

Simple Global Thresholding Neural Network for Shadow Detection

Guiyuan Li,^{1,2} Changfu Zong,¹ Dong Zhang,^{3*} Tianjun Zhu,⁴ and Jianying Li⁴

¹State Key Laboratory of Automotive Simulation and Control, Jilin University, Changchun 130025, P. R. China

²School of Automobile and Traffic Engineering, Liaoning University of Technology, Jinzhou 121001, P. R. China

³School of Electrical and Electronic Engineering, Nanyang Technological University, 639798 Singapore

⁴Department of Mechanical and Automotive Engineering, Zhaoqing University, Zhaoqing 526061, P. R. China

(Received March 24, 2021; accepted June 16, 2021)

Keywords: shadow detection, global threshold, neural network, binary fusion

Shadow detection based on vision sensors is widely used in image processing. Because of the variability of illumination and projection surface color, shadow detection based on a color image is a challenging problem. Aiming at solving the conflict between the complexity and robustness of current shadow detection algorithms, we established a new shadow detection network by combining the global thresholding method with a neural network, which realized the decoupling of the global threshold and binary fusion. Three public shadow detection datasets, large-scale shadow dataset of Stony Brook University (SBU), large-scale dataset with image shadow triplets (ISTD), and shadow detection for mobile robots features evaluation and datasets (SDMR), were utilized for its verification. Experimental results show that the performance of the proposed network approaches that of previous deep learning methods, both visually and in terms of objective indicators, but the proposed network has the advantages of a simple structure and good robustness.

1. Introduction

Vision sensors can provide high-resolution color information, which can accurately reflect the details of complex changes in light; therefore, shadow detection based on vision sensors is necessary in some image-processing-based applications, such as medical imaging diagnosis, remote sensing monitors, video surveillance, automatic driving, and other fields in which shadow detection is applied to identify targets or eliminate disturbances. However, owing to the diversity of lighting and projection surface colors, shadow detection is a difficult problem. Many methods have been proposed to improve the robustness of shadow detection algorithms.^(1–3) In this paper, existing methods are categorized into two major groups: methods based on empirical models and data-driven methods.

The methods based on empirical models can be summarized into two types: those based on experience and those based on a model. In the methods based on empirical models, differences in chromaticity or intensity between shadow and non-shadow in an image are enhanced to highlight the shadow. For instance, Tsai's research showed that pixels slightly change in hue but

*Corresponding author: e-mail: 2609582477@qq.com

<https://doi.org/10.18494/SAM.2021.3398>

greatly in intensity when shadow is generated; a ratio map that showed the ratio of intensity-equivalent components and hue-equivalent components was obtained, and the threshold of the ratio map was determined through the Otsu method to detect shadows.⁽⁴⁾ Murali and Govindan found that shadow pixels have lower intensity but more blue than non-shadow pixels. They used a double threshold method in the LAB color space for shadow detection, where the intensity threshold was calculated as the mean of the L channel and the hue threshold was calculated as the mean of the A and B channels.⁽⁵⁾ On the basis of the characteristic that the intensity of shadow pixels is lower than that of non-shadow pixels, Wehrwein *et al.* proposed an adjacent ratio map that showed the ratios of adjacent pixels of a gray image and used an empirical threshold for shadow detection.⁽⁶⁾ Jung *et al.* obtained a higher-order residual map in a log domain and detected shadows using an empirical threshold.⁽⁷⁾ Newey *et al.* found that texture features are invariant to illumination, and they adopted a local binary pattern feature and the Otsu method to detect shadows on the basis of image segmentation.⁽⁸⁾ In the model-based methods employing Retinex theory, the independence between chromaticity and intensity in an image is used to obtain invariant images, which are used to detect the edges of shadows. For example, Finlayson *et al.* proposed an invariant image through projecting 2D log chromaticity pixels in the direction (obtained by entropy minimization) orthogonal to the vector of luminance based on the assumptions of Lambertian shading and Planckian lighting.⁽⁹⁾ Nayar *et al.* obtained a light-independent image using a sequential labeling algorithm based on neighboring points on a smoothly curved surface with similar surface normal and illumination conditions.⁽¹⁰⁾ Tian and coworkers proposed a tricolor attenuation model and obtained a shadow-invariant red-blue difference image based on Planck's blackbody irradiance law.^(11,12)

Methods based on a data-driven approach can be divided into traditional machine learning and deep learning methods. Traditional machine learning methods usually need to extract histogram features on the basis of segmentation images and adopt the framework of cascade classifiers. For example, Zhu *et al.* adopted boosted decision trees integrated with a binary conditional random field as a classifier.⁽¹³⁾ Guo *et al.* proposed the detection framework of two cascaded support vector machines (SVMs).⁽¹⁴⁾ Hosseinzadeh *et al.* obtained the detection framework of an SVM combining two convolutional neural networks (CNNs).⁽¹⁵⁾ Deep learning methods with CNNs can be trained in an end-to-end way, which reduces the complexity of the staged training used in traditional machine learning methods. For instance, Dong *et al.* proposed a shadow detection network by combining the pre-trained deep network ResNeXt101 and an attention module⁽¹⁶⁾ with two NVIDIA GTX1080Ti cards used for training. The training took about 3.5 h. One GTX1080Ti card was used for testing, and the detection speed was 12 frames/s. Deep learning methods have achieved good detection results, but they are also generally computationally expensive.^(17–22)

It can be seen from the above discussion that empirical model methods have the advantages of simple calculation and high speed but low robustness. Traditional machine learning methods require hand-crafted features owing to the effects of various factors such as the object material, microsurface geometry, light field distribution, and sensor noise, making it very difficult to design shadow features with strong invariance. Deep learning methods use optimization algorithms to obtain shadow features from a dataset without manual feature design; however,

they require a large representative dataset, and the hyperparameter design is also a considerable challenge. With the purpose of finding a method with high detection accuracy and robustness similar to that of deep learning methods but with a simple structure and high speed, in this paper, a simple global threshold neural network (GTNN) for shadow detection based on a BP neural network is proposed. The key contributions of our work are outlined below:

- (1) A new network that combines the global threshold method with a neural network was proposed, without any assumptions about the lighting, surface, or camera.
- (2) Compared with other machine learning methods, this method has a simple structure and does not require hand-crafted features or a complex CNN.
- (3) The method has low dependence on training data and strong generalization capability, and is robust to different datasets once it is trained with a very small amount of training data from a single dataset.

The rest of this paper is organized as follows. In Sect. 2, we introduce the design concepts of the GTNN and its framework and detection and training processes. In Sect. 3, we conduct a comparative analysis of the experimental results. Conclusions are given in Sect. 4.

2. Proposed Method

Global threshold binarization for image segmentation is a widely used method in shadow detection. For a shadow image, when the difference in hue between shadow and non-shadow is significant, the shadow detection result based on the global threshold of hue will be better than that based on the threshold of intensity or saturation. However, owing to the diversity of light and projection surface colors, it is difficult to obtain good detection results by using any of the above thresholds separately. Fortunately, when we slightly adjusted the above thresholds and then fused the binary images obtained separately, we found that the detection results had been effectively improved. For example, in Fig. 1, the first column is the input image, the second to fourth columns are the binary images obtained through the thresholds of hue, saturation, and intensity histograms, respectively (white represents shadows and numbers in brackets represent thresholds), the fifth column of the first row is the result of the logic AND of the results of hue and intensity after the threshold adjustment, and the fifth column of the second row is the result

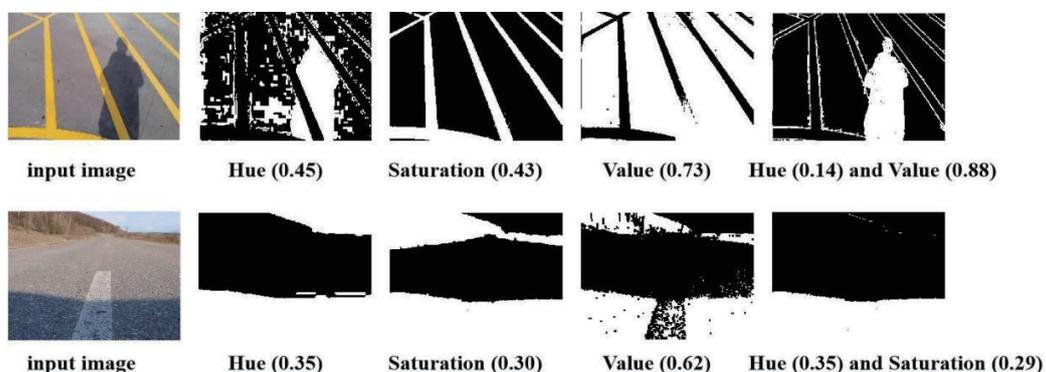


Fig. 1. (Color online) Detection results based on the global threshold.

of the logic AND of the results of hue and saturation. It can be seen that the results after fusion are better than those before fusion. If the above detection process is abstracted as a function that takes the global thresholds as the input and the binary images as the output, it is hypothesized that there is a strong coupling relationship between the thresholds; if threshold decoupling and binary image fusion are carried out by appropriate methods, better detection methods will be realized. On this basis, a simple GTNN for shadow detection is proposed in this paper.

2.1 Proposed shadow detection network

The overall architecture of the proposed GTNN for shadow detection is shown in Fig. 2. It mainly includes three parts: a front-end, a network part, and a back-end. The front-end is used to perform filtering on the feature maps to eliminate noise. Here, the feature map can be image data of different color spaces such as the image of each channel of the HSV color space, or it can be an image on the transform domain such as a ratio map, log ratio map, or segmentation image. The global threshold module is used to obtain the global threshold of different feature maps; the Otsu, Kittler, and entropy methods can be used in this module. The neutral network is used to decouple the global thresholds of different features; it is a fully connected neural network whose inputs are the global thresholds of the feature maps and whose outputs are the decoupled global thresholds. The BP neural network, RBF neural network, or another structure can be chosen as the neural network. The back-end is used to perform binarization through the decoupled global threshold, fuse the binary images, and perform morphological filtering for smoothing. Fusion methods can adopt various logical operations or their linear combinations.

2.2 Composition and training

The main modules of the proposed GTNN are shown in Table 1. Among them, three channels in HSV color space are used as feature maps, which are the input of the filter module, and the filter module consists of a maximum filter and a Gaussian filter in series. The Otsu method is adopted in the global threshold module. The network is composed of a BP neural network with three hidden layers, whose structure is shown in Fig. 3. In the back-end, the aggregation method is the logic AND operation, and the morphological filter module consists of morphological closing and opening operators in series.

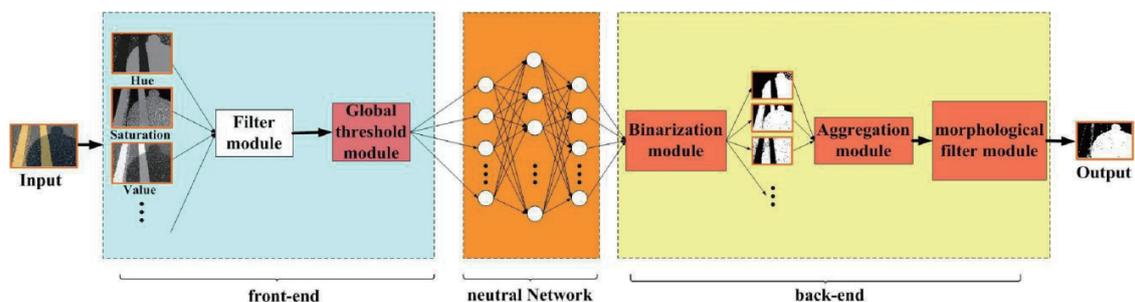


Fig. 2. (Color online) Architecture of GTNN.

Table 1
GTNN composition.

Part	Functional unit
Filter	Max filter: size = 3×3
	Gauss filter: size = 3×3
	Standard deviation = 1
Global threshold	Otsu method
Network	BP neural network
	Number of nodes of hidden layers: 100, 40, 30 Activation function of every hidden layer: linear
Aggregation	Logic AND operation
Morphological filter	Opening operator: threshold = 400, 8-connected neighborhood Closing operator: structuring element = $[1 \ 0 \ 1; 0 \ 1 \ 0; 1 \ 0 \ 1]$

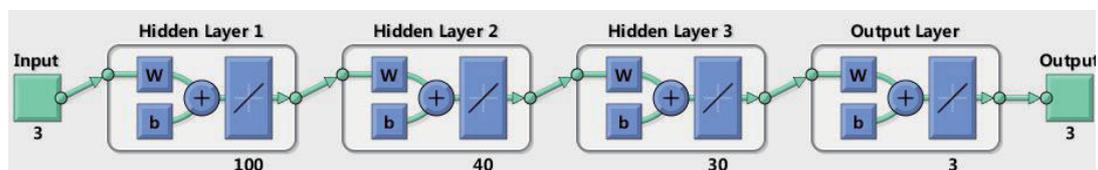


Fig. 3. (Color online) BP neural network structure.

To train the BP neural network, 100 shadow images were selected from the SBU training dataset.⁽²³⁾ These 100 images were mainly of outdoor environments and included changes in projection surface color. The global thresholds were obtained from the 100 images through the Otsu method in HSV color space and recorded as the input training data of the BP neural network. The thresholds of the H, S, and V channels were manually adjusted and inputted to the back-end for calculation until a satisfactory result was obtained, then the thresholds were recorded as the target values of training. The training data are shown in Fig. 4. The test data included 1000 images selected from the SBU dataset, 300 images selected from the ISTD dataset, and 400 images selected from the SDMR dataset.^(8,18) We adopted the mean squared error (MSE) as the loss function and the Levenberg–Marquardt algorithm as the training algorithm. The training data were split into training, validation, and test data at a ratio of 0.4:0.3:0.3.

The experimental hardware platform was an Intel (R) Core (TM) i7-6700K CPU (4.0 GHz, 16 GB RAM) and the experimental software platform was the Windows 7 operating system with MATLAB R2015a. The MSE curve used for the BP neural network training is shown in Fig. 5. The training time was about 5.58 s and the detection time of an image (480×640 pixels) was about 0.167 s.

3. Results and Discussion

3.1 Evaluation metrics

To evaluate the proposed method in terms of detection accuracy, we used two evaluation metrics as follows:

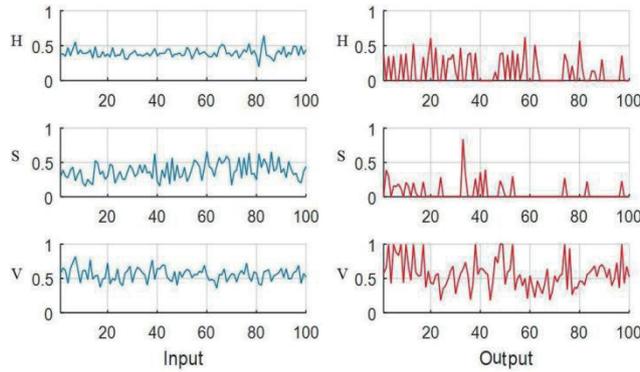


Fig. 4. (Color online) Global threshold data curve used for training.

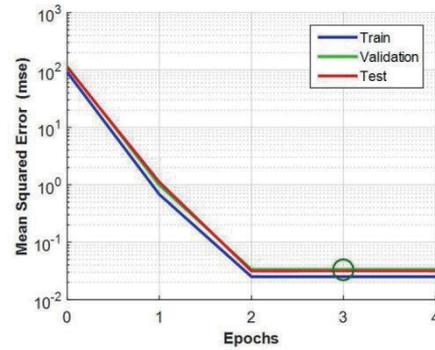


Fig. 5. (Color online) MSE curve of training.

$$Acc = \frac{TP + TN}{Np + Nn}, \quad (1)$$

$$BER = 1 - \frac{1}{2} \left(\frac{TP}{Np} + \frac{TN}{Nn} \right). \quad (2)$$

Here, TP , TN , Np , and Nn are the number of correctly detected shadow pixels, the number of correctly detected non-shadow pixels, the total number of shadow pixels, and the total number of non-shadow pixels, respectively.

3.2 Performance comparisons and analysis

We compared our method with four deep-learning-based methods, patched-CNN,⁽¹⁵⁾ Attention Res-Unet,⁽¹⁶⁾ scGAN,⁽¹⁷⁾ and stacked-CNN,⁽¹⁸⁾ as shown in Fig. 6. Our method achieves similar performance to Attention Res-Unet. Compared with scGAN, patched-GAN, and stacked-CNN, our method is more precise in detecting shadows cast on color areas such as the line mark. We provide further visual results tested on the SBU, ISTD, and SDMR datasets in Fig. 7, which show various challenging cases, such as changes in surface color and texture in the scene. The proposed method can accurately locate shadow areas, indicating that it has high robustness. Table 2 shows a comparison of quantitative results and Table 3 shows quantitative results of the test on the ISTD and SDMR datasets. As can be seen from Tables 2 and 3, although our shadow detection network is trained on 100 images from the SBU training set, it still obtained an accuracy close to that of the deep learning methods and also obtained higher accuracy on the ISTD and SDMR datasets, which demonstrates the generalization capability of our method.

The results show that the above-mentioned hypothesis of threshold coupling and binary fusion is reasonable, and the decoupling of the threshold by the BP neural network was successful. However, we can also see that the BER of our method is significantly higher than those of the other methods. Through the analysis of results, we found that there are three cases

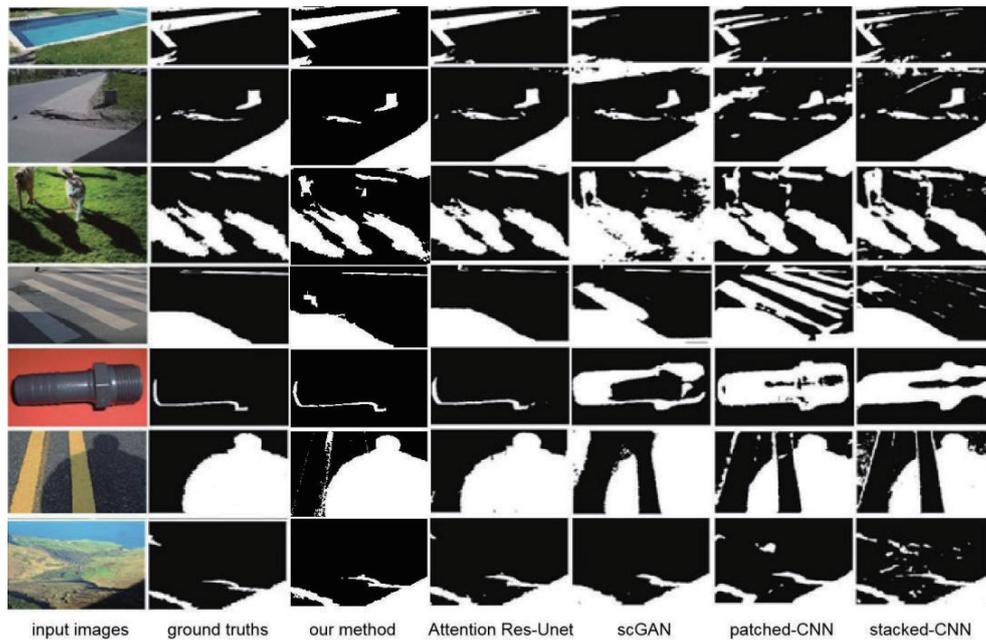


Fig. 6. (Color online) Visual comparison of test results obtained by our method and other methods.

