Master of Electrical Engineering

THE SMOKE DETECTION FOR EARLY FIRE-ALARMING SYSTEM BASED ON VIDEO PROCESSING AND CNN

The Graduate School
of the University of Ulsan

Department of Electrical Engineering
University of Ulsan, Korea
Tao Peng

THE SMOKE DETECTION FOR EARLY FIRE-ALARMING SYSTEM BASED ON VIDEO PROCESSING AND CNN

Supervisor: Professor Byeong-Woo Kim

A Dissertation

Submitted to
The Graduate School of the University of Ulsan
In partial fulfillment of the requirements
for the Degree of

Master of Electrical Engineering

by

Tao Peng

Department of Electrical Engineering
University of Ulsan, Korea
February 2021

THE SMOKE DETECTION FOR EARLY FIRE-ALARMING SYSTEM BASED ON VIDEO PROCESSING AND CNN

This certifies that the dissertation of Tao Peng is approved.

_____

Committee Chair Professor. Han-sil Kim

_____

Committee Member Professor. Byeong-Woo Kim

_____

Committee Member Dr. Su-jin Kwag

Department of Electrical Engineering

University of Ulsan, Korea

February 2021

ACKOWLEDGEMENTS

When this thesis was completed, I felt a lot of emotion! Many people have offered me valuable help in my thesis writing, include my supervisor, my colleagues and my parents.

Firstly, I would like to thank my supervisor Professor Kim and Professor Cho for their meticulous guidance on my thesis, the selection of the theme of the thesis, the construction of the thesis frame, the supply of relevant materials, and the revision of the later stage for their valuable suggestions helped me successfully completed this thesis. I am grateful to Professor Cho for his meticulous care in all aspects of my study, work, and life in the past two years. I also thank for Professor Kim for not only granting me knowledge without reservation, but also teaching me the principle of being a researcher and how to work. His scientific and rigorous academic spirit, conscientious work attitude, keen and innovative design thinking all influence me subtly, and let me benefit a lot from his precepts and deeds.

I am grateful to the Department of Electrical Engineering of University of Ulsan, for creating good learning conditions and a strong academic atmosphere for me, allowing me to spend two years of pleasant and unforgettable study time, I am grateful to all the teachers, the knowledge you impart will become my precious wealth, used for life.

I am grateful to my dear family, friends and colleagues in the laboratory. The two years of study and life have been both painful and joyous. I am grateful to everyone along the way!

Thanks again!

요약

최근, 세계에서 대형 산불이 잇따라 발생해 우리 생활과 밀접한 차량 연소사고를 연상케 한다. 새로운 에너지 자동차의 개발로, 전기 자동차의 수량은 해마다 증가하고 있다. 그러나 불확실한 요소가 많아 전기자동차의 자연연소 현상이 항상 발생하는데, 특히 야외주차장에 주차된 전기자동차가 자연 발화했을 때 제때 알아내기가 어렵다, 일단 불이 번지면 단시간에 진화에 어려움을 겪는다. 화재의 통제는 주로 예방에 있다. 기존 일부 기술을 활용해 화재가 발생했을 때 조기경보를 확인하고 발령할 경우 화재가 확대되기 전에 진화할 수 있다. 스모크 화염이 타기 전에 가장 눈에 띄는 특징이기 때문에 화재 감지가 화염이 타지 않도록 하는 역할을 할 수 있다. 스모크 반투명한 특성이 가지고 있으며, 외적인 간섭에 의해 형태와 질감 특성이 쉽게 변화한다. 이러한 특성들이 스모크 인식 난이도를 결정한다. 본 논문은 주로 컴퓨터 비전과 CNN 기술을 활용하여 주차장 내 초기 화재로 발생한 스모크 감지하고 인식한다. 본 논문의 주요 연구 내용은 다음과 같다.

1. 사전 준비단계에서 수집·정렬을 통해 알고리즘의 효과를 검출하기 위해 사용되는 12 개 동영상을 선택하였다. CNN 의 훈련·시험에 사용되는 12,470 개 스모크 이미지와 비 스모크 이미지 12,902 개가 포함된다.

2. 화재 스모크의 정적(색상)과 역동적인 측면에서 화재 초기 스모크의 일반적인 특성을 분석한다.

3. CNN 을 분류기로 삼아 스모크 이미지를 인식한다.

본 논문의 업적은 다음과 같은 혁신이 있다.

I. 데이터 세트. 본 논문에서 사용된 데이터 세트는 실외환경에서 녹화한 영상이다. 데이터 세트는 실제 상황에서 실제 화재 스모크 현장을 시뮬레이션 할 수 있는 실용적 있다.

II. 스모크 색채 특성 추출. 본 논문에서는 RGB 색 공간의 낮은 화질로 인한 특징 장애를 방지하는 스모크 속 HSV 색 공간의 특성에 초점을 맞춘다.

III. CNN 은 스모크 인식으로 사용된다. 신경망은 다수의 훈련 데이터를 통해 영상의 특성을 자동으로 학습할 수 있어 단일 형상의 인위적 선택 문제를 피할 수 있다.

IV. 마지막으로 완전한 야외 화재 스모크 인식 시스템이 설계되어 화재 스모크 조기 경고의 야외에서 시스템을 달성할 수 있다. 그런 다음 제안된 알고리즘의 효과를 검증하기 위해 네 그룹의 대조도 실험을 사용한다. 제안된 방법이 다른 방법보다 정확도가 높고, 허위 경보율이 낮으며, 인식 누락률이 높다는 실험 결과도 CNN 의 현장 적용이 일정한 연구 가치를 가지고 있음을 입증한다. 본 논문의 끝에는 제안된 알고리즘의 장단점을 요약하고, 알고리즘의 단점에 따라 그에 상응하는 해결책과 향후 연구의 방향을 제시한다.

키워드: **스모크 인식; 이미지 처리; GMM; HSV; CNN**

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 The Research Significance of the Subject

In 2019, Amazon forest fires occurred frequently. As of August 18, 2019, the number of Amazon forest fires has exceeded 72,000 since 2019. Beginning in September of the same year, a forest fire broke out in Australia, which lasted for nearly 5 months and was not extinguished until 2020. The fire caused very great losses to Australia and even the entire human race. From these news content can always remind people of fire accidents that are closely related to our lives. Especially the fire in the parking lot that also occurs outdoors, and something that will cause us large property losses. Through the investigation of many large-scale fire accidents in the parking lot, it was found that the initial stage of the fire was caused by the spontaneous combustion of individual vehicles and other reasons. The discovery was not timely, and the rapid spread of the fire caused the fire to intensify and finally caused more serious fire accidents.

On February 1, 2019. 17 cars catch fire at Newark Airport parking garage. February 23, 2019. 300 cars gutted in massive fire at parking lot near Aero India show in Bengaluru. April 17, 2019, vehicles in the parking lot of a community in Guangdong ignited spontaneously, lead to 15 cars were burned. June 29, 2019, a vehicle spontaneous combustion accident occurred in an underground parking lot in Qingdao, which burned more than 40 cars, April 5, 2020. 3,500 Rental Cars Destroyed by Massive Fire in Florida. April 27, 2020, a vehicle spontaneous combustion accident occurred in an outdoor parking lot in Shenzhen, and the incident caused 5 nearby cars to be burned. Vehicles in the parking lot are relatively dense. Once a fire takes shape, it will inevitably cause a large amount of property losses and even casualties. If there is a monitoring system that can detect the occurrence of fires and eliminate fire hazards as soon as possible, to eliminate the fire in an unburnt state can better protect the safety of human life and property. Smoke detection is the early detection of fires that occur. When the fire occurs, people are given a fire alarm, so that we can carry out put out a fire in time, and most reduce the loss caused by the fire. This thesis combines the static and dynamic characteristics of smoke and applies image recognition technology to early fire detection, which has important practical significance and application value for fire prevention, fire scale judgment and put out a fire.

Figure 1-1: 2019, Amazon forest and Australia forest fires.

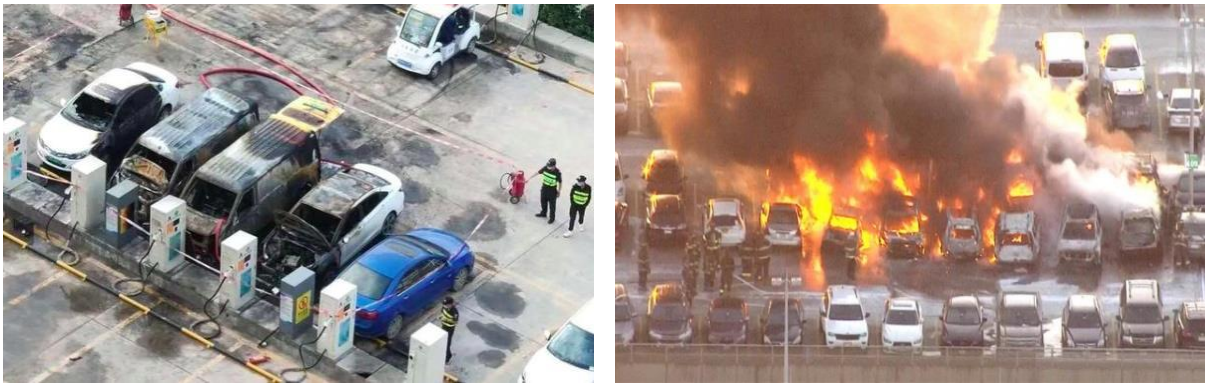


Figure 1-2: Spontaneous combustion of vehicles in outdoor charging stations and Newark Airport parking lot.

In recent years, with the development of computer science, computer vision and artificial intelligence have formed a research hotspot. Many new and improved algorithms have been continuously proposed [3-5], making the field of computer vision faster Development and application of computer vision technology to early fire smoke detection has important practical significance and application value for fire prevention, fire scale judgment and put out the fire.

1.2 Current Status of the Research

For fire smoke detection, sensors and infrared probes are now used more frequently. Sensors include smoke detector sensors, gas sensors, and temperature sensors. By monitoring the carbon dioxide and carbon monoxide in the air, and the temperature is monitored to achieve the purpose of detection. This detection method is effective in indoor enclosed areas and small space areas, such as tunnels. But for the outdoor, due to the limited sensing range of sensors, generally for spaces exceeding $12m$ [6], conventional point fire detectors can't play a very good role to detection the smoke. It is even impossible to install the large number of sensor equipment in the outdoor environment. In addition, the environment changes drastically in the outdoor environment, and the sensor equipment is easily damaged when exposed to the outdoors for a long time, causing false alarms. It is a cheap, convenient and accurate feasible solution to use a camera to monitor and use computer vision technology to detection smoke.

Scholars have been conducting research in this area since the 1990s. Early research was mainly aimed at the recognition of flames, most of which used its bright and color characteristics to be much obvious and easy to extract features. From the initial RGB color model [8-10], to the HIS [11], YCbCr [12] color model, later researchers have used these features to do corresponding research. For the convenience of improve accuracy, more researchers began to pay attention to other characteristics of flames and smoke. For flames, Wang Lin et al. [13] studied the characteristics of flame beats and found that the beat frequency of the flame's centroid was about $10HZ$, this feature can be used to distinguish flame and non-flame areas. Yuan Feiniu et al. [14] identified the smoke based on the characteristic that the smoke always moves upward after it is produced. Yaqin Zhao [15] and Yuan Feiniu [16] also used LBP and LBPV to extract local features to identify flames and smoke. Many scholars use wavelet analysis to extract edge features or wavelet energy [15] features for identification. For the convenience of further improve the accuracy of recognition, many scholars have also begun to integrate multiple features for recognition [21-23]. The main fusion features are color, motion state, background blur, wavelet transform, etc. Literature [24] uses the temporal and spatial features of smoke combined with dynamic texture model to detect smoke.

In recent years, many scientific research teams have begun to conduct research on fire smoke identification. The Smart Media Computing Laboratory of Sun Yat-sen University published a paper in 2014 [25], published the source code and examples on its website. Professor A. Enis Cetin of Bilkeng University in Turkey published several papers on fire video detection from 2008 to 2009 [26], put fire video detection examples and program installation packages on his website.

The Computer Vision and Pattern Recognition Laboratory of Keimyung University in South Korea released four versions of flame and smoke detection systems for four years from 2008 to 2013, and published several papers [27-28], Constantly put forward new detection methods, and the accuracy rate is constantly improving.

The Machine Intelligence Laboratory of Salerno, Italy established its own fire smoke video data set in 2012 and published papers in 2014 and 2015 [29-30]. The Multi Expert System (MES) designed by it is accurate in identifying flames. The rate reached 93.55%.

However, in recent years, most of the research on fire smoke detection is aimed at the detection of flames, and most of them are aimed at indoor spaces, and the detection of smoke is currently relatively small. In addition, because there is no standard video data set for fire and smoke detection, most of the video data used in the research are collected by some visual research laboratories [31-33]. Most of the flames and smoke in these videos are made indoors. The particularity of fire in the external natural environment. The Machine Intelligence Laboratory of the University of Salerno has collected a relatively complete set of outdoor smoke and fire video data sets, and conducted related research, but its main research is the detection after the fire has occurred, and there are fewer early detections.

With the popularity of deep learning, it has excellent performance in speech recognition [34-35], image recognition [36-38], natural language processing (NLP) [39-40] and other fields. Convolutional neural network (CNN) is a multi-layer perceptron designed to recognize two-dimensional shape data. It is a special deep neural network model. Its local connection and weight sharing characteristics reduce the complexity of the network model. CNN also have a lot of research and applications in the field of image recognition [41-45]. In recent years, some people have also started to talk about the application of CNN in fire and smoke recognition. In reference [46], the author designed a three-layer CNN as a recognition model for fire images, achieved better recognition results on small sample data sets using the parameter replacement method in reference [47]. The author uses the deep learning framework-Caffe to implement a CNN, compares the effects of a variety of different CNN and a relatively mature outdoor fire algorithm-fshell of the German Space Center on outdoor fire smoke video sets. The research results show that CNN can also perform well in outdoor fire recognition.

Early fire smoke detection can better realize fire monitoring and warning, has more practical significance and research value for fire prevention. However, there are few studies on this content at present, and the research on outdoor early fire smoke detection technology for many different types of fire and smoke videos, explore new algorithms for the complexity and strong interference of outdoor environments.

1.3 Introduction to Smoke Detection Technology

Generally, fire smoke recognition is mainly divided into four steps as shown in Figure 1-3 below: image preprocessing, extraction of the area to be detected, feature extraction, and classification and recognition.

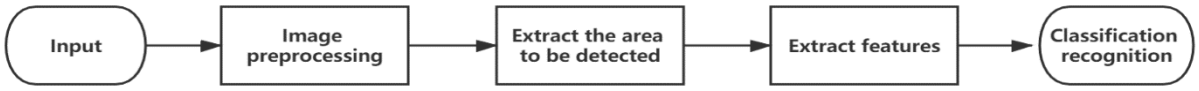


Figure 1-3: Diagram of the identification steps of fire smoke.

Image preprocessing: mainly for image graying, denoising, key frame extraction and other steps, for the convenience of eliminate interference and enhance the target area.

Extract the area to be detected: mainly use the obvious characteristics of the movement of the smoke, color, etc., to segment the suspected smoke area in the video, avoiding the calculation of the entire image to reduce the amount of calculation.

Extract features: It mainly extracts features such as texture, motion direction, color, and energy of the target area as the basis for the next step of classifying the images.

Classification recognition: mainly use Support Vector Machines (SVM), Decision Tree, Artificial Neural Network (ANN) and other classifier models to classify and recognize the extracted regions.

In traditional recognition methods, appropriate features must be selected manually, and then the features can be extracted through algorithms before they can be input into the classifier for classification. However, due to the non-rigidity, boundary uncertainty, and translucency of smoke, the characteristics of smoke are not so obvious, and it is easily interfered by other factors. With the development of CNN, neural networks are used in the classification of smoke images. The self-learning characteristics of CNN can find and learn the most suitable features for distinguishing smoke regions. In the CNN, feature extraction and classification recognition can be completed at the same time.

Experiments have proved that on the same smoke image data set, extracting the histogram of smoke orientation gradient (HOG) as a feature, using SVM as a classifier, the recognition rate of smoke images is only 82.26%, while the recognition rate of the CNN can reach 96.55%. To a certain extent, it reflects that the CNN's ability to classify images is much stronger than traditional methods, and the use of CNN for images classification eliminates the trouble of manual feature selection and feature extraction.

1.4 Introduction the Structure of Smoke Detection System in This Thesis

Different from traditional classifiers, convolutional neural networks include the functions of feature extraction and detection and recognition. For the convenience of make the system structure clear and facilitate the design of structured programming, this paper divides the system into four stages: input stage, preprocessing stage, recognition Stage, marking stage, the system framework diagram is shown in Figure 1-4. This system realizes the function of real-time video smoke detection. When using this system for smoke recognition, on the PC configured as described in section 5.1, On the training set described in section 4.2, after a short period of training, the specified video can be recognized.

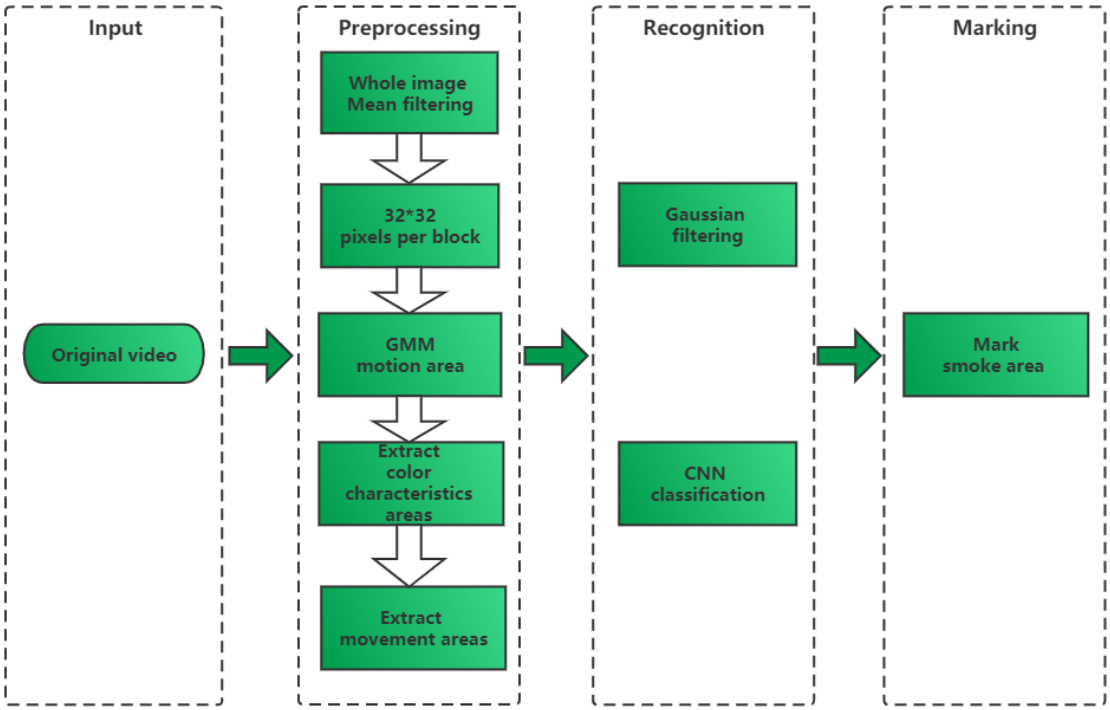


Figure 1-4: System framework diagram.

1）Read the video and decode it into a single frame image and input it into the system one by one,

2）Perform mean filtering on the input image, and then divide the image into regions according to the size of 32*24 pixels. Each region is a small image block; the Gaussian mixture model is used to extract the motion area in the entire video, and the motion area is included. Extract small image blocks, extract color features and motion features for each small image block, then use the image with both features in accordance with the smoke feature as the final image to be recognized,

3）The images to be recognized obtained in the preprocessing stage are processed by Gaussian filtering one by one, and then input into the trained convolutional neural network for classification and recognition,

4）Mark the position of the small image identified as smoke in the original video frame. The final output is as follows. The final output result is shown in Figure 1-5.

Figure 1-5: Schematic diagram of the final output.

1.5 The Structure of the Thesis and Main Work

This article summarizes some existing smoke recognition algorithms based on traditional methods. On this basis, an outdoor fire smoke detection method combining traditional methods and CNN is proposed, and a complete outdoor fire smoke detection system is completed. Compared with traditional methods, it proves that the CNN has a better effect in outdoor fire smoke recognition. The detailed structure of this thesis is as follows.

Chapter 1. This thesis introduces the research significance of fire smoke video recognition and the current research status at home and abroad, and briefly introduces the basic principles of fire smoke video detection technology and the basic structure of the system designed in this thesis.

Chapter 2. Introduced the source and sorting method of the smoke video data set used in the research.

Chapter 3. Introduces the pre-processing work that needs to be done before inputting the image into the CNN for classification, including filtering, extracting the smoke candidate area, and feature extraction. Through the comparison of different methods, the reasons for choosing the method used in this thesis are explained and given Comparison chart of the results.

Chapter 4. Introduced the basic principles of CNN and the smoke image recognition method based on CNN, compared the method of extracting HOG features of the image and using the SVM classification method to point out that the CNN has better recognition ability for smoke.

Chapter 5. Introduces the algorithm flow of the outdoor fire smoke detection system, verifies the effectiveness of the system through statistics and analysis of the four sets of comparative experimental results.

Chapter 6. Summarizes the advantages and disadvantages of the system, puts forward some ideas for algorithm improvement and some views on the prospects of smoke detection research.

CHAPTER 2

THE COMPILATION OF FIRE SMOKE VIDEO DATA SET

2.1 Collection of Data Sets

For image recognition research such as license plate recognition and face recognition, many institutions have established standard data sets during the research. However, due to the difficulty of obtaining smoke images and the relatively small number of researchers, smoke recognition is still no standard data set for research. The most data set used in the current research is the data set of the mobile phone of professors and students in the signal processing group of Bilkent University in Turkey[48], the official website of the Computer Vision and Pattern Recognition Laboratory of Keimyung University in Korea[49], In addition, the Machine Intelligence Laboratory of the University of Salerno in Italy established its own fire smoke video data set in 2012 [50], but there are still fewer people using this data set for research. From the University of Science and Technology in China, Professor Feiniu Yuan produced and published part of the data set, most of the video data sets be used are videos that recorded using burning objects outdoors to simulate smoke. The data set also includes tens of thousands of smoke and non-smoke images. In addition, other videos are mostly collected from YouTube and other video sites, and like the author of Reference [47], they make their own data sets. However, due to the uncertainty of outdoor fires and the constraints of conditions, it is difficult to make videos by self. Most of the data used in this thesis are collected from references. The data sources mainly include the following.

1) Download the public data set from the official website of the Signal Processing Group of Bilkent University in Turkey [48].

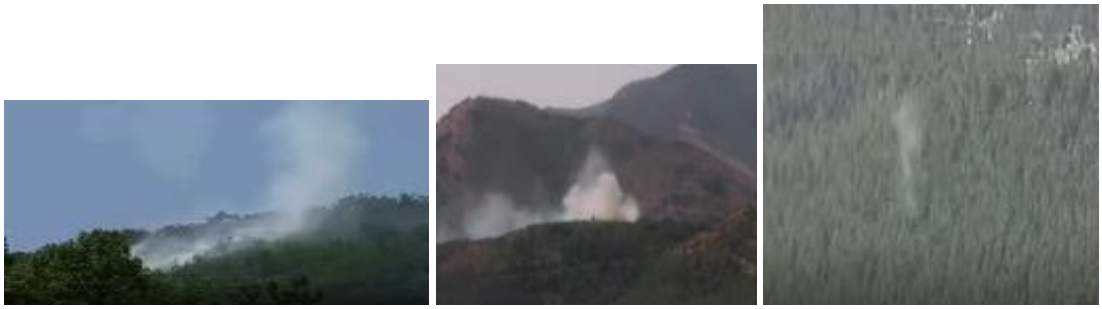
2) Download the public data set from the official website of the Computer Vision and Pattern Recognition Laboratory of Keimyung University in South Korea [49].

3) Download the public data set published by Professor Feiniu Yuan of University of Science and Technology in China.

4) Outdoor fire smoke data set from the Computer Vision Laboratory of the University of Nevada in Reno.

5) Part of the video collected by the video websites.

Some of the collected original video screenshots are shown in Figure 2-1. Can know that the size of the video is not consistent, and the quality of the video is also uneven. For the convenience of research, we will screen and crop the useful videos first, and then use the normalization method to sort out a set of video image data sets with uniform specifications and multiple environmental conditions.



(1) 16:9/720*405 (2) 4:3/640*480 (3) 1:1/240*240

(4) 4:3/640*480 (5) 4:3/640*480 (6) 16:9/1920*1080

Figure 2-1: The original video collected.

2.2 Organizing the Data Set

2.2.1 Screen

The collected video data sets are not all ideal videos that meet the requirements. For example, the image definition in Figure 2-1 (4) is too low, although the smoke area is visible to the naked eye, the noise in the video is too interference, not very researchable. Figure 2-1 (5) is a video in an indoor environment. The original video size of the image shown in Figure 2-1 (6) is 1920*1080 pixels, and the video size is too large, which is different from the current experimental conditions. In other words, the calculations are too large, difficult to meet real-time calculations, and the smoke area is small. If the video is scaled down, it will cause serious loss of the smoke area. Therefore, after collecting the data set, first perform the screening, only the ideal data set that meets the conditions is filtered out.

2.2.2 Crop

For videos that are too large, only the more experimentally significant parts are selected and cut. In Figure 2-2, the area in the lower left corner is where the smoke appears. The area in the lower left corner is reduced according to a rectangular area with an aspect ratio of 32 to 24 and saved as a new video clip. In addition, the upper left corner of the video is the floating cloud in the sky. In outdoor fire smoke recognition, the characteristics of cloud and smoke are very close, which is one of the important objects that interfere with recognition. Therefore, the part with clouds in the upper left corner is also cut out the rectangular section with an aspect ratio of 32 to 24. The other parts are mostly static backgrounds, and there is not much information displayed in the entire video, so it is not considered. In addition, for the convenience of normalization, the video is uniformly cropped to a video resolution of 320 pixels wide and 240 pixels high.

Figure 2-2: Schematic diagram of video cropping.

Regarding the crop size, the reason why a rectangle with an aspect ratio of 32:24 is selected during the cropping process is as follows.

Most video and monitors have an aspect ratio of 16:9 or 4:3, and 32:24 is 4:3. This size makes people looks comfortable.

32 and 24 are eight times of 4 and 3. Some algorithms in image processing often expand or shrink the highlight by multiples. 8 is the cubic of 2 can support the image to be reduced by multiples three times, which can meet the needs of the algorithm. However, the eight times of 16 and 9 are 144 and 72. Excessive values will cause the image size to be too large, which will increase the calculation during processing.

When cropping a video, the complete information in the original video should be retained, that is, when cropping the video, the object in the video must be completely cropped, not just a part of the object. When cropping a video, cannot just crop the smoke area, it must include the area around the smoke, especially some information that interferes with smoke recognition.

2.2.3 Normalized

Normalization is to change the video size to a uniform size through a scaling algorithm. Considering the calculation and video quality, the normalized uniform size is finally determined to be 320 pixels wide and 240 pixels high, which is convenient for dividing the video into $32 * 24$ image blocks in the subsequent processing. The division of the video reduce the calculation, is an effective means to improve the detection efficiency. In the collected original video set, there are some videos whose aspect ratio is not $32 * 24$, but the difference is not very large. After the image is scaled, the deformation has little effect on the recognition effect, and the original ratio of the video can be changed. The video size is uniformly $320 * 240$ pixels. The scaling algorithm used in this article to normalize the video is a bilinear interpolation algorithm.

2.2.4 Bilinear Interpolation Algorithm

The bilinear interpolation algorithm in the image scaling is an algorithm that has a better effect of scaling an image and a higher computational efficiency. Bilinear interpolation is to perform interpolation operations in both the x direction and the y direction. In the bilinear interpolation algorithm, the pixel of the target image is determined by its four points in the corresponding position of the original image, so the algorithm can better fill the missing pixels.

Before the interpolation operation, coordinate mapping must be performed first, assumption that $S(m, n)$ is a pixel point on the original image, $D(p, q)$ is the point corresponding to it on the scaled target image, then the relationship of S and D is.

$$p = \frac{SW * m}{DW} \tag{2-1}$$

$$q = \frac{SH * n}{DH} \tag{2-2}$$

Among them, SW is the width of the original image, SH is the height of the original image, DW is the width of the target image, and DH is the height of the target image. Let $i = [p], j = [q]$, then the value of $D(p, q)$ is determined by the four adjacent points $S(i, j), S(I, j + 1), S(i + 1, j)$, and $S(i + 1, j + 1)$, let $x = p, y = q, x_1 = i, x_2 = i + 1, y_1 = j, y_2 = j + 1$, then the interpolation calculation formula of D in the x direction is as follows.

$$S'(p, q) = \frac{x_2 - x}{x_2 - x_1} S(i, j) + \frac{x - x_1}{x_2 - x_1} S(i + 1, j) \tag{2-3}$$

$$S''(p, q) = \frac{x_2 - x}{x_2 - x_1} S(i, j + 1) + \frac{x - x_1}{x_2 - x_1} S(i + 1, j + 1) \tag{2-4}$$

The interpolation calculation formula in the y direction is as follows.

$$S(p, q) = \frac{y_2 - y}{y_2 - y_1} S'(p, q) + \frac{y - y_1}{y_2 - y_1} S''(p, q) \tag{2-5}$$

In the end.

$$S(p, q) = \frac{(x_2 - x)(y_2 - y)S(i, j)}{(x_2 - x_1)(y_2 - y_1)} + \frac{(x - x_1)(y_2 - y)S(i + 1, j)}{(x_2 - x_1)(y_2 - y_1)}$$

$$+ \frac{(x_2 - x)(y - y_1)S(i, j + 1)}{(x_2 - x_1)(y_2 - y_1)} + \frac{(x - x_1)(y - y_1)S(i + 1, j + 1)}{(x_2 - x_1)(y_2 - y_1)} \tag{2-6}$$

50*50 pixels 100*100 pixels 200*200 pixels

Figure 2-3: Bilinear interpolation method zoomed in the effect of the image.

The zooming effect diagram of the bilinear interpolation algorithm is shown in Figure 2-3, where the image in the middle is the original size of $100 * 100$ pixels, the image on the left is reduction to $50 * 50$ pixels after using the bilinear interpolation algorithm, and the image on the right side is magnification to $200 * 200$ pixels after using the bilinear interpolation algorithm. As show in the figure, the effect of using bilinear interpolation to process the image is better and can fill in the exact pixels, but for reduction images, jagged edges will appear, indicating the use of bilinear interpolation when the method is used to reduce the image, the problem of missing pixels cannot be avoided.

After screening, cropping, and normalization, 12 video clips with a size of $320 * 240$ pixels are finally produced, including 9 smoke videos and 3 non-smoke videos. The video set is shown in Figure 2-4. Among them, the short videos (1)-(3) are non-smoke videos, and the videos (4)-(12) are smoke videos. This video collections select smoke videos in a variety of outdoor environments, focuses on simulating a variety of outdoor environments through the experiment of this data set, the effectiveness of the system can be fully proved.

(1) (2) (3)

(4) (5) (6)

(7) (8) (9)

(10) (11) (12)

Figure 2-4: The final video set.

2.3 Chapter Summary

This chapter introduced the source of the data set in the process of this research, as well as the selection and tailoring of the data set for normalization. Finally, 12 video clips samples were produced as the research data set of this thesis. In the data set, the video of the occurrence of smoke under a variety of different conditions is included in the outdoor environments. It also contains some of the main factors that affect the identification of smoke in the outdoor environments. This data and research can fully prove the effectiveness of the algorithm in this paper. After the data set production and processing is completed, in the next chapter, will introduce how to preprocess the data set proposed in this thesis.

CHAPTER 3

PREPROCESSING

In the system designed in this thesis, the final input into the CNN for classification is small image blocks of $32 * 24$ pixels selected from the entire image. The preprocessing stage is to select these small image blocks from the entire image. This preprocessing stage selects the smoke area through four processes: filtering, extracting the motion area, extracting the color feature of the region of interest (ROI), and extracting the motion feature of the ROI. Through the preprocessing, it is avoided to input the entire image into the CNN, which reduces the amount of calculation of the system ensures the real-time performance of the system. At the same time, in the preprocessing stage, images that have been classified incorrectly can be filtered out, thereby improving the recognition accuracy of the system.

3.1 Filter

In the process of image collection, storage, and transmission, the quality of the image will be reduced by the interference and influence of various noise factors. The quality of the image will have a great impact on the subsequent segmentation, recognition, and recognition. The most common image noise is salt and pepper noise and gaussian noise. Salt and pepper noise include white salt noise and black pepper noise. The image of salt and pepper noise is shown in Figure 3-1 (2). The characteristic of salt and pepper noise is that the pixel depth of the noise is basically fixed, the position is random, and the mean filter can be used to remove it well. The image of gaussian noise is shown in Figure 3-1 (3). In contrast to salt and pepper noise, most pixels in gaussian noise have noise interference, and the pixel depth of the noise is uncertain. Gaussian filter eliminates gaussian noise can achieve better results. Mean filter and Gaussian filter are commonly used and effective filter methods in image processing. The following will compare the effects of these two filter methods in smoke detection applications, and finally determine the filter method used in this thesis.

(1) Original image (2) Salt and Pepper noise (3) Gaussian noise

Figure 3-1: Noise images comparison.

3.1.1 Mean Filter

Mean filter is a linear filter method, which has a relatively good effect on eliminating salt and pepper noise, is a common filter method with low computational complexity. Mean filter uses a mean filter template, which slides the template along the x-axis and y-axis in sequence on the original image, calculates the mean value of all pixels within the coverage of the template, and replaces the value of the center of the position covered by the template with this value filter out the noise. Since salt and pepper noise is often independent points, in this way, using the mean value of the surrounding points of the pixel to determine the value of the original pixel, the salt and pepper noise can be removed very well. However, because this method simply uses surrounding pixels the mean value of the point replaces the center point, does not consideration the excessive relationship between the pixels. Therefore, the Mean filter is not the best filter method, but due to its simple operation, when the filter effect is not very important, the mean filter is a very practical filter method.

The Mean filter template M_5 with a radius of 5 is as follows (3-1).

$$M_5 = \frac{1}{25}\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \tag{3-1}$$

Calculating the mean value of pixels can be done by means of image convolution. The formula for calculating the filtered image is

$$I_m = I * M_5 = \sum_{i=0}^{height} \sum_{j=0}^{width} (I(x+j, y+i) M_5(j,i)) \tag{3-2}$$

Where I_m represents the filtered image, I represents the original image, $height$ represents the height of the average filter template, $width$ represents the width of the average filter template, and x and y represent the offset of the current sliding window relative to the origin of the original image.

3.1.2 Gaussian Filter

Gaussian filter is also a linear filter method, which has a good effect on eliminating gaussian noise and is widely used in image denoising. Gaussian filter uses a filter template to perform a weighted average on each pixel of the entire image, and each weight in the filter template meets the gaussian distribution.

In the implementation of Gaussian filter, the image is gaussian blurred by image convolution. The convolution kernel slides along the original x and y directions. The convolution kernel and the image pixel values covered by it are convolved, and then used the convolution value replaces the pixel value at the center of the template. Assuming I is the original gray-scale image, G_σ is the convolution kernel or filter template, and I_σ is the image after convolution (filter), then I_σ can be express as

$$I_\sigma = I * G_\sigma \tag{3-3}$$

Among them, $*$ means convolution operation; G_σ is a two-dimensional gaussian kernel with standard deviation σ, which is defined as.

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{\frac{-(x-\mu_x)^2}{2\sigma_x^2} + \frac{-(y-\mu_y)^2}{2\sigma_y^2}} \tag{3-4}$$

Among them, μ is the mean value, σ is the variance, and the variables x and y represent the x-axis and y-axis of the two-dimensional image. Assuming that the coordinates of the center point is $(0,0)$, that is taking the center point as the origin, the equation (3-4) can be simplified to.

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{\frac{-(x^2+y^2)}{2\sigma^2}} \tag{3-5}$$

Therefore, the Gaussian filter template G_5 with a radius of 5 is as follows.

$$G_5 = \frac{1}{273}\begin{bmatrix} 1 & 4 & 7 & 4 & 1 \\ 4 & 16 & 26 & 16 & 4 \\ 7 & 26 & 41 & 26 & 7 \\ 4 & 16 & 26 & 16 & 4 \\ 1 & 4 & 7 & 4 & 1 \end{bmatrix} \tag{3-6}$$

Finally, the original image I and the Gaussian filter template G_σ are convolved according to the following formula (3-7) to obtain the final Gaussian filter result.

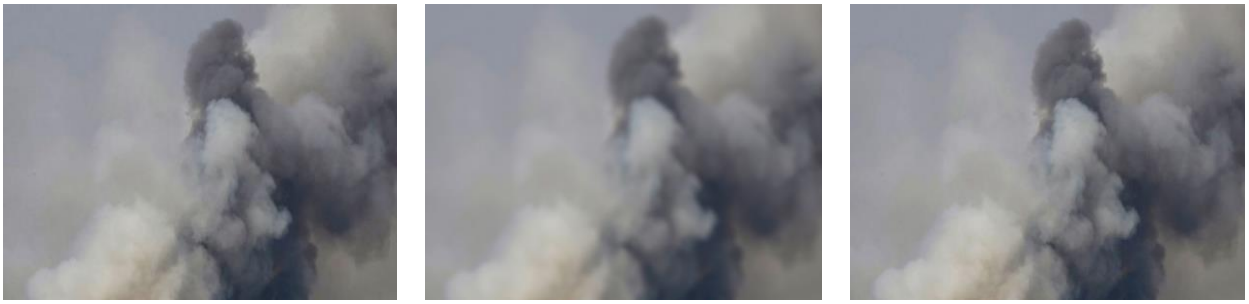
$$I_\sigma = I * G_\sigma = \sum_{i=0}^{height} \sum_{j=0}^{width} (I(x+j, j+j) G_\sigma(j, i)) \tag{3-7}$$

Among them, I_σ represents the filtered image, I represents the original image, G_σ represents the Gaussian filter template, $height$ is the height of the Gaussian filter template, $width$ is the width of the Gaussian filter template, and x and y are the deviations of the current sliding window relative to the original image origin shift.

In the process of Gaussian filter and average, there is a problem of processing image edge pixels. Since the center of the convolution kernel cannot be slid to the edge of the image when the image is convolved, the pixels at the edge of the image cannot be filtered. There are two common ways to solve this problem, which is to ignore the edge pixels and cannot perform filter operations. There are two commonly used ways to solve this problem. One is to directly ignore the edge pixels. After the filter is completed, add a layer of random values or original edge pixels on the outer edge of the image. Add a layer of random values or pixels at the edge of the original image. Another way is adjunction a layer of 0 or other data on the periphery of the original image, and then convolve the image so that the image after convolution is as large as the original image. Since there are many pixels in the image, and the amount of information contained in the pixels at the edge of the image is not large, the loss caused by the two processing methods to the image is acceptable.

3.1.3 Experiment and Compare

The experimental effects of Mean filter and Gaussian filter are shown in Table 3-1. Figures 3-2 and 3-3 show the filter effects of images with a size of $320 * 240$ pixels and $32 * 24$ pixels respectively. Figure (1) is the original image, Figure (2) is the effect diagram of Mean filter, and Figure (3) is the effect diagram of Gaussian filter.



(1) Original image (2) Mean filter (3) Gaussian filter

Figure 3-2: The filter effect of 320*240 pixels.

(1) Original image (2) Mean filter (3) Gaussian filter

Figure 3-3: The filter effect of 32*24 pixels.

It can be seen from Table 3-1 that, from the time consumed by the two filter methods, the time consumed by Gaussian filter is shorter than the time consumed by Mean filter, but the time consumed is only a few microseconds. The effect of the two filter methods is not large, from the effect of the two filter methods, the naked eye cannot judge which method is more suitable. After many tests during the experiment, it was found that before processing the video, a Mean filter with a kernel size of $5 * 5$ was used to filter the original image with a size of $320 * 240$ pixels, and then the kernel size was used in the recognition stage. The $5 * 5$ Gaussian filter performs Gaussian filter on the proposed target area with a size of $32 * 24$ pixels. This method has a better effect. Therefore, it is finally decided to use the Mean filter to perform the mean filter processing on the entire image in the preprocessing stage, and to use the Gaussian filter to perform the gaussian filter on the proposed target area in the recognition stage.

Table 3-1: Comparison results of mean filter and Gaussian filter.

Filter method	Nuclear radius	Image size	Filter effect	Time
Mean filter	$5 * 5$	$320 * 240$px	Figure3-2(2)	2.17ms
Gaussian filter	$5 * 5$	$320 * 240$px	Figure3-2(3)	0.81ms
Mean filter	$5 * 5$	$320 * 240$px	Figure3-3(2)	0.21ms
Gaussian filter	$5 * 5$	$320 * 240$px	Figure3-3(3)	0.13ms

For prove the influence of image filter on image recognition, the images with Gaussian filter and without Gaussian filter were used as input data. On the same training set and test set, the same structure of CNN was used for test. The experimental results are as show in Table 3-2.

Table 3-2: Comparison of the influence of image filtering on the recognition rate of CNN.

Filter method	Nuclear radius	Classifier	Image content	Recognition rate
Gaussian filter	5	CNN	Smoke and non-smoke	96.55%
Unfiltered	/	CNN	Smoke and non-smoke	94%

The CNN and data set used in the experiment here are consistent with those introduced in Section 4.2, and the recognition rate is the recognition rate of the CNN in the test data set. The results show that after the input image is processed by Gaussian filter, the recognition rate is 96.55%, and the recognition rate without processed by Gaussian filter is 94%, indicating that filter has a certain influence on the recognition rate of the CNN. At the same time, it is found in experiments that if the image input to the network is not filtered, the CNN is prone to overfitting during the training process, which causes the network to fail and cannot effectively determine the image type.

3.1.4 Summary

This section introduced the two filter algorithms of Mean filter and Gaussian filter. After comparative experiments and actual tests, comprehensively considering the influence of Mean filter and Gaussian filter on the final recognition effect of the system, it is finally determined to use Mean filter method to filter the entire image, use Gaussian filter as a method for filter candidate regions. A simple classification test illustrates the importance of filter in image processing. In the next chapter, will introduce how to extract the smoke ROI. The ROI is the image area initially extracted from the entire image, indicating that the area has a certain possibility of being a smoke area. The ROI is selected to reduce the input during classification. improve the efficiency of the system.

3.2 Extract the Motion Area

Because the computational complexity of image processing algorithms is relatively large, and the image data is two-dimensional, every time the image increases, the amount of calculation increases in exponential growth. However, video processing requires real-time performance. Generally, the ROI in the video only occupies a small part of the entire video image area, so for facilitate rapid processing of video data, can first extract the ROI in the video, and then perform further processing on the ROI.

In the research process of this thesis, found that after the appearance of smoke, due to the fluidity of the air, the smoke will always be in a moving state. According to this characteristic, the moving area in the video can be extracted first, but the moving area is not necessarily all smoke area. The area may also be moving cars, white cloud and smog, human and other objects, so after extracting the moving area, further processing is needed to determine whether the area is a smoke area.

Commonly used methods for extracting motion regions are optical flow method, inter-frame difference (IFD) method, Gaussian mixture model (GMM) method, etc. Due to the relatively large calculation in the optical flow method, this method is directly abandoned, and the method of extracting the motion area is selected in two methods: the IFD and the GMM method.

3.2.1 Inter-frame Difference Method to Extract Motion Regions

The basic idea of the inter-frame difference method for extracting the motion area is to compare the difference between the two consecutive images (or a few frames in between) in the video. If the difference is greater than a certain threshold, it is regarded as moving in the video part, and the difference is regarded as the movement area.

The principle of the IFD method can be expressed as the formula (3-8).

$$D(x,y) = \begin{cases} 1, & if\,|I_t - I_{t-1}| > T \\ 0, & others \end{cases} \tag{3-8}$$

Among them, D(x, y) represent the difference image of the two images before and after, 1 represent white, represent the motion area, 0 represent black, the non-motion area, I_t and I_{t-1} represents the original image at time t and $(t-1)$respectively, and T represent the threshold value, that is When the number of non-zero points in the image is greater than T, it means that the motion area is acquired, and the difference image is binarized, then the white area in the binarized image represents the motion area.

The advantage of the IFD method to extract the motion area is that it is simple in operation, sensitive to the moving part and has good adaptability to external interference factors such as light. However, the

disadvantage is that it cannot accurately extract the position of the motion area and cannot completely extract it, and the target area is prone to voids.

3.2.2 Gaussian Mixture Model to Extract Motion Regions

The GMM is a commonly used method to establish a background image model in the research of background filter. The pixel value of the background image is generally not really fixed. The reasons for the change of the background pixel value include two categories.

1) Object movement. Including the movement of objects in the background image, such as the movement of branches blown by the wind, the movement of clouds, the movement of cars, the movement of humans, etc., as well as the changes caused by the shaking of the camera itself.

2) Changes in brightness. In a still image, even if the objects in the image are not moving, changes in the light in the external environment can also cause changes in the background pixel values, such as direct sunlight, flashing lights, and other objects blocking light.

The pixel changes caused by these changes generally make small changes near the original pixel value, and the distribution is very close to the gaussian distribution, so it is very suitable to use the gaussian distribution to model the distribution of background pixel values. However, in many cases, the distribution of pixel values does not just change around one value, but changes around multiple values, such as shaking leaves, moving cars, or shadow changes caused by the movement of light happening, etc. Therefore, it is a reasonable way to use multiple gaussian distributions to model the background.

The basic idea of the GMM is to select K gaussian distributions, each gaussian distribution is called a component, and these components are linearly added to form the probability density function of GMM, as shown in (3-9).

$$\mathrm{P}(x) = \sum_{\kappa=1}^{\mathrm{K}} \pi_{\kappa} \mathcal{N}(x|\mu_{\kappa}, \sum_{\kappa}) \tag{3-9}$$

Among them, K represent the number of gaussian distributions, $\mathcal{N}(x|\mu_{\kappa}, \Sigma_{\mathrm{K}})$ represent the multivariate gaussian distribution, π_{κ} represent the mixed weight value (mixture coefficient), and π_{κ} satisfies$0 \leqq \pi_{\kappa} \leqq 1 \; and \; \sum_{\kappa=1}^{K} \pi_{\kappa} = 1$.

Calculate the probability distribution of GMM using the pixel values of the first N frame of images collection. Now that the probability density function is knowns, the parameters can be determined by maximum likelihood estimation. The likelihood function of GMM is as show in (3-10).

$$\sum_{i=1}^{k} \log\left\{\sum_{i=1}^{k} \pi_k N\,(X|\mu_k, \sum_k)\right\} \tag{3-10}$$

Finally, use the Expectation Maximum Algorithm (EM) to calculate the parameters.

$$\gamma(i,k) = \frac{\pi_k N(X_i|\mu_k, \sum_k)}{\sum_{j=1}^{k} \pi_j N(X_i|\mu_j, \sum_{kj})} \tag{3-11}$$

$$N_k = \sum_{i=1}^{N} \gamma(i,k) \tag{3-12}$$

$$\pi_k = \frac{N_k}{N} \tag{3-13}$$

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^{N} \gamma(i,k) X_i \tag{3-14}$$

$$\sum_k = \frac{1}{N_k} \sum_{i=1}^{N} \gamma(i,k)(X_i - \mu_k)(X_i - \mu_k)^T \tag{3-15}$$

In the formula (3-11), $\gamma(i,k)$ represent the probability that the i data has the k component.

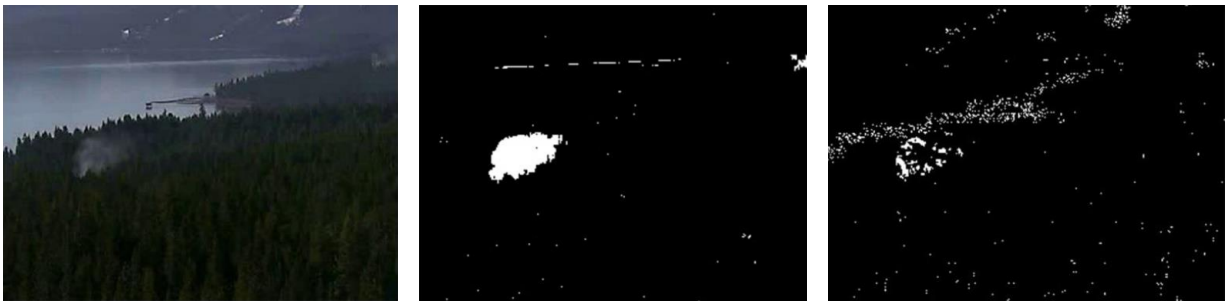
Starting from the $N+1$ frame, it is determined whether each pixel in the image matches the constructed k Gaussian model. If the matching is unsuccessful, it is determined as the front scenic spot, otherwise it is the background point.

3.2.3 Experiment and Compare

The effect of extracting background between frame difference method and GMM method is shown in the figure, where image (1) is the original image, image (2) is the motion area extracted by GMM method, and image (3) is the motion area extracted by IFD method.

From the comparison result, can get that the GMM can extract the motion area more accurately and completely, while the motion area extracted by the IFD method has holes and omissions. In addition, due

to the occurrence of leaf shaking in the wild environment, such as the regular and continuous movement of leaf shaking, for the inter-frame difference algorithm, each leaf shaking will be detected as a motion area, but like this regular and persistent movement is not the target area. For the GMM, a regular and continuous motion area like leaves, when a stable gaussian distribution is formed, the motion area extraction method based on the GMM will treat this area as a part of the background area, thus avoiding the interference of similar situations. The extracted background area will inevitably have some noises. These noises can be removed by mathematical morphological methods. For example, using "open operation" can remove isolated points in the image, and "close operation" can fill in small-scale holes in the image.



(1) (2) (3)

Figure 3-4: The motion area extracted by the GMM method (2) and the IFD method (3).

Because the location of the smoke is generally fixed in the video, when the smoke continues for a period of times, the GMM can easily determine the smoke area as a background image, which will cause the smoke area to be missed. The test found that it can pass reasonable increase the number of historical frames for gaussian modeling to alleviate this problem.

3.2.4 Summary

This section introduced how to use the algorithm model to extract the motion area in the video image, comprehensively analyzes the advantages and disadvantages of the optical flow method, the inter-frame difference method and the GMM method in extracting the motion area, the optical flow is complicated to calculate and is not suitable for real-time monitoring systems, the IFD method is simple to calculate, but the detection results are incomplete, the background subtraction effect is good, but the establishment of a good background model requires a lot of calculation and storage overhead. And finally determines the GMM through experiments, have better robustness, so the GMM is used as the algorithm to extract the motion area. In the next chapter, will analyze the color characteristics and motion characteristics of the smoke, and further filter the extracted motion regions.

3.3 Color Feature Analysis and Extraction of Smoke Area

The characteristics of smoke include color characteristics and motion characteristics, diffusion characteristics, wavelet energy, image entropy, etc. The color of the smoke produced by different burning materials is also different. For example, when petrochemical dyes are burned, black smoke will be emitted. After investigation and research in this thesis, the color of early fire smoke is mainly gray and white. In addition, when the smoke is generated, it is caused by burning. As the air rises, the cold air falls to form air convection, and the smoke generally has a tendency move to upwards. In this thesis, color and motion features are mainly selected as features for detecting smoke candidate regions.

3.3.1 Features in the RGB Color Space

In the RGB color model, the different superpositions of the three colors of R/G/B can present a variety of different colors, and the value range of each color component is [0,255], 0 means that the color occupies the proportion of is the smallest, 255 means the proportion of the color is the largest, (255,0,0) means red, (0,0,0) means black, and (255,255,255) means white. The RGB color space uses the RGB color model.

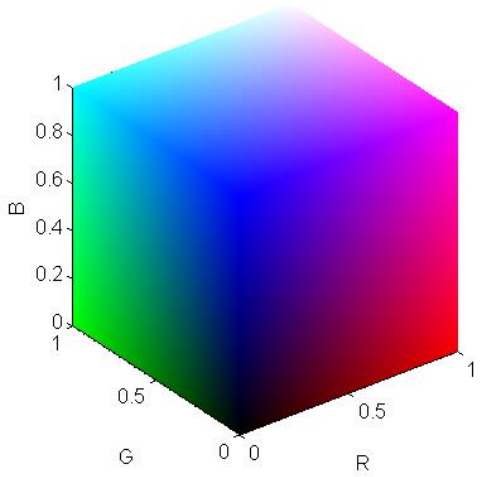


Figure 3-5: RGB color model.

Smoke images can find some characteristics in the RGB color space. In the reference [57], the color of the smoke proposed the generally presents the characteristics close to the grayscale image. The color of the smoke can be divided into light gray and dark gray according to the gray level, which also indicates the R, G, B of the smoke image, the values of the three components are equal or close. Thus, the algorithm of smoke image in RGB color space is proposed, as shown in formula (3-16).

Rule1:

$$R + \alpha = G + \alpha = B + \alpha$$

Rule2:

$$L_1 \leq I \leq L_2 \tag{3-16}$$

Rule3:

$$D_1 \leq I \leq D_2$$

If (rule1) AND [(rule2) OR (rule3)] = TRUE

Then smoke pixel

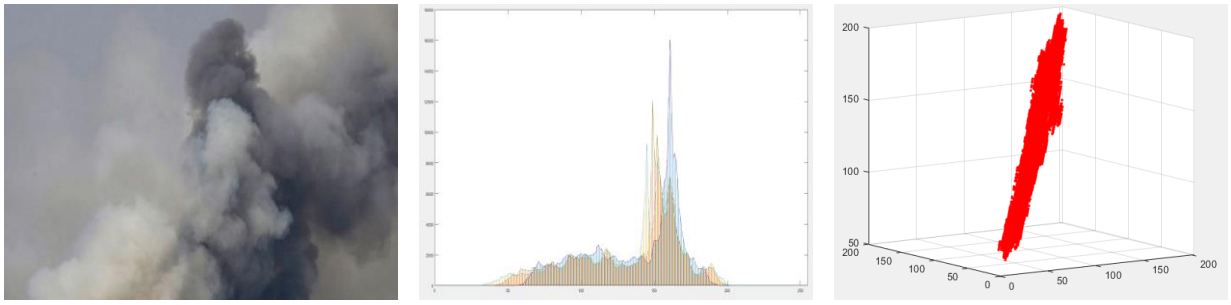
Else

not smoke pixel

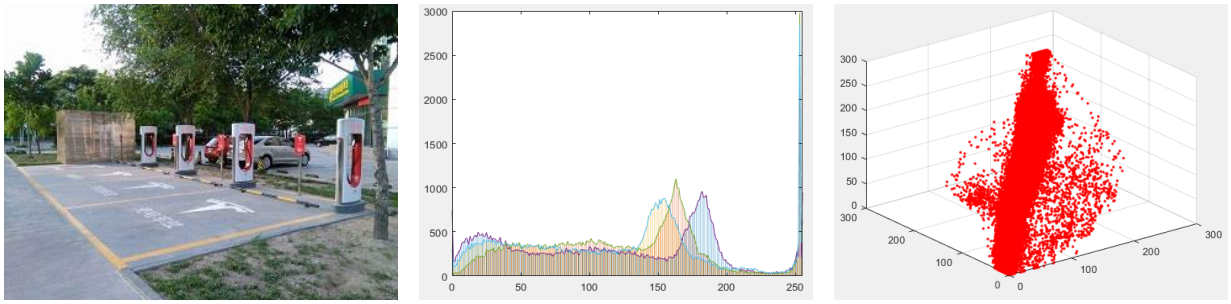
The value of α is 15~20, L_1=150, L_2=220, D_1=80, D_2=150. When the smoke is light gray, rule 2 is used, and when the smoke is dark gray, rule 3 is used.

This theory has a better recognition effect on smoke, but in outdoor smoke videos, due to the long shooting distance and low camera pixels, the image is blurred. In actual situations, non-smoke areas will also appear the color is close to the characteristics of the grayscale image, that is, the values of the three components of R, G, and B are close. This thesis did the following experiment to compare the characteristics of the smoke area and non-smoke area in the RGB color space in the outdoor fire smoke videos.

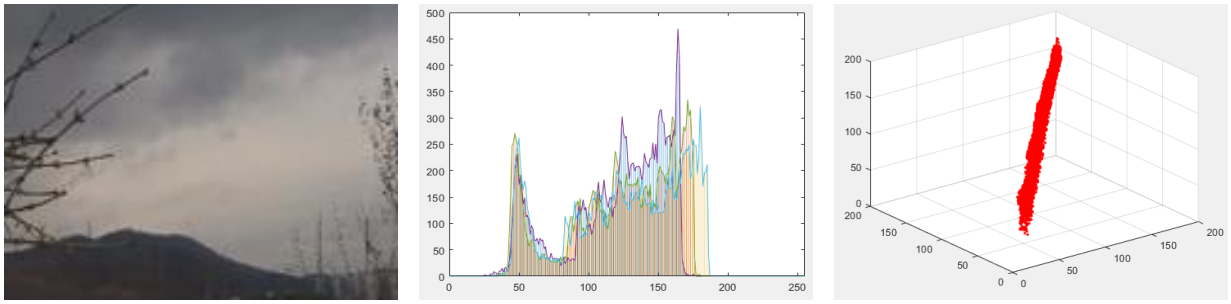
The above four groups of images are smoke images (Figure 3-6 (1)), non-smoke images in close-up (Figure 3-6 (2)), outdoor non-smoke images (Figure 3-6 (3)), and outdoor non-smoke images (Figure 3-6 (3)). The images after the smoke appears in the smoke area (Figure 3-6 (4)). The second image of each group is the statistics of the number of pixels in the R, G, and B color spaces of the first image. The horizontal axis represents all the pixels from the upper left corner to the lower right corner of the image. The vertical axis represents the size of the pixel value. The red, green, and blue lines respectively represent the three channels of each pixel in the image. The third image in each group is the representation of pixels in the RGB three-dimensional space. Through the analysis of the above chart, the following conclusions can be drawn.



