Science Heritage Journal (GWS)
DOI: http://doi.org/10.26480/gws.01.2019.24.31

ZIBELINE INTERNATIONAL PUBLISHING

ISSN: 2521-0858 (Print)
ISSN: 2521-0866 (Online)
CODEN: SHJCAS

CrossMark

RESEARCH ARTICLE

RESEARCH ON ARBITRARY VIRTUAL VIEW IMAGE SYNTHESIS METHOD OF TWO VIEWS

Sheng Nan[1,2]

[1]Huaibei Normal University, China
[2]Huaibei Vocational and Technical College, China
**Corresponding Author Email: wanxi.peng@qq.com*

ARTICLE DETAILS

Article History:

Received 1 February 2019
Accepted 19 March2019
Available online 22 March 2019

ABSTRACT

This paper proposes an improved virtual view image synthesis algorithm to overcome the drawbacks of conventional virtual view synthesis technologies in dealing with holes and artifacts. First, conduct edge detection and preparation operation on obtained depth images to reduce the holes caused by the step change of pixel value; second, use the Euclidean distance between pixels and the depth information of image to roughly integrate images; finally, conduct morphological swelling treatment on the obtained virtual images to fill the rest holes and remove artifacts. According to results of simulation experiment, the improved algorithm outperforms those applied in other papers in filling holes and removing artifacts in in the course of virtual view image synthesis.

KEYWORDS

Two views, morphological processing, image synthesis, DIBR

1. INTRODUCTION

At present, although 3D technology has already been sophisticated to some degree, its application still relies on wearable devices like 3D glasses. As glass-free 3D and glass-free virtual imaging technologies develop, there are space televisions requiring no wearable devices, such as MTV and FTV. However, these televisions cannot be viewed from multiple angles, which is a question in display technology and spatial spot calibration technology [1]. To realize multi-angle visual watch, large amount of location information must be labeled in space, involving challenges like collecting detailed information of spatial points, saving data and transmitting all sorts of data. The current glass-free labeling technology does not support multi-viewpoint watching.

Using spatial location to conduct image synthesis of virtual views is a technology to capture screen images and then label the virtual information of a location in the space in order to generate virtual view synthesis image [2]. However, the commonly used virtual view image synthesis technology is defective in dealing with holes and image artifacts, causing great problems on the quality of synthesized image.

To deal with the above-mentioned defects, this paper puts forward a new algorithm, including the following aspects: (1) Conduct edge detection and preparation operation on obtained depth images to modify the edge of depth image; (2) Introduce the Euclidean distance between pixels, considering the current spatial location of two view image and the according location of Euclidean distance; (3) Scan all of the image pixels and apply "Sudoko" to fill holes. Finally, prove the effectiveness of experiments with the actual fusion effect and quantitative analysis [3].

2. VIRTUAL VIEW SYNTHESIS METHOD BASED ON DIBR

2.1 DIBR Technology

The principle of DIBR technology is to use the two pixels on the image captured by camera, and then transform all pixels in the image to virtual ones through changing their spatial coordinates on the basis of certain reference and projection requirements, thereby synthesizing virtual views [4].

Transformation from real pixels to virtual ones by applying DIBR technology entails spatial coordinates conversion. The converted size and range is decided by the depth of the former image. Suppose the spatial coordinate of the location of camera equates the world coordinate system, and the two types of cameras represent virtual and real views respectively. The projections of any points $M = [X\ \ Y\ \ Z]^T$ in the spaces of the two cameras are $m = [u\ \ v]^T$ and $m' = [u'\ \ v']^T$. The conversion equations for the two coordinate systems are:

$$\tilde{m} \cong AP_n \tilde{M} \tag{1}$$

$$\tilde{m}' \cong A' P_n D' \tilde{M} \tag{2}$$

$\tilde{m} \in R^{3\times1}$, $\tilde{m}' \in R^{3\times1}$ and $\tilde{M} \in R^{4\times1}$ mark the coordinates in the spatial coordinates in the above-mentioned coordinate systems. They are the coordinate system of labeled camera, of virtual views to be synthesized and of the real world. The sign $\cong$ signifies that the equation holds when scale transformation parameter is not zero; $A \in R^{3\times3}$ and $A' \in R^{3\times3}$ stand for the internal parameter set of two cameras. $D' \in R^{4\times4}$ refers to the transformational relation of two cameras in the course of spatial coordinate projection, namely $D' = \begin{bmatrix} R & t \\ 0_3^T & 1 \end{bmatrix}$. The function of $P_n \in R^{3\times4}$ is to normalize the converted parameters.

