# Use of a vision model to quantify the significance of factors affecting target conspicuity.

M. A.Gilmore[a], C.K.Jones[a], A.Haynes[a], D.Tolhurst[b],T.Troscianko[c],G.Lovell[c], K.Pickavance[d].
[a]Defence Science and Technology Laboratory, Farnborough, Hants, GU14 OLX;
Cambridge
Bristol
Lockheed Martin UK INSYS Ltd,

## ABSTRACT

When designing camouflage it is important to understand how the human visual system processes the information to discriminate the target from the background scene. A vision model has been developed to compare two images and detect differences in local contrast in each spatial frequency channel. Observer experiments are being undertaken to validate this vision model so that the model can be used to quantify the relative significance of different factors affecting target conspicuity. Synthetic imagery can be used to design improved camouflage systems. The vision model is being used to compare different synthetic images to understand what features in the image are important to reproduce accurately and to identify the optimum way to render synthetic imagery for camouflage effectiveness assessment. This paper will describe the vision model and summarise the results obtained from the initial validation tests. The paper will also show how the model is being used to compare different synthetic images and discuss future work plans.

Keywords: Vision model, imagery fidelity, target conspicuity, camouflage

## 1. INTRODUCTION

In order to understand what makes an object detectable it is essential that the human visual system processes are understood. When there is an understanding of the relative significance of the different factors affecting target conspicuity it will be possible to design more effective camouflage systems. The approach adopted here has been to develop a computational vision model to evaluate the perceived differences between subtly different images. A model of visual difference prediction based on multi-scale analysis of local contrast has been developed and tested with psychophysical discrimination experiments on natural-scene stimuli. The model carries out a multiresolution analysis of the two pictures and detects differences in local contrast in each spatial frequency "channel". The model can account for differences in the chromatic domain by modelling differences in the luminance domain and in two opponent chromatic domains. A variety of psychophysical experiments, measuring thresholds for discriminating small changes in naturalistic images, have been undertaken to validate the model.

Computer graphics can be used to produce synthetic imagery for a variety of applications. For example the UK has developed the physically accurate scene generation system (CameoSim) to evaluate the effectiveness of different camouflage systems [1,2]. Synthetic images of different types of scene can be created to visualise the effect of different camouflage schemes under different conditions. In this way it is possible to design and evaluate new systems prior to production.

The vision model is being used to evaluate how similar synthetic images are to real world images to understand the significance of different aspects of synthetic image generation process, so that appropriate images can be generated for different applications. In addition, the model is being used to understand the different features that affect target detection, identification and recognition so that more effective camouflage systems can be designed.

This paper will describe the vision model and the initial validation tests undertaken. The paper will also describe how the model is being used to compare different synthetic images to understand the factors affecting target conspicuity.

## 2. VISION MODEL

### 2.1. Description of model

Computational vision models can attempt to predict perceived differences between images, but most of these models are derived only from theoretical considerations of human vision and are not based on psychophysical experimentation. This model of visual difference prediction is based on multi-scale analysis of local contrast in each spatial frequency 'channel' and has been extended to account for differences in the chromatic domain.

The model is based on knowledge of primary visual cortex and has much similarity with other models [3,4,5,6,7]. These models recognise that a visual image is processed in parallel (at least in the early stages of visual cortex processing) by channels or neurons with different optimal spatial frequencies but all with much the same bandwidth of about 1 octave [8,9,10,11,12].

The contrast at each point in an image, at each of several spatial frequency scales [13,14] is calculated. The contrast at the point (x,y) and in the frequency band F is defined as:

$$C_f(x,y) = \frac{a_f(x,y)}{l_f(x,y)}$$

where $a_F(x,y)$ is a bandpass filtered version of the original image convolved with a circularly-symmetric filter with frequency response given by:

$$A_F(f) = \exp\left[-\frac{(f-F)^2}{2\sigma^2}\right]$$

while $l_F(x,y)$ is the result of convolving the original image with a circularly-symmetric low pass operator with frequency response given by:

$$L_F(f) = \exp\left[-\frac{(f)^2}{2\sigma^2}\right]$$

f is spatial frequency and σ is the spread of the Gaussian frequency-response curves, and is chosen to be 0.3*F* so that the bandpass filters have a bandwidth of about 1 octave. Division of the bandpassed convolution by $l_F$ (the local mean luminance) is a model of the fact that the visual system encodes contrast rather than luminance per se; the mean luminance is calculated over an area proportional to the period of *F*.

