ICIC Express Letters
Part B: Applications
Volume **14**, Number **9**, September **2023**

ICIC International ©2023 ISSN 2185-2766
pp. 993–1000

A COMPARATIVE STUDY ON COW RECOGNITION: ANALYZING COLOUR SPACES, DISTANCE MEASURES AND DEEP NEURAL NETWORKS

Cho Cho Mar[1], Thi Thi Zin[1,*], Pyke Tin[1], Kazuyuki Honkawa[2]
Ikuo Kobayashi[3] and Yoichiro Horii[4]

[1]Graduate School of Engineering
[3]Sumiyoshi Livestock Science Station, Field Science Center, Faculty of Agriculture
[4]Center of Animal Disease Control
University of Miyazaki
1-1, Gakuen Kibanadai-Nishi, Miyazaki 889-2192, Japan
{ machochomar2020; pyketin11 }@gmail.com; { ikukob; horii }@cc.miyazaki-u.ac.jp
*Corresponding author: thithi@cc.miyazaki-u.ac.jp

[2]Honkawa Ranch
3898, Oaza-Takase, Hita City, Oita 877-0056, Japan
kazuyuki@honkawa.jp

Received January 2023; accepted April 2023

Abstract. *The implementation of autonomous cattle monitoring systems is becoming increasingly important in the livestock production and dairy farming industries. These systems require robust object detection and tracking systems to enhance their performance and reliability. The recognition task plays a crucial role in building a powerful tracking system. The aim of this study is to perform the recognition task for cow recognition by analyzing and comparing different colour spaces and distance measures to select the optimal ones. We extracted colour moment features and Co-occurrence Matrix (CM) features from various colour spaces: RGB, YCbCr, XYZ, HSV, CIELab, and grey level. We compared these features using different distance measures based on their accuracy values. We also conducted experiments on different pre-trained deep neural networks to extract Convolutional Neural Network (CNN) features and compared the accuracy values of the classification results with the Support Vector Machine (SVM) method. These experiments were conducted on the cow dataset created from a continuous 2-hour video. The results demonstrate that the combination of CM features extracted from the HSV colour space, and the Manhattan distance measure produced the highest accuracy values. Furthermore, we found that using the InceptionV3 pre-trained deep neural network produced the best accuracy results when combined with the SVM classifier. These findings provide insights into optimizing cow recognition for autonomous monitoring systems.*

Keywords: Co-occurrence matrix, CNN features, Colour moment features, Cow recognition, Distance measures, SVM

1. **Introduction.** The number of cattle is increasing year by year and the number of tasks to monitor individual cattle and analyze the condition and well-being state of cattle is rising in livestock and dairy farming. According to this, the workload for the cattle farm leads to challenging work for today's dairy communities. To cover this issue, different kinds of modern technologies are utilized for developing autonomous monitoring systems. Different kinds of sensors and IoT devices are very popular to accomplish this system. Among them, Radio Frequency Identification (RFID) is also an effective way to recognize and identify individual cattle by attaching an RFID tag to the ear of the cattle [1]. There are other sensors that are attached to cattle's bodies such as ear-tag and halter type sensors, neck collar type sensors, rumen bolus type sensors, leg-tag type sensors, and

DOI: 10.24507/icicelb.14.09.993

tail- and vaginal-mounted type sensors [2]. The disadvantages of these devices are making stress on cattle and high cost and sometimes producing unreliable results.

The popularity of visual-based monitoring systems is emerging and playing an important role in improving the productivity and profitability of cattle farms. The main objective of a non-contact sensor-based cow monitoring system that is very helpful for the agriculture and dairy community is to reduce the workload of employees, costs, and stress in cows due to the attached sensors to the cow bodies [3]. Those monitoring systems are composed of five main steps: data acquisition, object detection, object tracking, behavior recognition, and analysis and decision-making according to the results of previous parts.

Cow detection and tracking system is the root of a robust cow monitoring system to produce reliable information for varieties of cow behavior analysis like transition change detection, social relations between cows, rumination detection, and lameness detection. Valuable information from the behavior analysis can detect early disease symptoms and make appropriate prevention by assisting in the calving process, providing immediate care needed after calving, and reducing the calf death rate [1]. And recognition task plays a crucial part in building powerful tracking systems by matching the objects from time to time [4].

