International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol. 8, No. 4 (2015), pp. 99-112
http://dx.doi.org/10.14257/ijsip.2015.8.4.09

Automatic 2D-to-3D Image Conversion based on Depth Map Estimation

Fan Guo, Jin Tang and Hui Peng*

School of Information Science and Engineering, Central South University, Changsha, Hunan, China
tjin@mail.csu.edu.cn

Abstract

Recent advances in 3D have increased the importance of stereoscopic content creation and processing. Therefore, converting existing 2D contents into 3D contents is very important for growing 3D market. The most difficult task in 2D-to-3D conversion is estimating depth map from a single-view image. Thus, in this paper, we propose a novel algorithm to estimate the map by simulating haze as a global image feature. Besides, the visual artifacts of the synthesized left- and right-views can also be effectively eliminated by recovering the separation and loss of foreground objects in the proposed algorithm. Experimental results show that our algorithm can produce a good 3D stereoscopic effect and prevent the separation and loss artifacts with low computational complexity.

Keywords: *image, 2D-to-3D, conversion, depth estimation, separation, loss*

1. Introduction

A rapid growth of commercialization of 3D display has increased the demands of 3D media contents for supporting full utilities of 3D displays and has aspired humans to experience more realistic and unique 3D effects. In addition to generating better visual experiences than conventional 2D displays, emerging 3D display have many applications, including movies, gaming, photograph, education and so on. However, due to lack of 3D media contents, converting existing 2D contents into 3D contents for growing 3D markets is very necessary and meaningful.

How to generate or estimate the depth map using only a single-view image is the most important and difficult problem in 2D-to-3D conversion. Previous 2D-to-3D conversion methods are mainly divided into two classes: software-based method and depth cues-based method. The software-based method generates 3D content by using stereoscopic conversion tools, such as DDD's TriDef and ArcSoft's Media Converter, to retrieve depth maps. However, the stereoscopic visual effect produced by these tools is not obvious due to the limited information they used for conversion. A more feasible and effective method is the depth cues-based method. This kind of method is based on the key observation, that is, when observing the world, the human brain integrates various heuristic depth cues to generate the depth perception. The major depth perceptions are binocular depth cues from two eyes and monocular depth cues from a single eye [1]. The disparity of binocular visual system helps human eyes to converge and accommodate the object at the right distance. Monocular cues include focus/defocus, motion parallax, relative height/size, and texture gradient, providing various depth perceptions based on human experience. Therefore, humans can also perceive depth from the single-view image/video. The depth cues-based method assigns depth values using image classification [2], machine learning [3], depth from focus/defocus [4], depth from geometric perspective [5], depth from texture gradient, depth from relative height [6] and depth from multiscale local- and global-image features. For example, the computed image depth (CID) method [7] divides a single image into several sub-blocks and uses contrast and blurriness information to

ISSN: 2005-4254 IJSIP

generate depth information for each block. Han, *et al.*, [8] generated the depth map by employing both vanishing points and super-pixels as geometric and texture cues. Cheng, *et al.*, [9] assigned the depth map based on a hypothesized depth gradient model. The method can produce impressive results. However, if the assumption of the global depth does not hold or large foreground objects exists, the method may fail in the cases. Yang, *et al.*, [10] generated a feasible perceptual depth map by using the local depth hypothesis that based on the structural information of the input image and salient regions. However, user interaction is required for this method.

2D-to-3D depth generation algorithms generally face two challenges. One is the depth uniformity inside the same object. The other challenge involves retrieving an appropriate depth relationship among all objects. Generating a depth map from single 2D images is an ill-posed problem. Not all the depth cues can be retrieved from an image. To overcome these two challenges, this work presents a novel algorithm that uses a haze veil to generate a pseudo depth map rather than retrieving the depth value directly from the depth cue. Firstly, the proposed algorithm produces a simulated haze image to represent salient region segmentation. Then the pseudo depth map is automatically generated in single-view image using the transmission information. Experimental results indicate that the proposed algorithm may generate promising stereoscopic results with slight side effects.

2. Our Objective

An image veil is proposed in this paper to segment the saliency region from a single input image. This veil is generated based on the key observation that scene radiance is attenuated exponentially with depth, as indicated by the transmission map. If we can recover the transmission, then we can also recover depth information [11, 12]. Thus, depth information can be measured by the transmission map. To date, numerous studies have been done for estimating the transmission map from a haze image. Therefore, if we may transform a haze-free image into a haze image for the purpose of 2D to 3D conversion, then we may obtain depth information using various existing methods. A simple and effective method for removing haze is inspired by Retinex theory [13]. Based on this theory, the input image I with haze is the product of object reflectance R that can be regarded as haze-free image and scene illumination L that can be regarded as haze veil, that is:

$$I(x, y) = R(x, y) \cdot L(x, y) \quad . \tag{1}$$

where (x, y) is the position coordinate of a pixel. The main idea of the haze removal algorithm is to estimate the haze veil with the mean of the illumination component L that is obtained by convoluting the haze input image with a zero mean Gaussian smoothing function G. This process can be written as follows

