

Lightweight Distance Estimation from Single Images without Camera Intrinsics

Jarosław Bernacki

Department of Computer Science and Systems Engineering, Wrocław University of Science and Technology, Wyb. Wyspiańskiego 27, Wrocław, 50-370, Poland

E-mail: jaroslaw.bernacki@pwr.edu.pl

ORCID iD: <https://orcid.org/0000-0002-4488-3488>

Received: May 28, 2026; Revised: June 18, 2026; Accepted: July 8, 2026; Published: August 8, 2026

Abstract: In this paper, we address the problem of estimating the distance between a camera and a photographed object using minimal prior information. Specifically, the goal is to obtain distance estimates without access to intrinsic camera parameters such as focal length, sensor size, or lens distortion coefficients. We propose three simple heuristic methods, which can be viewed as lightweight variants of the classical camera pinhole model (CPH), but do not require calibration. Instead, they rely only on the approximate real height of a known object in the scene. The methods were validated on a representative dataset of images captured with several modern cameras and smartphones, using known object dimensions as ground truth. Their performance was compared against CPH using error metrics such as mean absolute error (MAE), mean absolute percentage error (MAPE), root mean squared error (RMSE), mean signed error (MSD), coefficient of determination R^2 , and supported by statistical testing (Shapiro–Wilk, ANOVA, Kruskal–Wallis). The analysis confirmed that, while less precise than fully calibrated approaches, the proposed heuristics achieve consistent and reliable distance estimates under minimal assumptions. These methods are particularly suited for lightweight applications and devices with fixed focal lengths, such as smartphones.

Index Terms: Digital forensics, camera pinhole model, image processing, privacy

1. Introduction

Photography is widely used in everyday life. Many people take photos with compact, mirrorless cameras (DSLM) or digital single-lens reflex (DSLR) cameras. For several years, other devices like smartphones and drones have seen a significant increase in popularity. A large number of users capture and instantly share their photos on social media. A huge number of photos encourage research in the field of digital and camera forensics. Camera forensics is a wide field that covers different areas, among others, hardwaremetry [1-4] or photogrammetry [5]. The popular problem in hardwaremetry is the camera's sensor recognition [6] that makes it possible to determine if a photo was taken by a particular camera or flatbed scanner recognition [7].

Digital forensics is a field that deals with the recovery, analysis, and securing of digital evidence in the goals of investigations. It is crucial in combating cybercrime, detecting forgeries, and analyzing incidents related to data breaches. One of its aspects is the analysis of digital images, including, among others, identifying the source of the photo or the authenticity of image files. Estimating the distance between the camera and the photographer can help determine the circumstances of taking the photo, which is important, for example, in the reconstruction of criminal events. These techniques use computer image processing methods and optical models, such as the pinhole model. Thanks to this, digital forensics provides tools for effective analysis of visual digital evidence [8,9,10].

In this paper, we deal with the problem of estimating the distance from the camera to the photographed object. This problem is generally solved by means of hardware devices, and there are a few purely algorithmic solutions. Previous research gives precise results but generally requires a large amount of auxiliary data, such as focal length, principal point, or image sensor format [11,12]. Moreover, this data may not be available or may be corrupted. Image distortion parameters (e.g., radial or tangential lens distortion) can also be used for distance estimation [13,14]. However, such an approach can be impractical because distortion depends both on the focal length and the distance to the object, requires prior calibration, and in some imaging setups, the distortion may be negligible or absent. Therefore, there is a motivation to develop a method that strikes a balance between accuracy and minimal reliance on camera or image metadata.

We propose three simple yet effective methods that do not use the camera's intrinsic parameters and rely only on information from a single image. In short, they are based on measuring the apparent height of the object in the image,

comparing it with known reference sizes, and applying a lightweight heuristic correction. The proposed methods are compared with the state-of-the-art camera pinhole model (CPH) [15] in terms of distance estimation performance. The experiments are conducted on a representative set of photographs from 10 different devices, where the real distance between the camera and the object was measured. Experimental results, enhanced by error measures analysis, including mean absolute error (MAE), mean percentage absolute error (MAPE), root mean squared error (RMSE), mean signed error (MSD), and coefficient of determination (R^2), show that the accuracy of the proposed methods is higher than the camera pinhole model. The obtained accuracy (approximately 18–20% MAPE) is lower than that of modern deep learning-based monocular depth estimation methods, but is achieved without camera calibration, camera metadata, or training data. Therefore, the proposed approach is intended for lightweight applications where such information is unavailable rather than for high-precision distance measurements.

1.1. Design goals and contribution

The goal of this work is to estimate the distance between the camera and the photographed object using the smallest possible amount of information. In particular:

- No intrinsic parameters of the camera (e.g., focal length, sensor size) are required, nor any additional calibration data such as lens distortion or aberration coefficients;
- Only the approximate real height of the photographed object is assumed to be known.

Unlike classical formulations based on the camera pinhole model, which rely on calibrated parameters, or recent learning-based monocular depth estimation approaches that require extensive training data, our methods occupy a complementary middle ground. The main contribution of this paper is the development of three simple but effective heuristics for distance estimation from a single image that minimizes the dependence on auxiliary information.

In contrast to prior pixel-based estimators that typically adapt the pinhole model with partial calibration data, the proposed approaches deliberately eliminate all dependencies on focal length, sensor size, or pixel scale, and instead reformulate the estimation problem in terms of directly observable image ratios. This makes the proposed methods distinct from existing geometric heuristics: they are tailored for situations where metadata is absent or cannot be trusted, and they provide interpretable formulas that trade some accuracy for maximum applicability. By using only basic image content, such as the apparent size of a known object, our approaches remain transparent, computationally inexpensive, and broadly applicable in situations where metadata is missing, unreliable, or deliberately removed, which is often the case in mobile imaging and forensic scenarios. Therefore, this paper examines the practical use of simple, calibration-free heuristics derived from pinhole geometry for estimating object distance in situations where intrinsic parameters or calibration data are unavailable. The proposed approach is not intended to compete with modern monocular depth estimation methods, but rather to provide a lightweight engineering solution for specific application scenarios. The emphasis is placed on operational simplicity, practical applicability, and a discussion of the method's assumptions and limitations.

1.2. Organization of the paper

This paper is organized as follows. In section 2, the previous and related works are described. Section 3 presents the proposed approaches, and in section 4, the results of distance estimation with the statistical analysis are described. The final section concludes with this work.

2. Related work

The estimation of object-to-camera distance is a long-studied problem in computer vision and photogrammetry. The most common theoretical foundation is the camera pinhole model (CPH) [15–19], which describes the geometric relationship between real-world points and their projections onto an image plane. The model originates from the concept of a pinhole camera (or camera obscura), where light passes through a small aperture and forms an inverted image on a photosensitive surface. The principles of this model have been known since antiquity, discussed by Aristotle and later formalized by Euclid in *Optics* [20].

Hardware-based and calibrated approaches Many existing distance estimation systems rely on dedicated hardware or calibrated setups. For example, fixed cameras with stable parameters have been used for controlled distance measurements [21]. Smith et al. [11] explicitly modeled lens distortion, showing its dependence on focal length and distance [14]. Magill [13] and Brown [12] formulated parametric distortion models to predict distance-dependent optical deviations. Matabosch et al. [22] combined a consumer-grade camera with a stripe laser to achieve distance estimation, highlighting the role of auxiliary illumination.

