

MODELADO 3D A PARTIR DE VIDEOS AÉREOS Y TERRESTRES ADQUIRIDOS CON CÁMARAS DE CONSUMO. LEVANTAMIENTO DE LA TORRE DE CALA MORESCA EN MONTE ARGENTARIO.

FROM AERIAL AND GROUND-BASED VIDEOS ACQUIRED WITH CONSUMER CAMERAS TO THE 3D MODEL OF THE CALA MORESCA TOWER ON MOUNT ARGENTARIO.

Fabio Lanfranchi; orcid 0000-0002-2768-8314

'SAPIENZA' UNIVERSITÀ DI ROMA

Piero Barlozzini; orcid 0000-0001-5341-466X

UNIVERSITÀ DEL MOLISE

Luca Martelli; orcid 0000-0003-3433-006X

'SAPIENZA' UNIVERSITÀ DI ROMA

doi: 10.4995/ega.2025.21756

En el ámbito del levantamiento 3D, el papel de los drones y cámaras estabilizadas resulta ser siempre más importante. Dado que existe la posibilidad de superar los límites representados por sitios de grabación que son normalmente inaccesibles y la oportunidad de desarrollar modelos digitales que son siempre más precisos, los aparatos de consumo como método de adquisición de datos se están confirmando como instrumentos eficaces y utilizables desde la escala arquitectónica a la territorial. En el ámbito del software, algunos programas de modelado Structure From Motion (SfM) permiten procesar videos automáticamente, seleccionando los fotogramas más adecuados para los algoritmos de procesamiento. El ámbito de aplicación elegido es la Torre de Cala Moresca, que está situada en el Monte Argentario en Toscana y fue construida por la Corona de España en el siglo XVI. Uno de los objetivos del estudio es ilustrar el flujo de trabajo adoptado para

definir un modelo de reconstrucción virtual de la torre, desarrollado a partir de grabaciones realizadas desde aparatos aéreos y terrestres equipados con sensores de bajo coste y procesadas automáticamente por el respectivo programa.

PALABRAS CLAVE: TORRE DE CALA MORESCA, SFM DE VIDEO, LEVANTAMIENTO INTEGRADO, VALIDACIÓN DEL MODELO, RECONSTRUCCIÓN DIGITAL

In the field of 3D surveying, drones and stabilized cameras are playing an increasingly important role. The ability to overcome limits regarding normally inaccessible points for taking pictures/videos and the opportunity to process increasingly precise digital models based on data acquired using consumer devices means such devices

are proving to be effective tools when dealing with the architectural and regional scales. With respect to software, some structure-from-motion (SfM) modelling programs allow videos to be processed automatically, selecting the most appropriate frames for the algorithms. The chosen case study is the Tower of Cala Moresca. This is situated on the Tuscan promontory of Mount Argentario and was built by the Spanish Crown in the 16th century. One of the objectives is to simplify the adopted workflow to define a model of the virtual reconstruction of the tower developed using only video filmed with aerial and ground-based devices equipped with low-cost sensors and processed automatically by the software.

KEYWORDS: TOWER OF CALA MORESCA, SFM FROM VIDEO, INTEGRATED SURVEY, MODEL VALIDATION, DIGITAL RECONSTRUCTION



Introducción

Este artículo corresponde al ámbito más amplio de una investigación universitaria en curso, focalizada en el conocimiento y divulgación del patrimonio cultural de los Presidios de Toscana. Una de las finalidades del trabajo es la realización de un modelo virtual de reconstrucción del estado original de la torre de Cala Moresca, desarrollado y basado en un modelado 3D SfM generado exclusivamente por medio de videos obtenidos con aparatos de consumo. En el ámbito del modelado para reconstruir el estado actual, los objetivos prefigurados incluyen la evaluación de la precisión del producto generado y la identificación de los posibles límites en términos de definición gráfica causados por el uso de sensores pequeños.

En los siguientes puntos, se puede resumir en pocas palabras el flujo de trabajo adoptado:

- Medios de adquisición para procesar fotogramas SfM sacados exclusivamente de videos filmados con aparatos aéreos y terrestres equipados con sensores pequeños.
- Medios de elaborar el modelado del levantamiento procesado por el software a través de la selección automática de fotogramas extraídos de los videos.
- Proceso de validación del modelo por medio de pruebas de precisión métrica, tanto puntuales como en general, confirmadas a través de campañas de adquisición puntual o de *Estación Total* (ET) y en masa por medio de *Terrestrial Laser Scanning* (TLS)

Estudio de caso, contexto histórico-geográfico

La torre de Cala Moresca pertenece al sistema defensivo de la costa,

construido a lo largo del siglo XVI entre Orbetello y parte de la isla de Elba (Fig. 1). Después de la victoria de Cosime I de Médici —aliado con la Corona de España— sobre la República de Siena —aliada con los franceses— en el 1557, Felipe II de España, virrey de Nápoles y sucesor de Carlos V, puso en marcha los acuerdos establecidos anteriormente, reservándose la franja litoral llamada Presidios de Toscana. En el enclave se establecieron militares para contrastar las incursiones, siempre más frecuentes, de piratas. Sucesivamente, Pedro Afán de Ribera mejoró el sistema de defensa en la zona durante la década de su mandato (1559–1571), renovando la red de torres litorales y construyendo muchas fortalezas. La Torre de Cala Moresca, ubicada en la costa occidental del promontorio de Monte Argentario, cerca de 35 metros sobre el nivel del mar y hoy en ruinas, probablemente perteneció a estas nuevas construcciones. (Fig. 2). Su función era ofrecer un punto de observación y defensa de la sección del mar debajo del promontorio y señalar la llegada de barcos enemigos a la guarnición desplegada para proteger la zona (LUISI, 1979). La ausencia de la construcción en la cartografía del siglo XVIII y entre los planos topográficos de las torres de la zona producidos por los ingenieros militares franceses en los primeros años del signo XIX puede sugerir su declive prematuro como puesto militar desde el punto de vista de importancia estratégica. Hoy en día, solo el basamento y una parte del alzado aún permanecen de la disposición original. El volumen en forma tronco-piramidal, caracterizado por la inclinación de 7° del escarpe, se desarrolla en una base rectangular de más o menos 11 x 8,5 metros colocada sobre eutinte-

Introduction

This contribution falls under the broader area of ongoing university research aimed at understanding and spreading the cultural heritage of the State of the Presidi. One of the goals of the work is to create a virtual model reconstructing the original state of the Tower of Cala Moresca, developed on the basis of a 3D SfM survey generated exclusively using videos acquired with consumer devices. With respect to reconstructing the current state, one of the objectives is to assess the precision of the final product and identify possible limits in terms of graphical definition due to the small size of the sensors.

Very briefly, the adopted workflow can be summarized in the following steps:

- Acquisition methods for processing SfM photographs taken exclusively from videos filmed using aerial and ground-based devices equipped with small sensors;
- Means of developing the survey model, processed with software to automatically selecting frames extracted from videos;
- Validating the model by checking the detailed and overall precision through point-like acquisition using a *total station* (TS) and large-scale acquisition via *terrestrial laser scanning* (TLS).

