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

ISO/IEC JTC1/SC29/WG11
MPEG2009/M16394
April 2009, Maui, USA

Source: GIST (Gwangju Institute of Science and Technology)
Status: Report
Title: EE1: Results of Depth Estimation on 'Pantomime' Sequence
Author: Cheon Lee and Yo-Sung Ho

1. Introduction

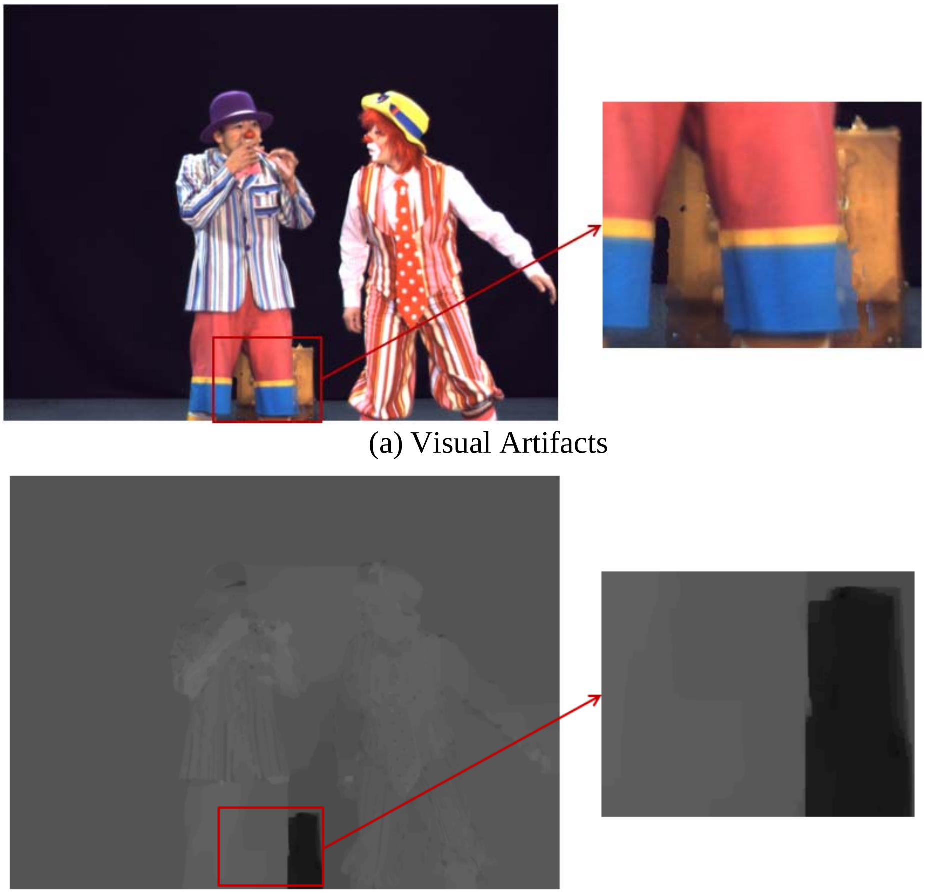
This document reports experimental results of the depth estimation software on 'Pantomime' sequence as a response to EE1 of 3D Video [1]. During the last Lausanne meeting, Nagoya Univ. has proposed a new algorithm using segmentation and block matching [2]. It is implemented in the updated DERS 3.0 software. In addition, ETRI has volunteered to provide the semi-automatic depth estimation reference software (SADERS). We, therefore, conducted experiments of depth estimation focusing on three additional methods. In order to find one best configuration, we tested only 20 frames from 80th frame to 99th on 'Pantomime' sequence. Then we generated depth maps for 500 frames. In this document, we describe all procedures and results.

2. Experiment of Semi-automatic Depth Estimation

During the Lausanne meeting, experts have agreed that the performance of depth estimation software is insufficient for 3D video system even though it has been improved significantly. In this sense, 3D video group called the 3D test materials including high quality depth videos [3]. ETRI has volunteered to provide a new semi-automatic method, and distributed SADERS on 2nd April 2009.

2.1. Visual Artifacts

The distributed software employs depth key-frames manually generated. Hence, we needed a depth key-frame. Before we generate it, we focused on the problem of depth map. 'Pantomime' sequence showed relatively stable performance for both depth estimation and view synthesis. Practically, the average PSNR value of synthesized results was higher than 34 dB [4], but it still has visual artifacts as shown in Fig. 1(a). Artifacts are induced by the erroneous and inconsistent depth estimation around depth discontinuity as shown in Fig. 1(b). When we make a key-frame manually, we tried to eliminate erroneous depth between trousers.



(a) Visual Artifacts

(b) Erroneous Depth Value

Fig. 1. Synthesis Artifacts on 'Pantomime' Sequence.

2.2. Generation of Key-frame

In order to generate a depth key-frame, we obtained a depth map generated by DERS 3.0. Then, we eliminated the unwanted depth values by hand from the estimated depth map. Figure 2 shows the original depth map by DERS 3.0 and the modified depth key-frame respectively. As you can see in the figure, we eliminated the depth values from issue region along the objects boundary. We made two key-frames for both left and right reference views.

2.3. Test on SADERS

SADERS uses three options;, 'ThresholdOfDepthDifference', 'MovingObjectsBSize' and 'MotionSearchBSize'. The first option 'ThresholdOfDepthDifference' is set to 15. Using other two options, we made nine combinations. After generating the depth videos, we synthesized two intermediate view images using the distributed view synthesis reference software 3 (VSRS 3), and then calculated their PSNR values. Table 1 shows the nine combinations of experiment and average PSNR values for 20 frames.

(a) Depth Map by DERS 3.0 (b) Manually Modified Depth Map

Fig. 2. Depth Key-frame for SADERS

Table. 1. Experimental Results on SADERS

Threshold Of Depth Difference	Motion Search BSize	Moving Objects BSize	Avg. PSNR of Left View (view 39)	Avg. PSNR of Right View (view 40)	Avg. PSNR
15	0	0	33.825 dB	33.336 dB	33.581 dB
	0	1	33.832 dB	33.343 dB	33.588 dB
	0	2	(invalid)		
	1	0	33.685 dB	33.153 dB	33.419 dB
	1	1	33.688 dB	33.155 dB	33.422 dB
	1	2	(invalid)		
	2	0	33.312 dB	32.867 dB	33.090 dB
	2	1	33.314 dB	32.865 dB	33.090 dB
	2	2	(invalid)		

Figure 3 demonstrates the results of depth maps and synthesized images. The generated depth map showed the effect of depth propagation; the background area is still black like depth key-frame as shown in Fig. 3(a), (b). Synthesized images in Fig. 3(c), (d) still have the visual artifacts.

(a) depth map of view 38 (b) depth map of view 41

(c) synthesized image of view 39 (d) synthesized image of view 40

Fig. 3. Results of SADERS for 98th frame

3. Experiment of DERS 3.0

Nagoya University proposed two new depth estimation methods: block matching and segmentation-based method. Assuming that the sub-pel precision and temporal enhancement are effective method, we conducted preliminary experiments to select a best parameter set for 20 frames. The used software is 'DERS 3.0', which is distributed on 2nd April 2009. Table 2 shows combinations of experiment and their results.

3.1. Test on Matching Block

The first added method is selection mode of matching block. When depth estimator calculates the matching cost, 3x3 block can be used instead of pixel matching. We checked the efficiency of the new method. Table 2 shows the experimental results for 20 frames. The PSNR values are the results of quality of synthesized images.

Table. 2. Experimental Results on Matching Block

	Pixel Matching		3x3 Block Matching	
Viewpoint	View 39	View 40	View 39	View 40
Synthesized Results (dB)	34.342	33.815	34.593	33.985
Average PSNR (dB)	34.079		34.289	

3.2. Test on Segmentation

Another added method is segmentation-based depth estimation. The updated DERS software exploits three segmentation methods: mean shift algorithm, phyramid segmentation, and K mean clustering. Table 3 shows the results of depth estimation. The best option was K mean clustering, of which average PSNR was about 33.4 dB. The worst one was mean shift algorithm. Comparing with the previous method, this segmentation method was not efficient. Since 'Pantomime' sequence has large black background, it could affect the segmentation method worse.

Table. 3. Experimental Results on Segmentation-based Method

	Mean Shift		Phyramid		K mean Clustering	
Viewpoint	View 39	View 40	View 39	View 40	View 39	View 40
Results (dB)	17.039	16.844	32.142	30.829	33.913	32.893
Avg. PSNR (dB)	16.942		31.4855		33.403	

(a) depth map and synthesized image using mean shift algorithm

(b) depth map and synthesized image using phyramid algorithm

(c) depth map and synthesized image using K mean clustering

Fig. 4. Results of SADERS for 98th frame

4. Best Configuration of Depth Estimation and Experimental Results

Based on testing 20 frames from 80th frame, we chose one best configuration for depth estimation. Table 4 describes the configuration that we used. The average PSNR of both two intermediate view videos is 34.7 dB.

Table. 4. Best Configuration and Experimental Results

<table>
<tr><th colspan="2"></th><th colspan="2">Left View</th><th colspan="2">Right View</th></tr>
<tr><td rowspan="11">Depth Estimation</td><td>View Position</td><td colspan="2">View 38</td><td colspan="2">View 41</td></tr>
<tr><td>Precision</td><td colspan="2">Half-pel</td><td colspan="2">Half-pel</td></tr>
<tr><td>SearchLevel</td><td colspan="2">Half-pel</td><td colspan="2">Half-pel</td></tr>
<tr><td>BaselineBasis</td><td colspan="2">1</td><td colspan="2">1</td></tr>
<tr><td>Filter</td><td colspan="2">MPEG-4 AVC 6-tap</td><td colspan="2">MPEG-4 AVC 6-tap</td></tr>
<tr><td>SmoothingCoefficient2</td><td colspan="2">3.00</td><td colspan="2">3.00</td></tr>
<tr><td>MatchingMethod</td><td colspan="2">Disparity-based</td><td colspan="2">Disparity-based</td></tr>
<tr><td rowspan="2">TemporalEnhancement</td><td colspan="2">On</td><td colspan="2">On</td></tr>
<tr><td>Threshold</td><td>1.00</td><td>Threshold</td><td>1.00</td></tr>
<tr><td>MatchingBlock</td><td colspan="2">3x3 Block matching</td><td colspan="2">3x3 Block matching</td></tr>
<tr><td>ImageSegmentation</td><td colspan="2">Off</td><td colspan="2">Off</td></tr>
<tr><td rowspan="3">View Synthesis</td><td>View Position</td><td colspan="2">View 39</td><td colspan="2">View 40</td></tr>
<tr><td rowspan="2">Average PSNR (dB)</td><td colspan="2">34.900</td><td colspan="2">34.500</td></tr>
<tr><td colspan="4">34.700</td></tr>
</table>

Although above configuration shows the best average PSNR value, the synthesized image still has some visual artifacts as shown in Fig. 5. The shape of bag located at the backside of the left clown has been distorted. This artifact was found from the previous method as wells. As a result, 3x3 block matching method improves the depth quality, but it does not eliminate the visual artifacts.

(a) synthesized image for view 39

(b) synthesized image for view 40

Fig. 5. Synthesis Results of 100th Frame

5. Conclusion

In this document, we reported the results of depth estimation software on ‘Pantomime’ sequence. Semi-automatic depth estimation reference software (SADERS) provided by ETRI is a new approach to enhance the performance of depth estimation, but the performance was not good enough. The segmentation based method showed interesting results, but the quality of depth map was not better than the previous method. On the other hand, 3x3 matching block method improved the quality of depth estimation. Therefore, we conducted the experiment using 3x3 matching block, temporal enhancement, and half-pel precision.

6. Acknowledgements

This work was supported in part by ITRC through RBRC at GIST (IITA-2009-C1090-0901-0017).

7. References

[1] ISO/IEC JTC1/SC29/WG11 “Description of Exploration Experiments in 3D Video Coding”, N10360, February 2009.
[2] ISO/IEC JTC1/SC29/WG11 “Depth Estimation Reference Software (DERS) with Image Segmentation and Block Matching”, M16092, February 2009.
[3] ISO/IEC JTC1/SC29/WG11 “Call for 3D Test Material: Depth Maps & Supplementary Information”, N10359, February 2009.
[4] ISO/IEC JTC1/SC29/WG11 “EE2: View Synthesis Results on 'Pantomime’ Sequence”, M16061, February 2009.