STATISTICS OF NATURAL FUSED IMAGE DISTORTIONS

David E. Moreno-Villamarín[⋆] *Hernán D. Benítez-Restrepo*[⋆] *Alan C. Bovik*[†]

[⋆]Departmento de Electrónica y Ciencias de la Computación
Pontificia Universidad Javeriana, Seccional Cali, Calle 18 No 118-250, Cali, Colombia
Email: david.moreno@ieee.org, benitez@ieee.org
[†]Department of Electrical and Computer Engineering
The University of Texas at Austin, TX 78712, Austin, USA
Email: bovik@ece.utexas.edu

ABSTRACT

The capability to automatically evaluate the quality of long wave infrared (LWIR) and visible light images has the potential to play an important role in determining and controlling the quality of a resulting fused LWIR-visible image. Extensive work has been conducted on studying the statistics of natural LWIR and visible light images. Nonetheless, there has been little work done on analyzing the statistics of fused images and associated distortions. In this paper, we study the natural scene statistics (NSS) of fused images and how they are affected by several common types of distortions, including blur, white noise, JPEG compression, and non-uniformity (NU). Based on the results of a separate subjective study on the quality of pristine and degraded fused images, we propose an opinion-aware (OA) fused image quality analyzer, whose relative predictions with respect to other state-of-the-art metrics correlate better with human perceptual evaluations.

Index Terms— NSS, LWIR, multi-resolution image fusion, fusion performance, image quality

1. INTRODUCTION

In recent years, increasing levels of uncertain global security, along with the availability of cheap, intelligent digital cameras is encouraging interest in the development of video systems capable of detecting anomalies or events that may affect the economics and safety of human activities [1]. Popular outdoor video surveillance systems that rely on electro-optical sensors are often prone to failures due to ambient illumination changes and weather conditions [2, 3]. One way of improving performance is to use alternate modes of sensing, such as infrared sensing. Decreasing costs and increasing miniaturization has made infrared sensing an interesting element in surveillance system design [4, 5]. Two main benefits of the joint use of thermal and visible sensors are: the complementary nature of the two modalities and the information redundancy captured by the sensors, which increases the reliability and robustness of a surveillance system. These advantages have motivated the computer vision community to study and investigate algorithms for fusing infrared and visible light videos for surveillance applications [5].

Due to a growing interest in LWIR and visible light image fusion, considerable efforts have been made to develop objective quality measures of fused images. The performance of different image fusion algorithms have been evaluated by image fusion quality metrics that are based on information theory [6], space- and frequency based image features [7], image structural similarity [8], and models of human perception [9,10]. Chen and Blum [10] investigated the performance of fusion metrics based on human vision system models assuming the presence of several levels of additive white Gaussian noise (AWGN). Liu et al [11] analyzed the impact of AWGN and blur on fused images. They found that the quality of fused images degrades with decreases in the quality of the images being fused. When the AWGN level was severe, the fused images were of almost the same quality, regardless of the fusion scheme used. These studies did not analyze important real distortions occuring in LWIR sensors, such as non uniformity (NU) and the "Halo Effect." Although extensive work has been conducted on studying the NSS of visible light images and their relationship to picture quality [12–15] and some studies have been done on the statistics of LWIR images [16, 17], very little work has been done on analyzing the statistics of fused LWIR-visible light images, and how those statistics might be affected by the presence of any of multiple possible impairments.

This study analyzes how image distortions such as AWGN, blur, JPEG compression, and non-uniformity noise in LWIR and visible light images affect the statistics (NSS) of fused LWIR-visible light images. We deploy previous bandpass image statistical models proposed in [16, 18] as a starting point, and create an 'opinion-aware' (OA) no-reference image quality prediction model using them. A deep comparison of the results obtained by the proposed OA model with those of state-of-the-art algorithms shows that our new model achieves

(a) (b) (c) (d)

Fig. 1. Example of fused images after the following distortions were applied to the constituent visible light and LWIR images. (a) AWGN. (b) NU. (c) Blur. (d) JPEG compression. Images obtained from [17].

highly competitive results.