Case study and historical/geographical context

The Tower of Cala Moresca pertains to the coastal defensive system that was developed in the 16th century between Orbetello and part of the Island of Elba (Fig. 1). After the victory of Cosimo I de' Medici (allied with the Spanish Crown) in 1557 over the Republic of Siena (allied with the French), Philip II of Spain, the heir of Charles V, implemented prior agreements, reserving for himself the coastal area of Tuscany, which was called State of the Presidi. The military was stationed in the enclave to combat the increasingly frequent pirate raids. During the decade of his rule, Pedro Afán de Ribera, the Viceroy of Naples, was urged by Philip II to improve the defensive system, renewing the network of coastal towers and building numerous imposing fortresses (DELLA MONACA et al., 1999). These new constructions probably included the Tower of Cala Moresca, in ruins today, located on the western coast of the Argentario Promontory

about 35 metres above sea level (Fig. 2). Its purpose was to offer a point of observation and defence for the stretch of sea around the promontory, as well as allowing troops to notify the garrison protecting the area of the arrival of enemy ships (LUISI, 1979). The absence of the construction in 18th-century maps and survey drawings of towers in the area made by French military engineers in the early 19th century would suggest its early decline as a military stronghold in terms of strategic importance. Today, only the base and a residual part of the elevation remain. The truncated pyramid shape, with a scarp inclined at 7° rises above a rectangular base measuring about 11 m x 8.5 m set on a partially visible euthyteria measuring 12.25 m x 9.92 m, corresponding to 21 x 17 Tuscan *braccia*. The maximum remaining height of the tower is about 7 metres. Today the structure is delimited at the top by a toroid-shaped oversailing course coinciding with the height of the first floor and marking the vertical faces. Within, there are two rooms that cannot be accessed due to safety reasons, one very probably destined to hold provisions (LICINIO, 1994) and the second arranged as a basin to collect rainwater falling from the roof.

Method, devices and software, workflow

The 3D survey model was developed via SfM reconstruction based on photographs automatically extracted by the software from aerial and ground-based videos. In the field of digital modelling, the *speeded up robust features* (SURF) and *scale invariant feature transform* (SIFT) algorithms are indispensable (LOWE, 1999; RESCH et al., 2015). The automated determination of homologous points between the photographs (TUYTELAARS AND MIKOLAJCZYK, 2007) enabled 3D point clouds to be generated even from unorganized collections of digital photographs (AGARWAL et al., 2011; SNAVELY et al., 2008). In SfM processing, refinement of the statistical calculation is not key, since the guiding parameters may be influenced by other factors, such as the type of optical sensor or ambient light. In practice, a comparison of the results obtained using a *full-frame* (FF) camera with those obtained using a digital video camera with





1. Imagen superior, localización geográfica. Imagen central, «Pianta del Littorale Toscano, dalla Torre S. Vincenzo, a tutto lo Stato dei Presidj» carta, tinta china y acuarela, 562 x 890 mm. Archivio di Stato di Firenze, Donación: Planos del Scrittoio delle Fortezze e Fabbriche, ID 6 (PRUCKER, s.f.). Imagen inferior, detalle del plano con indicación del sitio de la Torre de Cala Moresca.
2. Vista desde el este de la Torre de Cala Moresca.

1. Top: geographical setting. Middle: 'Pianta del Littorale Toscano, dalla Torre S. Vincenzo, a tutto lo Stato dei Presidj', paper, Indian ink, and watercolour, 562 mm x 890 mm. Archivio di Stato Firenze, Donation: Pianta dello Scrittoio delle Fortezze and Fabbriche, ID 6 (PRUCKER, undated). Bottom: detail of the plan indicating the location of the Tower of Cala Moresca.
2. View of the Cala Moresca Tower from the east.



2

rias parcialmente visibles de 12,25 x 9,95 metros, equivalentes a 21 x 17 *braccia toscanas*. La altura máxima sobrante de la torre es aproximadamente 7 metros. Hoy en día, la parte superior de la estructura está delimitada por una cornisa en forma de toroide que coincide con la altura del primer nivel del encofrado y marca las fachadas verticales. En el interior, se encuentran dos habitaciones a las cuales no se puede acceder debido a las condiciones de seguridad; una era muy probablemente destinada a guardar provisiones (LICINIO, 1994), mientras la segunda está estructurada como un tanque para recoger el agua de lluvia del techo.

Metodología, aparatos y software, flujo de trabajo

El modelado 3D fue desarrollado por medio de reconstrucción SfM

basada en fotogramas extraídos automáticamente por el software de las grabaciones aéreas y terrestres. En el ámbito del modelado digital, los algoritmos *Speeded Up Robust Features* (SURF) y *Scale Invariant Feature Transform* (SIFT) son componentes indispensables (LOWE, 1999; RESCH et al., 2015). La determinación automática de puntos homólogos entre fotogramas (TUYTELAARS y MIKOLAJCZYK, 2007) permitió la generación de nubes 3D de puntos, incluso de fotos digitales no organizadas (AGARWAL et al., 2011; SNAVELY, et al., 2008). En el procesamiento SfM, el nivel de precisión del cálculo estadístico no es el aspecto principal: los parámetros que lo guían pueden ser influidos por factores como, por ejemplo, el tipo de sensor óptico utilizado o la luz ambiental. En la práctica, haciendo una comparación entre los resulta-

lower resolution shows that in the latter case, the identification of homologous points is more fluid and linear. This is due not only to the greater number of images that can usually be extracted from the films and the consequent improved continuity among the photographs. Indeed, the lower resolution of photographs taken with cameras to frames taken from videos corresponds to a 'simplification' of the mathematical/statistical problem that the SURF and SIFT algorithms must address, thus making the processing more robust at the expense of detail. On the contrary, a lower sampling resolution and lower quality of focusing and lighting are obtained. It is therefore necessary to consider the expected level of detail in relation to the desired degree of representation. Given the above, the ability to perform SfM processing using videos taken with consumer devices, whether aerial or ground-based, is important in terms of containing the purchase price of the equipment, as well as the time needed for data acquisition in the field. In this specific case, the acquisition campaign was conducted using devices equipped with sensors smaller than 1". For the exterior

of the tower, an *unmanned aircraft system* (UAS) was used, while a small video camera equipped with electronically stabilized gimbals attached to a telescopic arm was used for the interior, which could not be accessed due to the precarious state of the construction. With respect to hardware and software, the topographic survey was conducted using a Leica TCR703 TS, while the TLS acquisitions were made with a FARO Focus M70 3D laser scanner. Aerial shots were taken from a DJI Phantom 4 UAS, equipped with a 35-mm-equivalent camera with aperture $f/2.8$ and $1/2.3''$ CMOS sensor. Shots inside the tower were taken using a DJI Osmo Pocket 3 with electronically stabilized camera, equivalent to 20 mm with aperture $f/2.0$ and $1''$ CMOS sensor. To balance the white among the various films, the calibrated Color Checker Video X-Rite table was framed before each run. With respect to software, the photograph models were processed using 3DF Zephir® v.7.013 by 3DFLOW Srl; the TLS model was constructed using JRC Reconstructor® v. 4.3.2 by Gexcel; Adobe Premiere Pro was used to edit the films and correct points of white and the primary colours; textures and 2D graphics were compared, modelled, and processed respectively using the open-source software CloudCompare v. 2.11.3 and Rhinoceros 8, while 2D models were developed using Autocad® by Autodesk.

To validate the measurements and assess the accuracy of the SfM model, the workflow was organized into three different phases (Table 1):

- acquisition
- processing
- verification (validation) of the results

Acquisition

The 3D acquisition using active sensors was designed to verify the accuracy of the models developed using data from passive sensor devices, specifically video cameras. With respect to the first acquisitions using a TS, a topographic polygon was defined in relation to 5 stations situated at the edges of the area of study (HARWIN and LUCIEER, 2012). The stations, marked as *ground control points* (GCP), were used to orient the products of the TLS and SfM (Fig. 3). An integral part of the topographic network also consisted of an additional 26 natural points collimated on

dos obtenidos con una cámara *Full Frame* (FF) y aquellos obtenidos con una videocámara digital de más baja resolución, se puede notar que en el segundo caso, la identificación de puntos homólogos es más fluida y lineal. Este resultado no se debe solo al número mayor de imágenes que se pueden extraer normalmente de las grabaciones videos ni a la consecuente mejor continuidad entre fotogramas. En realidad, la menor resolución de los fotogramas obtenidos con videocámaras más pequeñas corresponde a una «*simplificación*» del problema matemático-estadístico que los algoritmos SURF y SIFT deben afrontar, confirmando al procesamiento una mayor robustez, obviamente en detrimento del nivel de detalle. Al contrario, se obtiene una resolución de muestreo inferior y una calidad menor de enfoque e iluminación. Por lo tanto, es necesario considerar el nivel de detalle esperado en relación con el nivel de representación deseada. Teniendo en cuenta lo antedicho, la posibilidad de llevar a cabo el procesamiento SfM de grabaciones video con aparatos de consumo, tanto aéreos como terrestres, conlleva un valor significativo en términos de reducción de gastos para comprar el equipo y las horas de trabajo dedicados a la adquisición en el campo. En este caso específico, se llevó a cabo la campaña de adquisición por medio de aparatos equipados con sensores que no superan $1''$. Eso fue hecho al exterior de la torre por medio de un *Unmanned Aircraft System* (UAS), y al interno -donde no se podía acceder debido a las condiciones precarias de la construcción- por medio de una pequeña videocámara equipada con un cardán estabilizado electrónicamente y fijado a una vara telescópica. En relación con el

hardware utilizado, el levantamiento topográfico fue realizado con la ET Leica TCR703 mientras que el escáner láser 3D FARO Focus M70 fue utilizado para las adquisiciones TLS. Las tomas aéreas fueron realizadas con un UAS DJI Phantom 4 provisto de una cámara de 35 mm con abertura $f/2.8$ y sensor $1/2.3''$ CMOS. Para la grabación dentro de la torre, se utilizó un DJI Osmo Pocket 3 con cámara estabilizada electrónicamente, equivalente a 2 mm con abertura $f/2.0$ y sensor CMOS de $1''$. Para uniformizar el equilibrio de blanco entre las varias grabaciones, la tabla de colores calibrados Color Checker Video X-Rite fue enmarcada antes de cada grabación. En cuanto al software, el procesamiento de los modelos fotogramétricos fue realizado con 3DF Zephir® v.7.013 de 3DFLOW Srl; la construcción del modelo digital TLS con JRC Reconstructor® v. 4.3.2 de Gexcel; el montaje de las grabaciones, la corrección del punto de blanco y de los colores primarios con Adobe Premiere Pro; los procesamiento comparativos, el modelado y el procesamiento de texturas y gráficos 2D con el software de código abierto CloudCompare v. 2.11.3 y Rhinoceros 8; y los modelos 2D con el software Autocad® de Autodesk. Para la confirmación métrica y de precisión del modelo SfM, el flujo de trabajo fue organizado en tres fases distintas (Tabla 1):

- adquisición
- procesamiento
- verificación (validación) de los resultados

Fase de adquisición

El objetivo de la adquisición 3D con sensores activos era comprobar la precisión de los modelos elaborados

3. Panorámica nadiral del promontorio. En el centro, la torre de Cala Moresca; los puntos rojos indican los CGP (puntos de estación) correspondientes a los vértices del polígono. Tabla 1. Flujo de trabajo para definir y validar el modelado generado por medio de SfM. El hardware y software utilizados están indicados en el esquema.

3. Nadiral view of the promontory. The Tower of Cala Moresca is at the top, with the CGP (station points) corresponding to the vertices of the polygon shown in red. Table 1. Workflow for defining and validating the survey model generated using SfM. The diagram shows the hardware and software used.



3

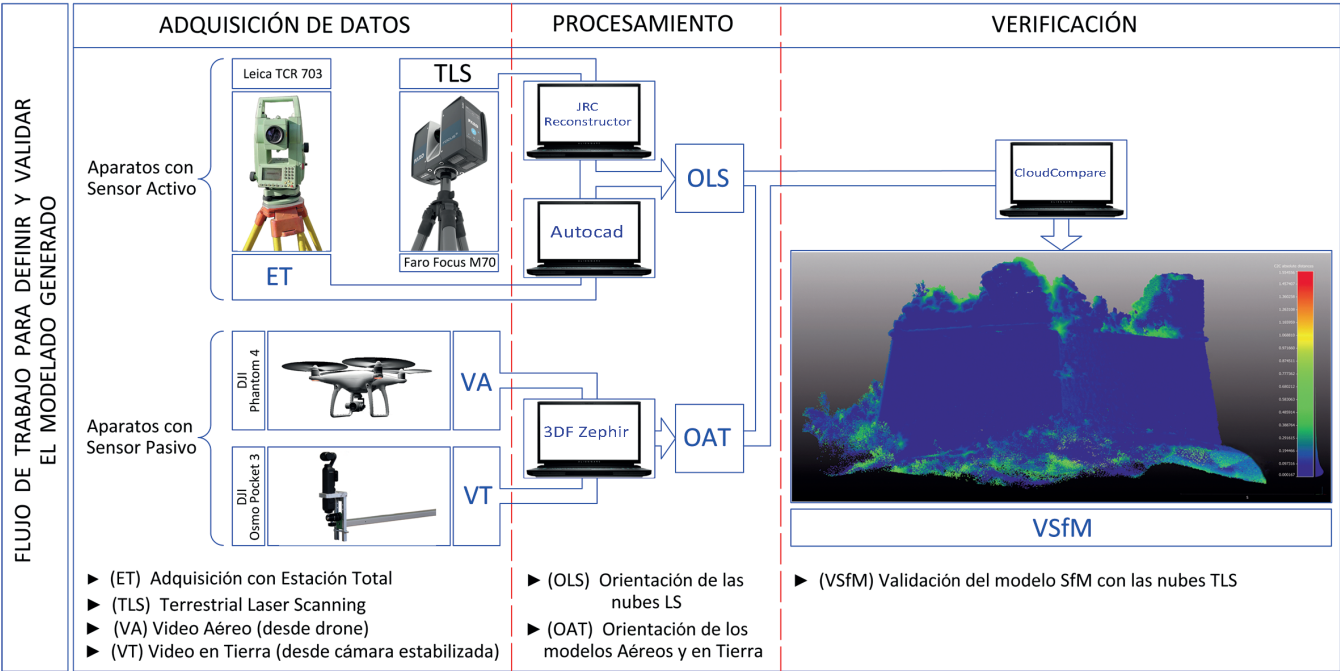


TABLA 1

the construction and partly used as GCPs and partly as *control points* (CP). When selecting the points, particular attention was paid to covering the tower completely to minimize both the overall alignment errors and possible inaccuracies (mixed pixels) typical of TLS acquisitions made from distant and separate centres of projection (LICHTI, 2007; WANG and CHENG, 2016). The TS and reflecting prism were stationed with forced centring to localize possible errors of eccentricity in the individual stations. For redundancy, further checks were made to verify the campaign measurements by confirming the spatial coincidence between the same points collimated by different stations. For the TLS acquisition campaign, eight scans were made for a total of just over 23 million points relating to the construction (Fig. 4). The resolution of seven scans was set on a spherical grid with a radius of 10 m and 7.7-mm step. The eighth, used to 'connect' the interior and exterior of the tower, was made using the TLS situated opposite to the only embrasure present in the base. In this case, the scan was set with a resolution of 7.7 mm at 20 m in order to detail the framed part of interior stonework with higher precision. As mentioned above, a UAS with passive sensor devices was used for the exterior, while a video camera managed via Wi-Fi was used for the interior. Contrasting lighting was a key feature of this phase of the work: natural, diffuse light for the exterior shots compared to critical conditions for the interior. Two missions were conducted for the UAS campaign: the first was dedicated to acquiring the tower from the nadir via a programmed flight (*waypoint*), while the other was dedicated to acquiring the elevations, which was done manually, given the presence of trees and thick brush, even against the tower. Particular care was taken regarding the percentage of overlap between the frames, given the ratio between *frames per second* (fps), set at 30, and the speed of the drone. The elevations of the construction were acquired by shooting one face at a time in the course of a single anti-clockwise revolution (Fig. 5). To balance the white among the various videos, the calibrated colour table was framed before each run. All UAS videos were joined in a single 7.51-GB MP4 file lasting 20:00 minutes. With respect to the interior videos, taken using

