SUBJECTIVE QUALITY ASSESSMENT OF THERMAL INFRARED IMAGES

Guanghui Yue[1], *Lixin Zhang*[1], *Jinxia Zhang*[2], *Zhaofei Xu*[2], *Shuigen Wang*[2], *Tianwei Zhou*[1], *Yuanhao Gong*[1], *and Wei Zhou*[3]

[1] Shenzhen University, Shenzhen 518060, China
[2] Yantai IRay Technology Co., Ltd., Yantai, 261400, China
[3] Cardiff University, United Kingdom

ABSTRACT

Thermal infrared images (TIIs) can be distorted by multiple factors, resulting in noise, low contrast, limited dynamic range, and fuzziness, which greatly impede their usefulness. It is crucial to evaluate the quality of TIIs. Unfortunately, there have been very few attempts to study this problem. In this study, we collected 1,000 authentically distorted TIIs using thermal infrared acquisition equipment and conducted strict subjective experiments to obtain a thermal infrared image quality assessment (IQA) database. Each image's quality score was obtained under strict scoring rules. Finally, we investigated the feasibility of several no-reference (NR) IQA methods in quality assessment of TIIs. We found that existing NR-IQA methods achieve ordinary performance in such a task, and there is an urgent need to develop a specific IQA methods for TIIs. The findings together with the constructed database are expected to pave the way for the development of more advanced IQA methods for further development of this field.

Index Terms— Thermal infrared images, image quality assessment, subjective assessment, no reference.

1. INTRODUCTION

Nowadays, the thermal infrared imaging technology is gradually spread to more areas, including but not limited to, the civil aviation, security, frontier defense, industry, building, transportation, and automation [1, 2]. Thermal infrared imaging is a technology that captures images of objects based on their emission, reflection, conduction, and radiation of infrared energy. However, due to the influence of equipment, environment and other factors, it inevitably causes image distortions, reduces the perceptual quality, and brings obstacles to the processing and subsequent applications of thermal infrared images (TIIs) [3–5]. Recently, proposing reliable quality assessment methods for TIIs has become an urgent requirement with the popularity of thermal infrared equipment.

Image quality assessment (IQA) is an essential task in computer vision and related fields because it plays a crucial role in various applications [6–8]. Generally, it can be divided into the subjective method and objective method [9–11]. The former evaluates the image quality by many qualified observers through the strict scoring rules and experiment protocols. It is often regarded as the golden standard of the objective method. Currently, the most common scoring rules for Natural Scene Images (NSIs) are based on a continuous impairment scale or five-grade rating scale, where the scale key points correspond to the quality level of 'Excellent', 'Good', 'Fair', 'Poor', and 'Bad'. Compared to NSIs, TIIs have several special characteristics, e.g., high background noise, low resolution, and low contrast, making the quality assessment task more challenging [12–14]. This motivates us to pay specific attention to the quality assessment of TIIs. However, despite the increasing research interest in IQA, research on the quality assessment of TIIs is still lacking.

In this study, we discuss the subjective quality assessment of TIIs and construct a new database, named TIIQAD. To keep the database consistent with the actual application, we collected 1,000 authentically distorted TIIs under different conditions. Considering the difference between TIIs and NSIs, we changed the scoring descriptions in the five-grade rating scale based on the physical principles and thermal characteristics of TIIs, paying more attention to information amount, contrast, and blur. The quality of each image is determined by the mean opinion score through rigorous subjective experiment and data processing. Based on the collected TIIQAD, we further investigate whether mainstream no-reference (NR) IQA methods designed for NSIs are competent for the quality assessment task of TIIs. Overall, the contributions of this paper are summarized as follows:

- We construct a new real-world database, named TIIQAD, for promoting the development of quality assessment of TIIs. The constructed TIIQAD includes 1,000 authentically distorted TIIs captured under different conditions, e.g., indoor/outdoor scenes and sunny/cloudy weather, using the thermal infrared acquisition equipment. Each image has a mean opinion score (MOS) obtained through strict subject experiments by 20 observers. The database will be released at `https://github.com/cheunglaihip`.

