



Communication Highway Visibility Estimation in Foggy Weather via Multi-Scale Fusion Network

Pengfei Xiao ^{1,2}, Zhendong Zhang ^{1,2}, Xiaochun Luo ^{1,2}, Jiaqing Sun ^{1,2}, Xuecheng Zhou ^{1,2}, Xixi Yang ^{1,2} and Liang Huang ^{1,2,*}

- ¹ Key Laboratory of Transportation Meteorology, China Meteorological Administration, Nanjing 210019, China; nuist_xiaopengfei@163.com (P.X.); 13851612368@163.com (Z.Z.); jqsun115@163.com (J.S.); zhouxuechengnju@163.com (X.Z.); yangqq2014@1zu.edu.cn (X.Y.)
- ² Jiangsu Provincial Meteorological Service Center, Nanjing 210019, China

* Correspondence: 20171100013@nuist.edu.cn

Abstract: Poor visibility has a significant impact on road safety and can even lead to traffic accidents. The traditional means of visibility monitoring no longer meet the current needs in terms of temporal and spatial accuracy. In this work, we propose a novel deep network architecture for estimating the visibility directly from highway surveillance images. Specifically, we employ several image feature extraction methods to extract detailed structural, spectral, and scene depth features from the images. Next, we design a multi-scale fusion network to adaptively extract and fuse vital features for the purpose of estimating visibility. Furthermore, we create a real-scene dataset for model learning and performance evaluation. Our experiments demonstrate the superiority of our proposed method to the existing methods.

Keywords: visibility estimation; image classification; multi-scale fusion network



Citation: Xiao, P.; Zhang, Z.; Luo, X.; Sun, J.; Zhou, X.; Yang, X.; Huang, L. Highway Visibility Estimation in Foggy Weather via Multi-Scale Fusion Network. *Sensors* **2023**, *23*, 9739. https://doi.org/10.3390/ s23249739

Academic Editors: Ofer Hadar and Yitzhak Yitzhaky

Received: 26 October 2023 Revised: 3 December 2023 Accepted: 6 December 2023 Published: 10 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

1. Introduction

Atmospheric visibility is a measure of how transparent the atmosphere is. For traffic safety, atmospheric visibility monitoring is one of the essential services [1]. Atmospheric visibility can be measured by sensors or visual perception. These sensors include meteorological sensors and forward scatter visibility sensors. However, the meteorological observation network is not dense enough to monitor agglomerate fog. Additionally, visibility meters are costly and inaccurate in non-uniform atmospheric conditions [2].

Another method for estimating visibility is based on visual perception. This method processes real-time images captured through surveillance to estimate visibility. Methods based on visual perception can be divided into three main types: multi-image-based methods, Koschmieder-based methods, and deep-learning-based methods.

The multi-image-based methods require multiple images to establish a relationship between image and visibility. Some researchers use the filtering methods (like the Sobel filter or homomorphic filter) to establish the relationship between image features and visibility [3,4]. However, these methods are sensitive to illumination variations. To overcome this disadvantage, Babari et al. [5] adopted the Lambertianess of the image as the reference to adjust the image contrast. Varjo and Hannuksela [6] proposed a new method based on feature vectors that were projections of the scene images with lighting normalization. Then, the new method was combined with the high-dynamic-range imaging to improve nighttime image quality. The above methods require ground truth data from sensors as a reference, so they are not suitable for real-time visibility estimation. To solve this problem, some methods based on Koschmieder's law are proposed [7]. Koschmieder's law describes the relationship between scene visibility and the extinction coefficient. Hautière et al. have conducted a series of studies along these lines [8–12]. These methods treat the distance between the camera and the furthest visible object as the scene visibility. The scene visibility can be obtained by calculating the Koschmieder's law from the geometric calibration model of the camera. Negru and Nedevschi [13] estimated the scene visibility by detecting the positions of the inflection point and the horizontal line in the image. These methods can estimate the scene visibility without ground truth data from sensors. Nevertheless, the precision of geometric calibration is critical to these approaches.

