

Perspective Transform-based Depth Estimation of Monocular Camera for Electrocutation Threat Determination of Construction Machinery

Tong Yang¹, Shuangfeng Wei¹, Liang Fan², Lei Zhang²

¹ School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, 102616 Beijing, China - 2108570022132@stu.bucea.edu.cn, weishuangfeng@bucea.edu.cn

² Guangzhou Zhongke Zhixun Technology Co., 510000 Guang Zhou, Guang Dong, China - fly6125@163.com, zhanglei@igiai.com

Keywords: Depth Estimation, Monocular Imagery, Perspective Projection, Point Cloud, Construction Safety.

Abstract

This paper discusses the application of perspective transform-based monocular camera depth estimation method in monitoring the safety distance between construction machinery and power lines. The traditional distance measurement relies on manual operation, which is inefficient and not precise enough, while the monocular depth estimation method based on deep learning can improve the accuracy, but faces high equipment cost and relies on a large amount of training data. In contrast, the depth estimation method based on perspective transformation proposed in this paper utilizes mathematical and physical principles to establish a mathematical model between the pixel distances of an image and the actual scene distances through perspective projection theory, which simplifies the hardware requirements and reduces the data dependence and improves the computational efficiency and applicability. Through experimental validation under different construction scenes and exposure conditions, the method demonstrates high accuracy and robustness, proving its practicality in construction safety management. This research provides a new technical direction and practical application possibility in the field of intelligent monitoring and 3D reconstruction.

1. Introduction

As China stands as the globe's preeminent electricity consumer, it boasts an expansive lattice of transmission lines, underscoring the paramount importance of their secure operation to protect urban functionalities (Shu, Y., and Chen, W., 2018). The relentless urbanization push renders engineering and construction endeavors in the proximity of these lines increasingly common, heightening the necessity to meticulously regulate the safe separation between construction machinery and power lines. Operating equipment too near these lines can precipitate safety mishaps, endangering construction workers' welfare and potentially inflicting substantial harm on the adjacent ecosystems and electrical infrastructures (Kiessling, F., Nefzger, P., Nolasco, J. F., 2014). Consequently, developing technology to monitor the distance between construction machinery and transmission lines in real-time with accuracy (Rao, A. S., Radanovic, M., Liu, Y., 2022) is of great practical significance and holds substantial application value.

Traditional methodologies for measuring distance are predominantly manual, marked by inefficiency and a lack of precision, rendering them unfit for the instantaneous monitoring demands of construction sites (Mu, W., Tong, D., 2020). While three-dimensional (3D) reconstruction technologies through binocular vision present a prospect for accurate distance gauging, practical implementations confront obstacles like elevated equipment expenditures and intricate calibration requisites (Kazerouni, I. A., Fitzgerald, L., Dooly, G., 2022). In contrast, monocular cameras, employing a singular imaging sensor to procure visuals akin to the unaided human perspective, mitigate hardware intricacy and financial outlay. Nevertheless, this modus operandi struggles with directly acquiring a scene's depth data (Macario Barros, A., Michel, M., Moline, Y., 2022).

Scholars have crafted a myriad of sophisticated techniques to deduce depth from single-lens imagery via deep learning, each distinguished by its architectural framework and foundational tenets. Certain approaches harness a comprehensive

Convolutional Neural Network (LeCun, Y., Bottou, L., Bengio, Y., 1998), which, through rigorously intensive deep learning regimens, educates these models to interpret the intricate correlations between imagery and its associated depth. The study (Fang, Z., Chen, X., Chen, Y., 2020) probes into the pivotal elements that influence monocular depth perception, scrutinizing a spectrum of encoder-decoder schemes, as well as fully supervised and autonomous supervisory losses. The methodology they introduced is lauded for its exceptional performance against the KITTI benchmark and its pioneering results in the NYU Depth v2 compendium. Nevertheless, it is prone to inaccuracies caused by fluctuations in luminosity and variable environments, yielding unreliable depth readings, notably in external landscapes or scenes under substantial light alterations. Complementary research directs its endeavors towards weaving additional contextual intelligence into the fabric of the depth approximation process, frequently leveraging multi-faceted learning paradigms to refine the model's interpretive acumen regarding depth data. The manuscript (Chen, Y., Zhao, H., Hu, Z., 2021) introduces a pioneering approach: a supervised, self-attention-oriented context aggregation network (ACAN) dedicated to monocular depth estimation. This innovative network is designed to dynamically assimilate task-oriented resemblances among disparate pixels, thereby sculpting a seamless tapestry of contextual data. Trials conducted on a recognized public benchmark for monocular depth estimation reveal that ACAN outstrips its predecessors in efficacy. Notwithstanding, its performance is overly tethered to the idiosyncrasies of the training corpus, which results in suboptimal applicability to novel or diverse data compilations.

To encapsulate, the endeavor to gauge depth through a monocular lens with deep learning techniques remains fraught with complexities. The article (Lahiri, S., Ren, J., Lin, X., 2024) articulate that, foremost, these deep learning constructs demand extensive volumes of annotated data for efficacious training, a pursuit that is inherently laborious and protracted. The research (Talaie Khoei, T., Ould Slimane, H., Kaabouch, N. 2023) also necessitates formidable computational power and considerable

durations for processing, which erects significant barriers to their deployment in real-time scenarios or on devices beset by resource constraints (Cremer, C. Z., 2021).

Addressing the challenges, the present study endeavors to investigate the application of mathematical and physical postulates (Tobler, W. R., 1963) to facilitate depth perception via a monocular camera, translating two-dimensional visual data into a 3D spatial understanding. This methodology aspires to enable real-time, precise assessments of the proximity between construction equipment and power lines, thereby elevating safety standards on construction sites. In contrast with deep learning paradigms, the technique proposed herein necessitates a leaner array of parameters and is anchored in straightforward geometrical and optical doctrines, proffering a distinct edge within engineering realms where computational capabilities are scarce. Furthermore, this approach exhibits a reduced dependency on voluminous training datasets and obviates the need for intricate, extensively trained neural networks, thus proving advantageous in scenarios where accumulating a copious amount of labeled data presents a formidable hurdle.

2. Methodology

In this research, the internal parameters of the camera apparatus is calibrated initially, proceeding to ascertain the extrinsic parameters through the deployment of the EPnP algorithm (Lepetit, V., Moreno-Noguer, F., Fua, P. 2009). After this calibration, compute the horizontal separation which is denoted as the depth value from the construction equipment to the camera, employing the mathematical framework devised. Utilizing these depth values, the pixel coordinates of the construction apparatus are reverse mapped onto the coordinates of the point cloud (Chen, K., Lai, Y. K., Hu, S. M., 2015). In advancing the analysis, to ascertain the distance between the construction equipment and the power lines, segment the power line point cloud from the scene point cloud applying the Principal Component Analysis algorithm to formulate the Oriented Bounding Box (O'Rourke, J., 1985). In computing process, derive the Euclidean distance between the construction equipment and the OBB of electrical power lines.

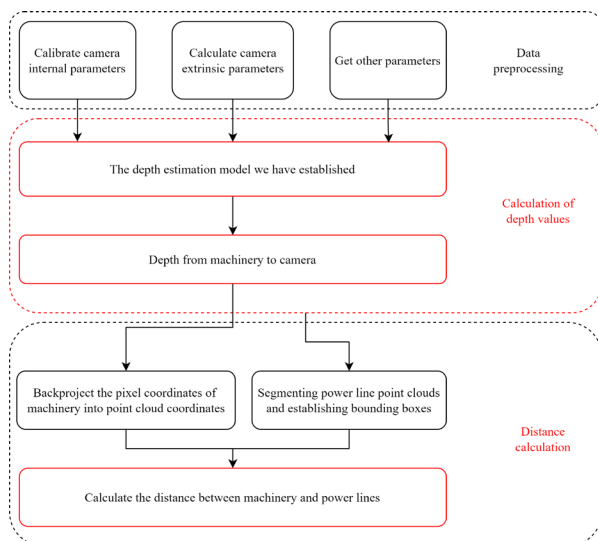


Figure 1. A block diagram illustrating the full pipeline of the proposed algorithm.

