COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 108 (2012) 679–688

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

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

Article history:
Received 18 October 2011
Received in revised form
5 December 2011
Accepted 5 December 2011

Keywords:
Chromosome analysis
Karyotyping
Image segmentation
Level set
Thresholding
Particle swarm

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

* *Corresponding author*. Tel.: +39 049 827 7953; fax: +39 049 827 7699.
E-mail address: enrico.grisan@dei.unipd.it (E. Grisan).
0169-2607/$ – see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.cmpb.2011.12.003

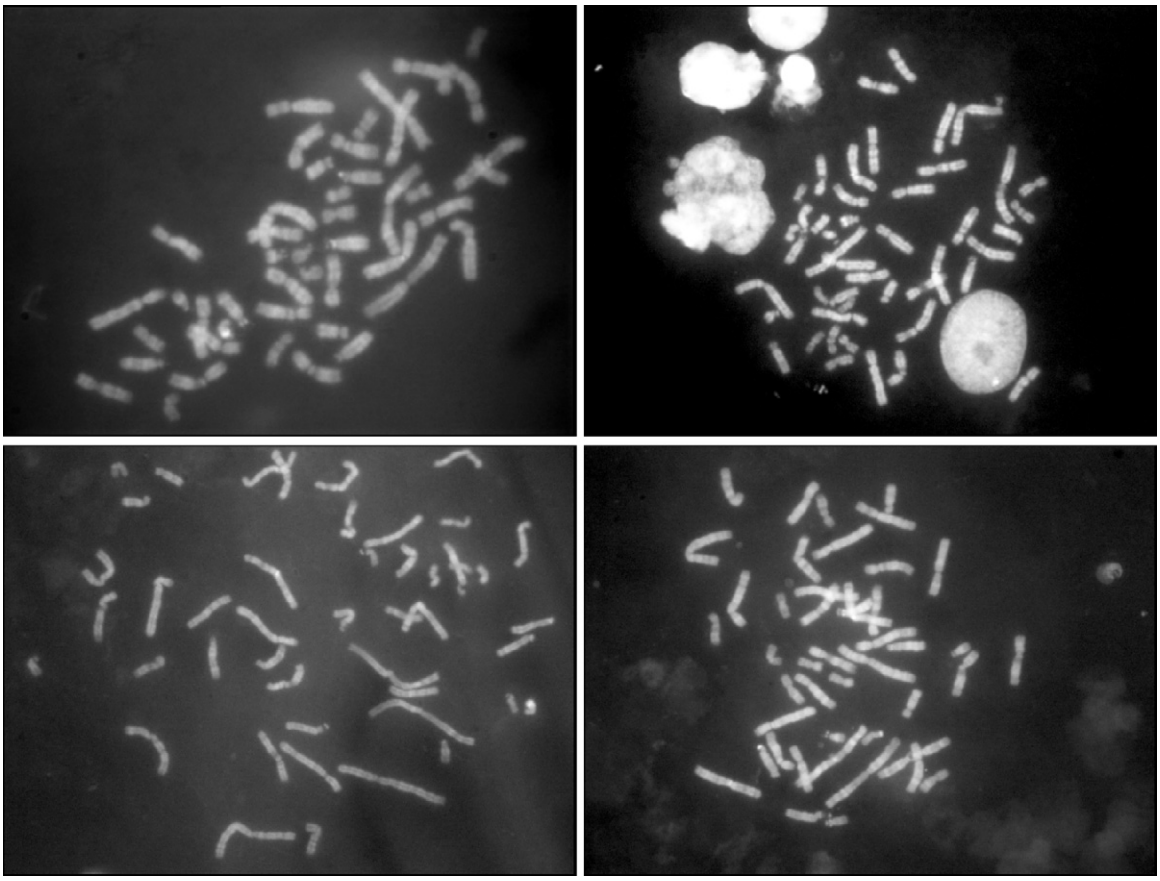

Fig. 1 – Typical Q-band prometaphase images acquired with PAL resolution.

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 (ISGA) 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.

The images thus appear as a dark background onto which the chromosomes stand out with bright and dark banding, as shown in Fig. 1.

The dataset used in this work is composed of 37 images with PAL resolution (768 × 576 pixels, 8 bits per pixel) acquired during routine laboratory analysis. In order to obtain a set of ground truth images as a reference against which to compare the results of the automatic methods, we manually annotated the 37 images. These ground truth images take binary values: 0 in the pixels belonging to the background and 1 in pixels that belong to a chromosome (with no distinction between single chromosomes or clusters of chromosomes). It is worth noting that the images do not necessarily contain a whole set of 46 chromosomes: as it may happen in routine laboratory acquisition, the whole set may be spread over different images. The dataset used in this work is publicly available for download at http://bioimlab.dei.unipd.it.

3. Methods

In this section we briefly describe all the thresholding techniques we have taken into account. We implemented them following the description provided in the original papers, highlighting the cases in which we arbitrarily set the value of free parameters.

Table 2 summarizes with brief descriptions the underlying rationales of each method employed in this work.

3.1. Otsu's Threshold

The idea beyond the Otsu's method [18] is that, given two classes of pixels in an image, it is desirable to find the separation that minimizes the combined intra-class variance (σ^2_{intra} in the following equations). Given two classes A and B, with numerosity n_A and n_B, the optimal threshold θ_{opt} is then:

$$\begin{aligned}\theta_{opt} &= arg\max\{\sigma^2_{intra}(\theta)\} \\ &= arg\max\{n_A(\theta)\sigma^2_A(\theta) + n_B(\theta)\sigma^2_B(\theta)\}\end{aligned} \quad (1)$$

3.2. Kapur's Threshold

Authors of [10] assume that an image is the outcome of two probability distributions, one for the object and the other for the background. In order to obtain the optimal threshold level, they maximize the total entropy of the partitioned image.

