Paper accepted 15/03/18 in Computers in Industry

Camoflage assessment: Machine and human

Timothy N. Volonakis[a, c, *], Olivia E. Matthews[a], Eric Liggins[d], Roland J. Baddeley[a], Nicholas E. Scott-Samuel[a], Innes C. Cuthill[b]

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

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

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

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

[*] Corresponding author: tim.volonakis@gmail.com (T.N. Volonakis)

15

16

17

Camouflage Assessment: Machine and Human

Timothy N. Volonakis[1*], Olivia E. Matthews[1*], Eric Liggins[3], 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] QinetiQ Ltd, Cody Technology Park, Farnborough, Hampshire, GU14 0LX, UK

18 ## Abstract

19 A vision model is designed using low-level vision principles so that it can perform as a
20 surrogate human observer. In a camouflage assessment task, using military patterns in an
21 outdoor environment, human performance at recognition and detection is compared with
22 the surrogate human observer. This involved field data acquisition and subsequent image
23 calibration, a human experiment, and the design of the vision model. Human and machine
24 performance, at recognition and detection, of military patterns in two environments was
25 found to correlate highly.

26

Key Words: Camouflage Assessment, Observer Modelling, Visual Search

27

28

Acknowledgements

This research was funded by Qinetiq. We thank Qinetiq; Eric Liggins, William Serle and Ian Moorhead, for their support and input.

1. Introduction

Military personnel and equipment need protection from detection during conflict. Camouflage is the primary method to achieve this, through coloured textures that match the background and/or disrupt the object's outline (Hartcup 2008; 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, if the camouflage is being designed for use in hostile theatres of war, 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 accessing inaccessible environments, and sometimes the location of conflict, by using 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 wavebands from the human visible. A computational approach is therefore required to overcome the limitations of assessing camouflage using human observers. Such a computational model should be ideally designed, in the first instance, in accordance 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 to human performance. More generally, however, such a system could be adapted to have a different 'front end' (e.g. infra-red, hyperspectral). 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 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 needs 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, (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 Grettag 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 ensure 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.

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 and participants received university course credits for their participation. 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 replicates 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. 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.

2.3 The Surrogate 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) describes 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.

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. 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).

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; 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)+iI(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 are Gaussian through small metric distortions such as shift and object pose.

Phase angles will cycle from to 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.

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.

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 local variation of that the 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 $P(A)$ 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 the 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 above 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:

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

2.3.3 Modelling the Detection Task

The model is trained on a series of crops. 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 1 - 6, 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.

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

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

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

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

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

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

Equations 1-6 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 Lab colour space are consistent with the judgements of human observers and are appropriate for discrimination purposes (Renoult et al., 2015). The model is a surrogate human observer. 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 Lab colour space, the Macleod-Boynton chromaticity space is used. The Macleod-Boynton colour space is another opponency colour 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 were in the fovea up to 20° away. Therefore colour patterns in the periphery 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, 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 Lab, 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 log probabilities were plotted against human performance to visualise the correlation.

3. Results

Human data was 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 Woodbudy 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

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),

375 we present physiological evidence and a computational argument for the use of log Gabor
376 filters. Such applicability of a surrogate human observer is high, because using human
377 participants is impractical given a variety of viewpoints, environments and objects. This
378 paper also defined a task that would allow a direct comparison between the biologically
379 motivated visual observer and human participants. The analysis of the behaviour from both
380 observers provides the necessary evidence to assess whether the model is an adequate
381 surrogate for a human observer. The task was to estimate the accuracy with which
382 camouflaged objects, military helmets with different coverings, could be detected and
383 recognised. The selection of a single object class with different colour patterns, rather than
384 an array of different objects, avoided the problem of object choice and allowed visibility to
385 be easily controlled through only coloration and textural properties. The visibilities of the
386 objects were unknown prior to the experiment because, to our knowledge, they had never
387 been evaluated in the two environments nor directly compared. However, a priori, the UN
388 PKB helmet was expected to be easy to detect, the Olive Drab harder to detect and the
389 three (Leigh Woods) or four (Woodbury Common) patterned camouflages hardest to detect.
390 It was essential that the visibility of the patterns varied. If human recognition and detection
391 for all camouflaged objects was at ceiling performance, or all the patterns were equally
392 visible, then we would lack any evidence that the model reflects what human subjects find
393 difficult and what they find effortless. There were clear differences in detectability of the
394 patterns to human subjects (Figs. 6 and 7) and the patterns do indeed provide a spectrum of
395 conspicuousness that is sufficient to draw conclusions from. The two different environments
396 did not contain bright blue elements and the texture of the pattern was smooth and
397 therefore UN PKB was, as predicted, very visible and the motivation for its inclusion as a
398 control was vindicated. Olive Drab is also texturally smooth and its colouration is