The depth calculation method this manuscript proposed establishes a mathematical model for the mathematical relationship between the pixel distances of two target points on the image and the distances in the real scene according to the

perspective projection theory, and the modeling process is as follows: The process of mapping points in 3D space into a camera image can be viewed as the projection of a 3D Euclidean space \mathbb{R}^3 onto a flat canvas S . First denote the mapping of the projection of the three-dimensional Euclidean space \mathbb{R}^3 onto the two-dimensional sphere S^2 as $f: \mathbb{R}^3 \mapsto S^2$, which is also known as the Apparent Spherical Projection (Flores, M., Valiente, D., Peidr , A., 2024). Then map the image on the optic sphere onto a planar canvas, denoted as a continuous injective: $g: S \mapsto U$, where S and U are the open sets in S^2 and \mathbb{R}^2 , respectively. Eq. (1) is the more common canvas projection, called the Spherical-Axis Plane Projection (Araujo, A. B., 2021):

$$\lambda\pi_\mu: \{S^2 \mid \theta < \arccos \mu\} \mapsto \mathbb{R}^2 \Rightarrow$$

$$(\theta, \varphi) \mapsto \lambda \frac{(1-\mu)\sin\theta}{\cos\theta-\mu} (\cos\varphi, \sin\varphi), \quad (1)$$

where $\mu =$ axis parameter, $-1 \leq \mu < 1$
 $\lambda =$ scaling factor, $\lambda > 0$

when $\mu = 0$, it is the Spherical Center Projection:

$$\lambda\pi_0: \{S^2 \mid \theta < \frac{\pi}{2}\} \mapsto \mathbb{R}^2 \Rightarrow$$

$$(\theta, \varphi) \mapsto \lambda \tan\theta (\cos\varphi, \sin\varphi), \quad (2)$$

Compounding the Apparent Spherical Projection with the Spherical Center Projection introduces:

$$\lambda\pi_0 \circ f: \begin{cases} x = r \sin\theta \cos\varphi \\ y = r \sin\theta \sin\varphi \\ z = r \cos\theta \end{cases} \Rightarrow$$

$$\begin{cases} x' = \lambda \tan\theta \cos\varphi \\ y' = \lambda \tan\theta \sin\varphi \end{cases} \mapsto \begin{cases} x' = \lambda \frac{x}{z} \\ y' = \lambda \frac{y}{z} \end{cases}, \quad (3)$$

Assume that the linear equations of the two markers determined in space are:

$$\begin{cases} x = x_0 + t \sin \alpha \cos \beta \\ y = y_0 + t \sin \alpha \sin \beta \\ z = z_0 + t \cos \alpha \end{cases} \quad (4)$$

where $x_0, y_0, z_0 =$ the coordinates of the first target
 $\alpha, \beta =$ spherical coordinates in the linear direction
 $t =$ parameter, indicating the distance from the point on the line to (x_0, y_0, z_0)

Projecting this line onto the canvas, one can introduce:

$$\begin{cases} x' = \lambda \tan \alpha \cos \beta + \lambda \frac{x_0 - z_0 \cos \beta \tan \alpha}{z_0 + t \cos \alpha} \\ y' = \lambda \tan \alpha \sin \beta + \lambda \frac{y_0 - z_0 \sin \beta \tan \alpha}{z_0 + t \cos \alpha} \end{cases}, \quad (5)$$

As $t \rightarrow +\infty$, it follows that:

$$\begin{cases} x' \rightarrow \lambda \tan \alpha \cos \beta \\ y' \rightarrow \lambda \tan \alpha \sin \beta \end{cases} \Rightarrow$$

$$(x', y') \rightarrow \lambda\pi_0[(\theta = \alpha, \varphi = \beta)], \quad (6)$$

When $\alpha = \frac{\pi}{2}$, (x', y') can be considered to tend to an infinity point. Let the distance the point on this line passes on the canvas after a distance ξ from (x_0, y_0, z_0) be $L(\xi)$. Then it can be deduced that:

$$L(\xi)^2 = [x'(\xi) - x'(0)]^2 + [y'(\xi) - y'(0)]^2$$

$$= \sigma \lambda^2 \left(\frac{1}{z_0 + \cos \alpha} - \frac{1}{z_0} \right)^2, \quad (7)$$

where

$$\sigma = x_0^2 + y_0^2 + z_0^2 \tan^2 \alpha - 2z_0 \tan \alpha (x_0 \cos \beta + y_0 \sin \beta)$$

