Pattern Recognition Letters 31 (2010) 2474–2488

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Identifying touching and overlapping chromosomes using the watershed transform and gradient paths

Petros Karvelis [a,b], Aristidis Likas [a], Dimitrios I. Fotiadis [b,*]

[a] Department of Computer Science, University of Ioannina, GR 45110 Ioannina, Greece
[b] Unit of Medical Technology and Intelligent Information Systems, Department of Materials Science and Engineering, University of Ioannina, GR 45110 Ioannina, Greece

ARTICLE INFO

Article history:
Received 23 December 2009
Available online 17 August 2010
Communicated by T. Vasilakos

Keywords:
Chromosome classification
Watershed transform
Karyotyping
Multiplex Fluorescent In Situ Hybridization

ABSTRACT

Automation of chromosome analysis has long been considered as a difficult task. However the advent of Multiplex Fluorescence In Situ Hybridization (M-FISH) made the analysis of chromosomes much easier. Nevertheless, the chromosomes in an M-FISH image do very often partially occlude each other; hence, their segmentation is not trivial and requires the application of a dedicated procedure. In this paper a method is presented for the segmentation of touching and overlapping groups of chromosomes in M-FISH images. Initially, the watershed transform is applied and the image is decomposed into watershed regions. Next, gradient paths starting from points of high concavity are computed for each produced region. Finally, adjacent regions are merged producing the final chromosome areas. To validate our method a benchmark database of 183 M-FISH images has been used. The proposed algorithm resulted in a 90.6% success rate for touching chromosomes and 80.4% for overlapping groups of chromosomes.

1. Introduction

Chromosomes are structures that contain the genetic information of cells. In a normal, nucleated human cell, there are 46 chromosomes represented in the clinical routine by a structure called the karyotype. The karyotype shows the complete set of chromosomes organized into 22 classes (each of which consists of a matching pair of two homologous chromosomes) and two sex chromosomes, XX in females or XY in males (Thompson et al., 1991). Producing a karyotype of a cell is of practical importance since it greatly facilitates the detection of abnormalities in the chromosome structure as shown in Fig. 1. Normally, the procedure of assigning each chromosome to a class (karyotyping) is based on the visual scanning of chromosome images by experts (biologists or cytogeneticists) (Thompson et al., 1991). This visual inspection is a time consuming and expensive process. Hence automated image chromosome analysis is still an important problem.

A technique was developed in the mid 90s to stain chromosomes with multiple colours so that each chromosome class appears with a distinct colour (Speicher et al., 1996). In this technique all chromosomes are labelled with five fluorophores. Also a DNA stain, called DAPI (4′,6-diamidino-2-phenylindole), is used to stain all the chromosomes with the same colour. The fluorophores attach to specific sequences of DNA, thus each pixel of the new multispectral image is represented as a five-dimensional vector, where each element of the vector represents the magnitude of the dye at that pixel of the image, Fig. 2. This technique not only facilitates the detection of subtle chromosomal aberrations (Veldman et al., 1997), but also makes the analysis of chromosome images easier; both for human inspection and computerized analysis. However, in practice, fluorophore absorption is not binary and there is significant overlap between each of the fluorophore absorptions along with variability in signal strength. This leads to a non-trivial classification problem, especially in the context of touching or overlapping regions (Schwartzkopf et al., 2005).

Many attempts have been made to automate parts of the chromosome M-FISH image analysis procedure (Schwartzkopf et al., 2005; Sampat et al., 2005; Wang and Castleman, 2005; Wang and Dandpat, 2006; Karvelis et al., 2008). However, chromosome images are inherent with the partial occlusion and touching of chromosomes, as shown in Fig. 3. This is one of the major factors hindering automatic analysis. Spectrum based methods use a pixel-by-pixel classifier to classify each pixel of the M-FISH image and this information may be sufficient to segment touching and overlapping chromosomes (Schwartzkopf et al., 2005). However the measured fluorescence at a pixel may be the combination of fluorescence in a neighbouring region leading many times to misclassification errors. These factors make the pixel spectral information of touching or overlapping chromosomes unreliable. Hence the spectral information alone cannot separate the touching and overlapping chromosomes efficiently.

On the other hand there is a variety of geometric separation based methods proposed in the literature for greyscale

* Corresponding author. Tel.: +30 26510 08803; fax: +30 2651007092.
E-mail address: fotiadis@cs.uoi.gr (D.I. Fotiadis).

0167-8655/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2010.08.002

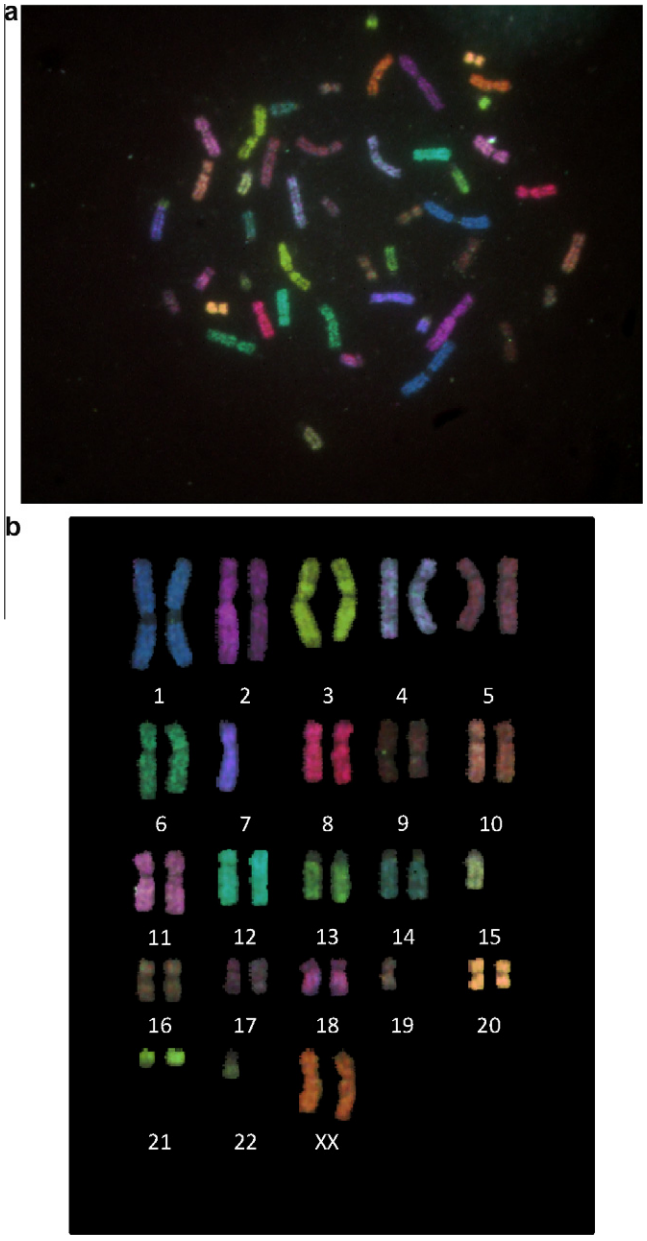


Fig. 1. (a) M-FISH chromosome image of a woman missing three chromosomes from classes 7, 15, 19, and (b) Karyotype of the M-FISH image: 43XX, -7,-15,-19.

chromosome images (Ji, 1989, 1994; Agam and Dinstein, 1997; Ritter and Schreib, 2001). The main idea of these methods is that they split the chromosome groups into segments and then they try to combine these segments into chromosomes. Valley searching techniques (Ji, 1989, 1994) attempt to find a "pale path" of grey values corresponding to a separation between touching–overlapping groups of chromosomes. Initially, all high concavity points (cut-points) are detected along the boundary of chromosomes. Next, a heuristic search is performed to detect the minimum density path between touching chromosomes. The chromosome group is split by the pale path and the segments are combined to form separate chromosomes. Agam and Dinstein (1997) used concave points to construct all the possible separation lines. In their work, they determined potential chromosomes using rectangle hypothesis testing. However this hypothesis does not always hold because of the existence of bended chromosomes that are touching or overlapping to each other and thus a straight line cannot split exactly the chromosomes.

We can conclude that when only the spectral information is used, the segmentation accuracy relies on the pixel-by-pixel classification accuracy. On the contrary, the geometry based methods assume that chromosome shape alone is sufficient for the purpose of separation. Thus both, geometry and spectral information, has to be merged in order to achieve better segmentation results for M-FISH chromosome images.

In this paper we present a novel segmentation method that tackles the problem of touching–overlapping group of chromosomes. Initially, the method uses the watershed transform to segment the DAPI image into watershed regions. The watershed transform has been widely used for the separation of touching/overlapping groups of objects from images (Beucher, 1992; Vincent and Soille, 1991; Malpica et al., 1997; Chen et al., 2003). In our case we propose the recursive application of the watershed transform to each watershed region. However there exist difficult cases of touching as also of overlapping groups of chromosomes that need separation. For this reason we use a geometry method such as the "gradient paths" to split each group of touching–overlapping chromosomes. However we do not compute the gradient paths using the intensity of pixels of the DAPI image, but we propose the computation of paths in the M-FISH image using pixels with high multichannel gradient magnitude values. This computation proves to be more efficient than the computation of the gradient path on the DAPI image since there are cases of touching or overlapping groups of chromosomes where the gradient path on the DAPI image is difficult to compute since the chromosomes are difficult to disentangle. Finally, after path computation, a region adjacency graph is computed and a region merging algorithm is used to merge all regions. In Section 2 the methods are presented describing in detail all the steps. Section 3 presents the results of the methodology. The discussion of the results is also presented in this section. Finally the conclusions of our work are presented in Section 4.

