. No standard procedures exist to evaluate camouflage effectiveness today, as few studies have validated the different approaches against realistic military scenarios.

*alexander.mikkelsen@ffi.no;

Target and Background Signatures IX, edited by Karin Stein, Ric Schleijpen,
Proc. of SPIE Vol. 12736, 127360B · © 2023 SPIE
0277-786X · doi: 10.1117/12.2678537

In recent years, FFI has developed a photosimulation methodology for evaluating the visual camouflage effectiveness of camouflages [12, 27-30] based on test methods used by other nations [26, 31-35]. The methodology requires soldiers to visually search for camouflage targets (uniforms) in various scenes, with detection time and probability of detection as metrics to quantify the camouflage effectiveness. The methodology is reproducible and has been developed to capture the "camouflage effectiveness" of different camouflage types under conditions similar to those in which the camouflage is anticipated to be employed by military units. In this paper, using the observer-based photosimulation method, we have assessed the camouflage effectiveness of five different winter camouflage patterns in Norwegian winter backgrounds: two entirely white camouflages and three camouflages with grey patterns printed on white. Combined with physical measurements (reflectance and color), we evaluated their performance related to their similarity to snow (color) and design (pattern).

2 METHODOLOGY

2.1 Spectral measurements

We used an HR1024 spectrophotometer (Spectra Vista Corporation, USA) to measure the diffuse reflectance of all textile and snow samples (Figure 1) for incoming light of wavelengths from 345 nm to 2500 nm. The spectrometer was connected to a laptop to control the measurements via the program SVC HR1024 (v. 1.7). The distance between the samples and the spectrophotometer was around 1 meter, yielding a measurement area of about 5x5 cm^2. Most measurements were outdoors under stable illumination conditions (sunny and clear sky) with little or no wind. The spectrophotometer was slightly tilted to reduce the degree of specular reflectance from the samples. A Lambertian white reference plate (Spectralon calibrated diffuse reflectance standard, PerkinElmer) was used between each measurement series to calibrate the instrument. We measured the samples at least five times, with a fixed measurement time of 1 second. The spectral results presented in the paper (section 3.1) are an average of five or more spectral measurements for each sample.

The snow types studied in this paper were dry powder snow, wet snow, and crusty snow, all measured in Norway. Figure 1 shows the measurement setup for the spectral measurements.

Figure 1. Measurement set-up for spectral measurements. Left: HR1024 spectrophotometer and a laptop for controlling the measurements. Middle: the spectrophotometer was calibrated using a white plate reflectance standard. Right: spectral measurements of a white textile.

2.2 Calculation of color difference

Color coordinates were calculated to quantify the color difference between the snow and textiles. We used one of the most common color spaces, the CIELAB color space, recommended by the International Commission on Illumination (CIE). We calculated the color coordinates from our spectral data on snow and textiles following reference [36], using CIE standard illuminant D65 and the CIE 1964 10° Standard Observer function. We discretized the function at 5 nm wavelength intervals between 380–780 nm. The color coordinates in the CIELAB color space are L* (lightness), a* (redness-greenness), and b* (yellowness-blueness). The color difference between two colors with a set of CIELAB color coordinates can be calculated by ΔE on a scale from 0 to 100, where 100 indicates total distortion, and 0 indicates complete color similarity:

$$\Delta E = \sqrt{\left(L^* - L^*_{ref}\right)^2 + \left(a^* - a^*_{ref}\right)^2 + (b^* - b^*_{ref})^2}. \quad (1)$$

Colors with ΔE-value less than 1.0 look similar to the human eye, while colors with a larger ΔE-value might be perceptible by an experienced observer (ΔE-value 1–2), inexperienced observer (ΔE-value larger than 2) [37].

2.3 Description of the evaluation method

The methodology used to assess winter camouflage was based on photosimulations with observers [26, 38, 39]. We have recently used the method for similar tests, as reported in [12, 27-29, 40]. For interested readers, we refer to these works for extensive details of the test methodology. We present the evaluation methodology for assessing camouflage in the bullet point list below and in Figure 1.

We used **observer-based photosimulations** to assess camouflage effectiveness. The main steps of the test method are illustrated in Figure 1. For the test method:

- We only photographed one camouflaged target per image.
- In a scene, all camouflaged targets were consecutively placed in the same spot.
- In total, 70 observers participated in the trials. Each observer observed each scene only once.

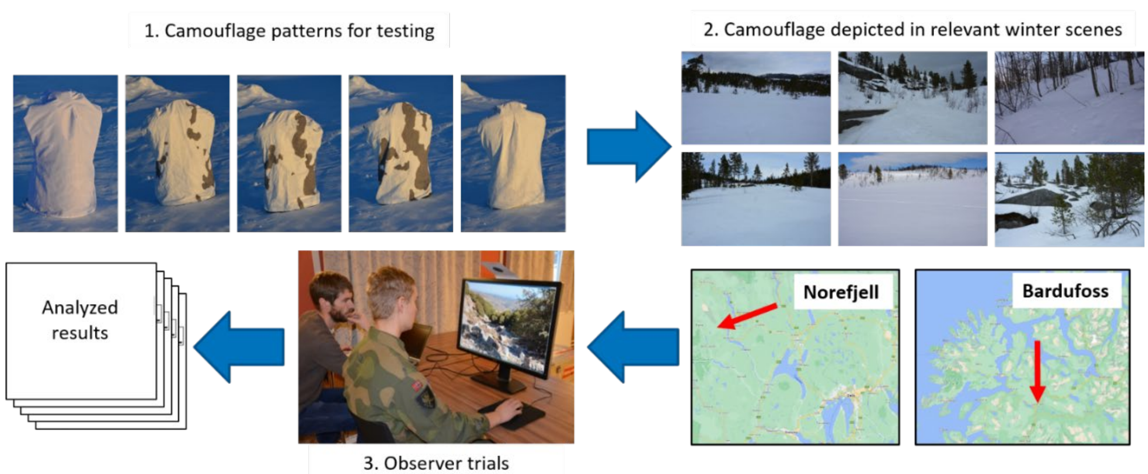

Figure 2. The main steps in the evaluation method: 1. select camouflage patterns to be tested, 2. depict camouflage in 11 distinct (and operatively relevant) Norwegian winter scenes, and 3. assess camouflage effectiveness by observer trials by 70 observers searching for camouflage targets in specified sets of images.

