Effect of Low SNR on Visual Image Quality

David R. Gerwe[1*], Carlos E. Luna[1], Brandoch Calef[2]

[1]*Boeing Phantomworks Adv. Network and Space Systems, 2260 East Imperial Highwa., El Segundo CA, 90245 USA*
[2]*Boeing Laser Technical Services, 535 Lipoa Pkwy., Kihei, HI, 96753 USA*
[*]*David.R.Gerwe@Boeing.com*

Abstract: The effect of noise on image quality is investigated using human evaluations of a set of images with a wide range of noise and blur levels. Quality is quantified in terms of the smallest contrast feature that can be reliably identified in a test pattern which can be equated to an effective visual resolution. It is found that the image starts to degrade as SNR falls below 50 and below 10 resolution becomes primarily determined by SNR and is insensitive to blur level. The Visual Information Fidelity metric is shown to accurately model the human response.

OCIS codes: 110.3000 image quality assessment 110.3055 information theoretic analysis 330.4060 vision modeling 110.2990 image formation theory

1. Introduction

This paper investigates the influence of noise on effective visual image resolution. Image resolution is often characterized by the minimum separation at which point light sources can be distinguished, for example the Sparrow and Rayleigh resolution criteria[1-2]. These resolution metrics relate to the point spread function (PSF) of the device. But they do not factor in the effect of noise. At low SNR levels the distance between and size of contrast features in the object must be greater than the Sparrow and Rayleigh criteria to be reliably distinguished. This applies whether the task is done visually or by computational processing. Conversely, at high SNR levels, signal processing techniques can be used to determine the separation between two point sources at much smaller separations than these limits[3]. This paper quantitatively characterizes degradation in visually perceived image resolution as a function of both PSF width and SNR level. Additionally a theoretic model based on the Visual Information Fidelity metric[4-7] is demonstrated to accurately match the human visual response at very low SNR levels at which other metrics such as the General Image Quality Equation (GIQE) fail[8-9].

2. Visual Experiment

A visual experiment was designed in which human observers would view and perform a task on a series of displayed test patterns that spanned a broad range of PSF blur and SNR levels. It was desirable that the task had as clear and direct link to quantifying image resolution as possible. It was intended that the test would relate well to images of complex extended scenes such as experienced in terrestrial surveillance, overhead remote sensing, resolved imaging of resident space objects, and medical imaging of tissues. As such the test patterns needed to involve distinguishing contrast features on larger patterns rather than simple pairs of point sources.

A variety of test patterns were considered involving determination of the orientation of 3-bar and other test shapes with periodic features that could be related to a spatial frequency. However, it was found that at blur levels well beyond the point at which there was any visual evidence of the primary contrast cycles across the pattern, the orientation could still be determined by elongation of the blurry blob in a specific direction. Images of the structure that was settled on for the tests are displayed in Fig. 1 at strong and weak blur and noise levels. A base unit pattern of a 2×2 arrangement of bright squares was used in which one of the squares is randomly chosen to contain a smaller darker square. Denoting the width of a dark square by W, the width of the bright squares is $3W$ and their separation is $1W$. The relative light levels in the pristine pattern are 3 for the squares, 2 for the dots, and 1 for the background. A test scene is comprised of a series of this basic pattern repeated over a broad range of scale sizes with largest at upper left, decreasing going down the leftmost column of elements, then decreasing further rastering down consecutive columns to the right.

The task of the human subjects for each displayed image was to find the smallest scale pattern at which they could identify the square containing the dark dot with 80-95% reliability. Practice tests were used to train the subjects to pick the most challenging scale size that they could achieve this detection rate, thus establishing a consistency between subjects in what was deemed the smallest size feature they could visually identify at a particular blur and SNR level.

The base scale of the 2×2 block patterns specified in terms of the width *W* of the dark square ranged from 8/9 to 16 image pixels, with a ~20% increase in size between each pattern and the next larger. The width *W* associated with the pattern scale selected by the human observers was equated to be the smallest identifiable feature size. The full combination evaluated by the observers consisted of 384 test images covering 16 SNR levels from 0.84 to 1000 and PSF widths from 0.2 to 4.75. PSF width was characterized by the half-width at 20% of the max. This measure of width was about 25% bigger than the full-width at exp(-1) and 47% bigger than the full width at half max (FWHM). The images were generated by convolving a 9× higher resolution pristine pattern with the PSF, downsampling to the pixelization of the displayed image, then adding noise. In half the images the noise was Gaussian and in the other half Poisson statistics were used representing photon shot noise. SNR was defined as ½ the peak signal after blurring, divided by the noise level at the brightest region of the image.

