

MIC Acceleration of Saliency Detection Algorithm

Changfei Zhou, Kai Lu, Wanqing Chi, Xiaoping Wang, Zhenhai Xiong
National University of Defense Technology
Changsha, China
zcf337889342@163.com

Abstract—Image processing has important influence during the development of visualization. With the scale and the complexity of image data expanding, it is becoming more and more difficult for image processing and target analysis. Saliency area detection of image processing, especially, visual attention is playing an important role the decreasing of computing burden and the improvement of the accuracy. Among this kind of algorithms, saliency detection based of structural similarity theory has more excellent performance than others through a new center-surround operator. In this paper, we realize the parallelization on data level and thread level of this algorithm on the MIC platform. The experiment results show that we could obtain a speed-up ratio of 5.6 in average with the help of the SIMD. We could also get a top performance speed-up ratio of 2.42 in average on an MIC coprocessor against a prevalent CPU.

Keywords: Visualization, saliency detection, structural similarity, MIC

I. INTRODUCTION

Saliency detection model based on visual attention is a multi-subject crossing and basic theory field, which involves visualization, computer vision, image processing, pattern recognition, artificial intelligence, psychology and neurobiology. Based on the different way of processing, saliency detection algorithms are divided into two categories, bottom-up visual saliency model and top-down visual saliency model. The former is driven by data, independent of any especial task. The latter is driven by recognizant experience, dependent on running tasks. The algorithm optimized by this paper provides a method to detect visual saliency by means of extracting ground feature, independent of upper-leveled information. So it belongs to the former one.

The Itti algorithm is proposed by Itti and Koch, based on Feature Integration Theory. Through detecting the image's multiple visual characteristics for their saliency respectively, it can merge the results and provide the final saliency figure. Inspired by the structural similarity theory and the numerical measurement of that value, the novel theory can also be utilized for the detection of visual saliency. Based on this theory, when computing a pixel's saliency, a 3*3 matrix with surrounding that pixel is introduced as the basic computing unit. However, this kind of algorithms has some common drawbacks as long executing time and calculated quantity.

Traditional single-core and multi-core processors couldn't satisfy our demand. Many Integrated Core computing architecture is based on Many Integrated Corecomputing model. Like cloud computing, Many Integrated Core computing has the similar framework and has been well developed since firstly proposed, which also has inspiration on cloud computing. In the future, we believe

that Many Integrated Corecomputing and cloud computing will motivate each other's further development.

Many Integrated Core(MIC) architecture has smaller kernel and hardware thread than universal multi-core processors so that it has higher computing resource density and lower communication cost on chip. It also has more transistors and computing power for more complex parallel applications. We can see that MIC architecture is specially designed for high performance computing and complex parallel applications.

In this paper, our algorithm to be improved is focused on the process of single and read-only pixel so that we could get it accelerated by means of MIC architecture. We take the global saliency detection algorithm as an example. We could also get a top performance speed-up ratio of 2.42 in average on an MIC coprocessor against a prevalent CPU.

II. BACKGROUND

A. Saliency detection

Traditional spatial model is not satisfying for saliency detection whose detecting results have mosaic and poor edge character. To resolve these problems, we can make use of the saliency area detecting algorithm based on structural similarity in which the contrast ratio is much higher.

Structural Similarity index (SSIM) is image quality evaluation theory proposed by Z.Wang based on human visual system. As human visual system (HVS) have many characteristics such as multi-channels, non-linear, covering effect, contrast sensitivity, mutual multi-channel stimulating effect, and so on. No accurate computing model has been proposed for some of these characteristics. Nowadays, most of the research on HVS is bottom-up. SSIM simulates the HVS from a relatively high level, utilizing the structural information to observe an object. It is a brand new image quality evaluation system. As a natural image's pixels are highly relevant to each other, especially for one pixel and its adjacent pixels. This relation includes structural similarity. And SSIM supposes that two images distortion is not serious if their structural similarity has not changed severely. In general, SSIM is a top-down image quality evaluation system, which treats the HVS as a black box, only simulates outer function rather than its inner components. So the SSIM system is much simpler, more effective and easier to realize, overcoming many disadvantages than bottom-up models.

In this paper, as figure1 shows, we use the saliency detection algorithm based on global structural similarity (SDGSS) as an example. The 3*3 matrix covering the center area is being processed. Figure 1 also shows our global saliency measurement scheme, the matrix is also used for computing the structural similarity, labeled black. The

top left corner is also black labeled because those pixels are multi-read by many areas.

