Extended Abstracts of the 2016 International Conference on Solid State Devices and Materials, Tsukuba, 2016, pp455-456

J-5-03

Cell-Scan-based Hardware Architecture for HOG-Feature Extraction

Xiangyu Zhang, Fengwei An, Ikki Nakashima, Lei Chen, and Hans Jürgen Mattausch

Hiroshima University, 1-3-1 Higashi-Hiroshima, Hiroshima 739-8530, Japan
Phone: +81-82-424-5730 E-mail: zhangxiangyu, anfengwei, chen, hjm@hiroshima-u.ac.jp

Abstract

Feature extraction on embedded systems is a challenging and important issue for object recognition. In this paper, we propose a hardware architecture with parallelized voting elements (PVEs) for cell-based histograms of oriented gradients (HOG) feature extraction, which is known to provide high efficiency in recognition accuracy. In addition, a cell-based scan synchronization with the image sensor accelerates the extraction speed and avoids using buffers for image frames or integral images. A test chip in 180 nm CMOS technology is fabricated to extract 3780-dimensional HOG feature vectors in windows with 128×64 pixels. Experimental results show that the proposed architecture achieves faster processing speed and larger flexibility for different image resolutions, with much less hardware and energy cost than previous works.

1. Introduction

Real-time processing is one of the fundamental problems in the very active and broad object-recognition research field. The histogram of oriented gradients (HOG) descriptor is one of the most widely used descriptors for object recognition due to its high recognition accuracy in a large variety of recognition tasks and its robustness against illumination changes. Unfortunately, traditional software or GPU-based implementations are insufficient for battery-limited mobile systems.

In this paper, we develop a hardware implementation for HOG descriptor extraction with cell-based scan manner. A prototype chip in 180 nm CMOS technology achieves fast processing speed as well as high flexibility for different image resolutions and small hardware costs.

2. Cell-based HOG-Descriptor Extraction

The HOG algorithm for object recognition was introduced in [1]. Gradient orientation and magnitude according (1) and (2) are calculated for each pixel. Then a histogram of gradient orientations of all pixels in an image cell of 8×8 pixels is computed. For each cell a 9-bin histogram for gradient angles between 0 and 180 degrees is generated by accumulating weighted gradient magnitudes in each of the bins.

$$O(i,j) = tan^{-1}\left(\frac{G_y(i,j)}{G_x(i,j)}\right) \qquad (1)$$

$$M(i,j) = \sqrt{G^2{}_x(i,j) + G^2{}_y(i,j)} \qquad (2)$$

Blocks of 16×16 pixels are defined to have horizontal and vertical overlaps of one cell. This results in 7×15=105 blocks for a detection window with 128×64 pixels. The histogram of a block is viewed as a 9×4=36 dimensional vector formed by concatenating the cell histograms within the block. Consequently, the HOG descriptor of a detection window forms can be represented as a vector with 105×36=3780 dimensions.

Instead of a block-based HOG-descriptor extraction, we propose a cell-based scan method, which exploits the cell-overlap characteristics of blocks for cell-histogram reuse in multiple block histograms. As indicated in Fig. 1 by the numbers in each of the 128 detection-window cells, corner (4), edge (40) and internal cells (84) can be used 1, 2 and 4 times, respectively, for HOG-descriptor-vector construction. Therefore, the proposed cell-based method reduces the construction time of the HOG-descriptor vector by a factor 3.28 (1×4+2×40+4×84=420/128) in comparison to the block-based method.

Furthermore, the cell-based scan method enables synchronized processing with the image sensor, since each transferred pixel data can be processed immediately. This means, that frame or integral-image buffers are not needed for its implementation. The histogram memory has to store only x/8 intermediate partial descriptor vectors of one cell row in case of an input image with x×y pixels. After the HOG-feature calculation of x/8 cells is completed, the corresponding memory locations can be overwritten by the intermediate data for the next cell row. This leads to the processing flexibility of input images with in principle unlimited height.

3. Hardware Architecture

The developed pipelined architecture with dual-port histogram memories for cell-based HOG descriptor extraction and synchronization to the input pixels from the image sensor is shown in Fig. 2. It can be divided into three main parts, namely memory-address control unit, calculation unit and vote unit with the dual-port histogram memories (DHM). The width of an input image can be freely specified and is only limited by the storage capacity of the DHMs, enabling high flexibility for processing of different input-image sizes.

The control unit is composed of two counters which generate pixel and cell addresses, respectively. Main responsibility of the calculation unit is bin decoding for each input pixel. To avoid square and root operation, a Manhattan approximation for eq. (2) with absolute values of $|G_x|$ and $|G_y|$ is used. The orientation-angle calculation is combined with the bin assignment to avoid the resource-expensive division for determining $\tan(O(i,j))$ in eq. (1). Firstly, we divide the range from $-90°$ to $90°$ into 9 bins covering $20°$ each, e.g. $-90°$ to $-70°$. Rather than initializing 9 tangent angles (bins) in a look-up-table [2], we adapt only 4 bins since the tangent function is a monotonically increasing odd function in this angle range. Accordingly, five 16-bit multipliers can be saved when compared to [2]. The bin assignment for each

gradient can be transformed from original eq. (1) to relation checks, i.e. $\tan(-90°) \leq G_y < \tan(-70°) \times G_x$. Only half of the interval is computed while the other half is obtained by comparing the signs of G_x and G_y. Dual-port histogram memories in the vote unit ensure that magnitude accumulation for updated bins can be completed in 1 clock cycle.

4. Experimental Results and Conclusions

A semi-custom test chip (see Fig. 3) was fabricated in 180 nm CMOS technology to implement the architecture for HOG descriptor extraction by the cell-based scan method in synchronization with pixel input. Total chip area is 1.59 mm^2, mainly needed for on-chip dual-port memory of 2.25 KB implementing the nine 128×16-bit DHMs. The word precision for G_x, G_y, and histogram values is 16 bit, which provides reasonable classification accuracy and minimizes hardware cost [3]. Power consumption is 42.3 mW at measured maximum frequency of 120 MHz at 1.8 V supply voltage.

The chosen DHM configuration for storing the x/8 intermediate HOG descriptors of an image row allows to handle a maximum input-image width of 1024 pixels, while the input-image height is unlimited. For the application example of XGA (1024 × 768) resolution videos, HOG-feature vectors can be extracted at 120 MHz operating frequency with a maximum frame rate of 122 fps.

Performance comparison to the state-of-the-art work [4] is presented in Table I. Instead of limitation to a fixed resolution, the reported prototype of the proposed architecture achieves a resolution flexibility of up to 1024 × ∞ pixels, where the maximum image width is only limited by the actually implemented dual-port-memory configuration. Even though a much less advanced CMOS technology is used, our prototype further demonstrates higher maximum operating frequency, a factor 67 less memory usage, a factor 360 lower energy consumption, and a factor 2.5 smaller Si-area.

In conclusion, this work presents a hardware architecture for HOG-descriptor extraction with a cell-based scan method and a prototype in 180 nm CMOS technology, which has high image-size processing flexibility (e.g. prototype can handle all images with of ≤1024 × ∞ pixels) and is suitable for low-power embedded systems.

Acknowledgements

This research was supported by grant 25420332 from the Ministry of Science and Education, Japan. The VLSI-chip was fabricated through the chip fabrication program of VDEC, the University of Tokyo in collaboration with, Rohm, Synopsys, and Cadence. The used standard cell library was developed by Tamaru/Onodera Laboratory of Kyoto University and released by Professor Kobayashi of Kyoto Institute of Technology.

References

[1] N. Dalal and B. Triggs, *In Computer Vision and Pattern Recognition, IEEE Computer Society Conference on* (2005) 886.

[2] M. Hahnle, et al., In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (2013) 629.

[3] K. Mizuno, *et al.*, *In Signal Processing Systems, IEEE Workshop on* (2012) 197.

[4] K. Takagi, *et al.*, *In Acoustics, Speech and Signal Processing, IEEE International Conference on* (2013) 2533.

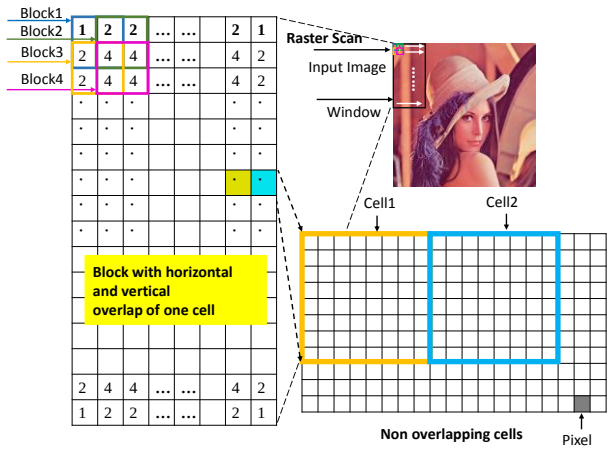

Fig.1 Block-overlap characteristics for cells in a detection window.

Table I Performance Comparison to Previous Work

	[4]	This Work
CMOS technology	65nm	180nm
Power dissipation (mW)	99.52 (42.9 MHZ @1.1V)	42.3 (120 MHZ @1.8V)
Storage size (M-bit)	1.22	0.018
Core size (mm^2)	3.96	1.59
Image resolution	HDTV (1920 × 1080 pixels) @30 fps@110 MHz	1024 × ∞ pixels Eg. (1024 × 768 pixels) @122 fps@120 MHz
Flexibility for image size (pixels)	only 1920×1080	≤1024 × ∞
Energy consumption *	1.6 nJ/pixel	4.4 pJ/pixel

*: Energy consumption = Power dissipation/(Image resolution*fps).

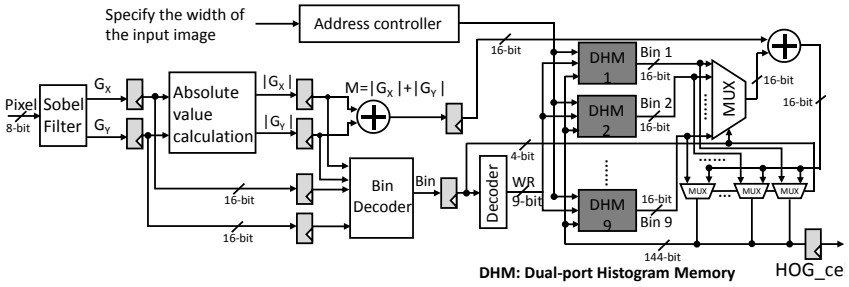

Fig.2 Block diagram of the hardware implementation for HOG-descriptor extraction with the proposed cell-based scan method.

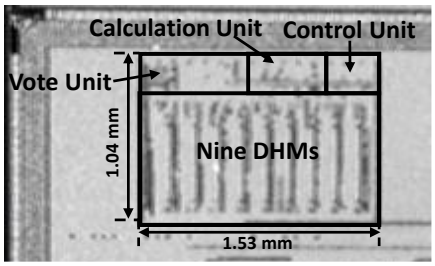

Fig.3 Micrograph of the fabricated chip in 180 nm CMOS technology.