A GUI allowed the human subjects to click on the smallest dark square they thought they could reliably identify, then proceeded to the next test image. Completing the full test generally took subjects 40 minutes to an hour not including encouraged breaks. Statistical outliers were investigated but no obvious ones were found. The fraction of correctly identified dark squares ranged from 86.9-93.2% with a mean of 89.4%. For each blur and SNR level the pattern scale sizes *W* identified by the 8 humans subjects were sorted, the highest and lowest values were tossed out, and the mean of the remaining was calculated and equated to resolution in units of pixels.

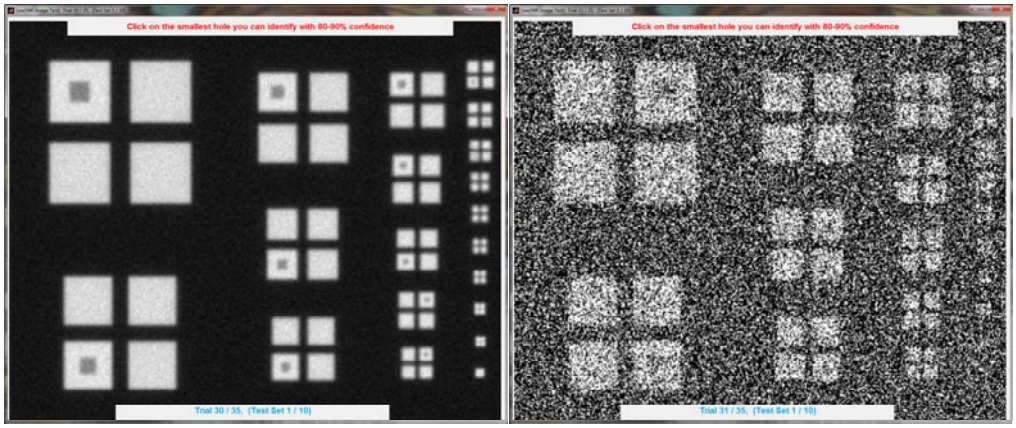

Fig. 1. Example image test patterns used in human subject experiments, (left) case with high SNR and small PSF, (right) case with low SNR and wide PSF.

3. Mathematical Resolution Models

VIF: The Visual Information Fidelty (VIF) metric was calculated for each image. VIF is related to the Shannon mutual information between the degraded and original pristine image. It is based in the wavelet domain and has been shown in many studies to agree well with the human response. Predicted resolution was related to VIF by,

$$\log_2(1 / \text{VIF\_Resolution}) = k_1 * \text{VIF} + k_2 \quad (1)$$

and a least squares regression was used to determine the values of k_1 and k_1 which produced the best fit.

FourierSNR: This resolution metric was predicted based on determining the highest spatial frequency at which the peak Fourier domain SNR was above some threshold. An SNR threshold of 3–4 was found to give a good fit to the human resolution using the following relationship,

$$(1 / \text{FourierSNR\_Resolution}) = k_1 * f_{\text{Thresh}} + k_2 \quad (2)$$

Rose: Resolution values were predicted based on the Rose model. The Rose SNR for each 2×2 block of squared in each blurred image was defined as follows. The contrast signal was calculated as the difference between the brightest value over the region subtended by the square and the signal at the center of the dark dot. This was divided by the average noise level of the pixels subtended by the square to get an SNR value. For each image this produces a curve of SNR values as a function of feature size. This curve was interpolated to determine the point at which it crosses a detection threshold, $\text{SNR}_{\text{RoseThresh}}$. This value is construed as the Rose resolution for the image.

Note that unlike the Rose and Fourier SNR methods, the VIF method is entirely general in that it only needs a pristine and final image, the calculations themselves to not require any knowledge about the imaging system or any post-processing performed to generate the image.

