THEORETICAL ADVANCES

# Robust automated multiple view inspection

**Luis Pizarro · Domingo Mery · Rafael Delpiano ·
Miguel Carrasco**

**Abstract** Recently, Automated Multiple View Inspection
(AMVI) has been developed for automated defect detection
of manufactured objects, and the framework was success-
fully implemented for calibrated image sequences. How-
ever, it is not easy to be implemented in industrial
environments because the calibration is a difficult and an
unstable process. To overcome these disadvantages, the
robust AMVI strategy, which assumes that an unknown
affine transformation exists between each pair of uncali-
brated images, is proposed. This transformation is esti-
mated using two complementary robust procedures: a
global approximation of the affine mapping is computed by
creating candidate correspondences via B-splines and
selecting those which better satisfy the epipolar constraint
for uncalibrated images. Then, we use this approximation
as initial estimate of a robust intensity-based matching
approach, which is applied locally on each potential defect.
The result is that false alarms are discarded, and the defects
of an industrial object are actually tracked along the un-
calibrated image sequence. The method is successful as
shown in our experiments on aluminum die castings.

## 1 Introduction

Recently, the *Automated Multiple View Inspection* (AMVI)
approach was developed for automated defect detection
[1]. This method is able to detect defects in two steps. In
the first step called *identification*, potential defects are
automatically identified in each image of the sequence
using a single filter without any prior knowledge of the test
object. The second step, called *tracking*, attempts to track
the identified potential defects along the image sequence.
As a result, only existing defects (and not the false detec-
tions) are successfully tracked in the image sequence
because they are located in positions dictated by the motion
of the test object. The preliminary results obtained using
AMVI methodology are promising for calibrated image
sequences. However, this approach is not suitable for all
industrial applications, because calibration is a difficult
process, and vibrations of the imaging system may induce
inaccuracies in the estimated parameters of the multiple
view geometric model. Thus, calibration is not stable and
the imaging system must be re-calibrated periodically. A
simple method was proposed in [2] to inspect objects on
uncalibrated image sequences, where structural points are
used to track the potential defects in the sequence via
bifocal constraints. The method achieves good perfor-

L. Pizarro (✉)
Mathematical Image Analysis Group, Faculty of Mathematics
and Computer Science, Saarland University, Building E11,
66041 Saarbrücken, Germany
e-mail: pizarro@mia.uni-saarland.de

L. Pizarro
Escuela de Ingeniería Informática, Facultad de Ingeniería,
Universidad Diego Portales, Av. Ejército 441, Santiago, Chile

D. Mery · R. Delpiano · M. Carrasco
Departamento de Ciencia de la Computación, Pontificia
Universidad Católica de Chile, Av. Vicuña Mackenna,
4860 Santiago, Chile
e-mail: dmery@ing.puc.cl

R. Delpiano
e-mail: delpiano@ing.puc.cl

M. Carrasco
e-mail: mlcarras@puc.cl

mance in some sequences, but fails when the structure points cannot be matched. In this case the estimation of the fundamental matrix is incorrect, therefore the tracking also fails.

Following the concept of camera multiplicity or multiple views, a reconfigurable array for machine inspection (RAMVI) was proposed in [3], where the calibration process requires manual intervention. The authors remark on the importance of the calibration for accurate inspection and propose a methodology to perform it automatically [4]. The advantage of using multiple views is also described in [5], where a visual inspection system that utilises a single camera and mirrors for simulating multiple cameras is proposed. A suitable pattern object is used to find the camera parameters before combining all views.

Calibration might be an extremely complicated procedure for real-time applications and manufacturing systems that cannot be halted for calibration purposes. Therefore, we aim to perform visual inspection avoiding an impracticable, expensive, and/or time-consuming calibration process. To overcome these drawbacks, we propose in this paper a new approach for automated visual inspection that can be directly applied on uncalibrated image sequences. To deal with the geometric distortions we assume that an unknown affine transformation exists between every pair of consecutive or non-consecutive images. We formulate the search for this affine mapping as a robust local estimation problem by means of an intensity-based matching approach. In implementing a good tracking algorithm, it is important to put special effort into finding the first (global) matching between every pair of images. This match is used to provide the initial estimate of the local optimisation process applied on each potential defect, which is crucial to attaining convergence. This is why another robust procedure is introduced at this step, which takes advantage of the geometric characteristics of the object being inspected. Using the RANSAC algorithm [6] we select the best candidate correspondences created via B-splines which better satisfy the epipolar constraint for uncalibrated images.

The rest of the paper is organised as follows: Section 2 explains our robust approach for uncalibrated AMVI. Section 3 shows preliminary results obtained with the proposed methodology. Finally, Sect. 4 delineates the concluding remarks and perspectives for future works.

## 2 Proposed method

Our proposed framework for automated visual inspection consists of five steps (A to E), which are outlined in Fig. 1. Before describing each step in detail, a brief introduction to each step is given.

The X-ray imaging systems are widely employed in non-destructive testing. They are particularly useful in automotive and aerospace industries for detecting different types of flaws: porosity, cracks, corrosion, inclusions, debris, rivets and thickness variations, among others [7–11]. The X-ray systems exploit the fact that most of the material defects are not visible. However, even in radioscopic images the signal-to-noise ratio (SNR) is low, so that the flaw signal is slightly greater than the background noise. For this reason, the identification of real defects with poor contrast can involve detection of false alarms as well. In some applications[1] one view is probably enough for examining material defects. However, the robustness of the inspection process can be increased when redundant information is used to validate flaw detection. Thus, two or more views of the same object taken from different viewpoints confirm and improve the diagnosis done by analysing only one image. This is a convenient and powerful alternative for examining complex objects where uncertainty leads to misinterpretation. A similar idea is also used by radiologists who analyse two different X-ray views of the same breast tissue to detect cancer at early stages. See for example [12], where the proposed method automatically finds correspondences between two views. Detection of delamination defects in rocket boosters is another example recommending the use of radiographic sequences [13]. Section 2.1 explains how the uncalibrated radioscopic image sequence in our experiments is obtained (Fig. 1, block A).

Once the image sequence is acquired, we search for potential defects on each view. Due to the low SNR of the images, detection of false alarms is likely. However, the detection of the real defects must be ensured in order to make the subsequent tracking possible. Potential defects are segmented and their features extracted in order to match them in a posterior correspondence analysis stage. Section 2.2 shows how the identification of potential defects is performed (Fig. 1, block B).

