A Hierarchical Type Segmentation Algorithm based on Support Vector Machine for Colorectal Endoscopic Images with NBI Magnification

Takumi Okamoto [1], Tetsushi Koide [1], Anh-Tuan Hoang [1], Koki Sugi [1], Tatsuya Shimizu [1],
Toru Tamaki [2], Bisser Raytchev [2], Kazufumi Kaneda [2], Yoko Kominami [3], Shigeto Yoshida [3], Shinji Tanaka[3]
1) Research Institute for Nanodevice and Bio Systems Hiroshima University
2) Graduate School of Engineering, Hiroshima University, 3) Graduate School of Biomedical Sciences, Hiroshima University
1-4-2 Kagamiyama, Higashi-Hiroshima, Japan
e-mail : { okamoto-rnbs, koide } @ hiroshima-u.ac.jp

Abstract - With the increase of colorectal cancer patients in recent years, the needs of quantitative evaluation of colorectal cancer are increased, and the computer-aided diagnosis (CAD) system which supports doctor's diagnosis is essential. In this paper, a hardware design of type identification module in CAD system for colorectal endoscopic images with narrow band imaging (NBI) magnification is proposed for real-time processing of full high definition image (1920x1080 pixel). A pyramid style identifier with SVMs for multi-size scan windows, which can be implemented with small circuit area and achieve high accuracy, is verified for actual complex colorectal endoscopic images.

I. Introduction

With the increase in the number of colorectal cancer patients, systems which support a doctor's diagnosis have been researched. The computer-aided diagnosis (CAD) system for colorectal endoscopic images with narrow band imaging (NBI) magnification [1] has already been proposed [2]. It can processes only the center 120x120 pixels of the endoscopic image for diagnosis and achieves 14.7 fps in throughput. In order to diagnosis a Full HD (1920x1080 pixels) endoscopic image, which may include several types of NBI magnification findings, that system needs to raster scan the image by scan window (SW). The raster scanning a Full HD image using 120x120 pixels SW size at every 10 pixels on software requires (1/14.7 fps) x (17,557 SWs) = 20 minutes, which is not acceptable in real-time diagnosis. Our proposed system processes Full HD images and shows the diagnosis results at the same time with the original input image for comparison. Hence, the processing time should be as small as one second to synchronize the original image and result. The proposed CAD system classifies colorectal endoscopic images obtained by classifying endoscopic diagnosis into three types (Types A, B and C3) based on the NBI magnification findings (Fig.1). The outline of the proposed CAD system is shown in Fig.2. It adopts Bag-of-Features (BoF) as a fundamental concept. The overview of processing flow of the system is as follows. First, the feature quantities for each type are extracted from training images for each type by Dense Scale-Invariant Feature Transform (D-SIFT) [3]. Then, those feature quantities are clustered and a Visual-Word (VW), which is effective in differentiation of each type image, is elected. In our system, 768-dimension (256x3 types) VWs are used for feature representation. Next, for each type image, a histogram is created by voting all the feature quantities for similar VWs. Then, the learning and identification steps by Support Vector Machine (SVM) are performed using the histograms generated above VWs. Finally a histogram is created from an input test image, and is identified by the learned SVM identifiers. In this paper, we propose a 2-step identifier based on SVM to realize a 3-class (3-type) identification and a pyramid-style bottom-up approach, which can be implemented with small circuit area and achieves high throughput and high accuracy, for the actual complex colorectal endoscopic images in real time. In order to verify the effectiveness of the proposed technique, we create several artificial images by combining several colorectal endoscopic images. On the other hand, we also verify the identification accuracy when number of voting key points for VW histogram creation increases by narrowing the key point sampling pitch.

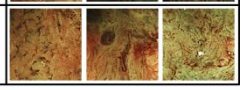
Type A (normal cells)	Microvessels are not observed or extremely opaque.	
Type B (tubular adenomas)	Fine microvessels are observed around pits, and clear pits can be observed via the nest of microvessels.	
Type C3 (cancer)	Pits via the misrovessels are invisible, irregular vessel diameter is thick, or the vessel distribution is heterogeneous, and avascular areas are observed.	

Fig.1. Narrow Band Imaging (NBI) magnification findings [2].

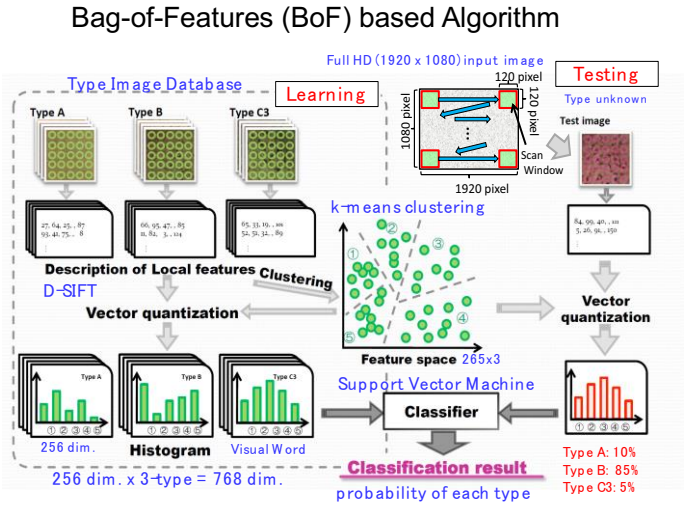


Fig.2. The outline of computer-aided diagnosis (CAD) system for colorectal endoscopic images with NBI magnification.

II. Type Identification Based on Support Vector Machine

A. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a technique of binary classification introduced in 1990s by Vapnik [4]. There are several kernels of SVM, which generate identification hyper-plane that maximizes a margin between a class *Y* and a class *Z*, allows the SVM to achieve high identification performance to an unlearned image. The accuracy difference in SVM computation using those different kernels is only 1% in our previous experiments [1]. Hence, the simplest one, the linear kernel is selected in our work. The linear kernel is also the best suitable for hardware implementation. Equation (1) shows a decision function of SVM with linear kernel in this work.

$$d_{Y:Z}(\vec{x}) = \sum_{i=1}^{N_Y+N_Z} coef_i \times \overrightarrow{sv_i} \cdot \vec{x} + \rho_{Y:Z} \qquad (1)$$

In which, $\boldsymbol{x}$ is a test data, $\boldsymbol{sv}_i$ is Support Vector (SV), which is obtained at the learning step for hyper plane identification between two corresponding classes *Y* and *Z*. $coef_i$ is a coefficient of each $\boldsymbol{sv}_i$, and $\rho_{Y:Z}$ is a coefficient of a decision function. If the identification function $d_{Y:Z}(\boldsymbol{x})$ is a positive value (>0), an input data $\boldsymbol{x}$ is determined as the class *Y*.

B. 3-Type Classification for Computer-Aided Diagnosis

In order to realize 3-class (3-type) identification, the one-versus-one technique [5] is taken. In our work, *f(x)* is a 2-class identifier with SVM. There are two methods in combining binary classifier. The first method uses majority vote of three binary identifiers and the second method uses probability estimation technique [6]. As shown in Fig.3, our proposed system receives the original endoscopic image online, extracts feature of each scan window (SW) before identifying the probability for each type (Type A, B and C3). Then, a color gradation map, which is converted from probabilities of all three types for each SW in the original endoscopic image is displayed on a support screen. It helps doctor to diagnose the type of each region of colorectal endoscopic images. In this work, we use the probability estimation technique among the results of 2-class identifiers (Fig. 4).

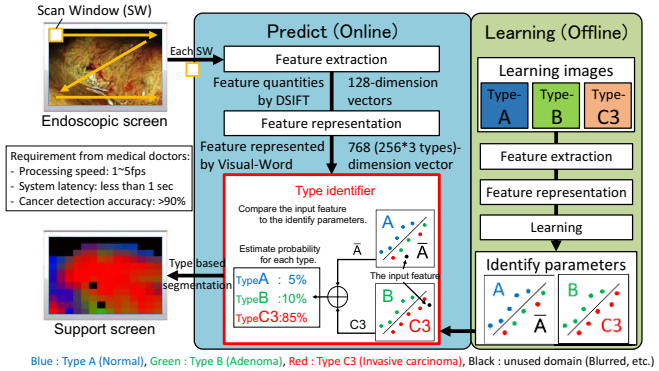


Fig.3. The proposed real-time processing for three types identifications and segmentation for CAD system with colorectal endoscopic image.

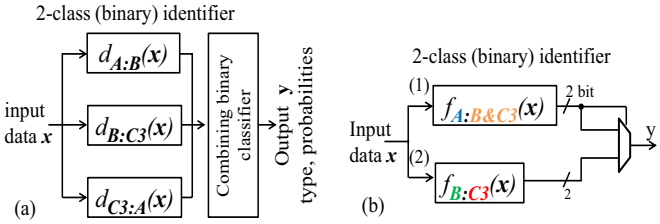


Fig. 4. 3-class identification techniques. (a) one-versus-one (1-step) identification, and (b) 2-step identification.

