Human Detection by Histogram of Oriented Gradients for Automatic Braking in Autonomous Cars

Christoporus Deo Putratama (13213008)  
Kevin Shidqi Prakoso (13213065)  
Bramantio Yuwono (13213128)

Date: 27/12/2016

EL4138- VLSI System Engineering

SEEI - ITB

1. Introduction

Recently, self-driving cars have been a trending topic in the development of the automotive industry. However, there have been several unfortunate events regarding the development of autonomous cars. There have been several accidents, and these events have had a negative effect on the public opinion. Therefore, more safety measures are in order if the autonomous car industry is to have a chance of thriving. One of the main concerns is aimed at the ability of the car to properly detect if an obstruction in ahead, and do the necessary manuevers to ensure minimum damage. The most pressing matter in this regard is of course, the ability to detect humans.

Human detection has been an important topic in image processing, and while many techniques have been developed, few have been able to produce results that are satisfactroy in terms of accuracy and speed. Naturally, the human detection process will need to be fast enough if it is to be effectively implemented in automatic braking. However, accurate algorithms usually dissapoint in terms of speed, and algorithms for automatic braking need to both accurate and fast, with perhaps a little more emphasis on the latter. One of the possible solutions is to develop a hardware-based system, in order to both have accuracy and speed.

  In this report, we aim to develop a circuit that implements a human detection algorithm by means of histograms of oriented gradients (HOG). HOG is a relatively new feature descriptor, introduced in 2005 by Dalal and Triggs [1]. This method is similar to edge orientation histograms, but this method differs in that it is computed on a grid of uniformly spaced cells which are consisted of pixels and uses local overlapping contrast normalization for accuracy.

The rest of this report is organized as follows. Section II explains fundamentals of the circuit and architecture, section III shows reduction techniques used in the circuit for efficient implementation of this algorithm as well as several other appeals, section IV explains the simulation results, while section V will give an overall evaluation of the implementation in an Altera DE2-115 FPGA.

1. Architecture description
   1. Overall architecture

The histograms of oriented gradients uses local histograms of oriented gradients of pixel luminance to characterize a given image. HOG description of an image roughly gives its structure and shape, enabling us to perform detection algorithms to rcognize human shapes. The process of HOG feature extraction is as follows:

1. Luminance is calculated from the color image, and the gradient strengths and angles are calculated from the luminance map.
2. Histograms of oriented gradients are calculated by summarizing the gradient magnitude and orientation for each group of pixels.
3. The histograms are normalized, and used to determine whether or not there is human in the image using SVM classifier.

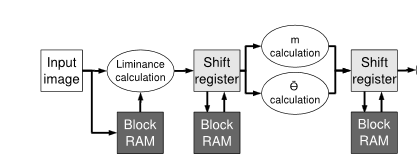


Figure 1: First half of architecture

Therefore, the design for the system is as follows. First, the input image needs to be converted to a luminance map, or in other words, converted to grayscale. Grayscaling is done by multiplying each color component of a pixel by a fraction, and then summing all the results. The final number is the luminance value of a pixel. To achieve this purpose, we use a simple adder combined with a constant multiplier consisting of shifters. The image is stored in a RAM block, as is the resulting grayscale image.

Second, we need to calculate the luminance gradient of the image, in directions x and y. The computations are done per block. Luminance gradients are easily computed by finding the difference of luminance between neighboring pixels, which is achieved using a simple subtractor. The luminance gradients are then used to calculate the gradient strengths and orientations. The results are used to fill the histogram of that particular block, with each pair of luminance gradients adding to a certain bar on the histogram depending on their orientation, by a value depending on their magnitude.

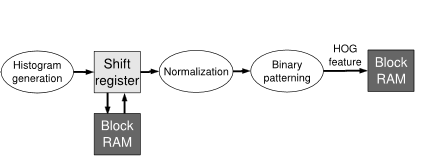


Figure 2: Second half of architecture

Third, the resulting histogram needs to be normalized and patterned. Histogram normalization causes the histogram to be as such, so that if all of its elements are summed, will produce unity as a result. Patterning involves combining several histograms from neighboring blocks. The resulting histogram is then saved in a RAM block, to be checked whether or not the histogram represents an image of a human.

