Application of the ITIQUE Image Quality Modeling Metric to SSA Domain Imagery

David. R. Gerwe
Boeing Phantom Works – Advanced Networks & Space Systems, Los Angeles, CA
Carlos E. Luna
Boeing Phantom Works – Advanced Networks & Space Systems, Los Angeles, CA
Brandoch Calef
Boeing Laser Technical Services, Maui, HI

CONFERENCE PAPER

This paper evaluates the ITIQUE image quality modeling framework for SSA applications. Based on Bovik and Sheik's VIF metric, ITIQUE evaluates the Shannon mutual information (MI) at multiple spatial scales between a pristine object and the image output from a detailed image formation chain simulation. Integrating the MI at each spatial scale and applying a calibration offset produces a prediction of NIIRS image quality indicating the level of interpretation tasks that could be supported. The model enables prediction of NIIRS quality obtainable as dependent on image collection conditions and image system design including both hardware and processing algorithms. The ITIQUE framework could facilitate concept evaluation and engineering design by quantitatively relating image formation performance directly in terms of end end-user mission needs. Previous work focused on overhead imagery of terrestrial scenes and linear processing only. This paper considers ground-based imaging of SSA objects and extends the previous study to include non-linear processing. A range of turbulence strengths and SNRs are included. ITIQUE predictions are shown to match well to results from a human visual assessment study in which a panel of observes rated NIIRS quality of the same imagery.

1. INTRODUCTION

Image quality metrics can be broadly classified into subjective and objective. The former refer to ratings of image quality obtained by a human observer. In contrast, objective image quality metrics are numerically computed from the image data or a mathematical description of the relevant imaging system. Objective image quality metrics can be a useful tool in the design and evaluation of new imaging systems. Of particular interest are objective quality metrics that predict the value of imagery in terms of image analysis tasks that can be successfully accomplished.

The overhead and tactical imaging communities have successfully used several task-based image quality metrics over the past few decades. For example, the National Imagery Interpretability Rating Scale (NIIRS) and the General Image Quality Equation can be used to evaluate an image in terms of its utility for detection/classification/identification tasks. The probability of combat ID (PCID) and the closely related Johnson criteria provide the probability of correct combat identification for objects based on the resolution elements across the object.

For Space Situational Awareness (SSA) applications, objective image quality metrics are sorely needed. There have been several attempts to develop subjective image quality metrics based loosely on the NIIRS approach. In this paper, we describe an information theoretic framework to numerically predict image quality (ITIQUE) and its extension to evaluate SSA imaging system performance in terms of end-user mission needs.

The rest of the paper is organized as follows: Section 2 presents background information on NIIRS and its predictive tool the GIQE. Also, an alternative method based on an Engineering NIIRS Ruler which allows for human subject studies without resorting to trained imagery analysts. Section 3 presents the ITIQUE framework and notes its successful application in previous studies focused on the overhead imagery domain. Section 4 presents results of a new study demonstrating its applicability to the SSA arena. Section 5 concludes the paper.

2. BACKGROUND

2.1 National Image Interpretability Rating Scale (NIIRS)

The NIIRS scale was introduced in 1974 and has had a long history of successful application in the evaluation of the informational potential of images[1-3]. The scale indicates the level of analysis task that can be performed with a particular image. NIIRS is a 10 level scale with ratings ranging from 0 (useless) to 9 with each level corresponds to a particular type of image analysis task. The criteria associated with each level have been adapted to provide different versions of the NIIRS scale for particular applications and sensing modalities[3].

Fig. 1 below has some example overhead images with corresponding NIIRS. The image on the left has a NIIRS rating of 4 which allows identification of farm buildings as barns, silos or residences for example. The center image is NIIRS 5 which allows identification of individual Christmas tree plantations. Finally, the image on the right is NIIRS 8 which allows identification of grill detailing and/or license plates on a passenger vehicle or truck. An increase in a level of NIIRS corresponds roughly to a doubling of resolution. This criteria-based scale directly relates NIIRS to human interpretability and mission utility and value. Unfortunately, to be accepted within large defense programs image quality performance studies typically require ratings by a panel of formally trained image analysts. This presents a practical limitation on the frequency with which NIIRS evaluation can be used as a tool for engineering analysis and system design.

