WATERSHED BASED DEPTH MAP MISALIGNMENT CORRECTION AND FOREGROUND BIASED DILATION FOR DIBR VIEW SYNTHESIS

Xuyuan Xu[1], Lai-Man Po[1], Kwok-Wai Cheung[2], Litong Feng[1], Chun-Ho Cheung[3]

1. Department of Electronic Engineering, City University of Hong Kong, Hong Kong SAR, China
2. Department of Computer Science, Chu Hai College of Higher Education, Hong Kong SAR, China
3. Department of Information Systems, City University of Hong Kong, Hong Kong SAR, China

ABSTRACT

The quality of the synthesized views by Depth Image Based Rendering (DIBR) highly depends on the accuracy of the depth map, especially the alignment of object boundaries of texture image. In practice, the misalignment of sharp depth map edges is the major cause of the annoying artifacts at the disoccluded regions of the synthesized views. In this paper, a new depth map preprocessing method using Watershed misalignment correction and dilation filter is proposed to align the foreground depth edges to cover the whole transitional color edge regions. This approach can handle the sharp depth map edges lying inside or outside the object boundaries in 2D sense. The quality of the disoccluded regions of the synthesized views can be significantly improved. Experimental results show that the proposed method achieves superior performance for view synthesis by DIBR especially for generating large baseline virtual views.

***Index Terms*—**DIBR, Misalignment, View Synthesis

1. INTRODUCTION

Nowadays 3D technology has been widely applied in various fields, like virtual reality, digital entertainment and immersive telecommunication, etc. 3D video improves the user's viewing experience by providing depth perception. Two of the high potential 3D representations are video-plus-depth format and multi-video plus multi-depth format. In video-plus-depth format, 3D video is represented by one color view and its corresponding depth map. It is a very compact format especially suitable for mobile 3D video applications. Stereo video and multi-view video can be synthesized with variable baseline through the Depth Image Based Rendering (DIBR) [1]. The synthesized videos can be used in the stereoscopic or auto-stereoscopic 3D display devices [2-4].

Generally, view synthesis using DIBR involves three major steps: (1) Depth Map Preprocessing, (2) 3D Image Warping, and (3) Hole Filling. Example images of the DIBR process using the well-known sequence of Lovebird1 are shown in Fig. 1. The enlarged color image and its corresponding depth map in view 4 of the "Lovebird1" sequence are shown in Fig. 1(a) and 1(b), respectively. The white regions, also called holes, in Fig. 1(c) are disoccluded regions of the synthesized view. Many hole filling methods have been developed to fill these disoccluded regions. Well-known hole filling methods include linear interpolation [5], depth-aid horizontal extrapolation [5], multi-directional extrapolation [6] and inpainting method [7]. The synthesized image using inpainting hole filling methods is shown in Fig. 1(d). However, annoying artifacts exist in large disoccluded regions of the synthesized images even advanced inpainting hole filling method is used.

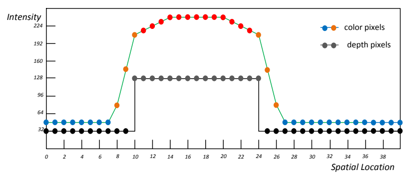
(a) (b) (c) (d)

Fig. 1: Examples of DIBR view synthesis using the enlarged 175th frame of the *Lovebird1* sequence: (a) Input color Image, (b) Input depth map, (c) Warped image with disoccluded regions (holes) in white pixels, and (d) Synthesized views using inpainting[7].

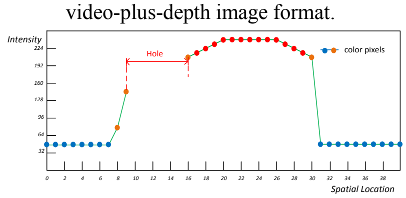
To minimize the artifacts of the synthesized view by DIBR, many depth preprocessing methods have been developed. One main approach is to use a smooth filter [2,8,9] to blur the sharp edge of the depth map, then the size of hole is reduced or even small hole can be eliminated. Unfortunately, the smoothing filter destroys not only the texture in the color image but also the geometrical depth information, degrading the 3D perception in the reconstructed 3D images/videos.

