IMPLEMENTATION OF SUPPORT VECTOR MACHINE AND BACKPROPAGATION CLASSIFICATION METHOD FOR CAR DETECTION USING HISTOGRAM OF ORIENTED GRADIENTS

Muhammad F R kaisupy[1], Galih Hermawan[2]
[1,2] Universitas Komputer Indonesia
Jl. Dipatiukur No 112-116 Bandung. 40132
E-mail : ramdhany.kaisupy@gmail.com, galih.hermawan@email.unikom.ac.id,

ABSTRACT

Detection system is a very helpful technology in various fields. In the field of traffic, detection systems are used for monitoring traffic traffic. Currently, traffic control has utilized CCTV cameras, namely RTMC (Road Traffic Monitoring Center). In addition, research on ITS (Intelligent Transportation System) also utilizes surveillance through the camera.

HOG is used in this study because it is a Feature Descriptor that takes edge or gradient structure that is characterized from local shape or edge direction and with good local gradient intensity distribution. The feature extraction results will be trained using SVM to generate a model for later detection. The detection results will be reclassified by neural network backpropagation.

The test was performed using 400 positive data and 1400 negative data showed an accuracy of 51.8% using 1 neuron on hidden layer.

Keywords: *Backpropagation*, Car Detection, *Support Vector Machine, Histogram of oriented Gradient*

1. INTRODUCTION

Some research on video detection on video ever done one of them is by using backgrund substraction method that is comparing the image of background without object with video, the problem gained in this research is the existence of shadow caused by sunlight about the vehicle will be detected as object.

There are several methods commonly used for detection on image processing, one of which is HOG (histogram of oriented gradient). HOG is used in this study because according to research conducted by ari kurniawan [1] HOG is a Feature Descriptor that takes edge or gradient structure that is characterized from local shape or edge direction and with good local gradient intensity distribution besides research conducted by Vanda Singh [2] The result of the feature of the extraction using histogram of oriented gradient method is better than using LBP.

In the research by Permata Cahyo, I Ketut eddy, and Muhtadin [3] using SVM classification, 99% of the test model accuracy was obtained, while the detection result of the car got 76.17% accuracy, therefore detection system still needed new method to improve the performance Accuracy, some research is also done to improve the histogram detection performance of oriented gradient that is by combining several methods of classification such as research conducted by KP Sanak Kumar and Dr. R. Bhavani [4] where the result of HOG with SVM gets 70.85% accuracy, while HOG with kNN gets 68.21% accuracy, and HOG with SVMkNN gets 72.48% accuracy.

In a study conducted by Indu Dokare and Naveeta Kant [5] comparing the performance of the SVM classification method, KNN and BPNN obtained classification results using BPNN got the highest accuracy, and the kNN classification method received the lowest accuracy. The combination of SVM and BPNN for text classification has been done in the research of Zhang Wen, Xijin Tang and Taketoshi Yoshida [6], where the result of classification using SVM and BPNN is better than using SVM or BPNN.

2. CONTENTS OF RESEARCH

2.1 Digital Image Processing

Image processing is image processing using computer for better quality. The goal is that images can be easily interpreted by humans or computers. In Figure 1 illustrated that image processing techniques manipulate the image into another image. So, the input is the image and the output is also the image. But the output image has better quality than the input image.

In general, operations on image processing are applied to images when [8]:

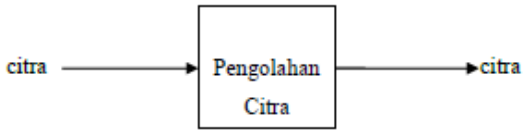


image 1. Image Processing

a. Improvement or image modification needs to be done to improve the quality Appearance or highlight some aspects of the information contained in the image,
b. The elements in the image need to be grouped, matched, or measured.
c. Some imagery needs to be merged with another image. The sub chapter is written with the new times roman font size 10.

Gmabar and the table can be inserted in this format in Figure 1.

2.2 Histogram of Oriented Gradients

Histograms of Oriented Gradients (HOG) is a feature descriptor used in computer vision and image processing for object detection. The HOG method was first introduced by Dalal and Triggs in 2005. The HOG method in this study will be used to test whether the test image is human or not. Figure 2 shows the working step of the feature extraction process of the HOG method.