1.1. Distortion Models

Several studies have characterized and modeled noise in the LWIR spectrum. Images obtained from focal plane arrays (FPA) can present NU fixed pattern noise [19], which produces a grid-like pattern. In [20] Pezoa and Medina describe an additive model of this type of noise in LWIR images. This spectral model is:

$$|\tilde{I}(u,v)| = B_u \exp\left(\frac{-(u-u_0)^2}{2\sigma_u^2}\right) + B_v \exp\left(\frac{-(v-v_0)^2}{2\sigma_v^2}\right) \quad (1)$$

$$\angle\tilde{I}(u,v) \sim U[-\pi,\pi] \quad (2)$$

where $\tilde{I}$ is the Fourier Transform of the non-uniformity noise. The parameters u_0 and v_0 represent the spectral location (horizontal and vertical noise center frequencies), $B_u = B_v = 5.2$ are the directional amplitudes, $\sigma_u = \sigma_v = 2.5$ are the scales of the respective horizontal and vertical bands, and $U[a,b]$ is the uniform distribution over the interval $[a,b]$. The distortion level is controlled using the standard deviation parameter σ_{NU}, which scales the dynamic range of the NU noise. Other common types of distortion which could affect both LWIR and visible light images are considered here, such as AWGN, blur, and JPEG compression.

Three distortion levels are used throughout the study for each distortion type, which were applied to the LWIR and visible light images of the three databases. For AWGN and NU the standard deviation was varied as $\sigma_{AWGN} = \sigma_{NU} = \{0.0025, 0.01375, 0.025\}$; for blur, a Gaussian blur kernel of size 15 $\times$ 15 pixels with $\sigma_{blur} = \{1, 2, 3\}$ was used; and for JPEG compression, the quality was set to 100, 90 and 80 percent using the "`imwrite`" Matlab algorithm. Figure 1 depicts several fused images obtained when both image sources were affected by the most severe distortion level.

This study of multimodal image fusion uses databases that we hereafter refer to as OSU [21], TNO [22], and MORRIS [17], which contain indoor and outdoor scenes of urban and rural environments.

1.2. Multi-resolution Fusion Methods

In night-vision applications, one of the most commonly-used tools is multi-resolution image fusion (MIF), which aims to retain the main features from the source images [23]. Our study considers the following MIF algorithms: average (AVG), gradient pyramid (GP) [24], and shift-invariant discrete wavelet transform with Haar wavelet (SIDWT) [25]. The decomposition level used in each of the algorithms was set to four, and the fusion rule in each case was the maximum of the high-pass pair of channels and the average of the low-pass channels. Our work deploys direct heterogeneous image fusion schemes based multi-resolution analysis.

2. NSS OF FUSED LWIR AND VISIBLE IMAGES

2.1. Processing Model

The most successful IQA models are based on bandpass statistical image models first observed by Ruderman. Here it is assumed that 'natural' visible light or LWIR images are captured by an optical camera, rather than being generated by a computer using artificial processing. Natural scene statistics are closely related to models of the responses of visual neurons [26, 27]. Ruderman found that removing the local sample means from a natural image and normalizing by the local sample standard deviations has a strong Gaussianizing and decorrelating effect on the resulting statistical distribution of the image. This operation produces the Mean-Subtracted Contrast Normalized (MSCN) coefficients of an image, which are computed as follows:

$$\hat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + C} \quad (3)$$

where I is a luminance image or image patch with $i \in 1,2,...M$ and $j \in 1,2,...N$, where M and N are the image height and width, respectively. The constant C is usually set as 1, thereby preventing division by zero or a small number. The local mean μ and standard deviation σ are defined as:

$$\mu(i,j) = \sum_{k=-K}^{K}\sum_{l=-L}^{L} w_{k,l} I_{k,l}(i,j) \quad (4)$$

$$\sigma(i,j) = \sqrt{\sum_{k=-K}^{K}\sum_{l=-L}^{L} w_{k,l}\left(I_{k,l}(i,j) - \mu(i,j)\right)^2} \quad (5)$$

where w is a 2D circularly-symmetric Gaussian weighting function sampled out to 3 standard deviations and normalized to unit volume.

We also model the distributions of four 'paired product' coefficients (pp) calculated as the products of adjoining MSCN coefficients [18], a set of log-derivative coefficients (pd) [28], which are intended to provide higher sensitivity to

high-frequency noise, and the divisively normalized steerable pyramid decomposition (sp) to describe oriented band-pass characteristics [14, 16]. By using closed form statistical models to parametrically fit these histograms it is possible to extract distortion-sensitive features.

