IMPROVING THE PERFORMANCE OF SIFT USING BILATERAL FILTER AND ITS APPLICATION TO GENERIC OBJECT RECOGNITION

Tomoaki YAMAZAKI, Tetsuya FUJIKAWA and Jiro KATTO

Department of Computer Science, Waseda University
E-Mail: {yama-t, fujikawa, katto}@katto.comm.waseda.ac.jp

ABSTRACT

Feature extraction of images can be applied to image matching, image searching, object recognition, image tracking etc. One of the effective methods to extract features of images is Scale-Invariant Feature Transform (SIFT) [1]. In this paper, we indicate problems of SIFT and propose a method to improve its performance by applying Bilateral Filter [2]. In addition, we implement its acceleration by GPGPU (general purpose GPU), apply this method to generic object recognition and perform a comparison experiment. We compare the proposed method with the original method using SIFT and confirm improvement of the identification rate by the proposed method.

***Index Terms*—** Feature extraction, Image edge detection, Object recognition

1. INTRODUCTION

In late years, with the evolution of computers, we can perform image processing to handle massive data. Especially, feature extraction techniques from an image are evolved and applied to various challenging tasks. One of the effective methods for this purpose is Scale-Invariant Feature Transform (SIFT) [1], which has been utilized in generic object recognition known as Bag-of-Keypoints (BoK) [3] and in structure from motion (SfM) for three-dimensional shape recovery [4]. On the other hand, the Bilateral filter was proposed for removing noise from images while preserving edges [2]. This filter can be considered as a locally adaptive Gaussian filter and can work in a noniterative manner dissimilar to anisotropic diffusion [5]. Theories also prove its effectiveness in a MAP (Maximum a posteriori) estimation framework [6,7]. In this paper, we present a simple but effective approach which replaces Gaussian filtering by the Bilateral filter in the scale space extrema detection of SIFT and examine its effectiveness from an aspect of the precision. We also try its fast implementation by using GPGPU [8], apply our approach to generic object recognition and investigate its performances extensively.

2. SIFT

SIFT [1] has robustness in rotation, scale and illumination changes of the images. It is mainly composed of keypoints detection and features description. Keypoints are detected by calculating extreme values of the difference images of smoothed images generated by Gaussian filters with different variances. A series of smoothed images is called scale space or image pyramid, and the difference image is called DoG (Difference of Gaussian). In the feature description, a feature vector is described by 128 dimensions, which are composed of sixteen orientation histograms, each of which has eight directions and is calculated by collecting brightness gradients in the adjacent area of the keypoints.
However, SIFT has a problem that keypoints on the edge or in low contrast regions tend to be deleted after detection, because the keypoint selection mechanisms of SIFT prefer corners and high contrast feature points in principle. Actually, the keypoints tend to fail matching on the edge in many cases.

3. PROPOSAL

In this paper, we focus on the smoothing process for DoG image generation. SIFT (and scale space) uses Gaussian filters for smoothing an original image. We propose a method to apply Bilateral Filter [2] in place of Gaussian Filter.
Bilateral Filter is known as an edge-preserving smoothing filter. It controls a filter weight, not only by pixel locations but by brightness gradients. For the points around which brightness changes rapidly, small weights are assigned to preserve an edge. On the other hand, for the points inside quiet regions, normal Gaussian weights are assigned to smooth the image.
Let $f(i,j)$ be an input image, and $g(i,j)$ be an output image. Bilateral filtering is given by

$$g(i,j)=\frac{\sum_{n=-w}^{w}\sum_{m=-w}^{w} f(i+m,j+n)\exp\left(-\frac{m^2+n^2}{2\sigma_1^2}\right)\exp\left(-\frac{\left(f(i,j)-f(i+m,j+n)\right)^2}{2\sigma_2^2}\right)}{\sum_{n=-w}^{w}\sum_{m=-w}^{w} \exp\left(-\frac{m^2+n^2}{2\sigma_1^2}\right)\exp\left(-\frac{\left(f(i,j)-f(i+m,j+n)\right)^2}{2\sigma_2^2}\right)}$$

where σ_1 is a parameter to control Gaussian shape in space, and σ_2 is a parameter to control effect of brightness change.

When σ_1 increases, filter's blur effect increases. When σ_2 decreases, edges tend to be preserved more emphatically. Since Bilateral Filter tries to keep more edges, it is expected that more keypoints on the edge can be distinguished and detected. Later in this paper, we call the image pyramid produced by Bilateral filtering DoB (Difference of Bilateral) images, which is analogous to DoG.

4. EXPERIMENTS

4.1. Comparison of DoG and DoB

We show examples of the DoG images and DoB images (256x256) in Figure 1, where the output values are emphasized. We can observe edges are preserved in the DoB images as expected. Magnitudes of output values of DoB are lower than those of DoG in general. However, in the case of DoB, we can observe several salient extreme values which are not observed (probably filtered out) in DoG.

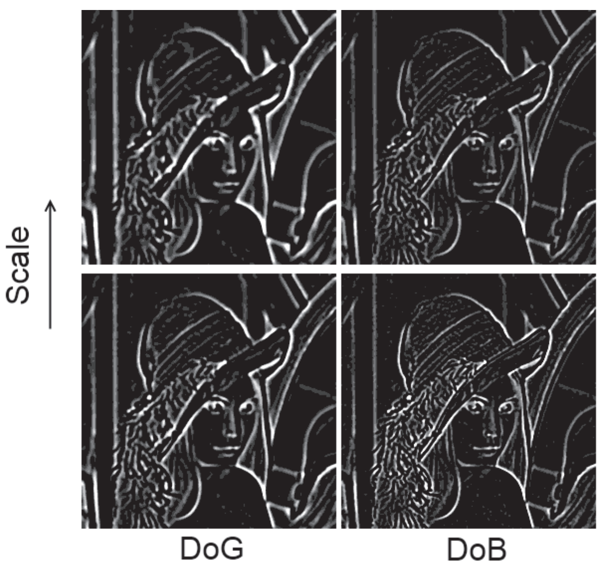


Figure 1: Example of the differential images

4.2. Comparison of keypoints

We compare the keypoints obtained by the original SIFT and the proposed method in Figure 2. Left images show all the detected keypoints, middle images show keypoints after removing keypoints on the edge, and right images show keypoints after removing low contrast ones. Upper images are results of the original SIFT, and lower images are those of the proposed method. We can observe that many keypoints remains in our proposal using the Bilateral filter.

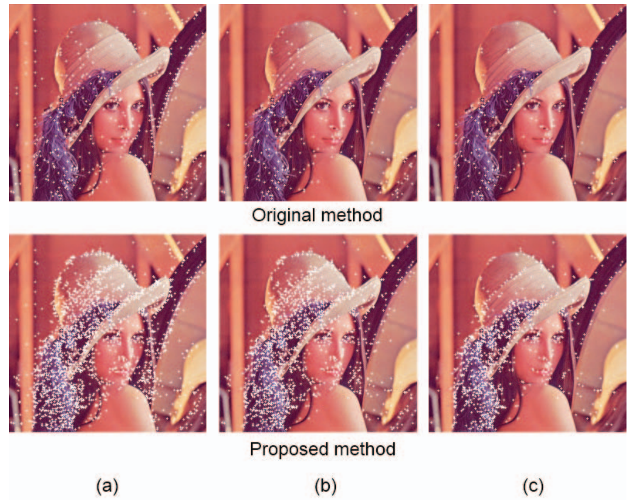
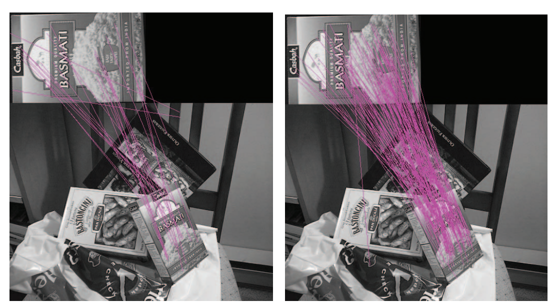


Figure 2: Example of keypoints selection: upper: original SIFT, lower: proposal, left: initial keypoints, middle: removing keypoints on the edges, and right: removing low contrast keypoints.

4.3. Comparison of keypoint matching

We show an experimental result that performs keypoint matching in Figure 3. It can be easily recognized that the matching errors occur in the original SIFT but they do not appear in the proposed method. Furthermore, corresponding keypoints are clearly increased because edge-preserving property of the Bilateral filter contributes to increase of information of the differences images.



Original：(30, 1112ms)　Proposed：(92, 4635ms)

Figure 3: Experimental result I (number of matched keypoints, computation time)

Then, we show another experiment result that performs feature matching between two images in Table 1. One image is original, and another image is clipped from the original and expanded to double. We prepare 10 sets of such images. Values in the Table 1 are the mean-values for 10 sets of images. The errors decrease to about a one-third whereas the number of correct matches increases to about 3 times.