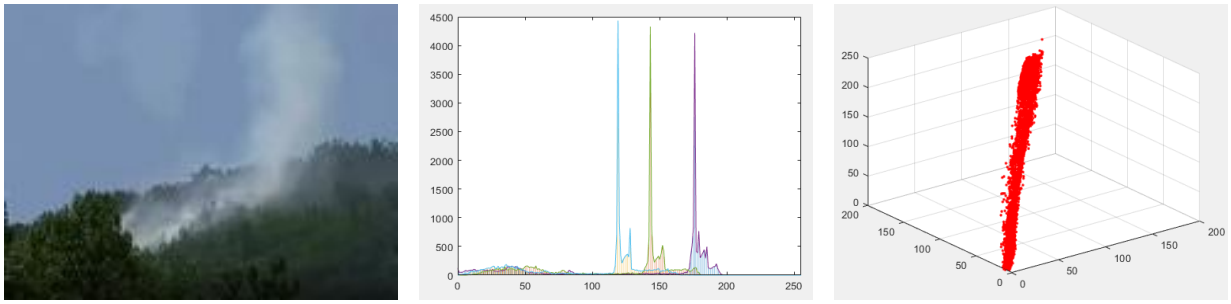
(1)



(2)



(3)



(4)

Figure 3-6: Four comparison images and the two-dimensional and three-dimensional images of the R, G, B components of each images.

The above four groups of images are smoke images (Figure 3-6 (1)), non-smoke images in close-up (Figure 3-6 (2)), outdoor non-smoke images (Figure 3-6 (3)), and the image after the smoke appears in the smoke

area (Figure 3-6 (4)). The second image of each group is the statistics of the number of pixels in the R, G, and B color spaces of the first image, the horizontal axis represents all the pixels from the upper left corner to the lower right corner of the image, the vertical axis represents the size of the pixel value, the red, green, and blue lines respectively represent the three channels of each pixel in the image. The third image in each group is the representation of pixels in the RGB three-dimensional space. Through the analysis of the above chart, the following conclusions can be drawn.

1) For the smoke images (Figure 3-6(1)), from the RGB two-dimensional image, the spacing between the three curves is basically the same, and the difference is not very large. From the RGB three-dimensional image, the pixels are evenly distributed the non-smoke images (Figure 3-6(2)) in the close-range position is centered on the diagonal, and the non-smoke images taken at a closer distance. The color of the image is better and clearer. There are three in the RGB two-dimensional image. The interval of the curve is relatively uneven, and the distribution of the pixels in the RGB three-dimensional image is not concentrated near the diagonal. In summary, the characteristics of smoke and fog area conform to the rules of literature [55].

2) For outdoor non-smoke images (Figure 3-6 (3)), in the RGB two-dimensional image, the spacing between the three bars is not uniform. From the RGB three-dimensional image, can get that the pixels are evenly distributed at the diagonal center. Indicating that the image is close to a grayscale image. But the difference between the three RGB channels is not consistent, that is, it does not satisfy the formula $R + \alpha = G + \alpha = B + \alpha$.

3) The images after the appearance of smoke in the outdoor non-smoke area (Figure 3-6 (4)), the RGB two-dimensional image and the RGB three-dimensional image show almost the same characteristics as the outdoor non-smoke images (Figure 3-6 (3)).

In summary, the RGB color space is not obvious for outdoor smoke characteristics, so the method of finding the characteristics of the smoke area in the RGB color space is abandoned. However, in the HSV color space, H represents the hue, which reflects the color of the image, S represents the saturation, which reflects the vividness of the image color, and V represents the value, which reflects the brightness of the image. Because of the appearance of smoke will make the image blurry, and the smoke area is generally white, it is reflected in the HSV color model that the saturation S is relatively low, and the value V will be relatively increased before and after the smoke appear. Reference [58] proposes that smoke will reduce the saturation of the background. Considering that the smoke cannot cover the background when the smoke is thin, the translucent characteristic of the smoke has this effect, but when the smoke is thick, it can completely cover the background area, so in this thesis chooses the smoke characteristics of is mainly for the color characteristics of the smoke, rather than the effect of the smoke on the background.

3.3.2 Features in the HSV Color Space

The HSV color space can express the effect of a specific style by adjusting the saturation and brightness. The HSV color model is closer to the way the human eye perceives colors. RGB is an additive primary color model that presents different colors in a combination of three primary colors. HSV is based on " What color? How is the shade? How is the light and dark?" The way encapsulates the information about the color. The color model of HSV can be represented by the cone model shown in Figure 3-7.

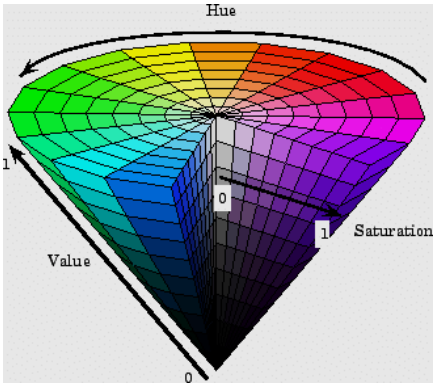


Figure 3-7: HSV cone model diagram.

The meaning of HSV cone model can be illustrated with the help of Figure 3-8:

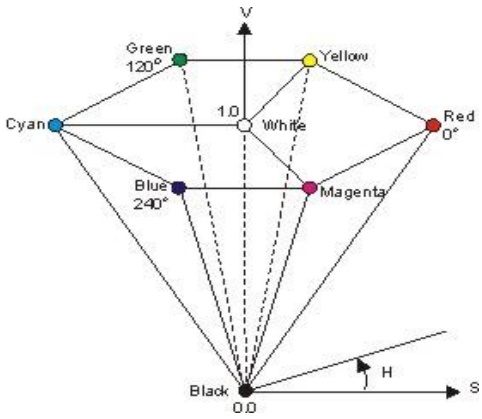


Figure 3-8: The meaning of HSV cone model.

1) The rotation angle at the top of the cone represents the size of the hue(H). The hue ranges from 0° to 360°, 0° represent red. Start from 0°, rotate in a counterclockwise direction, 120° is green and 240° is blue.

2) The radius of the circle at the top of the cone indicates the degree of saturation(S). The value of saturation ranges from 0 to 1, the value of the center position is 0, and the value of the edge position is 1. Each color can be regarded as a color that is a mixture of the spectrum color and white, and the saturation is reflected,

it is the closeness of a color to the spectral color. Greater saturation means greater proportion of the color spectrum and the more vivid of the color. When the saturation is 0, it means that the proportion of spectral colors in the color is 0%, that is, black and white, when the saturation is 1, it means that the proportion of spectral colors in this color is 100%.

3) The height of the cone indicates the degree of value (V). The value is also 0 to 1, value reflects the lightness and darkness of the color, when the value of the color is 0 (0%), that is, at the sharp corner of the cone, where the value of the color is the lowest, it appears as black, when the value of the color is 1 (100%), the color is brighter, and if the saturation is 0, it appears as white.

However, images are stored and displayed by RGB in the computer, and the stored pixel values represent the components of the three primary colors. If want to process the image in the HSV color space, need to convert the image from the RGB color space to the HSV color space. The conversion method of pixel value from RGB color space to HSV color space is as follows.

$$V = \max(R, G, B) \tag{3-17}$$

$$S = \begin{cases} \dfrac{V - \min(R, G, B)}{V}, & if\ V \neq 0 \\ 0 \quad , & otherwise \end{cases} \tag{3-18}$$

$$H = \begin{cases} \dfrac{60 * (G - B)}{(V - mim(R, G, B))}, & if\ V = R \\ 120 + \dfrac{60 * (B - R)}{(V - \min(R, G, B))}, & if\ V = G \\ 240 + \dfrac{60 * (R - G)}{(V - \min(R, G, B))}, & if\ V = B \end{cases} \tag{3-19}$$

After calculating the result.

$$H = \begin{cases} H + 360, & if\ H < 0 \\ H \quad , & if\ H \geq 0 \end{cases} \tag{3-20}$$

The final calculated result, $H \in [0^{\circ}, 360^{\circ}], S \in [0, 1], V \in [0, 1]$.

Since the image in the computer is stored in 24-bit format, each component of the pixel value should be a value from 0 to 255. For the convenience of calculation, the calculated values of the three components of H, S and V are mapped to from 0 to 255, the mapping formula is as follows.

$$H = \frac{H * 255}{360} \tag{3-21}$$

$$S = 255 * S \tag{3-22}$$

$$V = 255 * V \tag{3-23}$$

Because in the computer, the pixel values of all images are treated as values in the RGB space, so the image displayed on the monitor after converting the image to the HSV color space (Figure3-9(2)) will be quite different from the original image (Figure3-9(1)).

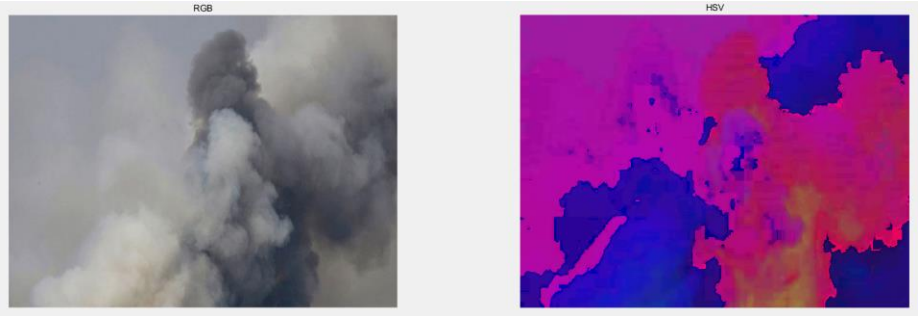


Figure 3-9: The comparison between the image converted to HSV space and the original image.

After converting the images from the RGB color space to the HSV color space, the characteristics of the smoke image in the HSV color space are analyzed in a similar way to the analysis of the RGB color space characteristics. As show in figure 3-10, take this image as an example to introduce and analyze the characteristics of smoke image in HSV color space.

Figure 3-10: Smoke image.

1) Map all pixels of the smoke image to the HSV three-dimensional space, where x, y and z represent the three components of H, S and V respectively. For convenience, the values of H, S, V are mapped to $0 \sim 255$. Within the range, the results shown in Figure 3-11. However, because the number of pixels in the image is

large, no rule can be seen from the image. Therefore, the points in the three-dimensional space are composed of x, y and z. Projected on two planes to generate three two-dimensional images in a pairwise combination of H, S and V.

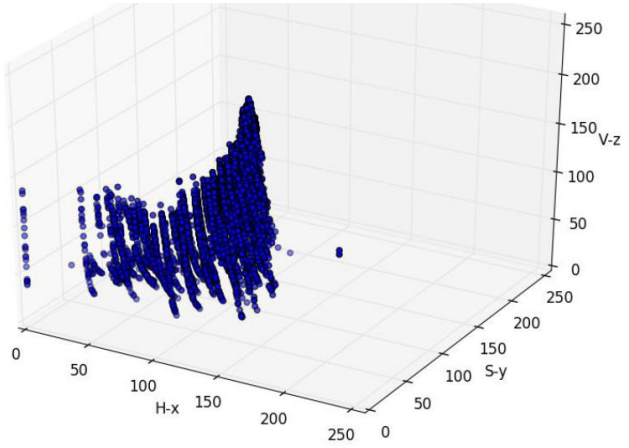


Figure 3-11: HSV three-dimensional image.

2) Figure 3-12 is the two-dimensional image obtained by projecting the HSV three-dimensional image onto the plane composed of the S and V axes. S is the X axis and V is the Y axis. It can be seen from the image that the pixels are mostly concentrated in the position where the S value is small. And the value of V is widely distributed and has no obvious characteristics.

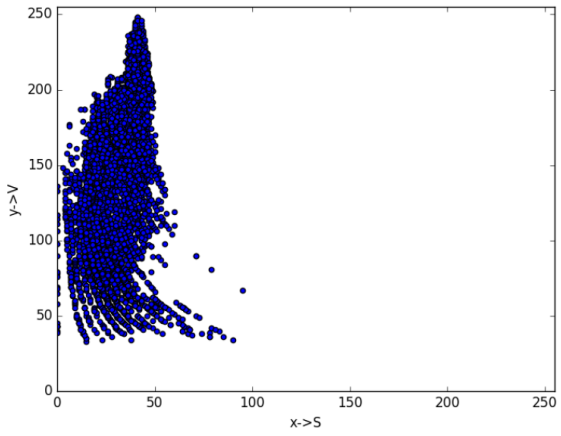


Figure 3-12: S-V projection image.

3) Figure 3-13 is the two-dimensional image obtained by projecting the HSV three-dimensional image onto the plane composed of S and H axes, where S is the x-axis and H is the y-axis, the image still shows that most of the pixels are concentrated in the smaller S value Location, and the value distribution of H is relatively scattered and has no obvious characteristics.

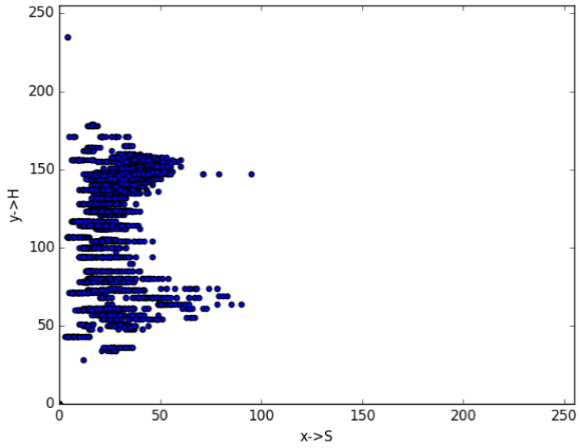


Figure 3-13: S-H projection image.

4) Figure 3-14 is a two-dimensional image obtained by projecting the HSV three-dimensional image onto the plane composed of the V and H axes, where V is the x-axis and H is the y-axis, the values of H and V in the image do not show obvious characteristics.

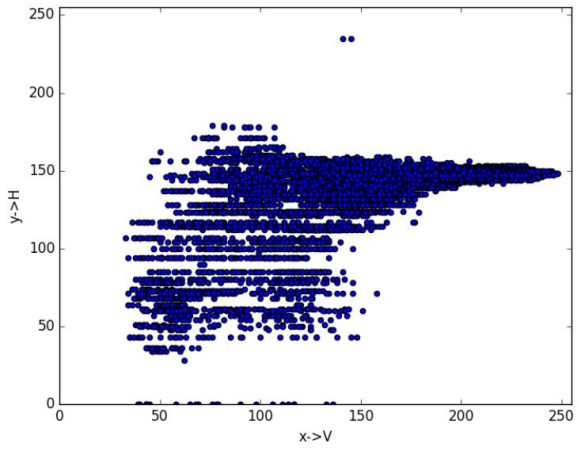


Figure 3-14: V-H projection image.

After analyzing large number of images in the HSV color space, final conclusions are the following three points.

1) In the HSV color space, the saturation(S) of the smoke image area is generally low. In this research, the final S is set to 70 after the experiment, that is, the saturation of the smoke area is generally lower than 70.

2) When smoke appears, the value(V) of the area will increase.

3) In this research, the average value V of 50 frames of the current video frame are use as the threshold. If the average value V of the area is greater than the threshold and the saturation S is lower than 70, the area considered to be a ROI area for smoke, and then the area will Enter to next step of motion feature analysis.

3.3.3 Experiment and Compare

The smoke area extracted based on the features in the RGB and HSV color space is shown in Figure 3-15, Figure (2) is the smoke area based on the feature extraction of the RGB color space, and Figure (3) is the feature extraction based on the HSV color space Area of smoke.

(1) original (2) RGB (3) HSV

Figure 3-15: Smoke area extraction results based on RGB (2) and HSV (3) color spaces.

It can be seen from the image that the target area extracted based on the features of the RGB color space does not cover all the smoke area, while the target area extracted based on the feature of the HSV color space covers most of the smoke area. Can know that the features of smoke in the RGB color space proposed in reference [57] are not very suitable for smoke videos in outdoor environments, and the features based on the HSV color space meet the requirements. Based on the area, combined with the motion characteristics of the smoke area, the smoke ROI area can be better extracted.

3.3.4 Summary

This section introduced the color characteristics of the smoke area. Color characteristics are the key smoke characteristics researched in this thesis. Use of color characteristics for smoke recognition is also a common method in early smoke recognition. However, the research found that in the RGB color space commonly used by human, the color characteristics of smoke are not ideal for outdoor fire smoke. Further research has found that the use of smoke in the HSV color space has a better recognition effect. This section gives the method of feature analysis and the comparison of recognition effects, which proves the HSV color the validity of features. In the next section, the analysis and extraction of smoke motion features will be introduction.

3.4 Analysis and Extraction of Motion Feature in Smoke Area

Because smoke is produced by combustion, the high temperature generated when the fuel burns will cause the air to form convection, causing the hot air with smoke to rise, so after the smoke is produced, there will generally be an upward movement trend, which is also the difference between smoke and clouds. The main characteristics of cloud and fog.

3.4.1 Motion Direction Detection Based on Optical Flow Method

Optical flow is about the concept of objects in the field of view and motion detection. It is a concept proposed by American physicist James J. Gibson in 1940 when he described the motion of objects to cause visual stimuli to animals. Optical flow is a two-dimensional vector field. By detecting the intensity changes of pixels in the image over time, the moving speed and direction of the object can be calculated, and a point can be displayed from one frame of image to the next frame between movement. As shown in Figure 3-16, it shows the movement of a point in 5 frames of continuous images, and the arrow represents the optical flow field vector. Optical flow plays a very important role in computer vision and image processing. It can be used for motion detection, image segmentation, motion compensation coding, video compression, etc.

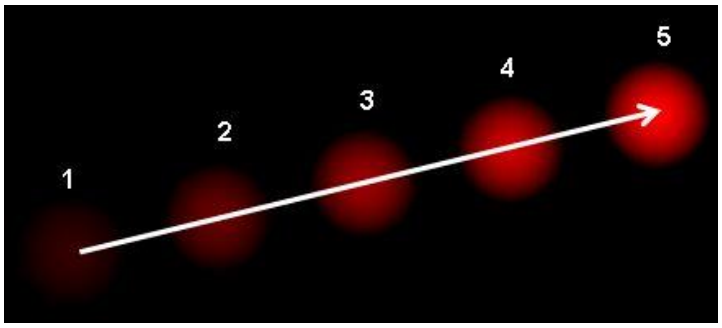


Figure 3-16: Point motion image in the optical flow field.

The optical flow method is based on the following three assumptions.

1) The brightness is constant. The detection target remains unchanged in appearance in the images of the previous frame and the next frame. For grayscale images, the gray level of the target image does not change.

2) The image motion amplitude between adjacent frames is small. The movement speed of the detection target in the image is relatively slow, and the position of the target object in the next frame of image is close enough to the position of the previous frame.

3) Consistent space. Pixels adjacent to the pixel point have similar motion, and the projection on the image plane is also in the adjacent area.

