FISH Image Analysis Using a Modified Radial Basis Function Network

Ioannis N. Kasampalidis*, Ioannis Pitas*, Georgia Karayannopoulou†, Ioannis Kostopoulos†, Paolo Aretini‡, Generoso Bevilacqua‡, Andrea Cavazzana‡, Michele Menicagli§, Antonina Starita¶ and Kleoniki Lyroudia‖
* Department of Informatics, School of Applied Sciences,
Aristotle University of Thessaloniki, Thessaloniki 54124, Greece E-mail: pitas@aiia.csd.auth.gr
† Department of Pathology, Medical School, Aristotle University of Thessaloniki, Thessaloniki , 54124, Greece
‡ Department of Oncology, Division of Surgical, Molecular and Ultrastructural Pathology,
University of Pisa and Pisa University Hospital, Pisa 56126, Italy.
§ MGM, Institute of Molecular Genetics and Medecine, Pisa 56126, Italy.
¶ Department of Computer Science, University of Pisa, Pisa 56127, Italy.
‖ Department of Endodontology, Dental School, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece

***Abstract*— Fluorescent in situ hybridization (FISH) is an valuable method for determining Her-2/neu status in breast carcinoma samples, an important prognostic indicator. Visual evaluation of FISH images is a difficult task which involves manual counting of dots in multiple images, a procedure which is both time consuming and prone to human error. A number of algorithms have recently been developed dealing with (semi)-automated analysis of FISH images. These algorithms are quite promising but further improvement is required in improving their accuracy. Here, we present a novel method for analyzing FISH images based on the statistical properties of Radial Basis Functions. Our method was evaluated on a data set of 100 breast carcinoma cases provided by the Aristotle University of Thessaloniki and the University of Pisa, with promising results.**

I. Introduction

Fluorescent in situ hybridization is an established and proven diagnostic method for gene status evaluation. Among other genes, it is essential in determining the status of Her-2/neu in breast samples, a valuable prognostic and diagnostic indicator [1]. The Her-2/neu (c-erbB2) oncogene is a tyrosine kinase receptor that is over-expressed in approximately 20-30% of high-grade invasive breast carcinomas. Her-2/neu positive tumours can be more aggressive and knowing that a cancer is Her-2/neu positive helps a medical team select the appropriate treatment. Fluorescence in situ hybridization (FISH) is one of the most widely used technologies to determine Her-2/neu status by determining gene-copy count.

The reading of FISH images is a difficult task since manual dot scoring over a large number of nuclei and over different tissue samples is a time consuming and fatiguing technique. Moreover, it is user-dependent in the clinical setting lacking specific training for this imaging technique. Current analysis of FISH signals in practice is performed in a semi-automated way with the aid of image processing software, which can display the different color channels of a FISH image and apply thresholds for nuclei segmentation. A study by Klijanienko et al. [2] has shown strong correlation of detection results using visual-only and semi-automated methods for evaluating the status of Her-2/neu in breast carcinomas samples. However, the counting of dots in a semi-automatic manner still remains an impractical procedure for a pathologist, since it requires user intervention for excluding poorly segmented, overlapping, clustered or nonepithelial cells [2].

Recently, a plurality of techniques have been proposed for analyzing FISH images targeted to a variety of genes. Most of these consist of a two-step process of nuclei segmentation and spot detection. Netten et al. [3], [4], perform nuclei segmentation via the ISODATA algorithm [5] and spot detection based on the top-hat transform followed by a number of variations for determining the threshold for detecting spots. Solrzano et al. [6] accomplish nuclei segmentation based on a combination of the ISODATA algorithm followed by the distance transform, while spot detection is based on the top-hat transform followed by the recursive reconstruction algorithm [7]. In Kozubek et al. [8], segmentation is carried out via bi-level histogram analysis [9] and morphological operations, while spot detection is performed based on a watershed-like technique dubbed gradual thresholding. Lerner et al. [10] , [11] perform nuclei segmentation on the blue channel using heuristically-derived thresholds and morphological operations, while spot detection is evaluated for a number of different techniques varying from Bayesian classifiers to neural networks. In Chawla et al. [12], nuclei segmentation is based on a variation of the watershed transform dubbed "gradient-weighted distance transform". O'Sullivan et al. [13] segment cell nuclei via the watershed transform, while spot detection is based on three different techniques ranging from intensity- to histogram- and watershed-transform-based methods. Finally, Raimondo et al. [14] proposed a method for nuclei segmentation based on thresholding of the blue channel using the algorithm the algorithm by Otsu, followed by the marked watershed algorithm. Spot detection is based on top-hat filtering followed by 2-d correlation with a suitable spot mask, which is derived from a small number of training spots.

