Pattern Recognition in the Automatic Inspection of Aluminium Castings

Domingo Mery[1], Romeu R. da Silva[2], Luiz P. Calôba[3], João M. A. Rebello[2]

1 Departamento de Ingeniería Informática
Universidad de Santiago de Chile
Av. Ecuador 3659, Santiago de Chile
EMail: dmery@ieee.org
http://www.diinf.usach.cl/~dmery

2 Engenheria Metalurgica e de Materiais
Escola de Engenharia e COPPE
Universidad Federal do Rio de Janeiro
PO Box: 68505, CEP 21945-970,
Rio de Janeiro
EMail: {romeu,jmarcos}@metalmat.ufrj.br
http://www.metalmat.ufrj.br

3 Engenheria Eletronica
Escola de Engenharia e COPPE
Universidad Federal do Rio de Janeiro
PO Box: 68504, CEP 21945-970,
Rio de Janeiro
EMail: caloba@lps.ufrj.br
http://www.lps.ufrj.br

ABSTRACT

In this paper, we report the results obtained recently by analysing around 400 features measured from 23.000 regions segmented in 50 radioscopic images of cast aluminium wheels with defects. The extracted features are divided into two groups: geometric features (area, perimeter, height, width, roundness, Fourier descriptors, invariant moments, and other shape factors) and grey value features (mean grey value, mean gradient in the boundary, mean second derivate in the region, radiographic contrasts, invariant moments with grey value information, local variance, textural features based on the co-occurrence matrix, coefficients of the discrete Fourier transform, the Karhunen Lòeve transform and the discrete cosine transform). We propose a new contrast feature obtained from grey level profiles along straight lines crossing the segmented potential defects. Analysing the area A_z under the receiver operation characteristic (ROC) curve, this feature presents the best individual detection performance yielding an area A_z=0.9944. In order to make a compact pattern representation and a simple decision strategy, the number of features are reduced using feature selection approaches. The relevance of the selected features is evaluated by calculation of linear correlation coefficient. The results are shown in tables of the linear correlation coefficients between features, and among features and class of defect. In addition, statistical classifiers and classifiers based on neural networks are implemented in order to establish decision boundaries in the space of the selected features which separate patterns (our segmented regions) belonging to two different classes (regular structure or defects).

1. Introduction

Automatic detection of failures in non-destructive testing is normally carried out by a well known pattern recognition technique, the steps of which are illustrated in Figure (1): image formation, pre-processing, segmentation, extraction of features, and classification (Mery et al, 2003a). In the *image formation* the attempt is made obtain a digital X-ray image of the object under test using flat panel detectors or image intesnifier with CCD camera. The *pre-processing* is devoted to improving the quality of the image in order to better recognize flaws (e.g., digital filters, integration, contrast enhancement, etc.). The *segmentation* process divides the digital image into disjoint regions with the purpose of separating the parts of interest from the rest of the scene. The present investigation uses the segmentation process oriented towards the detection of edges by employing the LoG filter (Mery & Filbert, 2002b). As can be seen in Fig. (2), this technique searches for changes in the grey values of the image (edges) thus identifying zones delimited by edges that indicate flaws. In the inspection of cast pieces, segmentation detects regions that are denominated as 'hypothetical defects', which may be flaws or structural features of the object. Subsequently, the *feature extraction* is centred principally around the measurement of geometric properties and on the intensity characteristics of regions. Finally, *classification* orders segmented regions in specific regions according to extracted features, assigning each region to one of a number of pre-established groups, which represent all possible types of regions expected in the image. Typically the classes that exist in detection of flaws in cast pieces are two: 'defects' or 'regular structures'[1].

[1] A sub-classification of flaws is possible to determine the type of flaw.

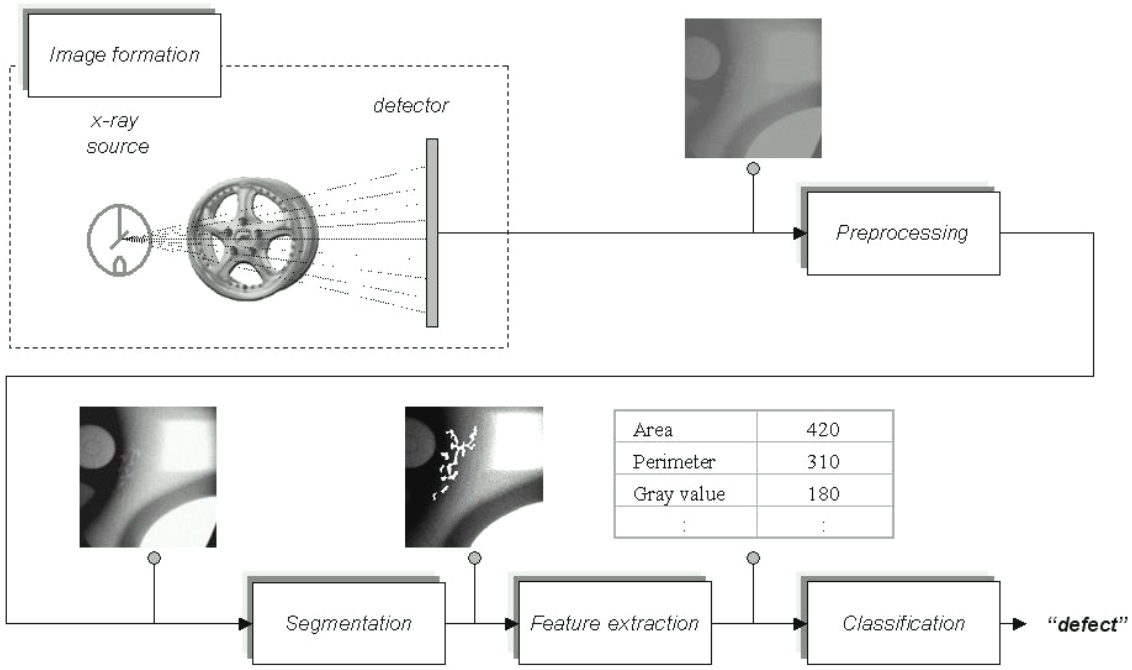


Figure 1. Pattern recognition in the automatic detection of flaws.

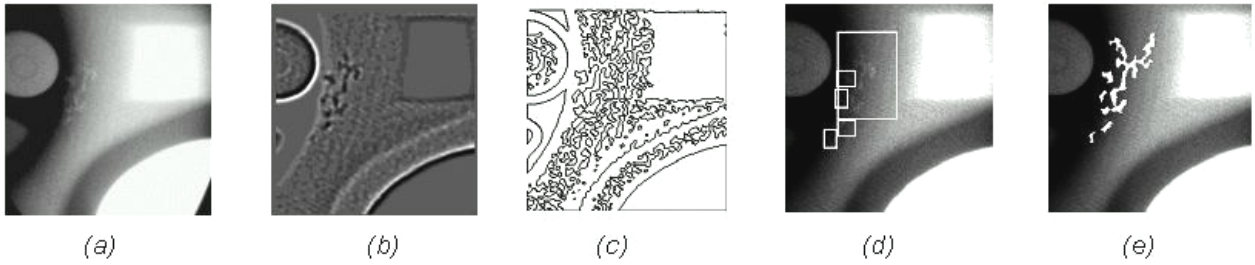


Figure 2: Detection of flaws using edge detection: (a) original image, (b) second derivative, (c) edges using a LoG filter (d) and (e) detection of flaws.

This paper presents an exhaustive analysis of over 400 features extracted from over 23,000 regions in 50 noisy radioscopic images. The paper is organised as follows. Section 2 briefly describes the features used. Section 3 carries out a feature selection in order to reduce computational cost of classification. Section 4 presents the linear correlation method for the selected features. Section 5 shows the results obtained by using statistical classifiers and classifiers based on neuronal networks, using original features and those obtained by an analysis of principal components. Finally, conclusions and suggestions for future research are presented[2].

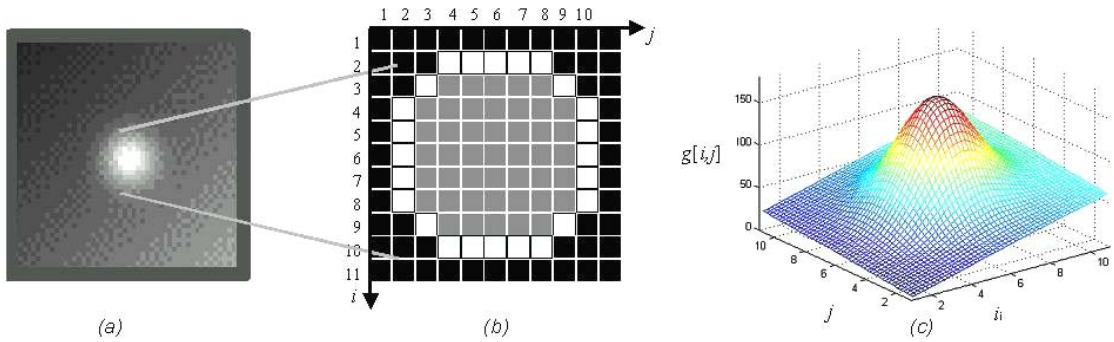


Figure 3: Example of a region. (a) X-Ray image, (b) segmented region, (c) 3d representation of the intensity (grey value) of the region and its surroundings.

[2] A extended version of this paper was recently published in (Mery et al, 2003).

Table 1. Extracted features.

Type	Variable	#	Description	Ref.
geo	I	1	number of the image	
geo	$(\bar{i}, \bar{j})$	2-3	centre of gravity	Castleman, 1996
geo	h, w	4-5	height and width	Castleman, 1996
geo	A, L	6-7	area and perimeter	Castleman, 1996
geo	R	8	roundness	Castleman, 1996
geo	$\phi_1 \cdots \phi_7$	9-15	Hu's moments	Sonka et al, 1998
geo	$\lvert DF_0\rvert \cdots \lvert DF_7\rvert$	16-23	Fourier descriptors	Zahn & Roskies, 1971
geo	$FM_1 \cdots FM_4$	200-203	Flusser and Suk invariant moments	Sonka et al, 1998
geo	$FZ_1 \cdots FZ_3$	204-206	Gupta and Srinath invariant moments	Sonka et al, 1998
geo	(a_e, b_e)	207-208	major and minor axis of fitted ellipse	Fitzibbon et al, 1999
geo	a_e / b_e	209	ratio major to minor axis of fitted ellipse	Fitzibbon et al, 1999
geo	α	210	orientation of the fitted ellipse	Fitzibbon et al, 1999
geo	(i_0, j_0)	211-212	centre of the fitted ellipse	Fitzibbon et al, 1999
geo	G_d	213	Danielsson form factor	Danielsson, 1978
int	G	24	mean grey value	Castleman, 1996
int	C	25	mean gradient in the boundary	Mery & Filbert, 2002b
int	D	26	mean second derivative	Mery & Filbert, 2002b
int	$K_1 \cdots K_3$	27-29	radiographic contrasts	Kamm, 1998
int	K_σ	30	deviation contrast	Mery & Filbert, 2002b
int	K	31	contrast based on CLP[3] at 0º and 90º	Mery & Filbert, 2002b
int	Δ_Q	32	difference between maximum and minimum of BCLP[3]	Mery, 2003
int	Δ'_Q	33	$\ln(\Delta_Q + 1)$	Mery, 2003
int	σ_Q	34	standard deviation of BCLP[3]	Mery, 2003
int	Δ''_Q	35	Δ_Q normalised with average of the extreme of BCLP[3]	Mery, 2003
int	$\overline{Q}$	36	mean of BCLP[3]	Mery, 2003
int	$F_1 \cdots F_{15}$	37-51	first components of DFT of BCLP[3]	Mery, 2003
int	$\phi'_1 \cdots \phi'_7$	52-58	Hu moments with grey value information	Sonka et al, 1998
int	σ_g^2	59	local variance	Mery & Filbert, 2002b
int	Tx_1	60-87	mean and range of 14 texture features[4] with d=1	Haralick et al, 1973
int	Tx_2	88-115	mean and range of 14 texture features[4] with d=2	Haralick et al, 1973
int	Tx_3	116-143	mean and range of 14 texture features[4] with d=3	Haralick et al, 1973
int	Tx_4	144-171	mean and range of 14 texture features[4] with d=4	Haralick et al, 1973
int	Tx_5	172-199	mean and range of 14 texture features[4] with d=5	Haralick et al, 1973
int	Y_{KL}	214-277	64 first components of the KL transform[5]	Castleman, 1996
int	Y_{DFT}	278-341	64 first components of the DFT transform[5]	Castleman, 1996
int	Y_{DCT}	342-405	64 first components of the DCT transform[5]	Castleman, 1996

2. Feature extraction

As mentioned above, segmentation is performed using an edge detection technique. All regions enclosed by edges in the binary image are considered 'hypothetical defects' (see example in Fig. (3)). During the feature extraction process the properties of each of the segmented regions are measured. The idea is to use the measured features to decide whether the hypothetical defect corresponds to a flaw or a regular structure.

[3] CLP: *Crossing line profile*, grey value function along a line that crosses the region at its centre of gravity. The term BCLP refers to the *best t* CLP, in other words the CLP that presents the best homogeneity at its extremes (Mery et al, 2003b).

[4] The following features are extracted based on a co-ocurrence matrix: second angular moment, contrast, correlation, cum of squares, inverse difference moment, mean sum, variance of the sum, entropy of the sum, entropy, variance of the difference, entropy of the difference, 2 measures of correlation information, and maximum correlation coefficient, for a distance of d pixels.

[5] The transformation takes a re-sized window of 32×32 pixels which includes the region and its surroundings.

The features extracted in this investigation are described below, and have been grouped in two categories: *geometric* (geo) and *intensity* (int) features. Geometric features provide information relative to the size and form of the segmented hypothetical flaws. Intensity features, on the other hand, provide information about the grey value of the segmented regions. Table (1) presents the set of features used in this investigation. The details of how these are calculated can be found in the references.

The total number of features extracted is 405 divided into 37 geometric features, and 368 intensity features.

3. Feature Selection

In order to reduce the computational time required for classification it is necessary to select features; this way the classifier only works with non-correlated features that provide flaw detection information. There are a variety of methods for evaluating the performance of the extracted features. The present Section includes the ROC analysis and the Fisher discriminator while the following Section explains the linear correlation analysis.

The ROC (*receiver operation characteristic*) analysis is commonly used to measure the performance of a two-class classification. In our case, each feature is analysed independently using a threshold classifier. This way, a hypothetical flaw is classified as a 'regular structure' (or 'defect') if the value of the feature is below (or above) a threshold value. The ROC curve represents a 'sensitivity' (S_n) versus '1-specificity' (1-S_p), defined as:

$$S_n = \frac{TP}{TP+FN}, \qquad 1-S_p = \frac{FP}{TN+FP} \tag{1}$$

in which *TP* is the number of true positives (correctly classified defects), *TN* is the number of true negatives (correctly classified regular structures), *FP* is the number of false positives (false alarms, or regular structures classified as defects) and *FN* false negatives, (flaws classified as regular structures)[6]. A graphic representation is presented in Fig. (4). Ideally, S_n = 1 and 1-S_p = 0, this means that all defects were found without any false alarms. The ROC curve makes it possible to evaluate the performance of the detection process at different points of operation (as defined for example by means of classification thresholds). The area under the curve (A_z) is normally used as a measure of this performance as it indicates how flaw detection can be carried out: a value of A_z = 1 indicates an ideal detection, while a value of A_z = 0.5 corresponds to random classification (Egan, 1975).

Another method for evaluating classifier performance is the Fisher discriminator, which evaluates the following function:

$$J = \mathrm{spur}\left(\mathbf{C}_w^{-1}\mathbf{C}_b\right) \tag{2}$$

where $\mathbf{C}_b$ and $\mathbf{C}_w$ represent the selected features' *interclass* (between) and *intraclass* (within) covariances respectively. A large value for *J* indicates a good separation of classes as it ensures a small intraclass variation and a large interclass variation in the feature space (Fukunaga, 1990).

The results presented in this article are derived from the analysis of 50 radioscopic images of cast aluminium pieces. The images have a high noise component due to the fact that they were taken without using averaging techniques. In these images a total of 22,936 regions were segmented. Visual inspection determined that 60 of these were real flaws corresponding to blow holes in the piece. These flaws were located in areas of the piece in which detection is very difficult (see examples in (Mery & Filbert, 2002b)). The 405 features mentioned in Table (1) were extracted from each one of the almost 23,000 regions.

[6] In the literature the following terms are also known *False Acceptance Rate* (*FAR*) and *False Rejected Rate* (*FRR)* defined as 1-S_p and 1-S_n respectively (Jain et al, 2000; Mery & Filbert, 2002a).

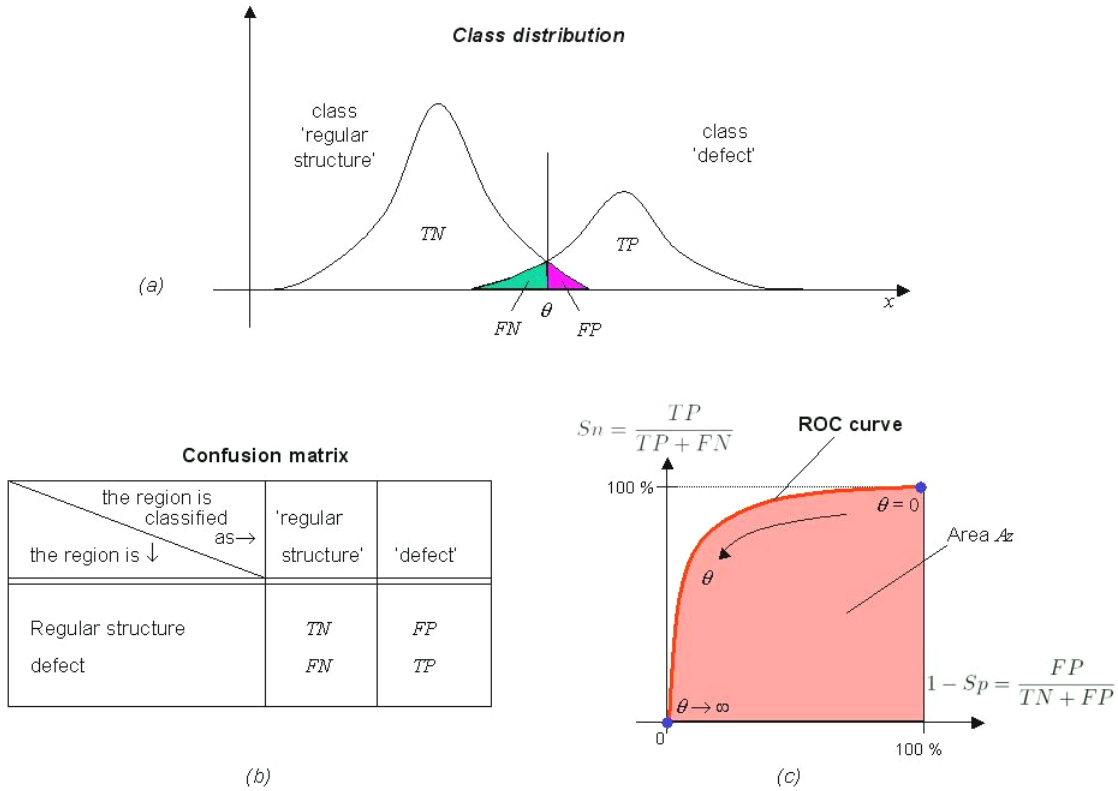


Figure 4: Defect Classification. a) Distribution of classes using a single feature *x*. b) Confusion matrix. c) ROC curve varying the threshold θ.

