Generation of Moire-Like Videos from RGB-D Videos Window

Sho Enomoto
Department of Information Systems, University of Nagasaki, Japan

Toru Hiraoka
Department of Information Systems, University of Nagasaki, Japan
E-mail: hiraoka@sun.ac.jp
https://sun.ac.jp/researchinfo/hiraoka/

Abstract

A non-photorealistic rendering method has been proposed to generate moire-like images from photographic images using bilateral filter and unsharp mask. Extensions to the conventional method have also been proposed to generate moire-like videos from videos or to generate moire-like images from RGB-D images. In this paper, a method is proposed to generate moire-like videos from RGB-D videos. Flickering is a problem in NPR videos, but the proposed method can suppress flicker. Through experiments using an RGB-D video taken by the authors, the flicker of moire-like videos generated by the proposed method was evaluated visually and quantitatively.

Keywords: Non-photorealistic rendering, Moire, Depth, Video, Bilateral filter, Unsharp mask

1. Introduction

Many non-photorealistic rendering (NPR) method using image processing have been proposed [1][2]. One of the NPRs has proposed to automatically generate moire-like images from photographic images [3]. Moire-like patterns are generated according to the change of the edges and the shading in the photographic images, and moire-like images are a type of op art expressed by superimposing moire-like patterns on photographic images. Moire-like images are generated by iterative calculations using bilateral filter[4] and unsharp mask. Extensions to the conventional method[3] have also been proposed to generate moire-like videos from videos[5] or to generate moire-like images from RGB-D images[6]. The conventional methods[5][6] added time and depth terms to the exponent part in the equation of bilateral filter, respectively. Moire-like videos[5] can improve the visual effect by moving moire-like patterns, and moire-like images[6] can generate more distorted moire-like patterns than moire-like images[3].

This paper proposes an NPR method to automatically generate moire-like videos from RGB-D videos. As far as the authors investigate, there is no research on NPR using RGB-D videos. The proposed method adds both time and depth terms to the exponent part in the equation of bilateral filter. Flickering is generally a problem in NPR videos[7], and when the conventional method[5] is applied to each frame of RGB-D videos to generate moire-like videos, flicker occurs in the same way. Hereinafter, moire-like videos generated by applying the conventional method[7] to each frame of RGB-D videos will be referred to as conventional moire-like videos. To confirm that the proposed method can generate moire-like videos with suppressed flicker, experiments were conducted using RGB-D videos taken by the authors.

2. Proposed Method

The proposed method is executed in two steps: Step 1 is the process using bilateral filter embedded the time and depth, and Step 2 is the process using unsharp mask and bilateral filter embedded the time. A flow chart of the proposed method is shown in Fig. 1.

Details of the steps in Fig. 1 are explained below.

Step 0: The input pixel values (R, G, B) and the depths

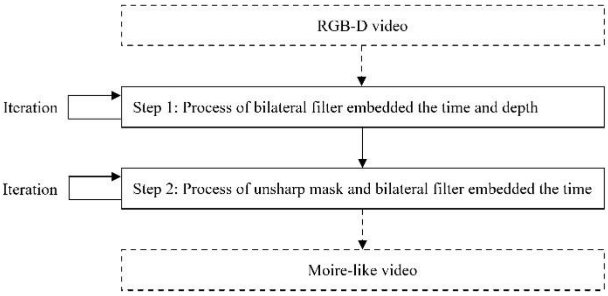


Fig.1. Flow chart of the proposed method

for spatial coordinates (i,j) in k-th frame of an RGB-D video are defined as $f_{R,i,j,k}$, $f_{G,i,j,k}$, $f_{B,i,j,k}$, and $f_{D,i,j,k}$, respectively, where $i=1,2,\cdots,I$, $j=1,2,\cdots,J$, and $k=1,2,\cdots,K$. The pixel values $f_{R,i,j,k}$, $f_{G,i,j,k}$, $f_{B,i,j,k}$, and $f_{D,i,j,k}$ have value of U gradation from 0 to $U-1$. The depths $f_{D,i,j,k}$ are linearly transformed so that the minimum distance becomes 0 and the maximum distance becomes $U-1$.

