

RESEARCH ARTICLE

Tracking guided actions recognition for COWS

Yun Liang*, Xiaoming Chen

Guangzhou Key Laboratory of Intelligent Agriculture, College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China

* Correspondence: sdliangyun@163.com

Received January 27, 2021; Revised April 28, 2021; Accepted July 26, 2021

Background: Cows actions are important factors of cows health and their well-being. By monitoring the individual cows actions, we prevent cows diseases and realize modern precision cows rearing. However, traditional cows actions monitoring is usually conducted through video recording or direct visual observation, which is time-consuming and laborious, and often lead to misjudgement due to the subjective consciousness or negligence.

Methods: This paper proposes a method of cows actions recognition based on tracked trajectories to automatically recognize and evaluate the actions of cows. First, we construct a dataset including 60 videos to describe the popular actions existing in the daily life of cows, providing the basic data for designing our actions recognition method. Second, eight famous trackers are used to track and obtain temporal and spatial information of targets. Third, after studying and analysing the tracked trajectories of different actions about cows, a rigorous and effective constraint method is designed to realize actions recognition by us.

Results: Many experiments demonstrate that our method of actions recognition performs favourably in detecting the actions of cows, and the proposed dataset basically satisfies the actions evaluation for farmers.

Conclusion: The proposed tracking guided actions recognition provides a feasible way to maintain and promote cows health and welfare.

Keywords: public framework; actions recognition; visual tracking; precision agriculture; cows health

Author summary: People often use cows actions to achieve scientific feeding and improve cows welfare, and then promote the quality and production of cows. However, traditional cows actions monitoring is usually conducted through video recording or direct visual observation, which is time-consuming and laborious, and often lead to misjudgment due to the subjective consciousness or negligence. We propose a method of cows actions recognition based on tracked trajectories to automatically recognize and evaluate the actions of cows. The method provides a feasible way to maintain and promote cows health and welfare.

INTRODUCTION

By employing various kinds of networks, the internet of things catches comprehensive sensing data anytime, anywhere, and quickly transmits the data with locations or tags to its center servers to monitor objects [1]. Recently, people often use it to track cows and monitor environment in breeding sheds (such as gas compositions, humidity and temperatures) to improve cows welfare and their products while reducing the cost and

labour force. As shown in Fig. 1, the information of cows is obtained by smart cameras or sensors firstly. Then, it is transmitted to different computing equipments to do data fusion, storage and intelligent monitor process. Finally, the real-time monitoring information is spread to different users such as farmers, managers and experts to achieve precision cows rearing.

In the modern precision cows rearing, an important problem is how to recognize the actions of cows because cows actions reflect the health status of cows and affect

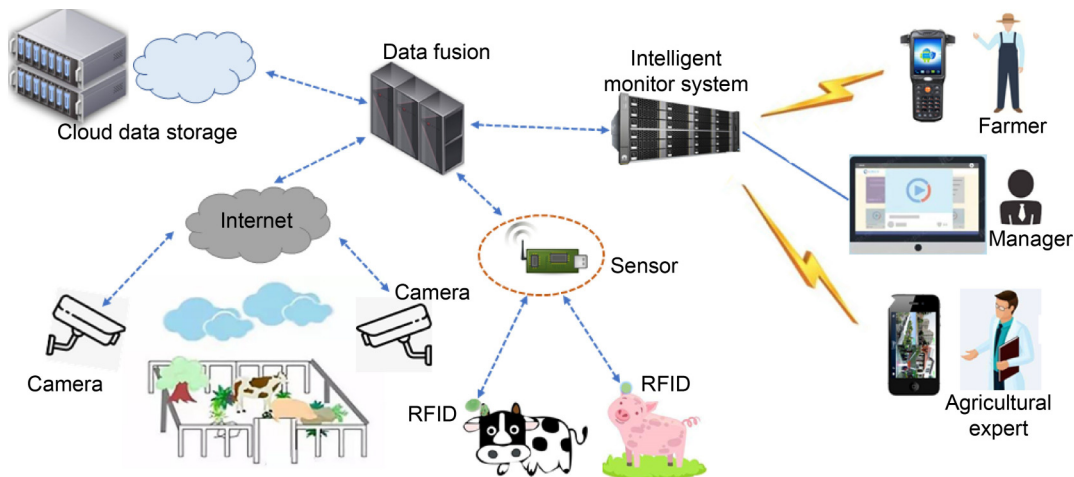


Figure 1. An example of precision cows rearing. By smart cameras and sensors, the sensing data is transmitted and computed by intelligent equipment on networks. Different users use the processed sensing data to monitor environment and breeding.

cows welfare. Real-time recognition of cows actions can detect diseases early and improve welfare [2–4]. For example, when a cow is suffering from lameness, it will cause a serious decrease in milk yield [5], and the performance is a significant reduction in standing time [6]; lying down for too long, the cow is more likely to be sick [7]; when the cow is in estrus, the walking time increases to four times that of normal condition [8,9]. Real-time recognition of the actions intensity of cows, such as quantity, duration, frequency, helps to grasp the health level and estrus status of cows in time, and provides a reference for improving the welfare of cows. Some actions recognition methods and framework have been proposed [10–12], but most of them focus on dealing with persons [13,14], motor vehicles [15–17] and so on. As far as we know, there are no researches about the actions recognition of cows. In fact, recognizing the actions of cows is one of the key factors about the precision rearing according to the expert opinion [18]. They help farmers to monitor the cows rearing and further to promote the quality and production of cows, such as milk.

In the traditional rearing model, people recognize the actions of cows by manual operation which leads to long implement time, high cost and manual errors [19]. This is unacceptable for farmers when they are striving between high profit and low cost. Therefore, it is extremely necessary to propose a framework to identify the actions of cows automatically and accurately [20]. By the identifying method from the framework, people monitor the health and welfare of cows based on the automatically computed actions, which is much more efficient, accurate and convenient with low cost than the

traditional methods. On one hand, with the actions information, farmers design scientific feeding schemes to improve the quality and quantity of cows, such as milk. On the other hand, in the research areas, cows actions are the important information for agricultural experts [18]. These experts use the actions information to study the comfort and emotion of cows, and the relationship between them and the produce capacity. Therefore, this paper proposes a framework for actions recognition of cows. Our main contributions are:

- A *public framework* is proposed to automatically recognize the popular actions of cows from their monitored videos. This framework performs as a part of intelligent monitor system about the intelligent agriculture which helps farmers to achieve precision cows rearing.
- For this framework, we set up a *dataset for cows*, which includes 60 videos with eleven challenges (such as occlusion and deformation). We define the ground truth (GT) for all videos which describes the target region on each frame of a video.
- For this framework, we define an algorithm to identify *five popular actions* of cows including *standing, walking, jumping, lying* and *running*. This proposed algorithm provides a normal form for people to define the method to recognize the actions of cows.
- We use eight trackers to track our cows dataset. From the tracking results, we find that different trackers are suitable for solving different challenges. Therefore, we summarize which tracker should be used for videos with different challenges.

RELATED WORK

This paper proposes a tracking guided actions recognition to automatically and accurately recognize the actions of cows. It is implemented in three phases: building a new dataset of cows, extracting actions information using related trackers, designing an algorithm to detect cows actions and validating the actions algorithm on the new dataset. Therefore, we describe the related work from the following three aspects.

