Optik 125 (2014) 115–120

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

An improved gradient-based dense stereo correspondence algorithm using guided filter

Huiyan Han, Xie Han*, Fengbao Yang

School of Information and Communication Engineer, North University of China, Taiyuan 030051, Shanxi, China

ARTICLE INFO

Article history:
Received 29 January 2013
Accepted 3 June 2013

Keywords:
Stereo correspondence
Guided filter
Gradient similarity measurement
Disparity

ABSTRACT

This paper presents an improved gradient-based real-time stereo correspondence algorithm using guided filter. Color intensity value is sensitive to radiometric distortions including exposure differences and illumination differences, thus the error correspondence rates of these methods are high. deMaeztu et al. [1] proposed to measure the similarity between pixels using the gradient value instead of color intensity. The method has better robustness to radiometric distortions than intensity-based local methods, but the running time is so long that it is not suitable for real-time applications, because the adaptive support weight of neighbor pixels depends on bilateral filter. Guided filter has edge-preserving character as bilateral filter, but runs faster than it, we use guided filter as adaptive support weight instead of bilateral filter of the neighbor pixels in a finite squared support window. The experimental results demonstrate that the improved algorithm performs much better compared with gradient-based method and other local methods, whether in accuracy or efficiency, according to the widely-used Middlebury stereo benchmarks, and the robustness to radiometric distortions of ours is better also.

1. Introduction

Dense stereo correspondence is one of the key and difficult problems in computer vision. In recent years, much research has been carried out for this problem, and great improvement has been achieved. According to [2], stereo correspondence algorithm is classified as local and global algorithms. Global methods can obtain high-accuracy disparity map, but it is difficult for users to determine the parameters with high complexity and it is not appropriate for real-time applications. Local approaches can achieve disparity map quickly, but the precision is low especially in depth discontinuity regions, and many researchers have focused on this problem.

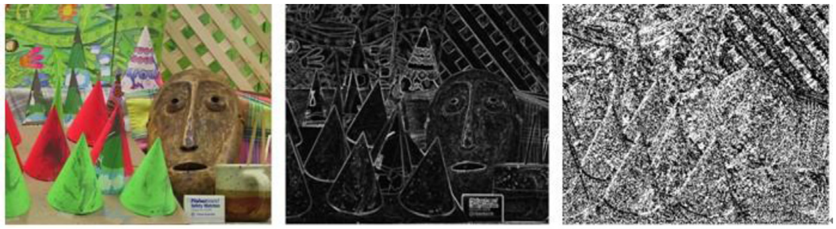
The similarity measurement used in most local stereo correspondence algorithms at present is based on pixel color intensity, and the corresponding pixels in the two views should have equal intensity. But the corresponding pixels in different views may not have the same intensity value due to radiometric distortions, image noise, and repetitive (or weakly) texture, which make the correspondence result extraordinarily sensitive to intensity changes. To overcome this problem, some local dense stereo correspondence methods often aggregate the correspondence cost of the pixels in the support region around every pixel, with the implication that all pixels in the region have the same disparity as the central pixel. By using the support window, the image ambiguity is reduced to a certain extent, however, if the support window locates on depth discontinuity region, the disparities of the pixels do not equal, which conflicts with the implication of the aggregating and results in the so-called "foreground fattening" phenomenon. To achieve correct correspondence result in both regions, the support window should vary adaptively for every pixel, and many approaches have been put forward. Adaptive support window methods [3,4] and multiple support window methods [5,6] improve the correspondence result to a certain extent, but the shape and size of the support window are restricted, and can not make each pixel truly adaptive in the entire image regions especially in depth discontinuity regions. To work this problem out, some segmentation based methods [7] segment the image in the preprocessing stage, but accurate color segmentation is difficult and time-consuming for rich texture images.

Another improved approach is adaptive support-weight [8] (the bilateral filter in fact), the size of the square support window is constant, and each pixel in the window has a different adaptive support weight, which depends on color similarity and proximity degree with the central pixel, and follow the principles of Gestalt visual. The correspondence result is satisfactory for a local approach and can be comparable to global ones. However, the running time is too long, and the method cannot be used in real-time applications. Some accelerated versions of the method have been proposed [9,10], but the improvement effects are not obvious.

The pixel-based correspondence cost in above-cited ones depend on the color intensity or gray level values only, which make the correspondence result error-prone. Four reasons have

* Corresponding author. Tel.: +86 13703516841.
E-mail address: hhy980344@163.com (H. Han).

0030-4026/$ – see front matter. Crown Copyright © 2013 Published by Elsevier GmbH. All rights reserved.
http://dx.doi.org/10.1016/j.ijleo.2013.06.038



(a)Left view of "Cones" (b)Gradient modulus map (c)Gradient phase map

Fig. 1. Gradient modulus and gradient phase map of a color image.

been introduced in [1] and proposed to measure the pixel similarity using the truncated pixel gradient ("Gradient" method for abbreviatory), the accuracy of the result disparity map is improved, and the method has better robustness to radiometric distortions. In 2011, guided filter aggregation approach [11,12] ("Costfilter" for abbreviatory) was proposed, the method is not only the best performing local stereo correspondence method, but also the best performing real-time method.

