IET Computer Vision

Research Article

IET Journals
The Institution of Engineering and Technology

Accuracy assessment of single viewing techniques for metric measurements on single images

ISSN 1751-9632
Received on 9th November 2017
Revised 4th March 2018
Accepted on 7th March 2018
E-First on 27th March 2018
doi: 10.1049/iet-cvi.2017.0549
www.ietdl.org

Ozan Arslan[1] ✉

[1]*Geomatics Department, Kocaeli University, Engineering Faculty, 41380, Kocaeli, Turkey*
✉ *E-mail: oarslan@kocaeli.edu.tr*

Abstract: Single view methodology has been a long-standing issue of extracting metric information from a single image in the relevant disciplines. To perform metric measurements on a single image, two methods, which are based on identifying vanishing points (VPs), were employed in the study. One of the two methods is based on the cross ratio and VPs computation and the other integrates robust statistical estimators and close range techniques in a multidisciplinary concept. The estimation of object distance from a single image taken in an outdoor environment was the main concern for a practical application in the study. As the main concern of the study, accuracy analysis of distance measurements was performed to compare the performance of two techniques and experimental results are presented.

1 Introduction

Getting precise three-dimensional (3D) metric information from the 2D images obtained from different viewing conditions has always been a popular research area in the field of photogrammetry and computer vision. Generally, the tools used for this purpose are based on traditional stereo vision methods which use stereo image pairs to extract precise 3D information. Yet it sometimes might not be possible to obtain such images with suitable viewing geometry for practical reasons or only single photographs are available. It is still possible to infer 3D information of the object geometry depicted in the single image by analysing certain image characteristics. When only one image portraying the scene exists, the primary problem is the recovery of the third dimension considering that this information was not captured in the image acquisition process. To sum up, then it is feasible to extract geometrical information of the original object and even to reconstruct the 3D object model from a single view. Some approaches and techniques developed in related fields for 3D object reconstruction based on a single image were briefly reviewed in [1].

There are a number of works which address the problem of 3D modelling or size estimation from single images using calibrated or uncalibrated cameras. One of the first researches was performed by Debevec *et al.* [2] developing a reconstruction system from single images by exploiting manually selected edge features to build a model from 3D primitives. A notable study was presented by El-Hakim [3], which is a flexible and accurate method which can be used within a wide range of structures. The approach does not need camera calibration or vanishing points (VPs) instead a variety of constraints based on topological relations and basic primitives are used. A widely known or cited study on the subject has been published by Criminisi *et al.* [4] and a technique for accurate visual metrology was presented from single uncalibrated images using some geometrical assumptions in this study. It is known that the affine 3D geometry of a scene can be measured by a single perspective image when the geometrical properties (e.g. perpendicularity and parallelism) of a reference plane are taken into account. The technique is generally based on the estimation of the vanishing line of the ground plane and the vertical VP using some geometric features in the image. In addition, some different methods have been recommended for distance or height estimation using a metric reference of an object in order to correctly scale affine distance measurements to Euclidean ones [5]. A visual metrology method has also been proposed by utilising uncalibrated images separated by an almost pure translation [6]. In another study, the height of objects has been inferred from an unmanned aerial vehicle using sequential images that are captured by a single camera and with known base distances of at least two objects [7]. A significant number of studies dealt with human height estimation have been a growing field of interest in the last few years with applications in forensics, video surveillance, and human tracking [8–10]. Some methods used a Bayesian-like statistical framework to estimate the height of human body without using a reference measure in the image plane. In that case, the intrinsic and extrinsic parameters of the camera need not be known (i.e. uncalibrated). The method has also been adopted in robot navigation for the extraction of vertical lines of an indoor environment to make the robot able to navigate autonomously [11, 12]. Some recent studies have focused on utilising smart mobile devices which are embedded with cameras and sensors for distance measurement [13, 14]. There are also some other approaches proposed in the literature but they are not included in this work because they have the inadequate capacity and have lower metric accuracy levels. There are relatively new flexible techniques which combine approaches from close range photogrammetry and computer vision to get metric information from a single image [1, 15, 16]. These methods include digital image analysis (edge detection, clustering etc.), robust VPs computation, camera self-calibration, and dimensional analysis in a multidisciplinary context.

The single view object modelling and measuring approaches are diverse in the related fields but a main categorisation for the single imaging methodology has not been accomplished so far. Hence, in this study, it was deemed appropriate to consider the methods used for extracting metric information from a single view in two basic categories. As is known VP detection is a cornerstone for the categorisation of all the methods. For this reason, single viewing techniques for acquiring 3D metric information were considered in the two main categories depending on the detection of VPs in this study. It is of importance to identify VPs correctly in order to obtain accurate 3D information. There is a common approach for 3D object reconstruction based on VPs from a single image with perspective projection in the related literature. The detection of VPs assuming perspective projection is a requirement for measuring and modelling of objects on single images (Fig. 1). VPs build the framework that supports the whole process in structural environments since they provide independent geometric constraints which can be exploited in many ways. The detected geometrical features are then used to define the different dimensions of the scene, i.e. ground plane and vanishing lines. This information is

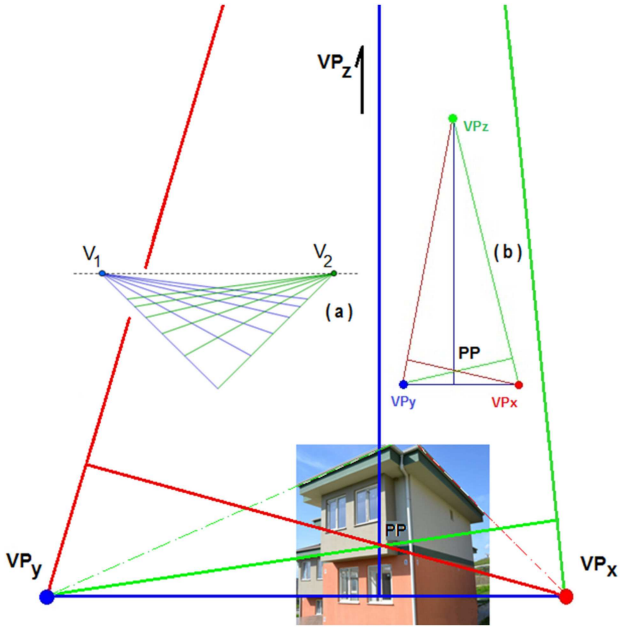

Fig. 1 *Vanishing point detection from a single image*
(a) VPs and vanishing line (dashed),
(b) PP and three orthogonal vanishing lines