Recent study on the artifacts of the DIBR view synthesis [10] found that the major cause of the annoying artifacts is the misalignment of the object boundaries between the depth map and the corresponding color image. In practice, the depth map estimated from image pairs [11] or captured from depth camera may not align with corresponding color image correctly. In addition, the object edges in color images are always fuzzy and contain transitional edge pixels. As a result, unprocessed depth map usually causes annoying artifacts after the hole filling process. The misalignment may cause the object boundaries of the depth map to lie outside or inside foreground objects in the color image. After the image warping process, foreground color pixels may be regarded as the background pixels and used to fill the disoccluded regions in the hole filling process. Then annoying artifacts may be resulted, especially for the large disoccluded region. The artifacts caused by misalignment may also occur when background pixels are regarded as foreground pixels. Unfortunately, the conventional depth map filters and advanced hole filling methods cannot eliminate these artifacts.

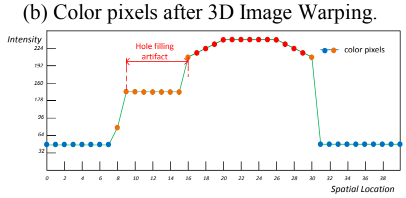
In [10], a simple foreground biased depth map refinement was proposed to tackle the depth map misalignment problem. It focuses on refining the positions of sharp depth map edges to the background regions based on the color edges information. However, this method only considers large horizontal depth discontinuities using very simple line-by-line processing, and therefore cannot be generalized to 2D depth map processing. In this paper, a more general solution for tackling the depth map misalignment problem is proposed using Watershed color segmentation and foreground biased depth map dilation. This paper is organized as follows. Section 2 describes the misalignment problem. Details of depth map misalignment correction are presented in section 3. Experimental results are shown in section 4. Finally, a conclusion is drawn in section 5.



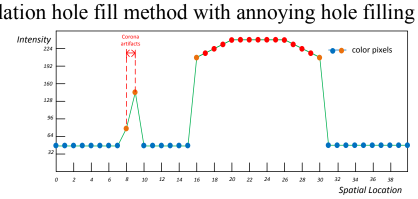
(a) Color pixel intensity values and depth values for a horizontal line in video-plus-depth image format.



(b) Color pixels after 3D Image Warping.



(c) Effect of hole filling from neighboring pixels using depth-aid horizontal extrapolation hole fill method with annoying hole filling artifact.

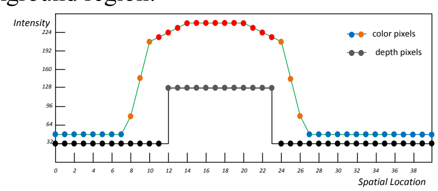


(d) Effect of hole filling with the pixels from the neighboring frame views.

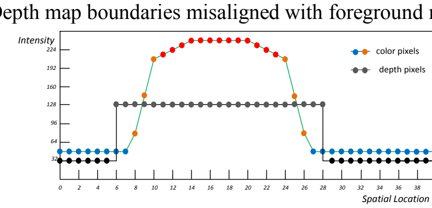
Fig. 2: Depth boundaries misaligned with transitional region of color edge.

2. DEPTH MAP MISALIGNMENT PROBLEM

The characteristics of depth map and natural color image are very different. Depth map is a piece-wise smooth image [12] that has large homogeneous regions within scene objects and contains sharp changes of depth at object boundaries. However, the edges in color image usually have intensity changing smoothly over transition regions. Fig. 2(a) shows the color intensity and depth pixel values of a horizontal line in a color and depth image. There are two depth map object boundaries aligned with the middle of the transitional texture edges. Depth map captured by depth camera or estimated from video frames may not align with color view correctly. These depth map edges may be misaligned at the foreground regions or background regions as shown in Fig. 3(a) and Fig. 3(b), respectively. After 3D Image Warping, a large hole is created in the synthesized view as shown in Fig. 2(b). In DIBR based view synthesis using one texture image, this hole is filled with the neighboring background pixels. Fig. 2(c) shows the effect of filling this hole using depth-aid horizontal extrapolation method. Object boundaries contain a combination of foreground and background color information. Incorrect depth values may be assigned to these edge pixels. That means foreground color pixels or transitional edge pixels with similar foreground colors are incorrectly treated as the background pixels and used to fill the holes, resulting in annoying artifacts in the synthesized views. The misalignment is also the cause of corona-like artifacts of the synthesized view when filling up the holes with the pixels from the neighboring view as Fig. 2(d). More annoying artifacts may be caused if the depth edges are misaligned with the foreground region as the case of Fig 3(a). However, if the depth map edges are misaligned with the background region without texture, the artifacts may not be very serious as the case of Fig 3(b). It is because the whole transitional edges are mapped to the foreground region with the synthesized view after 3D image warping. In addition, the holes are created in the background region. Thus, the holes have much higher chance to be filled up with pixels that are similar to the background region.



