



ARTICLE

Monocular Distance Estimated Based on PTZ Camera

Qirui Zhong¹, Xiaogang Cheng^{2,*}, Yuxin Song³ and Han Wang²

¹Bell Honors School Nanjing University of Posts and Telecommunications, Nanjing, 210003, China

²College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing, 210003, China

³College of Automation and College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing, 210003, China

*Corresponding Author: Xiaogang Cheng. Email: chengxg@njupt.edu.cn

Received: 24 January 2024 Accepted: 07 April 2024 Published: 15 May 2024

ABSTRACT

This paper introduces an intelligent computational approach for extracting salient objects from images and estimating their distance information with PTZ (Pan-Tilt-Zoom) cameras. PTZ cameras have found wide applications in numerous public places, serving various purposes such as public security management, natural disaster monitoring, and crisis alarms, particularly with the rapid development of Artificial Intelligence and global infrastructural projects. In this paper, we combine Gauss optical principles with the PTZ camera's capabilities of horizontal and pitch rotation, as well as optical zoom, to estimate the distance of the object. We present a novel monocular object distance estimation model based on the Focal Length-Target Pixel Size (FLTPS) relationship, achieving an accuracy rate of over 95% for objects within a 5 km range. The salient object extraction is achieved through a simplified convolution kernel and the utilization of the object's RGB features, which offer significantly faster computing speeds compared to Convolutional Neural Networks (CNNs). Additionally, we introduce the dark channel before the fog removal algorithm, resulting in a 20 dB increase in image definition, which significantly benefits distance estimation. Our system offers the advantages of stability and low device load, making it an asset for public security affairs and providing a reference point for future developments in surveillance hardware.

KEYWORDS

Public security; location information; PTZ camera; image processing

Nomenclature

f	Focal length of the lens
D	Distance between object and lens
u	Distance between object and lens in the equation
v	Distance between image and lens in the equation
$loss$	Value calculated by loss function
$offset$	Offset of each iteration
$weight_{[i][j]}$	Weight coefficient of coordinates (i, j)



1 Introduction

Reliable and accurate estimation of an object's distance using a monocular camera is crucial for enabling autonomous public security surveillance. With the widespread installation of surveillance cameras and the rapid development of infrastructural projects globally, there is an urgent need for effective surveillance methods. However, traditional approaches such as infrared mapping [1], RGB-D camera [2], Time of Flight camera [3] and Laser Radar [4] suffer from limitations such as high cost, narrow trial range, and low reliability.

In recent years, wildfires have caused significant economic losses, particularly in remote areas. This has highlighted the need for surveillance systems capable of not only monitoring but also locating emergency sites to mitigate the impact of natural disasters.

Pan-Tilt-Zoom (PTZ) cameras, with their horizontal and pitch rotation capabilities and optical zoom, offer a promising solution. This paper proposes a novel estimation model, FLTPS (Focal Length-Target Pixel Size), which leverages the optical zoom and target pixel size calculated from RGB images. FLTPS exploits the relationship between target pixel size and focal length to accurately estimate object distance, reducing hardware requirements and detection time.

To enhance detection accuracy, a novel salient object extraction algorithm is introduced, leveraging RGB layer features. Additionally, a dark channel prior fog removal algorithm [5,6] is introduced to enhance image contrast, which is beneficial to object detection using the Yolo algorithm [7] and salient object extraction.

The FLTPS model significantly improves estimation accuracy, achieving up to 95% accuracy when combined with the introduced algorithms. Generalization experiments demonstrate a 96% accuracy rate within a 5-km range using focal lengths ranging from 18 to 150.

As a novel monocular distance estimation method, FLTPS offers valuable insights for surveillance device manufacturers, suggesting improvements based on basic optical principles to enhance public security.

To make the schematic diagram and equation easy to understand, the notations used in this paper are listed above. More specific symbol used in this paper is explained where it is proposed.

This paper is structured into six sections. [Section 2](#) presents a comprehensive description of relevant literatures and compares them with our work. In [Section 3](#), we present the specific methodology we proposed and introduced to our work. In [Section 4](#), the outcomes of our algorithm are displayed. In [Section 5](#), the research discussion and suggestions for future research are presented. In [Section 6](#), we present our conclusion and the acknowledgements.

2 Related Work

Although single vision-only images could not offer useful distance information like Light Detection and Ranging (LiDAR) scanning pictures, they are also known as much evolving technology which obtain most of its data that is usable and accessible for humankind. As a result, several vision-only distance estimation methods were studied in last decade.

Every object in real world obeys perspective principle. The early study could use the principle to detect distance through vision-only image. Moreover, Jarvis [8] study this principle by striped lighting apparatus to detect the distance.

According to those light travels in straight line, which is showed in Fig. 1, we could get the clear formula:

$$h = \frac{H \cdot f}{D} \tag{1}$$

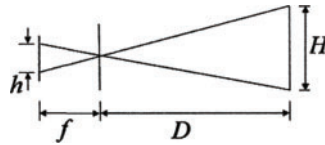


Figure 1: Perspective principle schematic diagram, indicating that the closer the object is to camera, the bigger it will be on the images. h and H represent the size of image and object

However, this approach heavily relies on the given information, such as the position of light in Jarvis’s study and the real size of object which could be quite hard to get in real situation. Moreover, Binocular ranging methods [9] are not considered because of their high price.

Also, some studies focus on the relative distance relation between already known object and unknown target to detect the distance [10] which could be used in the field of auto driving where camera captures the relative relation between obstacle and already known road lane. However, it is not very possible to find a suitable already known object as comparison in real situation.