- We investigated the performance of mainstream NR-IQA methods on TIIQAD. Experimental results showed that mainstream NR-IQA methods are not very suitable for quality assessment of TIIs because TIIs usually contain complex distortions caused by factors from equipment and environmental conditions. In addition, methods designed for images with the single distortion generally have poorer performance compared to those designed for images with multiple distortions.

2. THE THERMAL INFRARED IMAGE QUALITY ASSESSMENT DATABASE

2.1. Image Collection

Fig. 1 shows the image acquisition platform. Specifically, we used the FTII640 camera (a TII camera produced by Yantai IRay Technology Co., Ltd.) to collect TIIs. The captured TIIs were initially stored in the RAW format. We used a laptop computer to display the collected images in real time for better understanding of the captured scene during the image acquisition process. For the convenience of subsequent tasks, we converted the RAW data to the BMP format while maintaining the original resolution of 640×512 pixels using the software provided by Yantai IRay Technology Co., Ltd. To ensure the scene diversity, we specifically eliminated TIIs with similar content and ultimately selected 1,000 authentically distorted TIIs.

During image collection, we mainly collected images under representative conditions in view of the practical applicability. Fig. 2 shows some examples of the captured TIIs, in which each row contains images collected in indoors, outdoors, sunny day, and cloudy day, respectively. As can be seen, the contrast of TIIs between indoors and cloudy day is generally lower than that between outdoors and sunny day. There are two potential factors that lead to this phenomenon. On the one hand, the temperature difference between indoor objects is relatively small, and the thermal sensitivity of the infrared equipment is not enough to detect subtle temperature variations. On the other hand, on cloudy day, the presence of most air acts as a hindrance to infrared radiation, which weakens the energy reflection of the objects.

2.2. Analysis of Image Distortions

Affected by multiple factors, acquiring high-quality TIIs is a challenging task, and the captured TIIs are usually with complex distortions. For instance, thermal equilibrium of objects, long wavelength of radiation, and atmospheric attenuation can cause the strong spatial correlation, low contrast, and blur of TIIs. Random interference from the external environment and the imperfection of the thermal infrared imaging system could also lead to various types of noise in TIIs. Fig. 2 illustrates some examples with representative distortions.

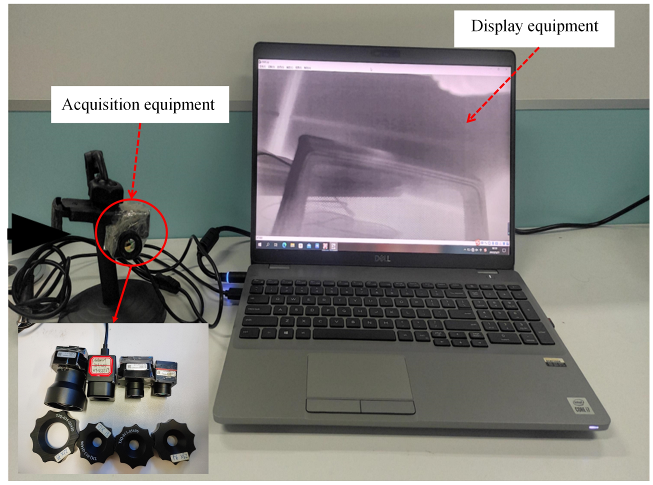

Fig. 1. TII acquisition platform.

- **Noise**: Noise in TIIs refers to random changes in brightness. The presence of noise reduces the definition and contrast of the image. In some cases, noise may cause incorrect recognition or misjudgment of targets.
- **Blur**: The primary causes of blur are improper settings of acquisition parameters, movement of equipment or targets, and issues of the optical lens. Blur can weaken the edge and detail information of TIIs, making the image unclear.
- **Low Contrast**: TIIs can suffer from low contrast distortion due to harsh environments, insufficient or uneven illumination, and low infrared radiation intensity of the object itself. Low contrast is not conducive to object recognition.
- **Other Distortions**: Apart from the distortions mentioned above, other distortions also occur during the TII collection. For instance, the limited thermal sensitivity can limit the dynamic range of TIIs, which affects thermal analysis. Reflection, refraction, and scattering can lead to information loss, which may affect the accurate detection of object surface temperature.

