Automatic Detection and Inspection Robot of All-Weather Urban Comprehensive Corridor Based on Voiceprint Detection

Wei Ding(✉) and Ruining Hu

Anhui Institute of Information Enginee, Wuhu 241000, China
1428150543@qq.com

Abstract. Aiming at the fire safety hazards of urban comprehensive corridors, an all-weather urban comprehensive corridor automatic detection and inspection robot is designed, and voiceprint recognition technology and computer vision technology are combined to realize the full coverage and all-weather automatic detection of urban comprehensive corridors. The robot can independently plan the route and realize the identification and positioning of the fire escape, thereby realizing the real-time detection of the fire escape. The experimental results show that the robot can not only detect the fire passage but also carry out real-time early warning and positioning of the fire passage. Through the application of voiceprint recognition technology, the all-weather automatic detection of urban comprehensive corridors is realized. At the same time, through the application of image recognition technology and computer vision technology, the robot can independently plan the route and realize the real-time detection of fire escape.

Keywords: urban integrated corridor · Automatic detection · Voiceprint recognition · robot

1 Introduction

With the continuous acceleration of urbanization, the fire hazard of urban comprehensive corridors is gradually increasing. Although the detection technology for fire exits has been developed for many years, it also brings certain difficulties to the detection of fire exits because urban integrated corridors often appear at the intersection of different functional areas.[1] At the same time, because urban integrated corridors usually use a large amount of energy-intensive equipment, this also leads to easy fire accidents in urban integrated corridors.

Traditionally, fire escapes have been detected using cameras to monitor them and transmit video and images in real time. Due to the large number of cables and communication cables in the urban comprehensive corridor and the existence of complex environments such as interlacing, it is difficult to achieve fire escape detection targets using camera monitoring. For the real-time detection of fire escapes, [2] the main method currently used is to combine image recognition technology and computer vision technology to achieve real-time detection of fire escapes. On this basis, domestic scholars

M. Zhang et al. (Eds.): CCF NCCA 2023, CCIS 1959, pp. 273–281, 2024.
https://doi.org/10.1007/978-981-99-8764-1_21

put forward a new urban comprehensive corridor automatic detection technology, that is, real-time detection technology for fire passages, which mainly uses voiceprint recognition technology to convert sound signals into electrical signals and complete real-time detection of fire passages [3]. The method first uses voiceprint recognition technology to classify and identify various sounds in the urban comprehensive corridor and then collects, sorts and different kinds of sounds through robots to realize real-time detection of fire escapes [4]. However, due to the difficulty of sound classification and recognition in this method, there are also certain restrictions on the collection range of sound in this method. Therefore, by the characteristics of sound in urban comprehensive corridors, this paper designs an all-weather urban comprehensive corridor automatic detection robot based on voiceprint recognition technology and computer vision technology. The robot is designed to autonomously plan paths and enable real-time detection of fire escapes, thus providing a solution for fire safety issues in urban integrated corridors [5].

2 Inspection Robot System Design

The robot system consists of 1 main control unit, 2 motor modules, 1 voiceprint recognition module, 1 path planning module and 2 actuators. Among them, the main control unit and the motor module interact with the host computer through wireless data communication. The robot can realize the automatic detection of the fire passage through voice control of the camera lens and can independently plan the robot's travel route in the fire passage and identify and locate the fire escape in real time [6]. In addition, the robot can warn of the fire escape according to the shape of the fire passage and can also locate the fire passage according to the location of the fire passage to realize real-time detection of the fire escape. At the same time, the robot can update and identify the route in real time according to information such as fire passages, fire emergency plans, and historical fire records to realize the automatic detection of urban comprehensive corridors around the clock [7].

The system sends data to the upper computer program through the wireless module during normal operation, and the upper computer then transmits the data to the STM32 lower computer. The lower computer transmits a signal to drive the motor to start, and the lower computer program accepts a feedback signal to realize the adjustment of the walking speed so that the robot can walk in the urban comprehensive corridor (as shown in Fig. 1).

