Voting-Based Depth Map Refinement and Propagation for 2D to 3D Conversion

Yu-Hsiang Chiu* , Ming-Sui Lee* and Wei-Kai Liao†

* National Taiwan University, Taipei, R.O.C

E-mail: {d00944012,mslee}@csie.ntu.edu.tw

† The White Rabbit Entertainment, Taipei, R.O.C

E-mail: wliao@twrglobal.com

***Abstract*—In this paper, a voting-based filter which is capable of enhancing the quality of depth map sequence and interpolates missing frames is proposed. The main concept is that if the information in the filter window is not consistent, then only the multitude decides the output. The minority will be considered as outliers. Compare to other depth map refinement method using joint bilateral filter, the outlier detection of the proposed scheme ensures that only appropriate information is involved so that the halo effects can be avoided. Moreover, in order to refine a sequence of depth maps, a space-time filtering extension is proposed. This extension refines each depth map according to the information of several adjacent frames rather than only one frame. As a result, the proposed method is capable of interpolating missing depth maps of a sequence and the errors on the depth maps are successfully reduced. The experimental results demonstrate that the voting-based filter not only provides a depth map sequence with good quality and temporal consistency but also provides several post-processing for depth maps in 2D to 3D conversion due to its flexibility.**

I. Introduction

Depth map is a two-dimensional array which contains per-pixel depth information for each pixel. It plays an important role in image and video 2D to 3D conversion. Depth-image-based rendering(DIBR)[7] is a method which synthesize virtual views according to a 2D image and its depth map. This technique is widely used in 3D movies and 3D TV industries. The quality of depth map influences the 3D visualization of images and videos which are converted by DIBR. There are two main factors to evaluate the quality of a depth map for 2D to 3D conversion. The first one is the precision of edge location. If the edges on depth maps are not aligned to the edges on the image, the foreground objects will not be spread from background correctly. The second one is temporal consistent. It effects the converted 3D video is stable or not.

Depth map refinement is a process that reduces the errors of depth map form noise, poor defined edges or other undesired artifacts. Compare to other approaches, the proposed method is robust in dealing with poor defined edges and fixing pixels whose depth information is incorrect. It worth mention the proposed filter is a space-time filter which could be used to refine a depth map sequence. There are two profits we can take from this approach. The first one is that the wrong depth values might take a chance of being fixed according to the information from neighbor frame. Another is that the depth value of specific pixel in different frames will be decided by almost same group of pixels, it ensures the filter output is temporal consistent. Another extended usage of space-time filter is interpolating missing frames. This function could achieve depth propagation.

The main function of depth propagation is synthesizing depth map sequence by only given key frames. Some old movies or TV programs are not captured by depth sensor, so the depth maps of those video is not available. In this situation, manually editing depth maps frame by frame is labor-intensive and difficult to maintain the temporal coherence. The main usage of depth propagation is producing depth maps for the video which do not have depth maps. It is an important application in 2D to 3D conversion.

The main contribution of this paper includes: (1) A depth map refinement procedure is proposed. This procedure aligns the edges of depth map with the edges in original image. (2) Based on the aiding of optical-flow, the proposed filter fixes depth map error by using the information from neighboring frames. (3) The refinement procedure and depth propagation scheme provide an easy way to manually create a depth map sequence.

This paper is organized as follows: Some related works are presented in Sec II. The details of the proposed method are described in Sec III. The experimental results of proposed method are demonstrated in Sec IV. The summarization of this paper is provided in Sec V.

II. Related Work

A. Edge Preserving Filter

The concept of edge preserving filter comes from bilateral filter[4]. The bilateral filter outputs a pixel as a weighted average of its neighboring pixels. It smoothes an image but preserves edges so that it has been widely used in noise reduction. Petschnigg et al. [10] suggested a modified version of bilateral filter called joint bilateral filter. The modification is that the weights are computed from another guidance image rather than the filter input. The joint bilateral filter is preferred if the edge information on the filter input is not reliable. He et al.[13] proposed a new type of explicit image filter which is called guided filter. It is derived from a local linear model, and the filter output is generated by considering the content of the guidance image. Compare to joint bilateral filter, the guided filter performs better near the edges.

B. Monocular Depth Map Estimation

If a video is not captured by stereo camera or 3d sensor, the depth maps of the video is not available. Some paper suggests a method to assign depth value automatically. Burazerovic et al.[8] propose a depth profiling scheme which generates a rough depth map by analyzing image content, then performing joint bilateral filter to refine the final result. Saxena et al.[1], [2] apply a Markov Random Field learning algorithm to capture monocular cues and corporate them into a stereo system. The quality of automatic estimated depth map might not well enough for 2D to 3D conversion. A proper depth map refinement scheme could be used to improve the results of automatic depth map estimation scheme.

