J. Vis. Commun. Image R. 40 (2016) 118–127

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Disparity map enhancement in pixel based stereo matching method using distance transform ☆

Yong-Jun Chang, Yo-Sung Ho *

Gwangju Institute of Science and Technology (GIST), 123 Cheomdan-gwagiro, Buk-gu, Gwangju 500-712, Republic of Korea

ARTICLE INFO

Article history:
Received 6 February 2016
Revised 24 May 2016
Accepted 20 June 2016
Available online 21 June 2016

Keywords:
Stereo matching
Disparity map
Distance transform
Disparity error detection
Disparity error correction

ABSTRACT

With the great success in three-dimensional (3D) movies, a lot of 3D content have been generated. Depth information is one of the important elements in 3D content generation. Stereo matching methods obtain depth information using the characteristic of binocular disparity. These methods find corresponding points between two images which have different viewpoints to calculate the disparity value. However, these methods have difficulties computing accurate disparity values in the textureless region. Smeared pixels near the edge region also make difficult for the stereo matching. In this paper, we propose a pixel based cost computation for the cross-scale stereo matching using the distance transform to improve these problems. In addition, the disparity error detection and correction methods are also proposed as a post-processing step. As a result, we obtain the enhanced disparity map which is robust to the textureless region and the edge region.

1. Introduction

In these days, 3D content is used in various fields such as 3D films, medical images, and 3D games. Depth information is an important requisite for 3D content generation. There are several ways to acquire depth information from the target image. Depth measurement using a depth camera is one of the methods to get the depth value from the object [1]. This method uses infrared rays to measure the distance between the depth camera and the object. Therefore, it can acquire the depth value of the object quickly and accurately. However, the depth camera is vulnerable to the outside because of sunlight. It has an effect on infrared rays of the depth camera.

On the other hand, a depth estimation from captured scenes is not restricted to the place. One of the ways to estimate the depth value from captured scenes is using stereo images. Most of 3D movies use stereo images to generate a 3D effect. These images are acquired by a stereo camera which captures scenes having two different viewpoints. Both images have same objects each other. Each object in stereo images has a disparity value. This value is determined by the distance between the camera and the object. If the object is located near the camera, it has a large disparity value. If the object is far from the camera, it has a small disparity value. For this reason, stereo images allow people to feel the 3D effect.

Stereo matching methods are typical ways to get depth information from stereo images. These methods acquire the disparity value of each pixel in both images. The disparity value is calculated by two corresponding points in stereo images. In order to find the disparity value easily, the image rectification is applied to captured images as a pre-processing algorithm [2]. Therefore, two corresponding points in both rectified images are searched in the same scan line by the epipolar geometry. The result of stereo matching methods is represented as a disparity map. There are two kinds of stereo matching methods. One is a local method and the other one is a global method.

The local method calculates a matching cost of each pixel in stereo images to estimate the optimal disparity value. The matching cost is computed by using similarity measures such as sum of absolute differences (SAD), sum of squared differences (SSD), and normalized cross correlation (NCC). This method considers a limited number of pixels in a specific region to acquire the disparity value of one pixel. Therefore, the local method generally has fast matching results. However, it usually has lower disparity accuracy in the disparity estimation than that of the global method.

On the other hand, the global method considers whole pixels in the image to determine the disparity value of one pixel. It uses an energy function which is based on Markov random field (MRF) to compute the energy between two corresponding points. The

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author.
E-mail addresses: yjchang@gist.ac.kr (Y.-J. Chang), hoyo@gist.ac.kr (Y.-S. Ho).
URLs: http://vclab.gist.ac.kr (Y.-J. Chang), http://vclab.gist.ac.kr (Y.-S. Ho).

http://dx.doi.org/10.1016/j.jvcir.2016.06.017
1047-3203/

energy function is composed of a data term and a smoothness term. The data term calculates the correlation between two corresponding points using similarity measures. This term is similar to the matching cost computation in the local method. The smoothness term checks the disparity consistency among neighboring pixels. The energy function is optimized by several optimization concepts such as belief propagation and graph cuts [3,4]. The global method generally estimates more accurate disparity values than the local method. However, this method is generally slower than the local method.

Both stereo matching methods have problems in the textureless region. Since this region does not have any textures, it is very difficult to find corresponding points in stereo images. For this reason, the stereo matching in this region is problematic. Even though the global method estimates more accurate disparity values than the local method in the textureless region, these problems still remain to be solved as homework yet.