$$\hat{L}(x, y) = I(x, y) * G(x, y), \tag{2}$$

$$\tilde{L}(x, y) = \frac{1}{\mathrm{H}\,\mathrm{W}} \sum_{x=1}^{\mathrm{H}} \sum_{y=1}^{\mathrm{W}} \hat{L}(x, y), \tag{3}$$

where $\tilde{L}$ is the estimated haze veil, and H and W denote the height and weight of the image, respectively. The haze veil is subtracted from the original input image in the logarithmic domain to remove the haze effect from the input image, and then the exponential transformation is used to obtain the final haze-removed result $\tilde{R}$, as showed below

$$\tilde{r}(x, y) = \ln I(x, y) - \ln \tilde{L}(x, y) = i(x, y) - l(x, y), \tag{4}$$

$$\tilde{R}(x, y) = \exp(\tilde{r}(x, y)). \tag{5}$$

Figure 1 shows the flowchart of the haze removal algorithm, and Figure 2 illustrates a haze removing example. From these Figures, one can clearly see that the method uses the

illumination component image obtained by the Retinex algorithm to remove the veil layer from the input image.

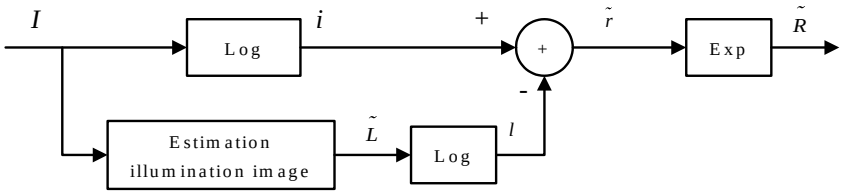


Figure 1. Flowchart of the Haze Removal Algorithm

(a) (b) (c)

Figure 2. Illustration of the Haze Removal Procedure (a) Input Image (b) Estimated Haze Veil (c) Haze Removal Result

Figure 3 illustrates the haze illusion by using two blocks to create the haze effect. Thus, we can deduce that the haze image [see Figure 3(c)] is obtained by adding the haze-free input image [see Figure 3(a)] to the haze veil [see Figure 3(b)]. Once the haze veil is derived from the input image, the haze effect is generated to compute the depth map of the input image.

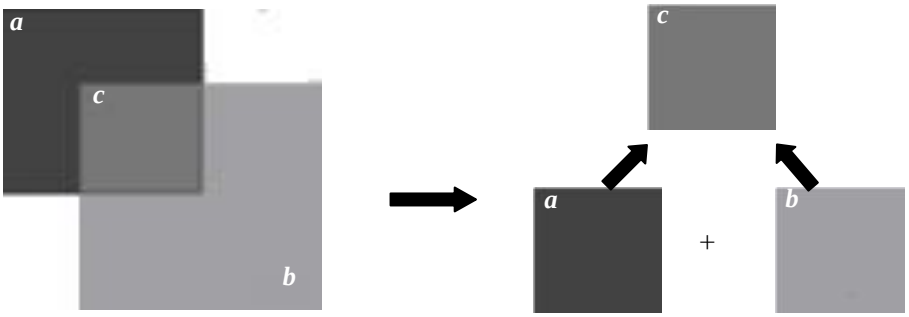


Figure 3. Generation of the Haze Illusion

3. Proposed Algorithm

3.1. Algorithm Procedure

Specifically, the proposed veil algorithm has three steps to automatically convert 2D image into 3D one. The first step is to generate a simulated haze image by adding a haze veil on the haze-free input image, and the haze image is used to represent salient region segmentation. The second step is computing the depth map by using the transmission map estimation in haze removal algorithm, which including initial depth map extraction, refined map estimation and final depth map estimation. The goal of the algorithm is to generate a depth map without using any heuristic depth cues or any user interaction. Finally, the 3D stereoscopic image is generated based on the estimated depth map. The overall procedure of this approach is depicted in Figure 4.

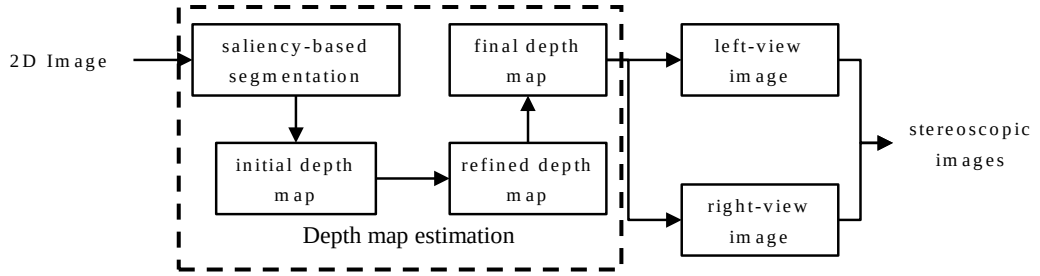


Figure 4. The Overall Procedure for 2D-To-3D Conversion

3.2. Pseudo Depth Map Estimation

In general, the 2D-to-3D conversion from a single image has been assigned to a problem of how to generate depth-map information from 2D images. The depth map estimation is automatic and consists in the following three stages: haze image simulation, initial depth map extraction, refined map estimation and final depth map estimation.

