JOURNAL OF FORENSIC SCIENCES · AAFS

Human height estimation from highly distorted surveillance image

Francesco Tosti Ing[1] | **Carla Nardinocchi PhD**[1] ⬤ | **Wissam Wahbeh PhD**[2] |
Claudio Ciampini TenCol[3] | **Maria Marsella PhD**[1] | **Pierpaolo Lopes Magg**[4] |
Samuele Giuliani Ing[1]

[1]DICEA, Sapienza University of Rome, Rome, Italy

[2]IDIBAU, University of Applied Sciences and Arts Northwestern Switzerland, Muttenz, Switzerland

[3]Reparto Investigazioni Scientifiche, Rome, Italy

[4]Comando Carabinieri Tutela Ambientale, Rome, Italy

Correspondence
Carla Nardinocchi, PhD, DICEA, Sapienza University of Rome, via Eudossiana 18, 00194 Rome, Italy.
Email: carla.nardinocchi@uniroma1.it

Abstract

Video surveillance camera (VSC) is an important source of information during investigations especially if used as a tool for the extraction of verified and reliable forensic measurements. In this study, some aspects of human height extraction from VSC video frames are analyzed with the aim of identifying and mitigating error sources that can strongly affect the measurement. More specifically, those introduced by lens distortion are present in wide-field-of-view lens such as VSCs. A weak model, which is not able to properly describe and correct the lens distortion, could introduce systematic errors. This study focuses on the aspect of camera calibration to verify human height extraction by Amped FIVE software, which is adopted by the Forensic science laboratories of Carabinieri Force (RaCIS), Italy. A stable and reliable approach of camera calibration is needed since investigators have to deal with different cameras while inspecting the crime scene. The performance of the software in correcting distorted images is compared with a technique of single view self-calibration. Both approaches were applied to several frames acquired by a fish-eye camera and then measuring the height of five different people. Moreover, two actual cases, both characterized by common low-resolution and distorted images, were also analyzed. The height of four known persons was measured and used as reference value for validation. Results show no significant difference between the two calibration approaches working with fish-eye camera in test field, while evidence of differences was found in the measurement on the actual cases.

KEYWORDS
camera calibration, height measurements, image distortion, photogrammetry, space resection, terrestrial laser scanner, vanishing line and point

Highlights

- A robust calibration approach is mandatory to extract metric information from video surveillance camera.
- A robust calibration approach improves measurement accuracy.

- Highly distorted frames were analyzed in terms of tangential and radial distortion.
- The approach was tested on several frames acquired from two actual cases.
- A comparison between two different calibration methods is performed.

1 | INTRODUCTION

Metric extraction from video surveillance is a fundamental aspect of forensics science. Video surveillance camera frames are an important source of information during investigations such as the height of an individual present at the crime scene. The greatest challenge is the measurement from a single image. It is common knowledge that highly accurate measurements can be extracted from two or more images at a certain distance from each other. The extraction of metric information from a single view requires additional information on the framed scene. In addition, today, surveillance cameras abound. Moreover, where they are installed and how they are installed (from the top, from the side) also vary. Therefore, investigators have to deal with very different contexts. Finally, considering the legal implications of the measurements in the investigation, verifiable and reliable methods must be used. Regardless of the applied method, the measurements' precision depends on the scene, on the quality, and on the arrangement of the camera. Besides, the individual's posture also impacts the accuracy of the measurement. Walking, standing, or resting positions will produce slightly different heights. Internal scenes usually present several advantages with respect to external ones. First of all, the quality of the image is better as they are not affected by sunlight. Moreover, the presence of easily identifiable lines and points helps to improve the 3D reconstruction of the scene. Regarding the camera, two situations may occur: either the camera has or has not been moved since the perpetrator was filmed. In the first instance, no further acquisitions can be made to improve the quality of the measurements, and therefore, the investigators can only use the already available frames. In the second instance, it is possible to produce further frames to assist the investigation. Finally, another challenging aspect is the distortion which often characterizes video surveillance camera frames. In fact, surveillance cameras mainly present a wide field of view (so-called fish-eye camera) to frame a large portion of the scene: it produces a not negligible distortion in the images, which require a more specific processing step. Even if several camera calibration methods exist, the constrained environment in which the video cameras are located and the investigators operative procedures require finding alternative solutions to obtain a rigorous calibration. Therefore, non-conventional methods can be used to reduce image distortion as illustrated in Wang [1]. In forensics, different methods are applied to extract height from a single image. These are based either on space resection to solve the single frame orientation by means of 3D ground control point (GCPs) or on the principle of projective geometry. Regarding the latter, the software used by the RaCIS is the Amped FIVE by Amped srl. This study has the aim to verify and confirm its accuracy in the presence of highly distorted images. We analyzed the video frames of two actual cases provided us by RaCIS (Raggruppamento Carabinieri Investigazioni Scientifiche) which were relative to two VSC video frames with high lens distortion. Amped FIVE requires the removal of distortion from the images to acquire measurements, and it provides a calibration tool for it. To establish the influence of the calibration process on the results, we implemented a single-view SC approach using the HALCON Library, and we compared results obtained from the same frames corrected with the two different calibration tools: Amped FIVE tool and our single-view SC approach. A field test designed with a large number of GCPs was used with the double goal of verifying Amped FIVE measuring the height of five persons from video sequences acquired with a fish-eye camera and validate our single-view calibration procedure. From a terrestrial laser scanner (TLS) survey of the field test, the signalized GCPs were automatically measured. Reference values, assumed as ground truth, of the camera calibration parameters were obtained using the calibration tools of Photomodeler by EOS working with a SC procedure based on a block of four images. Figure 1 shows the outline of the procedure.

2 | METHODS FOR METRIC EXTRACTION

Methods for metric extraction from a single image were based on either the principles of projective geometry or on space resection. Space resection determines the extrinsic orientation parameters of the VSC (three position parameters and three asset parameters) through the recognition of well identifiable points (GCPs) such as natural or marked points (target). These can be inserted a posteriori into the scene and framed by the VSC. The use of target points inserted a posteriori and measured during the survey assumes that the VSC has not been moved after the crime. Otherwise, it is only possible to use natural, clearly identifiable points and only the available frames.