Fig. 1. Sample overhead imagery and its corresponding NIIRS ratings illustrate the level of image analysis tasks that could be supported. Description of NIIRS level criteria are provided in the text above.

2.2 General Image Quality Equation (GIQE)

The NIIRS scale is very useful in describing the information utility of an image. However, it is not a practical tool for characterizing the performance of imaging systems given the need for trained imagery analysts. To overcome this, the GIQE[2] can be used to predict NIIRS values based on image system parameters. The original GIQE is given by,

$$NIIRS_{GIQE} = 11.81 + 3.32\log_{10}\left(\frac{RER_{GM}}{GSD_{GM}}\right) - 1.48H_{GM} - \frac{G}{SNR} \quad , \qquad (1)$$

where RER_{GM} is the geometric mean of the relative edge response of the system, GSD_{GM} is the geometric mean of the ground-sample-distance, H_{GM} is the geometric mean of the edge response overshoot caused by MTF compensation, G is the gain of the MTFC kernel and SNR is the signal-to-noise ratio. The equation captures the key trade between sharpness and noise-amplification.

The coefficients for each of the terms in Eq. 1 above were obtained by regression to fit the results of an image evaluation study conducted with 10 trained imagery analysts. A major restriction of the GIQE is that its predictions are only valid over the range of degradation types and levels considered for the underlying IA rating study. These were limited in scope to conventional systems with well-behaved stable PSFs without severe wavefront aberrations or motion induced blurring. New imaging conditions require additional IA studies and statistical analysis[4]. Furthermore, the form of the equation requires that the system can be accurately modeled as linear. Thus it is not applicable to systems that include non-linear image reconstruction or enhancement processing. In its current form, the GIQE has limited applicability to the evaluation of SSA imagery.

2.3 Engineering NIIRS Scale (ENS)

A major limitation in use of the NIIRS scale to imaging system performance evaluation is the limited availability of trained imagery analysts to obtain formally recognized ratings. A viable alternative is to utilize an engineering NIIRS scale to obtain image ratings using an untrained panel of observers. The images are evaluated by subjects using an image selection GUI such as the one shown in Fig. 2. The subjects are asked to choose from a sequence of reference images to find the one that most closely matches the image under evaluation on the basis of ability to match or detect features. The reference image sequence forms a calibrated "NIIRS ruler" consisting of a set of images generated at fixed Δ-NIIRS steps. Fig. 3 illustrates the NIIRS ruler concept. Thus each reference image translates directly to a NIIRS rating. In the illustrated GUI, the subjects use a slider to scan through the ruler to find the best match. The right panel of the GUI can be set to flicker between the reference ruler and test image to facilitate comparison.

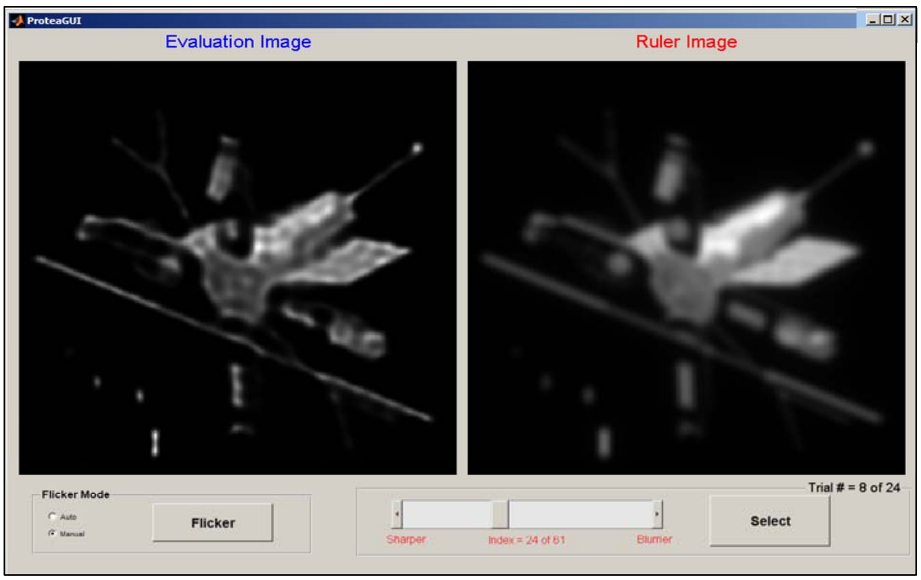

Fig. 2: Image selection GUI facilitates NIIRS assessment without the need for trained imagery analysts.

