

An FPGA Implementation of SVM for Type Identification with Colorectal Endoscopic Images

Takumi Okamoto^{1,*}, Tetsushi Koide¹, Anh-Tuan Hoang¹, Tatsuya Shimizu¹, Koki Sugi¹,
Toru Tamaki², Tsubasa Hirakawa², Bisser Raytchev², Kazufumi Kaneda²,
Shigeto Yoshida³, Hiroshi Mieno³, Shinji Tanaka⁴

¹ Research Institute for Nanodevice and Bio Systems (RNBS), Hiroshima University,

² Department of Information Engineering, Graduate School of Engineering, Hiroshima University,

³ Department of Gastroenterology, Hiroshima General Hospital of West Japan Railway Company,

⁴ Department of Endoscopy, Hiroshima University Hospital,

1-4-2 Kagamiyama, Higashi-Hiroshima 739-8527, Japan

*okamoto-rnbs@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 (1920 x 1080 pixel). A pyramid style image segmentation with SVMs for multi-size scan windows, which can be implemented on an FPGA with small circuit area and achieve high accuracy, is proposed 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]. 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).

Currently our software implementation of the system is able to identify with only the region (we call scan window) as small as 120x120 pixels at 14.7 fps and it takes about 20 minutes to scan and process a whole Full-HD (1920x1080) image. For further speed improvement for high resolution image, a hardware realization is indispensable because the computation time of software implementation is exponentially increased with the increase of image size. As a demand on a clinical doctors, the proposed CAD system satisfies the throughput of 1 - 5 fps and the latency is at least 1 sec for on-the-fly diagnostic supporting.

II. Overview of Computer-Aided Diagnosis System

Outline of the proposed CAD system is shown in Fig. 2. The system is based on a Bag-of-Features (BoF) representation of local features in the endoscopy image.

The system has two stages, learning and testing. The

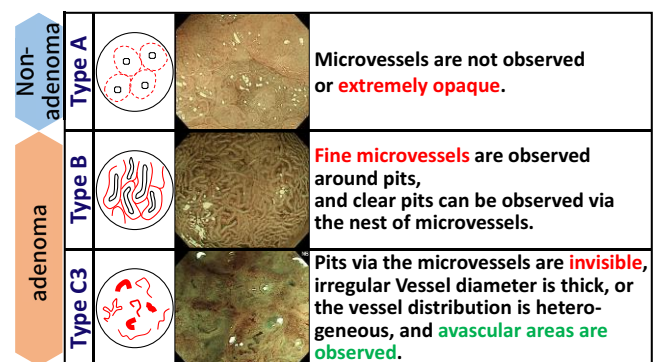


Figure 1. Narrow Band Imaging (NBI) magnification findings [1].

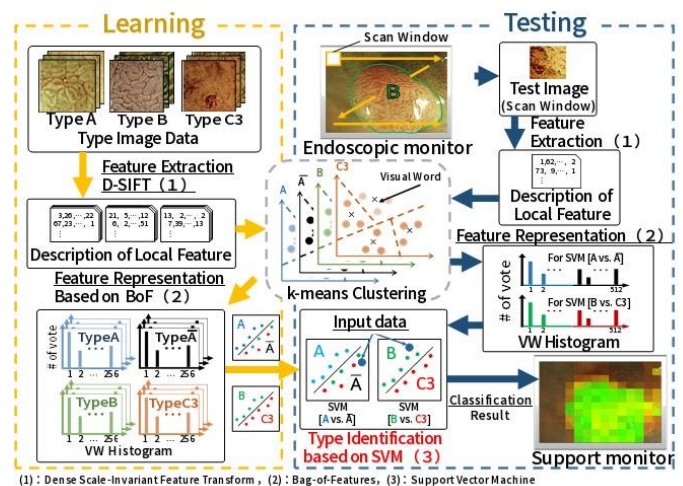


Figure 2. Computer-Aided Diagnosis system for endoscopic images.

overview of processing flow of the system is as follows.

First, the local feature quantities of the endoscopy images in each type are extracted based on Dense Scale-Invariant Feature Transform (D-SIFT) algorithm [3] because the pit patterns of endoscopic images (Fig. 1) are very complex and irregular comparing with object recognition such as face and pedestrian recognitions.

Then the features obtained from the learning phase are clustered and the center of each cluster is saved as a Visual

Word (VW) for each type, which are used for feature representation using k-means clustering. In the classifier module, support vector for support vector machine (SVM) is obtained at the learning phase using the type information of leaning image which is judged by the professional doctors.

