Application of Blind Image Quality Assessment Metrics To Pulsed Thermography

J. FLEURET^{1,2}, S. EBRAHIMI^{1,2}, C. IBARRA-CASTANEDO^{1,2}, X. MALDAGUE^{1,2}

¹Faculty of sciences and engineering, University Laval, Quebec City, Quebec, Canada, julien.fleuret.1@univ.laval.ca

²Laboratory of Computer Vision and Numerical Systems, Multipolar Infrared Vision team, Quebec City, Quebec, Canada

1 Abstract

This paper introduces an application of Blind Image Quality Assessment (BIQA) metrics to pulsed thermography. This study investigates the application of BIQA in association with state of the art methods in order to improve the detection of defects in PT sequences. The experiments show that BIQA may significantly improve the detection of defects when applied to steel, but fails to capture informative features on CFRP.

Keywords— Natural Scene Statistics, LWIR, BIQA, Blind Image Quality, active thermography, material, defect detection, Pulse Thermography, GMLOG, IRNDT

2 Introduction

Pulsed Thermography (PT) has become a very popular tool in the field of Non-Destructive Testing due to its relative low cost and its simplicity. PT consists in submitting a sample of interest to a high energy pulse. Often the pulse is produced with powerful flashes, and once the light reaches the sample surface, the sample will warm up and the heat will spread into the sample. The absorption of the energy into the material will vary significantly depending on the density of the material, the presence of inserts composed of other material, the presence of voids or cracks, as well as the structure of the material. This absorption is measured by a thermal cameras located behind the flashes. These phenomena will create artifacts, distorsion or noises and will influence the accuracy of the results of the processing methods. There is no PT database available which can provide a ground truth regarding the acquisition that could be used as a reference to assess the quality of an acquisition. Being able to assess the quality of images referenceless is the goal of Blind Image Quality Assessment (BIQA) metrics. BIQA for a natural image is still nowadays the subject of intensive research [1-4]. However the application of BIQA to other image modalities such as Long Wavelength InfraRed (LWIR) images is quite recent [5-7]. LWIR images have a higher dynamic range. Unlike natural images it is rare for a LWIR image to use the full scale of it dynamic range. However a LWIR generally uses several times more than a natural image which means than any approach developed for natural images that involve the of the texture is not suitable for LWIR images. Nevertheless some approaches initially developed for natural images can still be applied to any other modality as long as they do not extract features related to the texture or use statistical or dynamic range priors made for natural images. To the best of our knowledge BIQA has not been used in association with PT. This study offers a survey regarding the interest of application of BIQA in order to improve the detection of defects in image sequences related to PT experiments. The next section reviews the BIQA we chose for our study. The application of BIQA to the processing of data acquired during PT is detailed in section 4. Section 5 describes the experiment and details the methodology regarding the study. The results are introduced in section 6 and discussed in section 7. Section 8 concludes this study.

3 Blind Image Quality Assessment Metrics

BIQA involves any kind of algorithm that is able to associate a score to an image without needing any other source information. Among the many BIQA that the literature offers we selected two of them. Goodall et al. [5] have focussed on the application of Natural Scene Statistics (NSS) to Long Wavelength InfraRed (LWIR) images. NSS was used by Mittal et al. [2] in the method BRISQUE which was among the approaches able to give a quantitative result while not being computationally intensive. Goodall et al. try to offer an adaption of the approach described by Mittal et al. that fits with the statistics of LWIR images. We refer to the work of Goodall et al. [5] as LISA (LwIr nSs biqA). Another work of interest is the algorithm GMLOG-BIQA proposed by Xue et al. [3]. Xue et al. [3] compute the score based on two different gradient images computed from the input



Licence: https://creativecommons.org/licenses/by/4.0/deed.en

which are later normalized before the computation of the feature vector. This work does not directly use the texture of the image and normalizes the edges in a dynamic range of ten bins which gives this work a particular feature regarding multimodal applications. For example a statistical learner can be trained to associate vectors to scores using vectors computed on a database of colour and be able to give meaning to the full score when applied to LWIR images.

In the rest of this section we are offering an introduction to LISA and GMLOG-BIQA from a more theoretical point of view.