Step 1: The pixel values of the image at the t-th iteration number are defined as $f^{(t)}_{R,i,j,k}$, $f^{(t)}_{G,i,j,k}$, and $f^{(t)}_{B,i,j,k}$, where $f^{(0)}_{R,i,j,k}=f_{R,i,j,k}$, $f^{(0)}_{G,i,j,k}=f_{G,i,j,k}$, and $f^{(0)}_{B,i,j,k}=f_{B,i,j,k}$. The output pixel values $f^{(t)}_{R,i,j,k}$, $f^{(t)}_{G,i,j,k}$, and $f^{(t)}_{B,i,j,k}$ in bilateral filter embedded the time and depth are calculated by the following equations.

$$f^{(t)}_{R,i,j,k}=\frac{\sum_{l=i-W}^{i+W}\sum_{m=j-W}^{j+W}\sum_{n=k-O}^{k+O}E^{(t-1)}_{1,R,i,j,k,l,m,n}f^{(t-1)}_{R,l,m,n}}{\sum_{l=i-W}^{i+W}\sum_{m=j-W}^{j+W}\sum_{n=k-O}^{k+O}E^{(t-1)}_{1,R,i,j,k,l,m,n}} \quad (1)$$

$$E^{(t-1)}_{1,R,i,j,k,l,m,n}=e^{-\alpha((i-l)^2+(j-m)^2)}e^{-\beta(f^{(t-1)}_{R,i,j,k}-f^{(t-1)}_{R,l,m,n})^2}e^{-\gamma(f_{D,i,j,k}-f_{D,l,m,n})^2}e^{-\delta(k-n)^2} \quad (2)$$

$$f^{(t)}_{G,i,j,k}=\frac{\sum_{l=i-W}^{i+W}\sum_{m=j-W}^{j+W}\sum_{n=k-O}^{k+O}E^{(t-1)}_{1,G,i,j,k,l,m,n}f^{(t-1)}_{G,l,m,n}}{\sum_{l=i-W}^{i+W}\sum_{m=j-W}^{j+W}\sum_{n=k-O}^{k+O}E^{(t-1)}_{1,G,i,j,k,l,m,n}} \quad (3)$$

$$E^{(t-1)}_{1,G,i,j,k,l,m,n}=e^{-\alpha((i-l)^2+(j-m)^2)}e^{-\beta(f^{(t-1)}_{G,i,j,k}-f^{(t-1)}_{G,l,m,n})^2}e^{-\gamma(f_{D,i,j,k}-f_{D,l,m,n})^2}e^{-\delta(k-n)^2} \quad (4)$$

$$f^{(t)}_{B,i,j,k}=\frac{\sum_{l=i-W}^{i+W}\sum_{m=j-W}^{j+W}\sum_{n=k-O}^{k+O}E^{(t-1)}_{1,B,i,j,k,l,m,n}f^{(t-1)}_{B,l,m,n}}{\sum_{l=i-W}^{i+W}\sum_{m=j-W}^{j+W}\sum_{n=k-O}^{k+O}E^{(t-1)}_{1,B,i,j,k,l,m,n}} \quad (5)$$

$$E^{(t-1)}_{1,B,i,j,k,l,m,n}=e^{-\alpha((i-l)^2+(j-m)^2)}e^{-\beta(f^{(t-1)}_{B,i,j,k}-f^{(t-1)}_{B,l,m,n})^2}e^{-\gamma(f_{D,i,j,k}-f_{D,l,m,n})^2}e^{-\delta(k-n)^2} \quad (6)$$

where W is the window size, O is the number of the forward and backward frames used in the calculation, α, β, γ, and δ are positive constants, l and m are the positions in the window, and n is the position in forward and backward frames. In Equations (2), (4), and (6), the depth term $-\gamma(f_{D,i,j,k}-f_{D,l,m,n})^2$ and the time term $-\delta(k-n)^2$ are added to the coefficient of exponential function in the equation of bilateral filter. The larger the value of γ, the more different moire-like patterns from the conventional method[3] are generated. The smaller the value of δ, the more affected by the forward and backward frames. When the value of the frame number O is 0, the conventional moire-like videos are generated. The process of Step 1 is repeated T_1 times.