399 perceptually much closer to the environments used than UN PKB. The cost of pattern design
400 is expensive and if a simple olive green drab were effective this would have implications for
401 the design of camouflage. The other chosen patterns' visibilities could not be as easily
402 predicted as UN PKB, because they have never previously been compared in the two
403 environments. The PAGST helmet, the standard issue for the US Armed Forces from the
404 1980s to 2000s), was chosen as a typical item of camouflaged military equipment but
405 unvarying in shape (unlike a soldier or combat uniform) and easily portable. It is difficult to
406 predict how the model might perform with larger objects such as vehicles because these
407 objects would have to be placed much further away from the camera and so the spatial
408 scale of the background textures relative to the object would change. However, given the
409 success of the model in this task and the multiresolution nature of log Gabor filters, there
410 are grounds for thinking it has general applicability. The primary function of camouflage is
411 to avoid detection in plain sight by enemies. But it is also the case that friendly personnel
412 need to identify peers, and therefore there is a trade-off in visibility and identification such
413 that one needs not to be easily visible (to avoid attack) and yet remain identifiable (to avoid
414 friendly fire) (Talas et al. 2017). The framework elaborated here, where classification was
415 evaluated in a paired manner, helmet versus background, can be easily extended for this
416 problem as a multi-class classification task.
417
418 The model is an automatic and inexpensive process of evaluating camouflage given an
419 environment. This utility of a surrogate human observer is in removing human participants
420 from the process. A prime example of an experiment, in the context of camouflage,
421 involving using human participants or even wild animals (blue jays), as predators searching
422 for artificial prey, simulating natural evolution (Bond and Kamil, 2002, 2006; Reynolds,

2011). Using human participants as predators is not a limitation to simulate evolution, because natural selection in the wild can be rapid (Endler, 1986). However, human participants can be removed from the process, providing an objective and less expensive means of testing different environments and prey. Automating this procedure with a comprehensive vision model has a large impact for the design of camouflage patterns.

5. Conclusion

A surrogate human observer has been designed, and its behavior was compared with human participants. Its behavior correlated highly with human participants. There is large applicability for such a surrogate human observer, where it is impractical to use human participants. We have shown in a military application, 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. JOSA 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.

Bond, A. B. and Kamil, A. C. (2006). Spatial heterogeneity, predator cognition, and the evolution of color polymorphism in virtual prey. Proceedings of the National Academy of Sciences of the United States of America, 103(9):3214–3219.

Cai, J. and 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.

Endler, J. A. (1986). Natural Selection in the Wild. Princeton: Princeton University Press.

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

482 periphery of the visual field. Journal of Vision, 9(4):26–26.
483
484 Heinrich, D. H. and Selj, G. K. (2015). The effect of contrast in camouflage patterns
485 on detectability by human observers and camaeleon. In SPIE Defense and Security.
486 International Society for Optics and Photonics, pages 947604–947604.
487
488 Hubel, D. H. (1995). Eye, Brain, and Vision. Scientific American Library/Scientific American
489 Books.
490
491 Jones, J. P. and Palmer, L. A. (1987). An evaluation of the two-dimensional gabor filter
492 model of simple receptive fields in cat striate cortex. Journal of Neurophysiology,
493 58(6):1233–1258.
494
495 Kiltie, R. A., Fan, J., and Laine, A. F. (1995). A wavelet-based metric for visual texture
496 discrimination with applications in evolutionary ecology. Mathematical Biosciences,
497 126(1):21–39.
498
499 Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation
500 and model selection. International Joint Conference on Artificial Intelligence (IJCAI),
501 14(2):1137–1145.
502
503 Kovesi, P. (1999). Phase preserving denoising of images. Signal, 4(3):1.
504
505 Melin, A. D., Fedigan, L. M., Hiramatsu, C., Sendall, C. L., and Kawamura, S. (2007).
506 Effects of colour vision phenotype on insect capture by a free-ranging population of
507 white-faced capuchins, *Cebus capucinus*. Animal Behaviour, 73(1):205–214.
508
509 Morgan, M., Adam, A., and Mollon, J. (1992). Dichromats detect colour-camouflaged
510 objects that are not detected by trichromats. Proceedings of the Royal Society of London
511 248:291–295.
512
513 Mullen, K. T. (1985). The contrast sensitivity of human colour vision to red-green and
514 blue-yellow chromatic gratings. The Journal of Physiology, 359(1):381–400.
515
516 Pinto, N., Cox, D. D., and DiCarlo, J. J. (2008). Why is real-world visual object recognition
517 hard? PLoS Computational Biology, 4(1):e27.
518
519 Renoult, J. P., Kelber, A., and Schaefer, H. M. (2015). Colour spaces in ecology and
520 evolutionary biology. Biological Reviews 92: 292–315.
521
522 Reynolds, C. (2011). Interactive evolution of camouflage. Artificial Life, 17(2):123–136.
523
524 Sengottuvelan, P., Wahi, A., and Shanmugam, A. (2008). Performance of decamouflaging
525 through exploratory image analysis. First International Conference on Emerging Trends
526 in Engineering and Technology IEEE, pages 6–10.
527
528 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.