There are many ways to accomplish the object recognition task in deep learning and image processing methods. In a deep learning manner, the deep feature maps are extracted using different convolutional neural networks [5]. Those feature maps are used as input and classified with different classification methods such as Support Vector Machine (SVM), random forest, decision matrix, simple distance measures, k nearest neighbors, Multilayer Perceptron (MLP), and Naïve Bayes. Besides deep features, other features like color, texture, edge, shape corner, and point descriptors can be extracted to represent the object. There are numerous distance measures and classification methods to classify using those features [6].

We have previously proposed a cow detection and tracking system [7] in Lifetech 2020. In that work, we built multiple object tracking algorithms by assisting with recognition tasks via utilizing colour moments from RGB colour space, Gray Level Co-occurrence Matrix (GLCM), and CNN features using ResNet50 pre-trained network. In this paper, we proposed an object recognition step for a later cow tracking system using colour features, texture features, and CNN features. We also make numerous comparisons on different colour spaces, distance features, and CNN networks. This analysis of various kinds of features describes the optimal feature representation for the object tracking process. The proposed method for cow tracking systems has several advantages, including the use of multiple features such as colour features, texture features, and CNN features for object recognition. This approach helps to overcome some of the limitations of previous methods, which may have relied on a single type of feature or sensor.

By utilizing multiple features, the proposed method can provide better accuracy and reliability compared to traditional contact sensor-based cow monitoring systems. Additionally, this approach reduces the stress on the cows, as it does not require physical contact. Demonstrations have shown that the use of CNNs can help the system to learn and adapt to different environments and conditions, making it more versatile and adaptable. This means that the system can perform well even in challenging situations such as low-light conditions or crowded environments. Furthermore, the proposed method involves the identification of optimal feature representation for the object tracking process by analyzing different colour spaces, distance features, and CNN networks. This allows for a more thorough understanding of the underlying data and helps to improve the overall performance of the system. In summary, the proposed method for cow tracking systems offers several advantages, including better accuracy and reliability, reduced stress on the cows, adaptability to different environments, and improved feature representation analysis. These benefits make it a promising approach for improving cow monitoring and management.

The rest of this paper is composed of other three sections. Section 2 includes a detailed explanation of the proposed system and methods. In Section 3, the experimental results and evaluation for different features are presented. Finally, we discuss this paper and future work in Section 4.

2. Proposed System.

The overall proposed system and methods that are exploited for cow recognition are proposed in this section. First, we complete data acquisition on the cattle farms using fisheye cameras and 4K cameras installed on the large-scale cattle farm in Oita Prefecture, Japan. After getting the video from cattle farms, we performed some preprocessing steps to remove the unwanted regions according to the captured scenes. Then we continued to the detection stage using Hybrid Task Cascade (HTC) instance segmentation network which produces bounding box and mask predictions [8]. The cow regions from the detection are extracted by focusing on the mask detection results. The recognition process is performed on the detected cow regions from the HTC network as shown in Figure 1. The calculation of colour moments for different colour spaces is explained in Section 2.1. Moreover, the co-occurrence matrix and feature extraction are proposed in detail in Section 2.2. The information on various pre-trained networks is introduced briefly in Section 2.3. Then, the complete recognition scheme for this research work is described in Section 2.4.

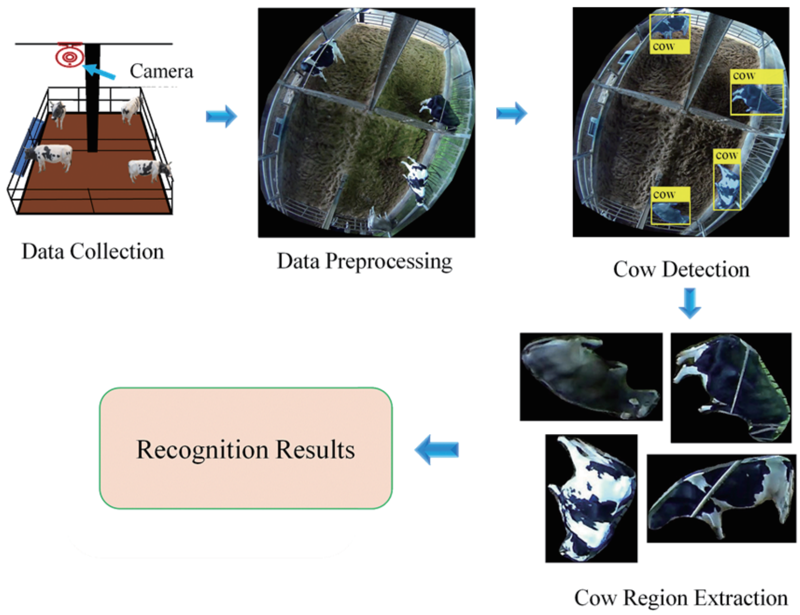


FIGURE 1. Flowchart for the proposed system

2.1. Colour moment features.

In literature, colour moments and colour histograms are used to perform object recognition [9,10]. We extracted three colour moments called mean, standard deviation, and skewness from the cow's region with Equation (1) from different colour spaces and made a comparison between them to find the optimal colour space for the cow tracking system. These moments are capable of scaling and rotation invariance between different cow regions. After feature extraction, the distance between two colour feature vectors between two cow regions at time t and $t+1$ is calculated using Equation (2).

