1. Introduction

In this paper, we propose a noise removal method from image sequences by spatio-temporal image processing. A spatio-temporal image can be generated by merging the acquired image sequence (Fig. 1(a)), and then cross-section images can be extracted from the spatio-temporal image (Fig. 1(b)). In these cross-section images, we can detect moving objects and estimate the motion of objects by tracing trajectories of their edges or lines.

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. 2). 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 (Kokaram et al., 1995, Masnou & Morel, 1998, Jolyeux et al., 1999, Bertalmio et al., 2000, Bertalmio et al., 2001, Kang et al., 2002, Bertalmio et al., 2003, Matsushita et al., 2005, Shen et al., 2006, Wexler et al., 2007). However, some of them can only treat with line-shape scratches (Kokaram et al., 1995, Masnou & Morel, 1998, Jolyeux et al., 1999), 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) (Bertalmio et al., 2000, Bertalmio et al., 2001, Kang et al., 2002, Bertalmio et al., 2003, Matsushita et al., 2005, Shen et al., 2006, Wexler et al., 2007). These methods are not suitable for surveillance systems and outdoor robots.

On the other hand, there are automatic methods that can remove noises without helps of human operators (Hase et al., 1999, Garg & Nayar, 2004). Hase et al. have proposed a real-time snowfall noise elimination method from moving pictures by using a special image processing hardware (Hase et al., 1999). 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 (Garg & Nayar, 2004). 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 (Yamashita et al., 2003, Tanaka et al., 2006).

Previous study (Yamashita et al., 2003) 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.
Removal of Adherent Noises from Image Sequences by Spatio-Temporal Image Processing

2. Overview of 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. 3). 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.

3. Spatio-temporal image

3.1 Image acquisition
An image sequence is acquired while a pan-tilt camera rotates. 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 \( \theta \) rad about the axis which is perpendicular to the ground and passes along the center of the lens. In the \( t \)-th step (frame \( t \)), the camera rotate \( \theta \) rad and the \( t \)-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 \( \theta \) makes a positive direction a counterclockwise rotation (the direction of Fig. 3).

The direction and the angle of the camera rotation are estimated only from image sequences. At first, they are estimated by an optical flow. However, the optical flow may contain error. Therefore, the rotation angle is estimated between two adjacent frames by an exploratory way. Finally, the rotation angle is estimated between each frame and base frame. The detail of the estimation method is explained in (Yamashita et al., 2009).

3.2 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 \(\kappa_1\) be the parameter of the radial distortion, respectively (Weng et al., 1992). The distortion of the image is corrected by Equations (1) and (2).

\[
\begin{align*}
    u_0 &= \tilde{u} + \kappa_1 \tilde{u} (\tilde{u}^2 + \tilde{v}^2) \\
    v_0 &= \tilde{v} + \kappa_1 \tilde{v} (\tilde{u}^2 + \tilde{v}^2)
\end{align*}
\]

3.3 Projective transformation
In the next step, the acquired \( t \)-th image (the image after \( \theta \) rad camera rotation) is transformed by using the projective transformation. The coordinate value after the transformation \((u_t, v_t)\) is expressed as follows (Fig. 4):

\[
    u_t = f \frac{f \tan \theta_t + \tilde{u}_t}{f - \tilde{u}_t \tan \theta_t}
\]
where \((\overline{u}, \overline{v})\) is the coordinate value of the \(t\)-th image before transformation, and \(f\) is the image distance (the distance between the center of lens and the image plane), respectively. The \(t\)-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. 5(b)). Procedures mentioned below are not applied for these regions.

![Projective transformation](image)

**Fig. 4. Projective transformation.**

### 3.4 Cross-section of spatio-temporal image

Spatio-temporal image \(I(u,v,t)\) is obtained by arraying all the images \((u,v)\) in chronological order (Fig. 5(a)). In Fig. 5(a), \(u\) is the horizontal axis that expresses \(u\), \(v\) is the vertical axis that expresses \(v\), and \(t\) is the depth axis that indicate the time (frame number \(t\)).

We can clip a cross-section image of \(I(u,v,t)\). For example, Fig. 5(b) shows the cross-section image of the spatio-temporal image in Fig. 5(a) along \(v = \overline{v}_1\).

Here, let \(S(u,t)\) be the cross-section spatio-temporal image. In this case, \(S(u,t) = I(u,\overline{v}_1,t)\).
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 Equations (3) and (4). Note that the trajectory of an adherent noise in Fig. 5(b) looks like a straight line, however, it is slightly-curved. In this way, there is a difference between trajectories of static objects and those of adherent noises. This difference helps to detect noises.

4. Noise detection

4.1 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. 6(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 Equations (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.

![Image](a) Median image $M(u,t)$.  
(b) Difference image $D(u,t)$.  
(c) Judgment image $H(u,t)$.  
(d) Noise region image $R(\tilde{u},\tilde{v})$.

Fig. 6. Noise detection.
4.2 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 Equation (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. 6(b)).

$$D(u,t) = \left| S(u,t) - M(u,t) \right|$$

4.3 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

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

The region of $H(u,t)=1$ is defined as noise candidate regions (Fig. 6(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 Equation (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 Equation (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).

After detecting noise regions in all cross-section spatio-temporal image $S(u,t)$, the noise region image $R(\bar{u},\bar{v})$ is generated by the inverse projective transformation from all $H(u,t)$ information (Fig. 6(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 (i.e., erosion and dilation) are executed for eliminating small noises.

5. Noise removal

Adherent noises are eliminated from the cross-section spatio-temporal image $S(u,t)$ by using the image restoration technique (Bertalmio et al., 2003) for the noise regions detected in Section 4.

At first, a original image $S(u,t)$ is decomposed into a structure image $f(u,t)$ and a texture image $g(u,t)$ (Rudin et al., 1992). Figure 7 shows an example of the structure image and the texture image. Note that contrast of the texture image (Fig. 7(c)) is fixed for the viewability.
After the decomposition, the image inpainting algorithm (Bertalmio et al., 2000) is applied for the noise regions of the structure image \( f(u,t) \), and the texture synthesis algorithm (Efros & Leung, 1999) is applied for the noise regions of the texture image \( g(u,t) \), respectively. This method (Bertalmio et al., 2000) overcomes the weak point that the original image inpainting technique (Bertalmio et al., 2000) 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.

![Image 7](https://example.com/image7.png)

Fig. 7. Image decomposition.

### 6. Experiment

Image sequence was acquired in a rainy day in the outdoor environment. Figure 8(a) shows an example of the original image when the rotation speed of the camera is constant, and Fig. 8(b) shows the result of the projective transformation. In this experiment, the frame rate was 30fps, the image size was 360x240pixels, 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. Parameters for the noise detection were set as \( T_n = 50 \), \( T_n = 10 \).

Figure 9 shows the intermediate result of the noise detection. Figures 9(a) and (b) show the cross-section spatio-temporal image \( S(u,t) \) in color and monochromic formats, respectively (red scanline in Fig. 8 (b), \( v = 150 \)). There is a moving object (a human with a red umbrella) in this image sequence. Figures 9(c), (d) and (e) show the median image \( M(u,t) \), the difference image \( D(u,t) \), and the judge image \( H(u,t) \), respectively. Figure 10(a) shows the noise region image \( R(\bar{u}, \bar{v}) \).

Figure 11 shows the noise removal result. Figures 11(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. 11(c) shows the noise removal result of the cross-section spatio-temporal image.

Figure 12 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.
Removal of Adherent Noises from Image Sequences by Spatio-Temporal Image Processing

(a) Original image.  
(b) Original image after the projective transformation.

Fig. 8. Acquired image.

(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 $H(u,t)$.

Fig. 9. Results of noise detection.
Fig. 10. Noise region image $R(\tilde{u}, \tilde{v})$.

(a) Noise region image $R(\tilde{u}, \tilde{v})$.  
(b) Detected by human.

(c) Comparison result.

Fig. 11. Results of noise removal.

(a) Structure image.

(b) Texture image.

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

Fig. 11. Results of noise removal.
To verify the accuracy of the noise detection, Fig. 10(a) is compared with the ground truth that is generated by a human operator manually (Fig. 10(b)). Figure 10(c) shows the comparison results. In Fig. 10(c), 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. 12(b)). This is because the image interpolation works well in the noise removal step.

![Original image](a) Original image. ![Result image](b) Result image.

Fig. 12. Results of noise removal (waterdrop).

Figure 13 shows comparison results of texture interpolation with an existing method. Figure 13(b) shows the result by the image inpainting technique (Bertalmio et al., 2000), and Fig. 13(c) shows the result by our method. The result by the existing method is not good (Fig. 13(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 robustly by using a spatio-temporal image processing (Fig. 13(c)).

Figure 14 shows results of mud blob removal, and Fig. 15 shows results of waterdrop and mud blob removal, respectively. Figure 16(a) shows an example of the original spatio-temporal image when the speed and the direction of the camera rotation is not constant, and Fig. 16(b) shows the result of the projective transformation, respectively. Figure 17 shows the intermediate results of the noise removal. Figures 17(a) shows the cross-section spatio-temporal image $S(u,t)$. There are moving objects (walking men) in this image sequence. Figures 17(b), (c), (d), and (e) show the median image $M(u,t)$, the difference image $D(u,t)$, the judgment image $H(u,t)$, and the noise region image $R(\vec{u},\vec{v})$, respectively. Figure 17(f) shows the noise removal result from the cross-section spatio-temporal image. Figure 18 shows the final results of noise removal for the image sequence. All noises are eliminated and the moving object can be seen very clearly in all frames. 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, existence of moving objects, and the speed and the direction of the camera rotation.

7. 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 $S(u,v,t)$ space. As to the camera motion, a camera translation should be considered in addition to a camera rotation (Haga et al., 1997). It is important to compare the performance of our method with recent space-time video completion methods (e.g., Matsushita et al., 2005, Shen et al., 2006, Wexler et al., 2007).

![Fig. 13. Comparison of noise removal results.](www.intechopen.com)
Removal of Adherent Noises from Image Sequences by Spatio-Temporal Image Processing

Fig. 14. Results of noise removal (mud blob).
Fig. 15. Results of noise removal (waterdrop and mud blob).

Fig. 16. Image sequence when the rotation speed of the camera is not constant.
Removal of Adherent Noises from Image Sequences by Spatio-Temporal Image Processing

Fig. 17. Result of noise detection and removal (camera motion is not constant).
Fig. 18. Results of noise removal (camera motion is not constant).
8. Acknowledgement

This research was partially supported by Special Project for Earthquake Disaster Mitigation in Urban Areas in cooperation with International Rescue System Institute (IRS) and National Research Institute for Earth Science and Disaster Prevention (NIED).

9. References


There are six sections in this book. The first section presents basic image processing techniques, such as image acquisition, storage, retrieval, transformation, filtering, and parallel computing. Then, some applications, such as road sign recognition, air quality monitoring, remote sensed image analysis, and diagnosis of industrial parts are considered. Subsequently, the application of image processing for the special eye examination and a newly three-dimensional digital camera are introduced. On the other hand, the section of medical imaging will show the applications of nuclear imaging, ultrasound imaging, and biology. The section of neural fuzzy presents the topics of image recognition, self-learning, image restoration, as well as evolutionary. The final section will show how to implement the hardware design based on the SoC or FPGA to accelerate image processing.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following: