Towards structural monitoring and 3D documentation of architectural heritage using UAV

Danila Germanese[1], Giuseppe Riccardo Leone[1], Davide Moroni[1], Maria Antonietta Pascali[1], and Marco Tampucci[1]

Institute of Information Science and Technologies,
National Research Council of Italy, Via Moruzzi, 1 – 56124-Pisa (IT)
`Name.Surname@isti.cnr.it`

Abstract. This paper describes how Unmanned Aerial Vehicles (UAVs) may support the architectural heritage preservation and dissemination. In detail, this work deals with the long-term monitoring of the crack pattern of historic structures, and with the reconstruction of interactive 3D scene in order to provide both the scholar and the general public with a simple and engaging tool to analyze or visit the historic structure.

Keywords: crack quantification methodology, crack monitoring, photogrammetry, UAV, 3D rendering

1 Introduction

Today the cultural heritage is considered a very efficient lever to create and enhance social capital, as it has proven to have a valuable impact on economy and society. In this perspective local and national policies more and more frequently allocate resources to the cultural heritage preservation, restoration, and dissemination to the general public.

The constant growth of digital technologies, also boosted by the advent of low-cost hardware and applications, plays a role in improving the surveying, modeling, and visualization of architectural heritage. Today, the digital representation is not only used to visualize data of an architectural heritage: often complemented with simulation, the digital representation is considered by professionals and academics a fundamental aspect. Indeed, the accurate representation of an ancient structure is a reliable and accessible documentation, also used to assess the mechanical stability or to monitor specific regions, preventing critical events.

The structural deterioration of cultural heritage structures is assessed by monitoring and measuring missing or deformed structural elements, cracks and fissures. Visual inspection remains the most commonly used technique to detect damage and evaluate their progress and severity. Nonetheless, such technique may be time consuming and expensive. Moreover, in some cases, access to critical locations may be difficult.

Apart from criticalities, the accurate monitoring of architectural structures provides, as a by-product, a set a data useful for the creation of informative models of the inspected structure, which are complex models made by geo-referenced architectural models and correlated databases of available information. Also, these models may be used to create an interactive scene of virtual reality for dissemination purpose.

The specific focus of the present work is devoted to: *(i)* monitor the structural health of architectural heritage via UAV; *(ii)* provide a clear and expressive documentation to both scholars and general public. The challenge is to acquire data using drones, hence increasing the safety of the monitoring and reducing time and cost, without loosing accuracy in measurements. Even if there are a number of promising methods in literature, the peculiarity of ancient masonry (generally showing very irregular patterns of shape and colour, or legal restrictions to the installation of sensors) poses an obstacle to the applicability of most of them. In the following, Section 2 provides a review of the literature about the methods used to measure and monitor crack pattern from visible images. Then we devote the remaining sections to the description of the ongoing work carried out in the framework of the MOSCARDO project, financed by the Tuscany region in the framework of the local actions devoted to the preservation and enhancement of the architectural heritage, including efficient dissemination to general public.

2 Related works

The rich literature devoted to the structural monitoring of buildings splits into two different groups: invasive and non-invasive methods. Here we focus on those methods which are more suitable to be applied in cultural heritage: *(i)* Close-range photogrammetry; *(ii)* Marker-based structural monitoring.

Close-range digital photogrammetry is a large family of methods (see [16] for a survey). At each acquisition, the acquired images are used to produce a 3D point cloud. The assessment of the crack opening is made by comparing the point clouds generated at different dates. Such comparison may be performed in many ways, e.g. : *(i)* by *conventional* analysis, i.e. comparing the estimated 3D coordinates of the same points by using statistical tests, [20]; *(ii)* by using shape analysis techniques (matching surfaces [7] or comparing their shape signatures [3]); *(iii)* by comparing a specific shape parameter (the surface area associated to each crack) complemented with a bootstrap testing to detect only statistical meaningful variations in crack opening [1].