The sensor equipped with the robot uses a high-definition infrared thermal imager and a visible light camera to achieve real-time detection of temperature and brightness. High-precision sensors are used to construct a three-dimensional model of the environment within the detection range of the robot, real-time calculation of spatial coordinates and motion attitude parameters through the environmental model, and real-time trajectory tracking, trajectory planning and obstacle avoidance functions on the three-dimensional map [8]. When an abnormal situation is found, the abnormal facility can be located and identified through the camera screen and voice, and the result can be fed back to the background command center. When abnormal conditions occur, video replay, panoramic linkage, manual confirmation and other operations can be performed. The robot is equipped with a set of data processing terminals based on voiceprint recognition

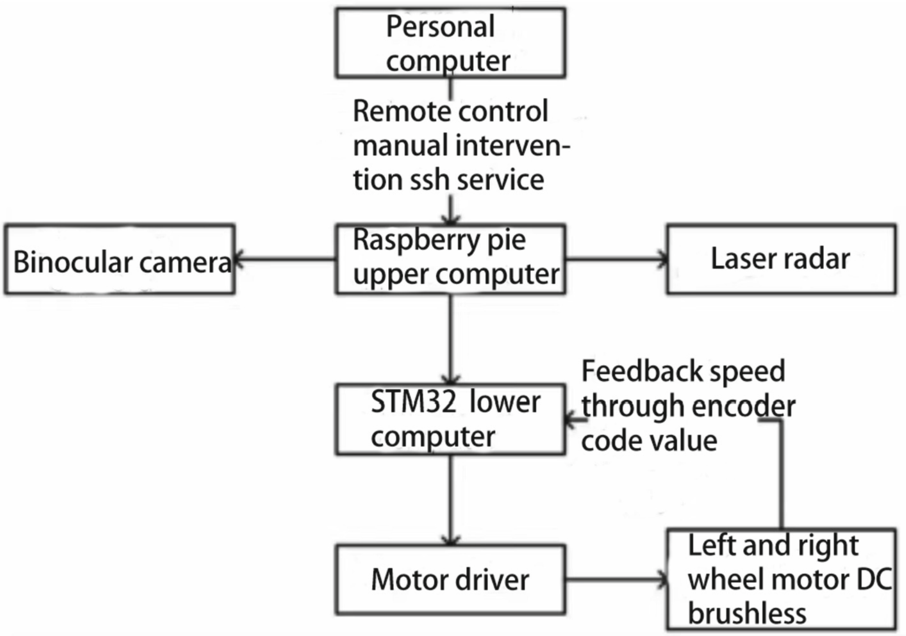

Fig. 1. Detecting the robot hardware system

technology, collects the sound in the environment through the microphone, and then uses the voiceprint recognition algorithm to process the collected sound signals to determine different sound sources [9]. Then, the sound signal heard is sent out through the speaker, and the voiceprint signal is collected through the microphone for processing. Finally, the processed voiceprint signal is sent to the host computer for judgment to realize the voiceprint recognition and early warning of the fire channel. At the same time, according to the length of the passage to be detected and the position information, the positioning of the fire escape is realized.

Path planning is one of the most important technologies to realize the autonomous walking of the robot. The path planning module includes three main parts of environment perception, obstacle perception and autonomous path planning, in which the environment perception module is responsible for collecting surrounding environmental information during the robot's travel, including sound, temperature and humidity, and processing this information after sending the information to the central processing unit through the wireless network [10]. The obstacle awareness module is responsible for sending the location and shape of obstacles to the central processing unit via the wireless network. The autonomous path planning module is the core module to realize robot path planning.

The path planning module obtains all speech samples in the current environment through the voiceprint recognition module, uses the Gaussian hybrid model method to perform speech classification recognition, and then obtains the speech samples in the current environment [11]. A Gaussian mixture model is a parametric probability density model expressed as a weighted sum of the density of Gaussian components. x is represented as a d-dimensional continuous value data vector, ω_i i = 1, M is the mixed weight, and its formula is:

$$P(x|\lambda) = \sum_{i=1}^{M} \omega_i g(x|\mu_i, \Sigma_i) \quad (1)$$

$$g(x|\mu_i, \Sigma_i) = \frac{1}{(2\Pi)^{D/2}|\Sigma_i|^{1/2}} exp\left\{-\frac{1}{2}(x-\mu_i)\prime\right\} \quad (2)$$

μ_i represents the mean vector Σ_i is the covariance $\sum_{i=1}^{M} \omega_i = 1$. After sampling by Gaussian mixture model, based on the analysis of the current environment, an optimal path planning scheme is formed, and the path information is sent to the central processing unit through the wireless network. The central processing unit makes decisions based on the path information obtained with the actual needs of the user and controls the robot to run according to a certain planned path.