Application-specific uses of pinhole geometry Several studies adapted the CPH for particular application domains. Loktev [23] used image blur and probabilistic analysis to infer geometric parameters of moving objects in monitoring systems. Kawakita et al. [24] designed a dual-mode camera that simultaneously captures color images and depth

information under varied lighting, reporting sensitivities to object reflectivity and illumination. Megalingam et al. [25] employed pinhole geometry and triangle similarity for automotive distance estimation after focal length calibration. Leorna et al. [26] applied a similar model to wildlife monitoring, validating accuracy through both simulation and field measurements. Saputra et al. [27] estimated pedestrian distances for adaptive street lighting using triangle similarity, achieving below 10% errors at short ranges.

Learning-based and feature-driven methods Other research explores vision-based detection pipelines that estimate both object presence and distance. Chou et al. [28] used YOLO-based object detection combined with a field-of-view model to assist navigation for visually impaired users. Gökçe et al. [29] focused on UAV detection, evaluating Haar-like, HOG, and LBP features with boosted classifiers. Natanael et al. [30] employed YOLO and color–edge cues for vehicle-following applications. More recently, deep learning has dominated monocular depth estimation: supervised and self-supervised models such as Monodepth2 [31] and MiDaS/DPT [32] predict dense depth maps with high accuracy, often combined with object detection for metric inference [33–35]. Despite excellent performance, these networks require large training datasets, calibration priors, and significant computational resources.

Relation to broader vision research Recent progress in robotics, remote sensing, and multimodal vision [36–42] has expanded scene understanding toward semantic and qualitative tasks such as navigation and object interaction. However, these methods rarely focus on direct, metric-scale distance estimation from a single uncalibrated image, which remains essential for practical fieldwork and low-resource applications.

3. Proposed method

3.1. State-of-the-art – Camera pinhole model

Let us shortly recall the camera pinhole model (CPH) [18]. It describes the geometric relationship between a 3D point and its 2D corresponding projection onto the image plane. The pinhole projection formula is defined as in Equation 1:

$$\frac{h}{f} = \frac{H}{d} \quad (1)$$

where:

- h is the size of the object on the sensor;
- f is the focal length of the lens;
- H is the real size of the object;
- d is the distance from a nodal point to the object.

All the above measures are expressed in millimeters.

3.2. The proposed methods

We propose three simple heuristics called EstI, EstII, and EstIII, which are lightweight variants inspired by the pinhole camera model and designed to estimate the camera–object distance without requiring calibration or metadata. The proposed estimators intentionally avoid explicit calibration parameters, which may introduce scale bias across devices; this trade-off is accepted in exchange for simplicity and ease of use. All these heuristics are motivated by the pinhole camera geometry, where the apparent image size of an object is inversely proportional to its distance from the camera. In its simplest form, this can be written as (Eq. 2):

$$d = C \cdot \frac{h_I}{h_o} \quad (2)$$

Where C is a proportionality constant that depends on focal length and sensor size. In practice, this constant can be estimated once for a device or device class using a simple reference measurement. EstII and EstIII extend this baseline with additional correction terms, which may be interpreted as compensating for tilt, perspective, or other scene factors. In the present work, these terms are treated as lightweight heuristic adjustments rather than fully calibrated geometric models, consistent with our focus on calibration-free and computationally inexpensive methods.

The EstI method The camera pinhole model is a reliable and precise approach for determining the distance between the camera and the object. However, it can be problematic in usage because the focal length of the photograph must be known. This information is typically stored as camera metadata (such as EXIF) and can be easily manipulated or entirely deleted.

Also, the size of the object at the camera's sensor must be determined, which may be a difficult task, most often due to no knowledge about the camera's sensor.

Let h_I and w_I denote image height and width in pixels, respectively. We have observed that the distance d between the camera and the photographed object can be estimated by a very simple heuristic called EstI, without data required in the camera pinhole model. Therefore, we propose a formula as shown in eq. 3.

$$d = \frac{h_I \cdot h_r}{h_o} \quad (3)$$

where:

- h_I is the height of the photograph in pixels;
- h_r is the approximation of the real height of the photographed object in millimeters;
- h_o is the height of the object in the photograph in pixels.

Distance d is expressed in millimeters. Such an approach gives an approximation of the real distance between the camera and the object. No particular knowledge about the image or camera is needed. The only required information that may not always be available is the real object height h_r , which is assumed to be known or measured beforehand. In practice, this requirement can often be satisfied, for example, when photographing engineering structures or other objects of standard size. The proposed method is fast and easy to implement: h_o may be easily obtained using simple graphic software, and h_I denotes the height of the photograph in pixels.

The EstII method Another proposed method, called EstII, is an extension of the EstI approach. In some cases, the EstII provides more accurate estimations of the d than the EstI. This approach is presented as eq. 4.

$$d = \frac{h_I \cdot h_r}{h_o + H_m \sqrt{\frac{w_I}{h_I}}} \quad (4)$$

where:

- h_I – image height in pixels;
- w_I – image width in pixels;
- h_r – is the approximation of the real height of the photographed object in millimeters;
- h_o – object height in the image (in pixels);
- and:

$$H_m = \frac{\frac{w_I}{h_I}}{\sqrt{1 + \left(\frac{h_I}{w_I}\right)^2}} \quad (5)$$

This method also does not require advanced data to generate the estimation of the distance between the camera and the photographed object. Again, we need the real approximation h_r of the photographed object. The other required values include the dimensions of the image h_I and w_I , and the object size on the image h_o , expressed in pixels. Finally, the H_m heuristic should be calculated in the manner, as shown in eq. 5.

The EstIII method If the type of the device that has captured a photograph is known (smartphone, compact camera, or other type), the methods EstI and EstII may be enhanced; however, the approximation of the camera's sensor diagonal is required. Let us consider the EstIII approach, which is defined as eq. 6.

$$d = \frac{h_I \cdot h_r}{h_o + H_m} \quad (6)$$

where:

- h_I – image height in pixels;
- w_I – image width in pixels;
- h_r – is the approximation of the real height of the photographed object in millimeters;

- h_o – object height in the image (in pixels);
- D_m – the diagonal of the sensor;
- and:

$$H_m = \frac{D_m}{\sqrt{1 + \left(\frac{h_I}{w_I}\right)^2}} \quad (7)$$

Similarly, as in the previous cases, the EstIII method uses image height and width in pixels, h_I and w_I , respectively, the approximation h_r of the real height of the photographed object in millimeters, and h_o , which is the object height in the image (in pixels). The EstIII requires calculating the heuristic H_m , as described in eq. 7. The most problematic is to obtain or estimate the sensor's diagonal D_m , particularly in the case of the nowadays smartphones, where the detailed parameters of the sensor are not known. The accuracy of the EstIII approach strictly depends on the value D_m . One may estimate the sensor's diagonal, but due to the very large number of used sensors, it might be biased with an error, particularly in the case of advanced cameras, including digital single lens mirrorless (DSLM) or digital single lens reflex (DSLR). Therefore, such an approximation is the easiest to reach in the case of smartphones or compact/sport cameras. The list of D_m estimations is presented in Tab. 1.

Table 1. Estimations of the D_m due to the type of the device

Device type	Approximation of the D_m [mm]
smartphone	4.5–5.5
compact camera	7.5–8.0
sport camera	7.5–8.0
DSLM	21.5–43
DSLR	27.5–43

The proposed heuristics do not rely on information such as distance estimation from a one-time field-of-view calibration or ground-plane homography, which makes them directly applicable in diverse and unconstrained environments. Other lightweight approaches can provide absolute distance estimations but require prior calibration or specific scene conditions.