2.4 Winter camouflage patterns evaluated in this test

This study tested and compared five uniform suits with different camouflage patterns, all considered relevant to arctic winter warfare. Two of the uniforms were all white but with different white colors. The other three uniforms had grey patterns printed onto white, with grey-to-white proportions of 10, 20 and 30 %. These patterns were generated at FFI to resemble the statistical "fingerprint patterns" or "noise" found in nature and to disrupt the targets' silhouettes. Below we have numbered and listed all camouflage patterns evaluated in the test. The camouflage uniforms are also shown in Figure 2.

- W00: Textile with 100 % white area. The white color was dissimilar to the white color of camouflage W01–W04.
- W01: FFI-produced camouflage pattern. 90 % white, 10 % grey. Same colors as W02 and W03.
- W02: FFI-produced camouflage pattern. 80 % white, 20 % grey. Same colors as W01 and W03.
- W03: FFI-produced camouflage pattern. 70 % white, 30 % grey. Same colors as W01 and W02.
- W04: Textile with 100 % white area. The camouflage had the same white color as camouflage W01–W03.

Figure 3. Winter camouflage patterns for evaluation in the visual range in Norwegian winter backgrounds. We captured the images at sunset, explaining the dark snow backgrounds dominated by shadows.

We sew all camouflage patterns chosen for evaluation to mannequin suits of identical shape and size (Figure 3). This allowed us to carry out image recordings of the various camouflage patterns in the winter sceneries we found to be relevant. For simplicity, in this paper, we will refer to the different winter camouflage uniforms as targets. Also, to ensure that all test targets had the same reference textile underneath, each target was mounted on top of a Norwegian woodland camouflage (Figure 4).

Figure 4. Before the photography, we mounted all winter camouflage patterns on top of a woodland combat suit and a mannequin. In this image, camouflage W00 was photographed at sunset.

2.5 Scene selection

The scenes used to evaluate the camouflaged targets in this study were all unique, capturing different aspects of Norwegian winter backgrounds. We placed each target in fixed positions within a winter scene and photographed them one after the other, ensuring that the mannequin's position in the snow remained fixed. We put effort into ensuring that tracks from the person placing out the targets were not visible in the image. We also ensured that the conditions, including target position, orientation, visual exposure, and illumination, were as equal as possible for each image capture. This was important to ensure that the observer's perception of the target and background contrast was accurate and to eliminate any potential image capture artifacts that could have impacted the final results.

Eleven scenes were chosen to cover a broad range of winter and arctic backgrounds. The geographical regions used for the image capture were all in Norway (north: Bardufoss, central part: Norefjell). Figure 5 depicts a map of the areas used for the different scenes.

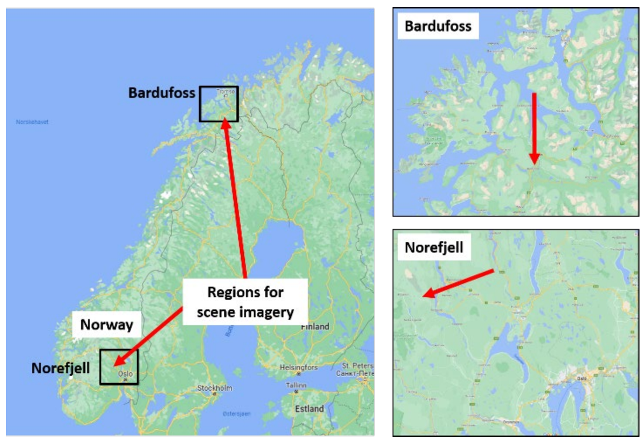

Figure 5. Areas used for scene image capture of snow camouflage in Norwegian winter backgrounds.

To decrease the chance of conditions affecting the camouflage effectiveness in the 11 distinct scenes, we photographed all the targets within minutes in each scene. We photographed the targets using a Nikon D5200 digital camera with an 18–105 mm lens, using a fixed zoom for each scene. The test methodology used in this study was based on evaluation by human detection of one isolated target at a time for every scene [27]. Therefore, we only photographed one target per image; consequently, each scene contained a set of images (i.e., one for each target).

We put effort into selecting optimal scenes so that the nature of the scene would not reveal or attract the observers' attention to the target position. This is illustrated in Figure 6, and for clarity, we underline that even if several target positions were considered for each scene, only one position was used to place targets for a given scene. Targets were also located in different positions in the image frame from one scene to the next, which means the observer should not expect the target to appear at any specific position from one scene to the next. As observers tend to start the search process in the center region of an image [41], we did not place the targets close to this area during the image-capturing process.

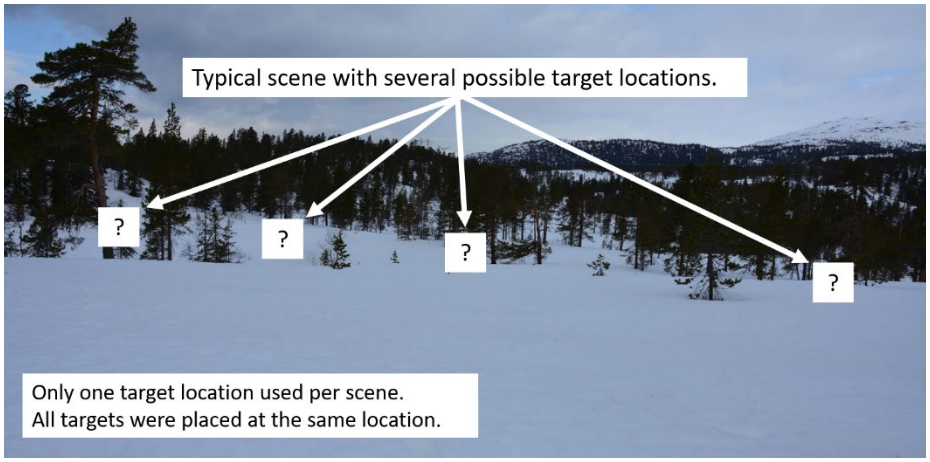

Figure 6. Illustration of the scene selection process. Terrain backgrounds with several possible locations of the targets were selected as scenes. This reduced the likelihood that observers could guess (from the scene type) where a target was likely to be positioned. When placing the targets in the selected position, we were careful not to make visible walking paths in the recorded images. The targets were only placed in one position (identical for all targets) per scene.

