# Single-Image Distance Measurement by a Smart Mobile Device

Shangwen Chen, Xianyong Fang, Jianbing Shen, *Senior Member, IEEE*, Linbo Wang, and Ling Shao, *Senior Member, IEEE*

***Abstract*—Existing distance measurement methods either require multiple images and special photographing poses or only measure the height with a special view configuration. We propose a novel image-based method that can measure various types of distance from single image captured by a smart mobile device. The embedded accelerometer is used to determine the view orientation of the device. Consequently, pixels can be back-projected to the ground, thanks to the efficient calibration method using two known distances. Then the distance in pixel is transformed to a real distance in centimeter with a linear model parameterized by the magnification ratio. Various types of distance specified in the image can be computed accordingly. Experimental results demonstrate the effectiveness of the proposed method.**

***Index Terms*—Accelerometer, distance measurement, single image, smart mobile device.**

## I. Introduction

DISTANCE measurement, e.g., measuring the distance on the ground or the height of objects, is very common in our daily life. While the methods using classical tools such as ruler, laser [1] and depth camera [2], [3] can be inconvenient or expensive, and image-based distance measurement methods [4]–[11], [19], [24] only require simple photographing and thus are cheap and easy to apply. In this paper, we focus on measuring various types of distance from a single image with a smart mobile device.

Measurement methods based on the principles of stereo vision rely on consumer cameras to fulfill the distance measurement [12], [13], [25], [27], [28]. Some studies [14], [15] capture images by two or more cameras aligned in a fixed position. Gao *et al.* [16] changed the focal length of the camera to capture two images for the depth estimation. Kim *et al.* [17] computed the distance with a sequence of reflected images obtained by a camera installed in front of a rotating mirror. These multi-image-based methods require fixed camera locations or to know the locations explicitly. Wu *et al.* [18] estimated front-vehicle distance by a mounted camera with a trained functional neuro-fuzzy network. Meanwhile, single-image-based methods with special constraints on the scene have also been introduced. One of them [20] uses two concentric or parallel circles to determine coplanar distances. Rahman *et al.* [21] proposed a learning-based approach to estimate the person-to-camera distance by the eye-distance statistics. All the above approaches aim at the personal computer platform and most of them only measure depth or height.

Recently, several studies have started to utilize smart mobile devices which are embedded with cameras and sensors for distance measurement. One example is the depth perception study of Holzmann and Hochgatterer [22], where two related images are captured to build a stereo vision system and displacement is obtained with the inner-sensor signals. Laotrakunchai *et al.* [23] simultaneously measured depth and object size in a similar way as Holzmann and Hochgatterer [22]. Those stereo vision-based methods require special photographing poses to capture two images. Han and Wang [26] measured tree-height from a single image through direct scaling the length by a parallel bench marking. Apparently, their method does not take the full advantages of a smart mobile device, such as the accelerometer which may be adopted to determine view orientation of the device. Most existing mobile device-based studies focus only on the depth and height measurement. We propose a new single-image-based approach for measuring more types of distance without the requirement of a special view or bench marking in existing methods. Our method just requires one known distance which can be any known length anywhere in the captured scene, and it can be prepared easily in advance (e.g., using any off-the-shelf ruler-like tool) in the case that it does not already have a known reference distance in the scene. Therefore, the proposed method is very convenient to apply in practice.

The proposed method utilizes the accelerometer which is the standard configuration of today's smart mobile devices

Manuscript received April 28, 2016; revised July 2, 2016; accepted September 15, 2016. Date of publication September 29, 2016; date of current version November 15, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61301295 and Grant 61502005, in part by the Natural Science Foundation of Anhui Province under Grant 1608085QF129 and Grant 1408085MF113, in part by the Key Program for Science and Technology Development of Anhui Province under Grant 1604d0802004, and in part by the Specialized Fund for Joint Building Program of Beijing Municipal Education Commission. This paper was recommended by Associate Editor L. Wang.

S. Chen, X. Fang, and L. Wang are with the Key Laboratory of Intelligent Computing and Signal Processing of MOE and the Institute of Media Computing, Anhui University, Hefei 230601, China (e-mail: chensw@ahu.edu.cn; fangxianyong@ahu.edu.cn; wanglb.2005@gmail.com).

J. Shen is with the Beijing Laboratory of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing 100081, China (e-mail: shenjianbing@bit.edu.cn).

L. Shao is with the Department of Computer Science and Digital Technologies, Northumbria University, Newcastle upon Tyne, NE1 8ST, U.K. (e-mail: ling.shao@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2016.2611599

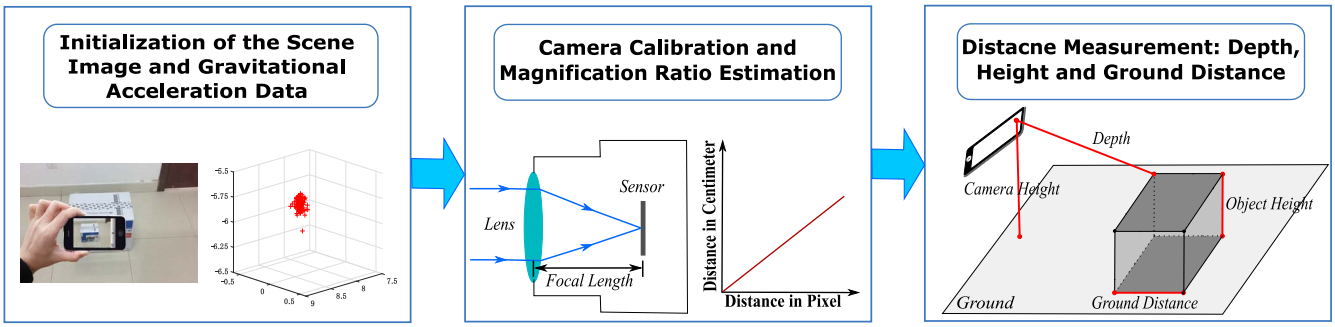



Fig. 1. Pipeline of the proposed method. (a) Scene is captured by a hand-held mobile device to obtain the acceleration data. (b) Camera is calibrated with the view orientation to estimate the magnification ratio. (c) Different types of distance (e.g., depth, height, and ground distance) are measured with the ratio.

to obtain the view direction of the device, and consequently, builds the geometric relationship between pixels in the image plane and their corresponding points on the ground. The distance between two points in pixel is converted into the real distance with a linear model parameterized by the magnification ratio. Three types of distance can be obtained subsequently: ground distance, depth, and height. Ground distance means the distance between two user-specified points on the ground. Depth is the distance from the camera optical center to the user-specified point in 3-D space. Height is the distance of two user-specified points representing the two ends of a line perpendicular to the ground. Other types of distance can also be measured with properly selected ground pixels based on the three types of distance. Our source code will be available at http://github.com/shenjianbing/distancemeasure.

The main contributions can be summarized as follows.

1) It is the first smart mobile device oriented single-image-based approach for distance measurement, which utilizes the embedded camera and accelerometer for photographing and recovering the scene geometry.
2) A new calibration method based on two known distances is proposed to obtain an accurate focal length, which is simpler and more efficient in comparison with the existing calibration methods.
3) Various types of distances including ground distance, depth, and height can be measured, while the existing methods can only handle one or two types.

## II. Proposed Method

Our method uses a smart mobile device configured with a camera and an accelerometer. The camera provides us a convenient way to capture the scene image. The accelerometer provides gravitational acceleration data to obtain the view orientation of the device. This direction helps back projecting the specified image pixels to the ground so that distance measurement can be fulfilled. Fig. 1 shows the pipeline of the proposed method. First, the image of the target scene and the corresponding gravitational acceleration data during photographing are prepared. The data are denoised to counter the camera jitter for better accuracy. Then, the camera is calibrated to perform back-projection operation and the magnification ratio which represents the ratio between the pixel distance and the real-world distance is computed in a linear model. Finally, different types of distance, e.g., ground distance, depth, and height, can be measured with the ratio by back projection.

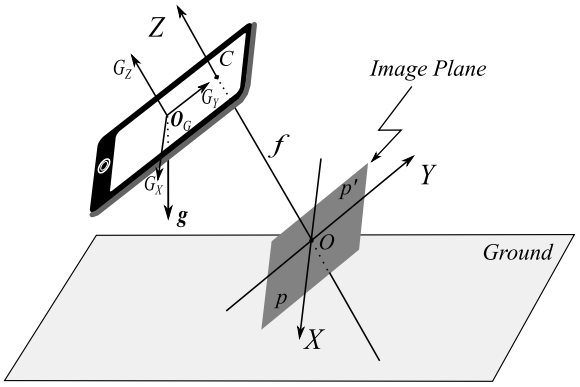



Fig. 2. Coordinate systems.

