A Novel Metric for Bone Marrow Cells Chromosome Pairing

Artem Khmelinskii*, Rodrigo Ventura, and João Sanches

***Abstract*—Karyotyping is a set of procedures, in the scope of the *cytogenetics*, that produces a visual representation of the 46 chromosomes observed during the *metaphase* step of the cellular division, called *mitosis*, paired and arranged in decreasing order of size. Automatic pairing of *bone marrow* cells is a difficult task because these chromosomes appear distorted, overlapped, and their images are usually blurred with undefined edges and low level of detail. In this paper, a new metric is proposed to compare this type of chromosome images toward the design of an automatic pairing algorithm for *leukemia* diagnostic purposes. Besides the features used in the traditional karyotyping procedures, a new feature, based on *mutual information*, is proposed to increase the discriminate power of the G-banding pattern dissimilarity between chromosomes and improve the performance of the classifier. The pairing algorithm is formulated as a combinatorial optimization problem where the distances between homologous chromosomes are minimized and the distances between nonhomologous ones are maximized. The optimization task is solved by using an integer programming approach. A new *bone marrow* chromosome dataset—Lisbon-K1 (LK1) chromosome dataset with 9200 chromosomes—was build for this study. These chromosomes have much lower quality than the classic Copenhagen, Edinburgh, and Philadelphia datasets, and its classification and pairing is therefore more difficult. Experiments using real images from the LK_1 and Grisan *et al.* datasets based on a leave-one-out cross-validation strategy are performed to test and validate the pairing algorithm.**

***Index Terms*—Bone marrow cells, chromosome pairing, classification, image processing, integer programming, leukemia, mutual information (MI), optical microscopy, optimization.**

I. Introduction

THE STUDY of chromosome morphology and its relation with some genetic diseases is the main goal of *cytogenetics*. Normal human cells have 23 classes of large linear nuclear chromosomes, in a total of 46 chromosomes per cell. The chromosomes contains approximately 30 000 genes (*genotype*) and large tracts of noncoding sequences. The analysis of genetic material can involve the examination of specific chromosomal regions using DNA probes, e.g., *fluorescent in situ hybridization* (FISH) [1], called *molecular cytogenetics*, *comparative genomic hybridization* (CGH) [2], or the morphological and pattern analysis of entire chromosomes, the *conventional cytogenetics*, which is the focus of this paper. These *cytogenetics* studies are very important in the detection of acquired chromosomal abnormalities, such as translocations, duplications, inversions, deletions, monosomies, or trisomies. These techniques are particularly useful in the diagnosis of cancerous diseases and are the preferred ones in the characterization of the different types of *leukemia*, which is the motivation of this paper [3].

The pairing of chromosomes is one of the main steps in conventional *cytogenetics* analysis where a correctly ordered *karyogram* is produced for diagnosis of genetic diseases based on the patient *karyotype*.

The *karyogram* is an image representation of the stained human chromosomes with the widely used Giemsa Stain metaphase spread (G-banding) [4], where the chromosomes are arranged in 22 pairs of somatic homologous elements plus two sex-determinative chromosomes (XX for the female or XY for the male), displayed in decreasing order of size. A *karyotype* is the set of characteristics extracted from the *karyogram* that may be used to detect chromosomal abnormalities. Fig. 1(a) shows a typical metaphasic plate for a normal male and Fig. 1(b) shows the respective *karyotype*. The *metaphase* is the step of the cellular division process where the chromosomes are in their most condensed state [5]. This is the most appropriated moment to its visualization and abnormality recognition because the chromosomes appear well defined and clear [3], [4], [6].

The pairing and *karyotyping* procedure, usually done manually by visual inspection, is time consuming and technically demanding. The application of the *G-banding* [4] procedure to the chromosomes generates a distinct transverse banding pattern characteristic for each class, which is the most important feature for chromosome classification and pairing. The *International System for Cytogenetic Nomenclature* (ISCN) [7] provides standard diagrams/ideograms of band profiles, as shown in Fig. 2, for all the chromosomes of a normal human, and the clinical staff is trained to pair and interpret each specific *karyogram* according to the ISCN information. Other features, related to the chromosome dimensions and shape, are also used to increase the discriminative power of the manual or automatic classifiers.

The design of automatic algorithms for pairing and classification has been an active field of research in the last two decades and still is an open problem today. The main aspects that must be taken into account in the design of an automatic pairing algorithm are common to the general classification problems: 1) feature selection; 2) classifier design; and 3) training and testing (validation). The most used features for chromosomes classification and pairing described in the literature are the following.

Manuscript received June 22, 2009; revised October 29, 2009; accepted December 14, 2009. Date of publication February 17, 2010; date of current version May 14, 2010. This work was supported in part by the Foundation for Science and Technology under ISR/IST plurianual funding. *Asterisk indicates corresponding author.*

*A. Khmelinskii is with the Institute for Systems and Robotics, Instituto Superior Técnico, 1049-001 Lisbon, Portugal (e-mail: aihmel@gmail.com).

R. Ventura and J. Sanches are with the Institute for Systems and Robotics, Instituto Superior Técnico, 1049-001 Lisbon, Portugal (e-mail: yoda@isr.ist.utl.pt; jmrs@ist.utl.pt).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2010.2040279

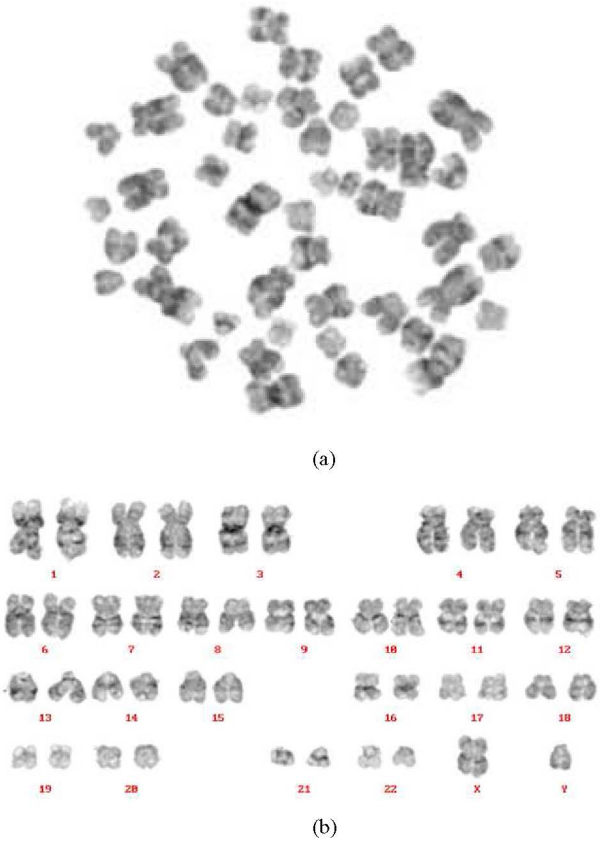


Fig. 1. Metaphase and corresponding karyogram. (a) Metaphase plate of a normal male. (b) Normal male karyotype.