Fourth, the detection process takes place. This involves multiplying a weighing matrix with the histogram, and adding it with a bias number. Systolic arrays are used to multiply matrices efficiently, and the resulting number determines whether or not a human is detected.If a human is indeed detected, a red box is drawn around the block which the histogram represents. As shown in the figures above, the process is fully pipelined.

* 1. NIOS II Processor and RAM

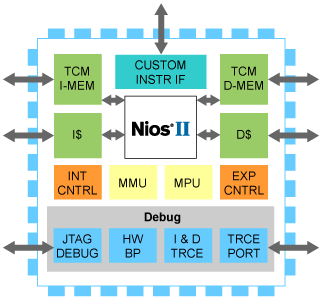


Figure 3: NIOS II Overview

The system is implemented on an Altera DE2-115 FPGA, which include 2MB of SRAM used in the architecture. To build the necessary components, the Altera Qsys is used. The SOPC built is as follows.

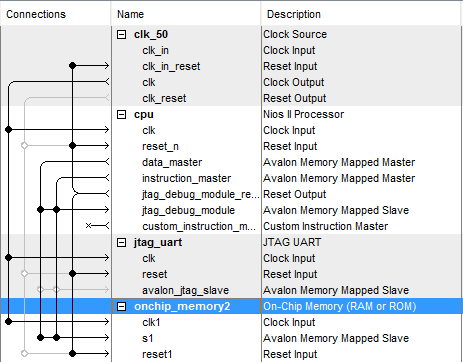


Figure 4: SOPC Overview

As a processor, NIOS II is too slow for our purpose, so its only responsibility is to manage RAM access and receive commands via the JTAG port, which enables switching of images, while all of the calculations are done on separate modules. The 2MB SRAM is more than enough to store the image, its grayscale counterpart, and any resulting luminance gradients as well as histograms. Several constants are also stored on the RAM, such as the weighing matrix.

* 1. Preprocessing

HOG is a feature descriptor to represents shapes, and as such, it ignores variations in color, which means an image needs to be converted to grayscale. A color image has three matrices, consists of R, G, and B matrix, so we must combine each of those matrices into a single matrix by using the following equation.

Y = 0.298912R + 0.586611G + 0.114478B

The resulting number is the luminance of a pixel, with a higher number representing a lighter shade of gray.

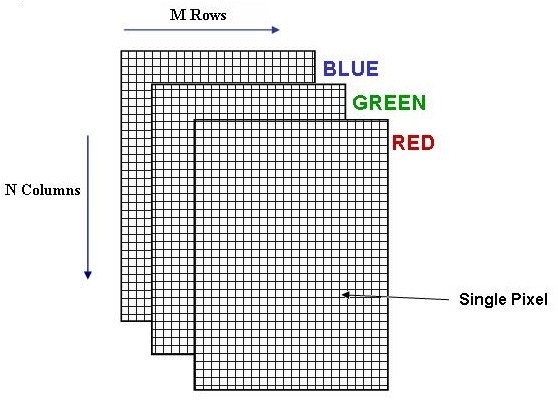


Figure 5:Representation of matrices from a color image

The implementation from this greyscaling process is to change all the three matrices into a 8-bit signal by the aforementioned equation. This involves multiplying and adding each component. However, because the multiplication factors are constant, the multiplication operation can be replaced by shifters, as follows.

0.58 is approximated by 0.100101, needing 3 shifters.

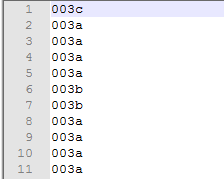
0.29 is approximated by 0.0100101 , needing 3 shifters.

0.11 is approximated by 0.000111 , needing 3 shifters.

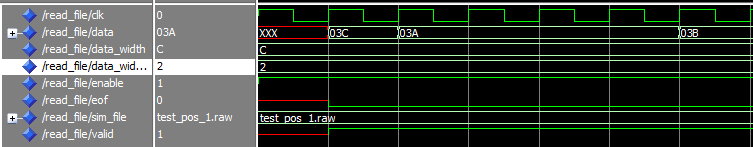
The resulting grayscale image is then stored on the RAM. To speed up the process, we execute grayscaling in parallel with 120 modules.



A sample of greyscaled image

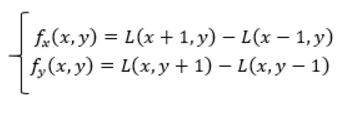


Sample of raw image represented in hex numbers.



Signal created from reading raw file in modelsim.

Gradient calculation is represented by the following expression:



Equation used to calculate gradients

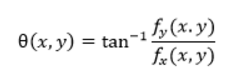
From the grayscaling process, the resulting image is then scanned to get the gradien from every pixel of the image. The gradient calculating process involves simple subtraction between neighbouring pixels. As before, calculation is done in parallel using 120 module, with each module continuing the work from a specific grayscaling module. As before, the results are then stored to the SRAM.

* 1. HOG Calculation

After getting *dx* and *dy*, we can calculate the magnitudes of the gradients by using following expression:



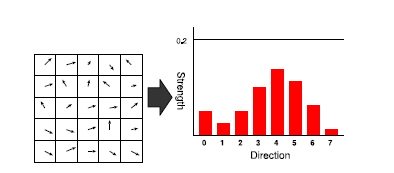
While the orientation of the gradients are calculated using the following formula



Equation used to calculate angle of gradien

However, implementing calculations such as square roots and trigonometric functions are not pratical in an FPGA. However, the exact value of the arctangent is not entirely necessary, because blocks in histograms represent a range of angles, not an exact number. Therefore, we approximate the arctangent calculation using shifter-multiplier and comparators. As for the magnitude calculations, approximations have been tried with dissapointing results. Therefore, we devised a new algorithm to calculate magnitudes efficiently. The details of the calculation are discussed in the following chapter.

Then the gradient strengths m are inserted into the histogram according to the corresponding quantized orientation label for every luminance value in a cell to make the histograms, as shown in the figure below.

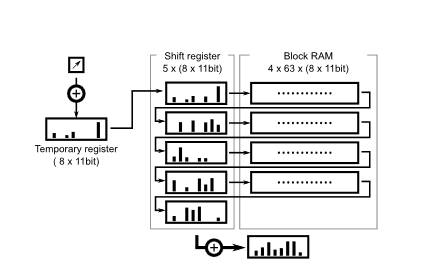


For example, when an angle of a certain pair of dx and dy is between Q1 and Q2, then the corresponding label becomes 2, and its magnitude is added onto bar with label 2. We take the range of angles from -π/2 to π/2, since HOG does not focus on gradient direction but orientation. The threshold values are as follows:



A histogram is generated for each cell, which contains 8x8 luminance values. Since we quantize the gradient orientation into eight labels, 8-dimension feature vectors are eventually generated. For a (w × h) input image, a total of w/8 × h/8 histograms are obtained since each cell consists of 8×8 luminance values.

Since quantizations have been made in the calculation steps of gradient orientations θd as mentioned in the previous parts, any other quantization process is not necessary in the histogram generation process. We have 640×480 values of luminance gradients(one for each pixel), which gives us a two-dimentional array with the width of 640/8 = 80 cells and the height of 480/8 = 60 cells. Since the maximum value of a gradient strength is 361, then the maximum size is 23,104 for a histogram of a single cell, data size for one cell becomes 14 bits ×8. Therefore, total histograms of luminance gradient for the entire single image are expressed with (14×8)×(80×60) = 537,600 bits.



We also take a stream process approach for histogram generation as shown in Figure . A specific histogram for a cell is obtained from a matirx of 64 (8×8) values of luminance gradients(both m and θd). A partial histogram of gradient orientations for 8 consecutive luminance values is made every 8 luminance values. Then the stream of the partial histograms goes through Block RAM line buffers so that partial histograms for eight lines are eventually summed up to make the full histogram for the cell.

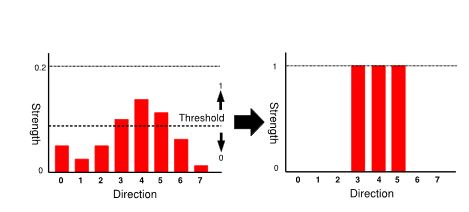
Since the maximum value for the gradient strength is 361, each orientation of a partial histogram of five luminance values can be expressed with 12 bits. Thus, the required capacity of the Block RAM buffers corresponds to 12 bits×8 orientations×63 cells×4 lines. Since the luminance values are streamedin every two clock cycles, full histograms are streamed out every 16 clock cycles.

* 1. Normalization and Binarization

The normalization process is carried out every block, which consists of 2x2 cells. With every histogram consisting of 9 parts, we have a two-dimensional array with a size of 36x36. Normalization invovles summing all 9 elements of a histogram, summing it with elements of other histograms, and calculating the root mean square value of the histograms. With a picture size of 640x480, we have a total of 78 x 58 x 72 dimensions from a single image data. Because of this complexity, as before we employ a simplification method to calculate normalization, as will be described in the next chapter. This process involves a maximum of three shifts.

Normalization will also use the streamed structure, similar to the one used for histogram generation. The process is carried out for a moving 2x2 window of cell histograms. For this process, three lines of 2-stage shift registers, along with 2 lines of 61-stage RAM blocks. Every time a new histogram is generated, 72 dimensions of histogram elements in the 2x2 cell are summed in two clock cycles. The next process, as mentioned, requires three shifts, and one conditional check, resulting in two cycles. Thus, the total amount of clock cycles required by the normalization process is 4 cycles.

The next part, binarization, strives to reduce the amount of memory required by the next processes. If we express the HOG as a 8-bit double precision floating point, it would require 78 x 58 x 72x 8-byte memory capacity, which is more than the SRAM can handle. In our scheme, each one dimension feature will be binarized using a threshold value so that every gradient can be represented as a single bit. Since the gradient orientations are quantized into 8 orientations, a feature amount obtained from a certain cell can be described by eight bits (or 0 to 255). An example can be seen in the figure below. Using this scheme, a 1/64 reduction can be achieved. However, this scheme will also reduce performance for several pictures, as will be shown in the last chapter.



* 1. Support Vectoring Machine

A support vectoring machine can be used when the data has exactly two classes. An SVM classifies the data by locating the best plane that separates the data points of one class from the other class. The *best* plane, in this sense, for an SVM is the one with the largest *margin* or euclidean distance between the two classes. Here, margin means the maximum width of the slab which is parallel to the plane that has no data points inside it. In other words, the best plane is the one whose distance from the nearest data point is the largest.

In this implementation, the SVM is represented by a weighing matrix and a biasing number. The histograms are multiplied with the weighing matrix, resulting in a single number, which is then added with the biasing number. If the result is a positive number, then the conclusion is that a human is detected.

As the SVM from the LSI-Contest website still has unpredictable flaws, we decided to train our own SVM classifier using two different datasets, namely MIT pedestrian database and INRIA database. A boosting algorithm is used with the SVM serving as the ‘weak’ classifier, based on [4]. The result of the training process is a new weighing kernel and a new bias number. These new components prove more effective at detecting humans in images outside of the one provided by the website.

* 1. VGA and Display Module

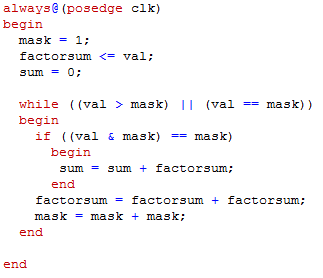
To display the image, we use a LG® Flatron L1742S monitor, using the ram as the screen buffer. The ram block in question is the original picture, which will be edited at the end of the detection process, should a human be detected. The editing process involves drawing a red box along the edges of the window being processed. This is achieved by simply changing the value of all the pixels at the edge of the box by a RGB value of FF,00,00. This means changing the value of the RAM block corresponding to the pixels at the edge of the box.

The Altera DE2-115 is equipped with a ADV7123 video DAC. Therefore, the ports that we must take into consideration are the Hsync, Vsync, and the RGB values, each with its own port.