2.2. Feature Models

It has been established that empirical distributions of both high-quality and distorted images that have been subjected to bandpass processing followed by divisive normalization, can be well modeled as following a Generalized Gaussian Distribution (GGD) [16, 18]. The standard method is to fit the histogram of the coefficients to a GGD probability density function:

$$f(x;\alpha,\sigma) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp\left(-\left(\frac{|x|}{\beta}^{\alpha}\right)\right) \tag{6}$$

where α is the shape parameter, σ the standard deviation, and Γ is the Gamma function. The products of spatially adjacent bandpass/normalized coefficients are well modeled as following an Asymmetric Gaussian Distribution (AGGD):

$$f(x;v,\sigma_l,\sigma_r) = \begin{cases} \frac{v}{(\beta_l+\beta_r)\Gamma(1/v)} \exp\left(-\left(\frac{|x|}{\beta_l}^{v}\right)\right) & x<0 \\ \frac{v}{(\beta_l+\beta_r)\Gamma(1/v)} \exp\left(-\left(\frac{|x|}{\beta_l}^{v}\right)\right) & x\geq 0 \end{cases} \tag{7}$$

where

$$\beta_l = \sigma_l \sqrt{\frac{\Gamma(1/v)}{\Gamma(3/v)}} \tag{8}$$

and

$$\beta_r = \sigma_r \sqrt{\frac{\Gamma(1/v)}{\Gamma(3/v)}} \tag{9}$$

Here v is the shape, and σ_l and σ_r are the spread parameters of the left (negative) and right (positive) sides of the model density. We estimate the GGD (α, σ) and the AGGD parameters (v, σ_l, σ_r) using the moment matching technique in [29] as in [16]. For each coefficient product image, a mean parameter is also computed:

$$\eta = (\beta_r - \beta_l)\frac{\Gamma(2/v)}{\Gamma(1/v)} \tag{10}$$

We obtain 46 features, which are later computed over three scales: the initial image scale, and the following two by reducing the resolution by a factor of two, yielding a total of 138. As a way of visualizing the features and the way that they cluster in response to the presence of distortion, we projected an exemplar set onto a two-dimensional space using Principal Component Analysis (PCA). Figure 2 depicts the two-dimensional PC space of features extracted from all of the fused images contained in the databases, and for each one of the considered fusion algorithms.

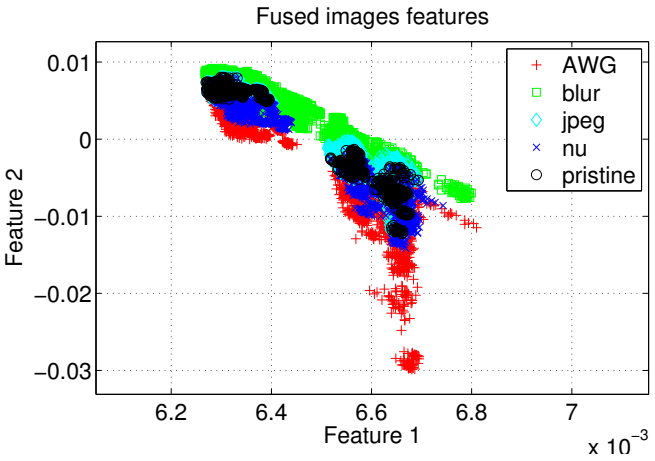

Fig. 2. A total of 138 features per image are projected in a 2D space using PCA with a cumulative variance of 0.9973. Distorted images tend to cluster away from pristine images.

3. QUALITY ASSESSMENT OF FUSED LWIR AND VISIBLE IMAGES

3.1. Subjective Study

We separately conducted a subjective study, which we use here to also create a trained opinion aware IQA model, and also to be able to assess how well it correlates with subjective judgments. We conducted the study on 27 volunteers. Each subject evaluated 150 single stimulus images over each of five testing sessions, yielding a total of 750 judged images apiece. The images were generated after the degradation and fusion of 25 pairs of pristine LWIR and visible gray-level images from the TNO and MORRIS databases. The distortions used were AWGN, NU, and blur, using the same degradation parameters described in subsection 1.1; and the fusion algorithms applied were average, gradient pyramid, and SIDWT. The test procedure was conducted following the recommendations in [30]. Each testing session was conducted under the same conditions and using the same equipment. The obtained subjective scores where converted to difference scores (between the pristine and distorted), later to Z-scores [16, 30] and finally to Difference Mean Opinion Scores (DMOS) for each distorted image.

3.2. Opinion Aware Fused Image Quality Analyzer

Our opinion aware (but otherwise blind) model was created by training on the aforementioned human subjective quality judgments of the images. To do this we employed a Support Vector Regression (SVR) algorithm to fit the NSS features to the DMOS, thereby obtaining a trained opinion aware quality model Q_{SVM}. This method has been previously applied to IQA using NSS-based features [16, 18].

