33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

An Algorithm for Cancer Nest Feature Extraction from Pathological Images

Tomoyuki Hiroyasu , Hiroaki Yamaguchi , Sosuke Fujita , Mitsunori Miki ,
Masato Yoshimi , Maki Ogura , Manabu Fukumoto

Abstract— **Here, we propose an algorithm to automatically obtain extraction filters for the affected regions from cancer images. The proposed algorithm consists of two steps: extraction of affected region candidates and elimination of false positives. Useful features of cancer images, such as the area and degree of circularity of cancer nests, etc., are extracted using the derived filters. These features are useful for supporting pathological diagnosis. Automatic Construction of Tree-structural Image Transformation (ACTIT) was used to construct these filters to extract the affected regions from the image. The proposed algorithm was applied to a mouth cancer pathological image. The results confirmed that the proposed algorithm can obtain good filters capable of extracting cancer nests. The derived filters were also applied to other images from the same specimen. The results also indicated that the generated filters show general versatility in extracting cancer nest candidates. The area and degree of circularity of the cancer nets were also derived automatically.**

I. Introduction

Pathological image diagnosis has become more important since new types of medical facilities are developed and utilized. However, pathologists make pathological diagnoses subjectively based on their knowledge and experience. In addition, there is an extreme shortage of pathologists in Japan. To address these issues, a pathological image support system has been developed. This system supports pathologists by automating the pathological diagnosis using a computer[1], [2]. There are several types of pathological image support system. One type of system involves specific image processing associated with each case and extracts affected regions from pathological images. The system supports the pathologist by deriving features for diagnosis automatically. These systems may improve the quantity and quality of diagnosis. Other groups have reported diagnosis support systems using not only pathological images but also other medical images[3], [4], [5].

As one method to extract the target region from images automatically, Nagao et al. proposed Automatic Construction of Tree-structural Image Transformation (ACTIT) to automatically construct image-processing filters[6], [7], [8]. Hiroyasu et al. constructed filters for pathological images using ACTIT[9]. To derive the amounts of affected region features, it is necessary to extract the affected regions from pathological images with high accuracy. However, tuning the algorithm to extract the affected regions with higher fraction results in more false positives. On the other hand, tuning the algorithm to reduce the false positives also reduced the true positives. Thus, there is a trade-off problem of accuracy in extracting affected regions. To overcome this problem, we propose an algorithm to construct image filters that extract the affected region automatically. The proposed algorithm consists of two types of image-processing filter, which extract the affected region candidates and eliminate the false positive regions. At the same time, we extract the amounts of useful features in extracted affected regions. Here, we extracted cancer nests and the feature amounts in a mouth cancer pathological image and verified the results.

T. Hiroyasu is with the Department of Life and Medical Sciences, Doshisha University.
H. Yamaguchi and S. Fujita are with the Graduate School of engineering, Doshisha University.
M. Miki and M. Yoshimi are with the Department of Science and Engineering, Doshisha University.
M. Ogura is with the Innovative Service Solutions Division, NEC Corporation and Graduate School of Medice, Tohoku University.
M. Fukumoto is with the Insitute of Development, Aging and Cancer, Tohoku University.

II. ACTIT

A. Outline

ACTIT proposed by Nagao et al. is a method used to construct image-processing filters automatically to obtain objective images. Generally, image processing can be expressed as a combination of known simple image-processing filters. Nagao et al. treated image processing as a combinatorial optimization problem of image-processing filters. In this method, a tree topology filter, which is shown in Fig 1, is constructed automatically by finding the solution of the combinatorial optimization problem.

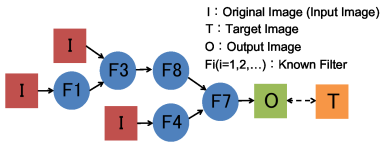


Fig. 1. Principle of automatic construction of tree topology filter

To construct the tree topology filter, the learning image set, consisting of the combination with the original image I and the target image T that is the ideal processing image for the original image, and multiple known filters are prepared in advance. The tree topology filter, which approximates from the original image to the target image, is constructed by optimizing the combination of known filters. To achieve image processing using the tree topology filter shown in Fig 1, the basic original image is input into the terminus nodes of this tree topology filter. Then, the known filter stored in each node is processed in order, and the output

image is obtained. To evaluate the tree topology filter, the following evaluation function (1) is used.

$$Evaluation\ Value = \frac{1}{k}\sum_{k=1}^{K}\left\{1 - \frac{\sum_{x=1}^{W_x}\sum_{y=1}^{W_y}|O(x,y) - T(x,y)|}{W_x W_y V_{max}}\right\} \quad (1)$$

The evaluation value is determined from the delta of each pixel value in the output image O and the target image T, and the maximum value is 1.0. K, W_x, W_y, and V_{max} indicate the number of learning image sets, the width of the image, the height of image, and the maximum tone value, respectively.

Table I and II show the prepared known filters[10]. These filters have only one or two inputs.

TABLE I
ONE INPUT KNOWN FILTERS

Number	Effect
f1	Edge emphasis(sobel)
f2	Edge emphasis(laplacian)
f3	Dilation
f4	Erosion
f5	Median
f6	Light Pixel
f7	Dark Pixel
f8	Large Area
f9	Small Area
f10	Binarization
f11	Inversion
f12	Fill

TABLE II
TWO INPUT KNOWN FILTERS

Number	Effect
F1	Bounded Sum
F2	Bounded Prod
F3	Logical Sum
F4	Logical Prod
F5	Algebraic Sum
F6	Algebraic Prod
F7	Sub

B. Optimization method for ACTIT

Koza proposed GP as an optimization method to determine the optimum solution for tree topology[11], [12]. GP is an optimization method in which the genetic algorithm has been extended to design for structural expression (tree topology, graph topology)[13].

In GP, a population consisting of multiple initial individuals (candidates for the solution) is first generated. An optimum tree topology for the problem is constructed by repeatedly applying genetic operators, such as evaluation, selection, crossover, and mutation, to the population. In this study, the evaluation value of each individual is obtained from the evaluation function (1).

III. ALGORITHM FOR EXTRACTION OF FEATURES OF CANCER NESTS

This section describes the algorithm used to extract the features of cancer nests from mouth cancer pathological images.

Step1. The cancer nest candidates in the pathological image are extracted from the mouth cancer pathological image. (First Filter)

Step2. The false positive regions are eliminated from the image obtained in Step 1. (Second Filter)

Step3. The feature amounts of elements are extracted from the image obtained in Step 2.

We construct the First Filter, which extracts the cancer nest candidates from the pathological images, and the Second Filter, which eliminates the false positive regions, using ACTIT. Then, we determine the degree of circularity and area as the amounts of cancer nest features in Step 3. As many of the cancer nests are circular or elliptical, the information regarding the degree of circularity helps the pathologist to make a diagnosis of cancer. As an example of use of features, the region has a high probability that the cancer nest if the degree of circularity of the region is high and it is possible to determine objectively whether the size of the cancer nest has increased by observing the size in time series.

Thus, it is possible to judge objectively by extracting the cancer lesion size as numerical data. The degree of circularity is calculated in the range of 0.0-1.0 by expression (2). S is the affected region area, L is the affected region boundary length, and the area is the number of pixels that make up the affected region. According to this equation, an area with a value of 1.0 is a circle.

$$Degree\ of\ Circularity = \frac{4\pi S}{L^2} \quad (2)$$

IV. VERIFICATION OF PROPOSED ALGORITHM ' S EFFECTIVENESS

We verified the effectiveness of the proposed algorithm to the mouth cancer pathological image shown in Fig 2. We verified the general versatility by applying two types of image-processing filter constructed by the proposed algorithm to multiple pathological images of the same specimen.