LISA

In their work Goodall et al. [5] investigate the application of NSS to LWIR images. To do so they try to update the method proposed by Mittal et al. [2]. They observed that LWIR images have statistical regularities which are also observed in visible images. Nevertheless the statistics related of each of the modalites are very different. To compute the NSS of an image the first step consists in calculating the Mean-Subtrated Contrast Normalized (MSCN) image (1).

$$\hat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + C}$$
(1)

In equation (1), the indices *i* and *j* are the spatial index over the height and width of the image $(i \in 1, \dots, M, j \in 1, \dots, N)$. The variables μ and σ represent respectively the weighted mean and weighted standard deviation of the local luminance taken in a window $(K \subset M, L \subset N)$.

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

$$\sigma(i,j) = \sqrt{\sum_{k=-K}^{K} \sum_{l=-L}^{L} \omega_{k,l} I_{k,l}(i,j) - \mu(i,j)}$$
(3)

In equations (2) and (3) the variable $\omega = \{\omega_{k,l} | k = -K, \dots, K, l = -L, \dots, L\}$ represents a circulary-symmetric normalized gaussian weighting function. Goodall et al. highlight the fact that the histogram of the resulting image is similar to the one of a natural image. Once the MSCN has been calculated Mittal et al. modeled the local distribution using Generalized Gaussian Distribution (GGD) based on the assumption that the distorsion of coefficients of the MSCN image could be enough to determine the type of distorsion present in the image.

$$f(x;a,\sigma^2) = \frac{\alpha}{2\beta\Gamma\alpha^{-1}}\exp(-(\frac{|x|}{\beta})^a)$$
(4)

$$\beta = \sigma \sqrt{\frac{\Gamma(\frac{1}{\alpha})}{\Gamma(\frac{3}{\alpha})}} \tag{5}$$

$$\Gamma(\alpha) = \int_{0}^{\infty} t^{a-1} \exp^{-t} \partial t, \ a > 0$$
(6)

Distorsions have an influence on the statistical relationship of a pixel with its neighbours. To take this into consideration in the feature vector, Mittal et al. compute the Approximate Generalized Gaussian Distribution for the paired products along each direction.

$$f(x;\nu,\sigma_l^2,\sigma_r^2) = \begin{cases} \frac{\nu}{(\beta_l+\beta_r)\Gamma(\frac{1}{\nu})} \exp\left(-\left(\frac{-x}{\beta_l}\right)^{\nu}\right) & x < 0\\ \frac{\nu}{(\beta_l+\beta_r)\Gamma(\frac{1}{\nu})} \exp\left(-\left(\frac{-x}{\beta_r}\right)^{\nu}\right) & x \ge 0 \end{cases}$$
(7)

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

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

In equations (8) and (9) the parameters σ_l and σ_r are scale parameters. In order to improve the feature vector Goodal et al. used the approach proposed by Zhang et al. [8] involving seven log-derivative coefficients. These coefficients are more representative of the high frequency distorsions.

$$J(i,j) = log(|I(i,j)| + K)$$

$$PD_1(i,j) = J(i,j+1) - J(i,j)$$
(10)
(11)

$$PD_{1}(i,j) = J(i,j+1) - J(i,j)$$

$$PD_{2}(i,j) = J(i+1,j) - J(i,j)$$
(11)
(12)

$$PD_{2}(i, j) = J(i+1, j) - J(i, j)$$

$$PD_{2}(i, j) = J(i+1, j+1) - J(j, j)$$
(12)
(13)

$$PD_{3}(i,j) = J(i+1,j+1) - J(i,j)$$

$$PD_{4}(i,j) = J(i+1,j-1) - J(i,j)$$
(13)
(14)

$$PD_{5}(i, j) = J(i-1, j) - J(i+1, j) - J(i, j-1) - J(i, j+1)$$
(11)

(11)

(11)

$$PD_{6}(i,j) = J(i,j) - J(i+1,j+1) - J(i,j+1) - J(i+1,j)$$
(16)

$$PD_{7}(i,j) = J(i-1,j-1) - J(i+1,j+1) - J(i-1,j+1) - J(i+1,j-1)$$
(17)

In equation (10) the variable K is used to stabilize the equation. The last step of the creation of the feature vector consists in using a perceptual model which captures the band-pass characteristics of the log-derivation by means of a steerable pyramid decomposition.

