Video-based Prediction for Header-height Control of a Combine Harvester

He Liu, Amy R. Reibman, Aaron C. Ault, and James V. Krogmeier
Department of Electrical and Computer Engineering, Purdue University, West Lafayette, USA

***Abstract*—Many automation applications in the agriculture industry focus on navigation or steering control. But there are few studies on functional controls such as header automation. This paper introduces a video-based prediction system for header-height control of a combine harvester. To achieve this goal, we propose a lighting-invariant spatial segmentation method to locate the field region. Crop presence detection is performed by training a classifier on texture features and the percentage of crops in the field can be estimated. Then the time to lift the header is predicted based on observing the trend of crop presence. The framework is tested on both bean and wheat harvesting video sequences and the decreasing crop percentage can be successfully estimated.**

***Index Terms*—video processing, combine harvester, automation**

I. Introduction

Autonomous driving is one of the most popular applications in the engineering industry. One objective of autonomous driving in an urban environment is to drive safely towards a target position and avoid all obstacles in the road [1]. To detect potential obstacles, different sensors such as LIDAR, RADAR, cameras and GPS are applied to sense the surrounding environment [2]. Automating farming vehicle also requires basic self-driving capability in the field, where steering automation control mainly focuses on alignment with field edges, such as [3] [4]. However, compared to autonomous driving for cars, farming vehicles automation has additional challenges.

Beside driving control, the machine operator also needs to perform the required farming work, for example during harvesting operations. For a combine harvester, the front reel (header) position should be adjusted based on crop[1] conditions in the field. However, there are not many studies on automating the front reel header in the literature. Towards the goal of fully automated farming vehicles, this paper explores the header-height prediction for a combine harvester.

In the combine harvester, the reel is a rotating wheel-like device that is attached to the front header that pushes the crop from the field. The header-height controls the position where the crops are cut off above the ground. In general, while the combine is harvesting, the header should be low to the ground so that no useful crops are left behind; when there are no crops in front, the header should be lifted to avoid potential damage [5]. Previous efforts on header control, such as [6] [7], try keep the header at a desirable constant height despite machine vibrations. However, they do not consider the additional adjustments described above that the farmer must still make. In this paper, we assume the combine is harvesting with the header in a lowered position and our goal is to predict the time when the reel should be lifted.

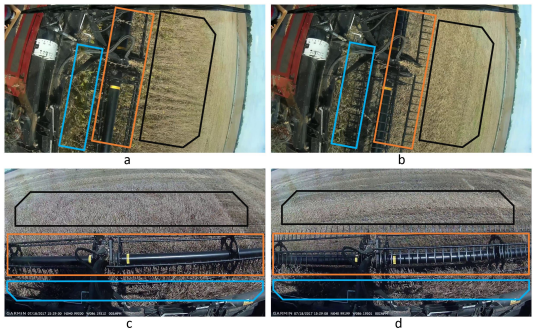

Fig. 1. Sample frames captured by dash camera mounted on combine harvester in bean field (top row) and wheat field (bottom row). Black box shows the front field region, orange box is the front reel and blue region is the conveyor belt carrying cut wheat or beans. On each row, left and right images shows the crop field and empty field in black box respectively.

The cue to lift the header is based on monitoring the upcoming field region of the combine harvester. Cameras have been used in agriculture. For example, [8] shows that the control systems of farming machines need vision-based guidance for monitoring nearby field conditions. In our settings, dash cameras are mounted on the front window by farmers. The cameras not only capture the field region, but also the header and other regions shown in Figure 1. The orange region shows the front reel (header) of the combine and the blue region is the conveyor belt that carries the cut crops into the middle. The field region, which we would like to monitor, is outlined in black. Therefore, the first task is to segment out the spatial position of the field region in the frame. This task is also useful for other operations, such as detecting the obstacles or measuring crop row alignments.

Spatial segmentation in agriculture videos is challenging. The unconstrained capturing environment, various video motions, and the limited amount of data are three general problems for processing videos of different agricultural activities [9]. Color-based segmentation methods rely heavily on lighting conditions in the outdoor environment. As a result, they are not effective when there are shadows or window reflections in the video frames. Instead of color, [9] uses video motion to select spatial regions of farming

This work was supported by the Foundation for Food and Agricultural Research Grant #534662 and the Open Ag Technologies and Systems (OATS) Center.