Although the results are in general satisfactory, further improvement is desirable, especially regarding nuclei segmentation for overlapping or out of focus nuclei and spot-detection

with excessive debris staining or high gene-copy number. Therefore, we hereby present a novel algorithm for FISH image analysis, which provides further improvements to previous successful efforts. The main advantages of our method is its simplicity and firm foundation on the statistical properties of radial basis functions. In addition, radial basis function exemplify an infinitive candidate for cell nuclei segmentation, since the shape of these objects is usually ellipsoid, thus providing significant aesthetic improvement in segmentation.

II. Materials

Formalin fixed-paraffin embedded tissue blocks from 100 cases were retrieved from the archives of Pathology Department of the Aristotle University Medical School and Pisa University Hospital. The majority of the cases had been submitted to our departments for the performance of the amplification of the HER-2/neu gene via fluorescence in situ hybridization (FISH) method. The copy number of HER-2/neu gene locus at 17q11.2-q12 and alpha satellite DNA located at band region 17p11.1-q11.1 (CEP17) was estimated by FISH in interphase cells on paraffin TMA sections (3.5μm), directly labeled with the PathVysionTM HER-2 DNA probe (Vysis) according to the manufacturer's instructions. Briefly, the sections were de-paraffinized by overnight heating at 60°C and by two xylene washes for 5 min each time, followed by dehydration in 100% ethanol for 5 min twice. The slides were air dried and immersed in pretreatment solution (NaSCN) at 80°C for 30 min. Proteolysis of neoplastic cells was performed by immersing the slides in protease solution at 37°C for 12 min. Denaturation of tissue sections mounted on the slides was performed by a solution of formamide in 70%, pH 7.5, at 72°C for 5 min. Hybridization was carried out by adding to the tissue sections 10L of LSI HER-2/CEP17 DNA probe for overnight incubation at 37°C in a moist chamber. Next day the slides were washed with post-hybridization buffer (2X SSC and 0.3% NP-40) at 72°C for 5 min. Hybridization signals were enumerated in a Zeiss microscope (Axioskop 2 plus HBO 100) equipped with a high quality x100 oil immersion objective, an appropriate filters set (EX BP360/51 for DAPI, EX BP485/17 for FITC/spectrum green, EX BP560/18 for rhodamine/spectrum orange) and a computerized imaging system. Sixty non-overlapping nuclei were selected randomly and scored for each tumor specimen. HER-2/neu probe is labeled in spectrum orange and the CEP 17 probe in spectrum green. Images were captured with a computer-controlled digital camera and processed with a software system (FISH Imager$^{\mathrm{TM}}$ METASYSTEMS). An example of a typical FISH image is shown in Fig. 1.

III. Methods

A. Nuclei Segmentation

Initially, the blue channel is extracted from the original RGB image and the image is subjected to adaptive histogram equalization with default input parameters as implemented in Matlab ®. Next, the image is thresholded based on the algorithm by Otsu et al. [15] and morphological opening and closing is applied with disk-shaped structural elements of radius 1 and 5 respectively. The last step of pre-processing includes image sub-sampling by a factor of 5 in each dimension, yielding an overall sub-sampling ratio of 25.

For nuclei segmentation, a large number, usually 100, of initial cluster centers are initialized at uniform intervals throughout the image, while the covariance matrix of each cluster is initialized at 0. This initial number of cluster centers is deemed large enough to cover the vast majority of cases encountered. Pixels of the sub-sampled image exceeding the threshold are introduced sequentially and the cluster closest to each point, as defined by the Euclidean distance, is updated as follows:

$$\boldsymbol{\mu_j}(t+1) = \boldsymbol{\mu_j}(t) + \eta[\boldsymbol{X_i} - \boldsymbol{\mu_j}(t)], \tag{1}$$

where $\boldsymbol{X_i}$ is the coordinate vector of the pixels, $\boldsymbol{\mu_j}$ is the vector estimate of the center of the cluster being updated and η is the learning rate, which is updated according to [16]

$$\eta = \frac{1}{n_j}, \tag{2}$$

where n_j is the number of pixel members of cluster j. The covariance matrix is calculated via the following formula

$$\boldsymbol{\Sigma_j}(t+1) = \frac{n_j - 2}{n_j - 1}\boldsymbol{\Sigma_j}(t) + \frac{[\boldsymbol{X_i} - \boldsymbol{\mu_j}]^T[\boldsymbol{X_i} - \boldsymbol{\mu_j}]}{n_j - 1}, \tag{3}$$

where $\boldsymbol{\Sigma_j}(t)$ is the covariance matrix estimate of cluster j at time t.