Then there is:

$$L(\xi) = \frac{\sigma \lambda}{|z_0| z_0 \sec \alpha + \xi} \quad (8)$$

Let the proportion of $L(\xi)$ to the full length of the image in the canvas be $S(\xi)$. Since $L(\infty) = \frac{\sigma \lambda}{z_0}$, then there is:

$$S(\xi) = \frac{L(\xi)}{L(\infty)} = \frac{\xi}{z_0 \sec \alpha + \xi} \quad (9)$$

Assume that the pixel coordinate of any point of the image is (u, v) , the difference between its vertical coordinate and the coordinate of the first target pixel, (u_1, v_1) , is $\Delta v = v_1 - v$. The vanishing point (u_0, v_0) of the line where the target is located in the image can be calculated from Eq. (6), and the difference between the vertical coordinates of this point and the coordinates of the first target pixel is $\Delta v' = v_1 - v_0$. From Eq. (9), $S(\xi) = \frac{\Delta v}{\Delta v'}$, where ξ , which is the horizontal distance between the 3D coordinates of the image point (u, v) in the actual scene and the first marker along the camera line-of-sight direction, can be found.

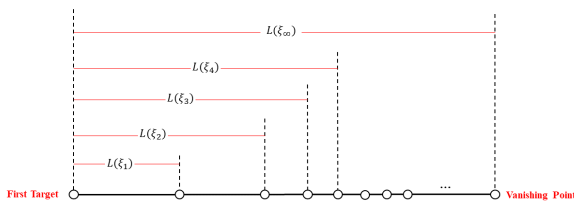


Figure 2. Layout diagram of experimental scene $L(\xi_i)$.

3. Experiments and Analysis

To validate the effectiveness of the proposed method, this study uses a single image captured by a monocular camera as input and evaluates its performance by calculating the distance between the construction machinery and the power line. To ensure the wide applicability of the experimental results, the researchers conducted experiments under several different construction scenarios. By analyzing the data obtained under these scenarios, the stability and accuracy of the proposed method can be verified under different environmental conditions. The experimental results show that the method has high accuracy and robustness on

the monocular camera depth estimation task. This research not only provides an important safety reference for construction machinery operators, but also provides new ideas and methods for researchers in related fields.

3.1 Experimental Scenario

A configuration was orchestrated at the construction site, as depicted in Figure 3. Firstly, the camera was fixed on the pole and its angle was adjusted so that the field of view of the camera could cover the measurement area. Then a number of targets are placed along the direction of the camera's line of sight, and the arrangement of the targets should follow the following conditions: 1) the targets should be placed along the direction of the camera's main optical axis, i.e., to ensure that the targets are in the same straight line with the camera's line of sight; 2) the bottom boundary of the first target is connected to the bottom edge of the image. After the camera is set up, a laser scanner is used to scan the scene in order to obtain the positional coordinates of the camera and the target, as well as the horizontal distance from the first target to the camera D_0 . According to the mathematical model established in this paper, the horizontal distance from the bottom point of the construction machinery along the direction of the camera realization to the first target can be estimated to get the depth of the point $Depth = \xi + D_0$.

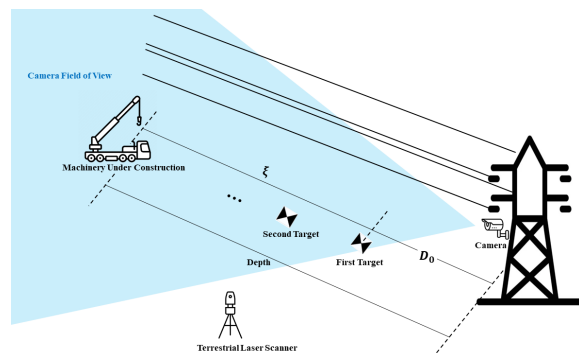


Figure 3. Layout diagram of experimental scene.

In order to evaluate the effect of exposure level on the image recognition algorithm, and to verify the adaptability and robustness of this paper's algorithm in practical applications, researchers select three different scenes of construction machinery from near and far, and consider the image Normal Exposure (NE), Overexposure (OE), and Underexposure (UE), as shown in Table 1, in order to simulate the lighting problems that may be encountered in the actual shooting.

Scene	Normal Exposure	Overexposure	Underexposure
Scene 1			



Table 1. Scene images under different exposures.

3.2 Depth Values Calculation and Error Analysis

The purpose of this experiment is to verify the effectiveness and accuracy of the proposed new monocular camera depth estimation algorithm. To evaluate the algorithm's wide-ranging applicability, it underwent rigorous testing across a spectrum of environments varying in scale and intricacy. By estimating the relative percentage error between the depth values and the true depth values, the accuracy of the depth estimation was able to be quantified and an objective evaluation of the performance of this method was provided accordingly.

In Table 2, the depth values of construction machinery to the camera as gauged, are systematically presented in ascending order from small to large. Utilizing the open-source point cloud processing software CloudCompare (Daniel Girardeau-Montaut, 2024), meticulously measured and documented these depth values to serve as the ground truth. Subsequently, our proposed depth estimation algorithm was implemented to calculate the depth values between the construction machinery and the camera across these three scenarios. This approach enabled a comparative analysis of the algorithmically determined depth values against the actual depth measurements, thereby facilitating the evaluation of our algorithm's performance and accuracy.

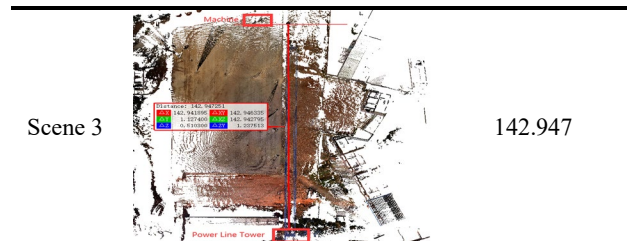
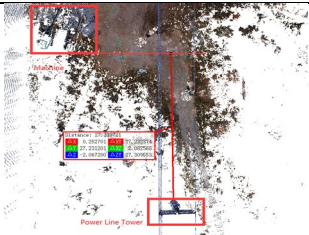
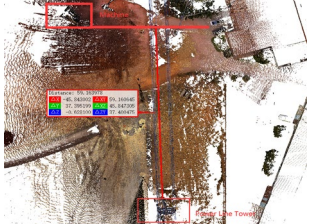


Table 2. The true depth from mechanical to camera in three scenarios.

Scene	Point cloud of scene	True depth(m)
Scene 1		27.310
Scene 2		59.164

Experimental outcomes are meticulously delineated in Table 3, illustrating the performance of the depth estimation between the construction machinery and the camera across various distance conditions. The depth estimation technique proposed herein demonstrates exceptional accuracy within all experimental frameworks, maintaining relative errors below 5%. As the magnitude of the experimental environments escalates, there is a discernible augmentation in the absolute error, escalating from approximately 1.218 meters in Scene 1 to roughly 6.771 meters in Scene 3. Concurrently, a marginal increase in relative error is observed, moving from approximately 4.5% in Scene 1 to about 4.7% in Scene 3.

Scene	Our Method(m)	Error(m)	Relative error(%)
Scene 1	28.528	1.218	4.456
Scene 2	56.447	2.717	4.592
Scene 3	136.176	6.771	4.737

Table 3. The absolute errors and relative errors of depth estimation algorithm proposed.

The insights conveyed through the data visualization in Figure 4 offer an intuitive perspective, revealing that notwithstanding the significant escalation in scene scale from Scene 2 to Scene 3, which culminates in an absolute error increment of approximately 5.553 meters, the fluctuation in relative error remains minimal. This outcome underscores the commendable stability of our proposed depth estimation approach across varying scene magnitudes, illustrating the algorithm's resilience against scale modifications and its capability to sustain elevated depth estimation precision across an extensive range.