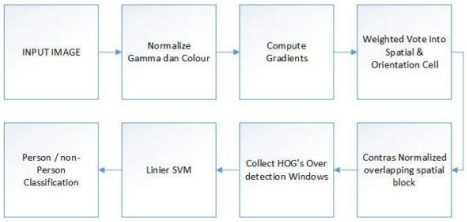


Image 2. HOG processing

2.3 Support vector machine

Support Vector Machine (SVM) was first introduced by Vapnik in 1992 as a harmonious series of eminent concepts in the field of pattern recognition. SVM is a learning machine method that works on the principle of Structural Risk Minimization (SRM) in order to find the best hyperplane that separates the two classes in input space [7].

SVM in this study included in the series of HOG method which is a tool to determine whether in detection window there is a human figure or not. SVM classifier serves to separate the human class and nonmanusia.

In Figure 2.10a there are several patterns that are divided into two classes-1 and class + 1. Patterns incorporated in class-1 are denoted by a red square, whereas the existing pattern in class + 1 is denoted by a yellow circle. The problem of classification can be translated by searching the line (hyperlane) that separates between the two groups. Some alternate line alternatives are shown in Figure 3a.

Figure 3b shows the hyperlane that separates class-1 with class + 1 that has a certain margin. Margin is the distance between hyperlane and the nearest pattern of each class. The nearest pattern is called vecto support

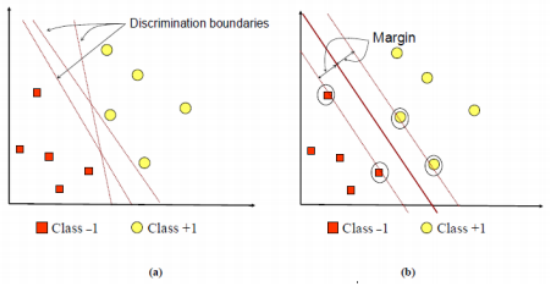


Image 3. SVM find best hyperlane

2.4 Backpropagation Neural Network

Backpropagation neural network is a type of artificial neural network using supervised learning method. In supervised learning there are pairs of input and output data used to train the Neural Network to obtain the desired weight. Backpropagation method uses output error to change the value of the weights in the backward direction. To get this error, the forward propagation phase must be done first. As the propagation progresses, neurons are activated by using the sigmoid activation function, as seen in equation 2.29.

The Backpropagation Neural Network architecture consists of three layers (layers) namely the input layer (input layer) xi, hidden layer (hidden layer) zj, and output layer (output output) yk. The hidden input layer and layer are connected with the weights of the vij and between the hidden layer and the output layer connected by the weights wjk

Architecture can be seen in figure 4

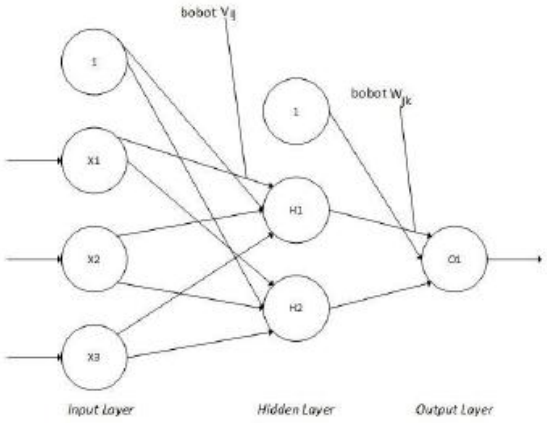


Image 4. Neural Network Architecture.

2.5 Problem analysis

Detection system is a very helpful technology in various fields. The most developed traffic monitoring system has been using magnetic loop detectors and radar as a vehicle monitoring tool.

Problems obtained on car detection using the background substraction method is the presence of shadows caused by sunlight on the vehicle will be detected as an object.

Based on the problem then made a car detection application by using histogram of oriented gradient, SVM and BPNN. Because SVM and BPNN method is a method that can do object classification.

2.6 Process Analysis

Analysis of the process to be conducted in this study is divided into several stages. First is the training phase, starting with feature extraction on the dataset, then the feature extraction results will be trained using svm and bpnn methods. After the training phase is done testing stage whether the system can recognize the car or not.

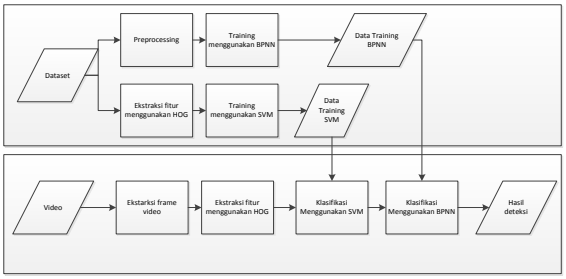


Image 5. Arsitektur Sistem

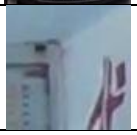
2.7 Input Data Analysis

Data input that will be used in this research is training data and test data.