$$M = [\mu, \delta, s]_d \tag{1}$$

$$M_{dist} = f(M_t, M_{t+1}) \tag{2}$$

where M is the feature vector of the mean (μ), standard deviation (δ), and skewness (s) for each colour channel (d) and M_{dist} is the distance between two colour moment feature vectors, M_t and M_{t+1}. And f is the distance function. Distance measures that are used in this paper are explained in Table 1. Briefly, we would like to introduce the distance measures described in Table 1. Distance measures, also known as distance metrics, are essential in machine learning and data science. Different distance metrics are chosen depending on the type of data being analyzed, making it crucial to understand the various metrics and the intuition behind them. Table 1 provides an overview of some of the most popular and useful distance measures used in various research fields. Using an effective distance metric can significantly improve the performance of our machine learning models, whether for classification tasks or clustering.

TABLE 1. Distance measures

Distance measure	**Equation**
Cosine	$1 - \frac{\sum_{i=1}^{k} M_{t,i} M_{t+1,i}}{\sqrt{\sum_{i=1}^{k} M_{t,i}^2}\sqrt{\sum_{i=1}^{k} M_{t+1,i}^2}}$
Manhattan	$\sum_{i=1}^{k} \lvert M_{t,i} - M_{t+1,i} \rvert$
Chebyshev	$\max_i \left(\lvert M_{t,i} - M_{t+1,i} \rvert \right)$
Euclidean	$\sqrt{\sum_{i=1}^{k} \left(M_{t,i} - M_{t+1,i} \right)^2}$
Chi-square	$\frac{1}{2} \sum_{i=1}^{k} \frac{(M_{t,i} - M_{t+1,i})^2}{(M_{t,i} + M_{t+1,i})}$

One of the most widely used distance metrics is Euclidean distance. It works based on the Pythagoras theorem and signifies the shortest distance between two points. The Euclidean distance is commonly used in applications such as image recognition, object detection, and speech recognition. Another important distance metric is cosine similarity. Cosine similarity is proportional to the dot product of two vectors and inversely proportional to the product of their magnitudes. It measures the cosine angle between the two vectors and ranges from 0 to 1, with 1 indicating perfect similarity. Cosine similarity only considers the angle between the two vectors and not the distance between them. It is commonly used in text analysis, recommendation systems, and image recognition.

Manhattan distance, also known as city block distance, is the sum of absolute differences between points across all dimensions. This distance metric is commonly used in computer vision, pattern recognition, and robotics. The Chebyshev distance, also referred to as the chessboard distance, is the greatest distance on any dimension between two real-valued vectors. This metric is often used in warehouse logistics, where the longest path determines the time it takes to get from one point to the next.

Chi-square distance is a weighted Euclidean distance that considers the difference between the observed and expected values of a variable. This distance metric is commonly used in statistical analysis, pattern recognition, and image processing. The chi-square distance is at the heart of correspondence analysis, which is extensively used in ecological research. This distance function is calculated on relative counts, not on the original ones, and is standardized by the mean, not by the variance.

In conclusion, distance metrics play a vital role in machine learning and data science. Understanding the various distance measures and their applications is crucial in selecting an appropriate distance metric for a specific task. An effective distance metric can significantly improve the performance of our machine learning models, leading to better results and more accurate predictions.