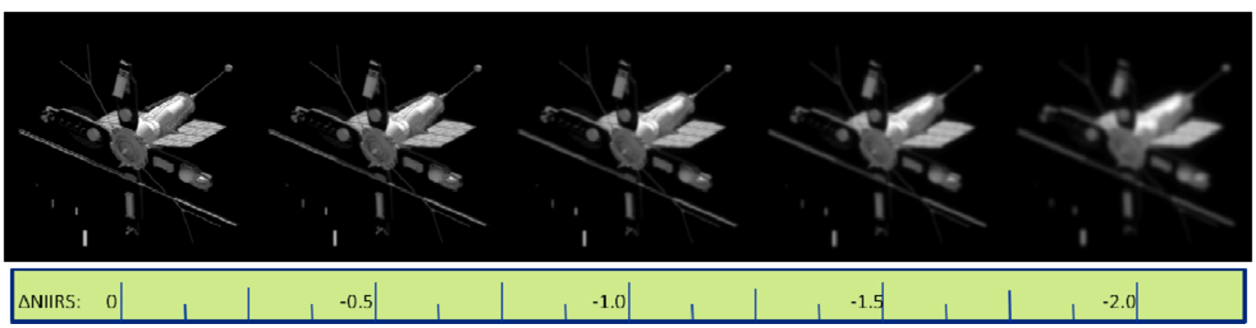

Fig. 3: Calibrated images at fixed Δ-NIIRS steps comprising a NIIRS ruler.

3. INFORMATION THEORETIC IMAGE QUALITY EQUATION (ITIQUE)

3.1 ITIQUE Framework

Information based image quality metrics have been introduced in [5-7]. In [8] the concept of an Information Theoretic Image Quality Equation (ITIQUE) relating mutual information to NIIRS was introduced. The ITIQUE framework utilizes the Visual Information Fidelity (VIF) introduced in [5] to measure the perceptually relevant mutual information between a reference and pristine image. The framework is depicted graphically in Fig. 4.

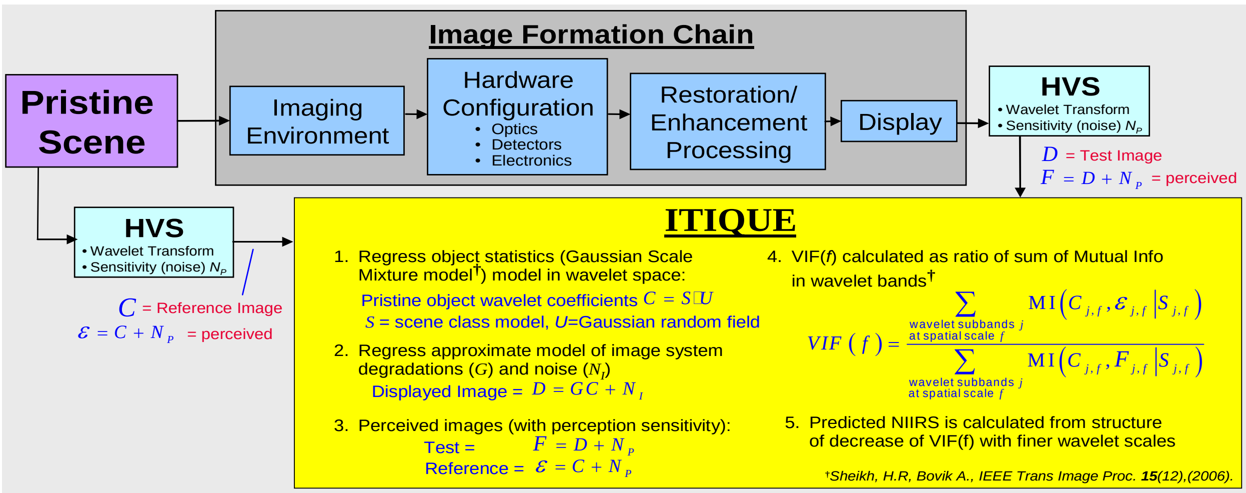

Fig. 4. ITIQUE framework uses VIF concept to measure perceptually relevant mutual information at several feature scale sizes and predict NIIRS

The reference image is the pristine input scene as seen through a model of the Human Visual System (HVS). The test image is the degraded image acquired by the imaging system, possibly post-processed by restoration or enhancement algorithms, as perceived by the HVS. The HVS model excludes any perceptually irrelevant information contained in the image data. The MI is calculated in the wavelet domain for both reference and test images.