The probability of the gray level $i \in \{0, \ldots, L-1\}$ in an image containing N pixels, where the number of pixels with gray level i is denoted by $h(i)$, is $p_i = h(t)/N$. Hence, given a threshold value t, the entropies of the two resulting classes are:

$$H(0,t) = -\sum_{i=0}^{t-1}\frac{p-i}{w_0}\ln\frac{p_i}{w_0}, w_0 = \sum_{i=0}^{t-1}p_i \quad (2)$$

$$H(t,L) = -\sum_{i=t}^{L-1}\frac{p-i}{w_1}\ln\frac{p_i}{w_1}, w_1 = \sum_{i=t}^{L-1}p_i \quad (3)$$

Given that the between-class entropy E_{bc} of the thresholded image is defined as

$$E_{bc}(t) = H(0,t) + H(t,L) \quad (4)$$

the optimal threshold t_{opt} is the one that maximizes the between-class entropy $E_{bc}(t)$, that is:

$$t_{opt} = arg\max\{E_{bc}(t)\} \quad (5)$$

3.3. Improved Sobel with Genetic Algorithm (IS-GA)

In [11] the segmentation is obtained through an optimized edge detection. This is achieved with an improved Sobel operator, coupled with a global threshold whose value is determined by the optimization of a fitness function via genetic algorithm.

The classical Sobel operator computes an approximation of the gradient of the image intensity by convolving a discrete differentiation operator T_x. To detect the edge of an image $f(x, y)$, the template T_x and $T_y = T_x'$ are convoluted with f, resulting in the matrix M_x and M_y representing the gradients along x and along y. Hence, the total gradient value is $G = |\nabla f(x,y)| = \sqrt{M_x^2 + M_y^2}$. If on the one hand this formulation is relatively inexpensive in computational terms, on the other hand the gradient approximation which it produces is relatively crude, in particular for high frequency variations in the image.

In order to compensate the Sobel operator's shortcomings, four 5 × 5 templates, oriented along both canonic and diagonal directions, are proposed. The value associated to each element is determined following the same criterion used in the classical Sobel approach. Once the total gradient is obtained, the edge of the image can be computed by choosing an appropriate threshold which defines the *chromosome* and *background* classes.

To this end, authors in [11] employed a Genetic Algorithm (GA), which provides the maximum of a fitness function. That function penalizes the difference between the two classes' numerosities and favors the difference between the two classes' means. The GA consists of the following steps (numerical values are taken from the original implementation):

- *Initialization*: given that a possible solution (threshold) is encoded by an 8-bit binary string, N strings are generated as initial population.
- For M times:
  - *Selection*: a number of solutions are selected with probability proportional to their fitness value.
  - *Single-point crossover*: a random number of adjacent bits, extracted by pairs of previously selected solutions, are swapped.
  - *Mutation*: a single bit of a randomly selected solution is altered.
- *Decoding*: the solution with the highest fitness is selected.

3.4. K-Means Clustering on Algebraic Moments (KM-AM)

Clustering of algebraic moments is proposed as the first stage of the Classification-driven Partially Occluded Object Segmentation (CPOOS) used in [8], where it is claimed that algebraic

moments, as other statistical features, provide a more comprehensive characterization than the pixel gray level itself. The (p, q) moment m_{pq}, of a pixel of an image I which is measured in an neighborhood around the pixel (x_0, y_0) of dimension $N \times N$ is:

$$m_{pq} = \sum_{x=x_0-N/2}^{N/2} \sum_{y=y_0-N/2}^{N/2} (x - x_0)^p (y - y_0)^q I(x, y) \quad (6)$$

To keep the computational requirements low, only the moments m_{01}, m_{10}, m_{02} and m_{20} with $N = 1$ were computed and employed as input features given to a K-means clustering [19], which separated the image into the *chromosome* and *background* classes.

3.5. Fuzzy C-Mean Clustering on Intensities (FCM)

Recently, mostly on M-FISH chromosome images, fuzzy clustering [20] approaches have been proposed for segmenting the chromosomes from the background to tackle the non sharp margins of the chromosomes, and the fact that the distributions of features of the pixels belonging to chromosomes and background are often not well separated. Even if the utilization of clustering on intensities values is not widespread for the segmentation of Q-band images, it is worth including the evaluation of its performance for the completeness of this study.

Briefly, the fuzzy c-means clustering technique provides the identification of the cluster centroids c_j, with $j \in 1, \ldots, C$, that minimize the objective function:

$$J = \sum_{i=1}^{N} \sum_{j=1}^{C} w_{ij}^m \|g_i - c_j\|^m \quad (7)$$

where N is the number of pixels in the image, C is the established number of classes, g_i is the gray level value of the image $I(x_i, y_i)$, $w_{i,j}$ is the degree of membership of g_i to the cluster represented by the centroid c_j, and m is a user-defined parameter; in this work, we set $m = 2$, and $C = 2$.

3.6. Multistage Adaptive Thresholding (MAT)

In [12], a general locally adaptive thresholding method using neighborhood processing is proposed. It makes use of two global thresholds, T_0 and T_1, obtained from the global intensity distribution (Eq. (9)) and of a local threshold based on neighborhood means and variances (Eq. (8)).

Pixels whose gray value is less than T_0 are classified as background, whereas those whose gray value is larger than T_1 are classified as foreground objects. Pixels that are below T_1 and above T_0 are processed by means of a dedicated function. The combined thresholding function is:

$$t(x) = \begin{cases} p(f(x)) - T_0, & \text{if } p(x) \le T_0 \\ p(f(x)) - \mu_S(x) - k\sigma_S(x), & \text{if } T_0 < p(x) < T_1 \\ p(f(x)) - T_1 & \text{if } p(x), \ge T_1 \end{cases} \quad (8)$$

where $p(\cdot)$ is a preprocessing function of the input image $f(x)$, μ_s and σ_s are respectively the mean and the standard deviation computed in a $b \times b$ neighborhood region S centered at x, and k is a constant coefficient that determines the ratio of pixels being classified as the objects and the background.

