Tasking on Natural Statistics of Infrared Images

Todd Goodall, Alan C. Bovik, Fellow, IEEE, and Nicholas G. Paulter Jr., Fellow, IEEE

Abstract—Natural Scene Statistics (NSS) provide powerful perceptually relevant tools that have been successfully used for image quality analysis of visible light images. NSS capture statistical regularities that arise in the physical world and thus are relevant to Long Wave Infrared (LWIR) images. LWIR images are similar to visible light images and mainly differ by the wavelengths captured by the sensors. The distortions unique to LWIR are of particular interest to current researchers. We analyze a few common LWIR distortions and how they relate to NSS models.

Humans are the most important factor for assessing distortion and quality in IR images, which are often used in perceptual tasks. Therefore, predicting human performance when a task involving LWIR images needs to be performed can be critical to improving task efficacy. The National Institute for Standards and Technology (NIST) characterizes human Targeting Task Performance (TTP) by asking firefighters to identify the locations of fire hazards in LWIR images under distorted conditions. We find that task performance can be predicted using NSS features. We also report the results of a human study.

We analyzed the NSS of LWIR images under pristine and distorted conditions using four databases of LWIR images. Each database was captured with a different camera allowing us to better evaluate the statistics of LWIR images independent of camera model. We find that models of NSS are also effective for measuring distortions in the presence of other independent distortions.

Index Terms—NSS, LWIR, NIST, TTP

I. INTRODUCTION

LONG Wavelength Infrared (LWIR) images have many uses in industry, military, medicine, and science. Non-destructive testing uses thermal imagers for detecting defect locations in manufactured materials, thereby allowing for better quality control [1]. Unmanned Airborne Vehicles (UAV) and security cameras often couple a thermal imager with a visible light camera to enhance night vision for scouting and to improve automatic threat detection over large distances [2]. Firefighters carry handheld imagers while scouting for critical burn points in burning buildings [3] [4] [5]. Thermographers use high-resolution thermal imagers for detecting inflammation, irregular blood-flow, and tumors [6].

A broad theme of this paper is the development and practical application of Natural Scene Statistics (NSS) models of LWIR images. NSS models describe statistical regularities that are observed on images taken of the natural world.[1] Examples of NSS of visible light images include the $\frac{1}{f}$ behavior of the amplitude spectrum [7] [8], the sparse coding characteristic of visual cortical filters in response to natural image stimuli [9], and the Gaussian exhibited by visual signals following band-pass filter and adaptive gain control operations [10]. Early cortical processing in higher mammalian visual systems appears to have adapted to these natural statistics [7], and much research into biological visual functioning has been guided by the "efficient coding" hypothesis which assumes that visual neurons have adapted to efficiently encode natural visual stimuli [11] [12].

Given their widespread use and application, LWIR images have been well studied. Mooney characterized sources of spatial noise [13] and the effect of noise on minimum resolvable temperature differences (MTD) as a function of frequency [14]. Lopez-Alonso further characterized spatial noise in IR images by using Principle Components Analysis (PCA) to separate spatial and temporal noise from a sequence of frames [15]. This led Pezoa and Medina to model the Non-Uniformity (NU) noise common in LWIR images in the frequency domain [16] as distinct from independent spatial noise. Using the NU model in place, Pérez *et al.* measured and compared the efficacy of several Non-Uniformity correction (NUC) algorithms [17] and developed methods for extracting the structure of the underlying fixed-pattern noise (FPN) [18].

Although NSS have proven to be highly successful tools in applications on visible light images, the development and use of similar models has not been nearly as widespread on LWIR images. Morris *et al.* compared LWIR image statistics with natural visible light image statistics, and found that the spectral power of LWIR images is more "heavy-tailed" and that LWIR wavelet histograms are generally peakier, likely due to the smooth behavior of infrared images. Kaser [19] and Goodall [20] modeled the fit of the BRISQUE [21] and NIQE [22] image quality models to LWIR images, and showed that these visible light models fit reasonably well to LWIR image data. To measure NU, noise, blur, and changes in brightness, Amon *et al.* developed four Image Quality Indicators (IQIs) [4] [5] which include making measurements on the amplitude spectrum. To measure NU in LWIR images, Hayat *et al.* introduced a Roughness Index computed from discrete spatial derivatives, while Martina *et al.* produced an improved version of this index called the Effective Roughness Index [23] which measures spatial derivatives computed on a high-pass image. To our knowledge, we have described the extent of previous work in modeling or utilizing NSS on LWIR images.

LWIR images are certainly 'natural' in the sense that we use the term, and understanding and modeling the NSS of LWIR images has the potential to underlie new algorithms for both identifying distortions present in any given LWIR image and for enhancing images to reduce the degree of distortion. Important distortions that we study here include Non-Uniformity (NU) noise, ferro-electric "Halo Effects," sensor noise, JPEG artifacts, blurring, and hotspots.

---

[1]In essence, captured photographically, of any real-world scenes, including both man-made and natural objects.

(a) Non-Uniformity   (b) "Halo Effect"   (c) Hotspots
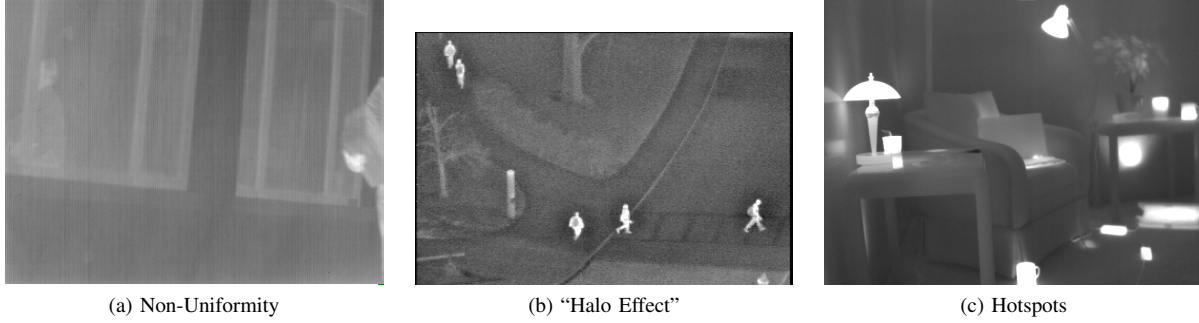
Fig. 1. Example artifacts of distorted LWIR images

A. Characteristics of LWIR Imagers and Images

Thermal imagers have a spectral sensitivity to the 7 μm to 14 μm wavelength band, although long wavelength infrared refers to wavelengths in the electromagnetic spectrum from around 4 μm to approximately 15 μm. Thermal imagers are sensitive to the radiation emitted by objects in the scene and the background radiation reflected by the objects. The relationship between the irradiation collected at the sensor and the temperatures of imaged materials is nontrivial. For example, the emissivity of different materials varies, and the surface properties of an object alter its emissivity. Similarly, the reflective properties of an object will vary the amount of background radiation reflected by the object and subsequently collected by the imager. All this variability can lead to errors in the measurement of an object's temperature.

Care must be taken when selecting a thermal imager for a given application. Two classes of thermal imagers exist - cooled and uncooled. Cooled detector imagers require expensive and bulky helium cryocoolers that need regular maintenance. Internal thermally-induced noise reduces the sensitivity of the sensors and the coolers regulate the temperature close to the phase-transition temperature. Currently, cooled imagers are most cost-effective for long range (≥ 5 km) surveillance imaging given their detector sensitivity, and they can be used to capture image data at high frame-rates as a result of modifiable integration time [24].

Uncooled thermal imagers are most popular. Also known as Focal Plane Array (FPA) imagers, they use either an array of resistors (called a microbolometer) or an array of ferro-electric ceramic transistors. The microbolometer works by measuring large changes in resistance corresponding to small changes in temperature. Unlike cooled detectors, uncooled imagers cannot modify their sensor integration time thus they usually capture images at a lower frame rate. The ferro-electric technology operates by measuring a temperature differential across a pyro-electric material which is refreshed by a constantly spinning wheel, called the "chopper." As a result of the "chopper," the images obtained by these detectors exhibit additional artifacts such as the "Halo Effect" which is lowering their adoption in the market. Overall, the main advantage of uncooled imagers over their cooled counterparts is their lower power consumption, cheaper components, and size [25] [26].

Non-Uniformity (NU) noise as exemplified in the image in Fig. 1(a) is a distortion specific to LWIR images. NU is an additive FPN which appears as a grid-like or striping pattern. These patterns result from manufacturing defects, dark current, and segmented sensor capture areas [16] [27] [15].

The "Halo Effect," depicted in Fig. 1(b), is another distortion which occurs mostly in thermal cameras equipped with ferro-electric sensors. This effect causes the region surrounding a bright object to grow darker and it causes the region around dark objects to grow lighter [28] [25]. This effect can be caused by both the physical operation of cameras containing ferro-electric sensors and back-reflection of IR illumination sources. The "chopper" which modulates the signal for ferro-electric detectors fails to entirely shield incoming infrared light which leads to overcompensation when subtracting the differential response from the average signal producing a halo. Reflective materials situated next to highly emissive materials within the scene have also been shown to produce a similar effect [29].

LWIR images commonly contain hot-spot areas exhibiting only small variations in image intensity, arising from highly emissive objects which stand out from the background as in Fig. 1(c). In general, LWIR images contain many smooth surfaces as a result of temperature diffusion. Hot-spots are less a distortion than a symptom of the environment, but they still produce interesting statistical regularities worthy of study.

Other unique degradations of LWIR images not covered later but still important for later study include radiometric distortions, geometric distortions, noise from reflective materials, and the history effect. Radiometric distortion refers to non-linear mapping of thermal energy to pixel values in an image which may destroy relevant sensitivity information. Geometric distortions occur when the sensors in the FPA are mis-aligned, causing blur. As discussed previously, materials imaged with an infrared camera are often assumed to be only emissive, but they can also be reflective which can produce false inferences. Lastly, heat in the Long Wavelength band can fluctuate faster than frame rate, which can be difficult to detect given the physical limits of infrared sensors. Geometric distortions are specific to individual imagers, radiometric distortions appear during the capture process, reflective noise measurements require knowledge of the captured objects, and the history effect is a time-varying distortion. These distortions and accompanying side information are not currently available for study thus they are not included here.

