Chemometrics and Intelligent Laboratory Systems 108 (2011) 2–12

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Extracting useful information from images

Sergey Kucheryavski

Aalborg University Esbjerg, ACABS research group, Niels Bohrs vej, 8, 6700 Esbjerg, Denmark

ARTICLE INFO

Article history:
Received 5 October 2010
Received in revised form 27 November 2010
Accepted 14 December 2010
Available online 22 December 2010

Keywords:
Image processing
Image analysis
Textures
Fractal analysis
Angle measure technique
Wavelets
Image morphology

ABSTRACT

The paper presents an overview of methods for extracting useful information from digital images. It covers various approaches that utilized different properties of images, like intensity distribution, spatial frequency content and several others. A few case studies including isotropic and heterogeneous, congruent and non-congruent images are used to illustrate how the described methods work and to compare some of them.

1. Introduction

Image analysis is used widely and successfully in many areas of science and industry. It is an essential and non-invasive tool for getting quantitative information about structure of investigated objects as well as for establishing a link between this information and various non-visual properties of the objects.

Any image contains a large amount of data. Just an ordinary color picture taken with a chip digital camera consists of several million pixels with three numeric values for each — color channels. Most of the basic mathematical methods and algorithms for processing and analysis of digital images were developed in 70–80th of the last century when first computers became available. Nowadays the modern methods allow to solve quite sophisticated problems, like recognition of face and emotions, analysis of moving objects and many others. Image analysis also has found its success in industry, for example for analysis of particles, for evaluation of homogeneity of mixings and in many other applications. The non-invasivity, simplicity and high speed of data acquisition procedure as well as relatively low price of the acquisition hardware make this approach quite attractive for analysis of industrial processes.

When someone mentions image analysis and chemometrics at the same time, in most of the cases the matter is MIA – Multivariate Image Analysis – an efficient and popular tool for processing and analysis of multichannel and hyperspectral images, invented by Geladi et al. [1]. Multivariate image analysis treats image pixels as objects and the corresponding color channels (or wavelength in case of multichannel and hyperspectral images) as variables. Being applied to a hyperspectral image it allows to clusterize it, find outliers and extreme pixels, and make classification and regression models for discovering parts with certain properties (for example evaluating quality of beef samples [2], detection disease marks in body tissues [3], estimation of hardness of maize kernels [4], and much more).

However a lot of information can be extracted even from simple grayscale images. Such images can be acquired by a camera with CCD or CMOS sensor, transformed from color or hyperspectral images (like score image, for example) or be a result of a simulation process. It is also possible to process color images by analyzing their color channels separately, so each channel is considered as a grayscale image. In any case all these images are represented by a matrix with intensity values.

This paper describes methods and algorithms for extracting useful information from grayscale images. The analytical procedure for such images is pretty straightforward and includes several steps besides the image acquisition.

The first step is image processing. It usually consists of two substeps — image enhancement and post-processing. Image enhancement aims to improve quality of an acquired image. Usually it includes correction of lightning conditions (brightness and contrast) and geometrical distortions if any as well as noise reduction. Image enhancement methods are very well described in various textbooks on image processing (for example [5,6]) and will be barely touched here.

Post-processing is mainly about how to prepare images for further analysis. The most used post-processing technique is segmentation — selection of image parts that are most important for the analysis. In the simplest case, segmentation just removes background pixels, but

E-mail address: svkucheryavski@gmail.com.

0169-7439/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2010.12.002

there are also cases where implementing of complex segmentation algorithms is necessary (for example detection of cells with predefined properties: color, shape and so on).

The second step is devoted to extracting features from images and this paper mainly focuses on this one. This step implies estimation of various quantitative properties of an image, like, for example, spatial frequencies content, roughness, geometrical properties of depicted objects and many others. The extracted features can be used as a final result, like, for example, distribution of particle size, or treated as an input for further analysis, including regression and classification.

This paper reviews this common approach for extracting relevant information from images. Several feature extraction methods are discussed. Most of these methods can be used widely for different purposes — since they are quite universal. Most of them also extract multivariate information from images and can be naturally integrated with chemometrics methods for data analysis. Each method is described briefly using one or two study cases and the description is supplemented with information and references about where and how the particular method has been used. It is important to underline that the paper does not claim to be exhaustive — the amount of different techniques for image features extracting is way larger. However the author hopes that the selected methods will give a decent illustration of main principles for extracting information from images.