At each wavelet sub-band, the VIF is calculated as the ratio of the MI between the pristine original object and the perceived image after measurement and enhancement processing by the imaging system to the MI between the pristine original object and the perceived object as viewed *in situ*. The VIF is only influenced by object features relevant to perception by the HVS and allows for possible information gain resulting from image enhancement.

ITIQUE predicts NIIRS values by combining the VIF at multiple feature size scales according to

$$NIIRS_{ITIQUE} = C_1 - C_2 \ \log2\left(\frac{1}{\sum_f VIF(f,\sigma_H)}\right), \qquad (2)$$

where C_1 is a bias coefficient, C_2 is a scaling coefficient, the sensitivity of the human visual system is characterized in terms of an effective noise level σ_H characterizes, and the denominator inside the logarithm is the sum of the VIF computed at various relevant feature scale sizes. The value of the coefficients in Eq. 2 can be obtained by regression against NIIRS ratings obtained by the ENS method described previously or a formal assessment using imagery analysts.

An assessment of the ITIQUE framework for predicting NIIRS was presented in [8]. In that work, the authors obtained NIIRS ratings using an ENS methodology including 13 subjects and 160 images. The images were a subset of a larger dataset comprising a variety of terrain types and imaging conditions relevant for overhead imaging of terrestrial scenes. The NIIRs predictions obtained with ITIQUE agreed well with the ENS scores of the study with root mean square error (RMSE) and mean absolute errors (MAE) of 0.230 and 0.182 ΔNIIRS respectively.

Over the wide range of blur types and strengths and SNR levels considered the ITIQUE metric exhibited a stronger correlation to the ENS scores than GIQE predictions.

4. EXTENSION OF THE ITIQUE FRAMEWORK TO THE SSA DOMAIN

A study was designed to assess the accuracy of the ITIQUE metric for SSA domain imagery. A set of 480 images were generated using a high-fidelity image simulation model. This set included two different objects, 6 levels of degradation for 4 different degradation mechanisms, and five methods of image reconstruction processing. Details are listed in Tables 1 and 2. The image set was comprised of four sub-sets. The first set contained 210 images and was designed to study the sensitivity the unprocessed and restored images to each single degradation mechanism (e.g. all but one degradation set to its minimum level). The second set of 20 were simply repeats of the first set, which were included to enable characterization of the self-consistency of each subject's scoring. The third set contained 125 images with the object, processing method, and level of each degradation type chosen completely randomly from the full space of possibilities listed in Tables 2 and 3. The last set also contained 125 images with degradation levels for each mechanism chosen randomly, but restricted so the sum of the degradation level indexes was less than 8 (where degradation level for each type was indexed from 1-6 from, larger numbers corresponding to more severe loss to image quality).

Degradation Type	Description	Range
Jitter	Gaussian blur (parameterized by full width at 1/e point)	{0,2,3.1,4.9,7.7,12} Units = pixels
Wavefront	Kolmogorov statistics scaled to desired D/r_o. Wavefronts were uncorrelated between each of 16 images used in a reconstruction.	{0,1.7,3.0,5.3,9.2,16} Units = D/r_o
Aperture size (Diffraction)	Diffraction from full circular aperture.	{0.25,0.38,0.57,0.87 1.3,2.0}, Units=meters
SNR	(Signal σ) / (noise σ) Noise is mixture of Gaussian read noise and Poisson shot noise. SNR scaled by adjusting exposure time.	{1,2.5,6.3,15.8,39.8, 100} Units = SNR

Table 1. Image degradation mechanisms and levels included in the study.

Processing Method	Description
None	
Shift-and-Add	Images registered by correlation then averaged
Wiener	Multi-frame linear Wiener filter deconvolution
Multi-Frame Deconvolution	Non-linear iterative Maximum Likelihood deconvolution with perfect knowledge of PSFs for each frame
Multi-Frame Blind Deconvolution	Same as MFD but PSFs are estimated as parameterized by wavefront Zernike coefficients

Table 2. Image restoration processing methods included in the study.