2.6 Preparation of human observers

Before the observer trial, we gave each observer an identical and consistent brief explaining the test procedure steps. To prepare the observers for the trial itself, each observer carried out a test run consisting of two practice scenes with a camouflaged target similar to the scenes to be shown later in the observer trial. The observers were allowed to ask questions during the briefing, sorting out potential misunderstandings before the main trial. However, the observers were not allowed to ask whether they had found the proper target during the trial itself, ensuring that the targets were assessed purely on the observers' performance. However, observers were free to choose their search strategy. We used an optimized distance from the observer to the widescreen used in the photosimulation and ensured that it was identical for each observer. The distance to the screen was approximately 40 cm for the screen to fill the observer's field of view. We also adjusted the screen's height so that the observers' eyes were leveling the center of the screen. The observer trial started after the adjustments, and the observers confirmed that they understood the procedures. The observers then carried out the trial individually.

2.7 Implementation of the observer trials

In the observer trial, a purpose-made software presented each observer a randomized image sequence of the 11 unique scenes. In a dimly lit room, the observers saw the scene images one at a time using a high-definition (HD, 2560 x 1600 pixels) pc screen. Each image in the sequence was a scene with one camouflaged (human-shaped) target. Individually, the observers searched for a target with a known shape but an unknown camouflage pattern and position in the image. Target detection was indicated when the observer mouse-clicked on the target as soon as the observer was convinced it was a real target. We established a minor tolerance surrounding each target for each scene to filter relevant detection data from anomalies or miss-detections. The dimension of the added tolerance around the target was typically lower than the dimension of the target itself, meaning that detection markings ("the observer's indication by a mouse click") just outside the target outline were assumed to be a correct detection. We chose this strategy intentionally to prevent false non-detections or the observer from spending too much time placing the mouse marker at the target center point. For each search, the software stored the corresponding detection time for further analysis. The software allowed the observers to be exposed to the same targets in more than one scene throughout the 11 scenes in the trial, and the observers were never asked to search for more than one target per scene.

In each scene, the observers had a maximum search time of 50 seconds to register a target detection. The whole search time gave the observer reasonable time to detect the target, but at the same time, not so much time that searches for an exceptionally well-hidden target would hamper the quality of the search process, preventing the observer from being de-motivated by tedious searches. The target (in the given scene) was stored as "non-detection" whenever the time limit of 50 seconds was exceeded. All detection times and non-detections were kept for further analysis.

2.8 Comparing camouflage effectiveness of targets in a single scene

In this trial, we chose detection time as the primary ranking criterion and probability of detection as a possible secondary criterion (as suggested in previous and similar studies [27-29]). A short justification of the rank procedure, in which we considered the median detection time the most important, is given below.

In general, detection times for a hidden or camouflaged target will be a distribution: a spread of numbers from low to high. Often the spread of detection times is non-symmetric around its mode (i.e., the most frequent detection time) and somewhat shifted towards larger values (i.e., a right-shifted distribution [42-44]). Because detection times can assign only positive values, limited "space" for deviations exists below the mode. On the contrary, above the mode, there is room for significant variations (at least until the fixed search time limit of 50 seconds). A mean value calculated based on a test target's detection times can thus give a false representation of the target's overall camouflage effectiveness, simply because single outliers potentially will shift the mean detection time towards a high value, but rarely towards a much lower value. The median, however, will not right-shift considerably by a few outliers with long detection times [45]. In this study, we therefore rely on our camouflage ranking of the targets on their corresponding median detection time, a ranking criterion that reflects the camouflage effectiveness of a target with higher precision than the mean. In addition, the median also accounts for the non-detections by simply counting them as high, undefined values.

2.9 Including non-detections in the evaluation

We also included the non-detections recorded during the observer trial in ranking the targets. By definition, the non-detections were not assigned any detection time value during the trial. Still, they were treated as some undefined value above the search time limit, which we defined as 50 seconds. Note that if the non-detections outnumbered the detection times, the median of that particular target was registered as a "non-detection."

If the distinct number of observers for a target was even, the median detection time was calculated by averaging the two detection times in the middle. For example, if there were four observers and the recorded detection times were: 1, 2, 3, and 4 seconds, the median detection time would be (2+3)/2 = 2.5 seconds. In the rare event that the median detection time of a target was the average of a well-defined detection time and the "first" non-detection, we assigned this particular non-detection the value of 50 seconds (the fixed search time), enabling a well-defined median (as the average of those two detections), albeit with a conservative estimate. In any case, the median as a ranking criterion preserved the valuable information, represented by non-detections, of the test targets.

2.10 Calculation of overall ranking

To find the overall rank for each target (overall 11 scenes), the approach was:

- A normalization of the median detection time for each target in each scene, representing the performance of the target relative to the other targets in a particular scene. A numeric value above 1.0 for a target reflected camouflage effectiveness above average for that scene, whereas a value below 1.0 reflected the opposite. Such an approach, similar to studies carried out earlier [12, 27-29, 40], also accounts for the targets' relative differences (not just their order).
- An assigned weight (higher, equal to, or less than one) for each scene. In this study, each of the scenes was given equal weight.

3 RESULTS

3.1 Spectral data and color likeness

To study the importance of color similarity to snow, we measured the spectral properties of different snow types in Norway. The reflectance curves of the snow types at visible wavelengths are presented in Figure 7: dry fresh snow (red curve), dry, coarse snow (blue curve), wet coarse snow (green curve), crusty snow (violet curve), and wet crusty snow (grey curve). The reflectance curves are calculated means of snow measurements performed at various places and at different times in Norway and thus yield a reasonable representation of snow properties.