2. Method

The proposed method consists of three stages as it is shown in Fig. 4: (a) the recursive watershed transform computation, (b) the computation of each gradient path and (c) the region merging process. The first stage consists of a number of steps. The first step is the conversion of the initial DAPI chromosome image to binary. In the second step, the Euclidean distance transform of the binary image is computed. The watershed transform is applied in the next step and an initial estimation of the segmented chromosome areas is obtained. The watershed transform is further applied separately to every segmented area until no more new areas are created. The first step of the second stage is the computation of the high concavity points along the boundary of each chromosome area. Next, all gradient paths are computed and the binary chromosome area is split along the gradient path. All gradient paths are computed using the multichannel gradient magnitude. In the final stage a recursive region merging procedure is applied as follows. A region adjacency graph is computed and also each region is classified independently using a region Bayes classifier. Then we merge all neighbouring regions that share the same class. The identification of the overlapping chromosomes takes place in the final step.

2.1. Recursive watershed segmentation

In the first step, the DAPI chromosome image is converted to binary using a well known automated threshold selection process (Otsu, 1979). Using the DAPI channel an initial estimation of the regions of the M-FISH image is produced. The threshold operation at grey level l partitions the pixel values of an image into two classes

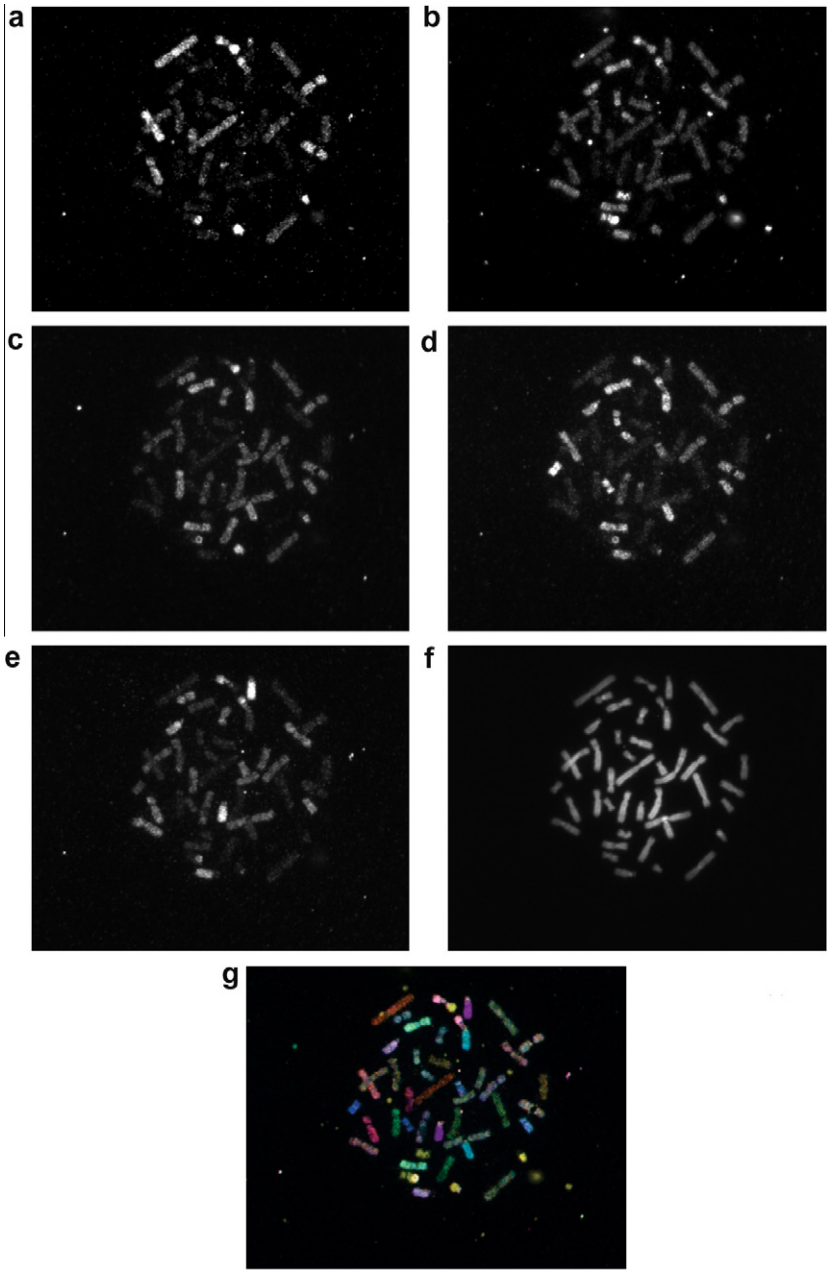


Fig. 2. An M-FISH image and its channels. (a) Cy 5.5 fluorophore, (b) Cy 5 fluorophore, (c) Spectrum Green fluorophore, (d) Spectrum Orange fluorophore, (e) Texas Red fluorophore, (f) DAPI fluorophore, and (e) M-FISH image.

K_0 and K_1 (representing background and object respectively), i.e., $K_0 = \{1, 2, \ldots, l\}$ and $K_1 = \{l+1, l+2, \ldots, L\}$, where L is the total number of grey levels in the image. An optimal threshold l^* can be determined by minimizing the following criterion function:

$$l^* = \arg\min_{l} \sigma_B^2(l),$$

where $\sigma_B^2(l)$ is the between-class variance for the threshold value l (Otsu, 1979).

After the computation of the threshold l^* the binary image B can be computed:

$$B(x,y) = \begin{cases} 0 & \text{if } DAPI(x,y) \leqslant l^*, \\ 1 & \text{if } DAPI(x,y) > l^*. \end{cases}$$

An example of the application of the threshold operation to a DAPI image is shown in Fig. 5.

In order to apply the watershed transform (WT) (Vincent and Soille, 1991) to the image B it is common to first compute the distance transform (DT) (Malpica et al., 1997; Chen et al., 2003). Given an $m \times n$ binary image B, its distance transform is a map that assigns to each on-pixel (p_1) (with coordinates (x_1, y_1)) the distance to the nearest off-pixel (p_2) (with coordinates (x_2, y_2)). The distance metric used is the Euclidean distance $D = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$.

The distance transform provides important information for the application of the watershed algorithm (Beucher, 1992). The number of regional minima of the negative distance transform constitutes indication of the number of areas that will be segmented by the WT. However a common problem is that the distance transform contains a large number of such minima leading the WT to over segment the initial image. On the other hand the greyscale reconstruction (Vincent, 1993) of the negative distance transform suppresses all minima whose depth are lower than or equal to a

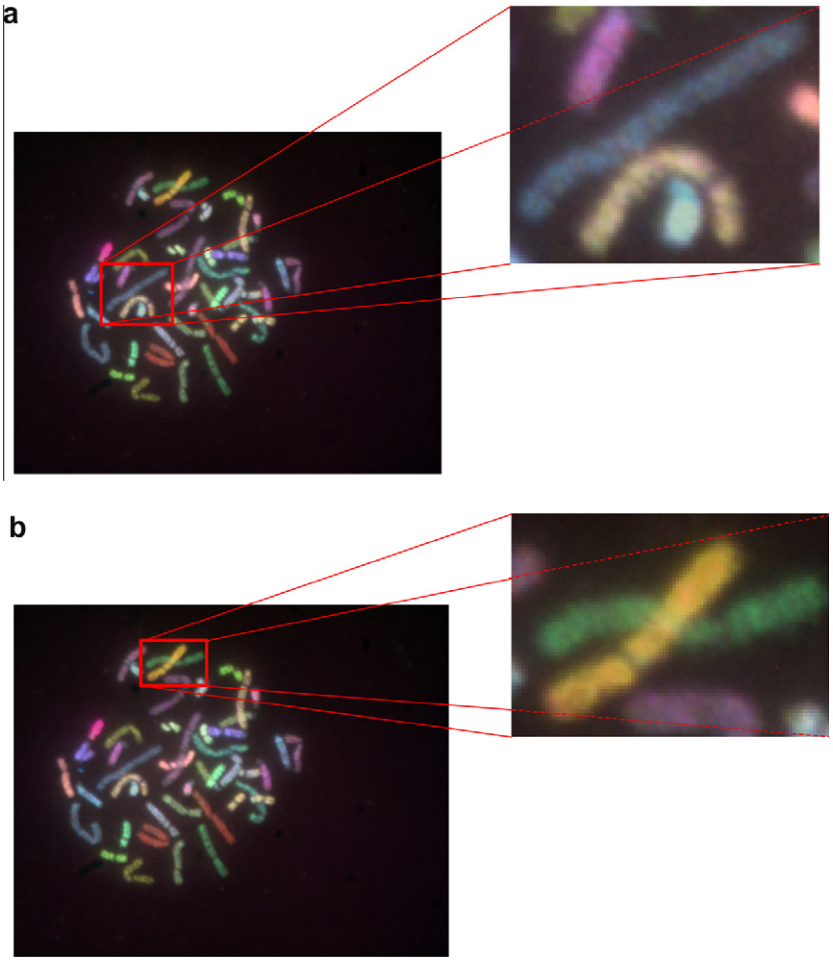


Fig. 3. Touching and overlapping group of chromosomes. (a) Three chromosomes that are touching each other, (b) two chromosomes that overlap.

threshold $h \in \Re$ Thus we apply this procedure in order to alleviate the over-segmentation problem. An alternative for the elimination of the over-segmentation effect could be the Gaussian blur of the gradient image (Gauch, 1999), however the choice of the width of the Gaussian kernel is a key parameter for these approaches.

The next step is the application of the WT. The watershed transform is a popular segmentation method originated in the field of mathematical morphology. The image is considered as a topographical relief, where the height of each point is related to its grey level. Imaginary rain falls on the terrain and water begins to rise filling the different catchment basins. The watersheds are the lines separating the catchment basins that form.