[1]In this paper, crop refers to different plants in the field, such as wheat, soybeans or corns.

videos. For a video captured on a straight moving vehicle, the field region should have the most consistent forward motion. And the motions generated from other regions are not consistent across time. As shown in Figure 1, the reel (orange) is rotating quickly and the conveyor belt (blue) is carrying the cut wheat or beans to the center. Both regions generate chaotic motions and occupy a large area of the frame.

In our application, the goal of monitoring the field region is to detect whether there are crops in front, and we use the fraction of the visible field that still has crops present to control the header position. Within the field region, we can classify the crops versus the empty field by training image classifiers. However, this image classification task is different than other typical classification problems. First, the camera is randomly mounted on top of the combine and it is far from field region. This long distance causes blurriness on the field region in the image. Secondly, from the classification point of view, the crop region and the empty field are very similar. As shown in Figure 1, the cut and uncut crop share a similar color for both wheat and beans. Even in the same farm field, the crops from different field regions can appear different because of the weather, lighting or the local soil condition.

In this work, we propose a novel video processing system for a combine harvester, which automatically predicts the time when the front reel should be lifted. There are three major contributions. Firstly, we propose a novel framework to automate the header of the combine harvester using video data. Secondly, we design a two-step spatial segmentation method for detecting the field region and we show it is robust to outdoor illumination changes. Thirdly, a crop presence classifier is designed to estimate the crop percentage in each frame.

This paper is organized as follows. In Section II, we review the previous farming automation systems and some related image processing techniques. Then we explain the proposed framework in detail, followed by introducing our field segmentation method and crop estimation method in Section III. The experiments of the crop classification and the header-height prediction are presented in Section IV. Finally we summarize the paper and discuss possible future work in Section V.

II. Previous Work

Autonomous driving systems such as [10] and [11] apply video-based approaches for steering control. For farming vehicles, video has mostly been applied for vehicle alignment automation. [8] develops a video-based algorithm to detect the lateral cutting edges when chopping corn. [12] applies motion analysis to adjust the tractor steering in the corn field. In addition, for the corn field, the Hough transform is used in [13] to estimate the most efficient chopping route. Besides videos, [14] reviews autonomous farming applications for tractors including navigation and steering control. But for studies related to the header of combine such as [5]–[7], they only focus on keeping the header at a constant height relative to the ground. They do not try to predict when the header-height should be adjusted.

Spatial segmentation is important because it allows us to distinguish the relevant visual cues from the proper area. Color-based graph cut methods such as [15]–[17] are popular for still image segmentation. But as we shown later in Section III, color features are not robust under different illuminations. [18] improves the graph cut method by adding temporal connection to the initial graph, which better separates the front object and the background. But as they stated, sensitivity to illumination changes is still a limitation of their method. [19] also uses motion information for segmentation. Their method segments out the foreground region by assuming only the foreground has large motion. But the desired field region in our application has smaller motions, which would be ignored by their method.

Identifying whether a crop is present or not is also an important component of controlling the header-height. We refer to this process here as crop presence classification. Feature selection is critical when performing this process. Interest points, color-based features and texture-based features are widely used in image classification. For dash camera videos in the farming environment, it is hard to find rich and consistent interest points in the field [9]. Also color features such as Color Co-occurrence Matrix (CCM) [20] and Content-Based Image Retrieval (CBIR) [21] are not effective due to lighting issues. Texture features are more robust and in this work, we apply the Local Binary Pattern (LBP) [22] and the Gray Level Co-occurrence Matrix (GLCM) [23].

III. Proposed Method

In this section, we first introduce the video-based reel prediction framework, and then explain the two-step field segmentation method. The crop presence classification and crop percentage estimation are discussed next.

A. Prediction framework

The goal of this framework is to analyze the field region in front of the combine harvester and predict when the header should be raised. As mentioned in Section I, we assume that the combine harvester is in its normal harvesting state, which means the front reel is rotating and harvesting crops. The output of the system is the predicted future time that indicates when the reel should be lifted. Furthermore, we assume the operator is making correct adjustments while operating the combine; therefore, the video contains the ground truth for the correct time to raise the header. Then the predicted time can be validated.

