

Defect characterization in thermographic non destructive testing with pattern recognition

Final assignment for advanced pattern recognition course TUDELFT

*Hernán D. Benítez ^a

^a*Pontificia Universidad Javeriana-Cali, Colombia*

Abstract

This report presents the use of pattern classifiers for defect depth estimation in composite materials by using Infrared Nondestructive Testing (IRNDT). First, the training and test datasets are explained. Additionally, the use of several features such as C_{max} and t_{max} , thermal contrast curves and dissimilarity representation of thermal contrast are shown. Finally, these features are used to train and test ldc, qdc and 1-NN classifiers and their performances in terms of classification errors are compared.

1. Training and test sets

The training and test sets are thermal contrast (TC) curves extracted from composite materials. In IRNDT thermal contrast is defined as:

$$\Delta T = T_{def} - T_s \quad (1)$$

in which T_{def} and T_s corresponds to time evolution of temperature in defective and sound areas of the inspected material. The thermal contrast type used in this work is reference-free which means that a sound area does not have to be defined a priori by an operator but it is defined by using a thermal model. More details of this thermal contrast can be found in [1].

The training dataset was extracted from the flat CFRP (Carbon Fiber Reinforced Plastic) sample CFRP006 shown in Figure 1 and the test dataset was obtained from sample CFRP007 with curved shape as shown in Figure 2. The test set has 226 x 236 observations which correspond to every pixel in the infrared image sequence. Figure 3 shows the TC curves for the five largest defects at different depths (15 mm in lateral size) in CFRP006.

The thickness of all the samples is 2 mm. These experiments were performed using two powerful flash lamps (Balcar FX 60, 6.4 kJ), with a 5 ms pulse as the excitation source. All the thermogram sequences were recorded using a FPA infrared camera (Santa Barbara Focalplane SBF125, 3 to 5 μ m), with a 320 x 256 pixel array. To generate training data twelve (12) experiments were carried out on a sample with 25 defects so that every defect had twelve observations producing 300 curves from defective pixels and 300 from sound pixels to complete 600 observations for the training set.

2. Features

The features studied in this assignment are:

- C_{max} and t_{max} which correspond to the maximum contrast value and the time at which this maximum contrast value is reached, respectively.
- TC curves
- Dissimilarities representation of TC curves.

The classifiers purpose is to estimate the depth at which a defect is inside the inspected sample. As can be observed in Figures 2 and 3 these depths are 0.2,

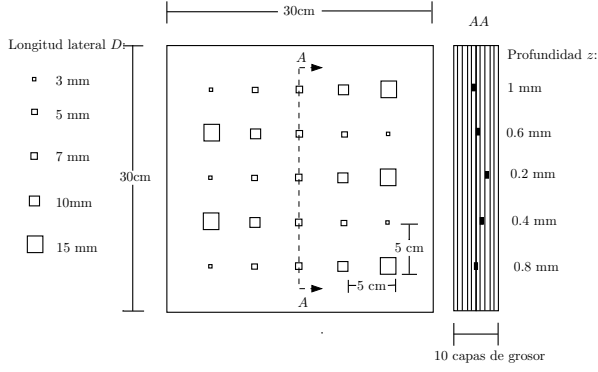


Fig. 1. CFRP sample with *Teflon*TM insertions for training (CFRP006)

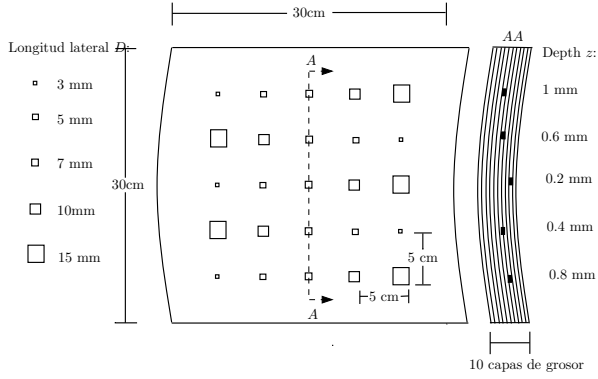


Fig. 2. CFRP sample with *Teflon*TM insertions for training (CFRP006)