In this paper, we propose a pixel based cost computation using the distance transform to improve the accuracy of disparity values in the textureless region. The distance transform gives the distance value from the edge region to the pixel [5]. Therefore, pixels in the textureless region have specific values by using this transform. This transform also gives a large weighting on the pixel in the edge region. Thus, this transform helps to estimate more accurate disparity value in both regions. Matching costs of the proposed method are aggregated by a cross-scale cost aggregation method [6]. As a result, we acquire an initial disparity map. In addition to this method, we also apply a disparity error correction algorithm to remove remaining disparity errors in the initial disparity map.

This paper is organized as follows. In Section 2, we introduce a disparity accuracy problem in the stereo matching. We also introduce a conventional method which is relevant to this problem. In Section 3, we explain a problem of the conventional algorithm and show a proposed method which has better experiment results than the conventional algorithm. After that, the experiment results are analyzed in Section 4 and we conclude this paper in Section 5.

2. Problem statement

2.1. Disparity accuracy problem in textureless region

In stereo matching methods, a feature detection is an essential step to estimate disparity values between two corresponding points in stereo images. The result of feature detection is affected by the characteristic of regions in the image. The textureless region in the captured scene does not have any features. Therefore, it causes matching ambiguities in the stereo matching. Fig. 1(a) shows the textureless region in the image. Fig. 1(b) shows disparity errors in that region.

In the local method, there is a pixel based stereo matching method to search the matching pixel. This matching method generally has the matching ambiguity problem. Since this method checks the pixel similarity in stereo images using only one pixel, a lot of similar pixels may be existed in the same scan line. For this reason, it is very weak for the textureless region and even some textured regions. Thus, the pixel based matching method generates noises in the disparity map. Fig. 2 shows the matching ambiguity problem of the pixel based matching.

In order to avoid this problem, the local method generally uses a window based matching. This matching method uses the window to find the corresponding pixel in stereo images. It checks the similarity of all pixels in the window. Hence, it usually finds more accurate disparity values in most regions than the pixel based matching method. The result of this method has different qualities depending on the window size. The larger window size is used, the more accurate disparity values in the textureless region are estimated. However, using the large sized window causes inaccurate discontinuity depth values in the edge region.

On the contrary, the global method has great matching results in many regions including the textureless region. However, this method sometimes has a smudged effect near the edge region because of the smoothness term. The smoothness term checks the disparity consistency among neighboring pixels. If the disparity value of current pixel has a large difference with that of neighboring pixel, then this term gives a penalty to the energy function to avoid choosing this disparity value.

2.2. Relevant work

The cross-scale cost aggregation method was proposed by Zheng et al. to improve the disparity accuracy in the textureless region [6]. This method uses multi-scale images to aggregate matching costs among different scale images. It bases on a coarse-to-fine (CTF) strategy [7]. In the process of stereo matching, there are cost noises in the cost computation result. All regions in stereo images have cost noises because of a mismatching problem. Especially, the textureless region has a lot of cost noises. These cost noises lead to estimate inaccurate disparity values in the disparity map.

The low-scale image usually has less cost noises than those of the large-scale image. Since the low-scale image has a lower resolution than the large-scale image, there are few pixel candidates which are used to the similarity measure. Therefore, the low-scale image is more likely to search exact corresponding points in the textureless region than the large-scale image. The work of Zheng et al. uses this characteristic to relive mismatching problem in that region.

In Zheng's method, the consistency checking is used to aggregate refined matching costs [6]. There are two types of consistencies. First one is an intra-scale cost consistency and the other one is a cross-scale cost consistency. The intra-scale cost consistency compares the matching cost of the current pixel with that of neighboring pixels to reduce the cost noise. In this consistency checking step, the least square optimization is used to find the optimal

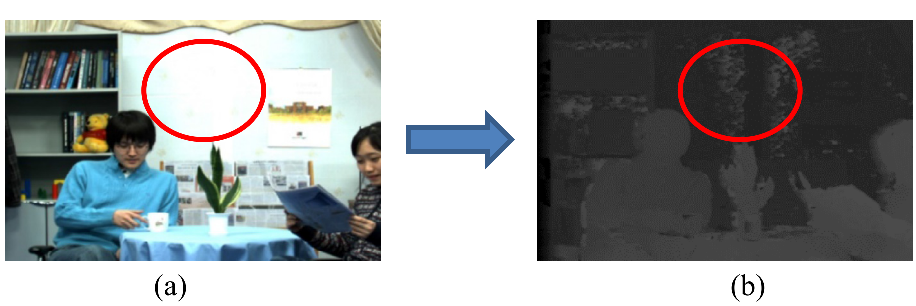

Fig. 1. Stereo matching in the textureless region. (a) The textureless region in the captured scene. (b) Disparity errors in the textureless region.