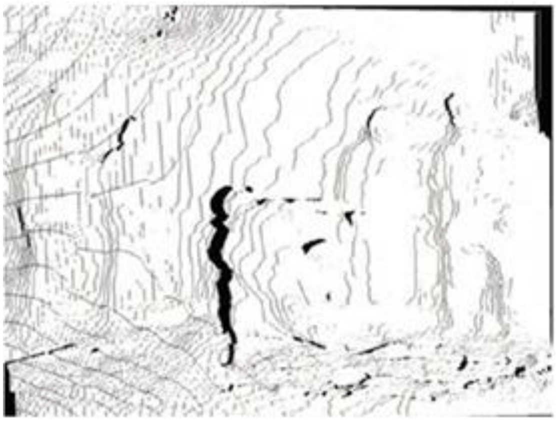
2.2 Holes

As shown in figure 1, take the video sequences on *Breaking Dancer* of Microsoft research as an example to illustrate the problem of holes that is likely to occur in the course of virtual view image synthesis. Figure 1 (a) is one frame of the current of color video captured by the camera; figure (b) is the coordinating in-depth image of figure (a); figure (c) is the virtual point synthesis image generated by conducting spatial location labeling on figure (a); figure (d) is the hole mark figure obtained by conducting binarization processing on figure (c). According to figure (c) and (d), holes can be easily identified in part of pixels after the binarization processing on the virtual view synthesis image obtained by converting the spatial coordinate of original color image. It can be deduced that the figure (c) obtained with traditional conversion methods are severely defective in image detail and quality [5].

(a) Calibration viewpoint color map

(b) Calibration viewpoint depth map

(b) Color map virtual viewpoint



(c) Virtual viewpoint empty mask pattern

Figure 1: Color image to a virtual viewpoint composite image conversion

3. IMPROVED VIRTUAL VIEW SYNTHESIS ALGORITHM

This paper proposes an improved image synthesis method to make up for the deficiency of conventional algorithms. The detailed process is as follows: first, remove the edge of the depth image converted by color image; second, project the modified depth image to spatial coordinate system to obtain virtual synthesized image; third, conduct binarization processing on the virtual view synthesis image obtain and apply swelling treatment on the pixels around the holes. The virtual view synthesis image can be obtained by filling the holes in synthesized image [6]. Figure 2 shows the whole process of synthesized algorithm used in tis paper. The step 1, 5, and 6 are new algorithms added to make up for the drawbacks of conventional algorithms.

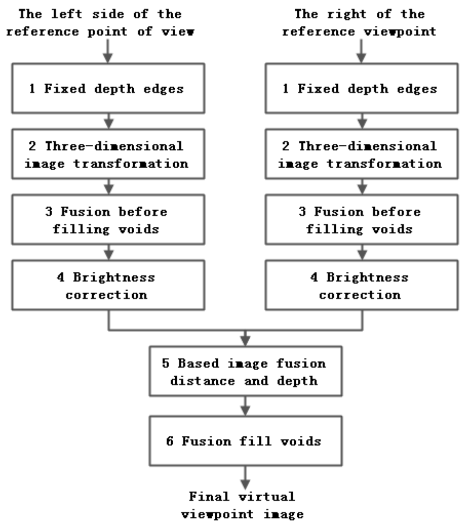


Figure 2: Algorithm flow of the proposed method

3.1 In-depth edge inspection and modification

The values of all pixels in the color images can lead to the coordinating depth image representing the space depth distance. As there is huge distance between the locations of object in the spatial coordinate system, the according depth value will greatly vary. Depth step and transition will occur at some junctions. These factors can cause huge holes in the post-synthesized virtual view image [7].

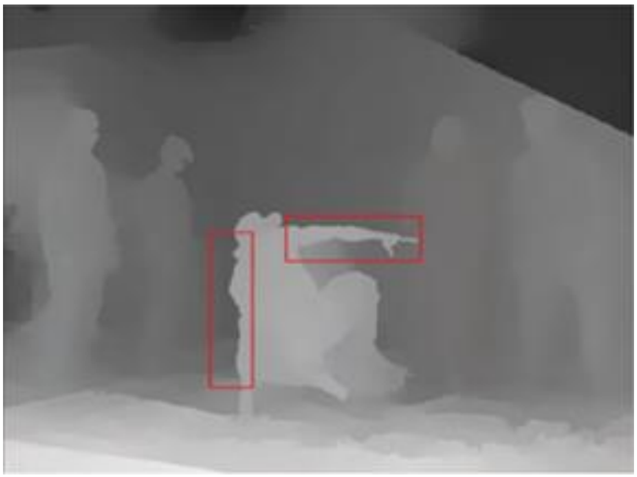
Image synthesis of two views involves image treatment of two angles. This paper uses left perspective to briefly describe the edge inspection and modification algorithm of depth image:

(1) Scan all pixels of depth image from left to right; places where pixel value is found to hop are regarded as edge pixel (u,v);
(2) Conduct spatial conversion of RGB and HSV on the color image obtained by camera;
(3) Take (u,v) as initial search point and search from left to right. Every N pixel is regarded as a traverse. Label the pixel $(u,v-n)$ with most apparent phase step as the initial point of the next traverse, and set all the depth value of pixels between $(u,v-n)$ and (u,v) as the depth value of the former (u,v). In the same way, the depth value modification on the right side of the image can be obtained through scanning from left to right. Set $g(u,v)$ as the depth value difference of the pixels in front of and behind (u,v), then

$$g(u,v) = L(u,v) - L(u,v-1) \tag{3}$$

The $L(u,v)$ in equation (3) represent the depth value of the present pixels.

Edge inspection and modification of depth image can revise the pixels in phase step and saltus step while retaining the basic information of original image, thereby decreasing the holes caused by image synthesis. Figure 3 present the effect comparison before and after edge inspection and modification [8].



(a) Get the camera depth map

(b) After edge detection and adjustment of the depth map

Figure 3: Before and after comparison of edge detection and depth map adjustment

3.2 Spatial image conversion

The standard for the equation (1) and (2) that are describe before is space coordinate system. Spatial visual limit should be broken on the basis of two views so as to realize image synthesis of any views. In standard space coordinate system, the location of camera is labeled as $\tilde{C}$:

$$\tilde{C} = -inv(R)t \tag{4}$$

Equation (4) can lead to 3D space coordinates:

$$\begin{cases} X = depth \bullet \dfrac{x - C_x}{z - C_z} + C_x , \\ Y = depth \bullet \dfrac{y - C_y}{z - C_z} + C_y , \\ Z = depth + C_z 。 \end{cases} \quad (5)$$

$depth$ in equation (5) is the depth value of depth image pixels obtained by camera. (x, y, z) is the conversion coordinate when spatial coordinate projection conversion is conducted; (C_x, C_y, C_z) is the labeled coordinate of camera's location.

3.3 Small hole filling

Figure 1 has described why conventional algorithms cause holes in the course of image synthesis. To solve the problem, this paper puts forward a way to fill the hole. The detailed algorithm is as follows:
First, traverse the whole depth hole masking image, label the first scanned empty pixel as a. Then, centering a, endow the maximal depth value of the pixels in Sudoko to pixel a so as to fill the hole of pixel in the depth masking image. Figure 4 present the virtual view synthesis image and according empty masking image after the hole is filled [9].

(a) Virtual view synthesis fill the void after FIG

(b) Fill the void after the mask pattern

Figure 4: Fill the void after the renderings

3.4 Improved image fusion method

Currently, the distance between spatial pixels is used to realise image synthesis. The algorithm of the method can be described with equation (6) and (7):

$$I_V(u,v) = \begin{cases} 0, & I_L(u,v) = 0 \;\&\&\; I_R(u,v) = 0, \\ I_L(u,v), & I_L(u,v) \neq 0 \;\&\&\; I_R(u,v) = 0, \\ I_R(u,v), & I_L(u,v) = 0 \;\&\&\; I_R(u,v) \neq 0, \\ (1-\alpha) I_L(u,v) + \alpha I_R(u,v), & I_L(u,v) \neq 0 \;\&\&\; I_R(u,v) \neq 0; \end{cases} \quad (6)$$

$$\alpha = \frac{|t_V - t_L|}{|t_V - t_L| + |t_V - t_R|} \quad (7)$$

$I_L(u,v)$ and $I_R(u,v)$ in the above equation is the pixel values of the two views of (u,v) whose grey levels have been modified. If the values of $I_L(u,v)$ and $I_R(u,v)$ are zero, there are holes in coordinating pixels. In addition, t_V, t_L and t_R are the amplitude index whose coordinates have been converted in multiple perspectives, and $I_V(u,v)$ is the depth value of a pixel (u,v) after virtual view image synthesis.

