

Comparison and ranking procedure for an objective assessment of thermographic NDE methods

by D.L. Balageas*, J.-M. Roche**, F.-H. Leroy**

* ENSAM ParisTech, I2M Bordeaux, TREFLE Dept., 33400 Talence cedex, France
daniel.balageas@u-bordeaux.fr

** ONERA, Composite Materials and Structures Dept., 92320 Châtillon, France

Abstract

The present work aims at setting up a procedure of comparison and ranking of thermographic NDE methods that achieves a substantial progress in terms of objectivity and completeness compared to the scarce comparative studies found in the literature. The comparison is based on three metrics instead of the sole signal-to-noise ratio and, starting from a high number of defect signatures, leads to a unique figure to characterize the merit of each assessed method. For illustration purposes, the procedure is applied to the comparison of four well-known thermographic methods (TSR, PCT, PPT and HOS) used for the processing of data resulting from a front-face pulse-heating experiment. The studied sample is a C/epoxy plate with artificial defects simulating delaminations.

1. Context

The reasoned choice of the optimal data processing of NDE experiments by infrared thermography is still unsettled because of the lack of comparative studies between all available methods. This choice, however, remains crucial, both for academic researchers and industry engineers, since it has a significant influence on the relevance of the resulting image which the final diagnosis is based on. Depending on the selected approach, the inspection by active thermography may lead to results close to the ultrasonic ones, or, on the contrary, be declared inefficient.

2. Philosophy of the approach

The present work wants to set up an objective, thorough, multiple metrics-based procedure of comparison and ranking of thermographic NDE methods. The philosophy of the approach, which clearly differs from the usual comparative studies found in the literature that are based on the sole signal-to-noise ratio (SNR), is the following:

- the diagnosis is supposed to be based on the observation of one or several grey-scale images;
- the defects are multiple and representative of the variability in size, depth and severity of real defects;
- the metrics characterizing the image qualities are: the SNR and the edge sharpness of the contrasts created by the defects, and the accuracy of the defect lateral extent estimated from the full-width half-maximum contrast (FWHM);
- the comparison leads to a unique figure to characterize each assessed method and rank them by order of merit.

Figure 1 represents the full procedure which will be detailed further.

3. Definition of the three metrics

To characterize the defect signature, almost all authors consider the mean value of the contrast in the defective zone, although these limits progressively become difficult to localise when the defect depth increases. In the present study, the analysis is focused on the fitted contrast profile along a line or a column of the image crossing the centre of the defect. This analysis consists in the assessment of the three metrics detailed in the following subsections.

3.1. Contrast-to-noise ratio

The contrast-to-noise ratio (CNR) is defined as the ratio between the maximum of the square of the contrast in the defect zone along the profile, $(C_{A_{max}})^2$, and the square of the standard deviation of the signal in the surrounding sound zone, $(\sigma_s)^2$: $CNR = (C_{A_{max}}/\sigma_s)^2$. It can also be defined in decibels, as follows: $CNR_{dB} = 10 \log_{10}[(C_{A_{max}}/\sigma_s)^2] = 20 \log_{10}(C_{A_{max}}/\sigma_s)$.

In practice, for the search of the maximum contrast of a defective zone, two cases are possible (see Figures 1a and 1a'): i) the defect has a high aspect ratio $AR = D_d/z_d$ (D_d being the lateral extent of the defect, and z_d its depth) so that its contrast signature shows a plateau (Figure 1a), ii) the defect has a low aspect ratio, so that its contrast signature is Gaussian-like shaped (Figure 1a').

The baseline, the plateau and the edges are locally fitted by least-mean square polynomials (generally of degree 1). When no plateau can be defined, a fitting operation is performed, using a polynomial of degree 2 or higher. The fitted values obtained for the central zone of the defect are then used to assess either the contrast between the plateau and the baseline, in the case of a high aspect ratio, or the maximum contrast if no plateau is present, in the case of a low or medium aspect ratio. The noise is assimilated to the standard deviation of the points of the baseline, σ_s , allowing the evaluation of the CNR.