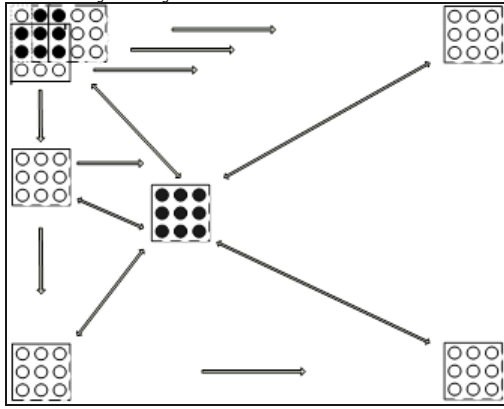


Figure 1. The Saliency detection algorithm based on global structural similarity

For the purpose of measuring one pixel's saliency of an image, we firstly build a 3*3 matrix surrounding this pixel as one of its characters, and compare the matrix and all the other same -scaled matrices. The structural similarity comparison result is based on human visual specialty and psychological feelings. After traversing all the same-scaled matrices' characters, the pixel's similarity value is acquired by weighted summing all these characters. The whole procedure includes all the 3*3 feature matrices' average value, variance and the covariance between the center matrix and surrounding matrices. Covariance between two different matrices is also an important feature to be computed. Equation shows how to compute the specific:

$$SSIM = \frac{(2u_x u_y + c_1)(2\sigma_{xy} + c_2)}{(u_x^2 + c_1)(\sigma_x^2 + c_2)} \quad (1)$$

Based on human visual character and psychological feelings, we would introduce a factor, L to compare the structural similarity between two matrices. The value of L, C1 and C2 could be computed by the equations in [1], we could also obtain a pixel's saliency value so that the image's saliency figure is created by traversing all the pixels of the image:

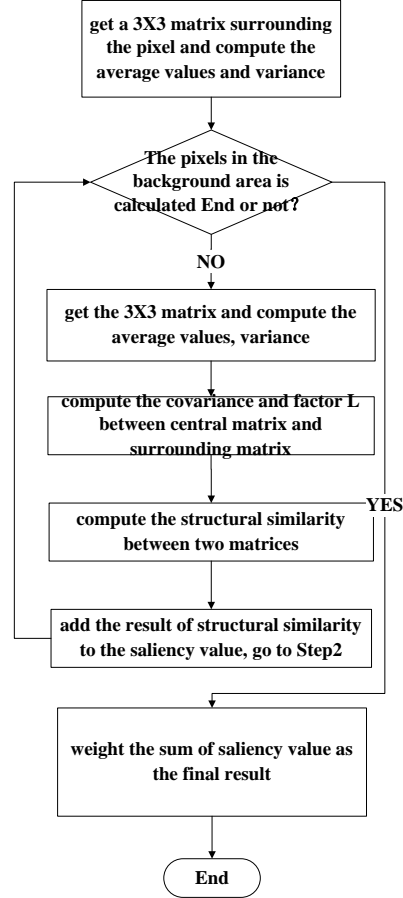


Figure 2. Algorithm Process Chart

The saliency map is obtained through traversing all pixels of the image. Regard to the characteristics of the algorithm. At run time, the more threads, the greater the speedup, although the growth rate will slow down as the increased number of threads, it is in a growth state. Therefore, in the experiment, taking the all of the thread operations.

B. MIC

Xeon Phi is the second generation MIC with the performance of above 1TFlops based on DGEMM tests. Figure 2 shows its architecture. Supporting the new SIMD (Knights Corner Instructions), vector calculations, Xeon Phi has a VPU and a vector register (Zmm register) on each processor as hardware. The bit wide of VPU and Zmm register is 64 bytes and 512 bits respectively. It also supports 16 single precision or 8 double precision flop vector calculations. As a SIMD instruction issued by VPU, 8 flop calculations have been completed. Currently, Intel has provided some methods of vector optimization such as the vectorization library, Intel intrinsic functions.

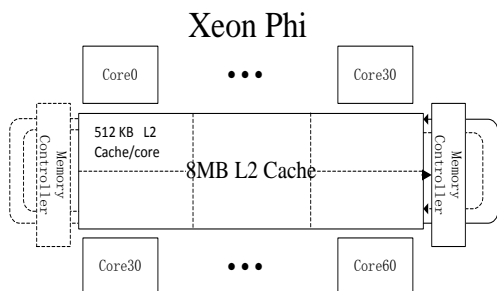


Figure 3. Architecture overview of an Xeon Phi coprocessor

Xeon Phi has 57 processors integrated and it could execute 4 threads at each processor simultaneously with the help of hyper-threading. Each processor has 32KB L1 level data and instruction cache and 512KB L2 level cache.

MIC framework provides three kinds of running models, Offload model, Co-processor-only model and Symmetric model.