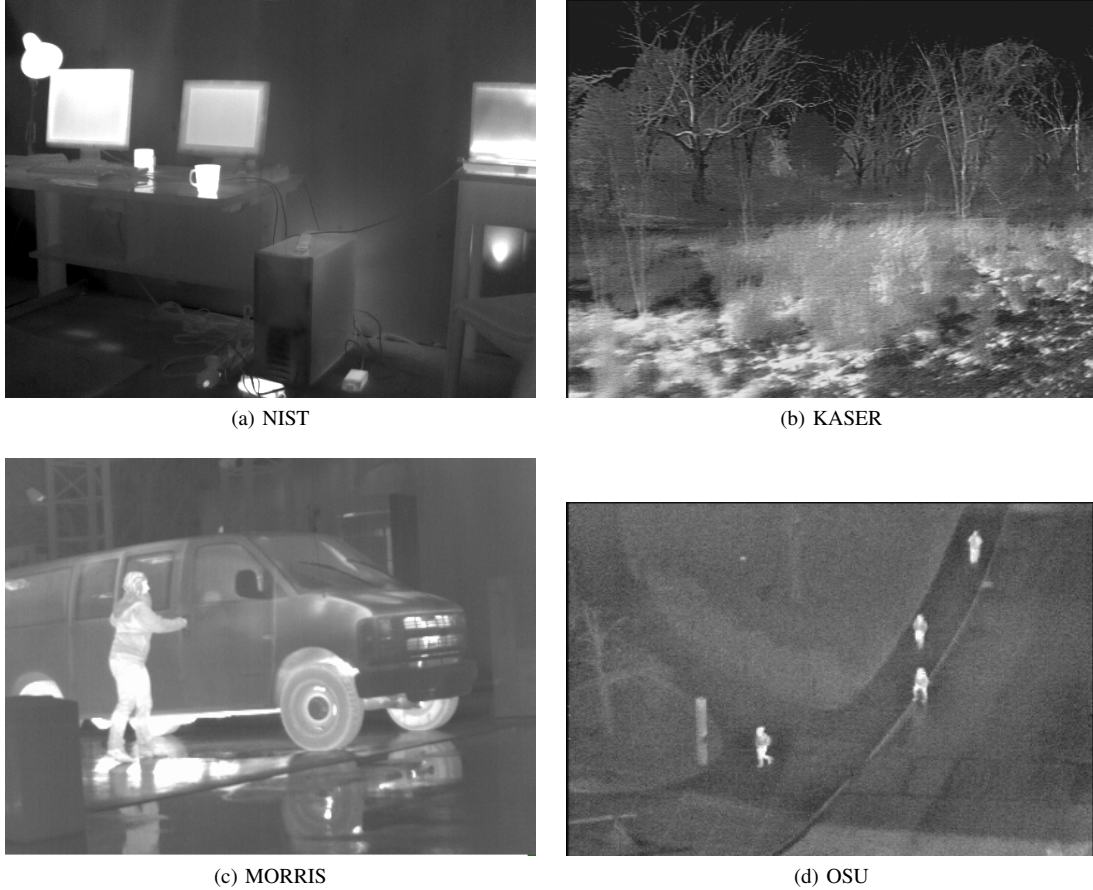The same distortions that occur in visible light images can

(a) NIST

(b) KASER

(c) MORRIS

(d) OSU

Fig. 2.   Example images from 4 LWIR databases

TABLE I
CAMERAS ASSOCIATED WITH THE 4 LWIR DATABASES

Database	Detector Material	Lens Diameter (mm)	Bit Depth	Resolution	Sensitivity (mK)
OSU	Ferro-Electric	75	8	360x240	100
MORRIS	Microbolometer	50	8	384x288	60
NIST	Microbolometer	10	14	640x480	55
KASER	unknown	unknown	8	640x480	unknown

of course also occur in LWIR images. For example, blur may arise from camera destabilization, especially in handheld devices, non-optimal lens focus moving the depth of field away from the object of interest, or object motion. Sensor noise may be induced by light sensitivity based on the integration times of the sensors. Over and under-exposure can occur as a function of exposure time, or from quantization or local saturation. JPEG distortion such as blocking artifacts and blurring can also be present, since most thermal cameras utilize the JPEG compression format.

B. LWIR Image Sources

Our study of the NSS of LWIR images has benefited from the availability of four separate IR image databases which we denote as NIST [30], KASER [19], MORRIS [31], and OSU [32]. Example images from each database are provided in Fig. 2. The NIST database includes 180 images from indoor office and home environments each containing multiple

hot objects and some containing fire hazards. The KASER database includes 37 images from outdoor environments taken using unknown camera models and suffering various distortions including Non-Uniformity, blur, and noise. The MORRIS database contains both indoor and outdoor images of urban environments including cars, pedestrians, and buildings. Finally, the OSU database contains images captured by a surveillance camera monitoring pathway intersections on the Ohio State University campus. Gathering a diverse set of images from a diversity of cameras allows for better generalization of the NSS of LWIR images.

The uncooled thermal cameras associated with each database are listed in Table I. Images from the microbolometer sensor types appear smoother and cleaner than the images from the ferro-electric sensor type used in OSU. Images obtained from this camera required processing by an additional non-linearity (using a log transform) in order that the NSS followed the same regularities observed in the images obtained from the
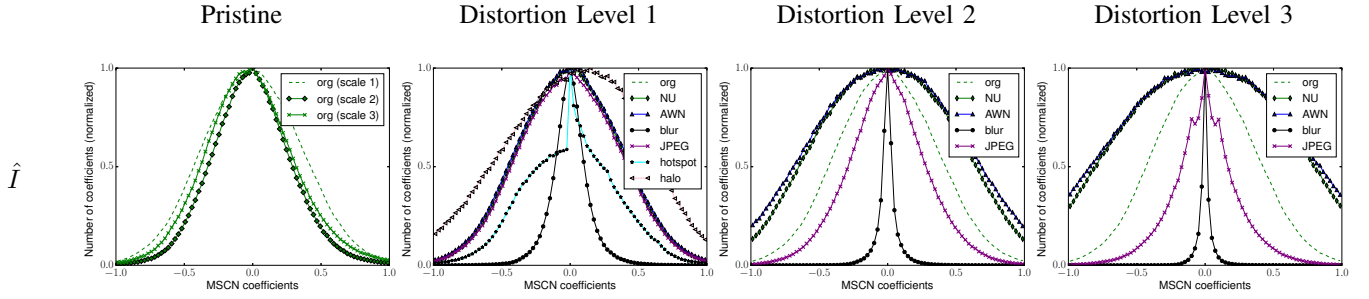
Fig. 3. MSCN histograms of center patches extracted from both NIST and MORRIS LWIR image databases. The left-most figure depicts the scale-invariant behavior of the MSCN statistics of non-distorted images. The remaining three figures depict increasing levels of distortion from left to right at the first scale. The terms org, NU, AWN, blur, JPEG, hotspot, and halo refer to pristine images, images with NU distortion, images with AWN distortion, images with blur distortion, images compressed with JPEG, hotspot image patches, and halo image patches, respectively.

other cameras. This non-linearity may be a result of the values being captured at the ferro-electric sensors being proportional to a non-linear function of luminance, like power.

All images were linearly mapped from their bit depth to the range 0 to 1 for comparability and ease of applying artificial distortions consistently. This does not change the image statistics beyond normalizing them to this range.

II. NSS OF LWIR IMAGES

A. Processing Model

In a pioneering deep study of the statistics of visible light images, Ruderman observed that applying a local bandpass filter combined with a non-linear operation to a natural image has a decorrelating and gaussianizing effect [10]. Highly successful Image Quality Assessment (IQA) models have used this property to measure distortions in images. Given an input luminance image, I, define its Mean-Subtracted Contrast Normalized (MSCN) coefficients

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

over spatial indices with $i \in 1, 2...M, j \in 1, 2...N$, where M and N are the image height and width, respectively, C is a constant which prevents instabilities when the denominator tends toward zero. The factors μ and σ are weighted estimates of the local luminance mean and standard deviation given by

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

and

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

where $w = \{w_{k,l} | k = -K, \cdots, K, l = -L, \cdots, L\}$ is a 2D circularly-symmetric weighting function sampled out to 3 standard deviations and normalized to unit volume.

The MSCN histograms of natural LWIR images appear similar (Gaussian) to those of visible light images, as shown in Fig. 3. To compute these histograms, coefficients were pooled by selecting center patches from images taken from the NIST and MORRIS databases.

The histograms of the pooled MSCN coefficients selected from center patches of LWIR images afflicted by three levels (severities) of common distortions (NU, AWN, blur, and JPEG) are compared in Fig. 3. Only one distortion level for hotspot and halo types is available.

In the Blind/Referenceless Image Spatial QUality Evaluator (BRISQUE) model [21], the MSCN histograms are supplemented by empirical paired product distributions which are computed by multiplying neighboring MSCN coefficients. Four directional coefficient products are computed at each coordinate

$$
\begin{aligned}
H(i,j) &= \hat{I}(i,j)\hat{I}(i,j+1) \\
V(i,j) &= \hat{I}(i,j)\hat{I}(i+1,j) \\
D1(i,j) &= \hat{I}(i,j)\hat{I}(i+1,j+1) \\
D2(i,j) &= \hat{I}(i,j)\hat{I}(i+1,j-1)
\end{aligned}
$$

The center patches of images from the NIST and MORRIS databases were used to compute the paired product histograms of both distortion-free and distorted images over multiple distortion levels as tabulated in Fig.4.

In an interesting extension of BRISQUE called the Derivative Statistics-based QUality Evaluator (DESIQUE) model [33], the MSCN histograms are supplemented by seven log-derivative distributions that are computed by differencing the logarithms of the intensities of neighboring pixel values. The following function is defined

$$J(i,j) = \log\left(\hat{I}(i,j) + K\right)$$

where K is a stabilizing constant, and the log-derivative coefficients are computed as

$$
\begin{aligned}
PD1(i,j) &= J(i,j+1) - J(i,j) \\
PD2(i,j) &= J(i+1,j) - J(i,j) \\
PD3(i,j) &= J(i+1,j+1) - J(i,j) \\
PD4(i,j) &= J(i+1,j-1) - J(i,j) \\
PD5(i,j) &= J(i-1,j) - J(i+1,j) \\
&\quad -J(i,j-1) - J(i,j+1) \\
PD6(i,j) &= J(i,j) + J(i+1,j+1) \\
&\quad -J(i,j+1) - J(i+1,j) \\
PD7(i,j) &= J(i-1,j-1) + J(i+1,j+1) \\
&\quad -J(i-1,j+1) - J(i+1,j-1)
\end{aligned}
$$

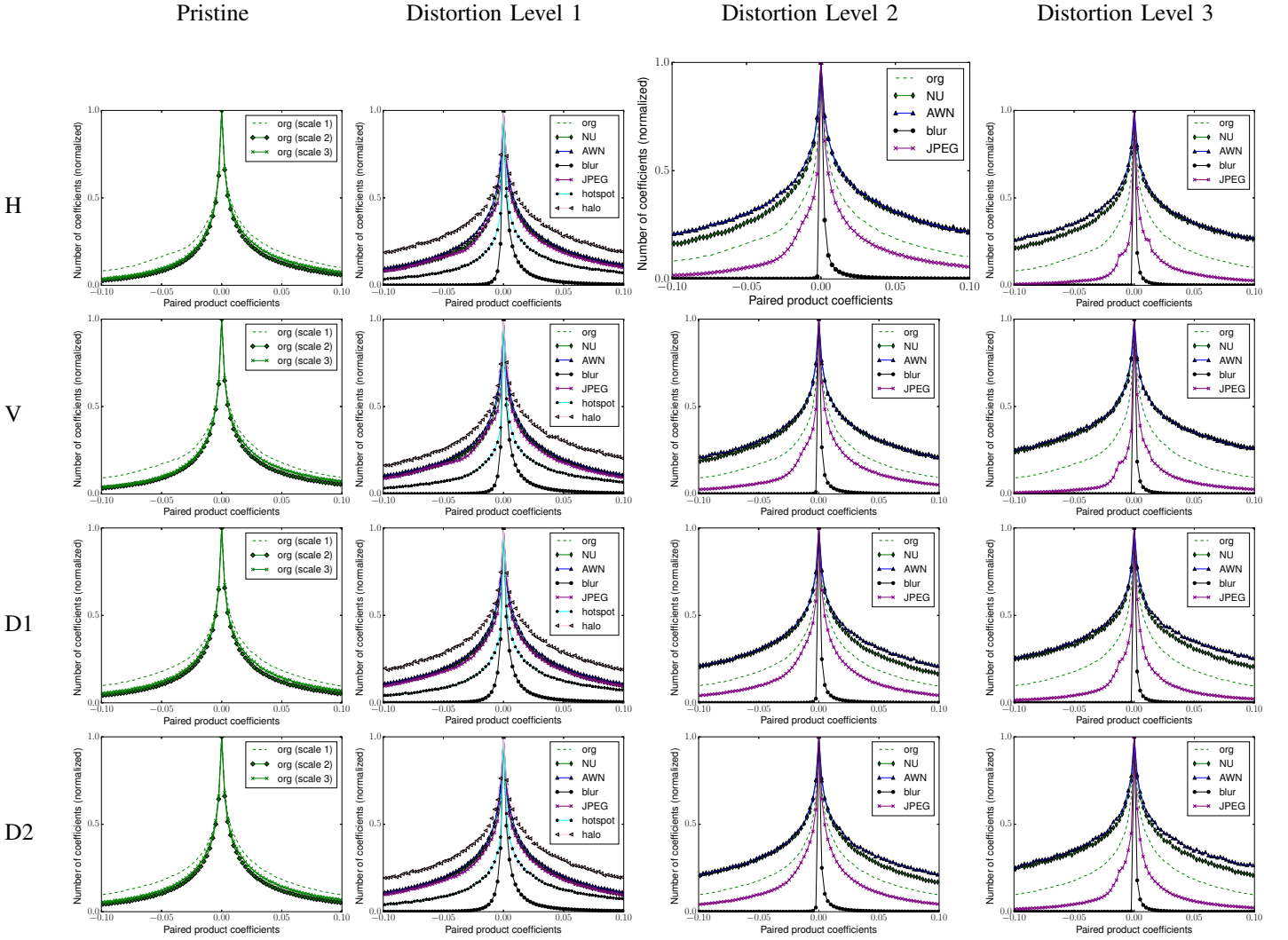The Log-Derivative distributions of both pristine and dis-

