Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

Individual identification of Holstein dairy cows based on detecting and matching feature points in body images

Kaixuan Zhao [a,b,c], Xin Jin [a,b], Jiangtao Ji [a,b,*], Jun Wang [a], Hao Ma [a], Xuefeng Zhu [a]

[a] College of Agricultural Equipment Engineering, Henan University of Science and Technology, Luoyang 471003, China

[b] Collaborative Innovation Center of Machinery Equipment Advanced Manufacturing of Henan Province, Luoyang 471003, China

[c] Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture Rural Affairs, Yangling, Shaanxi 712100, China

ARTICLE INFO

Article history:
Received 24 December 2018
Received in revised form
21 February 2019
Accepted 6 March 2019
Published online 29 March 2019

Keywords:
Cow identification
Feature detection
Feature descriptor
Image matching
Image processing

Image processing technology has been used in precision dairy farming to support management decisions. Vision-based animal identification systems can become a potential alternative to RFID. In this paper, a vision system is proposed to extract body images and identify Holstein cows. Side view videos of dairy cattle walking in a straight line were collected. Cow mask was detected using Adaptive SOM method. The largest inscribed rectangle was extracted to locate the cow's body area. A total of 528 videos were collected from 66 cows, and 3 videos were randomly selected for each cow to build template datasets, while the rest of the videos were used as test data. Feature points of the body image were extracted and matched with the template dataset to identify the unknown cow. Four feature extraction methods and two matching methods were investigated and evaluated. The results showed that the highest identification accuracy was 96.72% when the FAST, SIFT and FLANN methods were used for feature extraction, descriptor, and matching, respectively. However, the combination of ORB and BruteForce had better computational efficiency on the basis of high accuracy. Software was implemented and can realise accurate identification of dairy cattle in real-time.

* *Corresponding author*. College of Agricultural Equipment Engineering, Henan University of Science and Technology, Luoyang 471003, China.
E-mail address: jjt0907@163.com (J. Ji).
https://doi.org/10.1016/j.biosystemseng.2019.03.004
1537-5110/

1. Introduction

Video analysis technology originating from intelligent surveillance has gradually been applied into many fields of livestock and poultry farming (Ahrendt, Gregersen, & Karstoft, 2011; Dutta et al., 2015; Jover et al., 2009; Matthews, Miller, Clapp, Plötz, & Kyriazakis, 2016). Among them, intelligent perception and recognition of dairy cow behaviour can provide evidence for decision making in farm management and have become a hotspot in the field of precision livestock farming (Arcidiacono, Porto, Mancino, & Cascone, 2017; Nasirahmadi, Edwards, & Sturm, 2017; Zhao, Bewley, He, & Jin, 2018; Zheng et al., 2018). Individual identification was a technical prerequisite and application basis for automatic analysis of cow behaviour (Yajuvendra et al., 2013).

Radio frequency identification (RFID) has been proven as a prevalent solution for individual identification on farm (Cappai, Rubiu, Nieddu, Bitti, & Pinna, 2018; Samad, Murdeshwar, & Hameed, 2010). Many researchers also developed electric sensors based on this technology. Voulodimos, Patrikakis, Sideridis, Ntafis, and Xylouri (2010) described a platform for livestock management based on RFID-enabled mobile devices. RFID with tag numbers were associated with individual animals, and a data records system was developed to recognise basic information about each animal. Pang, He, Li, Huang, and Zheng (2011) proposed a new system integrated RFID technology with a wireless sensor network. Network structure and communication protocol were investigated. An information-acquiring and transmitting system was developed and can ensure the integrity of the cow record was complete and provided traceability.

The typical electronic identification device must be attached on the neck or ear by puncturing. Cow movements may affect the performance and the detection distance can be limiting. Installation of RFID may cause damage and affect animal welfare. Damage to the device or its loss may incur economic cost. Most importantly, for a computer vision system that needed individual identification, an additional RFID reader is needed to identify the animal in the video field, which increases complexity and cost of vision system used on the farm (Chapinal & Tucker, 2012; Hoffmann et al., 2013; Xiong, Qian, Luo, & Lu, 2005).

Therefore, a system that can realise individual identification of cows based on vision system could be valuable. Cai and Li (2013) presented a facial representation model of cattle based on local binary pattern (LBP) and extended LBP descriptors. Xia and Cai (2012) studied the cattle face recognition based on sparse representation classifier (SRC). Those studies demonstrated that vision-based animal identification could become a potential alternative to RFID. It could further improve the practicability of video analysis technology in the field of precision dairy farming. However, there is often little individual characteristic information contained in the face. In addition, the head position and view angle have great influence on the result of face recognition. The identification of individual dairy cattle using face recognition would not work for some side view vision systems.

Holstein cows are the most common in dairy farms in the US and many other countries. Holstein cows have a white-black body pattern that contains much individual information, which could be directly processed to identify the animal. Kim, Choi, Lee, and Yoon (2005) designed a computer vision system to recognise an individual Holstein cattle by processing images of their body patterns. This system involved image capture, image pre-processing, algorithm processing, and an artificial neural network recognition algorithm. Images of the 49 cows were analysed to learn input layer elements, and 10 cows were used to verify the output layer elements in the neural network. The system proved to be reliable for the individual recognition of cows in natural light. However, in this system, pure colour background was applied, which is far different from actual dairy environments. Furthermore, the artificial neural network recognition algorithm contained too many input and output neurons, which easily resulted in over-fitting for a large group.

By converting the problem of comparison and matching of individual cow images into a pattern recognition problem, each pattern in the data set corresponds to an individual cow. Convolution neural network (CNNs) developed in deep learning have been successfully applied in the research of pattern matching (Gao & Wang, 2014; Yu, Yin, Yin, & Du, 2014; Zhao & He, 2015). By using a large dataset containing thousands of manually matched images to train the network, the CNN approach was convenient and accurate to identify the individual cow (Li, Ji, Wang, Sun, & Yang, 2017). Based on this idea, Zhao and He (2015) proposed a model for individual identification of cows using CNNs under the actual environment of the dairy farm. The body images were interpolated and normalised into a matrix sized 48 $\times$ 48 as the input for the CNN network. The output patterns were corresponded with IDs of cows. After training the model by using 60,000 images, 90.55% of 21,730 testing images were recognised correctly.

Because the output of the pattern recognition model corresponds to the ID of cows in the herd, the size of the model will increase exponentially for large dairy farms. Meanwhile, the number of images needed for training will increase dramatically. In addition, because the intrinsic information of the image is stored in the neural network, the whole model must be retrained once new cows had joined the herd. This characteristic seriously reduced the practicability of the CNN network for identification of dairy cows.

Therefore, the objective of this paper is to develop an individual identification system for dairy cows based on machine vision. The system should meet two criteria: a) high identification accuracy; b) the database can be updated quickly and conveniently when new cows enter the herd. Therefore, a model based on feature points detection and image matching was proposed. On the basis of constructing the identification reference dataset of dairy cattle, the image in the database closest to the unknown cow was found as the result of individual identification by detecting and matching the feature points in the body image of cows. Different feature extraction and matching methods were evaluated to select the combination with best identification performance. Software was developed to implement real time identification of cows in farm use.

