FPGA hardware architecture with parallel data processing to detect moving objects using the background image subtraction technique

Ingrid Nunes Rodrigues, Charles Luiz Silva de Melo, Vanessa Mello da Frota Botinelly and Jozias Parente de Oliveira.

***Abstract*— Moving objects can be recognized through detection systems that use computer vision as it passes in front of a digital camera. However, a major challenge is to achieve recognition in real time, and to aid in visual analysis, a hardware architecture in FPGA (Field Programmable Gate Arrays) could be used for video processing focusing on the detection of moving objects using the background subtraction technique. And for that, some processes are critical as the digital image processing, video segmentation and real-time video processing. This paper's proposal is the study of Digital Image Processing, Background Subtraction techniques and FPGA to develop an architecture for video processing to detect moving objects using the Background Subtraction technique. First it is discussed the basics of Digital Image Processing, Background Subtraction and FPGA, covering its main characteristics. Then it is explained the methodology applied in the development work. Is then described in detail the implementation process, and the end shows the final results and main conclusions for the project.**

***Index Terms*—— Digital Image Processing, background subtraction, FPGA.**

I. Introduction

DUE to its wide range of use, digital image processing is very important because, comprises a set of hardware, software and theoretical foundations, all intertwined to obtain the information of interest observed from a scene where one of the digital processing steps is segmentation.

Video segmentation is a fundamental process where the image is subdivided into parts or constituent objects, up to the level needed for each use. That is, targeting occurs until the objects of interest have been isolated from the image forefront of belonging to the reference image. Thus, it is clear that this process determines success or failure in the analysis [1].

In digital video processing, real-time is defined as the interval time required for the system to process each pixel between two successive samples of each above. But one should consider the critical factors, as well as process each pixel has that there be transfer of each frame and perform other related operations data transfer to memory [2].

In video processing area, one of the fastest growing applications are among those that perform video segmentation for object detection. However, this becomes a challenging task when it has to be developed in real time, as it requires the processing of a significant amount of data and instructions. For example, given an image with the same resolution 640 x 480, 8-bit quantization, 30 frames per second and three color components (RGB) would need to process 26.36 Mbytes per second. When the processing is not performed in real time, the image may appear traces in accordance with the movement of people in the image, or even low resolution due to the reduction of the amount of processed pixels.

The hypothesis established here, is that it would be possible to implement a hardware architecture in FPGA with parallel data processing to detect moving objects in real time using a technique called subtraction of the background image. Where the capture of images is held by a video camera with a resolution equal to 640x480 pixels and a rate of 30 frames per second connected to an A/D converter whose digitized video data is processed in real time on the FPGA.

The implementation of an algorithm for detecting moving objects can contribute and have applications in several areas, such as entertainment, electronic surveillance, security systems, military applications, applications in industrial automation, trajectory planning, among others. The development of this research can contribute to improvements in the final image resolution quality processed since the video processing is done in real time.

II. Digital Image Processing

The term "image" refers to the two-dimensional light intensity function f (x, y), x and y are the spatial coordinates and the value of f at any point (x, y) is proportional to the brightness (or gray levels) of the image at that point, where x and y correspond to the spatial coordinates [1]. The gray level of a point of the image will depend on the amount of light falling on the scene. The function f (x, y) is characterized by

Ingrid Nunes Rodrigues is academic of Electrical Engineering at University of State of Amazonas and is formed in the technical course in Mechatronics from the Senai Antonio Simões School, Manaus, AM, Brazil (e-mail: inr.nunes@gmail.com).

Charles Luiz Silva de Melo holds a degree in Electrical Engineering from the Federal University of Amazonas and master's degree in electrical engineering from the Federal University of Amazonas. He is currently a professor at the Amazonas State University, Manaus, AM, Brazil (e-mail: charles.melo@yahoo.com).

Vanessa Mello da Frota Botinelly holds a degree in Electrical Engineering from the University of State of Amazonas, Manaus, AM, Brazil.