(a) Depth map boundaries misaligned with foreground region.



(b) Depth map boundaries misaligned with background region.

Fig. 3: Depth map misalignment.

3. DEPTH MAP MISALIGNMENT CORRECTION

According to the analysis in section 2, the depth map can be refined in the preprocessing stage in order to solve the misalignment problem. The idea is to have the foreground depth value to cover the whole transitional edges in color image. Then, the synthesized artifacts around the object boundaries should be significantly minimized in the synthesized views. To handle depth map edges misaligning at the foreground regions or background regions in horizontal, vertical and diagonal directions, a Watershed color segmentation based depth map misalignment correction method is proposed. Depth values around the sharp depth map edges are reassigned based on color pixels similarity to align the depth map edges within the transitional texture edge regions. And then a foreground biased depth map dilation process is used to extend the depth map object boundaries to cover the whole transitional texture edge regions in color image. This can minimize the annoying boundary artifacts generated during hole filling. The proposed method consists of three steps: (1) Unreliable region detection, (2) Watershed misalignment correction, and (3) Foreground biased depth map dilation. The details of these steps are discussed in the following subsections.

3.1. Unreliable Region Detection

Theoretically, synthesis artifacts are generated around object boundaries in the synthesized view. Depth discontinuity can generate disocclusion and occlusion after image warping. Most of the synthesized artifacts are the hole filling artifacts in disoccluded region. However, small and gradually changing of depth values does not give annoying artifacts because small depth discontinuities only generate very small holes in the warped color image. The annoying artifacts are related to large depth discontinuities of the depth map, which can generate large holes after the 3D warping process. To achieve high efficiency and maintain the depth map quality in smoothing region, only sharp depth map edges should be processed for misalignment correction. Sharp depth edges that create hole size larger than 3 pixels in image warping process are treated as unreliable regions as Fig. 4(a). However, these regions may not totally cover large misalignment regions outside or inside the object boundaries. Thus, a grey scale dilation filter [13] is repeated for 3 times to construct unreliable regions as Fig. 4(b).

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

Fig. 4: (a) Unreliable pixels protruded in the depth map (in white), (b) Unreliable regions are extended by repeating the dilation filter for 3 times, (c) Synthesized view using Watershed misalignment corrected depth map, and (d) Synthesized view using foreground biased dilated depth map.

3.2. Watershed Misalignment correction

There are three cases of depth map misalignment: (1) the sharp depth edge lies outside the object boundary, (2) the sharp depth edge lies inside the object boundary, and (3) the sharp depth edge lies within the transitional region of the texture edge. Ideally, these sharp depth map edges should be corrected to align to the middle of the transitional texture edges regions. The proposed Watershed segmentation based misalignment correction reassigns the depth values within the unreliable regions, by relocating the depth map edges to the maximum gradient positions of the transitional texture edge regions.

Watershed segmentation is a process to classify labels for each of the pixels in a given image. To speed up the Watershed segmentation process, a fast flooding process based on priority [14] queue is used. In the proposed misalignment correction method, Watershed algorithm uses depth values of the reliable regions of the depth map as segmentation labels to generate the labels of the unreliable regions based on the color pixels similarity in the texture image. Pixels in the unreliable regions are assigned with a label value (depth value) of the neighboring reliable depth pixels using color similarity.

3.3. Foreground Biased Depth Map Dilation

After the depth map misalignment correction, sharp depth map edges are located within the transitional texture edge regions of the color image. If such depth map is used for the DIBR view synthesis, annoying boundary artifacts are still easily observed around the disoccluded regions as shown in Fig. 4(c). The annoying boundary artifacts are due to sharp depth map edges located within the transitional texture edge regions. To enhance the quality of the synthesized view, a foreground biased depth map dilation process is applied to the Watershed misalignment corrected depth map. The dilation process is repeated in such a way that foreground depth values are extended to background regions until foreground depth values cover all the similar color pixels of the transitional texture edges or exceed a predefined extension limit.

In each dilation iteration, one-pixel width refinement regions are extended to the background regions that contain the similar color as foreground pixels. The foreground biased dilation process can be expressed as

$$U^t_{x,y} = B^t_{x,y} \cdot S^t_{x,y} \cdot \max_{\substack{-1 \le i \le 1 \\ -1 \le j \le 1}} U^{t-1}_{x+i,y+j} \qquad (1)$$

where U^t is the mask for representing the region after t^{th} iteration dilation. $U^t_{x,y}$ is the mask value at position (x,y). The initial mask U^0 denotes the barriers (edges of depth map segmented by Watershed in section 3.2). Value 1 in U^0 represents the barriers regions and value 0 is for other regions. B^t is background mask. $B^t_{x,y} = 1$ identifies the position (x,y) as the background region by comparing its neighboring depth values and $B^t_{x,y} = 0$ is for other region. S^t denotes the color similarity mask. The color distance is defined by the minimum Euclidean distance of color components and its neighboring foreground color pixels. If the color distance is less than the threshold W, value 1 is assigned to $S^t_{x,y}$ to indicate that color pixel $C_{x,y}$ at position (x,y) has the similar color to its neighboring foreground pixels. Otherwise, value 0 is assigned to $S^t_{x,y}$. In our experiments, the threshold W is set to 20 and the maximum number of iteration is set to 3.

The foreground biased dilation filter only extends regions that are the background regions and have similar color with foreground regions. The refinement process is described as

$$d^t_{x,y} = d^{t-1}_{\arg\min_{(m,n)\in\Omega} \mathrm{D}(C_{x,y},C_{m,n})}, \forall (\mathrm{x},\mathrm{y}) \in H^t, \qquad (2)$$

$$H^t = \{(x,y) \mid U^t_{x,y} = 1 \ \wedge \ U^{t-1}_{x,y} = 0\}, \qquad (3)$$

where $d^t_{x,y}$ and $d^{t-1}_{m,n}$ are the depth values of position (x,y) and (m,n) for the depth map at t^{th} iteration and depth map at $(t\text{-}1)^{th}$ iteration respectively. d^0 is the Watershed misalignment corrected depth map. Ω denotes the set of coordinate of the foreground neighboring pixels of the pixel at (x,y). $\mathrm{D}(C_{x,y}, C_{m,n})$ represent the Euclidean distance of two color pixels $C_{x,y}$ and $C_{m,n}$. Through the foreground biased depth map dilation process, depth pixels lying within transitional texture edge can be assigned with the foreground depth value and boundary artifacts can be removed as the results in Fig. 4(d).

4. EXPERIMETNAL RESULTS

To evaluate the performance of the proposed method, both subjective and objective evaluations are performed. The proposed method is compared with asymmetric Gaussian filter (AGF) [2], adaptive filter (AF) [8] and foreground biased depth map filter (FBDMF) [10]. The multi-views sequences of *Lovebird1*, *Mobile* and *Breakdance* are used for both subjective and objective evaluations. VSRS alpha 3.5 [15] is used to generate the virtual views and the common hole filling method [7] is used to recover the disoccluded region. For objective measurement of the performance, the synthesized virtual views are compared with the original views based on the Peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) metric.

From results of AGF and AF, boundary artifacts are reduced as compared with the synthesized views without depth map preprocessing. However, these boundary artifacts are still very annoying. In addition, these filtered depth maps are highly degraded around the object boundaries especially for using AGF and the 3D scene of the synthesized views may hardly be maintained. Since the AF has less smoothing effect on depth transitional regions and preserves the sharp transition of depth edge, depth map quality after AF is better than that after AGF.

From the results of FBDM, boundary artifacts are almost removed and misalignment between the color and depth image is corrected in horizontal direction. Unfortunately, boundary artifacts still exist due to misalignment in other direction of color and depth image, as the top regions of shoulder in Fig. 5(d), bottom region of hand in Fig. 5(d).