2. Materials and methods

2.1. Image acquisition

Videos of Holstein cows were captured when they walked on a path back to the barn after milking on a research dairy farm at the University of Kentucky, USA in early August, 2016. There were 66 cows in the group milked three times a day and walked back to the barn after milking through a narrow passage limited by electrical fences. A Nikon D5200 camera (Tokyo, Japan) was installed on a tripod 3.5 m away from path and 1.5 m above the ground. The camera used a 35 mm lens (Nikon AF-S DX 35 mm f/1.8G, Tokyo Japan) under ISO 400, auto exposure, and auto focus mode. Image resolution was 720 pixels (vertical) $\times$ 1280 pixels (horizontal). The camera was adjusted parallel to the walking direction of cattle by using a sensitive electric compass. There were two electric fences between the cow target and the camera, and the width was about 1 cm for each. Human interrupt was used to ensure every video only contained one cow.

Videos were collected during the time period 16:00 to 18:00 on sunny days. Cows for image capture had no contaminating mud on their bodies and were captured under natural lighting conditions. At least 10 s of background was recorded before the cow entered into the vision view for each video. The videos were continuously recorded until cows walked to the right edge of the vision field. The captured videos were stored in the camera's local memory card.

2.2. Dataset

The image of each cow was captured once a day for 8 consecutive days. Thus, each cow gave a total of 8 videos during the experiment. Among the 8 videos of each cow, 3 were randomly selected for the reference dataset, and the remaining 5 videos were used as test data. Therefore for all 66 cows, the reference dataset and test dataset contained 198 and 330 videos, respectively, totalling 528 videos. When the cow walked into the middle of the vision field in each video, the current frame image was processed to identify the cow in it.

2.3. Extraction of body image

Figure 1 shows image processing flow to extract the body image from a frame. The first step was to detect the cow target from background. BGSLibrary is an integrated C++ open library that implements different methods for foreground-background separation (Sobral, 2013). We used this library and tested 43 methods in it, including ViBe (Barnich & Van Droogenbroeck, 2011), Mixture of Gaussian (Zivkovic, 2004), Adaptive SOM (Maddalena & Petrosino, 2008), etc. According to human validation of the detection performance, Adaptive SOM was selected to segment the cow target as it had best separation performance and was adaptable to different light conditions. The largest region in the mask image was selected and optimised with a series of processing steps including dilating, eroding and hole-filling.

The next step was to extract body area from the mask image. The body area was defined as the largest inscribed rectangle within the cow mask image. We developed an algorithm to find the largest rectangle for any non-convex polygon. The pseudo code of the algorithm is:

```
rows = image's height
cols = image's width
area_max = 0
For i = 1 to rows
  For j = 1 to cols
    If image (i, j) is background
      Skip For
    End if
    If i is equal to 1
      h (i, j) = 1
    Else
      h (i, j) = h (i − 1, j) + 1
    End if
    If j is equal to 1
      w (i, j) = 1
    Else
      w (i, j) = w (i, j − 1) + 1
    End if
    minw = cols
    area_max = 0
    For k = 1 to h (i, j)
      minw = min(minw, w (i − k, j))
      area = (k + 1) × minw
      If area > area_max
        area_max = area
        rect.x = j − minw + 1
        rect.y = i − k
```

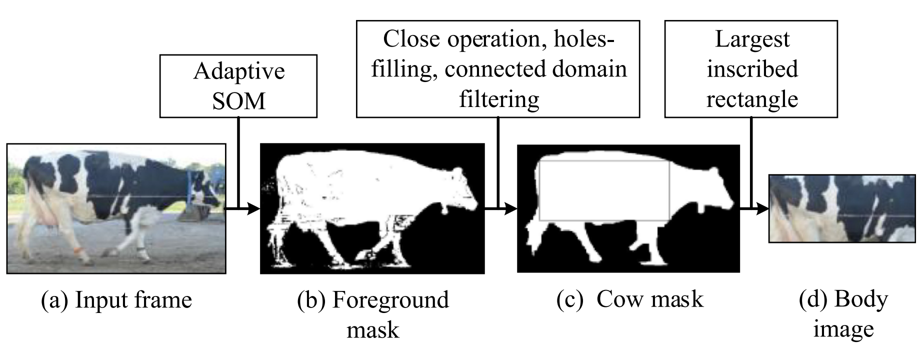


Fig. 1 – Steps for extracting body area from an input frame.

$rect.width = minw - 1$
$rect.height = k$
End if
End For
End For
End For
Output *rect*

2.4. Individual identification model

The principle and process of cow identification are similar to fingerprint identification. The fingerprint identification process follows the ACE-V (analysis, comparison, evaluation and verification) method to ensure the reliability of recognition. ACE-V is the description of a process used in comparing two different things and having someone else agree with the result of the comparison.

The identification model designed on the basis of the ACE-V principle is shown in Fig. 2. There were two main stages. The first stage was preparing the template dataset, and the second stage was identification of an unknown cow. The system consisted of four modules, including images database, features database, enrolment module and matching module.

The images database stored the body images. The features extracted from these body images were stored in the features database along with the cow ID. The objective of the enrolment module was to admit a cow using its ID and body pattern into an image database after the process of feature extraction. These features form the template that was used to determine or verify the identity of the unknown input, formulating the process of matching. The component of the model used for matching was referred to as the matching module. In this paper, four feature extraction methods and two matching methods were tested and analysed.

2.4.1. Feature extraction methods

2.4.1.1*FAST*. Features from Accelerated Segment Test (FAST) is a corner detection method, which can be used to extract feature points and to complete tracking and mapping objects (Rosten & Drummond, 2006). The most outstanding advantage of the algorithm is its sensitivity. The algorithm used 16 pixels (Bresenham circle with radius 3) to determine whether the centre pixel P was a corner point. Pixels around the pixel P were numbered in clockwise order from 1 to 16 on the circumference. A circular pixel was called a corner if there were N consecutive pixels on the circumference that were brighter than the sum of the centre pixel's brightness and a threshold t, or darker than the subtraction of centre pixel's brightness and a threshold t' (Rosten & Drummond, 2006). The FAST algorithm is easy to implement, and is especially featured for its high speed.

2.4.1.2*SIFT*. Scale Invariant Feature Transformation (SIFT) is an algorithm for detecting local features. This algorithm obtains features by finding interest points in a graph and its descriptors related to scale and orientation, and achieves good results in matching image feature points (Lowe, 1999). SIFT features are not only scale-invariant, but also give good detection results even if the rotation angle, image brightness or shooting angle are changed (Lowe, 2004). In the original paper, Lowe suggested that the descriptor used the gradient information of 8 directions calculated in the window of 4×4 in the critical point scale space, which represented a total of 128 ($4 \times 4 \times 8$)-dimension vector (Lowe, 1999).

