

Document downloaded from:

<http://hdl.handle.net/10251/120340>

This paper must be cited as:

Barbero-García, I.; Cabrelles, M.; Lerma, J.L.; Marqués-Mateu, À. (2018). Smartphone-based close-range photogrammetric assessment of spherical objects. *The Photogrammetric Record*. 33(162):283-299. <https://doi.org/10.1111/phor.12243>



The final publication is available at

<http://doi.org/10.1111/phor.12243>

Copyright Blackwell Publishing

Additional Information

SMARTPHONE-BASED CLOSE-RANGE PHOTOGRAMMETRIC ASSESSMENT OF SPHERICAL OBJECTS

INÉS BARBERO-GARCÍA* (inbargar@topo.upv.es)

MIRIAM CABRELLES (micablo@upvnet.upv.es)

JOSÉ LUIS LERMA (jllerma@cgf.upv.es)

ÁNGEL MARQUÉS-MATEU (amarques@cgf.upv.es)

Universitat Politècnica de València, Spain

* Corresponding author

Abstract

Smartphones have widened the possibilities for low-cost close-range image acquisition for three-dimensional (3D) modelling. They allow the rapid acquisition of large amounts of data for a wide range of applications. However, the accuracy of the models and the automation possibilities depend on the image acquisition conditions and application requirements. In this study, the accuracy and reliability of the derived photogrammetric 3D models are evaluated on a spherical setup for close-range applications (ca. 30 cm). Different numbers of images, network configurations, targets, devices and camera calibration methodologies are tested and evaluated. Results show that for this close-range application high accuracy (0.2 mm) and reliability can be achieved. The number of images did not significantly affect the accuracy but was vital for tie-point detection and image orientation. The use of artificial targets was found to be the key factor in increasing the final accuracy. In contrast, the image calibration strategy and the characteristics of the imaging device did not have a great impact on the results.

KEYWORDS: smartphone, network geometry, calibration, accuracy, reliability.

INTRODUCTION

THE PERFORMANCE OF SMARTPHONE CAMERAS has greatly increased over the past few years. Although their quality is still a limitation for obtaining of highly accurate images, they have the great advantage of being totally portable and are almost always to hand. Currently, they offer a low-cost option for close-range photogrammetric applications, including the creation of 3D models for a wide range of purposes, including structural monitoring (Wang et al., 2012), geomorphology (Micheletti et al., 2015), creative industries (Nocerino et al., 2017) and medical applications (Abreu de Souza, et al. 2012; Hellwich et al., 2016; Hernandez & Lemaire, 2016), among others. Other low-cost non-metric cameras,

such as webcams, have also been used for close-range modelling (Chong and Brownstein, 2010).

Smartphone video (more specifically slow-motion video) is a useful tool for acquiring large numbers of images, suitable even for fast moving objects. These images can be used for the creation of 3D models of moving objects (Barbero-García et al., 2017). With an image acquisition speed of 240 frames per second (fps) of many smartphones (still far beyond the ultra-high-speed cameras that reach up to 2000 fps), the computational cost is the main limitation given the number of images to be used for 3D modelling.

Despite their advantages, smartphone cameras present high internal instability that hampers their correct calibration. This problem is common to all non-metric digital cameras (Fraser, 2013), but is especially exacerbated when working with smartphones. The radiometric accuracy of smartphone cameras is lower than that of single-lens reflex (SLR) cameras but, despite their limitations, studies have concluded that these cameras can be used for photogrammetric tasks with a required accuracy of 1:10 000 (Akca and Gruen, 2009).

The development of useful tools, which could allow non-expert users to obtain accurate 3D models for different purposes, requires a high degree of automation (Remondino et al., 2014). However, most of the available automatic low-cost solutions provide low repeatability and reliability (Remondino et al., 2012). The development of fully automatic and reliable solutions for specific applications requires an extensive knowledge of the factors affecting the quality of the 3D models created using smartphones or other similar imaging devices, such as tablets. The most important factors include the determination of the ideal geometric network, the selection of the best video frames and their optimal number, as well as the accuracy requirements for camera calibration. These parameters can vary greatly depending on the characteristics and limitations of the image acquisition process for a specific application (such as moving or static target feature, camera-to-object distance, without forgetting the lighting conditions) and the required accuracy of the final 3D model.

The traditional working pipeline in photogrammetry includes a low number of images and the manual identification of tie points. In contrast, the automatic processes that are common nowadays require a large number of images and, therefore, short baselines. The high overlap allows the use of feature-detection algorithms such as the scale-invariant feature transform (SIFT; Lowe, 1999) and speeded-up robust features (SURF; Bay et al., 2007), so that tie points are detected automatically. The high speed of image acquisition also points to hyper-redundancy as a way to improve accuracy and compensate for the lack of internal stability. Some authors have studied the influence of hyper-redundancy and suggest using it as a tool to improve accuracy when the additional work to extract a large number of images is not an issue (Fraser et al., 2005). Current technologies, such as video, provide huge quantities of data. Therefore, the accurate determination of the number of images required for modelling, and the consequent filtering out of redundant images, has also become a necessity (Alsadik, et al., 2015).

Another common subject of study in photogrammetry is the possibility of accurate self-calibration, especially for digital cameras with poor internal stability. Although self-calibration, conducted simultaneously with the 3D-modelling process, is considered a powerful tool, a separate calibration process using a setup that ensures good image geometry is still recommended in many cases. A primary reason behind this is that the

optimal network for 3D reconstruction is usually not the best for camera calibration (Remondino and Fraser, 2006).

Lastly, it is important to take into account the variety of capabilities and characteristics of different smartphone cameras. Resolution and maximum frame rate vary greatly between devices, providing very different results.

This study assesses the different settings that can affect the creation of accurate smartphone 3D models for spherical objects in close-range applications. Spherical objects were selected as the final goal is to find the optimal setup for a specific application, namely the creation of 3D models for cranial deformation analysis in infants. In this application, the object to be modelled (the infant's head shape) is usually in movement and image acquisition is carried out by a doctor in a limited time frame (Barbero-García et al., 2017). However, as the real photogrammetric data-acquisition conditions are hard to assess and replicate, this study deals with a sphere whose shape emulates an infant's head. Therefore, it is possible to take as many shots as necessary in order to replicate the different conditions that can be encountered in real-life projects, usually in either a hospital or a clinic.

