Postprint of Stereo Matching Algorithm Based on Adaptive Weights and Occlusion Information

Authors: Yani Zhang, Diao Qinqing, Zhu Lingyun

Date: 2019-01-03T00:00:00+00:00

Abstract

To address the problem of high mismatch rates in traditional local stereo matching algorithms within depth-discontinuous regions, we propose a stereo matching algorithm based on adaptive weights and occlusion information. First, a left-right consistency check algorithm is employed to detect occlusion regions in both the reference and target images. Subsequently, occlusion information is utilized to reduce the weight of pixel points within occlusion regions during the cost aggregation stage. During the disparity optimization stage, a scanline propagation method is adopted to select the nearest point in the horizontal direction for filling the disparity values of occlusion regions. Finally, the mismatch rate is calculated for the disparity results based on the ground truth disparity maps provided by the Middlebury dataset. Experimental results demonstrate that the proposed occlusion-information matching algorithm based on adaptive weights reduces the mismatch rate by 16% compared to the adaptive weight algorithm, effectively solving the problem of high mismatch rates in local stereo matching algorithms within depth-discontinuous regions and improving the overall matching accuracy.

Full Text

Preamble

Vol. 37 No. 3
Application Research of Computers
ChinaXiv Cooperative Journal

Occlusion Information Stereo Matching Approach Based on Adaptive Weight

Zhang Yani[1], Diao Qinqing[1], Zhu Lingyun[2]†
(1. College of Computer Science & Engineering, Chongqing University of Technology, Chongqing 400054, China;

2. Liangjiang International College, Chongqing University of Technology, Chongqing 401135, China)

Abstract: To address the high mismatching rate of traditional local stereo matching algorithms in depth discontinuity regions, this paper proposes an occlusion information stereo matching approach based on adaptive weight. First, the left-right consistency check algorithm is employed to detect occlusion regions in both reference and target images. Then, occlusion information is utilized to reduce the weight of pixels within occlusion regions during cost aggregation, while scanline propagation is adopted in the disparity refinement stage to fill occlusion regions with horizontally nearest points. Finally, the mismatching rate is calculated based on the standard disparity maps provided by the Middlebury dataset. Experimental results demonstrate that compared with the adaptive weight approach, the proposed occlusion information stereo matching approach reduces the mismatching rate by 16%, effectively addressing the high mismatching rate issue in depth discontinuity regions and improving matching accuracy.

Keywords: stereo correspondence; adaptive weight; occlusion detection; cost aggregation; disparity refinement

---

0 Introduction

Stereo matching is the process of computing disparity maps from two images captured by two parallel cameras with identical parameters from different viewpoints. Scharstein and Szeliski conducted comprehensive research on stereo matching algorithms [?], categorizing them into four steps: matching cost computation, cost aggregation, disparity computation, and disparity refinement. Global stereo matching [?] performs disparity optimization across the entire stereo image pair by minimizing an energy function to compute disparity. Local stereo matching [?, ?] conducts cost aggregation within a support window and employs a simple WTA (winner-take-all) strategy for disparity selection. Both algorithms share a common assumption [?] that scenes are piecewise smooth —local methods assume smoothness within the support window, while global methods model smoothness directly. However, this assumption is often unrealistic in practice, as depth is typically discontinuous at object edges.

Depth discontinuity regions are usually accompanied by occlusion problems. In binocular stereo vision systems, occlusion phenomena [?] primarily arise from different viewing angles: the foreground region closer to the camera occludes portions of the background region farther from the camera, causing occluded background pixels to appear in only one image of the stereo pair. Due to the non-binocular visibility of occlusion problems, many algorithms treat occlusion recovery as a secondary issue either ignored entirely or handled simply during disparity refinement. Nevertheless, processing occlusion regions is essential for obtaining more accurate depth images.

Occlusion handling approaches generally fall into two categories. The first approach directly detects and recovers occlusion region pixels during disparity refinement, with common methods including left-right consistency checking (LRC) [?], sub-pixel interpolation [?], weighted median filtering [?], and image segmentation plane fitting [?]. Lu et al. [?] proposed a novel disparity refinement interpolation algorithm that employs LRC for occlusion detection and uses color and distance similarity for interpolation to improve occlusion estimation. Ye et al. [?] proposed a multi-step disparity refinement framework for occlusion handling that classifies outliers into three categories: leftmost occlusion, non-boundary occlusion, and mismatches, applying different recovery strategies for each type and introducing an order-based surface decision method for complex backgrounds or multiple background regions.