Under the above assumptions, suppose that the pixels (x, y, z) in the first frame of image move to the $(x + \Delta t, y + \Delta t, z + \Delta t)$ position of the second frame of image after the elapse of time Δt, based on the assumption that the first line of brightness is constant, the following conclusion can be drawn.

$$I(x, y, z) = I(x + \Delta t, y + \Delta t, z + \Delta t) \tag{3-24}$$

According to the second assumptions, the motion range of the object between adjacent image frames is very small. According to the Taylor series, expand the right side of the equal sign of the formula (3-24) to eliminate the similar terms, and divide the two sides by Δt at the same times to get the equation is as follows.

$$\frac{\partial I}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial I}{\partial y}\frac{\Delta y}{\Delta t} + \frac{\partial I}{\partial t}\frac{\Delta t}{\Delta t} = 0 \tag{3-25}$$

Finally, it can be concluded

$$\frac{\partial I}{\partial x}V_x + \frac{\partial I}{\partial y}V_y + \frac{\partial I}{\partial t} = 0 \tag{3-26}$$

Let

$$f_x = \frac{\partial I}{\partial x} \tag{3-27}$$

$$f_y = \frac{\partial I}{\partial y} \tag{3-28}$$

$$f_t = \frac{\partial I}{\partial t} \tag{3-29}$$

Then formula 3-26 can be simplified to

$$f_x V_x + f_y V_y + f_t = 0 \tag{3-30}$$

Among them, V_x and V_y represent the move speed of the pixel in the x direction and the y direction, or the optical flow of $I(x, y, z)$. f_x, f_y and f_t are the partial derivative of pixels (x, y, z) in the corresponding direction, f_x, f_y are the gradient of the image.

The formula (3-30) is the optical flow equation. However, both V_x and V_y in the equation are unknown parameters. We cannot solve two unknowns in one equation. Lucas-Kanade (L-K) algorithm can help us solve this problem.

The L-K algorithm is a widely used differential algorithm for optical flow estimation, invented by Bruce D. Lucas and Takeo Kanade. The algorithm assumes that the optical flow in the neighborhoods of the pixel

is a constant (that is, the third assumptions), and then uses the least square method to solve all the pixels in the neighborhood.

By combining the information of several adjacent pixels, the L-K algorithm can usually eliminate the ambiguity in the optical flow algorithm, compared with the point-by-point calculation method, the L-K algorithm is insensitive to image noise.

Based on the third assumptions, it is assumed that the optical flow equation formula (3-30) holds for all pixels in the window centered on P. That is, the flow vector (V_x, V_y) satisfies.

$$\begin{aligned} I_x(q_1)V_x + I_y(q_1)V_y + (I_t q_1) &= 0 \\ I_x(q_2)V_x + I_y(q_2)V_y + I_t(q_2) &= 0 \\ &\vdots \\ I_x(q_n)V_x + I_y(q_n)V_y + I_t(q_n) &= 0 \end{aligned} \tag{3-31}$$

Among them, q_1, q_2, ..., q_n are the pixels in the window, and $I_x(q_n)$, $I_y(q_n)$, $I_t(q_n)$ are the deviations of the image at the current moment to the current moment x, y and time t respectively at the point q_n derivative.

For the equations (3-30), the least squares method is used to fit, and the final solution result is

$$\begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} \sum_n I_x(q_n)^2 & \sum_n I_x(q_n)I_y(q_n) \\ \sum_n I_y(q_n)I_x(q_n) & \sum_n I_y(q_n)^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_n I_x(q_n)I_t(q_n) \\ \sum_n I_y(q_n)I_t(q_n) \end{bmatrix} \tag{3-32}$$

After finally obtaining V_x and V_y, the arctangent trigonometric function can be used to calculate the motion direction angle of the pixels in the area α.

The effect of using the optical flow method to detect the direction of movement is shown in Figure 3-17. The images on the left shows the direction of movement of a person when throwing a ball. The results show that it can better track the direction of movement when the basketball is thrown from the hand on the right. The image shows the detection of the smoke area. The results show that in the smoke area, the tracking of the movement direction of each feature point detected by the LK algorithm is not obvious, and the output result shows that the movement direction of the feature points is rather chaotic and not consistent. Towards

one direction. This is because the L-K algorithm is easier to track the points with obvious feature points (such as corner points), while the smoke is a non-rigid object, and the edge texture feature is not obvious. It is easier to track in the process of motion, and the smoke is a non-rigid object with edge texture features. It is not obvious. The texture features of the smoke are always changing during the movement, so the L-K algorithm cannot track the direction of the smoke well. Although the dense optical flow method can reduce the dependence on feature points, the amount of calculation is too large to be used in real-time detection algorithms.

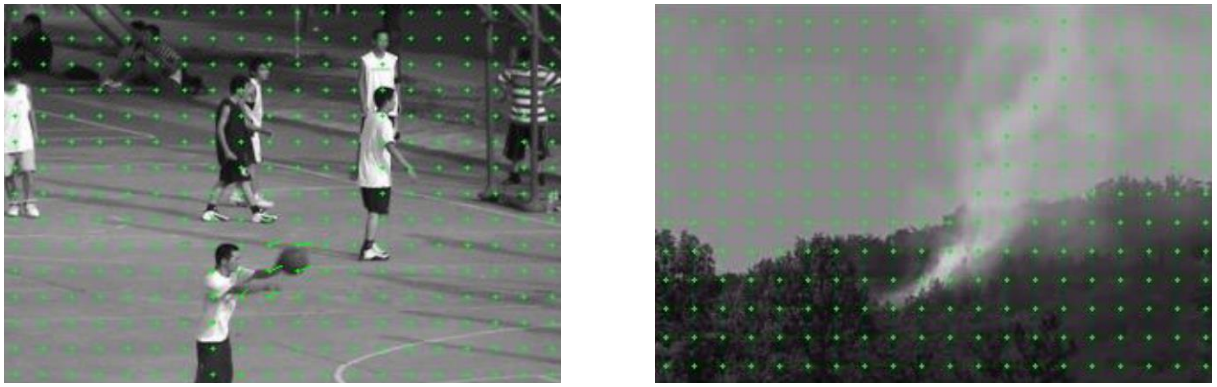


Figure 3-17: Use optical flow method to detect movement direction.

3.4.2 Based on the Motion Direction Detection of the Motion Block

The motion direction detection algorithm based on video blocks uses the feature of small image changes between adjacent frames of the video to determine the motion direction of the smoke. Although the smoke is constantly changing, the difference in the shape of the smoke in an image with a small number of frames is not very large. That is, when the frame interval is not large, the shape of the smoke in the direction of the movement of the smoke will change relatively close. Based on this feature of the smoke, and the fast-moving motion direction detection algorithm has a small amount of calculation and can meet the needs of real-time calculation, it is finally determined to use this algorithm to detect the motion direction of the image block.

In the detection process, the direction of the video is divided into 8 directions, with the horizontal to the right as the 0 degree direction, in a counterclockwise order, the interval of each direction is 45° degree direction, as shown in Figure 3-18, and then the 0° degree direction number is No. 1, and each direction is numbered in a counterclockwise order. No. 3 indicates a 90° degree direction, No. 5 indicates a 180° degree direction, and No. 7 indicates a 270° degree direction. The difference between the eight neighborhoods images at the corresponding positions in the center image and the next frame of image is calculated

respectively, and then the position with the smallest difference value is selected as the movement direction of the center image.

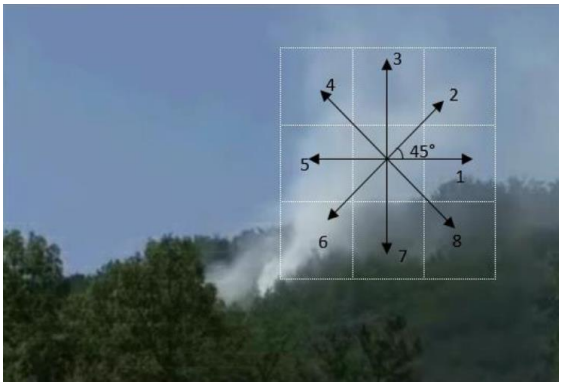


Figure 3-18: Schematic diagram of movement direction.

Let $I_t(x,y)$ be the central video block in the t-th frame of video, then $I_{t+1}(x+1,\ y)$, $I_{t+1}(x,\ y+1)$, $I_{t+1}(x+1,\ y-1)$, $I_{t+1}(x-1,\ y)$, $I_{t+1}(x-1,\ y-1)$, $I_{t+1}(x,\ y-1)$, $I_{t+1}(x+1,\ y-1)$ are respectively the eight-neighborhood video blocks located in the central video block 1st to 8th in the $t+1$-th frame of video. Then use the formula (3-34) to solve the difference between the 1st to 8th video block and the center video block respectively.

$$E = \frac{1}{w * h}\sum_{i=0}^{h}\sum_{j=0}^{w}(I_{t+1}(i,j) - I_t(i,j))^2 \tag{3-34}$$

Among them, E represent the size of the difference, w, h respectively represent the width and height of the video block, i, j respectively represent the size of the pixel value of the image block position coordinate (i,j). t represents the t-th frame image.

In the application of this thesis, first choose to use the GMM described in Chapter 4 to extract the motion area in the video, and then divide the video into blocks according to the pixel size of $32 * 24$, extract all the video blocks containing the motion area as the next step ROI areas detected. If the ROI area meets the color characteristics described in Section 3.3.2, then start to detect the motion direction of the ROI area, denote the ROI area as $C(x,y)$, and the specific detection steps are as follows.

1) The original image is RGB image. First, perform gray-scale processing on the ROI image according to formula (3-35) to obtain a gray-scale image $G(x,y)$.

$$G(x,y) = 0.299 * R(x,y) + 0.587 * G(x,y) + 0.114 * B(x,y) \tag{3-35}$$

2) Calculate the $C(x, y)$ and the difference $E1 \sim E8$ between the 8 neighborhoods image of the ROI area $C(x, y)$ in the t_{+1} frame.

3) Select the smallest neighborhood of E1~E8 as the movement direction of $C(x, y)$. If the movement direction range is 2, 3, 4, it is judged that the area has an upward trend, and $C(x, y)$ is passed as the final ROI area the first step is to classify and identify. If the direction of movement is not upward, the area is directly discarded.

3.4.3 Summary

This chapter analyzes the characteristics of the smoke area from the perspective of the color and movement characteristics of the smoke, concludes that the smoke presents a small saturation S in the HSV color space, and the appearance of smoke will increase the image value of the area. And using the characteristics of smoke generally moving upwards, after comparing the two methods of detecting the movement direction based on the optical flow method and the moving block, the movement direction detection based on the moving block is selected as the final detection algorithm. After the pre-processing and screening introduced in the second to fifth chapters, the ROI area for smoke recognition is finally selected. In the next chapter, will introduce the relevant theoretical knowledge of CNN, and how to use the CNN classifies and recognizes smoke images.

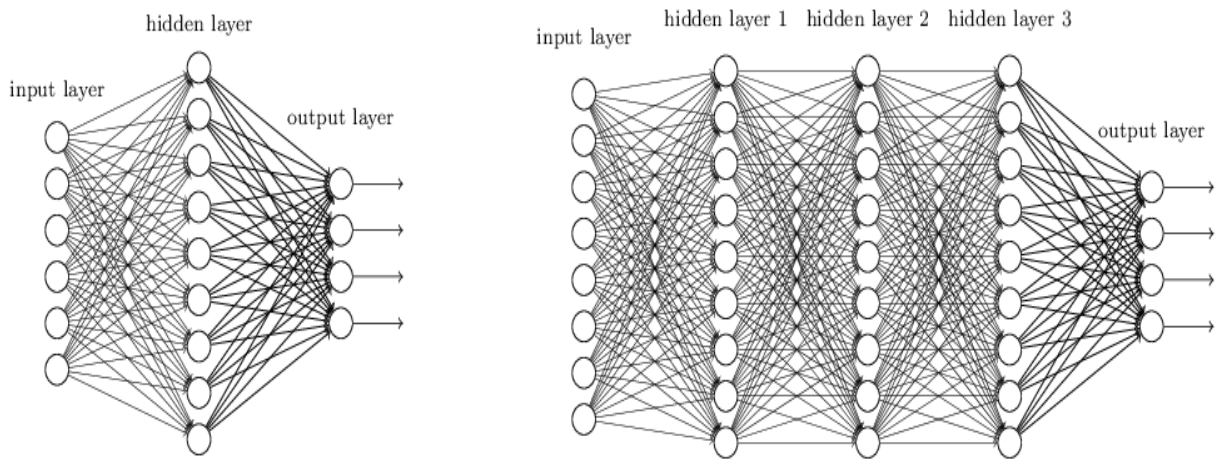
CHAPTER 4

SMOKE IMAGES RECOGNITION

With the continuous development of deep learning technology, end-to-end pattern recognition is becoming more and more popular. As long as the computing power is enough, you can always build a deep network of classification detection and prediction judgment suitable for a certain scene. However, in scenarios with short cycles, insufficient computing power, and high semantic understanding requirements, there is still a huge demand for feature-based pattern recognition. This section compares the classic features HOG+SVM and CNN to realize the intelligent recognition of smoke and non-smog. The pre-training cost, recognition speed and recognition rate of the two methods are compared.

4.1 CNN

In the past few years, the concept of deep learning has become more and more popular and has been applied in many technologies. The concept of deep learning was proposed by Professor Hiton and others in the paper published in "Science" in 2006 [59]. In the paper, the author described the deep learning neural network as a "deep autoencoding network". At the same time, the author also proposed to adopt train a multi-layer neural network with a small central layer to reconstruct a high-dimensional input vector, thereby converting the high-dimensional data into a low-level code. Since then, deep learning has been researched and applied by more and more researcher. Deep neural network is not a new artificial neural network, but a concept of network structure. Neural network is divided into input layer, hidden layer and output layer. If there are multiple hidden layers in a network structure, it is called a deep neural network, as shown in Figure 4-1. Convolutional neural network is a type of neural network, which is generally composed of multiple volumes and several layers (hidden layers), so CNN is also deep neural networks.



"Non-deep" feedforward neural network Deep neural network

Figure 4-1: Deep neural network and ordinary neural network.

The CNN is a kind of feed-forward neural network. Its special mechanism can be used to process two-dimensional data such as images and languages. The following chapters will Introduce CNN from the origin of neural network, network structure and characteristics of CNN.

4.1.1 The Origin of CNN

When Hubel and Wiesel [60-61] researched the visual cortex of cats, they found that there are two types of cells in the primary visual cortex of hair that bear different levels of visual perception functions, simple cells and complex cells, and they proposed the concept of receptive field [62]. The receptive field is the sensory receptor (the structure that senses external stimuli in biology). When stimulated and excited, the nerve impulse (various sensory information) is transmitted to the upper center through the centripetal neuron in the receptor, the stimulation area innervated by a neuron. For example, in the visual cortex, the receptive field of a neuron refers to a specific area on the retina that receives light stimulation, and only the stimulation in this area can activate the neuron. In the visual nervous system, the output of nerve cells in the visual cortex depends on photoreceptors on the retina. When the photoreceptors on the retina are stimulated and excited, they transmit nerve impulse signals to the visual cortex of the brain, but not all neurons in the visual cortex receive these signals, and only the receptive field of the receptor can be activated by the stimulus. Across the research found that in the visual cortex of cats, the receptive field of simple cells is long and narrow. Each simple cell is only sensitive to a certain angle of light in the receptive field, while complex cells are sensitive to a certain angle in the receptive field moving in a specific direction light band.

Inspired by this, Japanese scholar FukuShima proposed a multi-layer neural network with convolution and sub-sampling operations in 1980 [63], which uses training methods for unsupervised learning, it's considered to be the first volume implemented product neural network. In 1989, YannLeCun [64] introduced the back-propagation algorithm to the CNN, and in 1998 applied the CNN to the recognition of handwritten digits [65] and achieved great success.

In recent years, CNN have been used in image recognition, video analysis, natural language processing, drug discovery and other fields. At the same time, in March 2016, the artificial intelligence Go program AlphaGo developed by Google DeepMind in London, UK also used the relevant content of the CNN, showing the majors breakthrough of deep learning in the field of Go [66].

4.1.2 Network Structure of CNN

The conceptual diagram of the network structure of the CNN is shown in Figure 4-2.

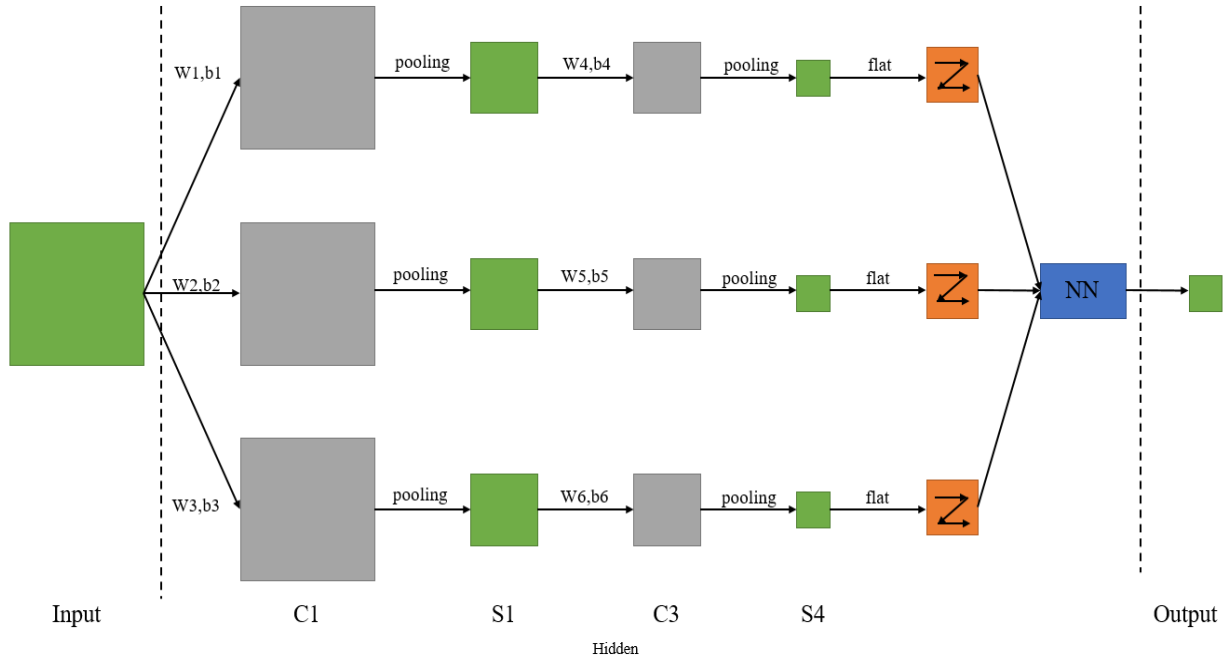


Figure 4-2: Conceptual diagram of convolutional neural network structure.

In the CNN, it is generally divided into input, convolutional layer, pooling layer, fully connected neural network layer and output layer. Among them, the convolutional layer, the pooling layer and the fully connected layer are the hidden layers in the corresponding neural network. The detailed steps of the convolutional neural network are as follows.

1) Through the input layer, input image I to the network.

