J. Vis. Commun. Image R. 25 (2014) 1595–1603

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Discontinuity preserving disparity estimation with occlusion handling

Woo-Seok Jang *, 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 28 October 2013
Accepted 23 July 2014
Available online 1 August 2014

Keywords:
Distance transform
Occlusion handling
Stereo vision
Energy optimization
Depth image-based rendering
Depth discontinuity
Hierarchical structure
3D content

ABSTRACT

In this paper, we propose a stereo matching algorithm based on distance transform to generate high-quality disparity maps with occlusion handling. In general, pixel intensities around object edges are smeared due to mixed values located between the object and its background. This leads to problems when identifying discontinuous disparities. In order to handle these problems, we present an edge control function according to distance transform values. Meanwhile, occluded regions occur, i.e., some portions are visible only in one image. An energy function is designed to detect such regions considering warping, cross check, and luminance difference constraints. Consequently, we replace the disparity in the occluded region with the one chosen from its neighboring disparities in the non-occluded region based on color and spatial correlations. In particular, the occlusion hole is filled according to region types. Experimental results show that the proposed method outperforms conventional stereo matching algorithms with occlusion handling.

1. Introduction

Depth images represent distance information between the camera and objects in the captured scene. The depth map is usually provided with its corresponding color image as a pair, often called video-plus-depth [1]. Recently, efficient image generation methods for arbitrary view positions have been vital due to the development of multi-view display devices and three-dimensional (3D) contents. In particular, depth image-based rendering (DIBR) is one of the most widely used methods which create a virtually-synthesized image by projecting color and depth data onto a target-view image plane [2]. The performance of DIBR mainly depends on the quality of depth information.

In general, active sensor-based and passive sensor-based methods exist for measuring depth information from a natural scene. The former employs physical sensors, e.g., infrared ray (IR) sensor, to directly acquire depth data based on the principles of time-of-flight [3]. Usually, the active sensor is more effective in producing high quality depth images than the passive sensor.

However, active sensors suffer from three inherent problems. First, depth data acquisition is difficult if the object is far from the sensor; off-the-shelf sensors allow measuring distances of within 10 m. Second, they are not applicable to outdoor environments. Finally, they produce low-resolution depth images, i.e., less than 640×480, due to challenging real-time distance measuring systems. Such inherent problems make active sensors not practical for various applications. In the industry, their usage is limited to applications mainly involving foreground extraction [4] and motion tracking [5] in indoor environment.

On the other hand, passive sensor-based methods indirectly estimate depth information from 2D images captured by cameras. Such methods can measure depth information of all objects in the captured scene unlike active sensor-based methods. In addition, indirect depth sensing of passive methods is applicable to both indoor and outdoor environments. Another advantage is that the depth image resolution depends on camera resolution, which is not limited to low resolution as in the active sensor. Due to such benefits, the ISO/IEC JTC1/SC29/WG11 Moving Picture Experts Group (MPEG) has utilized passive depth sensing rather than active depth sensing in the 3D video system standardization [6].

Stereo matching is one of the most widely used passive sensor-based methods. This process extracts 3D information from left and right images captured by a stereoscopic camera. In stereo matching, 3D information is calculated by examining the different perspective distortions of objects in the scene of two images. Consequently, in stereo normal case, the different image positions of corresponding image points called disparity is directly related to depth information based on camera parameters.

Over the past several decades, a variety of stereo matching methods have been developed to obtain high-quality disparity maps. However, accurate measurement of depth information from a natural scene still remains problematic due to difficult corre-

* Corresponding author. Tel.: +82 62 715 2258; fax: +82 62 715 3164.
E-mail addresses: jws@gist.ac.kr (W.-S. Jang), hoyo@gist.ac.kr (Y.-S. Ho).
URL: http://vclab.gist.ac.kr (W.-S. Jang).

http://dx.doi.org/10.1016/j.jvcir.2014.07.005
1047-3203/

spondence matching in three types of regions: textureless, discontinuous depth, and occluded areas [7]. First, since color data of the textureless region in left and right images are so similar each other in a wide range, correspondence matching often fails because of its ambiguousness. Second, in case of the depth discontinuous region, i.e., the edge region, smeared color values exist, which leads to ineffective correspondence matching. Lastly, in the occluded region, some pixels may appear in the left image but not in the right image; accordingly, there is no corresponding pixel in the right image.

In this paper, we propose a distance transform-based disparity estimation method with occlusion handling to solve the important problems of stereo matching. Distance transform (DT) [8] calculates the distance to the closest edge for each pixel of an input image. DT values of left and right images control the luminance weighting term for better correspondence matching in edge regions. In addition, an energy function is modeled with three constraints to detect occluded regions. Occlusion hole filling based on color and spatial weighting functions is presented as well. In particular, the proposed hole filling method utilizes different shape of referred windows according to occlusion types, i.e., leftmost occlusion and inner occlusion.