Going one step further, we postulate that only real defects can be followed along the image sequence, and logically false alarms discarded. Nevertheless, the uncalibrated imaging system generates images perturbed by geometric distortions, what makes any attempt to search for corresponding defects in two or more views cumbersome. To deal with this problem, we model geometric distortions as affine transformations[2]. Let $\mathbf{H}$ be a non-singular $3 \times 3$ matrix defining an affine mapping from all the homogeneous points $m_i$ in one view to the points $m_i''$ in another view, i.e. $m_i'' = \mathbf{H}\, m_i$. Three non-collinear corre-

---

[1] For instance in printed circuit board (PCB) inspection.

[2] In this paper we use affine transformations, although it is also possible to implement perspective transformations.
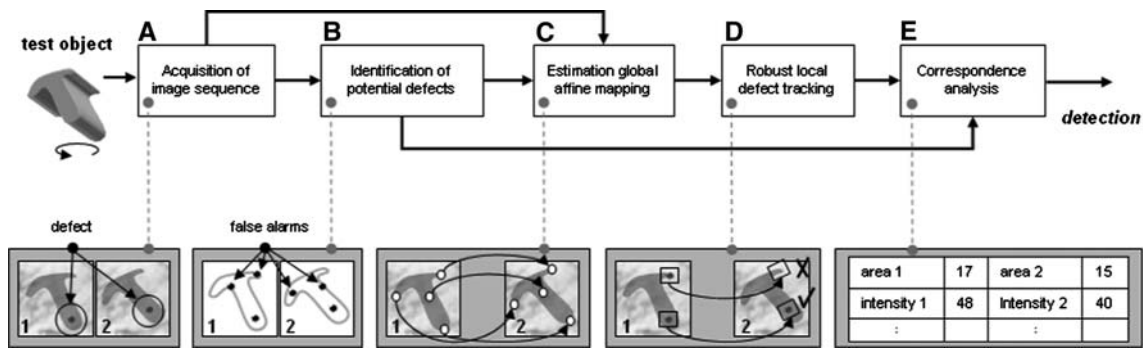
**Fig. 1** Block diagram of the proposed *robust automated multiple view inspection* system

sponding points form the following linear system of equations [chap.9, 14]

$$\underbrace{\begin{bmatrix} x_1' & x_2' & x_3' \\ y_1' & y_2' & y_3' \\ 1 & 1 & 1 \end{bmatrix}}_{\mathbf{M}'} = \mathbf{H} \underbrace{\begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{bmatrix}}_{\mathbf{M}}, \tag{1}$$

from which $\mathbf{H}$ can be computed as $\mathbf{H} = \mathbf{M}' \, \mathbf{M}^{-1}$. Additional corresponding points allow a more accurate approximation of the mapping. For $n$ points we compute

$$\mathbf{H} = \mathbf{M}'\mathbf{M}^T(\mathbf{M}\mathbf{M}^T)^{-1}. \tag{2}$$

If some unequivocal corresponding points between two consecutive or non-consecutive views were known, the problem of matching potential defective regions between these images would be solved by applying the mapping $\mathbf{H}$ to find corresponding coordinates for those regions, and comparing their extracted features obtained in the previous identification step. However, such corresponding points are not known. To find a reliable subset of such points, we first create candidate correspondences via B-splines, and then apply the RANSAC method to select the best points that satisfy the epipolar constraint. With the resulting points, a global mapping between two views is estimated by equation (2). Section 2.3 describes this procedure in detail (Fig. 1, block C).

The preceding stage presents a reliable mechanism to obtain a global approximation of the geometric mapping between two uncalibrated views. Furthermore, it is feasible to consider that the geometric distortion is non-uniform over the entire image. In fact, potential defects may be located in different parts of the image, where a slightly different distortion was induced by the uncalibrated imaging system. Thus, we want to estimate these local deformations considering the previous global computation as an initial local approximation. Therefore, we formulate the search for each potential defect from one view to another as an intensity matching problem, where the

intensities of the potential defect in the first view are to be iteratively tracked in the second view. Starting from the global affine transformation a local affine transformation for every potential defect is refined in each iteration. Moreover, it is possible to strengthen this process against illumination variations and partial occlusions[3] looking at the robust formulation of the visual matching problem. Section 2.4 details how the tracking of potential defects is performed (Fig. 1, block D).

The coordinates of every potential defective region are tracked from the first view to the second one where, in the best case, another potential defective region with similar feature was also found during the identification stage. As a direct result of the tracking process, three criteria must be fulfilled to consider a region as defective: (i) identifiability, (ii) spatial proximity, and (iii) feature proximity. Section 2.5 specifies the correspondence analysis carried out to verify the fulfillment of such requirements (Fig. 1, block E).

The proposed approach seems to be complicated, which is true from the computational point of view. However, the inspection itself is quite simple because the test object does not require placement accuracy; we only need to place and rotate the object, the rest is done by computer automatically. The bottom part of Fig. 1 shows a synthetic example where any two views of a test object are inspected. The object contains only one defect, but false alarms may appear. Ideally, the inspection system detects the flaw and discards all false alarms.

2.1 Acquisition of the image sequence

In order to facilitate the defect-tracking over the images, similar projections of the inspected object must be registered along the sequence. For this reason, and for sim-

---

[3] Occlusions appear when small flaws move in front (or behind) of a thick cross section of the casting, where X-rays are highly absorbed; and when flaws are located in the outer limits of the visible area of the casting.
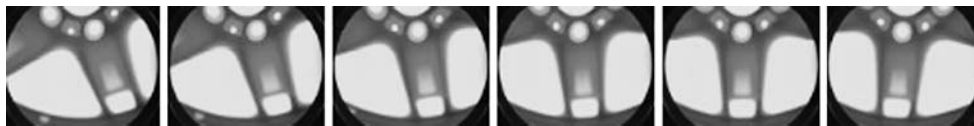
plicity, different views are taken by rotating the casting at smaller angular intervals (e.g. 5°, see Fig. 2). Each captured scene consists of only one rigid object in motion, whose 2D trajectories are smooth because there is no significant frame-to-frame motion, the velocity of the test object is constant, and the motion of the test object is generally only rotational or translational. Since many images are captured, the time of the data acquisition is reduced by taking the images without frame averaging[4]. The usual setup utilised for *automatic visual inspection* (AVI) on aluminum die castings is detailed in [1].

