Computers and Chemical Engineering 71 (2014) 501–511

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

Statistical process control based on Multivariate Image Analysis: A new proposal for monitoring and defect detection

J.M. Prats-Montalbán*, A. Ferrer

Multivariate Statistical Engineering Group, Department of Applied Statistics, Operations Research and Quality, Universidad Politécnica de Valencia, Cno. De Vera s/n, Edificio 7A, 46022 Valencia, Spain

ARTICLE INFO

Article history:
Received 17 April 2014
Received in revised form 31 July 2014
Accepted 26 September 2014
Available online 6 October 2014

Keywords:
Multivariate Image Analysis (MIA)
ARL
Control charts
RSS image
T^2 image
Wavelets

ABSTRACT

The monitoring, fault detection and visualization of defects are a strategic issue for product quality. This paper presents a novel methodology based on the integration of textural Multivariate Image Analysis (MIA) and multivariate statistical process control (MSPC) for process monitoring. The proposed approach combines MIA and *p*-control charts, as well as T^2 and *RSS* images for defect location and visualization. Simulated images of steel plates are used to illustrate the monitoring performance of it. Both approaches are also applied on real clover images.

1. Introduction

During the last 40 years visual inspection has been introduced in different industrial and technological fields for defect detection and classification. Two issues are critical for these purposes. One is the time required, because in processes usually many parts or samples are to be inspected, and inspection speed is a critical concern due to the high production rates (e.g. in ceramic tiles manufacturing). Another issue is the objectivity and consistency in the classification of the manufactured parts or in the monitoring of a process when establishing if it is in control or not. Visual inspection has been traditionally performed by well-trained operators in order to determine if the manufactured part meets the quality standards, presents any type of defect, fits a specific class, or if it comes from an in-control process. Examples can be found in the steel industry (Cardin et al., 2011); in the mining industry (Liu et al., 2005a,b), or in the ceramic tile industry (López et al., 2008; Prats-Montalbán et al., 2008).

However, this type of visual inspection is not exempt of subjective criteria, no matter the experience of the operator. The reasons for this subjectivity are related to (i) the variability in the classifications made by the same operator, due to e.g. fatigue or loss of attention (repeatability), and (ii) the variability in the classifications made by different operators, related to the different visual perception characteristics of each person (reproducibility). Thus, it is necessary to develop reliable image-based monitoring systems with low repeatability and reproducibility variation.

Nowadays, cameras-based visual systems capture information from the process without interfering with the product due to their non-destructive and non-invasive nature, avoiding the aforementioned problems of human visual inspection. This way, three different objectives are reached: (1) reducing expensive on-line sensors and laboratory sampling analyses; (2) avoiding tests that cannot be applied on the whole production as is the case of destructive tests to determine the stress endurance of a metallic, ceramic or any other kind of part; (3) allowing on-line characterization of products and processes that are not always amenable via traditional on-line instrumentation (Yu and MacGregor, 2003a,b).

Once an image has been acquired from the process, it is possible to apply image analysis techniques to analyze it. This information is used for classification, defect detection or image object characterization. A well-established image analysis technique in process monitoring is Multivariate Image Analysis (MIA) (Geladi and Grahn, 1996; Prats-Montalbán et al., 2011). Different works have treated the monitoring and classification problems based on MIA. An excellent review about the application of MIA in industrial environments for monitoring, prediction and control can be found in Duchesne et al. (2012).

* Corresponding author at: Departamento de Estadística e IO Aplicadas y Calidad, Universidad Politécnica de Valencia, Cno. De Vera s/n, Edificio 7A, 46022 Valencia, Spain. Tel.: +34 96 387 70 07x74949; fax: +34 96 387 74 99.
E-mail address: jopramon@eio.upv.es (J.M. Prats-Montalbán).

http://dx.doi.org/10.1016/j.compchemeng.2014.09.014

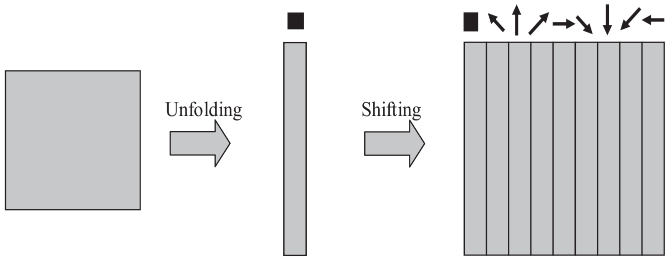

Fig. 1. Data structure used to perform the multivariate image analysis.

In a recent work, Megahed et al. (2011) have performed a review on the use of control charts with image data. Many different methods are presented and commented. Their main drawback is that they do not provide specific spatial control charts able to precisely localize the defect detected by the (whole) image analysis control systems. The paper claims for further study in this topic, in particular "*the study of the statistical properties and performance of many control charting methods*". It also states that there is a lack of study of the "*performance of existing methods as well as understand the differences between competing approaches*."