The images were rated using the GUI tool illustrated in Fig. 2 by a panel of 12 volunteer subjects drawn from several different divisions of Boeing (most of which had participated in previous studies). The images were presented in groups of 20 sub-sets each of 20. Image order was randomly shuffled between subjects. Each sub-test took around 10 minutes. The subjects were instructed to intersperse taking the sub-tests between other activities over a period of several days to ensure adequate breaks were taken between tests to avoid eye and mental fatigue or decline in attention. For each image the ENS scores over all subjects were averaged after throwing out the minimum and maximum scores to mitigate outliers. Evaluation of the repeated images did not indicate any subjects to be anomalously inconsistent. ITIQUE scores for each image were calculated using Eq. (2) with the parameters, C_1, C_2, and σ_H optimized to achieve the best fit to the ENS scores. The RMS error between the ITIQUE and ENS scores was not found to differ significantly whether the fit was performed using all 480 images or just the first set of 210. The scatter plot in Fig. 5 shows a strong correlation between ITIQUE predictions and ENS ratings over the full range of degradation types, levels, and restoration methods. The RMSE and MAE between the ITIQUE scores and ENS ratings were 0.34 and 0.26 respectively. This is not quite as good as the ~0.2 ΔNIIRS accuracy achieved with ITIQUE for the overhead imagery domain[8] but is still deemed an acceptable accuracy to be highly useful for engineering studies. However, direct comparison isn't fair since that the overhead domain studies did not include as severe range of degradation levels and did not include non-linear restoration methods. Note that 0.1 ΔNIIRS represents a just noticeable difference. There does appear to be a consistent bias for ITIQUE to over-predict image quality at the lowest range in image quality considered. A good understanding of the cause for this has not been reached.

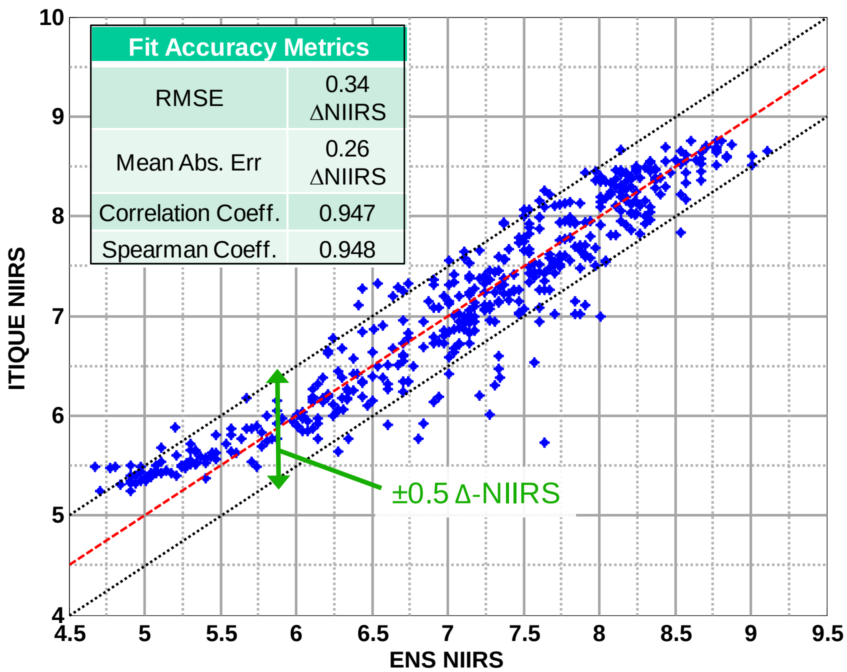

Fig. 5. ITIQUE metric scores are well correlated with ENS ratings.

Fig. 6 compares the ITIQUE metric scores against the ENS study trends for the impact of the each individual degradation type to restored image quality for the various algorithms considered. The trends are seen to be highly consistent with the discrepancies in the curves being dominated by displacement by a constant. Fig. 7 plots the differences between the ITIQUE and ENS scores. No systematic biases are observed with the exception of a relative decrease in the ITIQUE values for all processing methods at SNRs ≤ 4. Fig. 8 plots the improvement gained by the different processing methods over the raw imagery and shows strong agreement of the ITIQUE curves to those resulting from the ENS scores. This illustrates the utility of ITIQUE as an analysis tool for system and algorithm design, optimization, comparison, and selection.