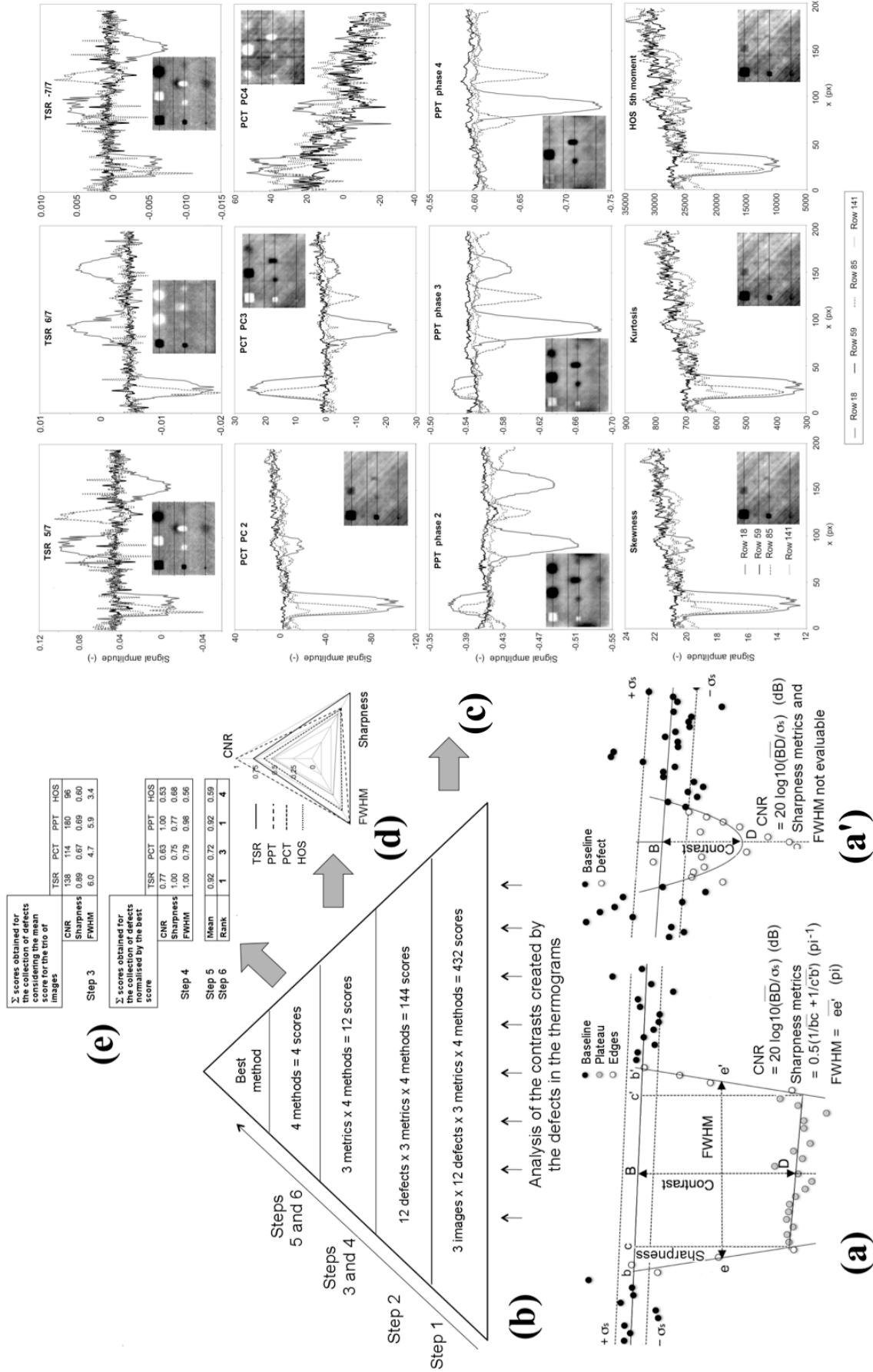


Figure 1. Flow diagram of the comparison and ranking procedure. (a) Analysis of the defect signatures in the case of a plateau; (a') Idem without plateau; (b) Data compression in 6 steps; (c) Assessed defect signatures (d) Results at the end of step 4; (e) Results of steps 3 to 6 leading to the final ranking.

3.2. Edge sharpness

Almost all of the works aiming at quantitatively comparing data processing methods for thermographic NDE are based on the sole CNR parameter. The sharpness is not taken into account, which may be explained by the difficulty to characterize it by a proper metrics and by the fact that in most studies, due to their standard shapes, square or circular, the defects are quite easy to detect. Yet, there is no denying that it is a relevant parameter regarding the quality of the infrared image.

One NDE study, in which the sharpness of defects detected in infrared images is measured, is found in the literature [1]. The metrics chosen to assess the defect sharpness is the magnitude of the gradient at the border between a defective zone and its sound zone. The gradients are calculated by finite differences from the pixel intensities of this zone, along orthogonal directions (columns and rows). The sharpness metrics is defined as the square root of the sum of the squares of the x and y gradients. This evaluation is made directly on the thermogram when the contrast is maximum so that the CNR and sharpness metrics are calculated at the same time.

In the present work, the sharpness metrics is calculated from the locally fitted signal, according to the formula given in Figure 1a. The metrics is normalised to make the comparison easier between the different defect signatures. For that, the gap between the plateau of contrast, when existing, and the sound material baseline is set equal to 1 for each defect, resulting a metrics value between 1 (when there is no pixel between the fitted plateau and the fitted baseline) and $1/(n+1)$ (when there are n such pixels). An arithmetic mean is calculated for both sides of the plateau for a given line (row or column). In the case of a medium or low aspect ratio, no identification of the sharpness is made, since there is no plateau.

3.3. Defect lateral extent

Pulse infrared thermographic images contain defect images of time-dependent dimensions. Defects seem to shrink when time increases, due to edge effects caused by lateral heat diffusion. Almond et al. [2,3] have shown that, in the case of pulse-heating, the lateral dimensions of a defect can be deduced from its full-width at half-maximum contrast (FWHM) linearly extrapolated to the "origin" of time, defined as the pulse occurrence time, although the present authors would strongly recommend to consider the time of contrast emergence instead. The FWHM contrast can be directly measured on the images as a function of the square root of time; the extrapolation, similarly to what is proposed by the early detection approach to identify defect depths [4], then mitigates the effects of the lateral diffusion. As pointed out by the authors, the method fails for defects of very low aspect ratios. In fact, it just cannot be applied when the plateau disappears (Gaussian-shaped contrasts).

In the present case, only one image is to be considered, a coefficient image, not a sequence of time-evolving thermogram images. By analogy with Almond's method, the FWHM contrast is measured in the TSR coefficient images. That would be *a priori* less efficient than the original Almond's method since the coefficient images are influenced by the full thermogram ($\Delta T(t)$), including late information suffering from diffusion effects. Furthermore, this identification is only applied to defect images presenting a plateau, such as defect #1-1, as illustrated by Figure 1a.

The metrics used to quantify the accuracy on the identified defect dimension is calculated as the logarithm of the inverse of the relative error on the identified defect dimension: $\text{Log}_{10}(D_d / \Delta D_d)$.

3.4. Profile analysis strategy

Considering the structure of the thermograms, the simplest way to analyse the contrasts using the three metrics consists in choosing the rows and/or the columns of pixels crossing the defect signature centres. A comparison of results obtained considering separately rows and columns is reported in [5], which concludes to the similarity of the results and the possibility to consider only the rows, so that the number of profiles to analyse can be reduced. This strategy is applied here.

4. Description of the procedure and application to pulse-heating data

4.1. Initial thermal data

The procedure is illustrated on data resulting from the front-face pulse-heating of a $270 \times 200 \times 5.25 \text{ mm}^3$ carbon/epoxy plate with artificial delamination-like defects of various sizes and depths (Figure 2a), already used for various studies [6,7]. The test was performed with two Elinchrom flash lamps, delivering 6 kJ in 4 ms; temperature was recorded by a CEDIP Jade LWIR camera, on a 320×256 pixels frame, with an integration time of 230 μs , over 60 s, at a frequency of 200 images per second. Four data processing methods were applied: the recently developed Thermographic Signal Reconstruction Polynomial Coefficients method (PC-TSR)¹ [8], Pulse Phase Thermography (PPT) [9], Principal Component Thermography (PCT) [10] and Higher Order Statistics (HOS) [11]. For each method it has been supposed that the diagnosis is based on a trio of images, as thoroughly explained in a previous study [8] and recalled here, in Figure 2b. Those trios of images are the ones on which the present comparison and ranking procedure is tested. All defects are taken into account, except defect #1-4* which is partially overlapping with defect #1/2-2* (overlap situations indeed alter the metrics values, especially for the deepest defects).

¹ This acronym is proposed to distinguish the method from the classical TSR method which uses the first and second logarithmic derivative images [12,13].