### A. Coordinate Systems

Let us first introduce the coordinate systems used in our method (Fig. 2). $\boldsymbol{OXYZ}$ is the image coordinate system of the device with the focal length $f$ being the distance between the camera optical center $\boldsymbol{C}(0, 0, f)$ and the origin $\boldsymbol{O}(0, 0, 0)$. For the gravitational acceleration data obtained by the accelerometer, each acceleration datum is a vector consisting of three orthogonally projected elements. Correspondingly, we can translate the image coordinate system to the device and then obtain the acceleration coordinate system $\boldsymbol{O_G G_X G_Y G_Z}$. $G_X$, $G_Y$, and $G_Z$ are axes corresponding to the three acceleration dimensions with $\boldsymbol{O_G}$ being the origin on the device. $\boldsymbol{g}$ in Fig. 2 denotes the gravitation, which is also the norm of ground. Since the accelerometer records the acceleration of the gravitation, for simplicity, we also denote the acceleration data by $\boldsymbol{g}$ in the following paragraphs.

### B. Initialization

We now discuss the initialization process. The scene image including the target distance to measure is captured with a hand-held smart mobile device. The accelerometer records

the gravitational acceleration at each spatial position when photographing the target scene. Each acceleration datum is a three-element vector representing the projection of the gravitation in $G_X$, $G_Y$, and $G_Z$ axes as Fig. 2 shows. Considering the gravitation is vertical to the ground, we can conclude that the view orientation of the device can be computed directly. For the distance measurement, this direction is required for fulfilling the back-projection which helps mapping the pixel distance to the real distance. Therefore, we record the gravitational acceleration when capturing the scene image.

However, the recorded acceleration data are highly unstable with heavy noises due to the unavoidably continuous camera jitter, even when we try our best to hold the device still. Fig. 3(a) shows ten such noisy example acceleration data sets recorded from ten captures. Each capture lasts for one second and is depicted in a different color. In this figure, $g_x$, $g_y$, and $g_z$ denote the three elements of each data in $G_X$, $G_Y$, and $G_Z$ dimensions, respectively. The dots in each set represent the recorded accelerations during the image capture. We can see the noticeable shaky movement of the device during capturing. Apparently, such noisy data cannot be directly used for view direction computation.

We propose a weighted average method to denoise the noisy sensor data. In this method, for a recorded acceleration data set $\boldsymbol{g}(t)$, $t \in [\tau - (T/2), \tau + (T/2)]$ at the exposure time $\tau$, the denoised data $\bar{\boldsymbol{g}}$ can be computed by weighted averaging all the data during the sample time $T$

$$\bar{\boldsymbol{g}} = \frac{1}{\sum_{\tau-\frac{T}{2}}^{\tau+\frac{T}{2}} w(t)} \sum_{t=\tau-\frac{T}{2}}^{\tau+\frac{T}{2}} w(t)\boldsymbol{g}(t)$$
$$w(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(t-\tau)^2}{2\sigma^2}\right\} \tag{1}$$

where $w(t)$ in (1) is the weight formulated by the Gaussian function, and $\sigma$ is the standard deviation. For the noisy data shown in Fig. 3(a), where $T = 1$ s, after setting $\sigma = 0.2$, we can obtain ten denoised acceleration data for the ten noisy sets, respectively, as shown in Fig. 3(b). It can be seen that the originally messy data are now condensed into a single datum. The view orientation of the device can then be easily computed with such a datum, which is the basis of back projection for computing the magnification ratio.

### C. Camera Calibration and Magnification Ratio Estimation

Magnification ratio transforms the distance in pixel to the real distance in centimeter. To compute the magnification ratio $e$, the pixels specified by the user in the image have to be back-projected to the ground. Focal length $f$ is required to perform the back-projection operation, hence, we calibrate the camera first to obtain the focal length. Once we know $f$, $e$ can be computed with a known distance.

*1) Calibrating the Camera:* The calibration method uses two unparallel known distances on the ground to estimate $f$. Normally, we only need to calibrate $f$ once due to the fixed camera lens of most smart devices. Therefore, it is easier to apply compared with the traditional calibration method [29] which requires a checkerboard with multiple images captured.

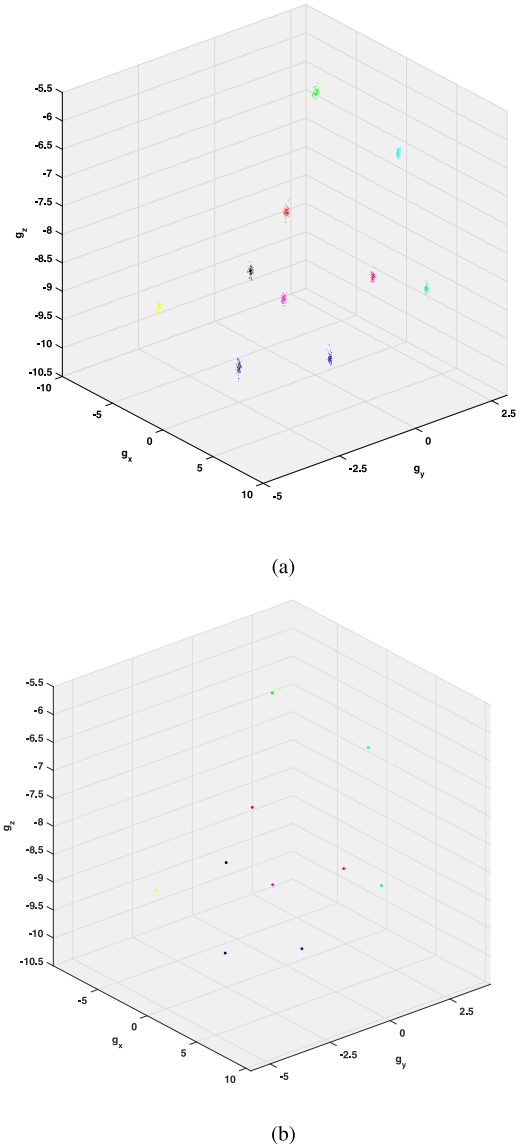



Fig. 3. Examples of gravitational acceleration data denoising. (a) Ten acceleration data sets shown in different colors for ten captures at different positions, with each set being obtained during the one-second photographing duration. (b) Denoised results of the ten data sets shown in (a). Note that the results (shown in dots) in (b) are zoomed in for clear viewing.

Assume that the image plane being $z = 0$ (Fig. 2). The ground can be described as

$$\begin{bmatrix} g_x & g_y & g_z & d \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = 0 \tag{2}$$

where $d$ is a parameter of ground equation and does not affect the measurement results, as explained in Section II-C2. $\boldsymbol{p}(p_x, p_y, 0)$ is a pixel corresponding to a ground point $\boldsymbol{P}(P_X, P_Y, P_Z)$.

The straight line $\boldsymbol{Cp}$ which passes $\boldsymbol{C}$ and $\boldsymbol{p}$ can be parameterized by $t$ as

$$\begin{cases} x' = tp_x \\ y' = tp_y \\ z' = -tf + f. \end{cases} \tag{3}$$

$\boldsymbol{P}$ also lies on the straight line $\boldsymbol{Cp}$ and, therefore

$$\boldsymbol{P} = \begin{bmatrix} P_X \\ P_Y \\ P_Z \end{bmatrix} = \begin{bmatrix} t_p & 0 & 0 & 0 \\ 0 & t_p & 0 & 0 \\ 0 & 0 & 0 & (1-t_p)f \end{bmatrix} \begin{bmatrix} p_x \\ p_y \\ 0 \\ 1 \end{bmatrix} = \boldsymbol{Hp} \tag{4}$$

where $t_p = [(-d - g_z f)/(g_x p_x + g_y p_y - g_z f]$.

Equation (4) shows that $\boldsymbol{p}$ and $\boldsymbol{P}$ are connected by a transform matrix $\boldsymbol{H}$. It can be used to compute the pixel distance between two pixels in the image. For instance, for the two pixels $\boldsymbol{p}(p_x, p_y, 0)$ and $\boldsymbol{p'}(p'_x, p'_y, 0)$ whose back-projection points are $\boldsymbol{P}(P_X, P_Y, P_Z)$ and $\boldsymbol{P'}(P'_X, P'_Y, P'_Z)$, their pixel distance is

$$\begin{aligned} \|\boldsymbol{PP'}\| &= \sqrt{(P_X - P'_X)^2 + (P_Y - P'_Y)^2 + (P_Z - P'_Z)^2} \\ &= sF(\boldsymbol{p}, \boldsymbol{p'}, f). \end{aligned} \tag{5}$$

In (5), $s = \|d + fg_z\|$ is a scalar and irrelevant to ground point and $F(\boldsymbol{p}, \boldsymbol{p'}, f)$ is a function of $\boldsymbol{p}$, $\boldsymbol{p'}$, and $f$

$$F(\boldsymbol{p}, \boldsymbol{p'}, f) = \frac{\sqrt{A_{pp'} f^2 + B_{pp'} f + C_{pp'}}}{\|D_{pp'} f^2 + E_{pp'} f + F_{pp'}\|} \tag{6}$$

where

$$\begin{aligned} A_{pp'} &= \left(g_x^2 + g_z^2\right)(p_x - p'_x)^2 + \left(g_y^2 + g_z^2\right)\left(p_y - p'_y\right)^2 \\ &\quad + 2g_x g_y (p'_x - p_x)\left(p'_y - p_y\right) \\ B_{pp'} &= 2g_z\left(p'_x p_y - p_x p'_y\right)\left(g_y(p_x - p'_x) + g_x\left(p'_y - p_y\right)\right) \\ C_{pp'} &= \left(g_x^2 + g_y^2\right)\left(p'_x p_y - p_x p'_y\right)^2 \\ D_{pp'} &= g_z^2 \\ E_{pp'} &= -\left(g_x g_z p'_x + g_y g_z p'_y + g_x g_z p_x + g_y g_z p_y\right) \\ F_{pp'} &= g_x^2 p_x p_y + g_x g_y p'_x p_y + g_x g_y p_x p'_y + g_y^2 p_y p'_y. \end{aligned}$$