The contributions of our work are as follows; (a) DT-based stereo matching is proposed to increase the accuracy of disparities in the edge region, (b) a new occlusion detection function is designed based on three constraints, and (c) occlusion hole filling is performed adaptively according to occlusion types.

The remainder of this paper is organized as follows. In Section 2, we state the problem in question and briefly introduce the related works. Then, Section 3 presents the proposed method in detail. Section 4 discusses the experimental results followed by conclusions in Section 5.

2. Problem statement

2.1. Occlusion and edge pixel problems

Over the past several decades, occlusion handling has been a challenging task in stereo matching. For left disparity map estimation, the occlusion region represents certain parts of an object that are visible in the left image but not in the right image, and vice versa. Fig. 1(a) illustrates the occlusion problem in left disparity map generation case. The red[1]-marked region appears in the left image only, which means occlusion. The occlusion problem leads to failure of finding corresponding pixels in the right image.

Accurate measurement of depth information in the edge region is important in stereo matching, because depth data of object borders are usually distinguishable. However, as shown in Fig. 1(b), pixels around edges in the left and right images have smeared color values. This affects measuring of discontinuous disparities in the associated area. For reduction of ambiguity in discontinuous regions, several approaches employ variable window sizes or adaptive window shapes via segmentation or pixel-wise similarity measures. The proposed method produces similar effects compared with the approaches using variable window size or adaptive window shapes. While the classical approaches alter the window to determine the pixels, the proposed approach keeps the window and controls the influence of pixels within the regular scope. The proposed method can reduce the effect of inaccurate pixel determination and reflect enough edge influence by distance transform.

In the approaches using segmentation, the influence on the segmentation quality is greater than the algorithm itself. On the other hand, the proposed method can simply calculate the disparity map without prior work such as segmentation. Pixel-wise similarity measure enables the acquisition of scene details. However, this produces poor results in textureless areas and is very sensitive to image noise.

2.2. Previous work

In general, stereo matching can be categorized into local and global methods. Local methods are processed by windows based on correlation where the disparity is assumed to be equal for all pixels within the correlation window [9]. Nevertheless, at discontinuities, this assumption generates blurred object borders and removes small details depending on the size of the correlation window. Thus, such an assumption should be disregarded for depth discontinuities.

In global methods [10], the task of computing disparities is cast as an energy minimization problem. Typically, an energy function for obtaining a disparity map D is formulated as

$$E(D) = E_D(D) + \lambda E_s(D), \tag{1}$$

where E_D is a data term which measures the pixel similarity and E_S is called the smoothness term which penalizes disparity variations. Belief propagation [11], dynamic programming [12] and graph cuts [13,14] are well-known methods for solving this energy function. Generally, global methods are computationally complex even for low resolution images with a small disparity range. Thus, they are not practical. Recently, several methods have been introduced to reduce the complexity of global methods [15–19]. However, the performance of the algorithms considering the practical use is insufficient. Thus, further refinement process is necessary.

In regards to occlusion handing, Kolmogorov and Zabih [14] have proposed an additional occlusion term for the energy function to penalize occluded pixels. Then, the energy function is optimized via graph cuts to compute final disparities. The drawback is that the penalty of the occlusion term depends on only the uniqueness constraint. Liu et al. [20] have presented a two-step local method; the initial matching cost is computed using contrast contest histogram descriptors. Consecutively, disparity estimation is performed via two-pass weighted cost aggregation considering segmentation-based adaptive support weights. In this algorithm, disparity similarities of neighboring pixels which prevent disparity variations are inapplicable to localized results. Ben-Ari and Sochen [21] have introduced a variational approach to find corresponding points. Two coupled energy functions are included for half-occlusion handling and discontinuity map generation. Since optimization is repeated, high complexity is induced. Even though Jang's method [22] generates high quality disparity maps, disparity information in edge regions are not estimated accurately due to its ambiguity. Furthermore, some errors in the non-occlusion region may propagate to the occlusion region during the disparity assignment process.

3. Proposed method

3.1. Overall framework

The proposed method is initially motivated by Yang's work [17] based on hierarchical belief propagation. Due to the hierarchical structure, the previous work computes disparities accurately in the textureless region. Execution speed-wise, their work is one of the most effective global algorithms. For practical use, we adopt this method. However, the quality is insufficient, especially in regards to occlusion and depth discontinuity due to their ambiguity. Thus, we sufficiently refine the results. Based on Yang's work,

[1] For interpretation of color in Figs. 1 and 5–7, the reader is referred to the web version of this article.