Compact and High-Speed Hardware Feature Extraction Accelerator for Dense Scale-Invariant Feature Transform

Tetsushi Koide[1], Takumi Okamoto[1], Tatsuya Shimizu[1], Koki Sugi[1], Anh-Tuan Hoang[1],
Toru Tamaki[2], Bisser Raytchev[2], Kazufumi Kaneda[2],
Shigeto Yoshida[3], Hiroshi Mieno[3], and Shinji Tanaka[4]
[1] Research Institute for Nanodevice and Bio Systems (RNBS), Hiroshima University
[2] Graduate School of Engineering, Hiroshima University
[3] Department of Gastroenterology, Hiroshima General Hospital of West Japan Railway Company
[4] Department of Endoscopy and Medicine Graduate School of Biomedical and Health Science, Hiroshima University

1-4-2 Kagamiyama, Higashi-Hiroshima, 739-8527, Japan
E-mail: koide@hiroshima-u.ac.jp

Abstract: This paper presents a D-SIFT based feature extraction hardware accelerator used in a real-time computer-aided diagnosis (CAD) system for endoscopic images. The FPGA implementation demonstrates that the proposed hardware oriented D-SIFT architecture was very compact due to no multipication is used and was very suitable for stream based image processing. The processing time for Full-HD (1920x1080) high resolution image is only *20 msec@100 MHz* and it is about *700 times faster* than that of software implementation (14 sec). The proposed D-SIFT accelarator can be also applicable for the feature extraction part for various types of image processing including 4 K and 8 K high resolusion images.

1. Introduction

With the increase in the number of colorectal cancer patients, systems which support a doctor's diagnosis have been researched. The CAD system for colorectal endoscopic images with NBI magnification [1] has already been proposed [2]. The proposed CAD system identifies 3 types of colorectal endoscopic image (Type A, Type B, and Type C3) as shown as 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 expo- nentially 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.

In this paper, we propose a hardware accelerator with an FPGA by implementing our hardware oriented Dense Scale-Invariant Feature Transform (D-SIFT) architecture [3] for up to 4K and 8K image sizes within real-time processing.

2. Outline 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 overview of processing flow of the system is as follows.

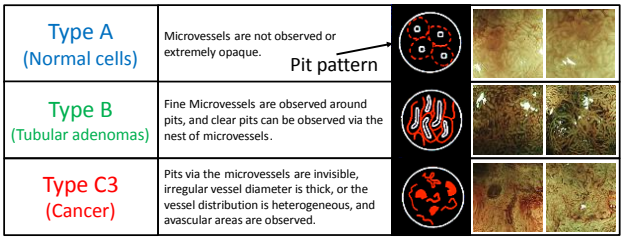


Figure. 1. Narrow Band Imaging (NBI) magnification findings [1].

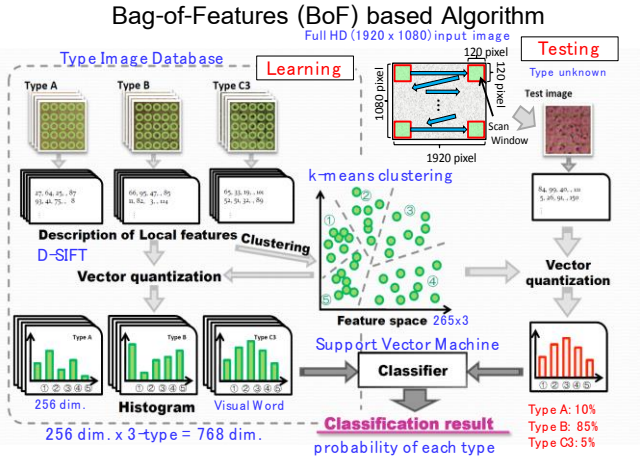


Figure.2. Computer-Aided Diagnosis System for Endoscopy Image.

First, the features 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 perfomed 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 a 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". In our software implimentation, 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.

3. Hardware Oriented D-SIFT Algorithm

We have proposed the original D-SIFT algorithm [5] so as to achive a stream based processing and a multiplication less implimentation by the following three considarations: (1) Improvement of the Gradient Direction Calculation Processing, (2) Simplification of the Weight for Convolution Processing and Convolution Features Sharing, and (3) Normalization Replacement by Threshold Processing.