Equation (5) contains two unknown variables, $s$ and $f$. Therefore, two different distances are required to estimate them. A more convenient way is to compute the ratio between two distances so that $s$ which contains $d$ is eliminated. The left variable $f$ can be computed by $\|\boldsymbol{PP'}\|$ and another distance $\|\boldsymbol{QQ'}\|$ for two pixels $\boldsymbol{q}$ and $\boldsymbol{q'}$

$$\frac{\|\boldsymbol{PP'}\|}{\|\boldsymbol{QQ'}\|} = \frac{F(\boldsymbol{p}, \boldsymbol{p'}, f)}{F(\boldsymbol{q}, \boldsymbol{q'}, f)}. \tag{7}$$

Equation (7) is a quartic function that every polynomial equation can be solved by radicals. We propose a robust prediction-based method to estimate $f$ from four roots of (7). It is based on the geometric relationships among focal length, field of view and the image plane of a CCD camera, which is shown in Fig. 4.

The relationship can be formulated as

$$\tan\frac{\alpha}{2} = \frac{h}{2f}. \tag{8}$$

The field of view $\alpha$ is almost fixed while the focal length $f$ changes for this imaging model and, in practice, the horizontal field of view $\alpha$ of the embedded camera is approximately 60°.

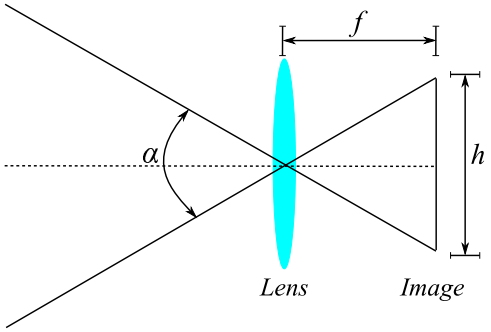



Fig. 4. Illustration of the principle of the CCD camera. $\alpha$ represents the horizontal or vertical view angle with $h$ being the corresponding width or height of the image.

Therefore, a prediction of $f$, $\bar{f}$, can be first estimated by setting $\alpha = 60°$

$$\bar{f} = \frac{h}{2\tan\left(\frac{1}{2}\alpha\right)}. \tag{9}$$

Then, the one closest to $\bar{f}$ among the four roots is selected as the finally estimated focal length.

*2) Computing the Magnification Ratio:* Assume that the relationship between the real distance of two pixels and their corresponding pixel distance is linear. For two pixels $\boldsymbol{p}$ and $\boldsymbol{p'}$, their real distance, $L_{pp'}$, can be estimated from their pixel distance, $\|\boldsymbol{PP'}\|$, via a linear model with a parameter, $e$ called magnification ratio

$$L_{pp'} = e * \|\boldsymbol{PP'}\|. \tag{10}$$

However, there is no information on the relationship between the real distance and the pixel distance and, therefore, a known distance in centimeter whose pixel distance can be estimated by (5) is required to determine $e$. Consequently $e$ can be obtained directly by solving (10) as

$$e = \frac{\widehat{L_{pp'}}}{\widehat{\|\boldsymbol{PP'}\|}} \tag{11}$$

where $\widehat{L_{pp'}}$ represents the known real distance with $\widehat{\|\boldsymbol{PP'}\|}$ being its pixel distance.

Such a known distance can be the ground distance, height or depth so that its pixel distance can be estimated by (5). This condition can be easily satisfied in the real measurement scenario. Furthermore, other kinds of known distance can also be adopted by (11) if some interactions can be taken to estimate its pixel distance (see Section IV for more details).

Although $e$ is scaled by $1/s$ with the unknown $s$, the scale does not affect the distance computation. That is, if there is another pair of points $\boldsymbol{u}$ and $\boldsymbol{u'}$, their real distance $L_{uu'}$ can be computed by (5), (10), and (11) as follows:

$$L_{uu'} = e * \|\boldsymbol{UU'}\| = \frac{F(\boldsymbol{u}, \boldsymbol{u'}, f)}{F(\boldsymbol{p}, \boldsymbol{p'}, f)} L_{pp'}. \tag{12}$$

Apparently, $s$ has no effect on distance measurement, and $d$ included in $s$ is useless as we have claimed before.

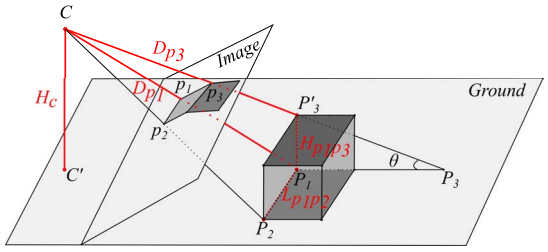



Fig. 5. Illustration of the measurements for different types of distance. $L_{p_1p_2}$ denotes the ground distance between $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$. $H_c$ and $H_{p_1p_3}$ represent the heights of camera and object $\boldsymbol{P_1P'_3}$. $e\|\boldsymbol{CP_1}\|$, $e\|\boldsymbol{CP_2}\|$, and $e\|\boldsymbol{CP'_3}\|$ are depths of $\boldsymbol{p_1}$, $\boldsymbol{p_2}$, and $\boldsymbol{p_3}$, respectively.

### D. Distance Measurement

Fig. 5 illustrates the three types of distance (ground distance, depth, and height) to measure and their corresponding notations. These three types of distance can be the base distances so that an arbitrary distance can be measured.

For depth, the bottom point is required to lie on the ground, e.g., point on the bottom of an object on the ground.

*1) Ground Distance:* Ground distance can be estimated in a relatively simple way according to the proposed method. For the two example ground points $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$ shown in Fig. 5, their corresponding image pixels are $\boldsymbol{p_1}$ and $\boldsymbol{p_2}$. To compute the ground distance $L_{p_1p_2}$, $\boldsymbol{p_1}$ and $\boldsymbol{p_2}$ are first back projected from the image plane to their ground points $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$ according to (4). Then $L_{p_1p_2}$ can be computed by

$$L_{p_1p_2} = e * \|\boldsymbol{P_1P_2}\| = \frac{F(\boldsymbol{p_1}, \boldsymbol{p_2}, f)}{F(\boldsymbol{p}, \boldsymbol{p'}, f)} L_{pp'}. \tag{13}$$

*2) Depth:* Depth is the distance between the camera optical center and the user-specified point in 3-D space. If the specified point is above the ground, its projection on the ground also has to be specified to determine its position in 3-D space. The depth is also computed with the magnification ratio using (10). For the ground point $\boldsymbol{P_1}$ in Fig. 5 whose corresponding image pixel is $\boldsymbol{p_1}$, its depth of $D_{p_1}$ can be computed as follows:

$$D_{p_1} = e * \|\boldsymbol{CP_1}\| = \frac{F(\boldsymbol{O}, \boldsymbol{p_1}, f)}{F(\boldsymbol{p}, \boldsymbol{p'}, f)} L_{pp'}.$$

It is difficult to compute the depth directly if the specified point is above the ground. For example, for the pixel $\boldsymbol{p_3}$ as shown in Fig. 5, its corresponding point in 3-D space is $\boldsymbol{P'_3}$ with back-projected point $\boldsymbol{P_3}$ on the ground. $\|\boldsymbol{CP'_3}\|$, the pixel distance of depth of point $\boldsymbol{p_3}$, cannot be computed directly. However, the distance between $\boldsymbol{C}$ and $\boldsymbol{P_3}$ can be computed in the similar way as $D_{p_1}$, supposing that the specified point is on the ground. The length of the line segment $\boldsymbol{P_3P'_3}$ also can be obtained with the ground distance $L_{p_1p_3}$ and $\theta$. Then the depth of $\boldsymbol{p_3}$ can be computed according to the geometrical relationship as follows:

$$D_{p_3} = e\left\|\boldsymbol{CP'_3}\right\| = e\|\boldsymbol{CP_3}\| - \frac{L_{p_1p_3}}{\cos(\theta)} \tag{14}$$

where $\theta$ is computed by the principle of the triangle geometry

$$\theta = \frac{\pi}{2} - \arccos \frac{\overrightarrow{\boldsymbol{CP_3}} * \boldsymbol{g}}{\left\|\overrightarrow{\boldsymbol{CP_3}}\right\| * \|\boldsymbol{g}\|}.$$

*3) Height:* Two kinds of height can be measured: 1) camera height and 2) object height. The former means the distance from camera optical center $\boldsymbol{C}$ to the ground while the latter means the distance between two specified points: one for the top of object and the other as the ground position for the bottom.