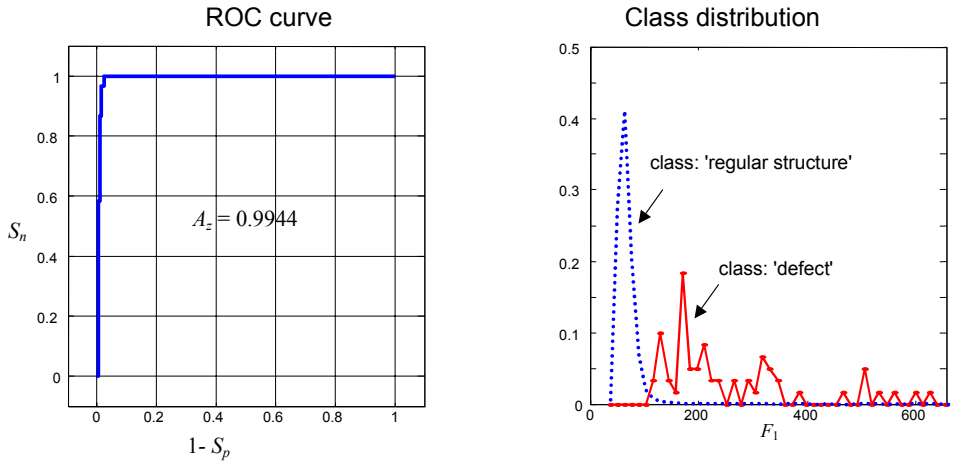
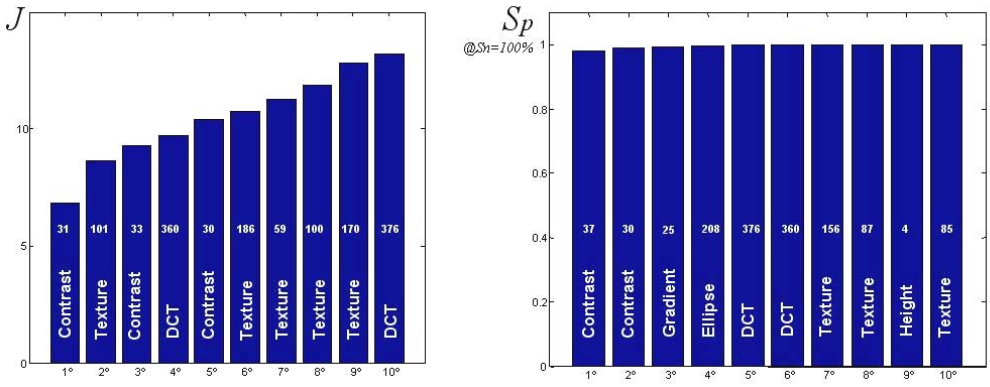


Figure 5: ROC curve and class distribution for the best contrast feature (# 37).

Table 2. Values of the area under the ROC curve A_z and Fischer discriminator J for the 28 pre-selected features.

#	var	A_z	J	#	var	A_z	J	#	var	A_z	J	#	var	A_z	J
4	h	0.94	1.5	59	σ_g^2	0.95	1.8	139	Tx_3	0.87	1.5	186	Tx_5	0.96	3.9
8	R	0.91	2.0	85	Tx_1	0.93	2.9	156	Tx_4	0.92	1.9	190	Tx_5	0.90	1.5
25	C	0.93	2.0	87	Tx_1	0.83	1.9	162	Tx_4	0.89	1.4	195	Tx_5	0.94	3.0
30	K_σ	0.99	3.0	100	Tx_2	0.97	4.6	167	Tx_4	0.91	2.2	208	b_e	0.96	1.7
31	K	0.99	6.8	101	Tx_2	0.83	1.6	170	Tx_4	0.90	1.9	285	Y_{DFT}	0.96	1.4
33	Δ'_Q	0.99	4.6	113	Tx_2	0.92	2.5	179	Tx_5	0.97	3.6	360	Y_{DCT}	0.95	2.5
37	F_1	0.99	1.9	128	Tx_3	0.96	3.2	182	Tx_5	0.94	2.8	376	Y_{DCT}	0.94	1.7



(a) (b)

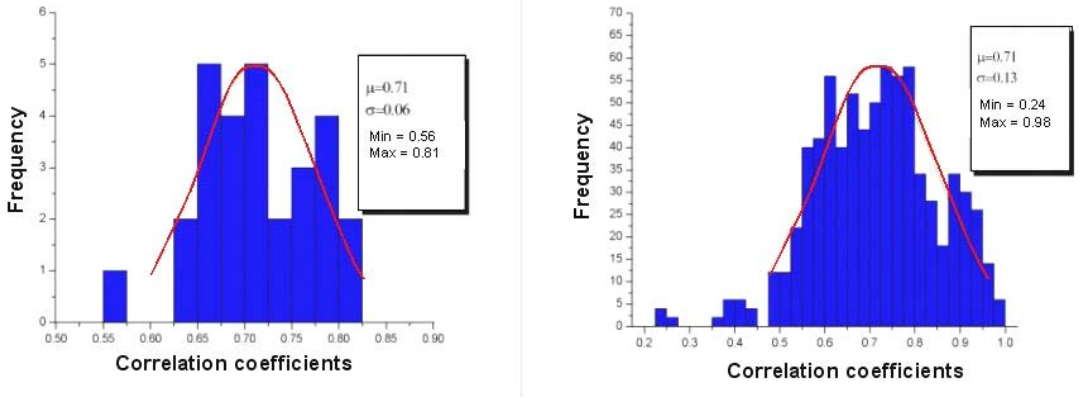
Figure 6: Selection of the first 10 features 10 by means of the SFS method using a) Fisher discriminator (*J*) and b) sensitivity specificity = 100%.

A pre-selection of features was obtained by eliminating those that presented an area under the ROC curve $A_z < 0.8$ and a Fischer discriminator $J < 0.2\ J_{max}$, where J_{max} corresponds to the maximum Fisher discriminator obtained upon evaluation of each of the features. Additionally, if two features present a correlation coefficient with an absolute value greater than or equal to 0.95 then the feature with a lower A_z is eliminated. In this manner of the 405 features, 376 were eliminated. Table (2) shows the 28 pre-selected features. As can be seen, the best results were obtained with features belonging to the contrast family. Texture also performed well. Additionally, some of the features were geometric, although better results were obtained with intensity features. Figure (5) shows the best ROC curve obtained for F_1, the contrast feature based on a Fourier analysis of grey values along the straight lines that cross the regions (Mery, 2003). In this case the area under the curve is $A_z = 0.9944$. The good class separation offered by this feature can also be appreciated.