Dataset for cows

This paper aims at recognizing and analysing the actions of cows. To track and obtain a large amount of available actions information of cows, a specialized dataset is required to describe the popular actions of cows. This dataset should include a great deal of videos which cover the popular actions existing in the daily life of cows. Then, people can design algorithms to recognize and analyse the actions of cows based on it.

However, although much effort has been devoted to the collection and annotation of image datasets containing thousands of image categories, cows actions datasets lag far behind. We learned about some public datasets to evaluate the actions and the trackers used in actions detections, such as the HMDB51 [11] and the UCF101 [12] for actions detection and the VOT dataset [14] and the TB-100 [16] for tracking. Regretfully, these actions datasets are constructed only for human actions [21]. However, these existing tracking datasets are mostly concerning human not cows (such as pedestrians and athletes [22]) and man-made objects (such as cars and bikes) [23]. Although some videos in VOT and TB-100 referring cows, which are too few and not good enough for the actions detection of cows, these actions data do not meet the needs of cows actions analysis. In addition, some people have constructed the data to deal with animals motions such as the work proposed in [19, 20]. However, these data are defined for catching the movements, they cannot calculate and analyse the trajectories of cows and identify their actions automatically.

To address this issue, we collected the most special actions video dataset of cows with five actions categories to-date. These videos are either shot at the cow farms or downloaded from the Internet. Similar to the existing datasets [14,16,24], we carefully choose these videos because they have some representative actions of cows and cover eleven kinds of challenges for tracking or actions recognizing (such as occlusion, motion blur, deformation). In the meanwhile, the number of these videos is 60 totally, which is enough for researchers to do the basic research of the actions

analysis of cows.

Tracking algorithms

We use some present famous tracking algorithms as the basic trackers to test our dataset and collect temporal and spatial information of the tracked cows. The correlation filter trackers [25,26] and the trackers based on deep learning [27–30] have recently shown excellent performance in tracking targets while dealing with some challenges such as target deformation, occlusion, illumination variation, cluster background, scale variation. In the visual tracker benchmark [8], totally about 29 trackers are described, which are also worth considering. Although many trackers have been proposed, we only select eight representative trackers for our framework to test our dataset. We select these trackers because they are very robust and fast in tracking target, and their codes or executable files have been published for free use. These eight trackers are the CSK [17], CT [31], DFT [32], KCF [33], RPT [34], BACF [35], SiamFC [36] and CREST [37].

To gain a better insight into the performance of trackers, we use each of the above trackers with their default and optimal parameters to test every video in our dataset. At the beginning, we manually specify the target region on the first frame of each video. Then in the rest frames, the states of target (namely the width, the height, the x-coordinate and y-coordinate of upper left point of target region) are automatically computed by one tracker. Finally, the recorded tracking results are used to achieve the automatic actions analysis of the related cows.

Actions recognition

Recently, many methods about actions recognition have been proposed [38, 39]. Most of them are defined for human but not for cows. Some most related works about actions recognition are as follows. Dollar *et al.* [40] defined the sparse spatio-temporal features to construct the method to recognize the behavior of human and rodents. Yin *et al.* [41] proposed a 3D facial expression database for facial behavior research, which focuses on 3D human facial expressions. Ben *et al.* [42] proposed an intrinsic sparse coding and dictionary learning formula for efficient coding of motion-recognized 3D skeleton sequences. Tang *et al.* [43] defined a deep progressive reinforcement learning method for actions recognition in skeleton-based videos. Gallego *et al.* [44] suggested a unified framework to solve several computer vision problems of event cameras: motion, depth, and optical flow estimation, which are used to find point trajectories to optimally align with event data for people

or cows. In addition, more methods about actions recognition can be reviewed from the work in [45–47].

A famous kind of actions detection method is based on the tracked trajectories of moving object [48, 49]. Following this way, we propose a new detection method: considering the tracking information, analysing the cow's trajectory and establishing a constraint rule to detect the cow's actions. We define five kinds of actions to describe the popular behaviours existing in the daily life of cows. These actions include *standing*, *walking*, *jumping*, *lying* and *running*. From the change in the target center, we derive target's displacement in the plane. Besides, we derive target's displacement outside the plane from the change of the target size. Constraint rules are designed to constrain physical quantities such as displacement and velocity, so as to achieve the purpose of detecting actions.

OUR FRAMEWORK AND ACTIONS RECOGNITION

In this section, we first propose a dataset about the actions of cows. Then, we define the method to recognize the actions based on the state changes of cows. Finally, we evaluate how to get the states of cows by the existing eight trackers.

Our dataset

Our dataset includes 60 videos, with which we recognize the representative and popular actions of cows. Figure 2 shows the first frames of these videos and the targets specified by the red rectangles. This dataset not

only provides some important references for cows breeding experts, but also represents the common challenges in the actions recognition of cows in their daily activities. Here, we use the same challenges defined in the framework for trackers [16], because all the challenges of the two frameworks are introduced by the same problems namely the changes of target and background. Totally, eleven challenges are used in this paper.

The videos of our dataset as shown in Fig. 2 are collected by downloading from Internet or shooting in different environments with various directions and angles. The yellow words are the name for each video. The red rectangle describes the target used to do actions recognition. We manually specify the target rectangle on each frame for every video, and record four parameters of its state including the width, height, the x-coordinate and y-coordinate of the rectangle center. The fifth parameter namely the depth of target is computed based on the above four parameters and the size of a video. This state record is the ideal value to analyse the actions. But when to automatically recognize the actions, we need a tracker to automatically compute the target rectangle of each frame based on the target object from the first frame. At that time, the specified rectangles of this dataset are used as the GT (ground truth) to evaluate the efficiency of all the trackers used in our framework.

Our actions include *standing*, *walking*, *jumping*, *lying* and *running*. For example, we construct seven videos (from Jump1 to Jump7 in Fig. 2) to describe the different kinds of jumping. We divide our dataset into five classifications based on the action of each video. Then, the proposed 60 videos are denoted by: seven of



Figure 2. Our dataset. Each image with red rectangle circled target is the first frame of a video.

them are for jumping, while another seven videos for lying, eight videos for running, eighteen videos for standing and twenty videos for walking.

Actions analysis

This paper proposes the algorithm to recognize actions of cows based on the state changes of target object. By the tracking trajectories and motion tendencies of a target, we obtain the changes of its states. This section utilizes the videos about cows as an example to describe our method. To avoid the interferences caused by moving camera and other man-made factors, the cameras are required to be fixed to shoot the videos. Here, a cow to be analyzed performs only one action in a video. If a cow does multiple actions in a video, we will first divide the video into multiple sub-videos, so that the cow performs only one action in each sub-video.

To analyse the changes of target states, we define five parameters to record the state. As each target state is described by a rectangle, we first define four parameters based on it. They are cx (x -coordinate of target rectangle center), cy (y -coordinate of target rectangle center), w (the width of target rectangle) and h (the height of target rectangle). These parameters vary greatly in different kinds of actions. For instance, when a cow walks horizontally, its cx changes a lot while its cy and size remain unchanged. Totally, we define five actions, namely *standing*, *walking*, *jumping*, *lying* and *running*. For the changes of states are related to the depth of shooting scene, we introduce a new parameter d to describe the depth of target. For example, when the cow is far away from the camera, its small change in size or displacement may lead to big movement. The depth d is defined by:

$$d = \vartheta_d * \sqrt{I_w * I_h / S}, \quad (1)$$

where S ($S = w * h$) is the area of target rectangle, I_w and I_h are separately the width and height of the image. ϑ_d is the coefficient and set to be 300 based on our experiments. As d is only related to the size of target and frame, it successfully describes the movement of a target along different directions under the fixed camera. Then d is an important parameter in addition to cx, cy, w, h .