3 Voiceprint Detection

Voiceprint refers to accurately identifying the voice and speaker to be recognized from the input voice and can also automatically determine the identity of the speaker according to the content and way of speaking. Voiceprint detection is mainly divided into two parts: target detection based on voiceprint recognition and segmentation based on target semantics, and its recognition system is composed of sound acquisition, preprocessing, feature extraction module and other modules, through the preprocessing, feature extraction and pattern classification of the collected voice signal, to realize the judgment of whether there is a specific person identity in the voice signal, and its system framework is shown in Fig. 2.

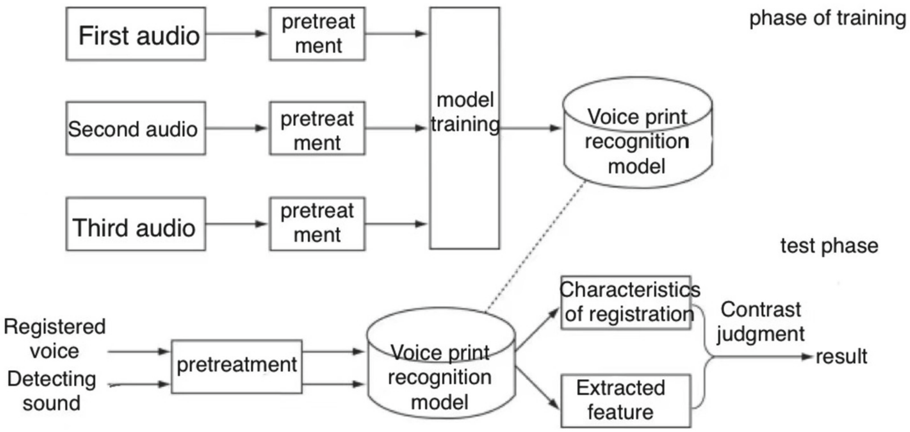

Fig. 2. Frame of voiceprint recognition system for inspection robot

This software uses to give full play to the system performance but also conducive to code modification and improvement, this software system is mainly divided into sound acquisition, main thread, registration and recognition overall process is shown in Fig. 3.

