Stereo Macthing

Yongshun Zhang

April 18, 2018

1 Preparation

1.1 Overview of Stereo Matching

According to Project_Stereo guide, I first read a brief guide of stereo matching [1].

I learn that there are local and global methods to calculate the disparities of each pixel depending on images from different perspectives and some constraints. Local methods, such as block matching use constraints on a small number of pixels surrounding the pixel of interest, and global methods, such as Dynamic programming and graph cuts, use constraints on scan-lines or the whole image [2]. Both of the method have their advantages and disadvantages, local methods calculate faster than global methods but accuracy is worse, and global methods manage to solve an optimization problem called cost aggregation, based on scan-lines or the whole image, which means greater computational complexity.

The semi-global matching algorithm we introduce next, which seems more like global methods, uses the pixels from whole image except the pixels with occlusions.

1.2 Semi-global Matching

It takes me lots of time to understand the main idea of semi-global matching, simply as SGM. Semi-global matching algorithm can be divided into three sessions: Cost Calculation, Cost Aggregation, Disparity Computation and Disparity Refinement [3], and in my comprehension, the two most important sessions are the cost calculation and cost aggregation, so below will introduce these sessions mainly.

Cost Calculation

The main method of calculating matching cost is based on image pixel p from its intensity I_{bp} and the suspected correspondence I_{mq} with q=e_{bm}(p,d) of the match image. SGM uses mutual information(MI) as the matching cost, which is insensitive to recording and illumination. The MI is defined as

$$MI_{I_1,I_2} = H_{I_1} + H_{I_2} - H_{I_1,I_2} \tag{1}$$

and the entropies, which represents degree of confusion of the intensity of images, are calculated from probability distribution

$$\begin{aligned} H_I &= -\int_0^1 P_I(i)\mathrm{d}i \\ H_{I_1,I_2} &= -\int_0^1 \int_0^1 P_I(i)\mathrm{d}i_1\mathrm{d}i_2 \end{aligned} \tag{2}$$

By using Taylor expansion [4], the calculation of joint entropy H_{I_1,I_2} can be transformed into a sum over pixels.
$\mathrm{H}_{I_1,I_2} = \sum_p h_{I_1,I_2}(I_{1p}, I_{2p})$
$Foreachpixel'sintensityinimages\mathrm{I}_1$ and I_2, data term h_{I_1,I_2} is calculated from the joint probability distribution

P_{I_1,I_2}. The number of corresponding pixels is n. And we use Parzen estimation with kernel function: 2D Gaussian [4].

$$\begin{aligned} h_{I_1,I_2}(i,k) &= -\frac{1}{n} log(P_{I_1,I_2} \otimes g(i,k)) \otimes g(i,k) \\ P_{I_1,I_2}(i,k) &= \frac{1}{n}\sum_p T[(i,k) = (I_{1p}, I_{2p})] \end{aligned} \tag{3}$$

In paper "Stereo Processing by Semi-global Matching and Mutual Information" written by Hirschmuller, it is said that in equation(3), gaussian smoothing is applied by convolution, but it has been found that using a small kernel(7 × 7) gives practically the same results as larger kernels, but is faster than before [8]. Concerned the occlusions, some of the intensities of I_1 and I_2 don't have a correspondence, therefore, it's suggested to calculate these entropies analog to the joint entropy.

$$H_I = \sum_p h_i(I_p) h_i = -log(P_I(i) \otimes g(i)) \otimes g(i) P_I(i) = \sum_k P_{I_1,I_2}(i,k) \tag{4}$$

Therefore, the resulting definition of Mutual Information is,

$$\begin{aligned} MI_{I_1,I_2} &= \sum_p mi_{I_1,I_2}(I_{1p}, I_{2p}) \\ mi_{I_1,I_2}(i,k) &= h_{I_1}(i) + h_{I_2}(k) - h_{I_1,I_2}(i,k) \end{aligned} \tag{5}$$

The Mutual Information leads to the matching cost,

$$C_{MI}(p,d) = -mi_{I_b, f_D(I_m)}(I_{bp}, I_{mq}) \tag{6}$$

Here the function f_D, as far as I'm concerned, can be replaced by stereo calibration and stereo rectification, but in the paper, it is said that f$_D iswarppingthematchimageaccordingtodisparityimage D 1001[3]. I don't understand how$f$_D$ is realized in the paper now, but in the coding project, I will figure out this problem.

Another important method of Cost Calculation is Hierarchically calculation, which apparently reduces the computation complexity of semi-global algorithm.