III. Simulation Result for Real Endoscopic Images

Since high dimension feature quantity (768 dimensions) is processed in our medical application, it is necessary to perform the sum-of-product operations in parallel. Hence, there is a subject that how to reduce hardware resources while increasing parallel degree. Therefore we implement the SVM identification module by a fixed-point number instead of floating point number. The bit length reduces from 64 bits to around 16 bits, and, so as reduces hardware resources. Then, we suggest an algorithm of 3-class identifier which has efficient processing and high accuracy.

A. 3-Type Classification for Computer-Aided Diagnosis

The Simple Gaussian Filter (GF) which is simplified for hardware implementation in [7] is used for feature extraction. We use LIBSVM [6] to create SVM classifiers in simulation. In order to implement the SVM identifier by fixed-point number, Equation (1), which contains floating-point number arithmetics, is changed into fixed-point number arithmetics based on LIBSVM. The *linear kernel* (*inner product*) is used as a kernel of SVM, and the number of Visual-Word is set to 768. 1260 real images in Table 1, which are taken by Hiroshima University Hospital Department of Endoscopy, are used as the learning and testing data set. These images are classified based on the NBI magnification findings [2] by at least two professional doctors in six years. Equations 2 is used for accuracy evaluation. True positive is used to determine whether identification of each type is carried out precisely.

TABLE I.
Narrow Band Imaging (NBI) magnification findings.

Type	A	B	C3	Total
Number of images	420	420	420	1,260

$$True\ Positive(i) = \frac{Posi\_Num(i)}{Img\_Num(i)} \times 100\ [\%] \qquad (2)$$

i : Type A, Type B, Type C3
Posi_Num(i) : Number of images identified correctly.
Img_Num(i) : Total numbser of image data of i.

B. Result for 10-fold Cross-Validation

The simulation results of true positive in Equation (2) for 3-class 1-step and 2-step identifications of Fig.3 with 10 times 10-fold cross validation are shown in Fig.5. Since the 2-step

classifier [A : $\overline{A}$] has been shown the sufficient identification performance of 98% in our previous experimental results, we consider a configuration in which 3-class identifier can be performed by 2-step identification. The first step performs with types [A : $\overline{A}$] identification and the second step performs by type [B : C3] identification as shown in Fig.4 (b). Figure 5 also shows the number of support vectors (SVs) of each identifier in this experiments. By employing 2-step identifier, it is possible to reduce by about 10% SVs without losing most of the identification performance, and also can reduce the hardware cost which is the number of identifiers.

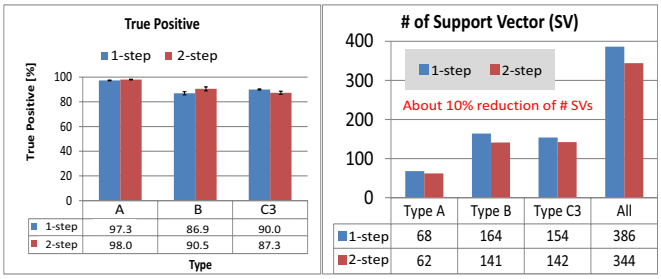


Fig.5. Simulation results of true positive and the number of support vectors for 3-class 1-step and 2-step identifications in Fig.4.

IV. Pyramid Style Hierarchical Identifier

A. Concept of Hierarchical Identification based on SVMs

Accordingly, it is difficult to catch the boundary of each type only by performing identification by single-size scan windows and to determine the image segmentation from the results. Thus we propose a hierarchical identification method using multiple-size SWs in Fig.6. In this paper, four multi-level scan window sizes, such as 60x60, 120x120, 180x180, and 240x240 pixel, are used, in which the unit scan window size (grid) is 60x60 pixel. The combination of SWs of arbitrary sizes is also possible.

In general, the multi-level scan does not necessarily need to process simultaneously. Since the sequential processing is suitable in software implementation, if the type identification result does not clearly determine in the current level scan window, then the scan window will be broken down to the lower level scan window size. This method is top-down hierarchical identification. In this paper, *a bottom-up hierarchical method* is used. We call this method as *pyramid style identification*. In the method, the type identification results of all levels are accumulated and the probability of each type of each unit region (unit scan window) is calculated. Finally an image segmentation result is obtained based on each grid probability.