The second approach improves matching algorithms to enhance robustness against occlusion. For instance, Zhou et al. [?] proposed an adaptive weight Census transform stereo matching algorithm that computes grayscale differences to differentiate pixels within the window during the Census transform stage and introduces adaptive weights to modify the aggregation window size during cost aggregation, thereby improving matching performance in depth discontinuity regions. In reality, due to the non-binocular visibility of occlusion regions, their matching costs are significantly higher than those of non-occluded pixels, which can affect the matching results of non-occluded regions. Integrating occlusion information into the matching process to reduce the influence of occlusion region pixels during cost aggregation yields better matching results.

This paper presents an improved method based on the classic adaptive window weight stereo matching algorithm [?]. Under experimental conditions with occlusion information, occlusion region pixels are processed in two aspects. First, during cost aggregation, smaller weights are assigned to occlusion region pixels to reduce their impact on cost aggregation for pixels being matched. Second, during disparity refinement, disparity is recomputed for occlusion region pixels based on the prior knowledge that foreground occludes background. Experimental comparisons demonstrate that the proposed algorithm not only improves stereo matching accuracy in depth discontinuity regions but also mitigates issues such as foreground expansion and edge blurring in the resulting disparity maps.

1 Adaptive Weight Stereo Matching Algorithm

Yoon et al.' s adaptive weight stereo matching algorithm [?] originates from Gestalt grouping principles of proximity and similarity—objects closer in relative distance are more likely to be perceived as a whole, and under equal conditions, more similar objects are more likely to be grouped together. The algorithm treats color similarity and spatial proximity as weighting factors and divides matching into three steps: adaptive window weight computation, similarity difference calculation, and disparity selection.

1.1 Adaptive Window Weight Computation

The algorithm adjusts each pixel' s weight within the support window based on geometric proximity and color similarity to reduce ambiguous matching. Pixels closer in distance receive larger weights, and pixels more similar in color receive larger weights, with both constraints operating independently. The weight formula is described as:

$$w(p,q) = \exp\left(-\frac{\Delta c_{pq}}{\gamma_c} - \frac{\Delta g_{pq}}{\gamma_p}\right)$$

where p is the center pixel of the support window, q is a pixel within p' s support window, Δc_{pq} is the RGB color distance between p and q, Δg_{pq} is the Euclidean distance between p and q, γ_c and γ_p are parameters controlling color and distance respectively, with γ_p proportional to the window size.

1.2 Similarity Difference Calculation

The key to stereo matching lies in cost aggregation within the support window—smaller cost indicates higher similarity between matching points. Unlike other matching algorithms, this approach considers weight computation in both support windows of the reference and target images during cost aggregation to reduce matching errors in depth discontinuity regions. The similarity difference during cost aggregation can be described as:

$$E(p,d) = \frac{\sum_{q\in S_p} w(p,q) \cdot \sum_{q\in S_d} w(p_d,q) \cdot e(q,q_d)}{\sum_{q\in S_p} w(p,q) \cdot \sum_{q\in S_d} w(p_d,q)}$$

where points p_d and q_d in the target image correspond to matching points of p and q in the reference image with disparity d. The denominator serves normalization purposes. The matching cost computation formula is:

$$e(q,q_d) = \min\left(\sum_{c\in\{r,g,b\}} |I_c(q) - I_c(q_d)|, T\right)$$

where I_c is the color intensity of channel c, and T is a parameter controlling the matching cost range. Using the TAD (truncated absolute difference) method enhances robustness against outliers.

1.3 Disparity Selection

Among all possible disparity candidates for point p, the point with the minimum similarity difference obtained through cost aggregation is identified as the correct match, and its corresponding disparity is assigned to point p.

2.1 Problems in Stereo Matching Under Occlusion Conditions

Among local stereo matching algorithms, the adaptive weight algorithm achieves significantly higher matching accuracy than adaptive window algorithms, yet still exhibits high mismatching rates in textureless and depth discontinuity regions, with the vast majority of mismatches occurring in depth discontinuity regions. Due to different binocular camera viewpoints, depth discontinuity regions are typically accompanied by occlusion problems. As shown in Figure 1 [Figure 1: see original paper], point p exists in the reference image, but its corresponding match in the target image is occluded and therefore does not exist.

