Application Object Detection Using Histogram of Oriented Gradient For Artificial Intelegence System Module of Nao Robot *(Control System Laboratory (LSKK) Bandung Institute of Technology)*

A K Saputra[1,], F Ariani[2], R Y Endra[3]
[1] Computer Science Faculty, Bandar Lampung University, Bandar Lampung, Indonesia
e-mail : ari.curnia.one@gmail.com
[2] Computer Science Faculty, Bandar Lampung University, Bandar Lampung, Indonesia
e-mail : fenty.ariani@ubl.ac.id
[3] Computer Science Faculty, Bandar Lampung University, Bandar Lampung, Indonesia
e-mail : robby.yuliendra@ubl.ac.id

Abstract. The introduction of the object (object recognition) is one of image processing to identify objects that will be identified for further processing in order to obtain an information data with the existence of the object recognition process. The purpose of this research is to develop intelligent system modules on NAO Robot to build applications that can detect objects around. Testing the accuracy of detection of the algorithm Feature Descriptor Histogram of Oriented Gradient used in building applications for the object detection intelligent system modules on NAO Robot. Results of Design Application Object Detection using the Histogram of Oriented Gradient algorithm can be used to add intelligent system modules on NAO Robot in because of the test results produce a test model tests show a success rate reached 99.10% in identifying the test image. On testing positive test image, the amount of training data usage greatly affect the object detection process becomes more accurate. In testing the combined image (image test positive and negative) the success rate rose 98.23% to 99.10%. *Keywords. Image Processing, Computer Vision, Feature Descriptor Algorithm, Histogram of Oriented Gradient, Support Vector Machine, Robot NAO, Artificial Intelegence For Robot.*

1. Introduction

Digital Image Processing or Image Processing is a study of science that process data in the form of images with the help of computer technology system, whether about image quality improvement or image recognition, all can be learned in this science. The development of this science can also be developed in various researches and in the field of Computer Vision. Computer Vision is a combination of image processing and pattern recognition. Computer vision relates to the theory on AI (Artificial Intelligent) which extracts information from images. In this research, image processing is applied to object detection application as intelligent system module on NAO Robot in The Control of System Control (LSKK) Bandung Institute of Technology in object recognition process. NAO is an autonomous and programmable medium-sized humanoid robot developed by Aldebaran Robotics, a French startup company headquartered in Paris. There are several feature descriptor methods for detecting objects, such as SIFF (Scale Invariant Feature Transformation), SURF (Speed Up Robust Feature), Color Based Object Detection, HOG (Histograms of Oriented Gradients), Viola Jones, Optical Flow, and etc. In this research, extraction feature method used is HOG (Histograms of Oriented Gradients) and this research used SVM (Support Vector Machine) as well as classification feature (feature classifier). HOG is used in this research because HOG is a Feature Descriptor that takes edge or gradient structure characterized from local shape or edge direction and with good local gradient intensity distribution. (Computer Vision, p. 01, Linda G. Shapiro & George C. Stockman, Prentice Hall, 2001).

2. Literature Review

2.1 Histogram of Oriented Gradient

Navneet Dalal and Bill Triggs explain that the HOG method makes it possible to detect humans by 6 basic steps as processes like, color normalization, gradient calculations, bin and cell orientation, normalization contrast with spatial block accumulation, histogram data collection, and classification by using SVM.

2.2 Steps in HOG

Here are the steps of Histogram of Oriented Gradient in the process of detecting objects:

A. Image Conversion or Color Normalization

A true color image is a color image representation that has three main components which are red, green and blue (RGB). Each component in the true color image has 256 possible values. The grayscale image has 28 (256) possible values on its pixel. The value starts from zero for black and 255 for white. The conversion of true color image to Grayscale changes the value of pixel which originally has 3 values Red, Green, and Blue which become one value that is gray. The equation used to get gray values is as follows:

$$L = 0.144 * R + 0.587 * G + 0.299 * B \quad (1)$$

Information:

L : gray value on pixels
0.144 : the weights for the red element (wR)
0.587 : the weights for the blue color element (wB)
0.299 : the weight for the green color element (wG)
R : the intensity value of the red element
B : the intensity value of the blue color element
G : the intensity value of the green color element

NTSC (National Television System Committee) defines weights for conversion of true color images to grayscale as follows: wR = 0.299, wB = 0.587, wG = 0.114.

