<table>
<tr><td>Project acronym: LADIO

Project number: 731970

Work package: Management

Deliverable number and name:

D6.3: Quarterly Management Report 3</td><td>Title: Quarterly Management Report 3

Work Package: WP6

Version: 1
Date: December 30, 2017

Author:
Carsten Griwodz</td></tr>
<tr><td>Type:
[X] Report
[ ] Demonstrator, pilot, prototype
[ ] Website, patent filings, videos, etc.
[ ] Other</td><td>Co-Author(s):
All partners

To:
Albert Gauthier, Project Officer</td></tr>
<tr><td>Status:
[ ] Draft
[ ] To be reviewed
[ ] Proposal
[X] Final / Released to EC</td><td>Confidentiality:
[X] PU – Public
[ ] CO – Confidential
[ ] CL - Classified</td></tr>
<tr><td colspan="2">Revision:
Final</td></tr>
<tr><td colspan="2">Contents:
Deliverable 6.3: Quarterly Management Report 3</td></tr>
</table>

Human Resources

SRL

The project is coordinated by Professor Carsten Griwodz. Konstantin Pogorelov, Jonas Markussen, Gunleik Groven, Stian Zeljko Vrba and Kristian Skarseth all contribute to the innovation and scientific work. Elisabeth Andersen contributes to the administrative work. Gunleik Groven, Stian Zeljko Vrba and Kristian Skarseth were formerly employed by LABO Mixed Realities.

MIK

The project is led at MIK by Benoit Maujean, R&D Manager. Fabien Castan, as senior lead developer is the technical manager of LADIO project, to coordinate the partnership in all development-related project work. Yann Lanthony and Nicolas Rondaud as senior software engineers as well as Jean Melou (PhD student) contribute to the development of LADIO tools. Thomas Eskénazi, as software architect, contributes to the data model and API definition. Grégoire De Lillo contributes on the SfM improvements as an R&D intern from IMAC engineering school.

INP

The project at INP is led by Pierre Gurdjos. The innovation and scientific work is supervised by Sylvie Chambon, Vincent Charvillat, Simone Gasparini and Géraldine Morin, all senior researchers. Hatem Rashwan was a full-time post-doc researcher dedicated to the project (namely the task T3.1.1 in WP3) and finished his employment contract on June 30, 2017. Clément Debize is a research engineer working on the project.

CTU

The project at CTU is led by Dr. Tomas Pajdla. Main innovative and scientific work is carried out by Ph.D. students Cenek Albl and Michal Polic. Additional contributions come from MSc. students Pavel Trutman, Jan Krcek, and Oleh Rybkin.

Project Progress

WP1 - Onset Data Acquisition

T1.2: Distributed Data Management
(planned duration months 7-9)

Distributed data management consists of a multitude of small components that are combined with the LADIO backend. These components are not assembled into a complete system before final integration.

The design of LADIO called for an integration with a peer-to-peer (P2P) storage management system that would allow CamBox, MiniBox and SetBox devices to act as fairly independent storage nodes. The envisioned platform for this was Infinit File Storage Platform (http://infinit.sh). The vision relied on the assumption that CamBox and MiniBox units would be always connected, a stable WiFi connectivity could be guaranteed at all times, and CamBox and MiniBox units would remain available throughout shooting days.

Through experiences that Quine made with the CamBox on the film sets of TV productions and preparation shoots, we had to accept the insight that several aspects of this approach are unrealistic. A CamBox connected to a main camera will on many sets record continuously, independent of a particular Take, and the same must be expected from 360 cameras. Other recordings devices, including secondary cameras whose recordings is desirable for the post-production team, may be used for a brief time and switch off immediately, not coming online again before being taken off-site. More frequently, the SD cards of secondary recording devices may be removed, breaking the link between the device and its content. All of these fairly unpredictable situations make it undesirable to maintain a distributed storage model that assumes continuous availability.

Furthermore, MIK requries that live previz, which was pursued by several LADIO partners in a previous project (POPART, H2020 #644874), is compatible with data collection.

Instead of relying on an infrastructure that makes decisions about the allocation of scarce WiFI network resources, LADIO partners agreed to follow a more controlled concept that puts a person (typically the DIT) in charge of making replication decision from the SetBox.

Instead of interpreting the CamBox / MiniBox units as equal peers and use a P2P storage concept, we interpret them as Internet-of-Things (IoT) devices that can be deployed freely, discovered by a compatible application, and whose function can be requested by remote units. The application requesting these services in this model is running on the SetBox. This maintains the original dataflow design of the LADIO application (Figure 1 and application section 1.3.4), in fact, it allows for a much more comprehensive implementation.

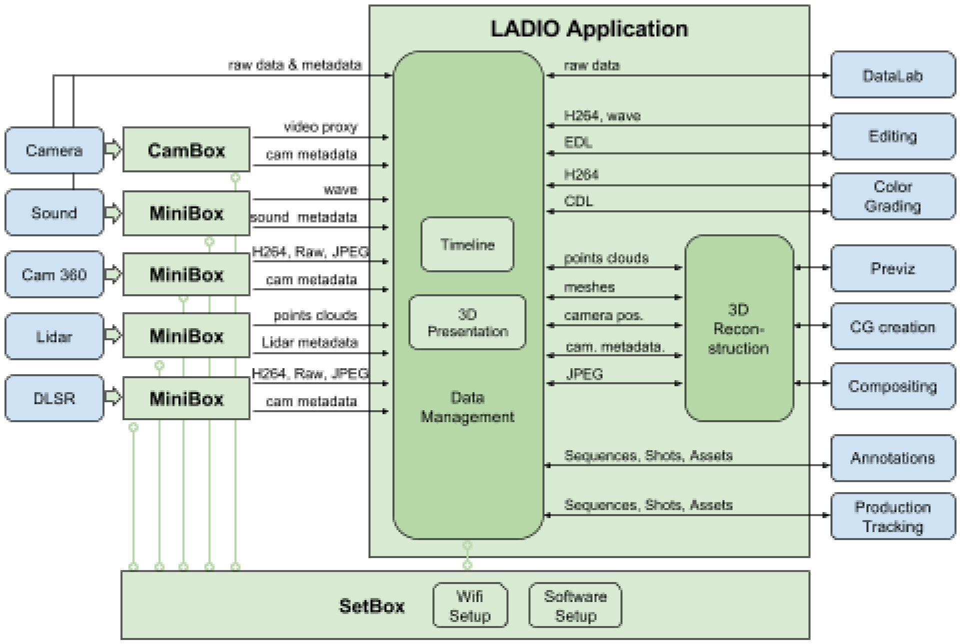


Figure 1: LADIO dataflow (from LADIO application section 1.3.4)

Whereas metadata is always delivered to the SetBox, which will host the central database for a all information related to a Scene on set, it becomes a human operator’s decision whether data from a particular CamBox / MiniBox should be (a) live streamed to the SetBox, (b) bulk transferred after a Take, or (c) whether data transfer will be handled by a collection of SD media or other removable storage media.

