Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Ultramicroscopy 103 (2005) 141–152
ultramicroscopy
www.elsevier.com/locate/ultramic

Advances in the segmentation of multi-component microanalytical images

Jérôme Cutrona[a,b], Noël Bonnet[a,b,*], Michel Herbin[b], Ferdinand Hofer[c]

[a]UMRS INSERM 514 (IFR53), Hôpital Maison Blanche, 45, rue Cognacq Jay, 51092 REIMS Cedex, France
[b]Laboratoire d'Etudes et de Recherches en Informatique (CReSTIC-LERI), University of Reims, France
[c]Research Institute for Electron Microscopy, Graz University of Technology, Steyrergasse 17, A-8010 Graz, Austria

Received 24 June 2004; received in revised form 10 November 2004; accepted 22 November 2004

Abstract

Segmenting multi-component microanalytical images consists in trying to find zones of the specimen with approximate homogeneous composition, representing different chemical phases. This can be done through pixel clustering. We first highlight some limitations of classical clustering algorithms (*C*-means and fuzzy *C*-means). Then, we describe a new algorithm we have contributed to develop: the Parzen-watersheds algorithm. This algorithm is based on the estimation of the probability density function of the whole data set in the feature space (through the Parzen approach) and its partitioning using a method inherited from mathematical morphology: the watersheds method. Next, we introduce a fuzzy version of this approach, where the pixels are characterized by their grades of membership to the different classes. Finally, we show how the definition of the grades of membership can be used to improve the results of clustering, through probabilistic relaxation in the image space. The different methods presented are illustrated through an example in the field of electron energy loss mapping, where four elemental maps are concentrated in a single chemical phase map.

PACS: 02.70.−c; 02.50.Sk; 68.55.Nq; 42.30.Va; 07.05.P

Keywords: Multi-component images; Segmentation; Clustering; Relaxation; Probability density function; Fuzzy logic

1. Introduction

The improvement of microanalytical techniques allows us to record, simultaneously or not, the maps of several elements constituting the specimen under study. The observation of these different maps permits us to interpret the content of the

*Corresponding author. Address. UMRS INSERM 514 (IFR53), Hôpital Maison Blanche, 45, rue Cognacq Jay, 51092 REIMS Cedex, France. Tel.: 333 26787771; fax: 333 26065861.
E-mail address: noel.bonnet@univ-reims.fr (N. Bonnet).

0304-3991/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.ultramic.2004.11.005

studied area of the specimen qualitatively. But quantitative information remains difficult to infer. One step towards a more quantitative interpretation will consist in concentrating the information contained in the different elemental maps into a single labelled map, where each label is associated with a class of pixels.[1] Each class of pixels is characterized by a set of concentrations, one for each of the studied elements.

There are several ways to go from the set of elemental maps to the unique class map. A first group of methods is interactive and sometimes called *interactive correlation partitioning*. It consists in selecting the different classes of pixels interactively, within the two- or three-dimensional scatterplot, and back-mapping the selected areas into the real image space.

The second group of methods, on which we will concentrate here, is automatic and we called it *automatic correlation partitioning*. It consists in automatically grouping similar pixels (that possess similar elemental composition) into one class, the number of classes being a priori unknown. This process is known as clustering in artificial intelligence and data analysis communities [1].

To the best of our knowledge, the first attempt to introduce the idea of clustering in the framework of imaging in microanalysis is in Ref. [2]. In this work, an algorithm well-known in data processing was introduced: the *K*-means algorithm. A little later, Bonnet et al. [3] expanded on this idea and presented more detailed applications of the *K*-means and the fuzzy *C*-means algorithms, a fuzzy version of the *K*-means algorithm.[2]

In parallel to their application to microanalytical data sets, clustering techniques are also evolving independently and we have contributed to the development of a new technique that overcomes some drawbacks of the previous ones. In this paper, we illustrate the capabilities of the new technique and its fuzzy version in the framework of multi-component microanalytical imaging.

The outline of the paper is the following. In the next section, we summarize the *C*-means and fuzzy *C*-means algorithms. We also emphasize their weaknesses. In the following section, we present the new algorithm we have developed (and called the Parzen-watersheds algorithm) and its fuzzy counterpart. We show how these algorithms surpass the classical algorithms. Then, we illustrate the application of the different algorithms to a set of four elemental maps recorded through electron energy-loss mapping. Finally, we draw some conclusions.

2. Classical clustering algorithms: *C*-means and fuzzy *C*-means

We start with a set of N elemental maps EM, each one of size $K = width \times height$:

$$EM_n(k), \quad n \in \{1,\ldots,N\}, \quad k \in \{1,\ldots,K\}. \tag{1}$$

Each pixel k of these maps can thus be described by a vector, $V_k \in \mathbb{Z}^N$, where $\mathbb{Z}^N$ is the N-dimensional space of positive integers:

$$V_k = [EM_1(k),\ldots,EM_n(k),\ldots,EM_N(k)]^{\mathrm{T}}, \quad k \in \{1,\ldots,K\}. \tag{2}$$

The purpose of the clustering process is to group pixels into a limited set of classes: $c \in \{1,\ldots,C\}$. This must be done on the basis of the data set alone, without the help of any additional training set.[3]

The classical *C*-means algorithm aims at grouping objects (here, objects are pixels) in order to minimize the objective function:

$$J_{C\text{-means}} = \sum_{c=1}^{C} \sum_{k=1}^{K} d^2(V_k, X_c), \tag{3}$$

where X_c is the centre of class c and d is the distance between object k and X_c.

[1] Or voxels for three-dimensional microanalytical techniques.

[2] The K in "K-means" and the C in "fuzzy C-means" have exactly the same meaning, i.e. the number of classes. So, in the remaining of the paper, we will use "C-means" and "fuzzy C-means", in order to avoid confusion.

[3] Other classification algorithms are based on the use of a training set. We do not consider them in this paper.