17th International Conference on Machine Vision Applications (MVA)
Fully Online, July 25–27, 2021.

AVM Image Quality Enhancement by Synthetic Image Learning for Supervised Deblurring

Kazutoshi Akita Masayoshi Hayama Haruya Kyutoku Norimichi Ukita
Toyota Technological Institute, Japan
{sd21501, sd17055, kyutoku, ukita}@toyota-ti.ac.jp

Abstract

An Around View Monitoring (AVM) system is widely used to allow a driver to watch the situation around a car. The AVM image is generated by image distortion correction and viewpoint transformation for images captured by wide view-angle cameras installed on the car. However, the AVM image is blurred due to these transformations. This blur impairs the visibility of the driver. While many deblurring methods based on CNN have been proposed, these general-purpose deblurring methods are not designed for the AVM image. (1) Since the blur level in the AVM image is region-dependent, deblurring for the AVM should also be region-dependent. (2) Furthermore, while supervised deblurring methods require a pair of input-blurred and output-deblurred images, it is not easy to collect the deblurred AVM image. This paper proposes a method for generating the pairs of training images that cope with the aforementioned two problems. These training images are generated by the inverse transformation of the AVM image generation process. Experimental results show that our method can suppress blur on AVM images. We also confirmed that even a very shallow CNN with the inference time of 2.1ms has the same performance as the SoTA model.

1 Introduction

Various advanced driver assistance systems (e.g.: automatic braking, lane deviation alarm) have been developed and put into practical use. These systems reduce the number of traffic accidents. On the other hand, the number of accidents in parking lots is unchanged [1]. Most of the accidents in parking lots are caused by a failure to check the safety of the surrounding. This is because many distraction factors (e.g.: steering, searching a parking spot) disturb drivers.

An Around View Monitoring (AVM) system, which is one of the advanced driver assistance systems, provides a solution for the problem in understanding the dynamic situation around the car. The AVM image is generated by the distortion correction and the viewpoint transformation of wide-angle camera images, as shown in Fig. 1. This system allows us to watch surroundings at once and compensate for the driver's blind spots. This system has the potential to prevent the accidents mentioned above, and many recent cars are

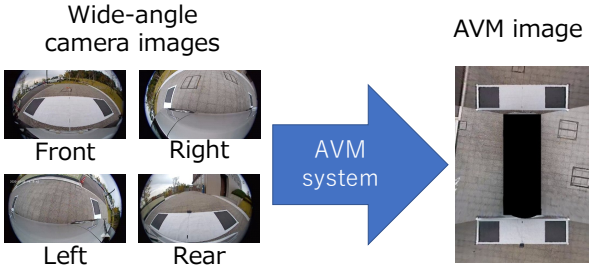

Figure 1. Overview of the AVM system. AVM images are generated from wide-angle camera images installed on all four sides of a car.

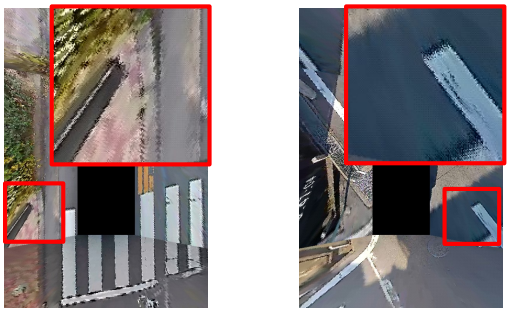

Figure 2. Examples of blurred AVM images. The large red box in the upper left corner is a zoomed-in version of the small red box for better visualization.

equipped with this system. Although this is very useful, sometimes the AVM image is severely blurred due to its generation process, as shown in Fig. 2). These blurs hurt the driver's perception for understanding the surrounding environment.

In recent years, with the development of CNNs, many high-performance deblurring methods are proposed [2, 3, 4, 5, 6, 7]. These methods might have the potential to suppress blur in the AVM image. However, these general-purpose deblurring methods are not designed for the AVM image. First, since the blur level in the AVM image is region-dependent, deblurring for the AVM image should also be region-dependent. Second, since basic CNNs are based on supervised learning, they require a pair of images with and without blur. While most of the deblurring methods use uniformly-and-artificially blurred images for training, it is re-

© 2021 MVA Organization

vealed that models trained on such images do not have sufficient performance on real images [6, 7] due to the gap between artificial and real blur. To perform well on real-world images, we need real-world blurry and non-blurry image pairs. In addition, these non-blurry images must be identical to the AVM image. However, in order to obtain such AVM images, it is necessary to prepare a camera fixed on the viewpoint of the AVM image without any support being visible in the camera image. It is, in reality, impossible to capture these AVM images.

Instead of the uniform-and-artificial blur generation process mentioned above, this paper proposes a method to synthetically reproduce the realistic blur in AVM images by the inverse transformation of the AVM image generation process. Experimental results show that the model trained on the images generated by our method performs sufficiently on severely blurred AVM images. We demonstrate that even a very lightweight CNN with a processing time of 2.1ms performs well for real-time processing.

2 Related Work

2.1 Deblur

Old deblurring methods are based on unsupervised learning, for example, using total variation, sparse coding, and self-similarity. Recently, many CNN-based methods are proposed and outperform the previous unsupervised methods. Since general CNN-based models are trained in a supervised manner, a pair of blurred and non-blurred images are required. Many previous methods [2, 3, 4, 5] synthetically give various blurs (e.g.: Gaussian blur, motion blur) and noise (e.g.: Gaussian noise, white noise) to high-quality images to obtain training pairs. However, models trained on data created in this way sometimes do not perform well on real images. This is because there are domain gaps between the blur/noise in the synthetic and real images. To tackle this problem, Zhang et al. [6] trains a model from unpaired blurred and non-blurred images using a discriminator that distinguishes blur or non-blur. Lehtinen et al. [7] trains the model with the pairs of differently-noised images in the same scene. However, these methods are inferior to supervised methods using images with no domain gap.

2.2 AVM Image Quality Enhancement

The main factor of blur in AVM images is the upscaling operations caused by its generation process. Suzuki et al. [8] proposed a method to optimize the camera parameters (e.g.: camera angle, view angle) to minimize the average upscaling rate of these transforms. Although this method improves the quality of the AVM image, sometimes it is impossible to install the camera with optimal parameters because of the car's size or shape. More essentially, even if this method optimizes the camera locations, blur due to the upscaling operation is unavoidable. Choi et al. [9] enhances the AVM images using super-resolution and sharpness enhancement. In this method, highly-upscaled regions in AVM images are generated with super-resolved and sharpened wide-angle images. However, in this method, super-resolution and sharpness enhancement are achieved by exploring the most appropriate patches in training images based on self-similarity, which also receives a bad influence from domain gaps, while super-resolution and sharpness enhancement can be also improved by CNN (e.g., [10, 11, 12, 13]).

3 Proposed Method

3.1 AVM Image Generation Process

The process of the AVM system is as follows. First, wide-angle camera images are rectified to perspective-camera images by distortion correction. The transformation parameters are obtained by calibrating each camera with a hand-held checkerboard pattern. Next, viewpoint transformation is applied to these perspective-camera images to obtain an AVM image. This viewpoint transformation is expressed by the Homography calibrated with an on-ground checkerboard pattern. Once these transformation processes are calibrated, we make the pixelwise look-up table for mapping each pixel in the wide-angle camera images to the AVM image for efficient online transformation.

3.2 Inverse AVM Image Generation for Realistic Blurred AVM Image Synthesis

Not only deblurring but also other image restoration and enhancement methods require the pairs of input degraded images (denoted by I^{AB}) and output high-quality images with no blur (denoted by I^{AN}) for supervised learning. However, I^{AN} is no available in reality. To train a deblurring model for the AVM image, our proposed method synthetically generates the degraded AVM image with blur (denoted by $\hat{I}^{AB}$) from the high-quality image with no blur (denoted by $\hat{I}^{AN}$).

The overview of our proposed method is shown in Fig. 3. The blur in the AVM image is caused by the image transformation process (denoted by "Transform" in the figure) from the images captured by wide-angle cameras (denoted by I^{C}) to the AVM image. Therefore, in order to reproduce the realistic blur in the degraded AVM image ($\hat{I}^{AB}$) from the image with no blur ($\hat{I}^{AN}$), we apply the inverse transformation of the AVM image generation process (described in Sec.3.1) to $\hat{I}^{AN}$. While any high-quality image can be used as $\hat{I}^{AN}$, images in the DIV2K dataset were used in our experiments. This inversely-transformed image (denoted by $\hat{I}^{C}$) is regarded as the wide-angle image. Then $\hat{I}^{C}$

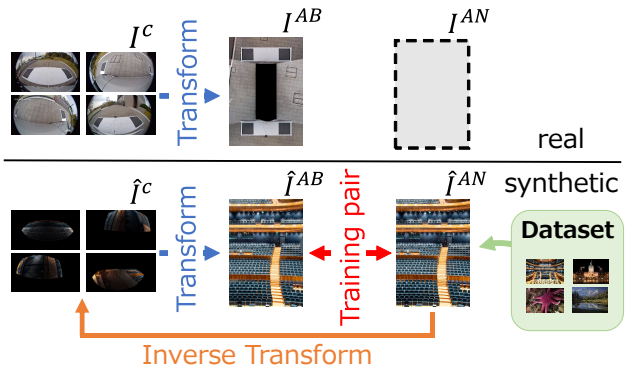

Figure 3. Overview of our proposed method.