Step 2: The pixel values $f^{(T_1)}_{R,i,j,k}$, $f^{(T_1)}_{G,i,j,k}$, and $f^{(T_1)}_{B,i,j,k}$ are defined as $g^{(0)}_{R,i,j,k}$, $g^{(0)}_{G,i,j,k}$, and $g^{(0)}_{B,i,j,k}$, respectively. The output pixel values $g^{(t)}_{R,i,j,k}$, $g^{(t)}_{G,i,j,k}$, and $g^{(t)}_{B,i,j,k}$ in unsharp mask using bilateral filter embedded the time are calculated by the following equations.

$$g^{(t)}_{R,i,j,k}=2g^{(t-1)}_{R,i,j,k}-\frac{\sum_{l=i-W}^{i+W}\sum_{m=j-W}^{j+W}\sum_{n=k-O}^{k+O}E^{(t-1)}_{2,R,i,j,k,l,m,n}g^{(t-1)}_{R,l,m,n}}{\sum_{l=i-W}^{i+W}\sum_{m=j-W}^{j+W}\sum_{n=k-O}^{k+O}E^{(t-1)}_{2,R,i,j,k,l,m,n}} \quad (7)$$

$$E^{(t-1)}_{2,R,i,j,k,l,m,n}=e^{-\alpha((i-l)^2+(j-m)^2)}e^{-\beta(g^{(t-1)}_{R,i,j,k}-g^{(t-1)}_{R,l,m,n})^2}e^{-\delta(k-n)^2} \quad (8)$$

$$g^{(t)}_{G,i,j,k}=2g^{(t-1)}_{G,i,j,k}-\frac{\sum_{l=i-W}^{i+W}\sum_{m=j-W}^{j+W}\sum_{n=k-O}^{k+O}E^{(t-1)}_{2,G,i,j,k,l,m,n}g^{(t-1)}_{G,l,m,n}}{\sum_{l=i-W}^{i+W}\sum_{m=j-W}^{j+W}\sum_{n=k-O}^{k+O}E^{(t-1)}_{2,G,i,j,k,l,m,n}} \quad (9)$$

$$E^{(t-1)}_{2,G,i,j,k,l,m,n}=e^{-\alpha((i-l)^2+(j-m)^2)}e^{-\beta(g^{(t-1)}_{G,i,j,k}-g^{(t-1)}_{G,l,m,n})^2}e^{-\delta(k-n)^2} \quad (10)$$

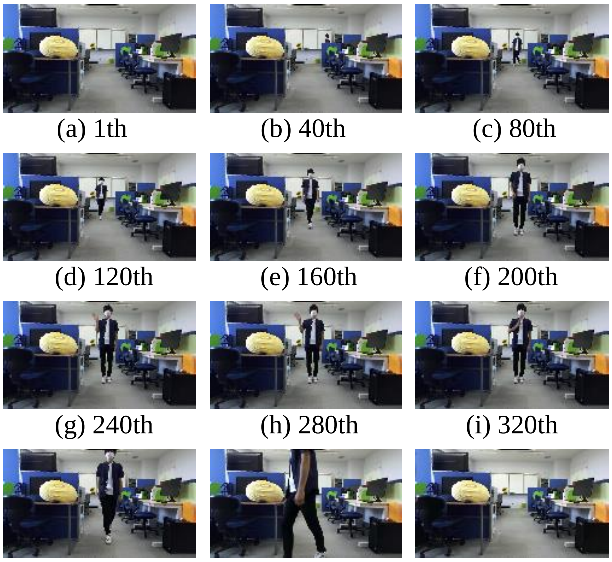
$$g^{(t)}_{B,i,j,k}=2g^{(t-1)}_{B,i,j,k}-\frac{\sum_{l=i-W}^{i+W}\sum_{m=j-W}^{j+W}\sum_{n=k-O}^{k+O}E^{(t-1)}_{2,B,i,j,k,l,m,n}g^{(t-1)}_{B,l,m,n}}{\sum_{l=i-W}^{i+W}\sum_{m=j-W}^{j+W}\sum_{n=k-O}^{k+O}E^{(t-1)}_{2,B,i,j,k,l,m,n}} \quad (11)$$

$$E^{(t-1)}_{2,B,i,j,k,l,m,n}=e^{-\alpha((i-l)^2+(j-m)^2)}e^{-\beta(g^{(t-1)}_{B,i,j,k}-g^{(t-1)}_{B,l,m,n})^2}e^{-\delta(k-n)^2} \quad (12)$$