Jozias Parente de Oliveira holds a doctorate in Electrical Engineering from the Federal University of Pará and a masters in Electrical Engineering from the Federal University of Campina Grande, is currently Coordinator of Electronic Engineering courses and Technology in Industrial Automation and Professor / Researcher at University of State of Amazonas, Manaus, AM, Brazil (e-mail: joziasparente@gmail.com).

978-1-4673-8756-9/15/$31.00 ©2015 IEEE

the amount of illumination falling on the scene and the amount of light that reflects from the scene through the object. These components are called illumination and reflectance and are designated by i(x, y) and r(x, y), respectively. The product of the two functions form f (x, y) as in equation (1) [3].

$$f(x, y) = i(x, y). r(x,y) \quad (1)$$

$$0 < i(x, y) < \infty$$

$$0 < r(x, y) < 1$$

A. Color model RGB and YCbCr

The RGB model or "true color" is a color model designed based on the graphics output devices with three primary colors: red, green and blue. The abbreviation derives from the English Red, Green and Blue. And each pixel has a particular color, which is described by the amount of red, green, and blue contained in color [3].

But the YCbCr color model was created because of a RGB failure, whereas the latter did not allow the compatibility between emissions of television systems in color and black and white receivers. This model is based on the separation of the RGB color signals into a luminance signal and two chroma signals or color difference.

B. Sampling and quantization

Sampling and quantization are scanning processes both spatially and in amplitude of a particular image.

Basically, we represent an image in computer, through an M by N matrix, where each component f of it represents a picture element or pixel.

Therefore, for the scanning process it is necessary to determine certain parameters, such as the values for M, N and the number of intensity levels. There are no restrictions placed to M and N other than that of being positive integers. However, due storage considerations and quantization hardware, the number of intensity levels is typically a power of 2. That way, as larger the parameters of M, N and G, bigger and better will be the image resolution. However, the higher the resolution is, the greater must be the space for processing and image storage [4].

C. Segmentation

The video segmentation is key to the detection and recognition of objects in a scene, whose function is to subdivide the image into its constituent parts. This function must be carried out until they have been isolated the application of interest objects. The object of interest, in turn, refers to the group of pixels that have a common feature in the image.

An example of object segmentation system can be represented by the subtraction technique of background image, where the first few frames are captured for construction of the reference model with no object to be detected, then a video input is used to capture new images, called the current frame. In the classification stage, the object is detected by subtracting each pixel of the current frame for each pixel of the reference frame or background. Finally, the resulting subtraction is segmented image to highlight regions of the image motion.

There are various segmentation techniques, none of which is able to target all types of images. An image gray levels, for example, can be segmented in two ways: either consider the similarities or differences between the gray levels. There are two most commonly used image segmentation techniques, which are Gaussian statistical method and the difference between adjacent frames.

In Gaussian statistical method, the reference image model is constructed pixel by pixel. The value of each pixel is the result of an independent process. The pixel values of the reference image are modeled as a Gaussian distribution, characterized by its mean value and its standard deviation, in the YCbCr color space. In the segmentation phase, for each frame processed, the observed value for each pixel is compared with its corresponding distribution in the reference image. The intensity information (Y) and chrominance (Cb, Cr) in YCbCr space will decide whether the pixel belongs or not to the reference image. Thus, for each pixel belonging to the background image it is necessary that the values of two components (Y and Cb) or (Y, Cr) are within the standard deviation of the values of the same. Each pixel classified as object is grouped, resulting in a list of regions representing the foreground image objects.

The method of difference between adjacent frames has various applications in detecting objects. It is based on the difference between the current image with the previous one, where each current image is subtracted from the previous image sequentially, resulting in a third image which is compared with a predetermined reference threshold to enable identification of local change in the scene. The subtraction with larger absolute values than the threshold is classified as foreground image.