B. Gradient Calculation

Calculating the value of the Gradient (Gradient Computing) is a process image after the conversion. A gradient is a result of the measurement of change in a function of intensity, and can be viewed as a collection of some of the functions of continuous intensity of the image. This process is used to obtain the line edges on the object in the image. Gradients of an image can be obtained by filtering 2 dimensional filters i.e. vertical and horizontal filters. The first image is converted in the form of grayscale to avoid the contribution of different intensities to every area of color (RGB). A method commonly used is the 1-D centered, with a matrix as follows:

[-1,0,1]

C. Spatial Orientation Binning

Creating a histogram requires a gradient value and the value is derived from the value of each pixel in an image. The images will then be divided into cells of predetermined size. So the histogram of every cell in the image will be made in order to know the value in each cell because each cell has a different value. Making the histograms require the existence of bin to know gradient value. The bin will be user-defined. In previous research the bin used is 4 bin orientation.

D. Block Normalization

Gradient values have different values; therefore it is necessary to group the cells into larger ones or called blocks. Blocks usually overlap because each cell contributes value more than once. In this block normalization there are two main block geometries that are R-HOG rectangle block and circular C-HOG but in this research the one will be used is R-HOG geometry. The final result in the normalization of this block is the feature. In this process, overlapping blocks are resolved with R-HOG. While blocks consist of 2 x 2 cells, there is 7 x 15 R-HOG in detector windows and it is 4 bin orientation so as to get 1680 vector in 1 detector windows. The number of these vectors in can be of 2x2x7x15x4 and this vector is called a feature. Figure 2.7 is an example of R-HOG geometry.

E. Windows Detector

Windows Detector is the windows sized 64 x 128 used for detection windows. (Navneet Dalal and Bill Triggs, 2003)

3. Research Methodology

3.1 Need Analysis

The analysis of functional requirement is performed to find out the function specification that can be done by the system. Needs analysis is a need that comes from stakeholders including the functions and features of a system. This Histogram of Oriented Gradient based object detection application has several capabilities, namely:

A. Object detection can be done with many objects or multidetect, in which the user makes data training according to object data will get detection or in this process called image acquisition, then the application will perform a process according to stages in Histogram Of Oriented Gradient algorithm. In this first stage, the normalization of color is the process of converting the colored image into grayscale image, and then Gradient Compute is the process of calculating the gradient value of each pixel in the image. The next process is Spatial Orientation Binning, the process where every cell in the image will be made histogram. Making the histograms requires the existence of bin to know gradient value. The bin will be user-defined. In the previous research bin used is 4 bin orientation, in this block normalization there are two main block geometries that is R-HOG rectangle block and circular C-HOG this research used R-HOG geometry. The final result in this block normalization is a feature. The next stage is the Windows Detector is a windows size 64 x 128 used for detection windows. This detection window consists of 8 x 8 pixels in each cell.

B. Application will send the data in the form of vector values to the robot module server according to the channel or server address of each module. After the data is successfully sent then it will be processed in Artificial Intelligence robot module to then the value is executed as command by NAO robot.

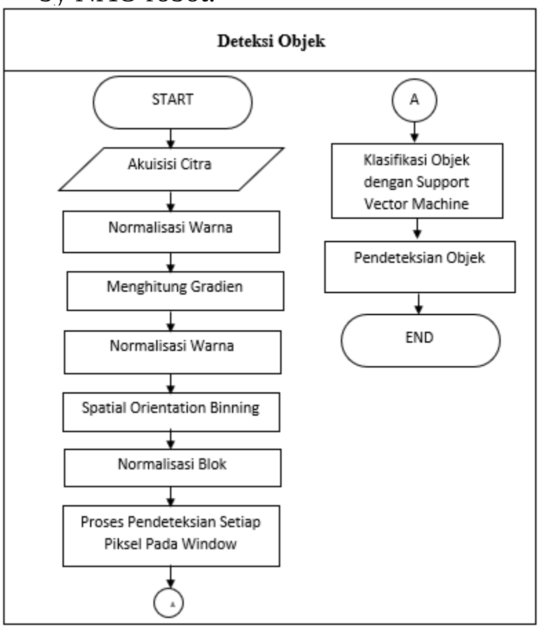

Figure 1. Flowchart of Object Detection Process Using Histogram of Oriented Gradient

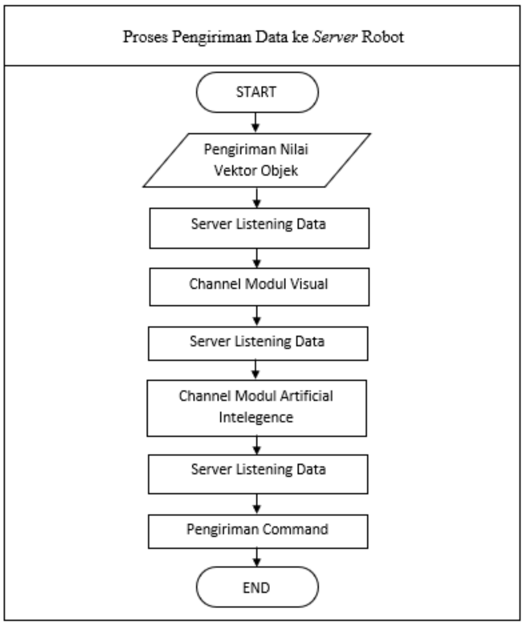

Figure 2. *Flowchart* Pengiriman Data ke *Server* ModuL

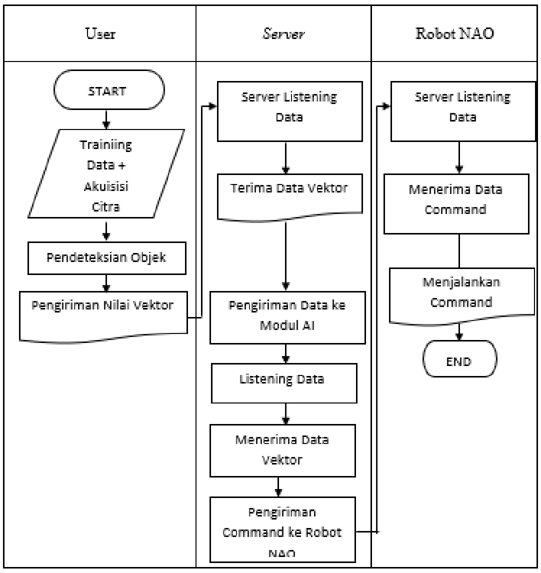

Figure 3. Flowchart of Object Detection Process Using Histogram of Oriented Gradient

4. Implementation

The implementation of object detection application for intelligent system module on NAO Robot is done because of the lack of information processing from object detection to the development of machine perception. Therefore, the authors apply this object detection application on intelligent system module NAO robot conducted in the Laboratory of Control Management system of Bandung Technological Institute (LSKK). LSKK was established since 1965 working in the field of education, research & development as well as professional support within the framework of Bandung Institute of Technology. Currently, LSKK consists of 16 staff from two Expert Group, namely KK Teknik Komputer & KK Full & Computer System.

4.1 Interfaces

Here are some views between users to interact with the Object Detection App for smart modules on the NAO Robot using the Feature of Descriptor Histogram of Oriented Gradient algorithm. This algorithm has several stages in the detection process, namely: Image Acquisition, Color Normalization, Gradient Compute, Spatial Orientation Binning, Block Normalization, Windows Detector, and Linear Support Vector Machine Classification.

