Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology
Vol.5, No.3, June (2015), pp. 37-46
http://dx.doi.org/10.14257/AJMAHS.2015.06.26

The Research of Depth Discontinuity Detection Using Adaptive DIP

Gil-Ja So[1], Sang-Hyun Kim[2]

Abstract

Depth discontinuities are the layers of the world and is used to track object, image retrieval, evaluation of the 3D visual fatigue. To extract depth discontinuities from disparity map, we adopt DIP operator that is the difference of inverse probabilities for extracting sketch features that contain valleys and edges in 2D image. However, DIP needs a threshold to determine depth discontinuities and this is threshold needs user interaction. In this paper, we propose the method that extracts depth discontinuities without user interaction using learned threshold. We learned the threshold with the Middlebury disparity maps and made the evaluation to decide the threshold.

Keywords : depth dicsontinuities, adaptive DIP, learned threshold, extracting edges, disparity map

1. Introduction

Depth discontinuities are defined as points on the image plane where the depth field is discontinuous[1]. Depth discontinuity is a powerful clue for many real-tasks like as tracking, detection, image retrieval. Little and Gillett[2] used an average of the local region and matched a left and right image of stereoscopy. Afterwards matching, they inferred depth discontinuities from pixels of occluded regions. Toh and Forrest[3] defined a depth discontinuity as a boundary which a left and right do not match.

However most of these research used information on texture in the image, and can not extract well depth discontinuity on the boundary of the occluded region. To solve this problem Stan Birchfield and And Carlo[4] proposed pixel to pixel based depth discontinuity extraction algorithm.

This algorithm used an dynamic programming for the faster running time. However, this method does not enforce the inter-scanline inconsistency, leads to the horizontal "streaking" artifacts[5][6]. To reduce this problem, [4] propagates information is need between scanlines in post production.

Received (April 12, 2015), Review Request(April 13, 2015), Review Result(May 01 , 2015)
Accepted(May 25, 2015), Published(June 30, 2015)

[1](Corresponding Author)626-790 Dept. Cyber & Police Science, YoungSan Univ. 288 Junam-ro, Yangsan, Gyeongnam, Korea
email: kjso@ysu.ac.kr
[2]626-790 Dept. Computer Engineering, YoungSan Univ. 288 Junam-ro, Yangsan, Gyeongnam, Korea
email: ksh50@ysu.ac.kr
* 이 논문은 영산대학교 교내 연구비 지원으로 작성되었습니다.

ISSN: 2383-5281 AJMAHS

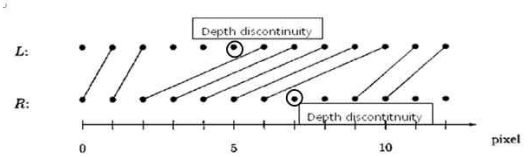
We adopt DIP(Difference of inverse probabilities) operator to get reliable depth discontinuity. DIP is an operator to extract edges subject to local intensities from 2D image[7][8]. The intensity of the 2D image is the brightness or materials. In contrast, the intensity of the disparity map is the depth or layers of the disparity map. Because of the similarity of the depth discontinuity in disparity map and the edge in the 2D image in the side of using the information of the gradation of the intensity, DIP could be applied for extracting depth discontinuities from disparity map. In this paper, we suppose the method to extract depth discontinuity from disparity map by DIP. DIP needs a appropriate threshold to get appropriate discontinuity from disparity map. This needs a user's interaction during process. We infer a threshold by characteristics of the distribution of the disparity map, that is, mean and standard deviation of disparity.

The paper is organized as follows. We would briefly review related previous works and describe our proposed method in section 2. We present our method to extract depth discontinuities using DIP operator in section 3. We show the effect of the threshold in DIP and the evaluation formula to get appropriate threshold for extracting depth discontinuity from disparity map using DIP operator in section 4. In section 5 contains some of our conclusion.

2. Related Works

2.1 Pixel to Pixel Depth Discontinuities

An algorithm which extract depth discontinuities based on Pixel to Pixel(throughout this paper, this algorithm would be called P2P label pixels which are greater than threshold, and extract these pixels as depth discontinuity. Advantage of this method is that depth discontinuity can be extracted without texture, is robust in image sampling, process time is fast by dynamic programming. Postprocessor of this method propagates disparity map vertically to get clearer disparity map.



[Fig. 1] The example of match sequence,
M=<(1,0),(2,1),(6,2),(7,3),(8,4),(9,5),(10,6),(11,9),(12,10),(13.11)>
The depth discontinuity pixels are circled

P2P produces the disparity map using left and right image of stereo images using intensity based stereo matching algorithms. P2P match pixels one scanline of the left or right image to pixels in the corresponding pixels of other. These matched pixels are arrayed in the form of match sequences. Fig. 1 shows a match

sequence on a short scanline. Unmatched pixels are occluded. depth discontinuity pixels occur at the right of occluded pixels in the left image and at the left of occluded pixels in the right image in stereoscopy.

P2P propose a simple cost function to select a best sequence as shown in (1). The best match sequence has the lowest cost.

$$\gamma(m) = N_{occ}K_{occ} - N_m K_r + \sum_{i=1}^{N_m} d(x_i, y_i) \quad (1)$$

In (1), K_{occ} is the occlusion penalty constant, K_r is the match reward constant, $d(x_i, y_i)$ is the distance between pixel x_i and pixel y_i, which is the pixel of the left and right stereo image, each. N_{occ} is the number of occlusion. N_m is the number of matched pairs.

P2P searches for the best possible path to make a cost of the sequence the lowest by the technique of dynamic programming. The process of finding a best match sequence in one scanline is independent of other scanlines. However, the intensity values of pixels from different scanlines are not independent. Thus P2P needs a post process which propagate information from rows and columns together to use all the information in the images. After post process, depth discontinuity is selected as a point which are accompanied by changes of at least two disparity levels.

2.2 DIP

DIP is the algorithm to detect edges which are the regions involving abrupt changes of intensity, and valleys which are the regions composed of local intensity minima. Valley is very important in vision. Various method to detect valley like as Laplacian, Pearson's logical valley use the gradient values of the pixels in a local region. However Laplacian is too sensitive to noise and Pearson's logical valley cannot sometimes extract valleys which have somewhat small rates of change of intensity.

The entropy operator computes the entropy of intensity in a local region. This method depends on the local intensities, therefore, can extract the edges of dark regions very well. The disadvantage of this method is that it extract edges as thick lines and cannot respond in valleys very well.

The human viewer is more sensitive to the edges and valleys in dark regions than those in bright regions. Therefore, to perceive and analyze objects in a manner akin to the human visual system, one must extract sketch features subject to the local intensities.

DIP satisfies the necessity for the perception and analysis of the human visual system as mentioned above. DIP is the in defined as (3).

$$DP(m,n) = \frac{I_m(m,n)}{I(m,n)} - \frac{I(m,n)}{I(m,n)} \quad (2)$$

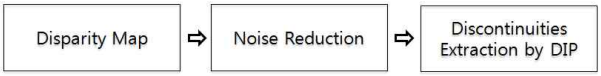
$$DIP(m,n) = DP\frac{\overline{I(m,n)}\,\overline{I(m,n)}}{I_m(m,n)I(m,n)} \quad (3)$$

$I(m,n)$ is the intensity of the pixel (m,n), $\overline{I(m,n)}$ is the sum of intensities and $I_m(m,n)$ is the maximum intensity in a window, in (2) and (3). Due to $\overline{I(m,n)}$, the value of DP depends on local intensities and DP responds well in valleys and edges owing to the difference being taken. However, DP has a defect of extracting neighbor pixels as well as the valleys themselves.

To remove this problem, DIP is proposed. Each $\overline{I(m,n)}$, $I_m(m,n)$ are much the same in valleys and neighbors, while $I(m,n)$ is smaller in the valleys than in their neighbors. Thus, the values of DIP are so much larger in the valleys themselves that it extracts valleys thinly. To prevent DIP from bing extremely large in spite of small rates of intensity change in very dark regions, DIP applied only when the difference $I_m(m,n)$ and $I(m,n)$ is greater than a given threshold.

3. Proposed Method

The proposed system is presented in Fig. 2. Disparity map is extracted from two images of the stereoscopy. In this paper we use the disparity map extracted using structured light in Middlebury dataset[10]. Holes which occur in occluded regions should be filled before the next step. Hole is filled with the interpolation value of neighbor pixels. In the next step, edge is detected from the pre-processed disparity map using adaptive DIP operator. Detected edges from the disparity map are depth discontinuities.



[Fig. 2] The process of the proposed system

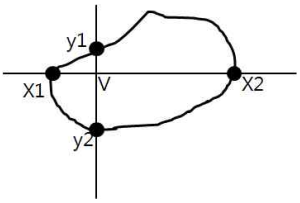
3.1 Disparity map

Disparity map is extracted by the method used by Middlebury[10]. They devised a method to acquire high-complexity stereo image pairs automatically with pixel-accurate correspondence information. In this approach, a pair of camera and one or more light projectors are used. Each camera uses the structured light sequence to determine a unique code for each pixel. Finding inter-image correspondence then trivally consists of finding the pixel in the corresponding image that has the same unique code. They present stereo data sets acquired with this method and demonstrate their suitability for stereo algorithm evaluation. We use these dataset for learning the threshold in DIP and evaluate the result of the discontinuity using this learned threshold.

3.2 Noise Reduction

There are many holes in the disparity map due to occluded region. In the next step, edge detection from the disparity map by DIP, the depth of pixel would be the denominator of the equation to calculate the gradient of the depth. If the depth of a pixel in disparity map is zero, this pixel is neglected in the extraction of depth discontinuity. However, neglecting the zero pixel, the hole, would be the cause of the disconnected edge in the discontinuity detection by DIP. Therefore, the hole should be filled with non-zero value.

To fill these holes, we calculate the value of pixel in holes using the horizontally and vertically interpolating value of the pixel in the holes' boundary. Equation (4) shows the interpolation of the pixel in the hole with the nearest neighbor pixels.



[Fig. 3] Pixel v in occluded region

The pixel v is a pixel in the hole, x_1, x_2, y_1, y_2 are pixels in the boundary which meet horizontally and vertically extended line. If v is in the occluded region, the depth of the v is interpolated with the differently weighted x_1, x_2, y_1, y_2.

$$I(v) = \frac{(x_1 w_1 + x_2 w_2) + y_1 w_3 + y_2 w4)}{2} \tag{4}$$
$$w_1 = \frac{dist(x_1,v)}{dist(x_1,x_2)} \quad w_2 = \frac{dist(x_2,v)}{dist(x_1,x_2)}$$
$$w_3 = \frac{dist(y_1,v)}{dist(y_1,y_2)} \quad w_4 = \frac{dist(y_2,v)}{dist(y_1,y_2)}$$

$dist(x,y)$ is the distance of pixel x and y. In (4), w_1, w_2, w_3, w_4 is obtained with the ratio of the distance between v and neighboring pixels.

3.3 The extraction of the depth discontinuities by DIP

The value of DIP is high at discontinuity pixel. So, we need a threshold to determine the value of DIP

which is a efficient value to extract discontinuity pixels. In other words, the pixel of which value of DIP is greater than a threshold would be considered as discontinuity one.

The result of the extraction is very various as to the value of threshold. As a threshold is so high, depth discontinuities would be extracted less. On the contrary to this, as a threshold is too low, depth discontinuities would be extracted too less. Therefore, threshold is very important in the extraction of depth discontinuities. Usually, threshold is decided by user during the process. The interaction of the user during the process is improper to extract a depth discontinuity in real time. To extract depth discontinuities in real time, a threshold should be decided without the interaction of the user efficiently.

The quality of the extraction depends on the ratio of discontinuity pixels(RD) to the image size of disparity map. So, the threshold needs to be determined using RD. For this reason, we select a threshold with the disparity level that make the right-hand side of CDF(cumulative distribution function) of DIP of disparity map to be greater than RD. However, we need which RD is efficient to make the extraction of discontinuities to be the best. In this paper, efficient RD(ERD) is learned with the mean and standard deviation of 10 disparity maps. After learning, ERD is determined by the evaluation formula in (5).

$$ERD = 0.029 * M(x) + 0.014 * S(x) + 4.884 \quad (5)$$

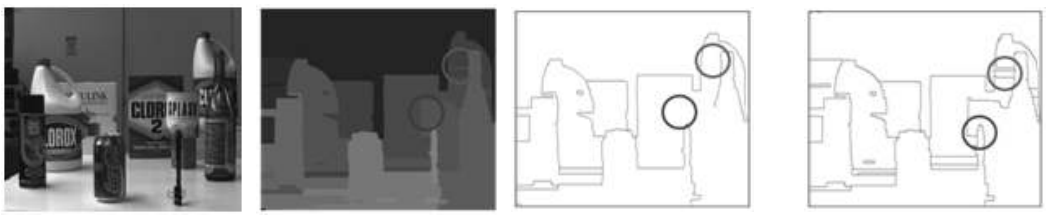
In (5), $M(x)$ is the mean and $S(x)$ is the standard deviation of disparity through the image x.

Through this way, threshold is determined automatically using ERD. As mentioned above, threshold is inverse proportion to ERD. In other words, if the ratio of discontinuity pixels is low, the threshold will be high and depth discontinuity pixels would be extracted less. On the contrary to this, if the ratio of discontinuity pixels is high, the threshold will be low and depth discontinuity pixels would be extracted more.

4. Experiment

4.1 The comparison of depth discontinuity of P2P and DIP

We extracted a depth discontinuity by P2P and DIP. (c) and (d) in Fig. 5 shows the result of P2P and DIP each.



(a) source of image (b) disparity map (c) depth discontinuity by P2P (d) depth discontinuity by DIP

[Fig. 5] The result of the depth discontinuity extracted by P2P and DIP

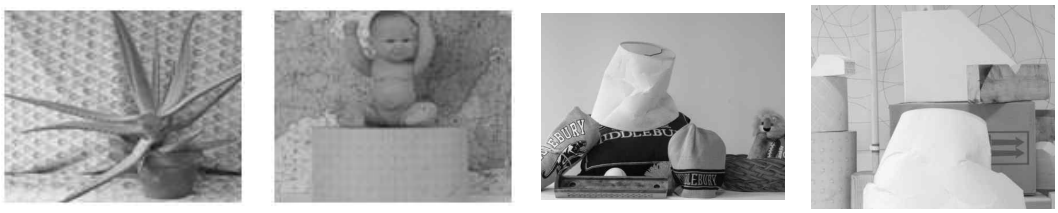
In Fig. 5, P2P cannot find the horizontal edge between box and can, in contrary, DIP can find it. As the result of the disparity map is clearer, depth discontinuities become more accurate. Therefore, we use disparity map made by Middlebury to learn the threshold of the DIP.

4.2 Learning the threshold of DIP

We present experimental results to show possibility for automatically learning the threshold used in DIP algorithm with closeness to the result of hand-tuned threshold. To obtain the more accurate learning data, we used disparity set in the Middlebury dataset.

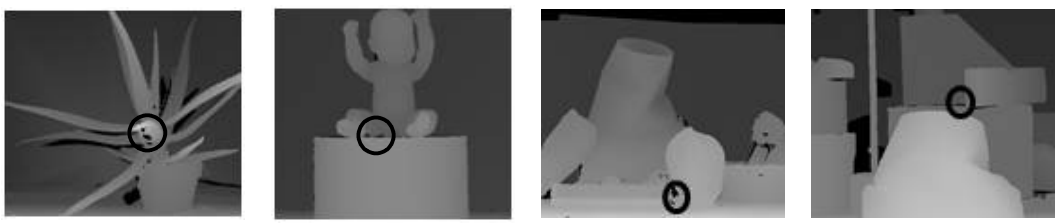
4.2.1 Environment of the experiment

Fig. 6 shows the left image set from the Middlebury dataset[10]. Fig. 7 shows the disparity map from the Middlebury dataset corresponding to the image of Fig. 6.



(a) Aloe (b) Baby (c) middle (d) lampshade

[Fig. 6] Left image of stereoscopy in Middlbury 3D database



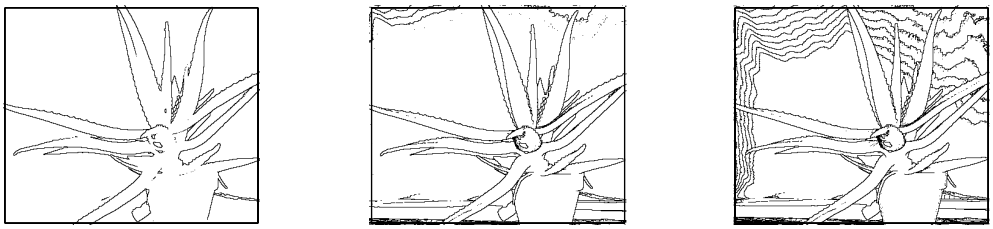
(a) Aloe (b) Baby (c) middle (d) lampshade

[Fig. 7] Disparity map

Disparity map in Fig. 7 is obtained by using technique of [10] and published in [11][12]. All of images have some occluded regions that are marked with circle. We filled this hole with the interpolation method that is presented in (4) before learning process.

4.2.2 The result of depth discontinuities using DIP

We extracted depth discontinuities from Aloe showed in Fig. 8. Because the threshold is various, the results are different as to the used threshold. Fig. 8 shows that there is the correlation between the threshold and the depth discontinuities extracted by DIP.



(a) RD=4, Th=255 (b) RD=8, Th=19 (c) RD=12, Th=17

[Fig. 8] The result of depth discontinuities by DIP as to various thresholds

In Fig. 8, the best result is the (a) which is extracted with RD 4, threshold 255. Depth discontinuities is expected less in (a) than in (c). As Fig. 8 shows, the result of the extraction is very different as to the value of the threshold. We need a way to learn the threshold automatically in real time to evaluate the visual fatigue.

4.2.3 Learning the threshold

We learned threshold with 10 images in Middlebury dataset. We infer a threshold by characteristics of the distribution of the depth map, that is, average and standard deviation. Table 1 shows average, standard deviation of the 10 images each and the result of the learned ERD and threshold.

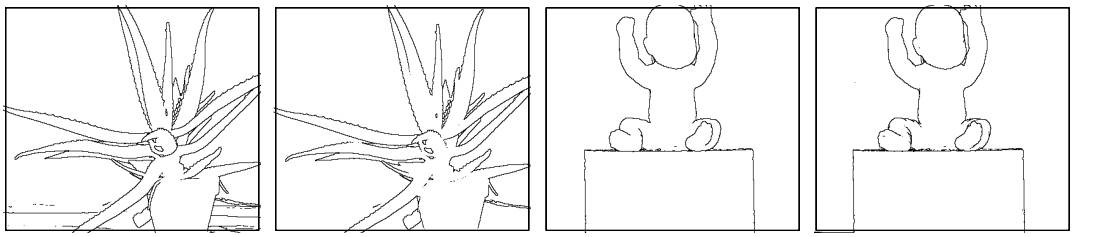
[Table 1] The result of the learned ERD

Name	Average (avg)	Standard Deviation	Hand-tuned ERD	Learned	
				ERD	Threshold
aloe	70.18	30.07	5.00	3.26	220
baby	83.03	32.18	2.00	2.92	20
bowling	115.53	59.46	2.00	2.35	55
flowerpots	124.35	54.34	3.00	2.03	25
lampshade	104.13	50.37	3.00	2.56	23
middle	93.98	48.95	3.00	2.83	20
monopoly	73.50	42.94	2.00	3.34	62
plastic	135.33	40.35	2.00	1.52	16
rocks	107.44	35.98	4.00	2.27	42
wood	121.33	39.38	2.00	1.91	22

After learning with average and standard deviation, we get an evaluation to decide the learned ERD, following (6).

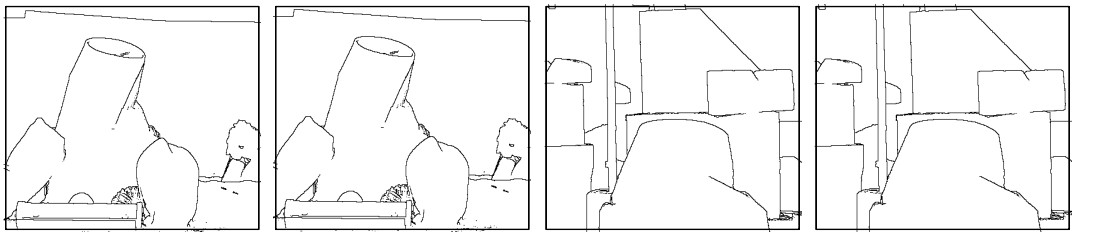
$$ERD = -0.029 * Avg + 0.014 * Std + 4.884 \quad (6)$$

Fig. 9 shows depth discontinuities by DIP used learned threshold for Aloe, Baby, middle, lampshade in Fig. 7.



(a-1)Hand tuned ERD 5 (a-2)Learned ERD 3.25 (b-1)Hand tuned ERD 2 (b-2)Learned ERD 2.92

(a) depth discontinuities of Aloe (b) depth discontinuities of Baby with learned ERD 2.92



(c-1) Hand tuned ERD 3 (c-2) Learned ERD 2.83 (d-1) Hand tuned ERD 3 (d-2) Learned ERD 2.56

(c) depth discontinuities of middle (d) depth discontinuities of lampshade

[Fig. 9] The result of depth discontinuities by DIP with learned threshold

Depth discontinuities of Fig. 9(a) are not sufficient compared with the hand-tuned result. From Fig. 9(b) to (d) is similar to the result of the hand-tuned threshold.

5. Conclusion

Depth discontinuities are important factor to let a human to make layers of the world. Extracting depth discontinuities from disparity map is similar to extraction of edges in 2D image. We employed DIP operator to extract depth discontinuities. However, DIP need a user interaction to decide a threshold. There are several applications that need to extract discontinuities like as evaluation of the 3D visual fatigue, tracking, detection, image retrieval. To apply depth discontinuities for these applications, there needs to obtain a threshold in real time based on the disparity map. We learned threshold using the Middlebury image sets and made evaluation to decide a threshold based on the average and standard deviation of the disparity map. The result of depth discontinuities using learned threshold is similar to that of hand-tuned threshold.

ISSN: 2383-5281 AJMAHS

References

[1] S. Birchfield , Depth and Motion discontinuities, Ph.D. dissertation, Stanford Univ., June **(1999).**

[2] J. J. Little and W. E. Gillett, Direct evidence for occlusion in stereo and motion, Imageand Vision Computing. **(1990)**, Vol 8, No.4, pp.328-340**.**

[3] P. S. Toh, and A. K. Forrest, Occlusion detection in early vision, Proceedings of the 3rd International conference on Computer Vision, **(1990)** pp.126-132**.**

[4] S. Birchfield and C. Tomasi, Depth discontinuities by pixel-to-pixel stereo, International Journal of Computer Vision.**(1999)**, Vol. 35, No. 3, pp. 269-293**.**

[5] S. Chen, B. Mulgrew, and P. M. Grant, A clustering technique for digital communications channel equalization using radial basis function networks, IEEE Trans. on Neural Networks. **(1993),** Vol. 4, pp.570-578**.**

[6] L. Wang, M. Gong, R. Yang, and D. Nister, High-quality Real-time Stereo using Adaptive Cost Aggregation and Dynamic Programming, 3D Data Processing, Visualization, and Transmission, Third International Symposium, **(2006),** pp. 798-805**.**

[7] Y. J. Ryoo and N. C. Kim, Valley operator extracting sketch features:DIP, Electron.Lett. **(1988),** Vol. 248, pp. 461-463**.**

[8] Y. D. Chun, S. Y. Seo, and N. C. Kim, Image retrieval using BDIP and BVLC moments, IEEE Transactions on circuits and systems for video technology. **(2003)**, Vol. 13, No. 9**.**

[9] A. Geiger, M. Roser, and R. Urtasun, Efficient large-scale stereo matching, Asian conference on computer vision, **(2010), V**ol. 6492, pp. 25-38**.**

[10] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003). **(2003)**, Vol. 1, pp. 195-202**.**

[11] D. Scharstein and C. Pal, Learning conditional random fields for stereo, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2007), **(2007) June**, Minneapolis, MN**.**

[12] H. Hirschmüller and D. Scharstein, Evaluation of cost functions for stereo matching, IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR2007), **(2007) June,** Minneapolis, MN**.**