TITLE PAGE

Camouflage Assessment: Machine and Human

Timothy N. Volonakis[1,3*], Olivia E. Matthews[1], Eric Liggins[4], Roland J. Baddeley[1], Nicholas E. Scott-Samuel[1], Innes C. Cuthill[2]

*Corresponding author

[1] School of Experimental Psychology, University of Bristol, 12a Priory Road, Bristol, BS8 1TU, UK

[2] School of Biological Sciences, University of Bristol, 24 Tyndall Avenue, Bristol, BS8 1TQ, UK

[3] Centre for Machine Vision, Bristol Robotics Laboratory, University of the West of England, Frenchay Campus, Coldharbour Lane, Bristol, BS16 1QY, UK

[4] QinetiQ Ltd, Cody Technology Park, Farnborough, Hampshire, GU14 0LX, UK

Abstract

A vision model is designed using low-level vision principles so that it can perform as a human observer model for camouflage assessment. In a camouflaged-object assessment task, using military patterns in an outdoor environment, human performance at detection and recognition is compared with the human observer model. This involved field data acquisition and subsequent image calibration, a human experiment, and the design of the vision model. Human and machine performance, at recognition and detection, of military patterns in two environments was found to correlate highly. Our model offers an inexpensive, automated, and objective method for the assessment of camouflage where it is impractical, or too expensive, to use human observers to evaluate the conspicuity of a large number of candidate patterns. Furthermore, the method should generalize to the assessment of visual conspicuity in non-military contexts.

Key Words: Camouflage Assessment, Observer Modelling, Visual Search

Declarations of interest

This research was funded by studentships to TV and OM funded by QinetiQ under the Dstl MAST programme.

Acknowledgements

We thank the Engineering & Physical Sciences Research Council, UK, (grant EP/M006905/1) for financial support. We thank QinetiQ; William Serle and Ian Moorhead, for their support and input. We thank three anonymous referees and the editor for their constructive comments

1. Introduction

Military personnel and equipment need protection from detection during conflict. Camouflage is the primary method to achieve this, frequently through coloured textures that match the background and/or disrupt the object's outline (Hartcup 2008; Merilaita et al. 2017; Talas et al. 2017). Assessment of effectiveness can be carried out in a number of ways. The most intuitive method is to use human participants as observers. Such an apparently straightforward procedure, however, is not only limited by uncontrollable conditions, such as the weather: it is also impractical given the large variety of objects/patterns that one might want to evaluate and the range of environments one might want them to be assessed in. Field trials are also expensive and, in some circumstances, may not even be possible. They also do not lend themselves to precise isolation of exactly what leads to the failure of camouflage, something that a paired comparison of otherwise identical target-present and target-absent scenes would allow. Photo-simulation attempts to overcome weather constraints and problems with inaccessible environment-types by using photographic or synthetic imagery. Recent advances in synthetic rendering are impressive; however, current methods are still computationally expensive and the images are unrealistic at small spatial scales due to the current limitations of simulating realistic ray scattering. Furthermore, human experiments are necessarily subjective and do not readily allow evaluation of camouflage against autonomous systems perhaps operating using different spectral bandwidths than the human vision. A computational approach is therefore helpful in overcoming the limitations of assessing camouflage when using human observers. Such a computational model should be ideally designed, in the first instance, in accordance with the human visual system, since it will be performing the task of a human observer and, if it is to replace subjective assessment, needs to be compared with human performance. More generally, however, such a system could be adapted to have a different ‘front end’ (e.g. infra-red sensor, hyperspectral sensor). Therefore it is surprising that a biologically motivated design for the assessment of camouflage has not been implemented.

This omission means that the confidence and extendibility of current models and metrics are low, falling short in their ability to cope with high dynamic range (i.e. natural) (Bhajantri and Nagabhushan, 2006; Hecker, 1992; Sengottuvelan et al., 2008), semi-automatic labelling or tracking of the target (Chandesa et al., 2009), non-probabilistic and non-scalable distance metrics to high dimensional data or multiple observations given many images (Birkemark, 1999; Heinrich and Selj, 2015; Kiltie et al., 1995). Human behavioural data need to be recorded to assess the coherence between human and model observers. This requires tasking human and model observers with the same experiment, based on a stimulus set from the real world: outdoor environments and militarily relevant objects.

2. Method

An experiment was devised so that human participants and a model observer could both be tasked with it, allowing for direct comparison. This method section is broken down into the three components that comprise this study: (i) images of objects placed in real world scenes were photographed and calibrated; (ii) a human experiment, using a protocol from psychophysics, recorded unbiased performance for recognition and detection of these objects; and (iii) the design of the visual observer model, and modelling the discrimination task.

2.1 Stimuli

Targets were photographed in two outdoor environments in the UK: Leigh Woods National Nature Reserve in North Somerset (2°38.6’ W, 51°27.8’ N), which is mixed deciduous woodland, and Woodbury Common in Devon (3°22' W, 50°40' N), a heathland used for Royal Marine training. A replica military PASGT helmet (Personnel Armor System for Ground Troops, the US Army’s combat helmet until the mid-2000’s) was the chosen object used in the experiment and visibility was manipulated by changes in helmet covers varying in both colour and textural appearance (Figure 1). The camouflage patterns worn by the helmet were United Nations Peacekeeper Blue (UN PKB), Olive Drab, Multi-Terrain Pattern (MTP, as used by the British Army since 2012), Disruptive Material Pattern (DPM, the dominant British Army pattern prior to the adoption of MTP), US Marine Pattern (MarPat) and, for the Woodbury Common experiment, Flecktarn (as used by the Bundeswehr, the German Army). These patterns were chosen not for the purpose of evaluation per se, but to reflect a range of styles (e.g. unpatterned Olive Drab, DPM as a subjective human design, MTP and MarPat based on spatio-chromatic analysis of natural scenes, but MarPat being ‘digital’ or pixellated), with UN PKB as a high visibility control.

For the computational approach to be useful, the spectrum of visibility across the patterns should be highly correlated in the model and human observers. Scene locations were selected on a meandering transect through the habitats, at 20 m intervals and alternating left and right. If the predetermined side was inaccessible or inappropriate due to occlusions then the opposite side of the transect path was used, and if neither side was accessible the interval was ignored and the next location in the transect was used. At each location the object was placed in a 3 × 3 grid resulting in nine images. The distance of each row of the grid was 3.5, 5 and 7.5 metres. The scene was also divided into 3 arcs: left, middle and right. The combination of distance and left-right positioning mean that, in the subsequent tests on humans, the location of the target within the scene was unpredictable. This resulted in nine images of each helmet per location for analysis, plus a scene including a Gretag-Macbeth Color Checker chart (X-Rite Inc., Grand Rapids, Michigan, USA) for calibration. The orientation of the helmet in each photograph was set an angle drawn randomly from the uniform distribution {0, 45, 90, 135, 180, 225, 270, 315°}. For efficiency of implementation, the list of random angles was generated before going into the field. Each scene was also photographed without a helmet present. Photographs were taken using a Nikon D80 digital SLR (Nikon Ltd., Tokyo, Japan) with focal length 35mm, exposure 1/30 and F-Number 16. RAW images (Nikon NEF format) were captured and these were subsequently converted to uncompressed 8-bit TIFF and calibrated. Images were calibrated by recording luminance and chromatic spectral values of the Gretag-Macbeth colour chart in the field using a Konica Minolta Chroma Meter CS - 100A colour and luminance meter (Konika, Tokyo, Japan). This process was repeated three times to average over the natural variation in lighting from moment to moment. The spectral values were transformed to the CIE sRGB colour space after first converting them to the CIE XYZ colour space. The process was then repeated in the lab from a projected image from the projector. A cubic polynomial approximated the relationship between the two sets of RGB measurements. Images were then calibrated using the coefficients of the polynomial for each RGB channel. Not only does this procedure avoid having a colour chart in every single image, but also it calibrates the entire pipeline in a single step: calibrating the camera, projector and images individually could result in over-fitting or multiplicative errors.

