Automated Inspection of Aluminum Castings using Classifier Fusion Strategies

Domingo Mery*[1], Max Chacón[2], Luis González[2], Luis Muñoz[2]

* Corresponding author:
[1] Departamento de Ciencia de la Computación
Pontificia Universidad Católica de Chile
Av. Vicuña Mackenna 4860(143), Santiago de Chile
e-mail: dmery@ing.puc.cl

[2] Departamento de Ingeniería Informática
Universidad de Santiago de Chile (USACH)
Av. Ecuador 3659, Santiago de Chile

September 7, 2004

Abstract

Generally, the flaw detection in automated visual inspection consists of two steps: a) identification of potential defects using image processing techniques, and b) classification of potential defects into 'defects' and 'regular structures' (false alarms) using a pattern recognition methodology. In the second step, since several features can be extracted from the potential defects, a feature selection must be performed. In this paper, several known classifiers are studied in the automated visual inspection: threshold, Euclidean, Mahalanobis, polynomial, support vector machine (SVM) and neural network. First, the performance of the classifiers is assessed individually. Second, the classifiers are combined in order to improve their performance. Seven fusion strategies in the combination were evaluated: 'and', 'or', 'majority vote', 'product', 'sum', 'max' and 'median'.

The proposed methodology was tested on real data acquired from 50 noisy radioscopic images of aluminum wheels, where 23000 potential defects (with only 60 real defects) were segmented and 405 features were extracted in each potential defect. Using fusion of classifiers a very good performance was achieved, yielding a sensitivity of 100% and specificity of 99.97%.
Keywords: automated visual inspection, fusion of classifiers, aluminum castings, radioscopy.

1 Introduction

Visual inspection is defined as a quality control task that uses visual data to determine if a product deviates from a given set of specifications[1]. Inspection usually involves measurement of specific part features such as assembly integrity, surface finish and geometric dimensions. If the measurement lies within a determined tolerance, the inspection process considers the product as accepted for use. In industrial environments, inspection is performed by human inspectors and/or automated visual inspection (AVI) systems. Human inspectors are not always consistent and effective evaluators of products, because inspection tasks are monotonous and exhausting. Typically, there is one rejected in hundreds of accepted products. It has been reported that human visual inspection is at best 80% effective. In addition, achieving human '100%-inspection', where it is necessary to check every product thoroughly, typically requires high level of redundancy, thus increasing the cost and time for inspection. For these reasons, in many applications a batch inspection is carried out. In this case, a representative set of products is selected in order to perform inferential reasoning about the total. Nevertheless, in certain applications a '100%-inspection' is required. This is the case of light-alloy castings produced for the automotive industry, where it is necessary to ensure the safety of consumers [10].

Light-alloy castings produced for the automotive industry, such as wheel rims, steering knuckles and steering gear boxes are considered important components for overall roadworthiness. In these objects, defects are manifested, for example, by bubble-shaped voids, cracks, slag formations or inclusions. Radioscopy rapidly became the accepted way for controlling the quality of die cast pieces through visual or computer-aided analysis of X-ray images. The purpose of this NDT method is to identify casting defects, which may be located within the piece and thus are undetectable to the naked eye. An example of such defects in a light-alloy

[1]For a comprehensive overview of automated visual inspection, the reader is referred to an excellent review paper by Newman and Jain [11]. The information given in this paragraph was extracted from this paper.

