An Evaluation of Objective Image Quality Assessment for Thermal Infrared Video Tone Mapping

Michael Teutsch[1], Simone Sedelmaier[1,2], Sebastian Moosbauer[1], Gabriel Eilertsen[3], Thomas Walter[2]

[1] Hensoldt Optronics GmbH, Oberkochen, Germany

{michael.teutsch, simone.sedelmaier, sebastian.moosbauer}@hensoldt.net

[2] Ulm University of Applied Sciences, Germany   [3] Linköping University, Sweden

thomas.walter@thu.de   gabriel.eilertsen@liu.se

Abstract

State-of-the-art thermal infrared cameras produce high quality images with a bit depth of up to 16 bits per pixel (bpp). In practice, the data often reach a bit depth of 14 bpp, which cannot be displayed naïvely to a standard monitor that is limited to 8 bpp. Therefore, the dynamic range of these images has to be compressed. This can be done with an operator called tone mapping. *There are many methods available for tone mapping, but the quality of the results can be extremely different. In this paper, we discuss and evaluate image quality assessment measures for tone mapping taken from the literature using thermal infrared videos. The usefulness of the measures is analyzed and effectively demonstrated by utilizing various reference Tone Mapping Operators (TMOs) based on traditional algorithm engineering on the one hand and deep learning on the other hand. We conclude that the chosen measures can objectively assess the quality of TMOs in thermal infrared videos.*

1. Introduction

Tone mapping is a method used in the field of computational imaging to map a High Dynamic Range (HDR) image [27] to a Low Dynamic Range (LDR) [10]. A camera that is able to acquire images with a value range (= bit depth) larger than 8 bits per pixel (bpp) can lead to suboptimal images when naïvely displayed on a standard monitor limited to 8 bpp. Such naïve displaying can be the result of linear downscaling and in this way, the rich information contained within the higher bit depth can get lost. Instead, intelligent and image content sensitive tone mapping can not only preserve all relevant information but even emphasize it. This approach is well-known for its application in computational photography [29], where visual-optical (VIS) HDR images and tone mapping can be used to make

details of the image content visible to the spectator even in case of difficult illumination conditions with image regions that appear to be close to saturation [14, 16, 28]. Similar demands raise for medical imaging [25] or visual surveillance applications, in which thermal infrared (IR) cameras are used [19]. Although thermal IR cameras do not see reflected but emitted light [12], different temperature ranges within the same observed scene can lead to image regions with weak contrast. Such regions can contain information crucial for scene understanding and need to be preserved by an appropriate Tone Mapping Operator (TMO).

Table 1. Motivation for IR Tone Mapping.

Linear downscaling	CAN24_AN [7]
CLAHE [41]	Reference TMO [9]

Various TMOs exist, but the quality of the results can be extremely different as demonstrated in Table 1: the first example shows naïve linear downscaling resulting in a dark image with weak contrast. The second example shows the result if the same 14-bit IR image is processed using a deep learning based Context Aggregation Network (CAN24_AN) that has been originally trained to improve the contrast of

8-bit color images [7]. Now it is possible to identify the image content, but the contrast especially in the background is still poor. Then, a standard approach for compressing the dynamic range of images is used called Contrast Limited Adaptive Histogram Equalization (CLAHE) [41] resulting in better contrast and more visible details in the background. Finally, the result of one reference TMO used in this paper is presented [9] providing best detail visibility. These examples not only demonstrate the need to study TMOs, but also to examine what it means to have a good result and how this can be objectively evaluated since human perception of quality is very subjective.

In this paper, we discuss and evaluate objective Image Quality Assessment (IQA) measures for tone mapping taken from the literature [8, 10, 37, 38]. Such measures consider both the tone mapped LDR image and the original HDR image. Hence, they are more suitable compared to conventional full reference metrics such as Peak Signal-To-Noise Ratio (PSNR) [40] or Structural Similarity (SSIM) [32]. While the measures were originally designed to evaluate tone mapping in the visual-optical spectrum, we analyze them for their applicability to thermal IR videos. To the best of our knowledge this is the first paper discussing tone mapping IQA in the IR spectrum. The contributions lie in (1) an in-depth analysis of the considered IQA measures exploring details missing in the original papers and including a mathematical derivation, (2) a detailed evaluation for two public thermal IR datasets [4, 11], and (3) the release of code[1] for four objective IQA measures [10] that has not been released before. In this way, we provide a tool enabling reproducibility and comparability of TMOs. While recent literature focuses on rather subjective IQA measures such as a purely qualitative evaluation [23] or the Tone Mapping Quality Index (TMQI) [16], we aim at evaluating objective measures to support human operators for surveillance applications. The main purpose of IR imaging in such applications is to convey information rather than producing visually appealing images.

The remainder of this paper is organized as follows: related work is presented in Section 2. The analyzed evaluation measures are introduced in Section 3. Experimental results are described in Section 4. We conclude in Section 5.

2. Related Work

IR Tone Mapping Operators: Tone Mapping is usually applied to an image's luminance channel in the log domain [8]. Since thermal infrared images consist of one intensity or luminance channel in general, typical color transformations or color restoring techniques [30] can be skipped here. An edge-preserving filter such as a bilateral filter [6] is applied to separate the image's base and detail layer. Then,

logistic or sigmoidal functions called tone curves are applied to the tiled base layer [9, 10, 20] to enhance the contrast of the signal but not the noise.

IR images usually belong to the group of HDR images. A typical infrared detector has a temperature resolution less than 0.02 K. However, in a natural scene the difference in temperature radiation can reach more than 30 K or sometimes even over 100 K. Consequently, the intensity value range of an infrared image often reaches more than 1,000 [19]. In general, however, there is not much literature that examines tone mapping specifically for IR images [17, 18, 19, 31]. Linear mapping and histogram adjustment based mapping are classic tone mapping methods that process an image globally [19]. With linear mapping, however, a larger range of radiation distribution lowers the radiation resolution of the output image leading to weak detail visibility. Global methods based on histogram adjustment such as histogram equalization are widely used due to their low computational complexity and effectiveness [18]. While a global tone-curve can easily compress the dynamic range, the contrast will be reduced. For IR images and videos, however, it can be important to preserve local contrasts to better convey information across an image and avoid losing important image features. In order to overcome the disadvantages of global approaches, methods were developed to enhance local contrast based on the characteristics of local regions [17].

Image Quality Assessment: There are no IR specific tone mapping evaluation measures. Hence, measures are adopted from tone mapping for visual-optical images. Such measures typically compare the original HDR with the processed LDR image to determine a loss of image quality during tone mapping. We skip subjective measures [21] and focus on objective measures instead. Yeganeh *et al.* [38] developed a semi-objective IQA model, the so-called TMQI, for tone mapped images using their corresponding HDR images as reference. The proposed quality index is based on two assessment methods. The SSIM [32] and its multi-scale extensions such as the MS-SSIM [33, 34] is the first approach claiming that structural fidelity is a good indicator of perceptual quality. This is based on the hypothesis that the main purpose of vision is to extract structural information from the visual scene. The second method is the so-called Natural Scene Statistics (NSS) model, which is based on the theory that the visual system is highly adapted to the natural visual environment. In this way, the variance from natural image statistics is used as a measure of perception quality [33]. However, several authors demonstrated that the TMQI is not sensitive to halo effects [26] or not sufficiently representing human perception [2]. Usually, the occurrence of artifacts during processing impairs the quality of the tone mapped video content. Objective evaluation measures aim at quantifying the presence of such artifacts

---

[1]https://github.com/HensoldtOptronicsCV/ToneMappingIQA

after tone mapping. Eilertsen *et al.* [10] propose four objective measures that assess the presence of (1) temporal incoherence artifacts such as local or global flickering are an important issue [3, 5], (2) the percentage of over- and underexposed image pixels, (3) the contrast and the (4) noise visibility of an image before and after tone mapping. These four objective measures together with the semi-objective TMQI are considered in the remainder of this paper.

3. Tone Mapping Image Quality Assessment

As already mentioned, the TMQI [38] is based on heuristics of human perception. However, not all complex processes involved in a subjective comparison can be considered. As a result, Eilertsen *et al.* [10] motivated to evaluate four specific properties and not the overall subjective quality as a single quality score. In the following sections, the four measurements covering the attributes *exposure*, *contrast*, *noise visibility* and *temporal incoherence* are presented in detail. The reason for the detailed description of the measures is that the authors describe them very briefly in their survey paper and hence a thorough understanding is not possible. In general, the aim of the measures is to indicate the performance of a TMO regarding (1) generating well-exposed tone mapped images with (2) good contrast, (3) reduced or at least not amplified camera noise and (4) show no temporal artifacts such as flickering. More details about the measures such as pseudo code or appropriate preprocessing can be found in the supplementary material.

3.1. Over-/Underexposure Measure

A well-exposed image has at best only few pixels in saturation. Hence, the goal of the exposure measure is to estimate if a TMO can be expected to calculate well-exposed images. Considering overexposure and thus the amount of saturation, the measure calculates the fraction of pixels in the normalized tone mapped image T with pixels $s \in [0, 1]$ which are above or equal 0.95. In the case of underexposure, the fraction of pixels below or equal 0.02 is taken into account. These two measures should have a low but unequal zero value indicating well-exposed tone mapped images capable of sufficiently compressing the dynamic range.

3.2. Loss Of Contrast Measure

Due to the difficulty of uniformly specifying contrast in different images, there are many ways to define contrast. A commonly used definition for measuring the contrast is the Michelson formula [22]

$$C_M = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}, \qquad (1)$$

where I_{max} and I_{min} represent the maximum and the minimum luminance values, respectively. C_M ranges from 0 to

1. However, the Michelson contrast measure does not represent image contrast well as it only depends on the maximum and minimum values of the image [1]. In IR imaging the socalled Infrared Focal Plane Arrays (IRFPA) are widely used to acquire IR images. Technical limitations and material defects in the manufacturing process lead to the fact that IRFPA usually contains defective pixels, which can affect the quality of IR images considerably [36]. These defective pixels can be seen as salt and pepper noise [13] and would have an erroneous impact on the Michelson contrast measure, as they are the maximum and the minimum luminance values. Hence, we need different contrast measures.

In general, the loss of contrast measure is divided into two calculations. First, the *local contrast* is determined. This means that the details are extracted with a bilateral filter by subtracting the filtered image F_{bi} from the tone mapped image T as shown in Eq. 2. Then, the absolute values of the subtraction are weighted with the input luminance L and the mean value over all pixels x is taken with the total number of pixels N leading to

$$C_{local}(T) = \frac{1}{N} \sum_x L_x \left| T_x - F_{bi_x} \right|. \qquad (2)$$

Equation 3 shows the definition of the bilateral filter, where the sum over Ω_x means that a local neighborhood of the pixel x is considered. Therefore, each pixel is a weighted average of its neighbors and the weight assigned to each neighbor decreases with the distance in the image plane $|T_p - T_x|$ and the distance on the intensity axis $\|p - x\|$. Additionally, a Gaussian function G is used as a decreasing function. σ_{sp} and σ_c are the corresponding standard deviations, where the index sp stands for space and the index c for color. In the context of this paper, color obviously means the one-channel luminance. W normalizes the sum of the weights.

