J. Vis. Commun. Image R. 68 (2020) 102767

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

A real time expert system for anomaly detection of aerators based on computer vision and surveillance cameras ☆

Yeqi Liu [a,c,d], Huihui Yu [b], Chuanyang Gong [a,c,d], Yingyi Chen [a,c,d,e,*]

[a] College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, PR China
[b] School of Information Science and Technology, Beijing Forestry University, Beijing 100091, PR China
[c] Key Laboratory of Agricultural Information Acquisition Technology, Ministry of Agriculture, Beijing 100083, PR China
[d] Beijing Engineering and Technology Research Center for Internet of Things in Agriculture, Beijing 100083, PR China
[e] National Innovation Center for Digital Fishery, China Agricultural University, Beijing 100083, PR China

ARTICLE INFO

Article history:
Received 20 March 2019
Revised 3 December 2019
Accepted 7 February 2020
Available online 8 February 2020

Keywords:
Computer vision
Surveillance camera
Anomaly detection
Optical flow
Object region detection
Application

ABSTRACT

Aerators are essential and crucial auxiliary devices in intensive culture, especially in industrial culture in China. In this paper, we propose a real-time expert system for anomaly detection of aerators based on computer vision technology and existing surveillance cameras. The expert system includes two modules, i.e., object region detection and working state detection. First, we present a small object region detection method based on the region proposal idea. Moreover, we propose a novel algorithm called reference frame Kanade-Lucas-Tomasi (RF-KLT) algorithm for motion feature extraction in fixed regions. Then, we describe a dimension reduction method of time series for establishing a feature dataset with obvious boundaries between classes. Finally, we use machine learning algorithms to build the feature classifier. The proposed expert system can realize real-time, robust and cost-free anomaly detection of aerators in both the actual video dataset and the augmented video dataset. Demo is available at https://youtu.be/xThHRwu_cnI.

1. Introduction

With the development of informatization and intelligentization, aquaculture is also increasingly tending to intensive culture and even industrial culture. Automated monitoring and control is critical to the development of this field. Dissolved oxygen is one of the key parameters of water quality in aquaculture which must be detected and regulated, because maintaining the stability of dissolved oxygen and related parameters is of great significance for ensuring the safe and rapid growth of aquatic animals and plants [1–3]. The impeller aerator is widely used in China to ensure the stability of dissolved oxygen and related parameters due to their low-cost and high-efficiency features. Hence, real-time anomaly detection of aerators is crucial to avoid economic losses and ecological problems caused by rapid death of fish in large areas due to lack of oxygen.

Traditionally, the detection of anomaly condition of aerators mainly depends on manual inspection, and some researchers also try to use circuit detection method [4]. However, the former method is labor-intensive and inefficient, and the latter often suffers from error detection or short circuit, which also cannot detect equipment problems caused by mechanical failure, e.g., the machine is stuck due to lubrication or foreign matter. Besides, the circuit maintenance requires professionals to operate, which is dangerous and difficult for ordinary workers. These methods cannot find anomaly condition timely and accurately.

Recently, computer vision technology has been one of the most rapidly developing and mature subfields in artificial intelligence. Due to the characteristics of no human contact, high efficiency and high precision, this technology has been widely used in various fields and realized industrialization, such as face recognition [5], automatic driving [6], plant identification [7], etc. Furthermore, from the investigation in aquaculture, surveillance cameras are widely used as the visual perception module of the Internet of Things in a variety of applications, but they are only used to detect suspicious behavior of human being in aquaculture of China. The video-based method further extends the application scope of computer vision technology, which can not only use a single image pro-

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author at: College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, PR China.
E-mail addresses: liuyeqi@cau.edu.cn (Y. Liu), chenyingyi@cau.edu.cn (Y. Chen).

https://doi.org/10.1016/j.jvcir.2020.102767
1047-3203/

cessing technology, but also provide more information in the temporal dimension of the inter-frame connection. Computer vision technology based on video images is used to detect the change of motion state in traffic monitoring and security fields [8,9].

Therefore, the use of existing surveillance cameras and computer vision technology can achieve real-time anomaly detection of aerator without any extra cost, which also can reduce labor intensity. Meanwhile, this methodology can realize real-time warning and recording of anomaly condition, providing useful reference and attempt for intelligent agriculture and agricultural expert systems. However, the methodology for robust object region detection and real-time working state detection of aerators using computer vision technology and existing surveillance cameras faces three major challenges, which have restricted the development of the methodology and application. The proposed system solves these practical problems:

(1) In the actual surveillance images, the height of the existing surveillance camera and its position relative to the target object are different, which results in a large difference in the size and shape of the object region, e.g., small object, long strip region, etc. To demonstrate the robustness of our method, the dataset is constructed from surveillance cameras of different positions and heights, which we call internal factors.
(2) There are many interference factors caused by different weather, different time, and different scenes, such as occlusion, brightness changes, camera jitter, background interference including ripple, suspended matter, pedestrian interference, etc. The constructed dataset also contains various common interference conditions, which we call external factors.
(3) For an application-oriented expert system, some system performance must be satisfied, e.g., stable, robust, fast, low-cost, easy to operate, etc. The proposed system is a real-time, accurate and automated expert system with existing surveillance cameras, which means no additional cost.

In short, this paper proposes the RF-KLT algorithm for robust motion feature extraction and presents a dimension reduction method for fast feature dataset establishment. The aim of this research is to provide a real-time and automatic expert system for anomaly detection of aerators with existing surveillance cameras. The study is conducted in two modules: (1) object region detection. For this purpose, the procedure of maximum contour region detection, candidate region detection, and object region detection is designed; (2) working state detection. The image feature of each frame is extracted by RF-KLT algorithm, and then the time series is converted into a two-dimensional feature dataset. Finally, the feature classification model established by SVM algorithm is used to detect the working state. The comparison between multiple foreground detection algorithms and machine learning algorithms shows that the proposed methods perform well in both modules.

1.1. Contributions

In this paper, the research object is impeller aerator, which is the most widely used type of aerator. The research purpose is to detect the aerator anomaly automatically and in real time, and the aerator anomaly is defined as the closed state under non-human operation. The expected research result is to realize real-time and accurate detection of the object region and the two movement states of aerators (open or closed). Main contributions of our study are four-fold:

(1) RF-KLT algorithm for motion feature extraction. To the best of our knowledge, we first propose a novel algorithm called RF-KLT algorithm combining with the reference idea and KLT algorithm [10]. Specifically, we assume that all motions are changes relative to a fixed reference frame, extending the conditions under which the KLT algorithm requires continuous inter-frame motion. We apply the RF-KLT algorithm to the object region detection and motion feature extraction of aerators, which proves that this algorithm can judge whether the current frame has a specific motion pattern relative to the reference frame.
(2) Object region detection method. Based on the latest region proposal idea [11] and the proposed RF-KLT algorithm, we present a small object region detection method consisting of three steps, i.e., maximum contour region detection, candidate region detection, and object region detection. In our application scenario, the problem of small object region caused by long distance and high position of surveillance cameras is solved.
(3) Time series dimension reduction method for feature dataset construction. We introduce a time series dimension reduction method for constructing feature dataset, i.e., we construct a two-dimensional dataset by using the average value of the window time series as another dimension based on the numerical distribution features of time series, thereby training a two-class classifier for detecting the working state of aerators.
(4) Expert system in practical applications. We present an expert system for anomaly detection of aerators, which consists of object region detection and working state detection. The object region detection method refers to the procedure of selecting target region from candidate regions, and the working state detection method includes motion feature extraction, feature dataset construction, and classifier training. In the actual application scenarios, as well as the augmented videos with artificial noise and interference in multiple ways, the proposed expert system can realize real-time, stable and accurate anomaly detection of aerators.

1.2. Paper organization

The rest of this paper is structured as follows. Section 2 presents an overall review of traditional and state-of-the-art related works. In Section 3, we first introduce the raw video data and the artificially augmented video data. Then, we organize the introduction of all methods according to the technical flow of object region detection and work state detection. Notably, the RF-KLT algorithm and the dataset construction method of time series are both presented in Section 3.2.3. In Section 4, we show the comparison, discussion and experimental results of the proposed expert system. Finally, Section 5 shows the conclusions and the future work of this study.

2. Related work

Because of the three major challenges we have summarized in the introduction section, there are few existing studies based on computer vision and surveillance cameras for aerator-related detection. He et al. [12] has made some attempts in this application scenario. They used Harris corner features based on the computer vision technology for working state detection between adjacent frames with the detection time of 4 FPS, which shows the real-time performance is not well. Moreover, they presented artificially selected object region, which inevitably affects the extraction of feature points. Additionally, the inter-class distance of the dataset

established using the statistical values of the motion feature distribution between adjacent frames is not obvious. By contrast, we extract the motion features based on RF-KLT algorithm to select the object region in the complex background, and propose a feature dataset construction method with obvious inter-class distance.

2.1. Expert systems based on computer vision technology and surveillance cameras

Expert systems based on computer vision technology and surveillance cameras are widely used in various fields, e.g., traffic detection [13], pedestrian identification [14], fire detection [15], etc. Expert systems that use surveillance cameras to detect suspicious behavior have also received attention from researchers [16,17]. In addition, expert systems based on computer vision technology and video information are used in agriculture for counting [18], tracking [19], and behavioral analysis of animals [20]. However, most of these methods, especially in agriculture, are based on a particular environment or require the relative position of the monitoring device to the monitored object, which limits the application value of this technology. Obviously, the real-time and on-the-ground monitoring combined with existing surveillance cameras and sophisticated computer vision technology is a low-cost, easy-to-promote, and real-time method in many application scenarios.

2.2. Object detection in video

Object detection methods in video are based on inter-frame features. The video adds time dimension to a single image, which helps to detect the foreground region [21]. Andrews et al. [22] compares 29 video-based background subtraction algorithms implemented in BGSLibrary for object detection. Nevertheless, the classical background modeling algorithms can detect a small foreground region [23], but they cannot feedback other features of the foreground region, so they are still difficult to further detect the object region from a large number of candidate object regions. Recently, the object detection in videos of ImageNet introduce a new challenge for this domain, i.e. VID, and all state-of-the-art methods are based on deep neural networks [24]. However, deep neural network-based methods require a large amount of training data and high-performance computing power, which is difficult to obtain in our scenario. The generalization capabilities of these models in our different application scenarios are also uncertain. Based on the idea of first generating candidate regions and then selecting the target regions, i.e., region proposal idea [11], we propose a selection procedure of the object regions through different motion features of different motion states in our scenario.

2.3. Optical flow

The latest advances in optical flow are combined with deep neural networks for fast and end-to-end object detection and object tracking [25,26], because optical flow can reflect temporal information of motion features between video frames. However, all of these optical flow methods are based on three major assumptions [27], which limits its application scenario. The pyramid method overcomes this obstacle in some extent [28], but it is still impossible to distinguish the temporal and spatial state of the movement from beginning to end. In contrast, we propose RF-KLT algorithm based on the reference idea, i.e., we do not assume the change between consecutive frames, but measure the change of optical flow between the current frame and a fixed reference frame. Experimental results show that the proposed RF-KLT algorithm can extract robust motion features for motion state detection in a fixed region.