To model how the visual system compares two images, the $C_F(x,y)$ for both images at all frequency scales is then calculated, and the contrasts in the two images, point by point within each frequency band are compared. The absolute value of the difference in contrast between the two pictures under comparison at each location and in each frequency band is calculated:

$$\Delta C_{F,j}(x,y) = \left| C_{F,j}(x,y) - C_{F,0}(x,y) \right|$$

where *j* is the picture number of the test stimulus and *j=0* represents the reference picture.

How each value of ΔC contributes towards the visibility of the difference between the pictures is estimated by evaluating each ΔC value against the "dipper function" for contrast discrimination for sinusoidal gratings [15,16,17]. Figure 1 shows such a dipper function. Each value of $\Delta C_F(x,y)$ is treated as if it is the contrast increment of a sinusoidal

grating of frequency $F$ to be compared against a reference or pedestal grating whose Michelson contrast is the average of the paired contrast values in the two pictures at that location and frequency band.

$$\overline{C}_{F,J}(x,y) = 0.5\left|C_{F,J}(x,y) + C_{F,0}(x,y)\right|$$

Is there a better picture?

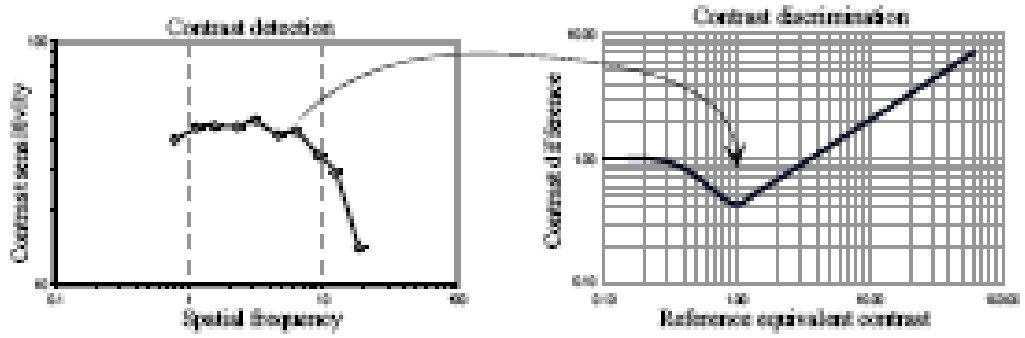


Figure 1. On the left is one observer's CSF – measures of the sensitivity for detecting the contrast of gratings. The sensitivity at a given spatial frequency determines the location of the contrast discrimination "dipper" on x and y axes.

The observer's contrast discrimination functions for achromatic gratings were estimated indirectly by adjusting the position on the x-axis (contrast reference) and y-axis (contrast difference) of a "dipper function" template for contrast discrimination according to the observer's contrast detection thresholds measured for a similar grating [18]. Thus, the model dipper functions were determined from each observer's contrast sensitivity functions (CSFs). Any differences between observer's abilities to discriminate between pictures should be accounted for by differences in their CSFs. Note that the linear, "Weber" part of the dipper function has a slope of only 0.7 on log/log axes rather than unity [15].

A measure (V) of how different two pictures might be at a single location and in a single frequency band is given by how far the calculated ΔC is above or below the dipper. There will be thousands of minute cues to discrimination, at the many locations and in the several frequency bands.

### 2.2. Pooling receptors and channels together.

The second stage in the model is to pool the many cues (V) provided at different locations and different frequency bands to give an overall assessment of whether or not the two pictures differ sufficiently for discrimination to be made. A weighted average of all the V cues, weighted across all locations and all frequency bands, is computed so that there is a single metric for a given pair of pictures rather than one measure per frequency band. A Minkowski sum with power of 4 is used [5]. The power of 4 derives from an empirical description of the amount of probability summation seen in grating detection experiments and relates to the steepness of the psychometric function [19,20]. It has been assumed that the same weighting will apply to contrast discrimination experiments. Thus, an overall cue V4 is given by:

$$V_4 = \sqrt[4]{\sum_F \sum_x \sum_y (V_4(x,y))^4}$$

$V_4$ is a single parameter that can be adjusted in order to optimize the fit of the model to many experimental threshold data.

