



High-Resolution Image Data Collection Scheme for Road Sensing Using Wide-Angle Cameras on General-Use Vehicle Criteria to Include/Exclude Collected Images for Super Resolution

Teruhisa Takano¹ · Shintaro Ono¹ · Hiroshi Kawasaki² · Katsushi Ikeuchi^{1,3}

Received: 9 June 2020 / Revised: 23 October 2020 / Accepted: 20 November 2020 / Published online: 7 January 2021
© The Author(s) 2021

Abstract

Fisheye cameras or wide-angle cameras used on automobiles will have various future applications as “distributed sensors in daily transport” – objects on roadside are frequently observed multiple times by multiple vehicles. Although the object regions are often captured with relatively low image quality (low resolution and large blur) resulting from the character of the lens, it is possible to enhance their quality by post-processing using super resolution (SR) technology. Here, it is required to decide which images to use as inputs for SR: a greater number of lower-quality images or fewer higher-quality images. We evaluated and discussed the input image quality necessary to obtain most effective SR results, especially focusing on degree of image blur. As a conclusion, we found a criterion that was related to the blur level of the initial input image of SR. Then, we considered its potential use as a requisite in observing road environments.

Keywords Reconstruction-based super resolution · On-vehicle fisheye/super wide-angle camera · Number of input images · Trade-off between image quality and quantity

1 Introduction

1.1 Background

Road sensing using general-purpose automobiles for business and personal use is becoming more feasible as well as using dedicated-purpose vehicles such as MMS (mobile mapping system). Particularly, web-connected on-vehicle cameras, driving video recorders, and parking assistance cameras are

going to gain popularity. Although image and position data collected by such sensing devices have relatively low quality or poor accuracy, the quantity of data is very large and the quality and accuracy can potentially be improved through post-processing.

Regarding on-vehicle cameras equipped on general-use vehicles for the purposes described above, super wide-angle lenses including fisheye lenses are typically used to observe wider areas using fewer cameras, leading to lower image quality especially around the peripheral area of the lens. It is even possible to generate images with higher resolution through post-processing, using super resolution (SR) technology.

However, because of its lens character, on-vehicle image sensing with such a super wide-angle camera faces a trade-off problem on the quality and quantity of the collected images and it is hard to meet both of them at the same time, as described in 3.1 with Fig. 1 in detail. In other words, collecting more images and putting all of them to the post process does not necessarily improve the output. Therefore, the criterion to include or exclude the collected images becomes an important issue.

1.2 Significance in Intelligent Transport Systems

Clarifying the criterion of input images have large significance towards smart city sensing along roads using

✉ Shintaro Ono
onoshin@its.iis.u-tokyo.ac.jp

Teruhisa Takano
teruhisa-takano@cvt.iis.u-tokyo.ac.jp

Hiroshi Kawasaki
kawasaki@ait.kyushu-u.ac.jp

Katsushi Ikeuchi
katsushi.ikeuchi@outlook.jp

¹ The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 1538505, Japan

² Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 8190385, Japan

³ Microsoft Research, 11801 97th Ln NE, Kirkland, WA 98034, USA

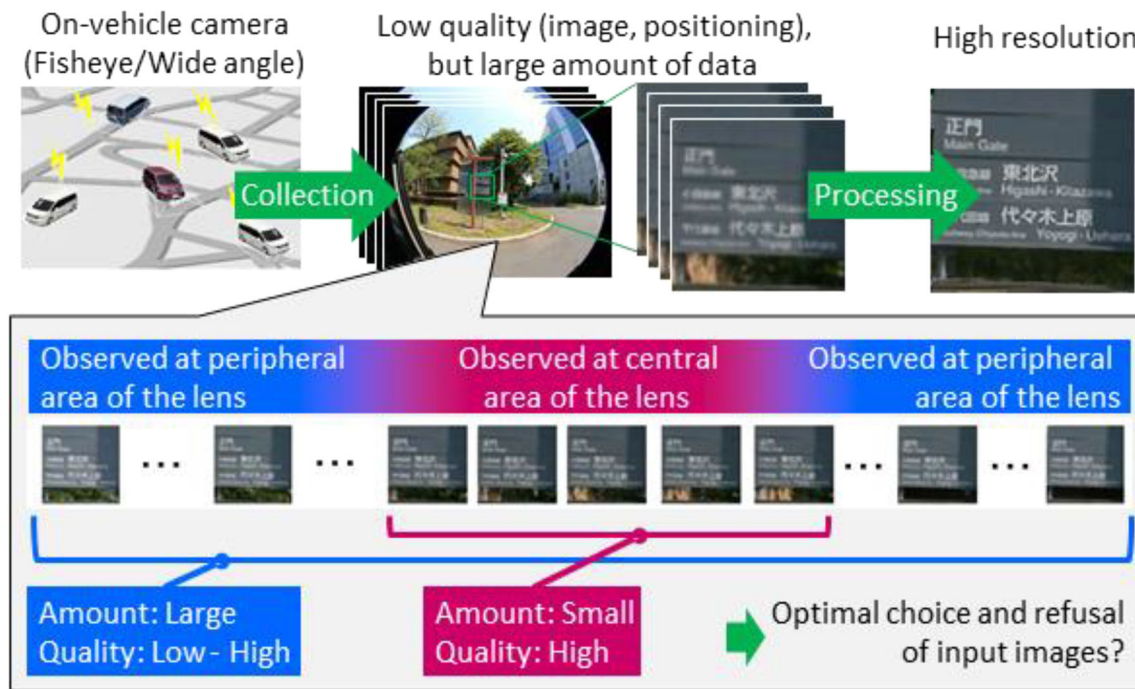


Fig. 1 Trade-off of quality and quantity in collecting images using an on-vehicle fisheye/wide-angle camera and SR (super resolution). Using more input images theoretically increases information for SR, but on the other

hand leads to estimation errors in SR parameters. Which images to include/exclude is the problem

general-use vehicles. It can assist to infer requisites for on-vehicle road sensing.

For example, in microscopic aspect, specifying the number of times of observations, maximum speed of vehicle in observation, and various camera specifications including frame rate and lens character, depending on the distance and size of the objects.

Moreover, in macroscopic aspect, it can also provide basic information for assessing the necessity for cloud-based sharing of data captured by other vehicles, quantity and frequency of the shared data, and design of the cloud-based sharing environment.

1.3 Contributions

Focusing on SR of the fisheye images collected by on-vehicle camera, we examine and consider the inclusion or exclusion of images that will enable better SR result. Then, we discuss the possibility of the blur level of the image region being a criterion for the inclusion/exclusion of input images. As a result, we find out a criterion that is related to the blur level of the initial input image of SR.

The contributions of this paper are:

1. Pointed out the trade-off problem on the quality and quantity of the images collected by typical on-vehicle cameras of general-use vehicles.
2. Developed the SR method that can deal with such collected images by introducing adaptive deformation and blur model.

3. Found out the criterion of inclusion/exclusion of input images for SR to achieve the highest image quality in result, focusing on the blur levels of them.
4. Opened up the possibility to establish or improve the scheme for on-vehicle road sensing by general-use vehicles and cloud-based data sharing.

The structure of this paper is composed of 6 sections. At first several related prior works are surveyed in Section 2. Next, Section 3 proposes the SR method using fisheye camera images and the evaluation method to examine the SR results. Section 4 shows results of the experiments, using a vehicle equipped with actual fisheye camera and shows the relation between the inclusion/exclusion of the images and SR results, focusing on the blur of the image. At Section 5, we discuss the relation between the blur of input images of SR and the output image quality of SR results. Finally, Section 6 gives a summary of this paper.

2 Related Works

2.1 Sensing by Vehicle for Dedicated Use or General Use

Regarding road sensing by a vehicle, a system composed of special sensing devices such as LiDARs and cameras has been commonly used on a dedicated-use vehicle. Such a system is generally called MMS (mobile mapping system) and widely

developed by survey companies. In this case, the specification of the sensor can be intentionally designed to obtain enough data quality; however, its cost makes it unrealistic to frequently survey wide area.

Conversely, exploiting general-use vehicles for probe survey became more feasible today. For example, it is quite practical to analyze motion data of general-use vehicles to sense traffic congestion, road-surface condition, dangerous driving, etc.

Moreover, web-connected on-vehicle cameras, including driving video recorders and parking assistance cameras, have gained popularity. Exploiting such online/offline data enables to visually share situations of congested spots [1], estimate per-lane as well as per-road traffic congestion [2], construct near-miss incident or traffic accident database [3, 4], etc.

Although image and position data collected by such a sensing scheme lacks high quality or accuracy than dedicated sensing scheme, it is realistic to sense wider area with higher frequency. Since the quantity of data is large, the issue of low quality and accuracy can potentially be improved in post-processing. These characters are summarized in Table 1. Image sensing by this scheme will lead to more expansive applications, such as map construction and criminal investigation, if issues regarding personal data protection are resolved.

2.2 Improving Resolution of Fisheye Images

Improving image resolution is a basic signal processing technique to generate a high-resolution image as output from inputted low-resolution images, called super resolution (SR). A method named reconstruction-based SR is known well. In this method, multiple images are provided as inputs and the image positions are registered in sub-pixel order.

Another well-known SR is a learning-based SR such as SRCNN [5], which is the technique to obtain a high-resolution image from even a single low-resolution image by learning the relation between the high- and low-resolution images. In this study we do not deal with this type of learning-based SR, where the output result is affected by learned cases that are not necessarily related with the observed images.

Several related studies have been conducted on SR for fisheye/super wide-angle cameras [6–9]. Although their

problem settings are in part similar to ours, they consider only limited cases in which the shape and blur level of the image region of interest (ROI) where the object is observed are constant throughout all of the input images. This differs from the conditions in our study. Moreover, to our knowledge, no other study has been reported discussing the relation between SR results and constraints on the input images.

3 SR Using Fisheye Camera Images

3.1 Trade-off of Input Images: Quality and Quantity

A super wide-angle lens, especially a fisheye lens, which is often used for on-vehicle devices of general-use vehicles, generally captures an image with greater distortion and lower resolution, especially in the peripheral area of the lens. Let’s assume a situation in which such a camera is on a vehicle moving directly forward, observing a static object.

As Fig. 1 illustrates, if an object is observed around peripheral area of the lens, its image quality becomes lower, and if observed in central area, its image quality becomes higher. Meanwhile, the peripheral-area observation has more chance to occur rather than the central-area observation. Therefore, the quality and the quantity cannot be compatible. This is inevitable for fisheye lens, unless some extra sensing scheme is assumed such as collecting sensing data by multiple vehicles.

Reconstruction-based SR technology theoretically requires sufficient information as input, i.e., sufficiently many input images. However, at the same time, low-quality input images cause errors in estimating image observation parameters (such as displacement or blur) required in the SR process and lower the quality of the result. In other words, there is a trade-off problem between the quality and quantity of the input image for SR. Whether to include or exclude the observed images from post-processing affects the final quality of the output image, however, it is unknown which images have to be included/excluded to obtain better results.

In this study, we focus on the size of the blur parameter of the image region as the criterion, and we discuss the problem

Table. 1 Sensing using dedicated-use vehicles and general-use vehicles

	Dedicated-use vehicles (MMS)	General-use vehicles
Sensing system specifications	Can be designed intentionally	An inexpensive and multi-purpose system is generally assumed
Accuracy of acquired data	Relatively high	Relatively low, but might be improved through post-processing
Quantity of acquired data	Depends on the operational opportunities of the vehicle itself	Potentially capable of collecting a large quantity of data through sharing, if the system becomes widely used
Sensing for wide areas and/or with frequent opportunities	Typically, costly and not very realistic	Relatively realistic, if the system becomes widely used

by comparing several inclusion/exclusion patterns based on this criterion.

3.2 SR Method Based on Adaptive Deformation and Blur Model

SR for fisheye image is fundamentally based on the conventional reconstruction-based SR method. This is a kind of completed technology for ordinary images, however, particularly in the case of fisheye images, some special processes is included through image observation: lens distortion/undistortion, deformation, and the blur with scaling caused by the undistortion. Please note that these effects are not uniform but dependent on position in the image. We have proposed an improved SR method based that can work even under such complex conditions in [10], by developing “adaptive degradation model”.

Figure 2 illustrates the problem of reconstruction-based SR for fisheye images using the adaptive degradation model. After successively shooting a real-world object and before it is stored as observed image data, a process with degradation model is assumed composed of displacement with deformation, blur, down sampling, and noise. The SR can be realized by finding an ideal high-resolution image (x) that minimizes the difference between the actual observed images (y_k) and the “degraded ideal image” using the same degradation model and parameters. This problem can be formulated as equations below:

$$\min_x \sum_k |DB_k x - z_k| \tag{1}$$

$$z_k = H_k y_k \tag{2}$$

$$H_c = I \tag{3}$$

($k = 0, 1, \dots, c, \dots, N-1$)

where k is number of successively taken images, N is whole number of images taken, x is an ideal image (in column-vector form), and y_k are rectangular regions in the observed images after correcting lens distortion. B_k and D are degradation matrix operations that represent blur and down sampling. D can be given as a constant matrix by problem setting.

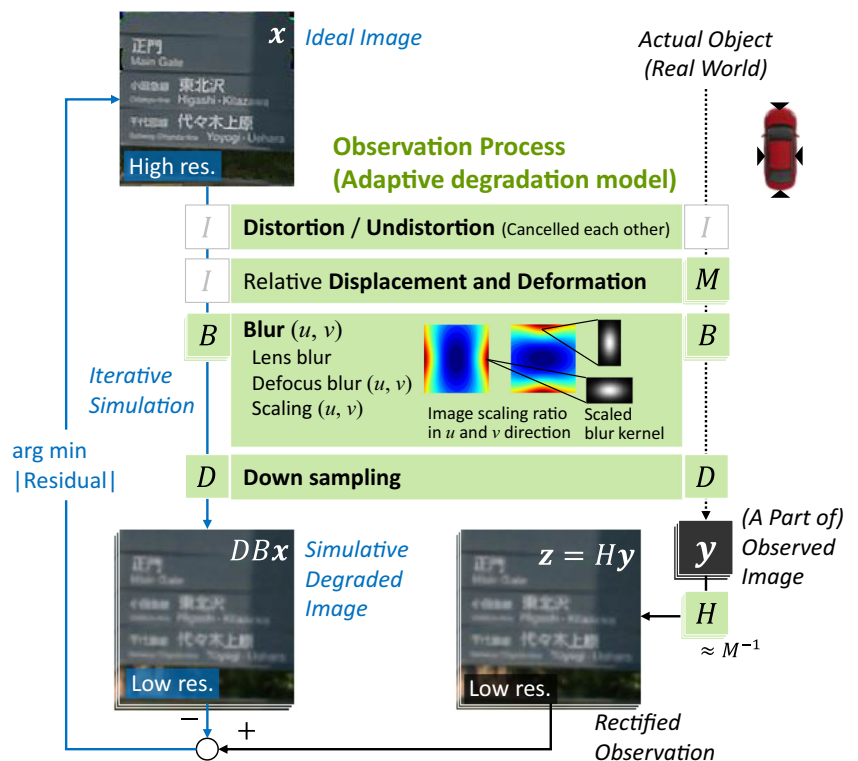
H_k absorbs the deformation and displacement of the images y_k and rectify them into z_k . c is the index of the image that has the lowest blur, i.e., y_c has the highest image quality among the observed images. It is typically given when the targeted object is observed at the nearest point to the lens center. Hereinafter, we call this image as *reference image*.

The approximate solution can be found through the following iterative calculation [11] as shown in Eq. (4),

$$^{n+1}x = ^n x \beta \left\{ \sum_k B_k^T D_k^T \text{sign}(D_k B_k ^n x - z_k) + \lambda \sum_l \sum_m \left\{ \alpha^{|l|+|m|} \left[I - S_y^{-m} S_x^{-l} \right] \text{sign} (^n x - S_x^l S_y^m nx) \right\} \right\} \tag{4}$$

($-p \leq l \leq p, 0 \leq m \leq p, l + m \geq 0$)

Fig. 2 Process of SR to enhance image quality. The adaptive degradation model in our prior work [10] enables to deal with fisheye images properly



where n is a count of iterations, β is a convergence step, λ is a smoothness constraint, α is an attenuation by distance, and S_x^l, S_y^m are image translations.

The detailed estimation of the deformable image registration matrix H_k and the blur matrix B_k is described in 3.3 and 3.4.

3.3 Blur Estimation

The blur matrix B_k consists of four components:

1. The blur generated by the lens itself by the optical interference (point spread function).
2. The defocus blur depending on the depth of field
3. Image scaling effect resulting from undistortion.
4. The motion blur.

These components are assumed to be integrated and represented approximately by convoluting a two-

Gaussian function, and the standard deviation of the Gaussian function corresponds to the degree of blur. We refer to this value as the “blur size”.

The lens blur is obtained by the edge-based MTF computation method [12] applied to the image observed in the center area of the lens. In this method, the lens blur is calculated by the following three steps: (1) find the region in which the edges continue diagonally, (2) calculate the differential value of the brightness in the region, (3) calculate an approximate curve by the Gaussian curve fitting and find a standard deviation.

The defocus blur is not assumed, i.e. the focus of the lens is fixed to the object depth. The image scaling ratio caused by undistortion can be calculated from lens distortion parameters for each pixel.

Motion blur should be reflected in B_k . In this study, we assume that the lighting and shooting condition is enough, i.e., illumination is high and/or the camera moves slow so that the motion blur can be negligible. In our prior work [13], we have proposed the integrated method to estimate motion blur and to perform SR, where the motion blur is not negligible.

3.4 Adaptive Deformable Image Registration

The deformation and the displacement can be represented by homography conversion, assuming that the targeted region locally be regarded as planar. By applying the homography matrices H_k to y_k , all the y_k are registered to the *reference image* y_c .

H_k can be found using image-to-image block matching and point-to-point correspondences inside the images by using Speeded-up Robust Features (SURF). Through this process, images are theoretically registered in sub-pixel order including rotation and deformation.

3.5 Evaluation Method

Peak signal-to-noise ratio (PSNR) is a well-known metric for evaluating image degradation and restoration. However, it requires a ground-truth image for comparison, and it is unknown in SR of real images. It is possible to capture images, perform SR using intentionally down-sampled images, and compare its result with the captured image. However, the captured image is not the ground truth since it includes a blur effect by the lens. In this sense, it is not easy to evaluate the performance of SR of real images.

In this study, we focus on edges in the image because SR generally restores the high-frequency components. From this viewpoint, we introduce the average edge intensity of the entire image for an evaluation metric for convenience. The improvement of the image quality by SR is judged by the magnitude of the edge intensity unless ringing effect (artifacts at edge boundaries, details in Fig. 10) occurs, since it should be large if the high frequency components of the objects are restored by SR. The metric used here does not need to be able to evaluate the absolute image quality, but the relative quality to analyze the trade-off problem described in the next section. Although it is possible to introduce other general metric such as the accuracy of character recognition, it is considerable only if the application is specified and not appropriate for basic analysis.

4 Quantitative Analysis of the Trade-off Problem

In order to quantitatively analyze the how choice of observed images affect result of SR, we acquire experimental data and test SR under several conditions on inclusion/exclusion of input images.

4.1 Acquisition of the Experiment Data

We successively observe signboards by an on-vehicle fisheye camera. The vehicle proceeds in a straight line parallel to a planar signboard. The signboard is on the left side of the path and the camera is directed to the left, where this arrangement is similar to the actual parking assistance camera equipped on the side mirror. The distance between the board and the line (depth to the object) are one, two, and four meters.

We used Canon EOS Kiss Digital X / Sigma 8 mm F3.5 EX DG Circular Fisheye for the camera and the lens, with its focal length fixed. Figure 3 shows some examples of the captured images, whose size is 3792×3792 pixels. We assume this specification is reasonable since on-vehicle

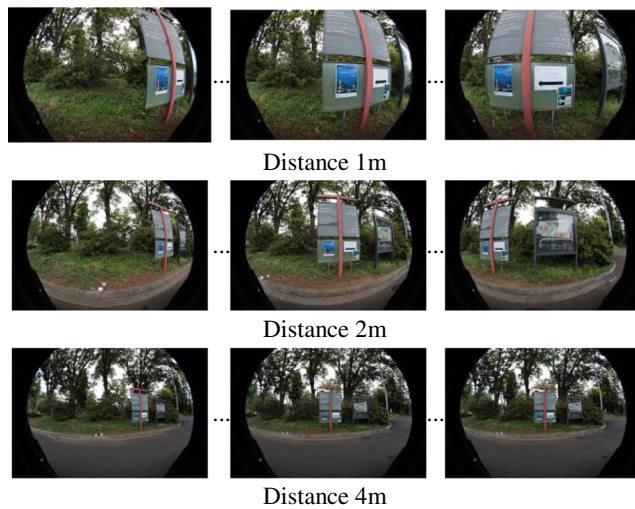


Fig. 3 Examples of fisheye images captured by the on-vehicle camera directed to lateral direction

camera are practically expected to have higher resolution in the future.

4.2 Deformable Image Registration and Blur Estimation

As pre-processing, the captured images are first undistorted, i.e. converted into perspective images by removing lens distortion. We use calculation method and tool of [14] for undistortion. Figure 4 shows the images after undistortion. In addition, to normalize roughly the apparent dimensions of the signboard for later comparison, the undistorted images captured at one meter and two-meter distances are resized into 1/4 and 1/2 scale, respectively.

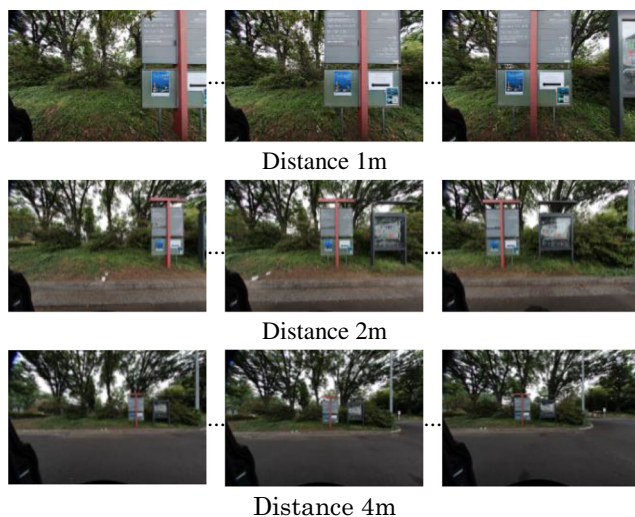


Fig. 4 Examples of the undistorted images

Then, five rectangular ROIs, A to E (each 64×64 pixels) for SR, are extracted from the observed images. These images include characters and patterns that are similar to the ones generally found in roadside signs and license plates. For each sequential frame, the regions corresponding to A to E are tracked using block matching and registered with image deformation. Hereafter, the registered sequential image sets of ROIs with the shooting depth d (m), are described as Z_d^k , i.e., $Z = \{A, B, C, D, E\}$ ($d = 1, 2, 4$). Figure 5 shows the examples.

The deformable image registration is realized by applying matrix H_k obtained by point correspondences of SURF feature between each input images Z_d^k and the *reference image* explained in 3.2. Figure 6 shows example of the correspondences. Since some input images captured at near the peripheral area of the fisheye lens are greatly influenced by the lens blur and the scaling due to the undistortion, it is generally not stable to align such images with the reference image correctly.

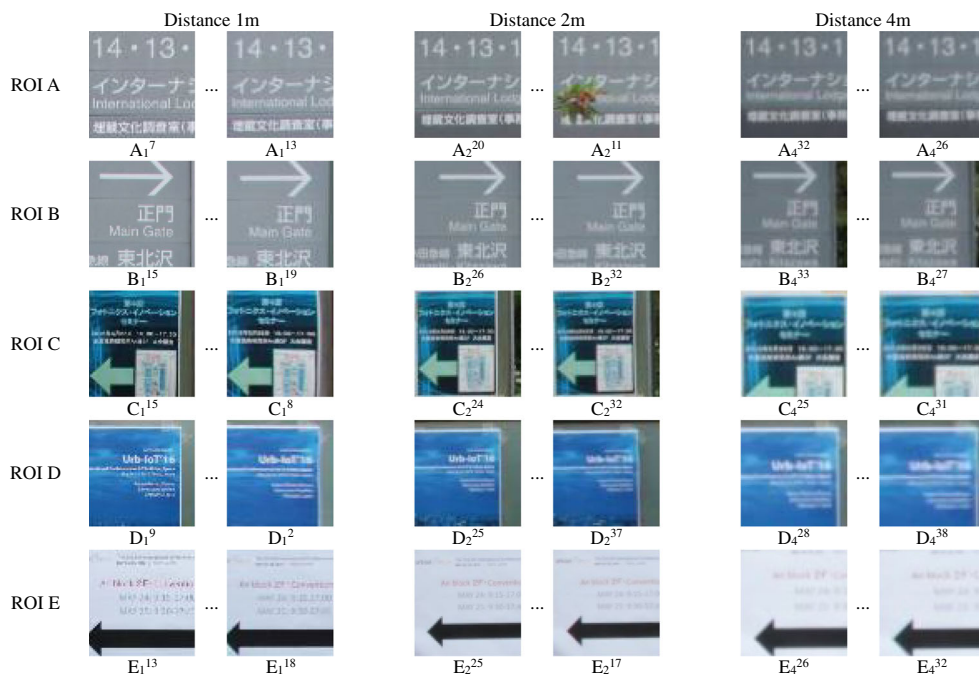
As mentioned in 3.3, the blur size of each Z_d^k is calculated by the convolution of the lens blur and the scaling factor. Table 2 shows the lens blur obtained from the edge-based MTF computation method. Since we apply this method to the images after resampling by undistortion process, the blur values should be corrected by the scaling factor. Figure 7 shows the per-pixel scaling factor calculated by the method of [14]. Figure 8 plots the average blur size of each Z_d^k (A_1 and B_1 for this example) obtained by the convolution of the lens blur and the scaling factor. In A_1 sequence, the blur size decreases from A_1^1 to A_1^7 and increases from A_1^7 to A_1^{16} sequentially (i.e., A_1^7 is the *reference image* of A_1).

4.3 Choosing Input Image Sets

We select several input images for SR from A_d to E_d and examine how the SR result varies depending on the input images, and especially their blur sizes. To select the input images for SR, the images are first sorted by ascending order of the blur size. As shown in Fig. 8, for example, in A_1 and B_1 , the orders are $(A_1^7, A_1^6, A_1^8, \dots, A_1^{16})$ and $(B_1^{14}, B_1^{13}, B_1^{15}, \dots, B_1^{23})$, respectively. In $A_2, B_2, A_4,$ and B_4 , only some reference images indicated with large dots with image indices in Fig. 8 are subject to selection.

Then with these orders, the first four images are selected as the first sets, and the first six, eight, ten, and further more images are selected as the second, third, fourth, and further sets. For example, in case of four image sets at A_1 , $(A_1^7, A_1^6, A_1^8, A_1^5)$ are selected and the minimum blur size is 0.7, the maximum blur size is 0.79, and the average blur size is 0.78. Furthermore, we examine in detail the image sets before and after the input image number where the average edge intensity of SR result becomes maximum. For example, if the average edge intensity becomes maximum at six of the input number, we examine five and seven additionally.

Fig. 5 Examples of registered image sets. Initial image (x_0) in SR and the boundary image that will maximize the edge intensity of a SR result are chosen for each image set



Throughout all of the inclusion/exclusion patterns, the minimum number of input images is four, and the maximum blur size is 2.72. The more input images are chosen, the larger the amount of information there is, but at the same time, low-quality image with larger blur size and lower registration accuracy are included.

4.4 Super Resolution

SR is performed to increase the resolution twice, generating a 128×128 pixels images as output from the 64×64 pixels input images.

Finding the ideal image x through the iterative calculation requires the initial image x_0 . For this, we use the image resized

by linear interpolation from the *reference image* explained in 3.2. For example, we choose A_1^7 and B_1^{14} in A_1 and B_1 sequence, respectively.

The parameters in (2) are set as $\beta = 1.0, \lambda = 0.2, \alpha = 0.5, p = 3$.

4.5 SR Result and Evaluation

Figure 9 shows examples of SR results, after 1000 times iteration, from A_1 to B_4 image sequences. To clarify the effect of SR, initial images x_0 (linear interpolation of the reference images) are also indicated for comparison. It can be recognized

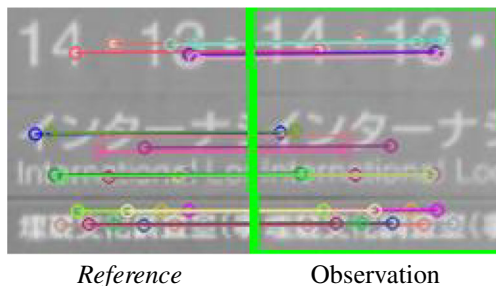


Fig. 6 Example of SURF feature correspondences that find the homography matrix for rectification of each. The left-hand image is A_1 which is the initial image in SR and the right-hand image is A_{14} . After rectifying the observed images, they are registered with the reference image as the green tetragons show the outline of the rectified images

Table 2 Lens blur size after scaling ratio correction, which is not available at A_4, D_1, D_2, D_4 due to the unclear edge

Image sets	Lens blur size	Image sets	Lens blur size
A_1	0.33	D_1	(0.39)
A_2	0.53	D_2	(0.48)
A_4	(0.77)	D_4	(0.83)
B_1	0.27	E_1	0.39
B_2	0.53	E_2	0.48
B_4	0.77	E_4	0.83
C_1	0.36		
C_2	0.52		
C_4	0.71		

Instead, they were substituted with that of $B_4, E_1, E_2,$ and E_4 respectively, since difference of the blur conditions (shooting distance and ROI position in the image) are negligible.

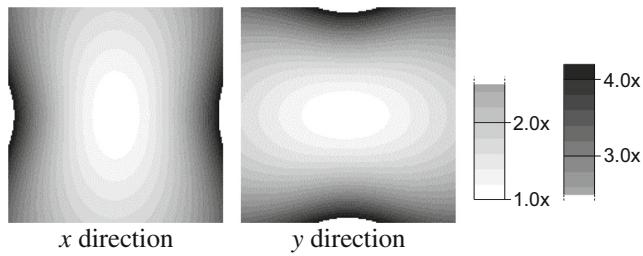


Fig. 7 Per-pixel image scaling factor caused by correcting distortion of the fisheye lens. While the central area is not scaled, the peripheral area is expanded more than four times

in appearance that the resolution is enhanced and high-frequency component is restored.

For evaluating these results quantitatively, the average of edge intensity within the entire image is calculated by Sobel

filtering. Although the edge intensity increases when some ringing effect occurs, it is not observed in Fig. 10. In this sense, this evaluation score can be effective to evaluate SR results.

Figure 11 shows results of the evaluation. The maximum score means the best SR result. For example, the highest score is achieved when eleven images are input from A_1 , and seven images from A_2 and A_4 . The same tendency can also be seen for B_1 to E_4 that the highest score is achieved at some number of input images and giving excessive number of input images leads the SR result even worse.

5 Discussion

5.1 Regulation to Improve Output Quality from Fisheye Images Collected by on-Vehicle Camera

Normally in reconstruction-based SR, the output quality Q is considered to be improved as the number of input images N whose blur size are almost the same increases because the edge components of the objects are restored by SR process with richer information. In that case the smaller the blur size of the input images is, the better the output quality Q can be. On the other hand, where there is a trade-off between number/quality of the input images, Q may not be improved even if N increases. This can be caused by the degradation of the input quality that occurs when the N increases, and accordingly the errors in deformable image registration and blur estimation are expanded.

In case where there is such a trade-off, we are quite interested in what kind of factor in the SR causes the output quality Q evaluated by the average edge intensity and how to find $N = N_0$ that maximizes the output quality Q in advance.

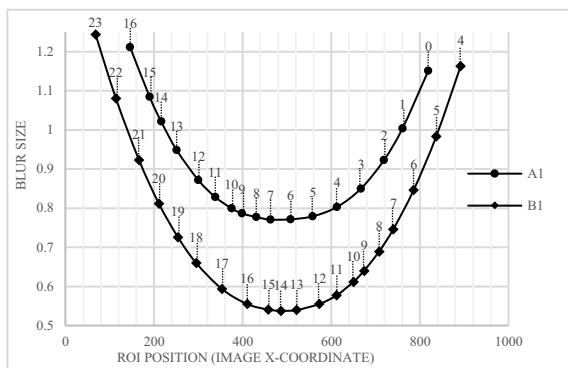
First of all, we quantify the input image quality by two measures: (1) the maximum blur size σ_{\max} and (2) the average blur size $\bar{\sigma}$, when Q is maximum (smaller value corresponds to better quality).

For example, in the A_1 case, from Fig. 11(a), Q becomes maximum when $N = N_0 = 11$ input images are used. From Fig. 8(a), the input images are $(A_1^2, A_1^3, \dots, A_1^{12})$ since images with smaller blur size are chosen for SR by priority. Here,

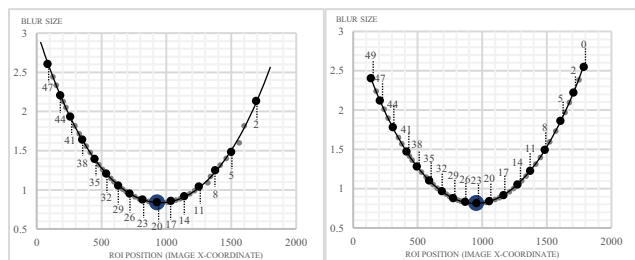
$$\sigma_{\max} = \max\{\sigma_2 \dots \sigma_{12}\} = \sigma_2 \tag{5}$$

$$\bar{\sigma} = \frac{1}{11}(\sigma_2 + \dots + \sigma_{12}) \tag{6}$$

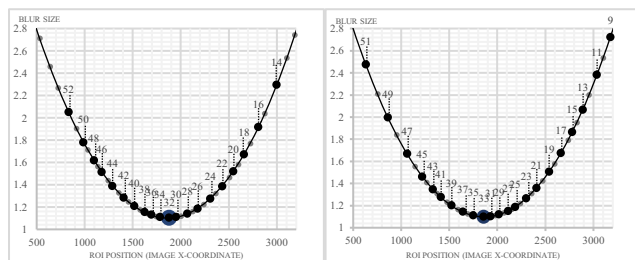
where σ_k represents the blur size of A_1^k . Fig. 12 plots the input quality for all image sequences. σ_{\max} and $\bar{\sigma}$ have variations from 0.68 to 1.48 and from 0.58 to 1.30 respectively. It denotes that the causative factor to specify the SR quality cannot be explained only by input quality.