The development of image recognition and deep learning technology has led to their gradual application in atmospheric visibility estimation. Deep-learning-based methods are superior for complex interaction processing compared to the above two methods. You et al. [14] combined CNNs (convolutional neural networks) and RNNs (recurrent neural networks) to estimate the relative atmospheric visibility from images, which significantly improved the estimation accuracy. However, the evaluation capacity of this model was only 300–800 m. Palvanov and Cho [15] proposed a deeply integrated convolutional neural network with three streams for visibility estimation. And this approach considered different image features as inputs to the model. Unlike the previous method using only visible light images, Wang et al. [16] used visible-infrared image pairs as the input and proposed a multimodal deep fusion model to learn the joint features from the input. Although the performance of the deep-learning-based methods was superior to the two methods mentioned above, the following problems still exist: (1) Most existing methods can not sufficiently extract fog features, which can provide valuable information for subsequent model learning; and (2) most existing methods are inadequate for extracting significant high-level multi-scale features for visibility estimation. Therefore, we propose a novel multi-scale fusion network for visibility estimation (Vis-MFN) from a single image. The main contributions are as follows:

- We propose a CNN-based method for highway visibility estimation from a single surveillance image. This method can provide low-cost and efficient support for intelligent highway management.
- (2) A multi-scale fusion network model is developed to estimate visibility from the input highway surveillance image. We are more concerned with the efficient transfer of low-level features to high-level features than with the design of complex network structures. Multiple image feature extraction methods are utilized to extract lowlevel visual features of fog, which can provide valuable information for subsequent model learning. The multi-scale fusion module is designed to extract the important high-level multi-scale features for the final visibility estimation, which can effectively improve the accuracy of the estimation.
- (3) We create a dataset of real-world highway surveillance images for model learning and performance evaluation. Each image in the dataset was labeled by professional traffic meteorology practitioners.

2. Proposed Method

We designed a deep multi-scale fusion network for highway visibility estimation. As we all know, extracting visibility features from a single image is difficult. To better understand the scene in the image, we extracted different types of features from the input image to provide critical information for the subsequent model learning. The multi-scale fusion module was further designed to jointly learn visibility from these image features of the same image. The overall architecture of the proposed network is shown in Figure 1. Multiple algorithms were adopted to process the input image to obtain detailed structural, spectral, and scene depth features. The multi-scale fusion module was further designed to fuse these features for adaptive visibility estimation.

2.1. Image Feature Extraction

2.1.1. Detailed Structural Feature Extraction

This branch is designed to extract the detailed structural features. The presence of fog can obscure detailed structural features in the image. Therefore, the richness of such features can reflect the visibility to some extent. We adopt the fast guided filter in [17] to decompose the image into a base layer and a detail layer. The detail layer is the specific structural feature that we desire. The guided filter is a technique for edge-aware image filtering. We choose the fast guided filter because of its superior visual output, fast processing speed, and ease of implementation. The guidance image, filtering input image, and filtering output image are represented as *I*, *p*, and *q*, respectively. The guided filter can be denoted as follows:

$$q_i = a_k I_i + b_k, \ \forall i \in \omega_k \tag{1}$$

$$a_{k} = \frac{\frac{1}{|\omega|} \sum_{i \in \omega_{k}} I_{i} p_{i} - \mu_{k} \overline{p}_{k}}{\sigma_{k}^{2} + \varepsilon}$$
(2)

$$b_k = \overline{p}_k - a_k \mu_k \tag{3}$$

where *i* is the index of a pixel and *k* is the index of a local square window ω with a radius *r*. μ_k , and σ_k are the mean and variance of *I* in the window *k*, and ε is a regularization parameter controlling the degree of smoothness. The filtering output is represented as follows:

$$q_i = \overline{a}_i I_i + b_i \tag{4}$$

where \bar{a}_i and b_i are the average of a and b, respectively, on the window ω_i at i. As shown in Figure 2, the detail layer mainly contains the detailed structural features, such as lane lines. Therefore, the fast guided filter is suitable for extracting detailed structural features of highway scene images.



Figure 1. The network structure of Vis-MFN.