Offload model: MIC is only part of accelerators which only participates computing rather than control. Algorithm and data are firstly organized at CPU, then transmitted to MIC through PCIe interface, finally the results is returned by MIC back to CPU.

Co-processor-only model: MIC acts like an independent processor, completing all the computing tasks without being controlled by CPU.

Symmetric model: MIC is equal to CPU and MIC is participating computing and control at the same time as CPU.

III. IMPLEMENTATION AND DESIGN

In this paper, we use Co-processor-only Model to achieve the visual salient region detection method based on global structural similarity algorithm on MIC platform. To fully use the MIC, first we employ 512-bits SIMD operation, then extend to multi-threading with OpenMP realizing data level parallel and thread level parallel.

A. Data level parallel

In order to exploit the data parallelism on each CPU core, we use 512-bits SIMD to vectorize the feature extraction and structural similarity metric calculation, including direction, color, and brightness computing.

data pack: MIC provides 512bit registers and vector operations, in order to use vector operation, we should pack the data into the Zmm register. Take color calculating as an example, if we need to detect i pixel's color similarity, the color information of 8 pixels around i should be packed into a vector register. Since the register length is 16 (float), we need to patch the vector by filling 0 into the remaining bits, and as a beginning the top left corner of the nine pixel color information should be packed into the vector register.

intrinsic function calculation: Call Intel intrinsic functions for color comparison, then generate the color similarity.

In order to reduce the pressure of vector packaging operation to the memory bandwidth, we use the method of partial register reuse. During the $i+1$ pixel detection, shift out the unneeded data in this iteration by a right-shift operation. Then read the needed data of $i+1$ iteration and insert it into the Zmm register. Details are shown in Figure 4:

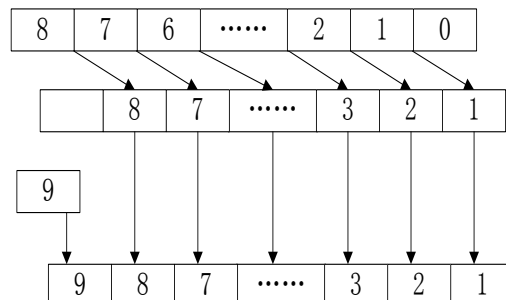


Figure 4. vector operations

B. Thread level parallel

After the vectorization of single-threaded algorithm, our work focused on the multithread parallelism for further optimized. MIC support OpenMP and MPI parallel programming model. Significant detection process is computationally intensive, and MIC is a shared memory architecture, we realize the visual salient region detection method based on global structural similarity algorithm using OpenMP on the MIC.

Using threads collaborative computing methods to compute tasks to each thread average. Firstly, when the program is created for visual feature extraction threads, put the computing tasks of color, direction and brightness to a certain number of threads to calculate at the same time. Because each part of the calculation is different, specific tasks dividing according to the test results is dynamical divided. After the feature extraction is completed, we set a barrier synchronization to ensure the completion of three-part feature extraction. As to structural similarity metric, the process is similar to feature extraction.

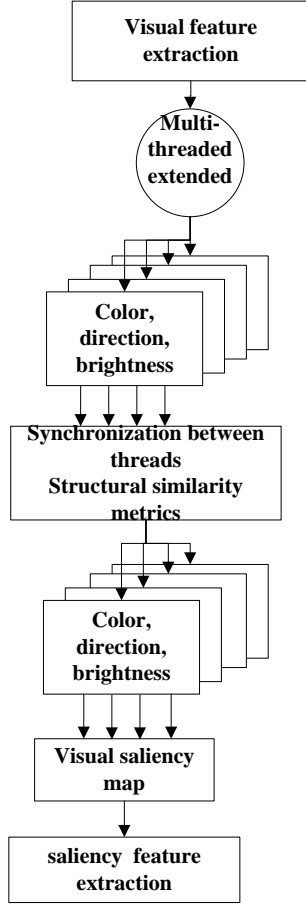


Figure 5. Visual salient region detection Parallel Algorithms based on structural similarity

IV. EVALUATION

In this section, we test the performance of visual saliency region detection method based on global structural similarity algorithm on Xeon Phi (KNC) coprocessor with 57 cores. As for comparison, we also test the time cost of algorithm implemented with one E5-2670(2.6GHz, 8-core). The configurations of Xeon Phi and E5-2670 are listed in Table 1. We use Intel C++ Compiler 13.0 to compile the code and OpenMP 3.1 to parallelize the code.

