PAPER

# Superpixel Based Depth Map Generation for Stereoscopic Video Conversion

**Jie FENG**[†a)], ***Member***, **Xiangyu LIN**[†b)], **Hanjie MA**[†], ***and*** **Jie HU**[†], ***Nonmembers***

**SUMMARY** In this paper, we propose a superpixel based depth map generation scheme for the application to monoscopic to stereoscopic video conversion. The proposed algorithm employs four main processes to generate depth maps for all frames in the video sequences. First, the depth maps of the key frames in the input sequence are generated by superpixel merging and some user interactions. Second, the frames in the input sequences are over-segmented by Simple Linear Iterative Clustering (SLIC) or depth aided SLIC method depending on whether or not they have the depth maps. Third, each superpixel in current frame is used to match the corresponding superpixel in its previous frame. Finally, depth map is propagated with a joint bilateral filter based on the estimated matching vector of each superpixel. We show an improved performance of the proposed algorithm through experimental results.

***key words:*** *depth map, superpixel, depth generation, 2D to 3D conversion*

## 1. Introduction

Interest in three-dimensional (3D) visualization and free viewpoint television is becoming stronger and stronger in recent years. However, 3D technology has not been very successful in commercial applications due to several problems. Lack of 3D video content is one of the biggest bottlenecks for the entire 3D industry. Stereoscopic conversion from 2D video is a means to satisfy the need for 3D video content. Most recent solutions start by extracting the depth maps from the original 2D video sequences. Then the 3D video contents can be generated by the depth image based rendering (DIBR) [1] technology.

Obviously, the quality of depth map is critical in successfully rendering 3D views. Most existing algorithms can generate an acceptable depth map in an automatic way [2], [3]. They explore various depth cues, such as motion parallax, texture gradients, linear perspective, relate height and geometric information, etc. However, the quality of the generated depth map is still not good enough since the existing computer vision algorithms at present are not able to infer accurate depth due to the complicated structures in common videos. As a result, automatic depth map generation schemes can only deliver a limited 3D perception for 3D video applications.

Semi-automatic 2D-to-3D conversion, which can improve the quality of generated depth maps, is attracting the attention of more and more researchers. It creates provide accurate depth maps for key frames by utilizing the high-level knowledge of humans and generates good-quality depth maps for non-key frames by computer vision algorithms. As a result, it can balance the conversion quality against the efficiency. Recently, some approaches have been proposed for semi-automatic depth map generation for 2D-to-3D conversion. Phan et al. [4] proposed a scheme which generates depth maps via Random Walks and Graph Cuts and combines the two maps into a single composite map. The study in [5] first over-segments and manually annotates the original image, then detects the edge and T-junction, and finally obtains the depth map by depth propagation and post-processing. Varekamp et al. [6] proposed a depth propagation method based on bilateral filtering through a block-based motion compensation algorithm. Wu et al. [7] use bi-direction optical flow and Mean Shift algorithm to extract foreground object and track depth information for non-key frames. Rzeszutek et al. [8] use user-defined strokes to label a number of key frames and then perform Random Walks segmentation framework for the depth map generation. Lie et al. [9] proposed a non-key frame depth propagation process which uses depth motion compensation and post trilateral filtering for a better depth contour of the dynamic foreground objects. Cao et al. [10] use a few user operations to segment multiple objects and assign proper depth to each object for key frames. Then, for non-key frames, the depth maps are generated automatically by a disparity propagation algorithm. The approaches in [5], [6] mainly aim at image semi-automatic depth generation. While approaches in [7]–[10] mainly focus on the video semi-automatic depth propagation based on the depth maps which have already been generated for the key frames with computational complexity. Considering that most existing methods only address segmentation quality or depth generation efficiency, our proposed scheme tries to strike a balance between them.

