Pig target tracking algorithm based on multi-channel color feature fusion

Longqing Sun*, Shuaihua Chen, Ting Liu, Chunhong Liu, Yan Liu
(*College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China*)

Abstract: In the process of tracking the target of the pig, with the change of the size of the tracking target in the video image, the estimated tracking target scale cannot be adaptively updated in real-time, resulting in the low accuracy of the tracking target. In this study, a multi-channel color feature adaptive fusion algorithm was proposed, and the target scale of the pig was updated in real-time by utilizing the contour information of the target pig. Experiments show that the proposed algorithm had a distance precision of 89.7% and an overlap precision of 87.5%, and the average running speed of this algorithm was 50.1 fps. The robustness of the proposed algorithm in tracking target deformation and scale variation were significantly improved, which satisfies the accuracy and real-time requirements of pig target tracking.
Keywords: pig tracking, color feature, correlation filter, ellipse fitting
DOI: 10.25165/j.ijabe.20201303.5346

Citation: Sun L Q, Chen S H, Liu T, Liu C H, Liu Y. Pig target tracking algorithm based on multi-channel color feature fusion. Int J Agric & Biol Eng, 2020; 13(3): 180–185.

1 Introduction

China is a big country for pig breeding. The pig breeding model is developing rapidly in the direction of scale, intelligence and welfare[1]. The monitoring of the growth status of pigs is the key to ensuring the healthy breeding of pigs. At present, it relies on artificial monitoring of the pig breeding process, with low efficiency and high error rate. The application of machine vision technology to monitor live pigs can effectively improve production efficiency. Real-time accurate detection and tracking of pigs is a prerequisite for the behavioral monitoring of pig eating, drinking, and excretion. Therefore, many researchers have done a lot of research on this. Van et al.[2] discussed the feasibility of using a computer vision system to improve livestock production for the low productivity and product quality. McFarlane et al.[3] used the blob edges of the pig in the segmented image to track the piglets in order to monitor the activity of the piglets in farrowing pens. Tillett et al.[4] used point distribution models to identify specific movements such as back bending and head nodding to study the behavioral characteristics of pigs. Perner et al.[5] addresses the occlusion problem in the pig tracking process through an object-oriented method to prevent objects from getting separated or combined by the blocks. Lind et al.[6] tracked target pigs based on a combination of image-subtraction and automatic threshold detection methods. Ahrendt et al.[7] proposed an algorithm for tracking multi-target pigs, using the support map segments to build up a 5D-Gaussian model of the individual pigs, but when the animals move quickly or overlap each other, the method is difficult to achieve accurate tracking. Kashiha et al.[8] painted basic patterns on the back of pigs, and used Otsu method and elliptic fitting method to track the target pigs. Li et al.[9] proposed an improved algorithm based on the adaptive Gaussian mixture model, which overcomes the shortcomings of the traditional Gaussian mixture model in pig detection. Sun et al.[10] proposed an algorithm for multi-target pigs tracking loss correction based on Faster R-CNN. Ju et al.[11] used the YOLO object detector to detect the pig target and found the possible boundary line between the touching-pigs by analyzing the shape. Many algorithms have been proposed in the field of pig target tracking. The problems such as target size change, deformation, occlusion, background clutters and real-time tracking are still not solved well.

Correlation filtering is a kind of high-efficiency target tracking algorithm developed based on the correlation principle of signal processing technology. Bolme et al.[12] applied correlation filtering to the tracking domain based on the Minimum Output Sum of Squared Error (MOSSE). Henriques et al.[13] used Circulant Structure of Tracking-by-detection with Kernels (CSK) to enhance the tracking accuracy of the correlation filter algorithm. In view of the insufficiency of the CSK algorithm feature representation, João et al. used the Kernelized Correlation Filter (KCF)[14] to extend the HOG feature of multi-channel gradients based on CSK. Danelljan uses the Color Name (CN)[15] algorithm to extend the features of multi-channel colors based on CSK. The above algorithm only considers the initial frame size and fails to implement the scale update as the tracking target changes in size in the video image. Based on the KCF algorithm and using HOG features and color features, Li et al.[16] proposed Scale Adaptive with Multiple Features (SAMF), which uses the translation filter to perform target detection on multi-scale scaled image blocks. Each frame size detection needs to sample 7 image blocks. Martin et al.[17] proposed the Discriminative Scale Space Tracker (DSST) algorithm, which divides the tracking process into translation tracking and scale tracking. Each frame size detection needs to sample 33 image blocks. Danelljan et al.[18] proposed the fDSST algorithm to reduce the computational cost for the DSST algorithm. Besides above mentioned hand-crafted features, some trackers employ deep features to better describe the tracked objects. DeepSRDCF[19] replaces hand-crafted features with CNN features in a spatially regularized DCF framework. CCOT[20] solves the

Received data: 2019-08-21 **Accepted date:** 2019-12-13
Biographies: Shuaihua Chen, Master student, research interests: video image processing technology, Email: cshanysys@163.com; **Ting Liu**, Master student, research interests: video image processing technology, Email: lt_christine@qq.com; **Chunhong Liu**, Associate Professor, research interest: agricultural internet of things research, Email: sophia_liu@cau.edu.cn; **Yan Liu**, Master student, research interests: video image processing technology, Email: 912034741@qq.com;
***Corresponding author: Longqing Sun**, Associate Professor, research interest: computer technology. China Agricultural University, Beijing 100083, China. Email: sunlq@cau.edu.cn.

problem of training in a continuous spatial domain. ECO[21] reduces the dimensions of the convolutional layer filter in the CNN features based on the CCOT algorithm. Recent works[22-25] have shown impressive performance using Siamese networks. Among them, SiameFC[22] employs the same CNN to extract features of two image patches and combine their representations using a similarity metric. In general, depth feature extraction requires more time than traditional feature extraction[26], and the real-time performance of tracking is also limited.

In the process of tracking the target of the pig, as the size of the tracking target changes in the video image, the tracking target scale cannot be adaptively updated in real-time, resulting in a low accuracy of the tracking target. In this paper, a multi-channel color feature adaptive fusion algorithm based on histogram color similarity is proposed to improve the feature representation ability of the algorithm. The target pig's contour information is used for ellipse fitting, and the ellipse's minimum positive circumscribed rectangle is used as the adaptive scale information of the pig target to improve the accuracy of the algorithm in scale estimation. In this paper, the pig tracking evaluation data set is constructed to comprehensively evaluate the tracking performance of the algorithm. The experiment shows that the robustness of the proposed algorithm in the tracking target deformation and scale change is significantly improved.

2 Materials and methods

2.1 Pig tracking data set

The experimental data was collected from a pig farm in Shandong Province, China. On August 12, 2014, four 24-hour videos of pig pens are captured. The image resolution is 640×480 and the frame rate is 25 frames per second. The video capture device is installed above the side of the pens. The captured video images are randomly divided into two groups, the first group is used for algorithm comparison research, and the second group is used for practical application testing. In order to test the effect of the algorithm on scale variation and deformation, according to the OTB (Online Object Tracking) tracking dataset[27,28] standard, a tracking dataset with 15 fully annotated sequences (numbered S1, S2, ..., S15) is built using the first group of video to the facilitate evaluation of tracking. The dataset contains a variety of factors that affect the accuracy of the target tracking algorithm, such as deformation, occlusion, fast motion, motion blur, and illumination variation. The dataset overview is shown in Table 1, and the partial images with ground-truth annotations are shown in Figure 1.

Table 1 Description of the dataset

Sequence	Description	Number of Frames
S1	SV, DEF, MB, BC	379
S2	SV, DEF, FM, IPR	119
S3	OCC, DEF, MB, FM, IPR	98
S4	IV, SV, OCC, DEF, IPR	760
S5	IV, SV	129
S6	IV, SV, DEF, NIG	211
S7	SV, DEF, IPR, NIG	558
S8	SV, DEF	112
S9	IV, SV, OCC, DEF,	450
S10	IV, SV, DEF, MB, FM, IPR,	110
S11	IV, SV, DEF, IPR	275
S12	IV, SV, OCC, DEF, BC	252
S13	IV, SV, DEF, BC	136
S14	SV, DEF, BC	275
S15	IV, SV, OCC, DEF, BC	360

Note: IV: Illumination Variation; SV: Scale Variation; OCC: Occlusion; DEF: Deformation; MB: Motion Blur; FM: Fast Motion; IPR: In-Plane Rotation; BC: Background Clutters; NIG: Night

Figure 1 Partial images with ground-truth annotations

2.2 Algorithm principle

2.2.1 Correlation filter tracking principle

The correlation filtering algorithm uses the initial tracking target gray information to train the filter h. Searching for the pixel coordinates of the maximum value in the target candidate region response output g by using the filter, that is the new position of the tracking target; The filter is updated in real-time based on the target tracking results. In order to speed up the operation, the above calculations are computed in the Fourier domain Fast Fourier Transform[29], which is expressed as follows:

$$\mathbf{F}(g)=\mathbf{F}(f)\odot\mathbf{F}(h)^{*} \tag{1}$$

The ⊙ symbol denotes element-wise multiplication, * indicates the complex conjugate, g indicates the response output, f indicates the input image, and h indicates the filter template, **F** is the fast Fourier transform.

2.2.2 Color multi-channel feature selection and adaptive weighted fusion algorithm

In the RGB space, the 12 channel color feature maps of the tracking target are extracted according to Equation (2), and the statistical tests are performed on each color channel, and then several color channels are further selected as the color features of the tracking target. The Center Location Error (CLE) is computed as the Euclidean distance between the estimated center location of the target and the ground-truth. The smaller the CLE is, the better the tracking performance. In this paper, the Average Center Location Error (ACLE) of the 15 annotated sequences shown in

Table 1 is used as the evaluation index of the color feature channel screening, and the calculation is shown in Equation (3). The test results are shown in Figure 2. As can be seen from the figure, the ACLE of the I_7, I_{11}, and I_{12} color channels is small. The algorithm extracts the I_7, I_{11}, and I_{12} color channels to enhance the performance of the algorithm.

$$\left\{\begin{matrix} I_1 = R & I_2 = G & I_3 = B \\ I_4 = (R+G)/2 & I_5 = (R+B)/2 & I_6 = (B+G)/2 \\ I_7 = \max(R,G,B) & I_8 = \min(R,G,B) & I_9 = (\max(R,G,B)+\min(R,G,B))/2 \\ I_{10} = (R+G+B)/3 & I_{11} = 0.2*R+0.3*G+0.5*B & I_{12} = 0.299*R+0.587*G+0.114*B \end{matrix}\right\} \tag{2}$$

$$ACLE = \frac{\sum_{i=1}^{m} \frac{\sum_{j=1}^{n} \sqrt{(x_{ij}-x_{ij}^{gt})^2+(y_{ij}-y_{ij}^{gt})^2}}{n}}{m} \tag{3}$$

where, m is the number of sequences; n is the number of video frames of the S_m sequence, (x_{ij},y_{ij}) is the tracking center coordinate, and (x_{ij}^{gt} , y_{ij}^{gt}) is the true ground-truth center coordinate.

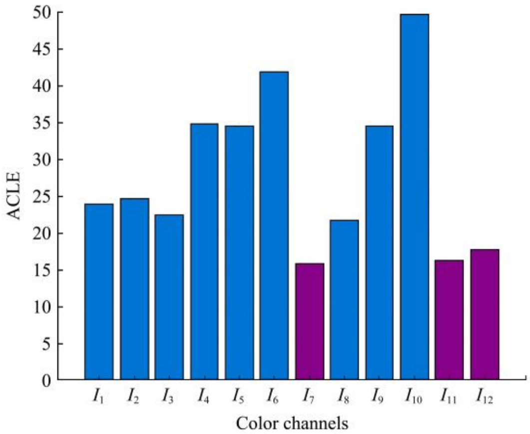

Figure 2 Average Center Location Error in 12 color channels

In order to further improve the tracking accuracy of the algorithm, the color histogram is used to establish the probability distribution of the tracking target color, and the Bhattacharyya Distance is used to measure the similarity of the color distribution of the target window of two adjacent frames[30]. The pixel value of the target window pixel is represented by $\{z_i\}$, and the histogram of the t-frame tracking target window is modeled according to Equation (4).

$$\hat{q}_u = C\sum_{i=1}^{d} K\left(\left\|\frac{z_i - z_t}{h}\right\|^2\right)\delta\left[\partial(z_i)-u\right], u=1,2,3,\ldots,k \tag{4}$$

where, z_t, z_i respectively represent the center point pixel and the ith pixel value of the tracking target window; d is the total number of pixels of the tracking target window; u is the number of color gray segments, k is the maximum number of segments of the color gray; $K(||z||^2)$ is the kernel function; h is the bandwidth of the kernel function; $\delta[\partial(z_i)–u]$ is the Kronecker pulse function, the function $\partial(z_i)$ maps each pixel into the corresponding feature space; C is a normalization constant. Similarly, the t+1 frame candidate tracking target histogram is modeled as shown in Equation (5), and z_{t+1} is the t+1 frame candidate tracking target center pixel value.

$$\hat{p}_u = C\sum_{i=1}^{d} K\left(\left\|\frac{z_i - z_{t+1}}{h}\right\|^2\right)\delta\left[\partial(z_i)-u\right], u=1,2,3,\ldots,k \tag{5}$$

The similarity between $\hat{q}_u$ and $\hat{p}_u$ is measured by the Bhattacharyya distance, ie

$$\rho = \sum_{u=1}^{k} \sqrt{\hat{q}_u \hat{p}_u} \tag{6}$$

The upper and lower frame similarities ρ_7, ρ_{11} and ρ_{12} of the I_7, I_{11}, and I_{12} color channels are calculated by the above equations.

The center coordinate $I_7(x_{t+1}, y_{t+1})$, $I_{11}(x_{t+1}, y_{t+1})$, $I_{12}(x_{t+1}, y_{t+1})$ of the t-th frame were calculated by the three color channels, and then the center position $I(x_{t+1}, y_{t+1})$ of the tracking target of the t+1th frame can be calculated as follows:

$$\begin{aligned} I(x_{t+1},y_{t+1}) &= \frac{\rho_7}{\rho_7+\rho_{11}+\rho_{12}} * I_7(x_{t+1},y_{t+1}) + \\ &\frac{\rho_{11}}{\rho_7+\rho_{11}+\rho_{12}} * I_{11}(x_{t+1},y_{t+1}) + \frac{\rho_{12}}{\rho_7+\rho_{11}+\rho_{12}} * I_{12}(x_{t+1},y_{t+1}) \end{aligned} \tag{7}$$

2.2.3 Adaptive scale update algorithm based on ellipse fitting

In order to enhance the robustness of the algorithm to the scale variation in the tracking process, the Otsu method[31] is used to segment the target pig in the current frame search region, and then the target pig edge contour coordinate $G(x, y)$ is extracted from the segmented image. The ellipse is fitted by the least squares method, and the minimum positive circumscribed rectangle of the ellipse is used as the adaptive scale information of the pig target. The elliptic equation is as shown in Equation (8) and the effect is shown in Figure 3. In order to maintain a robust filter, the filter scale is updated by the nearest neighbor interpolation method while the estimated scale is updated. Experiments show that this method can enhance the robustness of the algorithm to deformation and scale variation.

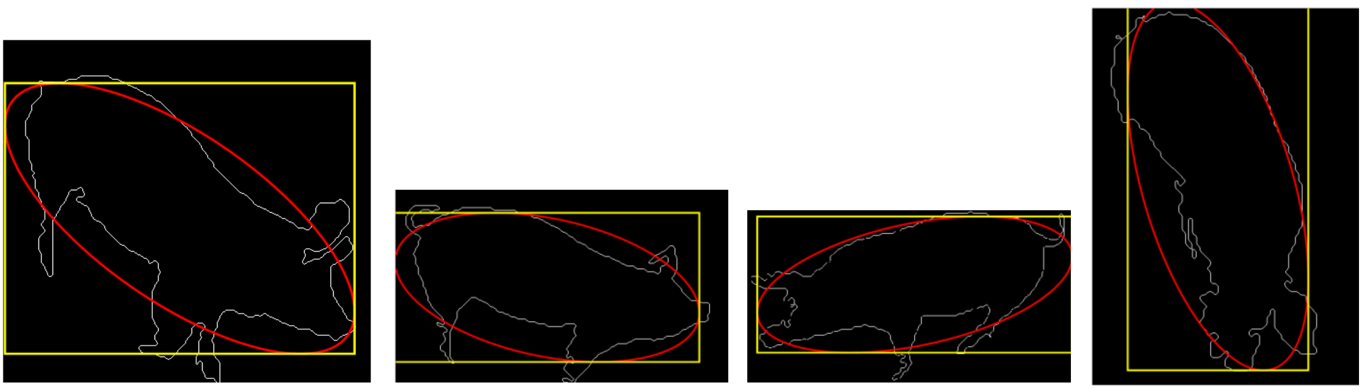

Figure 3 Scale information obtained by ellipse fitting

$$f(\boldsymbol{\alpha},\mathbf{X}) = \boldsymbol{\alpha}\mathbf{X} = Ax^2+Bxy+Cy^2+Dx+Ey+F=0 \quad (8)$$

where, $\boldsymbol{\alpha}$ is [A, B, C, D, E, F]; $\mathbf{X}_i$ indicates the ith contour coordinate $[x_i^2, x_i y_i, y_i^2, x_i, y_i, 1]$; substituting the edge contour coordinates $G(x, y)$ of the pig into Equation (8):

$$f(A,B,C,D,E,F) = \sum_{i=1}^{n}(Ax_i^2 + Bx_i y_i + Cy_i^2 + Dx_i + Ey_i + F)^2 \quad (9)$$

where, n is the total number of edge contour pixels, to minimize the above formula $\frac{\partial f}{\partial A} = \frac{\partial f}{\partial B} = \frac{\partial f}{\partial C} = \frac{\partial f}{\partial D} = \frac{\partial f}{\partial E} = \frac{\partial f}{\partial F} = 0$.

3 Results and discussion

3.1 Quantitative evaluation

In order to comprehensively evaluate the performance of the tracking algorithm in terms of deformation, scale variation and occlusion, the tracking dataset with 15 fully annotated sequences shown in Table 1 are selected to quantitatively evaluate the proposed method with baseline algorithms of KCF, fDSST, SiameFC and ECO. Two evaluation indices were compared: Distance Precision (DP) and Overlap Precision (OP). DP is the relative number of frames in the sequence where the center location error is smaller than a certain threshold. The threshold (d) is taken from 0 to 50 to draw the precision plot, and the calculation formula is as shown in Equation (10). OP is defined as the percentage of frames where the bounding box overlap exceeds a threshold $a\in[0, 1]$, and the calculation formula is as shown in Equation (11).

$$DP = \frac{\sum_{j=1}^{n} b(\sqrt{(x_j - x_j^{gt})^2 + (y_j - y_j^{gt})^2} \le d)}{n} \quad (10)$$

$$OP = \frac{\sum_{j=1}^{n} b(\frac{r_j I r_j^{gt}}{r_j U r_j^{gt}} \ge a)}{n} \quad (11)$$

In the above two formulas, $b(\alpha)$ is the indicator function, and When α is true or false, the value is 1,0 respectively; n is the total number of frames of the video sequence.

All the experiments are conducted on an Intel i7-6700 CPU (3.40 GHz) PC with 16 GB memory. We implemented the proposed tracker by native Matlab without optimization. The average distance and overlap precision of the algorithm are tested and compared on the 15 tracking sequences shown in Table 1. As shown in Figure 4, and the red curve represents our algorithm, and the blue, green, yellow and magenta curves represent the KCF, fDSST, SiameFC and ECO algorithms, respectively. It can be seen from the figure that the distance precision value and the overlap precision value of our algorithm under different thresholds perform better than other algorithms. The criterion is to use less than 20 pixels as the distance threshold and greater than the overlap threshold of 0.5 as the criterion for judging whether the tracking process is successful[15, 27-28]. On the distance precision (DP), the average precisions of the KCF, fDSST, SiameFC and ECO algorithms on 15 tracking sequences are 54.0%, 63.3%, 66.2% and 71.0%. The average precision of our algorithm is 74.3%, which is 20.3%, 11.0%, 8.1% and 3.3% higher than the KCF, fDSST, SiameFC and ECO algorithms respectively. On the overlap precision (OP), the average precisions of the KCF, fDSST, SiameFC and ECO algorithms on 15 tracking sequences are 42.9%, 57.9%, 60.6% and 62.9%. The average overlap precision of our algorithm is 64.9%, which is 22.0%, 7.0%, 4.3% and 2.0% higher than the average precisions of the KCF, fDSST, SiameFC and ECO algorithms, respectively. Our algorithm's running speed is 50.1FPS in CPU (Frames Per Second), which can track pig movement in real-time.

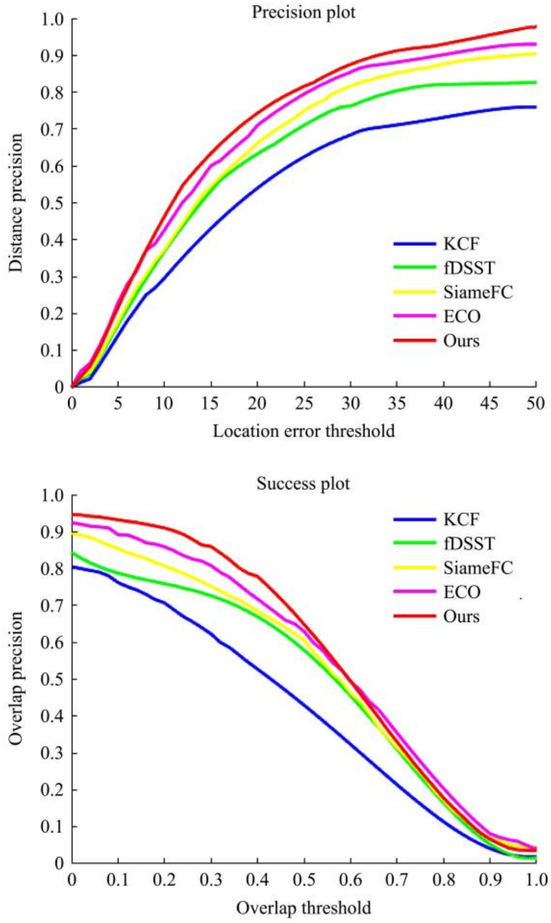

Figure 4 DP and OP values of different algorithms on the dataset

3.2 Qualitative evaluation

The results of the target tracking section are shown in Figure 5. The red rectangle represents the algorithm of our algorithm, while the blue, green, yellow and magenta rectangles represent the KCF, fDSST, SiameFC and ECO algorithms respectively.

In dealing with the problem of deformation, our approach takes full account of the characteristics of pig motion and proposes adaptive color feature fusion to improve the processing ability of the algorithm to deformation. As shown in the sequence of S1 in Figure 5, in the process of the transition of the live pig from the left and right motion state to the front and rear motion state, our approach can effectively deal with the influence of the deformation of the pig on the tracking performance; however, the KCF, fDSST, SiameFC and ECO algorithms fail to track. Our algorithms also perform well in sequence such as S9, S10, S11 and S14 in Figure 5.