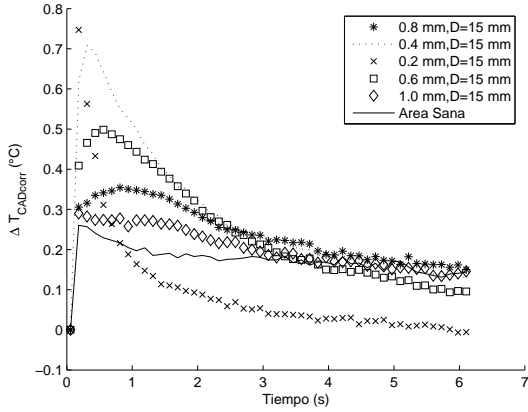


Fig. 3. Thermal contrast curves for the five largest defects at different depths (15 mm in lateral size) in CFRP006

0.4, 0.6, 0.8 and 1.0 mm. For this reason 6 classes are defined: 5 classes for defects (one per depth) and 0 mm that corresponds to the non-defective class.

For every case the TC curves have 32 points. Figure 4 shows the training data set for C_{max} (Fea-

ture 1) and t_{max} (Feature 2) features. It can be observed that the class overlap is stronger between the curves that represents deepest defects (1.0 mm and 0.8 mm) and the curves that represent non-defective areas since as long as the defect depth increases the TC curve tends to be similar to the TC curve of a non-defective area.

For the dissimilarity representation the routine *distm* of PRTOOLS [2] was used to get the square Euclidean distance matrix from training set matrix obtaining a dissimilarity matrix D_{train} of 600x600. The test set with 226x236 observations was subsampled with function *gendat* to get 1000 observations. Therefore a 1000x32 matrix is obtained which is used to calculate the square distance matrix between the training set and the test set obtaining D_{test} with size 1000x600. After getting these matrices a forward feature selection is applied with routine *featself* to do prototype selection [3] and obtain a subset of 5 squared Euclidean distances. Hence the D_{train} and D_{test} sizes are 600x5 and 1000x5, respectively.

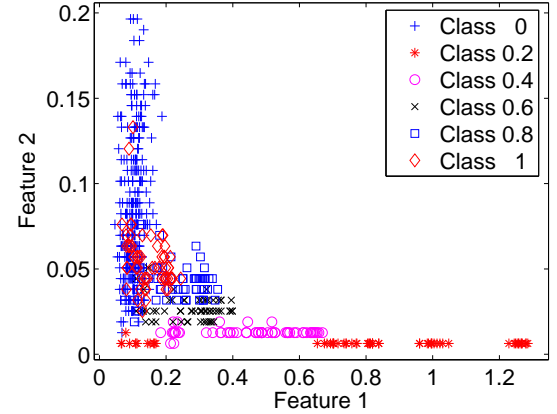


Fig. 4. Training data set for C_{max} (Feature 1) and t_{max} (Feature 2) features

3. Classifiers evaluation

In this section linear discriminant (ldc), quadratic discriminant (qdc) and 1 nearest-neighbor classifiers are trained with 600 observations from CFRP006 and tested with 226X236 observations extracted from sample CFRP007. These classifiers are trained and tested with three different features sets as shown in section 2. Figures 5, 6 and 7 show the decision boundaries computed by the classifiers ldc, qdc and 1-NN for feature space C_{max} and t_{max} .