3. SUBJECTIVE QUALITY ASSESSMENT OF THERMAL INFRARED IMAGES

3.1. Subjective Test

Twenty observers (with normal or corrected-to-normal vision, 21-26 years) participated in the subjective experiment. During the experiment, observers sat in front of a Dell 27-inch screen, which has a resolution of 1920×1080 pixels, in

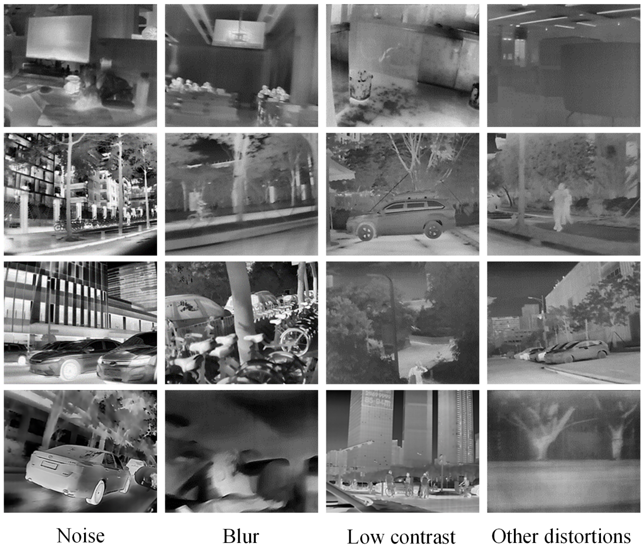

Fig. 2. Examples of TIIs in the TIIQAD.

a normal-lighting laboratory. A flexible viewing distance was set by considering practical applications, allowing observers to view the images from a distance of one to three times the image height [15].

The subjective experiment included two stages: a training stage and a testing stage. During the training stage, observers were required to understand the purpose, the operation process, as well as the scoring rules of the experiment. Before the formal testing phase, observers had to obtain 70% or higher accuracy on some prepared samples. It's important to note that these samples were not included in the test phase and the formal database. During the testing stage, the single stimulus method recommended by the international telecommunication union [16] was used. As there were no reference pairs for each test image, a five-level assessment scale was adopted, as shown in Table 1. The rating score was automatically record by a subjective scoring software, as shown in Fig. 3. Observers could click on the "Next" button to continue rating after completing the rating of the current image. To prevent visual fatigue and ensure the accuracy of experimental results, observers were encouraged to take breaks every 10 minutes. All images are presented randomly without repetition.

Table 1. Scoring rules for quality assessment of TIIs.

Ratings	Descriptions
1	Severely annoying, informationless.
2	Annoying, less information, low contrast.
3	Slightly annoying, little impaired information.
4	Distortion perceptible, full of information.
5	Distortion imperceptible, full of information.

3.2. Subjective data processing

Since the distortions of TIIs are complex, the observers may have distinct subjective rating scores due to their different interpretations of the task. Therefore, it is necessary to process subjective score data before determining the quality score of each image. For this purpose, we adopted the method recommended by SS in ITU-R BT.500 [16] to eliminate outliers and process the subjective score data, which improved the reliability of the obtained quality scores. After analysis, 3 out of 20 subjects were rejected. Then, a linear mapping function was used to scale the Z-Score to the range $[1, 5]$. Let $r_{p,q}$ denote the raw scores provided by the q-th subject on the p-th images, the Z-Scores are computed from the raw scores as follows:

$$Z_{p,q} = \frac{r_{p,q} - \mu_q}{\sigma_q}, \tag{1}$$

where $\mu_q = \frac{1}{N_q}\sum_{p=1}^{N_q} r_{p,q}, \sigma_q = \sqrt{\frac{1}{N_q-1}\sum_{p=1}^{N_q}(r_{p,q} - \mu_q)^2}$ and N_q is number of imgaes rated by the q-th subject. The MOS value for the p-th image was then calculated using this scaled score:

$$M_p = \frac{1}{Q}\sum_{q=1}^{Q} Z_{p,q}, \tag{2}$$

where $Q = 17$ represents the number of remaining valid subjects after eliminating outliers.