### 2.3. Chromatic model

Human vision processes luminance (brightness) information and colour information separately and in parallel [21], and the colour information is processed in red-green and blue-yellow opponent channels [22]. Therefore the model has been developed with three planes: a luminance plane, and red-green and yellow-blue colour opponent planes.

The coloured images (in a conventional RGB format) are first transformed in order to calculate how the three cone types of human vision (L, M and S) would respond to the images. The luminance signal is then taken as the sum of L+M, whereas the red-green opponent signal [23,24] is taken as (L-M)/(L+M) which is similar to one direction in Macleod and Boynton's [1979] colour space. By analogy, the blue-yellow opponent signal is calculated as (S-0.5(L+M))/(S+0.5(L+M)).

The image discrimination model is run three times on each pair of coloured images to get the overall discrimination variable $V_4$ for the luminance plane of the images and for the red-green and yellow blue planes. Sensitivity for colour signals is biased towards low spatial frequencies compared to luminance signals. The criterion values of $V_4$ for luminance and the two colour channels are allowed to vary separately in the optimisation to fit an observer's experimental data. The observer's ability to discriminate two images is set by the highest value of the three $V_4$ estimates (after accounting for their different criterion values).

## 3. A PSYCHOPHYSICAL EXPERIMENT

### 3.1. Initial model validation

Models must be validated against real psychophysical experimental data to determine how well they explain human discrimination performance. Generally, such validation has been carried out against psychophysical experiments performed with sinusoidal gratings. The initial validation tests used a morphing technique because it produces a set of stimuli where each one of the component pictures is an image of a plausible object (with slightly different shape, color and texture); each morphed image still shares the natural Fourier statistics of the original ones[25].

Experiments have been undertaken in which human observers attempt to discriminate small changes in the shape, brightness, texture and colour of images of fruit. The observers' real thresholds have been compared with those predicted by the low-level model of visual cortex processing. The purpose of this experiment was to obtain a large set of image-discrimination data on which the model could be optimised. In order to achieve this, two sets of images were produced. The first set was of a red pepper morphing gradually into a yellow lemon, all on the same background of leaves with dappled illumination. The morph from one fruit to the other was conducted in 40 steps, so that there were 41 images in a sequence. The second image set was produced by morphing a blue (re-coloured) pepper into the yellow lemon. Figure 2 shows typical basic stimuli (only 9 of the 40 steps are shown). In an experiment, a computer-controlled procedure determined how much morphing (in %) was needed for an observer to discriminate the initial pepper image from a morphed image.

The morphed image set was subjected to various filtering operations so that, in all, 49 different stimulus sequences were obtained. "Two-alternative forced-choice" techniques determined, for each of the 49 conditions, how much a filtered stimulus needed to be morphed in order for reliable discrimination (75% correct) from the parent pepper image [26]. In a single trial there were three time intervals of 0.5seconds each. The observers were free to look at whichever part of the image they wished and they were free to make eye movements within the 0.5 second image presentations. The middle interval was always known by the observer to contain a parent image. The first or third interval (chosen randomly by the computer) would also contain that same image, but the third or first interval (respectively) would contain the morphed image. The observer's task was to inform the computer whether the first or third interval contained the different image. If the observer chose the wrong interval too frequently the task was made easier by choosing a morphed image more different than the parent; if the observer chose the correct interval too frequently, the task was made harder. Thus, during an experiment the 'staircase' converged on that % morph which the observer could correctly identify approximately 75% of the time. The red-green morph sequences were from lemon to pepper, but the blue-yellow sequences were from bluish-pepper to lemon.

Fig. 2. Examples of morphed images, a lemon (top-left) is gradually morphed into a pepper (bottom-right).

The thresholds for the 49 conditions are mostly similar, except that performance was worst (higher thresholds) in the corners of the plot surface where either the luminance or the colour opponent planes, or both, had been subject to extreme filtering. This was especially so for stimuli with negative filtering of colour and positive filtering of luminance i.e. observers were relatively poor at discriminating changes in the image when the colour information had been sharpened or edge enhanced while the luminance information was blurred. This is consistent with the finding that the human visual system favours low spatial-frequency colour information (i.e. not sharpened) and high spatial frequency luminance information (i.e. not blurred) [27].

The model was applied to the psychophysical data. In general, the model matched the shape of the 49 data-point surface representing the experimental results. The model predicts that human thresholds should rise in the corners of the surface, where either the luminance or the color planes or both are highly filtered. However, the model produced a more exaggerated version of the threshold surface – predicting that the human observer should be even worse at discriminations for which they are poor, but even better for those stimuli for which their performance was good.

### 3.2. Further model validation

The results from the first experiment were used to refine the observer trial procedure and improve the image dataset. Therefore additional observer tests have been undertaken to compare 450 image pairs of everyday images and objects such as shown in Figure 3. In these tests the observer was asked to fixate on a small spot in the centre of the screen. The results predicted by the model are plotted against the average of two observer ratings in Figure 4. This shows a high degree of correlation between the measured and predicted image difference values. Value????

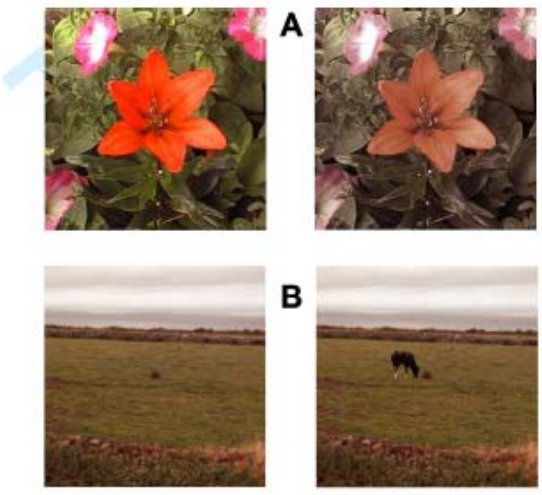


Figure 3. Examples of natural image pairs used for observer test to validate vision model.

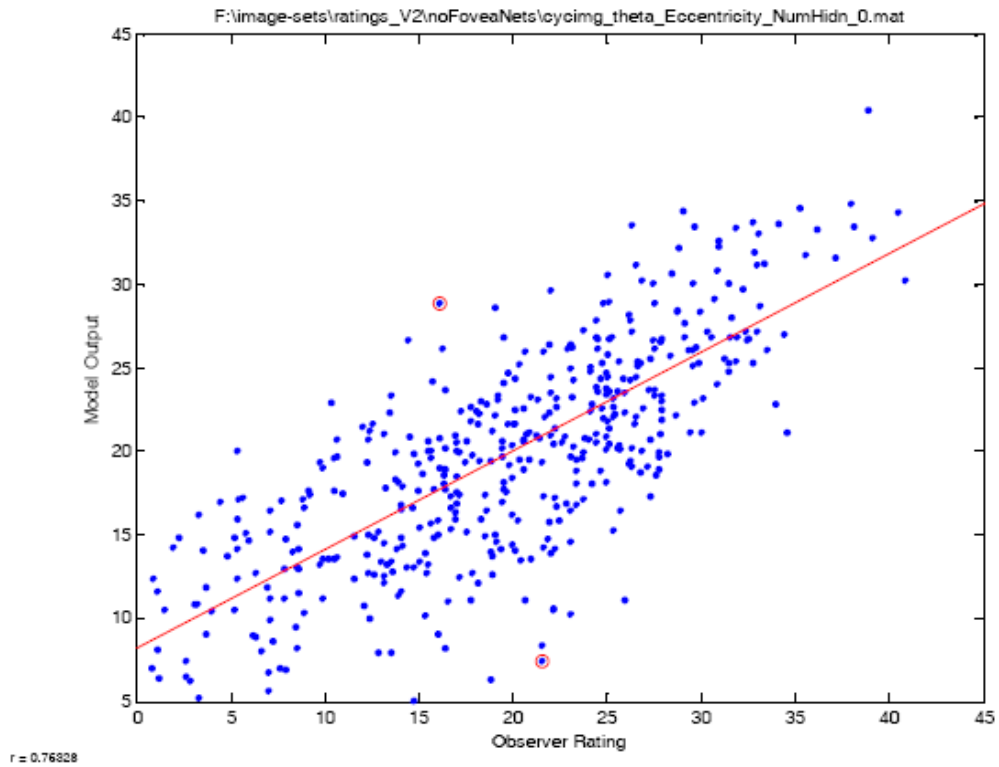


Figure 4. Observer trial results for comparing image pairs of everyday objects and scenes.