Original image

Base layer

Detail layer

Figure 2. The original image, base layer, and detail layer. The detail layer is the detailed structural feature which we want.

2.1.2. Spectral Feature Extraction

This branch was designed to extract the areas covered by fog. Images captured in dense fog generally have low contrast, resulting in indistinct visuals. Because of the low contrast, fog images closely resemble grayscale images. To improve the visual characteristics of fog, we used a spectral filter, as the human visual system is more responsive to colors than grayscale. To facilitate the extraction of the spectral features of fog, we converted the images from RGB color space to LAB color space, since it covers the full range of human color perception. As shown in Figure 3, the regions of fog in the filtered image are marked in pink. The filtered image presents fog regions more prominently than the original image. As a result, the following CNN-based model can easily identify the fog regions and focus on extracting local features.



Original image

Filtered image

Figure 3. The original image and the filtered image.

2.1.3. Scene Depth Feature Extraction

This branch is designed to extract the scene depth feature. The scene depth provides information about the depth of objects in the image, which is useful for visibility estimation. We utilized MiniNet [18] to extract the scene depth feature, as it is a lightweight and efficient network for unsupervised monocular depth prediction. As shown in Figure 4, the black area in the depth image is basically consistent with the fog area in the original image.



Original image

Depth image

Figure 4. Depth feature extraction.

2.2. Multi-Scale Fusion Module

The multi-scale fusion module is designed to adaptively fuse the detailed structural, spectral, and scene depth features. It consists of three parts, including a shallow feature representation block (SFRB), multiple multi-scale fusion blocks (MSFBs), and a global feature fusion block (GFB). Specifically, two 3×3 convolutional layers extracted shallow features from the structural, spectral, and scene depth features. This process is defined as follows:

$$S = f_s([X_1, X_2, X_3, X_4])$$
(5)

where f_s denotes the SFRB function; X_1, X_2, X_3 represent the structural, spectral, scene depth features, respectively; X_4 represents the original image; and [·] represents the concatenation operation. Then, several multi-scale fusion blocks (MSFBs) were designed to extract multi-scale features. This process can be defined as follows:

$$M_k = f_m(M_{k-1}), \ k = 1, \dots, n$$
 (6)

where f_m denotes the *m*-th MSFB function and M_k and M_{k-1} represent the input and output of the *k*-th MSFB, respectively. We fused these multi-scale features using the concatenation operation and further extracted high-level features through the convolution operation. This procedure can be formulated as follows:

$$G = f_g([M_1, M_2, \dots M_n]) \tag{7}$$

where $[M_1, M_2, ..., M_n]$ denotes the concatenation of the feature maps produced in each MSFN and f_g represents the following convolution operation.

The architecture of the multi-scale fusion block is shown in Figure 5. We constructed a three-bypass network using different convolutional kernels in each bypass for multi-scale feature extraction. We used multiple 3×3 dilated convolutions with varying dilation factors instead of convolutional kernels of various sizes to extract multi-scale features, which enlarged the receptive field and maintained the filter size [19]. We concatenated the features extracted by the dilated convolutions for multi-scale feature fusion and then used a 1×1 convolutional layer to reduce the dimension. Then, we concatenated the outputs of two 1×1 convolutional layers and used a 1×1 convolutional layer to reduce the dimension. This part can be denoted as follows:

$$C_1 = f_{c1}([f_{d1}(M_{n-1}), f_{d2}(M_{n-1}), f_{d3}(M_{n-1})])$$
(8)

$$C_2 = f_{c1}([f_{d1}(C_1), f_{d2}(C_1), f_{d3}(C_1)])$$
(9)

$$C_3 = f_{c1}([C_1, C_2]) \tag{10}$$

where f_{di} represents the dilated convolution with dilation factor *i*, and f_{c1} represents 1×1 convolution. Finally, the residual structure was adopted to increase information flow, which helped to reduce the computational complexity and improved the performance of the model. The output of *k*-th MSFB can be expressed as follows:

 $M_n = M_{n-1} + C_3$

Figure 5. The architecture of the multi-scale fusion block.