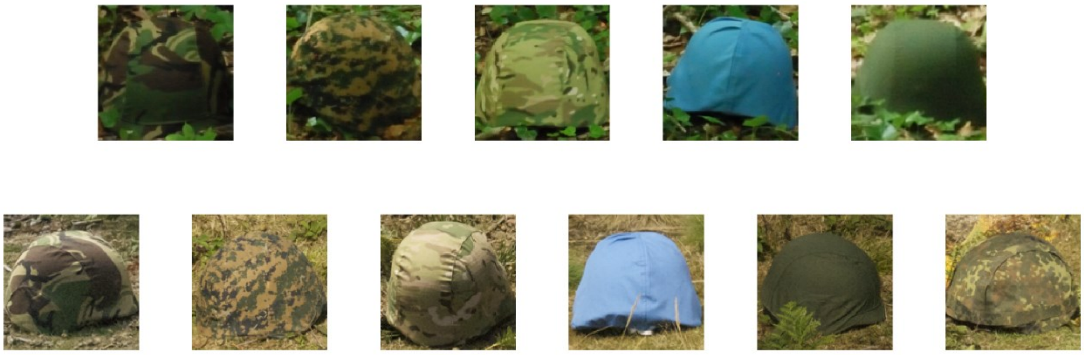

Figure 1. Example cropped helmet images from real world scenes

An example of each camouflaged helmet cropped for recognition purposes. From left to right the patterns that the helmet wears are DPM, MarPat, MTP, UN PKB, Olive drab and Flecktarn. The top row are the helmets from Leigh Woods and the bottom row are helmets from Woodbury Common. Flecktarn was only used in Woodbury Common.

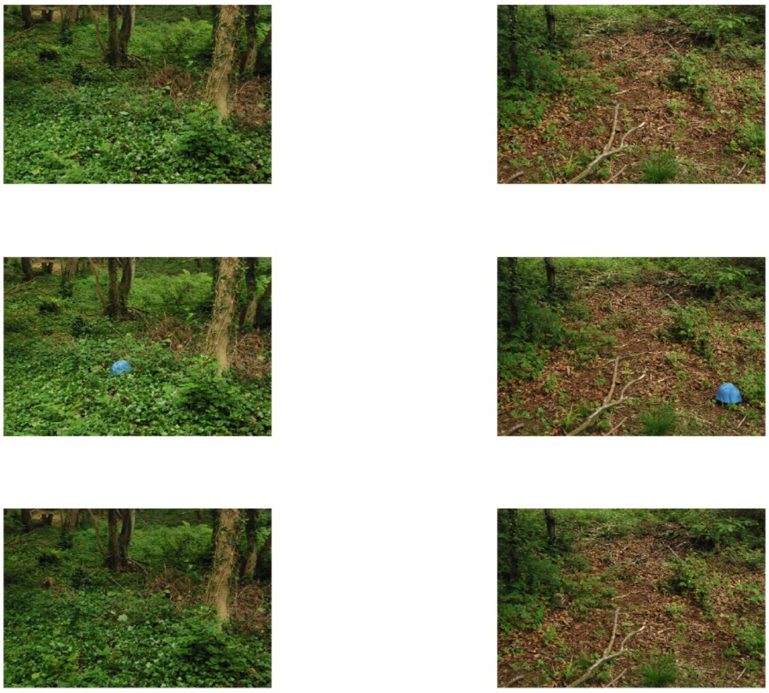

Figure 2. Example Leigh Woods scenes

Two example scenes from the Leigh Woods environment. The left column and the right column are two different scenes. The top two scenes do not contain a helmet. The middle two contain a UNPKB helmet. The bottom two contain the DPM helmet.

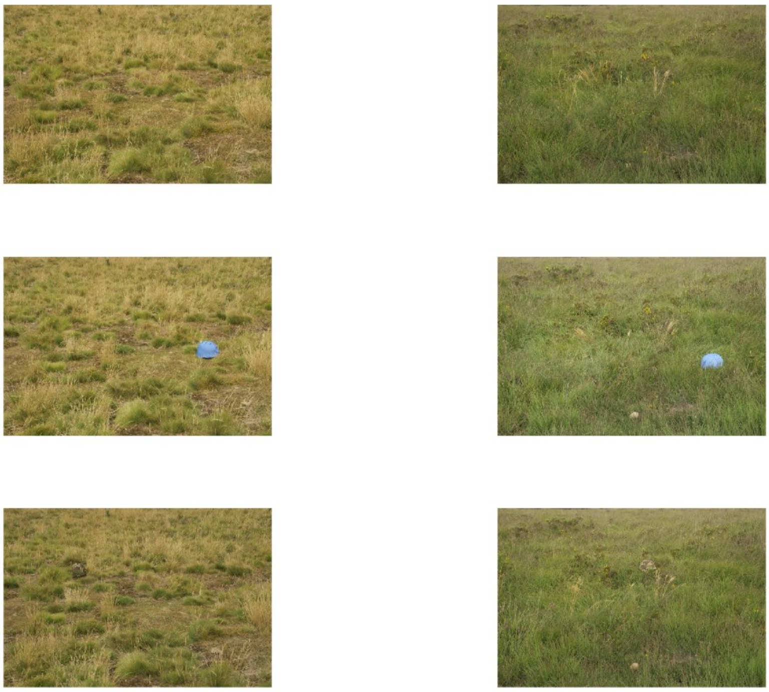

Figure 3. Example Woodbury Common scenes

Two example scenes from the Woodbury Common environment. The left column and the right column are two different scenes. The top two scenes do not contain a helmet. The middle two contain a UNPKB helmet. The bottom two contain the DPM helmet.

2.2 Human Experiment

2.2.1 Participants and Materials

A human experiment using 22 participants for the Leigh Woods dataset and another 20 participants for the Woodbury Common dataset was conducted. Each of the two experiments had an equal proportion of each gender. Images were projected onto a 190 × 107cm screen (Euroscreen, Halmstad, Sweden) from 310cm using a 1920 × 1080 pixel HD (contrast ratio 300,000:1) LCD Projector (PT-AE7000U; Panasonic Corporation, Kadoma, Japan). Participants were seated at a distance of 255 cm from the screen and therefore images subtended 41° horizontally and 24° vertically.

