Project acronym: POPART Project number: 644874 Work package: Management Deliverable number and name: D6.2: Quarterly Management Report 1 and 2	Title: Quarterly Management Report 1 Work Package: WP6 Version: 1 Date: April 1, 2015 Author: Håvard Espeland
Type: [X] Report [ ] Demonstrator, pilot, prototype [ ] Website, patent filings, videos, etc. [ ] Other	Co-Author(s): All partners To: Albert Gauthier, Project Officer
Status: [] Draft [ ] To be reviewed [ ] Proposal [X] Final / Released to EC	Confidentiality: [X] PU – Public [ ] CO – Confidential [ ] CL - Classified
Revision: Final	
Contents: Deliverable 6.2: Quarterly Management Report 1	

Human Resources

SRL

The project is coordinated by Håvard Espeland. Carsten Griwodz, Lilian Calvet contribute to the innovation and scientific work. Elisabeth Andersen contributes to the administrative work.

LAB

The project at LAB is led by Magne Eimot. Jon M. Puntervold, Gunleik Groven and Håvard Espeland also contribute to the innovation work. In addition, Zeljko Vrba will join LAB as a senior developer from May 1. and Kristian Skarseth as a developer from June 1.

MIK

The project at MIK is led by Benoit Maujean, R&D Manager at Mikros Image. Fabien Castan is the technical manager of Popart project; he is responsible for the 3D reconstruction library (Openvg) and photomodeling plugin for Autodesk Maya (MayaMVG) at Mikros Image. Elisa Prana, Nicolas Rondaud and Cyril Pichard, as software engineers contribute to the development of Popart tools, including the expected adaptations of OpenMVG and MayaMVG.

IRT

The project at IRT is led by Vincent Charvillat. Simone Gasparini is responsible for the Workpackage 1 and he's coordinating the tasks assigned to IRT, in particular on camera localization and the contribution to the OpenMVG libraries. Sylvie Chambon is also a staff member of the project and she will participate in the task involving the image matching for improving the performances of OpenMVG. Clement Aymard is a developer engineer who is developing the different contributions to the project.

BAN

The project at BAN is led by Christopher Hantel, CEO at Band Pro Munich. He is joined by Randy Wedick, contributing to the commercialization and productization of the POPART commercial product.

Person Months Contributed to the Project

Participant	Names of staff	WP1		WP2		WP3		WP4		WP5		WP6		Total	
		Planned	Actual	Planned	Actual	Planned	Actual	Planned	Actual	Planned	Actual	Planned	Actual	Planned	Actual
SRL	Håvard Espeland Lilian Calvet Carsten Griwodz Elisabeth Andersen	1.5	1,45	5.18	1,89	0	0	0	0	0	0	1.5	1,56	8.18	4,9
LAB	Magne Eimot Jon M. Puntervold Håvard Espeland Gunleik Groven	0	0	0	0	2.2	2.68	3.70	5.56	0	0	0	0	5.9	8.24
MIK	Benoit Maujean Fabien Castan Elisa Prana Nicolas Rondaud Cyril Pichard	1.13	0.27	0.55	0	3.9	3.60	2.20	2.45	0	0	0	0	7.77	6.32
IRT	Simone Gasparini Vincent Charvillat Clement Aymard	4.63	3.75	1.64	0	0	0	0.6	0	0	0	0	0	6.86	3.75
BAN	Randy Wedick Christopher Hantel	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total		**7,26**	**5,47**	**7.37**	**1,89**	**6.10**	**6,28**	**6,50**	**8,01**	**0**	**0**	**1.5**	**1,56**	**28.86**	**23,21**

Figure 1: The cumulative budgeted and contributed person-months since project start for each partner sorted by Work Package.

Project Progress

The project is now running as planned and we are working on all tasks expected implement POPART and its objectives. Details on each work package is given below.

WP1 - 3D Visual Database

Initial 3D visual database

IRT has worked on the first preliminary version of the visual database combining the 3D reconstructed points and their visual representation, namely the SIFT features. This version of the visual database is created in an off-line step from a collection of views of the scene. These views are processed by a structure-from-motion pipeline (OpenMVG) in order to calibrate the cameras and recover the relative poses of the cameras and the 3D reconstruction of the scene in terms of points cloud.

The visual database is composed of a tree structure, namely a vocabulary tree as described in the seminal work of [Nistér 2006]. This approach is normally used in image retrieval tasks, in order to find the images in a dataset that are visually similar to a given query image. In the case of image localization the query image is the current camera frame and the goal is to find the visually closest view among the set of initial views of the scene.

The initial 3d visual database deliverable has been submitted as D1.1. The source code is available at the following Github repository: https://github.com/poparteu/cameraLocalization

OpenMVG

IRT has begun to investigate possible strategies to improve the sequential reconstruction method, in particular considering two directions. A first contribution will be the use of a visual word approach to ease the burden of the matching among the images of the dataset. For that, approaches based on visual words and vocabulary tree will be experimented and test in the next months. A second contribution will be the implementation of an actual incremental method in which new images are added in order to update an existing reconstruction. Again, the approach will rely on visual word and vocabulary tree approaches in order to detect for each new image the set of matching images in the dataset, compute the pose of the new image and possibly update the 3D reconstruction with new 3D points.

Global SfM method
The target date for the global SfM method (T1.4) is month 12.
Regarding the tests done by MIK to compare the global vs incremental method, there is a huge performance improvement for a solution that “looks” better. Currently it’s not possible to really compare the accuracy, but we will be able to do it soon with the virtual dataset (T.4.3). The missing point to use it in production is the lack of support for radial distortion.
SRL has done a contribution to the OpenMVG library to support radial distortion in the bundle adjustment of the global method. (cf. pull request #236). In fact, it has been shown that the construction of the epipolar graph (global method) may result to be severely flawed as the geometric filtering step does not take into account the radial distortion within the camera model.
The last step to use it in production, is to provide an initial value for the radial distortion. MIK is working on the integration of lensfun library to get an estimation of the radial distortion based on the focal lens provided in the image metadatas.

WP2 - Real-time camera tracking

GPU artificial feature detection (SRL)

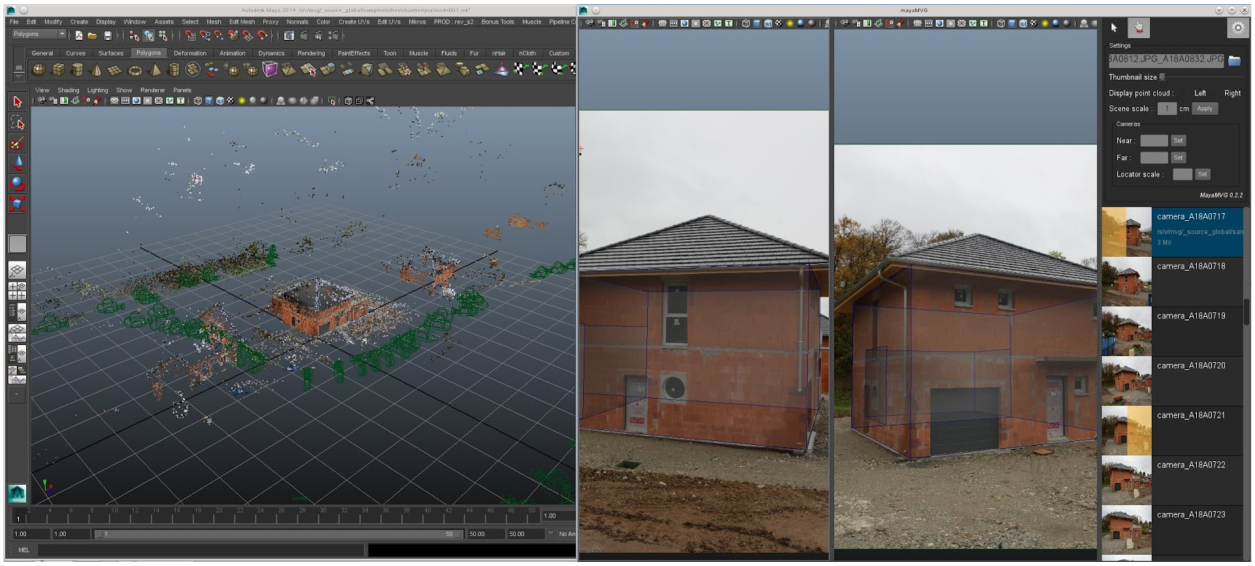
The target date for real-time performance of artificial features detection (task 2.1) is month 6. Artificial feature detection using CCTags has been improved to handle motion blur. The special situation found in the witness cameras that have been chosen for POPART, which use “rolling shutter with global reset”, is being investigated for further improvement of motion blur handling. The number of external dependencies in the code has been reduced, and a pipeline with GPU-centric decisions has been planned. Further algorithmic improvements that reduce memory transactions in favour of parallel work have been designed. The current state of the GPU implementation is incomplete, but promises that real-time detection can be achieved for 1K witness cameras.

Multi camera system calibration (SRL)

The target date for the multi-camera calibration (task 2.2) is month 7. The multi-camera calibration relies on the visual database, delivered by the SfM pipeline, which must contain both artificial (cctag) and natural features. Then, the integration of the cctag as feature in the SfM pipeline has been started but is still incomplete. Also, some prototyping have been done on computing the relative witness cameras poses based on a simulated synchronised multi-video streams.

WP3 - Shoot preparation, visualization and post production

T3.1a Shoot preparation : MayaMVG - Photomodeling Tool.
Before starting the POPART project, MIK created a maya plugin to load all openMVG data and allow the graphic artists to define faces (quads) which are snapped to the point cloud and extend this face. With the contribution of SRL, we added a new way to constrain the face to fit the point cloud but with the constraint of the connected edge in the mesh, which greatly improve the snapping.
We added a way to change the position and orientation of the reconstructed scene. We really improved the way to triangulate new points. We also added an exporter for the selected meshes and cameras in Alembic format, with a Nuke script to load it in Nuke and directly create a graph to project images.



We have installed the plugin in production and started to use it. We had a lot of feedbacks and we made many changes in the interaction, that includes a way to clean 2D information of the 3D mesh, the usage of maya to manage the shortcuts in order to get user editable shortcuts, fixes in the undo/redo, etc. We are currently working on a way to improve the performances of the intersection test that makes the plugin too slow to use with a big mesh.

T3.1b Shoot preparation : Tool to launch reconstructions offline or online (prototype)

We have done a prototype in collaboration with IRIT with a group of students:
http://oeufsdepie.github.io/MATRIX/index.html
https://github.com/OeufsDePie/MATRIX

This prototype retrieves the thumbnails from the camera, allows to download them, visualize the GPS position on a map and launch a reconstruction on a subset of pictures.

T3.2 Keying and compositing for on-set visualization

For the deliverable D3.2 due in month 4, we have decided to focus on the Maya integration to visualize the camera tracking in real-time. The keying and compositing developments will be done in the final visualization tool (D3.3).
MIK started a prototype to create a standalone application that can stream the camera position into Maya. This application will be the main POPART software to receive the video streams, compute the camera relocalization and stream the resulting camera pose over the network. Then we will receive this camera position right into Maya and also in a standalone POPART viewer as explained in D4.1.

WP4 - Integration and demonstration

Components API
The ambition of this workpackage is to design the workflow, dataflow and software components of the POPART Previz solution. In January, MIK experienced the SolidAnim previz solution on a real shooting. SolidAnim is the closest solution available on the market to the POPART ambition. That confirmed the need from our supervisors for such tools, but we also faced many troubles that show the need for a new tool integrated into the post-production workflow.
We have organized meetings with supervisors and graphic artists to analyze this experience and enlarge the scope to the whole POPART needs. This work has been the basis for defining the tools and their interactions.
The components API has been delivered as D4.1 on April 1st 2015.

Acquisition of multiple streams

LAB has worked on building a prototype of the acquisition system. An overview of the system is shown in Figure 2. The system is comprised of different sensors such as cameras, time code generator and our own embedded platform, which LAB have evaluated and integrated with our own developed hardware and software. The work have consisted of analysis of available hardware, development of custom hardware to integrate components, and finally develop software which enables capture of all data, where it can be safely stored as well as streamed with very low latency over the network via IP through LABs custom WiFi configuration or standard network cables. With regards to the witness cameras we have decided to use USB instead of IP cameras, as they are more robust, easier to operate and lower the total power consumption of the system.

The current state of the acquisition system is we have made good progress on the first portable prototype, as described in Task 4.4. We have put together all components in such a way that they are portable and can be mounted together with the camera rig described in the next section. The data sources are now synchronized with SMPTE time code, using external time code generation for robustness. Frame acquisition from the main camera, and the two witness cameras are now frame accurate, meaning all cameras capture pictures at the same instant. We have not yet finished acquiring lens meta-data, because we are waiting for test hardware to arrive, which BAN is helping us borrow. We are integrating with a what is called a FIZ (Focus, Iris, Zoom) unit, which is mounted on the camera, and operated by the focus responsible on the film set. This unit allows the acquisition system to read out the lens parameters, which will be stored time synchronized with the other data. We expect to deliver D4.2 on time in month 4.

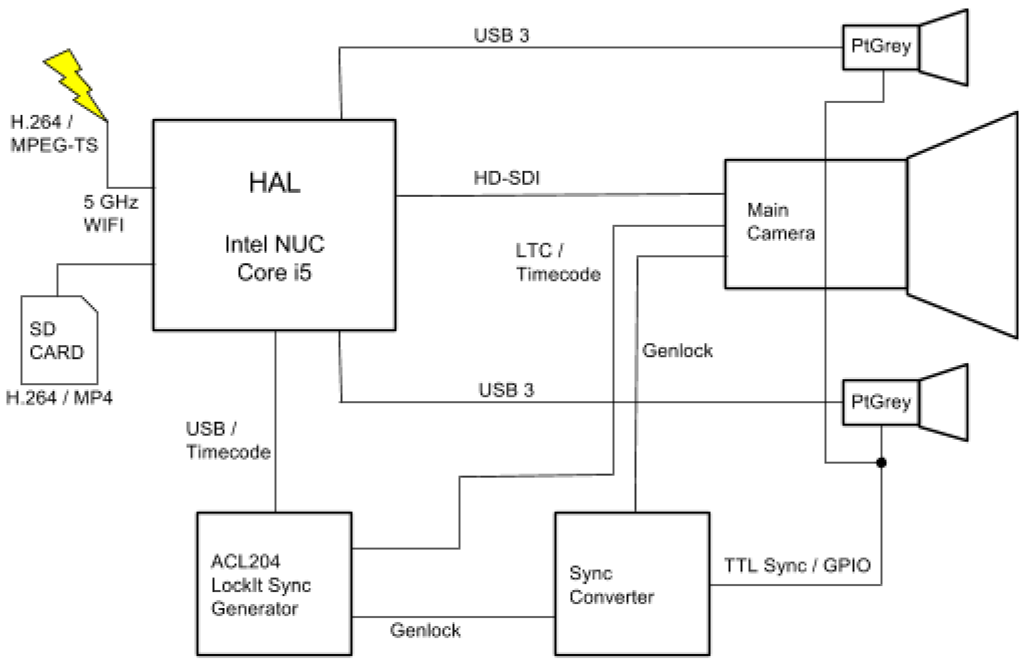


Figure 2: Overview of the camera acquisition system.

