Signal Processing: Image Communication 44 (2016) 44–56

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Stereo matching based on multi-direction polynomial model

Xiao Tan [a,b,*], Changming Sun [b], Tuan D. Pham [c]

[a] The University of New South Wales, Canberra, ACT 2600, Australia
[b] CSIRO Digital Productivity, PO Box 52, North Ryde, NSW 1670, Australia
[c] Aizu Research Cluster for Medical Engineering and Informatics, The University of Aizu, Fukushima 965-8580, Japan

ARTICLE INFO

Article history:
Received 24 March 2015
Received in revised form
29 October 2015
Accepted 23 December 2015
Available online 21 March 2016

Keywords:
Stereo matching
Multi-direction polynomial model
Energy optimization

ABSTRACT

This paper presents a segmentation based stereo matching algorithm. For the purposes of both preserving the shape of object surfaces and being robust to under segmentations, we introduce a new scene formulation where the reference image is divided into overlapping lines. The disparity value and the index of pixels on lines are modeled by polynomial functions. Polynomial functions are propagated among lines to obtain smooth surfaces via solving energy minimizing problems. Finally, the disparity of pixels is estimated from the disparity fields provided by lines. Because lines in multiple directions implicitly segment different objects in an under segmentation region, our method is robust for under segmented regions where it is usually difficult for conventional region based methods to produce satisfactory results. Experimental results demonstrate that the proposed method has an outstanding performance compared with the current state-of-the-art methods. The scene representation method in this work is also a powerful approach to surface based scene representations.

1. Introduction

Stereo matching is the process of finding the correspondences between pixels in two or more images which are captured of the same scene. Matching algorithms can be roughly categorized into pixel based and region based methods. Algorithms in the former category assign each pixel a disparity value according to a similarity measure between two matching pixels such as sum-of-squared-differences or mutual information [1,2] and some constraints such as smoothness constraint and uniqueness constraint. Finding correspondences is then formulated as a task of minimizing an energy function which can be solved by global optimization methods, such as Graph Cuts (GC) based [3], maximum-surface based [4], and Belief Propagation (BP) based methods [5–7] or formulated as a filtering like tasks such as cost volume filtering [8], soft aggregation [9], and information permeability [10]. In the region based algorithms, a region segmentation is carried out beforehand, then a fitting process is conducted on each segment according to a specified model such as the planar model [11–13], the B-splines model [14], and the plane-plus-parallax model [15].

Compared with the pixel based methods, the region based algorithms usually provide good results at textureless regions and at the disparity boundaries which usually coincide with region boundaries. Moreover, they perform well at highly slanted surfaces. However, the region based algorithms are subject to region segmentation results, particularly under segmentations where the assumption that the disparity in a segment follows a particular model is broken. A soft segmentation based method proposed in [14] decides if a pixel belongs to a surface model by minimizing an energy function. A similar method described in [15] updates the relationship between pixels and segments in the object level. Using the soft segmentation and the object information, these two methods are robust to under segmentations. However, as noticed in their work these types of methods are computationally expensive.

This paper proposes a stereo matching algorithm based on a new formulation: multi-direction polynomial model (MDPM) to represent the scene. In this model, the scene is decomposed into intersected lines in multi-directions using the boundaries from region segments. The disparity of pixels on these lines is fitted by a polynomial function. The intersected lines give disparity estimations to the intersecting pixel according to their polynomial functions. The disparities of pixels are chosen from the disparity values of all these estimation values so that the scene is both piecewise smooth and consistent with the local observation. As will be shown in the later parts of this paper, our model takes advantage of the region based methods in that it has good performance at highly slanted surfaces; another advantage over the conventional segmentation based methods is that our model is robust to under segmentations. Compared with scanline based methods [16–18], our method has the following advantages: (1)

* Corresponding author at: CSIRO Digital Productivity, PO Box 52, North Ryde, NSW 1670, Australia.
E-mail address: tanxchong@gmail.com (X. Tan).

http://dx.doi.org/10.1016/j.image.2015.12.008
0923-5965/

using the polynomial model gives better representation than using the linear model [18]; (2) by choosing the disparity value provided by lines rather than all disparity values within the search ranges, the assignment of pixel disparities is carried out efficiently.

2. Overview of the idea and notations

Unlike previous region based methods, this study models the relationship between the disparity of a pixel and its pixel index by a polynomial function on 1D lines. The lines are bounded by region boundaries from region segmentation results (see Fig. 1). Due to the existence of noises in cost aggregation results, the polynomial regression results on lines at some pixels may be incorrect. Fortunately, we find that by alternating the direction of lines at these incorrect pixels, and carrying out polynomial regression once again, the regression results will then provide correct results for many of them. An interesting question that follows is whether we can obtain better results by "fusing" results from polynomial functions in different directions. This paper answers this question by proposing a method to obtain a good disparity map. The proposed method consists of the following steps: (1) Polynomial regression: calculating a polynomial function along each line; (2) Polynomial propagation: obtaining disparity fields where object surfaces are smooth while different objects in under segmented regions are divided; (3) Disparity fusion: combining multiple disparity fields to obtain the final disparity results. Each disparity field corresponds to polynomial propagation results on lines in one specified direction.

Let I be the reference image of the input image pairs, and $L^{(\theta)}$ be a line at direction θ to the horizontal axis. In the proposed algorithm, four directions are used, where θ is equal to 0, $\pi/4$, $\pi/2$ and $3\pi/4$, respectively (see Fig. 2(a)). Lines passing through a pixel are bounded by the region boundaries containing this pixel (see Fig. 2 (b)). In this paper, a region segment is called a region for short.