Similar to the ground distance and depth, the height in pixel has to be computed first and then multiplied by the magnification ratio. For the camera height $H_c$ shown in Fig. 5, whose bottom $\boldsymbol{C'}$ is the projection of $\boldsymbol{C}$ on the ground, it can be computed as

$$H_c = e * \left\|\boldsymbol{CC'}\right\| = \frac{L_{pp'}}{F(\boldsymbol{p}, \boldsymbol{p'}, f)\|\boldsymbol{g}\|}. \tag{15}$$

For the example object height $H_{p_1p_3}$ as shown in Fig. 5, $\boldsymbol{p_1}$ and $\boldsymbol{p_3}$ are its bottom and top, respectively, with their corresponding back-projected points on the ground being $\boldsymbol{P_1}$ and $\boldsymbol{P_3}$. The corresponding point in 3-D space for $\boldsymbol{p_3}$ is $\boldsymbol{P'_3}$. Consequently, the object height $H_{p_1p_3}$ can be computed as

$$H_{p_1p_3} = e\left\|\boldsymbol{P_1P'_3}\right\| = e\|\boldsymbol{P_1P_3}\| \tan(\theta)$$

where $\theta$ is the angle between the straight line $\boldsymbol{CP_3}$ and the ground

$$\theta = \frac{\pi}{2} - \arccos \frac{\overrightarrow{\boldsymbol{CP_3}} * \boldsymbol{g}}{\left\|\overrightarrow{\boldsymbol{CP_3}}\right\| * \|\boldsymbol{g}\|}.$$

## III. Experiments and Results

The experimental results are shown qualitatively with real scenes and then quantitatively for analyzing the efficacy of the proposed method. The comparisons with the state-of-the-art methods are presented consequently. The measurement accuracy is evaluated by both absolute error and relative error. Absolute error, $\Psi_a$, is the difference between the estimated distances, $L_{\text{est}}$, and real distances, $L_{\text{real}}$

$$\Psi_a = \|L_{\text{real}} - L_{\text{est}}\|.$$

Relative error, $\Psi_r$, indicates the ratio between the absolute error and the real distances

$$\Psi_r = \frac{\|L_{\text{real}} - L_{\text{est}}\|}{L_{\text{real}}} * 100\%.$$

### A. Qualitative Results

In the real experimental scenes, an iPad Air 2 is used to validate the method by photographing one image of each experimental scene, with the sampling rate of gravitational acceleration signal being 100 Hz and the image resolution being $2592 \times 1936$. In all experiments, we set $\sigma = 0.2$ and $T = 1$ s to denoise the acceleremoter data. $d$ is set to zero since it has no effect on the measurement result.

The first experiment is an indoor scene which contains the regular floor blocks on the ground (Fig. 6). All the blocks are squares of $60 \times 60$ cm. The lengths of segments $\boldsymbol{AB}$ and $\boldsymbol{CD}$ in magenta are three and one times of the block length, respectively. They are used to calibrate the focal length $f$ and estimate the magnification ratio $e$. The distances to measure are marked in red. Among them, the lengths of segments $\boldsymbol{EF}$

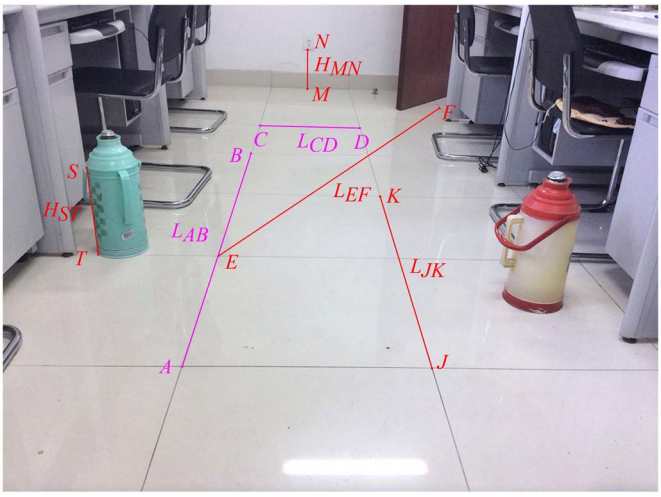



Fig. 6. Experimental scene 1. Segments ***AB*** and ***CD*** in magenta on the ground are used to calibrate the focal length $f$ and compute the magnification ratio $e$. $L_{EF}$ and $L_{JK}$ in red are ground distances to measure while $H_{MN}$ and $H_{ST}$ in red are object heights to measure. The camera height and depths of point cannot show explicitly in this figure.

and ***JK*** are the ground distances to measure for the two ground point pairs, ***E*** and ***F***, and ***J*** and ***K***, separately; the lengths of segments ***MN*** and ***ST*** are the object heights to measure. The camera height and the depth of point cannot be denoted explicitly in the image, so they are not shown in Fig. 6.

For this experiment, the denoised gravitational acceleration ***g*** is $(-0.02, 0.66, -0.32)$ after denoising the acceleration data. The focal length is 2300 after calibrating the camera by ***AB*** and ***CD***. Then, magnification ratio $e$ is obtained as 0.087. With the ratio, different types of distance can be measured by specifying their locations and types. For example, after clicking the locations of ***E*** and ***F*** and selecting its corresponding type being ground distance, $L_{EH}$ estimated by (13) is 269.35 cm. Other distances are measured similarly except that the height of camera has been computed automatically with the estimated magnification ratio according to (15). Table I lists the measurement results for scene 1 with real distance, absolute error and relative error. We can see that: 1) the maximum absolute error is 2.16 cm with its correspondingly relative error being 0.50 and 2) the maximum relative error is 1.93 and its correspondingly absolute error is only 0.58 cm, which is less than 1 cm.

Another three scenes are also experimented (Fig. 7), including two outdoor and one indoor scenes, where different types of distance can be measured in the same ways as those of scene 1 after denoising the acceleration data and computing the magnification ratio. The focal length $f$ which has been calibrated in the first experiment is reused and, therefore, only one known distance is required to compute the ratio. For the three scenes shown in Fig. 7, the distance between two ground points ***A*** and ***B***, $L_{AB}$ in magenta, is the known distance. We see that the known distances can be set by the objects having standard length, such as basketball court, floorboard, and road marking. Therefore, our method can be applied easily in practice due to the widely presented known distances.

TABLE I
MEASUREMENT RESULTS FOR SCENE 1

| Distance Types | Distance to measure | Real distance ($cm$) | Measured distance ($cm$) | Absolute error, $\Psi_a$ ($cm$) | Relative error, $\Psi_r$ (%) |
|---|---|---|---|---|---|
| Ground Distance | $L_{EF}$ | 268.00 | 269.35 | 1.35 | 0.50 |
| | $L_{JK}$ | 120.00 | 118.39 | 1.61 | 1.34 |
| Depth | $D_E$ | 197.00 | 196.83 | 0.17 | 0.09 |
| | $D_F$ | 430.00 | 432.16 | 2.16 | 0.50 |
| | $D_J$ | 147.00 | 147.73 | 0.73 | 0.50 |
| | $D_K$ | 252.00 | 250.64 | 1.36 | 0.54 |
| | $D_M$ | 483.00 | 481.94 | 1.06 | 0.22 |
| | $D_N$ | 478.00 | 477.53 | 0.47 | 0.10 |
| | $D_S$ | 195.00 | 194.01 | 0.99 | 0.51 |
| | $D_T$ | 205.00 | 203.62 | 1.38 | 0.67 |
| Height | $H_C$ | 87.00 | 87.38 | 0.38 | 0.44 |
| | $H_{MN}$ | 30.00 | 29.61 | 0.39 | 1.30 |
| | $H_{ST}$ | 30.00 | 29.42 | 0.58 | 1.93 |

TABLE II
MEASUREMENT RESULTS OF GROUND DISTANCE FOR SCENES 2–4

| Scene (#) | Distance to measure | Real distance ($cm$) | Measured distance ($cm$) | Absolute error, $\Psi_a$ ($cm$) | Relative error, $\Psi_r$ (%) |
|---|---|---|---|---|---|
| 2 | $L_{EF}$ | 350.00 | 348.14 | 1.86 | 0.53 |
| | $L_{JK}$ | 759.00 | 763.54 | 4.54 | 0.60 |
| 3 | $L_{EF}$ | 290.00 | 289.41 | 0.59 | 0.20 |
| | $L_{JK}$ | 200.00 | 202.76 | 2.76 | 1.38 |
| 4 | $L_{EF}$ | 152.50 | 151.88 | 0.62 | 0.41 |
| | $L_{JK}$ | 27.50 | 26.64 | 0.86 | 3.13 |

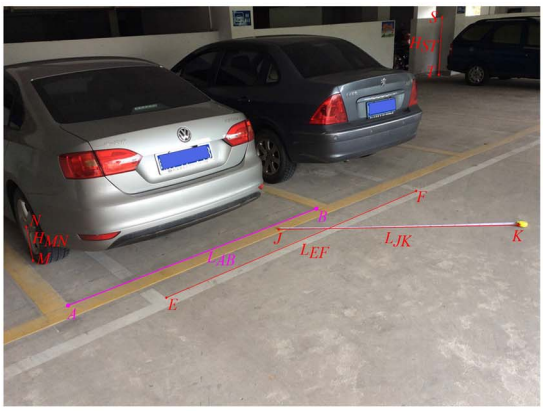
Table II shows the measurement results of ground distance for the three experimental scenes. The max relative error is less than 4% with most errors being less than 2%, which indicates the accuracy of the proposed method. We can also see that small measurement errors may produce large relative errors for short real distances because the absolute error takes more proportions of the real distance. For example, $L_{JK}$ in scene 4 is a short segment and obtains largest relative error (3.13%) even though its absolute error is less than 1 cm. The detailed statistical comparison between absolute error and relative error will be analyzed in Section III-B.

