Virtual Wiper –Removal of Adherent Noises from Images of Dynamic Scenes by Using a Pan-Tilt Camera–

Atsushi Yamashita, Tomoaki Harada, Toru Kaneko
and Kenjiro T. Miura

Abstract

In this paper, we propose a new method that can remove view-disturbing noises from images of dynamic scenes. One of the thorny problems in outdoor surveillance by a camera is that adherent noises such as waterdrops or mud blobs on the protecting glass surface lens disturb the view from the camera. Therefore, we propose a method for removing adherent noises from images of dynamic scenes taken by changing the direction of a pan-tilt camera, which is often used for surveillance. Our method is based on the comparison of two images, a reference image and a second image taken by a different camera angle. The latter image is transformed by a projective transformation and subtracted from the reference image to extract the regions of adherent noises and moving objects. The regions of adherent noises in the reference image are identified by examining the shapes and distances of regions existing in the subtracted image. Finally, regions of adherent noises can be eliminated by merging two images. Experimental results show the effectiveness of our proposed method.

keywords: image restoration, noise removal, adherent noise, moving object, pan-tilt camera

1 Introduction

Surveillance cameras are widely used for the observation of the traffic flow, the detection of trespassers, and so on. In these cases, automatic surveillance systems are expected because it is very difficult for human operators to check the situation at all times. In the near future, the task that mobile robots collect the information about the environment by using a camera also will become very significant and be in high demand for security or disaster response. Acquired images are very important especially in rescue robotics. However, in outdoor environments, it is often the case that scenes taken by the cameras are hard to see because of adherent noises on the surface of the lens-protecting glass of the camera. For example, waterdrops attached on the protecting glass may block the visual field in rainy days. Mud blobs may be also attached in outdoor environments. It would be desirable to remove adherent noises from images of such scenes for the surveillance and the environment recognition. The detection of noise positions in images and the interpolation of these adherent areas are essential techniques to solve this problem.

As to the detection of the position of noise areas in images, there are a lot of studies that detect moving objects or noises in images[1, 2, 3, 4]. These techniques remove moving objects or noises by taking the difference between the initial background scene and a current scene, or taking the difference between temporarily adjacent two frames. These methods are robust against the change of background[2], the change of the weather[3], or the change of the lighting condition[4]. However, it is difficult to apply these techniques to the above problem, because adherent noises such as waterdrops and mud blobs may be stationary noises in the images.

On the other hands, the image interpolation or restoration techniques for damaged and occluded images are also proposed[5, 6, 7, 8, 9, 10]. However, some of them can only treat with line-shape scratches[5, 6, 7], because they are the techniques for restoring old films. It is also required that human operators indicate the region of noises interactively (not automatically)[8, 9, 10]. At any hand, it is also very difficult to treat large noises and to duplicate the complex textures with these methods.

To solve these problems, we have proposed the method for the removal of view-disturbing noises such as waterdrops from images taken with multiple cameras to treat this problem[11]. However, the situation where this method can be applied is limited because multiple cameras cannot necessarily be prepared. Therefore, we have proposed a new method that can remove noises from images by using one camera[12]. In this method, a pan-tilt camera is used because there are not few cases where a pan-tilt camera system is used for the surveillance. However, this method has two big problems:

- This method can treat only static scenes. If there are moving objects in the image, they are regarded as the noise and eliminated from the image.
- The angle of the camera rotation must be precisely measured in this method. If there are little errors, the positions of noises cannot be extracted and the result does not become clear image.

Therefore, this paper propose a method for removing adherent noises from images of dynamic scenes that containing moving objects by using a pan-tilt camera[1]. The proposed method can also estimate the angle of the camera rotation.

Our method is based on the comparison of two images, a reference image and a second image taken by a different camera angle. The latter image is transformed by a projective transformation and subtracted from the reference image to extract the regions of adherent noises and moving objects. The regions of adherent noises in the reference image are identified by examining the shapes and distances of regions existing in the subtracted image. Finally, regions of adherent noises can be eliminated by merging two images.

The composition of this paper is detailed below. In Section II, the method of removing noises in images is explained. In Section III, experimental results are shown and Section IV discusses the effectiveness of our method. Finally, Section V describes conclusions and future works.