The authors of [12] choose as $p(\cdot)$ the bilateral filter proposed in [21], because of its edge-preserving properties. To derive the two global thresholds, a simple percentile measurement procedure is employed: with a given fraction A, they define the percentile as the value of the brightness a such that: $P_r(a) = A$, or equivalently in which $p_r(t)$ is the intensity distribution of the input image:

$$T_0 = \int_{-\infty}^{T_1} p_r(t)dt = A$$
$$T_1 = \int_{T_1}^{\infty} p_r(t)dt = A \quad (9)$$

As regards the methods for the determination of the coefficient k in Eq. (8), we refer to the original work [12], whilst in our implementation we set $A = 0.1$, and fixed the value for the neighborhood dimension $b = 7$.

3.7. Local Re-Thresholding (LRT)

In [7] a two-step process is proposed for the chromosome initial segmentation. In the first step the Otzu's threshold is computed and applied on the entire image, providing, as claimed by the authors, an under segmentation of the image. Then a new Otzu's threshold is computed again separately for each resulting connected component of the previously thresholded image.

3.8. Adaptive Thresholding (AdT)

In [9] a space-variant adaptive thresholding scheme is proposed: the image I is divided into a tessellation of squares of fixed dimension l_{tess}. For each square separately, the Otsu's threshold [18] is evaluated.

A matrix of thresholds M_{th} is therefor obtained, scaled by a factor $1/l_{tess}$ with respect to the size of I. Resizing M_{th} to the size of I using bilinear interpolation, an image representing the local threshold for each pixel in I is obtained. In our implementation, $l_{tess} = 100$.

3.9. Region Based Level Sets (RBLS)

Following [13], let us consider an image $I : \Omega \in \mathbb{R}^2 \to \mathbb{R}$, and let C be a contour in the image domain Ω. For each point $x \in \Omega$ an energy E_x is defined as:

$$F_1(x, y) = K(x - y)|I(y) - f_1(x)|^2 \quad (10)$$

$$F_2(x, y) = K(x - y)|I(y) - f_2(x)|^2 \quad (11)$$

$$E_x(C) = \lambda_1 \int_{in(C)} F_1(x, y)dy + \lambda_2 \int_{out(C)} F_2(x, y)dy \quad (12)$$

where λ_1 and λ_2 are positive constants, and K is a kernel function with a localization property, so that $K(u)$ decreases and approaches zero as $|u|$ increases, and $f_1(x)$ and $f_2(x)$ are used to fit image intensities near the point x.

The energy defined is therefore region based, meaning that depending on the velocity of decrease of the kernel K, points close to x contribute to the integral more than those further away that provide a vanishing contribute. The idea is that for each center point x, the locally defined energy E_x can be minimized when the contour C is exactly on the object boundary and the fitting values f_1 and f_2 are chosen optimally. However, to find the entire object boundary, we must minimize E_x for all the center points x in the image domain Ω:

$$E(C) = \int_{\Omega} E_x(C)dx \tag{13}$$

In order to allow the contour to automatically handle topological changes, the energy can be converted to an equivalent level set formulation, where the contour C is represented by the zero level set of a Lipschitz function $\phi : \Omega \to \mathbb{R}$. The local energy function can be then rewritten as:

$$E_x(\phi) = \lambda_1 \int F_1(x, y)H(\phi(y))dy + \lambda_2 \int F_2(x, y)(1 - H(\phi(y)))dy \tag{14}$$

where $H(\cdot)$ is the Heaviside function.

In order to ensure stable evolution of the level set function and to regularize the contour, two additional terms are added to the energy: the first is used to penalize the deviation of the level set function ϕ from a signed distance function, and the second penalizes overlong contours:

$$P(\phi) = \int_{\Omega} \frac{1}{2}(|\nabla\phi(x)| - 1)^2 dx \tag{15}$$

$$L(\phi) = \int_{\Omega} \delta(\phi(x))|\nabla\phi(x)|dx \tag{16}$$

The energy functional becomes:

$$\Lambda(\phi) = E(\phi) + \mu P(\phi) + \nu L(\phi) \tag{17}$$

where μ and ν are nonnegative constants. From this formulation we can deduce the Euler-Lagrange equation for ϕ which, introducing an artificial time t, it yields the evolution equation:

$$\frac{\partial\phi}{\partial t} = -\delta(\phi)(\lambda_1 e_1 - \lambda_2 e_2) + \nu\delta(\phi)div\left(\frac{\nabla\phi}{|\nabla\phi|}\right) + \mu\left(\nabla^2\phi - div\left(\frac{\nabla\phi}{|\nabla\phi|}\right)\right) \tag{18}$$

with

$$e_i(x) = \int_{\Omega} F_i(x, y)dy \tag{19}$$

3.10. Multi-Thresholding with Particle Swarm and Nelder–Mead Optimization (PSO-NM)

Under the assumption that the image histogram can be modeled as a mixture of Gaussians, authors of [15] present a hybrid Nelder–Mead (NM) simplex search method coupled with a particle swarm optimization (PSO) to solve the Gaussian curve fitting. The idea is to combine NM simplex method, which is a very efficient local search procedure but its convergence is excessively sensitive to the starting point selected, with PSO, that belongs to the class of global search procedures but requires much computational effort.

The NM simplex search method is a derivative-free line search method that was particularly designed for traditional unconstrained minimization scenarios [22]. First, function values at the $(N+1)$ vertices of an initial simplex are evaluated, which is a polyhedron in the factor space of N input variables. In the minimization case, the vertex with the highest function value is *replaced* by a newly reflected, better point, which would be approximately located in the negative gradient direction. Through iterations, the simplex can come sequentially closer to a local optimum point.

PSO is an evolutionary optimization techniques based on a metaphor of social interaction such as bird flocking and fish schooling [23]. It simulates a commonly observed social behavior, where members of a group tend to follow the lead of the best of the group. Similar to genetic algorithms (GA, see Section 3.3), PSO is also population-based and evolutionary in nature, with one major difference from GA that it does not implement selection; namely, all particles in the population survive through the entire search process.