The last step of the work of Goodall et al. was to conduct a psychovisual experiment on several datasets from which they could compute scores by using the Difference of Mean Opinion Score (DMOS). Then they trained a Singular Value Regressor using the features vector as predictor variable and the scores are given by the observers.

This work offered a quality assessment method that takes into account the statistical properties of a LWIR image. Some other works initially designed for natural image can still be used as long as they are based on a transformation or a derivative of the input image.

GMLOG-BIQA

In their work Xue et al. [3] use the Gradient Magnitude (GM) and the Laplacian of Gaussian (LoG) to compute a quality descriptor.

$$\mathbf{G}_{\mathbf{I}} = \sqrt{[\mathbf{I} \otimes \mathbf{h}_{\mathbf{x}}]^2 + [\mathbf{I} \otimes \mathbf{h}_{\mathbf{y}}]^2}$$
(18)

$$\mathbf{h}_{\mathbf{d} \in [\mathbf{x}, \mathbf{y}]} (\mathbf{x}, \mathbf{y} | \sigma) = \frac{\partial}{\partial \mathbf{d}} \mathbf{g}(\mathbf{x}, \mathbf{y} | \sigma)$$

$$= \frac{\partial}{\partial d} \left(-\frac{1}{2\pi\sigma} e^{-\frac{x^2 + y^2}{2\sigma^2}} \right)$$

$$= -\frac{1}{2\pi\sigma^2} \frac{d}{\sigma^2} e^{-\frac{x^2 + y^2}{2\sigma^2}}$$
(19)

$$\mathbf{L}_{\mathbf{I}} = \mathbf{I} \otimes \mathbf{h}_{\mathbf{LoG}} \tag{20}$$

$$\mathbf{h}_{\mathbf{LoG}}(\mathbf{x}, \mathbf{y}|\sigma) = \frac{\partial^2}{\partial x^2} \mathbf{g}(x, y|\sigma) + \frac{\partial^2}{\partial y^2} \mathbf{g}(x, y|\sigma)$$
(21)

$$= -\frac{1}{2\pi\sigma} \left[\frac{\partial^2}{\partial x^2} \left(e^{-\frac{x^2 + y^2}{2\sigma^2}} \right) + \frac{\partial^2}{\partial y^2} \left(e^{-\frac{x^2 + y^2}{2\sigma^2}} \right) \right]$$
$$= \frac{1}{2\pi\sigma} \frac{x^2 + y^2 - 2\sigma^2}{\sigma^4} e^{-\frac{x^2 + y^2}{2\sigma^2}}$$

The GM (18) gives information regarding the strength of luminant changes, the LoG (20) shows the intensity contrast considering a close neighbourhood. A joint adaptative normalization is used in order to obtain stable statistical features descriptors. The result of the joint adaptative normalization is organized as a two channel image. Each channel is quantified into a number of levels adapted to its dynamic.

$$\mathbf{p} \subset \mathbf{P} \in \mathbb{Z}^{2}$$

$$x = \mathbf{G}_{\mathbf{I}}(\mathbf{p})$$

$$y = \mathbf{L}_{\mathbf{I}}(\mathbf{p})$$

$$\mathbf{h}(y, x) = \mathbf{h}(y, x) + 1$$
(22)

Thus a bivariate histogram is computed as represented in equation (22) where \mathbf{P} is the set of all pixels. In order to reduce the dimensionality, the marginal distributions are calculated (23).

$$\mathbf{P}_{\mathbf{G}}(x) = \sum_{i=1}^{N} \mathbf{h}(x, i)$$
$$\mathbf{P}_{\mathbf{L}}(y) = \sum_{i=1}^{M} \mathbf{h}(i, y)$$
(23)

$$\mathbf{D} = \frac{\mathbf{h}}{P_G P_L} \tag{24}$$

$$\mathbf{Q}_{\mathbf{G}}(x) = \frac{\mathbf{P}_{\mathbf{G}}(x)}{N} \sum_{n=1}^{N} \mathbf{D}(x, n)$$
$$\mathbf{Q}_{\mathbf{L}}(y) = \frac{\mathbf{P}_{\mathbf{L}}(x)}{N} \sum_{m=1}^{M} \mathbf{D}(m, y)$$
(25)

The marginal distributions are then used to compute a dependency measure (24). The measure is finally used to compute the independency distributions (25).