In our case we apply the watershed method using the negative distance transform. The watershed algorithm produces a tessellation of the image into regions, these regions are called watershed regions and depicted in Fig. 6(a) and (b). Whereas several methods start with an over-segmentation of the image and iteratively merge regions based on some measures of similarity (Haris et al., 1998), our method introduces a new region splitting technique based on the watershed transform. All the steps of our method – which do not require any a priori knowledge – are recursively applied to every watershed area until no more new areas are produced. The result of the recursive watershed transform is shown Fig. 6(c)–(f).

2.2. Computation of gradient paths

The idea of paths has been introduced in early 90s (Ji, 1989, 1994) in order to separate touching groups of chromosomes for greyscale images such as the G-banded chromosome images (Sumner et al., 1971). It is based on two assumptions: (a) where chromosomes touch the cluster boundary tends to form an acute angle and (b) at points where chromosomes touch, the optical density is relatively low. The detection of the paths is computed via a search algorithm. The search begins at a cut-point and proceeds in the direction of the normal vector. A cut-point is a boundary point at which the boundary is highly concave. It then proceeds until another boundary point is found as follows: At the current point a list of candidates is found as it is shown in Fig. 7(a). A new trace point is found by choosing the candidate with the smallest intensity value. Finally, the searching direction is updated every d points to allow the path to follow the shape of its trace points, as it is shown in Fig. 7(b). The path that starts from the cut-point and ends to a boundary point was called a pale path.

The pale paths (Ji, 1989, 1994) were used to cut only touching groups of chromosomes without addressing the case of overlapping chromosomes. Moreover, these studies computed the pale paths only for greyscale images. Using a low intensity path the separation of touching chromosomes is feasible, but fails particularly in overlapping cases. Indeed as it is shown in Fig. 8(a) a pale path does not exist for the case of the overlapping group of chromosomes since the intensity of the overlapping region is homogenous and relatively high.

In this work, we propose a modification of the pale path approach in order to achieve separation of touching and overlapping chromosome groups in coloured M-FISH images. This modification uses the multichannel gradient of the M-FISH image (Karvelis et al.,

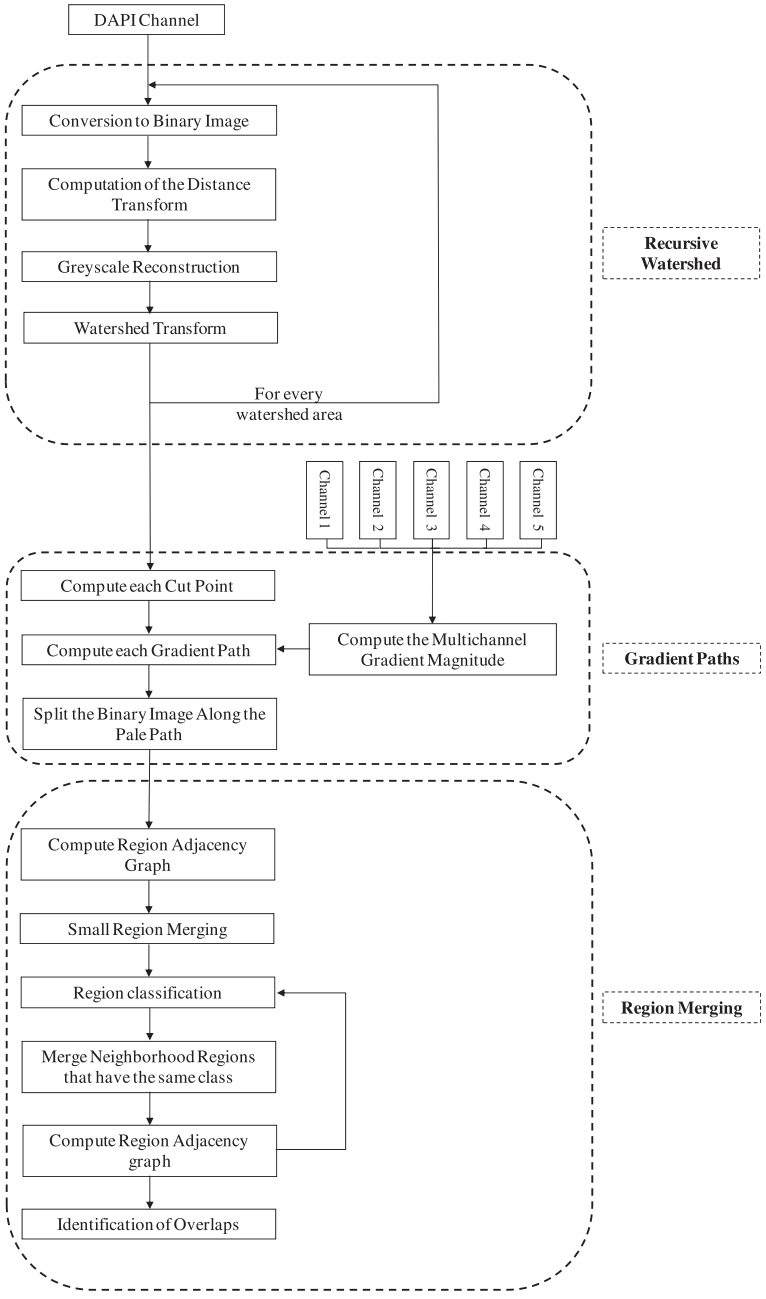


Fig. 4. Flowchart of our method.

2007; Drewniok, 1994). The basic idea is the following: instead of leading the path to follow low intensities pixels, the path now follows pixels of high multichannel gradient magnitude values. The computation of the multichannel gradient magnitude is based on the five channel coloured M-FISH image. This gives the advantage that the paths follow high gradient magnitude pixel values and these high values occur when chromosomes touch or overlap. The path that begins from a cut-point and follows pixels of high gradient magnitude values of the M-FISH image until it reaches a boundary point is now called a gradient path.

To compute the cut-points we first extract the boundary from the binary image B. Suppose that the pixels of the boundary of a segmented region define the set $(c_1, c_2, \ldots, c_{PB})$ where c_i, c_{i+1} are successive points of the boundary and PB the number of pixels of the region boundary. In order to compute the cut-points we compute the curvature of the boundary (Ji, 1989, 1994; Ritter and Schreib, 2001) since local maxima of the curvature indicate candidate positions of the cut-points. For each point of the boundary (c_i: $i = 1, \ldots, PB$) we consider the triangle that is defined from the three points c_{i-k}, c_i, c_{i+k} ($k = 3$) and compute the angle $a(i)$ defined by the triangle:

$$a(i) = \arccos\left(\frac{(c_i - c_{i-k}) \cdot (c_{i+k} - c_i)}{\|c_i - c_{i-k}\| \cdot \|c_{i+k} - c_i\|}\right) \cdot \operatorname{sgn}[\det(c_i - c_{i-k} \quad c_{i+k} - c_i)].$$

In Fig. 8, we demonstrate the steps for the computation of the cut-points in a group of touching and overlapping chromosomes. After the binarization of the chromosome group (Fig. 8(a) and (b)) the curvature of the boundary points (Fig. 8(c)) is computed and is illustrated in Fig. 8(d). All the cut-points are automatically computed by choosing the boundary points that exceed an angle threshold: $a(i) \geqslant 210°$, $i = 1, \ldots, PB$. The red points in Fig. 8(d)

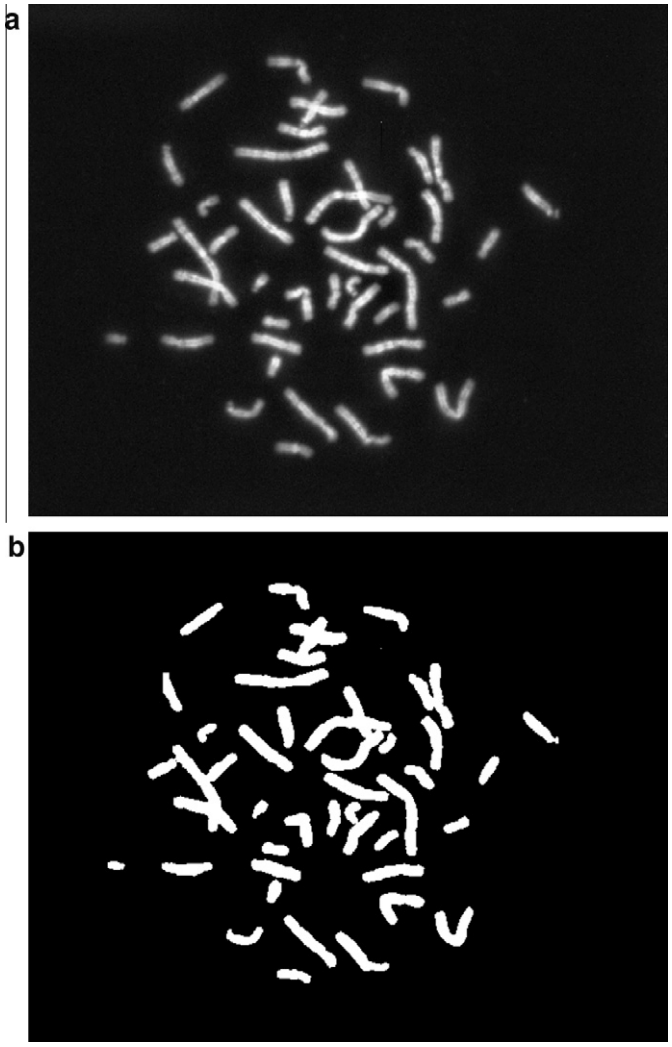


Fig. 5. The thresholding procedure for a greyscale DAPI image. (a) The DAPI image and (b) the binary image.

and (e) illustrate the cut-points that exceed this angle threshold. As we observe in Fig. 8(e), several candidate cut-points are computed. To overcome this problem, the neighbouring candidate cut-points are automatically grouped and from each group of candidate cut-points we extract the one that has the maximum angle as it is illustrated in Fig. 8(f).

