A Benchmark for Action Recognition of Large Animals

Yun Liang, Fuyou Xue, Xiaoming Chen, Zexin Wu, Xiangji Chen*
College of Mathematics and Informatics,
South China Agricultural University,
Guangzhou, China
e-mail: sdliangyun@163.com, checkie@vip.qq.com

***Abstract*: The action recognition of large animals plays an important role in the intelligent and modern farming. People often use the actions as the key factors to achieve scientific feeding and improve the animal welfare, and then the quality and productivity of animals are greatly promoted. However, most present action recognition methods focus on the actions of human (as pedestrian, athletes) or man-made objects (as cars, bikes). This paper proposes a benchmark to recognize and evaluate the actions of a kind of large animals namely the cows. First, we construct a dataset including 60 videos to describe the popular actions existing in the daily life of cows, and manually denote the target regions of cows on every frame in the dataset. Second, six famous trackers are evaluated on this dataset to compute the trajectory of cows which is the basis of actions recognition. Third, we define the method to recognize the actions of cows via the trajectories and validate the proposed method on our dataset. Many experiments demonstrate that our method of action recognition performs favorable in detecting the actions of cows, and the proposed dataset basically satisfies the action evaluations for farmers. The work in this paper provides an automatic and scientific method for famers to design a scheme to promote the quality and productivity of cows.**

Keywords; action recognition; intelligent farming; visual tracking; large animal; dataset

I. Introduction

Visual tracking and action recognition play an important role in computer vision. For their abilities in predicting the semantic information of moving objects, they are widely used in the intelligent robot, virtual reality, video surveillance, automatic drive and so on [1]. However, most of the present trackers, action identifiers and the related datasets are proposed for dealing with the moving persons (such as athletes), motor vehicles (such as the cars or motorbikes), some objects (such as a box or a doll) and so on. At the same time, there are very few researches about the action recognition of large animals. But in fact, the action recognitions are very important factors for scientific and modern feeding. They can help farms greatly promote the quality and productivity of animals.

In the traditional model, the health condition of animals is judged by the feeders which is not only poor efficiency but also high deviation. Meanwhile, it is hard to realize the real-time and accurate monitor of the health condition of the animals. Last but not least, while all the companies are striving for high profit with low cost, checking the health condition of the animals by hand is high cost. Therefore, it is extremely necessary for us to set up a benchmark for animal tracking and action recognition which can check the health condition of the animals automatically and accurately. The benchmark can track the animals and recognize their actions automatically, and use these data to check the health of the animals, which is much more efficient, accurate and convenient with low cost than the traditional methods. With the action information, feeders can design scientific schemes to feed animals which finally leads to better product quality as well as product quantity (such as milk for cows and meat for pigs) undoubtedly. On the other hand, in the research areas, animal action is an important analytical indicator for experts [2]. The experts use these data to study the animals' health and mood, and then analyze the relationship between these data and the production capacity. Therefore, we propose a benchmark for action recognition of large animals in this paper. Our main contributions are as follows:

1) We set up a dataset for large animals (cows), which include 60 videos with eleven challenges (including occlusion and deformation of the target animal). We define the GT (ground truth) for all the videos which describes the target region on each frame of every video.

2) We define the algorithm to recognize the five popular actions of cows which can guide us how to judge what the animals are doing briefly. The five actions include ***Stand, Lie, Run, Jump and Walk***.

3) We select and evaluate six trackers for our benchmark to compute the state changes of the animals. The state changes finally are the key factors in action recognition. The effectiveness and efficiency of the trackers are evaluated in this work by many experiments.

II. Related Work

This paper proposes a new dataset of large animals (mainly for cows), defines an algorithm to recognize the actions of the animals, and analyzes the effectiveness of the present trackers on the new dataset. Therefore, we describe the related work from the following three aspects.

A. Dataset for Large Animals

This paper aims at recognizing and analyzing the actions of large animals. Here, we take cows as the example of large animals. To deal with the actions of cows, a 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 the algorithms to recognize and analyze

978-1-5386-9497-8/18/$31.00 ©2018 IEEE
DOI 10.1109/ICDH.2018.00020

the actions of cows based on this dataset. However, there is no such public dataset for research people.

Recently, there are some public datasets to evaluate the trackers or actions, such as the VOT dataset [3], TB-100 [4]. These existing tracking datasets are mostly concerning human (such as pedestrians, athletes) and man-made objects (such as cars, bikes, dolls) and so on. Although some videos in VOT and TB-100 take some animals as the targets, such videos are too few and not specially defined for the research about the actions of large animals.

Therefore, we establish a dataset for cows by learning the existing datasets for actions or trackers. All the videos in our dataset are about cows. They are either shot at the cow farms or downloaded from the Internet. Similar to the existing datasets, we choose these videos because they have some representative actions of cows and cover the challenges for tracking or action recognizing (such as occlusion, motion blurred, deformation). In the meanwhile, the number of these videos is 60 totally, which are enough for researchers to do the basic research of the action analysis of cows.