2) Set n convolution kernels w and bias b, then use formula (4-1) to perform convolution operations on the input image I to obtain the $C1$ layer image. The convolution layer mainly plays a role in extracting image features. Through the convolution calculation of the image and large number of convolution kernels, the edges of the image can be extracted, gradient and other features.

$$C1 = (I * w) + b = \sum_{i=0}^{height} \sum_{j=0}^{width} (I(j,i)w(j,i) + b) \tag{4-1}$$

3) Through the pooling layer, the image of the $C1$ layer is reduced to half of the original. The methods of eating painting include maximum pooling and average pooling. Here, the maximum pooling with a size of $2 * 2$ is used as an example to explain how the pooling operation is performed. Suppose the original image size is $4 * 4$, as shown in Figure 4-3 (left), the pooled image is $2 * 2$. First select the $2 * 2$ area in the upper left corner of the original image, and then take the maximum value in this area as the value of the upper left corner of the image after pooling, fill in the position of the upper left corner of the right image. Then select the $2 * 2$ area in the upper right corner of the original image, then select the maximum value in this area again to fill in the upper right corner of the pooled image, and so on. The final pooled image is shown in Figure 4-3 (right). Since a maximum value is selected from the $2 * 2$ area each time as the pooling result, the final generated image becomes half of the original image. Through the pooling layer, amount of calculation data can be reduced. At the same time, because the image is compressed during pooling, part of the image information is lost, and some unnecessary details in the image can be erased, thus avoiding the occurrence of network overfitting.

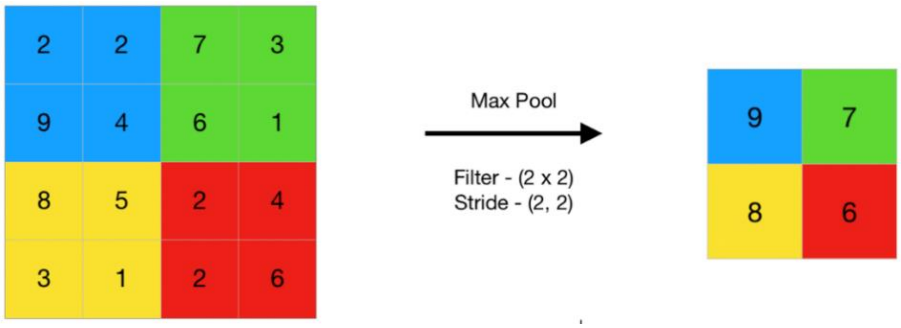


Figure 4-3: Example of max pooling, the left images is the original image, and the right images is the image after pooling.

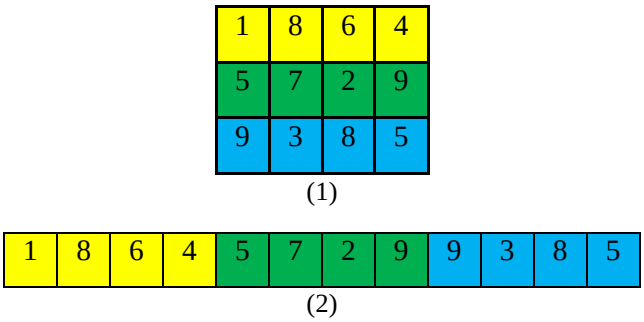


Figure 4-4: Schematic diagram of image data flattening.

4) Repeat steps 2 and 3 to perform several convolution and pooling operations. The specific number of times depends on the actual situation. Finally, the pooled result $S4$ is flattened. Flattening is to split the original image into rows, and then combine each row end to end to form a set of one-dimensional data. Suppose the image size of the $S4$ layer is $3 * 4$, as shown in Figure 4-4 (1), and the image size after flattening is $1 * 12$, as shown in Figure 4-4 (2).

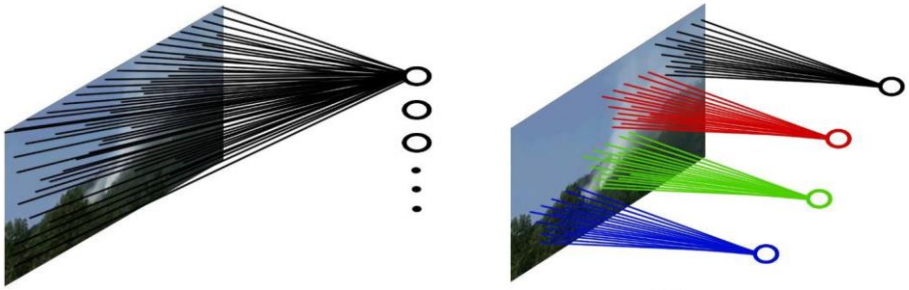
5) Input the flattened image to the fully CNN layer for calculation, and finally output the classification result to the output layer.

4.1.3 Local Connection and Weight Sharing

Compared with ordinary neural networks, the biggest advantage of CNN is local connection and weight sharing. These two characteristics greatly reduce the calculation amount of CNN, improve the computational efficiency of CNN, and maintain Original characteristics of image data.

First look at the local connection characteristics, as shown in Figure 4-5 (1), the left image is a schematic image of a fully connected network, and the right images is a schematic image of a partially connected network. Assuming the image size is $1000 * 1000$, if the network is a fully connected network, each pixel of the image needs to be connected to each neuron of the next layer of the network, and the total number of connections is $(1000 * 1000)\ (1000 * 1000) = 10^{12}$, which means that there are 10^{12} weight parameters in the network that need to participate in the calculation, and the number of layers in the network is generally more than one layer, so the number of parameters is extremely large. However, in the image, each individual pixel has no actual meaning, and only when combined with the surrounding pixels can it show specific information. Just like in the visual system, through the local receptive field to receive the stimulus of the external image, there is no need to perceive the global image, and the information felt by these local receptive fields can be synthesized at the high level to obtain the global image information. As shown in Figure 4-5(2), assuming that the size of the local receptive field is 10*10, the number of connections connected to the next layer of the network is $(10 * 10) * (1000 * 1000) = 10^8$, which is the weight the

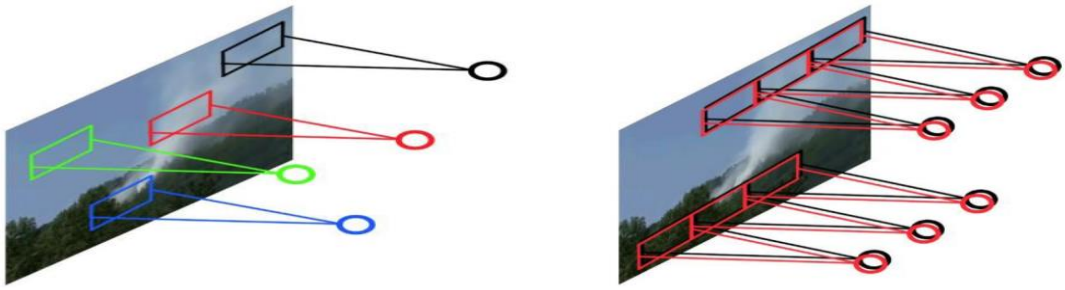
number of parameters has been changed to 10^8, which is 4 orders of magnitude less than the original number, and the amount of calculation is greatly reduced.



(1) (2)

Figure 4-5: Fully connected neural network (1) and partially connected neural network (2)

Look at the weight sharing feature, as shown in Figure 4-6, (1) is a schematic image of non-weight sharing, and (2) is a schematic image of weight sharing. On basis of the local connection, each neuron of the next layer of network must connect $10 * 10$ local receptive fields, that is, the local image of the image of the upper layer. If each receptive field is assigned a weight, there will be 108 weights in total. This is the result of non-weight sharing. If all receptive fields use the same weight, the total number of weights will become to $10 * 10 = 100$ weight parameters, and these 100 weights are the final parameters to be trained. In a CNN, the weight of each local receptive field is the value of the convolution kernel that convolves the receptive field. In each hidden layer, the number of convolution kernels is generally more than one, assuming that the network is convolved, there are 100 product cores, and the total number of weights need to be trained is $100 * 100 = 10^4$. Compared with the fully connected neural network, the complexity and amount of calculation are greatly reduced.



(1) (2)

Figure 4-6: Non-weight sharing (1) and weight sharing (2).

Under the combined effect of local connection and weight sharing, the parameters need to be trained in the CNN are reduced from 10^{12} at the beginning to 10^4 in the end, and the calculations for training the network is greatly reduced. This is also the characteristic of the CNN.

4.2 HOG+SVM

HOG + SVM gives us the first impression that they are used for pedestrian detection, but in fact they can be used for the detection of arbitrary target objects. What I want to introduce in this section is to use HOG + SVM to detect and classify smoke and non-smoke images.

4.2.1 Support Vector Machine（SVM）

In machine learning, the supervised learning model of support vector machine and related learning algorithms can be analyzed for data classification and regression analysis. Given a set of training examples, each training example is marked as belonging to one or the other of the two categories. The SVM training algorithm builds a model and assigns new examples to one category or the other, making it Non-probabilistic binary linear classifiers (although methods such as Platt scaling exist to use SVM in probabilistic classification settings). The SVM model represents these examples as a map of points in space, so as to divide the examples of each category by the widest possible obvious gap. Then map the new examples into the same space and predict that they belong to the category based on which side they fall on.

Since SVM is a convex optimization problem, the solution obtained must be the global optimum, not the local optimum. It is not only suitable for linear problems, but also for nonlinear problems. SVM can also be used for data with high-dimensional sample space, because the complexity of the data set depends only on the support vector rather than the dimensionality of the data set, which avoids the "dimensionality disaster" in a sense. The theoretical foundation is relatively complete.

4.2.2 Histogram of Oriented Gradient（HOG）

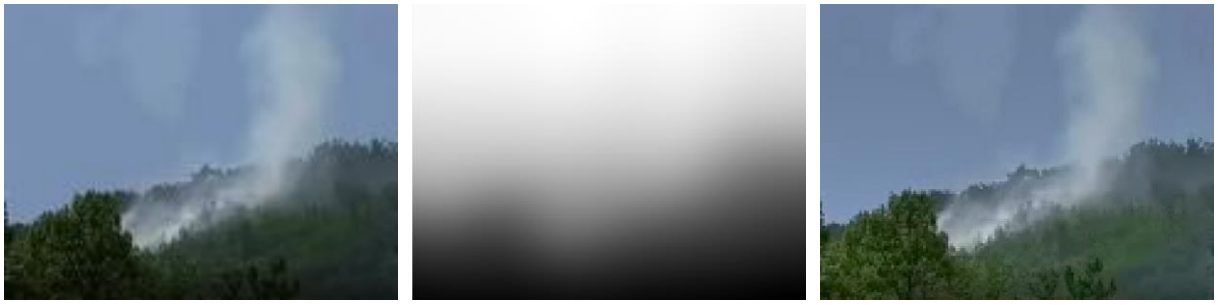
The HOG feature is a feature descriptor used for object detection in computer vision and image processing. It composes features by calculating and counting the histogram of the gradient direction of the local area of the image. Hog feature combined with SVM classifier has been widely used in image recognition, especially in pedestrian detection.

The main idea of HOG is that in an image, the appearance and shape of a local target can be well described by the directional density distribution of the gradient or edge. Compared with other characterization methods, HOG has many advantages. The important thing is that because HOG operates on the local grid cells of the image, it can maintain good invariance to the geometric and optical deformations of the image. These two types of deformations will only appear in a larger space.

The HOG feature extraction method is to put an image (the target or scan window wants to detect).

1) Grayscale (think the image as a three-dimensional image of x, y, z (grayscale)), for reduce the influence of lighting factors, the entire image needs to be normalized first. In the texture intensity of the image, the local surface exposure contributes a larger proportion, so this compression process can effectively reduce the local shadow and illumination changes in the image.

2) Use the Gamma correction method to standardize the color space of the input image (normalization), the purpose is to adjust the contrast of the image, reduce the impact of local shadows and light changes in the image, and suppress noise interference.



Original image Illumination component Applied Gamma correction

Figure 4-7: Gamma correction

3) Calculate the gradient of each pixel of the image (including size and direction), mainly to capture the contour information, while further weakening the interference of light.

4) Divide the image into small cells (for example, 4*4 pixels/cell),

5) Count the gradient histogram of each cell (the number of different gradients) to form the descriptor of each cell.

6) Combine every few cells to form a block (for example, $2 * 2$ cells/block), and concatenate the feature descriptors of all cells in a block to obtain the HOG feature descriptor of the block.

7) Series connect the HOG feature descriptors of all the blocks in the image to get the HOG feature descriptor of the image (the target we want to detect). This is the final feature vector available for classification.

The algorithm flow of using HOG+SVM to detect and classify smoke and non-smoke images is shown in Figure 4-8.

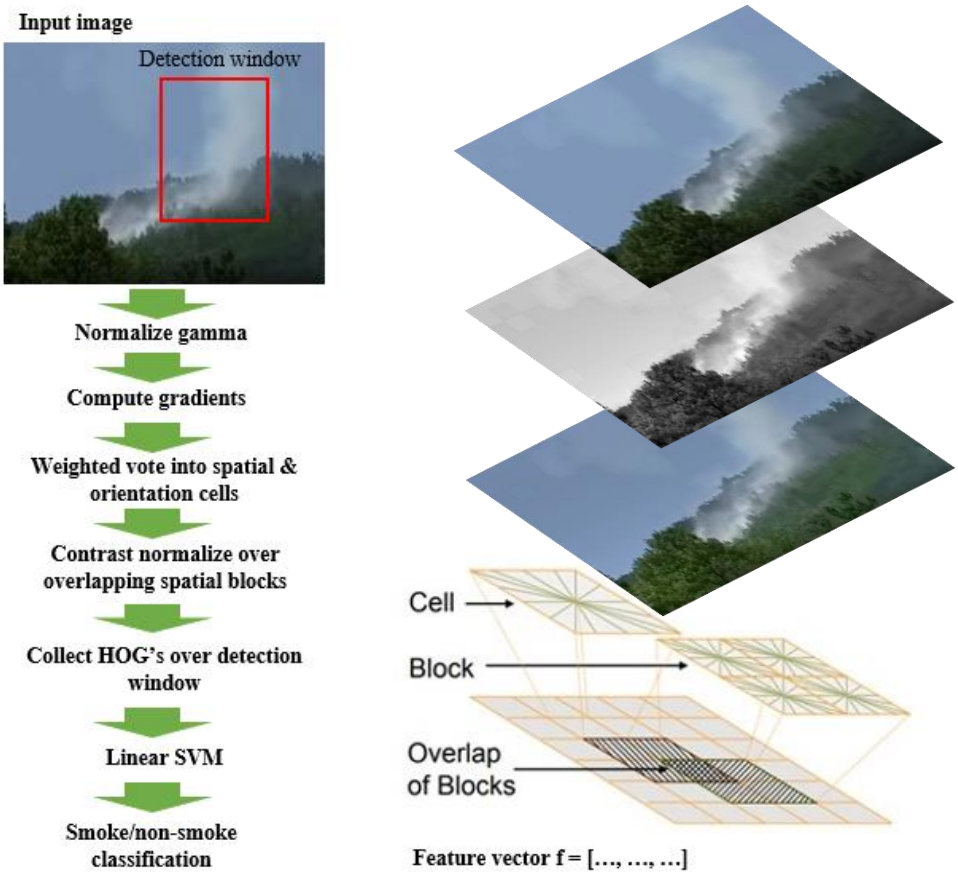


Figure 4-8: HOG feature extraction algorithm.

4.3 Experiment and Compare

In this section, CNN and HOG+SVM are used to identify the same group of smoke images, and the comparison proves that the CNN has better recognition ability for smoke images.

The design of the CNN is shown in section 5.2.3, and the experimental environment is shown in section 5.1. First, take the smoke image as a positive sample, and the non-smoke image as a negative sample. The data set used by CNN for training and testing are the same data set used by HOG+SVM method, the training data set include 12,470 smoke images (positive sample), 12,902 non-smoke images (negative sample), 500 smoke and non-smoke images in the test set, and the images in the test set and training set are different. The size of the images is $32 * 24$ pixels, all images are taken from screenshots of outdoor fire smoke videos. When the CNN is training in a random combination, that is, 100 images are randomly selected from the data set to form a training set to train the classifier for 1 training iterations. In this way, a small data set can be used to perform large numbers of trains on the network. Some non-smoke and smoke images in the data sets are shown in figure 4-9 and figure 4-10.

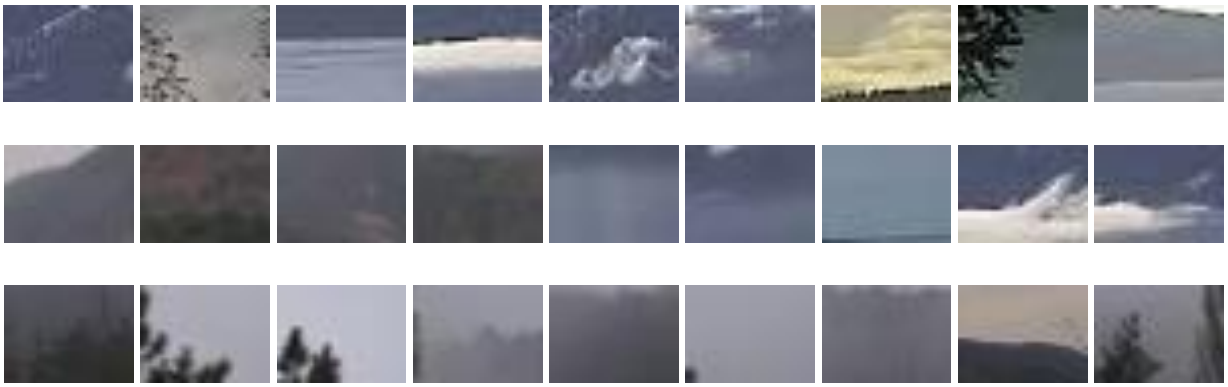


Figure 4-9: Some Non-smoke images (negative sample) in the data sets.

Figure 4-10: Some Smoke images (positive sample) in the data sets.

The result of the comparative experiment in the CNN and HOG+SVM as show in Table 4-1. The result is the classification effect on the test machine after the classifier is trained 5,000 iterations in a random combination.

Table 4-1: CNN and SVM classification experiment results.

Classification	Training iterations	Training time	Accuracy
HOG+SVM	5,000	28min	82.26%
CNN	5,000	220min	94%

(1) First 432, detected 444 (2) First 1, detected 48 (3) First 96, detected 120

Figure 4-11: The position of the first smoke image frame detected by CNN

(1) First 432, detected 480 (2) First 1, detected 84 (3) First 96, detected 180

Figure 4-12: The position of the first smoke image frame detected by HOG+SVM.

Using the fire smoke video collected by the CNN test after training, it is known that in the first test video, the first video frame where smoke appears is the 432th frame, and the smoke is detected at the 444th frame using the CNN method, and the HOG+SVM method detects inverted smoke at 480th frame. In the second test video, smoke starts to appear from the first frame, the CNN method starts to detect the inverted smoke at the 48th frame, and the HOG+SVM is at the 84th frame. In the third video, smoke appears from 96th frame, the CNN method is used to detect inverted smoke at 120th frame, and the HOG+SVM method is used to detect smoke at 180th frame. Can know that for the same test set, use of CNN can detect the smoke area

more quickly, which meets the real-time requirements of the fire-alarming system. The experimental comparison results are shown in the figure 4-11 and figure 4-12.

It can be seen from the table that the recognition accuracy of the CNN is 94%, and the recognition accuracy of HOG+SVM is 82.26%. CNN's ability to classify smoke images is better than HOG+SVM. The training time is not suitable for the GPU, because the calculation amount of CNN is relatively large, and the training time is relatively long.

The classification of images by SVM depends on the feature data of the images input to the classifier. The extraction of features plays a decisive role in the classification effect, while CNN can automatically learn features from the input images. As large as the number of images input to the classifier is large enough, CNN can learn the most suitable features for distinguishing different types of images.

4.4 Chapter Summary

This chapter introduced the relevant theories of CNN. and shows the advantages of CNN for classification of smoke images through comparison with the HOG+SVM method. Because the characteristics of smoke, difficult to extract the characteristics of smoke, CNN has its own advantages for smoke images recognition.

