Study on Personnel Detection Based on Retinex and YOLOv4 in Building Fire

Wei Li¹, Sen Li²,³*, Yeheng Wang² and Junying Yun²

¹Henan Vocational and Technical College of Architecture, Henan Zhengzhou 450064 China
²School of Building and Environmental Engineering, Zhengzhou University of Light Industry, Henan Zhengzhou 450002 China
*Email: lisen@zzuli.edu.cn

Abstract. When a fire occurs in a building, the internal environment is full of dense smoke, which will greatly hinder the evacuation and rescue of the trapped persons. If the evacuation and rescue are not in time, the life safety of the trapped persons will be seriously threatened. In response to this problem, this paper proposes a method for quickly detecting trapped persons in building fires. This method uses a combination of multi-scale Retinex image sharpening algorithm and YOLOv4 person detection algorithm. First obtain the image information of the fire scene, use the multi-scale Retinex algorithm based on the Gaussian pyramid to perform the sharpening process, and then use the YOLOv4 model to perform the personnel detection on the sharpened fire scene image. The experimental results show that the confidence of image person detection after Retinex sharpening processing has been significantly improved.

1. Introduction

When a fire occurs in a building, high temperature, dense smoke and toxic gases are generated inside, which seriously threaten the life safety of the trapped persons [1]. Building fires often cause casualties. In July 2021, a major fire accident occurred in a logistics warehouse in Changchun. As the people in the building failed to evacuate in time, 40 casualties were caused. Therefore, the study of personnel detection in building fire environment enables the estimation of the number of people trapped inside the building when a fire occurs, determines the specific location of the trapped people, and provides accurate information and technical support for the evacuation and rescue of people in the building. However, when a fire occurs, a large number of smoke particles will be produced, making the visibility in the building drastically reduced [2]. The acquired image information will also greatly reduce the detection of people, and it will not be able to accurately provide effective information, which will hinder the evacuation and rescue of people on the fire scene. To this end, this article explores a method that can quickly detect trapped people in a building fire scene with a harsh environment. Before detecting people on the building fire image, the image should be cleared.

Histogram equalization [3], Dark channel prior method [4], Retinex algorithm [5] and so on are commonly used in image clarity algorithms. The dark channel prior method uses the Sun as the light source to build the physical model, which is mainly applied to the outdoor environment. In comparison, the Retinex algorithm is more suitable for the indoor fire environment [6] [7]. Compared with other algorithms, YOLOv4 has the advantages of fast recognition and high accuracy [8], so it is one of the best algorithms for human detection. In order to improve the efficiency of evacuation and search the
trapped people, this paper combines the Retinex Algorithm and YOLOv4 Algorithm to detect the people in the fire scene video surveillance image.

2. Construction of building fire personnel detection model

2.1. Principle of multiscale Retinex (MSR) algorithm

In Retinex theory, the image seen by the human eye is determined by the incident light and the reflected light of the object [9]. As shown in formula (1), where $R(x, y)$ is the reflecting component of the object, $L(x, y)$ is the incident light, and $S(x, y)$ is the final image seen by the human eye.

$$S(x, y) = R(x, y) \ast L(x, y)$$ (1)

Retinex theory believes that the process of restoring the true color of an object is the process of retaining the reflection component of the object and removing the light component in the image [10] [11]. In order to improve the convenience of computation in image processing and to make the computation more in line with human visual perception, the image is converted into a logarithmic domain, as shown in formula (2):

$$R(x, y) = \log(I(x, y)) - \log(L(x, y))$$ (2)

Jobson et al. [12] studied the wrap function in the Retinex algorithm, and the results showed that the Gaussian function as a wrap function is better. Therefore, the single-scale Retinex algorithm (SSR) is proposed, as shown in formulas (3) and (4):

$$F(u, v) = \frac{1}{2\pi\sigma^2} \exp\left(\frac{-(u^2 + v^2)}{4\sigma^2}\right)$$ (3)

$$R(x, y) = \log(I(x, y)) - \log(I(x, y)) \ast F(u, v)$$ (4)

In formulas (3) and (4), “∗” stands for convolution computation; $F(u, v)$ is the Gaussian wrap function; $(u, v)$ is the coordinate position in the gauss template; $\sigma$ is the standard deviation of the Gauss function, also known as the scaling factor [13]. The multiscale Retinex algorithm combines weight methods under different parameter conditions, as shown in formula (5):

$$H_{MSR} = \sum_{n=1}^{N} \omega_n H_{SSRn}$$ (5)

In Formula (5), $n$ is the number of SSR operations; $\omega_n$ is the weight of the $n$th SSR; $H_{SSRn}$ is the restored image after SSR operations. A color recovery calculation is then added, as shown in Formula (6):

$$H = \sum_{i=R,G,B} C_i H_{MSRi}$$ (6)

In formula (6), $H$ is the image after processing the smoke; $C_i$ is similar to the concept of weight; $i$ is the color Channel R, G and B; Represents the current color channel brightness ratio to the total color channel brightness; $H_{MSRi}$ is the restored image after Multi-scale Retinex operation in $i$ channel [14] [15].

2.2. Gaussian Pyramid Substitution Method

The Retinex algorithm is very time consuming in Gaussian filtering, while the Gaussian pyramid significantly reduces the computation time by reducing the image dimension [16]. The Gaussian pyramid consists mainly of two parts: Dimensionality reduction and Gauss filtering. As shown in the Gaussian pyramid formula (7), $P$ is the order of the Gaussian pyramid; $G_p(x, y)$ is the filtered image of the pyramid of order $p$; $F(u, v)$ is still the gauss template, the size of the gauss template in this paper is 5.

$$G_p(x, y) = \sum_{n=-2}^{2} \sum_{m=-2}^{2} F(u, v) \ast G_{p-1}(2x + u, 2y + v)$$ (7)

The Gaussian pyramid filtering process is to first pass the 5×5 Gaussian template to perform dimensionality reduction filtering on the image to form the next-order filtered image. The same method is used to sequentially accumulate the filtered images to form a filtered image pyramid. When the
pyramid reaches the highest point, the dimensionality operation is performed until the image size reaches the original image size. The process is shown in figure 1.

The number of layers of the pyramid is determined by the size factor $\sigma$. The larger the scale factor $\sigma$, the number of layers of the Gaussian pyramid will continue to increase, and the corresponding processing time should also be increased. However, it is found through calculation that the filtering time of the Gaussian pyramid basically does not change with the scale factor. The main reason is that the increase in the number of Gaussian layers will not have much impact on the processing time. Retinex algorithm fused with Gaussian pyramid, its processing time can be shortened by 25 times [17].

2.3. YOLOv4 personnel detection algorithm

The YOLO algorithm uses the One-stage deep learning target detection method. One-stage does not need to delineate the target area first, and the target location and classification are performed on the same network at the same time. Therefore, there is a higher speed in the detection, which can meet the real-time detection of personnel. Based on the previous generation YOLO network framework, the YOLOv4 network has optimized the data processing, backbone network, network training, activation function, and loss function to varying degrees, and can be applied to the fast target detection system in the actual working environment [18]. Figure 2 shows the network structure of the YOLOv4 personnel detection model.

Figure 1. Gaussian pyramid filtering process.

Figure 2. Network structure of YOLOv4 target detection model.
CSPDarknet53, as the backbone network for YOLOv4 feature extraction, is based on Darknet53 and improved from the cross-stage peer-to-peer network structure. The CSP module contains a down-sampling convolution module, and the backbone network contains 5 CSP modules, which solves the waste of resources and increased calculations caused by repeated learning of gradient information. After the backbone network, SPP (Spatial Pyramid Pooling) module is added to increase the receiving range of features to separate the most significant features in the data. The input backbone network image of the module is 608*608. It undergoes an up-sampling operation while extracting features, and is spliced with Concat to output 19*19, 38*38, 76*76 feature maps. According to the feature map depth calculation formula 3*(5+C), where C represents the target detection type, and the value of C is 80. Therefore, the final output of the feature map prediction is 19*19*255, 38*38*255, and 76*76*255. The CBL module is composed of Conv (convolutional layer), BN (batch normalization), and Leaky_relu activation functions; Concat represents tensor splicing, and splicing will expand the dimension of tensor. The CBM module is composed of Conv (convolutional layer), BN (batch normalization), and Mish activation function.

The detection speed of YOLOv4 on Tesla V100 reached 65 fps, and the AP (Average Precision) on the MS COCO data set reached 43.5%. Compared with other target detection algorithms, the performance parameters of YOLOv4 are compared with other algorithms as shown in figure 3. The running speed of YOLOv4 is twice that of EfficientDet. The performance is comparable, compared with the AP and FPS of YOLOv3 of 10% and 12%, respectively [19].

Figure 3. Comparison of YOLOv4 with other detection algorithms.

Figure 4. Flow chart of personnel detection in building fire.

2.4. Building fire personnel detection model

The flow chart of building fire personnel detection explored in this paper is shown in figure 4. The acquired fire image information is cleared and personnel detected, and finally an image with personnel detection information is output.

3. Experimental Platform Construction

In order to verify the effectiveness of this algorithm in the detection of people on fire scene images, this paper builds a fire scene evacuation simulation experiment device. The device was shaped like an “L”, the overall outline of the device is shown in figure 5. A corridor space of 10m long, 1.6 m wide and 2.2 m high is formed inside the device. The device is equipped with a video monitoring system for obtaining evacuation images in the fire scene. The external console display of the camera is DH-PTZ1B203UE-GN-PD. The camera is a PTZ camera with an installation height of 2.2 m from the ground. It can be rotated 180° horizontally and 90° up and down, and can monitor the environment in the simulated corridor from multiple angles. The installation position is shown in figure 6.
The ignition device is placed out of the ignition window on the right side of the experimental device. Cotton rope is used as fuel in the experiment, and the cotton rope is ignited to produce a large amount of smoke, so as to achieve the effect of personnel evacuation under the simulated smoke environment. A smoke density measuring instrument is installed on the top of the middle of the device to monitor the change of the smoke concentration in the device and verify the image clearness processing effect of Retinex algorithm under different smoke concentrations. The smoke density meter is installed at a height of 2.20 m, as shown in figure 7.

Three LED explosion-proof lights, wall-mounted with a height of 175 cm, are installed in the simulated corridor to simulate emergency lighting in the fire environment, as shown in figure 7. The exit of the simulated corridor is installed with a fireproof rolling shutter door with a width of 1.2 m, and the door switch can be used to control the rolling shutter door. Close to the top of the rolling door and in front of the rolling door, a smoke exhaust port is set. After the outdoor smoke exhaust fan is activated, the corridor smoke is exhausted to the outdoors through the smoke exhaust port through the flue gas duct. A console and a display screen are set outside the simulation corridor to control the experimental equipment inside the corridor and display the information inside the simulation corridor.

### 4. Experimental results and analysis

The operating system of this article is Ubuntu 16.4, and the graphics card model is: NVIDIA Quadro K2000 4 G. According to the YOLOv4 algorithm, the reliability of the recognition of evacuated persons in the image is compared to judge the feasibility of the Retinex image sharpening algorithm. In order to simulate the corridor environment under fire conditions, one of the three LED explosion-proof lights is turned on to simulate dim light conditions. The smoke environment of the corridor is simulated by burning cotton ropes, and the evacuees pass through the corridor in a fire evacuation posture under different smoke concentrations. The experiment simulates the evacuation of people captured by surveillance cameras in the corridor under different smoke concentrations, and uses the YOLOv4 algorithm to detect people.
During the experiment, the cotton rope was ignited to produce smoke, and the smoke concentration gradually increased over time. This article selects five different concentrations of images for detailed analysis. As shown in figure 8, when the ambient smoke concentration is 0.304 dB, the smoke concentration is low at this time, and the contrast of the confidence of the detection before and after image processing is not obvious. In figure 9 and figure 10, the smoke concentration is 0.601 dB and 0.890 dB, respectively. After image processing, the human detection confidence is 0.87 and 0.93, respectively. Compared with the original image, the human detection confidence is increased by about 0.38 times and 0.43 times, respectively. As the smoke concentration increases, the YOLOv4 algorithm has lower and lower confidence in the detection of people before processing. The smoke concentration in figure 11 is 1.176 dB, the confidence in detection of people before processing is 0.52, and the confidence in detection of people after processing rises to 0.86. The confidence is increased by about 0.65 times compared to the original image. The smoke concentration in figure 12 is 1.407 dB. At this time, the upper body of the evacuated person in the original image is blurred, and the confidence of the person before processing is only 0.27. After the image processing, the confidence of the person detection is increased to 0.87, compared with the original image detection confidence of the staff increased by about 2.22 times.