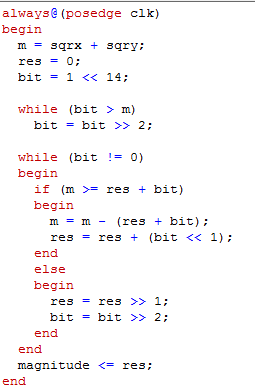
1. Simplification of the Calculation Process and Other Appeals

### 3.1 Magnitude Calculation

So, we will made an approximation to get these values of magnitudes and angles.



Verilog Algorithm to approximate a multiplier operation



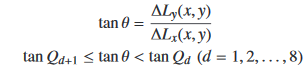
Verilog algorithm to approximate a root operation

### *3.2 Angle Calculation*

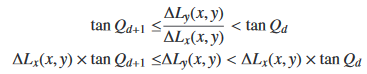
Angle calculation involves trigonometric functions such as the arctangent, which is fairly complicated to implement in an FPGA. One alternative is the CORDIC algorithm, however it takes many iterations to produce an accurate result. Combinatorial CORDIC circuits that can output a result in a single clock cycle exists, but they consume a lot of logic elements.

However, an accurate result is not needed because the histogram quantizes the angles into 9 ranges. Therefore, it is only needed to find out to which range the gradient belongs. Using another method, angle calculation can be achieved in two cycle using only shifters and conditional checks.

The method is as follows. The tangent values of the quantization threshold angles Qd(d = 1, 2, . . . , 8), can be precomputed since they are constant values. Instead of calculating the arctangent, we represent the tangent as dy/dx. Therefore, if the value of tan is known, we get a quantized gradient orientation by the conditional expression with tan Qd and tan Qd+1 as shown in the equations below.



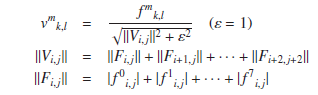
However, by multiplying dx with all the constant tangents, and by inserting the tan equation into the conditional statement, we get the following result.



Therefore, we need only to multiply dx with the constant tangent values which are precalculated, which can be done using shifters. Next, we need to evaluate the conditional statements to determine which bar in the histogram does the pixel in question belong to. Therefore, the angle calculation can be done in two clock cycles.

### 3.3 Normalization

The normalization process as used by MATLAB® and described in the equations below needs the calculation of a square root and division by fractional numbers, making FPGA implementation difficult.



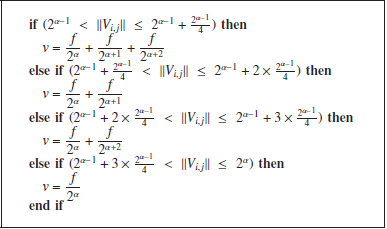
Therefore, we take an approximation approach as follows, which expands the method proposed in [5].If the denominator of the equation above (Sqr(||Vi, j||2 + ε2)) is approximated by 2α as 2α−1 <Sqr(||Vi, j||2 + ε2) ≤ 2α, the division for the normalization can be replaced by a shift operation. However, naive approximation to the nearest power-of-two value increases the normalization error. To mitigate the approximation error, we divided the interval between 2α−1 and 2α into four sub-intervals.

The range of 2α−1 to 2α can be divided into four intervals; (2α−1, 2α−1 + 2α−14 ], (2α−1 + (2^α−1)/4 , 2α−1 + 2 × 2^(α−1)/4 ], (2α−1 + 2 ×2^(α−1)/4 , 2α−1 + 3 × 2^(α−1)/4 ], and (2α−1 + 3 × 2^(α−1)/4 , 2α]. Here, ε = 1 and this is significantly smaller than ||Vi, j||2. Therefore, the conditional statements in the normalization equation can be derived when it is approximated that:

Sqr( ||Vi, j||^2 + ε^2) =||Vi, j||.

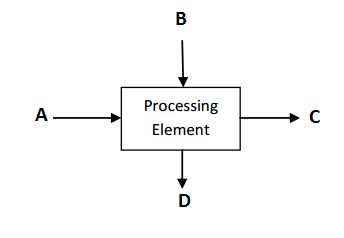
The figure below shows comparison results of approximation errors for our quadrisection approach and the naive power-of-two approach, in the case of f m k,l = 361. The results show that the normalization errors are effectively reduced with the relatively simple calculation process.