The next step of our method is the computation of the multichannel gradient magnitude, as the gradient path will follow pixels having high multichannel gradient values. The multichannel gradient magnitude is computed as follows. Assume a multichannel image $I(x,y): \mathbb{Z}^2 \to \Re^m$ for an M-FISH image and a direction n defined by the angle φ:

$$I(x,y) = \begin{bmatrix} I_1(x,y) \\ I_1(x,y) \\ \vdots \\ I_m(x,y) \end{bmatrix},$$

$$n = \begin{bmatrix} \cos\varphi \\ \sin\varphi \end{bmatrix},$$

where $I_i(x,y)$, $i = 1,\ldots,5$ are the components (channels) of the M-FISH image.

The directional derivative of $I(x,y)$ for a direction n is computed as follows:

$$\frac{\partial I}{\partial n} = \begin{bmatrix} \nabla I_1 \cdot n \\ \vdots \\ \nabla I_m \cdot n \end{bmatrix} = \begin{bmatrix} I_1^x & I_1^y \\ \vdots & \vdots \\ I_m^x & I_m^y \end{bmatrix} = J \cdot n,$$

where I_i^x and $I_i^y 1 \leqslant i \leqslant m$ are the derivatives of the ith component in the x and y direction respectively.

Next the direction n which corresponds to the maximum of the directional derivative is found by maximizing the Euclidean norm $\|J \cdot n\|^2 = (J \cdot n)^T (J \cdot n) = n^T (J^T J) n$.

The maximum of this norm $n^T(J^TJ)n$ is given by the maximum eigenvalue of the matrix (J^TJ) (Karvelis et al., 2008). The symmetric matrix (J^TJ) can be written as:

$$J^T J = \begin{bmatrix} \sum_{i=1}^{m} (I_i^x)^2 & \sum_{i=1}^{m} I_i^x \cdot I_i^y \\ \sum_{i=1}^{m} I_i^x \cdot I_i^y & \sum_{i=1}^{m} (I_i^y)^2 \end{bmatrix},$$

where finite differences (e.g. Sobel operators (Karvelis et al., 2008)) can be used to compute the directional derivatives I_i^x and I_i^y $1 \leqslant i \leqslant m$ in the x and y directions respectively.

Next we proceed to compute the gradient path. The initial direction of the gradient path is set as the bisector of the angle $a(i) = \angle c_{i-k}, c_i, c_{i+k}$ at the starting points c_{i-k}, c_i, c_{i+k}, as it is shown in Fig. 9(a) and (b), where c_i is the initial cut-point. The computation of the gradient path proceeds as follows: we choose from the pixel-candidates the one that has the maximum gradient value. We then proceed to the next pixel updating the current search direction every $d = 3$ points until we reach a boundary point. Finally, we delete points, of the binary image, along the gradient path. We present the computation of the gradient path for a touching–overlapping chromosome group in Fig. 9(c) and (d) and in Fig. 9(e) the final regions produced from the binary image by cutting along the gradient paths.

2.3. Region merging

The purpose of this stage is to connect regions that have been split by gradient paths. In our case we call a region small if it contains less than 25 pixels. This step was implemented by computing for each region of the binary image, the Region Adjacency Graph (Haris et al., 1998) (RAG), where two nodes (representing two distinct regions) are connected if the corresponding regions are adjacent. An example of a RAG is shown in Fig. 10.

Then a Region Bayes Classifier (RBC) (Karvelis et al., 2007; Landgrebe, 1980) was employed in order to classify all the regions of the watershed area as follows. Suppose a region $R = \{z_1, z_2, \ldots, z_w\}$ consists of w pixels. Each vector $z_i \in \Re^5$, $i = 1,\ldots,w$ measures the intensity of each of the five M-FISH channels. Also let ω_j, $j = 1,\ldots,N_C$ the chromosome classes, where $N_C = 24$.

The likelihood of the region R is computed as (Landgrebe, 1980):

$$p(R|\omega_i) = p(z_1, z_2, \ldots, z_w|\omega_j) = \prod_{i=1}^{w} p(z_i|\omega_j) = \left(\frac{1}{(2\pi)^{5/2}|\Sigma_j|^{1/2}}\right)^w \exp\left(-\frac{1}{2}\sum_{k=1}^{w}(z_k-\mu_j)^t \Sigma_j^{-1}(z_k-\mu_j)\right),$$

where μ_j is the five-component mean vector of class j, Σ_j is the 5×5 covariance matrix of class j, $|\Sigma_j|$ and Σ_j^{-1} are the determinant and inverse respectively. The mean vectors and the covariance matrixes for each class are computed by a training phase, from an annotated set of M-FISH images as follows:

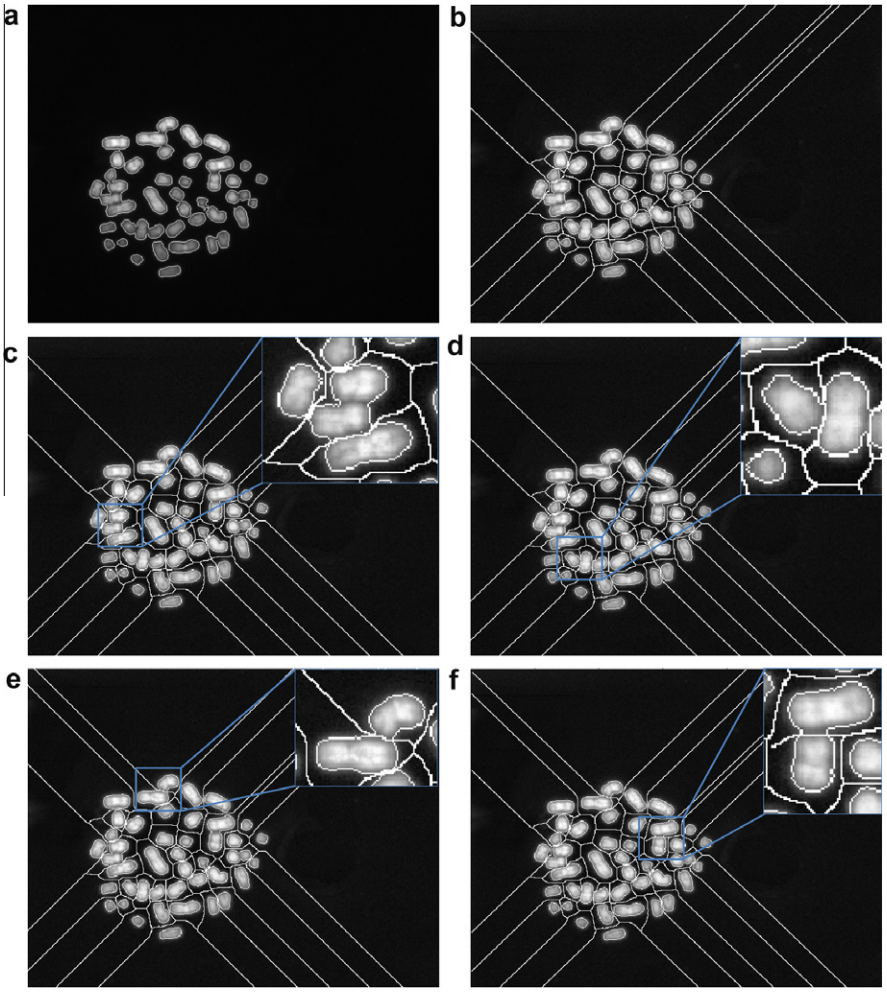


Fig. 6. An example of the application of the recursive watershed transform for a DAPI chromosome image. (a) DAPI image, (b) 1 iteration, (c) 2 iteration, (d) 3 iteration (e) 4 iteration and (f) 5 iteration.

$$\mu_j = \frac{1}{N_j} \sum_{k=1, z_j \in \omega_j}^{N_j} z_k, \quad j = 1, \ldots, N_C,$$

$$C_j = \frac{1}{N_j - 1} \sum_{k=1, z_j \in \omega_j}^{N_j} (z_k - \mu_j) \cdot (z_k - \mu_j)^T, \quad j = 1, \ldots, N_C,$$

where N_j the number of pixels of the jth class.

Using Bayes theorem (Karvelis et al., 2007; Fukunaga, 1990):

$$P(\omega_j|R) = \frac{p(R|\omega_j) \cdot P(\omega_j)}{\sum_{j=1}^{24} p(R|\omega_j) \cdot P(\omega_j)},$$

we can classify a region R to class j having maximum a posterior probability (Fukunaga, 1990):

$$P(\omega_j|R) > P(\omega_q|R) \quad \forall q = 1, \ldots, 24, \; q \neq j.$$

The prior probabilities are estimated from the same set of annotated images as follows:

$$P(\omega_q) = \frac{N_q}{\sum_{i=1}^{24} N_q}, \quad q = 1, \ldots, N_C.$$

For each small region R:(# *pixels of* $R \leqslant 25$) of the watershed area we use the RAG to select the neighbours N_R of R: $N_R = \{R_1, R_2, \ldots, R_k\}$. Let ω_i the class of region $R_i \in N_R$, $i = 1, \ldots, k$ and

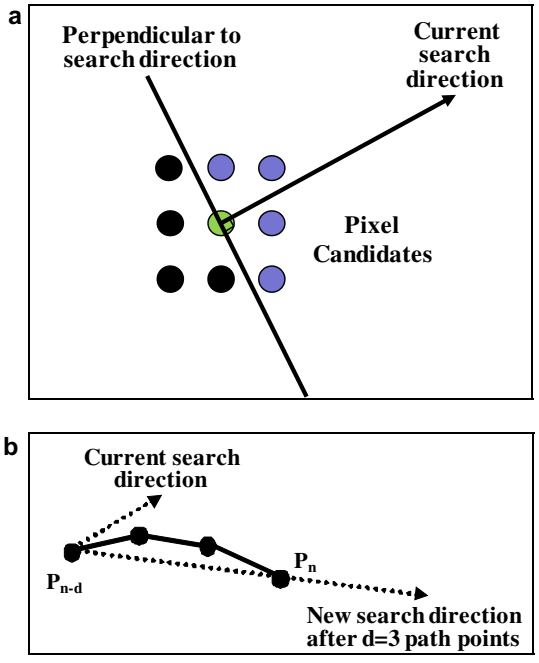


Fig. 7. Pale path computation: (a) candidates for the next path point and (b) update of path's direction after $d = 3$ points.