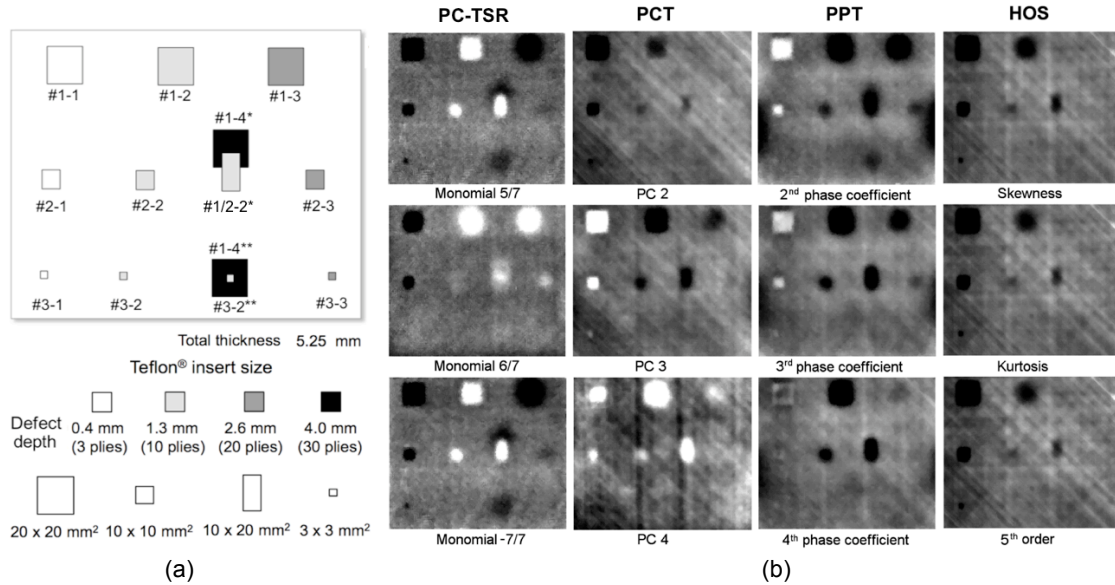


Figure 2. Tested carbon/epoxy coupon with its artificial defects (a) and trios of thermographic images obtained by the four data processing methods (PC-TSR, PCT, PPT, and HOS) (b). Figure taken from [8].

In these conditions, the ranking of the four thermographic NDE data processing methods is achieved over the three metrics, assessed for twelve defect contrasts, obtained on the three characteristic thermal images. Concretely, 432 defect signatures have to be compared: 4 methods \times 3 metrics \times 12 defect contrasts \times 3 images (see Figure 1c). Given this high number of signatures, a data compression is needed to finally reach 4 figures. Such data compression is achieved step by step as shown in Figure 1b.

4.2. Ranking procedure: step #1

The first step consists in assessing the 432 signatures. The choice is made to present final results in the most concise way, in radar charts (Figure 3), so that comparisons can easily be made. Each radar chart presents the results for a given method, a given metrics and the full collection of defects, following a clockwise distribution for decreasing aspect ratios (given in brackets right next to the defect identification numbers). Three curves are plotted, corresponding to the trio of images.

The magnitudes of the CNRs of the three PC-TSR images are quite close to each other, between 15 and 30 dB, up to aspect ratios from 7 to 8. Among the three defects #1-3, 2-2 and 3-1, which have almost the same aspect ratio, only the first two are detected with a CNR of about 20 dB contrarily to the last one detected with a CNR of 10 dB. This validates that the aspect ratio is not the only governing parameter of defect detection. For even lower aspect ratios (1.2 to 2.3), the defects cannot be detected. The sharpness of the even image (monomial 6/7) is almost lower than the one of the odd images. The influence of the aspect ratio is manifest: when it gets lower, sharpness decreases as well. The accuracy on the estimation of the defect lateral dimension follows the same trend: odd images lead to better results. The effect of the decrease of the aspect ratio is even more critical, so that it was impossible to properly evaluate the dimensions of defects #3-1 and 2-3.

Similar results are presented in Figure 3 for the other methods (rows 2, 3 and 4). The CNRs of the PCT images are slightly higher than the PC-TSR ones for defect aspect ratios higher than 7, but decrease for lower ones (the contrast is drowned in the noise). The sharpness has the same magnitude as the PC-TSR one, but quickly decreases with the aspect ratio (the contrast plateau no more exists for defects #3-1, 1-4, 2-3). The accuracy on the lateral defect dimensions is similar to the TSR one, except for defect #1/2-2* for which an very high, unexpected accuracy is found, probably due to the partial overlap with defect #1-4*.

The PPT sharpness and accuracy values found for defect dimensions are comparable to the ones of the previous two methods. On the contrary, the CNR reaches values generally higher than those of the other methods, except for the shallowest defects #1-1 and 2-1. The method appears to be the most efficient one when considering the sole CNR, but shows weaknesses regarding edge sharpness and defect dimension accuracy comparing to the TSR approach.

The results of the HOS method are generally not as good as those of the other three methods, in particular for deep defects ($AR \leq 10$), making this method the less performing one.

4.3. Ranking procedure: step #2

In a second step, the mean value of the metrics for each trio of images is calculated, reducing the total number of considered data to $432/3=144$. To get a more concise comparison, the same type of graphical chart has been applied, which enables a rapid overview of the compared performance of the four methods (see Figure 4).

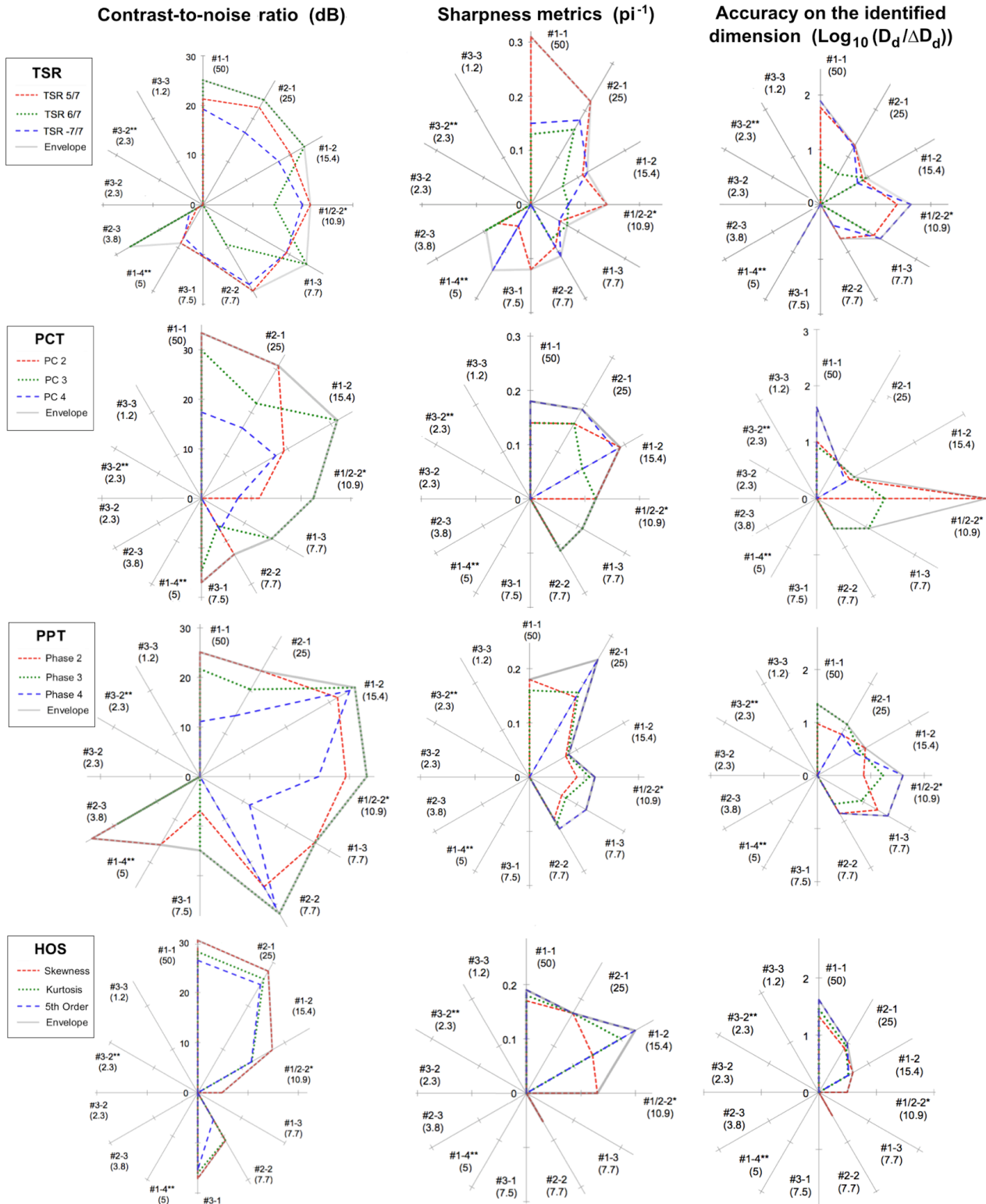


Figure 3. Assessment and comparison of PC-TSR, PCT, PPT and HOS image properties (defect CNR, edge sharpness and accuracy on the defect lateral dimension) given for decreasing defect aspect ratios. For this first step of the procedure the number of data to consider is equal to 432. The envelope curves correspond to the maximum score of each trio.

The TSR method does not appear to be the best one considering the CNR, but obtains the best results for edge sharpness and defect dimension accuracy. This limited performance in CNR is counterbalanced by the fact that, like PPT, it enables to detect deep defects, down to a 4 mm depth.

The previous results have been obtained considering the *mean value* of the three metrics for each trio of grey-scale images. Another possibility could be to consider *the maximum* of each trio (see Figure 5) which seems closer to the actual performance of the methods when the diagnosis is made from the examination of the trio of grey-scale images. Indeed, in this case, even if it is unseen in two images over three, a defect can still be detected as long as it appears in the third image.

In the case of methods based on the observation of a unique image, mean and maximum values merge and a unique set of three radar charts analogous to the ones of Figures 4 and 5 is directly obtained.

