IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 15, 2019, accepted August 23, 2019, date of publication August 26, 2019, date of current version September 9, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2937660

Research on Audio-Visual Detection Method for Conveyor Belt Longitudinal Tear

CHENGCHENG HOU[1,4], TIEZHU QIAO[1,4], MEIYING QIAO[2], XIAOYAN XIONG[1,3], YI YANG[1,4], AND HAITAO ZHANG[1,4]

[1]Key Laboratory of Advanced Transducers and Intelligent Control System, Ministry of Education, Taiyuan University of Technology, Taiyuan 030024, China
[2]School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo 454000, China
[3]College of Mechanical and Vehicle Engineering, Taiyuan University of Technology, Taiyuan 030024, China
[4]College of Physics and Optoelectronics, Taiyuan University of Technology, Taiyuan 030024, China

Corresponding author: Tiezhu Qiao (qiaotiezhu@tyut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China-Shanxi Coal-Based Low-Carbon Joint Fund under Grant U1810121, in part by the Development Plan for Industrial Science and Technology of Jinzhong under Grant 201803D03111003, and in part by the Natural Science Foundation of China-Shanxi under Grant 201801D121180.

ABSTRACT As a key equipment for production and transportation, conveyor belts are widely used in the coal mining industry. Once the conveyor belt longitudinal tear occurs, it will seriously affect the production and even cause personal injury. Therefore, the longitudinal tear detection of the conveyor belt is extremely important. In this paper, the sound detection is introduced into the longitudinal tear detection of the conveyor belt, and an audio-visual detection method for conveyor belt longitudinal tear is proposed. Camera and microphone array are used to collect the image and sound signals of conveyor belt, and the conveyor belt is detected from both image and sound, respectively. Then the image and sound analysis results are combined to comprehensively judge the status of the conveyor belt. The experimental results show that the audio-visual detection method can accurately identify the normal, abnormal, and longitudinal tear of the conveyor belt. The detection accuracy is over 86.72% and the sensitivity of longitudinal tear detection is greater than 92.59%. The proposed audio-visual detection method is verified to meet the requirements of longitudinal tear detection of coal mining industry conveyor belts.

INDEX TERMS Audio-visual detection method, conveyor belt, longitudinal tear.

I. INTRODUCTION

Conveyor belts are one of the important equipment for coal mine production, which are widely used in coal production and transportation. Due to the influence of the materials transported, such as gangue, anchors, etc., the conveyor belts are prone to longitudinal tear [1], [2]. Once the conveyor belt longitudinal tear occurs, it will seriously affect the production of coal mine and even cause personal injury. Therefore, longitudinal tear detection of conveyor belts is extremely important [3]. The existing detection methods for conveyor belt longitudinal tear are: mechanical detection method [4], electromagnetic detection method [5], ultrasonic detection method [6], and image-based detection method [7]. The current study on the longitudinal tear detection method for conveyor belt is mainly image-based detection method [8]. Yang *et al.* [9] developed a visual inspection system for detecting the longitudinal tear of conveyor belts by using the industrial line-array CCD camera and high-brightness linear light sources. Qiao *et al.* [10] proposed a method of Integrative Binocular Vision Detection that based on infrared and visible fusion to detect the longitudinal tear of conveyor belts. Yu *et al.* [11] detected the longitudinal tear of conveyor belts by combining the mid-infrared and long infrared vision. Andrejiova *et al.* [12] used the classification tree and the regression analysis to create the damage classification model for the analysis and the classification of the conveyor belt damage. The image-based detection method has a fast detection speed, and the recognition accuracy is high in the case where the industrial site environment is good [13]. However, the image-based detection method is easily affected by the surface condition of the conveyor belt. Once the surface condition of the conveyor belt is relatively bad, the recognition accuracy rate will drop rapidly, and wrong actions such as

The associate editor coordinating the review of this article and approving it for publication was Jingchang Huang.

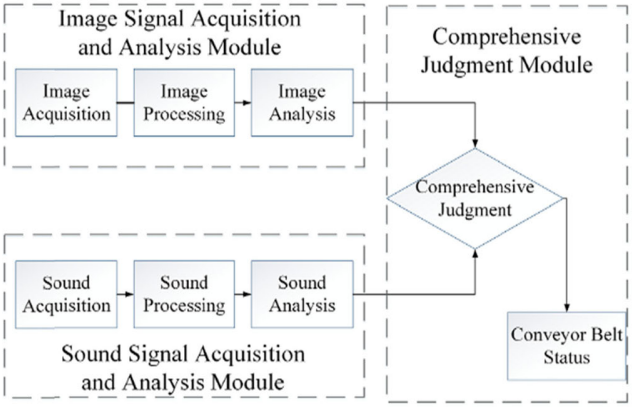

FIGURE 1. Overall flow of the audio-visual detection method.

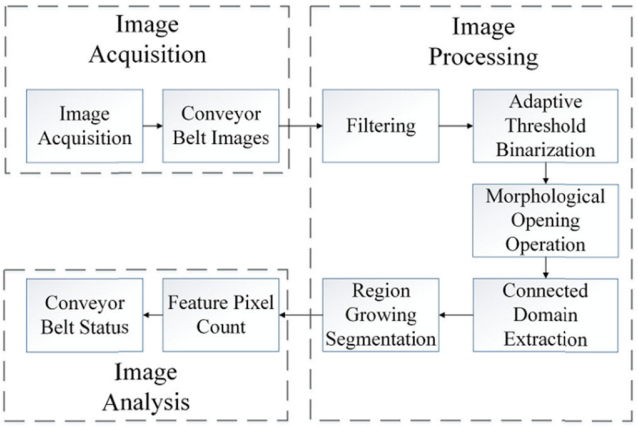

FIGURE 2. The specific procedure of image detection.

false alarms and false stops will occur, which will affect the normal production. Frequent emergency stops will also cause damage to the conveyor drive motor [14].

Same as the text, images, and film, sound is one of the important information carriers. It is widely distributed in daily production and life, and the information it transmits accounts for a large proportion of all information [15]. As an important source of information for human to perceive the environment, sound can reflect the sound events in the environment and judge whether the events in the environment are normal or not [16]. At present, once the conveyor belt longitudinal tear occurs, the nearby workers always found the longitudinal tear through abnormal sound. Therefore, it is feasible to introduce sound into the conveyor belt longitudinal tear detection. This paper proposes an audio-visual detection method for conveyor belt longitudinal tear. Camera and microphone array are used to collect the audiovisual signals of the conveyor belt. The collected images and sounds are processed separately by using the dual-thread. Then the current status of the conveyor belt is determined by combining the image and sound analysis results, and the corresponding operation is performed to the conveyor belt [17]. The audio-visual detection method for conveyor belt longitudinal tear is mainly divided into three modules: image signal acquisition and analysis module, sound signal acquisition and analysis module, and comprehensive judgment module based on audiovisual analysis results. The overall flow of the audio-visual detection method is shown in Fig.1.

The remainder of this paper is organized as follows. Section II describes the image detection of the conveyor belt. Section III describes the sound detection of the conveyor belt. Section IV shows the status analysis of the conveyor belt. Section V presents the experimental method and steps. The experimental results and analysis are shown in Section VI. Finally, the conclusions and prospects of this paper are given in Section VII.

II. IMAGE DETECTION

Conveyor belt longitudinal tear is generally caused by the falling impact of the long iron or hard flaky gangue that mixed in coal during the reloading [18]. Conveyor belt longitudinal tear occurs mostly at the conveyor belt loading point [19]. Therefore, we can use the camera to capture the lower surface image of the upper conveyor belt. The camera is installed at the loading point. Collect and analyze the conveyor belt image to determine whether the conveyor belt has longitudinal tear. The image detection of the conveyor belt longitudinal tear mainly includes three parts: image acquisition, image processing, and image analysis. The specific procedure of image detection is shown in Fig.2.