2.4. Time series analysis

The purpose of time series analysis is generally time series prediction, which can be divided into two categories: univariate time prediction [29,30] and multivariate time prediction [31,32]. Currently, machine learning algorithms, including deep neural networks [33,34], are widely used in time series prediction. Most of these methods need to be combined with external features to achieve more accurate predictions. In fact, univariate time analysis is more difficult because it does not extract relevant information from other external features. Also, we propose a method to construct a feature dataset which is only based on the numerical distribution characteristics of the single series. We directly convert the numerical distribution of the time series extracted from the video into a two-dimensional training dataset. The experimental results show that this method has achieved good results in the feature classification problem of time series.

3. Material and method

For each specific surveillance camera with different heights and distances in a practical scenario, the corresponding models are required to be established. Therefore, the purpose of object region detection is to detect the aerator region, and the purpose of working state detection is to train the classifier to judge the working state for anomaly detection. Notably, the module of work state detection extracts features from the object region in the reference frame. The relationship and purpose of each module are shown in Fig. 1, and the details of each module are given in Sections 3.2 and 3.3.

3.1. Video dataset

To build a universal dataset, we collected video data under different conditions, and each video includes 1–3 time periods of anomaly state. Considering the uncertainty of the installation position and height of surveillance cameras and the complexity of the interference, we notice two aspects: the first is internal factors, e.g., the location, height and type of surveillance cameras; the second is external factors, e.g., weather, time and possible interference. For instance, the presented videos were selected under complex scenes such as different surveillance cameras with different heights, locations and types, complicate weathers, different time periods, pedestrian interference, surveillance camera jitter, etc.

Specifically, the video dataset mainly comes from aerator videos which are within 5 min and extracted in real time on the Internet of Things (IoT) monitoring platform in Xiongan New Area, China. Eight typical complex scenarios selected from the video dataset are presented to demonstrate the effectiveness of the proposed methods, and these videos are labelled as shown in Table 1. The

Fig. 1. The relationship and purpose of each module.

Table 1
Dataset annotation.

Video number	1	2	3	4	5	6	7	8
Number of frames in the open state	1847	408	2067	177	458	1361	486	2498
Number of frames in the closed state	1944	57	1982	252	1220	1397	640	1021
Total number of frames	3791	465	4049	429	1660	2758	1126	3519

video data is from 3 different types of surveillance cameras, and a moving camera is also used to shoot some videos to simulate some specific scenes, i.e., videos 6 and 7. The description of video scenes under different conditions is shown in Table 2. In Table 2, the height of surveillance cameras relative to the object region is divided into three levels: high (H, 8 m), medium (M, 5 m) and low (L, 2–3 m); and the relative distance is also divided into three levels: far (F, >300 m), middle (M, 15–70 m), and near (N, <15 m). It also shows information including time, weather, video parameters, and background description. Notably, the video dataset used to detect the object region is a video clip cut in the corresponding video. In each clip, the working state of the aerator just changes once and the duration is between 10 and 20 s.

To further prove the robustness of the proposed method, video datasets are also augmented artificially through adding noise and varying brightness in various proportions to simulate some complex working environment, e.g., the illumination changes gradually or suddenly (e.g., Fig. 2). The gray change formula is shown in Formula (3–1).

$$p'(x,y)=\begin{cases} p(x,y)+v & (0\le p(x,y)+v\le 255)\\ 0 & (p(x,y)+v<0)\\ 255 & (p(x,y)+v>255)\end{cases} \qquad (3-1)$$

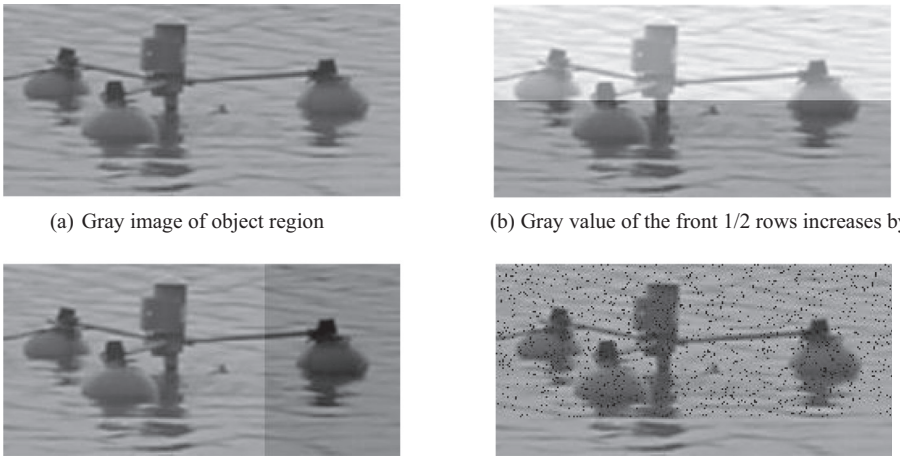
where p(x, y) is the original gray value at the image pixel (x, y), v is the amount of gray change, and $p'(x,y)$ is the changed gray value at the image pixel (x, y). In this study, the augmented video dataset was determined by changing the ratio of gray and noise within twice the area of object region. The random number is used in the ratio change. In Table 3, all augmented datasets are based on video 1, and it also shows the random ratio, step size, signal-noise ratio (SNR), and whether the reference frame makes the same change. For example, for video P3, the gray of the reference frame and current frame in the video is increased by 40 in all rows and all columns of the image pixel, and salt-and-pepper noise with the SNR in the range of [0.01, 0.1] is randomly added with 0.01 step size changing in all rows and columns of the image pixel.

3.2. Object region detection

Since the object region determined artificially will cause the detection of extra feature points and then affect the speed and precision of working state detection, this paper proposes an automatic detection method for aerator region. The method extracts integrated features from the motion feature and gray feature in inter-frames, and the contour feature in key frame. This method detects the object region in three steps, which is summarized as

Table 2
Raw dataset description.

Video number	Height	Distance	Frame rate (FPS)	Resolution	Weather	Time	Background description
1	M	M	25	720 * 1280	Overcast	a.m.	Ripple, drainage.
2	M	M	25	720 * 1280	Sun	p.m.	Direct sunlight, surveillance camera automatically set black and white mode.
3	M	M	25	720 * 1280	Overcast	a.m.	Human walking, ripple, drainage.
4	H	M	12	352 * 640	Overcast	p.m.	Occlusion.
5	H	M	12	352 * 640	Overcast	a.m.	Occlusion, human walking.
6	L	N	25	1080 * 1920	Sun	p.m.	Using a camera to simulate a close range scene.
7	L	N	25	1080 * 1920	Overcast	a.m.	Using camera to simulate; the strong wind causes camera shake, branch shaking and ripple.
8	H	F	25	352 * 640	Overcast	p.m.	Rocking branch, ripple.



(a) Gray image of object region (b) Gray value of the front 1/2 rows increases by 80

(c)Gray value of the post 1/3 columns reduces by 30 (d)Random salt and pepper noise in the front 4/5 rows(SNR=0.04)

Fig. 2. Example of artificially augmented dataset (the aerator is shown in the figure).

Table 3
Augmented dataset description.

Augmented video number	Raw video number	Change ratio of rows (step size is 0.1)	Change ratio of columns(step size is 0.1)	Change range of gray(step size is 1)	SNR of salt and pepper noise(step size is 0.01)	Whether reference frame makes the corresponding change
P1	1	[0:1]	[0:1]	[−80,80]	[0.01,0.1]	Yes
P2	1	[0:1]	[0:1]	+40	[0.01,0.1]	Yes
P3	1	All	All	+40	[0.01,0.1]	Yes
P4	1	All	All	+80	[0.01,0.1]	Yes

the detection of maximum contour regions, candidate regions and the object region: (1) an adaptive Gaussian mixture model is used to detect the maximum contour region of each frame; (2) the candidate region is detected from the maximum contour regions by integrating the gray feature in inter-frames and the contour feature in key frame; (3) the object region is detected through the candidate regions by calculating the maximum inter-class interval of the motion features dataset based on RF-KLT algorithm. The algorithm flow of this module is shown in Fig. 3.

3.2.1. Maximum contour detection

This research detects the set of maximum contour regions by the adaptive Gaussian mixture model [23,35]. Background is usually a fixed region with less change, and it is usually assumed that the background can be described by a statistical model. The Gaussian mixture model uses multiple weighted and mixed Gaussian models to analyze background characteristics. Next, the foreground can be detected by marking the part of the image that does not conform to this background model. The Gaussian mixture model [36] can detect the region with small change in the video image, and the adaptive Gaussian mixture model can increase the detection speed and increase the robustness under illumination and ghost regions [22]. The next two steps of the maximum contour region detection are based on the assumption that the aerator region can be detected by an adaptive Gaussian mixture model in the training video.

In this step, Gaussian smoothing is performed on each frame of the video as a preprocessing step firstly. The size of the Gaussian kernel is 5 * 5. The main purpose of Gaussian smoothing is to avoid detecting some inconspicuous corners. Secondly, an adaptive Gaussian mixture model is used for foreground detection of each frame. It models each pixel with an adaptive number of Gaussian distributions, using the color values of these pixels in the length of the time in the video as a mixed weight, because the color of the background generally lasts the longest and is more static. Thirdly, threshold each frame to obtain a binary image, and this

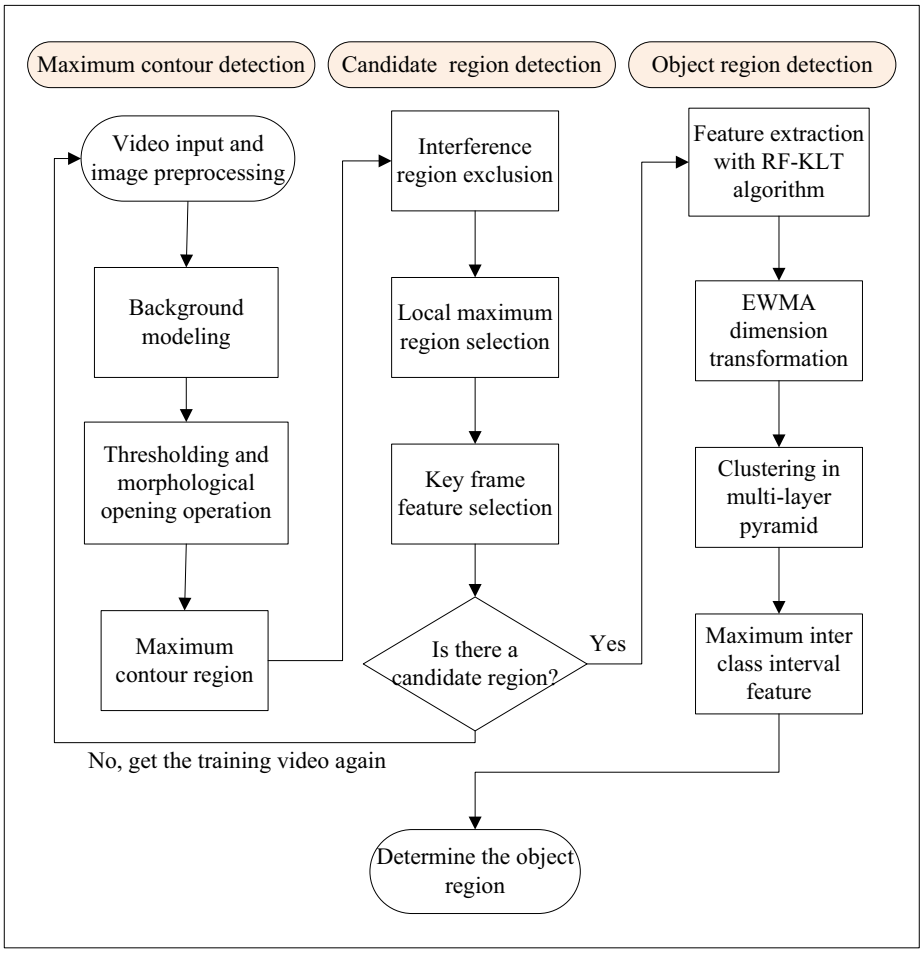


Fig. 3. The algorithm flow of object region detection module.

research selects a fixed threshold of 240 based on experience and experiments, which can avoid interference in shadowed regions. Fourthly, each frame is subjected to morphological operations to eliminate small regions and separate more obvious foregrounds. The advantage is to smooth the boundary of the larger object region without significantly changing the boundary of these regions. Finally, contour detections are performed on these foregrounds to obtain a region with the largest area from each frame in the video, which constitutes the set of the maximum contour regions.