3.2.1. Haze Image Simulation

In this section, we propose a method for simulating a haze image by adding a haze veil on the haze-free input image. The theory behind the haze simulation process is that if the haze veil can be subtracted from the degraded image to removal haze (see Figure 2), we can also simulate the haze image by adding the haze-free input image to the haze veil. In the haze removal experiments, we find that the veil estimated through a mean calculation of illumination component can only handle the uniform haze situation. If the haze is not uniform, the color distortion of haze removal result often occurs. However, it's not always true that the haze is evenly distributed at each position since the natural haze is dependent on the unknown depth information. Thus, we present a new way to estimate a non-uniform distributed haze veil in this paper.

According to the Koschmieder model [14], the apparent luminance of the scene objects at different distance is different, so different haze veil should be assigned according to their position. Therefore, we multiply the uniform veil $\tilde{L}$ by the original image and apply the color inversion operation to obtain a depth-like map. Considering that the intensity of an image reflects the amounts of photons received by every position of an image, furthermore, the smaller the distance between the scene points and the camera, the stronger the intensity will be, thus the haze veil reflected by the depth-like map may be measured by its intensity. Therefore, we extract the intensity component of the depth-like map to produce the haze veil whose distribution is according to real fog density of the scene. Thus the haze veil for the input image, $\tilde{L}'$ is estimated by

$$\tilde{L}'(x,y) = 255 - \omega_1 \times (R(x,y) \cdot \tilde{L}(x,y)), \quad (6)$$

where R is the input image without haze, $\tilde{L}$ is the mean of $\hat{L}$ obtained by Eq. (3) and ω_1 is an adjustment parameter set to 3 to generate a certain amount of haze in the input image. Then, we transform the image $\tilde{L}'$ from RGB to YCbCr color space, and extract the intensity component of the image, which stands for our final haze veil. Once the depth-like haze veil $\tilde{L}'$ is figured out, the haze veil can be added on the real input haze-free image R to get the log-haze image $\tilde{i}$ after the conditions are set. The process is expressed as follows

$$\tilde{i}(x,y) = \ln R(x,y) + \ln \tilde{L}'(x,y). \quad (7)$$

Finally, the simulated haze image I_{haze} can be obtained using exponential transformation, that is $I_{haze} = \exp(\tilde{i}(x,y))$. The saliency region is segmented from non-saliency regions (*e.g.*, the sky and objects or surfaces that are too dark or too light) in the image I_{haze}, such that haze image simulation is actually the image segmentation based on

saliency. For example, Figure 5(a) and Figure 5(b) show the original 2D image and the estimated haze veil. The simulated haze image is shown in Figure 5(c).

(a) (b) (c)

Figure 5. Process of Simulating Haze Image (a) Original 2D Image (b) Haze Veil (c) Simulated Haze Image

3.2.2. Pseudo Depth-map Estimation

Once the haze image I_{haze} is obtained, we can adopt the transmission estimation method that widely used in haze removal to obtain depth information. For this purpose, the dark channel prior [11, 12] and a guided filter [15] are used to estimate the depth map.

Specifically, we first estimate the atmospheric light A for the image I_{haze}. Most algorithms estimate A from the pixels with highest intensities, which is fast but not accurate. He, *et al.*, [11, 12] integrate the atmospheric light estimation with the dark channel prior and it makes the estimation result more accurate. This method is also adopted in this paper.

The depth map is calculated based on the image degradation model [14] and the dark channel prior proposed by He [11, 12]. For the haze image, we first estimate the initial depth map $\tilde{m}(x, y)$, this process can be written as

$$\tilde{m}(x, y) = 1 - \omega_2 \min_{c \in \{R, G, B\}} \left(\min_{(x', y') \in \Omega(x, y)} \left(\frac{I_{haze}^{c}(x', y')}{A^{c}} \right) \right) \tag{8}$$

where I_{haze}^{c} is a color channel of I_{haze}, $\Omega(x, y)$ is a local patch centered at (x, y), and (x', y') is the pixel location that belong to $\Omega(x, y)$. ω_2 is a constant parameter for adjusting the amount of haze for distant objects. The value of ω_2 is set to be 0.95 for all the results reported in this paper.

It should be noticed that there are obvious block effects and redundant details in the initial depth map. In order to handle these deficiencies, we thus use the guided filter [15] and bilateral filter to refine the initial depth map. The detailed estimation process of the final depth-map is described in the following steps.