$$F_{bi} = \frac{1}{W} \sum_{p \in \Omega_x} G_{\sigma_{sp}}(\|p - x\|) G_{\sigma_c}(|T_p - T_x|) T_p$$

with $\quad W = \sum_{p \in \Omega_x} G_{\sigma_{sp}}(\|p - x\|) G_{\sigma_c}(|T_p - T_x|)$

$$\qquad (3)$$

The second part of the contrast measure is the *global contrast*, which is estimated using a Gaussian kernel G leading to

$$C_{global}(T) = \frac{1}{N} \sum_x \sqrt{(G * T^2)_x - (G * T)_x^2}, \qquad (4)$$

where $*$ is the convolution operator and T is the same tone mapped image as used in Eq. 2 and 3. Equation 4 shows that the mean value of all pixels is calculated in the same way as with the local contrast measure. Finally, the two contrast

measures are applied to each image pair with input HDR image L and tone mapped image T. Afterwards, they get subtracted in order to estimate the loss of contrast after tone mapping. The local and global contrast coefficients should then be as large as possible, i.e. only minimal contrast is lost. This contrast difference or *loss of contrast* [10] can then be expressed with the formulations

$$c_{local}(L,T) = C_{local}(T) - C_{local}(L), \qquad (5)$$
$$c_{global}(L,T) = C_{global}(T) - C_{global}(L). \qquad (6)$$

3.3. Noise Visibility Measure

The motivation for the noise visibility measure is that camera noise can have a significant influence on the results of a tone mapping algorithm. Under the transformation of a tone curve $V : L \to T$, the noise visibility can vary. Eilertsen *et al.* [10] use an artificial, synthetic HDR input L without any noise to quantify the change in noise visibility. In this paper, we transfer this measure to the real world and apply it to all images considered for the evaluation framework. Starting from an input HDR image, artificial noise following a Poisson distribution is added and a corrupted HDR input is constructed. In thermal IR imaging it is common practice to favor Poisson distributed noise (or shot noise) instead of Gaussian distributed additive noise [15, 35]. Both the original HDR input L and the manipulated HDR input $\hat{L}$ are then tone mapped

$$\begin{aligned} T &= V(L), \\ \hat{T} &= V(\hat{L}), \end{aligned} \qquad (7)$$

where V describes the tone mapping operator or the tone curve, respectively. Then, the quality predictor Θ from the HDR-VDP 2.2.0 metric provided by Mantiuk *et al.* [21] is used to measure the visibility of the added noise. This metric needs a reference image and a test image as input. A result of 100 indicates best quality and that the visual differences are very small. A lower value, however, indicates lower image quality. Equation 8 shows the final definition of the noise visibility measure, where the quality metric is first calculated for the HDR input and the HDR input with additional noise and secondly for the corresponding tone mapped inputs. Subsequently, the difference is calculated as final result.

$$n = \Theta(\hat{L}, L) - \Theta(\hat{T}, T) \qquad (8)$$

A positive difference $n > 0$ means that the noise is more visible after tone mapping. Conversely, if the noise visibility measure becomes negative $n < 0$, the noise visibility is reduced after tone mapping.

3.4. Temporal Incoherence Measure

Various temporal artifacts can occur when performing video tone mapping if a TMO is applied for each frame individually. Global and local flickering artifacts appear, for example, when the image statistics used for a tone curve of a TMO change rapidly from frame to frame. As a result, the tone curve changes non-smoothly.

In order to investigate the temporal coherence of a TMO, a comparison of the tone mapped sequence and the input HDR sequence is performed to derive their correlation over time. However, the challenge is how to consider transformations to which the measure should be invariant such as scaling, exponentiation, temporal adjustment, etc. Therefore, the temporal incoherence measure quantitatively estimates the temporal inconsistencies with a natural measure that examines the linear dependence of the HDR image and the tone mapped image using their cross correlation. The sample Pearson correlation coefficient can be applied to calculate a normalized measure r_{LT} with

$$r_{LT} = \frac{\sum\limits_{k=t-d}^{t+d} (L_k - \mu_L)(T_k - \mu_T)}{\sqrt{\sum\limits_{k=t-d}^{t+d} (L_k - \mu_L)^2 \cdot \sum\limits_{k=t-d}^{t+d} (T_k - \mu_T)^2}} \qquad (9)$$

where L denotes the HDR and T the tone mapped images. An evaluation of the linear dependence is done over a local neighborhood in time with $k \in \{t-d, ..., t+d\}$. The means μ_L and μ_T and the standard deviations in the denominator are used to normalize the HDR and the tone mapped signals to have the same mean and standard deviation within the temporal neighborhood $\{t - d, ..., t + d\}$. In practice, the mean value or a pixel value of the tone mapped image can change over time in an opposite direction to the HDR image. This is due to the ability of a video TMO to adapt to the input over time based on statistics from the HDR sequence. Equation 9 does not take this into account. The aim is to make the measure invariant to local linear changes over time. Therefore, the normalization of Eq. 9 using the mean value μ is extended. As a first step, a normalization from a linear regression is performed, then the normalized HDR input is normalized again to have the same variance as the normalized tone mapped input.

