A review of thresholding strategies applied to human chromosome segmentation

Enea Poletti, Francesca Zappelli, Alfredo Ruggeri, Enrico Grisan*

Department of Information Engineering, University of Padova, Via Gradenigo 6/a, 35131 Padova, Italy

ARTICLE INFO

ABSTRACT

Karyotype analysis is a widespread procedure in cytogenetics to assess the presence of genetic defects by the visualization of the structure of chromosomes. The procedure is lengthy and repetitive and an effective automatic analysis would greatly help the cytogeneticist routine work. Still, automatic segmentation and the full disentangling of chromosomes are open issues. The first step in every automatic procedure is the thresholding step, which detect blobs that represent either single chromosomes or clusters of chromosomes. The better the thresholding step, the easier is the subsequent disentanglement of chromosome clusters into single entities.

We implemented eleven thresholding methods, i.e. the ones that appear in the literature as the best performers, and compared their performance in segmenting chromosomes and chromosome clusters in cytogenetic Q-band images. The images are affected by the presence of hyper- or hypo-fluorescent regions and by a contrast variability between the stained chromosomes and the background. A thorough analysis of the results highlights that, although every single algorithm shows peculiar strong/weak points, Adaptive Threshold and Region Based Level Set have the overall best performance. In order to provide the scientific community with a public dataset, the data and manual segmentation used in this paper are available for public download at http://bioimlab.dei.unipd.it

1. Introduction

Chromosome karyotyping analysis [1] is an important screening and diagnostic procedure routinely performed in clinical and cancer cytogenetic labs. Chromosomes are first stained with a fluorescent dye, and then imaged through a microscope for subsequent analysis and classification. Each chromosome in the image has to be identified and assigned to one of 24 classes: the result is the so-called karyotype image, in which all chromosomes in a cell are graphically arranged according to an international system for cytogenetic nomenclature (ISCN) classification [2]. Fig. 1 shows four typical PAL resolution (768 × 576, 8 bits/pixel) Q-banding prometaphase images.

Individual chromosomes only appear as distinct bodies towards the end of the cell division cycle, at prophase, when they are long string-like objects, contracting and separating at metaphase, just before cell division. Most of the studies aimed at the development of automatic cytogenetics systems for the analysis of banded chromosome preparations have concentrated on the prometaphase, the intermediate stage of contraction between prophase and metaphase [3,4].
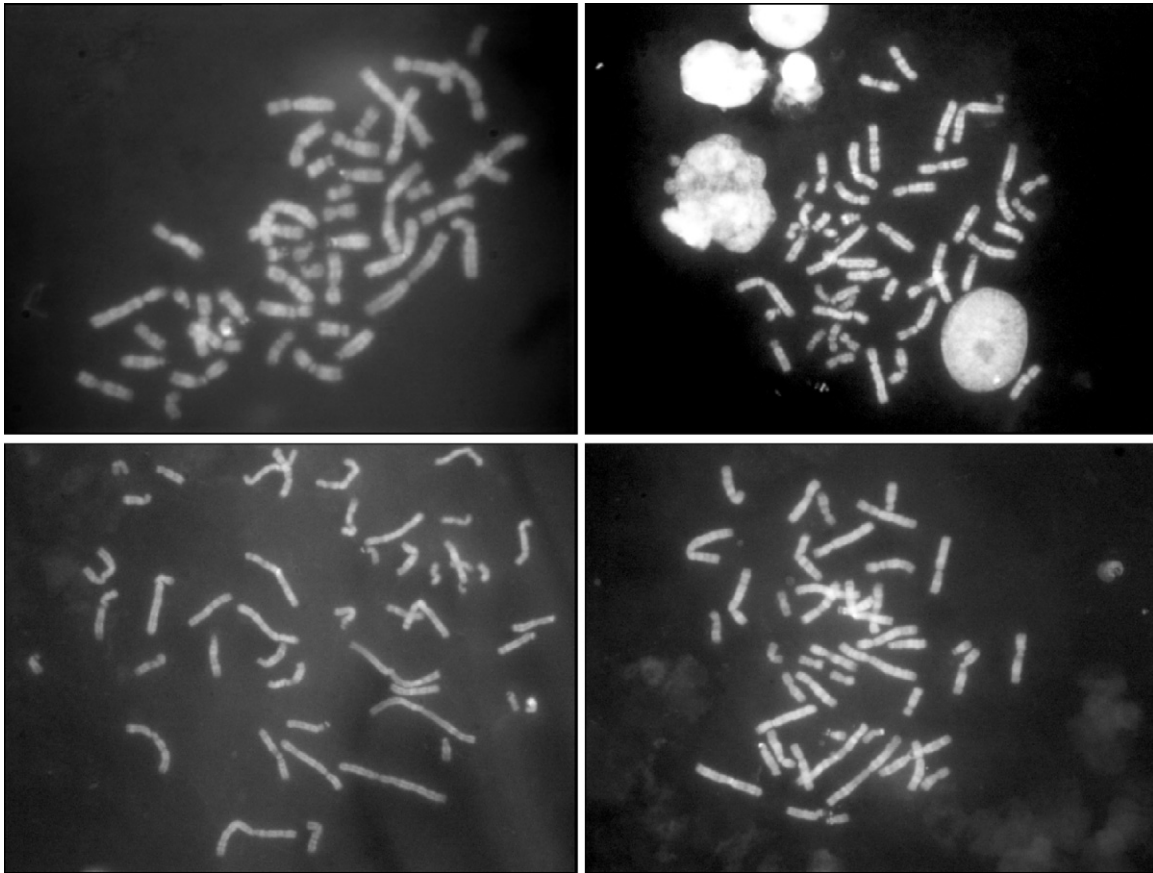
The first step to be taken in analyzing a chromosome image is the segmentation of chromosomes and chromosome clusters from the image background. Unfortunately, the high variability in chromosome and background fluorescence intensities makes the utilization of a global threshold

