

# Fast Adaptive Depth Estimation Algorithm Based on K-means Segmentation

Xin Dong<sup>1</sup>, Guozhong Wang, Tao Fan, Guoping Li, Haiwu Zhao, Guowei Teng

**Abstract.** A fast depth estimation algorithm based on the K-means image segmentation is proposed in this paper. Traditional global-based depth estimation methods using graph-cuts made good performance on the depth map estimation, however, the time cost is a big problem for real time application. A fast segment-based method was discussed in this paper, aiming to reduce time cost and at the same time maintain or even improve the performance of depth map estimation. Firstly, the reference image is segmented by K-means method and then mark each segment as different types. Secondly, apply different matching methods for each kind of segment to get the initial matching cost. Thirdly, correct the depth values of unreliable pixels in each segments. Finally, depth values of each segments are determined by using the final matching cost. The experience results demonstrate the superior performance of the proposed algorithm.<sup>2</sup>

**Keywords:** depth estimation; depth map; image segmentation

## 1 Introduction

Along with the development of the 3D technology, people realize the importance of 3DTV and it does give the users a sense of reality. However, compare 3DTV with traditional TV programme, the data size of 3DTV is much larger than the traditional ones, which brings us a big problem in 3D data translation. In order to meet the requirements of the limited bandwidth, it is not convenient to transport each view points of the 3D video, thus it is necessary to use

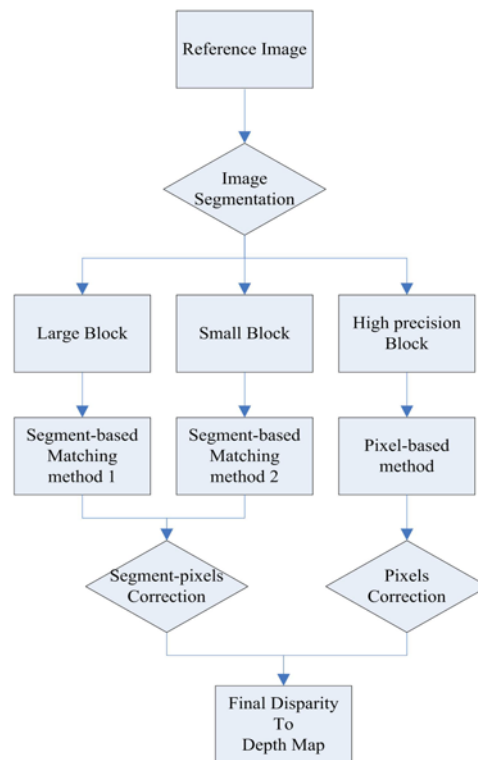
---

<sup>1</sup> Xin Dong (✉)  
Shanghai University, Shanghai China  
e-mail: ziseliuying@126.com

<sup>2</sup> This work was supported by the National Natural Science Foundation of China under grant No. 61271212, Key Project of Shanghai science and technology committee(12511502502)

an immediate 3D representation of view scene system. Using depth maps and synthesizing video point at the decoders can be much more efficient than transporting multi-view videos, thus a fast depth map estimation algorithm is needed.

There are many approaches have been taken to solve the problem of depth estimation, and the most popular methods in solving this problem are local-based method<sup>[1-3]</sup> and global-based<sup>[4]</sup> method. Global-based method defined the energy function to find the optimal parallax plane, it can achieve the optimal depth map in theory, but the algorithm complexity is too high to generate real time depth maps.<sup>[6,7,8]</sup> Local-based method is more efficient and more suit for real-time application than the Global-based method, but local-based method usually use a small window to make stereo matching and the precision of the pixels are related to the size of the window<sup>[10-13]</sup>, which has a seriously influence on the performance in edges. For these reasons, we dedicate to the local-based method and try to find a fast local-based algorithm using an adaptive window to achieve the same performance as the global-based algorithm.



**Fig. 1** Flow chat

## 2 Fast depth estimation Algorithm

The proposed method use the K-means<sup>[4]</sup> segmentation method which can significantly reduce the time cost and at the same time keep the good performance of segmentation. The flow chat is described in figure 1. For the flat areas we adopt adaptive methods to compute the disparity, which is depend on the size of each segment, for the high texture regions, a pixel-based algorithm is used to compute the disparity of each pixel to preserve the texture, then apply a correction step to refine the depth values. In the end, convert the disparities to depth values.<sup>[13-14]</sup>

### 2.1 Image Segmentation

Firstly, we assume the pixels in a flat region have the similar disparity and the disparity varies severely only in the pixels of the texture regions, we apply the K-means segmentation method to divide the reference image into several different segments and mark the segment as  $S_i$  ( $i = 0, 1, 2, \dots$ ).

Secondly, separate the segments into three categories according to the amount of the pixels in each segments. We define the smooth region as large block and small block, and the high texture region as high precision block.