Figure 2 shows the block diagram of the system. The system takes a block of video frames with length T as input. A spatial segmentation method is applied to detect the spatial position of the field. The field is first segmented with a coarse segmentation method based on motion consistency. Then we generate one single field region mask for

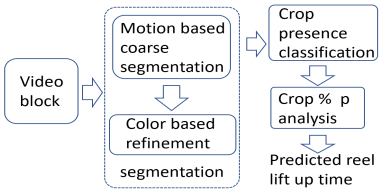

Fig. 2. The prediction framework. The dashed block represents the two-step field segmentation process.

all the frames in the block. Next we apply this mask on all the frames and learn their own specific color distributions. These color features are further used to refine the field segmentation results. Finally the segmentation process generates a field region mask for every image frame in the video block.

A pre-trained classifier is used to separate the crop region from the empty field. With these labels, we compute one percentage value p_t for frame t to represent the amount of crops in its field region, where $t \in \{1,2...T\}$. Then, the percentage of crop in the future, $p_{t+\Delta T}$, can be estimated by fitting a curve to the previously-computed p_t as a function of time. From this fitted curve, the desired time to lift the header can be estimated.

B. Two-step spatial segmentation

The goal of this section is to isolate the field region in front of the vehicle, which is the left dashed block in Figure 2. Color is useful for separating out the field, but it varies in different frames. Inspired by the idea from [9], motion analysis could provide a coarse but robust location of the field. Compared to the chaotic rotation in the header region, the motion in the field region is more consistent across time. So we analyze the consistency of motion at every spatial position in the video frame.

The motion consistency measure is based on optical flow analysis. The consistency measure C of a spatial position in a video block is computed by equation (1),

$$C = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\left[\frac{du_t}{dt} < \theta\right] \cdot \mathbb{1}\left[\frac{dv_t}{dt} < \theta\right] \quad (1)$$

where (u_t, v_t) is the optical flow at frame t, $\mathbb{1}$ represents the indicator function and θ is minimum motion threshold. This measure is counting the temporal gradient of motions, and it is based on the assumption that there are no sudden motion changes in a short time period. A small value of the measure represents the case when the motion at this position is rapidly changing, which corresponds to the header region. Large consistency means that this point has small motion changes or no motion at all. An extra motion magnitude thresholding process is added to eliminate the no-motion region such as the sky.

The above segmentation method only provides a coarse detection of the field, and typically only identifies the closer part of the field. But this incomplete field region helps to pinpoint the specific color distribution of the field in that particular frame. So in the second step, a RGB-based color histogram is generated from the coarsely-segmented field as reference, and we search the rest of field by comparing the local histogram with the reference. Based on the spatial connectivity, we follow a region growth searching method and start the refinement process from the coarsely-segmented mask.

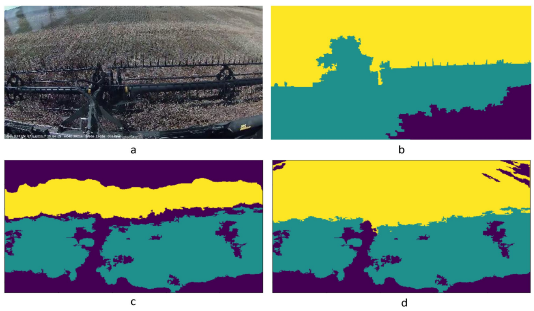

Fig. 3. The field segmentation result: yellow region is the target field region, green shows the front reel; a: the original image, b: segmentation result using [18], c: coarse segment result from motion consistency measure, d: segmentation after color refinement.

Figure 3 shows an example of the field segmentation result. Notice that Figure 3b using [18] is based purely on color information, but the green region mixes the field and header incorrectly because of the dark illumination. Figure 3c shows the result of coarse segmentation and it can be observed that the faraway field region is not included. After performing color-based refinement shown in Figure 3d, the field region has grown to include the whole field region. Notice the reel region (green) is not the target of this process, so no further refinement is performed.

C. Crop presence classification

The goal of this section is to analyze the field in front of the machine and estimate if there are crops to harvest. Since we want to measure the number of crops, directly classifying the whole field region is not possible. So we further divide the field region into smaller overlapping squares, which are used as the basic unit in this classification.