CHAPTER 5

SMOKE DETECTION SYSTEM DESIGN AND RESULT ANALYSIS

5.1 System Environment

The main parameters of the experimental environment used in the research process of this thesis are show in follow.

Hardware environment: Intel(R) Xeon(R) CPU E5620 @2.40GHz

RAM：12GB

operating system: Windows10

develop software: Matlab 2018b

5.2 System Flow Design

The system design process is shown in Figure 5-1,

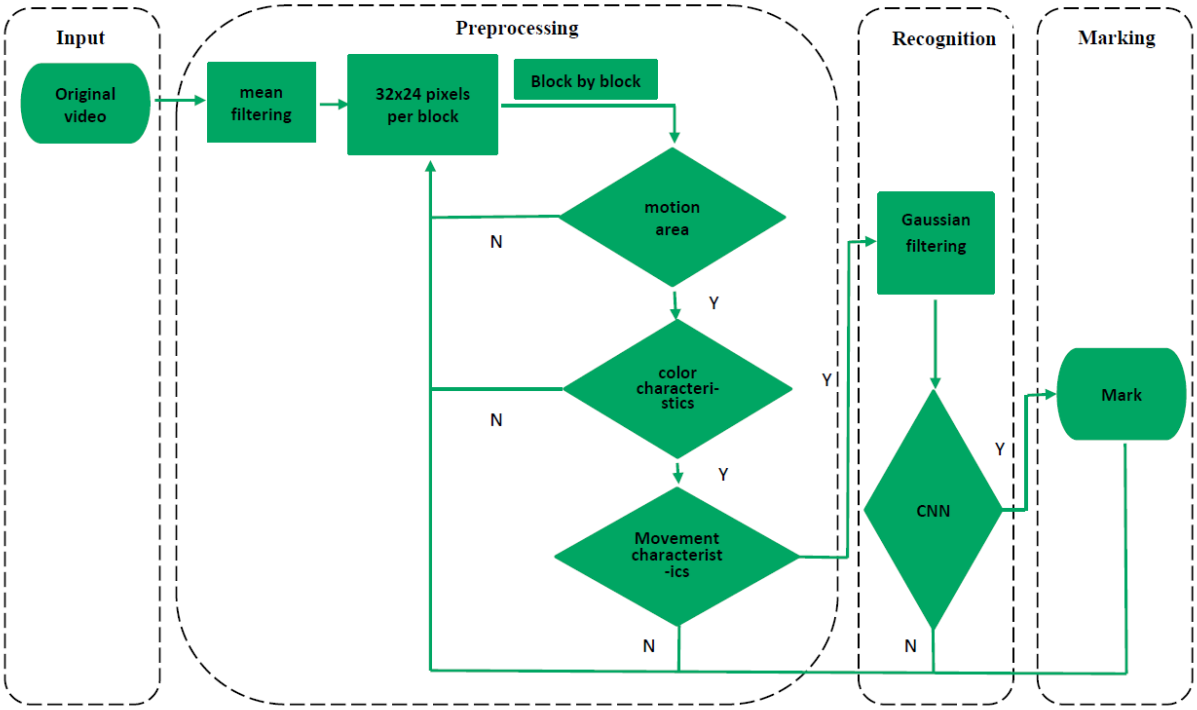


Figure 5-1: System design flow chart.

The system includes four stages, input stage, preprocessing stage, recognition stage and marking stage. The following subsections will introduce the four stages of functional realization and design methods in detail.

5.2.1 Input Stage

The input stage mainly realizes the function of input video data. In this stage, each frame of the video is decoded and input to the subsequent links frame by frame.

5.2.2 Preprocessing Stage

This stage includes all the processes of video processing before the images are input to the CNN, mainly include mean filter the entire image, image block, extract the motion area, extract the area that meets the HSV color characteristics, and extract the areas with upward movement tendency. The specific implementation process is as follows.

1) Use a $5 * 5$ kernel to perform the Mean filter on the entire image.

2) Divide the image into small block areas of $32 * 24$ pixels. and use the upper left corner of each block as the coordinates of the image block.

3) Use the GMM to extract the moving area in the image frame. and generate a binary mask image with a foreground tendency. The white pixels in the mask image are the foreground, the black pixels are the background. Select all the small block areas containing white pixels divided in step (2), input the selected small block areas one by one input to the next processing process.

4) According to the characteristics of the smoke in the HSV color space described in Section 3.3, first convert the image block input in step (3) to the HSV space, map the values of the three HSV channels to the range of 0 to 255. Then calculate the average saturation S of all pixels in the image block, if S is less than 70, go to step (5), otherwise judge the image block as a non-smoke image and continue processing the next image block.

5) Compare the average value V of the image block with the average value SV of the area 50 frames before this frame in the video. If V is greater than SV, the value V of the area has increased. Then proceed to the next step. Input the image block into step (6), otherwise it is judged that the image block is non-smoke image, continue to process the next image block.

6) According to the motion direction detection method based on block motion described in Section 3 and 4, the value of the image block in the RGB color space is used to determine the motion direction of the image block in combination with the image at that position in the previous frame. If the direction is any one of 2, 3, 4, it is determined that the image block has an upward movement trend, the position of the image block in the video is recorded, and the image block is intercepted in the original video, otherwise, the image block is judged to be a non-smoke block, continue to process the next image block.

In step 6), the image block input to this step is not used directly as the image of the input CNN, but the image block at that position in the original video is used, because when the CNN is trained, the input is input to the CNN. The images of the product neural network are all intercepted from the original video. There are no processed images, and the image block input to step 6) is the image after the mean filter. For ensure the consistency of the data, select the image in the original video as input to the CNN.

5.2.3 Recognition Stage

The recognition stage mainly uses the CNN to classify and recognize the ROI block obtained in the preprocessing stage. In chapter 4, the influence of filtering on the classification effect of CNN is mentioned,

so before inputting the image to the neural network, first use the $5 * 5$ check image to perform Gaussian filter, and then input the filtered image to recognition in the CNN.

The structure diagram of the CNN used in this system is shown in Figure 5-2. It includes one input layer, three convolutional layers, three pooling layers, one fully connected layer and one output layer. The specific network structure design is as follows.

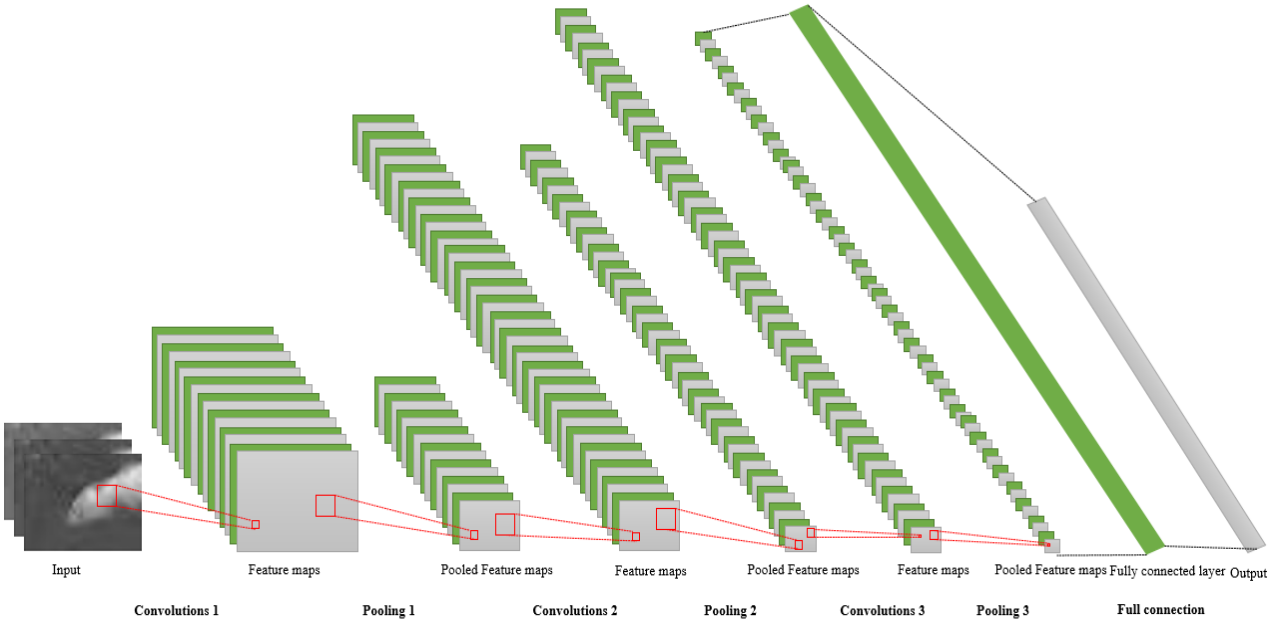


Figure 5-2: Schematic diagram of the convolutional neural network.

1) Input layer. The input data of the input layer is the image with a size of $32 * 24$ pixels. Since the input images is RGB three-channel data, the dimension of the input data is $32 * 24 * 3$.

2) Convolutional layer. The size of the convolution kernel is $5 * 5$, the step size is 1, the number of feature maps is 32, and the input image is convolved. After the Rectified linear unit (ReLU) function is used as the activation function, the final output result is $32 * 24 * 32$-dimensional data. The weight value of the convolution kernel is randomly generated using the normal distribution with a variance of 0.1, and the bias value is 1. The formula of the ReLU activation function is as follows.

$$f(x) = \max(0, x) \tag{5-1}$$

3) Pooling layer $S1$. The maximum pooling method is used to resample the output result of the convolutional layer $C1$, and the output result is $16 * 12 * 32$ dimensional data.

4) The functions of the convolutional layers $C2$ and $C3$ are the same as $C1$, the convolution kernels are both $5 * 5$, the step size is both 1, the number of feature maps are 96 and 128 respectively, and the dimensions of the output data are $16 * 12 * 96$ and $8 * 6 * 128$, the functions of pooling layers S2 and S3 are the same as S1, and the dimensions of the output data are respectively $8 * 6 * 96$ and $4 * 3 * 128$.

5) $F1$ is a fully connected layer, set 1024 neurons to connect with $S2$, and use ReLU as the activation function.

6) $F2$ is the output layer, set 2 neurons to connect with $F1$, use SoftMax as the activation function, and finally perform two classifications on the 1024-dimensional vector.

The parameter configuration of the CNN as show in Table 5-1.

Table 5-1: Network parameter configuration

Layer number	category	Feature map number	Convolution kernel size	Stride	Output dimension
Input	Input layer	3	/	/	32*24*3
C1	Convolutional layer	32	5*5	1	32*24*32
S1	Pooling layer	32	/	2	16*12*32
C2	Convolutional layer	96	5*5	1	16*12*96
S2	Pooling layer	96	/	2	8*6*96
C3	Convolutional layer	128	5*5	1	8*6*128
S3	Pooling layer	128	/	2	4*3*128
F1	Fully connected layer		Number of neurons: 1024		
F2	Output layer		Number of neurons: 2		

In the fully connected layer $F1$, the Dropout [67] method is used to avoid the occurrence of overfitting. Dropout method is proposed by Hintion, to improve the performance of neural networks by preventing the common effects of feature detectors. In the process of training the model, Dropout refers to randomly letting the weights of some hidden layer nodes in the network not involved in calculation, shielding these nodes from the network and not participating in the calculation, but retaining their weights, and only updating the weights of the working nodes. In this way, it can be ensured that any two nodes cannot work at the same

time, thereby avoiding the interaction of two implicit nodes with dependent relationships, preventing the situation that some features and other features are one inch apart. The optimization method used in the network is Adam optimizer [68]. The training image data set and test image data set of the CNN are different from each other.

The output layer outputs a one-dimensional array containing two weights each time. The first value W_1 in the array indicates the weight of the images belonging to smoke, and the second value W_2 indicates the weight of the images belonging to the non-smoke images, when $W_1 > W_2$ Indicates that the images is a smoke images, otherwise it is a non-smoke image.

5.2.4 Marking Stage

The marking stage means that the area judged as smoke in the recognition stage is marked in the original video. The recognized image areas are all small blocks divided in the original video, so the marked smoke areas are all $32 * 24$ rectangular areas, as show in Figure 1-4.

5.3 Analysis of Results

In this section, four sets of experiments will be conducted to verify the effectiveness of the algorithm designed in this thesis from different aspects. The four groups of experiments are as follows.

1) Do not use the ROI extraction method described in Chapter 3, and directly use CNN to detect the smoke video test set. The purpose is to verify the effectiveness of the preprocessing algorithm designed in this thesis for smoke video detection.

2) In the algorithm described in section 5.2, use HOG+SVM as a classifier instead of CNN, and then detect the video smoke test set. The purpose is to verify the effectiveness of CNN in identifying smoke images.

3) In the algorithm described in section 5.2, the images input to CNN is converted into HSV color space, and then the smoke video test set is detected. The purpose is to verify whether the different characteristics of smoke images in different color spaces have an impact on the classification of CNN.

4) Use the method described in this thesis to detect the smoke video test sets.

The smoke video test sets in the above experiment are all 12 video clips organized in section 2.2.

5.3.1 Experimental Method Description

The 12,470 smoke images and 12,902 non-smoke images in the data set section described in Section 4.2 are used as training data for training CNN and SVM. After experimentation, it is found that 100 images are randomly selected for training each iterations, 5000 iterations can make the network achieve better recognition effect, and the network can achieve better recognition effect after using it, and in the case of training without GPU, 5000 iterations training time is about 220min, and the recognition rate can reach more than 96% in the training set, so the network is trained for 5000 iterations each time, and the test video is directly recognized after the training is completed. The method described in this section experiment, the training method of SVM is same as CNN. For 5000 iterations of training, the time is about 28min, and the recognition rate is between 65% and 68% on the training set.

5.3.2 Evaluation Method of Experimental Results

The experimental results evaluate the quality of the model through 8 indicators True Positive (TP), False Positive (FP), True Negative (TN), False Negative (FN), accuracy rate, false alarm rate, sensitivity, and missed detection rate. The relationship among TP, FP, TN, and FN is shown in Table 5-2.

Table 5-2: Confusion matrix of classification results.

Reality	Forecast result	
	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

The concepts of the above eight indicators are as follows,

1) TP. Indicates the number of video frames of smoke identified as smoke in the smoke sample.

2) FP. Indicates the number of video frames of smoke that are not recognized as smoke in the smoke sample.

3) TN. Indicates the number of video frames in non-smoke samples that are not recognized as smoke.

4) FN. Indicates the number of video frames identified as smoke in non-smoke samples.

5) Accuracy rate. refers to the proportion of frames with real smoke in all alarm frames in all test videos, which indicates the probability that the system will issue a correct alarm. The calculation formula of accuracy is shown in (5-2),

$$Accuracy = \frac{TP}{TP + FN} \tag{5-2}$$

6) False alarm rate. refers to the proportion of the number of frames where non-smoke is detected as smoke among all the alarm frames in all test videos. It represents the probability that the system will issue a false alarm. The calculation is as shown in formula (5-3),

$$False = \frac{FN}{TP + FN} \tag{5-3}$$

7) Sensitivity. refers to the proportion of correctly detected smoke frames in the total smoke frames in all test videos. Indicates the system's sensitivity to smoke images. The higher the sensitivity, the stronger the system's ability to recognize smoke images. The calculation formula is shown in (5-4),

$$Sensitivity = \frac{TP}{FP + TP} \tag{5-4}$$

8) Missing detection rate. refers to the proportion of undetected smoke frames in the total image in all test videos. The lower the missed detection rate, the stronger the system's ability to recognize smoke. Its calculation formula is as shown in (5-5),

$$Miss = \frac{FP}{FP + TP} \tag{5-5}$$

5.3.3 Experimental Results and Analysis

The statistical results of the four groups of experiments are shown below. In the table, N represents non-smoke video and S represents smoke video. The unit of smoke frame number, non-smoke frame number, TP, FP, TN, and FN is frame.

Table 5-3: Experiment 1: Use CNN experiment statistical result.

Video number	Description	Smoke frames	Non-smoke frames	TP	FP	TN	FN
1	N	0	1,499	0	0	1,367	132
2	N	0	902	0	0	902	0
3	N	0	2,010	0	0	2,010	178
4	S	1,214	235	324	890	235	0
5	S	1,095	103	461	634	103	0
6	S	1,216	283	396	820	226	57
7	S	1,123	376	587	536	376	0
8	S	2,203	818	954	1,249	801	17
9	S	1,301	727	791	510	663	64
10	S	719	780	468	251	780	0
11	S	2,324	0	1,267	1,057	0	0
12	S	7,525	0	1,768	5,757	0	0
Total number of frames		18,720	7,733	7,016	11,704	7,285	448
Accuracy rate	94%	False alarm rate	6%	Sensitivity	37.48%	Missing detection rate	62.52%

Table 5-4: Experiment 2: Use HOG+SVM experiment statistical result.

Video number	Description	Smoke frames	Non-smoke frames	TP	FP	TN	FN
1	N	0	1,499	0	0	264	1,235
2	N	0	902	0	0	902	0
3	N	0	2,010	0	0	2,010	0
4	S	1,214	235	515	699	233	2
5	S	1,095	103	769	326	103	0
6	S	1,216	283	5	1,211	283	0
7	S	1,123	376	304	819	376	0
8	S	2,203	818	80	2,123	814	4
9	S	1,301	727	316	985	708	19
10	S	719	780	0	719	780	0
11	S	2,324	0	931	1,393	0	0
12	S	7,525	0	2,923	4,602	0	0
Total number of frames		18,720	7,733	5,843	12,877	6,473	1,260
Accuracy rate	82.26%	False alarm rate	17.74%	Sensitivity	31.21%	Missing detection rate	68.79%

Table 5-5: Experiment 3: Use HSV+CNN experimental statistical result.

Video number	Description	Smoke frames	Non-smoke frames	TP	FP	TN	FN
1	N	0	1,499	0	0	1,484	15
2	N	0	902	0	0	902	0
3	N	0	2,010	0	0	1,996	14
4	S	1,214	235	338	876	233	2
5	S	1,095	103	827	268	103	0
6	S	1,216	283	10	1,206	283	0
7	S	1,123	376	287	836	347	29
8	S	2,203	818	1,269	934	814	4
9	S	1,301	727	78	1,223	727	0
10	S	719	780	16	703	709	73
11	S	2,324	0	16	2,308	0	0
12	S	7,525	0	940	6,585	0	0
Total number of frames		18,720	7,733	3,781	14,939	7,598	135
Accuracy rate	96.55%	False alarm rate	3.45%	Sensitivity	20.20%	Missing detection rate	79.80%

Table 5-6: Experiment 4: Use Preprocessing + CNN experimental statistical result.

Video number	Description	Smoke frames	Non-smoke frames	TP	FP	TN	FN
1	N	0	1,499	0	0	1,335	164
2	N	0	902	0	0	888	14
3	N	0	2,010	0	0	2,010	0
4	S	1,214	235	264	950	235	0
5	S	1,095	103	786	309	103	0
6	S	1,216	283	51	1,165	283	0
7	S	1,123	376	511	612	315	61
8	S	2,203	818	1,503	700	813	5
9	S	1,301	727	866	435	706	21
10	S	719	780	30	689	701	79
11	S	2,324	0	395	1929	0	0
12	S	7,525	0	2,294	5,231	0	0
Total number of frames		18,720	7,733	6,700	12,020	7,389	344
Accuracy rate	95.11%	False alarm rate	4.88%	Sensitivity	35.79%	Missing detection rate	64.21%

Analysis of results,

1) Through the comparison of experiment 1 and experiment 4, can know the preprocessing process designed in this paper greatly reduces the amount of data input to CNN, improves the processing speed of CNN, and filters out some areas with strong interference, which is a guarantee the real-time nature and effectiveness of detection play an important role.