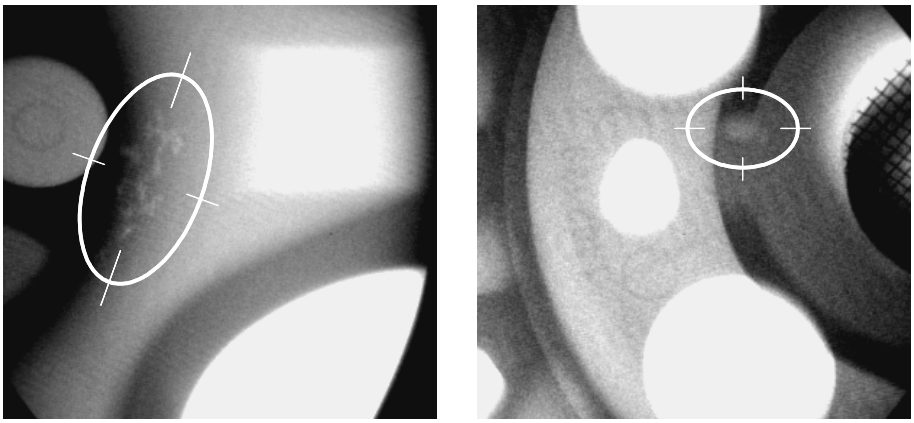


Figure 1: Radioscopic images of wheels with defects.

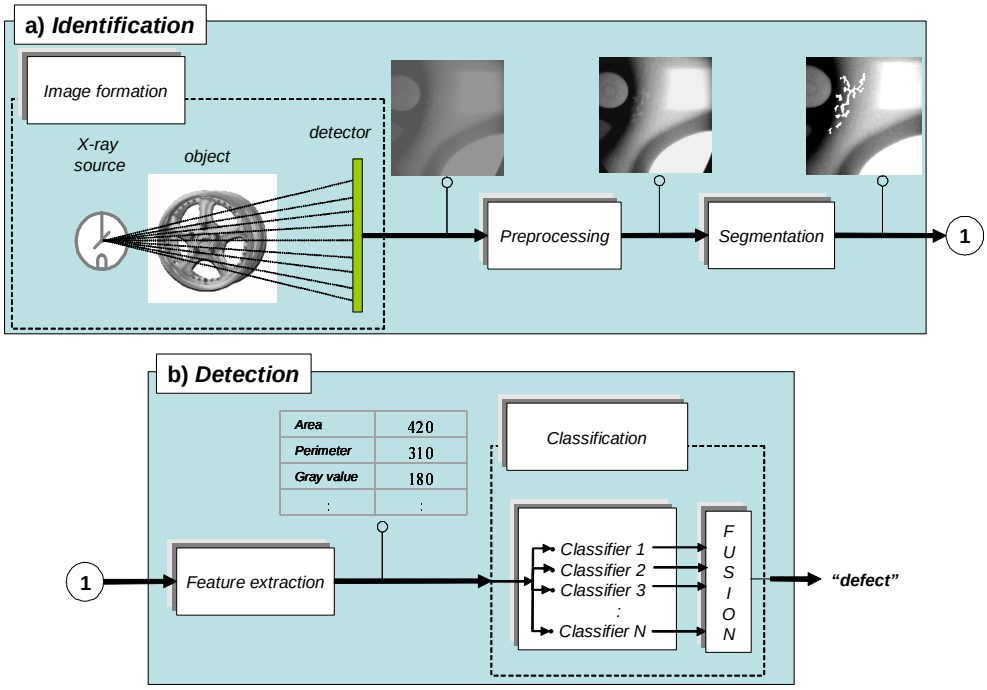


Figure 2: Inspection process using classifier fusion: a) identification of potential defects, b) detection of defects.

wheel is shown in the X-ray image in Fig. 1.

Typically, the automatic process used in automated visual inspection, as shown in Fig. 2, follows an image analysis methodology based on a) image processing and b) pattern recognition techniques that can be summarized as:

a) Identification of potential defects:

• *Image formation*: A digital image of the object being tested is taken and stored in the computer.

• *Image pre-processing*: The quality of the image is improved in order to enhance the details of the image.

• *Image segmentation*: Each potential flaw of the image is found and isolated from the rest of the scene.