4. Results

Fig. 2a shows a contour plot of the data from the human resolution evaluations as a function of PSF width and SNR. Tracing rightward across the plot along a horizontal line of constant PSF width, effective visual resolution is seen to asymptote to a best (minimum) value by about an SNR of 50. Tracing leftward at any specific PSF width, resolution is seen to degrade quickly as SNR falls below 10. Below SNR 10, the contours of constant resolution are nearly vertical in orientation indicating that resolution is primarily related to SNR in this region and only weakly influenced by the actual PSF size. What is happening here is that in order to see a contrast feature in the image, it needs to subtend a large enough region so that the eye can effectively average over many pixels to see that the mean signal level in that area is different than its surrounding. In mathematical terms, the effective SNR of a $N \times N$ pixel region is N-times higher than for a single pixel. In order to detect the feature reliably, it must be large enough to have an effective SNR that allows it to be distinguished.

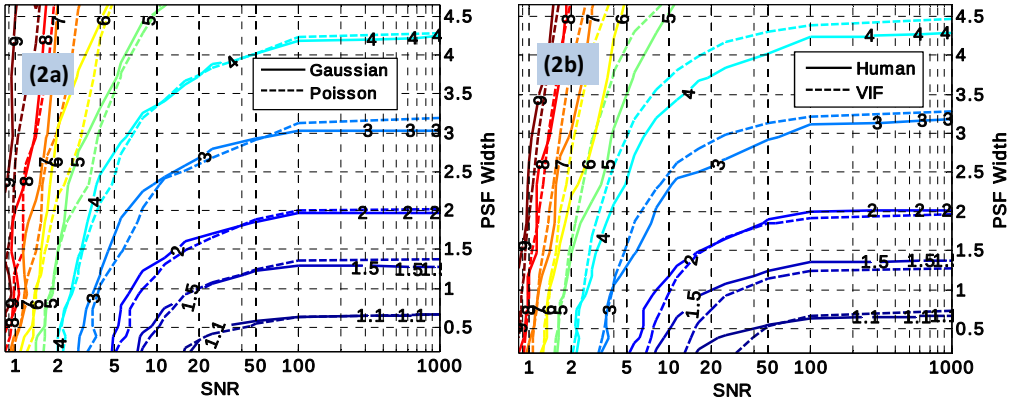

Fig. 2. Left (2a) contour plot of visual resolution vs. SNR and PSF width for Gaussian (solid line) and Poisson noise (dashed). Right (2b), similar to 2a but comparing visual resolution (solid line) to the VIF model prediction (dashed).

Fig. 2b compares the prediction using the VIF-based model to the actual human responses. They match well, generally within 10%. Fig. 3a shows a similar comparison of the Fourier model to human evaluated resolution. It agrees well below SNR of 100, but fails badly at high SNR. This is because the model does include any threshold mechanism and will indicate small features as can be discriminated at extremely large PSF widths as long as the SNR is high enough. However, since it predicts well up to SNR of 100, and the human evaluated resolution asymptotes at this level, this approach can be used as an accurate model by the following simple modification,

Resolution = $\min(\mathcal{F}(S,P), \mathrm{F}(100,P))$

where $\mathcal{F}$ is the Fourier model, S=SNR, P=PSF Width. Fig. 4 compares the Rose model predictions against human evaluated resolution. The accuracies of each model relative to the human evaluations are summarized in Table 1. The VIF model performed the best of the three approaches.

Table 1. Resolution Model Accuracy

Model	Mean Relative Error	Max Relative Error
Rose	16.5%	63.9%
Fourier SNR	11.9	51.4%
VIF	7.9%	36.8%

A second visual experiment was performed involving images of an airplane with a range of PSF widths and SNR levels were further processed by convolving with blur filters of various sizes. In this experiment the task of the human observers was to determine the filter strength which produced the best displayed image. VIF scores were calculated for each filtered image as well. Each graph in Fig. 4 plots the VIF score as a function of filter width for a particular SNR and PSF. In the 3×3 array of graphs, PSNR increases to the right and PSF width increases toward the top. Overlaid on each graph are blue circles indicating the filter sizes selected by 4 human observers, with the mean size indicated by the red stars. The peak of the VIF curve is seen to match well with the mean of the human

selections. This suggests that the VIF metric may could be used to optimize settings for various image enhancement algorithms.

