Pattern Recognition Letters 40 (2014) 66–71

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

A homography transform based higher-order MRF model for stereo matching ☆

Menglong Yang [a,b,*], Yiguang Liu [c], Zhisheng You [a,c], Xiaofeng Li [a,c], Yi Zhang [c]

[a] Key Laboratory of Fundamental Synthetic Vision Graphics and Image for National Defense, Sichuan University, Chengdu 610064, PR China
[b] School of Aeronautics and Astronautics, Sichuan University, Chengdu 610064, PR China
[c] School of Computer Science and Engineering, Sichuan University, Chengdu 610064, PR China

ARTICLE INFO

Article history:
Received 15 May 2013
Available online 10 January 2014

Keywords:
Stereo matching
Higher-order Markov random field
Belief propagation
Homography transform based stereo
Local-planar smoothness

ABSTRACT

Stereo matching is one of the most important and fundamental topics in computer vision. It is usually solved by minimizing an energy function, which includes a data term and a smoothness term. The data term consists of the matching cost, and the smoothness term encodes the prior assumption that the surfaces are piecewise smooth. In contrast to the traditional methods, in which the smoothness term is modeled by the pairwise interactions, the smoothness term is modeled with a higher-order model in this paper. With the prior assumption that a tiny piece of a smooth surface is approximately planar, a higher-order potential function based on the homography transformations is presented. Then the energy function defined on a factor graph is proposed, in which the coefficients of the factors depend on the color information of the input images so that the discontinuous edges are preserved. The belief propagation (BP) algorithm is adopted to minimize the energy function, and the experimental results tested on the Middlebury data set show the potential of the proposed method.

1. Introduction

Stereo is one of the most fundamental topics in computer vision. Stereo vision attracts many researchers and has been widely researched since there was the publicly available performance testing such as the Middlebury benchmark [14], which allows researchers to compare their algorithms against all the state-of-the-art algorithms.

Different from the feature matching [1,18], which matches sparse feature points in two images, the stereo matching can densely match the pixels. The stereo algorithm presented in this paper use the popular energy minimization framework that is the basis for most high-performance algorithms such as graph cuts [9,13,12] and belief propagation (BP) [8,22,21,23]. The stereo is achieved in these algorithms, essentially, by solving a Markov random field (MRF) model. The energy function typically includes a data term and a smoothness term, where the data term consists of the matching cost implied by the extracted disparity map and the smoothness term encodes the prior assumption that the world surfaces are piecewise smooth. Most smoothness terms are modeled using the well known pairwise interactions. The pairwise models are easily implemented by computer, and are thus widely applied.

However, the higher-order models have the ability to encode significantly more sophisticated priors and structural dependencies among image pixels. With the prior assumption that a tiny piece of a smooth surface could be considered approximately planar in 3D world, we propose a novel higher-order model based on the homography transformations. Then an algorithm to minimize the model is presented. The main contributions of this paper are summarized as follows.

The first contribution in this work is the definition of "smoothness" (Section 3.2), which can well express the local-planar bias, rather than the frontal-planar bias used in many traditional methods.

Based on the definition of "smoothness", the second contribution is naturally the proposition of a novel form of the smoothness term (Section 3.3). In contrast to the traditional smoothness terms modeled by the pairwise interactions, the smoothness term is modeled using a higher-order model in this paper. The new energy function is consequently obtained. In addition, we predefined a subset including all potential patterns of "smooth" surfaces for a neighborhood, the exponential computational complexity is avoided.

As the third contribution, we use different weights for the different neighborhoods, to represent different prior likelihoods of that the neighborhoods are "smooth". The weights depend on the

☆ This paper has been recommended for acceptance by Andrea Torsello.

* Corresponding author at: Key Laboratory of Fundamental Synthetic Vision Graphics and Image for National Defense, Sichuan University, Chengdu 610064, PR China. Tel.: +86 028 85372678; fax: +86 028 85372506.

E-mail address: steinbeck@163.com (M. Yang).

0167-8655/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2013.12.020

color information of the input images, and the details are introduced in Section 4.

This paper is organized as follows: Section 2 gives the formulation of the stereo problem and briefly reviews the factor graph. In Section 3, we propose the new smoothness term. Then the factor coefficients are introduced in detail in the Section 4. Section 5 presents the algorithm, Section 6 shows the results on the Middlebury data sets and Section 7 concludes.

2. Energy function

Following the notation in [3,15], the problem of stereo matching can be defined as follows. Let $\mathcal{P}$ be the set of pixels in an image and $\mathcal{L}$ be a set of labels which correspond to disparities. The aim is to find the disparity for each pixel, namely, assign a label $d_p \in \mathcal{L}$ to each pixel $p \in \mathcal{P}$. It typically requires solving the following minimization problem.

$$E(d) = \sum_{p\in\mathcal{P}} E_D(d_p) + \sum_{p\in\mathcal{P},\, s\in\mathcal{N}_p} S(d_p, D_s), \tag{1}$$

where $d = \{d_p\}$ includes the labels of all the pixels. $\mathcal{N}_p$ denotes the set of pixels that are neighbors of p, and D_s denotes the set of all labels for the neighboring pixels of p. $E(d)$ is the energy when the pixels are labeled as d. $E_D(d_p)$ is the cost of assigning label d_p to pixel p, and is referred to as the data term. $S(d_p, D_s)$ measures the cost of assigning labels d_p and D_s to the pixel p and its neighboring pixels, and is normally referred to as the smoothness term. Namely, we assume that the labels should vary smoothly almost everywhere but may change dramatically at some places such as pixels along object boundaries. Minimizing this energy corresponds to the maximum a posteriori (MAP) estimation problem for an appropriately defined MRF [11,2].