On the issue of dealing with scale variation, the segmentation results using Otsu method can effectively estimate the scale of the target pig individual, and the scale update has the characteristics of self-adaptation. As shown in the sequence of S6, S8 and S13 in Figure 5, when the motion of the pig is gradually approaching or away from the camera position, that is the moving pig target is reduced or expanded, our algorithm can adapt to the change of the scale information of the pig in real-time, and achieve accurate tracking of the target pig. The KCF algorithm has no scale update strategy. The scale update of fDSST, SiameFC and ECO algorithm fails to adapt to the change of the tracking target scale.

The oversized and undersized scale estimation will cause the tracking failure.

In this paper, the tracking performance of the algorithm under nighttime or low-light conditions is tested. The tracking performance of our algorithm under low-light conditions is good as shown in the S6 sequence in Figure 5.

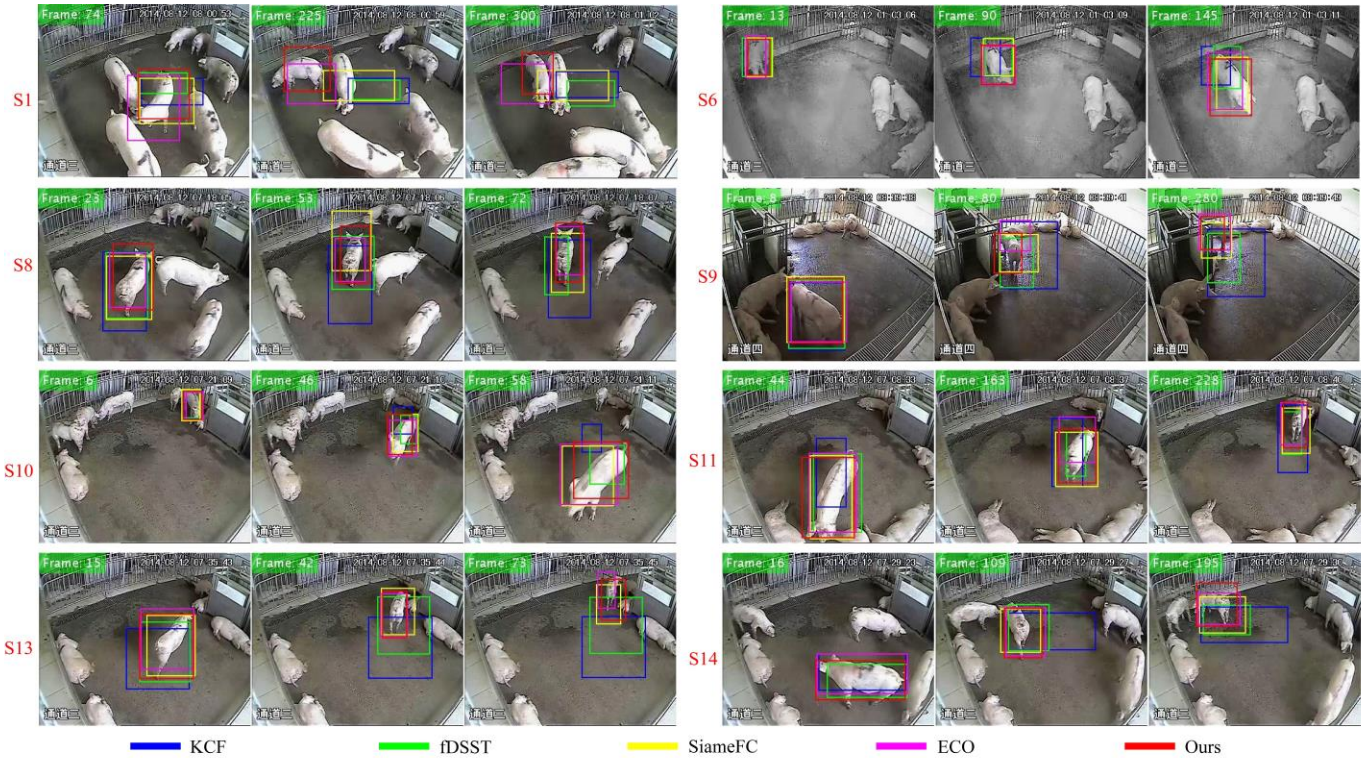

Figure 5 Different algorithm tracking results

3.3 Algorithm tracking application

The practical application performance of the algorithm is tested by using the second group of monitoring video data collected in the previous period. Figure 6 records the DP and OP values of the 45000 frame video sequence. It can be seen from the figure that the distance precisions of the KCF, fDSST, SiameFC and ECO algorithms are 63.6%, 77.5%, 82.9% and 86.8%, respectively. Our algorithm can accurately track the target, and the distance precision is 89.7%, which is increased by 26.1%, 12.2%, 6.8% and 2.9%, respectively. The overlap precisions of the KCF, fDSST, SiameFC and ECO algorithms are 59.2%, 70.6%, 77.3% and 83.2%. The overlap precision of the algorithm is 87.5%, which is increased by 28.3%, 16.9%, 10.2% and 4.3%, respectively, to meet the practical application requirements. The problem of scale variation and deformation is effectively solved, and it has better robustness.

