Regular Article

Real-Time Face Detection and Human Tracking System on FPGA Cyclone-V

Huu Luong Nguyen, Minh Son Nguyen, Tri Nhut Do

Faculty of Computer Engineering, University of Information Technology, Ho Chi Minh City, Vietnam

Correspondence: Tri Nhut Do, trinhutdo@uit.edu.vn

Communication: received 20 July 2021, revised 20 September 2021, accepted 21 September 2021
Online publication: 16 May 2022, Digital Object Identifier: 10.21553/rev-jec.297
The associate editor coordinating the review of this article and recommending it for publication was Prof. Tran Manh Ha.

Abstract– **Face detection in image sequence (real-time video stream) has been an active research area in the computer vision field in recent years due to its potential applications such as surveillance cameras, human computer interfaces, smart rooms, intelligent robots and biomedical image analysis. Face detection is a process that determines whether an image has a face or not. In this paper, an embedded system for detecting and tracking human faces in real-time video stream implemented on FPGA DE10-Nano is proposed. The system can be divided into two parts: data streaming, data processing. Experimental results show that the system is capable of accurately detecting faces of up to 5 different people at a distance of up to 1.5 meters from the camera, coexisting in the same frame in resolution of 320 × 240 pixels with a detection speed of only several hundred milliseconds prove the feasibility of the system. A comparison with similar existing projects will be discussed for evaluation and conclusion as well.**

Keywords– **Embedded system, human tracking system, face detection system, FPGA.**

1 Introduction

Field programmable gate arrays (FPGAs) [1] have become extremely popular in almost all the application domains such as computer vision, object detection and tracking. The FPGA is an ideal device to carry out works related to Video and Image processing by the parallel processing capability.

Paul Viola and Michael Jones introduced in [2] an effective method for object detection using Haar feature-based cascade classifiers. Objects in images are detected based on a machine learning algorithm in which a cascade function is trained from a lot of positive and negative images.

A Virtex-II 2V1000 using a MicroBlaze processor for face detection employed Viola-Jones algorithm introduced by Vinod Nair, Pierre-Olivier Laprise and James J. Clark in [3]. The results show that the system can detect people accurately at a rate of about 2.5 frames per second when it is running at 75 MHz, communicating with dedicated hardware over FSL links.

In addition, another system for face detection introduced by Hichem Ben Fakih, Ahmed Elhossini and Ben Juurlink in [4] was also employing Viola-Jones algorithm as well. The proposed design is able to discover faces in real-time with high accuracy. Speed-up is achieved by exploiting the parallelism in the design, where multiple classifier cores can be added. To maintain a flexible design, classifier cores can be assigned to different images. Moreover using different training data, every core is able to detect a different object type. The Zynq-7000 SoC from Xilinx is used, which features an ARM Cortex-A9 dual-core CPU and a programmable logic (FPGA). The current implementation focuses on the face detection and achieves a real-time detection at the rate of 16.53 FPS on image resolution of 640 × 480 pixels, which represents a speed-up of 6.46 times compared to the equivalent OpenCV software solution.

That's why the Viola-Jones algorithm is a well-known method for face detection systems utilizing Xilinx FPGA devices. Recently, the Haar feature-based cascade classifier was widely applied on processing to identify and count the oil palm trees [5]; to detect vehicles on moving for monitoring, avoiding accident and regulating traffic [6]; to recognize and verify handwritten signature with the accuracy of 92% for many different types of writers' languages and style [7]; to detect the defects of five types on different kinds of textiles [8].

In this paper, Viola-Jones algorithm using Haar-like features for face detection is employed in the proposed system. All images captured from the OV7670 camera module and stored in Image Frame Buffer will be processed for face detection. The proposed system is implemented on DE10-Nano Development Board [9]. This board has a robust hardware design platform built around the Intel System-on-Chip (SoC) FPGA with 5,570 Kbits embedded memory, 110 K programmable logic elements. Besides, it also supports other components such as: processor, peripherals, high-speed DDR3 memory, analog to digital capabilities, Ethernet networking, and much more that allow users to have not only simple but also flexible designs or even a high-performance, low-power processor system. The board top view along with its components descriptions are shown in Figure 1.