3.2.2. Candidate region detection

There are too many regions in the set of maximum contour regions, but the areas of the maximum contour regions in most video frames are very small. By excluding these interference regions, the model training speed can be accelerated and the accuracy will be improved. As shown in Fig. 4, the regularity curve between the areas of maximum contour regions and the number of frames obeys the law shown in Formula (3–2). In maximum contour regions, the first 1/4 regions with the largest area are selected to exclude interference regions with a small area. In order to avoid interference and repeated detection in adjacent regions, regions where the centroid distance of the contour is larger than a threshold are selected from the rest of maximum contour regions. The local maximum region is selected as a candidate region, and 1/10 of minimum value from video frame resolution (e.g., 720/10) is used as the threshold for excluding adjacent regions.

In short, the object region belongs to two states which include non-working state and working state. There are a lot of feature points in both states, because the body of aerators and the edge of the spray are both changed strongly on the gray value. The interference region, e.g., branch, is also a region of strongly gray change. In addition, the features of single frame image cannot detect the motion features of object region. Hence, the frame that appears in the maximum contour region is selected as a key frame for features extracted in this region. Also, when the aerator is closed, the first frame of the video is defined as the reference frame. Compared with maximum contour region in the key frame and corresponding region in the reference frame, features of the gray change in the spray region are extracted. To avoid the influence of illumination and improve the detection accuracy of small region, the feature of contour area is considered comprehensively. The feature function of the key frame and the reference frame is determined finally in Formula (3–3).

$$y = \frac{k}{x}(k > 0) \tag{3–2}$$

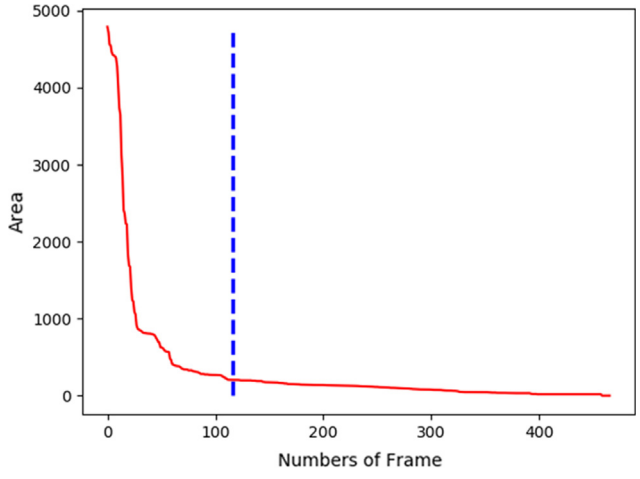


Fig. 4. Example of area distribution of maximum contour regions.

$$F = \frac{\sum_0^x \sum_0^y (I(x,y) - J(x,y))}{A_{max}} \tag{3–3}$$

where F is the feature value extracted in the selected region, $I(x,y)$ is the gray value of image point (x,y) in the selected region of the key frame, $J(x,y)$ is the gray value of image point (x,y) in the selected region of the reference frame, and A_{max} is the maximum contour area in the selected region of the key frame. F feature can be used to detect the spray region caused by the work of aerators, and it is also robust to the monitoring of small object region. If the candidate region is smaller, the area of detected contour is smaller and the corresponding F increases. This paper finally selects the set of regions where the F feature value is larger than the average value in the maximum contour regions as candidate regions.

3.2.3. Object region detection

The RF-KLT-based feature extraction method and the maximum inter-class interval measurement method are both used to determine the final object region from candidate regions. In KLT tracker [10] and its pyramidal implementation [28], considering an image point $\mathbf{u} = [x\ y]^T$ on the frame I, the objective of KLT tracker is to find the location $\mathbf{v} = \mathbf{u} + \mathbf{d} = [x + d_x\ y + d_y]^T$ on the next frame J such as the gray value $I(\mathbf{u})$ and $J(\mathbf{u})$ are "similar". The vector $\mathbf{d} = [d_x\ d_y]^T$ is the optical flow at point $\mathbf{u}$. It is essential to define the notion of similarity in a two-dimensional neighborhood sense because of the aperture problem. The optical flow $\mathbf{d}$ is defined as the vector that minimizes the residual function ϵ that defined in Formula (3–4), where the similarity function is measured on an image neighborhood of size $(2w_x + 1) \times (2w_y + 1)$. The value of w_x and w_y are integers, for which the typical values are 2, 3, 4, 5, 6, 7 pixels.

$$\epsilon(\mathbf{d}) = \epsilon(d_x, d_y) = \sum_{x=u_x-w_x}^{u_x+w_x} \sum_{y=u_y-w_y}^{u_y+w_y} (I(x,y) - J(x+d_x, y+d_y))^2 \tag{3–4}$$

The objective of pyramidal implementation of KLT tracker is also to find the location $\mathbf{v} = \mathbf{u} + \mathbf{d} = [x + d_x\ y + d_y]^T$ on the next frame J such as the gray value $I(\mathbf{u})$ and $J(\mathbf{u})$ are "similar", but the point $\mathbf{v}$ is not in the image neighborhood of size $(2w_x + 1) \times (2w_y + 1)$, and is far away from the point $\mathbf{u}$ in a larger range. It is preferable to have $d_x \le w_x$ and $d_y \le w_y$ in Formula (3–4) to find the accurate and robust point $\mathbf{v}$, so the pyramid is used instead of expanding the size of image neighborhood. The optical flow $\mathbf{d}^L = [d_x^L\ d_y^L]^T$ at level L in pyramid is defined as the vector that minimizes the new residual function ϵ^L that defined in Formula (3–5).

$$\begin{aligned}\epsilon^L(\mathbf{d}^L) &= \epsilon^L(d_x^L, d_y^L) \\ &= \sum_{x=u_x^L-w_x}^{u_x^L+w_x} \sum_{y=u_y^L-w_y}^{u_y^L+w_y} (I^L(x,y) - J^L(x + g_x^L + d_x^L, y + g_y^L + d_y^L))^2\end{aligned} \tag{3–5}$$

where $\mathbf{g}^L = [g_x^L\ g_y^L]^T$ is available from the computations done from level L to level L + 1 in pyramid, which is computed by Formula (3–6).

$$\mathbf{g}^{L-1} = 2(\mathbf{g}^L + \mathbf{d}^L) \tag{3–6}$$

For level L in pyramid, $\mathbf{d}^L$ is computed through the same procedure by Formula (3–5), which searches the finest point $\mathbf{v}$ to minimizes the functional $\epsilon^L(\mathbf{d}^L)$ in the neighborhood with constant

size $(2w_x+1)\times(2w_y+1)$. This procedure goes on until the bottom level is reached ($L=0$), and the initial value of the top level ($L=L_{max}$) in pyramid is initialized to zero.

$$\boldsymbol{g}^{L_{max}} = [0 \quad 0]^T \tag{3–7}$$

Therefore, the final optical flow $\mathbf{d}$ at point $\mathbf{u}$ is then defined in Formula (3–8).

$$\mathbf{d} = \boldsymbol{g}^0 + \boldsymbol{d}^0 = \sum_{L=0}^{L_{max}} 2^L \boldsymbol{d}^L \tag{3–8}$$

The objective of pyramidal implementation of RF-KLT tracker is the same as the above two methods, but it is changed to find more obvious motion feature, and guarantees the accuracy and robustness of feature point tracking. For each candidate region, RF-KLT algorithm proposed in this paper is divided into three steps, and the schematic diagram is shown in Fig. 5: (1) Shi-Tomas corner detection method is used to find the corners (point $\mathbf{u}$) in the fixed reference frame. The Shi-Tomas algorithm [37] determines the strong corner point through finding the maximum value from the minimum eigenvalue of each two eigenvalues in the pixel gradient matrix. It can ensure that the number of detected feature points is within a reasonable range, and the feature points in the object region all fall on the body of aerators. (2) The Lukas-Kanade algorithm is used to find the matching corners (point $\mathbf{v}$) corresponding to the reference frame in current frames. The search range of the matching corner points starts from the center of object region, and the radius of the diagonal of object region is selected as the radius. Lukas-Kanade algorithm instructs how this method is used for feature point tracking and has evolved into a practical KLT tracker [10]. (3) The average distance between matching corners of each frame is extracted as the final feature. Manhattan distance is used to measure this distance between matching corner points. The extraction of Dist features in each frame based on RF-KLT algorithm is shown in Formula (3–11).

$$\begin{aligned}\epsilon^L(\mathbf{d}^L) &= \epsilon^L\left(d_x^L, d_y^L\right) \\ &= \sum_{x=u_x^L-w_x}^{u_x^L+w_x}\sum_{y=u_y^L-w_y}^{u_y^L+w_y}\left(I_0^L(x,y) - J^L(x+g_x^L+d_x^L, y+g_y^L+d_y^L)\right)^2\end{aligned} \tag{3–9}$$

$$\mathbf{d} = \boldsymbol{g}^0 + \boldsymbol{d}^0 = \sum_{L=0}^{L_{max}} 2^L \boldsymbol{d}^L \quad \left(|\boldsymbol{d}| \le 2\sqrt{(l/2)^2+(h/2)^2}\right) \tag{3–10}$$

where $I_0^L(x,y)$ is a point $\mathbf{u}=(x,y)$ in a fixed reference frame in the level L of pyramid, and l and h are the length and width of the candidate region respectively. Compared with the conventional KLT algorithm, the RF-KLT algorithm is improved in three aspects. The calculation procedure is shown in Formula (3–9) and Formula (3–10), and schematic diagram and result formula are shown in Fig. 5 and Formula (3–11) respectively. The main procedure includes: (1) Finding the best matching point of the corresponding corner point in the fixed reference frame, which means that I_0^L is not changed along with J^L changed. The goal of the improvement is to obtain the movement state change characteristics with more distinct discrimination, and this is the most critical improvement. (2) Finding the best search radius under the multi-level image pyramid. It means that the range ($\mathbf{d}$) of searching matching point by optical flow method is limited. The goal of this improvement is to ensure that the matching corner is not lost and robust. (3) Using Manhattan distance to measure the distance between the corner points (point $\mathbf{u}$ and point $\mathbf{v}$), which is also to obtain more obvious movement characteristics. Notably, the precondition for matching corners using optical flow method based on the fixed reference frame is as follows: a multi-level pyramid and a fixed region. These two conditions ensure that the three basic assumptions of the optical flow method are still confirmed and the first and second assumptions are expanded.

$$\text{Dist} = \begin{cases} \left(\frac{1}{i}\right) * \sum_0^i |x_{i1} - x_{i0}| + |y_{i1} - y_{i0}| & (0 < i \le 5) \\ 0 & (i = 0) \end{cases} \tag{3 – 11}$$

where Dist is the distance feature value between matching corner points, (x_{i0}, y_{i0}) is the coordinates of the i-th corner point in the reference frame, and (x_{i1}, y_{i1}) is the coordinates of the i-th matching corner point in the current frame corresponds to the reference frame, and i is the number of corners. In order to ensure the real-time detection, and according to the number of strong corners on the body of aerators, the most obvious 5 corners with maximum eigenvalue are selected. Moreover, w_x and w_y are both set to 7 pixels in this study.