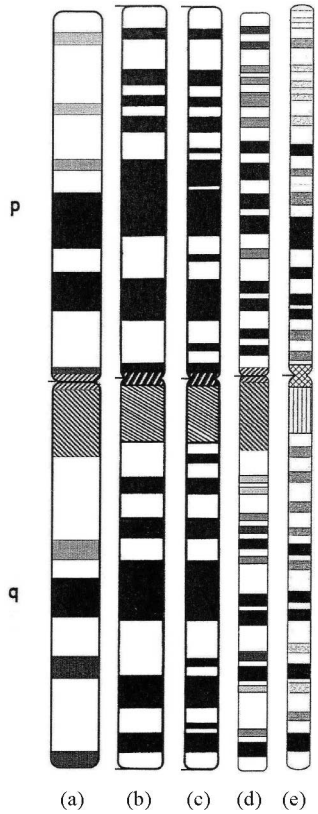


Fig. 2. ISCN Ideogram for the chromosomes of class 1 in various states of condensation. This picture shows in a more comprehensive way the difference between the chromosomes used in our paper and the traditional datasets. While the chromosome quality in the Edinburgh, Copenhagen, and Philadelphia datasets can be included in (b)–(e) interval, the quality of the chromosomes in our sets, extracted from *bone marrow* cells is below the (a) level of band description, which can be confirmed by analyzing the chromosomes of class 1 in the *karyogram* represented in Figs. 1(b), 3, and 5.

1) *Dimensions [8]:* Length, area, convex hull perimeter and centromeric index (ratio between the length of the short p-arm and total length of the chromosome).
2) *Geometrical:* Normalized area and fitted ellipsis axis dimensions and ratio.
3) *Pattern based:* The *G-banding* [4] profile was developed in the beginning of the 20th century and had become the most discriminative feature used in *karyotyping*. In order to grasp this relevant banding information, several strategies have been proposed, such as integrals or intensity average values computed over orthogonal directions to the medial axis of the chromosome [8], [9], *weighted density distributions* (WDDs) values of density and shape profiles [8]–[10], and *multiresolution analysis* of the banding pattern using wavelets [11]–[13].

Besides the feature selection, the classifier is the other crucial component in any automatic chromosome classification or pairing algorithm. Throughout the years, a significant number of approaches have been proposed and used in the design of classifiers, e.g., neural network and multilayer perceptron [9], [10], [14]–[17], Bayes [14], *hidden Markov models* (HMM) [18], template matching [19], wavelet [11], and fuzzy [19]. However, the classification rate obtained with automatic classifiers, typically in the range of 70%–80%, are still far from the performance reached by the human operator, typically with an approximate classification rate of 99.70% [16]. The main reasons for this difference is the difficulty to incorporate in the automatic classifiers the relevant subjective criteria used by the humans to achieve such high-classification performance, and eliminate the redundant and nondiscriminative features [16].

Recently, Wu *et al.* [20] and Wang *et al.* [21] proposed and evaluated a subspace approach based on *principal components analysis* (PCA), *linear discriminant analysis* (LDA), and *discrete cosine transform* (DCT) subspaces for automatic prototyping and classification of chromosome images. Another interesting approach was proposed by Karvelis *et al.* [22] to segment M-FISH chromosome images based on a multichannel watershed transform [23] and a Bayes classifier where a 82.40% classification rate is reported. This new technique is promising, but it is computationally expensive and demanding, and does not bring practical improvements to the *cytogenetics* analysis problem.

References to the pairing problem without classification are rare. Some of the most important ones are provided by Wu *et al.* [24] and Biyani *et al.* [25], where homologue pairing using maximum-weight graph matching is performed on 350 cells dataset and a 90.10% classification rate is reported.

In the traditional approach, the pairing procedure is performed after the classification of the chromosomes. Here, a different

approach is used. Instead of trying to accurately classify the chromosomes, the proposed algorithm attempts to pair them directly without knowing the class to which they belong. The algorithm was designed to process *bone marrow* cells chromosomes used in the scope of *leukemia* diagnosis. The images of this type of chromosomes present less quality than the ones used in the traditional genetic analysis using datasets such as Edinburgh [8], [16], Copenhagen [8]–[10], [17], [24], and Philadelphia [8], [10], namely, with respect to the centromere position, band profile description/discrimination, and level of condensation. Therefore, a new dataset, here called LK_1, with chromosomes not available in the traditional datasets, is used to evaluate the proposed algorithm.

The algorithm described in this paper is composed of the following three sequential steps.

1) *Image processing*—In this step, the chromosome images, extracted from the unordered karyogram, are processed by making histogram equalization, geometric distortion compensation, and dimensional scaling normalization (see Section II-A).
2) *Feature extraction*—In this step, discriminative features are extracted from the processed images, e.g., dimensions, normalized area, G-banding profile, and *mutual information* (MI) [26] between each pair of chromosomes in the karyogram (see Section II-B). The features extracted in this step are organized in a distance matrix containing the distances (using a given metric described later) between every two chromosomes in the *karyogram*.
3) *Pairing*—In this step, a combinatorial optimization problem is solved in order to obtain a permutation matrix that establishes the right correspondence between the chromosomes of each pair.

The images were acquired with a Leica Optical Microscope DM 2500. Some image preprocessing tasks, namely, noise reduction and chromosome segmentation, were manually performed with Leica continuous wave (CW) 4000 Karyo software used by the clinical staff. The pairing ground truth was obtained manually by the technical staff of the Institute of Molecular Medicine, Lisbon, and used to assess the accuracy of the proposed pairing algorithms.

Tests using real datasets with 4, 8, and 22 different classes of chromosomes (with increasing pairing difficulty, as shown in Fig. 3) extracted from 27 *karyograms* in a leave-one-out cross-validation and an 8-D feature space have been performed, and the error rate was computed. Results of the early stages of this study have already been published in [27] and [28], where the evaluation of the *MI* as a useful feature in the discrimination on the G-banding pattern is focused and the more appropriated metric to be used in the battery of classifiers is studied. This present paper puts together these results, proposes a new method to solve the optimization task of pairing, and tests the algorithm with an extended set of *bone marrow* chromosomes and with one other better quality dataset, as mentioned previously.

This paper is organized as follows. Section II formulates the problem, describing the real dataset, the extracted features used in a pairwise basis jointly with the training procedure and with the algorithm to compute the distances between chromosomes.

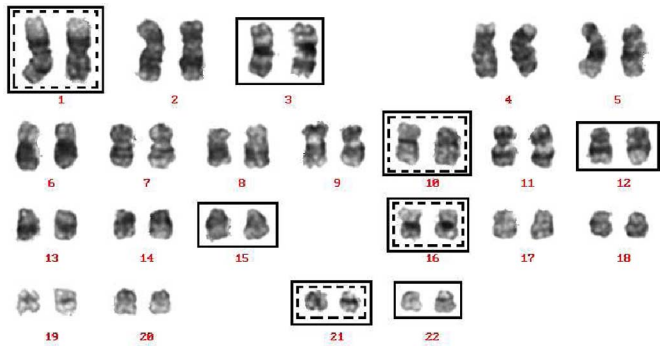


Fig. 3. Illustration of the three test sets extracted from each *karyogram*. (a) (Dashed line) Test set D_1. (b) (Continuous line) Test set D_2. (c) (All 22 classes) Test set D_3.