2.2.2 Procedure

At the start of each block participants were informed which helmet to search for by presenting an image of the helmet; only one camouflage type was present in any one block. There were 27 and 22 trials per block, respectively, for Leigh Woods and Woodbury Common, and the order of patterns across blocks and replicated within blocks were separately randomised for each participant. A trial consisted of sequentially presenting two scenes for 250 ms with a 250ms blank screen, of luminance and chromaticity equal to the mean of all the test images, immediately followed by a 250 ms cue screen, prior to each scene. One of the scenes presented contained a helmet and the other did not, the order being randomised. The participant's task was a two alternative force choice, reporting which of the two scenes contained the helmet. Responses were given using the number keys one and two on the keyboard, reporting the first scene or the second scene respectively during a 1000 ms response period after each pair of scenes. A presentation time of 250 ms is enough time to ensure a single saccade to the cued location, but not long enough to allow more complicated scan patterns. Since these scan patterns will be possibly highly variable, they will introduce variability into the responses above and beyond that due to the stimuli, and hence decrease the power of the study. One thousand milliseconds were allowed for subject's responses to allow more than adequate time to respond, but not so long as to increase the time of the total experiment. There were four general conditions of viewing, the factorial combination of two levels of colour information and two levels of location cueing. Cueing was of interest to separate effects of pattern recognition from detection, because the model was initially designed for recognition. Colour was of interest because it has been suggested that camouflage is more effective when there is chromatic as well as spatial noise (Melin et al., 2007; Morgan et al., 1992). In the first cueing condition, ('cued'), participants were cued to the location of the helmet. In the scene that did not contain the helmet, this cue's location was a random selection of one of nine possible pre-determined target locations. In the second condition, ('uncued'), the cue was presented in the centre of the screen for both scenes. The spatial cue was a white circle, 50 pixel diameter, 5 pixel line width, circle that was presented for 250ms. The whole experiment was repeated in greyscale and colour. As with pattern, the order of conditions for each participant was randomised.

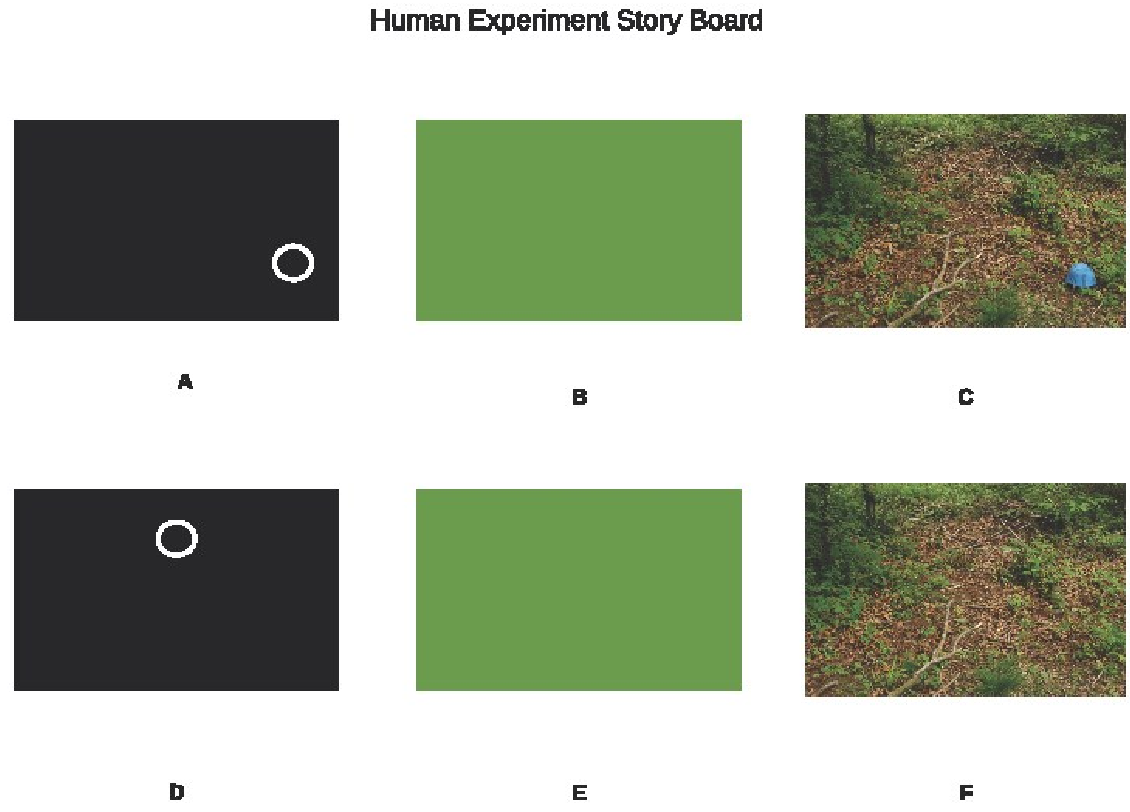

Figure 4. Human experiment storyboard

Storyboard for one trial in the experiment. Sequence is in alphabetical order. Duration of each interval was 250msec. Either C or F contains the helmet. Intervals A and D cue the participant to the spatial location of the helmet. Intervals B and E present a blank interval of average chromaticity across all scenes. At the end of the sequence, participants are asked which scene the helmet was in and are given 1000msec to respond. The procedure is identical for the uncued condition however the spatial cue in A and D are uninformative.

2.3 The Human Observer Model

2.3.1 The Model Framework

The model is a four-stage process as outlined below. By modelling low level visual processing, a side effect of the features chosen produces Gaussian variation from small metric distortions. The resultant Gaussian variation can then be approximated using a mixture of multivariate Gaussian distributions. The centre of each Gaussian distribution stores a familiar view. Probabilistic principal components (Tipping and Bishop, 1999b) describe the variability in an interpretable way to recognise unseen and unfamiliar views. Estimating the density and evaluating the maximum posterior probability determines the object class. This method turns the difficult problem of learning a complex invariant representation of an object into the simple problem of estimating parameters of a mixture of multivariate Gaussian distributions.

Stage 1. Filter Images with a Log Gabor Filter Bank

Grey scale images are cropped to a square and resized to 128×128 pixels, preserving the aspect ratio of the object. They are then filtered by a log Gabor wavelet filter bank, comprising three spatial scales (wavelengths of 16, 32 and 64 pixels), and four orientations (0, 45, 90 and 135°) (Kovesi, 2000). This first stage captures the early linear properties of the visual system. Whilst 2D Gabors can be used to approximate simple cells (Daugman, 1985; Jones and Palmer, 1987), we know that (i) simple cells are tuned to spatial frequency with a Gaussian bell-shaped tuning curve on a log frequency scale (De Valois et al., 1982; Field, 1987) and (ii) the Gabor filter has a D.C. component. The power in natural images is dominated by the D.C. component (Field, 1987), and, given that the cosine Gabor is sensitive to it and the sine Gabor is not, it will corrupt any computation of phase information in the next stage. The solution to both these problems is to employ log Gabors instead, which do not have a D.C. component (Kovesi, 1999).

