Pattern Recognition Letters 112 (2018) 168–175

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Deep self-guided cost aggregation for stereo matching

Williem [a], In Kyu Park [b,*]

[a] Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480
[b] Department of Information and Communication Engineering, Inha University, Incheon 22212, Republic of Korea

ARTICLE INFO

Article history:
Received 22 September 2017
Available online 6 July 2018

MSC:
41A05
41A10
65D05
65D17

Keywords:
stereo matching
cost aggregation
deep learning
guided-filter

ABSTRACT

In this paper, we present a deep self-guided cost aggregation method used to obtain an accurate disparity map from a pair of stereo images. Conventional cost aggregation methods typically perform joint image filtering on each cost volume slice. Thus, a guidance image is necessary for the conventional methods to work effectively. However, a guidance image might be unreliable due to several distortions, such as noise, blur, radiometric variation. Based on our observations, each cost volume slice can guide itself based on the internal features. However, finding a direct mapping function from the initial and filtered cost volume slice without any guidance image is difficult. To solve this problem, we use an advanced deep learning technique to perform self-guided cost aggregation. Because of the absence of ground truth cost volume, we offer the solution for the dataset generation. Our proposed deep learning network consists of two sub-networks: *dynamic weight network* and *descending filtering network*. We integrate the feature reconstruction loss and the pixelwise mean square loss function to preserve the edge property. Experimental results show that the proposed method achieves better results even though it does not employ a guidance image.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

During the last few decades, stereo matching has been a popular means of finding a correspondence map between two rectified images. Extensive studies have been conducted to obtain either fast or accurate stereo matching results [5,8,10,29,32]. In general, a stereo matching method consists of four steps: cost computation, cost aggregation, disparity computation, and disparity refinement [25]. In local or global stereo matching, cost aggregation plays a major role. Since Yoon and Kweon [33] introduced an adaptive method that achieves comparable results to those of global stereo matching, several cost aggregation methods have been developed [11,27]. Most of these methods have a similar concept to that of joint image filtering that utilizes a guidance image. Thus, the existence of a guidance image is inevitable. These methods assume that the weights of corresponding pixels on a color image are approximately constant to those on a disparity image. However, this assumption often fails to hold true when a patch with a high texture variance has a similar disparity value. In addition, the conventional methods fail when the guidance image is unreliable. Various stereo matching methods have been introduced to deal with unreliable images due to noise [7], radiometric variation [13,21], blur [2,30], and severe weather condition [19].

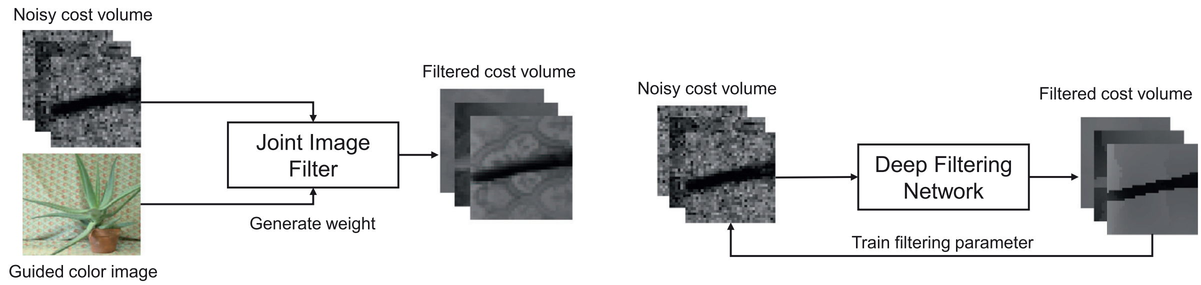
In this paper, we propose a novel cost aggregation method that removes the dependency on the guidance image. We observe that the initial cost volume possesses specific features that act as a guidance for itself. However, finding a correlation between non-filtered and filtered cost volume without any guidance image is complicated. Therefore, we employ a deep convolutional neural network (CNN) in our approach to enable a cost volume slice to perform a self-guided cost aggregation. Fig. 1 illustrates the conceptual differences between the conventional and proposed cost aggregation methods. We introduce two sub-networks, *dynamic weight network* and *descending filtering network*, to estimate the importance weight for each pixel and simultaneously to perform edge-aware filtering operator. To the best of our knowledge, the proposed work is both the first self-guided and first local cost aggregation method based on the deep learning approach. In summary, this paper makes the following contributions.

- We introduce a self-guided cost aggregation method for stereo matching that does not require any guidance color image.
- We propose a novel deep convolutional network consisting of two sub-networks: *dynamic weight network* and *descending filtering network*.