A. IMAGE ACQUISITION

The camera for collecting the images of the conveyor belt is mounted nearby the conveyor belt loading point and is mounted between the upper conveyor belt and the lower conveyor belt at a 45-degree angle to the conveyor belt. There is a large amount of dust and water in the operating environment of the conveyor belt, which will seriously affect the quality of the collected image, thus affecting the detection of the conveyor belt. In order to reduce the influence of water and dust on the collected images, it is necessary to install a dust removal and water removal equipment. The cleaning equipment is mounted on the front side of the camera to minimize the effects of environmental factors on the captured images. Fig.3 shows the installation of the camera and cleaning equipment.

The camera captures the lower surface image of the upper conveyor belt in real time. The normal image of the conveyor belt, the damaged image of the conveyor belt, and the longitudinal tear image of the conveyor belt are shown in Fig.4.

B. IMAGE PROCESSING

The collected conveyor belt images need to be pre-processed in real time to facilitate the analysis and judgment of the conveyor belt status in the later stage. The processing flow of the conveyor belt images is shown in Fig.2. Assuming that the image of the conveyor belt captured by the camera is $g(x, y)$, the procedure of image processing for $g(x, y)$ is as follows:

- Filtering. The original image $g(x, y)$ is smoothed by using median filtering [20] to eliminate noise, resulting in $m(x, y)$.

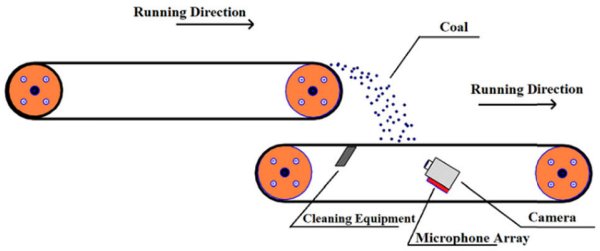

FIGURE 3. The installation of the camera and cleaning equipment.

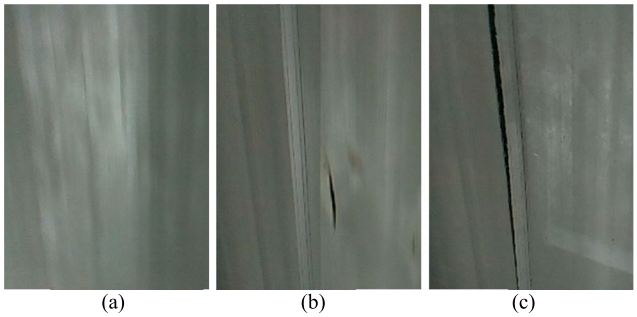

FIGURE 4. Original images of conveyor belt: (a) Normal image. (b) Damaged image. (c) Longitudinal tear image.

- Adaptive threshold binarization. For $m(x, y)$, the maximum inter-class variances method-Otsu's method [21] is used to generate the binary image $Otsu(x, y)$. Fig.5 shows the results of adaptive threshold binarization.
- Morphological opening operation [22]. It is judged whether there are white pixels on the binary image $Otsu(x, y)$. If it exists, the image $Otus(x, y)$ is subjected to morphological opening operation to further eliminate the noise point, resulting in $Open(x, y)$. The results of morphological opening operation are shown in Fig.6 (a) and Fig.6 (b).
- Connected domain extraction. 4-connections is used to find the connected area in the image $Open(x, y)$ [23]. Marking the center of gravity of the connected domain $c(i, j)$. The results of connected domain extraction are shown in Fig.6 (c) and Fig.6 (d).The red dot in the result image is the center of gravity.
- Region growing segmentation. Taking the center of gravity coordinate $c(i, j)$ of the connected domain as the growing seed point. The original image $g(x, y)$ is subjected to region growing segmentation to obtain $R(x, y)$ [24], [25]. The results of region growing segmentation are shown in Fig.6 (e) and Fig.6 (f).

C. IMAGE ANALYSIS

After image processing is completed, the image feature pixels are counted to determine the conveyor status. The number of white pixels ($number\_wp$) in the results image $R(x, y)$ of region growing segmentation is counted. Thresholds t_1 and t_2 are set to determine the status of the conveyor belt. The specific determination method is: when $number\_wp$ in the image $R(x, y)$ is less than the threshold t_1, the conveyor belt is determined to be normal. When $number\_wp$ in the image $R(x, y)$ is greater than the threshold t_1 and less than the thresholdt_2, it is determined that the conveyor belt is damaged. When $number\_wp$ in the image $R(x, y)$ is greater than the threshold t_2, it is determined that the conveyor belt is longitudinal tear. The judgment formula is as follows:

$$\begin{cases} number\_wp \le t_1, & Conveyor\ Belt\ Normal \\ t_1 < number\_wp \le t_2, & Conveyor\ Belt\ Damaged \\ number\_wp > t_2, & Conveyor\ Belt\ Longitudinal\ Tear \end{cases} \quad (1)$$

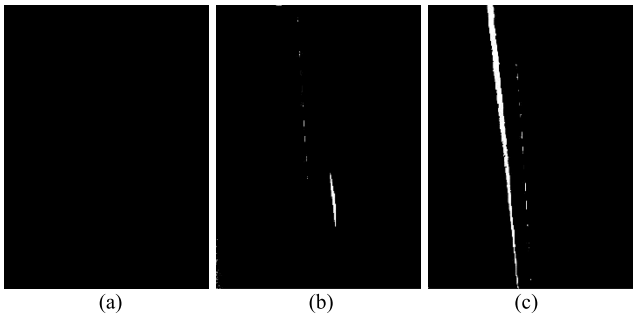

FIGURE 5. The results of adaptive threshold binarization: (a) Normal image. (b) Damaged image. (c) Longitudinal tear image.

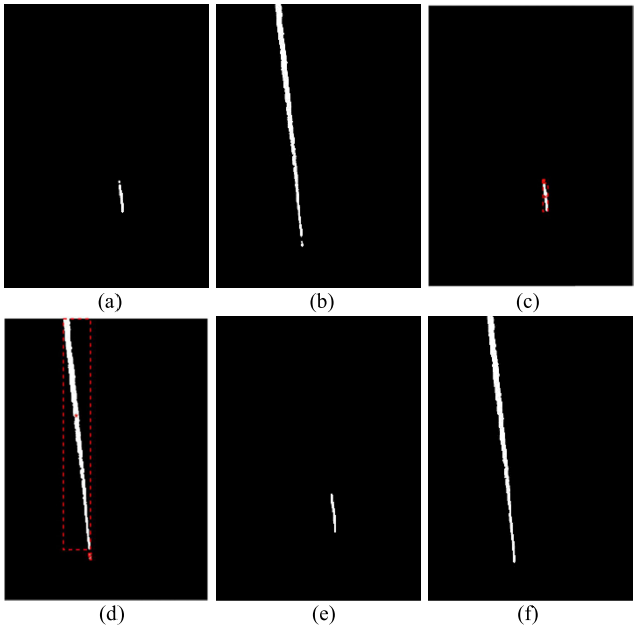

FIGURE 6. The results of image processing: (a) Damaged image morphological opening operation. (b) Longitudinal tear image morphological opening operation. (c) Damaged image connected domain extraction. (d) Longitudinal tear image connected domain extraction. (e) Damaged image region growing segmentation. (f) Longitudinal tear image region growing segmentation.