B. Tracking Algorithms

We use some existing tracking algorithms as the basic trackers to test our dataset and compute the moving trajectories of animals. Using the tracked trajectories, we design methods to recognize the actions of cows. In the Visual Tracker Benchmark [4], 29 trackers are evaluated. In this paper, we only select six representative trackers to test our dataset, because these trackers have published their codes or executable files and implemented very fast. These six trackers are CSK [5], CT [6], DFT [7], KCF [8], RPT [9], BACF [10].

We use each of the above trackers with its default parameters to test every video in our dataset. At the beginning, we manually specify the target region on the first frame. Then in the rest frames, the states of target (the width, the height, the x-coordinate and the y-coordinate of upper left point of target region) are computed by one tracker. The recorded states finally are used to achieve the action analysis of the related animal.

C. Action Recognition

Recently, many methods about action recognition have been proposed. Most of them are defined for human but not for animals. Some most related papers about action recognition are as follows. Dollar et al. [11] define the sparse spatio-temporal features to construct the method to recognize the behavior of human and rodents. Yin et al. [12] propose a 3D facial expression database for facial behavior research, which focuses on 3D human facial expressions. Ben et al. [13] propose an intrinsic sparse coding and dictionary learning formula for efficient coding of motion-recognized 3D skeleton sequences. Tang et al. [14] define a deep progressive reinforcement learning method for action recognition in skeleton-based videos. Gallego et al. [15] suggest a unified framework to solve several computer vision problems of event cameras: motion, depth, and optical flow estimation, which can be used to find point trajectories on the image plane that are optimally aligned with event data (people, animals).

Similar to the above work about human action recognition, we do researches on the action recognition of cows in this paper. We define five kinds of actions to describe the behaviors existing in daily life of cows. These actions include ***stand, walk, jump, lie and run.*** We use the state changes about a target (the specified cow), occurred in the tracked trajectory of a video to justify and analyze action. The state is described by four parameters namely the width (w), height (h), x-coordinate (cx) and y-coordinate (cy) of center point of target region.

III. Our Benchmark and Action 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 in the dataset. Finally, we evaluate how to get the states of cows by the existing trackers.

A. Our Dataset

Our dataset is as shown in Figure 1, with which we recognize the representative and popular actions of cows. The data in this dataset not only provide some important references for animal breeding experts [2], but also represent the common challenges in the action recognition of cows in their daily activities. Here, we use the same challenges defined in the benchmark for trackers [4], because both the challenges from the two benchmarks are introduced by the same problems namely the changes of the target and background. Totally, 11 challenges are used in this paper. The details about the challenges are given at section 3.3.

Figure 1 demonstrates our dataset of cows including 60 videos, which are collected by shooting in different environments with various directions and angles or downloading on the Internet. The yellow words are the name for each video. The red rectangle describes the target used to do action recognition. We manually specify the target rectangle on each frame for every video, and record four parameters (include the width, height, the x-coordinate and y-coordinate of the rectangle center) as its state description. This state record is the ideal value to analyze the actions. But when we automatic recognizing the actions, we need the trackers to automatically compute the state of target rectangle on each frame except the first frame. At that time, the specified rectangles can be used as the ground truth to compare the efficiency of each tracker implemented on the videos of cows.

The given dataset covers five popular actions of cows proposed in section 2. As shown in the names of videos, the actions about walk, run, jump stand and lie are all described. For example, we construct seven videos (from Jump1 to Jump7 in Figure 1) to describe the different kinds of jumping. We take the action of cows shown in a video to classify our dataset. Therefore, the proposed 60 videos are classified into five categories, of which 7 videos are for jumping, 7 videos are for lying, 8 videos are for running, 18 videos are for standing, 20 videos are for walking.

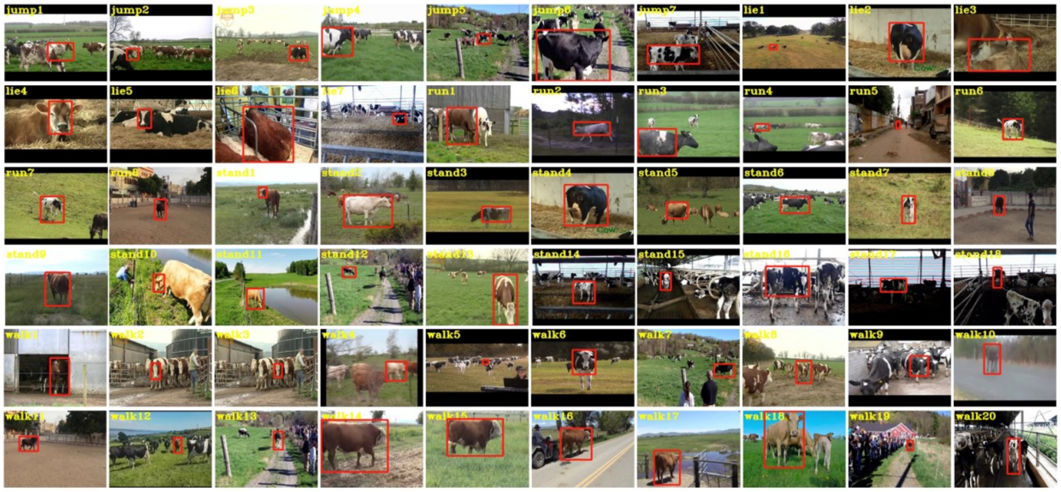


Figure 1. Our Dataset (60 videos). Each image is the first frame of a video. The red rectangle is the target of action recognition.