2) Through the comparison of experiment 2 and experiment 4, can know the average detection accuracy of experiment 2 is 82.61%, and the average accuracy of experiment 4 is 95.11%. To a certain extent, it reflects that CNN can better recognize smoke images compared to traditional classifiers.

3) Through the comparison of experiment 3 and experiment 4, can know the average accuracy of experiment 3 is 96.55%. Although outdoor smoke images have obvious characteristics in HSV color space, when inputting CNN data, the image is in RGB color Space and HSV color space do not have a great influence on the classification accuracy of CNN. At the same time, the missed detection rate of experiment 3 is 79.8%, and the missed detection rate of experiment 4 is 64.21%. The accuracy of the two experiments is not much different, but the missed detection rate of experiment 4 is significantly lower than that of experiment 3, indicating that in the HSV color space when outdoor smoke images are used as input data, images containing smoke can be detected more comprehensively.

4) The missed detection rates of the three groups of experiments were 68.79%, 79.8%, and 64.21%, reflecting that the overall recognition effect of the system for smoke is still not ideal. The main reasons for the high missed detection rate include, GMM cannot detect stable after Smoke area, the interference of floating clouds and fog, the small proportion of smoke in the video, etc.

In generally, the outdoor fire smoke video detection system designed in this thesis can accurately warn the video with smoke, and the system is robust. It can basically achieve the effect of real-time detection of smoke video, but the system leaks the detection rate is still high, which is the focus of further improvement in the future.

5.4 Chapter Summary

This chapter mainly introduced the design scheme of outdoor fire smoke detection system and the analysis of experimental results. Through the test on the test data set, it is finally shown that this outdoor fire smoke detection system has a good smoke detection effect, but the missed detection rate is still the focus of the next step.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Conclusion

Vehicles in the parking lot are relatively dense. Once a fire takes shape, it will inevitably cause a large amount of property losses and even casualties. If there is a monitoring system that can detect the occurrence of fires as soon as possible, and eliminate fire hazards as soon as possible, To eliminate the fire in an unburnt state can better protect the safety of human life and property. Smoke detection is the early detection and treatment of fires that occur. When the fire occurs, people are given a fire alarm, so that we can carry out put out a fire in time, and most reduce the loss caused by the fire. This thesis combines the static and dynamic characteristics of smoke and applies image recognition technology to early fire detection, Using the collected outdoor smoke video, the fire smoke detection and warning system proposed in this thesis is designed. This system can be applied to outdoor parking lots. The use of cameras arranged in the parking lot can reduce manpower input and identify fires in a timely and effective manner. And issued an early warning has important practical significance and application value for fire prevention, fire scale judgment and put out a fire.

6.2 Future Work

Video detection technology for outdoor fire smoke is not mature enough. Most of the research results are still limited to the research rooms of scientific research institutions, and the current market promotion value has not been clearly reflected, unlike the market demand for license plate recognition and fingerprint recognition. More extensively, market value is a powerful driving factor for research, which is also the reason for the slow progress in outdoor fire smoke identification. In addition, there is no standard data set for outdoor fire smoke detection, and personally, it is not easy to collect outdoor fire smoke data sets, which is also a major obstacle to outdoor smoke detection research. It is believed that with the promotion and development of video control technology, the technology of outdoor fire smoke video detection will also develop rapidly.

The current fire smoke detection algorithm is still based on traditional detection methods. It is necessary to manually select the most suitable smoke feature, and then use the algorithm to extract the feature and input

it into the classifier for classification to determine whether it is smoke. However, the characteristics of smoke are easily disturbed by the outside world. It is a better way to let the computer automatically learn the characteristics of smoke and then classify it, which can avoid the characteristics of smoke that are difficult to extract. The development of deep learning makes computers more and more intelligent. Through training data, computers can learn the characteristics of smoke by themselves. I believe that with the development of deep learning technology, more and more artificial intelligence technologies will be applied outdoors fire smoke detection technology.

In addition, the system proposed in this thesis still has a high rate of missed detection and false detection. In the future, it is still necessary to improve the intermediate pretreatment process, make full use of the characteristics of smoke, and distinguish between smoke and smog.

In the future research and study, I will continue to pay attention to the latest developments in the outdoor of smoke recognition, improve existing algorithms, and develop more practical outdoor fire smoke detection technologies.

REFERENCES

[1] Ojala T., Pietikinen M. and Harwood D. A Comparative Study of Texture Measures with Classification Based on Feature Distributions. Pattern Recognition, 29, 51-59, 1996.

[2] Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection[C] IEEE Computer Society Conference on Computer Vision & Pattern Recognition. IEEE, 2005.

[3] Platt J. Neural Information Processing System[M], Commonwealth of Massachusetts: MIT Press, 557-563, 1999.

[4] Healey G, Slater D, Lin T. A system for real-time fire detection[C] IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 1993.

[5] Chen T H, Wu P H, Chiou Y C. An early fire-detection method based on image processing[C] Image Processing, 2004. ICIP '04. 2004 International Conference on. IEEE, 2005.

[6] Liu Che-bin & Ahuja, Narendra Vision Based Fire Detection. 10.1109/ICPR.2004.979, 2004.

[7] Premal C E, Vinsley S S. Image processing based on forest fire detection using YCbCr colour model[C] International Conference on Circuit. IEEE, 2015.

[8] Yaqin Zhao, G.T. Fire Video Recognition Based on Flame and Smoke Characteristics[A], 2nd International Conference on Systems and Informatics [C], 113-118, 2014.

[9] Yuan F. Video-based smoke detection with histogram sequence of LBP and LBPV pyramids[J]. Fire Safety Journal, 132-139, 2011.

[10] Jayavardhana Gubbi, S. M. Marimuthu Palaniswami, Smoke detection in video using wavelets and support vector machines, Fire Safety Jourmal[J], 44:1110-1115, 2009.

[11] Panagiotis B., Kosmas D., Nikos G. Smoke detection using spario-temporal analysis, motion modeling and dynamic texture recognition[C]. In 2014 Proceedings of the 22nd European, Signal Processing Conference (EUSIPCO). 2076-1465, 2014.

[12] Bo Jiang, Y. L., Xiying Li, et al, Towards a solid solution of real-time fire and flame detection, Multimed Tools Appl[J], 2013.

[13] Enis Çetin, K. D., Benedict Gouverneur, Nikos Grammalidis, Osman Günay, Y.Hakan Habiboglua, B. Ugur Töreyin, and Verstockt, S., Video fire detection Review, Digital Signal Processing[J], 2013, 23:1827-1843.

[14] ByoungChul Ko, J.-Y. K., Jae-Yeal Nam, Wildfire smoke detection using temporospatial features and random forest classifiers, Optical Engineering[J], 2012, 51(1):017208_017201-017208_017210.

[15] ByoungChul Ko, S. K., Survey of computer vision-based natural disaster warning systems, Optical Engineering[J], 2012, 51(7):0709011-07090111.

[16] Pasquale Foggia, A. S. a. M. V., Real-time Fire Detection for Video Surveillance Applications using a Combination of Experts based on Color, Shape and Motion, IEEE Transactions on Circuits and Systems for Video Technology[J], 2015.

[17] Rosario Di Lascio, A. G., Alessia Saggese, et al, Improving fire detection reliability by a combination of videoanalytics[A], 11th International Conference on Image Analysis and Recognition, ICIAR 2014[C], 2014.

[18] CVPR Lab. at Keimyung University.Wildfire smoke ideo Database [DB/OL].htt-p://cvp-r.kmu.ac.kr/, 2012/2015-12-20.

[19] University of Salerno.Smoke Detection Dataset [DB/OL]. http://mivia.unisa.it/,2012-7/2015-12-20.

[20] http://signal.ee.bilkent.edu.tr/VisiFire/Demo/SmokeClips/ 2014/2015-12-20. Cetin, E., Computer Vision Based Fire Detection Dataset [DB/OL].

[21] Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives[J]. IEEE Transactions on Audio, Speech and Language Processing, 2012, 20(1):14-22

[22] Sainath T N. Improvements to deep neural networks for large vocabulary continus speech recognition tasks[R]. IBM T. J. Watson Research Center, 2014.

[23] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]. In Procceeding of 26th Annual Conference on Neural information processing system foundation, 2012:1097-1105.

[24] Sohn K, Jung D Y, Lee H, et al. Efficient learning of sparse, distributed, convolutional feature representations for object recognition[C]. In Proceeding of 2011 IEEE International Conference on Computer vision. Barcelona: IEEE, 2011:2643-2650.

[25] Cui Zhen, Chang Hong, Shan Shiguang, et al. Joint sparse representation for videobased face recognition[J]. Neurocomputing, 2014, 135:306-312.

[26] Cambria E, White B, Jumping NLP curves: a review of natural language processing research[J]. IEEE Computational Intelligence Magazine, 2014, 9(2):48-57.

[27] Collobert R, Weston J. Aunified architecture for natural language processing: deep neural networks with multitask learning[C]. In Procedings of 25th International Conference on Machine Learning. Helsinki, Finland: Association for Computing Machinery, 2008:160-167.

[28] Chongyuan Tao, Jian Zhang, Pan Wang, Smoke Detection Based on Deep Convolutional Neural Networks[C]. In: IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society, 2016:150-153

[29] CVPR Lab. at Keimyung University.Wildfire smoke ideo Database [DB/OL]. http://cvp- r.kmu.ac.kr/, 2012/2015-12-20.

[30] http://signal.ee.bilkent.edu.tr/VisiFire/Demo/SmokeClips/ 2014/2015-12-20. Cetin, E., Computer Vision Based Fire Detection Dataset [DB/OL].

[31] University of Salerno.Smoke Detection Dataset [DB/OL]. http://mivia.unisa.it/, 2012-7/2015-12-20.

[32] State Key Lab of Fire Science. At University of Science and Technology of China [DB/OL]. http://staff.ustc.edu.cn/~yfn/vsd.html

[33] P. KaewTraKulPong, R. Bowden.An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection[J]. Computer Vision and Distributed Processing.2002: 135-144

[34] Chris S, W.E.L Grimson. Adaptive background mixture models for real-time tracking[C]. In Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference.1999

[35] Thou-Ho Chen, Yen-Hui Y., Shi-Feng H. et al. The Smoke Detection for Early Fire-Alarming System Base on Video Processing[C]. In Proceedings of the 2006 International Conference on Intelligent Information Hiding and Multimedia Signal Processing.2006

[36] B., A. B. A novel video-based smoke detection method based on color invariants[C]. In 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).2016.

[37] Hinton, G. E. and Salakhutdinov, R. R.Reducing the dimensionality of data with neural networks. Science, Vol. 313. no. 5786, pp. 504 - 507, 28 July 2006.

[38] David H Hubel and Torsten N Wiesel. Receptive fields of single neurones in the cat's striate cortex. The Journal of physiology, 148(3):574–591, 1959.

[39] David H Hubel and Torsten N Wiesel. Receptive fields, binocular interactionand functional architecture in the cat's visual cortex. The Journal of physiology, 160(1):106–154, 1962.

[40] Hubel D H, Wiesel T N. Receptive fields of single neurones in the cat's striate cortex[J]. The Journal of Physiology, 1959, 148(3): 574-591.

[41] Kunihiko Fukushima. Neocognitron: A self-organizing neural network modelfor a mechanism of pattern recognition unaffected by shift in position. Biological cybernetics, 36(4):193–202, 1980.

[42] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard EHoward, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. Neural computation, 1(4):541–551, 1989.

[43] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradientbased earning applied to document recognition. Proceedings of the IEEE,86(11):2278–2324, 1998.

[44] Fei-Yue W., Jun Jason Z., Xinhu Z. et al. Where does AlphaGo go, from churchturing thesis to AlphaGo thesis and beyond. [J]. IEEE/CAA Journal of Automatica Sinica, 2016, 3(2): 113-120.

[45] Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing coadaptation of feature detectors[J]. arXiv preprint arXiv:1207.0580, 2012.

[46] Kingma D, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprintarXiv:1412.6980, 2014.

[47] C. Tao, J. Zhang and P. Wang, "Smoke Detection Based on Deep Convolutional Neural Networks," 2016 International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII), Wuhan, 2016, pp. 150-153, doi: 10.1109/ICIICII.2016.0045.

[48] Computer Vision and Pattern Recognition Laboratory, Keimyung University, Korea https://cvpr.kmu.ac.kr/

[49] Signal Processing Group, Bilkent University, http://signal.ee.bilkent.edu.tr/

[50] Mivia Lab, University of Salerno, https://mivia.unisa.it/datasets/

[51] State key lab of fire science, University of science and technology of China, http://staff.ustc.edu.cn/~yfn/

[52] Kaewtrakulpong, Pakorn & Bowden, Richard. (2002). An Improved Adaptive Background Mixture Model for Realtime Tracking with Shadow Detection. Proceedings of 2nd European Workshop on Advanced Video-Based Surveillance Systems; September 4, 2001; London, U.K. 10.1007/978-1-4615-0913-4_11.

[53] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149), Fort Collins, CO, USA, 1999, pp. 246-252 Vol. 2, doi: 10.1109/CVPR.1999.784637.

[54] Wangdan, Liuhuai. Background Extraction and Updating Based on Improved Gaussian Mixture Model. Journal of Nanjing Normal University, 2015(02): 60-64.

[55] GUO Wei, GAO Yuanyuan, LIU Xinyan. Improved moving object detection method based on mixture Gaussian model. Computer Engineering and Applications, 2016, 52(13): 195-200.

[56] Panchaofeng, Yangshusen, Chenning. Fire smoke recognition based on image entropy. Journal of Jiangsu University of Science and Technology, 2015, (01): 52-57.

[57] T. Chen, Y. Yin, S. Huang and Y. Ye, "The smoke detection for early fire-alarming system based on video processing," 2006 International Conference on Intelligent Information Hiding and Multimedia, Pasadena, CA, USA, 2006, pp. 427-430, doi: 10.1109/IIH-MSP.2006.265033.

[58] O. Besbes and A. Benazza-Benyahia, "A novel video-based smoke detection method based on color invariants," 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, 2016, pp. 1911-1915, doi: 10.1109/ICASSP.2016.7472009.

[59] Hinton, G.E. & Salakhutdinov, R.R.. (2006). Reducing the Dimensionality of Data with Neural Networks. Science (New York, N.Y.). 313. 504-7. 10.1126/science.1127647.

[60] HUBEL DH, WIESEL TN. Receptive fields of single neurones in the cat's striate cortex. J Physiol. 1959;148(3):574-591. doi:10.1113/jphysiol.1959.sp006308

[61] HUBEL DH, WIESEL TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol. 1962;160(1):106-154. doi:10.1113/jphysiol.1962.sp006837

[62] HUBEL DH, WIESEL TN. Receptive fields of single neurones in the cat's striate cortex. J Physiol. 1959;148(3):574-591. doi:10.1113/jphysiol.1959.sp006308

[63] Fukushima, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol. Cybernetics 36, 193–202 (1980). https://doi.org/10.1007/BF00344251

[64] Y. LeCun et al., "Backpropagation Applied to Handwritten Zip Code Recognition," in Neural Computation, vol. 1, no. 4, pp. 541-551, Dec. 1989, doi: 10.1162/neco.1989.1.4.541.

[65] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," in Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998, doi: 10.1109/5.726791.

[66] F. Wang et al., "Where does AlphaGo go: from church-turing thesis to AlphaGo thesis and beyond," in IEEE/CAA Journal of Automatica Sinica, vol. 3, no. 2, pp. 113-120, April 2016, doi: 10.1109/JAS.2016.7471613.

[67] Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov:

Improving neural networks by preventing co-adaptation of feature detectors. CoRR abs/1207.0580 (2012)

[68] Kingma, Diederik & Ba, Jimmy. (2014). Adam: A Method for Stochastic Optimization. International Conference on Learning Representations.

RESEARCH RESULTS DURING THE STUDY PERIOD

1）An Objective Image Quality Evaluation and Its Applications for Low Illumination and Sudden Illumination Changes

International Journal of Engineering Research and Technology. Number5 (2020). Co-author.

2）Concrete Crack Detection using Relative Standard Deviation for Image Thresholding

International Journal of Engineering Research and Technology. Number10 (2020). First author.

3）A Lane Centerline Recognition System Based on Improved High Efficiency Hough transform

International Journal of Engineering Research and Technology. Number11 (2020). First author.

4）Night-time Vehicle Tracking Based on Brake/Tail Light Color

2019 SOC 학술대회 17 May 2019. Co-author.

5）A FPGA Verification of Improvement Edge Detection using Separation and Buffer Line

2020 International Conference on Electronics, Information, and Communication（ICEIC）19-22 Jan. 2020. First author.

6）Face Tracking Using Unscented Kalman Filter

2020 International Conference on Electronics, Information, and Communication（ICEIC）19-22 Jan. 2020. Co-author.

ABSTRACT

In recent years, large-scale wildfires have occurred one after another around the world, which reminds us of vehicle spontaneous combustion accidents that are closely related to our lives. with the development of new energy vehicles, the number of electric vehicles has increased year by year. However, due to many uncertain factors, the phenomenon of spontaneous combustion of electric vehicles occurs always, especially when electric vehicles parked in the outdoor parking lot spontaneously ignite, when a fire is formed, it is difficult to find out in time, once the fire spreads, it is difficult to extinguish it in a short time. The control of fire mainly lies in prevention. If some existing technology is used to identify and issue an early warning when the fire occurs, it can be extinguished before the fire expands. Smoke is the most obvious feature of a flame before it burns, so smoke detection can play a role in preventing fires from being unburned. Smoke has semi-transparent characteristics，its shape and texture characteristics are easily changed by external interference. These characteristics determine the difficulty of smoke detection. This thesis mainly uses computer vision and CNN technology, to detect and recognize the smoke generated by the early fire in the parking lot. The main research contents of this thesis are as follows.

1. Through the collection and sorting in the preliminary preparation stage, 12 sections of smoke videos used to detect the effect of the algorithm, include 12,470 smoke images and 12,902 non-smoke images used to training and test the CNN.

2. Analyze the general characteristics of the smoke in the early stage of the fire from the static (color) and dynamic aspects of the fire smoke.

3. Use CNN as a classifier to recognize smoke images.

The work of this thesis has the following innovation:

I. Data set. The data sets used in this thesis are the video clips recorded in the outdoor environment. The data sets can simulate the scene of the real fire smoke in the real situation, and it is more practical.

II. Smoke color feature extraction. In this thesis, I focus on the characteristics of HSV color space in smoke, which avoids the feature failure caused by the low quality of video in RGB color space.

III. CNN is used to detection smoke. Neural network can automatically learn the characteristics of the images through large number of training data, to avoid the trouble of the artificial selection of a single feature.

IV. Finally, a complete set of outdoor fire smoke detection system is designed, the system can be achieved on the outdoor of fire smoke early warning. Then, four groups of contrast experiments are used to verify the effectiveness of the proposed algorithm. The experimental results show that the proposed method has higher accuracy rate, lower false alarm rate and missing detection rate than other methods, also proves that the application of CNN to the field has certain research value. At the end of this thesis, the advantages and disadvantages of the proposed algorithm are summarized, the corresponding solutions and the direction of further research are put forward according to the shortcomings of the algorithm.

Keywords: **Smoke Detection; Image Processing; GMM; HSV; CNN.**