It should be emphasized that the proposed equations are intended as engineering heuristics inspired by the classical pinhole camera geometry, in which the apparent height of an object decreases inversely with its distance from the camera, rather than as a rigorous geometric derivation. In practice, the device-specific proportionality constant (dependent on focal length and pixel scale) is implicitly absorbed into the ratio (h_I / h_o) , allowing the formulation to remain calibration-free and applicable across different devices. The design of the proposed models deliberately avoids the use of camera intrinsic parameters, focal length, and sensor characteristics in order to maximize applicability when such information is unavailable. Consequently, the method should be regarded as a lightweight empirical approximation whose performance is validated experimentally.

The corrective terms introduced in EstII and EstIII can be interpreted geometrically as simple compensations for perspective effects, such as object tilt or non-frontal orientation, which cause apparent height reduction in the image. These terms improve the estimation accuracy primarily for short and medium distances, where perspective distortion is more significant, while their influence becomes negligible for distant scenes.

4. Experimental evaluation

In this section, we experimentally evaluate the estimation accuracy of the proposed EstI, EstII, and EstIII methods and the camera pinhole model (CPH) state-of-the-art method. We compare the estimations of distance of all the mentioned methods with real distances to check their accuracy. For this purpose, we investigate two common error measures and statistical hypotheses verification. We also introduce a simple one-shot Field of View calibration method, which may be considered as an additional baseline approach.

The scripts for the statistical analysis were implemented in MATLAB R2024b with the support of the Statistics and Machine Learning Toolbox.

4.1. Devices and experiment setup

The devices used in the experiments, as well as their image resolution, are detailed in Table 2. The experiments were conducted with 10 devices at 15 different image resolutions in total, partially from the IMAGINE dataset [43]. Overall, the total number of 141 images was analyzed. All the photographs were shot at the widest angle of the lens. The dataset provides an initial cross-device validation, and the reported errors should be interpreted accordingly.

Table 2. Devices used in the experiments and their image resolutions

Brand	Model	Resolution [px]	Type	Focal length f [mm]
Acer	Liquid Jade S	1792 × 1008	smartphone	3.5
		2560 × 1440		
		2560 × 1920		
		3264 × 2448		
		4096 × 2304		
		4160 × 3120		
Apple	iPhone 5S	3264 × 2448	smartphone	4.15
DJI	Spark	3968 × 2976	drone	4.49
LG	K10	4160 × 2336	smartphone	3.81
NavRoad	myCAM	3264 × 2448	sport camera	2.28
Nikon	P100	3648 × 2736	compact	4.6
Samsung	Galaxy S7	4032 × 3024	smartphone	4.2
Samsung	Galaxy Tab A 10.1	3264 × 1836	tablet	2.91
Samsung	Galaxy Trend 2 Lite	2048 × 1232	smartphone	2.35
Yuneec	Breeze 4K	4160 × 3120	drone	4.4

Figs. 1–3 present sample photographing scenarios. The analyzed object height (in millimeters and pixels) is denoted as h_r and h_o , respectively. For each camera (or each image resolution from a camera), photographs of different objects were taken. The real (or approximate) height of the object was known, and in cases where no reference value was available, h_o was measured manually. Based on these values, the estimation of the distance was calculated according to the proposed methods and CPH.



Fig. 1. Sample photograph used in experiment (Samsung S7). $h_r = 17\text{mm}$, $h_o = 438\text{px}$, $h_r = 3024\text{px}$, the real distance from the camera is 100mm, distance estimated with CPH / EstI method: 96mm / 117mm



Fig. 2. Image (1): Yuneec Breeze 4K drone; $h_r = 50\,000\text{mm}$, $h_o = 1212\text{px}$, $h_l = 3120\text{px}$, real distance from the drone's camera is 98.3m (98310mm), distance estimated by proposed method is 128.7m (128713mm); image (2): image from DJI Spark drone; $h_r = 125\,000\text{mm}$, $h_o = 736\text{px}$, $h_l = 2976\text{px}$, real distance from the drone's camera is about 477m (477000mm), distance estimated by EstI method is 505.43m (505430mm)



Fig. 3. The size of the object (in pixels) according to horizontal (landscape) and vertical (portrait) orientation. $h_{r1} = 1500\text{mm}$; $h_{r2} = 1500\text{mm}$; $h_{r3} = 80\text{mm}$; $h_{r4} = 90\text{mm}$; $h_{o1} = 247\text{px}$; $h_{o2} = 256\text{px}$; $h_{o3} = 290\text{px}$; $h_{o4} = 315\text{px}$

Figure 2 presents sample images from drones. The height of the objects (in both cases, bridges) is known from their documentation. The real distance from the drone's camera to the object was taken on the GPS position of the drone and the object, and calculated by Google Maps. It is worth mentioning that measures of the real distance between a drone and an object can be slightly inaccurate. This inaccuracy can result from both the precision of the GPS module of the drone and the precision of distance calculation in Google Maps. However, we expect that such potential inaccuracies can be considered as negligible. We have also examined the impact of capturing images in horizontal (landscape) and vertical (portrait) orientations. It turns out that an object taken from the same place for both horizontal and vertical orientations differs in the size of pixels in the photo. Examples are presented in Fig. 3.

4.2. One-shot FOV calibration baseline

As a baseline, we introduce a minimal one-shot Field of View (FOV) baseline that requires only a single calibration image for a given device. Let h_r denote the real height of the object in the calibration image, h_o its pixel height, and d the corresponding ground-truth distance. Using the pinhole relation, an effective focal length in pixels is obtained from

$$f = \frac{h_o \cdot d}{h_r} \quad (8)$$

This value is then reused for all other images of the same device, yielding distance estimates:

$$\hat{d}_i = \frac{h_r \cdot f}{h_o^{(i)}} \quad (9)$$

Where $h_o^{(i)}$ is the pixel height of the same real object in image i . The one-shot FOV calibration yields improved accuracy for short distances, where the single-point calibration closely matches the operational range. However, its performance degrades for larger distances, where small errors in the calibration height or object measurement propagate linearly. This confirms the intended role of our calibration-free estimators: they offer more stable behavior across a wider distance range without any device-specific setup. Table 3 summarizes the calibration step and shows the resulting distance estimates for a representative subset of images from one device.

4.3. Individual inspection of the estimations

Sample and partial results for distance estimation can be seen in Tab. 3. Presented are real distances between the camera and the object and estimations of the proposed methods, as well as CPH. Estimates are expressed in millimeters.

Table 3. Sample results for distance estimation for images from Acer, the resolution of 1792×1008 , the d denotes the real distance [mm]

Image	d [mm]	EstI	EstII	EstIII	CPH	FOV
1.jpg	150	121	160	116	150	150
2.jpg	1200	1200	1598	1192	1488	1200
3.jpg	100	79	104	76	97	100
4.jpg	400	281	374	279	349	400
5.jpg	550	723	958	705	896	550
6.jpg	800	697	914	652	864	800
7.jpg	460	413	548	405	512	460
8.jpg	900	779	1037	772	966	900
9.jpg	570	509	676	503	631	570
10.jpg	250	179	239	178	222	250

The results indicate the overall accuracy of the distance estimation of the proposed and CPH methods. The performance may be considered satisfactory. In most cases, the estimations are quite accurate, for instance, for the image 2.jpg, the EstI and EstIII methods obtained the ideal values. On the other hand, 3.jpg and 7.jpg are less accurate, but still acceptable. The estimations for images 4.jpg and 10.jpg for EspII are nearly perfect.