## 2.2 Identification of potential defects

The identification of potential defects aims at segmenting regions that may correspond to real defects. Two general characteristics of the defects are used for identification: (i) a defect can be considered as a connected subset of the image, and (ii) the grey-level difference between a defect and its neighbourhood is significant. The potential defects are identified without prior knowledge. First, a Laplacian-of-Gaussian (LoG) kernel and a zero-crossing algorithm are used to detect edges on the X-ray image. In real defects, the resulting binary edge image should produce closed and connected contours which outline *regions*. However, a defect may not be perfectly enclosed if it is located at one edge of a regular structure as shown in Fig. 3c. In order to complete the remaining edges of these defects, a thickening of the edges in 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), and afterwards, each closed region is segmented. In order to identify the potential defects, features are extracted from crossing line profiles of each segmented region. Crossing line profiles are grey-level profiles along straight lines that cross each segmented region in the middle. If the variance of the crossing line profiles is high, the segmented region is classified as potential defect [15, 16]. Later on, the extracted features are used in the stage of correspondence analysis (Sect. 2.5) to match tracked

potential defects. This is a very simple detector with more than 85% detection rate

## 2.3 Estimation of the global affine mapping

At this stage we look for a global approximation of the affine mapping **H** between any two different views. As suggested in Eq. (2), such a transformation can be accurately estimated from a set of $n$ corresponding points. This set of points can be found by performing the following five steps:

i.  **Segmentation**. It consists of isolating object parts in which the intensity values are clearly distinguishable from the background. We use the Otsu's segmentation method [17] for this task, which estimates the best separation for bimodal histograms. See Fig. 4.

ii. **Feature extraction**. For every segmented region three features are extracted: area; centre of mass $(\bar{i},\bar{j}) = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}}\right)$, in terms of the statistical moment of order $(r + s)$ $m_{rs} = \sum_{(i,j) \in \Omega} i^r j^s$, where $\Omega$ is the set of pixels of the segmented region; and the group of four affine moment invariants derived by Flusser and Suk [18].

iii. **Region matching**. This step establishes correspondences among segmented regions by measuring their similarity. The smallest norm of the difference between the normalised feature vectors of two regions in different images is used to label those regions as corresponding (see Fig. 5). In accordance with how the image sequence is generated, it is plausible to consider that corresponding segmented regions in two consecutive frames have similar shapes, except for correspondences that run out of the limits of the visible area of the casting.

iv. **Introducing artificial points**. The corresponding centres of mass found in the previous step can be used to compute the mapping **H** as in (2). However, in practice we need more correspondences to improve the accuracy of such a computation. We increase the number of matches by interpolating artificial points among the centres of mass via B-splines[5]. The Cox-de Boor's recursive formulation of B-splines can be found in [19]. We use cubic B-splines for knots $t_i \in [0,1]$ with four control points $\{P_{-2}, P_{-1}, P_0, P_1\}$. Its matrix form is given by

---

[4] Digital radioscopic images are generated using a frame grabber, which averages $n$ samples of the scene taken at infinitesimal time intervals in order to reduce noise and improve the signal-to-noise ratio.

[5] B-splines are invariant under affine transformations. In practice, linear splines can also be utilised with enough number of knots.

**Fig. 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 = $11 \times 11$), **c** zero crossing image, **d** gradient image, **e** edge detection after adding high-gradient pixels, and **f** detected flaw using the variance of the crossing line profile
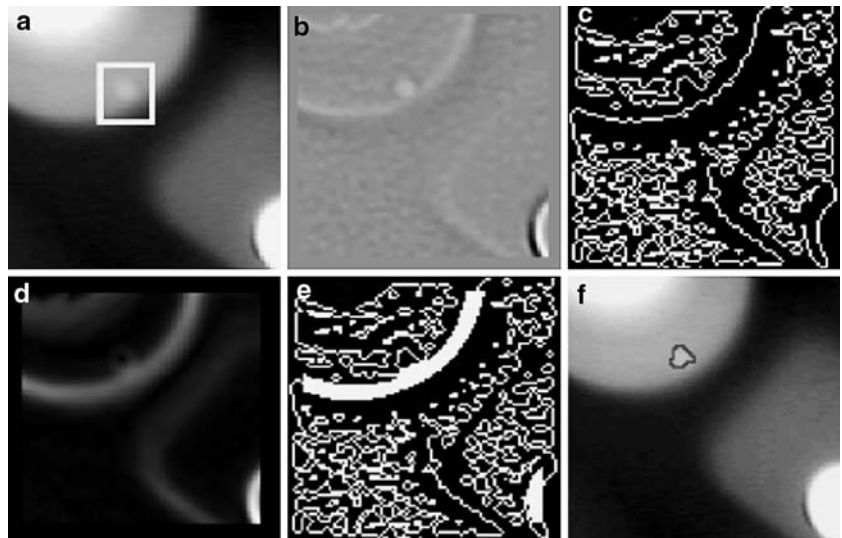


**Fig. 4** *Top* Three views of our real image sequence. *Bottom* Otsu's segmentation method applied on each view
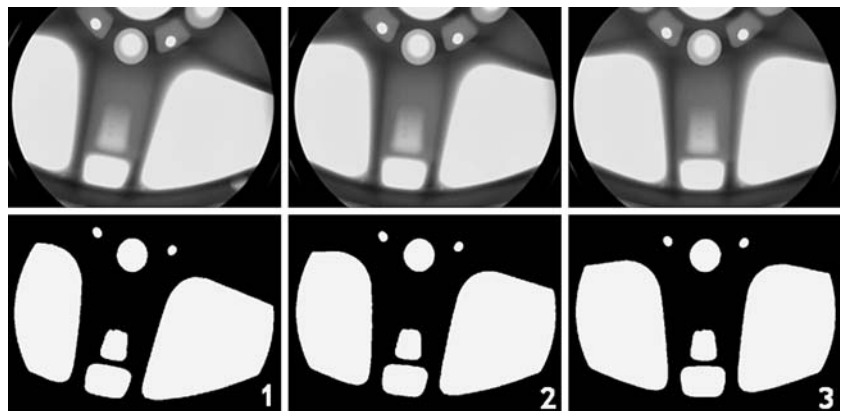