Fig. 1 – Typical Q-band prometaphase images acquired with PAL resolution.

impractical for a satisfactory segmentation of the image, since smaller chromosomes, and tails of the chromosomes, often appear with a lighter intensity than larger ones. Moreover, due to the blurred margins of the chromosomes, to the presence of staining debris, or to the the fact that long chromosomes may touch and overlap, the first segmentation step is usually unable to identify all chromosomes as single objects, but rather presents a number of clusters.

However, the main methods used to segment cytogenetic images are still based on the evaluation of a global threshold by means of the Otsu's method [5], on a global threshold with a Local Re-Thresholding (LRT) scheme [6,7], or on K-Means Clustering on Algebraic Moments (KM-AM) binarization [8]. In [9] a local adaptive thresholding (AdT) scheme has been proposed, but it seems to be sensitive to heavily clustered images, where the fluorescein leaking out of the chromosomes fills the region where the chromosomes are concentrated. Additionally, binary segmentation can be provided by optimizing the entropy separation between the two classes, as proposed in [10]. More recently, an adaptive segmentation scheme that globally optimizing the identification of region borders through Improved Sobel with Genetic Algorithm (IS-GA) has been proposed in [11], whereas a Multistage Adaptive Thresholding (MAT) preliminary dividing the image histogram into three segments is described in [12].

The level set technique has also proved a useful segmentation framework, especially for its ability in identifying smooth contours without topological constraint, so that the number of separate object identified and their shape may not be fixed. In order to deal with space variant contrast and luminosity heterogeneity, a Region-Based term has been introduced in the Level Set formulation (RBLS) [13,14].

Multilevel thresholding algorithm are computationally intensive, so that hybrid optimization methods combining classical Nelder–Mead (NM) simplex or Expectation-Maximization (EM) scheme with particle swarm (PS), are proposed respectively in [15] and in [16,17].

In the presence of so many techniques, each based on a very different rationale, there is little evidence about their relative performance when dealing with a specific problem. The aim of this work is to assess and compare the performance of all the algorithms quoted above when employed for human chromosome segmentation. We implemented the methods and then run them on a dataset of chromosome images, whose manual segmentation was performed in order to have a ground truth reference against which to compare the obtained results.

2. Chromosome data

Q-band prometaphase images are cytogenetic data obtained by staining the chromosomes with quinacrine, a fluorescent dye that concentrates in different regions of the chromosomes, giving rise to the characteristic banding patterns that identify the different chromosome types.