Two classes of regions are possible in a digital image taken in visual inspection: regions belonging to regular structures of the object, and those relating to defects. In the computer-aided visual inspection, the aim is to identify these two classes automatically. In step a), the identification of real defects must be ensured. Nevertheless, using this strategy an enormous number of regular structures (false alarms) are identified. For this reason, a second step is required:

b) Detection of defects:

• *Feature extraction*: The potential flaws are measured and some significant characteristics are quantified.

• *Classification*: The extracted features of each potential flaw are analysed and assigned to one of the two following classes: 'defect' or 'regular structure'. Different classifiers can be designed. The classification can be performed using only one classifier or a combination of several classifiers as shown in Fig. 2.

The detection attempts to separate the existing defects from the regular structures. In step b), since several features can be extracted from the potential defects, a feature selection must be performed eliminating features that are correlated or provide no information whatsoever. In addition, since the two classes show a skewed distribution (usually, there are more than 100 false alarms for each real defect), the classifier(s) must be carefully trained.

In this paper, we deal with the detection problem, i.e., which features can be selected, and how the two classes can be efficiently separated in a skewed class distribution. For the first task, the sequential forward selection (SFS) algorithm was investigated. For the second task, several known classifiers are studied: threshold, Eculidean, Mahalanobis, polynomial,

support vector machine (SVM) and neuronal network. The performance of the classifiers is firstly assessed individually. Afterwards, the classifiers are combined in order to improve the performance. Seven fusion strategies are studied: 'and', 'or', 'majority vote', 'product', 'sum', 'max' and 'median' [2]. The proposed methodology was tested on real data acquired from 50 noisy radioscopic images of aluminum wheels, where 23000 potential defects (with only 60 real defects) were segmented and 405 features were extracted in each potential defect. The hypothesis of our work is: the performance of the classification can be improved by combining classifiers that offer complementary information.

The rest of the paper is organized as follows: in Section 2 the identification of potential defects is briefly described. In Section 3 the feature extraction and selection is outlined. Section 4 presents the classifiers used in this investigation. In Section 5, the fusion strategies are explained. The experiments and results on real data are presented in Section 6. Finally, Section 7 gives concluding remarks.

2 Identification of potential defects

The contrast in the X-ray image between a flaw and a defect-free area of the specimen is distinctive. In an X-ray image we can see that the defects, such as voids, cracks or bubbles, show up as bright features. The reason is that the X-ray attenuation in these areas is shorter. Hence, according to the principle of differential absorption, the detection of flaws can be achieved automatically using image processing techniques that are able to identify unexpected regions in a digital X-ray image.

The X-ray image taken with an image intensifier and a CCD camera (or a flat panel detector), must be pre-processed to improve the quality of the image. In our approach, the pre-processing techniques are used to remove noise, enhance contrast, correct the shading effect and restore blur deformation.

The segmentation of potential flaws identifies regions in radioscopic images that may correspond to real defects. Two general characteristics of the defects are used to identify them: a) a flaw can be considered as a connected subset of the image, and b) the grey level difference between a flaw and its neighborhood is significant. According to the previously mentioned characteristics, a simple automated segmentation approach was suggested in [9] (see Fig. 3). First, a Laplacian-of-Gaussian (LoG) kernel and a zero crossing algorithm [2] are used to detect the edges of the X-ray images. The LoG-operator involves a Gaussian lowpass filter which is a good choice for the pre-smoothing of the noisy images that are obtained without frame averaging. The resulting binary edge image should produce closed and connected contours

[2] In the literature, *combination* and *fusion* of classifiers are used indistinctly: the idea is to design several classifiers and to combine their individual outputs in order to reach a consensus decision about the classification.