Table III shows the results of measured depth. We can see that the max error is less than 2% with most errors being less than 1%. Since depth is often a long distance, its relative error is small normally. Table IV shows the results of measured height, including camera height, $H_C$, and object height: $H_{MN}$ and $H_{ST}$. The largest relative error (4.41%) is produced by the shortest distance, $H_{ST}$ in scene 4, also because the absolute error takes more proportions of the real distance.

### B. Quantitatively Results

Measurement distance, view orientation, and known distance can be specified differently by different users and, therefore, in this section, statistical experiments on them are presented to explore the performance's variations. The error of the acceleration data reflects the noise degree of the device. Considering the acceleration vector ***g*** is composed of three elements $g_x$, $g_y$, and $g_z$ (Fig. 2), we can use the errors of one element by fixing other two elements for simulating the noises. In particular, we use the errors of $g_x$ or $g_y$ with the error range being between $-0.1$ and 0.1. The magnification


(a)




(b)


(c)

Fig. 7. Three other experimental scenes. Similar to Fig. 6, the segment ***AB*** in magenta in each scene image is used to compute the magnification ratio, while the segments in red are the distances to measure. (a) Scene 2. (b) Scene 3. (c) Scene 4.

ratio does not affect the errors and, therefore, it is set to one in the experiments.

We first focus on the measurement errors under different lengths of measurement distances. The view orientation is set to be coincident with the norm of the ground to minimize its affection. Fig. 8 gives the absolute error and relative error according to different distances. The figure shows that relative error is almost independent of distances although absolute error increases with longer distances. Clearly, large acceleration noise will lead to the exponential increases of both the relative error and absolute error. Fig. 8 also shows that absolute error in short distance is considerably smaller than the absolute error in the long distance even though the relative error is the same in each case.

TABLE III
MEASUREMENT RESULTS OF DEPTH FOR SCENES 2–4

| Scene (#) | Distance to measure | Real distance ($cm$) | Measured distance ($cm$) | Absolute error, $\Psi_a$ ($cm$) | Relative error, $\Psi_r$ (%) |
|---|---|---|---|---|---|
| 2 | $D_E$ | 490.00 | 492.89 | 2.89 | 0.59 |
| | $D_F$ | 490.00 | 490.61 | 0.61 | 0.12 |
| | $D_J$ | 522.00 | 519.99 | 2.01 | 0.39 |
| | $D_K$ | 1052.00 | 1056.92 | 4.92 | 0.47 |
| | $D_M$ | 1100.00 | 1098.67 | 1.33 | 0.12 |
| | $D_N$ | 1088.00 | 1087.01 | 0.99 | 0.09 |
| | $D_S$ | 1090.00 | 1089.16 | 0.84 | 0.08 |
| | $D_T$ | 1100.00 | 1098.67 | 1.33 | 0.12 |
| 3 | $D_E$ | 332.00 | 331.84 | 0.16 | 0.05 |
| | $D_F$ | 520.00 | 517.15 | 2.85 | 0.55 |
| | $D_J$ | 418.00 | 416.37 | 1.63 | 0.39 |
| | $D_K$ | 460.00 | 458.09 | 1.91 | 0.42 |
| | $D_M$ | 420.00 | 418.69 | 1.31 | 0.31 |
| | $D_N$ | 410.00 | 408.44 | 1.56 | 0.38 |
| | $D_S$ | 1518.00 | 1512.87 | 5.13 | 0.34 |
| | $D_T$ | 1525.00 | 1522.33 | 2.67 | 0.18 |
| 4 | $D_E$ | 218.00 | 216.89 | 1.11 | 0.51 |
| | $D_F$ | 220.00 | 223.50 | 3.50 | 1.59 |
| | $D_J$ | 172.00 | 172.12 | 0.12 | 0.07 |
| | $D_K$ | 188.00 | 189.53 | 1.53 | 0.81 |
| | $D_M$ | 207.00 | 207.84 | 0.84 | 0.41 |
| | $D_N$ | 204.00 | 202.85 | 1.15 | 0.56 |
| | $D_S$ | 166.00 | 165.46 | 0.54 | 0.33 |
| | $D_T$ | 167.00 | 166.86 | 0.14 | 0.08 |

TABLE IV
MEASUREMENT RESULTS OF HEIGHT FOR SCENES 2–4

| Scene (#) | Distance to measure | Real distance ($cm$) | Measured distance ($cm$) | Absolute error, $\Psi_a$ ($cm$) | Relative error, $\Psi_r$ (%) |
|---|---|---|---|---|---|
| 2 | $H_C$ | 165.00 | 163.96 | 1.04 | 0.63 |
| | $H_{MN}$ | 200.00 | 201.18 | 1.18 | 0.59 |
| | $H_{ST}$ | 145.00 | 144.12 | 0.88 | 0.61 |
| 3 | $H_C$ | 172.00 | 169.96 | 2.04 | 1.19 |
| | $H_{MN}$ | 30.00 | 29.56 | 0.44 | 1.47 |
| | $H_{ST}$ | 153.50 | 156.47 | 2.97 | 1.93 |
| 4 | $H_C$ | 76.00 | 75.48 | 0.52 | 0.68 |
| | $H_{MN}$ | 15.52 | 15.04 | 0.48 | 3.09 |
| | $H_{ST}$ | 3.40 | 3.25 | 0.15 | 4.41 |

The measurement performance with different view orientations is also analyzed. View orientation is decided by its optical axis with three degrees of freedom in the space and, consequently, it is inconvenient to analyze the orientation. However, the change of the view orientation will lead to the change of the intersection angle between the optical axis and the norm of the ground except when the view translates along the ***XOY*** plane or rotates around the optical axis. Therefore, we use the varying intersection angle between the optical axis and the norm of the ground to analyze the performance variation of view orientation. In this experiment, $g_x$ is set to zero and the angle is adjusted by $g_y$ and $g_z$. The target distance to measure is set to 1000 cm. Fig. 9 shows the statistical results when the angle changes from 0 to $\pi/3$. Only relative errors are shown since the same styles of figure can be obtained for both error metrics due to the fixed target distance.

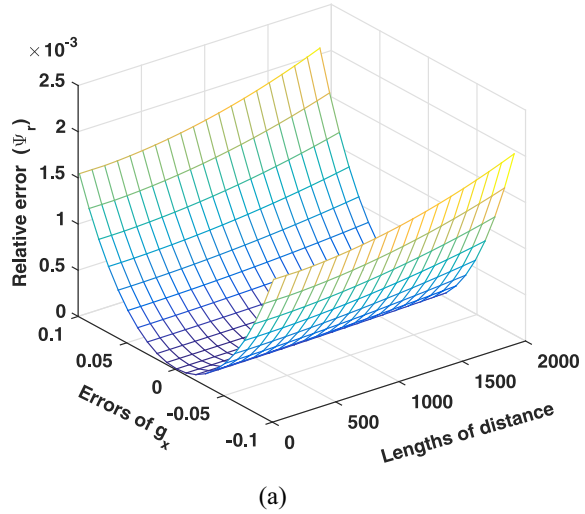



Fig. 8. Performance statistics for the length of distance to measure. (a) Relative errors. (b) Absolute errors.

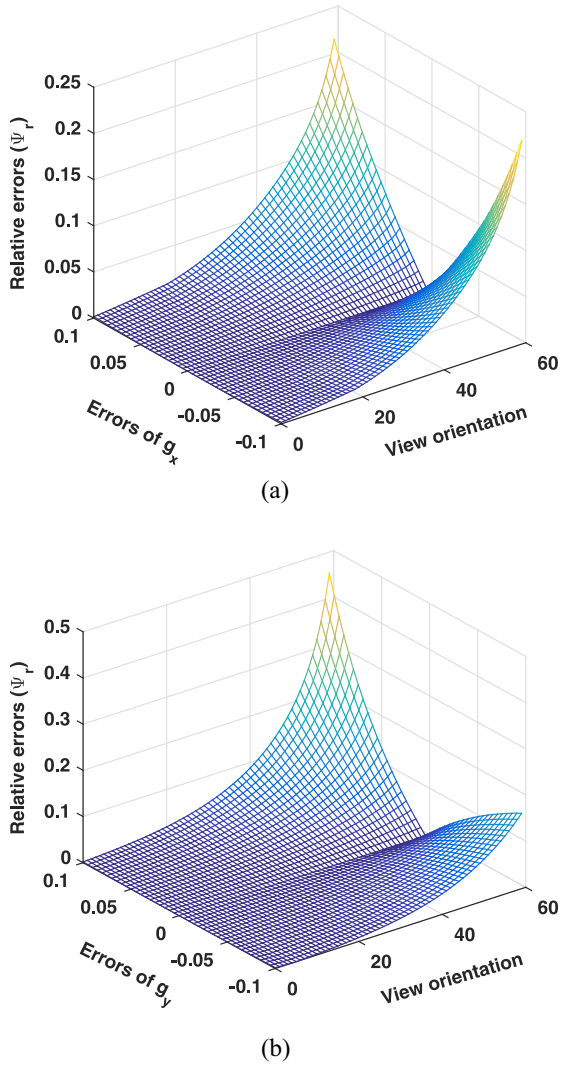



Fig. 9. Performance statistics of relative errors for the photographing angle.

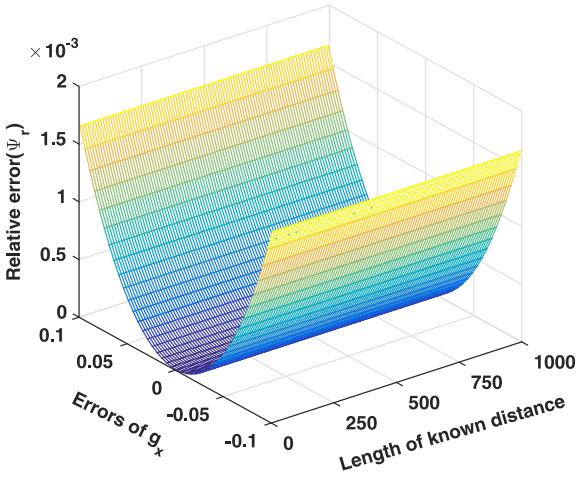
As shown in Fig. 9(a), symmetric results are obtained for the positive and negative errors of $g_x$, no matter what the view orientation is. However, the relative errors for positive error and negative error of $g_y$ are different [Fig. 9(b)]: they increase significantly faster for the positive error than the negative error when the view orientation increases. Hence, noisy $g_x$ and $g_y$ lead to different error performances. Fig. 9 also indicates that the measurement error is more sensitive to the noisy accelerometer data when the image is captured with an intersection angle larger than $\pi/4$.

One known ground distance in centimeter is required for obtaining the magnification ratio. Therefore, we also analyze whether the length of the known distance affects measurement accuracy. The view orientation is set to be coincident with the norm of the ground. Fig. 10 shows the statistical result, where the length of the known distance varies from 10 to 1000 cm. It can be seen that the known distance has little influence on relative error, however, a short known distance (less than 100 cm) seriously lowers the robustness of the method. This phenomenon is caused by the exponential growth of magnification ratio when the known distance decreases.

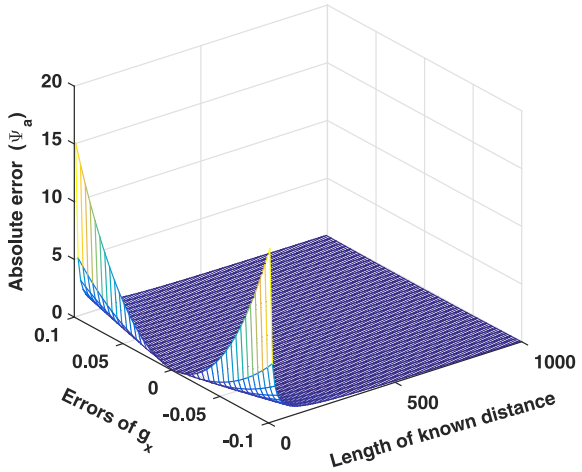
From the above quantitative analysis, we can summarize that the proposed method performs most robust when the ground is horizontal, the view orientation intersects the norm of the ground with an angle less than $\pi/4$ and the known distance is larger than 100 cm.

### C. Comparison With Existing Methods

We compare with three state-of-the-art methods, including Jiang and Jiang [20], Gao *et al.* [16], and Laotrakunchai *et al.* [23]. Both Jiang and Jiang [20] and Gao *et al.* [16] took the image-based approach by a general camera with some special pattern, where the former needed two concentric circles for distance computation and the latter used a checkerboard [29] to calibrate the camera. Laotrakunchai *et al.* [23] measured the distance directly with the acceleration data obtained from the accelerometer of the smart mobile device. However, a mapping function which builds the relationship between the real length and the observed length should be learned first. Therefore, this method is very time-consuming and unstable. In addition, it

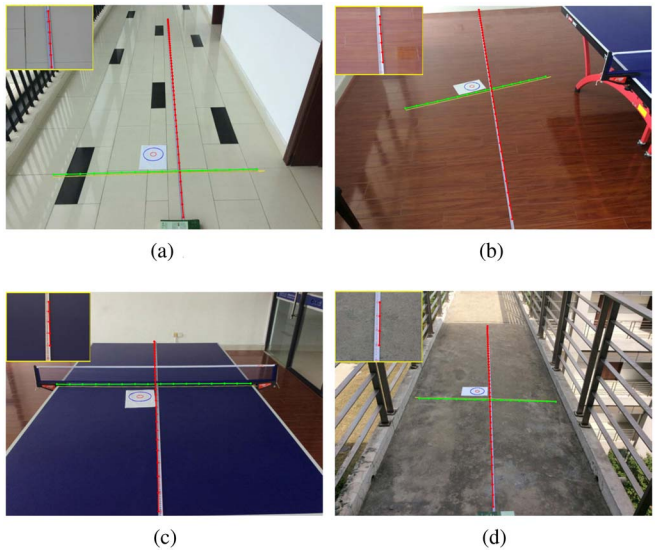



(a)




(b)

Fig. 10. Performance statistics for the length of the known ground distance. (a) Relative errors. (b) Absolute errors.



(a) (b) (c) (d)

Fig. 11. Four experimental scenes for the performance comparison. The small image in the top left corner of each subfigure gives an example one additionally captured with the long focal length (30.101 mm) for the depth to measure by the method of Gao *et al.* [16]. (a) Scene 5. (b) Scene 6. (c) Scene 7. (d) Scene 8.

requires a strict acceleration-deceleration dragging process to obtain the distance, i.e., first speed up the moving to the maximum velocity from the start position and then slow down continuously to finally stop at the target measurement position. Our method takes the image-based way but with only one shot of the scene by the smart mobile device, two known distances to calibrate the embedded camera and one easily accessed known distance to compute the distance. Apparently, our method is different from these representative methods and more convenient than them.

Table V summarizes the properties of these methods and ours. In the following, the comparison strategy based on this table is presented first and then the comparison results are shown and discussed.

*1) Comparison Strategy:* Jiang and Jiang [20] only measured the distance on a plane or the ground, where the assistant pattern of the two concentric circles lies. Therefore, this method can only measure the ground distance in comparison with ours, as shown in Table V. Using the checkerboard-based camera calibration, Gao *et al.* [16] could only measure depth and required to capture two images with different focal lengths at a fixed camera position. Laotrakunchai *et al.* [23] did not use the calibration pattern but purely relied on the data from the embedded accelerometer of the mobile device. Such data should be captured by the acceleration-deceleration style of device dragging. Apparently this strict requirement is difficult to satisfy in practice, especially when the moving distance is long. Therefore, this method is restricted to a very short distance (normally less than 50 cm). Its depth measurement is also difficult to do because it is nontrivial to: 1) restrict the hand motion in the acceleration-deceleration style for a point in the 3-D space and 2) separate the hand acceleration from the noisy accelerometer date affected by the gravity.

Relatively accurate ground distance, however, can be obtained by the method of Laotrakunchai *et al.* [23], if we can drag the device on the horizontal plane. In this case, if enough care is taken during the dragging, the hand can be supported stably with an acceleration-deceleration motion without the gravity affection on the motion acceleration. Therefore, ground distance can be taken for this method in comparison with ours. Similar configuration can also be applied to the vertical motion for height measurement by anchoring the hand along a vertical plane. But, considering that such a vertical motion is similar to the horizontal motion, we do not consider such a vertical motion in the comparison.

We now see that these methods cannot measure all types of distance as ours and consequently two separate comparisons are taken with them selectively, one for the ground distance comparison of our method with the methods of Jiang and Jiang [20], and Laotrakunchai *et al.* [23], and the other for the depth between the method of Gao *et al.* [16] and ours.

