A Large-scale VR Panoramic Dataset of QR Code and Improved Detecting Algorithm

Zhu Zehao, Zhai Guangtao, Zhang Jiahe, Jia Jun and Yi Fuwang

September 28, 2020

A Large-scale VR Panoramic Dataset of QR code and Improved Detecting Algorithm

Zehao Zhu, Guangtao Zhai, Jiahe Zhang, Jun Jia, and Fuwang Yi

[1] Institute of Image Communication and Network Engineering
[2] Shanghai Jiao Tong University
[3] Shanghai, China
[4] email:{zhuzehao, zhaiguangtao, jiajun0302,
zhangjiahe,yifuwang}@sjtu.edu.cn

Abstract. With the rapid development of mobile payment and scanning technology, QR code has become widespread in both consumer and enterprise domains. However, there is the lack of corresponding research on detecting QR code in panoramic video due to the lack of high-quality datasets. To fill this gap, in this work we propose a large-scale panoramic QR code dataset to facilitate relevant research. Our dataset includes the following characteristics:(1)It is by far the largest dataset in terms of image quantity. (2)Compared with the existing datasets, ours is closer to the realistic setting and can derive a variety of research problems. In addition to the dataset, we propose a QR code detecting approach in a complex environment based on deep learning improving accuracy of QR code detection.

Keywords: VR video panoramic detection· QR code · deep learning.

1   INTRODUCTION

Barcode has become an integral part of life and has played an important role in shopping, advertising, communication, etc. Barcodes were invented in 1949 by Bernard Silver and Norman Woodland [1]. Barcodes can be classified into two different main categories: 1D and 2D barcodes. 1D barcode is a graphic identifier that expresses a set of information by arranging a plurality of black bars and spaces of different widths according to certain coding rules [31–33]. 2D barcode is composed of black and white squares arranged in a certain regular pattern. The common types of the 2D barcode are QR Code, Micro QR Code, Data Matrix, Color Circular Code [28], etc.

QR code is greatly famous and popular in 2D barcodes and has a huge application space in mobile payment, advertising, and many automation scenes [20–22]. For example, Takashi Anezaki invented a robot that can identify QR code to track people [2, 26]. In the shopping field, BlindShopping is a piece of assisting equipment designed to help blind people shop in the supermarket. The blind can use the barcode detecting device to get the current location and product information [3, 25]. Library instruction and content can also be delivered

with smartphones by scanning QR codes, which is more reliable than augmented reality [4, 24]. QR Code can also be embeddef on digital screen via temporal psychovisual modulation (TPVM) [27]. QR code can be used as "invisible watermark" and "physical hyperlink" [14, 37].

Due to the rapid development of machine learning and deep learning, the accuracy of object detection has now been greatly improved. Today's barcodes are usually attached to interesting objects, such as products, express packages, corridors, and the object detection process can be simply replaced with barcode detection, which is easier and more accurate than ordinary object detection. Today, virtual vision has been applied to many fields such as video surveillance [15, 23], intelligent cameras [16, 29] and wearable vision assistance system [17, 30]. The application of QR code detection to the field of virtual vision is very promising [34–36].

When decoding these QR codes, the biggest challenge is that adverse environments, barcode such as blurring, high contrast, dark light, etc, will have various adverse effects on the detection of the barcode. How to detect barcode quickly and reliably is still a challenging task, and has become a critical topic in the computer vision community.

As far as we know, several barcode detection approaches have been proposed and the corresponding QR code datasets have been built, such as Melinda Katona [5], Gábor Sörös [6], Gallo [7], Dubská [8] and Zamberletti [9]. All of the above detection approaches have achieved very good detection results on the corresponding datasets. Unfortunately, these approachess are not effective in the detection of QR codes of VR panoramic videos. In addition, these datasets are far from the actual situation in terms of quantity and quality.