B. *Action Analysis*

In this paper, we design the algorithm to recognize the actions of big animals (cows) based on the state changes of target object. We use four parameters including cx (x-coordinate of target center), cy (y-coordinate of target center), w (the width of target), h (the height of target) to indicate the center position and size of target. This paper defines the actions of large animals (cows) based on the curves of cx, cy, w, and h.

To describe the range of variation of each parameter about target state, we define:

$$\begin{cases} \Delta_{cx} = cx_{max} - cx_{min} \\ \Delta_{cy} = cy_{max} - cy_{min} \\ \Delta_{w} \ = \ w_{max} - w_{min} \\ \Delta_{h} \ = \ h_{max} - h_{min} \end{cases} \qquad (1)$$

$\Delta_{cx}, \Delta_{cy}, \Delta_w, \Delta_h$ are the variation ranges of parameters $\mathrm{cx, cy, w, h}$; $cx_{max}, cy_{max}, w_{max}, h_{max}$ are the maximum value while $cx_{min}, cy_{min}, w_{min}, h_{min}$ are the minimum ones.

Then we define the threshold θ that distinguishes the range of parameter variation, because the change of parameter of an action is affected by the distance between the target and the camera device. When a cow jumps, for example, the closer it is to the camera, the greater the range of motion it shows in the camera, and the smaller it is the farther away it is. To rule out this effect, the definition of the threshold θ should take into account the size of the target:

$$\begin{cases} \theta_{cx} = \vartheta_{cx} * \ \overline{w} \\ \theta_{cy} = \vartheta_{cy} * \ \overline{h} \\ \theta_{w} \ = \vartheta_{w} \ * \ \overline{w} \\ \theta_{h} \ = \vartheta_{h} \ \ * \ \overline{h} \end{cases} \qquad (2)$$

$\theta_{cx}, \theta_{cy}, \theta_w, \theta_h$ are the thresholds for parameters ($\mathrm{cx, cy, w, h}$), $\vartheta_{cx}, \vartheta_{cy}, \vartheta_w, \vartheta_h$ are the proportional coefficients, and $\overline{w}\ and\ \overline{h}$ are the mean values of w and h in a video respectively.

In addition, we define variables v_{cx}, v_{cy}, v_w, v_h that describe the change rate of parameters:

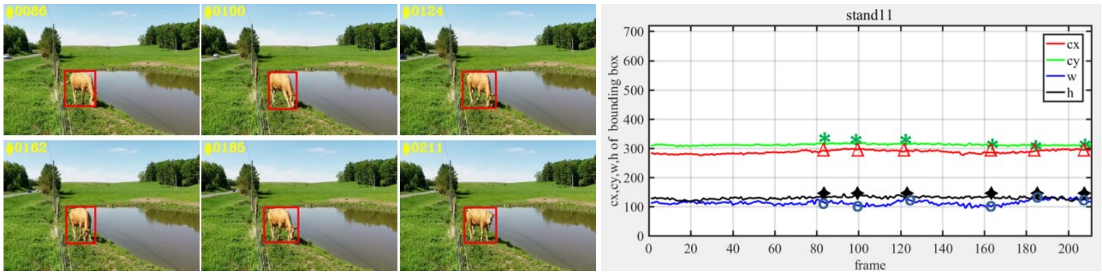
$$\begin{cases} v_{cx} = \left|\Delta_{cx}/(t_{\max\_cx} - t_{\min\_cx})\right| + W/\overline{w} \ + H/\overline{h} \\ v_{cy} = \left|\Delta_{cy}/(t_{\max\_cy} - t_{\min\_cy})\right| + W/\overline{w} + H/\overline{h} \\ v_{w} \ = \left|\Delta_{w} \ /(t_{\max\_w} \ - t_{\min\_w}) \ \right| + W/\overline{w} + H/\overline{h} \\ v_{h} \ = \left|\Delta_{h} \ \ /(t_{\max\_h} \ \ - t_{\min\_h}) \ \right| + W/\overline{w} + H/\overline{h} \end{cases} \qquad (3)$$

v_{cx}, v_{cy}, v_w, v_h reflect the change rate of $\mathrm{cx, cy, w, h}$ from minimum to maximum. $\Delta_{cx}, \Delta_{cy}, \Delta_w, \Delta_h$ indicate the range of variation of the parameters, the $t_{\max\_cx}, t_{\max\_cy}, t_{\max\_w}, t_{\max\_h}$ represent the time that the maximum value of $\mathrm{cx, cy, w, h}$ occur. The $t_{\min\_cx}, t_{\min\_cy}, t_{\min\_w}, t_{\min\_h}$ represent the time that the minimum value of $\mathrm{cx, cy, w, h}$ occur. In this paper, we use the frame number to describe the time of each frame. Since the true rate is independent of the distance from the target to the camera device, we add the $W/\overline{w}$ item and $H/\overline{h}$ item (W, H is the width and height of the video frame) to adjust the rate to be more acceptable.