*2) Comparison Results:* Fig. 11 shows the four comparison scenes. The method of Jiang and Jiang [20] will introduce

TABLE V
COMPARISON OF THE PROPERTIES OF OUR METHOD WITH THE METHODS OF JIANG AND JIANG [20], GAO *et al.* [16], AND LAOTRAKUNCHAI *et al.* [23]. NOTE: LAOTRAKUNCHAI *et al.* [23] REQUIRED THE DEVICE BE DRAGGED IN AN ACCELERATION-DECELERATION STYLE WHICH IS ONLY VALID FOR LESS THAN 50 cm AND, THEREFORE, IS NOT APPLICABLE FOR LONG DISTANCE

| Method | Number of images | Capture method | Type of distance in comparison with ours | | | Applicable to long distance |
|---|---|---|---|---|---|---|
| | | | Ground distance | Height | Depth | |
| Jiang and Jiang [20] | 1 | General camera with two concentric circles | √ | × | × | √ |
| Gao *et al.* [16] | 2 | General camera calibrated by a checkerboard | × | × | √ | √ |
| Laotrakunchai *et al.* [23] | 0 | Smart mobile device with a learnt mapping function | √ | √ | × | × |
| Our method | 1 | Smart mobile device with two known distances | √ | √ | √ | √ |

extra error when applying edge detection to determine the concentric circles and, therefore, in each scene, two known concentric circles (the blue and red ones shown in Fig. 11) with radii being 79 and 23.5 mm separately are printed in an A4 paper as the calibration pattern on the ground. The yellow and white rulers are used for measurement, where their scales are used as the start or end points of the measurement and, also provide the ground-truth ground distances. Both rulers are perpendicularly placed with the yellow one being almost horizontal for obtaining versatile distances to measure. The white ruler nearly vertically placed is longer than the yellow one to provide more scales due to the perspective effect.

For depth computation by the method of Gao *et al.* [16], two images are captured by a BenQ GH600 camera for each distance to measure, with the two calibrated focal lengths being 14.927 mm and 30.101 mm, respectively (see the small image on the top left corner of each subfigure in Fig. 11).

Considering the localization error when selecting each scale manually, for each method, we measure each distance eight times by carefully selecting the scale eight times and compute the final distance through averaging the best four measurements. Fig. 12 gives the comparison results on the ground distance. It can be seen that the errors from ours are smaller than those from the other two methods and decrease slower than theirs. This difference becomes significant when the distance increases. The method of Jiang and Jiang [20] is sensitive to large distance mainly because the reprojection error increases linearly when the point is far away from the concentric circles. The method of Laotrakunchai *et al.* [23] is also unstable because of its special dragging requirement, making the relative errors shaking significantly when the distance increases. Fig. 13 gives the results of the comparison between Gao *et al.*'s [16] method and ours, where it can be seen that our method always obtains smaller errors. Especially, the absolute errors of ours turns smaller than that from Gao *et al.* [16] when the depth turns larger. Their method requires to manually provide the locations of the target point in the two images. However, the accurate location becomes difficult to choose when the depth becomes large and, therefore, their absolute error becomes larger with the increase of the depth.

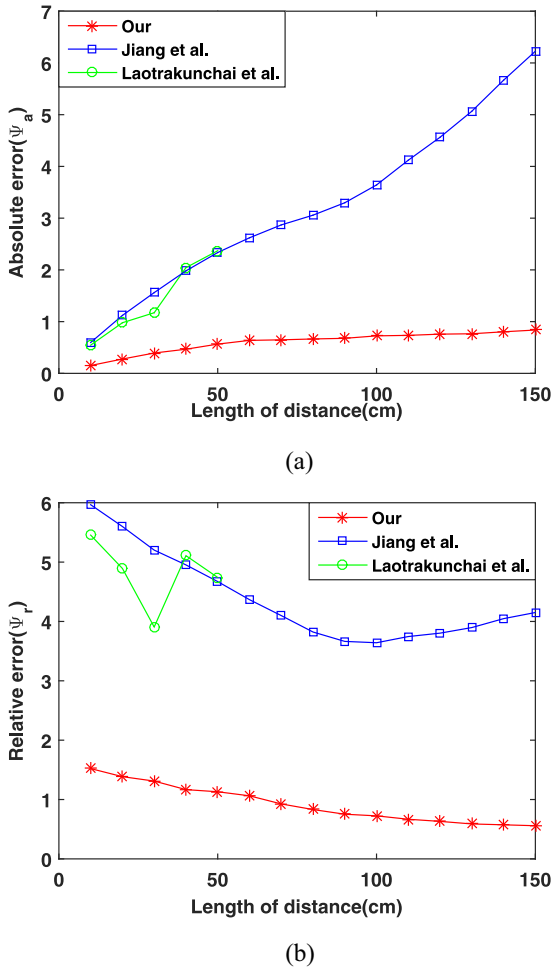
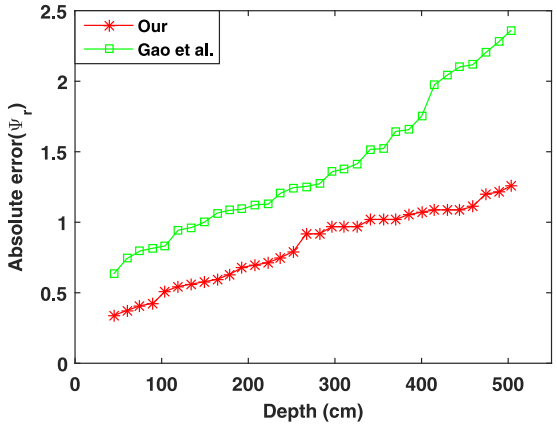


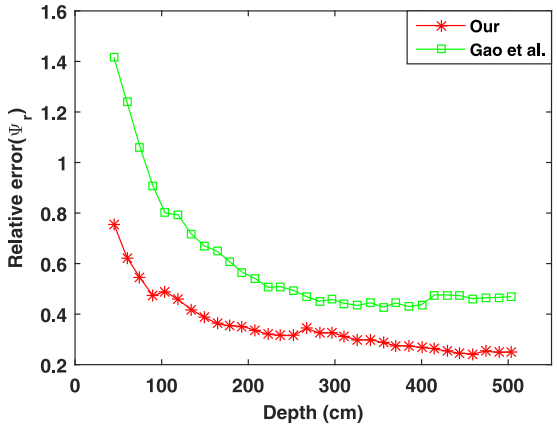

Fig. 12. Performance comparison of our method with the methods of Jiang and Jiang [20] and Laotrakunchai *et al.* [23] on ground distance measurement. Note: The method of Laotrakunchai *et al.* [23] can only measure up to 50 cm. (a) Absolute errors. (b) Relative errors.

## IV. ARBITRARY DISTANCE MEASUREMENT

Our method can also measure arbitrary types of distance on any object with proper user interactions in the image. The interactions are simply to choose some base pixels which should be the ground ones for obtaining the ground distance, height or depth as the base distances. Then the arbitrary distance can be measured according to the geometrical relationship between the target distance and the base distances. Fig. 14 shows the principle of this idea with a cuboid object in the scene. $\boldsymbol{p_1}$, $\boldsymbol{p_2}$, and $\boldsymbol{p_3}$ are image points of the vertexes $\boldsymbol{P_1}$, $\boldsymbol{P_2}$, and $\boldsymbol{P'_3}$ in 3-D space, respectively. $\boldsymbol{p_4}$ is the specified pixel on the edge of the cuboid, representing the projection of 3-D point $\boldsymbol{P'_4}$. If we want to compute the distance between $\boldsymbol{P'_3}$ and $\boldsymbol{P'_4}$, $L_{p_3p_4}$, we can choose $\boldsymbol{p_1}$ and $\boldsymbol{p_2}$ on the ground as base pixels and compute $L_{p_3p_4}$ via the base distances $D_{p_3}$ and $D_{p_4}$,




(a)




(b)

Fig. 13. Performance comparison of our method with the methods of Gao *et al.* [16] on depth measurement. (a) Absolute errors. (b) Relative errors.

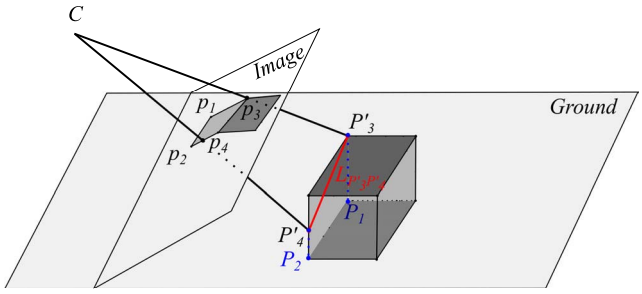



Fig. 14. Illustration of our method for arbitrary distance measurement.

respectively. Denoting $\boldsymbol{p_3}$ or $\boldsymbol{p_4}$ as $\boldsymbol{p}^+$ and $\boldsymbol{P'_3}$ or $\boldsymbol{P'_4}$ as $\boldsymbol{P}^+$. The relationship between the two correspondences, $\boldsymbol{p}^+$ and $\boldsymbol{P}^+$, is

$$D_{p^+} * \boldsymbol{p}^+ = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \boldsymbol{P}^+ \quad (16)$$

where $u_0$ and $v_0$ are the principal points of the camera. $D_{p^+}$ in (16) is the depth of $\boldsymbol{p}^+$ computed by (14) with the help of the base pixels $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$.