This paper builds a large-scale VR panoramic QR code dataset including QR codes and 1D barcodes. The captured environment includes blurring, excessive exposure, deformation, tilted filming angle, etc(see Fig. 1). Compared with the existing datasets, our dataset is much closer to a realistic scenario. In addtion, an improved detecting algorithm is proposed to detect QR codes. In our approach, we apply region proposal network (RPN) introduced by Faster R-CNN to generate region proposals [18].
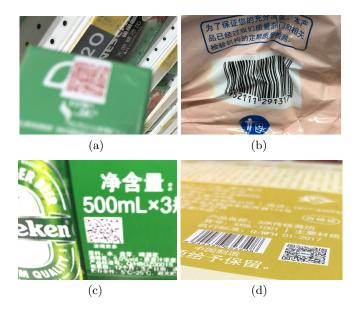
The rest of this paper is organized as follows. In Section 2, we We introduce the dataset of other authors and the corresponding detection methods. Characteristics of our dataset and evaluation metrics are given in Section 3. In Section 4, we describe the details of the proposed method. Section 4 concludes this paper and introduces the future work.

2   RELATED WORK

In this part, we review the existing datasets and detection approaches that are relevant to our task.

The dataset and detecting approach of Gábor Sörös   Gábor Söröss dataset [6] is a 1D barcode and QR code dataset which focuses on testing the

Fig. 1. Several types of low-quality barcodes. (a) blurring QR code; (b) deformed 1D barode; (c) the QR code under the case of high contrast; (d) the barcode under the case of tilted filming angle.

algorithm of detecting blurred 1D barcode and QR code. It includes 328 images (resolution:720×1280). Gábor Söros used the saliency map to detect areas with high a concentration of edge structures as well as for areas with high a concentration of corner structures. However, the environment in which the barcode is located is very simple. The detection algorithm obtained from this dataset is not well suited to Panoramic video detection.

The dataset and detecting approach of Dubská   Dubskás dataset [8] is a QR code dataset which focuses on testing the algorithm of detecting QR code. It includes 400 images (resolution:2560×1440). Most of the images in the database are very clear. Dubská used Hough transformation and parallel coordinates to detect QR code. In actual situations, the QR code may be blurred or deformed in various background. The database and algorithms can only be applied to Panoramic video detection.

The dataset and detecting approach of Zamberletti   Zamberlettis dataset [9] is a 1D barcode dataset. It includes 521 images (resolution:648×488). Among them, the training set includes 366 images, and the test set includes 155 images. Zamberletti presented a novel method based on a supervised machine learning algorithm that detects 1D barcodes in the twodimensional Hough Transform space. However, the dataset is the same as the two datasets above. The captured

environment of this dataset is very simple and is not well suited to Panoramic video detection. In addition, the localization of barcode is a rectangle with an angle, however, the barcode bounding boxes are not accurate when barcodes are distorted, and the background of their training images is not complex.

The dataset and detecting approach of Hansen   Hansen [10] did not build his own dataset. In the training of 1D Barcodes Zamberletti's dataset was used using the split into train and test as provided by the dataset. For the QR barcodes. the QR database provided by Gábor Sörös and the Dubská are used for training. Hansen applied YOLO [13] to detect tect QR code and 1D barcode at the same time. Due to the shortcomings of the data set, the detection results in the actual environment are not ideal.

3   A NEW DATASET FOR DETECTING QR CODE

In this section, we elaborate on the details of the QR code dataset. First, we introduce the characteristics of the dataset. Then, we propose evaluation metrics adopted for detecting QR codes.

3.1   Characteristics of Our Dataset

In this paper, we build QR code dataset to facilitate relevant research. The characteristics of the dataset can be summarized into three aspects.
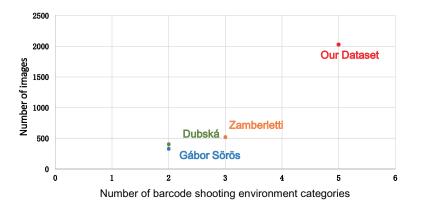

Fig. 2. Comparison with other related datasets in the literature

– **Large-scale:** As shown in Fig. 2, our dataset is the largest dataset so far for QR code detection. To collect this dataset, we capture 2024 images using

insta360 ONE X action camera(resolution:3024×4032). Among them, the training set includes 1300 images, and the test set includes 724 images. This is almost 4 times the size of the category of the previous largest dataset.

– **Multiple environments:** We made the following classifications of the shooting environment in which the images in the database are located. (1)Normal environment; (2)Blurring: The area of QR code is blurred due to inaccurate focus; (3)Excessive exposure: Due to excessive exposure, the black and white data area of the code is not obvious. (4)Deformation: The code on some products will change with the irregular shape of the product. (5)Tilted shooting angle: If the shooting angle is tilted, QR code are stretched at an angle. Their shapes are no longer rectangles, making them undetectable. (6)Shooting light is too dark. As shown in Fig. 2, Our database includes the above six environments. Table 1 shows the number of images in different environments.