Recently, the multi Hidden-Layer Neural Network, named DisNet [11] whose structure is showed in Fig. 2, trained using a supervised learning technique, which is used to learn and predicate the distance between the object and the camera sensor. However, the algorithm structure includes 3 hidden layers, in each there are 100 neurons. Moreover, the algorithm needs sufficient ground truth to achieve training demand, which is not friendly to device and daily affair.

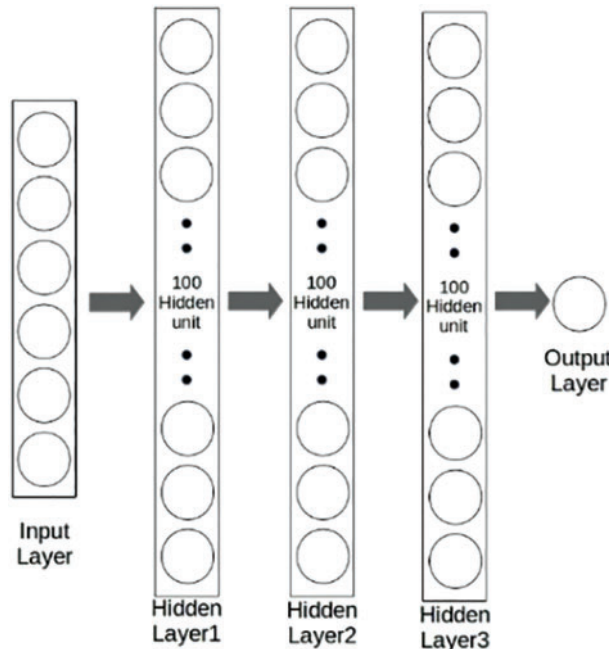


Figure 2: The structure of DisNet used for object distance prediction

Also, countless deep learning algorithms are proposed in the age of artificial intelligence. The decision of whether the emergency has happened is a prior of distance estimation. Several articles about deep learning based organizational decisions algorithms have been published [12]. Our system mostly used in the wild must take the environment and sustainability into consideration. Some articles have researched its current trends, challenges and future trajectories [13]. The procession of image captured by the PTZ camera utilize the algorithms related to convolution neural networks, feature extraction, which have been widely researched [14,15] and contribute to our work.

Above all, most vision-only distance estimated approaches used the information only from RGB images. However, PTZ cameras could not only offer the images but also the focal length which plays a decisive role in FLTPS model.

3 Novel System Description

3.1 FLTPS Model

It is well known to photographers that different focal length could cause different relative distance relation between objects which is known as visual phenomenon or optical illusion. The FLTPS model is constructed in the basis of this phenomenon. We took pictures of object in different distance with several focal length from 18 to 150 mm. The result is showed in Fig. 3.



Figure 3: The picture of object by 18, 24, 35, 50, 70, 100 mm

Obviously, the longer the focal length is, the larger the object is on the RGB image. After detecting the location of object by Yolo algorithm, the pixel size of object is calculated which has the trend below.

Clearly, the relation between Target Pixel Size and Focal length which is showed in Fig. 4 nearly meets parabola. It is easy to find the feature that the increasing rates of Target Pixel Size are quite different from each other. Specifically speaking, the pixel size of object which is close to camera has a higher speed of increasing.

This paper has researched this phenomenon on several objects in different distances ranging from 1 to 5 km. To process data of different objects, we standardize data in the way which could not change the relative relation between Target Pixel Size and Focal length. We choose the shortest focal length picture of the smallest object and save its pixel size which could be used as the base of standardization. We use $size_i$ to stand for the i th picture's salient object size in one group of pictures and $base$ to stand for

the smallest object's size in the shortest focal length. The data of object's pixel size after standardization will be saved in the new vector named $size'_i$. As a result, we could get the standardization formula:

$$size'_i = \frac{size_i}{base} \tag{2}$$

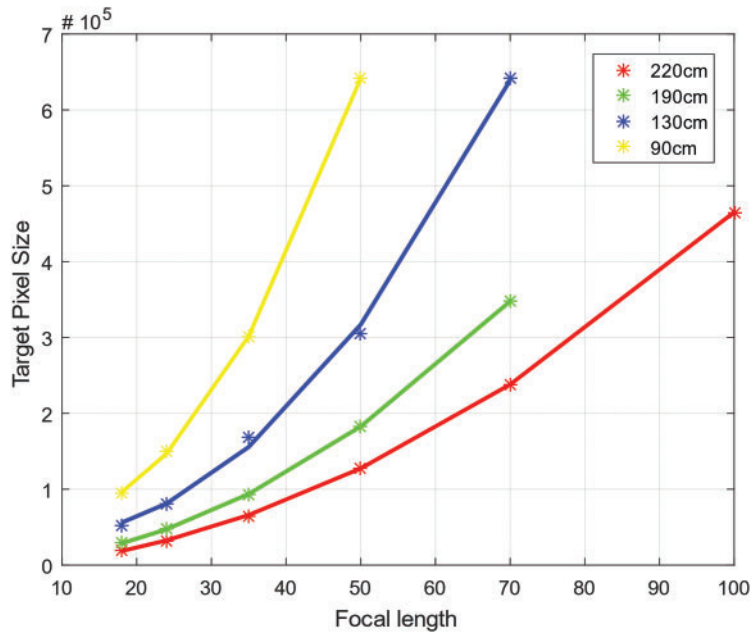


Figure 4: The relation between target pixel size and focal length

According to the parabola, we set a specific focal length and calculated object's pixel size which could stand for the increasing rate at the chosen focal length. In this paper, we collect the dataset by the camera using 18 to 150 mm focal length. The specific length chosen to predict the object size is 300 mm.

To find out the internal physical and mathematic theories of this phenomenon, this paper focuses on the basic optical imaging gauss formula. We use u to stand for object distance and v to stand for image distance. As usual, f stands for the focal distance.