A wide range of conditions were tested to determine the ideal network and setup. Due to the type of application, the setup was a simplification of real clinical conditions. Different numbers of image sets (19 to 95) were used to generate different models in order to evaluate the effect of hyper-redundancy. In addition, the influence of both (separate) camera calibration and (integral) self-calibration were tested, as both were performed for each of the network designs. The use of well-defined automatic coded markers as a tool to improve reliability and accuracy (Fraser, 1997; Luhmann et al., 2016) was also evaluated. Two different smartphones, with different camera characteristics, were tested. Reference data was obtained using a calibrated high-end SLR camera.

The coordinates of a set of control points were calculated for each of the created models. The distances between each pair of coordinates were calculated too; the differences from the reference data were then obtained. The completeness of the models was also evaluated.

The results of the study will be useful for the automation of the process and the creation of clear guidelines that will allow users (namely doctors) to successfully carry out the image acquisition.

METHODOLOGY

The evaluation of different parameters for 3D modelling, including network design, number of images, use of targets and calibration methodology, will drive the selection of the best methodology for 3D modelling close-range spherical objects, at an approximate camera-to-object distance of 30 cm with low imaging texture. Two smartphones were tested to assess the methodology: a Samsung Galaxy S7 Edge and a Samsung Galaxy Trend. The former can be considered a high-end smartphone; the latter a more conventional and cheaper smartphone. A high-end SLR camera, a Canon EOS-1Ds Mark III, was used to determine highly accurate coordinates of the discrete targeted points after self-calibration. These targeted points acted as control points and were used as reference data for 3D assessment.

The first step of the photogrammetric data processing started with the geometric calibration of each camera, to determine both the interior and the exterior orientation parameters, as well as the additional parameters. The setup consisted of: (1) one horizontal

and two vertical panels (in an L-shaped configuration), all of them with several targets already included on them, as well as additional coded targets); (2) the sphere to be modelled; and (3) a calibrated ruler. The data was processed using the in-house photogrammetric software FOTOGIFLE (Lerma et al., 2010, 2014), which allows users to determine the quality of all the estimates (interior and exterior orientation parameters, and object coordinates).

After the geometric calibration, one video of the object to be modelled was taken with each smartphone. Next, the Agisoft PhotoScan software was used to create several models for each video, varying the number of images, the presence of markers (targets) and the geometric calibration of the cameras (Fig. 1).

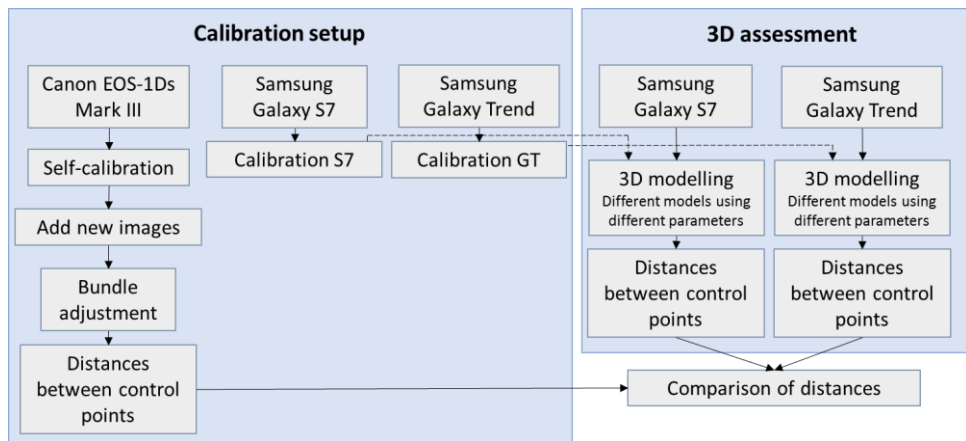


FIG. 1. Methodology workflow for calibration and 3D assessment.

Calibration

The camera calibration was carried out using a setup of panels with targets, with the object to be modelled placed inside the framework defined by the panels (Fig. 2). Although the inclusion of the object was not necessary for calibration purposes, it simplified the processing as the images for the self-calibration of the SLR camera were also used to create the reference 3D model.

For the Canon SLR camera calibration, a set of images were taken using a tripod and standard room illumination. The self-calibration setup included a set of convergent images, some of them rotated by 90°, good intersection angles of rays from the object points to the camera positions, and a sufficient number of targets well spread across the image format. This geometry was chosen to assure a good calibration, as stated by Fryer (2001) and Remondino and Fraser (2006). Finally, the images were calibrated using FOTOGIFLE with up to 10 additional parameters as proposed by Fraser (1997). A total of 89 circular and coded targets were used for the calibration, achieving an average calibration error of 0.26 pixels and a maximum error of 1 pixel. The resulting targeted points coordinates were set as fixed coordinates to calibrate both smartphones afterwards.

To calibrate the smartphone cameras, a video was recorded with each device. The conditions (distance, resolution, lighting...) used for this video calibration were the same

as used later for recording and carrying out the subsequent 3D modelling. A set of 11 frames from each video were manually selected to satisfy the best possible geometry. Most of the required targets were manually identified in each image set, due to the lack of automatic recognition by PhotoScan.

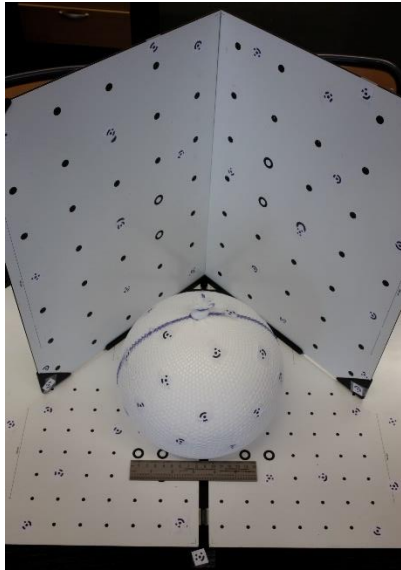


FIG. 2. Setup used for camera calibration that includes three orthogonal calibration panels, the sphere to be modelled and a calibrated ruler. Additional coded targets were incorporated to strengthen the geometry.

