# SEGMENTATION OF CHROMOSOME IMAGES BASED ON A RECURSIVE WATERSHED TRANSFORM

P.S. Karvelis*, D.I. Fotiadis*, **, M. Syrrou*** and I. Georgiou****

* Unit of Medical Technology and Intelligent Information Systems / Dept. of Computer Science, University of Ioannina, GR 451 10 Ioannina, Greece
** Biomedical Research Institute-FORTH, GR 451 10 Ioannina, Greece
*** Department of General Biology / Medical School, University of Ioannina, GR 451 10 Ioannina, Greece
**** Lab of Human Reproductive Genetics / Medical School, University of Ioannina, GR 451 10 Ioannina, Greece

pkarvel@cs.uoi.gr, fotiadis@cs.uoi.gr, msyrrou@cc.uoi.gr, igeorgiou@cc.uoi.gr

**Abstract: In this study, a method for the segmentation of touching groups of chromosomes is presented. In the first step an initial segmentation of the image is produced using the watershed transform. Next, every segmented area is further processed using the watershed transform and the procedure is recursively applied until no more new areas are produced. Finally, all paths starting from points of high concavity are computed and the final segmented chromosome areas are produced. In order to validate our method a dataset consisting of 940 chromosomes is used. From these, 515 were isolated chromosomes, 396 were touching and 29 were overlapping. The proposed method resulted in a success rate 92% for the touching chromosomes and 100% for the isolated chromosomes.**

## Introduction

Chromosome analysis is an essential procedure for detecting genetic abnormalities. Traditionally, cells are classified according to their karyotype, which is a tabular array in which the chromosomes are alligned in pairs (Fig. 1). Karyotyping is a useful tool to detect deviations from normal cell structure. Abnormal cells may have an excess or a deficit of chromosomes and/or structural defects which depict an exchange of genetic material [1].

Manual classification of human chromosomes is a slow and tedious procedure and much attention has been paid to make it automatic [2]. Most of the automated classification systems consist of the following stages: (i) image acquisition, (ii) image segmentation, (iii) feature extraction of segmented areas and (iv) classification of chromosomes.

The reason that prevents image segmentation from being a completely automatic procedure is its inability to automatically segment clusters of touching and overlapping chromosomes [3]. In chromosome analysis it is essential to segment the image into background and objects of interest. These objects often consist of two or more chromosomes, either touching or overlapping with each other. Therefore, these objects must be further divided into the consisting chromosomes.

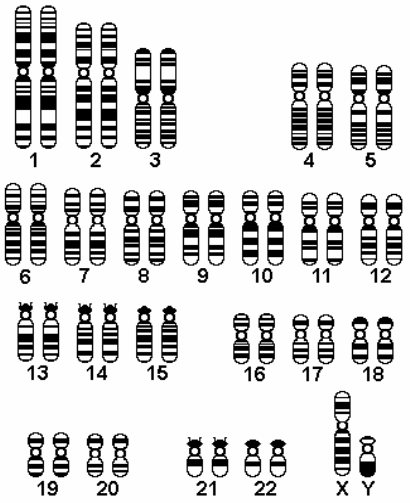



Figure 1: Schematic representation of chromosome classes.

Several methods have been proposed which tackle the problem of disentangling touching or overlapping chromosomes [2-13]. There are studies addressing only the touching/overlapping groups of chromosomes. Methods based on a threshold for segmenting chromosome pixels has been proposed [4-5]. Although, these methods are quite simple they cannot deal effectively with groups of chromosomes. Thus, their usage is limited to chromosome analysis.

Valley searching techniques [6-7] attempt to find a "valley" of grey values corresponding to a separation between touching groups of chromosomes. Initially, all high concavity points (cut-points) are detected along the boundary of chromosomes. Next, a heuristic searching is performed to detect the minimum density path between touching chromosomes. Although these methods report a high rate (90-95%) for segmenting touching and overlapping groups of chromosomes they make use of a large number of parameters. Which have to be tuned before the application of the method.