Based on the normal gauss formula [16]:

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v} \tag{3}$$

We could get the formula which represents the mathematic relationship between objects at different distances when the focal distance changes from f to f' . Also, v' stands for the new image distance.

According to Eq. (3), the new gauss formula as the focal length changes is showed below:

$$\frac{1}{v_1} + \frac{1}{u_1} = \frac{1}{v_2} + \frac{1}{u_2} = \frac{1}{f'} \tag{4}$$

We take a further step to find that:

$$v'_1 = \frac{1}{\frac{1}{u_2} - \frac{1}{u_1} + \frac{1}{v'_2}} \quad (5)$$

It is clear that difference of distance could cause a numerical deviation in the formula which reflects in $\frac{1}{u_2} - \frac{1}{u_1}$ part. As a result, the v'_1 could never equal v'_2 . When we set u_1 a static number and study the mathematic features of Eq. (5), we find that v'_2 is getting close to v'_1 as u_2 gets larger. Meanwhile, the changing rate of v'_2 is getting slowly when u_2 is much larger than u_1 known as the remained detected object's distance.

This mathematic derivation could be clearly known in Fig. 4 which shows that the increasing rate of pixel size of farthest object is the slowest. To make this optical law much more precise, we could transform Eq. (5) to:

$$v'_2 = \frac{u_2}{\left(\frac{1}{v'_1} + \frac{1}{u_1}\right) u_2 - 1} \quad (6)$$

The partial derivative is:

$$\frac{\partial v'_2}{\partial u_2} = -\frac{1}{\left[\left(\frac{1}{v'_1} + \frac{1}{u_1}\right) \cdot u_2 - 1\right]^2} \quad (7)$$

Since $\left(\frac{1}{v'_1} + \frac{1}{u_1}\right)$ is defined by the original device's set which could be considered as a constant, we replace it with A to make the formula easy to understand.

$$\frac{\partial v'_2}{\partial u_2} = -\frac{1}{[A \cdot u_2 - 1]^2} \quad (8)$$

To make a specific study of the optical rule, we take a further look at $A \cdot u_2$ part, also known as:

$$\begin{aligned} &\left(\frac{1}{v'_1} + \frac{1}{u_1}\right) \cdot u_2 \text{ part.} \\ &\left(\frac{1}{v'_1} + \frac{1}{u_1}\right) \cdot u_2 = \frac{u_2}{v'_1} + \frac{u_2}{u_1} \end{aligned} \quad (9)$$

Usually, the detected object is at a longer distance compared to the already known object, whose distance is u_1 . As a result, $[A \cdot u_2 - 1]$ is a positive number. According to the deduction above, we could find that the increasing rate of image distance, as object distance increases, gets more and more slowly as object distance gets larger which perfectly match the experimental result showed in Fig. 4.

The increasing rate of image distance will be considered as the most significant feature to detect the real distance of object. Also, the target size of object could reflect the image distance. As a result, we build Focal Length-Target Pixel Size model to get the object's distance. We take several groups of pictures of objects in different distance from 1 to 5 km, by the camera whose focal length ranges from 18 to 150 mm. Their target's pixel size is also satisfied with the parabola function. We set 300 mm focal length as a normal standard to get the pixel increasing rate. To removal the effect of object's physical size, we utilize the feature rescaling mentioned in Eq. (2).

The real-life target pixel size increasing rate data which gets from experiment is showed in Fig. 5.

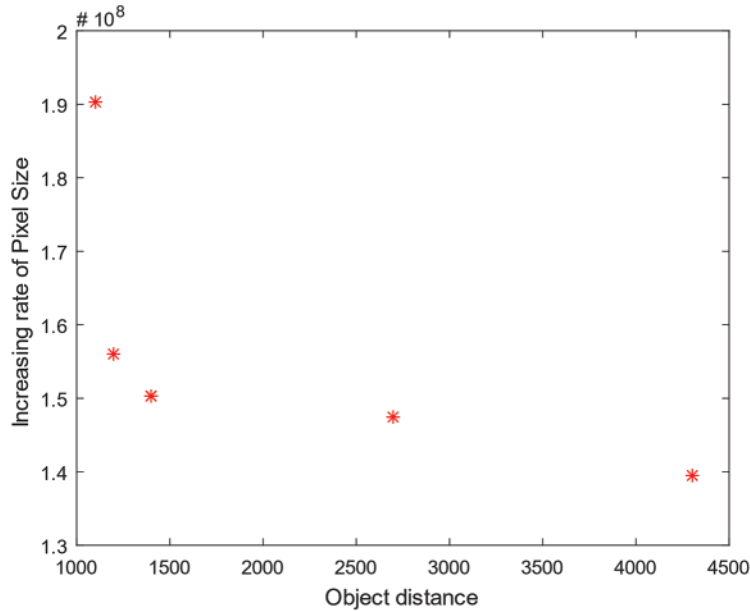


Figure 5: The mathematic relationship between increasing rate of target pixel size and object distance

The experimental result perfectly matches formula Eq. (8), which indicates that the increasing rate of target pixel size gets slowly as the object distance increases.

Due to the complex optical structure of camera lens, it is hard to use signal lens optical rule to build prediction formula. Based on Fig. 5, we establish inverse proportional function with offsets model to detect the unknown object. Back propagation neural networks [17] are introduced, combined with mean square error function, to amend the offsets until they match the already known target.

We set a, b, c as three different offsets to construct model:

$$y = \frac{b}{x - a} + c \quad (10)$$

The loss function is used to calculate the deviation of model during training steps:

$$loss = \sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n} \quad (11)$$

Back propagation algorithm propagates the loss factor to offsets to optimize detect model. We could introduce the learning rate factor to improve the efficiency of optimizing process:

$$offset' = offset - \frac{\partial loss}{\partial offset} \cdot lr_{offset} \quad (12)$$

We set a tiny number as standardization which means that the training step will stop if $loss$ factor minor than it. After calculating the satisfied offsets data, our FLTPS model is constructed successfully. The detecting model curve combined with the already known target increasing rate is showed below:

According to the detection curve, which perfectly matches the already known data which is showed in Fig. 6, we could know that the BP neural networks are equipped to achieve self-adaptive adjustment of offsets which could be devoted to the optimizer of FLTPS model.

