INTEGRATION, the VLSI journal 54 (2016) 10–23

Contents lists available at ScienceDirect

INTEGRATION, the VLSI journal

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

Frame buffer-less stream processor for accurate real-time interest point detection

Gian Domenico Licciardo [a,*], Thomas Boesch [b], Danilo Pau [c], Luigi Di Benedetto [a]

[a] Department of Industrial Engineering (D.I.In.), University of Salerno, Fisciano, Salerno, Italy
[b] Advanced System Technology-STMicroelectronics International N.V., Plan-les-Quates, Switzerland
[c] Advanced System Technology-STMICROELECTRONICS, Agrate Brianza, Milano, Italy

ARTICLE INFO

Article history:
Received 29 June 2015
Received in revised form
19 November 2015
Accepted 30 December 2015
Available online 2 February 2016

Keywords:
Field programmable gate arrays
Image elaboration
HW accelerator
Interest point detection

ABSTRACT

A high performance HW accelerator is proposed to extract and refine the Interest Points from images, by accurately calculating the Difference-of-Gaussian and using refinement algorithms from the SIFT method. Unique features of the accelerator consist in an accuracy comparable to the CDVS Test Model, reference software; in the capability to process the incoming pixel in streaming order to minimize the amount of embedded memory and avoid external frame buffers; in the possibility to configure the processor with different area/speed ratios. FPGA synthesis on a Xilinx XC7V2000T returns a maximum operation frequency up of 309 MHz at the fastest corner. Standard cell synthesis with the STMICROELECTRONICS FDSOI 28 nm technology, de-congestioned by the use of DPREG memories in place of SRAM, gives a maximum frequency of 1.2 GHz and a power dissipation of about 1 W at the typical conditions.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

The capability to initiate search queries from an object represented in a still or moving picture has enabled a large number of applications in the field of image retrieval, to be applied to e-commerce and entertainment, as well as to medical diagnostic, security, automotive and the like, so that several enabling methods have been proposed in the literature [1–5]. The plethora of proposed approaches has favored the birth of an official standardization process as part of MPEG-7 part-13, to which this work has contributed [6–8], with the purpose to collect the most effective methods to extract features from images and compactly describe them through a standard descriptor bitstream, in order to enable the efficient processing and transmission with the lowest bandwidth requirement [9]. Although the compliance of this last requirement is very important to simplify the server apparatus for retrieval and matching of VS queries, it implies that all the visual feature extraction and coding procedure must be executed locally to the clients, which must face the high computational demand to ensure the robustness of the descriptor. This, certainly, applies to the most referenced method, the Scale Invariant Feature Transform (SIFT) [3,4] that exhibits a very high stability against image modifications caused by resolution changes, affine transformations, illumination conditions and partial occlusions, at the cost, however, of a very high computational effort. This effort is mainly related to the scale invariant feature, obtained by extracting and refining the Interest Points (IPs) from a computational intensive scale-space pyramid of blurred, down sampled versions of the input image [10]. In turn, several alternatives to SIFT [5,9–12], either lack in accuracy, when simplifying the filtering and detection stages, or fail in the elaboration time. Indeed, SW implementations running on general purpose processors or, further worse, on resource constrained devices, exhibit very high latencies. From the profiling of Fig. 1, for example, appears clear the high computational demand of the filtering and detection phases, notwithstanding the running processor, ARM SoC (nVidia K1 with CUDA support), is strongly oriented to SIMD operations [6–8].

Similar limitations afflict the recent, all-HW solutions designed to reduce the computation time, whereas the accuracy is traded-off with area occupation, since a scale-space pyramid of adequate dimensions requires a large amount of memory for frame buffering and partial data storing, as well as a large number of arithmetic circuitry for parallel elaborations. In [13] part of the SIFT robustness is sacrificed by recurring to BRIEF for feature description; in [14] the area reduction has been obtained by reduced kernel dimensions of the input filters and by the elimination of the refinement phase; in [15] the memory requirement is increased to implements a task level parallelism in a ping-pong scheme; in [16] a simplified SIFT version is proposed to increase the parallelism; in [17] a large number of processing elements is used in a neural chip implemented by a Network-on-Chip (NoC) structure.

* Corresponding author. Tel.: +39 089962479.
E-mail address: gdlicciardo@unisa.it (G.D. Licciardo).

http://dx.doi.org/10.1016/j.vlsi.2015.12.010
0167-9260/© 2016 Elsevier B.V. All rights reserved.

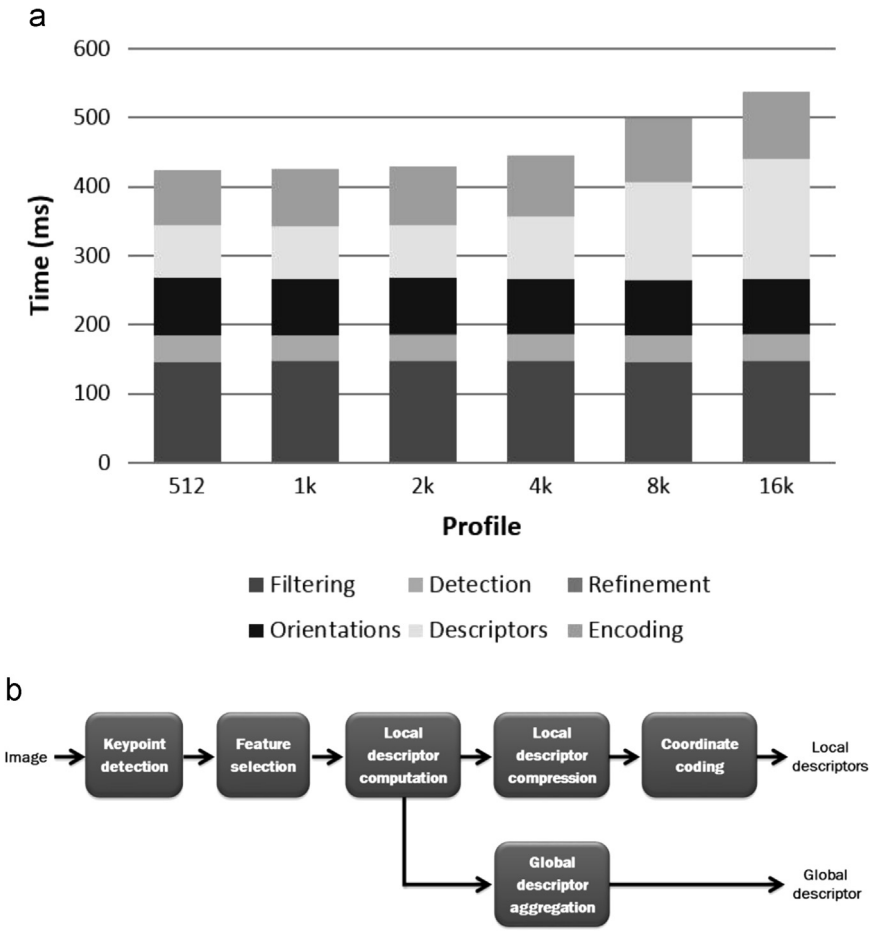

Fig. 1. (a) Extraction times on a nVidia K1 platform, Quad-Core ARM Cortex-A15 2.3 GHz. Filtering, detection and refinement times are all part of the keypoint detection stage. Although the processor is optimized for filtering operations, the computational effort of this part is evident; (b) Logical flow of the VS algorithm.

With the purpose to overcome the above accuracy limitations, in this work we propose an Application-Specific Processor (ASP) for the interest point detection and refinement, based on the Difference-of-Gaussian (DoG) calculation [18–20] and on the refinement methods proposed in SIFT [3,4], which effectively achieves the following features:

1. Streaming elaboration on input data received from image sensors [21].
2. The possibility to avoid frame buffers and off-chip memory, substituted by a reduced amount of on-chip SRAM or register files.
3. Platform independent implementation to operate with Field-Programmable Logics (FPL) as well as ASIC standard cells.
4. Accuracy comparable with the Compact Descriptor for Visual Search (CDVS) Test Model (TM) [22] to be useful for comparisons in the MPEG standardization process [6–8].
5. Scalable architecture that allows setting the area/speed ratio, in order to favor a platform independent implementation from Field-Programmable Logics (FPL) to ASIC standard cells.
6. At least 30 frame-per-seconds.

It is important to underline that this work is not SIFT, but it has been developed during the works of the ISO/IEC JTC1/SC29/WG11 MPEG7-Part 13 for VS standardization process, to demonstrate a HW implementation that could compete with the accuracy of the CDVS TM [22], but much less "hungry" of physical resources. Therefore, the proposed processor takes the best approaches from the existent literature to create the scale-space pyramid and the extraction and refinement of IPs [10]. Notwithstanding their computational complexity, these methods have been implemented in an efficient way, thus to be conveniently interfaced with different HW/SW systems for coding, as will be defined in the future by the standardization process. Further, the implemented features make the proposed solution useful in applications requiring smart sensors, which are capable to reduce the complexity of the control logic, as well as, the transmission bitrate, with the consequent simplification of the wiring apparatus. Synthesis on a Xilinx Virtex 7 [23] returns real-time performances on VGA (640×480 pixels) frames at a maximum frequency of 309 MHz in the fastest corner, while real-time performances can be obtained even on a much smaller Xilinx XCV5LX330 [24], which embeds a sufficient amount of BRAMs to fit the design. A STMICROELECTRONICS FDSOI 28 nm technology has been employed for standard cell synthesis at low temperature/low voltage, high temperature/low voltage, and typical temperature/voltage corners, in order to verify both speed and power dissipation. With an on-chip area of 4.4 mm^2, without any external memory, the ASIP consumes an overall power in the range [0.87, 1.37] W when the temperature varies between -40 °C and 125 °C, at the maximum frequency of 1.2 GHz, corresponding to theoretical 355 fps.

The remainder of the paper continues as follows: Section 2 describes the implemented algorithm and reports a numerical analysis to define the precision needed to represent partial data. Section 3 describes the architecture and the operations of the proposed design. Section 4 reports implementation results and comparisons with the existent literature. Section 5 concludes the paper.

Download English Version:

https://daneshyari.com/en/article/539436

Download Persian Version:

https://daneshyari.com/article/539436

Daneshyari.com