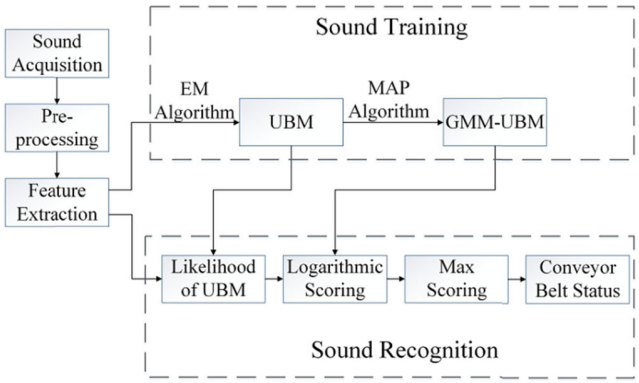

FIGURE 7. The specific procedure of sound detection.

III. SOUND DETECTION

Introducing the sound detection and classification into conveyor belt longitudinal tear detection is an important innovation in this paper. When the running conveyor belt is longitudinal tear, its sound will be very different from the normal running sound. Considering that the industrial site environment of the conveyor belts is more noise, it is difficult to collect a large number of conveyor belt longitudinal tear sounds for sound training to generate the longitudinal tear training database. In this paper, the GMM-UBM (Gaussian Mixture Model-Universal Background Model) [26] method based on the combined features is used for training and recognition of conveyor belt running sound and longitudinal tear sound. The sound detection of the conveyor belt longitudinal tear using GMM-UBM method is mainly divided into four parts: sound acquisition and pre-processing, feature extraction, sound training, and sound recognition. The specific procedure of sound detection of the conveyor belt longitudinal tear is shown in Fig.7.

A. SOUND ACQUISITION AND PRE-PROCESSING

We used a 4-microphone array to capture the conveyor belt running sounds in this paper. The microphone array was fixed at the bottom of the camera. The specific installation position of the microphone array is shown in Fig.3. During the sound detection, in order to obtain a sound signal conforming to the system standard, the original sound signals should be pre-processed. The pre-processing of the sound signals mainly includes pre-emphasis, framing, and windowing.

1) PRE-EMPHASIS

The collection of sound signals is often affected by various conditions. The sound power spectrum will drop dramatically as the frequency increases, so the collected sound signals need to be pre-emphasized. Pre-emphasis can enhance the spectrum of the falling part of the sound signals, thereby maintaining the flatness of the signals spectrum and preparing for the subsequent spectrum analysis work [27]. Pre-emphasis is usually implemented by using a first-order digital filter as shown in (2).

$$H(z) = 1 - \mu z^{-1} \tag{2}$$

where μ is a pre-emphasis factor, which is between 0 and 1. It is 0.96 in this paper. z is the digital frequency response. $H(z)$ is the transformation of the frequency response.

2) FRAMING AND WINDOWING

The collected conveyor belt sound signal is a non-stationary time-varying signal, but it can be considered as stable in a short period of time, i.e. the sound signal has a short-term quasi-stationary characteristic. A short-smooth sound segment can be intercepted from a continuous sound of a fixed characteristic as a frame, and the frame sound signal can be used to replace the continuous sound with a fixed characteristic, this process is called framing [28]. In this paper, the duration of each frame of sound is set as 20ms, the frame length is set as $N = 256$, and the frame shift is set as $T = 100$. After the framing processing of the sound, in order to maintain the smoothness of the two ends of each frame of the sound, the sound frames need to be further windowed. The commonly used window functions are the rectangular window and the Hamming window. The bandwidth of the Hamming window is large, and the high-frequency component is relatively small. Therefore, the window function is Hamming window in this paper. The Hamming window function is as follows:

$$w(n) = \begin{cases} 0.54 - 0.46\cos[2\pi n/(N-1)], & 0 \le n \le N \\ 0, & others \end{cases} \tag{3}$$

where N is the frame length. After the window function is determined, the pre-emphasized sound signal is framed and windowed, and the pre-processed sound signal is:

$$x_n(m) = x(n+m) * w(m) 0 \le m \le N-1,\ n = 0, 1T, 2T, \cdots \tag{4}$$

where $x(m)$ is the time domain signal of the conveyor belt sound, $w(m)$ is the window function, $x(n+m)$ is the n-th frame sound signal after the framing process, $x_n(m)$ is the n-th frame sound signal after the framing and windowing process, and N represents the frame length, T is the frame shift.

B. FEATURE EXTRACTION

During the short sound target recognition, single feature parameter cannot adequately characterize the target sound, resulting in poor performance. Using the combination of features instead of single feature can increase the effective feature dimension to compensate for the lack of feature samples. This paper uses a combination of short-time energy and Mel-frequency Cepstral Coefficients (MFCC) as the feature parameters.

1) SHORT-TIME ENERGY

From the perspective of sound energy, the sound energy of the conveyor belt longitudinal tear is different from the sound energy of the conveyor belt running. The energy of the longitudinal tear sound segment is greater than the energy of

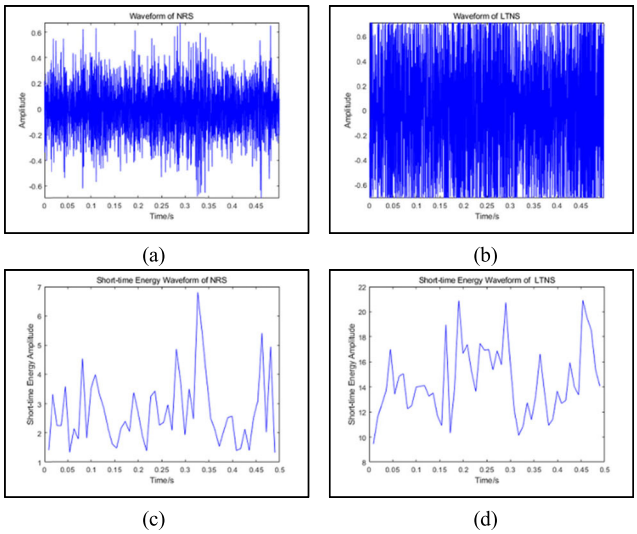

FIGURE 8. Conveyor belt sound waveform: (a) Waveform of NRS (No-load Running Sound). (b) Waveform of LTNS (Longitudinal Tear Sound with No-load). (c) Short-time Energy Waveform of NRS. (d) Short-time Energy Waveform of LTNS.

the running sound segment, wherein the energy of the longitudinal tear sound segment is a superposition of the energy of the running sound segment and the energy of the longitudinal tear sound segment. Therefore, we can distinguish between the longitudinal tear sound segment and the running sound segment by short-time energy. The sound segments of the conveyor belt mainly include the sound of the conveyor belt running at different speed with no-load and load. Fig.8 shows the waveform and the short-time energy waveform of the no-load running sound and the longitudinal tear sound with no-load. The calculation method of short-time energy is: assuming that the short-time energy of the n-th frame sound signal $x_{\mathrm{n}}(m)$ is E_{n}, the calculation formula is as follows [29]:

$$E_n = \sum_{m=0}^{N-1} x_n^2(m) \tag{5}$$

2) MFCC

The MFCC is based on the auditory mechanism of the human ear. The human ear just like a filter bank with non-uniform distribution. When the sound frequency is less than $1000Hz$, there are many filters. However, when the sound frequency is greater than $1000Hz$, the number of filters decreases, and the linear relationship becomes a logarithmic relationship. The Mel filter bank is a series of triangular bandpass filter bank that simulate human ear perception. The relationship between the Mel frequency and the actual frequency is: $Mel(f) = 2595 * \log(1 + f/700)$ [30]. The procedure of MFCC feature extraction is shown in Fig.9.

The specific steps for extracting the MFCC feature parameters from the collected conveyor belt sounds are as follows:

- Pre-processing. Pre-processing operation on the original sound signal to obtain the framed sound signal $x_n(m)$.