Next, in the testing phase, the D-SIFT based feature extraction is performed for a whole input image and a visual-words histogram is created by voting for the nearest VW. Then the CAD system classifies the testing image within an endoscopy movie (frame) by pre-learned SVM.

Finally, a color gradation map which is converted from the result of classifier for each SW displays for doctor as a “second opinion” (Fig. 3). In our software implementation, D-SIFT of Library VLFeat [3] is used for the feature extraction and Support Vector Machine (SVM) of LIBSVM [4] is used for type identification.

Also, as a utility of the real-time image recognition system, our research group applied the system to real clinical test. In the learning phase, a set of 2,247 cutout endoscopic images which were collected by a technical doctor are used. The learning images are separable into 504 Type A images and 1,743 Type B and C3 images, so in this experiment, the system identifies into type A or B&C3. The feature quantities extracted from D-SIFT algorithm are 128-dimensional vectors. Then, an output of the system is the SVM output value, and a cutoff value of the SVM output value is 0.5. Figure 4 shows that the result of evaluation and from this result, our group verified that the SVM output value of neoplastic lesions could identify the non-neoplastic lesions or neoplastic lesions [5].

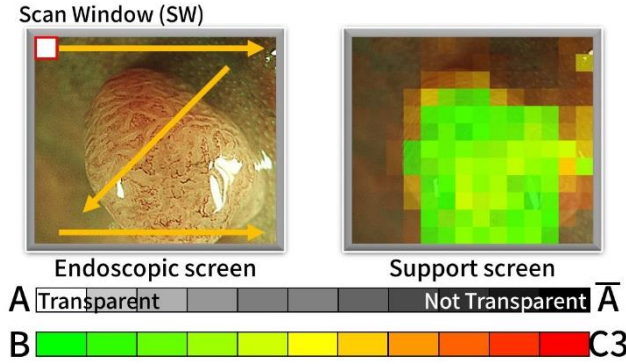


Figure 3. An Identification result for Type B NBI endoscopic image.

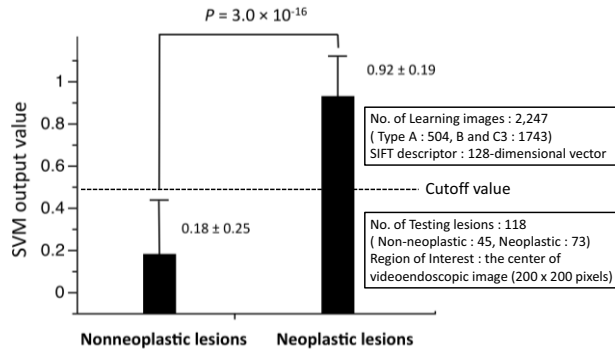


Figure 4. Support vector machine (SVM) output values for the images of non-neoplastic and neoplastic lesions [5].

III. Type Identifier 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.

$$d_{Y:Z}(\vec{x}) = \sum_{i=1}^{N_Y+N_Z} coef_i \times \overline{sv}_i \cdot \vec{x} + \rho_{Y:Z} \quad (1)$$

In which, x is a test data, 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 sv_i , and $\rho_{Y:Z}$ is a coefficient of a decision function. If the identification function $d_{Y:Z}(x)$ is a positive value (>0), an input data x is determined as the class Y .

In this work, it is better for the hardware implementation to calculate the decision function by fixed-point number rather than floating point one. Verification on identification 18-bit fixed-point number, with 16 bit fraction is acceptable without decrease in accuracy compared with identification using floating-point number [6].

In order to realize 3-class (3-type) identification, we adopt probability estimation technique with two recognition stages [6]. The first stage evaluate type A or B&C3, the second stage evaluate type B or C3 (we call 2-step evaluation).

B. Prediction Rate

Prediction rate shows that a result of identification from input data x is represented as a percentage each type [7]. Prediction rate is computed from the result of decision function as shown as eq. (1), and its distribution map with decision function values draws as sigmoid curve.