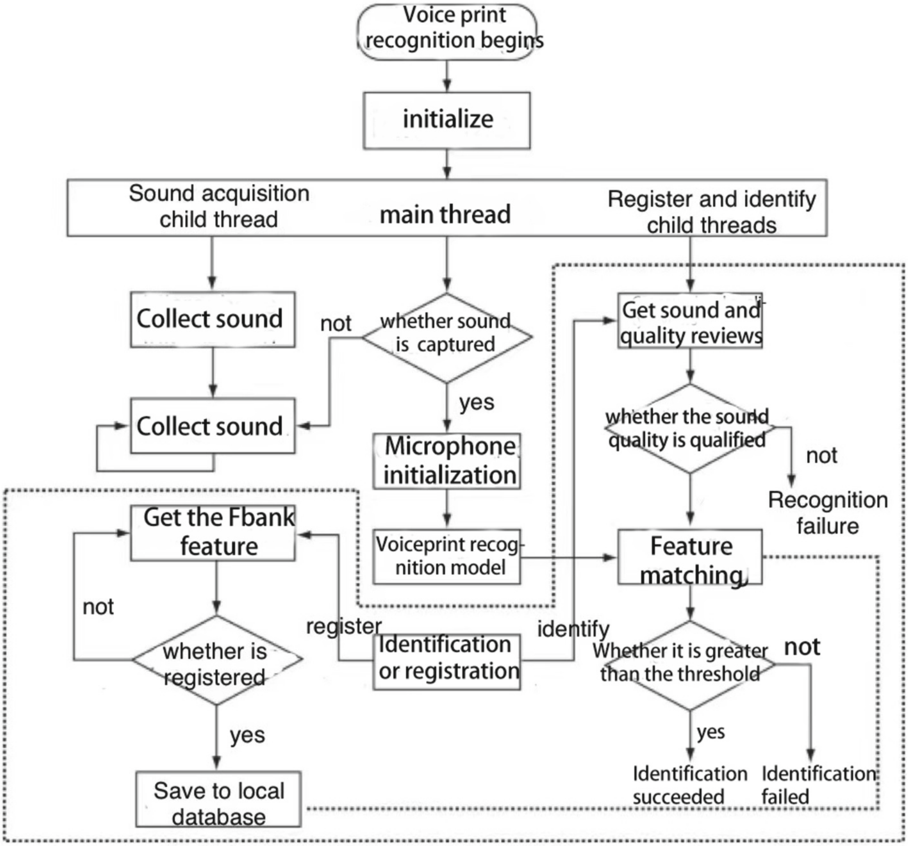

Fig. 3. Overall process of voiceprint recognition software

The technology has good robustness to various environments and background noise and can be effectively identified even in strong noise environments. Voiceprint recognition is a speech recognition method based on deep learning, computer vision, acoustics and other related technologies, which also has good performance in large noise environments. Voiceprint recognition technology can not only realize the determination of human identity but also realize the determination of people and objects and can distinguish between target objects and background sounds.

4 Key Technology of Voiceprint Recognition

The quality of the voiceprint recognition software system mainly depends on the voiceprint recognition algorithm, which refers to the application of speech recognition technology to process speech in a specific environment so that it can reflect the identity of the speaker. Similar to fingerprint recognition algorithms, voiceprint recognition algorithms are biotechnology that uses acoustic features for identity identification.

The voiceprint recognition is based on LFBank features use LFBank for feature extraction, firs emphasis the high-frequency part of the signal by a high-pass filter, and its formula is expressed as:

$$y(n) = x(n) - a * x(n-1) \quad (3)$$

where a is represented as the emphasis coefficient, and after the emphasis is performed, framing and windowing are carried out to increase the continuity between frames, reduce spectral leakage, and N represents the window length, its expression is:

$$\omega(n, a) = (1-a) - a * cos\left(\frac{2\pi n}{n-1}\right),\ 0 \leqq n \leqq N-1 \quad (4)$$

After framing and windowing, to perform a fast Fourier transform on each frame of data and transfer the signal from the time domain to the frequency domain for better identification of the sound signal, the formula is as follows:

$$x(k) = \sum\nolimits_{n=0}^{N-1} x(n)e^{\frac{-j2\pi k}{N}}, 0 \leqq k \leqq N-1 \quad (5)$$

$$p(k) = \frac{1}{N}|x(k)|^2 \quad (6)$$

The obtained signal spectrum is sent to the linear filter bank, the center frequency of the filter is represented by m, the resulting linear filter bank is shown in Eq. (7), and finally, the logarithmic energy processing of the formed spectrum can obtain the calculation formula of LFBank:

$$L_m(k) = \begin{cases} 0 & k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)}, f(m-1) & \leqq k \leqq f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)}, f(m) & < k \leqq f(m+1) \\ 0 & k > f(m+1) \end{cases} \quad (7)$$

$$S_L(m) = \ln\left(\sum\nolimits_{k=0}^{N-1} |X(k)|^2 L_m(k)\right), 0 \leqq m \leqq M \quad (8)$$

Since the working environment of the robot is dynamically changing, it is impossible to move according to the preset route every time, and the loss function is required to reduce the error rate, which combines the classification loss function based on the interval and the prototype cosine loss function based on small samples:

$$L = L_{AMS} + L_{CP} \quad (9)$$

$$L_{AMS} = -\frac{1}{N}\sum\nolimits_{n=1}^{N} log\frac{e^{S_{n,n}}}{\sum_{j=1}^{N} e^{S_{n,j}}} \quad (10)$$

$$L_{CP} = -\frac{1}{N}\sum\nolimits_{n=1}^{N} log \frac{e^{s\left(cos\theta_{y_i}-m\right)}}{e^{s\left(cos\theta_{y_i}-m\right)} + \sum\nolimits_{j\neq y_i} e^{scos\theta_j}} \tag{11}$$

The loss function clearly encourages the spacing of voiceprint classes to increase, and in the voiceprint feature, the metric space can be optimized by finding prototypes close to the target sample, the distance between the same kind can be reduced, and the loss function can effectively deal with the sample attribution problem and improve the robustness of the model.

5 Experimental Report

In this project, the functional test of the robot is carried out through experiments to verify the practicability of its system and the advanced algorithm of its algorithm. In this experiment, a highly sensitive PanasonicWM-62A acoustic sensor was used to acquire signals, and its parameters are shown in Table 1.