In the hybrid PSO-NM approach, given N parameters to be determined, an initial population of $3N+1$ particles is created in two steps: using a predetermined starting point, N particles are spawned in each coordinate direction, whereas the other $2N$ particles are randomly generated. A total of $3N+1$ particles are sorted by the fitness, and the top $N+1$ particles are then fed into the simplex search method to improve the $(N+1)$th particle. The other $2N$ particles are adjusted by the PSO method by taking into account the positions of the $N+1$ best particles.

In our implementation the number of level is $d=4$, so that $N=3d-1=11$ parameters are needed in order to identify the Gaussian mixture model. We considered as *chromosome* all those pixels whose probability of belonging to the Gaussian with the highest mean was the highest. We chose this assignation "rule" because it is the one that maximizes the method's performance.

3.11. Multi-Thresholding with Particle Swarm and EM Optimization (PSO-EM)

An alternative hybrid optimal estimation algorithm for solving multi-level thresholding problems in image segmentation is presented also in [16]. In this approach, the simplex local search i.e. replaced by Expectation-Maximization (EM) to update the best particles and to lead the remaining particles to seek optimal solution in search space. In the PSO-EM algorithm, the parameter estimates obtained from the PSO global search are expected to provide a suitable starting point for the EM step in fitting the Gaussian mixture model.

Unlike the real synergy of hybridization in PSO-NM (Section 3.10), the PSO-EM algorithm is more like a two-step search procedure by two different methods. The initial population is created by using a randomized starting point, that is, all particles are randomly generated in search space. Particles are sorted by the fitness, and the top one particle is then fed into the EM algorithm to update its location (solution) "once" in

parameter space. The remaining particles are adjusted by PSO taking into account the position of this elite particle if updated successfully. Note that if the global best particle could not be updated through the EM algorithm, the PSO-EM algorithm will generate a new one to replace the old one, and reselect the global best particle in the new swarm.

We chose $d = 4$ and adopted the same pixel class assignation "rule" described in Section 3.10 for the same reasons explained there.

4. Results

In order to assess the performance of the 11 methods, we run them on the dataset of 37 images and compared their output against the manual ground truth reference provided.

The performance of the various algorithms are evaluated considering both their ability to correctly identify the pixels belonging to chromosomes while rejecting pixels belonging to the background (Section 4.1), and also their ability to identify as separate objects all the blobs that have been manually segmented as separate objects (Section 4.2), being either chromosomes or clusters of chromosomes.

4.1. *Pixel classification*

Table 1 summarizes the most popular statistical measures of the performance of a binary classification test, and how they can be computed in terms of true positive, true negative, false positive, and false negative. In order to address the pixel classification capability of the methods we implemented, we considered four measures.

The first is the *sensitivity*, defined as the number of chromosome pixels identified as such over the number of ground truth chromosome pixels, thus accounting for the ability to correctly identify foreground pixels.

The second is the *positive predictive value* (PPV), which is the number of actual chromosome pixels identified as such over the total number of pixels that the algorithm identified as chromosomes. This last parameter is more useful than *specificity* in assessing the performance of the algorithms when the number of the pixels belonging to the two classes is highly unbalanced toward the negative class. Given the unbalance between the number of background pixels and chromosome pixels, the values of the *specificity* for all methods would saturate towards a meaningless 0.99. Sometimes it is useful to refer to the value of $1 - PPV$, which is called *false discovery rate* (FDR).

Finally, in order to obtain a global measure of the performance of the algorithms, and since the *accuracy* (fraction of correct classification, third row of Table 1) is not useful when dealing with classes of very different size, two parameters are computed. The first is the *distance from perfect classification* computed between the point (*sensitivity*, *specificity*) = (1, 1) and the corresponding nearest point on the ROC plane, having on the *x*-axis the value of *specificity* and on the *y*-axis the value of the *sensitivity*. The second is the *Matthews correlation coefficient* (MCC) [24], which is considered a good metric to evaluate the similarity between binary classifications [25].

Quantitative results are summarized in Table 3, were mean and standard deviation of the results on the 37 images

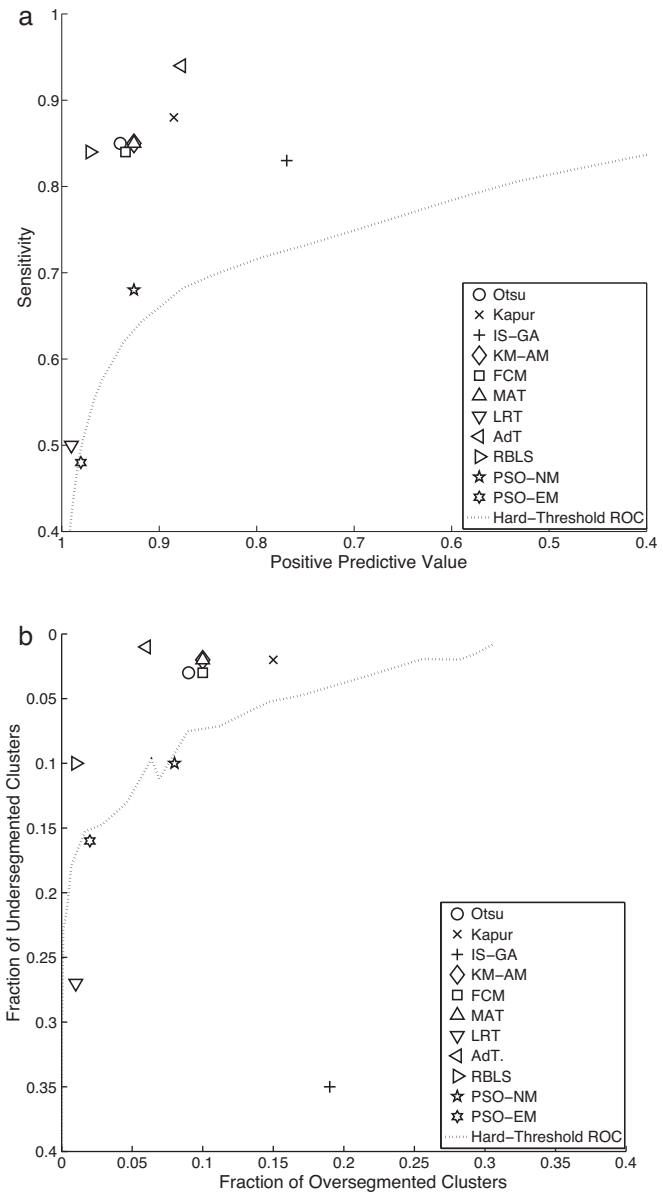

Fig. 2 – Plots of Positive Predictive Value against Sensitivity (a) and Fraction of Under-segmented Cluster (FUC) against Fraction of Over-segmented Clusters (FOC) (b). The reference dotted curve plot is obtained with different values of gray-level hard threshold, kept constant for all images in the database.

are computed for each method. Visual comparison of the various methods is shown in in Fig. 2a, where the performance are plotted in terms of *positive predicted value* (*x* axis) against *sensitivity* (*y* axis).