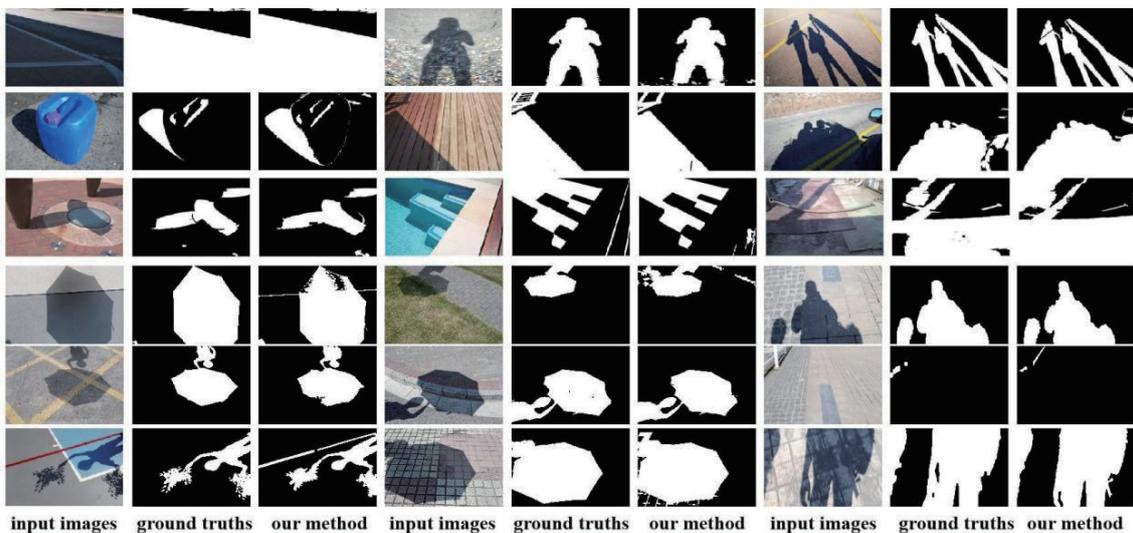


Fig. 7. (Color online) Qualitative results of our method tested on SBU, ISTD, and SDMR datasets.

Table 2
Comparison of quantitative results.

Methods	SBU	
	Acc	BER
GTNN (ours)	0.899	27.12
Attention Res-Unet	—	4.88
scGAN	0.90	9.10
patched-CNN	0.88	11.56
stacked-CNN	0.88	11.00

Table 3
Quantitative shadow detection results.

GTNN (ours)	Acc	BER
ISTD	0.925	16.13
SDMR	0.955	10.50

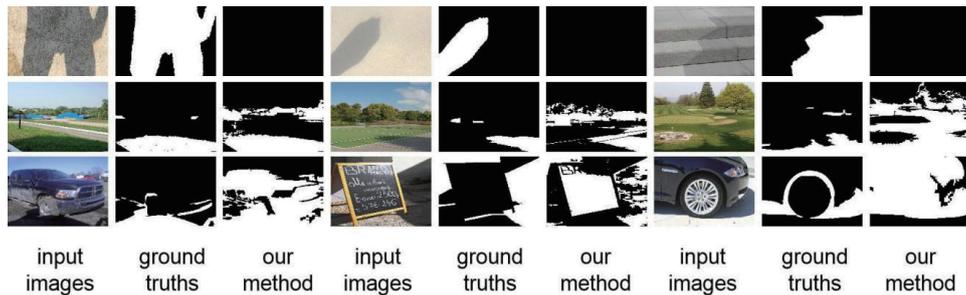


Fig. 8. (Color online) Three types of environment with failed detection.

where the detection is poor. The first case is shown in the first row of Fig. 8. In this case, the image environment is not complicated, but the shadow is relatively light; for this case, the shadow is incorrectly detected by our method. The second case is shown in the second row of Fig. 8. In this case, the image environment is dominated by lawns and woods. The third case is shown in the third row of Fig. 8. In this case, dark colors make up a large proportion of the images. For the latter two cases, false detection occurs with our method.

The result of these failures leads to an increase in BER. We believe that there are three main reasons for this result:

- (1) Small numbers and few types of training samples. In the 100 training images we selected from the SBU dataset, we mainly focused on images with obvious color changes on the projection surface, which were lacking in the above situations.
- (2) Lack of feature map types and the existence of threshold adjustment errors. We only used three feature maps as the input of the GTNN. In addition, the target training data for the BP network was obtained by manual adjustment, so the error of the target training data was relatively large.
- (3) The fusion rule is relatively simple. We only chose the logic AND operation to perform binary fusion, and the priority of fusion was not considered when the binary image fusion of any two of hue, saturation, and intensity can obtain a similar detection result.