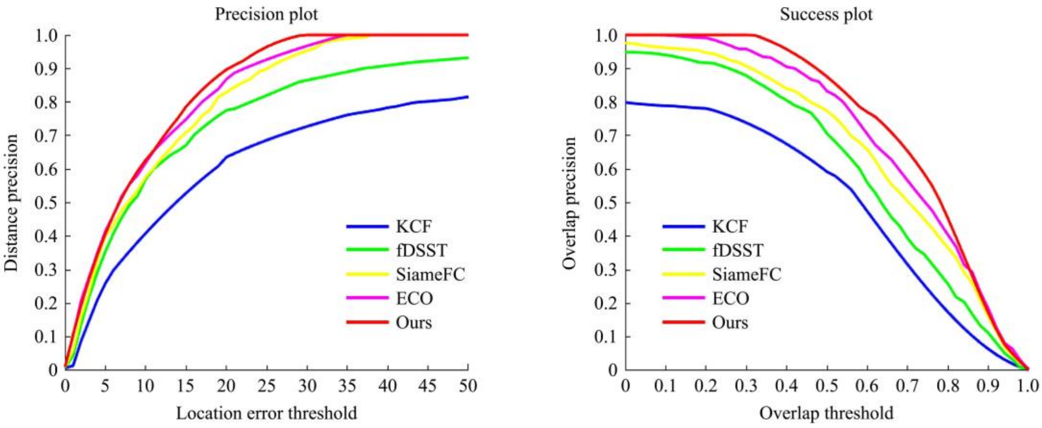

Figure 6 DP and OP values of different algorithms on random video data sets

4 Conclusions

In the process of tracking the target of the pig, as the size of the tracking target changes in the video image, the tracking target scale cannot be adaptively updated, resulting in a low accuracy of the tracking target. In this paper, a multi-channel color feature adaptive fusion algorithm is proposed, which uses the target pig Contour information to realize real-time adaptive pig target scale update. The main conclusions are as follows:

a) The evaluation index of the algorithm in the tracking data set is better than the similar algorithm, and the average operation speed is 50.1 FPS; From the qualitative evaluation, it can be seen that our algorithm can effectively deal with the problem of tracking loss caused by deformation and scale variation of a pig to improve the accuracy of the algorithm and meet the requirements of real-time.

b) It has good tracking performance on randomly selected actual monitoring data. The distance precision and overlap precisions are 89.7% and 87.5%, respectively, which can meet the practical application requirements of pig monitoring.

Acknowledgements

This work was supported in part by the National Key Research and Development Plan for the 13th Five-Year Plan under Grant 2016YFD0700200. This work was supported in part by the National High Technology Research and Development Program of China (2013AA102306).

[References]

[1] Nasirahmadi A, Sturm B, Olsson A C, Jeppsson K H, Müller S, Edwards S, et al. Automatic scoring of lateral and sternal lying posture in grouped pigs using image processing and Support Vector Machine. Computers and Electronics in Agriculture, 2019; 156: 475–481.

[2] Van der Stuyft E, Schofield C P, Randall J M, Wambacq P, Goedseels V. Development and application of computer vision systems for use in livestock production. Computers and Electronics in Agriculture, 1991; 6: 243–265.

[3] McFarlane N J B, Schofield C P. Segmentation and tracking of piglets in images. Machine Vision and Applications, 1995; 8: 187–193.

[4] Tillett R D, Onyango C M, Marchant J A. Using model-based image processing to track animal movements. Computers and Electronics in Agriculture, 1997; 17(2): 249–261.

[5] Perner P. Motion tracking of animals for behavior analysis. In: Arcelli C, Cordella L P, di Baja G S (Ed.). Visual Form 2001. IWVF 2001. Lecture Notes in Computer Science, Vol. 2059, Berlin, Heidelberg: Springer, 2001; pp.779–786.

[6] Lind N M, Vinther M, Hemmingsen R P, Hansen A K. Validation of a digital video tracking system for recording pig locomotor behaviour. Journal of Neuroscience Methods, 2005; 143(2): 123–132.

[7] Ahrendt P, Gregersen T, Karstoft H. Development of a real-time computer vision system for tracking loose-housed pigs. Computers and Electronics in Agriculture, 2011; 76(2): 169–174.

[8] Kashiha, M, Bahr C, Ott S, Moons C P H, Niewold T A, Ödberg, F O, et al. Automatic identification of marked pigs in a pen using image pattern recognition. Computers and Electronics in Agriculture, 2013; 93: 111–120.