We extract features from the divided squares. As shown in Figures 1 and 4, comparing the fields in the left and right columns, texture is more effective than color to separate the crops and the empty field. Between different texture features, we choose GLCM feature for this classification process because it better captures the directional textures in the field. As we show in Section IV, three different features including the color-based feature CBIR, the texture feature LBP and GLCM are all tested, and GLCM significantly outperforms the other two. In the implementation of GLCM, we apply the method from [24]: four possible directions of neighbor pixel pairs are collected in each square, and the histogram contrast and homogeneity are measured as the feature vector. Each square is described by an eight-dimensional feature, and all features are trained by a decision tree classifier.

D. Crop percentage analysis

The goal of this section is to estimate and predict the crop percentage over time. We use the pre-trained classifier mentioned above to classify every divided square into either

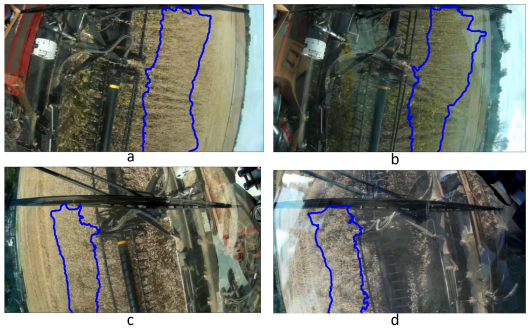

Fig. 4. Illumination variation of two video sets. The combine in the bean harvesting video (first row) is driving to the right and is driving to the left in the wheat harvesting videos (second row). The left column images are from the training set and right column are from the testing set. Notice the classifier is trained on squares which are selected from the coarsely-segmented field (blue).

crops or empty field. Then the results of this classification process are merged to generate one probability map of crops using a voting method: the probability of each field pixel that contains crop presence is voted by the number of squares which include that pixel position. In this case, each value in the probability map represents the confidence that the crop is present in that position. The weighted sum of the probability map is computed as the crop percentage p_t for that frame. Then we fit the percentage p_t as a function of time using a sigmoid function (2)

$$\hat{p}_t = \frac{s_0}{1 + e^{-\frac{s_1}{t - s_2}}} + s_3 \tag{2}$$

where all the s are parameters. This fitted curve represents the trend of the crop percentage over time. Based on the curve, the percentage at a future time $\hat{p}_{t+\Delta T}$ can be estimated. Also the time that the header needs to be lifted t_{up} can be predicted with equation (3),

$$t_{up} = arg(\hat{p}_t \to 0) + T_{delay} \tag{3}$$

which equals the time when the percentage reaches 0, plus some delay time T_{delay}. This delay time is a constant and depends on the speed of the vehicle. It represents the delay starting from the time when the camera is unable to see the crops close to the vehicle, until the machine harvests all the crops in front. We suggest T_{delay} equal 2 seconds in our implementation.

IV. Experiments

In this section, we first talk about the processing of source video data, and then show the results from the crop presence classifier. Next we discuss the analysis of crop percentage and header-height prediction.

A. Video preparation

All the source videos are collected from two different farms during a 3-year period. These videos are captured by a dash camera mounted on the front window of a combine harvester. We call a sequence of videos captured from the same day a set, and such videos normally have the same frame structure, because the camera does not move once mounted by the farmer that day. All the videos have resolution of 1920*1080 at 30 frames per second. Two sets of videos are selected in the experiment: one is on a soybean field and another is on a wheat field, as shown in Figure 4. For each set, we temporally segment the raw videos by hand and select those clips which contain a transition from crops to empty field. In total, we include 12 transition clips from the beans set and 20 clips from the wheat set.

TABLE I
Crop presence classification results. Notice the number below features is the feature dimension.

Feature	Trained dataset	Test on beans set	Test on wheat set
CBIR [21] (14D)	Both	0.498	0.641
	Beans	0.601	0.486
	Wheat	0.651	0.671
LBP (8D)	Both	0.601	0.783
	Beans	0.633	0.506
	Wheat	0.540	0.856
GLCM (8D)	Both	0.809	0.902
	Beans	0.961	0.712
	Wheat	0.655	0.935

Since the training and testing data are chosen from the same set of video frames, for validation we need to guarantee that the source frames of the training data and testing data are well separated in time. The different lighting conditions are considered as the separation boundary. Based on the illumination differences, we specify the first 9 clips in bean set and first 16 clips in wheat set as the source of training data, and the rest are the source for testing data. Several sample frames are shown in Figure 4. In the set of beans (top row), the training frame Figure 4a is much brighter than the testing frame Figure 4b, and the color of the fields are different. The lighting conditions are also different in the wheat frames set (bottom row), and the testing frame Figure 4d even contains a large area of glare.