Other techniques belonging to this family aims at automatically identifying and measuring structure damages and cracks by using image-based algorithms which allow for specifically filtering out the cracking patterns. In [5], for instance, the authors refer to two image processing methods to automatically and specifically filter out the cracking patterns: the first one evaluates the color level for each pixel, in order to add more "white" or more "black" and thus making the patterns related to structural discontinuities even darker (this method may fail when the walls of the structure are not clear). The second one is based on the detection of the edges by applying a Gaussian Blur to the original image, and subsequently subtracting the filtered image from the original one again. In this work, cracks were detected and inspected by using a rotary wing octocopter micro air vehicle (MAV) and a high resolution digital camera; nonetheless, no quantitative analysis of cracks was performed.

Jahanshahi et al. [10, 9] used a small cross-sectional cracks (0.4-1.4mm) detection method based on 3D reconstruction of the scene, image segmentation and binarization to isolate the pattern related to the structural defect, and finally two classifiers (SVM and NN) trained to distinguish crack from non-crack patterns. The used approach can be applied to images captured from any distance (20m in their experimental tests) and acquired using any resolution and focal length (600mm in this case). Nonetheless, it is suitable for detection of anomalies over homogeneous background (for instance, over concrete). In the work presented by Niemer and colleagues [13], a system based on a commercial camera and a dedicated software (Digital Rissmess-System, DRS) was purposely developed to monitor crack and fissures in civil structures. A cylindrical tube is fixed to the chamber which allows for a constant multi-spectral illumination. Three approaches were developed to extract crack parameters: (i) Fly-Fisher algorithm, which enables to monitor the crack over time and measure its dimensions automatically; (ii) manual measurement of the crack size at a pre-selected point and evaluation of crack profile; (iii) correlative approach, which infers crack parameters by the translation and rotation movements necessary to line up and join the two sides of the fissure.

The works of Jahanshahi [8] and Ellenberg [4] report and highlight the main challenges of cracks automatic detection in civil infrastructures performed with image-based methods. Several image processing techniques, including enhancement, noise removal, registration, edge detection, line detection, morphological functions, colour analysis, texture detection, wavelet transform, segmentation, clustering, and pattern recognition, are described and evaluated in [8]. Among the major challenges, the noise due to the edges of doors, windows, and buildings, that are sharpened when edge detection algorithms are performed.

In addition, in [4] the main problems related to the use of UAVs are reported: the environmental conditions, for instance, and the wind in particular. Moreover, the field of view of the camera, the angle of orientation of the UAV, and the GPS position must be well defined in order to perform reliable acquisitions.

It can be very useful to "mark" the most critical point of a discontinuity [19]. In the work of Nishiyama and colleagues [14], for instance, the so-called "reflective targets" were exploited. Such targets are made by glass droplets, so as to reflect the light as much as possible; they are usually positioned over the crack at points of interest. A number of images are acquired; by means of photogrammetric techniques, the coordinates of the targets can be calculated and the displacements (due to tensile and shear forces) of the two surface portions of the crack can be assessed. The main source of error is due to the calculation of the centroid coordinates. The greater the distance of the camera, the lower the accuracy with which the crack width will be calculated.

In [19], target detection is performed by Hough transform, in order to identify the geometric centers of the targets. Homography techniques are used to correct the perspective error and to identify the planar coordinates of the targets. Any displacement identified by the coordinates of the targets is used to calculate the force field along the discontinuities of interest.

Benning et al. [2] tested different structural elements of pre-stressed, reinforced and textile concrete. For the photogrammetric measurements, the surfaces were prepared by a grid of circular targets. Up to three digital cameras (Kodak DCS Pro 14n) captured images of the surface simultaneously; repeating the measurement in time intervals and calculating the relative distances between adjacent targets made it possible to monitor the cracks and discontinuities evolution. In addition, a Finite-Element-Module was developed, which simulated the test: thus, the results of photogrammetric measurements could be compared with the numeric tension calculation and iteratively improved. The markers can also be home-made; useful suggestions on their dimensions, materials, etc. can be taken from the study of Shortis et al. [17].

3 The solution

Main goal of our work is to design and develop a robotic system able to make the inspection of ancient buildings and structures as automatic as possible, not losing in measurement accuracy. To this aim, unmanned aerial vehicles (UAV) are endowed with cameras and acquire aerial images to allow (by means of advanced image processing techniques) for a contact-less and accurate inspection of cracks and fissures, hence in a cheaper, faster, and safer way.