1. Training Data

The data training used in divided into 2 types of positive training data and negative training data. Positive training data is data that contains pictures of cars, while the negative training data is data that contains images not cars. Here is an example of positive and negative training data

Table 1. training Data

Image	Data
	Positive
	Positive
	negative

2. Test Data

Test data in the form of static video camera footage on the highway. In the video recording will be done preprocessing the video is changed to frameframe for the conversion process to grayscale after it will go into feature extraction stage using histogram of oriented gradient. After that go into classification stages using SVM and then classify using BPNN.

2.8 Analysis of methods

The first step in this research is to extract the feature using HOG method by going through stages:

1. Normalization of gamma and color

The Normalize process of gamma and color or gamma normalization and color in the image frame of the video occurs in the HOG framework on EmguCV. Normalization of gamma and color in the frame images on this system using the instructions: void normalize_hists. The process of gamma normalization and color of the frame images in the HOG framework in the EmguCV library is performed on the frame images of the video. This EmguCV instruction will normalize the gamma and color in the frame images.

2. Compute Gradient

Compute gradients or gradient calculations on the image frame. In this second stage, the image was first converted into a gray scale image.

In this example the method used to convert pixel value from RGB image to grayscale is Luminosity method. Y (x, y) = (0.228 * R) + (0.518 * G) + (0.114 * B)

Here is an example of a car dataset image:


Image 6. Image car in rgb

To convert to grayscale the values are calculated using the following equation:

Y(1,1) = (0,299*90) + (0,587*94) + (0,114*79) = 91,094 ~ 91

Y(1,2) = (0,299*122) + (0,587*126) + (0,114*111) =123,09 ~ 123

Y(1,3) = (0,299*159) + (0,587*163) + (0,114*149) = 155,88 ~ 156

Y(1,4) = (0,299*167) + (0,587*151) + (0,114*147) = 155,33 ~ 155

Y(1,5) = (0,299*165) + (0,587*168) + (0,114*157) = 165,85 ~ 166

Y(1,6) = (0,299*86) + (0,587*79) + (0,114*78) = 86,849 ~ 87

Y(1,7) = (0,299*56) + (0,587*61) + (0,114*54) = 58,707 ~ 59

Y(1,8) = (0,299*61) + (0,587*66) + (0,114*59) = 63,707 ~ 64

Y(1,9) = (0,299*61) + (0,587*66) + (0,114*62) = 64,049 ~ 64

Y(1,10) = (0,299*72) + (0,587*76) + (0,114*75) = 74,69 ~ 75

Y(1,11) = (0,299*67) + (0,587*73) + (0,114*73) = 71,206 ~ 71

Y(1,12) = (0,299*87) + (0,587*92) + (0,114*95) = 90,847 ~ 91

Y(1,13) = (0,299*104) + (0,587*109) + (0,114*115) = 108,189 ~ 108

Y(1,14) = (0,299*88) + (0,587*93) + (0,114*99) = 92,189 ~ 92

Y(1,15) = (0,299*93) + (0,587*97) + (0,114*106) = 96,83 ~ 97

After conversion calculation from RGB to grayscale, we get the grayscale image as follows:



Image 7. grayscale result

After the image is converted to grayscale, the next step is to calculate the value of the image gradient. The method used is the center-difference method of gradient. Specifically this method performs filtering on grayscale image with kernels filter of the following equation:

$$D_2(x) = [-1 \quad 0 \quad 1\,]\; dan\; D_2\,(y) = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

Suppose the 3x3 pixel image matrix as follows:

$$i \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix} \rightarrow \begin{bmatrix} 91 & 123 & 156 \\ 157 & 166 & 182 \\ 155 & 159 & 114 \end{bmatrix}$$

For example, the convolution value of Ix (value I on the x-axis) and Iy (value I on the y-axis) at the p22 position as follows:

$$Ix(P_{22}) = [P_{21} \quad P_{22} \quad P_{23}] * [-1 \quad 0 \quad 1\,]$$

= [157 166 182] * [-1 0 1]
= (157 * -1) + (166 * 0) (182 * 1)
= (-157 + 0 + 182)
= 25
so Ix value in pixels is 25

$$Iy(P_{22}) = \begin{bmatrix} P_{12} \\ P_{22} \\ P_{32} \end{bmatrix} * \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 123 \\ 166 \\ 159 \end{bmatrix} * \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

= (123 * -1) + (166 * 0) + (159 * 1)
= -123 + 0 + 159
= 36
So the value of Iy on pixel is 36
Given the value of Ix and Iy on the coordinates it can be calculated the value of gradient magnitude and gradient orientation, How to calculate the magnitude gradient is to use the formula

$$|G| = \sqrt{{I_x}^2 + {I_y}^2}$$

$$G = \sqrt{25^2 + 36^2}$$

$$= 43.82921$$

Then the magnitude gradient is obtained 43.82821

$$\theta = arctan\frac{I_x}{I_y}$$

$$= \arctan\frac{25}{36} = 35°$$

Then get the gradient orientation value is 35°

3. Calculating Histogram and Orientation Bin

For each cell the local 1-D histogram will be accumulated or edge orientation on all pixels in the cell. The 1-D histogram cell-level combination forms a basic orientation histogram representation. Each orientation histogram divides the various angles of the gradient into fixed numbers in the specified bins. The amount of gradient of pixels in the cell is used to vote into the orientation histogram. For example a built-in histogram will be distributed via 0 ° - 180 ° with a number of channels equal to 9,

0	20	40	60	80	100	120	140	160

And so on At the gradient calculating stage the magnitude gradient is 43.8
and
The orientatiaon gradient is 35 °
Because 35 is between bin 20 and 40 then here is the calculation process determine the orientation binnig (Spatial Orientation Binning)

$$\frac{40 - 35}{20} \times 43.8 = 10.95$$

The above values are distributed to bin 20°

$$\frac{20 - 5}{20} \times 43.8 = 32.85$$

The above values are distributed to bin 40°

	10.95	32.85						
0	20	40	60	80	100	…	…	160

4. Block Normalization

At the time of the Compute Gradient process in obtaining different gradient values. So it is necessary to group each cell in the cells to be grouped again into a larger group called block. Whereas after grouping into blocks, this block is usually terjadi overlap atau tumpang tindih. Dalam This block normalization uses square block geometry R-HOG. This process is to be the final process of the HOG method that produces the feature. This process is done when the process of detector windows as in the process of calculating bin orientation bin. The size of windows detector used is 64x128 which consists of 8x8 pixels. Calculates the normalization of blocks ie takes the cell group and normalizes the overall contrast response. This is accomplished by accumulating a histogram size of a group of cells called blocks. The result will be used to normalize every cell in the block, here is the histogram counting in the block.
Suppose we have the color vector RGB [128, 64, 32]. The length of this vector is [1282 + 642 + 322} = 146.64. This is also called the L2 norm of the vector. Dividing each of these vector elements with 146.64 gives us a normalized vector [0.87, 0.43, 0.22]. Now consider the other vectors where the

element is twice the first vector value 2 x [128, 64, 32] = [256, 128, 64]. You can do it yourself to see that normalization [256, 128, 64] will produce [0.87, 0.43, 0.22]

5. Windows Detector

The final step in the HOG method is to collect the histograms of oriented gradients in the detector window. All histograms for each block will be dikonkatenasi to produce a feature vector value.

On EmguCV HOG Assemblies to the detection window on the image frame of the video using the instructions EmguCV :: gpu :: HOGDescriptor :: detectMultiScale. The HOG collection process of the image detection window of the HOG framework in the EmguCV library is performed on the entire frame of the image of the video. This EmguCV instruction will perform a HOG hanger to the detection window of the images.

At this stage will produce a feator vector of the processed image

0.657589 0.680929 0.641489 0.387971 0.991275 0.178049 0.799521 0.415526 0.651859 .... 0.066560

The result of this hog will be input from SVM. The size of the feature vector is generated from multiplication of block size (2x2 cell), number of bin 9, and number of blocks formed from images with size 128x64. This feature vector is used as input for SVM learning process. In the process of learning done 2 stages, namely training data and test models. In the data training, the HOG feature of the positive image is labeled +1 indicating that the trained image is a positive image feature and the negative image feature is labeled -1 which indicates the trained image is a negative image feature. This SVM training generates a model that will be tested using classify SVM.

SVM learning model is used as a reference in detecting cars. The detection process is done by doing a filter from the top left corner to the bottom right corner. If in the filter process there is a part of the image in accordance with the model, then the object is considered as a positive image which means a car object by giving a box on the object.