# 4. SYNTHETIC IMAGERY

## 4.1. CameoSim

CameoSim (the Camouflage Electro-Optical Simulator) has been developed over a number of years to produce a synthetic, high fidelity, physically accurate radiance map of 3D synthetic scenes for a wide range of operational scenarios, at any wavelength between 0.3 and 25 microns. Figure 5 shows a flow diagram of the processes involved in creating CameoSim imagery.

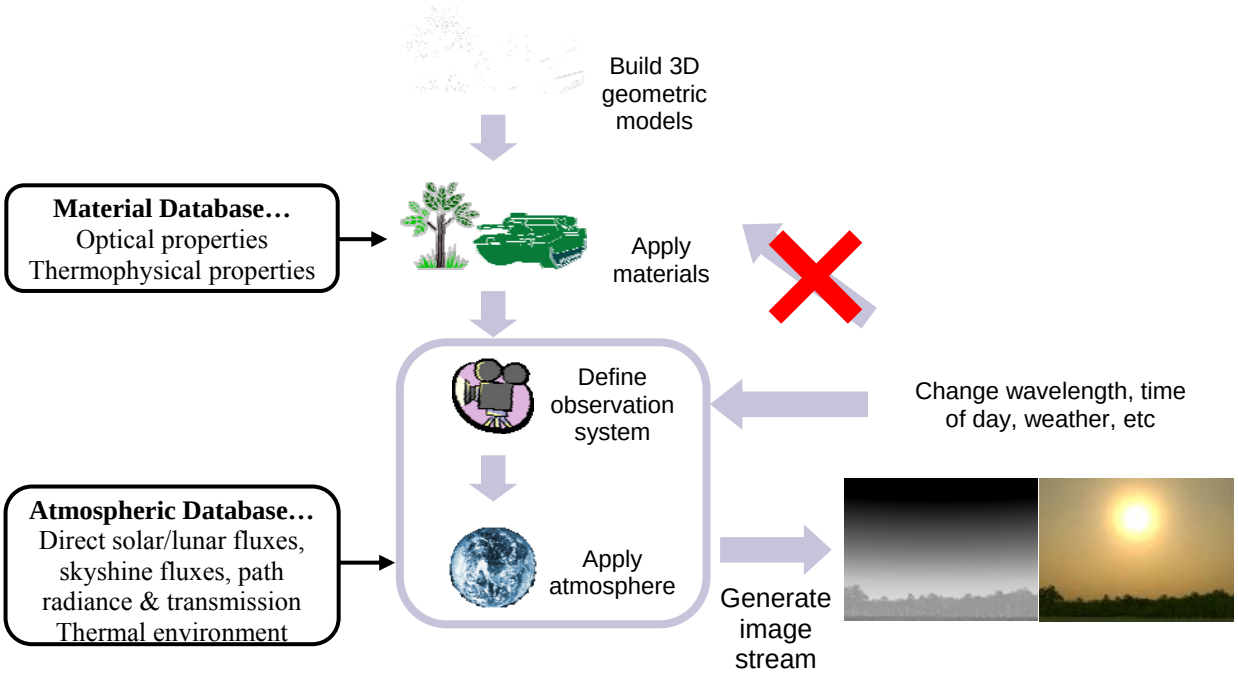


Figure 5. Flow diagram of the processes involved in producing CameoSim imagery

3D geometry can be imported into the CameoSim format using one of a number of available converters. These can then be manipulated to construct physically accurate representations of the object. There are three main building blocks: geometry, textures and physical materials. Materials based on physical properties can be constructed and applied via user-defined textures to the geometry. The user can define the various properties of an observation system and apply physics-based atmospherics. CameoSim is then able to render the scene and produce inband, hyperspectral and analysis imagery. If the image needs to be re-rendered for a different wavelength, weather, time of day, etc, there is no need to reconstruct the objects or materials. Rather, the user simply swaps the relevant input with another, a process that typically takes just seconds. This makes subsets of images with varying parameters easy to generate.