Stage 2. Process the Filtered Output

Next we compute local energy and phase from the filtered output in stage 1. Stage 2 accounts for two non-linear properties of the visual system, illumination invariance and shift invariance. The energy is logged and the effect is two-fold: (i) the energy is positive, and not symmetrical for Gaussian approximation in the fourth stage; and (ii) introducing logarithms will turn differences in illumination into additive offsets. Denoting the response of the real and imaginary filters as R(x,y) and I(x,y), where x and y indicate the index in the image and atan2 computes the four quadrant arc tangent, log local energy and phase can be computed as Energy = ln|R(x,y)+I(x,y)|+c and Phase = atan2(I(x,y),R(x,y)), where c is a small constant, 0.05, to avoid the undefined logarithm of zero and || is the absolute. The absolute is the magnitude of the real (cosine log Gabor) and imaginary (sine log Gabor) filters. The sum of the squared filter responses is the magnitude, since $sin^2 + cos^2 = 1$. The energy loses local position, but confers some translational invariance and therefore small shifts are turned into small variations. Local energy represents lines as symmetrical Gaussians. Therefore the variance of these features is Gaussian through small metric distortions such as shift and object pose.

Phase angles will cycle from π to −π as the distortion moves through sampling locations, resulting in correlated variation. Phase information is a polar, circular variable; in order to use this feature for Gaussian approximation one must convert this feature into Cartesian space. Therefore the sine and cosine of the phase are computed, doubling the number of dimensions required for phase information. Concatenating this sampled local logged energy, sine and cosine phase information creates the feature vector.

Stage 3. Sample the Local Energy and Phase.

A hexagonal lattice, of equal size to the image, is placed over the image and the local energy and phase is sampled at the centres of each hexagon. A hexagonal lattice provides optimal sampling where samples are equidistant from each other (Yfantis et al., 1987). Phase angles vary less at larger spatial scales and therefore, to avoid over complete and redundant sampling, hexagonal lattices at larger spatial scales have fewer hexagons.

Stage 4. Evaluate Recognition Decision Using Bayes' Rule

The Gaussian variation computed in stage 2 can now be approximated. A unimodal distribution can represent a single view of an object. A mixture of Gaussians can model a multimodal distribution where multiple views of an object are learnt. The dimensions of each Gaussian component should represent the local variation of that view. The concatenation of the local energy and phase results in a high-dimensional feature vector and therefore a mixture of probabilistic components (Tipping and Bishop, 1999a,b) or a mixture of factor analysers (Ghahramani and Hinton, 1996) provides a local subspace for each Gaussian component and approximates the high dimensional covariance structure. To evaluate the recognition of an object, a model is created explicitly for each class. Likelihoods are computed for each explicit class and the posterior probability that an unseen object came from each object class is then evaluated using Bayes' rule, $P(A|B) = P(A)P(B|A)$. Where $P(A|B)$ is the posterior probability that the data A is from the object class B and $P(B|A)$ is the likelihood of data A under the object class B. The prior probability $P(A)$ is equal for each object class and this therefore cancels out.

2.3.2 Modelling the 2AFC Recognition Task

Human participants were tasked with recognising a helmet given two different images. One of the images contained a helmet and the other did not. For a direct comparison, both observers need to be tasked in a similar way. Ten-fold cross-validation was used to assess the model's accuracy. However, instead of evaluating a single image at a time, two images, one with a helmet and one without, were both evaluated under both background and helmet models. Therefore each image needs to be evaluated under both models, producing four likelihoods (Fig. 5). There are two scenarios: either the helmet is in image A or it is in image B. In the first scenario the helmet is in image A, where there is a high likelihood that it came from the helmet model and so the likelihood that image B came from the background class will therefore have a high likelihood. Bayes' rule will integrate over the mutually exclusive probabilities as shown in the diagram by incorporating the four likelihoods P(A|Helmet), P(A|Background), P(B|Helmet) and P(B|Background). Using Bayes' rule, the probability that image A is a helmet is simply:

1. $P\ (Helmet|A) = \frac{P(A|Helmet)\times P(B|Background)}{P(A|Helmet)\times P(B|Background)+P(B|Helmet)\times P(A|Background)}$.

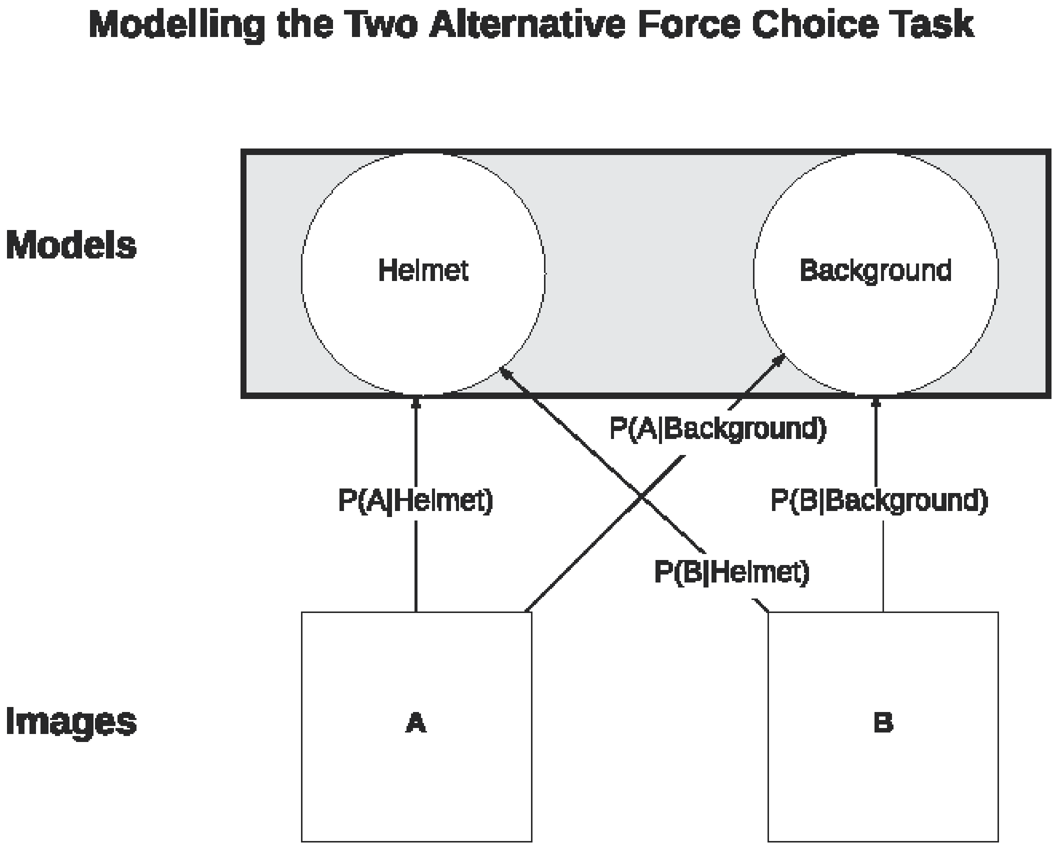

Figure 5. Graphical illustration at modelling the 2AFC procedure

To model the 2AFC task that humans were given, likelihoods under both models are computed for both images.