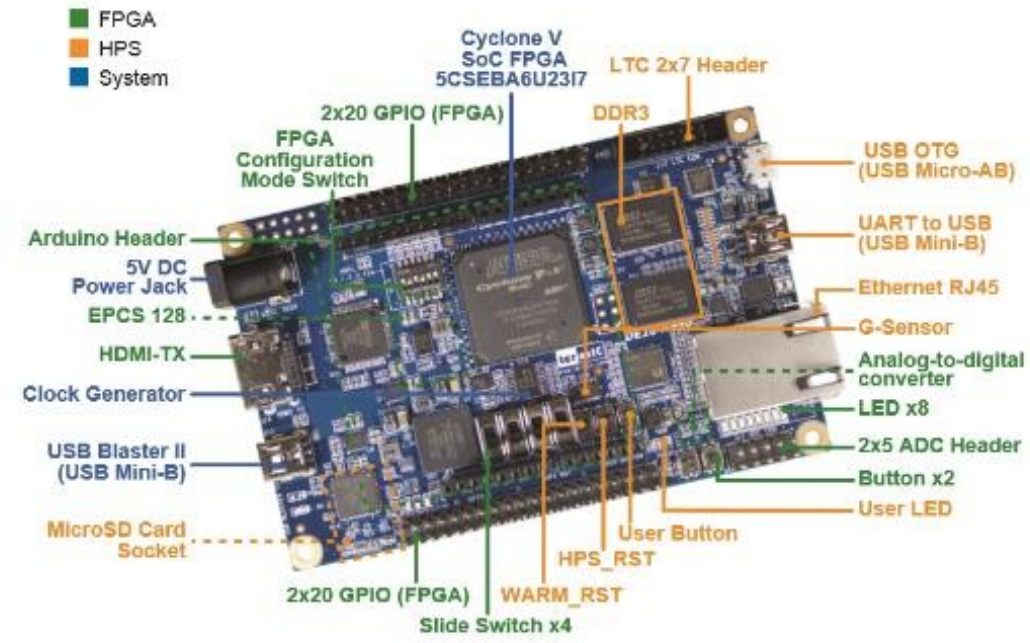

Figure 1. DE10-Nano Development Board.

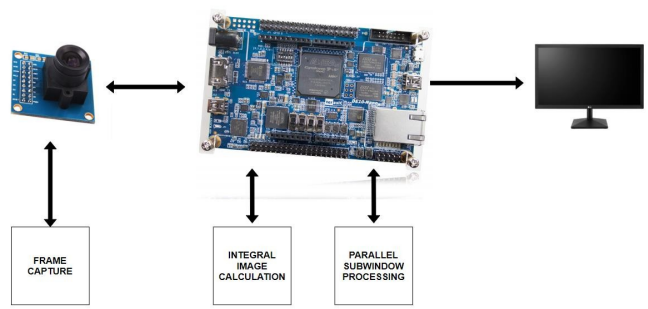

Figure 2. Overview of the proposed system.

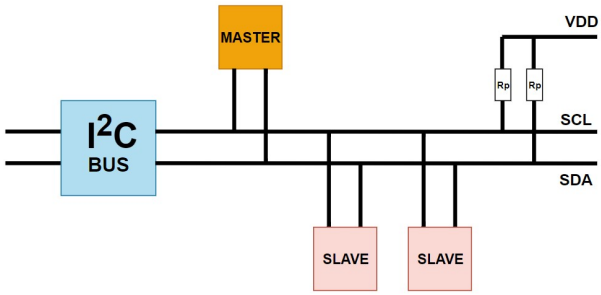

Figure 3. I2C protocol for camera sensor and SoC Cyclone-V in DE10-Nano.

The rest of this paper is organized as follows. Designs of the proposed system followed by a block diagram for face detection based on Viola-Jones algorithm using Haar-like features are described in Section 2. Section 3 describes the hardware system and software setup, an evaluation of experimental results is presented as well. Section 4 concludes the paper with future directions.

2 Proposed System

2.1 System Model

The design of the proposed system is made up of three main components: the OV7670 camera, the DE10-Nano development board, and a HDMI monitor (as shown in Figure 2). The processing step includes frame capture, integral image calculation, parallel sub window processing.

The OV7670 camera is connected to the DE10-Nano development board employing I2C protocol (see Figure 3). The communication between them is via SDA and SCL pins where the DE10-Nano development board acts as Master and the OV7670 camera acts as Slave. Images captured by the OV7670 camera will be converted into RGB444 format (ADC 12bits). Captured images are stored in the captured Block and then will be copied into the face detection Block in order to apply the Viola-Jones algorithm.

The FPGA consists of a number of components. These components are image frame capture, processing and display. All of these components are synchronized and controlled from a control unit. They are listed as in Table I. A phase lock loop (PLL) generates 100 MHz clock signals that are contributed to each component for synchronous processing in high speed between modules in the system. Figure 4 shows The block diagram of the real-time detection and tracking algorithms on FPGA Cyclone-V of DE10 Nano Board.

The proposed embedded system is designed that consists of a number of components. These components are image frame capturing, image processing and image displaying, etc. All components are synchronized

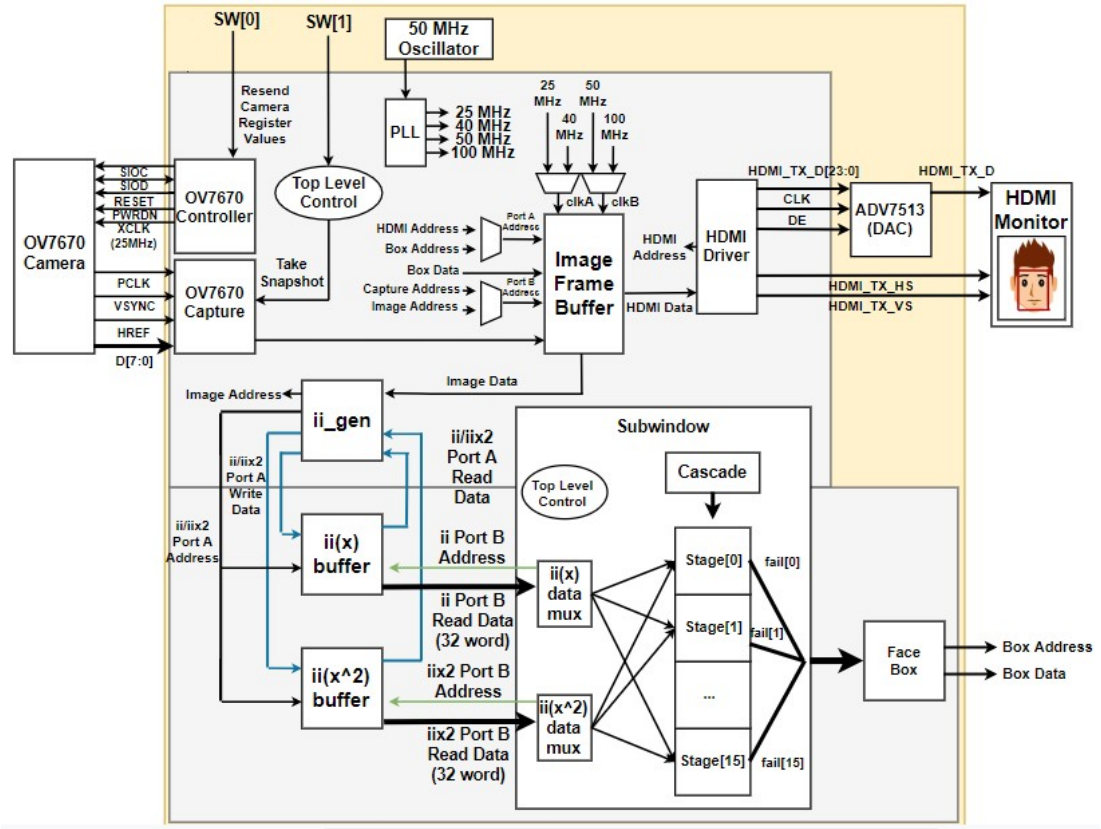

Figure 4. The block diagram of the real-time detection and tracking algorithms on FPGA Cyclone-V.