This method is suitable for hardware implementation. In the method, each unit probability is calculated using SVMs learned in each scan window size which is different on each level. Therefore, compared with the case where a SVM learned in single scan window size is used for scanning image, the number of scanning windows of the whole image is substantially reduced up to 1/10 (#SWs:17555 to 1901). Moreover, it is possible to improve the resolution of image segmentation by combining multi-level scan windows.

Generally, although the identification accuracy of a small scan window becomes low, by considering the identification result of scan window for each level, the pyramid style hierarchical identification can suppress the decrease of identification accuracy. Since the multi-level SVMs in the bottom-up method can be calculated by one image scan and does not require random access to an external memory, it is suitable for the hardware implementation of the streaming processing.

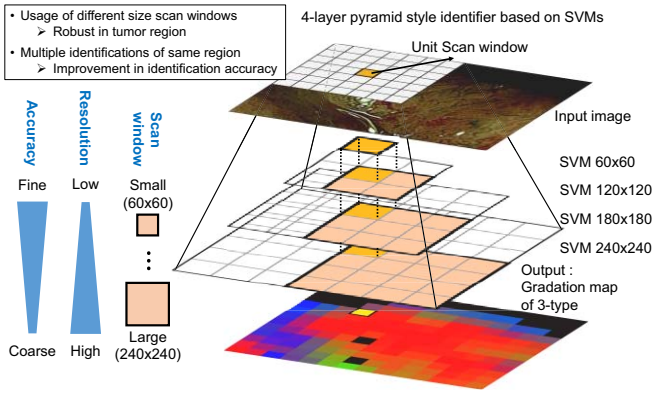


Fig.6. Concept of the image segmentation method based on bottom-up hierarchical (pyramid style) SVM identifiers.

B. Calculation of Type Probabilities with Multiple SVMs

The outline of the pyramid style hierarchical identification method which is suitable for hardware implementation is shown in Figs.7 and 8, in which the level number is assumed as three. We can obtain all input features for each level SVM by only one streaming scan using the hardware oriented feature extraction method which we have been proposed in [7, 8].

First, the type identifications are performed by each SVM for the three scan-size windows which are contained the same reference grid (60x60) (Fig.7). Next, the obtained type probabilities for overlapped grid (60x60) in each level are combined by averaging with different weights for each levels. The weight of each level SVM are determined by the true positive score (accuracy) in previous learning and testing experiments. In Fig.8, the weights of three SVMs, SVM60x60, SVM120x120, and SVM180x180 are 28%, 34%, and 37%, respectively. Then the final estimation probability of the upper-left corner grid is (P_A=0.75, P_B =0.22, P_{C3} =0.07). So this grid seems to be Type A and the gradation of the grid is calculated based on the above probability.

In this way, all grids (unit scan windows) of the input image are processed and the image segmentation result with gradation map can be obtained. Although in this example, the scan window size of each level is a multiple size of reference grid (60x60) and scan step (scan window displacement) of each level is 60 pixels, we can change them to arbitrary values. Then this can extend to obtain more complex and highly resolution gradation map.

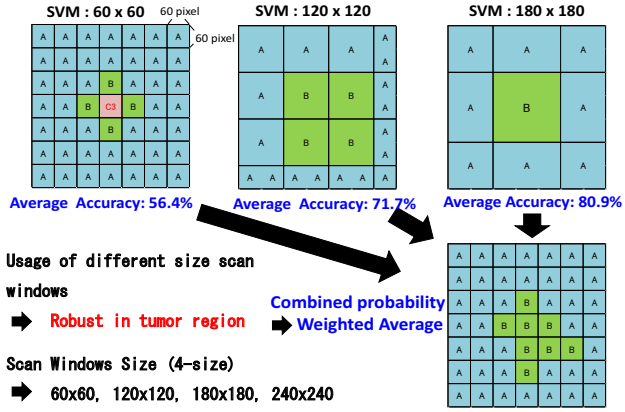


Fig.7. An example of three-level pyramid style identification.

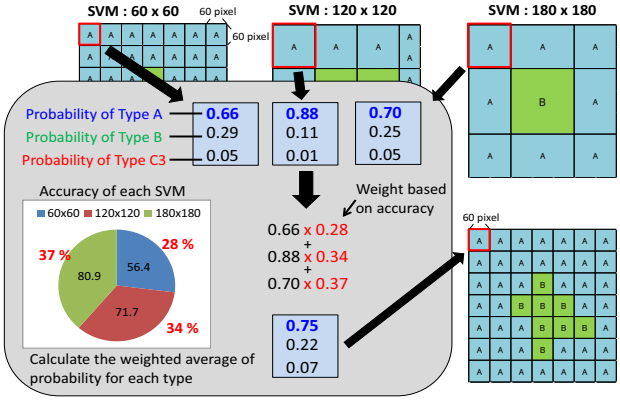


Fig.8. Calculation of probability of each gradation grid (60x60) in pyramid style identification.

C. *An approach to increase the accuracy of identification for each SW size*

In order to increase the accuracy of identification for each SW size, our approach increases the number of voting key points, and so increases the number of feature vectors for VW histogram per one scan window. In our works, sampling pitch in the feature extraction module is the distance between the key points. The total number of feature points per one SW increases by reducing sampling pitch. The number of key points in a 60x60 SW increases from 81 to 1682 if the sampling pitch changes from 5 to 1. The number of key points means the number of voting for VW histogram creation, so this number influences the latency of our system. However, we also introduce a high speed voting architecture [9] to overcome this problem. Figure 9 shows the true positive of 4 SW sizes of 60, 120, 180, and 240 on 1-step identification with sampling pitch changes from 5 to 1.

It shows that the true positive increases by narrowing the sampling pitch in some cases. For example, the true positive of type A with SW size 60x60 and that of type C3 with SW size 120x120 in Fig.9 increases when sampling pitch decrease. However, the true positive of type B in some SW sizes shows a decrease by narrowing the sampling pitch, so we have to consider and verify in more detail on 10-fold CV testing or using 2-step identification method.

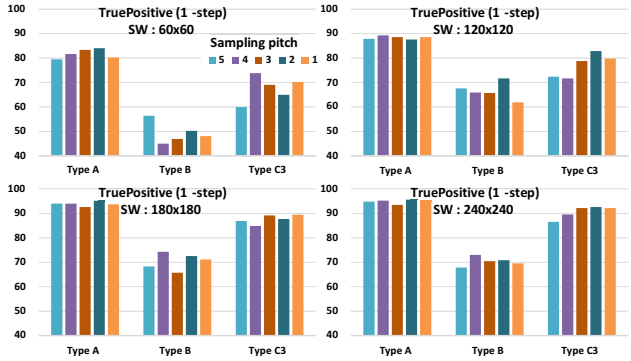


Fig.9. True positive on test evaluation for each sampling pitch.

D. *Basic Experimental Evaluation*

In order to verify the effectiveness of the pyramid style hierarchical identification method described above (bottom-up approach) and to verify the increase in accuracy of the identification, we create several artificial bench mark data by combining 60x60 (reference grid size) images. The four images of each type, which have the best accuracy on the test-evaluation are included to the learning phase with sampling pitch 5. They are rotated 90 degree to the right to create sample images of size 60x60 pixels. Those sample images are then combined to create the artificial patterns in this evaluation. The accuracies of Type A, Type B and Type C3 in those artificial patterns are about 97%, 92% and 91%, respectively. The artificial patterns which are imitated the elliptical tumor are created by pasting these 60x60 images together.

The example images are shown in Figs.10 and 11. The actual endoscopic images in Table 1 were used for learning of multiple SVMs. In this simulation, the probability of each grid is calculated by the average value of the sum of the probability of the corresponding grid in each layer. Figs.10 and 11 show the gradation maps based on the identification results. In these figures, if $P_A = 1.0$ (Type A) then the grid is colored by blue (B), if $P_B = 1.0$ (Type B) then the grid is colored by green (G), and if $P_{C3} = 1.0$ (Type C) then the grid is colored by red (R). So each gradation map is created by the combination of RGB colors based on the probability of each grid. Three numbers of each grid show the probability of each type (P_A, P_B, P_{C3}), respectively.