3. 1 Improvement of the Gradient Direction Calculation Processing

In our system, feature quantities of endoscopic images are extracted based on the direction of a gradient and the intensity of luminosity. The gradient of the luminosity value to *x*-direction is defined as *Gx*, and the gradient of the luminosity value to *y*-direction is defined as *Gy*. As shown in Fig. 3, the direction of a gradient and the intensity are calculated using *Gx* and *Gy* for each pixel. The most accurate method (original implementation) calculates the $Tan^{-1}(\frac{Gy}{Gx})$ angle and the gradient intensity of luminosity before assigning them to the 8 directions as shown in Fig. 4 (a). However, this calculation method is very complicated, and is not suitable for hardware implemen-tation.

In our implementation, the gradient of the luminosity value of the pixel are roughly classified into 4 directions according to the sign of *Gx* and *Gy*, and then, each rough direction is finely divided into two by comparing the absolute values of *Gx* and *Gy* as shown in Fig. 4 (b), equation (1), (2), and Table 1, respectively. By that, the pixels are classified in 8 directions based on their signs and absolute values in *x* and *y* directions. The number of dimensions relies on the number of directions that a gradient is divided.

$$Tan(0) = 0 < \frac{|Gy|}{|Gx|} < Tan\left(\frac{\pi}{4}\right) = 1 \qquad (1)$$

$$\Leftrightarrow \ |Gy| < |Gx| \qquad (2)$$

3. 2 Simplification of the Weight for Convolution Processing and Convolution Features Sharing

The gradient intensity of each computed direction is convolved. This convolution occurs with all blocks in each feature description unit. The coefficient for convolution process relies on the distance from the central point to the corresponding block as shown in Fig. 5.

A feature description unit is generated by multiplying the weighted factor in Fig. 5 with the corresponding 4 x 4 blocks. Hence, the feature of each block in a feature description unit is different from that in other feature description unit. By omitting this process, weighted features of the same block in all feature description units are similar regardless the feature description unit it belong to. Hence, those feature value are sharable among different units as shown in Fig. 6.

3. 3 Normalization Replacement by Threshold Processing.

The convolved block values are normalized in a unit of 16 blocks. This aims at obtaining the same feature quantity regardless of changing in brightness, as shown in Fig. 7. However, normalization process needs multiplication, division, and square route computation with many inputs, as shown in equation (3) and Table 2. In our implementation, the difference of the luminosity value was controlled by performing threshold comparison as shown in equation (4) and Table 2 at the output value of Gaussian Filter processing (Fig. 8). By applying threshold processing, we can omit the normalization process to reduce hardware size as well as shorten the computation critical path.

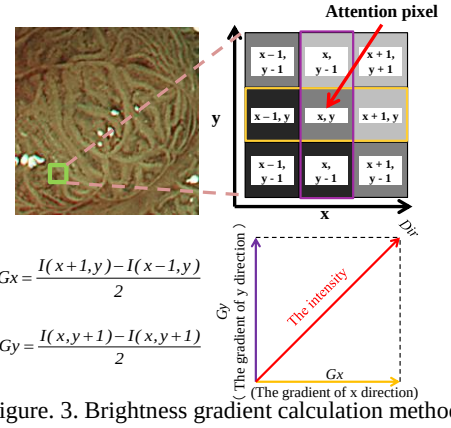


Figure. 3. Brightness gradient calculation method.

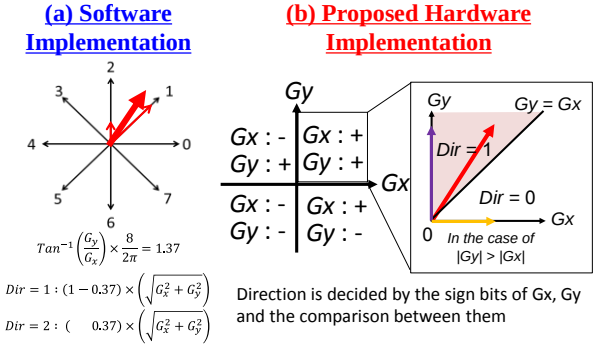


Figure. 4. Gradient direction calculation method.