Within CameoSim there is a programmable Sensor degradation facility that can degrade images to account for the optics and detector effects typical of a sensor. If the user requires it, the non-degraded image can be exported and passed to third party degradation models for more intensive or specific degradation.

The renderer is flexible enough to give the user the freedom to alter many parameters in the image generation process. Validation and "fit-for-purpose" was at the forefront of its development in order that for a particular scenario the relevant parameters can be tweaked to obtain the best image for the particular purpose.

CameoSim was designed as a physically accurate system, and the number of computations to produce an entire image is huge. Therefore, there are approximations and accuracy settings (e.g. rays per pixel) available that can produce images of various degrees of quality. A user can use these to trade fidelity against the time it takes to generate.

Inevitably the images created using a fast rendering method together with low fidelity geometric models, textures and material properties will be of a lower fidelity than those rendered using the more accurate rendering algorithms and detailed geometric models, textures and material properties. The fidelity level and hence the accuracy required will depend on what the imagery is to be used for.

### 4.2. Image dataset

A set of images of a landrover was positioned under trees has been created so that the significance of different parameters can be evaluated. The initial image set was created to assess the effects of:

- Time of day (different shadows and lighting effects) – Figure 6
- Target position – Figure 7
- Colour of the vehicle – Figure 8
- Rendering options – Figure 9

07:00 10:00 13:00 16:00 19:00

Figure 6. Synthetic images representing different times of day, i.e. different shadow and lighting effects

Base Moved None

Figure 7. Synthetic images representing different target positions.

Base - multi-colour Black Olive Grey

Figure 8: Synthetic images representing different colour vehicle.

High fidelity Medium fidelity Low fidelity

Figure 9. Synthetic images created with different rendering methods.

## 5. APPLICATION OF VISION MODEL TO SYNTHETIC IMAGERY

Observer tests have compared each of the images against the others and the results have been compared with the differences predicted by the vision model. Initial analysis of the results show good agreement between the relative predicted and measured results for different colour landrover and different rendering fidelity, as shown in Figures 10 and 11. However, these initial results have shown that the model is overpredicting the effect of shadows and a change in position of the landrover as shown in Figures 12 and 13.

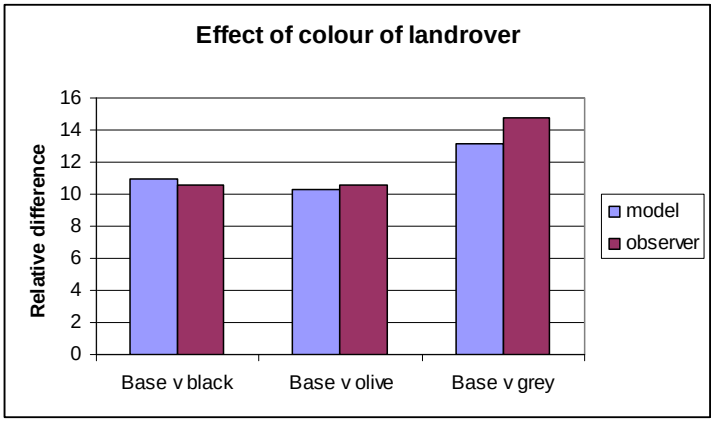


Figure 10. Predicted versus measured difference values for images showing different colour paint schemes on the landrover.

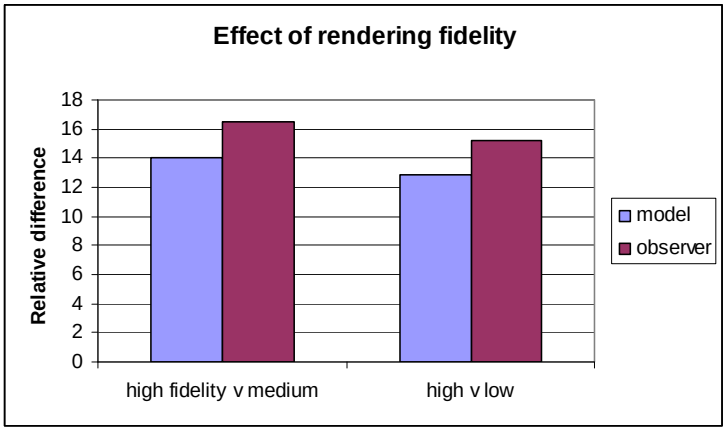


Figure 11. Predicted versus measured difference values for images rendered at different levels of fidelity.