To verify the performance of our model, we compared the

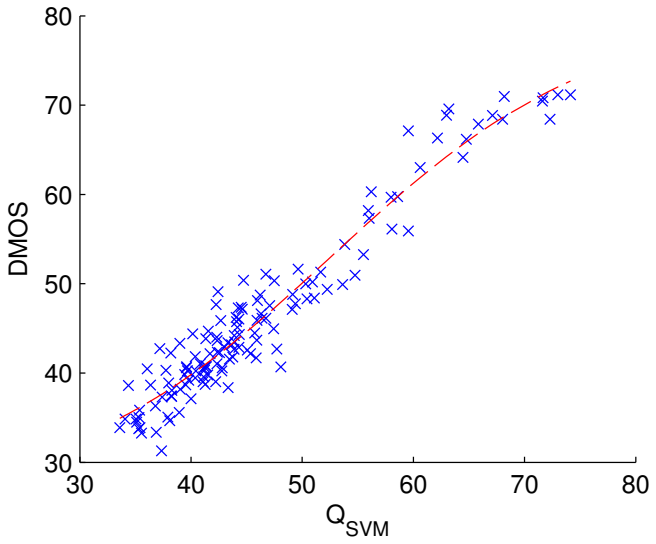

Fig. 3. Scatter plot of Q_{SVM} prediction scores versus the DMOS.

subjective scores obtained in the previous section to the fusion quality model predictions shown in Table 1, which were previously studied by Liu et al in [11]. Since Q_{SVM} requires a training procedure to calibrate, we divided the data from the subjective study into two random subsets, where 80% of the fused images and associated DMOS were used for training and 20% for testing, taking care not to overlap the train and test content. In order to account for a possible non linear relationship between predicted and actual scores, the algorithm scores were passed through a logistic function.

Table 1. Description of the fusion performance models studied in [11]

Model	Description
Q_{NCIE}	Nonlinear Correlation Information Entropy [6]
Q_P	Image Fusion Metric-Based on Phase Congruency [7]
Q_S	Piella's Metric [8]
Q_{CV}	Chen-Varshney Metric [9]
Q_{CB}	Chen-Blum Metric [10]

We repeated this process over 1000 iterations, computed SRCC, LCC, and RMSE for all models, and tabulated their median values in Table 2. Figure 3 depicts a scatter plot of the predicted scores delivered by our quality model Q_{SVM} versus DMOS for all the images evaluated in the subjective study, along with the best-fitting logistic function. Observe that the model Q_{SVM} has the highest correlation with human scores, while the other models yielded lower correlation.

Table 2. Median SRCC, LCC, and RMSE between DMOS and predicted DMOS measured over 1000 iterations

Model	SRCC	LCC	RMSE
Q_{NCIE}	0.194	0.270	9.713
Q_P	0.066	0.209	9.854
Q_S	0.267	0.303	9.614
Q_{CV}	0.042	0.050	10.067
Q_{CB}	0.070	0.079	10.048
Q_{SVM}	**0.932**	**0.961**	**2.813**

4. CONCLUSION

We found that fused LWIR-visible images created with multiresolution fusion algorithms such as Average, Gradient Pyramid, and Shift Invariant Discrete Wavelet Transform, possess statistical regularities when band-pass filtered and divisively normalized, and that these regularities can be modeled and used to characterize distortions and to predict fused image quality. NSS play an important role when analyzing distortions present in fused of LWIR and visible light images, as they have previously proved useful in modeling degradations of the visible and infrared pictures. We found that NSS are also potent descriptors of the quality of fused images affected by AWGN and NU. Therefore, we proposed an OA fused image quality analyzer that outperforms current fusion quality indexes, correlating better with human subjective evaluations.

5. REFERENCES

[1] J. Lee and A. Bovik, *The Essential Guide to Video Processing*, chapter 19 Video surveillance, pp. 619–649, Elsevier, 2009.

[2] A. Benkhalil, S. Ipson, and W. Booth, "A real-time video surveillance system using a field programmable gate array," *Int. J. Imaging Sys.Technol.*, vol. 11, pp. 130–137, 2000.

[3] G.L. Foresti, "A real-time system for video surveillance of unattended outdoor environments," *IEEE Trans. Circ. Sys. Video Technol*, vol. 8, pp. 697–704, 1998.