4.4. Analysis of error measures

As quality error evaluation, we use the mean absolute error (MAE), mean absolute percentage error (MAPE), root mean squared error (RMSE), mean signed error (MSD), and coefficient of determination (R^2) to examine the similarity of estimations of the proposed and CPH methods with real distance values. Their definitions are as shown in eq. 10-14.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - x_i| \quad (10)$$

$$MAPE = 100 \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - y_i}{x_i} \right| \quad (11)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2} \quad (12)$$

$$MSD = \frac{1}{n} \sum_{i=1}^n (y_i - x_i) \quad (13)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sum_{i=1}^n (x_i - \bar{y})^2} \quad (14)$$

Where x_i denotes the real distance from the object, and y_i is an estimated distance using the proposed and the CPH methods ($i = 1, 2, \dots$). The closer the value MAPE to 0, the better the accuracy, and the closer the R^2 value to 1, the better fit of x_i and y_i . The results are presented in Tab. 4.

Table 4. Error measures values

Error measure	EstI	EstII	EstIII	CPH	FOV
MAE	4949.70	9253.41	4662.79	12524.48	12637.69
MAPE	18.30	20.60	18.20	19.68	21.12
RMSE	27046.08	50408.35	24434.37	66455.53	66591.64
MSD	-3791.23	-8942.02	-3283.99	-11100	-12300
R^2	0.99	0.99	0.99	0.98	0.97

In many cases, the proposed and state-of-the-art method obtained similar results. Efficiency of all approaches oscillates at the level of 80.0%. The best results are observed in the case of EstI and EstIII, where the MAPE values obtained are 18.30% and 18.20%, respectively. The CPH reached 19.68%, and the EstII – 20.60%. These results confirm that all the methods perform the similar estimation accuracy, which we find satisfactory.

The results in terms of the R^2 measure clearly indicate that all methods represent a high level of estimation accuracy. The R^2 values obtained for all the proposed methods were close to 0.99, indicating a strong correlation between the estimated and reference distances. The state-of-the-art method also achieved a high R^2 value of 0.98, confirming a strong agreement between the estimated and real distances. These results even suggest that the proposed methods are a bit more accurate than the CPH approach.

The evaluation based on MAE, RMSE, and MSD further supports these observations. The lowest MAE and RMSE values are obtained for EstI and EstIII (MAE: 4949.70 mm and 4662.79 mm; RMSE: 27046.08 mm and 24434.37 mm, respectively), indicating that these methods not only produce accurate predictions on average but also maintain low variability in errors. In contrast, EstII, CPH, and FOV exhibit higher MAE and RMSE (MAE: 9253.41 mm and 12524.48 mm; RMSE: 50408.35 mm and 66455.53 mm), showing larger deviations from the real distances.

The MSD values reveal that all methods tend to slightly underestimate the actual distances, with mean signed differences ranging from -3283.99 mm (EstIII) to -12300.00 mm (FOV-baseline). Despite this small systematic underestimation, the combination of low MAE, RMSE, and high R^2 confirms that the proposed methods provide reliable and consistent estimations, outperforming the CPH baseline in most cases.

Table 5. Error measures values per-device

Device	Measure	EstI	EstII	EstIII	CPH	FOV
Acer Liquid Jade S	MAE	82.38	157.90	81.30	93.87	93.87
	MAPE	16.58	18.01	17.19	14.95	14.97
	RMSE	125.82	224.05	122.67	152.86	152.91
	MSD	-26.22	-153.87	-19.00	-59.70	-59.67
	R ²	0.89	0.92	0.90	0.89	0.89
Apple iPhone 5S	MAE	87.30	133.50	87.70	86.80	86.82
	MAPE	13.24	14.75	13.59	13.15	13.20
	RMSE	149.21	221.76	146.01	147.82	147.91
	MSD	-44.30	-133.10	-38.90	-41.60	-41.52
	R ²	0.86	0.86	0.86	0.86	0.93
DJI Spark	MAE	56675.25	105327.25	55283.50	164228.25	164228.34
	MAPE	15.68	16.92	17.13	20.40	20.46
	RMSE	71239.97	144601.87	66543.59	219440.35	219440.43
	MSD	-23157.75	-94662.25	-16149.00	-161446.75	-161446.73
	R ²	0.99	0.99	0.99	0.99	0.99
LG K10	MAE	111.10	130.80	112.30	113.20	113.24
	MAPE	28.09	15.35	28.91	19.96	20.00
	RMSE	146.57	210.87	147.93	148.36	148.41
	MSD	60.50	-97.80	64.50	-4.80	-4.70
	R ²	0.84	0.84	0.84	0.84	0.89
NavRoad myCam	MAE	365.80	498.90	352.40	311.70	311.71
	MAPE	35.29	41.70	34.71	32.40	32.47
	RMSE	454.65	617.09	438.29	388.31	388.33
	MSD	-362.60	-498.90	-348.60	-306.70	-306.65
	R ²	0.98	0.98	0.98	0.98	0.98
Nikon P100	MAE	98.80	148.80	99.30	544.50	544.51
	MAPE	33.74	30.33	34.63	43.68	43.76
	RMSE	137.90	209.42	135.23	670.66	670.69
	MSD	-38.40	-126.40	-33.50	-537.50	-537.49
	R ²	0.87	0.87	0.88	0.87	0.92
Samsung Galaxy S7	MAE	134.40	228.50	128.70	58.60	58.65
	MAPE	17.23	24.96	16.85	14.70	14.73
	RMSE	177.07	291.67	169.41	72.03	72.12
	MSD	-127.00	-228.50	-120.70	-4.40	-4.40
	R ²	0.96	0.96	0.96	0.96	0.99
Samsung Galaxy Tab A 10.1	MAE	106.20	130.40	108.00	151.30	151.34
	MAPE	32.34	17.64	33.44	18.43	18.52
	RMSE	136.12	199.55	138.81	233.83	233.91
	MSD	64.20	-92.20	69.00	-128.10	-128.06
	R ²	0.87	0.87	0.87	0.87	0.87
Samsung Galaxy Trend 2 Lite	MAE	65.60	129.50	71.40	303.70	303.78
	MAPE	19.34	14.61	21.17	30.41	30.46
	RMSE	85.93	194.17	87.68	401.53	401.57
	MSD	23.20	-123.10	32.20	-303.70	-303.70
	R ²	0.94	0.94	0.94	0.94	0.98
Yuneec Breeze 4K	MAE	111571.75	200452.00	101917.75	248771.50	248771.59
	MAPE	20.18	30.73	19.25	35.29	35.36
	RMSE	143814.83	260472.52	128710.94	324702.39	324702.42
	MSD	-111571.75	-200452.00	-101917.75	-248771.50	-248771.49
	R ²	0.98	0.98	0.98	0.98	0.98

The per-device error tables (Table 5) provide a detailed overview of the performance of all methods across different devices. Overall, EstI and EstIII consistently achieve lower absolute errors (MAE) and root mean square errors (RMSE) on most devices, indicating more precise distance estimates. EstII and the CPH baseline tend to exhibit larger deviations, particularly on devices with more challenging imaging conditions, such as the DJI Spark and Yuneec Breeze 4K. The higher RMSE values relative to MAE suggest the presence of occasional large errors in some measurements, which affect the overall error distribution.

The mean signed difference (MSD) values indicate the presence of small systematic biases. Negative MSD values for most methods reveal a tendency to underestimate distances, while a few devices, such as the LG K10 and Samsung Galaxy Tab A 10.1, exhibit slight overestimation. Among the methods, EstI and EstIII show smaller biases compared to EstII, CPH, and FOV, highlighting their improved reliability across devices.