After the training and testing sources are prepared, we label the ground truth of crops (positive) and empty field (negative) for all source data. Every training or testing transition clip has around 300 frames and they all have a decreasing trend of crops. So each transition clip produces one positive and one negative video block. For each clip, we label the first 30 frames as a positive video block and last 30 frames as a negative video block.

We prepare one pair of training data and testing data on the beans set, and another pair of training and testing data on the wheat set. There are 18 training video blocks and 6 testing blocks for the bean set, and 32 training blocks and 8 testing blocks in the wheat set. Each video block has 30 frames.

B. Crop presence classification

The basic units of the classifier are the squares divided from the field region in the image frames. To generate the squares from the labelled video blocks, first, the field region is located in each frame by performing the spatial segmentation method described above. Notice that to increase processing speed during segmentation, the frames are down-sampled for optical flow analysis. Then the segmented field mask is upsampled back to the original size, in

order to extract texture features in high-resolution images. Here we only use the motion-based coarse segmentation method to locate the closer field region. There are two reasons for this. Firstly, as described in Section I, the closer field region provides more robust texture features than the blurry faraway field. Secondly, the goal of this detection is to predict the crop percentage for the near future, so the distant field is less critical. The blue boxes in Figure 4 show the coarsely-segmented results of each frame.

Next the field region of each frame is divided into squares of 100 by 100 pixel with overlap of 50 pixels. Notice the pre-labelled positive video blocks only produce positive squares and all negative video squares are from negative video blocks. In total we prepared more than 60,000 squares for each training set. But for each video set, we randomly select 5000 positive squares and 5000 negative squares for training because the squares overlap.

To compare different features, for each square we extract three different types of features: CBIR, LBP and GLCM. CBIR computes the histogram in HSV color space with a fixed-bin range. The LBP feature is formed by computing the histogram of normalized LBP at every position in the square. GLCM is computed with the method in Section III. For each type of feature, we train three different classifiers. One general classifier is trained on the merged bean and wheat sets and other two classifiers are individually trained: one on the bean set and other on the wheat set. We generate two groups of testing data in the same way (one from each set). Finally all the classifiers are tested on both test groups.

The classification results are shown in Table I. This is a two-class classification so the base line accuracy is 0.5. Comparing three different features, the color feature CBIR does not work. LBP has a better performance on wheat than beans, because the wheat field texture is less directional, as shown in Figure 4. GLCM achieves the best performance in both sets of videos.

Comparing the three classifiers trained on the GLCM features, we can observe that the general classifier that was trained on both video sets has a worse performance than the specifically-trained classifiers. Also the classifier trained on beans does not work as well on wheat and vice versa. The reason could be the feature difference between the two types of fields. The specific conditions from different fields make it difficult to generate one single pre-trained classifier that solves the classification problem for all fields. So we need to develop a dynamic classifier that can adapt to the conditions in the specific fields. In future system design, the general trained classifier (as shown in the table) can be applied as an initialization step. Then based on the specific conditions of the field, the classifier can be improved and fine-tuned to achieve better accuracy.

C. Field analysis and prediction

This section shows how we apply the pre-trained classifier to analyze crop percentage and make predictions. This simulates an online field analysis of header-height prediction. For preparation, first, we use the testing transition clips in the bean set to represent the real-time captured videos. All these videos begin with driving towards the end of the field and end with the reel being lifted. We choose the pre-trained GLCM classifier (eighth row in Table I) for this prediction experiment. Notice the performance of the classifier in this section is not comparable with results in Table I, because in this experiment, the classifier is used to analyze all the squares from every frame, but most of them are not labelled. The target of this experiment is to predict the crop percentages across the video, and we apply the decreasing trend of crop percentage in every testing video sequence to validate the experiment results.