(a)



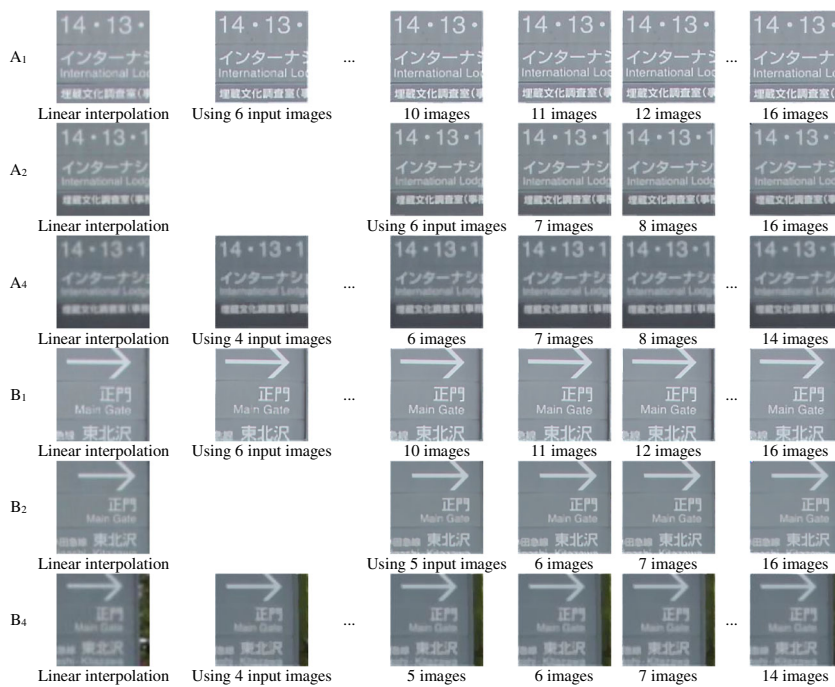
(b)



(c)

Fig. 8 Total blur size of ROI including both lens blur and scaling factor. The lens blur is constant, and the scaling factor comes from the lens character expressed by analytic model. The indices represents order of the capture in the sequence. Images chosen for SR are indicated by dots and indices. Images with smaller blur size are chosen by priority. **a** A_1 and B_1 , **b** A_2 and B_2 , **c** A_4 and B_4

Fig. 9 Result images of SR after 1000 times iterative calculation, using various number of input images (the linear interpolation of the reference image is just for comparison and not a result of SR)



Next, we examine the relation between the input quality and N_0 , number of input images when Q is maximum. From the result in Fig. 13, enough correlation was not recognized between these factors. The reason is assumed to be that if a lot of images with low quality are included

in the input of SR, a lot of errors will occur in deformable image registration, and the output quality cannot be certified any more.

Finally, we focus on quality of the *reference image*. Generally, when objects are captured in distance, the image quality becomes relatively low. For example, while range of the blur size of A_1 captured in 1 m distance is 0.72 to 1.24, that of A_4 is 1.10 to 1.52. We examined the relation between σ_{\min} and σ_{\max} , and between σ_{\min} and $\bar{\sigma}$. Here, σ_{\min} is the minimum blur size among input image sequence, which is equal to blur size of the *reference image*. As can be seen from the result in Fig. 14, quite a clear relation is found that both σ_{\max} and $\bar{\sigma}$ are almost proportional to σ_{\min} throughout all experimental cases. The proportionality factors were $\sigma_{\max}/\sigma_{\min} = 1.17$ and $\bar{\sigma}/\sigma_{\min} = 1.06$.

This means that when σ_{\min} is relatively small, lower-quality images should not be included for inputs to achieve better SR result, and conversely when σ_{\min} is relatively large, lower-quality images are acceptable for inputs. In this sense, the quality of the *reference image* that has the highest quality among the input images can be the key for criteria.

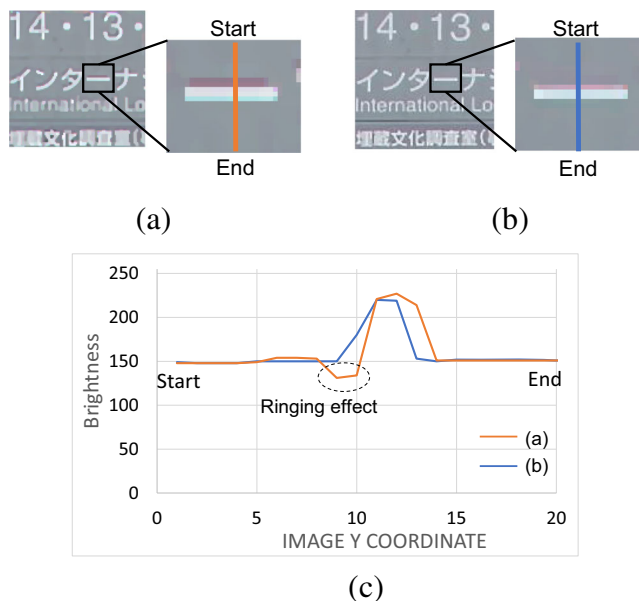


Fig. 10 Example of SR result with and without ringing, an artifact that appears at edge boundaries caused by overshoot of the brightness. From the Fig. 9, no ringing is observed in any SR results. **a** SR result with ringing, **b** SR result without ringing. (Fig. 9, A1 using 10 input images), **c** Brightness of the image **a** and **b**, cross section from top to bottom of the images

5.2 Reason for the Relation

Then, why σ_{\max} and $\bar{\sigma}$ can be proportional to σ_{\min} is the question. The *reference image* has the highest image quality in the input image set, and improvement of image quality by the SR starts relatively from 0x made of the *reference image*. It is considered that the improvement is

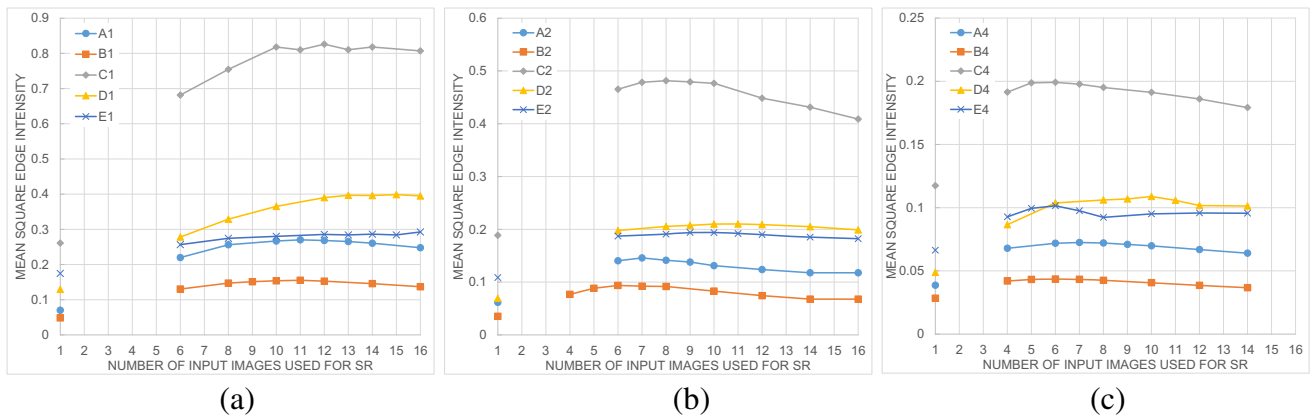
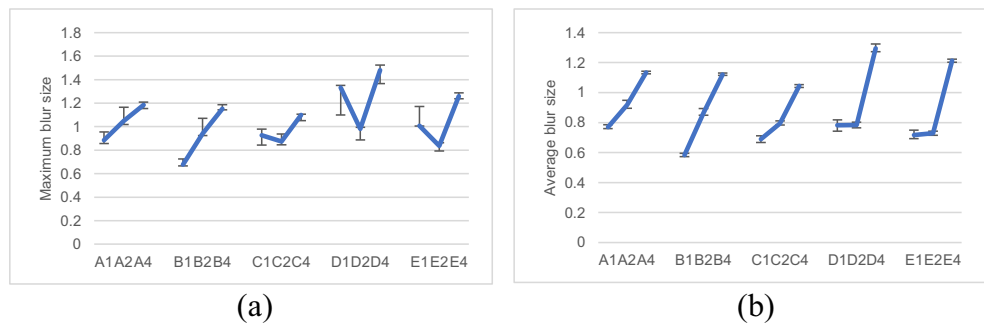


