Pattern Recognition Letters 31 (2010) 2389–2396

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Stereo depth estimation using synchronous optimization with segment based regularization

Tarkan Aydin [a,b,*], Yusuf Sinan Akgul [a]

[a] GIT Vision Lab, Department of Computer Engineering, Gebze Institute of Technology, Gebze, Kocaeli 41400, Turkey
[b] Bahcesehir University, Istanbul 34353, Turkey

ARTICLE INFO

Article history:
Received 28 January 2010
Available online 27 July 2010
Communicated by A. Fernandez-Caballero

Keywords:
Stereo
Optimization
Segment based regularization
Anisotropic smoothing

ABSTRACT

Stereo correspondence is inherently an ill-posed problem, which is addressed by regularization methods. This paper introduces a novel stereo correspondence method that uses two synchronous interdependent optimizations. The regularization of the correspondence problem is done adaptively by considering the image segments and the intermediate disparity maps of the two optimizations. Our adaptive regularization allows inter-segment diffusion at the beginning of the optimizations to be robust against local minima. When the two optimizations start producing similar disparity maps, our regularization prevents inter-segment diffusion to recover the depth discontinuities. Our experimental results showed that the proposed algorithm can handle sharp discontinuities well and provides disparity maps with accuracy comparable to the state of the art stereo methods.

1. Introduction

Stereo correspondence is one of the fundamental problems of computer vision. The typical result of the correspondence problem is expressed as a disparity map, i.e. spatial shifts in the pixel positions of the corresponding points. The main difficulties of the correspondence problem are the ambiguity due to the image noise, repeated texture, and occlusions. These problems make the stereo correspondence an ill-posed problem, which is classically addressed by a regularization method to stabilize the solution. The main role of the regularization is to incorporate a priori information to handle image noise and to fill-in missing and ambiguous data.

Classically, the regularization is employed by local and global methods. The local methods perform regularization directly on the data space by employing some aggregation scheme (Intille and Bobick, 1994; Kanade and Okutomi, 1994; Scharstein and Szeliski, 1998; Yoon and Kweon, 2006). The global methods, on the other hand, formulate the problem as an energy functional that needs to be minimized to produce the desired solution. The regularization is performed on the disparity space by introducing an explicit smoothness criteria so that reliable disparity values are propagated to ambiguous image regions. Dynamic programming was tried by enforcing smoothness only along the epipolar lines in order to obtain a globally optimal solution to the discrete form of the energy functional (Ohta and Kanade, 1985; Gong and Yang, 2005). However, the resulting disparity maps contain well-known streaking effects due to the inconsistency between epipolar lines. Alternatively, a global minimum of the functional can also be obtained in polynomial time via graph cuts (Roy and Cox, 1998; Ishikawa, 2003) by using a convex smoothness term. However, these methods oversmooth the depth discontinuities. A discontinuity preserving regularizer might produce a good solution but it is known that introducing a discontinuity preserving smoothness term makes the problem NP-complete (Kolmogorov and Zabih, 2004). Therefore, using an approximate optimization method for the functional with non-convex smoothness terms became more popular, such as graph cuts (Boykov et al., 2001), belief propagation (Sun et al., 2003), and genetic algorithm (Saito and Mori, 1995). However, these methods only produce integer valued disparity maps due to their discrete nature. This restriction is a severe drawback if curved or slanted surfaces are present in the scene (Li and Zucker, 2006).

Another class of global approaches, as a counterpart, use related partial differential equations (PDE) and variational methods in order to find the minimizer of the continuous form of the energy functional. These methods can achieve a continuous solution by iteratively evaluating the associated Euler–Lagrange equation. An inherent advantage of these methods is the capability of making sub-pixel disparity estimations due to the continuous solution they provide.

* Corresponding author at: GIT Vision Lab, Department of Computer Engineering, Gebze Institute of Technology, Gebze, Kocaeli 41400, Turkey. Tel.: +90 262 6052243; fax: +90 262 6052205.

E-mail addresses: tarkan.aydin@bahcesehir.edu.tr (T. Aydin), akgul@bilmuh.gyte.edu.tr (Y.S. Akgul).

URL: http://vision.gyte.edu.tr (T. Aydin).

0167-8655/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2010.07.012

The minimization process of these continuous methods is characterized by the choice of the regularizer. Using a disparity driven isotropic regularizer with a quadratic term makes the minimization robust against local minima (Robert et al., 1992). However, the depth discontinuities in the resulting disparity maps would be oversmoothed. Although it is possible to use a non-quadratic smoothing term, such as total variation regularizer, to inhibit the oversmoothing of discontinuities (Slesareva et al., 2005), it cannot handle the discontinuities adequately (Ben-Ari and Sochen, 2007).

There are several other regularization methods for the handling of the discontinuities. Shah (1993) uses nonlinear diffusion to extract stereo matches and occluded regions simultaneously in conjunction with a gradient descent minimization. Similarly, Robert and Deriche (1996) use anisotropic disparity driven regularization in order to prevent smoothing of the disparity map at the estimated discontinuities. Therefore, it tends preserve the discontinuities present at the initialization. Alternatively, image driven regularizers were used to align depth discontinuities along edges and inhibit smoothing across edges (Alvarez et al., 2002; Kim et al., 2004). Min et al. (2006) employed the image segments to perform anisotropic smoothing at the segment boundaries depending on the magnitude of image gradients. The problem with image driven regularizers is that they have to work with over-segmented images when the images are highly textured. In addition, the boundary leakage problem becomes an issue when there are gaps at the object boundaries.

Nevertheless, these discontinuity preserving approaches require sufficiently reliable initialization in order to converge to the desired solution. In most cases, the initialization errors cannot be recovered, especially for noisy and occluded regions.

In this paper, we introduce a novel initialization insensitive regularization method that preserves the depth discontinuities. Our framework employs two separate but dependent energy functionals (Akgul and Kambhamettu, 1999; Aydin and Akgul, 2006) which are intended to be minimized synchronously until converging to the same solution. Because of the interaction between the optimizations, the overall result of our system is always better than the results achievable by a single optimization. Reliable convergence is ensured by starting each optimization with different initial conditions.

In order to handle depth discontinuities robustly, we employ image segments to align the depth discontinuities with the segment boundaries. Unlike the previous image based smoothing techniques, the proposed method adjusts the smoothing by utilizing not only the segment information but also the positional differences between the synchronous optimizations. These two means of adjusting the smoothing make it possible to use isotropic and anisotropic smoothing adaptively. As a result, we produce more robust depth discontinuity positions. Note that our employment of synchronous optimizations is very different from that of Aydin and Akgul (2006), which does not use the optimizations for the regularization and completely ignores the depth discontinuities.

Selecting an appropriate stopping criteria is crucial for many diffusion techniques in order to avoid oversmoothing and insufficient regularization (Scharstein and Szeliski, 1998). Since our discontinuity preserving regularization method relies on the positional difference between the solutions of each optimization, the diffusion between the segments are prevented when both optimizations find the same disparity map, hence smoothing of the discontinuities is inhibited even at superfluous iterations. This inherent stopping criteria of our framework is an important advantage over the similar systems against problems like sensitivity to extra iterations.

The rest of this paper is organized as follows. Section 2 reviews the synchronous energy functional. Section 3 describes the proposed regularization that preserves the depth discontinuities. Section 4 describes the system validation and experiments. Finally, we provide concluding remarks in Section 5.

2. Overview of the approach

2.1. Energy-based global stereo formulation

Traditional global stereo energy formulation is written as the sum of the data term and a regularization term. Consequently, the stereo correspondence problem is formulated as the minimization of the following energy functional,

$$E(D) = \int \alpha\phi(D) + \beta\psi(|\nabla D|)dp, \quad (1)$$

where D is the disparity map which assigns disparity values to each pixel p in the reference image. α and β are weighting coefficients for adjusting the relative weights of each term.

The *data term* ϕ computes the image similarity measure by means of commonly used similarity metrics, such as the sum of squared differences (SSD), the sum of absolute differences (SAD), and the normalized cross correlation (NCC). The *smoothness* or *regularization* term ψ is introduced to impose a priori information (smoothness) on the desired disparity map by penalizing disparity gradients (∇D).

2.2. Synchronous energy formulation

Based on the classical stereo energy functional, the synchronous optimizations are formulated as the minimization of two energy functionals by introducing a new tension term φ as in the following equations.

$$E(D_1) = \int \alpha\phi(D_1) + \beta\psi(|\nabla D_1|^2) + \lambda\varphi((D_1 - D_2)^2)dp, \quad (2)$$

$$E(D_2) = \int \alpha\phi(D_2) + \beta\psi(|\nabla D_2|^2) + \lambda\varphi((D_2 - D_1)^2)dp, \quad (3)$$

where D_1 and D_2 are the disparity maps obtained from each optimization.

The tension term φ is for the interaction between the two minimizations and it is the core idea of the synchronous optimization method. The main function of this term is to lower the difference between the two disparity maps D_1 and D_2. Note that without the tension term, minimization of the energy functionals defined by Eqs. (2) and (3) by starting from different initial configurations would produce a different disparity map for each equation. However, if the equations are optimized in synchronization with the help of the tension term, they would end up finding the same disparity map.

The disparity maps are computed by searching the minimizers of the energy functionals defined in Eqs. (2) and (3). Minimization of the functionals via the gradient descent method by introducing an artificial evolution parameter t yields the equations,

$$\frac{\partial D_1}{\partial t} = \gamma(\alpha\phi'(D_1) + \beta\nabla\cdot(\psi'\nabla D_1) + \lambda\varphi'(D_1 - D_2)), \quad (4)$$

$$\frac{\partial D_2}{\partial t} = \gamma(\alpha\phi'(D_2) + \beta\nabla\cdot(\psi'\nabla D_2) + \lambda\varphi'(D_2 - D_1)), \quad (5)$$

where ϕ', ψ' and φ' are the derivatives of the functions ϕ, ψ and φ, respectively. ψ' is also called diffusion or conduction coefficient (Perona and Malik, 1990). The minimizers are found by computing asymptotic states ($t \to \infty$) of the solutions D_1^t and D_2^t, which are the disparity maps produced by first and second optimizations at iteration t.

The function of the diffusion term is to produce disparity maps that assign similar values to neighboring pixels if there is no depth discontinuity between the pixels, which is called regularization.