The very last step consists to train a linear Suport Vector Regression (SVR) between the independency distributions and marginal distribution variables with the Difference Mean Opinion Score (DMOS).

4 Application of the BIQA to pulsed thermography

In the pulsed thermography experiment the acquisition starts as a very short moment before the emission of the pulse. The aim of this short moment is to ensure that the acquired data does contains absorption of the whole energy by the sample of interest. However during the processing, the number of frames to process is reduced based on the experience of the user regarding the material and the sample. The study does not have the goal of automatically finding the perfect subsample of the sequence which will give the best results. However we instead wish to compare the difference between reducing the data acquisition to N frames starting from moment t and selecting the N best images starting from moment t.

5 Method

The aim of this study is to evaluate the interest of BIQA to improve the results of the detection of defects in materials. The experiments have been conducted on two different types of materials: a CFRP sample and a steel sample. A CFRP sample contains twenty-five Teflon inserts. These defects are divided into five batches of five defects. Each batch has five inserts having the same depth but different sizes (from 3x3 mm² to 15x15 mm²). Also, as illustated in figure 1, and 2 each batch has been positioned at a specific depth (from 0.2 to 1 mm). The steel sample contains four flat-bottomed holes, each of them has a size of 30x30 mm², each one located at a different depth (from 1 to 2.5 mm). The materials have been investigated by a classic pulse thermographic procedure. Each material has been stimulated from the front side by a pulse generated by two photographic flashes (Balcar FX 60, 5 ms thermal pulse, 6.4 kJ/flash). A mid-wave infrared (MWIR) camera FLIR Phoenix (1.5 μm to 5.0 μm , 345 Hz, 14 bit per pixel, 640×512) was used for data acquisition. For each experiment, a sequence of the 2000 frames was acquired during a period of 30 seconds. That duration ensured that both the warm-up and the cool-down period had been acquired by the sensor.

Once the raw data has been acquired, three test sets are generated. The first processing consists in computing the cold image, shrinking the datasets in order to only keep the data that has been acquired after the pulse, and subtracting the cold image. Then the dataset was trimmed a second time in order to try to keep the N most representative images. Three datasets were generated, containing respectifvely the N first images, the N images with the best score considering the LISA metric and the N images with the best score considering the GMLOG-BIQA metric. Pulsed Phase Thermograpy (PPT) [9] and Principal Component Thermography (PCT) [10] algorithms were used to process each dataset. In our experiments we chose N = 400.

In order to assess the interest of with the proposed approach, it was compared with the state of the art two metrics were selected. In order to assess the quality of the detection we chose to use the Signal over Noise



Figure 1: CFRP sample used for the experiments.



Figure 2: Steel sample used for the experiments.

Ratio (SNR). We use the formulation proposed by Usamentiaga et al. [11]

$$SNR(roi_s, roi_n) = \frac{|\mu_s - \mu_n|}{\sqrt{\frac{(\sigma_s^2 + \sigma_n^2)}{2}}}$$
(26)

6 Results

In order to illustrate the interest of the proposed approach, we computed the the SNR for each case and for each defect. The results are reported in tab 1 and tab 2. The experiments consisted in applying PCT and PPT to the acquired sequences. From both we kept the best images from a visual point of view and then computed the metrics. Figure 5 shows the score computed for all the frames of the acquired sequence from both GMLOG and LISA metrics, for each sample. Figures 3 and 4 show the best image selected for the processing of both PCT and PPT for each situation.



Figure 3: These images represent the best component image of the PCT. In the first column the PCT has been applied on a sequence of 400 images following the pulse. In the second column the PCT has been applied on the 400 best images following the pulse considering metric LISA. The last column corresponds to the PCT having been applied on the 400 best images following the pulse considering metric GMLOG. The sample used in the first row is a steel sample, while in the second row it is a CFRP sample.





Figure 4: These images represent the best image of the PPT. In the first column the PPT has been applied on a sequence of 400 images following the pulse. The second column the PPT has been applied on the 400 best images following the pulse considering metric LISA. The last column corresponds to the PPT has been applied on the 400 best images following the pulse considering metric GMLOG. The sample used in the first row is a steel sample, that of the second row is a CFRP sample.

7 Discussion

As visible in figure 5, both GMLOG and LISA make it easy to find the pulse from the raw thermal data for both the CFRP and the steel sample. The results obtained using LISA seem to be easier to interpret. It is interesting to see that the CFRP sample seems to show quasi periodical variations of energy diffusion even late after the end of the stimulation. These variations are visible with both metrics, nevertheless the cause of these variations is hard to explain. Their quasi-periodic apparition may highlight a feature from the thermal camera in these experiments. However such a feature should have also been visible on the acquisition related to the steel sample which is not the case. Further investigation is needed in order to see the cause as well as the potential interest of these variations in energy. Figures 4 and 3 show the best images chosen after the PCT and PPT on each sequence. It is visible on both PPT and PCT data that the dataset created using LISA and GMLOG provides a significant improvement in comparison to the traditional approach on the steel sample, while having almost no improvement at all on the CFRP. These observations are confirmed by table 1 and table 2. Table 1 shows that the datasets created using LISA and GMLOG have a higher mean SNR and a lower standard deviation. One can note that in both datasets the SNR for the smallest defects are higher than on the results computed from the arbitrarily selected images.

The results obtained do not allow us to definitely conclude concerning the interest of BIQA in NDT. An obvious conclusion would be that the proposed study shows significant improvement in metalic material, but very poor results in composites. However in order to reach such a conclusion more materials must be investigated, which will be done in further research.

8 Conclusion

In this paper we investigated the use of two blind image quality assessment metrics to PT. These metrics were used to select a subset composed of the 400 best images selected after the emission of the pulse based on their score. The results were compared with a subset of 400 images selected after the pulse. The experiments were conducted on two reference samples one of CFRP and one made of steel. This study shows that BIQA

		Lateral size (mm)	0.2	0.4	0.6	0.8	1.0
		Depth (mm)		••••	0.0	0.0	
PCT	GMLOG	3	2.5457	2.1235	2.2839	1.9186	1.8704
		5	2.2169	2.3039	1.9359	1.9937	2.0865
		7	2.1131	2.5404	1.9554	2.408	1.9497
		10	2.7378	1.9537	1.9813	2.2446	2.3126
		15	2.0757	2.2433	2.3975	2.3277	2.2852
	LISA	3	2.6225	2.0606	2.4509	1.9099	1.8749
		5	2.139	2.3749	1.9444	2.0829	2.1787
		7	2.157	2.6531	1.9736	2.3791	1.9433
		10	2.5292	1.9534	1.9988	2.3513	2.2414
		15	2.0806	2.1902	2.3906	2.2958	2.2778
	Raw	3	2.8926	2.0936	2.3501	1.3413	1.2959
		5	2.1216	2.2928	1.6331	1.7372	2.1451
		7	2.179	1.9468	1.6074	2.1424	1.9296
		10	2.6596	1.6982	1.8191	1.4873	1.8921
		15	2.2947	1.9256	2.1939	2.2048	2.2208
	GMLOG	3	2.416	2.1601	2.4525	1.995	2.0978
		5	2.4933	2.4715	1.8654	2.1702	2.1406
		7	2.6442	2.2192	1.9385	2.4401	2.0686
		10	2.9587	1.8626	2.3717	2.1763	2.292
PPT		15	2.4697	2.5563	2.9108	2.2157	2.649
	LISA	3	2.4953	2.2114	2.3799	1.9632	1.9124
		5	2.0145	2.3849	1.9613	2.1887	2.218
		7	2.267	2.4227	1.951	2.4706	2.0763
		10	2.8645	2.061	2.197	2.4544	2.3762
		15	2.0865	1.9825	2.4136	2.3934	2.3947
	Raw	3	2.5529	1.9458	2.3	1.7424	1.8191
		5	2.2609	2.5225	1.9058	1.8785	1.5492
		7	1.9798	2.3988	1.9275	1.9844	1.4886
		10	3.0723	2.0748	2.234	1.9355	1.749
		15	2.0383	2.5733	2.3275	2.141	1.8662

Table 1: SNR computed for the CFRP sample.