Devices are detectable using an IoT middleware, AllJoyn, which firstly allows all boxes to announce their presence in a WiFi network using multicast DNS. Their announcements are formulated in an XML datastructure that contains information about the box’s recording devices (camera, 360 cam, microphone, …) and their feature set. Control messages sent from the SetBox (acting as an AllJoyn client) are formulated as REST queries to an nginx web server running on the CamBox / MiniBox. This is included in more detail in Deliverable 1.2.

WP2 - Data Model and API

T2.3: File and database representation for models and metadata (planned duration months 7-9)

The LADIO data model as defined by D2.1 has been implemented as an XML schema that derives directly from the EBUCore data model. The XML schema has been validated in its RDF form using the W3C validator and is available from:
https://github.com/griwodz/ladioproject/ontologies/ladio

The information about relative camera positions is encoded in metadata that can be associated with a sequence of frames in a MediaResource, which may be as short as a single frame. As decided in D2.1, we rely on the MediaInfo resources that have been adopted by the EBU for encoding per-frame meta information, and where we have given input to the EBU working group that was taken into consideration in selecting to “per frame-sequence” subexpression of MediaInfo. To express relative position encoding of one camera to another, we could use either the 4x3 matrix approach using the [R|t] (rotation and translation) expression, the Essential matrix expression that is encoded in a 3x3 matrix, the Fundamental matrix that is also encoded in a 3x3 matrix, or a quaternion, which consists of 4 complex numbers. We settled on the Essential matrix expression, which is most frequently used when applying transformations on calibrated cameras.

Unfortunately, we did not find a sensible and readable way in which the original MediaInfo schema, available from https://github.com/MediaArea/MediaAreaXml.git can be supplied with extended attributes through inheritance extensions. We are therefore currently copying the mediainfo schema into ladioproject. We will approach the maintainers with this extension when we can actually demonstrate use of this feature in a complete system. It is also found in:
https://github.com/griwodz/ladioproject/ontologies/ladio

Currently still in separate branches of our OpenMVG, we have implemented code that is useful in branches of the OpenMVG 3D reconstruction pipeline when 360 videos are integrated. In the *dev_rig* branch, code that loads a 360 panorama in dual fisheye representation and uses the knowledge of the camera structure to describe the pair of cameras as a static rig, storing the information in Alembic files. This fixed structure is exploited later in the pipeline. The reason for assuming a dual fisheye is that all partners are currently using the Ricoh Theta, which consists of two fisheye cameras. Its raw image output, which is most useful for reconstruction, is consequently a dual-fisheye image. The second piece of code is located in the branch popart_develop_de265 and translates videos from the currently proliferating equirectangular projection (which is actually also the primary output of the Ricoh Theta) into a sequence of CubeMap-projected images. Since the image plane shape, radial distortion and focal length are implicitly known after this transformation, it become feasible to use the groups of 6 output frames in the regular OpenMVG pipeline. Both elements are available in:
https://github.com/alicevision/openMVG.git

Feasible LiDAR solutions have not become available so far in the project. Consequently, LADIO has used the existing datasets that are generated from Velodyne scanners. These are not following any standard file format, but require the user to process network packet dumps. The code that processes these packet dumps and creating a point cloud has already been contributed to the Point Cloud Library (PCL), and is documented in http://pointclouds.org/documentation/tutorials/hdl_grabber.php . Since LADIO uses PCL intensely for our 3D registration work, it has been most feasible to use this library.

The idea of implementing an Alembic encoding of accuracy information within the WP2 timeline has appeared to be problematic. Since the extracting and use of accuracy information is still actively researched and developed, it has not been feasible to implement a fixed binary format. Instead, this information is so far stored in an ad-hoc textual format generated by the software in the private repository https://github.com/alicevision/uncertainty . Its integration into OpenMVG has not started yet and the initial assumption about means of encoding have not been verified.

The use of accuracy information on a large scale is large limited by the available memory in today's computers, and the need to move it quickly into and out of compute memory. Our activity on direct GPU-disk communication as well as direct mapping of devices between computers over PCIe interconnection links has been part of this task. The direct GPU-disk communication is going to be integrated into the CMPMVS code in WP4. The current code is working (though not always entirely stable) and is available at https://github.com/enfiskutensykkel/ssd-gpu-dma

WP3 - Advanced 3D reconstruction

T3.1 - SfM : LIDAR integration

Registering all LiDAR clouds in a common 3D frame (**3D-3D registration**). Many algorithms have been tested such as the classical ICP and sparse ICP (ICP referring to "Iterative Closest Point"). Sparse ICP provides results better than ICP. ICP is sensitive to outliers and missing data often observed in 3D scans. Sparse ICP is an extension algorithm of ICP to cope with these difficulties.

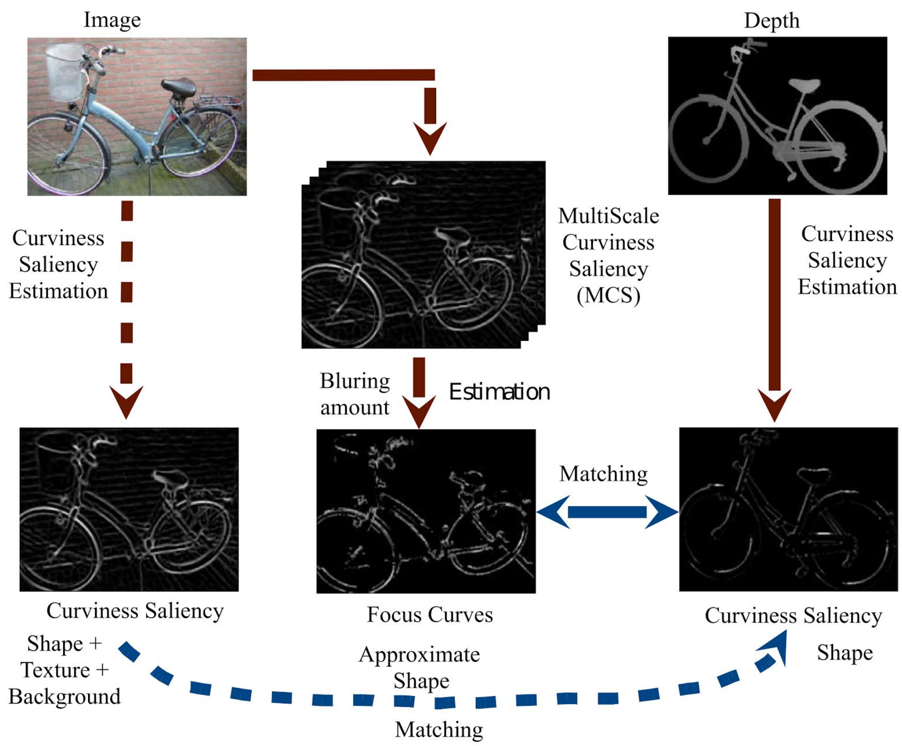
About registering one SfM point cloud to the LiDAR point cloud (**3D-3D registration**). The two main difficulties of registering SfM to LiDAR point clouds are the difference in the density and scale. For density, the two points are needed to downsample (voxel grid is used) with different voxel sizes. For downsampling, Principal Component Analysis (PCA) is used to provide an approximation the scale difference between the two point clouds.

The source code is available on our private github:
https://github.com/alicevision/pointCloudAlignment

Regarding the proposed method of **2D/3D registration** (see figure T3.1.1) we represent each 3D model with a set of depth images rendered from a sphere around the object. Curviness saliency features are extracted from the rendered depth images and compatible features (i.e., with same order derivatives) "in focus" are extracted from the intensity image and then matched. The camera pose is that of the depth image that best matches the intensity image.

.

We worked on the theoretical aspects of the proposed curviness saliency feature. For the 3D model, we defined the curviness saliency feature as a function of the difference of the ordered principal curvatures of the depth surface $\mathcal{D}(\mathbf{x}) = \left[\mathbf{x}^\top, Z(\mathbf{x})\right]^\top$, where Z refers to the 3D point depth w.r.t. a given camera frame, at each 2D point $\mathbf{x}$. We want to detect features in the intensity surface $\mathcal{I}$ and check whether they are good candidates to be matched to detected curvilinear features in the depth surface $\mathcal{D}(\mathbf{x})$. We show that that one key issue here is that detected features in $\mathcal{I}$ can be matched to detected features in $\mathcal{D}$ on the condition that both are based on measurements with the same order of derivation in $Z(x,y)$, in order to yield a ``compatible'' matching that ensures repeatability. The fact that $\mathcal{I}$ depends on $Z(x,y)$ and its derivatives up to order-1 entails that the detection of features in $\mathcal{I}$ must rely on order-1 variations of the surface $\mathcal{I}$, e.g., on its differential along some adequate direction. To this end, for the intensity image, we showed that we can define the curviness saliency feature as the norm of the image gradient which has such property. We obtained the best results so far.



FigureT3.1.1 : Outline of the proposed method for registering an intensity image to a 3D object, using "curvilinear multiscale features in focus".

T3.2 - SfM: 360° cameras and camera rigs integration (planned duration months 3-9)

As explained in details in the Deliverable 3.2, we have added a new software to preprocess the 360 images to integrate them into the pipeline. Most of the 360 cameras are based on a combination of multiple image sensors that are combine together to generate the 360 image. To integrate them into the pipeline as standard images, we split the 360 image into one image per sensor. The 360 images can be in equirectangular or dual-fisheye format. So the preprocessing software has 2 modes:

- Equirectangular

Figure 3 : Same image with cut lines in overlay.
We can see that a margin between images is skipped to avoid stitching areas.

- Dual-fisheye

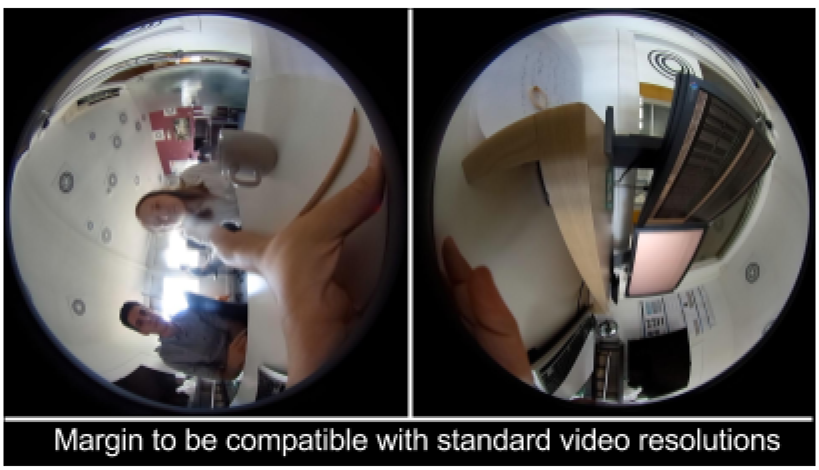


Figure 6 : the 2 output images exported and declared
as a rig of 2 cameras with fisheye optics.

As explain in more details in D3.2, we have also added the support for rig of cameras. It has required a major refactoring, as it is a basic notion of the "View" class.
The notion has been added in the following steps:

- imageListing

The input json project file can now contains groups of groups of input images to represent the images from each camera of the rig.

- Structure-from-Motion

We have modified the "incremental" pipeline to take the rig into account but we have not added the rig support into the "global" pipeline as it is not used in production. First, it changes the initial image pair selection, as it creates some particular configuration case (we cannot use different cameras of the rig at different times for the initialization). It also change the next best view selection, with the same kind of

particular configurations. After each successful resectioning, we check the status of the rig and update the rig pose or the sub-pose according to what is already solved.
Finally, we have added the rig notion into the Bundle Adjustment to declare the new constraints between cameras on rig poses and sub-poses.

We have good preliminary results that allows to retrieve more cameras and the number of 3D points reconstructed is slightly increased. But the increase of the number of 3D points is less than expected and we have also faced some datasets where the usage of rig can failed and drastically reduce the quality achieved without.
We plan to implement a new strategy, a bit more complex, to localize cameras individually and only declare the rig when we have enough cameras to declare the rig with a good enough initialization.

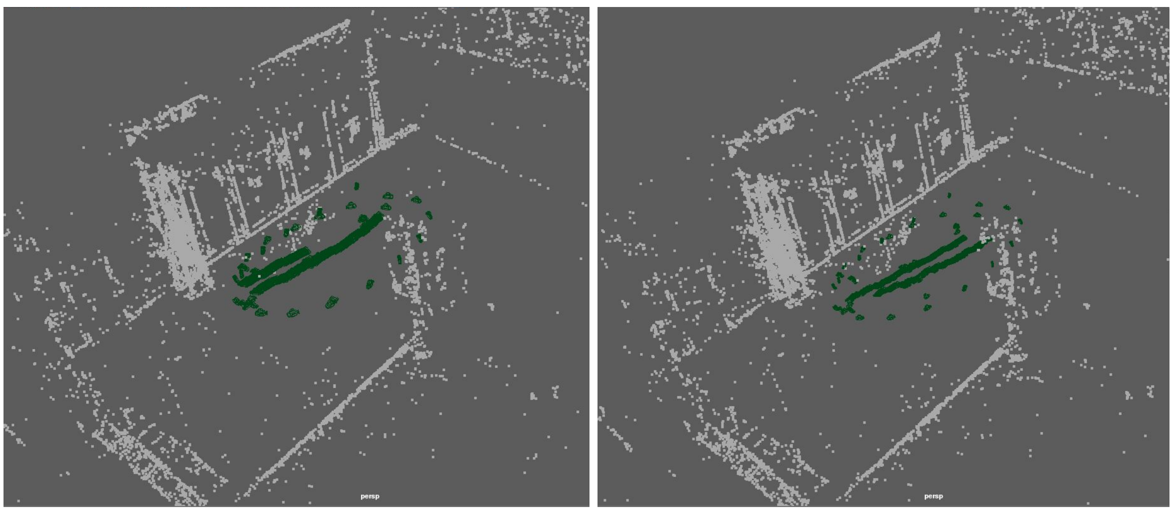


Figure 1 : Dataset eth3d_electro: reconstructed point cloud and cameras, without rig (left), with rig (right)

The source code is available on our public github:
https://github.com/alicevision/openMVG/commits/dev_rig
And it will be merged soon in the "develop" branch.

T3.3 - Improve SfM accuracy and precision
(planned duration months 1-9)

Bundle adjustment (BA) is performed through iterations of the Levenberg-Marquardt algorithm. At each iteration, the normal linear equation $H\Delta x = g$ must be solved for the parameter-vector increment Δx, where H is the "regularized Hessian matrix" and g some vector. In order to deal with large-scale BA data, the so-called "Schur Complement" technique consists in replacing, using Gaussian elimination, the normal linear equation by a system of two new linear equations. Both are linear in two subvectors Δy and Δz taken from the partition $\Delta x = [\Delta y^T, \Delta z^T]^T$ and must be cheaper to solve.

- Parameter ordering

One must carefully partition $\Delta x = [\Delta y^T, \Delta z^T]^T$ in such a way that the resolution of the two new linear equations has a true lower complexity. This issue is referred to as *parameter ordering* in Ceres Solver. Theory suggests that the choice of Δy and its own internal ordering have a significant of impact on the efficiency and accuracy of the algorithm. Even if Ceres Solver provides an option (default option) for automatic *parameter ordering*, we re-ordered the whole set of parameters such that the first parameter group to be eliminated (Δz) is that of the parameters of the m points (also called 3D landmarks) so the parameters Δy to be estimated first coincide with the parameters of the n cameras.
We applied two "parameter ordering" methods into the code and tested it on 4 datasets: (1) the Ceres Solver method by default ("automatic") *vs.* the proposed one ("ordering"). Results are given in figure T.3.3.1.

Dataset name	m = # points	n = # images	Time (s) "automatic"	Time (s) "ordering"	Time saving
Lastpant	*375 000*	*1050*	*7000*	*6350*	***9.3%***
Lou	*285 000*	*510*	*1600*	*1480*	***7.5%***
Cirque	*120 000*	*500*	*2800*	*2700*	***3.6%***
Levallois	*120 000*	*540*	*2550*	*2500*	***2%***

Table T.3.3.1: Results of parameter ordering on 4 datasets.

- Local Bundle Adjustment

In OpenMVG, the incremental SfM pipeline step integrates a new group of unprocessed views at each step. The corresponding cameras are resected and the new 3D points are added by triangulation. The whole 3D reconstruction (old plus new data) is then fully refined by Bundle Adjustment (BA). The time spent in this minimisation problem increases exponentially all along the reconstruction (the number of parameters to estimate basically is $9\#cameras + 3\#points$). To decrease the computation time allowed to BA, we proposed a "local approach" that reduces the number of parameters to be refined similar to [Mouragnon:2006][Mei:2011].

All the reconstructed cameras are stored in a "camera graph" where vertices are cameras and edges link camera pairs only if the two cameras have a sufficient number of matched measurements of 3D landmarks. The Local Bundle Adjustment (LBA) classifies each camera as belonging to a *ActiveRegion*, *PassiveRegion* or *SkippedRegion* class according to its distance to the newly resected cameras. We define LBA as the problem of jointly estimating the camera parameters *only* associated to views in *ActiveRegion* and the parameters of 3D landmarks *only* seen in “active views” by minimizing a (robust) sum of squared reprojection errors in *both* *ActiveRegion* and *PassiveRegion*. In other words, the LBA algorithm refines all parameters in *ActiveRegion*; Parameters included in the *PassiveRegion* are set to constant. The parameters included in the Skipped Region are not added to the algorithm.

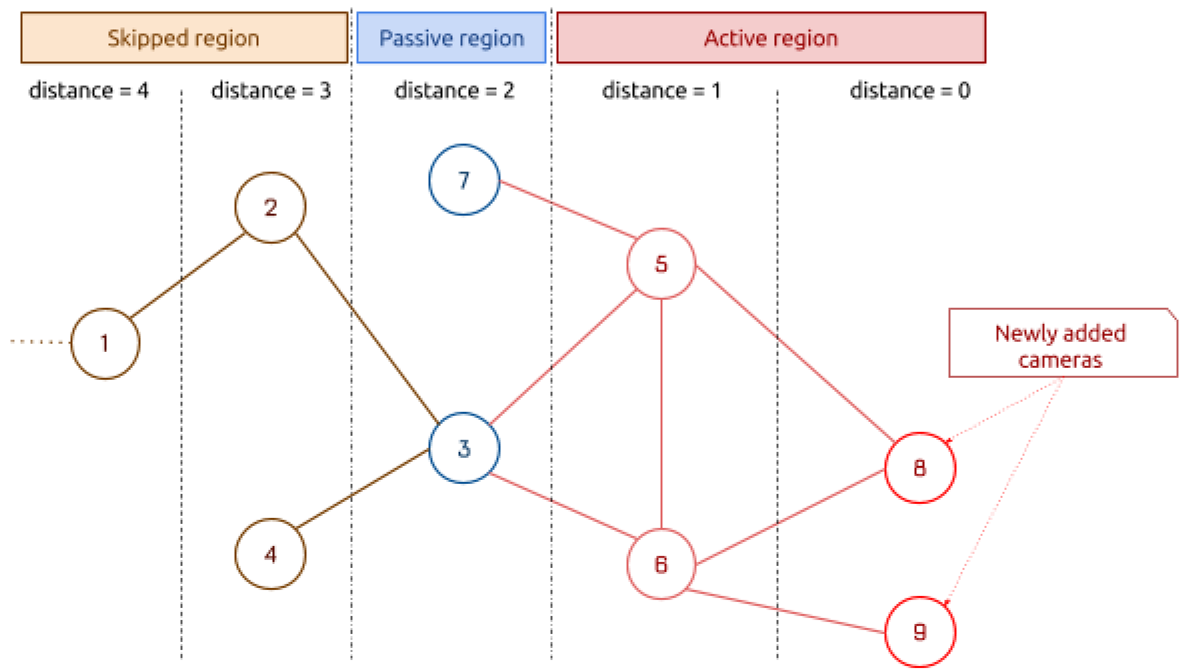


Figure 1 - Example of regions distribution.

This technique reduces drastically the number of refined parameters at each incremental step, ignoring the parameters far from the newly added camera. In fact, we can argue that the ignored parameters are usually very little affected by standard BA refinement so LBA thereby reduces the time spent in each BA part as seen in in figure 2 with very close reconstruction accuracy.

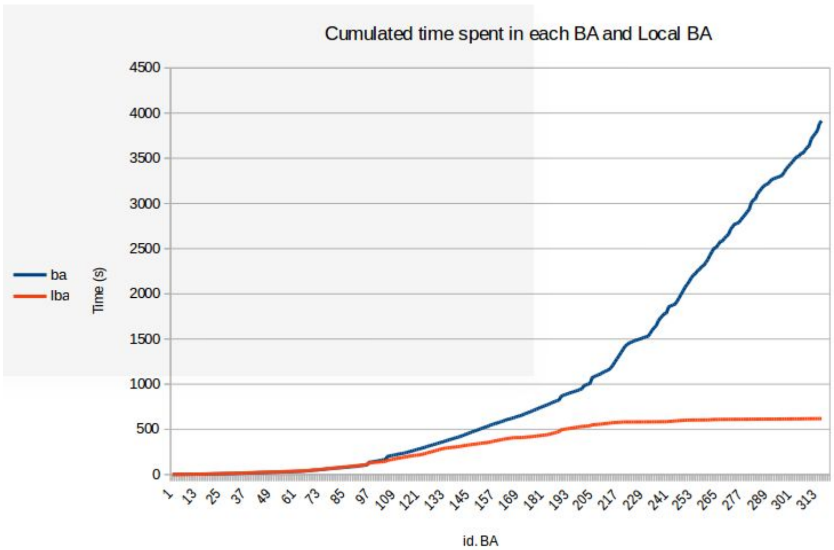


*Figure 2. Cumulated time spent in the refinement for the **Cirque** dataset, with the classic BA (blue) and the Local BA (red).*

The Local BA has been tested on several datasets of images to assess the gain in running time [Table 1] and its impact on the reconstruction accuracy [Table 2].

Dataset name	*Lastpants*	*Levallois*	*Cirque*	*Lou*
#images	1077	595	543	513
#intrinsics	13	28	27	2
BA	3 h 28	47 min	1 h 5	1 h 7
Local BA	33 min	8 min	10 min	20 min
Gain	**x 6.25**	**x 5.6**	**x 6.3**	**x 3.4**

Table 1. Gain in speed evaluation of the Local BA approach against the classic BA method.

Dataset name	*Lastpants*	*Levallois*	*Cirque*	*Lou*
#images	1077	595	543	513
#intrinsics	13	28	27	2
#poses BA	1044	534	494	513
#poses Local BA	1039	494	479	513
Ratio	**99.5%**	**92.5%**	**97%**	**100%**

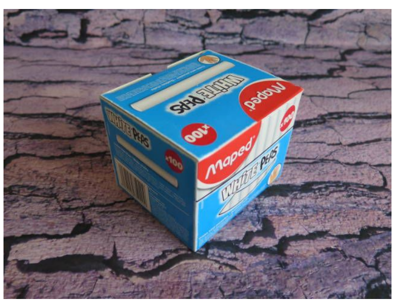
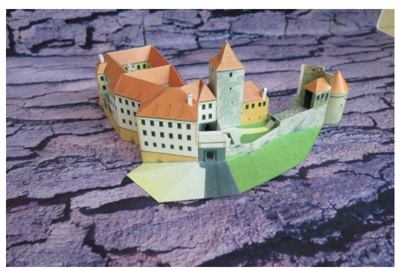
#points BA	376167	120481	118878	288166
#points Local BA	373530	114594	113626	285780
Ratio	**99.3%**	**95.1%**	**95.6%**	**99.2%**

Table 2. Comparison of the reconstruction quality between the Local BA approach and the classic BA method.

T3.4 - SfM : Multi-body

(planned duration months 9-14)

We have started collecting data for a full 3D reconstruction of an object moving in-front of a static background with convenient markers for camera reconstruction. Three data sets have been acquired, see figure below, and standard pipelines (COLMAP, OpenMVG, YASFM) are currently being tested to evaluated what results can be obtained by the standard approaches and to evaluate how suitable are the data sets for further development of Multi-Body SfM and MVS.

		
“Box” data set.	“Daliborka” data set.	“Lican” data set.

WP4 - Detailed 3D representation

T4.1 MVS Baseline

(planned duration months 3-9)

As presented in details in the Deliverable 4.1, we have continued the refactoring of the MVS pipeline started in Q1.

In Q3, we have rewritten the tetrahedralization which was based on the CGAL library (licensed under GPL) by new code based on the GEOGRAM library (licensed under the 3-clauses BSD license).

GEOGRAM Library : http://alice.loria.fr/index.php/software/4-library/75-geogram.html

We have rewritten some mesh cleaning procedures to remove code using methods from CGAL that have no equivalent in GEOGRAM. We have also replaced the Maxflow library which was released under GPL, by the boost::graph implementation (released under the Boost license).

Boost Graph Library : http://www.boost.org/doc/libs/1_66_0/libs/graph/doc/index.html

Finally, we have removed the CUDATemplates library released under GPL by new code directly based on CUDA.

We have also refactored the meshing and texturing algorithms and improved the code quality: *use const*, *use references* instead of pointers when possible, fuse duplicated code and add parameters to replace the hard-coded file paths originally used in low level functions. We added an option to retexture an external mesh from the original images, by remapping the visibilities from the original mesh provided by the reconstruction to the customised user mesh. As presented in D4.1, we made extensive tests on public datasets and production datasets.

It is available on our private github and in the deliverable D4.1
https://github.com/alicevision/CMPMVS

As explained in the deliverable D4.1, we decided to postpone the release of the source code on GitHub, in order to merge our SfM and MVS pipelines into one repository and to release it with the new version of Meshroom to maximize the marketing impact of this open source release.

T4.2 MVS GPU Optimization
(planned duration months 10-18)

SRL contributed early to MIK's and CTU's efforts on T4.1 because SRL requires a tight collaboration to learn about and understand the algorithms and code of CMPMVS. This early effort led to some insights about possible improvements, whose delayed implementation would also have penalized efforts in T4.1.

First, SRL has removed the CMPMVS dependency of the CUDATemplates library released under GPL. Initially, we have isolated all the used CUDATemplates library data structures, functions and classes. Than, we have wrote the set of wrappers around remaining CUDATemplates code to ensure correctness of the following code replacement and enabling of the data flow analysis. Next, we have created an independent set of replacement functions and classes implement the original CUDATemplates functionality with direct using of CUDA API. Than, we have debugged our new code with a help of specially created raw data comparators for the input and output of the functions and classes' methods being replaced. Finally, we have removed all the code related to the CUDATemplates library, removed all the library references in the CMPMVS and removed all the helping and wrapping code.

Next, SRL continued CMPMVS code exploration and performance analysis. With the main focus on the GPU profiling in terms of individual GPU kernels. Currently profiling of the GPU code leads to an enormous increase in the whole pipeline processing time up to 10 times depending on an used profiler mode. We continued to research in order to find the best profiling strategies can be used efficiently with the CMPMVS pipeline. For the now, we the best candidate is an original nVidia *nvprof* with a disabled CPU and CUDA system calls profiling in a conjunction with the custom timing anchors wrapping time-critical GPU-intensive parts of the code.

WP5 - LADIO Application

T5.1 - LADIO Application Timeline (planned duration months 3-12)

The development in this work package has been seriously delayed while we are conducting our discussions with the Norwegian SME partner (Quine) who we expect to join through a second Amendment.

The implementation of the LADIO backend is tightly integrated with the T1.2 (Distributed Data Management). The LADIO backend is concerned with the control of data collection from the CamBox/MiniBox as well as reception and storage of data and metadata. The technical decisions for data transfer (relying on an IoT concept, data transfer as in the three possible modes live streaming, upload and manual memory card transfer) have been discussed with T1.2.

Contentious points have been the choice of database that would be suited for the data model of LADIO, transfer and storage efficiency, and compliance with EBUCore as a metadata standard. Specifially:

- **Efficient transfer and import of information**:
  It is given that media data including audio and video recordings, would be compressed according to standard compression schemes, in particular versions of MPEG. A majority of metadata information relating to a Take would never exist on the CamBox / MiniBox devices, but only on the SetBox, and therefore not require any transfer. The SetBox assigns a new UUID to a CamBox / MiniBox for every Scene. This allows a straightforward identification of data belonging to a particular scene as well as a device. Furthermore, in case of communication failure on set, the unique device ID that exists for every CamBox / MiniBox can be used to manually associate content recorded by it with a a Scene and Take.
- **Camera metadata as data or metadata**:
  LADIO uses an extended form of the MediaInfo description of camera metadata, which can describe changes on a per-frame basis. The MediaInfo proposal includes dynamic parameters such as gain, aperture, focal and length (of a particular zoom lens), LADIO adds to this relative physical position of the camera as well as the reference for every

frame. In some contexts, this information is considered metadata of video frames, and treated as such. It would be inserted into the LADIO database on the SetBox, it could be used by processing tools through database queries, and it should be exported according to the MediaInfo specification for export to other tools. In other contexts is should be considered data, in the sense of being a time-dependent medium that is synchronized with a video stream. The two streams of information together become part of the post-processing pipeline, not the least because these camera parameters influence how virtual elements must be rendered. The LADIO decision is to treat the transfer of this metadata from a CamBox to the SetBox like data transfer (streamed, uploaded or transferred offline), and use an import process on the SetBox to import it into the database for later XML export.

- **Choice of database**:
  The decision about a specific database (DB) has been a recurring topic in telcos involving MIK and SRL as well as Quine.
  One basic decision referred to the choice of a memory DB versus a disk-backed DB. Essentially, a DB can consider content stored on disk as true reference and RAM as cache, or it can consider RAM as true reference and disk as backup. For LADIO purposes, a disk-backed DB is essential to cope with power less and other uncertainties on set. Partners agreed that the DB should not be a classical relational DB, since large parts of the EBUCore schema are optional when used on set, and since the most suitable fields for primary keys are not known without practical experience. Consequently, a variant of NoSQL DB must be chosen.
  A major point of contention is still the need for integrating the database instance on the SetBox with the databases at the “home site” of a production, ie. whether the database on the SetBox should be considered a replica of a distributed database, a cached part of the database, a separate instance, or a reference instance that is copied to the home site. There is no clear decision yet.
  Three concrete DBs that were proposed are: MongoDB (A mixed model memory/disk DB that relies on JSON documents both internally and for external interfaces. It has a many language bindings and a large user community. It does not deal with XML export.), RethinkDB (An active memory DB, which enables independent processes to interact through DB operations.), and BaseX (A native XML database used through an XPath / XQuery processor, supporting also REST interfaces.). All of them can in principle solve the LADIO requirements, but the specific implementation requires a serious effort. Discussions are concerned with the advantages and disadvantages of the candidates, including performance, crash-resistance, replicability, ease-of-use, user community, and programmability. The question has not been resolved yet.
- **Point of conversion between proprietary and standardized (EBUCore) form**:
  The LADIO application running on the SetBox must export a standardized form of the data collected on set for use with other applications. This export follows the EBUCore schema. The question is whether the same form should be used for communication between CamBox/MiniBox and SetBox. The advantages would be the avoidance of conversion and the ability to easily import data that had be retrieved manually from a

CamBox/MiniBox in case of problems. Disadvantages are the size of XML representations and the problem of formulating an XML document in a way that is suited both for streaming and storing locally (streaming means either overhead due to repetition or a permanently incomplete document). Since the SetBox must anyway include an import process ("ingest") for reading SD cards, the advantages of exchanging XML between CamBox/MiniBox and SetBox are not convincing. As a side-effect, the importance of choosing a DB with native XML support is reduced.

Task 5.4: 3D Reconstruction benchmark

Two INP trainees (Marc Camillière and Johary Rakotomalala) in second year of ENSEEIHT engineering school are working on the creation of an academic dataset for the registration of a 2D image to 3D model. This data will consist of web-collected data and include:

A. A selection of publicly available 3D object models,
B. A set of synthetically rendered images per object model coupled with camera poses,
C. A set of web-selected real images of object model instances coupled with camera poses.

In A) the 3D object models are taken from sites sharing online free 3D models. In LADIO, the 3D object are given by a set of 3D depth maps, which describes how the original object surface is "shortened" by a perspective viewing. We provide a code along with the dataset for generating the depth map from the 3D models, given a camera pose. On the other hand, B) and C) will be used as test intensity images. In B) camera poses are known and will correspond to ground-truth. In c) the camera poses camera poses will be determined by manually registering the real images to the 3D model and will be considered as ground-truth.

WP6 - Management

T6.3 - Platform for dissemination of public project information

Discussions with a new SME partner who would take over a large part of WP5 work from SRL after joining the project has consumed considerable time and attention.

Official project website (http://ladioproject.eu) has been updated and a new one has been published around Alice VIsion : https://alicevision.github.io/, describing each step of our open source photogrammetry pipeline.

T6.4 - Data management planning

Datasets have been updated by completing the LiDAR scans from Velodyne scanners, studio recordings with the iPhone 6s and the Red Dragon 4kHD: http://ladioproject.eu/#datasets

Workshops

In the third quarter, there have been workshops where the partners met to make progress on the following developments:

- June 19-21th in Toulouse: small workshop with MIK (Fabien Castan) and IRT (Pierre Gurdjos and Clément Debize) to work on T3.3
- June 22-23th in Toulouse: IRT (Pierre Gurdjos and Clément Debize) and MIK (Fabien Castan, Yann Lanthony, Gregoire De Lillo) attended the CNES workshop on “3D Reconstruction” http://cct.cnes.fr/content/animations-du-cct-17
- June 26-27th in Toulouse : small workshop with SRL, MIK and IRT before the plenary meeting
- July 6-7th in Paris: small workshop with MIK (Fabien Castan) and CTU (Tomas Pajdla) to work on T4.1 on the MVS voting strategy

Plenary meetings

We held the 2nd plenary meeting in Toulouse at INP on June 29-30th and the first meeting with the Advisory Board on June 28th. Three advisory board members were available for the meeting, Ben Hagen (Kamerawerk GmbH), Lalo Nielsen (Zentropa) and Jon Michael Puntervold (J.M. Puntervold Filmproduksjon).

LADIO members gave to the advisory board members a project overview, explained the objectives, approach taken, consortium, the structure of the projects and the planned milestones. We presented Quine as a potential new partner to the advisory board, who would take up the project role that was covered by former partner LABO.

Meeting the advisors in Toulouse

The advisory board agreed with the LADIO vision of maintain al potential information form set in a structured, standardized manner. They are predicting an upgrade of the role of the DIT on set, and the importance of keeping track of written reports as well as all recorded raw material. However, beyond this raw material, we should also keep track of material that comes from the edit system.

The discussion made it clear that tracking of recording devices is relevant. Although it is not a central topic of LADIO, it is an inherent feature provided by the LADIO CamBox. An intense discussion dealt with the need for live (visual) feedback for the success of such tracking. The advisors supported LADIO partners' use of marker-based tracking for difficult cases as well as for understanding the scale of scenes, but the wish to avoid markers for easier cases. Assistance by inertia-based tracking for this case is expected to be necessary at this time. The advisors were very interested to motivate LADIO partners to put more effort into live previz.

Synchronization between recording devices, as well as of recording devices with other equipment on set was another discussion topic. It is considered essential for actually using results for reconstruction and insertion of virtual elements. There was general agreement that Ambient units are the reference for synchronizing all kinds of material.

Material itself should be delivered as raw images wherever possible. Experience with 360 cameras (LADIO partners used the Ricoh Theta) has shown dissatisfying stitching of the individual cameras, while the individual cameras can be interpreted as a Rig of barrel-distorted cameras. Not only transformation of 360 cameras is problematic. Advisors confirmed the observation that also high-end DSLR cameras transform images in undocumented ways, and unmodified original material would be desirable. The advisors stated that the average number of cameras on shooting set is 2 cameras. More cameras need more work for lighting. And the cameras keep rolling between takes. The average time that is actually recorded per day is 1.5 hours (maximum 3 or 4 hours), with around 1 useful minute for 25 minutes shot.

It is important to handle both raw and compressed files. While CODEX camera recorder exists and is extremely fast (500 MB/s soon 750 MB/s), it is also too expensive for most productions, and not portable to sites where cheap network connectivity is unavailable.

A challenge raised during the meeting was the acceptance of VFX equipment on set. The advisors made it clear that such equipment must be as small and unobtrusive as possible. Due to the need of mounting part of the system on the camera, it is considered most important to convince the DOP of the need. The easiest way of doing this would be to introduce the equipment in pre-production, where the DOP himself can gain information earlier than with classical VFX. It should furthermore be clarified to the team on set that the additional equipment is meant for the improvement of the post-production pipeline, and neither threatening nor changing their roles on the set. It was pointed out that VFX in movies has become a requirement that influences the earning of movies, the monetary success of the production.

Not only script notes are meta-information. Information about everything that occurs on set can be relevant in assessing VFX requirements, and having a global video recording (360 camera for instance) can be very valuable to retain this information. As an example that was discussed, sudden lighting changes due to clouds, or even shadows cast by a plane, may make a difference in the selection of the takes for a shot, and influence the amount of required VFX work. Preliminary VFX work should be done on set to detect such details, without requiring a later return to the set or heavy work in the VFX studio. The advisors did therefore ask whether LADIO could include additional research on previs with tracking into the LADIO project.

On the subject of tracking, simple solutions like ARKit from Apple were mentioned. ARKit does realtime camera tracking with plane detection (and odometry), however, in spite of a good plane detection, ARKit is not capable of tracking the plans scale and shift. LADIO can rely on our GPU SIFT implemented PopSift as part of a tracking pipeline, but this must be augmented by IMU-based position prediction: by itself, the SIFT trackers can achieve around 1 frame per second (with natural features).

LADIO could report that our lens calibration from 3D reconstructed environments had undergone that first tests. Advisors pointed out that anamorphic lenses are currently “hot” on sets. This kind of lens has not been considered by LADIO yet, and we have to check, as well as doing some real production tests.

LADIO could also present first results of registering SfM point cloud with LiDAR cloud within the LADIO pipeline. A new approach is under development for LADIO, which consists of matching photos with 3D models. It does not require textured models, but relies on structures, which is achievable with LiDARs. The results are promising, but there are still major research challenges.

In terms of the LADIO frontend, the project used its own application, but it includes also an integration with Maya. It would be desirable to integrate the previz system with a game engine (such as Unity), for its ability to rendering into 3D models and interact quickly with them. It was recommended to consider also an integration with tools such as SketchUp for the integration between previz and props construction. In creating such props, the interaction with the real universe is important for the credibility of cinematography (animation, framing, etc…).

Plenary audio / video conferences

In addition to these face to face meetings, the LADIO consortium holds a video conference every 4th week (with an exception in August due to the number of people on vacation) to keep track of the project progress:

- July 12th 2017 - Video conference
- June 14th 2017 - Video conference

Status of deliverables

The following deliverables were submitted in this quarter.

- D1.2 - Distributed Network - Implementation of the Distributed Data Management system. Deliverable submitted on August 31 2017.
- D2.3 - File format implementations verified in LADIO prototypes. Deliverable submitted on August 31 2017.
- D2.4 - Release implementations as open source. Deliverable submitted on August 31 2017.
- D3.2 - SfM integration for 360° cameras and camera rigs. Deliverable submitted on August 31 2017.
- D3.3 - Improve SfM accuracy and precision. Deliverable submitted on August 31 2017.
- D4.1 - Implement OpenCMPMVS. Deliverable submitted on August 31 2017

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	Carsten Griwodz Konstantin Pogorelov Jonas Markussen Gunleik Groven Stian Z. Vrba Kristian Skarseth Elisabeth Andersen	9,00	12,48	9,00	7,79	4,33	1,41	0,00	0,00	3,50	3,15	2,50	1,53	**28,33**	**26,36**
MIK	Benoit Maujean Fabien Castan Yann Lanthony Nicolas Rondaud Jean Mélou Thomas Eskénazi	1,00	1,21	7,00	1,41	6,67	1,84	5,00	10,74	3,50	3,12	1,50	2,0	**24,67**	**20,32**
INP	Vincent Charvillat Sylvie Chambon Simone Gasparini Pierre Gurdjos Geraldine Morin Clément Debize Marc Camillière Johary Rakotomalala	0,00	0,00	0,00	0,00	11,67	13,1	0,00	0,00	0,00	4,00	0,50	1.1	**12,17**	**18,20**
CTU	Tomas Pajdla, Cenek Albl, Michal Polic Oleh Rybkin Pavle Trutman	0,00	0,00	0,00	0,00	7,33	11.26	4,00	7,9	0,00	0,00	0,50	0,00	**11,83**	**19,16**
Total		**10,0**	**13,69**	**16,00**	**9,20**	**30,00**	**27,61**	**9,00**	**18,64**	**7,00**	**10,27**	**5,00**	**4,63**	**77,00**	**84,04**

Figure 10: The cumulative budgeted and contributed person-months since project start for each partner sorted by Work Package.

References

[Abeles:2013] Peter Abeles, “Examination of Hybrid Image Feature Trackers”

[Ferber:2016] Marvin Ferber, Mark Sastuba, Steve Grehl, Bernhard Jung, “Combining SURF and SIFT for Challenging Indoor Localization using a Feature Cloud”,

[Mei:2011] C. Mei *et al.* “RSLAM: A System for Large-Scale Mapping in Constant Time Using Stereo.” IJCV 2011.

[Mouragnon:2006] É. Mouragnon *et al.* “Real TIme Localization and 3D Reconstruction.” CVPR 2006.

[Sibley:2010] D. Sibley *et al.* “Vast-scale Outdoor Navigation Using Adaptive Relative Bundle Adjustment”, IJRS 2010.

[Steedly:2001] D. Steedly *et al.* “Propagation of Innovative Information in Non-Linear Least-Squares Structure-from-Motion”, CVPR 2001.

[Srajer:2016] F. Srajer. Image Matching for Dynamic Scenes. Master Thesis. CTU in Prague 2016. (https://dspace.cvut.cz/bitstream/handle/10467/64771/F3-DP-2016-Srajer-Filip-filip-srajer-diploma-thesis.pdf?sequence=-1&isAllowed=y)

[Krcek:2017] J. Krcek. Multi-Body Structure from Motion. Master Thesis. CTU in Prague 2017. (https://dspace.cvut.cz/bitstream/handle/10467/68618/F3-DP-2017-Krcek-Jan-Master_s_thesis.pdf?sequence=-1&isAllowed=y)