– **Close to realistic scene:** The images in our database are all taken in the wild. These QR codes exist in products, advertisements, posters, books, etc. Compared to the single scene of other datasets, our dataset is closed to a realistic scene. The network trained with this data set is more suitable for the detection of the actual environment.

We use the dataset provided by Gábor Sörös and Dubská. These two datasets contain 328 images and 114 images, respectively. We hand-labeled our dataset and the two author's dataset , since they do not provide ground truth. The background of these images is more complex, and some distortions, such as excessive exposure, blur, tilted shooting angle, also exist on images to simulate the real challenging situations. Besides, we also take some images containing small barcodes. We split these 2466 images into a training set and a test set randomly, and the number of the training set is 1742. We evaluate our method on the test set.

Table 1. Comparison with other related datasets in the literature

	Ours	Gábor Sörös	Dubská	Zamberletti
Total number	2024	328	114	521
Normal environment	1304	86	92	450
Blurring	280	242	-	43
Excessive exposure	156	-	-	-
Deformation	110	-	-	-
Tilted filming angle	94	-	12	-
Dark shooting light	80	-	12	28

3.2   Evaluation Metrics

In this section, we use two metrics to evaluate our experiment results.

Intersection over Union (IoU): IoU is a measure of the accuracy of detecting a corresponding object in a particular dataset. It can reflect whether the predicted bounding box matches the ground truth. IoU can be calculated as follow:

$$IoU(GT, DR) = \frac{GT \cap DR}{GT \cup DR} \tag{1}$$

where GT is ground truth and DR is detection result. In general, if $IoU \geq 0.5$, the result is a good test result. We think the detection result is correct.

Mean Average Precision (mAP): The first metric is mean Average Precision. In the target detection algorithm, mAP is often used as a benchmark to measure the accuracy of the algorithm. The nature of mAP is actually an average of each category of average precision(AP). Average precision is calculated from the precision-recall curve. The precision recall are defined as follow:

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$R = \frac{TP}{TP + FN} \tag{3}$$

Calculate the average precision using the following formula:

$$AP = \sum_{n}(R_n - R_{n-1})P_n \tag{4}$$

where n is rank order, R_n and P_n are precision and recall at rank n. mAP is the average of each class of AP:

$$mAP = \frac{1}{N}\sum_{n}^{N} AP_n \tag{5}$$

where N is the total number of class.

Mean Area Ratio (mAR): The second metric that we propose is mean Area Ratio. In the actual situation, it is not enough to locate only the barcode. The final purpose is to obtain the decoding result of the barcode. The more accurate the prediction bounding box is, the higher the decoding rate will be. In order to evaluate the accuracy of barcode locating, we use mAR as another evaluation metric. The mAR is defined as follows:

$$AR = \begin{cases} \frac{|A_p - A_t|}{A_p} & if \ IoU > 0.5 \\ 0 & otherwise \end{cases} \tag{6}$$

$$mAR = \frac{1}{N}\sum_{n}^{N} AR_n \tag{7}$$

where A_p is prediction bounding box and A_t the area of barcode in prediction bounding box. AE is the proportion of bar code in the prediction bounding box. N is the number of prediction bounding boxes. For one ground truth, $AE \in [0,1]$, so that if IoU is less than 0.5, we think it is a false prediction and set AE to 0

4   OUR METHODOLOGY AND EXPERIMENT

In this section, we use two metrics to evaluate our experiment results.

4.1   Our Methodology

In this section, we will introduce our approach for barcode detection. Firstly the VR Panoramic video will be divided into video frames(panoramic images), then we will use Faster R-CNN algorithm [10] to detect QR code in the panoramic image, and finally we will stitch the detected images according to the frame rate of the original video. The network structure is shown in Fig. 3. The whole process can be separated into two stages: one is region proposal network (RPN) and another is a classification network and a bounding box regression network.
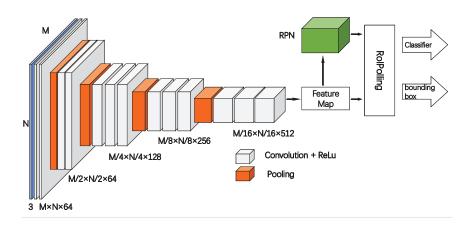

Fig. 3. Network structure. Input image is on the left with size $M \times N \times 3$, and filters are consistent with VGG16 [11]

We also use a classification layer to predict region proposals, and a regression layer to refine region proposals. The shape of region proposals is a horizontal rectangle which simplifies the RoI pooling. The rectangle region proposal comes from the circumscribed rectangle of the quadrilateral.

The network process is as follows. In Convolution Layers, the input is images and the output is the feature maps of the extracted image, which is used as a shared feature of the Region Proposal Network(RPN) and the fully connected

layer. In RPN, generate region proposals based on feature maps. Get the ideal proposals by bounding box regression on anchors [12]. In RoI Pooling, the input is feature map and proposals, and the output is the extracted proposal feature maps, which are used for fully connected layer pair category determination. In Classifier, the proposal feature maps are used to calculate the category of the proposal, and the bounding box regression is performed again to get the exact object detection frame position. In the last step, train classification probability and bounding tree regression using Softmax Loss (Detection classification probability) and Smooth L_1 Loss (Detection border regression). By regressing the bounding box, we can get the detection frame position of the object. Softmax Loss is defined as follows.

$$L = \sum_{j=1}^{T} y_j \log S_j \tag{8}$$

L is a loss. S_j is the jth value of the output vector S of softmax. The range of j is 1 to the number of categories T. y is a $1 \times T$ vector with T values inside, and only 1 value is 1, and other T-1 values are 0. The flow chart of the algorithm is shown Table 2.

Smooth L_1 is defined as follows.

$$SmoothL_1 = \begin{cases} 0.5x^2 & |x| < 1 \\ |x| - 0.5 & otherwise \end{cases} \tag{9}$$

$x = f(x_i) - y_i$ is the difference between the true value and the predicted value

We apply RoIAlign method proposed by He [13] to get each Region of Interest (RoI), and extract feature from integrated feature map according to each RoI.

4.2   Experimental Details

In our experiments, we use VGG16 [10] network as our base network. The VGG16 network is initialized with an ImageNet-pre-trained model. For testing and training the same computer has been used with an Intel(R) Xeon(R) E5-1620 v3 3.50 GHz processor, 16 GB DDR4 memory, 1TB SSD hard drive and Nvidia GeForce GTX 1080 with Windows 10 as the operating system. Before the images are fed into the network, we resize the shortest side to 600. We use four scales: 512, 256, 128 and 64, and three aspect ratio: 1:1, 1:2, and 2:1, for our anchors. The number of candidate region proposal is 300. We use insta360 ONE X to shoot VR panoramic videos with QR codes. The frame rate of this video is 30.

Data augmentation techniques are applied to expand the training set in our experiment. Inspired by Zhang [19], We use image flipping and image rotation to process our images during the training procedure. Image flipping directions are horizontal and vertical, image rotation angles are 90, 180 and 270. For each image, we randomly choose image flipping or image rotation. If image flipping is chosen, we randomly choose the flipping direction. Similarly, if image rotation is chosen, we randomly choose the rotation angle.
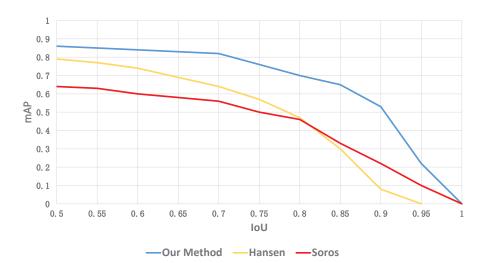
Fig. 4. mAP vs: IoU overlap ratio

Table 2. Comparison with different methods

Method	mAP	mAR
Gábor Sörös [6]	0.538	0.588
Hansen [10]	0.642	0.685
Dubská [8]	0.581	0.601
Our method	**0.886**	**0.811**