Step 1. For the initial depth map, we first compute the linear coefficients a_k and b_k for the guided filter:

$$a_k = \frac{\frac{1}{|\omega|} \sum_{(x,y) \in \omega_k} I_{haze}(x, y) \tilde{m}(x, y) - u_k \bar{m}_k}{\sigma_k^2 + \varepsilon}$$
$$b_k = \bar{m}_k - a_k u_k \tag{9}$$

where I_{haze} is the guidance image and $\tilde{m}$ is the input image of the guided filter since the filter is a general linear translation-variant filtering process, which involves a guidance image and an input image [11]. In Eq. (9), ε is a regularization parameter preserving a_k from being too large. u_k and σ_k^2 are the mean and variance of I_{haze} in a window ω_k that

centered at the pixel *k*. $|\omega|$ is the number of pixels in ω_k, and $\bar{m}_k = (1/|\omega|)\sum_{i\in\omega_k} \tilde{m}_i$ is the mean of $\tilde{m}$ in ω_k.

Step 2. Once the linear coefficients (a_k, b_k) are obtained, we can compute the filter output by

$$m'(x,y) = \bar{a}_k \tilde{m}(x,y) + \bar{b}_k \tag{10}$$

where $\bar{a}_k = (1/|\omega|)\sum_{i\in\omega_k} a_i$ and $\bar{b}_k = (1/|\omega|)\sum_{i\in\omega_k} b_i$. $\tilde{m}$ is the initial depth map, and the filter output m' is the refined depth map.

Step 3. A bilateral filter is used here to remove the redundant details for the refined depth map m' since the bilateral filter can smooth images while preserving edges. Thus, the redundant details of the refined depth map m' estimated by the algorithm presented above can be effectively removed. This process can be written as:

$$\hat{m}(\mathbf{u}) = \frac{\sum_{\mathbf{p}\in N(\mathbf{u})} W_c(\|\mathbf{p}-\mathbf{u}\|) W_s(|m'(\mathbf{u})-m'(\mathbf{p})|) m'(\mathbf{p})}{\sum_{\mathbf{p}\in N(\mathbf{u})} W_c(\|\mathbf{p}-\mathbf{u}\|) W_s(|m'(\mathbf{u})-m'(\mathbf{p})|)} \tag{11}$$

where $m'(\mathbf{u})$ is the refined depth map corresponding to the pixel **u**=(*x*, *y*), $N(\mathbf{u})$ is the neighbors of **u**. The spatial domain similarity function $W_c(x)$ is a Gaussian filter with the standard deviation is σ_c: $W_c(x) = e^{-x^2/2\sigma_c^2}$, and the intensity similarity function $W_s(x)$ is a Gaussian filter with the standard deviation is σ_s, it can be defined as: $W_s(x) = e^{-x^2/2\sigma_s^2}$. In our experiments, the value of σ_c and σ_s is set as 3 and 0.4, respectively. Thus, we can obtain the final depth map $\hat{m}(x,y)$.

Figure 6 shows the corresponding initial depth map, refined depth map and the final depth map for the original image in Figure 5(a). From these figures, one can see that the final depth map [see Figure 6(c)] generated using the proposed method reflects the relative positions between scene objects and their neighboring regions. Thus, the map is a pseudo depth map instead of a recovery of real depth information. Generally, the pseudo map is based on the visual attention of mapping the saliency regions from the position close to the viewer while mapping the non-saliency regions from farther positions. Thus, images from the saliency region attract more visual attention and can be regarded as the final depth map for 2D to 3D stereoscopic conversion.

(a) (b) (c)

Figure 6. Process of Estimating the Depth Map (a) Initial Depth Map (b) Refined Depth Map (c) Final Depth Map

3.3 3D Image Visualization using Depth Map-based Rendering

Once the depth map is obtained, the left-view and the right-view images can be synthesized by the following steps. Firstly, we compute the parallax value *Parallax*(*x*, *y*) from each pixel (*x*, *y*) in the estimated depth map. The computation of the parallax value can be written as

$$Parallax(x,y) = \omega_3 \times \left(1 - \frac{\hat{m}(x,y)}{\text{ZPP}}\right), \tag{12}$$

where $\hat{m}(x,y)$ is the final depth map for the single image, ω_3 is the maximum parallax value. As can be seen in Figure 7(a), we can get the value of ω_3 by the similar triangle principle. Specifically, *V* is the distance between screen and viewer, and the inter-ocular distance *E* is about 6.35cm. *D* is the Max depth into the screen, and it is set to 10cm. Thus, the computed ω_3 value is 0.578cm. Next, we should express the value ω_3 in the form of pixel. In our experiment, 17'' monitor (1280 × 1024 Resolutions) is used here, so 1cm on the monitor is corresponding to 38 pixels. Thus, the maximum parallax value ω_3 is approximately 30 pixels for the image having a width size approximate to 1000. The zero parallax plane (ZPP) is set as the region with the depth value of *Th*, which is computed by $Th = \max(\hat{m}(x,y)) - 10$ to prevent separation and loss of artifacts.

