QUANTIFICATION OF VISUAL CLUTTER USING A COMPUTATIONAL MODEL OF HUMAN PERCEPTION: AN APPLICATION FOR HEAD-UP DISPLAYS

Michael Zuschlag
Volpe National Transportation Systems Center
Cambridge, MA, USA

ABSTRACT

A means of quantifying the cluttering effects of symbols is needed to evaluate the impact of displaying an increasing volume of information on aviation displays such as head-up displays. Human visual perception has been successfully modeled by algorithms that process an image through a bank of visual filters for a range of spatial frequencies and orientations. The model proposed here derives a vector of "feature density" values from these filtered images where each value represents the degree to which the image contains a particular spatial frequency and orientation. Differences in these feature densities between a target and a context is used to calculate the degree the target is salient relative to the context.

Keywords: Head-up displays; Clutter; Image analysis; Visual perception models

INTRODUCTION

Advanced technology is bringing an increasing volume of information to the flight deck that must be displayed in the relatively limited area of the flight deck displays. Display symbols must be designed such that the symbols are salient, and thus easy to read, but not so dominant that they create clutter by visually interfering with other significant objects. The compromises a designer must make between salience and clutter can be seen in head-up displays (HUDs), which are quite sensitive to the cluttering effects of symbology (e.g., Ververs and Wickens, 1998). While qualitative design guidelines emphasize minimizing HUD clutter (Newman, 1995), new technologies such as enhanced vision systems imply HUDs will required to display even more information. Designers and other display evaluators would be greatly aided if a means of quantifying the level of clutter were available so that the salience of symbols can be more optimally balanced.

A MODEL TO CALCULATE VISUAL SALIENCE

Salience as Average Color Difference

As a first approximation, assume a monochrome display, as is the case with current aviation HUDs. The degree a monochrome target, *o* (i.e., a HUD symbol), is salient with respect to a context, *i*, is related to the color contrast between the color of the target and the color of the context, where color includes both luminance and chromatic components. The perceptual difference in any two colors can be represented by their Euclidean distance in L*u*v* space (Wyszeski and Stiles, 1982). It is assumed that perceptual salience has an inverse exponential relationship to perceptual color difference. Thus, the salience of target *o* in the context of *i* should be related to average salience of the color differences of each point of *i*:

$$S_{oi}(0) = \frac{1}{A_i}\left[\iint 1 - \exp\left(-\beta\, p_{\Delta,xy}\right) dx\, dy \right],$$

where A_i is the area of the context for the target, $p_{\Delta,xy}$ is the L*u*v* distance between the target color and the color of a point at coordinates x and y in the context, and β is a constant to be empirically determined.

This inverse exponential relationship implies that after a certain level of color contrast, additional contrast has little effect on human performance, which is consistent with experimental research on HUDs (Weintraub and Ensing, 1992).

An application of this formula is illustrated in Figure 1, where a HUD symbol, a Bray-style flight path marker (FPM) (Weintraub and Ensing, 1992), is compared to uniform backgrounds of 0%, 75%, and 87% gray. In this example, the Red-Green-Blue (RGB) color values of the background images were assumed to be of the sRGB color space (International Electrotechnical Commission, 1999) in order to convert them to L*u*v* difference distances. The resulting $S_{oi}(0)$'s are shown, where higher value represents greater salience. The parameter β was rather arbitrarily set to 0.05. In practice, this value would be determined by fitting the model to human performance.

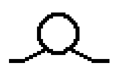
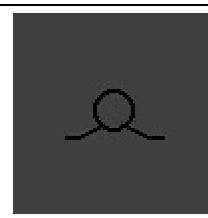
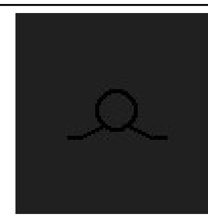
	(a)	(b)	(c)
Target and Context Note: For printing purposes, a black rather than bright green FPM is used.			
Context gray scale	0%	75%	87%
$S_{oi}(0)$	0.993	0.729	0.372

Figure 1. Calculated salience that compares average background color to target color.

The calculated salience agrees with intuition; the value of $S_{oi}(0)$ decreases as the contrast between the target and its context decreases. However, $S_{oi}(0)$ itself does not take into account the cluttering effects of any visual features that may reside within the context. In a HUD, these features may represent variations in the background texture (e.g., features of cloud or terrain), objects within the out the window (OTW) scene (e.g., traffic and runways), other nearby HUD symbols, and possibly overlaying textures from an enhanced vision or synthetic vision system. Consider Figure 2. The value of $S_{oi}(0)$ for (a) is about the same as (b). However, one would probably expect the target in (b) to be more difficult to see. Thus, in addition to $S_{oi}(0)$, one needs to account for the degree the context has features similar in shape and color to the target.