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.

Figure legends

*Asterix * denotes Figure to be in colour*

***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. 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 cued in A and D are uninformative.

***Figure 3. 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 4. 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.

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.

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: Woodbury Common

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

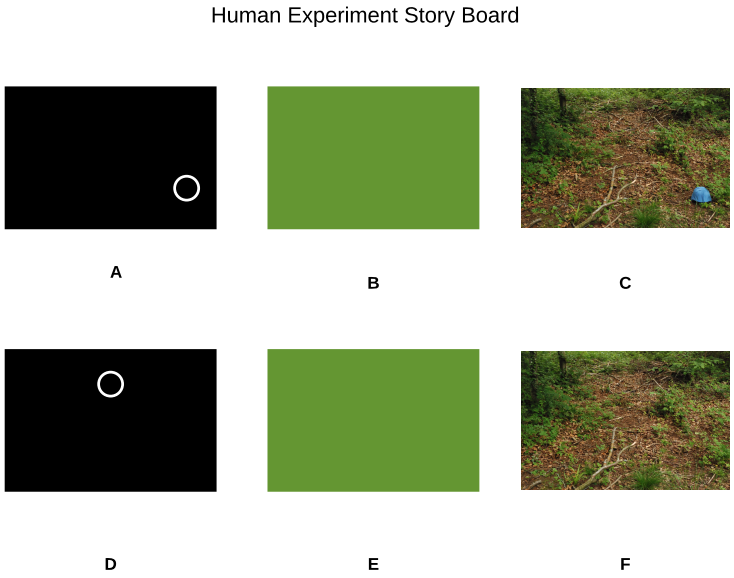
Figure 9. Human and model detection accuracy: Woodbury Common