C. Pyramid-style Type Identification

Accordingly, due to the complexity of the boundary of each type as shown in Fig. 5, 6 and 7, upper left part, 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 proposed a pyramid-style identification method using multiple-size SWs in Fig.8 [8]. In this paper, four multi-level scan window sizes, such as 60x60, 120x120, 180x180, and 240x240 pixel, are used. 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

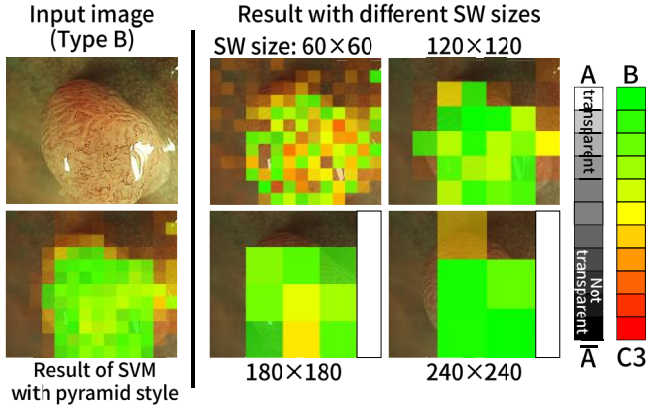


Figure 5. Result of the real endoscopic image (Type B).

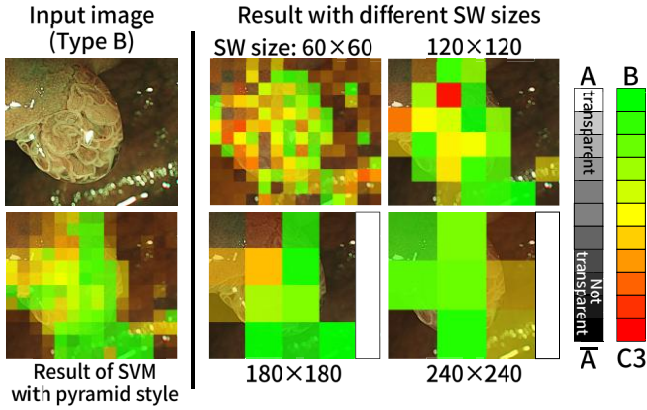


Figure 6. Result of the real endoscopic image (Type B).

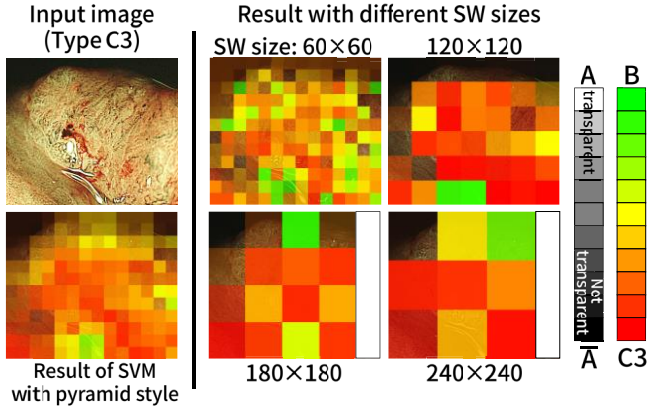


Figure 7. Result of the real endoscopic image (Type C3).

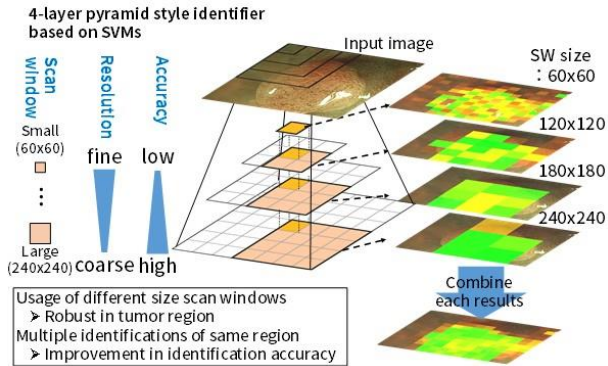


Figure 8. Concept of the image segmentation with 4-layered pyramid style SVM.

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 prediction rate.

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.

IV. An Architecture of Decision Function Calculator

Figure 9 shows the block diagram of the decision function module in our type identification module. In the case of the 2-step evaluation method, we calculate two decision function, $d_{A:\bar{A}}(\vec{x})$ and $d_{B:C3}(\vec{x})$, and so we need two decision function calculation modules. Since high dimension feature quantity (512 dimensions) is processed in our medical application, it is necessary to perform the sum-of-product operations, which has included the decision function (eq.1), in parallel.