The GCPs coordinates can be carried out by means of a total station or a TLS. TLS has the considerable advantage in that a three-dimensional (3D) reconstruction of the scene can be performed. A 3D model provides much more information than the 3D coordinates of signalized (or natural) GCPs. De Angelis et al. [2] and Hoogeboom et al. [3] use a TLS to acquire coordinates of GCPs and to produce a 3D model of the scene. Then, they inserted the 3D model in the Autodesk 3D Studio Max environment and reproduced a virtual camera with the same orientation parameters (internal and external) as those obtained from space resection. Frames are acquired by means of the virtual camera within the Autodesk 3D Studio Max environment. With this system, the original image can be superimposed on the virtual one using selected reference points. Height is obtained based on the correspondence between the image created
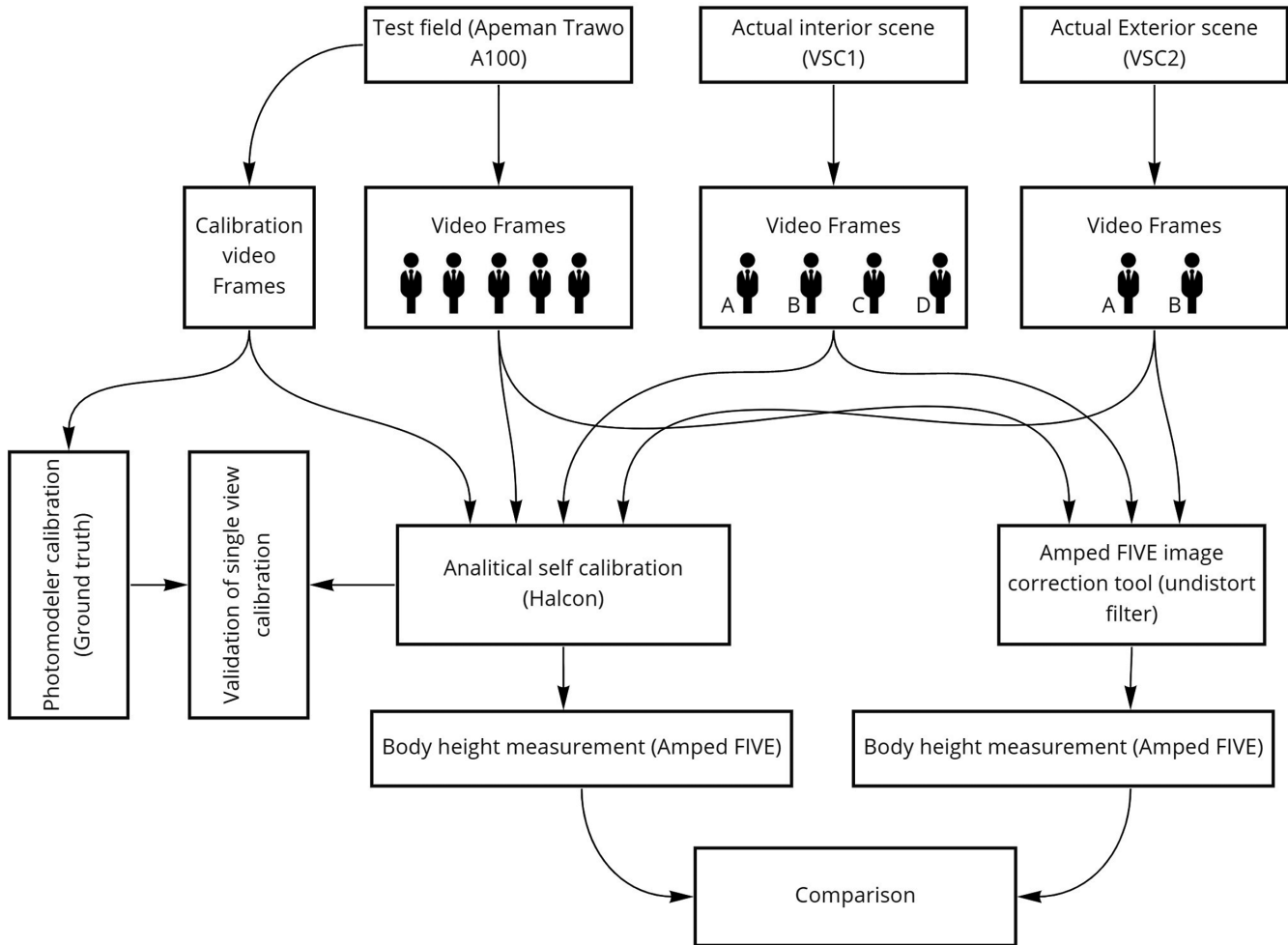
334 | | TOSTI ET AL.

FIGURE 1 Outline of the procedure

within the virtual environment and the real image. The assumption is that the original image has to be distortion-free or at least negligible. Momemi et al. [4] present an approach to estimate the height of an object based on the assumption that the height of the camera and camera focal length are known, and a vanishing point can be extracted. The key of the methodology is to compute the ratio of two adjacent objects in image planes which represents the ratio of the two adjacent objects in real 3D environments. The study by Johnson and Liscio [5] uses the software Scene by FARO. They demonstrated how TLS is a technique that could be potentially used by investigators to determine the suspect's height from video footage obtaining a mean difference between measured and known height to less than a centimeter. Here, it is necessary to georeference the still frame with the scan identifying features (f.i. points) on the image that can also be identified on the 3D model obtained from TLS survey. The still frame is used to texture the 3D model, and then the point at the bottom of the foot and the point at the vertex on the head (projected on the scan model) are recognized. Thanks to knowing the camera's orientation by means of a simple proportion, height can be established. High distortion in the still frame should be corrected while georeferencing the still frame to the scan, but no mention was made of this aspect. A totally
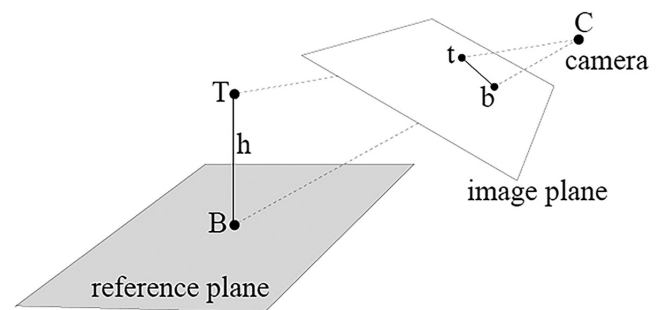

FIGURE 2 Reference plane, unknown segment **TB** in the real world and segment **tb** to measure on the image plane [7]

different approach is the use of projective geometry described by Criminisi et al. [6,7]. Neither the focal length nor position and asset parameters are required and it is merely based on the projectivity of the image. Given a reference plane, it is possible to determine the height of every segment (**tb** in Figure 2) once the projectivity line and vertical vanishing point are known.

To measure the height, the reference plane is horizontal while the distance belongs to a vertical line. This method requires the presence of well-identifiable lines and also the knowledge of a reference

height, which starts from the reference plane, necessary to solve the scale of the image. This method is based on the principles of projective geometry. Once the plane vanishing line and the vertical vanishing point (Figure 3) are detected, the reference height (**tr–br**) is projected along the line to which the height to be measured (**t–b**) belongs, that is, the segment **i–b** in Figure 4. The measurement is taken by means of the line–line homography or cross-ratio between the points' base point on the reference plane, the top point of both the reference height and the searched height, and the vertical vanishing point on the projective line.

This approach relies on the ability to reliably estimate the vanishing line and the vertical vanishing point using the line present in the view. The technique requires a known reference height in the scene. This method is widely used and also provides centimeter accuracy. Forensic science is focused on the accuracy of the measurement that depends on several factors: the camera's arrangement, an individual's posture; the clothing worn such as hats and shoes, the identification of image points corresponding to the subject's head, and feet. Viswanath [8] examines the height error as a function of a subject's location and camera height. They observe the height variation of a pole of known height moved across the scene. An error distribution model is defined to correct the
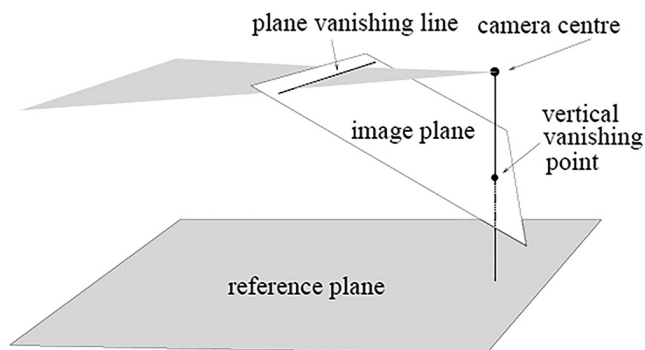
height measured in an image. Edelman [9] compares the methods based on projective geometry and those on space resection. They are comparable and both accurate to within 2 cm. Liscio [10] describes a measurement approach from single surveillance camera image based on the EOS System's Photomodeler. This commercial software, which has a robust bundle adjustment calibration algorithm, has recently updated the space resection software (known as "inverse camera" by Photomodeler) to allow users to correct the camera's distortion using GCPs provided, for instance, by a TLS survey. Twenty or more well-distributed GCPs are recommended to correct the image and calibrate the single view. Height is measured on a virtual plane built in the 3D model in correspondence to the subject's feet. Ljungberg and Sönnerstam [11] have a mean error of 2.30 cm with the single-view metrology approach that also includes the calibration of the low-resolution web camera. The estimated height is lower than actual height probably because of the subject's posture. Moreover, they analyze the influence of the individual's posture while walking and conclude that height has to be measured from the frame corresponding to midstance in walking. Hoogeboom et al. [3] compare the frontal, lateral, and posterior views of subjects, the lateral view being the best to extract height. Finally, different studies (Benabdelkader and Yacoob [12] and Edelman and Alberink [13]) focus on the reliability and accuracy of body height estimations.

FIGURE 3 Vanishing line and vertical vanishing point [7]
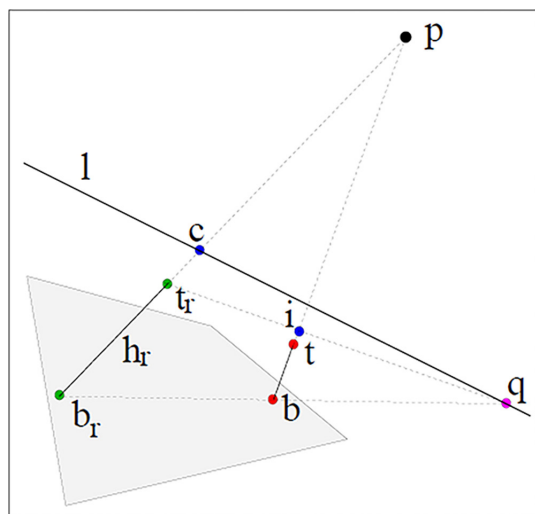
FIGURE 4 Measurement of the segment **tb** along the line Pb [6]

3 | MATERIALS AND METHODS

The camera calibration process recovers the intrinsic and extrinsic camera's parameters. The first depends on the lens and sensor characteristic, while the second on the relative pose between frame and ground reference system. All the projective rays pass through the projective center. Its projection on the image plane is known as the principal point (PP), and its distance to the image plane is called the principal distance. PP coordinates and principal distance are the intrinsic camera parameters.

In addition, because of the camera lens distortion, the three points of a projective ray, an object point, its corresponding image point, and the projective centre, do not belong to the same straight line. The image points are translated by a distance dt producing distorted images. The distortion dt has two main components: the radial distortion (dr) and the decentering distortion (dt), also called tangential distortion as it is perpendicular to the first one. The radial distortion dr is typically larger than the tangential one dt, and it varies with the principal distance. Radial (k_1, k_2, k_3) and tangential (p_1, p_2) coefficients are the intrinsic lens parameters which describe the image distortion [14]. Symmetric radial distortion is represented as an odd-ordered polynomial series: $\Delta r = k_1 r^3 + k_2 r^5 + k_3 r^7$. Tangential distortion can be modeled by Brown correction equation. A useful means of representing the magnitude of decentring distortion is via the profile function: $p(r) = (p_1^2 + p_2^2)^{1/2} r^2$. So, the effects of tangential distortion can be often neglected, but not the radial distortion effects, which produce a typical cushion or barrel distortion. Analytical SC offers a robust method to calculate the orientation parameter as illustrated by Fraser [15], during which both intrinsic and extrinsic parameters are estimated. To rigorously self-calibrate

a digital camera, the photogrammetrist needs only to collect four or more images of a field of a few tens of distinct targets. The calibration of a VSC may only be based on a single-view approach. Moreover, VSC are often characterized by high distortion because their wide field of view and therefore images have a strong lens distortion: straight lines are visibly represented curved in the images reaching thousands of pixel translation at the corner as in the image of Figure 4. The image point translation due to the distortion introduce significant systematic error. A larger number of GCPs is essential to properly describe and correct such lens distortion.

A single-view SC approach was implemented using the HALCON Library [16]. To verify its reliability, a field test (Figure 5 left) with about 90 targets was built. GCP's coordinates were measured with a TLS survey (Figure 5 right) with automatic target recognition obtaining a standard deviation of ±0.7 mm. Several field test frames were acquired from four different positions by a fish-eye camera, the Apeman Trawo A 100, with fixed focal length.

The intrinsic and extrinsic parameters of the camera were solved using the SC algorithm of the Photomodeler [17] in a photogrammetric block of four images, and these calibration parameters were considered as reference (ground truth) and compared with those obtained from the calibration by means of a single frame with the HALCON Library. Because a single image is used and distortion parameters are contemporarily estimated, a high number of GCPs are required. The results of SC have been analyzed as a function of the number and the kind of GCPs used.

3.1 | Reference calibration

The software Photomodeler was used to estimate the calibration parameter of the Apeman Trawo A100. The Apeman Trawo A100 is not a VSC but rather an action camera. But similarly to VSC, it is characterized by a very wide field of view (170°), focal length of 3 mm, and a sensor size of 1/2.33″ (about 6 mm × 4.5 mm) with a resolution of 5120 × 3840 pixels. A four-image block (Figure 6) was adjusted using all the field test targets present in the images and intrinsic and extrinsic camera's parameters were estimated. Intrinsic camera parameters were assumed as ground truth in the following analysis. The first column of Table 1 shows the parameters and their corresponding precision which is of the order of few microns. The PP is relative to the upper-left corner of the sensor.

Tangential distortion is negligible compared to radial distortion: it produces a 1-pixel distortion at the image's border compared with about 1000 pixels with the latter. To further check the orientation's accuracy, 50 check points, such as the door frame or floor corner, were plotted, and their distance from the TLS point cloud, assumed as reference, is calculated. The mean difference is 3.5 mm with a standard deviation of 2.3 mm. To test HALCON software performance, the calibration parameters were solved with HALCON software using the four images, which agrees to tens of millimeters to the ground truth. The second column of Table 1 shows the error of the PP coordinate and of the focal distance obtained with HALCON

TABLE 1 Principal point and principal distance of Apeman Trawo A100 Camera assumed as ground truth and error obtained with HALCON calibration using four images

	Photomodeler (ground truth and its rms)	HALCON-MVTec error (mm)
Principal point O_x (mm)	3.11 ± 0.006	0.010
Principal point O_y (mm)	2.343 ± 0.006	0.007
Principal distance c (mm)	3.29 ± 0.004	0.020

FIGURE 5 Test field with 92 target (left), Point Cloud (right) [Color figure can be viewed at wileyonlinelibrary.com]

FIGURE 6 Four frames used for analytical self-calibration of APEMAN camera [Color figure can be viewed at wileyonlinelibrary.com]

estimation. The distortion profiles of HALCON differ for 10 pixels at 1.8 mm from PP from the distortion profile obtained with the ground truth.

