

Joint Bilateral Filter and Multi-Scale Cost Aggregation in Stereo Matching

Ye You-ping^{1,a}, Zheng Hong^{1,b}, Chen Hao^{1,c}, Yang Yu^{1,d}

¹School of automation and electrical engineering, Beihang University, Beijing 100191, China

^asxdfyyp@163.com, ^bjulyanna@vip.sina.com, ^c772143295@qq.com, ^d506283896@qq.com

Keyword: bilateral filter, multi-scale, stereo matching, cost aggregation

Abstract. Dense correspondence is a key problem in binocular stereo vision. The existing solution for this problem can be divided into local method and global or semi-global method. Bilaterally weighted patches matching is a classical local method, while it is computationally expensive and its accuracy need to be improved. In this paper, a new method is proposed which uses bilaterally weighted method to calculate matching cost and multi-scale method to produce cost aggregation. Use Gauss down-sampling method to produce multi-scale image, calculate matching cost in every sample scale. Then add inter-scale regularization into optimization function and solve the new optimization problem. This new algorithm is evaluated on Middlebury dataset, and it presents significant improvement than single bilaterally weighted method.

Introduction

Stereo matching algorithm generally comprises four steps: cost computation, cost aggregation, disparity computation, disparity refinement [1]. In cost computation, the matching cost of each pixel will be calculated, and then, the costs are aggregated, finally, the disparity of each pixel is calculated by different methods and the disparity will be refined by some post-processing methods.

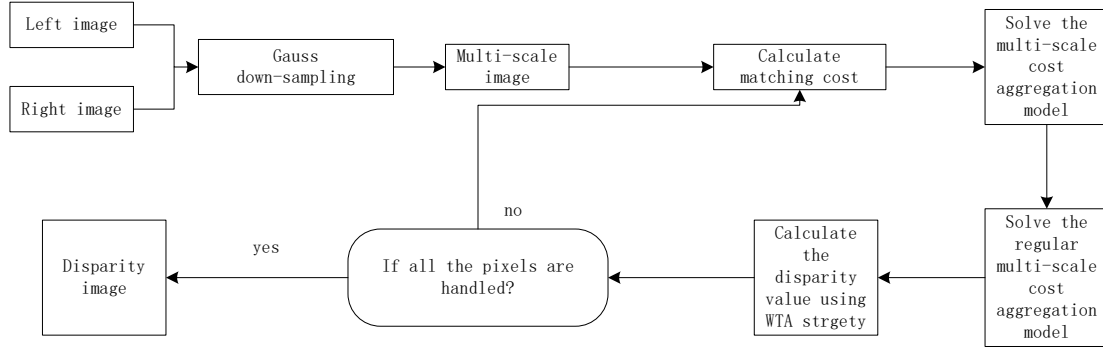
In every steps, most methods have been proposed, different methods have different impact on stereo matching algorithm. Actually, the choice of cost aggregation methods has most significant impact on the performance of matching algorithm. Some simple linear filters all can be used to produce cost aggregation such as box and Gaussian filter. Yoon and Kweon [2] use bilateral filter method to compute cost aggregation. Rhemann et al. [3] adopted the guided image filter into cost aggregation. Yang [4] proposed a non-local method in recent which is different from the local method and its kernel size is the entire image.

So far, all the algorithms are only processing the input images at the normal scale. Considering that eye processing information is a process from rough to precision, so we proposed to compute the cost aggregation in multiple scale. In this paper, our algorithm aim to improve the performance of the bilateral filter method. The algorithm can be summarized in following four parts:

- 1) compute sampling images at different scale
- 2) compute match cost and aggregate cost using bilateral filter kernel
- 3) calculate robust cost aggregation combining cost aggregation computed in step 2)
- 4) produce the disparity and refine the disparity

The method

The method can be divided into four parts as introduced above. Firstly, use Gauss down-sampling method to produce the multi-scale images. Then calculate the matching cost of every pixel. And next, aggregate the matching cost. The most important part of our method is using the multi-scale cost aggregation model and inter-scale regularization model to calculate the matching cost of every pixel. Finally, select the lowest cost as the matching cost of each pixel. The frame of our algorithm can be described as follows:



Cost computation

There are many methods used to calculate matching cost, such as SD, AD, CEN, CG, GRD. In this paper, we use the intensity + gradient cost function [3, 4]. And it can be formulated as:

$$c(i, l) = (1 - \alpha) \cdot \min(\|I(i) - I'(i_l)\|, \tau_1) + \alpha \cdot \min(\|\nabla_x I(i) - \nabla_x I'(i_l)\|, \tau_2) \quad (1)$$

Here $I(i)$ represents the color vector of pixel i . ∇_x is the gradient operator along X direction and the image must be grayscale. i_l is the matching pixel with i , and they have a disparity l . α is a term which is used to balance the gradient and color. τ_1, τ_2 are truncation values.