Hence, there is a subject that how to reduce hardware resources while increasing parallel degree. The decision function calculation module is implemented using 2 pipelined stage architecture. The 1st stage is used for sum-of-product computation and the 2nd stage is used for summation. Support vector (SV) and coefficients obtained by learning step are read from memories. The timing chart of the calculation one scan window is shown in Fig.10. The sum-of-product calculation in

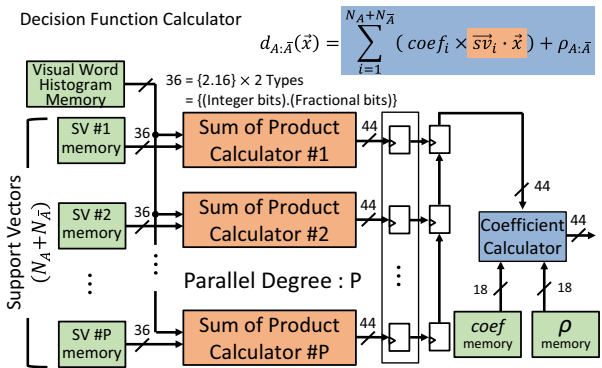


Figure 9. Decision function calculation architecture.

the decision function is computed by sum-of-product calculator blocks which are colored to pale red in parallel. When the 1st sum-of-product calculation is finish, the other support vectors are read from sv memory for the next sum-of-product calculation. At the same time, coefficient calculator reads the result of each sum-of-product, multiplies it with $coef_i$ before summing them up for summation. Finally, the coefficient $\rho_{A:\bar{A}}$ is added to compute the calculation of the decision function.

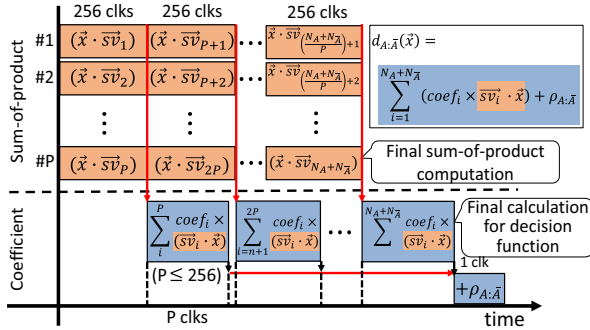


Figure 10. Timing chart of calculating the decision function.

V. An Architecture of Prediction Rate Calculator

Prediction rate is calculated repeatedly till minimizing error, so it is not taken constant processing time [7]. Therefore, it is desirable to develop hardware algorithm which can compute the prediction rate in fixed processing time for realization hardware implementation. Thus, distribution curve is approximated to linear as shown as eq. (2). The original sigmoid curves are colored by gray, and approximated lines are colored by blue and green, $d_{A:\bar{A}}(\vec{x})$ and $d_{B:C3}(\vec{x})$, respectively as shown as Fig. 11.

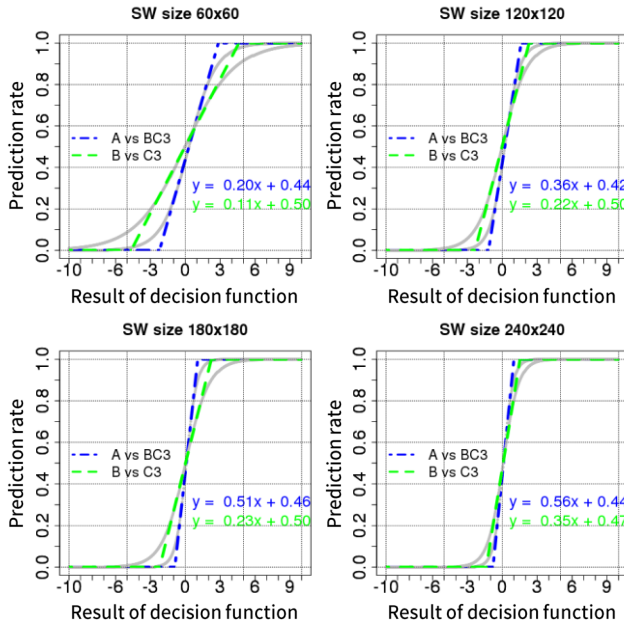


Figure 11. Timing chart of calculating the decision function.