After each new pixel is introduced, a statistical test for cluster splitting is performed, as described in Kotropoulos et al. [16]. More specifically, the sum of squared errors of each cluster is defined as

$$E_j^1 = \sum_{\boldsymbol{X} \in C_j} ||\boldsymbol{X} - \boldsymbol{\mu_j}||^2, \tag{4}$$

while if the cluster is split, the sum of squared errors is defined as

$$E_j^2 = \sum_{\boldsymbol{X} \in C_{j1}} ||\boldsymbol{X} - \boldsymbol{\mu_{j1}}||^2 + \sum_{\boldsymbol{X} \in C_{j2}} ||\boldsymbol{X} - \boldsymbol{\mu_{j2}}||^2, \tag{5}$$

where C_j,C_{j1} and C_{j2} are the original cluster j and the two clusters resulting from its splitting, respectively. The centers resulting from the split cluster are calculated as the mean of their pixels members, while the pixel members of each resulting cluster are found depending on the sign of the following quantity:

$$e_j^T(\boldsymbol{X} - \boldsymbol{\mu_j}), \tag{6}$$

where e_j is the principal normalized eigenvector of cluster j. If the quantity in (6) is positive, the pixel is assigned to cluster C_{j1}, while if negative, the pixel is assigned to cluster C_{j2}. Finally, splitting of the original cluster is accepted if

$$\frac{E_j^2}{E_j^1} < 1 - \frac{2}{p\pi} - \beta\sqrt{\frac{2(1 - \frac{8}{p\pi^2})}{pn_j}}, \tag{7}$$

where p is the dimensionality of the data set, i.e. 2, and β is the upper value corresponding to the 95% percentile of the normal distribution.

When all pixels above the threshold have been introduced, the center and covariance matrix of each cluster are updated by calculating the mean and covariance matrix of its member pixels respectively. Then, clusters without any members are deleted and the mean change of the cluster center estimates since the last iteration is calculated. If the mean change is less than 1 pixel no more iterations are performed. If these conditions are not met, another iteration is performed, up to a maximum of 10, where the number of pixel members and the covariance matrix of each cluster are both re-set to 0. The maximum number of iterations was determined heuristically and is suitable for most of the cases encountered.

The next step consists of hole filling, where connected component analysis is performed on the inverted thresholded image and holes with pixel count greater than $max(n_j)$ are removed. For connected-component analysis, 4-connectivity is used. This is followed by nuclei merging, where initially, we calculate the area of each cluster by including all pixels satisfying

$$[\boldsymbol{X_i} - \boldsymbol{\mu_j}]^T {\boldsymbol{\Sigma_j}}^{-1} [\boldsymbol{X_i} - \boldsymbol{\mu_j}] < \beta, \tag{8}$$

where β is the upper value corresponding to the 95% percentile of the normal distribution. Next, we calculate the fraction of pixels of that cluster below the threshold, i.e. the "black" pixels, as defined by the algorithm by Otsu et al. [15] and we examine the neighbors of each cluster for possible mergers, where neighbor clusters are defined as those meeting the following criterion

$$||\boldsymbol{\mu_i} - \boldsymbol{\mu_j}|| < \beta \times (v_i + v_j), \tag{9}$$

where $\boldsymbol{\mu_i}$ and $\boldsymbol{\mu_j}$ are the centers and v_i and v_j the principal eigenvalues of clusters i and j respectively, while β is the upper value corresponding to the 95% percentile of the normal distribution. Merging of clusters i and j is decided if

$$BlackPixelFraction_{i,j} < \beta, \tag{10}$$

where the black pixel fraction is calculation on both clusters i and j and β is the 50% percentile of the exponential distribution with μ equal to the mean of the black pixel fraction of all initial nuclei before merging. The exponential distribution was used because it was experimentally found to be a better fit for the distribution of the black pixel fraction among all nuclei. The cluster resulting from the merger of two clusters cannot be merged again during the same iteration and this procedure iterates until no more clusters can be merged. Next, we remove outlier clusters satisfying

$$BPF_{i,j} > \beta, \tag{11}$$

where BPF is the fraction of black pixels, β is the is the 95% percentile of the exponential distribution with μ equal to the mean of the black pixel fraction of all nuclei after merging. Finally, clusters touching the image border are also deleted.

B. Spot Detection

Spot detection for the Her-2/neu and CEP-17 probes is performed on the red and green channel separately. Each channel is pre-processed with a succession of steps commencing with top-hat transform with a disk-shaped structural element of radius 4, followed by thresholding, where the threshold is determined as described in Raimondo et al. [14]. Her-2/neu and CEP-17 probes appear in both the red and green channels, with Her-2 having higher intensity in the red channel and CEP-17 in the green channel. For that reason, when detecting red spots, we keep only the pixels of the red channel whose intensity is at least 10% higher than that of the intensity of the green channel. The last step of pre-processing includes morphological opening with a disk-shaped structural element of radius 1.

Next, a number of cluster centers equal to 900 are initialized at uniform intervals throughout the image, with the covariance matrices of the clusters initialized at 0. The initial number of cluster centers is deemed large enough to cover the vast majority of cases encountered. Pixels above the determined threshold are introduced sequentially and the corresponding cluster is updated according to (1) and (3), while clusters without any members are deleted. After introduction of each new pixel, cluster splitting is considered based on (7) and iteration continues until the mean cluster center estimate change is less than 1 or a total number of 10 iterations is reached. Finally, cluster merging is performed, where the merging of each cluster to its neighbors is examined and merging is decided with the closest cluster satisfying

$$||\boldsymbol{\mu_i} - \boldsymbol{\mu_j}|| < \beta \times max(v_i, v_j), \tag{12}$$

where β is the upper value corresponding to the 99% percentile of the normal distribution and v_i and v_j the principal eigenvalues of clusters i and j, respectively.

Finally, the pixels in a 9×9 neighborhood around the spot are examined and if the average intensity of the red channel is not greater by at least 10% than the average intensity of the green channel, when detecting red spots, and vice versa, then the spot is deleted. In addition, spots with fewer than 5 or more than 400 pixel elements are also removed. The 10%-intensity and 5- and 400-pixel thresholds were derived heuristically and are adequate for most of our test cases.

IV. Results

Determination of gene amplification status in standard medical practice involves calculation of the average Her-2/CEP17 ratio in every nucleus and averaging that ratio over a subset of nuclei. Our team of medical experts averages the ratio over 60 nuclei. In order to calculate the average spot ratio, we only include all nuclei with at least one red spot; these nuclei are labeled valid nuclei. For these nuclei, the number of green spots is the maximum of 1 and the actual number of green spots detected in the nucleus. For classification, we follow two approaches. In the first, we classify nuclei as positive or negative, depending on whether the ratio is greater than or less than 2 respectively. In the second approach, we ignore nuclei

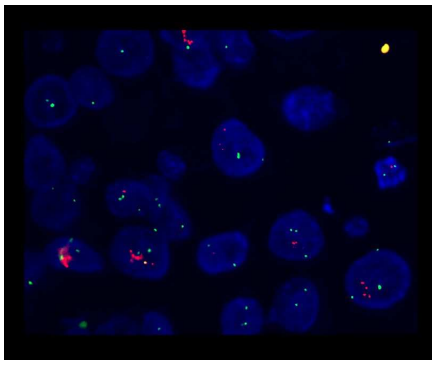

Fig. 1. Original FISH image. Nuclei are shown in blue, while Her-2/neu and CEP-17 probes in red and green respectively.

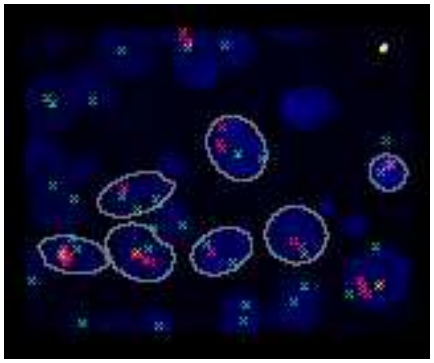

Fig. 2. FISH image analysis results. Segmented nuclei are outlined, while detected Her-2/neu and CEP-17 fluorescent probes are marked in red and green 'X' respectively.

with a ratio between 1.9 and 2.1, placing them in a special category that requires manual classification by the experts.

The test set provided by AUTH Medical School and the University of Pisa comprised of 100 cases of breast cancer carcinoma. As determined by the medical experts, 25 cases were identified as positive and 75 as negative. For the first approach, the overall accuracy 95%, 84% for the posivive cases and 99% for the negative cases. Following the second approach, where we place the cases with ratio in the 1.9-2.1 interval in the "unknown" category, our method is is able to correctly classify all the remaining cases, while only 10% of cases are placed in the unknown category.

V. Conclusion

We have presented an novel method for analyzing FISH images based on the statistical properties of Radial Basis Functions with promising results. Our method also provides for a gray area, where cases need to be further evaluated manually by an expert clinician. Our results show than when this approach is used, the accuracy of our method is 100%, while the percentage of cases that need to be further evaluated is 10%. The clinicians in our team consider this to be an acceptable compromise. Further tests and comparisons to other methods on a larger data are necessary in order to adopt our approach for clinical practice.

Acknowledgment

This work was supported by the EU project Biopattern: Computational Intelligence for biopattern analysis in Support of eHealthcare, Network of Excellence Project No. 508803.

References

[1] P. P. Osin and S. R. Lakhani, "The pathology of familial breast cancer. immunohistochemistry and molecular analysis," *Breast Cancer Research*, vol. 1, no. 1, pp. 36–40, 1999.

[2] J. Klijanienko, J. Couturier, M. Galut, A. K. El-Naggar, Z. Maciorowski, E. Padoy, V. Mosseri, and P. Vieth, "Detection and quantification by fluorescence in situ hybridization (fish) and image analysis of her-2/neu gene amplification in breast cancer fine-needle samples," *Cancer Cytopathology*, vol. 87, no. 5, pp. 312–318, 1999.

[3] H. Netten, I. T. Young, L. J. van Vliet, H. J. Tanke, H. Vroljik, and W. C. R. Sloos, "Fish and chips: Automation of fluorescent dot counting in interphase cell nuclei." *Cytometry*, vol. 28, no. 1, pp. 1–10, 1997.

[4] H. Netten, L. J. van Vliet, H. Vroljik, W. C. R. Sloos, H. J. Tanke, and I. T. Young, "Fluorescent dot counting in interphase cell nuclei." *Bioimaging*, vol. 4, pp. 93–106, 1996.

[5] T. W. Ridler and S. Calvard, "Picture thresholding using an iterative selection method." *IEEE Trans. on Systems, Man, and Cybernetics*, vol. SMC-8, no. 8, pp. 630–632, 1978.

[6] C. O. de Solorzano, A. Santos, I. Vallcorba, J.-M. Garcia-Sagredo, and F. del Pozo, "Automated fish spot counting in interphase nuclei: Statistical validation and data correction." *Cytometry*, vol. 31, no. 2, pp. 93–99, 1998.

[7] L. Vincent, "Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms." *IEEE Trans. on Image Proc.*, vol. 2, no. 2, pp. 176–201, 1993.

[8] M. Kozubek, S. Kozubek, E. Lukasova, A. Mareckova, E. Bartova, M. Skalnikova, and A. Jergova, "High-resolution cytometry of fish dots in interphase nucleus nuclei." *Cytometry*, vol. 36, no. 4, pp. 279–293, 1999.

[9] W. K. Pratt, *Digital image processing*. New York: John Wiley & Sons Inc, 1991.

[10] B. Lerner, W. F. Clocksin, S. Dhanjal, M. A. Hulten, and C. M. Bishop, "Feature representation and signal classification in fluorescence in-situ hybridization image analysis." *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 31, no. 6, pp. 655–665, 2001.

[11] W. F. Clocksin and B. Lerner, "Automatic analysis of fluorescence in-situ hybridization images." in *Proceedings of the 11th British Machine Vision Conference*, Bristol, England, September 2000, pp. 666–674.

[12] M. K. Chawla, G. Lin, K. Olson, A. Vazdarjanova, and e. a. S. N Burke, "3d-catfish: a system for automated quantitative three-dimensional compartmental analysis of temporal gene transcription activity imaged by fluorescence in situ hybridization." *Journal of Neuroscience Methods*, vol. 139, pp. 13–24, 2004.

[13] J. N. O'Sullivan, J. C. Finley, R. A. Risques, W. Shen, K. A. Gollahon, A. H. Moskovitz, S. Gryaznov, C. B. Harley, and P. S. Rabinovitch, "Telomere length assessment in tissue sections by quantitative fish: image analysis algorithms." *Cytometry*, vol. 58, no. 2, pp. 120–131, 2004.

[14] F. Raimondo, M. A. Gavrielides, G. Karayannopoulou, K. Lyroudia, I. Pitas, and I. Kostopoulos, "Automated evaluation of her-2/neu status in breast tissue from fluorescent in situ hybridization images." *IEEE Transactions on Image Processing*, vol. 14, no. 9, pp. 1288–1299, 2005.

[15] N. Otsu, "A thresholding selection method from graylevel histogram," *IEEE Trans. on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[16] C. Kotropoulos and I. Pitas, "A variant of learning vector quantizer based on split-merge statistical tests," in *Lecture Notes in Computer Science:Computer Analysis of Images and Patterns*, D. Chetverikov and W. G. Kropatsch, Eds. Springer Verlang, 1993, pp. 822–829.