Cost aggregation

Considering that $c(i, l)$ is noisy, so the cost aggregation can be formulated as:

$$\bar{c}(i, l) = \arg \min_z \frac{1}{Z_i} \sum_{j \in N_i} K(i, j) \|z - c(j, l)\|^2 \quad (2)$$

Here $K(i, j)$ is a filter kernel which is used to measure the similarity between pixel i and pixel j . In this paper, it is a bilateral filter kernel. And $Z_i = \sum_{j \in N_i} K(i, j)$, is a normalized constant. We can

solve the problem (2) through derivation. So the solution can be described as:

$$\bar{c}(i, l) = \frac{1}{Z_i} \sum_{j \in N_i} K(i, j) c(j, l) \quad (3)$$

And the kernel $K(i, j)$ can be formulated as :

$$K(i, j) = \exp\left(\frac{-\Delta c_{ij}}{h_c^2}\right) \exp\left(\frac{-\Delta g_{ij}}{h_g^2}\right) \quad (4)$$

$$\Delta c_{ij} = \frac{1}{3} \|I(i) - I(j)\| = \frac{1}{3} \sum_{c \in \{r, g, b\}} |I_c(i) - I_c(j)| \quad (5)$$

$$\Delta g_{ij} = \sqrt{(i_x - j_x)^2 + (i_y - j_y)^2} \quad (6)$$

Multi-scale cost aggregation

We add s superscript s to the c, means the cost as different scale. So, the multi-scale cost formula can be expressed as:

$$\bar{c}^s(i^s, l^s) = \arg \min_{\{z^s\}_{s=0}^S} \sum_{s=0}^S \frac{1}{Z_{i^s}^s} \sum_{j^s \in N_{i^s}^s} K(i^s, j^s) \|z^s - c(j^s, l^s)\|^2 \quad (7)$$

Similarly, the problem can be solved as:

$$\forall s, \bar{c}^s(i^s, l^s) = \frac{1}{Z_{i^s}^s} \sum_{j^s \in N_{i^s}^s} K(i^s, j^s) c(j^s, l^s) \quad (8)$$

Inter-scale regularization

We use a vector $\bar{v} = (\bar{c}^0(i^0, l^0), \bar{c}^1(i^1, l^1), \dots, \bar{c}^S(i^S, l^S))$ represent the cost aggregation at every scale, and $\bar{v} = (\bar{c}^0(i^0, l^0), \bar{c}^1(i^1, l^1), \dots, \bar{c}^S(i^S, l^S))$ the cost aggregation with regularization.

Considering the relationship between the multi - scale images, we join the regularization term.

Then the problem can be shown as follows:

$$\bar{c} = \arg \min_{\{z^s\}_{s=0}^S} \left(\sum_{s=0}^S \frac{1}{Z_{i^s}^s} \sum_{j^s \in N_{i^s}^s} K(i^s, j^s) \|z^s - c(j^s, l^s)\|^2 + \lambda \sum_{s=0}^S \|z^s - z^{s-1}\|^2 \right) \quad (9)$$

In order to solve the problem (9), we can set the partial derivation of (9) to be equal with zero, so we can get following function.

$$-\lambda z^{s-1} + (1 + 2\lambda)z^s - \lambda z^{s+1} = \bar{c}^s(i^s, l^s) \quad (10)$$

The equation (10) can be expressed as a matrix.