ISSN: 1727-1983 

Several works on the segmentation of touching chromosomes are based on connecting two adjacent cut-points [3]. The same approach has also been proposed for the overlapping chromosomes [9-11]. However, this approach depends on heuristics and thresholds and makes use of a multistage complex procedure.

Although the above methods have achieved satisfactory results improvements can be made. Most of these methods either make use of a large number of parameters or require a set of prototypes. The proposed methodology can handle these two aspects effectively.

In this paper a method for the segmentation of chromosomes is presented. The method uses the watershed transform to recursively segment touching chromosomes. It also combines the valley searching [6, 7] technique to resolve the problem with touching chromosomes.

**Materials and Methods**

Our proposed method consists of a number of steps as it is shown in Fig. 2. The first step is the conversion of the initial chromosome image to binary. The second step is the removal of very small objects. In the third step, the Euclidean distance transform of the binary image is computed and the grayscale image is reconstructed. The watershed transform is applied in the next step and a first estimation of the segmented chromosome areas is obtained. Every segmented area is processed by local histogram equalization and the watershed transform is further applied to that area – instead of the whole image– until no more new areas are created. Finally, all pale paths [6] are computed and the final segmented chromosome areas are produced.

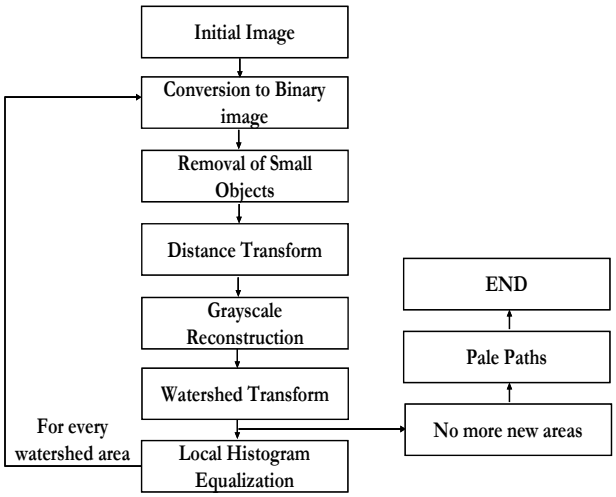



Figure 2: The proposed segmentation method.

In the first step the chromosome image is converted to binary using a well known automated threshold selection process [15] (Otsu's method). The threshold operation is the partition of the pixels of an image into two classes $C_0$ and $C_1$ (representing background and object respectively) at gray level $t$, i.e., $C_0 = \{0,1,\cdots,t\}$ and $C_1 = \{t+1, t+2, \cdots, L-1\}$, where $L$ is the total number of gray levels in the image. Let $\sigma_B^2$, $\sigma_T^2$ be the between-class variance, and the total variance, respectively. An optimal threshold can be determined by minimizing the following function with respect to $t$:

$$\eta = \frac{\sigma_B^2}{\sigma_T^2}. \tag{1}$$

Thus, the optimal threshold $t^*$ is defined as

$$t^* = \arg\min_t \eta, \tag{2}$$

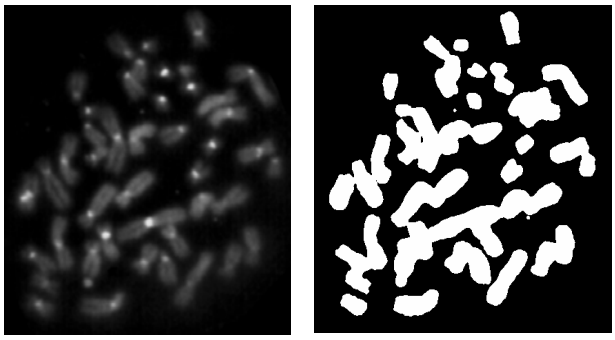
where

$$\sigma_T^2 = \sum_{i=0}^{L-1} (i-\mu_T)^2 P_i,\ \mu_T = \sum_{i=0}^{L-1} iP_i, \tag{3}$$

$$\sigma_B^2 = w_0 w_1 (\mu_1 \mu_0)^2,\ \mathrm{w}_0 = \sum_{i=0}^{t} P_i,\ \mathrm{w}_1 = 1 - w_0, \tag{4}$$

$$\mu_1 = \frac{\mu_T - \mu_t}{1-\mu_0},\ \mu_0 = \frac{\mu_T}{w_0},\ \mu_t = \sum_{i=0}^{t} iP_i,\ P_i = \frac{n_i}{n}. \tag{5}$$