We model the disparity value of pixels and the index of pixels by a polynomial function: $d_p^{(\theta)} = \mathcal{F}^{(\theta)}(C_p)$, where C_p is the pixel index of p; $\mathcal{F}^{(\theta)}$ is the polynomial function of p; and θ indicates the direction of the line. The polynomial function is explicitly expressed as $\mathcal{F}^{(\theta)}(C_p) = \sum_{i=0}^{\Theta} \alpha_i C_p^i$ where α_i is the coefficient of the ith order term, and Θ is the highest order of the polynomial function. In the rest of the paper, we call polynomial function as polynomial for simplicity.

3. Polynomial regression and polynomial propagation

Polynomials are calculated separately for lines in each of the four directions. We take horizontal direction ($\theta = 0$) as an example to present polynomial regression and the propagation. Methods shown in this section are similarly applied to the other directions. In this research, we assume that the stereo image pairs are rectified and take the left image as the reference image.

3.1. Polynomial regression

Polynomials are computed on individual lines using least squares regression with a random sampling method. More specifically, we randomly sample $n + 1$ non-occluded pixels and solve a linear system between their image coordinate and initial disparity values to calculate a nth order polynomial function. The initial disparity values can be obtained by directly applying the winner-take-all (WTA) method to a cost volume D or by using low level processing methods such as the box-filtering, the edge-aware filtering [8], or the hierarchical belief propagation approach [6] on D to suppress noises. The occluded pixels are detected before performing the polynomial regression. To this end, a cross checking process is carried out on the initial disparity maps of the left and the right images. We distinguish a pixel as an occluded pixel or a non-occluded pixel because the cost volume of an occluded pixel contains the mis-leading information instead of information about true disparity value of this pixel, which will degrade the performance of polynomial regression especially in regions where occluded pixels dominate. For the sake of robustness in high order regression, we scale pixel coordinates to the range of $[0, 1]$ and use the scaled coordinates $\hat{C}_p$ for regression. After regression we adjust the obtained coefficients $\hat{\alpha}_i$ to α_i so that $\sum_{i=0}^{\Theta} \hat{\alpha}_i \hat{C}_p = \sum_{i=0}^{\Theta} \alpha_i C_p$ is satisfied for all pixels on this line. We perform several trials for random sampling. The number of total trails and the

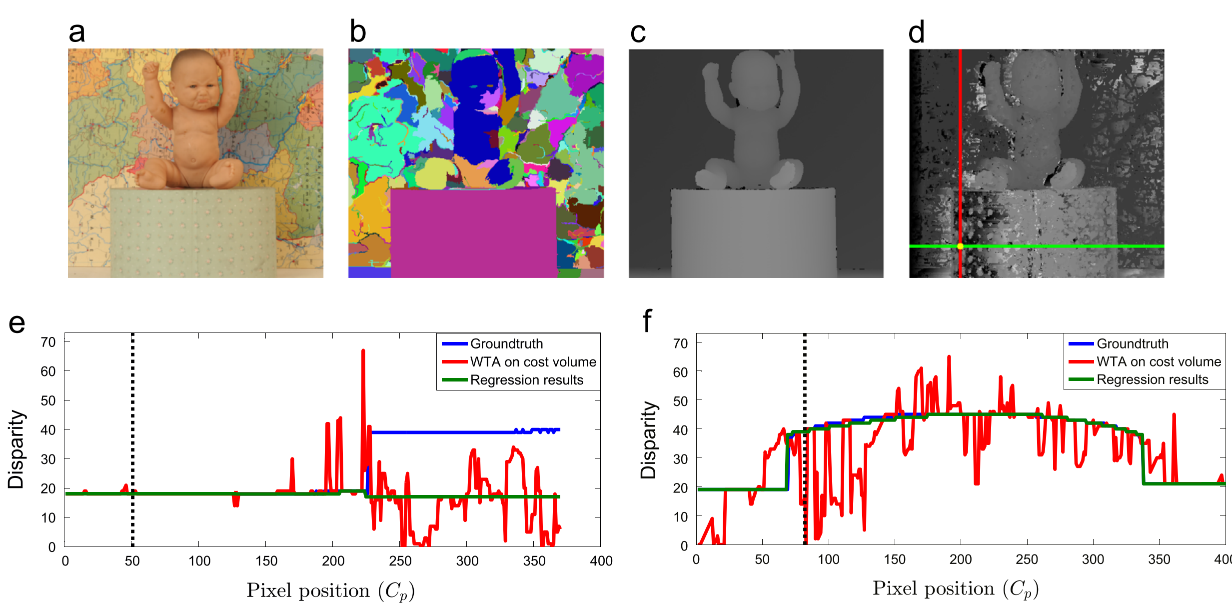

Fig. 1. Multi-direction polynomial regression. (a) Reference image. (b) Region segmentation results. (c) Groundtruth. (d) Polynomial regression along vertical and horizontal directions at the pixel colored in yellow. (e) Polynomial regression results in vertical direction (along the red line in (d)). (f) Polynomial regression results in horizontal direction (along the green line in (d)). The disparity of the yellow pixel is denoted in the vertical dash line in (e) and (f). Regression results along vertical direction are incorrect. After altering the line to the horizontal direction, the regression result at the yellow pixel becomes correct.