Mathematical Biosciences and Engineering

MBE, 21(3): 3618–3630.
DOI: 10.3934/mbe.2024159
Received: 12 December 2023
Revised: 30 January2024
Accepted: 31 January2024
Published: 06 February 2024

http://www.aimspress.com/journal/MBE

Research article

SFEMM: A cotton binocular matching method based on YOLOv7x

Zhang Guohui, Gulbahar Tohti*, Chen Ping, Mamtimin Geni and Fan Yixuan

School of Mechanical Engineering, Xinjiang University, Urumqi, Xinjiang 830047, China

* **Correspondence:** Email: gulbahart@xju.edu.cn.

Abstract: The cotton-picking robot needs to locate the target object in space in the process of picking in the field and other outdoor strong light complex environments. The difficulty of this process was binocular matching. Therefore, this paper proposes an accurate and fast binocular matching method. This method used the deep learning model to obtain the position and shape of the target object, and then used the matching equation proposed in this paper to match the target object. Matching precision of this method for cotton matching was much higher than that of similar algorithms. It was 54.11, 45.37, 6.15, and 12.21% higher than block matching (BM), semi global block matching (SGBM), pyramid stereo matching network (PSMNet), and geometry and context for deep stereo regression (GC-net) respectively, and its speed was also the fastest. Using this new matching method, the cotton was matched and located in space. Experimental results show the effectiveness and feasibility of the algorithm.

Keywords: deep learning; binocular ranging; binocular stereo matching; cotton picking robot; cotton centroid positioning.

1. Introduction

Cotton is an important economic crop in China [1]. Selected Xinjiang long staple cotton fibers can reach a length of over 40 mm, with high uniformity, and can be spun into ultra-fine yarns. The yarn lines are dry and uniform, with high strength and a smooth surface with less fuzz. Large cotton-picking machines can damage the integrity of cotton fibers and increase the impurity content of cotton after picking due to the high-speed rotation of the picking spindle. In order to protect the excellent fiber characteristics of long staple cotton, manual picking can only be carried out at present [2].

The precise cotton-picking robot can effectively alleviate the problem of cotton fiber damage

during the machine picking process. However, the precise cotton-picking robot's spatial positioning of cotton before picking has always been a challenge and there is no perfect solution. There are many researchers engaged in related research on the spatial positioning of cotton or other crops. In 2014, Wang et al. used lasers to scan cotton and then binarize the images to obtain cotton coordinates [3]. However, this method was limited to single cotton plants and did not study the situation of multiple cotton plants. In 2021, Liu et al. applied monocular ranging technology to the field of agriculture [4]. On the premise of finding the average size of grapefruit, they used monocular ranging technology to estimate the distance of grapefruit. With the continuous iteration and updating of classic object detection algorithms such as Faster R-CNN (region-CNN, convolutional neural networks) and YOLO (you only look once) series [5–10], many researchers have applied these algorithms to the spatial localization of crops [11]. In 2023, Gharakhani et al. and others applied structured-light cameras to robot picking cotton [12]. Liu et al. applied a combination of depth camera and improved YOLOv5 to the recognition and localization of chili peppers [13]. However, the complex lighting conditions in cotton fields are not conducive to the use of structured light and TOF (time of flight) cameras.

Considering the complex environment of farmland, this study selected binocular cameras to measure the depth information of cotton and then spatially locate it. The key aspect of using binocular cameras to perform this entire process on cotton is binocular matching. There are already mature algorithms for other aspects, including target detection, calculation of depth information, and determination of crop spatial position. Therefore, this article proposes a binocular matching algorithm, SFEMM (shape feature extraction matching method).

The characteristic of SFEMM is that it does not match all pixels in the view and obtain parallax, but ignores the background pixels, only matches the parts closely related to the target object, and calculates the parallax as the central position of the target object. This can reduce the interference of background factors on matching and improve the accuracy. Obtaining the parallax of an object directly instead of getting the parallax of the whole picture can reduce the running time. Compared with the traditional method, it can reduce the interference of other factors except the target object in the picture, and compared with the matching ranging method based on deep learning, it has low requirements for data sets. Also, the training pressure of the deep learning model is small. The model only needs the function of target detection. It can quickly and accurately match the target object in the picture, which is convenient for subsequent work.