Three-dimensional Model Setup

A sphere 18 cm in diameter was used for the tests. This object tries to emulate, in a simplified form, an infant's head modelled for cranial deformation analysis. The clinical conditions were also replicated and thus neither special lighting nor a tripod were used. To further imitate clinical conditions and facilitate the 3D modelling by adding some texture, a fitted cap, similar to those used on the patients, was placed on the sphere. A total of 24 coded targets, with an approximate diameter of 0.5 cm, were placed on the panels and sphere. These coded targets were recognised automatically by the PhotoScan software as markers; they were also be used as control points for accuracy calculations. A calibrated ruler was placed next to the sphere to allow the scaling of the reference dataset (Fig. 3).

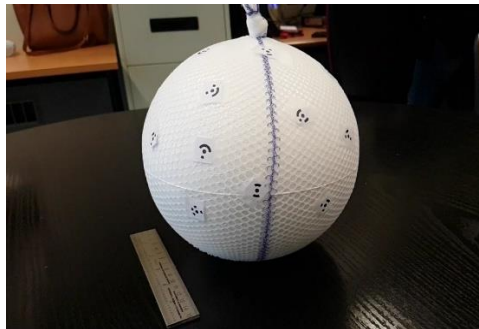


FIG. 3. Three-dimensional model setup.

Reference Data

A high-accuracy photogrammetric solution was chosen as the best methodology to obtain accurate coordinates of the targets. The Canon EOS-1Ds Mark III SLR was used with a pixel count of 5616 x 3744, a focal length of 35 mm and ISO speed of 100 (Table I).

A total of 40 images were taken covering the sphere and targets around it to assure the best possible accuracy. A tripod was employed to achieve neat pictures. The setup was the same as used for calibration purposes, and the images used for the calibration were also included. 29 further images were added, covering the back side of the sphere after removing the vertical panels used for calibration.

The coordinated control points (coded targets) were computed with the software FOTOGIFLE, using a bundle adjustment and the camera calibration parameters obtained in a previous step during the camera calibration (Fig. 1). Sub-millimetre precision was obtained for every control point.

Image Acquisition

A video was recorded with each smartphone. The maximum frame rate for each device was chosen (Table I). A slow-motion video with a 1280x720 pixel count and 239 fps was obtained using the Samsung Galaxy S7 smartphone. No slow-motion functionality was available on the Samsung Galaxy Trend Plus. Therefore, a video with a similar pixel count and 30 fps was obtained. The length of both videos was under 1 minute. The specifications of the cameras can be consulted in Table I, although not all the parameters are provided by the manufacturer for the Samsung Galaxy Trend.

TABLE I. Camera parameters. The mode actually used is highlighted in **bold**.

<i>Camera</i>	<i>Focal length (mm)</i>	<i>Maximum frame rate (fps)</i>	<i>Effective format size (mm)</i>	<i>Pixel count (pixels)</i>	<i>Pixel pitch (mm)</i>
Canon EOS-1Ds Mark III	35		35.94 x 23.96	Still: 5616 x 3744	0.0064
Samsung Galaxy S7	4.17	239	5.37 x 3.02	Still: 3840 x 2160 Video: 1280 x 720	0.0014 0.0042
Samsung Galaxy Trend Plus	Unknown	30	Unknown	Still: 2048 x 1536 Video: 1280 x 738	Unknown

Network Design

The determination of the optimal network design is especially important for the application of the methodology in real working environments. Under clinical conditions, simplicity and speed of data acquisition are vital to present a realistic and useful methodology.

Two different camera configurations were tested:

- (1) The first network consisted of a single ring of nearly horizontal convergent images taken around the object (Fig. 4(a)). Two models were created: (i) using 26 images; and (ii) using 52 images.
- (2) The second network consisted of two sets of rings taken at different heights around the object, together with a third ring of zenithal (nadir) images (Fig. 4(b)). Four models were tested: (i) 19 images (9 at low height + 9 at medium height + 1 at the top); (ii) 26 images (16+9+1); (iii) 52 images (30+20+2); and (iv) 95 images (55+25+15).

The model with 19 images (2(i) above) was considered the generic network design, similar to the one presented by Kraus (1997) for similar shapes, as it is the one where the optimal multi-ray intersection is obtained with as few images as possible. Therefore, the other networks should be considered redundant. The set of images for each model was manually selected to obtain the optimal geometry.

In order to evaluate the effect of the calibration, every model was calculated three times using different geometric camera calibration methodologies:

- (1) Self-calibration from PhotoScan.
- (2) Fixed interior (inner) calibration from FOTOGIFLE on an optimal network, and image orientation in PhotoScan.
- (3) Self-calibration and image orientation in PhotoScan using interior calibration parameters from FOTOGIFLE as pre-calibration data.

For methodology (3), the calibration parameters obtained using FOTOGIFLE were imported into PhotoScan in Australis format (Fraser and Edmundson, 2000) and then transformed by PhotoScan into its own format. This process was necessary as the calibration parameters are defined differently in each software (Drap and Lefèvre, 2016).

Although different workflows were followed to determine the interior orientation parameters, the external orientation was always carried out using PhotoScan. To evaluate the effect of the presence of coded targets, every model was calculated both with and without them. The whole process consisted, therefore, of a total of 36 3D models (6

networks; 3 calibrations; 2 target types) for each smartphone. Every model was scaled using five ground control points (GCPs) whose coordinates were extracted from the reference data.

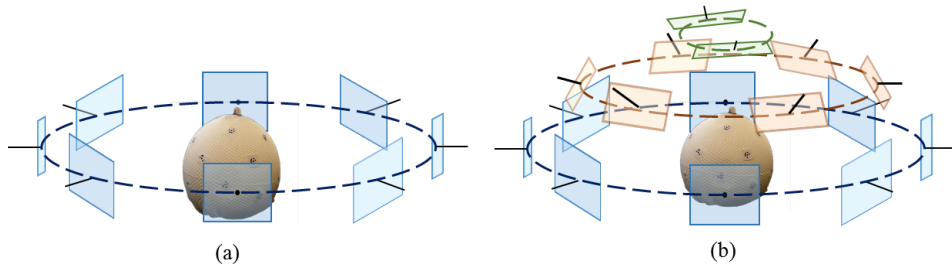


FIG. 4. Camera network geometry. (a) Single ring of nearly horizontal images at a constant height around the object. (b) Three rings of images at different altitudes and employing horizontal, oblique and zenithal (nadir) attitudes.

Model Creation

In the first place, the images were masked to exclude the background. When using images of moving objects in live environments, the background changes its position in relation to the object. Therefore, the background must be excluded to allow the creation of the 3D model. In this particular case, using the background would allow the identification of points outside the object and the results would not represent the true condition that would be achieved on site. After masking, some targets were manually identified, as PhotoScan was unable to detect all of them automatically. Lastly, the model was created using the conventional software pipeline: (1) image orientation; (2) dense matching; and (3) meshing. Finally, every model was scaled using the five GCPs whose coordinates were provided by the reference model. An additional step was also undertaken: model texturing for visualisation purposes.

Assessment

The assessment was carried out considering two parameters: (1) the accuracy of the targets coordinates; and (2) the completeness of the model. The accuracy of the coordinates was evaluated by comparing the distances between targets. For each model, a total of 171 distances between all possible pairs of targets were calculated. Later, the distances were compared to those of the reference dataset. This process, as presented by Luhmann and Wendt (2000), does not need registration of the models and, consequently, registration errors are avoided.

The completeness of the models was checked visually. The models were classified as complete when the whole sphere was correctly represented, and as not complete when holes were apparent in a model (Fig. 5). In addition, a further category was considered for incomplete 3D models that presented only small imperfections on the top; these are termed ‘bare’ models. This latter case is due to the lack of zenithal images during the single ring data acquisition (Fig. 4(a)).

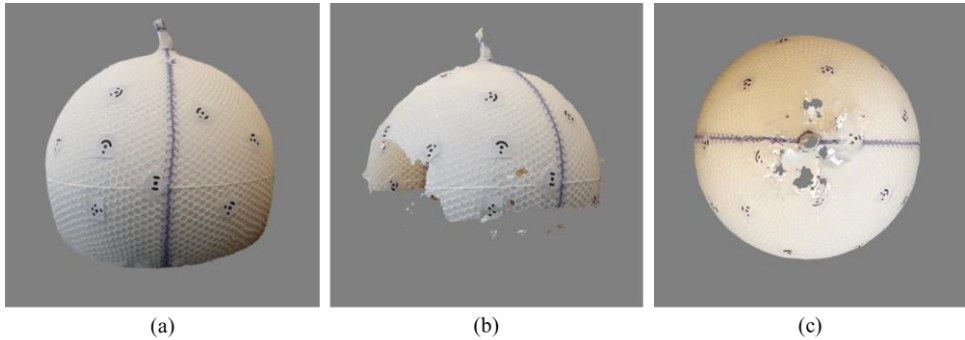


FIG. 5. Examples of: (a) a complete three-dimensional model; (b), incomplete model; and (c) a model with small imperfections (bare).

RESULTS

Calibration

Calibration results are presented in Table II. For both smartphone cameras, in addition to the elementary interior orientation parameters (x_0 , y_0 , f), two radial lens distortion parameters were requested (k_1 , k_2), as well as the first decentring lens distortion parameter (p_1) and the in-plane differential scaling between the horizontal and the vertical pixel spacing (b_1). Every parameter was checked for statistical determinability and those below 99% were removed. In total, the Samsung Galaxy S7 required seven additional parameters while Samsung Galaxy Trend Plus required nine parameters (two more than the other smartphone: p_2 – the second decentring lens distortion parameter; and b_2 – non-orthogonality between the x and y axes). The mean of the residuals of the bundle adjustment obtained using FOTOGIFLE for the Samsung Galaxy Trend smartphone was below 1.39 pixels, with an average of 0.36 pixels. For the Samsung Galaxy S7 smartphone calibration, the residuals were 1 pixel higher, with an average of 1.36 pixels. These residuals could be reduced only by performing an independent calibration for each photograph. However, this option was discarded as independent calibration parameters, obtained using an ideal calibration network, would not be applicable for the rear-side 3D modelling. Worth noting is that the standard deviation of the additional calibration parameters with the Samsung Galaxy Trend was always better than on the Galaxy S7 counterpart. The average intersection angle was 41.9° for the Samsung Galaxy S7 and 26.3° for the Samsung Galaxy Trend. The calibration distortion patterns (Fig. 6) do not show large differences between the two devices.

TABLE II. Additional calibration parameters and standard deviation (σ) obtained using FOTOGIFLE (f , x_0 , y_0 and x , y errors in pixels).

	<i>Galaxy S7</i>		<i>Trend Plus</i>	
	<i>Mean</i>	σ	<i>Mean</i>	σ
x_0	16.68	0.47	15.65	0.24
y_0	-24.16	0.35	32.61	0.27

f	-1029.28	0.44	-1292.32	0.22
k_1	-1.30^{07}	2.09^{09}	-1.38^{07}	6.97^{10}
k_2	2.50^{-13}	3.70^{-15}	2.20^{-13}	1.36^{-15}
k_3	0	0	0	0
p_1	7.56^{-07}	1.60^{-07}	1.06^{-06}	4.92^{-08}
p_2	0	0	-9.48^{-07}	6.51^{-08}
b_1	0.0235	1.40^{-04}	0.0255	4.71^{-05}
b_2	0	0	0.0017	4.52^{-05}
Mean error in x		1.36		0.34
Mean error in y		1.13		0.36

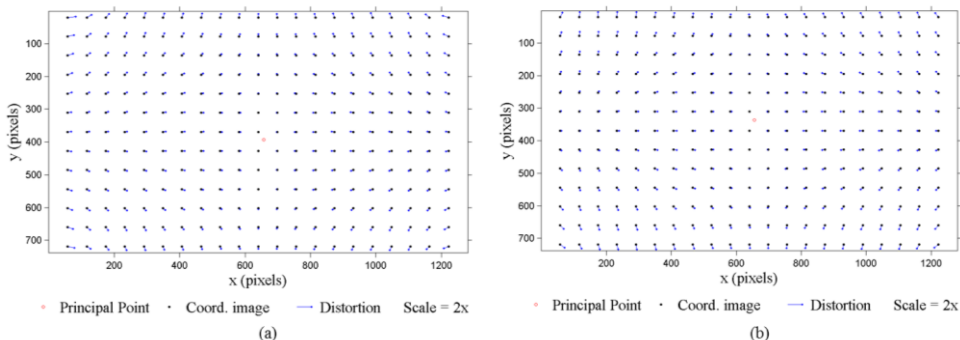


FIG. 6 Calibration distortion patterns for (a) the Samsung Galaxy S7 and (b) the Samsung Galaxy Trend.