3. Experiments

3.1. Dataset

To our knowledge, there is no publicly available dataset of actual highway fog scenes. For this reason, we created a real-scene dataset for highway visibility estimation. Specifically, we selected about 30 cameras located near traffic weather stations in the Jiangsu section of the Beijing-Shanghai Expressway. These cameras were positioned more than

(11)

10 km from each other to guarantee diverse shooting scenarios. Then, we collected more than 18,000 surveillance images of real highway scenes obtained from these cameras. These images were collected in different time periods, further ensuring the diversity of the dataset. After similarity elimination and quality control, 15,000 images were selected as the training dataset and 3000 images were chosen as the test dataset. Meteorology professionals classified the fog intensity into five levels according to visibility. The detailed visibility level standard is shown in Table 1. All dataset images were automatically labeled based on observation data from traffic weather stations and subsequently adjusted by skilled traffic meteorology practitioners. Some sample images from the dataset are presented in Figure 6. The dataset is currently not accessible for public use due to security concerns regarding road data.

Table 1. Visibility level standard.

Visibility Level	1	2	3	4	5
Visibility distance	0–50 m	50–100 m	100–200 m	200–500 m	500+ m



Figure 6. Some sample images in our dataset.

3.2. Implementation and Training Details

We used Adam [20] for optimization. The momentum and weight decay were set to 0.9 and 0.0001, respectively. The learning rate was initialized to 0.0001 and decreased by a factor of 10 every 40 epochs. We implemented our models with Pytorch1.5.1 on a single GPU of NVIDIA RTX 3090. Cross-entropy was adopted as the loss function.

3.3. Comparison Experiments

We compared the proposed method with several deep-learning-based methods, including two image classification methods (AlexNet [21] and VGG16 [22]) and two atmospheric visibility estimation methods (relative CNN-RNN [14] and STCN-Net [23]). We re-trained these three deep-learning-based methods on our dataset, where the parameters were set according to the recommendations in the paper.

Table 2 shows the accuracy of the multiple methods tested on the test dataset. The experimental results indicate that the AlexNet and VGG16 methods showed poor performance, primarily because the AlexNet and VGG16 models are designed for natural image classification and do not take into account the characteristics unique to road scene images. Consequently, extracting visibility features effectively is quite challenging. The relative CNN-RNN method outperformed AlexNet and VGG16 because the CNN-RNN module was able to capture the global view while approximating human attention shift, which enabled it to learn more effective visual features compared to AlexNet and VGG16. However, the relative CNN-RNN estimated visibility from only the original image, so it was challenging to extract multi-scale features. The STCN-Net achieved a better performance than the relative CNN-RNN method, since this method designed a novel 3D multi-feature stream matrix, which provided rich low-level features. However, the STCN-Net performed slightly less well in terms of visibility below 500 m. The proposed Vis-MFN achieved the best estimation accuracy. Figure 7 shows the confusion matrices of multiple methods on the test dataset. It can be seen that the classification labels of the proposed method are mainly clustered along the main diagonal of the matrix, demonstrating the stability of the classification performance.

Figure 8 shows the estimated results of different methods on test images with different visibility levels. The proposed method was able to make correct estimates on sample images with different visibility levels, since the proposed method had two main advantages. The first advantage was that three image feature extraction algorithms could adequately extract visibility-related features from the images. The second advantage was that the multi-scale fusion module could adaptively extract useful features for visibility estimation.

3.4. Ablation Experiments

We further conducted ablation experiments to verify the effectiveness of the image feature extraction algorithms and the multi-scale fusion module.

- (1) Vis-MFN-NF: No image feature extraction algorithm was used in the model.
- (2) Vis-MFN-NM: The multi-scale fusion blocks were replaced by multiple convolutions in series. Meanwhile, the receptive field of the new network remained unchanged.
- (3) Vis-MFN-M2: Only two multi-scale blocks were used in the network.
- (4) Vis-MFN-M4: Four multi-scale blocks were used in the network.