In this paper, on the basis of analysis of "Costfilter" method and "Gradient" method, we propose an improved stereo correspondence algorithm. The pixel-based correspondence measurement depends on the gradient values, jointly with the aggregation step of adaptive support weight using guided filter, disparity is calculated based on WTA principle, left-right cross consistency check and post processing. Compared with "Gradient" method, the accuracy of the resulting disparity maps is improved, the time complexity is reduced, while robustness to radiometric distortion keeps constant. The improved algorithm is described in Section 2. In Section 3, the stereo correspondence performance of the new proposed algorithm is compared with other local methods, the robustness to radiometric distortion is compared with three well known similarity measurements, "Gradient" method, "Costfilter" method and color intensity-based method ("Intensity" for abbreviatory). We conclude our paper in Section 4.

2. The improved stereo correspondence algorithm

We use gradient values as pixel-based correspondence measurement, aggregation using the guided filter, compute the disparity value by selecting the minimal aggregated correspondence value for each pixel using WTA (Winner-Take-All) manner, after left-right cross consistency check and post processing, obtain the ultimate disparity map.

2.1. The gradient-based similarity measurement

The gradient of an image corresponds to the direction along which the gray value of the image changes most remarkably. As edge detection is seeking the local maximum and the direction of image gradient, gradient value can be used to reflect image edge or skeleton to some extent. This can be observed from Fig. 1, the gradient modulus map and gradient phase map of image "cones" are shown. Gradient has better robustness against image noise, differences in sampling and local brightness changes between image views [1] (three important sources of error in methods relying on color intensity similarity of corresponding pixels).

Gradient value is composed of gradient modulus m and gradient phase p. Suppose $f(x,y)$ represents a gray image, the gradient of $f(x,y)$ is defined as a vector $\nabla f = \begin{bmatrix} Gx \\ Gy \end{bmatrix}$, Gx and Gy is the differences in x (horizontal) and y (vertical) direction. The modulus m of ∇f is $m = \sqrt{Gx^2 + Gy^2}$ (the form [1] used). For simplicity, we use an approximate form $m = |Gx| + |Gy|$ instead to calculate m. The phase of ∇f is $\phi = \mathrm{arctg}(Gy/Gx)$.

There are some commonly-used similarity cost functions, such as summed absolute differences (SAD), summed squared differences (SSD), normalized cross correlation (NCC). It has been shown that SAD is the fastest one in computation and has some advantages over NCC and SSD [13], so we use SAD of the gradient values as the similarity measurement.

The rate of change (modulus m) and the direction of the greatest rate of change (phase ϕ) provide different information about neighborhood of a pixel and have different invariance properties with respect to radiometric distortion. Neither the modulus nor the phase is affected by additive (offset) changes in the input images. Multiplitive variations (gain) affect the modulus but not the phase. So phase and modulus are separating and the weights of both are not equal, as expressed in Eq. (1):

$$G(p,d) = \sum_{c \in \{r,g,b\}} (\alpha |m_l^c(p) - m_r^c(p-d)| + f(|\phi_l^c(p) - \phi_r^c(p-d)|)) \quad (1)$$

where $G(p,d)$ is the correspondence cost of pixel p when disparity is d, m^c and ϕ^c are the modulus and phase of the gradient operator applied to the color band $c(r,g,b)$ respectively, α is the weight of modulus and controls the sensitivity of the algorithm to radiometric differences between images. Through our practical experiments, the best value for α is 0.12 (the same to [1]), and f is a function to limit the range of difference of phases of the two corresponding pixels to $[0,\pi]$ (as in Eq. (2)). To reduce influences of outliers, we use truncation value of gradient value, as expressed in Eq. (3):

$$f(x) = \begin{cases} x, & \text{if}(0 \le x \le \pi) \\ 2\pi - x, & \text{if}(\pi < x < 2\pi) \end{cases} \quad (2)$$

$$C(p,d) = \min\{G(p,d), Tg\} \quad (3)$$

2.2. Aggregation using guided filter

The bilateral filter is time-consuming and loses the real-time advantage of local over global approaches. The bilateral filter weight is expressed as:

$$W_{p,q}^{bf} = \frac{1}{K_p} \exp\left(-\frac{|p-q|^2}{\sigma_s^2}\right) \exp\left(-\frac{|I_p - I_q|^2}{\sigma_c^2}\right) \quad (4)$$

where K_p is the normalization factor and σ_c, σ_s are constant parameters which adjust the spatial and color similarity respectively [10] proposed an approximate and fast method, but space complexity is huge and the accuracy of the correspondence result is lower than the original method [11,12] proposed to use guided filter instead of bilateral filter, this is mainly due to the following reasons [14]: