FPGA Design of Real Time Hardware for Face Detection

Kibum Suh*

1Department of Rail System, Woosong University, Deajeon, South Korea, kbsuh@wsu.ac.kr

*Correspondence: kbsuh@wsu.ac.kr; Tel.: (+82-42-630-4657)

ABSTRACT - This paper proposes the hardware architecture of face detection FPGA hardware system using the AdaBoost algorithm. The proposed structure of face detection hardware system is possible to work in 30 frames per second and in real time. And the AdaBoost algorithm is adopted to learn and generate the characteristics of the face data by MATLAB, and finally detected the face using this data. This paper describes the face detection hardware structure composed of image scaler, integral image extraction, face comparing, memory interface, data grouper and detected result display. The proposed circuit is so designed to process one point in one cycle that the proposed design can process full HD (1920x1080) image at 70MHz, which is approximate 2316087 x 30 cycle.

General Terms: Pattern Recognition, Security, Algorithms et. al.

Keywords: FPGA, Adaboost, Real time, face detection

1. INTRODUCTION

Recently, research on biometric technologies that can perform personal authentication, information protection, and identity verification using the human body such as face, fingerprint, and iris has been actively conducted. Object recognition in video sequence or images is an important field in computer vision, image processing applications, security, bioinformatics, and artificial intelligence.

Object detection involves extracting information from an image (or a sequence of) frames, processing the information, and locating specific objects and images in the information. This process is computationally intensive, and there have been several attempts to implement object detection algorithms in hardware, especially in embedded and real-time systems [1-6]. Most of the proposed works are designed for FPGA target implementations. Until now, many hardware design methods of adaboost algorithm using FPGA have been proposed. The method of proposing the face recognition hardware using the Adaboost method includes [7-13].

In this paper, face recognition system using FPGA board is proposed based on the Adaboost algorithm by Viola and Jones [14]. For the implementation of hardware, HAAR-like feature used for face detection, an integral image designed for speed improvement, and a method of boosting from a weak classifier to a strong classifier were used. Through comparison with the existing hardware structure [7][9][13], the advantages of this hardware structure will be verified.

2. THE TRAINING USING ADABOOST ALGORITHM

In this paper, we used the HAAR-like feature that detects and compares face regions in images using Viola and Jones's Adaboost algorithm. Figure 1 shows the feature vectors obtained through learning the Adaboost algorithm of Viola and Jones and the feature vectors used in this paper. The Adaboost algorithm needs to generate feature vectors for the pre-processing process. As shown in figure 1, for a 20x20 image, 17,100 vectors were created for Type 0 and Type 1 patterns, 7600 vectors were created for Type2 and Type3 patterns, and 8100 vectors were created for Type4. For training, training was performed using MATLAB®, and the MIT CBCL Face Data Set was used. The Face data set consists of 6,977 cropped images (2,429 faces and 4,548 non-faces).

The total number of feature vectors used in this paper is 57,500, of which only 200 feature vectors are used for face recognition.

The weights of the classifier are modified while repeating step by step. At each step, learning proceeds in such a way that only the features that play a decisive role in detecting faces from among the feature sets are left and the rest are removed [16]. Therefore, this module is designed to receive the data output from the integral image processing process and directly compare the current data and the trained data value to determine whether it is a human face or not.
3. HARDWARE ARCHITECTURE

With respect to the YCbCr signal of the image input from the camera, the VIM module detects only the Y signal, which is a black and white signal, and stores it in the first memory (org_sram) to input data to the module for face recognition.

Figure 2 shows the process of performing the image size reduction section and integral image extraction section on the stored data, and outputting the facial region area on the LCD screen of the Vertex5 LX330 board through the Adaboost classifier.

In this paper, since the image size is reduced to 0.75 times that of the previous image, the reduction process is repeated several times depending on the size of the screen to detect the face area.

In order to apply the Adaboost algorithm, it is necessary to calculate the integral image. Therefore, in the data flow, the image reduction process and the integral image extraction process are simultaneously performed on the first frame start signal, and the face region extraction, image reduction and integral image extraction processes are repeatedly performed from the second frame start signal. Since the pipeline structure was applied, time wasted in the data processing process was minimized and an efficient hardware structure was constructed. Figure 3 shows the pipeline flow diagram of hardware for face recognition.

3.1 Integral image calculation module

Many image processing filters perform window operations, and as the window increases, the time complexity increases accordingly. One of the methods to reduce the time complexity is the integral image processing method. When the average filter is used in the integral image processing, a window of a certain size is set in each pixel, and a new pixel value is set by taking the average value of the pixels in the window. If the integral image processing method is used, the sum of the pixel values for a rectangle of any size can be calculated within a certain time regardless of the size.

In addition, the process for detecting a specific object in the image has the advantage that it can quickly obtain the Haar-like feature value in the next pipeline after converting the data value of the original image into an integral image. Figure 4 shows the data flow of the integral image processing designed in this paper and the operation to obtain the pixel value of region D.

3.2 Image resizing

In the image resizing process, as shown in Figure 4, the input image data is resized in units of 4 pixels. The first pixel value to be resized is the sum of the first pixel value before resizing multiplied by 3/4 and the second pixel value multiplied by 1/4. The second pixel value to be resized is multiplied by 1/2 and added to the second and third pixel values before resizing.

The third pixel value to be resized was used as the sum by multiplying the third pixel value by 1/4 and multiplying the fourth pixel value by 3/4. A 4x4 image was reduced to 3x3 by performing a resizing process for each pixel value of the image. As shown in Figure 5, if the size of the original image is 480x272, the image after the image resizing process has a size of 360x204.

Figure 1: The set of feature vector

Figure 2: The overall architecture of face detection hardware

Figure 3: The pipeline flow diagram of face detection

Figure 4: Integral image calculator and data flow

Figure 5: Image resizing process
Figure 5: Image resizing operation

Figure 6 shows the actual hardware output screen through the image reduction process proposed in this paper.

Figure 6: Image resizing results

3.3 Saving default size windows

The data value that has undergone the integral image processing process is stored in the memory. In the next pipeline, the integral image data is read in units of a basic window size of 20×20, stored in a register, and then supplied to the face detection module. As shown in Figure 7, this module reads the 20×20 integral image and stores it in the register when the start signal input from the face region detection controller is generated. When a start signal is generated, the execution process is reversely stored from registers 19 to 0, and when all data in each register is filled, the data is supplied to the face area detection module to perform the face recognition process.

Figure 7: Saving default size windows

Table 1 shows the results of calculating the required time for a 480x272 size image. Based on the real-time processing method, as 30 frames per second must be processed, the total number of cycles is 132,606 × 30 = 3,978,180, which operates at about 4 MHz.

<table>
<thead>
<tr>
<th>Calculation method</th>
<th>Width-20/2</th>
<th>Height-20/2</th>
<th>First row processing cycles(s)</th>
<th>Row change cycle (B)</th>
<th>Total cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1920×1080</td>
<td>995</td>
<td>100500</td>
<td>52010</td>
<td>3023130</td>
</tr>
<tr>
<td>2</td>
<td>1440×960</td>
<td>710</td>
<td>13400</td>
<td>57250</td>
<td>677250</td>
</tr>
<tr>
<td>3</td>
<td>1080×720</td>
<td>530</td>
<td>10700</td>
<td>320650</td>
<td>320650</td>
</tr>
<tr>
<td>4</td>
<td>810×455</td>
<td>395</td>
<td>7505</td>
<td>178935</td>
<td>178935</td>
</tr>
<tr>
<td>5</td>
<td>607×341</td>
<td>294</td>
<td>5586</td>
<td>99666</td>
<td>99666</td>
</tr>
<tr>
<td>6</td>
<td>455×225</td>
<td>118</td>
<td>4142</td>
<td>55154</td>
<td>55154</td>
</tr>
<tr>
<td>7</td>
<td>341×161</td>
<td>86</td>
<td>27370</td>
<td>80429</td>
<td>80429</td>
</tr>
<tr>
<td>8</td>
<td>255×131</td>
<td>52</td>
<td>1495</td>
<td>16658</td>
<td>16658</td>
</tr>
<tr>
<td>9</td>
<td>193×107</td>
<td>44</td>
<td>8050</td>
<td>6030</td>
<td>6030</td>
</tr>
<tr>
<td>10</td>
<td>143×80</td>
<td>30</td>
<td>1178</td>
<td>1209</td>
<td>1209</td>
</tr>
<tr>
<td>11</td>
<td>107×60</td>
<td>20</td>
<td>1672</td>
<td>2508</td>
<td>2508</td>
</tr>
<tr>
<td>12</td>
<td>80×45</td>
<td>13</td>
<td>720</td>
<td>1290</td>
<td>1290</td>
</tr>
<tr>
<td>13</td>
<td>60×31</td>
<td>7</td>
<td>240</td>
<td>620</td>
<td>620</td>
</tr>
<tr>
<td>14</td>
<td>45×24</td>
<td>2</td>
<td>36</td>
<td>373</td>
<td>373</td>
</tr>
<tr>
<td>Sum of Cycles/frame</td>
<td>14400</td>
<td>118304</td>
<td>5326600</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In Table 2, in the same way as Table 1, the number of cycles for full HD is 2,316,087x30=69,482,610, indicating that it operates at approximately 70MHz and real-time processing is possible.