2.2. **Co-occurrence Matrix (CM) features.** In the case of cow regions that are similar in colour, the textural information must be considered for object recognition to produce more accurate results. Gray Level Co-occurrence Matrix (GLCM) is well known textural feature for image representation [11,12]. However, some cows in our data are similar in colour. Therefore, we extract the textural features from CM on the grey-level images as well as on colour images and make comparisons among those colour spaces. Then, we calculate four types of features: contrast (Con), correlation ($Corr$), energy (Eng), and homogeneity (H), with four degrees of orientation and a 1-pixel distance. When we compare two cow regions by means of CM, the minimum distance between any two features is calculated for all orientations to act as a rotation-invariant approach. The two main parameters, orientation (θ) and distance, are required to calculate the features from the CM matrix. In this paper, we calculate these features for each colour channel (d) with all orientations ($\theta = 0°$, $45°$, $90°$, $135°$) and distance of 1 pixel as shown in Equation (3).

$$CM_{d,dist} = [Con_{d,\min}, Corr_{d,\min}, Eng_{d,\min}, H_{d,\min}]_{t,t+1} \tag{3}$$

$$Con_{d,\min} = \min\left(|Con_{\alpha,t} - Con_{\beta,t+1}|_d \ \forall\alpha, \beta \in \theta\right) \tag{4}$$

$$Corr_{d,\min} = \min\left(|Corr_{\alpha,t} - Corr_{\beta,t+1}|_d \ \forall\alpha, \beta \in \theta\right) \tag{5}$$

$$Eng_{d,\min} = \min\left(|Eng_{\alpha,t} - Eng_{\beta,t+1}|_d \forall\alpha, \beta \in \theta\right) \tag{6}$$

$$H_{d,\min} = \min\left(|H_{\alpha,t} - H_{\beta,t+1}|_d \ \forall\alpha, \beta \in \theta\right) \tag{7}$$

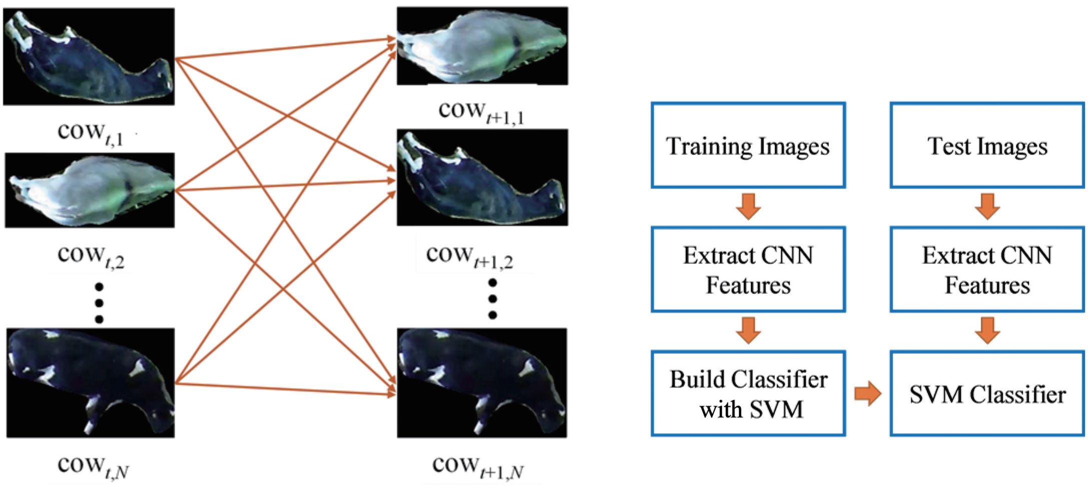
$CM_{d,dist}$ is the vector composed of four minimum distances of CM features for each d at time t and $t+1$. The minimum distance between two cow regions at time t and $t+1$ by comparing with all the orientation degrees is shown in Equations (4)-(7).

2.3. **CNN features.** For object recognition, deep CNN features are also important to represent the object's appearance [13,14]. The pre-trained deep neural networks are used in feature extraction and performed classification using SVM, one of the popular machine learning algorithms. Among the different pre-trained networks, we compared the 16 networks presented in Table 2.

Table 2. Pre-trained networks

Pre-trained network	**Depth**	**Number of parameters ($\times 10^5$)**	**Input size**
squeezenet	18	1.24	227×227
googlenet	22	7	224×224
inceptionv3	48	23.9	299×299
densenet201	201	20	224×224
mobilenetv2	53	3.5	224×224
resnet18	18	11.7	224×224
resnet50	50	25.6	224×224
resnet101	101	44.6	224×224
xception	71	22.9	299×299
incetpionresnetv2	164	55.9	299×299
shufflenet	50	1.4	224×224
darknet19	19	20.8	256×256
darknet53	53	41.6	256×256
alexnet	8	5.3	227×227
vgg16	16	138	224×224
vgg19	19	144	224×224

2.4. **Recognition scheme.** Recognition is also the identification of objects among the classes they belong to according to the previously learned knowledge. In the recognition task, there are two different ways according to the features. For colour moments and CM features, we make a comparison between cow regions in two consecutive frames as the way most of the tracking algorithms behave. Each cow region in the previous frame at time t is compared with all the cow regions at time $t+1$ and find the pair of cows with minimum distance. The number of maximum pairs of cows is $N \times N$ as shown in Figure 2(a). N is the number of cow classes. Figure 2(b) shows the recognition scheme using CNN features and the SVM classifier. In this case, we split into two stages: training and testing. In the training, we extract CNN features using different pre-trained networks and build the classifier with SVM. In the testing stage, we extract CNN features from input images and classify them using the pre-built SVM model.