C. Depth Map Refinement

The quality of depth map is an important factor of the application which need depth information as input. There are several works are proposed to enhance the quality of depth map. Kornprobst et al.[18] mentions depth map reconstruction methods based on joint bilateral filter. Mueller et al.[15] presents a mechanism for improving depth maps by adaptive cross-trilateral median filtering. In this approach, the weighted median is used rather than the weighted average, and the cross-trilateral filter is an improved version of joint bilateral filter. Some times, the depth map is down-sampled for computational or compression issues. Kopf et al.[12] introduces a joint bilateral up-sampling procedure which could used to enhance the resolution of down-sampled depth map. Yang et al.[19] provides a optimization solution based on cost-volume filter to reconstruct high resolution depth map. Considerate temporal consistent is also an important factor of depth map quality. Smirnov et al.[20] extents Yang's work to process depth map both spatial and temporal domain. It provides a solution to improve the temporal consistent of depth map sequence.

D. Depth propagation

Depth information can be propagated from key frame to other frames if the quality of key frames is good. Harman et al.[17] suggested a machine learning based depth propagation. A neural network or a decision tree is trained to determine the depth value by given a pixel's position and its color. Varekamp et al.[5] introduced a framework that interpolates next frame by a joint bilateral filter and then uses block duplication to reduce undesired artifacts.

III. Methodology

A. Overview

Since it is difficult to obtain an accurate depth map of an image in most cases, several algorithms based on joint bilateral filter were proposed to overcome this problem. However, those algorithms may cause halo effects if incorrect information are involved as the traditional joint bilateral filter calculates the weighted average in a neighborhood. Here a voting-based process is proposed to avoid those undesired effects. The flowchart of the proposed method is shown in Fig. 1. Taking a video and its corresponding depth maps as input, the first step is to construct a space-time filtering window. With the aid of temporal correspondence information provided by optical flow, the proposed filter is performed both spatially and temporally. Then outlier detection is applied followed by computing the weighted average of depth values within the space-time window. As a result, the depth maps of the input sequence are successfully refined.

The choices of this window provide flexibility in dealing with either an image or a video. If the window includes only the current frame, the systems outputs one depth map. One practical application is that a user could simply draw a depth map roughly and the proposed method is able to refine it to improve the quality, which makes 2D to 3D conversion an easy task. If the window contains several previous frames, the proposed system is capable of generating the depth map of current frame. In other words, it serves as depth propagation since the depth maps of a sequence can be interpolated automatically from depth maps of key frames.

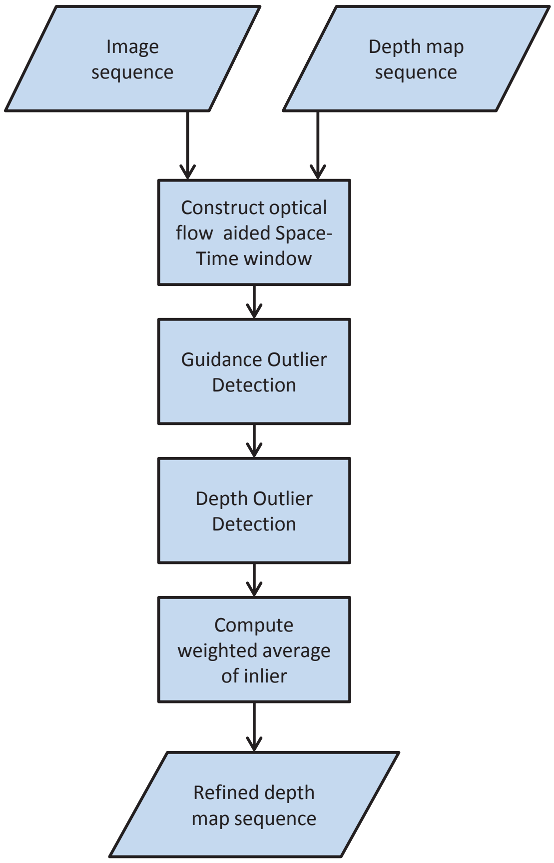


Fig. 1: Flow chart of proposed method.

B. Voting Based Edge Preserving Filtering

The joint bilateral filter calculates a weighted average of the pixels' values. When it is applied to a depth map, it is defined as:

$$D_p^{'} = \frac{1}{Z_p} \sum_{q \in W_p} w_{p,q} D_q \quad (1)$$

where D_q denotes the depth value of input at q and $D_p^{'}$ represents the depth value of filter output at p. W_p is the filter window centered at p. Z_p is a normalization factor that ensures pixel weights sum to 1:

$$Z_p = \sum_{q \in W_p} w_{p,q} \quad (2)$$

$w_{p,q}$ is the joint bilateral kernel which is defined as:

$$w_{p,q} = G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(\|I_p - I_q\|) \quad (3)$$

where I_p denotes the intensity value of the guidance image at p. G_{σ_s} is the zero-mean spatial Gaussian function that decreases the influence according to the distance between p and q. G_{σ_r} is the zero-mean range Gaussian function that reduces the influence caused by the color difference between I_q and I_p.

Though the joint bilateral filter smoothes an image and preserves edges simultaneously, it usually generates halo artifacts when adopted to enhance or reconstruct depth maps. The reason why it causes halo effects is that there are usually many outliers involved in computing the weighted average. The outliers mentioned above include guidance outlier and depth outlier. The first one means that I_q is much different from I_p. In this case, the weight of the outlier controlled by the σ_r. However, if the σ_r is too small, the weight becomes small as well and the effect of smoothing is reduced. The second outlier is the one whose edges of input image and the guidance image are misalignment. In this case, two or more different groups of depth values may weigh high, which causes blur artifacts in the filter output.