In case $g^{(t)}_{R,i,j,k}$, $g^{(t)}_{G,i,j,k}$, and $g^{(t)}_{B,i,j,k}$ are less than 0, then $g^{(t)}_{R,i,j,k}$, $g^{(t)}_{G,i,j,k}$, and $g^{(t)}_{B,i,j,k}$ must be set to 0, respectively. In case $g^{(t)}_{R,i,j,k}$, $g^{(t)}_{G,i,j,k}$, and $g^{(t)}_{B,i,j,k}$ are greater than $U-1$, then $g^{(t)}_{R,i,j,k}$, $g^{(t)}_{G,i,j,k}$, and $g^{(t)}_{B,i,j,k}$ must be set to $U-1$, respectively. The process of Step 2 is repeated T_2 times, and a video composed of the pixel values $g^{(T_2)}_{R,i,j,k}$, $g^{(T_2)}_{G,i,j,k}$, and $g^{(T_2)}_{B,i,j,k}$ is the moire-like video.

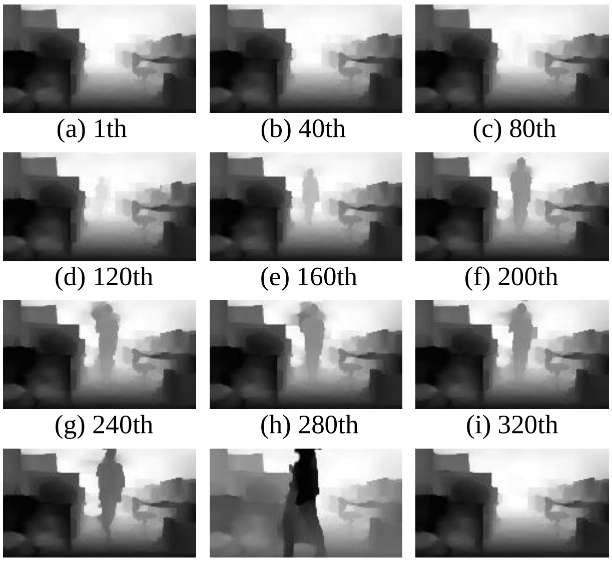
3. Experiments

Experiments were conducted using an RGB-D video which consists of 440 frames, 49 frames / second, 320 * 180 pixels, and 256 gradations. The RGB-D video was shot using ZED stereo camera to capture a scene of a man moving indoor. The farthest distance from the camera was 5.987 meters. The RGB and depth images at the 1th, 40th, 80th, 120th, 160th, 200th, 240th, 280th, 320th, 360th, 400th, and 440th frames of the RGB-D videos are shown in Fig. 2 and Fig. 3, respectively. In reference to the literatures[3][5][6], the values of the parameters α, β, γ, δ, W, O, T_1, and T_2 used in all experiments were set to 0.01, 0.01, 0.001, 0.1, 10, 3, 10, and 20, respectively. Hereinafter, the moire-like video generated by the proposed method under the above conditions is referred to as a proposed moire-like video, and the moire-like video generated by applying the conventional method[6] to each frame of the RGB-D video under the above conditions is referred to as a conventional moire-like video.



(a) 1th (b) 40th (c) 80th
(d) 120th (e) 160th (f) 200th
(g) 240th (h) 280th (i) 320th
(j) 360th (k) 400th (l) 440th

Fig.2. RGB images



(a) 1th (b) 40th (c) 80th
(d) 120th (e) 160th (f) 200th
(g) 240th (h) 280th (i) 320th
(j) 360th (k) 400th (l) 440th

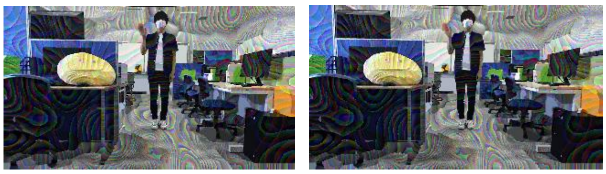
Fig.3. Depth images

The proposed and conventional moire-like videos were visually evaluated. The conventional moire-like video had many flicker. On the other hand, in the proposed video, flicker was suppressed and the movement of moire-like patterns was smooth. As an example, the proposed and conventional moire-like images of the 300th and 301th adjacent frames are shown in Fig. 4 and Fig. 5, respectively. From Fig. 4 and Fig. 5, the proposed moire-like patterns changed less between the two adjacent frames than the conventional moire-like patterns.

