2008 IEEE International Conference on
Robotics and Automation
Pasadena, CA, USA, May 19-23, 2008

Removal of Adherent Noises from Image Sequences by Spatio-Temporal Image Processing

Atsushi Yamashita, Isao Fukuchi, Toru Kaneko and Kenjiro T. Miura

Abstract— **This paper describes a method for removing adherent noises from image sequences. In outdoor environments, it is often the case that scenes taken by a camera are deteriorated because of adherent noises such as waterdrops on the surface of the lens-protecting glass of the camera. To solve this problem, our method takes advantage of image sequences captured with a moving camera. The method makes a spatio-temporal image to extract the regions of adherent noises by examining differences of track slopes in cross section images between adherent noises and other objects. Finally, regions of noises are eliminated by replacing with image data corresponding to object regions. Experimental results show the effectiveness of our method.**

I. INTRODUCTION

In this paper, we propose a noise removal method from image sequences by spatio-temporal image processing.

In recent years, cameras are widely used for surveillance systems in outdoor environments such as the traffic flow observation, the trespassers detection, and so on. It is also one of the fundamental sensors for outdoor robots. However, the qualities of images taken through cameras depend on environmental conditions. It is often the case that scenes taken by the cameras in outdoor environments are difficult to see because of adherent noises on the surface of the lens-protecting glass of the camera.

For example, waterdrops or mud blobs attached on the protecting glass may interrupt a field of view in rainy days (Fig. 1). It would be desirable to remove adherent noises from images of such scenes for surveillance systems and outdoor robots.

Professional photographers use lens hoods or put special water-repellent oil on lens to avoid this problem. Even in these cases, waterdrops are still attached on the lens. Cars are equipped with windscreen wipers to wipe rain from their windscreens. However, there is a problem that a part of the scenery is not in sight when a wiper crosses.

Therefore, this paper proposes a new noise removal method from images by using image processing techniques.

A lot of image interpolation or restoration techniques for damaged and occluded images have been also proposed in image processing and computer vision societies [1]–[10]. However, some of them can only treat with line-shape scratches [1]–[3], because they are the techniques for restoring old damaged films. It is also required that human operators indicate the region of noises interactively (not automatically) [4]–[10]. These methods are not suitable for surveillance systems and outdoor robots.

(a) Waterdrop. (b) Mud blob.

Fig. 1. Example of adherent noise.

On the other hand, there are automatic methods that can remove noises without helps of human operators [11], [12]. Hase *et al.* have proposed a real-time snowfall noise elimination method from moving pictures by using a special image processing hardware [11]. Garg and Nayar have proposed an efficient algorithm for detecting and removing rain from videos based on a physics-based motion blur model that explains the photometry of rain [12]. These techniques work well under the assumptions that snow particles or raindrops are always falling. In other words, they can detect snow particles or raindrops because they move constantly.

However, adherent noises such as waterdrops on the surface of the lens-protecting glass may be stationary noises in the images. Therefore, it is difficult to apply these techniques to our problem because adherent noises that must be eliminated do not move in images.

To solve the static noise problem, we have proposed the method that can remove view-disturbing noises from images taken with multiple cameras [13], [14]. Previous study [13] is based on the comparison of images that are taken with multiple cameras. However, it cannot be used for close scenes that have disparities between different viewpoints, because it is based on the difference between images. Stereo camera systems are widely used for robot sensors, and they must of course observe both distant scenes and close scenes. Therefore, we have proposed a method that can remove waterdrops from stereo image pairs that contain objects both in a distant scene and in a close range scene [14]. This method utilizes the information of corresponding points between stereo image pairs, and thereby sometimes cannot work well when appearance of waterdrops differs from each other between left and right images.

We have also proposed a noise removal method by using a

This research was partially supported by Special Project for Earthquake Disaster Mitigation in Urban Areas.