Table I
FPGA Resources

Family	Cyclone V
Device	5CSEBA6U23I7
Logic utilization (in ALMs)	10,865/41,910 (26 %)
Total registers	16284
Total pins	60 / 314 (19 %)
Total block memory bits	2,826,805/5,662,720 (50%)
Total DSP Blocks	87 / 112 (78 %)
Total PLLs	1 / 6 (17 %)

Table II
FPGA Configuration Modes

MSEL[4:0]	Configuration Modes	Description
10010	AS	FPGA Configure from EPCS
01010	FPPx32/Compressi on Enabled/Fast POR	FPGA Configure from HPS software: U-Boot, with image stored on the SD card, like LXDE Desktop (default)
00000	FPPx16/Compressi on Disabled/Fast POR	FPGA Configure from HPS software: U-Boot, with image stored on the SD Card

and controlled from a control unit. A phase lock loop (PLL) generates clock signals that are contributed to each component for synchronous processing between modules in the system.

2.2 FPGA Configuration Mode Setting

Figure 5 shows the pins for configuring operation modes of the DE10-Nano board. When it is powered on, it can be configured from EPCS or HPS. The MSEL[4:0] pins are used to select the configuration scheme. It is implemented as a 6-pin DIP switch SW10 on the DE10-Nano board.

When the board is powered on and MSEL[4:0] set to “10010”, the FPGA is configured from EPCS, which is pre-programmed with the default code. If developers use the "Linux LXDE Desktop" SD Card image, the MSEL[4:0] needs to be set to “01010” before the board is powered on. Table II describes modes according to MSEL (SW10) codes.

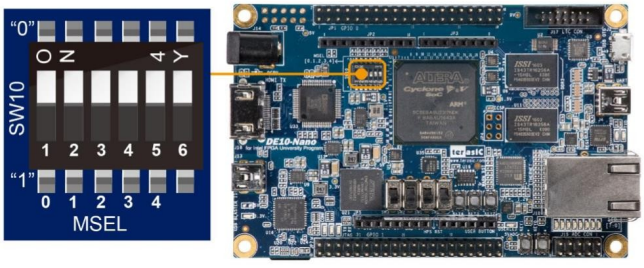

Figure 5. FPGA Configuration Mode Setting.

2.3 Viola-Jones Algorithm

Viola-Jone method (as shown in Figure 6) that is employed in this paper consists of elements of Haar Cascade classifier, integral image and adaBoost algorithm. The algorithm needs a lot of positive images (images of faces) and negative images (images without faces) to train the classifier. In order to detect faces, features must be extracted. Each feature is a single

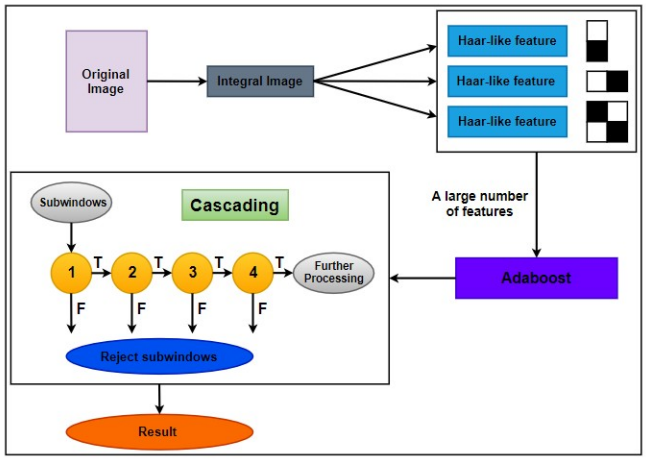

Figure 6. Viola-Jones algorithm overview.