In terms of relative errors (MAPE) and the coefficient of determination (R^2), EstI and EstIII provide the most accurate and consistent predictions, with R^2 values generally ranging from 0.84 to 0.98. This demonstrates strong agreement with real distances across all devices. Consumer smartphones, such as Acer, iPhone, and Samsung models, tend to produce

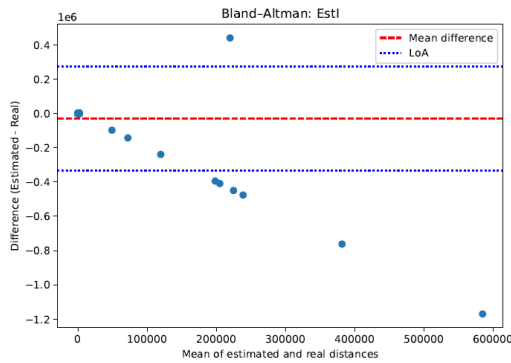
lower absolute errors and smaller biases, whereas specialized devices like drones display higher variability due to larger measurement ranges or less controlled imaging conditions.

Overall, the results indicate that EstI and EstIII provide stable performance through a variety of devices, with lower errors and reduced bias compared to the other evaluated methods. While high correlation with ground-truth distances is observed for most estimators, we note that correlation alone does not imply absolute accuracy and may coexist with scale-related errors over wide distance ranges. In this context, EstII, CPH, and FOV exhibit higher variability and systematic underestimation on several devices, whereas EstI and EstIII show more consistent behaviour across operating conditions.

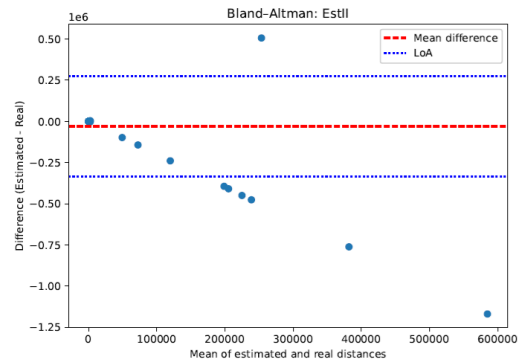
4.5. Bland-Altman plots

The Bland–Altman plots were used to assess agreement between the estimated and real distances. For each method, the differences between estimated and real values were plotted against their mean. The mean difference (MD) indicates whether a method systematically overestimates or underestimates the measurements.

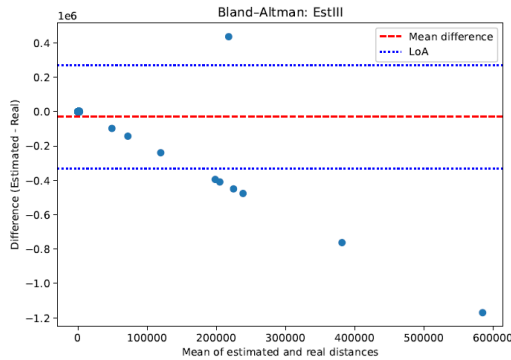
Limits of Agreement (LoA) were calculated as $MD \pm 1.96 \cdot SD$, where SD is the standard deviation of the differences. Under the assumption of normally distributed differences, approximately 95% of the points are expected to lie within these limits. Points outside the LoA highlight extreme deviations or cases where the method is less reliable. These plots allow a clear visualization of both systematic bias and variability in the estimation error (Fig. 4).



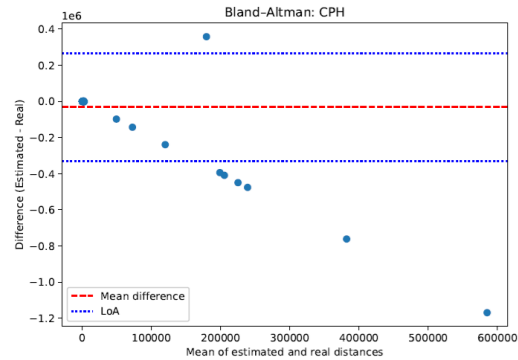
(a) Bland-Altman plot for the EstI method



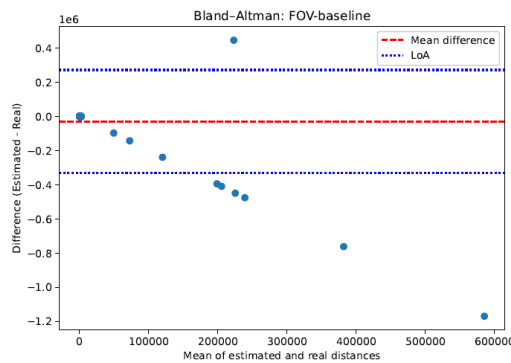
(b) Bland-Altman plot for the EstII method



(c) Bland-Altman plot for the EstIII method



(d) Bland-Altman plot for the CPH method



(e) Bland-Altman plot for the FOV-baseline method

Fig. 4. Bland-Altman plots for all considered methods

From the plots, it can be observed that most points lie within the LoA, indicating that the estimation errors are generally consistent and predictable. EstI and EstIII show smaller spreads around the mean difference, reflecting lower variability in errors, while EstII, CPH, and FOV exhibit a slightly larger dispersion, indicating higher deviations for certain measurements. The negative mean differences across all methods suggest a slight tendency to underestimate distances. Overall, the plots confirm that the proposed methods perform reliably across the range of tested distances.

4.6. Statistical hypotheses verification

In this subsection, we provide the results of the statistical hypotheses verification. In particular, we are interested in whether the real distance values differ significantly from the estimations obtained with the proposed and the CPH methods. This analysis can be used to evaluate the relative performance of the considered methods. The analysis was conducted at a significance level of $\alpha = 0.05$. All statistical tests were performed in MATLAB.

Normality tests The analysis begins with the normality tests in order to check whether the analyzed data follow a normal distribution. We propose to use the Shapiro-Wilk test, whose hypotheses are defined as follows [44]:

- H_0 : The real distances and estimated distances data follows the standard distribution;
- H_1 : The real distances and estimated distances data does not follow the standard distribution.

To accept or reject the null hypothesis, we calculate the Shapiro-Wilk test statistic W and then calculate the p -value. The p -value is compared with the significance level, and if greater, we accept the null hypothesis H_0 . Otherwise, the alternative hypothesis H_1 is accepted:

- If $p \leq \alpha$: H_0 is rejected and H_1 is accepted;
- If $p > \alpha$: there is no reason to reject H_0 .

Alternatively, if W is greater than the test critical value, H_0 may be accepted. The critical value for 141 input samples is equal to 0.97. The results of normality tests are presented in Tab. 6.

Table 6. Normality tests

Method	W	p -value
real distance	0.22	0.00
EstI	0.21	0.00
EstII	0.21	0.00
EstIII	0.21	0.00
CPH	0.20	0.00
FOV	0.22	0.00

The test statistic W is smaller than the critical value (approximately 0.20-0.22), indicating that the null hypothesis cannot be retained. Moreover, the corresponding p -values are below the significance level ($p < 0.05$), providing sufficient evidence to reject the null hypothesis and conclude that the analyzed data do not follow a normal distribution. Moreover, the p -values reached 0.00, so compared with the significance level, it also confirms that the analyzed data does not follow the normal distribution. Therefore, for further analysis, the non-parametric Kruskal-Wallis ANOVA test should be used.

Kruskal-Wallis ANOVA test The non-parametric Kruskal-Wallis ANOVA one-way analysis of variance for ranks is an extension of the U Mann-Whitney test for more than two populations. This test is used for the verification of the differences of the medians in $k > 2$ populations [45,46]. The hypotheses for k populations are defined as follows:

- H_0 : $\theta_1 = \theta_2 = \dots = \theta_k$ (all the medians of real distances and estimated distances are equal);
- H_1 : Not all medians θ_j are equal ($j = 1, 2, \dots, k$).