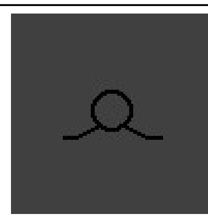
	(a)	(b)
Target and context		
$S_{oi}(0)$	0.729	0.752

Figure 2. Failure of average color difference in accounting for cluttering effects of texture.

Salience as Differences in Features

In artificial intelligence research, certain successful models of computational visual feature detection are based on results from low-level human and primate visual perception studies (Doll, McWhorter, Wasilewski, and Schmieder, 1998; Wilson, 1991). These models analyze an image for a range of spatial frequencies and orientations (Bergen and Landy, 1989). The greater a target differs from its context in the amplitudes of the spatial frequencies across the orientations, the more salient the target (Itti, Koch, and Niebur, 1998).

Specifically, let **I** be a two-dimensional array representing the perceptual salience of each pixel in an image compared to the target's color (i.e., $1 - \exp(-\beta\, p_{\Delta,xy})$), where the image may be the context or the target itself. Given a monochrome and transparent HUD, the array element values for any HUD symbol, including the target, are all 0 except for the background, which is set to 1, so the array represents the HUD symbol against a high contrast background.

The features of an image with respect to the target's color are then quantified as illustrated in Figure 3.

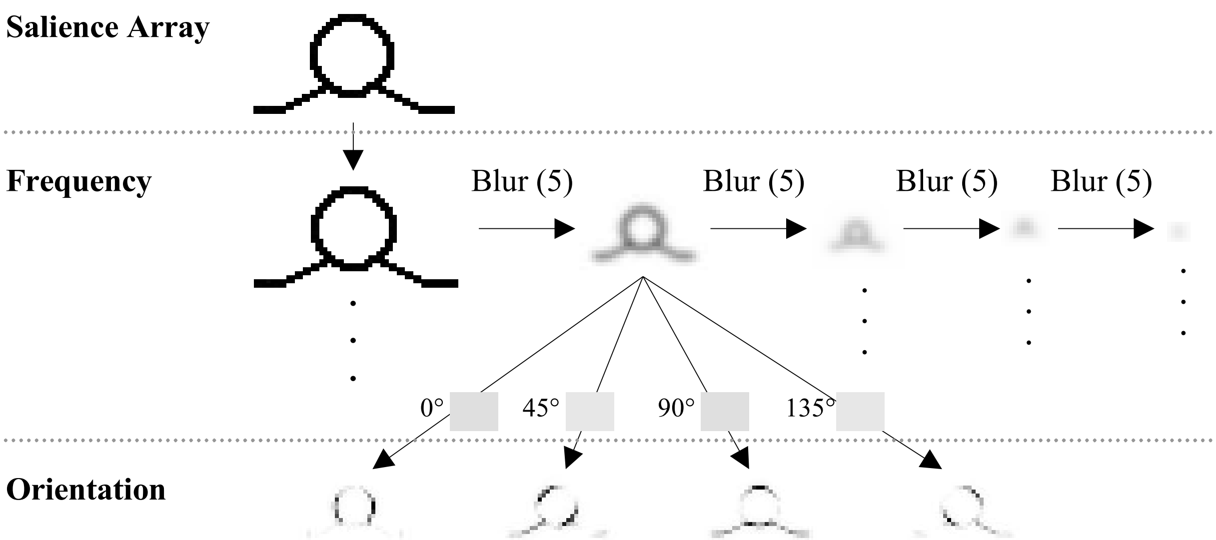


Figure 3. Algorithm for feature detection.

First, a range of frequencies for the spatial filtering is accomplished by building a "pyramid" of images of successively lower frequency v, where each successive image $\mathbf{I}_{v/2}$ is half the width and height of its predecessor $\mathbf{I}_v$. This done by first blurring a copy of the predecessor image as follows:

$$\mathbf{I}_{blurred} = \mathbf{I}_v * b * b^{\mathrm{T}}, \quad b = \{\, 0.05, 0.25, 0.40, 0.25, 0.05 \,\}.$$

Then shrinking it to half its dimensions by summing the value of each set of four adjacent pixels:

$$p_{v/2,\, x/2,\, y/2} = (\, p_{blurred,xy} + p_{blurred,x+1,y} + p_{blurred,xy+1} + p_{blurred,x+1,y+1} \,),$$

where p_{xy} is a pixel at position (x,y) in **I**.

This is done four times resulting in five octaves of frequency filtering, spanning the detectors for spatial frequencies found in the visual cortex (Wilson and Gelb, 1984).

