Signal Processing: *Image Communication* 29 (2014) 1232–1244

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Stereo matching using cost volume watershed and region merging

Xiao Tan [a,b,*], Changming Sun [b], Xavier Sirault [c], Robert Furbank [c], Tuan D. Pham [d]

[a] The University of New South Wales, Canberra, ACT 2600, Australia
[b] CSIRO Computational Informatics, Locked Bag 17, North Ryde, NSW 1670, Australia
[c] CSIRO Plant Industry, Clunies Ross Street, Canberra, ACT 2601, Australia
[d] Aizu Research Cluster for Medical Engineering and Informatics, The University of Aizu, Fukushima 965-8580, Japan

ARTICLE INFO

Article history:
Received 17 January 2014
Received in revised form
22 April 2014
Accepted 5 June 2014
Available online 17 June 2014

Keywords:
3D/stereo scene analysis
Stereo matching
Watershed segmentation
Region merging
Graph cuts

ABSTRACT

Segment based disparity estimation methods have been proposed in many different ways. Most of these studies are built upon the hypothesis that no large disparity jump exists within a segment. When this hypothesis does not hold, it is difficult for these methods to estimate disparities correctly. Therefore, these methods work well only when the images are initially over segmented but do not work well for under segmented cases. To solve this problem, we present a new segment based stereo matching method which consists of two algorithms: a cost volume watershed algorithm (CVW) and a region merging (RM) algorithm. For incorrectly under segmented regions where pixels on different objects are grouped into one segment, the CVW algorithm regroups the pixels on different objects into different segments and provides disparity estimation to the pixels in different segments accordingly. For unreliable and occluded regions, we merge them into neighboring reliable segments for robust disparity estimation. The comparison between our method and the current state-of-the-art methods shows that our method is very competitive and is robust particularly when the images are initially under segmented.

1. Introduction

Dense stereo matching is one of the key issues in computer vision. Many factors such as noise, distortion, reflection of light, lack of texture and occlusion will affect the matching process and make the matching results unreliable. A survey of studies on stereo matching and a broad range of stereo matching methods can be found in [1]. Stereo matching methods can be roughly categorized into two classes, local and global methods. Methods in the first class build on the hypothesis that neighboring pixels have similar disparities. These methods determine disparities by measuring the similarity between pixels in two local windows in stereo images. To perform well at boundaries, many techniques such as multiple windows [2], shiftable window [3], adaptive window [4], and weighted window [5] have been proposed. Global methods usually formulate the stereo matching problem using a Markov random field (MRF) model for integrating constraints such as the ordering constraint [6–9], the uniqueness constraint [10], the visibility constraint [11], the piecewise smoothness constraint [3,12–14], and the ground control point constraint [15–17]. Some semi-global methods address the stereo matching problem by recursively diffusing the cost to neighboring pixels such as the methods in [18,19] and [20].

In recent studies, most of the advanced methods employ the segmentation constraint [11,21–25]. The segment based methods preserve the boundary of disparity jumps and perform well in textureless regions where

* Corresponding author.

http://dx.doi.org/10.1016/j.image.2014.06.002
0923-5965/

direct matching may suffer from the lack of intensity variation. The segment based methods typically have two main steps, image segmentation and disparity assignment [21,23,26,27]. In the first step, pixels are grouped into segments based on their color or texture. The disparities of pixels in these segments are then fitted according to some models such as a single disparity value [27,28], a disparity plane [22,24,29,30], or more sophisticated functions [25,31].

In [21], a segment based method is described based on the hypothesis that no large disparity jump exists within an image segment. Under this hypothesis, the disparity can be calculated by warping the reference image to the source image. Thus, the stereo matching problem is solved by minimizing a global image warping energy. Hong and Chen [29] proposed another segment based stereo matching method using graph cuts to assign the plane parameters by minimizing an energy function which considers both the observed data and the discontinuity between the segments. As the smoothness between segments is taken into account, this method can handle over segmentation; but it is still built upon the hypothesis that disparities in each segment change smoothly. Sun et al. [11] and Yang et al. [24] used a plane to approximate each segment obtained from a color based segmentation method and then plane approximation is used as a soft constraint to control the labeling of pixel disparities. In [27,28], stereo matching is carried out on the segments obtained from over segmentation results. These two methods do not perform well for slant or curved surfaces because disparities of pixels in a segment are forced to be the same.