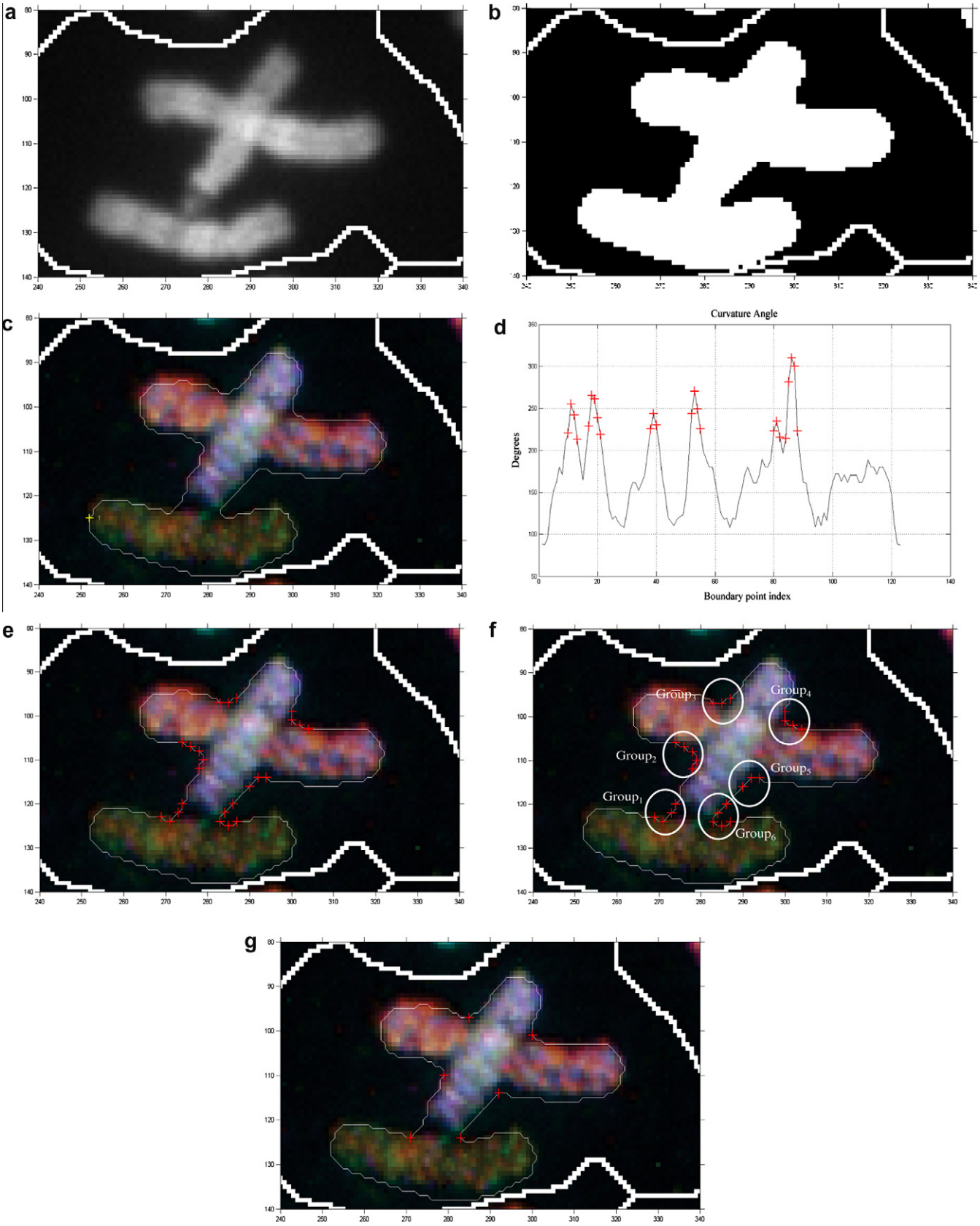


Fig. 8. The computation of the cut-points for a touching–overlapping group of chromosomes. (a) The DAPI watershed area, (b) the binary image, (c) the boundary of the group of chromosomes overimposed on the M-FISH image (the yellow point depicts the first boundary point), (d) the curvature along the boundary points with red points are depicted the cut-points that exceed the angle threshold, (e) the cut-points (red points) overimposed on the M-FISH image, (f) the groups of the candidate cut-points and (g) the final computed cut-points. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

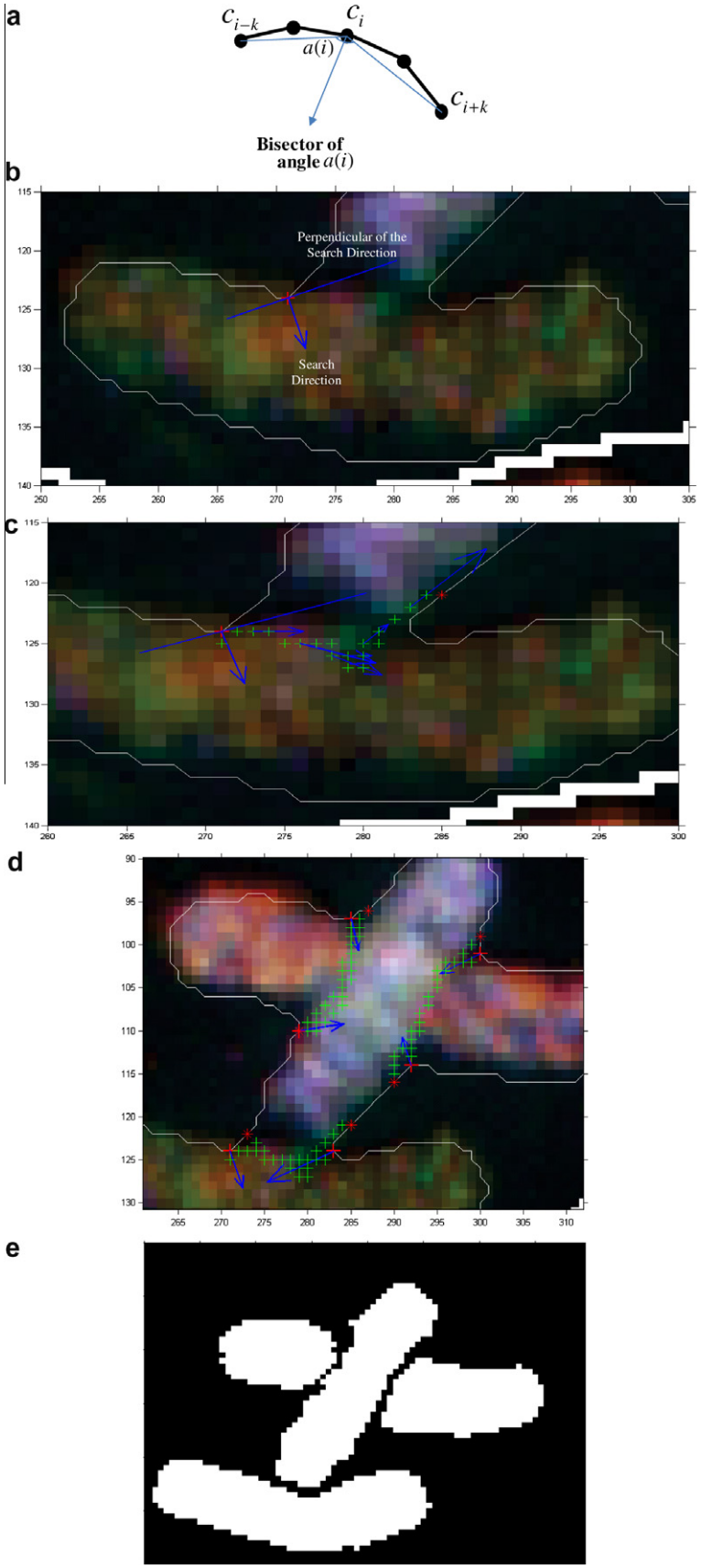


Fig. 9. The computation of the gradient paths. (a) The initial search direction and the perpendicular of the search direction, (b) a gradient path reaching the other side of the boundary (the green points depict the points of the gradient path), (c) all the gradient paths computed for all the cut-points of the chromosome group, and (d) the binary image after the binary image has been cut by the gradient paths.