**Fig. 5** Result of matching regions according to their similarity
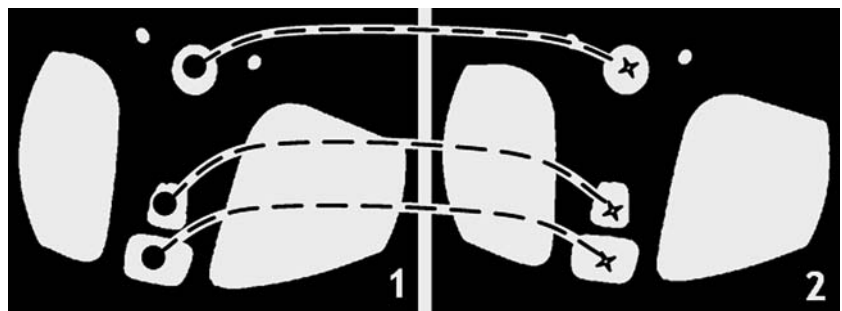


$$B(t_i) = \begin{bmatrix} t_i^3 & t_i^2 & t_i^1 & 1 \end{bmatrix} \frac{1}{6} \begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 0 & 3 & 0 \\ 1 & 4 & 1 & 0 \end{bmatrix} \begin{bmatrix} P_{-2} \\ P_{-1} \\ P_0 \\ P_1 \end{bmatrix}. \tag{3}$$

Varying the number of knots $(t_1,...,t_k)$ among the control points (centres of mass) we regulate the set of artificial points which act as candidate corresponding points (see Fig. 6).

v. **Selection of corresponding points**. According to the *principle of multiple view geometry* [6], all corresponding coordinates between two views are related by the *fundamental matrix*[6] **F**, such that

$$m_i'^T \, \mathbf{F} \, m_i = 0. \tag{4}$$

This relation is known as *epipolar constraint for uncalibrated images* and indicates that the point $m_i'$ can

---

[6] Do not confuse the fundamental matrix **F** with the affine mapping **H**.
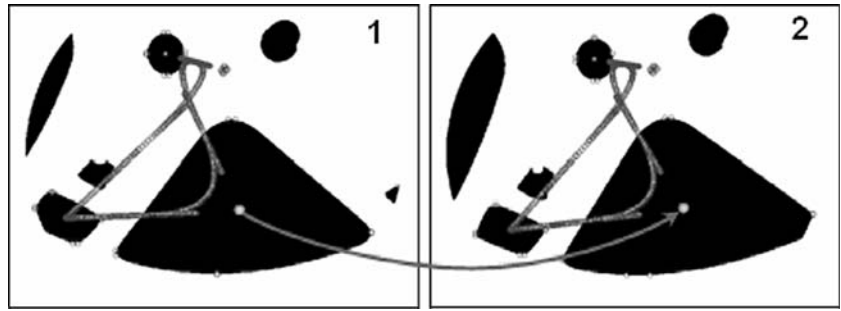
only lie on the epipolar line $l' = \mathbf{F}\, m_i$ of the point $m_i$. Then, from the $k$ candidate points created via B-splines we choose the set of $n$ ($n < k$) correspondences that allow the most accurate computation of the fundamental matrix. This is done by the well-known RANSAC [6] algorithm, which is robust against noise perturbations of the data. This algorithm requires three parameters: the number $N$ of samples/iterations, the threshold $t$ that measures the maximum distance at which a pair of correspondences satisfy (4), and the number $n$ of expected correspondences. We use the Sampson distance [20] and set $t = 2$ pixels. $N$ can be computed as

$$N = \frac{log(1 - p)}{log(1 - (1 - \epsilon)^s)}, \tag{5}$$

using a probability $p = 0.99$ to ensure that at least one sample of $s$ points is free from outliers, and $s = 7$ points necessary to compute the matrix $\mathbf{F}$; and the pessimistic case of having a fraction $\varepsilon = 0.5$ of contaminated correspondences in the input data. From the set of $k$ potential matching points generated by B-splines, we expect to find $n = (1-\varepsilon)k$ pair of correspondences. For $k = 1,000$ knots, $n = 500$ correspondences are expected. See [6] for implementation details. Finally, the selected $n$ points are used to compute the global approximation of the affine mapping $\mathbf{H}$ via equation (2). This approximation of the affine distortion is iteratively refined at every potential defect found in the identification stage (Sect. 2.2), as we will see next.

## 2.4 Robust local defect tracking

Once potential defective regions have been identified in two consecutive or non-consecutive uncalibrated images, and given a preliminary estimation of the global geometric distortion between them, we attempt to track the intensities of each potential defect from the first view onto the second view. Only real flaws should be tracked, whereas false alarms must be consequently discarded. Here, we face the well-known *visual matching problem*, which has been dealt with in literature by means of two approaches: feature-based matching (e.g. [21, 22]) and intensity-based matching (e.g. [23, 24]). Our inspection system combines both strategies.

Using the notation presented in [25], the goal of our intensity-based matching algorithm is to align a template image $T(\mathbf{x})$ with another image $I(\mathbf{x})$, where $\mathbf{x} = (x,y,1)^T$ is a column vector of homogeneous pixel coordinates. A template $T(\mathbf{x})$ represents a potential defective region in the first view and $I(\mathbf{x})$ is the second view where the template has to match. Classical formulations aim to minimise the *sum-of-squared-differences* (SSD) of the intensities between the template $T$ and image $I$ warped onto the coordinate frame of the template, which is known as the *least-squares* (LS) formulation

$$\sum_{\mathbf{x}} [I(\mathbf{W}(\mathbf{x}; \mathbf{p})) - T(\mathbf{x})]^2, \tag{6}$$

where the sum is performed over all pixels in the template image, and $\mathbf{W}(\mathbf{x};\mathbf{p})$ is the warping map obtained by applying the affine transformation $\mathbf{H}$ to the template coordinates, i.e.

$$\mathbf{W}(\mathbf{x}; \mathbf{p}) := \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 + p_1 & p3 & p5 \\ p_2 & 1 + p_4 & p6 \\ 0 & 0 & 1 \end{pmatrix}}_{\mathbf{H}} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}. \tag{7}$$