4.3 Performance

In this section, we evaluate our model. In the first experiment, we compare our method with others' methods. Henson [10] used a *You Only Look Once* network(YOLO) [13] to detect QR codes. Compared to YOLO, although the calculating speed of Faster R-CNN is slower, the accuracy is higher. The methods of Dubsk*á* [8] and Zamberletti [9] only detect one barcode category, as show table 2. Dubsk*ás* method relies on Hough Transform to detect lines. But our experimental dataset contains a distorted image, therefore, their line detector cannot detect straight lines effectively. This restricts the performance of their method. Although the method proposed by G*á*bor S*örös* [6] can resist blur, it is less effective in complex environments. In the second experiment, we compare our model with others' models. The result is shown in Fig. 4. The superiority of the proposed method is obvious. With IoU overlap ratio increasing, the mAP of the Faster R-CNN model decreases significantly. The qualitative results can be seen in Fig. 5. we localize the barcodes and classify them. The picture in Fig. 5 represents a certain video frame in the test video. From the results, we can see that the network can detect distorted QR codes such as blur and deformation.
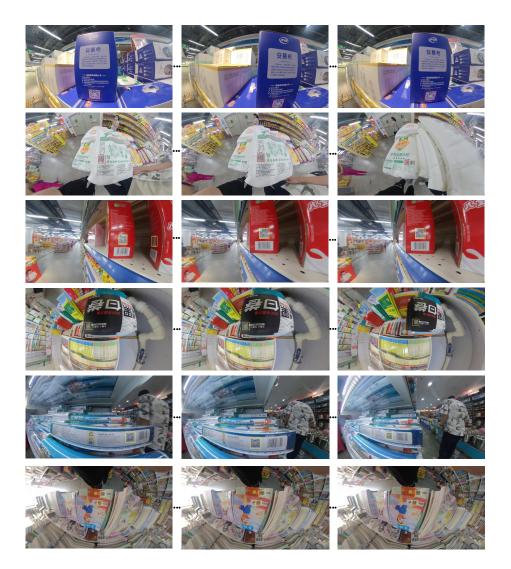
Fig. 5. Example video snippets of object instances

5   CONCLUSION AND FUTURE WORK

This paper presents a large-scale VR panoramic QR code dataset. Comparing with the existing datasets, ours is closer to the realistic setting. At the same time, we propose our method for detecting QR codes. After testing, our model is superior to other models. Compared with others, they have certain advantages. In the future, We will expand the number of images of the dataset. In addition, the algorithm will be extended to detect more barcodes.

6   ACKNOWLEDGE

References

1. Tuinstra T R. Reading barcodes from digital imagery[D]. PhD thesis, Cedarville University, 2006.
2. Anezaki T, Eimon K, Tansuriyavong S, et al. Development of a human-tracking robot using QR code recognition[C]//2011 17th Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV). IEEE, 2011: 1-6.
3. Lpez-de-Ipia D, Lorido T, Lpez U. Blindshopping: enabling accessible shopping for visually impaired people through mobile technologies[C]//International Conference on Smart Homes and Health Telematics. Springer, Berlin, Heidelberg, 2011: 266-270.
4. Walsh A. QR Codesusing mobile phones to deliver library instruction and help at the point of need[J]. Journal of information literacy, 2010, 4(1): 55-65.
5. Katona M, Nyl L G. Efficient 1D and 2D barcode detection using mathematical morphology[C]//International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing. Springer, Berlin, Heidelberg, 2013: 464-475.
6. Sörös G, Flrkemeier C. Blur-resistant joint 1D and 2D barcode localization for smartphones[C]//Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia. 2013: 1-8.
7. O. Gallo and R. Manduchi, Reading 1d barcodes with mobile phones using deformable templates, IEEE transactions on pattern analysis and machine intelligence, vol. 33, no. 9, pp. 18341843, 2011.
8. Dubská M, Herout A, Havel J. Real-time precise detection of regular grids and matrix codes[J]. Journal of Real-Time Image Processing, 2016, 11(1): 193-200.
9. Zamberletti A, Gallo I, Albertini S. Robust angle invariant 1d barcode detection[C]//2013 2nd IAPR Asian Conference on Pattern Recognition. IEEE, 2013: 160-164.
10. Hansen D K, Nasrollahi K, Rasmussen C B, et al. Real-Time Barcode Detection and Classification using Deep Learning[C]//IJCCI. 2017: 321-327.
11. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
12. Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[C]//Advances in neural information processing systems. 2015: 91-99.

