Road Recognition System with Heuristic Method and Machine Learning

Hagai Raja Sinulingga
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Medan, Indonesia
hagairaja@gmail.com

Rinaldi Munir
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
rinaldi@informatika.org

***Abstract*—Road recognition is one of essential information for determining an Autonomous Vehicle movement. Latest research has shown that machine learning could be used to obtain the information from images. Nevertheless, the system could be improved by effectivity and efficiency. This research proposed finding better feature combinations and using Artificial Neural Network algorithm to build higher accuracy road detection model for better effectivity. Region of Interest module using heuristic method also applied to reduce computation for better efficiency. These three new modules are implemented and combined with road recognition module to become road recognition system. The proposed method performance then tested and compared with the latest research. The experiment results shown that Artificial Neural Network cannot increase the system effectiveness. Nonetheless, with right feature and region of interest module, the proposed system successfully gives better performance. The prototype has accuracy increased from F1-score 0,94 to 0,95 and speed increased from 99 to 112 frames processed per second.**

Keywords*—*Road recognition, machine learning, artificial neural network, autonomous vehicle

I. INTRODUCTION

Like a driver, Autonomous Vehicle (AV) need some essential information to make decision when moving on the road. The more accurate and complete information that AV has, the more error-free the decision will be so it can prevent car accident. Some of the essential information are the road information, surrounding vehicle, traffic sign and lights, and pedestrian or other things around the AV. This research will be focused only for the road recognition information. It includes the detection of the road on the image and determining what is the road based on the detected road. Obtaining the information for AV must be done quickly because driving is a continuous process, so if there is lack of speed on the retrieval could result a car accident itself.

Nowadays, there are many researches have been done trying to solve the road detection problem. The solutions can be grouped into two big approach which are heuristic and machine learning [1]. Heuristic method is image processing techniques that making direct pixel changes on image. In other hand, machine learning method used statistical theory for machine could learn from data to perform a decision. For the heuristic approach, there are researchers who utilize lane following assisted by Otsu's method and Hough's transformation method [5] but this kind of approach has a shortcoming when the road have unclear road lines, while these are commonly found on Indonesian roads. There are also researchers that detect through texture information [6] and road color using Color Space [4] as well as RGB Histogram approaches[3]. Unseparable from shortcomings, this approach still unable to handle the problem when there are changes of color characteristics of the road such as when it is bright or dark. Also, the problem when there are texture changes of the road such as the presence of holes or puddles.

Looking at the many factors that can affect a successful road detection heuristically, came the idea to solve it through machine learning. This approach is carried out because with machine learning decisions taken by the program can be adjusted to the current conditions. Recent research conducted shows that the performance using several machine learning algorithms such as KNN, Random Forest, SVM gives satisfactory results with an average F1-Score of 0.958 and execution time per frame of 16ms for the Random Forest model [1].

However, this can still be improved in terms of accuracy, namely the re-election of features as machine learning inputs. In addition, in terms of execution time can also be streamlined, namely through the determination of the Region of Interest (ROI) so that not all parts of the image must be classified. This feature extraction is done by using the heuristic method and assisted by machine learning while for the determination of ROI is done simply by the heuristic method. Lastly, in terms of the machine learning model itself, Winarto is still limited in comparing old algorithms such as KNN, SVM, and Random Forest and has not made comparisons with newer neural network learning algorithms and has the same relative level of computation for ever greater amounts of data. This is reasonable because in Winarto's research, he is still doing learning with a small amount of data which is a total of 300 images and not the streets of Indonesia. For this reason, an analysis will be conducted on road features that are more precise in detecting road, determining ROI, applying heuristic methods, and comparing them with neural network-based machine learning to help improve the accuracy and speed of model predictions. The neural network algorithm used in this study is the simplest form, Artificial Neural Network (ANN). This was chosen because ANN has low complexity and is easy to implement. The classification model and methods designed by the experimental process will be tested with the methods used in previous studies [1] to determine whether there is an improvement in the system both in terms of accuracy and execution time.

The system will be built following the method of building a general road recognition system as practiced by Hazeem [5]. The system will begin with road detection and then determine the way and last display it again as video. Thus, the road and the way information could be used by the AV.

978-1-7281-8038-0/20/$31.00 ©2020 IEEE

II. RELATED WORKS

A. Road Texture Approach

Graovac, et al [2] detected roads with a heuristic approach using pattern matching on the texture of the road. The path texture differentiator value is extracted from N * N pixels by finding the Standard Deviation values of light intensity, Third Moment, Uniformity, and Average Entropy of the accumulative amount of gray starting from 0-255 and based on the pixel position by calculating Contrast, Inverse moment of the first order, Uniformity, Entropy. Furthermore, a sampling of parts of the image consisting of the road, background, and sky are carried out.

Graovac did not show the accuracy of the results of his research. However, the conclusion from the research that was conducted, it would be difficult if in the picture there were various textures other than the background and sky such as pedestrians, walls, etc. because they had to find a threshold value for each class feature. This research is also still being done on static images.

B. Histogram Approach

Enjat M, et al [3] in their research carried out road detection with a heuristic approach using image histogram values. First, the image is converted into a grayscale image for each RGB value, i.e. an image with one color channel with a range of values 0-255. Then from the histogram, the peak value of the dominant color is considered to be the color of the road.

Research by Enjat still requires further research related to conditions in low light conditions. In addition, in this study itself, it is assumed that the dominant road color and different from the background color. This is quite difficult to maintain because the AV may be going to run on areas that have the same colored background as the road such as a house wall or when the AV is moving on an inclined road so that what looks dominant is the sky.

C. Color Space Approach

Komang, et al [4] in their research also did a heuristic approach that tried to make a better method than template matching. The proposed method is to compare the value of the color range in the block with the road sample in general. The color range value is the minimum-maximum distance value of an image channel.

The idea remains the same, namely by dividing the image into blocks, then detecting and checking the correlation between segments classified as roads. This research tests the value of Precision, Recall, F-Measure, and Accuracy. The results show that the method proposed by Enjat is better than the template matching technique. However, with this approach, there are still problems when the brightness level of the image changes resulting in the color range value changing and the background color condition that has the same color range as the color range of the road sample so that the detection results become inaccurate.

D. Machine Learning Approach

Winarto [1] in October 2019 had just finished conducting road detection research using the machine learning approach. Winarto in his research compared what algorithms have high accuracy in classifying small blocks of an image to detect roads. The algorithm used by Winarto is KNN, Random Forest, and SVM. In his research, Winarto has used block normalization, main color value, neighbors color value, color type and local binary pattern as image extracted features.

In his research, Winarto tried to detect three parts of the road which are road, roadside, and background. To detect these three parts, Winarto divides the two road detection cases by classifying two classes and three classes. Both have road and background classes, but in the case of three classes, there are roadside classes. For the case of the two classes referred to as hybrids because according to Winarto actually the roadside class can be derived from the border between the road and the background. In other words, for the hybrid class, the detection of the three parts is done with the assumption, while the detection of machine learning is done through classification. The labeling done for each case is also different but from the same system flow i.e. the model is trained with the image that has been labeled and the model that has been trained is tested on the image of the test path.

In Winarto's research, two aspects were tested, which were accuracy that was seen from the F-1 Score and speed that was seen from how many frames (FPS) were generated in one second. How to calculate the F-1 Score can be seen in equation (1). The TP, FP, FN number is the count of blocks that are fulfilled the criteria. For example TP is when the block is a road and predicted as road or the block is not road and predicted as not road.

$$F-1 = 2 \times (Precision \times Recall)/(Precision + Recall) \quad (1)$$

$$Precision = TP\ /\ (FP + TP)$$

$$Recall = TP\ /\ (FN + TP)$$

$$TP(TruePositive) = Prediction\_true\_actual\_true$$

$$FP(FalsePositive) = Prediction\_false\_actual\_true$$

$$FN(FalseNegative) = Prediction\_false\_actual\_false$$

However, in this research there are still things that are not optimal, in terms of the amount of data and image features used. The images that become training data are still too little, which are 300 images with the case of Indonesian roads only 15 images. There are only 4 features used to detect road and not all of them are combined. In addition, other machine learning algorithms as will be done in this study is Artificial Neural Network (ANN), as a newer machine learning algorithm have also not been tested.

III. THE DATASET

Table 1. Distribution of dataset

Case	Training		Tuning		Testing		Total
	#	%	#	%	#	%	
1. Toll	50	71	13	18	7	10	70
2. Straight	207	71	51	17	30	10	288
3. Intersection three	32	72	8	18	4	9	44
4. Intersection four	19	73	4	15	3	11	26
5. Turn left	33	70	10	21	4	8	47
6. Turn right	93	72	23	17	13	10	129
7. Good road	311	71	78	18	43	9	432
8. Roads are flooded	53	71	13	17	8	10	74
9. Potholes	20	71	5	17	3	10	28
10. The road is quiet	233	71	58	17	33	10	324
11. Crowded road	151	71	38	18	21	10	210
Total image	384	72	96	18	54	10	534

The data used in this study was prepared specifically for the streets of Indonesia. Road images are retrieved from videos taken in front of vehicles on Indonesian roads. There are a total of 536 images taken as datasets that will be used as training data, tuning data, and test data. The image is separated into 11 cases and divided by the number as shown in table 1. The total images is not as the same as the sum of all cases images because of the images on the cases are overlapping.

The image is prepared along with ground truth as the basis of machine learning for learning. Examples of prepared images are shown in Figure 1. It appears that the road is labeled in purple and the background is labeled in red.

Figure 1. Image and its ground truth

IV. PROPOSED METHOD

The stages in building a road recognition system can be seen in Figure 2. These stages can be divided into four major stages. These stages are:

A. Preparation

At this stage, the frame is split in the input video into images. The image with the RGB color type will then be the basic input for processing at a later stage. If the first image has been completed, then continue with the second, third and so on.

B. Heuristic Method

At this stage the image will be processed using heuristic methods. Heuristic methods are image processing techniques that make physical changes directly to the image pixel. This stage start by determining ROI to make the detection of parts more efficient. ROI is horizontally-cutted bottom part of image which contain road. ROI is determined by finding first occurrence of detected road block on the first frame from top of the image. Then for the next frame, the position only check each one block on top and bottom from the previous position to see if the ROI has moved or not. Video is a continuous image, so the movement of the edge of the road is just a slight different from the previous frame.

Then the image is cropped on the ROI line and proceed with the feature extraction stage. At this stage a feature value calculation for images per n x n block is calculated for the features to be used in the classification at the machine learning stage. This feature will later become one of the proper feature considerations in building a road recognition system.

C. Machine Learning

Feature extraction can also use the CNN machine learning algorithm [7]. This feature extracts features from the image using the pretrained VGG16 model. Then of course, at this stage testing will be done for each machine learning method being compared which are Random Forest and Artificial Neural Network. For each method a model is trained using training data and for each model will also be tuned by tuning data and tested using test data. Later the detection results will be compared to the output of each model.

D. Display

After each block has been classified using a machine learning method, the next stage is determining the roadway. The direction is concluded based on the curve that the blocks which determined as road occur. This is done by simple rule and iteration.

Then after the information about the location of the road and the direction of the road has been obtained, the image will be prepared to be displayed again into a video at the post-processing stage.

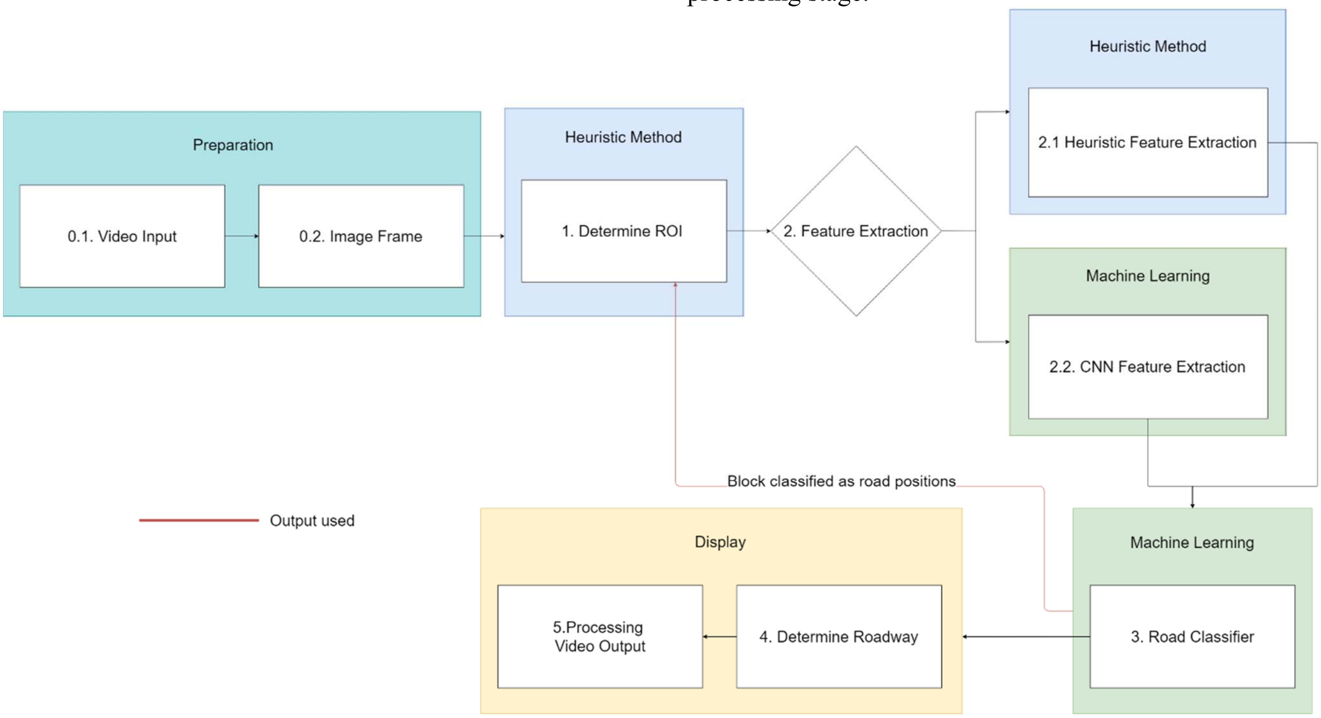

Figure 2. Proposed architecture

V. Evaluation

A. Scenario

Testing is divided into two major stages namely testing the search for the best system solutions and comparison testing with previous research [1]. Both of these tests are carried out in series. In general, the two tests will look at two aspects namely the accuracy from the F1-Score and the speed from the FPS. Testing the search for the best system solution is done to answer the problem formulation about what are the right features in building a machine learning model for road detection and see if ANN can increase the effectiveness of road detection. Whereas comparative testing with previous research was carried out to answer the problem formulation on how to determine the right ROI to improve road detection efficiency.

Table 2. Combination of Features

No	Feature configuration	Notes
1.	CNN features (64 features/block)	The six features are combined into 63 feature combinations
2.	Combination of heuristic features	
	2.1. Normalization (2 features / blocks)	
	2.2. Main color (3 features / blocks)	
	2.3. Neighbors color (24 features / blocks)	
	2.4. Local binary pattern (64 features / blocks)	
	2.5. Log Chromaticity Space (1 features / blocks)	

In testing the search for the best solution, the model development is performed with a combination of features as shown in table 2. Each model built has tuned parameters based on data tuning. The model with the best Random Forest and ANN algorithm for each feature is then tested with each test data for the cases in table 1. Thus there is performance per cases and also overall performance which using the 54 images. The combination of features and algorithms that provide the optimum F1-score and FPS is chosen as the best system solution. Optimum Solution is a solution that is able to process a frame with a fast time so it does not produce a broken effect on the video but also has high accuracy. F1-score is seen from the results of performance in predicting test data while FPS is measured by looking at the average time taken by the model to predict a frame.

Furthermore, in comparison testing with previous research only a comparison was made between the best solution model from previous testing with the model proposed by Winarto [1]. So that the testing between the two models is fair, the Winarto model is repeated learning with the same data used for the best solution model. In addition, the environment used is also equated. After that to see the effectiveness produced, it is done in the same way by comparing the F1-scores of the two models in predicting test data. Meanwhile, to see the efficiency side, it is done by seeing the results of the ROI movement in the road video trailer experiment. Tests carried out on live video to see whether ROI has indeed succeeded in increasing the efficiency of road detection.

B. Result and Analysis

In the first phase of testing, the results show that there are four combinations of features that can produce a maximum F1-Score of 0.98. The combination of features are:

1. Normalization + Neighbors
2. Normalization + Neighbors + LCS
3. Normalization + Neighbors + Main
4. Normalization + Neighbors + LCS + Main

But the drawback of these combination is that their FPS score is very small. The highest is 2.947 FPS with the combination of block normalization (Normalization) and neighboring color values (Neighbors). While one normal video is producing 20-30 FPS. This can produce a broken effect on the detection system when detecting videos. While more and more features are used, the smaller the FPS generated. This can be seen from the FPS of combination 1 is higher than combination 2 (FPS = 2.710) and 3 (FPS = 2.710) and the smallest combination of FPS values is combination 4 (FPS = 2,435).

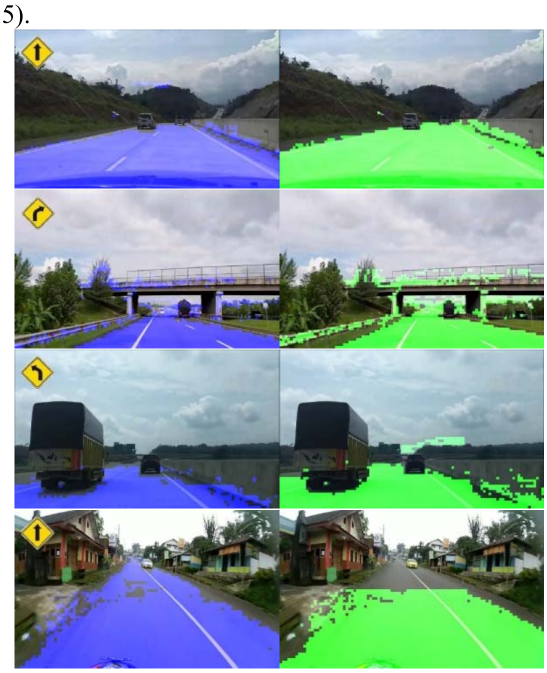

Figure 3. The effect of determining ROI, in order of rows 1-4 is a case of ROI position: in the middle - below (bad) - below (good) – above.

Because of this low FPS value, a combination of features with a significant FPS is sought with the highest F-1 Score and what is found is a combination of block normalization features and the main color value. This combination produces 16.8 FPS and has an F-1 Score of 0.95. The next significant combination was block normalization, main color values (Main) and LCS with 9,882 FPS and F1-Score 0.95. The rest with a higher FPS actually has a lower F1-Score. For this solution the best feature is the block normalization feature and the main color value feature. It turns out that this combination of features is the same as the best feature in Winarto's research.

From the first phase of testing it can also be seen that the CNN feature, which extracted by VGG16 model, has worse performance than the heuristic feature in terms of accuracy and speed. In terms of accuracy, if only using the CNN feature, the result is an F1-Score of 0.91. When combined with the block normalization and neighbors color values, the highest F1-Score is 0.97. While without the CNN feature, the combination of the two features has reached F1-Score 0.98. In terms of speed, models using the CNN feature have the highest FPS of 0.03 with an F1-Score of 0.94 while those without a CNN feature with an F1-Score of 0.95 have a FPS of 16.8 (see Figure 4).

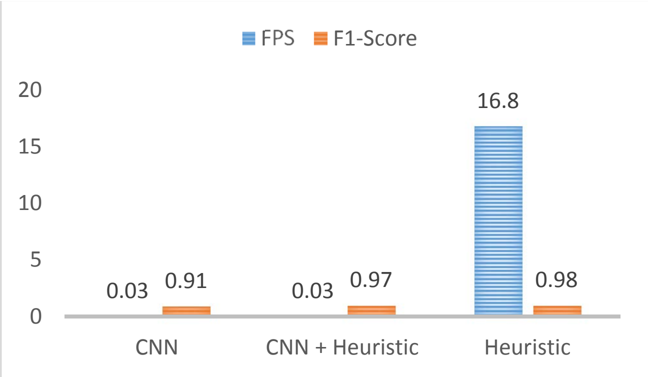

Figure 4. Performance of CNN and heuristic features

Overall, the results of the first stage of the test show the Random Forest algorithm has better accuracy and speed than the ANN algorithm. The Random Forest algorithm has the highest F1-Score of 0.98 with a combination of Normalization + Neighbors features while the ANN algorithm has the highest F-1 Score of 0.95 with several feature combinations. The combination of features for the ANN algorithm with the highest F1-Score and the highest FPS is Normalization + Neighbors + LCS with FPS 2.847 (see Figure 5). Despite the low F1-Score, all feature combinations have a higher ANN FPS algorithm than the Random Forest algorithm. However, because FPS is meaningless without good F1-Score, the best model from the first test phase is using Random Forest algorithm with Normalization + Main colors as the features.

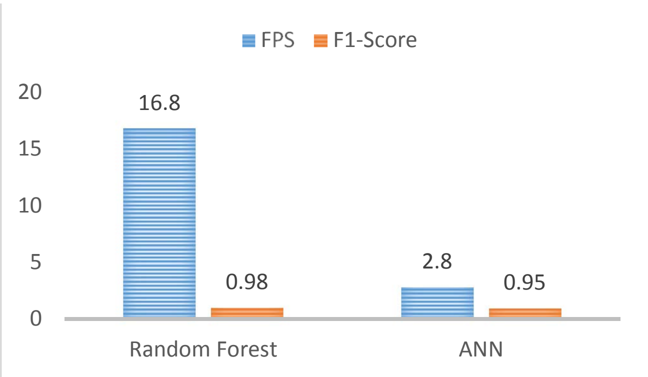

Figure 5. Performance of random forest and ANN

In the second phase of testing, the results show that there were increasing performance from the Winarto model to the best model. Overall the best model has a higher accuracy of 0.95 compared to Winarto 0.94. However, 2 out of 11 cases of imagery obtained Winarto's model has higher accuracy, namely in the case of three intersections and turning left. Nevertheless, this result cannot be used as a benchmark because the number of test cases, each of which is only 4 images so that the potential change in results is still very high if the number of test cases is added. Winarto's model has a lower accuracy than his research which is 0.958 to 0.94 after being tested using this research dataset. In terms of speed in each image case, the best model always has a higher speed than the Winarto model, which is in the range of 8-24 FPS. Overall the image case, the average speed of the best model is 112 FPS while the Winarto model only 99 FPS (see Figure 6). To make it clear, this test focuses on the model determination time and not include the feature extraction process, because the feature is same so it is better to focus on model time only.

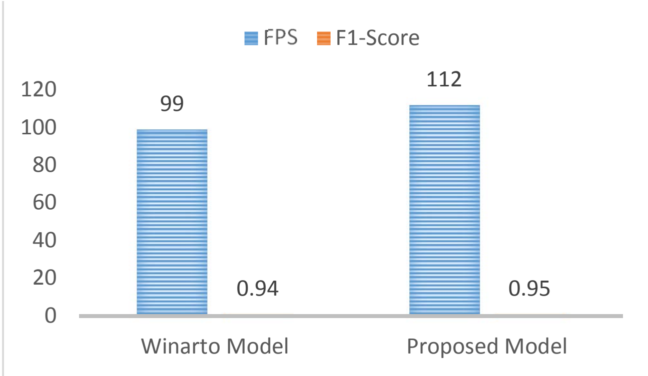

Figure 6. Performance of Winarto model and proposed model

The second stage test results also show that the best model has advantages and disadvantages in detecting the road. The model has high accuracy when detecting roads with characteristics (see Figure 7):

1. The road has a color that is quite different from the background color.
2. The vehicle in the image is not blurred and has a different color from the road.
3. The path in the image gives a consistent color shape to the end of the road.

Figure 7. Sample of good detection.

In general, the model is good at detecting if the image has a clear enough difference between the road and the background. Whereas on the contrary, the model has low accuracy (see figure 8) when detecting roads with characteristics:

1. The image is foggy so as to make the image blurry
2. There is a bridge with a relative position in the middle of the image.
3. There are coloured vehicles similar to the road.
4. Uneven wet streets.

To determine the proposed ROI, it can be seen in Figure 3 that the proposed system has a dynamic ROI while the Winarto system has a static ROI. In this test the F1-Score calculation between the proposed system and the Winarto system cannot be compared because there is no ground truth video. However, it can be seen that for the case of ROI above the static ROI, it is found that road detection has higher accuracy. As for the case of ROI in the middle or the same as the static ROI similar results are obtained. Finally for cases under static ROI, sometimes it results in higher accuracy and sometimes the same. So in terms of overall accuracy, the ROI position case shows that the system has better accuracy than Winarto.