From these results, we have succeeded in catching the great portion of tumor in the artificial patterns. In addition, the gradation maps of sampling pitch 1 in Figs.10 and 11 are clearer than that of sampling pitch 5, respectively. Each cell's gradation is near to primary color: red (R), green (G) and blue (B). It means that smaller sampling pitch gives clearer gradation color in our proposed system. Fig.12 shown the result of the test pattern 4 for each SW size of 60x60, 120x120, 180x180 and 240x240 in sampling pitch 5 and 1, respectively. If we look in more detail in case of sampling pitch 5 in Fig.12, the identification is unclear and could not identify correctly. However, the final combination result of sampling pitch 5 in Fig.11, which is created by averaging identification results of all SW sizes with some weight shows a close identification result with the test pattern 4 in Fig.11 (left). It proves that our

proposed method can generate a high accuracy identification result even if incorrect identification occurs on individual SW size. Also, the grid color is mixed at the boundary portions of each type and the difference of each color in RGB is not so large, the probability of grid at the boundary seems to be near values. It proves that a boundary portion roughly can be identified by the proposed method. An example of the result applied to the actual endoscopic image is shown in Fig.13, with sampling pitch 5. This result also introduces the identification capability of SVM with reflection and blur.

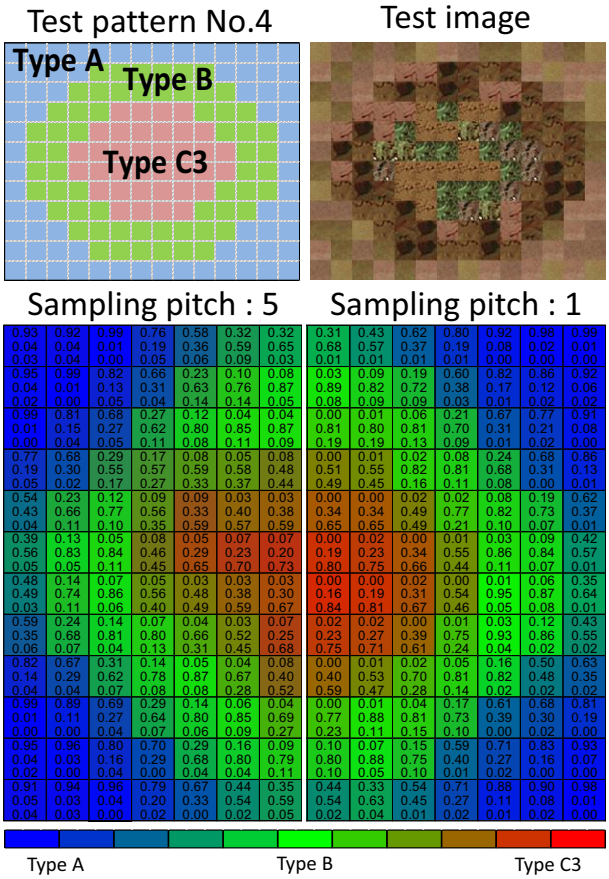


Fig.10. Result of the artificial images (Test pattern 4).
Bottom left : Sampling pitch 5. Bottom right : Sampling pitch1

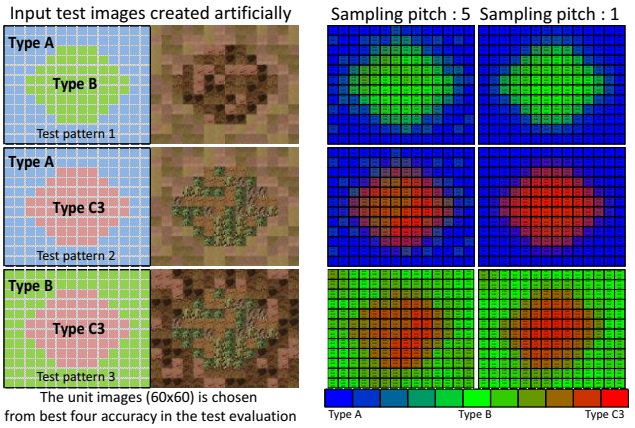
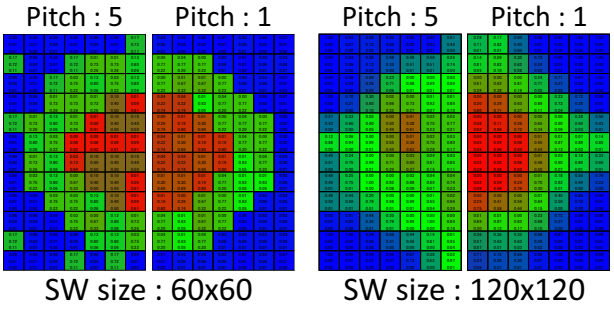
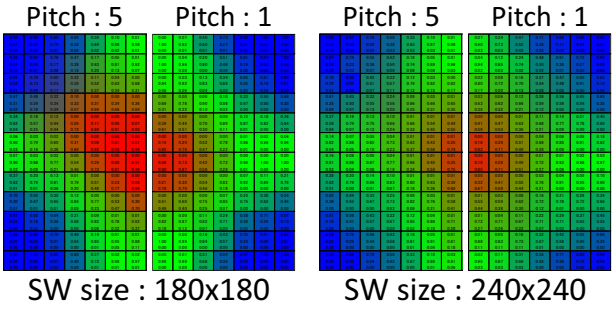


Fig.11. Result of the artificial images (Test pattern 1-3).