2. Case studies

To show how the chosen methods work and to compare some of them, four case studies will be used, with both isotropic and non-isotropic images.

2.1. Segmentation of biochips

Biochip is a rectangular substrate with a number of chemical microreactors — typically small spots of gel with different reagents inside. If one deposits a solution with another reagent (conjugated with fluorescent label) over a chip surface, reactions will happen in some cells and further, having a proper post-processing, it is possible to see these cells as light spots on the chip. Biochips allow to analyze reactions among hundreds of reagents at the same time and are used widely, for example, in medicine for discovering different mutations of a disease.

For automatic analysis a biochip image has to be segmented to individual cells. That could be done easily since the cells are ordered on a chip. However due to different reasons, a segmented cell can be shifted on a resulted image. In Fig. 1 a result of such segmentation with some correctly and incorrectly segmented cells is shown. The main problem in this case study is to find how the segmentation errors can be corrected.

2.2. Recognition of plastic particle concentrations

This case study is mostly devoted to illustrate how to extract and use textural features. Texture recognition and analysis is a large separate part of image analysis. Textures have no exact definitions but can be considered as images that contain some stochastically or regularly repeated patterns. Most of stochastic textures are isotropic and invariant to rotations.

In this case study 32 images of dark and light plastic particles, mixed in different proportions: 10:1 (10 volumes of dark and 1 volume of light particles), 10:2, 10:3 and 10:4 were prepared as an example of stochastic textures. The particles had the same size and shape. For each of the four cases eight pictures were taken. The mixtures were shaken for about 1 min before each shot. Finally the images were cropped and downsampled to the resolution of 512×512 pixels.

Examples of the images from each set are shown in Fig. 2. The main task for this case is to discriminate images with different concentrations of white pellets. It can be noticed that the last two sets give almost similar images, so some problems can be expected.

2.3. Morphological analysis of pharmaceutical pellets

Analysis of morphological properties of particles is also a quite common problem. It can be used, for instance, in microscopy for recognition and analysis of different cells and bacteria, for monitoring processes where particles are involved (like controlling of size and shape of pellets) and in many other areas.

In this case study an image with about 1000 small pellets of saccharine with coated layer of paracetamol will be used to show main steps of morphological analysis. The pellets have physical size from 30 to 40 μm. The original image has a resolution of 2048×2048 pixels. The part of the image is shown in Fig. 3.

2.4. Classification of brain tomograms

In this case study a set of images corresponds to one sample. The images are scans of magnetic resonance imaging (MRI) tomograms of human brains with early stage of Alzheimer disease (AD) and without it. So features must be calculated for all images from a particular set and combined to a vector.

The MRI images were taken from OASIS public database (http://www.oasis-brains.org). This dataset is a part of project “Open Access

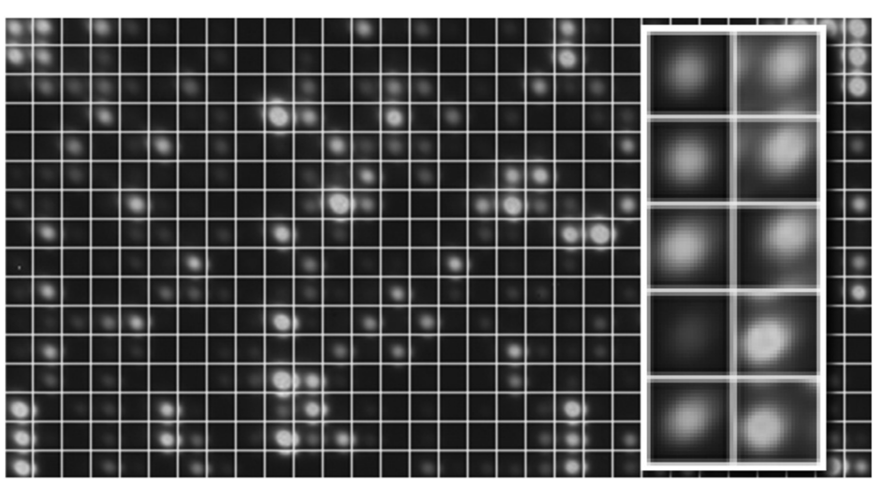

Fig. 1. Digitized image of segmented biochip and examples of correctly and incorrectly segmented cells.