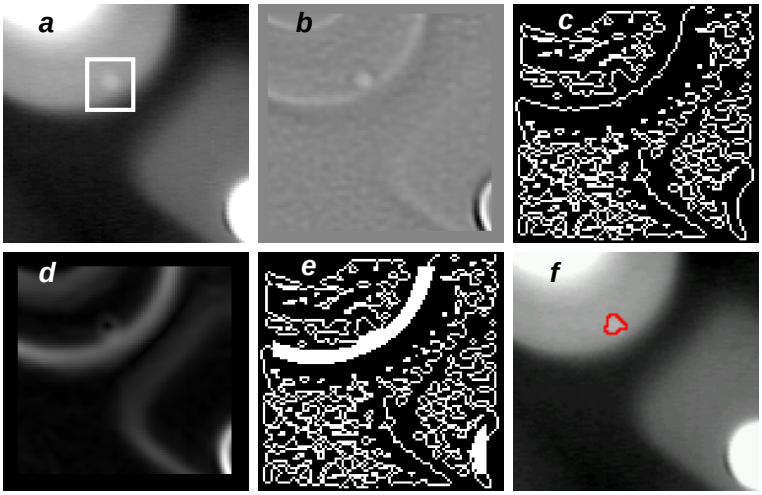


Figure 3: Detection of flaws: a) radioscopic image with a small flaw at an edge of a regular structure, b) Laplacian-filtered image with $\sigma = 1.25$ pixels (kernel size = 11×11), c) zero crossing image, d) gradient image, e) edge detection after adding high gradient pixels, and f) detected flaw after classification step.

which demarcate *regions* at real flaws. However, a flaw may not be perfectly enclosed if it is located on the edge of a regular structure as shown in Fig. 3c. In order to complete the remaining edges of these flaws, a thickening of the edges of the regular structure is performed as follows: a) the gradient of the original image is calculated (see Fig. 3d); b) by thresholding the gradient image at a high grey level a new binary image is obtained; and c) the resulting image is added to the zero crossing image (see Fig. 3e). Afterwards, each closed region is segmented as potential flaw. For details see a description of the method in [9].

This is a very simple detector of potential flaws with a large number of false alarms flagged erroneously. However, the advantages are as follows: a) it is a single detector (it is the same detector for each image), b) it is able to identify potential defects independent of the placement and the structure of the specimen, i.e., without a-priori information of the design structure of the test piece, and c) the detection rate of real flaws is very high (approximately 90%). In order to reduce the number of false positives, the segmented regions must be measured and classified.

In our experiments, 50 X-ray images of aluminum wheels were analyzed. In the segmentation 22936 potential flaws were obtained, in which there were only 60 real flaws, i.e., the skew is 381:1. Some of the real defects were existing blow holes. The other defects were produced by drilling small holes in positions of the casting which were known to be difficult to detect.

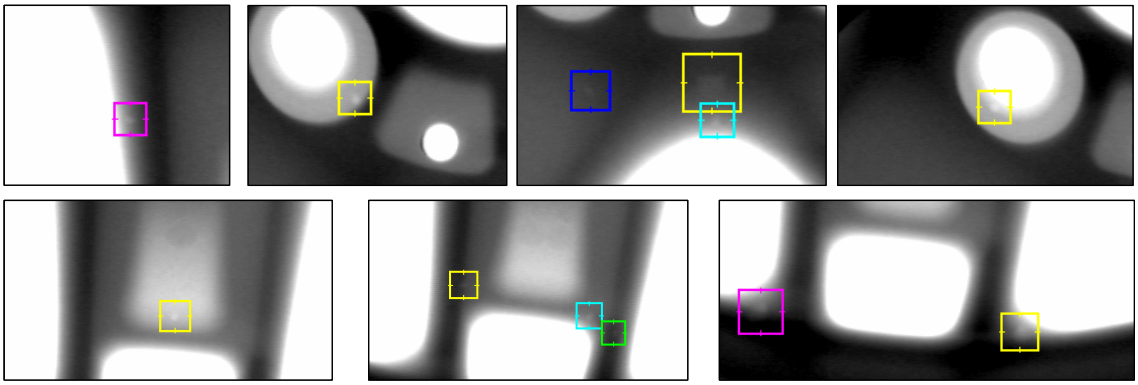


Figure 4: X-ray images used in our experiments.