This paper uses the improved image synthesis method in which the Euclidean distance between two pixels and the depth value of pixels is used to conduct image synthesis of virtual views. Compared with conventional synthesis methods, this paper puts forward a new algorithm that has two additional approach sides: spatial distance and image depth. The algorithm is shown in the equation below:

$$I_V(u,v) = \frac{(1-\alpha)\cdot d_L(u,v)\cdot I_L(u,v) + \alpha \cdot d_R(u,v)\cdot I_R(u,v)}{(1-\alpha)\cdot d_L(u,v) + \alpha \cdot d_R(u,v)} \quad (8)$$

$$\alpha = \frac{|t_V - t_L|}{|t_V - t_L| + |t_V - t_R|} \quad (9)$$

The $d_L(u,v)$ and $d_R(u,v)$ in equation (8) represent the depth values of a pixel in original image converted in various perspectives.
In order to ensure that a small number of holes remain after image synthesis, this paper adopts the working mechanism of swelling and erosion in morphology to conduct morphological treatment on the according depth mask figure [10]. Virtual synthesis of treated mask figure can effectively deal with artifacts and remaining holes. Effect figure is shown below:

(a) After synthesis after morphological processing of FIG

(b) Final composite image corresponding masks

Figure 5: The effect of improving the use of image fusion algorithm implemented

4. RESEARCH RESULT ANALYSIS

In order to verify the effectiveness of the method put forward in this paper, *Breaking Dancers*, a video test sequence of Microsoft is used as the research object. This test requires three cameras, two of which are respectively used to label the video imaging of left and right views [11]. The other one is used to label the location relation after virtual view image synthesis. The computer processor used in this paper is i5-3470, whose dominant frequency is 3.2GHz and memory is 16G. To ensure the reliability of the experiment quantitatively and qualitatively, this paper compares experimental results in the following two aspects:

4.1 Comparison of fusion effects

To compare the effects of synthesis, this paper compares the results of three methods, conventional virtual view synthesis method, *Yang* mentioned in literature and the method proposed in this paper. As shown in figure 6, figure （d） is the original color image obtained by labeled camera. Figure (a), (b) and (c) respectively refer to the results of three methods. It is found that holes in figure (a) is relatively more than in figure (b) and (c). Generally speaking, the effect of figure (c) is better than figure (b) in detail treatment and reproduction degree.

Figure 7 indicates that the method proposed in this paper far outperforms the *Yang* in coping with hop of pixel value at junction generated by the obviously different depth values. In figure (a), there is obvious color difference at junctions; in figure (b),the shape of a part of human body changes; in figure (c), there is no such problems. The figure 8 presents the virtual view synthesis effect of *Yang* and the method advised in this paper.

(a) Original DIBR method

(b) Yang method

(c) Method of this paper

(d) Real image

Figure 6: Comparative effects of two methods of synthesis and the original image

(a) Local contrast (1)

(b) Local contrast (2)

Figure 7: Yang method, the proposed method and the effect of amplifying the original image of the three comparison

(a) Comparative Results of Article 2

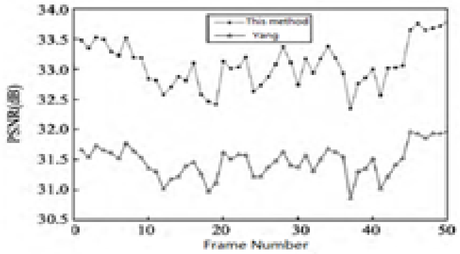
(b) Comparative Results of Article 16

(c) Comparative Results of Article 50

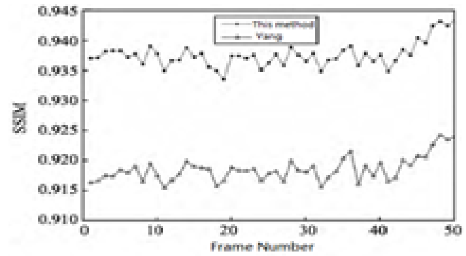
Figure 8: Contrast effect of using of Yang method and the method of paper to synthesis other image frames

4.2 Quantitative analysis and comparison

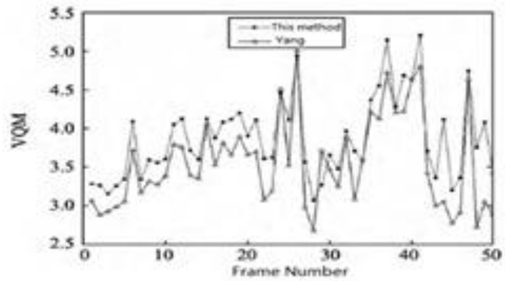
The three standards commonly used to judge evaluate the quality of image are peak signal-to-noise ratio *PSNR*, structural similarity index measurement *SSIM* and video quality measurement *VQM*. With these three standards, this paper compares several virtual view synthesis methods in a quantitative manner. The experiment collects the first 50 frames of original color images in the camera as test sample. Data comparison shown in figure 9 shows that the method applied in this paper is better than *Yang* in the three standards.