Table. 1. Parameter of the Gradient Direction Calculation..

I (i, j)	Luminosity value at (*i, j*)
Gx	The gradient of the luminosity value to x-direction
Gy	The gradient of the luminosity value to y-direction
Dir	Value to express an gradient direction

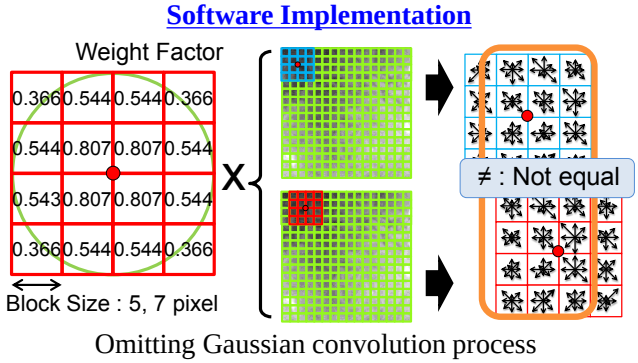


Figure. 5. Convolution Process by software implementation.

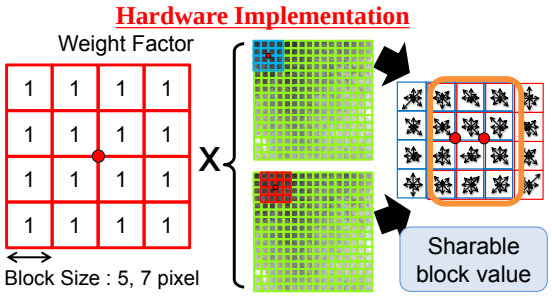


Figure. 6. Convolution Process by hardware implementation.

$$Dst_n = \frac{Tmp_n}{\sqrt{Tmp_1^2 + Tmp_2^2 + \cdots + Tmp_{128}^2}} \quad (3)$$

$$Img''(x,y) = \min\{Img'(x,y) \gg 10,\ 255\} \quad (4)$$

Table. 2. The parameters of normalization.

n	The number of dimensions
Tmp_n	The value of dimension n before normalization
Dst_n	The value of dimension n after normalization
Img''	The value which smoothed the input image
Img'	The value which performed threshold processing

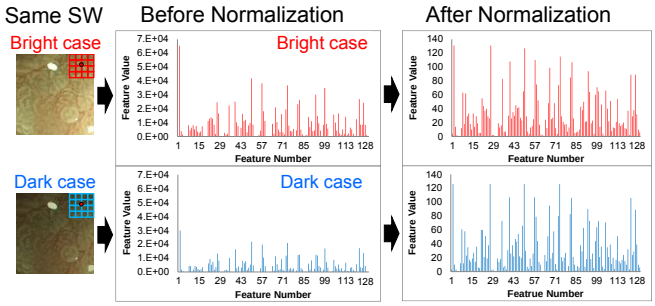


Figure. 7. Aims of normalization.

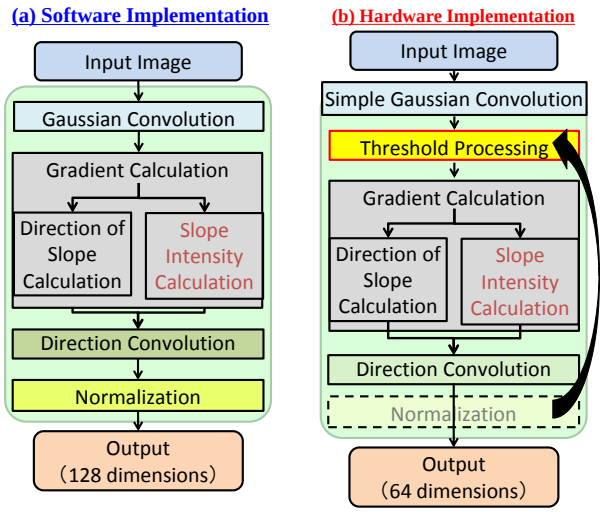


Figure. 9. D-SIFT Flow Chart and Proposed threshold processing for normalization replacement.