The estimation of global temporal incoherence is described below. In this context, the mathematical elaboration is more detailed than it was done by Eilertsen *et al.* [10] in order to better understand the individual steps. The measure is applied using the mean values in each frame $L_t = \frac{1}{N} \sum_x L_{t,x}$, where x are the pixel values and N is the number of pixels. The first step is to calculate the sample means of all HDR images L and tone mapped images T within the temporal neighborhood,

$$\boldsymbol{\mu_L} = \begin{pmatrix} \sum\limits_x \frac{L_{-d,x}}{N} \\ \vdots \\ \sum\limits_x \frac{L_{d,x}}{N} \end{pmatrix} = \begin{pmatrix} \mu_{L_{-d}} \\ \vdots \\ \mu_{L_d} \end{pmatrix}, \qquad (10)$$

$$\boldsymbol{\mu_T} = \begin{pmatrix} \sum_x \frac{T_{-d,x}}{N} \\ \vdots \\ \sum_x \frac{T_{d,x}}{N} \end{pmatrix} = \begin{pmatrix} \mu_{T_{-d}} \\ \vdots \\ \mu_{T_d} \end{pmatrix}. \qquad (11)$$

The sample distances are defined by

$$\boldsymbol{X} = \begin{pmatrix} -d & \cdots & d \end{pmatrix}. \qquad (12)$$

Equations 13 and 14 describe the calculation of the weights w_L and w_T of the linear regression:

$$w_L = \frac{\sum\limits_{k=t-d}^{t+d} (\boldsymbol{\mu_L^T} \cdot \boldsymbol{X})}{\boldsymbol{X} \cdot \boldsymbol{X}^T} = \frac{\sum \left(\mu_{L_{-d}} \cdots \mu_{L_d} \right) \cdot (-d \cdots d)}{(-d \cdots d) \cdot \begin{pmatrix} -d \\ \vdots \\ d \end{pmatrix}} \qquad (13)$$

$$w_T = \frac{\sum\limits_{k=t-d}^{t+d} (\boldsymbol{\mu_T^T} \cdot \boldsymbol{X})}{\boldsymbol{X} \cdot \boldsymbol{X}^T} = \frac{\sum \left(\mu_{T_{-d}} \cdots \mu_{T_d} \right) \cdot (-d \cdots d)}{(-d \cdots d) \cdot \begin{pmatrix} -d \\ \vdots \\ d \end{pmatrix}} \qquad (14)$$

The results of Eqs. 13 and 14 are scalar values. In general, a temporal averaging kernel is defined by uniform weights K with

$$\boldsymbol{K} = \begin{pmatrix} \frac{1}{2d+1} \\ \vdots \\ \frac{1}{2d+1} \end{pmatrix} \qquad \text{with} \qquad \boldsymbol{K} \in \mathbb{R}^{(2d+1) \times 1}. \qquad (15)$$

With the help of the kernel K the fitted lines after the regression are calculated for the HDR input and the tone mapped input, respectively:

$$\boldsymbol{y_L} = w_L \cdot \boldsymbol{X} + \sum_{k=t-d}^{t+d} \boldsymbol{K} \cdot \boldsymbol{\mu_L} \qquad (16)$$

$$\boldsymbol{y_T} = w_T \cdot \boldsymbol{X} + \sum_{k=t-d}^{t+d} \boldsymbol{K} \cdot \boldsymbol{\mu_T} \qquad (17)$$

Subsequently, the mean values are normalized with the inserted lines:

$$\hat{\boldsymbol{t}}_L = \boldsymbol{\mu_L} - \boldsymbol{y_L^T} \qquad (18)$$

$$\hat{\boldsymbol{t}}_T = \boldsymbol{\mu_T} - \boldsymbol{y_T^T} \qquad (19)$$

For the second part of the normalization the variances are calculated for the normalized HDR and tone mapped input:

$$\sigma_L^2 = \sum_{k=t-d}^{t+d} \boldsymbol{K} \cdot (\hat{\boldsymbol{t}}_L)^2 \qquad (20)$$

$$\sigma_T^2 = \sum_{k=t-d}^{t+d} \boldsymbol{K} \cdot (\hat{\boldsymbol{t}}_T)^2 \qquad (21)$$

Now, the normalization of the HDR signals is performed on the basis of the standard deviations with

$$\tilde{\boldsymbol{t}}_L = \frac{\hat{\boldsymbol{t}}_L \cdot \sigma_T}{\sigma_L}. \qquad (22)$$

Then, the linear slope of the HDR signal L is put back:

$$\boldsymbol{t_L} = 0.25 \cdot \boldsymbol{X}^T + \tilde{\boldsymbol{t}}_L \qquad (23)$$

$$\boldsymbol{t_T} = 0.25 \cdot \boldsymbol{X}^T + \hat{\boldsymbol{t}}_T \qquad (24)$$

The slope of 0.25 was chosen by Eilertsen *et al.* [10] and additionally applied to the signals. Now the HDR signal has the same slope and variance as the tone mapped signal. As in Eq. 9, the entries of the sample Pearson correlation coefficient q_1, q_2 and q_3 are calculated.

$$q_1 = \sum_{k=t-d}^{t+d} \boldsymbol{K} \cdot (\boldsymbol{t_L})^2 \qquad (25)$$

$$q_2 = \sum_{k=t-d}^{t+d} \boldsymbol{K} \cdot (\boldsymbol{t_T})^2 \qquad (26)$$

$$q_3 = \sum_{k=t-d}^{t+d} \boldsymbol{K} \cdot \boldsymbol{t_L} \cdot \boldsymbol{t_T} \qquad (27)$$

$$cf_{LT} = 1 - \max \left(0, \frac{q_3}{\sqrt{q_1 \cdot q_2}} \right) \qquad (28)$$

Finally, the incoherence measure cf_{LT} is estimated, so that negative correlations are clamped and considered as equivalent to no correlation. The measure is calculated for each stack of $2d + 1$ HDR and the corresponding tone mapped images in an image sequence consisting of M frames. Since a temporal neighborhood is considered, a temporal offset must be introduced starting with the $(d+1)$-th frame and ending with the $(M - d - 1)$-th frame.

