Pattern Recognition 55 (2016) 198–206

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

An $O(1)$ disparity refinement method for stereo matching

Xiaoming Huang, Yu-Jin Zhang *

Department of Electronic Engineering, Tsinghua University, Tsinghua National Laboratory for Information Science and Technology, Beijing 100084, China

ARTICLE INFO

Article history:
Received 10 March 2015
Received in revised form
9 December 2015
Accepted 21 January 2016
Available online 5 February 2016

Keywords:
Stereo matching
Disparity refinement
Belief aggregation
Belief propagation
Non-local

ABSTRACT

Disparity refinement is the final step but also the timing bottleneck of stereo matching due to its high computational complexity. Weighted media filter refinement method and non-local refinement method are two typical refinement methods with $O(N)$ computational complexity for each pixel where N indicates the maximum disparity. This paper presents an $O(1)$ disparity refinement method based on *belief aggregation* and *belief propagation*. The *aggregated belief*, which means the possibility of correct disparity value, is efficiently computed on a minimum spanning tree first, and then the *belief aggregation* is fast performed on another minimum spanning tree in two sequential passes (first from leaf nodes to root, then from root to leaf nodes). Only 2 additions and 4 multiplications are required for each pixel at all disparity levels, so the computational complexity is $O(1)$. Performance evaluation on Middlebury data sets shows that the proposed method has good performances both in accuracy and speed.

1. Introduction

Stereo matching algorithms usually consists of four steps [1]: matching cost computation, cost aggregation, disparity computation and disparity refinement.

A lot of work has been spent on developing robust cost computation [2,3] and cost aggregation methods [4–10], while the high complexity of disparity refinement becomes the timing bottleneck in stereo matching algorithm. For example, for the guided image based algorithm [11], the average runtime of disparity refinement is about 6.5 seconds on the Middlebury data sets [12], reported by Yang [13]. Traditional refinement step include left–right check [14], hole filling and a median filter.

Weighted media filter refinement method is widely adopted by many stereo matching algorithms (*e.g.*, [11,15]). But the high complexity of this filter becomes the timing bottleneck. Recently, a constant time weighted median filtering [16] was proposed. The author proposed refinement with the help of a histogram whose size is maximum disparity N. This algorithm is driven by the recent progress on fast median filtering [17–19], fast algorithms [20–23] for bilateral filtering [24], and other fast edge-aware filtering [25,26], these existing $O(1)$ edge-aware filters can be performed on each histogram bin. The method shows good accuracy at slow speed due to $O(N)$ computational complexity for each pixel, the speed evaluation of this method is shown in Section 4.2. Moreover, one main shortage of this refinement method is that the support windows are of fixed size.

Yang proposed a non-local refinement method [13] using a non-local aggregation method on MST (minimum spanning tree) structure. All pixels are firstly divided into stable or unstable pixels after left–right disparity check, a new cost volume will be computed based on this checked disparity map, then followed by non-local aggregation at each disparity level and a winner-take-all operation to propagate the disparity values from stable pixels to unstable pixels. But the speed is still very slow due to $O(N)$ computational complexity for each pixel, the speed evaluation of this method is shown in Section 4.2.

In previous work [27], we presented one fast disparity refinement method based only on belief propagation. The method shows good performance on complicated cost aggregation methods (*e.g.*, guided filter aggregation [11], non-local aggregation [13]) but failed in simple cost aggregation methods such as box-filter aggregation [1].

In this paper, we proposed a fast refinement based on *belief aggregation* and *belief propagation*.

All pixels firstly have initial disparity belief. We build a hybrid MST whose edge weight is determined by disparity distance and color distance, *belief aggregation* is efficiently computed on this hybrid MST in two sequential passes (same to cost aggregation in [13], first from leaf nodes to root, then from root to leaf nodes). The pixel has greater *aggregated belief* if it has larger close neighbors both in disparity and color. Then, we build another MST whose edge weight is only determined by color distance, *belief propagation* is fast performed on this MST in two sequential passes (first from leaf nodes to root, then from root to leaf nodes). The pixel having lower *aggregated belief* will receive propagation from the pixel with

* Correspondence author. Tel.: +86 10 62798540; fax: +86 10 62770317.
E-mail addresses: huangxm14@mails.tsinghua.edu.cn (X. Huang), zhang-yj@mail.tsinghua.edu.cn (Y.-J. Zhang).