Table 1: Experimental result II

	Correct (number)	False (number)	Time (ms)
Original	3.4	1.0	438.5
Proposed	10.7	0.3	4371.7

5. FAST IMPLEMENTATION WITH GPGPU

As shown in the previous subsection, computational complexity of the Bilateral filter increases in comparison with Gaussian Filter, and results in total processing time increase. This motivates us to implement the proposed method by CUDA which enables us to accelerate by GPGPU through C/C++ APIs [8]. We show the graph of the computing speed with and without the speedup in Figures 4 and 5. In the case of 512*512 images, we succeeded in speeding up the filter processing part to about 14 times and total processing to about 6 times. Without the speedup, the computing time of the Bilateral Filter accounts for most of total processing time but, with the speedup, it decreases to around 40%. As a result, processing time of the proposal becomes comparable to the original one.

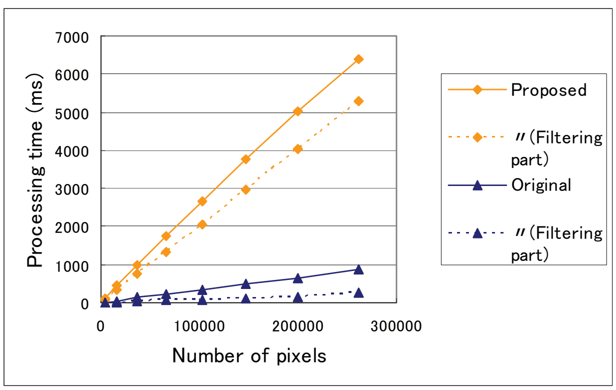


Figure 4: Computing speed without GPGPU acceleration

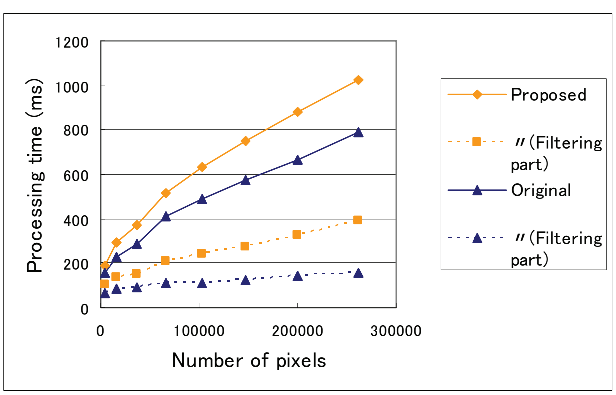


Figure 5: Computing speed with GPGPU acceleration

6. APPLICATION TO GENERIC OBJECT RECOGNITION

In this section, we apply the proposed method to generic object recognition and investigate its performance.

6.1. Experiment outline

We compare the identification rate of generic object recognition known as Bag of Keypoints (or visual words) [3], in which original SIFT is used (noted by Gaussian) or our extension is used (noted by Bilateral). We perform experiments about next three items.

A) Comparison of identification rates when the number of visual words is changed.
B) Comparison of identification rates per category.
C) Comparison of identification rates when two thresholds are changed.

We use 20 categories of Caltech-256 image dataset [9]. As learning images and test images, we use 40 pieces for each. We use k-means clustering for vector quantization and SVM (support vector machine) for classification. We change the number of visual words between 100~1000 in experiment (A) and plot the maximum identification rates in experiment (B). The experiment (C) has two controllable thresholds; one for keypoint exclusion on the edges and another for the low contrast. We compare identification rates when we changed each threshold. In the case of (A) and (B), we fix these thresholds to 10.0 and 0.4. Both are the typical values in the original method.

6.2. Experimental results

Figures 6 and 7 show results of experiments (A) and (B), respectively. In the experiment (A), identification rates are always higher in the proposed method than the original method as expected when the number of visual words are changed. However, in the experiment (B), there are some categories in which identification rates become lower than the original method. The images belonging to these categories tend to have simple shapes. For these categories, we may not need the features of complex parts and our approach seems not to contribute to performance improvement. Figure 8 shows results of experiment (C). In this experiment, we plot the identification rates when two thresholds for keypoint removal are changed. It is observed that the variation of the identification rates becomes large in the proposed method in comparison with the original method. Figure 9 shows the total number of keypoints counted for the two thresholds corresponding to Figure 8. From these figures, it can be concluded that the threshold setting has more influences on the identification rate than the number of keypoints.