Fig. 4. Paired product histograms of center patches extracted from both NIST and MORRIS LWIR image databases. The left-most column depicts the scale-invariant behavior of paired products extracted from non-distorted images. The remaining 3 columns depict increasing levels of distortion from left to right at the first scale. The terms org, NU, AWN, blur, JPEG, hotspot, and halo refer to pristine images, images with NU distortion, images with AWN distortion, images with blur distortion, images compressed with JPEG, hotspot image patches, and halo image patches, respectively.

torted images over multiple distortion levels are plotted in Fig. 5.

Perceptual neurons in the early stages of the human visual system form responses that capture information over multiple orientations and scales. These responses have been successfully approximated by steerable filters, with the steerable pyramid decomposition being most popular [34] [35] [36].

The Distortion Identification-based Image Verity and INtegrity Evaluation (DIIVINE) [36] index predicts image quality using coefficients generated from the steerable pyramid overcomplete wavelet decomposition. Oriented image subbands are divisively normalized by dividing the local contrast estimated from neighboring subbands and scales.

The divisively normalized steerable pyramid orientation subbands for center patches extracted from one scale and six orientations for both distortion-free and distorted images are plotted in Fig. 6. Each band is denoted d_α^θ where α denotes the level and $\theta \in \{0°, 30°, 60°, 90°, 120°, 150°\}$.

B. Distortion Models

We next describe the generative noise models used to create distorted LWIR images. Pezoa and Medina developed a model of Non-Uniformity which can be used to artificially distort pristine images [16]. Based on a spectral analysis of NU, they proposed the model

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

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

where $\tilde{I}$ is the Fourier Transform representation of the noise image, $B_u = B_v = 5.2$, $\sigma_u = \sigma_v = 2.5$, and where $U[a,b]$ denotes the uniform distribution on $[a,b]$. The severity of NU can be controlled by scaling the dynamic range using
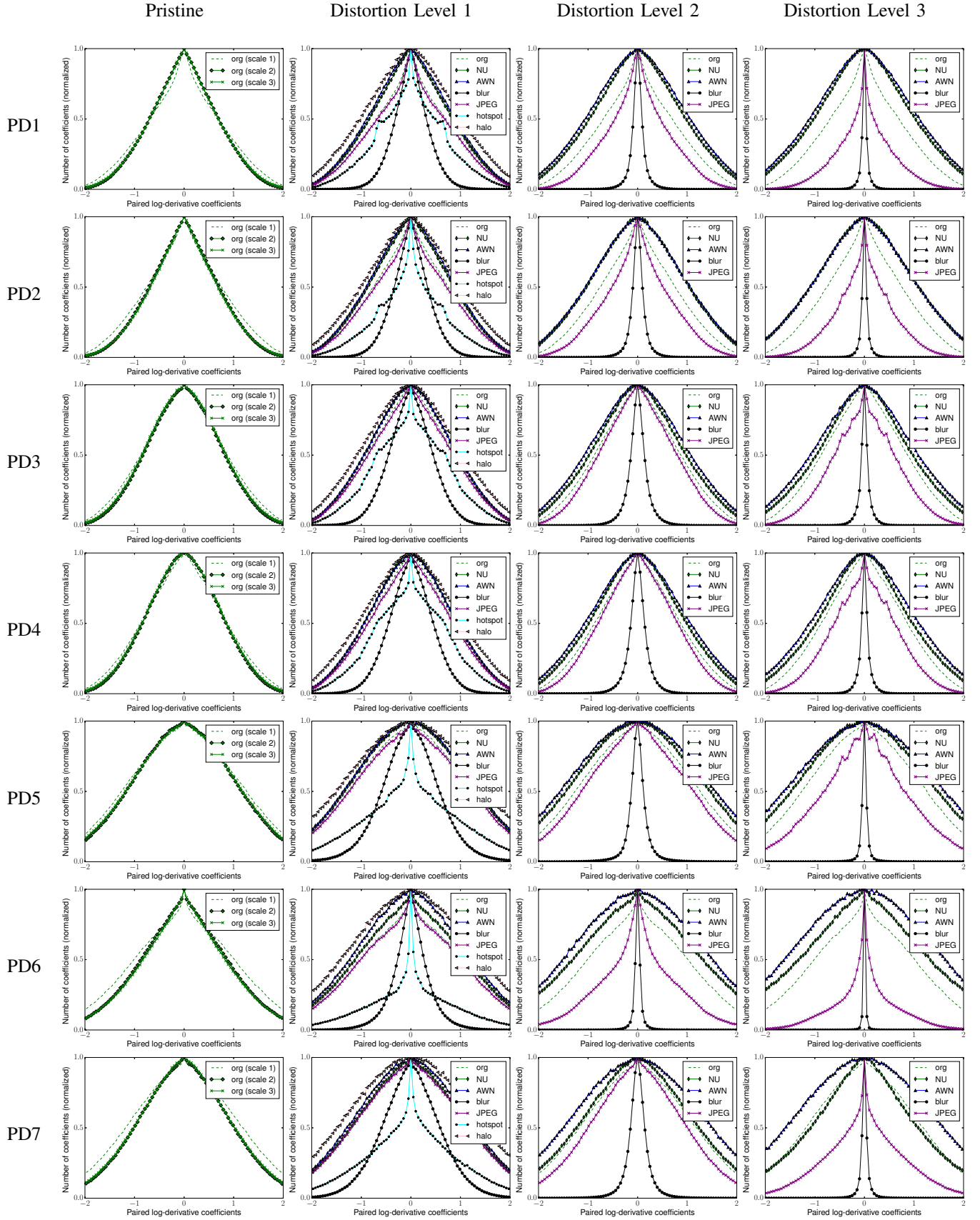
Fig. 5. Paired Log-Derivative histograms of center patches extracted from both NIST and MORRIS LWIR image databases. The left-most column depicts the scale-invariant behavior of the Log-Derivative coefficients extracted from non-distorted images. The remaining 3 columns depict increasing levels of distortion from left to right at the first scale. The terms org, NU, AWN, blur, JPEG, hotspot, and halo refer to pristine images, images with NU distortion, images with AWN distortion, images with blur distortion, images compressed with JPEG, hotspot image patches, and halo image patches, respectively.
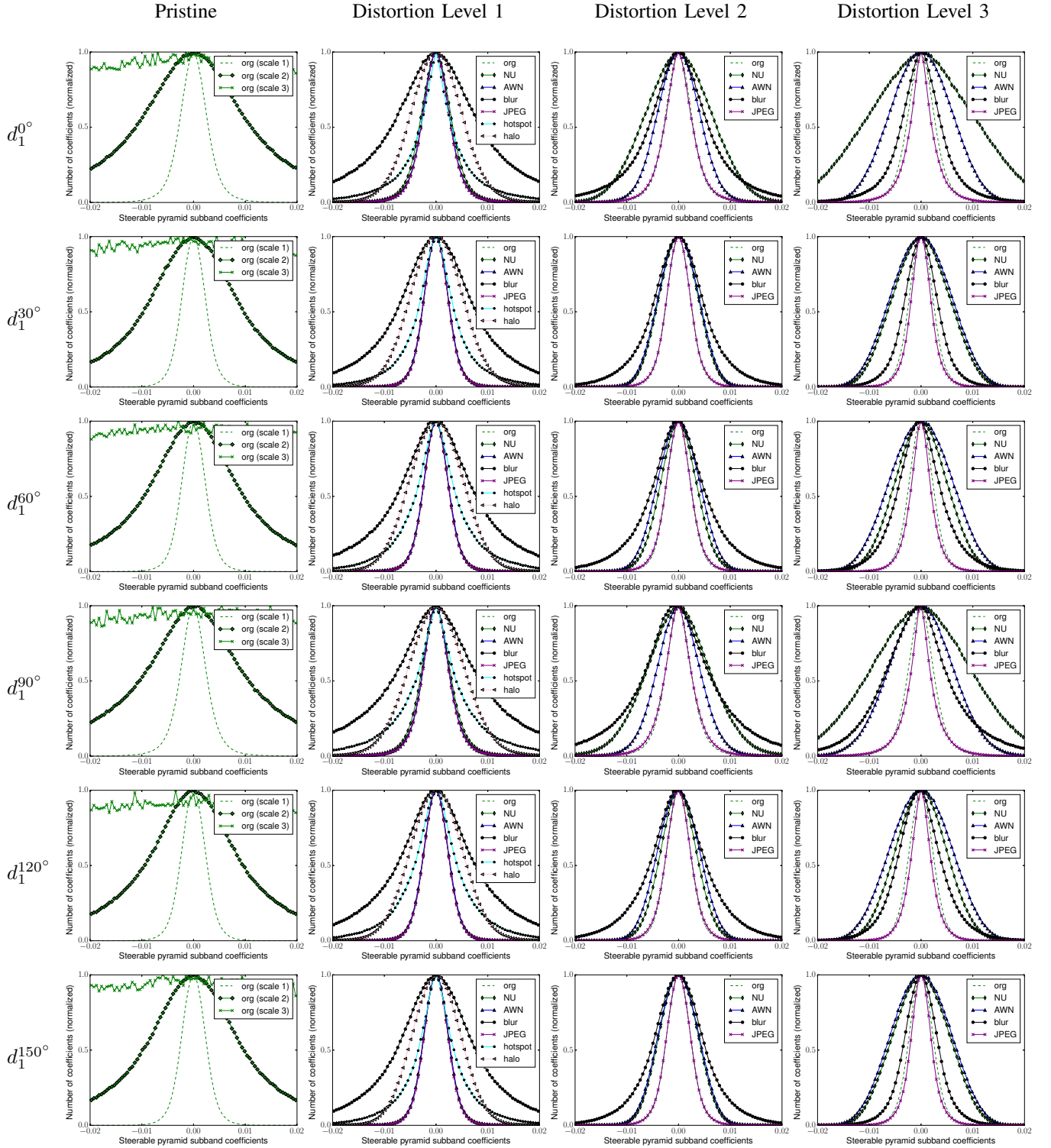
Fig. 6. Steerable Pyramid histograms of center patches extracted from both NIST and MORRIS LWIR image databases and divisively normalized. Each of the six orientations are shown. The left-most column depicting changes in standard deviation of the coefficients across scales for non-distorted images. The remaining three columns indicate a qualitative change in histogram shape for changes in distortion level. The terms org, NU, AWN, blur, JPEG, hotspot, and halo refer to pristine images, images with NU distortion, images with AWN distortion, images with blur distortion, images compressed with JPEG, hotspot image patches, and halo image patches, respectively.

a standard deviation parameter σ_{NU}. Levels 1-3 of distortion were generated by setting $\sigma_{\mathrm{NU}} = \{0.0025, 0.01375, 0.025\}$. The marginal histograms of images distorted by NU post-processed by MSCN, paired products, paired log-derivatives,

and steerable pyramid subbands are depicted in Figs. 3, 4, 5, and 6. These histograms match AWN behavior in MSCN histograms, appear to differ with respect to AWN asymmetrically in paired product histograms, appears distinctive in the log-

derivative histograms, and has a distinctively large standard deviation in the horizontal and vertical subbands, d_1^0 and d_1^{90}, of the steerable pyramid.

