IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 5 January 2024, accepted 8 March 2024, date of publication 19 March 2024, date of current version 26 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3379277

RESEARCH ARTICLE

Video-Based Analysis of Cattle Behaviors: Improved Classification Using FlowEQ Transform

JUNG-WOO CHAE[1], HYEON-SEOK SIM[1], CHANG-WOO LEE[2], CHANG-SIK CHOI[2], AND HYUN-CHONG CHO[1,3], (Member, IEEE)

[1]Department Graduate Program for BIT Medical Convergence, Kangwon National University, Chuncheon-si 24341, Republic of Korea
[2]Gangwon State Livestock Research Institute, Hoengseong-gun 25266, Republic of Korea
[3]Department of Electronics Engineering, Kangwon National University, Chuncheon-si 24341, Republic of Korea

Corresponding author: Hyun-Chong Cho (hyuncho@kangwon.ac.kr)

This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (MOE) under Grant 2022R1I1A3053872; in part by the Regional Innovation Strategy (RIS) through NRF funded by MOE under Grant 2022RIS-005; and in part by the Rural Development Administration, Republic of Korea, under Grant 00260110.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Animal Care and Use Committee (IACUC) of Rural Development Administration.

ABSTRACT Cattle management plays a crucial role in determining the productivity of livestock farms. With the expansion of large-scale livestock operations, it has become increasingly impractical for livestock managers to rely on traditional visual observations for comprehensive monitoring of cattle behaviors, encompassing health and overall welfare. Consequently, the incorporation of automation technology in livestock management is emphasized. The objective of this study is the video-based identification of cattle behavior that can be utilized in automated cattle management systems. With a specific focus on behaviors closely associated with their management, the study employs deep learning-based action classification methods over the commonly used object detection. This approach enables the classification of intricate, repetitive, and slow behaviors that were challenging to detect. Furthermore, a novel method named FlowEQ transform was introduced, incorporating temporal information into the input data. This enhancement proved instrumental in providing valuable insights for inferring cattle behavior, resulting in an impressive 8% improvement in classification performance and achieving a high accuracy rate of 91.5%. The utilization of action classification and the introduction of the innovative FlowEQ transform mark a significant advancement in automated cattle management. This approach is poised to enhance the efficiency of behavior monitoring on livestock farms.

INDEX TERMS Action classification, automation technology, cattle behavior, cattle management, deep learning, FlowEQ transform.

I. INTRODUCTION

Cattle stand as one of the foremost livestock species globally, wielding a significant impact on agriculture and related industries [1]. Their well-being, productivity, and health carry substantial implications for human livelihood, given their pivotal role as a primary source of food. This significance is underscored by the escalating global population and its consequential rise in food requirements. Over recent decades, the production of beef and dairy from cattle has consistently surged worldwide. As per the Food and Agriculture Organization of the United Nations (FAO), the combined meat yield from beef and buffalo has exhibited a steady upward trajectory [2]. Starting at 47.2 million tonnes in 1980, this

The associate editor coordinating the review of this manuscript and approving it for publication was Okyay Kaynak.

production escalated to 59.6 million tonnes in 2000, further reaching 75.9 million tonnes by 2020. This sustained 25% increase every two decades indicates a continuous and gradual rise in output.

Effective cattle management and health serve as pivotal factors directly impacting the productivity of livestock farms. Traditionally, monitoring cattle behavior for health assessment has relied on direct observation by farm workers [3]. However, technological advancements have revolutionized livestock farming. Driven by efficiency, smaller cattle operations are dwindling in favor of expanding automated facilities in large-scale farms. Consequently, managing numerous cattle with fewer workers makes direct observation of their behavior and health an increasingly impractical endeavor. Even when feasible, this method can be time-consuming, labor-intensive, and susceptible to subjective interpretations, leading to incomplete assessments.

Automated systems for livestock management emerge as a potential remedy for these challenges [4]. Modern farms leverage automation for heightened productivity and competitiveness, integrating automated systems for feed distribution, watering, and environmental control [5], [6]. Notably, most of these advancements are based on the technologies that enable real-time monitoring of cattle behavior. Such systems predominantly rely on sensors and cameras to collect data [7], [8], employing analytical algorithms to decipher cattle behavior patterns. This proves especially beneficial in large-scale farms where direct oversight of numerous cattle by workers poses logistical hurdles. Additionally, the adoption of these automated systems facilitates the rapid and objective identification of potential risks, such as diseases, while being non-intrusive, thereby significantly contributing to cattle welfare by minimizing stress.

A. RELATED WORKS

Several studies have delved into identifying cattle behavior, serving as the groundwork for automated cattle management technology. Fuentes et al. employed Faster R-CNN and YOLOv3 object detection algorithms, successfully detecting 15 behaviors in cows by integrating frame-level and spatio-temporal data [9], [10], [11]. Additionally, their exploration extends to monitoring and recognizing individual cattle behavior in enclosed barn environments, utilizing YOLOv5 as an action detector [12], [13]. This comprehensive approach involves analyzing video data from multiple camera angles to ensure thorough monitoring and accurate behavior recognition.

Zheng and Qin utilized the YOLOv5 object detection algorithm for cow behavior detection, incorporating the Cascaded-Buffered IoU (C-BIoU) for Multi-Object Tracking (MOT) [14], [15]. Meanwhile, Wang et al. focused specifically on detecting cow estrus behavior in natural settings, employing an enhanced YOLOv5 algorithm [16]. Wu et al. applied CNN models with Bi-LSTM networks to classify five specific behaviors in individual dairy cows within complex environments [17], [18].

Moreover, Nguyen et al. introduced a deep learning approach for cow welfare monitoring. Their method utilizes Cascade R-CNN for cow identification and Temporal Segment Networks (TSN) for action recognition, achieving high accuracies in detecting behaviors like drinking and grazing [19], [20], [21]. These diverse studies collectively underscore the significance of identifying cattle behavior. Furthermore, ongoing research continues to explore the potential for automating cattle management.

B. NOVELTY AND CONTRIBUTIONS

Automating cattle management relies on monitoring specific behavior frequencies and abnormalities. Hence, the accuracy of behavior identification methods remains crucial in previous technologies. The proposed study was conducted with a focus on using video-based methods to identify cattle behaviors closely associated with their management. To achieve this, we utilize action classification methods based on deep learning, departing from traditional object detection, to identify cattle behaviors [22]. Action classification, by analyzing multiple video frames, adeptly detects intricate or repetitive behaviors challenging for frame-based object detection. Additionally, the study introduces the FlowEQ transform method tailored for action classification, enhancing cattle behavior classification performance without significantly inflating computational costs through input data transformation.

All data used in this research were internally collected and classified. Collaborating with the Gangwon-do Livestock Research Institute in South Korea, we collected, refined, and labeled essential cattle behaviors based on expert guidance and review from the same institute. This study profoundly acknowledges the significance of automated livestock management systems. The adoption of action classification methods for identifying cattle behavior and the novel FlowEQ transform method stand as promising endeavors, aiming to bolster accuracy and efficiency within automated livestock management systems. Moreover, leveraging high-quality validated data, this research is poised to substantially contribute to the field, ensuring high reliability.

II. MATERIALS AND METHODS

This study aims to differentiate cattle behavior using action classification. The research unfolds across three key phases: First, data collection and dataset creation took place by installing cameras at the Gangwon-do Livestock Research Institute research pens. Subsequently, deep learning-based action classification was utilized to identify five distinct cattle behaviors: normal state, rumination, lactation, calf interaction, and cow interaction. Finally, we introduced the FlowEQ transform, a novel preprocessing method that enhances action classification performance through input data modification. The following sections provide comprehensive insights and explanations into these processes.

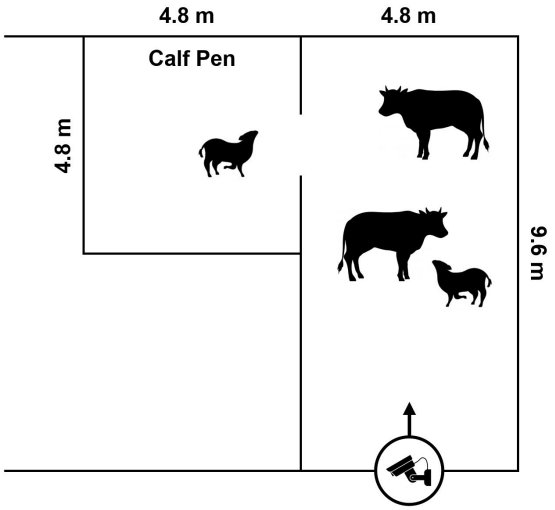


FIGURE 1. Structure of the research pen, and position of the camera.

TABLE 1. Composition of the Constructed Dataset.

Number of cattle behavior videos	
Normal state	400
Rumination	400
Lactation	400
Calf interaction	400
Cow interaction	400
Specifications	
Data format	AVI
Resolution	1280 x 720 (HD)
Video length	150 frames, 30 fps

A. DATA COLLECTION AND DATASET CONSTRUCTION

Research data collection for the study was a collaborative effort with the Gangwon-do Livestock Research Institute in South Korea. Network IP cameras (GB-CDX04, GASI) were installed in 4.8m x 9.6m research pens accommodating two pairs of cows and calves. These cameras were positioned 3 m high, centrally placed along one side of the pen. Adjacent to the camera-installed pen, the upper left corner of the camera's field of view captures the calves' enclosure, while the lower left corner contains part of the other cattle's pens. Fig. 1 illustrates the layout of the pen with the camera location.

Research data was collected in AVI video format using the installed camera, from December 1 to December 11, 2021. The collected data were then meticulously categorized into five behaviors crucial to cattle management, guided by insights from experts at the Gangwon-do Livestock Research Institute. Under expert supervision, this categorization process identified five behaviors: rumination (repeated mouth movements while sitting or standing), lactation, calf interaction (calf licking or sniffing the mother cow), cow interaction (mother cow licking or sniffing the calf), and a normal state (walking around, standing still, or lying without any specific behavior like rumination). Fig. 2 showcases examples of these targeted cattle behaviors observed in the collected data, while detailed specifications of the dataset are outlined in Table 1.

B. VIDEO-BASED CLASSIFICATION OF CATTLE BEHAVIOR

The aim of this study was to identify five distinct cattle behaviors: rumination, lactation, calf interaction, cow interaction, and the normal state. However, these behaviors lack specific, conspicuous postures, such as mounting [23]. Some involve actions that cannot be assessed from a single image, such as rumination observable through mouth movement while standing or lying [24]. Consequently, defining these behaviors necessitates observing the situation over a duration, a task impossible with frame-by-frame object detection. To address this, we proposed employing action classification. Unlike frame-based analysis, action classification operates on a video-level basis, examining multiple frames within a defined range to draw inferences. This approach offers advantages in deciphering sequential or repetitive behaviors challenging to discern in a single scene.

For this study, TimeSformer, a deep learning-based action classification, served as the baseline algorithm [25]. Notably, it is the first model to employ the transformer architecture for video analysis, a pivotal advancement in video comprehension that has spurred diverse model explorations. Utilizing the same input as (1), TimeSformer processes F 3-channel (RGB) frames of size $H \times W$ for video analysis.

$$X \in \mathbb{R}^{H \times W \times 3 \times F} \tag{1}$$

Next, the frame is decomposed into $P \times P$ patches, resulting in N patches covering the entire frame ($N = HW/P^2$). These patches are then flattened into vectors $x_{(p,t)} \in \mathbb{R}^{3P^2}$, where $p = 1 \ldots N$ denotes spatial locations and $t = 1 \ldots F$ signifies an index across frames. Following this step, linear embedding operations are conducted, yielding the embedding vector $z^{(0)}_{(p,t)}$ as depicted in (2).

$$z^{(0)}_{(p,t)} = Ex_{(p,t)} + e^{pos}_{(p,t)} \tag{2}$$

Each patch $x_{(p,t)}$ undergoes linear mapping into an embedding vector $z^{(0)}_{(p,t)} \in \mathbb{R}^D$ using a learnable matrix $E \in \mathbb{R}^{D \times 3P^2}$, while $e^{pos}_{(p,t)} \in \mathbb{R}^D$ represents a learnable positional embedding. This positional embedding encodes the spatiotemporal position of each patch. TimeSformer operates on self-attention by computing queries, keys, and values from the sequence of embedding vectors $z^{(0)}_{(p,t)}$. This mechanism enables TimeSformer to handle temporal dependencies in videos, integrating time-based self-attention within the transformer architecture. This capability allows the model to process sequences of frames over time, enhancing precision in distinguishing various actions. Moreover, TimeSformer incorporates specialized positional encoding to capture temporal ordering information critical for understanding action sequences in videos. Furthermore, it employs Divided Space-Time Attention (divST), a split attention design that independently processes spatial and temporal information. This design significantly enhances the model's ability to discern subtle motion details within video frames.

FIGURE 2. Five cattle behaviors utilized in the study.

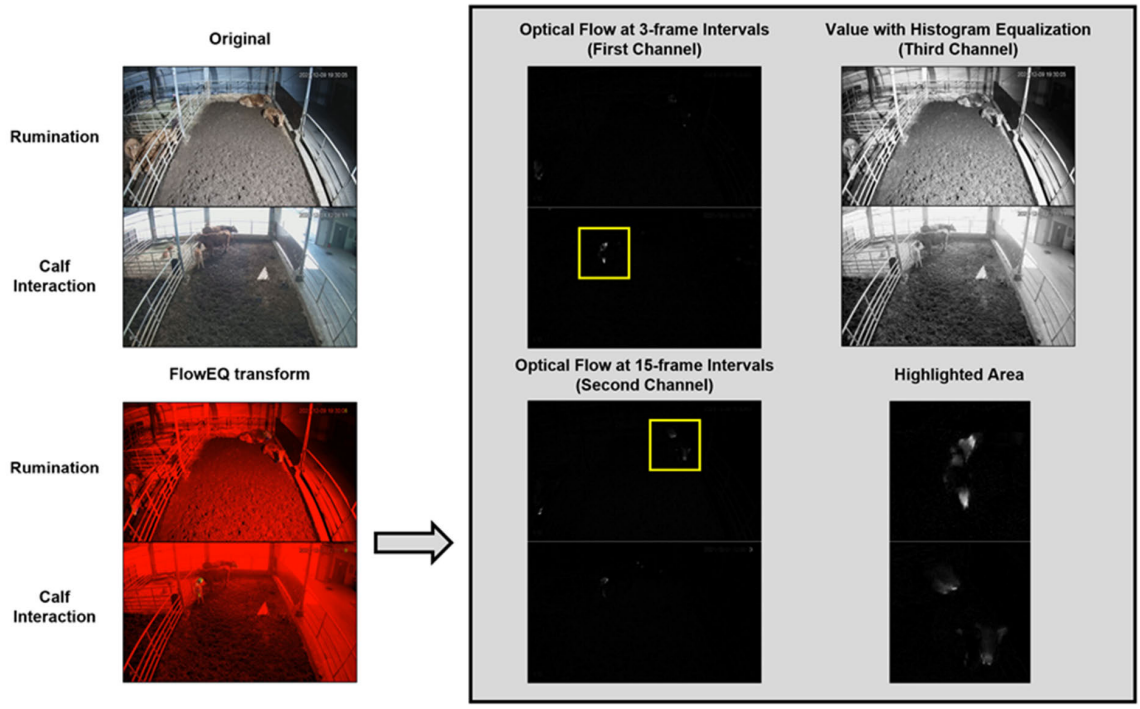


FIGURE 3. Examples and characteristics of the FlowEQ transform application.

C. FlowEQ TRANSFORM FOR IMPROVED ACTION CLASSIFICATION

Images inherently contain spatial information. In the context of videos, incorporating a temporal axis introduces both spatial and temporal data. Dynamic vision algorithms, like action classification, rely on discerning these differences between spatial and temporal information to infer targets effectively. However, temporal cues are not discernible within a single image. Hence, dynamic vision algorithms predominantly operate on video sequences, extracting and leveraging temporal information by analyzing consecutive frames alongside individual frame features. However, if temporal information could be integrated into a single image, it would significantly augment dynamic vision algorithms.

In this study, a novel method called the FlowEQ transform has been introduced to elevate action classification performance while simultaneously identifying cattle behavior. The FlowEQ transform is a preprocessing technique that modifies input data for action classification, transforming conventional 3-channel RGB images into data with distinct channels. In this process, the first channel encapsulates temporal information corresponding to rapid movements, while the second channel captures temporal details of slower or typical movements. The last channel preserves the spatial information from the original image, encompassing its shape and contours.

To extract temporal information for the first and second channels, optical flow was applied to consecutive frames along the same time axis [26]. Optical flow tracks brightness changes based on object movement, creating a motion field that describes object motions. At this point, the motion field can be considered to include the temporal information, which changes over time. Optical flow is divided into local and global methods, with techniques from traditional computational approaches to utilizing deep learning [27]. In this study, the focus is on real-time applicability while minimizing computational costs [28]. Consequently, optical flow methods that are based on deep learning or require high processing costs were excluded, and the Dual TV-L1 optical flow was used [29]. In the current landscape, when even basic

entry-level graphic hardware can perform computations of approximately 1,000 GFLOPs, and servers are capable of up to 50,000 GFLOPs, the Dual TV-L1 optical flow algorithm, which utilizes approximately 9.22 GFLOPs for processing 640 × 480 images at 30 fps, along with the FlowEQ transform that is based on it, does not present a challenge for real-time processing. This method demonstrated superior accuracy and lower noise among various traditional methods such as Lucas-Kanade, Horn-Schunck, Farneback, Brox and EpicFlow [30], [31], [32].

The first channel captures rapid movements by applying optical flow between the current frame and one taken three frames earlier, while the second channel portrays slower movements by comparing the current frame with one taken fifteen frames earlier. For the final channel, the information was derived from the value channel after converting RGB to HSV and applying histogram equalization. Transformer-based deep learning models exhibit significantly less texture bias compared with CNNs [33]. This decreases the reliance on image texture, pattern, or color to identify subjects, instead showing increased sensitivity to higher-dimensional features such as shape and contours. Therefore, after converting to the HSV channel, hue and saturation information were discarded, focusing solely on the value information that highlights shape and contours. Additionally, histogram equalization was employed to intensify the distinctiveness of shapes and contours. This alteration from the original input data is demonstrated in Fig. 3, showcasing examples of applying the FlowEQ transform to rumination and calf interaction. In the highlighted area, the first channel captures rapid movement details of a calf licking a cow in the calf interaction video, while the second channel captures repetitive and slow movement details in the rumination video.

III. RESULTS AND DISCUSSION

All training and performance evaluations were conducted on a system using Windows 10, CUDA 11.3 with cuDNN, Python 3.9.6, and PyTorch 1.12, with the following configuration: Intel® Core™ i9-12900KS Processor, NVIDIA RTX A6000, and 64GB RAM. For evaluation, the study employed a confusion matrix to assess the classification performance of five cattle behaviors, with Table 2 outlining the dataset configurations. The classification results, including TP, TN, FP, and FN, were utilized to derive evaluation metrics such as precision, recall, specificity, f1-score, and accuracy. Finally, all assessments were conducted using a 3-fold cross-validation approach.

Table 3 presents a performance comparison according to the proposed application method, alongside the classification performance of two original action classification models, TimeSformer and I3D. First, using a simple TimeSformer model, the classification performance appeared relatively low. Confusion primarily arose between the normal state and rumination, attributed to the challenge in distinguishing their subtle and non-drastic movements. Also, confusion occurred between calf interaction and cow interaction, differentiation being reliant on the subject of the action. In this case, distinguishing these behaviors posed a challenge due to the visual similarity between adjacent calves and cows. Despite the simplicity of the process, converting from RGB to HSV and applying histogram equalization resulted in an overall performance enhancement. However, similar to employing solely a simple TimeSformer, confusion between the normal state and rumination, as well as between calf and cow interactions, persisted. The utilization of the proposed FlowEQ transform yielded the most significant improvements. It notably succeeded in differentiating between the normal state and rumination previously challenging to discern and accurately classified calf interaction and cow interaction with high efficiency. This enhancement stems from the FlowEQ transform's integration of temporal data about subtle and repetitive rumination movements, enhancing differentiation from the normal state. Additionally, for interactions, it incorporated information that distinctly identified the subject of the action, aiding in more accurate classification. These observations confirm that the proposed FlowEQ transform significantly enhances action classification performance.

TABLE 2. Datasets Configuration used for Performance Evaluation.

Cattle behaviors video data				
	Train	Validation	Test	Total
Normal state	240	80	80	400
Rumination	240	80	80	400
Lactation	240	80	80	400
Calf interaction	240	80	80	400
Cow interaction	240	80	80	400
Extracted frame data				
	Train	Validation	Test	Total
Normal state	2,400	800	800	4,000
Rumination	2,400	800	800	4,000
Lactation	2,400	800	800	4,000
Calf interaction	2,400	800	800	4,000
Cow interaction	2,400	800	800	4,000

An investigation into whether the type of optical flow, crucial to the FlowEQ transform, impacts performance was conducted. For comparison, Farneback, a local optical flow, was applied instead of Dual TV-L1, a previously utilized global optical flow. The 'TimeSformer (divST) & FlowEQ transform (based on Farneback)' section of Table 3 depicts performance with the Farneback-applied FlowEQ transform. Farneback exhibited relatively more noise compared to Dual TV-L1 and lacked a clear motion field for small actions. Consequently, even when integrated into the FlowEQ method, it only marginally improved performance over the original due to the absence of precise temporal information. However, the FlowEQ transform's performance, similar to the HSV application in Table 3 but without the H and S channels, suggests that transformer-based action classification can effectively infer subjects using solely the value component, devoid of the H and S channels.

TABLE 3. Classification Performance of Cattle Behaviors Video Data.

	Precision	Recall	Specificity	F1-score	Accuracy
TimeSformer (divST)					
Normal state	0.882	0.750	0.972	0.811	
Rumination	0.802	0.913	0.935	0.854	
Lactation	0.962	0.938	0.989	0.949	0.831
Calf interaction	0.922	0.588	0.986	0.718	
Cow interaction	0.705	0.988	0.885	0.823	
TimeSformer (divST) & HSV with Histogram Equalization					
Normal state	0.872	0.850	0.966	0.861	
Rumination	0.820	0.913	0.946	0.864	
Lactation	0.974	0.950	0.993	0.962	0.878
Calf interaction	0.870	0.750	0.971	0.805	
Cow interaction	0.860	0.925	0.958	0.892	
TimeSformer (divST) & FlowEQ transform (based on Farneback)					
Normal state	0.841	0.725	0.963	0.779	
Rumination	0.855	0.888	0.958	0.871	
Lactation	0.951	0.963	0.985	0.957	0.872
Calf interaction	0.830	0.913	0.948	0.869	
Cow interaction	0.873	0.863	0.965	0.868	
TimeSformer (divST) & FlowEQ transform (Ours)					
Normal state	0.907	0.850	0.977	0.877	
Rumination	0.902	0.925	0.973	0.914	
Lactation	0.987	0.975	0.997	0.981	0.915
Calf interaction	0.867	0.900	0.964	0.883	
Cow interaction	0.914	0.925	0.977	0.919	
I3D					
Normal state	0.878	0.538	0.972	0.667	
Rumination	0.689	0.913	0.844	0.785	
Lactation	0.977	0.538	0.995	0.694	0.628
Calf interaction	0.493	0.413	0.865	0.449	
Cow interaction	0.440	0.738	0.719	0.551	
I3D & FlowEQ transform (Ours)					
Normal state	0.734	0.588	0.934	0.653	
Rumination	0.890	0.913	0.959	0.901	
Lactation	0.827	0.538	0.964	0.652	0.715
Calf interaction	0.564	0.775	0.824	0.653	
Cow interaction	0.663	0.763	0.879	0.709	

Subsequently, an assessment was conducted to evaluate the proposed FlowEQ transform's efficacy in other action classification models. For this purpose, Inflated 3D ConvNets (I3D), a representative 3D CNN model, was utilized [34]. Unlike TimeSformer, I3D adopts a CNN architecture for action classification. It transforms a 2D CNN into 3D, enhancing its capability to capture temporal information in video data. The 'I3D' and 'I3D & FlowEQ transform (Ours)' sections in Table 3 represent the performance of the original I3D and the I3D integrated with the FlowEQ transform. While the original I3D displayed some classification capability, it performed less effectively than TimeSformer. It encountered challenges distinguishing between the normal state and lactation and struggled notably with differentiating calf and cow interactions, impacting its overall performance. However, incorporating the FlowEQ transform with I3D enhanced the classification performance by approximately 8%. Although the FlowEQ transform did not significantly improve the classification of the normal state and lactation, it notably reduced confusion between the two interactions, emphasizing its effectiveness in action classification.

An additional experiment was conducted to assess the FlowEQ transform's efficacy in encapsulating temporal information within each frame. Hypothesizing that frames

TABLE 4. Classification performance of extracted frame data.

	ViT-B/16				
	Precision	Recall	Specificity	F1-score	Accuracy
Normal state	0.411	0.380	0.714	0.395	
Rumination	0.833	0.031	0.996	0.060	
Lactation	0.808	0.379	0.938	0.516	0.348
Calf interaction	0.207	0.585	0.340	0.306	
Cow interaction	0.483	0.363	0.780	0.414	
	ViT-B/16 & FlowEQ transform (Ours)				
	Precision	Recall	Specificity	F1-score	Accuracy
Normal state	0.657	0.605	0.902	0.630	
Rumination	0.709	0.894	0.877	0.791	
Lactation	0.876	0.996	0.947	0.932	0.703
Calf interaction	0.513	0.455	0.877	0.482	
Cow interaction	0.712	0.566	0.928	0.631	

processed through FlowEQ, containing temporal data, could enable action analysis even without explicitly learning the temporal axis, 10 frames were regularly extracted from the videos, forming the extracted frame data in Table 2. Table 4 showcases the classification results for this extracted frame data across different methods. ViT-B/16, the classification algorithm underlying TimeSformer, was utilized for simple image classification without leveraging action classification that learns temporal axis information [35]. However, when solely trained through ViT-B/16, proper classification was not achieved overall. Particularly, behaviors like rumination, indistinct in a single frame, remained unidentified. In conclusion, the failure to differentiate between classes led to an overall classification failure. However, classification with the applied FlowEQ transform remarkably improved behavior differentiation, almost doubling the classification performance.

Notably, it successfully identified rumination, a behavior previously undetected. Moreover, behaviors like lactation, calf interaction, and cow interaction where visual cues merely depict a calf and cow in proximity exhibited notably reduced confusion owing to the incorporation of temporal information. This underscores the diverse potential of the FlowEQ transform method and its significance in dynamic vision analysis.

IV. CONCLUSION

The primary objective of this study was to identify cattle behaviors using action classification while concurrently enhancing performance with the introduction of the FlowEQ transform. The application of action classification effectively discerned cattle behaviors that proved challenging or indistinct at the frame level. Moreover, the newly proposed FlowEQ transform modified the input data of the classification model, introducing a motion field representing movement. This allowed the incorporation of temporal information into the frames constituting the video, enabling the action classification to learn from more informative data and achieve heightened inferential performance without significant increases in computational costs, owing to its straightforward procedures. Furthermore, the inclusion of temporal information in the images was confirmed through verification using a simple image classification algorithm. We anticipate that these advancements in development techniques could be utilized in a wide range of applications, where enhanced understanding and classification of complex behaviors are crucial, such as in automated monitoring and management systems for livestock, wildlife observation, and even in enhancing surveillance and security measures.

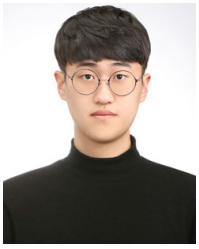
While the application of action classification successfully classified cattle behavior and improved performance through the FlowEQ transform, there are areas that warrant further experimentation. First, despite collecting 400 videos per class for cattle behavior, this number, when divided into train, validation, and test sets, may not be considered extensive. Therefore, we are continuing to collect data and plan to conduct further research with an expanded dataset. Second, the proposed FlowEQ transform will be applied to various action classification models to assess its effectiveness. Concurrently, the performance in action classification based on architectures other than the transformer will be evaluated to confirm broader applicability. Finally, building on the observed potential of the FlowEQ transform in frame-level analysis, we plan to develop new deep learning models that incorporate this method.

REFERENCES

[1] F. Napolitano, A. Bragaglio, E. Sabia, F. Serrapica, A. Braghieri, and G. De Rosa, "The human–animal relationship in dairy animals," *J. Dairy Res.*, vol. 87, no. S1, pp. 47–52, Aug. 2020.
[2] *Food and Agriculture Organization of the United Nations*. Accessed: Jan. 5, 2024. [Online]. Available: https://www.fao.org/home/en/
[3] S. Paudyal, "Using rumination time to manage health and reproduction in dairy cattle: A review," *Veterinary Quart.*, vol. 41, no. 1, pp. 292–300, Jan. 2021.
[4] M. Crociati, L. Sylla, A. De Vincenzi, G. Stradaioli, and M. Monaci, "How to predict parturition in cattle? A literature review of automatic devices and technologies for remote monitoring and calving prediction," *Animals*, vol. 12, no. 3, p. 405, Feb. 2022.
[5] P. Karn, P. Sitikhu, and N. Somai, "Automatic cattle feeding system," *in Proc. 2nd Int. Conf. Eng. Technol.*, vol. 2, Dhapakhel, Nepal, Sep. 2019, pp. 138–142.
[6] Y. Qu, G. Sun, B. Zheng, and W. Liu, "Environment monitoring system of dairy cattle farming based on multi parameter fusion," *Information*, vol. 12, no. 7, Jul. 2021, Art. no. 273.