In this paper, we proposed a novel semi-automatic depth map generation scheme. The rest of this paper is organized as follows. Section 2 details the proposed scheme. The experiments and results are presented in Sect. 3. Finally, the paper is concluded in Sect. 4.

## 2. The Proposed Scheme

Our proposal is based on segmenting each video frame into superpixels of arbitrary shape and then generating the depth map of each frame via superpixel merging and match-

Manuscript received October 16, 2013.
Manuscript revised March 16, 2014.
[†]The authors are with the School of Information Science and Technology, Zhejiang Sci-Tech University, China.
a) E-mail: arlose@zstu.edu.cn
b) E-mail: linxiangyu@zstu.edu.cn (Corresponding author)
DOI: 10.1587/transinf.E97.D.2131

ing techniques. The depth map generation approach consists of the following steps: (1) Key frame's depth map is generated by superpixel merging and some user interaction. (2) Key frame or the frame with generated depth map is over-segmented by depth aided superpixel segmentation method and other frame is segmented by the SLIC method with the same parameters. (3) A motion vector for each segmented superpixel is estimated by a color based region matching algorithm so as to find the corresponding depth aided superpixel in the processed frame. (4) Depth map is propagated with a joint bilateral filter based on the matching vector of each superpixel.

### 2.1 Key Frame Depth Map Generation

For the most general 2D video sequences or movies which have no corresponding depth map images, at the authors' best knowledge, the best way of generating a good depth map is to adopt the semi-automated technique. Some frames in the sequences are selected as the key frames whose depth maps are generated via human interaction so as to achieve the best quality. Other frames are treated as non-key frames whose depth maps are propagated by the generated depth maps of key or non-key frames.

Therefore, as the first step, we apply a key-frame extraction method as described in [14], which includes block-based histogram difference shot segmentation and cumulative occlusion based key frame selection. It can guarantee relatively fewer depth propagation errors in all frames and is more robust than the traditional temporal interval-based method.

As shown in Figs. 1 and 2, the extracted key frames' depth maps are generated by superpixel merging and some user interactions. First the key frames are over-segmented by the Simple Linear Iterative Clustering (SLIC) [11] method, which is a good implementation of the superpixel algorithm [12]. It can output a desired number of regular, compact regions, which are called superpixels, with a low computational overhead. The pixels in one superpixel are usually most likely uniform in color and texture as can be seen in Fig. 2 (b). The average pixel number in a superpixel is a very important parameter in the SLIC segmentation procedure and is defined as superpixel's size $\delta$ in this paper.

After the over-segmented stage, most of the small superpixels are merged according to their similarity. We use the color feature of each superpixel as the similarity measure characteristics, and calculate the average R, G, B color value of all the pixels in each superpixel region as $R_a$, $G_a$ and $B_a$, the color difference $Diff_{color}(i, j)$ between the $i$th superpixel and $j$th superpixel in a frame can be calculated as follows:

$$Diff_{color}(i, j) = |R_a(i) - R_a(j)| + |G_a(i) - G_a(j)| + |B_a(i) - B_a(j)| \quad (1)$$

If two neighboring superpixel's color difference value $Diff_{color}$ is less than a predefined threshold $T_{color}$, then these two superpixels are merged together into one larger region. The bigger the $T_{color}$ is, the more regions are merged, the less user interactions are required; otherwise, the smaller the $T_{color}$ is, the less regions are merged, the more user interactions are required, while the more accurate segmentation results are achieved. After the merging process, the number of regions in a frame will be much smaller then that of the initial superpixel regions, as shown in Fig. 2 (c). Next, a few user interactions are imposed on the regions. The user can scribble on the regions including merged regions and single superpixel with a mouse to label them as different layers with the aid of his own experience. Each layer usually contains one object. With the help of the previous merging process, the number of the user's scribbles is greatly decreased down to about 5 to 30 for one key frame. Sometimes scribbles are required to distinguish pixels from different objects with similar color. Figure 2 (d) shows the result of this step where different layers are indicated by different alpha channel masks.