To accept or reject the null hypothesis, we calculate the test statistic H and then calculate the p -value. Then, the p -value is compared with the significance level, and if greater, we may accept the null hypothesis H_0 . Otherwise, the alternative hypothesis H_1 is accepted:

- If $p \leq \alpha$: H_0 is rejected and H_1 is accepted;
- If $p > \alpha$: there is no reason to reject H_0 .

Alternatively, one may compare the test statistic H with the test critical value. If the test statistic is greater than the critical value, the H_0 should be rejected. The test follows the χ^2 distribution with $k - 1$ degrees of freedom.

In our case, we compare the medians of the sequence representing the real distances with the sequences corresponding to the proposed and state-of-the-art methods. We analyze the 5 populations, so the number of degrees of freedom between groups is equal to $6 - 1 = 5$; the number of degrees of freedom within groups reaches $6 \cdot 141 - 6 = 840$.

Thus, the critical value of the χ^2 distribution for 5 degrees of freedom is equal to 11.07. To accept the null hypothesis, the calculated value of the test statistic H should not be greater than the critical value.

The calculated test statistic $H = 0.11 < 11.07$, so it means that there is no reason to reject H_0 . The p -value is equal to 0.12. It leads to the conclusion that there is no statistical difference between the real distances and estimations generated with the EstI, EstII, EstIII, and the CPH and FOV state-of-the-art methods. For clarity, we may also analyze the mean of ranks, which may be seen in Tab. 7.

Table 7. Mean values of ranks

Method	Mean of ranks
real distance	402
EstI	413
EstII	467
EstIII	408
CPH	445
FOV	406

The obtained results clearly indicate that the means of ranks are similar in cases of the analyzed data. It confirms their high level of similarity. This means that the proposed methods generate approximate distance estimation under minimal assumptions.

4.7. Limitations

Although the proposed methods demonstrate that useful distance estimates can be obtained without access to intrinsic camera parameters, several limitations should be acknowledged. First, the accuracy of our heuristics is lower than that of modern monocular depth estimation approaches based on deep learning. While the reported results are comparable to the classical pinhole model under realistic constraints, they do not reach the precision of learning-based techniques trained on large-scale datasets.

Another limitation concerns the dataset size and diversity. Our evaluation was performed on 141 images collected from 10 consumer devices, which provides a reasonable but still limited basis for generalization. A broader dataset covering more camera models, environmental conditions, and object types would be necessary to confirm the robustness of the methods. Furthermore, the approaches require prior knowledge of the approximate real-world dimensions of at least one object in the scene. This assumption, while often practical, may not be valid in all scenarios, particularly when no reference objects are available.

A potential source of uncertainty is the measurement of the object height (h_r or h_o). In this study, the objects were either of known dimensions (e.g., standardized items) or measured manually, which kept this error small compared to other sources. A more detailed perturbation study could be considered in future work.

Another limitation concerns the accuracy of the ground truth distances. In our dataset, reference values were obtained from GPS readings and online maps, which themselves are subject to measurement uncertainty. Typical GPS errors are on the order of a few meters in horizontal positioning and can be larger in vertical measurements. This uncertainty propagates into the evaluation of our methods and should be considered when interpreting the reported error metrics. Ground truth errors are small relative to the studied distances and do not affect the overall conclusions, but they still add some variability. The estimation error depends on distance and scene type. For ground-level scenes up to about 30–40 m, the proposed estimators remain consistent, while for long-range drone imagery, the accuracy decreases due to the small apparent object size in pixels. To sum up, the ground truth distances obtained from GPS and online maps may include errors of several meters, which contribute to the uncertainty of MAE and RMSE values. The estimators are linearly sensitive to errors in the assumed real-world object height (h_r): small perturbations of $\pm 5 - 10\%$ in h_r produce approximately proportional changes in MAPE and MSD. On the contrary, deep learning-based monocular depth methods and homography-based approaches typically yield lower errors when properly trained or when scene priors are available, but they require substantial training data or explicit calibration/scene markers. Our heuristics are intended as a complementary, low-data alternative for situations where those resources are not available.

Another limitation stems from the simplified geometric formulation adopted by the proposed estimators. The methods intentionally avoid the use of intrinsic camera parameters, such as focal length, sensor size, or pixel pitch, in order to remain calibration-free and applicable to heterogeneous imaging devices. Consequently, differences in optical characteristics between devices are not explicitly modeled and may introduce systematic bias, particularly when cameras with substantially different fields of view or imaging pipelines are compared. The present study partially decreases this effect through evaluation on multiple consumer devices with different image resolutions; however, explicit modeling of intrinsic parameters or device-specific calibration could further improve estimation accuracy and constitutes an interesting direction for future research.

Finally, the present work does not include an extensive robustness analysis with respect to factors such as illumination changes, occlusions, or extreme viewing angles. The discussion of potential applications in robotics, environmental monitoring, and forensics should also be seen as preliminary. Although these domains appear promising, real-world case studies and deployment tests are needed to fully validate the applicability of the proposed methods.

Despite these limitations, the presented heuristics fill an important methodological gap. They provide a lightweight, interpretable, and resource-efficient alternative that remains useful when metadata is missing, unreliable, or deliberately removed, and when advanced data-driven methods cannot be applied.

4.8. Discussion and summary

Experiments showed that the proposed methods reliably estimate the distance between the camera and the object. Proposed methods work well both for small objects photographed from small distances as well as for big objects at long distances. Tests were conducted with cameras that produce images at approximately 25-30 mm focal length, such as smartphones and drones. The changes of the focal length are not taken into account; therefore, the estimation of the distance will not be correct if the focal length changes from 30 millimeters to higher values, which are available in zoom compact cameras or DSLRs. However, smartphones and drones are becoming increasingly popular, and many people have abandoned traditional cameras, so this method can be very useful for these kinds of photographic devices, since they have a limited possibility to change the focal length. Obviously, the obtained results are not precise, but note that although the camera pinhole model requires much more parameters, it also gives approximate results. The required knowledge about the focal length and the size of the object at the camera's sensor is a strong disadvantage of the CPH method. The focal length is usually stored as EXIF metadata, which can be easily manipulated or deleted. The information about the camera's sensor size, which is used to determine the size of the object, should be given by camera manufacturers. Unfortunately, most, especially smartphone manufacturers, do not disclose about precise type and size of the sensor. The proposed methods do not require the aforementioned information.

The proposed heuristics are deliberately simple and are not intended to compete with advanced deep learning-based monocular depth estimation methods. Instead, they are designed as lightweight, geometry-inspired engineering approximations that can estimate object distance without requiring camera calibration, intrinsic parameters, or image metadata. This makes them suitable for practical scenarios where such information is unavailable, for example when analyzing images acquired from heterogeneous mobile devices or multimedia evidence in forensic investigations. The trade-off for this simplicity is reduced accuracy in challenging imaging conditions, particularly when strong lens distortion or non-ideal imaging geometry is present.

The proposed heuristics, despite their simplicity, can be useful in scenarios where camera metadata is missing or unreliable. Potential applications include low-cost mobile robotics (e.g., obstacle approximation using a known object), environmental monitoring (estimating distances to reference markers or vegetation), or forensic imaging, where metadata is often stripped intentionally. While a detailed case study lies beyond the scope of this work, these examples illustrate the practical utility of lightweight distance estimation in diverse fields.

For reference, we compared our estimator with a simple single-device FOV calibration baseline. The calibrated variant reduced bias by approximately 10%, at the cost of requiring a one-time setup. Compared to monocular depth networks, our approach trades absolute accuracy for transparency and independence from training data, which is beneficial in constrained or forensic settings.