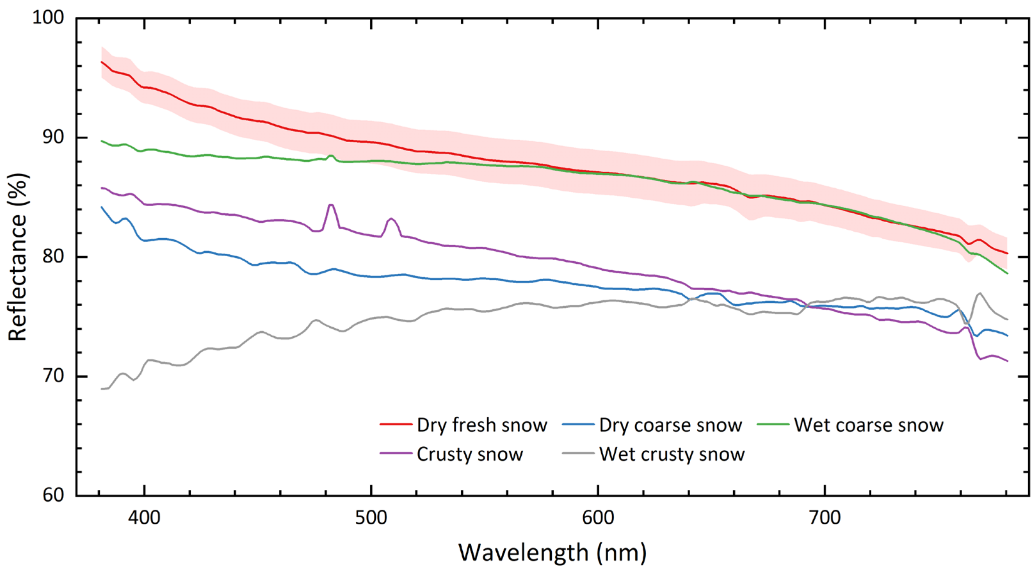

Figure 7. The reflectance of different snow types plotted against visible wavelengths: dry fresh snow (red curve), dry, coarse snow (blue curve), wet coarse snow (green curve), crusty snow (violet curve), and wet crusty snow (grey curve). The colored red area around the red curve indicates the standard error (SE) of the measurements on dry fresh snow. The other snow types had a similar magnitude of standard error, but for clearness, their errors are not shown here.

We found that dry fresh snow generally reflects more visible light than dry, coarse snow and that increasing moisture content lowers the snow reflectance. Depending on the snow type and wavelength, the snow reflectance varied between 70 and 96 %, where wet crusty snow reflected the least. To compare the color of the snow with the white color of the camouflage targets, we used these measurements to calculate the color coordinates of the snow (see section 2.2 for the method).

Table 1 presents the color difference between the white color of camouflage targets W00–W04. The white color of target W00 was significantly more dissimilar to the snow than that of targets W01–04, with ΔE values of 12.4 ± 2.0 and 5.5 ± 2.0, respectively. The white color difference between snow and targets is thus perceptible at a glance with human eyes; however, the white color of target W00 stood out the most.

Table 1. Color difference (ΔE) and estimated standard error (SE) between snow and the white color of camouflage targets W00–W04.

Target	ΔE	SE
W00	12.4	2.0
W01–W04	5.5	2.0

3.2 Camouflage effectiveness in selected winter scenes

In this section, we will go through the results of the observer trial test in detail. The test results contain detection times distributions for all test targets in 11 different winter scenes, and it will be too overwhelming to present results from each scene in detail. Thus, we present detailed results from only two of the eleven scenes in this section. The overall result, a sum of the performance in all eleven scenes, is given in the next section (Figure 10).

The result figures (Figure 8 and Figure 9) representing the detection results from the observer test contain much information that requires a proper explanation. To guide the reader, we briefly explain how to interpret the information in the figures. At the left of the results figures, we have named the different targets (W00–04). The blue dots and red circles in the figures located horizontally to the right of each target represent the targets' corresponding camouflage effectiveness. The blue dots each represent a detection time from one of the observers in the observer trial. If the detection time from one observer was more than 50 seconds, there is no blue dot for that detection time. Each target had 7–8 observers, giving about eight unique detection times (blue dots) per target. The blue asterisk symbol indicates the median detection time, characteristic of each target's camouflage effectiveness. A long detection time generally represents high camouflage effectiveness, whereas short times yield poorer target performance. Hence, by looking at and comparing the relative values of the blue asterisk symbols in a figure, it is possible to get a first impression of which target performed better (and poorer) in a particular scene.

The red circles in Figure 8 and Figure 9 show the probability for each specific target to remain undetected throughout the entire search time of 50 seconds (the number of non-detections divided by the number of observers per target in all scenes). Generally, the further the red circle is located to the right, the better the corresponding target performed.

