- APROVIS3D -

Analog PROcessing of bioinspired VIsion Sensors for 3D reconstruction

Document Reference:		
Title: **D2.2.1 Stereo System with 2 standard DVS sensors designed**		
Contractor: **UCA**		
Prepared by: **Jean Martinet**		
Document Type: **Deliverable**		
Version: **1**		Pages: **8**
Classification: **External document**		

Document Track

Version	Date	Remarks and Authors
1.0	15/01/2022	First draft (J. Martinet - UCA)
1.1	31/03/2022	Final version (J. Martinet - UCA)

Authors

	Role / Function	Name	Organisation
Prepared by	Project Coordinator/WP0L	J. Martinet	UCA
Checked by	Quality check	J. Martinet	UCA
Released by	Project Coordinator/WP0L	J. Martinet	UCA
Approved by	Project Coordinator/WP0L	J. Martinet	UCA

TABLE OF CONTENTS

FIGURES

TABLES

No table

1 Introduction

During the implementation of the APROVIS3D project, WP2 targets the sensing with event sensors. Task 2.2 more specifically focuses on stereo sensing.

1.1 Purpose

This document D2.2.1 deliverable of the APROVIS3D project describes the technical activities related to designing a stereo-DVS sensing system.

2 Documentation

2.1 Applicable and Referenced Documents

#	Id	Description	Identifier (Ed Rev)	Date
AD1	FPP	Full Project Proposal	1.0	15.01.2019

2.2 Glossary and Terminology

Acronym	Definition
CMOS	Complementary Metal-oxite-semiconductor
APS	Active Pixel Sensor
RGB-D	Red-Green-Blue-Depth
ROS	Robot Operating System
CSV	Comma-Separated Value
UAV	Unmanned Aerial Vehicle

3 Stereo sensing system designed

3.1 Stereovision

Stereo vision is the computation of depth based on the binocular disparity between the images of an object in left and right eyes. The 3D information can be obtained from a pair of images, also known as a stereo pair, by estimating the relative depth of points in the scene. These estimates are represented in a stereo disparity map, which is constructed by matching corresponding points in the stereo pair. The disparity is computed as:

where x and x_i are the distance between points in the image plane corresponding to the 3D scene point and their camera center. B is the baseline that represents the distance between two cameras, Z is the depth, and f is the focal length of the camera (see Figure 1).

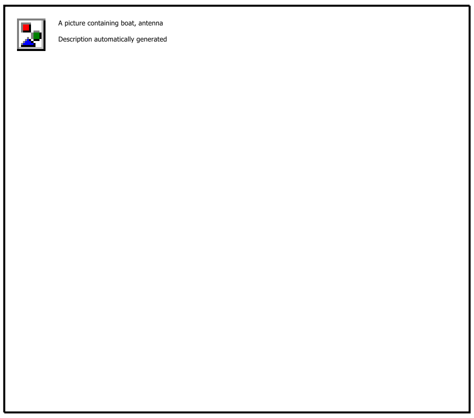

Figure 1: Stereo vision geometry

Stereo matching or the “correspondence problem” is the problem of finding a pair of corresponding pixels or patches in a set of stereo images. Having the matching objects or pixels, we can calculate the depth which is inversely proportional to the disparity—the displacement along the epipolar lines—between the matches.

3.2 Hardware description

For this first setup, we have used a pair of commercial cameras by Prophesee. The pair of cameras purchased from Prophesee (Figure 2) is a CSD3SVCD sensor belongs to Prophesee’s 3rd generation product family. They provide the support for real-time viewing and recording of events stream captured by the sensor. The evaluation kit features a VGA resolution of 640x480 pixels, a wide dynamic range up to

120dB, lens D-FOV 70° C mount (ref. VSTechnology SV-0813V), with power supply and data exchange with standard USB 3.0 interface, and event timestamping with microsecond (µs) precision.

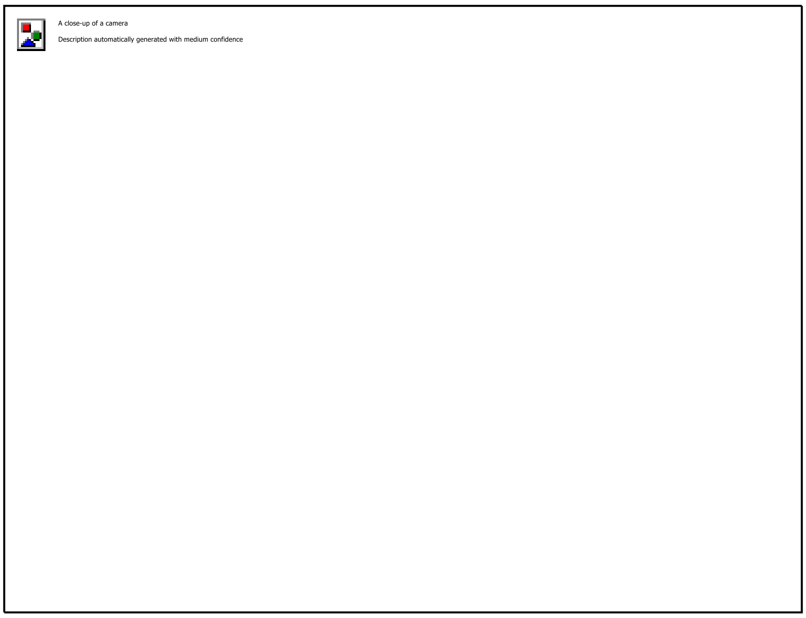

Figure 2: Prophesee Event Cameras. Here the cameras are mounted on a technical support in a stereo configuration mode.

Each camera has a pixel array containing 640 x 480 event-driven pixels CMOS vision sensor with a pixel pitch (distance in millimeters from the center of a pixel to the center of the adjacent pixel) of 15um, so a pixel has a size of 15um x 15um. The cameras have been mounted on a technical tripods, calibrated and used to record test datasets.

After the cameras were plugged in (drivers and software installations) on Ubuntu Xenial 16.04, they were calibrated using Prophesee libraries. The full calibration of the cameras consists in two steps: the intrinsic of both devices, and successively the extrinsic calibration of the two cameras synchronized together.

3.3 Mono-calibration (intrinsic)

The goal of this type of calibration is to compute the instrinsics camera parameters. Generally, these parameters can be divided in two parts: the camera matrix (Equation A.1) and the distortion parameters. The camera matrix is described using the pinhole camera model, that describes the mathematical relationship between the coordinates of a point in three-dimensional space and its projection onto the image plane of an ideal pinhole camera, where the camera aperture is described as a point and no lenses are used to focus light.

The intrinsic calibration should not move too much during the lifetime of the camera, given that the optics stay untouched. If one modifies the aperture or focus, recalibration might be needed.

To do the calibration, a blinking chessboard has been used in front of the camera, paying attention to use different position in terms of angulation and depth. For a correct calibration between 40 and 50 linking point frames are required, as displayed in Figure 3.

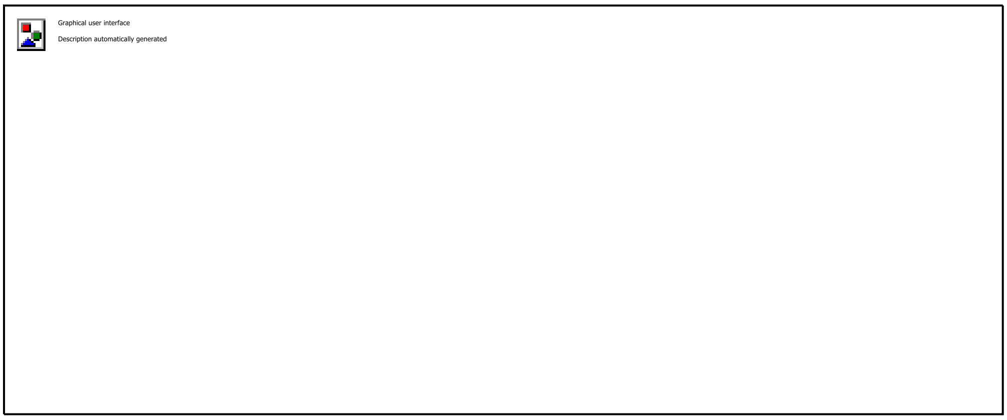

Figure 3: Monocalibration of an event camera. On the left, we can see what is captured by the camera, and on the right, the linking points captured by the system. These points are converted from the system into information to compute the calibration.

3.4 Stereo-calibration (extrinsic)

The goal of the extrinsic calibration is to compute the geometric transformation (rotation and translation) between the pair of cameras. The principle of this type of calibration process is to move a target, blinking chessboard in our case, in front of the two cameras; when the target is detected in both images, a relative pose between the two cameras can be computed. During multicamera acquisition, data streams must be synchronized by a synchronization cable to allow correct processing (Figure 4).

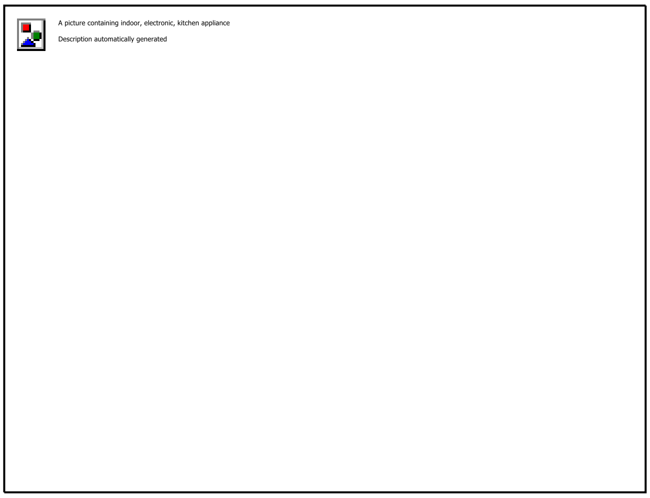

Figure 4: Synchronization cable for the extrinsic calibration. Each pin has a particular function, so it is important to link the correct ones to make the cameras working correctly.

During the calibration and the future recording, one camera it will be the master and the other one the slave. Both camera clocks are now sharing the same clock and start simultaneously insuring synchronous timestamping.

3.5 Recording data and access the events

To record data using our stereo system, the record scripts and the calibration files have been used obtaining in output a file .raw. These .raw files are translated in post processing extension .dat (a tool that allowed to record directly .dat was deprecated because of the extremely large size of the extension).

At this point, for what concerns the mono-recording and stereo-recording, we are allowed to access the events by a Matlab script provided by Prophesee that creates in output a data structure having all the events information needed. This structure can be then saved as Hdf5 file or other formats. Instead, being the .dat output structure different in stereo-recording with disparity information, the same script cannot be used, so after an analysis of the data structure, we implemented a parser (not provided by Prophesee).

In details, we selected a convenient data format, ensuring that we had a little-endian ordering, then we created a Hex conversion to implement the correct parser. This parser was then added to Aertb (https://github.com/rfma23/aertb), a tool that allows us to load events from files with different event camera file extensions into Python.