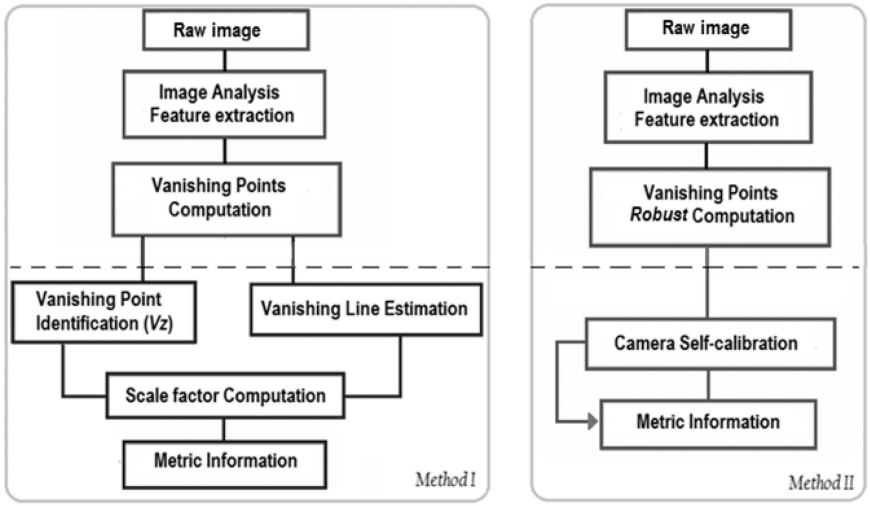

Fig. 2 *Methods of obtaining metric information from a single image*

lastly used in a standard photogrammetric procedure with the final goal of measuring the object dimensions in single images. To estimate the locations of three main VPs in an effective manner in the image, digital line segmentation and clustering algorithms are generally utilised following an automatic vectorisation process. One of the most critical steps in this process is concerned with the identification of the structural elements with a high precision belonging to a raw or un-oriented image. Noteworthy to say that camera calibration need not be known in this process. If the calibration process is used, interior orientation of the camera can be computed directly using the geometric relation between the VPs and the three main object directions. To make precise metric measurements on a single image, the two VP-based methods were adopted in the study as previously explained. As one of the main focuses of the study, accuracy analysis of distance determinations from a single image was performed to compare the performance of the two techniques and experimental results are presented to estimate uncertainties of measurements in this work. One of the methods is based on the cross ratio and VP computation and the other one uses robust statistical estimators for VPs detection. In the study, the measurement of the distances marked on a building was the main concern for practical applications taken in our university campus environment. The measured distances will be compared in a statistical context for determining the accuracy levels of the two methods.

2 Materials and method

Computer vision and photogrammetry has always been used in acquiring precise 3D metric information of objects from images. These disciplines can be divided into several categories in view of different perspectives. The approaches can generally be divided into two categories: the methods that exploit 3D information from multiple (or stereo) images and the ones that analyse geometric properties in a single image. The first category is a widely known method and deals with the 3D reconstruction of the object scene using stereo vision, to estimate homography or to calibrate the camera [6, 17, 18]. Single viewing methods have generally based on the analysis of geometric properties [19–21] but, in this scenario, they often use previously positioned markers for the calibration process. Several methods have been suggested to get metric information from single images in the literature [22, 23]. In the study, the two main categories adopted might be addressed by the two methods as shown in Fig. 2.

The common feature of both methods is that they use VPs detection producing higher accuracy values of metric information. The method illustrated in the left (Fig. 2) is generally based on the

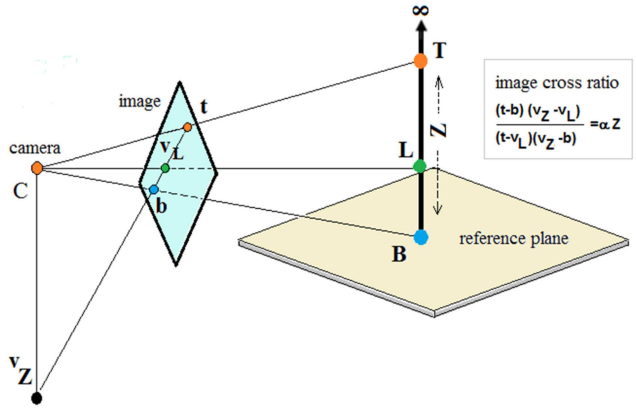

Fig. 3 *Perspective imaging: the vanishing line of the reference plane*

works of Criminisi and it does not require camera calibration. It is called as method I in the subsequent sections of the paper. The other method shown in the right (Fig. 2) is mainly adopted from Aguilera [16, 24] and includes camera self-calibration. This method is also called as method II in the subsequent sections. Method II incorporates several robust and statistical techniques in order to guarantee reliability in the automation of the process. The interior orientation (or calibration) parameters (i.e. the focal length, the principal point (PP), and the lens distortion) can be recovered automatically based on VPs geometry (Fig. 1*a*). The method combines several approaches (i.e. photogrammetry, computer vision, and robust estimators) for 3D reconstruction [15].

The study focuses on getting metric information (i.e. distance determination on the objects) from single images in practice as previously mentioned. Several camera calibration methods based on varied purposes rely on estimating VPs have been proposed. It has been reported that estimating the VPs is usually the bottleneck of these methods because it is sensitive to noise. There has been an extensive research on the problem of detection of VPs with a high-quality degree. VPs emerge from the intersection of a bundle of parallel lines in the scene, which become perspective lines in its representation. Therefore VPs are traditionally estimated from the intersections of hypothetic perspective lines or by minimising the distance to projections of parallel lines. These procedures cannot be regarded as robust for the photogrammetric perspective. In fact, a VP can only be determined with only strong perspectives. Complex and time-consuming techniques have been used to solve the general problem of VP detection. Some of the VP detection techniques use a Hough transform approach in which the parameter space is located on a so-called Gaussian sphere [25, 26]. Note that most techniques incorporate variations of the iterative optimisation techniques. Kosecka and Zhang [27] use the expectation maximisation algorithm that estimates the coordinates of the VPs in an iterative fashion and the probabilities of individual line segments for a given vanishing direction. Tardif [28] and Xu *et al.* [29] use an algorithm for estimating VP hypotheses in the image plane using pairs of edges and compute consensus sets using the J-linkage algorithm. Wildenauer and Hanbury [30] have proposed a RANSAC-based approach using a solution to estimate three orthogonal VPs from four aligned lines with two or three orthogonal directions. Tretyak *et al.* [31] use different types of geometric primitives and construct energy functions for each layer with respect to hypothesised VPs [31]. In the majority of the algorithms described above, the RANSAC method and iterative optimisation schemes were used, which can lead to complex parameter tuning and high computational cost. Lezama *et al.* [32] proposed a deterministic and non-iterative method but the method still has a high computational cost.