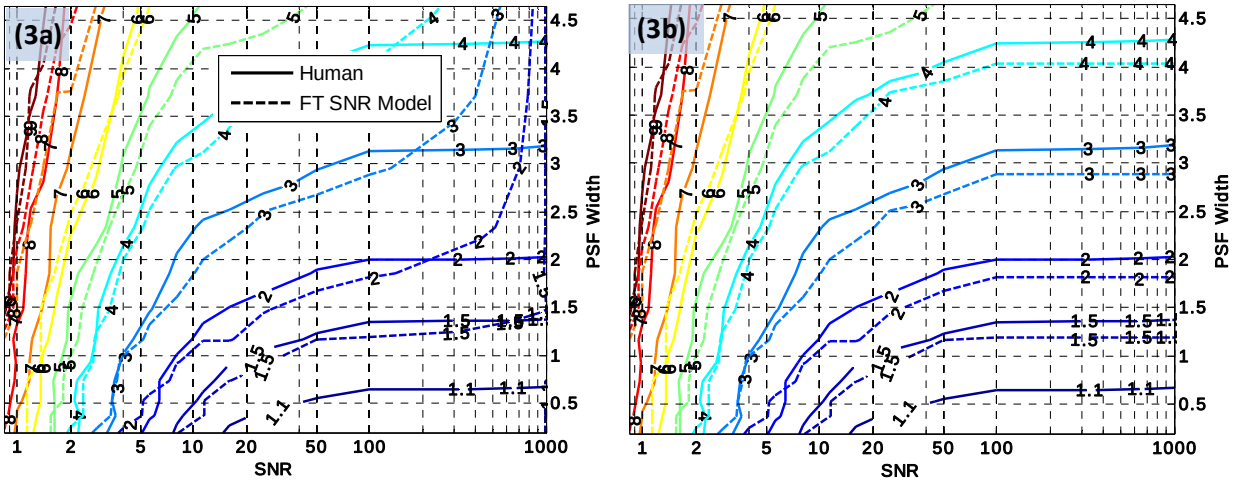

Fig. 3. Comparison of Fourier SNR model to human evaluated resolution before (left) and after (right) adjustment to achieve good match at high SNR levels.

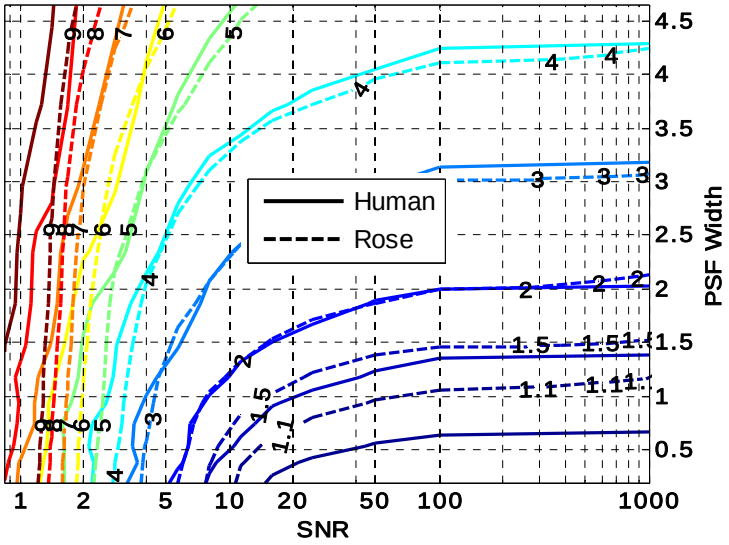

Fig. 4. Comparison of Rose model to human evaluated resolution.