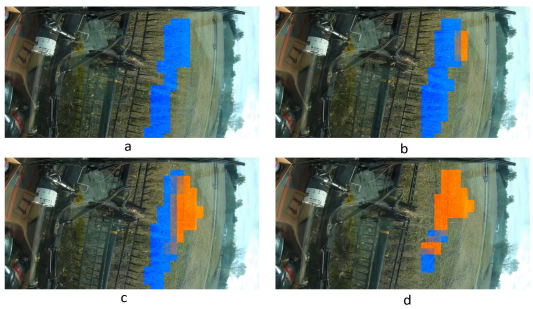

Fig. 5. Probability maps of beans in the field, color represent the probability of beans: blue indicates beans and orange indicates empty field. (a) frame 90, (b) frame 120, (c) frame 150, (d) frame 180.

The experiment steps are explained as follows. We first temporally divide the input testing clips into video blocks and perform spatial segmentation to find the field region. Then for every frame in the block, the classifier is used to separate all the squares that can be extracted from the field region. Next we apply the voting method explained in Section III-D and generate a probability map. This process is repeated for every frame in the whole video sequence, and finally we generate a series of probability maps.

Figure 5 shows some probability maps at frame number 90, 120, 150 and 180 in one testing clip. The colored regions (both orange and blue) represent the coarsely-segmented field region, and the color shows the probability: blue indicates more likely to be crop area; orange indicates no crop. Notice the combine harvester is driving to the right, and the crop region (blue) is gradually shrinking from right to left over time. In Figure 5b and 5c, we can clearly observe the borderline between the crops and the empty field. But in Figure 5d, there are some uncertain regions in the lower half of the field. One possible reason could be the pre-trained classifier has a bias that makes it more effective at classifying crops than classifying empty field.

The crop percentage p_t can be estimated based on the probability maps and we can fit the percentage using equation (2). The percentage plot of one testing clip is shown in Figure 6, where the x axis represents the frame number and the y axis shows the percentage p_t. We can observe that before frame 110, the crop percentage is around 100% and after the decrease between frame 110 to around 220, the percentage reaches 0 after frame 220. The fitted sigmoid

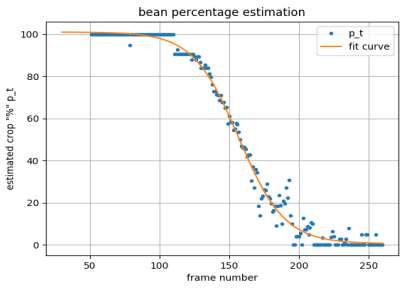

Fig. 6. The fitted curve of crop percentage p_t

function of $\hat{p}_t$ is shown in orange, which correctly captures the decreasing trend of the bean percentage in the field.

Considering the prediction of time to lift the header t_{up}, according to equation (3), we need to find the time when $\hat{p}_t$ reaches to 0. By setting a low cut-off percentage, we estimate that this time is at frame 220. Adding the constant time offset $T_{dealy} = 2$ seconds, the predicted header lifted time is around 280, which agrees with the actual time when the header is lifted. We applied this prediction, both on the 32 transition clips and an additional 38 normal driving clips. All videos have the correct prediction regarding whether the header needs to be lifted. Among 32 transition clips, 29 clips have the correct prediction time.

The performance of our overall system depends on both the classifier and the segmentation process. Errors from the classifier lead to inaccurate probability estimation. The spatial segmentation result is also critical, and the inconsistency of segmentation affects the later steps. The colored (blue and orange) regions from Figure 5 show the coarse segmentation results in consecutive frames. The field region varies slightly between the four different frames, which introduces errors when counting the crop percentage. One reason is that the global motion magnitude at different times is changing, but the threshold value θ in equation (1) is fixed. To minimize the error in future work, designing a dynamic threshold value that is adaptive to the global motion magnitude can smooth the segmentation difference and further improve the overall system robustness.

V. Conclusion

This paper introduces a video-based header-height prediction framework for a combine harvester. The time to lift the header is predicted based on the crop estimation using a crop presence classifier. In addition, a two-step image spatial segmentation method is proposed to search the field region. This segmentation method is robust under varying lighting conditions, which can be applied to other outdoor video processing applications.