Table 3 shows the estimation accuracy of the proposed method and two ablation methods. It can be seen that both the multi-scale fusion module and the image feature extraction methods significantly improved the estimation accuracy. Although the performance of Vis-MFN-M4 was better than that of Vis-MFN-M2, the complexity of Vis-MFN-M4 was significantly increased. Therefore, we chose the Vis-MFN-M2 for operational application.

Table 2. The estimation accuracy of multiple methods on the test dataset. The best result is marked in bold.

	AlexNet	VGG16	Relative CNN-RNN	STCN-Net	Vis-MFN
Accuracy	69.21%	68.72%	78.58%	81.10%	81.76%

Table 3. The results of the ablation experiment. The best result is marked in bold.

Setting	Vis-MFN-NM	Vis-MFN-NF	Vis-MFN-M2	Vis-MFN-M4
Image feature extraction methods	1	×	✓	✓
Multi-scale fusion module	×	✓	\checkmark	\checkmark
The number of MSFBs	2	2	2	4
Accuracy	75.36%	72.85%	81.76%	82.49%





Figure 7. The confusion matrices of multiple methods on the test dataset.



Figure 8. Visualization of the estimated results of different methods on images with different visibility levels. The estimated results and ground truth are marked on the image.

4. Conclusions

Highway visibility estimation in foggy weather is important for traffic safety and intelligent highway management. Traditional methods utilize visibility sensors to directly measure visibility. However, the meteorological observation network is not dense enough to monitor agglomerate fog. Additionally, visibility sensors are costly and inaccurate in nonuniform atmospheric conditions. Although deep learning methods have been applied by many scholars for visibility estimation, the following problems still exist: (1) Most existing methods cannot sufficiently extract fog features, which can provide valuable information for subsequent model learning. (2) Most existing methods are inadequate in extracting significant, high-level, multi-scale features for visibility estimation. To close these research gaps, we propose a multi-scale fusion network for estimating highway visibility in foggy weather. The network can estimate the visibility level from a single image of a highway scene. To achieve this, we used multiple image feature extraction methods to extract the structural, spectral, and scene depth features from the image. Then, a multi-scale fusion module was designed to extract and fuse relevant features to estimate visibility adaptively. The experiments demonstrate the effectiveness of the proposed method. Our future work aims to improve the performance of our method and propose a benchmark for highway visibility estimation.

Author Contributions: Conceptualization, L.H.; methodology, P.X.; software, P.X. and J.S.; validation, Z.Z. and X.Y.; formal analysis, X.Z.; investigation, X.L.; resources, P.X.; data curation, P.X. and X.L.; writing—original draft preparation, P.X.; writing—review and editing, X.L.; visualization, X.L.; supervision, X.L.; project administration, L.H.; funding acquisition, L.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the China Meteorological Administration Innovation and Development Project (CXFZ2022J069) and the China Meteorological Administration Joint Research Project on Meteorological Capacity Enhancement (22NLTSQ006).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The corresponding author can provide data supporting the findings of this study upon request. Due to ethical and privacy concerns, the data are not publicly available.

