Optik 137 (2017) 124–134

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Original research article

Segmentation-based stereo matching using combinatorial similarity measurement and adaptive support region

Ning Ma [a,b,*], Yubo Men [b], Chaoguang Men [b], Xiang Li [b]

[a] College of Computer Science and Information Engineering, Harbin Normal University, Harbin 150025, China
[b] College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China

ARTICLE INFO

Article history:
Received 24 September 2014
Received in revised form 19 February 2017
Accepted 4 March 2017

Keywords:
Stereo matching
Combinatorial similarity measurement
Belief propagation
Disparity

ABSTRACT

An accurate segmentation-based algorithm using combinatorial similarity measurement and adaptive support aggregation strategy is developed for stereo matching. The advantage of our method which based on a segmentation framework is generating disparity map in textureless regions correctly and localize depth boundaries precisely. Initial disparity map is estimated by a local correspondence approach which consisting of a combinatorial similarity measurement function and an adaptive window with arbitrary shape and size. The accuracy of the initial disparity map generated by combinatorial similarity measurement is better than that generated by individual similarity measurement. The plane parameters are fitted utilizing the initial disparity map. In this process, a method named "MC-RANSAC" which consisting of mutual consistency check criterion and random sample consensus algorithm is proposed to filter out the outliers and obtain the stable disparity plane parameters. Lastly, the disparity plane is optimized by belief propagation. Experiments with stereo image pairs show the validity of the proposed method.

1. Introduction

Computational stereo for extracting three-dimensional scene structure is one of the most widely studied and fundamental topics in computer vision. The goal of stereo matching is to determine the disparity map between two or more digital images taken from distinct viewpoints for the same scene [1]. Generally, stereo matching is used to extract the 3D information in terms of a disparity map from a pair of stereo images. It serves as an important step in many applications such as robot navigation, industrial control, medical diagnosis, teleconferencing, 3D environment reconstruction, image and video based rendering, and stereo video coding [2]. Many existing dense two-frame stereo correspondence algorithms are evaluated by a taxonomy and categorization scheme on the Middlebury stereo evaluation website [3]. A complete stereo matching algorithm performs the following four steps: matching cost initialization, cost (support) aggregation, disparity calculation, and disparity refinement. The estimation of accurate disparity map is still challenging work due to the ill-posed nature of the stereo matching problem, such as textureless regions, occlusion, perspective distortion, repetitive patterns, reflections, shadows, illumination variations and poor image quality, sensory noise and high computing load.

The existing stereo algorithms can be classified into 2 categories: local approaches and global approaches. Local approaches work in the way that they aggregate matching costs of neighboring points within a finite support window and

* Corresponding author at: College of Computer Science and Information Engineering, Harbin Normal University, Harbin, 150025, China.
E-mail address: maninghsd@163.com (N. Ma).

http://dx.doi.org/10.1016/j.ijleo.2017.03.018

utilize a local winner-takes-all (WTA) strategy for disparity estimation at each pixel [4]. It implicitly models the assumption that the disparity of all the pixels in the support window is consistent, and the scene of stereo vision is piecewise smooth. Usually, this category approaches are simple and effective, which could easily capture accurate disparity in high textured regions. The aggregation strategy of support window and the similarity measurement function are two major research topics for local approaches. For aggregation aspect, adaptive support window [5,6] and adaptive support weight [7] are the two mainstream strategy. Adaptive support window based cost aggregation methods select an appropriate window with arbitrary size and shape for each pixel by evaluating the local variation of intensity or color. Adaptive support weight based cost aggregation methods compute the support weights for all pixels within the window using color similarity and geometric proximity. For similarity measurement function, the sum of absolute difference (SAD), sum of squared difference (SSD), normalized cross correlation (NCC) are commonly used similarity measurement functions. Moreover, the census nonparametric transform and mutual information are the advanced similarity measurement functions [8]. The global approaches make explicit smoothness assumptions of the disparity space image and carry out optimized stereo matching by minimizing the global energy function. Due to the excellent experimental results, the global optimization have attracted much attention. The dynamic programming (DP), belief propagation (BP), graph cuts (GC) and semi-global are the representative algorithms. DP approach performs a global optimum of the cost function in one-dimensional disparity space, this ensures that DP is a very efficient algorithm [9]. However, the well-known horizontal "streaks" artifacts affects the accuracy of the DP algorithm due to the smoothness consistency inter-scanlines cannot be well enforced. BP and GC approaches obtain disparity map by optimizing the two-dimensional energy formulated on a data term and a smoothness term [10,11]. Although a number of excellent results have been obtained, both BP and GC approaches are typically expensive of computation and storage. Another disadvantage of these approaches is that there are so many parameters need to determine sometimes. The semi-global method employs the mutual information cost function and executes semi-global optimization [12].

In the existing stereo matching, the segmentation-based approaches have attracted attention due to their excellent performance [13–18]. Generally speaking, segmentation-based approaches performs four steps: first, the reference image is decomposed into non-overlapping homogeneous color segments; second, the initial disparity is estimated by local match method; third, the disparity plane is fitting by a plane fitting technique; finally, the disparity plane is optimized. In general, this category approaches perform well in textureless regions and depth discontinuity regions. However, there are two main factors that will affect the capability of segmentation-based algorithm. The first factor is the accuracy of initial disparity, and the second factor is how to effectively filter out outliers. In this work, we propose a segmentation-based stereo matching algorithm using combinatorial similarity measurement function and adaptive support aggregation strategy, which can reduce the effects of two factors above mentioned. Firstly, we segment the reference image utilizing the mean-shift algorithm. Secondly, in order to get accurate initial disparity map, we employ a local approach which consisting of a combinatorial similarity measurement function and an adaptive window with arbitrary shape and size. Thirdly, we assign a disparity plane model to each segment by MC-RANSAC fitting method using the initial disparity map. Finally, an energy function is formulated in segmented regions, and the optimal disparity plane labeling is approximated by BP.

This paper is organized as follow: In Section 2, we present the novel stereo matching approach and the overall algorithm is described in details. In Section 3, experimental results and analysis are given using an extensive evaluation data set which includes Middlebury standard data, synthesized images and real-world stereo pair. The paper is concluded in Section 4.

2. Algorithm description

The proposed stereo matching approach is shown in Fig. 1. The input data of our algorithm are two rectified stereo images in epipolar geometry. The entire algorithm involves four steps, in the first step we segment the reference image into a number of non-overlapping regions utilizing the mean-shift algorithm [19]. Second step, the Adaptive Combinatorial Similarity Measurement algorithm is designed to estimate the initial disparity map. Third step, we filter out outliers and assign a disparity plane model to each segment by MC-RANSAC algorithm. Last step, we formulate new energy function in segmented regions and use belief propagation to refine the disparity map. Output of our algorithm is the final disparity map. The main steps are explained in the following paragraphs.

2.1. Color based segmentation

The mean-shift color segmentation algorithm is employed to divide the reference image into regions of homogeneous color. In our paper, it is assumed that the disparity in each color segment is continuous, thus, the over-segmentation is adopted since it helps to meet this assumption in practice. The mean-shift approach is essentially defined as a gradient ascent search for maximum kernel density estimation in a high dimensional feature space. The main advantage of the mean-shift approach is based on the fact that edge information is incorporated as well which ensures our method to estimate disparity in textureless regions and depth discontinuities precisely. The output results of color segmentation on 4 stereo images used for testing are shown in Fig. 2 and pixels belong to the same segment are assigned the same color randomly.