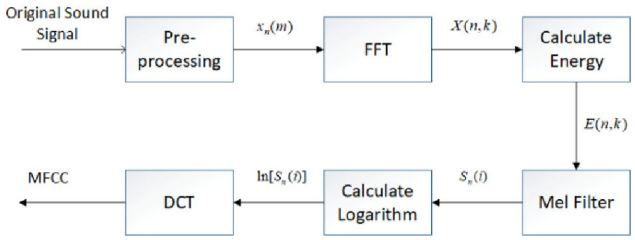

FIGURE 9. The procedure of MFCC feature extraction.

- Fast Fourier Transform (FFT) [31], which converts the sound signal $x_n(m)$ from a time domain signal to a frequency domain signal to obtain linear spectrum $X(n, k)$:

$$X(n, k) = FFT[x_n(m)] \tag{6}$$

  where k represents the k-th spectral line in the frequency domain.
- Calculate the energy of the spectral line $X(n, k)$ and obtain the discrete power spectrum $E(n, k)$:

$$E(n, k) = [X(n, k)]^2 \tag{7}$$

- Transform the discrete power spectrum $E(n, k)$ through the Mel filter and calculate the energy $S_n(i)$ in the Mel filter:

$$S_n(i) = \sum_{k=0}^{N-1} E(n, k)H_i(k), \quad 0 \le i \le M \tag{8}$$

  where H_i(k) is the frequency response of each triangle filter and its function is:

$$H_i(k) = \begin{cases} 0, & k < f(i-1) \\ \dfrac{k - f(i-1)}{f(i) - f(i-1)}, & f(i-1) \le k \le f(i) \\ \dfrac{f(i+1) - k}{f(i+1) - f(i)}, & f(i) \le k \le f(i+1) \\ 0, & k > f(i+1) \end{cases} \tag{9}$$

  where $f(i)$ is the center frequency, M is the number of triangular filters, which is generally between 24 and 40. It is 30 in this paper.
- Calculate the logarithm. Take the natural logarithm of the energy $S_n(i)$ in the Mel filter, obtain $\ln[S_n(i)]$.
- Discrete cosine transform (DCT) of $\ln[S_n(i)]$ and obtain the MFCC feature parameter $C(n)$ which characterizing the sound characteristics. The specific formula is:

$$C(n) = \sum_{i=0}^{M-1} \ln[S_n(i)] \cos\left(\frac{\pi l(i-0.5)}{M}\right), \quad l = 0, 1, 2, \cdots, L \tag{10}$$

  where L refers to the order of the MFCC feature parameter, which is generally between 12 and 16. In this paper $L = 12$, i.e. the extracted MFCC is a 12-dimensional feature vector.

In this paper, the short-time energy E_n and MFCC feature parameter are mixed to obtain the feature vector E_n_MFCC.

The sequence of the feature vector E_n_MFCC is: the short-time energy E_n is the first dimension of the feature vector E_n_MFCC, and the 12-dimensional MFCC is the second dimension to the thirteenth dimension of the feature vector E_n_MFCC.

C. SOUND TRAINING

GMM-UBM is a high-order GMM (Gaussian Mixture Model), which is adaptively obtained based on training sounds. It can reduce the different between recognition sounds and training sounds, and improve the system recognition rate [32]. During the sound training, the collected conveyor belt sounds are first pre-processed, and the 13-dimensional feature vector E_n_MFCC of the conveyor belt sound is extracted. Then the partial sounds are used to train the UBM (Universal Background Model). The GMM about conveyor belt status are then adapted from the UBM using the maximum a posteriori (MAP) estimation. The adapted GMM established in this paper mainly includes three types: conveyor belt normal GMM, conveyor belt unusual GMM, and conveyor belt longitudinal tear GMM. Finally, during the sound recognition, the features of the sound to be recognized are extracted and matched with the adapted GMM. The final output score is the difference between the recognition sounds and adapted GMM, which is used to determine the conveyor belt status. The following description of the GMM-UBM method cites the methods described in references [26] and [28].

GMM is a linear weighted combination of a certain number of Gaussian probability density functions. The probability density function of an N-order GMM is a linear combination of N single Gaussian distributions, which is used to describe the distribution of frame features in the feature space [33], as shown in (11).

$$P(X/\lambda) = \sum_{i=1}^{N} \omega_i b_i(X), \quad i = 1, 2, \cdots, N \tag{11}$$

where X is a D-dimensional random vector, herein is the 13-dimensional feature vector E_n_MFCC extracted from conveyor belt sound. ω_i is a mixed weight that satisfies $\sum_{i=1}^{M} \omega_i = 1$. Each sub-distribution $b_i(X_t)$ is a D-dimensional joint Gaussian probability distribution, which can be expressed as:

$$b_i(X) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(X-\mu_i)^t \sum\nolimits_i^{-1} (X-\mu_i)\right\} \tag{12}$$

where $\sum_i$ is the covariance matrix, μ_i is the mean vector. A complete Gaussian Mixture Model parameter λ consists of mixed weight ω_i, mean vector μ_i, and covariance matrix $\sum_i$ [34], as shown in (13).

$$\lambda = \{\omega_i, \mu_i, \sum_i | i = 1, 2, \cdots, N\} \tag{13}$$

UBM is also a Gaussian mixture model, which is trained by many types of sound data. Therefore, UBM is a common reflection of all sound characteristics and environmental channels. The more training data sets of UBM and the wider the coverage, the better the final recognition. During sound training, it is first necessary to pre-processing and feature extract of the collected conveyor belt sounds, then train an UBM that is independent of the conveyor belt status sounds by using the Expectation Maximization (EM) algorithm, and finally obtain the adapted GMM of each conveyor belt status sound by using the MAP algorithm.

1) GENERATE UBM USING EM ALGORITHM

The parameter estimation algorithm of the GMM is a process of determining the model parameter λ according to the EM algorithm for a given set of training sound data. The EM algorithm is an iterative algorithm that estimates a new parameter $\hat{\lambda}$ through E step and M step. Make the likelihood under the new model parameters be greater than the likelihood under the original model parameters $P(X/\hat{\lambda}) \geq P(X/\lambda)$. The new model parameter is re-estimated as the current parameter for the next round. Repeat the iterative operation until the model converges [35]. Thereby ensuring the monotonous increase of the model likelihood, and finally generating a stable UBM model.

Assuming that a set of sound feature sequences of length T is $X = \{X_1, X_2, \cdots, X_T\}$, the functions for estimating the GMM parameters by using the EM algorithm are as follows:

Mixed weight ω_i of the i-th GMM:

$$\omega_i = \frac{1}{T} \sum_{t=1}^{T} P(i|X_t, \lambda) \tag{14}$$

Mean vector μ_i of the i-th GMM:

$$\mu_i = \frac{\Sigma_{t=1}^{T} P(i|X_t, \lambda) X_t}{\Sigma_{t=1}^{T} P(i|X_t, \lambda)} \tag{15}$$

Variance δ_i^2 of the i-th GMM:

$$\delta_i^2 = \frac{\Sigma_{t=1}^{T} P(i|X_t, \lambda) X_t^2}{\Sigma_{t=1}^{T} P(i|X_t, \lambda)} - \mu_i^2 \tag{16}$$

The posterior probability of the i-th component is:

$$P(i|X_t, \lambda) = \frac{\omega_i b_i(X_t)}{\Sigma_{i=1}^{N} \omega_i b_i(X_t)} \tag{17}$$

2) GENERATE ADAPTED GMM USING MAP ALGORITHM

Assuming that the feature vector of the target type sound is a sequence $X = \{x_t, t = 1, 2, \cdots T\}$, the specific steps of the MAP algorithm for generating the adapted target type sound GMM are as follows [36]:

- Calculate the probability of the t-th eigenvector x_t obtained from the i-th Gaussian component:

$$P(i|x_t, \lambda_{UBM}) = \frac{\omega_i b_i(x_t)}{\Sigma_{j=1}^{N} \omega_j b_j(x_t)} \tag{18}$$

- Using the $P(i|x_t, \lambda_{UBM})$ obtained by (18), the mixed weight, mean vector, and variance of the target type sound are calculated separately:

$$\omega_i = \sum_{t=1}^{T} P(i|x_t, \lambda_{UBM}) \tag{19}$$

$$\mu_i = \frac{1}{\omega_i} \sum_{t=1}^{T} P(i|x_t, \lambda_{UBM}) x_t \tag{20}$$

$$\delta_i^2 = \frac{1}{\omega_i} \sum_{t=1}^{T} P(i|x_t, \lambda_{UBM}) x_t^2 \tag{21}$$

- Update the GMM of the target type sound with new statistical parameters:

$$\omega_i^{'} = \left[\frac{\alpha_i^{\omega} n_i}{T} + (1 - \alpha_i^{\omega}) \omega_i \right] \gamma \tag{22}$$

$$\mu_i^{'} = \alpha_i^{m} E_i(x) + (1 - \alpha_i^{m}) \mu_i \tag{23}$$

$$\delta_i^{'2} = \alpha_i^{\upsilon} E_i(x^2) + (1 - \alpha_i^{\upsilon})(\delta_i^2 + \mu_i^2) - \mu_i^{'2} \tag{24}$$

where γ is used to adjust $\omega_i^{'}$ to make $\Sigma_{i=1}^{M} \omega_i^{'} = 1$. α_i is an adaptive parameter that regulates the proportion of new and old parameters, $\alpha_i = \omega_i/(\omega_i + \tau)$. Where τ refers to the degree of connection between the new target type sound model and the old UBM model. Usually, the value of τ is set between 8 and 20, which is 16 in this paper.

D. SOUND RECOGNITION

After the above steps, the adapted target type sound GMM are finally obtained: conveyor belt normal GMM, conveyor belt unusual GMM, and conveyor belt longitudinal tear GMM. During sound recognition, the likelihood output ratio $P(X_t|\lambda_{UBM})$ of the sound feature vector and the UBM model is first calculated, and then the likelihood output ratio $P(X_t|\lambda_{GMM-UBM})$ of the sound feature vector and the adapted target type sound GMM is calculated. Using logarithmic scoring, the likelihood of the sound to be recognized is the difference between the likelihood of the two models. The calculation formula is as follows:

$$S = \frac{1}{T} \sum_{t=1}^{T} \left[\log P(X_t|\lambda_{GMM-UBM}) - \left[\log P(X_t|\lambda_{UBM}) \right] \right] \tag{25}$$

where X_t is the combined feature vector of one frame of the sound to be recognized, $\lambda_{GMM-UBM}$ and λ_{UBM} represent the adapted target type sounds GMM and UBM, respectively, and S is the likelihood score of the final recognition. The larger the S value, the more likely the recognition sound comes from the target type sound. The decision threshold is set according to the experimental situation.

IV. CONVEYOR BELT STATUS ANALYSIS

Combined with the above image detection results and sound detection results, the current status of the conveyor belt can be obtained by comprehensive analysis and judgment. According to the analysis results, the status of the conveyor belt is mainly divided into conveyor belt normal, conveyor belt abnormal, and conveyor belt longitudinal tear. In order to quickly view the specific reasons for the conveyor belt abnormal and the conveyor belt longitudinal tear in the analysis results, the conveyor belt abnormal is subdivided into conveyor belt abnormal/sound (abnormalities caused by sound affect), conveyor belt abnormal/image (abnormalities caused by image affect), and conveyor belt abnormal (abnormalities caused by image and sound affect). Conveyor belt longitudinal tear is subdivided into conveyor belt longitudinal tear/sound (longitudinal tear caused by sound affect), conveyor belt longitudinal tear/image (longitudinal tear caused by image affect), and conveyor belt longitudinal tear (longitudinal tear caused by image and sound affect).

According to the analysis results of the conveyor belt, the corresponding operations are carried out to the conveyor belt. The operations are divided into conveyor belt keep running, common alarm, emergency alarm, and stop. The conveyor belt status analysis and the corresponding operations are shown in Table 1.

Fig.10 shows the overall flowchart of the audio-visual detection method for conveyor belt longitudinal tear.

V. EXPERIMENT

In order to verify the validity and accuracy of the proposed audio-visual detection method for conveyor belt longitudinal tear, an experimental conveyor belt platform was built in the laboratory. Fig.11 shows the experimental conveyor belt platform. The experimental platform uses a common steel cord conveyor belt. The length of conveyor belt is 23 m, width is 0.8 m, and thickness is 15 mm. The maximum running speed of the conveyor belt is 4 m/s. A steel chisel was installed between the conveyor belt frames. The steel chisel was installed between the upper and lower conveyor belt for causing longitudinal tear of the upper conveyor belt. The image acquisition sensor used in the experiment was a visible light camera with a resolution of 1200∗800, which meets the conveyor belt image detection requirements. The sound acquisition sensor was a 4-microphone array. The technical parameters of the microphone are: the sensitivity is -26dBFS (Decibels Full Scale), the acoustic overload point is 120 dBSPL(Decibels Sound Pressure Levels) ,and the SNR is 61dB. A personal PC was used for audio-visual signals analysis and processing. The specific configuration of the PC is: CPU is Inter Core i5-8400 2.8GHz, memory is 8GB, and the software development platform is Pycharm. The experiment used dual-thread to collect the images and sounds of the conveyor belt simultaneously. The image acquisition frequency was set as 10 frames/second. The sound sampling rate was set as 44100 Hz, and the duration of each sound segment was 500 ms. The schematic diagram of the experimental platform is shown in Fig.12.

The resolution of captured conveyor belt image is 800∗600. Each image corresponds to the conveyor belt range

TABLE 1. Conveyor belt status analysis and operation.

Image detection result	Sound detection result	Conveyor belt status	Operation
Normal	Normal	Normal	Keep Running
Normal	Unusual	Abnormal/Sound	Common Alarm
Normal	Longitudinal Tear	Longitudinal Tear/Sound	Emergency Alarm
Damaged	Normal	Abnormal/Image	Common Alarm
Damaged	Unusual	Abnormal	Common Alarm
Damaged	Longitudinal Tear	Longitudinal Tear/Sound	Emergency Alarm
Longitudinal Tear	Normal	Longitudinal Tear/ Image	Emergency Alarm
Longitudinal Tear	Unusual	Longitudinal Tear/ Image	Emergency Alarm
Longitudinal Tear	Longitudinal Tear	Longitudinal Tear	Stop