4. The Proposed D-SIFT Architecture

The hardware architecture of the proposed algorithm is shown in Fig. 10. The architecture consists of four units, (a) simple gaussian filter processing unit, (b) gradient calculation unit, (c) direction and intensity of gradient calculation unit and (d) directions convolution unit. In the proposed architecture, pipeline processing is realized by using FIFO. Moreover, by reducing the number of directions of the gradient from 8 to 4, the amount of memories is reduced by about 20%. In addition, pipeline processing is realized by performing block line gradient computation. Each pixel will get to the system for block line gradient computation before storing into block buffer. When the pixel in the next line of the same block comes, the corresponding intermediate block line gradient is read from the block buffer to continue the gradient computation for that block. In the system, two block sizes of 5- and 7-pixel are necessary, unit (c) and (d) in Fig. 10 are duplicated for Block Size = 5, 7 pixels processing in parallel.

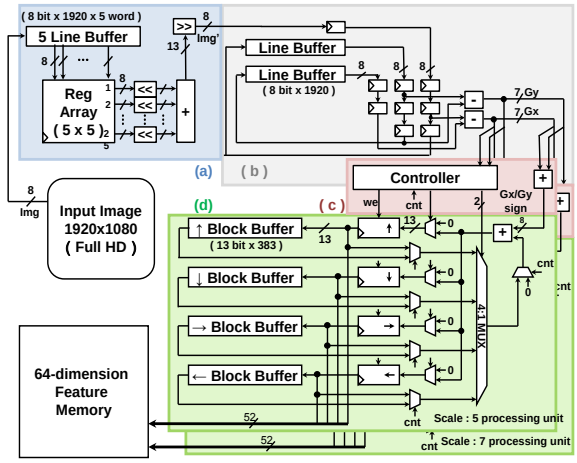


Fig. 10: Compact and high-speed D-SIFT architecture.

5. FPGA Implementation and Evaluation

We have implemented the D-SIFT architecture on FPGA, Altera Stratix IV (EP4SE530H35C2) device. The occupied resources and processing time are shown in Fig. 11. The DSP (Digital Signal Processing) block in Altera's FPGA is the dedicated block used to calculate the fixed-point multiplication in high speed and it is suitable for our SVM based classifier [6]. Hence, the DSP less D-SIFT implementation is the best for our CAD system.

An evaluation platform is shown in Fig. 12. The platform receives the input image with capture board on a PC via a HD-SDI cable. Then the D-SIFT feature extraction occurs on the FPGA board, then feature quantities are sent to feature transformation and identification modules on the PC for cluster searching, voting to create and identifying using SVM. Finally, the result of SVM module is displayed as the supporting image.

The result of latency comparison with software (original) implemantation and the proposed hardware implementation is shown in Fig. 13. The table in Fig. 13 shows the relationship between number of pixels and Scan Window (SW) which is the region for feature extraction from the input image. The plots on the graph shows processing latency by

software inplementation and the proposed hardware. The performance is estimated for Full-HD (2 M pixels) image, the proposed hardware D-SIFT is about 700 times faster than that of software implementation. From this result, our proposed hardware is applicable for real-time processing for 4K and 8K high resolution image, which takes 80 msec and 160 msec, respectively. The proposed hardware can be used as a D-SIFT feature extraction accelarator for general images beyond the endoscopic images. Also, we estimate the performance of the whole system with the estimation results of other modules, feature transformation [7], and type identifier module [8]. From the implementation results, the throghtput is *16.7 fps* and latency is *60 msec*. So real-time processing is achievable for the on-the-fly diagnostic support system for clinical doctor (demand throghtput: >5 fps, latency: <1 sec).