Table 1. Physical parameters of sensors

parameter	Parameter value
sensitivity	-45 ± 4 dB(0 dB = 1 V/Pa,1 kHz)
Signal-to-noise ratio(S/N)	$>$ 58 dB
impedance	$<$ 2.2k Ω
Detection frequency range	20 ~ 16000 Hz

In order to verify the effectiveness of the algorithm proposed in this paper, a robot platform is built in the laboratory, and the proposed algorithm is experimentally verified. In the experiment, a camera and a microphone were used to collect voiceprint images of the fire scene and fire escape, respectively, and saved as a database. By filtering and threshold segmentation of the images in the database, the voiceprint images of the fire passage and fire escape are obtained. A simulated environment was set up in the laboratory, and three fire escapes were set up accuracy of the voiceprint recognition effect in different pipeline environments was explored using a variety of environments and voiceprint recognition as independent variables, and the percentages in Table 2 the accuracy a total of 540 experiments.

Table 2 that the recognition rate reaches more than 90% under the effect of slight noise in the pipeline, and the robot can effectively detect the fire passage and fire passage and can also warn the fire passage after identifying it.

Table 2. Effects of voiceprint recognition system in various environments

environment	Time		
	3s	6s	10s
Silence in the pipe (approx. 25 dB)	95.0%	99.2%	99.5%
Slight noise in the pipe (approx. 15 dB)	90.5%	95.2%	97.2%
Large noise in the pipeline (approx. 0 dB)	80.8%	88.2%	92.1%

6 Conclusion

Aiming at the fire safety hazards of urban comprehensive corridors, this paper proposes an all-weather urban comprehensive corridor automatic detection and inspection robot, which has the following advantages:

(1) All-weather detection: Due to the characteristics of fixed position of sound sources when the robot extracts sound signals proposed in this paper, the robot can detect urban comprehensive corridors around the clock.

(2) Real-time early warning: when abnormal situations such as fire are detected, the robot can carry out real-time early warning and positioning of the fire escape;

(3) Autonomous route planning: The robot can autonomously plan the route and detect the fire escape in real time.

References

1. CUI Lin,WANG Zhiyue. Research on voiceprint recognition based on LFBank and FBank hybrid features. Comput. Sci.2022,49(S2):621–625
2. GUO Xin,LUO Chengfang,DENG Aiwen. Design and implementation of voiceprint recognition system in open scene based on deep learning [J] J Nanjing Univ. Inform. Sci. Technol. (Natural Science Edition) 2021, 13(05):526–532.DOI:https://doi.org/10.13878/j.cnki.jnuist.2021.05.003
3. Cheung S Y, Pravin v. Traffic surveillance by wireless sensor networks: final report[R].University of California,Berkeley,Jan.2007
4. LIN Linghui,ZHANG Yongfu,ZHANG Xinyue,ZHANG Fangwei,XIA Mingqian. Design of security system based on voiceprint recognition [J].Inf. Communication, 2020(06):112–114
5. Sun, Z., Bebis, G., Miller, R.: On-road vehicle detection: a review. IEEE Trans. Pattern Anal. Mach. Intell. **28**(5), 694–711 (2006). https://doi.org/10.1109/TPAMI.2006.104
6. ZHANG Yujie,ZHANG Zan. Application of DenseNet in voiceprint recognition [J].Computer Engineering and Science,2022,44(01), pp .132–137
7. Limin, L.I.U., Jiwu, J.I.N.G.: Rapidly developing biometric identification standard specification in China [J]. China Inform. Secur. **2**, 68–72 (2019)
8. Reynolds, D.A., Rose, R.C.: Robust text-independent speaker identification using Gaussian mixture speaker models [J]. IEEE Trans. Speech Audio Process.S **3**(1), 72–83 (1995)
9. Desplanques B, Thienpondt J, Demuynck K. ECAPATDNN: emphasized channel attention, propagation and aggregation in TDNN Based Speaker Verification [J] . arXiv e-print, 2020, arXiv: 2005, p. 07143

10. Nagrani A, Chung J S, Zisserman A. VoxCeleb: a largescale speaker identification dataset[C] ‖ Interspeech.In: ISCA: ISCA, 2017. DOI: 10. 21437/interspeech., pp.2017–950 (2017)
11. Kye S M, Kwon Y, Chung J S. Cross attentive pooling for speaker verification [C] ‖2021 IEEE Spoken Language Technology Workshop (SLT) . January 19 - 22, 2021, Shenzhen, China. IEEE, 2021: 294-300