2.1 Method I

Metric rectification of a plane in a single image also partially determines camera calibration. It has been shown that three VPs for orthogonal directions allow partial calibration of a camera from a single view. The PP of the camera is at the orthocentre of the triangle which has the VPs as its vertices (Fig. 1*b*).

As is known from Criminisi [33] when the required vanishing line and VP are determined together with the image of the reference line segment, the vertical reference height can be measured based on the cross ratio (Fig. 3). In this figure, point C represents the perspective centre, V_z indicates the vertical VP in image space and point V_L indicates the intersection between the vanishing line and the image plane. The distance between points t and b of an object can be estimated using a reference distance on the scene.

As is known many studies have verified the robustness of camera calibration methods based on VPs [34]. Given VPs and reference height information, any distance on the object can be computed straightforwardly. The projection of a world point $X \in R3$ into an image point $x \in R2$, considering perspective projection, is described by the projection matrix

$$\tilde{\boldsymbol{x}} = \boldsymbol{P}\tilde{X} = \boldsymbol{K}\,[\boldsymbol{R}|\boldsymbol{T}]\,\tilde{\boldsymbol{X}}, \tag{1}$$

where $\tilde{\boldsymbol{X}}$ and $\tilde{\boldsymbol{x}}$ are the points $\boldsymbol{X}$ and $\boldsymbol{x}$ in homogeneous coordinates, respectively; $\boldsymbol{K}$ is the matrix representing the intrinsic parameters (i.e. interior orientation) of the camera; the extrinsic parameters are $\boldsymbol{R}$ – rotation matrix – and $\boldsymbol{T}$ – translation vector from the world to the camera system (i.e. exterior orientation). Equation (1) can also be simplified in the following form:

$$\tilde{\boldsymbol{x}} = [p_1\ p_2\ p_3\ p_4]\tilde{\boldsymbol{X}}, \tag{2}$$

where p_1, p_2, p_3, and p_4 are the columns of P; and the equality is up to scale. Wang *et al.* [35] prove that p_1, p_2, and p_3 are the orthogonal VPs corresponding to the world (object) coordinate system and that p_4 is the image of the world origin. Here we expressed these orthogonal VPs as v_x, v_y, v_z [36]. The final projection matrix is

$$\boldsymbol{P} = \left[v_x v_y \alpha v_z \tilde{m}\right], \tag{3}$$

where α is an unknown scalar referred to as the *metric factor.* If v_z and m are available, then the metric factor α is the only unknown value. To measure heights in images, we need to compute the distance between points in two different planes. Let v_z be the VP that indicates the scene vertical direction, and m the ground vanishing line. It has been proved that for an arbitrary object with height Z and delimited by image points t and b (top and bottom points), it holds that

$$\alpha Z = \frac{-\|\,\boldsymbol{b}\times\boldsymbol{t}\,\|}{\tilde{m}\boldsymbol{b}\|\,v_z\times\boldsymbol{t}\,\|}. \tag{4}$$

It is simple to compute Z via a reference measurement in the image with a known length. Thus, if we want to measure height Z_{obj} of an object, and Z_{ref} is a known distance on the same image, then (4) allows the computation of the scale factor α and subsequently the distance Z_{obj} between t and b object points.

The process of computing distances signalised on the building in images is illustrated in Fig. 4. Given VPs and reference length information, the distances on the wall of the building can be computed straightforwardly. Reliable measurements in images can be done by accurately detecting the VPs and estimating vertical VP V_z and vanishing line m that represents the ground plane efficiently. The cornerstone of this calculation is an accurate estimation of the vertical VP and the vanishing line as explained before.

2.2 Method II

Method II presents an integrated approach combining close range photogrammetry, computer vision and robust statistics in an efficient manner. The methodology includes some well-known steps in close-range photogrammetry, i.e. features extraction, VP computation, camera self-calibration, and dimensional analysis [15, 16, 22, 24, 37]. Structural elements like straight lines are detected with high precision using a hierarchical and hybrid approach.

Fig. 4 *Concept of measuring the distances on the building depicted in a single image*

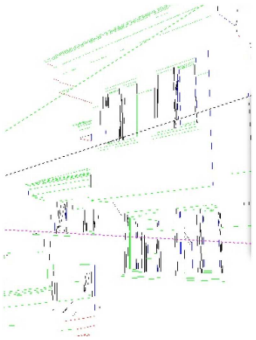

Fig. 5 *Line segmentation and clustering in the single image*

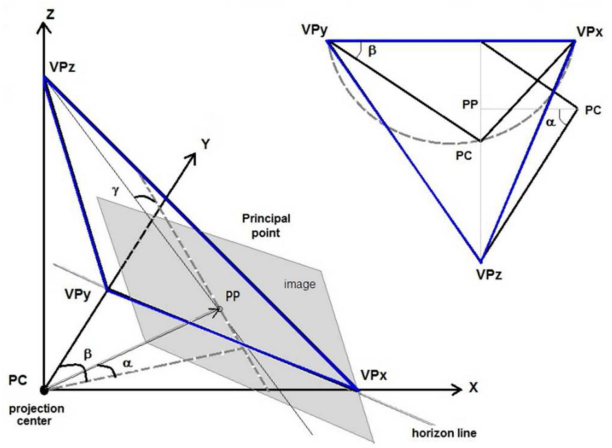

Fig. 6 *Camera parameters defined in a single image (adapted from Aguilera)*

Canny and Burns operators are used for line segment extraction with high accuracy and reliability. Fig. 5 illustrates the result of line segmentation and clustering process on the selected single image in this study.

Segments are clustered by analysing the slope and distance between them in order to satisfy the orientation constraint. At this stage, radial distortion is also provided. Robust clustering of segments is performed according to the three VPs through a robust estimator (RANSAC) to detect possible outliers (i.e. wrong vanishing lines). After the vanishing lines are extracted and clustered, a robust VP computation is performed. Owing to the presence of mini-segments and weakness of the geometry unfavourable intersection conditions for VP computation might occur. In such cases, since the least squares method is not a good choice to check possible inaccurate observations (e.g. false vanishing lines), robust estimators such as M-estimators have been used in the method. Method II uses a twofold approach for a precise VP computation, an estimation step which inherits some of the analytical concepts of the traditional Gauss sphere method combined with RANSAC and followed by a computation step which applies a reweighted least squares adjustment supported by a modified Danish estimator [24, 37]. A hybrid self-calibration method is used for the single image calibration combining some approaches related to close range photogrammetry and computer vision. VP geometry is used to directly recover projection matrices by analysing the simple properties of VPs while adding geometric constraints derived from the image. A complete camera model is recovered in a consecutive fashion, i.e. internal and external parameters are separately estimated. In the first step, the intrinsic parameters (i.e. the focal length, the PP, and the lens distortion) are recovered automatically based on VPs geometry and image analysis. The (PP of the camera is identified by the orthocentre of the triangle formed from the three VPs of the three mutually orthogonal directions through the vectoral cross product of the segments of the triangle and its heights (Fig. 6)

The focal length can be computed using the distances from the PP to any of the triangle's vertices and the opposite side as in the following form:

$$f = \sqrt{|(x_{VP} - x_{PP})(y_{VP} - x_{PP}) + (x_{VP} - y_{PP})(y_{VP} - y_{PP})|}. \quad (5)$$

The radial lens distortion parameters (k_1, k_2) are estimated by exploiting the presence of short segments that satisfy the orientation constraint through the collinearity condition. In the next step, the extrinsic parameters (i.e. the perspective rotation matrix and the translation vector) are estimated in a two-pass process [16, 22]. First, the rotation matrix (camera orientation) is obtained automatically based on the correspondence between the VPs and the three main object directions. This relationship allows the cosine vectors of the optical axis to be extracted, obtaining the three angles (α, β, γ) directly. These rotation angles (α, β, γ) are called the 'axis, tilt and swing', respectively, identify the rotation of object-space to image-space. The *axis* is a clockwise rotation of the nadir direction (Z). The *tilt* angle is a rotation of a line parallel to the true horizon line. The *swing* angle is a rotation about the image z-axis (Fig. 6). Then, the translation vector $\boldsymbol{x}$ (i.e. the absolute camera pose) can be estimated based on some a priori object information (e.g. a distance combined with a geometric constraint defined by the user). The well-known Gauss–Markov model of least squares is normally used for the estimation of $\boldsymbol{x}$

$$\boldsymbol{V} = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{l}, \quad (6)$$

(see equation below) where r_{ij} are (3 × 3) rotation matrix elements, f is focal length, (x_{PP}, y_{PP}) is the PP coordinates, and (X_i, Y_i, Z_i) represent a priori known information about the object. The object coordinates of a point are simply recovered by means of the collinearity model.

Once the camera model has been estimated accurately, a dimensional analysis of the object based on distances, angles and areas can be performed. Therefore, the entire metric measurement problem could be reduced to the problem of computing the coordinates of the object through the collinearity condition

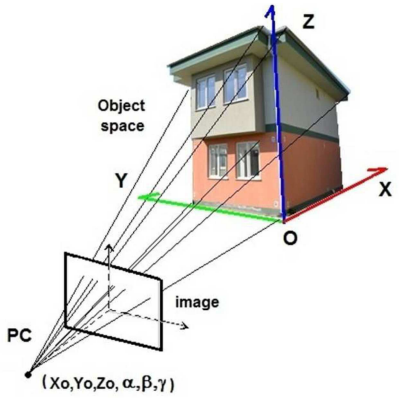

Fig. 7 *Metric measurement concept from a single image through collinearity condition*

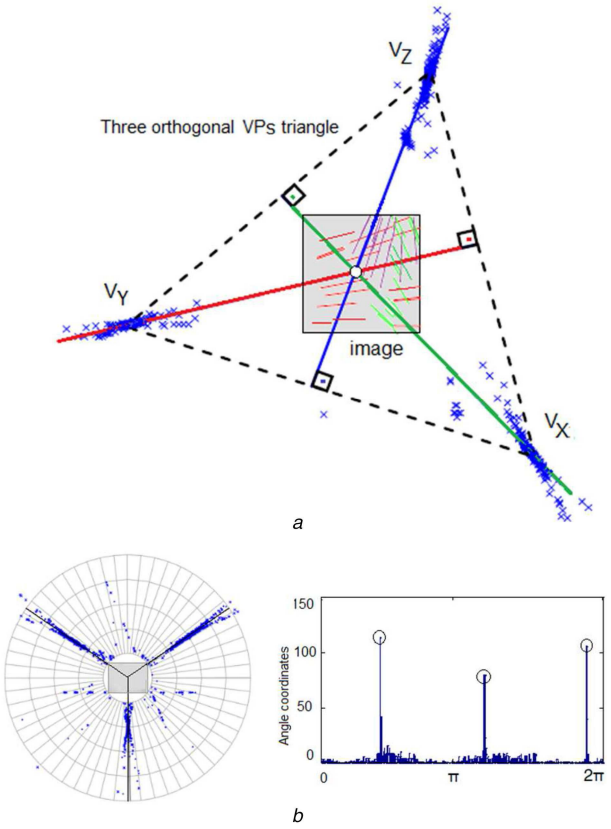

Fig. 8 *Estimation of three orthogonal VPs*
(a) *Image of converging lines and three orthogonal VPs triangle,*
(b) Polar coordinate system referenced to the image centre and histogram of angle coordinates of intersections (adapted from Li *et al.* [38])

$$X = X_{\mathrm{o}} + (Z - Z_{\mathrm{o}})\frac{r_{11}(x - x_{\mathrm{PP}}) + r_{12}\,(y - y_{\mathrm{PP}}) - r_{13}f}{r_{31}(x - x_{\mathrm{PP}}) + r_{32}\,(y - y_{\mathrm{PP}}) - r_{33}f}, \tag{7}$$

$$Y = Y_{\mathrm{o}} + (Z - Z_{\mathrm{o}})\frac{r_{21}(x - x_{\mathrm{PP}}) + r_{22}\,(y - y_{\mathrm{PP}}) - r_{23}f}{r_{31}(x - x_{\mathrm{PP}}) + r_{32}\,(y - y_{\mathrm{PP}}) - r_{33}f}. \tag{8}$$

The collinearities (7) and (8) relate image measurements to the object coordinate system (X, Y, Z) through the camera constant f, the image coordinates (x, y), the PP (x_{PP}, y_{PP}), the projective centre (X_0, Y_0, Z_0), and the rotation matrix $\boldsymbol{R}$ (r_{ij}) that serves the misalignment of both reference frames. It should be noted that geometric constraints (perpendicularity, coplanarity etc.) or image invariants (e.g. distances) are needed to provide metric information from a single image (Fig. 7).

3 Estimating VPs in a single image

The problem of locating VPs in 2D perspective projection images is relatively old and has been studied for some time to time. This task is generally assessed as the computation of line intersections, however by virtue of quantisation errors, lines that correspond to a single VP intersect inside an area termed as vanishing region. To tackle the vanishing region problem, the methods of the VP detection can be divided into three main stages: detection of line segments in the image, clustering of these segments in groups that converge to a VP, and VP estimation for each cluster. As previously mentioned a reliable and accurate 3D object reconstruction depends strongly on a robust estimation of VPs. VPs can be precisely computed from the bundles of perspective lines using the theoretical framework of the projective geometry. Since a set of parallel lines converge to an intersection point (i.e. VP) in the scene, VPs are conventionally estimated from intersections of hypothetic perspective lines or by minimising the distance to projections of parallel lines [1]. There are many tools available in the related fields for detection of VPs. A simple Matlab code developed by Li *et al.* [38] was employed for detection of VPs in this study. The method uses an accumulator cell to collect lines passing through the corresponding image point. Peaks in the accumulator space represent the potential VPs. The method accurately and simultaneously estimates the three orthogonal VPs and focal length from single images. The distribution of intersections of converging lines corresponding to one VP in a polar coordinate system with the origin at the image centre is considered. Fig. 8*a* shows an image with line segments converging to three orthogonal VPs. Lines corresponding to different VPs are shown in different colours. Blue crosses represent intersections of all pairs of detected lines.