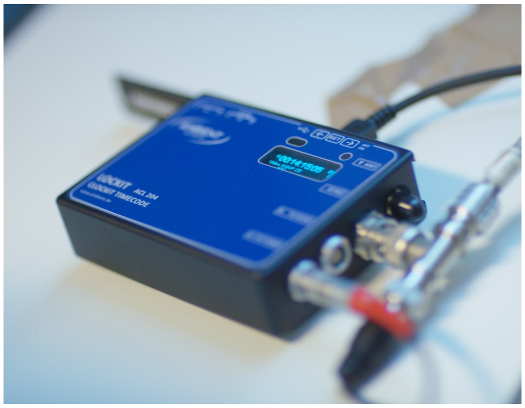
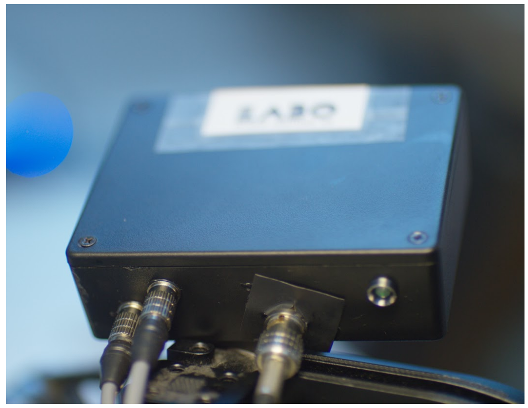
Prototype camera rig

The prototype Camera Rig as of March 2015 is ongoing work. We have identified and tested several components for a first prototype, as described below.

The initial prototype witness-camera-rig consist of a standard 19mm bridge for mounting on rods with standard rosettes and two arms to extend from the bridge and a rosette turn adaptor for mounting the witness cameras and turning them in all directions.

We have considered many different options for witness cameras, and have decided to use the Point Grey Greyscale 1-60 fps Flea 3 cameras with rolling shutter and global reset.

Timecode (TC) and Genlock/shuttersync is distributed from an Ambient TC unit (picture left below), delivering TC over USB to the HAL unit, and Genlock/shutter sync to the witness cameras and main camera.

The Genlock signal is distributed to a proprietary sync-trigger device (right picture) for the Point Grey cameras developed by LABO for this project and to the standard sync/genlock input on the main camera.

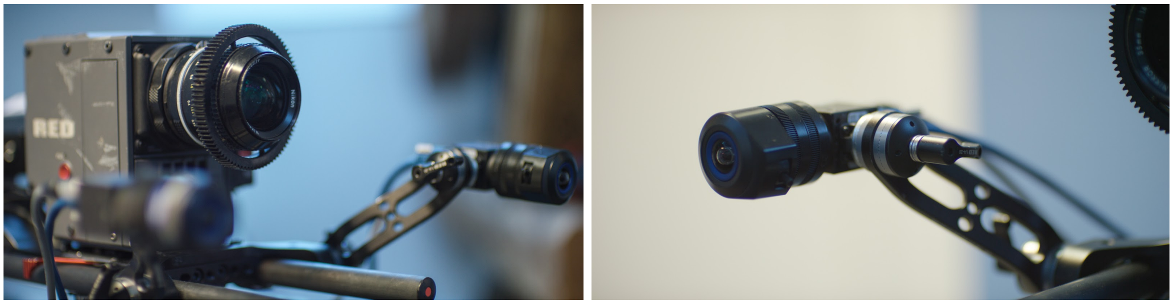
And then all three cameras are encoded with timecode embedded to H.264 on HAL and distributed through wifi to external units, like the live compositor/tracker down the line in the POPART project.

HAL (the capture and control unit pictured above) is based on a modified Intel NUC device, with a Black Magic Design SDI capture card (for main camera) and a custom-built Linux kernel to accommodate hardware transcoding of the video streams while still capturing synchronized video from USB3 (the witness cameras) and SDI (main camera).

The system will draw power from a standard D-tap on a V-mount battery (or similar), with an inverting unit in case of accidental polarity messups.

In addition to a solution like this, we'll have a solution where we mount the witness cameras on top of the main camera.

This is the 1st iteration of the capturing system, and we expect to make several changes going forward. We anticipate having a fully functional prototype rig for D4.4, due July 1.



The current iteration of the prototype rig with two witness cameras.

WP5 - Productisation and Distribution

This work package will start in month 7.

WP6 - Management

Project management has commenced without big obstacles to the progress of the project. The Kick off meeting was held at SRL in Oslo, and both physical meetings and video conferences were planned according to the Description of Work.

All partners present at the Kick-off meeting in Oslo, January 2015.

The consortium has agreed through an extraordinary General Assembly, to move 3PM for IRT from WP1 to WP4 (T4.6) to include IRT after 12 months in the project allowing them to contribute to the integration work.
The European Commission stated that such a transfer can be done without revising the Grant Agreement.

The pre-financing was distributed to all the partners in the middle of February.

Plenary meetings

Project meetings are held at least every 6 months. This is the meeting held in the reporting period in question:

- January 13-14. Kick off meeting at Simula Research Laboratory, Oslo Norway.

Plenary audio / video conferences

In addition to meetings, the POPART consortium holds a video conference every 4th week to keep track of the project progress:

- February 19th. Video conference
- March 19th. Video conference

More spontaneous teleconferences are also held supporting the different work packages and the partners involved.

Status of deliverables

Deadlines for internal review and final submission of each of the deliverables for this period have all been met.

- D1.1 - Initial 3D Visual Database library. Deliverable submitted on April 1st 2015.
- D4.1 - API Definition. Deliverable submitted on April 1st 2015.
- D6.2 - Quarterly management report #1. Completed QMR1 on April 1st 2015 and will be submitted as part of D6.2 in month 6.
- D6.1 - Tool documentation. Deliverable submitted on February 1st 2015.

Advisory board

For the POPART kickoff, only the leader of the advisory board, Gudrun Austli, were able to attend the meeting. Fortunately, Randy Wedick from BandPro and the LA-scene were able to introduce the development team of the workflow challenges and constraints that are inherent in real world film sets. Miss Austli's experience as a producer also gave valuable input to the team regarding safe insurable data infrastructure. The POPART consortium hope all the members of the advisory board will be able to attend at the next consortium meeting in Toulouse, but due to the busy schedules in the film world, we are not yet sure if this will be possible. An alternative we consider is to visit the advisory board members, and show them prototypes and collect feedback for project direction in addition to meetings online.

