Real-Time Cattle Action Recognition for Estrus Detection

Eui-Ju Heo, Sung-Jin Ahn, and Kang-Sun Choi*
Interdisciplinary Program on Creative Engineering, KOREATECH
Cheonan, 31253 – South Korea
[e-mail: lobelia@koreatech.ac.kr]
[e-mail: dkstjdwls199@koreatech.ac.kr]
[e-mail: ks.choi@koreatech.ac.kr]
*Corresponding author: Kang-Sun Choi

Received June 18, 2018; revised August 29, 2018; accepted September 26, 2018; published April 30, 2019

Abstract

In this paper, we present a real-time cattle action recognition algorithm to detect the estrus phase of cattle from a live video stream. In order to classify cattle movement, specifically, to detect the mounting action, the most observable sign of the estrus phase, a simple yet effective feature description exploiting motion history images (MHI) is designed. By learning the proposed features using the support vector machine framework, various representative cattle actions, such as mounting, walking, tail wagging, and foot stamping, can be recognized robustly in complex scenes. Thanks to low complexity of the proposed action recognition algorithm, multiple cattle in three enclosures can be monitored simultaneously using a single fisheye camera. Through extensive experiments with real video streams, we confirmed that the proposed algorithm outperforms a conventional human action recognition algorithm by 18% in terms of recognition accuracy even with much smaller dimensional feature description.

Keywords: Action Recognition, Cow Behavior, Surveillance, Monitoring Cattle, SVM

A preliminary version of this paper appeared in IEEE ICCE 2018, Jan. 12, Las Vegas, USA. This version includes an improvement on the proposed features and extensive experimental results. This work was supported by the BK21 plus program through the National Research Foundation (NRF) funded by the Ministry of Education of Korea.

http://doi.org/10.3837/tiis.2019.04.023 **ISSN : 1976-7277**

1. Introduction

On livestock farms, it is important to monitor the health and anomalies of the livestock in order to reduce costs and welfare impacts [1]. One of the measures to assess animal health is movement. Unhealthy animals suffering from limb disorders or disease are much less active than healthy animals. Animal movement also indicates estrus. The most observable sign of the estrus phase is the mounting action, where a cow in estrus follows another female cow for a few seconds, and then mounts the cow. Estrus detection in cattle helps determine the optimal time for artificial insemination, which leads to increasing not only conception rates for the herd but also milk production. In this paper, we propose a machine vision–based cattle action recognition algorithm to detect mounting events during estrus and present a real-time cattle monitoring system exploiting the algorithm.

There have been several attempts to recognize the movement of animals [2-7]. Radio frequency transmitters have been used to determine presence and duration at feeding and watering locations, with this data being correlated to health status [2-4]. Data loggers with accelerometers have been used to analyze postural behavior patterns of cattle [6].

In order to detect lameness in cows, wireless 3D accelerometers were used to measure temporal gait characteristics on all four limbs of the cows [8], and accelerometer data from cow collar sensors were collected by clustering them in accordance with activity level. Then, mounting events were identified using a change-detection technique on the high-activity index derived from clustered time series data.

However, attachment of sensors may cause stress and, in some cases, is impractical to use due to their cost and vulnerability. Automatic computer vision systems can be an effective alternative to avoid such problems in monitoring animals [9].

Poursaberi *et al*. [10] evaluated the status of lameness in an individual cow by extracting the arc of the back from side-view videos during standing and walking. Kashiha *et al*. [11] quantified pig locomotion associated with lameness with top-view videos. Specifically, foreground objects representing pigs were extracted using image processing, and then each pig was located by fitting each object into an ellipse model. In order to extract pigs more accurately in an unevenly illuminated pigpen, Guo *et al*. [12] applied adaptive partitioning and multilevel thresholding. By extracting object contour information and utilizing it, the method can observe several objects at the same time, but the contour information becomes severely degraded in complex backgrounds.

Recently, automatic estrus detection requiring more complicated motion recognition has been investigated [9], [13], [14].

Nasirahmadi *et al*. [9] presented a mounting-detection algorithm for pigs, where moving objects are initially fitted into ellipses, and then, the major and minor axis length ratio of the fitted ellipse is utilized to determine mounting events. However, ellipse fitting is not suitable for cow localization because cattle usually move their heads to a much larger extent than pigs, as shown in **Fig. 1**. In addition, not only adult cows but also young calves live in the barn. As a result, the major and minor axis lengths can be estimated inaccurately, and thus mounting detection can fail.

Chung *et al*. [13] presented a cattle mounting detection system in which the height of the moving object is examined with side-view videos, since the mounting action leads to a sudden increase in object height. However, there are usually many livestock in a single barn, which leads to complex scenes in which animals are frequently occluded by each other and move