Occlusion-induced matching problems manifest in two primary aspects. First, occlusion region pixels have high matching costs. When cost aggregation for non-occluded pixels incorporates occlusion region pixels, the similarity difference becomes abnormally high, leading to incorrect matches. Second, occlusion pixels inherently lack corresponding matches, making disparity computation through other pixels inherently inaccurate. This paper improves the stereo matching algorithm by addressing both aspects: assigning lower weights to occlusion region pixels during cost aggregation and recomputing disparity for occlusion region pixels during disparity refinement.

2.2 Adaptive Weight Calculation Under Occlusion Conditions

For pixels located in occlusion regions, their corresponding matches are occluded by other regions and therefore do not exist. Consequently, matching cost computation involves comparing occlusion region pixels with pixels from other regions, typically yielding high cost values. During cost aggregation, if the support window of a pixel being matched contains occlusion region pixels, the similarity difference becomes elevated, causing incorrect matches. Therefore, occlusion information can be utilized to assign lower weights to occlusion region pixels during cost aggregation, thereby mitigating their adverse effects.

This paper employs left-right consistency checking (LRC) for occlusion region detection. Since the cost aggregation stage considers weight computation in both reference and target image support windows, occlusion detection must be performed separately for both images. The following describes the reference image occlusion region detection method. First, initial disparity maps for both reference and target images must be obtained. Then, each pixel p in the reference image is traversed, with its disparity denoted as d_p. The corresponding point p_d in the target image is located. If $|d_p - d_{p_d}| >$ threshold, pixel p is identified as an occlusion region pixel. Finally, the reference image occlusion map is obtained. The detection pseudocode is as follows:

```
Input: Reference image disparity map dR, Target image disparity map dT
```

```
      The width of dR, the height of dR, the channels of dR
Output: Reference occlusion image oR
for y = 0 to h
  for x = 0 to w
    d1 = dR[image->data + y*step + x*nchs]
    d2 = dT[image->data + y*step + (x-d1)*nchs]
    if abs(d1-d2) > threshold
      oR[image->data + y*step + x*nchs] = 255
    else
      oR[image->data + y*step + x*nchs] = 0
return oR
```

Using the detected occlusion information, smaller weights are assigned to occlusion region pixels during cost aggregation. The weight formula is described as:

$$w(p,q) = \begin{cases} \varepsilon, & \text{if } q \text{ is in occlusion region} \\ \exp\left(-\frac{\Delta c_{pq}}{\gamma_c} - \frac{\Delta g_{pq}}{\gamma_p}\right), & \text{otherwise} \end{cases}$$

where ε is a small non-zero value. Even corresponding pixels exhibit matching costs due to slight color variations (lighting effects). If the support window of a pixel being matched contains numerous occlusion region pixels, reducing the number of pixels in cost aggregation would lower the similarity difference, potentially causing incorrect disparity selection. The occlusion region detection and weight computation for the target image follow the same procedure as for the reference image.

2.3 Disparity Recalculation Under Occlusion Conditions

For occlusion region pixels, their corresponding matches do not exist in the other image. Therefore, computing disparity through other pixels is inherently inaccurate. This paper recomputes disparity for occlusion region pixels during the disparity refinement stage.

In practice, occlusion regions generally exist in background areas. Due to different viewing angles, foreground regions occlude portions of background regions. As shown in Figure 2 [Figure 2: see original paper], the red diagonal texture region is visible from the left reference view but belongs to the blind zone for the right view, making it an occlusion region in the reference image. Similarly, the green upper texture region is visible from the right target view but occluded from the left view, making it an occlusion region in the target image. Observation reveals that occlusion regions always exist in background areas.

When capturing two objects from the same viewpoint, the object closer to the camera belongs to the foreground region, while the farther object belongs to the background region. Depth discontinuity regions are typically accompanied by occlusion problems, where the occluded area is the background region closest to the foreground. In the binocular stereo vision model, the relationship between distance and disparity can be derived from triangulation principles:

$$Z = \frac{b \cdot f}{d}$$

where Z represents the object' s depth from the camera, f is the camera focal length, b is the baseline, and d is disparity. When b and f remain constant, objects closer to the camera have larger disparity values; therefore, background regions have smaller disparity values than foreground regions. Based on this principle, disparity is recomputed for occlusion region pixels during disparity refinement. First, the reference image disparity map d_R is obtained using the adaptive weight stereo matching algorithm with the new weighting scheme. Then, the initially detected reference image occlusion map o_R is scanned. If pixel p belongs to an occlusion region, horizontal scanning is performed left and right from this pixel to find the first non-occluded pixels on each side, denoted as p_a and p_b. Finally, corresponding to the reference image, the disparities $d_R(p_a)$ and $d_R(p_b)$ of these two pixels are compared, with the smaller disparity assigned to the occlusion pixel p. The algorithm flowchart is shown in Figure 3 [Figure 3: see original paper].