Following the pre-selection of features a selection process was carried out based on the *Sequential Forward Selection* (SFS) method (Jain, et al, 2000). This method requires an objective function *f* that evaluates the performance of the classification using *m* features. This function could be, for example, the Fisher discriminator defined in Eq. (2). The method begins with one feature (m =1), and a search is performed for the feature that maximises the function *f*. Subsequently a second search is carried out for that feature that maximises the function *f* with two features (m =2). This method ensures that neither features that are correlated with the already selected feature nor those that do not maximise *f* are considered. This process is repeated until the best *n* features are obtained. This approach works best with normalised features, i.e. those that have been linearly transformed in such a way as to obtain a mean value equal to zero, and a variance equal to one.

It is also possible to use another objective function to evaluate separation between classes. An alternative function would be to evaluate the specificity for a threshold classification that obtains a sensitivity of 100%, known in the literature as S_p @ S_n=100%. The first 10 features selected with the SFS method are shown in Fig. (6) for both objective functions. As can be seen the objective function increases in the measure that new features are added with the SFS method. For example, and using only one feature, feature # 31 (K) maximises function *J*, thus obtaining a value of J=6.8 (see first column in Figure (6a)). Likewise, when using this feature with feature # 101 (Tx_2) the objective function *J* increases to 8.5 (see second column in Fig. (6a)). It should be mentioned that this texture is the feature that most increases *J* in combination with feature K. This process is repeated until the 10 *best* features are found.

It can be concluded that there are a number of selected features that show a significant performance. An appropriate mixture of contrast and texture features obtains the best separation of classes. When designing a classifier it is also necessary to consider the computational cost of calculating the features. It is recommended that the selected features should be able to be computed quickly in order that the detection of flaws can be carried out in the time required for the automated inspection process.



(a) (b)

Figure 7: a) Histogram of the correlation coefficients between features and the 'defect' and 'regular structure' classes b) histogram of correlation coefficients between features.

4. Linear correlation analysis and relevance criterion

The analysis of the two linear correlation coefficients is a method for evaluating the correlation that exists between selected features, and also the correlation between features and the ideal classification supervision variable y_k, in which y_k is 1 (or 0) if the sample k belongs to the class 'defect' (or 'regular structure'). A detailed explanation for calculating linear correlation coefficients is given in (Silva et al, 2002a).

In the present investigation, after the pre-selection of the 28 most relevant features according to the Fisher and ROC criteria, (summarised in Table (2)), a linear correlation analysis was performed on a set of the 60 existing samples from the 'defective' class, and 180 randomly selected samples taken from 22,786 samples belonging to the 'regular structure' class. The limiting value for a variable to have a 95% probability of being correlated with another is given by $2/\sqrt{N}$ where N is the number of samples of the variable (Silva et al, 2002a). Thus the limiting values are $2/\sqrt{60}$ = 0.26 for correlation in the 'defect' class, and $2/\sqrt{180}$ = 0.15 for correlation in the 'regular structure' class, while for correlation between classes the limiting value was $2/\sqrt{60+180}$ = 0.13. Figure (7a) shows the resulting histogram for the correlation coefficients found between the features and the variable y_k. Likewise, Fig. (7b) shows the histogram of the correlation coefficients obtained between all features. It is evident from Fig. (7a) that the 28 features are correlated with the ideal classification supervision variable y_k, because the limits of 0.26 or 0.15 are surpassed by the minimum value of 0.56 (obtained for feature # 25). The maximum correlation coefficient, 0.81, was obtained for feature # 100.

Figure (7b) shows that there exists a strong correlation between the majority of the features because the limit of 0.13 is surpassed by all the correlation coefficients found. The graph also shows that some features are highly correlated, and therefore their use as classifiers would be redundant.

A recommended method of feature selection is to use those that show a good cost-benefit ratio, in other words those whose extraction does not imply a high computational cost while at the same time show a good performance in classification. The present investigation selects those features that satisfy the previously mentioned criteria, and are easy to extract. The objective is to reduce the dimensions of the input data of the classifiers, and to reduce the computation time necessary for the extraction of features, thus reducing total inspection process time.

5. Classification

This Section presents the results obtained when using statistical classifiers and classifiers based on neuronal networks. The principal problem manifested in the design of these classifiers is the paucity of information regarding the 'defect' class. This is due to the fact that in data obtained from segmentation there are approximately 381 samples in the 'regular structure' class for every sample in the 'defect' class. In this type of classification, performance cannot be optimised by minimising classification error because a classifier that indicates that all samples belong to the 'defect' class will

have, in our case, a success rate of 381/382 = 99.74%. Nevertheless, this classifier will have detected none of the defects. For this reason classification performance must be measured according to the sensitivity and specificity values explained in Section 3. During training the Neyman-Pearson criterion is recommended (Kay, 1998), in which sensitivity is maximised for a given specificity.

In order to train classifiers, the data sets shown in Table (3) were used. These sets were taken from the total of samples in the 'defect' class. Set I corresponds to the entire sample. In sets II, III, IV and V the quantity of samples from the 'regular structure' class were reduced. In these cases, the samples from the 'regular structure' class were selected randomly and there are no common elements among the selected sets. In sets V and VI the samples from the 'defect' class were repeated until they were equal in number to the samples from the 'regular structure' class.

Table 3: Data sets used in the experiments.

	Set I	Set II	Set III	Set IV	Set V	Set VI
Defects	60	60	60	60	2,000[7]	22,876[7]
Regular structures	22,876	180	500	2,000	2,000	22,876

5.1 Statistical classifiers

In the statistical pattern recognition, classification is achieved using the concept of similarity: patterns that are *similar* are assigned to the same class (Jain et al, 2000). Although this methodology is quite simple it is necessary to establish a good metric of similarity. Using a representative sample, in which the expected classification is known, it is possible to perform a *supervised classification* by finding a discriminating function that can supply information about the degree of similarity between the n features to be evaluated (contained in the vector of features $\mathbf{x} = [x_1,..., x_n]^T$) and the features that represent each class. In this investigation experiments are carried out with the following statistical classifiers threshold-, linear-, nearest neighbour- and Mahalanobis-classifier (Mery & Filbert, 2002a; Mery et al, 2003b).

The experiments first used a threshold based statistical classifier. Using a single feature (# 37), the rule is to classify a region as 'defect' if the feature is above a threshold value. In this way a value for S_n = 95.0% is obtained, with a specificity of 1-S_p = 1.4% using Set I of the data. This means that of the 60 existing real flaws, TP = 57 have been identified correctly, nonetheless FP (false alarms) = 325 are also obtained. A threshold classifier with two features classifies a region as a 'defect' if each feature exceeds a threshold value independently. The vector space as well as the decision lines for this classification can be seen in Figure (8). The performance obtained is S_n = 95.0%, 1-S_p = 1.0% (TP = 57 and FP = 230) using Set I of the data. The use of more sophisticated statistical classifiers did not improve classification performance. This is because the classes are superimposed on the feature space. The results obtained with the different statistical classifiers are summarised in Table (4), Groups 1 to 5.