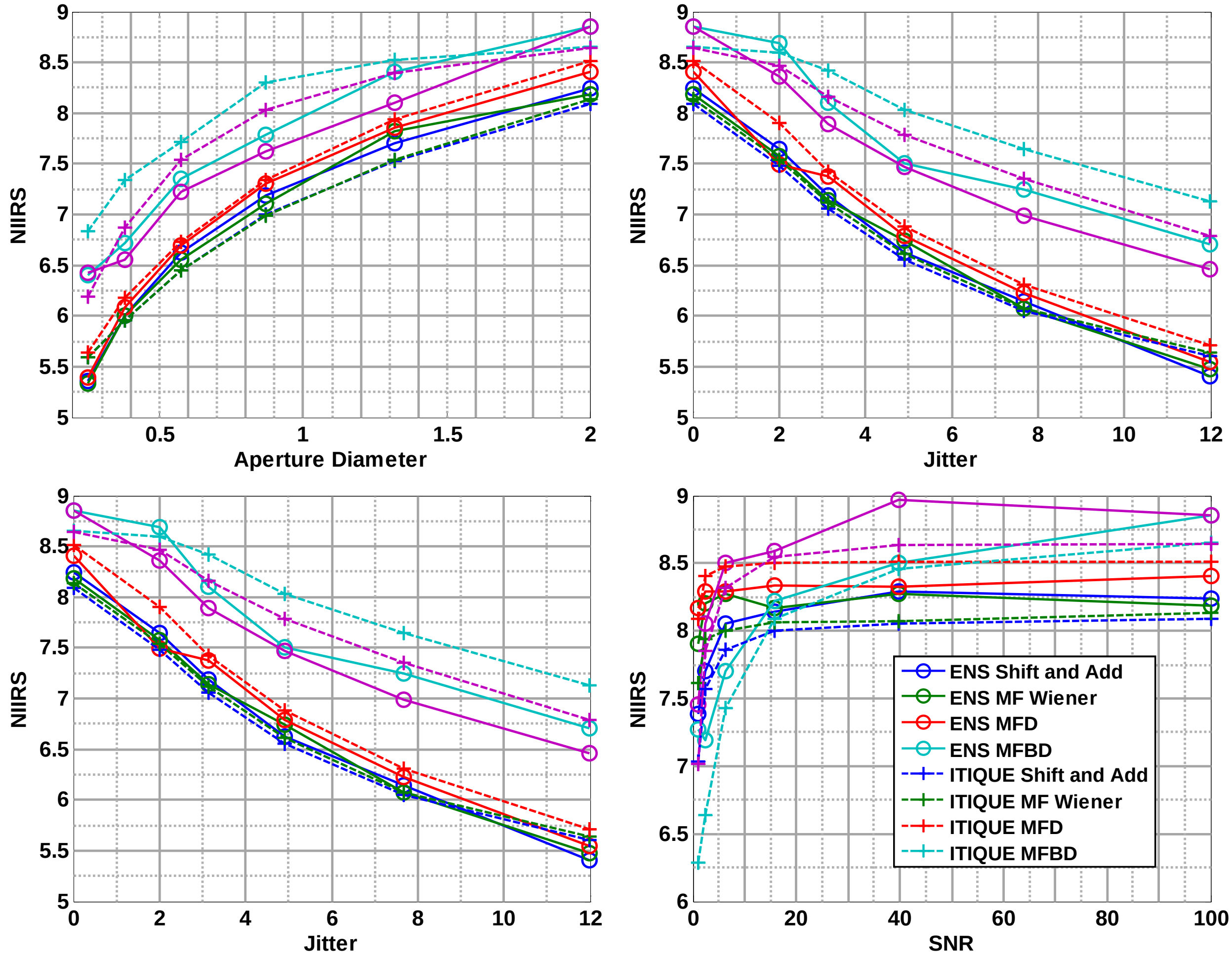

Fig. 6. Comparison ITIQUE to ENS curves versus degradation levels show highly consistent trends between them for all processing methods.