$n_i$ is the number of pixels in grey-level $i$ and $n$ is the total number of pixels in a given image defined as $n = \sum_{i=0}^{L-1} n_i$. Moreover, $P_i$ is the probability of the occurrence of grey level $i$. The application of the Otsu method to the image in Fig. 3(a) results in image of Fig. 3(b).



(a) (b)

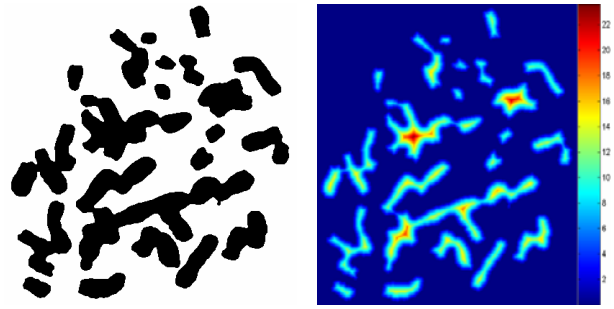
Figure 3: Image segmentation using Otsu's method: (a) Original image, and (b) Thresholded image.

Rejection of very small objects due to debris or other artefacts takes place in the second stage. The area of every segmented object is measured and very small objects are rejected.

In the third step a measure of the separation of points in the image is produced by computing the distance transform (DT) of the complement binary image [16]. Given an $m \times n$ binary image $I$, the distance transform of $I$ is a map that assigns to each pixel the distance to the nearest white pixel. The distance metrics used to compute the distance transform include the $L_1: |x_1 - x_2| + |y_1 - y_2|$, $L_2: \sqrt{(x_1-x_2)^2 + (y_1-y_2)^2}$ and


ISSN: 1727-1983 

$L_\infty : \max\left(|x_1 - x_2|, |y_1 - y_2|\right)$, with $L_2$ metric being the most often used. The distance transform of the complement binary image (Fig. 4(a)) is shown in Fig. 4(b).



(a) (b)

Figure 4: Euclidean distance transform: (a) Complement binary image, and (b) Distance transform.

The distance transform plays an important role in the application of the watershed transform [17, 18]. The number of regional minima† of the negative distance transform acts as a marker to the watershed transform depicting the number of objects that will be segmented. Due to the fact that chromosomes are not completely circular to present exactly one minimum to the distance transform, over segmentation results when applying the watershed transform.

A very efficient method to reduce the number of maxima-minima in a grayscale image is grayscale reconstruction [19]. Let $I$ and $J$ be two grayscale images defined on the same domain, taking their values in the set $\{0,1,\ldots,N\}$, such that $J \le I$ (i.e., for each pixel $p \in D_I$, $J(p) \le I(p)$ )). The grayscale reconstruction $\rho_I(J)$ of (mask) $I$ from (marker) $J$ is:

$$\forall p \in D_I,$$
$$\rho_I(J) = \left\{ k \in [0, N-1] \,\middle|\, p \in \rho_{T_k(I)}(T_k(J)) \right\}, \tag{9}$$

where

$$T_k(I) = \left\{ p \in D_I \,|\, I(p) \ge k \right\}. \tag{10}$$

In order to reduce the number of maxima of the distance transform we apply the grayscale transform as follows:

$$D_h(D) = \rho_D(D - h). \tag{11}$$

where $D$ the distance transform of the binary image, $D_h$ the grayscale reconstructed distance transform and $h \in \Re$.

For the 1-D case the grayscale reconstruction is illustrated in Fig. 5.