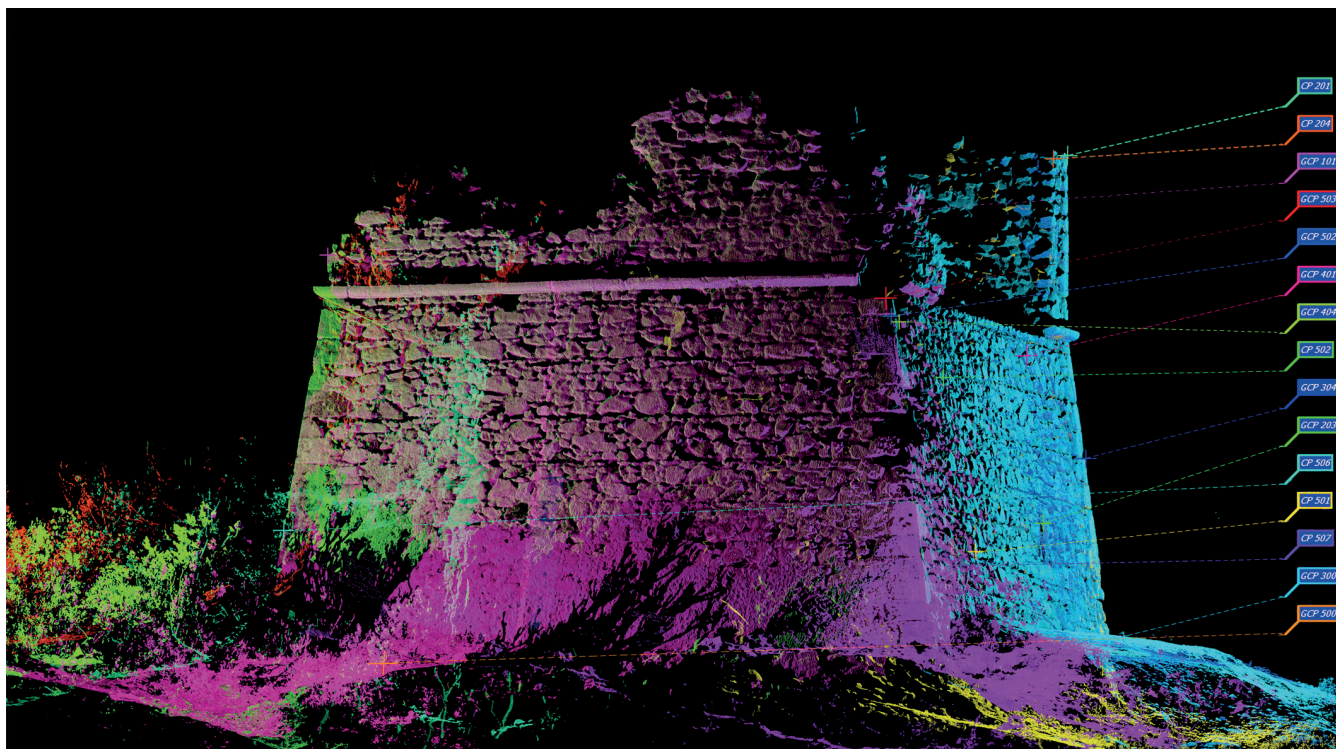
con datos adquiridos por medio de aparatos con sensor pasivo, videocámaras específicamente. En relación con los sensores activos, una ET fue utilizada para definir un polígono topográfico basado en las cinco estaciones posicionadas en los bordes de la zona de estudio (HARWIN y LUCIEER, 2012). Las estaciones, señaladas como puntos de control terrestre (GCP por la sigla in inglés), fueron utilizadas para orientar los productos del TLS y del modelo SfM (Fig. 3). Una parte integral de la red topográfica consistía en 26 puntos naturales colimados usando la construcción; algunos de ellos fueron utilizados como limitación GCP y otros como *Puntos de Control* (PC). Para seleccionar los puntos, particular atención fue dedicada a la cobertura completa de la torre para minimizar tanto los errores de alineamiento como las posibles imprecisiones «*mixed pixel*» típicas de las adquisiciones TLS realizadas desde centros de proyección lejanos y escondidos (LICHTI, 2007; WANG y CHENG, 2016). La ET y el prisma reflectante fueron colocados con centramiento forzado para localizar posibles errores de excentricidad en la estación individual. Más inspecciones redundantes fueron hechas para verificar las mediciones de campo por medio de comparación de la coincidencia espacial entre puntos iguales colimados por estaciones distintas. Para la campaña de adquisición TLS, ocho escaneos fueron hechos por un número total —relativos solo a la construcción— de poco más de 23 millones de puntos (Fig. 4). La resolución de siete escaneos fue estructurada en una malla esférica con un radio de 10 metros y paso de 7,7 mm. El octavo escaneo fue utilizado como «*conexión*» entre exterior e interior

de la torre, y fue realizado con el TLS situado enfrente de la única tronera presente en el basamento. En este caso, el escaneo fue fijado con una resolución de 7,7 mm a 20 metros para detallar la mampostería interna enmarcada con más precisión. Como antes mencionado, para la adquisición a través de aparatos con sensor pasivo, un UAS fue utilizado para el exterior de la torre y una videocámara gestionada a través de Wi-Fi fue utilizada para el interior. Esta fase del trabajo fue caracterizado por situaciones de iluminación contrastada: luz natural y difusa para las tomas externas en contraste con condiciones críticas en el interior. La campaña UAS se organizó en dos tareas: adquirir una vista nadiral de la torre por medio de un vuelo programado (*waypoint*) y obtener los alzados por medio de control manual, considerando la presencia de árboles y arbustos abundantes, incluso apoyados en la torre. Atención particular fue dedicada a los porcentajes de cobertura entre fotogramas, considerando la relación entre *fotogramas por segundo* (fps) —fijado en 30— y la velocidad del dron. Los alzados de la construcción fueron obtenidos filmando las fachadas una por una durante una única revolución en sentido antihorario (Fig. 5). Con la finalidad de uniformizar los parámetros relativos al equilibrio del blanco entre las varias tomas, la tabla de colores calibrados fue enmarcada antes de cada filmación. Todas las tomas del UAS fueron incorporadas en una única grabación MP4 de 7,51 GB y de una duración de 28:00 minutos. En relación con las tomas del interior, hechas por medio de la videocámara introducida en la tronera del basamento, particular atención fue dedicada a la coordinación de los modos



4. Vista total de la torre con mapas de gama de los ocho escaneos TLS.

4. Overall view of the tower with gamma maps of the eight TLS scans.



4

de movimiento entre aquellos aptos para la filmación y aquellos aptos para la maniobra de la vara telescópica de 10 metros, en la cual estaba sujeta la videocámara (Fig. 6). Dadas las condiciones de luz baja y discontinua en el interior, simulaciones de filmación fueron hechas antes de las adquisiciones para identificar un valor de *sensor light sensitivity* (ISO) eficaz para el procesamiento SfM. Lamentablemente, las condiciones operativas no posibilitaron el uso de iluminación artificial para reducir los valores ISO. En este caso específico, el valor fijado en 3200 permitió grabaciones nítidas y con un nivel de perturbación (estático) aceptable durante el procesamiento. Teniendo en cuenta la limitación de rotación azimutal del objetivo (-235° – 58°), la adquisición fue hecha en dos tomas, rodadas a 30 fps y luego incorpora-

das en un único MP4 de 7,85 GB con una duración de 14:00 minutos.

Fase de procesamiento

En cuanto al procesamiento de datos ETS, una vez que se eliminaron los valores anómalos, como por ejemplo puntos repetidos e incongruentes, las nubes fueron alineadas con procesos del punto más cercano iterativo (ICP por la sigla en inglés) relativos a los objetivos CGP y PC para reducir las incertidumbres del cálculo. Considerando también la buena visibilidad entre las estaciones, las relativas nubes fueron procesadas con el método de ajuste de paquetes «*bundle adjustment*»; el error medio de alineamiento de los puntos correspondientes fue de 0,4 mm, y el error medio ICP fue igual a 2 mm. Con respecto a los modelos SfM, para evaluar la capacidad automática de la recons-

the video camera inserted in the embrasure of the base, particular attention was paid to coordinating the movement between what was suitable for the shot and what was suitable for swinging the 10-m telescopic arm bearing the video camera (Fig. 6). Given the low, discontinuous light in the interior environment, filming simulations were made before the acquisitions to identify a *sensor light sensitivity* (ISO) suitable for the SfM processing. Unfortunately, the operating conditions did not enable the use of artificial lighting to reduce the ISO values. In this case, the value set at 3200 yielded clear shots with an acceptable level of noise that was eliminated during processing. Due to the limit in azimuthal rotation of the video camera lens (-235° – 58°), the films were acquired in 2 runs set at 30 fps and then encompassed in a single 7.85-GB MP4 file lasting 14:00 minutes.

