SCIRES it

SCIentific RESearch and Information Technology
Ricerca Scientifica e Tecnologie dell'Informazione
Vol 11, Issue 2 (2021), 125-138
e-ISSN 2239-4303, DOI 10.2423/i22394303v11n2p125

CASPUR-CIBER Publishing, http://www.sciresit.it

COMPUTER VISION ALGORITHMS PERFORMANCE IN ARCHITECTURAL HERITAGE MULTI-IMAGE BASED PROJECTS. GENERAL OVERVIEW AND OPERATIVE EVALUATION: THE NORTH TOWER OF BUÑOL´S CASTLE (SPAIN)

Jose Luis Cabanes, Carlos Bonafé**

*Universidad Politécnica de Valencia – Valencia, España.

Abstract

Multi-image based modeling has proven to be effective providing solutions for surveying and documenting cultural heritage, and in particular architectural heritage. In addition to the issues related with instruments and captation strategy, the operativity of these projects is supported by three bases: Computer Vision (C.V.) algorithms, analytical close-range photogrammetry, and theory of errors. In this work we propose an approach that examines the importance of the first, from two points of view. On one hand, we present a brief overview of its intervention in the different processing stages, both in photomodeling as in photograms stitching projects, thus reviewing the fundaments regarding the two classic branches of architectural photogrammetry. On the other, we present a review of the operational strategy with these algorithms, through a case study that evaluates the results of two software applications, advancing some methodological improvements.

Keywords

Multi-Image based projects, Computer Vision algorithms, Close Range Photogrammetry, 3D Modeling, Photograms stitching

1. *Introduction*

Multi-image based techniques have meant a considerable improvement regarding to surveying and documenting architectural heritage. Especially in overlapping sets of photos, because quite automated processes have been developed in the recent years, which include C. V. algorithms, close range photogrammetry, and optimization of results by adjustment of errors.

With these three supports, a good number of software applications allow to solve the formation of 3D models of buildings and objects, as well as the automatic stitching of photos for the implementation of panoramas, "synths" and virtual visits.

On the other hand, we must add to that other "parameters that affect the quality of photogrammetric results including used software, number of photos, and control points" (Elkhrachy, 2020), and particularly, the efficiency in collecting information that is achieved today with the multi-sensor techniques, that include image and range based captures. So that, thanks to the combination of hybrid captures and CV-based software, has been possible to generalize the use of multi-image based techniques, and achieve a reliable range of results for the analysis and documentation of cultural heritage, and in particular architectural heritage. Among the three basic supports, it is surely the operativity with the first of them, the one that results more relevant, since it largely conditions the effectiveness of the others. As Debevec, Taylor and Malik (1996) affirm: "these systems are only as strong as the underlying stereo algorithms".

In the present work we propose, on the one hand, to review the intervention of this automatic technology in the processing stages, and on the other, to analyze the results in a specific Case Study, in terms of reliability, since its operational strategy is one of the keys to achieving optimal results

2. *Intervention levels of these operators*

Since its inception in the 1970s, a wide variety of applications use Computer Vision techniques in automatic Stereo/Multi-vision projects through passive and other types of cameras, to extract 3D information from the environment, even with real-time data processing. (Szeliski, 2011).

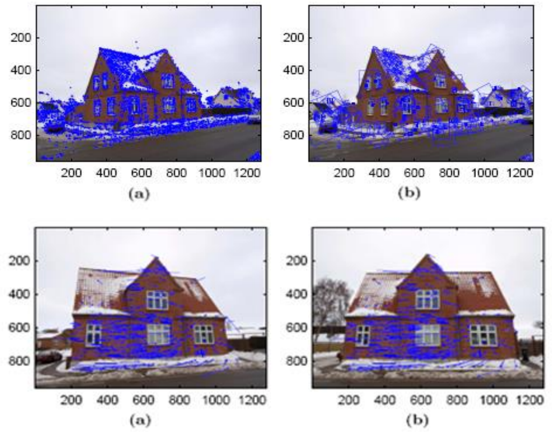

Fig. 1: Example of *keypoints* detection with SIFT algorithm. Top: "The 4966 keypoints detected by SIFT in a 1280 × 960 pixel image. a) Location of all detected keypoints. b) Differently sized and oriented matrices associated to the area-based search (only 10% are shown to avoid clutter)". Bottom: "Keypoint matches between the two images shown in (a) and (b), using the threshold τ = 0,6 in the correlation. The dots represent the location of keypoints (only 25% of the 1527 matches are shown)" (Lindequist, 2010)

In the first decade of the 21st century, some semiautomatic close-range photogrammetic applications were launched, that handled the reconstruction of a model using coded or backlit targets to identify orientation points. However, the "substantially increasing level of automation in the photogrammetric process (has been) due to the considerable algorithms improvement from the Computer Vision community", that brought a radical change by allowing to work with extense sets of neighbouring photos. (Skarlatos & Kiparissi, 2012). These operators intervene at three levels: feature detection in images, correlation between the photopair points, and filtering processes, as we will explain below. Among the former, Scale Invariant Feature Transform (SIFT), which was the first to be published in 2004, and Speeded Up Robust Features (SURF), are noteworthy for feature detection and matching. Both algorithms operate in two stages: in the first, through the examination of the grey-scale channel of the entire image, looking for highly contrasted pixels with respect to the neighbours, through an area-based search, and according to a Gaussian-type function. The "key points" thus obtained are defined by their pixel coordinates, and by a "descriptor", a 64 or 128 dimension vector, that represents a gradient histogram of the features of the neighbouring pixels. In this way, the results are quite robust to changes in the nearby photos of the sequence.

