Heterogeneous vision chip and LBP-based algorithm for high-speed tracking

Jie Yang, Cong Shi, Liyuan Liu and Nanjian Wu

A local binary pattern (LBP)-based tracking algorithm and a novel heterogeneous vision chip architecture that implements the algorithm in high speed are proposed. The algorithm is more robust than previously reported high-speed tracking algorithms. The proposed vision chip architecture adopts multiple levels of parallel processors to execute the algorithm in pixel-parallel and patch-parallel ways. Experimental results show that the proposed implementation can achieve 1000 fps robust tracking.

Introduction: A vision chip adopts massively parallel SIMD processing elements (PEs) to carry out image processing algorithms in a pixel-parallel fashion [1–4]. It is widely applied in the field of high-speed object tracking. However, reported vision chips can only perform simple and less robust algorithms such as background subtraction and self-window; thus they can only be applied to certain scenarios with a clear background and constant illumination. In this Letter, we propose a local binary pattern (LBP)-based tracking algorithm. The feature of the tracked object is characterised using an LBP descriptor that is discriminative and invariant to the illumination change. The algorithm is more robust than previously reported vision chip tracking algorithms. Moreover, we develop a heterogeneous vision chip architecture that can perform the algorithm in multiple levels of parallelism. Experiments and performance analysis are carried out. The results indicate that our proposed implementation can achieve 1000 fps robust and high-speed tracking.

Proposed tracking algorithm: The LBP is one of the best performing texture descriptors and it has wide applications [5]. A well accepted way to form a global description with a LBP is to divide the image into several regions from which the LBP histograms are extracted and concatenated into an enhanced vector [6]. The main components of the proposed LBP-based high-speed tracking algorithm are shown in Fig. 1. The object window is first divided into a grid of sub-regions. The LBP histogram of each sub-region is then concatenated to form a global feature of the object, as shown in Fig. 1*a*. This method has been successively applied to facial image analysis, and it has proven to be discriminative and robust. The tracking procedure is shown in Fig. 1*b*. In the successive frame, the global features of the search windows surrounding the object location are generated and compared with the feature of the object. The sum of absolute difference (SAD) between the features of the search window and the feature of the object window is calculated. The search window with the smallest SAD value is updated as the object window. If the proposed algorithm is processed in high speed, the difference between two successive frames will be limited to a few pixels, even with rotation and scale changes. The feature is updated for every frame to adapt to affine and scale changes of the object. In the following Section, the algorithm is implemented on our vision chip in a multiple parallel way.

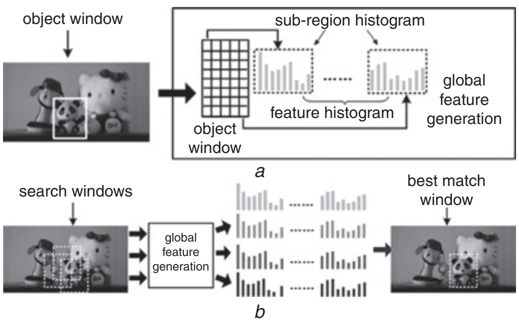

Fig. 1 *Main components of our LBP-based tracking algorithm*
a LBP feature extraction procedure
b Find object among search windows

Proposed vision chip architecture: The proposed vision chip architecture is shown in Fig. 2. It is mainly composed of an $N \times N$ macro block (MB), a RISC core and some other necessary logic. The RISC core controls the overall system. The MB consists of $N_1 \times N_1$ PEs and a patch processing unit (PPU). The PE contains a local memory and an arithmetic and logic unit (ALU) that can perform 1 bit *and, or, addition* and *inversion* operations. The $(N_1 \times N)^2$ PE array has an $O(N \times N)$ parallelism and can perform some algorithms in a pixel-parallel way [2, 3]. The PPU has an RISC-like structure. It consists of an instruction sequencer, an 8 bit ALU, a general purpose register, a PE input buffer register and a local memory. The PPU can perform arithmetic operations and index addressing that cannot be accomplished by the PE. Similar to the PE array, the $N \times N$ PPU array has an $O(N \times N)$ parallelism, and it can process algorithms in a patch-parallel way.

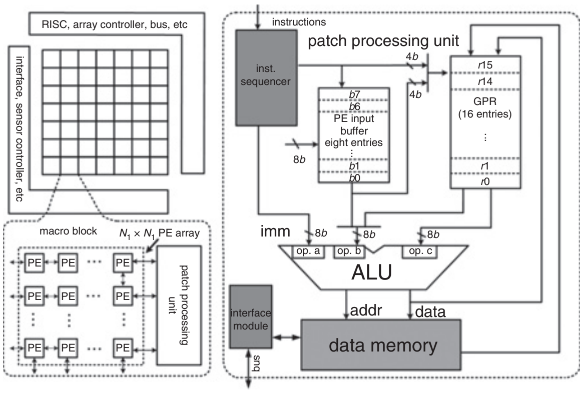

Fig. 2 *Proposed vision chip architecture, schematics of MB and PPU*

Compared with the previously reported architecture with row-parallelism [3], the topology of this architecture is more suitable for patch-parallel processing. It is highly coupled with the proposed algorithm. The LBP generation and sub-region histogram calculation procedures are performed and fully speeded up by the PE array and PPU array in the pixel-parallel way and the patch-parallel way, respectively. The PE array can finish the LBP generation of an $(N_1 \times N)^2$ resolution image within 800 instruction cycles. The PPU array can generate the histograms of $N \times N$ image patches simultaneously. Furthermore, the image patch size processed by PPU is not limited by the $N_1 \times N_1$ PE size of the MB. Data in one MB can be easily moved to the adjacent MB with the help of the PE array data transfer operation. The feature match procedure can also be finished by the PPUs in parallel. Each PPU can store the histograms of the same sub-region of different frames. Then the histogram SAD difference of all sub-regions can be calculated in PPUs in parallel. At last, the RISC accesses the PPUs, and only a few operations are needed before determining the best match window and the object location.

Implementation and experimental results: A tracking system was built as shown in Fig. 3*a*. It contains a camera, two actuators and a field programmable gate array (FPGA) which is behind the camera. The proposed architecture is realised on the FPGA. It consists of a 128 × 64 PE array and 16 × 8 PPUs uniformly distributed along the PE array. Each PE includes 64 bits memory and each PPU can access 4096 (8 × 8 PE, 64 bits per PE) bits PE memory. The PE array and the PPU array can give a peak performance of 56 GOPS (giga-operations per second) at a clock frequency of 50 MHz when performing 8-bit arithmetic operations. The prototype system continues capturing the target and extracting the position of the target. According to the position of the target, the actuators adjust to locate the target in the centre of view. Fig. 3*b* gives the record of the motion of the target during the target tracking experiment. The collection of 1110 data points clearly exhibits the sinusoidal motion of the target. A more detailed performance analysis of the proposed architecture is shown in Fig. 4. The breakdown of time and computation for the proposed vision chip performing the tracking algorithm on a 128 × 128 pixel image are shown in Fig. 4*a*. Thanks to the pixel-parallel PE array, 10.8 million operations are finished consuming only 194 µs. Data transfer and histogram operations are performed by the PPU array. It takes 309 µs to finish about 2.1 million operations. The RISC is used to control the whole system and issue some of the SIMD instructions for the PE array and the PPU array. It consumes 490 µs. Fig. 4*b* shows the instruction count of some kernel functions of the algorithm. LBP operators with different scales are realised. The

$\mathrm{LBP}_{8,1}$ and $\mathrm{LBP}_{16,2}$ operators are implemented in 793 and 1673 PE instruction cycles, respectively. An 8 × 8 patch histogram is finished with a PPU and its corresponding 8 × 8 PEs. It takes 64 PE and 222 PPU instruction cycles, respectively. Histograms with bigger patch size can be finished using additional PE data transfer instructions. The SAD of two histograms can be finished in 100 PPU instruction cycles.

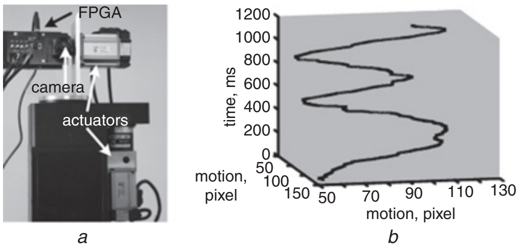

Fig. 3 *Experiment setup and result*

a Prototype tracking system
b Trajectory of target

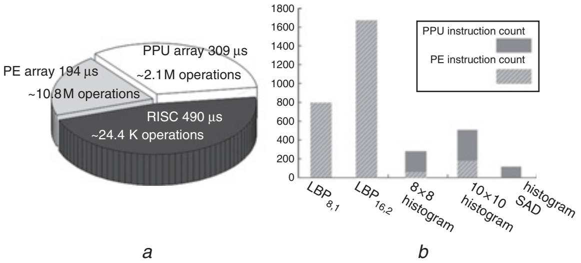

Fig. 4 *Detailed performance analysis of proposed architecture*

a Breakdown of time and computation for proposed vision chip performing tracking algorithm on 128 × 128 pixel image
b Instruction count of some kernel functions

Table 1 shows a comparison with previously reported vision chips. The algorithms proposed in these studies either utilise a threshold value to obtain a binary image [3, 4] or capture a binary image in the first place [1, 2]. These algorithms can all be classified as binary image-based methods. The algorithms can only work in artificial environments with clear a background and constant illumination. The proposed algorithm adopts a robust LBP feature that can work in much more complex environments. Furthermore, the proposed architecture has two levels of parallelism. The integration of PPUs greatly enhances the system flexibility and performance for the proposed algorithm. The architecture in [3] has the highest tracking speed among those listed in Table 1, however the algorithm is simple. If our proposed algorithm is implemented using the architecture described in [3], the tracking speed will be below 1000 fps.

Table 1: Comparison with previous studies

	This work	Komuro *et al.* [1]	Miao *et al.* [2]	Zhang *et al.* [3]	Zhao *et al.* [4]
Image greyscale (bit)	8	2	2	8	8
Algorithm	LBP based	Binary image based	Binary image based	Binary image based	Binary image based
Architecture	PE array PPU array	PE array	PE array	PE array RP array	N/A
Resolution	128 × 128	128 × 128	16 × 16	128 × 128	64 × 64
Frequency (MHz)	50	40	20	100	N/A
Tracking speed (fps)	1000	1000	1000	10 000	100

Conclusion: This Letter introduces a high-speed tracking implementation composed of an LBP-based tracking algorithm and a heterogeneous vision chip architecture. Taking advantage of the LBP texture descriptor, the proposed algorithm is more robust than those in previously reported studies. In addition to the pixel-parallel PE array, the proposed architecture utilises the PPU to finish complex operations and gains the patch-parallel processing ability. Experimental results show that the proposed high-speed tracking implementation can achieve 1000 fps tracking. Our analysis indicates that the pixel-parallel and patch-parallel designs of the architecture greatly improve system performance and reduce the algorithm's execution time.

Acknowledgments: This work was supported by the National Natural Science Foundation of China (grant nos 60976023 and 61234003) and the Special Funds for the Major State Basic Research Project of China (grant no. 2011CB93-2902).

13 January 2014
doi: 10.1049/el.2014.0033

Jie Yang, Cong Shi, Liyuan Liu and Nanjian Wu (*Institute of Semiconductors, Chinese Academy of Sciences, No. 35 QinHua East Road, Beijing 100083, People's Republic of China*)

E-mail: nanjian@red.semi.ac.cn

References

1 Komuro, T., Ishii, I., Ishikawa, M., and Yoshida, A.: 'A digital vision chip specialized for high-speed target tracking', *IEEE Trans. Electron Devices*, 2003, **50**, (1), pp. 191–199, doi: 10.1109/TED.2002.807255
2 Miao, W., Lin, Q., Zhang, W., and Wu, N.-J.: 'A programmable SIMD vision chip for real-time vision applications', *IEEE J. Solid-State Circuits*, 2008, **43**, (6), pp. 1470–1479, doi: 10.1109/JSSC.2008.923621
3 Zhang, W., Fu, Q., and Wu, N.-J.: 'A programmable vision chip based on multiple levels of parallel processors', *IEEE J. Solid-State Circuits*, 2011, **46**, (9), pp. 2132–2147, doi: 10.1109/JSSC.2011.2158-024
4 Zhao, B., Zhang, X., Chen, S., Low, K.-S., and Zhuang, H.: 'A 64 × 64 CMOS image sensor with on-chip moving object detection and localization', *IEEE Trans. Circuits Syst. Video Technol.*, 2012, **22**, (4), pp. 581–588, doi: 10.1109/TCSVT.2011.2170119
5 Ojala, T., Pietikainen, M., and Maenpaa, T.: 'Multiresolution gray-scale and rotation invariant texture classification with local binary patterns', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, (7), pp. 971–987, doi: 10.1109/TPAMI.2002.107623
6 Ahonen, T., Hadid, A., and Pietikainen, M.: 'Face description with local binary patterns: application to face recognition', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2006, **28**, (12), pp. 2037–2041, doi: 10.1109/TPAMI.2006.244