TABLE I. TESTING PLATFORMS

Configuration	E5-2670 CPU	Xeon Phi
cores	8	57
Total threads	16	228
Frequency	2.6GHZ	1.1GHZ
Memory	128GB	5GB

Firstly we test the performance enhancement obtained by vectorization. Table2 shows the results before and after the vectorization. The size of the image is 128*128.We

extend the thread number from 1 to 16. We can see the vectorization for the code achieves average speedup ratio of 5.6 and the multi-thread parallelization obtains an effect close to linear increment. The test results are shown in Table 3.

TABLE II. EXECUTION TIME WITH AND WITHOUT SIMD VECTORIZATION (MS)

Num_threads	MIC (without SIMD)	MIC (SIMD)	Speedup
1	65200	11800	5.3
2	34350	6133	5.6
4	18620	3350	5.5
8	10420	1903	5.5
16	5895	981	6.1

TABLE III. MULTI-THREADED PERFORMANCE COMPARISON

Num_threads	MIC (SIMD)	Speedup
1	11800	1
2	6133	1.92
4	3350	3.41
8	1903	6.3
16	981	12.6

We test the time cost of visual salient region detection method based on global structural similarity algorithm running on one Xeon Phi coprocessor and one CPU. On the CPU, we adapt 16 threads. On Xeon Phi coprocessor, we use 210 threads the code vectorized by 512-bit SIMD. From Table 4, we can see that the performance of our algorithm running on one Xeon Phi coprocessor achieves an average speedup ratio of 2.42 times compared to one E5-2670 CPU.

TABLE IV. COMPARISON BETWEEN CPU AND XEON PHI COPROCESSOR (MS)

Im_size	CPU(8 threads)	MIC (210 threads)	Speedup
256*256	8170	3350	2.3
256*512	17550	7970	2.2
512*512	296330	123530	2.4
512*768	454810	168510	2.7
768*768	1922920	770350	2.5

V. CONCLUSION

We realize the transplantation of the algorithm of saliency area detection based on the global structural similarity onto the Xeon Phi platform. Then we test the performance of this new version. Firstly, we explore the data parallelism of the original algorithm on the new platform by way of manual vectorization and acquire the speed up by a factor of 5.6. Then, we introduce the OpenMP multi-thread method to get it effectively realized on multi-Xeon Phi processors. Through the data-level and thread-level parallelism, this algorithm has acquired the acceleration by a factor of 2.42 on multi-Xeon Phi platform against Xeon E5-260 CPU. We believe this method is effective in the field of image processing and other algorithms involving saliency detection.

However, there are also some obstacles to overcome. Firstly, MIC couldn't provide satisfying degree of

parallelism as the background area chosen for saliency computing is sometimes too large, we will move forward a single step of optimizing the algorithm to get it resolved. Then, we will attempt the image segmentation to acquire more acceleration. Finally, we will also have more other relevant algorithms transplanted onto this new platform and have more confirmatory experiments.

ACKNOWLEDGMENT

This work is partially supported by NCET, and National Science Foundation (NSF) China under grant numbered 61272142, 61170261, 61103082, 61103193 and 61003705, and a grant from the National High-tech R&D Program of China (863 Program) numbered 2012AA01A301 and 2012AA010901.

REFERENCES

- [1] Itti L, Koch C and Niebur E. "A model of saliency-based visual attention for rapid scene analysis". In IEEE Trans. On Pattern Analysis and Machine Intelligence, 1998, 20(11):1254-1259
- [2] Zhou Wang, A C Bovik. "A universal image quality index", IEEE Signal Processing Letters, 2002, 9(3):81-84
- [3] Zhou Wang, A C Bovik, H R Sheikh, E P Simoncelli. "Image quality assessment: from error visibility to structural similarity", IEEE Transactions on Image processing, 2004, 13(4):600-612
- [4] Wikipedia --Structural_similarity. [Online].
- [5] Kai lu, Zhenhai Xiong, Chongfei Li, Gen Li "Significant detection based on global structure similarity metrics", Computer Engineering and Science
- [6] Chongfei Li, ZhiguoQu, Kai Lu, YinghuiGao. "Visual salient region detection method based on structural similarity", Computer engineering & science.
- [7] [10] T. Elgar, "Intel Many Integrated Core (MIC) Architecture," in 2nd UK GPU Computing Conference, December 2010. [Online]. Available: <http://www.many-core.group.cam.ac.uk/ukgpucc2/programme.shtml>,
- [8] Intel® Xeon Phi™ Coprocessor System Software Developers Guide. SKU# 328207-001EN November, 2012.
- [9] [14] QIAN Xinglong, ZANG Bimyu, ZHU Chnamqi, "Partial Reuse of the Vector Registers in SIMD Optimization", COMPUTER ENGINEERING&SCIENCE, Vol.29,No.5,2007,