For the local temporal incoherence coefficient, the same calculation steps are performed, but they are applied per pixel and not with the mean values in each frame. Consequently, the temporal averaging kernel K for the local measure is a matrix with $K \in \mathbb{R}^{h \times v \times (2d+1)}$ where h and v are the horizontal and vertical image dimensions. In addition, the normalization steps are performed elementwise, which results in a matrix for the correlation coefficient in Eq. 28.

Therefore, the local per pixel measure is only calculated on values of the tone mapped image that are not under- or over-exposed. Subsequently, an averaging is performed.

In general, the temporal incoherence measure gives a result of 0 if the HDR and the tone mapped signals are fully linearly correlated. On the other side, a value of 1 means that the signals are either completely uncorrelated or negatively correlated. Consequently, the goal of a TMO is to produce an incoherence coefficient close to 0.

4. Experiments and Results

In this section, we first describe the datasets used for our experiments. Then, we show results for artificially degrading reference HDR and LDR IR images to verify that the evaluation measures are sensitive to typical IR image artifacts. Four reference TMOs are introduced and qualitatively and quantitatively evaluated in the remainder of this section.

4.1. Datasets

Two public thermal IR datasets that provide 8- and 16-bit image pairs are used for the experiments. The Linköping Thermal IR (LTIR) dataset provided by Berg *et al*. [4] contains image sequences in the PNG format. Seven different scenes with 20 sequences in total are available for different IR cameras. However, the resolution of the images varies across the sequences between 320×256 and 640×480 pixels. The FLIR dataset [11] is a thermal dataset for training DCNNs in an automotive environment. This dataset is available in TIFF format with a bit depth of 16 bpp, of which 14 bpp are effectively used, and in JPEG format with a bit depth of 8 bpp, where an automatic gain control is applied. All images have a resolution of 640×512 pixels.

4.2. Evaluation of the Tone Mapping IQA Measures

Now, we use artificial image quality degradation techniques to verify that the presented evaluation measures are able to detect their target artifacts for thermal IR images. Each evaluation is carried out with five samples taken from each dataset. In this way, we can present mean and standard deviation of the evaluation measures. The chosen samples for the FLIR dataset are frames 08866, 09032, 09119, 09780, and 10087 for the exposure, loss of contrast, and noise visibility measures and five sequences of 13 frames each starting at the already mentioned frames (e.g., 08866–08877) for the temporal incoherence measure. From the LTIR dataset we chose the first frame of the sequences (4) horse, (10) crouching, (15) depthwise crossing, (18) quadcopter2, and (19) selma, respectively. We use the first 13 frames of each sequence for the temporal incoherence measure. Since both datasets provide 8- and 16-bit images, we directly use these image pairs for our evaluations in this section. Example images for all image quality degradations can be found in the supplementary material.

Table 2. Five IQA measures for four TMOs on the FLIR test sub-dataset. The qualitative evaluation in Fig. 2 confirms the relative quantitative evaluation behind the numbers of this table. The best result for each measure is underlined.

TMO	[11]	[9]	[7]	[16]
TMQI ↑	.593	.519	.517	.427
Underexposure (in %) ↓	.54	.18	.16	3.61
Overexposure (in %) ↓	.82	.06	.13	1.81
Loss of Global Contrast ↓	-0.16	-0.104	-0.105	-0.061
Loss of Local Contrast ↓	-0.046	-0.011	-0.014	-0.008
Noise Visibility ↓	-	18.59	11.22	3.74
Global Temporal Incoherence ↓	.0002	.0003	.0003	.0005
Local Temporal Incoherence ↓	.0253	.0365	.0325	.009

Over-/Underexposure: Over- and underexposure can be measured well as we can see in Fig. 1 (a-b). The image quality is artificially impaired using clipping. For example 0.9 means that we clip the 10 % brightest pixels and equalize the image intensities again, i.e. we bring more bright pixels to saturation. As a consequence, the overexposure measure increases for both datasets. A similar behavior can be observed for underexposure, where we bring more and more dark pixels to 0 intensity.

Noise Visibility: The verification of the noise visibility measure is more complicated since we need four images for the comparison: the original HDR and LDR images as well as the noisy HDR and the resulting tone mapped noisy LDR image. As we do not have access to the TMO used to tone map the FLIR and the LTIR dataset, we simple apply Poisson noise to both the HDR and the LDR image to generate the noisy images. Then, we successively add Gaussian noise to the noisy LDR image to test the measure. Figure 1 (c) shows the result. With increasing sigma for Gaussian noise, the noise visibility measure increases, too.

Loss Of Contrast: In Fig. 1 (d-g), we can see the results for the verification of the loss of contrast measure. We use two techniques to degrade the image quality: clipping and Gaussian blur. For clipping, we continuously reduce the image value range leading to a flat image with weak contrast. Both degradations show an expected behavior of the measure: monotonically increasing values. However, all the values are negative. This means that we always gain contrast after tone mapping compared to the original HDR image. This may be surprising at the beginning. But since typical scenes do not cover the entire temperature range of thermal IR cameras, the HDR intensity value range is not fully occupied. Hence, negative values must actually be expected.
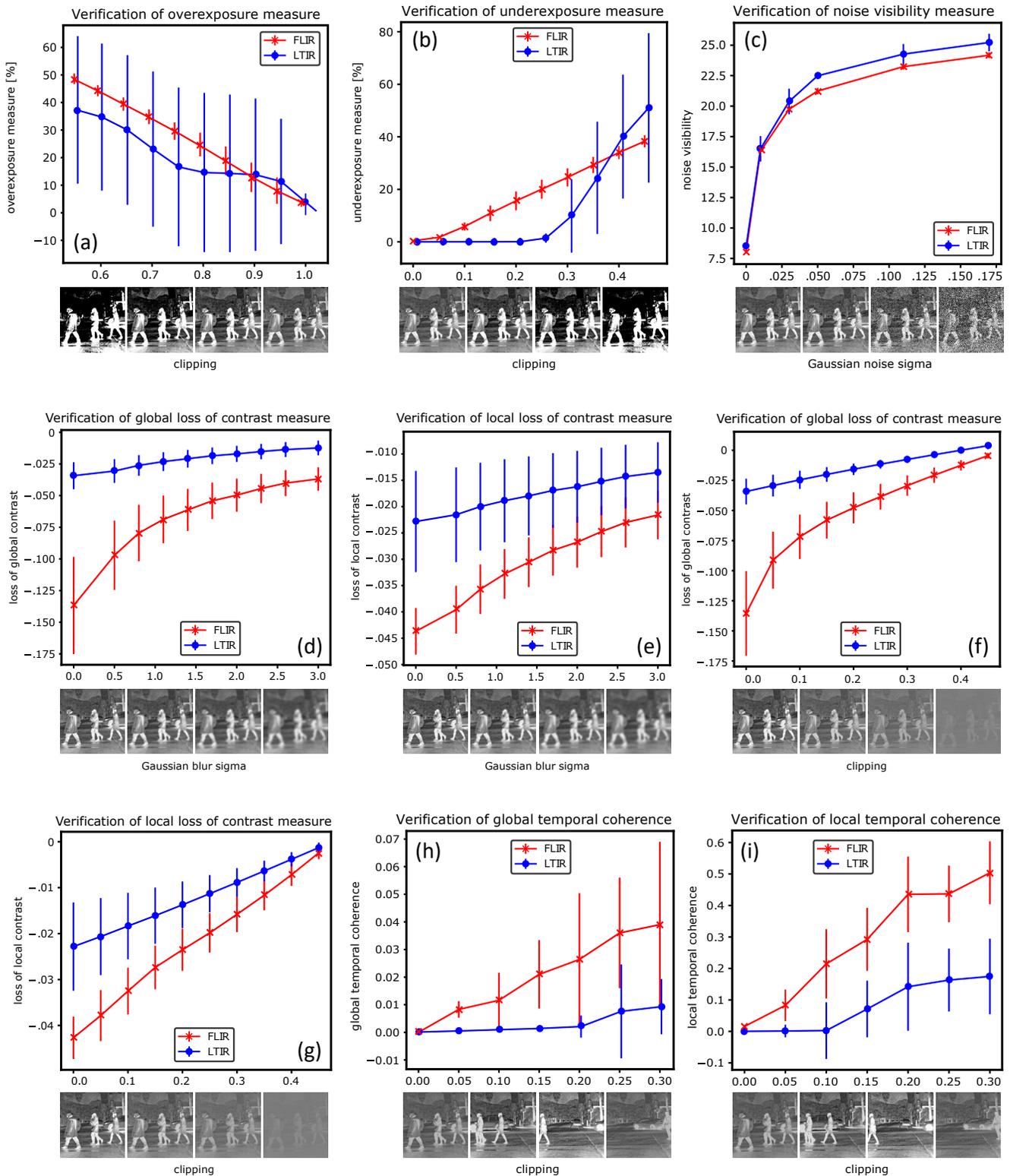
Figure 1. Verification results for the four discussed objective IQA measures. For artificially impaired images, the evaluation measures behave as expected. We consider two different IR datasets and took five samples from each dataset. This indicates that the IQA measures are applicable to evaluate IR video TMOs. The vertical bars show the standard deviation across the five samples. The small images that demonstrate the degradation of the image quality are cropped from one sample of the FLIR dataset. The original images and their impaired versions can be found in the supplementary material. The exact numbers of the plots can be found on the website mentioned in Footnote 1.
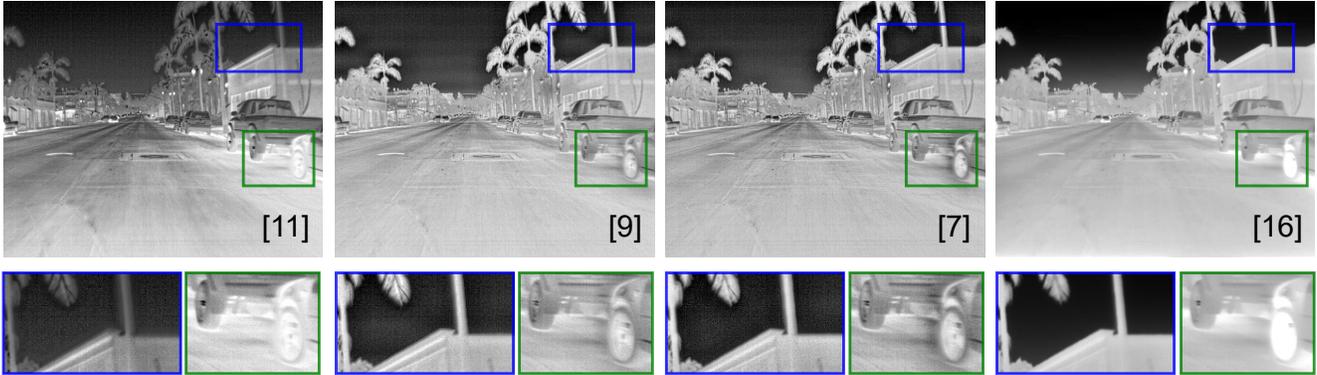
Figure 2. Processed image confirming Table 2 by showing us that (1) [16] do good noise reduction at the price of contrast, (2) the DCNN [7] was successfully trained to perform like [9], and (3) the original image [11] shows the best tone mapping result w.r.t. detail visibility.