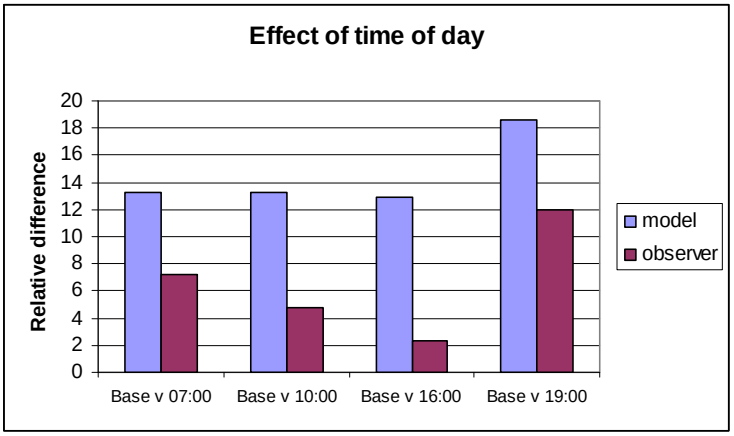


Figure 12. Predicted versus measured difference values for images representing different times of day.

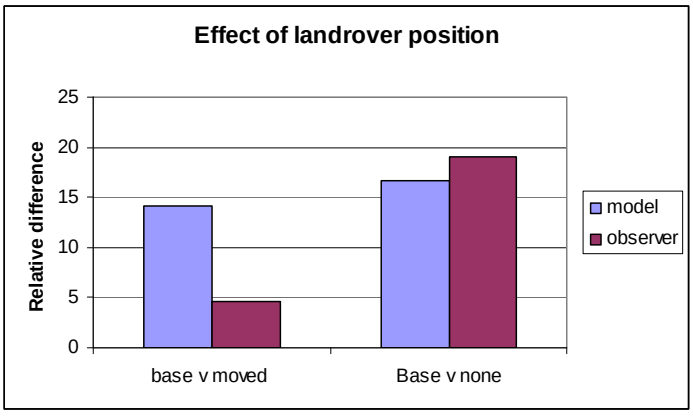


Figure 13. Predicted versus measured difference for images representing different landrover positions.

## 6. FUTURE PLANS

The results from analysis of the synthetic imagery are still being analysed as the understanding of the effects has a fundamental effect on the interpretation of the results. The results may be because the effects are viewed by peripheral vision and are not well seen, or they may be because the effects are capricious ephemeral things that say nothing about the context of the scene. Further tests will be undertaken to clarify these findings.

Therefore the vision model, that has been developed to quantify how much difference there has to be between two serially-presented visual images, is identifying important aspects to understand when designing camouflage systems and when creating synthetic imagery for military applications. The vision model will allow analysis of imagery to understand what features of a target have the greatest effect on detection, recognition and identification by human observers. Human psychophysical data for a natural-image discrimination task is being used to validate the model. The fit, although not perfect, is very promising and further validation tests will be undertaken work to test the model against detection performance in a greater variety of naturalistic tasks.

# Acknowledgements

The work reported in this paper has been jointly funded by UK MoD and EPSRC.