Dist feature is an unlabeled dataset that changes with the number of frames, and it can be considered as time series data. To smooth the data and perform dimensional transformation, the exponentially weighted moving average (EWMA) model [38] is used to process original dataset, and its calculation formula is shown in Formula (3–12). EWMA model maintains coherence of the inter-frame feature value in a fixed length window. Furthermore, the most important is that the two-dimensional feature reconstructed by deleting the time dimension is more robust. In the constructed dataset, one class of data is close to the coordinate origin, while the other class of data is far away from the coordinate origin. The two classes of datasets constructed in this way also have good flatness in shape.

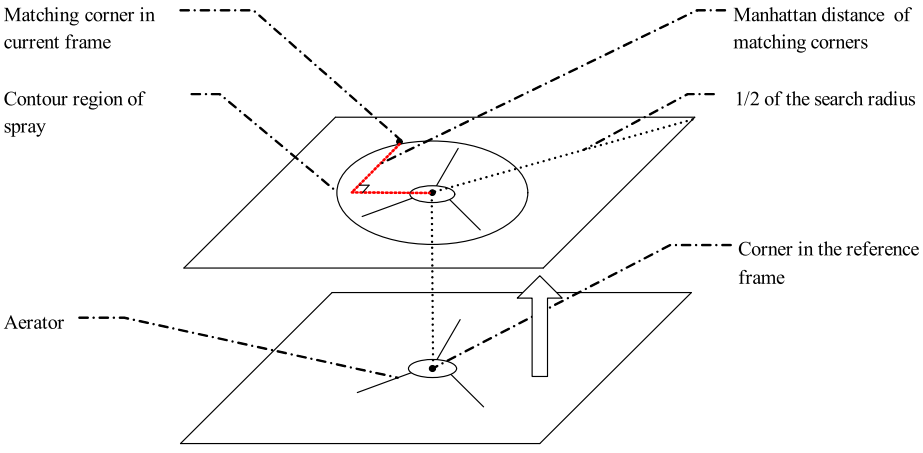


Fig. 5. Dist feature extraction based on RF-KLT algorithm.

$$y_t = \frac{x_t + (1-\alpha)x_{t-1} + (1-\alpha)^2 x_{t-2} + \cdots + (1-\alpha)^t x_0}{1 + (1-\alpha) + (1-\alpha)^2 + \cdots + (1-\alpha)^t} \quad (3-12)$$

where y_t is the EWMA feature corresponding to the Dist feature in the current frame; x_t is the Dist feature in the current frame, which means the last frame of the sliding window. x_{t-1} is the Dist feature in the previous frame of the sliding window; etc. x_0 is the Dist feature in the first frame of the sliding window; $0 < \alpha < 1$, and $\alpha = \frac{2}{s+1}$, where $s \geq 1$, and s is the frame rate of video. t is the length of the sliding window, and its selection is also automatically obtained based on the frame rate of video.

After data smoothing and dimensional transformation, the dataset has a distinct distribution rule of two classifications. Because the dataset is flat and the numbers of class are known, the K-means clustering method (K = 2) can accurately obtain the centroid of each class of dataset and thus calculate the inter-class interval feature [39]. The maximum inter-class interval feature under the multi-level pyramid can prevent the size of object region from affecting experimental results. As shown in the Formula (3–13), the region with the largest centFeature is selected as the final object region from candidate regions. The centFeature under the high level pyramid can avoid a small feature value between two classes in the small region. At the same time, by reducing the proportion of distance features under the high level pyramid, large feature value between two classes in the large region can also be avoided.

$$\text{CentFeature} = \sum_0^i \frac{Diff}{2^i} \quad (0 \leq i \leq 4) \quad (3-13)$$

where the CentFeature means inter-class interval feature of the selected candidate region. Diff is the difference of abscissa values between two class centroids under the i-th pyramid of the selected region, and i is the number of pyramid levels, increasing from 0 to 4 levels.

3.3. Working state detection

The purpose of this section is to achieve automatic detection of aerator working state. The algorithm flow of working state detection is shown in Fig. 6. In this module, the main procedures include feature extraction, feature dataset construction (including EWMA data processing and dimension transformation), data labelling, and classifier training. The details of RF-KLT algorithm and EWMA model are described in Section 3.2.3. In the reference frame of RF-KLT algorithm, corner detection is performed in the target region obtained by object region detection module.

3.3.1. Feature extraction and dataset construction

The extraction of image features in computer vision is a critical step. The purpose of RF-KLT algorithm is to extract robust features for the motion state change. Note that corner matching is performed under the 4-level pyramid. The high level pyramid can increase the search range of corner matching, and get larger features of corner distance variation, which increases the discrimination of datasets with two classes. The conventional optical flow method is based on the detection of adjacent frames, but the corner change features are not obvious in the motion state detection of the fixed region. Therefore, the proposed RF-KLT algorithm can be used to extract more obvious features of motion state change in scenes of brightness change tempestuously and large scale motions. This algorithm can also be used in other conditions to detect motion state change in a fixed region.

The EWMA model is used to transform time series data into the same dimensionality to construct a dataset that is smoother and more distinguishable, which can avoid the influence of single error on the experimental results, and can also reduce the influence of the change of working state on the experimental data. The constructed dataset has flat and highly discriminating features.

3.3.2. Unlabeled data annotation

Clustering algorithm is widely used in the learning of unlabeled samples, which is used to reveal the inherent laws of data and provide a basis for further data analysis. It is also a typical unsupervised learning algorithm, which is mainly used to automatically cluster similar samples into one class. The number of classes and inter-class characteristics of the datasets constructed in this paper are known, including the on and off of two classes of feature data. One class is close to the coordinate origin, while the other class is far away from the coordinate origin. To achieve data labeling, and because of the flatness and aggregation shape of dataset distribution, K-means clustering methods are used to annotate data [39].

In this study, label 1 is assigned to the dataset that is far from the coordinate origin corresponding to the open state, and label 0 is assigned to the dataset that is close to the coordinate origin corresponding to the closed state. The feature data in the state change process and the feature data due to the corner matching errors are determined based on the distance from the centroids of the two classes. This enables automatic labeling of unlabeled datasets. To avoid classifying high-error data into one class, the feature dataset constructed by the two working states of aerators must be a balanced dataset in training videos. This study sets the ratio of two classes of dataset between 1/5–5. The object region is small, and thus the change of features is also small. Therefore, if the corner matching is not stable, the mismatched feature value will be too large. Then the error data becomes a class of data, and the trained classifier cannot correctly detect the open state. However, the balance of training dataset can avoid such errors.

3.3.3. Classifier training

Through the feature extraction from different working states, the construction of datasets, and the labeling of unlabeled data, the final dataset is a triple, which includes two-dimensional feature data and their corresponding labels. The SVM algorithm has a complete theoretical basis and good classification ability,

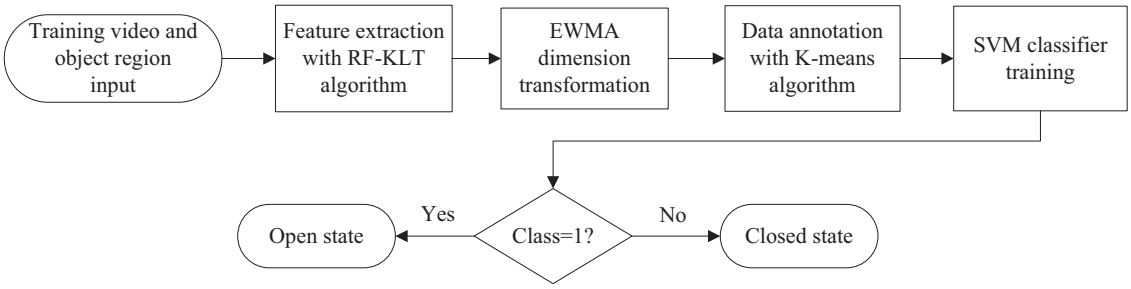


Fig. 6. The algorithm flow of working state detection module.

and is widely used in the data classification with limited samples [40]. This research constructs a dataset that is linearly separable, so the linear SVM algorithm is used to effectively classify the dataset.

At last, the application procedure of our methods is divided into three steps: (1) the corner distance feature (Dist feature) of real-time frame is extracted according to the matching corner points in the object region of the reference frame; (2) constructing two-dimensional dataset using the EWMA model; and (3) classifying feature dataset according to trained classifiers. If the class label is one, it indicates that the aerator is currently in the working state; while if the class label is zero, the aerator currently is in the closed state.

4. Result and discussion

4.1. Software and algorithm parameter

In this study, the software is programmed in Python 3.6.3 with the Numpy and Pandas scientific computing libraries. Opencv 3.3.1 in python is used for the basic image processing and video analysis [41]. Scikit-learn 0.19.1 in python is used for machine learning algorithm comparison [42]. The algorithms development and experiments are based on the 64-bit Intel 3rd-Generation Core i5 CPU (i5-3230 M, 2.60 GHz). This is a personal laptop with a common configuration, indicating that the experimental results have a good reference value in the general monitoring system.

The precision, recall and f1-score are used to evaluate the results of working state detection, and they are defined as Formula (4–1), Formula (4–2) and Formula (4–3), respectively.

$$\text{precision} = \frac{tp}{tp + fp} \tag{4—1}$$

$$\text{recall} = \frac{tp}{tp + fn} \tag{4—2}$$

$$\text{f1} = 2 * \text{presion} * \text{recall}/(\text{precision} + \text{recall}) \tag{4—3}$$

where *tp* is the number of correct samples to be classified correctly, i.e., the number of frames in working state is correctly classified as working state. *fp* is the number of incorrect samples to be classified correctly, and *fn* is the number of incorrect samples to be classified incorrectly. In the comparison experiment of the algorithm, the key parameters and reference literature in the involved algorithms are given in Table 4. Most of these algorithms are from OpenCV library and Scikit-learn library respectively. The specific meaning and details of the parameters in the algorithm refer to the official documentation, and abbreviations of some algorithm are used in this study.