The orthocentre of the three orthogonal VPs triangle coincides with the image centre, which is considered as the origin of the polar coordinate system (Fig. 8*b*). Therefore, if angle coordinates of the VPs are detected, altitudes of the triangle could be obtained. Angle coordinates of true VPs could be obtained by seeking peaks in the histogram of angle coordinates of intersections. Fig. 8*b* shows the histogram of angle coordinates of intersections of all pairs of lines in Fig. 8*a*. Three significant peaks correspond to angle coordinates of the three VPs in Fig. 8*a*. In the method, the altitudes of the triangle are detected firstly. Some novel constraints on both VPs and focal length are obtained from these three altitudes. Then constraints from these altitudes are used to detect radial coordinates of the three VPs and focal length simultaneously. The detected altitudes of the triangle are shown as coloured solid lines in Fig. 8*a*. The three peaks correspond to the three orthogonal VPs in Fig. 8*b*.The method decomposes a 2D Hough parameter space into two cascaded 1D parameter spaces which makes the method more robust and faster than the previous methods without losing accuracy [38].

It is clear that since the constraints are used as an additional criterion to reduce errors, the flexibility of the method relies on these constraints provided by the three orthogonal VPs. In this work, it was aimed to make distance measurements from a single image taken by a NIKON D3100 digital camera (focal length: 18.5512 mm, format size: w: 24.0174 h: 16.00) in our university campus environment. Fig. 9 denotes the detected three VPs in the image using the Matlab tool. Intersections of detected lines in the image are marked by blue crosses and detected VPs are marked by black squares in this figure.

$$\boldsymbol{A} = \begin{bmatrix} r_{31}(x_i - x_{\mathrm{PP}}) + r_{11}f & r_{32}(x_i - x_{\mathrm{PP}}) + r_{12}f & r_{33}(x_i - x_{\mathrm{PP}}) + r_{13}f \\ r_{31}(y_i - y_{\mathrm{PP}}) + r_{21}f & r_{32}(y_i - y_{\mathrm{PP}}) + r_{22}f & r_{33}(y_i - y_{\mathrm{PP}}) + r_{23}f \end{bmatrix},$$

$$\boldsymbol{l} = \begin{bmatrix} (x_i - x_{\mathrm{PP}}) + (r_{31}X_i + r_{32}Y_i + r_{33}Z_i) + f(r_{11}X_i + r_{12}Y_i + r_{13}Z_i) \\ (y_i - y_{\mathrm{PP}}) + (r_{31}X_i + r_{32}Y_i + r_{33}Z_i) + f(r_{21}X_i + r_{22}Y_i + r_{23}Z_i) \end{bmatrix},$$

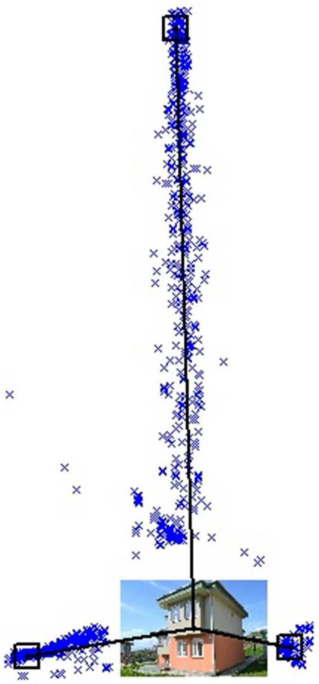

Fig. 9 *VPs detected in the single image (for method I)*

Table 1 VPs computed for method I

	x, mm	***y***, mm
v_x	2661.300092	1071.448513
v_y	−1462.579757	1198.489030
v_z	816.111462	−10,135.324495

VP coordinates calculated from the tool are given in Table 1.

After computing the VPs, the scale factor can be calculated from (4) and finally object distances will directly be measured as explained in the previous sections.

4 Results and discussion

The experiment was arranged as shown in Fig. 10. Two vertical reference distances and six different measurement distances (d1, d2,..., d6) marked on the wall of the building were used in the experiments. These distances were selected as aligned by both horizontally and vertically so as to form a perspective projection. To compare the distance variations or accuracies of the two methods, all the distances were measured in single images using both the VPs based method (method I) and robust VPs and camera self-calibration based method (method II). Experiments were conducted to evaluate the accuracy of the two methods used for getting metric information from single images. The reference distances with 2.00 m length were marked on the walls of the building to retrieve the projection matrix according to the method using target points as illustrated in the figure. The number of single images used in the experiments was 12 in the study. Shown in Fig. 10 is one of the images showing the building face on which some control measurements were performed. The lines along the building face/corners were used to find VPs.

The actual (or reference) distances were carefully measured by a precise measuring tape. The summary of descriptive statistical results of the distances computed using the two methods is given in Tables 2 and 3. It can easily be seen that the accuracy of the distances from method II is considerably higher than method I. The average standard deviations of the distances computed from method I and method II are ±0.018 and ±0.005 m, respectively. When the 'shape' of the distributions of two datasets is examined from Tables 2 and 3 asymmetrical distributions can be observed. Especially the dataset including distance measures obtained from method I have strong skewness values. The dataset from method II is more likely to yield a histogram that reflects the normal distribution as it has smaller skewness values.

The 'coefficient of variation' parameters are examined it can easily be seen that the degree of variation of the dataset from method II is smaller than the dataset of method I for all distance measures. This result is consistent with the results obtained from the standard deviation measure. Generally, it is possible to say that the distances were more accurately estimated when using method II. To observe the measured distance variations, the computed distances in all images are plotted on the graphs in Figs. 11*a* and *b*.

Since the mean value of the distances is different (i.e. 3.00 and 2.00 m) two separate graphs were created to present relative differences of the distances. Figs. 11*a* and *b* show the measured distance variations of six different marked lengths on the walls estimated by using the two methods. It is clear that the results of differences from the actual distance value computed from method II which is shown on the left part of the figure are smaller than method I. Standard deviation values of distance measurements also prove this result.

Note that these results mainly obtained from method I included no camera calibration process. When method II including robust camera calibration was used, the distances were estimated more accurately.