2.4.1.3*SURF*. Speeded Up Robust Features (SURF) is a robust algorithm for detection and description of local feature point (Bay, Tuytelaars, & Van Gool, 2006). As with SIFT algorithm, the basic SURF algorithm can be divided into three parts: extraction of local feature points, description of feature points, matching feature points. However, there were two major innovations in the performance efficiency of SURF (Bay, Ess, Tuytelaars, & Van Gool, 2008). One is the use of integral graphs on Hessian (Hessian matrix), and the other is the use of dimension-reduced feature descriptors. Taking 5×5 pixels as a sub-region, a range of 20×20 pixels around the feature point was selected and contains 16 sub-regions. The sum of the direction of Hal wavelet transform in the x, y direction of the sub-region (where the direction parallel to the feature point is x, and the direction vertical to the feature point is y) is calculated as Σdx and Σdy, and the sum of the length of that is calculated as $\Sigma|dx|$ and

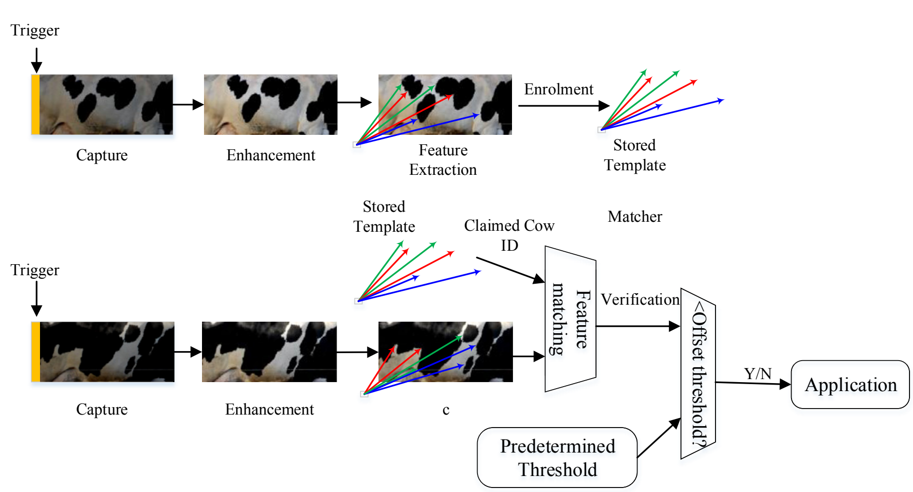


Fig. 2 – Individual identification model for Holstein cows.

Σ|dy|. Therefore, for each sub-regions, 4 descriptors are calculated. Finally, a 64-dimension feature is generated.

2.4.1.4ORB. Oriented FAST and Rotated BRIEF (ORB) is another fast feature point extraction and description algorithm (Rublee, Rabaud, Konolige, & Bradski, 2011). ORB algorithm is divided into feature point extraction and feature point description. Feature extraction is developed by FAST algorithm, and feature point description is improved according to BRIEF (Binary Robust Independent Elementary Feature) description algorithm. The main idea is to select *n* groups of point pairs randomly near the feature points, and combine the grey value of these points into a binary string as the feature descriptor of the feature point. ORB feature combines the FAST feature point detection method with the BRIEF feature descriptor. The ORB algorithm is 100 times faster than the SIFT algorithm and 10 times faster than the SURF algorithm (Rublee et al., 2011).

2.4.2. *Matching methods*

2.4.2.1FLANN. The FLANN algorithm is based on the K-means tree or KD-TREE search operation (Muja & Lowe, 2009). In that paper, KD-TREE was used to classify data points in n-dimension space R^n as specific parts, the object of which was to find the nearest Euclidean distance between searching points and query points in KD-TREE (Chum, Philbin, & Zisserman, 2008). After all Euclidean distances $d\ (p,\ q)$ in vector space R^n have been stored in KD-TREE structure (also known as train of KD-TREE), the nearest point to the reference point can be searched effectively. The whole search process is a top-down recursive process within the KD-TREE. Firstly, the value of the target point and the segmentation point are compared based on a certain dimension to distinguish whether the target point is in the left region or the right region. The loop is then compared with the corresponding nodes until the feature points are matched successfully. The cow body image for querying was matched with the images in template dataset respectively. The more key points in the two images that were successfully matched, the more similar the two images were. In the template dataset, each cow had 3 template images. The query images were matched with the three images separately, and the average number of feature points during the three matches was taken as the match result for that group of template images. The template groups with the most matching points referred to the ID of the cow for querying.

2.4.2.2BruteForce. Brute Force (BF) algorithm is a common pattern matching algorithm. Taking the feature points in query image and candidate image in dataset as set *U* and *T* respectively, the first step of the BF algorithm was to match the first point in *U* with the first point in *T*. If they were matched, then continue to compare the second point in *U* and the second point in *T*. If they were not matched, then the second point of *U* and the first point of *T* were compared successively until an acceptable matching was obtained. When using BruteForce for SIFT and SURF features, Euclidean distance was used for measurement. For binary string-based descriptors like ORB, BRIEF, BRISK etc., Hamming distance was needed for measurement.

2.4.3. *Combination of algorithms*

FAST is an algorithm for feature point detection. It cannot extract the descriptor of feature points. Therefore, FAST needs to be combined with other descriptors to match feature points. SIFT, SURF, ORB algorithm can not only detect feature points, but also generate the feature descriptor. In theory, one of those three algorithms can be used to detect and match the feature points at the same time. In order to fully compare the matching performance of different algorithms, one algorithm was combined with a new generation one. In addition, the FAST method was combined with the ORB method to increase the feature points detected. Finally, four groups of different combinations were determined, i.e. FAST + SIFT, SIFT + SURF, SURF + ORB, and FAST + ORB. FLANN was used for nearest neighbour search of high-dimensional features in big data sets. BruteForce tried all possible matches so that it can find the best match. The FLANN algorithm can reduce the number of comparisons for a large dataset, but may lose correct matching. When there were fewer matching points or the time of a single match was relatively short, BruteForce was used to obtain a better match result. Otherwise, FLANN was used to improve matching efficiency.

2.5. *Selection of matching points*

In the process of feature matching, error matching of feature points may occur due to obstructions in front of cows and complex background. Because of the high dimension of feature points, the probability of error matching was considerable. Lowe proposed a filter method to compare the nearest neighbour distance with the next nearest neighbour distance in order to exclude invalid feature points in the matching results (Muja & Lowe, 2009). If the ratio of nearest distance and the second nearest distance was less than a threshold *T*, then the matching point was accepted. Less SIFT points were matched using a smaller proportional threshold, but the stability was enhanced. Lowe recommended a ratio threshold of 0.8 for the best match (Muja & Lowe, 2009).

However, it was found that the filtering performance of the same ratio value for different cows was not always stable. In most cases, even if the ratio value was very large, a great number of correct matching points were still filtered out. In addition, the body pattern of cows always has similar appearance, which may result in frequent error matches. In order to solve those problems, this paper proposed a filtering method based on the relative offset of matched points.

The situation of similar pattern of dairy cows was analysed with the example shown in Fig. 3. Figure 3a and b was supposed to be the body images of two cows, where cow *a* had a large triangle spot (S) in the left, and cow *b* had a scaled and rotated triangle spot (S′) in the upper right corner. Because of the invariance of the translation, rotation and scaling of the algorithms, the feature points in the two patterns were matched to form the point pairs of P_i–P_i'. These unexpected matches had considerable impact on image matching and identification. Because the body images from the same cow in this study had little change of translation, rotation and scaling, the coordinate offset D_i between matched points were relatively small when the feature points were correctly

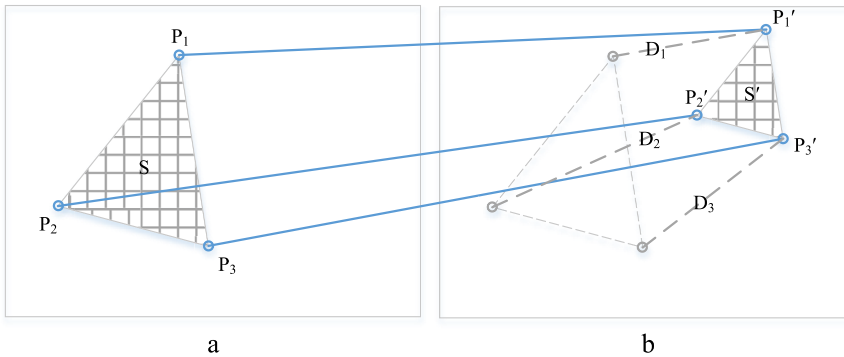


Fig. 3 – Unexpected matching of similar shapes.

matched. Therefore, a threshold *DT* was determined to filter out wrong matching. If the relative displacement D_i was smaller than *DT*, P_i' can be considered as a correct match to P_i. Otherwise, P_i' was filtered from the matching result.

Because of the influence of the shooting angle, the distance between cows and camera and the walking posture of cows, the images of the same cow were slightly shifted, rotated and scaled when captured at a different time. Therefore, when the value of *DT* was too small, the correct matching result may be discarded. Conversely, when the value of *DT* was too large, the wrong matching pairs cannot be filtered out successfully. Through theoretical analysis and experimental verification, when *DT* was taken as 100 pixels, the effective matching point pairs can be retained to the maximum extent on the basis of filtering most error matches.

2.6. *Identification software*

Software was developed on PC to implement the model described above. The processor of the PC was Inter i5-4210U, with main frequency of 2.39 GHz, 8 GB memory, and 256 GB SSD. The algorithm development platforms were Visual Studio 2013 and OpenCV 2.3.9. The flow chart of the software is shown in Fig. 4. Before the loop, the dataset update module calculated features of template images on the hard drive and stored them in memory. During the loop, the program checked the position of the cow and only processed the frame when the cow was in the middle area of the field. Therefore, only one image was processed for identification for each video. Before running the software, the template dataset should be prepared. If no images in the template dataset were matched with the current frame, then the software entered the enrolment stage.

After obtaining the body image, the centroid coordinate of the body image was calculated. The offset between the x coordinate and longitudinal symmetry line of the entire image was compared. When the absolute value of the offset was less than 10% of the width of the vision field, the cow was considered to be in the middle region of the image. Therefore, the following processing can be continued as long as the deviation was within the above range.

The interface of the software is shown in Fig. 5. The software had three areas, which were bottom area, showing area and message area. The software supported manual operation and fully automatic operation (auto). In auto mode, the software processed every frame it read from a camera or file, and stored the result with the body image for human inspection. To use this software, it was necessary to keep the background static and ensure enough illumination. This software has already been put into use in a commercial farm in Yinchuan, Ningxia.

3. Results and discussions

3.1. *Detecting and matching feature points*

Figure 6 shows the identification process of an unknown image using FAST + SIFT + FLANN method. Firstly, FAST was

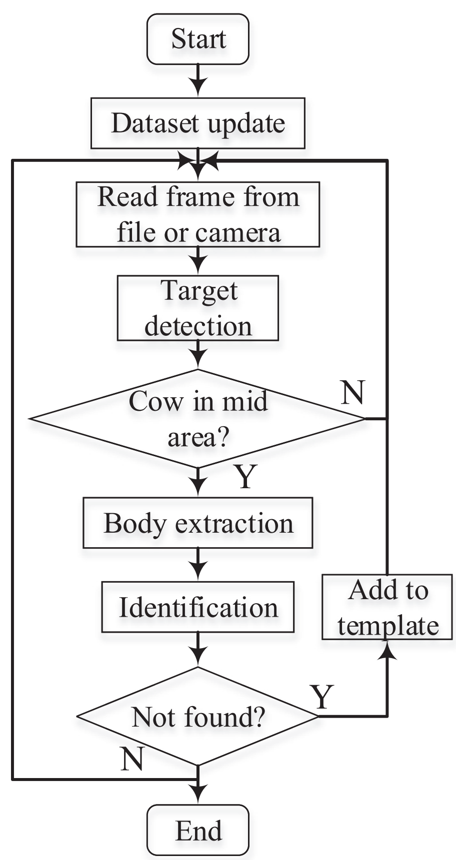


Fig. 4 – Flow chart of cow identification software.

Fig. 5 – Interface of cow identification software.

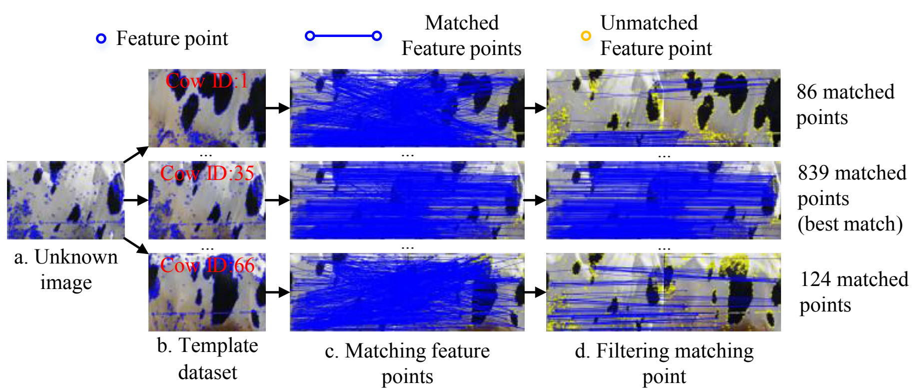


Fig. 6 – Extracting, matching and filtering feature points.

used to detect the feature points in both unknown image (Cow_u) and image in reference dataset (Cow_r). The detection results are shown in Fig. 6a and b, where the points detected can be divided into correct feature points and interference (unexpected) points. The correct feature points were mainly located at the edge of the black and white spot of the cow. The interference points located on the contamination on cow's body, the edges of the electronic fence, and highlight region in black spot. After the feature points were detected, the SIFT operators of all feature points in Cow_u and Cow_r were calculated. For each feature point in Cow_u, the FLANN algorithm was used to find the best match among all the feature points in Cow_r, and the matching result was connected with blue solid line. Each unknown image was matched with all the images in the reference dataset, as shown in Fig. 6c. Figure 6d shows the result of filtering matches using offset thresholds. The results show that the filtering algorithm can filter out most of the error matching effectively and retain the correct matching to a large extent. For each image Cow_r in reference dataset, if the number of matched points between Cow_r and Cow_u was the maximum among all images in reference dataset, the ID of Cow_r was assigned to Cow_u.

3.2. *Identification result*

Figure 7 shows the identification results using three algorithms individually to detect and match feature points. The identification accuracies of SIFT and SURF were 91.39% and 91.46%, respectively. However, the detection efficiency of SURF method was obviously higher than that of SIFT. In comparison with SIFT, the loading time of the reference dataset was reduced by about 50%, and the average identification time for each image was reduced from 16.19 s to the 10.6 s. Among the three algorithms, ORB had the highest identification accuracy and efficiency. The overall detection accuracy was 95.41% and the loading time was less than 5 s. In particular, the time of identification for a single frame was only 0.77 s, which was much lower than the other algorithms and ensures real time identification. The ORB algorithm had highest detection performance as 366 feature points were detected on average for each image.