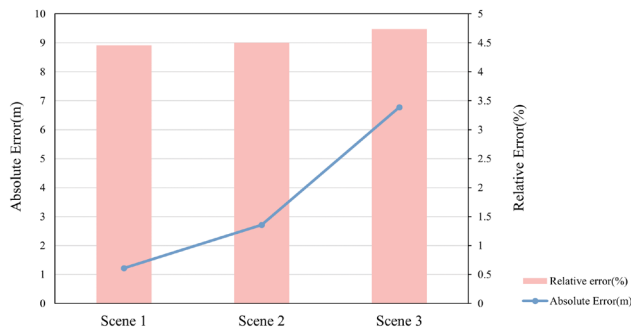


Figure 4. The bar and line chart statistics of the error in our proposed depth estimation algorithm.

3.3 Distance Calculation and Error Analysis

In this experiment, our initial task entailed segmenting the power lines in proximity to the construction machinery from the scene's point cloud. Following this, the Oriented Bounding Box around the power lines was constructed. The computation of the coordinates of the eight vertices of this enclosing box was achieved through the application of 3D geometric principles, with these coordinates serving as the foundational data for subsequent distance computations. Thereupon, the depth value from the construction machinery to the power line was harnessed, as assessed in Section 3.1, to establish the distance separating the construction machinery from the power line. For the purposes of accuracy verification, the software, CloudCompare, was employed to gauge the shortest distance between the machinery and the power line, subsequently adopting this measurement as the baseline truth, which is represented in Table 4. By juxtaposing the computed distance against the true shortest distance, the researchers are furnished with the means to assess the efficacy of our depth estimation algorithm within real-world applications, evaluating its performance.

Scene	Point cloud of scene	True distance(m)
Scene 1		25.815
Scene 2		25.201
Scene 3		24.881

Table 4. The real distance between construction machinery and power lines.

The information, set forth in Table 5, provides the outcomes derived from computing the distance between the construction machinery and the power line, along with an appraisal of the errors encountered in three distinct scenarios subjected to varying exposure conditions. The depth estimation algorithm that has been formulated in this research shows impressive accuracy and robustness across all scenarios. Furthermore, the absolute error of our algorithm is maintained within a 3-meter threshold, and the relative error is consistently confined under 10%, even amidst the demanding challenges posed by normal exposure, overexposure, and underexposure conditions. The precision achieved here fulfills the safety criteria concerning the distance that ought to be upheld between construction machinery and power lines, hence demonstrating the practical applicability and the reliability of the proposed depth estimation algorithm.

Scene	Our Method(m)	Error(m)	Relative Error(%)
Scene 1	NE	23.935	1.88
	OE	23.708	2.107
	UE	23.895	1.92
Scene 2	NE	23.144	2.057
	OE	23.130	2.071
	UE	23.133	2.068
Scene 3	NE	22.521	2.360
	OE	22.495	2.386
	UE	22.507	2.374

Table 5. The real distance between construction machinery and power lines.

The visual analysis presented in Figure 5 delineates the distribution of distance computation errors across three distinct scenes, each under varying exposure conditions including normal exposure, overexposure, and underexposure. The figure incorporates a line graph to represent absolute error and a bar graph for displaying relative error. A notable observation from this graphical representation is the relatively stable level of error across each scene, irrespective of the exposure conditions. This highlights the consistent accuracy of our proposed algorithm, even amidst challenging lighting scenarios. Furthermore, the minimal variance in the error of our algorithm as the scene scale expands underscores the algorithm's commendable applicability and stability across different magnitudes. Such stability assumes critical importance in real-world applications, primarily because it ensures that the algorithm can furnish reliable depth estimations across a diverse array of environments and conditions. This reliability is paramount for bolstering construction safety, enhancing operational efficiency, and mitigating the likelihood of potential accidents.

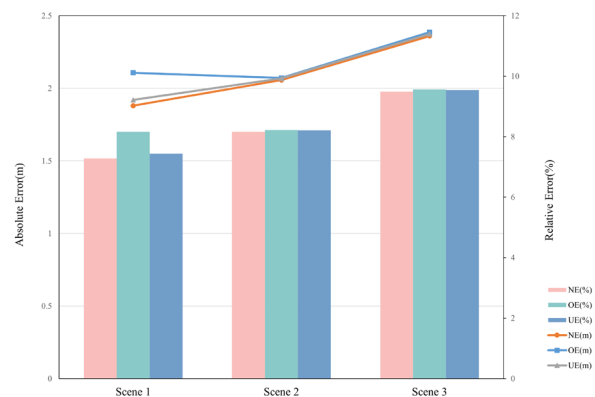


Figure 5. The bar chart and current situation diagram of the calculation error of the distance from machinery to power lines.