In the following step, we also need some user interactions to assign the depth value of one frame in a layer by layer way. Only three parameters, including the maximum depth value of the layer, the minimum depth value of

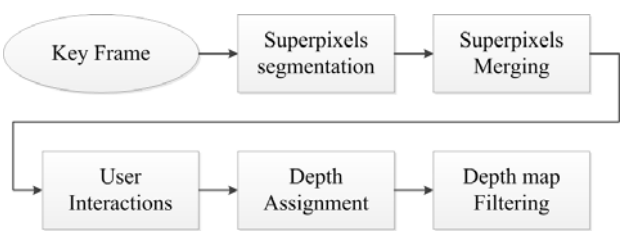


**Fig. 1** Key frame's depth map generation process.

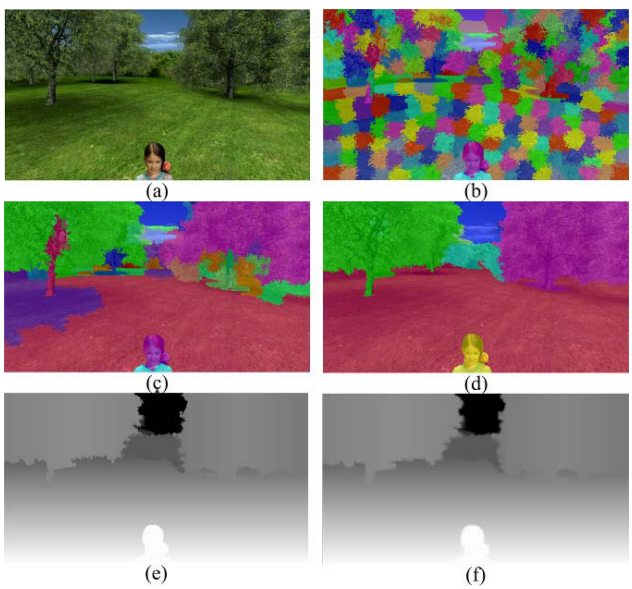


**Fig. 2** Key frame's depth map generation process. (a) the original frame. (b) result of over-segmenting into superpixels. (c) segmentation result after superpixels merging. (d) segmentation result after user interactions. (e) depth assignment result. (f) depth result after Gaussian filtering.

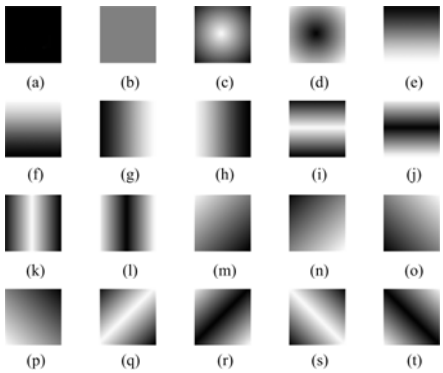


**Fig. 3** Key frame's depth assignment templates.

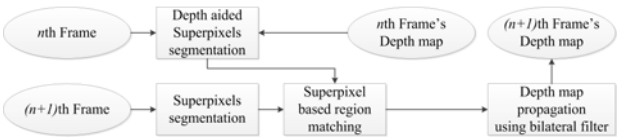


**Fig. 4** Depth map propagation process.

the layer and the depth models index, are needed to set for each layer. There are 20 depth templates used for this key frame's layer depth assignment process. They can be set up in advance and are shown in Fig. 3. In these templates, the whiter the color, the larger the depth value. For example, in Fig. 3 (e), the color in this depth model is whiter and whiter from top to bottom, that means the depth values in the layer applying this model are larger and larger from top to bottom. The depth values of the pixels from the top line in this layer are set as the minimum depth value while the depth values of the pixels from the bottom line are set as the maximum depth value. The depth values in the other lines are set according to a linear function. Moreover, to reduce the disocclusion artifacts in the rendering process [13], the generated depth map of the key frame is smoothed by a Gaussian filter to obtain the final one as shown in Fig. 2 (f).