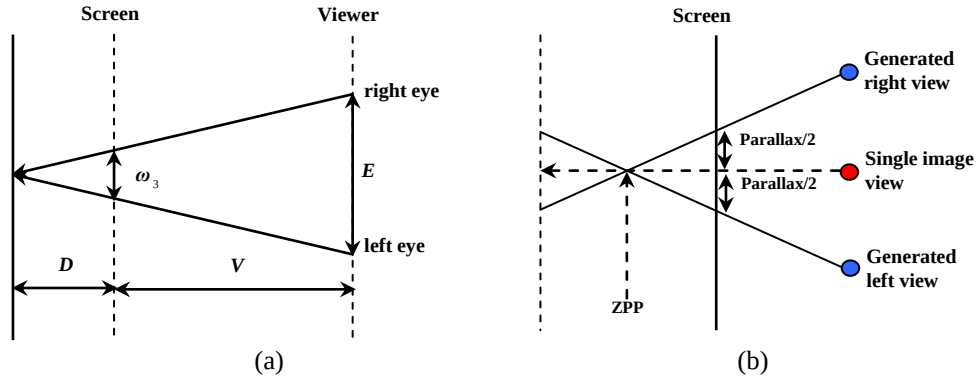


Figure 7. Stereoscopic Generation (a) Max Parallax Computation (b) Right View and Left View Generation

Then, we consider the input image as the center view of the stereoscopic pair, as shown in Figure 7(b). In order to produce the left or right-view image, each pixel of the input image is shifted by the amount of $Parallax(x,y)/2$ to left or right direction. The missing pixels at the image boundary will be filled to synthesize a right-view or left-view image with the same size of input original image. At last, the anaglyph images can be generated by using these left or right-view images. Viewers can feel the sense of depth with the help of anaglyph glasses (Left: red, Right: cyan) to see these images. For example, Figures 8(a) and 8(b) is the Left-/right-view image produced by using the proposed approach for the input 2D image [Figure 5(a)], and Figure 8(c) shows our final 3D conversion result.

(a) (b) (c)

Figure 8. Process of 3D Image Visualization (a) Left-view Image (b) Right-view Images (c) Stereoscopic Image

Another example for illustrating the process of generating stereoscopic images is shown in Figure 9. Here, virtual left- and right-eye views [Figure 9(d) and Figure 9(e)] are

rendered using the depth map [Figure 9(c)] obtained from the proposed technique to evaluate performance. In most cases, the pseudo depth map is generated based on the visual attention, as shown in Figure 6(c). However, it should be noticed that the estimated depth area corresponding to some people in Figure 9(c) seems to be incorrect by saliency detection, but the approach may still provide good 3D perception, as shown in Figure 9(f).

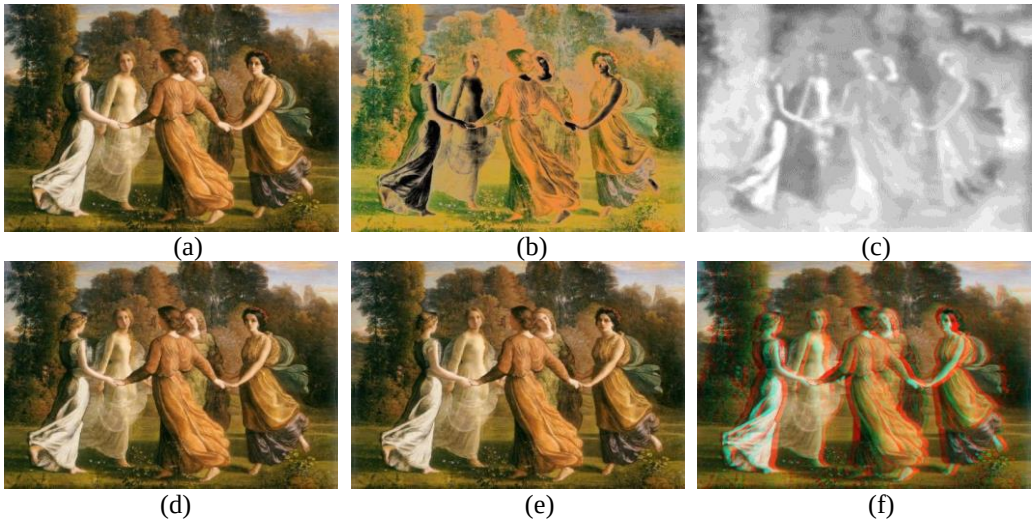


Figure 9. Process of Generating Stereoscopic Images (a) Input Image (b) Simulated Haze Image (c) Estimated Depth Map (d) and (e) Left-view and Right-view Images, Respectively (f) Stereoscopic Conversion Result

From our experiment, we find that perceptual and cognitive consistency is the most important factor in 2D to 3D stereoscopic conversion. Although the depth map estimated by the proposed method just provides a new way to simulate the virtual left- or right-view image and it does not recover the real depth information, human visual system may overwrite the depth perception and make the depth cues consist with our daily life experience. In other words, our eyes are 'cheated' by our brain. To our knowledge, there has been no formal investigation that has attempted to explain why the side effects of these methods are hard to be discovered even the depth is inverted. Therefore, 2D to 3D stereoscopic conversion does not require accurate metric depth map since human visual perception can generate correct results even when the depth map of object is inverted. Besides, when the light gradient and the relative position between the salient objects and other parts of the scene are preserved, human visual system may overwrite the depth perception with daily life experience. Hence, the light gradient and the relative position between the saliency region and non-saliency region play an important role on the depth perception. This could also explain the side effects of proposed algorithm are hard to discover when the saliency region or the depth is inverted. Our observations on Figure 9(c) and Figure 9(f) confirm this conclusion.