Processing occlusion region pixels in both cost aggregation and disparity refinement stages aims to improve the matching accuracy of the traditional adaptive weight algorithm in depth discontinuity regions and mitigate issues such as foreground expansion and edge blurring in the resulting disparity maps. Specific experimental procedures and results are analyzed in Section 3.

2.4 Mismatch Rate Calculation

This paper evaluates matching algorithms by computing the error percentage of disparity maps as follows:

$$\delta = \frac{1}{N} \sum_{(x,y)} \left(|d(x,y) - d_g(x,y)| > \theta \right)$$

where N is the total number of pixels in the disparity image, $d(x,y)$ is the disparity value of pixel (x,y) obtained by the algorithm, $d_g(x,y)$ is the disparity value of the corresponding pixel in the ground truth provided by the Middlebury dataset, and θ is a threshold (typically 1). For better evaluation, error percentages are usually computed for three regions: non-occluded regions (nocc), all regions (all), and depth discontinuity regions (disc).

3 Analysis and Discussion

To evaluate the proposed algorithm' s matching performance, experiments were conducted from two perspectives: (a) detecting occlusion regions in reference and target images using left-right consistency checking; and (b) processing occlusion region pixels during cost aggregation and disparity refinement under occlusion-aware conditions to obtain new disparity maps. The experimental software platform was VS2010 with algorithms implemented in C++, using the OpenCV library for image reading, display, and storage. The hardware platform consisted of an Intel Core i7-8550U processor at 1.80GHz with 8GB RAM. Experimental test data were sourced from four standard image pairs (Tsukuba, Venus, Teddy, Cones) in the Middlebury standard image dataset.

3.1 Occlusion Detection Results and Analysis

Occlusion detection employed the left-right consistency check (LRC) method using disparity maps obtained from the classic adaptive window weight stereo matching algorithm [**?**]. Since the classic adaptive weight algorithm considers weights in both reference and target image support windows during cost aggregation, occlusion regions must be detected separately for both images. The experimental parameters are shown in Table 1 , where T controls the matching cost range, W is the support window width, γ_p and γ_c control distance and color weights respectively, and Threshold is the threshold value for the LRC detection algorithm.

In occlusion images, occlusion region pixels are black while other regions are white. Figure 4 [Figure 4: see original paper] shows that occlusion typically occurs in depth discontinuity regions, and objects closer to the imaging device (with larger disparity) have larger occlusion areas. It also reveals that the classic adaptive weight stereo matching algorithm still suffers from high mismatching rates in depth discontinuity regions.

3.2 Stereo Matching Results and Analysis

After obtaining occlusion region images, the improved algorithm assigns lower weights to occlusion region pixels during cost aggregation. Here, $\varepsilon = 0.1$, with the reason for not setting it to 0 being to balance the matching costs that should exist, preventing mismatches caused by reducing the number of pixels in cost aggregation. During disparity refinement, disparity is recomputed for occlusion region pixels based on the prior knowledge that foreground occludes background. The disparity maps obtained by the proposed algorithm are shown in Figure 5 [Figure 5: see original paper].

The results demonstrate satisfactory disparity maps with clear differentiation of objects in depth discontinuity regions. Given disparity maps, baseline, and calibrated focal length information, 3D world coordinates can be computed via triangulation. In disparity maps, stronger color intensity indicates larger disparity values and closer distance. Compared with ground truth, the proposed algorithm still exhibits some mismatches in low-texture and texture-repetition regions, which warrants further research.

3.3 Algorithm Comparison and Analysis

To verify matching accuracy, the obtained disparity maps were compared against ground truth, computing error percentages for non-occluded regions (nocc), all regions (all), and depth discontinuity regions (disc), along with average error percentages. Comparisons were made against the classic adaptive window stereo matching algorithm [?], an improved adaptive weight algorithm [?], and an adaptive weight Census transform stereo matching algorithm [?]. Results are presented in Table 2 .

As shown in Table 2, the proposed algorithm reduces the average mismatching rate by 16% compared to the classic adaptive weight stereo matching algorithm [?]. Paper [?] represents an improved adaptive weight algorithm, while [?] targets high mismatching rates in depth discontinuity regions. The proposed algorithm achieves lower mismatching rates than both. Since this work specifically addresses occlusion detection and handling in depth discontinuity regions —detecting occlusion areas and recovering them based on foreground-occluding-background principles—the error percentage in depth discontinuity regions is reduced. Simultaneously, reducing the influence of occlusion region pixels during cost aggregation for non-occluded pixels improves overall matching accuracy.