The proposed and conventional moire-like videos were quantitatively evaluated. The averages of the absolute values of the differences between the pixel values between adjacent frames (hereinafter, frame difference averages) were calculated, and the average of the frame difference averages of all frames (hereinafter, all frame difference average) was calculated. The all frame difference average P is calculated by the following equations.

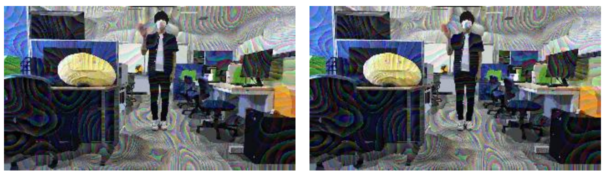
$$P = \frac{\sum_{k=1}^{K-1}\sum_{i=1}^{I}\sum_{j=1}^{J} E_{3,i,j,k}}{3(K-1)} \quad (13)$$

$$E_{3,i,j,k} = |g_{R,i,j,k}^{(T_2)} - g_{R,i,j,k+1}^{(T_2)}| + |g_{G,i,j,k}^{(T_2)} - g_{G,i,j,k+1}^{(T_2)}| + |g_{G,i,j,k}^{(T_2)} - g_{G,i,j,k+1}^{(T_2)}| \quad (14)$$



(a) 300th (b) 301th

Fig.4. Proposed moire-like images



(a) 300th (b) 301th

Fig.5. Conventional moire-like images

Table 1. All frame difference averages

Proposed moire-like video	Conventional moire-like video
485216.922	935160.140

It is judged that the smaller the all frame difference average, the less the flicker of moire-like videos. The all frame difference averages of the proposed and conventional moire-like videos are shown in Table 1. The all frame difference averages of the proposed and conventional moire-like videos are 485216.922 and 935160.140, respectively. The proposed moire-like video had a smaller all frame difference average than the conventional moire-like video. Thus, the proposed moire-like video had less flicker than the conventional moire-like video.

4. Conclusion

This paper proposed an NPR method to automatically generate moire-like videos from RGB-D videos. The proposed method added both time and depth terms to the exponent part in the equation of bilateral filter in the conventional method[3]. Through experiments using an RGB-D video taken by the authors, the appearance of moire-like videos generated by the proposed method was evaluated visually and quantitatively. As a result of the experiments, it was found that the proposed method can suppress flicker.

The future task is to apply the proposed method to more RGB-D videos, although this paper applied the proposed method to one type of RGB-D video.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP19K12664.

References

1. P. Haeberli. Paint by numbers: abstract image representations. ACM SIGGRAPH Computer Graphics, 1990, 24(4), 1990: 207-214.
2. J. Lansdown and S. Schofield. Expressive rendering: a review of nonphotorealistic techniques. IEEE Computer Graphics and Applications, 1995, 15(3): 29-37.
3. T. Hiraoka and K. Urahama. Generation of moire-picture-like color images by bilateral filter. IEICE Transactions on Information and Systems, 2013, E96-D(8): 1862-1866.
4. S. Paris, P. Kornprobst, J. Tumblin, and F. Durand. Bilateral filtering: theory and applications. Foundations and Trends in Computer Graphics and Vision, 2008, 4(2): 1-73.
5. T. Hiraoka and N. Ataka. Generating moire-like animation by bilateral filter. The Japanese Journal of the Institute of Industrial Applications Engineers, 2018, 6(1): 17-22.
6. R. Takaki and T. Hiraoka. Generation of moire-like images from RGB-D images. ICIC Express Letters, 2021, 15(1): 37-42.
7. B. J. Meier. Painterly rendering for animation. Proceedings of SIGGRAPH 96, 1996: 477-484.

Authors Introduction

Mr. Sho Enomoto

He is a student in the Department of Information Systems, University of Nagasaki. His research interests inculede machine learning and non-photorealistic rendering.

Dr. Toru Hiraoka

He received B.Des., M.Des. and D.Eng. degrees from Kyushu Institute of Design in 1995, 1997 and 2005, respectively. He is currently a Professor in University of Nagasaki. His research interests include non-photorealistic rendering and disaster prevention.