The "Halo Effect" occurs naturally in the images in the OSU database. Davis's method [37], which is based on background subtraction and morphological techniques, was used to isolate moving objects (often people) in the images. Since not all objects extracted using this method exhibited the "Halo Effect," patches with a clear visible "Halo Effect" were isolated by hand. A total of 188 example patches were thus selected from the OSU database for use here. The marginal histograms computed from MSCN coefficients exhibit a slight skew in Fig. 3, the paired product and paired log-derivative coefficients exhibit heavier tails in Figs. 4 and 5, and the steerable pyramid coefficients exhibit fatter histograms as depicted in Fig. 6. These histogram comparisons may not only reflect the "Halo Effect" in isolation since these artifacts are combined with the non-linearity associated with Ferro-Electric sensors.

Hotspots were isolated by hand from the NIST and MORRIS databases. A total of 135 hotspot patches including people, environmental hazards, and other miscellaneous objects were extracted. When comparing to the natural LWIR image histrograms, the hotspot histogram shapes computed using MSCN coefficients demonstrate an asymmetry in Fig. 3, paired product and paired log-derivative coefficients exhibit peakiness in Figs. 4 and 5, and steerable pyramid coefficients exhibit heavier tails in Fig. 6.

JPEG, Additive White Noise (AWN), and blur distortions were compared using the same set of images drawn from the NIST and MORRIS databases. JPEG images were generated at the 100, 90, and 80 percent quality settings corresponding to distortion levels 1, 2, and 3 producing average respective bit rates of 3.6, 1.0, and 0.5 bpp. Distortion levels involving Gaussian white noise matched the levels of NU mentioned previously for comparability, using $\sigma_{\text{AWN}} = \{0.0025, 0.01375, 0.025\}$ (recall the gray-scale range is $[0, 1]$). Blur was generated with a Gaussian blur kernel with scale parameters $\sigma_b = \{1, 2, 3\}$.

JPEG distortions cause the MSCN, paired product, paired log-derivative, and steerable pyramid histograms to become narrower. These same histograms for AWN become wider. Blur distortion histograms become narrower as in JPEG, with the exception of the steerable pyramid histograms.

C. Feature Models

A parametric General Gaussian Distribution (GGD) [38] has been used to model the MSCN, Paired Log-Derivative, and steerable pyramid subband coefficients. The associated GGD probability density function is

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

where

$$\Gamma(x) = \int_0^\infty s^{x-1}e^{-s}ds.$$

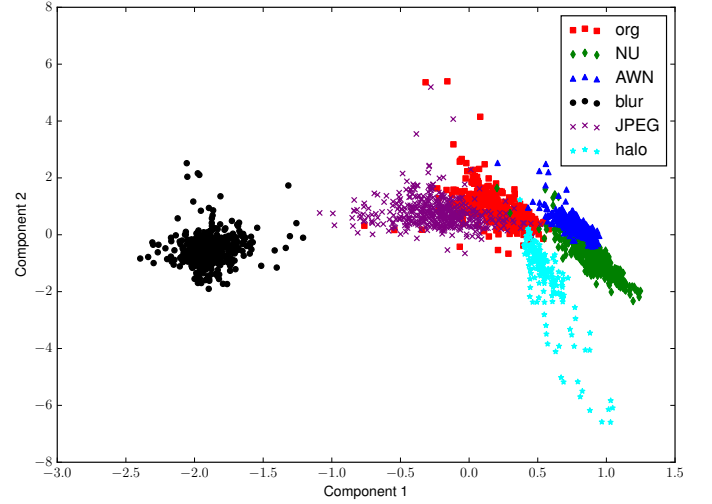An Asymmetric Gaussian Distribution (AGGD) [39] has

Fig. 7. A total of 46 features over 3 scales yields 138 features per image, projected here into 2 dimensional space using PCA. Even though the total explained variance ratio of top two components is 0.734, distorted images cluster away from the natural images. Note that hotspots were not included in the projection because they significantly produce a sparse distribution likely resulting from the limited size of the image patches.

been used to effectively model to the paired product coefficients. The pdf is

$$f(x; v, \sigma_l^2, \sigma_r^2) = \begin{cases} \frac{v}{(\beta_l+\beta_r)\Gamma(\frac{1}{v})} exp\left(-\left(\frac{-x}{\beta_l}\right)^v\right) & x < 0 \\ \frac{v}{(\beta_l+\beta_r)\Gamma(\frac{1}{v})} exp\left(-\left(\frac{x}{\beta_r}\right)^v\right) & x \geq 0 \end{cases}$$

where β_l and β_r are given by

$$\beta_l = \sigma_l\sqrt{\frac{\Gamma\left(\frac{1}{v}\right)}{\Gamma\left(\frac{3}{v}\right)}}$$

and

$$\beta_r = \sigma_r\sqrt{\frac{\Gamma\left(\frac{1}{v}\right)}{\Gamma\left(\frac{3}{v}\right)}}$$

respectively.

The parameters (α, σ^2) of the GGD model fit can be estimated using the technique described in [38]. The parameters $(v, \sigma_l^2, \sigma_r^2)$ of the AGGD model fits can be estimated using the moment matching technique described in [39]. Another parameter, η, given by

$$\eta = (\beta_r - \beta_l)\frac{\Gamma\left(\frac{2}{v}\right)}{\Gamma\left(\frac{1}{v}\right)}$$

is also computed for each product image using the estimates of the other parameters. Therefore, the best-fit model of each set of paired product coefficients yields 4 features (η, v, σ_l^2, σ_r^2).

Since the hotspot images exhibit asymmetric histograms, negative and positive MSCN coefficients were measured separately. Negative and positive coefficients correspond to the left and right halves of the histograms. Therefore, four parameters (α_l, σ_l^2, α_r, and σ_r^2) were extracted from the MSCN coefficients. The differences in value between the left and right halves, $\alpha_r - \alpha_l$ and $\sigma_r - \sigma_l$, are used to capture the asymmetry.

TABLE II
FEATURE SUMMARY FOR MSCN (f), PAIRWISE PRODUCTS (pp), PAIRED LOG-DERIVATIVES (pd), AND STEERABLE PYRAMID SUBBANDS (sp) FOR THE FIRST SCALE

Feature ID	Feature Description	Computation Procedure
f_1 - f_2	Shape and Variance	GGD fit to MSCN coefficients
f_3 - f_4	Shape and Variance difference	GGD fit to right and left halves of MSCN coefficients
pp_1 - pp_4	Shape, mean, left variance, right variance	AGGD fit to H pairwise products
pp_5 - pp_8	Shape, mean, left variance, right variance	AGGD fit to V pairwise products
pp_9 - pp_{12}	Shape, mean, left variance, right variance	AGGD fit to D1 pairwise products
pp_{13} - pp_{16}	Shape, mean, left variance, right variance	AGGD fit to D2 pairwise products
pd_1 - pd_2	Shape and Variance	GGD fit to PD1 pairwise log-derivative
pd_3 - pd_4	Shape and Variance	GGD fit to PD2 pairwise log-derivative
pd_5 - pd_6	Shape and Variance	GGD fit to PD3 pairwise log-derivative
pd_7 - pd_8	Shape and Variance	GGD fit to PD4 pairwise log-derivative
pd_9 - pd_{10}	Shape and Variance	GGD fit to PD5 pairwise log-derivative
pd_{11} - pd_{12}	Shape and Variance	GGD fit to PD6 pairwise log-derivative
pd_{13} - pd_{14}	Shape and Variance	GGD fit to PD6 pairwise log-derivative
sp_1 - sp_2	Shape and Variance	GGD fit to $d_1^{0^\circ}$ subband
sp_3 - sp_4	Shape and Variance	GGD fit to $d_1^{30^\circ}$ subband
sp_5 - sp_6	Shape and Variance	GGD fit to $d_1^{60^\circ}$ subband
sp_7 - sp_8	Shape and Variance	GGD fit to $d_1^{90^\circ}$ subband
sp_9 - sp_{10}	Shape and Variance	GGD fit to $d_1^{120^\circ}$ subband
sp_{11} - sp_{12}	Shape and Variance	GGD fit to $d_1^{150^\circ}$ subband

A overview of the MSCN (f), paired product (pp), paired log-derivative (pd), and steerable pyramid subband (sp) features is provided in Table II.

To visualize the clustering of the features over three scales, the features for each distortion class were projected into a two-dimensional space using Principle Component Analysis (PCA) as depicted in Fig. 7. The distorted images appear to cluster in this projection which reasonably preserves their class groupings.

A boxplot comparing the features in Table II between pristine LWIR images and pristine visible light images is provided in Fig. 8. A total of 29 pristine visible light images were obtained from the LIVE Image Quality Assessment Database [40] [41] [42]. The MSCN shape parameter, f_1, is not significantly different between visible and LWIR images when using 95 percent confidence intervals. Comparing f_3, we can infer that LWIR images provide more symmetrically shaped MSCN histograms with 95 percent confidence.

The mean parameter, η, for each of the paired product features differs between LWIR and visible light images. Additionally most of the standard deviation parameters, σ_l and σ_r, differ between the modalities. Most shape parameters for paired products do not appear to differ between LWIR and visible light images. By contrast, most of the shapes and standard deviation parameters for pd and sp are significantly different from visible light images. Note that individual parameter differences are bound to exist by chance with a low number of pristine images, but there does seem to be a difference between the two groups overall.

D. NIST Descriptors

Previous work by NIST has produced four Image Quality Indicators (IQIs) [3] [4] [5] which are described as Brightness (B), Contrast (C), Spatial Resolution (SR), and Non-Uniformity ($\widehat{NU}$) defined as
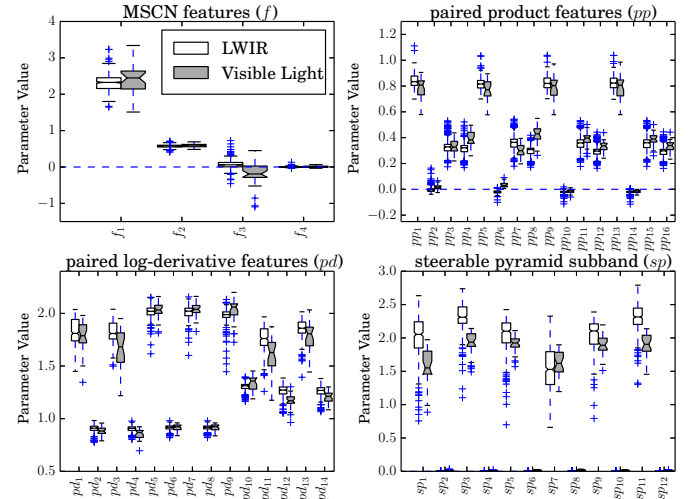

Fig. 8. Box plot comparison of features between natural LWIR and natural visible light images. The notches indicate 95 percent confidence intervals about the median.