4.2. *Object segmentation*

In order to assess the correct identification of separate clusters and chromosomes in the images, with respect to the manual segmentation, two further parameters are evaluated.

The first counts the number of separate clusters that have been over-segmented with respect to the number of

Table 1 – Statistical measures for the performance of a binary classification test. P, N, $\widehat{P}$ and $\widehat{N}$ stand respectively for positives, negatives, estimated positives, and estimated negatives, while TP, TN, FP and FN stand respectively for true positive, true negative, false positive, and false negative.

Sensitivity	$= \frac{TP}{P} = \frac{TP}{TP+FN}$	*Specificity*	$= \frac{TN}{N} = \frac{TN}{TN+FP}$
PPV	$= \frac{TP}{\widehat{P}} = \frac{TP}{TP+FP}$	*FDR*	$= \frac{FP}{\widehat{P}} = \frac{FP}{FP+TP}$
Accuracy	$= \frac{TP+TN}{P+N}$	*MCC*	$= \frac{TP \times TN - FP \times FN}{\sqrt{P \times N \times \widehat{P} \times \widehat{N}}}$

Table 2 – Methods employed and brief description of their underlying rationale.

Method		Rationale
Global	Otsu	Minimization of the intra-class variance
	Kapur	Maximization of the between-class entropy
	IS-GA	Edge detection + fitness function optimization *via* GA
	KM-AM	Minimization of the distance between algebraic moments and clusters' centroids
	FCM	Minimization of the distance between intensities and clusters' centroids, weighted by a probabilistic membership function
Local	MAT	Two global percentile thresholds + local threshold based on mean and variance
	LRT	Global Otsu threshold + local Otsu thresholds on resulting connected components
	AdT	Interpolation of sampled local Otsu thresholds
	RBLS	Minimization of the Energy of a Curve that fits the contour of the objects of interest
Multi-th.	PSO-NM	Fitting of the histogram with Gaussian mixture model *via* simplex + PSO
	PSO-EM	Fitting of the histogram with Gaussian mixture model *via* EM + PSO

Table 3 – Performance on the *pixel classification*, mean (standard deviation). The best results are in bold.

Method		Sensitivity	Positive predictive value	Distance from perfect classification	Matthews correlation coefficient
Global methods	Otsu	0.85 (0.09)	0.94 (0.08)	0.16 (0.09)	0.87 (0.06)
	Kapur	0.88 (0.22)	0.89 (0.08)	0.13 (0.22)	0.83 (0.18)
	IS-GA	0.83 (0.08)	0.77 (0.08)	0.24 (0.08)	0.72 (0.06)
	KM-AM	0.85 (0.10)	0.93 (0.10)	0.16 (0.09)	0.86 (0.06)
	FCM	0.84 (0.10)	0.93 (0.10)	0.17 (0.09)	0.86 (0.06)
Local methods	MAT	0.85 (0.31)	0.93 (0.33)	0.17 (0.36)	0.86 (0.32)
	LRT	0.50 (0.10)	**0.99 (0.02)**	0.51 (0.10)	0.67 (0.08)
	AdT	**0.94 (0.03)**	0.88 (0.10)	**0.07 (0.03)**	**0.88 (0.06)**
	RBLS	0.84 (0.10)	**0.97 (0.02)**	0.16 (0.11)	**0.88 (0.07)**
Multi-threshold	PSO-NM	0.68 (0.28)	0.93 (0.15)	0.34 (0.27)	0.73 (0.21)
	PSO-EM	0.48 (0.32)	**0.98 (0.03)**	0.56 (0.32)	0.56 (0.31)

manually identified clusters (fraction of over-segmented clusters, FOC), assessing the percentage of clusters that are wrongly split into different objects by the different methods. The second parameter measures the fraction of multiple manual cluster that are identified as a single object by the segmentation method (fraction of under-segmented clusters, FUC).

Quantitative results are summarized in Table 4, were mean and standard deviation of both FOC and FUC on the 37 images are computed for each method. Visual comparison of the performance is shown in Fig. 2b, where the FOC (*x* axis) is plotted versus the FUC (*y* axis). It is worth noting that for none of the methods there are segmented objects in the images that do not overlap with the manually identified chromosomes.

5. Discussion

The eleven thresholding methods we considered in this work can be classified in three categories: the *global methods* provide a threshold that is constant all over the image, the *local methods* provide a space-variant threshold, and the *multi-threshold* identify a number of grey-level intervals separated by different thresholds. Although every single method has its own peculiarities (weaknesses or strengths), the ones belonging to the same category tend to share similar values in performance assessment.

Q-band prometaphase images present all the typical problems of the image segmentation task: both intra- and inter-image variability of contrast, luminosity, and appearance of the objects of interest (in both scale and shape), and background inhomogeneity.

The aim, in the specific application of the karyotyping, is to provide an accurate description of chromosome location in order to reduce the computational burden of the *disentanglement* step (i.e. the subsequent stage in which clusters are separated into the constituent chromosomes, e.g. as in [5,8,9,26–30]). In this context, the under-segmentation of chromosomes results in an unnecessary burden for this disentangling step, but the over-segmentation makes the recovering of the possibly wrongly divided chromosomes usually impossible. Thus, a slight under-segmentation (larger FUC) with negligible over-segmentation (smaller FOC) is much more desirable than the opposite situation.