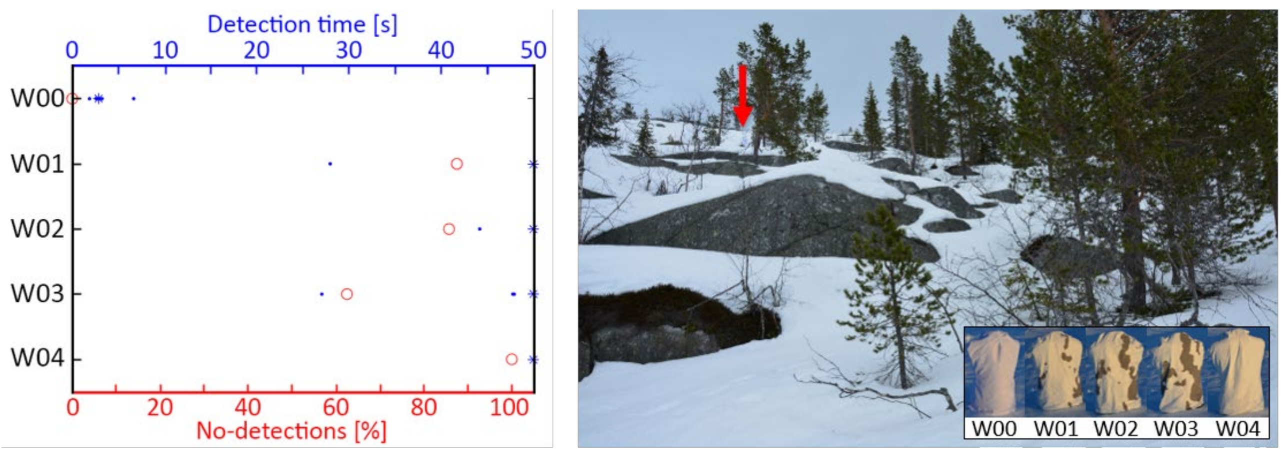

Figure 8. Distributions of detection times (blue dots) for the camouflaged targets in scene 9. The blue asterisk symbol indicates the median detection time for each target, while the red circle indicates the probability of non-detection within 50 seconds. Right: An image of the scene with a target (marked by the red arrow). The inset images show target W00–04.

In scene 9 (Figure 8), we found that target W00 performed worst with a median detection time of around 4 seconds, and all observers could spot the target within 50 seconds (i.e., no-detections at 0 %). The other targets were much more challenging for the observers to detect, with a high percentage of no-detections and long detection times. From the results in scene 9, it was not possible to find notable differences in performance between targets W01–04, i.e., the targets with the same white color but different patterns.

Compared to scene 9, the targets in scene 11 (Figure 9) were generally easier for the observers to detect, with shorter detection times and a lower percentage of no-detections. In scene 11, targets with patterns (W01–03) performed significantly better than targets without patterns, i.e., entirely white targets (W00 and W04). All observers managed to detect each of the entirely white targets, and the detection time was short, with mean detection times of less than five seconds. Target W03 performed best in scene 11 with 100 % no-detections, i.e., none of the observers managed to detect the target. Target W01 and W02 had a lower grey-to-white ratio than target W03 and were more straightforward for the observers to see in scene 11. These targets had median detection times of around 15 seconds and a low percentage of no-detections (6 and 11 %).

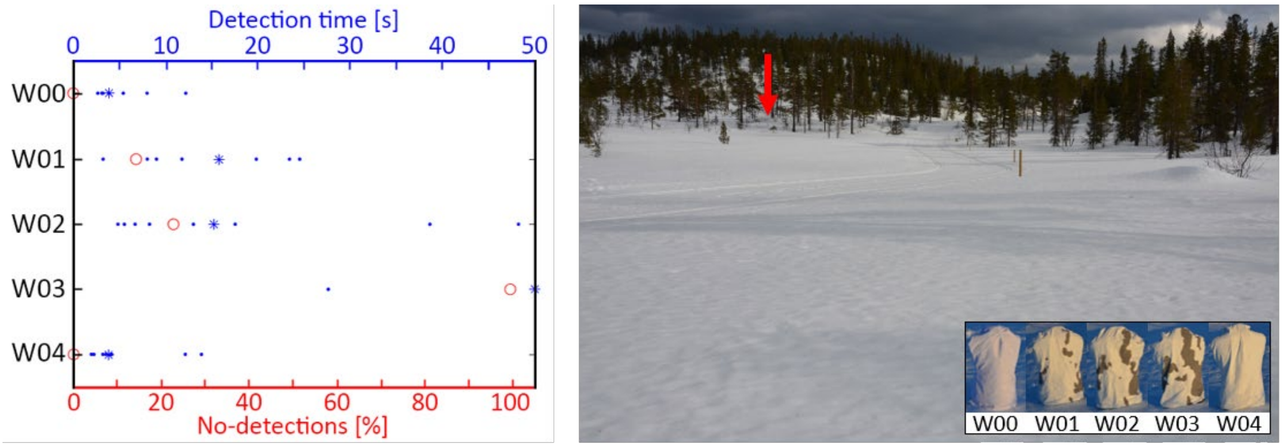

Figure 9. Distributions of detection times (blue dots) for the camouflaged targets in scene 11. The blue asterisk symbol indicates the median detection time for each target, while the red circle indicates the probability of non-detection within 50 seconds. Right: An image of the scene with a target (marked by the red arrow). The inset images show target W00–04.

3.3 Overall camouflage effectiveness

Figure 10 shows the overall result from the camouflage tests and is the most important result of the winter camouflage test. Each target is pictured in the right part of the figure and has result columns in the left result figure. Generally, the results shown in Figure 10 should be interpreted the following way: targets assigned with long columns performed better than targets assigned with shorter columns.

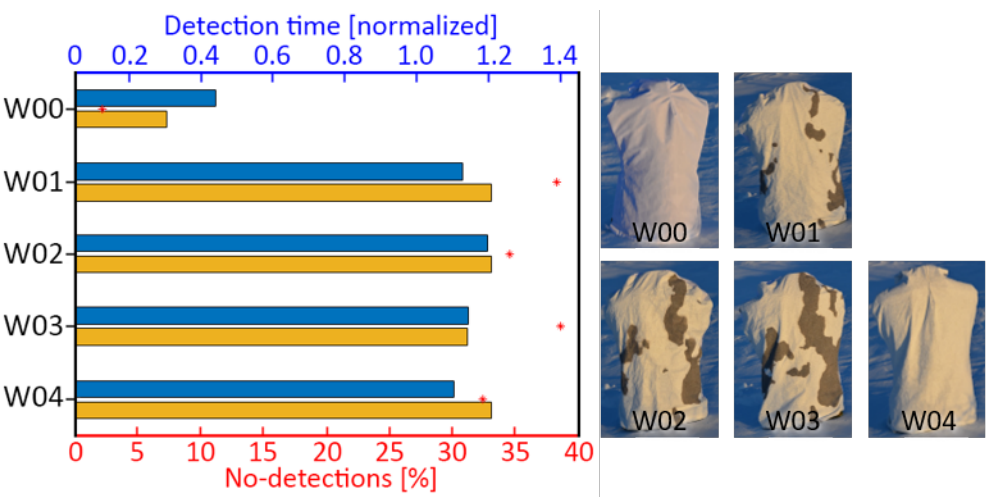

Figure 10. Effectiveness of the different camouflage uniforms in all scenes given by normalized detection time. Generally, longer columns relate to higher camouflage effectiveness. A column of double length compared to another (of the same color) corresponds to the target achieving a doubling in camouflage effectiveness. Red asterisk symbols show the probability of not being detected (high values = high camouflage effectiveness). The blue columns are the normalized averages of the median detection times from each scene, while the yellow columns are the normalized medians of all the medians from each scene.

Because Figure 10 contains compressed and important results, we will explain the figure content in detail. Figure 10 includes the same kind of information as all the figures representing individual scenes (as shown in Figure 8 and Figure 9), where each target is ranked based on median detection times (the blue and yellow columns) and based on the probability of being non-detected (red asterisk symbol). Figure 10 is essentially a compressed version of all 11 scenes added together. Each target was assigned a performance number given by the length of the corresponding two columns (blue and yellow). As there is no standardized method of adding results from different scenes, resulting in some overall performance, we present two columns that are assumed relevant for the overall performance. The blue columns are the normalized averages of the median detection times from each scene, while the yellow columns are the normalized medians of all the medians from each scene. For simplicity, we will refer to the blue columns when explaining how to interpret the results below.

The column length assigned with a target measures camouflage effectiveness in all scenes, where the length of the columns should be interpreted as the performance numbers of a given target. The relative differences between the target performances are preserved because the lengths of the columns are related to detection times (albeit given some normalized values). To extract reliable results from the extensive detection data set from the different scenes, we must preserve the individual performance differences amongst the targets for each scene and then add them together. This means that a column that achieved double length compared with another column should be interpreted as the observers of that corresponding target needing twice as long a search time to detect the target (relative to the target assigned with the short column). Hence, the differences in camouflage effectiveness amongst the targets can be read off directly as the differences in the length of the columns. In particular, this means that columns achieving higher detection time values than 1 in Figure 10 correspond to the target performing better than the average of all targets in the test, and columns achieving lower detection time values than 1 in Figure 10 correspond to the target performing poorer than the average of all targets in the test.

We have assigned a red asterisk symbol to each target in Figure 10. The asterisk indicates the average probability of a target remaining undetected (the number of non-detections of a target for all scenes divided by the total number of observers exposed to that particular target). From Figure 10, we see that the detection time (blue/yellow columns) and the probability of non-detections (red asterisk symbols) correlate, meaning that targets that had high median detection times (yellow columns) also had a correspondingly high percentage of non-detections, and thus supporting the validity of the results. Similar to the detection time columns in the figure, the targets' probability of remaining undetected followed the same trends, where long detection times (i.e., high camouflage effectiveness) generally correlated with high probabilities of non-detections (i.e., high camouflage effectiveness).

The results presented in Figure 10 show that target W00 performed significantly worse than the other targets, with a high probability of being detected and short detection times. The other targets (W01–04), with a more snow-like white color, performed comparably and had similar detection times and probability of being detected. Target W01 had a slightly higher percentage of no-detections than target W02–04, but likely within measurement errors. In section 4.1, we discuss the results further.

4 DISCUSSION

4.1 How to interpret the result columns

As shown in Figure 10, each target is assigned two columns, one blue and one yellow. Both columns represent the detection times of a given target in all scenes but with two different ways of "summing" over the scenes. The blue column shows the normalized average of the median detection times from all the scenes for one specific target. For each target, we added all median times (one per scene) together and then calculated the average of the sum. We performed this calculation for all targets individually. Because each scene was unique and with significant variations in the degree of difficulty (regarding target detection), we divided the median times of each scene by a difficulty factor. We defined the difficulty factor as the average of all the targets' medians for each scene. For example, the difficulty factor in scene 11 (Figure 9) is then: $(4+15+14+50+4)/5 \sim 17$.

The yellow column shows the normalized median of the median detection times from all the scenes for a given target. The yellow columns contain the same information as the blue ones, only in a slightly different wrapping. Both types of columns are calculated from distributions of median times of each target specifically. As it is not obvious how to pick one single number to represent such distributions of median times [45, 46], we have calculated both the average (blue columns) and median (yellow columns). The median-based columns (yellow) are typically shorter than those based on the averages (blue). Yet, being uncertain which of the two abovementioned measures is the most correct, we note that switching from one of them (blue/yellow) to the other (yellow/blue) does not alter the rank of the targets, nor does it change the overall impression that some targets performed much better than others in Norwegian winter backgrounds.

4.2 Camouflage performance related to pattern and color

From the results of the observer test (Figure 10), we found that target W00 overall performed significantly poorer than all the other targets, i.e., the target was easier to detect by the observers, both due to short detection times and high probability of detection. We believe the poor camouflage performance of target W00 was directly related to the large color difference from snow. In Table 1, we presented the quantified color difference between the white color of the targets and the snow. We found that target W00 had a significantly more significant color difference to snow than the other targets. A low color difference is essential to achieve low contrast to arctic winter backgrounds where snow dominates the scenes. Moreover, target W00 did not have a camouflage pattern to disrupt the target's unnatural shape, which also is assumed to have contributed negatively to the target's camouflage performance.

Target W04 did not have a camouflage pattern like target W00 but had a much better color resemblance to snow. For the scenes used in this study, this indicates that color contrast to snow was more important than camouflage pattern. Targets with camouflage patterns (W01–03) performed slightly better than target W04 with the same white colors as these targets but without a pattern. Additional studies are required to differentiate the camouflage performance of the patterns, for example, in scenes with direct sunlight or in scenes where the targets are more exposed.