## REFERENCES

1. AW Haynes, M A Gilmore, D R Filbee, C Stroud. Accurate scene modelling using synthetic imagery. SPIE Conference 5075. *Targets and Backgrounds IX: Characterisation and Representation.* 85-96. 2003
2. D R Filbee, C A Stroud, G Hutchings, A Kirk, T Ward, D Brunnen. 'Modelling of high fidelity synthetic inagery for defence applications. *SPIE Conference 4718. 2003.* 2002
3. DALY, S. 1993. The visible differences predictor: an algorithm for the assesment of image fidelity. In Digital images and human vision. ed. WATSON AB, pp. 179-206. MIT Press, Cambridge, Mass.
4. DOLL, T.J., MCWORTER, S.W., WASILEWSKI, A.A. AND SCHMIEDER, D.E. 1998. Robust, sensor-independent target detection and recognition based on computational models of human vision. Optical Engineering 37, 2006-2021.
5. ROHALY, A.M., AHUMADA, A.J. AND WATSON, A.B. 1997. Object detection in natural backgrounds predicted by discrimination performance and models. Vision Research 37, 3225-3235.
6. WATSON, A.B. 1987. Efficiency of a Model Human Image Code. Journal of the Optical Society of America A 4, 2401-2417.
7. WATSON, A.B. 2000. Visual detection of spatial contrast patterns: Evaluation of five simple models. Optical Express 6, 12-33.
8. BLAKEMORE, C. AND CAMPBELL, F.W. 1969. On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images. Journal of Physiology-London 203, 237-260.
9. DE VALOIS, R.L., ALBRECHT, D.G. AND THORELL, L.G. 1982. Spatial-frequency selectivity of cells in macaque visual cortex. Vision Research 22, 545-559.
10. MOVSHON, J.A., THOMPSON, I.D. AND TOLHURST, D.J. 1978. Spatial and temporal contrast sensitivity of neurons in areas 17 and 18 of the cat's visual cortex. Journal of Physiology 283, 101-120.
11. TOLHURST, D.J. AND THOMPSON, I.D. 1981. On the variety of spatial-frequency selectivities shown by neurons in area-17 of the cat. Proceedings of the Royal Society of London Series B-Biological Sciences 213, 183-199.
12. WATSON, A.B. AND ROBSON, J.G. 1981. Discrimination at threshold: labelled detectors in human vision. Vision Research 21, 1115-1122.
13. PELI, E. 1990. Contrast in Complex Images. Journal of the Optical Society of America A 7, 2032-2040.
14. TADMOR, Y. AND TOLHURST, D.J. 1994. Discrimination of changes in the second-order statistics of natural and synthetic-images. Vision Research 34, 541-554.
15. LEGGE, G.E. 1981. A Power Law For Contrast Discrimination. Vision Research 21, 457-467.

16. LEGGE, G.E. AND FOLEY, J.M. 1980. Contrast masking in human vision. Journal of the Optical Society of America 70, 1456-1471.
17. NACHMIAS, J. AND SANSBURY, R.V. 1974. Grating contrast discrimination may be better than detection. Vision Research 14, 1039-1042.
18. PÁRRAGA, C.A. AND TOLHURST, D.J. 2000. The effect of contrast randomisation on the discrimination of changes in the slopes of the amplitude spectra of natural scenes. Perception 29, 1101-1116.
19. QUICK,R.F. 1974 A vector magnitude model of contrast detection. *Kybernetik,* 16.65-67
20. ROBSON,J.G. and GRAHAM,N. 1981 Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research,* 21. 409-418
21. MULLEN, K.T. AND LOSADA, M.A. 1994. Evidence for separate pathways for color and luminance detection mechanisms. Journal of the Optical Society of America A, 11, 3136-3151.
22. HURVICH, L.M. AND JAMESON, D. 1957. An opponent-process theory of colour vision. Psychological review 64, 384-404.
23. PÁRRAGA, C.A., BRELSTAFF,G., TROSCIANKO, T and MOORHEAD,I.R. 1998 Color and luminance information in natural scenes. *Journal of the Optical Society of America A*, 15. 563-569
24. OLMOS, A. AND KINGDOM, F.A.A. 2004. A biologically inspired algorithm for the recovery of shading and reflectance images. Perception 33, 1463-1473.
25. PÁRRAGA, C.A. TROSCIANKO, T AND TOLHURST, D.J. 2000. The human visual system is optimised for processing the spatial information in natural visual images. Current Biology 10, 35-38.
26. PÁRRAGA, C.A., TROSCIANKO, T. AND TOLHURST, D.J. 2005. The effects of amplitude-spectrum statistics on foveal and peripheral discrimination of changes in natural images, and a multi-resolution model. Vision Research submitted.
27. MULLEN, K.T. 1985. The contrast sensitivity of human color vision to red-green and blue-yellow chromatic gratings. Journal of Physiology-London 359, 381-400.