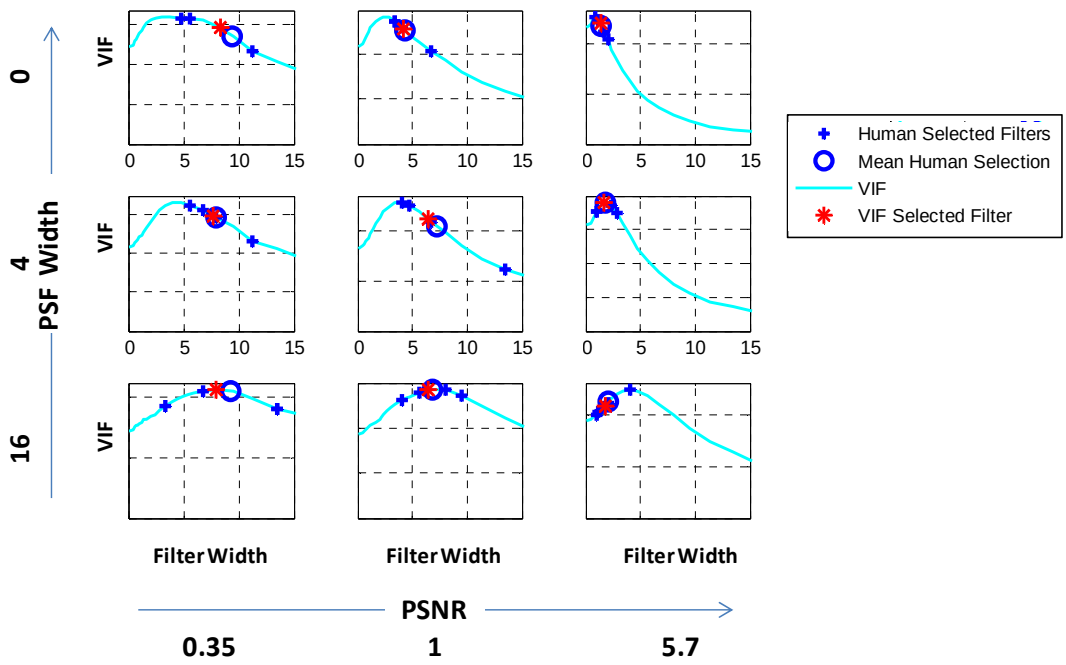

Fig. 4. Filter selected by VIF metric agrees well with filter deemed best by human evaluation.

4. Summary

This study quantified the degradation to visual image resolution at extremely low SNR levels by human evaluation of test patterns. It was shown that as SNR decreases below 10, effective resolution is dominantly determined by SNR with only weak sensitivity to PSF size (for PSFs up to 5 pixels width). The VIF metric was shown to provide an accurate model to human resolution. Furthermore the VIF metric was found to accurately pick filter settings that were near the values deemed by human observers to maximize the visibility of small features on an image of an airplane. This indicates the VIF could be highly useful in predicting image quality at low SNR levels and for automating optimization of input settings for image enhancement algorithms.

5. Acknowledgement

This research was supported by funding from the Air Force Office of Scientific Research (AFOSR) and performed under the Air Force Research Lab IROSS Contract # FA9451-05-C-0257.

References

1. C. M. Sparrow, 1916, ApJ, 44, 76.
2. E. Hecht, *Optics*, 4th edition, Addison-Wesley, 2001.
3. M. Shahram and P. Milanfar, "Imaging below the diffraction limit: a statistical analysis," IEEE Trans. Image Proc. **13**(5) 677-689 (2004).
4. H. R. Sheikh and A. Bovik, "Image Information and Visual Quality", IEEE Trans. Image Proc. **15**(12) 430-444 (2006).
5. H.R. Sheikh, M.F. Sabir, and A. Bovik, "A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms," IEEE Trans. Image Proc. **15**(11) 3441-3452 (2006).
6. D. R. Gerwe, C. E. Luna, B. Calef, "Information Theoretic Based Image Quality Evaluation," (invited) OSA Signal and Recovery Conference STuC1 (2009).
7. D. Gerwe, B. Calef, C. Luna, "Applications of the ITIQUE Image Quality Modeling Metric to SSA Domain Imagery," AMOS Tech. Conf. Proceedings (2012).
8. J. C. Leachtenauer, W. Malila, J. Irvine *et al*, "General Image-Quality Equation: GIQE," Appl. Opt. 3**6**(32) 8322-8328 (1997).
9. S. T. Thurman, J. R. Fienup, "Analysis of the general image quality equation," Proc. SPIE **6978** F1-13 (2008).