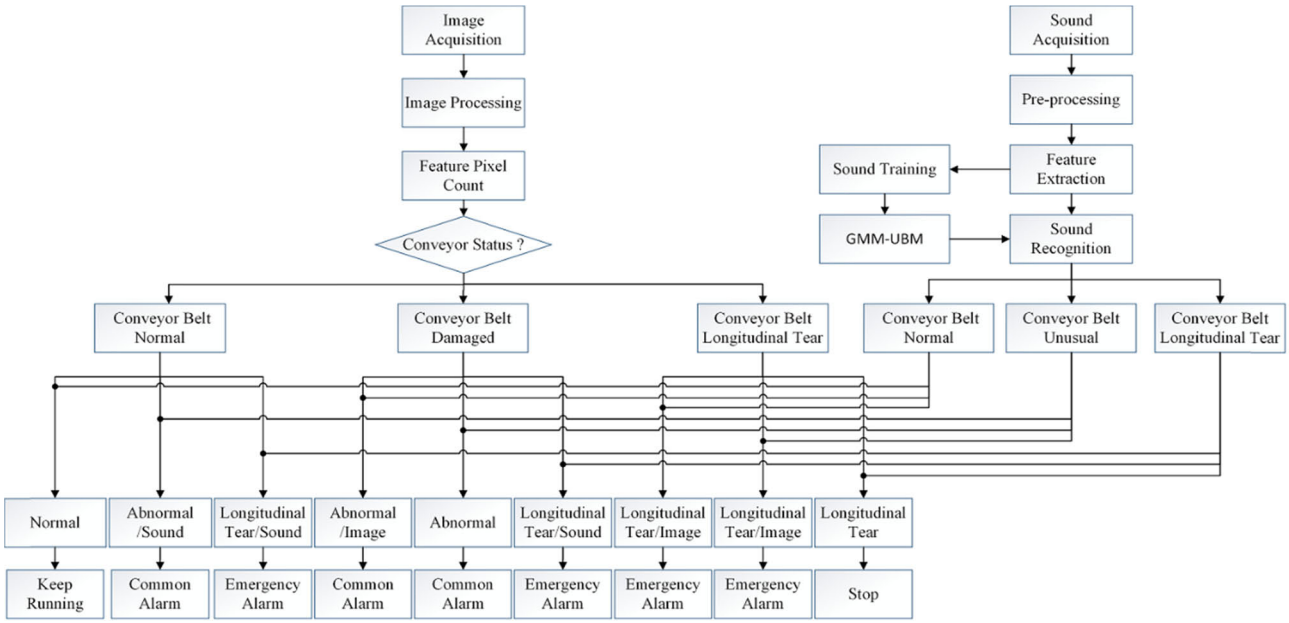

FIGURE 10. The flowchart of audio-visual detection method for conveyor belt longitudinal tear.

of 650mm∗487mm. In the image analysis, when the threshold t_1 is set to 350 and the threshold t_2 is set to 2450, the conveyor belt normal, damaged, and longitudinal tear can be effectively detected. The minimum width of the conveyor belt longitudinal tear that can be identified using the image detection method described herein is 3.318 mm.

At the same speed, the load running sound of the conveyor belt is different from the no-load running sound. However, we cannot simulate the running status of the conveyor belt under heavy load in the laboratory. Therefore, the sounds of the conveyor belt collected in the experiment mainly includes: no-load running sound, disturbing no-load running sound, no-load running longitudinal tear sound, disturbing no-load running longitudinal tear sound, and disturbing no-load running damaged sound. Conveyor belt running sound is different at various speeds. Therefore, it is necessary to collect the running sound of the conveyor belt at different speeds. Specifically, the speed of the conveyor belt is adjusted to increase by 0.5 m/s each time, i.e. 0.5m/s, 1.0m/s, 1.5m/s, ... 4.0m/s. Conveyor belt no-load running sounds at each speed were collected 100 segments. Then, the conveyor belt disturbing no-load running sounds at each speed, the conveyor belt no-load running longitudinal tear sounds at each speed, the conveyor belt disturbing no-load running longitudinal tear sounds at each speed, and conveyor belt disturbing no-load running damaged sounds at each speed were collected 100 segments, respectively. The UBM is generated by the EM algorithm using the first 70 segments of each state sound. The sound used for UBM training totals 2,800 segments, 1,400s. The adapted target type sound GMM are generated by the MAP algorithm using the remaining 30 segments of each state sound. The sound used to train and generate adapted GMM totals 1,200 segments, 600s.

After each adapted target type sound GMM has been generated by using the GMM-UBM method, the experiment begin. The specific steps of the experiment are:

- Open the audio-visual detection program for conveyor belt longitudinal tear. Start the conveyor belt at a speed

TABLE 2. The statistics of the image data and sound data collected in the experiment.

Experiment	Normal Running		Disturbing Sound Running		Longitudinal Tear		Disturbing Sound Longitudinal Tear		Disturbing Sound Damaged	
	Normal Image	Normal Sound	Normal Image	Unusual Sound	Longitudinal Tear Image	Longitudinal Tear Sound	Longitudinal Tear Image	Longitudinal Tear Sound	Damage d Image	Unusual Sound
1	465	96	663	135	528	108	558	114	437	89
2	639	132	519	105	714	144	801	162	782	158
3	546	114	594	120	555	111	507	102	604	122

TABLE 3. analysis results of the audio-visual detection method for conveyor belt longitudinal tear.

Experiment	Normal Running			Disturbing Sound Running			Longitudinal Tear			Disturbing Sound Longitudinal Tear			Disturbing Sound Damaged		
	CBN	CBA	CBLT	CBN	CBA	CBLT	CBN	CBA	CBLT	CBN	CBA	CBLT	CBN	CBA	CBLT
1	93	3	0	101	31	3	0	8	100	5	101	8	7	75	7
2	128	4	0	80	18	7	2	3	139	10	141	11	11	143	4
3	113	1	0	98	20	2	0	5	106	3	91	8	8	113	1

CBN: Conveyor belt normal; CBA: Conveyor belt abnormal; CBLT: Conveyor belt longitudinal tear.

TABLE 4. Accuracy statistics of audio-visual detection method.

Experiment	Accuracy
1	86.72%
2	90.01%
3	91.56%

FIGURE 11. Experimental conveyor belt platform.

of 1 m/s to verify the detection effect of the audio-visual detection method on the normal running of the conveyor belt;

- Use a metal rod to rub the conveyor belt frame to produce a disturbing sound to verify the detection effect of the audio-visual detection method on the disturbing sound running of the conveyor belt;
- Stop the conveyor belt. Adjust the steel chisel so that it is tightly pressed to the lower surface of the upper conveyor belt. Restart the conveyor belt to verify the detection effect of the audio-visual detection method on the conveyor belt longitudinal tear;
- Use metal rod to rub the conveyor belt frame to produce a disturbing sound to verify the detection effect of the audio-visual detection method on the disturbing sound conveyor belt longitudinal tear;
- Stop the conveyor belt. Adjust the steel chisel to away from the conveyor belt, and then restart the conveyor belt at 1m/s speed. Use the metal rod to rub the conveyor belt frame to produce disturbing sound to verify the detection effect of the audio-visual detection method on the disturbing sound conveyor belt damaged.

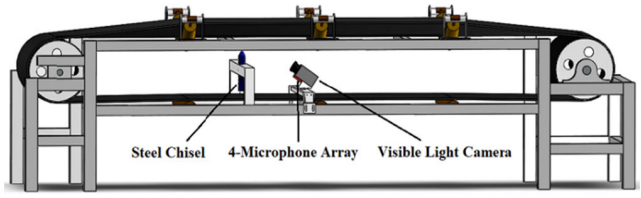

FIGURE 12. Schematic diagram of the experimental platform.

Adjust the conveyor belt speed to 2 m/s and 3 m/s, and the above experiment steps are repeated, respectively. Table 2 shows the statistics of the image data and sound data collected in the experiment.

VI. EXPERIMENTAL RESULTS AND ANALYSIS

The processing analysis time of each sound segment was 30 ms in the experiment. In the audio-visual detection method for conveyor belt longitudinal tear, the analysis result of one segment of sound and the analysis result of five frames

TABLE 5. Sensitivity statistics of audio-visual detection method.

Experiment	Normal Running	Disturbing Sound Running	Longitudinal Tear	Disturbing Sound Longitudinal Tear	Disturbing Sound Damaged
1	96.88%	74.81%	92.59%	88.60%	84.27%
2	96.97%	76.19%	96.53%	87.04%	90.51%
3	99.12%	81.67%	95.50%	89.22%	92.62%

TABLE 6. Analysis results of the image-based detection method.