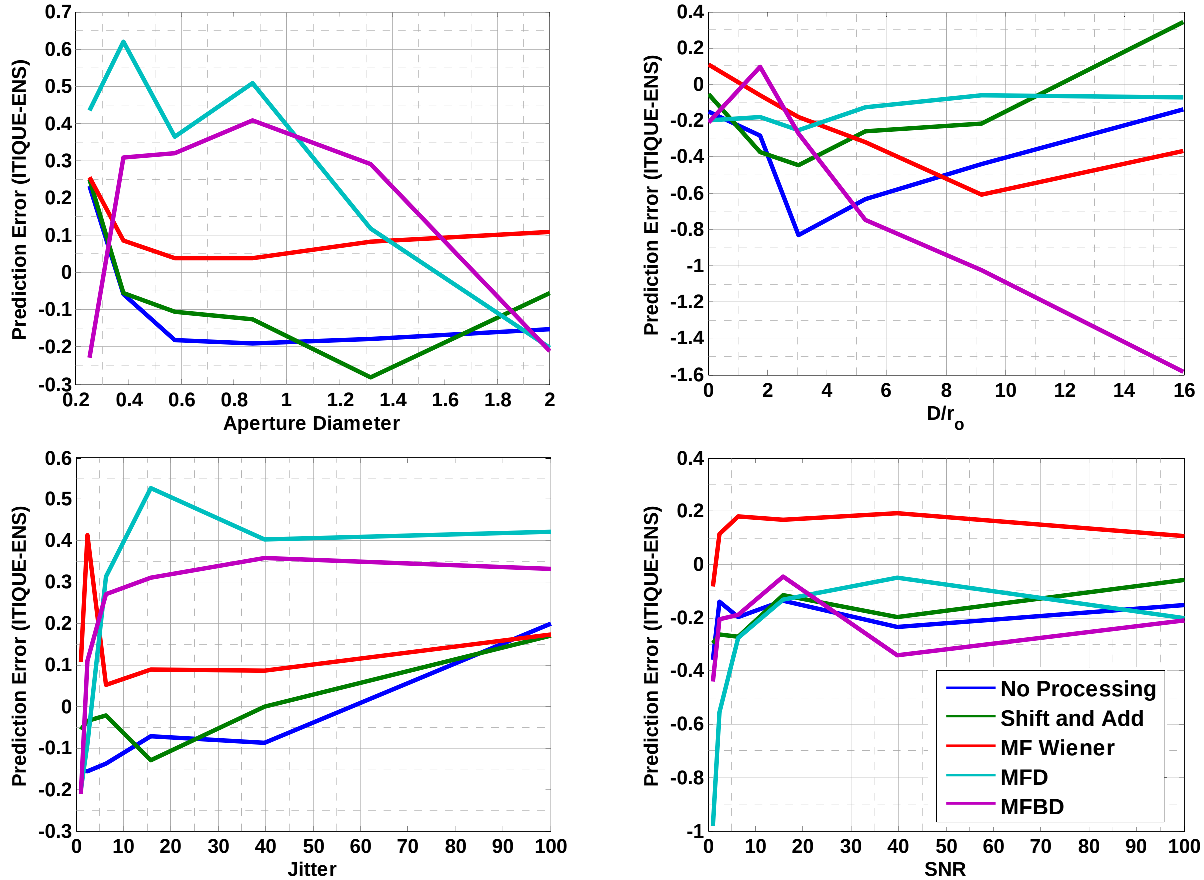

Fig. 7. Differences between ITIQUE and ENS scores do not indicate systematic biases.