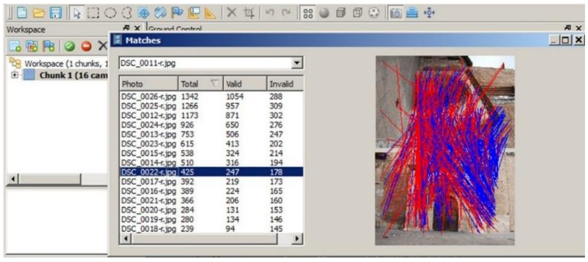

Fig. 2.a: Two-dimensional matching algorithms overview: a photo-modeling software application. Blue lines show acceptable matches between photos 0011 and 0022 (up and down in the thumbnails on the right). In general, the highest proportion occurs among pairs of neighbouring photos (Agisoft LLC, 2015)

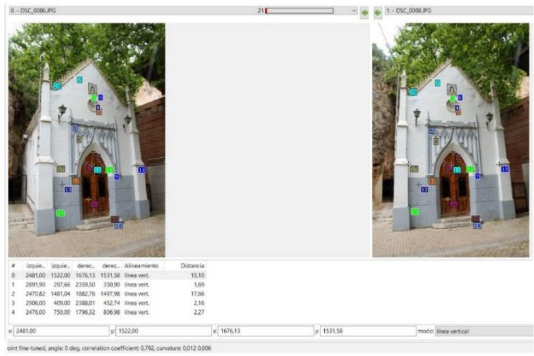

Fig. 2.b: Control points: the same concept in panorama-stitching applications. After a process called "Fine-Tune", to accurately "estimate the corresponding (position of a) point up to one tenth of a pixel", for each pair detected the correlation coefficient is shown (in the lower part of the screeen). "Typically values over 0.8 indicate that the image areas around each point of the pair are very similar (an 80% correlation)" (Hugin, 2012)

The next stage is to resolve the correspondence between the keypoints detected in pairs of overlapped photos. For this purpose, SIFT uses the "nearest neighbour method", i.e., the point in the second image located at the minimum distance in pixel coordinates, improving the search by an additional comparison with the distance of the second nearest neighbouring point. That is, if the point in the initial image is d_i, and the closest in the target image are d'_j and d'_k , the i-j match is accepted if the threshold

$$\tau > \frac{\|d_i - d'_j\|}{\|d_i - d'_k\|}$$

is smaller than a certain value, which ensures discarding a high percentage of false matches (see Figs. 1, 2.a and 2.b).

Software applications allow different adjustments in the parameters of these technologies. The intensity of the "Smart Feature

Detection" (SFD) controls the quantity and reliability of the resulting "keypoints", by means of the level of subsampling, and by the characteristics of the search matrices. The reliability of the two-dimensional matching process is usually controlled by two settings. On one hand, by the preselection of pairs, to avoid analyzing those with an insufficient overlap. And likewise, by means of other area-based algorithms, such as the "Correlation coefficient", the "Quadratic matching", or the "Least Squares Matching" (LSM), among others, that use descriptor comparison, with the aim of discard false correspondences. (Re et al., 2014). Similar matching algorithms allow filtering the orientation process, and solving the linear correlation in the metric reconstruction phase, as we will see below.

The matching results must be filtered according to "strategies for outlier rejection", based on robust adjustment algorithms, such as RANSAC (Random Sample Consensus) or MAPSAC (Previtali et al., 2011). This way, the inevitable errors caused by excessive disparity between shots, shadows, glare, or other causes, are filtered by checking the alignment between the key points in the photopair, by means of linear regression techniques, and discarding those that exceed a certain threshold that, therefore, reveal false matches (filtering-1 stage in Fig. 3).

3. *General overview of the C.V. algorithms performance in SfM projects*

3.1. Projective Reconstruction phase

In a multi-image based project, aimed to create a 3D model or a panoramic photography, firstly the projective-based reconstruction of the sequence must be resolved, accurately determining the location and rotation of the cameras (the "Structure from Motion" phase, aka SfM), followed by a metric reconstruction phase, in order to obtain a dense point cloud of the model, or else to stitch the images together to form a panorama or a "synth", according to the type of project involved. Computer Vision based techniques are so versatile, that they result decisive in both types of projects, and also in both phases, leading to a general workflow like the one shown in the following diagram, which we will briefly discuss below. (see Fig. 3).

The projective reconstruction phase begins with an initial reconstruction, known as "relative orientation", in which "the two first images of the sequence are used to determine a reference frame" (Pollefeys et al., 2000). The spatial position of the two first cameras can be calculated by three main approaches, used according to the project characteristics, developed from the photogrammetric theory of photo pairs, which due to their importance we will briefly comment.

One of these principles is to consider the "coplanarity condition" between the two vectors associated with the "nuclear plane" of a point M, in the Terrero-Hauck configuration, and the translation vector between the two cameras O-O´, so that their mixed product results null. A classic focusing of this formulation is based on the development of the corresponding null determinant, involving the internal parameters of the images through the "collinearity condition" or Direct Linear Transformation (DLT). As a consequence, this "image condition" between a 3D object point and its two image vectors, can be written in projective coordinates:

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \frac{-1}{f}\left[K \cdot [R \cdot T] \cdot \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \right] = \frac{-1}{f} \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{pmatrix} \cdot \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

with (K) the "calibration matrix" of the cameras, and (R T) the product of matrices that represents the extrinsic parameters of rotation and translation between the exterior system and the image system, also known as the "projection matrix". (Förstner, 2004)