[7] H. Dohi, A. Yamada, S. Tsuda, T. Sumikawa, and S. Entsu, "Technical note: A pressure-sensitive sensor for measuring the characteristics of standing mounts of cattle," *J. Animal Sci.*, vol. 71, no. 2, pp. 369–372, Feb. 1993.
[8] T.-K. Dao, T.-L. Le, D. Harle, P. Murray, C. Tachtatzis, S. Marshall, C. Michie, and I. Andonovic, "Automatic cattle location tracking using image processing," in *Proc. 23rd Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2015, pp. 2636–2640.
[9] A. Fuentes, S. Yoon, J. Park, and D. S. Park, "Deep learning-based hierarchical cattle behavior recognition with spatio-temporal information," *Comput. Electron. Agricult.*, vol. 177, Oct. 2020, Art. no. 105627.
[10] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
[11] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
[12] A. Fuentes, S. Han, M. F. Nasir, J. Park, S. Yoon, and D. S. Park, "Multi-view monitoring of individual cattle behavior based on action recognition in closed barns using deep learning," *Animals*, vol. 13, no. 12, Jun. 2023, Art. no. 2020.
[13] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, "TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 2778–2788.
[14] Z. Zheng and L. Qin, "PrunedYOLO-Tracker: An efficient multi-cows basic behavior recognition and tracking technique," *Comput. Electron. Agricult.*, vol. 213, Oct. 2023, Art. no. 108172.
[15] P. Voigtlaender, M. Krause, A. Osep, J. Luiten, B. B. G. Sekar, A. Geiger, and B. Leibe, "MOTS: Multi-object tracking and segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Feb. 2019, pp. 7942–7951.
[16] R. Wang, Z. Gao, Q. Li, C. Zhao, R. Gao, H. Zhang, S. Li, and L. Feng, "Detection method of cow estrus behavior in natural scenes based on improved YOLOv5," *Agriculture*, vol. 12, no. 9, Aug. 2022, Art. no. 1339.
[17] D. Wu, Y. Wang, M. Han, L. Song, Y. Shang, X. Zhang, and H. Song, "Using a CNN-LSTM for basic behaviors detection of a single dairy cow in a complex environment," *Comput. Electron. Agricult.*, vol. 182, Mar. 2021, Art. no. 106016.
[18] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging," 2015, *arXiv:1508.01991*.
[19] C. Nguyen, D. Wang, K. Von Richter, P. Valencia, F. A. P. Alvarenga, and G. Bishop-Hurley, "Video-based cattle identification and action recognition," in *Proc. Digit. Image Computing: Techn. Appl. (DICTA)*, Nov. 2021, pp. 01–05.
[20] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6154–6162.
[21] L. Wang, Y. Xiong, Z. Wang, Y. Qiao, D. Lin, X. Tang, and L. Van Gool, "Temporal segment networks: Towards good practices for deep action recognition," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 20–36.
[22] T. Lodkaew, K. Pasupa, and C. K. Loo, "CowXNet: An automated cow estrus detection system," *Expert Syst. Appl.*, vol. 211, Jan. 2023, Art. no. 118550.
[23] J.-W. Chae and H.-C. Cho, "Identifying the mating posture of cattle using deep learning-based object detection with networks of various settings," *J. Electr. Eng. Technol.*, vol. 16, no. 3, pp. 1685–1692, May 2021.
[24] J. H. M. Metz, "Time patterns of feeding and rumination in domestic cattle," Wageningen Univ. Res., Wageningen, The Netherlands, Tech. Rep. 28228271, 1975.
[25] G. Bertasius, H. Wang, and L. Torresani, "Is space-time attention all you need for video understanding?" in *Proc. ICML*, Jul. 2021, vol. 2, no. 3, pp. 1–4.
[26] S. S. Beauchemin and J. L. Barron, "The computation of optical flow," *ACM Comput. Surv.*, vol. 27, no. 3, pp. 433–466, 1995.
[27] J. Hur and S. Roth, "Optical flow estimation in the deep learning age," in *Modelling Human Motion: From Human Perception to Robot Design*, N. Noceti, A. Sciutti, and F. Rea, Eds. Springer, 2020, pp. 119–140.
[28] S. T. H. Shah and X. Xuezhi, "Traditional and modern strategies for optical flow: An investigation," *Social Netw. Appl. Sci.*, vol. 3, no. 3, Mar. 2021, Art. no. 1.
[29] C. Zach, T. Pock, and H. Bischof, "A duality based approach for realtime TV-L^1 optical flow," in *Proc. 29th DAGM Symp. Pattern Recognit.*, vol. 29, Heidelberg, Germany, Sep. 2007, pp. 214–223.
[30] A. Bruhn, J. Weickert, and C. Schnörr, "Lucas/Kanade meets Horn/schunck: Combining local and global optic flow methods," *Int. J. Comput. Vis.*, vol. 61, no. 3, pp. 1–21, Feb. 2005.
[31] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," in *Proc. 13th Scandin. Conf. Image Anal. (SCIA)*, vol. 13, Halmstad, Sweden, Jun. 2003, pp. 363–370.
[32] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *Proc. 8th Eur. Conf. Comput. Vis. (ECCV)*, vol. 8, Prague, Czech Republic, May 2004, pp. 25–36.
[33] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," 2015, *arXiv:1511.08458*.
[34] J. Carreira and A. Zisserman, "Quo vadis, action recognition? A new model and the kinetics dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4724–4733.
[35] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16 × 16 words: Transformers for image recognition at scale," 2010, *arXiv:2010.11929*.

JUNG-WOO CHAE received the B.S. degree from the Department of Electronics Engineering, Kangwon National University, South Korea, in 2019, and the M.S. degree from the Department of BIT Medical Convergence, Kangwon National University, in 2021, where he is currently pursuing the Ph.D. degree.



HYEON-SEOK SIM is currently pursuing the combined B.S. and M.S. degree with the Department of Electronics Engineering, Interdisciplinary Graduate Program for BIT Medical Convergence, Kangwon National University, South Korea.

CHANG-WOO LEE received the M.S. degree in dairy science and the Ph.D. degree in animal breeding and genetics from Kangwon National University, South Korea, in 1999 and 2003, respectively. From 2003 to 2006, he did postdoctoral research with the National Institute of Animal Science. He is currently a Chief Researcher with Gangwon State Livestock Research Institute, South Korea.

CHANG-SIK CHOI received the B.S. degree in animal resource science and the B.S. degree in animal science from Kangwon National University, South Korea, in 2013 and 2015, respectively. He is currently pursuing the Ph.D. degree with Gangwon State Livestock Research Institute, South Korea.

HYUN-CHONG CHO (Member, IEEE) received the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Florida, USA, in 2009. From 2010 to 2011, he was a Research Fellow with the University of Michigan, Ann Arbor, MI, USA. From 2012 to 2013, he was a Chief Research Engineer with LG Electronics, South Korea. He is currently a Professor with the Department of Electronics Engineering and the Interdisciplinary Graduate Program for BIT Medical, Kangwon National University, South Korea.