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

607 Fig. 1

608
609

610 Fig. 2



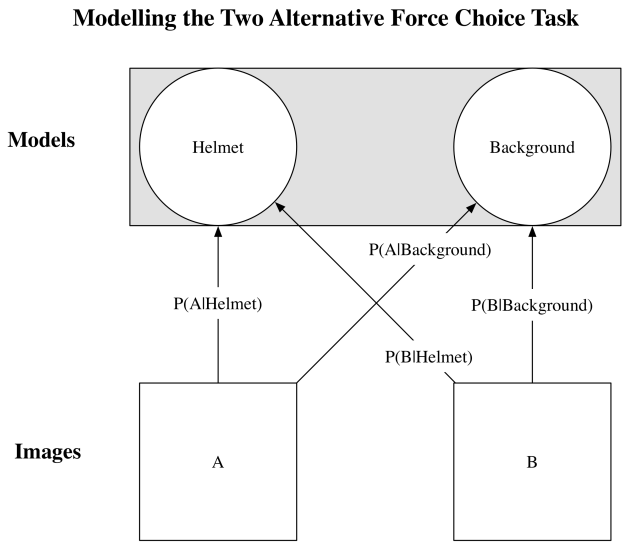
611
612

613 Fig. 3

614
615

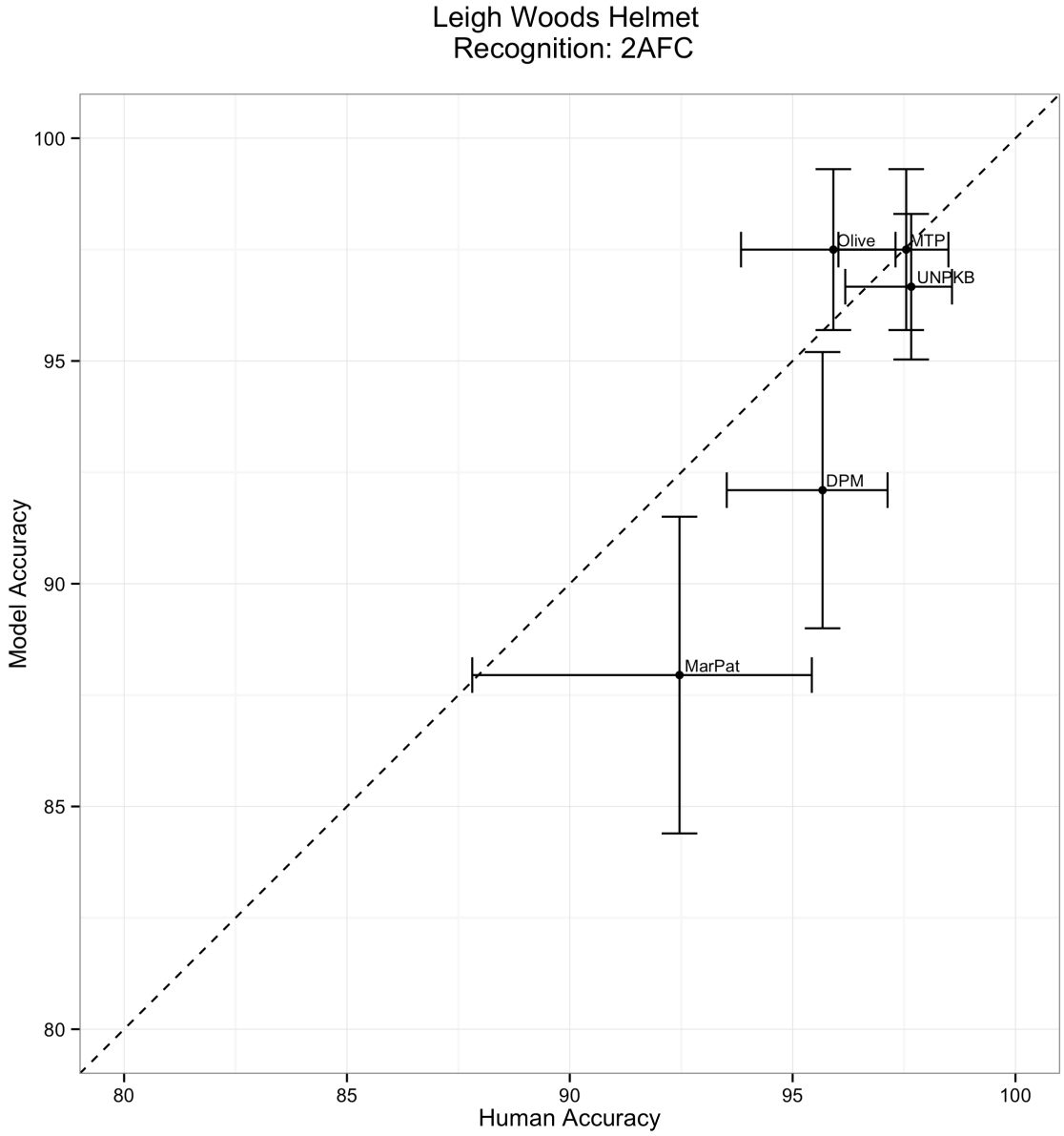
Fig. 4

619 Fig. 5



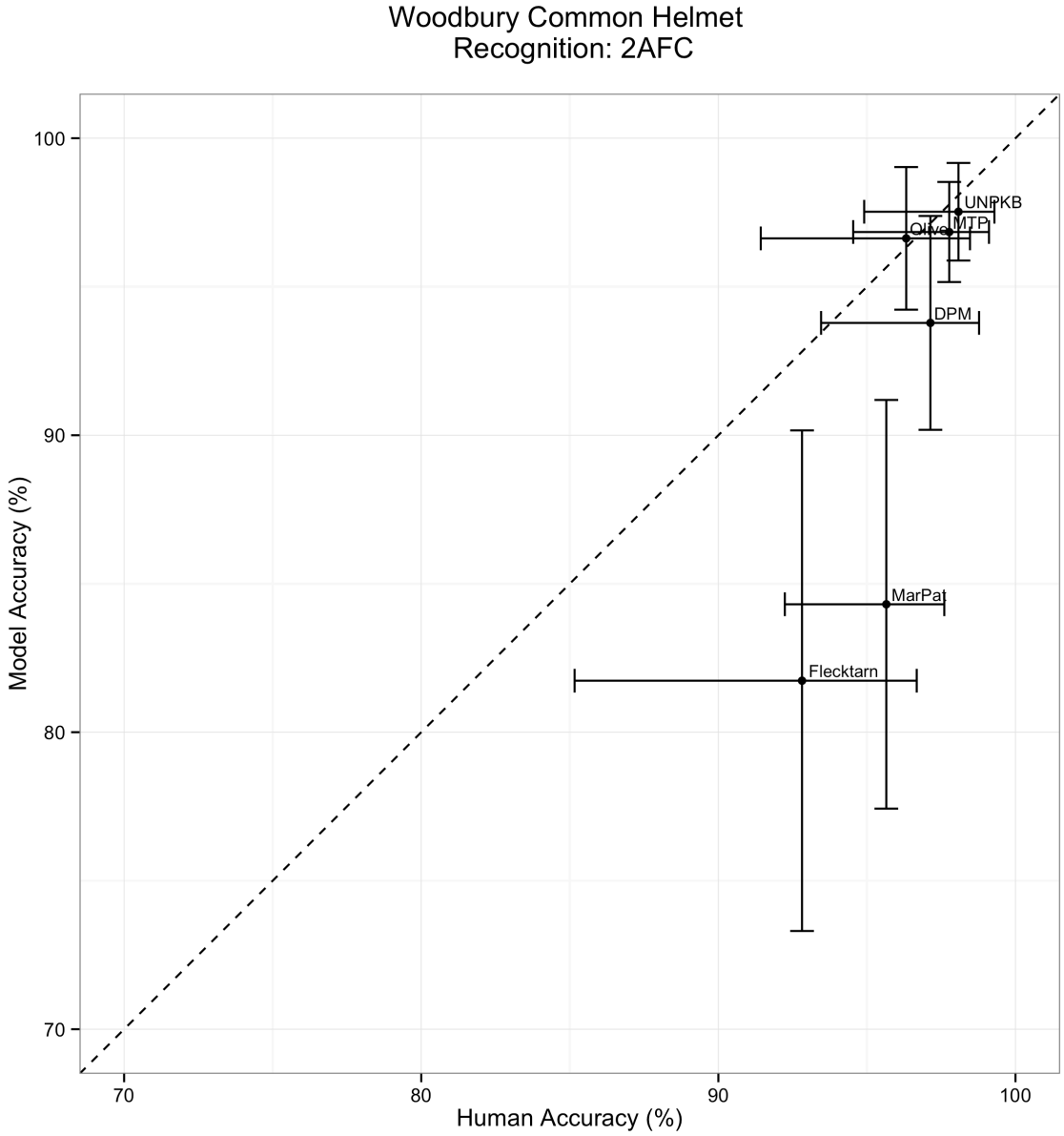
620
621

622 Fig. 6



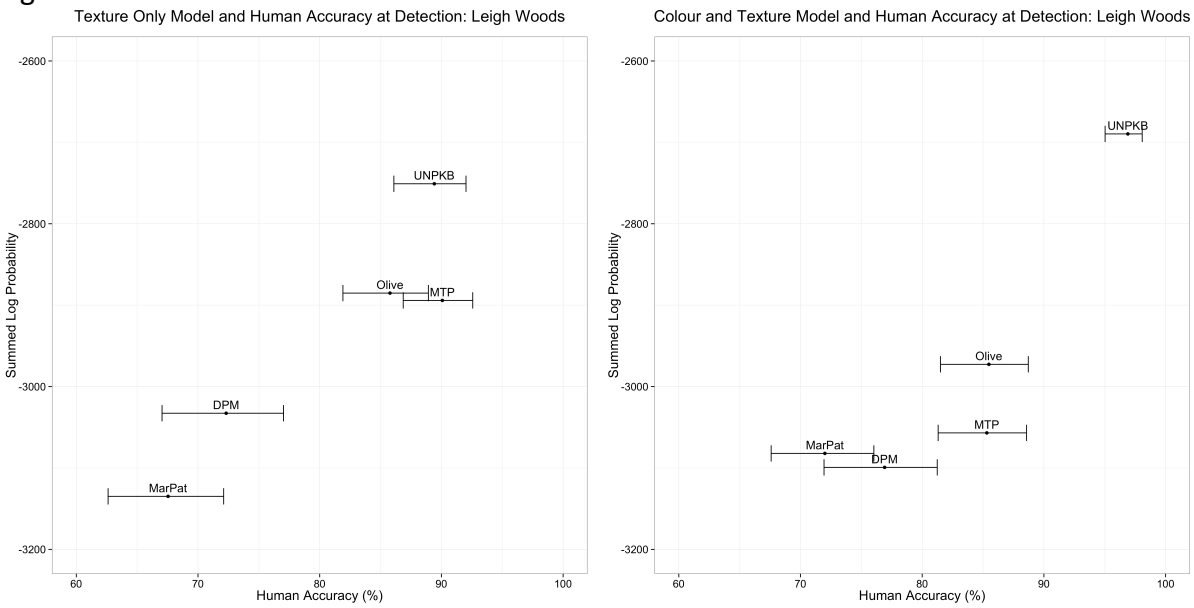
623
624

625 Fig. 7



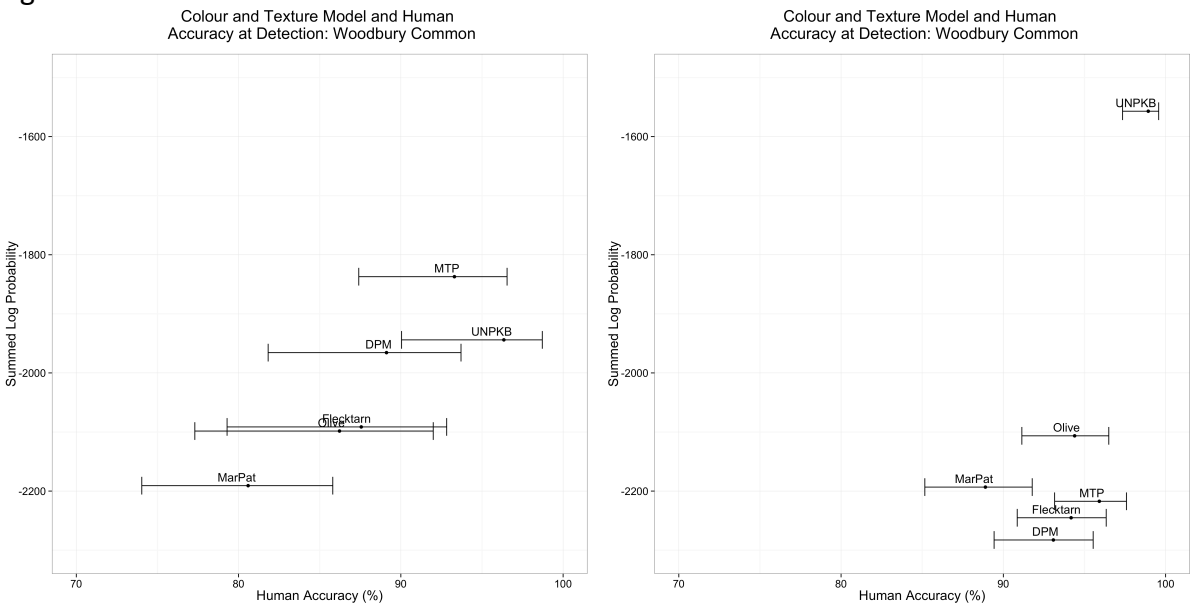
626
627

628 Fig. 8


Texture Only Model and Human Accuracy at Detection: Leigh Woods
Summed Log Probability
-2600
-2800
-3000
-3200
UNPKB
Olive
MTP
DPM
MarPat
60
70
80
90
100
Human Accuracy (%)
Colour and Texture Model and Human Accuracy at Detection: Leigh Woods
Summed Log Probability
-2600
-2800
-3000
-3200
UNPKB
Olive
MTP
MarPat
DPM
60
70
80
90
100
Human Accuracy (%)

629

630 Fig. 9


Colour and Texture Model and Human
Accuracy at Detection: Woodbury Common
Summed Log Probability
-1600
-1800
-2000
-2200
MTP
UNPKB
DPM
Flecktarn
Olive
MarPat
70
80
90
100
Human Accuracy (%)
Colour and Texture Model and Human
Accuracy at Detection: Woodbury Common
Summed Log Probability
-1600
-1800
-2000
-2200
UNPKB
Olive
MarPat
MTP
Flecktarn
DPM
70
80
90
100
Human Accuracy (%)

631