Method Article

An FPGA-based IP for recognizing violence against children

Muhammad Alhammami a,∗, Samir Marwan Hammami b

a Higher Institute for Applied Sciences and Technology, Damascus, Syria
b Dhofar University, Salalah, Oman

A B S T R A C T

We present in this paper an FPGA-based IP for recognizing most common violent actions against children (VACR IP). VACR IP uses only skeleton joints data as inputs to keep the privacy of people inside their homes or in the schools. The proposed hardware achieved a detection rate of 97.72%, processing speed 761FPS, and a latency value equals to 2.79msec using a 50MHz system clock.

In sum, this research method presents:

• First FPGA-based IP for recognizing most common child abuses without any privacy breach of people real images by using only skeleton joints data.
• The IP can detect the violence and the type of the violent action.
• The IP can be embedded in complete systems to be installed in schools by school psychologists and counselors.

© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

A R T I C L E   I N F O
Method name: An FPGA-based IP for Recognizing Violence Against Children
Keywords: FPGA-based hardware IP, Children abuse detection, Action recognition, Machine learning, Fog-based system
Article history: Received 8 November 2020; Accepted 29 April 2021; Available online 8 May 2021

∗ Co-submissions are papers that have been submitted alongside an original research paper accepted for publication by another Elsevier journal
∗ Corresponding author.
E-mail address: muhammad.alhammami@hiast.edu.sy (M. Alhammami).

https://doi.org/10.1016/j.mex.2021.101378
2215-0161/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Introduction

Detecting violence against children is a critical application in human action recognition field as it is distinguished by the necessity of low latency, real-time decision-making, and it should work even if there is no network available. Even more, it is preferable not to send data to a cloud so users can feel that they are keeping their privacy. Hence, using fog-based processing and computing concepts to implement the objective system is more practical than building a cloud-based solution. The VACR FPGA IP core is developed in this project to ease the design porting to FPGA based fog-based alerting systems that help families and authorities to rescue children under physical abuses as fast as possible. The IP core takes the skeletal joints data of the adult and the child and produces two outputs. The first output is a logical output indicates whether the child is being physically abused or not. The second output is related to the type of action being done by the adult toward the child. Violent actions include kicking, punching, throwing, shoving, strangling, and slapping. Non-violent actions include touching, hugging, lifting, laying down, approaching, departing. There is no previous FPGA-based IP for detecting violence against children [3].

Method details

To build VACR hardware IP, the methodology customizes an existing machine-learning based model for detecting violence against children in [1,2] by reducing the number of required features to fit the structure of resources of the used FPGA device. It also uses the MMU VAAC dataset [4] developed by the Digital Home Research Centre at Multimedia University.

The main blocks of Violence Against Children Recognition (VACR) IP are shown in Fig. 1:

1. Receiving the incoming skeleton joints data serially and shaping it in parallel frames: using a counter and a registered demultiplexer.
2. Features computation: This features computation block implements the needed functions to calculate the used features in [4] applied on the new data. Each RTL block of 2D Euclidean distance consists of three adders, two multipliers, and one square root computer. This structure needs six clock cycles to finish the computation, so a pipeline of 6 steps is introduced inside the structure.
3. 1-NN classifier: By scaling up the previous structure, it is possible to compute any k-dimensional Euclidean distance and will be used in the classification block.
4. ROM memory which contains the training data: Originally, the ML-based model in [2] used 25 features. However, the configuration of the memory of the on-chip ROM in the used FPGA device does not allow to store all of the 25 features easily. Hence, only the top-ranked 20 features available in [2] are stored in the ROM and will be used in the rest of the implementation of VACR IP.
5. Synchronizing and controlling blocks.
6. FIFO blocks to store intermediate values wherever there are slower functions after faster functions.

The input data of the block contain information about positions of 20 joints belong to one adult and one child represented as integers using 16bits. Hence, the actual input bitrate is given by Eq. (1)
Fig 1. The block diagram of VACR IP.