Since the maximum value of ||Vi, j|| is 81,225, the maximum number of shift operations for the division is 19. Thus, as a fraction part, 19 bits of 0s are appended to the LSB side of f m k,l in advance of shifting, and obtained vm k,l is also expressed with a fixed point arithmetic number with 14 + 19 = 33 bits.

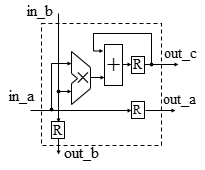


### 3.4 Matrix Calculation

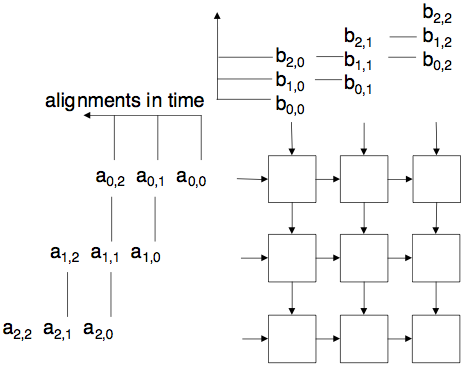
Matrix calculation is required for the SVM detection and the mask with magnitude calculation. To make matrix calculation more efficient, systolic arrays are used. Systolic architecture consists mainly of an two-dimentional array of uniform processing elements, in which data flows between the neighboring processing elements, synchronously, from four different directions. Processing elements take data from Top and Left and output the results to Right and Bottom.

[](http://1.bp.blogspot.com/-Vb9RAHN1sV8/VbulUV-2zBI/AAAAAAAAAu8/SRXsADo8ZCI/s1600/processing%2Belement.PNG)

Here, each processing element, consisting of a multiplier and an adder, performs four operations, namely fetching, multiplication, shifting & addition**.** As the following figure shows, "in\_a", "in\_b" are inputs to the processing element and "out\_a", "out\_b" are outputs to the processing element. "out\_c" is to get the output result of each processing element, connected directly to the output of the module.

[](http://3.bp.blogspot.com/-gOyEfuAfNHk/Vbxt05BG_YI/AAAAAAAAAvQ/MRCKX4ULZzg/s1600/pe.PNG)

The processing elements are arranged in the form of a matrix. In this case we use, multiplication of 3x3 matrices, because by examining the weighing matrix, we find that there is a pattern in the array, in which 3 zeros are followed by 3 numbers. The following figure depicts how matrix A and B are fed into PE(processing element) array.

[](http://2.bp.blogspot.com/-XvlZz8VB6lE/VbyeTqj1TeI/AAAAAAAAAvg/oZysEUhCfO8/s1600/PE%2Barray.png)

### 3.5 Scaling

Scaling is needed to detect humans with different sizes in images. The number of scaling processes depends on the size of the image. By following the algorithm in the MATLAB® file,we compare the image to a 134 x 70 pixels image and then we scale it up by 1.25n each time, until it reaches the full image size.

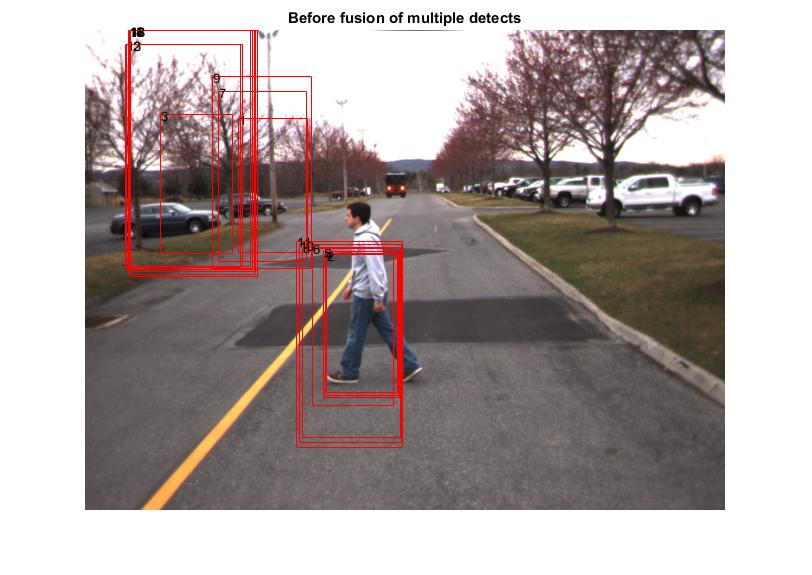
After we get the scale number, we will iterate the process of human detection up to n times. With each of iteration, the scaled image will have it’s matrix rescaled too. The new matrix’s index number are multiplier of the scaler itself.