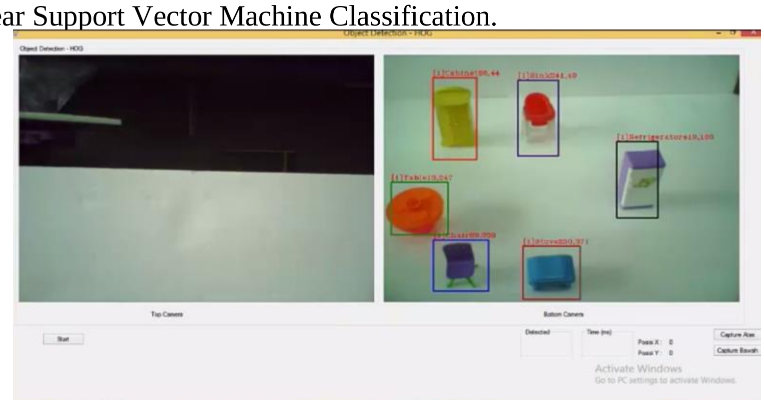

Figure 4. Display of Application Object Detection

4.2. Image Acquisition

Image acquisition is the first step to get a digital image. The purpose of image acquisition is to determine the required data and choose the method of recording digital images. In this study the acquisition of images in the object detection uses video camera from NAO Robot

Figure 5. Image Acquisition Process from NAO Robot Camera

In the process of image acquisition as a data retrieval training, image acquisition training data is divided into 2 ie Data Training of Postitive Image Acquisition and Data Training on Negative Image Acquisition.

4.2.1 Positive Image Acquisition Data

Positive image acquisition is data of training object data that will be in detection, here are the positive training data

Figure 6. Data Training on Positive Image Acquisition

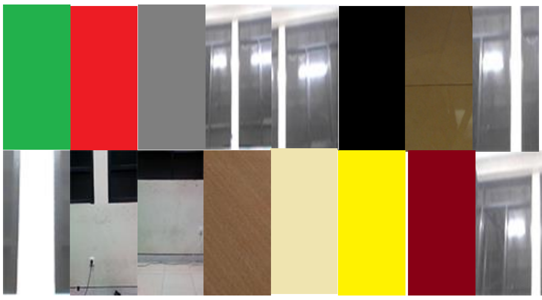

Figure7. Data Training on Negative Image Acquisition

4.2.2 Data of Training Image of Negative Image Acquisition

Negative image acquisition is a training data object that is not in detection, here is the negative training data

4.3 Color Normalization

This stage is the process of converting the original image into a grayscale image. This process can be done by taking all pixels in the picture then the color of each pixel will be taken the information about the 3 basic colors of red, blue and green (through RGB color function). These three basic colors will be added then divided by three to obtain the average value. This is the average value that will be used to colorize the image pixels so that the colors become grayscale. The three basic colors of a pixel will be set to the average value (through the RGB function). Getting the grayscale value can be found with the following conditions:

$$f_0(x,y) = \frac{f_i^R(x,y) + f_i^G(x,y) + f_i^B(x,y)}{3} \qquad (2)$$

4.4. Calculating Gradient Value

This process is used to get a border on the object in the image. The gradient of an image can be obtained by doing filtration with a 2 dimensional filter that is a vertical and horizontal filter. Gradient is the result of measuring changes in an intensity function, and an image can be viewed as a collection of several continuous intensity functions of the image. The first thing to do is converting the image in the form of grayscale to avoid having to consider the contribution of different intensities to each color field (RGB). The commonly used method is 1-D centered, with the following matrix:

[-1,0,1]

$$\frac{\partial f}{\partial x} = \frac{f(x+h) - f(x-h)}{2h} \qquad (3)$$

1. Using partial derivative formulas for imag function f (x, y):
   a) For the x axis:

$$\frac{\partial f}{\partial y} = \frac{f(y+h) - f(y-h)}{2h} \qquad (4)$$

   b) For the y axis:

$$R = \sqrt{x^2 + y^2}$$

$$\theta = arc\ \tan\left(\frac{y}{x}\right) \qquad (5)$$

2) then, will get the values of x and y used to calculate the gradient:
   a. Magnitude (large *gradient*) :
   b. Orientation of gradient (in angle):

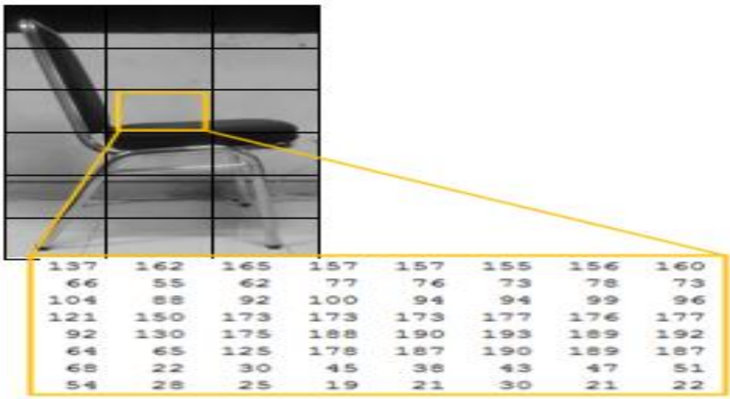

Figure 8. Values In Picture Cells

   c. The gradient calculation process uses the conventional 1-D centered method

137	162	165	157	157	155	156	160
66	55	62	77	76	73	78	73
104	88	92	100	94	94	99	96
121	150	173	173	173	177	176	177
92	130	175	188	190	193	189	192
64	65	125	178	187	190	189	187
68	22	30	45	38	43	47	51
54	28	25	19	21	30	21	22

Filter mask = -1 0 1 for Dx

(-1 0 1)T for Dy

Dx = -150 + 173 = 23

$\rightarrow |\nabla f| = \sqrt{Dx^2 + Dy^2} = 86{,}12$ Dy = -92 + 175 = 83 (6)

Magnitude

Orientation $\rightarrow \theta = \arctan\left(\frac{Dx}{Dy}\right) \approx 15^0$

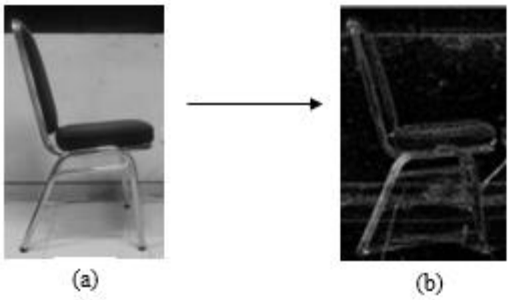

Figure 9. Grayscale Image (b) Image of Gradient Calculation Result

4.5 Spatial Orientation Binning

This process produces sensitive encoding of localized image content, but remains resistant to small changes in pose and appearance. Images are divided into several regions with small spaces called cells. 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 edges of the gradient into fixed numbers within specified bins. The amount of gradient of pixels in the cell is used to vote into the orientation histogram. For example, a histogram will be built distributed through 0 ° -180 ° with a number of channels equal to 9, so the vote in the histogram is as follows:

a) All gradients with large angles [0 ° -20 °] provide a vote for channel 1.
b) All gradients with large angles [20 ° -40 °] provide a vote for channel 2.
c) And so on.