4. Conclusion

In this paper, a novel monocular camera depth estimation method pivoted on the concept of perspective transformation is proposed, which caters to the exigencies of safety surveillance in the proximity of construction machinery and power lines. This method meticulously constructs mathematical models pertaining to optic sphere projection and the spherical axis plane projection, thereby enabling a more precise evaluation of the depth between construction machinery and cameras, and subsequently, calculating the safety distance to power lines. The empirical evidence from our experiments attests to the efficacy of our algorithm—it excels not only under single exposure conditions, but it also retains exceptional accuracy and robustness across multiple exposure settings, as well as various scene scales. When compared to conventional measurement methods, our strategy boasts significant benefits, such as lower costs, heightened efficiency and simplicity, which collectively serve to amplify the safety management standards at construction environments. Looking ahead, the researchers aim to further refine our algorithm. Prospective research endeavours might involve adopting optimization strategies informed by gradient descent or genetic algorithms to meticulously fine-tune the parameters within the perspective transformation procedure. The objective of such enhancements would be to curb errors during the coordinate transformation phase and to amplify the precision and timeliness of depth estimation. The end goal is to facilitate even more accurate monitoring of construction site safety.

Acknowledgment

This work is supported by China Southern Power Grid Corporation Technology Project Fund (Project Number: [031300KK52220013]).

References

- Araújo, A. B., 2021. Spherical perspective. *Handbook of the Mathematics of the Arts and Sciences*. Cham: Springer International Publishing, 527-587.
- Chen, K., Lai, Y. K., Hu, S. M., 2015. 3D indoor scene modeling from RGB-D data: a survey. *Computational Visual Media*, 1, 267-278.
- Chen, Y., Zhao, H., Hu, Z., 2021. Attention-based context aggregation network for monocular depth estimation. *International Journal of Machine Learning and Cybernetics*, 12, 1583-1596.
- Cremer, C. Z., 2021. Deep limitations? Examining expert disagreement over deep learning. *Progress in Artificial Intelligence*, 10, 449-464.
- Daniel Girardeau-Montaut, 2024. CloudCompare Software, Version 2.13.0. <http://www.cloudcompare.org> (14 February 2024).
- Fang, Z., Chen, X., Chen, Y., 2020. Towards good practice for CNN-based monocular depth estimation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 1091-1100.
- Flores, M., Valiente, D., Peidró, A., 2024. Generating a full spherical view by modeling the relation between two fisheye images. *The Visual Computer*, 1-26.
- Kazerouni, I. A., Fitzgerald, L., Dooly, G., 2022. A survey of state-of-the-art on visual SLAM. *Expert Systems with Applications*, 205, 117734.
- Kiessling, F., Nefzger, P., Nolasco, J. F., 2014. Overhead power lines: planning, design, construction. *Springer*.
- L., Bengio, Y., 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.
- Lahiri, S., Ren, J., Lin, X., 2024. Deep learning-based stereopsis and monocular depth estimation techniques: a review. *Vehicles*, 6(1), 305-351.
- Lepetit, V., Moreno-Noguer, F., Fua, P. 2009. EP n P: An accurate O(n) solution to the P n P problem. *International journal of computer vision*, 81, 155-166.
- Macario Barros, A., Michel, M., Moline, Y., 2022. A comprehensive survey of visual slam algorithms. *Robotics*, 11(1), 24.
- Mu, W., Tong, D., 2020. Distance in spatial analysis: measurement, bias, and alternatives. *Geographical Analysis*, 52(4), 511-536.
- Rao, A. S., Radanovic, M., Liu, Y., 2022. Real-time monitoring of construction sites: Sensors, methods, and applications. *Automation in Construction*, 136, 104099.
- Shu, Y., Chen, W., 2018. Research and application of UHV power transmission in China. *High voltage*, 3(1),1-13.
- Talaei Khoei, T., Ould Slimane, H., Kaabouch, N. 2023. Deep learning: Systematic review, models, challenges, and research directions. *Neural Computing and Applications*, 35(31), 23103-23124.