### 2.2 Depth Aided Superpixel Segmentation

The non-key frames' depth maps are propagated frame by frame in an automatic way, as denoted in Fig. 4.

In traditional SLIC superpixel image segmentation algorithm, only the color and the position's correlation information are considered to classify different regions. In some cases, however, the pixels with similar color are not belonging to the same object, such as the part marked by the blue circle in Fig. 5. In this case, other information is needed to distinguish between these pixels with similar color. Depth information is a good choice to distinguish one object from another under the assumption that different objects often have different depth values. Based on this observation, we add depth information in the SLIC algorithm to get a better segmentation result.

The Euclidean distance $D_s$ in the 5D space is used as the distance measure in SLIC algorithm. $D_s$ is defined as

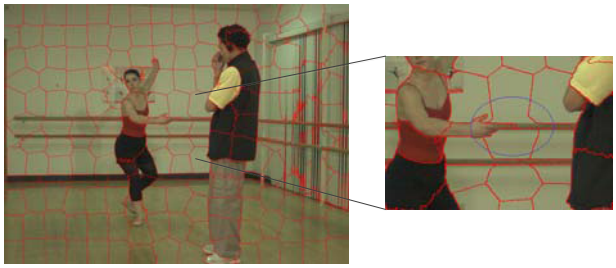

**Fig. 5** Traditional SLIC superpixel segmentation to different objects with similar color.

follows [11]:

$$\begin{aligned} d_{lab} &= \sqrt{(l_k - l_i)^2 + (a_k - a_i)^2 + (b_k - b_i)^2} \\ d_{xy} &= \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2} \\ D_s &= d_{lab} + \frac{m}{S} d_{xy} \end{aligned} \tag{2}$$

where $l$, $a$, $b$ are the Lab color space value, $x$ and $y$ are the x-coordinate and y-coordinate value. The subscript $k$ and $i$ represent the different pixels. $D_s$ is the sum of the lab distance $d_{lab}$ and the normalized position coordinate distance $d_{xy}$. $S$ is the grid interval. The variable $m$ is used to control the compactness of a superpixel. The greater the value of $m$, the more spatial proximity is emphasized and the more compact the cluster.

We take the depth information as the sixth dimension and modified the Euclidean distance $D_s$:

$$\begin{aligned} d_d &= |d_k - d_i| \\ D_s &= d_{lab} + \frac{m}{S} d_{xy} + d_d \end{aligned} \tag{3}$$

Then the simple linear iterative clustering algorithm is performed as described in [11]. The only difference is the vector used for computing the image gradients. In our algorithm, the depth information is taken into account and added to the *lab* vector.

In our scheme, the depth aided superpixel segmentation (D-SLIC) described above is only applied to the key frames and the non-key frames with depth map. Conversely, the traditional SLIC algorithm is applied to the other non-key frames which don't have depth maps yet.

### 2.3 Superpixel Region Matching

For depth propagation, we consider a situation of two adjacent frames in which the prior has a generated depth map while the latter hasn't. We segmented these frames by the D-SLIC as described in Sect. 2.2 and the SLIC [11] using the same parameters, respectively.

Then for each superpixel region in the latter frame, we calculate the difference value between the current superpixel and the reference superpixel in the window size of M × M of the prior frame. The difference value can be calculated as follows:

$$\begin{aligned} Diff(i,j) = &\left|R(i) - R_{ref}(j)\right| + \left|G(i) - G_{ref}(j)\right| \\ &+ \left|B(i) - B_{ref}(j)\right| \\ &+ \lambda \times \left(\left|X(i) - X_{ref}(j)\right| + \left|Y(i) - Y_{ref}(j)\right|\right) \end{aligned} \tag{4}$$

