論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Depth Estimation and Object Detection in Different Scales by Using PCA and Multi-pooling Layers
著者(和文)	GUOYuxiang
Author(English)	Yuxiang Guo
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第12541号, 授与年月日:2023年9月22日, 学位の種別:課程博士, 審査員:熊澤 逸夫,山口 雅浩,鈴木 賢治,小尾 高史,篠﨑 隆宏
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第12541号, Conferred date:2023/9/22, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

系・コース： Department of, Graduate major in	Information and Communications Engineering	系 コース	申請学位（専攻分野）： Academic Degree Requested	博士 Doctor of	（Engineering）
学生氏名： Student's Name	Guo Yuxiang		審査員主査： Chief Examiner	Kumazawa Itsuo	

要旨（英文 800 語程度）
Thesis Summary（approx.800 English Words ）

In order to enable CNN to recognize depth information, first we use the texture cue and blur cue information obtained from the pre-processing of the 2-D image as the clues of the depth information, and judge the depth of different positions on the 2-D image by comparing the local clues. This obtains the relative depth of each block on the entire 2-D image. Depth information is inferred from a monocular camera 2-D image. And it is taken as the classification of "near" or "far" region when two blocks in the image are compared with respect to their distances. As the focus principle of perspective view, the texture cue is rich around the focus spot, and the blur cue is surrounding the monocular camera spot. Therefore, the depth information of the entire image can be obtained from the relative depth comparison of all blocks. In the absence of disparity information, the relative depth information obtained by this method is more reliable.

PCA (principal component analysis) process is conducted to extract the edge lines corresponding to eigenvector orientation. The proposed classified edge line orientation is very helpful in extracting local depth cues because of the enhanced effectiveness of the spatial frequency analysis process. The depth information of the entire image can be obtained from the relative depth comparisons of all blocks. To adapt to the real world application, this depth estimation method is verified by BSD (blind spot detection) function implementation for ADAS (advanced driver assistant system) with acceptable detection performance under the limited conditions of cost-effective monocular image processing technology. In the absence of disparity information, the relative depth information obtained by this method is more reliable.

The object scale on 2-D image is changing along with the depth. The larger the depth is, the smaller the object scale; and the vice versa. Therefore, it is necessary to take into account the change of object scale when CNN recognizes the scale variant object. Many object recognition methods at different scales are benefited from the features extracted from deep CNN structure. One of the most widely used structures is the pyramid structure, in which the image is resized to be processed in multiple levels. Different object scale belongs to a level in the pyramid. And no matter which pyramid level an object belongs to, its classification attribute keeps unchanged. The coarse-to-fine pyramid can also support feature extraction of texture cue and blur cues.

The eigenvalue calculated through PCA can represent the object point distribution, i.e. the object shape. The eigenvalue changes along with the scale variation. In order to get the scale invariant shape feature, the eigenvalue normalization is executed to remove the influence of the object scale dependency. Through theoretical calculation and experiment verification, it is proved that the normalized eigenvalues can represent the scale invariant shape features of the object. Consequently, in order to enhance the performance and robustness of scale variant object detection, a novel network structure is proposed in which

the module of multi-pooling structure and PCA processing is embedded before fully connected layers. The convolutional process is equal to apply noise filter to the input image data. And then the feature map becomes more concentrated on the activated object shape. In this case, the feature of scale invariance is extracted through module of the parallel PCA processing with multi-pooling layers. The fully connected layers produce the Softmax accuracy prediction for scale variant object recognition.

This modified deep CNN model is trained with the dataset of scale variant vehicle images captured from road vehicle traffic scene. In the experiments, different network structures are implemented: VGG-16, VGG-16 with SPP, VGG-16 with multi-pooling module and VGG-16 with proposed structure. The detection performance of variant vehicle scale is compared between basic VGG16 network and the spatial pyramid pooling network as well as the proposed network structure of module with PCA and multi-pooling layers. The other networks of the current popularity are also tested by our dataset. These networks are ViT-L/16(2023), Nasnet-A(6)(2018) and Resnet-152(2016). Through these comparison studies, it is proved that the proposed module achieves the state-of-the-art result. Moreover, we embedded the proposed module into YOLO network, and the detection performance is evaluated by mean average precision. The experiment is carried out by using dataset PASCAL VOC. The YOLO network embedded by our proposed module achieved better accuracy than YOLO itself. Through the comparisons, the proposed module contributes the feature extraction of scale invariance and enhances the recognition accuracy for scale variant object classification. The testing results demonstrated that the proposed network structure can achieve the comparable detection accuracy as per the state-of-the-art result. The module of multi-pooling loop parallel to PCA processing loop can be embedded to other deep network structure and the comprehensive capability can be expected. The more efforts are necessary to verify the proposed module practicality with generality of other network applicability and with a bigger scale of dataset in application. It is also necessary to further validate the algorithms with regarding to the problems of scale variant object recognition including semantic segmentation, and further step of scale variant object detection.