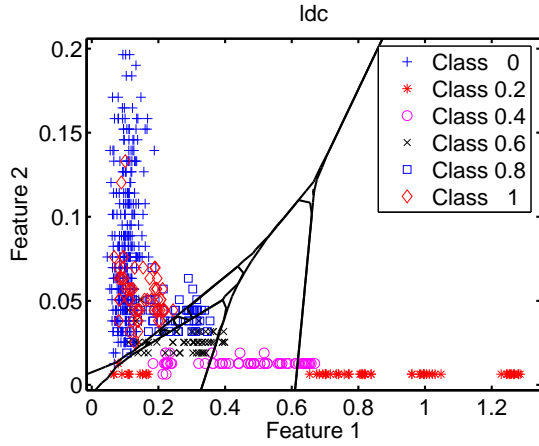


Fig. 5. Linear classifier decision boundary for C_{max} and t_{max} features

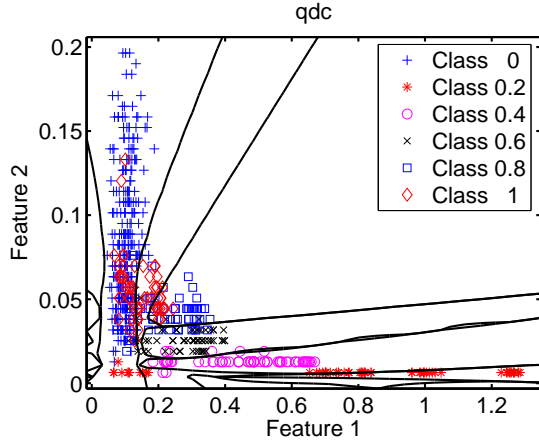


Fig. 6. Quadratic classifier decision boundary for C_{max} and t_{max} features

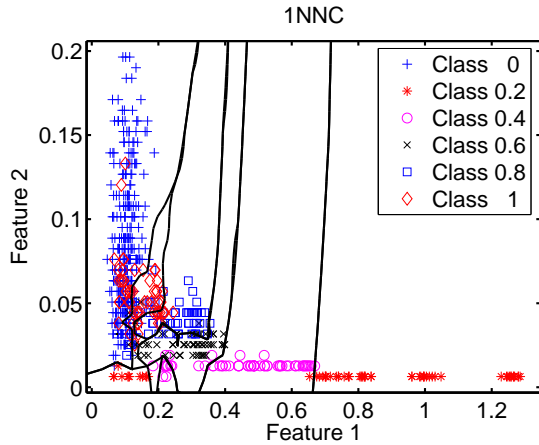


Fig. 7. 1 Nearest-neighbor classifier decision boundary for C_{max} and t_{max} features

Table 3 shows the classification errors for the different sets of features obtained with classifiers ldc, qdc, 1-NN and MLP. The number of hidden neurons for the neural classifier was found by using a cross-validation (routine *crossval* in PRTOOLS) procedure on the training set by choosing $k=10$ and a range of [5:40] for the number of units in the hidden layer explored. For this case the number of units found for the hidden layer is 25. The errors obtained with features C_{max} , t_{max} and dissimilarities among thermal contrast curves are discouraging. However, for the TC curves feature set a lower error (20,8 %) is obtained with the ldc classifier. For this case the qdc and 1-NN classifiers provide lower error rates than in other cases, although their error rates are still very high (40,17 % and 90 % !). For the MLP case only the thermal curves feature set was used to train and test obtaining a classification error of 23.8 %.

Table 1

Classification errors for different feature sets and classifiers

Features	Classif Errors(%)			
	ldc	qdc	1-NN	MLP
C_{max}, t_{max}	96.9	98.1	98.5	-
Dissimilarities	86.1	98.8	98.5	-
TC curves	20.8	40.2	98.5	23.8

4. Conclusions

From the classification errors comparison it can be concluded that the TC curves is the best representation in order to reduce error rate. Since for the C_{max} and t_{max} features the observations are sensitive to changes because of noise. Additionally, the dissimilarity representation expressed by square Euclidean distances seems not to be a good feature set in this case and it is necessary to test with other dissimilarity measures such as area differences or correlation coefficients to explore a possible reduction of error rates with these dissimilarity measures. On the other hand, the lowest error in the TC curves feature space was obtained with the ldc classifier followed by the neural classifier. This means that the depth estimation problem could be partially solved by a simple classifier that does not require long and complex training process.

Finally, the disadvantage of using learning machines in TNDT is that this approach requires at

least one sample with a known set of defects to obtain the training dataset and that once the system is trained this can only be used for the same material with the same experimental platform.

References

- [1] Benítez HD, Ibarra-Castanedo C, Bendada H, Maldague X, Loaiza H, Caicedo E. Definition of a new thermal contrast and pulse correction for defect quantification in pulsed thermography. *Infrared Physics & Technology* 2008;51(3):160-167.
- [2] PRTools4, A Matlab Toolbox for Pattern Recognition, Delft University of Technology, Duin RPW, Juszczak P, Paclik P, Pekalska E, de Ridder D, Tax DMJ.
- [3] Pekalska E, Duin R, Paclik P, Prototype selection for dissimilarity-based classifiers, *Pattern recognition* 2006; 39: 189-208.
- [4] Heijden F, Duin R, De Ridder D, Tax DMJ, *Classification, parameter estimation and state estimation - an engineering approach using Matlab*, John Wiley & Sons, 2004.