Examples of the radioscopic images used in our investigation are shown in Fig. 4.

3 Feature extraction and feature selection

In order to reduce the number of false alarms of the segmentation, the segmented regions must be measured and classified into one of the two classes: 'regular structure' or 'defect'. Feature extraction is centered principally around the measurement of geometric properties (area, perimeter, form factors, Fourier descriptors, invariant moments, etc.), and on the intensity characteristics of regions (level of grey, contrasts, gradient, second derivative, texture features, etc.). Geometric features provide information relative to the size and form of the segmented potential flaws. Intensity features, on the other hand, provide information about the grey value of the segmented regions. Thus, a large number of features can be extracted.

In our experiments, for each potential defect, 405 features were extracted (f_1, f_2, ..., f_{405}), of which were 37 geometric features and 368 intensity features. The reader is referred to [8] for details of the extracted features. However, not all features provide information for detection. In order to build a compact and accurate model, irrelevant and redundant features must be removed. A pre-selection of features can be performed by eliminating those that present an area under the *receiver operation characteristic* (ROC) curve [3] $A_z < A_{z\,\mathrm{min}}$ (e.g., $A_{z\,\mathrm{min}} = 0.8$) and a Fischer discriminator $J < k\ J_{\mathrm{max}}$ (e.g., $k = 0.2$), where J_{max} corresponds to the maximum Fisher discriminator obtained upon evaluation of each of the features. Additionally, if two features present a high correlation coefficient C (e.g., $|C| > 0.9$) then the feature with a lower A_z can be eliminated. In our experiments, only 28 (of 405) features were pre-selected using this methodology.

From the pre-selected features, a set of features can be selected using well known feature selection algorithms, e.g, *Sequential Forward Selection* (SFS) [5]. This method selects the best single feature and then adds one feature at a time that, in combination with the selected features, maximizes classification performance. The iteration is stopped once no considerable improvement in the performance is achieved on adding a new feature. By evaluating selection performance we can use the Fisher discriminator, which ensures: i) a small intraclass variation and ii) a large interclass variation in the space of the selected features. In the feature selection, the best ten features obtained with SFS algorithm were chosen for the classification. They correspond to four contrast features and six texture features.

4 Classifiers

Once the proper features are selected, a classifier can be designed. Typically, the classifier assigns the selected features, arranged in a *feature vector* $\mathbf{x}$, to one of the two classes: 'regular structure' or 'defect', that are denoted by ω_0 and ω_1 respectively. Some of the most important classifiers used in pattern recognition were evaluated in our investigation, namely:

• **Threshold** (Thr): The decision boundaries of class ω_1 define a hypercube in feature space, i.e., if $\mathbf{x}$ is located between decision thresholds then the feature vector is assigned to class ω_1. The thresholds are chosen from the representative sample [3].

• **Euclidean** (Euc): A mean value of each class ($\bar{\mathbf{x}}_0$ and $\bar{\mathbf{x}}_1$) of the representative sample is calculated. A feature vector $\mathbf{x}$ is assigned to class ω_j if the Euclidean distance metric $\| \mathbf{x}-\bar{\mathbf{x}}_j \|$ is minimal, for $j = 0, 1$ [3].

• **Mahalanobis** (Mah): The Mahalanobis classifier employs the same concept as the Euclidean classifier. However, it uses the distance metric called the 'Mahalanobis distance' [3].

• **Polynomial** (Pol): in which a quadratic combination of the selected features is used for a polynomial expansion of the discriminant function d. If $d > \theta$ then the feature vector is assigned to class ω_1, otherwise to class ω_0. Using a least-squares approach, function d can be estimated from an ideal known function d^*, that is obtained from a representative sample [1].

• **Linear SVM** (SVM_L): Support vector machine (SVM) finds an optimal hyperplane that correctly classifies feature vectors as much as possible and separates the feature vectors of the two classes as far as possible [12].