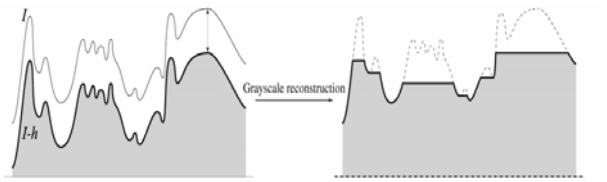



Figure 5: Grayscale Reconstruction of mask $I$ from marker $I - h$.

The computation of the watershed transform (WT) is the next step of our method. 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. The watersheds are the lines separating the catchment basins, as it can be seen in Fig. 6.

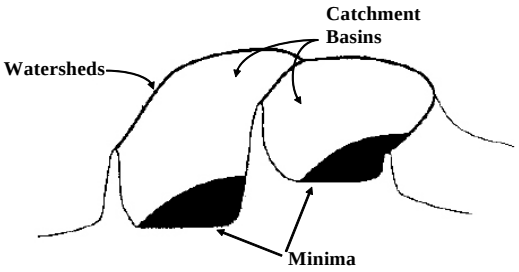



Figure 6: Minima, catchment basins and watershed lines.

The watershed computation algorithm used is based on immersion simulations [17]. The algorithm consists of two basic steps: sorting and flooding. At the first step, all the image pixels are sorted in increase order according to their intensities. At the flooding step, the pixels are accessed in increasing intensity order using the sorted image and labels are assigned to catchment basins.

The output of the watershed algorithm is a tessellation of the input image into its different catchment basins, each one characterized by a unique label. Only the pixels belonging to the watershed transform are assigned a special label to distinguish them from the catchment basins.

In our case the watershed transform of the negative distance transform produces lines which cut the touching chromosomes. An example of the application of the watershed transform to the grayscale reconstructed distance transform is shown in Fig. 7.

The final step of our methodology is Local Histogram Equalization (LHE) [20]. LHE is a contrast enhancing technique which defines a map of gray levels $p$ into gray levels $q$ such that the distribution of gray level $q$ is uniform.

†A regional maxima, minima $M$ of a grayscale image is a connected component of pixels with a given intensity value $h$ such that every pixel in the neighborhood of $M$ has a strictly lower, higher respectively value.


ISSN: 1727-1983 

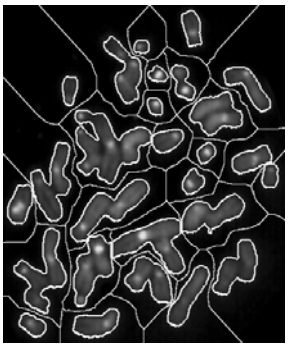


Figure 7: Watershed segmentation applied to the distance transform.

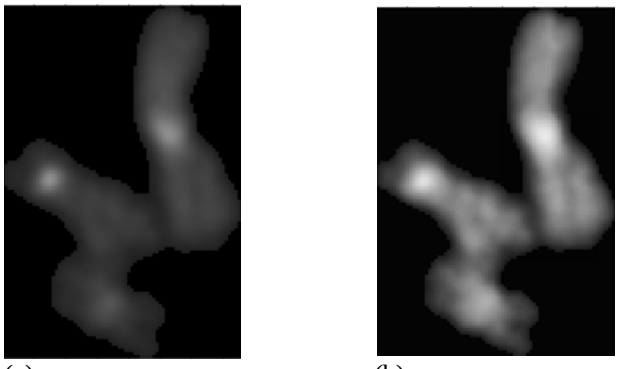
The probability density function of a pixel at intensity level $r_k$ is given as:

$$p_r(r_k) = \frac{n_k}{n}, \qquad (12)$$

where $0 \le r_k \le 1$, $k = 0,1,\ldots,255$, $n_k$ is the number of pixels at intensity level $r_k$ and $n$ is the total number of pixels. A mapped intensity $s_k$ of level $k$ is defined as:

$$s_k = \sum_{j=0}^{k} \frac{n_j}{n} = \sum_{j=0}^{k} p_r(r_j). \qquad (13)$$

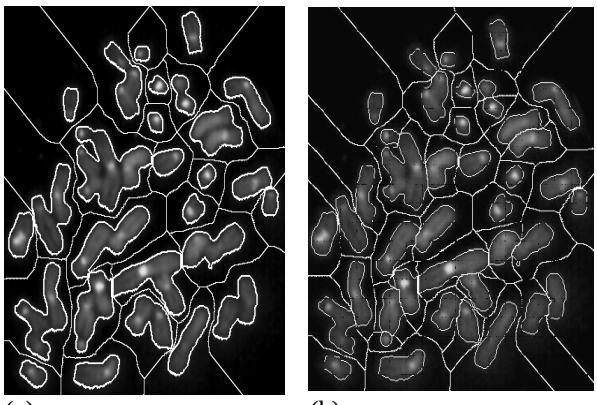
We apply the histogram equalization locally by using a local window $7 \times 7$ centred at each pixel to every watershed area. This results in expanding the contrast locally, and changing the intensity of each pixel according to its local neighbour. The use of LHE is illustrated in Fig 8.



(a) (b)

Figure 8: Local Histogram Equalization (LHE) of a watershed area: (a) Original image, and (b) Local histogram equalized image.

Whereas several methods start with an oversegmentation of the image and iteratively merge regions based on some measure of similarity [21, 22] our method introduces a new region splitting technique based on the watershed transform. All the steps of our method, which does not acquire any a-priori knowledge, are recursively applied to every segmented area until no more new areas are produced. The result of the recursive watershed transform is shown Fig 9.

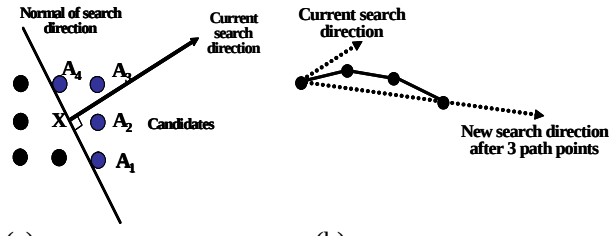


(a) (b)

Figure 9: Recursive watershed transform. (a) Initial segmentation. (b) Final segmentation after 2 iterations.

In order to separate very difficult cases of touching chromosomes the method of "pale paths" [6, 7] is implemented. The basic hypotheses of the method are: (a) At points where chromosomes touch, the optical density is relatively low and (b) where chromosomes touch the cluster boundary tends to form an acute angle.

The algorithm detects pale paths 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. 10(a). The intensity of each candidate is weighted so that candidates lying in the current direction are favored. 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. 10(b).




(a) (b)

Figure 10: Pale path computation: (a) Candidates for the next path point, and (b) Update of path's direction after 3 path points.

We must finally find a cut point and a 4-connected path to the opposite boundary, which passes through an area of relatively low density. By deleting points, of the binary image, along the pale path and refining connected components two separate objects are obtained. The computation of a pale path is illustrated in Fig. 11.


ISSN: 1727-1983 

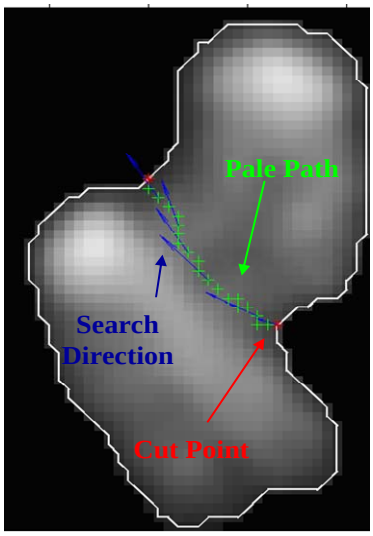



Figure 11: Pale path separating two chromosomes

**Results**

We have validated our methodology with a database of chromosome images which were collected at the Department of General Biology, Medical School of University of Ioannina. The database contains 940 chromosomes. More specifically the database consists of 940 chromosomes: 515 isolated, 396 touching and 29 overlapping. A touching group of chromosomes was present if the chromosomes were touching each other in the binary image (Fig. 3 (b)).