Experiment	Normal Running			Disturbing Sound Running			Longitudinal Tear			Disturbing Sound Longitudinal Tear			Disturbing Sound Damaged		
	CBN	CBD	CBLT	CBN	CBD	CBLT	CBN	CBD	CBLT	CBN	CBD	CBLT	CBN	CBD	CBLT
1	375	48	42	578	51	34	45	61	422	38	52	468	23	317	97
2	536	72	31	417	57	45	32	75	607	52	53	696	31	624	127
3	453	59	34	501	62	31	44	58	453	43	61	403	15	503	86

CBN: Conveyor belt normal; CBD: Conveyor belt damaged; CBLT: Conveyor belt longitudinal tear.

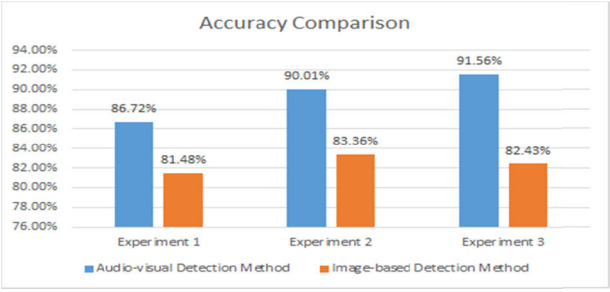

FIGURE 13. Accuracy comparison of the audio-visual detection method and image-based detection method.

of image were comprehensively analyzed to obtain the status of the conveyor belt. Table 3 shows the analysis results of the audio-visual detection method for conveyor belt longitudinal tear. Wherein the conveyor belt abnormal include conveyor belt abnormal/sound and conveyor belt abnormal /image; the conveyor belt longitudinal tear include conveyor belt longitudinal tear/sound and conveyor belt longitudinal tear/image.

We calculated the accuracy, sensitivity, and the uncertainty of the accuracy, respectively. Accuracy is the ratio of the total number of samples correctly identified by the audio-visual detection method (in Table 3) to the sum of all samples in the experiment (in Table 2). Sensitivity is the ratio of the number of correctly identified samples in one type (in Table 3) to the actual total number of samples in that type (in Table 2). The uncertainty of the accuracy rate refers to the maximum distance between the each accuracy and the average of all accuracy. Table 4 shows the accuracy of each experiment. Table 5 shows the sensitivity of each experiment. In the three experiments, the highest accuracy of audio-visual detection method was 91.56% in experiment 3. The highest sensitivity of longitudinal tear detection was 96.53% in experiment 2. The uncertainty of the experimental accuracy was 2.72% in experiment 1.

In order to compare the recognition accuracy of the audio-visual detection method for conveyor belt longitudinal tear and the image-based detection method for conveyor belt longitudinal tear. Counted the analysis results in the image analysis process. Table 6 shows the number of images in the analysis results of the image-based detection method.

Calculated the accuracy of the image-based detection method. Compared with the accuracy of the audio-visual detection method, the comparison results are shown in Fig. 13. It can be found that the accuracy of the audio-visual detection method is at least 5.24% higher than that of the image-based detection method.

VII. CONCLUSION

In order to improve the safety of coal mining conveyor belts, an audio-visual detection method for conveyor belt longitudinal tear is proposed in this paper. This method has never been proposed in this field. The visible light camera and microphone array are respectively used to collect the image signals and the sound signals of the conveyor belt. Then the dual-thread analysis to obtain the image detection results and the sound detection results about the conveyor belt, respectively. Finally, combined with the detection results of the audiovisual signals, the current status of conveyor belt is obtained. An experimental conveyor belt platform was built in the laboratory to verify the audio-visual detection method. The experimental results show that the proposed audio-visual detection method can accurately identify the normal, abnormal and longitudinal tear of the conveyor belt. The detection accuracy is over 86.72% and the uncertainty of accuracy is 2.72%. The sensitivity of longitudinal tear detection is greater than 92.59%. By comparing with the image-based detection method for longitudinal tear, the accuracy of audio-visual detection is at least 3.95% higher. The audio-visual detection method for conveyor belt longitudinal tear meets the

requirements of industrial site reliability. However, the audio-visual detection method proposed in this paper needs to be improved from many aspects. Future research will include: 1) improving the detection speed of the proposed method to ensure the real-time performance of the longitudinal tear detection; 2)studying the specific information of band, spectrum, and energy of the conveyor belt longitudinal tear sound; 3) increasing the diversity of the experimental methods to verify the effectiveness of the proposed method from various aspects, such as changing the size or position of steel chisel; 4)applying the audio-visual detection method to the coal mine site for further testing, correcting various parameters, and improving the applicability of the audio-visual detection method.

