FPGA Based Speeded Up Robust Features

Tomáš Krajník, Jan Šváb, and Libor Přeučil

The Gerstner Laboratory for Intelligent Decision Making and Control
Department of Cybernetics, Faculty of Electrical Engineering
Czech Technical University in Prague
Karlovo namesti 13,
121 35 Prague 2,
Czech republic
Email: tkrajnik@labe.felk.cvut.cz
Phone: +420 22435 5754
Fax: +420 22492 3677

Many tasks of mobile robot navigation and mapping are based on image processing. Machine vision algorithms have been investigated for a long time, resulting not only in reactive navigation techniques [1] but also in elaborated mapping and localization approaches [2]. Mapping approaches in unstructured environments often rely on feature extraction algorithms (e.g. Scale Invariant Feature Transformation (SIFT) [3]). These algorithms provide descriptors of significant image areas, which are immune to illumination and camera position changes. One of the most efficient algorithms from this family is Speeded Up Robust Features (SURF) [4], which outperforms current SIFT implementations in speed and robustness. This algorithm has been reported to achieve around 2 frames per second on an ordinary CPU. Implementations of these algorithms utilizing GPUs ([5, 6]) have achieved performance over 30 FPS, which is considered as suitable for real-time applications. Although these implementations are based on an affordable computational hardware, small robotic platforms usually can not afford to carry entire PC system. As a consequence small robots have to use small intelligent cameras, e.g. CMUCam [1] with on-board image processing. These intelligent cameras can rely on standard microcontrollers, digital signal processors or specialized integrated circuits. A popular choice for embedded machine vision is usage of Field Programmable Gate Arrays (FPGA). Several authors [7, 8, 9, 10] reported successful implementation of robotic vision algorithms on a FPGA-based hardware. Among these, some conclude that their SIFT [3] algorithm implementation performs better than on conventional CPUs [10] and some demonstrate their implementation in a real robotic system [9]. We present

[1] http://www.cmucam.org

an implementation of the SURF algorithm massively accelerated by the FPGA logic.

The original SURF algorithm is composed of three stages. In the first stage, interest points are found in the image by a Fast-Hessian Detector. In the second stage, Haar wavelet responses for both x and y directions are calculated around the interest point and the most dominant direction is chosen to achieve rotation invariance. In the last stage, Haar wavelet functions are used to calculate descriptor of interest point's surrounding area. Obtained descriptor is invariant against changes in scale, rotation and brightness. The SURF algorithm contains several optimizations. Most significant improvement in calculation speed is achieved by use of "Integral Image", which allows fast calculation of filter responses used in all previously mentioned stages.

The implementation proposed in this article follows the original definition as closely as possible. Considering nature of FPGA architecture, it is obvious that only some parts of the original SURF algorithm are suitable for implementation using logic. It has been decided to implement the Fast-Hessian detector using FPGA logic and to perform descriptor calculations by ordinary CPU with floating-point accelerator. However, proposed implementation of the Fast-Hessian algorithm applies optimizations, which affect precision of the detector and therefore repeatability and distinguishness of the whole algorithm.

This paper presents a comparison of the herein sketched approach against the original SURF implementation on the same test dataset[2] as the authors of SURF. Moreover, we compare the SURF, GPU-SURF and FPGA-SURF on a dataset of 2000 images captured by a mobile robot moving in a semi-urban environment. Mentioned experiments prove similar repeatability and distinctiveness of SURF and FPGA-SURF descriptors.

A use-case type verification of presented implementation is done by creating an embedded navigation system using a monocular camera vision and FPGA-SURF technology. This navigation system first creates an environment map during a teleoperated drive. After that, the system is able to localize a mobile robot in the environment by matching detected and mapped features. It offers reliable and robust performance in indoor as well as outdoor environments and processes 10 images (1024x768 pixels) per second with a total power consumption of less than 10 W.

References

1. Thorpe, C., Hebert, M., Kanade, T., Shafer, S.: Vision and navigation for the Carnegie-Mellon Navlab. IEEE Transactions on Pattern Analysis and Machine Intelligence **10**(3) (1988) 362 – 373

[2] http://www.robots.ox.ac.uk/~vgg/research/affine/

2. Davison, A.J., Reid, I.D., Molton, N.D., Stasse, O.: MonoSLAM: Real-time single camera SLAM. IEEE Transactions on Pattern Analysis and Machine Intelligence **29**(6) (2007) 1052–1067
3. Se, S., Lowe, D., Little, J.: Vision-based mobile robot localization and mapping using scale-invariant features. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Seoul, Korea (2001) 2051–2058
4. Bay, H., Tuytelaars, T., Gool, L.: Surf: Speeded up robust features. In: Proceedings of the ninth European Conference on Computer Vision. (2006)
5. Cornelis, N., Gool, L.V.: Fast scale invariant feature detection and matching on programmable graphics hardware. In: CVPR 2008 Workshop (June 27th), Anchorage Alaska (2008)
6. Wu, C.: SiftGPU: A GPU implementation of scale invariant feature transform (SIFT). online (2007)
7. Diaz, J., Ros, E., Mota, S., Rodriguez-Gomez, R.: FPGA-based architecture for motion sequence extraction. International Journal of Electronics **94**(5) (2007) 435–450
8. Diaz, J., Ros, E., Pelayo, F., Ortigosa, E., Mota, S.: FPGA-based real-time optical-flow system. IEEE Transactions on Circuits and Systems for Video Technology **16**(2) (2006) 274–279
9. Bonato, V., Marques, E., Constantinides, G.A.: A Parallel Hardware Architecture for Scale and Rotation Invariant Feature Detection. IEEE Transactions on Circuits and Systems for Video Technology **18**(12) (2008) 1703–1712
10. Tornow, M., Kaszubiak, J., Kuhn, R.W., Michaelis, B., Schindler, T.: Hardware approach for real time machine stereo vision. In: WMSCI 2005: 9th World Multi-Conference on Systemics, Cybernetics and Informatics, Orlando, FL, Jul 10-13, 2005, Vol 5. (2005) 111–116