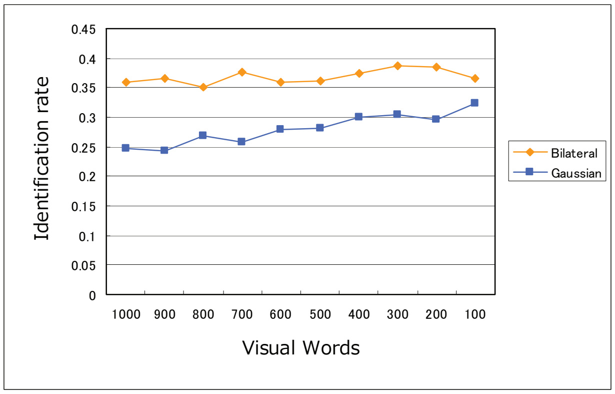


Figure 6: Result of experiment (A)

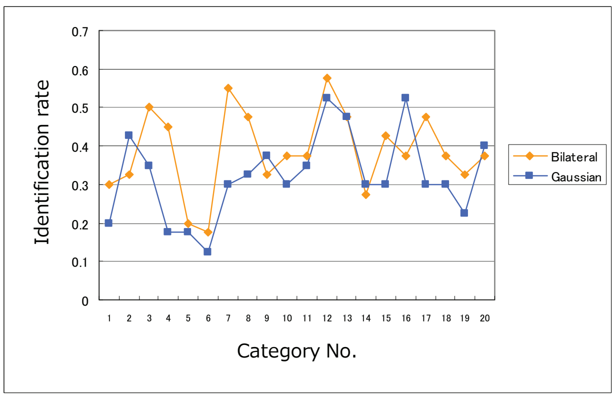


Figure 7: Result of experiment (B)

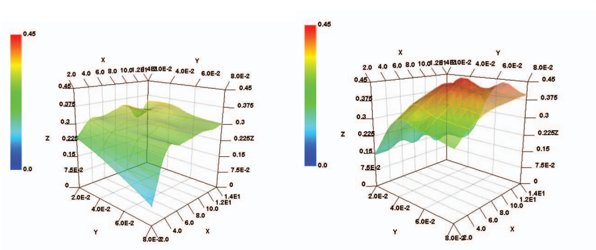


Figure 8: Result of experiment (C) (Left: Original method, Right: Proposed method)

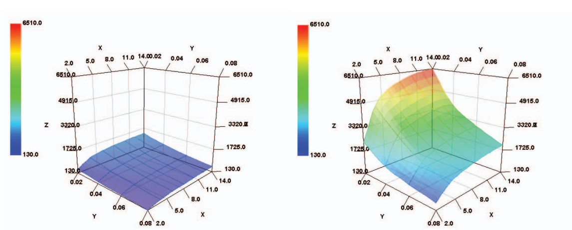


Figure 9: Total number of keypoints in each threshold

7. CONCLUSIONS

In this paper, we propose a method to apply the Bilateral filter instead of Gaussian filter in SIFT. In the result, matching-errors on the edges can be decreased and, because the Bilateral filter controls weighting according to spatial distance as well as brightness and amount of information of the difference-images is increased, many features came to be extracted against the original one. Furthermore, we implemented GPGPU acceleration and compared identification rates of generic object recognition as an application of the proposal. We confirmed that identification rates can be improved generally. However, for some categories, it is not always advantageous to have many keypoints because identification rates are sensitive to appropriate threshold setting. For future prospects, we try to develop an adaptive method which classifies images according to their properties (details, edges and so on) and allocates adequate parameters for each class.

8. REFERENCES

[1] D. Lowe, "Distinctive image features from scaleinvariant keypoints," *Proc. of International Journal of Computer Vision (IJCV)*, 60(2), pp. 91-110, 2004.

[2] C. Tomasi and R. Manduchi, "Bilateral Filtering for Gray and Color Images", *Proc. of ICCV*, pp.839-846, 1998.

[3] G. Csurka et al., "Visual Categorization with Bags of Keypoints," *ECCV Workshop on Statistical Learning in Computer Vision*, pp. 1-22, 2004.

[4] N. Snavely et al., "Photo tourism: Exploring photo collections in 3D," *ACM Trans. on Graphics*, 25(3), pp.835-846, 2006.

[5] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. on PAMI*, 12(7), pp.629-639, 1990.

[6] M. Elad, "On the origin of the bilateral filter and ways to improve it," *IEEE Trans. on IP*, 11(10), pp.1141-1151, 2002.

[7] M. Tanaka and M. Okutomi, "Latent common origin of bilateral filter and non-local means filter", *Image Processing: Algorithms and Systems*, 2010.

[8] "CUDA Technical Training", http://developer.nvidia.com/object/cuda_training.htm

[9] G. Griffin et al., "Caltech 256 object category dataset," Technical Report UCB/CSD-04-1366, California Institute of Technology, 2007.