Three-dimensional Models

A total of 36 models were created for each device, so 72 models in total. The coordinates of the coded targets in each model were obtained by combining automatic and manual procedures whenever the fully-automatic solution did not work. Later, the distances between every pair of markers were calculated and compared to those of the reference dataset derived from the high-end SLR digital camera. The mean difference in distance and the 65th percentile are shown in Table III. This table also specifies if the model was complete and acceptable (in green); incomplete (in yellow); or the additional ‘bare’ category (in blue) for those models that were almost correct but presented small imperfections, mainly in the upper area of the object. This bare category was necessary as some network designs, that achieved high accuracies, failed to produce useful 3D models without any additional processing. Therefore, this bare category highlighted those 3D models demanding additional photogrammetric and/or editing processing. In addition, 3D models that were not achieved after the automatic photogrammetric workflow are highlighted in orange.

TABLE III. Accuracy of the models. Mean and 65th percentile (P_{65}) of the differences in distance. Incomplete models shown in yellow. Bare models with small imperfections in blue. Complete models in green. Models not achieved after the automatic photogrammetric workflow in orange.

			Single ring models				Three ring (horizontal + oblique + nadir) models							
			26 images		52 images		19 images (9+9+1)		26 images (16+9+1)		52 images (30+20+2)		95 images (55+25+15)	
	Interior Calibration	Coded targets	Mean	P_{65}	Mean	P_{65}	Mean	P_{65}	Mean	P_{65}	Mean	P_{65}	Mean	P_{65}
Samsung Galaxy S7	PhotoScan	Yes	0.6	0.6	0.5	0.8	0.3	0.4	0.3	0.3	0.2	0.2	0.2	0.3
		No	0.9	1	0.8	0.9	Not aligned	Not aligned	0.9	1	0.8	0.9		
	FOTOGIFLE	Yes	0.4	0.4	0.4	0.4	0.4	0.5	0.4	0.5	0.4	0.4	0.4	0.5
		No	0.6	0.8	0.4	0.5	Not aligned	Not aligned	0.8	0.9	0.8	1.1		
	FOTOGIFLE + PhotoScan	Yes	0.1	0.2	0.4	0.4	0.2	0.2	0.1	0.2	0.1	0.2	0.2	0.2
		No	0.5	0.6	0.5	0.6	Not aligned	Not aligned	0.6	0.7	0.5	0.6		
Samsung Galaxy Trend	PhotoScan	Yes	0.5	0.6	0.7	0.8	0.2	0.3	0.2	0.2	0.2	0.2	0.2	0.2
		No	Not aligned	0.7	0.8	Not aligned	Not aligned	1.4	1.6	0.7	0.8			
	FOTOGIFLE	Yes	0.6	0.7	0.7	0.9	0.7	0.8	0.7	0.8	0.6	0.7	0.7	0.8
		No	Not aligned	2.3	2.6	Not aligned	Not aligned	Not aligned	1.2	1.4				
	FOTOGIFLE + PhotoScan	Yes	0.2	0.2	0.6	0.7	0.1	0.1	0.1	0.1	0.2	0.3	0.2	0.2
		No	Not aligned	1.4	1.8	Not aligned	Not aligned	Not aligned	2.9	3.6				

For the Samsung Galaxy S7 and the simple network design with a single ring of nearly horizontal images, models were nearly complete but presented small imperfections, mostly in the upper part which was not correctly covered by the images. However, the calculated distance error was acceptable, being below 1 mm for 65% of the points in every case. The results were almost independent of the type of calibration and the presence of targets. For the second network design with three rings of imagery, no models were obtained with 19 images and some problems appeared in areas without targets using 26 and 52 images. In the last case which used 95 images, all models were complete; the error was below 1 mm in every case and under 0.5 mm whenever targets were used. Despite the differences registered in the completeness of the various models, no significant improvement in the accuracy of the models was achieved by increasing the number of images.

For the Samsung Galaxy Trend smartphone, with a significantly less powerful camera, no complete models could be obtained with fewer than 95 images. On the one hand, the best result was obtained in the 95-image network, using targets and with no fixed independent calibration, yielding an average error of 0.2 mm. On the other hand, the error increases without targets, the lowest mean being 0.7 mm.

The influence of the individual parameters is presented in Fig. 6. The usage of coded targets is the most important (key) factor to improve accuracy (Fig. 6(a)). The models created without targets have a maximum error of 1.4 mm while the models with targets have an error below 0.7 mm in every case. Targets also helped to increase the reliability. The self-calibration and the fixed FOTOGIFLE calibration provided similar accuracy (Fig. 6(c)), while the non-fixed calibration, using first the calibration from FOTOGIFLE and

afterwards slightly improvements from PhotoScan, provided a significant improvement. This type of calibration also improved the reliability slightly. The two different devices showed very different reliability values. The Samsung Galaxy S7 provided correct models in 40% of the cases while only 10% of the models were correct when using the Samsung Galaxy Trend. The accuracy was also significantly better on the S7 device. The two network geometries used presented very different results for both accuracy and reliability. The one-ring network geometry provided slightly smaller errors in coordinates. Nevertheless, the reliability of the results was much better for the three-ring geometry.

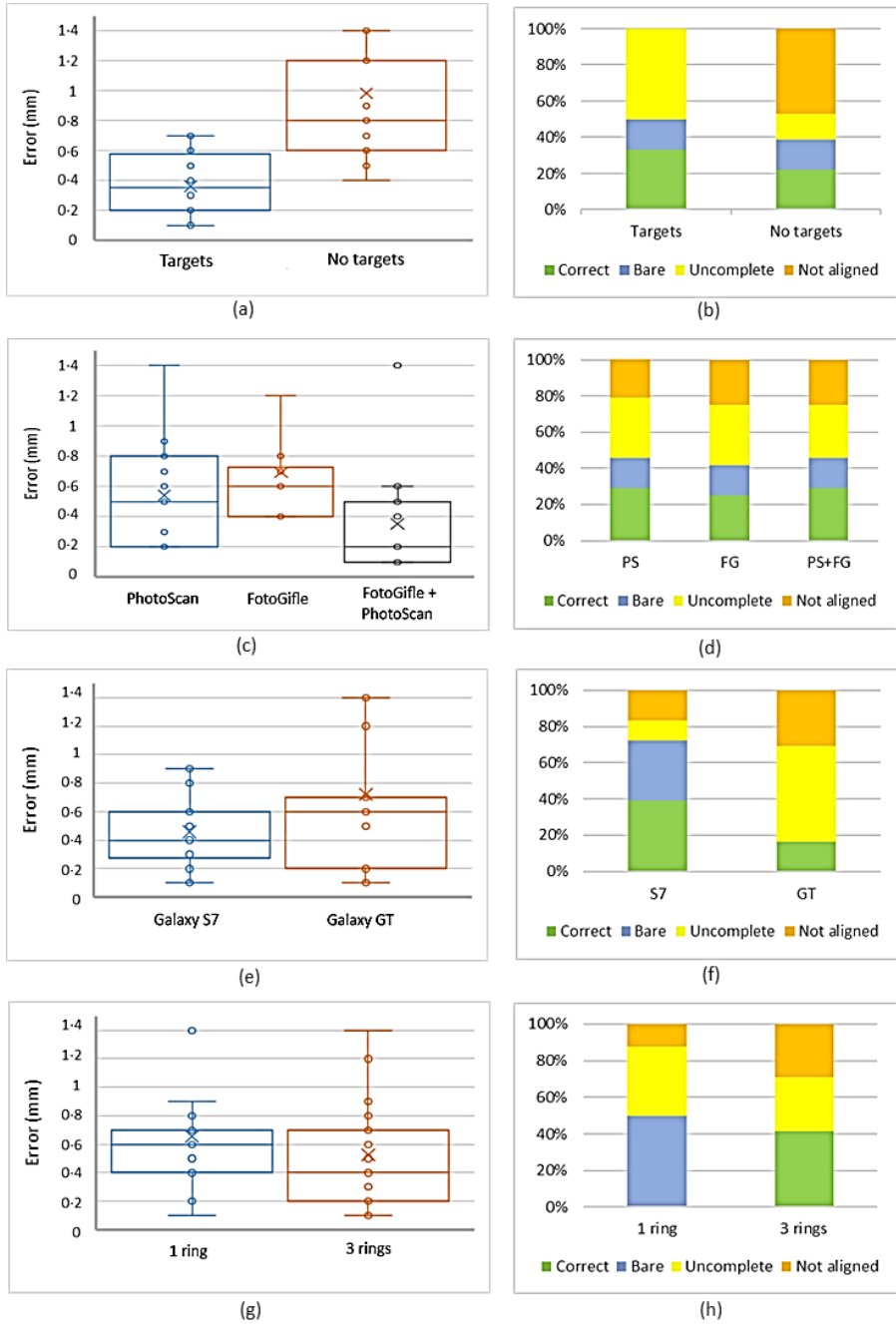


FIG. 6. Box-and-whisker plots showing the accuracy of the models (using the mean error for each one) grouped by: (a) the presence/absence of targets; (c) calibration approach; (e) device; and (e) camera network geometry. Reliability of the models for the same set of parameters (b, d, f, h).

Due to the high influence of the coded targets in the final 3D modelling assessment, it is necessary to analyse the results of the different networks and the number of images, with and without targets, to obtain meaningful results (Fig. 7). The accuracy was better than 1 mm for every network geometry and number of images whenever targets were included. However, the reliability was really affected by both the geometry and the number of images; only the models with 95 images were correct in every case.

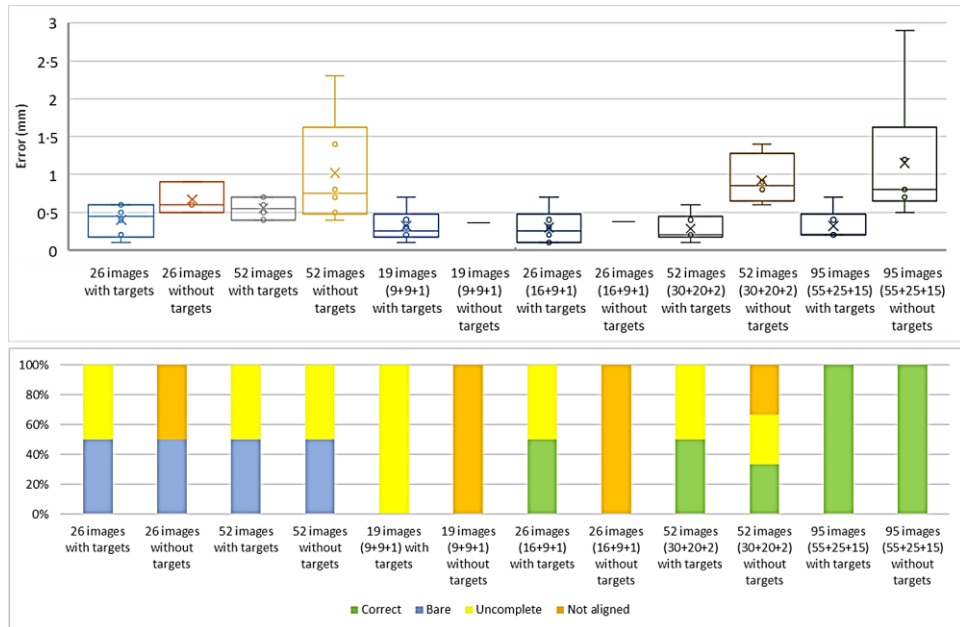


FIG. 7. Accuracy (top) and reliability (bottom) for the different network geometries based on the number of images, with and without targets.

DISCUSSION

Smartphone cameras in slow-motion video mode have proven to be a useful tool for the quickly obtaining images that can be used for various close-range photogrammetric applications. This study has evaluated the possibilities of two smartphone cameras for close range photogrammetry (at approximately a 30 cm camera-to-object distance) and 3D modelling of a small spherical object with low image texture. In particular, the effects of calibration, hyper-redundancy and the presence of well-defined targets have been evaluated. This has been achieved by keeping in mind a particular application: the modelling of infants' heads for cranial deformation analysis. Nevertheless, this close-range photogrammetric experience can be extrapolated to various other scenarios that require the use of low-resolution sensors imaging small and round objects, for instance industrial

applications modelling a series of moving spherical objects, or the documentation of small sculptures in cultural heritage with video recording using smartphones.

The number of images was found to be vital to obtain complete models of acceptable accuracy. A high hyper-redundancy photogrammetric approach provided good results in almost every situation. For a low number of images, especially for the second network type with three rings and using the Samsung Galaxy Trend device, insufficient tie points were obtained and image orientation was not possible. Probably an increase in the object's texture would allow completeness of the models with fewer images, as the tie point detection would be easier and more successful. The use of coded targets for the image-orientation step allowed partial models to be obtained in many cases, even when no image orientation was possible without them. For well-defined imaging networks of the three-ring configuration, the targets improved the accuracy significantly. In a particular case, the mean distance changed from 2.9 mm (unacceptable) to 0.2 mm, which turned out to be the highest accuracy of all models. The calibration of the cameras was found to be the least important aspect of all the parameters evaluated. The calibration methodology (self-calibration with PhotoScan, camera calibration with FOTOGIFLE, and a combination of both) did not improve the final accuracy of the 3D models, even considering both a large number of images and a strong geometric network. One single ring of images was not enough to recover accurately the internal orientation parameters of the camera, even with targets. Therefore, the self-calibration approach can be considered a suitable method for slow-motion smartphone video-image acquisition for close-range photogrammetric applications, whenever configurations of the three-ring type are considered.

Different results were obtained for the two devices. The Samsung Galaxy S7, equipped with a more powerful camera, provided better accuracy and much higher reliability than the Samsung Galaxy Trend. However, both devices achieved good results in optimal conditions (high redundancy and coded targets). Therefore, it can be stated that acceptable reliability and accuracy can be obtained using most smartphones in the market, as long as the network conditions are adequate.

The network configuration based on the single ring with nearly horizontal images (Fig. 4(a)) is only recommended when the number of images to be acquired needs to be low and no well-defined targets can be used. For these cases, this geometry improved the chance of correct image orientation, but small imperfections in the upper part of the model should be expected. These results can be explained since the images cover mainly the medium-height part of the object, resulting in a high overlap in this area. Because of this fact, the feature detection algorithms performed easier and better, allowing the detection of a higher number of tie points.

The hyper-redundancy of the network allows users to achieve complete models even with low-resolution cameras and low-texture objects. However, the larger number of images does not significantly affect the accuracy in the coordinates of targets. Results suggest that the best methodology for this application is the use of a large number of images (95 or more) distributed at different heights around the object. The use of coded targets should be considered, even if it requires extra development to automate the methodology. Under these conditions, it is possible to obtain an error below 0.2 mm, which can be considered totally

acceptable as the most common methodologies currently in use have a precision worse than 1 mm (Schaaf et al., 2010).

CONCLUSIONS

This paper has assessed the factors affecting the 3D modelling process in order to develop fully automatic and accurate solutions for specific close-range applications. The usage of smartphones for close range photogrammetry can provide sub-millimetre accuracy whenever slow-motion video is selected and camera-to-object distances of up to 30 cm are considered.

The number of images required to obtain a fully automatic image orientation is much higher than that needed to obtain high accuracy in classic photogrammetric networks with a small number of manually measured tie points and highly convergent shots. The inclusion of coded targets (whose measurement and matching can be partially automated) has proven to increase significantly the accuracy of the final 3D model. In addition, coded targets facilitate enormously the image orientation, and this is independent of whether a low-end or high-end smartphone camera is used. Three rings of images with coded targets are considered optimal for achieving maximum accuracy with a self-calibration approach, (at least for slow-motion smartphone video image acquisition of a spherical object). Last but not least, highly redundant overlapping images are required to achieve correct reliability in the 3D modelling.

Further evaluations will be carried out in the future using different cameras, such as stable high-speed cameras, that might yield better calibration results and even better metric deliverables.

ACKNOWLEDGEMENTS

This study was partially supported by grant number ACIF/2017/056 from the Conselleria d'Educació of the Generalitat Valenciana and the European Social Fund. The authors also acknowledge the support from the 2017 Subprogram B No. B03 (HEAD3D++) to the development of coordinated actions between the Universitat Politècnica de València and the Hospital Universitari i Politécnic/Instituto de Investigación Sanitaria La Fe.

REFERENCES

- ABREU DE SOUZA, M., ROBSON, S. and HEBDEN, J. C., 2012. A photogrammetric technique for acquiring accurate head surfaces of newborn infants for optical tomography under clinical conditions. *Photogrammetric Record*, 27(139): 253–271.
- AKCA, D. and GRUEN, A., 2009. Comparative geometric and radiometric evaluation of mobile phone and still video cameras. *Photogrammetric Record*, 24(127): 217–245.
- ALSADIK, B., GERKE, M. and VOSSELMAN, G., 2015. Efficient use of video for 3D modelling of cultural heritage objects. *International Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2(3/W4): 1–8.
- BARBERO-GARCÍA, I., LERMA, J. L., MARQUÉS-MATEU, Á. and MIRANDA, P., 2017. Low-cost smartphone-based photogrammetry for the analysis of cranial deformation in infants. *World Neurosurgery*, 102: 545–554.
- BAY, H., ESS, A., TUYTELAARS, T. and VAN GOOL, L., 2007. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3): 346–359.
- CHONG, A. K. and BROWNSTEIN, G., 2010. A technique for improving webcam precision in biological plant measurement. *Photogrammetric Record*. 25(130): 180–196.