REFERENCES

[1] F. Hakami, A. Pramanik, N. Ridgway, and A. K. Basak, "Developments of rubber material wear in conveyer belt system," *Tribol. Int.*, vol. 111, pp. 148–158, Mar. 2017.
[2] Y. Yang, C. Hou, T. Qiao, H. Zhang, and L. Ma, "Longitudinal tear early-warning method for conveyor belt based on infrared vision," *Measurement*, vol. 147, Dec. 2019, Art. no. 106817.
[3] M. Andrejiova, A. Grincova, and D. Marasova, "Measurement and simulation of impact wear damage to industrial conveyor belts," *Wear*, vols. 368–369, pp. 400–407, Dec. 2016.
[4] X. Peng, "A novel image-based method for conveyor belt rip detection," in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput. (ICSPCC)*, Aug. 2013, pp. 1–4.
[5] Y. Pang and G. Lodewijks, "A novel embedded conductive detection system for intelligent conveyor belt monitoring," in *Proc. IEEE Int. Conf. Service Oper. Logistics, Inform.*, Jun. 2006, pp. 803–808.
[6] D. Dobrotă, "Vulcanization of rubber conveyor belts with metallic insertion using ultrasounds," *Procedia Eng.*, vol. 100, pp. 1160–1166, Jan. 2015.
[7] X.-L. Hao and H. Liang, "A multi-class support vector machine real-time detection system for surface damage of conveyor belts based on visual saliency," *Measurement*, vol. 146, pp. 125–132, Nov. 2019.
[8] D. Xiao, X. Li, and K. He, "Power balance of starting process for pipe belt conveyor based on master-slave control," *IEEE Access*, vol. 6, pp. 16924–16931, 2018.
[9] Y. Yang, C. Miao, X. Li, and X. Mei, "On-line conveyor belts inspection based on machine vision," *Optik*, vol. 125, no. 19, pp. 5803–5807, Oct. 2014.
[10] T. Qiao, L. Chen, Y. Pang, G. Yan, and C. Miao, "Integrative binocular vision detection method based on infrared and visible light fusion for conveyor belts longitudinal tear," *Measurement*, vol. 110, pp. 192–201, Nov. 2017.
[11] B. Yu, T. Qiao, H. Zhang, and G. Yan, "Dual band infrared detection method based on mid-infrared and long infrared vision for conveyor belts longitudinal tear," *Measurement*, vol. 120, pp. 140–149, May 2018.
[12] M. Andrejiova, A. Grincova, and D. Marasova, "Failure analysis of the rubber-textile conveyor belts using classification models," *Eng. Failure Anal.*, vol. 101, pp. 407–417, Jul. 2019.
[13] Y. Yang, Y. Zhao, C. Miao, and L. Wang, "On-line longitudinal rip detection of conveyor belts based on machine vision," in *Proc. IEEE Int. Conf. Signal Image Process. (ICSIP)*, Aug. 2016, pp. 315–318.
[14] C. Hou, T. Qiao, H. Zhang, Y. Pang, and X. Xiong, "Multispectral visual detection method for conveyor belt longitudinal tear," *Measurement*, vol. 143, pp. 246–257, Sep. 2019.
[15] E. Benetos, G. Lafay, M. Lagrange, and M. D. Plumbley, "Polyphonic sound event tracking using linear dynamical systems," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, no. 6, pp. 1266–1277, Jun. 2017.
[16] D. Y. Oh and I. D. Yun, "Residual error based anomaly detection using auto-encoder in SMD machine sound," *Sensors*, vol. 18, no. 5, p. 1308, Apr. 2018.
[17] S. Cabon, F. Porée, A. Simon, B. Met-Montot, P. Pladys, O. Rosec, N. Nardi, and G. Carrault, "Audio- and video-based estimation of the sleep stages of newborns in Neonatal Intensive Care Unit," *Biomed. Signal Process. Control*, vol. 52, pp. 362–370, Jul. 2019.
[18] V. Molnár, G. Fedorko, M. Andrejiová, A. Grinčová, and M. Tomašková, "Analysis of influence of conveyor belt overhang and cranking on pipe conveyor operational characteristics," *Measurement*, vol. 63, pp. 168–175, Mar. 2015.
[19] L. Xianguo, S. Lifang, M. Zixu, Z. Can, and J. Hangqi, "Laser-based on-line machine vision detection for longitudinal rip of conveyor belt," *Optik*, vol. 168, pp. 360–369, Sep. 2018.
[20] M. Storath and A. Weinmann, "Fast median filtering for phase or orientation data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 3, pp. 639–652, Mar. 2018.
[21] M. Huang, W. Yu, and D. Zhu, "An improved image segmentation algorithm based on the Otsu method," in *Proc. 13th ACIS Int. Conf. Softw. Eng., Artif. Intell., Netw., Parallel/Distrib. Comput.*, Aug. 2012, pp. 135–139.
[22] S. Kollem and B. Panlal, "Enhancement of images using morphological transformations," *Int. J. Comput. Sci. Inf. Technol.*, vol. 4, no. 1, pp. 33–50, Mar. 2012.
[23] J. Zhao, H. Yu, X. Gu, and S. Wang, "The edge detection of river model based on self-adaptive canny algorithm and connected domain segmentation," in *Proc. World Congr. Intell. Control Autom.*, Jul. 2010, vol. 2, no. 1, pp. 1333–1336.
[24] X. Lu, J. Wu, X. Ren, B. Zhang, and Y. Li, "The study and application of the improved region growing algorithm for liver segmentation," *Optik*, vol. 125, no. 9, pp. 2142–2147, May 2014.
[25] B. Ma, Z. Liu, F. Jiang, Y. Yan, J. Yuan, and S. Bu, "Vehicle detection in aerial images using rotation-invariant cascaded forest," *IEEE Access*, vol. 7, pp. 59613–59623, 2019.
[26] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digit. Signal Process.*, vol. 10, nos. 1–3, pp. 19–41, 2000.
[27] X. Zhou, H. Hu, L. Wei, J. Wang, W. Dai, and H. Mao, "Recognition of infant's emotions and needs from speech signals," in *Proc. IEEE Int. Conf. Syst. Man, Cybern. (SMC)*, Oct. 2016, pp. 4620–4625.
[28] J. He, J. C. Yang, J. Xiong, G. Sun, and M. Xiao, "An abnormal speech detection algorithm based on GMM-UBM," *J. Multimed.*, vol. 9, no. 5, pp. 660–667, May 2014.
[29] M. Jalil, F. A. Butt, and A. Malik, "Short-time energy, magnitude, zero crossing rate and autocorrelation measurement for discriminating voiced and unvoiced segments of speech signals," in *Proc. Int. Conf. Technol. Adv. Electr. Electron. Comput. Eng. (TAEECE)*, May 2013, pp. 208–212.
[30] K. Mannepalli, P. N. Sastry, and M. Suman, "MFCC-GMM based accent recognition system for Telugu speech signals," *Int. J. Speech Technol.*, vol. 19, no. 1, pp. 87–93, Mar. 2016.
[31] S. Vázquez, M. Amor, and B. B. Fraguela, "Portable and efficient FFT and DCT algorithms with the heterogeneous butterfly processing library," *J. Parallel Distrib. Comput.*, vol. 125, pp. 135–146, Mar. 2019.
[32] S.-H. Jung and Y.-J. Chung, "Screaming sound detection based on UBM-GMM," *Int. J. Grid Distrib. Comput.*, vol. 10, no. 11, pp. 11–22, Nov. 2017.
[33] H. Khlaifi, D. Istrate, J. Demongeot, and D. Malouche, "Swallowing sound recognition at home using GMM," *IRBM*, vol. 39, no. 6, pp. 407–412, Dec. 2018.
[34] Y. Koizumi, S. Saito, H. Uematsu, and N. Harada, "Optimizing acoustic feature extractor for anomalous sound detection based on Neyman-Pearson lemma," in *Proc. 25th Eur. Signal Process. Conf. (EUSIPCO)*, Aug./Sep. 2017, pp. 698–702.
[35] C. L. Byrne, "The EM algorithm: Theory, applications and related methods," Dept. Math. Sci., Univ. Massachusetts Lowell, Lowell, MA, USA, Tech. Rep., Mar. 2017.
[36] Y. Tsao and Y.-H. Lai, "Generalized maximum a posteriori spectral amplitude estimation for speech enhancement," *Speech Commun.*, vol. 76, pp. 112–126, Feb. 2016.

CHENGCHENG HOU received the B.S. degree in information engineering from North Minzu University, Yinchuan, China, in 2011. He is currently pursuing the M.S. degree in control engineering with the Taiyuan University of Technology, Taiyuan, China. His research interests include audiovisual signal processing and coal mine safety detection.

TIEZHU QIAO received the B.S. degree in industrial electrical automation from the Shanxi Mining Institute, in 1990, and the M.S. degree in control science and engineering and the Ph.D. degree in circuits and systems from the Taiyuan University of Technology, Taiyuan, China, in 2004 and 2015, respectively, where he is currently a Professor and the Director of the Key Laboratory of Advanced Transducers and Intelligent Control System, Ministry of Education. His research interests include machine vision, advanced transducers, and intelligent control systems.

MEIYING QIAO received the Ph.D. degree in control theory and control engineering from the China University of Mining and Technology, Xuzhou, China, in 2012. She is currently an Associate Professor and the Deputy Director of the Department of Automation, Henan Polytechnic University, Jiaozuo, China. Her research interests include intelligent control, machine learning, and ac speed control.

XIAOYAN XIONG received the Ph.D. degree in mechatronic engineering from the Taiyuan University of Technology, Taiyuan, China, in 2008, where she is currently a Professor and the Vice President of the Department of Mechanical and Electronic Engineering, School of Mechanical Engineering. Her research interests include mechanical fault diagnosis, modern signal processing, and related hardware and software implementation.

YI YANG received the Ph.D. degree in instrument science and technology from Tianjin University, Tianjin, China, in 2018. He is currently with the Key Laboratory of Advanced Transducers and Intelligent Control System, Ministry of Education, Taiyuan University of Technology, Taiyuan, China. His research interests include computer vision, image processing, and pattern recognition.

HAITAO ZHANG received the B.S. degree in measuring and controlling technologies and instruments from the Jiangsu University of Science and Technology, Jiangsu, China, in 2009, and the M.S. and Ph.D. degrees in instrumentation science and technology from Tianjin University, Tianjin, China, in 2015, where he was a Postdoctoral in mechanical engineering, from 2015 to 2017. He is currently with the Key Laboratory of Advanced Transducers and Intelligent Control System, Ministry of Education, Taiyuan University of Technology, Taiyuan, China. His research interests include the development of precision measuring technology and instruments, precision measurement of geometrical parameters, error calibration, and compensation of precision instrumentations.