One way to extract the background objects is by selecting a T threshold that separates the image into two groups. However, the threshold values are extremely critical , since these values are not appropriate occur difficulty in the isolation of regions of interest so low values cause detecting artifacts in the image, while high values suppress significant changes in the image[5].

A histogram is a diagram representing the frequency in which they occur brightness in an image.

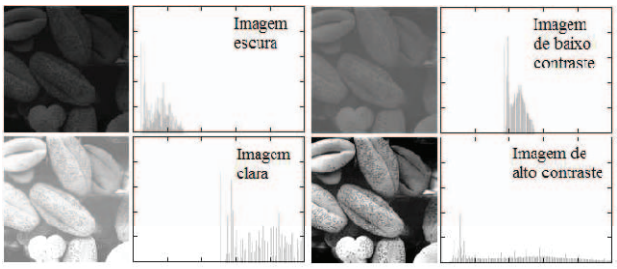

Fig. 1.Histograms corresponding to four basic types of images.

Histograms are the basis for numerous spatial domain processing techniques. And the histogram manipulation can be used effectively to improve image, as shown in Fig. 1. It is noted that the darker the image histogram components are

concentrated on the low side (dark) gray scale. Likewise, the histogram of the image light components are inclined upward side of the gray scale. An image with low contrast has a histogram is narrow and be centered toward the middle of grayscale. And the histogram components in high-contrast image spread in a wide range of shades of gray, and also that the distribution of pixels is not too far from uniform, with few vertical lines being much larger than the others [1].

III. FPGA (FIELD PROGRAMMABLE GATE ARRAYS)

The FPGA is a leading logic devices, its main features are being reprogrammable and use SRAM technology or the antifuse programmable connections, as well as being flexible and adaptable to the needs of each application.

It can be set by software, and is described as a highly condensed arrangement of small blocks of circuitry, consisting of some logic gates and flip-flops, with some interface signals. So it has the advantage that in a single chip can be embedded complex systems with low, high and medium levels of algorithms [2].

The fundamental structure of a FPGA is composed of three basic elements: interconnections, configurable logic block (CLB) and input / output blocks (I / O).

The CLB (Configuration Logical Blocks) is the configurable logic block, built by the meeting of flip-flops and the use of combinational logic. The IOB (Input and Output Blocks) is the block inputs, outputs or bidirectional access to the outside world the device, selectable individually, which are responsible for interfacing the outputs from the outputs of the CLBs combinations. Distributed array of programmable interconnects (or routing matrix or programmable keys or interconnections keys) provides the interconnection between the configurable logic blocks and the connection to the inputs and outputs. The largest FPGAs can have tens of thousands of CLB's in addition to memory and other resources [6].

IV. HARDWARE DESCRIPTION LANGUAGE – VERILOG

There are several hardware description languages and the main feature of which is to describe the function of a piece and independent of the application form of hardware. The most common are: ABEL, AHDL, VHDL and Verilog.

The Verilog is currently one of the leading hardware languages (Hardware Description Languages - HDL). This language was created by Gateway Design Automation in 1985 and later developed by Cadence Design Systems, is specified in the IEEE standard 1364, 1995. The Verilog abstraction levels are divided into Behavioral, Register Transfer Level (RTL), Port Logic and Layout. Each defines how much detail on the design there is a particular description [7].

The layout represents the description that has greater detail, may be chosen any temporal information.

Logic Gate represents the information about the layout that is lost, but there is detailed information about the time, the implemented function and architecture. This module is implemented in terms of logic gates and interconnections between these ports.

The RTL represents the temporal information that is restricted to clock cycle, records and logic among them is set, existing information about the architecture, but not about the technology.

The Behavioral represents the highest level of abstraction of Verilog. A module can be implemented in terms of design, without worrying about the hardware implementation details. Your kind of description is very similar to the C programming. There is only information in terms of function, the information relating to the architecture or the records do not exist and there may be temporal information.

