The Journal of Systems and Software 86 (2013) 198–210

Contents lists available at SciVerse ScienceDirect

The Journal of Systems and Software

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

Improving Graph Cuts algorithm to transform sequence of stereo image to depth map

Wei-Ming Chen *, Sheng-Hao Jhang

Institute of Computer Science and Information Engineering, National Ilan University, NIU, Ilan, Taiwan

ARTICLE INFO

Article history:
Received 24 April 2012
Received in revised form 9 July 2012
Accepted 16 July 2012
Available online 15 August 2012

Keywords:
Autostereoscopic
Depth map
Graph Cuts
Mean Shift
Stereoscopic Image Sequence
DIBR (Depth Image Based Rendering)

ABSTRACT

Recently, 3D display systems are getting considerable attentions not only from theater but also from home. 3D multimedia content development plays very important role in helping to setup a visual reality entertainment program. Lenticular Autostereoscopic display is one of the 3D-TV having the following advantages such as improving 3D viewing experience, supporting wider viewing angles for multiple viewers, and no requiring any special glasses. However, most of the current 3D movie and camera do not support the Autostereoscopic function. Therefore, we proposed a system that can transform the current 3D stereoscopic image sequence to the depth map sequence. These sequences can be warped into the multiplexed image by DIBR (Depth Image Based Rendering), and show with Autostereoscopic.

Some recent techniques that transform the stereoscopic correspondence problem are based on Graph Cuts. They transform the matching problem to a minimization of a global energy function. However, it has been difficult to include high level information in the formulation of the Graph Cut. In this paper, we describe a new technique for generating depth map sequence from stereoscopic image sequence. We improve the Graph Cuts from pixel-based matching to region-based by using the Mean Shift 3D regions clustering to link the features of images before segmentation. And we also use the result of 3D regions clustering to assign depth values to time domain. After the sequence of depth map has been obtained, the DIBR method was used in transformation process. The experimental result shows that our system not only establishes a mechanism of depth transformation but also improves the accuracy and effectiveness on traditional Graph Cuts.

1. Introduction

There are many ways to show the 3D scenes. In general, user can watch the scenes with or without specific glasses dependent on each different display technology. These two methods will bring slightly different sense to users. Usually, the 3D scenes watched without glasses are more real and fun than another. Therefore, Autostereoscopic display is being the main research field of 3D vision. It is worth noting that Autostereoscopic display needs six to nine views of different degrees to generate one scene. The cost of generating content is more expensive than traditional stereo images. Therefore, according to the problem mentioned above, some researchers (Liu et al., 2007; Liang and Tam, 2005) applied a method called Depth-Image-Based-Rendering to transform the stereo images to multi views of different degree. The DIBR use the depth value obtained from "Depth Map" to measure the coordinate of pixels from different degree. Therefore, no matter the Autostereoscopic needs six or nine views. They can be obtained if the depth of pixels is catchable.

The generation of depth map is a popular research field in 3D vision. Cipolla and Robertson (1999) used the vanish point which is produced due to optical theory to estimate the stereoscopic views. Schneider et al. (1994) proposed the assumption that different depth of pixel should have different blur rate. These two methods were reconstructed the depth value from single view. Due to the restricted information or special scenes, there could have some wrong estimation in depth. And they will affect the outcome of DIBR, too. On the other hand, some researchers applied the method of using calibrated camera to take the left and right images, and reconstructing the depth map from triangle geometry (Falkenhagen, 1994). Moreover, Boykov and Kolmogorov (2001) used Graph Cuts to solve the stereo problem by using data term and smooth term to assign the disparity between left and right images. The energy function is shown in Eq. (1).

$$E(d) = \sum_{p \in P} D(p, d_p) + \sum_{(p,q) \in N} K(p, q) \cdot T(d_p \neq d_q), \quad (1)$$

* Corresponding author.
E-mail addresses: wmchen@niu.edu.tw, R9843001@ms.niu.edu.tw (W.-M. Chen).

0164-1212/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jss.2012.07.044

where P is a set of all pixels in image and $d = \{d_p | p \in P\}$, d_p is a disparity label of pixel p in the left image, N is a set of all pairs of neighboring pixels, and $D(p, d)$ is data term i.e. a penalty for assigning a label d to a pixel p. The function K and T are the penalty of discontinues between d_p and d_q, also called Smooth term. In traditional method, the K function is usually specified by the Euclidean distance between two pixels.