The change rate of each parameter is an important measurement for actions identification. For example, both *walking* and *running* have great changes in the size and depth of target, but *running* has much bigger change rate than *walking*. We first define the range of those parameters:

$$\Delta_f = f_{\max} - f_{\min}. \quad (2)$$

f refers to parameters (t, cx, cy, w, h, d). If f is t , $\Delta_f = \Delta_t$

is the number of frames, in which t_{\min}, t_{\max} are the first and the last frame. If f is cx , $\Delta_f = \Delta_{cx}$ is the range for the cx and cx_{\min}, cx_{\max} is its minimum and maximum in a video. Similarly, $\Delta_f = \Delta_{cy}$ is for cy , while $\Delta_f = \Delta_w$ is for target width, $\Delta_f = \Delta_h$ is for target height and $\Delta_f = \Delta_d$ is for dept. Then, we define the change rate $v_d, v_{cx}, v_{cy}, v_{total}$ by:

$$\begin{cases} v_d = (\Delta_d / \Delta_t) * \sqrt{I_w * I_h / \bar{S}}, \\ v_{cx} = (\Delta_{cx} / \Delta_t) * \sqrt{I_w * I_h / \bar{S}}, \\ v_{cy} = (\Delta_{cy} / \Delta_t) * \sqrt{I_w * I_h / \bar{S}}, \\ v_{total} = \sqrt{v_{cx}^2 + v_{cy}^2 + v_d^2}, \end{cases} \quad (3)$$

where v_d, v_{cx}, v_{cy} are the change rates of d, cx, cy for a video. v_{total} is their combination and describes the move rate in any direction. We use $\sqrt{I_w * I_h / \bar{S}}$ as the weight to adjust the change rates where \bar{S} is the average area of target rectangle to reduce the interference from the different distance between target and camera. By using the above five parameters and its change rates, we define the actions of cows.

The action of *standing*

Standing describes the action that the body of a cow remains in a small region in a video. Then in this action, the change of parameters is very small even equal to zero. Figure 3 describes the action of a standing cow. From Fig. 3A, we know the state of the cow nearly remain same from #0086 to #0211. We define *standing* action by:

$$\begin{cases} \Delta_{cx} < \vartheta_{cx} * \bar{w}, \\ \Delta_{cy} < \vartheta_{cy} * \bar{h}, \\ \Delta_w < \vartheta_w * \bar{w}, \\ \Delta_h < \vartheta_h * \bar{h}. \end{cases} \quad (4)$$

$\vartheta_{cx}, \vartheta_{cy}, \vartheta_w, \vartheta_h$ are set to 0.15 based on our experiments. \bar{w}, \bar{h} are the average of w and h in a video. The first item in Eq. (4) requires that the change of cx must be limited in a small range based on the target width. When \bar{w} is small, it means the cow is far from camera. At this situation, the small-scale change of cx refers to big-scale displacement. Therefore, we use \bar{w} to define the reasonable range of cx . The situations of Δ_{cy}, Δ_w , and Δ_h are similar. Meanwhile, d remains unchanged. In this paper, we describe the changes of target states for a video by the curves as Fig. 3B. The red rectangles as shown in Fig. 3A describe the target regions. The red, green, blue, black and pink curves in Fig. 3B separately describe the changes of five parameters frame by frame.

The $\ast\blacklozenge\blacklozenge$ represent the positions of the frames in Fig. 3A on these curves. In this paper, we take the upper left corner of an image as the origin point of coordinates. The yellow point in an image describes the center location of the target. As shown in Fig. 3B, the curves of cx, cy, w, h and d nearly maintain unchanged in the whole video which means that the cow is in the state of *standing* by Eq. (4).

The action of *walking*

Walking describes the action that a cow moves in a small speed. Taking cows as example, for the movement occurs in any direction, its cx can change on small-scale or large-scale (as Fig. 4A, a cow walks horizontally). However, the body of cow fluctuates little in this action, then the cy varies in a small region. At the same time, the moving speed is neither too fast nor too slow. Otherwise when the speed is very high, the cow is running. Therefore, we define the action of *walking* as:

$$\begin{cases} \Delta_{cy} < \vartheta_{cy} \ast \bar{h}, & \vartheta_{cy} = 0.25, \\ \alpha_1 < v_{total} < \alpha_2, & \alpha_1 = 10, \alpha_2 = 30. \end{cases} \quad (5)$$

The first item of Eq. (5) limits the range of Δ_{cy} within a certain threshold decided by the coefficient ϑ_{cy} and the average of target height \bar{h} . We set $\vartheta_{cy} = 0.25$ based on the experimental experience. By the second item, the

walking speed v_{total} is also within a certain range, which we set it between 10 and 30 for the cow. Figure 4 describes the walking action of a cow from #0006 to #0121. For this video, the Δ_{cy} is $17 < 49(\vartheta_{cy} \ast \bar{h})$ and the v_{total} is 21.3 (< 30), so we conclude that the cow is walking base on Eq. (5).

Figures 5 and 6 describe the action of *walking* along random direction with similar performance. Take Fig. 5 as an example, the curves in Fig. 5B change as follows, The d curve declines slowly which means the target is coming to camera. The cx curve ascends slowly which indicates the target is moving to right, and the cy curve slightly increases which indicates the body fluctuation is very small. For this video, by its $\Delta_{cy} = 46 < 73(\vartheta_{cy} \ast \bar{h})$, $v_{total} = 24.1$ from #0042 to #0100 frame, we conclude the *walking* action by Eq. (5).

The action of *jumping*

Jumping describes the action that a cow jumps up and down greatly in a period. The main characteristic for this action is that the cy of target rectangle changes in a large-scale. In vertical jumping, the cy of cows fluctuates a lot in a short time. Generally, the jumping of cows usually are repeated in a short time, which caused many continuous local maximums and minimums on the curve of cy . For example, when jumping arrives at the



Figure 3. The action of *standing* about video (stand11). (A) Six target rectangles and their centres; (B) The five curves of state parameters.

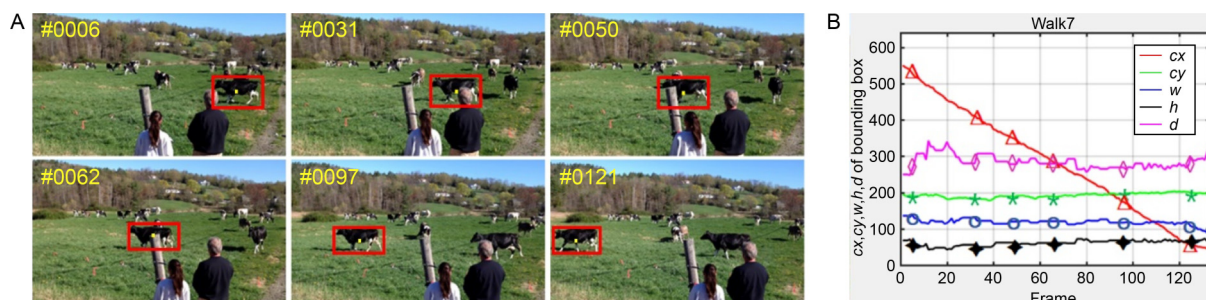


Figure 4. The action of *walking* about video (walk7). (A) Six target rectangles and their centres; (B) The five curves of state parameters.

lowest point, a local minimum appears. To differentiate with the action of *running*, the speed of cy is limited in a small range. Therefore, we define the action of *jumping* by:

$$\begin{cases} \Delta_{cy} > \vartheta_{cy}^1 * \bar{h}, & \vartheta_{cy}^1 = 0.5, \\ \exists t(cy_t + \vartheta_{cy}^2 * \bar{h}) < \cup(cy_{t-k}, cy_{t+k}), k > 0, & \beta_1 = 10, \beta_2 = 30. \\ \beta_1 < v_{total} < \beta_2, \end{cases} \quad (6)$$

The first item in Eq. (6) requires that the Δ_{cy} must be bigger than the average height of target \bar{h} . The second item describes the minimum at time t by setting cy_t ,

smaller than the element in set \cup which includes cy_{t-k} at time $(t-k)$ and $cy_{t+k}(t+k)$. The third item limits the speed of cows in a small range in the action of *jumping*. In this paper, we set $\vartheta_{cy}^2 = 0.3$, $k = 5$ for the second item based on many experiments.