Prediction Rate

$$= \begin{cases} 1 & (decvalue \times a + b > 1) \\ decvalue \times a + b & \\ 0 & (decvalue \times a + b < 0) \end{cases} \quad (2)$$

We also suggest a hardware-oriented method for probability estimation using 5-bit 2-stage look up table. A result of verification shows that the proposed probability estimation with 2nd stage in the range of 0.45 ~ 0.55 performs the same accuracy ($\pm 2\%$) with software implementation. Figure 12 shows a prediction rate calculation architecture with 2-stage look-up table.

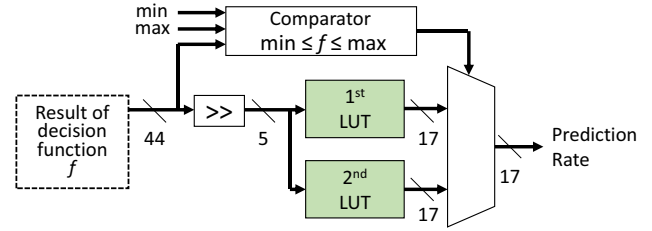


Figure 12. A prediction rate calculation architecture.

VI. Performance Verification of CAD system

We have implemented the decision function calculation architecture on FPGA, Altera Stratix IV (EP4SE530H35C2) device. The occupied resources several number of parallel multiplier DSPs, 32, 64 and 128, and processing time are shown in Table 1. Also, we verified that a performance of the whole CAD system which computes one Full-HD endoscopic image from input an image data into feature extraction module [8] to output the result from type identification module via a feature transformation module [9] as shown as Fig. 13.

The processing time of feature transformation module in Fig. 13 is a critical path in our system, so the number of parallel multipliers as shown as Fig. 13 at the bottom can reduce to 64 in the same processing time from 128. From this performance verification results, throughput is 16.7 fps and latency is 60 msec at 100 MHz. So it is achievable about the real time and on-the-fly diagnostic support for clinical doctor (demand throughput: >5 fps, latency: <1 sec).

VII. Estimation Platform Construction for SVM Identifier with Hardware-Software Co-design

We constructed an estimation platform with hardware-software co-design for the proposed SVM architecture. Overview of the platform used in estimation is shown in Fig. 14 and 15. Real endoscopic recorded movie which provided from Hiroshima University hospital is input to the system via HD-SDI (Serial Digital Interface) cable. We use a Decklink mini recorder as a capture board, and a ProceIV 530-A Development board for implementation. The development environment is shown in Table 2.

Once, a scan window (SW) is trimmed from the input frame, and feature vectors are extracted by feature extraction module. The feature vectors are transformed to Visual Word (VW)

histogram, and input to the type identification module implemented on FPGA. PCI Express Gen3 x4 connection between PC and FPGA board using Direct Memory Access (DMA) method. Other SWs of the input frame are processed in the same manner.

Table 3 shows the comparison in processing time of the type identification module (shown in red rectangle on Fig. 15) to process one SW in hardware and software, in which, processing time on hardware is defined as time for VW histogram transmission from software to hardware using DMA method and computation time on hardware. Since processing time of one SW relies on the number of dimensions of VW histogram, this time is unchanged regardless the size of each SW.

Type identification time of one Full-HD frame with pitch 60 and pitch 30 pixels are shown in Figs 16 and 17. The parallel degree P is set to 64. The total processing time of one frame lineally increases with the number of SW inside the frame. Hence processing time for frame with bigger SW size gradually decrease in both hardware and software implementation, from 771 msec to 667 msec and from 35.7 msec to 30.7 msec on software and hardware implementation, respectively for raster scan with pitch 60 pixels as shown in Fig. 16. These time increase 4 times when raster scan with pitch 30 pixels as shown in Fig. 17. In all the situation, the type identification implemented on FPGA is 20 times faster than that on software.

Table 1. Resource Utilization of the decision function.

Altera Stratix VI Available (EP4SE530)	No. of parallel multiplier DSPs						
	32		64		128		
# ALUTs	424,960	2,154	1%	2,992	1%	8,079	2%
# registers	424,960	2,256	1%	3,759	1%	7,200	2%
# Logic Array Blocks	21,248	207	1%	299	1%	795	4%
# M9K RAM Blocks	1,280	512	40%	520	41%	512	40%
# M144K RAM Blocks	64	64	100%	64	100%	64	100%
# 18 x 18 DSPs	1,024	288	28%	544	53%	1024	100%
Max Frequency [MHz]	163.9 MHz		133.0 MHz		122.61 MHz		
Processing time (a Full-HD image)	15,368 clocks		7,696 clocks		3,872 clocks		

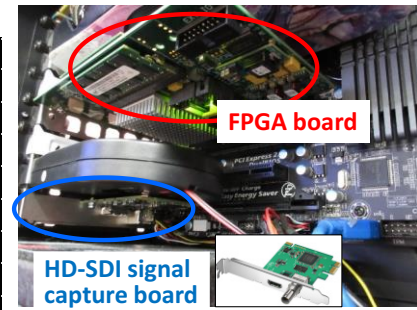


Figure 14. FPGA and Capture board.