2.3.3 Modelling the Detection Task

The model is trained on a series of cropped images, where the object fills the crop. If the model is presented with an image of the target at a different spatial scale, i.e. the object does not fill the crop, it would be unable to recognise the object. To accommodate scale, likelihoods are computed for both the helmet and background classes at different spatial scales, at intervals of 10 ranging from the smallest helmet to the largest helmet across all images. Weightings are computed for each scale using Bayes' rule by evaluating which scale is most probable from the helmet class whilst evaluating that the other spatial scales belong to the background class. The weightings are multiplied with the likelihoods from each scale and summed. In short this procedure integrates probabilities over all spatial scales into a single likelihood for classification. This probabilistic approach, graphically demonstrated below where A and B denote two different sized crops at location in an image, is superior over simply taking the maximum, because the maximum only considers one model and if two scales are likely under the probabilistic approach the maximum would be too brittle and would ignore one of the likely scales. Equations below 2 - 7, show how Bayes' rule integrates the likelihoods over all the spatial scales, denoting two spatial scales A and B. Detection was modelled using leave-one-out cross-validation instead of the 2AFC approach. This was because there were too few scenes to compare the helmet scenes with. Problematically, if one were to compare likely peaks between two scenes, one scene would always have the same area of interest and this would be compared to many helmets. Leave-one-out cross-validation also provides a straightforward way to manipulate the training data so that the model did not see any of the scene whilst detecting the helmet.

$$2.\quad P(Helmet|A,B) = \frac{P(A|Helmet) \times P(B|Background)}{P(A|Helmet) \times P(B|Background) + \ P(B|Helmet) \times P(A|Background)}$$

$$3.\quad P(Helmet|B,A) = \frac{P(B|Helmet) \times P(A|Background)}{P(A|Helmet) \times P(B|Background) + \ P(B|Helmet) \times P(A|Background)}$$

$$4.\quad L1 = P(A|Helmet) \times P(Helmet|A,B) + \ P(B|Helmet) \times P(Helmet|B,A)$$

$$5.\quad L2 = P(A|Background) \times P(Helmet|A,B) + \ P(B|Background) \times P(Helmet|B,A)$$

$$6.\quad Posterior\ probability\ that\ helmet\ is\ at\ (x,y) = \frac{L1}{L1 + L2}$$

$$7.\quad Posterior\ probability\ that\ helmet\ is\ absent\ at\ (x,y) = \frac{L2}{L1+L2}$$

Equations 2-7 elaborate an example of how the model evaluates over spatial scale, where ***A*** and ***B*** denote two images each at a different spatial scale.

2.3.4 Colour

There are three main issues to consider when including colour: i) colour in the periphery, ii) efficient feature combination of texture and colour and iii) appropriate choice of colour space for measuring the distance between colours. The representation of short, medium and long wavelength receptors on its own is insufficient because computed distances in the colour space do not correlate with human perception (Tkaclic and Tasic, 2003; Wyszecki and Stiles, 1982). Projections in the CIE L*a*b* colour space are consistent with the judgements of human observers and are appropriate for discrimination purposes (Renoult et al., 2015). The model is a human observer model. Whilst recognition accuracy should be high, similar to human observers, it should not be able to recognise camouflaged objects all the time. The aim of the model is not to break camouflage and achieve perfect recognition. Therefore, instead of opting to use the CIE L*a*b* colour space, the MacLeod-Boynton chromaticity diagram is used. The MacLeod-Boynton chromaticity diagram (MacLeod and Boynton, 1979) is an isoluminant cone excitation space that is particularly good at discriminating large chromatic differences (Renoult et al., 2015). Modelling the detection of camouflaged helmets therefore is being treated as evaluating saliency, which this colour space has been shown to be successful at (Tatler et al., 2005). Colour is perceived differently in the periphery, because there are fewer cone receptors outside of the fovea (Hubel, 1995). The receptive field sizes in the periphery increase with eccentricity (Abramov et al., 1991), and therefore for objects to appear chromatically similar as if they were in the fovea, they must be spatially larger (Hansen et al., 2009; Vakrou et al., 2005). Given that an object is big enough to be scaled, the upper bound of eccentricity has been found to be 40° to 50° (Abramov et al., 1991; Hansen et al., 2009), after which it has not been found to be possible to simulate chromaticity as if it were in the fovea. An object that subtends 2° of visual angle has been found to appear approximately chromatically similar as if it were in the fovea up to 20° away. Therefore colour patterns can be simulated by low-pass-filtering the image (Mullen, 1985). Given the approximate appearance of foveal chromaticity with eccentricity up to 20° (half of the display), of objects that subtend 2° of visual angle, the scene was convolved with a Gaussian, whose standard deviation was measured to be 1° of visual angle, which was chosen so that it was comfortably smaller than 2°. It must be noted that the Gaussian blur is only an approximation and does not accommodate larger receptive fields as objects are more distant. The brightness varies the most across an image. Without processing the luminance, the mixture of Gaussians will have to explain this large variation, which will result in noisy likelihoods. The luminance information across all images could be normalised between one and zero, however that would no longer be Gaussian and, because we are only interested in chromaticity and not luminance at this point, the luminance channel was excluded and was therefore not modelled. Excluding the luminance channel is straightforward to do using some colour spaces such as hue, saturation and value (HSV), where luminance is represented in the channel named value, or opponency colour spaces such as the Macleod and Boynton or L*a*b*, where again the luminance is represented in its own channel. Removing the luminance channel is a standard method to avoid the large variance of brightness in images (Cai and Goshtasby, 1999; Shadeed et al., 2003). Instead of concatenating colour onto the feature vector of energy and phase, another Gaussian mixture model was trained for colour, allowing probabilities of colour and texture to be independent and a full covariance structure of colour to be modelled rather than a mixture of factor analysers. For each posterior map, the probabilities in the region where the target was located were logged and the maximum was taken. The summed log probabilities were plotted against human performance to visualise the correlation.

3. Results

Human data were not normally distributed and therefore a Generalised Linear Mixed (Effects) Model with binomial error and logit link function was used to generate interpretable means and error for analysis.   Figures 6 - 9 compare the model accuracy with that of human accuracy and below in table 1 are the correlation coefficients between the model and human observers for each condition. Correlations coefficients are very high, all above 0.85 with the exception of detection in Woodbury Common in colour.

Condition	Correlation
Leigh Woods	
Recognition	0.90
Detection Greyscale	0.93
Detection Colour	0.89
Woodbury Common	
Recognition	0.91
Detection Greyscale	0.87
Detection Colour	0.68

Table 1. The correlation coefficients between the model and human participants at 3 different conditions in two different environments, Leigh Woods and Woodbury Common

