Optik 126 (2015) 1052–1056

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Stereo matching based on guided filter and segmentation

Yuan Zhou*, Chunping Hou

School of Electronic Information Engineering, Tianjin University, Tianjin 300072, China

ARTICLE INFO

Article history:
Received 23 January 2014
Accepted 27 January 2015

Keywords:
Stereo matching
Guided filter
Markov random field
Mean shift

ABSTRACT

For improving the accuracy of stereo matching and maintaining discontinuity of object edge and continuity of non-edge area in the matching result, a stereo matching method based on guided filter and mean shift is proposed in this paper. Matching cost function based on Markov random field (MRF) and guided filter are established, and the initial disparity value of each pixel is calculated by minimizing the corresponding matching cost function. Mean shift algorithm is used to segment the stereo images into independent areas, and improve the final disparity map. Results of the proposed method are tested on the international standard Middlebury stereo benchmark and compared with other methods. Comparative results show high accuracy of the proposed algorithm and its superiority to some prevailing algorithms.

1. Introduction

From the earliest inquiries into visual perception, it was known that we perceive depth based on the differences in appearance between the left and right eye. Under simple imaging configurations, namely two eyes or cameras looking straight ahead, the amount of horizontal motion or disparity is inversely proportional to the distance from the observer. While the basic physics and geometry relating visual disparity to scene structure, such as epipolar geometry, rectification and plane sweep, are well understood, automatically measuring this disparity by establishing dense and accurate inter-image correspondences is a challenging task. Stereo matching is the process of taking two or more images and estimating a 3D model of the scene by finding matching pixels in the images and converting their 2D positions into 3D depths. The process result is commonly a sparse or dense depth map that assigns relative depths to pixels in the input images. In image processing and computer vision, the topic of stereo matching has been one of the most widely studied and fundamental problems, and continues to be one the most active research areas.

Early stereo matching image locations, using stereo matching algorithms, were feature-based, i.e., they first extracted a set of potentially matchable image locations, using either interest operators or edge detectors, and then searched for corresponding locations in other images using a patch-based metric. The limitation to sparse correspondences was partially due to computational resource limitations, but was also driven by a desire to limit the answers produced by stereo algorithms to match with high certainty. In some applications, there was also a desire to match scenes with potentially very different illuminations, where edges might be the only stable features. While sparse matching algorithms are still occasionally used, most stereo matching algorithms today focus on dense correspondence, since this is required for applications such as image-based rendering or modeling. Dense correspondence algorithms generally perform the following four steps:

1. matching cost computation;
2. cost aggregation;
3. disparity computation;
4. disparity refinement.

Scharstein and Szeliski [10] have done some research on a taxonomy of dense correspondence algorithms. They have separated these algorithms into two broad classes, namely local methods and global methods.

Local or window-based algorithms, where the disparity computation at a given point depends only on intensity values within a finite window, usually make implicit smoothness assumptions by aggregating support.

Global stereo matching methods perform some optimization or iteration steps after the disparity computation phase and often skip the aggregation step altogether, because the global smoothness constraints perform a similar function. Many global methods are formulated in an energy-minimization framework. Once the global energy has been defined, variety of algorithms can be used to find a minimum. Traditional approaches are associated with regularization and Markov random fields. Prevailing global algorithms include dynamic programming [1,12,13], belief propagation

* Corresponding author. Tel.: +86 13820375613.
E-mail address: zhouyuan@tju.edu.cn (Y. Zhou).

http://dx.doi.org/10.1016/j.ijleo.2015.01.030
0030-4026/

[5,6,11,14] and graph cuts [2]. Dynamic programming algorithm is used in [1] to improve the disparity result in occluded areas. Improvements are made by bringing in a connected, undirected graph and constraints are given by the relevance between pixels in the graph in [12,13]. The belief propagation algorithm which works by passing messages around the image is used in [5,6,11], and optimization is achieved by minimizing the matching cost function.

Global algorithms generally have a better robustness than the local algorithms and can figure out more reliable disparity maps. That is because when computing the disparity of each pixel, global algorithms comprehensively take disparities of other pixels into consideration, which leads to a stronger correlation between each pixel, but local algorithms are just different. However, the smoothness term existed in global optimizations makes disparity smooth everywhere and may lead to poor results at object boundaries. The quality of the edges and occluded points enormously influences quality of the disparity result. Till now, few algorithms specific to edges or occluded points are proposed. In this paper, guided filter will be brought in to solve this problem.

When it comes to the disparity refinement, the method in this paper first segments the images into regions and then tries to label each region with a disparity, instead of performing computations on a per-pixel basis used in many other methods [8]. And mean shift algorithm is used to do the segmentation. Refinement and smoothing are done within the areas which come up with mean shift. This algorithm can achieve a better weigh between the continuity of non-edge area and the discontinuity of object edge.

Broadly speaking, in this paper, a global dense correspondence algorithm based on guided filter and Markov random fields (MRF) is proposed. Because of the high distinctiveness of the guided filter, edges in the disparity maps are kept clearly. After that, in order to keep the smoothness of the disparity map, the image segmentation method called mean shift is used to find the sections with similar color, and smoothing is done within each result section.