4.3 Validity of the results

The number of specific scenes in this study was 11, comparable to those used in similar studies and higher than similar camouflage studies where isolated sub-topics were studied [27, 29, 30, 33, 34, 47-50]. No exact number of scenes is considered a minimum for a test to be reliable. Generally, the larger number of distinct scenes, the higher reliability of the results can be assumed. Conversely, a much larger number of scenes than in this study is time-consuming and additional scenes will eventually look similar to those already captured. Also, many scenes will increase the trial time per observer and potentially reduce the data quality if the test is too long or too tedious, reducing the observers' focus and concentration.

The results in the 11 distinct scenes show large variations in performance amongst the targets from one scene to the next. This was expected because we have many different winter scenes in Norway, and in this study, we tried to vary the types of scenes as much as possible with our 11 scenes. We have selected scenes from different regions in Norway (north and south) where the regions were considered relevant to winter camouflage purposes. We have yet to try to classify the 11 scenes to see which target performed better in any specific location or terrain type, as this is both difficult and irrelevant to the primary purpose of this study. Note that all of the eleven scenes were recorded in overcast conditions.

In this study, the number of observers per target in each scene was 7–8, slightly fewer than usual [27, 29, 30, 33, 34, 47-50]). However, we believe the decent number of scenes (11) and the homogenous group of observers (soldiers) should compensate for the relatively low number of observers and smooth out statistical variations due to potential differences in observance skills and prerequisites.

This study aimed to evaluate the performance of different snow camouflage patterns and colors. The overall results (Figure 10) clearly show substantial differences in performance between the camouflage target W00 and the other targets (W01–04). There is no reason to believe such a performance gap would be altered significantly by adding more scenes and observers to the test. Therefore, the main results in this study are considered reliable and serve as a basis for follow-up development and evaluation to improve the snow camouflage in use.

4.4 Alternative ways of evaluating specific camouflage performances

In this study, we chose detection time as our primary evaluation criterion and detection probability as a supporting secondary criterion. Our results show that the overall rank of the targets exhibited the same general trends independent of whether detection time or probability was used to assess them (Figure 8–Figure 10). This correlation strengthens the main findings because they did not depend on which criterion we used. It also supports what FFI has experienced from other camouflage evaluation studies [40]. Our preferred ranking criterion (detection time) is also supported by similar studies conducted by other NATO nations that are upfront in evaluating camouflage performance [7, 10, 11, 25, 26, 32, 34, 35, 38, 39].

As discussed in previous FFI studies and also shown by other studies conducted by NATO nations [26], there is no golden standard on how to assess the camouflage performance of combat uniforms. Likely, there are also other criteria well suited. The optimal criterion may be related to the primary purpose of the camouflage (long detection times, remaining undetected, challenging to track when in motion, or other).

5 CONCLUSION

This study evaluated the performance of five winter camouflage targets in visible light. Two targets were white (W00 and W04) but with different white colors, whereas the three other targets had grey patterns on white (W01-03). The targets were photographed in 11 distinct scenes in Norway, followed by an observer-based photosimulation trail undertaken by a large group of Norwegian soldiers: detection time and probability of detection were used to quantify the performance of the targets. In the trail, W00, of the white targets, was overall the easiest target for the soldiers to see. Most soldiers detected the target fast and usually within the maximum allowed detection time for each scene. The other white target, W04, was much more challenging for the soldiers to detect, i.e., the target had a significantly better concealment performance. To study the effect of color-likeness to snow on the performance, we measured the color coordinates of the textiles and compared them to snow. We found that the white color on target W04 was markedly more similar to snow than that of target W00.

Targets W01–03 had the same white color as target W04 but with additional grey patterns on top. Our study shows that these targets were slightly more challenging for the soldiers to detect than target W04. The differences in camouflage performance between the targets with patterns (W01–03) were insignificant and might require further studies on a wider variety of winter scenes to be substantial. Therefore, we conclude that color matching to snow was more critical than camouflage patterns in the scenes used for the photosimulation test.

The scenes used for the photosimulation trial were all captured during overcast conditions. Future work should include scenes with direct sunlight and compare the camouflage performance to cloudy conditions. As sensors operate at a wide range of spectra, performing a similar study of targets captured at infrared wavelengths would also be interesting.

REFERENCES

[1] A. Beevor, "Russia's New Winter War," Foreign Affairs, (2022).

[2] T. Koivurova, M. Heikkilä, J. Ikävalko *et al.*, "Arctic cooperation in a new situation: Analysis on the impacts of the Russian war of aggression," (2022).

[3] C. Wall, and N. Wegge, "The Russian Arctic Threat," (2023).

[4] J. A. Endler, "Disruptive and cryptic coloration," Proceedings. Biological sciences, 273(1600), 2425-2426 (2006).

[5] J. A. Endler, "Progressive background in moths, and a quantitative measure of crypsis," Biological Journal of the Linnean Society, 22(3), 187-231 (1984).

[6] S. Merilaita, N. E. Scott-Samuel, and I. C. Cuthill, "How camouflage works," Philosophical Transactions of the Royal Society B: Biological Sciences, 372(1724), 20160341 (2017).

[7] C. Michalis, N. E. Scott-Samuel, D. P. Gibson *et al.*, "Optimal background matching camouflage," Proceedings of the Royal Society B: Biological Sciences, 284(1858), 20170709 (2017).

[8] I. C. Cuthill, M. Stevens, J. Sheppard *et al.*, "Disruptive coloration and background pattern matching," Nature, 434(7029), 72-74 (2005).

[9] R. Hecker, "Camaelon - Camouflage assessment by evaluation of local energy, spatial frequency and orientation," SPIE, 1687, 342-349 (1992).

[10] M. Friškovec, and H. Gabrijelčič, "Development of a procedure for camouflage pattern design," Fibres & Textiles in Eastern Europe, 18(4), 81 (2010).

[11] A. Toet, and M. Hogervorst, "Urban camouflage assessment through visual search and computational saliency," Optical Engineering, 52(4), 041103 (2012).

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

[13] M. Stevens, and S. Merilaita, "Defining disruptive coloration and distinguishing its functions," Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1516), 481-488 (2009).