1) *The action of Stand*

A standing cow is in a static state and its body remains in a small region in an image. For a period of time, the relative displacement of the four parameters about its state is small and not obvious. Figure 2 describes an example of the action of a standing cow. From Figure 2(a), we know the state of the cow nearly remain the same from #0086 to #0211. This is also verified by Figure 2(b). We define the action of Stand for a cow as:

$$\begin{cases} \Delta_{cx} < \ \theta_{cx} \ , & \vartheta_{cx} = 0.15 \\ \Delta_{cy} \ < \ \theta_{cy} \ , & \vartheta_{cy} = 0.15 \\ \Delta_{w} \ \ < \ \theta_{w} \ , & \vartheta_{w} \ = 0.15 \\ \Delta_{h} \ \ \ < \ \theta_{h} \ , & \vartheta_{h} \ \ = 0.15 \end{cases} \qquad (4)$$

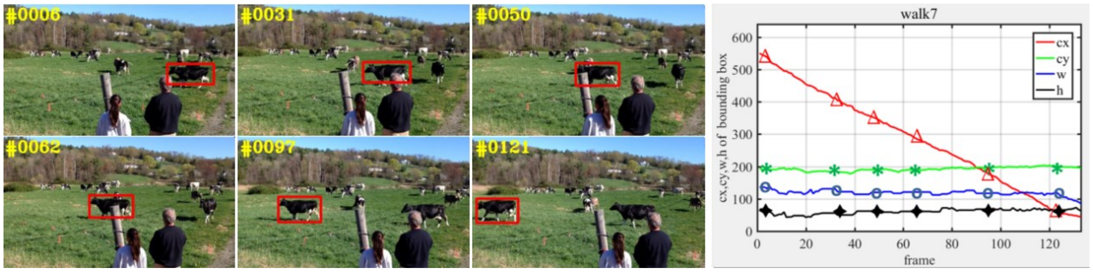


(a) Six ideal target regions of video (stand11)

(b) The state changes of video (stand11)

Figure. 2 The action of ***Stand*** about video (stand11).

In this paper, we describe the state changes of a video as the curves shown in Figure 2. The bounding boxes namely the red rectangles in Figure 2(a) describe the ideal target regions of a cow. The red, green, blue and black curves in Figure 2(b) separately describe the state change of target cow frame by frame. the ✱,Δ,✦,○ represent the positions on these curves of the frames in Figure 2(a). We use the same method to design Figure 3, Figure 4, Figure 5 and Figure 6. The $\vartheta_{cx}, \vartheta_{cy}, \vartheta_{w}, \vartheta_{h}$ are set based on a large number of experiments for the Equation 4, Equation 5, Equation 6, Equation 7 and Equation 8.



(a) Six ideal target regions of video (walk7)

(b) The state changes of video (walk7)

Figure 3. The action of ***Walk*** about video (walk7).

2) *The action of Walk*

The walking action of a cow can be classified into horizontal walking (relative displacement of target changes greatly, size change varies a little), and vertical direction walking (target position changes a little, size changes greatly). Figure 3 describes the action of horizontal walking of the cow denoted by the red rectangle. As shown in Figure 3(b), the width and height of the bounding boxes of target are not significantly changed. From #0006 to #0121, the w only changes by 17 pixels and the h only changes 28 pixels. However, the curve of cy about center position remains a straight line without any changes while the curve of cx sharply reduces as the red curve describes. The changes of its states demonstrate that the cow moves in horizontal direction at a constant speed. Following this, we define the definition of cow horizontally walking by:

$$\begin{cases} \theta_{cx} < \Delta_{cx}\,, & \vartheta_{cx} = 0.25 \\ \Delta_{cy} < \theta_{cy}\,, & \vartheta_{cy} = 0.25 \\ \Delta_{w} < \theta_{w}\,, & \vartheta_{w} = 0.4 \\ \Delta_{h} < \theta_{h}\,, & \vartheta_{h} = 0.4 \\ \alpha_1 < v_{cx} < \alpha_2\,, & \alpha_1 = 10\,, \alpha_2 = 20 \end{cases} \tag{5}$$

According to Equation 5, we calculate the parameters for video walk7 demonstrated in Figure 3. $\theta_{cx} = 30 < \Delta_{cx} = 471$; $\Delta_{cy} = 10 < \theta_{cy} = 14$; $\Delta_{w} = 50 < \theta_{w} = 61$; $\Delta_{h} = 23 < \theta_{h} = 29$; $v_{cx} = 14.47$, we can conclude that the cow is walking.

3) *The action of Jump*

Similar to the walking action, there are two kinds of jumping action namely jumping in the horizontal direction and jumping in the vertical direction. In this paper, we take the vertical jumping as an example to design the method to recognize jumping action of cows. In the vertical jumping, the cy of a cow fluctuates greatly in a short period, but the value changes of cx remain only in a kind of trend. Generally, the jumping may occur several times in a short period, and multiple local maximums and local minimums appear continuously on the curves describing the state changes about cy. When a cow makes jump to the highest point, a local maximum value appears in the figure. When a cow arises the lowest point in jumping, a local minimum value appears. Figure 4 describes a typical jumping action of a cow. It is known from the figure that the cx value rises, indicating that the target moves to the right. The cy curve has multiple minimum values, which indicates that the target is jumping. According to the above analysis, we define the jumping action by:

$$\begin{cases} \theta_{cx} < \Delta_{cx}\,, & \vartheta_{cx} = 0.15 \\ \theta_{cy}^{1} < \Delta_{cy} < \theta_{cy}^{2}\,, & \vartheta_{cy}^{1} = 0.15\,, \vartheta_{cy}^{2} = 0.5 \\ \exists \mathrm{t}\,(cy_t + \theta_{cy} < cy_{t-k}) \wedge \left(cy_t + \theta_{cy} < cy_{t+k}\right) & \\ \vartheta_{cy} = 0.3\,, k = 5 & \end{cases} \tag{6}$$

The change range of cy namely the Δ_{cy} is controlled between 0.15 and 0.5 of the target's height, and there is a minimum point. We use the third item of Equation 6 to justify a minimum point. It is verified by a large number of experiments that $\vartheta_{cy} = 0.3, k = 5$, and the range of cx namely the Δ_{cx} is greater than 0.15 of the target width.

4) The action of Lie

The action of lie is finished by two steps. First, the cy of a cow undergoes a great change while the cx changed in a trend. In this course, the cow changes its state from standing to lying. Second, the state of the cow remains unchanged for a long time. Figure 5 describes a classic action of cow lying. According to Figure 5(b), from #60 to #90, the cy value becomes larger, indicating that the target has a sudden drop in motion. The cx has a related change because the cow swings his body when lies down. Based on the above analysis, we define this kind of lying action by:

$$\begin{cases} \Delta_{cy}/(t_{\max\_cy} - t_{\min\_cy}) > 0 \\ \Delta_w/(t_{\max\_w} - t_{\min\_w}) < 0 \\ \Delta_h/\left(t_{\max\_h} - t_{\min\_h}\right) < 0 \\ \beta_1 < v_{cy} < \beta_2 \ , \quad \beta_1 = 10\,, \beta_2 = 20 \end{cases} \tag{7}$$

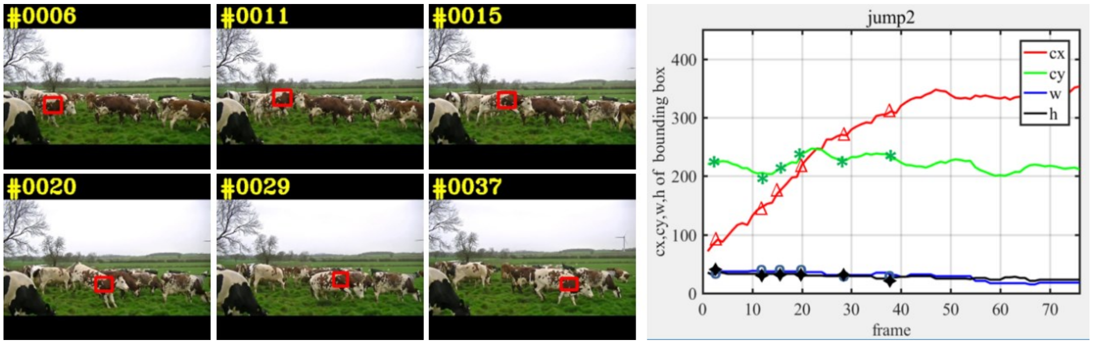
The change rate of cy namely the v_{cy} is controlled between 10 and 20. The slope of the line formed by the maximum and minimum values of cy is greater than 0, indicating that the target moves downward and the size is reduced. For the cow in video lie2 shown by Figure 5, we calculate the $v_{cy} = 14.16$, and the cy value increases, w and h separately drop by 42 pixels and 58 pixels. Therefore, we conclude that the cow in Figure 5 is in a lying state.

5) The action of Run

The cows in the running action have a fast change of state of the cow. There are still two kinds of running action, namely along the horizontal direction and the vertical direction. Usually, for the horizontal running, the cx changes greatly in a short period. For the vertical running, the size of target cow is scaled greatly in a short period. Here, we take the horizontal running as an example to define the action recognition of cow running. Figure 6 describes a classic action of cow running in horizontal direction. As shown in Figure 6(b), from #70 to #123, the cx value is significantly increased. The slope of cx line is large enough and the cy line just has a fluctuating change. The variations of cx and cy indicate that the target moves rapidly in the x-axis direction. According to the above analysis, we define the running in horizontal direction by:

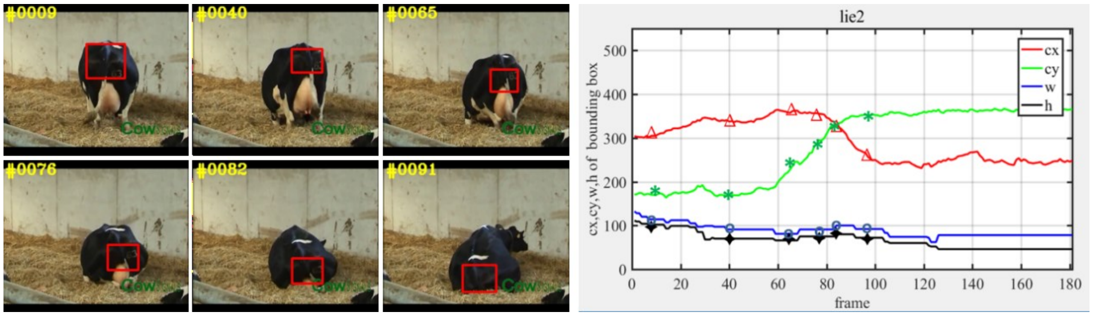
$$\begin{cases} \gamma < v_{cx}, & \gamma = 20 \\ \theta_w < \Delta_w\,, & \vartheta_w = 0.15 \\ \theta_h < \Delta_h\,, & \vartheta_h = 0.15 \\ \theta_{cx} < \Delta_{cx}\,, & \vartheta_{cx} = 0.15 \\ \theta_{cy}^1 < \Delta_{cy} < \theta_{cy}^2\,, & \vartheta_{cy}^1 = 0.15\,,\ \vartheta_{cy}^2 = 0.5 \\ \exists \mathrm{t}(cy_t + \theta_{cy} < cy_{t-k}) \wedge \left(cy_t + \theta_{cy} < cy_{t+k}\right) \\ \vartheta_{cy} = 0.3\ , k = 5 \end{cases} \tag{8}$$

Here, we define that the change range of cx, w are greater than 0.15 of $\bar{w}$, the range of h is greater than 0.15 of $\bar{h}$, the range of cy is controlled between 0.15 and 0.5 of $\bar{h}$, and v_{cx} is greater than 20. For Figure 6, we calculate the cow's $v_{cx} = 22.7$ and the change of cy is very little. Then, we conclude that the cow is running in the horizontal direction.



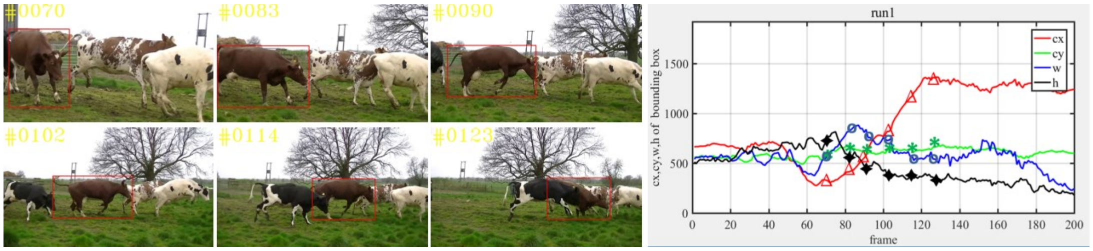
(a) Six ideal target regions of video (jump2) (b) The state changes of video (jump2)

Figure 4. The action of ***Jump*** about video (jump2).



(a) Six ideal target regions of video (lie2) (b) The state changes of video (lie2)

Figure 5. The action of ***Lie*** about video (lie2).

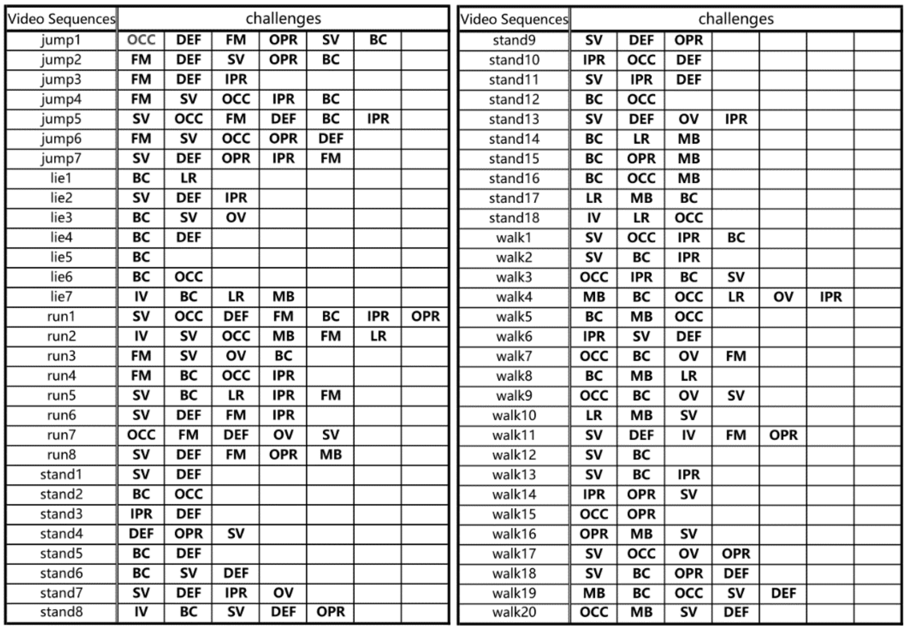


(a) Six ideal target regions of video (run1) (b) The state changes of video (run1)

Figure 6. The action of ***Run*** about video (run1).