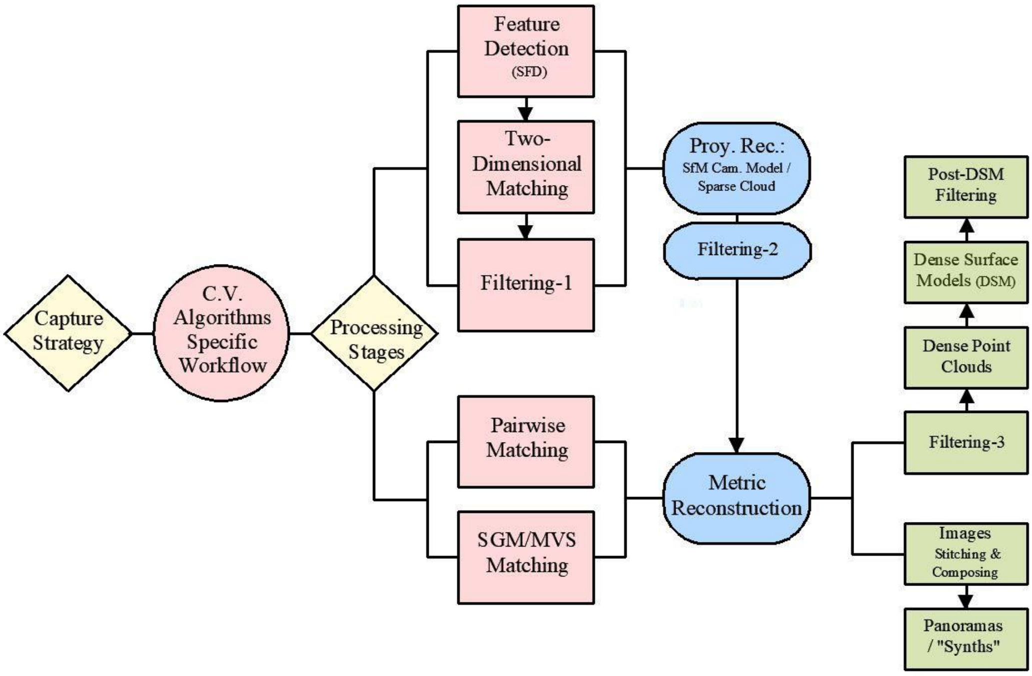

Fig. 3: General chart of the C.V. algorithms involved in the processing stages of a multi-view based project

Another useful development of the coplanarity condition starts from expressing this mixed product according to the coordinate system associated with the first camera O:

$$(x'_O \quad y'_O \quad z'_O)\cdot\left[(T_O)\wedge\begin{pmatrix}x\\y\\z\end{pmatrix}\right]=0$$

with To the translation vector between images O and O´.

It is verified that this relationship leads to a simple expression that sums up all the transformations involved:

$$(x' \quad y' \quad z')\cdot(E)\cdot\begin{pmatrix}x\\y\\z\end{pmatrix}=0$$

with (E) called the "essential matrix", thus associated with the image coordinates, and with the chance to involve the calibration parameters of the camera.

It is also possible to solve a photo pair from the Terrero-Hauck nuclear beam of planes, with a strictly projective approach. If we consider the two epipolar lines associated with a point M, lying in a nuclear plane of this beam. The expression of their projectivity will be:

$$r'_{ME}=\rho\cdot(A)\cdot r_{ME}$$

From this, transforming the tangential coordinates of the epipolar line r'_{ME}, this relation can be formulated:

$$r'_{ME}=(x' \quad y' \quad t')\cdot\begin{pmatrix}u'_{ME}\\v'_{ME}\\w'_{ME}\end{pmatrix}=\rho\cdot(x' \quad y' \quad t')\cdot(F)\cdot\begin{pmatrix}x\\y\\t\end{pmatrix}=0$$

that corresponds to a bilinear expression of the homography between epipolar series of corresponding points in each image, therefore in pixel coordinates. The matrix (F) is known as the "fundamental matrix" of the epipolarity, and "describes the projective relation between two uncalibrated views". (Rodehorst et al., 2008). It yelds an obvious practical utility, since its calculation can be carried out from image coordinates in both images.

As we can see, these three main derivations to analytically solve a pair of photos, are based on the control points previously obtained with the C.V. algorithms for SFD and two-dimensional matching. With the "relative pose" of the first pair of photos resolved, the workflow continues, in short, adding cameras until the "multi-photo orientation of isolated beams" stage is complete.