Figure 8 shows the identification results using different combinations of algorithms. The number of extracted points in the graph shows that FAST has highest performance for feature detection. When it was combined with SIFT and ORB, the numbers of feature points detected were 793 and 490,

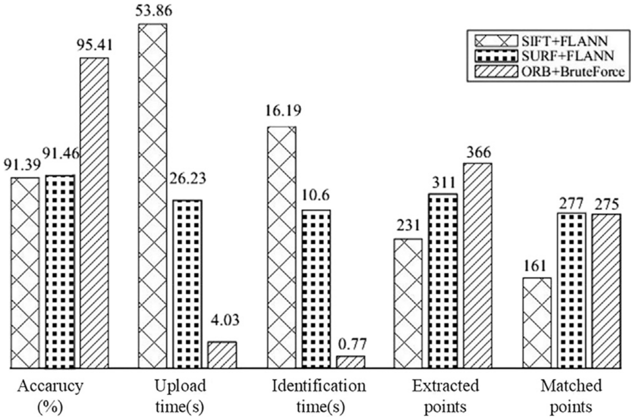


Fig. 7 – Identification results using three algorithms separately.

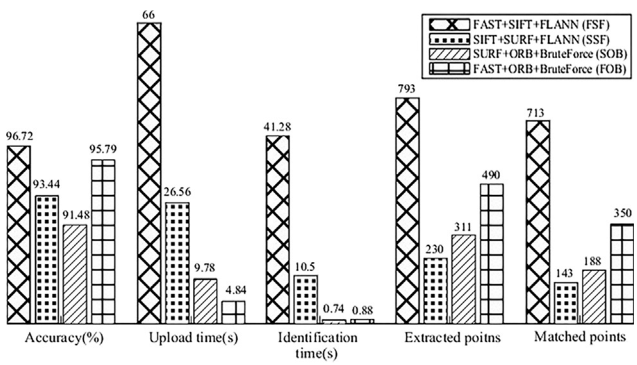


Fig. 8 – Individual identification results based on different combinations of algorithms.

respectively. Those two combinations also achieved higher identification accuracy, at 96.72% and 95.79%, respectively. Due to the calculation complexity of the SIFT descriptor, the loading time and the identification time of the FSF were longest among all combinations of algorithms. In the FOB combination, the computational efficiency was greatly improved as ORB was a binary operator. However, because the FAST algorithm detected more feature points than the ORB algorithm did, the average identification time of FOB was increased to 0.88 s. The combination of SIFT and SURF slightly improved identification accuracy and efficiency compared with using SIFT alone, which benefited from optimisation and improvement of the SURF algorithm. The identification accuracy of SOB combination was relatively lower than others. In summary, the combination of FAST algorithm with other algorithms can improve the identification accuracy. However, because more feature points were detected, the identification time was increased to varying degrees.

3.3. *Wrong matching*

Figure 9 shows one case of error matching for cows whose body images are similar. As the area of cows' body was close to pure black, it was difficult to distinguish them from each other and failure in identification occurred.

Because of the similarity of body images between those cows, it was difficult to discriminate them correctly even by

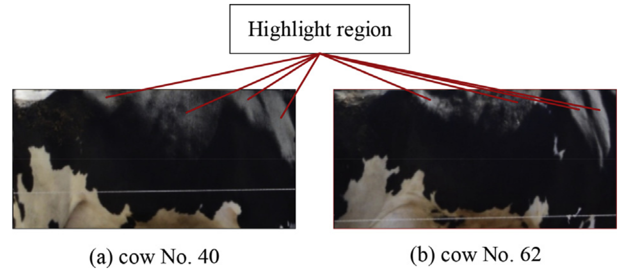


Fig. 9 – Error matching of cows with similar pattern and highlight regions.

human observation. When the images were influenced by light conditions, the similar cows were incorrectly identified with higher probability. The cows in Fig. 9 had a large black area, and the position and shape of the white spot were similar. Normally, the feature points only exist at the boundary line between black and white areas, rather than inside the solid colour region. However, due to large solar inclination, highlights appeared on each cow's body, which led to a large number of unexpected feature points at the edge of the highlights region, and resulted in error matching.

Figure 10 shows the influence of the highlight region on feature point detection when using different algorithms. The detection results showed that the sequence of algorithms according to descending number of unexpected feature points generated in the highlight region were: FAST, SURF, SIFT, and ORB. In the detection results of FAST and SURF algorithms, there were many feature points in the edge and interior of the highlight region. These unexpected feature points were mismatched frequently with the correct feature points at the edges of the black and white spot in the body image. Therefore, a large number of mismatches were produced, which affected the result of individual identification. The SIFT algorithm showed high performance to resist the highlight area. However, it was sensitive to the change of grey level in the white body spot, and some unexpected feature points were

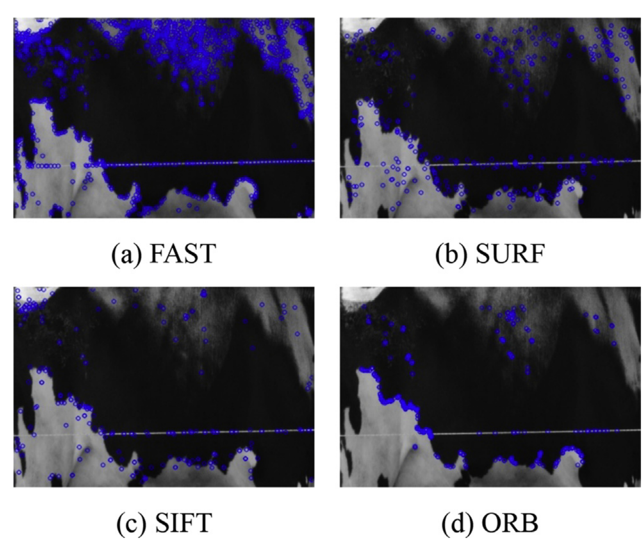


Fig. 10 – Influence of highlight region on feature point detection using different algorithms.

generated in the white body spot. The ORB algorithm was insensitive to the grey change of the white spot, but its ability to resist the highlight was lower than that of the SIFT algorithm. In addition to producing highlights, illumination also had great impact on the exposure of the image. Using histogram equalisation can reduce the influence of exposure under different illumination conditions.

To check the matching ranking of unknown images with images in the reference dataset, 66 cows were indexed from 1st to 66th. The number of matched feature points between the query image and ith in reference library was taken as $N_{q\text{-}i}$ ($i = 1$ to 66) which were then sorted in descending order and taken as $\mathbf{Q_N} = \{Qi, i = 1,2,3, \ldots, 66\}$. Assume real ID of the query image was c, then index of $N_{q\text{-}c}$ in $\mathbf{Q_N}$ was taken as I_r. Theoretically $N_{q\text{-}c}$ is the maximum value and in the first place in $\mathbf{Q_N}$ ($I_r = 1$). However, when a wrong match occurred, the index of $N_{q\text{-}c}$ was no longer the maximum of $\mathbf{Q_N}$. In the case that the real ID of the query image is 40, the sorted $N_{40\text{-}i}$ is listed in Fig. 11. The results showed that the best match of the query image was cow No. 62, followed by cow No. 40. The numbers of matched feature points were very close.