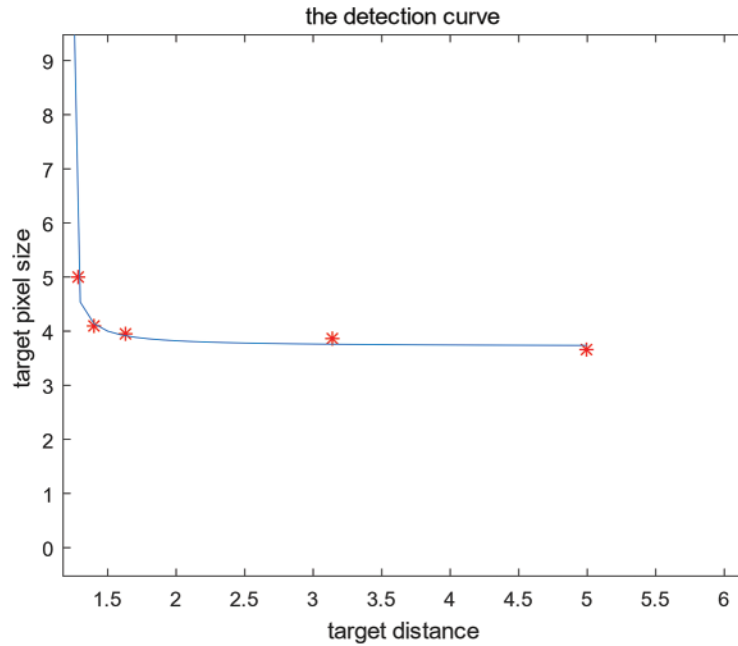


Figure 6: The detection curve combined with the already known target information

FLTPS model relies on the extraction of target object and calculates its pixel size as accurate as possible. To improve the detection accuracy of model, we introduce *dark channel prior fog removal* and salient object extraction methods to process the pre-data process.

3.2 Dark Channel Prior Haze Removal

A clear input image with high contrast plays a significant role in later algorithm process. Also, the accurate calculation or target pixel size influence the accuracy the distance detection directly. As a result, we introduce dark channel prior haze removal algorithm from He to improve the quality of images and accuracy of distance detection. Dark Channel Prior algorithm combines dark channel, estimation of atmospheric light and haze concentration to removal its negative effect.

To find out the effect of haze removal algorithm, Laplace variable variance and Peak Signal-to-Noise Ratio (PSNR) algorithms [18] are introduced.

Laplace variable variance [19] is the algorithm which could measure the amount of detail or definition of one single picture by comparing the contrast among pixels and gradient. The higher the result is, the clearer the image is. Here are 4 groups of pictures showed in Fig. 7.

The quantized data is showed in Fig. 8.

PSNR is used to compare the difference between raw picture and the processed picture pixel by pixel obeying the following formula:

$$MSE = \sum \frac{(img_1[i,j] - img_{ij}[i,j])^2}{size} \quad (13)$$



Figure 7: Four groups of pictures captured by PTZ cameras in the wild and processed by dark channel prior algorithm

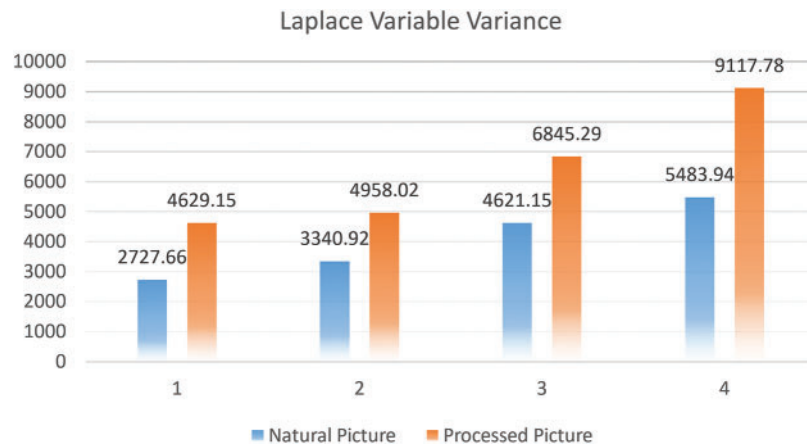


Figure 8: Laplace variable variance comparing the amount of detail of pictures

The quantized data is showed in Fig. 9.

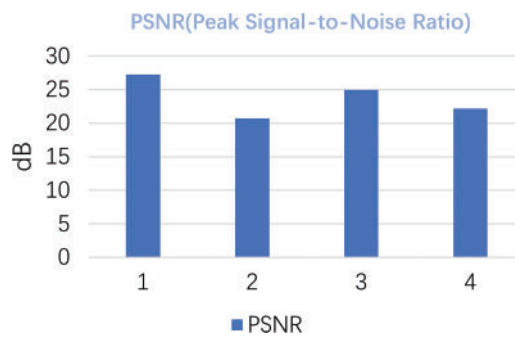


Figure 9: Peak signal-to-noise ration comparing the processed pictures to original pictures

According to the double standards of pictures' quality, we could infer that the dark channel prior algorithm dose makes an improvement of picture's quality which about 20 dB gain in picture definition.

3.3 Salient Object Extraction

As described above, FLTPS relies on the accuracy of the calculation of target pixel size. As a result, the extraction of salient object to get its pixel size is greatly important to detection accuracy. The mainstream extraction algorithms [20] are based on morphology, which focus on the object contour detection [21] or object features [22].