Fig. 11 Effect of SR evaluated by mean square of edge intensity (using one input image simply means linear interpolation of the reference image). Existence of a peak for each graph quantitatively denotes that the best SR result is realized by some number of input images and giving excessive number of input images leads worse SR result. **a** Distance 1 m, **b** Distance 2 m, **c** Distance 4 m

Fig. 12 Quality of the input image sets when the highest SR quality Q in terms of edge intensity is achieved. The variation of the values means that the causative factor to specify the SR quality cannot be explained only by input quality. **a** Maximum blur σ_{max} of input, **b** Average blur $\bar{\sigma}$ of input



contributed by two possible factors: (1) the number effect and (2) the initial quality effect. Here, the number effect is the effect that the output quality is improved as the input number is increased, and the initial quality effect is the effect that the higher output quality is achieved as the quality of the initial image σ_0 is higher. From the result,

the knowledge in the following is inferred and this can be the reason of the question above.

- The number effect appears dominantly when the input quality is relatively low, i.e., both the *reference image* and other images do not have high image quality.

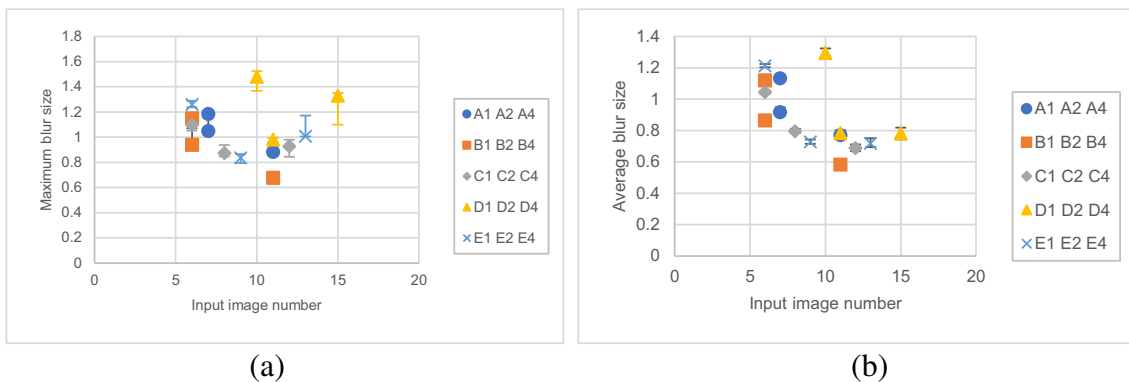


Fig. 13 Relation between input image quality (σ_{max} and $\bar{\sigma}$) and number of input images when the best SR quality is achieved (N_0) for each input image set. Not enough correlation is recognized between these factors. **a** σ_{max} vs N_0 , **b** $\bar{\sigma}$ vs N_0

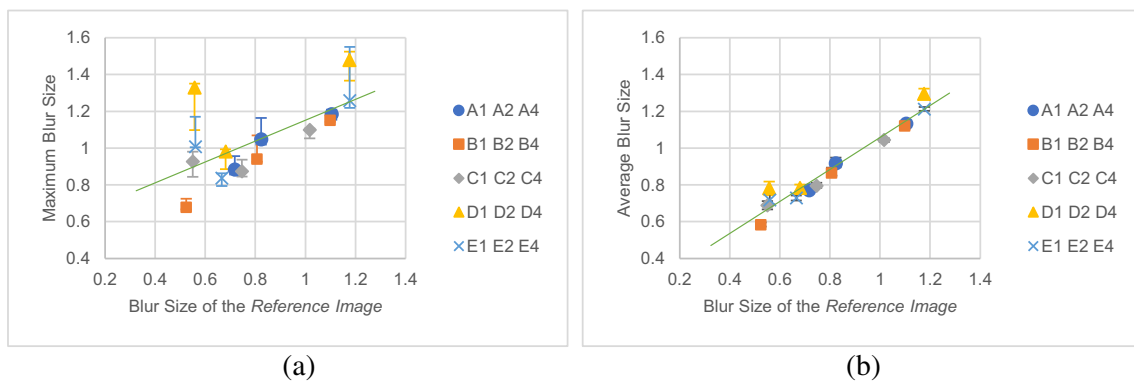


Fig. 14 Relation between quality of input image (σ_{\max} and $\bar{\sigma}$) and quality of the *reference image* σ_{\min} when the best SR quality is achieved for each input image set. Clear proportional relations are found, where the proportionality factors were $\sigma_{\max}/\sigma_{\min} = 1.17$ and $\bar{\sigma}/\sigma_{\min} = 1.06$. **a** σ_{\max} vs σ_{\min} , **b** $\bar{\sigma}$ vs σ_{\min}

- The initial quality effect appears dominantly when the *reference image* has relatively high quality.

6 Conclusion

In this study, in view of the possibility of road sensing by general-use vehicles, we considered and solved several issues on acquiring high-quality images from low-quality images collected by the on-vehicle camera.

First, we pointed out the problem of the trade-off problem on the quality and quantity of the images collected by typical on-vehicle cameras of general-use vehicles. Second, we developed the SR method that can deal with such collected images by introducing adaptive deformation and blur model. And third, we found out the criterion of inclusion/exclusion of input images for SR to achieve highest image quality in result can be clearly given by the ratio of maximum or average blur size and minimum blur size among the input images.

The image with minimum blur size is typically captured at the center of the lens, which is the key for criteria. If the input images have relatively high image quality when a targeted object is close to the camera for example, even higher image quality will be realized by SR, even if images with lower quality captured at the peripheral area of the lens are input to SR. On the other hand, if the input images have relatively low quality and even the image captured at the center of the lens does not have high quality, giving relatively more input images to SR will make the result worse.

The future work in technical aspect includes the followings: improving the blur estimation process and the regulation process, comparison with other SR methods, experiments using more commonly-used inexpensive cameras, and consideration of motion blur. It is also possible to apply this method to the image achieved by the camera installed in front shield of

vehicle. Taking advantage of these knowledge, constructing a visual sensing system in road environment by general-use vehicles is also the future work in the aspect of applied use.