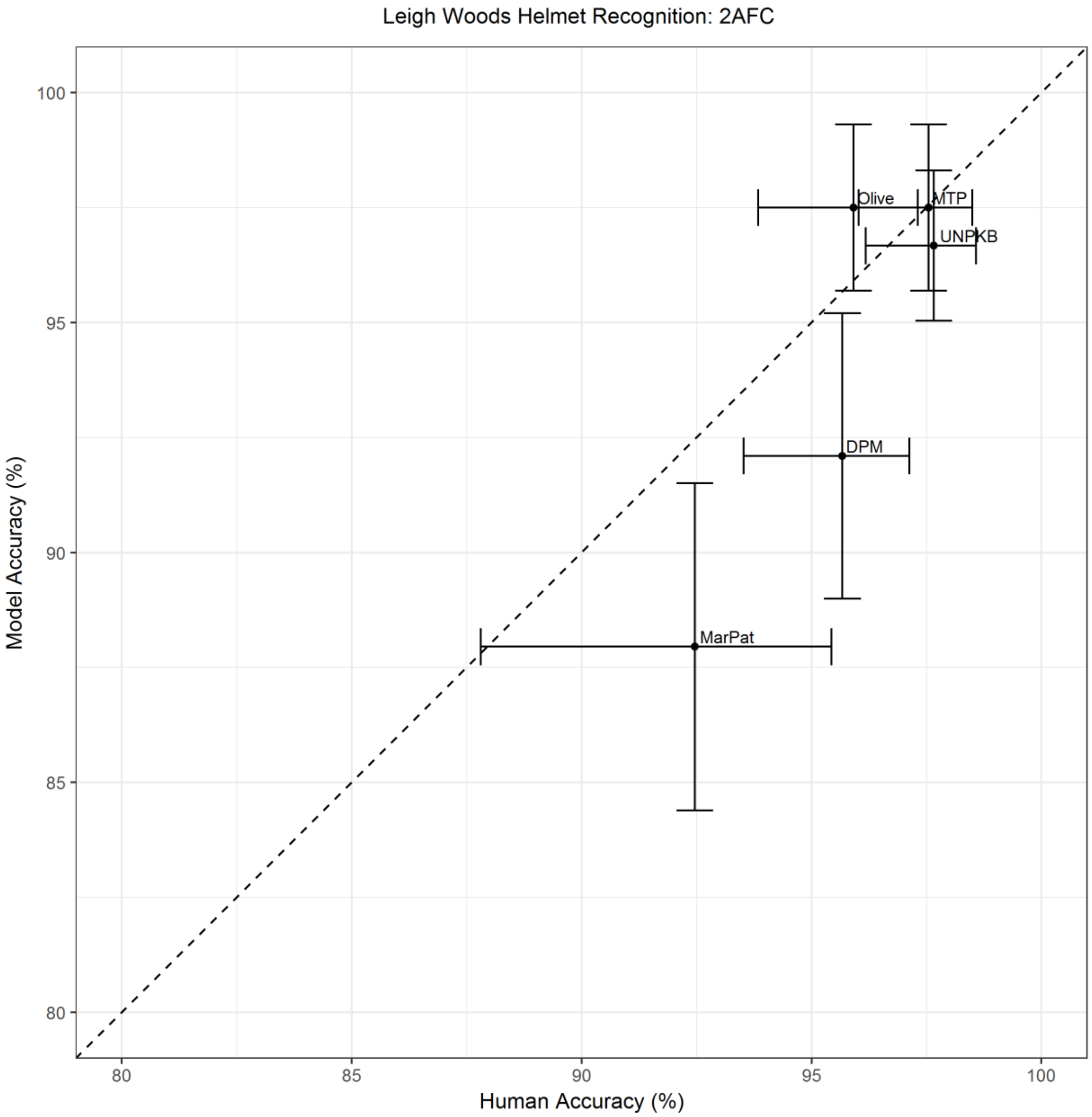

Figure 6. Human and model recognition accuracy: Leigh Woods

Leigh Woods model accuracy at recognition in greyscale plotted against human accuracy at recognition in greyscale. Correlation coefficient: 0.937. Error bars are 95% confidence intervals.

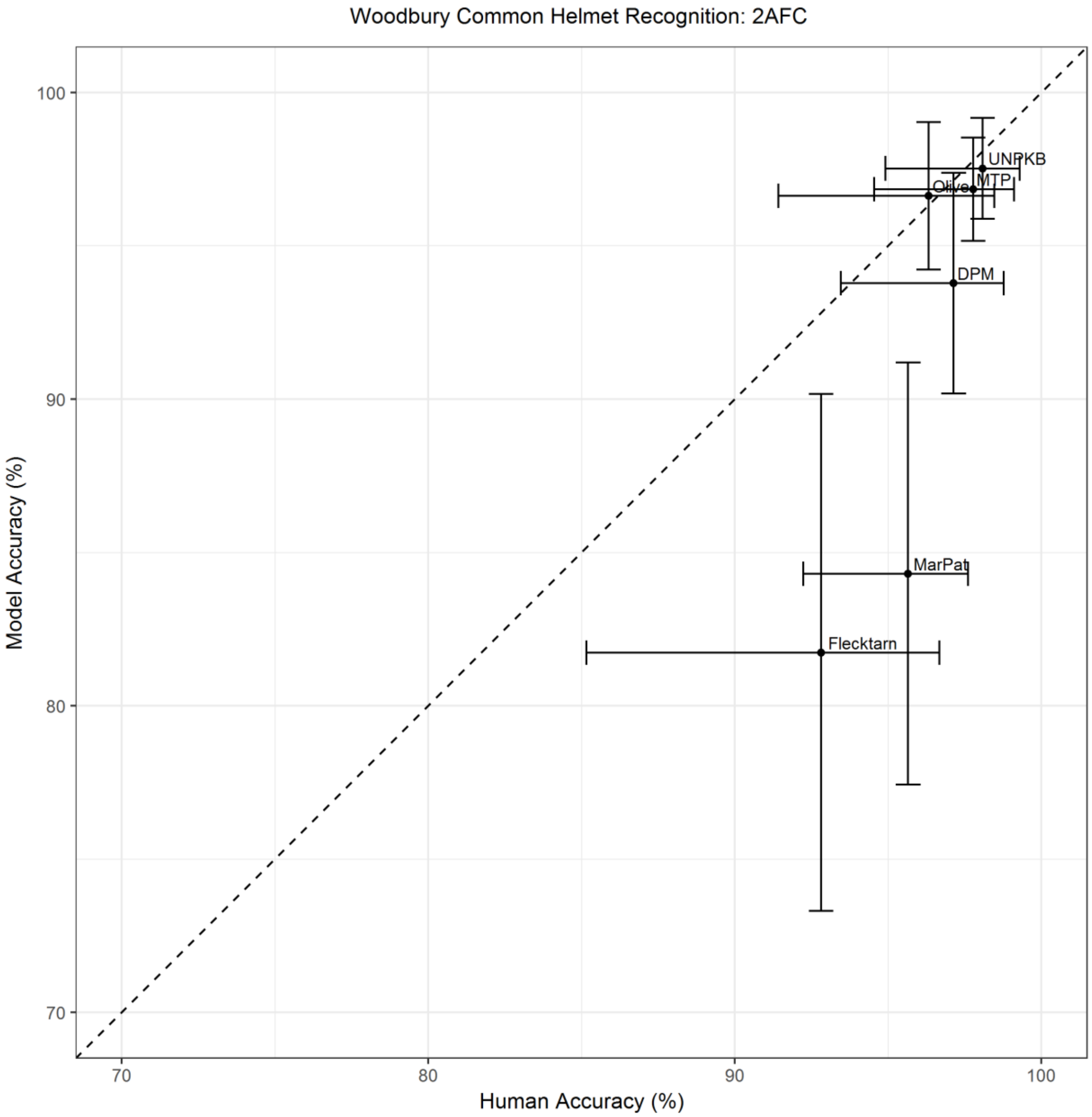

Figure 7. Human and model recognition accuracy: Woodbury Common
Woodbury Common model accuracy at recognition in greyscale plotted against human accuracy at recognition in greyscale. Correlation coefficient: 0.859.