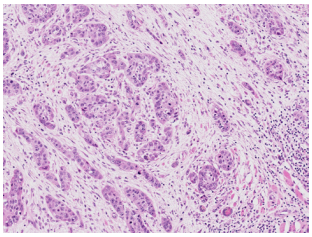


Fig. 2. Pathological Image

Fig 2 shows a 10 $\times$ microscopic image of a pathological slide. This pathological slide was produced by Hematoxylin and Eosin staining of cells harvested from the human body and sliced. To evaluate the region extracted by the proposed algorithm, we defined the true positive fraction (TPF), which is the probability of correctly extracting the target, and false positive fraction (FPF), which is the probability of falsely extracting non-target, according to expressions (3) and (4).

$$TPF = \frac{A_C \cap A_R}{A_R} \quad (3)$$

$$FPF = \frac{A_C \cap \bar{A_R}}{A_C} \quad (4)$$

A_C is the region extracted by the proposed algorithm and A_R is the extraction target region. Therefore, TPF and FPF approach 1.0 and 0.0 as the accuracy of extraction increases, respectively, and the best extraction result is obtained when (TPF, FPF) is (1.0, 0.0). According to the information of authority doctor, the target values were set as TPF $\geq$ 0.9 and FPF $\leq$ 0.2. Therefore, the extraction result is judged to be good if it fulfills these values.

Table I and II show known simple image-processing filters, which are parts of the derived filter. The parameters of GP are summarized in Table III.

TABLE III
GP PARAMETERS

Parameter	Value
Number of Generations	500
Population Size	500
Selected Method	Tournament
Tournament Size	2
Crossover Rate	0.9
Mutation Rate	0.1
Tree Depth Restriction	25

The Large Area filter shown in Table I extracts the corresponding regions when the area of arbitrary regions is larger than the threshold, while the Small Area filter extracts regions smaller than the threshold. These filters are used to eliminate noise for small areas. Here, we set the thresholds of these filters to 500 as the optimum threshold to eliminate noise without eliminating the cancer nest regions. This value was derived by preliminary experiment.

A. Extraction of the cancer nest candidates (Step1)

We constructed an image-processing filter (First Filter) to extract the cancer nest candidate from Fig 2. The learning image sets used for filter construction is shown in Fig 3. The learning image sets shown in Fig 3 were smalls area of the cancer nest regions extracted from Fig 2.

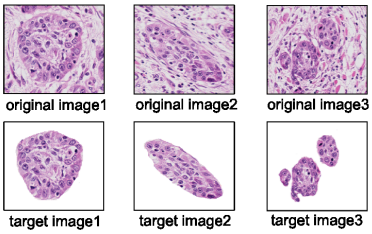


Fig. 3. Learning image sets for First Filter construction

We constructed the First Filter by ACTIT using these image sets. The images shown in Fig 3 were converted to grayscale images when they were input into filters. The constructed First Filter is shown in Fig 4. The output image, which was obtained by applying the First Filter to Fig 2, is shown in Fig 5.

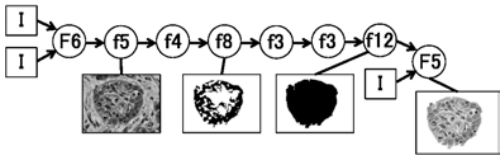


Fig. 4. The First Filter

The filter numbers shown in Fig 4 corresponding to those shown in Table I and II.

As shown in Fig 4, we confirmed that the First Filter extracted cancer nest candidates from the learning image. As shown in Fig 5, TPF was 0.982, which fulfilled the target value. On the other hand, FPF was 0.310, and many false positive regions were extracted.

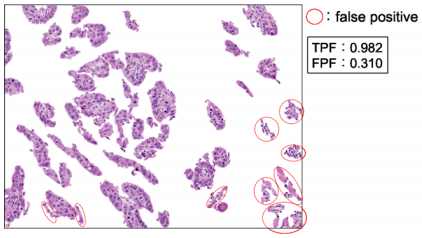


Fig. 5. Extraction of cancer nest candidates (First Filter)

B. Elimination of false positive regions (Step 2)

We constructed the Second Filter to eliminate the false positive regions from Fig 5 using ACTIT. The learning image set is shown in Fig 6.

