EUROGRAPHICS Workshop on Graphics and Cultural Heritage (2014)
R. Klein and P. Santos (Editors)

Short Paper

Cloud-based 3D Reconstruction of Cultural Heritage Monuments using Open Access Image Repositories

Andreas Hadjiprocopis[†a], Konrad Wenzel[b], Mathias Rothermel[b], Marinos Ioannides[a], Dieter Fritsch[b], Michael Klein[c], Paul S.Johnsons[c], Guenther Weinlinger[c], Anastasios Doulamis[d], Eftychios Protopapadakis[d], Georgia Kyriakaki[d], Kostas Makantasis[d], Dieter Fellner[f], Andre Stork[f], Pedro Santos[f]

[a]Cyprus University of Technology, Digital Heritage Lab, Cyprus; [b]University of Stuttgart, Germany; [c]7Reasons GmbH, Austria; [d]Technical University of Crete, Greece; [f]Fraunhofer Institute for Computer Graphics Research IGD, Germany

Abstract

A large number of photographs of cultural heritage items and monuments is publicly available in various Open Access Image Repositories (OAIR) and social media sites. Metadata inserted by camera, user and host site may help to determine the photograph content, geo-location and date of capture, thus allowing us, with relative success, to localise photos in space and time. Additionally, developments in Photogrammetry and Computer Vision, such as Structure from Motion (SfM), provide a simple and cost-effective method of generating relatively accurate camera orientations and sparse and dense 3D point clouds from 2D images. Our main goal is to provide a software tool able to run on desktop or cluster computers or as a back end of a cloud-based service, enabling historians, architects, archaeologists and the general public to search, download and reconstruct 3D point clouds of historical monuments from hundreds of images from the web in a cost-effective manner. The end products can be further enriched with metadata and published. This paper describes a workflow for searching and retrieving photographs of historical monuments from OAIR, such as Flickr and Picasa, and using them to build dense point clouds using SfM and dense image matching techniques. Computational efficiency is improved by a technique which reduces image matching time by using an image connectivity prior derived from low-resolution versions of the original images. Benchmarks for two large datasets showing the respective efficiency gains are presented.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Digitizing and Scanning— I.3.7: Three-Dimensional Graphics and Realism— I.3.7: Virtual reality— I.3.8: Applications—

1. Introduction and Contributions

Photogrammetry aims to reconstruct the 3D geometry of an object based solely on 2D images. Structure from Motion (SfM) algorithms [HZ04] use data from unstructured sources, e.g. photographs or video taken by off-the-shelf cameras in order to recover the structure of a scene in the form of a 3D point cloud, as well as the camera positions. Additionally, the amount of photographs available online and for free in social media sites and Open Access Image Repositories (OAIR), such as Flickr and Picasa, is enormous. One challenge in applying SfM to images harvested from the web is that camera characteristics, resolution and content of these images may be uncertain. However, the presence of metadata inserted by camera, user and OAIR in the form of Exif tags improves search accuracy. Additionally, comparing the features of $O(n^2)$ image pairs in order to establish connectivity among n images is time consuming because, for OAIR, n is very large and image connectivity sparse. Our main goal is to enable architects, archaeologists, but also the general public to reconstruct in 3D, views of historical monuments from images harvested from the web in a simple, robust and computationally efficient manner on desktop or cluster computers or via a cloud-based service. Focusing on cloud-based environments and simple-to-use web front ends is suitable for the end user with limited computational resources. Beside the benefit of decentralised processing and storage, such cloud-based services will have significant impact on the documentation of cultural heritage when professionals in the field share high-quality 2D images

† a.hadjiprocopis@cut.ac.cy, andreashad2@gmail.com

DOI: 10.2312/gch.20141316

of historical sites and their 3D models, produced and enriched with metadata collectively. This paper describes our implemented Image Search and Retrieval Engine and 3D Reconstruction Pipeline with optimisations to shorten the image matching process. It is capable of producing 3D dense point clouds from hundreds or thousands of photographs retrieved from Open Access Image Repositories. The system can be run on a variety of hardware; from medium- to high-end desktops to dedicated computer clusters depending on the number of input photographs. Its usage is as simple as defining a few keywords for searching for images in OAIR. Our system will then automatically analyse image Exif data, retrieve relevant and suitable images based on tags and produce a 3D reconstruction of the scene as a dense 3D point cloud. The implemented 3D Pipeline uses public domain software: [VF08] for SIFT (Scale Invariant Feature Transform), Bundler for Bundle Adjustment [SSS08b] and SURE [RWFH12] for dense point cloud derivation.