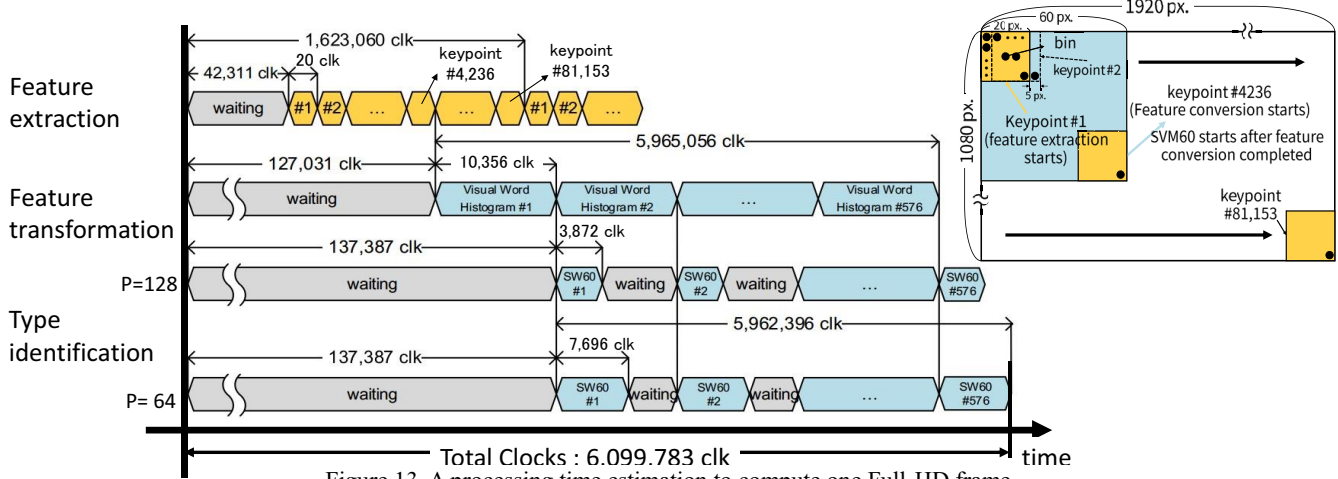


Figure 13. A processing time estimation to compute one Full-HD frame.

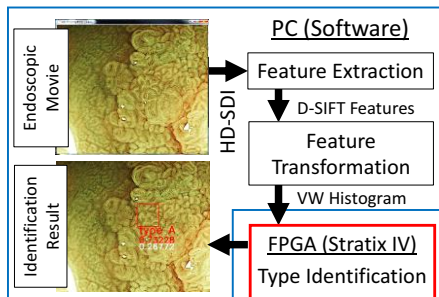


Figure 15. Evaluation Environment Overview.

Table 2. Elements of Evaluation Environment.

FPGA Board	PROCe IV 530-A (GIDEL)
Memory	On-board DDRII 512MB DDRII SODIMM 1GB x2
Installed FPGA	Altera Stratix IV EP4SE530H35C2
Host I/F	PCI-Express Gen 3.0
OS	Windows7 Enterprise SP1 64bit
CPU	AMD FX-8120 8Core @3.10GHz
Memory	DDR3-1333 8GB x 4
Language	C++
IDE	Microsoft VisualStudio 2012
Library	OpenCV 3.0

Table 3. Processing Time Estimation of a SW.

ScanWindow Size	60 (msec)	120 (msec)	180 (msec)	240 (msec)
Hardware	0.06	0.06	0.06	0.06
Software	1.2	1.2	1.2	1.2

VIII. Conclusion

In this paper, we introduce the hardware architecture of type identification module with SVM for endoscopic images. Estimation on processing time shows that the system can process one frame in 61 msec@100 MHz (with 64 parallel multiplier DSPs). We also evaluate the processing time of a scan window on the evaluation environment using hardware-software co-design method. The processing time of type identification module for one scan window on hardware is about 0.06 msec, and is about 20 times faster than that of scan window implementation, which needs 1.2 msec to process one scan window. Processing time of type identification for one Full-HD frame is also verified.