4.2. Comparison and discussion of experiments

4.2.1. Comparison between RF-KLT algorithm and KLT algorithm

Compared with the classical KLT algorithm, the RF-KLT algorithm proposed in this paper has a significant discrimination degree of motion features in a fixed object region. The comparison of the differences in the motion state changes between KLT and RF-KLT is shown in Fig. 7. For the video with same work state changes, the feature change relative to the reference frame is more obvious than that of the consecutive frame. The KLT algorithm has continuous and dynamic corner motion, and thus its feature change between matching corner points is small and the distinction degree is not high. In contrast, the matching corners in the reference frame are fixed and stable in the RF-KLT algorithm, resulting in a significant but stable feature change.

Specifically, when the working state of two adjacent frames does not change, the matching corner point is itself, so the values of the generated features are equal and small (Fig. 7, line 1); and when the working state changes relative to reference frames, there is no matching corner point in the spray region. Then, the matching corners will follow the boundary between the spray and the calm water surface. Finally, when the spray region is stable in the maximum region, the matching corner position is fixed and stable, and thus the values of the generated features are relatively stable and large (Fig. 7, line 2). In short, feature changes between adjacent frames only occur when the working state changes, so they change instantaneously; while feature changes between the reference frames are relatively fixed, so they maintain significant but stable values. Therefore, the Dist feature based on the RF-KLT algorithm have a good discrimination degree on the change of motion state. At the same time, the RF-KLT algorithm selects the Shi-Tomas corner with a largest eigenvalue, and thus it has a fast speed of 0.002 FPS in video 1 compared to the Harris corner point [49].

The image pyramid can match feature points in a larger pixel range, and the RF-KLT algorithm is also built on the multi-level pyramid. The relationship between the number of pyramid levels and the distance feature of matching corner points is shown in Fig. 8. In Fig. 8, when the pyramid is higher than four levels, the distinguishability of the characteristics no longer increases. Therefore, the RF-KLT algorithm in the module of working state detection is to match the corner points under the 4-level pyramid. The detection time of each frame and the degree of instability of corner matching increase with the increase of the number of pyramid levels, so higher-level pyramid does not need to be constructed without improving feature discrimination.

Since the selection of the reference frame leads to an obvious distinction of the extracted features, the determination of the reference frame is critical. The reference frame of the RF-KLT algorithm can be determined in one frame, i.e., any frame of the non-working state of aerators can be selected. Fig. 9 shows the accuracy of working state detection with different reference frame in video 1. The experimental results show that there is no effect on the experimental results to select 1^{st}, 10^{th}, 20^{th}, 40^{th} or 80^{th} frame as the reference frame, which proves that the RF-KLT algorithm is stable and robust.

4.2.2. Comparison with different object detection algorithms

The object region detection of videos in this study is divided into three steps. Table 5 shows the influence of the different object detection method on the final detection result in different videos. In this experiment, the foreground detection algorithm is used in the maximum contour region detection module, and the subsequent two-step detection methods are not changed. This paper compares the influence of AFD, WRS-HSV, FBS, GMB, KNN-B and AGMB (details are in Table 4) on the experimental results in the maximum contour region detection, and the last method is our selected algorithm. In Table 5, the object region is correctly identified in various complex scenarios based on the KNN algorithm and our model, which shows that the principle of regional selection procedure constructed in this paper is correct. For simple background, most algorithms can accurately detect the object region. However, in complex background, e.g., the object region is too small or the interference is serious, the KNN algorithm and our algorithm have good stability. In the time performance of each frame, as shown in Fig. 10, our method is faster than KNN in all scenarios.

4.2.3. Comparison with different machine learning algorithms

After the data is labelled by the clustering algorithm, a classifier with high precision and high generalization ability can be trained for detecting the working state. The result of the combination of

Table 4
Algorithms and key parameters in the comparison of experiments.

Algorithm	Abbreviation	Key parameters in this study	Reference	Document
Support vector machine	SVM	C = 1, kernel: linear	[40]	Scikit-learn,
Linear regression	LR	C = 1, solver: liblinear	[43],	[42]
Linear discriminant analysis	LDA	solver: svd	[44]	
K-Nearest Neighbor	KNN	n_neighbors = 5, weights: uniform	[45]	
Classification and regression trees	CART	criterion: gini	(B. LI, Friedman, Olshen, & Stone, 1984)	
Naive bayes	NB	priors: None	[46]	
K-means clustering	/	n_clusters = 2, init: k-means++, max_iter = 300	[39]	
Spectral clustering	/	n_clusters = 2, affinity: rbf, gamma = 1.0, assign_labels: kmeans	[47]	
Agglomerative clustering	/	n_clusters = 2, linkage: ward	[48]	
Adjacent frame difference	AFD	Gaussian smoothness(kernel: 5 * 5), threshold of binary image:25	/	/
White region selection in HSV space	WRS-HSV	Gaussian smoothness(kernel: 5 * 5), H: 0–180, S: 0–30, V: 220–255	/	
Fixed background subtraction	FBS	Gaussian smoothness(kernel: 5 * 5), threshold of binary image: 25	/	
Gaussian mixture background model	GMB	Gaussian smoothness(kernel: 5 * 5), threshold of binary image: 240	[36]	Opencv, [41]
K-Nearest Neighbor algorithm	KNN-B	history = 30, Gaussian smoothness(kernel: 5 * 5), threshold of binary image: 240	[23]	
Adaptive Gaussian mixture background model	AGMB	Gaussian smoothness(kernel: 5 * 5), threshold of binary image: 240	[23,35]	

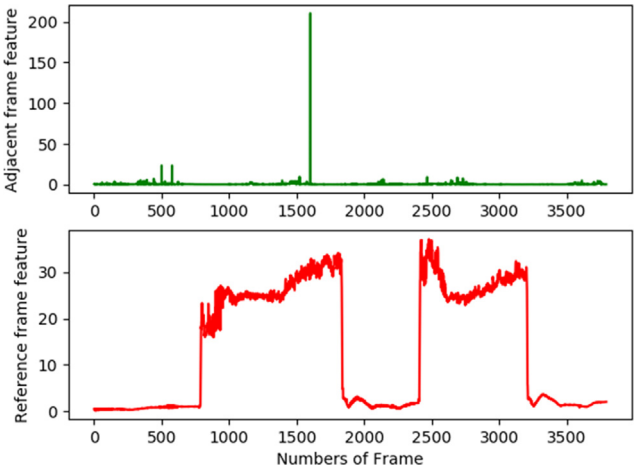


Fig. 7. Comparison of Dist value between conventional KLT algorithm and RF-KLT algorithm (video 1).

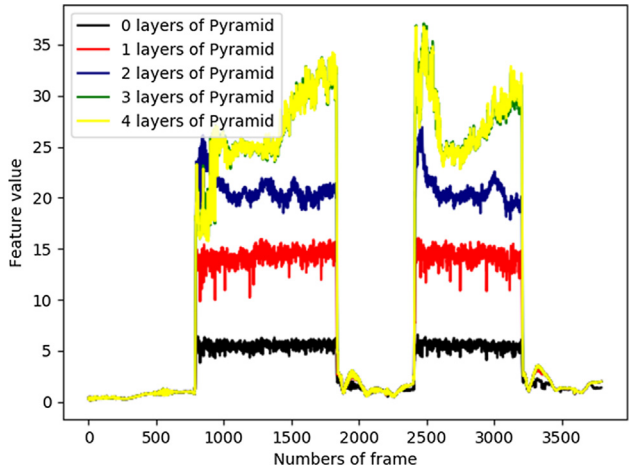


Fig. 8. Comparison of Dist value in different pyramid levels (video 1).

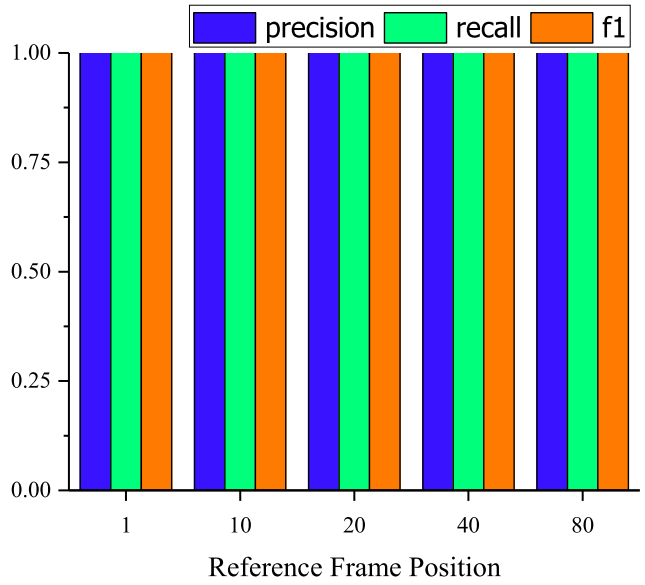


Fig. 9. Comparison of the influence of different reference frame in working state detection module in video 1.

Table 5
Comparison of different foreground detection algorithms on object region detection.

Video number	AFD	WRS-HSV	FBS	GMB	KNN-B	AGMB(Ours)
1	☑		☑	☑	☑	☑
2	☑	☑	☑	☑	☑	☑
3	☑	☑			☑	☑
4		☑	☑	☑	☑	☑
5	☑		☑	☑	☑	☑
6	☑	☑	☑	☑	☑	☑
7	☑	⊙	☑	☑	☑	☑
8	⊙	⊙	⊙	⊙	☑	☑

(☑ indicates that the final result of object region detection is accurate; ⊙ indicates that the final result of object region detection is inaccurate, but the object region is detected in the candidate regions; and empty table indicates that object region is not detected in the maximum contour regions.)

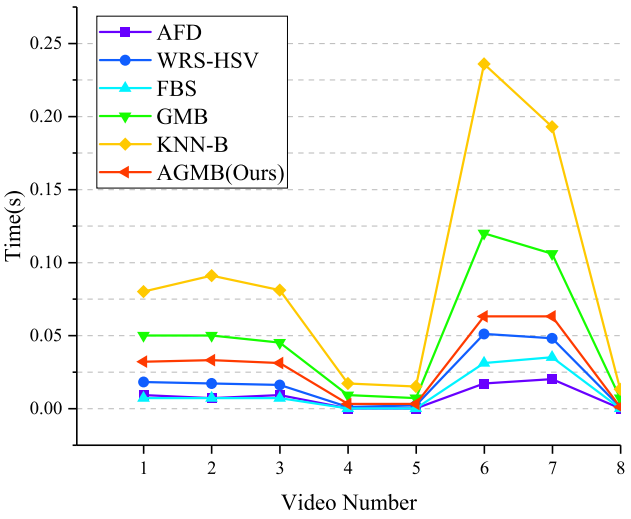


Fig. 10. Time cost comparison of different foreground detection algorithms.

several clustering algorithms and classification algorithms in video 1 and video 8, which represents simple and complex scenes respectively, is shown in Fig. 11. In this paper, 4/5 of feature dataset is used as training dataset, and the rest is used as test dataset. The average precision and standard deviation of the training dataset classification result are determined using a five-fold cross validation method. In Fig. 11, K-means clustering, spectral clustering and agglomerative clustering with ward linkage are used to label data. LR, LDA, KNN, CART, NB and SVM with linear kernel (details are in Table 4) are used to build the classifier for the annotated dataset.