V. RESULTS AND DISCUSSION

After being conducted literature searches in digital processing picture area, object detection and programmable logic devices, we selected two main areas of study: Video segmentation and reprogrammable logic devices. These technologies are the focus of this work, and as seen, they are the theoretical basis of this framework.

Then, after an analysis of programming platforms, computational software MATLAB and Quartus II were selected for simulation and application testing throughout the project design. The reason for such choices are the MATLAB has a structured language, that is, has high-level programming, facilitating the handling of data and objects, and offering numerous support tools and learning accessible and organized, and the Quartus II has the essential design to project various types of hardware language, is compatible with all household devices Altera® FPGA and CPLD, and offers numerous support tools and learning accessible and organized. In addition, it is widely used and accepted in scientific works of digital processing picture area and programmable logic devices.

After the search period of the algorithm, the first activities using MATLAB software were made, making the algorithm validation to the method of Floating Point and Fixed Point method using, for both, the same images from the database, resulting in two different approaches. Also occurred the hardware architecture definition required for implementation of the proposed method, and subsequent implementation and validation in the Quartus II software and testing hardware (FPGA).

The algorithm consists of image acquisition, selection of the threshold value of each color component of the image, separating the color components Y, Cb and Cr, statistical calculation using the image gray levels, calculating the difference between the average image and the current image in each color component Y, Cb and Cr making the pixel by pixel difference for each. The classifying picture elements as belonging to the object or the background image are done by comparison with the predetermined threshold. Finally, it has been the replacement of picture elements for their corresponding values from the test object image for each color component, but when the pixels belong to the background color does not happen replacement.

A. *Floating Point results*

Are presented below, the results obtained with the proposed method implemented in MATLAB with operations in floating point. First, the proposed method is evaluated in relation to the variation of thresholds.

It can be seen in Fig. 2, the original image and the result of subtracting the background image. In the first analysis it is observed that many pixels in the reference image are classified as the object, since the threshold values are too low, while in the third analysis several pixels belonging to the object were classified as the reference image when the values of thresholds are too high. And in the second analysis, there is a better result, which does not occur so detection of image pixels being classified as reference object, that is, there was less detection of false positives.

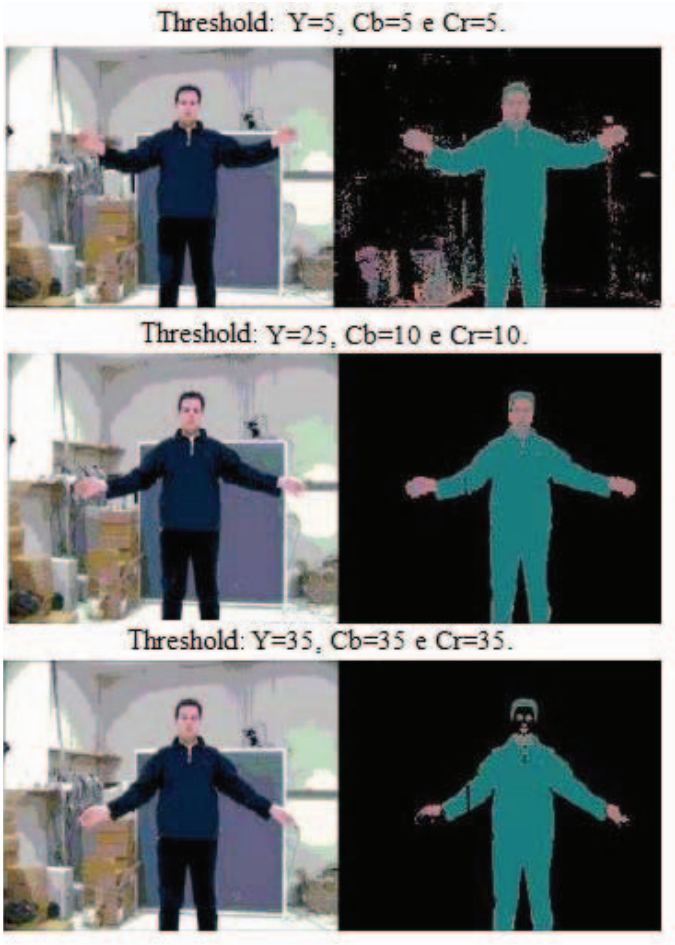

Fig. 2. Input image and the resulting image in floating point.

To understand how the difference of the thresholds acts on each component of YCbCr color space, it is noted that in Fig. 3 in the first frame of the four images is represented Y component representing luminance information, and the following frames are represented by components Cb and Cr that represent the blue and red colors, respectively, and the fourth image is the resulting image of the three components together. One blank line must be left before and after each figure.

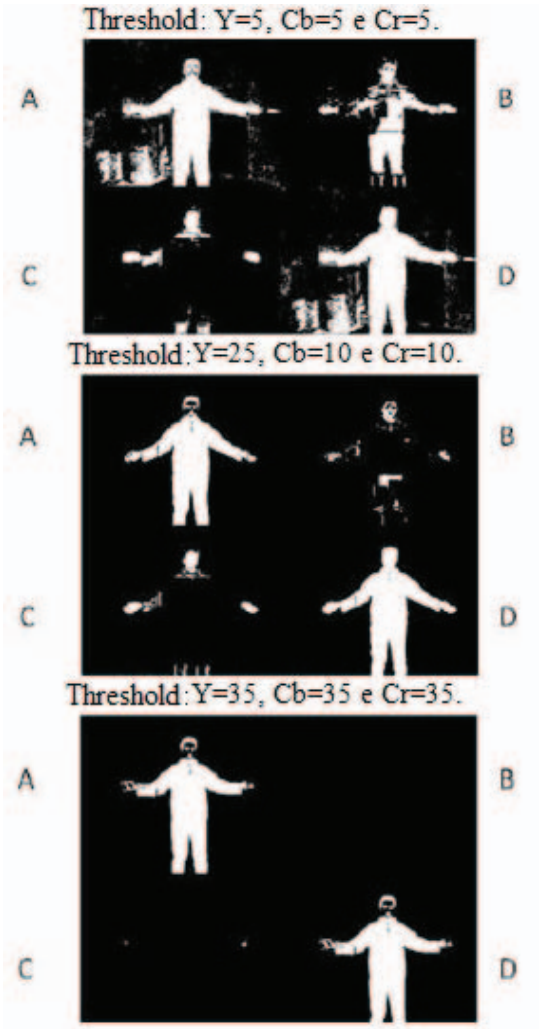

Fig. 3. Resulting images in floating point where a) Y component, b) Cb component, c) Cr component d) YCbCr component.

B. *Fixed point results*

In Fig. 4 it can be seen that several pixels in the reference image are classified as object when the threshold values are too low, while in the third analysis several pixels belonging to the object are classified as the reference image when the threshold values are too high. And in the second analysis, there was a better outcome quantitatively where the object is detected and does not occur much identifying pixels of the reference image being classified as the object, i.e., is not observed a lot of false positives.

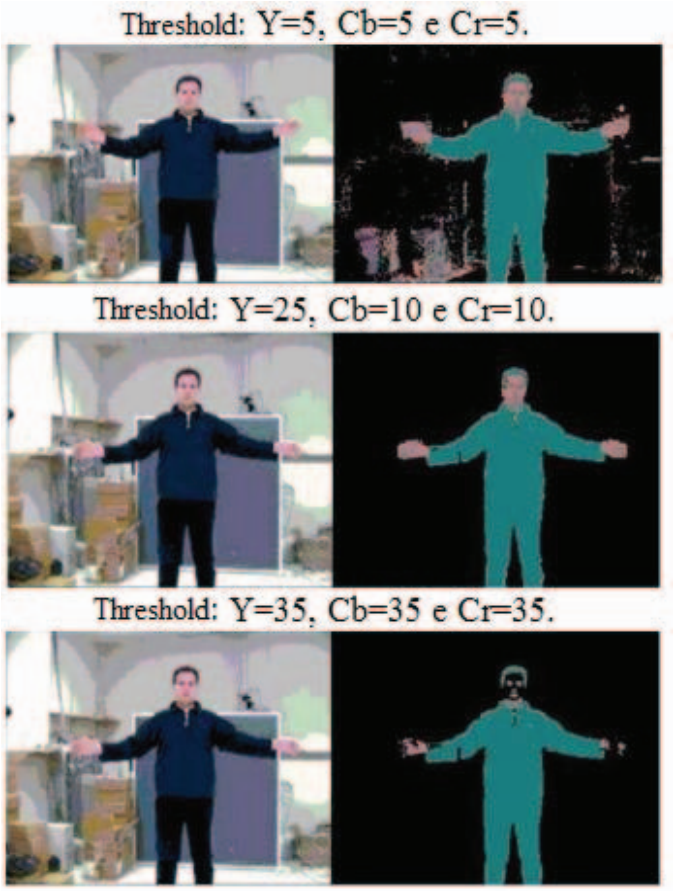

Fig. 4. Input image and the resulting image in Fixed Point with different thresholds.

It is observed in Fig. 5 in the first frame of the four images is represented Y component representing luminance information, and the following frames are represented by Cb and Cr components that represent the blue and red colors, respectively, and fourth image is the resulting image of the three components together.

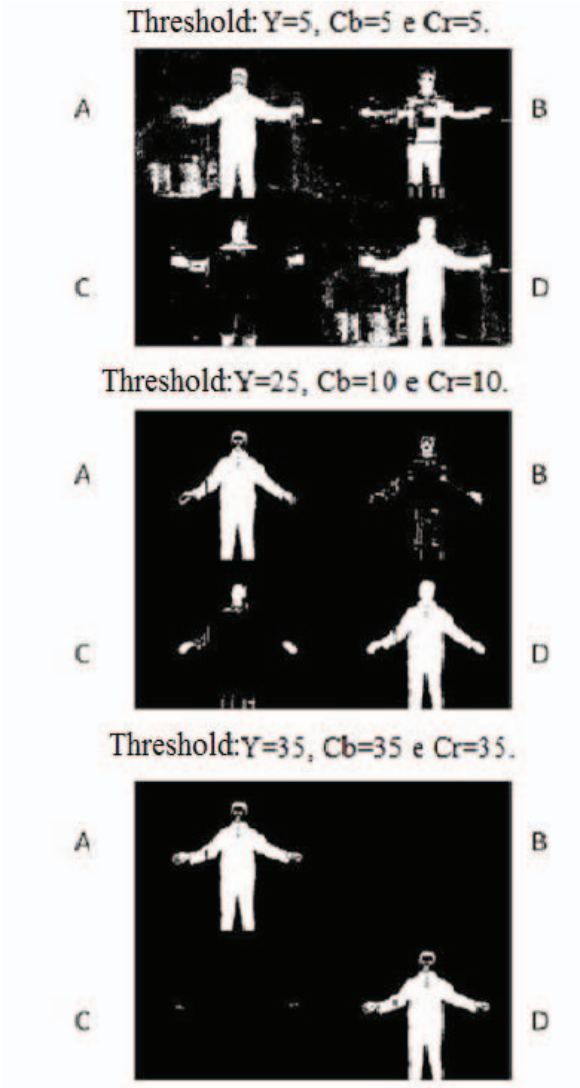

Fig. 5. Resulting images in fixed point with different thresholds where a) Y component, b) Cb component, c) Cr component d) YCbCr component.

Thus, the analysis made on the two results there is no significant difference between the measurements, since the difference between them is in the millimeter scale values, and therefore do not reflect a true difference in the population.

C. Hardware results

In this chapter we present the experimental results of the project proposal. The architecture has been implemented in an educational kit and development of Altera called DE-2. This kit has an FPGA, of the Cyclone II (EP2C35F672C6) family, with 33,000 logic elements. The proposed architecture was modeled with the hardware description language Verilog. The simulation and synthesis tool used was the Quartus II Altera. The implemented architecture can process 30 frames per second and an image with a resolution equal to 640 x 480 pixels.

The result of the segmentation is illustrated in Fig. 6, where occurs the capture of the current images whose pixels are classified as belonging or not the reference image.

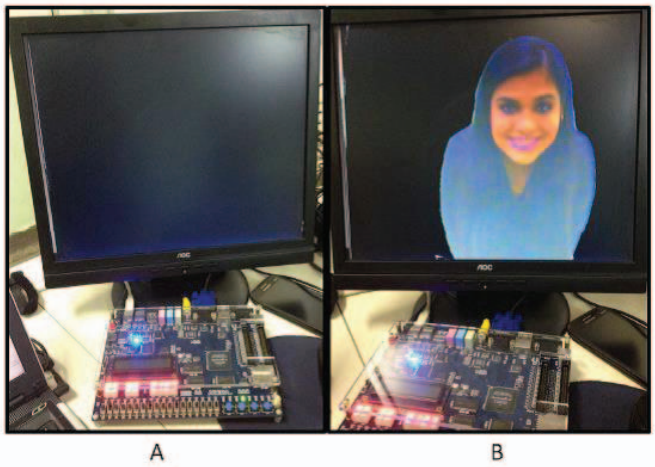

Fig. 6. Segmentation Results: A) no object in scenario B) with object in the scene.

VI. Conclusion

In view of what has been discussed, it was observed that the results of evaluation of the method for detecting objects implemented in MATLAB with operations in floating point and fixed point were key to evaluate the parameters of the proposed methods, and thus performance comparisons in different threshold values.

The object detection method implemented in the first stage of this work was evaluated using three types of thresholds. The first low threshold values used for each component resulted in the detection of many false positives. The second uses the balanced threshold values for each component and resulted in the detection of many true positives. The third threshold uses high values for each component and resulted in the detection of many false negatives.

The experimental results in hardware were obtained in the FPGA testing kit, since the simulations in the Quartus II do not allow observing a video signal or an image in the simulation output.

The method developed during this research achieved better results with reference images without objects. Why should improve detection for indoor environments with objects in the scene.

Given the above, it can be stated that the objective was achieved, that is, you can implement a hardware architecture in FPGA with parallel data processing to detect real-time moving objects using image a technique called subtraction background.

References

[1] GONZALEZ, R. e WOODS, R. Processamento de imagens digitais. Wilmington, USA : s.n.,1996.

[2] FERREIRA, Camilo Sánchez. Implementação do algoritmo de subtração de fundo para detecção de objetos em movimento, usando sistemas reconfiguráveis. Brasilia:[s.n.], 2012.

[3] FERNANDES COSTA, M. G. Programa de Mestrado em Engenharia Elétrica: Processamento Digital de Imagens, Universidade Federal do Amazonas, Manaus : s.n., 2008.

[4] MARQUES FILHO, Ogê e VIEIRA NETO, Hugo. Processamento Digital de Imagens. Rio de Janeiro : Brasport, 1999.

[5] OLIVEIRA, Jozias. Método para extração de objetos de uma imagem de referência estatica com estimativa das variações de iluminação.2008. Proposta de Tese (Doutorado em Engenharia Elétrica), Universidade Federal do Pará, Belém,2008.

[6] FREIRE, T. S. S. Interfaceamento de entrada e saída em aplicações com uso de FPGA. 2010. Monografia apresentada como pré-requisito para conclusão do Curso de Engenharia Elétrica, Centro de Tecnologia, Universidade Federal do Ceará, Fortaleza, 2003.

[7] PALNITKAR, Samir Verilog HDL - A guide to Digital Design and Synthesis. [s.l.] : SunSoft Press, 1996.