[14] M. Stevens, and S. Merilaita, "Animal camouflage: current issues and new perspectives," Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1516), 423-427 (2009).

[15] T. Troscianko, C. P. Benton, P. G. Lovell *et al.*, "Camouflage and visual perception," Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1516), 449-461 (2009).

[16] K. B. Toh, and P. Todd, "Camouflage that is spot on! Optimization of spot size in prey-background matching," Evolutionary Ecology, 31(4), 447-461 (2017).

[17] C. A. Párraga, T. Troscianko, and D. J. Tolhurst, "The human visual system is optimised for processing the spatial information in natural visual images," Current Biology, 10(1), 35-38 (2000).

[18] A. Mondal, "Camouflage design, assessment and breaking techniques: a survey," Multimedia Systems, 28(1), 141-160 (2022).

[19] A. Mikkelsen, and G. Selj, "Spectral characteristics of moist birch leaves and synthetic materials: experimental studies and evaluation of models," Optical Engineering, 60(11), 117102 (2021).

[20] G. Selj, and A. Mikkelsen, [Spectral reflectance measurements of snow and snow covered objects: experimental studies compared with mathematical models] SPIE, ESD (2021).

[21] M. Winkelmann, [Analysis of exploitable spectral features of target and background materials] SPIE, ESD (2015).

[22] H. Kariis, and C. Åkerlind, [Overview of the adaptive camouflage for the soldier II (ACAMSII)] SPIE, ESD (2021).

[23] A. Schwegmann, "Camouflage evaluation by bio-inspired local conspicuity quantification," Proc. SPIE, 10794, 107940H (2018).

[24] E. Peli, "Contrast in complex images," Journal of the Optical Society of America A, 7(10), 2032-2040 (1990).

[25] M. A. Hogervorst, A. Toet, and P. Jacobs, "Design and evaluation of (urban) camouflage," SPIE Defense, Security, and Sensing, 7662, (2010).

[26] A. Toet, and M. Hoogervorst, "Review of camouflage assessment techniques," Proc. SPIE, 11536, 1153604 (2019).

[27] G. K. Selj, and M. Søderblom, "Discriminating between camouflaged targets by their time of detection by a human-based observer assessment method," SPIE Security + Defence, 9653, (2015).

[28] D. Heinrich, and G. K. Selj, "Evaluation of camouflage pattern performance of textiles by human observers and CAMAELEON," Proc. SPIE, 10432, 1043206 (2017).
[29] G. K. Selj, and D. Heinrich, "Search by photo methodology for signature properties assessment by human observers," Proc. SPIE, 9474, 947411 (2015).
[30] G. Selj, and D. Heinrich, [Disruptive coloration in woodland camouflage: evaluation of camouflage effectiveness due to minor disruptive patches] SPIE, ESD (2016).
[31] C. M. Birkemark, "CAMEVA: a methodology for computerized evaluation of camouflage effectiveness and estimation of target detectability," AeroSense '99, 3699, (1999).
[32] J. E. Peek, L. Hepfinger, R. Balma *et al.*, [Guidelines for camouflage assessment using observers (instructions pour les evaluations de camouflage faisant appel a des observateurs)(cd-rom)] NATO RESEARCH AND TECHNOLOGY ORGANIZATION (FRANCE), (2006).
[33] F. Racek, A. Jobánek, T. Baláž *et al.*, [Evaluation of validity of observer test for testing of camouflage patterns] SPIE, ESD (2018).
[34] M. A. Hogervorst, A. Toet, and P. Bijl, "Human search with a limited field of view: effects of scan speed, aperture size, and target conspicuity," Optical Engineering, 52(4), 1-14, 14 (2013).
[35] M. A. Hogervorst, P. Bijl, and A. Toet, "On the relationship between human search strategies, conspicuity, and search performance," Defense and Security, 5784, (2005).
[36] R. T. Marcus, [chapter 2 - The Measurement of Color] North-Holland, (1998).
[37] W. Mokrzycki, and M. Tatol, "Colour differenceΔ E-A survey," Mach. Graph. Vis, 20(4), 383-411 (2011).
[38] U. S. A. N. S. Research, [Photosimulation camouflage detection test], (2009).
[39] F. M. Gretzmacher, G. S. Ruppert, and S. Nyberg, "Camouflage assessment considering human perception data," Aerospace/Defense Sensing and Controls, 3375, (1998).
[40] D. Heinrich, and G. K. Selj, [Evaluation methods of signature effectiveness - a first evaluation of camouflage assessments by CAMAELEON compared to human observers], (2016).
[41] S. K. Mannan, "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).
[42] A. Toet, P. Bijl, F. Kooi *et al.*, "A high-resolution image data set for testing search and detection models," TNO-report TM-98-A020, (1998).
[43] K. Cooke, "The sources of variability in the search process," RTO MP-45, (2000).
[44] L. G. Williams, "Target conspicuity and visual search," Human Factors, 8, 80-92 (1966).
[45] P. J. Bickel, and K. A. Doksum, [Mathematical statistics - Basic ideas and selected topics] Chapman and Hall/CRC, (2015).
[46] S. S. Sawilowsky, "Misconceptions leading to to choosing the t test over the Wilcoxon Mann-Whitney test for shift in location parameter," J Mod Appl Stat Method, 4, 598-600 (2005).
[47] G. Selj, [Disruptive camouflage tricks the human eye: a study of detection times of two near-similar targets in natural backgrounds] SPIE, ESD (2015).
[48] Q. Jia, X. L. Lv, W. D. Xu *et al.*, "Intelligent design of gradual disruptive pattern painting and comparison of camouflage effectiveness," Cluster Computing, 22(4), 9293-9301 (2019).
[49] J. Culpepper, and V. Wheaton, [A target detection model predicting field observer performance in maritime scenes] SPIE, ESD (2014).
[50] C.-C. Chang, Y.-H. Lee, C. J. Lin *et al.*, "Visual Assessment of Camouflaged Targets with Different Background Similarities," Perceptual and Motor Skills, 114(2), 527-541 (2012).