Fig. 2 illustrates an example of edge misalignment between a guidance image and a depth map. Fig. 2(a) shows a window centered at p in the guidance image. Fig. 2(b) is the window centered at p in the depth map. As illustrated in Fig. 2(c), combining Fig. 2(a) and Fig. 2(b) results in edge misalignment. Region A represents the background pixels with background depth values. Also, since p belongs to foreground, all of pixels in region A should not participate in determining the depth value of p and pixels in region A are defined as guidance outliers. Region B is the foreground pixels with background depth values. Obviously, the depth values of the pixels in region B are incorrect so that pixels in region B are defined as depth outliers. Region C is the area of foreground pixels with foreground depth values. Thus, an appropriate depth value of point p should be determined by the weighted average of depth values in region C only.

In order to solve the problem of edge misalignment, a voting-based process is developed. Pixels whose color are similar to the target pixels have the right to vote. The pixels who don't have the right to vote are seen as guidance outliers. Under the assumption that the depth values of most pixels are

(a) (b) (c)

Fig. 2: An example of outlier. (a) a window in the guidance image, (b) the corresponding window in the depth map, (c) combine (a) and (b).

correct, since pixels in region C are more than that in region B, the depth values in region C dominate and pixels in region B are treated as depth outliers. Two outlier detection algorithms are proposed: guidance outlier detection and depth outlier detection. Modified from mean-shift[9], these two methods are mode seeking algorithms. A mode seeking algorithm finds the multitude of pixel's color or depth values in a filter window. The two outlier detection algorithms are described in Fig. 3 and Fig. 4, respectively.

procedure GUIDANCEOUTLIERDETECTION(I, W_p)
  $Z = \sum_{q \in W_p} w_{p,q}$
  $\bar{I}_p = \frac{1}{Z} \sum_{q \in W_p} w_{p,q} I_q$
  $var_p = \frac{1}{Z} \sum_{q \in W_p} w_{p,q} (I_q - \bar{I}_p)^2$
  while $var_p > \sigma_r^2$ and var_p is changing **do**
    $O_p^{'} = \{q | q \in W_p and (I_q - \bar{I}_p)^2 > \alpha * var_p\}$
    $Z = \sum_{q \in W_p \setminus O_p^{'}} w_{p,q}$
    $\bar{I}_p = \frac{1}{Z} \sum_{q \in W_p \setminus O_p^{'}} w_{p,q} I_q$
    $var_p = \frac{1}{Z} \sum_{q \in W_p \setminus O_p^{'}} w_{p,q} (I_q - \bar{I}_p)^2$
  end while
  return $O_p^{'}$
end procedure

Fig. 3: The algorithm of guidance outlier detection.

In Fig. 3, I denotes the original image. The notations of W_p, $w_{p,q}$ and I_p are the same as (1). $O_p^{'}$ denotes the set of guidance outliers and var_p represents the color variance of current inlier. Z is the normalization factor and α is s parameter which controls the convergence speed. σ_r is the standard deviation of color difference which determines the convergence. The original mean-shift algorithm searches local maxima of a probability density by given samples. In the proposed algorithm, bilateral weighting function is used to replace original kernel function of mean-shift. This modification makes the proposed method shifts the mean to the global maxima gradually. When the variance of inliers is lower than a pre-defined threshold, or the variance stops decreasing, the iteration terminates. Notice that if α is too small, it is easy to reach local maxima. The suggestion of selecting α is in the range of 1.0 to 1.5.

The depth outlier detection algorithm is similar to guidance outlier detection. In Fig. 4, D denotes an input depth map, D_p represents the depth value at p, $O_p^{''}$ means the depth outliers

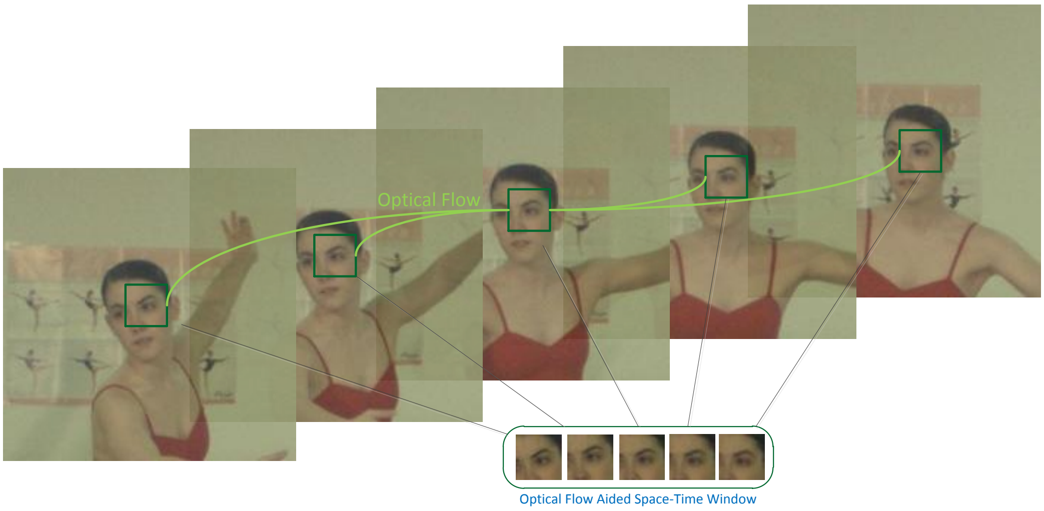


Fig. 5: Illustration of space-time window.