(a) Confrontation of PSNR



(b) Confrontation of SSIM



(c) Confrontation of VQM

Figure 9: Qualitative comparison of the two algorithms

Figure 1 lists the comparative data based on the three standards of the first 50 frames color images processed by the algorithm in this paper and *Yang* algorithm. It is found that the average value, maximal value and minimal value of the algorithm applied in this paper are all better than the *Yang*.

Table 1: Two algorithms comparison of data in three evaluation categories

Method Comparison	PSNR			SSIM			VQM		
	average value	Max value	Min value	average value	Max value	Min value	average value	Max value	Min value
Yang	32.1352	30.7261	30.3289	0.7362	0.9032	0.9157	3.7827	5.1896	2.9861
This paper	32.7862	31.7836	32.1728	0.8261	0.8927	0.9411	3.2896	4.9861	2.5618

Comparing the data in figure 2 shows that the values of *PNSR* and *SSIM* obtained in the improved algorithm equate or overtake the experimental results used in other research papers. Although the *SSIM* of the algorithm used in a paper exceeds that of this paper by 0.0067, its *PNSR* is lower by 0.794.

Table 2: Comparison with other literature algorithms for quantitative

Method Comparison	*PNSR*	*SSIM*
Methods of Literature [6]	29.5289	0.9137
Methods of Literature [7]	30.1013	n/a
Methods of Literature [11]	31.7523	0.9480
Methods of this paper	30.9613	0.9413

5. CONCLUSION

Based on conventional virtual view image synthesis algorithm, this paper comes up with an improved one that applies the Euclidean Distance between spatial pixels and the depth information to conduct image synthesis between two views to make up for the defects of conventional methods in dealing with holes and artifacts. First, the algorithm conducts edge inspection on the obtained depth image to pretreat the prixels showing phase steps of luminosity on the edge, thus lowering the possibility of causing empty pixels; then project the view images coordinated with the cameras labeled by the two views based on spatial coordinate conversion mechanism; fill the holes and modify luminosity of the obtained images; at last, use the Euclidean Distance and depth information conversion to synthesize images, and conduct swelling treatment and artifacts removal with morphology on the synthesized images. Simulative experiments show that the method proposed in this paper can deal with the problems that are likely to occur in treating image synthesis. In addition, experiment data also show that the effect of the algorithm used in this paper is better than any used in other papers.

REFERENCES

[1] Bavirisetti, D.P. 2013.Two-scale image fusion of visible and infrared images using saliency detection[J]. Infrared Physics and Technology.

[2] Li, H.F. 2011. Performance improvement scheme of multifocus image fusion derived by difference images[J]. Signal Processing.

[3] Xi, M. 2013. Depth-image-based rendering with spatial and temporal texture synthesis for TV[J]. EURASIP Journal on Image and Video Processing.

[4] Suryanarayana, M. 2012. Muddala.Spatio-temporal consistent depth-image-based rendering using layered depth image and inpainting[J]. EURASIP Journal on Image and Video Processing,2012.

[5] Azevedo, R.G.A. 2010. Real-Time Depth-Image-Based Rendering for TV Using OpenCL[J]. Maitre Matthieu.

[6] Yang, Y. 2014. Multi-focus Image Fusion Using an Effective Discrete

Wavelet Transform Based Algorithm[J].Measurement Science Review.

[7] Bavirisetti, D.P. 2012. Infrared and visible image fusion scheme based on NSCT and low-level visual features[J]. Information Fusion.

[8] Abdipour, M. 2006. Multi-focus image fusion using sharpness criteria for visual sensor networks in wavelet domain[J]. Computers and Electrical Engineering.

[9] Hrušovský, B. 2011. Error Concealment Algorithms Applied on Multi-View Video Sequences[J]. Acta Electrotechnica et Informatica.

[10] Peng, M.J.Q. 2008. Multimodality medical image fusion algorithm based on gradient minimization smoothing filter and pulse coupled neural network[J]. Biomedical Signal Processing and Control.

[11] Yoon, H.S. Network-driven low complexity coding for wireless multi-view video system[J]. Advanced Materials Research.

[12] Bosc, E. 2010. Adaptive data-driven parallelization of multi-view video coding on multi-core processor[J]. Science in China Series F: Information Sciences.