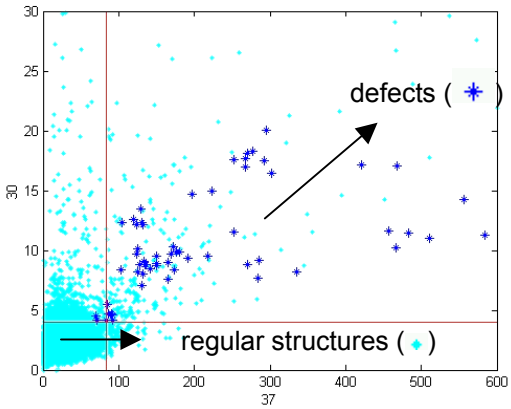


Figure 8: Vector space of features F_1 (# 37) and K_σ (# 30), the lines represent the thresholds.

[7] The 60 samples of the 'defect' class were duplicated until they equalled the number of samples in the 'regular structure' class.

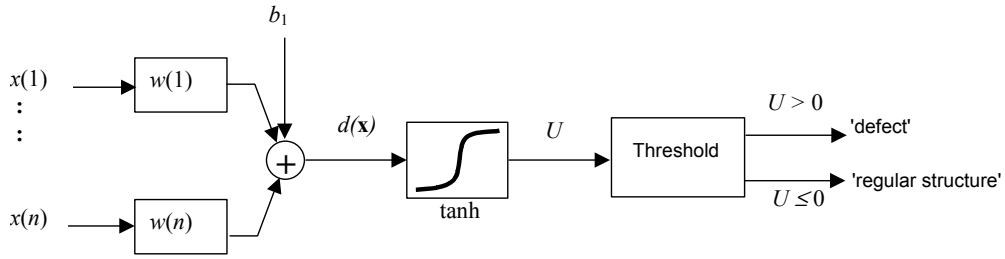


Figure 9: Model of the supervised network with retro-propagation of error. The network has only one neurone for implementation of linear classifiers

5.2 Linear classifiers using neuronal networks (NN)

When neuronal networks are used for pattern recognition criteria, it is usual to start with a study of linear classifiers with the purpose of establishing the possibility of linear separation of the classes. This is due to the operational and developmental simplicity of these classifiers compared those that are non-linear. Linear classifiers' functioning can be explained as follows: A linear discriminator decides if the sample, represented by its features grouped in a vector **x**, belongs to the class 'defect' if $d(\mathbf{x}) > 0$ where

$$d(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x} + b. \qquad (3)$$

If this is not the case, **x** is assigned to the class 'regular structure'. As can be seen, the discriminator is defined by the vector **w**, which weights the input **x,** and b which represents an off-set value. In the domain of the input, the discriminating function represents the geometric space of the points that satisfy $d(\mathbf{x}) = 0$, which form a hyperplane perpendicular to the vector **w** and at a distance (in the direction of **w**) of $-b / \|\mathbf{w}\|$ from the origin. Usually it is normalised so that $\|\mathbf{w}\|=1$, adjusting the value of b so as to not alter the inequality $d(\mathbf{x}) > 0$. In this case $d(\mathbf{x})$ measures the distance from the input **x** to the hyperplane, and is a measure of the probability of success of the classification for that specific input (Silva et al, 2001). An optimal discriminator is one that maximises the probability of success of classification. Optimal linear discriminators are a much-used technique in statistics known as Fisher Discriminators. A practical way of implementing these is through a neuronal network with one layer, where this layer contains one unique neurone per class, as described by Haykin (1994). In our work, linear classifiers were implemented using a supervised network with only neurone of the hyperbolic tangent type: $U = \tanh(d(\mathbf{x}))$. Learning was carried out using the error retro-propagation algorithm with cascade training (Haykin, 1994). Figure (9) illustrates the neuronal implementation model of the classifier that was used.

As the neuronal learning process used minimises the total mean quadratic error, training in which there are many samples of the 'regular structure' class and very few of the 'defect' class (in our case the ratio is 381:1) is not recommended because the 'regular structure' class would be dominant in this learning process, and logically the network would focus on learning to minimise the error in the 'regular structure' instead of the 'defect' class. There are two solutions for this problem: the number of samples of the 'regular structure' class can be reduced (but thus losing the network's capacity to generalise), or duplicate the samples from the smaller class thus diminishing the dominant tendency of the larger class (Haykin, 1994). In our case both possibilities were tested.

Table (4), Groups 6, 7 and 8, present the results of S_n and 1-S_p obtained with the linear classifiers for the three sets of data using the 28 features from Table (2). As can be seen, all defects were detected in all three cases. Additionally the number of false alarms (FP) is only 2 for sets II and IV, and zero for set III, which implies that the value of 1-S_p is very low for all cases. These are excellent results and show that linear separation using neuronal networks is possible between both classes when using the 28 features from Table (2).

Table 4: Performance of the statistical classifiers and neuronal networks in diverse experiments.

Group	Features	Classifier	Set	*TP*	*FP*	S_n	1-S_p
1	# 37	threshold	I	57 / 60	324 / 22,876	95.0%	1.42%
2	# 30, 37	threshold	I	57 / 60	230 / 22,876	95.0%	1.00%
3	# 30, 31, 37	linear	I	57 / 60	310 / 22,876	95.0%	1.33%
4	# 30, 37, 101	Euclidean	I	57 / 60	326 / 22,876	95.0%	1.59%
5	# 30, 37, 101	Mahalanobis	I	57 / 60	363 / 22,876	95.0%	1.59%
6	28 features of Table (2)	NN	II	60 / 60	0 / 180	100.0%	0.00%
7	28 features of Table (2)	NN	III	60 / 60	0 / 500	100.0%	0.00%
8	28 features of Table (2)	NN	IV	60 / 60	2 / 2,000	100.0%	0.10%
9	# 25, 30, 31, 33, 37, 101, 186, 208, 360, 376	NN	IV	60 / 60	2 / 2,000	100.0%	0.10%
10	# 4, 8, 25, 30, 31, 33, 37	NN	IV	58 / 60	8 / 2,000	99.6%	0.40%
11	# 30, 37	NN	IV	52 / 60	7 / 2,000	86.5%	0.35%
12	# 30, 37	NN	V	60 / 60	2 / 2,000	100.0%	0.10%
13	# 30, 37	NN	VI	60 / 60	558 / 22,876	100.0%	2.44%
14	P_1, P_2 of PCA of 28 features	NN	IV	60 / 60	2 / 2,000	100.0%	0.10%
15	P_1 of PCA of features # 30 and 37	NN	IV	60 / 60	2 / 2,000	100.0%	0.10%

With the purpose of finding a classifier that uses fewer features, the ten most important features found with the SFS method using the Fisher discriminator (see Fig. (6a)) and the variable S_p @ S_n=100% (see Fig. (6b)), a set of 10 features made up of the 6 first features of each method, were selected (as can be seen in Fig. (6) features # 30 and #360 are common to both selections). The performance of these features on a linear classifier with Set IV of the data, was evaluated. As seen in Table (4), Group 9, the results obtained were also S_n = 100% and 1-S_p = 0.1% (*TP* = 60, *FP* = 2). Therefore, with a dimensional reduction from 28 to 10 features, the results are identical, showing that a reduction in the dimensions of the input data is both possible and recommendable.