Section III describes the classification/pairing procedure and the classifier itself. Section IV illustrates the several experiments performed to evaluate the performance of the algorithm and the comparison results with other methods and datasets. Section V concludes the paper.

II. Material and Methods

In a typical *cytogenetic* diagnosis based on the *karyogram* at least 20 *karyograms* must be analyzed (when possible), while in a typical patient sample, the number of *metaphases* observed can reach 200. Therefore, the set of *metaphases* from which the *karyograms* are extracted for the diagnosis is composed of the ones presenting better quality for the trained human eye, namely, the ones that present less overlaps and blur. The automatic selection of *metaphases* from each *plate* is not considered in this paper. The chromosomes used in the *leukemia* diagnosis are obtained from *bone marrow* cells and present much less quality than the ones used in the traditional *cytogenetics*. The images from the Edinburgh and Copenhagen datasets are based on routinely acquired peripheral blood cells (constitutional *cytogenetics*), while in the Philadelphia dataset, the images are based on cells extracted from chorionic villus (prenatal *cytogenetics*). In both constitutional and prenatal *cytogenetics*, the observed cells are all equal, meaning that the same *karyotype* is always observed, independently of which cell is being analyzed, making it possible to choose those *metaphases* that present better image quality. On the contrary, in tumoral *cytogenetics* (*leukemia* in this case), a mixture of both normal and cancerous cells is observed, with significant differences not only between normal and tumoral cells, but also within the tumoral cells, which are the key cells for the diagnosis. In addition, while in prenatal and constitutional *cytogenetics*, it is possible to accurately control the cell division cycle in order to obtain chromosomes with the best possible morphology, in tumoral *cytogenetics*, this is not possible because it is much more difficult to predict the behavior of these cancerous cells.

A new chromosome dataset LK_1 [29] was created in collaboration with the Institute of Molecular Medicine, Lisbon, to test the classification and pairing algorithms of this type of "low" quality chromosomes for *leukemia* diagnosis purposes. The

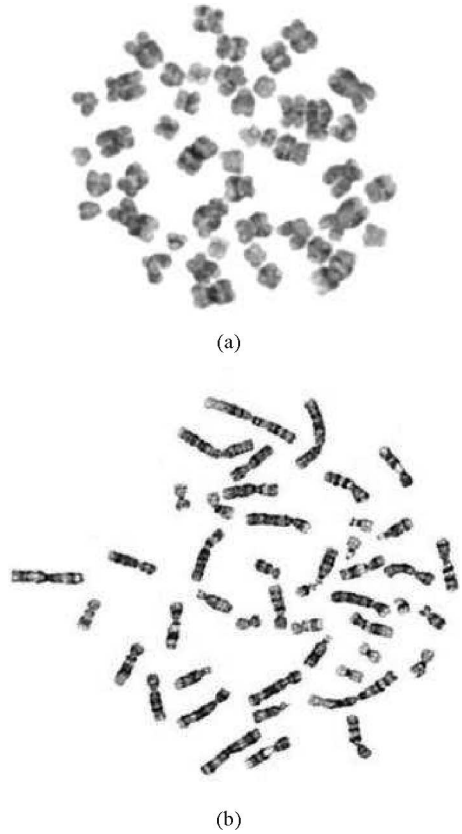


Fig. 4. (a) Two different metaphase plates containing *bone marrow* chromosomes and (b) chromosomes from the Copenhagen dataset [24].

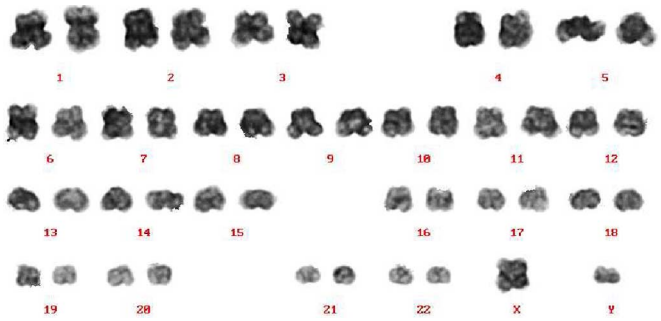


Fig. 5. Example of a very "low" quality *karyogram* based on *bone marrow* cells.

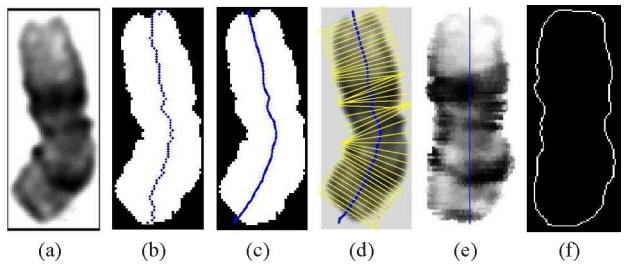


Fig. 6. Geometrical compensation. (a) Original image. (b) Chromosome and medial axis segmentation. (c) Axis smoothing. (d) and (e) Interpolation along orthogonal lines to the smoothed medial axis. (f) Border regularization.

bone marrow cell chromosomes in this new dataset were manually segmented, correctly oriented, ordered and annotated by the clinical staff to be used as ground truth data in the conducted tests. Two different quality *metaphase plates* are displayed in Fig. 4 to illustrate the differences between the chromosomes of our dataset and the chromosomes of the other standard datasets. The ideogram for the chromosomes of class 1, in various states of condensation, displayed in Fig. 2, shows in a schematic and more comprehensive way the difference between the chromosomes used in our study and the traditional datasets, while the chromosomes in the Edinburgh, Copenhagen, and Philadelphia datasets can be observed at different levels of detail shown in Fig. 2(b)–(e), the detail level of our chromosomes is below the level shown in Fig. 2(a), which can be confirmed by analyzing the chromosomes of class 1 in the *karyograms* represented in Figs. 1(b), 3, and 5.

To further validate the proposed algorithm, experiments were made by using Grisan *et al.* dataset [30]. This dataset is of the same nature and quality as the Philadelphia, Edinburgh, and Copenhagen datasets because the images are based on cells extracted from the amniotic fluid and choroidal villi (prenatal *cytogenetics*).

In this paper, the data are organized in *karyograms*, each one composed by $2N$ chromosomes images obtained with a Leica Optical Microscope DM 2500, where $N = 22$ is the number of homologous pairs, as shown in Fig. 1(b). The sex chromosomes were put aside and only karyograms that present no numerical or structural abnormalities were used at this stage of the paper.

The automatic pairing algorithm is composed of four main steps: 1) chromosome image extraction from the unordered *karyogram* and image processing; 2) feature extraction; 3) classifier training; and 4) pairing. In the next sections, these components are described in detail.

A. Image Processing

The image processing step aims at image contrast enhancement and compensation of geometric distortions observed in each chromosome not related with its intrinsic shape or size. The image brightness and contrast depend on the specific tuning of the microscope and the particular geometric shape of each chromosome depends on the specific *metaphase plaque* from which the chromosomes were extracted. These effects must be compensated to improve the results of the pairing algorithm.

The image processing step is composed of the following operations.

1) *Chromosome extraction*—Each chromosome is isolated from the unordered *karyogram*.
2) *Geometrical compensation*—The geometric compensation, performed by using the algorithm described in [8], is needed to obtain chromosomes with vertical medial axis, as shown in Fig. 6. This compensation algorithm is composed of the following main steps:
   a) chromosome and medial axis segmentation [see Fig. 6(b)];
   b) axis smoothing [see Fig. 6(c)];

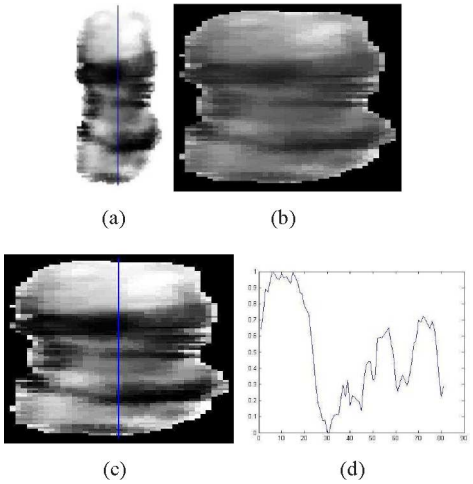


Fig. 7. Dimension and shape normalization and intensity equalization. (a) Geometrically compensated image. (b) Spatial normalization. (c) Histogram equalization. (d) Band profile.

c) interpolation along orthogonal lines to the smoothed medial axis [see Fig. 6(d)–(e)];
d) border regularization [see Fig. 6(f)].

3) *Shape normalization*—The features used in the comparison of chromosomes are grouped into two classes: 1) geometric based and 2) pattern based (G-banding). To compare chromosomes from a band pattern point of view, geometrical and dimensional differences must be removed, or at least attenuated. Therefore, a dimensional scaling is performed before the pattern features is extracted to make all the chromosome with the same size and aspect ratio by interpolating the original images, as shown in Fig. 7(a) and (b).
4) *Intensity compensation*—The *metaphase plaque* from which the chromosomes are extracted does not present a uniform brightness and contrast. To compensate for this inhomogeneity, the spatially scaled images are histogram equalized [31], as shown in Figs. 7(b) and (c).

B. Feature Extraction

The processed images are used to extract discriminative features to pair the chromosomes. We are using some of the most used features in the classification of chromosomes, but others, such as the centromere location, are not used due the very poor quality of the images. The extracted features, used to compute the distance between two chromosomes in the pairing process according to a metric defined later, are the following.

1) *Size/Area:* This class of features includes the *area* in pixels of each chromosome, its *perimeter*, bounding box *dimensions*, and *aspect ratio*, extracted from the nonnormalized shape images.
2) *Shape: Normalized area* is computed as the ratio between the perimeter and the area of the normalized shape images.
3) *Pattern:* Two classes of features are used to discriminate chromosomes pairs with respect to its pattern characteristics are the following.

a) *Band profile*—Band profiles, like the one displayed in Fig. 7(d), are computed as the average intensity values across each line of the shape normalized processed image, $h(n) = (1/N)\sum_{i=1}^{N} I(n,i)$, where N is the number of columns of the image. To avoid measurement degradation due to misalignment during the comparison step, the band profiles of two chromosomes are aligned. A shift constant $\hat{\tau}$ is estimated by maximizing the cross correlation function of the two profile vectors $h_i(n)$ and $h_j(n)$ as

$$\hat{\tau} = \arg\max_{\tau} \phi_{i,j}(\tau) \tag{1}$$

where $\phi_{i,j}(\tau) = \phi(h_i(n), h_j(n-\tau))$ is the cross correlation function [32]. The maximum of this function, when both profiles are aligned, occurs when $\tau = 0$. The distance between the chromosomes with respect to the band profile is the Euclidean distance between one profile and the other, shifted by $\hat{\tau}$

$$d(i,j) = \|h_i(n) - h_j(n-\hat{\tau})\|_2. \tag{2}$$

b) *Mutual Information*—The MI is proposed in this paper as a new feature for chromosome pairing that aims at increasing the discriminative power of the classifier with respect to the band pattern (G-banding) that characterizes each class of chromosomes. This measure is widely used in medical image processing, namely, in medical image registration [26] and is particularly suitable to compare pattern similarities based on the histograms of two images [31], such as chromosome images. This is a valid assumption since given two chromosomes from the same class, the corresponding G-banding will overlap and maximal dependence between the gray value of the images will be obtained [26].
The MI associated with two chromosome shape normalized images $I_A(i,j)$ and $I_B(i,j)$ is defined as follows [26]:

$$\mathrm{MI}(I_A, I_B) = \sum_{a,b} p_{AB}(a,b) \log\left[\frac{p_{AB}(a,b)}{p_A(a)p_B(b)}\right] \tag{3}$$

where $p_A(a)$ and $p_A(b)$ are the histograms of the images I_A and I_B, respectively, and $p_{AB}(a,b)$ is the joint histogram of both images. Note that this feature is not associated with each chromosome individually, as the previous ones, but is calculated for every pair. This property is particularly useful in our approach where the chromosomes are not individually classified.

The features extracted in this step are used to compute a $(44)^2 \times L$ matrix of distances, as shown in (8), where 44^2 is the total number of chromosome pairs in a given *karyogram*, excluding the sexual pair and L is the total number of features. The distance between two chromosomes with respect to each kth feature f_k is the absolute difference of both

features when they are scalars, $d_k(a,b) = |fa_k - fb_k|$, an Euclidean distance when they are vectors (e.g., band profile), $d_k(a,b) = \sqrt{\sum_n (fa_k(n) - fb_k(n))^2}$, and a single scalar in the case of the MI.

C. Distance Between Chromosomes

The overall distance between two chromosomes involving all features, given a vector of weights, $\mathbf{w}$, is defined as a weighted distance computed as follows:

$$D(i,j;\mathbf{w}) = \sum_{k=1}^{L} w(k) d_k(i,j) \tag{4}$$

where $w(k)$ is the weight associated with the kth feature and $d_k(i,j)$ is the distance between the ith and jth chromosomes with respect to the kth feature.

The proposed pairing algorithm is based on a supervised classifier, previously trained with manually paired images provided by experts. During the training step, a set of vector weights $\mathbf{w}_r$ with $1 \le r \le N = 22$ are estimated by using all possible pairs of the training set.

The distance between two chromosomes is assumed to be the smallest one among all weight vectors $\mathbf{w}_r$

$$\mathcal{D}(i,j) = \min_{r \in \{1,\ldots,22\}} D(i,j;\mathbf{w}_r). \tag{5}$$

The vectors $\mathbf{w}_r$, obtained during the training step, are computed by minimizing an energy function under the constraint $\|\mathbf{w}\| = 1$

$$\mathbf{w}_r = \arg\min_{\mathbf{w}:\|\mathbf{w}\|=1} E(\mathbf{w}) \tag{6}$$

where

$$E(\mathbf{w}_i) = \underbrace{\sum_{(a,b)\in V(i)} D(a,b;\mathbf{w}_i)}_{intra\text{class distance}} - \underbrace{\sum_{(a,b)\in U(i)} D(a,b;\mathbf{w}_i)}_{inter\text{class distance}} \tag{7}$$

where $V(i)$ is the set of all pairs of chromosomes of the ith class and $U(i)$ is the set of all chromosomes where at most one chromosome in each pair belongs to the ith class. Each weight vector $\mathbf{w}_r$ is computed by minimizing the sum of *intra*class distances (between chromosomes of the same class) and maximizing the sum of *inter*class distances (between chromosomes where at most one of them belongs to that class).