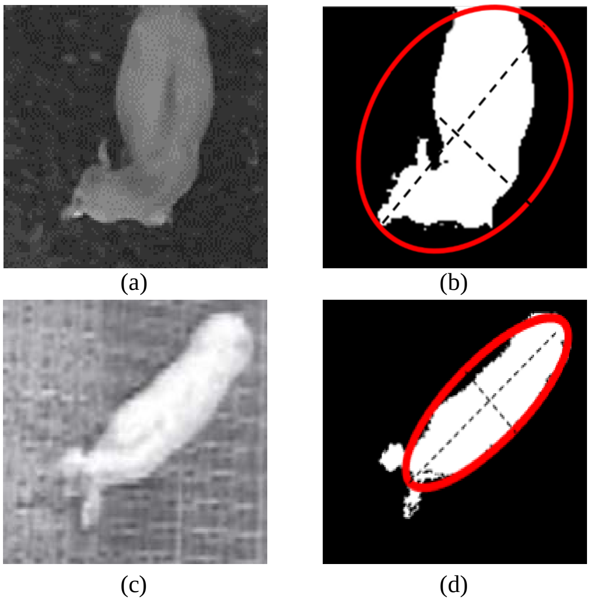

Fig. 1. Feasibility of ellipse fitting for cows and pigs. (a) A cow image, and (b) ellipse fitting result for (a). (c) A pig image, and (d) ellipse fitting result for (c).

around continuously. Despite easy mounting detection when using side-view videos, object size varies largely depending on distance to the camera and the cow closest to the camera can occlude the mounting event. In addition, side-view videos can be affected by direct sunlight and are easily degraded.

Tsai and Huang [14] detected estrus behavior of Holstein cattle having black-and-white markings based on the most observable sign of estrus, where a cow follows another cow for a few seconds, and then mounts the cow. Using motion detection and morphological operations, moving objects were extracted from top-view videos. Then, the mounting event is detected based on the changes of moving object lengths, i.e. the extracted moving object is much longer, compared to the length of a normal cow during the following behavior, and then, the moving object shrinks suddenly during the mounting event. However, foreground object extraction is easily affected by changes, such as sudden light changes, dynamic backgrounds, ground urine stains. That is why the method requires many parameters to be tuned. Furthermore, if the object of interest in the herd includes taurine cattle having no special patterns on the bodies, simple image processing tends to fail to extract foreground object.

In order to deal with these problems, motion recognition for a practical estrus detection system needs to be more robust to dynamic complex scenes while keeping computational complexity low.

As computing power increases exponentially, sophisticated computer vision–based motion recognition, mainly for human movement, has been researched and developed [15-23].

Among them, a robust representation of movements, called the motion history image (MHI), has been widely used to describe motion information within a video. The MHI can represent not only the location of object motion generated in the scene but also the process of how the object moves, and with very low complexity [15]. Furthermore, it is suitable for expressing human posture, gait, and gestures, because it is not sensitive to silhouette noises, such as holes, shadows, and missing parts [17].

In this paper, we present a vision-based cattle monitoring system for estrus detection as shown in **Fig. 2**. Using a top-down fisheye camera, cattle in the three enclosures are monitored in real-time. If any cattle action of interest is recognized, the system stores a short video clip on the server and notifies the person in charge of the event by sending a message. Then, the person can access the video clip to check the event.

In order to recognize cattle actions, we propose novel MHI-based features that can effectively describe cattle movements, especially mounting action. Since mounting action occurs during the rare estrus period, it is difficult to acquire a number of training dataset. Through learning the proposed features with the support vector machine (SVM) framework that is well suited for small or medium sized dataset, various cattle actions, such as mounting, walking, tail wagging, and foot stamping, can be recognized in complex scenes thanks to MHI-inherited robust representation and silhouette noise insensitivity.

Furthermore, since the proposed cattle action recognition algorithm is very efficient, a visual cattle monitoring system using it can recognize the actions of cattle in three enclosures simultaneously, which reduces the system cost.

The main contributions of this paper can be summarized as follows:

- We propose an effective feature description for cattle action recognition, which is robust to complex scenes. Compared with a conventional action recognition method using the same MHI, the proposed feature description improves the recognition accuracy significantly.
- Since the proposed feature description has a sufficiently small dimension, multiple cattle can be monitored simultaneously in real time, which leads to the system cost reduction.

To the best of our knowledge, the proposed algorithm is the first vision-based action classification for livestock.

In the following section, before presenting the proposed features, MHI (the intermediate

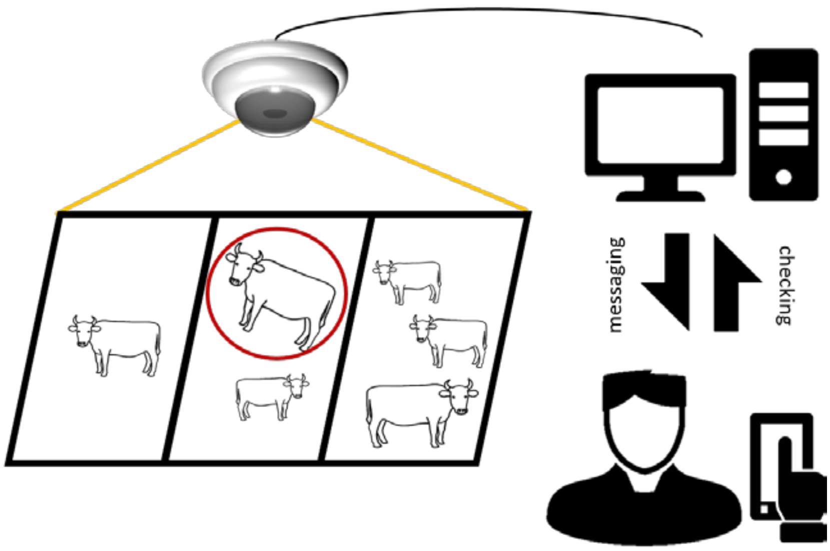

Fig. 2. The configuration of the proposed cattle monitoring system. Cattle in the three enclosures are monitored using a fisheye camera. When interesting actions including mounting action are recognized, a short video clip is stored in the server, and a message is sent to the person in charge for checking the events.

information from which the proposed features are extracted) is first investigated. In Section 3, we present the proposed cattle action recognition algorithm in detail. Through extensive experiments, the proposed algorithm is compared to conventional motion recognition algorithms in Section 4. Finally, concluding remarks are given in Section 5.

2. Motion History Image

Thanks to both robust representation and the simplicity of MHI, several features based on MHI have been proposed to effectively represent characteristics of motion [18-20], [22]. Bradski and Davis [19] applied Hu moments [24] to the gradient of the MHI silhouettes to recognize an object pose.

Weinland *et al.* [18] proposed a motion history volume that consists of multiple MHIs obtained from multiple cameras for free-viewpoint motion representation.

Kim *et al*. [20] presented depth motion appearance to describe the global 3-D shape of body movement using a modified MHI. An entire sequence of depth maps was encoded to a 4096-dimensional histogram of gradients (HoG) descriptor for motion recognition.

Since cattle actions are well represented in MHI, we also exploit MHI to produce the proposed feature. Therefore, we review MHI and examine its distinguishable patterns with respect to cattle actions.

Let $I(x,y,n)$ and $D(x,y,n)$ be an image sequence with temporal index n, and successive frame difference $D(x,y,n) = |I(x,y,n) - I(x,y,n-1)|$, respectively.

Then, MHI denoted by $H^{\Gamma}(x,y,n)$ is obtained as follows:

$$H^{\Gamma}(x,y,n) = \begin{cases} \Gamma, & if\ D(x,y,n) > \delta, \\ \max(0, H^{\Gamma}(x,y,n-1) - 1), & otherwise, \end{cases} \tag{1}$$

where Γ indicates the maximal temporal extent of the movement and δ is a threshold value for detecting motion occurrence.

If the motion occurrence is sufficiently certain, i.e. $D(x,y,n) > \delta$, then $H^{\Gamma}(x,y,n)$ is set to its maximal value, Γ. H^{Γ} decreases gradually to zero over time unless another motion occurrence is detected at the position. We omit the positional indices for simple notation unless ambiguity may be caused.

Fig. 3 visualizes the process of MHI generation. Given an input image sequence $I(n)$, motion occurrence is detected as shown in **Fig. 3** (b). The latest motion-detected regions have the maximal value, Γ, while other regions are diminished. Since MHI leaves trails about any movements, not only the location of the object's current motion but also the process of how the object moves are represented in MHI.

Fig. 4 demonstrates MHIs for typical cattle actions. Letting a connected region having positive values within $H^{\Gamma}(n)$ be a motion pattern $M_i(\mathrm{n})$, where i indicates the pattern index, each MHI has very unique and distinguishable patterns. In the MHI for mounting action, the motion pattern is quite a bit larger than in other actions. The motion pattern is enlarged instantaneously, and then decreases gradually for 3 ~ 7 seconds. Walking action produces a directional motion pattern. Tail wagging and foot stamping occurring over a very short term yield small and disperse motion patterns. In the following section, we constitute a small dimensional feature that effectively represents such unique patterns in MHI.

3. Proposed Cattle Action Recognition Method

A block diagram of the proposed cattle action recognition method is shown in **Fig. 5**. Initially, the image frame obtained using a fisheye camera is geometrically undistorted. After MHI is obtained as explained in the previous section, moving objects are searched for based on the MHI.

If a moving object is detected, a candidate motion pattern series (CMPS) for the object is gathered, and a feature vector representing the object movement is obtained using the CMPS. Finally, the action for the CMPS is determined using an SVM classifier. A detailed description of each processing step is given in the following subsections.

3.1 Preprocessing

In the proposed system, a single fisheye camera is utilized to monitor multiple enclosures. This system configuration reduces costs effectively, but the fisheye lens causes geometric distortion leading to inaccurate motion information. In the preprocessing step, the fisheye lens distortion is corrected, and the region of the enclosures to be monitored is extracted from each video frame, $F(n)$. The extracted and undistorted region image is denoted by $I(n)$.

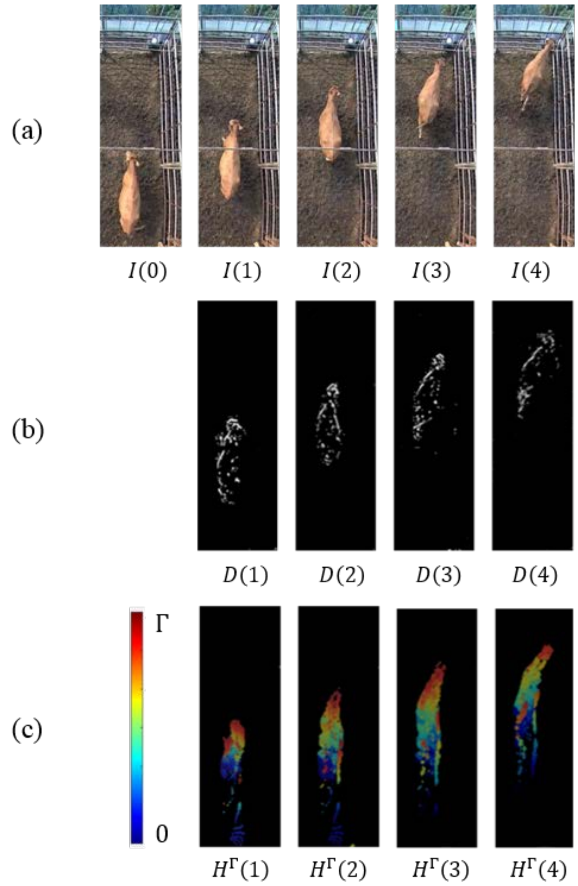

Fig. 3. The process of MHI generation. (a) An input image sequence I. (b) Binary maps indicating $D > \delta$. (c) MHIs H^{Γ}.

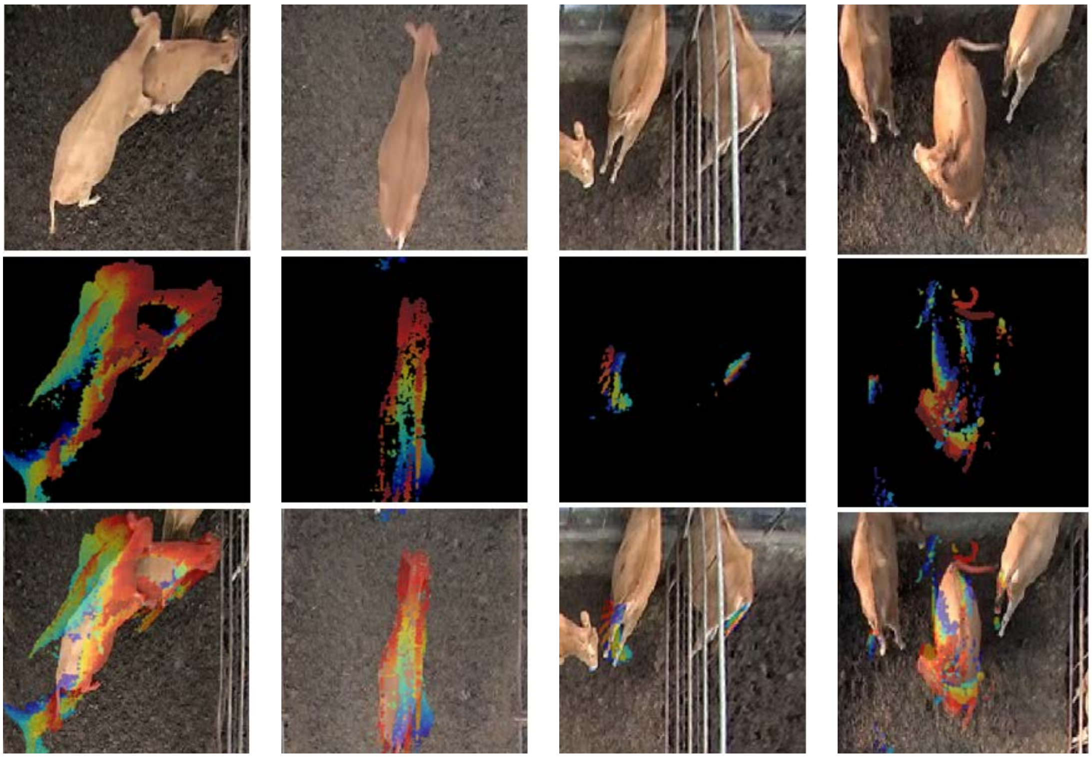

Fig. 4. MHIs for the cattle actions of interest: (a) mounting, (b) walking, (c) tail wagging, and (d) foot stamping. The first and second rows show input images and corresponding MHIs, respectively. Motion patterns are overlaid on the input images in the third row.

Fig. 6 shows an example of the preprocessing step. Since cattle in both the left and right enclosures look shrunken due to the fisheye distortion, their motion is also represented as smaller than the actual movement. Through the preprocessing step, the sizes of adult cattle become similar, regardless of location, as shown in **Fig. 6** (b).

Since the installed camera is fixed, the geometric transformation for correcting the lens distortion does not change over time. Therefore, the coordinate mapping for the transformation can be reused to alleviate huge amounts of re-computation, once it is calculated.

3.2 Candidate Motion Pattern Detection

In a dynamic complex scene, a number of motion patterns can appear in the MHI. In order to improve both computational efficiency and the robustness of the recognition method, only a large motion pattern is regarded as cattle movement. Specifically, if the size of $M_i(n)$ becomes larger than a predefined threshold ξ at the n_0-th frame, we hypothesize that a certain action of interest is about to begin.

As a candidate for an action of interest, a series of motion patterns $M_{i^*}(n_0+n)$'s corresponding to $M_i(n_0)$, called CMPS, are gathered for the following N frames ($N < \Gamma$). Then, the proposed features are extracted from the CMPS.

Object tracking has been one of the most important and still active research areas in the field of computer vision [26-29]**.** A recent research [28] can detect and track multiple objects successfully even in complex background. However, thanks to the slow cattle movement, a simple tracking algorithm is used in our method.

Since the corresponding motion patterns in successive frames typically overlap due to intrinsic slow motion of the cattle, the Jaccard similarity coefficient $\mathcal{J}$ can be a simple but effective measure for temporal association. Given two regions, R_i and R_j, $\mathcal{J}$ is obtained as

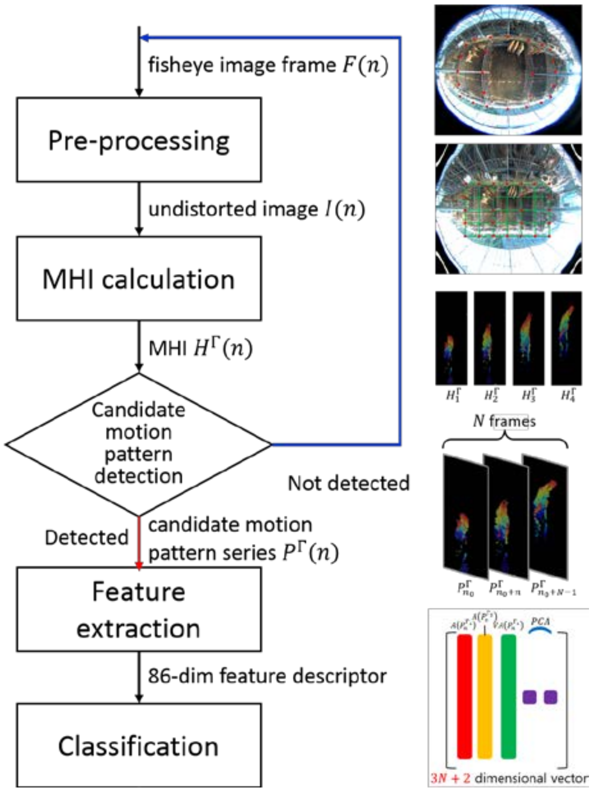

Fig. 5. Block diagram of the proposed cattle action recognition method.

(a) (b)

Fig. 6. Preprocessing result: (a) a fisheye image, and (b) an distorted image.

$$\mathcal{J}(R_i, R_j) = \frac{|R_i \cap R_j|}{|R_i \cup R_j|}, \tag{2}$$

where $|\cdot|$ indicates the region area. That is, the more the two regions overlap, the larger $\mathcal{J}$ becomes.

For a motion pattern $M_i(n)$, the corresponding motion pattern $M_{i^*}(n+1)$ within $H^\Gamma(n+1)$ is determined as follows:

$$i^* = \underset{j}{\operatorname{argmax}}\, \mathcal{J}\big(M_i(n), M_j(n+1)\big) \tag{3}$$

The CMPS obtained from an MHI sequence generated with Γ is denoted by $P_n^\Gamma = M_{i^*}(n_0+n)$, $(0 \le n \le N)$. The CMPS tracked in an MHI sequence generated with Γ is denoted by $P^\Gamma(n) = M_{i*}(n_0+n)$, $(0 \le n < N)$.

3.3 Proposed Features

Although MHI provides a rich representation of object movement, tremendous amounts of data and training time are required for learning MHI itself, due to its high dimensionality. In order to alleviate this problem, we extract low-dimensional features consisting of unique and distinguishable information from MHI, and an SVM classifier is trained using the features.

As mentioned above, cattle behavior can be characterized by the movement area that is represented in the size of the motion pattern in the MHI. The tail wagging and foot stamping behaviors show small-sized motion patterns, whereas walking and mounting behaviors have large motion patterns because the entire body moves. In contrast to the walking behavior showing continuous locomotion in a certain direction, the large amount of movement in the mounting behavior occurs instantaneously, and then rapidly disappears.

In order to discriminate such motion patterns, information about both the amount and direction of movement is utilized as features.

Three motion area-related values $\mathcal{A}(P_n^{\Gamma_1})$, $\mathcal{A}(P_n^{\Gamma_2})$, and $\nabla(\mathcal{A}(P_n^{\Gamma_1}))$ are exploited as the features for the amount of movement, where $\mathcal{A}(\cdot)$ represents the region size and $\nabla(\mathcal{A}(P_n^{\Gamma_1}))$ is obtained as

$$\nabla\big(\mathcal{A}(P_n^{\Gamma_1})\big) = \mathcal{A}\big(P_n^{\Gamma_1}\big) - \mathcal{A}\big(P_{n-1}^{\Gamma_1}\big). \tag{4}$$

When $\Gamma_1 > \Gamma_2$, $\mathcal{A}\big(P_n^{\Gamma_1}\big)$ represents the amount of all the movements detected within the past Γ_1 frames, whereas $\mathcal{A}(P_n^{\Gamma_2})$, indicates the amount of more recent movement, and thus $\mathcal{A}\big(P_n^{\Gamma_1}\big) \ge \mathcal{A}(P_n^{\Gamma_2})$. The ratio between $\mathcal{A}\big(P_n^{\Gamma_1}\big)$ and $\mathcal{A}\big(P_n^{\Gamma_2}\big)$ can describe a unique characteristic of the action. If $\mathcal{A}\big(P_n^{\Gamma_2}\big)$ is close to $\mathcal{A}\big(P_n^{\Gamma_1}\big)$, it implies that abrupt large movement has occurred recently. Although both the mounting and walking actions produce large motion patterns, walking action tends to keep $\mathcal{A}\big(P_n^{\Gamma_2}\big)$ small, compared to $\mathcal{A}\big(P_n^{\Gamma_1}\big)$, whereas with mounting action $\mathcal{A}\big(P_n^{\Gamma_2}\big)$ fluctuates greatly.

To determine whether a specific direction exists in an action, positional information (i.e. the centroid of the motion pattern) is exploited. With mounting action, the center of the motion pattern does not move much, in spite of a large amount of movement. On the other hand, for walking action, the centroid of the motion pattern continuously moves in one direction.

First, the MHI-weighted centroid position of the candidate motion pattern obtained with Γ_1 is calculated as

$$\bar{x}_n = \frac{1}{z} \sum_{(x_k, y_k) \in P_n^{\Gamma_1}} x_k \cdot H^{\Gamma_1}(x_k, y_k, n),$$
$$\bar{y}_n = \frac{1}{z} \sum_{(x_k, y_k) \in P_n^{\Gamma_1}} y_k \cdot H^{\Gamma_1}(x_k, y_k, n), \tag{5}$$

where $z = \sum_{P_n^{\Gamma_1}} H^{\Gamma_1}(x_k, y_k, n)$ is the normalization factor. $\bar{x}_n$ and $\bar{y}_n$, $(0 \leq n \leq N)$, contain the history of center position of the candidate motion pattern.

Then, using the principal component analysis with a set of points consisting of $\bar{x}_n$ and $\bar{y}_n$, two eigenvalues, λ_0 and λ_1, are obtained. Each eigenvalue is proportional to the variance of the given data along the direction of each principal component. As a result, if λ_0 is quite greater than λ_1, it can be determined that the centroid of the motion pattern moves in a specific direction. Therefore, the two eigenvalues are also used as features to represent existence of directionality in the candidate motion pattern.

Conclusively, once a candidate motion pattern is detected from $H^{\Gamma_1}(n_0)$, i.e. $|M_i(n_0)| > \xi$, $M_i(n_0)$ is tracked for the following N frames to obtain $P_n^{\Gamma_1}$ and $P_n^{\Gamma_2}$, $(0 \leq n \leq N)$. The three motion area-related feature values, $\mathcal{A}(P_n^{\Gamma_1}), \mathcal{A}(P_n^{\Gamma_2})$, and $\nabla(\mathcal{A}(P_n^{\Gamma_1}))$ are determined in each frame, while the two motion directionality–related feature values, λ_0 and λ_1, are obtained once for the candidate motion pattern series. Consequently, for each candidate motion pattern series, a $(3N + 2)$-dimensional vector is produced as the feature vector, and is then fed into the SVM classifier to identify the cow action pattern.

4. Experimental Results

In the experiments, a total of 104 video clips, 26 each for the four cattle behaviors, were used. The video sequences were taken at 20 frames per second at a farm in Cheonan, South Korea, from August 2016 to April 2017. Each video clip includes one of the four behaviors.

The proposed method was executed every fifth frame in the video clip (that is, at four frames per second) for effective dimension reduction of the feature vector without degrading performance. The image frame of the three adjacent enclosures, each of which was 8*m* long and 4*m* width, were undistorted to I_n with 400×600 resolution in the preprocessing step.

δ and ξ were set to 20 and 2000, respectively. The detected movement remains within MHI for at least three seconds by setting Γ_1 to 12; and Γ_2 (for keeping more recent movement) was set to 1. Once a candidate motion pattern was detected, it was tracked for 7 seconds according to the mounting behavior duration. That is, the candidate motion pattern series consists of $N = 28$ temporally corresponding motion patterns, and thus, 86-dimensional feature vectors were used in the experiments.

In order to evaluate the performance of the proposed cattle action recognition method, the MHI-histogram of gradients (MHIHoG)-based method [25] was implemented and compared. In this implementation, a 108-dimensional MHIHoG feature vector was extracted from each candidate motion pattern within the same undistorted MHI as in the proposed method.

Table 1. Confusion matrix of the MHIHOG method.

		Actual class			
		Mounting	Walking	Tail wagging	Foot stamping
predicted	Mounting	**0.6667**	0.1667	0.0556	0.2222
	Walking	0.2778	**0.7222**	0.0556	0.1111
	Tail wagging	0.0556	0.1111	**0.7778**	0.2222
	Foot stamping	0	0	0.1111	**0.4444**

Table 2. Confusion matrix of the proposed method.

		Actual class			
		Mounting	Walking	Tail wagging	Foot stamping
predicted	Mounting	**0.8889**	0.1111	0.0556	0
	Walking	0.0556	**0.8333**	0	0.0556
	Tail wagging	0	0	**0.7778**	0.1111
	Foot stamping	0.0556	0.0556	0.1667	**0.8333**

Table 3. Comparison of the proposed and the MHHOG method.

	Proposed	MHIHoG
Accuracy	0.9167	0.8264
Precision	0.8373	0.6714
Recall	0.8333	0.6528

For both the proposed method and MHIHoG identically, 28 video clips, seven each for the four actions, were employed to train the SVM classifier, and the other 76 clips were used for recognition test.

Tables 1 and 2 summarize the feature matrices of the MHIHoG-based method and the proposed method, respectively, obtained using 5-fold cross validation. The general idea of confusion matrix is to count the number of times instances of class actual are classified as class predicted. The MHIHoG-based method had low recognition rates for mounting and foot stamping actions. Mounting and walking actions were confused with each other, and 22% of the two actions were misclassified with MHIHoG features.

The proposed method classifies cattle action patterns with a much higher recognition percentage than the MHIHoG-based method. In particular, for the mounting action (the main action to be recognized for estrus detection), the true positive rate became 0.89, whereas both false positive and false negative were relatively small. In particular, mutual misclassification between the mounting and walking actions was just 8% of the two actions.

The proposed method successfully classified cattle actions in terms of the amount of movement. The mounting/walking actions involving large amounts of locomotion, and the tail wagging/foot stamping actions with small amounts of motion, were classified with 94% accuracy. In contrast, the MHIHoG method misclassified 15% of the data, in terms of the amount of movement, and in particular, the misclassification percentage increased to 33% for the foot stamping action. **Table 3** shows the overall recognition rates for the two methods in terms of accuracy, precision, and recall. For each action, the three metrics are calculated as a binary classification problem. Then, the overall evaluation for a certain metric is determined by averaging the corresponding metrics for the four actions. The proposed method outperformed the MHIHoG-based method in all the three metrics.

The processing time of the proposed method was 25ms for each frame. Specifically, the preprocessing that geometrically undistorts the fisheye image frame requires about 21.5ms. The remaining processes including MHI calculation, feature extraction, and classification take only 3~4ms, ensuring real-time monitoring.

5. Conclusion

In this paper, we presented MHI-based features representing cattle behaviors. The SVM classifier learned with the proposed features recognized cattle mounting action with an accuracy of 89%. Compared with the conventional motion recognition method for human motion, the proposed feature effectively distinguishes the action of the cattle even with a lower dimension. It was confirmed that the proposed action recognition method outperformed the conventional method. Thanks to its low complexity, the proposed method can be applied to monitor multiple enclosures simultaneously. The proposed method is also inexpensive and can be applied in real complex environment.

References

[1] M. S. Dawkins, "A user's guide to animal welfare science," *Trends in Ecology & Evolution*, vol. 21, no. 2, pp. 77–82, 2006. Article (CrossRef Link).

[2] B. F. Sowell, J. G. P. Bowman, M. E. Branine, and M. E. Hubbert, "Radio frequency technology to measure feeding behavior and health of feedlot steers," *Applied Animal Behaviour Science*, vol. 59, no. 4, pp. 277–284, 1998. Article (CrossRef Link).

[3] W. F. Quimby, B. F. Sowell, J. G. P. Bowman, M. E. Branine, M. E. Hubbert, and H. W. Sherwood, "Application of feeding behaviour to predict morbidity of newly received calves in a commercial feedlot," *Canadian Journal of Animal Science*, vol. 81, no. 3, pp. 315–320, 2001. Article (CrossRef Link).

[4] B. Wolfger, B. J. Manns, H. W. Barkema, K. S. Schwartzkopf-Genswein, C. Dorin, and K. Orsel, "Evaluating the cost implications of a radio frequency identification feeding system for early detection of bovine respiratory disease in feedlot cattle," *Preventive veterinary medicine*, vol. 118, no. 4, pp. 285–292, 2015. Article (CrossRef Link).

[5] D. D. Shane, B. J. White, R. L. Larson, D. E. Amrine, and J. L. Kramer, "Probabilities of cattle participating in eating and drinking behavior when located at feeding and watering locations by a real time location system," *Computers and Electronics in Agriculture*, vol. 127, pp. 460–466, 2016. Article (CrossRef Link).

[6] B. Robert, B. J. White, D. G. Renter, and R. L. Larson, "Evaluation of three-dimensional accelerometers to monitor and classify behavior patterns in cattle," *Computers and Electronics in Agriculture* vol. 67, no. 1, pp. 80–84, 2009. Article (CrossRef Link).

[7] C. J. Rutten, A. G. J. Velthuis, W. Steeneveld, and H. Hogeveen, "Invited review: Sensors to support health management on dairy farms," *Journal of Dairy Science*, vol. 96, no. 4, pp. 1928–1952, 2013.

[8] M. Pastell, J. Tiusanen, M. Hakojärvi, and L. Hänninen, "A wireless accelerometer system with wavelet analysis for assessing lameness in cattle," *Biosystems Engineering*, vol. 104 no. 4, pp. 545–551, 2009. Article (CrossRef Link).

[9] A. Nasirahmadi, S. A. Edwards, and B. Sturm, "Implementation of machine vision for detecting behaviour of cattle and pigs," *Livestock Science*, vol. 202, pp. 25-38, 2017. Article (CrossRef Link).

[10] A. Poursaberi, C. Bahr, A. Pluk, A. V. Nuffel, and D. Berckmans, "Real-time automatic lameness detection based on back posture extraction in dairy cattle: Shape analysis of cow with image processing techniques," *Computers and Electronics in Agriculture*, vol. 74, no. 1, pp.110–119, 2010. Article (CrossRef Link).

[11] M. A. Kashiha, C. Bahr, S. Ott, C. P. H. Moons, T. A. Niewold, F. Tuyttens, and D. Berckmans, "Automatic monitoring of pig locomotion using image analysis," *Livestock Science*, vol. 159, pp. 141–148, 2014. Article (CrossRef Link).

[12] Y.-Z. Guo, W.-X. Zhu, P.-P. Jiao, C.-H. Ma, and J.-J. Yang, "Multi-object extraction from top view group-housed pig images based on adaptive partitioning and multilevel thresholding segmentation," *Biosystems Engineering*, vol. 135, pp. 54–60, 2015. Article (CrossRef Link).

[13] Y.-W. Chung, D.-W. Choi, H.-S. Choi, D.-H. Park, H.-H. Chang, and S. Kim, "Automated detection of cattle mounting using side-view camera," *KSII Trans. on Internet and Information Systems*, vol. 9, pp. 3151–3168, 2015. Article (CrossRef Link).

[14] D.-M. Tsai and C.-Y. Huang, "A motion and image analysis method for automatic detection of estrus and mating behavior in cattle," *Computers and Electronics in Agriculture*, vol. 104, pp. 25–31, 2014. Article (CrossRef Link).

[15] A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, pp. 257–267, 2001. Article (CrossRef Link).

[16] D.-Y. Chen, S.-W. Shih, and M.-Y. Liao, "Human action recognition using 2D spatio-temporal templates," in *Proc. of IEEE Int. Conf. on Multimedia and Expo, Beijing, China*, pp. 667–670, 2007. Article (CrossRef Link).

[17] J. Liu and N. Zheng, "Gait history image: a novel temporal template for gait recognition," in *Proc. of IEEE Int. Conf. on Multimedia and Expo, Beijing, China*, pp. 663–666, 2007. Article (CrossRef Link).

[18] D. Weinland, R. Ronfard, and E. Boyer, "Free viewpoint action recognition using motion history volumes," *Computer Vision and Image Understanding*, vol. 104, no. 2, pp. 249–257, 2006. Article (CrossRef Link).

[19] G. R. Bradski and J. W. Davis, "Motion segmentation and pose recognition with motion history gradients," in *Proc. of Machine Vision and Applications*, pp. 238-244, 2000. Article (CrossRef Link).

[20] D.-H. Kim, W.-H. Yun, H.-S. Yoon, and J.-H. Kim, "Action recognition with depth maps using HOG descriptors of multi-view motion appearance and history," *Int. Conf. on Mobile Ubiquitous Computing, Systems, Services and Technologies*, Rome, Italy, 2014. Article (CrossRef Link).

[21] J. K. Aggarwal and Q. Cai, "Human motion analysis: A review," in *Proc. of Nonrigid and Articulated Motion Workshop*, pp. 90–102, 1997. Article (CrossRef Link).

[22] J. K. Aggarwal and S.-H. Park, "Human motion: Modeling and recognition of actions and interactions," in *Proc. of IEEE Int. Symposium on 3D Data Processing, Visualization and Transmission*, pp. 640–647, 2004. Article (CrossRef Link).

[23] A. Farooq and C. S. Won, "A survey of human action recognition approaches that use an RGB-D sensor," *IEIE Trans. on Smart Processing and Computing*, vol. 4, no. 4, pp. 281-290, Aug. 2015. Article (CrossRef Link).

[24] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. on information theory*, vol. 8, no. 2, pp. 179–187, 1962. Article (CrossRef Link).

[25] C.-P. Huang, C.-H. Hsieh, K.-T. Lai, and W.-Y. Huang, "Human action recognition using histogram of oriented gradient of motion history image," in *Proc. of IEEE Int. Conf. on Instrumentation, Measurement, Computer, Communication and Control*, pp. 353–356, 2011. Article (CrossRef Link).
[26] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 2544-2550, 2010. Article (CrossRef Link).
[27] Z. Kalal, M. Krystian, and M. Jiri, "Forward-backward error: Automatic detection of tracking failures," in *Proc. of IEEE Int. Conf. on Pattern recognition*, pp. 2756-2759, 2010. Article (CrossRef Link).
[28] Y. Ma, Y. Liu, S. Liu, and Z. Zhang, "Multiple object detection and tracking in complex background," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 31, no. 2, 2017. Article (CrossRef Link).
[29] W. Chantara, J.-H. Mun, D.-W. Shin, and Y.-S. Ho, "Object tracking using adaptive template matching," *IEIE Trans. on Smart Processing and Computing*, vol. 4, no. 1, pp. 1-9, Feb. 2015. Article (CrossRef Link).

Eui-Ju Heo was born in Daejeon, South Korea, in 1993. She received the B.S. degree in Electronic Engineering from the KOREATECH, South Korea, in 2017. Currently, she is pursuing an M.S. degree in Interdisciplinary Program in Creative Engineering at KOREATECH. Her research interests include image segmentation, real-time image processing, and 3D data processing.

Sung-Jin Ahn was born in Cheongju, South Korea, in 1991. He received the B.S. degree in Electronic Engineering from the KOREATECH, South Korea, in 2016. He received the M.S. degree in Interdisciplinary Program in Creative Engineering form KOREATECH. His research interests include machine learning, real-time image processing, and 3D data processing.

Kang-Sun Choi received the Ph.D. degree in nonlinear filter design in 2003, the M.S. in 1999, and B.S. in 1997, in electronic engineering from Korea University. In 2011, he joined the School of Electrical, Electronics & Communication Engineering at Korea University of Technology and Education where he is currently an Assistant Professor. From 2008 to 2010, he was a Research Professor of the Department of Electronic Engineering at Korea University. From 2005 to 2008, he worked in Samsung Electronics, Korea, as a Senior Software Engineer. From 2003 to 2005, he was a visiting scholar at the University of Southern California. His research interests are in the areas of multimedia compression, video processing, and computational photography. He is the recipient of an IEEE International Conference on Consumer Electronics Special Merit Award (2012).