3.2 | SC from single view

All the processing of single-view calibration was realized with the HALCON Library, and each result was compared with the ground truth produced by means of Photomodeler. All the available targets were used in each frame, respectively, 52, 76, 65, and 41 GCPs for images 1, 2, 3, and 4. Table 2 and Figure 7 show the result obtained

TABLE 2 Error of the intrinsic parameter of the four single views, that is the differences between ground truth and the four single view calibration results (one for each frame)

	Image 1	Image 2	Image 3	Image 4
Number of points	52	76	65	41
O_x (µm)	−1	−16	−6	−8
O_y (µm)	1	−8	−3	−5
c (µm)	40	10	20	30

by each frame. Table 2 summarizes the error of the intrinsic camera's parameters, which are the differences between parameters obtained with the single frames and the ground truth (Table 1). The PP and focal distance were estimated with an error in the order of microns–tens of microns. Figure 7 compares the four radial distortion profiles. The black curve in Figure 7 corresponds to the ground truth distortion profile. All the curves are somewhat overlapped within 2.7 mm from PP, at 2.3 mm, we have the half of the image vertical side (vertical line in Figure 7). The yellow and green curves, which respectively correspond to images 2 and 3, fit until the image corner with the ground truth (differences of only 60 pixels at the corner). On the other side, the images 1 (orange-dashed line) and 4 (red-dashed line) starting from 2.3 mm have increasing differences reaching the value of about 450 pixels at the corner. It is important to underline that, in the presence of highly distorted images, it is recommended to measure in the central area of the image, say within the borders of the image. Amped FIVE, for example, does not correct all points of the original image from distortion as shown in Figure 8. Anyway, adopting the single-view SC, the mean errors of all the four calibration are <25 pixels at the image border.

To understand the importance of the number of GCPs, the calibration parameters were then solved with a decreasing number of well-distributed GCP on the image 2 which has the most GCPs. So,

FIGURE 7 Comparison of radial distortion profile of Apeman Trawo produced in a test field. Dashed line corresponds to the ground truth. The yellow and green (dashed) curves correspond to images 2 and 3, the dashed orange and red lines correspond to images 1 and 4. At 2.3 mm, we have the half of the image vertical side. The vertical line indicates half the height of the image (image border) [Color figure can be viewed at wileyonlinelibrary.com]
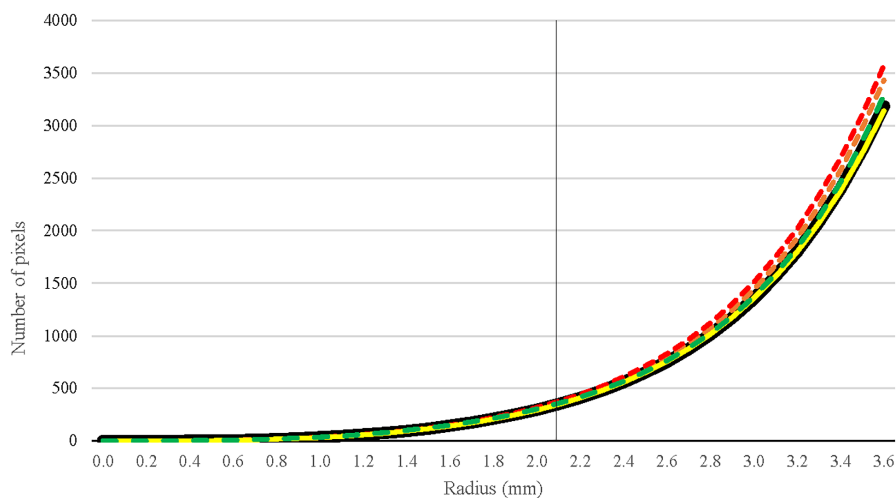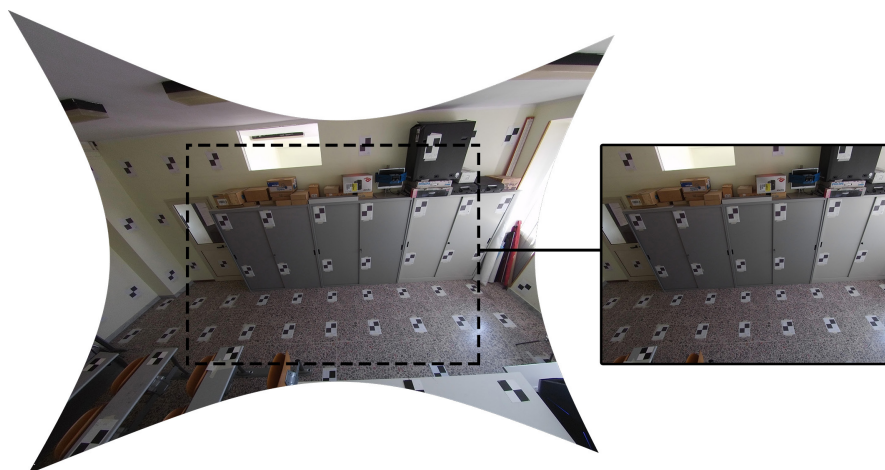

FIGURE 8 Distortion free image produced by the HALCON software (left) and Amped FIVE (right). While HALCON correct the whole image, Amped FIVE produces a cropped corrected image ignoring highly distorted borders [Color figure can be viewed at wileyonlinelibrary.com]

decreasing the number of well-distributed GCPs, we checked the consequently degradation of calibration result (Table 3). We used 60, 40, 20, 10, and 8 GCPs, and from the results, we can infer that at least 20 GCPs are necessary to correctly estimate the radial distortion parameters. That is more evident by looking at the radial distortion profile of Figure 9 where only the central part of the sensor is plotted in the profile to better highlight the differences. Green, yellow, and red profiles correspond, respectively, to the solution obtained with 20, 40, and 60 GCPs, while the black and the purple profiles to those with 10 and 8 targets. The latter two are clearly distinct from the other.

Finally, the calibration is performed only using natural points, which is the only option when the camera was moved after acquiring the image of the perpetrator. In this case, the calibration result depends on the point precision, which may be excellent in indoor situations (like those in the field test) and of lower quality in external scenarios. From the TLS point cloud, 23 natural points, well-identifiable

and evenly distributed, were selected, such as the corner of a piece of furniture or a door frame. Sub-millimeter difference was achieved (−0.12 mm for O_x, −0.23 mm for O_y, and 0.04 mm in the principal distance) in the camera's calibration parameters and a difference of about 15 pixels at the image border in the radial distortion profile. The final step was the distortion correction of the images produced with the HALCON Library. All visible points in the original image are present in the modified (rectified) image. Figure 10 shows the original image and its correspondent distortion-free image. The corrected image was used in Amped FIVE to measure the person's height.