The implementation of resizing the input image is similar with MATLAB algorithm reference provided by the lsi contest website, with some change such as in the Scaler module, instead using bilinear interpolation , we use iteration of scaling loop.

|  |
| --- |
| function pict = Scaler(img, scale)  [rows, cols ] = size(img);  row1 = floor(rows/scale);  col1 = floor(cols/scale);  pict= zeros(row1,col1,'uint8');    for i=1 : row1  for j=1 : col1  a = floor(scale\*i);  b = floor(scale\*j);  pict(i,j) = img(a,b);  end  end    end |

### 3.6 Adaboost

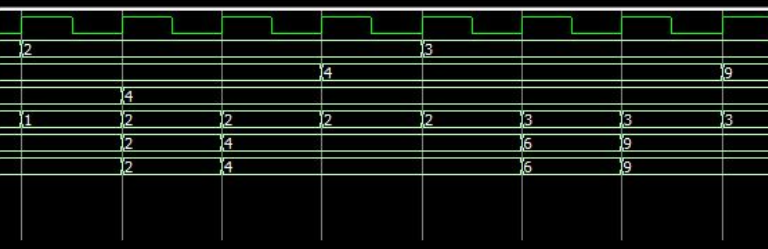
From experimenting with different images, we have found that the SVM provided by lsi contest website still has several flaws, as several pictures show that the SVM either makes a false positive or does not detect humans in a picture. An example is provided below.



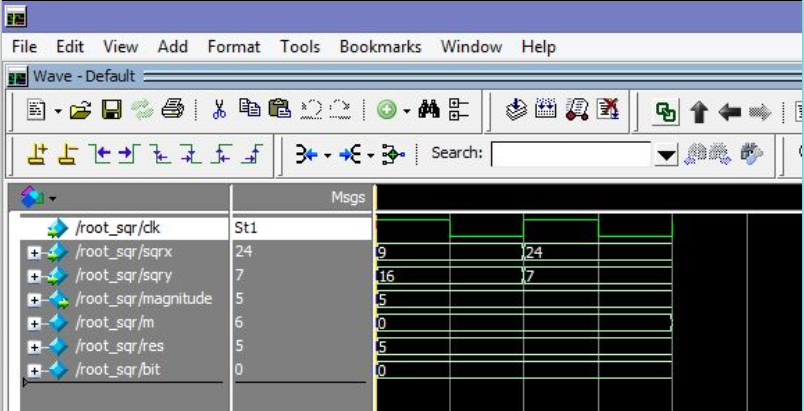
Therefore, a new SVM was trained, using both the INRIA database and the MIT pedestrian database. The new SVM is a “weakened” version of the SVM, but we also used boosting algorithms to improve performance of the SVM. the size of the weights assigned by Adaboost to the weak SVMs, serve as an indication of which data points are likely to become support vectors in the final model, and hence can be useful in the implementation of editing algorithms.