- **B** is the average of the luminosity intensities:

$$B = \frac{1}{MN} \sum_{i \in N} \sum_{j \in M} I(i,j)$$

- **C** is defined as RMS contrast:

$$C = \sqrt{\frac{1}{MN} \sum_{i \in N} \sum_{j \in M} (I(i,j) - B)^2}$$

- **SR** (cycles/pixel) is computed by

$$SR = \int_0^{f_c} (MTF_{curve}(u) - NEM) \, du$$

where $MTF_{curve}(u)$ is the modulation transfer function

defined by the Butterworth filter

$$\widetilde{H}(u) = 1/\left(1 + \left(\frac{u}{W_n}\right)^4\right)$$

of order 2. The cutoff frequency is

$$f_c = W_n \left[(1 - \text{NEM})/\text{NEM}\right]^{0.25}$$

where $\text{NEM} = 0.02861$ is the Noise Equivalent Modulation.

- $\widehat{\text{NU}}$ is given by $\widehat{\text{NU}} = \mu/\sigma = \text{B/C}$, the SNR of the image.

As currently defined, the SR statistic, which depends directly on the parameter W_n, is not implementable. This dependency on W_n assumes that any loss of spatial resolution can be modeled based on the response of a Butterworth filter. According to Morris *et al.* [31], the log of the radial spectral power of LWIR images can be well described as following a GGD probability law. Unfortunately, this fit does not generalize when distortions are present in an image, thus a 10th order polynomial approximation was used to yield a much better fit. Overall, the IQIs provide a total of 13 features that are extracted from each image. Unlike the other features, the IQI features are not model based, but rather are sample statistics.

III. TASKING ON NSS

In this section, we study the practical usefulness of the LWIR NSS and IQI features just described for solving five different visual LWIR tasks. First, we use the features to develop a measure of NU on LWIR images. Second, we devise a method to determine presence of the "Halo Effect." The third task is automated prediction of the ability of human experts to detect targets of interest on LWIR images. Fourth, we describe a human study which obtained subjective quality scores on LWIR images, and show that the NSS features are highly predictive of subjective image quality. Lastly, we will show how the LWIR NSS can be used to create localized distortion maps which can aid the identification of local distortions such as hotspots and occurrences of the "Halo Effect."

A. Measuring NU

In NUC algorithms, producing a no-reference estimate of the NU in an image is essential [17]. State-of-the-art methods for estimating the magnitude of NU include the Roughness index, Effective Roughness Index, and SNR. LWIR images commonly contain both fixed pattern noise and additive white noise, and the level of both types of noise should be estimated.

The most common method for estimating NU is the spatial SNR of the image defined as μ/σ where σ and μ are the standard deviation and mean pixel intensities within a user-defined area. Another common and popular method is the Roughness [43] index:

$$\text{Ro}(I) = \frac{\|h_1 * I\|_1 + \|h_2 * I\|_1}{\|I\|_1}$$

where h_1 is the 1-D differencing filter with impulse response $[1, -1]$, $h_2 = h_1{}^T$, and $\|\cdot\|_1$ is the L_1 norm. The Effective
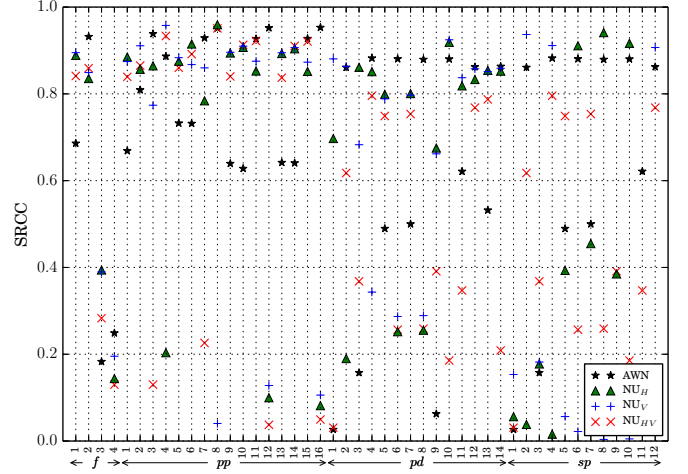

Fig. 9. SRCC of features against two isolated distortions. When correlating features against NU, only the images degraded by NU were used. When correlating features against AWN, only the images degraded by AWN were used.

Roughness [23] index:

$$\text{ERo}(I) = \text{Ro}(g * I)$$

where g is a high-pass filter, with the additional modification that the L_2 norm is used in place of the L_1 norm.

Two weaknesses of current NU estimation approaches are their inability to capture spatial structure and assumption of grid-like patterns of FPN, which, in reality, can often be striping [17]. Additionally, these approaches generally assume that NU is the only distortion within the image. Often, other noise is present that can seriously hinder effectiveness in estimating NU.

A new approach that we have devised to measure the performance of a NUC algorithm utilizes the proposed NSS features listed in Table II. We find that these features are capable of capturing the type of NU, the magnitude of that NU, and the amount of Gaussian white noise that is present.

To compare existing NU estimation techniques, we degraded the images in the MORRIS and NIST databases by taking two samples $\sigma_{\text{NU}}, \sigma_{\text{AWN}} \in \text{U}[0.025, 0.0025]$ where σ_{NU} and σ_{AWN} are the standard deviations of the Non-Uniformity and additive white noise respectively. Three categories of degradations were produced, those with just NU distortions, those with AWN distortions, and those with combined NU and AWN distortions. The two types of noise are thus independently embedded into the signal.

Using these three sets of degraded images, we compared the performances of the state-of-the-art NU metrics. A Support Vector Regressor (SVR) was used to map the features to independently predict NU and AWN on each image. The images in each set were split into non-overlapping subsets: 80 percent for training, and 20 percent for testing. The Spearman Rank Correlation Coefficient (SRCC) was used as a measure of non-linear monotonicity between the actual and predicted values, and (Pearson's) linear correlation coefficient (LCC) was used as a measure of linear correlation between actual and predicted values. Random 80/20 splits were produced and

TABLE III

PREDICTING FOREGROUND AWN WITH BACKGROUND DISTORTION. SRCC AND LCC MEASURED OVER 1000 ITERATIONS USING 80/20 TRAIN/TEST SPLITS. "NONE" INDICATES NO BACKGROUND DISTORTION, $\mathrm{NU_H}$ INDICATES PRESENCE OF HORIZONTAL STRIPING NU BACKGROUND DISTORTIONS, $\mathrm{NU_V}$ INDICATES PRESENCE OF VERTICAL STRIPING NU BACKGROUND DISTORTIONS, AND $\mathrm{NU_{HV}}$ INDICATES PRESENCE OF GRID-LIKE NU BACKGROUND DISTORTIONS. L_1 AND L_2 REFERS TO L_1 AND L_2 NORMS RESPECTIVELY. THE IQIS WERE USED IN PLACE OF SNR BECAUSE SNR ALONE PERFORMED EXTREMELY POORLY.

NR Method	SRCC				LCC			
	None	$\mathrm{NU_H}$	$\mathrm{NU_V}$	$\mathrm{NU_{HV}}$	None	$\mathrm{NU_H}$	$\mathrm{NU_V}$	$\mathrm{NU_{HV}}$
$f + pp + pd + sp$	0.974	0.966	0.964	0.966	0.977	0.969	0.967	0.969
$f + pp + pd$	0.975	0.964	0.966	0.965	0.977	0.967	0.969	0.969
$f + pp$	0.972	0.960	0.960	0.961	0.975	0.963	0.963	0.965
$f + pd$	0.969	0.955	0.957	0.960	0.972	0.959	0.960	0.963
f	0.963	0.950	0.952	0.954	0.966	0.953	0.957	0.958
pp	0.967	0.962	0.961	0.961	0.971	0.965	0.965	0.965
pd	0.955	0.948	0.953	0.948	0.959	0.953	0.956	0.952
sp	0.957	0.964	0.955	0.957	0.957	0.960	0.952	0.956
Ro, L_1	0.697	0.504	0.499	0.509	0.747	0.571	0.569	0.578
Ro, L_2	0.714	0.567	0.556	0.593	0.718	0.583	0.565	0.600
ERo, L_1	0.651	0.709	0.703	0.663	0.693	0.761	0.756	0.702
ERo, L_2	0.795	0.695	0.619	0.736	0.786	0.693	0.609	0.710
IQIs	0.601	0.653	0.615	0.629	0.589	0.637	0.603	0.612

TABLE IV

PREDICTING FOREGROUND NU WITH BACKGROUND DISTORTION. SRCC AND LCC MEASURED OVER 1000 ITERATIONS USING 80/20 TRAIN/TEST SPLITS. $\mathrm{NU_H}$ REFERS TO HORIZONTAL STRIPING NU FOREGROUND DISTORTIONS, $\mathrm{NU_V}$ REFERS TO VERTICAL STRIPING NU FOREGROUND DISTORTIONS, AND $\mathrm{NU_{HV}}$ REFERS TO GRID-LIKE NU FOREGROUND DISTORTIONS. "NONE" REFERS TO ABSENCE OF BACKGROUND DISTORTION, AND "AWN" REFERS TO PRESENCE OF AWN BACKGROUND DISTORTION. L_1 AND L_2 REFERS TO L_1 AND L_2 NORMS RESPECTIVELY. THE IQIS WERE USED IN PLACE OF SNR BECAUSE SNR PERFORMED EXTREMELY POORLY.

NR Method	SRCC						LCC					
	$\mathrm{NU_H}$		$\mathrm{NU_V}$		$\mathrm{NU_{HV}}$		$\mathrm{NU_H}$		$\mathrm{NU_V}$		$\mathrm{NU_{HV}}$	
	None	AWN	None	AWN	None	AWN	None	AWN	None	AWN	None	AWN
$f + pp + pd + sp$	0.975	0.973	0.970	0.969	0.977	0.969	0.976	0.973	0.972	0.971	0.978	0.969
$f + pp + pd$	0.975	0.973	0.969	0.969	0.977	0.969	0.976	0.972	0.971	0.969	0.978	0.970
$f + pp$	0.971	0.973	0.964	0.967	0.977	0.973	0.972	0.972	0.966	0.966	0.978	0.975
$f + pd$	0.967	0.938	0.961	0.940	0.971	0.923	0.968	0.939	0.964	0.944	0.973	0.919
f	0.940	0.897	0.949	0.895	0.951	0.866	0.942	0.888	0.951	0.891	0.952	0.851
pp	0.970	0.972	0.966	0.969	0.976	0.975	0.971	0.972	0.968	0.968	0.977	0.976
pd	0.961	0.930	0.957	0.939	0.965	0.916	0.962	0.932	0.959	0.942	0.966	0.916
sp	0.961	0.964	0.967	0.965	0.973	0.965	0.962	0.963	0.966	0.966	0.973	0.962
Ro, L_1	0.548	0.157	0.552	0.151	0.556	0.136	0.626	0.239	0.621	0.236	0.625	0.229
Ro, L_2	0.572	0.213	0.609	0.244	0.548	0.183	0.533	0.237	0.575	0.274	0.502	0.212
ERo, L_1	0.424	0.400	0.404	0.393	0.464	0.268	0.417	0.414	0.404	0.413	0.468	0.328
ERo, L_2	0.565	0.191	0.642	0.336	0.646	0.222	0.565	0.283	0.678	0.401	0.647	0.308
IQIs	0.005	0.140	0.004	0.108	0.025	0.061	0.004	0.127	0.006	0.086	0.024	0.041

scored 1000 times, and the median SRCCs and LCCs are reported in Tables III and IV. Table III shows the correlation between the actual and predicted white noise variance in images with and without background NU distortion. Table IV shows the correlation between actual and predicted NU magnitude in images with and without presence of background AWN distortion.

