Assignment 2: Disparity Map for Stereo Matching

LI Xiaoyang

April 19, 2023

1 Introduction

In this project, semi-global block matching (SGBM) [2] is adopted to compute disparity map for stereo matching task. Moreover, weighted least square (WLS) [1] filtering is used for image filtering. OpenCV library is utilized in the implementation of SGBM and WLS. Methodology and experiment analysis will be covered in the following sections. In terms of evaluation metrics, peak signal-to-noise ratio (PSNR) is selected to evaluate the difference between the ground truth disparity maps and results. In my trial and error, the highest PSNR for three test cases, **Art**, **Dolls** and **Reindeer** is **17.00**, **17.27** and **15.51** respectively. In comparison, empirical PSNR value in image compression ranges from 20 to 40.

2 Methodology

2.1 Semi-global block matching

In general, OpenCV provides four algorithms for stereo matching, including block matching (BM, SteroBM), semi-global block matching (SGBM, StereoSGBM), graph cut (GC, cvStereoGCState) and dynamic programming (DP, cvFindStereoCorrespondence). To trade-off between accuracy and time efficiency, semi-global block matching is adopted in the implementation.

In this section, SGBM [2] will be mainly explained. Firstly, Sobel operator is used to compute gradient of both left and right images before matching cost calculation. After pre-processing, block matching cost is derived based on previous results. Next, block matching cost is directly calculated and aggregated with previous block matching cost.

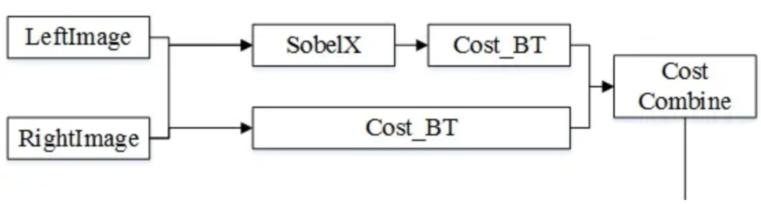


Figure 1: Flow chart for step I, II

Notably, pre-processing module reserves more edge details and information, thereby enhancing the accuracy. Secondly, the cost of each and every pixel is replaced cost values of neighboring pixels, which is similar to Sum of Absolute Difference algorithm (SAD). Here the neighboring window is introduced, where larger window size leads to smoother edges in image. In order to achieve similar effects with global stereo matching algorithms, all pixels in the image are supposed to be involved in the constraints of the current pixel. Therefore, multi-direction cost aggregation is introduced to simulating the dependence of pixels in multiple paths on the current pixel. We formulate the aggregation cost as follows.

$$E(D) = \sum_{i}(c(i, D_j) + \sum_{j \in N_i})P1 * 1[|D_i - D_j = 1|] + \sum_{j \in N_i} P2 * 1[|D_i - D_j| > 1])$$

(a) before

(b) after

Figure 2: WLS example

In the formula, N_i denotes the neighboring window for pixel i. P1, P2 are penalty terms for cases where disparity difference is equal to 1 or larger than 1 respectively. After trial and error, empirical values for P1 and P2 are **10** and **120** in my implementation. Intuitively, larger P1 value leads to smoother image with fixed P2 value. After semi-global matching (SGM) optimization, winner-take-all strategy is used to calculate the disparity. Last but not least, post-processing method including weighted least square filtering and uniqueness check are introduced to improve performance of the algorithm.

2.2 Weighted least square filtering

The weighted least squares (WLS) [1] filter is a non-linear, edge-preserving, smoothing filter, which is first proposed in [17]. WLS filter can effectively capture details at multiple scales via multi-scale edge-preserving decomposition. It is designed to smooth the image without losing edge details and information. The result is similar to the source image whereas the region with small gradient is more smooth. Denote the source image as g, result image as u, weight matrix along x axis and y axis as A_x, A_y respectively. The loss function $f(u)$ is given as:

$$f(u) = (u-g)^T(u-g) + \lambda(u_x^T D_x^T A_x D_x u_x + u_y^T D_y^T A_y D_y u_y)$$

The original weighted least squares filtering is used to process two-dimensional images, and its purpose is to filter out weak edges and retain strong edges, thus giving stronger gradients greater weight and smaller gradients greater weight. Now our goal is to reduce strong interference and make the curve smoother. Instead, smaller weight can be assigned to the region with smaller weight and larger weight can be assigned to the region with larger weight.

3 Experiment

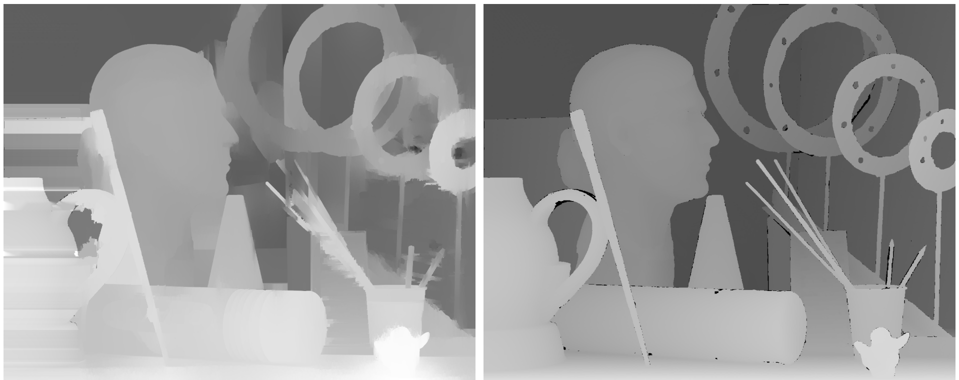
In my experiment, parameters including neighboring window size and number of disparity largely account for performance of the algorithm. The optimal setting for test cases **Art**, **Dolls** and **Reindeer** are as follows. Notably, number of disparity is supposed to be divisible by 16 in OpenCV implementation.
Art: number of disparity = 176, window size = 5, psnr = 17.00
Dolls: number of disparity = 176, window size = 3, psnr = 17.27
Reindeer: number of disparity = 176, window size = 11, psnr = 15.51

The quality of disparity map is evaluated by peak signal-to-noise ratio (PSNR), which is calculated as follows.

$$PSNR = 20 * \log_{10} \frac{max_i}{\sqrt{mse}}$$

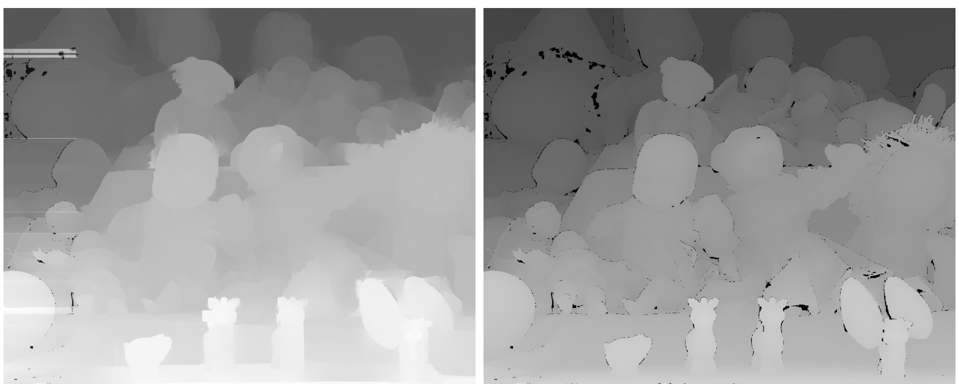
In the above formula, max_i denotes peak pixel value of the image and mean squared error (mse) is defined as:

$$mse = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} [I_1(i,j) - I_2(i,j)]^2$$



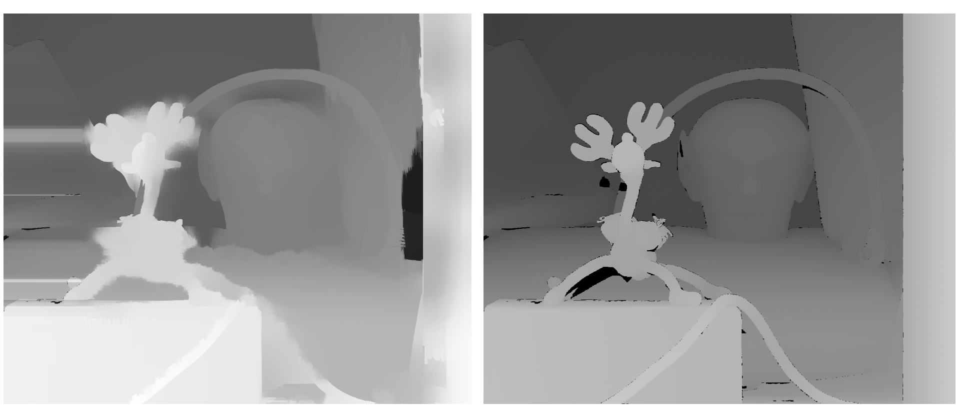
(a) psnr = 17.00 (b) GT

Figure 3: Art



(a) psnr = 17.27 (b) GT

Figure 4: Dolls



(a) psnr = 15.51 (b) GT

Figure 5: Reindeer

References

[1] Farbman, Z., Fattal, R., Lischinski, D., Szeliski, R., 2008. Edge-preserving decompositions for multi-scale tone and detail manipulation. ACM Trans. Graph. 27, 1–10. URL: https://doi.org/10.1145/1360612.1360666, doi:10.1145/1360612.1360666.

[2] Hirschmuller, H., 2008. Stereo processing by semiglobal matching and mutual information. IEEE Transactions on Pattern Analysis and Machine Intelligence 30, 328–341. doi:10.1109/TPAMI.2007.1166.