3.4 Face detection module

The face detection method using the Adaboost algorithm is learned using a feature set extracted from a learning image set composed of face and non-face images. The learning process is performed through repeated calculations, and consists of the steps of feature selection, error rate calculation corresponding to the observed value, and weight updating.

The classification ability is strengthened at each iteration because weak classifiers are combined to form a strong classifier through stages.

Table 1. Cycle calculation for 480x272 image

Table 2. Cycle calculation for 1920x1080 image
By repeating the process of increasing the weight of the misclassified training image and decreasing the weight of the correctly classified training image, the HAAR-like feature with the minimum error rate is selected.

\[
h_n(x) = \begin{cases} 
1 & \alpha_i h_i(x) + \ldots + \alpha_n h_n(x) \geq \frac{1}{2} (\alpha_i + \ldots + \alpha_n) \\
0 & \text{otherwise}
\end{cases} 
\]

Equation 3.1 is a strong classifier, meaning that it is classified by linearly combining \( n \) features. The weight of the classifier is modified while repeating step by step. At each step, learning proceeds in such a way that only the features that play a decisive role in detecting faces from among the feature sets are left and the rest are removed.

In Figure 8, as the critical path of the adder tree used in the face region detection module using Equation (3.1) increases, F/F is used to reduce the critical path of the adder tree. Through this, it was possible to reduce the delay of the critical path used in the face region detection module to 5 ns.

3.5 Reducing bit of integral image

In the process of calculating the integral image on the actual hardware operation board, the size of the RAM increases. To solve this problem, a calculation method through modulo operation was used [15].

3.6 Removing Overlap

The Overlap module is designed to make a simple expression by treating two or more rectangles as one rectangle when they are located inside the same position, and is performed as shown in Figure 9. The position of the inner rectangle is found with the x-position and y-position values output from the face recognition controller module. In Figure 9, in order to know the inclusion relationship of rectangular boxes such as (1), (2), (3), it is designed to store up to 20 coordinates output from the face recognition controller and find the inner rectangle.

The result obtained by applying the overlap removal module is output on the screen of the target board, and Figure 10 shows the actual hardware output screen.

4. EXPERIMENTAL RESULTS

This paper verified hardware by using MODELSIM and comparing the results from MATLAB, to the output data values of each module, the integral image part, the image reduction part, and the last face region detection part as shown in Figure 11 using MODELSIM.

Table 3 shows the results of comparative analysis of this paper with other systems. In this paper, a 480x272 (VGA) image and a clock frequency of 24 MHz were used for test. Other studies aim to process 360x240 images and show low frame rate performance. Compared to other researches, it was confirmed that it operates at a high frame rate and low frequency clock for a large screen size.
Table 3. The comparison with others

<table>
<thead>
<tr>
<th></th>
<th>FPGA</th>
<th>Logic elements</th>
<th>Window size</th>
<th>Image size</th>
<th>Frame/sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.Kyu[7]</td>
<td>XCV2930</td>
<td>116238</td>
<td>24x24</td>
<td>220x240</td>
<td>60fps</td>
</tr>
<tr>
<td>J.Ch[9]</td>
<td>XCV3X11 0</td>
<td>62,890</td>
<td>20x20</td>
<td>220x240</td>
<td>26fps</td>
</tr>
<tr>
<td>K.-C. Chang[13]</td>
<td>XCA7A200</td>
<td>Not Conga parallel</td>
<td>80X80 And 60X60</td>
<td>360X240</td>
<td>1014fps/825MIO.</td>
</tr>
<tr>
<td>Our System</td>
<td>XCV3X03 0</td>
<td>74,757</td>
<td>20x20</td>
<td>40X272, 150X108</td>
<td>112Bps/499-272/20MHz</td>
</tr>
</tbody>
</table>

5. CONCLUSION

In this paper, we designed a face recognition hardware system for DVR system that detects a human face from an image and outputs it on the screen. The hardware designed by creating a reference vector using MATLAB was verified using MODELSIM, and the synthesis was performed by FPGA verification using Xilinx ISE Design Suite and by Synopsys’ Design Compiler in 0.18um process, 35% of Vertex5 LX330 Slice LUT 74,757 was used. It is designed to operate at an operating frequency of 24MHz, so it is possible to process 112 frames per second for VGA-level images and 30 frames per second in the case of 70MHz clock system for full HD images.

6. ACKNOWLEDGMENTS

This research is based on the support of 2021 Woosong University Academic Research Funding.

REFERENCES


© 2022 by the Kibum suh. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).