MVA2015 IAPR International Conference on Machine Vision Applications, May 18-22, 2015, Tokyo, JAPAN

# 3D Reconstruction from Photographs by CMP SfM Web Service


Jan Heller
CTU in Prague, FEE
hellej1@cmp.felk.cvut.cz

Michal Havlena
ETH Zurich, IGP
michal.havlena@geod.baug.ethz.ch

Michal Jancosek
Capturing Reality
michal@capturingreality.com

Akihiko Torii
Tokyo Institute of Technology
torii@ctrl.titech.ac.jp

Tomas Pajdla
CTU in Prague, FEE
pajdla@cmp.felk.cvut.cz



## Abstract

*CMP SfM Web Service is a remote procedure call service operated at the Center of Machine Perception (CMP) of the Czech Technical University in Prague. The majority of available procedures are implementations of Computer Vision methods developed at CMP. The service can be accessed through web page and command line scripting interfaces. This paper presents the service through users' perspective, providing a brief description of the CMP Structure-from-Motion methods and the respective job types. We also explain the technology behind the web service and the most common use cases of the browser interface.*



Figure 1: "One-button" reconstruction of LION dataset as seen from the viewpoint of one of the original images. Input image (top left), 3D points created during plane-sweeping (top right), resulting shaded mesh (bottom left), textured mesh (bottom right).

## 1 Introduction

CMP[1] SfM Web Service is a general remote procedure call service operated by Center for Machine Perception. The majority of available procedures (in the parlance of the service known as *job types*) fall—as its name suggests—into the category usually known as *Structure-from-Motion* in the Computer Vision. Indeed, the primary impetus for the development of the service was the need to provide a 'painless' access to Structure-from-Motion methods developed at CMP to Computer Vision and Photogrammetry communities. Nowadays (beginning of 2015), its user base has grown beyond these two communities to contain users from fields such as archeology and architecture. However, the service still remains an academic endeavor without general public access.

## 2 Methods

This section presents a brief overview of the most interesting CMP developed methods accessible through the web service together with the references to scientific papers where the respective methods appeared. For detailed descriptions of the job types of the individual methods in the service, see Section 4. For user convenience, the service contains also so called "One-button" methods which simply run the default sparse reconstruction method followed by the default dense reconstruction method without requesting any further parameter setting. Sample output of such fully automatic processing is shown in Figure 1.

### 2.1 Large-Scale Sequential Data Processing

Sequential datasets containing thousands of perspective, fish-eye, or panoramic images can be reconstructed using a large-scale sequential SfM tool implemented at CMP [21]. The method makes use of MSER [15], SIFT [14], and SURF [2] image features to detect and describe salient image feature points, FLANN [17] to perform approximate nearest neighbour search in the descriptor spaces of consecutive images giving raise to tentative feature matches, and PROSAC [3] with voting to find the largest subsets of tentative matches consistent with the epipolar geometries for internally calibrated cameras. The Dominant Apical Angle criterion [21] is used to select a subsequence of images suitable for incremental sequential 3D model generation by chaining individual epipolar geometries, the remaining cameras are connected to the model later by camera resectioning [7]. Finally, global non-linear optimization, SBA [13] with angular reprojection error, is performed to refine both camera poses and 3D point positions. The overall consistence of the 3D model for long overlapping trajectories is enforced by using loop-closing techniques based on SURF-based visual words and tf-idf scores [18] followed by another run of SBA, see Figure 2.

### 2.2 Large-Scale Unordered Data Processing

Unordered datasets containing thousands of images can be reconstructed using a large-scale randomized SfM tool also implemented at CMP [8]. The method avoids performing costly exhaustive pairwise matching of all images by selecting similar ones using SURF-based visual words and tf-idf scores. The computation

[1]This work was supported by EU FP7-SPACE-2012-312377 PRoViDE Project.

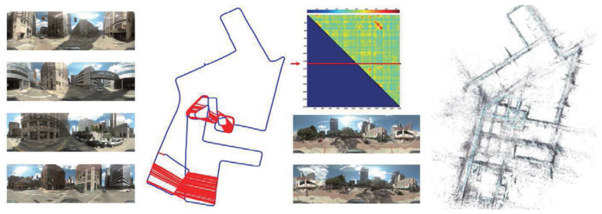


Figure 2: Visualization of the computation of the sequential SfM tool [21]. Input panoramic images from Ladybug camera (left) are used to construct a trajectory consisting of 4,799 camera poses. Further geometric constraints are obtained using SURF-based visual words and tf-idf scores (middle). Resulting 3D model after the final SBA run (right) is consistent with the observed reality.

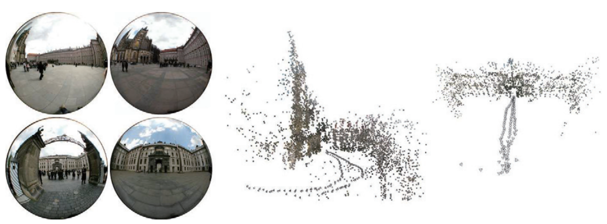


Figure 3: Randomized SfM tool [8]. 4,472 fish-eye images acquired by Nikon FC-E8 lens (left) were input into the pipeline resulting in several 3D models, the largest of them consisting of around 100 camera poses (right).

consists of three stages. First, atomic 3D models from camera triplets are constructed for images with similar tf-idf scores using MSER, SIFT, and SURF features, FLANN matching, and PROSAC with voting again. Several criteria have to be fulfilled in order to accept the created atomic 3D model, namely: (i) the three epipolar geometries between the pairs of images in the triplet must be consistent, (ii) matches accepted by PROSAC must be distributed in the whole image, and (iii) there must be a sufficient number of triangulated 3D points having Dominant Apical Angle larger than a threshold. Triplets containing images with similar tf-idf scores are further merged into larger reconstruction segments and, finally, remaining cameras are connected to the largest segment using camera resectioning. Extensive use of SBA improves the overall quality of the obtained 3D models. See Figure 3 for sample reconstructions obtained by the method.

### 2.3 Classical Unordered Data Processing

Datasets containing up to 1,000 perspective images can be reconstructed using a classical SfM tool Bundler [20] by N. Snavely. The original method performing exhaustive pairwise matching of SIFT image feature points, track construction from matches verified by RANSAC [5], and 3D model construction from the most promising image pair has been improved in several ways: (i) SURF-based visual words and tf-idf scores help to avoid exhaustive pairwise matching, (ii) SURF features with adjustable detection threshold can be used instead of SIFT, (iii) feature detection and matching has been parallelized to run on multiple cores.

Figure 4: Several models created by method [10]. Note that no a priori assumption on the geometry of the scene is being used so even extremely articulated structures like the statue of a lykan (left) or the statue of the dragon (right) can be reconstructed correctly.

### 2.4 Multi-View Dense 3D Reconstruction

Multi-view dense 3D reconstruction tool implemented at CMP [10] can be used to create colored or textured surface models of the scenes from perspective datasets once the cameras are calibrated, making it an optional successive step of any of the aforementioned SfM methods providing external camera calibrations. The method itself is an extension of the method of P. Labatut [12] which is based on the s-t cut of the graph based on the Delaunay tetrahedralization of the point cloud obtained from plane-sweeping. The baseline method accumulates energy to individual tetrahedrons based on the visibility of swept 3D points in the respective cameras. The main improvement of our method over the baseline one lies in the ability to reconstruct so called "weakly-supported surfaces", i.e. the surfaces which do not have strong support by the data, by accentuating small jumps in the energy function. This allows for reconstructing also semi-transparent or poorly-textured surfaces, see Figure 4.

## 3 Technology

CMP SfM Web Service was originally conceived as a simple remote procedure call service. Remote procedure call paradigm means that there is no need to install any code on the user's computer and all the computations are performed on a server. From the provider's point of view, this arrangement provides easier control and management of the code that is always executed on the same hardware and software configuration. However, after introducing the browser interface, CMP SfM Web Service became a proper web application for remote image data management and multi-view stereo reconstruction. The browser part of the service is available at:

`http://ptak.felk.cvut.cz/sfmservice`.

Both browser and command line scripting interfaces work based on the following client/server paradigm:

- A 'dataset' is created on the server and a dataset ID is assigned,
- data is uploaded (typically image files) to the dataset based on the dataset ID,
- an XML file describing the data and various parameters (*job XML file*) is uploaded and a job ID is assigned,

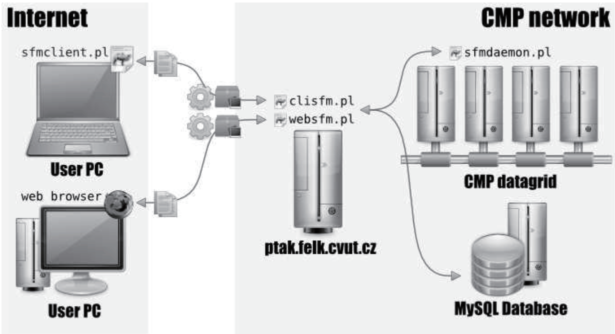
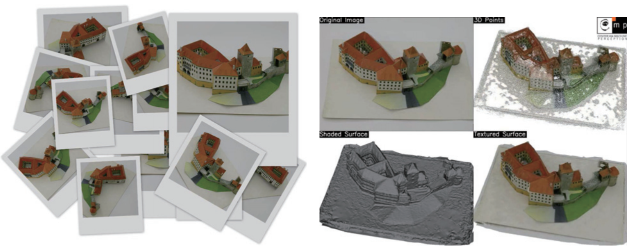



Figure 5: Dataflow in the web service.



(a) (b)

Figure 6: "One-button" reconstruction. (a) An example of an unordered set of input images. (b) Resulting reconstruction (lower right corner) as seen from the viewpoint of one of the original images.

- the procedure (job) is executed on the server,
- the client queries the server on the job's progress using the job ID,
- once a job execution is finished, results (in a form of a zip file) can be downloaded,
- and finally, the client should remove any no longer needed datasets or jobs from the server.

See Figure 5 for more details on the dataflow in CMP SfM Web service. A thorough technical guide to the service can be found in report [9].

### 3.1 Command Line Scripting Interface Client

The scripting interface of the service can be accessed through a command line application called `sfmclient`. The client is a Perl script and should work with Perl5 and libwww-perl collection installed. The client script does not need to be installed; it can be run from any convenient location. Current version can be downloaded from: `http://ptak.felk.cvut.cz/sfmservice/data/sfmclient-0.1.2.zip`.

Technically, `sfmclient` uses HTTP to comunicate with a CGI Perl script residing on a CMP server. SfM web service scripting interface workflow follows the pattern outlined above—data upload, a job XML file upload, job execution, job status querying and download of the results. These steps are referred to as *actions* are identified by the action option of `sfmclient`. However, not every action possible through browser interface is available through the scripting interface. For example, it is currently not possible to list user's datasets or jobs.

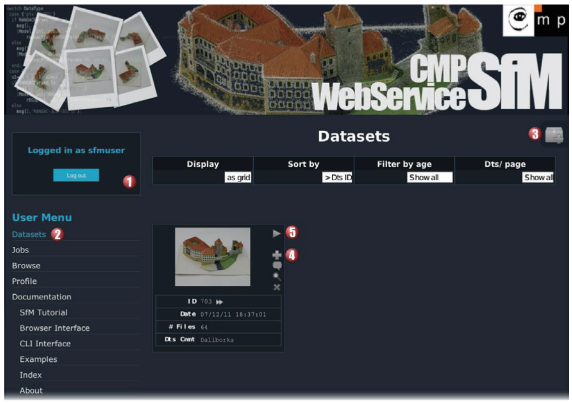



Figure 7: Datasets section.


Figure 8: Jobs section.

### 3.2 Web Browser Interface

The browser interface provides a more convenient way of submitting the jobs and fetching the results once the computation is finished. Here we concentrate on two common use cases: (i) using the "One-button" feature and (ii) submitting a custom XML file.

#### 3.2.1 "One-button" Feature

Reconstruction using the "One-button" feature, *i.e.*, creation of a 3D textured mesh representing an object visible in an unordered set of input images, is the most common use case of CMP SfM Web Service, see Figure 6. Note that the input images need to contain the correct EXIF tags information for the service to be able to properly initialize the internal camera calibration process. As can be seen in Figure 7, running this job in the browser interface is very easy, therefore it can be used by users that have only a little knowledge about 3D reconstruction, too.

One just has to log in (**1**), go to 'Datasets' section (**2**), create a new dataset (**3**), add files to the dataset (**4**), and start the "One-button" reconstruction by pressing icon (**5**). The information on job progress as well as its results can be obtained as shown in Figure 8.

#### 3.2.2 Running Jobs Using Job XML File

The browser interface also supports execution of jobs defined by a custom job XML files. After creating a dataset, uploading data and creating a job XML file, the user should follow Figure 8 to run the job, namely go to 'Jobs' section (**2**) and press icon (**5**). Available job types are described in detail in Section 4.

| Category | Job type | Details |
|---|---|---|
| *Camera Calibration* | `calradv1` | Implements several radial distortion estimation methods and can be used to undistort the images [6, 11]. |
| | `calkinv1` | Implements a calibration procedure for Microsoft Kinect described in [19]. |
| | `sbav16` | A Generic Sparse Bundle Adjustment C/C++ Package Based on the Levenberg-Marquardt Algorithm [13]. |
| | `xbav1` | Implements a constrained bundle adjustment routine for panoramic cameras described in [1]. |
| *Sparse Reconstruction* | `sfmseqv1` `sfmseqv2` | The sequential SfM tool with loop closing [21] for sequential image sets, *e.g.* video sequences. |
| | `sfmonlv1` | The online version of sequential SfM [21]. |
| | `sfm3drv1` | The randomized SfM tool [8] can be used to obtain camera poses and to create 3D models from unorganized image sets. |
| | `bundlerv03` `bundlerv04` | Bundler by N. Snavely [20] modified to allow for avoiding exhaustive pairwise image feature matching. |
| | `vsfmv05` `visualsfmv05` | VisualSFM by C. Wu [22] providing fast reconstructions by employing a GPU. |
| *Dense Reconstruction* | `multireconv1` `multireconv2` `multireconv3` | These jobs are based on the s-t cut of the graph based on the Delaunay tetrahedralization of the point cloud obtained from planesweeping [10]. |
| *One-button* | `onebuttonv1` `onebuttonv2` `onebuttonv3` `vsfmmvsv1` | Combination of `bundlerv04` or `vsfmv05` and `multireconv3` job types for simple sparse and dense 3D reconstruction. The "One-button" feature of the browser interface actually runs `onebuttonv3` job on the respective dataset. |
| *Camera Model Conversion* | `cahv2krc` `krc2cahv` | Converts between CAHV camera model and KRC camera model [4]. |

Table 1: Job types

## 4 Job Types

There are in total 21 job types available in five categories. Table 1 lists the job types as well as provides their short descriptions.

In the following, four job types for sparse and dense reconstructions are briefly presented. The description may not be necessarily up to date, since the interface is subject to change. A more detailed presentation together with the description of the rest of the job types can be found at the service's web site.

### 4.1 Job Type sfmseqv2

The job type `sfmseqv2` is based on the sequential SfM pipeline with loop closing [21] and can be used to obtain camera poses and to create 3D models from sequential image sets, *e.g.* video sequences. It can work both with perspective and omnidirectional images as long as the internal camera calibrations are known or encoded in the EXIF image metadata. The order of the images in the sequence is determined by the lexicographical ordering of their names.

To create a job XML file, one has to provide the ID of the dataset containing the images in the `xmljob` node. Further, the file has to contain one or more calibration nodes, describing the internal calibrations of the cameras that acquired particular images. For omnidirectional images, the pipeline supports two different projection models. The first one is the two-parameter non-linear camera model described in [16] which is the generalization of the equi-angular model. The second supported omnidirectional projection model is the equi-rectangular model often used by 360° panoramic images.

The result is a zip file consisting of several files including the VRML model containing camera poses and a sparse 3D point cloud and various text files which can be used to import the estimated camera calibrations to custom SW.

### 4.2 Job Type sfm3drv1

The job type `sfm3drv1` is based on the randomized SfM pipeline [8] and can be used to obtain camera poses and to create 3D models from unorganized, *i.e.* non-sequential, image sets. Similarly to the job type `sfmseqv2`, it can work both with perspective and omnidirectional images as long as the internal camera calibrations are known or encoded in the EXIF image metadata. The result file is equivalent to the one of job type `sfmseqv2`, too.

### 4.3 Job Type bundlerv04

The job type `bundlerv04` is based on Bundler v0.4 by N. Snavely [20] with a slight modification which allows to avoid matching visually dissimilar image pairs by comparing detected visual words, so it can be used to obtain camera poses and to create 3D models from larger unordered image collections. It can work with perspective images only and the focal lengths of the cameras used need to be known or encoded in the EXIF image metadata.

### 4.4 Job Type multireconv3

The job type `multireconv3` is based on the s-t cut of the graph based on the Delaunay tetrahedralization of the point cloud obtained from planesweeping [10] and can be used to obtain textured dense 3D meshes from image sets. It works with calibrated perspective images only (this job needs results output by a sparse reconstruction job).

To create a job XML file, one has to provide the ID of the dataset containing the images in the `xmljob` node. Further, the file has to contain one or more calibration nodes, describing the internal and external calibrations of the cameras (P-matrices in MATLAB format) that acquired particular images. All input images have to be the same resolution and have to have a respective `calibration` node.

The result is a zip file consisting of several files, most importantly a textured and a colored VRML models of the final mesh, a textured and a colored VRML models of the simplified mesh which is more suitable for fast viewing and a video with a fly-over containing all camera viewpoints. The video can be also viewed directly in the browser interface thanks to conversion to a format playable by Adobe Flash.

## 5 Conclusions

We have presented CMP SfM Web Service, a remote procedure call service operated at the Center of Machine Perception (CMP) of the Czech Technical University in Prague. Especially thanks to the browser interface and the implementation of the "One-button" reconstruction feature, the service is successfully used by people from different fields, mainly cultural heritage preservation (archaeologists) and landscape surveying.

Although the service interface is fixed, the number of available jobs is growing in time, allowing the users to benefit from using the state-of-the-art methods for the given problem, *e.g.* the recently integrated GPU-enabled sparse 3D reconstruction method – VisualSFM by C. Wu [22].

## References


[1] C. Albl and T. Pajdla. Constrained bundle adjustment for panoramic cameras. In *Computer Vision Winter Workshop*, pages 1–7, February 2013.

[2] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. In *Proceedings of European Conference on Computer Vision (ECCV)*, pages 404–417, 2006.

[3] O. Chum and J. Matas. Matching with PROSAC - progressive sample consensus. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 220–226, June 2005.

[4] K. Di and R. Li. CAHVOR camera model and its photogrammetric conversion for planetary applications. *J. Geophys. Res.*, 2004.

[5] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. ACM*, 24(6):381–395, June 1981.

[6] A.W. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 125–132, 2001.

[7] R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2003.

[8] M. Havlena, A. Torii, J. Knopp, and T. Pajdla. Randomized structure from motion based on atomic 3d models from camera triplets. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2874–2881, June 2009.

[9] J. Heller, M. Havlena, A. Torii, and T. Pajdla. CMP SfM Web Service v1.0. Research Report CTU–CMP–2010–01, Center for Machine Perception, K13133 FEE Czech Technical University, Prague, Czech Republic, January 2010.

[10] M. Jancosek and T. Pajdla. Multi-view reconstruction preserving weakly-supported surfaces. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3121–3128, June 2011.

[11] Z. Kukelova and T. Pajdla. A minimal solution to the autocalibration of radial distortion. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–7, June 2007.

[12] P. Labatut, J.-P. Pons, and R. Keriven. Efficient multi-view reconstruction of large-scale scenes using interest points, Delaunay triangulation and graph cuts. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, June 2007.

[13] M.I.A. Lourakis and A.A. Argyros. SBA: A Software Package for Generic Sparse Bundle Adjustment. *ACM Trans. Math. Software*, 36(1):1–30, 2009.

[14] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)*, 60(2):91–110, November 2004.

[15] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *British Machine Vision Conference (BMVC)*, volume 1, pages 384–393, 2002.

[16] B. Micusik and T. Pajdla. Structure from motion with wide circular field of view cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(7):1135–1149, 2006.

[17] M. Muja and D. Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. In *Conference on Computer Vision Theory and Applications (VISAPP)*, pages 331–340, 2009.

[18] J. Šivic and A. Zisserman. Video Google: Efficient visual search of videos. In *Toward Category-Level Object Recognition (CLOR)*, pages 127–144, 2006.

[19] J. Smisek, M. Jancosek, and T. Pajdla. 3D with Kinect. In *IEEE Workshop on Consumer Depth Cameras for Computer Vision (ICCVW)*, pages 1154–1160, 2011.

[20] N. Snavely, S. Seitz, and R. Szeliski. Modeling the world from Internet photo collections. *International Journal of Computer Vision (IJCV)*, 80(2):189–210, 2008.

[21] A. Torii, M. Havlena, and T. Pajdla. From Google street view to 3D city models. In *IEEE Workshop on Omnidirectional Vision, Camera Networks and Non-classical cameras (ICCVW)*, pages 2188–2195, October 2009.

[22] C. Wu. Towards linear-time incremental structure from motion. In *International Conference on 3D Vision (3DV)*, pages 127–134, 2013.