where $Diff(i, j)$ represents the difference value between the $i$th superpixel in the latter frame and the $j$th superpixel in the prior frame. $R(i)$, $G(i)$, $B(i)$ are the average R, G, B color value of the pixels in the $i$th superpixel and $X(i)$, $Y(i)$ are the center coordinates value of the $i$th superpixel in the latter frame. Accordingly, $R_{ref}(j)$, $G_{ref}(j)$, $B_{ref}(j)$ are the average R, G, B color value of the pixels in the $j$th superpixel and $X_{ref}(j)$, $Y_{ref}(j)$ are the center coordinates value of the $j$th superpixel in the prior frame. $\lambda$ is the weighting factor for spatial proximity matching. The spatial influence increases with increasing $\lambda$. The color values of each pixels in a superpixel tend to be similar. In our evaluations, it has been found that the color factor is more important than the position factor, thus, $\lambda$ is set to 0.5 are selected to represent a typical result in our experiment.

Next, in the window size of M×M, we select the $j_{min}$th superpixel which has the minimum value of $Diff(i, j)$ as the matching superpixel to the $i$th superpixel of the latter frame.

$$j_{min} = \arg\min_{j} Diff(i, j) \tag{5}$$

Then the matching vector $MV_x(i)$ and $MV_y(i)$ is calculated as follows:

$$\begin{aligned} MV_x(i) &= X(i) - X(j_{min}) \\ MV_y(i) &= Y(i) - Y(j_{min}) \end{aligned} \tag{6}$$

### 2.4 Depth Map Propagation Using Bilateral Filter

Bilateral filter which combines domain filtering and range filtering is very effective for depth map propagation [6]. In [6], the geometric distance and the color difference are used to determine the Gaussian weights in bilateral filter. This can't distinguish different objects with similar color values. In order to eliminate this unfavorable effect, we use the matching vector of each superpixel as a parameter to calculate the depth value in the latter frame:

$$D^i(x,y) = \frac{\sum_{m=-N}^{N}\sum_{n=-N}^{N} w^i(m,n) D_{ref}(x + MV_x(i) + m, y + MV_y(i) + n)}{\sum_{m=-N}^{N}\sum_{n=-N}^{N} w^i(m,n)}$$

$$w^i(m,n) = \begin{cases} 2^{\Delta I^i(m,n)/0.125}, & -N \le m,n \le N \\ 0, & otherwise \end{cases}$$

$$\Delta I^i(m,n) = \sum_{c=r,g,b} \left|I^c(x,y) - I^c_{ref}(x + MV_x(i) + m, y + MV_y(i) + n)\right| \tag{7}$$

where $D^i(x, y)$ is the estimated depth value at pixel $(x, y)$ in the $i$th superpixel and $D_{ref}(x, y)$ is the depth value at pixel $(x, y)$ from the prior reference frame. The weights $w^i(m, n)$ depend on the color difference between pixel $(x, y)$ in the latter frame and the neighbor pixels $(x + MV_x(i), y + MV_y(i))$ in the prior frame. $N$ is the filter window size. The bigger the $N$ value is, the more pixels are filtered, and the more complexity of the algorithm is. To balance the conversion quality against the efficiency, we set $N$ to 9, the same value as that in reference [6].

We go through all the pixels in the latter frame using the filter and finally obtain the depth map. The depth map are used as the reference for propagating the next frame's depth map as described above. Finally, all depth maps in the sequence are generated.

## 3. Experiment

The proposed depth map generation algorithm is implemented using VS2010 and is tested on an Intel® Core™ i5 CPU@2.4 GHz personal computer with 4 GB memory. To verify our proposed algorithm, various test sequences including well-known video plus depth sequences are used for objective and subjective evaluations.