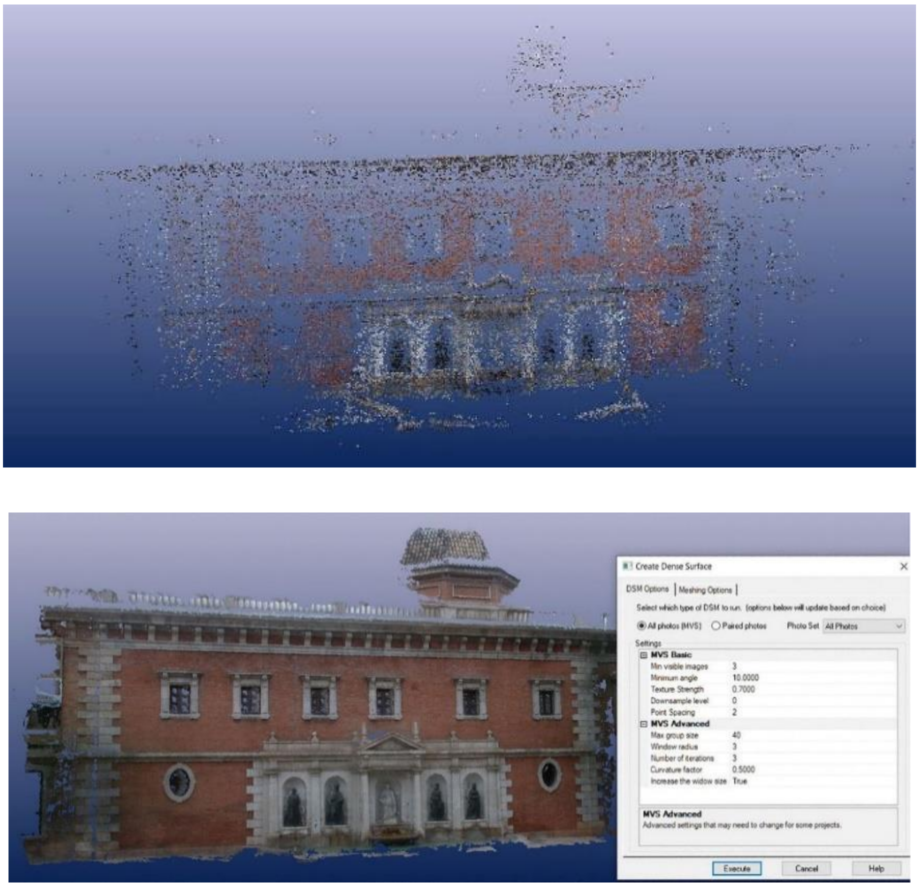

Fig. 4: Facade of the Old University of Valencia in the Patriarch Square, captured with cameras on tripod. Top: "Sparse Cloud", obtained with SFD and area-based matching algorithms. Bottom: Dense point cloud obtained with a MVS algorithm. (PhotoModeler Technologies, 2017)

The three previous orientation foundations can also intervene in this complex process, according to different methodologies. Successive results are filtered by different methods, such as the "epipolar condition", which allows the use of linear correlation operators (now usually under a "trifocal tensor" configuration), robust correction technologies (such as RANSAC), removing points with greatest residual before recalculating, or other like the Singular Value Decomposition (SVD), which in the cases of E and F matrices, helps to "clean up" its two-rank geometric condition (filtering-2 stage in Fig. 3). (Barazzetti et al., 2011)

Complementarily a least squares optimization calculation can be executed, intended to minimize the errors of the involved variables, usually operating with the residuals of the image coordinates obtained by alternating spatial intersections and resections (retroprojection errors). The least squares fitting of all the beams in block calculation, known as "bundle adjustment", yelds the optimization of the photograms positions, and leave the SfM model ready for exploitation of results. This overall redistribution of errors is important because "the estimated parameters are prone to inaccuracies caused by wrong correspondences, critical camera configurations (e.g. small baselines), measurement noise, or calibration errors" (Hänsch et al., 2016). Apart from the image coordinates, the internal orientation parameters (optical and radiometric), are normally also involved, implementing a "project calibration" for better results, in a process then known as "hybrid block adjustment". (Dorffner & Forkert, 1998).

The projective reconstruction phase provides thus a model of oriented / calibrated cameras, with proven quality, as well as a set of 3D points calculated at the same time, usually called "sparse point cloud" or "low density point cloud".

3.2. Metric Reconstruction phase

On the other hand, C.V. based matching algorithms are no less important in the metric reconstruction phase, in both types of projects similarly. In photo-modeling ones, they first give rise to the “deep recovery” or “depth map” of the sequence, “...generated by estimating 3D coordinates for additional matches between images”. To this end, two types of calculations are used, depending on the characteristics of the project. The first one was the “Pairwise Matching” method, registering and merging the partial results of a linear matching, by means of the above mentioned area-based algorithms, but now with less uncertainty, due to the restriction of the “one dimensional correlation”, derived from the epipolar condition in a photopair. (Kraus, 1993).

The depth map can also be obtained by Semi-global Matching / Multi-View Stereo technologies (SGM / MVS), which have proven their satisfactory operativity specially in UAV-based projects, in which “the processing pipeline can be hindered due to the limited quality of the data”. (Haala & Rothermel, 2012). The basic idea behind is “to use more than two images within the matching process” (photo subsets), and “combine the pairwise results afterwards to create the final solution”. (Bethmann & Luhmann, 2015). (see Fig. 4).

A “dense” or “stereo” cloud” is then achieved, characterized by a heterogeneous spatial distribution of the points, and a variable quantity of clusters of outliers. To improve its quality, a pre-DSM filtering of the cloud is required, with techniques such as elimination of redundant points and outliers, among others (filtering-3 stage in Fig. 3). (Altunas et al., 2019).

If the aim is to document or visualise the model, the 3D point cloud is normally triangulated to achieve a phototextured network, called Dense Surface Model (DSM). For this purpose, firstly algorithms such as Delaunay or Poisson perform a spatial triangulation, and then, others solve the texture projection on the mesh, normally by means of its previous parameterization in portions, to be re-projected from the oriented images, and assigning them a color according to criteria such as the relative frontality with respect to the source images (Visual Computing Laboratory, 2016). The triangulation must also be cleaned with post-DSM filtering to eliminate errors such as crossed or non-manifold faces, among others.

Fig.5: Facade of the Old University of Valencia in the Patriarch Square. Up: initial images with divergent directions. Middle: rigid-solid transformation over planar mapping. The photos result aligned to ensure that the overlapped content coincides with no deformation (no stretching or distortion). Below: general homography transformation over planar mapping. The images are linearily deformed so that the two laterals result aligned with the central, to recover the base-plane in the model

Instead, in a photogram-transformation project, once solved the stages: i) spatial location and filtering of the control points, through C.V. algorithms; ii) orientation of the images, by means of any of the three commented approaches; and iii), filtering and global bundle or hybrid block adjustment, the projective transformation that best adapts to camera movement, can be automatically deduced from the previous SfM recovery. (Microsoft Research, 2015).

In a project for the formation of a panoramic image, the shooting sequence may correspond to a planar or a rotation motion. In the first case, in theory, the most restrictive rigid-solid homography transformation, usually works well to stitch the images on a homothetic plane with respect to the model basic plane.

In the second, the shots correspond to the same theoretical gnomic projection, so that a rigid-solid homography allows all the depth planes of the model to be correctly linked with no “parallax” error.[1] (Gao et al., 2011). However, the difficulty in achieving parallelism between all the shots in a planar motion, and the slightly divergent location

[1] This is due to the projective concept of conical projection as a complete spherical radiation.

of the optical center of the camera lens, with respect to the device´s center of rotation, leads to a general projective transformation being the most suitable for achieving a stitching surface that matches well with the shooting sequence. (Xiang et al, 2016)

Fig. 6: Basic elements of a "Synth" visualisation. Up: low-quality point cloud obtained with C.V. algorithms. Down: overlay of isolated photos (in this case from a Spin motion). On the bottom of both images: navigation bar with the usual controls. (Microsoft Research, 2014)

In any case, the result of these projective transformations, automatically determined by the specific software applications from C.V. algorithms, allows identifying the basic surface on which to map the images, involving therefore a "metric reconstruction" of the scene.

After images stitching, two composing formats can be adopted: (i) a smooth transition between isolated photos, optimizing the number of pixels required at each moment, which results in a fairly continuous trip with a surprisingly realistic effect (a "Synth"); or (ii) a seamless composition of all the images on a plane, cylinder or sphere.

The projection modality (previous to its final planar view), is also involved in the metric reconstruction of the rotation motion, as "there is no single, unique projection for representing sections of the sphere on the globe. Instead, all projections have various attributes and limitations" (Hugin, 2012). (Figs. 5 and 6)

4. *Case Study*

4.1. Location, equipment and methodology

As a practical application, we are going to analyze a photo-modeling project of the Buñol´s Castle Tower (Spain) located at the north entrance, and accessible from the outside through a stone bridge.

For data capture, the following material was used:

- Ground equipment: Nikon 7100d semi-professional camera with tripod and self-timer, avoiding blurred photos.
- Aerial equipment: Drone F550 with modifications made by ourselves, such as the addition of an autopilot system, and an automatic shooting system (with custom 3D printed support). A Xiamomi Yi camera (type Go Pro) was used.

For the processing of the project, a terrestrial photogrammetry + UAV customized methodology was used. On each side of the Tower (exterior and interior), a base model has been provided by the cameras on tripod. The aerial sub-model, with lower quality shots, was then registered also on each side.

The control of the reliability of each base model was performed with a methodology like the one we expose here (see section 4.2), and finally the two faces of the Tower, were registered.

4.2. Analysis of the operativity with C.V. algorithms

For our analysis of the operativity with C.V. algorithms, we have processed only the photos corresponding to the interior façade of the Tower, taken with the Nikon D-7100, for being of higher quality than those obtained from the drone.

These are 80 images with a dimension of 3872 x 2592 pixel, taken with minimum focal (18 mm), minimum aperture (f / 22), and self-timer, thus ensuring a good radiometric quality, level of detail, and uniform dynamic range.

We have processed them with PhotoModeler v.1.2.0. 2127 (aka PM), and PhotoScan v. 2017.1.1.2199 (aka PS) software products. (see figs. 7 and 8)

Fig.7: North Tower of Buñol´s Castle. Top: sequence of terrestrial and aerial shots describing respectively arcs and spirals. We have checked the effectiveness of these sequences, with nadiral and azimuthal angles of different signs, so that the orientations of the model are registered from all possible positions with enough redundancy, in order to optimize the SFD and Matching algorithms. (Cabanes & Bonafé, 2019). Bottom: interior views of the filtered 3D point cloud of the full model

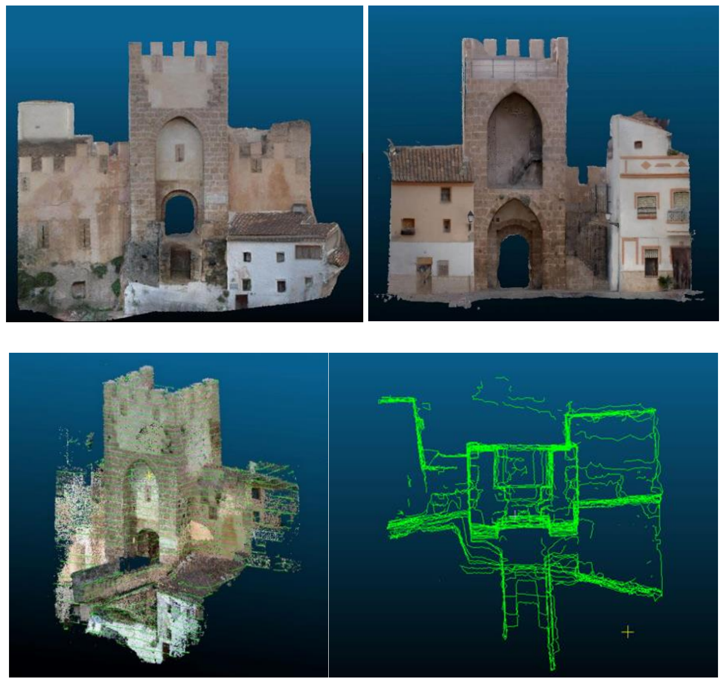

Fig. 8: North Tower of Buñol´s Castle. Yop: orthophotos exterior / interior from the complete DSM. Bottom: automatic level curves as polylines, obtained from the dense point cloud (edited with Cloud Compare, 2019)

In accordance with the stages exposed in the general overview of multi-image based projects, we present four tables below. In Tables 1 and 2 we have evaluated the operativity in the projective reconstruction phase, with both software applications, establishing, in general, two levels of intensity.

Tab. 1: SFD levels analysis.