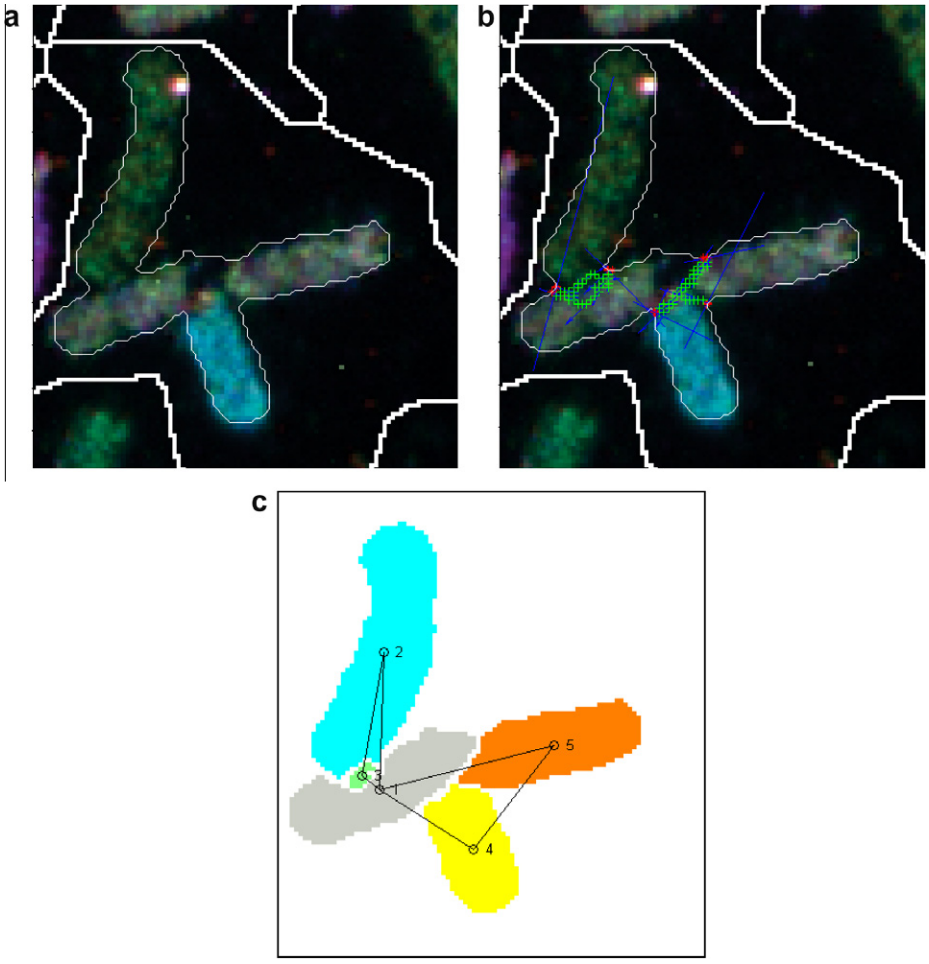
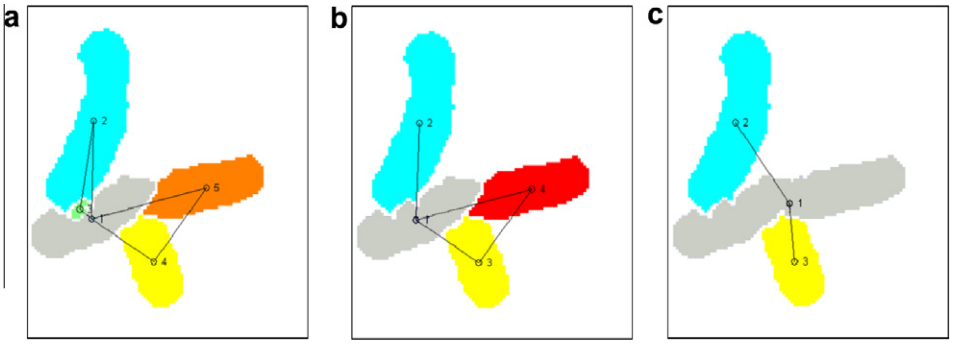


Fig. 10. The Region Adjacency Graph after the gradient paths split the binary image. (a) The M-FISH image for a touching group of chromosomes (the thin white line depicts the boundary of the binary image), (b) the gradient paths overimposed on the M-FISH image, and (c) the Region Adjacency Graph after the gradient paths split the chromosomes.



#Region	Class
R_1	ω_1
R_2	ω_4
R_3	ω_{16}
R_4	ω_{13}
R_5	ω_1

#Region	Class
R_1	ω_1
R_2	ω_4
R_3	ω_{13}
R_4	ω_1

#Region	Class
R_1	ω_1
R_2	ω_4
R_3	ω_1

Fig. 11. Region merging of the binary image. (a) Initial region adjacency graph and the classes of each of the regions of the binary image: the small region R_3 is merged with region R_1, since the posterior probability $P(\omega_1|R_3) \geqslant P(\omega_i|R_3)$, $i = \{1, 4\}$, (b) the region R_1 and R_4 are merged since they share the same class ω_1 and (c) the final region merging result.

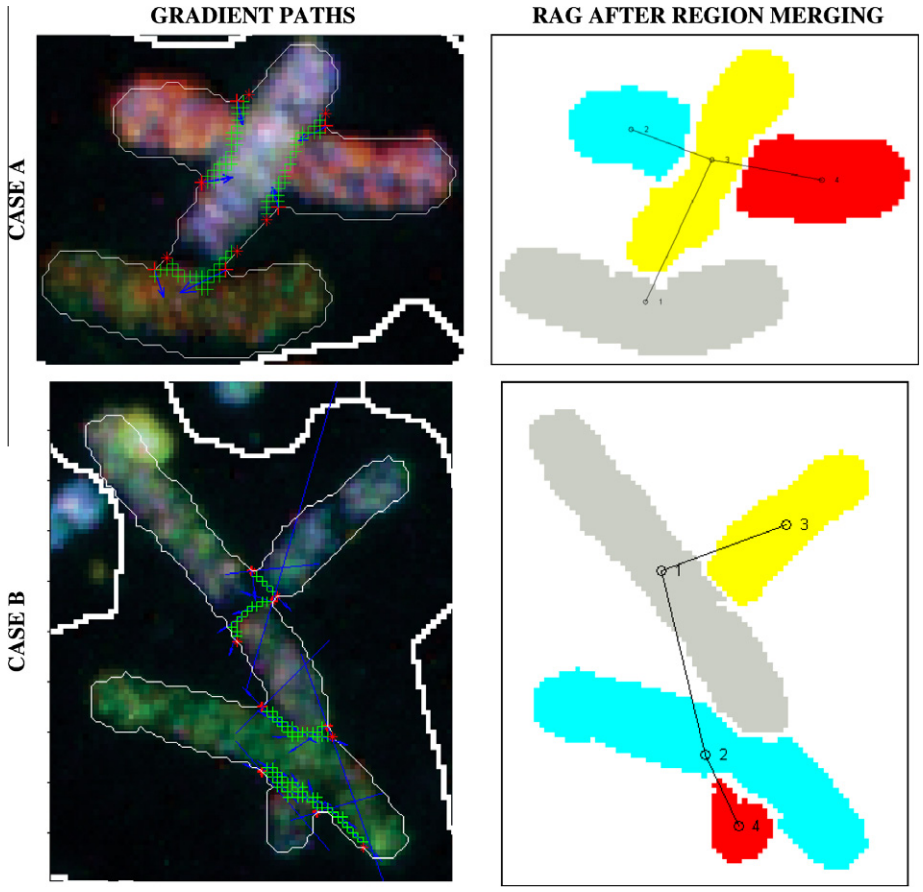


Fig. 12. Two overlapping chromosome cases. The proposed method identifies them correctly. CASE A: Regions R_2 and R_4 are identified as an overlap since the region R_3 is connected with two regions of the class ω_{23}. CASE B: Regions R_1 and R_4 are identified as an overlap since the region R_2 is connected with two regions of the class ω_1.

Table 1
Number of touching and overlapping chromosomes in the M-FISH database.

	Touching chromosomes	Overlapping chromosomes
Total number	1178	189

$\Omega_R = \{\omega_1, \omega_2, \ldots, \omega_i, \ldots, \omega_k\}$ the set of these classes. Then we compute the posterior probabilities $P(\omega_i|R)$ of region R for $\omega_i \in \Omega_R$ and region R is merged to region R_j whose class ω_j has maximum posterior: $\Omega_R = \{\omega_1, \omega_2, \ldots, \omega_i, \ldots, \omega_k\}$

$$P(\omega_j|R) \geq P(\omega_i|R) \quad \forall \omega_i \in \Omega_R.$$

An example of small region merging is shown in Fig. 11(a) and (b).

The next step of our method is to classify all the regions (including the merged ones) of the binary image using the Region Bayes Classifier described previously. Then all regions that are adjacent and share the same class are connected. Finally, the RAG is computed and the procedure is repeated until no more regions are connected.

The final step of our method is the identification of the overlapping chromosomes. The key idea in this step is that when two chromosomes overlap, a cross shaped object is formed. In this case our method splits the binary image in way that one region of the image separates two regions that share the same class. This is illustrated in Fig. 12. Thus the final step of our method is to identify these overlapping cases by checking for each region of the binary image whether it has two neighbouring regions of the same class.

Table 2
Comparison of our work with other works presented in the literature for the touching group of chromosomes.

		Schwartzkopf et al. (2005)	Ji (1989)	The proposed method
Separation accuracy	Accuracy (%)	77	84.2	90.6
Dataset description	#Images	183	183	183
	#Touches	720	1178	1178

3. Results and discussion

3.1. Dataset

To validate our method we used the ADIR M-FISH database (Schwartzkopf et al., 2005) which is a database of M-FISH images. The database consists of 183 multispectral M-FISH images. Each image has 517 × 645 pixels of 8-bit resolution. The database contains five-channel image sets recorded at different wavelengths. In addition, a DAPI image file is included for each M-FISH image. The dataset includes also a classification map (karyotype image), stored as an image file established by experienced cytogeneticists. This image is labelled so that the grey level of each pixel represents its class number (chromosome class). In addition, background pixels are 0, and pixels in a region of overlap are −1. This data file serves as "ground truth" to test the accuracy of our method.

As a ground truth for the touching chromosomes, we used the binary image produced by the DAPI image to identify the cases

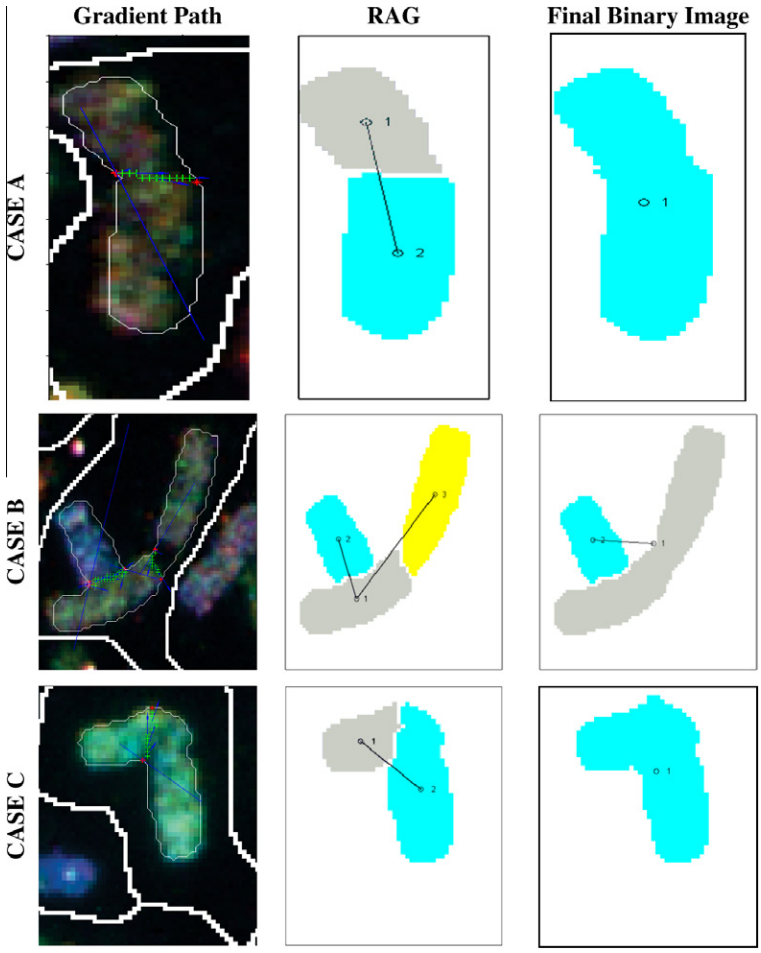


Fig. 13. Examples of three different cases of bended chromosomes. The proposed method handles them successfully after the region merging stage.

of touching chromosomes in an M-FISH image. For each object produced by the binarization procedure we determined the cases of touching. Finally as a ground truth for the cases of overlapping we used the characterized karyotype image of the M-FISH database since an overlapping region is represented in that image by pixels having the value of −1. The number of touches and overlaps in the M-FISH database is shown in Table 1.

3.2. Touching chromosomes

The separation accuracy for the touching group of chromosomes was measured by our method. A correct separation occurs when two or more touching chromosomes are segmented correctly. The results of our method for the touching groups of chromosomes are shown in Table 2. We have also compared our method with the method of pale paths (Ji, 1989) for the touching groups of chromosomes as the method of pale paths cannot handle overlapping cases. In order to compute the pale path we have used the DAPI image since the pale path uses a greyscale image.

It is interesting to mention the robust behaviour of our method in the case of isolated bended chromosomes. It is common in the M-FISH chromosome database to find cases where isolated chromosomes bend, as shown in Fig. 13. For these cases, cut-points are found and gradient paths begin to split the chromosome into two regions. However the region merging stage merges these regions to form one chromosome again.

Table 3
Comparison of our work with other works presented in the literature for the overlapping group of chromosomes.

		Schwartzkopf et al. (2005)	The proposed method
Separation accuracy	Accuracy (%)	34	80.4
Dataset description	#Images	183	183
	#Overlaps	189	189

3.3. Overlapping chromosomes

The separation accuracy for the overlapping group of chromosomes is also measured. The results of our method for the overlapping groups of chromosomes are shown in Table 3.

We have described a novel method for the separation of touching and overlapping groups of M-FISH chromosome images. Our method is based on the recursive application of the watershed transform and the computation of gradient paths for each watershed area. A region merging stage is finally applied to merge regions that have been wrongly split by the gradient paths. Our method is evaluated using an M-FISH chromosome image database and an overall separation accuracy of 90.6% and 80.4% for the touching and overlapping groups of chromosomes respectively has been found.

In fact, only one method has been presented in the literature for the separation of M-FISH images testing its ability to separate

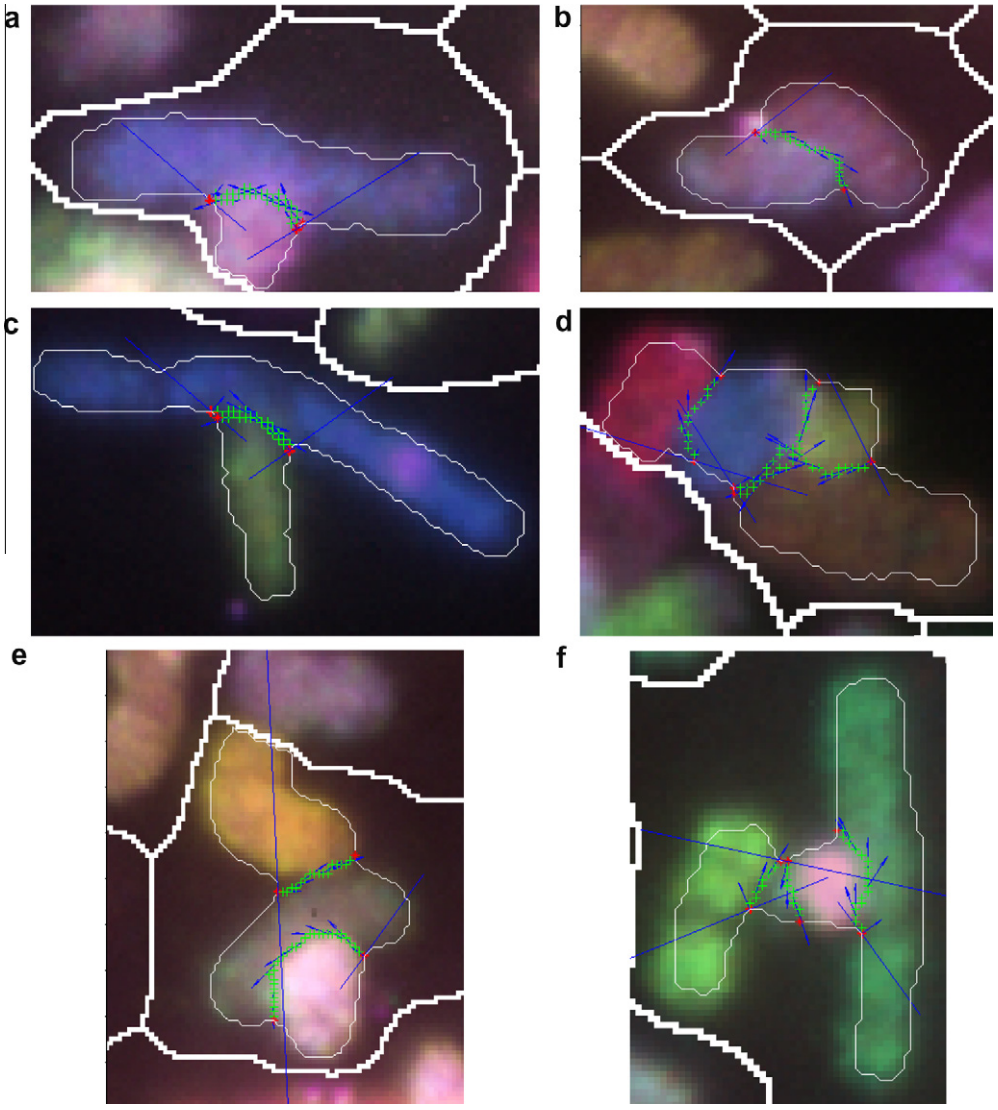


Fig. 14. Six examples of touching and overlapping groups of chromosomes where straight lines between cut-points cannot separate correctly the chromosomes.

touching and overlapping groups of chromosomes for the whole M-FISH database (Schwartzkopf et al., 2005). Our method uses the information from all the channels (the five channel M-FISH image including the DAPI image) whereas Schwartzkopf et al. (2005) use only the information provided by the five channel M-FISH image.

To best of our knowledge the pale paths were able to separate only touching groups of chromosomes without handling overlapping chromosomes (Ji, 1989, 1994). In this paper, we expand the idea of the paths in order to address also the case of overlapping groups. More specifically we introduce the gradient paths which more effectively segment not only touching but also overlapping groups of chromosomes for the M-FISH images. The gradient path is superior to other proposed splitting techniques for two reasons:

(A) Unlike other methods (Agam and Dinstein, 1997), we do not assume that a path is a straight line between two cut-points. Fig. 14 depicts some examples of touching and overlapping chromosome groups, none of which can be split by a straight line without fragmenting a chromosome. Such cases usually happen where more than two chromosomes are involved in a group of touching–overlapping or one of the chromosomes is bent.

(B) The paths have been appropriately modified in order to separate overlapping and touching groups of chromosomes in M-FISH images by computing the multichannel gradient and the running of the path through high gradient magnitude values. Fig. 15 illustrates some examples of the computation of the pale path (Ji, 1989, 1994) using the DAPI image versus the gradient path which uses the multichannel gradient magnitude of the M-FISH image.

While previous methods were based only on region merging techniques (Karvelis et al., 2008), the proposed method utilizes region merging and region splitting techniques to correctly segment the chromosome touching and/or overlapping groups. More specifically the proposed method starts with an initial number of segmented regions and recursively split or merge the regions until no more regions are produced. Fig. 13 is an indicative example of how the region merging step corrects erroneous separations of the region splitting step. Although, the region splitting technique separated the chromosome into two distinct regions, region merging corrected this error by merging the regions again to form one chromosome.