[4] Marie Freebody, "Consumers and cost are driving infrared imagers into new markets," *Photonics Spectra*, vol. 49, no. 4, pp. 40–44, april 2015.

[5] A. Torabi, G. Massé, and G. A Bilodeau, "An iterative integrated framework for thermal–visible image registration, sensor fusion, and people tracking for video surveillance applications," *Computer Vision and Image Understanding*, vol. 116, pp. 210–221, 2012.

[6] Q. Wang, Y. Shen, and J. Jin, *Performance evaluation of image fusion techniques*, chapter 19, pp. 469–492, Elsevier, 2008.

[7] J. Zhao, R. Laganiere, and Z. Liu, "Performance assessment of combinative pixel-level image fusion based on an absolute feature measurement," *International Journal of Innovative Computing, Information and Control*, vol. 3, no. 6, pp. 1433–1447, Dec. 2007.

[8] G. Piella and H. Heijmans, "A new quality metric for image fusion," in *Proc. International Conference on Image Processing*, 2003.

[9] H. Chen and P. K. Varshney, "A human perception inspired quality metric for image fusion based on regional information," *Information Fusion*, vol. 8, pp. 193–207, 2007.

[10] Y. Chen and R. S. Blum, "A new automated quality assessment algorithm for image fusion," *Image and Vision Computing*, vol. 27, no. 10, pp. 1421–1432, 2009.

[11] Z. Liu, Senior Member, E. Blasch, and Z. Xue, "Objective Assessment of Multiresolution Image Fusion Algorithms for Context Enhancement in Night Vision : A Comparative Study," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 94–109, 2012.

[12] F. Yuming, M. Kede, W. Zhou, L. Weisi, F. Zhijun, and Z. Guangtao, "No-reference quality assessment of contrast distorted images based on natural scene statistics," *IEEE Signal Process. Lett*, vol. 22, no. 7, pp. 838–842, 2015.

[13] A. C. Bovik, "Automatic prediction of perceptual image and video quality," *Proceedings of the IEEE*, vol. 101, no. 9, pp. 2008–2024, Sept. 2013.

[14] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: from natural scene statistics to perceptual quality.," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350–64, 2011.

[15] C. C. Su, A. C. Bovik, and L. K. Cormack, "Natural scene statistics of color and range," in *18th IEEE International Conference on Image Processing (ICIP)*, 2011, pp. 257–260.

[16] T. Goodall, A. C. Bovik, and N. G. Paulter, "Tasking on natural statistics of infrared images," *IEEE Transactions on Image Processing*, pp. 1–17, 2015.

[17] N. J. W. Morris, S. Avidan, W. Matusik, and H. Pfister, "Statistics of infrared images," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2007, pp. 1–7.

[18] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–708, 2012.

[19] N. Rajic, *Noise in Infrared Thermography*, chapter 5, 2001.

[20] J. E. Pezoa and O. J. Medina, "Spectral model for fixed-pattern-noise in infrared focal-plane arrays," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. 2011, pp. 55–63, Springer.

[21] J. W. Davis and M. A. Keck, "A Two-Stage Template Approach to Person Detection in Thermal Imagery," in *2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05)*, 2005, pp. 364–369.

[22] A. Toet, J. K IJspeert, A. M. Waxman, and M. Aguilar, "Fusion of visible and thermal imagery improves situational awareness," *Displays*, vol. 18, no. 2, pp. 85–95, 1997.

[23] R. S. Blum and Z. Liu, *Multi-sensor Image Fusion and its Applications*, CRC press, 2005.

[24] P.J. Burt and R.J. Kolczynski, "Enhanced image capture through fusion," *ICCV*, pp. 173 – 182, 1993.

[25] O. Rockinger and T. Fechner, "Pixel-level Image Fusion: The Case of Image Sequences," *Proceeding of SPIE*, vol. 3374, pp. 378–388, 1998.

[26] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.

[27] D. L. Ruderman, "The statistics of natural images," *Network: Computation in Neural Systems*, vol. 5, no. 4, pp. 517–548, 1994.

[28] Y. Zhang and D. M. Chandler, "An algorithm for no-reference image quality assessment based on log-derivative statistics of natural scenes," *Image Quality and System Performance*, vol. 8653, pp. 86530J, 2013.

[29] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized gaussian distributions in subband decompositions of video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, no. 1, pp. 52–56, 1995.

[30] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1427–1441, 2010.