procedure DEPTHOUTLIERDETECTION($D, W_p, O_p^{'}$)
  $Z = \sum_{q \in W_p \backslash O_p^{'}} 1$
  $\bar{D}_p = \frac{1}{Z} \sum_{q \in W_p \backslash O_p^{'}} D_q$
  $var_p = \frac{1}{Z} \sum_{q \in W_p \backslash O_p^{'}} (D_q - \bar{D}_p)^2$
  while $var_p > \sigma_d^2$ and var_p is changing **do**
    $O_p^{''} = \{q | q \in W_p \backslash O_p^{'} and (D_q - \bar{D}_p)^2 > \alpha * var_p\}$
    $Z = \sum_{q \in W_p \backslash (O_p^{'} \cup O_p^{''})} 1$
    $\bar{D}_p = \frac{1}{Z} \sum_{q \in W_p \backslash (O_p^{'} \cup O_p^{''})} D_q$
    $var_p = \frac{1}{Z} \sum_{q \in W_p \backslash (O_p^{'} \cup O_p^{''})} (D_q - \bar{D}_p)^2$
  end while
  return $O_p^{''}$
end procedure

Fig. 4: The algorithm of depth outlier detection.

and var_p is the depth variance of current inlier. Since the depth value around the target point might not be correct, we just find the mode of depth values in $W_p \backslash O_p^{'}$. σ_d is the standard deviation of depth values. This parameter controls when to stop and it also controls the smoothness of output. The depth outlier is represented by region B in Fig. 2(c). Our voting-based filter is then defined as:

$$D_p^{'} = \frac{1}{Z_p} \sum_{q \in W_p \backslash (O_p^{'} \cup O_p^{''})} w_{p,q} D_q \tag{4}$$

where the normalization factor Z_p is defined as:

$$Z_p = \sum_{q \in W_p \backslash (O_p^{'} \cup O_p^{''})} w_{p,q} \tag{5}$$

This mode seeking algorithm provides a robust analysis for excluding outliers.

C. Space-Time Filtering

Generating depth maps of a video frame by frame usually results in temporal inconsistency. Trembling effects are inevitable for a 3D video which is converted from a 2D sequence based on these depth maps. In order to enhance the visual quality of 3D videos, how to deal with the inconsistency of depth maps of adjacent frames becomes an important issue. A space-time filter which concerns both previous and following frames is proposed to solve this problem. Observing that there might exists a non-rigid transformation between neighboring frames, the center of the proposed filtering window is adjusted according to optical flow [11]. Fig. 5 shows an example of the proposed space-time filtering window and the space-time filter is defined as

$$D_p^{'(t)} = \frac{1}{Z_p} \sum_{(q,t^{'}) \in W_p^{(t)} \backslash (O_p^{'(t)} \cup O_p^{''(t)})} w_{p,q,t,t^{'}} D_q^{(t^{'})} \tag{6}$$

where t and $t^{'}$ are frame indices and $w_{p,q,t,t^{'}}$ is the space-time kernel expressed as follows:

$$w_{p,q,t,t^{'}} = G_{\sigma_s}(\|p + f_p^{(t^{'})} - q\|) G_{\sigma_r}(\|I_p^{(t)} - I_q^{(t^{'})}\|) \tag{7}$$

Moreover, $W_p^{(t)}$ is an optical-flow aided space-time window:

$$W_p^{(t)} = \{(q, t^{'}) \mid \|t^{'} - t\| < T \, and \, \|p + f_p^{(t^{'})} - q\| < R\} \tag{8}$$

where $f_p^{(t^{'})}$ represents the flow vector from frame t to frame t' at p, T denotes the temporal radius and R is the spatial radius. The choice of radii depends on the precision of optical flow. The higher the precision, the smaller the radius. As it can be seen from above equations, since the center of the window moves according to the optical flow, the depth value of a pixel is determined by almost the same group of pixels. That means the depth values of the same point in different frames are similar thus the proposed filter successfully preserves temporal consistency. Moreover, a voting process is performed in the space-time window so that the output is still reliable even though there exist some uncertainties of optical flow.

In addition to preserving temporal consistency of depth maps of a sequence, how to save computational complexity of generating depth maps also attracts attentions. A concept of depth propagation is proposed here to target this issue. Different from traditional process, here we tend to obtain only the depth maps of several key frames and try propagate those depth information to other frames according to the characteristic of a sequence. The beauty of the proposed space-time filtering is that it also provides depth propagation by slightly modifying the filter window to includes only previous frames:

$$D_p^{'(t)} = \frac{1}{Z_p} \sum_{(q,t') \in (W_p^{(t)} \cap \{(q,t')|t'<t\}) \setminus (O_p^{'(t)} \cup O_p^{''(t)})} w_{p,q,t,t'} D_q^{(t')} \tag{9}$$

The modified version not only interpolates current depth map from previous frames but also maintains temporal consistency. Moreover, the proposed framework is capable of propagating depth information in a shot even with large movements.