	Smart Feature Detection Key points detected /used per photo, avg.	
SFD Downsampling level	Medium (85%)	High (75 % PM / 55% PS)
PhotoModeler (PM)	25.000 / 14.500	45.000 / 25.000
PhotoScan (PS)	20.500 / -	35.000 / -

In Table 1 we have tested two types of downsamplig for Smart Feature Detection (SFD), which is the only setting available in both applications. To work with comparable values we have established the categories:

- SFD Medium, which corresponds to approximately 85% of area reduction ("Very High" level in PM, and "Low" level in PS).
- SFD High, less homogeneous, since it corresponds to the "Extra High" level in PM (75%) and "High" level in PS (55%).

In Table 2 we test the Matching algorithms, and we treat two categories:

- Match Medium, with GOC = 10 and MQT = 0.5, which are the PM default values. And pair preselection "Generic" in PS.
- Match High, with GOC = 15 and MQT = 0.4, with a higher level of demand only in PM, since PS does not allow resetting the default option.

Good Overlap Count (GOC) values the overlap between two photos to apply the matching algorithm, while Match Quality Threshold (MQT) refers to the number of accepted tracks.

In Table 3 we show a summary of the main parameters of the SfM models obtained, combining options from the previous tables. We include the most representative values of the sparse reconstruction: number of 3D points, maximum retroprojection residual (υ_{max}) and its disparity, as well as the redundancy in the depth recovery, and the proportion of weak points. The first orientation obtained in each case, including a cameras field calibration, ("Orientation" rows in the Table), has been improved through a sequential process of removing the points with greatest residuals, and then running a bundle adjustment of all variables, until achieving a new residual in the worst point that does not exceed the initial one. Thus, the filtered cameras model ("R + B.A." rows in the Table) obeys to a more weighted distribution of the retroprojection error.

From all this we deduce that there are three characteristics to highlight, that allow us to propose a methodological improvement of project adjustments, relative to these technologies. Firstly, Table 1 shows an increase in the number of key points detected with a lower level of downsampling. Despite this, if we compare the two levels of matching in Table 2 with PM, we see that a higher quality requirement causes a significant decrease in the tracks used, in the order of 50%, which can lead to a problematic depth map, if it turns out below a certain threshold.

Secondly, after considering the filtered results of the SFM reconstruction, we see that it is necessary to previously asses whether it is appropiate to readjust the SFD + Matching algorithms, or even reconsider the selection ofphotos. In our case, we have found a problematic SFD / Matching Medium configuration in PM, which, despite having an acceptable "Matching

Tab. 2: Matching levels analysis.

	Total photopairs		**Matches detected / used per photo (avg) / max**	
SFD Match *Match*	Medium Medium *High*	High Medium *High*	Medium Medium *High*	High Medium *High*
PhotoModeler (PM)	1.319 *1.319*	2.290 *2.290*	850 / 800 / 1.000 *350 / 300 / 500*	1.100 / 950 / 1.500 *475 / 475 /1.500*
PhotoScan (PS)	--	--	4.885 / 280 / 4.000 -- / -- / --	41.130 / 3.600 / 4.000 -- / -- / --

Tab. 3: Main Orientation Parameters

	Quantity Nº of 3D points		**Residuals** υ_{max} / RMS υ_i / Max RMS υ_i (pixels)		**Redundancy** Rays per 3D point (avg) / 2D Points per photo / 2D Points in two photos only	
SFD *Match* Orientation ***Filtering-2*** Match Orientation ***Filtering-2***	Medium *Medium* First ***R +B.A.*** *High* First ***R+B.A.***	High *High* First ***R +B.A.*** *High* First ***R+B.A.***	Medium *Medium* First ***R +B.A.*** *High* First ***R+B.A***	High *High* First ***R +B.A.*** *High* First ***R+B.A.***	Medium *Medium* First ***R +B.A.*** *High* First ***R+B.A***	High *High* First ***R +B.A.*** *High* First ***R+B.A.***
PhotoModeler (PM)	37.678 36.139 58.006 40.158(*)	61.185 ***59.893*** 64.496 *63.452*	2,15 / 0,65 / 2,29 ***2,05/ 0,79 / 2,06*** 2,10 / 0,75 / 1,98 ***1,99 / 0,51 / 1,93***	2,10 / 0,75 / 2,09 ***2,00 / 0,57 / 1,87*** 2,10 / 0,74 / 2,07 ***1,96 / 0,56 / 1,87***	5 / 2.495 / 39 ***5 / 2.212 / 40*** 5 / 3.718 / 39 ***3 / 1.643 / 54***	5 / 3.583 / 42 ***5 / 3.406 / 42*** 5 / 4.162 / 40 ***5 / 3.967 / 40***
PhotoScan (PS)	11.208 ***10.527***	22.555 ***18.187***	6,54 / 2,48 / - ***5,58 / 2,27 / -***	6,60 / 1,22 / - ***5.44 / 1.06 / -***	5 / 696 / -- 4 / 595 / --	9 / 2.650
Notes	(*): Only 76 / 80 cameras oriented					

used (avg.)" of 800/1000 in Table 2, only orients 76/80 cameras. Therefore, the combination of medium level of subsampling + not very demanding matching, fails to locate enough points of orientation in the depth map, due to an excessive disparity, and causes 4 cameras orientation failure, and, in addition, a minimum redundancy level of 3 "Rays per 3D point (avg.) [2] " in the others (Table 3).

Third, we see that a project workflow oriented to a robust SfM configuration, to ensure the metric quality of the 3D model, must essentially meet: (i) optimal values for υ_{max} and for its dispersion, as well as an acceptable redundancy in the number of "Rays per 3D point (avg)"[3]; and (ii) the value of υ_{max} must correspond, at any case, to the average photoscale, and to an allowable tolerance in the 3D model. In general, it is observed that the results shown in Table 3 present lower errors, with PM and a complete workflow (SFD / Matching / Orientation + Filtering), although a more aggressive filtering methodology of the Sparse Cloud in PS can lead to an average equivalent results. In all cases, the configuration SFD Medium / Matching Medium / Orientation First, produces results that are clearly worse compared to the others

As for the algorithms involved in the metric reconstruction, both applications share the SGM / MVS technology, and we have also defined two categories:

- SGM / MVS Medium, adjusting: (i) in PM: Downsampling level 1 (a level 0 implies full image resolution), and Point Spacing 2 (standard sampling rate), which are the default options; (ii) in PS: Quality "Low" (subsampling equivalent to that selected for the SFD), and Depth Filtering "Moderate" (level 3/4) regarding the sampling range.