This paper is organized as follows. We introduce the related works in Section 2. Section 3 describes the detail of the proposed

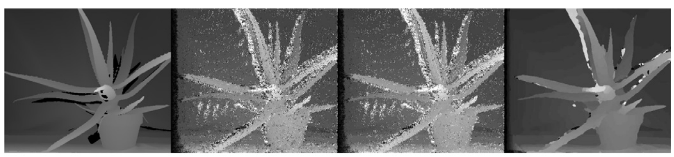
* Corresponding author.
E-mail addresses: william003@binus.ac.id (Williem), pik@inha.ac.kr (I.K. Park).

https://doi.org/10.1016/j.patrec.2018.07.010
0167-8655/© 2018 Elsevier B.V. All rights reserved.



(a) (b)

Fig. 1. The concept of the (a) conventional and (b) proposed cost aggregation methods.



(a) (b) (c) (d)

Fig. 2. Disparity maps of Aloe data. (a) Ground truth; (b) From the initial cost volume; (c) From the self-guided filtered cost volume (Guided filter [12] using the input cost volume as its guidance); (d) From the deep self-guided filtered cost volume (Proposed method).

method. Experimental results are presented in Section 4. Finally, we conclude in Section 5.

2. Related works

Stereo matching has been an active research topic during the last few decades. Scharstein and Szeliski [25] surveyed various stereo matching techniques, a study that later became the main guideline for stereo matching research. Although four main steps are used in stereo matching, the focus of this study is solely on the cost aggregation step. Tombari et al. [27] were the first to evaluate the cost aggregation step. They classified various cost aggregation strategies and performed numerous experiments on known techniques. Hosni et al. [11] extended the work in [27] by adding new adaptive support weight (ASW) techniques, evaluating larger datasets, and providing general insights about ASW techniques. They showed that cost aggregation using a guided filter [12] performed better than other methods in terms of accuracy and computational time.

In early studies on cost aggregation, Yoon and Kweon [33] utilized a symmetric bilateral filtering method to filter 3D cost volume during the cost aggregation step. The weight for each pixel was computed based on the color similarity and spatial distance from the center pixel inside the support window. Their method performed better than previous local algorithms and comparably to global matching algorithms. Zhang et al. [36] introduced a method to generate a shape-adaptive support region based on the upright cross local support skeleton. The cross skeleton was reconstructed based on color similarity and connectivity constraints. In addition, the researchers introduced an efficient means to perform cost aggregation using orthogonal integral images. Hosni et al. [12] proposed an efficient cost volume filtering method that outperformed [33] in terms of both quality and computational speed. They employed the most popular edge-aware filter known at the time, called *guided filter* [6]. Pham and Jeon [20] employed the domain transform [4] to perform edge-aware cost aggregation.

Instead of reducing the complexity with respect to matching window size, Min et al. [18] reduced the computational redundancy among the search spaces. They proposed an efficient cost aggregation method based on a joint histogram. Two approximations of a search space and window size were used simultaneously for fast computation. Yang [31] introduced a non-local cost aggregation method based on a minimum spanning tree (MST). They inserted a full image into a tree structure so that each pixel in the image could efficiently contribute to other similar pixels. However, their method depends on the local weight which is set as an edge in the tree structure. To improve performance, Mei et al. [15] proposed a novel tree structure called a segment tree (ST). They generated a sub-tree for each segment in an image and then connected those sub-trees to form a fully segmented tree. They considered not only local edge weights but also non-local segment properties for each pixel contribution.

Although most methods have been developed in a single scale, [35] and [3] proposed an integrated solution of various scale cost aggregation methods. Zhang et al. [35] introduced an optimization function that employs a regularizer to aggregate cost across multiple scales. Choi and Chang [3] determined pixelwise mixing coefficients of each scale filter adaptively. These coefficients were obtained by performing supervised learning. Both of the aforementioned methods outperformed the single scale cost aggregation methods.

Note that most conventional methods require a guidance image in order to perform cost aggregation. In this paper, our focus is to develop a novel cost aggregation method that is free of a guidance image. Note that the absence of the guidance image is a big handicap because much less information exists to perform edge-aware filtering. Therefore, we adopt the deep learning approach during the cost aggregation step. To the best of our knowledge, the proposed method is the first deep learning based and self-guided cost aggregation method.

3. Proposed method

3.1. Stereo matching

In our study, we focus on local stereo matching to evaluate the performance of the proposed method. We employ *Census filter* [34] to compute the matching cost, which is proven to be robust when illumination varies [9]. The matching cost is described as follows.

$$F_j|_{j \in N_i} = \begin{cases} 1 & \text{if } I_j < I_i \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where i and N_i are the coordinate of a center pixel and its pixel set in the support region, respectively. I_j and F_j denote the intensity and bit value of pixel j, consecutively. A bit string is generated from a 9×7 window and stored in 64-bit integer. Using the bit string, we measure the matching cost $C^O_{i,l}$ for each pixel i and disparity label candidate l. Hirschmüller and Scharstein [9] utilize the

Download English Version:

https://daneshyari.com/en/article/6940133

Download Persian Version:

https://daneshyari.com/article/6940133

Daneshyari.com