The common disadvantage of all segment based methods mentioned above is that they are all built upon the hypothesis that no large disparity jump exists within an image segment. In other words, the 3D surface boundaries are exactly aligned with the segment boundaries. Therefore, they all prefer an over segmentation input to prevent missing any boundaries of 3D surfaces. By exploiting a soft segmentation, the method described in [25] does not require 3D surface boundaries to coincide with segment boundaries. However, it also suffers from under segmentation, as the 3D models are calculated directly from the initial segmentations without using any method to identify the disparity jump within a segment. Unfortunately, not all segments can be well represented by a plane model (see Fig. 1) and not all 3D surface boundaries coincide with segment boundaries. Therefore, over segmentation cannot be guaranteed without some prior knowledge about the image. Furthermore, as stated in [29], it is difficult to correctly estimate the 3D model for small segments which are typically obtained from an over segmentation. The computational cost will also be high if images are highly over segmented. Therefore, such over segmentation may not be a good choice. In this paper, we propose a new segment based stereo matching method. Comparing with the prior arts, the proposed method is very robust to

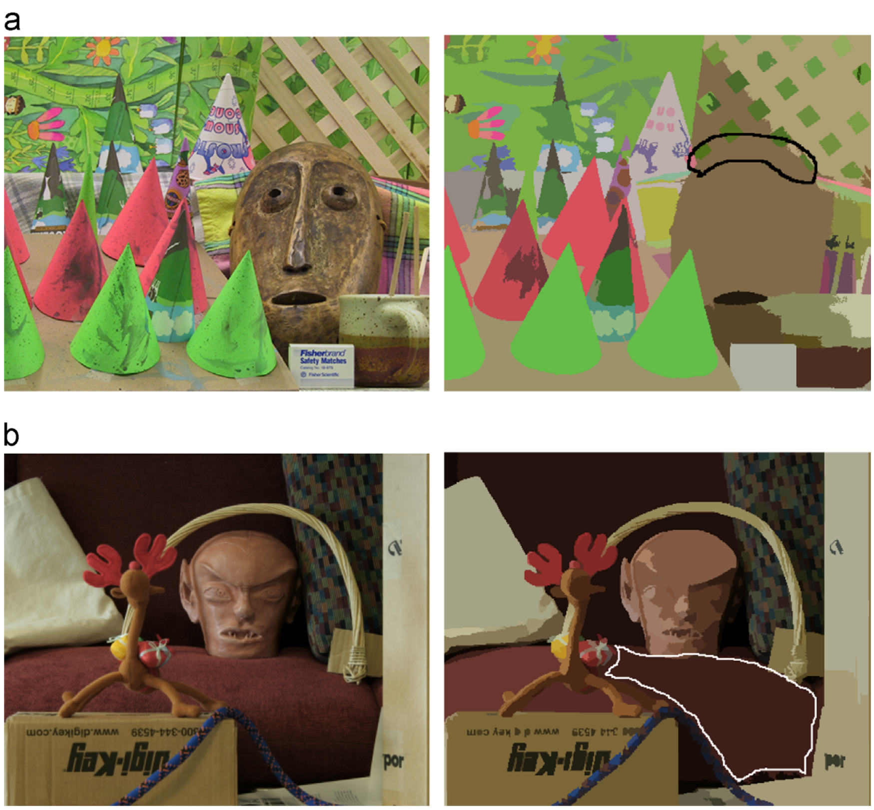

Fig. 1. (a) Left image and initial segmentation of the “Cones” dataset. Regions of two objects within the black contour are incorrectly grouped into one segment. (b) Left image and initial segmentation of the “Reindeer” dataset. The segment within the white contour is a 3D curve on the cushion.