All the non invasive methods found in literature require high resolution cameras, good environmental and lighting conditions. The most used approaches based on image processing may be divided into two groups: marker-less and marker-based methods. Generally, methods in the first group are based on close-range photogrammetry and require the availability of feature points in the images (e.g. too homogeneous textures generally do not have enough feature points), and high quality images. Even if in literature there are some promising results, e.g. [1], we preferred to implement a marker-based approach, as the peculiarity of the considered case study of the project makes it difficult and complex to evaluate the crack opening at the required level of accuracy. In more detail, our solution was inspired by the ArUco framework, described by Salinas et al. in [12]: the authors define a dictionary of coded planar square markers and tackle the mapping problem as a variant of the sparse bundle adjustment problem, by solving the corresponding graph-pose problem; the optimization is done thanks to the sub-pixel detection of the corners of the markers placed in the scene, by minimizing their re-projection errors in all the observed frames. Such method showed to outperform the well-known Structure from Motion and visual SLAM

techniques in indoor experiments.
Our preliminary experiments have been performed in a controlled setting, in order to verify the accuracy of the solution, and to define the acquisition procedure to be tested outdoor. The experiments and the results are reported in the next section. In the management of the cultural heritage, the 3D recording and documentation is a fundamental task [6] and an accurate 3D rendering is the first step towards the enhancement of cultural heritage. On the one hand the scholar may access a rich set of functionalities to explore all the digital information extracted about the building; on the other hand, the general user will enjoy an interactive survey of the building, possibly including not accessible areas, and possibly enriched with complementary information.

Various technologies can be used to build a 3D digital model [15]. Many ground system are based on LiDAR (Light Detection And Ranging) which is a remote sensing method that uses light in the form of a pulsed laser to measure ranges. Despite the LiDAR result quality, we preferred to explore less expensive solutions. A simple and lightweight monocular camera can be used for Structure from motion (SfM), the photogrammetric technique which estimate the 3D model of the structure from a sequence of different views of the object. The images could be also the still frames of a video. The number, the quality, and the resolution of the images can affect very much the time needed to obtain a full 3D reconstruction. In general, these algorithms are very time consuming, taking hours or days using a normal pc. The using of multiple parallel GPUs or of cloud computing are strongly recommended to speed up the whole process. SfM algorithm, in most of its implementations, consists of three main steps: detection of the control points, building of a dense point-cloud, and finally the surface reconstruction as a polygonal mesh. The very starting point is the feature point selection in all the images, generally obtained by applying SIFT (scale-invariant feature transform) or SURF (speeded-up robust features), very popular and performing feature detectors. Then the corresponding features are matched, hence a registration between images is provided; incorrect matches are usually filtered out with specific algorithms, e.g. RANSAC (random sample consensus). All the matched points are called control points: this set is a sparse point cloud. In the second step, using wide baseline stereo correspondence [18], a dense point cloud is built. The final step is devoted to the definition of a polygonal mesh, the final 3D model of the object. The most popular software of 3D reconstruction and some preliminary results are described in the next section.

4 Experimental setup and preliminary results

The data acquisition system in made up by several optical and electronic devices mounted on the custom ISTI-CNR MAV. This drone is a Micro Air Vehicle designed and assembled at the Institute of Information Science and Technologies of the National Research Council of Italy. This drone has two flight modes: the usual free flight mode, controlled by the pilot, and the programmed flight mode, when the drone flight is based on a predefined set of GPS coordinates of

Fig. 1. The ISTI-CNR drone used for this research (left) and the Ximea XiQ high speed camera used to run SLAM algorithms on the NVIDIA TX1 processor (right)

waypoints. The latter modality is quite interesting, for our purpose, as it allows to repeat the same flight over time; hence it may support the creation of a large dataset of the site of interest over time.