Table 4 – Performance on the *object segmentation*, mean (standard deviation). The best results are in bold.

Method		1-Fraction of Over-segmented Clusters (FOC)	1-Fraction of Under-segmented Clusters (FUC)
Global methods	Otsu	0.97 (0.05)	0.91 (0.05)
	Kapur	**0.98 (0.05)**	0.85 (0.04)
	IS-GA	0.65 (0.14)	0.81 (0.03)
	KM-AM	**0.98 (0.05)**	0.90 (0.04)
	FCM	0.97 (0.05)	0.90 (0.05)
Local methods	MAT	0.98 (0.42)	0.90 (0.04)
	LRT	0.73 (0.05)	**0.99 (0.03)**
	AdT	**0.99 (0.03)**	0.94 (0.07)
	RBLS	0.90 (0.02)	**0.99 (0.01)**
Multi-threshold	PSO-NM	0.90 (0.10)	0.92 (0.07)
	PSO-EM	0.84 (0.06)	**0.98 (0.02)**

5.1. *Global methods*

By analyzing the performance metrics, it clearly appears that the classical segmentation algorithms that provide a global threshold (Otsu, Kapur, IS-GA, KM-AM, and FCM) perform well (even if they are not the best) and robustly with respect to all metrics.

The best performer in this category is the **Otsu's threshold**, which outperforms all other global methods in terms of *false discovery rate* (0.06) and *Matthews correlation coefficient* (0.87). However, both FCM and KM-AM are inferior to Otsu only by less than 0.01 in those two metrics. Kapur and KM-AM both show the second best FOC of the overall comparison (0.02, which is worse only than the 0.01 reached by the AdT). The worst performance in this category is the one provided by the IS-GA, which shows an unacceptable *false discovery rate* (0.30) as can be appreciated even by visual inspection in Fig. 3d and 4d.

5.2. *Local methods*

MAT, LRT, AdT and RBLS are all methods that fall under the *Local Methods* category, because the threshold they provide is space-variant. No surprise if the best overall performance are all within the local methods, since they are able to accommodate the background variation.

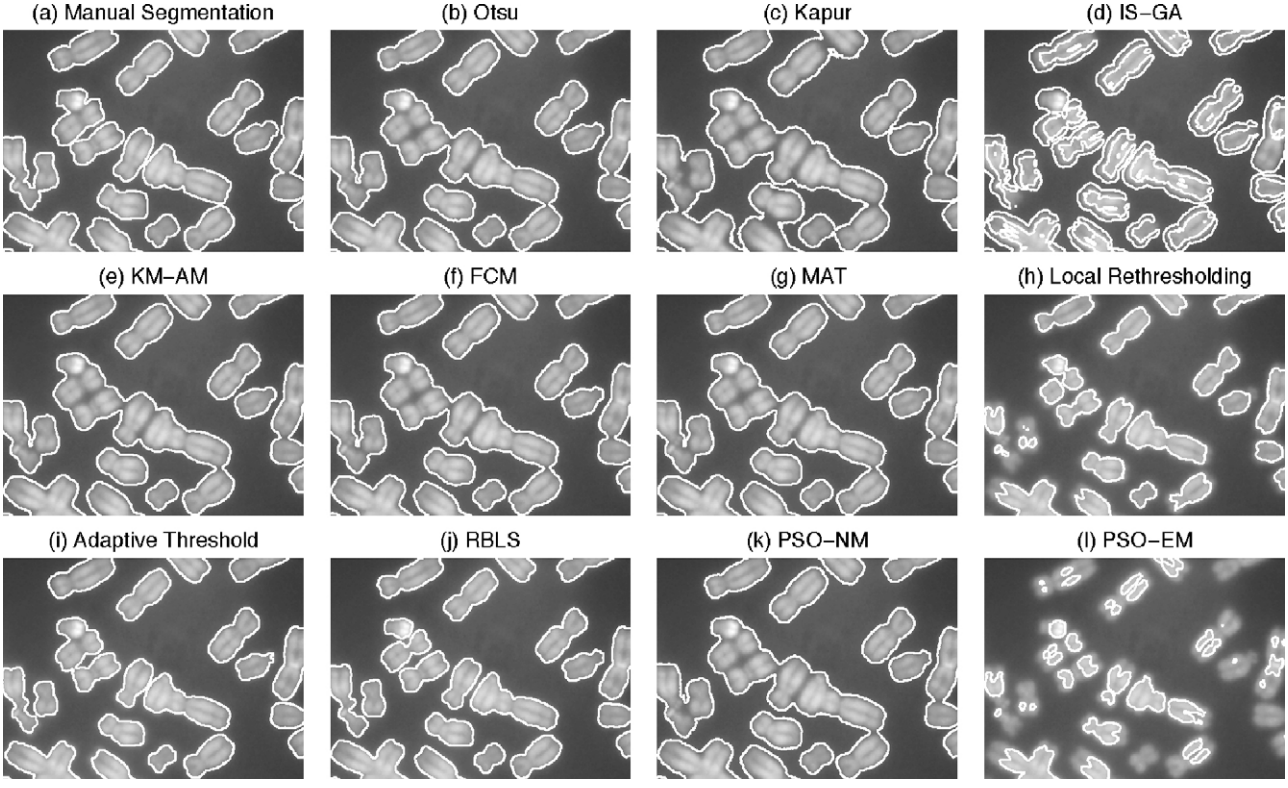

Fig. 3 – Detail of a karyotype image and the resulting boundary segmentation obtained by using the presented methods: (a) manual segmentation, (b) Otsu Threshold [18], (c) Kapur Threshold [10], (d) Improved Sobel with Genetic Algorithm [11], (e) K-Means Clustering on Algebraic Moments [8], (f) Fuzzy C-Mean Clustering [20], (g) Multistage Adaptive Thresholding [12], (h) Local Re-Thresholding [7], (i) Adaptive Thresholding [9], (j) Region Based Level Sets [13], (k) Multi-Thresholding by Nelder–Mead and Particle Swarm Optimization [15], (l) Multi-Thresholding by Expectation-Maximization and Particle Swarm Optimization [16,17].