Figure 7 describes an example of *jumping* action for a cow. As shown in Fig. 7B, according to the green curve, we know that the cy fluctuates greatly, and many minimums and maximums appear. At the same time, the moving speed ($v_{total} = 22.4$) is not fast. Therefore, we detect the cow in Fig. 7 is jumping based on Eq. (6) and this is also satisfying the intuitive judgements of users.



Figure 5. The action of *walking* about video (walk14). (A) Six target rectangles and their centres; (B) The five curves of state parameters.

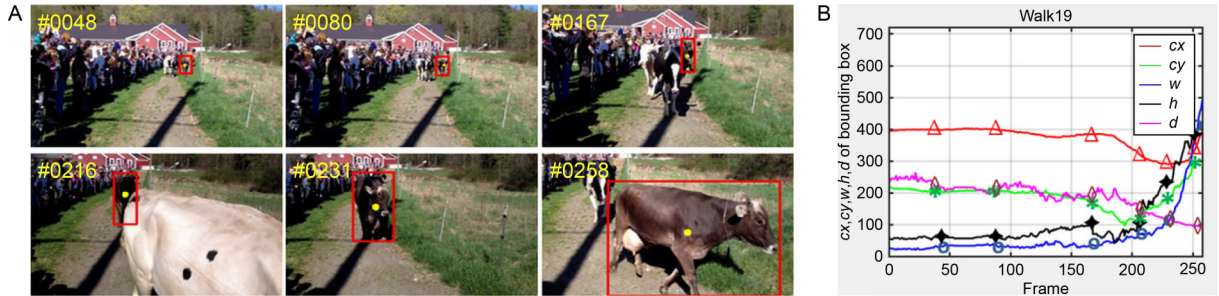


Figure 6. The action of *walking* about video (walk19). (A) Six target rectangles and their centres; (B) The five curves of state parameters.



Figure 7. The action of *jumping* about video (jump2). (A) Six target rectangles and their centres; (B) The five curves of state parameters.

The action of *lying*

Lying describes the action that how a cow achieves lying on land or its bed from standing state. We decide this action by three continuous steps. First, the cow is standing. We decide it by the small changes of the center point and size of the cow. Second, the cow is changing its state from standing to lying. We decide it by the great declines about the cy . Third, the cow is lying on land or its bed and just shaking its body in a little range. The center position and size remain unchanged in this step. Therefore, we define the action of *lying* by:

$$\begin{cases} t_{cy_min} < t_{cy_max}, \\ \Delta_{cy}^{fir} < \vartheta_{cy}^{fir} * \bar{h}, & \vartheta_{cy}^{fir} = 0.15, \\ \Delta_{cy}^{sec} > \vartheta_{cy}^{sec} * \bar{h}, & \vartheta_{cy}^{sec} = 0.5, \\ \Delta_{cy}^{thi} < \vartheta_{cy}^{thi} * \bar{h}, & \vartheta_{cy}^{thi} = 0.15. \end{cases} \quad (7)$$

As the coordinates for an image start from its upper left corner, the coordinates become bigger and bigger from left to right and from top to bottom. Therefore, the first item of Eq. (7) indicates that the cy changes from small to large to describe the state change from standing to lying. Here, we use $\Delta_{cy}^{fir}, \Delta_{cy}^{sec}, \Delta_{cy}^{thi}$ to separately describe the Δ_{cy} in the first, second and third step. Then, the second item describes the first step of lying down which requires Δ_{cy}^{fir} is less than a threshold namely $\vartheta_{cy}^{fir} * \bar{h}$. The third item describes the second step which requires the Δ_{cy}^{sec} is larger than a threshold namely $\vartheta_{cy}^{sec} * \bar{h}$. Meanwhile, the cy changes considerably in a short time before reaching t_{cy_max} . The fourth item describes the third step which requires the Δ_{cy}^{thi} is less than the threshold $\vartheta_{cy}^{thi} * \bar{h}$. It means that cy is unchanged from t_{cy_max} to the last frame. Finally, we set $\vartheta_{cy}^{fir} = 0.15$, $\vartheta_{cy}^{sec} = 0.5$, $\vartheta_{cy}^{thi} = 0.15$ based on our experiments.

Figure 8A shows an example of *lying*. Figure 8B describes the parameters curves for this action.

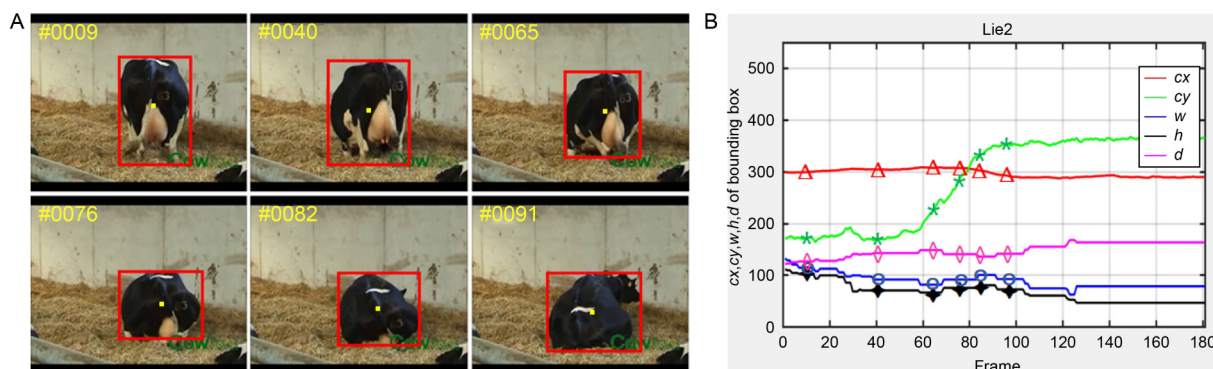


Figure 8. The action of *lying* about video (lie2). (A) Six target rectangles and their centres; (B) The five curves of state parameters.

According to the green curve in Fig. 8B, the cy is almost unchanged from #0001 to #0055. Then, from #0056 to #0096 frame, it increases greatly. Finally, from #0097 to the last frame, the cy is unchanged again. Then, by Eq. (7), we detect that it is the action of *lying*.