Future work includes the development of the whole CAD system including our D-SIFT architecture [8] and our feature transformation architecture [9] in one FPGA board.

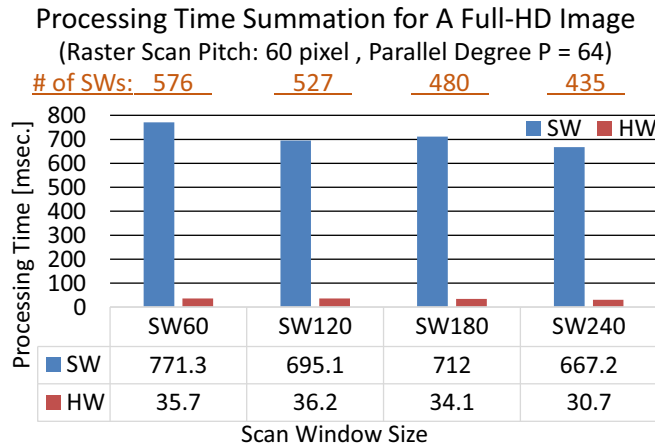


Figure 16. Result of Processing Time Evaluation (60 px pitch).

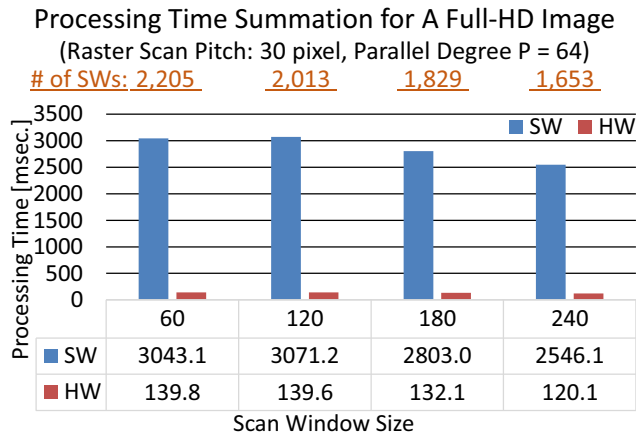


Figure 17. Result of Processing Time Evaluation (30 px pitch).

Acknowledgement

Part of this work was supported by Grant-in-Aid for Scientific Research (B) JSPS KAKENHI and JSPS Fellows, Grant Numbers 26280015 and 16J06130, respectively, and was with the help of a grant by Chugoku Industrial Innovation Center. The FPGA design tools in this work have been supported by the Altera University Program, the Mentor Graphics Higher Education Program and Cadence Design Systems, Japan.

References

- [1] 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.
- [2] T. Tamaki, et al., "Computer-aided colorectal tumor classification in NBI endoscopy using local features", *Medical Image Analysis*, Vol. 17, No. 1, pp. 78-100, 2013.
- [3] A. Vedaldi, and B. Fulkerson, "Vlfeat: an open and portable library of computer vision algorithms," <http://www.vlfeat.org/>
- [4] V. Vapnik, "Statistical learning theory," John Wiley & Sons Inc., New York, 1998.
- [5] Y. Kominami, et al., "Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy," *Gastrointestinal Endoscopy*, Vol. 83, Issue 3, pp. 643-649, 2016.
- [6] S. Shigemi, et al., "An FPGA implementation of support vector machine identifier for colorectal endoscopic images with NBI magnification," *Proc. of the 28th Intl. Conf. on Circuits / Systems, Computers and Communications*, pp. 571-572, 2013.
- [7] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *Journal of Machine Learning Research*, 5:975-1005, 2004.
- [8] T. Koide, et al., "A hardware accelerator for bag-of-features based visual word transformation in computer aided diagnosis for colorectal endoscopic images", *Proc. of the 31th International Technical Conference on Circuits/Systems, Computers and Communications 2016*, pp. 233-236, July 2016.
- [9] T. Koide, et al., "A hardware accelerator for bag-of-features based visual word transformation in computer aided diagnosis for colorectal endoscopic images", *Proc. of the International Technical Conference on Circuits/Systems, Computers and Communications*, pp. 387-390, July 2016.