[1]Note that we use only 1 DOF rotation.

2 Image Restoration

The difference between images of same scene is very small where adherent noises do not exist, and it is large where adherent noises and moving objects exist. 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 (Figure 1). This is because noises are attached to the surface of the protecting glass of the camera and move together with the camera. On the other hand, the positions of moving objects in images change while the camera rotates.

Therefore, we can obtain two images in which only the positions of adherent noises and moving objects are different from each other when the image after the camera rotation is transformed to the image whose direction of eyeshot is same with that before the camera rotation. By taking into consideration the relationship of the adherent noises' positions in two images, the positions of adherent noises in each image can be estimated. The positions of moving objects can also be estimated in the similar way of adherent noises. Finally, the parts of images where no adherent noises exist are merged to construct a clear image.

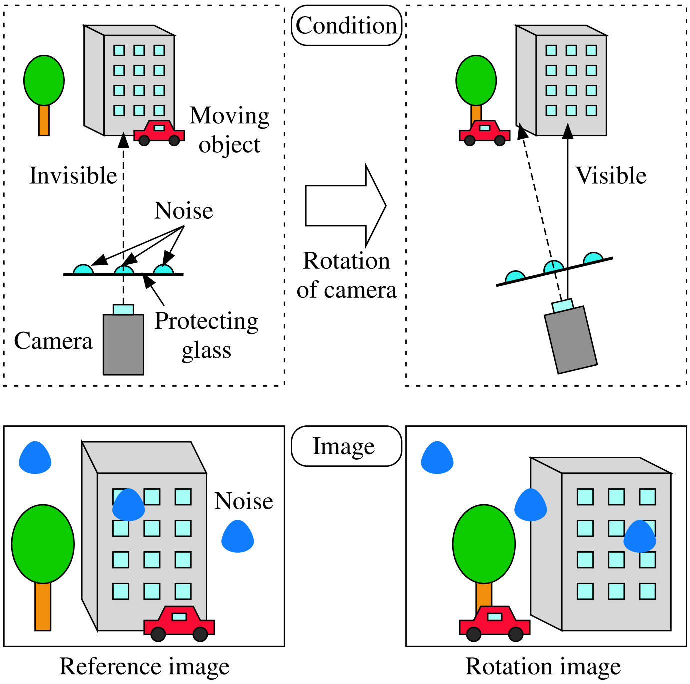


Figure 1: Image acquisition.

The procedure of noise removal is shown in Figure 2.

In this paper, we call the image before the camera rotation "reference image", and the image after the camera rotation "rotation image", respectively.

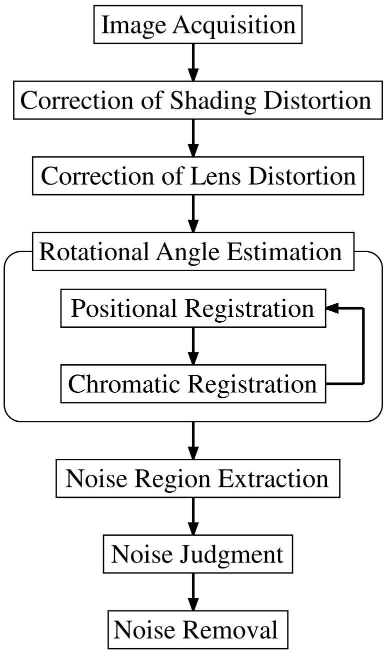


Figure 2: Procedure.

2.1 Image Acquisition

At first, one image is acquired where the camera is fixed. Next, another image is taken after the camera rotates θ degree about the axis which is perpendicular to the ground and passes along the center of the lens. The rotation angle θ makes a positive direction a counterclockwise rotation (the direction of Figure 1).

2.2 Image Registration

The image registration for two obtained images must be executed to generate two images describing the same scene.

In the first step, the shading distortion and the lens distortion are corrected. In the next step, the positional registration of two images is carried out. Finally, the chromatic registration about the common field of view of two images is executed after the positional one.

2.2.1 Correction of Shading Distortion

The brightness of the image varies from space to space because of the shading distortion. The brightness around the edge of the image is darker than that at the center. This distortion originates in the characteristic of a lens. Therefore, we must correct the shading distortion.

The shading distortion becomes large according to the distance from the image center[13]. The irradiance (brightness) $E(u,v)$ at pixel (u,v) can be calculated as follows:

$$E(u,v) \quad = \quad L\frac{\pi}{4}\frac{D}{f}\cos^4\alpha, \tag{1}$$

$$\alpha = \arctan \frac{\sqrt{u^2+v^2}}{f}, \tag{2}$$

where L is a radiance (luminosity energy) per unit volume, D is a diameter of a lens, and f is a image distance[2]. The value of f can be obtained from the camera calibration using the planner pattern on which surface checked patterns are drawn. It is not necessary to know the value of L and D explicitly because they disappears in the process of calculation.

We can correct the shading distortion by considering this characteristic.

2.2.2 Correction of Lens Distortion

The distortion from the lens aberration two images is rectified. Let (u, v) be the coordinate value without distortion, (u', v') be the coordinate value with distortion (observed coordinate value), and κ_1 be the parameter of the radial distortion[14]. The distortion of the image is corrected by (3), (4).

$$u' = u + \kappa_1 u(u^2+v^2), \tag{3}$$

$$v' = v + \kappa_1 v(u^2+v^2). \tag{4}$$

2.2.3 Positional Registration

In the next step, the rotation image (the image after the camera rotation) is transformed by using the projective transformation. The coordinate value after transformation (u_2, v_2) is expressed as follows (Figure 3):

$$u_2 = f\frac{f\tan\theta + u_1}{f - u_1\tan\theta}, \tag{5}$$

$$v_2 = f\frac{\sqrt{1+\tan^2\theta}}{f - u_1\tan\theta}v_1, \tag{6}$$

where (u_1, v_1) is the coordinate value before transformation, θ is the rotation angle of the camera, and f is the image distance.

The image after the camera rotation is transformed to the image whose direction of eyeshot is same with that before the camera rotation.

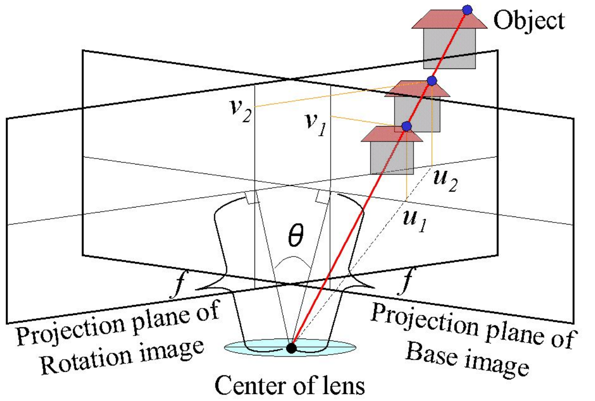
We call the rotation image after the projective transformation “transformed image”.

2.2.4 Chromatic Registration

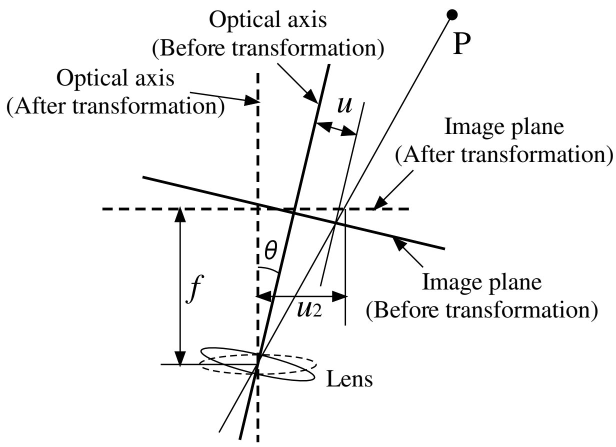
It is often the case that the chromatic tone of the transformed image (the rotation image after the transformation) changes from the reference image (the image before the camera rotation) under the influence of the change of the lighting condition.

Here, let A_1 be the average brightness of the reference image, and A_2 be that of the transformed image about the common field of view of two images, respectively. The common field of view can be

[2] The image distance is equal to the distance between the center of lens and the image plane. Although it is confusable, the image distance is not same as the focal length. When an image of an infinitely (or at least sufficiently) distant object is created on the sensor, this distance is equal to the focal length of the lens[15].



(a) Schematic view.



(b) Top view.

Figure 3: Projective transformation.

easily calculated from the rotational angle of the camera. The chromatic registration can is done as follows:

$$\begin{pmatrix} R'(u,v) \\ G'(u,v) \\ B'(u,v) \end{pmatrix} = \frac{A_1}{A_2} \begin{pmatrix} R(u,v) \\ G(u,v) \\ B(u,v) \end{pmatrix}, \tag{7}$$

where $(R'(u,v), G'(u,v), B'(u,v))$ are the corrected RGB values of the transformed image at pixel (u,v), and $(R(u,v), G(u,v), B(u,v))$ are the original values of the transformed image.

2.3 Rotational Angle Estimation

The image registration goes wrong when the rotational angle of the camera is not known precisely. However, the angle cannot be measured with very high accuracy because of the delicate error of a pan-tilt camera and the limitation of the resolution of the camera's encoder. Therefore, we must estimate the rotational angle of the pan-tilt camera from two acquired images.

The reference image becomes same as the transformed image when the rotational angle of the camera is known precisely in the case that there are no adherent noises and moving objects in images. The difference between the reference image and the transformed image may become little also when there are noises and moving objects in images. Therefore, the rotational angle of the camera can be estimate to check the difference between two images. The difference $D(u,v)$ that indicates the difference between two images at pixel (u,v) can be calculated as follows:

$$D(u,v)^2 = (R_1(u,v) - R_2(u,v))^2 + (G_1(u,v) - G_2(u,v))^2 + (B_1(u,v) - B_2(u,v))^2, \tag{8}$$

where $(R_1(u,v), G_1(u,v), B_1(u,v))$ are the RGB values of the reference image at pixel (u,v), and $(R_2(u,v), G_2(u,v), B_2(u,v))$ are the RGB values of the transformed image after the chromatic registration.

Usually, gray-scale images may be used when comparing two images. The gray-scale image is calculated by $M = 0.299R + 0.587G + 0.114B$ where M is the brightness. However, the ratio of blue (B) is smallest among RGB values in this equation. In the case of adherent waterdrops, small difference of B value must be considered because the color of waterdrops is sensitive about the blue color. Therefore, we adopt the distance in RGB space to treat waterdrops successfully.

The total sum of the difference between two images about the common field of view when the rotational angle is θ is also calculated as follows:

$$S(\theta) = \frac{1}{C} \sum_u \sum_v D(u,v), \tag{9}$$

where C is the total pixel number of the common field of view.

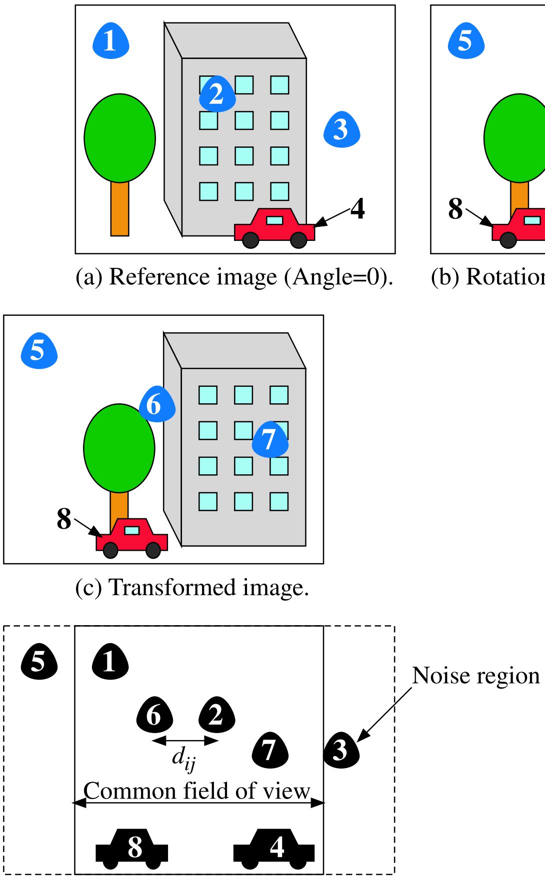
The rotational angle of the camera θ_{opt} is estimated to minimize $S(\theta)$.

$$\theta_{opt} = \arg\min_\theta S(\theta). \tag{10}$$

The optimization is accomplished by the exploratory search. We must iterate the positional and chromatic registrations until estimating the rotation angle.

2.4 Noise Region Extraction

The positions where adherent noises and moving objects exist are estimated by comparing the reference image and the transformed image after the rotational angle estimation. The thresholding process gives a difference image where noise and moving-object regions and the rest regions are binarized (Figure 4).



(a) Reference image (Angle=0). (b) Rotation image (Angle= θ).

(c) Transformed image.

(d) Difference image between (a) and (c).

Figure 4: Projective transformation and difference between two images.

We define regions where the differences between two gray-scale images are larger than a certain threshold T as the noise regions of two images. The difference image $H(u,v)$ is obtained by

$$H(u,v) = \begin{cases} 0, & D(u,v) \leq T_1 \\ 1, & D(u,v) > T_1 \end{cases} . \tag{11}$$

The region of $H(u,v) = 1$ is defined as noise regions. Ideally, the noise regions consist of adherent noises and moving objects. However, the regions where adherent noises or moving objects don't exist are extracted in this process because of other image noises. Therefore, the morphological operations (opening operation: erosion and dilation) are executed for eliminating small noises.

2.5 Noise Judgment

It is judged to which image each noise region is attached between the reference image and the transformed image, or judged that the region is derived from a moving object. The noise attached to the transformed image looks moving leftward to the same noise attached to the reference image when the direction of the camera rotation is counterclockwise ($\theta \geq 0$). And the distance between these noise regions is also calculated by (5) and (6).

When the noise region is derived from a moving object, another noise region that has a similar shape exists. However, the relationships of the positions between these two noise regions are different from the case of adherent noises.

Therefore, we can distinguish the noise regions into three cases: (i) the noise region is derived from the adherent noise on the reference image, (ii) that is derived from the adherent noise on the transformed image, and (iii) that is derived from a moving object.

Here, let d_{ij} be the distance between noise region i and j. When d_{ij} satisfies (5) and (6), it can be judged to which image each noise region is attached by distinguishing whether there are the same size noise region that is distant from another noise region. If there is same size region on the left side of the observed noise region, the observed region is derived from the reference image. If there is same region on the right side, the observed region is derived from the transformed image.

The judgment of "same size" is executed by the pixel-based comparison (Figure 5).

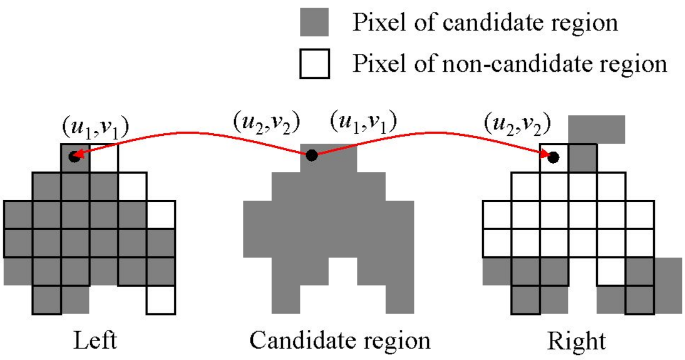


Figure 5: Judgment by pixel correspondence.

When the observed noise region is compared with the left region, the existence rate E_l is calculated. In the case that pixel (u_2, v_2) in the observed region, 1 point is added to the left score n_l, if pixel (u_1, v_1) belongs to the noise region ($H(u_1, v_1) = 1$). If pixel (u_1, v_1) does not belong to the noise region ($H(u_1, v_1) = 0$), point is not added to the left score n_l. If (u_1, v_1) is out of the common field of view, 0.5 points is added to the left score n_l (Table 1).

After changing (u_2, v_2) to all pixel in the observed noise region, the existence rate on the left side E_l is calculated as follows:

$$E_l = \frac{n_l}{N}, \tag{12}$$

Table 1: Score of Decision.

Situation of pixel	Score
Candidate region	1
Non-candidate region	0
Out of common field of view	0.5

where N is the number of the pixel of the observed noise region.

In the same way, the right score n_r can be measured and the existence rate on the right side E_r is calculated.

$$E_r = \frac{n_r}{N}. \tag{13}$$

After that, we can distinguish the noise region into three cases. If E_l and E_r are small, there is no noise region that has same size on the left and right sides. Therefore, the noise region is regarded as being derived from a moving object. If $E_l > E_r$, the noise region is derived from adherent noise on the reference image. The rule of decision when the rotational angle of the camera $\theta \geq 0$ is shown in Table 2. In this table, parenthetic conditions mean the cases when $\theta < 0$.

Table 2: Rule of Decision.

Condition	Decision
$E_l < T_2$ and $E_r < T_2$	Moving object
$E_l \geq T_2$ and $E_l > E_r$ ($E_r \geq T_2$ and $E_r > E_l$)	Adherent noise on reference image
$E_r \geq T_2$ and $E_l \leq E_r$ ($E_l \geq T_2$ and $E_r \leq E_l$)	Adherent noise on transformed image

For example, in Figure 4(d), noise region 2 can be judged as the adherent noise in the reference image because E_l is large. Noise region 7 can also be judged as the adherent noise in the transformed image because E_r becomes large by adding n_r to 0.5 points for the pixel number of this region.

The distance between noise region 4 and 8 is too large and this condition does not satisfy (5) and (6). Therefore, E_l and E_r become small values and these regions are judged as moving objects.

2.6 Noise Removal

Noise removal is performed by using another image data for the adherent noise regions. It occurs that the removal of noise contour cannot be performed because the difference between the area around the contour and the scenery is small and then the noise region becomes small. Therefore, the dilation operation is executed about each noise region.

3 Experiments

We verified the effectiveness of the proposed method through experiments. The resolution of images was set as 640×480pixel. The number of the morphological operations was two times in this resolution. All process in the algorithm is done automatically.

Figure 6 shows the results when there are mud blobs on the protecting glass. Figure 6(a) shows the reference image in which a man walked in the right side of the image, and Figure 6(b) shows the rotation image in which he moved to the center of the image. Figure 6(c) shows the positions of moving objects, and Figure 6(d) shows the positions of adherent noises on the reference image. The rotational angle of the camera is estimated that $\theta = 5.07$deg, while the initial given angle of the camera is 5.0deg. Almost all computation time is spent on angle estimation because this process needs iterations. Moving objects can be extracted from the images and it can be said that the positions of adherent noises in the reference image are judged properly by comparing Figure 6(a) and (d).

Figure 6(e) and (f) show the improved reference and rotation images, respectively. The black regions in these figures are out of common field of view. These results indicate that our method can remove mud blobs from images that contain moving objects.

Figure 7 shows the results when there are waterdrops on the protecting glass. In this situation, a man walked from the center of the image to the left side. In this case, the rotational angle of the camera is estimated that $\theta = 5.18$deg. These results indicate that our method can remove waterdrops from images.

Although there are several false noise regions in Figure 6 and Figure 7, the improved images are well recovered. This is because these false noise regions are too small for us to feel unnaturalness.

Figure 8 shows a part of another image. Figure 8(a) shows the original image in which waterdrop attached on the edge of the background object, Figure 8(b) is the result by our proposed method. A result with "image inpainting" algorithm [8] is shown in Figure 8(d). In the case of image inpainting, human operator indicates the position of adherent waterdrop (Figure 8(c)). Figure 8(b) is correctly restored, although the edge of the background object is not correctly restored in Figure 8(d)[3].

From these results, it is verified that our method can remove adherent noises without reference to the colors and the sizes. Our method can also treat with images of dynamic scenes that contain moving object.

4 Conclusions

In this paper, we propose a new method that can remove view-disturbing noises from images by processing images taken with a pan-tilt camera system that can change the direction of eyeshot. In our method, an image of a distant prospect is taken at first and another image is taken after changing the direction of eyeshot. The new image is transformed with the projective transformation and compared

[3]Note that all parameters of "image inpainting" algorithm [8] were not perfectly set correctly in our experiments.

(a) Reference image.

(b) Rotation image.

(c) Moving objects.

(d) Adherent noises in the reference image.

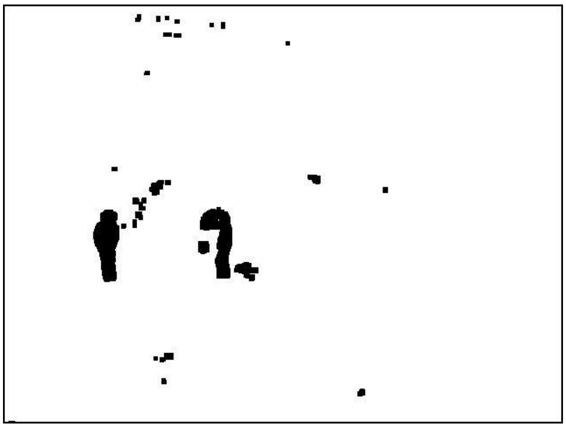
(e) Improved reference image.

(f) Improved rotation image.

Figure 6: Result of mud blobs removal.

(a) Reference image.

(b) Rotation image.



(c) Moving objects.

(d) Adherent noises in the reference image.

(e) Improved reference image.

(f) Improved rotation image.

Figure 7: Result of waterdrop removal I.

(a) Original image.

(b) Result image.

(c) Waterdrop area.

(d) Result image by [8].

Figure 8: Result of waterdrop removal II.

with the first one to detect the region where noises may exist. The region where noises exist can be eliminated to merge two images.

Moving objects are also extracted from images of dynamic scenes and adherent noise on the protecting glass can be decided. The rotational angle of the camera can be also estimated not from the encoder value of the camera but only from two acquired images.

The proposed method is simple and effective, and therefore it has the potential to be used in several places include surveillance and rescue robotics.

Of course, our method cannot work well in the case that the area and the number of adherent noises are very large such as cloudburst. Additionally, it is also difficult to generate clear images when the speeds of the moving objects in the image are same with those of adherent noises. In these cases, more than three images will be utilized for estimating the positions of adherent noises with high accuracy in the future work. An automatic determination of threshold values (T_1 and T_2) is also needed. The chromatic registration method in this paper must be sophisticated, e.g.[16], too.

Acknowledgment

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)), and the Ministry of Education, Culture, Sports, Science and Technology, Grant-in-Aid for Young Scientists (B), 15700153, 2003.

REFERENCES

[1] Anil C. Kokaram, Robin D. Morris, William J. Fitzgerald and Peter J. W. Rayner: "Detection of Missing Data in Image Sequences," *IEEE Transactions on Image Processing*, Vol.4, No.11, pp.1496–1508, 1995.

[2] Atsushi Nagai, Yoshinori Kuno and Yoshiaki Shirai: "Surveillance System Based on Spatio-Temporal Information," *Proceedings of the 1996 IEEE International Conference on Image Processing* (*ICIP1996*), Vol.2, pp.593–596, 1996.

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

[4] Takashi Matsuyama, Takashi Ohya and Hitoshi Habe: "Background Subtraction for Non-Stationary Scenes," *Proceedings of the 4th Asian Conference on Computer Vision* (*ACCV2002*), pp.662–667, 2002.

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

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

[7] 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.

[8] Marcelo Bertalmio, Guillermo Sapiro, Vicent Caselles and Coloma Ballester: "Image Inpainting," *Computer Graphics* (*SIGGRAPH 2000*), pp.417–424, 2000.

[9] Sung Ha Kang, Tony F. Chan and Stefano Soatto: "Inpainting from Multiple Views," *Proceedings of the 1st International Symposium on 3D Data Processing Visualization and Transmission*, pp.622–625, 2002.

[10] Tony F. Chan and Jianhong Shen: "Variational Image Inpainting," *IMA Preprint*, 1868, pp.1–28, 2002.

[11] Atsushi Yamashita, Masayuki Kuramoto, Toru Kaneko and Kenjiro 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.

[12] Atsushi Yamashita, Toru Kaneko and Kenjiro 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.

[13] B. K. P. Horn: *Robot Vision*, The MIT Press, 1986.

[14] Juyang Weng, Paul Cohen and Marc 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.

[15] S. G. Lipson, H. Lipson and D. S. Tannhauser: *Optical Physics Third Edition*, Cambridge University Press, 1995.

[16] Lisa Gottesfeld Brown: "A Survey of Image Registration Techniques," *ACM Computing Surveys*, Vol.24, No.4, pp.325–376, 1992.