Online stereo camera calibration in embedded processing context.

Michal SZCZEPANSKI[1], Mehdi DAROUICH[1], Erwan PIRIOU[1], Frédéric CHAUSSE[2]

[1]CEA, LIST, Algorithm and Architecture Codesign Laboratory - 91191 Gif-sur-Yvette CEDEX, France
[2]Institut Pascal - A4 impasse Blaise Pascal-63178 Aubire Cedex , France

firstname.lastname@cea.fr, frederic.CHAUSSE@univ-bpclermont.fr

Résumé – Cet article décrit le travail en cours sur la conception d'une nouvelle approche de calibration en ligne de caméra en contexte embarqué. L'objectif de ce travail est de proposer un système capable de fonctionner en temps réel, et de se recalibrer en utilisant les données de la scène (sans motifs contrôlés). Les éléments utilisés pour la calibration sont extraits du flux de données générées pour l'application principale, afin de limiter son impact sur le temps d'exécution.

Abstract – This paper describes the work in progress on the design of a new approach of online method for camera calibration on an embedded system. The goal is to propose a system able to work in real time and to autocalibrate using data from the scene (without controlled patterns). The features used by the calibration process are extracted from the application pipeline to limit its impact on the global execution time.

1 Introduction and context

Nowadays one of the main challenge for researchers and engineers is to create modern electronic systems like robot, autonomous vehicles, smart glasses that are more independent and smarter. To reach this goal, those systems have to be able to interact with the surrounding world. Analyse and understanding the environment may be based on images provided by cameras that can replace most of costly sensors. Present-day visual information delivers color wide spectrum of data, which can enhance the perception, improves safety and precision of many tasks. The use of a set of at least 2 cameras, stereo cameras, allows to extract depth information from the scene and delivers important data for localisation and exploration in real time. This has found use in many applications like : automotive, robotics, entertainment and virtual reality. The challenges of these modern applications are real time computation, portability, higher reliability, lower power, need for efficient calculations and performance, especially for embedded solutions. We can diversify the various applications due to objective but many of them is based on the similar low-level functions. The driver assistant system, navigation of the mobile robot, a sand-blind assistant on the head-set or a head component for game accessory, all of these high-level applications may be based on the same basic functions as feature (corners) and objects detection, localization and recognition. For automotive applications it may be the road infrastructure like traffic sign, lights, for applications working into a building it can be the equipment elements like doors. For each of these applications the constraints are the same, the high accuracy and reliability during real-time operation are very important. Therefore these applications require a permanent calibration capability of the stereo vision system. It has to be flexible, without initial processes, no overhead which can be seen as separate module in robot operator system and fit for embedded constraints. The online stereo camera calibration provides the relation between camera's position, it is the critical problem during stereo vision processing. Nowadays accurate calibration parameters are found offline by constructors during manufacturing and delivered with equipment. Most of time the cameras in stereo vision platform are connected very rigidly into the metal boxes. This solution raises price and makes the solution less profitable. Loosely attached cameras may reduce cost, however, in this case, we cannot rely on calibration made a single time because of the drift appearing in case of long-term usage. Even the most accurate initial parameters can change, influenced by many factors, for example the engine vibration, the changing weather conditions or the mechanical stress. For many applications wrong calibration parameters will cause computing error and lead to mistakes and dangerous situations.

The purpose of our study is to provide a real time online continuous stereo camera calibration in the context of embedded systems. We strongly believe that this would raise the attractiveness and reduce cost of platform with stereo vision cameras. The next section presents the state of art of calibration methods. Then we present the proposed approach. Finally, we present the results of our test.

2 State of art

Camera calibration is a necessary step of stereo vision to extract correct 3D information from 2D images. In the last decade, researchers have been studying different calibration methods of extrinsic camera parameters which characterize the transformation between the unknown cameras reference frames and the known world reference frame. This is referred to translation vector between the relative position of the origins of the two cameras and rotation matrix that can cover the corresponding axes of the two frames into alignment. There are many different ways to estimate these parameters which we can divide into two main groups : offline and online methods. Most work consider the stereo rigs, as two cameras which provide overlapping views. It means that cameras simultaneously observe mostly the same parts of a scene. In this chapter some interesting approaches will be presented.

The most popular calibration procedures are based on offline methods that make use of known special patterns the chessboards, viewed in two cameras from different positions. Many work presents these kind of methods which are based on algebraic projective geometry. These are well known, tested by community and give accurate and stable results. The advantage over other methods is that it simultaneously provides intrinsic and extrinsic parameters. However, it implies the use of pattern in front of the cameras, which is a constraint when dealing with real time applications.

Researchers tried to get rid of calibration patterns and make the process of calibration in a stable unknown scene. This group is known as self-calibration methods and some of them are based on epipolar geometry constraints and rely on eight correspondences between the same points in a pair of uncalibrated images [2], [1], [4]. Good recognition, perfect matching and the initial estimation of the parameters are necessary. The singular value decomposition or linear methods are used to estimate the relative transformation between cameras however the initial estimation is required to estimate scale thereby proper interpretation results. This method is simple and fast which would fit to embedded constraints. However it is sensitive to noise in input data. Another subgroup of self-calibration method based on local bundle adjustment which minimizes the reprojection error between the image location of observed and predicted image points on several consecutive frame. Warren [6] has shown ability to compute camera parameters with high precision, however in the embedded context considering the amount calculation needed, the use of this kind of optimisation method may be incompatible with the real time constraint.

Another interesting approach is to combine camera offline calibration and self-calibration methods. These methods try use the static landmarks from environment as patterns. Xu's [5] proposes a stereo self-calibration method which takes on-road marks cross walk corners as the feature points. Another interesting example is Carrera's [3] work where calibration is done after the combine data from monocular Simultaneous Localization and Mapping (SLAM) with local and global bundle adjustment. This method provides more accurate calibration results than standard self-calibration algorithms. However this method implies complex optimization method and heavily depends on scenes with high-quality textures and set of particular motion during calibration.

Considering the application domain our target calibration method has to be online without special patterns (to be applicable everywhere) and compatible with embedded systems. The subgroup of self-calibration methods which is based on epipolar geometry seems to fit the constraint of low processing power. After the analyse of the state of art we choose the eight point algorithm with normalization proposed by Hartley [1]. This algorithm seems to be efficient and need small amount of calculations. However this algorithm has some drawbacks described in literature : it is considered as very sensitive and requires the scale factor from other source. To tackle sensitivity problem we will use an iterative RANSAC method to filter input data.

3 Approach

In this section we present our approach on one particular example, the pedestrian navigation on smart glasses equipped with stereo cameras. The purpose of this application is to navigate and lead the operator according to scenario, in a city or in a building with ability to recognize and avoid potential hazards or obstacles. The complexity and precision of this application demands for accurate information provided by set of well calibrated cameras attached to the eyeglasses frames. However, the construction of the glasses can not be a metal rigid cage and the position of the camera could change at any time. Thus, the system requires continuous monitoring of stereo camera calibration. Moreover, as it is targeted to be embedded on a small device as a Cortex A9 processor or a Raspberry Pi board, limited processing power and energy are available. Therefore the online camera calibration task should imply a limited processing overhead, and a controlled execution frequency.

The whole processing pipeline presented in the figure 1 is divided into 3 blocks : the low-level functions, high-level application and calibration part. The first part, low-level functions are scene analysis task, that can be performed on each monocular video stream separately and thus one not impacted in case of uncalibration.

In our study case, it is composed at least of Point of Interest (POI) detection, description and temporal matching and basic object detection (traffic sign). A stereo matching is performed between POI of both streams afterwards. It will be used by both high level SLAM and calibration process. On the other hand, the high-level application takes as the input two video streams and data processed by both monocular blocks. A vision-based odometry [10] function is implemented as high level application to provides navigation data to the user it takes tracked POIs as input and generates a trajectory. Knowing the calibration, input video streams are rectified to respect epipolar constraint and processed to generate a depth map. To enhance this functionality, a traffic sign recognition [8] and will be used to detect obstacles. The geometrical dependence between these two streams known as stereo-calibration parameters are mandatory to obtain correct results.

The calibration block is responsible for the online continuous calibration and the monitoring of the parameters. The purpose of this block is to verify that the calibration is correct and if it is not the case, provide a new one. To avoid increasing the computing power, our approach will use the same input data as high-level application. The calibration processing is based on the eight point algorithm presented in state of art. Input data are selected from stereo matched POIs delivered from low-level function. To choose the best 8 POIs among the set of matched POIs provided, a RANSAC method is used. This iterative method allow us determine the inliers among the data

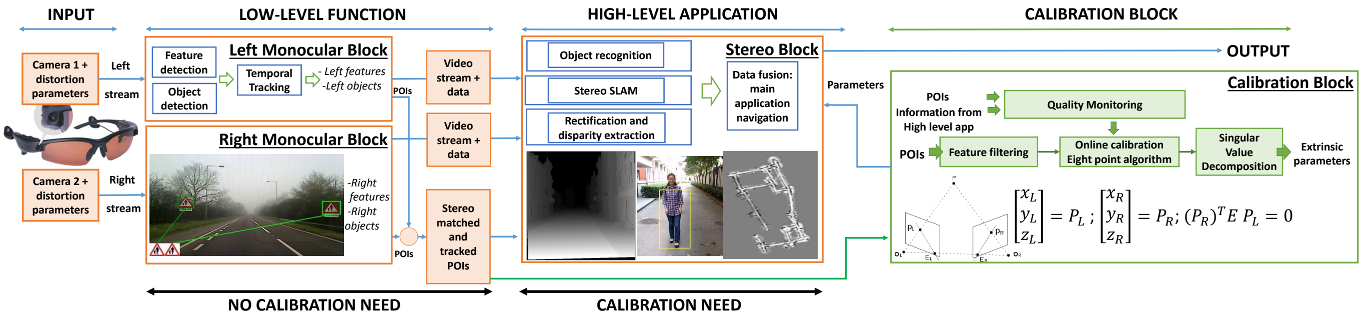


FIGURE 1 – Pipeline of approach.

extracted in the image. The current model estimated by algorithm is continuously monitored by fulfilling the geometrical constraints condition. If a certain amount of POIs do not fill this condition, the system considers a new calibration is necessary. The best model calculated by eight point algorithm from available points will be converted by singular value decomposition to new parameters. However to generate complete set of parameters the scale information is necessary. One solution we plan to use is to extract this information from detected object in the scene from which we know the real size. A simpler option, implemented in the following experiments is to derive the scale from the measured distance between the cameras (assuming it will not vary). Before performing the calibration, a filtering pass is done to select a set of the most valuable POIs from the high quantity of matched POIs delivered by the low-level function. This filtering may be based on the temporal stability of the POIs or on the fact that they are detected on known objects in the scene.

4 Results

In this section we present the results of calibration in term of quality and processing time based on the StLucia [7] dataset. Firstly we tested our eight point algorithm on corner points detected from the chessboard pattern presented in StLucia database (fig 2 originally for offline calibration). The left part of figure 3a presents how we compare our approach with the offline method implemented in the Matlab *Stereo Calibration App* [9], in term of precision. The purpose of this test is to minimize the effect of wrong points matching and detection by delivering 40 correct pairs of points by images pairs and the scale factor. All the points from several pairs of images are common for both methods, and seen as will be displayed at the same image. During these tests we run four different datasets which contain 4, 5, 8 or 10 images with different configuration of structure which allows us to evaluate and analyse the impact of distinct arrangement and number of points in a scene. Additionally to take into account the randomness introduced by RANSAC, we run an algorithm ten times for each group. On the figure 4 we show results of the tests where the median of the errors for all three axis are presented. Each of them is calculated as the subtraction between the initial value delivered by StLucia dataset and the parameters calculated by offline and online method. We can see that with a sufficient amount of perfect POIs, at least 5 group of chessboard, which means 200 perfect POIs, we obtain the same level of precision as offline method in two axes. The figure 4a presents the error in X which is smaller than 0.01%. The fig 4b shows error in Y parameters which is equal around 10%. These higher differences are due to construction of stereo system of cameras where distance around X axis is much bigger than the two other axis. The error rate on the Z axis (4c) is much higher than for the two other axis for both offline and online method. This is due to the fact that the axis is along the optical axis of the camera. As a results of using several different dataset, we found that those with high rotation of planar structures and high distribution around whole image gave much more accurate results. That is why the good distribution and view from different rotated planar structure is a key to choose the POIs to inject in the calibration pipeline.

The second experiment is performed on the real dataset, without any pattern. The right part of figure 2 presents examples of views from the database. In this test, we do not compare results to offline method because the offline method will not work without patterns. We compare the obtained parameters with the stereobase delivered from references. We assume that it does not vary so we try to find back the stereo parameters and compute error based on initial calibration. In each pair of image, there are around 200 pairs of points detected and matched. Like in the previous test, all data coming from several pairs of images are injected as all points will be seen in the same image. The different number of images 6,7,15 and 50 allow us evaluate and analyse the impact of the number of POIs vs. the level of noise in data. As in the previous test, in order to account the impact of randomness we run ten times each database. From the results presented on figure 4, we can see that due to construction of stereo camera we can still calculate close results in X parameter (fig 4d). We start to loose accuracy in Y (4e) where the error is between 20% and 140%depending on the number of injected POIs, higher than in a previous test. The interesting results shows Z axis where the calibration is able to find better results on real dataset than previous test. However we can conclude that increasing the number of images without patterns cause a higher number of outliers which lead to higher error rate.

The results of the last test presented on the figure 5 are about the execution time of one iteration of algorithm and the whole model calculation. To calculate the parameters, the eight point

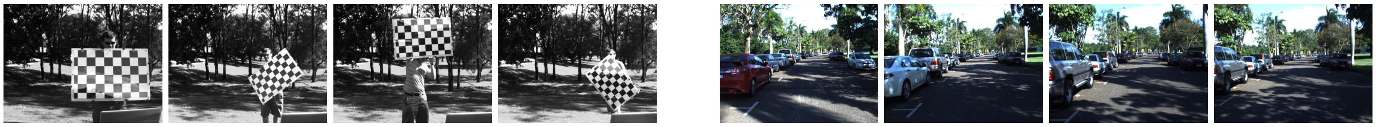
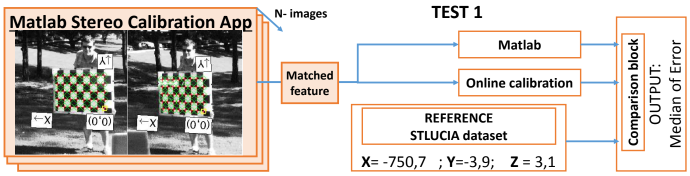
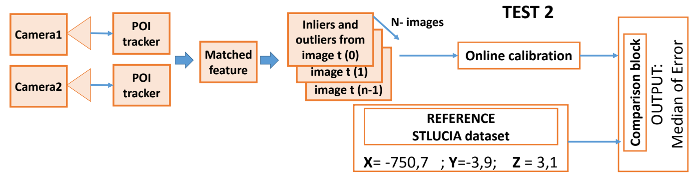


FIGURE 2 – Two database left with perfect point, right with real data.

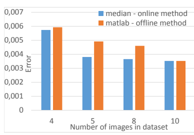


(a) Online and offline method comparison on perfect features.

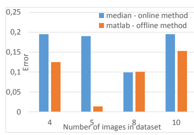


(b) Online method on real data.

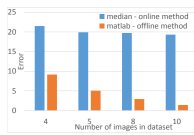
FIGURE 3 – Pipeline of test approach.



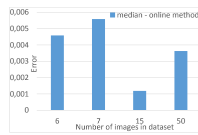
(a) Error X.



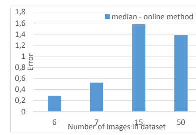
(b) Error Y.



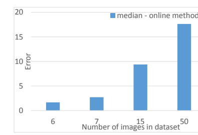
(c) Error Z.



(d) Error X.



(e) Error Y.



(f) Error Z.

FIGURE 4 – Results for both dataset. (a-c : Test 1 ; d-f : Test 2)

Number of points		Time in [ms]		Number of iteration	Precision of results
whole	inliers	one iteration(median)	whole model estimation		
1159	795	0,044	84,221	442	Good
1797	1317	0,028	9,122	87	Good
4005	2763	0,035	14,615	101	Good
12732	7887	0,031	116,983	372	Good

FIGURE 5 – Execution Time and number of POIs for several images.

algorithm uses non-deterministic RANSAC method. Due to this the time spent each time on each estimation is different. We did test to show the time required to calculate the model relative to the number of iteration and points. Lower number of iterations allows us to calculate the model faster. A lower number of points has impact on time needed to calculate the single iteration, however less points lead us to a higher instability and a lower quality of results. For the moment the execution times have been measured on a standard PC, but in further study will be performed on an embedded processor.

5 Conclusion and Future Work

This paper presents the work in progress on an online calibration method in real-time embedded context. We choose the 8 point algorithm, a well-know and rather light method from the state of art. Our approach is based on the reuse of data extracted for the main application to process the calibration during execution with a limited overhead in processing power. The experiments shows that injecting aggregated matched features is not enough to extract precise extrinsic parameters. However, the tests performed with controlled structures (patterns) show that injecting features on well known structures may provide a precision close to offline methods. In real application, structures like traffic signs or equipment would be used for this purpose. A better filtering of the input POIs will be performed to increase the calibration accuracy and stability. First results shown, this is not trivial problem for both groups of method : the online and offline. We would like to increase the precision as close to offline method as it is possible. Considering this problem we have to explore and study the influence of the precision of calibration on the output of the stereo high-level application. This investigation should allow us to determine the level of precision we need to achieve. In parallel to these studies, a deeper characterization of the execution time and memory consumption of the calibration process will be performed on embedded processor. Strategies of recalibration will be studied to lower impact of this process on the overall application execution time. Our ambition is to provide new view at the problem of calibration and allows on the faster and wider spread of stereo cameras systems.

Références

[1] R. Hartley, A. Zisserman *Multiple View Geometry in Computer Vision*. 2003

[2] F. Luong, Quang-Tuan, et al. *Self-calibration of a stereo rig from unknown camera motions and point correspondences*. 1993

[3] G. Carrera, A. Angeli et al. *SLAM-based automatic extrinsic calibration of a multi-camera rig*. 2011

[4] T. Gee et al. *Tsai camera calibration enhanced*. 2015

[5] D. Xu, Q. Zeng et al. *Online stereovision calibration using on-road markings*. 2014

[6] M. Warren, D. McKinnon et al. *Online calibration of stereo rigs for long-term autonomy*. 2013

[7] M. Warren, D. McKinnon et al. *Unaided stereo vision based pose estimation*. 2011

[8] J. Greenhalgh, M. Mirmehdi *Real Time Detection and Recognition of Road Traffic Signs*. 2012

[9] The MathWorks, Inc. *Stereo Calibration App*.

[10] F. Fraundorfer, D. Scaramuzza *Visual Odometry : Matching, Robustness, Optimization, and Applications* 2012