A. Yamashita, I. Fukuchi, T. Kaneko and K. T. Miura are with Department of Mechanical Engineering, Shizuoka University, 3–5–1 Johoku, Naka-ku, Hamamatsu-shi, Shizuoka 432–8561, Japan yamashita@ieee.org

A. Yamashita is with Department of Mechanical Engineering, California Institute of Technology, Pasadena, CA 91125, USA

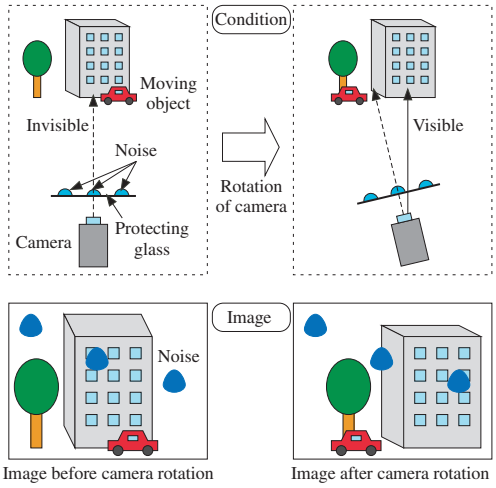


Fig. 2. Image acquisition by using camera rotation.

single camera [15], [16]. These methods use a pan-tilt camera, and eliminate adherent noises based on the comparison of two images; a first image and a second image taken by a different camera angle (Fig. 2). However, adherent noises cannot be eliminated if a background object is blocked by a waterdrop in the first image and is also blocked by another waterdrop in the second image.

In this paper, we use not only two images at certain two frames but all of the image sequence to remove adherent noises in the image sequence. We generate a spatio-temporal image by merging the acquired image sequence, and then detect and remove adherent noises.

II. Noise Detection and Removal Method

As to adherent noises on the protecting glasses of the camera, the positions of noises in images do not change when the direction of the camera changes (Fig. 2). This is because adherent noises are attached to the surface of the protecting glass of the camera and move together with the camera. On the other hand, the position of static background scenery and that of moving objects change while the camera rotates.

We transform the image after the camera rotation to the image whose gaze direction (direction of the principal axis) is same with that before the camera rotation. Accordingly, we can obtain a new image in which only the positions of adherent noises and moving objects are different from the image before the camera rotates.

A spatio-temporal image is obtained by merging these transformed images. In the spatio-temporal image, trajectories of adherent noises can be calculated. Therefore, positions of noises can be also detected in the image sequence from the spatio-temporal image. Finally, we can obtain a noise-free image sequence by estimating textures on adherent noise regions.

A. Spatio-Temporal Image

1) Image Acquisition: An image sequence is acquired while a pan-tilt camera rotates at a constant speed θ rad/frame (Fig. 2).

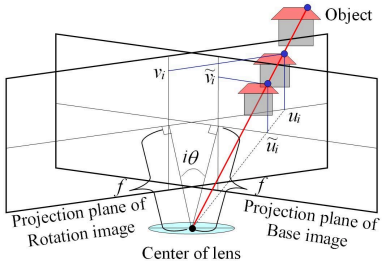


Fig. 3. Projective transformation.

At first (frame 0), one image is acquired where the camera is fixed. In the next step (frame 1), another image is taken after the camera rotates θ rad about the axis which is perpendicular to the ground and passes along the center of the lens. In the i-th step (frame i), the camera rotate $i\theta$ rad and the i-th image is taken. To repeat this procedure n times, we can acquire $n/30$ second movie if we use a 30fps camera.

Note that the rotation angle θ makes a positive direction a counterclockwise rotation (the direction of Fig. 2).

2) Distortion Correction: The distortion from the lens aberration of images is rectified. Let $(\tilde{u}, \tilde{v})$ be the coordinate value without distortion, (u_0, v_0) be the coordinate value with distortion (observed coordinate value), and κ_1 be the parameter of the radial distortion, respectively [17]. The distortion of the image is corrected by (1) and (2).

$$u_0 = \tilde{u} + \kappa_1 \tilde{u} \left(\tilde{u}^2 + \tilde{v}^2\right), \quad (1)$$

$$v_0 = \tilde{v} + \kappa_1 \tilde{v} \left(\tilde{u}^2 + \tilde{v}^2\right). \quad (2)$$