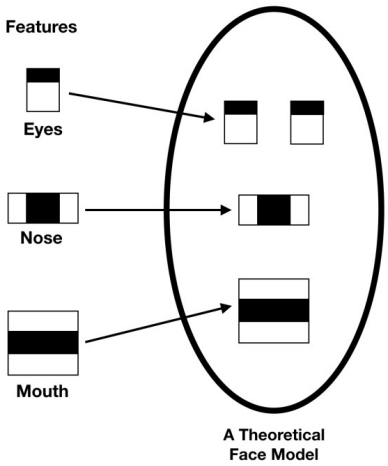

Figure 7. Haar features for a face.

value obtained by subtracting the sum of pixels under the white rectangle from the sum of pixels under the black rectangle. Haar cascades method processes on gray images. Color captured image will be converted into gray image, each pixel will have a value from 0 to 255 (depending on how gray or black it is). Adaboost, short for Adaptive Boosting, is a machine learning meta-algorithm formulated by Yoav Freund and Robert Schapire. It can be used in conjunction with many other types of learning algorithm to improve its performance.

Haar-like features are digital image features used in object recognition. In this project, it is black and white rectangle to represent face characteristics as shown in Figure 7.

The filters including Haar features are assigned into an image in order to capture features in the face like the nose, the distance between two eyebrows, etc. For face detection, Viola-Jones uses 4 basic features as shown in Figure 8.

Beside these basic features, more complex features including edge feature (as shown in Figure 9), line feature (as shown in Figure 10) and center surround feature (as shown in Figure 11) are added in order to capture more details of the human face as well.

Figure 8. Simple features that are used by Viola-Jones.

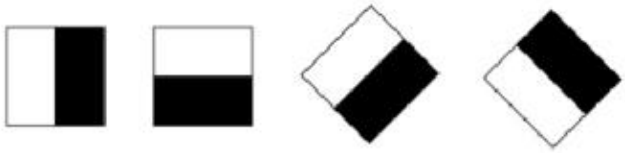

Figure 9. Edge feature.

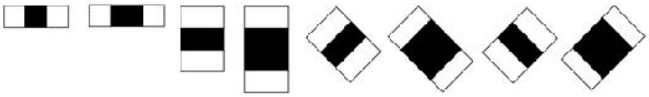

Figure 10. Line feature.

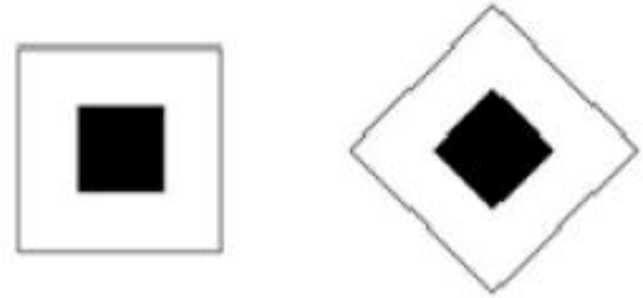

Figure 11. Center surround feature.

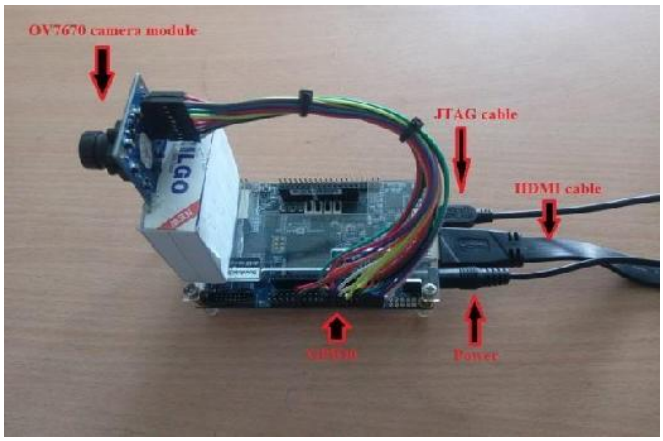

Figure 12. The proposed real-time face detection system.

3 Experimental Results

3.1 Hardware Setting

The proposed system hardware is shown as in Figure 12. The USB Blaster (JTAG cable) is utilized to connect the system to PC for programming. HDMI cable is utilized to connect the system to a monitor for display. The camera OV7670 is connected to the system by wires.

Table III
Experiment Scenarios

Distance	0.5 m	0.7 m	1 m	1.5 m
One person	X	N.A.	X	X
Two people	X	N.A.	X	X
Three people	X	N.A.	X	X
Four people	X	N.A.	X	X
Five people	N.A.	X	N.A.	X

Table IV
Experimental Results

Scenario	No. of detected faces	No. of undetected faces	Ratio
One person	1	0	1/1
Two people	2	0	2/2
Three people	2	1	2/3
Three people	3	0	3/3
Four people	4	0	4/4
Five people	4	1	4/5
Seven people	5	2	5/7
A man wearing glasses	1	0	1/1
Two people	2	0	2/2
wearing glasses	1	1	1/2
Three people	2	1	2/3
wearing glasses	1	2	1/3
Four people	2	2	2/4
wearing glasses	1	3	1/4
Person wearing mask	1	0	1/1
Person with a drawn face	1	0	1/1

Camera OV7670 resolution is scaled down to 320 × 240 pixels due to a large amount of data for face detection algorithms that need to be processed in order to specify detected face location in the video stream. Then a rectangle surrounding the face will be drawn.

3.2 Experimental Results

Many experiments are conducted in order to verify the system capability of detecting faces in a frame that contains up to 7 people. A face of a person, 2 faces among 3 people, and 4 faces among seven people in a frame are detected. The system can also perform more difficult tasks such as a man face wearing black glasses, mask, and drawn face. However, the system is not capable of detecting half of a face because there are not enough features of a face including eyebrows, eyes, nose and mouth. People taking part in doing experiments are tasked to appear in camera view with a distance range from 0.5 m to 1.5 m according to many scenarios as described in Table II. For example, there are 3 scenarios of camera distance of 0.5 m, 1 m and 1.5 m, respectively. "N.A." in the Table III stands for "Not Applicable" while "X" is stands for "chosen distance". There're 14 scenarios (selected by X) for experiment with the combination between number of people and distance. The experiments are done in order to find out the detection ratio of the system for conclusion and evaluation (the best case, the worst case, the best distance, the worst distance, and maximum people) of the system. Note that each scenario is tested 100 times.

Figure 13. One person scenario (0.5 m - 1 m - 1.5 m).

Figure 14. Two people scenario (0.5 m - 1 m - 1.5 m).

Figure 15. Three people scenario (0.5 m - 1 m - 1.5 m).

Figure 16. Four people scenario (0.5 m - 1 m - 1.5 m).

Figure 17. Five people scenario (0.7 m - 1.5 m).

Experimental results according to scenarios are listed in Table IV.

Several results are illustrated as in from Figure 13 to Figure 17.

Table V
Experimental Results

Scenario	No. Faces	Detected Faces	Detection Rate (%)	Min-Max Time (ms)
One-0.5m	100	96	89	105 - 286
One-1m	100	84	78	132 - 487
One-1.5m	100	51	43	241 - 568
Two-0.5m	200	171	81.5	234 - 887
Two-1m	200	157	72.5	165 - 879
Two-1.5m	200	86	38	360 - 637
Three-0.5m	300	193	58.3	228 - 451
Three-1m	300	165	48	171 - 638
Three-1.5m	300	102	26.7	206 - 428
Four-0.5m	400	185	39.5	134 - 443
Four-1m	400	169	35	179 - 379
Four-1.5m	400	114	20.5	280 - 336
Five-0.7m	500	216	36.6	178 - 349
Five-1.5m	500	173	27.4	187 - 581

Table VI
Comparison Between Hardware and Software in Terms of Detection Rate within 05.m Distance of Face and Camera

Tool	Detected Faces	Detection Rate (%)
Hardware on DE10-Nano	590	92
OpenCV on Intel 7 CPU	619	97

All obtained experimental results in terms of detection rate and speed are summarized in Table V.

According to Table V, the proposed system works on the optimal performance at the distance 0.5 m - 1 m and 1 - 2 people. Detection rate of the system is evaluated by each case. It is not likely to combine cases in order to calculate a common value because detection rate will be affected by the detection distance and the number of people. Maximum people that the system is capable of detecting is 5 with a low detection rate. In the case of one person, we tried testing to find out the maximum distance which the system is capable of detecting faces. That maximum distance is 1.8 m. Minimum time the system is able to detect faces is 105 ms. Maximum time the system is capable of detecting faces is 887 ms.

In addition, several experiments were done in order to evaluate the proposed system performance to other existing systems such as OpenCV and MATLAB Tools for face detection in terms of processing time (speed) and detection rate. A set of images (484 images containing 641 faces) were used to perform measurements. Table VI shows the comparison of the proposed hardware and software tool (OpenCV) in terms of detection rate. The results show that the software tool is more accurate than the proposed hardware.

Above set of images were repeatedly used for another experiment with the same purpose but in terms of speed. Table VII shows the comparison of the proposed hardware and software tool (MATLAB) in terms of processing time (speed). The results show that the proposed hardware is faster than the MATLAB tool because of the parallel architecture of the processor.

Table VII
Comparison Between Hardware and Software in Terms of Speed within 05.m Distance of Face and Camera

Tool	Image Nature	Time (ms)
Hardware	Image with single face	113.3
	Image with multiple face	134.0
Software	Image with single face	1734
	Image with multiple face	2289

Table VIII
Result Performed with 320×240 Resolution Images

Total Faces	Software	Hardware
1	1.256 ms (0.79 fps)	34.712 ms (28.80 fps)
6	1.402 ms (0.71 fps)	37.378 ms (26.75 fps)
11	1.538 ms (0.65 fps)	41.711 ms (23.97 fps)

Table IX
Result Performed with 640×480 Resolution Images

Total Faces	Software	Hardware
1	2.165 ms (0.46 fps)	133.143 ms (7.51 fps)
6	2.919 ms (0.34 fps)	146.745 ms (6.81 fps)
11	3.129 ms (0.31 fps)	152.664 ms (6.55 fps)

After doing all the experiments, it is concluded that the software face detection system is capable of processing the images at speeds of an average of 0.71 fps with 320×240 pixel images and 0.37 fps with 640×480 pixel images. Moreover, the hardware face detection system has the performance improvement up to 37.33 times faster than the software face detection system with 320×240 pixel images and up to 18.8 times faster than the software face detection system with 640×480 pixel image. Table VIII shows the performance with 320×240 resolution images while Table IX shows the performance with 640×480 resolution images.

4 Conclusion

In this paper, a real-time face detection and human tracking system that utilizes the fpga DE10-NANO controller board with Viola-Jones algorithm embedded is proposed. The project is carried out on Quartus tool for data transmission with 2 main functions: from camera to the board and from the board output to the screen. The proposed system performs detection rate up to 92% and the fastest processing time 113.3 ms. The proposed system is not as accurate as the software tool, but it is flexible, easy to install and use because there is only one controller board. Moreover, the proposed system gives results much faster than the software tool.

Acknowledgment

The authors would like to thank to students Phan Truong Khang and La Ngoc Le from the Department of Computer Engineering for implementing and testing the proposed system.

References

[1] Y. Wei, X. Bing, and C. Chareonsak, "FPGA implementation of adaboost algorithm for detection of face biometrics," in *Proceedings of the IEEE International Workshop on Biomedical Circuits and Systems*. IEEE, 2004, pp. S1–6.