Let us consider the following matrix where each element $d_i(k)$ is the distance associated with the kth feature of the ith pair of chromosomes in the *karyogram*

$$\mathbf{\Theta}_r = \begin{pmatrix} d_1(1) & d_1(2) & d_1(3) & \ldots & d_1(L) \\ d_2(1) & d_2(2) & d_2(3) & \ldots & d_2(L) \\ d_3(1) & d_3(2) & d_3(3) & \ldots & d_3(L) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ d_R(1) & d_R(2) & d_R(3) & \ldots & d_R(L) \end{pmatrix}. \tag{8}$$

$\mathbf{\Theta}_r$ is a $R \times L$ matrix, where L is the number of features used in the pairing process and R the number of different pairs of chromosomes in the training set from class r. Let us also consider the matrix $\tilde{\mathbf{\Theta}}_r$ with the same structure of $\mathbf{\Theta}_r$, but now involving all pairs of the training set where at most one chromosome in each pair belongs to the rth class.

By using the Lagrange method, the energy function may be written as follows:

$$E(\mathbf{w}_r) = \Phi_r \mathbf{w}_r + \gamma \mathbf{w}_r^T \mathbf{w}_r \tag{9}$$

where $\Phi_r = \mathbf{1}^T \mathbf{\Theta}_r - \tilde{\mathbf{1}}^T \tilde{\mathbf{\Theta}}_r$ is a line vector with length L, $\mathbf{1}$ is a column vector of ones, and γ is the Lagrange multiplier. The minimizer of $E(\mathbf{w}_r)$ is

$$\mathbf{w}_r = \Phi_r^T / \sqrt{\Phi_r \Phi_r^T} = \mathrm{vers}(\Phi_r) \tag{10}$$

where $\mathrm{vers}(\Phi_r)$ is the unit length vector aligned with Φ_r.

Equation (10) is used in the training step to compute the set of vectors $\mathbf{w}_r$, with $1 \le r \le 22$, which are then used, in turn, to compute the distance between two chromosomes using (5).

The distances computed using (5) form a symmetric matrix of distances $\mathbf{D}$, where each element, $\mathcal{D}(i,j)$, is the distance between the ith and the jth chromosomes

$$\mathbf{D} = \begin{pmatrix} \mathcal{D}(1,1) & \mathcal{D}(1,2) & \mathcal{D}(1,3) & \ldots & \mathcal{D}(1,22) \\ \mathcal{D}(2,1) & \mathcal{D}(2,2) & \mathcal{D}(2,3) & \ldots & \mathcal{D}(2,22) \\ \mathcal{D}(3,1) & \mathcal{D}(3,2) & \mathcal{D}(3,3) & \ldots & \mathcal{D}(3,22) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \mathcal{D}(22,1) & \mathcal{D}(22,2) & \mathcal{D}(22,3) & \ldots & \mathcal{D}(22,22) \end{pmatrix}.$$

III. Classifier

The pairing process is a computationally hard problem because the optimal pairing must minimize the overall distance, i.e., the solution is the global minimum of the cost function. This problem can be stated as a combinatorial optimization problem. Moreover, it can be formulated as an integer programming problem, thus allowing for very efficient optimization methods. To do so, the cost function, as well as the constraints, have to be expressed by linear functions of the variables.

Considering n chromosomes (for n even), a pairing assignment $\mathcal{P}$ is defined as a set of ordered pairs (i,j), such that: 1) $i \neq j$ holds for any pair and 2) any given index i appears in no more than one pair of the set. A pairing assignment is said to be *total* if and only if, for any $i = 1,\ldots,n$, there is exactly one pair (r,s) in the set such that either $i = r$ or $i = s$. The sum of distances implied by a pairing $\mathcal{P}$ can be written as

$$C(\mathcal{P}) = \sum_{(i,j)\in\mathcal{P}} \mathcal{D}(i,j) \tag{11}$$

and the goal of the pairing process is to find a total pairing $\mathcal{P}$ that minimizes $C(\mathcal{P})$.

Note that the cost function (11) can be reformulated as a matrix inner product between the distance matrix $\mathbf{D}$ and a pairing matrix $\mathbf{X} = \{x(i,j)\}$, where

$$x(i,j) = \begin{cases} 1, & (i,j) \in \mathcal{P} \ \text{ or } \ (j,i) \in \mathcal{P} \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

Thus, (11) can be rewritten as $C(\mathcal{P}) = (1/2)\mathbf{D} \cdot \mathbf{X}$ where "$\cdot$" denotes the usual matrix inner product, which is defined as

follows:

$$\mathbf{D} \cdot \mathbf{X} := \sum_i \sum_j \mathcal{D}(i,j)x(i,j). \tag{13}$$

The cost function then becomes linear with the pairing matrix $\mathbf{X}$. The entries of this matrix are the parameters with respect to which (13) is to be minimized.

In order for the matrix $\mathbf{X}$ to represent a valid total pairing, this matrix has to satisfy constraints 1) and 2) mentiones earlier, which can be expressed in linear form as follows: constraint 1) is equivalent to state that the main diagonal of $\mathbf{D}$ is all zeros and constraint 2) corresponds to having one and only one entry equal to 1 in each row, as well as in each column. Constraining the domain of the matrix entries to be Boolean (i.e., $x(i,j) \in \{0,1\}$), the latter is the same to say that

$$\forall_i \sum_j x(i,j) = 1 \qquad \text{and} \qquad \forall_j \sum_i x(i,j) = 1. \tag{14}$$

The combinatorial optimization problem can then be restated as a integer programming problem, consisting of

$$\begin{aligned} &\text{minimize } \mathbf{D} \cdot \mathbf{X} \\ &\quad \text{where } \forall_i \forall_j x(i,j) \in \{0,1\} \\ &\text{subject to } \mathbf{X} = \mathbf{X}^T \\ &\qquad \forall_i x(i,i) = 0 \\ &\qquad \mathbf{X}\mathbf{1} = \mathbf{1} \\ &\qquad \mathbf{X}^T\mathbf{1} = \mathbf{1} \end{aligned} \tag{15}$$

where $\mathbf{1}$ is a column of ones of appropriate dimension.

This integer programming problem can be solved by standard numeric optimization packages, such as the GNU Linear Programming Kit[1] (GLPK) that was used here.