As can be seen from Table III, each of the NSS feature groups, f, pp, pd, and sp, produce better predictors of AWN both with and without presence of NU as compared to Ro, ERo, and the IQIs. Combinations among these NSS feature groups do not increase predictor accuracy by much. In Table IV, each NSS feature group again produces better predictors of NU both with and without presence of AWN as compared to Ro, ERo, and the IQIs. The f group which does not measure directionality performs several percentage points lower than the other groups, pp, pd, and sp, which do. Comparing each of the previous methods, Ro, ERo, and the IQIs using

Table III, note a large difference between conditions with and without presence of background NU distortion. Similarly, Table IV indicates that these methods have low performance when detecting the amount of NU even without background distortion. ERo using the L_2 norm performs better than the other previous methods, but it is still heavily influenced by the level of background noise present in the image signal. It is important to note that the IQIs have almost no correlation with the amount of NU distortion present for this test, and they were a mediocre predictor of the presence of white noise.

Fig 9 depicts the SRCC of each feature with the amount of NU and AWN in images without background distortion. Fig 10 depicts the SRCC as in Fig 9, but with background distortion. For AWN alone, the standard deviation feature, f_2, produced the highest correlation whereas the asymmetry features, f_3 and f_4, did not correlate well with the listed distortions. With the presence of background distortion, the standard deviation, f_2 was again the most predictive feature. For $\mathrm{NU_H}$ and $\mathrm{NU_V}$
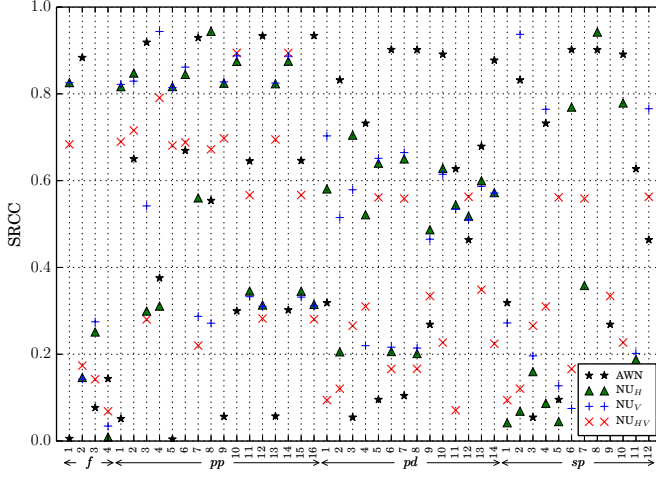
Fig. 10. SRCC of features against two combined distortions. When correlating features against NU, the images degraded by both NU and AWN were used. When correlating features against AWN, this same image set was used.

with and without background distortion, the shape parameter f_1 was the best predictor.

Since NU_H and NU_V are striping effects, they are highly oriented distortions. The sp group features show significant correlation with directionality, with vertical striping effects being highly correlated with the d_1^0 subband standard deviation, and horizontal striping effects being highly correlated with the d_1^{90} subband standard deviation. The paired product features indicate a similar oriented correlation, the horizontal paired product σ_r, pp_4, correlates highly with vertical striping, and the vertical paired product σ_r, pp_8, correlates highly with horizontal striping. This high degree of correlation between predicted and actual degree of distortion in single features is useful.

B. Discriminating the "Halo Effect"

The authors of [28] developed a person-detector which used the statistical gradients of estimated halos to enhance the detection task. To our knowledge, no methods exist for detecting halo artifacts in LWIR images.

To study how well the "Halo Effect" can be discriminated using our feature models, two sets of image patches (with and without halos) were constructed using background subtraction and manual classification to develop a supervised learner. Most of the image patches were of size 110x110. A total of 415 image patches were contained in both sets, with 227 image patches being halo-free, and 188 patches containing halos.

AWN and NU distortions were applied to each patch in both sets to reduce the dependence on the correlation between "Halo Effect" and the level of other common noise distortions. Each of these 415 image patches thus contained two artificial distortions in addition to the halo effect distortions. The distortion magnitudes $\sigma_{NU}, \sigma_{AWN} \in U[0.0025, 0.025]$ were randomly sampled and used as the variance of the white noise and Non-Uniformity distortions for each patch. The intervals for this uniform distribution were selected to scale the distortion from a just-noticeable to a significant difference.
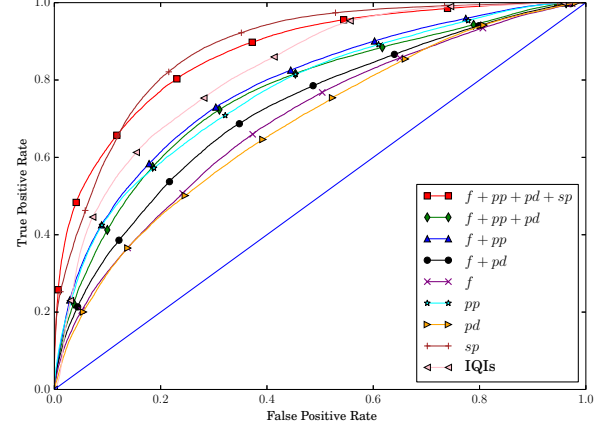

Fig. 11. ROC indicating the ability of NR algorithms to sort patches as either containing halos or as non-halo patches. Curves computed from 1000 train/test iterations using 415 total patches from the OSU dataset without content overlap.

TABLE V
AREAS UNDER THE ROC CURVES IN FIG. 11

NR Feature Set	Area Under ROC Curve
$f + pp + pd + sp$	0.795
$f + pp + pd$	0.711
$f + pp$	0.723
$f + pd$	0.675
f	0.651
pp	0.699
pd	0.639
sp	0.795
IQIs	0.735

Given these two distorted sets, those containing halos and those without, we devised a binary classification task. As in section A, we split the dataset into two non-overlapping subsets: 80 percent for training and 20 percent for testing. A Support Vector Classifier (SVC) was used to map the features between two classes. Random 80/20 splits were produced and classified with associated class probability estimates 1000 times.

Receiver Operating Characteristic (ROC) curves for the binary classification task using the proposed feature groups and the IQIs are shown in Fig. 11. The areas under the ROC curves are provided in Table V. The proposed NSS-based feature groups, except for sp and combinations of sp, achieved worse performance as compared to the IQIs for this discrimination task. Specifically, the sp performed significantly above the IQIs providing the largest discrimination capability both alone and when combined with with f, pp, and pd feature groups.

The steerable pyramid transform provides directionality of distortion which provides a great deal of information especially for the provided halo effect patches. Most objects in a scene are not circularly symmetric, thus their associated halo effect will not be symmetric. The steerable pyramid provides smooth directional features which are highly useful for the task.

TABLE VI
MEDIAN SRCC AND LCC BETWEEN ACTUAL AND PREDICTED TTP FROM
1000 ITERATIONS

NR Feature Set	SRCC	LCC
$f + pp + pd + sp$	0.665	0.671
$f + pp + pd$	0.640	0.646
$f + pp$	0.582	0.601
$f + pd$	0.609	0.613
f	0.504	0.527
pp	0.562	0.582
pd	0.566	0.568
sp	0.340	0.367
IQIs	0.621	0.630

Fig. 12. Example stimulus

Fig. 13. Sliding Quality Bar

C. TTP of Firefighters and Hazards

Researchers at NIST conducted a study involving firefighters whose task was two-fold [4]. First, given an LWIR image, the expert determined whether a hazard was present. Second, if a hazard was present, the expert was asked to identify the location of the hazard. This study was broken up into two phases. The phase 1 study used 4500 images. These images were created by degrading 180 pristine images. Five different levels of degradation corresponding to each IQI were generated and 25 sets of the four IQIs were used (for a total of 100 unique arrangements of the five values of each of the four IQIs). These 25 sets were deemed sufficient to represent the defined IQI space (5^4). Phase 2 used 55 sets of the four IQIs (for a total of 9900 images). The larger number of sets served also to extend the range of IQIs to include more extreme values. Note that the IQIs in this study were used as distortion-generating settings, allowing for direct measurement of distortion with TTP.

In this study, the experts were given a stimulus image, and tasked to either identify the location of the environmental hazard by clicking on it, or by indicating that there is no distortion. To better isolate detectability, we converted the dataset into patches centered about the hazards. Images with no hazards were discarded. Next, only the scores of observers that attempted to identify the location of the present environmental hazard were kept. Hits and misses were measured depending on whether the cursor click was near the hazard. The probability of hit was computed over all observers. By modifying the dataset in this way, SRCC and LCC correlations between target quality and target detectability could be more directly measured.

Using the probability of hit, the NSS quality features, and the IQIs, we used a SVR to estimate TTP. As a way of comparing the features, the median SRCC and LCC coefficients are reported in Table VI from 1000 iterations. Combinations of features provide the best estimators of TTP, with the combination of all natural features providing the highest correlations for TTP. Note that the IQIs in Table VI use the 13 features, while the degradations to the images provided in the study made modifications based on the original 4 parameters.

D. Blind Image Quality Assessment of LWIR Images

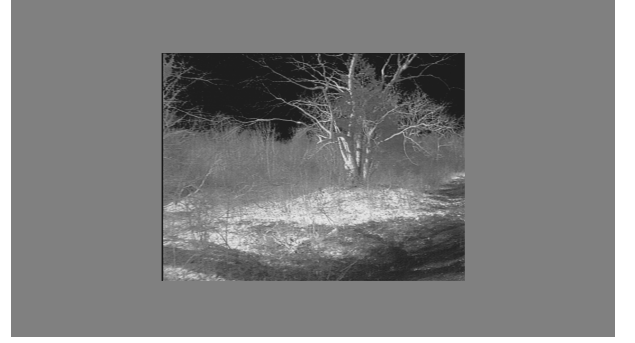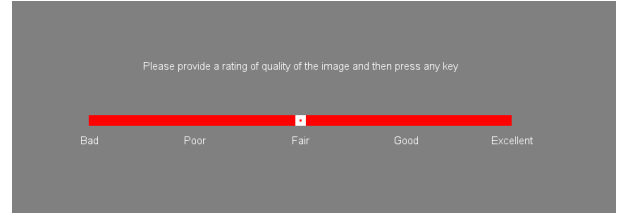We conducted a lengthy and sizeable human study, the results of which we used to assess how well NSS-based blind image quality prediction models designed for LWIR images correlate with subjective quality scores. A collection of 28 indoor and outdoor images were selected from the NIST and KASER databases as "pristine" images. Artificial blur and noise distortions were applied to the pristine images. Three levels of blur, three levels of noise, and combinations of blur and noise produced a total of 252 distorted images.

The subject test procedure was written using Matlab and the PsychToolbox [44] [45]. Each subject was first presented with a training session in which 10 images were shown before the main testing session, to give them an understanding of how to score images. Two testing sessions were performed with each session containing a unique set of 126 images. Subjects were presented with a single stimulus image for 10 seconds as depicted in Fig. 12. At the end of the 10 seconds, a continuous sliding quality bar with the labels "Bad", "Poor", "Fair", "Good", or "Excellent" was presented, as shown in Fig. 13.