[2] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, vol. 1. IEEE, 2001, pp. I–I.

[3] V. Nair, P.-O. Laprise, and J. J. Clark, "An FPGA-based people detection system," *EURASIP Journal on Advances in Signal Processing*, vol. 2005, no. 7, pp. 1–15, 2005.

[4] H. Ben Fakih, A. Elhossini, and B. Juurlink, "An efficient and flexible FPGA implementation of a face detection system," in *Proceedings of the ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*, 2015, p. 261.

[5] R. F. Rahmat, Y. Azzakirot, and T. Z. Lini, "Tree identification to calculate the amount of palm trees using Haar-Cascade classifier algorithm," in *Proceedings of the 3rd International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM)*. IEEE, 2019, pp. 36–39.

[6] S. Choudhury, S. P. Chattopadhyay, and T. K. Hazra, "Vehicle detection and counting using Haar feature-based classifier," in *Proceedings of the 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON)*. IEEE, 2017, pp. 106–109.

[7] A. AbdelRaouf and D. Salama, "Handwritten signature verification using Haar cascade classifier approach," in *Proceedings of the 13th International Conference on Computer Engineering and Systems (ICCES)*. IEEE, 2018, pp. 319–326.

[8] Y. Wang, L. Li, X. Wan, and J. Wang, "Woven fabric defect detection based on the cascade classifier," in *Proceedings of the 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. IEEE, 2019, pp. 1–5.

[9] Terasic Inc, "DE10-Nano user manual, 2003-2017," last accessed: 2019/10/27.

Huu Luong Nguyen received engineer from Hanoi University of Science and Technology, Vietnam in 1985 and received Ph.D. degree in Computer Modeling for Structural metal processes at The St. Petersburg Polytechnic University in 1993. In currently, Dr. Nguyen is an Lecturers and Researcher of Faculty of Computer Engineering at University of Information Technology – Vietnam National University at Ho Chi Minh City. His research is focussed on fields of Image analysis, AI for IoT.

Minh Son Nguyen received B.Engr. and M.Engr. in Computer Engineering from Ho Chi Minh City University of Technology, Vietnam in 2001 and 2005 respectively and received Ph.D. degree in Electrical Engineering at The University of Ulsan, Korea in 2010. In currently, Dr. Nguyen is Dean of Faculty of Computer Engineering at University of Information Technology – VietNam National University at Ho Chi Minh City. Dr. Nguyen is also Director of Automotive R&D LAB of UIT and Member of Commitee of Science and Technology of SaiGon High-Tech Park. His research is focussed on fields of Wireless Embedded Internet, AI for IoT, Smart System and System-on-Chip Design.

Tri Nhut Do received the B.Eng. degree in electrical electronics engineering and the M.Eng. degree in technology cybernetics engineering from the Ho Chi Minh City University of Technology, Vietnam, in 2002 and 2005, respectively, and the Ph.D. degree in electrical engineering from the University of Ulsan, South Korea, in 2012. From 2013 to 2014, he was a Post-Doctoral Fellow under the supervision of Prof. R. Phan with the Security Laboratory, Multimedia University, Malaysia. From 2015 to 2017, he was a Post-Doctoral Fellow under the supervision of Prof. U-X. Tan and Prof. C. Yuen with the Robotics Innovation Laboratory, Pillar of Engineering Product Development, Singapore University of Technology and Design, Singapore. From 2018 to 2020, he was a lecturer with the Faculty of Technology and Engineering, Thu Dau Mot University, Thu Dau Mot City Vietnam. Since 2021, he has been a lecturer with the Faculty of Computer Engineering, University of Information Technology, Vietnam National University in Ho Chi Minh City, Vietnam. His research interests include indoor localization, human daily activities tracking, robotics, sensor fusion, and human emotion recognition for security. He is a recipient of the Outstanding Paper Award from ICCAS-SICE 2009.