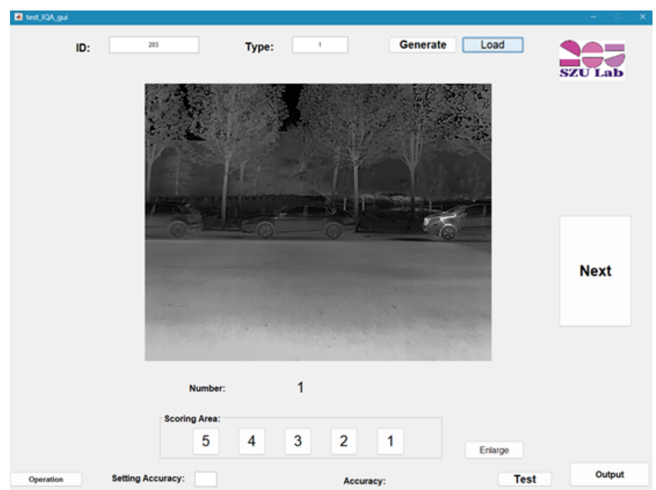

Fig. 3. Graphic user interface of the subjective test.

Fig. 4 presents the distribution of MOSs. As can be seen, TIIQAD covers a wide range of image quality levels, from extremely poor to satisfactory, and shows good differentiation across different quality levels. Notably, there are relatively few images with high scores, with the majority of images concentrated in the middle grades. This is because it's hard to obtain perfect image during the image collection process, due to various distortions. Furthermore, observers may be hesitant to rate an image too low or too high during the image observation process.

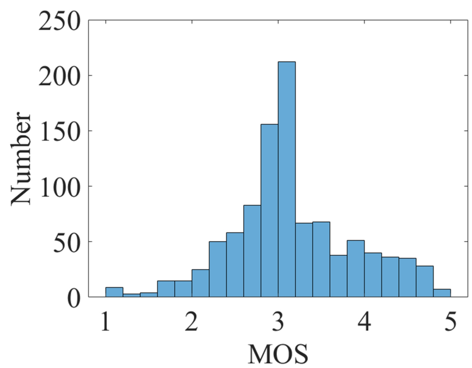

Fig. 4. The distribution of MOS values of TIIQAD.

4. PERFORMANCE COMPARISON OF MAINSTREAM NR-IQA METHODS

In this section, we investigate the feasibility of mainstream NR-IQA methods in tackling the quality assessment task of TIIs on the proposed TIIQAD. We first introduce the experimental settings, including the compared methods and the evaluation metrics, and then report the experiments results as well as the associated discussions.

4.1. Experimental Settings

Due to the limited number of objective methods designed specifically for TIIs, we select eight well-known NR-IQA methods designed for NSIs to investigate their feasibility in assessing the quality of TIIs. These objective methods can be classified into three categories according to their scope. The first category includes quality assessment methods designed specifically for contrast change, such as MDM [17] and NR-CDIQA [18]. The second category includes quality assessment methods designed for images with the synthetic distortion, such as BRISQUE [19], BIQME [20], GM-LOG [21], OG-IQA [22], and SSQE [23]. The last category includes a quality assessment method, namely GWH-GLBP [24], proposed for images with the authentic distortion.

To assess the performance of these NR- IQA methods, we use four commonly used evaluation metrics: Pearson linear correlation coefficient (PLCC), Spearman rank correlation coefficient (SRCC), Kendall rank correlation coefficient (KRCC), and root mean square error (RMSE). Before calculating PLCC and RMSE, a non-linear fitting of the predicted scores was necessary as recommended by the Video Quality Expert Group [25]:

$$f(x) = \kappa_1 \left[\frac{1}{2} - \frac{1}{e^{\kappa_2 (x-\kappa_3)} + 1} \right] + \kappa_4 \cdot x + \kappa_5, \quad (3)$$

where x and $f(x)$ are the predicted score obtained by an objective method and the corresponding subjective score, respectively. k_i is the parameter to be fitted using iterative least squares estimation. Theoretically, higher values of SRCC, KRCC, and PLCC, in contrast to RMSE, indicate a superior performance of the tested objective method.

4.2. Experimental Results and Analysis

Table 2 presents the experimental results of eight objective methods, from which we have the following observations. Firstly, BRISQUE, GWH-GLBP, BIQME and GM-LOG have SRCC and PLCC values over 0.85, and BRISQUE shows the best performance with SRCC of 0.895 and PLCC of 0.877. While these results are promising, there are still large room for performance improvement in quality assessment of TIIs. Secondly, all listed methods exhibit generally low KRCC values, ranging from 0.354 to 0.702. This suggests that the predicted scores of these objective methods have weak correlation with subjective scores. Finally, methods designed for images with the single distortion (e.g., NR-CDIQA and MDM) generally have poorer performance compared to those designed for images with multiple distortions (e.g., BRISQUE, BIQME, and GLBP). The above observed phenomenon can be attributed to several possible reasons. Firstly, TIIs often contain complex distortions that appear in a mixed way rather than a single occurrence. It is hard to effectively characterize the complex distortions based on a small number of features such as contrast and gradient may not be adequate. Most objective methods have limited potential in expressing complex distortions. Secondly, methods that consider multi-feature fusion and statistical features are more conducive to characterizing the distortions in TIIs. Nevertheless, the extracted features are still unable to fully characterize the complex distortions in TIIs. Thirdly, the scene content of TIIs is varying, and these methods do not comprehensively consider image complexity, and scene changes, which makes it difficult for existing objective methods to understand the real authentic distortions.

Table 2. Performance comparison of NR-IQA methods on the proposed TIIQAD.

METHODS	Evaluation metrics			
	PLCC	SRCC	KRCC	RMSE
SSQE [23]	0.498	0.503	0.354	0.621
NR-CDIQA [18]	0.568	0.561	0.399	0.586
MDM [17]	0.644	0.608	0.438	0.544
OG-IQA [22]	0.852	0.822	0.640	0.372
GM-LOG [21]	0.887	0.854	0.676	0.328
GWH-GLBP [24]	0.892	0.871	0.694	0.322
BIQME [20]	0.890	0.868	0.695	0.323
BRISQUE [19]	0.895	0.877	0.702	0.316

5. CONCLUSION AND FUTURE WORKS

Thermal infrared technology enables non-contact, high-efficiency, and high-precision thermal measurement and imaging, with a wide range of applications in fields such as industry, medicine, and security. However, the captured TIIs are usually with complex distortions caused by multiple factors. Filtering out low-quality TIIs can provide users with more realistic images, improve work efficiency, help manufacturers optimize infrared equipment parameters. This paper presents a preliminary discussion on the subjective quality assessment of TIIs. Firstly, we constructed a quality assessment database, which consists of 1,000 authentically distorted TIIs and theirs associated MOSs, by strict subjective experiments in a laboratory environment. We then investigated the performance of eight mainstream NR-IQA methods on this database. Experimental results show that these methods are not very suitable for the quality assessment of TIIs. Future works can start by analyzing and quantifying distortion characteristics of TIIs, particularly those caused by acquisition equipment, operators, and harsh weather conditions, and designing specific objective methods for TIIs.

6. ACKNOWLEDGEMENT

This work was supported in part by National Natural Science Foundation of China (No. 62371305), in part by Natural Science Foundation of Shenzhen (No. JCYJ20230808105906 013), in part by Guangdong Basic and Applied Basic Research Foundation (No. 2024A1515030025 and No. 2024A15 15030278), in part by Medicine Plus Program of Shenzhen University (No. 2024YG007), and in part by the Tencent "Rhinoceros Birds"-Scientific Research Foundation for Young Teachers of Shenzhen University.

7. REFERENCES

[1] W. Zhang, X. Sui, G. Gu, Q. Chen, and H. Cao, "Infrared thermal imaging super-resolution via multiscale spatio-temporal feature fusion network," *IEEE Sensors Journal*, vol. 21, no. 17, pp. 19 176–19 185, 2021.

[2] G. Zhai and X. Min, "Perceptual image quality assessment: a survey," *Science China Information Sciences*, vol. 63, pp. 1–52, 2020.

[3] G. Yue, C. Hou, K. Gu, T. Zhou, and G. Zhai, "Combining local and global measures for dibr-synthesized image quality evaluation," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 2075–2088, 2018.

[4] J. Zhang, X. Zhou, L. Li, T. Hu, and C. Fansheng, "A combined stripe noise removal and deblurring recovering method for thermal infrared remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.

[5] G. Yue, C. Hou, K. Gu, T. Zhou, and H. Liu, "No-reference quality evaluator of transparently encrypted images," *IEEE Transactions on Multimedia*, vol. 21, no. 9, pp. 2184–2194, 2019.