- SGM / MVS High, adjusting: (i) in PM: Downsampling 0, and Point Spacing 1 (maximum density); (ii) in PS: Quality "High" (subsampling equivalent to that selected for the SFD), and Depth Filtering "Aggressive" (level 1/4), also maximum density. This yelds an adequate level for models that require a detailed reconstruction, as recommended by both Manuals.

The results of the tests that we collect in Table 4, in general, confirm the greater performance achieved with low levels of subsampling in the images, high ranges of "point spacing" and filtering (*SFD High / Matching Medium / Orientation + Filtering*). But they must be supported on a robust

[2] The SFD default option in PM produces an even bigger subsampling, so the expected result would be even more unfavourable.

[3] In general υ_{max} <2 and its dispersion <0.5 (pixel) are supported, as well as a minimum of 4 "Rays per 3D Point".

Tab. 4: SGM / MVS results overview.

	Dense Cloud: **Nº of 3D points**			
SFD Match ***Orient. + Filter.-2*** **SGM**	Medium Medium ***R +B.A.*** Basic	High Medium ***R +B.A.*** High	Medium Medium ***R +B.A.*** Basic	High Medium ***R +B.A.*** High
PhotoModeler (PM)	4.131.365	15.419.773	6.129.992	18.002.975
PhotoScan (PS)	972.787	13.603.801	15.626.945	14.739.299

previous configuration of the cameras model, as we have previously commented. (fig. 9)

In our surveying project, we have combined exigent SFD + Matching configurations, along with an adequate level of filtering of the Sparse Cloud on each side of the Tower, so that when registering the aerial cameras, and filtering / optimizing the joint model, the errors found have been acceptable.

Conclusions

So far, we have briefly discussed an approach of the C. V. algorithms intervention and operativity in Multi-Image based projects for architectural documentation that allows some conclusions to be drawn in both aspects.

As regards the first, we have seen how these operators, plus the geometric base provided by projective relations, as well as the optimizing of errors techniques, constitute the three supports, that we have considered with a wide focusing, recovering the unity of the two classic branches of photogrammetry, surveys and photograms transformation. We have shown how they result decisive in the two stages, projective-based and metric-based, of these projects. The first one involves SFD, two-dimensional matching, and filtering (1 and 2) algorithms, that jointly provide reliable data to calculate the SfM recovery. While its performance in the second stage is based on one-dimensional correlation techniques between photopairs, especially through the latest SGM / MVS, along with the pre-DSM depth map filtering operators. In both stages, the combination of these filtering C.V. technologies, together with the optimization of errors, or "subpixel refinements", allow to improve the different partial calculations.

The experimental basis of these operators is constantly evolving, and allows the resolution of increasingly ambitious projects, especially like those based on hybrid-capture data, due to “their high repeatability for close-range applications” (Barazzetti et al., 2010). In summary, in relation with this first aspect, our analysis results in a brief review of these image-based projects, from the point of view of C.V. technologies.

The second question, regarding its operativity, shows us how these algorithms concentrate the processes in which the operator's intervention options are relevant. From the analysis of our Case Study, we can draw useful conclusions, regarding three aspects. Firstly, we have seen how the relationship between the subsampling levels of SFD and matching quality, can become problematic for a strong orientation, either due to an insufficient number of trackings, or an excessive disparity in the depth map.

Regarding the orientation stage, we have identified some basic markers to ensure a strong SfM reconstruction, such as the SFD+Matching algorithms configuration, and the filtering intensity of the Sparse Cloud before the re-calculation and bundle adjustment processes. These markers result particularly important to ensure the reliability of the “new methodologies to collect large amounts of data from various sources, that must be accurately registered and integrated” (Remondino et al., 2009), as we have verified in our Case Study.

And thirdly, in relation to SGM/MVS technology, we have verified that its highest performance must be supported by a robust cameras model that provides metric reliability and results adjusted to the average photoscale of the capture, on the one hand, and to the conditions of the commission on the other.

In summary, our research underlines the importance of C.V. operators to automatically reconstruct models of buildings with a high level of detail, and "giving more possibilities to study objects thanks to the complex and absolute interactivity between the real object (point clouds and photographs), and virtual system of 2D and 3D digital models" (Attenni et al., 2017).

We have given here an overall review and some methodological improvements for its operativity, an unusual consideration, which can contribute effectively to achieve more accurate results in multi-image based projects related with heritage architectural models. "Computer vision techniques and procedures permitted to add a major automation (even to photogrammetric processes), but, in particular, they allow to model more complex objects, only exploiting the algorithms implemented in the common software tools." (Aicardi et al., 2018).

Fig. 9: Stereo point clouds obtained with Medium / High settings in PM (top) and PS (bottom), respectively, without pre-DSM filtering. It can be seen how a greater demand in the SFD + Matching and SGM / MVS algorithms, generally produces denser models, with a more uniform distribution of points and fewer clusters of outliers. This ensures a more precise metric evaluation, and a more sharp and uniform posterior mesh. Note: the displayed models are incomplete because the aerial photos are missing

REFERENCES

Agisoft LLC. (2015). *Agisoft Photoscan Professional (Ver. 1.2.0)*. Retrieved from http://www.agisoft.com

Aicardi, I., Chiabrando, F., Lingua, A., & Noardo, F. (2018). Recent trends in cultural heritage 3D survey: The photogrammetric computer visión approach. *Journal of Cultural Heritage*, 32, pp. 265.