Contour Detection and Hierarchical Image Segmentation is a traditional algorithm that processes large amounts of pixel calculations to combine the effect of brightness, color, and texture gradient. With the thriving of deep neural networks, CNN frame [23] has become a hit in computer vision. Object Contour Detection with a Fully Convolutional Encoder-Decoder Network [24] is an algorithm that relies on a deep convolutional network to detect an object's contour. However, mass computing power is a huge test of hardware that is not available to devices pursuing lightweight mounting. Also, the CNN frame needs countless training data. However, the result of the model is quite hard to satisfy, as shown in Fig. 10.



Figure 10: Object contour detection with 16 fully convolution. Model trains 15000 times with this picture

Compositing demand, FLTPS only needs a part of object to detect its distance, such as a color stain. As a result, we come up with salient object extraction algorithm based on RGBs channels.

Also, the RGBs channels are stable when the focal length changes which may put huge pressure on extraction algorithms based on morphology. For example, detecting the distance of wildfire site is one of the top missions of our model. What extraction algorithm needs is extracting the body of fire or smoke, which is showed in Fig. 11.

Salient object extraction algorithm focuses on finding the RGBs features of single picture according to the RGBs histogram. Comparing to the normal histogram [25], salient object detection needs to put higher weight on pixels at the center of the picture, because object picture is segmented based on Yolo algorithm. As a result, we design the brand-new algorithm to calculate the three channels. The weight of each pixel ranging from 0.5 which is at the edge of picture, to 1 which is

at the center of the picture.

$$weight_{[i][j]} = 1 - \frac{\max [abs (i - mid_y), abs (i - mid_x)]}{2 \cdot \max [mid_y, mid_x]} \tag{14}$$

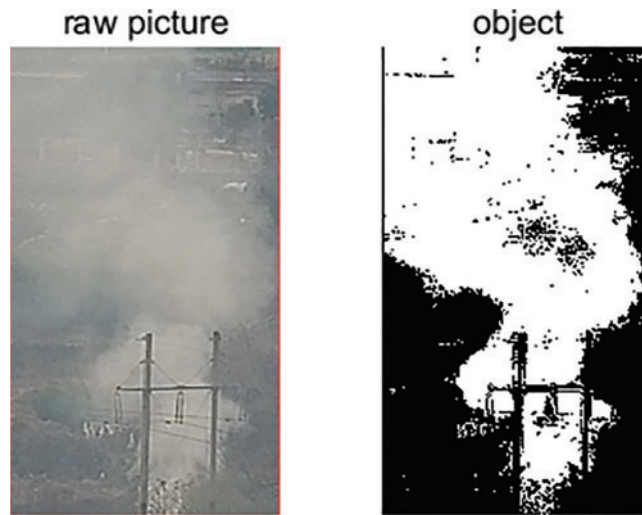


Figure 11: Wildfire emergency captured by PTZ camera and the main object extracted by salient object extraction algorithm based on RGBs channel

According to it, we could get the weight graphic which is used to dot multiply each channel of RGB pictures' pixel brightness data. And then, the novel RGBs channel histogram could be got which is showed below.

Histogram is the most important tool to assist us to find the features of one single picture. It is easy to find that there one obvious peak in Fig. 12 which is most likely to be the main body of object.

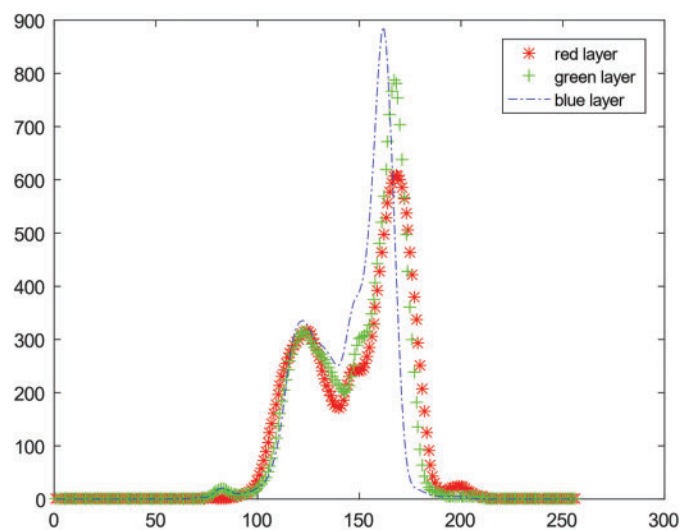


Figure 12: Histogram calculated by novel algorithm

The second part of extracting object is selecting the RGBs features in Histogram. Due to the discreteness of unit 8 type data, we process the histogram data channel by channel following the rule:

$$save_i = \sum_{j=i}^{i+\varphi} channel_j \quad (15)$$

In the formula, φ indicates the calculating width of each batch. The larger φ is, the wider the final chosen part is. Through Eq. (15), we could get a new vector. The index of max data in *save* is the beginning index of channel. And φ is the width of chosen part. When we set φ is 15, the chosen part is showed below.

Through the chosen features, which is showed in Fig. 13 in red, we could extract pixel which has all features. The processed result is showed in Fig. 14.