After the result svm out system will crop on detected object and will do backpropagation classification with stages.

Objects detected using Histogram and SVM will be detected using Backpropagation Neural Network. Because this classification method can not accept the input data in the form of vectors smaller than -1 and the input bigger than +1, the input data used is the image of the object detected using HOG and SVM. Then it will do preprosesing like Grayscale, edge detection and tresholding. So the data entered in the form of binary numbers 0 and 1.

Edge detection

The edge detection process is executed by using the sobel method, the sobel method uses 2 matrices of 3x3 matrix Gx and matrix Gy both matrix is used to calculate the pixel color difference being calculated around it vertically and horizontal matrix Gx is used to calculate horizontally and matrix Gy is used to calculate vertically.

Matrix Gx is shown in the table

Table 3.9 Matrix Gx

-1	0	+1
-2	0	+2
-1	0	+1

Matrix Gy is shown in the table

Table 3.10 Matrix Gy

+1	+2	+1
0	0	0
-1	-2	-1

In the calculation of this edge detection, all pixels are counted except the pixels that are on the edges of the image, this example will be calculated on pixels (1,1) of the grayscale results in table 3.2 taken 5 pixels.

X/Y	0	1	2	3
0	91	123	156	155
1	157	166	182	156
2	155	159	114	49
3	164	128	50	63

Here are the calculations for each pixel:

Sx= (91 * -1) + (123 * 0) + (156 * 1) + (157 * -2) + (166 * 0) + (182 * 2) + (155 * -1) + (159 * 0) + (114 * 1) = 74

Sy = (91 * 1) + (123 * 2) + (156 * 1) + (157 * 0) + (166 * 0) + (182 * 0) + (155 * -1) + (159 * -2) + (114 * -1)= -94

so M = |Sx| + |Sy| =74 + 91 =165

New Pixel Value is 165

Table 3.11 New Pixel Value

X/Y	0	1	2	3
0				
1		165		
2				
3				

After the new pixel value can be obtained, the next stage is done tersholding.The tresholding steps are

1. Determine the threshold value. In this example the limit is taken 128
2. Read each pixel of the image, if the pixel value of the image is greater than 128, then the matrix on the same pixel value is changed to 1, otherwise if the

pixel value is less than 128 then the value in the same pixel matrix is filled with 0. The result example Tresholding on pixels (1,1)

Table 3.10 Pixel Value of Tresholding Results

X/Y	0	1	2	3
0				
1		1		
2				
3				

After preprocessing, learning is done using backpropagation.

The backrpopagation architecture used is

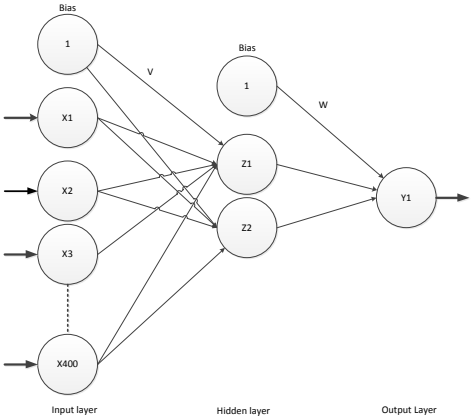


Image 8. architecture backpropagation

It uses 1 hidden layer with 400 input neurons. Obtained from the number of pixels detected as car objects.

2.9 Test results.

Testing is done on the software by comparing the output data of the system and the data of human observation. This test is performed to determine the extent to which the accuracy of the system can calculate the number of recorded car objects.

The test is done using confolution matrix

2.9.1 Test scenario 1

In the test of scenario 1, use 1 neuron in the hidden layer on 3 videos each duration 4 seconds and 10 seconds. Get results

Table 4.8 Video Detection Result Pengujian1.avi

Detik	True Positiv	True negativ	False Positiv
00:01	3	1	2
00:02	3	1	2
00:03	2	0	2
00:04	2	0	2
total	10	2	8

Table 4.9 Video Detection Result Pengujian2.avi

Detik	True Positiv	True negativ	False Positiv
00:01	7	3	2
00:02	6	4	2
00:03	6	3	1
00:04	5	3	2
total	24	13	7

Table 4.10 Video Detection Result Pengujian3.avi

Detik	True Positiv	True negativ	False Positiv
00:01	5	3	0
00:02	5	6	0
00:03	5	4	0
00:04	5	4	0
00:05	3	4	0
00:06	4	2	1
00:07	4	4	0
00:08	4	4	0
00:09	2	4	0
total	37	35	1

Table 4.11 Confusion Matrix Scenario Testing 1

Class		System	
		Car	Not Car
Actual	Car	10+24+37	2+13+35
	Not Car	1+7+8	0

Accuracy $= \frac{10+24+37}{10+24+37+2+13+35+1+7+8} * 100\ \% = \frac{71}{137} * 100\% = 51.8\%$

Scenario Testing 2

In the 2nd scenario test, use 2 neurons in the hidden layer on 3 videos each 4 seconds and 10 seconds duration.