(a). left : 60x60 and right : 120x120.



(b). left : 180x180 and right : 240x240.

Fig.12. Simulation result of the artificial test pattern No.4.

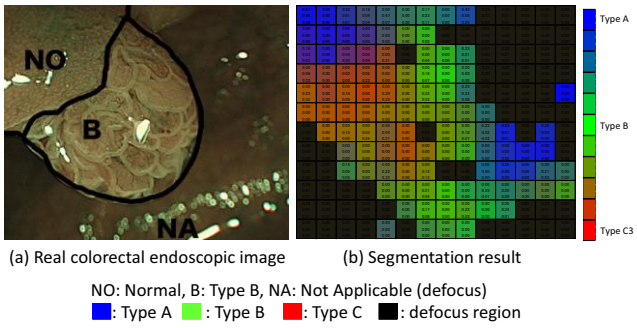


Fig.13. Results of the actual endoscopic image of the SVM with identification capability of reflections and blur.

V. FPGA Implementation of 2-classes SVM Identifier

We have implemented the SVM based 2-class identifiers into FPGA (Altera StratixIV EP4SE360F35C2). The platform has been developed by our group for hardware and software co-design system [10]. The PROC Wizard [11] development tool is used for the interface code generation between host (PC) and FPGA board. Table 3 shows the implementation results and resource utilizations for the number of processing units of inner product calculation part [12]. From this results, it is possible to achieve throughput > 21.2 fps@100 MHz and latency < 47.2 msec@100 MHz in the processing of a Full HD (1920 x 1080) image (Condition: SW: 120x120, Scan-Step: 10 pixel, #SWs: 17,557). In order to implement the pyramid style identifier, we need to consider the number of scan windows in each layer, that is, the tradeoff between parallelism and hardware size. In our simulation, the number of scan windows at scan-step: 60 pixel are 576 (60x60), 527 (120x120), 480 (180x180), and 435 (240x240), respectively. Since the total

number of SWs to be processed in pyramid style identifier is 1,901, about 1/10 of the conventional method, it can be implemented with high throughput and high accuracy for actual complex colorectal endoscopic images.

We calculated the pyramid style identification processing speed per Full HD image in worst case to verify the latency of our system approximately. In our system, SW size 60x60 SVM starts its scan process first, then SW size 120x120, 180x180 and 240x240 start their process latter (Fig.14(a)). In the case of the operating frequency is 100 MHz and one pixel data is read in one clock, waiting time for type segmentation data preparation for SW size=60x60 is about 1920x60=1.2 msec. After that, type segmentation starts and finished after 28 msec for image processing with SW size=60x60 shown in Fig.14(b). Data preparation time and type segmentation time for other SW sizes are also shown in Fig.14 (b).

We assume that the number of support vectors is 300 and the parallel inner product processing units used by each SW size is 150. Since the four SW sizes work in parallel, the total number of inner product processing units (DSP) used in this case is 600. The worst case approximate calculation in Fig.14(b) shows that the latency in the type identification module is about 29.2 msec. It proves that our proposed method can perform real time identification. This approximate calculation is used in the worst case, so optimized implementation can give improvement on hardware size and latency. In order to achieve an optimized implementation, the number of DSP units used by each SW size should be changed to make balance in latency among those SW size processing and so increase the final latency.

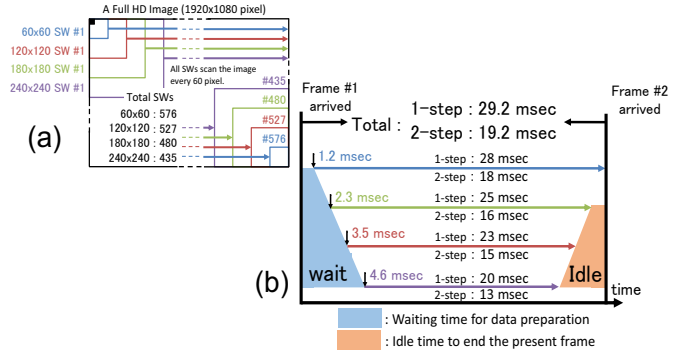


Fig.14. (a) Stream processing a Full HD image for each SW size, (b) An approximate calculation of worst case latency both 3-classes identification techniques.