Acknowledgments: We are very grateful to Jiangsu Transportation Holdings Co., Ltd. (Nanjing, China) for providing us with much help in data collection and processing in the research.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Fabbian, D.; Dear, R.D.; Lellyett, S. Application of artificial neural network forecasts to predict fog at Canberra International Airport. *Weather Forecast.* 2007, *22*, 372–381. [CrossRef]
- Kwon, T.M. Atmospheric Visibility Measurements Using Video Cameras: Relative Visibility. Retrieved from the University of Minnesota Digital Conservancy. 2004. Available online: https://hdl.handle.net/11299/1027 (accessed on 1 July 2004).
- Liaw, J.J.; Lian, S.B.; Huang, Y.F.; Chen, R.C. Atmospheric visibility monitoring using digital image analysis techniques. In Proceedings of the International Conference on Computer Analysis of Images and Patterns, Münster, Germany, 2–4 September 2009; pp. 1204–1211.
- 4. Luo, C.H.; Wen, C.Y.; Yuan, C.S.; Liaw, J.J.; Lo, C.C.; Chiu, S.H. Investigation of urban atmospheric visibility by high-frequency extraction: Model development and field test. *Atmos. Environ.* **2005**, *39*, 2545–2552. [CrossRef]
- Babari, R.; Hautière, N.; Dumont, E.; Papelard, J.P.; Paparoditis, N. Computer vision for the remote sensing of atmospheric visibility. In Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV Workshops), Barcelona, Spain, 6–13 November 2011; pp. 219–226.
- Varjo, S.; Hannuksela, J. Image based visibility estimation during day and night. In Proceedings of the Asian Conference on Computer Vision, Singapore, 1–5 November 2014; pp. 277–289.
- 7. Lee, Z.; Shang, S. Visibility: How applicable is the century-old Koschmieder model? J. Atmos. Sci. 2016, 73, 4573–4581. [CrossRef]
- Hautiére, N.; Babari, R.; Dumont, E.; Brémond, R.; Paparoditis, N. Estimating meteorological visibility using cameras: A probabilistic model-driven approach. In Proceedings of the Asian Conference on Computer Vision, Queenstown, New Zealand, 8–12 November 2010; pp. 243–254.
- Gallen, R.; Cord, A.; Hautière, N.; Aubert, D. Towards night fog detection through use of in-vehicle multipurpose cameras. In Proceedings of the IEEE Intelligent Vehicles Symposium, Baden-Baden, Germany, 5–9 June 2011; pp. 399–404.
- 10. Babari, R.; Hautiere, N.; Dumont, E.; Brémond, R.; Paparoditis, N. A model-driven approach to estimate atmospheric visibility with ordinary cameras. *Atmos. Environ.* **2011**, *45*, 5316–5324. [CrossRef]
- 11. Hautiere, N.; Tarel, J.P.; Lavenant, J.; Aubert, D. Automatic fog detection and estimation of visibility distance through use of an onboard camera. *Mach. Vis. Appl.* **2006**, *17*, 8–20. [CrossRef]
- 12. Hautière, N.; Labayrade, R.; Aubert, D. Estimation of the visibility distance by stereovision: A generic approach. *IEICE Trans. Inf. Syst.* **2006**, *89*, 2084–2091. [CrossRef]
- Negru, M.; Nedevschi, S. Image based fog detection and visibility estimation for driving assistance systems. In Proceedings of the IEEE 9th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 5–7 September 2013; pp. 163–168.
- 14. You, Y.; Lu, C.; Wang, W.; Tang, C.K. Relative CNN-RNN: Learning relative atmospheric visibility from images. *IEEE Trans. Image Process.* **2018**, *28*, 45–55. [CrossRef] [PubMed]
- 15. Palvanov, A.; Cho, Y.I. VisNet: Deep Convolutional Neural Networks for Forecasting Atmospheric Visibility. *Sensors* 2019, 19, 1343. [CrossRef] [PubMed]
- Wang, H.; Shen, K.; Yu, P.; Shi, Q.; Ko, H. Multimodal Deep Fusion Network for Visibility Assessment With a Small Training Dataset. *IEEE Access* 2020, *8*, 217057–217067. [CrossRef]
- 17. He, K.; Sun, J. Fast guided filter. *arXiv* 2015, arXiv:1505.00996.
- 18. Liu, J.; Li, Q.; Cao, R.; Tang, W.; Qiu, G. MiniNet: An extremely lightweight convolutional neural network for real-time unsupervised monocular depth estimation. *ISPRS J. Photogramm. Remote Sens.* **2020**, *166*, 255–267. [CrossRef]
- 19. Yu, F.; Koltun, V. Multi-scale context aggregation by dilated convolutions. *arXiv* **2015**, arXiv:1511.07122.
- 20. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* 2014, arXiv:1412.6980.
- 21. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* 2017, 60, 84–90. [CrossRef]
- 22. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv 2014, arXiv:1409.1556.
- 23. Liu, J.; Chang, X.; Li, Y.; Li, Y.; Ji, Y.; Fu, J.; Zhong, J. STCN-Net: A novel multi-feature stream fusion visibility estimation approach. *IEEE Access* 2022, *10*, 120329–120342. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.