2. Materials and methods

In this paper, the SFEMM algorithm was proposed and applied to the spatial location of cotton. The spatial positioning scheme was generally divided into three parts: target detection of cotton ball, binocular matching of cotton balls using SFEMM, and space positioning. The process is shown in Figure 1.

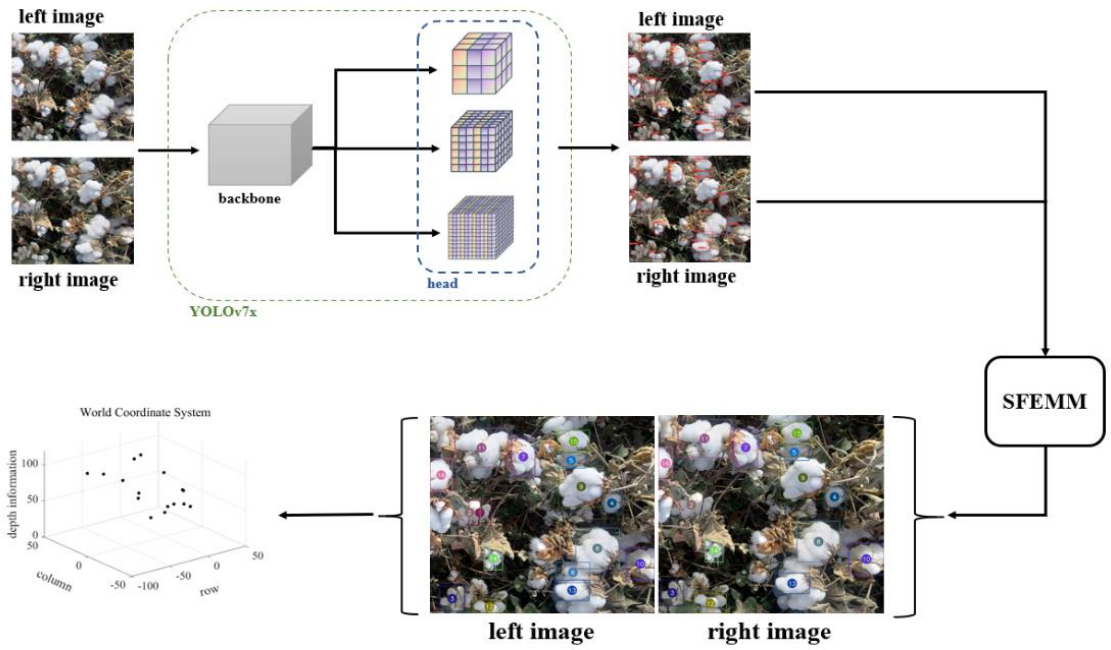


Figure 1. Scheme process.

2.1. SFEMM

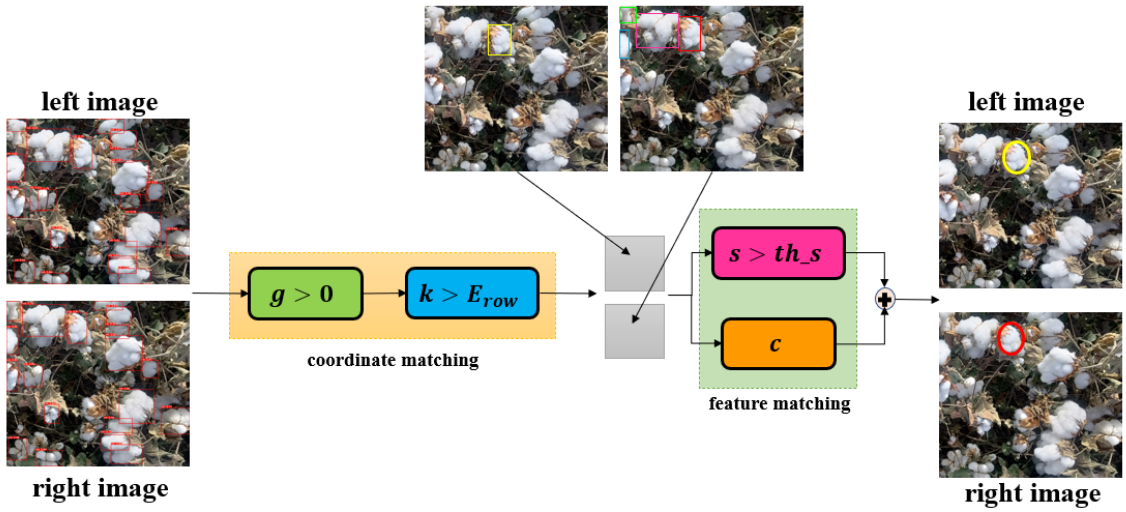


Figure 2. Scheme of the SFEMM algorithm.