TABLE III.
Resource utilization of Stratix IV FPGA..

		Altera Stratix IV Available	Number of Inner Product Processing Units							
			50		100		200		500	
Max Frequency [MHz]			150.1	%	95.2	%	136.4	%	112.7	%
Resource Usage	# ALUT	424,960	2,689	0.63	5,190	1.22	13.641	3.21	250.552	58.96
	# registers	424,960	4,344	10.2	8,596	2.02	17,098	4.02	42,600	10.02
	# Logic Array Blocks	21,248	437	2.06	572	2.69	1,187	5.59	5,930	27.91
	# M9K RAM Blocks	1,280	102	7.97	202	15.78	402	31.41	1,001	78.20
	# 18 x 18 DSP Units	1,024	52	5.08	102	9.96	202	19.73	502	49.02

VI. Conclusion

This paper presents the hardware design of type identifier in CAD system for colorectal endoscopic images with NBI magnification for real-time processing of Full HD images. The newly proposed pyramid style identifier with SVMs for multi-size and multi-layer scan windows enables smaller circuit area and higher throughput with higher identification accuracy than the conventional method for actual complex colorectal endoscopic images. We verified the increase of the identification accuracy of pyramid style identification method by the simulation with various sampling pitch. The proposed method also could achieve high accuracy in type segmentation even if individual SW sizes could not identify correctly by combining segmentation results from other SW sizes. The proposed CAD system with the gradation map for the corresponding colorectal endoscopic image could be very effective as a diagnostic support system for both non-expert and expert doctors.

Acknowledgement

Part of this work was supported by Grant-in-Aid for Scientific Research (C) and Research (B) JSPS KAKENHI Grant Numbers 2459102 and 26280015, respectively.

References

[1] T. Tamaki, et al., “Computer-aided colorectal tumor classification in NBI endoscopy using local features, medical image analysis,” Available online 13 September 2012, ISSN 1361-8415, 10.1016/j.media. 2012.08.003. (2012).

[2] H. Kanao, et al., “Narrow–band imaging magnification predicts the histology and invasion depth of colorectal tumors,” Journal of Gastrointestinal Endoscopy, vol. 69, no.3, pp. 631-636, 2009.

[3] The VLFeat open source library, http://www.vlfeat.org/

[4] V. Vapnik, “Statistical learning theory,” John Wiley & Sons Inc., New York, 1998.

[5] C. Platt. John, et al., “Large margin dags for multiclass classification,” Neural Information Processing Systems 12 (NIPS1999), pp. 547-553, 1999.

[6] LIBSVM:A Library for Support Vector Machines
http://www.csie.ntu.edu.tw/~cjlin/libsvm/

[7] T. Mishima, et al., “A simple and effective hardware oriented feature extraction algorithm for colorectal endoscopic images with NBI magnification,” Proc. of the 28th International Conference on Circuits / Systems, Computers and Communications (ITC-CSCC2013), pp. 567-570, 2013.

[8] T. Mishima, et al., “FPGA implementation of feature extraction for colorectal endoscopic images with nbi magnification,” Proc. of the IEEE International Symposium on Circuits and Systems (ISCAS2014), pp.2515-2518, 2014.

[9] K. Sugi, et al., “Consideration for Acceleration of Feature Transformation based on the Bag-of-Features for Colorectal Endoscopic Images,” IEICE Tech. Rep., vol. 114, no. 302, CPSY2014-55, pp. 7-12, Nov. 2014.

[10] M. Omori, et al., "HW/SW Co-design of Region Growing Image Segmentation", Proceedings of the 26th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC2011), pp. 322-325, 2011.

[11] PROCWizard, http://www.gidel.com/procwizard.htm.

[12] T. Koide, et al., “FPGA Implementation of Type Identifier for Colorectal Endoscopic Images with NBIMagnification,” Proceedings of the 12th IEEE Asia Pacific Conference on Circuits and Systems (APCCAS 2014), pp. 651–654, 2014.