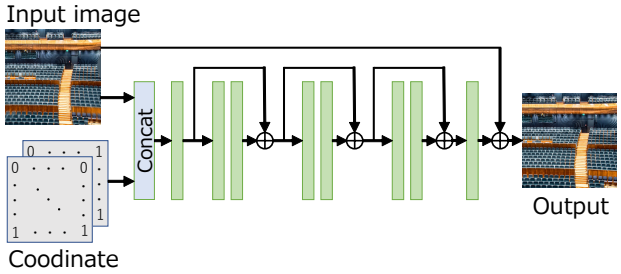

Figure 4. The shallow model architecture used in our experiments. The green rectangle indicates the convolution layer with activation function. $\oplus$ indicates pixel-wise add operation.

is provided to the AVM image generation process for obtaining the synthetic AVM image ($\hat{I}^{AB}$). With the pairs of $\hat{I}^{AB}$ and $\hat{I}^{AN}$, we train our CNN-based deblurring model.

More specifically, the inverse transformation of the AVM image generation process is conducted based on the look-up table obtained in Section 3.1. Since we use nearest-neighbor interpolation for the distortion correction and the Homography transformation in the AVM image generation process, the look-up table has reproducible correspondences between the integer coordinates of the wide-angle camera image and the AVM image. This look-up table allows us to achieve the inverse transformation easily and efficiently.

3.3 Deblurring Network and its Training

While the core contribution of this paper is realistic blurred AVM image generation described in Sec. 3.2, any deblurring method is applicable. In our experiments, we used our proposed shallow model (shown in Fig. 4) and MPRNet [5], which is one of the SoTA models for supervised deblurring. Since the blur in the AVM image depends on the image coordinates (i.e., region-dependent), the deblurring network should be trained in accordance with this region dependency. This region dependency is trained by concatenating the map of the normalized coordinates to the input image. This image concatenated with the coordinate map is fed into the network.

Table 1. Quantitative evaluation. Here, the GFLOPs is given for an input image of 780 × 530 pixels. The unit of the model size is MB.

	Synthetic		Real	Model	
	PSNR	PI	PI	size	GFLOPs
Shallow	33.03	2.843	2.999	0.3	124.3
MPRNet	33.16	2.910	3.002	20.1	4754.3
Input	29.62	3.052	4.078	-	-

For training the models, the following two losses are employed: the reconstruction loss l_{recon} and the perceptual loss l_{VGG}.

The reconstruction loss is given by the following equation:

$$l_{recon} = \frac{1}{WH} \sum_{x=1}^{W} \sum_{y=1}^{H} \left| I_{x,y}^{GT} - I_{x,y}^{Pred} \right|, \tag{1}$$

where I^{GT} and I^{Pred} denote the ground-truth and predicted images, respectively. W and H denote the dimensions of the image.

The perceptual loss is given by the following equation:

$$l_{VGG}^{i} = \frac{1}{W_i H_i} \sum_{x=1}^{W_i} \sum_{y=1}^{H_i} \left| \phi_i(I^{GT})_{x,y} - \phi_i(I^{Pred})_{x,y} \right|, \tag{2}$$

where ϕ_i denotes the feature map obtained by the i-th maxpooling layer within the pretrained VGG19 network. W_i and H_i are the dimensions of the i-th feature map. In our experiments, $i = 5$.

The total loss used to train the models is given by the following equation:

$$l_{total} = l_{recon} + \alpha l_{VGG}^{5}, \tag{3}$$

where α is a weight factor of perceptual loss. In our experiment, we use $\alpha = 0.1$.

For training, we used RAdam [14] optimizer with β=(0.9, 0.999) and the mini-batch size was 32. The learning rate was initialized to 1e-4 and multiplied by 1/10 at 600,000 iterations while total iterations are 750,000. During training, blurred and non-blurred images were randomly cropped into 128 × 128 pixels from 780 × 530 image due to memory constraints. We trained the deblurring models with images given by DIV2K [15]. We split 900 images of DIV2K into 800 training images and 100 evaluation images. In our shallow model, we used parametric ReLU [16] after each convolution layer. Any normalization (e.g., batch normalization, instance normalization) was not used.