Temporal Incoherence: Since the temporal incoherence measure is mainly sensitive to flickering, we simulate this artifact by randomly performing clipping to each image of a stack of 13 images. Clipping is applied in the same way as for the over-/underexposure measure, but the range for random clipping is successively increased from 0 to 0.3. The latter means that 30 % of the darkest and the brightest pixels are clipped and equalized leading to brighter or darker images. In Fig. 1 (h-i), we can see that the temporal incoherence gets larger with an increasing value range for intensity clipping. One negative aspect is the small value range especially for global temporal incoherence. Scaling could help here but we wanted to stay as close as possible to the original measure.

4.3. Reference TMOs

Four reference TMOs are used to further analyze the IR tone mapping IQA measures. The baseline comes from the datasets themselves as both provide 8- and 16-bit image pairs. The second is a video TMO proposed by Eilertsen *et al.* [9] that performed best in a recent survey [10]. We use our own implementation as there is no public code available. Then, we consider an approach based on Deep Convolutional Neural Networks (DCNNs) [7]. In the original paper, the goal is multitask learning to perform multiple image quality improvement techniques end-to-end. However, as mentioned in the introduction, the DCNN expects visual-optical 8-bit images as input and hence cannot perform tone mapping. So, we apply transfer learning and fine-tuning to the network and use IR images tone mapped by the second approach [9] as ground truth during training. Assuming successful training after 180 epochs (the loss curve looked fine), both TMOs should then roughly perform similar. The intention here is to discover whether the IQA measures together with a qualitative evaluation can verify this similar performance or not. Finally, we consider a recent approach [16] and use the provided Matlab code.

4.4. TMO Evaluation using the IQA measures

Table 2 shows the measures (1) TMQI, (2) over-/underexposure, (3) loss of contrast, (4) noise visibility, and (5) temporal incoherence for the four considered TMOs applied to the FLIR dataset. Since we used the training and the validation subset for DCNN training, we only evaluate on the test subset that consists of 4,224 frames in total. The numbers indicate multiple findings: the semi-objective TMQI measure clearly favors the FLIR baseline TMO. Liang's [16] images are relatively over-/undersaturated and weakly contrasted but with very good noise reduction. The DCNN [7] successfully learned to perform like [9] but produces more noise in comparison. The temporal incoherence is similar across all tested TMOs. Figure 2 shows qualitative results that confirm the numbers of Table 2. We consider this another indication that the evaluation measures are suitable to assess the quality of tone mapped images.

Similar numbers for the LTIR dataset actually did not provide as much insight as the numbers here. One reason might be that the LTIR dataset is too diverse as it is acquired by 7 different cameras. Hence, it is difficult to see clear tendencies in the TMOs and/or the IQA measures.

5. Conclusion and Future Work

We presented IQA measures for thermal IR video tone mapping. Among them we identified four that can be used for objective quality assessment, which is favored compared to subjective measures for real world applications such as surveillance. These four measures were analyzed in depth and applied to two public IR video datasets. The measures work as expected and can be used to assess the quality of IR video TMOs. Reference results on the FLIR dataset were provided for four state-of-the-art TMOs. We mostly discussed full reference measures that require the original HDR image as input. Since the TMQI performed considerably well, however, further blind tone mapping IQA measures [24, 39] should be evaluated in the future.

References

[1] Ayman Abaza, Mary Ann Harrison, Thirimachos Bourlai, and Arun Ross. Design and evaluation of photometric image quality measures for effective face recognition. *IET Biometrics*, 3(4):314–324, 2014. 3

[2] Nikola Banić and Sven Lončarić. Sensitivity of tone mapped image quality metrics to perceptually hardly noticeable differences. In *Fifth Croatian Computer Vision Workshop*, 2016. 2

[3] Francesco Banterle, Alessandro Artusi, Kurt Debattista, and Alan Chalmers. *Advanced high dynamic range imaging*. AK Peters/CRC Press, 2017. 3

[4] Amanda Berg, Jörgen Ahlberg, and Michael Felsberg. A thermal object tracking benchmark. In *IEEE AVSS*, 2015. 2, 6

[5] Ronan Boitard, Rémi Cozot, Dominique Thoreau, and Kadi Bouatouch. Survey of temporal brightness artifacts in video tone mapping. In *HDRi2014 - Second International Conference and SME Workshop on HDR imaging*, 2014. 3

[6] Jiawen Chen, Sylvain Paris, and Frédo Durand. Real-time edge-aware image processing with the bilateral grid. In *ACM Transactions on Graphics (TOG)*, volume 26, page 103. ACM, 2007. 2

[7] Qifeng Chen, Jia Xu, and Vladlen Koltun. Fast image processing with fully-convolutional networks. In *IEEE ICCV*, 2017. 1, 2, 6, 8

[8] Gabriel Eilertsen. *The high dynamic range imaging pipeline*. PhD thesis, Linköping University, Sweden, 2018. 2

[9] Gabriel Eilertsen, Rafał K. Mantiuk, and Jonas Unger. Real-time noise-aware tone mapping. *ACM Transactions on Graphics (TOG)*, 34(6):198:1–198:15, 2015. 1, 2, 6, 8

[10] Gabriel Eilertsen, Rafal Mantiuk, and Jonas Unger. A comparative review of tone-mapping algorithms for high dynamic range video. In *Eurographics*, 2017. 1, 2, 3, 4, 5, 8

[11] FLIR Systems. FREE FLIR Thermal Dataset for Algorithm Training. https://www.flir.com/oem/adas/adas-dataset-form/. [Accessed 8 March 2020]. 2, 6, 8

[12] Rikke Gade and Thomas B. Moeslund. Thermal cameras and applications: a survey. *Machine Vision and Applications*, 25(1):245–262, 2014. 1

[13] Andrs Restrepo Girn and Humberto Loaiza Correa. A new algorithm for detecting and correcting bad pixels in infrared images. *Ingeniera e Investigacin*, 30(2):197–207, 2010. 3

[14] Bo Gu, Wujing Li, Minyun Zhu, and Minghui Wang. Local edge-preserving multiscale decomposition for high dynamic range image tone mapping. *IEEE Transactions on Image Processing*, 22(1):70–79, 2013. 1

[15] Howard V. Kennedy. Modeling noise in thermal imaging systems. In *Proceedings of SPIE*, volume 1969, 1993. 4

[16] Zhetong Liang, Jun Xu, David Zhang, Zisheng Cao, and Lei Zhang. A hybrid l1-l0 layer decomposition model for tone mapping. In *IEEE CVPR*, 2018. 1, 2, 6, 8

[17] Chengwei Liu, Xiubao Sui, Xiaodong Kuang, Yuan Liu, Guohua Gu, and Qian Chen. Adaptive contrast enhancement for infrared images based on the neighborhood conditional histogram. *Remote Sensing*, 11(11):1381, 2019. 2

[18] Chengwei Liu, Xiubao Sui, Xiaodong Kuang, Yuan Liu, Guohua Gu, and Qian Chen. Optimized contrast enhancement for infrared images based on global and local histogram specification. *Remote Sensing*, 11(7):849, 2019. 2

[19] Haibo Luo, Lingyun Xu, Bin Hui, and Zheng Chang. Tone mapping infrared images using conditional filtering-based multi-scale retinex. In *Proceedings of SPIE*, volume 9675, 2015. 1, 2

[20] Rafal Mantiuk, Scott Daly, and Louis Kerofsky. Display adaptive tone mapping. *ACM Transactions on Graphics*, 27(3), 2008. 2

[21] Rafał K. Mantiuk, Kil Joong Kim, Allan G Rempel, and Wolfgang Heidrich. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Transactions on graphics (TOG)*, 30(4):40, 2011. 2, 4

[22] Albert A. Michelson. *Studies in Optics*. The University of Chicago Science Series. University Press, 1927. 3

[23] Rico Montulet and Alexia Briassouli. Deep learning for robust end-to-end tone mapping. In *BMVC*, 2019. 2

[24] Hossein Ziaei Nafchi, Atena Shahkolaei, Reza Farrahi Moghaddam, and Mohamed Cheriet. Fsitm: A feature similarity index for tone-mapped images. *IEEE Signal Processing Letters*, 22(8):1026–1029, 2014. 8

[25] Sung Ho Park and Ethan D. Montag. Evaluating tone mapping algorithms for rendering non-pictorial (scientific) high-dynamic-range images. *Journal of Visual Communication and Image Representation*, 18(5):415–428, 2007. 1

[26] Aakanksha Rana, Praveer Singh, Giuseppe Valenzise, Frederic Dufaux, Nikos Komodakis, and Aljosa Smolic. Deep tone mapping operator for high dynamic range images. *arXiv preprint arXiv:1908.04197*, 2019. 2

[27] Erik Reinhard, Greg Ward, Sumanta Pattanaik, and Paul Debevec. *High Dynamic Range Imaging: Acquisition, Display, and Image-based Lighting*. Morgan Kaufmann, 2010. 1

[28] Qi Shan, Jiaya Jia, and Michael S Brown. Globally optimized linear windowed tone mapping. *IEEE Transactions on Visualization and Computer Graphics*, 16(4):663–675, 2010. 1

[29] Richard Szeliski. *Computer Vision: Algorithms and Applications*. Springer, 2010. 1

[30] Jack Tumblin, Jessica K. Hodgins, and Brian K. Guenter. Two methods for display of high contrast images. *ACM Transactions on Graphics (TOG)*, 18(1):56–94, 1999. 2

[31] Minjie Wan, Guohua Gu, Weixian Qian, Kan Ren, Qian Chen, and Xavier Maldague. Infrared image enhancement using adaptive histogram partition and brightness correction. *Remote Sensing*, 10(5):682, 2018. 2

[32] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 2

[33] Zhou Wang and Alan C Bovik. Reduced-and no-reference image quality assessment. *IEEE Signal Processing Magazine*, 28(6):29–40, 2011. 2

[34] Zhou Wang and Qiang Li. Information content weighting for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 20(5):1185–1198, 2010. 2

[35] William L. Wolfe. *Handbook of military infrared technology*. Office of Naval Research, University of Michigan, 1965. 4

[36] Minghui Yang, Sihai Chen, Xin Wu, Wen Fu, and Zhangli Huang. Identification and replacement of defective pixel based on matlab for ir sensor. *Frontiers of Optoelectronics in China*, 4(4):434–437, 2011. 3

[37] Hojatollah Yeganeh. *Cross Dynamic Range And Cross Resolution Objective Image Quality Assessment With Applications*. PhD thesis, University of Waterloo, 2014. 2

[38] Hojatollah Yeganeh and Zhou Wang. Objective quality assessment of tone-mapped images. *IEEE Transactions on Image Processing*, 22(2):657–667, 2012. 2, 3

[39] Guanghui Yue, Chunping Hou, Ke Gu, Shasha Mao, and Wenjun Zhang. Biologically inspired blind quality assessment of tone-mapped images. *IEEE Transactions on Industrial Electronics*, 65(3):2525–2536, 2017. 8

[40] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. A comprehensive evaluation of full reference image quality assessment algorithms. In *IEEE ICIP*, 2012. 2

[41] Karel Zuiderveld. Contrast limited adaptive histogram equalization. In Paul S. Heckbert, editor, *Graphics gems IV*, pages 474–485. Academic Press Professional, Inc., 1994. 1, 2