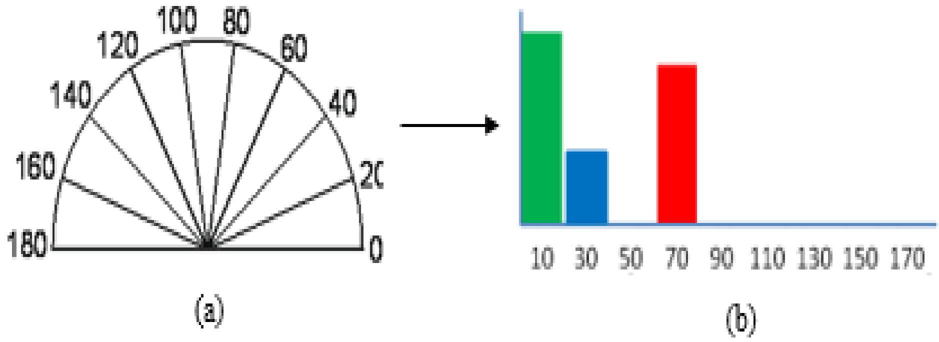

Figure 10. Gradient Angle, (b) Histogram Graph on Cells

Fig. (A) is the angle of the direction gradient of illumination in an image, the image (b) is the Histogram graph of each cell in the image, each orientation histogram divides the various angles of the gradient into fixed numbers in specified bins. The amount of gradient of pixels in the cell is used to vote into the orientation histogram. Here is the calculation process determines the orientation of binnig (Spatial Orientation Binning) on the image of the chair:

Orientation = 70 o
Orientation = 70 o
Orientation = 15 o
Magnitude = 35
Magnitude = 20
Magnitude = 86.1

$$\frac{30-15}{20} * 86 = 64{,}5\ bin\ 10^{o}$$

$$\frac{15-10}{20} * 86 = 21{,}5\ bin\ 30^{o}$$

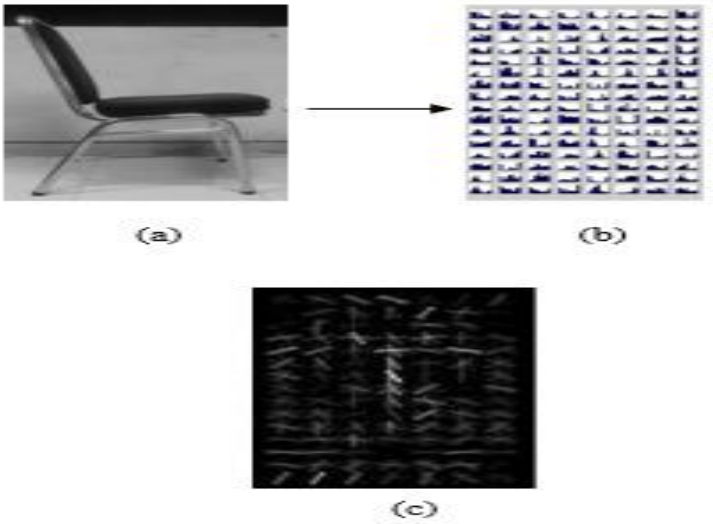

Figure 11. Grayscale Image, (b) Histogram Value, (c) Magnitude, (d) Visualization of Histogram Feature of Oriented Gradient

4.6 Block Normalization

At the time of the Compute Gradient process it is obtained different gradient values. So it is necessary to group each cell in the cells into a larger group called block. Whereas after grouping into blocks, these blocks usually overlap. In normalization, this block uses R-HOG square block geometry. This process is the final process of the HOG method that produces the feature. This process is done when the process of windows detector as in the process of calculating bin orientation bin. The size of the windows detector used is 64x128 which consists of 8x8 pixels. Block normalization calculation is done by taking the cell group and normalizing the overall contrast response. This is done 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:

1) $\mathbf{L1}\ -\ sqrt\colon v \rightarrow \sqrt{v\ /\ (\|v\|_1\ +\ \varepsilon)}$;
2) ie L1 -norm followed by square root, the amount to treat the vector descriptor as an opportunity distribution
3) L2 - Hys, L2 -norm followed by clipping (limited to maximum value) and renormalizing which is an unnormalized vector. The k-standard and a small constant.