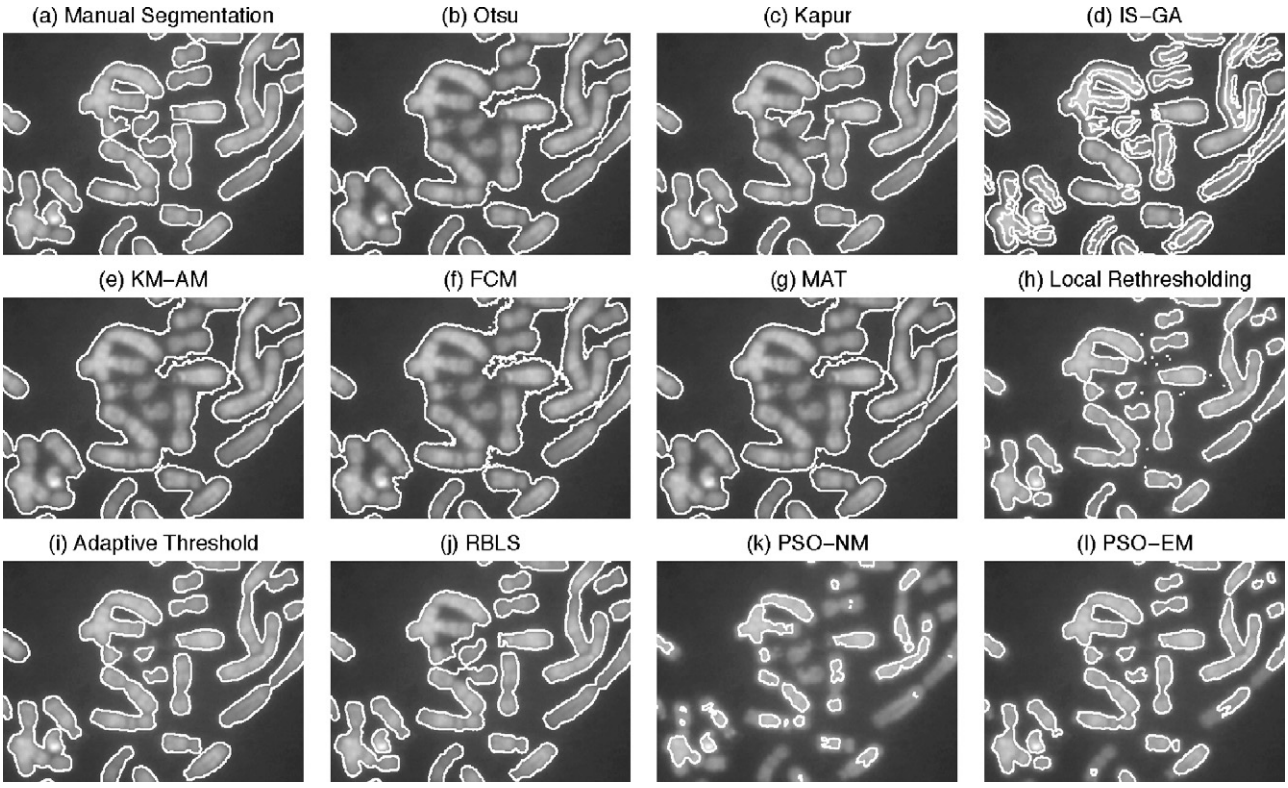

Fig. 4 – Detail of a karyotype image and the resulting boundary segmentation obtained by using the presented methods: (a) manual segmentation, (b) Otsu Threshold [18], (c) Kapur Threshold [10], (d) Improved Sobel with Genetic Algorithm [11], (e) K-Means Clustering on Algebraic Moments [8], (f) Fuzzy C-Mean Clustering [20], (g) Multistage Adaptive Thresholding [12], (h) Local Rethresholding [7], (i) Adaptive Thresholding [9], (j) Region-Based Level Sets [13], (k) Multi-Thresholding by Nelder–Mead and Particle Swarm Optimization [15], (l) Multi-Thresholding by Expectation-Maximization and Particle Swarm Optimization [16,17].

AdT shows the best *sensitivity* (0.94), *distance from perfect classification* (0.07) and *Matthews correlation coefficient* (0.88) among all the tested methods. It has also the lowest FOC amongst all methods (0.01). **RBLS** have the same *Matthews correlation coefficient* and is superior as the *false discovery rate* and FUC are concerned (0.03 and 0.01 respectively), but have also a lower *sensitivity* (0.84, see Fig. 4i *vs* 4j)). RBLS and MAT show about the same performance, which are similar to the best reached by the best among the Global Methods. LRT have at the same time almost the lowest overall *sensitivity* (0.5) and the best overall *Positive predictive value* (0.99), suggesting that this method segment as foreground very few pixels, regardless of the information they yield (see Fig. 3h and 4h). This is the reason why LRT share with the RBLS the best performance in terms of FUC (0.01).

5.3. *Multi-threshold*

Surprisingly, the multi-thresholding methods fail in correctly separating the chromosomes from the background, as the *sensitivity* they provide is very low (<0.68), probably because the lack of spatial constraints makes parts of the chromosomes to be classified as background.

Even if the number of classes identified by the multiple thresholds is sufficient to take into account the variability within an image, at the same time spatially distant regions with similar gray levels are classified similarly, even if in one case they may be a bright object on a dark background, and in the other a uniformly bright background.

It is worth noting that, like the LRT, the good overall results in terms of *positive prediction value* (0.98) and FUC (0.02) obtained with the PSO-EM are counterbalanced by a very poor *sensitivity* (0.48), which is the worst amongst all methods (see Fig. 3l and 4k).

6. Conclusion

In this paper we compared a variety of thresholding strategies by testing them on the specific problem of segmenting chromosomes (either single or in clusters) in Q-band prometaphase images. These images present two typical problems that arise in image segmentation: appearance variability of the objects of interest throughout the image and in different images, and background inhomogeneity.