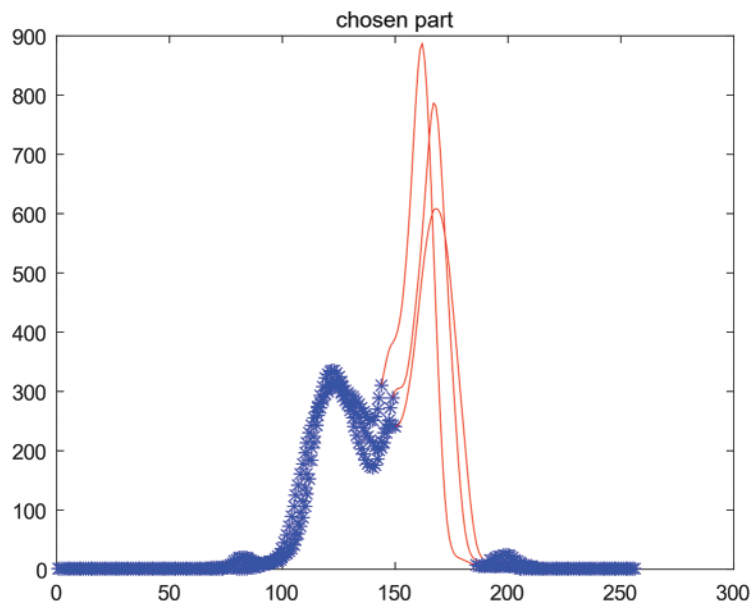


Figure 13: The selected features of each RGB channel through our novel algorithm

Also, φ is a significant variable which could change its number to fit different environment. We use different φ to process the same image, the result is showed below.

Variable φ could amend model according to the contrast ratio [26] of pictures. Usually, high φ is used in pictures with low contrast ratio, such as those taken in haze of frog weather, or the object is classified as smoke which is showed in Figs. 15 and 16. Low contrast ratio is used in pictures with high contrast ratio, such as pictures with fire as object, which is showed below.

With the assistance of novel object extraction algorithm based on RGBs channels, FLTPS model can not only calculate object's pixel size and detect their distance much more accurately, but also it puts much less pressure on hard device.



Figure 14: The extracted object through our novel algorithm

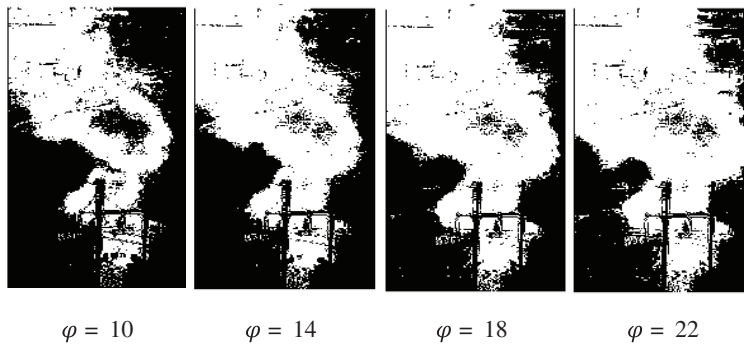


Figure 15: The processed results with different φ



Figure 16: Salient object detection in picture with high contrast ratio

3.4 Function of Each Module

The core of our work is the FLTPS model, which is used to estimate the distance of target object. The dark channel prior haze removal and salient object extracting algorithm is introduced and proposed to improve the accuracy of the FLTPS model. The quality of image could be improved by dark channel prior haze removal in the condition of poor visibility. The effectiveness of this algorithm is showed in Fig. 7. The extraction of salient object is finished in Eq. (4), which could significantly improve the calculation of target pixel number. The effectiveness of this algorithm is showed in Fig. 15.

4 Generalization Experiment Result

Above all, Monocular distance estimated based on PTZ camera's detail algorithms and tetchiness are described. FLTPS model is core of this system, which is aimed at estimating object's distance based on a group of optical images. Dark channel prior algorithm could improve the quality of images. Also, salient object extraction improves the accuracy of calculating target pixel size which is beneficial to estimation. The whole frame is showed in Fig. 17.

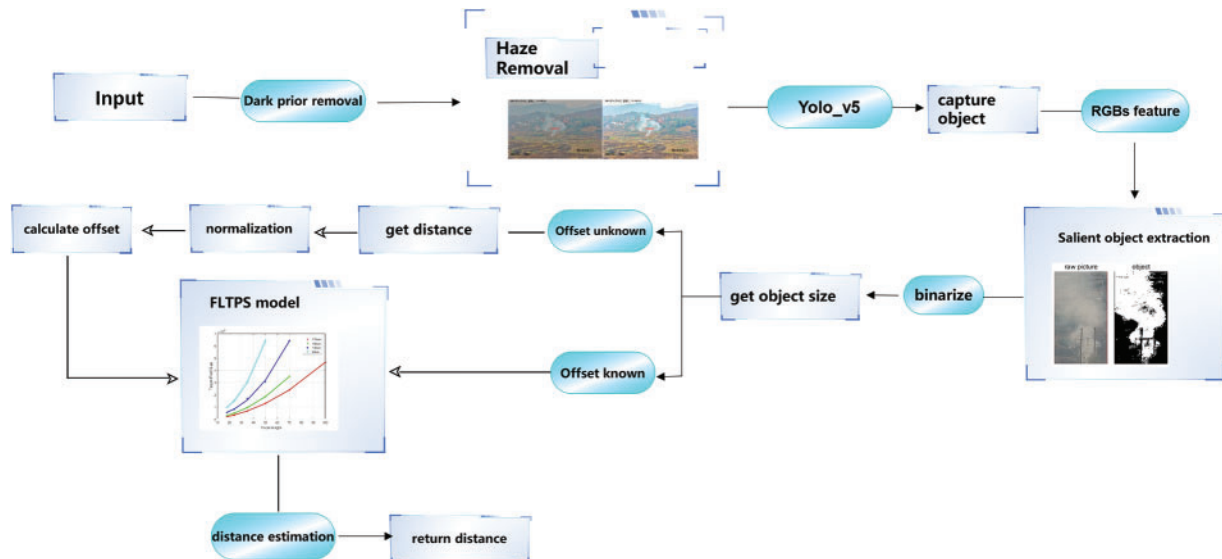


Figure 17: The whole system frames

Also, we apply the novel monocular distance estimation algorithm into practical use cases. Greenland Square Zi Feng Tower, which is 2.7 km away from experiment site, is used to test distance estimation model. We choose a part of tower as salient object, which is showed in Fig. 18.