3) Projective Transformation: In the next step, the acquired i-th image (the image after $i\theta$ rad camera rotation) is transformed by using the projective transformation. The coordinate value of the i-th image after the transformation (u_i, v_i) is expressed as follows (Fig. 3):

$$u_i = f\frac{f\tan i\theta + \tilde{u}_i}{f - \tilde{u}_i \tan i\theta}, \quad (3)$$

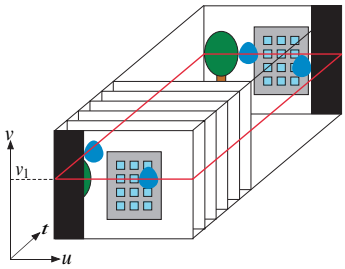
$$v_i = f\frac{\sqrt{1+\tan^2 i\theta}}{f - \tilde{u}_i \tan i\theta}\tilde{v}_i, \quad (4)$$

where $(\tilde{u}_i, \tilde{v}_i)$ is the coordinate value of the i-th image before transformation, θ is the rotation angle of the camera in one frame, and f is the image distance (the distance between the center of lens and the image plane), respectively.

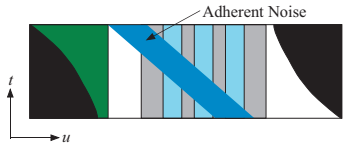
The i-th image after the camera rotation is transformed to the image whose gaze direction is same with that before the camera rotation.

After the projective transformation, there are regions that have no texture in verge area of images (Black regions in Fig. 4(b)). Procedures mentioned below are not applied for these regions.

4) Cross-Section of Spatio-Temporal Image: Spatio-temporal image $I(u, v, t)$ is obtained by arraying all the images (u_i, v_i) in chronological order (Fig. 4(a)). In Fig. 4(a), u is the horizontal axis that expresses u_i, v is the vertical axis that expresses v_i, and t is the depth axis that indicate the time (frame number i).

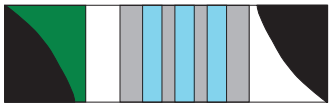


(a) Spatio-temporal image $I(u, v, t)$.

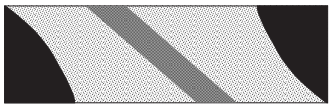


(b) Cross-section image $S(u, t)$.

Fig. 4. Spatio-temporal image.

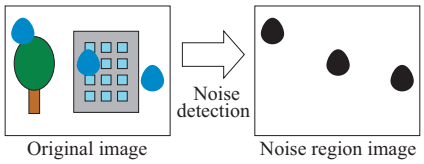


(a) Median image $M(u, t)$.



(b) Difference image $D(u, t)$.

(c) Judgment image $H(u, t)$.



(d) Noise region image $R(\tilde{u}, \tilde{t})$.

Fig. 5. Noise detection.

We can clip a cross-section image of $I(u, v, t)$. For example, Fig. 4(b) shows the cross-section image of the spatio-temporal image in Fig. 4(a) along $v = v_1$.

Here, let $S(u, t)$ be the cross-section spatio-temporal image. In this case, $S(u, t) = I(u, v_1, t)$.

In the cross-section spatio-temporal image $S(u, t)$, the trajectories of the static background scenery become vertical straight lines owing to the effect of the projective transformation. On the other hand, the trajectories of adherent noises in $S(u, t)$ become curves whose shapes can be calculated by (3) and (4)[1].

In this way, there is a difference between trajectories of static objects and those of adherent noises. This difference helps to detect noises.

B. Noise Detection

1) Median Image: Median values along time axis t are calculated in the cross-section spatio-temporal image $S(u, t)$. After that, a median image $M(u, t)$ can be generated by replacing the original pixel values by the median values (Fig. 5(a)).

Adherent noises are eliminated in $M(u, t)$, because these noises in $S(u, t)$ are small in area as compared to the background scenery.