Fig 2. An example of the integration of the FPGA-based VACR IP in a real-time system.

where the capturing frame rate of the Kinect sensor is 30fps and the joints are in 2D space according to [1,2].

20 joints × 2coordinates × 16bits × 30frames = 19.2 Kbit/sec

The utilized FPGA resources for this IP are 14% of logic elements and 60% of memory bit based on Cyclone IV GX – EP4CGX150DF31C7 device. The latency of the IP equals to 139500 cycles, and next for each 65703 cycles there are new values on outputs (Output ready signal is also available). The skeleton data is fed to the VACR FPGA IP block serially. The IP firstly changes this data to parallel form, then it calculates the necessary Euclidean distances-based features which will be used for the classification. The IP outputs two types: “isViolent” and “Action Name”. “isViolent” output indicates whether this action is violent or not. “Action Name” output is raising one of 12 signals which are correspondence to the 12 action names including kicking, punching, throwing, shoving, strangling and slapping, touching, hugging, carrying, laying down, departing, and approaching [4].
Fig 3. Confusion Matrix, True Positive rate and False Negative rate of action name recognition based on the outputs of the FPGA-based VACR IP.

The characterization of the IP was done using SignalTap tool in the Intel Quartus software:

- Detection rate of violence: 97.72%
- Recognition rate of action: 95.41%
- High Performance: 761FPS at 50MHz clock signal while it achieved only 9fps on a desktop computer with Intel i7 processor at 2.67GHz and 8GB RAM using C++ based software.
- Latency is 2.79msec at 50MHz clock signal.

To run and test the VACR IP in real-time, Fig. 2 shows an example of the integration of the FPGA-based VACR IP in a real-time system which consists of [5]:

- Microsoft© Kinect device for data capturing.
- Intel© Atom based system for configuring, controlling, and interfacing the Kinect sensor to the FPGA device by using the Microsoft SDK library provided on a Microsoft Windows operating system.
- FPGA-based SoC device from Altera© as accelerated processing elements.
- PCI-e link to transfer data from Kinect device to FPGA element.

This research work used a child mannequin for testing to not let any child face any kind of violence. Figs. 3 and 4 shows the confusion matrix of the both resulted outputs, i.e. violent/non-violent and the type of action (i.e. name of action).
Fig 4. Confusion Matrix, True Positive rate and False Negative rate of Violence Recognition based on the outputs of the FPGA-based VACR IP.

Conclusion

The promising results of experimenting VACR IP motivates us to develop a fog-based system which can accept many Kinect sensors which are near to each other’s, like in a school or a hospital, by using the ability to replicate VACR IP according to the number of the necessary Kinect sensors and working in parallel with only one FPGA device.

Declaration of Competing Interest

The authors of this paper certify that they have NO affiliations with or involvement in any organization or entity with any financial interest, or non-financial in the subject matter or materials discussed in this manuscript.

References

[1] M. Alhammami, S.M. Hammami, C.-P. Ooi, W.-H. Tan, Optimised ml-based system model for adult-child actions recognition, KSII Trans. Internet Inf. Syst. 13 (2) (2019) 929–944.
[2] S.M. Hammami, M. Alhammami, Vision-based system model for detecting violence against children, MethodsX 7 (2020) 100744.
[3] I. Sittón-Candanedo, R.S. Alonso, J.M. Corchado, S. Rodríguez-González, R. Casado-Vara, A review of edge computing reference architectures and a new global edge proposal, Future Gen. Comput. Syst. 99 (2019) 278–294.
[4] M. Alhammami, C.-P. Ooi, W.-H. Tan, Violent actions against children, Data in Brief 12 (2017) 480–484.
[5] M. Alhammami, O.C. Pun, T.W. Haw, Hardware/software co-design for accelerating human action recognition, in: 2015 IEEE Conference on Sustainable Utilization And Development In Engineering and Technology (CSUDET), IEEE, 2015, pp. 1–5.