The action of *running*

Running describes the action that a cow moves in a high speed on any direction. For this action, the state of cows changes faster. It means that the speed of running cow is higher than that of walking and jumping. In addition, compared to walking, the body of running cow fluctuates much greater and faster especially on the parameter cy . Figures 9A and 10A separately show the running in horizontal direction and top-right direction. Based on the above analysis, we define the action of *running*:

$$\begin{cases} \Delta_{cy} > \vartheta_{cy} * \bar{h}, & \vartheta_{cy} = 0.25, \\ v_{total} > \gamma, & \gamma = 30. \end{cases} \quad (8)$$

The first item of Eq. (8) makes the change range of cy larger than a certain threshold decided by the coefficient ϑ_{cy} and the average height of target \bar{h} . The second item of Eq. (8) requires that the speed of v_{total} must be greater than the threshold γ . According to many experiments, we set $\vartheta_{cy} = 0.25$ and $\gamma = 30$. As $v_{total} = 37$ in Fig. 9, we denote it as *running* by Eq. (8).

As shown in Fig. 10B, the curve of d slowly declines, and the curve of w increases greatly. This indicates that the target is close to camera. At the same time, both the curves of cx and cy increases. It means the movement of target occurs along both horizontal and vertical direction. In addition, its v_{total} is equal to 36.8. Then, the action is detected as *running* by Eq. (8). On the contrary, an increase in d in Fig. 11B indicates that the target is moving away from the camera. Moreover, its v_{total} is equal to 34.2 that the action is also detected as *running* by Eq. (8).

Challenges of videos in our dataset

We list the challenges of every video in our dataset as shown in Table 1. We use the same eleven challenges from the framework [16] to denote the challenges of our dataset. These challenges include illumination variation (IV), scale variation (SV), occlusion (OCC), deformation (DEF), motion blur (MB), fast motion (FM), in-plane rotation (IPR), out-of-plane rotation (OPR), out-of-view (OV), background clutters (BC), and low resolution (LR). Please review more details about these challenges in Tracking framework [16].

As shown in Table 1, the challenges are corresponding to different actions of cows, meanwhile the same kind of actions usually have similar challenges. For example, the challenges for the actions about moving are much more complicated than those in static state. Specifically speaking, the running or jumping cows will always bring more challenges especially MB, FM, SV and so on. In addition, as cows are gregarious animals, the challenges such as BC, OCC and OV are very common in the videos of its daily life. Moreover, cows often walk outdoors so that the variation of illumination also affects the performance of the tracking algorithms.



Figure 9. The action of *running* about video (run1). (A) Six target rectangles and their centres; (B) The five curves of state parameters.

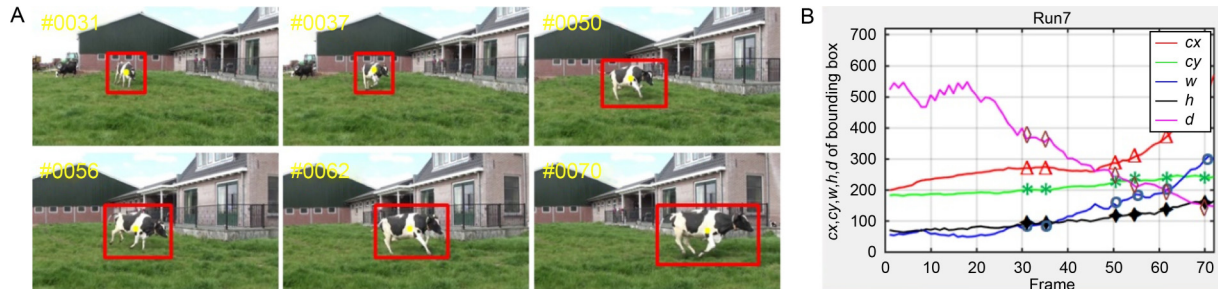


Figure 10. The action of *running* about video (run7). (A) Six target rectangles and their centres; (B) The five curves of state parameters.

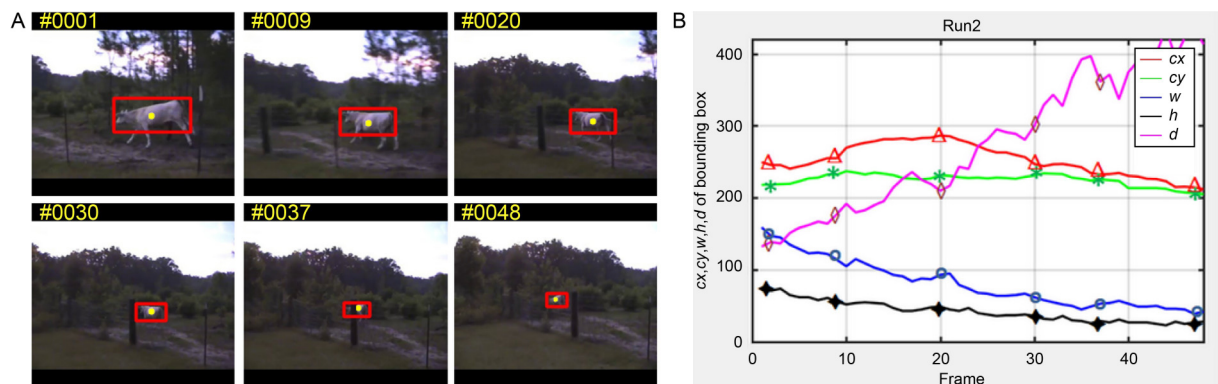


Figure 11. The action of *running* about video (run2). (A) Six target rectangles and their centres; (B) The five curves of state parameters.

Table 1 The challenges of videos in our dataset shown in Fig. 2

Videos	Challenges							
lie1	BC	LR						
lie2	SV	DEF	IPR					
lie3	BC	SV	OV					
lie4	BC	DEF						
lie5	BC							
lie6	BC	OCC						
lie7	IV	BC	LR	MB				
run1	SV	OCC	DEF	FM	BC	IPR	OPR	
run2	IV	SV	OCC	MB	FM	LR		
run3	FM	SV	OV	BC				
run4	FM	BC	OCC	IPR				
run5	SV	BC	LR	IPR	FM			
run6	SV	DEF	FM	IPR				
run7	SV	FM	MB	IPR				
run8	SV	DEF	FM	OPR	MB			
walk1	SV	OCC	IPR	BC				
walk2	SV	BC	IPR					
walk3	OCC	IPR	BC	SV				
walk4	MB	BC	OCC	LR	OV	IPR	SV	
walk5	BC	MB	OCC					
walk6	IPR	SV	DEF					
walk7	OCC	BC	OV	FM				
walk8	BC	MB	LR					
walk9	OCC	BC	OV	SV				
walk10	LR	MB	SV					
walk11	SV	DEF	IV	FM	OPR			
walk12	SV	BC						
walk13	SV	BC	IPR					
walk14	IPR	BC	SV					
walk15	OCC	OPR						
walk16	OPR	MB	SV					
walk17	SV	OCC	OV	OPR				
walk18	SV	BC	OPR	DEF				
walk19	MB	BC	OCC	SV	DEF			
walk20	OCC	MB	SV	DEF				
jump1	OCC	DEF	FM	OPR	SV	BC		
jump2	FM	DEF	SV	OPR	BC			
jump3	FM	DEF	IPR					
jump4	FM	SV	OCC	IPR	BC			
jump5	SV	OCC	FM	DEF	BC	IPR		
jump6	FM	SV	OCC	OPR	DEF			
jump7	SV	DEF	OPR	IPR	FM			
stand1	SV	DEF						
stand2	BC	OCC						
stand3	IPR	DEF						
stand4	DEF	OPR	SV					
stand5	BC	DEF						
stand6	BC	SV	DEF					
stand7	SV	DEF	IPR	OV				
stand8	IV	BC	SV	DEF	OPR			
stand9	SV	DEF	OPR					
stand10	IPR	OCC	DEF					
stand11	SV	IPR	DEF					
stand12	BC	OCC						
stand13	SV	DEF	OV	IPR				
stand14	BC	LR	MB					
stand15	BC	OPR	MB					
stand16	BC	OCC	MB					
stand17	LR	MB	BC					
stand18	IV	LR	OCC					