(a) (b)

FIGURE 2. Recognition scheme for (a) colour moments and CM features, and (b) CNN features

3. **Experiments and Results.** This section involves all the explanations about datasets, experiments, and results in pursuing this research.

3.1. **Dataset.** We prepared the cow data from two cattle farms with a total of 12 cows. Then, we collect 7,200 images for each cow from continuous 2-hour-long videos with 1 frame per second (fps). Figure 3 shows the sample images of all cow classes. We can see that some cows have similar patterns and colour values. For building the SVM classifier, we used 70% of the dataset for training and the rest 30% for testing for each class. We calculated the accuracy of the classification results of test data.

3.2. **Experiments.** Three main experiments are accomplished in this section. The first one is the comparisons among the colour spaces by means of colour moments using the different distance measures. The second experiment is also the comparison of different colour spaces by the way of CM features. The final experiment is the comparison of various pre-trained deep learning networks based on the extracted CNN features and SVM classifier.

3.3. **Results.** We conducted experiments on various colour spaces, distance measurements, and deep learning networks for cow recognition. The accuracy values for each combination are presented in Table 3. Table 3(a) shows that the CIELab colour space outperforms other colour spaces in representing objects using cosine distance measure

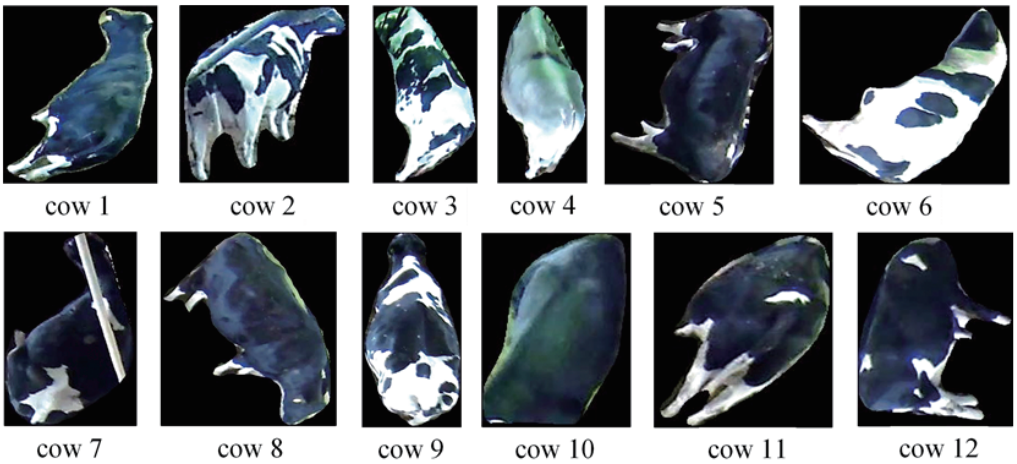


FIGURE 3. Cow dataset: Each image represents a class

TABLE 3. Experimental results

(a) Accuracy comparison for colour spaces and distance measures using colour moments

	Accuracy				
Distance / Colour space	**Cosine**	**Manhattan**	**Chebyshev**	**Euclidean**	**Chi-square**
RGB	0.9776	0.9822	0.9718	0.9773	0.9853
YCbCr	0.9784	0.9775	0.9788	0.9787	0.9802
HSV	0.9599	0.9718	0.9440	0.9571	0.6688
XYZ	0.9609	0.9652	0.9651	0.9648	0.8239
CIELab	0.9866	0.9829	0.9818	0.9839	0.8643
Gray Level	0.8843	0.9635	0.9658	0.9652	0.9645

(b) Accuracy comparison for colour spaces using CM features

Colour space	**Accuracy**
RGB	0.9650
YCbCr	0.9678
HSV	0.9488
XYZ	0.9666
CIELab	0.9408
Gray Level	0.9429