In the original setting on the bottom of the drone there is a stabilized component, named gimbal, hosting a digital camera for video recording, as shown in Fig. 1. The main optical camera used for 3D reconstruction and photogrammetry is the Canon EOS M, a 18 Mega-pixel mirror-less with a sensor APS-C of $22.3x15$ mm (aspect ratio $3:2$). The maximum video resolution is of $1920x1080$ pixel at 30 fps. It weighs 298 g and has dimensions $108x66x32$ mm. The focal length varies in the range 18-55 mm. E.g., setting the focal length at 24mm, and the target at 1.5 mt, the field of view will be of 1.39 m (width) and 0.93 m (height), and the pixel resolution (computed from the camera fact-sheet) will be of 0.27 mm. Beside the main device we added a lightweight camera connected to an embedded video processing unit. The aim of this extra hardware is mainly to experiment some on-line SLAM algorithm (Simultaneous Localization And Mapping) which is useful to display some real-time information on the augmented reality console of the operator. The camera is 4 MegaPixel HighSpeed USB3 Ximea XiQ[1] (Fig. 1). Its weight is only 32g and it can acquire at 90fps in gray scale. The embedded processor is the NVIDIA Jetson TX1[2]. It is an ARM processor couples with parallel GPU and it has an outstanding processing power for such size and consumption. It is already the de-facto standard among similar custom drone projects. It is mounted on a credit card sized carrier board from Auvidea, because the original development kit is too large and too heavy for this purpose. The connection with the ground station uses standard Wi-fi channels.

In order to assess the accuracy of the ArUco marker detection and the repeatability of such measurements, we performed some tests in our laboratory. The camera Canon Eos M has been calibrated using a ChArUco board, with a focal length of 24 mm and image dimension of $5184x3456$ pixel. The re-projection error estimated is of 2.8 pixel. Six markers, with side length of 5.5 cm, were fixed

---

[1] www.ximea.com/en/products/usb3-vision-cameras-xiq-line/mq042cg-cm

[2] https://developer.nvidia.com/embedded/buy/jetson-tx1

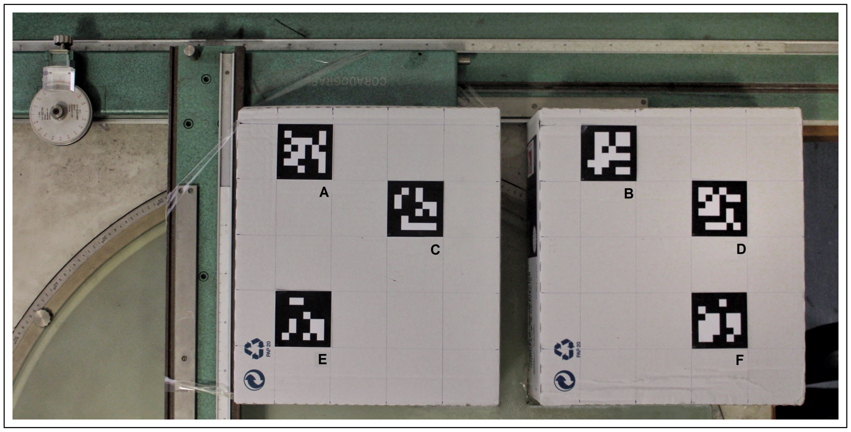

Fig. 2. Simulation of the crack opening:the coordinatographer and the six ArUco markers on it

on two identical boxes (three markers on left, three on right). The left box was fixed to a mobile axis of a coordinatographer, while the right one to the table (Fig. 2). The accuracy of the coordinatographer is of 0.1 mm.

The camera acquires images at about 150 cm from the coordinatographer table. We simulated a crack opening by moving the left box along one direction far from the fixed box. The opening of the crack is performed in 10 steps: 5 by 5 mm, and 5 by 1 mm. The distance between the markers is computed following the same procedure described in [12]: a set of six frames of the same scene is acquired and at each frame the graph-pose is estimated minimizing the re-projection error in the detection of the corners of the planar markers visible. The output of the algorithm are the 3D coordinates of the markers' corners, with ids. The resulting distances between the markers' barycenters are then computed and showed in Table 1. This preliminary test pointed out that there is a need for significant improvement: the error, within 1 mm in most cases, has to be reduced. In order to reduce the error, we plan to both improve the camera calibration to subpixel accuracy and the marker mapping, by using the ChArUco diamond markers instead of the single markers. A diamond marker is a chessboard composed by

Table 1. Simulation of the crack opening: three pairs of markers (A and B, C and D, E and F) moving away from each other, in five steps by 5 mm ($T_1, \ldots, T_5$) and 5 steps by 1 mm ($T_6, \ldots, T_{10}$). All the distance values are expressed in mm.