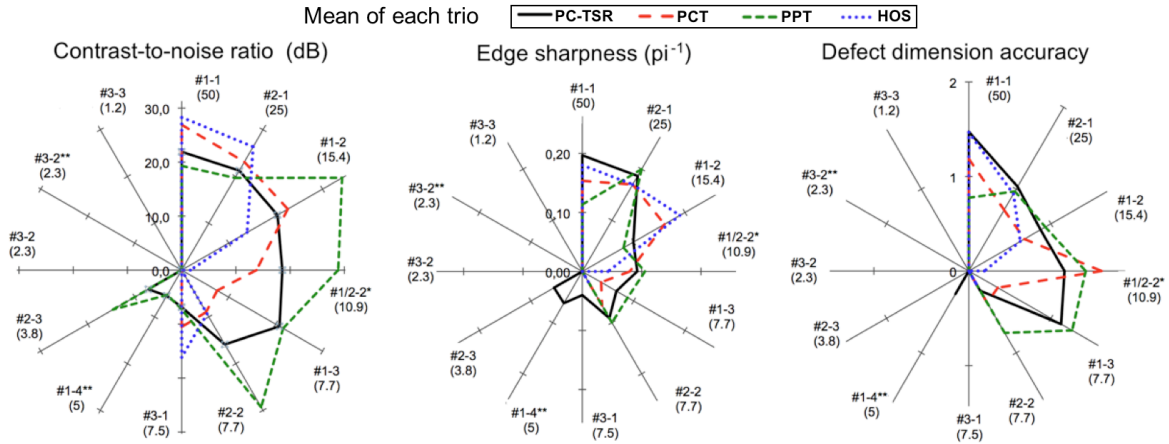


Figure 4. Step #2: evaluation of the mean values of the metrics for each trio of grey-scale images (144 data).

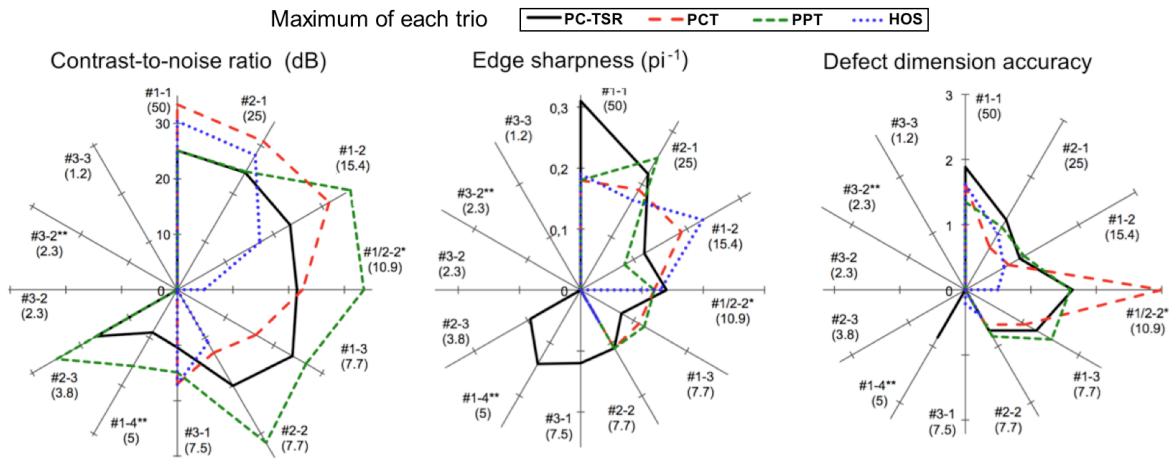


Figure 5. Alternative for step #2: evaluation of the maximum values of the metrics for each trio of grey-scale images.

4.4. Ranking procedure: steps #3 to #6

At the third step, the sums of the marks for the full collection of defects are calculated for each metrics and method, reducing the number of data to 144/12=12. The fourth step does not change this number since it consists in normalizing the marks obtained for each metrics by the highest of them. At the fifth step, the means of the normalized marks are calculated, leading to 12/3=4 data, each one characterizing a method. The sixth and last step consists in ranking the four methods, based on these four marks, and thereby pointing out the best method, or at least the most relevant method for the considered material, defects and experimental conditions.

These marks are presented in Table 1, considering both possible choices for step #2: mean or maximum of the scores obtained for each trio of grey-scale images. In the present case, the rankings by decreasing order of merit are the following:

- i) considering the mean score of each trio, PC-TSR comes first, PPT second, PCT third, and HOS last;
 - ii) considering the maximum score of each trio, PC-TSR and PPT come first, ex-æquo, followed by PCT and HOS.
- The choice of set of data does not alter the conclusion: the PC-TSR approach gets the best mark, alone or ex-æquo with PPT. Then come PCT and finally HOS. Such results are in very good agreement with the qualitative assessment

proposed in [8], which stated that the TSR method was at least as good as the other processing methods, based on the number of detected defects.