Subsequently, a new training set was developed from Set IV of the data, using the first 8 features in Table (2) that represent the lowest computational cost. The results obtained are shown in Table (4), Group 10. As can be seen S_n = 96.6% and 1-S_p = 0.4% (*TP* = 58, *FP* = 8), which represents a slightly inferior index from that obtained with the 28 features used initially. Nevertheless, the percent reduction is small compared to the difficulty represented by the extraction of the totality of the features.

Figure (8), which illustrates the vector space with features # 30 and # 37, shows the good separation achieved between classes using these features. For this reason a new experiment was carried out using Set III, but using only the two features mentioned. Initially, for this set of data, the results were only S_n = 86.5% and 1-S_p = 0.35% (see Table (4), Group 11). In this case the reduction in sensitivity is quite significant. The reason for this reduction is probably due to the difference in the number of samples per class, which is accentuated even more by the elimination of 26 features. In order to test this hypothesis the number of 'defect' data were doubled until 2.000 samples sere obtained (Set 4 of the data), thus equalling the data belonging to the 'regular structure' and giving equal importance to both classes during the network training process. The results can be seen in Table (4), Group 12. Surprisingly, the performance was the same as that obtained with the set composed of the 28 features, that is S_n = 100% and 1-S_p = 0.1%. Thus it is evident that the reduction in the input data is satisfactory.

As a final test, the neuronal network was trained using Set IV composed of 22,876 samples from the class 'defect' (duplicates) and 22,876 samples from the 'regular structure' class. As the training of this network represents an elevated computational cost, only two features were considered, # 30 and # 37 which showed a very good performance in their evaluation. The results were S_n = 100% and 1-S_p = 2.44% (see Table (4), Group 13).

5.3 Principal Components of Linear Discrimination

It is difficult to visualize the vector space of the features when the system has more than three dimensions. This is the case shown in Table (2) in which there is a set of 28 features. One way of visualising this space is by means of its projection onto another two or three-dimensional space. Commonly this reduction is achieved by means of the Karhunen-Lòeve Transformation, also known as Principal Component Analysis (PCA) (Castleman, 1996).

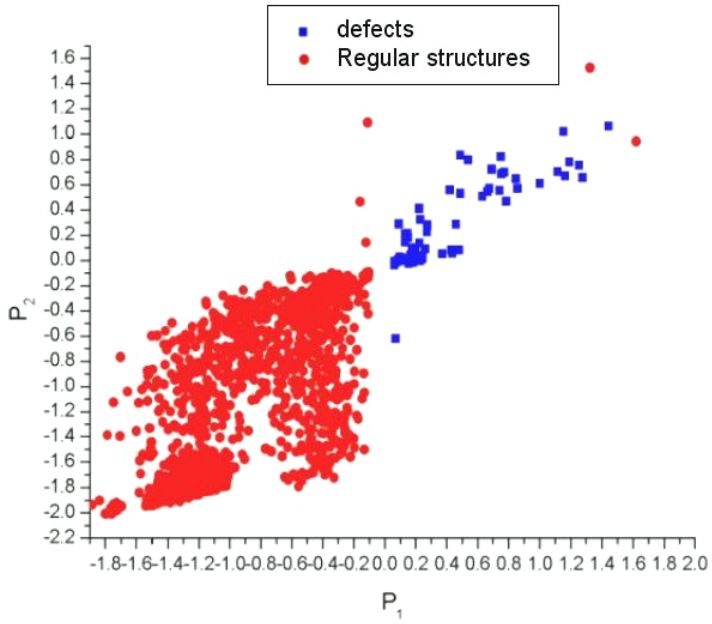


Figure 10: Two-dimensional graph of the two principal components (P_1 and P_2) of linear discrimination based on the 28 features of Table (2).

In this investigation the two principal linear discrimination components with independent activity were used, implemented by an error retro-propagation type network with only one neurone, (Silva et al, 2003), to visualise the discrimination between both classes using Set IV of the data. Figure (10) shows the two-dimensional graph made up of the two principal components (P_1 and P_2) obtained from the 28 features. As can be seen it is possible to separate both classes using a simple linear classifier. In this case there are only two false alarms, and the results are summarised in Table (4), Group 14.

Another application of principal discrimination components is the reduction in the dimensions of the classifier input data; in our case the candidate features total 405 (see Table (1)). These features can be substituted by the principal discrimination components reducing the system's dimensions to 2, 3 or 4 dimensions[8]. In this case the classification result obtained showed that # 30 and # 37 are extremely important and sufficient for obtaining a high success index. On the basis of a principal component analysis for these two features, and using the first principal component as an input variable for the linear classifier, a new experiment can be carried out. In other words with the first resulting component, defined as the projection of the input in the principal direction of linear discrimination (Silva et al, 2003), we now have a new feature which is the linear combination of the first two. A new performance evaluation was carried out using this component and the result indicates the same S_n = 100% and S_p = 99.9% (see Table (4), Group 15). Summarising, on the basis of an initial 405 features, we arrive at only one. These results show once more what Silva et al (2002b) concluded in their work: what is important is the *quality* and not the *quantity* of the features used.

6. Conclusions

This work presents an exhaustive analysis of segmented regions of noisy radioscopic images with the purpose of carrying out an automatic detection of flaws in cast aluminium pieces. This study presents over 400 features extracted from approximately 23,000 regions. The analysis of the extracted features indicates that information about flaws is found more often in intensity features, especially contrast and texture features, and less so in geometric features. Diverse experiments were

[8] This reduction implies a simplification only in the design of the classifier. It should be remembered that to obtaing the principal components it is necessary to have access to all the original features. In the example mentioned, to obtain the two, three, or four components, it is necessary to first extract the 405 initial features.

carried out under different conditions, evaluating the performance of statistical classifiers and those based on neuronal networks in varied combination of features and training data sets. The results are similar, and in almost all experiments it was possible to detect a large number of real flaws with a small number of false alarms (S_n > 95%, 1-S_p<2.5%). Although it is true that these results are very good in percentage terms, they are not so in absolute terms in that the number of false alarms remains high. Considering that there are approximately 23,000 hypothetical flaws in 50 images, a value of 1-S_p = 1% indicates that on average there are approximately 5 false alarms per image, which is unacceptable in industry. Nevertheless, these results are a substantial improvement over those presented by Mery & Filbert (2002a), in which, for the same images, the results obtained had a value of 1-S_p = 8% which indicates approximately 15 false alarms per image. The improvement in performance is due principally to the development of new contrast features presented in (Mery, 2003). This leads us to the conclusion previously mentioned that it is the *quality* and not the *quantity* of the features used that is of importance.