Our contribution is threefold. Firstly, we provide a software tool which seamlessly and robustly searches and downloads photographs and then produces dense 3D point clouds of the scene they depict. Our implementation is highly parallelisable and simple to use. Image search is quite powerful because image metadata is also searched. Thus, relevancy of search results can be improved by optionally specifying a geographical bounding box, date period, image resolution, camera model and characteristics, etc. Additionally, we have already used our software tool as the back end of a cloud-based service, currently under testing. Secondly, we reduce image matching time by utilising lower-resolution versions of the original input images to estimate image connectivity priors and investigate how this method affects quality and speed [AWWF]. In particular, image matching time can be shortened by first working on a low-resolution version of the input images in order to quickly establish image connectivity. When the images are processed at their original resolution, this image connectivity prior helps to avoid matching un-connected pairs. Based on experiments presented in this article, this method can reduce processing time by up to 67% with little quality loss, depending on input images. Thirdly, we enrich the final 3D product with additional metadata such as historical context and links to related artefacts and deliver it to Digital Libraries, for example EUROPEANA and UNESCO Memory of the World, and Virtual Museums. We provide benchmark results over two image datasets and for various image reductions showing the improvement in processing time versus cameras "missed" – an indication of quality loss. The use of our implemented pipeline from monument search to 3D reconstruction is demonstrated with the presentation of the 3D dense point cloud from the complete search results of the keyword *Colisseum* from Flickr.

2. Related Work

The first system to apply SfM algorithms to online, unorganised photo collections was *Photo Tourism* [SSS06] which computed the viewpoint of each photo in the collection and a sparse 3D model of the scene. Scalability was an issue because of the exhaustive pairwise image matching employed. An improvement of the matching process was introduced by Snavely et al. [SSS08a] who constructed skeletal sets of images whose reconstruction approximates the full reconstruction of the whole dataset. A similar project, *Building Rome in a Day* [AFS*11] performs dense modelling from internet photo collections consisting of millions of images. The increased speeds obtained are due to some refinements e.g. filtering by early 2D appearance-based constraints as well as parallelising the matching process and using Approximate Nearest Neighbour (ANN) [AMN*94] search whose results are pruned by RANdom SAmple Consensus (RANSAC) [FB81]. An important factor for the impressive performance of this system is the utilisation of graphics processors and parallel computation on multi-core computer architectures. Frahm et al., in [FFGG*10] improved the performance by an order of magnitude larger dataset by combining image appearance and colour constraints with 3D multi-view geometry constraints. This provides an initial registration to be used in computing the dense geometry of the scene using fast plane sweeping stereo [YP03]. In the framework of partitioning methods, Farenzena et al., [FFG09], introduced a new hierarchical scheme for SfM where matched images are placed in a hierarchical cluster structure, thus revealing the sub-problems which can be dealt in parallel.

3. Search, Download and 3D Reconstruction Pipeline

Today there exist a variety of OAIR over the Internet, for example Flickr and Picasa. Typically they offer an API for searching metadata and retrieving stored images. Using images from OAIR for 3D reconstruction presents some challenges: (**a**) size, resolution and quality of images varies enormously; (**b**) irrelevant objects and people obscure the main subject; (**c**) images matching the geo-location requirement do not necessarily contain the site in question because the camera is looking somewhere else, usually away; (**d**) dates set in cameras and inherited by the image are not always accurate; (**e**) users do not always tag or describe their images or they do so incorrectly; (**f**) Image processing software may be used to alter the photographs.

The steps in our 3D reconstruction pipeline are: (**1**) Convert all input images to a common lossless format and calculate sensor width in pixels for each image. (**2**) Reduce the size of each input image by factor r. (**3**) For each image, extract SIFT features. (**4**) Match SIFT features among all image pairs, $O(n^2)$, yielding a connectivity matrix (CM_1) to be passed to Bundle Adjustment and be pruned with RANSAC. (**5**) Perform BA using Bundler, [SSS08a]. This is done incrementally adding a few images at every iteration, resulting in the determination of camera orientation and sparse scene geometry. The connectivity matrix CM_1 will be used initially. (**6**) Analyse the output of BA and deduce connectivity matrix CM_2. (**7**) Repeat step 3 but this time with the original, full-resolution images. (**8**) Using our

own feature matching implementation (based on KeyMatchFull.cpp of [SSS08a]) match features among only those pairs of images which are connected according to CM_2, calculated in step 6. This will result in connectivity matrix CM_3. (**9**) Perform BA on the original images using CM_3 calculated in step 8. (**10**) Reconstruct the scene as a dense point cloud utilising the sparse scene geometry derived in the previous step. For this purpose, we utilise SURE [RWFH12].

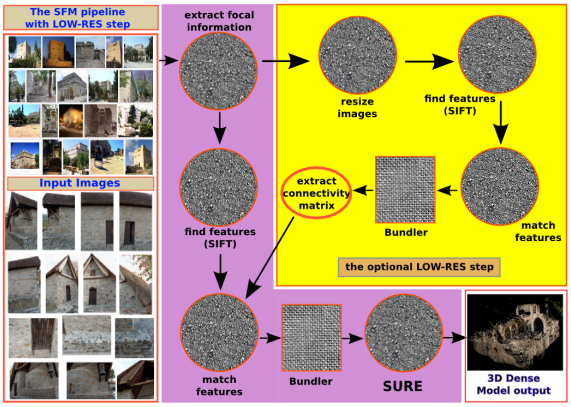

Figure 1: *The implemented 3D Reconstruction Pipeline*

Figure 1 depicts the steps in our pipeline. The procedure in the yellow/light area (Low-Res step) applies steps 2 to 6 on lower-resolution versions of the original images. This stage can be quite fast depending on scaling factor r. The purple/dark area represents a typical SfM pipeline modified to accept an image connectivity matrix in order to speed up the image matching stage by skipping the image pairs which are not connected according to the CM_3 calculated in step 8. The implemented 3D Reconstruction Pipeline can exploit all the available processing cores at its disposal. Optionally, the SIFT detection stage can be performed on GPU.

4. Results

Two image sets are used in order to quantify improvements in the total processing time and quality cost, for various image area reduction factors r (5% and 95% of the original). The two datasets, captured on-site, are from the interior of the Asinou Church, Cyprus (S_1), and the exterior of the Archangelos Church, Pedoulas, Cyprus (S_2). Processing was done on a single node comprising 2 Intel X5650@2.66GHz hexa-core CPU and 48GB RAM under Linux. Experiments were run twice. Mean values are presented in table 1. Each row in tables 1(a) and 1(b) corresponds to a run through the 3D reconstruction pipeline for a given image area reduction (r). The first column ($\%r$) shows the *new image area as a percentage of the original*. Thus, the first row ($r = 100\%$) corresponds to the original image size and the last row ($r = 5\%$) to a 95% reduction by area. The 3^{rd} column (T_{cpu}) shows the CPU time in seconds taken for one complete run for a given image size. It uses UNIX's `time` command and takes into account activity on all cores.

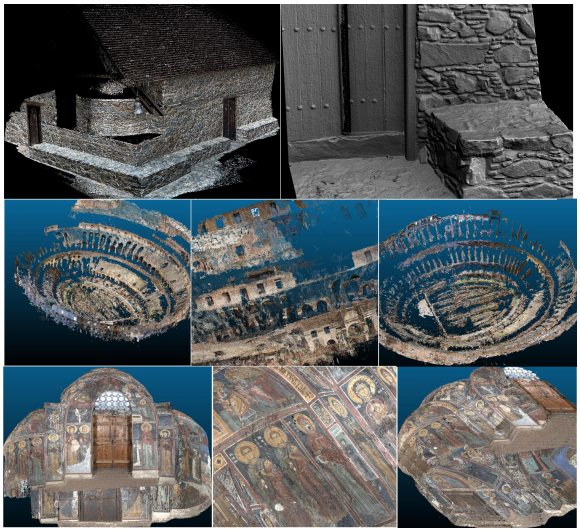

Figure 2: ***Top****: Arch. Michael, Pedoulas, Cyprus.* ***Middle****: Colisseum, Rome.* ***Bottom****: Asinou Church Interior, Cyprus*

The 4^{th} column ($T_{100\%}$) shows run time as a percentage of that of the original size images ($T_{cpu}|r = 100\%$): $T_{100\%} = 100\% \times (\frac{T_{cpu}}{T_{cpu}|r=100\%} - 1)$. A positive $T_{100\%}$ means longer run time than the original by $T_{100\%}$ %, while a negative number means faster run time. The greyed rows show where processing improvement occurs first. The mean reprojection error and the number of cameras "missed" by BA are shown as E_P and E_M. The mean number of features per image after BA is shown as K/C. The total number of 3D points produced by BA is E_{3D}. For both datasets our proposed methodology shows an improvement in run time without significant loss of quality. In particular, for S_1 reducing image area to 25% of the original, results in 36% faster processing and with only 2 cameras missed. For S_2, reducing image area to 10% of the original, results in 67% faster processing and also a reduction in cameras missed by 32%. In general, the relationship between processing time and scaling factor, r, is linear. The use of our pipeline from search to 3D reconstruction is demonstrated with the presentation of the 3D dense point cloud from all the search results for the keyword "*Colisseum*" from Flickr. This is a set of 1,090 images which includes a lot of irrelevant entries because no filtering was applied. 3D models of the Arch. Michael (mesh), Asinou interior and the Colisseum (point clouds) are shown in figure 2.

5. Conclusions and future work

We have presented a simple, robust and fully parallelisable software tool which searches for images on Open Access Image Repositories based on user-defined keywords and other constraints such as geo-location, dates, copyright and camera model characteristics, downloads them and constructs a 3D dense point cloud using our implementation of an SfM pipeline modified in order to accelerate the image matching process with the Low-Res step. Our software can be used as

$r(\%)$	MP	T_{cpu} (s)	$T_{100\%}(\%)$	K/C	E_P	E_{3D}	E_M
100	0.81	46327	0	8814	0.194	553538	0
95	0.77	65643	42	8575	0.193	548133	2
90	0.73	63171	36	8141	0.193	548482	1
85	0.69	59994	30	7707	0.193	548609	1
80	0.65	58112	25	7261	0.193	548337	2
75	0.61	55855	21	6822	0.193	548348	2
70	0.57	52574	13	6394	0.193	548268	2
65	0.53	49745	7	5947	0.193	548369	2
60	0.49	46873	1	5509	0.193	548221	2
55	0.45	44312	-4	5063	0.193	548191	2
50	0.41	41679	-10	4581	0.193	548245	2
45	0.37	39530	-15	4097	0.193	548231	2
40	0.33	36780	-21	3602	0.193	548165	2
35	0.28	34668	-25	3108	0.193	548287	2
30	0.24	32492	-30	2622	0.193	548228	2
25	0.20	29637	-36	2135	0.193	548042	2
20	0.16	27371	-41	1685	0.193	546230	7
15	0.12	25011	-46	1228	0.193	546125	8
10	0.08	22615	-51	801	0.193	544077	11

(a) Archangelos Michael Church, S_1

$r(\%)$	MP	T_{cpu} (s)	$T_{100\%}(\%)$	K/C	E_P	E_{3D}	E_M
100	8.27	105151	0	44015	0.556	611267	76
95	7.86	123851	18	43374	0.540	561586	79
90	7.44	120274	14	42640	0.540	561557	79
85	7.03	114291	9	41751	0.541	560832	79
80	6.62	108878	4	40882	0.540	562141	79
75	6.20	102169	-3	39735	0.541	561054	79
70	5.79	96301	-8	38452	0.540	562534	79
65	5.38	89702	-15	37030	0.541	561580	79
59	4.88	83172	-21	34999	0.541	560574	79
55	4.55	78271	-26	33482	0.542	562517	79
54	4.46	77218	-27	33044	0.541	561302	79
50	4.13	72642	-31	31412	0.541	561740	79
44	3.64	70998	-32	23445	0.551	789872	51
40	3.31	67233	-36	21758	0.552	788621	51
35	2.89	60845	-42	19353	0.552	785942	51
30	2.48	54977	-48	16807	0.548	787998	51
25	2.07	49156	-53	14120	0.554	791019	51
20	1.65	43491	-59	11266	0.550	790518	51
15	1.24	39251	-63	8342	0.549	789746	51
10	0.83	34976	-67	5415	0.551	788231	51
5	0.41	24692	-77	3473	0.540	560024	79

(b) Asinou Interior Church, S_2

Table 1: *Results of running the 3D reconstruction pipeline on (**a**) the Archangelos Michael dataset, S_1 and (**b**) the Asinou dataset, S_2, for various image reductions, %r yielding MP mega-pixels. Completion time (sec) and time relative to original in columns T_{cpu} and $T_{100\%}$. Negative $T_{100\%}$ indicates faster completion compared to original. Greyed row indicates when speed up begins. K/C is mean number of keypoints per camera used in BA E_P is mean reprojection error of cameras in BA, E_{3D} is the number of 3D points estimated in BA and E_M is the number of cameras missed.*

a back-end to a cloud-based service where end users share processing time and storage costs as well as 2D images and metadata for the enrichment of collectively produced 3D models. Tested on two image datasets (not from OAIR), our 3D reconstruction pipeline yielded up to 67% faster processing when using image connectivity based on lower-resolution versions of the originals. At the same time, the increase in the cameras missed during bundle adjustment was very small. A 3D point cloud obtained by searching on OAIR and reconstructing without any manual intervention is also presented. We aim to use Feature Selection techniques, such as clustering and Principal Component Analysis for reducing the size of feature space and compacting SIFT descriptors. We also aim to express image connectivity within a stochastic framework and utilise different ways to estimate it.

Acknowledgements • We wish to thank Noah Snavely for the discussion on SfM quality metrics and reprojection error. • This work was supported by the Cy-Tera Project (NEA ΥΠΟΔΟΜΗ/ΣΤΡΑΤΗ/0308/31) co-funded by the EU Reg. Dev. Fund and the Republic of Cyprus through the Research Promotion Foundation. • The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the EU 7^{th} Framework Programme FP7-PEOPLE 2007-2013 under REA grant agreement IAPP2012 n^o 324523.

References

[AFS*11] AGARWAL S., FURUKAWA Y., SNAVELY N., SIMON I., CURLESS B., SEITZ S. M., SZELISKI R.: Building rome in a day. *Communications of the ACM 54*, 10 (2011), 105–112. 2

[AMN*94] ARYA S., MOUNT D. M., NETANYAHU N. S., SILVERMAN R., WU A. Y.: An optimal algorithm for approximate nearest neighbor searching in fixed dimensions. In *ACM-SIAM Symposium on Discrete Algorithms* (1994), pp. 573–582. 2

[AWWF] ABDEL-WAHAB M., WENZEL K., FRITSCH D.: Automated and accurate orientation of large unordered image datasets for close-range cultural heritage data recording. *Photogrammetrie - Fernerkundung - Geoinformation 2012*, 6, 679–689. 2

[FB81] FISCHLER M. A., BOLLES R. C.: RANSAC: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM 24*, 6 (1981), 381–395. 2

[FFG09] FARENZENA M., FUSIELLO A., GHERARDI R.: Structure-and-motion pipeline on a hierarchical cluster tree. In *IEEE 12th Intl. Conf. on Comp. Vision* (2009), pp. 1489–96. 2

[FFGG*10] FRAHM J.-M., FITE-GEORGEL P., GALLUP D., JOHNSON T., RAGURAM R., WU C., JEN Y.-H., DUNN E., CLIPP B., LAZEBNIK S., POLLEFEYS M.: Building Rome on a Cloudless Dy. In *Computer Vision - ECCV*, vol. 6314 of *Lecture Notes in Computer Science*. Springer, 2010, pp. 368–381. 2

[HZ04] HARTLEY R. I., ZISSERMAN A.: *Multiple View Geometry in Computer Vision*, 2^{nd} ed. Cambr. Univ. Press, 2004. 1

[RWFH12] ROTHERMEL M., WENZEL K., FRITSCH D., HAALA N.: Sure: Photogrammetric surface reconstruction from imagery. In *LC3D Workshop* (december 2012). 2, 3

[SSS06] SNAVELY N., SEITZ S. M., SZELISKI R.: Photo tourism: exploring photo collections in 3D. *ACM Trans. Graph. 25*, 3 (July 2006), 835–846. 2

[SSS08a] SNAVELY N., SEITZ S. M., SZELISKI R.: Modeling the world from internet photo collections. *International Journal of Computer Vision 80*, 2 (2008), 189–210. 2, 3

[SSS08b] SNAVELY N., SEITZ S. M., SZELISKI R.: Skeletal graphs for efficient structure from motion. In *Proc. Computer Vision and Pattern Recognition* (2008). 2

[VF08] VEDALDI A., FULKERSON B.: VLFeat: An open and portable library of computer vision algorithms, 2008. 2

[YP03] YANG R., POLLEFEYS M.: Multi-resolution real-time stereo on commodity graphics hardware. In *Computer Vision and Pattern Recognition* (June 2003), vol. 1, pp. 211–217. 2