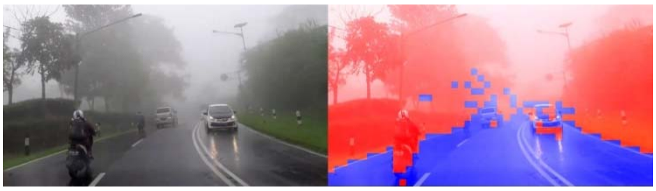

Figure 8. Sample of bad detection.

For the speed seen from the FPS show different results. In the case of roads that have an ROI below the static ROI, a higher FPS is obtained. Conversely, for ROI above static ROI, a lower FPS is obtained. As for ROI that is close enough to static ROI, it has FPS which is also close. This is caused by differences in the number of blocks predicted due to the movement of the ROI position. Even though the ROI above ROI status provides lower efficiency, overall the proposed ROI system provides better efficiency because it can detect all parts of the road in all cases without having to detect all parts of the image. While in that case the static ROI did not succeed in detecting all parts of the road in the image.

The module determining the direction of the road has successfully determined each direction in accordance with the results of road detection. But if the results of road detection have a lot of noise and are not continuous as in the ground truth image, then the determination of the direction of the road also becomes inappropriate. This also results in the determination of the direction of the road to be easily changed from one direction to another depending on the position of noise in the image. These noises happened usually when there is other part of image that has similar colour with the road. It can happen in every cases of road direction. Figure 9 shows an example of the noise effect which causes a change in detection of road direction from straight to turn left.

Figure 9. Effect of noise on determining the direction of the road.

In addition there are also deficiencies in the detection of left turn and right turn. The criteria for determining the direction of the road depend on the shape of the basin that appears at the end of the road. While in some cases of images, the detection system failed to detect the exact end of the road. This causes the module to conclude that the road direction is straight. Figure 10 shows an example of this weakness.

Figure 10. Detection of end of road failed (should be turn left instead of straight).

VI. Conclusion

The prototype of the road recognition system was successfully built to recognize the static video of the road during the day. But for the determination of the direction of the road still has shortcomings caused by an imperfect road detection model. Even so, the road detection system that was built was better than the system proposed in previous research by Winarto. The proposed model is built using the random forest algorithm with the main color value and the block normalization position as features. The model built is better in terms of accuracy and speed compared to the Winarto model. For overall accuracy the Winarto model has an accuracy of 0.94 and the model built was 0.95 with 11 of the 13 cases won by the model built. While the speed for all cases, the model built has a faster time of 112 FPS and 99 FPS Winarto model.

Future studies can do things like: testing at night, using parallel programming, and building prototypes that can process real-time video.

References

[1] Winarto, & DH Widyantoro, "Road and Roadside Detection using Machine Learning Approaches," International Conference on Data and Software Engineering (ICoDSE), 2019.

[2] Graovac, S. & Goma, A, "Detection of Road Image Borders Based on Texture Classification," International Journal of Advanced Robotic Systems, Vol. 9, 242, 2012.

[3] M.D. Enjat Munajat, Dwi H. Widyantoro, Rinaldi Munir, "Road Detection System based on RGB Histogram Filterization and Boundary Classifier," ICACSIS, 2015.

[4] I. Komang S. & Fitri U, "Road Detection Based on the Color Space and Cluster Connecting," IEEE International Conference on Signal and Image Processing, 2016.

[5] Hazeem B. Taher, "Proposed Method for Road Detection and Following Boundaries," Journal of Theoretical and Applied Information Technology (JATIT), University of Thi-Qar, Irak, 2018.

[6] Stevica G, & Ahmed G., "Detection of Road Image Borders Based on Texture Classification," International Journal of Advanced Robotic Systems, 2012.

[7] Rohit Thakur, "Step by step VGG16 implementation in Keras for beginners," Toward Data Science, retrieve from https://towardsdatascience.com/step-by-step-vgg16-implementation-in-keras-for-beginners-a833c686ae6c, 2019.