While the methods inherit the limitations of heuristic approaches and are less accurate than state-of-the-art learning-based techniques, they fill an important methodological gap. Specifically, they provide a transparent, computationally inexpensive alternative that remains applicable when data-driven methods are infeasible due to resource constraints or lack of training data.

The potential forensic use cases assume the scenario where only a single image is available and calibration is lacking; however, real-world forensic deployment would require dedicated validation and adherence to chain-of-custody practices. Therefore, the proposed methods are intended for exploratory forensic scenarios involving single images without access to camera calibration data.

5. Conclusions and future work

In this paper, an approach for estimating the distance between the camera and the photographed object was proposed. The main objective was to estimate the distance only from a single photograph, without the use of such camera parameters as focal length or image sensor format. The experiments were conducted on a representative set of images from different devices, such as smartphones, compact cameras, and drones. The statistical analysis confirmed the comparable efficiency with the state-of-the-art approach. Therefore, the proposed methods can be successfully used for distance estimation for devices with no possibility to change the focal length, like smartphones or drones.

The proposed approach also has several limitations. Its accuracy depends on the quality of object detection, the estimation of the reference height, image resolution, and the imaging geometry, with performance generally deteriorating for long object distances and in the presence of strong lens distortion. Therefore, the proposed estimators should be

regarded as lightweight engineering approximations intended for scenarios where camera calibration and intrinsic parameters are unavailable.

Future work will focus on improving robustness across different camera types by incorporating simplified camera models, accounting for lens characteristics, and evaluating the proposed heuristics on larger and more diverse datasets.

All the Declarations and Statements

Author Contributions Statement

Jarosław Bernacki – Conceptualization, Data Curation, Methodology, Supervision, Software Implementation, Formal Analysis, Visualization, Statistical Analysis, Writing – the original draft, Writing – Review and Editing.

Conflict of Interest Statement

The authors declare no conflicts of interest.

Funding Declaration

This research received no funding.

Data Availability Statement

Data will be available on request.

Ethical Declarations

No human subjects and/or animals were involved in the studies.

Acknowledgments

We sincerely thank the experts for their professional evaluation and valuable recommendations, which have contributed to improving the quality of the experiment and the reliability of its results.

Declaration of Generative AI in Scholarly Writing

The generative AI was used for language polishing.

Abbreviations

None.

Appendix A/B/C..., with appendix title

None.

References

- [1] Mo Chen, Jessica Fridrich, Miroslav Goljan, and Jan Lukas. Determining image origin and integrity using sensor noise. *IEEE Transactions on information forensics and security*, 3(1):74–90, 2008. DOI: 10.1109/TIFS.2007.916285
- [2] Miroslav Goljan. Digital camera identification from images–estimating false acceptance probability. In *International workshop on digital watermarking*, pages 454–468. Springer, 2008. DOI: 10.1007/978-3-642-04438-0_38
- [3] Chiara Galdi, Michele Nappi, and Jean-Luc Dugelay. Multimodal authentication on smartphones: Combining iris and sensor recognition for a double check of user identity. *Pattern Recognition Letters*, 82:144–153, 2016. An insight on eye biometrics. DOI: 10.1016/j.patrec.2015.09.009
- [4] Chiara Galdi, Michele Nappi, and Jean-Luc Dugelay. Combining hardwaremetry and biometry for human authentication via smartphones. In Vittorio Murino and Enrico Puppo, editors, *Image Analysis and Processing — ICIAP 2015*, pages 406–416, Cham, 2015. Springer International Publishing. DOI: 10.1007/978-3-319-23234-8_38
- [5] Ugur Baysal and Gokhan Sengul. Single camera photogrammetry system for eeg electrode identification and localization. *Annals of biomedical engineering*, 38:1539–1547, 2010. DOI: 10.1007/s10439-010-9950-4
- [6] Jan Lukas, Jessica Fridrich, and Miroslav Goljan. Digital camera identification from sensor pattern noise. *IEEE Transactions on Information Forensics and Security*, 1(2):205–214, 2006. DOI: 10.1109/TIFS.2006.873602
- [7] Nitin Khanna, Aravind K Mikkilineni, George TC Chiu, Jan P Allebach, and Edward J Delp. Scanner identification using sensor pattern noise. In *Security, steganography, and watermarking of multimedia contents IX*, volume 6505, pages 563–573. SPIE, 2007. DOI: 10.1117/12.705837
- [8] Arafat Al-Dhaqm, Richard Adeyemi Ikuesan, Victor R Kbande, Shukor Abd Razak, George Grispos, Kim-Kwang Raymond Choo, Bander Ali Saleh Al-Rimy, and Abdulrahman A Alsewari. Digital forensics subdomains: the state of the art and future directions. *IEEE Access*, 9:152476–152502, 2021. DOI: 10.1109/ACCESS.2021.3124262