The affine mapping $\mathbf{H}$ is parameterised by an unknown vector $\mathbf{p} = (p_1,...,p_6)^T$. In literature there are several methods for minimising (6). In particular, the Lucas–Kanade algorithm [26] assumes that a current estimation of $\mathbf{p}$ is known and then it solves iteratively for additive increments $\Delta\mathbf{p}$:

$$\sum_{\mathbf{x}} [I(\mathbf{W}(\mathbf{x}; \mathbf{p} + \Delta\mathbf{p})) - T(\mathbf{x})]^2, \tag{8}$$

updating the parameter vector as $\mathbf{p} \leftarrow \mathbf{p} + \Delta\mathbf{p}$. In general, the Eq. (8) is not robust in presence of *outliers* like

occlusions, illumination changes and non-gaussian noise, because its quadratic error measure assigns a high influence to gross errors, i.e. large deviations cause undesirable distortions in the resulting matching process. In order to downweigh the effect of outliers in the minimisation process, we derive a *robust* formulation of the matching problem. We seek for the M-estimator of $\Delta \mathbf{p}$ as the minimum of the global energy function

$$\Delta \hat{\mathbf{p}} = \arg \min_{\Delta \mathbf{p}} \mathrm{E}(\Delta \mathbf{p}), \qquad (9)$$

where the energy function $\mathrm{E}(\Delta \mathbf{p})$ is defined in terms of a symmetric, positive-definite robust loss function[7] $\rho$, which has an unique minimum at zero, and it is chosen to be less increasing than square [28], i.e.

$$\mathrm{E}(\Delta \mathbf{p}) = \sum_{\mathbf{x}} \rho(z_{\mathbf{x}}), \qquad (10)$$

where $z_{\mathbf{x}}$ is the normalised residue given by

$$z_{\mathbf{x}} = \frac{r_{\mathbf{x}} - \mathrm{Median}(\mathbf{r})}{\hat{\sigma}}. \qquad (11)$$

$\hat{\sigma}$ is the robust standard deviation of the residual vector $\mathbf{r} = \mathbf{I}(\mathbf{W}(\mathbf{x}; \mathbf{p} + \Delta \mathbf{p})) - \mathbf{T}(\mathbf{x})$, and it is computed through the *median absolute deviation* (MAD) [29] as

$$\hat{\sigma} = \zeta \, \mathrm{Median}(|\mathbf{r} - \mathrm{Median}(\mathbf{r})|). \qquad (12)$$

The factor $\zeta = 1/\phi^{-1}(0.75) = 1.4826$ (where $\phi$ is the cumulative distribution function of the standard normal distribution) is introduced in the Eq. (12) to obtain a consistent estimator of $\sigma$, which reaches the same efficiency as the least-squares estimator when only Gaussian noise exists. Moreover, it has been statistically proven that the median is more robust against outliers than the mean as estimator of the central tendency [29].

To solve the robust estimation problem we use the *iteratively reweighted least squares* (IRLS) algorithm proposed in [30]. Performing a first-order Taylor expansion, the residual vector is linearised as

$$\mathbf{r} = \mathbf{I}(\mathbf{W}(\mathbf{x}; \mathbf{p})) + \frac{\partial \mathbf{I}}{\partial \mathbf{p}} \Delta \mathbf{p} - \mathbf{T}(\mathbf{x}), \qquad (13)$$

and setting to zero the partial derivative of the expression (10) with respect to delta_p, we obtain

$$\sum_{\mathbf{x}} w(z_{\mathbf{x}}) \left[ \frac{\partial I}{\partial \mathbf{p}} \right]^T \left[ I(\mathbf{W}(\mathbf{x}; \mathbf{p})) + \frac{\partial I}{\partial \mathbf{p}} \Delta \mathbf{p} - T(\mathbf{x}) \right] = \mathbf{0}, \qquad (14)$$

where $\psi(u) = \frac{\partial \rho(u)}{\partial u}$ and $w(u) = \frac{\psi(u)}{u}$ are the first partial derivative and the weight of the robust loss function $\rho(u)$, respectively. Finally, the solution of the Eq. (9) is given by

$$\Delta \hat{\mathbf{p}} = -\mathbf{H}^{-1} \sum_{\mathbf{x}} w(z_{\mathbf{x}}) \left[ \frac{\partial I}{\partial \mathbf{p}} \right]^T [I(\mathbf{W}(\mathbf{x}; \mathbf{p})) - T(\mathbf{x})], \qquad (15)$$

where the Jacobian and the Hessian are respectively defined as

$$\frac{\partial I}{\partial \mathbf{p}} = \frac{\partial I}{\partial \mathbf{W}} \frac{\partial \mathbf{W}}{\partial \mathbf{p}} = \nabla I \frac{\partial \mathbf{W}}{\partial \mathbf{p}},$$

$$\mathbf{H} = \sum_{\mathbf{x}} w(z_{\mathbf{x}}) \left[ \frac{\partial I}{\partial \mathbf{p}} \right]^T \left[ \frac{\partial I}{\partial \mathbf{p}} \right]^T.$$

As commented before, we consider the geometric distortions induced by the uncalibrated imaging system over the image domain as non-uniform. Therefore, it is necessary to apply the intensity-matching algorithm on each potential defect in order to estimate more accurately the local deformation at that location. In addition, the set of features extracted from each potential defect during the identification stage are now used to distinguish between true and false flaws. The following section describes such a procedure.

## 2.5 Correspondence analysis

Once individual projections for each hypothetical flaw have been found in both views by applying the local matching algorithm, a correspondence analysis is carried out to determine which of them are real and which are false alarms. A region will be classified as defective if the following three criteria are fulfilled:

i.  **Identifiability**. The detection of existing defects must be ensured in the stage of identification of potential defects (Sect. 2.2). If we do not segment the defects at that step, we cannot detect them later on[8]. Thus, to be considered as a flaw, a potential defect must be detected on both views.

ii. **Spatial proximity**. A discontinuity in the first image must be projected to a position in the second image near a hypothetical defect with similar characteristics. To be considered in the vicinity of a flaw in the second image, the projected centre of mass of the defect can be at the most 5 pixels apart—on each coordinate—from its candidate correspondence.

---

[7] Alternatives to choose $\rho$, for instance, are: Cauchy, Huber, Tukey, Geman-McClure and Lorentzian robust functions [27]. In our experiments we use the Geman-McClure one.

[8] Inspection approaches which make use of only one view are also affected by this problem.

iii. **Feature proximity**. To be considered similar to a flaw in the second image, at least four out of six shape characteristics of the projected region might differ by at most 30%, which is measured by taking the norm of the difference between the two normalised vectors of features. The following characteristics are taken into account: area of the segmented defect, average grey value, second derivative, and three different values of contrast.

To overcome the identifiability problem, correspondences in more views can be investigated. For instance, even if we identify a defect in the frames 1 and 3 (but not in frame 2), we can track it if we check the correspondences between views 1 and 3. This strategy was implemented by Mery and Filbert in [1] under the calibrated approach. They were able to track correspondences between a frame $i$ and the following frames $i + 1$, $i + 2$ and $i + 3$. However, the trade-off between the computational time demanded by these calculation and the performance requirements imposed by a particular application, must be carefully analysed.