3.3 | Amped FIVE calibration

Amped FIVE has two approaches to calibrate the camera. The first is an "ad oculum" calibration. Visible distortion can be removed manually adjusting the image until the cushion or barrel effects are removed. It goes without saying that this procedure is not recommended, and we have not evaluated its results on the heights measured. The second approach is based on the automatically correction of the cushion and barrel effect after that the operator has provided up to three lines that should be straight in the scene. To compensate for optical distortion, we used the filter "Undistort," both polynomial and rational mapping function were used, and no difference come out. No result is given except the corrected image. In the distortion-free image, parts of the

TABLE 3 Error of intrinsic parameters of self-calibration single view using a decreasing number of targets

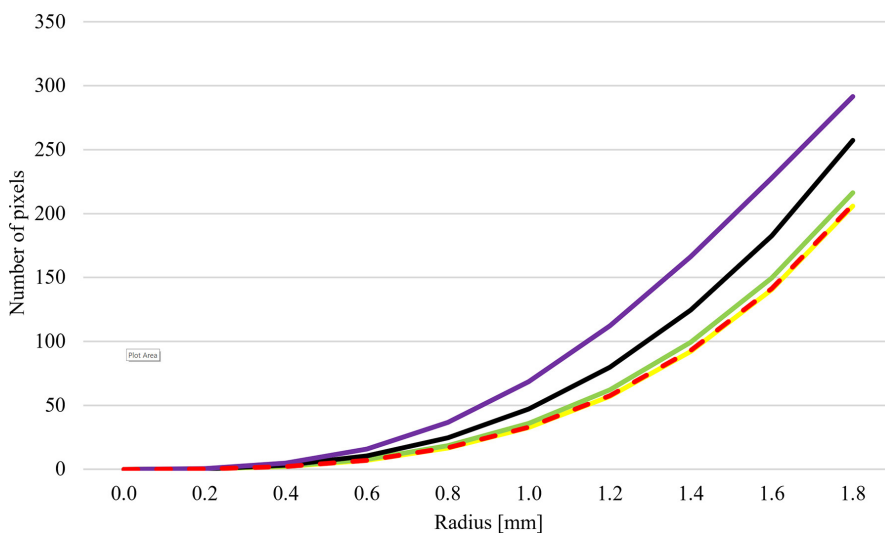|	Single image calibration and GCPs number				
Number of points	60	40	20	10	8
O_x (μm)	−20	−20	−10	10	−120
O_y (μm)	−10	−10	0	−20	−330
c (μm)	10	10	20	90	170

FIGURE 9 Comparison of radial distortion profiles with decreasing number of targets: purple line (8 GCPs), black line (10 GCPs), green line (20 GCPs), yellow line (40 GCPs), and red-dashed line (60 GCPs) [Color figure can be viewed at wileyonlinelibrary.com]

FIGURE 10 Original (left) and distortion free (right) image [Color figure can be viewed at wileyonlinelibrary.com]

original image are cut. Figure 11 shows the original image with depicted one line used for the calibration and the distortion-free image.

4 | RESULTS

Human height was measured using distortion-free still-frames in Amped FIVE. To understand the influence of a rigorous calibration approach, both approaches to image distortion correction are applied: the SC approach and the calibration tool of Amped FIVE. So, for each frame, we produced a double distortion-free image. We analyzed 18 frames acquired in the test field and 15 frames on two actual scenes. In the first, the heights of five people, and in the second, the height of four people, were measured. In the field test, people were asked to walk around the room, were informed of the experiment, and they assumed a straight posture. They were measured in a standing position thus not influencing posture. In the actual video, we selected frames with the investigators engaged in the crime scene. They did not know that they were to be measured. So, in this case, posture may vary, and usually, it is not always straight. Nobody wore hats or shoes with heels. The plane vanishing lines and vertical vanishing point are obtained by means of a couple of lines, and we always used the same couple of lines in the free-distortion images correspondent to the same frame.

4.1 | Human height from test field frames

Body height was measured in 18 frames (Figure 12). The camera was placed at four different points, and five individuals were taken at different distances from the camera. Projectivity detecting two lines for each direction was solved. It is important to underline that we were working with a high-resolution camera and in an environment where lines are very well identifiable and therefore an ideal scene for projectivity analysis. Moreover, measurements were acquired in standing position, and people are aware of the experiment. The Amped FIVE approach provided a mean accuracy of the 18 measurement of 0.2 cm with a maximum error of 0.4 cm and a standard deviation of 0.12 cm. Similar findings were obtained from the image corrected previously with a SC.

4.2 | Human height from actual VSC frames

Body height was measured from two actual videos acquired by the RaCIS during their investigation. We only have a portion of the footage and the points cloud acquired by RaCIS during the TLS survey of the scene. To safeguard anonymity, the still-frames presented were obscured with time and location references, and no frames with the perpetrator were provided by the RaCIS. The two video surveillance cameras share a large angle of view which produces highly distorted images. The first has a low resolution of

FIGURE 11 The line used in Amped FIVE to calibrate the camera and the distortion-free image produced by Amped FIVE [Color figure can be viewed at wileyonlinelibrary.com]
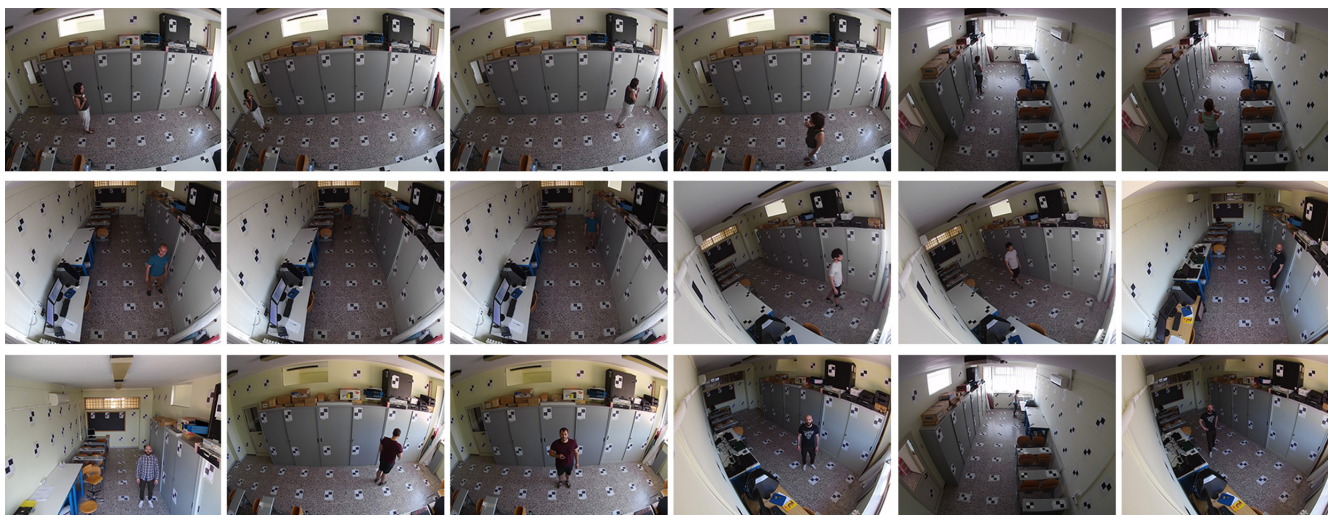

FIGURE 12 The frames acquired in the test field [Color figure can be viewed at wileyonlinelibrary.com]

about 0.5 Mpixel (960 × 480 pixel) and was installed in an interior environment, while, the second, which has double the resolution (1280 × 960 pixel), was placed externally. Figure 13 shows the points used for the SC. More specifically, there are 23 GCPs for the first scene, 8 of which were signalized targets, and 22 natural GCPs for the second. Figure 14 shows the distortion corrected images. It is important to underline that GCP selection is driven by its visibility both in the TLS point cloud and in the still-frame.