3.4. *Feature point selection*

Figure 12 shows the results of filtering matching point using different methods. The original matching results showed that there were many errors matching the body images from the same cow. When the ratio value was 0.8, the matching results were accurate and correct, but too many correct matching points were filtered out. When the displacement threshold method was used to filter, it can filter out the wrong matching points and retain the correct matching results as much as possible, which can meet the matching requirements in this study.

The filtering method proposed by Lowe compared the best matching of feature points with the second best matching, and accepted the matching result if the degree of difference between the two matches was large enough. Otherwise, the result was discarded. In this study, only black and white colours were included in the body images of dairy cows, and most of the feature points were distributed at the edges of the black and white areas. Therefore, the difference of descriptors among feature points was minor. In this case, most matching point pairs were filtered out by using the Lowe filtering method. In this study, the position deviation of correct paired points in two images was limited. By using the offset between matched points as the basis for filtering the matching results, the algorithm proposed in this paper can retain the correct matched point pairs to the maximum extent and filtered out the unexpected matching effectively. Therefore, filtering results in this paper were better than when using Lowe's method.

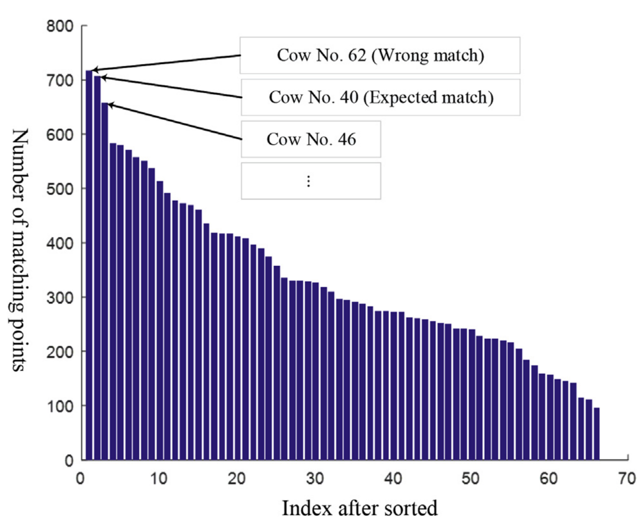
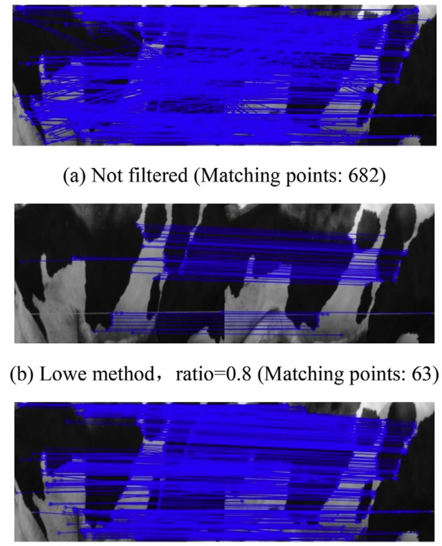


Fig. 11 – Number of matched potions between cow No. 40 and other cows (using FAST + SIFT + FLANN).



(a) Not filtered (Matching points: 682)

(b) Lowe method，ratio=0.8 (Matching points: 63)

(c) Displacement threshold method，r<=100 pixels (Matching points: 473)

Fig. 12 – Comparison of filtering method proposed in this paper with Lowe's method.

3.5. *Discussion*

3.5.1. *Comparison of features*

Feature-based image matching was divided into two steps: feature point extraction and feature point matching. This paper mainly compared three feature descriptors, namely SIFT, SURF and ORB. SURF was basically an update version of SIFT. SURF was always a better choice comparing with SIFT. ORB had the advantage of less computing time. Table 1 shows the comparison of performance of four methods.

In the introduction to SURF above, it was suggested that SURF was a similar feature descriptor to SIFT. Compared with SIFT, SURF used square filter and determinant value of Hessian matrix to detect extreme values, and used integral graph to accelerate calculation. The feature descriptor was a vector of $16 \times 4 = 64$ dimension. Therefore, the speed of feature point detection was greatly improved. In the field of computer vision, the overall performance of the ORB algorithm is better than other feature extraction algorithms. As an improved algorithm of BRIEF, ORB is invariant to rotation and not sensitive to noise, but ORB is not invariant to scale. In the application of cow's individual identification, because the dairy cows had

Table 1 – Comparison of performance of three feature descriptors.

Indicators	Ranking by performance
Number of feature points	ORB > SURF > SIFT
Computational efficiency	ORB > SURF >> SIFT
Rotation robustness	SURF > ORB ~ SIFT (approximately)
Fuzzy robustness	SURF > ORB ~ SIFT
Scale transformation robustness	SURF > SIFT > ORB

little transformation in scale and rotation during walking, and the image clarity was good, the ORB method can achieve a better comprehensive performance. As it has high processing efficiency, the ORB algorithm is suitable for the scene that requires good real-time performance.

3.5.2. *Improvement measurement*

As shown in Fig. 11, the first three values of matched points were significantly closer than the others. Therefore, the cows numbered 62, 40 and 46 were all similar to the query cow, so they can be regarded as one group. Therefore, for similar cows, several references were selected from the matching results, and then the final decision was made by a second manual checking to improve the identification accuracy. The cows with the matching numbers close to the maximum value in the list were selected using Eq. (1).

$$G = \left\{ i \in Z \left| \frac{|N_{q-i} - N_{\max}|}{N_{\max}} < Td \right. \right\}, \quad Z = \{1, 2, \ldots, 66\} \tag{1}$$

where $N_{\max}$ is the maximum in the list, and Td is the threshold for selection.

However, the redundancy of identification result increased, and one additional step was needed to decide the query cow manually when there was more than one element in the group. If there was only one element in the group, that one was decided as the identification result directly. Table 2 shows the result of OOB (ORB + ORB + BruteForce) when Td was 0.1 using two-step identification. In the two-step matching method, the result of each match was no longer a unique output, but a group of IDs. This group may or may not contain the correct ID of the query cow. Table 2 illustrates that proportion of groups only containing one element is 87.54% out of all groups, and 86.56% of the results is actually the correct ID of the query cow. The proportion of groups having two elements and three or more elements were 6.56% and 5.9%, respectively. Among all identification results, 98.36% of the groups contained the correct ID. Therefore, it can be considered that the identification accuracy of the two-step method can be improved to 98.36% using two-step identification with little manual participation. The average manual check frequency was 16.07 per 100 images.

3.5.3. *Future study*

Through the evaluation and test of various algorithms for feature point detection and matching, it was found that the highest number of feature points detected was 793, of which 713 were correctly matched. The new algorithms only improved the efficiency of detection and matching. This indicated that the detection and matching rate of feature points had already reached a very high level, and it was difficult to further improve the accuracy of individual identification of dairy cows through point matching. Therefore, future study should focus on the methods for improving image quality and developing speckle features, e.g. highlight removal, binarisation and contour matching, to future improve the identification accuracy.