Our experimental result shows that simple energy function can obtain the great semantic segmentation, and more accuracy depth map than other methods. Although Graph Cuts is useful for eliminating noise in the image, the traditional Graph Cuts algorithm was time-consuming due to the pixel-based computing process. In addition, the segmented result was not smooth enough when formulating from simple energy function. To solve these problems, we proposed an improved algorithm that can be used in transformation between L/R images, or say, side-by-side images, and depth map. In our paper, we linked the relation of image sequence by using 3D segmentation to speed up the process of reconstruction. Finally, the output of our system can be transformed from DIBR and shown with Autostereoscopic.

2. Overview

In Fig. 1, our system includes the following two main stages. The function of stage 1 is to link scene object from image sequence on time domain. The function of stage 2 is the reconstruction of depth map. In stage 1, the block motion estimation is employed to provide region information to Mean Shift algorithm. In stage 2, the pre-process step will reduce the inner noise and color complexity of regions of the stereoscopic image sequence to speed up Graph Cuts algorithm first. As the flow shows, Mean Shift 3D region clustering step combines the pre-processed images and the results of block motion estimation to produce the depth map by using Graph Cuts algorithm. Furthermore, these results will be reused to assign the depth value of the following images in sequence.

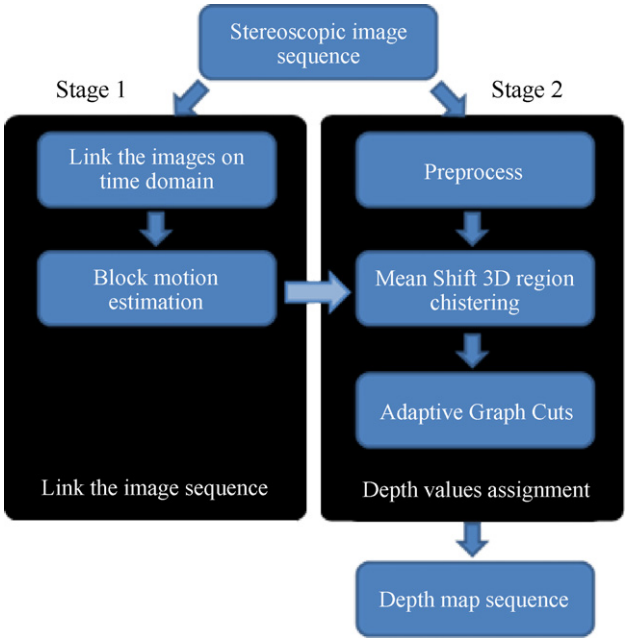

Fig. 1. System flow of depth-map generation.

3. Method

In the proposed method, the transformation can be divided into two main processes. These processes will be described in this chapter in detail. First of all, system reduces the size of the input images to speed up the computing time as well as reserve the accuracy of reconstructions. Next, hexagon-based fast search algorithm (Zhu et al., 2002) is employed to estimate the motion vector of scene that will be further referenced from 3D segmentation. In addition, the pre-process of reconstruction was also applied to enhance the contour and minimize the inner noise and color complexity in each region. Segmentation is hard if an image contains noise. The Gaussian smoothing filter is used for eliminating noise in the image. Since the Gaussian filter blurs the whole image include contours that make the image unclear. In other words, the edges are distorted so that the localization of boundary of regions will not be accurate. This may also reduces the accuracy of region segmentation. We proposed an adaptive algorithm called Separate Gaussian blur since it is derived from Gaussian blur filter. Unlike the traditional Gaussian blur, the proposed filter can reduce the noise and complexity of color but also keep the contour feature that is the important information during reconstruction phase.

The traditional Graph Cuts algorithm is also a time-consuming work for it computes one pixel and their neighbors on whole image each iteration. Followed Mean Shift 3D segmentation step combined the motion vectors and pre-processed images to cluster pixels with similar features into same region. These segmented regions will become the key source of Graph Cuts. This process turns the processing base into 3D area to reduce the computing time and improve the accuracy. Our experimental result shows it can enhance the efficiency and accuracy of depth map reconstruction.

Fig. 2. Left: Block set of original image. Right: Motion estimation of blocks after scene move.