[6] W. Lin and C.-C. J. Kuo, "Perceptual visual quality metrics: A survey," *Journal of visual communication and image representation*, vol. 22, no. 4, pp. 297–312, 2011.

[7] T. Zhou, S. Tan, B. Zhao, and G. Yue, "Multitask deep neural network with knowledge-guided attention for blind image quality assessment," *IEEE Transactions on Circuits and Systems for Video Technology*, 2024.

[8] G. Yue, H. Wu, W. Yan, T. Zhou, H. Liu, and W. Zhou, "Subjective and objective quality assessment of multi-attribute retouched face images," *IEEE Transactions on Broadcasting*, 2024.

[9] Y. Zhao, J. Cheng, W. Zhou, C. Zhang, and X. Pan, "Infrared pedestrian detection with converted temperature map," in *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE, 2019, pp. 2025–2031.

[10] G. Yue, C. Hou, and T. Zhou, "Blind quality assessment of tone-mapped images considering colorfulness, naturalness, and structure," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 5, pp. 3784–3793, 2018.

[11] Y. He, B. Deng, H. Wang, L. Cheng, K. Zhou, S. Cai, and F. Ciampa, "Infrared machine vision and infrared thermography with deep learning: A review," *Infrared physics & technology*, vol. 116, p. 103754, 2021.

[12] T. Trongtirakul and S. Agaian, "Unsupervised and optimized thermal image quality enhancement and visual surveillance applications," *Signal Processing: Image Communication*, vol. 105, p. 116714, 2022.

[13] L. Lai, J. Chu, L. Leng *et al.*, "No-reference image quality assessment based on quality awareness feature and multi-task training," *Journal of Multimedia Information System*, vol. 9, no. 2, pp. 75–86, 2022.

[14] F. Erlenbusch, C. Merkt, B. de Oliveira, A. Gatter, F. Schwenker, U. Klauck, and M. Teutsch, "Thermal infrared single image dehazing and blind image quality assessment," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 459–469.

[15] T. Xiang, Y. Yang, and S. Guo, "Blind night-time image quality assessment: Subjective and objective approaches," *IEEE Transactions on Multimedia*, vol. 22, no. 5, pp. 1259–1272, 2019.

[16] B. Series, "Methodology for the subjective assessment of the quality of television pictures," *Recommendation ITU-R BT*, vol. 500, no. 13, 2012.

[17] H. Z. Nafchi and M. Cheriet, "Efficient no-reference quality assessment and classification model for contrast distorted images," *IEEE Transactions on Broadcasting*, vol. 64, no. 2, pp. 518–523, 2018.

[18] Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang, and G. Zhai, "No-reference quality assessment of contrast-distorted images based on natural scene statistics," *IEEE Signal Processing Letters*, vol. 22, no. 7, pp. 838–842, 2014.

[19] 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.

[20] K. Gu, D. Tao, J.-F. Qiao, and W. Lin, "Learning a no-reference quality assessment model of enhanced images with big data," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 4, pp. 1301–1313, 2017.

[21] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and laplacian features," *IEEE Transactions on Image Processing*, vol. 23, no. 11, pp. 4850–4862, 2014.

[22] L. Liu, Y. Hua, Q. Zhao, H. Huang, and A. C. Bovik, "Blind image quality assessment by relative gradient statistics and adaboosting neural network," *Signal Processing: Image Communication*, vol. 40, pp. 1–15, 2016.

[23] L. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal processing: Image communication*, vol. 29, no. 8, pp. 856–863, 2014.

[24] Q. Li, W. Lin, and Y. Fang, "No-reference quality assessment for multiply-distorted images in gradient domain," *IEEE Signal Processing Letters*, vol. 23, no. 4, pp. 541–545, 2016.

[25] J. Antkowiak, T. J. Baina, F. V. Baroncini, N. Chateau, F. FranceTelecom, A. C. F. Pessoa, F. S. Colonnese, I. L. Contin, J. Caviedes, and F. Philips, "Final report from the video quality experts group on the validation of objective models of video quality assessment march 2000," *Final report from the video quality experts group on the validation of objective models of video quality assessment march*, vol. 2000, 2000.