A closer look at the intermediate results coming from step #4 points out the strengths and weaknesses of each studied processing method: based on the sole CNR parameter, PPT is the best method; but as soon as the sharpness and the accuracy of the assessment of the lateral dimensions (deduced from the FWHM measurement) are taken into account, PC-TSR takes the lead. To clarify them, these results can be graphically presented (see Figure 6).

Σ of the scores obtained for the full collection of defects considering the mean scores of the trios of grey scale images		TSR	PCT	PPT	HOS
		Step 3	CNR	138	114
	Sharpness	0.89	0.67	0.69	0.60
	FWHM	6.0	4.7	5.9	3.4

Σ of the scores obtained for the full collection of defects, normalized by the same obtained by the best method for the given metrics		TSR	PCT	PPT	HOS
		Step 4	CNR	0.77	0.63
	Sharpness	1.00	0.75	0.77	0.68
	FWHM	1.00	0.79	0.98	0.56

Step 5	Mean	0.92	0.72	0.92	0.59
Step 6	Ranking	1	3	1	4

Σ of the scores obtained for the full collection of defects considering the maximum scores of the trios of grey scale images		TSR	PCT	PPT	HOS
		Step 3	CNR	174	165
	Sharpness	1.33	0.90	0.86	0.78
	FWHM	8.5	7.8	7.4	4.5

Σ of the scores obtained for the full collection of defects, normalized by the same obtained by the best method for the given metrics		TSR	PCT	PPT	HOS
		Step 4	CNR	0.75	0.70
	Sharpness	1.00	0.68	0.65	0.59
	FWHM	1.00	0.91	0.87	0.52

Step 5	Mean	0.92	0.77	0.84	0.52
Step 6	Ranking	1	3	2	4

Table 1. Steps #3 to #6 of the procedure for comparison and ranking of the four NDE methods when considering in the second step the mean (left) and the maximum (right) of the scores obtained for each trio of grey-scale thermal images.

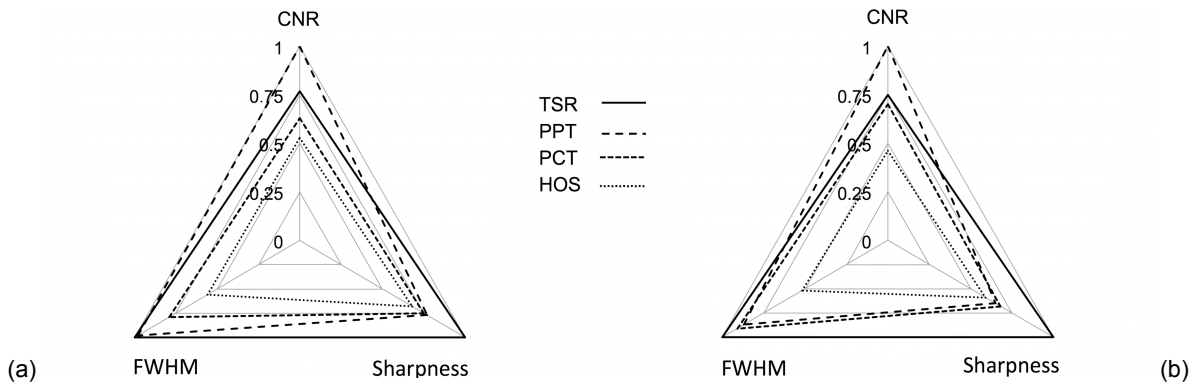


Figure 6. Normalized mean sums of the scores obtained for the full collection of defects (step #4) when considering the mean scores (a) or the maximum scores (b) for each trio of grey-scale thermal images.

5. Discussion

It is clear that the final results of such an assessment and ranking cannot be considered as absolute. The results depend on experimental conditions: infrared camera performances, heating source power and duration, recording frequency and duration, material properties, defect types and depth locations. Any change of any of these parameters may affect the score values and perhaps the final ranking.

Yet, the question of the objectivity of this ranking procedure still remains open and deserves to be discussed. A real effort has been done to make it as unbiased as possible, in particular with the multiplication of the number of metrics, each one of which is being assigned a similar weight, contrarily to what is found in the literature. Most of the previous works are indeed based on the sole SNR metrics. In the present procedure, that would mean to give a zero-weight to the edge sharpness and FWHM metrics, which would lead to rank PPT ahead of the other methods. The wisest path to follow now would be to assign proper, meaningful weights to the different metrics, which would inevitably require taking into account the physiology of the human eyesight and the image processing of the human brain in a diagnosis based on visual

inspection. In such a process what are the role and the importance of the chosen metrics? Here is probably a direction in which the research must be refined.