Each image was scored by 24 subjects with each score discretized to integers on [0, 100]. In order to account for differences in image content, we computed the Difference Mean Opinion Scores (DMOS). Let s_{ijk} be the opinion score given by subject i, on image j during session $k = \{1, 2\}$. Then the difference score for subject i, image j, and session k is given by

$$d_{ijk} = s_{ij_{ref}k} - s_{ijk}, s_{ij_{ref}k} \neq s_{ijk},$$

where $s_{ij_{ref}k}$ is the score given to the (hidden) pristine image corresponding to the distorted one. The difference scores from each session were then converted to Z-scores:

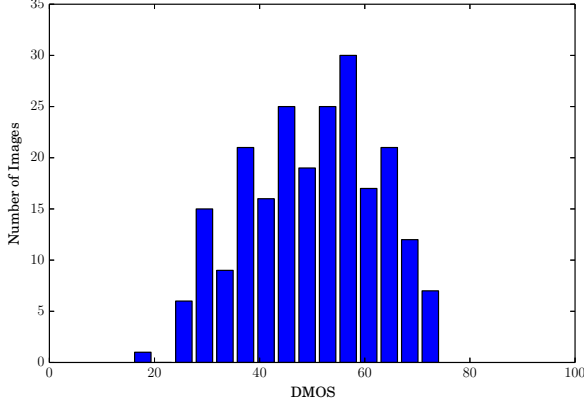$$z_{ijk} = \frac{d_{ijk} - \mu_{ik}}{\sigma_{ik}}$$

Fig. 14. Histogram of DMOS scores

TABLE VII
MEDIAN SRCC AND LCC BETWEEN DMOS AND PREDICTED DMOS
MEASURED OVER 1000 ITERATIONS

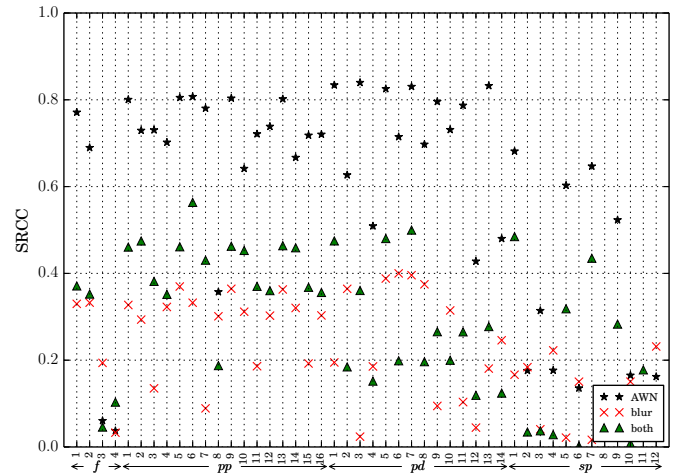| NR Feature Set	Median SRCC	Median LCC
$f + pp + pd + sp$	0.815	0.820
$f + pp + pd$	0.794	0.809
$f + pp$	0.809	0.817
$f + pd$	0.727	0.742
f	0.714	0.736
pp	0.794	0.809
pd	0.696	0.732
sp	0.825	0.828
IQIs	0.726	0.705
Ro, L_1	0.135	0.189
Ro, L_2	0.162	0.221
ERo, L_1	0.570	0.576
ERo, L_2	0.616	0.667

Fig. 15. SRCC of NSS features against DMOS scores. The performance against each distortion (noise and blur) was isolated for the purposes of comparison.

where

$$\mu_{ik} = \frac{1}{N_{ik}} \sum_{j=1}^{N_{ik}} d_{ijk}$$

and

$$\sigma_{ik} = \sqrt{\frac{1}{N_{ik} - 1} \sum_{j=1}^{N_{ik}} (d_{ijk} - \mu_{ik})^2}$$

and where N_{ik} is the number of test images seen by subject i in session k.

The subject rejection procedure specified in the ITU-R BT 500.11 recommendation is useful for discarding scores from unreliable subjects. Z-scores are considered normally distributed if their kurtosis falls between the values of 2 and 4. The recommendation is to reject if more than 5 percent of the Z-scores lie outside two standard deviations of the mean. Using this procedure, all except one subject was found to be acceptable. The one outlier chose the same value of 50 for all images. Thus only one subject was rejected [45] [46].

After the subject rejection procedure, the values of z_{ijk} fell into a range on $[-3, 3]$. A linear rescaling was used to remap the scores onto $[0, 100]$ using

$$z'_{ij} = \frac{100(z_{ij} + 3)}{6}$$

Finally the Difference Mean Opinion Score (DMOS) of each image was computed as the mean of the $M = 24$ rescaled Z-scores:

$$\text{DMOS}_j = \frac{1}{M} \sum_{i=1}^{M} z'_{ij}.$$

A plot of the histogram of the DMOS scores is shown in Fig. 14, indicating a reasonably broad distribution of the DMOS scores.

Table VII shows the Spearman's Rank Correlation Coefficient (SRCC) and (Pearson's) linear correlation coefficient (LCC) between the subjective scores and the model predictions for NR feature groups. The results were computed using 1000 iterations of randomly sampled training and testing groups. As in the previous sections, 80 percent of the data is used for training and the remainder for testing. Care was taken

to not overlap training and testing on the same content in any iteration since such an overlap could inflate performance results by training on the content rather than distortion. An SVR was used to fit the NSS feature parameters to the DMOS scores.

We observe that the steerable pyramid group features provide the highest correlation with the human subjective scores which is only a slight improvement over the BRISQUE model, $f + pp$. The combinations of feature groups performs worse compared to the individual groups indicating possible overfitting with the training set. For these blur and AWN distortions, the directional feature groups provide the highest correlation with DMOS scores with the IQIs and NU distortion models providing comparatively low correlation. The proposed models provide a great deal of predictive capability with human opinion scores, but there appears to be additional variation not accounted for in our proposed models.

Fig 15 depicts the SRCC of each feature's value with the human opinion scores. The highest individual feature correlations occur in the paired Log-derivative feature group, pd, but VII indicates that individual feature correlations are not as powerful as groups of features for predicting quality scores.

In fact, the sp feature group provides the highest correlations with DMOS scores when used together in a regression, but individually, they appear to make poor predictors.

E. Local Distortion Maps

Local distortion maps can be useful for identifying local distorted regions, which can occur as particular local distortions such as hotspots or halos, or they may arise from some unknown (combination of) distortions. It is possible to automatically find local distorted regions of LWIR images using NSS-based features.

A distortion map can be generated using a sliding window to capture patches from the image being analyzed. We used a 96x96 sliding square window scanned along the image in 12 pixel step increments (strides). Thus each captured patch overlapped with 87.5 percent of the last patch in sequence. Each patch was classified using multiple probabilistic SVCs, one per distortion type, to determine the likelihood that the patch belonged to that distorted class or to the natural image class. The probabilities of distortion were gathered and mapped into an image which highlights distorted areas. Example distortion maps are shown in Fig. 16. Some distortion maps, such as JPEG appear to provide false positives, but this is an artifact of relative probability within the map and full-scale contrast stretching. This technique could be useful for both identifying likely distortions, and localizing them in an LWIR image.

IV. CONCLUSION AND FUTURE WORK

LWIR images possess statistical regularities similar to those of visible light images. The NSS of LWIR images are powerful descriptors that can be used to find localized distortions, provide global distortion estimates, predict human task performance, and predict human subjective quality scores. Maps of distortions are potentially useful for designing correction algorithms or LWIR image compression methods.

Other distortions not studied here include geometric distortions, infrared reflections, and radiometric distortions. It's possible that studying these distortions in the context of LWIR NSS could also prove fruitful.

The NSS of LWIR videos are also of great interest. The LWIR videos used in surveillance could be modeled and studied to provide better compression techniques, better detection algorithms, and better overall video quality. The common visible light video compression formats including H.264/MPEG4 could be studied in light of thermal sensitivity requirements. Thermal variance, which is separate from and poorly represented by models of motion would be worthy of analysis as well.

ACKNOWLEDGMENT

REFERENCES

[1] C. Ibarra-Castanedo, J.-M. Piau, S. Guilbert, N. P. Avdelidis, M. Genest, A. Bendada, and X. P. Maldague, "Comparative study of active thermography techniques for the nondestructive evaluation of honeycomb structures," *Research in Nondestructive Evaluation*, vol. 20, no. 1, pp. 1–31, 2009.

[2] "Aeryon labs inc." http://www.aeryon.com/, 2014.

[3] F. K. Amon, N. P. Bryner, and A. Hamins, *Thermal Imaging Research Needs for First Responders: Workshop Proceedings*. US Government Printing Office, 2005.

[4] F. K. Amon, *Evaluation of Image Quality of Thermal Imagers used by the Fire Service*. US Department of Commerce, National Institute of Standards and Technology, 2009.

[5] F. K. Amon, D. Leber, and N. Paulter, "Objective evaluation of imager performance," *Fifth International Conference on Sensing Technology*, pp. 47–52, 2011.

[6] N. A. Diakides and J. D. Bronzino, *Medical Infrared Imaging*. CRC press, 2007.

[7] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer. A*, vol. 4, no. 12, pp. 2379–2394, 1987.

[8] D. Tolhurst, Y. Tadmor, and T. Chao, "Amplitude spectra of natural images," *Ophthalmic and Physiological Optics*, vol. 12, no. 2, pp. 229–232, 1992.

[9] B. A. Olshausen *et al.*, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

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

[11] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annual Review of Neuroscience*, vol. 24, no. 1, pp. 1193–1216, 2001.

[12] B. A. Olshausen and D. J. Field, "Natural image statistics and efficient coding," *Network: Computation in Neural Systems*, vol. 7, no. 2, pp. 333–339, 1996.

[13] J. M. Mooney, F. D. Sheppard, W. S. Ewing, J. E. Ewing, and J. Silverman, "Responsivity nonuniformity limited performance of infrared staring cameras," *Optical Engineering*, vol. 28, no. 11, pp. 281 151–281 151, 1989.

[14] J. M. Mooney, "Effect of spatial noise on the minimum resolvable temperature of a staring sensor," *Applied Optics*, vol. 30, no. 23, pp. 3324–3332, 1991.

[15] J. Lopez-Alonso, J. Alda, and E. Bernabeu, "Principal-component characterization of noise for infrared images," *Applied Optics*, vol. 41, no. 2, pp. 320–331, 2002.

[16] 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*. Springer, 2011, pp. 55–63.

[17] F. Pérez, M. Nova, J. E. Pezoa, M. Figueroa, and S. N. Torres, "Spatial and frequency domain metrics for assessing fixed-pattern noise in infrared images," *SBMO/IEEE MTT-S International Microwave & Optoelectronics Conference*, pp. 1–5, 2013.

[18] F. Pérez, J. E. Pezoa, M. Figueroa, and S. N. Torres, "Empirical frequency domain model for fixed-pattern noise in infrared focal plane arrays," submitted to *Infrared Physics & Technology*. June 2014.

[19] J. Y. Kaser, "Utilizing natural scene statistics and blind image quality analysis of infrared imagery," http://repositories.lib.utexas.edu/handle/2152/22601, 2013, [Online; accessed 12-Dec-2013].

[20] T. Goodall and A. C. Bovik, "No-reference task performance prediction on distorted LWIR images," *IEEE Southwest Symposium on Image Analysis and Interpretation*, pp. 89–92, 2014.

[21] 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. 4695–4708, 2012.

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

[23] C. San Martina, S. N. Torresa, and J. E. Pezoa, "An effective reference-free performance metric for non-uniformity correction algorithms in infrared imaging system," in *Proc. of the 20th Annual Meeting of the Laser and Electro-optic Society*, Lake Buena Vista, Florida, USA, 2007, pp. 21–25.