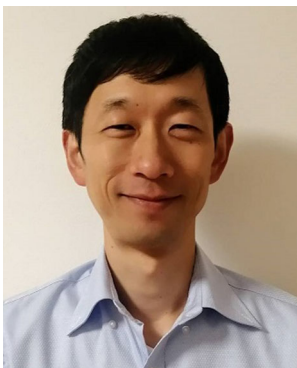
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Pioneer Corporation: Pioneer Introduces New CYBER NAVI Car Navigation Systems for Japan Market", <https://global.pioneer/en/news/press/index/1617> (2013). Accessed 1 Jun 2020
2. Toyota Motor Corporation: Toyota Plans Trial Provision of Lane-specific Traffic-con-gestion Information Based on Driving Video Data from 500 Tokyo Taxis", <https://newsroom.toyota.co.jp/en/detail/19231576/> (2017). Accessed 1 Jun 2020
3. Kataoka, H., Suzuki, T., Oikawa, S., Matsui, Y., Satoh, Y.: Drive Video Analysis for the Detection of Traffic Near-Miss Incidents", IEEE International Conference on Robotics and Automation (ICRA). <https://ieeexplore.ieee.org/document/8460812> (2018)
4. Chan, F.-H., Chen, Y.-T., Xiang, Y., Sun, M.: Anticipating Accidents in Dash-cam Videos, ECCV, pp 184–199 (2016)
5. Dong, C., Change Loy, C., He, K., Tnag, X.: Image Super-Resolution Using Deep Convolutional Networks. IEEE Trans. Pattern Anal. Mach. Intell. **38**(2), 295–307 (2016)
6. Kawasaki, H., Ikeuchi, K., Sakauchi, M.: Super-resolution omnidirectional camera images using Spatio-temporal analysis. Electro. Commun. Japan. **89**(6), 47–59 (2006)
7. Nagahara, H., Yagi, Y., Yachida, M.: Super-Resolution from an Omnidirectional Image Sequence. Proc. 26th Annual Conf. of the IEEE Industrial Electronics Society (IECON). **4**, 2559–2564 (2000)

8. Fan, Z., Qi-dan, Z.: Super-Resolution Image Reconstruction for Omni-Vision Based on POCs. Proc. 21st Annual Int'l Conf. Chin. Control. Decis (CCDC), 5045–5049. <https://ieeexplore.ieee.org/document/5194962> (2009)
9. Arican, Z., Frossard, P.: Joint Registration and Super-Resolution With Omnidirectional Images. IEEE Trans. Image Process. **20**(11), 3151–3162 (2011)
10. Takano, T., Ono, S., Matsushita, Y., Kawasaki, H., Ikeuchi, K.: Super Resolution with Fisheye Camera Images for Visibility Support of Vehicle, IEEE International Conf. on Vehicular Electronics and Safety (ICVES). <https://ieeexplore.ieee.org/document/7396905> (2015)
11. Farsiu, S., Robinson, D., Elad, M., Milanfar, P.: Fast and robust multiframe super resolution: IEEE trans. Image Process. **13**(10), 1327–1344 (2004)
12. Photography –Electronic Still Picture Cameras–Resolution Measurements, ISO Standard 12233 2000
13. Matsushita, Y., Kawasaki, H., Takano, T., Ono, S., Ikeuchi, K.: Joint Technique of Fine Object Boundary Recovery and Foreground Image Deblur for Video Including Moving Objects: 13th International Conference on Quality Control by Artificial Vision. <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10338/103380R/Joint-technique-of-fine-object-boundary-recovery-and-foreground-image/10.1117/12.2266746.full> (2017)
14. Scaramuzza, D., Martinelli, A., Siegwart, R.: A Toolbox for Easy Calibrating Omnidirectional Cameras, Proceedings to IEEE International Conference on Intelligent Robots and Systems (IROS). <https://ieeexplore.ieee.org/document/4059340> (2006)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Teruhisa Takano is formerly a Ph.D candidate student in Graduate School of Interdisciplinary Information Studies, The University of Tokyo, Japan, and currently working in Nissan Motor Corporation. His research interest includes reconstructing the real world from image data in real time, using visual processing. He received Master degree in Engineering from Department of Applied Physics, The University of Tokyo, in 2000.



Honda Research Institute Japan. He is a member of IEEE, IEICE, IPSJ, IEEJ and JSAE.

Shintaro Ono is a project associate professor of Institute of Industrial Science, The University of Tokyo, Japan, and a visiting researcher of Aichi Prefectural University, Japan. His research interest includes sensing and visualization technology for advanced mobility and infrastructure. He received Master and Dr. degree in Information Science and Technology from The Univ. of Tokyo in 2003 and 2006, respectively. During 2017–2020, he worked as a senior researcher in



Kagoshima Univ. and associate professor at Saitama Univ. He was also a visiting professor at INRIA Rhone Alpes in France, Rice and Columbia Univ. in the US. He has published over 200 reviewed papers including ICCV, CVPR, ECCV, PAMI, IJCV and Eurographics in computer vision and graphics areas. He won several outstanding awards including Songde Ma Prize (the best paper of ACCV) in 2007, the best paper on PSIVT in 2009, Nagao Prize (the best paper on MIRU) in 2011, yearly best paper award on IPSJ in 2013, the best paper at Interaction 2015, the best paper honorable mention at WACV 2017 and the best paper on PSIVT in 2019. He is a member of VRSJ, IPSJ, IEICE and IEEE.

Hiroshi Kawasaki is a full professor of the Department of Information Science and Electrical Engineering, Kyushu University, Japan. He is currently researching on capturing 3D shape and texture of moving objects. Applications on VR and AR systems for medical, robotics, underwater and transportation systems are also important research topics. He received Ph.D degree from Univ. of Tokyo, Japan, in 2003. Prior to Kyushu Univ., he was a full professor at



Katsushi Ikeuchi Sr. Principal Research Manager, Applied Robotics Research, Microsoft, and Emeritus Professor, The University of Tokyo, Japan. He received the BE degree in Mechanical Engineering from Kyoto University in 1973 and the Ph.D degree in Information Engineering from The University of Tokyo in 1978. After working at the Artificial Intelligence Laboratory, Massachusetts Institute of Technology for three years, Electrotechnical

Laboratory, Ministry of International Trade and Industry, Japanese government for five years, and Robotics Institute, Carnegie Mellon University for 10 years, Institute of Industrial Science, The University of Tokyo for 19 years, he joined Microsoft in 2015. His research interest spans computer vision, robotics, and computer graphics. In the field of computer vision, he proposed the so-called “smoothness constraints,” near-by pixels have similar visual characteristics and a variational approach with this constraint in shape-from-shading algorithm. Later, this smoothness constraint is extended as the regularization term. His original paper, “Numerical Shape from Shading and Occluding Boundaries,” was selected as one of the most influential papers to have appeared in the Artificial Intelligence Journal within the past 10 years. In the field of

robotics, he created the field of learning-from-observation, where a robot learns how to do tasks from observing human performance. One of his papers on this topic, “Toward Automatic Robot Instruction from Perception-Mapping Human Grasps to Manipulator Grasps” won IEEE-RAS K-S Fu memorial best transaction paper award. He also started an area, referred to as e-Heritage, to digitally preserve cultural heritage by using computer vision and computer graphics technique. The U Tokyo team, led by him, successfully digitize Bayon, Angkor wat and Preah Vihear temples in Cambodia and through this project, his team won best paper awards three times from VSMM conference series. His community service includes: general chair of IROS95, ITSC99, IV01, ICCV05, ACCV07, ICCV17; program chair of CVPR96, ICCV03, ICRA09, ICPR12, ICCV15; EIC of IJCV (2000-2017), EIC of IJ ITS (2012-2014), associate editor of IEEE Trans. RA, IEEE Trans. PAMI; and a distinguished lecturer of IEEE Signal Processing Society in 2000-2002, IEEE Robotics and Automation Society in 2004-2006 and IEEE Computer Society in 2008-2010. Through these research and society service, he was awarded a (life) fellow from IEEE, IEICE, IPSJ and RSJ. He received the Distinguished Researcher Award from IEEE-PAMI, the Medal of Honor with Purple Ribbon from Japanese Emperor, the Okawa prize from Okawa Foundation as well as a couple of research achievement awards from Japanese professional societies.