For clearer demonstration of improvement effects, Figure 6 [Figure 6: see original paper] compares the proposed algorithm with the adaptive weight algorithm [?]. The results show that the proposed algorithm not only reduces mismatching rates in depth discontinuity regions but also improves overall matching accuracy. The algorithm structure is simple, and future experiments can introduce multi-threading to enhance computational efficiency.

4 Conclusion

This paper assigns lower weights to occlusion region pixels during cost aggregation, reducing their impact on cost aggregation for pixels being matched and improving overall stereo matching accuracy. During disparity refinement, disparity is recomputed for occlusion region pixels based on the foreground-occluding-background prior knowledge, enhancing accuracy for occlusion region pixels and reducing mismatching rates in depth discontinuity regions. Experimental results demonstrate that the proposed algorithm mitigates foreground expansion and edge blurring issues in disparity maps. Compared with classic adaptive stereo matching algorithms and their improvements, the proposed algorithm enhances the accuracy of local stereo matching algorithms in depth discontinuity regions.

While the proposed algorithm achieves higher matching accuracy than the adaptive weight stereo matching algorithm, its computational efficiency is slightly lower. Future work will focus on improving algorithm runtime efficiency through further research and experimentation.

References

[1] Scharstein D, Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms [J]. International Journal of Computer Vision, 2002, 47 (1-3): 7-42.

[2] Zhang Gexiang, Rong Haina, Song Hui, et al. Multi-aperture stereo reconstruction for artificial compound eye with cross image belief propagation [J]. Applied Optics, 2018, 57(7): B160.

[3] Wang Qiong, Jiang Xiaoliang, Yuan Jianying, et al. Regional stereo matching algorithm based on improved Census transform [J]. Application Research of Computer, 2015, 32(10): 3164-3167.

[4] Lee S, Hong H. Near-real-time stereo matching method using both cross-based support regions in stereo views [J]. Optical Engineering, 2018, 57(2): 023103.

[5] Mattoccia S, Giardino S, Gambini A. Accurate and efficient cost aggregation strategy for stereo correspondence based on approximated joint bilateral filtering [C]// Proc of the 9th Asian Conference on Computer Vision. New York: ACM Press, 2012: 371-380.

[6] Malathi T, Bhuyan M K. Asymmetric occlusion detection using linear regression and weight-based filling for stereo disparity map estimation [J]. IET Computer Vision, 2017, 10 (7): 679-688.

[7] Lü Niqi, Song Guanghua, Yang Bowei. Semi-global stereo matching algorithm based on feature fusion and its CUDA implementation [J]. Journal of Image and Graphics, 2018, 23(6): 0874-0886.

[8] Bian Jilong, Li Jinfeng, Luo Siqing. Sub-pixel stereo matching method based on variational principle [J]. Application Research of Computer, 2014, 31(9): 2846-2849.

[9] Zhang Qi, Xu Li, Jia Jiaya. 100+times faster weighted median filter (WMF) [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2014: 2830-2837.

[10] Xie Yuechao, Zeng Siyu, Chen Long. A novel disparity refinement method based on semi-global matching algorithm [C]// Proc of IEEE International Conference on Data Mining Workshop. Piscataway, NJ: IEEE Press, 2015: 1135-1142.

[11] Lu Jun, Fang Ying, Zhang Xin. Stereo matching and occlusion handling based on dynamic programming using AD and Census transform [J]. Transactions of Beijing Institute of Technology, 2015, 35(12): 1274-1279.

[12] Ye Xiao, Gu Yuzhang, Chen Lili, et al. Order-based disparity refinement including occlusion handling for stereo matching [J]. IEEE Signal Processing Letters, 2017, 24 (10): 1483-1487.

[13] Zhou Wangwei, Jin Wenguang. Novel stereo matching algorithm for adaptive weight Census transform [J]. Computer Engineering and Applications, 2016, 52(16): 192-197.

[14] Yoon K J, Kweon I S. Adaptive support-weight approach for correspondence search [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2006, 28(4): 650-656.

[15] Yang Kui, Zhao Yan, Su Qinghua, et al. Fast recursive adaptive weight stereo matching [J]. Journal of Beijing University of Aeronautics and Astronautics, 2013, 39(7): 963-967.

Note: Figure translations are in progress. See original paper for figures.