## 3 Experimental results

In this section we apply our approach for automated defect detection in a sequence of 72 uncalibrated radioscopic images of an aluminum wheel. A segment of six views was shown in Fig. 2. The dimensions of the wheel are 470 [mm] diameter and 200 [mm] height. The image size is $572 \times 768$ pixels with a dynamic range of 8 bits. The wheel has 12 known flaws. Three of these defects are existing blow holes with diameter $\varnothing = 2.0 - 7.5$ [mm] (see Fig. 7). They were initially detected by a visual human inspection. The remaining nine flaws were produced by drilling small holes ($\varnothing = 2.0 - 4.0$ [mm]) in positions of the casting which were known to be difficult to detect (see Fig. 3). A pattern of 1 [mm] in the middle of the wheel is projected as a pattern of three pixels in the image, i.e. the defects are actually very small. In addition, because the signal-to-noise ratio in our radioscopic images is low, the flaws signal is slightly greater than the background noise, as illustrated in Fig. 7. In our experiments, the mean grey level of the flaw signal (without background) ranges from 2.4 to 28.8 grey values with a standard deviation of 6.1. Analysing a homogeneous background in different areas of interest we obtain a noise signal within ±13 grey values with a standard deviation of 2.5. For this reason, the segmentation of real defects with poor contrast can as well involve the detection of false alarms.

The results of the segmentation stage are summarised in Table 1 and partially shown in Fig. 8. One observes that there are 7.74 false alarms per image. Nevertheless, the detection performance in this experiment is still good, because it is possible to identify 86% of all projected flaws along the sequence, whereas 14% of the existing 238 flaws are not identified because of their poor contrast with the background or because they are located at edges of regular structures.

The performance of both least-squares (6) and robust matching (10) algorithms are detailed in Table 2. Notice that the least-squares method might be equivalently obtained by setting $\rho(z) = z^2$ in the robust formulation (10). Each column subdivides the set of detected potential flaws into separate categories according to their actual condition and the classification given by each algorithm. The first four rows show the false negatives, i.e. real defects that could not be matched in the second frame. The first row enumerates the real defects that were not detected in the segmentation stage; the second row reflects the real defects that were impossible to register because they ran out of the image, which made looking for correspondences unfeasible. In both cases the errors do not count as bad performance of either the algorithms or the correspondence analysis, but they are considered as segmentation problems. The third and fourth rows show these defects in which the matching method either diverges or converges on a wrong location, respectively. In both cases the errors count as limitations of the matching algorithm. Finally, the fifth, sixth and seventh rows enumerate the detected defects, false alarms remaining after the matching, and the false alarms of the segmentation eliminated through the matching process, respectively. The detection performance of the matching algorithms is computed as the ratio number of detected defects to the number of detectable defects, excluding those not identified by the segmentation. The robust algorithm detects almost 87% of the flaws with only 0.76 false alarms per image. The computation time required to process one image pair was in average 24.3 [s] on a Pentium 4, 2.8 GHz desktop computer.

We have used the same image sequence as in [1], where flaws were tracked over previously calibrated images. Since our approach implements tracking over uncalibrated images, the results might not be fairly comparable. However, Table 3 outlines the comparative performance of both approaches. The methodology based on calibrated images detects 100% of the defects when using three views, whereas 83% when using five views. This is because it is more probable to find flaws correctly segmented in three views than in five views. Nevertheless, using more views helps to reduce the percentage of false alarms (relative to the number of potential defects). On the other hand, our robust approach for uncalibrated images makes use of only two views and achieves an acceptable compromise between

**Fig. 7** Radioscopic image of a casting with grey-level profiles of three defects
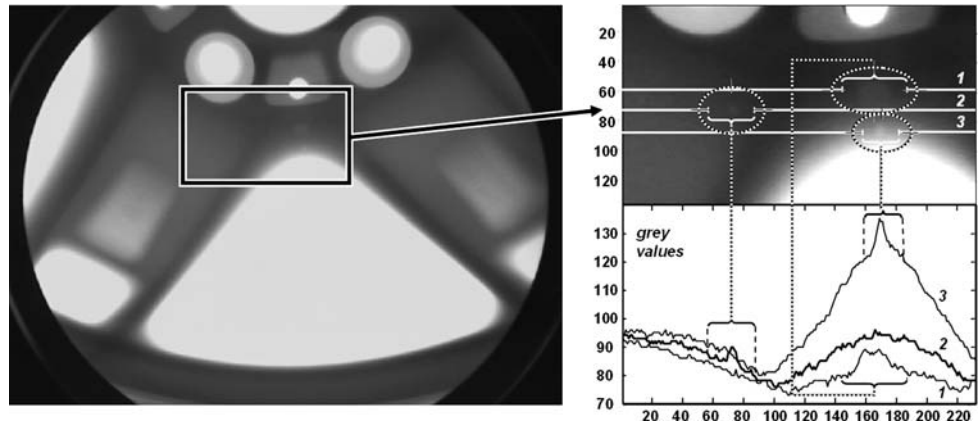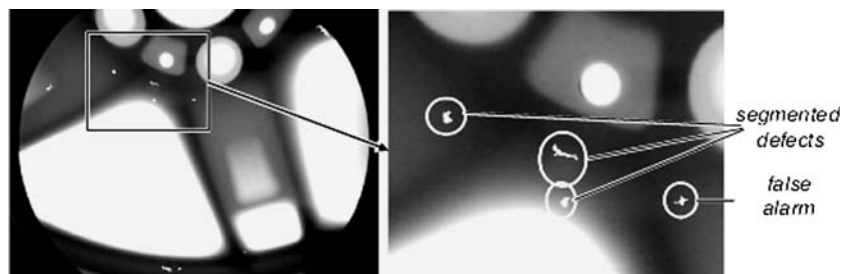
**Table 1** Performance of the identification step

| Existing defects ($D$) | 238 |
|---|---|
| Segmented potential defects | 761 |
| Detected defects ($TP$) | 204 |
| False alarms ($FP$) | 557 |
| Detection performance ($TP/D$) | 85.71% |
| False alarms per image | 7.74 |

detected defects and false alarms. Ways to improve these results are discussed further in the concluding section.