C. Challenges of Videos in Our Dataset

Table 1. The challenges of videos in our dataset shown in Figure 1



Video Sequences	challenges						
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		
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	OCC	FM	DEF	OV	SV		
run8	SV	DEF	FM	OPR	MB		
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				
walk1	SV	OCC	IPR	BC			
walk2	SV	BC	IPR				
walk3	OCC	IPR	BC	SV			
walk4	MB	BC	OCC	LR	OV	IPR	
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	OPR	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			

We list the challenges of every video in our dataset as shown in Table 1. We use the same 11 challenges from the benchmark [4] to denote the challenges of our dataset. These challenges include IV (Illumination Variation), SV (Scale Variation), OCC (Occlusion), DEF (Deformation), MB (Motion Blur), FM (Fast Motion), IPR (In-Plane Rotation), OPR (Out-of-Plane Rotation), OV (Out-of-View), BC (Background Clutters), and LR (Low Resolution). Please review more details about these challenges in Tracking benchmark [4].

As shown in Table 1, the challenges corresponding to different actions of cows vary, meanwhile the same actions have similar challenges. For example, when cows are moving, the challenges to track one of them 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, we will encounter such challenges as BC, OCC, and OV, when recognizing and tracking a specific cow in a large group of highly similar cows, such is a problem to be solved in our tracking algorithms. Moreover, cows often walk outdoors so that the variation of illumination will also affect the performance of the tracking algorithms.

IV. Tracking Results of Existing Trackers

Our experiments are implemented based on matlab2016a on a regular PC (64-bit win10 operating system, Intel Core i5-4200H 2.80GHz processor, NVIDIA GeForce GTX 950M graphics card, 4GB RAM).

To verify the reliability, rationality, and diversity of our dataset, we selected six trackers from the visual tracker benchmark [4], including the BACF [10], CSK [5], CT [6], DFT [7], KCF [8], RPT [9], and analyzed the robustness and accuracy of these trackers on the videos of our dataset. We compare the two center coordinates (cx, cy) of the targets tracked by different algorithms with the ground truth of our dataset annotations. Calculating the standard deviation can reflect the degree of dispersion of the tracking data. Four videos related to four actions are used to evaluate the six selected trackers and the results are as follows.

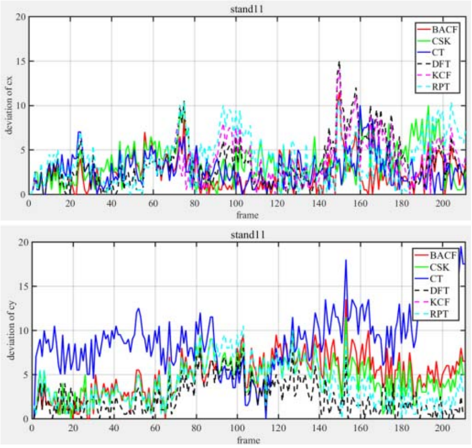


Figure 7. Tracking deviation of video stand11

Figure 7 describes the tracking results of a stand action video (stand11). The deviation of the tracking results of the six trackers is kept within 20, and results of BACF, DFT, RPT are closer to our ground truth.

Figure 8 describes the tracked results of a lying action video(lie2). Compared with stand action, the deviation of the tracking result is increased.

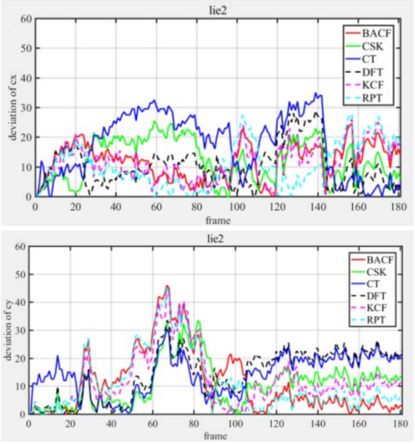


Figure 8. Tracking deviation of video lie2

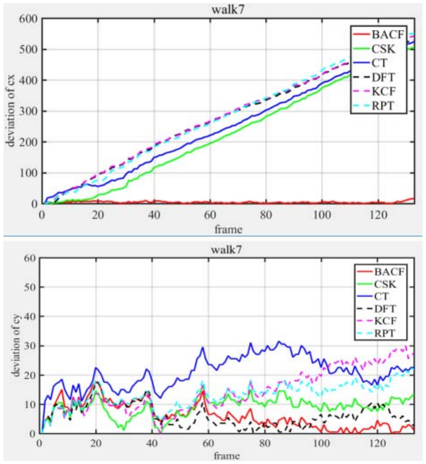


Figure 9. Tracking deviation of video walk7

Figure 9 describes the tracked results of a walk action video(walk7). The deviation value of cx reflects the higher robustness and accuracy of BACF, while other trackers lose their targets.