• **Polynomial SVM** (SVM_P): This classifier employs the same concept as the linear SVM. However, the separation is obtained by a polynomial function [12].

• **Radial Basis Function** (SVM_R): The separation in this SVM approach is obtained by a radial basis function based on a Gaussian function [12].

• **Neural network** (NN): The selected features are inputs of a neural network that is trained in order to obtain the desired class of the training data. In our study, we used a Multi Layer

Perceptron trained with back-propagation algorithm [4].

Results obtained in the classification are presented in Section 6.

5 Fusion strategies

By evaluating the performance of several classifiers, it had been observed that although one of the classifiers would yield the best performance, the sets of patterns misclassified by the other classifiers would not necessarily overlap. Thus, the other classifiers offer complementary information, which can be useful to better classify a pattern. It has been reported that by combining the individual outputs, using simple fusion methods, a higher performance is obtained [7]. In our study, the following fusion strategies are investigated in automated visual inspection:

• **'and':** a feature vector $\mathbf{x}$ is assigned to class ω_1 ('defect') if all individual classifiers assign the feature vector to this class.

• **'or':** a feature vector $\mathbf{x}$ is assigned to class ω_1 if at least one individual classifier assigns the feature vector to this class.

• **'majority vote':** a feature vector $\mathbf{x}$ is assigned to class ω_1 if the majority of the individual classifiers assigns the feature vector to this class.

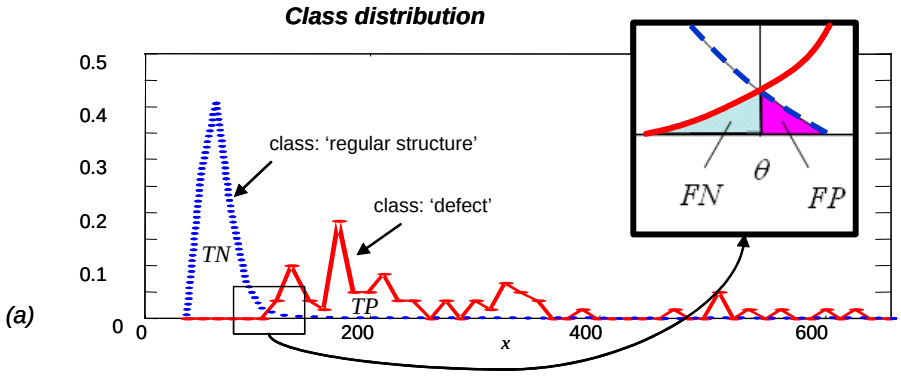
• **'sum':** According to the Bayesian decision theory, a pattern $\mathbf{x}$ should be assigned to class ω_j for which a posteriori probability $P(\omega_j|\mathbf{x})$ is maximum. In practice, the j-th classifier can provide only an estimate $P_j(\omega_i|\mathbf{x})$ of the true a posteriori class probability $P(\omega_i|\mathbf{x})$. The idea of classifier combination is to obtain a better estimate of the a posteriori class probabilities by combining all of the individual estimates and, thus, reducing the classification error. In this case, the estimation is obtained by averaging the $P_j(\omega_i|\mathbf{x})$ for $j = 1, .., N$, where N is the number of classifiers. The fused decision is obtained by applying the maximum value selector to the class dependent averages of the outputs of individual estimations [6].

• **'product'**, **'max'**, **'median'**: These strategies employ the same concept as 'sum'. However, the a posteriori class probabilities are estimated by computing the product, the maximum and the median of the individual estimates [12].

Results obtained in the fusion of classifiers are presented in the next Section.