The SFEMM algorithm used information such as the position of the same object in the field of view of two cameras and the characteristics of the object itself to achieve matching. The operation process of this method was divided into two parts. One part used the target detection technology based on deep learning to extract the position, shape, and color category of crops in the view; The other part was to process and screen the obtained crop information to complete the matching. As shown in Figure 2, the input image was a cotton image that has been detected. The coordinate matching part consists of two steps: horizontal coordinate matching and vertical coordinate matching [14–16]. The feature

matching part consists of two parts in parallel: the shape features of the target and other features such as color and category [16–20]. For this target detection model, this paper selected YOLOv7x [21]. SFEMM is an auxiliary module, which can give the model binocular matching ability after connecting with the target detection model.

2.1.1. Determination of column coordinate characteristics of cotton ball

Vertical coordinate matching can be called column coordinate matching. For the same seed cotton ball, its column coordinates in the left camera view are greater than those in the right camera view. It is commonly understood that the position of the same object in the left image is larger than in the right image. This phenomenon is also the basic condition for obtaining the parallax of objects. The column coordinates of the seed cotton ball can be determined by the following equation.

$$g = x_l^j - x_r^i,\ i,j = 1,2 \ldots n,\ g > 0 \tag{1}$$

where x_l^j is the column coordinate of the jth seed cotton for binocular matching in the left image, x_r^i is the column coordinate of the ith seed cotton in the right image, and g is the matching of cotton balls in column coordinates.

2.1.2. Determination of row coordinate characteristics of cotton ball

Horizontal coordinate matching can be called row coordinate matching. According to the limit constraint principle, the row coordinates of the same seed cotton ball in the left and right views are the same. However, in actual work, there will be a little error in the coordinates of two view lines due to the influence of working conditions and hardware equipment. Therefore, during binocular matching, the row coordinate of the same object in the left and right pictures are very close, and the proximity of the row coordinates can be expressed by the absolute value of the difference between the two row coordinates. In order to make this algorithm meet the binocular images with different resolutions, the absolute value is divided by the height of the image, which can intuitively show the proximity of the horizontal coordinates of the object in the global.

$$k = e^{-|y_l^j - y_r^i|/h},\ i,j = 1,2 \ldots n,\ k > E_{row} \tag{2}$$

where y_l^j is the row coordinate of the jth seed cotton for binocular matching in the left image, y_r^i is the row coordinate of the ith seed cotton in the right image, h is the height of the image, k is the matching of cotton balls in row coordinates (the larger the value of k, the closer the two row coordinates are), and E_{row} is the threshold of row coordinate matching with a value range of 0.85–0.95.

2.1.3. Determination of cotton ball shape and size characteristics

The idea of object shape feature matching is that the width and height of the same object in the left picture and the right picture are very close. By comparing the width and height of the object in the left image with the width and height of the object in the right image, the similarity of the shape features

of the matched object can be directly expressed as following equation.

$$s = \frac{\min\left(w_l^j, w_r^i\right)}{\max\left(w_l^j, w_r^i\right)} + \frac{\min\left(h_l^j, h_r^i\right)}{\max\left(h_l^j, h_r^i\right)}, i, j = 1,2 \ldots n, \quad s > th\_s \tag{3}$$

where w_l^j, h_l^j are the width and height of the jth object in the left image respectively, w_r^i, h_r^i are the width and height of the ith object in the right image respectively, s is the similarity of the objects in the left and right views for matching (the larger the s value, the more similar the shape characteristics of the object), and th_s represents the threshold of crop similarity with a value range of 1.5–1.8.