A clear image sequence can be obtained from $M(u, t)$ by using the inverse transformation of (3) and (4) if there is no moving object in the original image. However, if the original image contains moving objects, the textures of these objects blur owing to the effect of the median filtering. Therefore, the regions of adherent noises are detected explicitly, and image restoration is executed for the noise regions to generate a clear image sequence around the moving objects.

2) Difference Image: A difference between the cross-section spatio-temporal monochrome image and the median monochrome image is calculated for obtaining the difference image $S(u, t)$ by (5).

$$D(u, t) = |S(u, t) - M(u, t)| . \tag{5}$$

Pixel values in regions of $D(u, t)$ where adherent noises exist become large, while pixel values of $D(u, t)$ in the background regions are small (Fig. 5(b)).

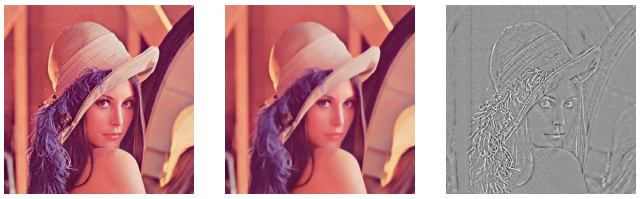
3) Noise Region Image: The regions where the pixel values of the difference images are larger than a certain threshold T_b are defined as the noise candidate regions. The judgment image $H(u, t)$ is obtained by

$$H(u, t) = \begin{cases} 0, & D(u, t) < T_b \\ 1, & D(u, t) \geq T_b \end{cases} . \tag{6}$$

The region of $H(u, t) = 1$ is defined as noise candidate regions (Fig. 5(c)). Note that an adherent noise does not exist on the same cross-section image when time t increases, because v-coordinate value of the adherent noise changes owing to the influence of the projective transformation in (4). Therefore, we consider the influence of this change and generate $H(u, t)$ in the way that the same adherent noise is on the same cross-section image.

In the next step, regions of adherent noises are detected by using $H(u, t)$. The trajectories of adherent noises are expressed by (3). Therefore, the trajectory of each curve is tracked and the number of pixel where $H(u, t)$ is equal to 1 is counted. If the total counted number is more than the threshold value T_n, this curve is regarded as the noise region. As mentioned above, this tracking procedure is executed in 3-D (u, v, t) space. This process can detect adherent noise regions precisely, even when there are moving objects in the original image sequence thanks to the probability voting (counting).

[1]In Fig. 4(b), the trajectory of an adherent noise looks like a straight line, however, it is slightly-curved.



(a) Original image. (b) Structure image. (c) Texture image.

Fig. 6. Image decomposition.

After detecting noise regions in all cross-section spatio-temporal image $S(u,t)$, the noise region image $R(\tilde{u},\tilde{v})$ is generated by the inverse projective transformation from all $H(u,t)$ information (Fig. 5(d)).

Ideally, the noise regions consist of adherent noises. However, the regions where adherent noises don't exist are extracted in this process because of other image noises. Therefore, the morphological operations (opening, *i.e.* erosion and dilation) are executed for eliminating small noises.

C. Noise Removal

Adherent noises are eliminated from the cross-section spatio-temporal image $S(u,t)$ by using the image restoration technique [7] for the noise regions detected in Section II-B.

At first, a original image $S(u,t)$ is decomposed into a structure image $f(u,t)$ and a texture image $g(u,t)$ [18]. Fig. 6 shows an example of the structure image and the texture image[2].

After the decomposition, the image inpainting algorithm [4] is applied for the noise regions of the structure image $f(u,t)$, and the texture synthesis algorithm [19] is applied for the noise regions of the texture image $g(u,t)$, respectively. This method [7] overcomes the weak point that the original image inpainting technique [4] has the poor reproducibility for a complicated texture. After that, noise-free image can be obtained by merging two images.

Finally, a clear image sequence without adherent noises is created with the inverse projective transformation.