This line of investigation belongs to the area of detection of flaws without a priori knowledge of the structure of the image and with noisy images, i.e., images taken without using integration techniques. It is known that for these images false alarms can be eliminated using a posterior analysis based on image sequence analysis (Mery & Filbert, 2002b). Nonetheless, is would be interesting to test the methods described in this article on the analysis of radioscopic images in which noise has been reduced by means of an integration process by averaging various images of the same scene. It is quite probable that false alarms could be further diminished using this process.

In future work we plan to investigate the relevance of the extracted features according to relevance criteria using neuronal networks described in (Silva et al, 2002b); delve into other feature selection techniques (Jain et al, 2000); implement other classifiers (based on more complex neuronal networks, or on fuzzy logic, or a fusion of classifiers) and experiment with new features that improve classification performance.

It is also necessary to find new data with the purpose of having access to a more representative sample of the 'defect' class as this would allow a better training of classifiers and access to separate data sets for training and independent tests. Given that in the present investigation the number of samples in the 'defect' class was very small, tests were performed on the same data set that was used for training.

In order that other members of the NDT community can make contributions in this area, the data used in this investigation are available in:

http://www.diinf.usach.cl/~dmery/papers/DATA_PANNDT2003a.zip

The data are in a plain text format. The table contains 22,936 rows and 29 columns. Each row is one sample. The first 28 columns correspond to the 28 features of Table (2). For example column 7 is feature F_1. Column 29 constitutes the ideal classification supervision variable y_k, where y_k is 1 (or 0) if the sample k (k=1...22,936) belongs to the class 'defect' (or 'regular structure').

Acknowledgements

The authors wish to thank the Department for Research and Development of the Universidad de Santiago de Chile (project DICYT 06-0119MQ), and the CNPq, CAPES and FAPERJ for financing this investigation, and the Technical University of Berlin for financing the participation in this Symposium.

References

Bishop C. M., 1995: *Neuronal Networks for Pattern Recognition.* New York. Oxford University Press. 1995.

Castleman, K., 1996, *Digital Image Processing.* Prentice-Hall, Englewood Cliffs, New Jersey.

Danielsson, P.Y., 1978: A new shape factor. *Computer Graphics and Image Processing* 7, 292–299.

Egan, J., 1975: *Signal detection theory and ROC analysis*. Academic Press, New York.

Fitzibbon, A., Pilu, M., Fisher, R.B., 1999, Direct Least Square Fitting Ellipses , *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(5):476-480.

Fukunaga, K., 1990, *Introduction to statistical pattern recognition* , Academic Press, Inc., Second Edition, San Diego.

Haralick, R., Shanmugam, K., Dinstein, I., 1973, Textural features for image clasification. *IEEE Trans. on Systems, Man, and Cybernetics* SMC-3, 610–621.

Haykin S. *Neuronal Networks – A Comprehensive Foundation*. Macmillian College Publishing. Inc. 1994.

Jain, A.K., Duin, R.P.W., Mao, J., 2000, Statistical Pattern Recognition: A Review , *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(1):4-37.

Kamm, K.F., 1998, Grundlagen der Röntgenabbildung. In *Moderne Bildgebung: Physik, Gerätetechnik, Bildbearbeitung und -kommunikation, Strahlenschutz, Qualitätskontrolle*, K. Ewen (Ed.), 45-62, Georg Thieme Verlag, Stuttgart, New York, 1998.

Kay, S.M., 1998, *Fundamentals of Statistical Signal Processing: Detection Theory.* Prentice Hall Signal Processing Series, Volume 2. 1998.

Mery, D., Filbert, D., 2002a, Clasification of Potential Defects in Automated Inspection of Aluminium Castings Using Statistical Pattern Recognition. In *Proceedings of 8th European Conference on Non-Destructive Testing (ECNDT 2002)*, Jun. 17-21, Barcelona, Spain.

Mery, D., Filbert, D., 2002b, Automated Flaw Detection in Aluminum Castings Based on the Tracking of Potential Defects in a Radioscopic Image Sequence. *IEEE Transactions on Robotics and Automation*, 18(6): 890-901.

Mery, D., 2003, Crossing line profile: a new approach to detecting defects in aluminium castings. In *Proceedings of the Scandinavian Conference on Image Analysis (SCIA-2003)*, June 29 – July 2, Göteborg, Sweden. To be published in *Lecture Notes in Computer Science*, 2003.

Mery, D., Filbert, D., Jaeger, Th., 2003a, Image Processing for Fault Detection in Aluminum Castings. In *Analytical Characterization of Aluminum and Its Alloys* , Ed. C.S. MacKenzie and G.Y. Totten, Marcel Dekker Inc. , New York (in Press).

Mery, D., da Silva, R., Calôba, L.P., Rebello, J.M.A, 2003b: Pattern Recognition in the Automatic Inspection of Aluminium Castings. *Insight, Journal of the British Institute of Non-destructive Testing*, 45(7):441-449.

Silva, R.R., Siqueira, M. H. S., Calôba, L.P. et al, 2001, Radiographics Pattern Recognition of Welding Defects using Linear Clasifiers. *Insight, Journal of the British Institute of Non-destructive Testing*, 43(10):669-674.

Silva R R, Siqueira M H. S, Calôba L P, et al., 2002a, Contribution to the Development of a Radiographic Inspection Automated System. In *Proceedings of 8th European Conference on Non-Destructive Testing (ECNDT 2002)*, Jun. 17-21, Barcelona, Spain.

Silva R R, Siqueira M H. S, Calôba L P, et al., 2002b Evaluation of the Revelant Characteristic Parameters of Welding Defects and Probability of Correct Clasification using Linear Clasifiers. *Insight, Journal of the British Institute of Non-destructive Testing*, 44(10): 616-622.

Silva R R, Calôba L P, Siqueira M H. S, Rebello, J.M.A., 2003, Patterns nonlinear clasifier of weld defects in industrial radiographies. In *Proceedings of 3rd Panamerican Conference for Nondestructive Testing (PANNDT 2003)*, Rio de Janeiro, 02-07 Junio, 2003.

Sonka, M., Hlavac, V., Boyle, R., 1998: *Image Processing, Analysis, and Machine Vision* , 2 edn. PWS Publishing, Pacific Grove, CA.

Zahn, C., Roskies, R., 1971, Fourier descriptors for plane closed curves. *IEEE Trans. Computers*, C-21 269–281.