Project acronym: POPART Project number: 644874 Work package: Management Deliverable number and name: D6.2: Quarterly Management Report 1 and 2	Title: Quarterly Management Report 2 Work Package: WP6 Version: 1 Date: July 1, 2015 Author: Håvard Espeland
Type: [X] Report [ ] Demonstrator, pilot, prototype [ ] Website, patent filings, videos, etc. [ ] Other	Co-Author(s): All partners To: Albert Gauthier, Project Officer
Status: [ ] Draft [ ] To be reviewed [ ] Proposal [X] Final / Released to EC	Confidentiality: [X] PU – Public [ ] CO – Confidential [ ] CL - Classified
Revision:	
Contents: Deliverable 6.2: Quarterly Management Report 2	

Human Resources

SRL

The project is coordinated by Håvard Espeland. Carsten Griwodz, Lilian Calvet contribute to the innovation and scientific work. Elisabeth Andersen contributes to the administrative work.

LAB

The project at LAB is led by Magne Eimot. Jon M. Puntervold, Gunleik Groven and Håvard Espeland also contribute to the innovation work. Kristian Skarseth has recently joined LAB as a developer.

MIK

The project at MIK is led by Benoit Maujean, R&D Manager at Mikros Image. Fabien Castan is the technical manager of Popart project; he is responsible for the 3D reconstruction library (Openvg) and photomodeling plugin for Autodesk Maya (MayaMVG) at Mikros Image. Elisa Prana, Nicolas Rondaud and Cyril Pichard, as software engineers contribute to the development of Popart tools, including the expected adaptations of OpenMVG and MayaMVG. Furthermore, some Matte Painters of Mikros Image studio have contributed to the generation of the virtual dataset.

IRT

The project at IRT is led by Vincent Charvillat. Simone Gasparini is responsible for the Work package 1 and he's coordinating the tasks assigned to IRT, in particular on camera localization and the contribution to the OpenMVG libraries. Sylvie Chambon is also a staff member of the project and she will participate in the task involving the image matching for improving the performances of OpenMVG. Clement Aymard is a developer engineer who is developing the different contributions to the project.

BAN

The project at BAN is led by Christopher Hantel, CEO at Band Pro Munich. He is joined by Randy Wedick, contributing to the commercialization and productization of the POPART commercial product.

Person Months Contributed to the Project

Participant	Names of staff	WP1		WP2		WP3		WP4		WP5		WP6		Total	
		Planned	Actual	Planned	Actual	Planned	Actual	Planned	Actual	Planned	Actual	Planned	Actual	Planned	Actual
SRL	Håvard Espeland Lilian Calvet Carsten Griwodz Elisabeth Andersen	3,0	2,12	10,36	5,71	0	0	0	0	0	0	3	3,56	16,36	11,39
LAB	Magne Eimot Jon M. Puntervold Håvard Espeland Gunleik Groven Kristian Skarseth	0	0	0	0	4,4	4.81	7,4	7.93	0	0	0	0	11,8	12.74
MIK	Benoit Maujean Fabien Castan Elisa Prana Nicolas Rondaud Cyril Pichard	2,26	1,24	1,1	0	7,8	5,78	4,4	5,03	0	0	0	0	15,54	12,04
IRT	Simone Gasparini Vincent Charvillat Clement Aymard Sylvie Chambon Nicolas Bertrand	9,25	7.95	3,27	3	0	0	1,2	0	0	0	0	0	13,72	10.95
BAN	Randy Wedick Christopher Hantel	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total		**14,52**	**11,31**	**14,74**	**8,71**	**12,2**	**10,59**	**13**	**12,96**	**0**	**0**	**3**	**3,56**	**57,43**	**47,13**

Figure 1: The cumulative budgeted and contributed person-months since project start for each partner sorted by Work Package.

Project Progress

The project is running as planned and we are working on all tasks expected to implement POPART and its objectives. Details on each work package is given below.

WP1 - 3D Visual Database

3D visual database

A first version of the 3D visual database has been delivered as deliverable D1.1 and the relevant source code is available at the following Github repository:
https://github.com/poparteu/cameraLocalization
Since then, several improvements have been addressed in order to speed up the vocabulary tree creation, and query and to maintain the compatibility with the OpenMVG data format. In particular most of the code is now parallelized on CPU, which can reduce the time required to

generate the vocabulary tree. As for the input data format, OpenMVG 0.81, released on May 30th, introduced a new improved data format to collect the data of the SfM pipeline, which in turn required to adapt the voctree library to use the same data format.
On going works on the visual database are aiming at implementing some variants of the vocabulary tree methods in order to speed up the query process, and to ease the computational burden of the matching process in OpenMVG. These two aspects are explained more in detail in the OpenMVG section and the Database Query section of WP2.

OpenMVG

The OpenMVG SfM pipeline is based on the brute force matching of all the images provided in order to establish the matching relation among the images and thus computing the epipolar geometry relating each pair of images. This has a highly computational cost (O(n^2)) which severely affect the time required for the computation of the structure for large dataset of images: for example a dataset of 600 images, a common size for a real case scenario, the brute force matching among all the images may take up to more than 10 hours. This time constraint cannot be sustained during the shooting of a movie, where a reliable reconstruction of the scene has to be done the day before the actual shooting. In order to ease this burden we decided to use the vocabulary tree approach which can provide, for a query image, a set of the best matching images of a dataset. In particular we decided to adopt the general approach VocMatch recently introduced by [Havlena2014]. The main idea of this approach is to train a large visual vocabulary from a large set of images in order to obtain a large set of visual words (16 million visual words from a vocabulary tree with 2 levels and a branching factor of 4000): this allows a finer classification of the SIFT features into visual words, where, ideally, each visual word corresponds to an unique feature. Thus the feature matching problem can be cast as image indexing problem: beside evaluating the similarity of two images using the usual approach of comparing the visual word histogram, the methods allows to recover the potential matching features between the two images. This allows to reduce significantly the computational cost of the matching process.

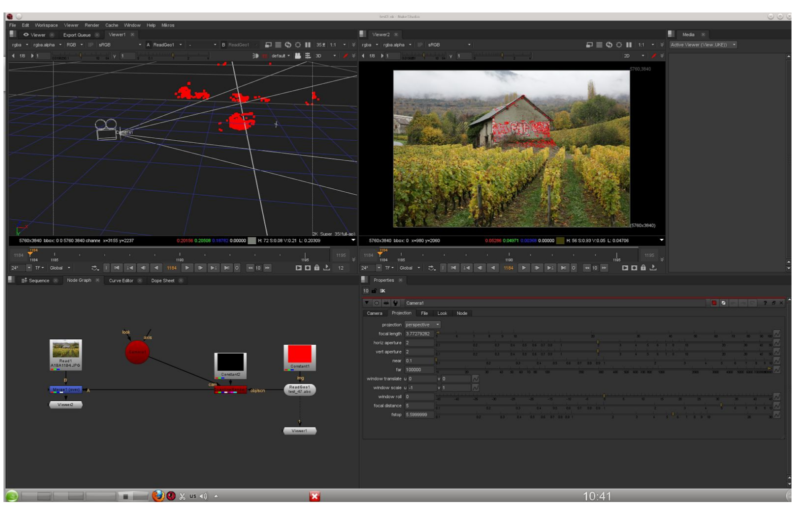
Some preliminary experiments has been carried out between IRT and MIK on a dataset of 514 images using a preliminary and incomplete version of the method. We trained a large visual vocabulary using 15000 random images downloaded from Flickr and we used the vocabulary to find the matching images for each of the 514 images. For each image of the dataset we then get a list of N best matching images, with which we fed the OpenMVG incremental pipeline (thus skipping the brute force O(n^2) image matching process and limiting the match to the list obtained from the vocabulary tree. This preliminary experiments showed quite promising results as they showed a drastical reduction of the computational time yet preserving the final quality in the 3D reconstruction. The brute force matching took around 10 hours for the 514 images, while using the list generated with the vocabulary tree the matching process took up to 1 hour, as the matching time depends on the value of N (the larger N is, the more it takes). The table below summarize some of the results in term of the number of best matching image provided by the vocabulary tree (N), the relevant matching time inside the OpenMVG incremental pipeline and the number of camera reconstructed (expressed also as percentage of the number of camera reconstructed with the brute force approach)

N	matching time	Number of reconstructed cameras
10	35 min	402 (78%)
30	50 min	449 (87%)
60	~ 1 h	477 (92%)

On going work is aimed at fully implementing the VocMatch method in order to obtain directly the candidate matches between two matching images. More experiments will be also carried out on larger dataset in order to asses the validity of the method and the gain in performance wrt the current OpenMVG implementation.

[Havlena2014] Havlena, M., & Schindler, K. (2014). VocMatch: Efficient Multiview Correspondence for Structure from Motion. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Proceedings of the 2014 European Conference on Computer Vision (ECCV 2014)* (Vol. 8691, pp. 46–60). Zurich, Switzerland,: Springer International Publishing. doi:10.1007/978-3-319-10578-9

MIK added the support of Alembic in openMVG. Alembic is a 3D file format which supports the common geometric representations used in the industry, including polygon meshes, subdivision surface, parametric curves, NURBS patches and particles. Alembic also has support for transform hierarchies and cameras. We are now able to generate an Alembic file directly from openMVG and load it in Nuke as you can see in the illustration below.



The Point Cloud and Cameras exported from openMVG are loaded in Nuke

We faced many challenges because there are some differences in the interpretation of the Alembic file in the different versions of Nuke and Maya. With the support of Alembic in openMVG we drastically improve the file storage size and the record performances. The existing file format used in openMVG is text based without any compression or optimization. On the other hand, Alembic contains two storage back-ends: HDF5 and Ogawa with good compression and multi-threaded reading and writing.

About Alembic: http://www.alembic.io
About HDF5: https://www.hdfgroup.org/HDF5/whatishdf5.html
About Ogawa: http://www.alembic.io/updates.html

Alembic supports custom attributes allowing us to store openMVG specific information.
Relying on a standard file format allows us to use it directly in most of the 3D softwares used in post-production instead of writing specific loader for all these softwares.

WP2 - Real-time camera tracking

GPU artificial feature detection (SRL)

Feature tracking in POPART relies on CCTags, whose theoretical development was conducted at IRT, but which is continued by POPART participants both at IRT and SRL.

CCTags are planar markers that consist of a well-known number of concentric circles. These markers are to be pre-installed on a set in sections that will not be obscured but are also outside the expected viewing angles of the main camera. Their physical position on the set is pre-determined and stored. The CCTags are then during the shoot tracked by witness cameras that are mounted on the main camera but tend to point into other directions.

By using a sufficiently large number of concentric rings, it is possible to determine a parametric curve from the other edge of the outer ring to the common center of all rings, even though the planar marker is not seen in an orthogonal view. The CCTag algorithm builds parametric curves that converge onto this center, which leads to an algorithm for the localization of the CCTags' centers that is robust to challenging images, in particular to noise, blur and strong occlusions. Simultaneously, the width of the concentric rings provides a low-resolution barcode-like identifier that allows each CCTag to have a specific ID.

Importantly for the POPART project, CCTags remain identifiable both in their geometric and photometric characteristics even in case of strong motion blur, to the extent that some multi-directional motion blur can be tolerated. The reason for this is that likelihood that uni-directional motion blur leaves those sharp edges of circles sharp that are orthogonal to the motion direction. Thus, concentric circles allow the partial preservation of edges located along the virtual line passing through their imaged center and whose the direction is orthogonal to the (supposed unidirectional) motion blur direction. Experiments have shown that this property is still true even in presence of perspective distortions.

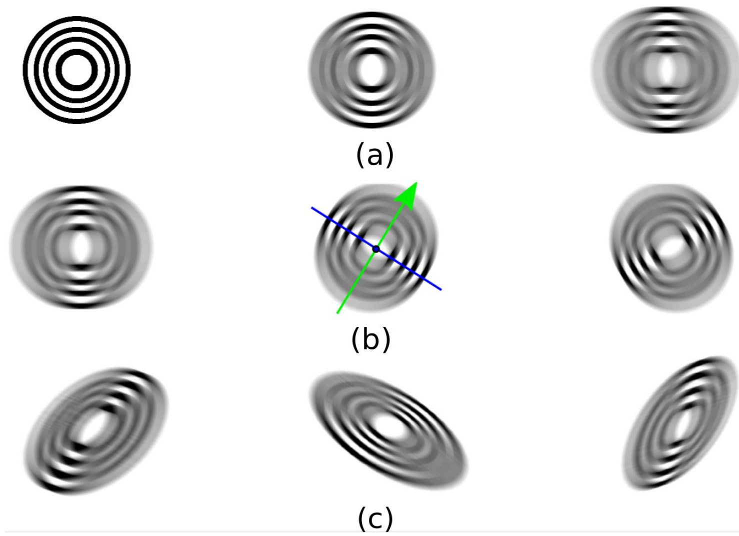
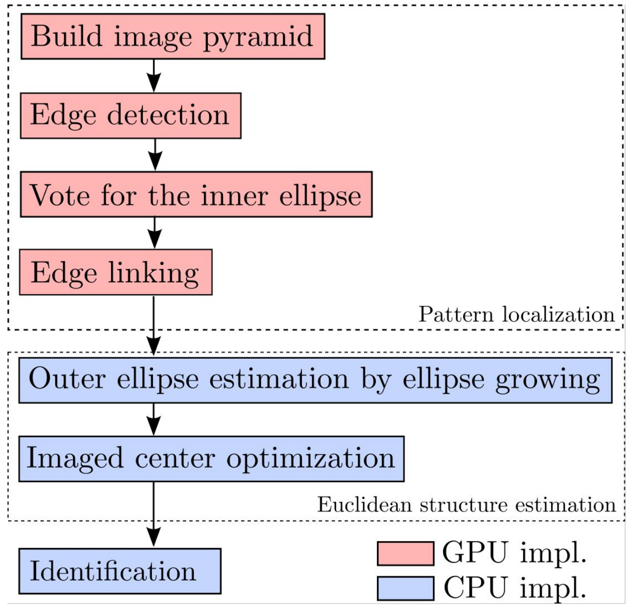


Illustration of CCTags with blur applied: (a) horizontal, (b) multi-directional, (c) in the presence of perspective distortion

The optimized CPU code for the computation of CCTags is on the scale of 100ms of milliseconds for each image even at SD resolutions. It was therefore necessary to move compute-intensive parts of the CCTag computation to the GPU. The choice of the project fell on a particular GPU of the most recent NVidia GPU generation, the GeForce GTX 980. Although it provides less parallel computing cores than NVidia's Titan family of cards, it is considerably cheaper, while still offering compute capability 5.2 and retaining all the architectural benefits of the latest generation of cards. For the implementation, CUDA (SDK 7.0) was chosen.

The design of CCTag is well-suited for partial parallelization, due to several pixel-centric steps at the beginning of the process, followed by a sequence of point selection, which requires knowledge about all candidates for CCTag identification, optimization of the center selection, and finally, an identification of the particular CCTag ID from an estimation of ring size ratios. The optimal split between GPU-side and CPU-side code is not certain yet, for two reasons:

a. The optimization of both CPU and GPU code is still progressing with the intent of supporting an even larger number of cameras and higher resolutions at real-time;
b. The initial assumption that the edge image, which must certainly be generated on the GPU side, would only be required by the GPU, was found to be untrue. Since it must be transferred from the GPU to the CPU anyway and optimized CPU code exists for all stages, there is room for experimentation.



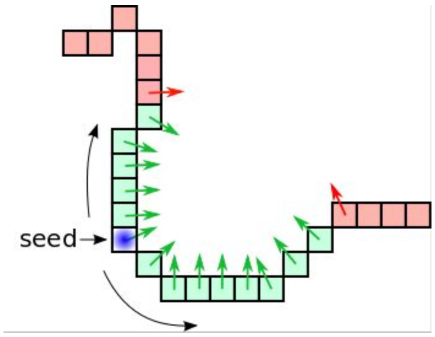
Overview of responsibility split for CCTag steps between CPU and GPU

The overview of steps shows the sequence that has to be performed for every single camera image. Identification in this step provides a list of IDs for all CCTags that have been identified from an image, where an increase of markers in a image affects all steps from edge linking to identification.

Building the image pyramid is a means of reducing the effort of estimating circle radii and circle widths by creating lower resolution images. In the GPU implementation, this makes firstly use of the GPU's texture engine that can use normalized coordinates to create enlarged as well as reduced versions of the input image through hardware bilinear filtering. This is followed by Gaussian filtering to reduce noise for each resolution. Gaussian filtering is well-suited for GPU implementation because it can be implemented by separated convolution filters, which reduces the number of required memory accesses and floating point operations.

Edge detection for CCTags relies on the Canny edge detector, which is mostly comprised of localized decisions. The most time-consuming step of the Canny edge detector is the edge hysteresis step, which requires multiple sweeps of the entire image plane to upgrade possible edge points to probably edge points based on neighbourhoods. The small step consumes currently 80% of the GPU computation time for the Canny algorithm. The thinning step, on the other hands, could be optimized by improved lookup tables.

Voting requires, first, a sweep of the edge image, as well as gradient images, to identify stretches that may be spanning the thickness of a black or white circle of a CCTag in an image. These candidates are stored in a list (an array on the GPU), and are subsequently chained to determine whether they may comprise a chain of stretches pointing towards an image center. While the creation is localized and well-suited for GPU operation, chaining is not, because global pointers should be avoided on GPUs. Instead, lookup tables are required, which are rather cache-unfriendly because of the disconnection of candidates in the list and the coordinates of potentially neighbouring pixels in the edge image. Furthermore, chain length is defined by a run-time configuration parameter, which reduces optimization options further. In spite of this, there is such a large number of steps that consist mainly of 1D and 2D index lookup operations that this step is still much better suited for the GPU than for the CPU. The output of this step is a very small number of CCTag centers and points located on outer circles of those tags.



An illustration of Edge Linking. The blue dot represents the position of the seed.
The black arrows illustrate the two directions followed by the edge linking algorithm into both directions. The set of green pixels represents the convex edge segment obtained. The linking is stopped when the gradient direction are “diverging” (red arrows).

Edge linking has not been moved to the GPU yet. Because of the small number of CCTags that are expected in an image, the small number of points that are evaluated in every step of the edge linking loop, and the unknown duration of the loop, it is highly uncertain whether moving it to the GPU is a good idea. However, all decisions that are taken are strictly local to a known edge point, and there are supposedly many points of the same outer ellipse of a CCTag that are growing together to form an arc segment. It is thus still a matter of experimentation whether this can lead to a further performance improvement.

In terms of computing efficiency, we are achieving the following average computing times for example images of size 3264x2448:

	single-core CPU	GPU
Uploading of the image (only on the GPU), building the Pyramid, Gaussian filtering, Canny edge detection	519ms	9.17ms
Voting procedure	221ms	9.84ms

The table demonstrates how the efficiency of the GPU suffers as random memory access and loops start to dominate the workload, but it is still achieving a 20-times speed-up compare to a single-core CPU version. Unfortunately the connection to the full CPU pipeline is still missing.

The code is available at https://github.com/poparteu/CCTag

Multi camera system calibration (SRL)

To support multi-camera calibration, SRL has undertaken the development of a new, GPU-based SIFT implementation, which provides required real-time functionality for the natural feature tracking requirement for the main camera. This tracking is supposed to be used as the primary source of tracking information for the POPART system, with augmentation by the CCTags in difficult cases; here in particular green screen scenarios.

SIFT (Scale-Invariant Feature Transform) is a rather heavy-weight, but classical operation that extracts a large number of 128-float feature vectors from images, each of which describes a visually dominant real-world point within its pixel neighbourhood. This point can be matched with features extracted from other images containing the same point, and it is known to achieve a considerable number of well-matching points across viewing angles as large as 30 degrees and a large range of image scales.

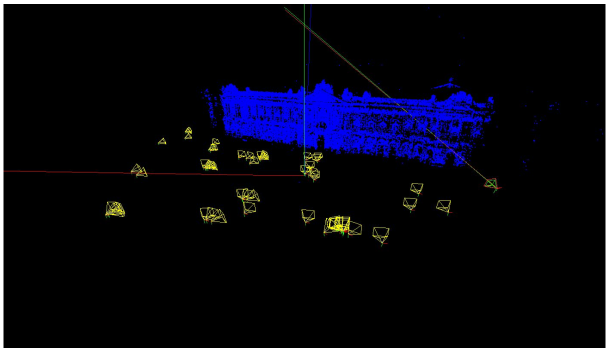
Although SIFT is currently the dominant feature extractor in research works, it is, inconveniently, both computationally heavy and patented in the USA. Of the implementations that are publicly available, there is one that achieves real-time speed for 1024p images. Unfortunately, this particular implementation cannot be adopted for POPART because it is exclusively licensed for use in research, with no option for acquiring a commercial license.

Because of this, SRL with help from IRT created a new implementation of the SIFT feature detector in CUDA on the GeForce GTX 980, which works currently just in real-time (39ms / image) for images of 1920x1200, when about 10000 features are detected. Most of the code is newly written based on the original SIFT paper. The final two steps, computation of the feature point orientation, and final computation of the feature vector, have been ported from the OpenCL implementation of the BeMap project, which is available under a BSD-like license, and which is turn derived from the VLFeat implementation.

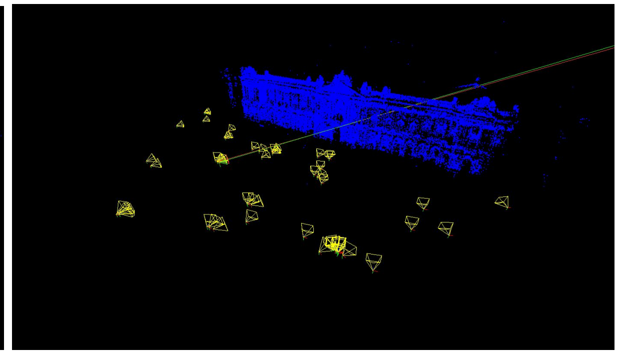
The code is available at https://github.com/poparteu/popsift

According to theoretical considerations, there should be potential for considerable further speedup in the code structure, by changing the order in the creation of the SIFT image pyramid. While this would not make any different in analog signal processing, it is currently unclear whether it would change the quality of the detected feature vectors. This is matter of further investigation.

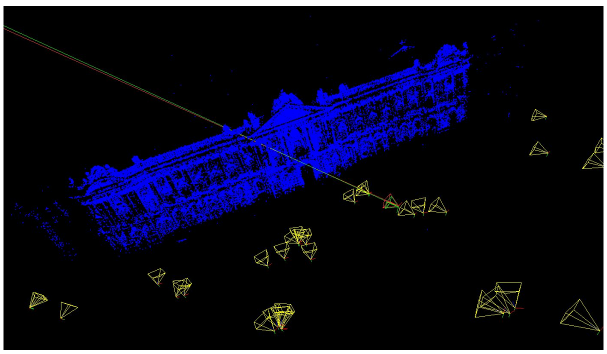
Query from the database (IRT)
IRT is currently developing and evaluating different strategies for the query to the database. A first naive and simple method consists on taking the first best result of the query and use it for the camera localization: the best matching image is matched with the query image and the 3D-2D correspondences are then computed and used to estimate the camera position. This methods, even if it is very basic and simple works in many cases, but it is not robust to outliers and wrong matches.



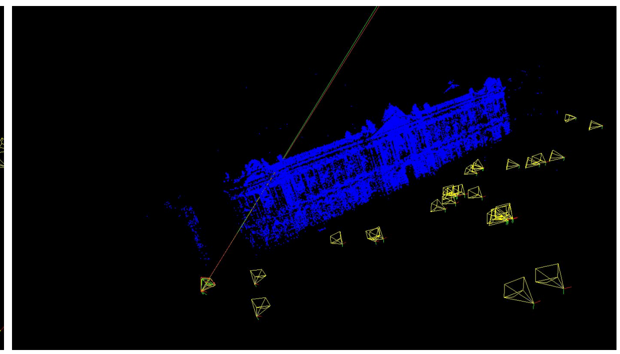
mean reprojection error : 0.043 px



mean reprojection error : 0.13 px



mean reprojection error : 0.24 px



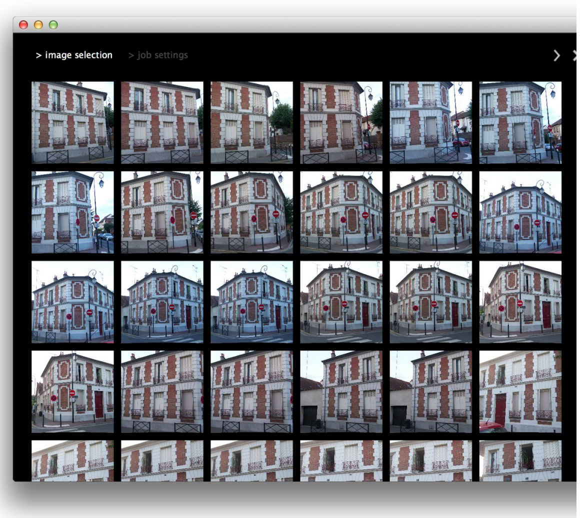
mean reprojection error : 0.018 px

Some examples of successful image localization for Toulouse Capitole dataset with the naive method and the relevant mean reprojection errors.

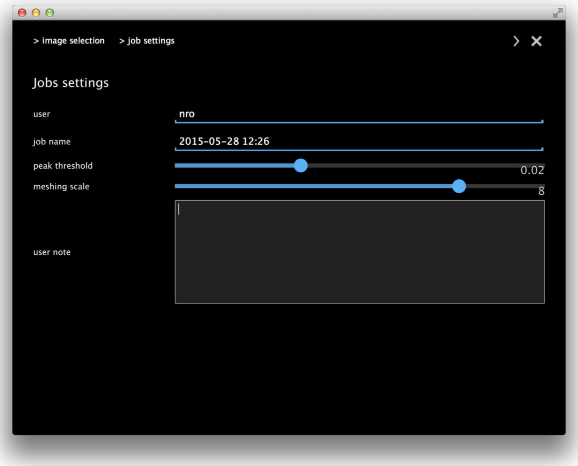
A different approach consists in retrieving the N best matches for the query images, and from them retrieve the 3D points of the scene that are visible by the N images. Then the 3D points that are less visible (ie they are visible by less than k images) are discarded. For each of the remaining points, a mean SIFT descriptor is computed from the SIFT features associated to that point in the images in which the point is visible. Then a FLANN matching between these mean SIFT features and the features of the query image is performed, thus retrieving the 3D-2D point associations in order to estimate the pose. This method should be more robust to outliers as it intrinsically considers the results images that are likely to belong to the same cluster of images depicting the same part of the scene. This second method is being developed and soon tested in order to asses the effective robustness and the computational cost of creating the data structure at each query. Both methods will be part of the deliverable D2.2 due on month 7.
Other interesting methods that we will consider will be the use of the VocMatch approach to retrieve at the same time the best matching image and the correspondences as described in the section OpenMVG of the WP1. This approach could be interesting in order to reduce the query time, even if the query to a larger database may be more costly in terms of computation (for each feature that needs to be compared in the vocabulary tree we need 4000*2 comparisons, instead of the classical 10*6 comparisons).

WP3 - Shoot preparation, visualization and post production

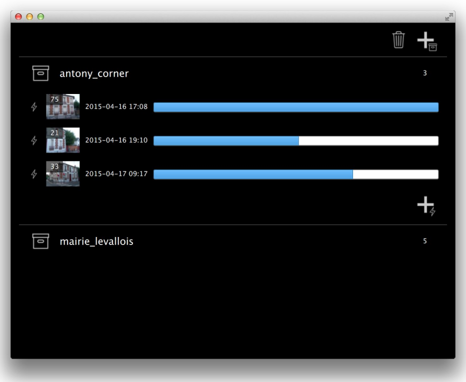
MIK has worked on the main reconstruction software. The UI allows to add images by drag & drop from your prefered file manager, choose the reconstruction options, compute locally or on renderfarm and follow the progression.



Select input images



Choose the reconstruction options



Follow the progression of the reconstructions

For the delivery D3.2, as explained in QMR1, we have decided to focus more on the reconstruction software and video acquisition before putting to much effort on the initial visualization tool.

As we have defined in the API definition, 2 softwares will receive the camera poses in live: one is the visualization tool and the other one is the 3D editing software. MIK has started with the easiest one: streaming the the camera pose into Autodesk Maya.

On the other hand, LAB has developed a software to visualize the 3 video streams. LAB has added a basic real-time 3D rendering which is able to load FBX files. LAB has also integrated the support of the commercial keyer called Primatte.

See D3.2 for more details.

WP4 - Integration and demonstration

In WP4, the three deliverables that have been worked on during the second quarter are the 2 Datasets, the Prototype Camera Rig and the Acquisition of multiple streams.

Mixed Reality Dataset

The mixed reality dataset is now completed as planned with the exception that we are using offline tracking with SynthEyes instead of Intersense as comparison. The dataset has been released to the public in accordance with the proposal. The shooting dataset is detailed in report D4.3.

Making of the Mixed reality dataset in Grimstad, Norway 2015

The dataset has been recorded with the acquisition system (D4.2) and the prototype camera rig (D4.4). Both these deliveries have worked out as intended, and shown promise for the viability of the POPART system. The studio was fitted with CCtags (WP2), and initial tests show good marker recognition, even under harsh lighting conditions.

The data-sets ambition has been somewhat scaled back, in accordance to the delay of the on-set visualisation tools, a new dataset will be done this autumn in tandem with an advisory board workshop, exploring working on-set with a complex mixed reality scenario with POPART-technology.

The physical assets for this scene are already under construction, and the virtual assets have been created.

Virtual dataset

MIK has created a complete virtual set with photo-realistic rendering and camera motion, in order to dispose of a controlled environment to be used as a reference set with a real ground truth.

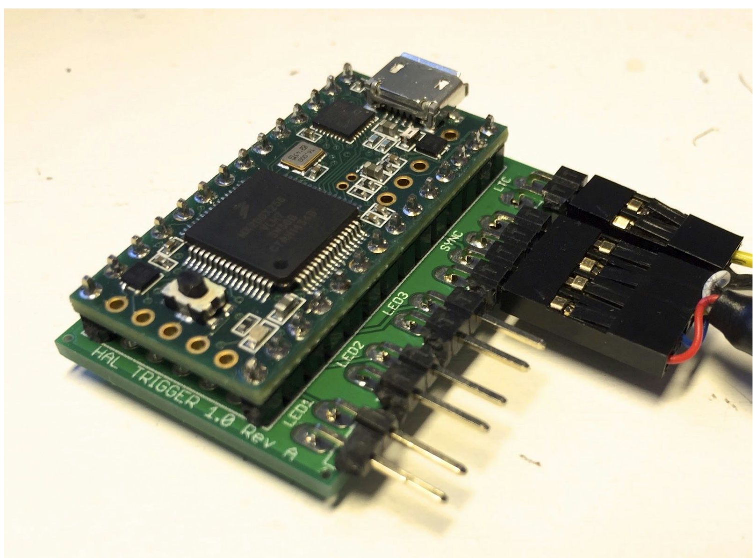
It contains two kinds of renders: a set of still pictures to evaluate the 3D reconstruction and a camera motion to evaluation the camera tracking/localization.

For the camera motion we have computed images with several settings of the 3D render (motion blur and/or depth of field enabled) and several camera focals.

View from a reconstructed camera in Autodesk Maya, with the 3D modeling on top of the image plane

Prototype Camera Rig
LAB has completed two revisions of the camera rig, and have now designed a functionally complete prototype. LAB is now ready to begin the productification of a custom made camera rig. Further information about the Camera Rig is described in report D4.4.

Acquisition of multiple streams
The acquisition system has been completed to fill the needs of the POPART after plan in month 5, and is detailed in report D4.2. After the deliverable, LAB has continued to improve the acquisition system to improve reliability and use software trigger instead of genlock to synchronize the witness cameras to the main camera. This enables removal of a trigger cable to each witness camera and reduces risk of having the wrong settings on the genlock generator, ruining the recording. The improved trigger system is still in development, and will be integrated directly with the RED brain, further camera models to be added later. Until this new system is complete, we continue to use the trigger system delivered in D4.2.



Improved trigger system for the witness cameras developed by LAB in June 2015.

WP6 - Management

During this quarter, the partners were gathered in Toulouse for the second plenary meeting.
The Advisory board was also invited and was present at this two-day meeting.
A General assembly was held, where we decided on the location for the next plenary, discussed the review meeting and the open source licenses.

Consortium meeting with all partners and Advisory board, May 28th in Toulouse.

Plenary meetings

Project meetings are held at least every 6 months. This is the meeting held in the reporting period in question:

- May 28th - 29th: 2nd Plenary at Institut National Polytechnique De Toulouse, France.

Plenary audio / video conferences

In addition to meetings, the POPART consortium holds a video conference every 4th week to keep track of the project progress:

- May 20th Video conference
- June 30th Video conference

More spontaneous teleconferences are also held supporting the different work packages and the partners involved.

Status of deliverables

Deadlines for internal review and final submission of each of the deliverables for this period have all been met.

- D3.1 - Shoot preparation tool. Deliverable submitted on May 1st 2015.
- D4.2 - Acquisition of multiple streams. Deliverable submitted on May 1st 2015.
- D3.2 - Initial visualization tools. Deliverable submitted on July 1st 2015.
- D2.1 - Artificial feature detector library. Deliverable submitted on July 1st 2015.
- D4.3 - Deliver dataset. Deliverable submitted on July 1st 2015.
- D4.4 - Prototype camera rig. Deliverable submitted on July st 2015.
- D6.2 - Quarterly management report #1-2. Deliverables submitted on July 1st 2015.
- D6.5 - Data management plan. Deliverable submitted on July 1st 2015.

Adjustments to the DoW

To accommodate to the progress of the project and the schedule, we have moved deliverable D3.2 from May 1 to July 1. This was approved by the PO.

The reason for requesting this change was that we wanted to build a solid acquisition system as a strong foundation for the rest of the project before starting the development of the initial visualization tools.

Advisory board

By request of advisory board member Henning Lind Eriksen, Lars “Lalo” Werner, the head of post-production at Zentropa in Copenhagen (Klippegangen and Filmbyen) joined the advisory board. His merits include being one of the visual effects supervisor on Melancholia, supervising the critically acclaimed Criterion Collection edition of Breaking the Waves (together with Lars von Trier), as well as supervising dozen and dozens of Zentropa feature-films through the world-renowned Klippegangen post-production facility, having hands on experience with VFX, rotoscoping, workflow/pipeline integration and pre-viz.

Max Penzel has moved to Singapore, and has been replaced on the board by Randy Wedick of Band Pro, who through his company, among other things, serves the technical ambitious and world-renowned director David Fincher and his team with cutting edge solutions. Randy also works as a director and DoP on a documentary on the late Dennis Hopper, being shot on special monochrome Red Dragon cameras.

The advisory board was present in the plenary Toulouse meeting and joined the discussions on the development status, progress and future work. Many technical suggestions and comments were made, but the main take-away was that the Advisory board members would like to experience using the system as soon as possible. We will therefore host an ambitious workshop to enable the advisory board to explore hands-on possibilities of the technology.

POPART –advisory board comments, summary penned by Advisory board leader Gudrun Austli after the Toulouse meeting:

For POPART to be successful we believe it needs to

1. *Simplify the process artistically, giving the director a greater opportunity to do creative choices on set.*
2. *A cheaper and easier way of making set extensions, than already existing solutions.*
3. *Store all the metadata in to one bank, making it easy to find all data about one shot.*
4. *Be easy to use on set. Without a lot of extra wires or technical crew on set during shooting.*

Testing – focusing on end results.
The main target for the advisory board at this stage of the development is to understand where this technology will be a helpful tool. To do that, it is essential to the advisory board to see some test, focusing on the end results.

The tests should be showing both what this technology is able to do and where the limits are. Examples:

- *Backlights.*
- *Flare.*
- *Handheld camera.*
- *On-set interaction with virtual assets*

For the next stage of the development, after seeing the tests, the advisory board should be able to do some case studies. Going through examples from previous productions, trying to figure out where POPART technology would have been a better choice. It´s also important for the advisory board to look at what situations POPART would not be the preferred choice, and why.

Previz – on set.
This is essential for POPART to be successful and we expect this to have the main focus after testing and understanding the main use of POPART.

Conclusion

The POPART project is overall on schedule with all partners contributing to their respective work packages. All deliverables have been submitted in time. We have made some minor adjustments to the DoW, as explained above. Additionally, we have enhanced the advisory board with well-known industry professionals. The new members have extensive experience in working with film productions where POPART applies. In addition to giving valuable feedback to the project, this can also increase the impact of the Results by opening doors to high-profile projects.