[9] Li Y Y, Sun L Q, Zou Y B, Li Y. Individual pig object detection algorithm based on Gaussian mixture model. Int J Agric & Biol Eng, 2017; 10(5): 186–193.

[10] Sun L Q, Zou Y B, Li Y, Cai Z D, Li Y, Luo B, et al. Multi target pigs tracking loss correction algorithm based on Faster R-CNN. Int J Agric & Biol Eng, 2018; 11(5): 192–197.

[11] Ju M, Choi Y, Seo J, Sa J, Lee S, Chung Y, et al. A Kinect-based segmentation of touching-pigs for real-time monitoring. Sensors, 2018; 18(6): 1746.

[12] Bolme D S, Beveridge J R, Draper B A, Lui Y M. Visual object tracking using adaptive correlation filters. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Fancisco: IEEE. 2010; pp.2544–2550.

[13] Henriques J F, Caseiro R, Martins P, Batista J. Exploiting the circulant structure of tracking-by-detection with kernels. In: Fitzgibbon A., Lazebnik S., Perona P., Sato Y., Schmid C (Ed.). Computer Vision-ECCV 2012. Lecture Notes in Computer Science, Vol. 7575, Berlin, Heidelberg: Springer, 2012; pp.702–715.

[14] Henriques J F, Caseiro R, Martins P, Batista J. High-speed tracking with kernelized correlation filters. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015; 37(3): 583–596.

[15] Danelljan M, Shahbaz Khan F, Felsberg M, van de Weijer J. Adaptive color attributes for real-time visual tracking. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014; pp.1090–1097.

[16] Li Y, Zhu J K. A scale adaptive kernel correlation filter tracker with feature integration. In: Agapito L, Bronstein M, Rother C (Ed.). Computer Vision-ECCV 2014 Workshops. Lecture Notes in Computer Science, Vol. 8926, Cham: Springer, 2014; pp.254–265.

[17] Danelljan M, Häger G, Shahbaz Khan F, Felsberg M. Accurate scale estimation for robust visual tracking. In: Proceeding of the British Machine Vision Conference, Nottingham: BMVA Press, 2014.

[18] Danelljan M, Häger G, Shahbaz Khan F, Felsberg M. Discriminative scale space tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017; 39(8): 1561–1575.

[19] Danelljan M, Hager G, Shahbaz Khan F, Felsberg M. Convolutional features for correlation filter based visual tracking. In: The IEEE International Conference on Computer Vision (ICCV) Workshops, 2015; pp.58–66.

[20] Danelljan M, Robinson A, Khan F S, Felsberg M. Beyond correlation filters: Learning continuous convolution operators for visual tracking. In: Leibe B, Matas J, Sebe N, Welling M (Ed.). Computer Vision-ECCV2016. Lecture Notes in Computer Science, Vol. 9909, Cham: Springer, 2016; pp.472–488.

[21] Danelljan M, Bhat G, Shahbaz Khan F, Felsberg M. Eco: Efficient convolution operators for tracking (CVPR). In: The IEEE Conference on Computer Vision and Pattern Recognition, 2017; pp.6638–6646.

[22] Bertinetto L, Valmadre J, Henriques J F, Vedaldi A, Torr P H S. Fully-convolutional siamese networks for object tracking. In: Hua G, Jégou H (Ed.). Computer Vision-ECCV2016 Workshops. Lecture Notes in Computer Science, Vol. 9914, Cham: Springer, 2016; pp.850–865.

[23] Valmadre J, Bertinetto L, Henriques J, Vedaldi A, Torr P H S. End-to-end representation learning for correlation filter based tracking. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017; pp.2805–2813.

[24] Zhang Y H, Wang L J, Qi J Q, Wang D, Feng M Y, Lu H C. Structured siamese network for real-time visual tracking. In: The European Conference on Computer Vision (ECCV), 2018; pp.351–366.

[25] Li B, Yan J J, Wu W, Zhu Z, Hu X L. High performance visual tracking with siamese region proposal network. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018. pp. 8971–8980.

[26] Li P X, Wang D, Wang L J, Lu H C. Deep visual tracking: Review and experimental comparison. Pattern Recognition, 2018; 76: 323–338.

[27] Wu Y, Lim J, Yang M H. Online object tracking: A benchmark. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013. pp.2411–2418.

[28] Wu Y, Lim J, Yang M H. Object Tracking Benchmark. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015; 37: 1834–1848.

[29] Flannery B P, Press W H, Teukolsky S A. Numerical recipes in C. New York: Press Syndicate of the University of Cambridge, 1992; 24: 78.

[30] Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003; 25(5): 564–575.

[31] Otsu N. A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics, 1979; 9(1): 62–66.