TRACKING RESULTS AND ACTIONS RECOGNITION RESULTS

Our experiments are implemented based on matlab 2016a on a regular PC (64-bit win10 operating system, Intel Core i5-4200 H 2.80 GHz processor, NVIDIA GeForce GTX 950 M graphics card, 4 GB RAM). To verify the reliability, rationality, and diversity of our dataset, we choose eight trackers, including five classical trackers from visual tracker benchmark [16]: CSK [17], KCF [33], CT [31], DFT [32], RPT [34], a correlation filter tracker BACF [35], two deep learning trackers SiamFC [36] and CREST [37], and analyse the robustness and accuracy of these trackers on the videos of our dataset. Then, the users of our benchmark can select the right tracker according to the evaluations of these trackers.

We use precision and success rate to quantitatively analyse the tracking effect of the trackers. In tracking precision evaluation, a widely used standard is the center position error, which is defined as the euclidean distance between the center position of the tracking target and the ground truth. We calculate the percentage of frames whose center position error is less than the error threshold to evaluate the tracking precision of a tracker. We use a precision error threshold equal to 20 pixels.

Accurate tracking not only needs to determine the center of the target, but also to determine the size of the target. Another evaluation criterion is the overlap rate of the bounding box. Assuming that the tracking bounding box is B_t and the ground truth bounding box is B_{gt} , the overlap rate is defined as $S = |B_t \cap B_{gt}| / |B_t \cup B_{gt}|$. We calculate the percentage of frames whose overlap rate S is greater than the given threshold to evaluate the tracking success rate of a tracker. The threshold of overlap rate we used is 0.5.

Figure 12 shows the average tracking precision and success rate of eight trackers on 60 cows videos. The tracker BACF based on correlation filtering achieves the best results both in tracking precision and success rate. In terms of precision, it's 0.24 higher than the second-ranked SiamFC; in terms of success rate, it's 0.27 higher than the second-ranked CREST, which means that BACF is more suitable for tracking our cows dataset. Trackers based on deep learning, SiamFC and CREST, ranked second in precision and success rate, respectively. This shows that trackers based on deep learning are more accurate than traditional trackers, but when tracking new objects, they take time to retrain to fully show their advantages.

Tables 2 and 3 respectively show the tracking precision and success rate of eight trackers for different

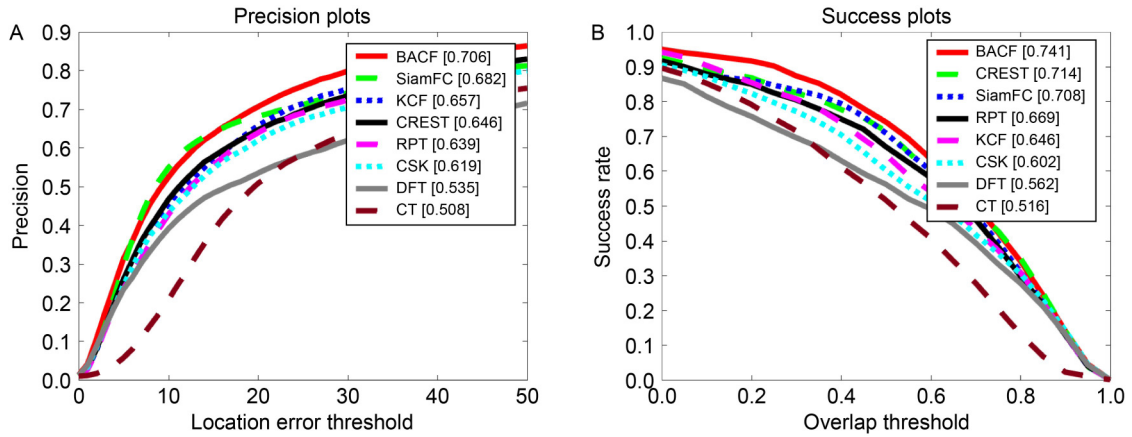


Figure 12. Precision and success rate of eight trackers on the cows dataset. (A) Precision of eight trackers; (B) Success rate of eight trackers.

Table 2 Precision of eight trackers for eleven challenges

Challenges	Trackers							
	SiamFC	CREST	BACF	KCF	RPT	CSK	DFT	CT
IV	1.000	0.671	0.736	0.695	0.741	0.964	0.708	0.711
SV	0.605	0.539	0.590	0.553	0.532	0.486	0.419	0.448
OCC	0.536	0.525	0.599	0.556	0.536	0.539	0.457	0.366
DEF	0.667	0.586	0.630	0.603	0.622	0.586	0.523	0.466
MB	0.813	0.748	0.854	0.744	0.706	0.705	0.520	0.623
FM	0.520	0.497	0.534	0.474	0.493	0.463	0.378	0.340
IPR	0.585	0.488	0.588	0.594	0.531	0.507	0.525	0.392
OPR	0.599	0.473	0.548	0.453	0.454	0.453	0.243	0.390
OV	0.348	0.484	0.531	0.456	0.331	0.331	0.243	0.226
BC	0.675	0.672	0.739	0.700	0.665	0.665	0.553	0.520
LR	0.859	0.774	0.915	0.845	0.742	0.781	0.676	0.568

Table 3 Success rate of eight trackers for eleven challenges

Challenges	Trackers							
	SiamFC	CREST	BACF	KCF	RPT	CSK	DFT	CT
IV	0.842	0.807	0.818	0.646	0.756	0.669	0.647	0.490
SV	0.619	0.613	0.646	0.532	0.582	0.463	0.433	0.423
OCC	0.598	0.631	0.682	0.612	0.602	0.577	0.539	0.459
DEF	0.704	0.693	0.710	0.611	0.645	0.576	0.540	0.480
MB	0.694	0.707	0.769	0.663	0.662	0.588	0.542	0.526
FM	0.513	0.523	0.553	0.448	0.435	0.392	0.318	0.242
IPR	0.651	0.606	0.647	0.589	0.593	0.532	0.486	0.390
OPR	0.686	0.636	0.638	0.501	0.539	0.480	0.403	0.456
OV	0.493	0.539	0.672	0.456	0.426	0.354	0.254	0.282
BC	0.705	0.698	0.743	0.657	0.677	0.637	0.581	0.535
LR	0.747	0.757	0.771	0.736	0.689	0.597	0.567	0.503

challenges. For the challenges of IV, SV, DEF and OPR, SiamFC performs better than other trackers because it

finds an object similar to the target in the first frame every time it tracks, so as to reduce the impact of target

deformation and illumination variation. BACF outperforms other trackers for the challenges of OCC, MB, FM, OV, BC and LR, because BACF considers large enough background around the target, so that when the target disappears or is occluded temporarily, or the frame is blurred, BACF can still track the target when it appears again. Therefore, it is more appropriate to use SiamFC to track cows that are active in places with illumination variation, such as grassland. For *running* and *jumping* actions that may cause motion blur, it's better to use BACF to track. For gentle movements, such as *standing*, *walking* and *lying down*, BACF, SiamFC, CREST are all suitable.