D. Coarse-to-Fine Refinement

How to choose the radius of filter window depends on the quality of input depth map and the precision of optical flow. An ideal radius should be adequately large to cover sufficient information, otherwise the correct depth values will be recognized as outliers. However, when the window size is getting large, the process becomes time consuming. Therefore, a speed-up scheme is further proposed. Rather than performing the filter directly on the depth map of the input image, pyramids of the image and its depth map are constructed first. The filter is applied to the coarsest level and the output is upsampled and serves as the input to the next level. This process repeats until it reaches the finest level. In other words, adopting a filter with the same size to the coarser level is equivalent to applying a filter with larger kernel size to the original image. The proposed multi-level structure accelerates the process drastically while preserving satisfactory results. Fig. 6 details the coarse-to-fine algorithm. r is the radius of the filter window and L represents the number of layers. I^c and D^c denote the down-sampled original image and its depth map.

procedure PYRAMIDVBJ(I, D, r, L)
    if $L! = 1$ **then**
        $I^c = Gaussian\_downsampling(I)$
        $D^c = NN\_downsampling(D)$
        $D^c = PyramidVBJ(I^c, D^c, r, L-1)$
        $D = NN\_upsampling(D^c)$
    end if
    return $VotingBasedFilter(I, D, r)$;
end procedure

Fig. 6: The algorithm of coarse-to-fine depth map refinement.

IV. Experimental Results

A. Single Image Depth Map Refinement

The depth maps refined by the proposed method are compared to the results refined by two other existing filters: joint bilateral filter [8] and guided image filter [13]. Fig. 7(a) show the original images searched from internet and the Fig. 7(b) are the depth maps drawn roughly by the user. We can clearly see that these two depth maps are not precise as the edges are not well-aligned to the ones of the original images. These two roughly drawn depth maps are then refined by the joint bilateral filter, guided image filter and the proposed filter, and the results are demonstrated in Figs. 7(c), (d), and (e), respectively. The depth map generated by the proposed method outperforms the others as the edges are more precise and there are no halo effects.

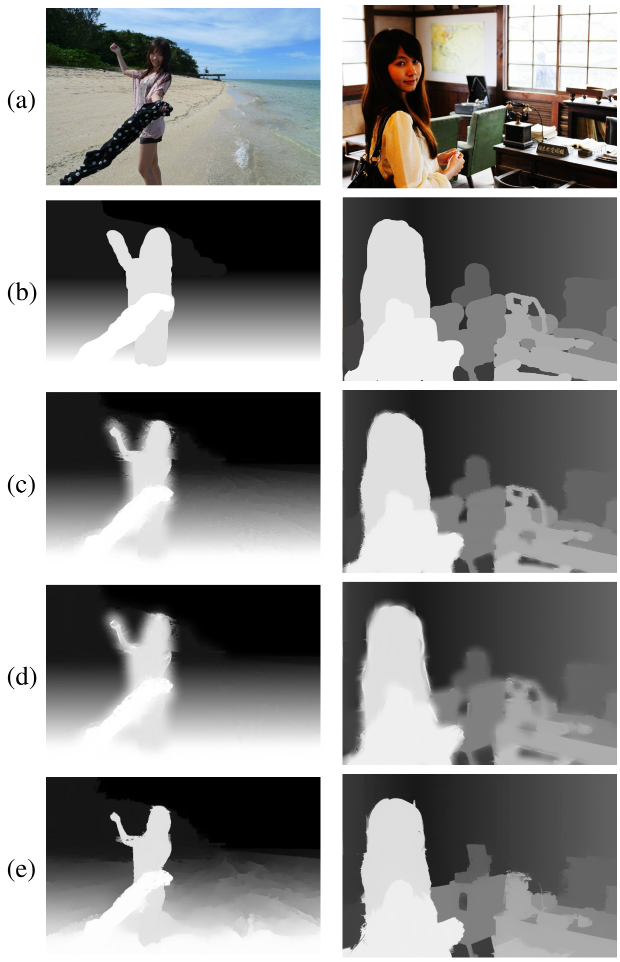
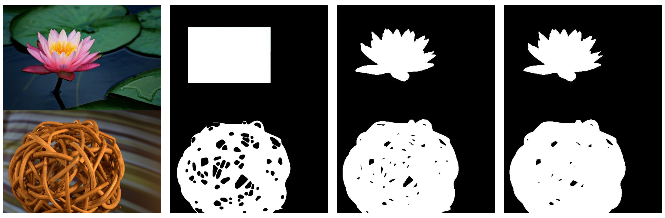


Fig. 7: Results of depth map refinement. (a) original images, (b) user's quick sketch, (c) refined by the bilateral filter, (d) refined by the guided image filter, (e) refined by the proposed method.

Fig. 8: Results of performing DIBR on the images of Fig. 7(a) and the refined depth maps by the proposed method.

Another case shows the performance of the edge alignment of the proposed filter. Suppose that the input is a binary image which is roughly marked as foreground and background. As we can see from Fig. 9, the proposed filter outputs a better result of foreground and background. In other words, it also able to refine the segmentation result and its performance is similar to GrabCut[6]. However, the proposed method is not good at dealing with small holes since they may be treated as outliers in the down-sampled version.