Then, four spatially filtered arrays are generated for each frequency by convolving the array first by a five-element Gaussian vector then an orthogonal three-element approximately Gaborian vector. This is done for vectors angled at 0, 45, 90, and 135 degrees, which again roughly corresponds to detectors in the cortex. An absolute value is taken of the resulting element values. For example, the 0 degree filtering of an image corresponding to spatial frequency v is:

$$\mathbf{I}_{v,0} = |\, \mathbf{I}_v * b * g^{\mathrm{T}} \,|, \quad g = \{\, -0.5, 1.0, -0.5 \,\}.$$

While the 90 degree filtering is:

$$\mathbf{I}_{v,90} = |\, \mathbf{I}_v * b^{\mathrm{T}} * g \,|.$$

Thus, for each input image $\mathbf{I}$, the image analysis yields 20 output arrays, $\mathbf{I}_{v\theta}$ (5 frequencies · 4 orientations). In a sense, high values of the elements of $\mathbf{I}_{v\theta}$ represent an edge at orientation θ where image color changes with respect to the target color. For the same L*u*v* distances, changes towards the target color are weighted more than changes away owing to the transforming the L*u*v* distances by $1 - \exp(-\beta\, p_{\Delta,xy})$. Uniform images have no edges, so all elements of such a $\mathbf{I}_{v\theta}$ are 0.

Let the feature density of an image $f_{i,v\theta}$, represent the degree that image i has features per unit area of spatial frequency v and orientation θ that are similar in color to the target color. This is calculated by summing all array elements, $p_{v\theta,xy}$, of $\mathbf{I}_{v\theta}$ and dividing by the area of the image, A_i:

$$f_{i,v\theta} = \frac{1}{A_i} \sum_x \sum_y p_{v\theta,xy}\,.$$

In this manner, the 20 feature density values are calculated for both the target and its context. Let $S_{oi}(v,\theta)$ be the salience of target o in context image i with respect to features of v and θ. For a target that overlays and combines with the context, much as a HUD symbol would combine with the OTW view or enhanced vision system imagery, target salience is considered to be proportional to the degree the target adds features to the context:

$$S_{oi}(v,\theta) = w_{v\theta}\, \frac{f_{o,v\theta}}{(f_{o,v\theta} + f_{i,v\theta})}\,,$$

where $w_{v\theta}$ is an empirically derived weight representing the significance of the corresponding feature in perceptions of salience. Note that when o is compared to a featureless uniform context, i, each $S_{oi}(v,\theta)$ simply equals $w_{v\theta}$.

For a target that is presented proximal to the context, such as a HUD symbol with respect to other HUD symbols, target salience is considered to be proportional to the degree the target has different features from the context, weighted by a function of the target and context spatial separation $d(i,o)$:

$$S_{oi}(v,\theta) = d(i,o)\, w_{v\theta}\, \frac{|\, f_{o,v\theta} - f_{i,v\theta} \,|}{(f_{o,v\theta} + f_{i,v\theta})}\,.$$

Overall, the salience of a target o with respect to the context i is the combined effects of $S_{oi}(0)$ and all $S_{oi}(v,\theta)$. That is, the salience of the features must be weighted by the background salience, compensating, in a sense, for the salience of target's background pixels being set to 1.0. Thus, total salience, S_{oi}, is:

$$S_{oi} = S_{oi}(0) \sum_v \sum_\theta S_{oi}(v,\theta)\,.$$

MODEL PERFORMANCE

As a demonstration of this model, consider Figure 4. With the FPM symbol acting as the target and three backgrounds of varying clutter each acting as contexts, $S_{oi}(0)$ and $\Sigma\Sigma S_{oi}(v,\theta)$ are calculated with respect to the target's color. Relatively arbitrary parameters are used: all $w_{v\theta} = 0.05$, $\beta = 0.05$.

As can be seen in Figure 4 moving from (a) to (c), $S_{oi}(0)$ decreases as more cluttering features are added to the context, as the average color becomes darker and thus more like the color of the target (i.e., black). Note also how the $\Sigma\Sigma S_{oi}(v,\theta)$ sharply decreases with additional features, with the total calculated salience of Figure 4(c) being 0.274. Contrast that now to Figure 2(a), for which $S_{oi}(0) = 0.729$, and $\Sigma\Sigma S_{oi}(v,\theta) = 1$ (all 20 $S_{oi}(v,\theta) = w_{v,\theta} = 0.05$), resulting in a substantially higher calculated overall salience of 0.729. Indeed, using these arbitrary parameters, Figure 4(c) is rated as less salient than even Figure 1(c) (total salience = 0.372), which more or less corresponds to intuition.

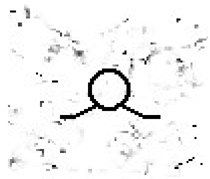
	(a)	(b)	(c)
Target and context			
$S_{oi}(0)$	0.988	0.920	0.752
$\Sigma\ \Sigma\ S_{oi}(\nu,\theta)$	0.926	0.591	0.364
Total Salience	0.916	0.544	0.274

Figure 4. Salience calculated from differences in average color and features.

As another illustration, consider Figure 5. Here, average grayscale and $S_{oi}(0)$ for the contexts are relatively constant and only the features of each context are varied. The context for Figure 5(a), dominated by high vertical frequencies, has few features in common with the FPM, so the model rates the FPM to be more salient there. In contrast, the FPM has strong diagonal features of high to low frequencies, and thus the salience is rated lower in Figure 5(b). This is fairly consistent with intuition; a better correspondence to human experience can be expected with more systematic parameter fitting.

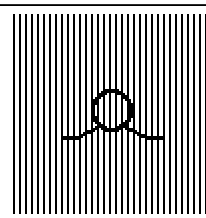
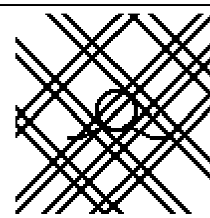
	(a)	(b)
Target and context		
Average grayscale	67%	67%
$S_{oi}(0)$	0.665	0.694
$\Sigma\ \Sigma\ S_{oi}(\nu,\theta)$	0.568	0.343
Total Salience	0.378	0.238

Figure 5. Effects of different features on calculated salience, holding average color difference constant.

CONCLUSION

The model presented here performs in accordance with one's intuition in accounting for the degree clutter interferes with symbol salience, suggesting that this approach is promising. The final verdict will depend on experimental validation in which the model's predictions will be compared to human performance. Before it can be used to evaluate actual HUDs, however, a number of details need to be addressed. Firstly, the spatial separation function $d(i,o)$ must be specified. Secondly, most objects viewed in and/or through a HUD vary in shape and color, thus ultimately a sample of representative of images for each object is necessary. Thirdly, for the sake of fast processing, it is preferred if one can evaluate the feature densities, $f_{i,v\theta}$, of each *component* of the context then somehow calculate their joint effect on each target; this calculation may be effectively approximated by a simple sum of the individual effects. Fourthly, in actual HUDs, the true color of HUD symbology is affected by the color of the background. Furthermore, HUDs are designed to vary in brightness to maintain a constant contrast ratio. The significance of these characteristics needs to be addressed. These characteristics may simplify the implementation of the model: a constant contrast ratio implies the luminance contribution to $S_{oi}(0)$ is constant so that one only needs to estimate the chromatic differences between the OTW view and the HUD symbols.

If successful, this approach can ultimately be generalized to other aviation displays such as navigation displays. Application to more traditional aviation displays may be on the one hand simpler, as most aviation displays have a uniform background (typically black). On the other hand, most aviation displays are not monochrome, implying a need to evaluate the salience of each object with respect to multiple target colors.

ACKNOWLEDGEMENTS

This study was supported by FAA Office of the Chief Scientific and Technical Advisor for Human Factors, AAR-100, working in cooperation with the FAA Transport Airplane Directorate, AMN-111. The author would like to express appreciation to the following people who made this document possible: Dale Dunford of the FAA, Chuck Oman, Miwa Hayashi, Pawan Sinha, and James DiCarlo of the MIT, and Dan Hannon of the Volpe Center.

REFERENCES

Bergen, J. R. and Landy, M. S. (1989). Computational modeling of visual texture segregation. In M. S. Landy, and J. A. Movshon (Eds.), *Computational models of visual processing* (pp253-271). Cambridge, MA: MIT Press.

Doll, T. J., McWhorter, 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.

International Electrotechnical Commission (1999). *Multimedia systems and equipment - Colour measurement and management - Part 2-1: Colour management - Default RGB colour space - sRGB* (IEC 61966-2-1). Geneva: International Electrotechnical Commission.

Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*, pp1254-1259.

Newman, R. L. (1995). *Head-up displays: Designing the way ahead*. Hants, England: Ashgate Publishing.

Ververs, P. M. and Wickens, C. D. (1998). Head-up displays: Effects of clutter, display intensity, and display location on pilot performance. *The International Journal of Aviation Psychology, 8,* 377-403.

Weintraub, D. J. and Ensing, M. (1992). *Human factors issues in head-up display design: The book of HUD.* CSERIAC State-of-the-art report SOAR 92-2.

Wilson, H. R. (1991). Psychophysical models of spatial vision and hyperacuity. In D. M. Regan (Ed.) *Spatial Vision* (pp64-86). New York: MacMillan Publishers.

Wilson, H. R. and Gelb, D. J. (1984). Modified line-element theory for spatial-frequency and width discrimination. *Journal of the Optical Society of America, 1,* 124-131.

Wyszecki, G., and Stiles, W. S. (1982). *Color science: Concepts and methods, Quantitative data and formulae,* (2nd ed.). New York: John Wiley & Sons.