IV. Experiments and Results

Three kind of experiments were performed in this paper to assess the performance of the proposed algorithm, here called *linear combination of distances* (LCD): 1) in the first stage, the introduction of the MI is evaluated as a valuable discriminative factor to the pairing procedure; 2) in the second experiment, a comparison of the proposed method with three standard classification techniques—*nearest neighbor* (NN), *support vector machines* (SVMs), and *Euclidean distances* (EDs)—is done; and 3) in the last experiment, the performance of the LCD algorithm, designed to deal with *bone marrow* chromosomes, is assessed with a better quality Grisan *et al.* dataset [30]. This dataset is of the same nature and similar quality as the Philadelphia, Edinburgh, and Copenhagen datasets because these images are based on cells extracted from the amniotic fluid and choroidal villi.

Two subsets from the LK_1 dataset containing 27 and 19 real *karyograms* are used to form three test sets. From the first subset, two different test sets with different pairing difficulty are extracted. The first set, D_1, contains chromosomes only from classes 1, 10, 16, and 21 (framed with a dashed line in Fig. 3), presenting a low pairing difficulty level mainly due to the great size difference between the four classes of chromosomes. The second test set, D_2, is formed by adding chromosomes from classes 3, 12, 15, and 22 to D_1 (framed with a continuous line in Fig. 3). The added chromosomes are similar to the ones already present in D_1, and therefore, the pairing task difficulty increases. The last test set, D_3, contains all 22 homologous pairs extracted from the second subset, containing 19 *karyograms*. Therefore, the first testing set is composed of $27 \times 4 \times 2 = 216$ chromosomes, the second of 432, and the third one of 836 chromosomes, as shown in Table I.

TABLE I
Different Chromosome Test Sets Built From LK_1 and Grisan *et al.* Dataset Used to Evaluate Implemented Algorithms

Datasets	LK_1			Grisan *et al.*
Tissue	*bone marrow*			amniotic fluid, choroidal villi
Test sets	D_1	D_2	D_3	
# classes	4	8	22	22
# chromosomes	216	432	836	836

See Fig. 3, where D_1 (dashed line), D_2 (continuous line), and D_3 (all) are displayed.

The pairing results using different methods, test sets, and datasets are displayed in Table II. This table displays the number of pairing errors in each experiment, where each line corresponds to a single test using a leave-one-out cross-validation strategy (LOOCV), i.e., where all, but one, *karyogram* are used for training and the remaining one is used for testing. The symbol $\surd$ is used to indicate that a completely correct pairing was obtained (0 errors).

It is concluded that the introduction of the MI as a new feature, in the case of the LCD method proposed here, leads to an improvement in the pairing results in most of the cases. This can be observed in the case of experiments 3, 12, and 26 in D_2 with eight classes (see Table II) and 2, 13, 14, and 16 in the D_3 with 22 classes (see Table II).

For comparison purposes, three standard classification/pairing methods were also used with the D_3 dataset. In two of them, a two-stage process was followed: first, the chromosomes were individually classified, and then, it was evaluated whether both chromosomes of each (ground truth) class were classified in the same class. The two classification methods are: 1) the NN, where each chromosome is classified in the same class of the nearest chromosome in the training set, using the Euclidean metric, 2) the SVM classifier, trained with the training set. An open-source implementation of this classifier [33] is used. The goal of this experiment is to compare the performance of the proposed method, involving only a pairing procedure, with the classical approach where the pairing task is based on a previous classification step of each chromosome in one of the 22 admissible classes. As mentioned previously, the LOOCV approach was used in the training and testing steps. In the last classification/pairing method, here called 3) ED, the same optimization method as described in Section III was used, but the distance between chromosomes is simply the Euclidean distance computed with its respective feature vectors. The goal is to evaluate the benefits of the proposed metric (see Section II-C).

[1] http://www.gnu.org/software/glpk

TABLE II
CLASSIFICATION RESULTS

	LK_1									Grisan *et al.*
	D_1		D_2		D_3					
	LCD						NN	SVM	ED	LCD
	wo MI	MI	wo MI	MI	wo MI	MI				
1	√	√	2	2	13	13	18	13	17	5
2	√	√	√	√	10	8	17	12	15	9
3	√	√	4	2	2	2	15	10	8	7
4	√	√	√	√	2	2	17	14	8	2
5	√	√	√	√	5	5	18	14	10	8
6	√	√	√	√	13	13	18	16	18	4
7	√	√	√	√	√	√	15	6	2	2
8	√	√	√	√	8	8	6	15	10	11
9	√	√	√	√	2	2	13	7	4	2
10	√	√	√	2	8	10	14	12	12	√
11	√	√	√	√	5	5	14	11	11	4
12	√	√	2	√	10	10	12	9	12	8
13	√	√	√	√	7	6	14	10	9	5
14	√	√	√	√	15	13	17	13	18	9
15	2	2	√	√	√	√	16	10	4	3
16	√	√	2	2	4	2	16	11	12	5
17	√	√	√	√	2	2	15	10	9	2
18	2	2	√	√	13	13	15	11	12	7
19	√	√	√	√	11	11	20	11	16	7
20	2	2	√	√						
21–25	√	√	√	√						
26	√	√	4	2						
27	√	√	√	√						
total	6	6	14	10	130	125	290	215	207	100

(Columns 1-6) Number of pairing errors for the test sets D_1, D_2, and D_3 to assess the improvement of the algorithm due to the introduction of the MI as a discriminative factor in the distance function. wo = without MI. (Columns 6–9) Number of pairing errors for the test set D_3 to compare the proposed method - LCD with standard methods like the Nearest Neighbour (NN), Support Vector Machines (SVM) and Euclidean Distances (ED). (Columns 6 and 10) Number of pairing errors for the test set D_3 to validate the proposed algorithm—LCD written to process *bone marrow* cells chromosomes (LK_1 dataset) with a better quality Grisan *et al.* dataset.

TABLE III
COMPARISON OF CLASSIFICATION RATES FOR ALL TRAINING/TEST DATASETS

LK_1									Grisan *et al.*
D_1		D_2		D_3					
LCD						NN	SVM	ED	LCD
wo MI	MI	wo MI	MI	wo MI	MI				
94.44	94.44	93.52	95.37	68.90	**70.10**	30.60	48.60	50.50	76.10

(Columns 1–6) Without (wo) and with MI. (Columns 6–9) Comparison with other methods: NN algorithm, SVM, and ED. (Columns 6 and 10) Comparison with the Grisan *et al.* chromosome dataset.

Because our proposed method was developed to process *bone marrow* cells chromosomes with a much lower quality than the standard datasets (Copenhagen, Philadelphia, and Edinburgh), it was of our interest to assess how the algorithm performs with a higher quality dataset. As expected, the algorithm performs better for the Grisan *et al.* dataset. See columns 6 and 10 in Tables II and III. For the D_3 test set, the algorithm achieved a 76.10% classification rate for the Grisan *et al.* dataset, which is higher than the rate achieved for the LK_1 dataset, and although not presented here, this behavior is consistent across various experiments performed for the D_1 and D_2 test sets for the same dataset where 100% classification rates were obtained. It should be stressed, however, that even though the algorithm performed better on the Grisan *et al.* dataset, the final result could be improved even more if the images from this dataset had better resolution.