(a) (b) (c) (d)

Fig. 9: Results of image segmentation. (a) original images (b) masks drawn by a user, (c) result obtained by using GrabCut[6], (d) result obtained by using the proposed method.

B. Temporal Depth Map Refinement

In this section, rather than dealing with a single image, the proposed method is performed to refine a sequence of depth maps. Fig. 10(a) shows several frames of 'ballet' sequence and Fig. 10(b) are the corresponding depth maps[3]. As it can be seen from those figures, the depth values of some regions are incorrect due to inappropriate depth estimation . However, the depth information of these regions in neighboring frames are correct. Thus, by considering the temporal consistency, the proposed filter is able to correct the depth values of these regions effectively. The results obtained by using the proposed filter are presented in Fig. 10(c).

C. Depth Propagation

An image sequence is chosen from a real movie to demonstrate the proposed algorithm. In the first experiment, a shot from a Taiwanese movie called "BLACK & WHITE" is used. The challenge in this shot is that there is an object moving with a complex background. Note that the airplane is rotating so that its shape is changing in different frames. To maintain the shape of the object on the depth map in the depth propagation procedure is a challenging task. This shot includes 19 frames.

(a) (b) (c)

Fig. 10: Results of temporal depth map refinement. (a) original frames from 'ballet' sequence, (b) the depth maps from the dataset, (c) refined depth maps by applying the proposed method.

Fig. 11 shows the first frame and a hand-drawn depth map; the frame is referred to the key frame. The results obtained by using the proposed method and other methods such as the joint bilateral filter based method[5] and the direct propagation of the depth values according to optical flow[11] are resented in Fig. 12. As we can see from those resultant images, the main problem of other methods is the accumulated error. The accumulated errors in the case are reduced by a voting process. This is conducive to the propagation of the depth values over a longer. Fig. 13 demonstrates the results of DIBR on original frames and the depth maps obtained by using the proposed method.

Fig. 11: Key frame of shot "Airplane" and its depth map.

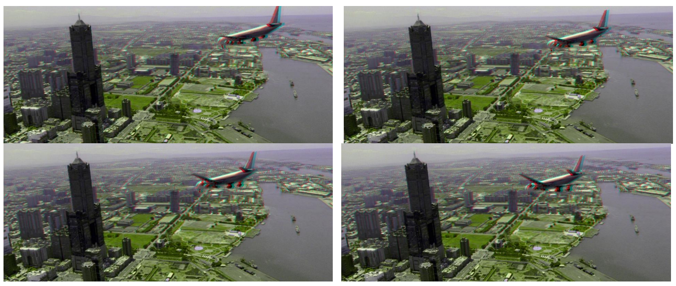


Fig. 13: Anaglyphs of frames 5, 10, 15 and 19 which are converted based on the propagated depth map by the proposed method.

Another shot is also chosen from the movie "BLACK & WHITE." The challenge of this shot is that it involves camera motion and the movements of multiple objects. Fig. 14 shows the first frame and its hand-drawn depth map which is treated as the key frame. This shot includes 38 frames. Fig. 15 demonstrates the results of depth propagation of the proposed method and others. It is observed that the proposed algorithm is also effective in propagating the depth values of multiple moving objects. Fig. 16 shows anaglyphs generated by DIBR and the depth maps acquired by using the proposed method.

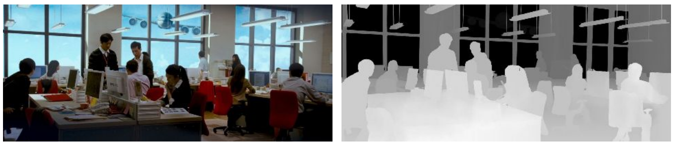


Fig. 14: Key frame of shot "Office" and its depth map.

Fig. 16: Anaglyphs of frames 10, 20, 30 and 38 which are converted based on the propagated depth map by the proposed method.

V. CONCLUSION

A voting-based filter is proposed to refine a depth map sequence in this paper. The voting process ensures that only appropriate information is involved in determining the filter output so that it is able to correct erroneous regions, especially edges, of the depth map. Moreover, since the proposed space-time filter is performed adaptively by considering information of neighboring frames, it successfully reduces the influence of temporal inconsistency of a sequence of depth maps. By slightly modifying the filter window, the proposed scheme can also propagate depth information from key frames to others, which substantially reduces the computational complexity. Furthermore, the process can be accelerated by adopting multi-level structure. In conclusion, the proposed scheme provides a flexible and convenient way to obtain a sequence of depth maps with satisfactory quality at low cost. It generates good depth maps as input for 2D to 3D conversion and is expected to be widely used in 3D movie and 3DTV industries.

ACKNOWLEDGMENT

This research is sponsored by The White Rabbit Entertainment. Special thanks are given to Mr. Yueh-Hsun Tsai and Prajna Works for providing the multimedia contents of Black & White for experiments. Through the assistance of The White Rabbit Entertainment and Prajna Works, this research was able to accomplish successfully.

REFERENCES

[1] A. Saxena, J. Schulte, and A. Y. Ng. Depth estimation using monocular and stereo cues. In *In IJCAI*, 2007.

