INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11 MPEG2017/ M42432
April 2018, San Diego, USA

Source Gwangju Institute of Science and Technology (GIST)
Status Report
Title [MPEG-I-Visual]: Disparity Estimation using TSS Algorithm
Author Ji-Hun Mun (GIST, jhm@gist.ac.kr)
Yong-Jun Chang(GIST, yjchang@gist.ac.kr)
Yo-Sung Ho (GIST, hoyo@gist.ac.kr)

Abstract

When we estimate a disparity map from stereo images, maximum disparity search ranges are considered to find an optimal matching cost value. However, searching the maximum disparity search range consumes a huge computational power. In order to overcome this problem, we adopt the three step search (TSS) algorithm in the general local stereo matching procedure.

1 Introduction

Currently used MPEG-I-Visual light field sequences can be considered as a set of stereo images. Except for the light field scene which captured by microlens array embedded light field camera, they have enough baseline in each sub-aperture image. Because of that reason, we can apply the TSS based stereo matching algorithm to MPEG-I-Visual light field sequences.

If we conduct the maximum disparity search range method to get an accurate disparity map, that causes a high computational complexity. In this document, we will show how the TSS algorithm handles the maximum disparity search range problem by restricting a disparity search range.

2 Fast Depth Searching Range Selection

The TSS algorithm usually used in conventional video coding technique. This algorithm searches the target pixel coordinate throughout the most efficient motion data with coarse to fine searching pattern. TSS method finds a target image coordinate by applying different step size for matching cost computation.

The left image pixel value $I_L(\mathrm{x},\mathrm{y})$ compares the pixel value on the right side at $I_R(\mathrm{x},\mathrm{y})$ initially. Then next comparison pixel coordinate on right side image would be $I_R(\mathrm{x}-\mathrm{S}_i,\mathrm{y})$, where S_i indicates the i^{th} searching step size. Additionally, the next comparison image coordinate is selected by applying the second searching step size $\frac{S_i}{2}$. As a result of that, second image coordinate on right side image would be $I_R((\mathrm{x}-\mathrm{S}_i)-\frac{S_i}{2},\mathrm{y})$. This process is continuously performed until the step size S_i is smaller than 1. The matching coordinate searching procedure is demonstrated in Fig. 1.

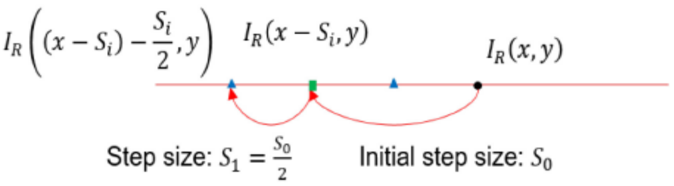

Figure 1. TSS-based target coordinate searching procedure

3 Initial Step Size Determination and Cost function

Depending on the initial step size, estimated disparity map quality takes a serious effect. Too larger step size causes a local minimum problem while searching the target coordinate. In order to solve that issue, we propose adaptive initial step searching method and it is defined in (1).

$$S_{i(x,y)} = \begin{cases} \alpha \cdot (d_{(x-1,y)}+1), & when\ d_{(x-1,y)} < \tau \\ e^{-\frac{var(x,y)}{\varepsilon_v}} \cdot d_{(x-1,y)} + \left(1-e^{-\frac{var(x,y)}{\varepsilon_v}}\right) \cdot PDV, & when\ d_{(x-1,y)} \geq \tau \end{cases} \quad (1)$$

where, $S_{i(x,y)}$ is the initial step size at (x,y), $d_{(x-1,y)}$ is the disparity value at $(x-1,y)$, $var(x,y)$ is the variance of matching block at (x,y), and *PDV* is the predicted disparity value at (x,y). In (1), the below equation is balanced via weighting factor $e^{\frac{var(x,y)}{\varepsilon_v}}$. If $d_{(x-1,y)}$ is smaller than τ, we assume that previous pixel disparity includes error which coming from local minimum issue. To handle this problem, we increase the value of the initial step size. However, if $d_{(x-1,y)}$ is larger than τ, we adopt the initial step size which composed of $d_{(x-1,y)}$ and *PDV*.