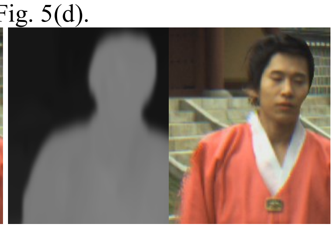
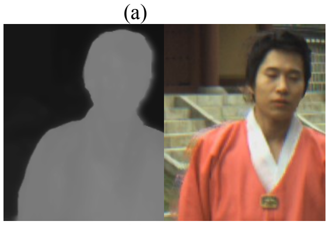
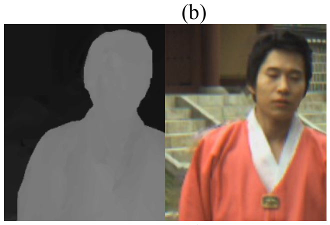
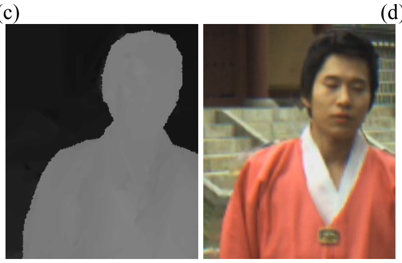

(a)

(b)

(c)

(d)

(e)

Fig. 5: Results of *Lovebird1* sequence. (a) Unprocessed depth map and its synthesized view, (b) Result of AGF, (c) Result of AF, (d) Results of FBDMF, and (e) Results of proposed method.

Results of proposed method show the quality of synthesized views is much better than that of AGF and AF. This is mainly due to the fact that the hole filling processing mainly uses the background pixels information with color dissimilar to the foreground regions. Compared with the FBDMF, the misalignment for all direction can be corrected and boundary artifacts are removed. The irregular boundaries of the depth map using the proposed method is due to the dilation process for covering the transitional pixels of the texture edges. These irregular boundaries do not degrade the 3D perception of the synthesized views as they just make the pixels of the transitional texture edge projected to the foreground regions.

For the objective evaluation, Table 1 and Table 2 show the average PSNR and SSIM of the test sequences. Both PSNR and SSIM are improved using the proposed method. The synthesized views from proposed method are the most similar to their original views among all compared methods. The most significant improvement is found for the *Breakdance* sequence. It is because hole sizes in this rendering are relatively large compared with other test sequences. Among these methods, the asymmetric filter has the worst performance since depth map is highly blurred with the drawback of geometrical depth information is significantly destroyed.

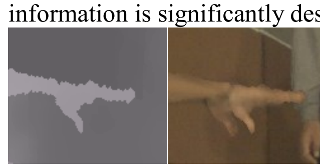
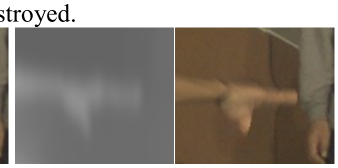
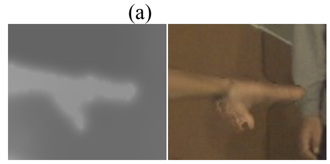
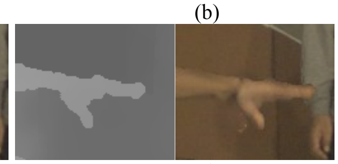
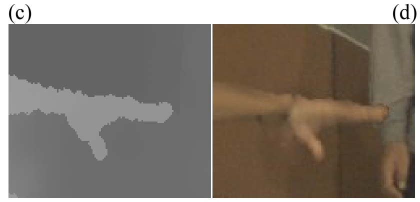

(a)

(b)

(c)

(d)

(e)

Fig. 6: Results of *Breakdance* sequence. (a) Unprocessed depth map and its synthesized view, (b) Result of AGF, (c) Result of AF, (d) Results of FBDMF, and (e) Results of proposed method.

Preprocess / Sequence	Without preprocess(dB)	AGF (dB)	AF (dB)	FBDMF (dB)	Proposed (dB)
Lovebird1	32.7230	32.3913	32.6542	32.7431	**32.7784**
Breakdance	32.5265	31.2710	32.6816	32.7456	**32.8433**
Mobile	36.8611	36.1228	36.6884	36.9147	**36.9691**

Table 1: Average PSNR of the synthesized views.

Preprocess / Sequence	Without preprocess	AGF	AF	FBDMF	Proposed
Lovebird1	0.9395	0.9368	0.9392	0.9397	**0.9402**
Breakdance	0.9476	0.9357	0.9490	0.9496	**0.9502**
Mobile	0.9826	0.9814	0.9821	0.9825	**0.9827**

Table 2: Average SSIM of the synthesized.