The principle of selecting a clustering algorithm is based on the regularity of dataset: (1) two classes, (2) linear separation and obvious boundaries between classes, and (3) determination of the class labels corresponding to the different working state of aerators. Other clustering algorithms such as density clustering algorithms label data into multiple classes because the number of classes cannot be preset. This study compares three linear classifiers: LR, LDA, and SVM with linear kernel. At the same time, three nonlinear classification algorithms have been added for comparison. From the clustering result of video 1 in Fig. 11(a), there are no obvious differences in the three methods of the result in the clustering algorithm; and in the classification algorithm, the accuracy and standard deviation of KNN, CART, and SVM algorithms are all good. However, the KNN algorithm is worse in time cost, and the CART algorithm is also slightly slower than the SVM algorithm. As shown in Fig. 11(b), for the complex dataset such as the small object region in video 8, the SVM algorithm is also an optimal algorithm in accuracy, standard deviation, and time cost.

The average time performance of each frame used in clustering algorithms is shown in Fig. 12. The time performance of the K-

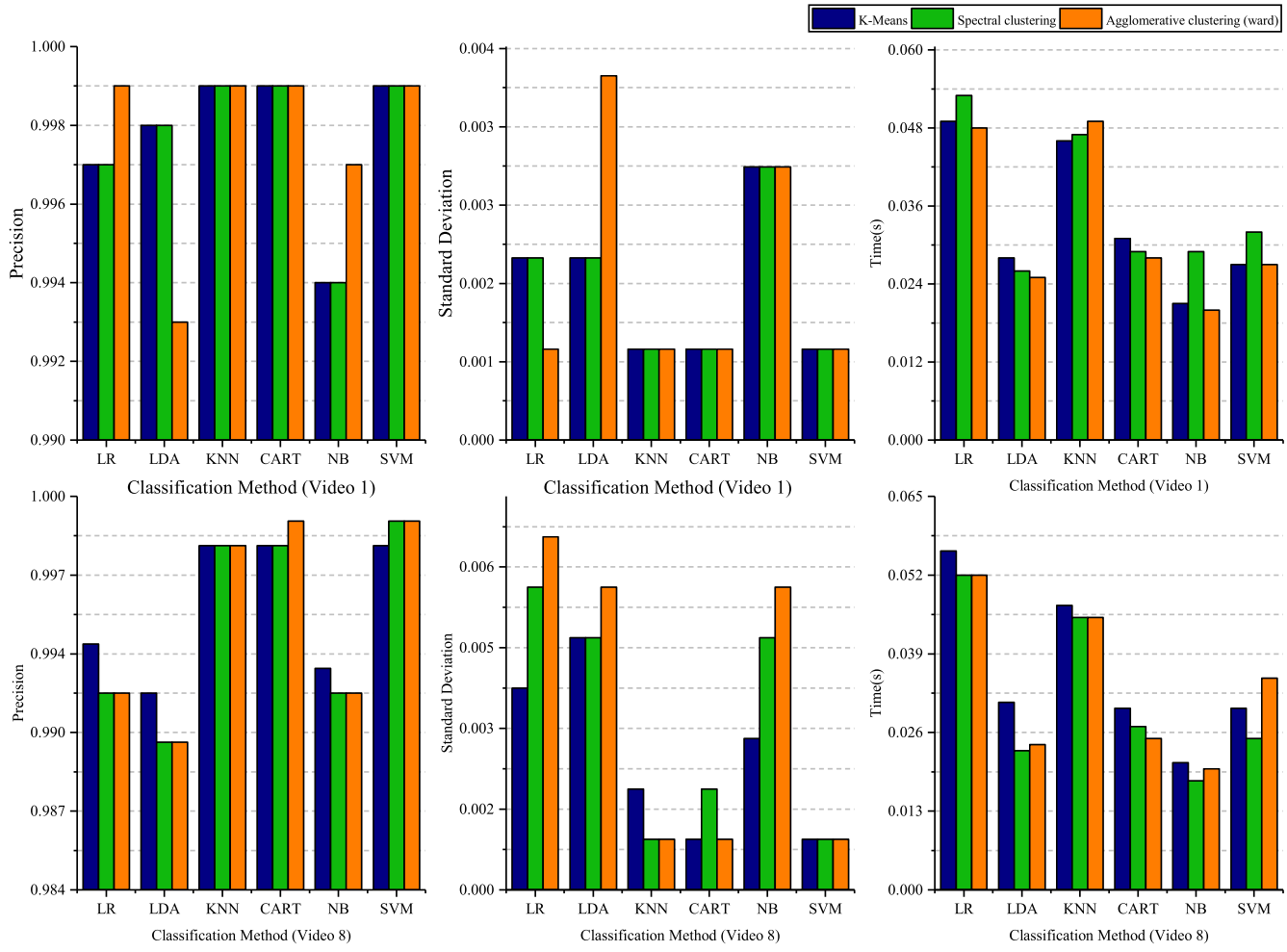


Fig. 11. Comparison of different machine learning algorithm combinations in video 1 and video 8.

means algorithm is one or two orders of magnitude faster than the other two clustering algorithms. According to the accuracy and time analysis, this paper selects the K-means algorithm for data annotation, and selects the linear SVM algorithm to train the classifier. Notably, the results of algorithm comparison also show the robustness and stability of the feature extraction method and dataset reconstruction method in this study, because the accuracy of dataset classification is closed to 100% in the combination of all the different machine learning algorithms. For example, the accuracy of our algorithm combination in training dataset is 99.9%.

As shown in Formula (4–4), in order to avoid labeling the error data as one class, and to prove that the two classes of data labeled by clustering algorithms are closed state and open state respectively, this paper designs a kind of evaluation index—class coefficient (CR), according to the Jaccard coefficient, to measure whether the data is correctly labeled.

$$CR = \frac{|N_{negative} - N_{openframe}|}{N_{frames}} \quad (4\text{–}4)$$

where $N_{negative}$ is the number of negative samples, i.e., the number of samples near the coordinate origin; $N_{openframe}$ is the frame number when the object region is detected in the training video; N_{frames} is the total number of frames in training video. Note that the CR coefficient is calculated based on working state changed once. It can measure the proportion of a correct sample in the total sample. Specifically, when it is detected that the frame of the object region appears, i.e., the aerator has been opened steadily, the sample data is converted to positive class, thereby the value of $N_{negative}$ is determined. The degree to which the CR is close to zero indicates how accurately the construction dataset was labeled. However, CR cannot be 0, because $N_{openframe}$ is determined when the object region is detected, the matching corner distance feature has reached its maximum value. Accordingly, some of the samples in the state

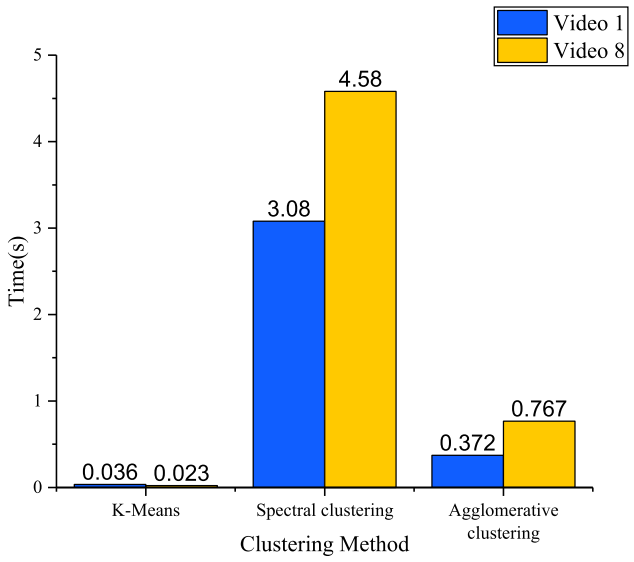


Fig. 12. Time cost comparison of different clustering algorithms.

Table 6
CR coefficients of different clustering algorithms in typical video dataset.

Video number	CR		
	K-means	Spectral clustering	Agglomerative clustering
1	0.023	0.023	0.023
8	0.059	0.059	0.059

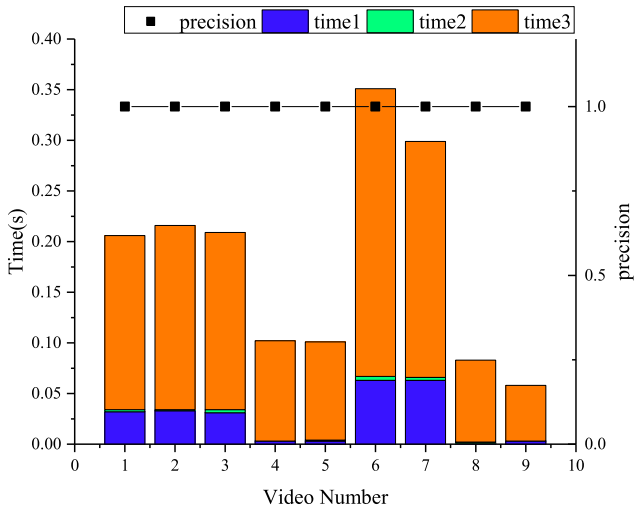


Fig. 13. Time cost and precision of object region detection module in different video datasets.

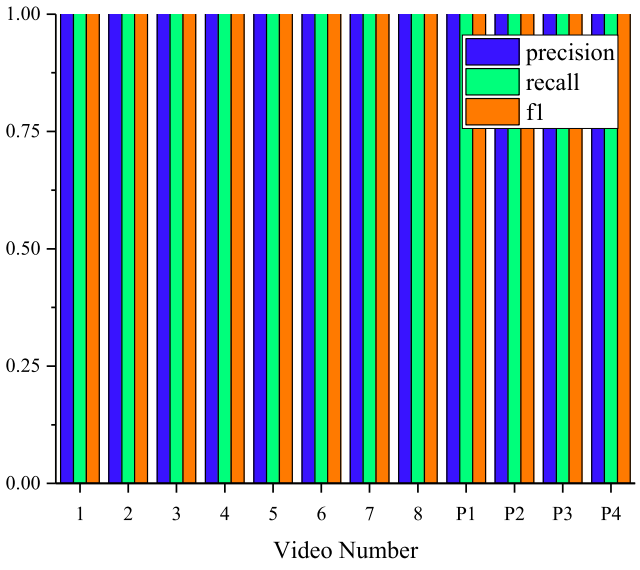


Fig. 14. Performance of working state detection for different video datasets.