For the purpose of comparison of accuracies, box plot graphs are also plotted for studying distributional characteristics of the dataset (Figs. 12*a* and *b*). As it can easily be seen the boxplot of method I have generally larger errors in comparison with method II containing larger distance variations. Standard errors of the mean for the second method are smaller than method I since the range of measurement variations is small. More specifically the distribution of measurements estimated from method I showed asymmetrical shape (i.e. skewed distribution) showing different patterns as illustrated in Fig. 12*a*.

When the two methods were compared in terms of accuracy and reliability, the following points were determined: it is worthy to note that both the methods are only effective with images of man-made environments, in which it is possible to extract image invariants (e.g. distances and straight segments) corresponding to different 3D orientations [15, 22, 33]. For almost all single image modelling approaches the ability to accurately determine the VP and vanishing line directly affect the result and accuracy of the distance measurement. From this point of view, method II seems to have advantageous as it has developed a robust VP computation. Method I is based on single image metrology that uses some assumptions such as 'three orthogonal sets of parallel lines, a priori known points on the ground plane and one height' in the object space. The advantages of the method might be outlined as (i) it is possible to make diverse starting assumptions about the space in the image in a mathematical fashion, (ii) the inconsistencies of the spatial representation can be limited using information available directly from the image itself, (iii) the colour information and texture of the original image are preserved, (iv) errors can be detected by reprojection and then eliminated. The disadvantages are (i) it is applicable only to those objects which were constructed with strong perspective rules, (ii) the quality of resolution is closely related to the quality of the source image, (iii) it is not practical to identify corresponding points between planes manually. In general, uncertainty in the method might be assumed in the reference heights, vertical VP and plane vanishing line. It is plausible to think that as the number of reference distances is increased, the uncertainty of the measured distance will decrease. The advantages of method II are briefly summed as below: (i) In method II, it can be noticed that some relevant tasks have been automated or partially automated relative to the other method, (ii) original strategies and algorithms integrating robust estimators have been adapted in each step guaranteeing more precision and reliability, (iii) the uncertainty of the measurements has been taken into account and a final statistical analysis is performed based on error propagation, (iv) the developed approach is also based on line segment extraction but incorporates clustering strategies together with robust estimators, (v) a robust VP computation is applied to guarantee more precision and reliability. The disadvantages of the method might be summarised as below: (i) the method is

Fig. 10 *Target points and marked distance measurements on the single image*

Table 2 Statistical summary results of measurements obtained from method I

	d1	d2	d3	d5	d4	d6
min	2.9870	2.9700	2.9680	2.9700	1.9810	1.9680
max	3.0410	3.0250	3.0250	3.0230	2.0340	2.0250
mean	3.0176	2.9964	3.0017	2.9971	2.0074	2.0017
variance	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
std. dev	0.0179	0.0181	0.0181	0.0180	0.0180	0.0181
skewness	−0.7089	0.3514	−0.6617	0.1703	0.1331	−0.6617
kurtosis	−0.6617	−1.1041	−0.7458	−1.3923	−1.4348	−0.7458
coeff. var	0.5936	0.6028	0.6041	0.6002	0.8978	0.9059

Table 3 Statistical summary results of measurements obtained from method II

	d1	d2	d3	d5	d4	d6
min	2.9920	2.9900	2.9900	2.9900	1.9910	1.9910
max	3.0080	3.0080	3.0080	3.0070	2.0080	2.0050
mean	2.9989	2.9998	2.9988	2.9985	2.0003	1.9988
variance	0.00003	0.00004	0.00002	0.00002	0.00003	0.00002
std. dev	0.0052	0.0060	0.0050	0.0048	0.0053	0.0045
skewness	0.2050	−0.3758	0.0332	0.2218	−0.3764	−0.3794
kurtosis	−0.9145	−0.6403	0.0531	0.0382	−0.9256	−0.6819
coeff. var	0.1721	0.1996	0.1651	0.1587	0.2672	0.2239

applicable in scenes with strong geometric contents (i.e. presence of structural planes and lines), (ii) the images acquired have to be oblique with three VPs (i.e. no more VPs should exist), (iii) it is necessary to have some a priori information about the object (e.g. a distance and parallelism of object edges) to overcome the ''ill-conditioning'' problem.

To say more generally experimental results show that method II provided higher accuracies than the previous method (i.e. method I). At this point, it is plausible to think that VP determination seems to have more importance than the other relevant factors (i.e. reference points, geometry primitives etc.) for measurement accuracy of single images. It is obvious that any slight inaccuracy in measuring VPs will result in large errors. Small variations in feature geometry and some other uncertainties might yield a numerically unstable or ill-conditioned structure for the subject. The other point is on camera calibration which is still an essential process for defining 3D viewing geometry for the related problems. The inclusion of the calibration process which is performed more robustly in method II might have contributed an increased accuracy. It should be emphasised that extracting metric or 3D information from a single view is still a process that offers an intrinsic complexity. Hence, it is notable to say that the findings of this study could be only valid with the inclusion of some strong geometric features in the image, i.e. regular structures, perspective rules, planes or lines. This kind of information should be known a priori. Images have to be geometrically appropriate for the well detection of VPs in the image (as seen in Fig. 4).

When the practical applications are considered the following issues can be stated for the two methods in terms of the achieved accuracy values for geometrical information recovery. First of all, it should be noted that these two methods do not have as high accuracy values (i.e. with a precision of mm levels) as stereovision techniques for metric measurement purposes. In some applications where very high accuracy values are not desired, it is possible to use these two methods provided that there are some geometric image invariants in scenes mentioned previously. Method II can be used effectively in civil engineering, cultural heritage conservation, criminology and urban planning applications when the accuracy values obtained in the study are taken into account. This method can be used for getting metric measurements from the images aiming digital archiving of construction structures such as bridges, viaducts with a centimetre (or below) level accuracy. As a technical assistance, it might also be utilised for the documentation, virtual reconstruction and digital preservation of cultural heritage objects with an acceptable precision. Method I which is simpler to use is applicable for the applications where accuracy requirements are lower than Method II. 3D graphical modelling, forensic investigations (human height determination, accident analysis etc.) and virtual modelling are some of the applications that are possible to employ with Method I with a few centimetres accuracy. Estimating human height is one of the popular tasks used in video surveillance and the method seems to be practical for this application.