In order to verify the effectiveness of our proposed cows actions recognition algorithm, we use overall accuracy and single type of actions accuracy to evaluate our actions recognition algorithm. The overall accuracy refers to the ratio of the number of videos where cows actions are correctly recognized to the total number of videos, and single type of actions accuracy refers to the ratio of the number of videos where one type of cows actions (such as *standing*) is correctly recognized to the total number of videos for that type of actions. We first use the artificially annotated ground truth data representing the cows' motion trajectory in 20% of the videos of each type of actions on the cows dataset to adjust the coefficients and thresholds in the cows actions recognition algorithm, so that the actions recognition algorithm can achieve the highest accuracy on the ground truth data, and the remaining 80% of the videos

are used to verify the actions recognition algorithm. Next, we use the selected eight trackers to obtain the tracking trajectories of cows in videos respectively, and then use the recognition algorithm to determine the actions corresponding to the input tracking trajectories, that is, to recognize the actions of cows in videos.

Table 4 shows the overall accuracy and five type of actions accuracy of our proposed cows actions recognition algorithm under eight trackers. In terms of overall accuracy, the accuracy of the BACF-based actions recognition algorithm reaches 0.91, achieving the best results, while the CREST-based and SiamFC-based actions recognition algorithms rank second and third respectively. The result shows that, under the premise that the tracker can accurately track the target, the cows actions recognition algorithm we proposed can correctly recognize the cows' actions in the videos in most cases. In terms of single type of actions accuracy, the trackers can track gentle actions stably, such as *standing* and *lying*, so that the recognition accuracy will be higher. However, when tracking such violent actions as *running*, there are often challenges such as motion blur, deformation, which will lead to tracking instability or failure, thus reducing the recognition accuracy.

In addition, we demonstrate the efficiency of the eight selected trackers by Table 5. Usually, for the same tracker, the process time will become longer when the target region becomes bigger. The numbers in Table 5 are the frames per second (FPS) for different trackers referring to different videos. The bigger number means

Table 4 The accuracy (acc) of the proposed cows actions recognition algorithm under eight trackers

Trackers	Acc					
	Overall acc	Standing acc	Walking acc	Running acc	Jumping acc	Lying acc
SiamFC	0.87	0.94	0.88	0.50	0.80	1.00
CREST	0.89	0.94	0.82	0.83	0.80	1.00
BACF	0.91	0.94	0.82	0.83	1.00	1.00
KCF	0.81	0.94	0.65	0.50	1.00	1.00
RPT	0.85	1.00	0.76	0.50	0.80	1.00
CSK	0.74	0.94	0.59	0.33	0.60	1.00
DFT	0.68	1.00	0.53	0.00	0.40	1.00
CT	0.70	0.83	0.59	0.50	0.40	1.00

Table 5 The FPS of different trackers for five videos

Videos	Trackers							
	SiamFC	CREST	BACF	CSK	CT	DFT	KCF	RPT
run1	86.5	2.2	6.5	100.5	25.6	10.1	90.4	5.0
lie2	97.1	3.8	13.8	229.5	63.6	3.3	80.5	3.4
jump2	112.0	4.2	26.8	519.6	77.3	15.2	230.6	6.0
stand11	88.7	2.5	14.8	219.4	40.5	4.9	104.9	2.9
walk7	85.3	2.0	16.1	81.0	66.7	6.8	88.2	1.0

faster process. SiamFC, CSK and KCF for tracking will have higher real-time performance.

CONCLUSIONS

In this work, we propose a new method of tracking guided actions recognition for cows. This is an important part of the intelligent monitoring in precision cows rearing. We establish a dataset on cows actions, which enriches the dataset on cows actions. Moreover, we propose a set of theoretical methods to analyse the actions of the tracked cows from the data tracked by the trackers. At the same time, we summarize which tracker should be used for videos with different challenges. SiamFC is more reliable for tracking objects whose appearance changes due to the challenges like deformation (DEF), scale variation (SV), out-of-plane rotation (OPR). For the target with occlusion (OCC), motion blur (MB), fast motion (FM), BACF is more suitable for tracking. In a word, for the videos monitored precision cows rearing, our proposed method automatically and accurately recognizes the actions of cows which provides important information for experts and farmers to improve the quality and production of cows. At present, we are dealing with short videos captured by fixed cameras, and a cow to be analyzed performs only one action in a video. However, in the actual production scene, the camera may be shaken, which makes the video quality decline, resulting in tracking deviation. It is necessary to use video pre-processing method to deblur or denoise the video to improve the video quality, and then recognize the cow's actions. In the future, we will propose more actions about cows such as *drinking*, *eating*, *following*, *sleeping*, which are very important for cows rearing. However, these actions need some assisted information such as the recognition of trough and water tank.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (No. 61772209), Science and Technology Planning Project of Guangdong Province (Nos. 2019A050510034 and 2019B020219001), the Production Project of Ministry Education China (No. 201901240030), the College Students Innovations Special Project of China (No. 202010564026), and Guangzhou Key Laboratory of Intelligent Agriculture (No. 201902010081).

COMPLIANCE WITH ETHICS GUIDELINES

The authors Yun Liang and Xiaoming Chen declare that they have no conflict of interest or financial conflicts to disclose. All procedures performed in studies involving animals were in accordance with the ethical standards of the institution or practice at which the studies were conducted, and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

OPEN ACCESS

This article is licensed by the CC By under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

REFERENCES

1. Qiu, T., Chen, N., Li, K., Atiquzzaman, M. and Zhao, W. (2018) How can heterogeneous internet of things build our future: a survey. *IEEE Comm. Surv. and Tutor.*, 20, 2011–2027
2. Fregonesi, J. and Leaver, J. (2001) Behaviour, performance and health indicators of welfare for dairy cows housed in strawyard or cubicle systems. *Livest. Prod. Sci.*, 68, 205–216
3. Haley, D. B., de Passillé, A. M. and Rushen, J. (2001) Assessing cow comfort: effects of two floor types and two tie stall designs on the behaviour of lactating dairy cows. *Appl. Anim. Behav. Sci.*, 71, 105–117
4. Krohn, C. (1994) Behaviour of dairy cows kept in extensive (loose housing/pasture) or intensive (tie stall) environments. iii. grooming, exploration and abnormal behaviour. *Appl. Anim. Behav. Sci.*, 42, 73–86
5. Green, L. E., Hedges, V. J., Schukken, Y. H., Blowey, R. W. and Packington, A. J. (2002) The impact of clinical lameness on the milk yield of dairy cows. *J. Dairy Sci.*, 85, 2250–2256
6. Mattachini, G., Riva, E., Bisaglia, C., Pompe, J. C. and Provolo, G. (2013) Methodology for quantifying the behavioral activity of dairy cows in freestall barns. *J. Anim. Sci.*, 91, 4899–4907
7. Thorup, V. M., Munksgaard, L., Robert, P. E., Erhard, H. W., Thomsen, P. T. and Friggens, N. C. (2015) Lameness detection via leg-mounted accelerometers on dairy cows on four commercial farms. *Animal*, 9, 1704–1712
8. López-Gatius, F., Santolaria, P., Mundet, I. and Yániz, J. L. (2005) Walking activity at estrus and subsequent fertility in dairy cows. *Theriogenology*, 63, 1419–1429
9. Kiddy, C. A. (1977) Variation in physical activity as an indication of estrus in dairy cows. *J. Dairy Sci.*, 60, 235–243
10. Hou, R., Chen, C. and Shah, M. (2017) Tube convolutional neural network (T-CNN) for action detection in videos. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 5823–5832
11. Kuehne, H., Jhuang, H., Garrote, E., Poggio, T. and Serre, T. (2011) Hmdb: a large video database for human motion recognition. In: *Proceedings of the IEEE international conference on computer vision (ICCV)*, pp. 2556–2563
12. Soomro, K., Zamir, A. R. and Shah, M. (2012) Ucf101: a dataset