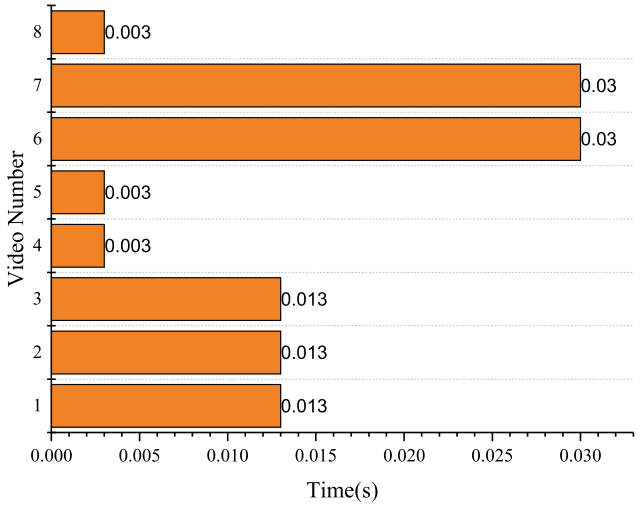


Fig. 15. Time cost of working state detection for different video datasets.

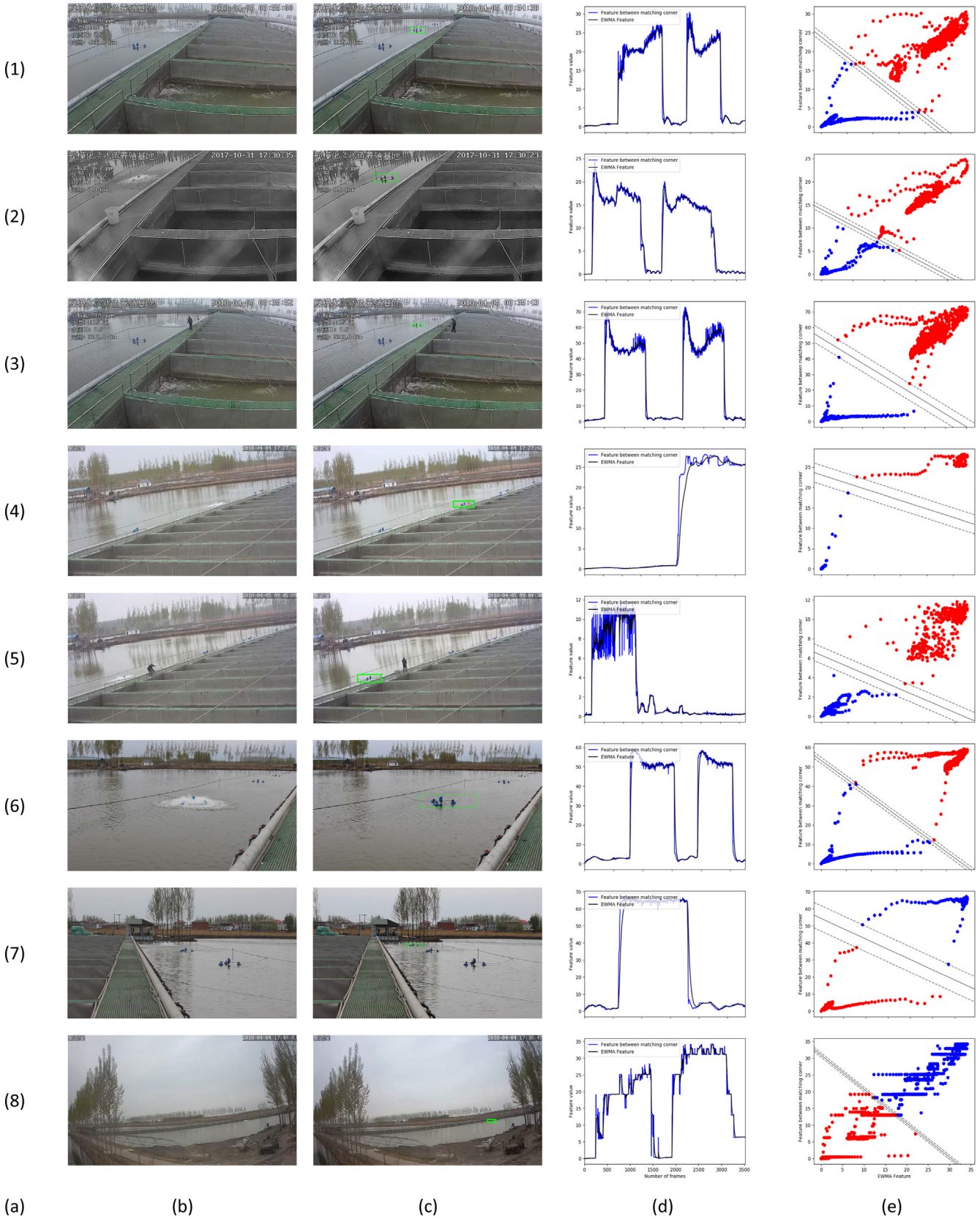


Fig. 16. The result of key algorithm flow in video dataset: (a) is the video number, (b) is a frame of the opening state of aerators in video dataset. (c) is the result of the object region detection. (d) is the feature curve based on the RF-KLT algorithm and the data after EWMA processing, and (e) is the result of dataset construction, data annotation and classification.

change process have been classified as positive class. Table 6 is the CR value of video 1 and video 8 corresponding to the above comparison experiment of machine learning algorithms, in which the video 8 only intercepts segments whose working state of object region changes once. The result shows that the clustering methods of this paper have higher accuracy in labeling samples, which also shows the obvious distance between classes of the constructed dataset.

4.3. The performance of our expert system

4.3.1. Object region detection

In Fig. 13, the time of each step in the object region detection module and the precision of feature point detection results in the test dataset are shown. Time1 is the detection time of each frame in maximum contour region detection step, time2 is the detection time of each frame in candidate regions detection step, and time3 is the detection time of each region in object region detection step. The detection of the maximum contour region takes less time, and the time for selecting the candidate object region is even more negligible. This is because the value of gray change in the F feature of object region is calculated along with the contour area, so this step is only a judging procedure. The proportion of the time taken to obtain the largest inter-class interval under the multi-level pyramid is large, which is why we suggest that the training video of the object region detection module should be between 10 and 20 s. Also, another reason is, as the number of training video frames increases, the number of candidate regions increases accordingly, resulting in an increase in the time for determining the final object region from the candidate regions.

The object region detection is mainly to prepare for the work state detection. The most crucial step in the feature extraction of work state detection is corner matching. The best result is that all detection corners fall on the body of aerators. In this study, the detection performance is evaluated according to Formula (4–1). However, *tp* is the corner point in object region that is correctly detected and *fp* is the corner point in object region that was detected incorrectly. As shown in Fig. 13, the object region is fully detectable according to the precision result, which means that the precision of object region detection is nearly 100%. Visualization of object region detection module is shown in Fig. 16(b).

4.3.2. Working state detection

The evaluation indicators, including precision, recall and f1-score, of the working state detection in the video datasets are shown in Fig. 14, and the detection time of each frame is shown in Fig. 15. In Fig. 14, the accuracy of discriminant model can reach 100% in test dataset in all scenes including videos with artificially increasing brightness and noise. This results mean that the features extracted based on the RF-KLT algorithm, the method of datasets constructed, and labeling method are all robust. In Fig. 15, videos 6 and 7 are supplement scenes with a moving camera. Due to the highest resolution, the detection time of each frame is longer. In short, the detection time from different types of surveillance cameras is different and the detection time increases with the higher resolution of the surveillance camera. The results show that the detection speed of real surveillance cameras is between 77 and 333 FPS, which is much higher than 25 FPS of general surveillance cameras. The result shows this method is real-time in many complex scenarios. Also, the results of the augmented video dataset artificially prove the stability of our method.

The result of object region detection, feature extraction, and EWMA data processing, data labelling, and classification in the algorithm flow of this paper are shown in Fig. 16. Fig. 16(c) shows that the detection of the object region by our method is accurate. Fig. 16(d) shows that features extracted based on the RF-KLT algorithm have distinct discrimination and stability under various scenarios. After the K-means clustering method labels the data, the support vectors of the SVM algorithm are all in the process of the working state change, that is, the feature vector from the beginning of the spray to reach the maximum region. There is a clear separation between the two classes of constructed dataset in Fig. 16(e).

However, in a very small object region, the number of detected corner points is small and the area of the spray region is small, resulting in a smaller corner distance feature. Therefore, the discrimination of datasets is not obvious. Fig. 17 shows the some results in algorithm flow for P1-P4, which are video datasets with artificially augmented interference. The results show that the influence of brightness is greater than that of noise, and the randomly changed brightness does not affect the experimental results. Obviously, a non-skipping increase in brightness does not affect the detection results, and even better situations may occur. Nevertheless, when the changed range of brightness is extreme and con-

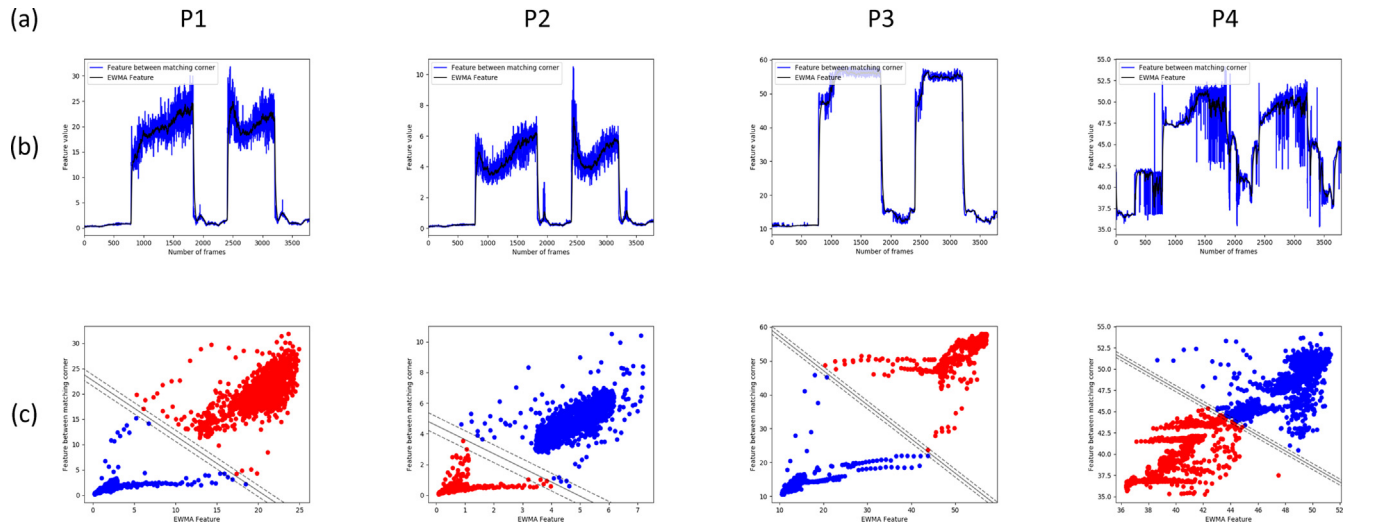


Fig. 17. The result of key algorithm flow in the artificially augmented dataset: (a) is the video number, (b) is the feature curve based on the RF-KLT algorithm and the data after EWMA processing, and (c) is the result of dataset construction, data annotation and classification. The artificially augmented video dataset is based on video 1, so the object region is the same as the object region in video 1.

stant, the instability of the feature data is increased. For example, when we increase the gray level of video 1 by 80 units, the feature distribution curve begins to be not smooth. This is because when the brightness of the object region changes sharply, the corner detection based on the feature value of gray variation will change, resulting in a change of the distance features between matching corners.

