High-speed visual tracking with mixed rotation invariant description

Yongxing Yang, Jie Yang, Zhongxing Zhang, Liyuan Liu and Nanjian Wu✉

A mixed rotation invariant description (MRID)-based tracking algorithm and a novel high-speed visual tracking system that implements the algorithm are proposed. MRID is a novel rotation invariant description of texture and edge information by annular histograms and dominant direction. It overcomes rotation variant and large computation issues in conventional local binary pattern histograms of oriented gradient (LBP-HOG) feature description. The proposed tracking system contains an image sensor, a hierarchical vision processor and a two dimension of freedom actuator. The vision processor integrates processors with pixel and row-level parallelism to speed up the tracking algorithm. Experiment results show that the proposed system can achieve over 1000 fps processing speed of the tracking algorithm.

Introduction: Visual target tracking is widely applied in visual surveillance, human–computer interaction, visual navigation and activity analysis. However, the response speed of conventional tracking systems is limited to <60 fps due to serial processing. Some researchers adopt parallel single-instruction-multiple-data (SIMD) processors to speed up tracking algorithms [1–3]. However, these processors can only carry out simple algorithms such as background subtraction, segmentation and motion detection, thus they can only be applied to certain sceneries with a clean background. The local binary pattern (LBP) histogram of gradient (HOG) feature description is widely used in target detection and tracking [4, 5]. However, both the HOG and LBP histograms are rotation variant, which results in target shifting and tracking failure. In this Letter, we propose a mixed rotation invariant description (MRID)-based tracking algorithm and a novel high-speed visual tracking system. This MRID is invariant to rotation and illumination changes so that it achieves more robust tracking than previously reported fast tracking algorithms. The proposed tracking system integrates processors with pixel and row-level parallelism to speed up the tracking algorithm. The system with hierarchical parallelism can achieve over 1000 fps processing speed.

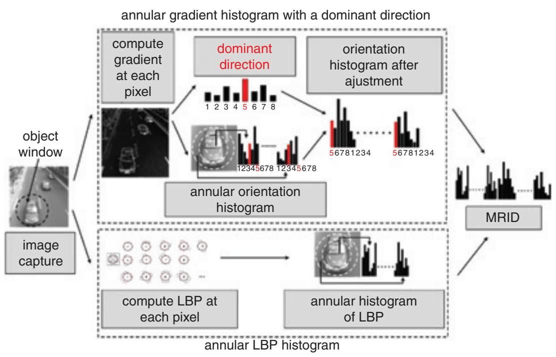

Fig. 1 *Process of proposed MRID*

MRID-based tracking algorithm: The proposed MRID is a novel rotation invariant description. It adopts a mixed annular LBP histogram and an annular gradient histogram with a dominant direction to describe the texture and edge information of an object, as shown in Fig. 1. In the annular gradient histogram, first the horizontal and vertical gradients of each pixel are calculated by masks $[-1\ 0\ 1]$ and $[-1\ 0\ 1]^T$, respectively. Then, the magnitude and orientation of the gradient are calculated. The orientation histogram is obtained with the range of 45° in each orientation bin. Next, the highest peak is detected as the dominant direction of the object. Compared with the conventional cellular histogram, the object window is a circular patch instead of a rectangle region and it is subdivided into annular spatial bins to improve distinctiveness while maintaining rotation invariance. The annular histogram of orientation is obtained from the object window and the order of bins is adjusted according the dominant direction. In the annular LBP histogram, the uniform LBP of each pixel is calculated as in [6]. Then, the LBP histogram of the annular object window is obtained. Lastly, the annular LBP histogram and annular gradient histogram with a dominant direction are concatenated to form a rotation invariant feature of the object. Compared with the conventional LBP-HOG feature, MRID overcomes rotation variant and large computation issues.

The tracking algorithm is shown in Fig. 2. In the successive frames, the MRID of search windows are generated and classified by the Adaboost classifier [7]. It combines K weak classifiers $\boldsymbol{g}(\boldsymbol{i})$ with a specific weight $\boldsymbol{\alpha}(\boldsymbol{i})$ to form a strong classifier as shown in formula (1)

$$\boldsymbol{C} = \sum_{i=0}^{K-1} \boldsymbol{\alpha}(\boldsymbol{i}) \times \boldsymbol{g}(\boldsymbol{MRID}(\boldsymbol{i}) - \boldsymbol{T}(\boldsymbol{i})) \tag{1}$$

where

$$\boldsymbol{g}(\boldsymbol{MRID}(\boldsymbol{i}) - \boldsymbol{T}(\boldsymbol{i})) = \begin{cases} 1 & \boldsymbol{MRID}(\boldsymbol{i}) - \boldsymbol{T}(\boldsymbol{i}) \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

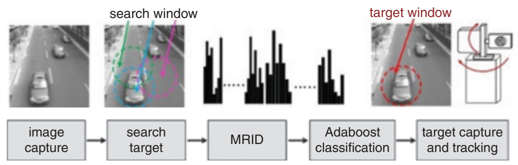

Fig. 2 *Process of MRID-based tracking algorithm*

The weak classifier $\boldsymbol{g}(\boldsymbol{x})$ consists of a feature **MRID(*i*)** which corresponds to a single MRID histogram bin and a threshold vector $\boldsymbol{T}(\boldsymbol{i})$ corresponding to **MRID(*i*)**. Both $\boldsymbol{\alpha}(\boldsymbol{i})$ and $\boldsymbol{T}(\boldsymbol{i})$ are obtained with the Adaboost training algorithm [7]. The value $\boldsymbol{C}$ of the search window shows the classification result: target or not. If $\boldsymbol{C}$ is larger, the search window is more likely to be the target. Then, the search window with the largest $\boldsymbol{C}$ is updated as the target window during tracking.

Hierarchical vision processor: As is shown in Fig. 3, the proposed vision processor contains an $M \times M$ processing element (PE) array, an $M \times 1$ row processor (RP), a dual-core MPU and a motor controller. The pixel-parallel PE array and row-parallel RP array can speed up low-level and middle-level image processing operations by $O(M^2)$ and $O(M)$, respectively. Each PE consists of a piece of local memory, a temporary register, several multiplexers and a 1-bit arithmetic logic unit (ALU) performing operations such as *and, or, addition* and *inversion*. The RP consists of a local memory, a register file, some multiplexers and an 8-bit ALU. The ALU can carry out operations such as max/min selection, absolute value calculation, subtraction and addition. RP can also perform register index addressing to accelerate middle-level algorithms like histogram extraction and vector calculation in row paraller manner.

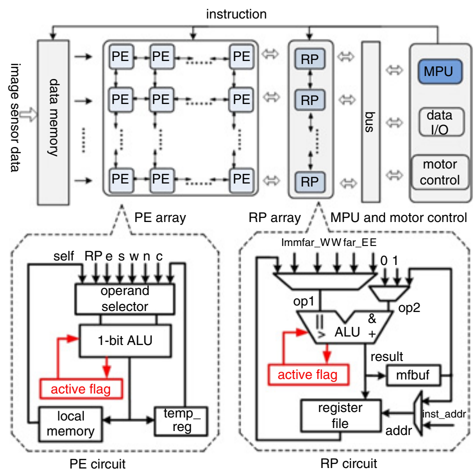

Fig. 3 *Proposed vision processor*

Compared with previous SIMD processors [1–3], this processor improves its flexibility with little circuit cost. For some operations such as annular histogram calculation, the SIMD manner will result in inefficiency and even errors (e.g. a rectangle histogram instead of an annular histogram is calculated because all the PE/RP works simultaneously). Multiple-instruction-multiple-data is a solution but it costs a lot. In the proposed processor, there is an 'active flag' (red rectangle in Fig. 3) in each PE and RP. The PE/RP works only when the active flag equals to 1. This flag can be set in advance or calculated during tracking. This design only costs several additional logic gates compared with SIMD processors but it achieves better flexibility than the original processor.

System implementation: The proposed target tracking system is shown in Fig. 4*a*. It contains an image sensor, a vision processor implemented on a FPGA and an actuator. In this prototype, the processor consists of a 128×128 PE array, a 128 RP array and an MPU; the operating frequency is 100 MHz. The proposed algorithm on a 750×480 pixel image is implemented on the proposed processor. In detail, LBP and gradient calculation only need the greyscale values of the calculated pixel and its surrounding pixels. These operations can be carried out by the PE array with pixel parallelism. The RP array can extract the annular histogram and calculate the classifier $\boldsymbol{C}$ in a row-parallel way. The MPU captures the target position and controls the motors tracking. The breakdown of time and computation for the proposed vision processor performing the tracking algorithm are shown in Fig. 4*b*. Thanks to the hierarchical parallelism and flexible architecture, the total processing speed reaches 1094 fps (914 µs). Fig. 4*c* shows the execution time of some kernel functions of the algorithm.

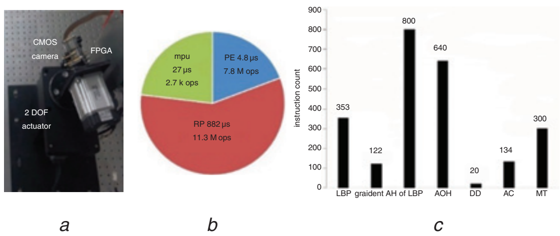

Fig. 4 *System implementation*

a Photo of proposed tracking system
b Breakdown of time and computation
c Instruction count of some kernel functions: AH of LBP (annular histogram of LBP), AOH (annular orientation histogram), DD (dominant direction), AC (Adaboost classification), MT (motor tracking)

Experiment results: Some experiments of tracking a moving car model were made to evaluate our tracking system. Fig. 5 shows the results of the target tracking. The actuator keeps tracking the target at the centre of view. Fig. 5*a* shows the x and y coordinates of its centre positions during tracking. It can be seen that the centre position of the target image was always controlled around the centre of the camera view. The trajectories of the actuator are also given in Fig. 5*b*. The tilt (red line) and pan (blue line) movement correspond to the periodical motion of the target.

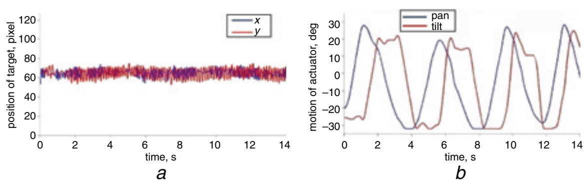

Fig. 5 *Tracking result*

a x and y coordinates of target centre
b Trajectories of actuator

Table 1: Comparison with related works

	This work	Ref. [2]	Ref. [3]	Ref. [4]
Image	8-bit mono	8-bit mono	8-bit colour	8-bit mono
Algorithm	MRID based	LBP based	Colour tracking	HOG-LBP
Resolution	750×480	128×128	512×512	N/A
Frequency (MHz)	100	50	151.2	N/A
Processing speed (fps)	1000	1000	2000	N/A

Table 1 shows comparison with previously reported tracking systems. In [3], the colour images are used to achieve an extremely fast tracking, they adopt several thresholds to convert a HSV image to a binary image. This method will lose the target when the colour of the background and the target are similar. In [2] and [4], an LBP and LBP-HOG-based tracking algorithm is implemented. They are more robust in environments with a complex background and changing illumination. However, both of them use the cellular histogram and neither is rotation invariant. This will result in failure when tracking objects with sudden rotation. Our proposed tracking algorithm adopts a robust MRID feature which can work in a more complex environment with large degrees of rotation and abrupt motion. Furthermore, the hierarchical processor is more flexible than the original SIMD processor and it greatly enhances the speed of our tracking algorithm to over 1000 fps.

Conclusion: This Letter proposes a novel MRID tracking algorithm implemented on a hierarchical vision processor. Thanks to the MRID feature, the tracking algorithm works more robustly than previously reported studies. The processor implements the algorithm with multiple parallelisms. The experiment result shows that our tracking system achieves over 1000 fps tracking in a complex environment.

Acknowledgments: This work was supported by the National Natural Science Foundation of China (grant nos. 61234003, 61434004) and the Special Funds for the Major State Basic Research Project of China (grant no. 2011CB932902).

Submitted: *3 September 2015*   E-first: *29 February 2016*
doi: 10.1049/el.2015.3071
One or more of the Figures in this Letter are available in colour online.

Yongxing Yang, Jie Yang, Zhongxing Zhang, Liyuan Liu and Nanjian Wu (*Institute of Semiconductors, Chinese Academy of Sciences, Beijing 100083, People's Republic of China*)

✉ E-mail: nanjian@red.semi.ac.cn

References

1 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
2 Yang, J., Shi, C., Liu, L., and Wu, N.: 'Heterogeneous vision chip and LBP-based algorithm for high-speed tracking', *Electron. Lett.*, 2014, **50**, (6), pp. 438–439, doi: 10.1049/el.2014.0033
3 Gu, Q., Al Noman, A., Aoyama, T., Takaki, T., and Ishii, I.: 'A fast color tracking system with automatic exposure control'. IEEE Int. Conf. on Information and Automation (ICIA), Yinchuan, China, August 2013, pp. 1302–1307, doi: 10.1109/ICInfA.2013.6720495
4 Baopu, L., Can, Y., Qi, Z., and Guoqing, X.: 'Condensation-based multi-person detection and tracking with HOG and LBP'. IEEE Int. Conf. on Information and Automation (ICIA), Hailar, China, July 2014, pp. 267–272, doi: 10.1109/ICInfA.2014.6932665
5 Qing, C., Dickinson, P., Lawson, S., and Freeman, R.: 'Automatic nesting seabird detection based on boosted HOG-LBP descriptors'. IEEE Int. Conf. on Image Processing (ICIP), Brussels, Belgium, September 2011, pp. 3577–3580, doi: 10.1109/ICIP.2011.6116489
6 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
7 Viola, P., and Jones, M.J.: 'Robust real-time face detection', *Int. J. Comput. Vis.*, 2004, **57**, (2), pp. 137–154, doi: 10.1023/B:VISI.0000013087.49260.fb