First, we investigate the objective quality of the generated depth map by the Peak Signal-to-Noise Ratio (PSNR) results using *Ballet* and *Breakdancers* sequences [15], which are provided by Microsoft Research (MSR) group with both color and associated depth maps at 1024 × 768 pixels resolution. Each sequence consists of 100 frames at 15 frames per second with 8 different camera views. Every 20 frames in each sequence are treated as the key frames while the depth map of other non-key frames are chosen for reference purposes. The PSNR of the depth map can be calculated via the mean squared error (MSE) between the reference depth map and propagated depth map of the other 95 non-key frames, respectively, for each sequence. Figure 6 shows the PSNR comparison results. It can be confirmed from Fig. 6 that the proposed algorithm (with superpixel size of 64) are always superior to the algorithm in [9] with a gain of up to 4.2 dB and can achieve the average PSNR gains of 1.91 dB and 1.12 dB for *Ballet* and *Breakdancers* respectively.

Next, we exhibit the generated depth map of the previous sequence *Ballet* and sequences *Philips-the-3D-experience*. Sequence *Philips-the-3D-experience*, which

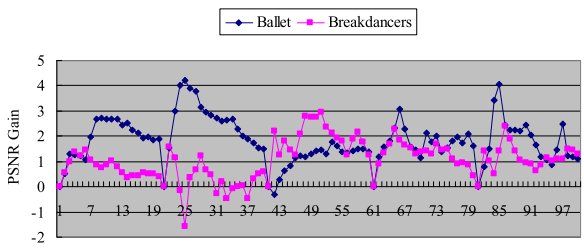


**Fig. 6** Propagated depth maps' PSNR gain.

has the feature of large displacement and textureless regions, is from the Philips WoWvx web site [16]. It has 61 frames from frame number 380 to 440. The depth maps with the number of 380, 400, 420 and 440 are already provided. Based on these provided or generated depth maps of the key frames in these sequences, we propagated the rest depth maps of the non-key frames. Some results of these snap-shots are shown from Figs. 7 and 8. The first line of Figs 7 and 8 are the original images, while the second and the last line are the associated depth maps generated by the algorithm in [9] and the proposed algorithm respectively. As it is shown in these images, the proposed depth generation algorithm preserves more carefully the depth discontinuities and the contours of the objects. In Fig. 8, for example, the outline of the girl's body is very blurred in the depth map of the second line which is generated by the algorithm in [9]. On the other hand, our proposed method preserves the body boundaries very well. Our method might slightly blurred contours due to errors in the segmentation (the trees in the background) but, in overall, it presents better subjective quality than [9].

**Fig. 7** The original images and associated depth maps generated by the algorithm in [9] and the proposed algorithm of the sequence *Ballet*.

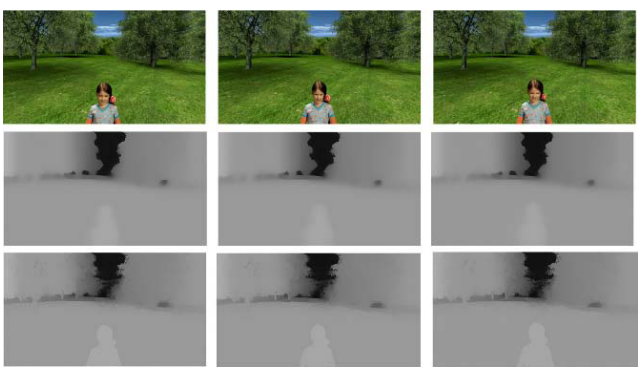

**Fig. 8** The original images and associated depth maps generated by the algorithm in [9] and the proposed algorithm of the sequence *Philips-the-3D-experience*.

Moreover, the parameter $\delta$ of the superpixel's size in segmentation procedure influences the quality of generated depth map and the consuming time of propagation process. We select six typical values at 16, 32, 64, 128, 256 and 512 for $\delta$ and show the subjective quality in Fig. 9 and the ratio of the propagation consuming time in Fig. 10. Table 1 lists the PSNR gains against the algorithm in [9] with different $\delta$ value. Table 1 indicates that the PSNR gains are similar with $\delta$ less to 64, when $\delta$ is larger than 64, the PSNR gains are decreasing. By subjective observation, it can also be confirmed from Fig. 9 that the smaller the superpixel's size is, the more detailed depth maps can be produced.

The algorithm in reference [9] consumes about 100 s per frame. While the base consuming time of the proposed algorithm with $\delta$ equal to 64 is 98.9 s per frame including 60.8 s for over-segmentation and 39.1 s for superpixel matching and depth map propagation in the same computing environment.

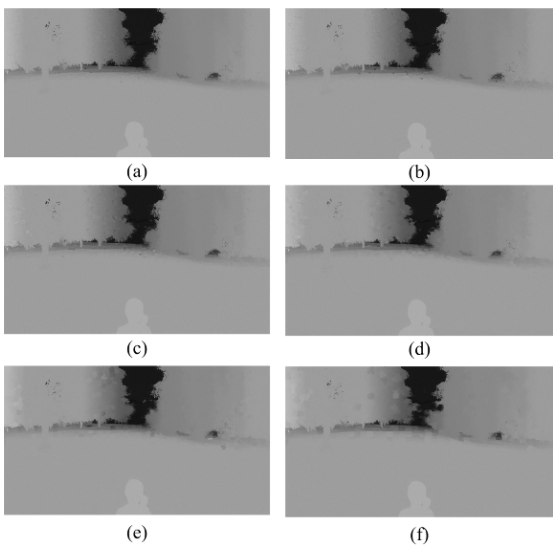


**Fig. 9** Depth propagation result with different number of pixels in a superpixel. (a) 16 (b) 32 (c) 64 (d) 128 (e) 256 (f) 512.

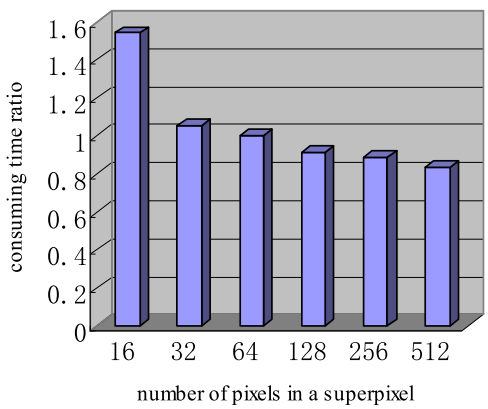


**Fig. 10** The ratio of the propagation consuming time with different superpixel's size.

**Table 1** Average PSNR gain (dB) with different $\delta$.

| $\delta$ | 16 | 32 | 64 | 128 | 256 | 512 |
|---|---|---|---|---|---|---|
| PSNR gain *for Ballet* | 1.95 | 1.92 | 1.91 | 1.69 | 1.60 | 1.43 |
| PSNR gain *for Philips-the-3D-experience* | 1.11 | 1.09 | 1.12 | 1.03 | 0.91 | 0.84 |

## 4. Conclusion

In this work, we have proposed a depth map generation algorithm for converting monoscopic video to stereoscopic 3D video. To improve the quality of the generated depth maps, the frames are over-segmented and the segmented regions are matched by superpixel matching algorithm. Depth maps are propagated using bilateral filter according to the matching vectors. Experimental results illustrate the robustness and effectiveness of this approach. We will consider a bi-directional region matching algorithm for accurate depth propagation in the future work.

## Acknowledgments

This work is supported by Natural Science Foundation of Zhejiang Province under Grant No.Y1100632 and the Zhejiang Province Public Technology Research Program under Grant No.2013C31021.

### References

[1] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," Proc. SPIE, vol.5291, pp.93–104, 2004.
[2] W.J. Tam and L. Zhang, "3D-TV content generation: 2D-to-3D conversion," 2006 IEEE International Conference on Multimedia and Expo (ICME), pp.1869–1872, 2006.
[3] L. Zhang, C. Vázquez, and S. Knorr, "3D-TV content creation: Automatic 2D-to-3D video conversion," IEEE Trans. Broadcast., vol.57, no.2, pp.372–383, 2011.
[4] R. Phan, R. Rzeszutek, and D. Androutsos, "Semi-automatic 2D to 3D image conversion using a hybrid random walks and graph cuts based approach," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp.897–900, 2011.
[5] X. Yan, Y. Yang, G. Er, and Q. Dai, "Depth map generation for 2D-to-3D conversion by limited user inputs and depth propagation," 3DTV Conference: The True Vision — Capture, Transmission and Display of 3D Video (3DTV-CON), pp.1–4, 2011.
[6] C. Varekamp and B. Barenbrug, "Improved depth propagation for 2D to 3D video conversion using key-frames," 4th European Conference on Visual Media Production (IETCVMP), pp.1–7, 2007.
[7] Y. Wu, P. An, P. Wang, and Z. Zhang, "Stereoscopic video conversion based on depth tracking," IEEE International Conference on Signal Processing (ICSP), pp.1190–1193, 2010.
[8] R. Rzeszutek, R. Phan, and D. Androutsos, "Semi-automatic synthetic depth map generation for video using random walks," IEEE International Conference on Multimedia and Expo (ICME), pp.1–6, 2011.
[9] W. Lie, C. Chen, and W. Chen, "2D to 3D video conversion with key-frame depth propagation and trilateral filtering," Electron. Lett., vol.47, no.5, pp.319–321, 2011.
[10] X. Cao, Z. Li, and Q. Dai, "Semi-automatic 2D-to-3D conversion using disparity propagation," IEEE Trans. Broadcast., vol.57, no.2, pp.491–499, 2011.
[11] A. Radhakrishna, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "Slic superpixels," Technical Report 149300, EPFL, 2010.
[12] A. Radhakrishna, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," IEEE Trans. Pattern Anal. Mach. Intell., vol.34, no.11, pp.2274–2282, 2012.
[13] W.J. Tam, G. Alain, L. Zhang, and T. Martin, "Smoothing depth maps for improved stereoscopic image quality," Proc. SPIE, vol.5599, pp.162–172, 2004.
[14] D. Wang, J. Liu, J. Sun, W. Liu, and Y. Li, "A novel key-frame extraction method for semi-automatic 2D-to-3D video conversion," IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), pp.1–5, June 2012.
[15] C. Zitnick, S. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," ACM Siggraph and Trans. Graph, vol.23, pp.600–608, 2004.
[16] "WoWvx," [Online]. Available: http://www.WoWvx.com/, accessed March 6, 2013.

**Jie Feng** was born in Jinzhou, Liaoning, China, in 1980. He received the B.Sc. and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 2003 and 2009, respectively. Since 2009, he has been joining with Zhejiang Sci-Tech University, Hangzhou, China. His major research field is video coding, video analysis and 2D to 3D video conversion.

**Xiangyu Lin** was born in Ningbo, Zhejiang, China, in 1983. He received the B.Sc. and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 2006 and 2012, respectively. Since 2012, he has been joining with Zhejiang Sci-Tech University, Hangzhou, China. His major research field is video coding and video quality evaluation.

**Hanjie Ma** was born in Huangshi, Hubei, China, in 1982. He received the B.Sc. and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 2003 and 2009, respectively. Since 2011, he has been joining with Zhejiang Sci-Tech University, Hangzhou, China. His major research field is video coding and video transmission.

**Jie Hu** was born in Ningbo, Zhejiang, China, in 1977. She received the B.Sc. degree from Zhejiang University, Hangzhou, China, in 2002. Since 2002, she has been joining with Zhejiang Sci-Tech University, Hangzhou, China. Her major research field is video coding, digital image processing.