Procedure for Efficient Depth Estimation With Stereo Event Cameras

Graduate Candidate

Steven Wegmann

Introduction: Depth estimation is an important topic in computer vision applications. Humans perceive depth by the slight horizontal difference between both eyes when looking at an object. This difference is called disparity. It is inversely proportional to the distance of the object. The same principle can be applied with two cameras each simulating a human eye. The disparity can be calculated by finding the same point of interest between these two cameras. Conventional cameras have proven to be a viable and cost-efficient method to solve these stereo-matching problems, but they do not give good results in certain scene properties.

New milestones in event cameras provide a novel approach to solve these problems. These cameras are in possession of a special pixel architecture that allows each pixel to asynchronously respond to lighting changes. Each pixel can generate an event if this change reaches a certain adjustable threshold. Event cameras provide a sparse amount of relevant data at object borders.

The goal of this thesis is twofold. To evaluate the compatibility of event cameras with embedded systems which are limited in computational power and to implement event-based depth estimation that strives towards a good trade-off between accuracy and speed.

Approach: Many existing algorithms only utilize the most recent event at every pixel location. This completely disregards the high time resolution event cameras are capable of. Most methods with high accuracy usually need to consider a large spatial neighborhood around each pixel in a costly operation. Algorithms that match individual events often lead to inaccurate results in more complex scenes.

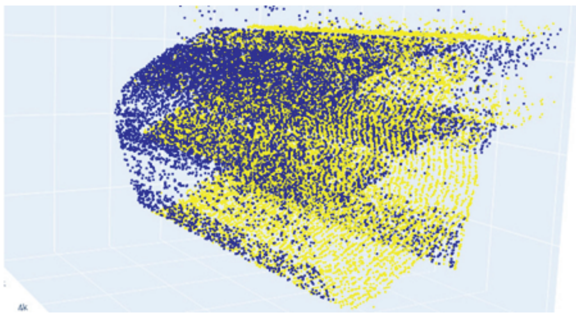
The proposed method matches pixels between cameras by utilizing past events in a certain time frame. Downsampling the event stream into coarser blocks of data decreases the amount of individual matching procedures and adds an additional parameter in how different regions of data can be compared.

Conclusion: The sparse event stream removes a lot of redundant data, which decreases the amount of information an algorithm has to process. On the other hand, this sparsity may not be able to provide enough information to differentiate image regions with similar event generation. These regions require costly algorithms not feasible for embedded systems.

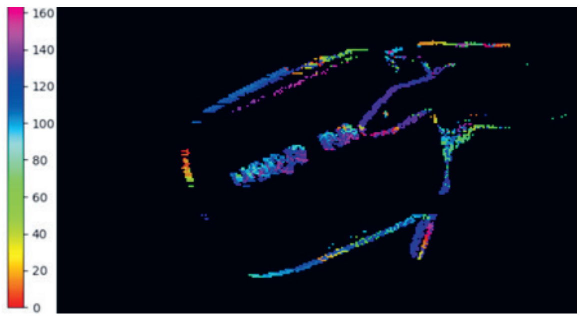
The proposed method provides a trade-off between accuracy and speed with an additional reduction in spatial resolution. Accuracy may be improved by using more sophisticated methods for calculating similarity. The disparity map could also be improved by removing disparity outliers in a post-processing step. Speed improvements could certainly be made by optimizing the code implementation. Future work may also look into improving matching accuracy by adjusting the event generation itself.

Events generated with an event camera.
Own presentment

A time frame of events in three-dimensional space.
Own presentment



Disparity map created by the proposed method.
Own presentment



Advisor
Prof. Dr. Christian Werner

Co-Examiner
Michael Hvelplund Schneider, Premium Design Group AG, Altendorf, SZ

Subject Area
Embedded Software Engineering