	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
$d(A, B)$	4.76	5.56	6.19	4.35	4.01	1.05	1.32	0.41	0.86	1.87
$d(C, D)$	4.57	5.38	5.66	4.63	4.33	0.87	0.77	1.02	0.78	1.64
$d(E, F)$	4.45	5.96	5.36	4.59	4.27	1.07	1.03	0.70	0.92	1.46
actual Δ	5	5	5	5	5	1	1	1	1	1

3x3 squares and 4 ArUco markers inside the white squares. The detection of a diamond marker take advantage of the known relative position of the markers in it, and it will improve the robustness and accuracy of the pose estimation in the marker mapping algorithm.

A first acquisition has been carried out flying close to an old tower in Ghezzano, near Pisa. The drone was equipped with the Canon EOS M camera. This flight was useful to verify that the ArUco markers of different dimensions (side length of 5.5 cm, 8 cm, 12 cm, and 20 cm) are detected and correctly recognized. Also, the frames extracted from the acquired video were used to compare the result of the most popular software for the 3D reconstruction. We chose to test the following software:

- **Agisoft Photoscan**[3]: it is maybe the first photogrammetric software, it is proprietary software, exploiting the CUDA parallel GPU technologies. It allows the user tuning the reconstruction parameters during the procedure to increase the quality, depending on the input data, on the user preferences, and on the computing resources.
- **COLMAP**[4]: it is an open-source software, exploiting CUDA technologies. The user can configure the reconstruction settings, but cannot interact with the middle result of the reconstruction phases.
- **Autodesk Recap Photo**[5]: it is proprietary software which exploits cloud technologies in order to perform the reconstruction remotely without burden user calculators. It processes up to 100 photos at once, and it is not possible to configure any setting relative to the reconstruction phases.

Beyond computing dense detailed models [11], we got the best result from Agisoft Photoscan because the optional interaction during the selection of the key points is very useful to process only the interesting part of the image and delete the rest. A brief result of these three stages is depicted in Fig. 3. Then, a virtual scene containing the reconstructed object has been created. The virtual scene has been realized exploiting the Unity[6] engine. The exploitation of such type of engine guarantees the easiness of navigation and, at the same time, the overall representation quality. Inside the scene, users can easily navigate around the reconstructed object and have a quick-look of all the regions of interest of the structure. In the virtual environment, structure cracks are highlighted and labelled with latest measurements. It is also possible to interact with the cracks in order to retrieve past calculated values or visualize charts representing the crack opening evolution over time.

5 Conclusions

We believe that the accuracy of the algorithm mapping of pairs of planar markers, installed along fissures and cracks of ancient buildings and structures, may

---

[3] www.agisoft.com

[4] colmap.github.io

[5] www.autodesk.com/products/recap

[6] www.unity3d.com

a) Original frame

b) Sparse cloud (keypoints)

c) Dense cloud

d) Mesh with texture (final 3D model)

Fig. 3. a) the original aerial view from the drone; b) the first step of the algorithm selects reliable key-points; c) the dense point cloud is computed; d) the final 3D model is a polygonal mesh with texture from the original frame

be improved to enable the monitoring of the crack opening, using drones. Also, the usage of planar markers provides useful 3D information about how the two sides of a crack are moving. The testing phase showed that even if the survey carried out by the drone is detailed enough to have a 3D model usable by the expert, and enjoyable for people, the acquisition procedure, the hardware setting, and the algorithms used to assess the crack opening need to be further optimized to achieve at least the accuracy of a few tenths of a millimeter.

6 Acknowledgement

This work is being carried out in the framework of the Tuscany Regional Project MOSCARDO (FAR-FAS 2014). The Jetson TX1 embedded processor used for this research was donated by the NVIDIA Corporation.