3.5.3.1*Highlight removal*. It was found that when the pure black area of dairy cattle was large, it was common to produce a highlight area. A large number of pseudo feature points were generated at the edge of these regions, resulting in identification errors. Therefore, the removal or elimination of the highlight region in the image can effectively reduce the pseudo feature points and further improve the identification accuracy. Yang, Tang, and Ahuja (2015) proposed a highlight removal method using a single input image. This method estimated the maximum diffuse chromaticity values of the specular pixels and propagated the maximum diffuse chromaticity values from the diffuse pixels to the specular pixels. Shen and Zheng (2013) proposed an efficient method to separate the diffuse and specular reflection components from a single image. This method was developed based on the principle that the intensity ratios between the maximum values and range values were independent of surface geometry. The algorithm was simple and easy to implement, but it was not suitable for the elimination of large region highlights. Fan, Yang, Hua, Chen, and Wipf (2017) proposed a deep neural network structure that exploits edge information in addressing layer separation and image filtering. Two CNN networks, E-CNN and I-CNN, were used for edge prediction and image reconstruction, respectively. This deep architecture may have great potential to remove highlight region in cows' images captured under natural conditions.

3.5.3.2*Binarisation*. Holstein cow body spot only contains black and white colours. But due to changes of light and white balance, the colour of the image may deviate. These deviations were also an important source of mismatch. Therefore, binarisation of the image can enhance the appearance of cow body spot to the maximum extent. Adaptive binary processing is generally aimed at the case where only white and black colours are included in the image. This situation is similar to the ORC process in scanned documents. The OTSU method is an adaptive threshold method and is widely used in binary processing of scanned document grayscale images

Table 2 – Result of two-step identification for the combination of ORB + ORB + BruteForce.

Group categories	1 element	2 elements	≥3 elements	Sum
Contains correct match	86.56%	6.23%	5.57%	98.36%
Not contains correct match	0.98%	0.33%	0.33%	1.64%
Sum	87.54%	6.56%	5.9%	100%

(Galar, Fernandez, Barrenechea, Bustince, & Herrera, 2011). In recent years, researchers have studied the use of deep learning for the binarisation of images, e.g. CNNs (Saidane & Garcia, 2007), RNN (Westphal, Lavesson, & Grahn, 2018), FCN (Barakat & El-Sana, 2018). In the process of image binarisation using deep neural network, the traditional binary threshold segmentation become a pattern recognition problem. By supervised learning, the white and black areas can be distinguished accurately.

3.5.3.3*Contour matching*. The SIFT method can find the feature points at the edge between black and white regions. However, the dimensions of feature points were high, which affected the matching efficiency. When the database is very large, it is difficult to match in real time. When the cow image is converted to a binary one, the edges of the white spot in the image form a specific contour. These contours rotate little because the cows move their bodies horizontally as they walk. Therefore, it is feasible to match the contour of spots. This requires that the matching algorithm is scale-invariant. The contour matching algorithm can be divided into three steps (Chatbri, Kameyama, & Kwan, 2016): ***S1.*** Zoom the two contours (generally point set or binary image) to the same scale, ***S2.*** Calculate the contour description of the two contours. ***S3.*** Calculate the distance of the two groups of contours using Euclidean distance, earth mover's distance (Grauman & Darrell, 2004), Hausdorff distance (Huttenlocher, Klanderman, & Rucklidge, 1993), etc.

3.5.3.4*Farm use*. In this study, cow's images from the right side were used for individual identification. In practice, some cows may have very similar body patterns on one side but different on the other side, especially in large groups. Therefore, bilateral side-view images could be collected and merged into a complete unfolded cow body image, which could further improve the accuracy of the system when used on larger farms. In addition, for dairy farms where side-view image systems are difficult to install, top view images could be another option. Obviously, the algorithm proposed in this paper can be simply adjusted to fit the top view image system. At present, the machine vision system in this paper can only be used to identify dairy cow breeds with a body pattern (Holstein, Guernsey, Ayrshire, etc.). For other breeds without speckled information, a logo or ID number can be artificially painted the animal's body, or even a QR code. In fact, there are some dairy farms that already do this. The dairy farm in this study is a research dairy farm with only 66 lactating cows. The accuracy and efficiency of the algorithm should be tested in a larger herd in the future.

4. Conclusion

In this paper, a computer vision system for dairy cow individual identification was proposed. The system detected the object cow and located the body area as the individual identity information using the videos of walking cows in side view. By creating a template database, unknown images were matched and compared to determine their ID. Results showed that the SIFT method can accurately calculate the feature points in the body pattern. The feature point selection method proposed in this paper can not only filter the error matching, but also retain the effective matching to the maximum extent. When FAST, SIFT and FLANN were used for point detector, extractor and matcher, the accuracy of one-step identification was 96.72%. The combination of ORB and BruteForce methods had higher matching efficiency, but the identification accuracy dropped to 95.41%. If a two-step match method was used, the probability of the group containing the correct match increased to 98.36%. This study demonstrated that the method of image matching was feasible for individual identification of Holstein cows. In the system, the template dataset can be updated easily and quickly, which improved the practicability of vision system for individual identification in farm. In order to improve the accuracy and efficiency of the system, methods for extracting the binary information of cow body pattern are worthy of study in the future.

Acknowledgements

The authors thank the Coldstream farm, University of Kentucky, USA for their cooperation in the project. The authors also thank Dr. Jeffrey Bewley for building the experiment setup on the farm and providing the access to the farm. The work was sponsored by the Key R&D and Promotion Projects in Henan Province (Science and Technology Development, No. 192102110089), National Key Research and Development Program of China Sub-project (No. 2017YFD0700800), the Innovation Scientists and Technicians Troop Construction Projects of Henan Province (No. 184200510017), and Open Funding Project of Key Laboratory of Agricultural Internet of Things (No. 2018AIOT-07).

REFERENCES

Ahrendt, P., Gregersen, T., & Karstoft, H. (2011). Development of a real-time computer vision system for tracking loose-housed pigs. *Computers and Electronics in Agriculture, 76*(2), 169–174.
Arcidiacono, C., Porto, S. M., Mancino, M., & Cascone, G. (2017). A threshold-based algorithm for the development of inertial sensor-based systems to perform real-time cow step counting in free-stall barns. *Biosystems Engineering, 153*, 99–109.
Barakat, B. K., & El-Sana, J. (2018). Binarization free layout analysis for Arabic historical documents using fully convolutional networks. In *2018 IEEE 2nd international workshop on Arabic and derived script analysis and recognition (ASAR)*.
Barnich, O., & Van Droogenbroeck, M. (2011). ViBe: A universal background subtraction algorithm for video sequences. *IEEE Transactions on Image Processing, 20*(6), 1709–1724.
Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). *Computer Vision and Image Understanding, 110*(3), 346–359.
Bay, H., Tuytelaars, T., & Van Gool, L. (2006). Surf: Speeded up robust features. In *European conference on computer vision*.
Cai, C., & Li, J. (2013). Cattle face recognition using local binary pattern descriptor. In *Signal and information processing association annual summit and conference (APSIPA), 2013 Asia-Pacific, Kaohsiung, Taiwan*.
Cappai, M. G., Rubiu, N. G., Nieddu, G., Bitti, M., & Pinna, W. (2018). Analysis of fieldwork activities during milk production recording in dairy ewes by means of individual ear tag (ET)

alone or plus RFID based electronic identification (EID). *Computers and Electronics in Agriculture, 144*, 324–328.
Chapinal, N., & Tucker, C. B. (2012). Validation of an automated method to count steps while cows stand on a weighing platform and its application as a measure to detect lameness. *Journal of Dairy Science*, 95(11), 6523–6528. https://doi.org/10.3168/jds.2012-5742.
Chatbri, H., Kameyama, K., & Kwan, P. (2016). A comparative study using contours and skeletons as shape representations for binary image matching. *Pattern Recognition Letters, 76*, 59–66.
Chum, O., Philbin, J., & Zisserman, A. (2008). Near duplicate image detection: min-Hash and tf-idf weighting. In *, 810. British machine vision conference, BMVC 2008* (pp. 812–815).
Dutta, R., Smith, D., Rawnsley, R., Bishop-Hurley, G., Hills, J., Timms, G., et al. (2015). Dynamic cattle behavioural classification using supervised ensemble classifiers. *Computers and Electronics in Agriculture, 111*, 18–28.
Fan, Q. N., Yang, J. L., Hua, G., Chen, B. Q., & Wipf, D. (2017). A generic deep architecture for single image reflection removal and image smoothing. In *IEEE international conference on computer vision, 2017* (pp. 3238–3247).
Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., & Herrera, F. (2011). An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. *Pattern Recognition*, 44(8), 1761–1776. https://doi.org/10.1016/j.patcog.2011.01.017.
Gao, X., & Wang, Y. (2014). Recognition of similar handwritten Chinese characters based on CNN and random elastic deformation. *Journal of South China University of Technology (Nat Sci Ed), 42*(1), 72–76.
Grauman, K., & Darrell, T. (2004). Fast contour matching using approximate earth mover's distance. In *, vol. 1. Proceedings of the 2004 IEEE computer society conference on computer vision and pattern recognition, CVPR 2004* (p. I-I).
Hoffmann, G., Schmidt, M., Ammon, C., Rose-Meierhofer, S., Burfeind, O., Heuwieser, W., et al. (2013). Monitoring the body temperature of cows and calves using video recordings from an infrared thermography camera. *Veterinary Research Communications, 37*(2), 91–99.
Huttenlocher, D. P., Klanderman, G. A., & Rucklidge, W. J. (1993). Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 15*(9), 850–863.
Jover, J. N., Alcaniz-Raya, M., Gomez, V., Balasch, S., Moreno, J. R., Colomer, V. G., et al. (2009). An automatic colour-based computer vision algorithm for tracking the position of piglets. *Spanish Journal of Agricultural Research, 7*(3), 535–549.
Kim, H. T., Choi, H. L., Lee, D. W., & Yoon, Y. C. (2005). Recognition of individual Holstein cattle by imaging body patterns. *Asian-Australasian Journal of Animal Sciences, 18*(8), 1194–1198.
Li, W., Ji, Z., Wang, L., Sun, C., & Yang, X. (2017). Automatic individual identification of Holstein dairy cows using tailhead images. *Computers and Electronics in Agriculture, 142*, 622–631.
Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision, ICCV 1999* (p. 1150).
Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 60*(2), 91–110.
Maddalena, L., & Petrosino, A. (2008). A self-organizing approach to background subtraction for visual surveillance applications. *IEEE Transactions on Image Processing, 17*(7), 1168.
Matthews, S. G., Miller, A. L., Clapp, J., Plötz, T., & Kyriazakis, I. (2016). Early detection of health and welfare compromises through automated detection of behavioural changes in pigs. *The Veterinary Journal, 217*, 43–51.
Muja, M., & Lowe, D. G. (2009). Fast approximate nearest neighbors with automatic algorithm configuration. In *, vol. 2. The international conference on computer vision theory and applications, VISAPP 2009* (pp. 331–340).
Nasirahmadi, A., Edwards, S. A., & Sturm, B. (2017). Implementation of machine vision for detecting behaviour of cattle and pigs. *Livestock Science, 202*, 25–38.
Pang, C., He, D., Li, C., Huang, C., & Zheng, L. (2011). Method of traceability information acquisition and transmission for dairy cattle based on integrating of RFID and WSN. *Transactions of the Chinese Society of Agricultural Engineering, 27*(9).
Rosten, E., & Drummond, T. (2006). Machine learning for high-speed corner detection. In *European conference on computer vision, 2006* (pp. 430–443).
Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. In *IEEE international conference on computer vision, ICCV 2011* (pp. 2564–2571).
Saidane, Z., & Garcia, C. (2007). Robust binarization for video text recognition. In *, vol. 2. Ninth international conference on document analysis and recognition, ICDAR 2007* (pp. 874–879).
Samad, A., Murdeshwar, P., & Hameed, Z. (2010). High-credibility RFID-based animal data recording system suitable for small-holding rural dairy farmers. *Computers and Electronics in Agriculture, 73*(2), 213–218.
Shen, H., & Zheng, Z. (2013). Real-time highlight removal using intensity ratio. *Applied Optics*, 52(19), 4483–4493. https://doi.org/10.1364/AO.52.004483.
Sobral, A. (2013). BGSLibrary: An OpenCV C++ background subtraction library. In *IX Workshop de Visão Computacional, WVC 2013 (p. 7), 2*(6).
Voulodimos, A. S., Patrikakis, C. Z., Sideridis, A. B., Ntafis, V. A., & Xylouri, E. M. (2010). A complete farm management system based on animal identification using RFID technology. *Computers and Electronics in Agriculture*, 70(2, SI), 380–388. https://doi.org/10.1016/j.compag.2009.07.009.
Westphal, F., Lavesson, N., & Grahn, H. (2018). Document image binarization using recurrent neural networks. In *13th IAPR international workshop on document analysis systems, DAS 2018* (pp. 263–268).
Xia, M., & Cai, C. (2012). Cattle face recognition using sparse representation classifier. *ICIC Express Letters, Part B: Applications, 3*(6), 1499–1505.
Xiong, B., Qian, P., Luo, Q., & Lu, J. (2005). Design and realization of solution to precision feeding of dairy cattle based on single body status. *Transaction of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 21*(10), 118–123.
Yajuvendra, S., Lathwal, S. S., Rajput, N., Raja, T. V., Gupta, A. K., Mohanty, T. K., et al. (2013). Effective and accurate discrimination of individual dairy cattle through acoustic sensing. *Applied Animal Behaviour Science*, 146(1–4), 11–18. https://doi.org/10.1016/j.applanim.2013.03.008.
Yang, Q., Tang, J., & Ahuja, N. (2015). Efficient and robust specular highlight removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(6), 1304–1311. https://doi.org/10.1109/TPAMI.2014.2360402.
Yu, Y., Yin, G., Yin, Y., & Du, L. (2014). Defect recognition for radiographic image based on deep learning network. *Chinese Journal of Scientific Instrument, 35*(9), 2012–2019.
Zhao, K., Bewley, J. M., He, D., & Jin, X. (2018). Automatic lameness detection in dairy cattle based on leg swing analysis with an image processing technique. *Computers and Electronics in Agriculture, 148*, 226–236.
Zhao, K., & He, D. (2015). Recognition of individual dairy cattle based on convolutional neural networks. *Transactions of the Chinese Society of Agricultural Engineering, 31*(05), 181–187.
Zheng, C., Zhu, X., Yang, X., Wang, L., Tu, S., & Xue, Y. (2018). Automatic recognition of lactating sow postures from depth images by deep learning detector. *Computers and Electronics in Agriculture, 147*, 51–63.
Zivkovic, Z. (2004). Improved adaptive Gaussian mixture model for background subtraction. In *Proceedings of the 17th international conference on pattern recognition, ICPR 2004* (pp. 28–31).