Microprocessors and Microsystems 39 (2015) 494–503

Contents lists available at ScienceDirect

Microprocessors and Microsystems

journal homepage: www.elsevier.com/locate/micpro

# Analysis of the efficiency of the census transform algorithm implemented on FPGA

C.-A. Tavera-Vaca, D.-L. Almanza-Ojeda, M.-A. Ibarra-Manzano *

*Department of Electronics Engineering, DICIS-Universidad de Guanajuato, Salamanca, Guanajuato 36885, Mexico*

A R T I C L E   I N F O

*Article history:*
Available online 10 August 2015

*Keywords:*
Stereo vision
Field programmable gate arrays
Census transform
Hamming distance
Tanimoto distance
Dixon Koehler distance

A B S T R A C T

Over the course of the last two decades, continuous advances in the stereo vision field have been documented. In this paper we present an analysis of the efficiency for the stereo vision algorithm of the Census Transform algorithm. In addition to the conventional correlation method based on Hamming distance minimization, we use two similarity measures: the Tanimoto and the Dixon-Koehler distances. Then, we compare its performance in terms of accuracy and hardware resources needed for implementation. These comparisons are performed by introducing a generalized model for each hardware architecture, scalable depending on design parameters such as Census Transform window size and maximum disparity range.



## 1. Introduction

Stereo vision is a computational method used to recover the three-dimensional information from an image, which is lost in the process of capture. The method makes use of (at least) two images of the same scene, by convenience referred to as left and right images. The images are slightly separated at a certain distance, known as baseline. By correlating the pixels on one image with the pixels on the other, that is, finding the distance between a reference pixel and a target pixel on the other image, known as disparity, it is possible to compute the depth information of the objects in the scene.

In the last two decades the field of computational stereo has been object of an extensive study and popularity given the many applications it has in robotics [1], obstacle detection [2,3], autonomous vehicles [4,5], 3D model acquisition [6] and surveillance [7]. Currently, there are many approaches to solve the stereo correspondence problem, and the mainstream algorithms and methods can be reviewed in [8–10].

However, one of the main limiting restrictions to stereo methods is the significant computational resources needed to calculate the stereo correspondence between the pixels [11]. In sequential architectures, big computations imply big computation times. This restriction is the main reason why real-time applications are challenging to accomplish on software (PC) architectures.

Reconfigurable architectures have been proposed as a solution for real-time applications due to the concurrent processing inherent to hardware-based processors. These approaches involve the use of digital signal processors (DSP), application-specific integrated circuits (ASIC), field programmable gate arrays (FPGA) and some graphics processing units (GPU) architectures. Due to the rapid development of programmable logic devices, FPGA devices have been widely used in a range of applications. They offer the best compromise between the processing speed of the ASIC and the flexibility of reprogramming the device to custom fit a design for a specific application [12]. In [13], it is shown that FPGA has a slightly superior performance than GPU architectures for certain configurations. Nevertheless, the amount of logic resources for any device is limited and usually accurate methods require more resources. Because of this, finding a compromise between the resources available and the accuracy desired is needed.

Due to the concurrent processing nature of the FPGA, some algorithms can be implemented more efficiently than others when designed for these devices. Local or area-based algorithms are usually the ones that adapt best to FPGA. Among these, the architectures based on Census Transform [14] are particularly compatible with FPGA due to the logic nature of the operators, which define the transform, as well as the parallel computations required by the method.

In this work we present an analysis of the accuracy of disparity maps, computed on FPGA device by using a standard Census Transform stereo architecture. This analysis is a function of the size of the transform window and the equivalent quantity of hardware resources needed for the implementation. In order to carry out this

* Corresponding author. Tel.: +52 4646479940.
*E-mail addresses:* ca.taveravaca@ugto.mx (C.-A. Tavera-Vaca), luzdora@ieee.org (D.-L. Almanza-Ojeda), ibarram@ugto.mx (M.-A. Ibarra-Manzano).

http://dx.doi.org/10.1016/j.micpro.2015.08.002
0141-9331/

analysis, we programmed a parameterized core able to generate architectures in Very-High-Speed Hardware Description Language (VHDL), based on a set of design parameters such as Census Transform window size, image dimensions, maximum disparity range and size in bits of the pixel word. In addition, we present three different similarity measures between sets or bit vectors: Hamming, Tanimoto and Dixon-Koehler distances. Finally, we propose a set of mathematical functions that try to estimate the amount of logic resources needed to implement an architecture based on the design parameters. This work seeks to provide designers with the tools to obtain, beforehand, an estimate of how many hardware resources are needed to implement an architecture similar to the one described in this work and an expected accuracy performance.

This work has been organized in the following manner. In Section 2 we present the works related to stereo vision, Census Transform, and similarity measures. Section 3 describes the hardware architectures and the generalized models for the Census Transform and the correlation modules. In Section 4, we present the results for accuracy of the methods and their costs in terms of hardware. Finally, in Section 5 we present our conclusions.

## 2. Related works

For the last two decades, stereo vision has been intensively studied and many algorithms and methods have been developed, optimized and tested. In the references [8–10] the current mainstream approaches for stereo vision are described. Among these approaches, the algorithms based on the Census Transform [14] have been particularly popular for hardware implementations, namely on FPGA devices.

Census Transform-based implementations have been well adapted particularly for obstacle detection and mobile robotics applications as shown in [15]. In the cited work an implementation of an optimized architecture based on Census Transform is presented, as well as a performance comparison between three different devices: FPGA, DSP and PC for three different Census window sizes. The FPGA implementation results are shown to be considerable superior in terms of the processing time required, able to generate 130 disparity maps of $640 \times 480$ pixels. In [16] a similar architecture is designed using high-level synthesis software (GAUT), and a comparison among other works in terms of the resources needed for implementation is described. The architectures presented in [15,16] are then applied to an obstacle detection application on a mobile robot [1].

The work presented in [17] describes a parameterized stereo vision core designed to generate architectures based on Census Transform. The authors conclude that the hardware resources usage and the maximum clock frequency supported by the device depend on of the transform window size and the maximum disparity range. Whereas the focus taken on this work is similar to ours, their conclusions only show in a graphical form the hardware percent usage and no information about the estimated accuracy of the maps is provided.

In [18] a full stereo vision system is presented. The system includes all the blocks needed to carry out the stereo correspondence process: rectification, stereo matching and post-processing using a FPGA. In that work the authors use the Census Transform algorithm for the stereo matching block, with an $11 \times 11$ window size.

In [19] an ample survey on several area-based algorithms and their hardware cost is presented. They present a comparative study of the Sum of Absolute Differences (SAD), Sum of Squared Differences (SSD), Normalized Cross-Correlation (NCC), Census and E-Census algorithms for a fixed maximum disparity and different window sizes. The respective correct match percentage of the disparity maps is obtained too. The architectures used in that work were generated by the Xilinx System Generator and the Resource Estimator Block included in the software.

For our study, our architecture generator core has been programmed using the software MATLAB to generate architectures in VHDL. MATLAB is only used as an interface to input the different parameters whereas the cores generates the architectures in VHDL in the same manner a person would. All the generating functions have been programmed as "text files generating functions" that contain the hardware descriptions. The purpose of this approach is to have a degree of customization that is not possible to achieve in high-level description generators such as GAUT or Xilinx System Generator. The interest in a fully customizable architecture is to analyze each section and, if deemed necessary, to custom a specific section for a particular application. Our work focuses only on the stereo matching process; therefore, rectification and post-processing are not considered.

### 2.1. Overview of passive stereo vision

The general algorithm for stereo vision systems based on Census Transform is shown in Fig. 1. The left and right images are processed independently. In order to decrease the complexity of the algorithm and the number of operations in the matching process, our approach assumes that the epipolar restriction is used. In this restriction, the main axes of the cameras are aligned in parallel. So that, the epipolar lines between the two cameras correspond to the displacement of the position between one point on the left image with respect to the right one. If any pair of pixels is visible in both cameras and assuming they are the projection of a single point in the scene, then both pixels must be aligned on the same epipolar line. In order to minimize the effect caused by any minimal misalignment in the camera axes, a rectification process is carried out. Once the images have been rectified and distortion corrected, the stereo matching process is initiated.

The non-parametric transforms Rank and Census were proposed by Zabith and Woodfill [14] as an alternative approach for the stereo correspondence problem algorithms based on statistical methods. The non-parametric transform, unlike other methods such as SAD, SSD, or NCC, does not operate over the pixel values of the image, but quantify the way a pixel is related to its

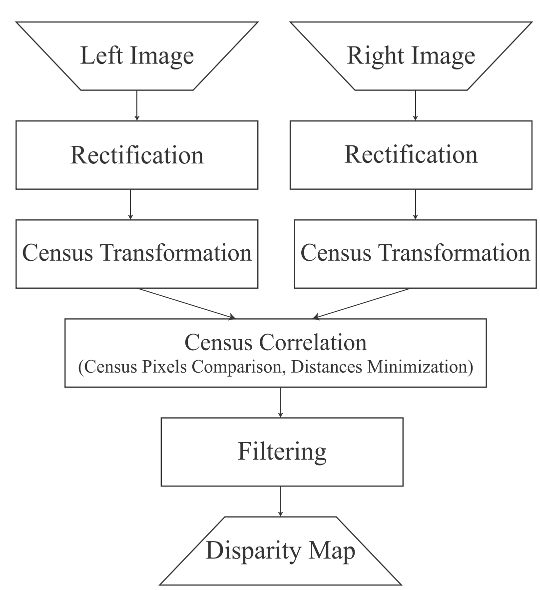


**Fig. 1.** Stereo vision algorithm.