III. Experiment

Image sequence was acquired in a rainy day in the outdoor environment. Fig. 7(a) shows an example of the original image, and Fig. 7(b) shows the result of the projective transformation. In this experiment, the frame rate was 30fps, the image size was 360×240pixels, and the length of the movie was 100frames, respectively. We used a pan-tilt-zoom camera (Sony EVI-D100) whose image distance f was calibrated as 261pixel, and the rotation speed of the camera was set as 2.79×10^{-2}rad/frame. Parameters for the noise detection were set as $T_b = 50$, $T_n = 10$. These parameters were constant through all experiments.

Fig. 8 shows the intermediate result of the noise detection. Figs. 8(a) and (b) show the cross-section spatio-temporal image $S(u,t)$ in color and monochromic formats, respectively (red scanline in 7(b), $v = 150$). There is a moving object

[2]Contrast of the texture image (Fig. 6(c)) is fixed for the viewability.

(a) Original image.

(b) Original image after the projective transformation.

Fig. 7. Acquired image.

(a human with a red umbrella) in this image sequence. Figs. 8(c), (d), (e) show the median image $M(u,t)$, the difference image $D(u,t)$, and the judge image $H(u,t)$, respectively. Fig. 8(f) shows the noise region image $R(\tilde{u},\tilde{v})$.

Fig. 9 shows the noise removal result. Figs. 9(a) and (b) show the structure image after applying the image inpainting algorithm and the texture image after applying the texture synthesis algorithm, respectively, while Fig. 9(c) shows the noise removal result of the cross-section spatio-temporal image.

Fig. 10 shows the final results of noise removal for the image sequence. All waterdrops are eliminated and the moving object can be seen very clearly in all frames. Fig. 11 shows results of mud blob removal, and Fig. 12 shows results of waterdrop and mud blob removal, respectively.

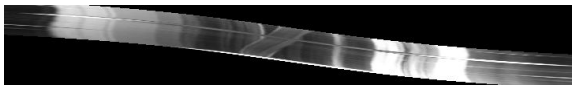
From these results, it is verified that our method can remove adherent noises on the protecting glass of the camera regardless of their positions, colors, sizes, and existence of moving objects.

IV. Discussion

To verify the accuracy of the noise detection, Fig. 8(f) is compared with the ground truth that is generated by a human operator manually (Fig. 13(a)). In Fig. 13(b), red regions indicate the correct detection, blue regions mean undetected noises, and green regions are exceeded detection regions. Actually, undetected noises are hard to detect when we see the final result (Fig. 10(b)). This is because the image interpolation works well in the noise removal step.

Fig. 14 shows comparison results of texture interpolation with an existing method. Fig. 14(b) shows the result by the image inpainting technique [4], and Fig. 14(c) shows the result by our method. The result by the existing method is not good (Fig. 14(b)), because texture of the noise region is estimated only from adjacent region. In principle, it is difficult to estimate texture in several cases from only a single image. On the other hand, our method can estimate texture

(a) Cross-section spatio-temporal image $S(u,t)$ (color).

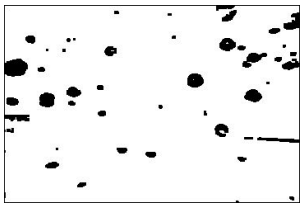


(b) Cross-section spatio-temporal image $S(u,t)$ (gray scale).

(c) Median image $M(u,t)$.

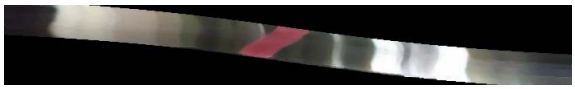
(d) Difference image $D(u,t)$.

(e) Judgment image $H(u,t)$.

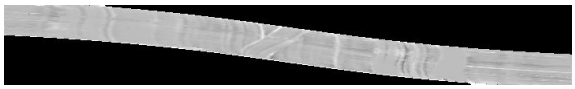


(f) Noise region image $R(\tilde{u},\tilde{t})$.

Fig. 8. Results of noise detection.



(a) Structure image.



(b) Texture image.

(c) Cross-section spatio-temporal image after noise removal.

Fig. 9. Results of noise removal.

(a) Original image. (b) Result image.

Fig. 10. Results of noise removal (waterdrop).

robustly by using a spatio-temporal image processing (Fig. 14(c)).

As to threshold values for the noise detection, we determined them by trial and error. From experimental results (*i.e.* Fig. 8(d)), the difference between the noise region and the background scenery is discriminating. Therefore, they can be decided automatically by using a simple thresholding technique like [20].

As to the camera rotation, the position of principle center of the camera must correspond to that of the rotational center of the camera, although it is very difficult to fit them. It was expected that the disagreement occurred in our experiment. However, the influence of the disagreement between the principle point and the rotational center is very small as compared with the influence of other factors such as the image distortion, and it can be disregarded.

V. Conclusion

In this paper, we propose a noise removal method from image sequence acquired with a pan-tilt camera. We makes a spatio-temporal image to extract the regions of adherent noises by examining differences of track slopes in cross section images between adherent noises and other objects. Regions of adherent noises are interpolated from the spatio-temporal image data. Experimental results show the effectiveness of our method.

As future works, the quality of the final result will be improved for interpolating noise regions in (u,v,t) space. As to the camera motion, a camera translation should be considered in addition to a camera rotation [21]. It is important to compare the performance of our method with recent space-time video completion methods (*e.g.* [8]–[10]).

References

[1] A. C. Kokaram, R. D. Morris, W. J. Fitzgerald and P. J. W. Rayner: "Interpolation of Missing Data in Image Sequences," *IEEE Transactions on Image Processing*, Vol.4, No.11, pp.1509–1519, 1995.

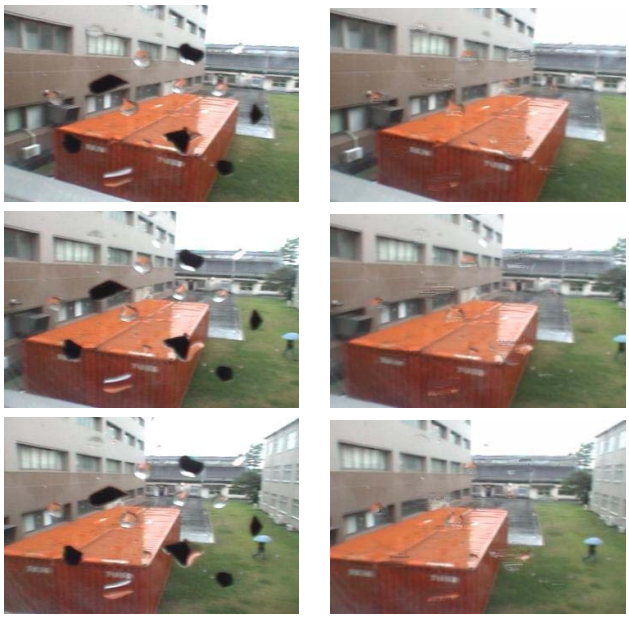
[2] S. Masnou and J.-M. Morel: "Level Lines Based Disocclusion," *Proceedings of the 5th IEEE International Conference on Image Processing* (*ICIP1998*), pp.259–263, 1998.

[3] L. Joyeux, O. Buisson, B. Besserer and S. Boukir: "Detection and Removal of Line Scratches in Motion Picture Films," *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition* (*CVPR1999*), pp.548–553, 1999.

[4] M. Bertalmio, G. Sapiro, V. Caselles and C. Ballester: "Image Inpainting," *ACM Transactions on Computer Graphics* (*Proceedings of SIGGRAPH2000*), pp.417–424, 2000.

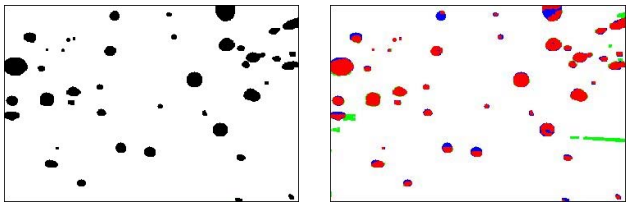
(a) Original image. (b) Result image.

Fig. 11. Results of noise removal (mud blob).



(a) Original image. (b) Result image.

Fig. 12. Results of noise removal (waterdrop and mud blob).



(a) Detected by human. (b) Comparison result.

Fig. 13. Result validation.

[5] M. Bertalmio, A. L. Bertozzi and G. Sapiro: "Navier-Stokes, Fluid Dynamics, and Image and Video Inpainting," *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (*CVPR2001*), Vol.1, pp.355–362, 2001.

[6] S. H. Kang, T. F. Chan and S. Soatto: "Inpainting from Multiple Views," *Proceedings of the 1st International Symposium on 3D Data Processing Visualization and Transmission*, pp.622–625, 2002.

[7] M. Bertalmio, L. Vese, G. Sapiro and S. Osher: "Simultaneous Structure and Texture Image Inpainting," *IEEE Transactions on Image Processing*, Vol.12, No.8, pp.882–889, 2003.

[8] Y. Matsushita, E. Ofek, X. Tang and H.-Y. Shum: "Full-frame Video Stabilization," *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (*CVPR2005*), Vol.1, pp.50–57, 2005.

[9] Y. Shen, F. Lu, X. Cao and H. Foroosh: "Video Completion for Perspective Camera Under Constrained Motion, *Proceedings of the 18th International Conference on Pattern Recognition* (*ICPR2006*), Vol.3, pp.63–66, 2006.

[10] Y. Wexler, E. Shechtman and M. Irani: "Space-Time Completion of Video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.29, No.3, pp.463–476, 2007.

[11] H. Hase, K. Miyake and M. Yoneda: "Real-time Snowfall Noise Elimination," *Proceedings of the 1999 IEEE International Conference on Image Processing* (*ICIP1999*), Vol.2, pp.406–409, 1999.

[12] K. Garg, S. K. Nayar: "Detection and Removal of Rain from Videos," *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (*CVPR2004*), Vol.1, pp.528–535, 2004.

[13] A. Yamashita, M. Kuramoto, T. Kaneko and K. T. Miura: "A Virtual Wiper –Restoration of Deteriorated Images by Using Multiple Cameras–," *Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS2003*), pp.3126–3131, 2003.

[14] Y. Tanaka, A. Yamashita, T. Kaneko and K. T. Miura: "Removal of Adherent Waterdrops from Images Acquired with a Stereo Camera System," *IEICE Transactions on Information and Systems*, Vol.89-D, No.7, pp.2021–2027, 2006.

[15] A. Yamashita, T. Kaneko and K. T. Miura: "A Virtual Wiper – Restoration of Deteriorated Images by Using a Pan-Tilt Camera–," *Proceedings of the 2004 IEEE International Conference on Robotics and Automation* (*ICRA2004*), pp.4724–4729, 2004.

[16] A. Yamashita, T. Harada, T. Kaneko and K. T. Miura: "Virtual Wiper –Removal of Adherent Noises from Images of Dynamic Scenes by Using a Pan-Tilt Camera–," *Advanced Robotics*, Vol.19, No.3, pp.295–310, 2005.

[17] J. Weng, P. Cohen and M. Herniou: "Camera Calibration with Distortion Models and Accuracy Evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.14, No.10, pp.965–980, 1992.

[18] L. I. Rudin, S. Osher and E. Fatemi: "Nonlinear Total Variation Based Noise Removal Algorithms," *Physica D*, Vol.60, pp.259–268, 1992.

[19] A. A. Efros and T. K. Leung: "Texture Synthesis by Non-parametric Sampling," *Proceedings of the 7th IEEE International Conference on Computer Vision* (*ICCV1999*), Vol.2, pp.1033–1038, 1999.

[20] N. Otsu: "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on System, Man, and Cybernetics*, Vol.9, pp.62–66, 1979.

[21] T. Haga, K. Sumi, M. Hashimoto, A. Seki and S. Kuroda: "Monitoring System with Depth Based Object Emphasis Using Spatiotemporal Image Processing," *Technical Report of IEICE* (*PRMU*97-126), Vol.97, No.325, pp.41–46, 1997.

(a) Noise region. (b) Inpainting. (c) Our method.

Fig. 14. Comparison of noise removal results.