In fact, the actual remaining subjective part lies in the choice, for each method, of the composition of the trio of images. Here, the selected images were: the three coefficients of higher orders of the degree 7 polynomial for PC-TSR; the second, third and fourth components for PCT; the second, third and fourth phase coefficients for PPT; the skewness, kurtosis and fifth statistic moment for HOS. The selection of a different trio or the decision to pick only one image (in that latter case, the procedure would directly start at step #2) might result in a different ranking. However, such choices are integral parts of the application of the processing techniques, and clearly independent of the ranking procedure itself.

Finally, it is worth noticing that the proposed procedure can also be used to take the most out of a method, by comparing the merits of several images derived from the thermograms. For the PC-TSR method, for instance, the choice of the optimal degree of the fitting logarithmic polynomial could be refined, as well as the number and ranks of the coefficient images.

6. Conclusion

The proposed procedure constitutes an improvement in the way to compare and rank by order of merit several methods of thermographic nondestructive evaluation. The procedure insures an increased objectivity and a high flexibility by taking into account several metrics instead of the sole contrast-to-noise ratio. It is illustrated for the particular case of four processing methods exploiting a trio of thermal images built from the full collection of thermograms recorded during a pulse-heating thermographic experiment on a carbon/epoxy composite with delamination-type artificial defects.

The procedure could be applied to other types of configurations (for instance when the number of images per method is reduced to one), to a unique method applied to different types of thermal solicitations, such as pulse- and step-heating, or to the optimisation of a unique method. Conversely, the procedure could be applied to the comparison of images generated by NDE techniques based on different physical phenomena, such as ultrasonic, eddy-current and thermal images of a same structure with defects. The diversity of the potential applications highlights the universality of the present approach.

REFERENCES

- [1] Usamentiaga R, Venegas P, Guerediaga J, Vega L, López I. A quantitative comparison of stimulation and post-processing thermographic inspection methods applied to aeronautical carbon fibre reinforced polymer. *Quantitative Infrared Thermography Journal*. 2013;10(1):55-73.
- [2] Almond DP, Lau SK. Edge effects and a method of defect sizing for transient thermography. *Applied Physics Letters*. 1993;62(25):3369-71.
- [3] Almond DP, Saintey MB, Lau SK. Edge effects and defect sizing by transient thermography. QIRT 94 Conf., Sorrento, Italy, 1994. Ed. Europ. Therm. & Industrie, Paris, 1995: 247-252. QIRT Opens Archives: www.qirt.org, Paper #1994-037.
- [4] Balageas DL. In search of early time - an original approach in the thermographic identification of thermophysical properties and defects. *Advances in Optical Techniques*. 2013: Article ID 314906. Available from: <http://dx.doi.org/10.1155/2013/314906>.
- [5] Balageas DL, Roche J-M, Leroy F-H. A procedure for the comparative assessment of thermal NDE data processing techniques: application to TSR, PPT, PCT and HOS methods. Soumis à *Materials Evaluation*, 2016.
- [6] Balageas DL, Déom AA, Boscher DM. Characterization and nondestructive testing of carbon-epoxy composites by a pulsed photothermal method. *Materials Evaluation*. 1987;45(4):461-465.
- [7] Krapez J-C, Boscher D, Delpech Ph, Déom A, Gardette G, Balageas D. Time-resolved pulsed stimulated infrared thermography applied to carbon-epoxy nondestructive evaluation. QIRT 92 Conf., Châtenay-Malabry, France, 1992, pp. 195-200. QIRT Open Archives: <http://qirt.org>, Paper QIRT 1992-029.
- [8] Roche J-M, Leroy F-H, Balageas DL. Images of thermographic signal reconstruction coefficients: a simple way for rapid and efficient detection of discontinuities. *Materials Evaluation*. 2014;72(1): 73-82.
- [9] Maldague X, Marinetti S. Pulse phase infrared thermography. *Journal of Applied Physics*. 1996;79:2694-2698.
- [10] Rajic N. Principal component thermography for flaw contrast enhancement and flaw depth characterisation in composite structures. *Composite Structures*. 2002;58(4):521-528.
- [11] Madruga FJ, Ibarra-Castanedo C, Conde OM, López-Higuera JM, Maldague X. Infrared thermography processing based on higher-order statistics. *NDT&E International*. 2010;43:661-666.
- [12] Oswald-Tranta B, Shepard S.M. Comparison of pulse phase and thermographic signal reconstruction processing. Proc. SPIE 8705, Thermosense: Thermal Infrared Applications XXXV, Vol. 87050S. GR Stockton, FP Colbert, Eds. May 22, 2013; doi:10.1117/12.2017899
- [13] Shepard SM, Hou YL, Lhota JR, Golden JM. Automated processing of thermographic derivatives for quality assurance. *Optical Engineering*. 2007;46(5):051008_1-6.