As an extended step of our method, postprocessing of detected flaws could be introduced. For example, defective objects can be automatically taken out from the manufacturing line if the quality control system requires the production of flawless pieces only; or, before removing them, a human inspector could verify the identified flaws by means of a computer-assisted tool. Moreover, in automated inspection of castings we should identify flaws with diameter greater than 2 [mm], which were imaged as regions of approximately 12 pixels in our experiments early on. This allowed us to segment them correctly together with many other false alarms that should be discarded. Our application on aluminum wheels requires that every flaw be detected, i.e. no defects of certain size should remain at any particular location of the casting. However, for other inspection tasks this requirement might be relaxed.

## 4 Conclusions and future work

The multiple view strategy is opening up new possibilities for non-destructive testing by taking into account correspondences between different views of a test object. In this paper, we present the *Robust AMVI* strategy for tracking potential defects on uncalibrated image sequences. Modelling the geometric distortion between each pair of consecutive or non-consecutive views as an unknown affine mapping, this framework introduces two complementary robust procedures to accurately estimate such a transformation. Firstly, a global approximation of the mapping is computed through a set of selected corresponding points of the inspected object. Secondly, the intensities of each potential defect in the first view are iteratively matched onto the second view. As a result, only real defects are successfully tracked and false alarms are discarded. The practical importance of our method lies in avoiding the calibration process. The defect detection is carried out directly on the distorted views produced by an uncalibrated imaging system. This might help in manufacturing processes or in real-time applications that cannot be halted for calibration purposes, or it entails a difficult, unstable, and time-consuming process.

In our experimental results on aluminum die castings we have shown that flaw detection in uncalibrated images is promising. Our framework recognises 86.7% of all existing defects with only 0.76 false alarms per image. The utilised



**Fig. 8** Segmentation of flaws. The existing defects were successfully detected; however, there are also false alarms

**Table 2** Performance of the matching algorithms

| Method | Least-squares | Robust |
| --- | --- | --- |
| No segmentation | 13 | 13 |
| Occlusion | 10 | 10 |
| Divergence | 14 | 2 |
| Wrong convergence | 22 | 22 |
| Detected defects | 145 | 157 |
| False alarms | 52 | 55 |
| Eliminated | 505 | 502 |
| Total | 761 | 761 |
| Detection performance | 80.1% | 86.7% |
| False alarms per image | 0.72 | 0.76 |

**Table 3** (C)alibrated versus (U)ncalibrated approaches

| Method | Detected defects (%) | False alarms (%) |
| --- | --- | --- |
| C—3 views | 100 | 25 |
| C—5 views | 83 | 0 |
| U—least-squares | 80.1 | 26.4 |
| U—robust | 86.7 | 25.9 |

image sequence is truly representative of those employed in labs to test algorithms for detecting potential flaws. Each image along the sequence contained 12 physical defects synthetically placed in a way that their detection is not evident due to their form, miniature size, location, intensity, and deepness; some of them are almost imperceptible. The proposed methodology is straightly generalisable to any manufacturing system of regular structures. Indeed, this framework is not limited to X-rays, and can be employed in uncalibrated sequences acquired from other imaging systems.

Our tracking scheme is based merely on two views because the quote of false alarms is low, but it can be run for three or more views. In the future we plan to extend this framework from image-pairs to image-triplets by means of trifocal tensors. Instead of using centres of mass to generate artificial points, we will consider structural edges of the objects to avoid lack of closed regions at the outer limits of the visible casting area, where most of the mismatches took place.

To improve the computational performance of our approach it would be worth considering more efficient algorithms that solve the visual matching problem. In particular, the *inverse compositional* approach proposed in [31] is an interesting alternative that pre-computes the Jacobian and Hessian matrixes, which are updated at each iteration in our implementation. On the other hand, it is also valuable to look into coarse-to-fine strategies like

*multigrid methods*. These allow the implementation of highly efficient real-time applications. See for example [32], where speedups of several hundreds are reached in estimating real-time motion.

## 5 Originality and contribution

In the last 15 years, imaging systems have revolutionised many industrial processes. These systems make use of cameras to obtain discretised visual representations of the process being inspected, then specialised algorithms perform the necessary computations and the corresponding data analysis. Such systems make it possible to correct deviations from the expected behaviour of the process automatically or in a computer-assisted human way. In order to obtain accurate results, the camera parameters must be precisely calibrated. Nevertheless, calibration of an imaging system is often a difficult, an unstable and time-consuming process. Moreover, in certain environments the systems must be re-calibrated periodically. To avoid the calibration process, in this paper we propose the *Robust AMVI* strategy, a methodology to work with uncalibrated images on industrial visual inspection problems. Applying two complementary robust procedures, we attempt to track identified potential defects of an inspected object along an image sequence. As a result, only existing defects can be successfully tracked while false alarms are discarded. We show experiments on aluminum die casting in the context of X-ray imaging, although this framework is straightly generalisable to any manufacturing system of regular structures. Since no calibration is required, it is indeed possible to implement an automated multiple view inspection for industrial environments where calibration is not available or it cannot be afforded.

## References

1. Mery D, Filbert D (2002) Automated flaw detection in aluminum castings based on the tracking of potential defects in a radioscopic image sequence. IEEE Trans Rob Autom 18(6):890–901
2. Mery D, Carrasco M (2005) Automated multiple view inspection based on uncalibrated image sequences. Lect Notes Comput Sci 3540:1238–1247
3. Abramovich G, Barhak J, Spicer P (2005) Reconfigurable array for machine vision inspection (RAMVI). In: Proceedings of the 3rd international conference on reconfigurable manufacturing, Ann Arbor
4. Spicer P, Bohl K, Abramovich G, Barhak J (2006) Robust calibration of a reconfigurable camera array for machine vision inspection (RAMVI) using rule-based colour recognition. In: Proceedings of the 1st international conference on computer vision theory and applications, Setúbal, pp 131–138

5. Gumustekin S (2004) A visual inspection system using a single camera and mirrors. In: Proceedings of the 12th IEEE signal processing and communications applications conference, Kusadasi, pp 257–260

6. Hartley RI, Zisserman A (2002) Multiple view geometry in computer vision. Cambridge University Press, Cambridge

7. Boerner H, Strecker H (1988) Automated X-ray inspection of aluminum casting. IEEE Trans Pattern Anal Mach Intell 10(1):79–91

8. Filbert D, Klatte R, Heinrich W, Purschke M (1987) Computer aided inspection of castings. In: IEEE-IAS Annual Meeting, Atlanta, USA, pp 1087–1095

9. Heinrich W (1988) Automatische Röntgenserienprüfung von Gußteilen. PhD Thesis, Institut für Allgemeine Elektrotechnik, Technische Universität Berlin