4.7 *Windows Descriptor*

This process is to collect descriptor of all blocks which are the overlapping grid enclosing the detection window to the nature of the combined vector feature used in the classifier. The one used as the series parameter descriptor is R-HOG. The R-HOG block descriptor uses square or rectangular grid overlap cells. R-HOG calculates the grid (which defines the cell number of each block) of the pixel cells containing bins each, which is a parameter.

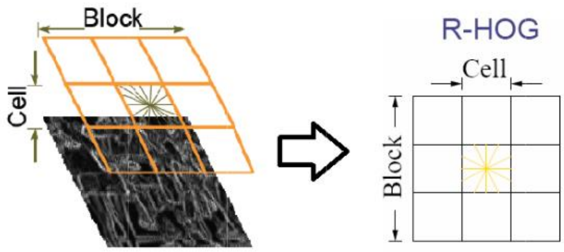

Figure 12. R-HOG Cell

So if it is assumed that it has:

1) Window Size
2) The pixel cells (tital R-HOG cells)
3) Block cells without overlap (as many as 7 blocks)

Then the final size descriptor will be obtained: (7 X 15) X (2 X 2) X 9 = 3780

4.8 Linear Support Vector Machine (SVM) Classifier

Support Vector Machine (SVM) is a learning machine method that works on the pronsip of Structural Risk Minimization (SRM) in order to find the best hyperplane that separates the two classes (categories) in the input space. The basic principle of Support Vector Machine is a linear classifier, and is being developed in order to work on nonlinearity. To check whether in the window there is an object to be detected or not, SVM Classifier is used to separate human and non-human or in this research is object detection. In SVM Classifier and classification algorithms that attempt to separate an optimal hyperplane (Cristianini & Shawe-Taylor, 2000).

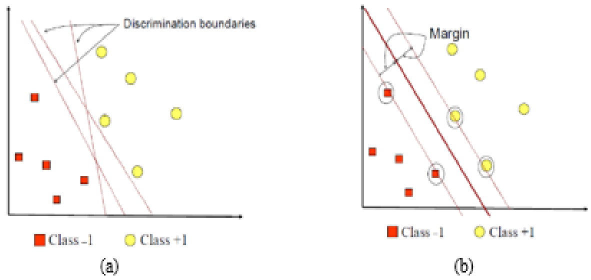

Figure 13. SVM in its attempt to find the best hyperlane to separate classes -1 and +1

4. Results

Results Accuracy of Object Detection in Object Detection App Using Algorithm Feature Descriptor Histogram of Oriented Gradient In the process of learning on SVM it was 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 signifies the image being trained is a negative image feature.

a. Data Set

Data set is a required component for data learning using SVM. Data learning required in the form of positive data training is data containing car objects, training negative the data containing non-car objects (background), positive test data, negative test data, and image to be detected. The images used are 64x128, except the image to detect, has a size of 680x480, larger than the image training and image test. Here is the image data to be detected:

1. 4 positive image data for seats
2. 4 positive image data for monitor
3. 4 positive image data for bucket
4. 4 positive image data for table

Negative image is obtained with random crop image with size 64x128. The number of negative images used are as follows:

1. 36 negative image training
2. 36 negative image testing

Overall, the composition of the image data set is shown in the following table

Table 1. Image Test Data in Test Data Training

Image Positif		Image Negatif	
Training Image	Test Image	Training Image	Test Image
16	16	36	36

a. Testing Model

Testing image is a test of learning result model to know the level of model accuracy in recognizing test image. Test images have the same pixels as training data, but have different image variations. From Figure 8, prior to testing the model is done the extraction feature of HOG from the tested image. Furthermore, the feature is changed into the form of feature vector of the same size with the feature vector training data is 4608x1. The vector feature of the positive image is combined with the negative image vector feature that produces a test feature. The vector feature of this test image is classified using classify SVM with the training model as its reference. From this test it is known that the number of successful images detected in accordance with the predefined class (positive image is detected as positive image, negative image is detected as negative image) and also known as well the number of images that cannot be detected in accordance with the specified class.

b. Object Detection

The detection process is done by filtering the detected image with the model (SVM Model) as a filter which is by moving the model from the top left corner of the detected image until the lower right corner. If in filtering process there is a feature HOG weight (weight) above the specified threshold value then the feature is detected as a positive image by giving a box on the part. However, if the feature weight is detected below the specified threshold value then the feature is considered as a negative image.

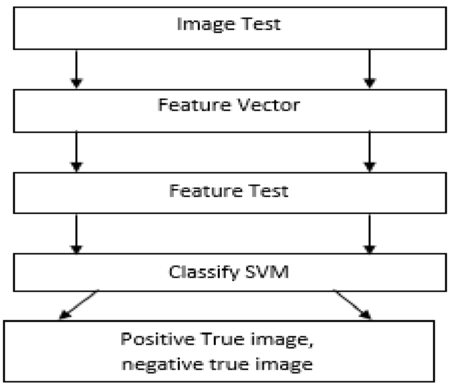

Figure 14. *Testing Image*

In this test, the amount of training data used is made varied. Training results are stored in model form then tested using SVM classify. The test was done with the composition of training data with the number of positive image tests as many as 16 positive image tests and 36 negative image tests. Testing is done 2 times with initial training data as much as 16 positive images and 36 negative images. For the first test, it generates the following data.

First experiment:

Data *image training*:

1. *Positive image:* 16 *image*
2. *Negative image:* 36 *image*

Test results:

1. Positive image test : 16 correct, 3 incorrect, 19 total
2. Negative image test : 30 correct, 8 incorrect, 38 total
3. Positive and negative image test : 46 correct, 11 incorrect, 57 total

Second experiment:

Data *image training* :

1. *Positive image:* 16 *image*
2. *Negative image:* 36 *image*

Test results:

1. Positive image test : 16 *correct*, 3 *incorrect*, 19 total
2. Negative image test : 30 *correct*, 3 *incorrect*, 19 total
3. Positive and negative image test : 46 *correct*, 11 *incorrect*, 57 total

The test is continued for up to 2 tests. The results obtained from the test are as follows.

Table 2. Results of Training Data Testing

Training Positif	Training Negaitif	Hasil Tes Positif	Hasil Tes Negaitif	Hasil Tes Positif dan Negatif
16	36	95,39%	99,60%	98,23%
16	36	95,39%	99,60%	98,23%

Positive test results show an increase when the amount of training data is added. The last percentage shows that the success rate of the test with classify SVM is 99.10% with positive training data as much as 16 images and negative training data of 36 images. However, during the second test, the number of training data 16 positive images and 36 negative images get the same result, because the model resulting from the second training able to detect the image with the composition 30 correct, 3 incorrect, 57 total. While the negative test results show a success rate of 99.60% of the image test as many as 36 images. Linearly, the test results show an increase with the added amount of training data.

5. Conclusion

Based on the results of the research and discussion, which is done from the designing stage to testing the Application Object Detection Using Histogram of Oriented Gradient For Intelligent System Module On NAO Robot, the conclusions drawn from the research are as follows:

1. The test result of model test shows the success rate reaches 99.10% in recognizing the test image.
2. In testing positive test images, the amount of use of training data greatly affects the process of detecting objects to be more accurate.
3. In the combined image test (positive and negative test image) the success rate increases 98.23% to 99.10%.

References

[1] Linda G. Shapiro & George C. Stockman. 2001. *Computer Vision.* United States : Prentice Hall.

[2] Navneet Dalal and Bill Triggs. 2003. *Histogram of Oriented Gradient for Human Detection*. Perancis : INRIA Rhone-Alps, 655 avenue de l'Europe, Montbonnot 38334.

[3] Cristianini & Shawe-Taylor. 2000. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods.* Inggris : Cambridge University Press.