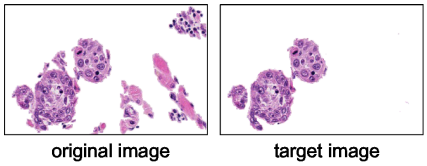


Fig. 6. Learning image set for Second Filter construction

The original image shown in Fig 6 is a part of Fig 5. The target image was created by eliminating false positive regions from the original image. The constructed Second Filter is shown in Fig 7. The result obtained by application of the Second Filter to Fig 5 is shown in Fig 8.

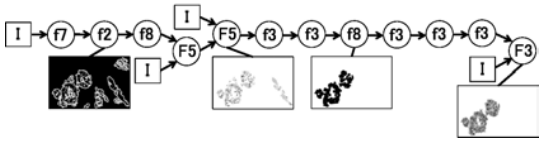


Fig. 7. The Second Filter

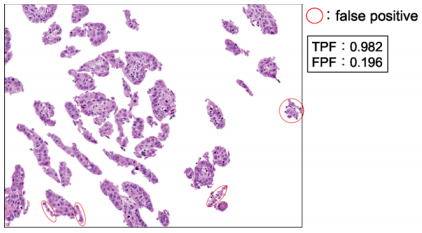


Fig. 8. Elimination of false positive regions

As shown in Fig 7, we confirmed that the Second Filter eliminated the false positive regions of the cancer nest regions. As shown in Fig 8, FPF changed from 0.310 to 0.196 with maintenance of high TPF, and the target values were fulfilled. From Fig 5 and 8, it was confirmed that the proposed algorithm produced filters capable of appropriately extracting the cancer nests.

C. Application of the filters to pathological images of the same specimen

We applied the First Filter and the Second Filter shown in Fig 4 and 7 to nine other pathological images (A-1 - A-9) obtained from the same specimen (specimen A). The results are shown in Table IV.

TABLE IV

APPLICATION OF FILTERS TO THE IMAGES OF THE SAME SPECIMEN

	TPF		FPF	
	First Filter	Second Filter	First Filter	Second Filter
A-1	0.9432	0.9432	0.1672	0.1504
A-2	0.9805	0.9805	0.3770	0.2714
A-3	0.9777	0.9583	0.1425	0.1425
A-4	0.9472	0.9255	0.1324	0.1205
A-5	0.9801	0.9753	0.1856	0.1695
A-6	0.9295	0.9166	0.3149	0.2790
A-7	0.9761	0.9761	0.4806	0.3076
A-8	0.9294	0.9139	0.1861	0.1557
A-9	0.9775	0.9775	0.5402	0.5064
Mean	0.9601	0.9502	0.2807	0.2337
Median	0.9761	0.9583	0.1861	0.1695

As shown in Table IV, TPF $\geq$ 0.9 in all images. Therefore, the First Filter showed high general versatility for extraction of cancer nest candidates. Moreover, FPF $\leq$ 0.2 in A-1, A-3, A-4, A-5, and A-8, and thus the target values were fulfilled. The false positive regions of A-9, which had higher FPF than the other regions, were examined, and the results confirmed that there were many false positive regions that were not present in the learning images. The false positive regions using learning images and false positive images of A-9 are shown in Fig 9.

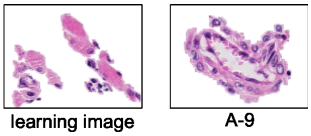
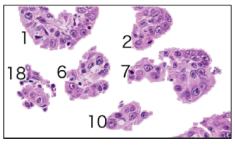


Fig. 9. Extraction of false positive regions

In this study, we used only one set. However, to increase the general versatility with the Second Filter, several types of learning image set with a wide range of diversity of false positive regions should be prepared.

D. Extraction of cancer nest feature amounts (Step3)

We extracted feature amounts of extracted regions from Fig 8. The feature amounts extracted in this study were the area and degree of circularity of the regions. A part of the extracted image and the feature amounts are shown in Fig 10.



number	AREA[px]	CIRCULATORY
1	5792	0.45
2	3056	0.64
6	3750	0.48
7	6801	0.27
10	1968	0.51
18	2350	0.23

Fig. 10. Results of feature amount extraction

In No.18, we determined that both the area and degree of circularity were lower than in the others. In fact, this region was not a false positive region. Thus, we confirmed the features of cancer nests, with large area or degree of circularity, by extracting feature amounts. Thus, extracting and analyzing these feature amounts supports pathological diagnosis and allows objective judgment.