Then, $L_{p_3p_4}$ can be obtained by

$$L_{p_3p_4} = e\left\|\boldsymbol{P'_3P'_4}\right\|.$$

## V. Conclusion

In this paper, we have proposed a novel method for conveniently measuring various distances from a single image captured by a smart mobile device. The accelerometer integrated in the smart device is used to estimate the view direction which thus helps back-projecting the image pixel to the ground. The back projection is performed based on a new camera calibration method which can estimate the focal length accurately with two known distances. Then, the magnification ratio can be computed for converting the pixel distance into real distance. With back-projection and the magnification ratio, various types of distance including ground distance, depth, and height can be measured accurately. Experimental results show the effectiveness of the proposed method. Currently, our method requires a known distance for the measurement. How to measure the distance without the known distance will be considered in our future work.

## References

[1] H. Zhao and R. Shibasaki, "A vehicle-borne urban 3-D acquisition system using single-row laser range scanners," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 33, no. 4, pp. 658–666, Aug. 2003.

[2] J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced computer vision with microsoft kinect sensor: A review," *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1318–1334, Jun. 2013.

[3] B. Langmann, W. Weihs, K. Hartmann, and O. Loffeld, "Development and investigation of a long-range time-of-flight and color imaging system," *IEEE Trans. Cybern.*, vol. 44, no. 8, pp. 1372–1382, Aug. 2014.

[4] Z. Jiang, N. Jiang, Y. Wang, and B. Zang, "Distance measurement in panorama," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, San Antonio, TX, USA, 2007, pp. VI-393–VI-396.

[5] M.-C. Lu, W.-Y. Wang, and C.-Y. Chu, "Image-based distance and area measuring systems," *IEEE Sensors J.*, vol. 6, no. 2, pp. 495–503, Apr. 2006.

[6] C.-M. Wang and W.-Y. Chen, "The human-height measurement scheme by using image processing techniques," in *Proc. Int. Conf. Inf. Security Intell. Control (ISIC)*, 2012, pp. 186–189.

[7] C.-T. Chuang, W.-Y. Wang, C.-P. Tsai, Y.-H. Chien, and M.-C. Lu, "An image-based area measurement system," in *Proc. Int. Conf. Syst. Sci. Eng. (ICSSE)*, 2011, pp. 644–648.

[8] M.-C. Lu, C.-C. Hsu, and Y. Y. Lu, "Improvements and application of the image-based distance measuring system," in *Proc. WSEAS Int. Conf. (CISST)*, 2007, pp. 17–19.

[9] W.-Y. Wang, M.-C. Lu, C.-T. Chuang, and J.-C. Cheng, "Image-based height measuring system," in *Proc. 7th WSEAS Int. Conf. Signal Process. Comput. Geometry Artifical Vis. (ISCGAV)*, 2007, pp. 147–152.

[10] T.-H. Wang, M.-C. Lu, W.-Y. Wang, and C.-Y. Tsai, "Distance measurement using single non-metric CCD camera," in *Proc. 7th WSEAS Int. Conf. Signal Process. Comput. Geometry Artif. Vis.*, 2007, pp. 24–26.

[11] C.-C. Chen, M.-C. Lu, C.-T. Chuang, and C.-P. Tsai, "Vision-based distance and area measurement system," *WSEAS Trans. Signal Process.*, vol. 4, no. 2, pp. 36–43, 2008.

[12] N. Pauly and N. I. Rafla, "An automated embedded computer vision system for object measurement," in *Proc. Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Columbus, OH, USA, 2013, pp. 1108–1111.

[13] M.-C. Lu, C.-C. Hsu, and Y.-Y. Lu, "Image-based system for measuring objects on an oblique plane and its applications in 2-D localization," *IEEE Sensors J.*, vol. 12, no. 6, pp. 2249–2261, Jun. 2012.

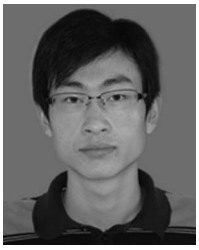
[14] Y. Wang and Z. Jiang, "Distance measurement in panoramic video," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, New York, NY, USA, 2009, pp. 426–429.

[15] Y. M. Mustafah, R. Noor, H. Hasbi, and A. W. Azma, "Stereo vision images processing for real-time object distance and size measurements," in *Proc. Int. Conf. Comput. Commun. Eng. (ICCCE)*, 2012, pp. 659–663.

[16] H. Gao, J. Liu, Y. Yu, and Y. Li, "Distance measurement of zooming image for a mobile robot," *Int. J. Control Autom. Syst.*, vol. 11, no. 4, pp. 782–789, 2013.

[17] H. Kim, C.-S. Lin, J. Song, and H. Chae, "Distance measurement using a single camera with a rotating mirror," *Int. J. Control Autom. Syst.*, vol. 3, no. 4, pp. 542–551, 2005.

[18] C.-F. Wu, C.-J. Lin, and C.-Y. Lee, "Applying a functional neurofuzzy network to real-time lane detection and front-vehicle distance measurement," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 4, pp. 577–589, Jul. 2012.
[19] W. Wang, J. Shen, X. Li, and F. Porikli, "Robust video object cosegmentation," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3137–3148, Oct. 2015.
[20] N. Jiang and Z. Jiang, "Distance measurement from single image based on circles," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Honolulu, HI, USA, 2007, pp. 809–812.
[21] K. A. Rahman *et al.*, "Person to camera distance measurement based on eye-distance," in *Proc. 3rd Int. Conf. Multimedia Ubiquitous Eng. (MUE)*, 2009, pp. 137–141.
[22] C. Holzmann and M. Hochgatterer, "Measuring distance with mobile phones using single-camera stereo vision," in *Proc. Int. Conf. Distrib. Comput. Syst. Workshops (ICDCSW)*, 2012, pp. 88–93.
[23] S. Laotrakunchai, A. Wongkaew, and K. Patanukhom, "Measurement of size and distance of objects using mobile devices," in *Proc. Signal Image Technol. Internet Based Syst. (SITIS)*, Kyoto, Japan, 2013, pp. 156–161.
[24] X. Dong, J. Shen, L. Shao, and L. Van Gool, "Sub-Markov random walk for image segmentation," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 516–527, Feb. 2016.
[25] J. Han *et al.*, "Two-stage learning to predict human eye fixations via SDAEs," *IEEE Trans. Cybern.*, vol. 46, no. 2, pp. 487–498, Feb. 2016.
[26] D. Han and C. Wang, "Tree height measurement based on image processing embedded in smart mobile phone," in *Proc. Int. Conf. Multimedia Technol. (ICMT)*, Zürich, Switzerland, 2011, pp. 3293–3296.
[27] X. Lu, X. Li, and L. Mou, "Semi-supervised multitask learning for scene recognition," *IEEE Trans. Cybern.*, vol. 45, no. 9, pp. 1967–1976, Sep. 2015.
[28] B. Du *et al.*, "Exploring representativeness and informativeness for active learning," *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2015.2496974.
[29] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.



**Shangwen Chen** is currently pursuing the Ph.D. degree with the School of Computer Science and Technology, Anhui University, Hefei, China.

His current research interests include mobile computing and computer vision.

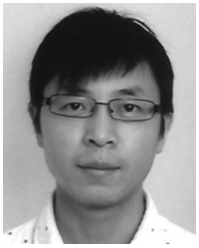
**Xianyong Fang** is a Full Professor with the School of Computer Science and Technology and also the Director of the Institute of Media Computing, Anhui University, Hefei, China. His current research interests include computer graphics and computer vision.

**Jianbing Shen** (M'11–SM'12) is a Professor with the School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China. He has published about 50 journal and conference papers such as the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, the IEEE TRANSACTIONS ON CYBERNETICS, the IEEE CVPR, and the IEEE ICCV. His current research interest includes computer vision.

Dr. Shen has also obtained many flagship honors, including the Fok Ying Tung Education Foundation from Ministry of Education, the Program for Beijing Excellent Youth Talents from Beijing Municipal Education Commission, and the Program for New Century Excellent Talents from Ministry of Education.

**Linbo Wang** received the Ph.D. degree in computer science and technology, Nanjing University, Nanjing, China, in 2014.

He is currently a Lecturer with the School of Computer Science and Technology, Anhui university, Hefei, China. His current research interests include image processing and computer vision.



**Ling Shao** (M'09–SM'10) received the Ph.D. degree in computer vision from the Robotics Research Group, University of Oxford, Oxford, U.K.

He is a Full Professor and the Head of the Computer Vision and Artificial Intelligence Group with the Department of Computer Science and Digital Technologies, Northumbria University, Newcastle upon Tyne, U.K., and an Advanced Visiting Fellow with the Department of Electronic and Electrical Engineering, University of Sheffield, Sheffield, U.K. His current research interests include computer vision, image processing, pattern recognition, and machine learning.

Dr. Shao is an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CYBERNETICS, and other journals. He is a fellow of the British Computer Society, IET, and a Life Member of ACM.