4. Separation and Loss Problem

Despite the advantages mentioned above, in our experiment we find that the left- and right-views of our proposed method are sometimes not perfectly synthesized. That's because the estimated depth image may sometimes cause visual artifacts such as the separation and loss of foreground objects, as shown in Figures 10(b) and 10(c). Here, the separation occurs when a foreground object which has a larger parallax value moves further than background and the loss occurs in the direction in which foreground objects move during the virtual left- and right-views synthesis procedure. Although many works have been done to synthesize virtual view images, conventional algorithms always

assume a perfect boundary match between a color image and the corresponding depth image, so the separation and loss effect are not considered [16].

(a) (b) (c)

Figure 10. Synthesized Virtual View Images with the Separation and the Loss Effects (a) Original Image (b) Left-view Image (c) Right-view Image (Red Ellipse: Separation Area, Green Ellipse: Loss Area)

To solve the separation and loss problems, we perform a one-tap IIR filter to raise the parallax values of the foreground region of the separation area or the parallax values of the background region of the loss area to parallax that are similar to those of the background or the foreground. Since the pixel only move horizontally depending on the value of the corresponding parallax during the view synthesis step, the process can thus be written as:

$$L(i, j) = (1 - \lambda)\, parallax(i, j) + \lambda L(i, j + 1) \quad (13)$$

$$R(i, j) = (1 - \lambda)\, parallax(i, j) + \lambda R(i, j + 1) \quad (14)$$

where $L(i, j)$ and $R(i, j)$ are the intensity values of the left- and right-view images, respectively. λ $(0 \le \lambda \le 1)$ is a parameter that controls the parallax propagation strength. In our experiment, we set $\lambda = 0.95$. By using this method, we can reduce the separation and loss in the synthesized view with low computational complexity, as shown in Figure 11.

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

Figure 11. Synthesized Left- and Right-view Images Obtained by using One-tap IIR Filter (a) Left-view Image (b) Zooming Region for the Separation Area (c) Right-view Image (b) Zooming Region for the Loss Area (Red Ellipse: Interested Area)

5. Experimental Results

5.1 Qualitative Comparison

Figure 12 shows some experimental results for single image, including data on seven sets of the original images, depth maps and red-cyan images. The method generates fair stereoscopic images which offer a good depth effect to viewers, as shown in Figures 12(a-g).

Figure 12. Examples of 2D to 3D Stereoscopic Conversion Results for a Set of Images

We also evaluate the visual quality of the proposed algorithm by comparing with other 2D-to-3D conversion approaches: the commercial software of ArcSoft's Media Converter 7 and the Cheng's method [9]. Figure 13 shows the comparison of generated Red-Cyan images for the three testing images. From the figure, one can see that the 3D effects produced by Media Converter 7 is not obvious compared with the results generated by other two algorithms, since the simple and easy-to-use media converter mainly utilizes the color cue to extract depth information. Cheng's algorithm [9] can produce vivid and realistic visual effects. However, a hypothesized depth gradient model is required for the method. When the assumption of the global depth does not hold or large foreground objects exist, the depth gradient hypothesis is invalid. Notice that the results obtained with

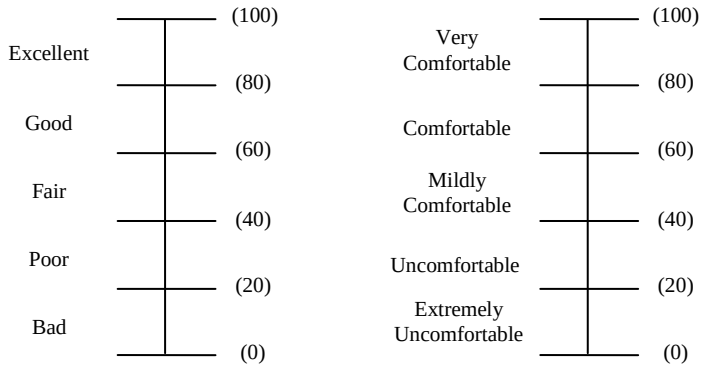
our algorithm seems visually close to the results obtained by Cheng's algorithm without using any hypothesis or user interaction, thanks to the haze image simulation.

Figure 13. Comparison of Red-Cyan Images for Three Testing Images. First column: Original 2D image. Second Column: Results Obtained by Arcsoft's Media Converter 7. Third column: Results Obtained by Cheng's method [9]. Forth Column: Results Obtained by the Proposed Method

5.2. Quantitative Evaluation

To quantitative assess and rate the different 2D-To-3D conversion methods, the seven images in Figure 12 and the three testing images in Figure 13 were used to perform the quantitative evaluation. Both depth quality and visual comfort were evaluated using a single - stimulus presentation method that is a slightly modified version of that described in previous works [9, 17]. Depth quality measures the sense of depth felt by the viewer, and visual comfort refers to the subjective sensation of comfort that accompanies the physiological change [18]. Thus, depth quality and visual comfort can be both measured by asking the viewer to report his/her level of perceived depth quality and visual comfort.