V. CONCLUSIONS AND FUTURE WORKS

In this paper, an algorithm to extract cancer nest features from mouth cancer pathological images was proposed. This algorithm supports pathological diagnosis. The proposed algorithm uses two types of image-processing filter to extract the target cancer nests effectively: a filter for extracting cancer nest candidates and a filter for eliminating false positive areas using ACTIT. On application of the proposed algorithm to a mouth cancer pathological image, target extracted images with TPF = 0.982 and FPF = 0.195 were derived, which satisfied the target values. Moreover, application of the derived filters to other images from the same specimen yielded target extracted images with TPF $\geq$ 0.9. Thus, the results demonstrated the high general versatility of the derived filter. On the other hand, the FPF values of some images were high. This defect should be overcome in future studies. One solution is to examine the learning image set to improve the general versatility of the Second Filter.

The area and degree of circularity of the cancer nests were extracted as cancer features in this study. It is assumed that these values of cancer features will help the pathologist to make objective judgments. Furthermore, we expect this algorithm to be useful not only to check the affected region but also to determine the type and degree of disease progression. In future, increases in the types of extracted features will be expected to provide further support for pathological diagnosis. Thus, the relationship between cancer status and features of cancer images should be examined in future studies.

REFERENCES

[1] NEC. Pathology picture analysis[in japanese]. http://www.nec.co.jp/solution/bio/rd/.

[2] Maki Ogura, Akira Saito. Cancer diagnosis assistance system[in japanese]. *Pathology and clinical*, Vol. 24, No. 4, pp. 411–415, 2006.

[3] E-Liang Chen, Pau-Choo Chung, Ching-Liang Chen, Hong-Ming Tsai, and Chein-I Chang. An automatic diagnostic system for ct liver image classification.

[4] A. Banumathi, J. Praylin Mallika, S. Raju, and V. Abhai Kumar. Automated diagnosis and severity measurement of cyst in dental x-ray images using neural network. *Biomedical Soft Computing and Human Sciences*, Vol. 14, No. 2, pp. 105–110, 2009.

[5] Metin N. Gurchan, Jun Kong, Olcay Sertel, B.Barla Chambazoglu, Joel Saltz, and Umit Catalyurek. Computerized pathological image analysis for neuroblastoma prognosis. *AMIMA Annu Symp Proc*, Vol. 2007, pp. 304–308, 2007.

[6] Aoki and T.Nagao. Actit; automatic construction of tree-structural image transformation. *The Institute of Image Information and Television Engineers*, Vol. 53, No. 6, pp. 888–894, 1999.

[7] T. Nagao S. Masunaga. Automatic generalization of image transformation processes using genetic algorithm. *Proceedings of the IWANVMT-97*, pp. 101–106, 1997.

[8] Y.Nakao and T.Nagao. 3d medical image processing using 3d-actit; automatic construction of tree-structural image transformation. *IEIC Technical Report*, Vol. 103, No. 540, pp. 49–53, 2004.

[9] T.Hiroyasu, S.Fujita, A.Watanabe, M.Miki, M.Ogura, and M.Fukumoto. Comparison of gp and sap in the image-processing filter construction using pathology images. *CISP*, Vol. 9, No. 16, pp. 904–908, 2010.

[10] T. Nagao S. Aoki. Automatic construction of tree-structural image transformation using genetic programming, international conference of image processing. pp. 529–533, 1999.

[11] J.Koza. *Genetic programming, on the programming of conputers by means of natural selection*. MIT Press, 1992.

[12] J.Koza. *Genetic programming, Automatic Discovery of Reusable Programs*. MIT Press, 1994.

[13] D.E.Goldberg. Genetic algorithms in search; optomization; machine learning. *Addison Wesley*, 1989.