# 4. Simulation Results and Analysis

## 4.1 Magnitude Calculation

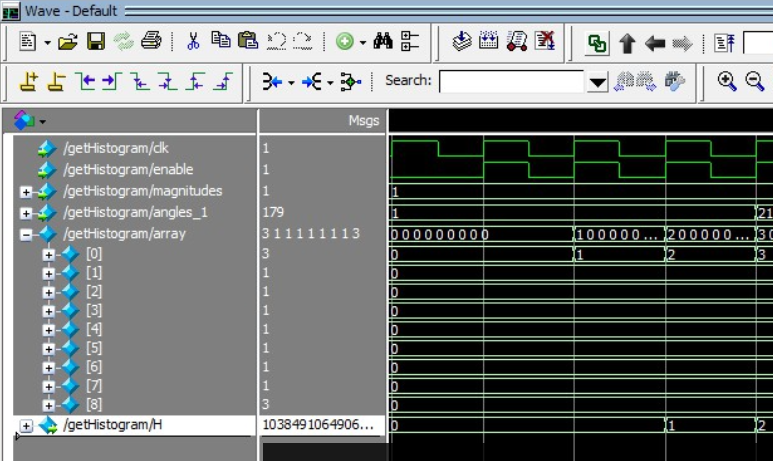


Square

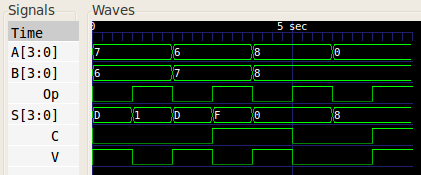


Square Root

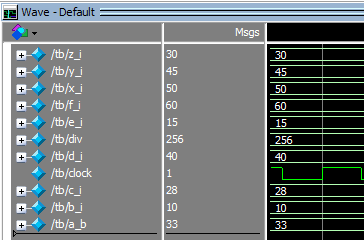
## 4.2 Histogram Calculation



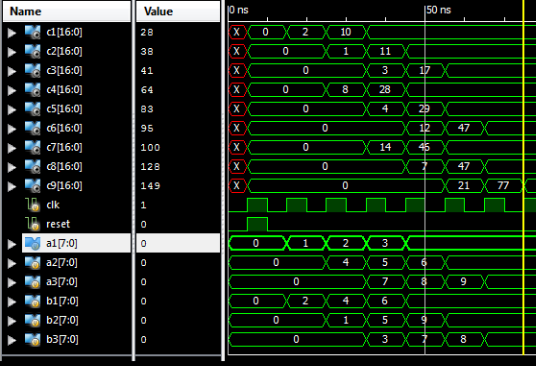
## 4.3 Preprocessing



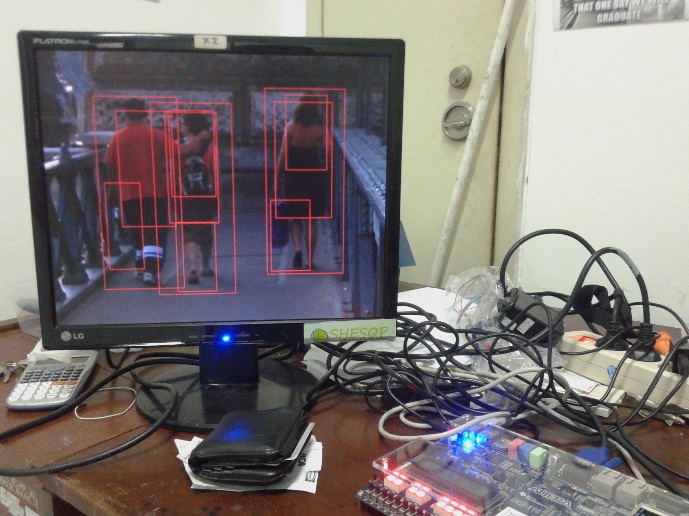
## 4.4 Histogram Normalization



## 4.5 Matrix Multiplication



## 4.6 Overall





## 4.7 Critical Path & Analysis

5. Implementation and Conclusion

The conclusions in this report are:

* We use gradien to compute histogram so we can conclude if there is an object in an image by using SVM.
* We’re succeed to approximate multiplier, divider and root operation by using only shifter and iteration.

References

1. <http://www.lsi-contest.com/shiyou_1e.html>, accessed in 27 December 2016
2. <https://www.mathworks.com/help/fixedpoint/ug/convert-cartesian-to-polar-using-cordic-vectoring-kernel.html>, accessed in 27 December 2016Adiono, Trio. *Perancangan Sistem VLSI,* Pusat Mikroelektronika ITB, Bandung, 2012.
3. Ciletti, M.D. *Advanced Digital Design With the Verilog HDL,* Publishing House of Electonics Industry, Beijing, 2010.