Although every single algorithm has its peculiar strong/weak points, local methods have generally better performance than the other categories, because of their implicit capability of adapting to the background variation. Adaptive Threshold and Region Based Level Set showed noticeably better performance than their competitors. It

is worth noting that the behavior of the two methods is somehow symmetric: the former excels for its sensitivity and capability to avoid over-segmentation, the latter shines in positive prediction value and capability to avoid under-segmentation.

Conflict of interest

All authors have no conflict of interest.

Acknowledgment

The authors wish to thank TesiImaging Srl for financial support and for having kindly provided chromosome images.

REFERENCES

[1] J. Graham, J. Piper, Automatic Karyotype Analysis, 1994, pp. 141–185.
[2] Intarnational Standing Committee on Human Cytogenetic Nomenclature, ISCN: An International System for Human Cytogenetic Nomenclature (2005), Karger and Cytogenetics and Genome Research, 2005.
[3] E. Poletti, E. Grisan, A. Ruggeri, A modular framework for the automatic classification of chromosomes in q-band images, Computer Methods and Programs in Biomedicine (2011), September 30 [Epub ahead of print].
[4] E. Poletti, A. Ruggeri, E. Grisan, An improved classification scheme for chromosomes with missing data, in: Proc. 33rd Annual International Conference of the IEEE-EMBC, August 30–September 3, Boston, Massachusetts, USA, 2011.
[5] L. Ji, Intelligent splitting in the chromosome domain, Pattern Recognition 22 (5) (1989) 519–532.
[6] L. Ji, Fully automatic chromosome segmentation, Cytometry 17 (1994) 196–208.
[7] R.J. Stanley, J.M. Keller, P. Gader, C.W. Caldwell, Data-driven homologue matching for chromosome identification, IEEE Transactions on Medical Imaging 17 (3) (1998) 451–462.
[8] B. Lerner, Toward a completely automatic neural-network-based human chromosome analysis, IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics 28 (4) (1998) 544–552.
[9] E. Grisan, E. Poletti, A. Ruggeri, Automatic segmentation and disentangling of chromosomes in q-band prometaphase images, IEEE Transactions on Information Technology in Biomedicine 13 (4) (2009) 575–581.
[10] J. Kapur, P. Sahoo, A. Wong, A new method for graylevel picture thersholding using the entropy of the histogram, Computer Graphics and Image Processing 29 (3) (1985) 273–285.
[11] J. Zhang, Y. Chen, X. Huang, Edge detection of images based on improved sobel operator and genetic algorithms, in: International Conference on Image Analysis and Signal Processing ISAP09, China, 2009, pp. 31–35.
[12] F. Yan, H. Zhang, R. Kube, A multistage adaptive thresholding method, Pattern Recognition Letters (2005) 1183–1191.
[13] C. Li, C.-Y. Kao, J.C. Gore, Z. Ding, Minimization of region-scalable fitting energy for image segmentation, IEEE Transactions on Image Processing 17 (10) (2008) 1940–1949.
[14] E. Grisan, E. Poletti, A. Ruggeri, An improved segmentation of chromosomes in q-band prometaphase images using a region based level set, in: World Congress on Medical Physics and Biomedical Engineering, September 7–12, 2009, Munich, Germany, vol. 25/4, Springer, Berlin, Heidelberg, 2010, pp. 748–751.
[15] E. Zahara, S.-K.S. Fan, D.-M. Tsai, Optimal multi-thresholding using an hybrid optimization approach, Pattern Recognition Letters (2005) 1082–1095.
[16] S.-K.S. Fan, Y. Lin, A multi-level thresholding approach using a hybrid optimal estimation algorithm, Pattern Recognition Letters (2007) 662–669.
[17] S.-K.S. Fan, E. Zahara, A hybrid simplex search and particle swarm optimization for unconstrained optimization, European Journal of Operation Research 181 (2007) 527–548.
[18] N. Otsu, A threshold selection method from gray-level histogram, IEEE Transactions on System Man and Cybernetics 9 (1979) 62–66.
[19] MJB, Some methods for classification and analysis of multivariate observations, in: Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, pp. 1, 281–297, 1967.
[20] B.J.C., Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.
[21] C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images, in: Sixth International Conference on Computer Vision, 1998, 1998, pp. 839–846.
[22] J.A. Nelder, R. Mead, A simplex method for function minimization, The Computer Journal 7 (4) (1965) 308–313, doi:10.1093/comjnl/7.4.308.
[23] J. Kennedy, R. Eberhart, Particle swarm optimization, in: IEEE International Conference on Neural Networks, 1995. Proceedings, 1995, vol. 4, 1995, 1942–1948. doi:10.1109/ICNN.1995.488968.
[24] B. Matthews, Comparison of the predicted and observed secondary structure of t4 phage lysozyme, Biochimica et Biophysica Acta 405 (1975) 442–451.
[25] P. Baldi, S. Brunak, Y. Chauvin, C. Andersen, H. Nielsen, Assessing the accuracy of prediction algorithms for classification: an overview., Bioinformatics 16 (2000) 412–424.
[26] B. Lerner, H. Guterman, I. Dinstein, A classification-driven partially occluded object segmentation (CPOOS) method with application to chromosome analysis, IEEE Transactions on Signal Processing 46 (10) (1998) 2841–2847.
[27] G. Agam, I. Dinstein, Geometric separation of partially overlapping nonrigid objects applied to automatic chromosome segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (1997) 1212–1222.
[28] C. Urdiales García, A. Bandera Rubio, F. Arrebola Pérez, F. Sandoval Hernández, A curvature-based multiresolution automatic karyotyping system, Machine Vision and Applications 14 (2003) 145–156.
[29] X. Shunren, X. Weidong, S. Yutang, Two intelligent algorithms applied to automatic chromosome incision, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003 (ICASSP'03), 2003, pp. 697–700.
[30] G. Ritter, L. Gao, Automatic segmentation of metaphase cells based on global context and variant analysis, Pattern Recognition 41 (2008) 38–55.