There is a tradeoff between the choice of this angle threshold and the number of cut-points. If we choose a smaller threshold a

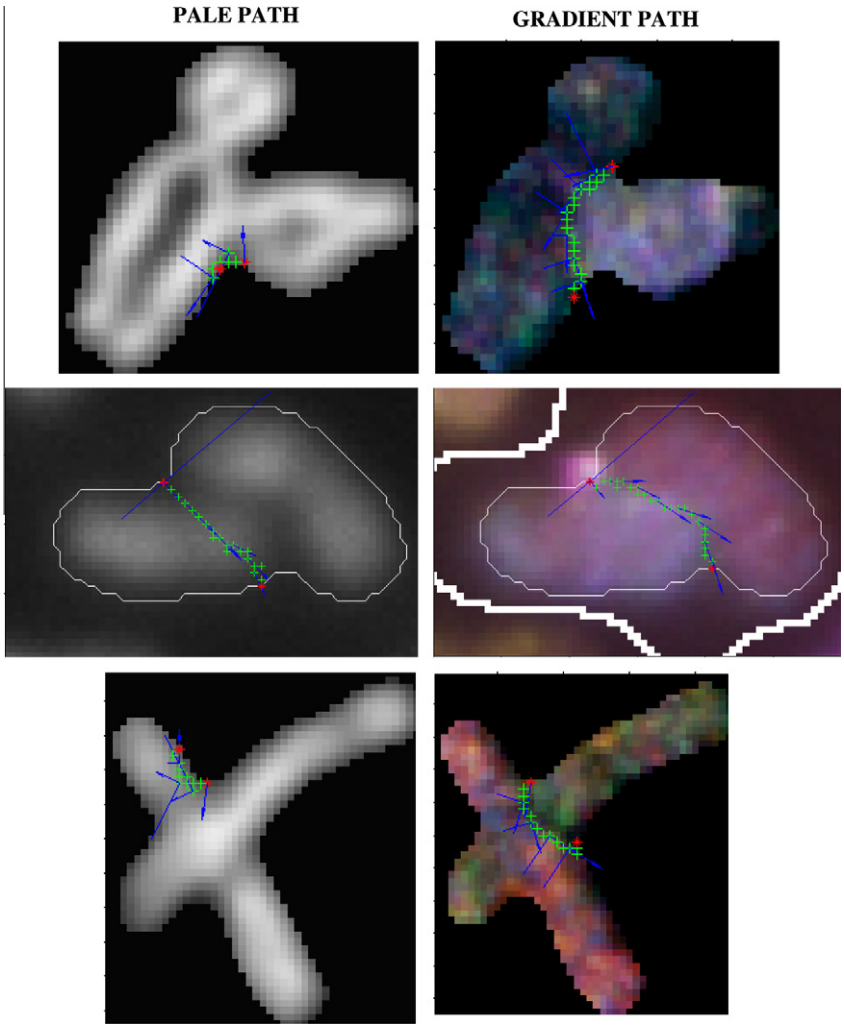


Fig. 15. Pale path versus gradient path. Three examples where the pale path fails to separate correctly the chromosomes while the gradient path correctly separates them.

larger number of cut-points is computed. However, the grouping of neighbouring cut-points and the selection of those with the maximum angle (among their neighbours) result in the same final cut-points. On the other hand the choice of a larger angle threshold will result to a smaller number of cut-points and thus less gradient paths will be computed. Thus, the separation of the touching or overlapping group of chromosomes will not be feasible. In our experiments we have heuristically determined the best value for this parameter.

Tables 2 and 3 show a comparison between the proposed study and the method which proposed by Schwartzkopf et al. (2005). While the number of overlapping cases is the same, the number of touches differs between the two methods. This is done due to the different approaches employed by the two methods for the determination of touching groups of chromosomes. In our case we used the binary image produced by the DAPI image to identify the number of touching chromosomes in an M-FISH image whereas Schwartzkopf et al. (2005) has manually chosen the number of touches. In general, it is difficult to compare the two methods directly since they are not handling the same number of touching chromosomes. However our method is employed in the same M-FISH database and the number of touches is higher than that reported in Schwartzkopf et al. (2005). We have also compared our method with the method of pale paths (Ji, 1989, 1994) for the touching groups of chromosomes as the method of pale paths cannot handle overlapping cases. Touching and/or overlapping chromosome groups are highly likeable to appear in a chromosome image making time consuming and difficult the effort of the cytogeneticist. Therefore, the efficient treatment of this problem is very important for the biologists. Since our method provides sufficiently accurate results in identifying touching and overlapping chromosomes, it is expected to be a very useful tool that will greatly facilitate the task of the cytogeneticist.

4. Conclusions

A method for the separation of touching and overlapping groups of chromosomes in M-FISH images was presented. The method is based on the recursive application of the watershed transform and the splitting of groups of chromosomes by multichannel gradient paths. These paths split these groups of chromosomes from points of high concavity and follow pixels of high multichannel gradient magnitude. Then a region merging stage is applied to merge neighbouring regions producing the final chromosome areas. The computation of the gradient path based on multichannel gradient values instead of the computation of the pale path which is based on low intensity values proves to be more robust. Also the gradient paths can now handle overlapping cases of chromosomes whereas the pale paths could only handle touching groups of chromosomes. As for future work an unsupervised method will be developed for the classification of M-FISH images. In this way we

avoid the use of already characterized M-FISH images in order to train the Bayes classifier.

References

Agam, G., Dinstein, I., 1997. Geometric separation of partially overlapping non rigid objects applied to automatic chromosome classification. IEEE Trans. Pattern Anal. Machine Intell. 19, 1212–1222.

Beucher, S., 1992. The watershed transformation applied to image segmentation. Scanning Microscope. 6, 299–314.

Chen, Q., Yang, X., Petriu, E., 2003. Watershed segmentation for binary images with different distance transforms. In: 3rd IEEE Internat. Workshop on Haptic, Audio and Visual Environments and their Applications, 20–21 September 2003, Ottawa, Canada.

Drewniok, C., 1994. Multi-spectral edge detection – Some experiments on data from Landsat-TM. Internat. J. Remote Sens. 15, 3743–3765.

Fukunaga, K., 1990. Introduction to Statistical Pattern Recognition. Academic Press, San Diego.

Gauch, J., 1999. Image segmentation and analysis via multiscale gradient watershed hierarchies. IEEE Trans. Image Process. 8 (1), 69–79.

Haris, K., Efstratiadis, S., Maglaveras, N., Katsaggelos, A., 1998. Hybrid image segmentation using watersheds and fast region merging. IEEE Trans. Image Process. 7 (12), 1684–1699.

Ji, L., 1989. Intelligent splitting in the chromosome domain. Pattern Recognition 22 (5), 519–532.

Ji, L., 1994. Fully automatic chromosome segmentation. Cytometry 17, 196–208.

Karvelis, P., Tzallas, A., Fotiadis, D., Georgiou, I., 2008. A multichannel watershed-based segmentation method for multispectral chromosome classification. IEEE Trans. Med. Imaging 27 (5), 697–708.

Karvelis, P., Tzallas, A., Fotiadis, D., Georgiou, I., 2007. Region based segmentation and classification of multispectral chromosome images. In: 20th IEEE International Symposium on Computer-Based Medical Systems (CBMS), 20–27 June 2007, Maribor, Slovenia.

Landgrebe, D., 1980. The development of a spectral-spatial classifier for earth observational data. Pattern Recognition 12 (3), 165–175.

Malpica, N., Solorzano, C., Vaquero, J., Santos, A., Vallcorba, I., Sagredo, J., Pozo, F., 1997. Applying watershed algorithms to the segmentation of clustered nuclei. Cytometry 28, 289–297.

Otsu, N., 1979. A threshold selection method for gray-levels histograms. IEEE Trans. Pattern Anal. Machine Intell. 13, 583–598.

Ritter, G., Schreib, G., 2001. Using dominant points and variants for profile extraction from chromosomes. Pattern Recognition 34, 923–938.

Sampat, M., Bovik, A., Aggarwal, J., Castleman, K., 2005. Supervised parametric and non-parametric classification of chromosome images. Pattern Recognition 38, 1209–1223.

Schwartzkopf, W., Bovik, A., Evans, B., 2005. Maximum-likelihood techniques for joint segmentation–classification of multispectral chromosome images. IEEE Trans. Med. Imaging 24, 1593–1610.

Speicher, M., Ballard, S., Ward, D., 1996. Karyotyping human chromosomes by combinatorial multi-fluor FISH. Nature Genet. 12, 368–375.

Sumner, T., Evans, J., Buckland, A., 1971. A new technique for distinguishing between human chromosomes. Nature New Biol. 232, 31–32.

Thompson, M., McInnes, R., Willard, H., 1991. Genetics in Medicine. Saunders, Ontario.

Veldman, T., Vignon, C., Schröck, E., Rowley, J., Ried, T., 1997. Hidden chromosome abnormalities in hematological malignancies detected by multicolor spectral karyotyping. Nature Genet. 15, 406–410.

Vincent, L., Soille, P., 1991. Watershed in digital spaces: An efficient algorithm based on immersion simulations. IEEE Trans. Pattern Anal. Machine Intell. 13, 583–598.

Vincent, L., 1993. Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms. IEEE Trans. Image Process. 2, 176–201.

Wang, Y., Castleman, K., 2005. Normalization of multicolour fluorescence in situ hybridization (M-FISH) images for improving colour karyotyping. Cytometry 64, 101–109.

Wang, Y., Dandpat, A., 2006. A hybrid approach of using wavelets and fuzzy clustering for classifying multispectral florescence in situ hybridization images. Int. J. Biomed. Imaging, 1–11.