6. Conclusion

This paper introduces our hardware accelerator implementation for the fundamental D-SIFT feature extraction in real time. The feature extraction time is linearly changed with the image size, in which D-SIFT features of 8K image can be extracted in as short as 160 msec and that for 4K image takes relatively short as 80 msec. Hence, this fundamental real-time DSIFT feature extraction accelerator implementation can be used in any application regardless the image size and scan window size. In addition, the relatively small hardware size occupation (0.5%) of the proposed implementation leaves many spaces for other application implementation. In particular, applying the proposed accelerator to CAD system with Full-HD image significantly increases the DSIFT feature extraction speed up 700 times compared with the software implementation. Processing time for a Full-HD image reduces from 14 sec in software to 20 msec @ 100MHz frequency.

Our future work includes the development of the whole CAD system including our D-SIFT to VW feature transformation architecture and our SVM architecture in one FPGA board.

Acknowledgment

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 and the Mentor Graphics Higher Education Program.

References

[1] H. Kanao, et al., "Narrow–band imaging magni fication 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] Chin-Chung Chang, Chin-Jen Lin, "Livsvm – a library for support vector machins," http://www. csie.ntu.edu.tw/~cjlin/ libsvm/

[5] T. Mishima, et al., "FPGA implementation of feature extra-ction for colorectal endoscopic images with NBI magnifi-cation," Proc. of the IEEE International Symposium on Circuits and Systems (ISCAS2014), pp. 2515-2518, June 1-5 2014.

[6] T. Okamoto, et al., "A hierarchical type segmentation algorithm based on support vector machine for colorectal endoscopic images with NBI magnification," Proc. of the 19th Workshop on SASIMI2015, pp. 374-379, Mar 16-17, 2015.

[7] T. Koide, K. Sugi, et al., "A hardware accelarator 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, July 2016 (to appear).

[8] T. Okamoto, et al., "Image segmentation of pyramid style identifier based on support vector machine for colorectal endoscopic images", Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC2015), pp.2997 – 3000 , Milano Conference Center, Milano, Italy, August 25-29 2015.

FPGA Board	PROCe IV 530-A (GiDEL)
Memory	On-borad DDRII 512MB DDRIISODIMM 1GB x 2
Installed FPGA	Altera Stratix IV EP4SE530H35C2
# of FPGAs	1
Host Interface	PCI-Express Gen 3.0

Resources	Available	SGF	GC	BFC	Total
# of ALUTs	424,960	919	603	473	1995 (0.47 %)
# of Registers	424,960	683	374	350	1407 (0.33 %)
Total RAM [bit]	21,233,664	65,536	32,768	26,624	124928 (0.59 %)
# of DSP blocks	1,024	0	0	0	0 (0 %)

Fig. 11: FPGA implementation results.

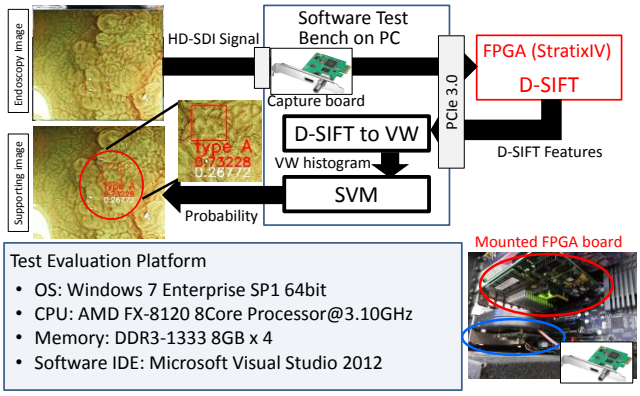
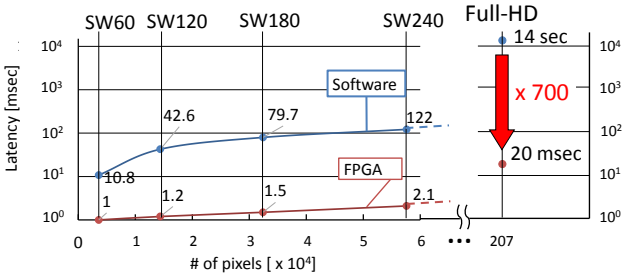


Fig. 12: The evaluation environment.



SW size	SW60	SW120	SW180	SW240
# of pixels	3,600	14,400	32,400	57,600

Fig. 13: Latency comparison with software and proposed implementation.