- of 101 human actions classes from videos in the wild. arXiv, 1212.0402
13. Liu, M., Meng, F., Chen, C. and Wu, S. (2019) Joint dynamic pose image and space time reversal for human action recognition from videos. In: Proceedings of the 33rd AAAI Conference on Artificial Intelligence, pp. 8762–8769
 14. Kristan, M., Pflugfelder, R. and Matas, J. (2015) The visual object tracking VOT2015 challenge results. In: Proceedings of the IEEE International Conference on Computer Vision Workshop (ICCVW), pp. 564–586
 15. Saif, S., Tehseen, S. and Kausar, S. (2018) A survey of the techniques for the identification and classification of human actions from visual data. *Sensors (Basel)*, 18, 3979
 16. Wu, Y., Lim, J. and Yang, M. H. (2015) Object tracking benchmark. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37, 1834–1848
 17. Henriques, J. F., Caseiro, R., Martins, P. and Batista, J. (2012) Exploiting the circulant structure of tracking-by-detection with kernels. In: Proceedings of the 12th European conference on Computer Vision (ECCV), pp. 702–715
 18. Hart, B. L. (1988) Biological basis of the behavior of sick animals. *Neurosci. Biobehav. Rev.*, 12, 123–137
 19. Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C. and Clune, J. (2018) Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. In: Proceedings of the National Academy of Science of the United States of America, pp. E5716–E5725
 20. Sharma, S. U. and Shah, D. J. (2016) A practical animal detection and collision avoidance system using computer vision technique. *IEEE Access*, 5, 347–358
 21. Andriluka, M. and Iqbal, U. Insafutdinov, Eldar., Pishchulin, L., Milan, A., Gall, J., Schiele, B. (2018) PoseTrack: A Benchmark for Human Pose Estimation and Tracking. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5167–5176
 22. Andriluka, M., Pishchulin, L., Gehler, P. and Schiele, B. (2014) 2D human pose estimation: New benchmark and state of the art analysis. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3686–3693
 23. Kalogeiton, V., Weinzaepfel, P., Ferrari, V. and Schmid, C. (2017) Joint learning of object and action detectors. In: The IEEE International Conference on Computer Vision (ICCV), pp. 2001–2010
 24. Gilani, S. O., Subramanian, R., Yan, Y., Melcher, D., Sebe, N. and Winkler, S. (2015) PET: An eye-tracking dataset for animal-centric Pascal object classes. In: Proceeding of the IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6
 25. Jeong, S., Kim, G. and Lee, S. (2017) Effective visual tracking using multi-block and scale space based on kernelized correlation filters. *Sensors (Basel)*, 17, 433
 26. Henriques, J. F., Caseiro, R., Martins, P. and Batista, J. (2015) High-speed tracking with kernelized correlation filters. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37, 583–596
 27. Wang, H., Liu, P., Du, Y., Liu, X. (2019) Online convolution network tracking via spatio-temporal context. *Multimed. Tools Appl.*, 78, 257–270
 28. Liu, N., Hong, H., Tao, F. (2019) Robust object tracking via constrained online dictionary learning. *Multimed. Tools Appl.*, 78, 3689–3703
 29. Gao, J., Wang, Q., Xing, J., Ling, H., Hu, W., Maybank, S.J. (2018) Tracking-by-fusion via Gaussian process regression extended to transfer learning. *IEEE Trans. Pattern Anal. Mach. Intell.*, 42, 939–955
 30. Li, F., Tian, C., Zuo, W., Zhang, L. and Yang, M. H. (2018) Learning spatial-temporal regularized correlation filters for visual tracking. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4904–4913
 31. Zhang, K., Zhang, L. and Yang, M. H. (2012) Real-time compressive tracking. In: Proceedings of the 12th European conference on Computer Vision (ECCV), pp. 864–877
 32. Sevilla-Lara, L. and Learned-Miller, E. (2012) Distribution fields for tracking. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1910–1917
 33. Henriques, J.F., Caseiro, R., Martins, P., Batista, J. (2015) High-speed tracking with kernelized correlation filters. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37, 583–596
 34. Li, Y., Zhu, J. and Hoi, S. C. H. (2015) Reliable Patch Trackers: Robust visual tracking by exploiting reliable patches. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 353–361
 35. Galoogahi, H. K., Fagg, A. and Lucey, S. (2017) Learning backgroundaware correlation filters for visual tracking. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 1144–1152
 36. Bertinetto, L., Valmadre, J., Henriques, J., Vedaldi, A. and Torr, P. (2016) Fully-convolutional siamese networks for object tracking. In: European Conference on Computer Vision, pp. 850–865
 37. Song, Y., Chao, M., Gong, L., Zhang, J. and Yang, M. (2017) Crest: convolutional residual learning for visual tracking. In: IEEE International Conference on Computer Vision (ICCV), pp. 2574–2583
 38. Zhang, Y., Cheng, L., Wu, J., Cai, J., Do, M. N. and Lu, J. (2016) Action recognition in still images with minimum annotation efforts. *IEEE Trans. Image Process.*, 25, 5479–5490
 39. Poppe, R. (2010) A survey on vision-based human action recognition. *Image Vis. Comput.*, 28, 976–990
 40. Dollar, P., Rabaud, V. and Cottrell, G. (2005) Behavior recognition via sparse spatio-temporal features. In: Proceedings of the IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, pp. 65–72
 41. Yin, L., Wei, X. and Sun, Y. (2006) A 3D facial expression database for facial behavior research. In: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, pp. 211–216

42. Tanfous, A. B., Drira, H. and Amor, B. B. (2018) Coding kendall's shape trajectories for 3D action recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2840–2849
43. Tang, Y., Tian, Y., Lu, J., Li, P. and Zhou, J. (2018) Deep progressive reinforcement learning for skeleton-based action recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5323–5332
44. Gallego, G., Rebecq, H. and Scaramuzza, D. (2018) A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3867–3876
45. Guo, G. and Lai, A. (2014) A survey on still image based human action recognition. *Pattern Recognit.*, 47, 3343–3361
46. Zhang, T., Zhang, Y., Cai, J. and Kot, A. C. (2016) Efficient object feature selection for action recognition. In: Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2707–2711
47. Feichtenhofer, C., Pinz, A. and Wildes, R. P. (2017) Spatiotemporal multiplier networks for video action recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7445–7454
48. Girdhar, R., Gkioxari, G., Torresani, L., Paluri, M. and Tran, D. (2018) Detect-and-track: Efficient pose estimation in videos. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 350–359
49. Lu, X., Yao, H., Zhao, S., Sun, X., Zhang, S. (2019) Action recognition with multi-scale trajectory-pooled 3D convolutional descriptors. *Multimed. Tools Appl.*, 78, 507–523