		Lateral size (mm)	20	
		Depth (mm)	30	
	GMLOG	1.0	0.3517	
		1.5	0.3511	
		2.0	0.027	
		2.5	0.0457	
Ŀ	-	1.0	1.5123	
Ğ	S	1.5	1.5635	
H	L1	2.0	1.7049	
		2.5	1.113	
		1.0	0.3146	
	aw	1.5	0.3235	
	R	2.0	0.1401	
		2.5	0.4404	
	TOG	1.0	0.6558	
		1.5	0.9648	
	M'	2.0	0.7028	
	9	2.5	0.6907	
T	-	1.0	0.3539	
, D	$S_{\vec{r}}$	1.5	0.4872	
H	ΓI	2.0	0.2923	
		2.5	0.4935	
		1.0	0.4104	
	aw	1.5	0.1676	
	8	2.0	0.2321	
		2.5	0.288	

Table 2: SNR computed for the steel sample.

we selected for this study allows a significant improvement in terms of defect detections on the steel sample, but failed to capture informative data on the CFRP sample. The interest of using BIQA in the processing of PT data will depend mostly of the materials as well as the kind on features the BIQA uses in order to provide its score.

Acknowledgement

This research was supported by the Fonds Quebecois de Recherche - Nature et Technologie (FQRNT). The authors would like to thank Annette Schwerdtfeger for the time and energy she spent reviewing this article.

References

- [1] N Venkatanath, D Praneeth, Maruthi Chandrasekhar Bh, Sumohana S Channappayya, and Swarup S Medasani. Blind image quality evaluation using perception based features. In 2015 Twenty First National Conference on Communications (NCC), pages 1–6. IEEE, 2015. 1
- [2] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, 2012. 1, 2
- [3] Wufeng Xue, Xuanqin Mou, Lei Zhang, Alan C Bovik, and Xiangchu Feng. Blind image quality assessment using joint statistics of gradient magnitude and laplacian features. *IEEE Transactions on Image Processing*, 23(11):4850–4862, 2014. 1, 3
- [4] Aleksandr Shnayderman, Alexander Gusev, and Ahmet M Eskicioglu. An svd-based grayscale image quality measure for local and global assessment. *IEEE transactions on Image Processing*, 15(2):422– 429, 2006. 1
- [5] Todd Richard Goodall, Alan Conrad Bovik, and Nicholas G Paulter. Tasking on natural statistics of infrared images. *IEEE transactions on image processing*, 25(1):65–79, 2016. 1, 2
- [6] David Eduardo Moreno-Villamarín, Hernán Darío Benítez-Restrepo, and Alan Conrad Bovik. Predicting the quality of fused long wave infrared and visible light images. *IEEE Transactions on Image Processing*, 26(7):3479–3491, 2017.
- [7] Camilo Gerardo Rodríguez-Pulecio, Hernán Darío Benítez-Restrepo, and Alan Conrad Bovik. Making long-wave infrared face recognition robust against image quality degradations. *Quantitative InfraRed Thermography Journal*, pages 1–25, 2019. 1
- [8] Yi Zhang and Damon M Chandler. No-reference image quality assessment based on log-derivative statistics of natural scenes. *Journal of Electronic Imaging*, 22(4):043025, 2013. 3
- [9] Xavier Maldague and Sergio Marinetti. Pulse phase infrared thermography. *Journal of applied physics*, 79(5):2694–2698, 1996. 4
- [10] Nikolas Rajic. Principal component thermography. Technical report, DEFENCE SCIENCE AND TECH-NOLOGY ORGANISATION VICTORIA (AUSTRALIA ..., 2002. 4
- [11] R Usamentiaga, Clemente Ibarra-Castanedo, and X Maldague. More than fifty shades of grey: Quantitative characterization of defects and interpretation using snr and cnr. *Journal of Nondestructive Evaluation*, 37(2):25, 2018. 5