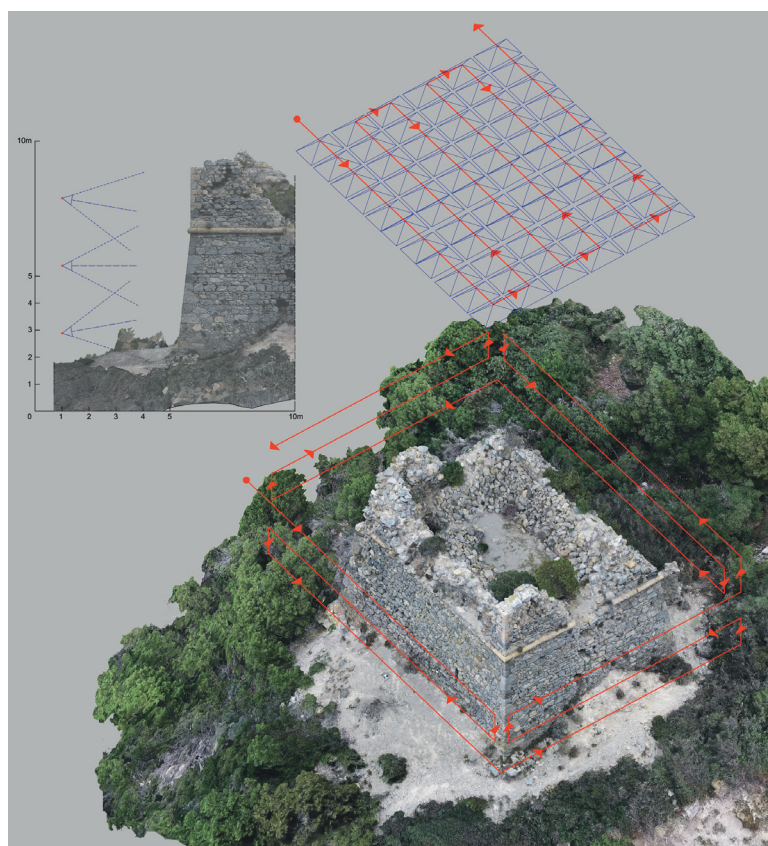
Processing

With regard to processing the TLS data, the anomalous values — such as duplicate and inconsistent points — were removed to reduce the calculation uncertainties, and

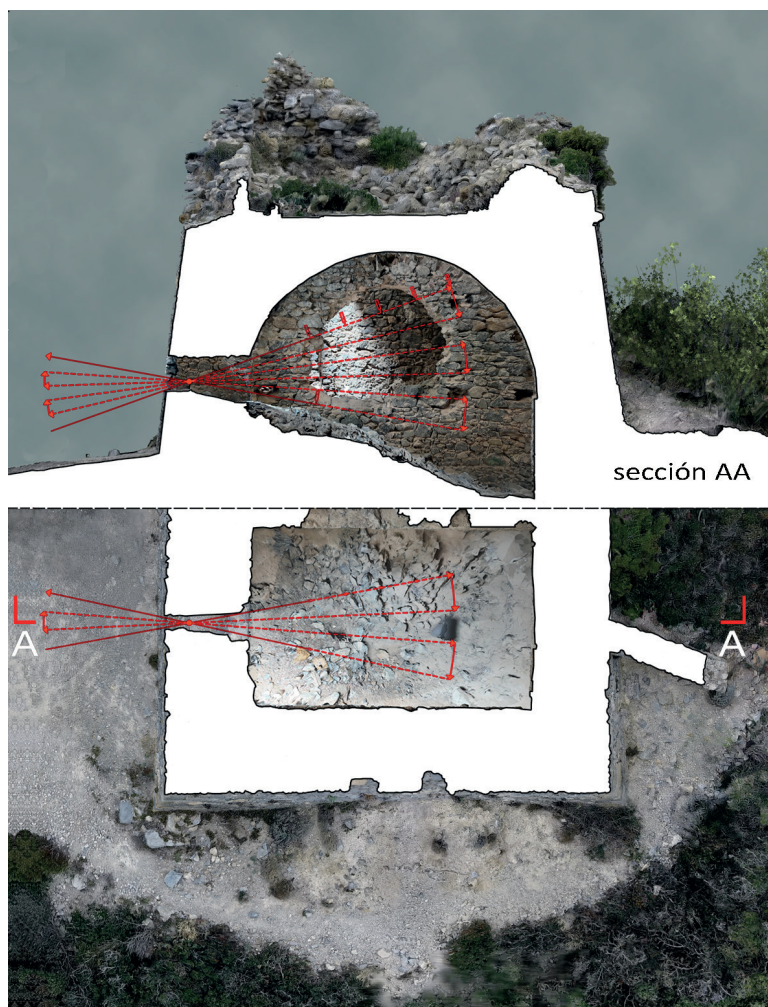
the point clouds were aligned using *iterative closest point* (ICP) procedures referring to the GCP and CP targets. In addition, given the good visibility between the stations, the relative clouds were processed using the bundle adjustment method, with the average alignment error between corresponding points equal to 0.4 mm, and an average ICP error equal to 2 mm. With respect to the SfM models, any corrections of the parameters of the input films was avoided in order to assess the automatic photo reconstruction capabilities of the software. Two different methods were adopted to minimize errors with respect to orientation, alignment, and metric adjustment of the sparse clouds of the aerial and ground-based models, defined respectively using 1865 photos extracted from the software in the first case, for slightly more than 367,000 points, and 1019 photos for just over 263,000 points in the second case. For the aerial model, bundle adjustment was used with the GCP station points in the topographic network, along with natural CPs collimated on the construction (Fig. 7). For the alignment, the *final average residual* was equal to 0.229 cm and the *root mean square error* (RMSE) was equal to 0.261 cm. The point cloud of the interior was aligned automatically by the software and subsequent use of the homologous point coordinates identified in the TLS cloud of the interior enabled its orientation with respect to the local coordinate system. A further check was made by exporting the SfM model to the software used to process the TLS clouds. In this respect, the dense cloud was joined with the cloned TLS 'connection' cloud and reduced to just the interior portion to optimize the joining process using only the points of interest (Fig. 8). The segmented cloud used as a reference was joined with the interior SfM cloud using *bundle adjustment*. The results of the alignment yielded 0.0132 m of average error for the ICP and 0.0000 m of alignment error between corresponding points. Once the interior SfM model was verified, it was definitively joined to the exterior model.

Verification and validation of the SfM survey model

The overall accuracy of the SfM model was analysed by comparing the dense cloud, consisting of slightly more than 2 million



5



6



5. Abajo, modelo SfM de la torre con indicación del plan de vuelo definido para la adquisición video; izquierda, una sección de la torre con las altitudes de vuelo y orientación de la videocámara; derecha, plan de vuelo para la adquisición de vistas nadirales.

6. Plano (extracto) y sección transversal de la torre: esquemas de movimiento de la vara telescópica utilizados para la adquisición video del interior con cámara remota estabilizada.

7. Nube densa del modelo SfM: rojo, algunos puntos GCP y PC utilizados para la orientación.

8. Arriba a la izquierda, vista 2D del escaneo de «conexión» exterior-interior. Derecha, detalle de la nube de puntos TLS en blanco; en ocre, la nube densa de puntos SfM. Abajo, detalle de la unión entre la nube TLS segmentada y la nube SfM.

5. Bottom: SfM model of the tower, indicating the flight plan defined for the video acquisitions. Top left: extract of a section of the tower with the flight heights and video camera orientation. Top right: flight plan for acquisitions from the nadir.