The detection of Zi Feng Tower's distance is 2.7588359 km, with 2.179% error. Except it, we have chosen several different objects to test our model's stability, the result is showed in Table 1.

According to data, we could infer that the error of model can be lower than 5% when it detect the object whose distance ranging from 1 to 5 km.

The accuracy of our work could be compared to DisNet algorithms, even though we put much less pressure on the device.

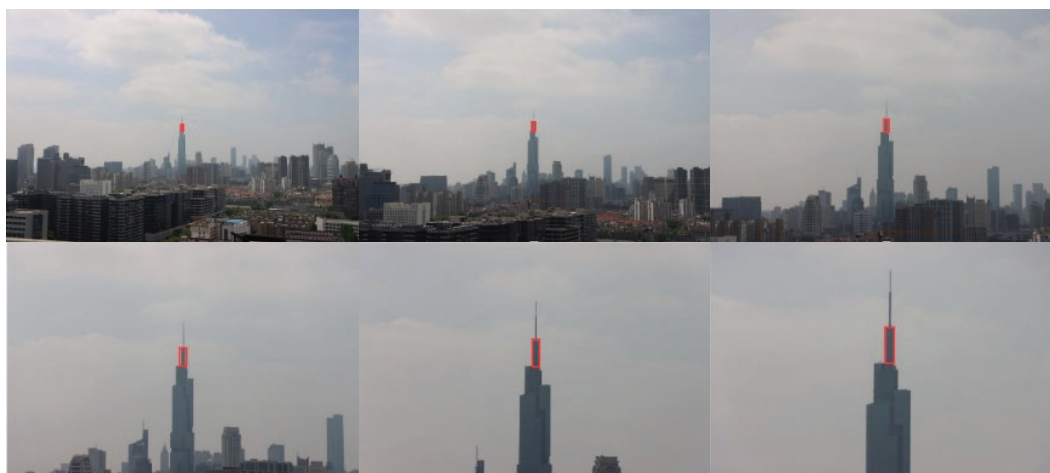


Figure 18: The data captured by PTZ camera for generalization experiments

Table 1: Experimental result

Physical distance	Estimated distance	Error
4300 m	4500 m	4.65%
1100 m	1084.8 m	1.38%
1400 m	1345.7 m	3.88%
1200 m	1222.2 m	1.85%

5 Discussion and Future Work

5.1 Public Security

In 2020, it was reported that there were over 236 thousand hectares of wild land suffering from the damage of wildfire, most of which are remote areas where public security affairs are hard to access. One monocular distance estimated algorithm based on a PTZ camera can not only reduce the pressure on humanity but also protect public security and reduce the loss of life and property. This novel algorithm could be devoted to protecting public security affairs.

5.2 The Future Development of Device

Recently, more and more advanced algorithms based on everyday devices put much more pressure on the CPU or GPU, which brings high costs to the device and reduces the speed of processing. Our novel algorithm utilizes the physical optical laws to realize the lightweight mounting. Also, we know that the complex constructed lens is built based on optical laws. The device manufacturers could refer to Eq. (7) to design lenses, which may perform better in our model. From Eq. (7), we know that the accuracy of detecting could be higher if the lens has a smaller image distance, and the range of focal length could be more extensive. Even though more cameras are aimed at small sizes, our discovery offers a brand new idea to design specific PTZ cameras to improve the accuracy of detection.

5.3 The Improvement of Processed Image

The salient object extraction algorithm and Dark Prior Removal algorithm do make a lot of effort in improving accuracy, but there is still a long way to improve the quality of the images and the accuracy of calculating, such as CNN networks. The simplified kernel could already extract the salient object based on RGBs' features; deep learning still has the potential to create a more efficient convolution kernel to extract objects with the basis of sufficient training data.

6 Conclusion

This paper focuses on monocular distance estimation and introduces a novel algorithm with low computational pressure and high detection accuracy. Additionally, a new salient object extraction algorithm, combined with He K's Dark Prior Removal algorithm, is proposed to enhance the accuracy of distance estimation in the system.

The findings of this research hold significant implications for public security affairs, particularly considering the rapid development of infrastructural projects worldwide. As governance capabilities increase, surveillance coverage will expand, making our work highly applicable and valuable for future reference. Moreover, there is a need for further research and development in more effective estimation models and object extraction algorithms, as these are hot topics in Computer Vision (CV) and can significantly mitigate the impact of emergencies.

Acknowledgement: We would like to thank Nanjing University of Post and Telecommunication for supporting this work.

Funding Statement: The project is financially supported by the Social Development Project of Jiangsu Key R&D Program (BE2022680), the National Natural Science Foundation of China (Nos. 62371253, 52278119).

Author Contributions: The authors confirm contribution to the paper as follows: Study conception and design: Qirui Zhong Author; data collection: Qirui Zhong; analysis and interpretation of results: Qirui Zhong, Yuxin Song, Xiaogang Cheng, Han Wang; draft manuscript preparation: Yuxin Song. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and materials: Data and materials will be provided if required.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] H. Kurokawa *et al.*, "A probabilistic approach to determination of ceres' average surface composition from dawn visible-infrared mapping spectrometer and Gamma Ray and Neutron detector data," *J. Geophys. Res.: Planets*, vol. 125, no. 12, 2020. doi: [10.1029/2020JE006606](https://doi.org/10.1029/2020JE006606).
- [2] C. Choi and H. I. Christensen, "3D pose estimation of daily objects using an RGB-D camera," presented at the IEEE/RSJ Int. Conf. Intell. Robots Syst., Vilamoura, Algarve, Portugal, Oct. 7–12, 2012, pp. 3342–3349.
- [3] L. Li, "Time-of-flight camera—An introduction," in *Tech. White Pap.*, SLOA190B, 2014.
- [4] R. Schneider, P. Thümel, and M. Stockmann, "Distance measurement of moving objects by frequency modulated laser radar," *Opt. Eng.*, vol. 40, no. 1, pp. 33–37, Jan. 2001. doi: [10.1117/1.1332772](https://doi.org/10.1117/1.1332772).