http://dx.doi.org/10.1016/j.patcog.2016.01.025
0031-3203/

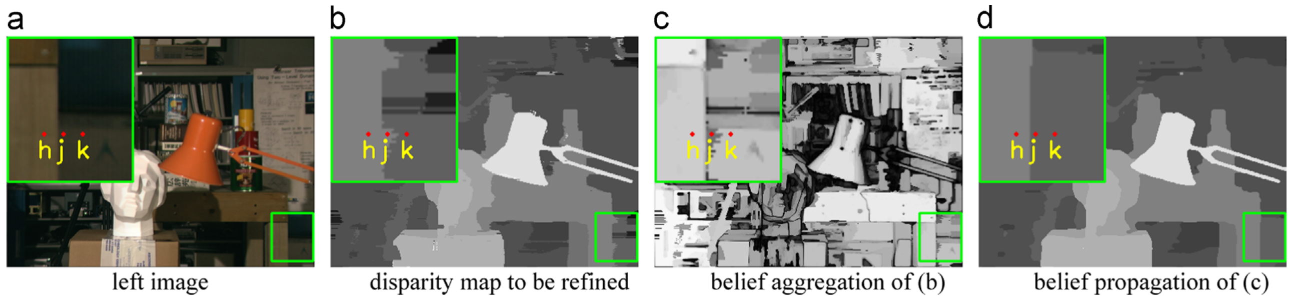

Fig. 1. Proposed refinement process of *tsukuba* data set. (a) Left image of *tsukuba* data set. (b) Guided filter aggregation [11] followed by left–right check and hole filling. (c) *belief aggregation* of (b), brighter color indicates higher belief. (d). *belief propagation* of (c). Although the disparity of h and j is same in (b), but h has larger close neighbors both in disparity and color, so h has greater *aggregated belief* in (c). Compared with k, pixel j has close color in (a) but smaller *aggregated belief* in (c), so j receives propagation from k, final disparity of j is assigned by the disparity of k. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

similar color and higher *aggregated belief*. Proposed refinement process on *tsukuba* data set is demonstrated in Fig. 1. Left image of *tsukuba* and disparity map to be refined is shown in Fig. 1(a) and (b), *belief aggregation* result is presented in Fig. 1(c), *belief propagation* result is presented in Fig. 1(d).

In proposed refinement method, only 2 additions and 4 multiplications are required for each pixel at all disparity levels, so the computational complexity is $O(1)$. Performance evaluation tested on Middlebury data set [12] shows good performance both in accuracy and speed.

The remaining of the paper is organized as following: Previous work is introduced in Section 2. Section 3 gives the details of the proposed disparity refinement algorithm. The experimental results are presented in Section 4. Finally, the paper concludes in Section 5.

2. Previous work

In this section, we mainly review the MST (minimum spanning tree) based non-local filter [13].

The reference color/intensity image I is represented as a connected, undirected graph $G=(V, E)$, where each node in V corresponds to a pixel in I, and each edge in E connects a pair of neighboring pixels. The graph G is thus simply the standard 4-connected or 8-connected grid. For an edge e connecting pixels s and r, its weight is determined as follows:

$$w(s,r)=w(r,s)=|I(s)-I(r)| \tag{1}$$

A tree T can be constructed by selecting a subset of edges from E. Yang [13] proposed to construct a MST connecting all the pixels and the sum of its weights is minimized. For any two pixels p and q, their distance $W(p,q)$ is determined by the sum of the edge weights along the path in T, and

$$S(p,q)=\exp\left(-\frac{W(p,q)}{\sigma}\right) \tag{2}$$

denotes the similarity between p and q where σ is a parameter to adjust the similarity between two nodes.

Let $C_d(p)$ denote the matching cost for pixel p at disparity level d. The final aggregated cost of pixel p at disparity level d is computed as follows:

$$C_d^A(p)=\sum_{q\in I}S(p,q)C_d(q) \tag{3}$$

Different from the local filtering-based methods, in the non-local cost aggregation method, p gets support weights from all the pixels in I. Yang proved that the non-local cost aggregation can be accomplished in exactly linear time by traversing the tree structure in two sequential passes: first from leaf to root, then from root to leaf.

In the first pass from leaf to root, the intermediate aggregated cost $C_d^{A\uparrow}(p)$ of each node p can be computed:

$$C_d^{A\uparrow}(p)=C_d(p)+\sum_{q \text{ is child of } p}S(p,q)C_d^{A\uparrow}(q) \tag{4}$$

Note that $C_d^{A\uparrow}(p)$ equals to the final cost aggregation $C_d^A(p)$ if p is the root node.

In the second pass from root to leaf, the final aggregated cost $C_d^A(p)$ of each node p can be computed:

$$C_d^A(p)=S(p,q)C_d^A(q)+(1-S^2(p,q))C_d^{A\uparrow}(q) \tag{5}$$

where pixel q is parent of pixel p.

The non-local filter is an efficient filter with the following advantages:

1. It provides a non-local solution, which theoretically and experimentally outperforms local cost aggregation methods.
2. It has low computational complexity: only 2 addition/subtraction operations and 3 multiplication operations are required for each pixel at each disparity level.
3. It can be used for non-local disparity refinement, which is proved to be more robust and effective than weighted media filter refinement method presented in [11].

3. Proposed disparity refinement

In this section, we first propose *belief aggregation* method on a hybrid MST whose edge weight is determined by disparity distance and color distance, then we present *belief propagation* method on another MST whose edge weight is only decided by color distance, and lastly we discuss the computational complexity of the algorithm.

3.1. Belief Aggregation on Hybrid MST

In a general local stereo matching algorithm, left and right disparity maps are obtained separately. Then left–right disparity check divides all the pixels into stable or unstable pixels. If left disparity is equal to the corresponding right disparity, the pixel is regarded as a stable pixel, otherwise it is considered as an unstable pixel. In addition, a hole filling step estimates disparity from stable pixels to unstable pixels. Fig. 1(b) shows one disparity map with guided filter aggregation method [11] followed by left–right check and hole filling.

In this paper, we first propose *disparity belief* for each pixel to represent the possibility of correct disparity value. Obviously, the pixel has greater *disparity belief* if it has large number of close neighbors both in disparity and color. If we construct a hybrid MST