SCIENCE CHINA
Information Sciences

• LETTER •

May 2023, Vol. 66 159401:1–159401:2
https://doi.org/10.1007/s11432-021-3473-1

Low-cost real-time VLSI system for high-accuracy optical flow estimation using biological motion features and random forests

Cong SHI[1], Junxian HE[1], Shrinivas PUNDLIK[2],
Xichuan ZHOU[1*], Nanjian WU[3] & Gang LUO[2]

[1]*School of Microelectronics and Communication Engineering, Chongqing University, Chongqing* 400044*, China;*
[2]*Schepens Eye Research Institute, Massachusetts Eye and Ear, Harvard Medical School, Boston* MA 02115*, USA;*
[3]*State Key Laboratory for Superlattices and Microstructures, Institute of Semiconductors, Chinese Academy of Sciences, Beijing* 100083*, China*

Received 14 September 2021/Revised 15 December 2021/Accepted 2 April 2022/Published online 9 February 2023

Citation Shi C, He J X, Pundlik S, et al. Low-cost real-time VLSI system for high-accuracy optical flow estimation using biological motion features and random forests. Sci China Inf Sci, 2023, 66(5): 159401, https://doi.org/10.1007/s11432-021-3473-1

This study proposes a low-cost, real-time, very large-scale integration (VLSI) architecture for optical flow estimation. The architecture adopts parallel spatiotemporal filters to extract bio-inspired motion features at each pixel location and uses hardware random forests to infer the motion speed. Our system achieves higher estimation accuracy at low computational hardware costs under real-time constraint than previous biological motion estimation systems. A field-programmable gate array (FPGA) prototype of our VLSI system was implemented on a Xilinx Zynq-7045 FPGA chip. It achieved 30 frame/s motion estimation on 320 × 240 image sequences. The mean endpoint error was only 0.5 pixels for the horizontal translation at 8 pixels/frame, 0.7 pixels for in-plane rotation at 3°/frame, and 0.8 pixels for fast looming at a rate of 6%/frame, respectively.

The real-time and low-cost estimation of optical flows (i.e., motion vectors on two-dimensional (2D) image pixel arrays) is fundamental to many computer vision tasks [1]. To achieve this goal, biological optical flow VLSI systems leverage spatiotemporal filters to extract motion energy features [2] for pixel motion speed inference [3–6], which need much less computational costs than deep learning systems [7]. However, they could not handle high motion velocities accurately due to their lack of counting in temporal frequency aliasing in the motion energy (Appendix B). To break this limitation, this study proposes a low-cost, high-accuracy real-time optical flow VLSI system with embedded random forests to infer motion vectors from the motion energy. Our random forest core adaptively and implicitly dealiases temporal frequencies to provide accurate estimates on high velocities at a high processing throughput.

Algorithm flow. Figure 1(a) presents our optical algorithm flow using motion energy features and the random forest model. It consists of four steps. (i) Difference-of-Gaussian (DoG) filtering. A 2D DoG filter is applied to input image frames, mimicking human retina ganglion cell functions to suppress detrimental spatial components. (ii) 3D Gabor filtering and motion energy computation. It mimics the function of human cortical V1 cells. Separated horizontal (H) and vertical (V) channels apply 3D spatiotemporal Gabor filters to the DoG-filtered frames. In the H channel, each 3D Gabor filter is tuned to a particular spatiotemporal frequency pair (f_x, f_t) (with f_y fixed at 0). Meanwhile, in the V channel, each 3D Gabor filter is tuned to a spatiotemporal frequency pair (f_y, f_t) (with f_x fixed at 0). The horizontal and vertical motion energy features at each pixel are computed from the pixel values of the Gabor filtered images in the two channels. The 3D filter is decomposed into a 1D temporal Gabor filter, a 1D spatial Gaussian filter, and a 1D spatial Gabor filter for low-cost hardware implementation. (iii) Velocity inference. The random forest model (an ensemble of decision trees) [8] is used to robustly infer the motion vector of each pixel from their horizontal and vertical motion energies. Each decision tree infers independently, and their averaged result is the outcome of the random forests. The decision tree involves only a few comparisons without complicated multiplications commonly met in other intelligence models. (iv) Confidence labeling. An inferred motion velocity is labeled as confident only if its motion energy vector has an L1 norm above a threshold. Confidence flags are valuable in high-level visual tasks. Details about the algorithm are presented in Appendix B.

VLSI hardware design. The proposed optical flow VLSI hardware architecture is depicted in Figure 1(b). It consists of a DoG filter block, a temporal Gabor unit (TGaU) array, a steerable Gaussian unit (stGU) array, a spatial Gabor unit (SGaU) array, a motion energy compute unit (MEU) array, an L1-norm block, a random forest core containing a

* Corresponding author (email: zxc@cqu.edu.cn)

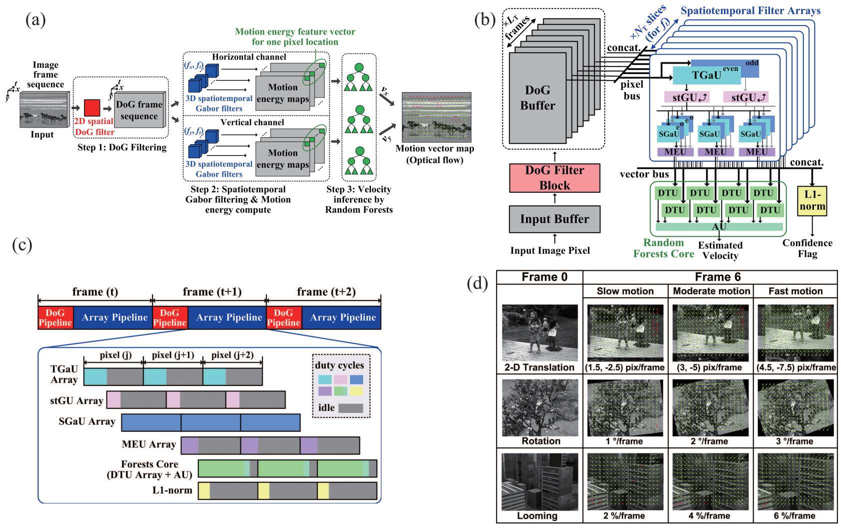

Figure 1 (Color online) (a) Our optical flow algorithm flow using biological motion energy features and random forests; (b) proposed optical flow VLSI system architecture; (c) timing diagram of the array-level pipeline; (d) optical flows estimated by our FPGA prototype. Red arrows in (d) represent motion vectors without the confidence flag.

decision tree unit (DTU) array with an averaging unit (AU), and some memory buffers. The pipelined working flow of the computational units is illustrated in Figure 1(c). The stGU unit can be dynamically steered (i.e., reconfigured) as either a vertical or a horizontal 1D Gaussian filter at runtime. Hence, the pipelined arrays can be reused for horizontal and vertical velocity component estimations to significantly save hardware costs. The DTU only holds a few simple logic components (i.e., comparators, multiplexers, and registers) and a moderate amount of memories. Its parameters are learned off-chip before inference. Our hardware architecture is scalable for different tradeoffs among the estimation accuracy, processing latency, and resource costs. Details on the architecture and circuit design are presented in Appendix C.

FPGA implementation. We prototyped the optical flow VLSI architecture on a Xilinx Zynq-7045 FPGA chip. It ran at a 100 MHz clock frequency, processed 320 × 240 images at 30 frame/s in real time, and consumed only < 20% logic resources and half of the memories on the FPGA chip. The experimental platform and prototype hyperparameter configurations can be found in Appendix D. To evaluate the FPGA prototype, we selected 20 real-world urban images from the Internet and resized them to a 320 × 240 grayscale format. Then, we used PC software to generate motion sequences under various motion patterns and speeds. Some optical flow results estimated by the FPGA prototype are shown in Figure 1(d). We adopted the standard endpoint error (EE) metric [9] to measure the optical flow accuracy. The mean EE of our prototype was 0.5 pixels for the rapid horizontal translation at 8 pixels/frame, 0.7 pixels for 3°/frame for the fast rotation, and 0.8 pixels for the quick looming at a rate of 6%/frame, respectively. For relative work comparisons and in-depth discussion about our algorithm and hardware design, refer to Appendix D.

Conclusion. This study proposes a low-cost real-time VLSI hardware system for accurate optical flow estimations based on biological motion energy features and embedded random forests. We employed pipeline processing arrays to improve the hardware system throughput. A 100 MHz FPGA prototype was implemented and achieved 30 frame/s real-time performance on 320 × 240 images. It consumed only < 20% logic resources and half of the memory resources on the FPGA. The optical flow error was only 0.5 pixels for the 1D horizontal translation at a high velocity of 8 pixels/frame and merely 0.7 and 0.8 pixels for the challenging 3°/frame in-plane rotation and 6%/frame looming, respectively.

Acknowledgements This work was supported in part by National Key Research and Development Program of China (Grant No. 2019YFB2204303), Science and Technology on Analogue Integrated Circuit Laboratory (Grant No. JCKY2019210C060), National Natural Science Foundation of China (Grant No. 61434004), and Key Project of Chongqing Science and Technology Foundation (Grant Nos. cstc2019jcyj-zdxmX0017, cstc2021ycjh-bgzxm0031)

Supporting information Appendixes A–E. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Fortun D, Bouthemy P, Kervrann C. Optical flow modeling and computation: a survey. Comput Vision Image Understanding, 2015, 134: 1–21
2 Grzywacz N M, Yuille A L. A model for the estimate of local image velocity by cells in the visual cortex. Proc R Soc Lond B, 1990, 239: 129–161
3 Shi C, Luo G. A compact VLSI system for bio-inspired visual motion estimation. IEEE Trans Circ Syst Video Technol, 2018, 28: 1021–1036
4 Etienne-Cummings R, van der Spiegel J, Mueller P. Hardware implementation of a visual-motion pixel using oriented spatiotemporal neural filters. IEEE Trans Circ Syst II, 1999, 46: 1121–1136
5 Orchard G, Thakor N V, Etienne-Cummings R. Real-time motion estimation using spatiotemporal filtering in FPGA. In: Proceedings of IEEE Biomedical Circuits and Systems Conference, 2013. 306–309
6 Orchard G, Etienne-Cummings R. Bioinspired visual motion estimation. Proc IEEE, 2014, 102: 1520–1536
7 Li Z J, Zhang Y F, Wang J, et al. A survey of FPGA design for AI era. J Semicond, 2020, 41: 021402
8 Breiman L. Random forests. Machine Learn, 2001, 45: 5–32
9 Baker S, Scharstein D, Lewis J P, et al. A database and evaluation methodology for optical flow. Int J Comput Vis, 2011, 92: 1–31