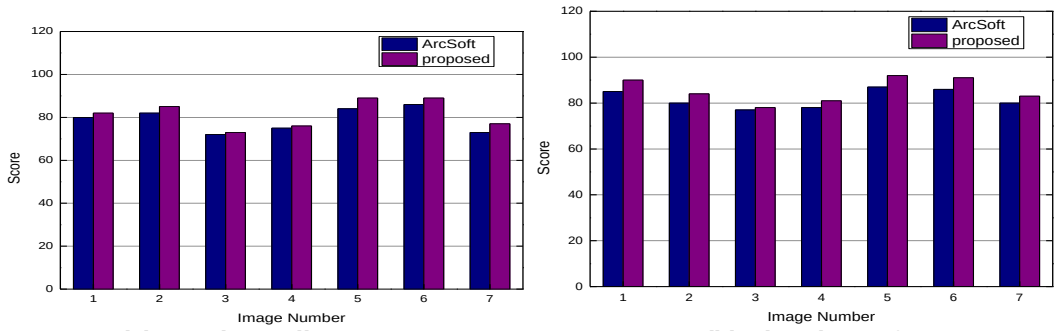
The synthesized Red-Cyan images were viewed with anaglyph glasses. The subjective evaluation was performed by 10 individuals with normal or correct-to-normal visual acuity and stereo acuity. The participants watched the stereoscopic images in a random order and were asked to rate each image based on two factors, depth quality and visual comfort. The overall quality of depth quality was assessed using a five-segment scale, as shown in Figure 14(a), and that for visual comfort was assessed by Figure 14(b).



(a) (b)

Figure 14. Rating Scales used for Evaluation (a) Depth Quality and (b) Visual Comfort

Figures 15(a) and 15(b) show the values of the two factors acquired by Media Converter 7 and our proposed algorithm for the test images shown in Figure 12. From Figure 15, one can deduce that in the depth quality, viewers can feel a better sense of depth using the proposed algorithm compared to the results obtained by ArcSoft's software.

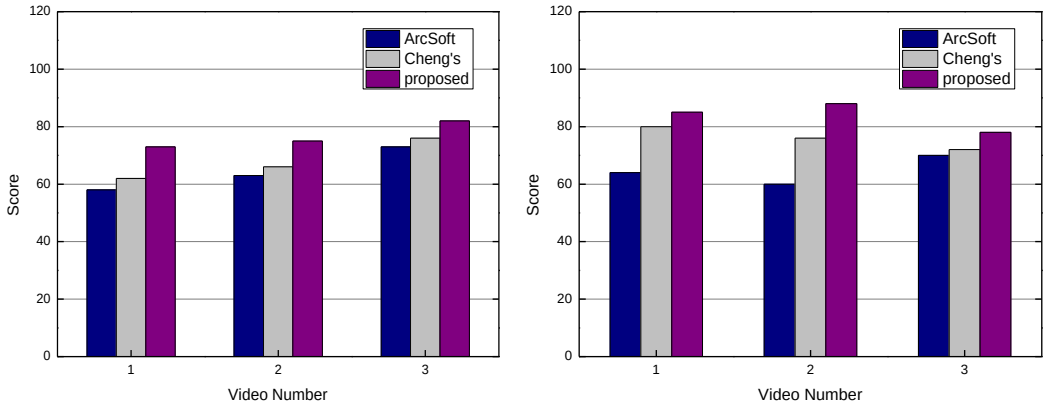


(a) Depth Quality **(b) Visual Comfort**

Figure 15. Quantitative Evaluation Results for Test Images Shown in Figure 12

For the three testing images shown in Figure 13, we compare our results with other two conversion methods: ArcSoft's software and Cheng's method [9]. Evaluation results demonstrate that the proposed algorithm produces a similar or better 3D visual effects compared with Cheng's method for the input image, as shown in Figure 16. One can clearly see that the proposed algorithm has the best scores in depth quality and visual comfort. This confirms our observations on Figure 13. Furthermore, when the images vary in the lighting source, the proposed algorithm still works well in this case. This is because the illumination on the surface is well-preserved in the proposed method, so human can have depth perception with daily life experience. Thus, we can rank the three

methods for the test images in decreasing order with respect to visual comfort: the proposed method, Cheng's method and ArcSoft's software.



(a) Depth Quality **(b) Visual Comfort**

Figure 16. Quantitative Evaluation Results for the Test Videos Shown in Figure 13

6. Conclusion

A novel and automatic method was proposed to generate a pseudo depth map in single-view image using the estimated haze veil. A haze image was simulated by adding a haze veil on the input image to represent salient region segmentation, and then it estimated pseudo depth map by using the transmission estimation method in haze removal algorithm. Using the depth map, left- and right-view images were synthesized, and finally the stereoscopic images were generated to provide a sense of depth to the viewers with the help of anaglyph glasses. Besides, the separation and loss artifacts of the synthesized results can also be effectively prevented with low computational complexity. The whole process of the proposed algorithm could be performed automatically without any heuristic cues or user interaction. The future work includes improving the computational efficiency of the proposed algorithm using GPU implementation or parallel computation.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 91220301), the China Postdoctoral Science Foundation (No. 2014M552154), the Hunan Postdoctoral Scientific Program (No. 2014RS4026), and the Postdoctoral Science Foundation of Central South University (No. 126648).