4. Conclusions

In this paper, we proposed a new GTNN for shadow detection that combines the global threshold method with the BP neural network. Through this method, different global features with strong illumination and texture invariance are effectively utilized. Through the neural network framework proposed in this paper, threshold decoupling and binary fusion are realized. Compared with the mainstream CNN method, the network has a simple structure, fast calculation, and high robustness. The main problem at present is that the detected BER is high. By increasing the numbers of types of feature map and training sample and improving the priority of fusion, there is still plenty of scope to improve the performance. In contrast with the end-to-end CNN method, our proposed method requires the manual acquisition of thresholds to train the BP neural network, which increases the artificial errors and workload. In future research, to achieve end-to-end training, the error of the BP mechanism in the fusion layer will be further studied.

Acknowledgments

This work is supported by the Scientific Study Project for Institutes of Higher Learning of Liaoning Provincial Department of Education (JP2016018), Characteristic Innovation Project of Guangdong Provincial Department of Education (2019KTSCX201), Zhaoqing Research and Development Technology and Application of Energy Conservation and Environmental Protection Ecological Governance (2020SN004), Teaching Quality and Reform of Higher Vocational Education Project of Guangdong Province (GDJG2019463), 2021 Special Projects in Key Fields of Colleges and Universities of Guangdong Province (2021ZDZX1061), and Youth Innovative Talents Project of Guangdong Provincial Department of Education (2018KQNCX290).

References

- 1 K. R. M. Adeline, M. Chen, X. Briottet, S. K. Pang, and N. Paparoditis: *Isprs J. Photogram. Remote Sens.* **80** (2012) 21. <https://doi.org/10.1016/j.isprsjprs.2013.02.003>
- 2 A. Sanin, C. Sanderso, and B. C. Lovel: *Pattern Recognit.* **45** (2012) 1684. <https://doi.org/10.1016/j.patcog.2011.10.001>
- 3 J. Tian, Z. Wang, and Y. Tang: *Inf. Control* **44** (2015) 21. <https://doi.org/10.13976/j.cnki.xk.2015.0215>
- 4 V. J. D. Tsai: *IEEE Trans. Geosci. Remote Sens.* **44** (2006) 1661. <https://doi.org/10.1109/TGRS.2006.869980>
- 5 S. Murali and V. K. Govindan: *Cybern. Inf. Technol.* **13** (2013) 95. <https://doi.org/10.2478/cait-2013-0009>
- 6 S. Wehrwein, K. Bala, and N. Snavely: *Proc. 2015 Int. Conf. 3D Vision (3DV, 2015)* 580–587. <https://doi.org/10.1109/3DV.2015.58>
- 7 C. Jung, W. Kim, and C. Kim: *Opt. Lett.* **36** (2011) 4428. <https://doi.org/10.1364/OL.36.004428>
- 8 C. C. Newey, C. Charles, O. D. Jones, and H. M. Dee: *Spat. Cogn. Comput.* **18** (2017) 1. <https://doi.org/10.1080/13875868.2017.1322088>
- 9 G. D. Finlayson, D. Graham, M. S. Drew, and C. Lu: *Int. J. Comput. Vision* **85** (2009) 35. <https://doi.org/10.1007/s11263-009-0243-z>
- 10 S. K. Nayar and R. M. Bolle: *Int. J. Comput. Vision* **17** (1996) 219. <https://doi.org/10.1007/BF00128232>
- 11 J. Tian, J. Sun, and Y. Tang: *IEEE Trans. Image Process.* **18** (2009) 2355. <https://doi.org/10.1109/TIP.2009.2026682>
- 12 J. Tian, L. Zh, and Y. Tang: *EURASIP J. Adv. Sig. Proc.* **116** (2012) 1. <https://doi.org/10.1186/1687-6180-2012-116>
- 13 J. Zhu, K. G. G. Samuel, S. Z. Masood, and M. F. Tappen: *Proc. 2010 IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR, 2010)* 223–230. <https://doi.org/10.1109/CVPR.2010.5540209>
- 14 R. Guo, Q. Dai, and D. Hoiem: *Proc. 2011 IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR, 2011)* 2033–2040. <https://doi.org/10.1109/CVPR.2011.5995725>
- 15 S. Hosseinzadeh, M. Shaker, and H. Zhang: *Proc. 2018 IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS, 2018)* 3124–3129. <https://doi.org/10.1109/IROS.2018.8594050>
- 16 Y. Dong, H. Feng, Z. Xu, Y. Chen, and Q. Li: *J. Zhejiang Univ.* **53** (2019) 373. <https://doi.org/10.3785/j.issn.1008-973X.2019.02.021>
- 17 V. Nguyen, T. F. Y. Vicente, M. Zhao, M. Hoai, and D. Samaras: *Proc. 2017 IEEE Int. Conf. Computer Vision (ICCV, 2017)* 4520–4528. <https://doi.org/10.1109/ICCV.2017.483>
- 18 J. Wang, X. Li, and J. Yang: *Proc. 2018 IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR, 2018)* 1788–1797. <https://doi.org/10.1109/CVPR.2018.00192>
- 19 X. Hu, L. Zhu, C. Fu, J. Qin, and P. A. Heng: *Proc. 2018 IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR, 2018)* 7454–7462. <https://doi.org/10.1109/CVPR.2018.00778>
- 20 L. Zhu, Z. Deng, X. Hu, and C. Fu: *Proc. 2018 European Conf. Computer Vision (ECCV, 2018)* 1–16. https://doi.org/10.1007/978-3-030-01231-1_8
- 21 L. Qu, J. Tian, S. He, Y. Tan, and R. W. H. Lau: *Proc. 2017 IEEE Conf. Computer Vision and Pattern Recognition (CVPR, 2017)* 2308–2316. <https://doi.org/10.1109/CVPR.2017.248>
- 22 S. H. Khan, M. Bennamoun, F. Sohel, and R. Togneri: *Proc. 2014 IEEE Conf. Computer Vision and Pattern Recognition (CVPR, 2014)* 1939–1946. <https://doi.org/10.1109/CVPR.2014.249>
- 23 Y. Vicente, F. Tomas, L. Hou, C. P. Yu, M. Hoai, and D. Samaras: *Proc. 2016 European Conf. Computer Vision (ECCV, 2016)* 816–832. <https://doi.org/10.1007/978-3-319-46466-449>