5. CONCLUSION

In depth-enhanced 3D video formats, the misalignment of depth map edge and texture edges is one of key problems for generating high quality synthesized views, especially for synthesizing large baseline virtual views. The misaligned depth map may cause foreground or background pixels warped to incorrect positions during the 3D image warping process as inappropriate pixels are used to predict the disocclusion. Thus, annoying artifacts always appear in the object boundary of the synthesized views. Conventional pre-filtering methods by smoothing depth map edges are ineffective with drawback of degrading the 3D perception or even destroying the textures in background regions. The proposed method can correct sharp depth map edges lying inside or outside the object boundary of the color image by the Watershed segmentation. The misalignment corrected depth map is further enhanced by a foreground biased dilation filter which makes the whole transitional texture edges assigned with foreground depth values. Experimental results show that the proposed method can significantly improve the rendered image quality. Most annoying artifacts are removed and texture distortion can also be avoided without using any smooth technique compared with the conventional pre-filtering methods.

6. REFERENCES

[1] C. Fehn, "Depth-Image-Based Rendering (DIBR), Compression, and Transmission for a New Approach on 3D-TV", Proc. SPIE Conf. Stereoscopic Displays and Virtual Reality Systems, vol. 5291, pp. 93-104, Jan. 2004.

[2] L. Zhang and W.J. Tam, "Stereoscopic Image Generation Based on Depth Images for 3D TV," IEEE Trans. broadcast., vol. 51, no. 2, pp. 191-199, Jun. 2005.

[3] K. Müller, P. Merkle, T. Wiegand, "3-D Video Representation Using Depth Map", IEEE Proceeding, vol. 99, no. 4, pp. 643-656, Apr. 2011

[4] K. Müller, A. Smolic, K. Dix, P. Merkle, P. Kauff, and T. Wiegand, "View Synthesis for Advanced 3D Video Systems," EURASIP J. Image and Video Process., vol. 2008, p.11, 2008.

[5] C. Vázquez, W. J. Tam and F. Speranza, "Stereoscopic Imageing: Filling Disoccluded Areas in Depth Image-Based Rendering", Proc. SPIE Conf. 3-DTV, Video, Display V, Vol. 6392, 63920D, 2006.

[6] L.M. Po, S. Zhang, X. Xu, Y. Zhu, "A new multidirectional extrapolation hole-filling method for depth-image-based rendering" , Proc. IEEEConf. Image Processing (ICIP), pp. 2637-2640, Sept. 2011.

[7] A. Telea, "An image inpainting technique based on the fast marching method", J. Graphics Tools, vol.9, no.1, pp.25–36, 2004.

[8] Y. K. Park, K. Jung Y. Oh, S. Lee, J. K. Kim, G. Lee, H. Lee, K. Yun, N. Hur, J. Kim, "Depth-Image Base Rendering for 3DTV Service over T-DMB", Signal Processing: Image Communication, vol. 24, pp. 122-136, Jan 2009.

[9] P. Lai, D. Tian, P. Lopez, "Depth map processing with iterative joint multilateral filtering", Proc. IEEE conf. Picture Coding Symposium (PCS), pp.9-12, Dec. 2010.

[10] X. Xu, L. M. Po, K. W. Cheung, K.H. Ng, K. M. Wong, C. W. Ting, "A Foreground Biased Depth Map Refinement", Proc. IEEE Conf. Acoustics, Speech, and Signal Processing (ICASSP), pp.805-808, Apr. 2012.

[11] D. Scharstein, R. Szeliski, and R. Zabih, "A Taxonomy and Evaluatoin of Dense Two-frame Stereo Correspondence Algorithms", Proc. IEEE Workshop Stereo and Multi-Basedline Vision, pp. 131-140, 2001.

[12] Y. K. Park, K. Jung Y. Oh, S. Lee, J. K. Kim, G. Lee, H. Lee, K. Yun, N. Hur, J. Kim, "Depth-Image Base Rendering for 3DTV Service over T-DMB", Signal Process. Image Commun, voi 24(1-2), pp. 122-136, Jan 2009.

[13] K. Zhang, S.C. Chen, D. Whitman, M.L. Shyu, J. Yan, C. Zhang, "A progressive morphological filter for removing nonground measurements from airborne LIDAR data", IEEE Trans. Geosci. Remote Sens., vol. 41, no.4, pp. 872- 882, Apr. 2003.

[14] F. Meyer, "Color image segmentation", Proc. IEEE Conf. Image Processing and its Applications, pp. 303- 306. 1992

[15] G. Bang, M.S. Ko, J. Yoo, G.M. Um, W.S. Cheong, N. Hur, "Boundary Noise Removal and Hole Filling for VSRS3.5 Alpha", ISO/IEC JTC1/SC29/WG11, M19992, Mar 2011.