We would like to highlight the differences between these VSCs and the Apeman Trawo. The first is the different sensor size. Despite the fact that no specific information about the cameras were available, by analyzing the scale image and its pixel resolution, we can assume a sensor of about 1/4″ and 1/3″ (half diagonal of the sensor is about 2.1 mm for the first VSC and 2.8 mm for the second one). Moreover, the first video camera had a resolution of 960 × 480, with an uncommon ratio width/height equal to 2:1, and the second 1280 × 960, with a ratio width/height equal to 4:3. Figure 15 shows the radial and tangential distortion profiles of the two VSCs compared to that of the Apeman Trawo. Black curves correspond to Apeman Trawo A100, the red curves correspond to the first

VSC, and the blue curves correspond to the second VSC. The radial distortion profile is plotted until 200 pixels of distortion to better compare the curves. Tangential distortion is plotted with dashed lines. Both VSCs also had a slightly larger tangential distortion compared with the Apeman, which has quasi-zero tangential distortion. The first VSC has the largest radial distortion that requires a very accurate distortion correction.

After the distortion correction, it is possible to solve the image projectivity. Amped FIVE requires the identification of the vanishing line and vertical vanishing point by means of two lines for each direction. The use of more than two lines should improve the results (Creminisi et. al., 2002), but from our experience, the addition of a line can worsen the least square solution of the vanishing point and line. For instance, in the external environment (Figure 14 right), there are only a few valid lines along the transversal direction (green lines). Therefore, only the two better-defined lines for direction were selected. Figure 14 shows the lines used to solve the projectivity that were the same in both distortion-free images. Before measuring height, a fixed element belonging to the environment was observed to ensure the correct

FIGURE 13 GCPs used for the self-calibration: VSC1 (left) and VSC2 (right) [Color figure can be viewed at wileyonlinelibrary.com]
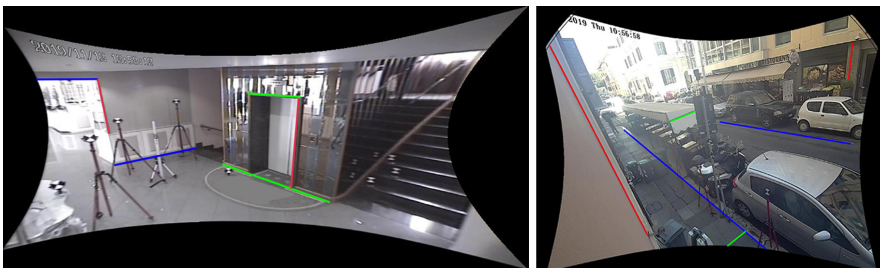

FIGURE 14 Distortion corrected images and line used in Amped FIVE to estimate vanishing line and vanishing vertical point: VSC1 (left), VSC2 (right) [Color figure can be viewed at wileyonlinelibrary.com]
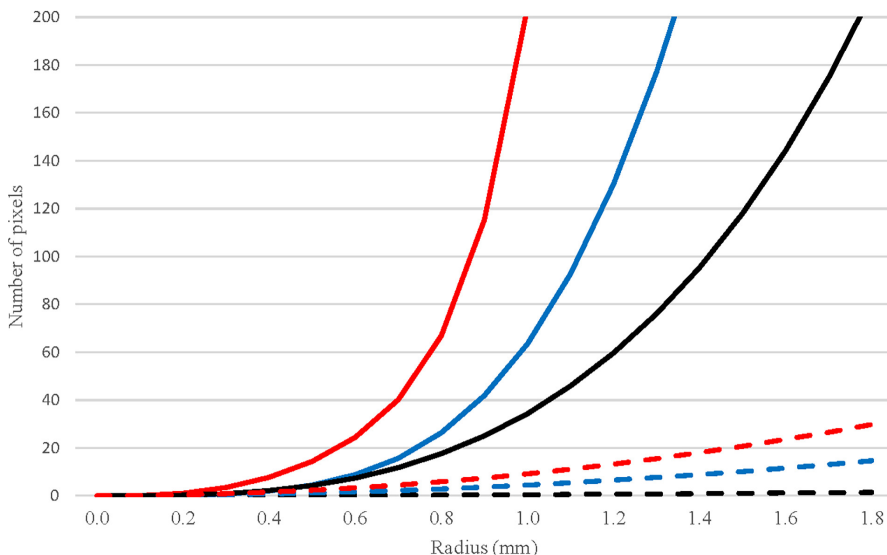

FIGURE 15 Radial and tangential distortion profile of the three cameras. Black lines correspond to Apeman Trawo A100, the red lines correspond to the first VSC, and the blue lines correspond to the second VSC. Dotted lines correspond to tangential distortion [Color figure can be viewed at wileyonlinelibrary.com]

extraction of the vanishing point and line. It was an element of 90 cm in the first scenes and of 225 cm in the second scene. A well-defined element reduces the measurement errors introduced during the selection of the points corresponding to the feet and head. The measurements were repeated three times, and the mean error, root mean square, and minimum and maximum errors were calculated (Table 4). In both videos, the element was underestimated with an error of −3.22 and −4.9 cm, respectively, in the first and the second VSC with a maximal error of

−5.88 cm in the frames corrected inside Amped FIVE. By adopting the analytical SC, a sub-centimeter accuracy was confirmed. Finally, 15 frames including investigators (of known identity) during their inspection were selected. More specifically, four investigators were recognized, hence called (Figure 16) persons A, B, C, and D. Frames with subjects in the straightest possible positions were selected. Each person appeared in more than one frame and two (i.e. persons A and B) were present in both VSCs. Figure 17 shows the Amped FIVE interface

TABLE 4 Mean error, root mean square error, maximum error, and minimum error in the measurement of a well-defined fixed element present into the two scenes