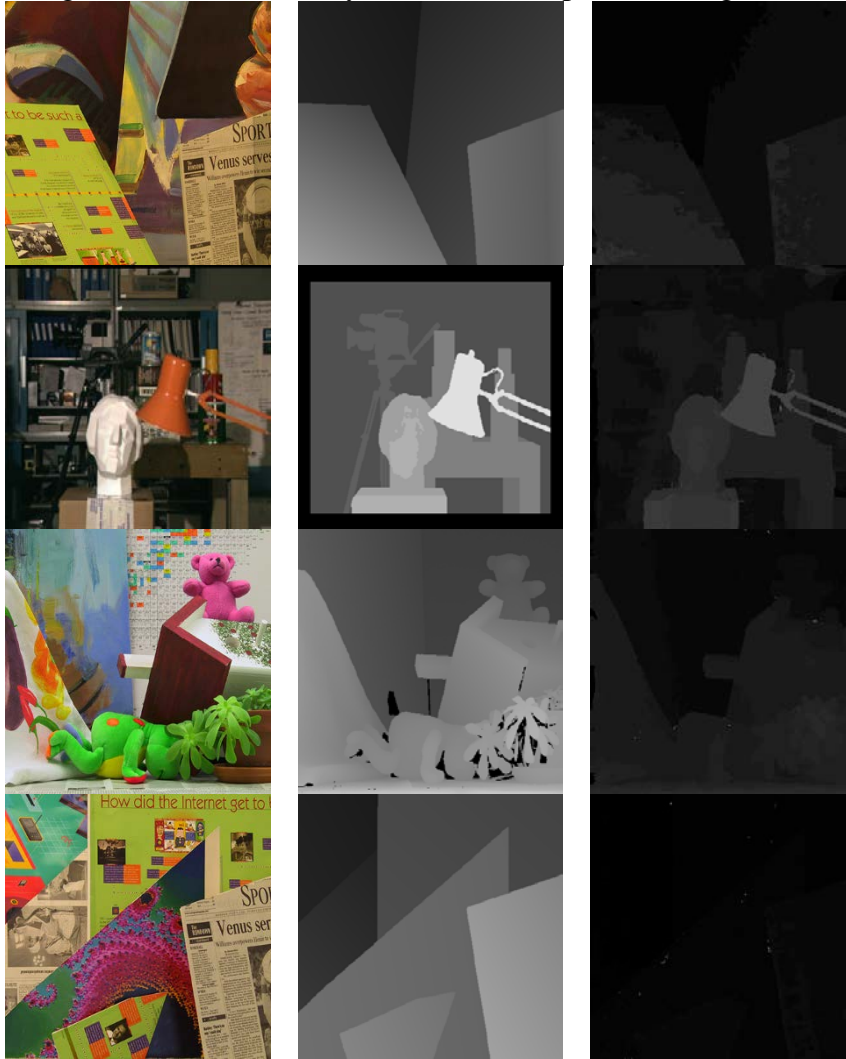
$$A\bar{v} = \bar{v} \quad (11)$$

Here, A is a constant $(S+1) \times (S+1)$ matrix. Finally, we can get the cost aggregation in normal scale which means the scale value s is equal to zero.

$$\bar{c}^0(i^0, l^0) = \sum_{s=0}^S A^{-1}(0, s) \bar{c}^s(i^s, l^s) \quad (12)$$

Experiment

We test our algorithm on Middlebury dataset, the output of our algorithm are shown as follows.



Left image

ground truth

our result

The first column are standard input images and the second are standard ground truth data. And the last column are the disparity images produced by our algorithm.

Conclusion

In this paper, we proposed an algorithm to improve the matching accuracy of traditional stereo matching method. Especially, the algorithm uses the characteristics of the human eye to see the image from clear to fuzzy. Use the method of Gauss pyramid to generate the images of different scale. And then use GRD method to calculate matching cost, BF method to aggregate cost and WTA(winner take all) method to produce disparity value. All these methods are applied in the images in every scale. Moreover, considering the relationship among images in different scale, we add the regularization term into optimization function. Finally, we test our algorithm on Middlebury dataset, it can get good performance no matter in time consumption or matching accuracy.

Reference

- [1] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. IJCV, 2002.
- [2] K.-J. Yoon and I. S. Kweon. Adaptive support-weight approach for correspondence search. TPAMI, 2006.
- [3] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz. Fast cost-volume filtering for visual correspondence and beyond. In CVPR, 2011.
- [4] Q. Yang. A non-local cost aggregation method for stereo matching. In CVPR, 2012.
- [5] Julia, L. F., & Monasse, P. (2015). Bilaterally Weighted Patches for Disparity Map Computation. Image Processing On Line, 5, 73-89.