Altunas, C., Mert, S., Yaman, G., Cengiz, Y., & Sonmez, M. (2019). Photogrammetric wireframe and dense point cloud 3d modelling of historical structures: the study of sultan Selim mosque and Yusuf Aga library in Konya, Turkey. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XLII-2/W11, pp. 79.

Atteni, M., Bartolomei, C., Inglese, C., Ippolito, A., Morganti, C., & Preari, G. (2017). Low Cost Survey and Heritage Value. *Scires IT - SCIentific RESearch and Information Technology*, Vol 7-2, 115-132.

Barazzetti, L., Remondino, F., & Scaioni, M. (2010). Extraction of accurate tie points for automated pose estimation of close range blocks. *ISPRS*, Vol. XXXVIII, Part 3A, Commission III – WG 1, pp. 152. Doi: 10.2423/i22394303v7n2p115

Barazzetti, L., Forlani, G., Remondino, F., Roncella, R., & Scaioni, M. (2011). Experiences and achievements in automated image sequence orientation for close-range photogrammetric projects. *Proc. of SPIE,* Vol. 8085. doi: 10.1117/12.890116.

Bethmann, F., & Luhmann T. (2015). Semi-Global Matching in Object Space. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XL-3/W2, 23-30.

Cabanes, J.L., & Bonafé, C. (2019). Toward Hybrid Modeling and Automatic Planimetry for Graphic Documentation of the Archaeological Heritage: The Cortina Family Pantheon in the Cemetery of Valencia. *International Journal of Architectural Heritage*, 14(8), 1210-1220. doi: 10.1080/15583058.2019.1597214.

Cloud Compare (2019). *Cloud Compare V2*. Retrieved from http://www.cloudcompare.org

Debevec, P.E., Taylor, C.J., & Malik, J. (1996). Modeling and Rendering Architecture from photographs: A hybrid geometry- and image-based approach. *SIGGRAPH 96 Conference proceedings*, 1-10.

Dorffner, L., & Forkert, G. (1998). Generation and Visualization of 3D photo-models using hybrid block adjustments with assumptions on the object shape. *ISPRS Journal of Photogrammetry and Remote Sensing, Issue*, 53-6, 369-378.

Elkhrachy, I. (2020). Modeling and Visualization of three dimensional objects using Low-Cost terrestrial photogrammetry. *International Journal of Architectural Heritage*, 14 (10), 1456–1467. https://doi.org/10.1080/15583058.2019.1613454

Förstner, W. (2004). Projective Geometry for Photogrammetric Orientation Procedures. *ISPRS Congress, Istanbul*, 1-164

Gao, J., Seon, K., & Brown, M.S. (2011). Constructing image panoramas using dual-homography warping. *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition,* 49-56.

Haala, N., & Rothermel, M. (2012). Dense multiple stereo matching of highly overlapping UAV imagery. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XXXIX-B1, 387-392.

Hänsch, R., Drude, I., & Hellwich, O. (2016). Modern methods of bundle adjustment on the GPU. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume III-3, 43-48.

Hugin, (2012). *Hugin Panorama Stitcher*. Retrieved from http://hugin.sourceforge.net/docs/manual/Projections.html

Kraus, K. (1993). *Photogrammetry, Vol. 1*. Ferd. Umler Ed., Bonn, pp. 367. ISBN 3-427-78684-6.

Lindequist, C. (2010). *3D Reconstruction of buildings from images with automatic façade refinement*. Master's thesis in Vision, Graphics and Interactive Systems, Aalborg University, 41-46. Retrieved from http://es.aau.dk,

Microsoft Research (2014). *A New Spin for Photosynth*. Retrieved from https://www.microsoft.com/en-us/research/blog/new-spin-photosynth/

Microsoft Research (2015). *Image composite editor*. Retrieved from http://research.microsoft.com/en-us/um/redmond/projects/ice/.

Pollefeys, M., Vergauwen, M., & Van Gool, L. (2000). Automatic 3d modeling from image sequences. *ISPRS*, Vol. XXXIII, 3-10.

PhotoModeler Technologies (2017). *PhotoModeler Scanner*. Retrieved from https://www.photomodeler.com/

Previtali, M., Barazzetti, L., & Scaioni, M. (2011). Multi-step and Multi-photo matching for accurate 3D Reconstruction. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 38 (3/W22), 103-108.

Re, C., Roncella, R., Forlani, G., Cremonese, G., & Naletto, G. (2014). Evaluation of an Area-based matching algorithm with advanced shape models. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XL-4, 215-221.

Remondino, F., El-Hakim, S., Girardi, S., Rizzi, A., Benedetti, S., & Gonzo, L. (2009). 3D Virtual Reconstruction and Visualization of Complex Architectures. *Proceedings of the ISPRS* Working Group V/4.

Rodehorst, V., Heinrichs, M., & Hellwich O. (2008). Evaluation of relative pose estimation methods for multi-camera setups. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVII, Part B3b, 135- 140.

Skarlatos, D. & Kiparissi, S. (2012). Comparison of laser scanning, photogrammetry and SFM-MVS pipeline applied in structures and artificial surfaces. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume I-3, 299-304.

Szeliski, R. (2011). *Computer Vision. Algorithms an Applications*. Ed. Springer, Switzerland, 3-28. eBook ISBN 978-1-84882-935-0

Visual Computing Laboratory (2016). *MeshLab software*. Retrieved from http://www.meshlab.net/

Xiang, T., Xia, G.S., & Zhang, L. (2016). Image stitching with perspective-preserving warping. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*. WG III/3, 287-294.