Cost Aggregation

In order to efficiently calculate the aggregating matching cost, aggregating matching cost is discribed in the sum of each pixel with 1D aggregating cost from directions arround the pixel, the number of directions can be 4, 8, 16 and so on, and 8 or 16 directions are often chosen, as the following picture shows

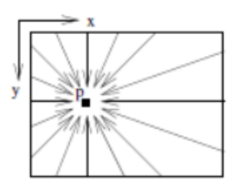

Figure 1: 16 Paths from all Drection r

Let L_r be a path that is traversed in the direction r. The cost $L_r(p,d)$ of the pixel p at disparity d is defined recursively as

$$L_r(p,d) = C(p,d) + min(L_r(p-r,d), L_r(p-r,d-1)+P_1, L_r(p-r,d_1)+P_1, \min_i L_r(p-r,i)+P_2) - \min_k L_r(p-r,k) \tag{7}$$

P_1 is constant penalty for all pixels q in the neighborhood N_p of p, for which the disparity changes a little bit. P_2 is a larger penalty than P_1, and the value of P_2 is changed based on the disparity changes and intensity changes, because we should take small penalty of discontinuities in order to protect the borderedges of the image and discontinuities often has larger disparity changes and larger intensity changes, so the value of P_2 is defined as $P_2 = \frac{P'_2}{|I_{bp} - I_{bq}|}$. The last data term is to keep the aggregating cost small. And the summed cost is

$$S(p, d) = \sum_r L_r(p, d) \tag{8}$$

The proccessing steps of semi-global matching algorithm can be discribed as Figure 2.

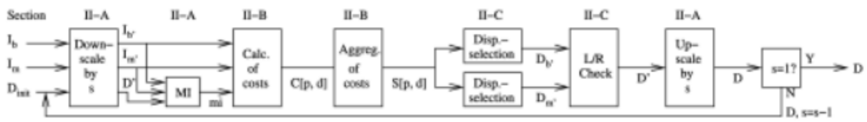

Figure 2: Steps of SGM

Problems

The paper of SGM is quite obscure and difficult to understand, and I read the paper "Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information" at least 6 times until now, but I am still lack of confidence to say that I understand and master all the methods in this paper. There are also some mathmatical knowledge, including parzen estimation, filters, gaussian distrubution and so, which are mentioned in the papers [3] [4] [8].

In "Visual correspondence using energy minimization and mutual information", the author introduce how to use parzen estimation and Taylor Expansion to make estimation of pixels' probability and calculate mutual information. And I also read some blogs [7] and lectures [6] about parzen estimation, which explain the theory of parzen estimation clearly. Besides, I also learn gaussian blur, image intensity, gaussian distribution, delta funtion and other new knowledge from wikipedia and other papers.

1.3 CUDA

I learn the basic grammar of CUDA programming mainly based on the book "GPU 高性能编程 CUDA 实现" [14] and Prof. Mohamed Zahran' lectures [15]. Besides, I also search for details about CUDA on Virtual Workshop [16].

With the learning of CUDA programming on GPU, I understand the importance of fundamental classes, such as Computer Composition Principle and Operating System. When we use GPU, its important to understand the principle of process and threads, in other words, how the program works, which is different from programming on cpu. When we use CUDA, we can realize parallel and make cooperation between threads to accelerate speed of program.

References

[1] **Stereo Matching: an Overview** Andrea Fusiello http://www.diegm.uniud.it/fusiello/teaching/mvg/stereo.pdf

[2] Myron Z.Brown, Darius Burschka, Gregory D.Hager, "Advances in Computational Stereo", Graphics and Scientific Visualization (CS 525), 2007.

[3] H. Hirschmuller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 2. IEEE, 2005, pp. 807-814.

[4] Kim, J., Kolmogorov, V., Zabih, R. ”Visual correspondence using energy minimization and mutual information,” In: Proceedings of IEEE International Conference on Computer Vision. vol. 2, pp. 1033–1040 (2003).

[5] **Gaussian Blur** Wikipedia https://en.wikipedia.org/wiki/Gaussian_blur

[6] **Lecture 10: Density estimation** Ricardo Gutierrez-Osuna, Wright State University, Introduction to Pattern Recognition. http://www.diegm.uniud.it/fusiello/teaching/mvg/stereo.pdf

[7] **Introduction to kernel density estimation (Parzen window method)** milania's blog https://milania.de/blog/Introduction_to_kernel_density_estimation_%28Parzen_window_method%29

[8] H. Hirschmuller, ”Stereo Processing by Semiglobal Matching and Mutual Information,” IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 30, NO. 2, FEBRUARY 2008

[9] **What is Image Intensity** IGI Global, https://www.igi-global.com/dictionary/image-intensity/13822

[10] **Mulitvariate Normal Distribution** Wikipedia https://en.wikipedia.org/wiki/Multivariate_normal_distribution

[11] **Delta Function** Wikipedia https://en.wikipedia.org/wiki/Delta_function_(disambiguation)

[12] **Median Filter** Wikipedia https://en.wikipedia.org/wiki/Median_filter

[13] Boykov, Veksler, Zabih, ”Fast Approximate Energy Minimization via Graph Cuts”, IEEE PAMI 23(11), pp 1222ff, 2001

[14] Jason Sanders, Edword Kandrot. GPU 高性能编程 CUDA 实现 [M]. 聂雪军等译. 机械工业出版社, 2011

[15] **Graphics Processing Units (GPUs): Architecture and Programming** Prof. Mohamed Zahran, NewYork University, https://cs.nyu.edu/courses/fall17/CSCI-GA.3033-004/

[16] **Introduction to GPU and CUDA Programming** Virtual Workshop https://cvw.cac.cornell.edu/gpu/structure

[17] 郑宝东, 王忠英线性代数与空间解析几何 [M]. 第 4 版. 高等教育出版社, 2013 .04

[18] David C. Lay. 线性代数及其应用 [M]. 刘深泉, 洪毅等译. 第 3 版. 机械工业出版社, 2005 .08