[2] A. Saxena, M. Sun, and A. Y. Ng. Make3d: Learning 3d scene structure from a single still image. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(5):824–840, May 2009.

[3] C. Lawrence Zitnick, S. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High-quality video view interpolation using a layered representation. *ACM Trans. Graph.*, 23(3):600–608, August 2004.

[4] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proceedings of the Sixth International Conference on Computer Vision*, ICCV '98, pages 839–, Washington, DC, USA, 1998. IEEE Computer Society.

[5] C. Varekamp and B. Barenbrug. Improved depth propagation for 2d to 3d video conversion using key-frames. In *Visual Media Production, 2007. IETCVMP. 4th European Conference on*, pages 1 –7, nov. 2007.

[6] C. Rother, V. Kolmogorov, and A. Blake. "grabcut": interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23(3):309–314, 2004.

[7] C. Fehn. Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV. In *Electronic Imaging 2004*, volume 5291 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pages 93–104. The International Society for Optical Engineering., May 2004.

[8] D. Burazerovic, P. Vandewalle, and R.-P. Berretty. Automatic depth profiling of 2d cinema - and photographic images. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 2365 –2368, nov. 2009.

[9] D. Comaniciu, P. Meer, and S. Member. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:603–619, 2002.

[10] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama. Digital photography with flash and no-flash image pairs. In *ACM SIGGRAPH 2004 Papers*, SIGGRAPH '04, pages 664–672, New York, NY, USA, 2004. ACM.

[11] G. Farnebäck. Two-frame motion estimation based on polynomial expansion. In *Proceedings of the 13th Scandinavian conference on Image analysis*, SCIA'03, pages 363–370, Berlin, Heidelberg, 2003. Springer-Verlag.

[12] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele. Joint bilateral upsampling. In *ACM SIGGRAPH 2007 papers*, SIGGRAPH '07, New York, NY, USA, 2007. ACM.

[13] K. He, J. Sun, and X. Tang. Guided image filtering. In *Proceedings of the 11th European conference on Computer vision: Part I*, ECCV'10, pages 1–14, Berlin, Heidelberg, 2010. Springer-Verlag.

[14] M. Guttmann, L. Wolf, and D. Cohen-Or. Semi-automatic stereo extraction from video footage. In *ICCV*, pages 136–142. IEEE, 2009.

[15] M. Mueller, F. Zilly, P. Kauff. Adaptive cross-trilateral depth map filtering. In *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, 2010.

[16] P. Kauff, N. Atzpadin, C. Fehn, M. Müller, O. Schreer, A. Smolic, and R. Tanger. Depth map creation and image-based rendering for advanced 3dtv services providing interoperability and scalability. *Image Commun.*, 22(2):217–234, February 2007.

[17] P. Harman, J. Flack, S. Fox, and M. Dowley. Rapid 2d to 3d conversion. In *in Stereoscopic Displays and Virtual Reality Systems IX, Andrew*, pages 78–86, 2002.

[18] P. Kornprobst, J. Tumblin, and F. Durand. Bilateral filtering: Theory and applications. *Foundations and Trends in Computer Graphics and Vision*, 4(1):1–74, 2009.

[19] Q. Yang, R. Yang, J. Davis, and D. Nistér. Spatial-depth super resolution for range images. In *CVPR*, 2007.

[20] S. Smirnov, A. P. Gotchev, and K. O. Egiazarian. A memory-efficient and time-consistent filtering of depth map sequences. In Jaakko Astola and Karen O. Egiazarian, editors, *Image Processing: Algorithms and Systems*, volume 7532 of *SPIE Proceedings*, page 753217. SPIE, 2010.