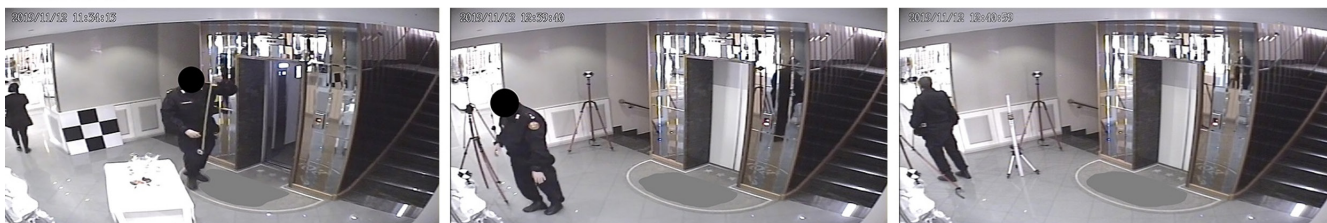|	HALCON distortion correction (cm)				Amped FIVE distortion correction (cm)			
VSC	Mean	rmse	max	min	Mean	rmse	max	min
1	0.40	0.22	0.09	0.72	−3.22	0.47	−3.56	−2.89
2	0.87	0.30	0.53	1.07	−4.9	1.38	−5.88	−3.92

Person A
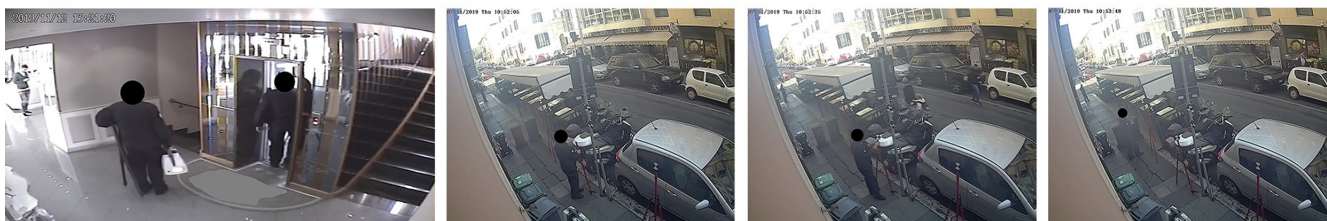
Person B
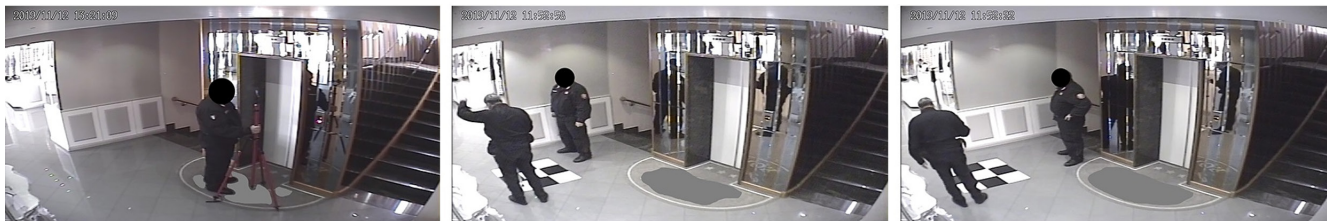

Person C

Person D

FIGURE 16 The frames extracted from the two VSC footage [Color figure can be viewed at wileyonlinelibrary.com]

FIGURE 17 Height measurement of person C by means of Amped FIVE. (Left) distortion correction with self-calibration, (right) distortion correction with Amped FIVE [Color figure can be viewed at wileyonlinelibrary.com]

during the measurements of the individuals' heights. The operator had to identify a point in correspondence to the feet and the other point corresponding to the head's vertex. The feet had to belong to the reference plane, and the point is the intersection of the two lines connecting the toes of one foot with the heel of the other. Toes and heels were not always well identifiable. This depends on posture and on the body side present in the frame. In some situation, the operator can be uncertain about which point identifies the feet; therefore, several measurements were realized according to the most probable head-foot line. We observed that the height changes always <1 cm with respect to the final mean value with a standard deviation of half a centimeter.

The measurement in each frame was realized by Amped FIVE in the image corrected by the distortion both with the SC approach (left image of Figure 17) and with the Amped FIVE calibration tool (Right image of Figure 17). Table 5 shows the height of the four people deducted from the two distortion-free images and actual height that correspond to the mean value of all the measured frame, that is four measurements for person A, five for person B, and three for both persons C and D. It is important to underline that the operator did not know the height of the four persons in advance.

The presence of persons A and B in both videos allows to do a more detailed examination of the single observations in both VSCs footage. The number in parenthesis in Table 6 is the number of frame where the corresponding person was measured in videos 1 and 2. More specifically, person A was measured once in the first VSC and three times in the other. Therefore, the mean value of 173.8 cm (Table 6) corresponds to the unique height observed of person A in the first VSC. The person B was measured three times in the first VSC and twice in the second. We observed smaller differences between the height measured in both persons A and B in the image which distortion is corrected by means of the SC approach, that are 3 mm for person A and 9 mm for person B. Instead, there are 6.3 cm differences between the height of person A measured on frames of the first and the second video corrected with Amped FIVE calibration tool. The difference is 9 mm for person B. We also observed larger error (with respect to true height) for the person B with both calibration methods: more specifically 3.5 cm using the SC and 7.5 cm with Amped FIVE calibration that could be explained by a systematic error caused by the individual's posture.

A better understanding of the result may be given taken in account the dispersion of the measurement (Table 7). Person A has the largest standard deviation in both calibration methods, 1.8 and 4.9 cm, respectively, as evidenced by the minimum and maximum value. All the other people have a sub-centimetric precisions. To

TABLE 5 Subject's height deducted from the analysis of the frame corrected with the self-calibration, with Amped FIVE line projectivity-based approach and the actual height

	SC Amped FIVE (cm)	Amped FIVE (cm)	True height (cm)
A	174.1	174.0	173
B	174.6	170.4	178
C	187.3	181.9	187
D	172.9	170.3	173

Note: It corresponds to the mean value of all the measured frames.

TABLE 6 Heights of individual A and B obtained in the frames corresponding to the two different video cameras

	Person A				Person B			
	Self-calibration		Amped FIVE		Self-calibration		Amped FIVE	
		Mean (cm)	Mean (cm)			Mean (cm)	Mean (cm)	
VSC1	(1)	173.8	169.3		(3)	175.1	170.2	
VSC2	(3)	174.1	175.6		(2)	176.0	171.1	
Difference (cm)		0.9	6.5			0.9	0.9	

Note: Number in parenthesis is the number of frame where the corresponding person was measured in videos 1 and 2. In the last raw. True height for person A is 173 cm and for person B is 178 cm.

TABLE 7 Error, mean value of the height, standard deviation of the measurements, and minimum and maximum value

	Self-calibration					Amped FIVE				
	Error (cm)	Mean (cm)	rms (cm)	Min (cm)	Max (cm)	Error (cm)	Mean (cm)	rms (cm)	Min (cm)	Max (cm)
A (4)	1.1	174.1	1.8	172.1	176.5	1.0	174.0	4.9	169.3	180.2
B (5)	−3.4	174.6	0.9	173.4	175.5	−7.6	170.4	0.6	169.8	171.1
C (3)	0.3	187.3	0.4	187.0	187.5	−5.1	181.9	0.7	181.4	182.4
D (3)	−0.1	172.9	0.4	172.5	173.2	−2.7	170.3	0.3	169.9	170.6

Note: Number in parenthesis is the total number of frames where the corresponding person was measured.

conclude, applying a more rigorous SC to the images improves the measurement's accuracy by several centimeters.

