

Reconfigurable Block-based Normalization Circuit for On-chip Object Detection

Aiwen Luo¹, Fengwei An², Xiangyu Zhang², Lei Chen and Hans Jürgen Mattausch^{1,3}

Hiroshima University

¹Grad. School of Adv. Sciences of Matter, ²Grad. School of Engineering, ³Research Inst. for Nanodevice and Bio Systems
1-3-1 Higashi-Hiroshima, Hiroshima 739-8530, Japan

Phone: +81-82-424-5730 E-mail: aiwen-luo, anfengwei, zhangxiangyu, chen, hjm@hiroshima-u.ac.jp

Abstract

Normalization is critical for feature-vector optimization in vision-based detection systems. The presented block-based L1-norm-circuit architecture is configurable for different image-cell sizes, cell-based feature descriptors and image resolutions. The applied data-storage scheme includes flexible regulation according to customization parameters from the input. Pedestrian detection accuracy comparable to the L2-norm is obtained with greatly reduced computing complexity. An object-detection prototype system for performance evaluation in 65 nm CMOS implements the developed L1-norm circuit together with a cell-based HOG descriptor.

1. Introduction

Illumination intensity of light source, foreground-background contrast and the automatic gain control from a camera, etc., limit the performance of vision-based detection systems. To avoid degradation of recognition performance due to above issues, an effective local contrast normalization method turns out to be essential.

We present a general-purpose normalization-circuit architecture which is implemented by a reconfigurable ASIC-based solution in 65 nm CMOS technology. The simpler L1-norm instead of the L2-norm or the L2-Hys-norm is applied in our hardware design and is verified to meet a favorable tradeoff between computing complexity and detection accuracy. The developed L1-norm-circuit architecture is reconfigurable for any application using image-cell-based feature vectors (FVs), such as histograms of oriented gradients (HOG) [1] or Haar-like [2] descriptors. Additionally, strongly reduced on-chip memory requirements and wide performance-optimized applicability to multiple FVs are achieved.

2. L1-norm-circuit architecture and its optimization

Equation (1) defines the local contrast Lp-norm operation for a block of 2×2 image cells. Blocks slide with a fixed half-block stride (i.e., one cell stride) in our design. The rectangle cell size (CS) is only limited by the on-chip memory capacity.

$$d_i' = d_i / \left(|d_{i(\text{cell}0)}|^p + |d_{i(\text{cell}1)}|^p + |d_{i(\text{cell}2)}|^p + |d_{i(\text{cell}3)}|^p \right)^{1/p} \quad (1)$$

Here, d_i ($i \in [0, n-1]$) refers to one component of the n-dimensional cell-FV (e.g. $n=9$ in [1], while $n=4$ in [2], and $CS=8$). Thus each cell response contributes to several components of the final normalized block-descriptor vector.

The developed reconfigurable circuit architecture for

block-based L1-normalization consists of three main parts, as illustrated in Fig.1. The upper 'Block part' is used for caching and updating of intermediate block-summation results in the currently processed row of image blocks, which depends on assigned image resolution and CS in different applications. The necessary storage space for one row of blocks as well as the corresponding storing locations, i.e. read and write block addresses (BAs) for the block memory, are regulated in accordance with the cell number in the horizontal (CNH) and vertical (CNV) direction of the input image. Fig.2 illustrates more reconfigurable-architecture details of the 'block address decoder', which generates the BAs related to a given image cell according to eq. (2).

$$BA = \begin{cases} BA_0 = k - k/c \\ BA_1 = k - k/c - 1 \\ BA_2 = k - k/c - c \\ BA_3 = k - k/c - c + 1 \end{cases} \quad (2)$$

Here k and c represent the current-cell order number in the whole image and its CNH value, respectively. Cells marked identically in Fig.3 are overlapped by the same number of blocks (1, 2 and 4 blocks for cells marked in black, gray and white, respectively), but contribute to different block types (i.e. different BA_i in eq. (2)).

Cache and extension for one row of cells are handled in the lower 'Cell part' of the circuit architecture in Fig.1. The four cells related to a same block are outputted successively. To ensure a fixed-point computation in the middle 'Pipelined L1-norm processing' part of Fig.1, each cell-FV dimension is multiplied by a factor of 2^m ($m=12$ in the practical design). The final descriptor-FV of a detection window is constructed by combining the normalized block-FVs (i.e., d_0', d_1', \dots, d_n') of all related blocks. Important is, that the block and cell memories are reutilized after completion of the processing of the current block and cell rows, respectively. Consequently, on-chip memory requirements are strongly reduced.

3. Performance evaluation

An object-detection system with the developed general-purpose normalization circuit is prototyped in 65 nm CMOS and has 2.395 mm × 1.195 mm core area, as shown in Fig. 4. The layout of L1-norm circuit in Fig. 5 verifies its logical density. Table I summarizes the properties of our general-purpose L1-norm circuit in comparison to [3]. Figures 6 and 7 show, that a detection system with L1-norm achieves comparable accuracy performance to the case where L2-norm is

applied and works with much better accuracy than without normalization (no-norm) for different feature descriptors and classifiers. It is verified that 98.23% pedestrian-detection accuracy with L1-norm can be achieved by a pre-trained linear support vector machine (SVM) classifier, which allocates a trained threshold and a fixed weight for each dimension of a 3780-dimensional HOG descriptor for 5656 scan-windows of 64×128 -pixel size.

4. Conclusions

The reported reconfigurable circuit architecture for L1-normalization is compatible with various cell-based feature

descriptors, flexibly adapts data-storage resources to input parameters for customization, and enables a favorable trade-off between computing complexity and detection accuracy.

Acknowledgements

The VLSI-chip was fabricated through the chip fabrication program of VDEC, the University of Tokyo in collaboration with, Renesas Electronics, Synopsys, and Cadence.

References

- [1] X. Zhang, *et al.*, J. Appl. Phys. **56.4S** (2017) 04CF01.
- [2] A. Luo, *et al.*, J. Appl. Phys. **56.4S** (2017) 04CF06.
- [3] K. Mizuno, *et al.*, IEICE Trans. Electron., **E96-C** (2013) 433

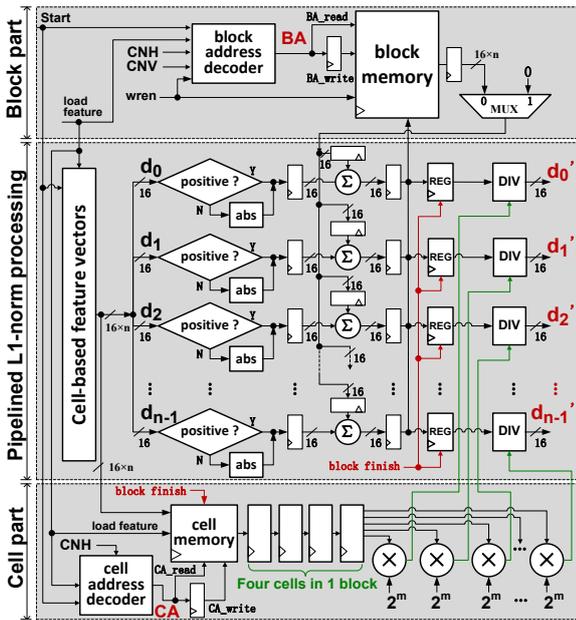


Fig.1 Reconfigurable circuit architecture for block-based L1-normalization of cell-based feature vectors.

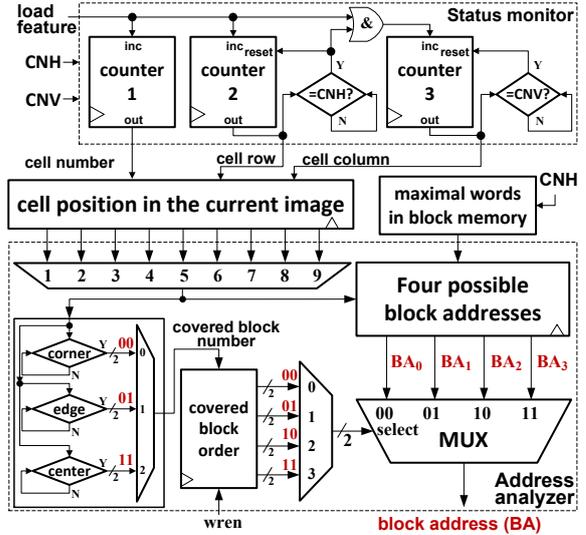


Fig.2 More detailed structure of the 'block address decoder' control circuit (see Fig.1) for coordinating the reconfigurable normalization with cell position and image size, designated by the configuration parameters CNH (up to 128) and CNV (unlimited).

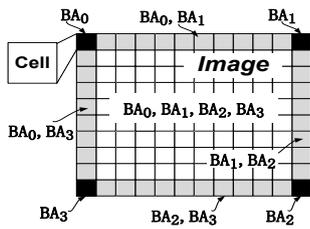


Fig.3 The corresponding BAs for all cells according to their positions.

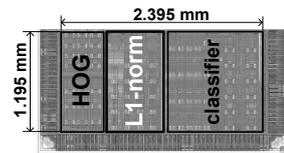


Fig.4 Micrograph of the prototype chip in 65 nm CMOS technology.

Table I Comparison with previous work

	IEICE [3]	This work
Technology	65 nm CMOS	65 nm CMOS
Memory on chip	610 kb (one core)	328 kb (36kb for norm)
Normalization	L2-Hys-norm	L1-norm
Feature descriptor	Cell-based HOG	Cell-based HOG [1]
Resolution	1980×1080 pixels	128cells×∞ (up to 8×8 pixel cell)
Frame rate	30 fps (HD)	31 fps (XGA)

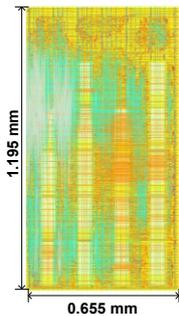


Fig.5 Layout of the developed circuit for L1-normalization.

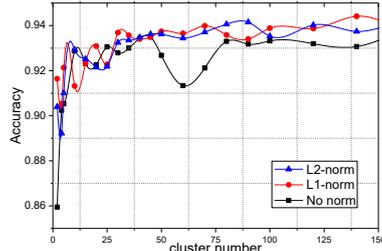


Fig.6 Pedestrian-detection accuracy as a function of the cluster (reference) number. A cell-based Haar-like descriptor [2] and a nearest neighbor search (NNS) classifier are used.

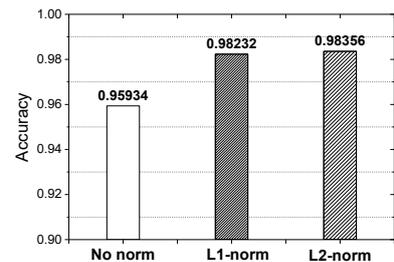


Fig.7 Pedestrian-detection accuracy with cell-based HOG feature descriptor [1] and a pre-trained SVM classifier, based on 5656 scan-windows of 64×128 -pixel size.