2. Stereo matching

In this section, the guided filter is introduced first. Then, the matching cost function is computed based on MRF and weighted matching cost is achieved after guided filtering. After that, winner-take-all algorithm is used to get the minimum cost function for each pixel.

2.1. Guided filter

Guided filter was first proposed by He [7]. Unlike bilateral filter [3], its runtime is linear in the number of image pixels, and it has the edge-preserving smoothing property. For a grayscale image, the kernel weight of a pixel can be defined as:

$$W_{i,j} = \frac{1}{|w|^2} \sum_{k:(i,j)\in w_k} \left(1 + \frac{(I_i-\mu_k)(I_j-\mu_k)}{\sigma_k^2+\varepsilon} \right). \tag{1}$$

Here, i and j are pixel indexes, μ_k and σ_k^2 are the mean and variance of I in w_k, $|\omega|$ is the number of pixels in w_k, ε is a regularization parameter.

The terms $(I_i-\mu_k)$ and $(I_j-\mu_k)$ have the same sign (+/−) when I_i and I_j are on the same side of an edge, while they have opposite signs when the two pixels are on different sides. So in (1) the term $1+\frac{(I_i-\mu_k)(I_j-\mu_k)}{\sigma_k^2+\varepsilon}$ is much smaller (and close to zero) for two pixels on different sides than on the same side. This means that the pixels across an edge are almost not averaged together. Just as the ideal step edge of a 1-D signal in Fig. 1 shows, for a window that exactly center on the edge, the variables and are as indicated.

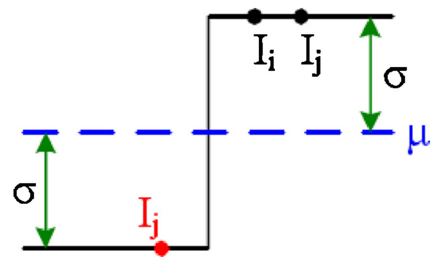

Fig. 1. An ideal step edge.

Similarly, when it comes to color pictures, we can define the weight as:

$$W_{i,j} = \frac{1}{|w|^2} \sum_{k:(i,j)\in w_k} (1+(I_i-\mu_k)^T(\sum k+\varepsilon U)^{-1}(I_j-\mu_k)). \tag{2}$$

Here, I_i, I_j and μ_k are no longer luminance, but 3 × 1 color vectors of RGB. $\sum k$ is the 3 × 3 covariance matrix of I in w_k, and U is a 3 × 3 identity matrix. According to the principle, the guided filter is an edge-preserving smoothing filter, so less blurring occurs at the edges. As a result, a more clear disparity map can be achieved after bringing in guided filter.

2.2. Matching cost based on MRF

Markov random field (MRF) [9] theory provides a basis for modeling contextual constraints in visual processing and interpretation. It enables us to develop optimal vision algorithms systematically when used with optimization principles. How to give a suitable characteristic value to corresponding pixels is what MRF theory tries to solve. Traditional MRF theory uses the maximum posteriori probability to obtain the accurate characteristic value for each pixel.

In stereo matching problems, figuring out the disparity for each pixel is the main purpose. Given two images: left image I_l and right image I_r. I_l is chosen as the reference image. According to the MRF theory, the disparity set is defined as $L_d=\{l_1, l_2, l_3, \ldots, l_M, \}$. Stereo matching is just the process of figuring out the correct disparity value $l\in L_d$ for the pixel with coordinate (x, y). As a result, our matching cost function consists of three variables, namely coordinate (x, y) for pixel i and the corresponding disparity value l. The absolute luminance difference is defined as:

$$\overline{C_{SAD(i,\ell)}} = \frac{1}{P\times Q} \sum_{j\in N_{(i)}} |I_{r(j+\ell)} - I_{l(j)}|. \tag{3}$$

And the absolute gradient difference is:

$$\overline{C_{GRAD(i,\ell)}} = \frac{1}{P\times Q} \sum_{j\in N_{(i)}} |\nabla_x I_{r(j+\ell)} - \nabla_x I_{l(j)}|. \tag{4}$$

Here, I is the luminance of the pixel, $N_{(i)}$ is the set of pixels in a size-fixed window centered at the pixel (x, y), $P\times Q$ is the size of the window, $|\cdot|$ means absolute value operation, j is the index of the pixel in set $N_{(i)}$, ∇_x means the gradient operation of horizontal direction. Because the disparity only exists in the horizontal direction, only horizontal gradient is computed here. As a result, our matching cost function is:

$$CF_{(i,l)} = (1-\lambda)\cdot \min[\overline{C_{SAD(i,\ell)}}, \tau_1] + \lambda \cdot \min[\overline{C_{GRAD(i,\ell)}}, \tau_2]. \tag{5}$$

Here, $\lambda\in(0, 1)$ is a harmonic coefficient, used for balancing luminance difference and gradient difference. When $\overline{C_{SAD(i,l)}}$ (or $\overline{C_{GRAD(i,l)}}$) is larger than a threshold value τ_1 (or τ_2), τ_1 (or τ_2) is used to replace $\overline{C_{SAD(i,l)}}$ (or $\overline{C_{GRAD(i,l)}}$), in order to reduce the