5 | CONCLUSION

Height extraction from VSCs is an important aspect of forensic science. Image distortion affects accuracy which is compromised when distortion is not correctly modeled. This study aimed to analyze the application of the Amped FIVE software when high distortion images were used. We showed how a SC approach improves accuracy by several centimeters compared with a calibration based on a vanishing line used by the Amped FIVE software. To realize that, the influence of camera calibration and distortion correction was analyzed both with frames of a footage acquired in a field test ad hoc designed with about 90 GCPs and with frames from the footage of two VSCs used by the RaCIS during their enquiry. In the field test, a footage was acquired by the ApemanTrawo A100, and intrinsic camera parameters were estimated both with a rigorous bundle adjustment, using the Photomodeler photogrammetric software, and by a single view calibration approach, using the HALCON Library software. We observed that, using the SC approach, parameters were correctly estimated with at least 20 evenly distributed GCPs. Then, to confirm the consistency of the Amped FIVE approach, the heights of five persons were measured from the frames of the Apeman Trawo, and the heights of a further four persons were measured using frames from the footage of two VSCs used by the RaCIS. Despite all the cameras produce high distortion images, we observed a different performance of Amped FIVE with the Apeman Trawo and with the two VSC. In fact, heights measured in the field test (based on frames acquired with the Apeman Trawo) had a sub-centimeter mean error, with a standard deviation of 0.2 cm and a maximum error of 0.39 cm, while heights measured in the frames extracted from the VSCs footage had a total mean error of 4.1 cm considering the mean error obtained with the four people (that is 1.0 cm for the person A, 7.6 cm for the person B, 5.1 cm for the person C, and 2.7 cm for the person D), with a maximum error of 7.6 cm (person B) and a standard deviation of 2.9 cm. To understand the different performance of the Amped FIVE software, we had compared the distortion profile of the three cameras. All three cameras, the Apeman Trawo and the two VSCs, produced significantly distorted frames. Differences among them regarded the absence of tangential distortion using Apeman Trawo compared with a very low tangential distortion with the other two VSCs. Moreover, VSCs have a low-resolution sensor corresponding to 1/4″ and 1/3″ and a higher radial distortion. Applying the distortion correction based on SC to the frame, the total mean error was 1.2 cm with a standard deviation 1.9 cm. A larger error of 3.5 cm was obtained for person B, which could also depend on the systematic error given by his posture during the acquisition. To conclude, considering that investigators have to use different cameras while inspecting the crime scene, often using low resolution and with high-distortion lens camera, a stable and reliable approach is needed. Since, a TLS survey is used in their routine work to fix and register the

crime scene, it is easy to provide the measurements of several (about 20) GCPs to achieve a more rigorous camera calibration. To permit GCP detection and simplify its recognition, the scan must be as complete as possible to avoid data gaps caused by the presence of objects along the scan line. Amped FIVE is based on a robust approach, whereby height measurement from video surveillance camera with centimetric accuracy was obtained. Its calibration, also based on projective geometry, can affect such accuracy when it cannot correctly model the distortion and, therefore, the use of distortion corrected images produced with a SC approach is recommended.

ORCID

Carla Nardinocchi 🔵 https://orcid.org/0000-0003-2644-4569

REFERENCES

1. Wang Z. Pin-hole modelled camera calibration from a single image. ACM J. 2009;V:1–36.
2. De Angelis D, Sala R, Cantatore A, Poppa P, Dufour M, Grandi M, et al. New method for height estimation of subjects represented in photograms taken from video surveillance systems. Int J Legal Med. 2007;121:489–92. https://doi.org/10.1007/s00414-007-0176-4.
3. Hoogeboom B, Alberink I, Goos M. Body height measurements in images. J Forensic Sci. 2009;54(6):1365–75. https://doi.org/10.1111/j.1556-4029.2009.01179.x.
4. Momeni-K M, Diamantas S, Ruggiero F, Siciliano B. Height estimation from a single camera view. In: Csurka G, Braz J, editors. Proceedings of the 2012 international conference on computer vision theory and applications (VISAPP 2012); 2012 Feb 24–26; Rome, Italy. Lisbon, Portugal: Institute for Systems and Technologies of Information and Communication; 2012. p. 358–64. https://doi.org/10.5220/0003866203580364.
5. Johnson M, Liscio E. Suspect height estimation using the Faro Focus3D Laser Scanner. J Forensic Sci. 2015;60(6):1582–8. https://doi.org/10.1111/1556-4029.12829.
6. Criminisi A, Reid I, Zisserman A. Single view metrology. In: Proceedings of the seventh IEEE international conference on computer vision; 1999 Sep 20–25; Kerkyra, Greece. Piscataway, NJ: IEEE; 1999. p. 434–41. 10.1109.ICCV/1999.791253
7. Criminisi A, Zisserman A, Van Gool L, Bramble S, Compton D. A new approach to obtain height measurements from video. In: Higgins K, editor. Proceedings SPIE3576, investigation and forensic science technologies. Bellingham, WA: SPIE; 1999. p. 227–38. https://doi.org/10.1117/12.334540.
8. Viswanath P, Kakadiaris IA, Shah SK. A simplified error model for height estimation using a single camera. In: Proceedings of the international conference on computer vision workshops; 2009 Sep 27–Oct 4; Kyoto, Japan. Los Alamitos, CA: IEEE Computer Society Publications; 2009. p. 1259–66. https://doi.org/10.1109/ICCVW.2009.5457466.
9. Edelman G, Alberink I, Hoogeboom B. Comparison of the performance of two methods for height estimation. J Forensic Sci. 2010;55(2):358–65. https://doi.org/10.1111/j.1556-4029.2009.01296.x.
10. Liscio E. Extracting accurate measurements from single surveillance camera images. Photomodeler. 06/15/2017. https://www.

photomodeler.com/article/extracting-accurate-measurements-from-single-surveillance-camera-images/. Accessed 27 Jul 2021.

11. Ljungberg J, Sönnerstam J. Estimation of human height from surveillance camera footage – A reliability study [dissertation]. Jönköping, Sweden: School of Health Sciences, University of Jönköping; 2008. https://www.diva-portal.org/smash/get/diva2:67847/FULLTEXT01.pdf. Accessed 25 Aug 2021.

12. Benabdelkader C, Yacoob Y. Statistical body height estimation from a single image. In: Proceedings of the international conference on automatic face & gesture recognition; 2008 Sep 17–19; Amsterdam, Netherlands. Los Alamitos, CA: IEEE Computer Society Publications; 2008. p. 1–7. https://doi.org/10.1109/AFGR.2008.4813453.

13. Edelman G, Alberink I. Height measurements in images: how to deal with measurement uncertainty correlated to actual height. Law Probab Risk. 2010;9(2):91–102. https://doi.org/10.1093/lpr/mgp033.

14. Cattaneo C, Mainetti G, Sala R. The importance of camera calibration and distortion correction to obtain measurements with video surveillance systems. J Phys Conf Ser. 2015;658:012009.

15. Fraser CS. Digital camera self-calibration. ISPRS J Photogramm Remote Sens. 1997;52(4):149–59. https://doi.org/10.1016/S0924-2716(97)00005-1.

16. Halcon. https://www.mvtec.com/products/halcon. Accessed 27 Jul 2021.

17. Photomodeler. https://www.photomodeler.com. Accessed 27 Jul 2021.