5 Conclusion

Single image metrology algorithms based on VPs and lines are presented to assess the accuracy of distance measurements from single photographs. An accuracy analysis for the distance determination from single perspective images was included to estimate uncertainties of measurements in this work. One of the most crucial steps in this process concerns identifying VPs with a

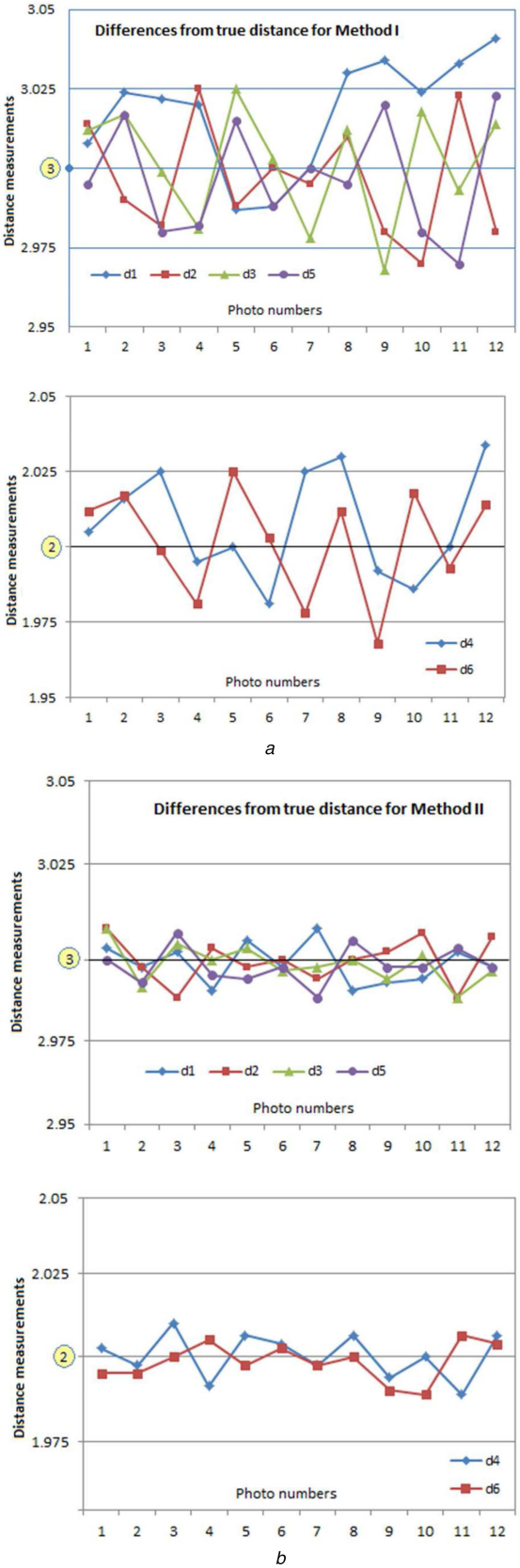

Fig. 11 *Distance measurements obtained from the single images*
(a) Distribution of relative distance estimated for method I,
(b) Distribution of relative distance estimated for method II

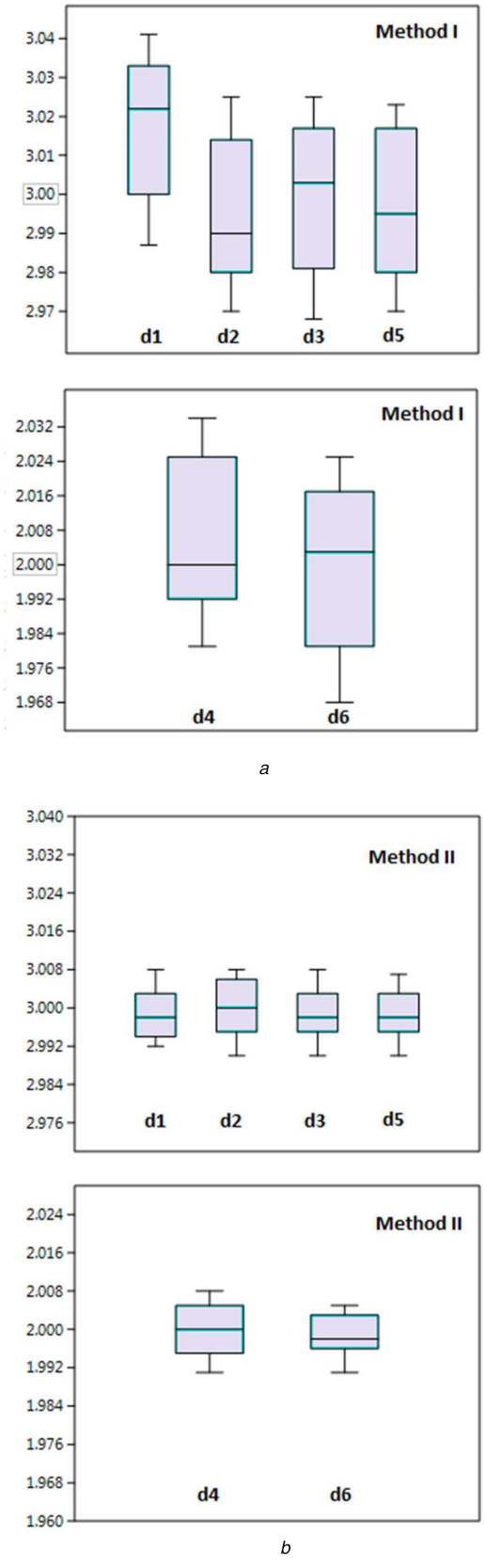

Fig. 12 *Distance measurement variations for accuracy comparison of the two methods*
(a) Box plot of the relative distance variations for method I,
(b) Box plot of the relative distance variations for method II

high precision belonging to a raw or un-oriented image. The two methods were employed to measure the distances of a reference wall in an outdoor environment. Experimental results show that the method including robust camera calibration and statistical estimators provided higher accuracy than the other method which is based on the cross ratio. To conclude careful attention should be given on estimating the VPs since any inaccuracy on detection of VPs can cause large uncertainties for the object size measurement. The method would be applicable in scenes with geometrically appropriate conditions. Considering that some research studies on the configuration and accuracy of metric measurements from single images taken by smartphone devices have begun, a multi-

disciplinary approach integrating the other disciplines will be needed for the automation of single view measurements.