We set two thresholds  $B_1$  and  $B_2$  ( $B_1 > B_2$ ), if the size of  $S_i > B_1$ , we define this segment as large block, if the size of  $S_i > B_2$  and  $S_i < B_1$ , we define the segment as small block and if the size of  $S_i < B_2$ , we define the segment as high precision block.

### 2.2 Calculate the initial matching cost

In this step, traverse the disparity values from  $d_{\min}$  to  $d_{\max}$ , calculate the initial matching cost  $C_1$ , and the corresponding disparity is  $d_1$ .

The simple SAD(sum of absolute differences) measure is performed in the large blocks to aggregate matching cost, as shown in equation 1,  $C_{SAD}(x, y, d)$  is the matching cost,  $W(x, y)$  represents the segment of the large block and  $(x, y)$  is the coordinate value,  $I(u, v)$  is the luminance value of the pixel  $(u, v)$ ,  $T_{\cos t}$  is the time cost between two frames.

$$C_{SAD}(x, y, d) = \sum_{(u,v) \in W(x,y)} |I_L(u, v) - I_R(u-d, v)| + T_{cost} \quad (1)$$

In order to avoid the error in edges, we also combined the SAD with a single direction gradient based measure for the small blocks. In equation 2 ,  $C_{SGRAD}(x, y, d)$  is the matching cost of gradient function ,  $\nabla_x I(x, y)$  represents the gradient function in x direction. Then we can gain the the matching cost of the small block in equation 3.

$$C_{SGRAD}(x, y, d) = \sum_{(u,v) \in W(x,y)} |\nabla_x I_L(u, v) - \nabla_x I_R(u, v)| \quad (2)$$

$$C(x, y, d) = \omega_1 * C_{SAD}(x, y, d) + \omega_2 * C_{SGRAD}(x, y, d) + T_{cost} \quad (3)$$

In the high precision block, unlike the first two steps, these segments need to be more precise, we calculate the cost in pixel-wise, and compute the error cost point by point. We combined MAD(mean absolute differences) with an eight direction gradient based measure for the high precision blocks.<sup>[15]</sup> As shown in equation 4 and 5 ,  $C_{MAD}(x, y)$  is the matching cost of MAD, and in equation 6 and 7,  $C_{SGRAD}(x, y)$  is the matching cost in eight direction gradient based measure ,then the final matching cost is  $C(x, y, d)$  in equation 8.

$$M(x, y) = \frac{1}{9} * \sum I(x-a, y-b) \quad a, b \in \{-1, 0, 1\} \quad (4)$$

$$C_{MAD}(x, y) = M_L(x, y) - M_R(x-d, y) \quad (5)$$

$$G(x, y) = I(x, y) - I(x-a, y-b) \quad a, b \in \{-1, 0, 1\}, a|b \neq 0 \quad (6)$$

$$C_{SGRAD}(x, y) = \frac{1}{8} * \sum |G_L(x, y) - G_R(x-d, y)| \quad (7)$$

$$C(x, y, d) = \omega_1 * C_{MAD}(x, y, d) + \omega_2 * C_{SGRAD}(x, y, d) + T_{cost} \quad (8)$$

In the end, we will get the initial matching cost  $C_1$  by applying the WTA(winner-take-all) and get the corresponding disparity  $d_1$  in equation 9, at the same time ,we take down the minimum matching cost  $C_{min1}$  and the second minimum matching cost  $C_{min2}$  ,if  $|C_{min1} - C_{min2}| > threshold1$  , we mark the pixel as unreliable pixel.

$$d_1(x, y) = \min_{d \in (d_{min}, d_{max})} C(x, y, d) \quad (9)$$

## ***2.3 Calculate the final matching cost***

### **2.3.1 Correction of the unreliable pixels in large blocks and small blocks**

We need to make a refinement in each segment, this is because we have computed the initial disparity values which are achieved by the previous segmentation step, and the precision is relatively low, the pixels in the same segment have the single disparity values, this may not match the real disparity values, so we make a pixel wise refinement to correct the unreliable pixels.

In this step, we will find the unreliable pixels in the blocks and mark these pixels by a new value. In order to get more precision, traverse the disparity values from  $d_1 - 1$  to  $d_1 + 1$ , calculate the minimum matching cost  $C_2$  in pixel-wise, and the corresponding disparity is  $d_2$ . We use the MAD in the pixel-wise based method and compute the matching cost point by point.

Compare  $d_1$  and  $d_2$ :

If  $d_1 = d_2$ , this means the result in the first and second process are the same and then we assume the disparity value is correct and mark the pixel as reliable pixel, remain  $d_1$  unchanged, assign the final disparity value  $d(x, y) = d_1(x, y)$ ;

If  $d_1 \neq d_2$  this means the result in the first and second process are different, so mark the pixel as unreliable pixel;

If a pixel is marked as unreliable, compare the corresponding  $C_1$  and  $C_2$ , if  $abs(C_1 - C_2) > threshold2$ , this means the two results in the first and second process vary severely, then we assign the final disparity value of this pixel as the second result:  $d(x, y) = d_2(x, y)$ .

### **2.3.2 Correction of the unreliable pixels in high precision blocks**

In this step, to achieve reliable pixels we refine the disparity values by using a weighted bilateral filtering.

When we get the final disparity  $d(x, y)$ , transform the disparity to depth values  $z(x, y)$  using the method in reference [15].

### 3 Experiment Results

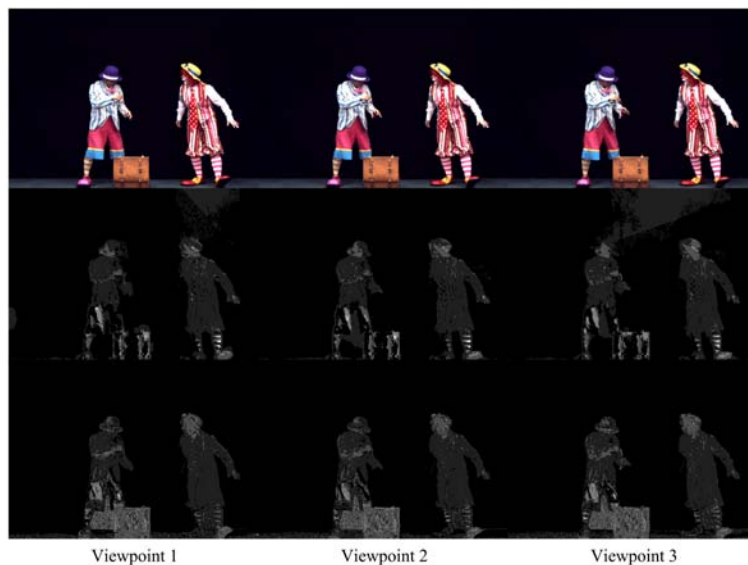
In this section, we will compare our result with the results of the depth estimation reference software(DERS) in reference[5].

For image segmentation, we choose the K-means segmentation method for time saving, we have compared the several matching methods such as mean-shift[5] and pyramid segmentation[5], K-means method achieves the fast segmentation and at the same time keep the border better.

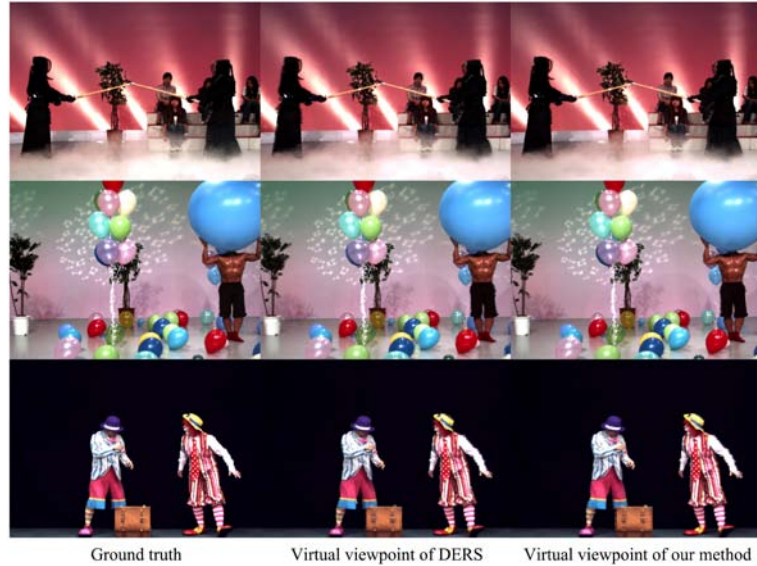
In image segmentation step, we set  $B_1 = 50, B_2 = 10000$ ; in calculate the matching cost step, we set  $\omega_1=0.5$  and  $\omega_2=0.5$ ,  $threshold1=5$ ,  $threshold2=2$ .

Figure 2 is the comparison of the results in reference[5] and in our method, the first line is the three views of video sequence Pantomime, the second and third line are the corresponding depth maps generated by reference[5] and our method, respectively.

Figure 3 shows the comparison of the virtual images of video sequence Balloons, Kendo and Pantomime between the results in reference [5] and our method. From the subjective assessment we can conclude that our method achieves better results in depth map generation.



**Fig. 2** Comparison of depth maps of three views. From the first line to the last line are three viewpoints of ground truth, depth maps in reference [5] and depth maps in our method.



**Fig. 3** Comparison of virtual viewpoints in reference [5] and our results, from the first column to the last column is ground truth picture, virtual image in reference [5] and our method.

The objective evaluation in Table 1 is the average objective results. From table 1 we can conclude that average PSNR in the proposed method is 0.2dB higher than the results in reference [5], what's more, the computational time costs of the proposed depth estimation method are reduced by 61-70% in comparison with the global-based method in reference [5].

**Table 1** Objective evaluation of the average PSNR and running time

Sequence	PSNR (our method)	PSNR (DERS)	Running Time (our method)	Running Time (DERS)
Balloon	41.323567dB	40.7489292dB	34.6130 s	89.4325 s
Kendo	41.788944dB	41.4412448dB	30.7231 s	93.3654 s
Pantomime	39.219330dB	39.2116935dB	41.6325 s	143.9337 s
Champagne	38.972679dB	38.8973295dB	94.947 s	408.809 s

## 4 Conclusion

In this paper, we have presented a fast depth estimation method based on the image segmentation, the results obtained in our method is compared with the depth estimation reference software (DERS), and the matching window is adaptive

with the segment, we combined the block-based and pixel-based matching method to acquire accurate matching cost. The Signal To Noise Ratio of the virtual image is higher than DERS and the computational time cost is lesser. The results show that the proposed method presents a high quality depth map with a faster speed.

## 5 References

1. Cancino-Suarez, S., Klimaszewski, K., Stankiewicz, O., Domanski, M. Enhancement of stereoscopic depth estimation by the use of motion information[J]. System Theory (SSST), 44th Southeastern Symposium on, 2012:94-98.
2. Ekmekcioglu, E., Velisavljevic, V., Worrall, S.T. Edge and motion-adaptive median filtering for multi-view depth map enhancement[J]. Picture Coding Symposium, 2009:1-4.
3. Ekmekcioglu, E., Velisavljević, V., Worrall, S.T. Content Adaptive Enhancement of Multi-View Depth Maps for Free Viewpoint Video[J]. Selected Topics in Signal Processing, IEEE Journal, 2011, 5(2):352-361.
4. Andreas Klaus, Mario Sormann and Konrad Karner. Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure[J]. Pattern Recognition, ICPR 2006, 18th International Conference on, 2006, 3:15-18.
5. Olgierd Stankiewicz, Krzysztof Wegner, Menno Wildeboer, International Organisation For Standardisation Organisation International De Normalisation ISO/IEC JTC1/SC29/WG11 Coding Of Moving Pictures And Audio ISO/IEC JTC1/SC29/WG11 MPEG/M17049 October 2009, Xian, China.
6. X Mi, F Cheng, H Zhang - Image and Signal Processing (CISP), Stereo Matching based on Global Edge Constraint and Variable Window Propagation, Image and Signal Processing (CISP), 2012 5th International Congress on, 16-18 Oct. 2012, Page(s): 936 - 940.
7. Y Yin, M Jin, SY Xie, A Stereo Pairs Disparity Matching Algorithm by Mean-Shift
8. Segmentation, Image and Signal Processing (CISP), 2012 5th International Congress on, 16-18 Oct. 2012, Page(s): 936 - 940.
9. Anders Olofsson, Modern Stereo Correspondence Algorithms: Investigation and evaluation, 2010.
10. X Bai, S Kamata, An Efficient Window-Based Stereo Matching Algorithm using Foreground Disparity Concentration, Control Automation Robotics & Vision (ICARCV), 2012 12th International Conference on, 5-7 Dec. 2012, Page(s): 1352 - 1357.
11. X Liu, ZH Li, DM Li, LY Wang, Computing Stereo Correspondence based on Motion Detection and Graph Cuts, Instrumentation, Measurement, Computer, Communication and Control (IMCCC), 2012 Second International Conference on, 8-10 Dec. 2012, Page(s): 1468 - 1471.
12. Gorkem Saygili, Laurens van der Maaten, Emile A. Hendriks, Improving segment based stereo matching using SURF key points, Image Processing (ICIP), 2012 19th IEEE International Conference on, Sept. 30 2012-Oct. 3 2012, Page(s): 2973 - 2976.
13. X Mi, F Cheng, H Zhang, Stereo Matching based on Global Edge Constraint and Variable Window Propagation, Image and Signal Processing (CISP), 2012 5th International Congress on, 16-18 Oct. 2012, Page(s): 936 - 940.
14. Y Boykov, O Veksler, R Zabih, Fast Approximate Energy Minimization via Graph Cuts, Pattern Analysis and Machine Intelligence, IEEE Transactions on (Volume:23, Issue: 11), Nov 2001, Page(s):1222 - 1239.
15. H Hirschmuller, D Scharstein, Evaluation of Stereo Matching Costs on Images with Radiometric Differences, Pattern Analysis and Machine Intelligence, IEEE Transactions on (Volume:31, Issue: 9), Sept. 2009, Page(s): 1582 - 1599.