6 Experiments and results

As explained in Sections 2 and 3, 50 radioscopic images of aluminum wheels were analyzed as follows: i) in the segmentation 22936 potential flaws were obtained, in which there were only 60 real flaws; ii) for each potential defect, 405 features were extracted; iii) only 10 of them



Confusion matrix

the region is classified as→ / the region is ↓	'regular structure'	'defect'
Regular structure	*TN*	*FP*
defect	*FN*	*TP*

(b)

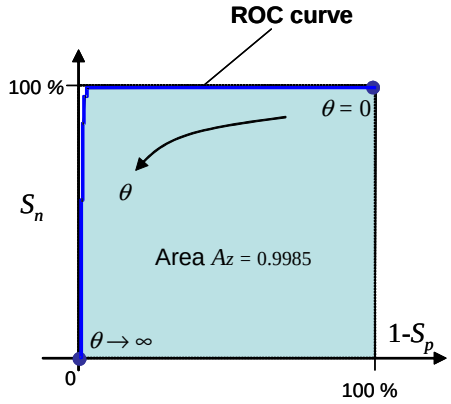


Figure 5: Defect Classification using threshold classifier and a single feature. a) Distribution of classes using a single feature x. b) Confusion matrix. c) ROC curve varying a threshold θ.

were selected. In this Section we present the results obtained with this data using fusion of classifiers.

The classifiers were designed using a training data set defined as 50% of the segmented defects (30 samples) and 50% of the segmented regular structures (11438 samples). The parameters involved in each classifier were adjusted in order to obtain the best performance in the training data.

The performance was assessed by computing the receiver operation characteristic (ROC) curve [3]. The ROC curve is defined as a plot of the 'sensitivity' (Sn) against the '1-specificity'

$(1 - Sp)$ as shown in Fig. 5:

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

where

TP is the number of true positives (flaws correctly classified);

TN is the number of true negatives (regular structures correctly classified);

FP is the number of false positives (false alarms, i.e., regular structures classified as defects); and

FN is the number of false negatives (flaws classified as regular structures).

Ideally, $S_n = 1$ and $1 - S_p = 0$, i.e., all flaws are detected without flagging false alarms. The ROC curve permits the assessment of the detection performance at various operating points. The area under the ROC curve (A_z) is normally used as performance measure because it indicates how reliably the detection can be performed. A value of $A_z = 1$ gives perfect classification, whereas $A_z = 0.5$ corresponds to random guessing. A real example is illustrated in Fig. 5c, where a high performance is achieved.

The area A_z is presented for each individual classifier in Table 1. In addition, the best operation point –sensitivity (S_n) and 1-specificity ($1-S_p$)– is presented in the last two columns. We observe that the best performance was achieved by the classifier NN yielding $A_z = 0.9994$. However, there are other classifiers that have obtained similar performances (see for example SVM_P with only 0.12 false positives per image, or SVM_L where all defects were detected).

Table 1: Performance of individual classifiers

Classifier	Name	A_z	TP	FP	FP/image	S_n	$1 - S_p$
Thr	Threshold	0.9985	58	223	4.46	96.6%	0.97%
Euc	Euclidean	0.9789	57	448	8.96	95.0%	1.95%
Mah	Mahalanobis	0.9987	58	122	2.44	96.6%	0.53%
Pol	Plynomial	0.9989	58	69	1.38	96.6%	0.30%
SVM_L	Linear SVM	0.9967	60	151	3.02	100.0%	0.66%
SVM_P	Polynomial SVM	0.9985	58	6	0.12	96.6%	0.03%
SVM_R	Radial Basis Function SVM	0.9969	57	33	0.66	95.0%	0.14%
NN	Neural Network	0.9994	59	36	0.72	98.3%	0.15%

Table 2: Performance of fusion strategies (shaded(yes)/non-shaded(no) indicate used/not used in the fusion)