13. K. He, G. Gkioxari, P. Dollar, and R. Girshick, Mask r-cnn, in Proceedings of the IEEE international conference on computer vision, pp. 29612969, 2017.
14. Gao Z, Zhai G, Hu C. The invisible qr code[C]//Proceedings of the 23rd ACM international conference on Multimedia. 2015: 1047-1050.
15. Qureshi F Z, Terzopoulos D. Surveillance camera scheduling: A virtual vision approach[J]. Multimedia systems, 2006, 12(3): 269-283.
16. Qureshi F Z, Terzopoulos D. Towards intelligent camera networks: a virtual vision approach[C]//2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance. IEEE, 2005: 177-184.
17. Jiang B, Yang J, Lv Z, et al. Wearable vision assistance system based on binocular sensors for visually impaired users[J]. IEEE Internet of Things Journal, 2018, 6(2): 1375-1383.
18. Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[C]//Advances in neural information processing systems. 2015: 91-99.
19. Zhang J, Jia J, Zhu Z, et al. Fine detection and classification of multi-class barcode in complex environments[C]//2019 IEEE International Conference on Multimedia Expo Workshops (ICMEW). IEEE, 2019: 306-311.
20. Liu Y, Yang J, Liu M. Recognition of QR Code with mobile phones[C]//2008 Chinese control and decision conference. IEEE, 2008: 203-206.
21. DeSoto D B, Peskin M A. Login using QR code: U.S. Patent 8,935,777[P]. 2015-1-13.
22. Kan T W, Teng C H, Chou W S. Applying QR code in augmented reality applications[C]//Proceedings of the 8th International Conference on Virtual Reality Continuum and its Applications in Industry. 2009: 253-257.
23. Jia J, Zhai G, Zhang J, et al. EMBDN: An Efficient Multiclass Barcode Detection Network for Complicated Environments[J]. IEEE Internet of Things Journal, 2019, 6(6): 9919-9933.
24. Zhang J, Li D, Jia J, et al. Protection and Hiding Algorithm of QR Code Based on Multi-channel Visual Masking[C]//2019 IEEE Visual Communications and Image Processing (VCIP). IEEE, 2019: 1-4.
25. Zhang J, Min X, Jia J, et al. Fine localization and distortion resistant detection of multi-class barcode in complex environments[J]. Multimedia Tools and Applications, 2020: 1-20.
26. Jia J, Zhai G, Ren P, et al. Tiny-BDN: An Efficient and Compact Barcode Detection Network[J]. IEEE Journal of Selected Topics in Signal Processing, 2020.
27. Song K, Liu N, Gao Z, et al. Deep Restoration of Invisible QR Code from TPVM Display[C]//2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). IEEE, 2020: 1-6.
28. Yi F, Zhai G, Zhu Z. A Robust Circular Two-Dimensional Barcode and Decoding Method[C]//2019 Picture Coding Symposium (PCS). IEEE, 2019: 1-5.
29. Zhu Y, Zhai G, Min X. The prediction of head and eye movement for 360 degree images[J]. Signal Processing: Image Communication, 2018, 69: 15-25.
30. Zhai G, Min X, Liu N. Free-energy principle inspired visual quality assessment: An overview[J]. Digital Signal Processing, 2019, 91: 11-20.
31. Sharif A, Zhai G, Jia J, et al. An Accurate and Efficient 1D Barcode Detector for Medium of Deployment in IoT Systems[J]. IEEE Internet of Things Journal, 2020.
32. Zhai G, Min X. Perceptual image quality assessment: a survey[J]. SCIENCE CHINA Information Sciences, 2020, 63(11): 211301.

33. Sun W, Min X, Zhai G, et al. MC360IQA: A Multi-channel CNN for Blind 360-Degree Image Quality Assessment[J]. IEEE Journal of Selected Topics in Signal Processing, 2019, 14(1): 64-77.
34. Zhu Y, Zhai G, Min X, et al. The prediction of saliency map for head and eye movements in 360 degree images[J]. IEEE Transactions on Multimedia, 2019.
35. Shen W, Ding L, Zhai G, et al. A QoE-oriented Saliency-aware Approach for 360-degree Video Transmission[C]//2019 IEEE Visual Communications and Image Processing (VCIP). IEEE, 2019: 1-4.
36. Zhai G, Min X. Perceptual image quality assessment: a survey[J]. SCIENCE CHINA Information Sciences, 2020, 63(11): 211301.
37. Jia J, Gao Z, Chen K, et al. Robust Invisible Hyperlinks in Physical Photographs Based on 3D Rendering Attacks[J]. arXiv preprint arXiv:1912.01224, 2019.