References

1. Armesto, J., Arias, P., Roca, J., Lorenzo, H.: Monitoring and assessing structural damage in historical buildings. The Photogrammetric Record **21**, 269–291 (2006)

2. Benning, W., Görtz, S., Lange, J., Schwermann, R., Chudoba, R.: Development of an algorithm for automatic analysis of deformation of reinforced concrete structures using photogrammetry. VDI Berichte **1757**, 411–418 (2003)
3. Cardone, A., Gupta, S., Karnik, M.: A survey of shape similarity assessment algorithms for product desing and manifacturing applications. Journal of Computing and Information Science in Engineering **3**, 109–118 (2003)
4. Ellenberg, A., Kontsos, A., Bartoli, I., Pradhan, A.: Masonry crack detection application of an unmanned aerial vehicle. In: Proceedings of Computing in Civil and Building Engineering, ASCE2014 (2014)
5. Eschmann, C., Kuo, C., Kuo, C., Boller, C.: Unmanned aircraft systems for remote building inspection and monitoring. In: 6th European workshop on structural health monitoring (2012)
6. Georgopoulos, A., Stathopoulou, E.K.: Data acquisition for 3D geometric recording: state of the art and recent innovations, pp. 1–26. Springer (2017)
7. Gruen, A., Akca, D.: Least square 3d surface and curve matching. ISPRS Journal of Photogrammetry & Remote Sensing **59**, 151–174 (2005)
8. Jahanshahi, M.R., Kelly, J.S., Masri, S.F., Sukhatme, G.S.: A survey and evaluation of promising approaches for automatic image-based defect detection of bridge structures. Structure and Infrastructure Engineering **5**(6), 455–486 (2009)
9. Jahanshahi, M.R., Masri, S.F.: A new methodology for non-contact accurate crack width measurement through photogrammetry for automated structural safety evaluation. Smart materials and structures **22**(3), 035,019 (2013)
10. Jahanshahi, M.R., Masri, S.F., Padgett, C.W., Sukhatme, G.S.: An innovative methodology for detection and quantification of cracks through incorporation of depth perception. Machine vision and applications **24**(2), 227–241 (2011)
11. Majdik, A.L., Tizedes, L., Bartus, M., Sziranyi, T.: Photogrammetric 3d reconstruction of the old slaughterhouse in budapest. In: Computational Intelligence for Multimedia Understanding (IWCIM), 2016 International Workshop on (2016)
12. Munoz-Salinas, R., Marin-Jimenez M.J. andYeguas-Bolivar, E., Medina-Carnicer, R.: Mapping and localization from planar markers. P. Recog. **73**, 158–171 (2018)
13. Niemeier, W., Riedel, B., Fraser, C., Neuss, H., Stratmann, R., Ziem, E.: New digital crack monitoring system or measuring ad documentation of width of cracks in concrete structures. In: Measuring the Changes: 13th FIG Symposium on Deformation Measurements and Analysis (2008)
14. Nishiyama, S., Minakata, N., Kikuchi, T., Yano, T.: Improved digital photogrammetry technique for crack monitoring. Adv. Eng. Informatics **29**(4), 851–858 (2015)
15. Remondino, F., Campana, S.: 3D Recording and Modelling in Archaeology and Cultural Heritage - Theory and Best Practices. Archaeopress BAR (2014)
16. Remondino, F., El-Hakim, S.: Image-based 3d modelling: a review. The Photogrammetric Record **21**, 269–291 (2006)
17. Shortis, M.R., Seager, J.W.: A practical target recognition system for close range photogrammetry. The Photogrammetric Record **29**(147), 337–355 (2014)
18. Strecha, C., Fransens, R., Gool, L.V.: Wide-baseline stereo from multiple views: A probabilistic account. In: In CVPR, pp. 552–559 (2004)
19. Valença, J., Dias-da Costa, D., Júlio, E., Araújo, H., Costa, H.: Automatic crack monitoring using photogrammetry and image processing. Measurement **46**(1), 433–441 (2013)
20. Welsch, W., Heunecke, O.: Models and terminology for the analysis of geodetic monitoring observations. In: FIG 10th International Symposium on Deformation Measurements. International Federation of Surveyors, vol. 25, p. 22 (2001)