In the results presented here, the mean classification rate (MCR) corresponds to the average of the ratio between the number of correctly paired chromosomes and $N/2$ (the total number of pairs of chromosomes considered). Table III presents the comparison of the MCRs for all the performed experiments, which also allow us to conclude that the introduction of this new MI feature, indeed, leads to an increase in the correctly paired chromosomes rate, namely, when the pairing difficulty level is high. Additionally, the proposed method outperforms the standard NN, SVM, and ED algorithms that do not achieve a classification rate higher than 50.50%.

The classification time is dependent on the distance matrix $\mathbf{D}$, but in all tests performed here, it ranges from few milliseconds (four classes test sets) up to few tenth of a second for the 22 classes test sets. This is a major improvement against much higher execution times of the previously used classification

method by Khmelinskii *et al.* [27], [28] that could reach up to minutes when performed on the 22 classes test sets.

A 70.10% mean classification rate, reached when using the MI feature for the most realistic dataset D_3, with the 22 classes of chromosomes, is particularly relevant. These results, although below the 90.10% classification rate presented by Wu *et al.* [24], [25] are very optimistic and promising, given the poor quality of the images and the reduced amount of features used with this type of chromosomes.

V. Conclusion

In this paper, a new metric is proposed to measure the distance between chromosomes to be used in the automatic chromosome pairing procedure, in the scope of *karyotyping* process used in *cytogentic* analysis. The proposed algorithm is based on the traditional features extracted from the *karyogram*, such as, dimensions and banding profiles, plus a new one, based on the MI, to improve the discriminative power of the pairing algorithm with respect to the the G-banding pattern.

The main goal of this paper is to provide useful contributions toward the design of a fully automatic chromosomes pairing algorithm of *bone marrow* cells to be used in the diagnosis of *leukemia*. The images of these chromosomes present less quality and level of detail than the ones usually used in traditional genetic analysis using datasets such as Edinburgh, Copenhagen, and Philadelphia.

The algorithm is composed by four main steps: 1) image processing of the *karyograms* provided by the technicians; 2) feature extraction from the processed images characterizing the size, shape and band pattern; 3) training of a classifier (performed once) where similarity among chromosomes are characterized; and finally, 4) pairing.

In the image processing step, the chromosome images, extracted from the unordered *karyogram*, are processed in order to compensate for geometrical and intensity distortions, and to normalize their dimensions. This normalization is needed to make it possible the band pattern comparison between chromosomes. The features extracted from the processed images discriminate each pair with respect to their size, shape, and band pattern.

Here, a novel metric distance is proposed to be used in the pairing procedure that linearly combines the distances associated with each feature. The coefficients of the linear combination are obtained through a training step using chromosomes paired manually by experts. Vectors of coefficients associated with each one of the 22 classes are computed and the distance between two arbitrary chromosomes is the minimum one among all distances obtained with these 22 vectors.

The training process consists in the estimation of each vector of coefficients, from the chromosomes in the training set, by minimizing the overall distances between chromosomes of the same class (*intra*class) and by maximizing the distances between chromosomes when at least one of them does not belong to that class (*inter*class).

The pairing process is performed by efficiently solving a combinatorial problem where a permutation matrix is obtained from the distance matrix computed with the extracted features associated with each pair of chromosomes in the *karyogram*.

Tests using 19 *karyograms* based on *bone marrow* cells, working with 22 classes of chromosomes and a LOOCV strategy allow us to conclude that the proposed pairing algorithms, working within an 8-D feature space, achieves a 70.10% mean classification rate. The addition of the MI feature to the traditional geometrical and band profile features described in the literature leads to a clear improvement in the performance of the classifier. Executing the algorithm on a higher quality dataset, a 76.10% classification rate was obtained. Using 27 *karyograms* and working with a limited number of classes (≤ 8), a mean classification rate larger than 93% was obtained in all experiments.

Qualitative comparisons with the results obtained with the Leica CW 4000 Karyopairing software using the same data were performed and have shown relevant improvements.

In addition, a new chromosome dataset with 9200 chromosomes from *bone marrow* cells, called LK_1, was built to provide a ground truth to test classification and pairing algorithms for this type of "low" image quality chromosomes. This dataset was made publicly available [29].

The results presented in this paper are promising. In fact, despite the low quality of this type of chromosomes, it was shown that it is possible to achieve comparable classification rates to the ones obtained with the classical chromosome dataset, such as Edinburgh, Copenhagen, or Philadelphia, whose images are of significantly higher quality, presenting a uniform level of condensation, and from which it is possible to extract additional features, e.g., centromere position.

Acknowledgment

The authors would like to thank the clinical staff of the Cytogenetics/Virology Laboratory, Institute of Molecular Medicine (IMM)—GenoMed of Lisbon (namely, S. Santos, C. Sousa, and P. Costa) for providing us with the karyograms much needed for the testing of the implemented algorithms and Prof. M. do C. Fonseca (IMM), for providing us with this challenging task and all the support needed. They would also like to thank J. Xavier (ISR-IST) for helping us formulating the pairing as an integer programming problem.

References

[1] C. M. Price, "Fluorescence in situ hybridization," *Blood Rev.*, vol. 7, no. 2, pp. 127–134, 1993.

[2] J. C. Tan, J. J. Patel, A. Tan, C. J. Blain, T. J. Albert, N. F. Lobo, and M. T. Ferdig, "Optimizing comparative genomic hybridization probes for genotyping and SNP detection in plasmodium falciparum," *Genomics*, vol. 93, no. 6, pp. 543–550, Jun. 2009.

[3] D. E. Rooney and B. H. Czepulkowski, *Human Cytogenetics, A Practical Approach*, vol. II, 2nd ed. Ithaca, NY: IRL Press, 1992.

[4] J. Swansbury, *Cancer Cytogenetics: Methods and Protocols (Methods in Molecular Biology)*. Totowa, NJ: Humana Press, 2003.

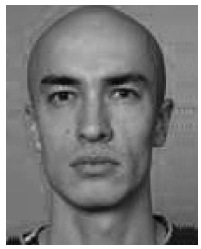
[5] H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, and J. E. Darnell, *Molecular Cell Biology*, 4th ed. San Francisco, CA: Freeman, 2004.

[6] B. Czepulkowski, *Analyzing Chromosomes*. Abingdon, U.K.: BIOS Scientific Publishing, 2001.

[7] L. G. Shaffer and N. Tommerup, Eds., *An International System for Human Cytogenetic Nomenclature (ISCN), 2005*. Basel, Switzerland: Karger and Cytogenetic and Genome Research, 2004, 130 pp. ISBN 3-8055-8019-3.

[8] J. Piper and E. Granum, "On fully automatic feature measurement for banded chromosome classification," *Cytometry*, vol. 10, pp. 242–255, 1989.

