Signal Processing: *Image Communication* 39 (2015) 173–184

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Robust non-local stereo matching for outdoor driving images using segment-simple-tree

Cuong Cao Pham, Vinh Quang Dinh, Jae Wook Jeon *

College of Information and Communication Engineering, Sungkyunkwan University, Suwon, Republic of Korea

ARTICLE INFO

Article history:
Received 30 April 2015
Received in revised form
17 August 2015
Accepted 18 September 2015
Available online 3 October 2015

Keywords:
Stereo matching
Non-local cost aggregation
Tree structure
Driver assistance systems

ABSTRACT

Non-local cost aggregation has recently emerged as a promising approach for stereo-matching and has attracted much interest over the past few years. Most non-local algorithms are reportedly better than state-of-the-art local algorithms for high-quality indoor images. However, the accuracy of non-local algorithms is still limited for outdoor images. Computing disparity maps for outdoor images in driver assistance systems is one of the most actively researched topics in the field of stereo vision. In this paper, we present a robust non-local stereo matching algorithm that improves the performance of non-local approaches for outdoor driving images. The proposed algorithm is inspired by the non-local cost aggregation method based on a minimum spanning tree, and it improves the estimation accuracy by introducing an alternate, effective segment-simple-tree that is more adequate for outdoor driving images than the minimum spanning tree. Experimental results showed that the proposed algorithm is superior to the existing local and non-local algorithms, and is comparable to semi-global matching.

1. Introduction

Binocular stereo matching is one of the most important algorithms in computer vision because it provides computers with the depth perception ability similar to that of humans. The goal is to estimate the disparity maps from two rectified images of the same scene taken from left and right viewpoints. Binocular stereo matching has been extensively studied over the past few decades, and numerous algorithms have been proposed [1–3]. Nevertheless, it is still an active area of research because new challenges have surfaced and there are a variety of problems that have not yet been solved. Among these challenges, computing the disparity maps for outdoor images in driver assistance systems (DAS) is one of the most actively researched topics in this field [4–6].

In general, a stereo matching algorithm consists of four steps: cost computation, cost aggregation, disparity optimization, and post-processing. In this paper, we mainly focus on the cost aggregation step since it has the greatest impact on the accuracy of the estimated disparity map. Most existing cost aggregation algorithms define a local 2-D support window for each pixel and perform summing/averaging operations using the information obtained from the pixels inside of that window. State-of-the-art local algorithms include the adaptive support-weight approach [7], geodesic diffusion [8], and fast cost-volume filtering [9]. These were respectively inspired by bilateral filtering [10], anisotropic diffusion [11], and guided image filtering [12], which are the three most well-known edge-aware image filtering methods.

Semi-global matching (SGM) [13] is one of the best performing stereo matching algorithms for outdoor driving images [4]. SGM performs the cost aggregation step using several global paths instead of the local pixel similarities used in the local approach. Even though this is an optimization-based approach that uses dynamic programming for each path, it can still be considered as a cost aggregation step [14].

* Corresponding author.
E-mail addresses: cuongpc@skku.edu (C.C. Pham), dqvinh@skku.edu (V.Q. Dinh), jwjeon@yurim.skku.ac.kr (J.W. Jeon).

http://dx.doi.org/10.1016/j.image.2015.09.012
0923-5965/

In the past few years, non-local cost aggregation has recently emerged as a promising approach since the accuracy and the computation are reportedly better and faster than existing local algorithms. The key idea behind its success is that it aggregates the matching cost using the information from all of the pixels in the image rather than only those inside a specific window, as the aforementioned local algorithms do [7–9]. Cigla et al. [15] proposed an information permeability algorithm with separable successive weighted summation along the horizontal and vertical directions, and the improved spatio-temporal information permeability filtering is presented in [16]. Pham et al. [17] employed a sequence of 1-D filters to reduce the computation and memory costs relative to conventional 2-D filters by applying domain transform [18].

Yang's method [19,20], the most notable non-local algorithm, carries out the cost aggregation on a minimum spanning tree (MST). The MST is constructed using a subset of edges whose sum of edge-weights is minimal, and the matching cost is then aggregated on this tree within two passes, leaf-to-root and root-to-leaf. Mei et al. [21] improved upon this algorithm by incorporating image segmentation [22] in the tree construction procedure. This results in the new tree structure, which is called segment-tree (ST). Although these methods generate accurate disparity maps for indoor images as shown in [19–21], they fail to do the same for outdoor driving images. Both MST and ST are not adequate for outdoor driving images, as will be demonstrated by analysis in the later sections. In this paper, we present a robust non-local algorithm that can meet this challenge. Specifically, we perform cost aggregation on an alternate, effective segment-simple-tree (SST) that is more suitable for outdoor driving images than MST and ST.

The remainder of this paper is structured as follows. Section 2 presents the non-local cost aggregation on a MST, which is the inspiration for our algorithm. Section 3 presents the proposed algorithm. Section 4 presents the cost computation, disparity optimization and post-processing. Section 5 presents the experimental results that compare our algorithm to those proposed in prior literature. Section 6 provides the conclusions for this paper.

2. Non-local cost aggregation on MST

In this section, we present the MST-based non-local algorithm [19,20], which directly inspired our algorithm. Specifically, we summarize the construction of the MST and the two-pass cost aggregation. The proposed algorithm performs a two-pass cost aggregation similar to that of the MST. However, the tree is constructed differently by employing an alternate, effective segment-simple-tree structure (SST) that is more suitable than MST for outdoor driving images.

To begin, we denote $G=(V,E)$ as a connected, undirected graph that represents the guidance image I. Each vertex $v \in V$ corresponds to a pixel in I. Each edge $e \in E$ connects two neighboring pixels p and q. The weight w_e associated with this edge is computed using the absolute difference of the intensity between the two pixels. Formally, the edge weight is expressed as follows:

$$w_e = w(p,q) = |I(p)-I(q)| \tag{1}$$

A set of edges $E_{MST} \subset E$ connecting all vertices, whose sum of the weights is minimal compared to the sums of the weights of all other sets of edges, is selected from G to construct a minimum spanning tree $T=(V, E_{MST})$. Consequently, the distance $D(s,r)$ between two arbitrary nodes s and r in the tree can be computed as the sum of the weights along the path connecting them.

The similarity $W(s,r)$ between two nodes can then be computed as a Gaussian function as follows:

$$W(s,r) = \exp\left(-\frac{D(s,r)}{\sigma}\right) \tag{2}$$

where σ is the user input parameter for adjusting the similarity. Accordingly, the aggregated cost $C_d^A(s)$ for pixel s at disparity level d is computed as a weighted sum of the raw matching cost C_d of all other pixels. It is formally computed by

$$C_d^A(s) = \sum_{r \in I} W(s,r) C_d(r) \tag{3}$$

Here, we note that this strategy is similar to the adaptive support-weight approach [7], except that r covers all pixels in the image rather than only the pixels inside a support window. This modification improves the estimation accuracy when tested with indoor images [19,20].

Another advantage of this method is that it can be implemented using a linear time algorithm by traversing the tree T in two passes. First, the raw matching cost C is aggregated from the leaf nodes to the root node to produce the intermediate aggregated cost $C^{A\uparrow}$. Each node s receives supports from its subtrees, while the root node receives supports from all of the other nodes. This process is formally expressed by

$$C_d^{A\uparrow}(s) = C_d(s) + \sum_{r \in Ch(s)} W(s,r) C_d^{A\uparrow}(r) \tag{4}$$

where $Ch(s)$ is the set that contains the children of node s. Second, the reverse pass traverses from the root node to the leaf nodes, so that each node s receives supports from nodes other than its subtrees. This is to ensure that every node receives supports from all other nodes in a manner similar to that for the root node. This process is computed as follows:

$$C_d^A(s) = W(Pr(s),s) \cdot C_d^A(Pr(s)) + (1-W^2(Pr(s),s)) \cdot C_d^{A\uparrow}(s) \tag{5}$$

where $Pr(s)$ denotes the parent of node s. An intuitive representation of a MST generated from an input image and its two-pass cost aggregation is shown in Fig. 1. The computational complexity for each pixel is $O(1)$. Although this algorithm is able to produce good results for indoor images [19,20], its performance for outdoor images is limited, as will be presented in Section 5.

3. Proposed algorithm

3.1. Analysis of outdoor driving images

In this subsection, we first analyze the disparity characteristics of outdoor driving images to understand the key idea behind the SST proposed for non-local stereo matching in DAS. Fig. 2 illustrates this analysis using a synthesized stereo image pair obtained from the EISATS dataset [4]. Fig. 2(a) and (b) shows the left image and the corresponding