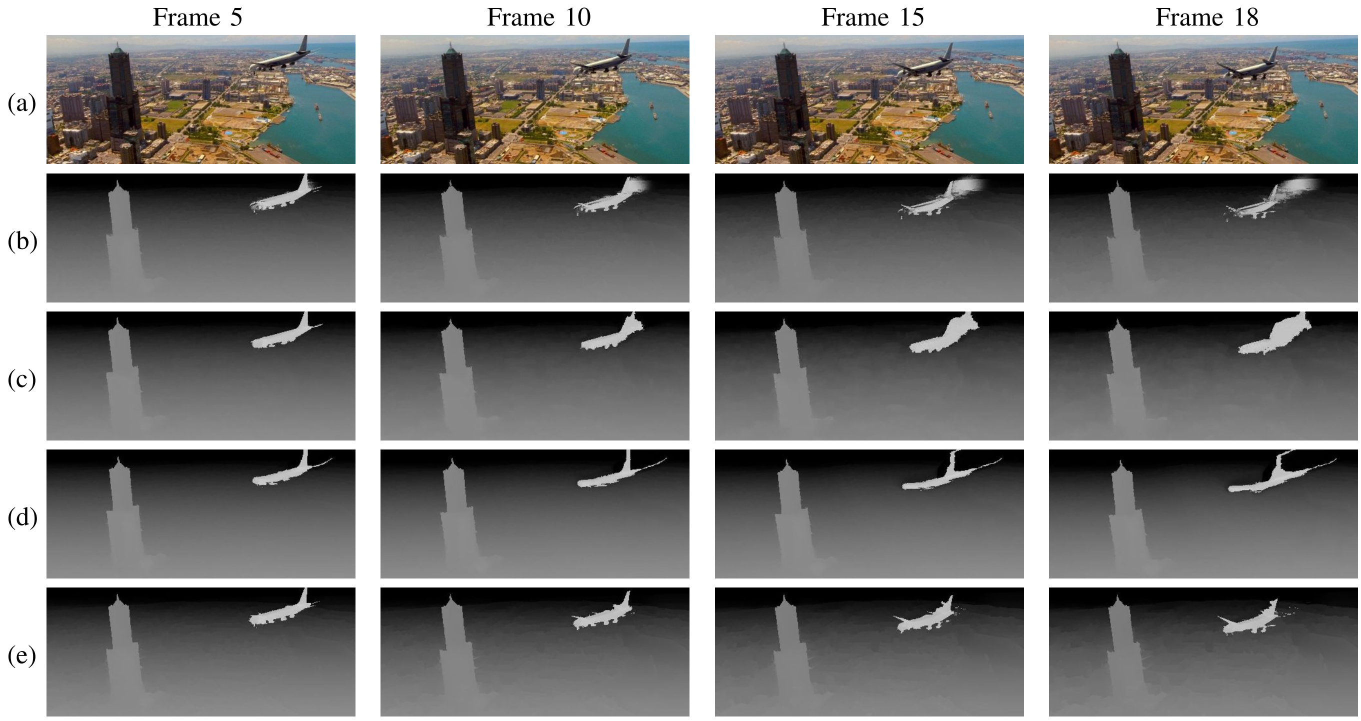


Fig. 12: The propagation results of shot ”Airplane”. (a) original frames, (b) propagated by joint bilateral filter, (c) the results of [5], (d) propagated by optical flow only[11], (e) the proposed method.

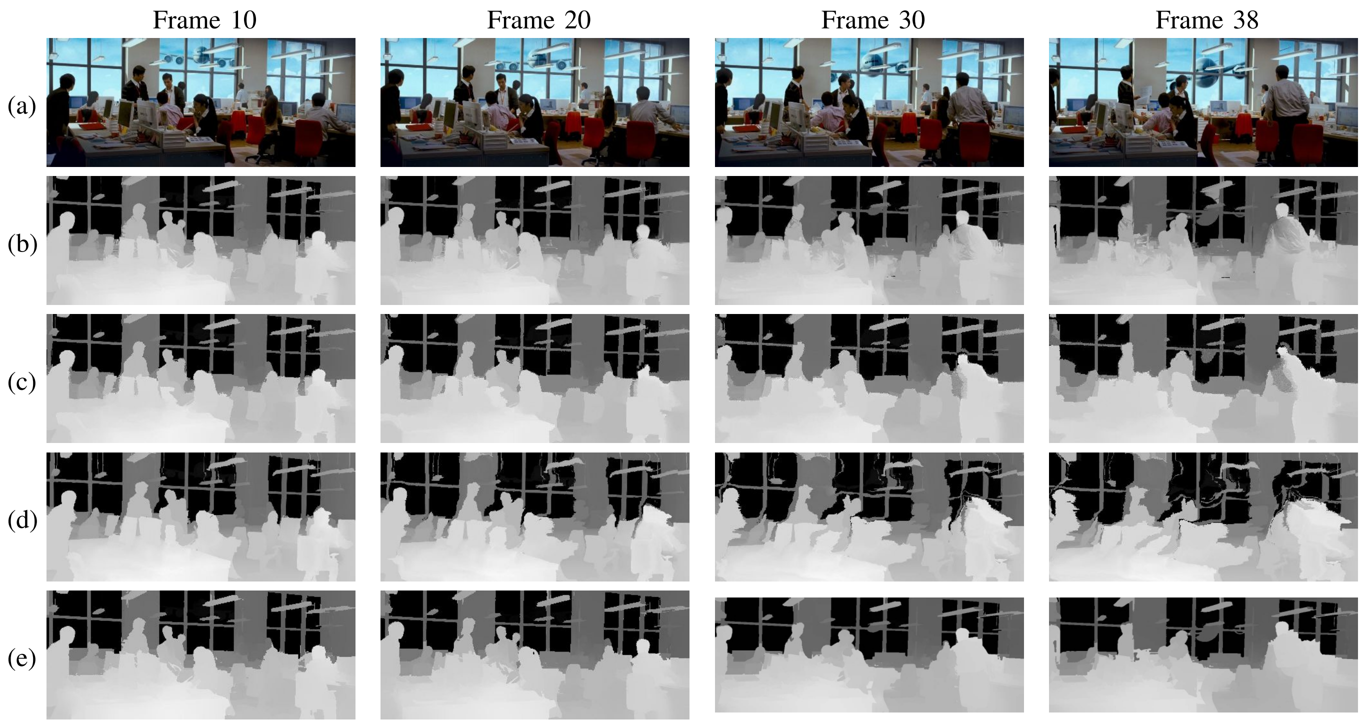


Fig. 15: The propagation results of shot ”Office”. (a) original frames, (b) propagated by joint bilateral filter, (c) the results of [5], (d) propagated by optical flow directly[11], (e) the proposed method.