Fusion strategy	Thr	Euc	Mah	Pol	SVM_L	SVM_P	SVM_R	NN	*TP*	*FP*	*FP*/image	S_n	1-S_p
'and'	yes	no	yes	yes	no	yes	no	yes	58	1	0.02	96.6%	0.004%
'or'	no	no	no	no	no	yes	yes	yes	59	8	0.16	98.3%	0.034%
'majority vote'	no	no	no	no	no	yes	yes	yes	60	7	0.14	100.0%	0.03%
'sum'	no	no	no	no	no	no	yes	yes	57	6	0.12	95.0%	0.026%
'product'	no	no	no	no	no	no	yes	yes	59	21	0.42	98.3%	0.091%
'max'	no	no	no	no	no	yes	yes	yes	57	6	0.12	95.0%	0.026%
'median'	no	no	no	no	no	yes	yes	yes	57	6	0.12	95.0%	0.026%

By combining the individual classifiers, better performances were achieved. However, the best results were not obtained by combining *all* of them. It was observed that the performance was improved when individual classifiers offered complementary information. The best results evaluating each fusion strategy are presented in Table 2. For example in fusion strategy 'or', the best performance was obtained by using classifiers SVM_P, SVM_R and NN, i.e., by combining these three individual classifiers and another one (e.g., threshold classifier) no higher performance was achieved. In six of seven of fusion strategies, the number of false positives was reduced considerably yielding less than 0.2 false alarms per image. In addition, we observe the relevance of the neural network classifier because it is present in each combination. The best performance was achieved by fusion 'majority vote', where all defects were detected flagging only 0.12 false alarms per image.

7 Conclusions

In this paper, a methodology based on combination of classifiers was suggested for the automated inspection of aluminum die castings. Eight different classifiers were combined using seven fusion strategies. The proposed methodology was tested on real data acquired from radioscopic images of aluminum wheels. It has been observed that although one of the individual classifiers would yield the best performance, the sets of patterns misclassified by the other classifiers would not necessarily overlap. Thus, the other classifiers offer complementary information, which was useful to better detection. Therefore, our initial hypothesis remains defensible, because in comparison with the best performance obtained by evaluating individual classifiers, the best combination increments the sensitivity from 98.3% to 100% and the

specificity from 99.85% to 99.97%. The significance of this result is very important in this automated visual inspection, where it is necessary i) to detect all defects in order to ensure the safety of consumers, and ii) to obtain a low rate of false alarms in order to reduce false rejects.

Acknowledgment

This work was supported in part by FONDECYT (Chilean Fund for Scientific and Technological Development) under grant no. 1040210.

References

[1] H. Boerner and H. Strecker. Automated X-ray inspection of aluminum casting. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 10(1):79–91, 1988.

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

[3] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. John Wiley & Sons, Inc., New York, 2 edition, 2001.

[4] S. Haykin. *Neural Networks A Comprehensive Foundation*. Macmillan College Publishing, Inc., Boston, 1994.

[5] A.K. Jain, R.P.W. Duin, and J. Mao. Statistical pattern recognition: A review. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(1):4–37, 2000.

[6] J. Kittler and F.M. Alkoot. Sum versus vote fusion in multiple classifier systems. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25(1):110–115, 2003.

[7] L.I. Kuncheva. A theoretical study on six classifier fusion strategies. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(2):281–286, 2002.

[8] D. Mery, R. da Silva, L.P. Caloba, and J.M.A. Rebello. Pattern recognition in the automatic inspection of aluminium castings. *Insight*, 45(7):475–483, 2003.

[9] D. Mery and D. Filbert. Automated flaw detection in aluminum castings based on the tracking of potential defects in a radioscopic image sequence. *IEEE Trans. Robotics and Automation*, 18(6):890–901, December 2002.

[10] D. Mery, Th. Jaeger, and D. Filbert. A review of methods for automated recognition of casting defects. *Insight*, 44(7):428–436, 2002.

[11] T.S. Newman and A.K. Jain. A survey of automated visual inspection. *Computer Vision and Image Understanding*, 61(2):231–262, 1995.

[12] X. Wang and Y. Zhong. Statistical learning theory and state of the art in SVM. In *IEEE International Conference on Cognitive Informatics (ICCI'03)*, pages 55–59, London, 2003.