References

[1] W. J. Tam and L. Zhang, "3D-TV content generation: 2D-to-3D conversion", Proceedings of IEEE International Conference on Multimedia and Expo, **(2006)** July 9-12, Toronto, Canada.
[2] S. Battiato, S. Curti, E. Scordato, M. Tortora and M. La Cascia, "Three-Dimentional Image Capture and Applications V1", vol. 5302, **(2004)**.
[3] D. Hoiem, A. A. Efros and M. Hebert, ACM Transactions on Graphics, vol. 3, no. 24, **(2005)**.
[4] J. Park and C. Kim, Proceedings of SPIE, vol. 6077, **(2006)**.
[5] Y.-M. Tsai, Y.-L. Chang and L.-G. Chen, "Block-based vanishing line and vanishing point detection for 3d scene reconstruction", Proceedings of International Symposium on Intelligent Signal Processing and Communications, **(2006)** December 12-15, Yonago, Japan.
[6] Y. J. Jung, A. Baik, J. Kim and D. Park, "A novel 2D-to-3D conversion technique based on relative height depth cue", Proceedings of SPIE - The International Society for Optical Engineering, **(2009)** January 19-21, San Jose, United states.
[7] H. Murata, SID Digest of Technical Papers, vol. 1, no. 29, **(1998)**.

[8] K. Han and K. Hong, "Geometric and texture cue based depth-map estimation for 2D to 3D image conversion", Proceedings of IEEE International Conference on Consumer Electronics, **(2011)** January 9-12, Las Vegas, United states.
[9] C.-C. Cheng, C.-T. Li and L.-G. Chen, IEEE Transactions on Consumer Electronics, vol. 3, no. 56, **(2010)**.
[10] N.-E. Yang, J.W. Lee and R.-H. Park, "Depth map generation from a single image using local depth hypothesis", Proceedings of IEEE International Conference on Consumer Electronics, **(2012)** January 13-16, Las Vegas, United states.
[11] K. He, J. Sun and X. Tang, "Single image haze removal using dark channel prior", Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, **(2009)** June 20-25, Miami, United states.
[12] K. M. He, J. Sun and X. O. Tang, IEEE Trans. Pattern Anal. Mach. Intell., vol. 12, no. 33, **(2011)**.
[13] E. H. Land, Proc. of the National Academy of Science USA, vol. 83, **(1986)**.
[14] N. Hautiere, J.-P. Tarel, J. Lavenant and D. Aubert, Machine Vision and Applications, vol. 1, no. 17, **(2006)**.
[15] K. M. He, J. Sun and X. O. Tang, "Guided Image Filtering", Proceedings of 11th European Conference on Computer Vision, **(2010)** September 5-11, Berlin, Germany.
[16] H.-W. Cho, S.-W. Chung, M.-K. Song and W.-J. Song, "Depth-image-based 3D rendering with edge dependent preprocessing", Proceedings of IEEE International Midwest Symposium on Circuits and Systems, **(2011)** August 7-10, Seoul, Korea.
[17] "ITU-R Recommendation BT.500-10", Methodology for the subjective assessment of the quality of television pictures, **(2000)**.
[18] W. J. Tam, F. Speranza, S. Yano and H. Ono, IEEE Transactions on broadcasting, vol. 2, no. 57, **(2011)**.

Authors

Fan Guo, received the B.S. degree in computer science and technology from the Central South University (CSU), Changsha, China, in 2005, and the M.S. degree and the Ph.D. degree in computer application technology from CSU, Changsha, China, in 2008 and 2012, respectively. Currently, she is a postdoctoral fellow and Lecturer with the School of Information Science and Engineering, CSU. Her main research interests include image processing, pattern recognition and virtual reality.

Jin Tang, received the B.S. degree and the M.S. degree from Peking University, Beijing, China, in 1987 and 1990, respectively, and the Ph.D. degree in pattern recognition and intelligence system from Central South University (CSU), Hunan, China, in 2002. He is currently a Professor in the School of Information Science and Engineering, CSU, Changsha. His current research interests are focused on image processing, pattern recognition and computer vision.

Hui Peng, received the B.Eng. and M.Eng. degrees in Control Engineering from Central South University (CSU), Changsha, China, in 1983 and 1986, respectively. He obtained the Ph.D. degree in Statistical Science from the Graduate University for Advanced Studies, Hayama, Japan, in 2003. He is currently a Professor in the School of Information Science and Engineering, CSU, Changsha. His research interests include nonlinear system modeling, statistical modeling, system identification, parameter optimization, signal processing, predictive control, robust control and process control.