[9] J. R. Stanley, M. J. Keller, P. Gader, and W. C. Caldwell, "Data-driven homologue matching for chromosome identification," *IEEE Trans. Med. Imag.*, vol. 17, no. 3, pp. 451–462, Jun. 1998.
[10] M. Zardoshti-Kermani and A. Afshordi, "Classification of chromosomes using higher-order neural networks," in *Proc. IEEE Int. Conf. Neural Netw.*, Nov./Dec., 1995, vol. 5, pp. 2587–2591.
[11] Q. Wu and K. R. Castleman, "Automated chromosome classification using wavelet-based band pattern descriptors," in *Proc. 13th IEEE Symp. Comput.-Based Med. Syst.*, Jun. 2000, pp. 189–194.
[12] N. Sweeney, R. L. Becker, and B. Sweeney, "A comparison of wavelet and Fourier descriptors for a neural network chromosome classifier," in *Proc. 19th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Oct. 1997, vol. 3, pp. 1359–1362.
[13] L. V. Guimaraes, A. Schuck, and A. Elbern, "Chromosome classification for karyotype composing applying shape representation on wavelet packet transform," in *Proc. 25th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Sep. 2003, vol. 1, pp. 941–943.
[14] B. Lerner, H. Guterman, I. Dinstein, and Y. Romem, "A comparison of multilayer perceptron neural network and bayes piecewise classifier for chromosome classification," in *Proc. IEEE World Congr. Neural Netw., IEEE Int. Conf. Comput. Intell.*, Jun./Jul. 1994, vol. 6, pp. 3472–3477.
[15] B. Lerner, M. Levinstein, B. Rosenberg, H. Guterman, L. Dinstein, and Y. Romem, "Feature selection and chromosome classification using a multilayer perceptron neural network," in *Proc. IEEE World Congr. Comput. Intell., IEEE Int. Conf. Neural Netw.*, Jun./Jul. 1994, vol. 6, pp. 3540–3545.
[16] B. Lerner, "Toward a completely automatic neural-network-based human chromosome analysis," *IEEE Trans. Syst., Man Cybern. B, Cybern.*, vol. 28, no. 4, pp. 544–552, Aug. 1998.
[17] J. M. Cho, "Chromosome classification using backpropagation neural networks," *IEEE Eng. Med. Biol. Mag.*, vol. 19, no. 1, pp. 28–33, Jan./Feb. 2000.
[18] J. M. Conroy, R. L. Becker, Jr, W. Lefkowitz, K. L. Christopher, R. B. Surana, T. O'Leary, D. P. O'Leary, and T. G. Kolda, "Hidden Markov models for chromosome identification," in *Proc. 14th IEEE Symp. Comput.-Based Med. Syst.*, Jul. 2001, pp. 473–477.
[19] A. M. Badawi, K. G. Hasan, E. A. Aly, and R. A. Messiha, "Chromosomes classification based on neural networks, fuzzy rule based, and template matching classifiers," in *Proc. 46th IEEE Int. Midwest Symp. Circuits Syst.*, Dec. 2003, vol. 1, pp. 383–387.
[20] Q. Wu, Z. Liu, T. Chen, Z. Xiong, and K. R. Castleman, "Subspace-based prototyping and classification of chromosome images," *IEEE Trans. Image Process.*, vol. 14, no. 9, pp. 1277–1287, Sep. 2005.
[21] Y. Wang, Q. Wu, K. R. Castleman, and Z. Xiong, "Chromosome image enhancement using multiscale differential operators," *IEEE Trans. Med. Imag.*, vol. 22, no. 5, pp. 685–693, May 2003.
[22] P. S. Karvelis, A. T. Tzallas, D. I. Fotiadis, and I. Georgiou, "A multichannel watershed-based segmentation method for multispectral chromosome classification," *IEEE Trans. Med. Imag.*, vol. 27, no. 5, pp. 697–708, May 2008.
[23] S. Beucher, "Watershed, hierarchical segmentation and waterfall algorithm," in *Mathematical Morphology and its Applications to Image Processing*, E. Dougherty, Ed. Boston, MA: Kluwer, 1994.
[24] X. Wu, P. Biyani, S. Dumitrescu, and Q. Wu, "Globally optimal classification and pairing of human chromosomes," in *Proc. 26th Annu. Int. Conf. IEEE EMBS*, Sep. 2004, pp. 2789–2792.
[25] P. Biyani, X. Wu, and A. Sinha, "Joint classification and pairing of human chromosomes," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 2, no. 2, pp. 102–109, Apr./Jun. 2005.
[26] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: A survey," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.
[27] A. Khmelinskii, R. Ventura, and J. Sanches, "Chromosome pairing for karyotyping purposes using mutual information," in *Proc. 5th IEEE Int. Symp. Biomed. Imag.: Nano Macro*, May 2008, pp. 484–487.
[28] A. Khmelinskii, R. Ventura, and J. Sanches, "Automatic chromosome pairing using mutual information," in *Proc. IEEE 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC 2008)*, Aug., pp. 1918–1921.
[29] Lisbon K1—Chromosome Dataset. (2010). [Online]. Available: http://mediawiki.isr.ist.utl.pt/wiki/Lisbon-K_Chromosome_Dataset
[30] E. Grisan, E. Poletti, and A. Ruggeri, "Automatic segmentation and disentangling of chromosome in Q-band prometaphase images," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 4, pp. 575–581, 2009.
[31] A. K. Jain, *Fundamentals of Digital Image Processing*. Upper Saddle River, NJ: Prentice-Hall, 1989.
[32] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Upper Saddle River, NJ: Prentice-Hall, 2000.
[33] C. C. Chang and C. J. Lin. (2001). LIBSVM: A library for support vector machines [Online]. Available: http://www.csie.ntu.edu.tw/cjlin/libsvm



Artem Khmelinskii received the M.Sc. degree in biomedical engineering from the Instituto Superior Técnico (IST), Lisbon, Portugal, in 2007. He is currently working toward the Ph.D. degree at the Section of Knowledge Guided Image Processing (KGB), Division of Image Processing (LKEB), Department of Radiology, Leiden University Medical Center, Leiden, The Netherlands.

He has been engaged in multimodality image registration in small animal imaging. He is also a Researcher with the Institute for Systems and Robotics, IST, where he is engaged in the chromosome pairing problem. His current research interests include (bio)medical image processing, preclinical research, and molecular imaging analysis.

Rodrigo Ventura received the Licenciatura, M.Sc., and Ph.D. degrees from the Instituto Superior Técnico, Lisbon, Portugal, in 1996, 2000, and 2008, respectively, all in electrical engineering and computers.

Since 1998, he has been an Assistant Professor in IST, where he is currently a member of the Institute for Systems and Robotics, a Portuguese private research institution. His current research interests include biologically inspired cognitive architectures, field robotics, and artificial intelligence in general.

João Sanches received the E.E., M.Sc., and Ph.D. degrees from the Technical University of Lisbon, Lisbon, Portugal, in 1991, 1996, and 2003, respectively.

He is currently an Assistant Professor in the Electrical and Computer Engineering Department, Technical University of Lisbon, where he is also a Researcher in the Institute for Systems and Robotics. He was engaged in teaching signal processing and control. He has been engaged in functional MRI, 3-D ultrasound, confocal imaging microscopy, and neurophysiology, mostly in collaboration with several groups from the Medical School at the Hospital de Santa Maria, Lisbon. He has authored or coauthored numerous papers in the field of biomedical engineering. His current research interests include biomedical engineering.