- DRAP, P. and LEFÈVRE, J., 2016. An exact formula for calculating inverse radial lens distortions. *Sensors*, 16(6): article no. 807.
- FRASER, C. S., 1997a. Digital camera self-calibration. *ISPRS Journal of Photogrammetry and Remote Sensing*, 52(4): 149–159.
- FRASER, C. S., 1997b. Innovations in automation for vision metrology systems. *Photogrammetric Record*, 15(90): 901–911.
- FRASER, C. S., 2013. 'Automatic camera calibration in close range photogrammetry', *Photogrammetric Engineering & Remote Sensing*, 79(4): 381–388.
- FRASER, C. S. and EDMUNDSON, K. L., 2000. Design and implementation of a computational processing system for off-line digital close-range photogrammetry. *ISPRS Journal of Photogrammetry and Remote Sensing*, 55(2): 94–104.
- FRASER, C. S., WOODS, A. and BRIZZI, D. (2005) 'Hyper redundancy for accuracy enhancement in automated close range photogrammetry. *Photogrammetric Record*, 20(111): 205–217.
- HELLWICH, O., ROSE, A., BIEN, T., MALOLEPSZY, C., MUCHA, D. and KRÜGER, T., 2016. Patient registration using photogrammetric surface reconstruction for smartphone imagery. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 41(B5): 829–833.
- HERNANDEZ, A. and LEMAIRE, E., 2016. A smartphone photogrammetry method for digitizing prosthetic socket interiors. *Prosthetics and Orthotics International*, 41(2): 210–214.
- KRAUS, K., 1997. *Photogrammetry: Volume 2: Advanced Methods and Applications*. Fourth edition. Dümmler, Bonn, Germany.
- LERMA, J. L., NAVARRO, S., CABRELLES, M. and SEGUÍ, A. E., 2010. Camera calibration with baseline distance constraints. *Photogrammetric Record*, 25(130): 140–158.
- LERMA, J. L., NAVARRO, S., SEGUÍ, A. E. and CABRELLES, M., 2014. Range-based versus automated markerless image-based techniques for rock art documentation. *Photogrammetric Record*, 29(145): 30–48.
- LOWE, D. G., 1999. Object recognition from local scale-invariant features. *IEEE Computer Society International Conference on Computer Vision*. Volume 2. Washington, DC, USA. Pages 1150–1157.
- LUHMANN, T. and WENDT, K., 2000. Recommendations for an acceptance and verification test of optical 3-D measurement systems. *International Archives of Photogrammetry and Remote Sensing*, 33(B5): 493–500.
- LUHMANN, T., FRASER, C. and MAAS, H.-G., 2016. Sensor modelling and camera calibration for close-range photogrammetry. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115: 37–46.
- MICHELETTI, N., CHANDLER, J. H. and LANE, S. N., 2015. Investigating the geomorphological potential of freely available and accessible structure-from-motion photogrammetry using a smartphone. *Earth Surface Processes and Landforms*, 40(4): 473–486.
- NOCERINO, E., LAGO, F., MORABITO, D., REMONDINO, F., PORZI, L., POIESI, F., ROTA BULO, S., CHIPPENDALE, P., LOCHER, A., HAVLENA, M., VAN GOOL, L., EDER, M., FÖTSCHL, A., HILSMANN, A., KAUSCH, L. and EISERT, P., 2017. A smartphone-based 3D pipeline for the creative industry - the replicate EU project. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42(2/W3): 535–541.
- REMONDINO, F. and FRASER, C. S., 2006. Digital camera calibration methods: considerations and comparisons. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(5): 266–272.
- REMONDINO, F., DEL PIZZO, S., KERSTEN, T. P. and TROISI, S., 2012. Low-cost and open-source solutions for automated image orientation– a critical overview. *Lecture Notes in Computer Science*, 7616: 40–54.
- REMONDINO, F., SPERA, M. G., NOCERINO, E., MENNA, F. and NEX, F., 2014. State of the art in high density image matching. *Photogrammetric Record*, 29(146): 144–166.
- SCHAAF, H., PONS-KUEHNEMANN, J., MALIK, C. Y., STRECKBEIN, P., PREUSS, M., HOWALDT, H. P. and WILBRAND, J. F., 2010. Accuracy of three-dimensional photogrammetric images in non-synostotic cranial deformities. *Neuropediatrics*, 41(1): 24–29.
- WANG, C.-H., MILLS, J. P. and MILLER, P. E., 2012. Automated low-cost photogrammetry for flexible structure monitoring. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 39(B5): 393–398.

Zusammenfassung

Smartphones haben die Möglichkeiten für kostengünstige Erfassung von 3D-Modellen im Nahbereich erweitert. Große Datenmengen können in sehr kurzer Zeit für eine Vielzahl von Anwendungen erfasst werden. Allerdings hängen die

Genauigkeit der Modelle und die Möglichkeiten einer Automatisierung sehr stark von den Aufnahmebedingungen und den Anforderungen der Anwendung ab. In diesem Beitrag werden Genauigkeits- und Zuverlässigkeitsaspekte der abgeleiteten 3D-Modelle in einer sphärischen Messanordnung für Nahbereichsanwendungen (ca. 30 cm) evaluiert. Unterschiedliche Bildanzahl, Aufnahmeanordnungen, Zielmarken, Geräte und Kamerakalibrierungsmethoden werden geprüft und ausgewertet. Die Ergebnisse zeigen, dass für diese Anwendung eine sehr hohe Genauigkeit (0.2 mm) und Zuverlässigkeit erzielt werden kann. Die Zahl der Aufnahmen hat keinen signifikanten Einfluss auf die Genauigkeit, war aber Voraussetzung für die Detektion von Verknüpfungspunkten und die Bildorientierung. Der Einfluss künstlicher Zielmarken hatte einen entscheidenden Einfluss, um die Genauigkeit des Ergebnisses zu erhöhen. Hingegen hatten die Strategie zur Bildkalibrierung und die Eigenschaften des Bildaufnahmesystems keinen großen Einfluss auf die Ergebnisse.

Resumen

Los teléfonos inteligentes han ampliado las posibilidades en la toma de imágenes para modelado tridimensional (3D) de objeto cercano con bajo coste. Estos dispositivos permiten la obtención de gran cantidad de imágenes que pueden usarse en diferentes aplicaciones. La precisión de los modelos y la posibilidad de automatización dependen de las condiciones durante la toma de datos y las necesidades de la aplicación. En este estudio la precisión y la fiabilidad de los modelos fotogramétricos 3D se evalúan para una aplicación de objeto cercano (30 cm) sobre una superficie esférica. Se ha evaluado diferente número de conjuntos de imágenes, la geometría de la red, el dispositivo, la existencia de dianas y la metodología de calibración. Los resultados muestran que en esta aplicación de objeto cercano pueden obtenerse altas precisiones (0.2 mm) y una alta fiabilidad. El número de imágenes no afecta en gran medida a la precisión de los resultados, pero sí a la posibilidad de obtener suficientes puntos homólogos para la creación del modelo. El uso de dianas es el factor que más ha aumentado la precisión. Por otro lado, la metodología de calibración de la cámara apenas ha mejorado la precisión de los resultados.