(c) Accuracy comparison for pre-trained networks using CNN features and SVM

Pre-trained networks	**Accuracy**	**Pre-trained networks**	**Accuracy**
squeezenet	0.8429	xception	0.8748
googlenet	0.7984	incetpionresnetv2	0.8640
inceptionv3	0.8437	shufflenet	0.8827
densenet201	0.9038	darknet19	0.8204
mobilenetv2	0.8288	darknet53	0.8843
resnet18	0.8182	alexnet	0.8086
resnet50	0.8739	vgg16	0.7675
resnet101	0.8768	vgg19	0.7829

with colour moment features. Table 3(b) presents the accuracy values for all colour spaces using CM features, and the YCbCr colour space achieves the highest accuracy. Table 3(c) compares the accuracy of different pre-trained networks for CNN feature extraction and classification with SVM. The results indicate that densenet201 is the most suitable network for object recognition.

4. **Conclusions.** This study presents a comparative analysis of various colour spaces, distance measures, and deep learning networks for cow recognition. Our findings suggest that the CIELab colour space with cosine distance measure for colour moments, the YCbCr colour space for CM features, and the densenet201 network for CNN feature extraction are the optimal representations for cow recognition. In future studies, we plan to combine multiple feature representations to identify the best combination for even more accurate recognition results. Moreover, based on this analysis of different comparisons, we aim to build a robust object tracking system to provide cost-effective monitoring of dairy cattle farms.

Acknowledgment. We thank K. Honkawa, D. V. M., for giving us every convenience of studying on the ranch and for his valuable advice.

REFERENCES

[1] T. T. Zin, C. N. Phyo, P. Tin, H. Hama and I. Kobayashi, Image technology based cow identification system using deep learning, *Proc. of the International MultiConference of Engineers and Computer Scientists*, vol.1, pp.236-247, 2018.

[2] M. Lee and S. Seo, Wearable wireless biosensor technology for monitoring cattle: A review, *Animals*, vol.11, no.11, 2779, 2021.

[3] T. T. Zin, P. Tin, I. Kobayashi and H. Hama, Markov chain techniques for cow behavior analysis in video-based monitoring system, *Proc. of the International MultiConference of Engineers and Computer Scientists*, vol.1, 2018.

[4] A. Biglari, *A Vision-Based Cattle Recognition System Using Tensor-Flow for Livestock Water Intake Monitoring*, Ph.D. Thesis, New Mexico State University, 2022.

[5] Z. Weng et al., Facial recognition of dairy cattle based on improved convolutional neural network, *IEICE Transactions on Information and Systems*, vol.105, no.6, pp.1234-1238, 2022.

[6] M. Bansal, M. Kumar and M. Kumar, 2D object recognition techniques: State-of-the-art work, *Archives of Computational Methods in Engineering*, vol.28, no.3, pp.1147-1161, 2021.

[7] C. C. Mar, T. T. Zin, I. Kobayashi and Y. Horii, A hybrid approach: Image processing techniques and deep learning method for cow detection and tracking system, *2022 IEEE 4th Global Conference on Life Sciences and Technologies (LifeTech)*, pp.566-567, 2022.

[8] K. Chen, J. Pang, J. Wang, Y. Xiong, X. Li, S. Sun and D. Lin, Hybrid task cascade for instance segmentation, *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.4974-4983, 2019.

[9] M. Chandrajit, R. Girisha and T. Vasudev, Multiple objects tracking in surveillance video using color and Hu moments, *Signal & Image Processing: An International Journal*, vol.7, pp.15-27, 2016.

[10] S. Sergyan, Color histogram features based image classification in content-based image retrieval systems, *2008 6th International Symposium on Applied Machine Intelligence and Informatics*, pp.221-224, 2008.

[11] B. Khaldi, O. Aiadi and M. L. Kherfi, Combining colour and grey-level co-occurrence matrix features: A comparative study, *IET Image Processing*, vol.13, no.9, pp.1401-1410, 2019.

[12] D. O. Aborisade, J. A. Ojo, A. O. Amole and A. O. Durodola, Comparative analysis of textural features derived from GLCM for ultrasound liver image classification, *Energy*, vol.2, no.10, 2014.

[13] F. de Lima Weber et al., Recognition of pantaneira cattle breed using computer vision and convolutional neural networks, *Computers and Electronics in Agriculture*, vol.175, 105548, 2020.

[14] K. K. Anilkumar, V. J. Manoj and T. M. Sagi, Automated detection of leukemia by pretrained deep neural networks and transfer learning: A comparison, *Medical Engineering & Physics*, vol.98, pp.8-19, 2021.