For future work, we can further automate the hand labelling work of temporal segmentation in the video preparation step. This would allow us to design a fully automated and adaptive prediction system which directly analyzes raw video sequences. There are some limitations of this work. The camera-based system is only able to predict the time when the reel should be lifted, but the time when reel should be lowered is not considered. Also we only consider the header-height to be in either high or low position, but for an accurate control system, predicting precise height positions needs further exploration.

References

[1] J. Levinson *et al.*, "Towards fully autonomous driving: Systems and algorithms," in *2011 IEEE Intelligent Vehicles Symposium (IV)*, pp. 163–168.

[2] H. Cho *et al.*, "A multi-sensor fusion system for moving object detection and tracking in urban driving environments," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1836–1843.

[3] M. O'Connor *et al.*, "Automatic steering of farm vehicles using GPS," *Precision Agriculture*, pp. 767–777, 1996.

[4] R. Lenain *et al.*, "High accuracy path tracking for vehicles in presence of sliding: Application to farm vehicle automatic guidance for agricultural tasks," *Autonomous Robots*, vol. 21, no. 1, pp. 79–97, 2006.

[5] Y. Xie *et al.*, "Fundamental limits in combine harvester header height control," *Journal of Dynamic Systems, Measurement, and Control*, vol. 135, no. 3, p. 034503, 2013.

[6] G. Lopes *et al.*, "Ae–automation and engineering technologies: Optimal header height control system for combine harvesters," *Biosystems Engineering*, vol. 81, no. 3, pp. 261–272, 2002.

[7] Y. Xie and A. Alleyne, "Two degrees of freedom control for combine harvester header height control," in *ASME 2012 5th Annual Dynamic Systems and Control Conference joint with the JSME 2012 11th Motion and Vibration Conference*. American Society of Mechanical Engineers, 2012, pp. 539–547.

[8] E. Benson, J. Reid, and Q. Zhang, "Machine vision-based guidance system for agricultural grain harvesters using cut-edge detection," *Biosystems Engineering*, vol. 86, no. 4, pp. 389–398, 2003.

[9] H. Liu *et al.*, "Video classification of farming activities with motion-adaptive feature sampling," in *IEEE 20th International Workshop on Multimedia Signal Processing (MMSP)*, 2018.

[10] H. Xu *et al.*, "End-to-end learning of driving models from large-scale video datasets," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2174–2182.

[11] M. Bojarski *et al.*, "End to end learning for self-driving cars," *arXiv preprint arXiv:1604.07316*, 2016.

[12] N. Sainz-Costa *et al.*, "Mapping wide row crops with video sequences acquired from a tractor moving at treatment speed," *Sensors*, vol. 11, no. 7, pp. 7095–7109, 2011.

[13] F. Rovira-Más *et al.*, "Hough-transform-based vision algorithm for crop row detection of an automated agricultural vehicle," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 219, no. 8, pp. 999–1010, 2005.

[14] M. Li *et al.*, "Review of research on agricultural vehicle autonomous guidance," *International Journal of Agricultural and Biological Engineering*, vol. 2, no. 3, pp. 1–16, 2009.

[15] B. L. Price *et al.*, "Geodesic graph cut for interactive image segmentation," in *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2010, pp. 3161–3168.

[16] P. Arbelaez *et al.*, "Contour detection and hierarchical image segmentation," *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 898–916, 2011.

[17] P. F. Felzenszwalb *et al.*, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, 2004.

[18] M. Grundmann *et al.*, "Efficient hierarchical graph-based video segmentation," in *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2010, pp. 2141–2148.

[19] W. Wang *et al.*, "Saliency-aware geodesic video object segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3395–3402.

[20] C.-H. Lin *et al.*, "A smart content-based image retrieval system based on color and texture feature," *Image and Vision Computing*, vol. 27, no. 6, pp. 658–665, 2009.

[21] J. Yue *et al.*, "Content-based image retrieval using color and texture fused features," *Mathematical and Computer Modelling*, vol. 54, no. 3-4, pp. 1121–1127, 2011.

[22] T. Ahonen *et al.*, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 12, pp. 2037–2041, 2006.

[23] A. Baraldi and F. Parmiggiani, "An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, no. 2, pp. 293–304, 1995.

[24] M. Hall-Beyer, "GLCM texture: A tutorial v. 3.0 March 2017." https://prism.ucalgary.ca/handle/1880/51900.