All the isolated chromosomes were segmented correctly by the Otsu method. The performance of our method is measured by means of correct separation of touching chromosomes ( *CSTC* ) which is given as:

$$CSTC = \frac{\#of\ \ correctly\ separated\ chromosomes}{total\ \ \#\ of\ \ touching\ chromosomes}. \quad (14)$$

A correct separation occurs when two or more touching chromosomes are segmented correctly. Indicative results are shown in Table 1.

Table 1: Performance of our method and comparison to the non recursive watershed transform.

| Methodology | CSTC (%) | Erroneous separation (%) |
|---|---|---|
| Watershed (Non Recursive) | 47 | 1 |
| Watershed (Recursive) | 82 | 3 |
| Watershed (Recursive) with Pale Paths | 92 | 5 |

The overall performance of our method is measured by means of correct segmentation of chromosomes ( *CSC* ) which is given by:

$$CST = \frac{\#of\ \ correctly\ segmented\ chromosomes}{total\ \ \#\ of\ \ chromosomes}. \quad (14)$$

where the number of correctly segmented chromosomes is the number of correctly separated chromosomes plus the correctly segmented isolated ones. Indicative results are shown in Table 2.

Table 2: Overall performance of our method.

| Methodology | CSC (%) | Erroneous segmented (%) |
|---|---|---|
| Watershed (Recursive) | 89 | 3 |
| Watershed (Recursive) with Pale Paths | 93 | 5 |

**Discussion**

An automated method for the segmentation of touching group of chromosomes has been presented. It is based on the watershed transform and its application recursively. The application of all the steps of our method locally results in an increase of the *CSTC* as we can see in Table 1. The basic difference from other watershed transform methodologies [17, 18, 21, 22] is the use of WT locally and recursively to every watershed area.

Our approach employs pale paths for separation of difficult cases after the recursive watershed transform. The performance obtained by Liang [6, 7] is marginally better than ours. However, our method can be adopted to other applications, where separation between touching objects is required [14].

Table 3 presents the results of various methods proposed in the literature which deal with touching groups of chromosomes. It is difficult to compare methods directly since published methods rarely use rates directly comparable with others.

Table 3: Related works for segmenting touching groups of chromosomes.

| Authors | Year | CSTC (%) | Erroneous separation (%) |
|---|---|---|---|
| Ji [6] | 1989 | 95 | 4 |
| Ji [7] | 1994 | 95 | 4 |
| Agam et al [3] | 1997 | 82 | 7 |
| Popescu et al [14] | 1997 | 82 | - |
| Lerner et al [9] | 1998 | 83 | 9 |

The segmentation step is a very important step for chromosome analysis. One must segment correctly all chromosomes to proceed to the classification step. It is obvious that the need of an automatic method for the segmentation of touching and overlapping group of chromosomes is vital. Our future work will focus on resolving overlapping groups of chromosomes and proceed to the classification step.

**Conclusions**

We have presented an innovative methodology for


ISSN: 1727-1983 

resolving touching groups of chromosomes. The method is based on the recursive use of the watershed transform. The use of the pale paths method was necessary since cases of touching groups were reported. The proposed method will be further tested to other applications since the separation between touching objects is required in several problems (e.g. touching nuclei).

**References**

[1] PARIS CONFERENCE (1971): 'Standardization in Human Cytogenetics', National Foundation, New York.

[2] LERNER B., (1998): 'Toward A Completely Automatic Neural-Network-Based Human Chromosome Analysis', *IEEE Transactions on Systems, Man and Cybernetics*, **28**, pp. 544-552.

[3] AGAM G., and DINSTEIN I. (1997): 'Geometric Separation of Partially Overlapping Nonrigid Objects Applied to Automatic Chromosome Classification', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**, pp. 1212-1222.

[4] LUNDSTEEN C., GERDES T., MAAHR J., and PHILIP J., (1987): Clinical Performance of A System For Semi Automated Chromosome Analysis, *American Journal of Human Genetics*, **41**, pp. 493-502.