The energy function is minimized using belief propagation (BP) algorithm in this paper, because it can well solve the discrete labeling problem in MRF [11]. Belief propagation is commonly used in artificial intelligence and information theory and has demonstrated empirical success in numerous applications including low-density parity-check codes, turbo codes, free energy approximation, and satisfiability [4]. For stereo matching, it has been widely applied and gets good performances [14]. The BP algorithm will be briefly reviewed in Section 5. BP operates on a factor graph, i.e. bipartite graph nodes corresponding to variables $\mathcal{P}$ and factors $\mathcal{F}$, with edges between variables and the factors, as shown in Fig. 1. In terms of the definition in Fig. 1, the energy function can be written as:

$$E(d) = \sum_{f_p\in\mathcal{F}} f_p(\mathbf{d}_p) \tag{2}$$

Here

$$f_p(\mathbf{d}_p) = E_D(d_p) + S(\mathbf{d}_p) \tag{3}$$

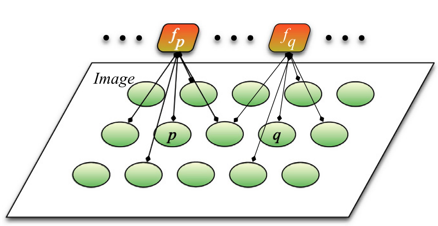

Fig. 1. Bipartite graph. The ellipses represent the variables (image pixels) and the quadrangles represent the factors. Every edge connects a pair of neighbors, i.e. a variable and a factor. Note the number of the factors equals the number of the pixels. We define the sequence number of each factor as the sequence number of the pixel, which is the central one of the pixels connected to the factor.

where $\mathbf{d}_p$ is a vector of labels for the variable nodes connected to the factor node f_p.

3. Smoothness term

3.1. Smoothness in traditional methods

In most algorithms, pairwise interactions are adopted as the smoothness prior, i.e. the smoothness term is $\sum_{p,q\in\mathcal{N}} S(d_p, d_q)$ and measures the cost of assigning labels d_p and d_q to two neighboring pixels p and q [6,8,21–23]. The common pairwise interactions are defined as the jump cost based on the degree of difference between labels. The truncated linear model is commonly used, i.e.

$$S(d_p, d_q) = \rho \min(|d_p - d_q|, \eta), \tag{4}$$

where ρ and η are scalar constants. This equation is defined under the assumption of piecewise-smooth surfaces and the smoothness here implies a prior expectation that the neighboring pixels have the disparities as close as possible. It is conceivable that this smoothness prior assumption could work well for frontal-planar surfaces.

Some algorithms using second-order smoothness priors models get good results [17,10]. In these algorithms, the smoothness terms involve three variables, i.e. the pixel p, its left neighbor pixel l and its right neighbor r. The smoothness term is $\sum_{p\in\mathcal{P}} S(d_l - 2d_p + d_r)$ to robustly favor piecewise linear solutions. In other words, it assume the disparities of pixels on a smooth surface should satisfy the linear constrains as possible.

3.2. Definition of smoothness

In practice, if we take a tiny piece of a smooth surface, it is (approximatively) a plane with few exceptions. Therefore in this work, we assume "smoothness" means that the image pixels in a neighborhood represent the 3D points on the (approximatively) same plane, rather than only the close 3D points. See Fig. 2 for example.

Given a set of points X in 3D space, with their corresponding image points x_i in the reference image and x_i' in the target image, if all the points are on a same plane, there should be a linear transformation, i.e. homography transformation [7] between the points x_i and x_i', i.e.,

$$x_i' = Hx_i \tag{5}$$

where x and x' are homogeneous coordinates of the image points. H is a 3×3 homography matrix, whose elements are

$$H = \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{bmatrix}. \tag{6}$$

Eq. (5) can be expressed in terms of the vector cross product as another form $x_i' \times x_i = 0$. The following equation can be thus deduced.

$$A_i\mathbf{h} = \mathbf{0} \tag{7}$$

where $\mathbf{h} = [H_{11}, H_{12}, H_{13}, H_{21}, H_{22}, H_{23}, H_{31}, H_{32}, H_{33}]^T$, and A_i is a 2×9 matrix calculated by x_i and x_i'. If defining the homogeneous coordinates $x_i = (x, y, z)^T$ and $x_i' = (x', y', z')^T$, A_i could be calculated as

$$A_i = \begin{bmatrix} -xz' & -yz' & -z'z & 0 & 0 & 0 & x'x & x'y & x' \\ 0 & 0 & 0 & xz' & -yz' & -z'z & y'x & y'y & y' \end{bmatrix}. \tag{8}$$