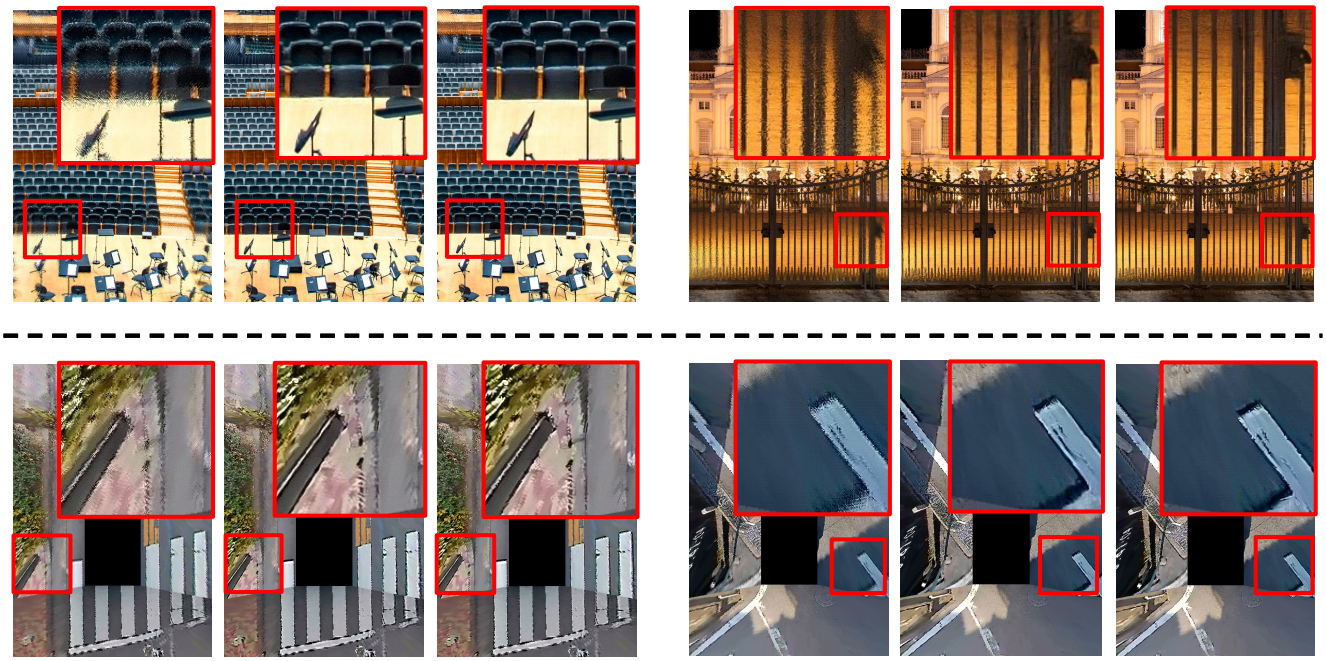

Figure 5. Synthetic and real images are shown in the top and bottom, respectively. The left, center, and right columns in each example are the input, the output of our shallow CNN, and the output of the MPRNet, respectively. Each large red box is a zoomed-in version of each small red box for better visualization.

4 Experiments

We conducted evaluation experiments with two kinds of AVM images. The first one was generated completely as the same way as the blurred AVM image generation process described in Sec. 3.1. Since the training and test images were generated in the same way, deblurring this image is easy. However, there can be a domain gap between the above AVM image and that generated from real wide-angle images. The second one is this AVM image generated from real wide-angle images. In what follows, the first and second AVM images are called the synthetic and real images, respectively. While 800 images in the DIV2K dataset were used for training the deblurring models, other 100 images in the dataset were used for the above synthetic test images. The real images are 111 AVM images generated by real images captured by a commercial AVM system (SABROC SYSTEM) on roads.

The examples of deblurred AVM images are shown in Fig.5. CNN trained by our method can successfully suppress the severe blur in the AVM images. Furthermore, the visibility of test results for the shallow CNNs is comparable to MPRNet. For quantitative evaluation, we use PSNR and the perceptual index (PI) [17], which are the standard metric of the image reconstruction accuracy and the perceptual quality, respectively. PI combines the no-reference image quality measures of Ma et al. [18] and NIQE [19] as follows:

$$PI = \frac{1}{2}((10 - Ma) + NIQE) \tag{4}$$

A lower PI indicates better perceptual quality, i.e., less blur and noise. The results of the quantitative evaluation are shown in Table 1.

From Fig. 5 and Table 1, we can see that the shallow model and MPRNet suppress blur and noise from the input image very well in both synthetic and real images. Furthermore, the performance of the shallow model is comparable to MPRNet, even though the shallow model has a significantly smaller model size and GFLOPs. This should be because our proposed blurred AVM image generation can successfully imitate the image blurring process for generating real AVM images (i.e., less domain gap between the synthetic and real AVM images), and therefore the deblurring process for the real AVM image is easy. In such a simple deblurring problem, even a very shallow model has sufficient performance. Our shallow model takes only 2.1ms per frame to enhance AVM image with 780×530 pixels on the NVIDIA GeForce GTX 1080Ti GPU.

5 Concluding Remarks

This paper proposed a realistic AVM image generation method for supervised deblurring to improve the AVM image quality. Since blurred pixels are generated by image enlargement, future work includes joint learning of deblurring and super-resolution (e.g., perceptual quality [17, 20] and video [21, 22, 23, 24]) for more high-fidelity AVM image generation. This work was supported by JSPS KAKENHI Grant Number 19K12129.

References

[1] Car park accident investigation. https://www.tmpc.or.jp/Portals/0/images/03_business/business/index_01/h30_2_g.pdf. (Accessed on 12/27/2020).

[2] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep CNN denoiser prior for image restoration. In *CVPR*, 2017.

[3] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image restoration. *arXiv*, abs/1812.10477, 2018.

[4] Xingang Pan, Xiaohang Zhan, Bo Dai, Dahua Lin, Chen Change Loy, and Ping Luo. Exploiting deep generative prior for versatile image restoration and manipulation. In *ECCV*, 2020.

[5] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. *arXiv*, abs/2102.02808, 2021.

[6] Kaihao Zhang, Wenhan Luo, Yiran Zhong, Lin Ma, Björn Stenger, Wei Liu, and Hongdong Li. Deblurring by realistic blurring. In *CVPR*, 2020.

[7] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. In *ICML*, 2018.

[8] Masayasu Suzuki. Development of arround view system. *JSAE Annual Congress*, 116(07):17–22, 2007.

[9] Dong-Yoon Choi, Ji-Hoon Choi, Jinwook Choi, and Byung Cheol Song. Sharpness enhancement and super-resolution of around-view monitor images. *T-ITS*, 19(8):2650–2662, 2017.

[10] Radu Timofte et al. NTIRE 2018 challenge on single image super-resolution: Methods and results. In *CVPR Workshop*, 2018.

[11] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Deep back-projection networks for super-resolution. In *CVPR*, 2018.

[12] Muhammad Haris, Greg Shakhnarovich, and Norimichi Ukita. Deep back-projection networks for single image super-resolution. *arXiv*, 1904.05677, 2019.

[13] Shuhang Gu et al. AIM 2019 challenge on image extreme super-resolution: Methods and results. In *ICCV Workshop*, 2019.

[14] Liyuan Liu, Haoming Jiang, Pengcheng He, Weizhu Chen, Xiaodong Liu, Jianfeng Gao, and Jiawei Han. On the variance of the adaptive learning rate and beyond. In *ICLR*, 2020.

[15] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *CVPR Workshops*, 2017.

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *ICCV*, 2015.

[17] Yochai Blau, Roey Mechrez, Radu Timofte, Tomer Michaeli, and Lihi Zelnik-Manor. The 2018 PIRM challenge on perceptual image super-resolution. In *ECCV Workshops*, 2018.

[18] Chao Ma, Chih-Yuan Yang, Xiaokang Yang, and Ming-Hsuan Yang. Learning a no-reference quality metric for single-image super-resolution. *CVIU*, 158:1–16, 2017.

[19] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a "completely blind" image quality analyzer. *Signal Process Letters*, 20(3):209–212, 2013.

[20] Kai Zhang et al. NTIRE 2020 challenge on perceptual extreme super-resolution: Methods and results. In *CVPR Workshop*, 2020.

[21] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Recurrent back-projection network for video super-resolution. In *CVPR*, 2019.

[22] Seungjun Nah et al. NTIRE 2019 challenge on video super-resolution: Methods and results. In *CVPRW*, 2019.

[23] Dario Fuoli et al. AIM 2020 challenge on video extreme super-resolution: Methods and results. In *ECCVW*, 2020.

[24] Muhammad Haris, Greg Shakhnarovich, and Norimichi Ukita. Space-time-aware multi-resolution video enhancement. In *CVPR*, 2020.