About the Authors



Guiyuan Li received his M.S. degree from Liaoning University of Technology, China, in 2006, and his Ph.D. degree from Jilin University, China, in 2021. Since 2006, he has been a lecturer at Liaoning University of Technology. His research interests are system control and simulation, image processing, and self-driving vehicle applications. (2609582477@qq.com)



Changfu Zong received his M.S. and Ph.D. degrees from Jilin University, China, in 1994 and 1998, respectively. Since 2001, he has been a professor at Jilin University. His research interests are in vehicle system dynamics and vehicle chassis control. (378811297@qq.com)



Dong Zhang received his M.S. degree in 2015 from Jilin University, Changchun, China, and his Ph.D. degree in 2019 from University of Lincoln, Lincoln, UK. He is currently a research fellow at the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. His research interests are in vehicle dynamic control and advanced driver assistance systems. (dr_dongzhang@outlook.com)



Tianjun Zhu received his Ph.D. degree in vehicle engineering from Jilin University, Changchun, China, in 2010. Since 2017, he has been a professor at Zhaoqing University with the Mechanical and Automotive Engineering College in Guangdong Province. His research interests include vehicle dynamic control, new energy vehicles, and self-driving vehicle applications. (happy.adam2012@hotmail.com)



Jianying Li received his M.S. degree in safety technology and engineering from South China University of Technology, Guangzhou, China, in 2010 and his Ph.D. degree in mechanical design and theory from South China University of Technology in 2016. From 2010 to 2013, he worked in an enterprise. He is currently a lecturer at School of Mechanical and Automotive Engineering of Zhaoqing University. His main research focuses on vehicle dynamics and robot design and control. (360103919@qq.com)