10. Mohr G, Fock T (2004) X-ray inspection in the aerospace industry—state of the art, challenges and emerging techonologies. In: Proceedings of the 16th world conference on nondestructive testing, Montreal

11. Mahi Uddin Khan Md (1998) Non-destructive testing applications in commercial aircraft maintenance. In: Proceedings of the 7th European conference on non-destructive testing, Copenhagen

12. Kita Y, Highnam R, Brady M (2001) Correspondence between different view breast X-rays using curved epipolar lines. Comput Vis Image Underst 83(1):38–56

13. Rebuffel V, Pires S, Caplier A, Lamarque P (2003) Automatic delamination defects detection in radiographic sequences of rocket boosters. In: Proceedings of the international symposium on computed tomography and image processing for industrial radiology, Berlin

14. Jähne B, Haußecker H, Geißler P (1999) Handbook of computer vision and applications, vol 2. Signal processing and pattern recognition. Academic, San Diego

15. Mery D (2003) Crossing line profile: a new approach to detecting defects in aluminium castings. Lect Notes Comput Sci 2749:725–732

16. Mery D, da Silva R, Caloba LP, Rebello JMA (2003) Pattern recognition in the automatic inspection of aluminium castings. Insight 45(7):475–483

17. Haralick RM, Shapiro LG (1992) Computer and robot vision. Addison-Wesley, New York

18. Sonka M, Hlavac V, Boyle R (1998) Image processing, analysis, and machine vision, 2nd edn. PWS Publishing, Pacific Grove

19. de Boor C (1978) A practical guide to splines. Springer, New York

20. Sampson PD (1982) Fitting conic sections to 'very scattered' data: an iterative refinement of the bookstein algorithm. Comput Vis Graph Image Process 18:97–108

21. Wells WM III (1997) Statistical approaches to feature-based object recognition. Int J Comput Vis 21(1–2):63–98

22. Hsieh JW, Liao HYM, Fan KC, Ko MT, Hung YP (1997) Image registration using a new edge-based approach. Comput Vis Image Underst 67:112–130

23. Black MJ, Jepson AD (1998) Eigen-tracking: robust matching and tracking of articulated objects using a view-based representation. Int J Comput Vis 26(1):63–84

24. Hager GD, Belhumeur PN (1998) Efficient region tracking with parametric models of geometry and illumination. IEEE Trans Pattern Anal Mach Intell 20(10):1025–1039

25. Baker S, Matthews I (2004) Lukas-Kanade 20 years on: a unifying framework. Int J Comput Vis 56(3):221–255

26. Lucas BD, Kanade T (1981) An iterative image registration technique with an application to stereo vision. In: Proceedings of the 7th international joint conference on artificial intelligence, Vancouver, pp 674–679

27. Black MJ, Anandan P (1996) The robust estimation of multiple motions: parametric and piecewise-smooth flow fields. Comput Vis Image Underst 63(1):75–104

28. Stewart CV (1999) Robust parameter estimation in computer vision. SIAM Rev 41(3):513–537

29. Rousseeuw PJ, Leroy AM (1987) Robust regression and outlier detection. Wiley, New York

30. Holland PW, Welsch RE (1977) Robust regression using iteratively reweighted least-squares. SIAM Rev 6:813–827

31. Baker S, Matthews I (2001) Equivalence and efficiency of image alignment algorithms. In: Proceedings of the 8th IEEE conference on computer vision and pattern recognition, Vancouver, vol 1, pp 1090–1097

32. Bruhn A, Weickert J, Kohlberger T, Schnörr C (2006) A multigrid platform for real-time motion computation with discontinuity-preserving variational methods. Int J Comput Vis 70(3):257–277

**Author Biographies**

**Luis Pizarro** was born in Valparaíso, Chile in 1977. He received the B.E. and M.Sc. degrees in Informatics Engineering from Universidad Técnica Federico Santa María, Chile, in 2003. From 2001 to 2003, he worked as a research assistant at the same university and was a consulting engineer at Anglo American Chile S.A. From June 2004 to June 2005, he worked as a researcher at the Mining and Metallurgy Innovation Institute S.A., Chile. He is currently pursuing his PhD in Computer Science at Saarland University, Germany, sponsored by the German Academic Exchange Service (DAAD). His research interests include image analysis, computer vision, pattern recognition, robust statistics, and industrial inspection.

**Domingo Mery** was born in Santiago, Chile in 1965. He received the Diploma (M.Sc.) degree in Electrical Engineering from the Technical University of Karlsruhe, Germany, in 1992, and the PhD degree with distinction at the Technical University of Berlin, in 2000. From 1993 to 1996, he was a Research Assistant at the Department of Electrical Engineering at Universidad Católica de Chile. From 1996 to 2000, he was a Research Scientist at the Institute for Measurement and Automation Technology at the Technical University of Berlin with the collaboration of YXLON X-Ray International. From 1989 to 1992, he was a recipient of a Scholarship from the Konrad-Adenauer-Foundation, and from 1996 to 2000 of a Scholarship from the German Academic Exchange Service (DAAD) for his PhD work. From 2001 to 2003 he was Associate Researcher at the Department of Computer Engineering at the University of Santiago, Chile. Since 2004 he is Assistant Professor at the Department of Computer Science at the Pontificia Universidad Católica de Chile, Chile. Since 2005 he is Chair of the Computer Science Department.

His research interests include image processing for fault detection in aluminum castings, X-ray imaging, real-time programming, and computer vision.

**Rafael Delpiano** was born in Santiago, Chile in 1982. He received the B.E. degree in 2006 from Pontificia Universidad Católica de Chile (PUC). As a student, he founded the ACM Student Chapter at PUC. From 2001 to 2006 he worked as a research assistant in the Department of Computer Science at PUC. Currently, he is lecturer on computer programming at the same university and is the head of the IT Department of Providencia Mayorship in Santiago de Chile.

**Miguel A. Carrasco** was born in Santiago, Chile in 1977. He received the B.S. in Informatics Engineering and M.Sc. in Computer Engineering degrees from Universidad de Santiago de Chile in 2004. Due to his work on a robust algorithm for non-destructive testing of weld seams, his master's thesis received the highest honors. He is currently pursuing his PhD in Computer Science at Pontificia Universidad Católica de Chile, sponsored by the National Commission for Scientific and Technological Research (CONICYT). His research interests include image processing, automatic visual inspection, computer vision, and pattern recognition.