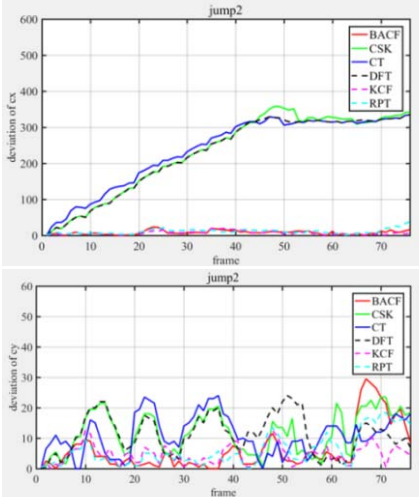


Figure 10. Tracking deviation of video jump2

Table2.The different trackers' FPS referring to different videos

Tracker	BACF	CSK	CT	DFT	KCF	RPT
Run1	6.5	100.5	25.6	10.1	90.4	5.0
Lie2	13.8	229.5	63.6	3.3	80.5	3.4
Jump2	26.8	519.6	77.3	15.2	230.6	6.0
Stand11	14.8	219.4	40.5	4.9	104.9	2.9
Walk7	16.1	81.0	66.7	6.8	38.2	1.0

Figure 10 describes the tracked result of a jump action video (jump2), in which the target can be followed by BACF, KCF, RPT, and the average deviation of BACF is smaller than that of the other two trackers. We conclude that challenges have a great impact on the tracked results especially the challenges introduced by great deformation such as BC, SV, OCC and so on.

We demonstrate the efficiency of the six selected trackers by Table 2. Usually, for the same tracker, the process time will become longer when the target region becomes bigger. The numbers in Table2 are the frames per second (FPS) for different trackers referring to different videos. The bigger number means faster process.

V. Conclusion

This paper proposes a new benchmark for actions recognition of a kind of large animals namely the cows. The benchmark includes 60 videos whose target bounding boxes are specified by us. These videos cover the popular actions of cows and include 11 challenges in action recognition. This benchmark also provides the algorithms to recognize the four actions of cows including: ***Stand, Walk, Jump, Lie*** and ***Run***. At the same time, we select six trackers for this benchmark and evaluate them on our dataset. The evaluation results show that the trackers of learning background information and Correlation Filters (BACF, KCF, RPT) are better ranked. The tracker based on sparse representation (CT) is not ideal in large scale variation and fast motion. In the future work, we will design a more detailed analysis of the actions of large animals and provide the principles that are more conducive to tracking and actions identification of large animals.

Acknowledges

This work is supported by the National Natural Science Fund of China (61772209, 61472335), the Science and Technology Planning Project of Guangdong Province (2016A050502050) and the Innovation and entrepreneurship project (20180564107).

References

[1] G. R. Bradski, "Intel's Computer Vision Library applications in calibration, stereo segmentation, tracking, gesture, face and object recognition", IEEE-CVPR, Hilton Head Island, SC, USA, June 2000, doi:10.1109/CVPR.2000.854964.

[2] B. L. Hart, "Biological basis of the behavior of sick animals", In Neuroscience & Biobehavioral Reviews, vol 12, Issue 2, pp. 123-137, Summer 1988, doi:10.1016/S0149-7634(88)80004-6.

[3] M. Kristan, "The Visual Object Tracking VOT2015 Challenge Results", ICCV, pp. 1-23, Dec. 2015.

[4] Y. Wu, "Object tracking benchmark", IEEE Trans, vol 37, Issue 9, pp. 1834–1848, Sept. 2015, doi: 10.1109/TPAMI.2014.2388226.

[5] J. F. Henriques, "Exploiting the Circulant Structure of Tracking-by-Detection with Kernels", ECCV, pp. 702-715, 2012.

[6] K. Zhang, "Real-time Compressive Tracking", ECCV, pp. 864-877, 2012.

[7] L. S. Lara, "Distribution Fields for Tracking", CVPR, Providence, RI, USA, June 2012, doi: 10.1109/CVPR.2012.6247891.

[8] J. F. Henriques, "High-Speed Tracking with Kernelized Correlation Filters", IEEE TPAMI, vol 37, Issue 3, pp. 583-596, March 2015, doi:10.1109/TPAMI.2014.2345390.

[9] Y. Li, "Reliable Patch Trackers: Robust Visual Tracking by Exploiting Reliable Patches", CVPR, pp. 353-361, June 2015.

[10] H. K. Galoogahi, "Learning Background-Aware Correlation Filters for Visual Tracking", CVPR, pp. 1135-1143, 2017.

[11] P. Dollar, "Behavior recognition via sparse spatio-temporal features", IEEE, Beijing, China, Oct. 2005, doi:10.1109/VSPETS.2005.1570899.

[12] L. Yin, "A 3D facial expression database for facial behavior research", IEEE, Southampton, UK, April 2006, doi:10.1109/FGR.2006.6.

[13] A. B. Tanfous, "Coding Kendall's Shape Trajectories for 3D Action Recognition", IEEE-CVPR, pp. 2840-2849, June 2018.

[14] Y. Tang, "Deep Progressive Reinforcement Learning for Skeleton-based Action Recognition", CVPR, Tsinghua University, China, 2018.

[15] G. Gallego, "A Unifying Contrast Maximization Framework for Event Cameras, with Applications to Motion, Depth, and Optical Flow Estimation", IEEE-CVPR, ETH Zurich, Switzerland, 2018.