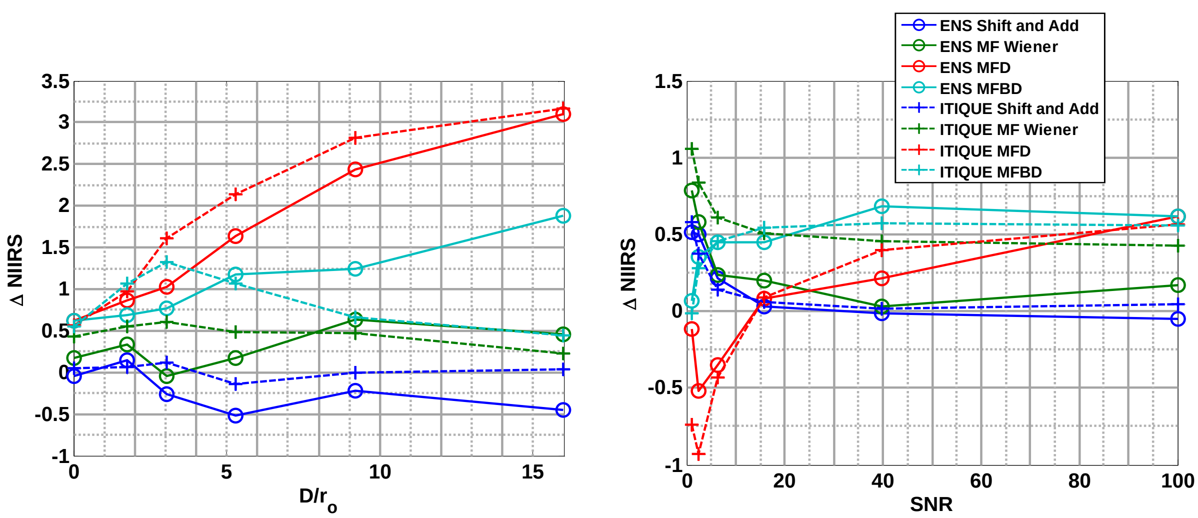

Fig. 8 plots the improvement gained by different image restoration algorithms over the raw imagery, illustrating the utility of ITIQUE as an analysis tool for system and algorithm design, optimization, comparison, and selection.

5. CONCLUSION

The described study establishes ITIQUE as a viable metric and engineering tool for system for supporting analysis of Space Situational Awareness domain imaging systems and image reconstruction algorithms. ITIQUE scores agreed well with ENS scores obtained using a panel of human subjects and spanning a wide range of degradation types and levels and image restoration methods. ITIQUE reliably predicted quantitative trends of the impact of different degradations on each reconstruction algorithm considered. Of strong importance this included severe wavefront aberrations and non-linear algorithms, neither of which can be treated by the current GIQE NIIRS prediction model. This is a strong step forward in filling a deficiency in computer calculable metrics that correlate well with the level of an image's visual information content.

6. ACKNOWLEDGEMENT

This research was supported by funding from the Air Force Office of Scientific Re search (AFOSR) and performed under the Air Force Research Lab IROSS Contract # FA9451-05-C-0257. The authors also wish to express their appreciation to Prof. Sudhakar Prasad for conversations that motivated this work, to Dr. Doug Hope for his involvement at the early stages of this research on options for expressing and calculating MI between images, and to Dr. Greg Pavlin and Dr. Chuck Matson for their many constructive comments.

7. REFERENCES

1. Driggers, R. G., Cox, P. and Kelley, M., "*National imagery interpretation rating system and the probabilities of detection, recognition and identification*," Opt. Eng. **36**(7), 1952-1059 (1997).
2 Leachtenauer, J. C. et al, "*General Image Quality Equation: GIQE*", Appl. Opt. **36**(32), 8322-8328 (1997).
3 J. M. Irvine, "National Imagery Interpretability Rating Scales (NIIRS): Overview and Methodology," SPIE Proc. **2128**, 93-103 (1997).
4 Fiete, R. D., Tantalo, T. A., Calus, J. R., Mooney, J. A. "Image quality of sparse-aperture designs for remote sensing," Opt. Eng. **41**(8) 1957-1969 (2002).
5 Sheikh, H. R., Bovik A. C. and Veciana, G., "*An Information Fidelity Criterion for Image Quality Assessment Using Natural Scene Statistics*," IEEE Trans. Image Proc., **14**(12), 2117-2128 (2005).
6 Sheikh, H. R. and Bovik A. C., "*Image Information and Visual Quality*," IEEE Trans. Image Proc., **15**(2), 430-444 (2005).
7 Prasad, S., "*Statistical-information-based performance criteria for Richardson-Lucy image deblurring*", JOSA A, **19**(7), 1286-1296 (2002).
8 D. R. Gerwe, C. E. Luna, B. Calef, "*Information Theoretic Based Image Quality Evaluation*," OSA Signal and Recovery Conference STuC1 (2009).