5. Conclusion and future work

In this paper, we presented a real-time expert system with existing surveillance cameras for anomaly detection of aerators, which consists of object region detection and working state detection. In the object region detection module, we proposed a regional selection method for detecting small object region from complex background based on region proposal idea [11]. This method includes three steps, i.e., maximum contour region detection, candidate region detection, and target region detection. In the working state detection module, we proposed the RF-KLT algorithm for robust motion feature extraction in fixed regions, which is simply based on the reference idea. The RF-KLT algorithm extends the applicability of the conventional optical flow method and breaks the limitation of adjacent frames. Moreover, we presented a dimension reduction method of time series for extracting the numerical distribution characteristics. Finally, we established an SVM classifier for work state judgment of aerators. The experimental results show that the accuracy of object region detection is 100%, the average accuracy of working state detection is 99.9%, and the detection speed is between 77 and 333 FPS according to different surveillance cameras. The proposed expert system can achieve real-time, zero-cost and accurate anomaly detection of aerators in complex and different application scenarios, which can be easily applied as part of the intelligent agriculture and agricultural expert system.

In future, we aim to improve this expert system so that it can be monitored for twenty-four hours, because this study was based on video dataset during the day and could not be monitored at night unless lighting equipment with high power was used, but it is a waste of energy. Accordingly, we expect to use surveillance cameras with infrared night vision equipment or other night vision equipment for imaging and analysis. Besides, the reference idea shows that the matching method of feature points with a certain reference frame can be used to detect motion state in videos. Therefore, different feature detection algorithms (e.g., Harris, SIFT, ORB, etc.) can be attempted. Finally, the proposed method for detecting small regions with fixed motion features from complex scenes, and the fast modeling method for numerical distribution of time series can be both applied to more practical scenarios.

CRediT authorship contribution statement

Yeqi Liu: Conceptualization, Methodology, Writing - original draft. **Huihui Yu:** Supervision, Writing - review & editing. **Chuanyang Gong:** Data curation, Writing - review & editing. **Yingyi Chen:** Funding acquisition, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declared that there is no conflict of interest.

Acknowledgments

This work is supported by the National Key R&D Program of China "Next generation precision aquaculture: R&D on intelligent measurement, control and equipment technologies" (No. 2017YFE0122100), and the Key R&D Program "Research and development of intelligent model and precise monitoring of shrimp processing" (No. 2018YFD0700904-2).

References

[1] A. Bardon-Albaret, E.A. Saillant, Effects of hypoxia and elevated ammonia concentration on the viability of red snapper embryos and early larvae, Aquaculture 459 (2016) 148–155.
[2] D. Solstorm, T. Oldham, F. Solstorm, P. Klebert, L.H. Stien, T. Vågseth, F. Oppedal, Dissolved oxygen variability in a commercial sea-cage exposes farmed Atlantic salmon to growth limiting conditions, Aquaculture 486 (2018) 122–129.
[3] Y. Chen, J. Xu, H. Yu, Z. Zhen, D. Li, Three-dimensional short-term prediction model of dissolved oxygen content based on PSO-BPANN algorithm coupled with kriging interpolation, Math. Probl. Eng. (2016) 1–10.
[4] C. Ma, D. Zhao, J. Wang, Y. Chen, Y. Li, Intelligent monitoring system for aquaculture dissolved oxygen in pond based on wireless sensor network, Trans. Chin. Soc. Agricul. Eng. 31 (2015) 193–200.
[5] C. Lu, X. Tang, Surpassing human-level face verification performance on LFW with Gaussianface, Comput. Sci. (2014).
[6] J. Janai, F. Güney, A. Behl, A. Geiger, Computer Vision for Autonomous Vehicles: Problems, Datasets and State-of-the-Art, arXiv.org, 2017.
[7] J. Wäldchen, P. Mäder, Plant species identification using computer vision techniques: a systematic literature review, Arch. Comput. Methods Eng. 25 (2018) 1–37.
[8] K.H.S. Murugan, V. Jacintha, S.A. Shifani, Security system using raspberry Pi, in: International Conference on Science Technology Engineering & Management, 2017, pp. 863-864.
[9] G. Jun, J.K. Aggarwal, M. Gokmen, Tracking and Segmentation of Highway Vehicles in Cluttered and Crowded Scenes (2016) 1–6.
[10] J. Shi, Good Feature to Track, in: IEEE Conference on Computer Vision and Pattern Recognition, 1994.
[11] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 580-587.
[12] Y.X. Jinhui He et al., Aerator state detection based on corner optical flow and SVM algorithom, Comput. Eng. Sci. 37 (2015) 1566–1572.
[13] S.K. Kumaran, D.P. Dogra, P.P. Roy, Queuing theory guided intelligent traffic scheduling through video analysis using Dirichlet process mixture model, Expert Syst. Appl. (2018).
[14] M.U.K. Khan, H.S. Park, C.M. Kyung, Rejecting Motion Outliers for Efficient Crowd Anomaly Detection, IEEE Transactions on Information Forensics & Security, PP, 2018, pp. 541–556.
[15] S. Frizzi, R. Kaabi, M. Bouchouicha, J.M. Ginoux, E. Moreau, F. Fnaiech, Convolutional neural network for video fire and smoke detection, Indust. Electron. Soc. IECON 2016 – Conf. IEEE (2016) 877–882.
[16] W.K. Lee, C.F. Leong, W.K. Lai, L.K. Leow, T.H. Yap, ArchCam: real time expert system for suspicious behaviour detection in ATM site, Expert Syst. Appl. (2018).
[17] R. Arroyo, J.J. Yebes, L.M. Bergasa, I.G. Daza, J. Almazán, Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls, Expert Syst. Appl. 42 (2015) 7991–8005.
[18] M. Wadhai, V.V. Gohokar, A. Khaparde, Agriculture pest detection using video processing technique, International Conference on Information Processing, 2015, pp. 544–547.
[19] M.C. Chuang, J.N. Hwang, J.H. Ye, S.C. Huang, K. Williams, Underwater Fish Tracking for Moving Cameras Based on Deformable Multiple Kernels, IEEE Transactions on Systems Man & Cybernetics Systems, 2017, pp. 1–11.
[20] K. Pasupa, N. Pantuwong, S. Nopparit, A comparative study of feature point matching versus foreground detection for computer detection of dairy cows in video frames, Springer-Verlag, New York Inc., 2015.
[21] K. Toyama, J. Krumm, B. Brumitt, B. Meyers, Wallflower: Principles and Practice of Background Maintenance, in: The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999, pp. 255-261 vol. 251.
[22] A. Sobral, A. Vacavant, A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos, Comput. Vis. Image Underst. 122 (2014) 4–21.
[23] Z. Zivkovic, V.D.H. Ferdinand, Efficient adaptive density estimation per image pixel for the task of background subtraction, Pattern Recogn. Lett. 27 (2006) 773–780.
[24] W.O. Li Liu, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu, Matti Pietikäinen, Deep learning for generic object detection: a survey, arXiv:1809.02165, 2018.
[25] X. Zhu, Y. Xiong, J. Dai, L. Yuan, Y. Wei, Deep feature flow for video recognition, CVPR (2016) 4141–4150.
[26] X. Zhu, Y. Wang, J. Dai, L. Yuan, Y. Wei, Flow-Guided Feature Aggregation for Video Object Detection, Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 408–417.
[27] S. Baker, I. Matthews, Lucas-Kanade 20 years on: a unifying framework, Int. J. Comput. Vision 56 (2004) 221–255.
[28] J.-Y. Bouguet, Pyramidal implementation of the Lucas Kanade feature tracker description of the algorithm, Opencv Documents 22 (1999) 363–381.

[29] T. Zhou, S. Gao, J. Wang, C. Chu, Y. Todo, Z. Tang, Financial time series prediction using a dendritic neuron model, Knowl.-Based Syst. 105 (2016) 214–224.
[30] M. Xu, M. Han, C.L.P. Chen, T. Qiu, Recurrent Broad Learning Systems for Time Series Prediction, IEEE Trans. Cybernet. (2018) 1–13.
[31] R. Dipietro, N. Navab, G.D. Hager, Revisiting NARX recurrent neural networks for long-term dependencies, arXiv: 1702.07805, 2017.
[32] Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang, G. Cottrell, A dual-stage attention-based recurrent neural network for time series prediction, Proc. Int. Joint Conf. Artif. Intell. (2017) 2627–2633.
[33] Y. Zheng, Y. Liang, S. Ke, J. Zhang, X. Yi, GeoMAN: multi-level attention networks for geo-sensory time series prediction, Proc. Int. Joint Conf. Artif. Intell. (2018).
[34] Y. Liu, C. Gong, L. Yang, Y. Chen, DSTP-RNN: a dual-stage two-phase attention-based recurrent neural network for long-term and multivariate time series prediction, Expert Syst. Appl. 143 (2019) 113082.
[35] Z. Zivkovic, Improved adaptive gaussian mixture model for background subtraction, Pattern Recog. Int. Conf. (2004) 28–31.
[36] P. Kaewtrakulpong, R. Bowden, An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection (2002) 135–144.
[37] B.D. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, in: International Joint Conference on Artificial Intelligence, 1981, pp. 674-679.
[38] H.Y. Kim, H.W. Chang, Forecasting the volatility of stock price index: a hybrid model integrating LSTM with multiple GARCH-type models, Expert Syst. Appl. 103 (2018).
[39] D. Arthur, S. Vassilvitskii, k-means++:the advantages of careful seeding, in: Eighteenth Acm-Siam Symposium on Discrete Algorithms, New Orleans, Louisiana, 2007, pp. 1027-1035.
[40] T.F. Wu, C.J. Lin, R.C. Weng, Probability estimates for multi-class classification by pairwise coupling, J. Mach. Learn. Res. 5 (2004) 975–1005.
[41] OpenCV, Open Source Computer Vision, Available at: http://opencv.org/, 2017.
[42] Scikit-learn, Machine learning in Python, Available at: http://scikit-learn.org/, 2017.
[43] M. Schmidt, N.L. Roux, F. Bach, Minimizing finite sums with the stochastic average gradient, Math. Program. 162 (2013) 1–30.
[44] B. Fan, Z. Lei, S.Z. Li, Normalized LDA for semi-supervised learning, in: IEEE International Conference on Automatic Face & Gesture Recognition, 2008, pp. 1-6.
[45] A. Munther, R. Razif, M. Abualhaj, M. Anbar, S. Nizam, A preliminary performance evaluation of K-means, KNN and EM unsupervised machine learning methods for network flow classification, international, J. Elect. Comput. Eng. 6 (2016) 778–784.
[46] J. He, Y. Zhang, X. Li, Y. Wang, Bayesian classifiers for positive unlabeled learning, International Conference on Web-Age Information Management, 2011, pp. 81–93.
[47] U.V. Luxburg, A tutorial on spectral clustering, Statist. Comput., 17, 2007, 395-416.
[48] G.J. Szekely, M.L. Rizzo, Hierarchical clustering via joint between-within distances: extending ward's minimum variance method, J. Classif. 22 (2005) 151–183.
[49] C.J. Harris, A combined corner and edge detector, Proc. Alvey Vision Conf. 1988 (1988) 147–151.