- [9] Fran Casino, Thomas K Dasaklis, Georgios P Spathoulas, Marios Anagnostopoulos, Amrita Ghosal, Istvan Borocz, Agusti Solanas, Mauro Conti, and Constantinos Patsakis. Research trends, challenges, and emerging topics in digital forensics: A review of reviews. *IEEE Access*, 10:25464–25493, 2022. DOI: 10.1109/ACCESS.2022.3154059
- [10] Jean-Paul A Yaacoub, Hassan N Noura, Ola Salman, and Ali Chehab. Advanced digital forensics and anti-digital forensics for iot systems: Techniques, limitations and recommendations. *Internet of Things*, 19:100544, 2022. DOI: 10.1016/j.iot.2022.100544
- [11] Christian Brauer-Burchardt, Matthias Heinze, Christoph Munkelt, Peter Kuhmstedt, Gunther Notni, and IOF Fraunhofer. Distance dependent lens distortion variation in 3d measuring systems using fringe projection. In *BMVC*, page 327, 2006. DOI: 10.5244/C.20.34
- [12] C Brown Duane. Close-range camera calibration. *Photogramm. Eng*, 37(8):855–866, 1971.
- [13] Arthur A Magill. Variation in distortion with magnification. *Journal of the Optical Society of America*, 45(3):148–152, 1955. DOI: 10.1364/JOSA.45.000148
- [14] John G Fryer and Duane C Brown. Lens distortion for close-range photogrammetry. *Photogrammetric engineering and remote sensing, International Archives of Photogrammetry and Remote Sensing Volume*: 26(5):30-37, 1986.
- [15] Richard Hartley and Andrew Zisserman. Multiple view geometry in computer vision. Cambridge university press, 2003. DOI: 10.1017/CBO9780511811685
- [16] David A Forsyth and Jean Ponce. Computer vision: a modern approach. Pearson Education, Prentice Hall professional technical reference, ISBN-10: 0273764144, 2012.
- [17] Juho Kannala and Sami S Brandt. A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. *IEEE transactions on pattern analysis and machine intelligence*, 28(8):1335–1340, 2006. DOI: 10.1109/TPAMI.2006.153
- [18] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11):1330–1334, 2000. DOI: 10.1109/34.888718
- [19] Peter Sturm. Pinhole camera model. In *Computer Vision: A Reference Guide*, pages 983–986. Springer, 2021. DOI: 10.1007/978-3-030-63416-2_472
- [20] Nicholas J Wade and Stanley Finger. The eye as an optical instrument: from camera obscura to helmholtz’s perspective. *Perception*, 30(10):1157–1177, 2001. DOI: 10.1068/p3210
- [21] Krzysztof Murawski. Method of measuring the distance to an object based on one shot obtained from a motionless camera with a fixed-focus lens. *Acta Physica Polonica A*, 127(6):1591–1595, 2015. DOI: 10.12693/APhysPolA.127.1591
- [22] Carles Matabosch, David Fofi, Joaquim Salvi, and Josep Forest. Registration of moving surfaces by means of oneshot laser projection. In *Pattern Recognition and Image Analysis: Second Iberian Conference, IbPRIA 2005, Estoril, Portugal, June 7-9, 2005, Proceedings, Part I 2*, pages 145–152. Springer, 2005. DOI: 10.1007/11492429_18
- [23] Daniil A Loktev and Alexey A Loktev. Estimation of measurement of distance to the object by analyzing the blur of its image series. In *2016 International Siberian Conference on Control and Communications (SIBCON)*, pages 1–6. IEEE, 2016. DOI: 10.1109/SIBCON.2016.7491683
- [24] Masahiro Kawakita, Keigo Iizuka, Tahito Aida, Hiroshi Kikuchi, Hideo Fujikake, Jun Yonai, and Kuniharu Takizawa. Axi-vision camera (real-time distance-mapping camera). *Applied Optics*, 39(22):3931–3939, 2000. DOI: 10.1364/AO.39.003931
- [25] Rajesh Kannan Megalingam, Vignesh Shriram, Bommu Likhith, Gangireddy Rajesh, and Sriharsha Ghanta. Monocular distance estimation using pinhole camera approximation to avoid vehicle crash and back-over accidents. In *2016 10th international conference on intelligent systems and control (ISCO)*, pages 1–5. IEEE, 2016. DOI: 10.1109/ISCO.2016.7727017
- [26] Scott Leorna, Todd Brinkman, and Timothy Fullman. Estimating animal size or distance in camera trap images: Photogrammetry using the pinhole camera model. *Methods in Ecology and Evolution*, 13(8):1707–1718, 2022. DOI: 10.1111/2041-210X.13880Digital Object Identifier (DOI)
- [27] Dany Eka Saputra, Aloysius Shendy Mulyana Senjaya, Joshua Ivander, and Alex William Chandra. Experiment on distance measurement using single camera. In *2021 4th International Conference on Information and Communications Technology (ICOIACT)*, pages 80–85. IEEE, 2021. DOI: 10.1109/ICOIACT53268.2021.9564010
- [28] Ka Seng Chou, Teng Lai Wong, Kei Long Wong, Lu Shen, Davide Aguiari, Rita Tse, Su-Kit Tang, and Giovanni Pau. A lightweight robust distance estimation method for navigation aiding in unsupervised environment using monocular camera. *Applied Sciences*, 13(19):11038, 2023. DOI: 10.3390/app131911038
- [29] Fatih Gokce, Gokturk Ucoluk, Erol Sahin, and Sinan Kalkan. Vision-based detection and distance estimation of micro unmanned aerial vehicles. *Sensors*, 15(9):23805–23846, 2015. DOI: 10.3390/s150923805
- [30] Gafencu Natanael, Cristian Zet, and Cristian Fosalau. Estimating the distance to an object based on image processing. In *2018 International Conference and Exposition on Electrical and Power Engineering (EPE)*, pages 0211–0216. IEEE, 2018. DOI: 10.1109/ICEPE.2018.8559642
- [31] Clement Godard, Oisin Mac Aodha, Michael Firman, and Gabriel Brostow. Digging into self-supervised monocular depth estimation, 2019. DOI: 10.48550/arXiv.1806.01260
- [32] Rene Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer, 2020. DOI: 10.48550/arXiv.1907.01341
- [33] Hong Liang, Zizhen Ma, and Qian Zhang. Self-supervised object distance estimation using a monocular camera. *Sensors*, 22(8), 2022. DOI: 10.3390/s22082936
- [34] Armin Masoumian, David G.F. Marei, Saddam Abdulwahab, Julian Cristiano, Domenec Puig, and Hatem A. Rashwan. Absolute Distance Prediction Based on Deep Learning Object Detection and Monocular Depth Estimation Models. *IOS Press*, October 2021. DOI: 10.3233/FAIA210151
- [35] Rene Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. Vision transformers for dense prediction, 2021. DOI: 10.1109/ICCV48922.2021.01196
- [36] Rafael Hidalgo, Aparna S. Varde, Jesse Parron, and Weitian Wang. Incorporating commonsense knowledge to enhance robot perception. *IEEE Transactions on Automation Science and Engineering*, 22:15488–15501, 2025. DOI: 10.1109/TASE.2025.3565191

- [37] Jan Pauls, Max Zimmer, Una M. Kelly, Martin Schwartz, Sassan Saatchi, Philippe Ciais, Sebastian Pokutta, Martin Brandt, and Fabian Gieseke. Estimating canopy height at scale, 2024. Proceedings of the 41st International Conference on Machine Learning, PMLR 235:39972-39988, 2024. DOI: 10.48550/arXiv.2406.01076
- [38] Aparna Varde, Elke Rundensteiner, Giti Javidi, Ehsan Sheybani, and Jianyu Liang. Learning the relative importance of features in image data. In 2007 IEEE 23rd International Conference on Data Engineering Workshop, pages 237–244, 2007. DOI: 10.1109/ICDEW.2007.4400998
- [39] Hongxiang Jiang, Jihao Yin, Qixiong Wang, Jiaqi Feng, and Guo Chen. Eaglevision: Object-level attribute multimodal llm for remote sensing, 2025. DOI: 10.48550/arXiv.2503.23330
- [40] Blessing Austin-Gabriel, Cristian Noriega Monsalve, and Aparna S. Varde. Power plant detection for energy estimation using gis with remote sensing, cnn & vision transformers, 2024. DOI: 10.48550/arXiv.2412.04986
- [41] Tim Havers, Lukas Masur, Eduard Isenmann, Stephan Geisler, Christoph Zinner, Billy Sperlich, and Peter Deking. Reproducibility and quality of hypertrophy-related training plans generated by gpt-4 and google gemini as evaluated by coaching experts. *Biology of Sport*, 42(2):289–329, 2025. DOI: 10.5114/biolsport.2025.145911
- [42] Zeying Gong, Rong Li, Tianshuai Hu, Ronghe Qiu, Lingdong Kong, Lingfeng Zhang, Yiyi Ding, Leying Zhang, and Junwei Liang. Stairway to success: Zero-shot floor-aware object-goal navigation via llm-driven coarse-to-fine exploration, 2025. DOI: 10.48550/arXiv.2505.23019
- [43] Jarosław Bernacki and Rafał Scherer. Imagine dataset: Digital camera identification image benchmarking dataset. In Proceedings of the 20th International Conference on Security and Cryptography – SECRYPT, pages 799–804. INSTICC, SciTePress, 2023. DOI: 10.5220/0012130300003555
- [44] Samuel Sanford Shapiro and Martin B Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965. DOI: 10.1093/biomet/52.3-4.591
- [45] WH Kruskal and WA Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, pages 583–621, 1952. DOI: 10.2307/2280779
- [46] William H. Kruskal. A Nonparametric test for the Several Sample Problem. *The Annals of Mathematical Statistics*, 23(4):525 – 540, 1952. DOI: 10.1214/aoms/1177729332

Authors' Profiles



Jarosław Bernacki is an assistant professor at Wrocław University of Science and Technology (Poland) in the Department of Computer Science and Systems Engineering (Faculty of Information and Communication Technology). His main research interests include digital forensics (in particular digital camera identification), hardwaremetry and computational photography. He is also interested in image processing, deep learning methods, air quality, and sensors networks.

How to cite this paper

Jarosław Bernacki, "Lightweight Distance Estimation from Single Images without Camera Intrinsics", *International Journal of Wireless and Microwave Technologies(IJWMT)*, Vol.16, No.4, pp. 346-362, 2026. DOI:10.5815/ijwmt.2026.04.20