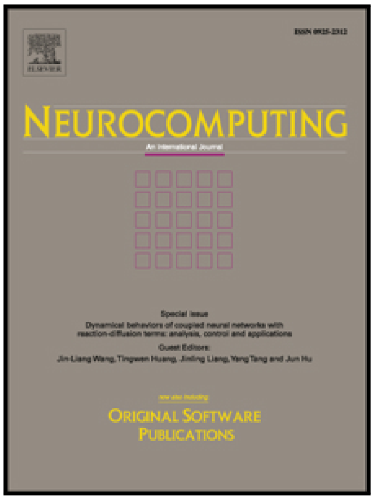
Communicated by Bin Fan

Accepted Manuscript



Learning Both Matching Cost and Smoothness Constraint for Stereo Matching

Menglong Yang, Xuebin Lv

PII: S0925-2312(18)30537-X
DOI: 10.1016/j.neucom.2018.05.008
Reference: NEUCOM 19556

To appear in: *Neurocomputing*

Received date: 17 July 2017
Revised date: 27 April 2018
Accepted date: 1 May 2018

Please cite this article as: Menglong Yang, Xuebin Lv, Learning Both Matching Cost and Smoothness Constraint for Stereo Matching, *Neurocomputing* (2018), doi: 10.1016/j.neucom.2018.05.008

Learning Both Matching Cost and Smoothness Constraint for Stereo Matching

Menglong Yang[a], Xuebin Lv[b,*]

[a]School of Aeronautics and Astronautics, Sichuan University, Chengdu 610064, P.R. China
[b]School of Computer Science and Engineering, Sichuan University, Chengdu 610064, P.R. China

Abstract

A typical stereo matching algorithm involves calculating of the matching cost and smoothness constraint. Recently some algorithms show the good performances by learning the matching cost, but the learning of smoothness constraint have been little researched. This paper focuses on both aspects by using convolutional neural networks. The proposed method first learns a Euclidean embedding for each image using a convolutional neural network with a triplet-based loss function, where the matching cost is directly computed by the squared L2 distances between two vectors in the embedding space. Then we use similar convolutional neural networks to learn the smoothness constraint, and a generalized Semigloba Matching algorithm is proposed to estimate the disparity. The proposed method has a comparable performance with the state-of-the-art algorithms by using smaller and shallower networks, and the experiments show that learning both of matching cost and smoothness constraint has positive effects for stereo matching.

Keywords: Stereo Matching, Convolutional Neural Network, Smoothness Constraint, Semiglobal Matching

1. Introduction

As an fundamental and important step for stereo vision, stereo matching attracts many researchers. It has been widely researched since there was the publicly available testing dataset such as the Middlebury [1] and KITTI 2012 stereo benchmark [2], which allow researchers to compare their algorithms against all the state-of-the-art algorithms.

A type stereo matching algorithm contains four steps summarized by Scharstein and Szeliski [1], i.e. matching cost computation, cost aggregation, optimization, and disparity refinement, respectively. Local stereo methods focused on the first two steps[3, 4] and often fail in challenging scenarios of weakly textured, saturated or reflective regions. Many global methods focused on the research on steps (3) and (4) are thus researched and perform well on Middlebury benchmark, such as graph cuts [5, 6, 7] and belief propagation (BP) [8, 9, 10, 11, 12]. The stereo is achieved in these global algorithms, essentially, by solving a Markov Random Field (MRF) model, including the assumptions of photo-consistency and smoothness. However, stereo for the complex scenarios is still a challenging issue[13].

1.1. Related Work

Many learning-based stereo algorithms[14, 15, 16, 17, 18, 19, 20, 21, 22] have been proposed since the introduction of large stereo datasets [23, 18].

A class of training methods aim to compute or refine matching cost. For example, Kong and Tao [14, 15] initialized the matching cost with sum of squared distances, and using a trained model to predict the probability that whether the initial disparity is correct. The initial matching cost was refined based on the predicted probabilities. Some other works[19, 20] focused on estimating the confidence of the computed matching cost. Recently, J. Zbontar and Y. LeCun [24] trained a convolutional neural network (CNN) to compute the matching cost and their method outperformed on KITTI 2012 benchmark. However calculating the matching cost for a single image pair need more than 1 minute with a Nvidia GeForce GTX Titan GPU. Z. Chen et. al [25] used Siamese structure to train a CNN, where the matching cost is fast computed as the cosine distance of two feature vectors.

Except the matching cost, the stereo algorithms often involve a smoothness assumption to reduce noise by spurious or ambiguous mismatches. Many global algorithms show the effectiveness of encouraging similar pixels to be assigned to the same label, where a typical way is performing an image segmentation and re-

*Corresponding author. Email: lvxb@scu.edu.cn