6. Plan (excerpt) and cross section of the tower: diagrams showing the movement of the telescopic arm used for the video acquisition of the interior with the stabilized camera.

7. Dense cloud for the SfM model. Some GCPs and CPs used for the orientation are shown in red.

8. Top left: 2D view of the scan to 'connect' the interior and exterior. Right: detail of the TLS point cloud in white, SfM dense point cloud in ochre. Bottom: detail of the union between the segmented TLS cloud and the SfM cloud.

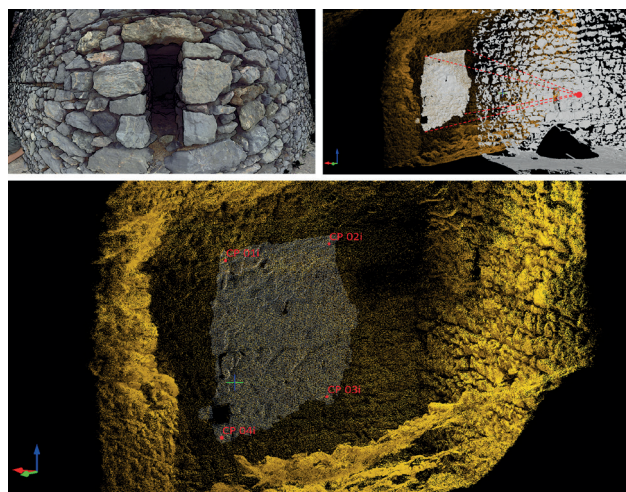
trucción fotogramétrica del software, ninguna corrección fue aportada a los parámetros de procesamiento. Dos métodos distintos fueron adoptados para minimizar los errores de orientación, de alineamiento y de ajuste métrico de las nubes escasas de los modelos aéreos y terrestres, definidos respectivamente por 1865 fotogramas extraídos del software para poco más de 367.000 puntos en el primer caso, y por 1019 fotogramas para alrededor de 263.000 puntos en el segundo. Para el modelo aéreo, una ajuste de paquetes fue implementada con los puntos de estación GCP de la red topográfica y con algunos PC naturales colimados en la construcción (Fig. 7). Los datos de alineamiento obtendidos resultaron ser igual a 0,229 cm para el *Residuo Medio Final* (RMF) e igual a 0,261 cm para la *Raíz del Error Cuadrático*

co Medio (RECM). El alineamiento de la nube de puntos de los interiores fue efectuado automáticamente por el software y el uso de las coordenadas de puntos homólogos identificados en la nube TLS del interior posibilitó la orientación con respecto al sistema de coordenadas locales. Otra inspección fue hecha exportando el modelo SfM en el software utilizado para procesar las nubes TLS. En este ámbito, la nube densa de fotogramas fue unida con la nube de «conexión» TLS, clonada y reducida a la porción interna, con la finalidad de optimizar el proceso de unión de solo los puntos de interés (Fig. 8). La nube segmentada utilizada como referencia fue unida con la nube SfM del interior por medio del ajuste de paquetes. Se confirmaron los resultados del alineamiento con 0,0132 m de error medio para el ICP y 0,0000 m de

points, and the joined TLS clouds, reduced to a number of points similar to that of the SfM model. The results show satisfactory congruence, given that the average deviations for just the masonry partitions are less than 1 cm. The most significant variations, as highlighted by colours deviating from the predominant blue, are found in the areas affected by vegetation rooted in the masonry partitions correspond to areas hidden in the shadows with respect to the TLS centres of acquisition, for example, the top of the construction. The deviations detected between the TLS and SfM clouds ultimately show values of overall congruence under the threshold of the so-called 'graphical uncertainty' permitted for renderings in 1:50 scale. With respect to the communicational effectiveness of the mesh model in video mode, the result seems satisfactory, especially given the complete absence of any intervention in the possible optimization options. In contrast, this refinement is essential if the model is used for educational purposes. In fact, as clear as the model is and given its fluid navigability with suitable hardware and specific software, the digital weight is not compatible with the requirements of the *real-time rendering* platform. Having verified the precision of the model after generating the textures, a series of horizontal and vertical planes were established to define the polygonal chains of the plans and sections (Fig. 10). Finally, the



7



8

communicational rendering was completed by generating a series of orthophotos relating to the plan and elevations (Fig. 11–14).

Conclusion

The results show that when studying a small architectural element with a variety of volumetric and location characteristics, the results of SfM modelling using video processing fall within the expressed scale limits. In this specific case, the experimentation yielded a quick, reliable, automatic model whose definition is suitable for both 'conventional' representations, i.e. those pertaining to 'traditional' communication, and for dynamic 3D models used for educational purposes. It is clear that the procedures used here require verification, especially in more complex or extensive areas. Of course, rapid technological development, even related to the small sensors contained in consumer devices -a key feature in the case of light UAS's and small stabilized cameras- presents an opportunity to achieve increasingly efficient, reliable results in the short term. Despite the agility and effectiveness of the SfM method, it must still be used in synergy with established 3D surveying methods, such as massive TLS acquisitions or 'classical' techniques such as TS surveys. Despite the need for orientation and sizing, this need may, as in this case, be indispensable for solving critical aspects depending on the underlying origin of the procedure: light. ■

References

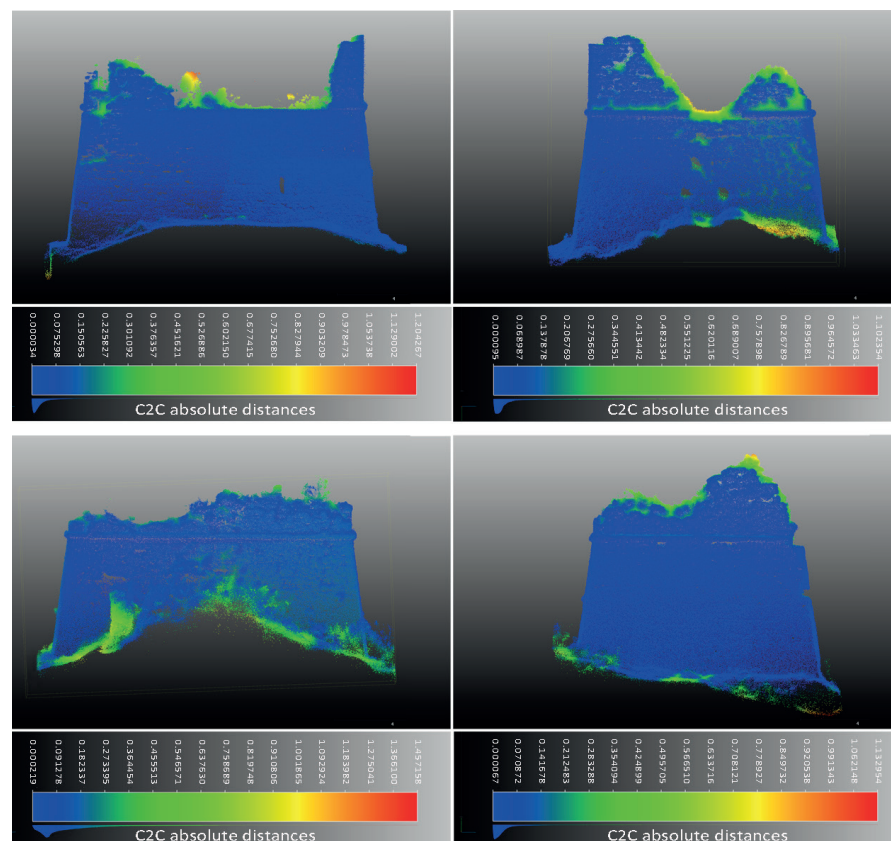
- AGARWAL, S., FURUKAWA, Y., SNAVELY, N., SIMON, I., CURLESS, B., SEITZ, S.M., SZELISKI, R., 2011. Building Rome in a day. *Communications of the ACM*, October 2011, Vol. 54 N. 10, pp. 105-112. <https://doi.org/10.1145/2001269.2001293>
- DELLA MONACA, G., ROSELLI, D., TOSI, G., 1999. *Fortezze e torri costiere dell'Argentario, Giglio e Giannutri*. Pitigliano: Editrice Laurum.
- HARWIN, S., LUCIEER, A., 2012. Assessing the Accuracy of Georeferenced Points Clouds produced via multi view stereopsis from Unmanned Aerial Vehicle (UAV) imagery. *Remote Sens.* 2012, 4, pp. 1573–1599. <https://doi.org/10.3390/rs4061573>

error de alineamiento entre puntos correspondientes. Después de la verificación, el modelo SfM del interior fue definitivamente integrado con los modelos del exterior.