![Figure 8](image1.png)  
(a) Before treatment (confidence level 0.94) (b) After treatment (confidence level 0.96)

![Figure 9](image2.png)  
(a) Before treatment (confidence level 0.63) (b) After treatment (confidence level 0.87)

![Figure 10](image3.png)  
(a) Before treatment (confidence level 0.65) (b) After treatment (confidence level 0.93)

![Figure 11](image4.png)  
(a) Before treatment (confidence level 0.52) (b) After treatment (confidence level 0.86)

![Figure 12](image5.png)  
(a) Before treatment (confidence level 0.27) (b) After treatment (confidence level 0.87)

![Figure 13](image6.png)  
Figure 13. Line chart of confidence comparison before and after processing.
According to the above experimental results, the line graph is drawn as shown in figure 13 to compare the confidence of the person detection before and after the processing. The horizontal axis is the flue gas concentration, and the vertical axis is the person detection confidence. It can be clearly seen that as the flue gas concentration increases, it is processed by the multi-scale Retinex algorithm proposed in this paper. When the flue gas concentration is 0.60-1.10 dB, the confidence of personnel detection is increased by about 0.5 times, and when the concentration is about 1.10-1.40, the confidence of personnel detection is increased by about 2 times.

5. Conclusion
This paper proposes a method for rapid detection of people in building fires. The multi-scale Retinex fire field image clearing algorithm based on Gaussian pyramid is used to clear the images of evacuated people in smoke. The clarified image is then used YOLOv4 target detection algorithm to detect people. The experimental results show that when the smoke gas concentration is 0.50-1.40 dB, the confidence of the personnel detection after the image sharpening process can be increased by about 0.90 times compared with the original image. When the flue gas concentration reaches about 1.40 dB, the processed person detection confidence can be increased by about 2.22 times compared with the original image. It greatly improves the utilization efficiency of visual information in fire rescue, and provides important technical support for speeding up building fire rescue.

Acknowledgments
This paper was supported by the National Natural Science Foundation of China (51804278), the Key Science and Technology Project of Henan Province (212102210535), and the Key Scientific Research Project of Henan Province (20A620005).

References
[1] Park H, Meacham B J and Dembsey N A, Goulthorpe M 2014 Building Research & Information 42 696-709
[2] Park S, Han KW and Lee K. 2020 Int Conf on Information and Communication Technology Convergence (Jeju South Korea: 345 E 47th ST, New York 10017 USA) pp 1563-65
[3] Patel S and Goswami M 2014 Int Conf on Contemporary Computing and Informatics (mysuru India; 345 E 47th ST, New York 10017 USA) pp 167-8
[4] Liu F and Yang C M 2014 Int Conf on Signal Proc, Communications and Computing ( Guilin China: 345 E 47th ST, New York 10017 USA) pp 483-6
[5] Land E H and McCann J J 1971 Journal of the Optical Society of America 61 1-11
[6] Hassan N, Ullah S, Bhatti N, Mahmood H and Zia M 2021 Multimedia Tool and Applications 80 1839-18
[7] Shao G, Gao F, Li TJ, Zhu R and Pan T 2020 Chinese Automation Congress 6294-99
[8] Wang S 2021 Int Conf on Big Data. Artificial Intelligence and Int of Things Engineering (Chongqing China: Institute of Electrical and Electronics Engineers Inc) pp 877-881
[9] Xu K and Jung C 2017 International Conference on Acoustics Speech and Signal Processing (New Orleans LA: 345 E 47th ST, New York 10017 USA) pp 1363-67
[10] Parihar A S and Singh K 2018 Int Conf on Inventive Systems and Control (Coimbatore India: 345 E 47th ST, New York 10017 USA) pp 619-624
[11] Shao G, Gao F, Li T, Zhu R, Pan T and Chen Y 2020 Chinese Automation Congress (Shanghai China: 345 E 47th ST, New York 10017 USA) pp 6294-99
[12] Alexander Z 2018 Procedia Computer Science 131 6-14
[13] Lin C, Zhou H and Chen W 2020 Laser Optoelectronics Progress 57 16
[14] Michela L, Gabriele S, Cristian B and Alessandro R 2018 Iet Image Processing 12 833-849
[15] Zhu M Y, Su F and Li W J 2011 Applied Mechanics & Materials 39 32-37
[16] Wang L, Horiuchi T and Kotera H 2007 Journal of Imaging Science and Technology 51 34-43
[17] Li S, Wang S, Zhang D, Feng C and Shi L 2019 Image and Video Processing 13 1037-43
[18] Bochkovskiy A and Wang C Y, Liao H Y M 2020 *arXiv preprint arXiv:2004*
[19] Boyuan W and Muqing W 2020 *Int Conf on Computer and Communications* (Chengdu China: Institute of Electrical and Electronics Engineers Inc) pp 1198-1202