- [5] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2010.
- [6] S. Serikawa and H. Lu, "Underwater image dehazing using joint trilateral filter," *Comput. Electr. Eng.*, vol. 40, no. 1, pp. 41–50, Jan. 2014. doi: [10.1016/j.compeleceng.2013.10.016](https://doi.org/10.1016/j.compeleceng.2013.10.016).
- [7] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A review of yolo algorithm developments," *Procedia. Comput. Sci.*, vol. 199, pp. 1066–1073, 2022.
- [8] R. A. Jarvis, "A perspective on range finding techniques for computer vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-5, no. 2, pp. 122–139, 1983. doi: [10.1109/TPAMI.1983.4767365](https://doi.org/10.1109/TPAMI.1983.4767365).
- [9] X. Sun *et al.*, "Distance measurement system based on binocular stereo vision," presented at the IOP Conf. Ser.: Earth Environ. Sci., 2019, vol. 252, no. 5, pp. 52051. [10.1088/1755-1315/252/5/052051](https://doi.org/10.1088/1755-1315/252/5/052051).
- [10] H. Königshof, N. O. Salscheider, and C. Stiller, "Realtime 3D object detection for automated driving using stereo vision and semantic information," presented at the IEEE ITSC, Auckland, New Zealand, Oct. 27–30, 2019, pp. 1405–1410.
- [11] M. A. Haseeb, J. Guan, D. Ristic-Durrant, and A. Gräser, "DisNet: A novel method for distance estimation from monocular camera," in *10th Planning, Perception and Navigation for Intelligent Vehicles (PPNIV18)*, IROS, 2018.
- [12] M. Mohamed, "Empowering deep learning based organizational decision making: A survey," *Sustain. Mach. Intell. J.*, vol. 3, no. 2, pp. 13, 2023. doi: [10.61185/SMIJ.2023.33105](https://doi.org/10.61185/SMIJ.2023.33105).
- [13] M. Mohamed, "Agricultural sustainability in the age of deep learning: Current trends, challenges, and future trajectories," *Sustain. Mach. Intell. J.*, vol. 4, no. 2, pp. 20, 2023. doi: [10.61185/SMIJ.2023.44102](https://doi.org/10.61185/SMIJ.2023.44102).
- [14] Y. Ding *et al.*, "Unsupervised self-correlated learning smoothly enhanced locality preserving graph convolution embedding clustering for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2022.
- [15] Y. Ding *et al.*, "Multi-scale receptive fields: Graph attention neural network for hyperspectral image classification," *Expert. Syst. Appl.*, vol. 223, pp. 119858, 2023. doi: [10.1016/j.eswa.2023.119858](https://doi.org/10.1016/j.eswa.2023.119858).
- [16] C. F. Gauss, "Dioptric investigations," (in Deutsch), in *Dioptric Investigations*, 1st ed. Leipzig, Germany: Printing and Publishing of Dieterichsche Bookstore, 1841, vol. 1.
- [17] M. Buscema, "Back propagation neural networks," *Subst. Use Misuse*, vol. 33, no. 2, pp. 233–270, 1998. doi: [10.3109/10826089809115863](https://doi.org/10.3109/10826089809115863).
- [18] J. Korhonen and J. You, "Peak signal-to-noise ratio revisited: Is simple beautiful?" in *2012 Fourth International Workshop on Quality of Multimedia Experience*, IEEE, 2012, pp. 37–38.
- [19] R. Bansal, G. Raj, and T. Choudhury, "Blur image detection using Laplacian operator and open-CV," in *2016 Int. Conf. Syst. Model. Adv. Res. Trends (SMART)*, IEEE, 2016, pp. 63–67.
- [20] D. Ziou and S. Tabbone, "Edge detection techniques-an overview," *Pattern Recognit. Image Anal. CIC Raspoznavaniye Obrazov I Analiz Izobrazhenii*, vol. 8, pp. 537–559, 1998.
- [21] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, 2010. doi: [10.1109/TPAMI.2010.161](https://doi.org/10.1109/TPAMI.2010.161).
- [22] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Electr. Eng.*, vol. 40, no. 1, pp. 16–28, 2014. doi: [10.1016/j.compeleceng.2013.11.024](https://doi.org/10.1016/j.compeleceng.2013.11.024).
- [23] L. Alzubaidi *et al.*, "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, pp. 1–74, 2021. doi: [10.1186/s40537-021-00444-8](https://doi.org/10.1186/s40537-021-00444-8).
- [24] J. Yang, B. Price, S. Cohen, H. Lee, and M. H. Yang, "Object contour detection with a fully convolutional encoder-decoder network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 193–202.
- [25] H. P. Dembinski, J. Pivarski, and H. Schreiner, "Recent developments in histogram libraries," presented at the EPJ Web Conf., 2020, vol. 245, pp. 05014.
- [26] Stelzer, "Contrast, resolution, pixelation, dynamic range and signal-to-noise ratio: Fundamental limits to resolution in fluorescence light microscopy," *J. Microsc.*, vol. 189, no. 1, pp. 15–24, 1998. doi: [10.1046/j.1365-2818.1998.00290.x](https://doi.org/10.1046/j.1365-2818.1998.00290.x).