High Speed vSLAM Using System-on-Chip Based Vision

Jörgen Lidholm
Mälardalen University
Västerås, Sweden
jorgen.lidholm@mdh.se

February 28, 2007
1 The ChipVision Project

Within the ChipVision project we are focusing on implementing vision algorithms in reconfigurable hardware called Field Programmable Gate Arrays (FPGA). Computers are getting faster and faster but for real-time image processing computers are today still not fast enough, at least not for the more complex algorithms. We define real-time image processing as "image processing at the frame rate of the camera". Which in standard CMOS cameras is about 30 frames per second.

2 My Research

My focus within the ChipVision project is high speed stereo vision based Simultaneous Localization and Mapping (SLAM) for robots.

Self localization is the most fundamental part in mobile robotics. If a robot does not know where it is, it can not plan and act in a sensible way.

Humans navigate by looking around with our eyes and positioning ourselves in reference to the surrounding objects. For a robot to do this it is necessary to find objects, features or landmarks in the image as references. My research will focus on high speed vSLAM using special purpose hardware, where the goal is to have real-time performance. Testing will be the main research method, algorithms for image feature extraction must evaluated regarding if it can be implemented in hardware for real-time processing. A corner detection algorithm developed by Moravec [1] and further improved by Stephen and Harris [2] is already implemented in our hardware and should be useable. It can today perform at 14 frames per second. Depending on the performance of the corner detector some other algorithms may be evaluated.

The the next step is to match extracted image features between two cameras to be able to position each landmark in euclidean space and put it in a map. There are plenty of algorithms for map representation. Many mapping algorithms have only been tested in offline processing, in my research these can be tested in an online system with the limited resources robots often have. And the hard requirements on speed and quality an online system have.

3 Related work

My research will be in the area of mobile robot navigation using stereo vision. This is a very hot area and the problem can be divided into the sub problems:
landmark detection and navigation. I will start by presenting related work on robot navigation and then on vision landmark detection.

3.1 SLAM

For the robot to plan and act in its environment it needs to know what the environment looks like. For sensing the environment usually ultrasonic-, laser range finders or vision systems, sometimes a combination of these, are used.

The seminal SLAM paper was presented by Smith, et al. [3]. The authors proposed a probabilistic procedure to build a stochastic map and localize the robot in the map [4].

While a robot is moving around it builds a map of the proximity, using an odometry system and range finders to find the proximity. This map can later be used for navigation. Articles handling SLAM often focus on reducing the errors inferred by odometry and time-of-flight range finders. SLAM is a non-trivial but very important problem in mobile robotics.

Some SLAM algorithms presented uses filters to reduce measurement noise and algorithms for matching which are sometimes in the order of $O(N^3)$ in complexity, where $N$ is the number of landmarks. These are not an option for a high speed vSLAM system. Montemerlo, et al. [3] presents a SLAM algorithm which only has a complexity of $O(M \log N)$ where $M$ is the number of particles in a particle filter and $N$ is the total amount of landmarks. Which is a gigantic improvement.

3.2 Multiagent SLAM

Multiagent SLAM is a version of SLAM where several, often very small and simple, robot units cooperates to build a map. Sometimes these robots don’t use odometry but rather measures the distance to each other where one or more of the robots serves as a reference point by standing still. Multiagent SLAM can also be used for parallel mapping to increase speed [6]. This is for example useful when mapping a catastrophe area, which often requires the mapping to be as quick as possible to rescue people in need. Though multiagent SLAM adds complexity, how can the map be built reliably with several robots not sharing the exact same world view. Every individual robot is affected by miss-readings independently. No single robot has a perfect view of the environment.
3.3 Detecting Image Features

In regular SLAM the range from the robot to some object, like walls can be found using proximity sensors. With a camera system for sensing the proximity it is necessary to extract distinctive features from an image as reference landmarks. Just like light-houses are used in nautical navigation. In 1981 Moravec presented his work on stereo matching using a corner detector [1]. Harris and Stephens presented their work on an improved version of Moravec's corner detector in [2].

The corner detection is quite simple and does not give distinctive features. But a given set of corners can be considered to be a distinctive feature. By thresholding, the amount of detected corners can be controlled, more corners gives more information of the current image but also increases the amount of data needed to be processed. This algorithm is not completely invariant to scale. Some corners may be detected in one image scale but not in another.

D. G. Lowe developed an algorithm named Scale Invariant Feature Transform (SIFT) [7] for finding distinctive features in an image. The algorithm is invariant to scale, translation and rotation. This algorithm is unfortunately very computationally intense and tests have shown that it could only run at 2 fps using a Intel PIH 700 MHz [8]. It also requires a lot of data for each feature detected. Though, the algorithm gives very good resulting output.

In [9] the authors implemented the feature extraction part of SIFT in an FPGA and reduced the calculation time required for the extraction from 600 to 60 ms. This also shows the potential of special purpose hardware.

In [10] Bay, et al. presents an algorithm called Speeded Up Robust Features (SURF), which is similar to SIFT. SURF requires approximately one third of the computational time SIFT requires [11]. With the same quality on the result.

Sure SIFT and SURF are very complex, but the current trend is going towards these kind of feature detectors for vSLAM.

3.4 vSLAM

Visual SLAM is probably the most interesting SLAM problem. Vision can provide so much information of the environment. For example the SIFT and SURF algorithms provide a descriptor for each extracted feature containing a lot of information about the object like position and rotation. This helps dropping the need for an odometry system. In [12], Davison presents a real-time vSLAM algorithm using a singel b/w camera and a desktop computer which does not need an odometry system. A SLAM system which uses range
finders has so many error prone sources that it needs odometry to improve quality, but still it is common with accumulated errors from the odometry which makes the robot loses track of its position after a longer time. This can be dismissed with a high speed system like Davison presents.

4 The industry

On the market today there are a few robot products which use vision systems for navigation like Honda’s ASIMO, Sony’s AIBO and others.

ASIMO, named after the science fiction writer Isaac Asimov [13], can detect moving objects like people and stationary objects and obstacles like tables and stairs using a vision system. ASIMO can even recognize peoples faces.

The AIBO robot dog, unfortunately discontinued by Sony, is a very interesting robot. A highly embedded system with advanced features like a simple camera running SIFT.

Evolution Robotics Inc. has developed SIFT based vSLAM algorithm which they use in their products [14, 15].

CMOS cameras is becoming cheaper and cheaper. Today you can buy a 5 Megapixel camera capable of up to 96 fps, depending on resolution, for $36. So the technology to do amazing things exists and it is cheap.

5 Key Conferences and Leading Research Groups

Keith Price maintains an excellent web page listing over 70’000 scientific papers on computer vision [16]. He also lists research groups in the computer vision area.

My research area is covered by both vision and robotics conferences. The robotics conferences also tend to have a wide interest area, since robotics covers so many research areas from mechanics to vision to AI.

5.1 Robotics Conferences

The International Conference on Intelligent Robots and Systems (IROS) is one of the bigger robotics conferences with a broad area of interest, for example:

- Evolutionary Robotics
- Humanoid Robotics
• Robot Surgery
• Autonomous Vehicles
• Underwater Robots
• Search and Rescue Robots
• Entertainment Robots

All highly related to vision [17].

The International Conference on Robotics and Automation (ICRA) is also one of the bigger conferences in robotics. The 2007 year conference will for example cover:

• Automation
• Cognitive robotics
• Field and service robotics
• Human-Centered and life-like robotics
• Mechanics, design and control
• Mobile and distributed robotics
• Sensing and perception

Where some of the keywords are computer vision, SLAM, localization, mapping and more [18].

In ICRA 2008 the theme of the conference will be Human-Centered Robotics, technology which aids humans in their everyday life [19].

5.2 Vision related conferences

There are several vision conferences with a variety of foci. The conferences CVPR and ICCV [20, 21] have identical lists of topical areas. These have high relevance to my research, and have for example the following points in the topics list:

• Sensors and Early Vision
• Motion and Tracking
• Stereo and Structure from Motion
• Object Recognition
• Performance Evaluation

At the bi-annual European Conference on Computer Vision (ECCV) [22] in 2006 they had a similar agenda to the one they have at CVPR and ICCV. But for ECCV 2006 they explicitly specified "Active and Robot Vision".

5.3 Leading Research Groups

At the University of British Columbia, the Laboratory for Computational Intelligence (LCI) is a very successful research team, with people like David Lowe, (the father of SIFT) and Jim Little. The focus of this research group is artificial intelligence in robotics.

At the AI lab at Stanford University, there is a Robotics and Machine Learning group lead by Sebastian Thrun which are prominent in the area of SLAM. And many of Thrun’s algorithms define the state of the art in robotics perception and control.

I would also like to mention the research team at Evolution Robotics. Evolution Robotics is a company with close connection to the research world and thereby having bleeding edge technology.

References