[24] "Cooled versus uncooled cameras for long range surveillance," *FLIR Technical Note, TN_0109_0005_EN*, http://www.flir.com/uploadedfiles/Eurasia/MMC/Tech_Notes/TN_0005_EN.pdf.

(a) Source Image

(b) Heatmaps

(c) Source Image

(d) Heatmaps

(e) Source Image

(f) Heatmaps

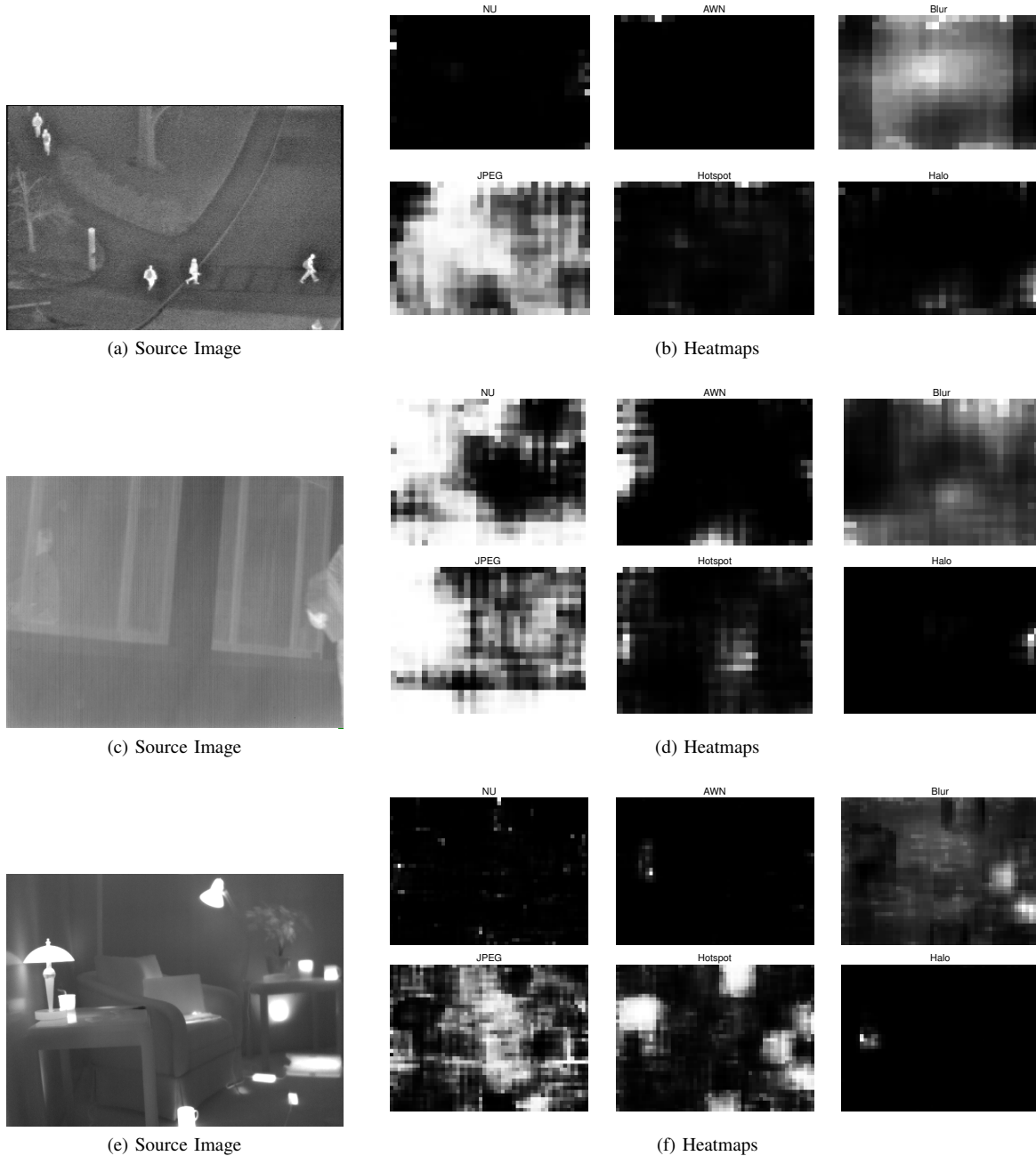Fig. 16. Identifying distorted regions

[25] J. V. Anda, "Uncooled detectors for thermal imaging cameras," *Flir Technical Note 16*, http://www.flir.com/uploadedFiles/Eurasia/Cores_and_Components/Technical_Notes/uncooled%20detectors%20BST.pdf.

[26] C. M. Hanson, H. Beratan, R. A. Owen, M. Corbin, and S. McKenney, "Uncooled thermal imaging at texas instruments," in *Infrared Detectors: State of the Art*, vol. 1735, 1992, pp. 17–26.

[27] G. C. Holst, *CCD Arrays, Cameras, and Displays*, JCD Publishing, Winter Park, Florida 32789 and SPIE, 1998.

[28] J. Davis and V. Sharma, "Robust background-subtraction for person detection in thermal imagery," *Proceedings of Computer Vision and Pattern Recognition Workshop*, vol. 8, p. 128, 2004.

[29] I. Schwartz, K. Snail, and J. Schott, "Infrared halo effects around ships," DTIC Document, Tech. Rep., 1985.

[30] J. L. Rowe, *The Impact of Thermal Imaging Camera Display Quality on Fire Fighter Task Performance*. ProQuest, 2008.

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

[32] J. W. Davis and M. A. Keck, "A two-stage template approach to person detection in thermal imagery," *Seventh IEEE Workshops on Application of Computer Vision*, pp. 364–369, 2005.

[33] Y. Zhang and D. M. Chandler, "An algorithm for no-reference image quality assessment based on log-derivative statistics of natural scenes," *IS&T/SPIE Electronic Imaging*, pp. 86 530J–86 530J, 2013.

[34] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891–906, 1991.

[35] E. P. Simoncelli and D. J. Heeger, "A model of neuronal responses in visual area MT," *Vision Research*, vol. 38, no. 5, pp. 743–761, 1998.

[36] 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. 3350–3364, 2011.

[37] J. W. Davis and V. Sharma, "Background-subtraction using contour-based fusion of thermal and visible imagery," in *Computer Vision and*

Image Understanding, vol. 106, no. 2. Elsevier, 2007, pp. 162–182.

[38] 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.

[39] N.-E. Lasmar, Y. Stitou, and Y. Berthoumieu, "Multiscale skewed heavy tailed model for texture analysis," *IEEE International Conference on Image Processing*, pp. 2281–2284, 2009.

[40] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "Live image quality assessment database release 2," http://live.ece.utexas.edu/research/quality.

[41] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, 2006.

[42] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[43] M. M. Hayat, S. N. Torres, E. Armstrong, S. C. Cain, and B. Yasuda, "Statistical algorithm for nonuniformity correction in focal-plane arrays," *Applied Optics*, vol. 38, no. 5, pp. 772–780, 1999.

[44] D. H. Brainard, "The psychophysics toolbox," *Spatial Vision*, vol. 10, pp. 433–436, 1997.

[45] 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.

[46] ITU-R, "Recommendation BT.500-11. methodology for the subjective assessment of the quality of television pictures," *International Telecommunication Union, Geneva, Switzerland*, 2002.

Alan Conrad Bovik holds the Cockrell Family Endowed Regents Chair in Engineering at The University of Texas at Austin, where he is Director of the Laboratory for Image and Video Engineering (LIVE). He is a faculty member in the Department of Electrical and Computer Engineering and the Institute for Neuroscience. His research interests include image and video processing, computational vision, and visual perception. He has published over 700 technical articles in these areas and holds several U.S. patents. His publications have been cited more than 35,000 times in the literature, his current H-index is over 70, and he is listed as a Highly-Cited Researcher by Thompson Reuters. His several books include the companion volumes *The Essential Guides to Image and Video Processing* (Academic Press, 2009).

Dr. Bovik has received a number of major awards from the IEEE Signal Processing Society, including: the Society Award (2013); the Technical Achievement Award (2005); the Best Paper Award (2009); the Signal Processing Magazine Best Paper Award (2013); the Education Award (2007); the Meritorious Service Award (1998) and (co-author) the Young Author Best Paper Award (2013). He also was named recipient of the Honorary Member Award of the Society for Imaging Science and Technology for 2013, received the SPIE Technology Achievement Award for 2012, and was the IS&T/SPIE Imaging Scientist of the Year for 2011. He is also a recipient of the Hocott Award for Distinguished Engineering Research from the Cockrell School of Engineering at The University of Texas at Austin (2008) and the Distinguished Alumni Award from the University of Illinois at Champaign-Urbana (2008). He is a Fellow of the IEEE, the Optical Society of America (OSA), and the Society of Photo-Optical and Instrumentation Engineers (SPIE). He co-founded and was the longest-serving Editor-in-Chief of the *IEEE Transactions on Image Processing* (1996-2002), created and served as the General Chairman of the *First IEEE International Conference on Image Processing*, held in Austin, Texas, in November, 1994, along with numerous other professional society activities, including: Board of Governors, IEEE Signal Processing Society, 1996-1998;; Editorial Board, *The Proceedings of the IEEE*, 1998-2004; and Series Editor for Image, Video, and Multimedia Processing, Morgan and Claypool Publishing Company, 2003-present.

Dr. Bovik is a registered Professional Engineer in the State of Texas and is a frequent consultant to legal, industrial and academic institutions.

Nicholas G. Paulter Jr. is the Group Leader for the Security Technologies Group and is a former Program Manager for the Detection, Enforcement, and Inspection program of the Law Enforcement Standards Office (OLES) at NIST in Gaithersburg, MD. He develops and oversees metrology programs related to concealed-weapon detection, through-wall surveillance, traffic control devices, imaging for security and emergency response applications, biometrics for identification, and weapon output characterization. In this role, he has initiated two NIST programs in sub-millimeter-wave concealed weapon imaging, facilitated the development of visible-light hyper-spectral image projection, initiated a NIST-centric imaging metrology program for security and emergency applications, started two through-wall surveillance/sensing metrology activities, and started a program for characterization of the high-voltage output of electroshock weapons. He was awarded a NIST Bronze Medal in 2003 for his work in developing minimum performance requirements for metal detectors. Prior to being a Program Manager in OLES, from 1990 to 2005, Mr. Paulter was the Leader for the High-Speed Pulse Metrology Project at NIST. In that capacity, he developed several high-speed electrical pulse generation and sampling systems, electro-optic based measurement systems, and short optical pulse laser systems for use in pulse metrology. Moreover, during his tenure as the Project Leader, the metrology services provided by his project became the best in the world.

From 1980 to 1989, Mr. Paulter was with Los Alamos National Laboratory, Los Alamos, NM, and was involved in the study of fast electrical and optical phenomena. Mr. Paulter has authored or co-authored over 100 peer-reviewed technical articles and provided numerous presentations at a variety of technical conferences. He is a 2008-2009 Commerce Science and Technology Fellow and a 2010 IEEE Fellow; and received a B.S. in biological sciences from the University of California, Santa Barbara in 1980; an M.S. in chemistry from the University of New Mexico in 1988; and an M.S. in electronics engineering from the University of Colorado Boulder in 1990.

Todd Richard Goodall received the B.S. degree in Computer Engineering from Clemson University, Clemson, SC, in May 2012. He joined the Laboratory for Image and Video Engineering (LIVE) in August 2013, and is currently pursuing the M.S. and Ph.D. degrees in electrical and computer engineering.

He is the recipient of the NDIA UWD Academic Fellowship for 2012-2013. His research interests include the statistical modeling of images and videos, design of image and video quality assessment algorithms, visual perception, and computational vision.