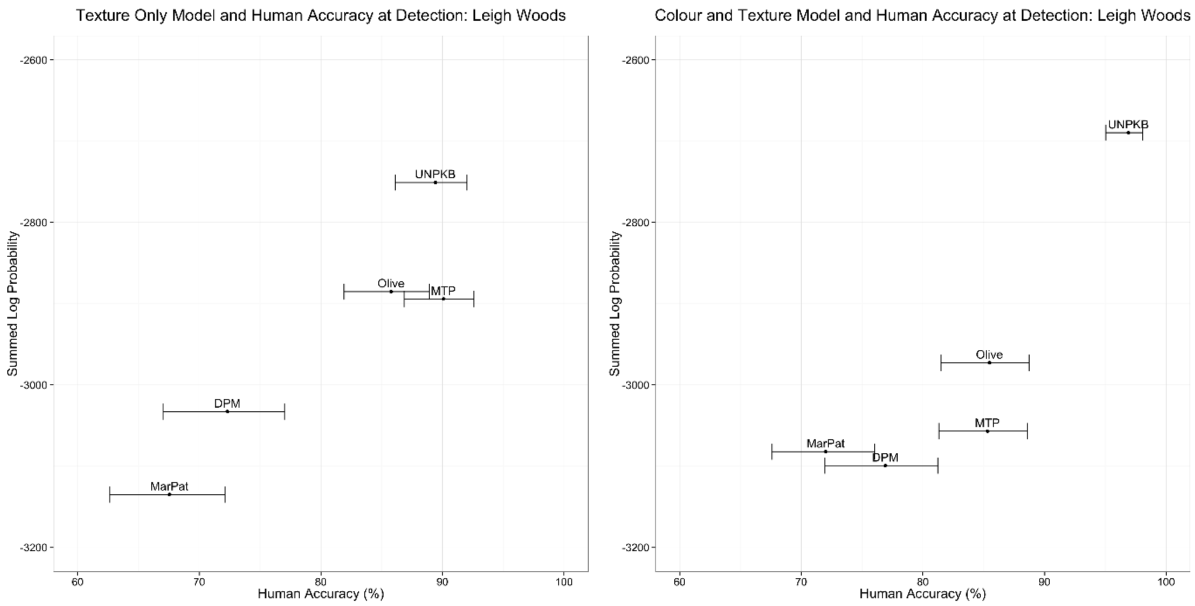

Figure 8. Human and model detection accuracy: Leigh Woods

Model and Human Accuracy at Detection in Leigh Woods. Left: Texture Only, Right: Colour and texture. Error bars are 95% confidence intervals.

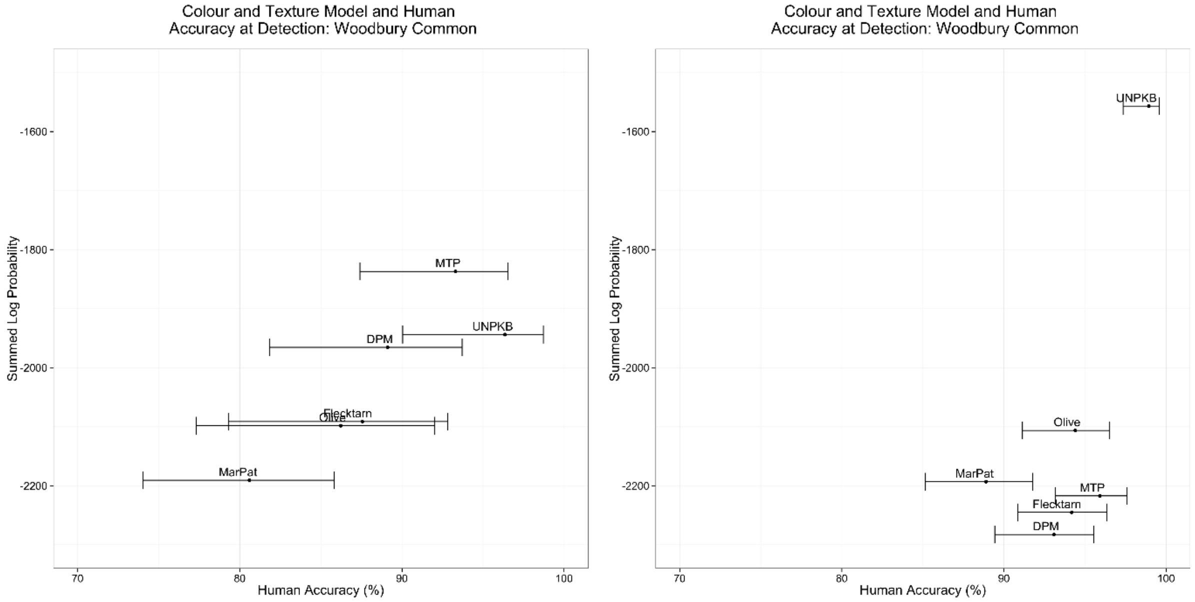

Figure 9. Human and model detection accuracy: Woodbury Common

Model and Human Accuracy at Detection in Woodbury Common. Left: Texture Only, Right: Colour and texture. Error bars are 95% confidence intervals.

4. Discussion

This paper has described and validated a visual recognition system that is designed to behave in a similar way to humans. The principles of its design are based upon low-level visual processing in the primary visual cortex. Although it is well-known that Gabor filters can approximate simple cells found in the primary visual cortex, and simple models using Gabor filters can achieve high recognition accuracy on simple datasets (Pinto et al., 2008), we present physiological evidence and a computational argument for the use of log Gabor filters. Such applicability of a human observer model is high, because using human participants is impractical given a variety of viewpoints, environments and objects. This paper also defined a task, a judgement of whether a target is present or absent in a scene, that would allow a direct comparison between the biologically motivated visual observer and human participants. The analysis of the behavior from both observers provides the necessary evidence to assess whether the model is an adequate surrogate for a human observer. The task was to estimate the accuracy with which camouflaged objects, military helmets with different coverings, could be detected and recognised. The selection of a single object class with different colour patterns, rather than an array of different objects, avoided the problem of object choice and allowed visibility to be easily controlled through only colouration and textural properties. The visibilities of the objects were unknown prior to the experiment because, to our knowledge, they had never been evaluated in the two environments nor directly compared. However, a priori, the UN PKB helmet was expected to be easy to detect, the Olive Drab harder to detect and the three (Leigh Woods) or four (Woodbury Common) patterned camouflages hardest to detect. It was essential that the visibility of the patterns varied. If human recognition and detection for all camouflaged objects was at ceiling performance, or all the patterns were equally visible, then we would lack any evidence that the model reflects what human subjects find difficult and what they find effortless.

There were clear differences in detectability of the patterns to human subjects (Figs. 6 and 7) and the patterns do indeed provide a spectrum of conspicuousness that is sufficient to draw conclusions from. The two different environments did not contain bright blue elements and the texture of the pattern was smooth and therefore UN PKB was, as predicted, very visible and the motivation for its inclusion as a control was vindicated. Olive Drab is also texturally smooth and its colouration is perceptually much closer to the environments used than UN PKB. The cost of pattern design is expensive and if simple olive drab were effective this would have implications for the design of camouflage; in fact this was not the case, with the patterned Flecktarn, Marpat and DPM performing better in most contexts. These patterns' visibilities could not be as easily predicted as UN PKB, because they have never previously been compared in the two environments. We should not over-interpret their relative effectiveness in our experiment, as the experiment was not designed with this goal. Multiple replicates of each pattern type, and habitat class, would be needed before we could conclude that, say, Marpat was better than MTP for these environments. Similarly, we cannot be sure that tendency of humans to outperform the model for Flecktarn, Marpat and DPM, but not MTP or the untextured patterns, is due to specifics of the textures or colours involved.

The PASGT helmet, the standard issue for the US Armed Forces from the 1980s to 2000s, was chosen as a typical item of camouflaged military equipment but unvarying in shape (unlike a soldier or combat uniform) and easily portable. It is difficult to predict how the model might perform with larger objects such as vehicles because these objects would have to be placed much further away from the camera and so the spatial scale of the background textures relative to the object would change. However, given the success of the model in this task and the multiresolution nature of log Gabor filters, there are grounds for thinking it has general applicability. The primary function of camouflage is to avoid detection in plain sight by enemies. But it is also the case that friendly personnel need to identify peers, and therefore there is a trade off in visibility and identification such that one needs, not to be easily visible (to avoid attack) and yet remain identifiable (to avoid friendly fire) (Talas et al. 2017). The framework elaborated here, where classification was evaluated in a paired manner, helmet versus background, can be easily extended for this problem as a multi-class classification task.

5. Conclusion

A human observer model has been designed, and its detection and recognition behavior was compared with human participants. Its behavior correlated highly with human participants. There is large applicability for such a human observer model, where it is impractical to use human participants. We have shown that an inexpensive and automated objective assessment of camouflage effectiveness is possible in a real-world setting.

References

Abramov, I., Gordon, J., and Chan, H. (1991). Color appearance in the peripheral retina: effects of stimulus size. Journal of the Optical Society of America, A, 8(2):404–414.

Bhajantri, N. U. and Nagabhushan, P. (2006). Camouflage defect identification: a novel approach. ICIT'06 9th International Conference on Information Technology, pages 145–148.

Birkemark, C. M. (1999). Cameva: a methodology for computerized evaluation of camouflage effectiveness and estimation of target detectability. International Society for Optics and Photonics. In AeroSense 1999, pages 229–238.

Cai, J. & Goshtasby, A. (1999). Detecting human faces in color images. Image and Vision Computing, 18(1):63–75.

Chandesa, T., Pridmore, T., and Bargiela, A. (2009). Detecting occlusion and camouflage during visual tracking. IEEE International Conference on Signal and Image Processing Applications (ICSIPA), pages 468–473.

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. Optical Society of America, 2(7):1160–1169.

De Valois, R. L., Albrecht, D. G., and Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. Vision Research, 22(5):545–559.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. Journal of the Optical Society of America, 4(12):2397–2394.

Ghahramani, Z. and Hinton, G. E. (1996). The EM algorithm for mixtures of factor analyzers. Technical report, University of Toronto.

Hartcup G, 2008. Camouflage: The History of Concealment and Deception in War. Barnsley, UK: Pen and Sword.

Hecker, R. (1992). Camaeleon – camouflage assessment by evaluation of local energy, Spatial frequency, and orientation. In Aerospace Sensing. International Society for Optics and Photonics, pages 343–349.

Hansen, T., Pracejus, L., and Gegenfurtner, K. R. (2009). Color perception in the intermediate periphery of the visual field. Journal of Vision, 9(4):26–26.

Heinrich, D. H. and Selj, G. K. (2015). The effect of contrast in camouflage patterns on detectability by human observers and camaeleon. In SPIE Defense and Security. International Society for Optics and Photonics, pages 947604–947604.

Hubel, D. H. (1995). Eye, Brain, and Vision. Scientific American Library/Scientific American Books.

Jones, J. P. and Palmer, L. A. (1987). An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. Journal of Neurophysiology, 58(6):1233–1258.

Kiltie, R. A., Fan, J., and Laine, A. F. (1995). A wavelet-based metric for visual texture discrimination with applications in evolutionary ecology. Mathematical Biosciences, 126(1):21–39.

Kovesi, P. (1999). Phase preserving denoising of images. Signal, 4(3):1.

Kovesi, P. D. (2000). MATLAB and Octave functions for computer vision and image processing. Centre for Exploration Targeting, School of Earth and Environment, The University of Western Australia.

MacLeod, D.I.A. and Boynton, R.M. (1979). A chromaticity diagram showing cone excitation by stimuli of equal luminance. Journal of the Optical Society of America, A, 69:1183-1186.

Melin, A. D., Fedigan, L. M., Hiramatsu, C., Sendall, C. L., and Kawamura, S. (2007). Effects of colour vision phenotype on insect capture by a free-ranging population of white-faced capuchins, *Cebus capucinus*. Animal Behaviour, 73(1):205–214.

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

Morgan, M., Adam, A., and Mollon, J. (1992). Dichromats detect colour-camouflaged objects that are not detected by trichromats. Proceedings of the Royal Society of London B248:291–295.

Mullen, K. T. (1985). The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. The Journal of Physiology, 359(1):381–400.

Pinto, N., Cox, D. D., and DiCarlo, J. J. (2008). Why is real-world visual object recognition hard? PLoS Computational Biology, 4(1):e27.

Renoult, J. P., Kelber, A., and Schaefer, H. M. (2015). Colour spaces in ecology and evolutionary biology. Biological Reviews 92: 292–315.

Shadeed, W., Abu-Al-Nadi, D. I., and Mismar, M. J. (2003). Road traffic sign detection in color images. ICECS 2003. Proceedings of the 2003 10th IEEE International Conference on Electronics, Circuits and Systems, 2003., 2:890–893.

Sengottuvelan, P., Wahi, A., & Shanmugam, A. (2008). Performance of decamouflaging through exploratory image analysis. First International Conference on Emerging Trends in Engineering and Technology, 2008. ICETET'08 (pp. 6-10).

Talas L, Baddeley R, Cuthill IC, (2017). Cultural evolution of military camouflage. Phil Trans R Soc B 372, 20160351.

Tatler, B. W., Baddeley, R. J., and Gilchrist, I. D. (2005). Visual correlates of fixation selection: effects of scale and time. Vision Research, 45(5):643–659.

Tkaclic, M. and Tasic, J. F. (2003). Colour spaces: perceptual, historical and application. EUROCON 2008. Computer as a Tool. The IEEE Region, 8 (1), 304–308.

Tipping, M. E. and Bishop, C. M. (1999a). Mixtures of probabilistic principal component analyzers. Neural Computation, 11(2):443–482.

Tipping, M. E. and Bishop, C. M. (1999b). Probabilistic principal component analysis. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 61(3):611–622.

Vakrou, C., Whitaker, D., McGraw, P. V., and McKeefry, D. (2005). Functional evidence for cone-specific connectivity in the human retina. The Journal of Physiology, 566(1):93–102.

Wyszecki, G. and Stiles, W. S. (1982). Color Science: Concepts and Methods, Quantitative Data and Formulae. 2nd edition. New York: Wiley.

Yfantis, E. A., Flatman, G. T., and Behar, J. V. (1987). Efficiency of kriging estimation for square, triangular, and hexagonal grids. Mathematical Geology, 19(3):183–205.