Tabel 4.12 Hasil Deteksi Video Pengujian1.avi

Detik	True Positiv	True negativ	False Positiv
00:01	3	1	2
00:02	3	3	2
00:03	2	0	2
00:04	2	0	2

total	10	4	8

Tabel 4.13 Hasil Deteksi Video Pengujian2.avi

Detik	True Positiv	True negativ	False Positiv
00:01	5	3	4
00:02	5	2	3
00:03	6	4	1
00:04	5	4	2
total	21	13	10

Tabel 4.14 Hasil Deteksi Video Pengujian3.avi

Detik	True Positiv	True negativ	False Positiv
00:01	4	2	1
00:02	5	3	0
00:03	4	4	1
00:04	3	3	2
00:05	3	4	0
00:06	3	3	2
00:07	2	2	2
00:08	4	2	0
00:09	2	4	0
total	30	26	6

Tabel 4.15 Confusion Matrix Skenario Pengujian 2

kelas		Sistem	
		Car	Not Car
Actual	Car	10+21+30	4+13+26
	Not Car	8+10+6	0

$$\text{Accuracy} = \frac{10+21+30}{10+21+30+4+13+26+8+10+6} * 100\ \% = \frac{61}{138} * 100\% = 44.2\%$$

The greatest accuracy achieved from this performance test is in test scenario 1 using a hidden layer, with an accuracy of 51.8%.

6. END

3.1 Conclusions

Based on the results of research, analysis, system design, and implementation and testing, it can be concluded that the simulator of car detection built using histogram method of oriented gradients with SVM and BPNN classification method has a fairly small accuracy of 51.8% by using 1 neuron in hidden Layer.

3.2 Suggestions

Based on the results of research and implementation accuracy of this research has not achieved a high accuracy. To improve performance and better results, it is proposed some suggestions as follows:

1. Using a combination of other classification methods such as naïve bayes and svm
2. Using other feature extraction methods such as pca.
3. Using more data sets

BIBLIOGRAPHY

[1] Ari Kurniawan Saputra, "Aplikasi Deteksi objek menggunakan Histogram Of Oriented Gradien Untuk Modul Sistem Cerdas Pada Robot NAO," Universitas Bandar Lampung, Bandar Lampung, Skripsi 2016.

[2] Dr. Vinod Shokeen Vandna Singh, "Compariosn of feature Extraction Algorithm For Gender Classification From Face Images," IJERT, vol. 2, no. 5, may 2013.

[3] Cahyo Permata, I Ketut Eddy Purnama, and Muhtadin, "Deteksi Mobil Menggunakan Histogram Of Oriented Gradient," Institut Teknologi Sepuluh November, Surabaya, Skripsi 2010.

[4] K.P Sanak Kumar and Dr.R Bhavani, "Activity Recognition in Egocentic Video Using SVM, kNN and Combined SVMkNN Classifiers," in Internation Conference on Advanced Material Technologies, Visakhapatnam, 2016.

[5] Indu Dokare and Naveeta Kant, "Performance Analysis of SVM, k-NN and BPNN Classifiers for Motor Imagery," International Journal Of Engineering Trends and Technology, vol. 10, pp. 19-23, April 2014.

[6] Wen Zhang, Xijin Tang, and Taketoshi Yoshida, "Text Classification With Support Vector Machine And Backpropagation Neural Network," in ICCS, Berlin, 2007, pp. 150-157.

[7] A. Gunawan, E. Kusnandar and F. E. Gunawan, V. B. Wewengkang, "Visual Traffic Surveillance Using Feature Point Approach," in Procedia Engineering, vol. 50, 2012, pp. 354-360.

[8] D. Zhou, "Bus Passenger Recognition and Track of Video Sequence," Journal of Multimedia, vol. 8, pp. 262-269, Juni 2013.