As depicted in (1), when $d_{(x-1,y)}$ is larger than τ, the initial step size is affected by variance of matching block which defined in (2). If the matching block is located in homogeneous region, then previous disparity value $d_{(x-1,y)}$ will take more effect than *PDV* and vice versa on textured region. The *PDV* value is computed between current pixel and the three neighboring pixels using the sum of absolute differences.

$$C(x,y) = \frac{\sum_{i=-||\frac{N}{2}||}^{|\frac{N}{2}|} \sum_{j=-||\frac{N}{2}||}^{|\frac{N}{2}|} |G(x+i,y+j)-G(x,y)|}{N^2} \quad (2)$$

In order to generate a disparity map while considering the circumstance of current pixel, we define the matching cost function in (3). ∇I indicates the color difference between current pixel and previous pixel, *DC* is the disparity cost term, and *CC* is the color cost term.

$$Cost(x,y) = e^{-\frac{\nabla I}{\varepsilon_C}} \cdot DC(x,y) + f\left(1-e^{-\frac{\nabla I}{\varepsilon_C}}\right) \cdot CC(x,y) \quad (3)$$

If ∇I has small value, we assume that the disparity value of the current pixel may has a similar disparity value with previous pixel. Therefore, *CC* term will take high weight value than *DC*. Similarly, *DC* term will have higher weight value than *CC* term when the difference of current pixel and previous pixel value has a large gap.

4 Experimental Results

Disparity map estimation test is performed with ULB dense light field sequences [2]. Especially, we choose the stereo images from band 1 region where each of the image frame has 2mm distance. By considering the general scene distance between stereo image, we use image pair which has 2.5cm and 5cm distance for each image.

To demonstrate an efficiency of the TSS-based disparity estimation method, we compare the performance with maximum disparity search range-based method in terms of computational complexity. Fig. 2 shows the estimated disparity map from 2.5cm distance image pair.

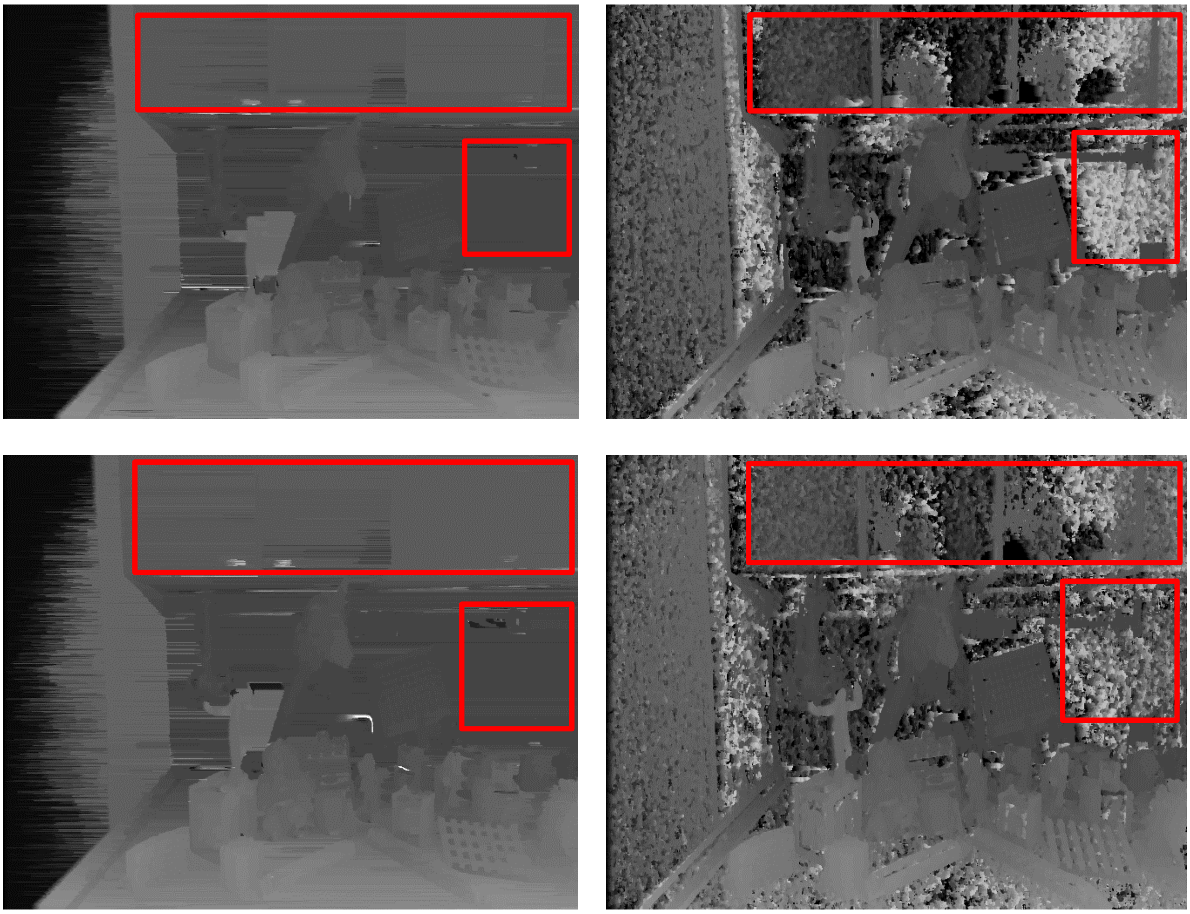

Figure 2. Estimated disparity map from 2.5cm distance stereo image
TSS search result (Left column) / maximum search result (Right column)

Since the TSS-based disparity estimation method considers initial step size for cost computation, the left region in estimated disparity result includes stripe errors. When the TSS search disparity result compare with the maximum search result, it shows more accurate disparity value in homogeneous region as depicted in red boxes.

The computation complexity performance between two methods are evaluated by measuring a processing time. In case of maximum search shows 198.01 sec for each image pair, and TSS presents 42.79 sec in 2.5cm distance image pair.

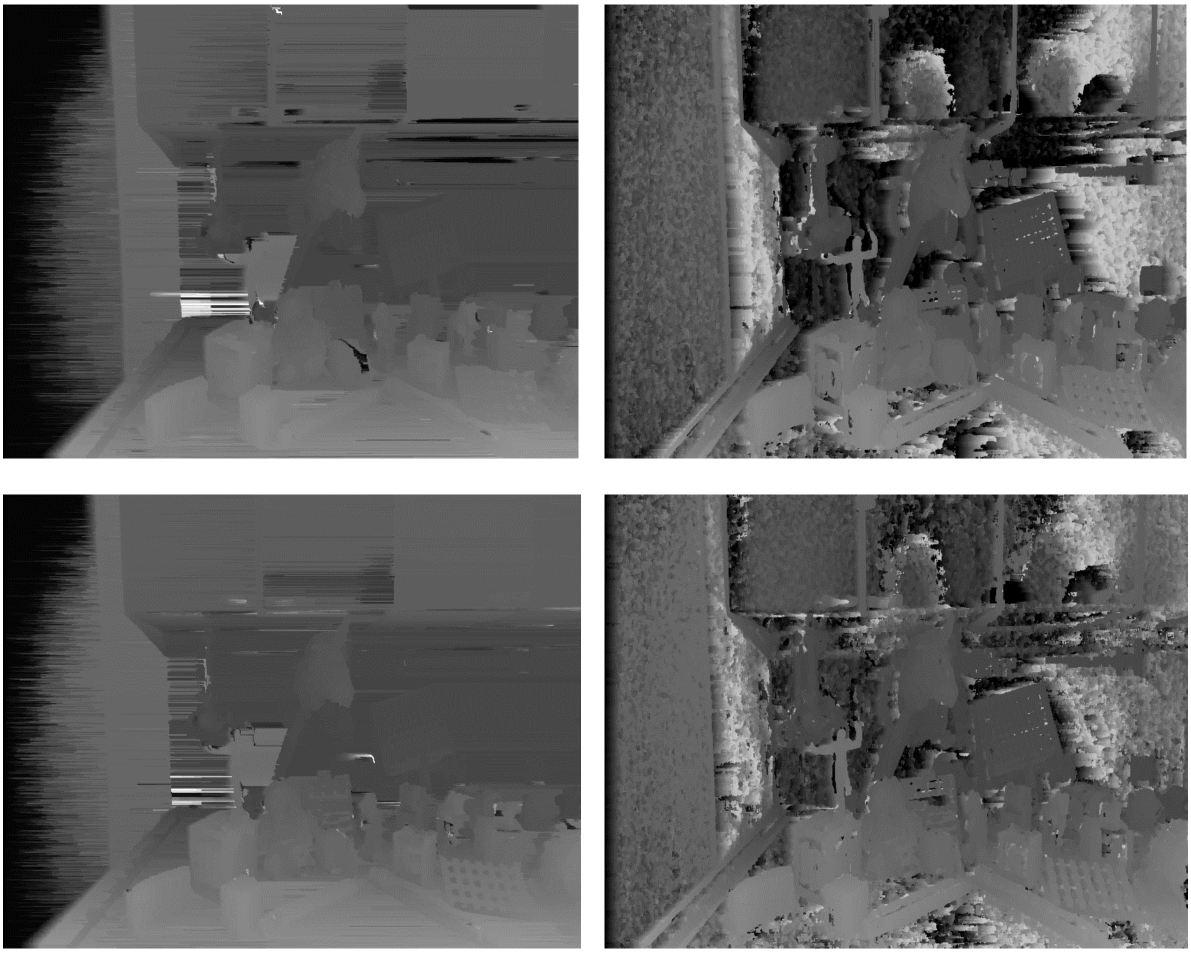

Figure 3. Estimated disparity map from 5cm distance stereo image
TSS search result (Left column) / maximum search result (Right column)

When we use an image pair which has 5cm distance, it shows degraded disparity quality especially in object boundary accuracy as indicated in Fig. 3. However, in homogeneous region, it presents similar quality except for some stripe errors.

The computation complexity performance between two methods are evaluated by measuring a processing time. In case of maximum search shows 394.60 sec for each image pair, and TSS presents 50.40 sec in 5cm distance image pair.

Average time consumption for disparity estimation of each method shows that TSS algorithm explicitly reduce the computational time. TSS-based disparity estimation method takes 46.59 sec, but maximum disparity estimation method consumes 296.25 sec averagely. From these computational time comparison result, we notice that TSS method shows better computational efficiency about 84.3% than maximum disparity searching method.

5 Conclusion

In this document, we propose a TSS-based disparity estimation method. This method can reduce the computation time efficiently when generating the disparity map, compare to other maximum range disparity searching methods. From our experimental results, we expect that the TSS method can be extended to estimate the disparity map in light field scenes.

Acknowledgement

This work was partially supported by National Research Foundation of Korea(NRF) Grant funded by the Korean Government (MSIP) (No. 2011-0030079), and partially supported by 'The Cross-Ministry Giga KOREA Project' grant funded by the Korea government (MSIT) (GK17C0100, Development of Interactive and Realistic Massive Giga-Content Technology).

References

[1] Jon Karafin, Didier Doyen, Gauthier Lafruit, “Lytro Cinema Dense Light Field Video Sequence Specification,” ISO/IEC JTC1/SC29/WG11 MPEG2017/M40357, April 2017, Hobart, AU.

[2] D. Bonattto, A. Schenkel, T. Lenertz, Y. Li, and G. Lafruit, “ULB High Density 2D/3D Camera Array Data Set, version2,” ISO/IEC JTC1/SC29/WG11 MPEG2017/M14083, July 2017.