6 References

[1] Arslan, O.: '3D object reconstruction from a single image', *Int. J. Environ. Geoinformatics*, 2014, **1**, pp. 20–29

[2] Debevec, P.E., Taylor, C.J., Malik, J.: 'Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach'. Proc. SIGGRAPH'96, 1996, pp. 11–20

[3] El-Hakim, S.F.: 'A practical approach to creating precise and detailed 3D models from single and multiple views', *Int. Arch. Photogramm. Remote Sens.*, 2000, **33**, (5), pp. 202–209

[4] Criminisi, A., Reid, I., Zisserman, A.: 'Single view metrology', *Int. J. Comput. Vis.*, 2000, **40**, pp. 123–148

[5] Pears, N., Liang, B., Chen, Z.: 'Mobile robot visual navigation using multiple features', *J. Appl. Signal Process.*, 2005, **14**, pp. 2250–2259

[6] Chen, Z., Pears, N., Liang, B.: 'A method of visual metrology from uncalibrated images', *Pattern Recognit. Lett.*, 2006, **27**, (13), pp. 1447–1456

[7] Cai, J., Walker, R.: 'Height estimation from monocular image sequences using dynamic programming with explicit occlusions', *IET Comput. Vis.*, 2010, **4**, (3), pp. 149–161

[8] BenAbdelkader, C., Yacoob, Y.: 'Statistical body height estimation from a single image'. Proc. 8th IEEE Int. Conf. on Automatic Face and Gesture Recognition, 2008, pp. 1–7

[9] Jeges, E., Kispal, I., Hornak, Z.: 'Measuring human height using calibrated cameras'. Proc. 2008 Conf. on Human Systems Interactions, 2008, pp. 755–760

[10] Fernanda, A.A., Taubin, G., Goldenstein, S.: 'Efficient height measurements in single images based on the detection of vanishing points', *Comput. Vis. Image Underst.*, 2015, **138**, pp. 51–60

[11] Guan, Y.P.: 'Unsupervised human height estimation from a single image', *J. Biomed. Sci. Eng.*, 2009, **2**, (6), pp. 425–430

[12] Zhou, J., Li, B.: 'Exploiting vertical lines in vision-based navigation for mobile robot platforms'. Proc. IEEE Conf. on Acoustics, Speech and Signal Processing, 2007, pp. 465–468

[13] Chen, S., Fang, X., Shen, J., *et al.*: 'Single-image distance measurement by a smart mobile device' *IEEE Trans. Cybern.*, 2017, **47**, (12), pp. 4451–4462

[14] Holzmann, C., Hochgatterer, M.: 'Measuring distance with mobile phones using single-camera stereo vision'. Proc. Int. Conf. on Distributed Computing Systems Workshops (ICDCSW), 2012, pp. 88–93

[15] Aguilera, D.G., Gomez-Lahoz, J.: 'From 2D to 3D through modelling based on a single image'. *Photogramm. Rec.*, 2008, **23**, (122), pp. 208–227

[16] Aguilera, D.G., Gomez-Lahoz, J.: 'Forensic terrestrial photogrammetry from a single image', *J. Forensic Sci.*, 2009, **54**, (6), pp. 1376–1387

[17] Madden, C., Piccardi, M.: 'Height measurement as a session-based biometric for people matching across disjoint camera views'. Proc. Conf. on Image and Vision Computing, New Zealand (IVCNZ), 2005, pp. 282–286

[18] Khan, S., Shah, M.: 'A multiview approach to tracking people in crowded scenes using a planar homography constraint'. Proc. 9th European Conf. on Computer Vision (ECCV), 2006, pp. 133–146

[19] Criminisi, A.: 'Single-view metrology: algorithms and applications'. Proc. 24th DAGM Symp. on Pattern Recognition, 2002, pp. 224–239

[20] Park, S.W., Kim, T.E., Choi, J.S.: 'Robust estimation of heights of moving people using a single camera'. Proc. Int. Conf. on IT Convergence and Security (ICITCS), 2011, pp. 389–405

[21] Lee, K.Z.: 'A simple calibration approach to single view height estimation'. Proc. Ninth Conf. on Computer and Robot Vision (CRV), 2012, pp. 161–166

[22] Aguilera, D.G., Gómez-Lahoz, J.: 'Dimensional analysis of bridges from a single image'. *J. Comput. Civ. Eng.*, 2009, **23**, pp. 319–329

[23] Arslan, O.: 'Metric measurements from a single image with the estimation of vanishing points'. Int. Scientific Conf. on Applied Sciences, Antalya Turkey, 27–30 September 2016

[24] Aguilera, D.G., Gómez, J., Finat, J.: 'A new method for vanishing points detection in 3D reconstruction from a single view', ISPRS Commission V, WG V/2. Mestre, Venice, 2005

[25] Lutton, E., Maitre, H., Lopez-Krahe, J.: 'Contribution to the determination of vanishing points using Hough transform', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1994, **16**, (4), pp. 430–438

[26] Shufelt, J.A.: 'Performance evaluation and analysis of vanishing point detection techniques'. ARPA Image Understanding Workshop, Morgan Kaufmann Publishers, Palm Springs, 1996, pp. 1113–1132

[27] Košecká, J., Zhang, W.: 'Video compass'. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds) '*Computer vision – ECCV 2002, lecture notes in computer science*' (Springer, Berlin, Heidelberg, 2002), vol. **2353**, pp. 476–490

[28] Tardif, J.P.: 'Non-iterative approach for fast and accurate vanishing point detection'. Int. Conf. on Computer Vision, 2009

[29] Xu, Y., Oh, S., Hoogs, A.: 'A minimum error vanishing point detection approach for uncalibrated monocular images of man-made environments'. Computer Vision and Pattern Recognition, 2013

[30] Wildenauer, H., Hanbury, A.: 'Robust camera self-calibration from monocular images of Manhattan worlds'. Computer Vision and Pattern Recognition, 2012

[31] Tretyak, E., Barinova, O., Kohli, P., *et al.*: 'Geometric image parsing in man-made environments', *Int. J. Comput. Vis.*, 2012, **97**, (3), pp. 305–321

[32] Lezama, J., Grompone Von Gioi, R., Randall, G., *et al.*: 'Finding vanishing points via point alignments in image primal and dual domains'. Computer Vision and Pattern Recognition, 2014

[33] Criminisi, A., Zisserman, A., Van Gool, L., *et al.*: 'A new approach to obtain height measurements from video'. Proc. SPIE, Boston, Massachusetts, USA, November 1998, vol. 3576

[34] Heuvel, F.A.: 'Vanishing point detection for architectural photogrammetry', *Int. Arch. Photogramm. Remote Sens.*, 1998, **XXXII**, part 5, pp. 652–659

[35] Wang, G., Tsui, H., Wu, Q.: 'What can we learn about the scene structure from three orthogonal vanishing points in images', *Pattern Recognit. Lett.*, 2009, **30**, (3), pp. 192–202

[36] Criminisi, A.: '*Accurate visual metrology from single and multiple uncalibrated images*' (Springer-Verlag, New York, USA, 2001)

[37] Aguilera, D.G., Gomez Lahoz, J., Finat, J., *et al.*: 'A comparative analysis for the vanishing points estimation in urban scenes from a single view'. CIPA Int. Workshop: Vision Techniques Applied to the Rehabilitation of City Centres, Lisbon, Portugal, 2004

[38] Li, B., Peng, K., Ying, X., *et al.*: 'Simultaneous vanishing point detection and camera calibration from single images'. Proc. Int. Symp. on Visual Computing, Part II (LNCS, **6454**), 2010, pp. 151–160