Verificación y validación del modelado SfM

El análisis comparativo de la precisión del modelo SfM se llevó a cabo por medio de una comparación entre la nube densa de fotogramas, constituida por poco más de 2 millones de puntos, y la unión de las nubes TLS reducida a un número de puntos parecido al número del modelo SfM. Los resultados demuestran una congruencia satisfactoria considerando las desviaciones que, en promedio y relativas solo a las divisiones de

mampostería, eran menores de 1 cm. Las variaciones más importantes han sido resaltadas por colores que se extienden desde el azul predominante, en las zonas afectadas por la presencia de vegetación arraigada en las divisiones de mampostería (Fig. 9). Las otras partes afectadas por las desviaciones corresponden a zonas escondidas o en las sombras con respecto a los centros de adquisición TLS, como por ejemplo la parte más alta de la construcción. Las desviaciones detectadas entre las nubes TLS y SfM indican valores de congruencia global debajo del umbral denominado «tolerancia gráfica», aceptable para las imágenes renderizadas en escala 1:50. También desde el punto de vista de la eficacia comunicativa del





modelo de malla en modo video, el resultado parece ser satisfactorio, considerando sobre todo la ausencia total de intervención en el proceso de optimización. Esta mejora es imprescindible si se usa el modelo para finalidades divulgativas. Dado que el modelo resulta ser nítido y navegable de manera fluida con hardware y software específicos, el peso digital no es compatible con los requisitos de las plataformas de navegación de imágenes renderizadas en tiempo real. Después de generar las texturas y verificar la precisión del modelo, se estableció una serie de planos de cortes horizontales y verticales para definir la cadena poligonal de planos y secciones (Fig. 10). Por último, el sistema comunicativo fue completa-

do con la generación de una serie de ortofotografías relativas al plano y a las fachadas (Fig. 11–14).

Conclusiones

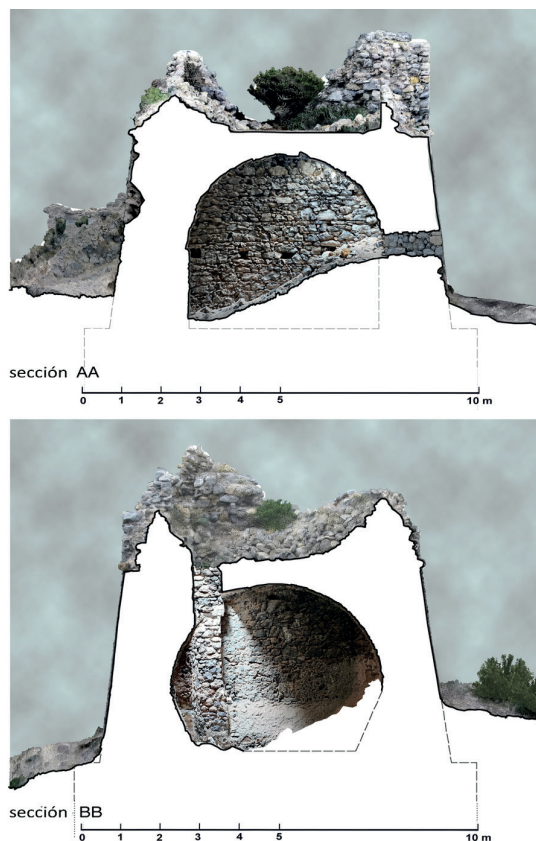
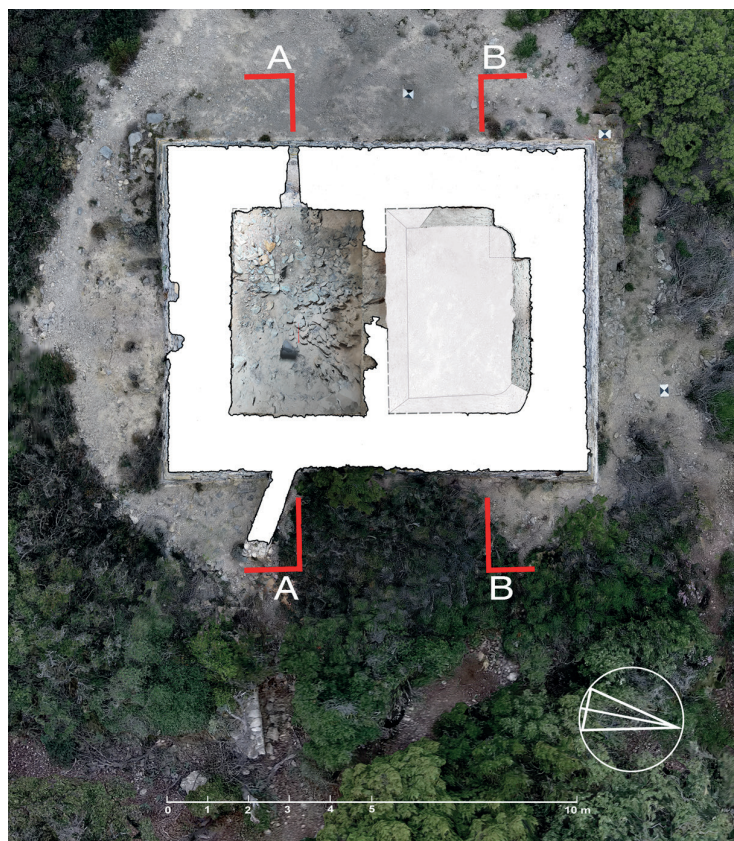
En la investigación de un elemento arquitectónico de dimensiones reducidas pero provisto de características volumétricas y de localización articuladas, los productos del modelo SfM obtenidos por medio del procesamiento de videos resultan ser satisfactorios. En este caso específico, la experimentación permitió definir un modelo automático, rápido y fiable, caracterizado por una definición adecuada, tanto para representaciones «convencionales» -perteneciente al ámbito comunicativo «tradicional»- como para modelos 3D que

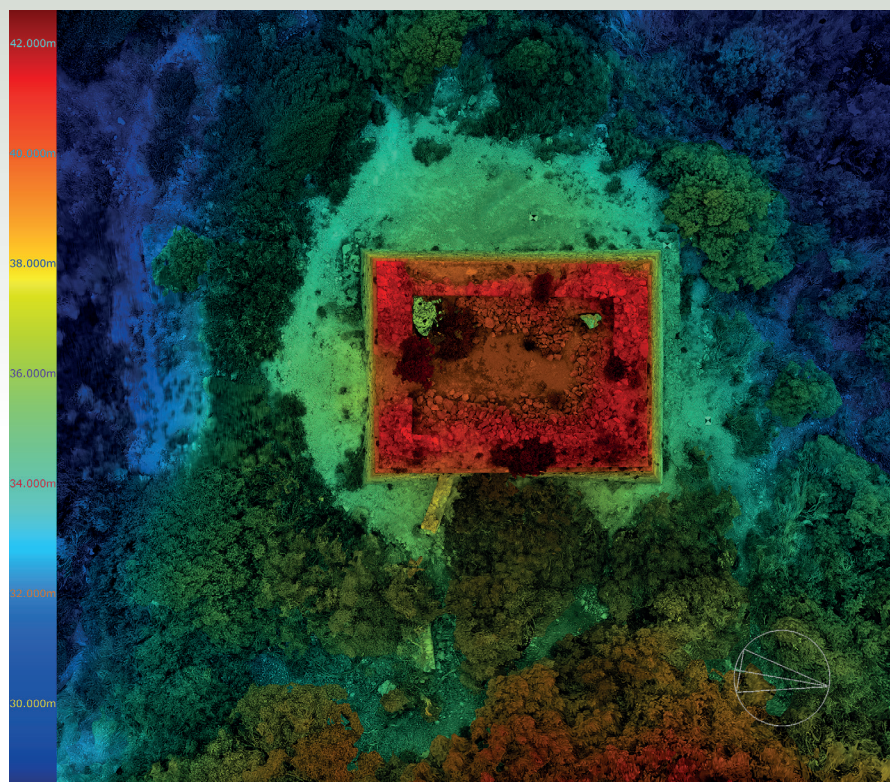
9. Mapas de la desviación entre las nubes TLS y SfM para cada fachada. En sentido horario desde la izquierda arriba: fachadas oeste, sur, este y norte.

10. Plano del basamento y secciones transversales.

9. Maps of the deviation between the different TLS and SfM clouds for each elevation, clockwise from top left: west, south, east, north.

10. Plan of the base and cross sections.





11

11. Ortofotografía horizontal de la torre y el contexto orográfico. La imagen demuestra la gama cromática de las altitudes sobre el nivel del mar.

12. Ortofotografía de las fachadas oeste (arriba) y sur (abajo).

13. Ortofotografía de las fachadas este (arriba) y norte (abajo).

11. Horizontal orthophoto of the tower and orographic context. The image highlights the chromatic range relating to the height above sea level.

12. Orthophotos of the western and southern elevations (above and below, respectively).

13. Orthophotos of the eastern and northern elevations (above and below, respectively).



12



13





pueden ser utilizados dinámicamente para finalidades divulgativas. Queda claro que los procedimientos utilizados requieren verificación, sobre todo en ámbitos más complejos y más amplios. Ciertamente, el rápido desarrollo tecnológico de los sensores pequeños que caracterizan los aparatos de consumo —en este caso UAS ligeros y videocámaras pequeñas estabilizadas— conlleva conseguir resultados siempre más eficaces y fiables a corto plazo. Aunque estos sean ágiles y eficaces, permanece la necesidad de utilizar el método SfM en combinación con métodos consolidados de levantamiento 3D, como por ejemplo las adquisiciones masivas TLS o aquellas «clásicas» como el levantamiento por medio de ET. Este requisito —aparte de las necesidades de orientación y dimensionamiento— puede ser indispensable para resolver las criticidades que dependen del mismo fundamento en el cual se base el procedimiento: la luz. ■

Referencias

- AGARWAL, S., FURUKAWA, Y., SNAVELY, N., SIMON, I., CURLESS, B., SEITZ, S.M., SZELISKI, R., 2011. Building Rome in a day. *Communications of the ACM*, October 2011, Vol. 54 N. 10, pp. 105-112. <https://doi.org/10.1145/2001269.2001293>
- DELLA MONACA, G., ROSELLI, D., TOSI, G., 1999. *Fortezze e torri costiere dell'Argentario, Giglio e Giannutri*. Pitigliano: Editrice Laurum.
- HARWIN, S., LUCIEER, A., 2012. Assessing the Accuracy of Georeferenced Points Clouds produced via multi view stereopsis from Unmanned Aerial Vehicle (UAV) imagery. *Remote Sens.* 2012, 4, pp. 1573-1599. <https://doi.org/10.3390/rs4061573>
- LICHTI, D.D., 2007. Error modelling, calibration and analysis of an AM-CW terrestrial laser scanner system. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61(5), pp. 307-324. <https://doi.org/10.1016/j.isprsjprs.2006.10.004>
- LICINIO, R., 1994. *Castelli medievali. Puglia e Basilicata: dai Normanni a Federico II e Carlo I d'Angiò*. Bari: edizioni Dedalo.
- LOWE, G. G., 1999. Object recognition from local scale-invariant features. *Proceedings of the Seventh IEEE International Conference on Computer Vision: Vol. 2*, pp. 1150-1157. <https://doi.org/10.1109/ICCV.1999.790410>
- LUISI, R., 1979. *Scudi di pietra I castelli e l'arte della guerra tra Medioevo e Rinascimento*. Roma-Bari: Editori Laterza.
- PRUCKER, U., (s.f.). [Pianta del Littorale Toscano, dalla Torre S. Vincenzo, a tutto lo Stato dei Presidj]. Pianta dello Scrittoio delle Fortezze e Fabbriche (6), Archivio di Stato di Firenze, Florencia, Italia. http://www502.regione.toscana.it/searcherlite/cartografia_storica_regionale_scheda_dettaglio.jsp?imgid=15504
- RESCH, B., LENSCH, P.A.H., WANG, O., POLLEFEYS, M., SORKINE-HORNUNG, A., 2015. Scalable Structure From Motion for Densely Sampled Videos. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3936-3944. <https://doi.org/10.1109/CVPR.2015.7299019>
- SNAVELY, N., SEITZ, S. M., SZELISKI, R., 2008. Modeling the world from Internet photo collections. *International Journal of Computer Vision*, 80, pp. 189- 210. <https://doi.org/10.1007/s11263-007-0107-3>
- TUYTELAARS, Tinne, MIKOLJCZYK, Krystian., 2008. Local Invariant Feature Detectors: a Survey. *Foundations and Trends® in Computer Graphics and Vision: Vol. 3: No. 3*, pp. 177-280. <https://doi.org/10.1561/06000000017>
- WANG, Q., CHENG, J., 2016. Development of a mixed pixel filter for improved dimension estimation using AMCW laser scanner. *ISPRS Journal of Photogrammetry and Remote Sensing*. 119, pp. 246-258. <https://doi.org/10.1016/j.isprsjprs.2016.06.004>
- LICHTI, D.D., 2007. Error modelling, calibration and analysis of an AM-CW terrestrial laser scanner system. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61(5), pp. 307-324. <https://doi.org/10.1016/j.isprsjprs.2006.10.004>
- LICINIO, R., 1994. *Castelli medievali. Puglia e Basilicata: dai Normanni a Federico II e Carlo I d'Angiò*. Bari: edizioni Dedalo.
- LOWE, G. G., 1999. Object recognition from local scale-invariant features. *Proceedings of the Seventh IEEE International Conference on Computer Vision: Vol. 2*, pp. 1150-1157. <https://doi.org/10.1109/ICCV.1999.790410>
- LUISI, R., 1979. *Scudi di pietra I castelli e l'arte della guerra tra Medioevo e Rinascimento*. Roma-Bari: Editori Laterza.
- PRUCKER, U., (s.f.). [Pianta del Littorale Toscano, dalla Torre S. Vincenzo, a tutto lo Stato dei Presidj]. Pianta dello Scrittoio delle Fortezze e Fabbriche (6), Archivio di Stato di Firenze, Florencia, Italia. http://www502.regione.toscana.it/searcherlite/cartografia_storica_regionale_scheda_dettaglio.jsp?imgid=15504
- RESCH, B., LENSCH, P.A.H., WANG, O., POLLEFEYS, M., SORKINE-HORNUNG, A., 2015. Scalable Structure From Motion for Densely Sampled Videos. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3936-3944. <https://doi.org/10.1109/CVPR.2015.7299019>
- SNAVELY, N., SEITZ, S. M., SZELISKI, R., 2008. Modeling the world from Internet photo collections. *International Journal of Computer Vision*, 80, pp. 189- 210. <https://doi.org/10.1007/s11263-007-0107-3>
- TUYTELAARS, Tinne, MIKOLJCZYK, Krystian., 2008. Local Invariant Feature Detectors: a Survey. *Foundations and Trends® in Computer Graphics and Vision: Vol. 3: No. 3*, pp. 177-280. <https://doi.org/10.1561/06000000017>
- WANG, Q., CHENG, J., 2016. Development of a mixed pixel filter for improved dimension estimation using AMCW laser scanner. *ISPRS Journal of Photogrammetry and Remote Sensing*. 119, pp. 246-258. <https://doi.org/10.1016/j.isprsjprs.2016.06.004>