[5] PIPER J., and GRANUM E., (1989): On Fully Automatic Feature Measurement For Banded Chromosome Classification, *Cytometry*, **10**, pp. 242-255.

[6] JI L., (1989): Intelligent Splitting in the Chromosome Domain, *Pattern Recognition*, **22**, pp. 519-532.

[7] JI. L., (1994): Fully Automatic Chromosome Segmentation, *Cytometry*, **17**, pp. 196-208.

[8] POPESCU M., GADER P., KELLER C., KLEIN C., STANLEY J., and CALDWELL C., (1999): Automatic Karyotyping of Metaphase Cells with Overlapping Chromosomes, *Computers in Biology and Medicine*, **29**, pp. 61-82.

[9] LERNER B., GUTERMAN H., and DINSTEIN I., (1998): A Classification-Driven Partially Occluded Object Segmentation (CPOOS) Method with Application to Chromosome Analysis, *IEEE Transactions on Signal Processing*, **46**, pp. 2841-2847.

[10] CHARTERS C., and GRAHAM J., (1999): Trainable Grey-Levels Models for Disentangling Overlapping Chromosomes, *Pattern Recognition*, **32**, pp. 1335-1349.

[11] CHARTERS C., and GRAHAM J., (2002): Disentangling Chromosome Overlaps by Combining Trainable Shape Models With Classification Evidence, *IEEE Transactions on Signal Processing*, **50**, pp. 2080-2085.

[12] STANLEY R., KELLER J., GADER P., and CALDWELL C., (1998): Data-Driven Homologue Matching for Chromosome Identification, *IEEE Transactions on Medical Imaging*, **17**, pp. 451-462.

[13] URDIALES C., RUBIO A., PEREZ A., and HERNANDEZ F., (2003): A Curvature-Based Multiresolution Automatic Karyotyping System, *Machine Vision and Applications*, **14**, pp. 145-156.

[14] MALPICA N., SOLORZANO C., VAQUERO J., SANTOS A., VALLOCORBA I., SAGREDO J., and POZO F., (1997): Applying Watershed Algorithms to the Segmentation of Clustered Nuclei, *Cytometry*, **28**, pp. 289-297.

[15] Otsu N., (1979): A Threshold Selection Method for Gray-Levels Histograms, *IEEE Transactions on Pattern Analysis Machine Intelligent*, **13**, pp. 583-598.

[16] BREU H., GIL J., KIRKPATRICK D., and WERMAN M., (1995): Linear Time Euclidean Distance Transform Algorithms, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17**, pp. 529-533.

[17] VINCENT L., and SOILLE P., (1991): Watershed in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **13**, pp. 583-598.

[18] BEUCHER S, (1992): The Watershed Transformation Applied to Image Segmentation, *Scanning Microscope*, **6**, pp. 299-314.

[19] VINCENT L., (1993): Morphological Grayscale Reconstruction in Image Analysis: Applications and Efficient Algorithms, *IEEE Transactions on Image Processing*, **2**, pp. 176-201.

[20] PIZER S., AMBURN D., AUSTIN D., CROMATIE R., GESELOWITZ A., GREER T., ROMENY B., ZIMMERMAN J., and ZUIDERVELD K., (1987):Adaptive Histogram Equalization and Its Variations, *Computer Vision Graphics and Image Processing*, **39**, pp. 355-368.

[21] HARIS K., EFSTRATIADIS S., MAGLAVERAS N., and KATSAGELOS A., (1998): Hybrid Image Segmentation Using Watersheds and Fast Region Merging, *IEEE Transactions on Image Processing*, **7**, pp. 1684-1699.

[22] LIN G., CHAWLA M., OLSON K., GUZOWSKI J., BARNES C., and ROUSAM B., (2005): Hierarchical, Model-Based Merging of Multiple Fragments for Improved Three-Dimensional Segmentation of Nuclei, Cytometry, **63**, pp. 20-33.


ISSN: 1727-1983 