The present work tries to address these last two goals. In particular the paper is focused on the development of a MIA-based statistical process control (SPC) monitoring scheme able to check the adequacy of the product with regards to the established standards and, furthermore, locate and isolate any kind of defect. We do not address the problems related to illuminating conditions changes. Different ways for dealing with this problem are treated by Reis and Bauer (2009), and very recently by Ottavian et al. (2013, 2014). This paper presents and develops a novel Multivariate Image Analysis technique, and deeply analyzes its potential in comparison with a competing approach, in terms of control charts performance and statistical properties; trying to understand the differences between them.

Steel plate simulated images are used to illustrate the monitoring performance of the proposed approach. Section 2 presents the data structure linked to these images, the different image preprocessing techniques required for on-line implementation and the proposed MIA-based SPC monitoring scheme. Section 3 illustrates the statistical comparison study with a sound competing approach (Liu and MacGregor, 2005, 2006) in terms of average run length (ARL), based on simulated images of steel plates. Section 4 presents the results of both methodologies applied on a real data set (clover images). Finally, Section 5 outlines the conclusions of this paper.

2. Materials and methods

2.1. Image data structure

The first type of information when dealing with digital images is formed by the intensities of the $I = n_1 \times n_2$ pixels arranged in the two dimensions that shape the images. In the case of gray level images, each pixel summarizes the global level of spectral information in it. The second type is the spatial information associated to the variations of intensity levels existing in a local zone (a neighborhood) of the image (i.e. texture). Bharati (2002) and Bharati and MacGregor (2000) provide an approach for easily gathering this type of information.

Bharati and MacGregor's approach consists of unfolding each image into just one column vector having the intensity levels of all the pixels, and registering, for each one of them, the intensities of the neighboring ones. This is illustrated in Fig. 1, where a 3×3 neighborhood window is used. In this study, it is sensible to this window size, since the unspecific defects analyzed can be of different sizes and shapes, and also random textured. However, in those cases were one is more or less aware about the type of defects (nature, size, shape, orientation, etc.) that may appear in any process, a rational way to approach would be to design an experiment with different window sizes and directions (image analysis kernels) being analyzed, afterwards deciding which treatment to use for optimizing some or various types-of-defects detections.

This leads to a highly correlated and complex data structure suitable to be analyzed by multivariate statistical projection methods such as principal component analysis (PCA) (Jackson, 2003). This approach can be applied in a straightforward way to color images by simply applying this procedure to each color band, afterwards stacking each data structure related to each color one beside the other (Prats-Montalbán and Ferrer, 2007).

2.2. MIA-based SPC scheme

The MIA-based SPC monitoring scheme proposed in this work, as any SPC scheme, is carried out in two phases. In Phase I (model building) monitoring charts are built according to a set of historical in-control data, once the performance of the process has been modeled, and the assumptions of its behavior are checked.

In Phase II (model exploitation) these charts are used to monitor the process using on-line images, assuming the form of the distribution of the monitoring statistics to be known along with its values of the in-control parameters (Woodall, 2000). This monitoring scheme constitutes what the authors have named in the MIA field as the *fit to a pattern model approach* (FPM) (Prats-Montalbán and Ferrer, 2007, 2011). In this paper we have improved the FPM approach and named as *Percentage-based FPM*, PFPM. For a general overview of the application of PCA to multivariate SPC see e.g. Ferrer (2007). The basics of this scheme are explained in the following.

2.2.1. Phase I (model building): off-line process monitoring

Phase I of PFPM starts by building a PCA model (Jackson, 2003). PCA projects the original variables onto new ones, called latent variables, orthogonal and arranged according to their eigenvalues. Applying a PCA model to the textural matrix $\mathbf{X}$ $(I \times J)$ can be expressed as:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \tag{1}$$

where $\mathbf{T}$ $(I \times R)$ and $\mathbf{P}$ $(J \times R)$ are the scores and loading matrices for R principal components ($R \leq \text{rank}(\mathbf{X})$), respectively; and $\mathbf{E}$ $(I \times J)$ is the residual matrix of the PCA model.

The number of R components to extract is a non-trivial issue to deal with, and it depends on the problem at hand (Camacho and Ferrer, 2012, 2014). One possible approach is trying to extract as many components as necessary for appropriate modeling of the main spatial features of the images, e.g. by screen plot and further loadings inspection if looking for any deviation from these behaviors; or by looking for those components gathering some specific defects, when a priori known; or to choose that number of components that maximizes detection capacity (or classification accuracy) when defect images are available in addition to NOC images (Prats-Montalbán et al., 2008). Note that the widely used cross-validation approach is not appropriate to address this part of the modeling stage (Camacho and Ferrer, 2012, 2014), since minimizing the squared prediction error in cross validation (objective function in CV) is not necessarily related to maximizing the fault detection power (objective function in process monitoring and fault detection).

The model is built using one or some in-control or process pattern images. These images are collected when the process is operated in-control and assuming independence. Once the loading