2.1.4. Characteristics, color, and category of the target

The characteristics, color, and category of the target can be expressed as following equation.

$$c = 1/n \sum_{j=1}^{n} d(l_j, r_j) \tag{4}$$

$$d(l_j, r_j) = \begin{cases} 1, & l_j = r_j \\ 0, & l_j \neq r_j \end{cases} \tag{5}$$

where l_j, r_j represent the characteristics of the object in the left view and the characteristics of the object in the right view respectively (such as the color of the target, the aspect ratio of the target in the image, and category of target, etc), $d(l_j, r_j)$ is whether the j-th characteristics of the object in the left and right views are the same or similar, and c is the similarity of multiple characteristics of the object in the left and right views. There are many ways to obtain the characteristics of object, and the following are introduced in this paper: using deep learning model to identify the category of objects, using the characteristics of RGB (red, green, blue) pictures to judge the color of objects, and using the filter to obtain the texture of the object.

The specific operation process of the cotton ball matching part is as shown in Figure 2: Selecting a seed cotton ball in the left camera view as the target, and then find the matching seed cotton ball from the right camera view. Each seed cotton ball in the right camera image will be considered as an alternative only if it satisfies Eqs (1)–(3) at the same time (there is a certain probability that the target in the left camera view and the target in the left camera view are the same seed cotton ball in real space), that is, $g > 0, k > th\_j, s > th\_s$. Among all the alternatives, only the target with the closest feature, that is, the matching target with the maximum of s and c, will be considered as the final matching result. This target is the target most likely to be the same as the seed cotton ball in the left camera view in real space.

2.2. Spatialization

After matching, parallax could be obtained for ranging and then used for spatial positioning. Position cotton in the world coordinate system, with the camera center as the origin, and the coordinate axis was shown in Figure 3. To distinguish between the world coordinate system and the image coordinate system, in Figure 3, *u* was used instead of the x-axis of the image coordinate system, and *v* was used to represent the y-axis of the image coordinate system [22]. The positioning equations are

shown in Eqs (6)–(8).

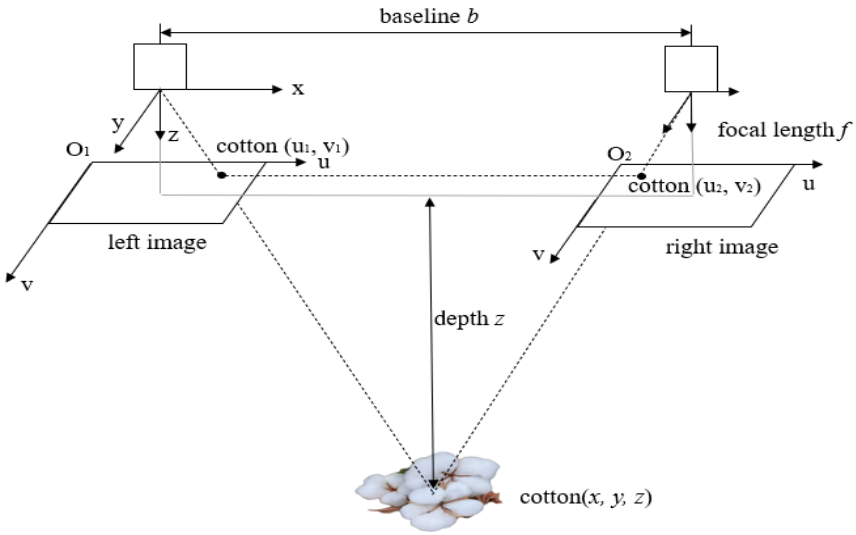


Figure 3. Coordinate system setting.

2.2.1. Z-axis coordinate calculation equation

$$z = f \times b/d \tag{6}$$

where z is the depth information and Z-axis coordinate (mm), f is the focal length (pixel), b is the baseline (mm), and d is parallax (pixel).

2.2.2. X-axis coordinate calculation equation

$$x = (x_{img} - x_0) \times z/f \tag{7}$$

where x is the x-axis coordinate of the object in the world coordinate system (mm), x_{img} is the coordinates of the object in the picture (pixels), and x_0 is the x-axis coordinate of the camera center in the picture coordinate system (pixels).

2.2.3. Y-axis coordinate calculation equation

$$y = (y_{img} - y_0) \times z/f \tag{8}$$

where y is the y-axis coordinate of the object in the world coordinate system (mm), y_{img} is the coordinates of the object in the picture (pixels), and y_0 is the y-axis coordinate of the camera center in the picture coordinate system (pixels).

2.3. Dataset of cotton

In order to make the model have better detection effect, this study collected cotton photos from farmland. When collecting cotton photos, the dataset collection refers to the actual situation in the field

and takes photos of cotton from different angles and distances. In order to increase the anti-interference ability of the model to the light after training, the cotton was photographed in different light environments. Considering that defoliant is needed during cotton picking, this paper takes photos of cotton before and after pesticide spraying respectively. The example image of cotton dataset is shown in Figure 4. Then, the collected images are filtered and sorted out. Labeling software was used to label 3000 pictures in total.

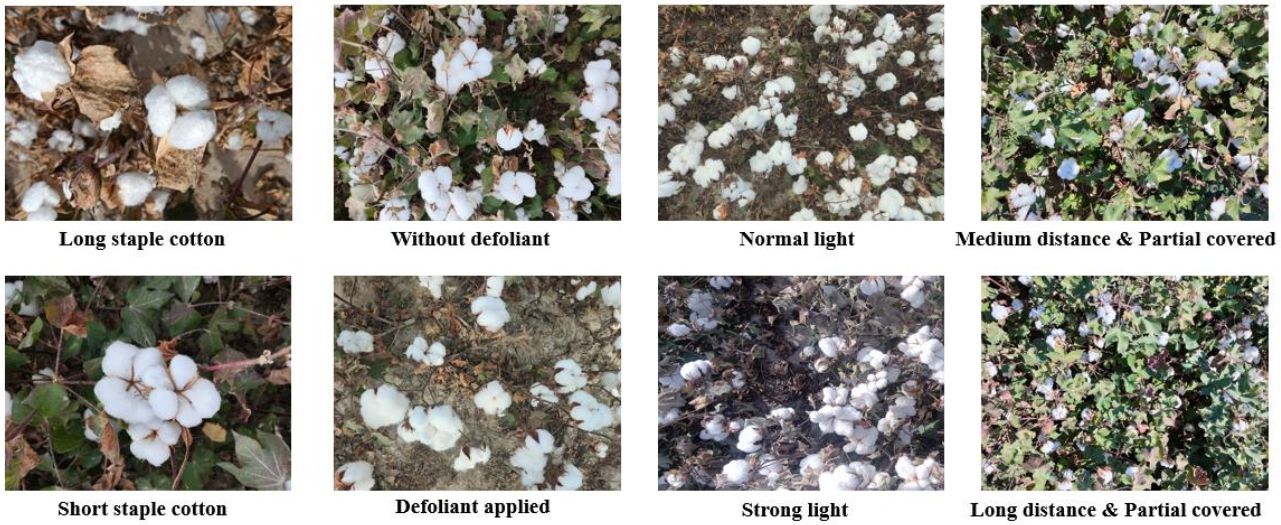


Figure 4. Example of cotton dataset.

2.4. Evaluation index

2.4.1. Evaluation of target detection model

Precision (P), recall (R), and mean average precision (mAP) were used as the evaluation criteria of the model. Precision was used to evaluate the accuracy of cotton detection. Recall was used to evaluate the comprehensiveness of detection. Mean average precision was the mean of the average accuracy (AP) under all categories [23].

$$P = TP/(TP + FP) \tag{9}$$

$$R = TP/(TP + FN) \tag{10}$$

$$AP = (TP + TN)/(TP + TN + FP) \tag{11}$$

$$mAP = \sum_{i=1}^{C} AP_i/C \tag{12}$$

where
TP(true positives) = positive example which is correctly predicted;
FP (false positives) = positive example which is falsely predicted;
FN (false negative) = negative example which is falsely predicted;
TN (true negative) = negative example which is correctly predicted;
C = number of target categories.

2.4.2. Evaluation of binocular matching algorithm

Matching precision (MP) and matching recall (MR) were used as the evaluation criteria of the binocular matching algorithm. Matching precision was used to evaluate the accuracy of cotton matching. Matching recall was used to evaluate the comprehensiveness of cotton matching.

$$MP = TM/(TM + FM) \tag{13}$$

$$MR = TM/CM \tag{14}$$

where
TM (true matching) = targets which is correctly matched;
FM (false matching) = targets which is falsely matched;
CM (can matching) = targets that can be matched.

2.5. Experimental platform

The experimental platform used in this research is CPU: i5-12490F; GPU: RTX3070. Platform of binocular camera was built with 500W pixels industrial cameras of Ruiertushi company. As shown in Table 1, it presents the technical parameters of the binocular camera, including hardware performance and calibration parameters. The binocular camera is shown in Figure 5. It is assembled by two industrial cameras, and the bracket is made by 3D printing technology.

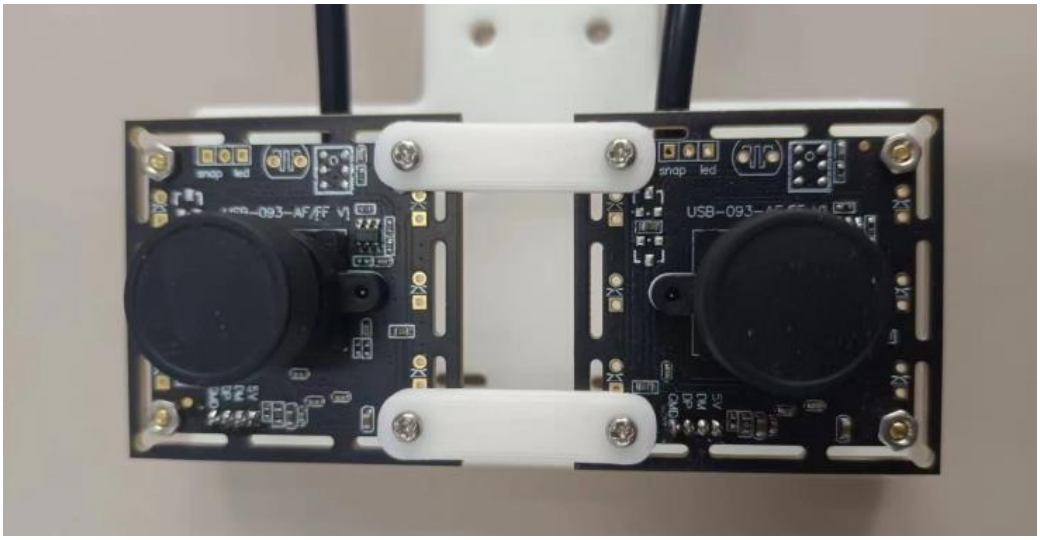


Figure 5. Assembled binocular camera.

Table 1. technical parameters of the binocular camera.

Hardware performance		Calibration parameters	
output format	MJPEG/YUV	focal distance(left)	2043.875
frame rate	30 FPS	focal distance(right)	2064.712
maximum resolution	2592 * 1944	image center(left)	1151.527 * 925.810
baseline	5cm	image center(right)	1280.002 * 962.588

3. Results and discussion

3.1. Comparison between SFEMM and other matching algorithms

On this experimental platform, the running time of SFEMM algorithm proposed in this study is 12 ms. This paper compared the running time of SFEMM with other matching algorithms, using the image size of 2311*2087. Under the condition of the same platform and the same size picture, the operation time of SFEMM is the shortest, as shown in Table 2.

Table 2. Comparison of running time of several algorithms.

Algorithm	Running time (s)	MP (%)	MR (%)
SFEMM	**0.013**	**87.52**	**83.73**
BM	0.062	33.41	28.69
PSMNet	0.41	81.37	78.25
GC-Net	0.91	75.31	70.12
SGBM	1.97	42.15	37.82

As for the matching precision, SFEMM was 54.11, 45.37, 6.15, and 12.21% higher than BM (block matching), SGBM (semi global block matching), PSMNet (pyramid stereo matching network), and GC-net (geometry and context for deep stereo regression) respectively, and its speed was also 4.8, 31.5, 70, and 151.5 times faster than BM, PSMNet, GC-net, and SGBM respectively. The matching effect of SFEMM was shown in Figure 6. In addition to matching cotton, this article also provided its application to matching other types of objects, such as potted plants and marigold. The images of cotton and marigold in Figure 6 were taken in the outdoor natural environment, and the images of potted plants were taken in the indoor environment. This shows that SFEMM had a matching effect on crops in outdoor environment.

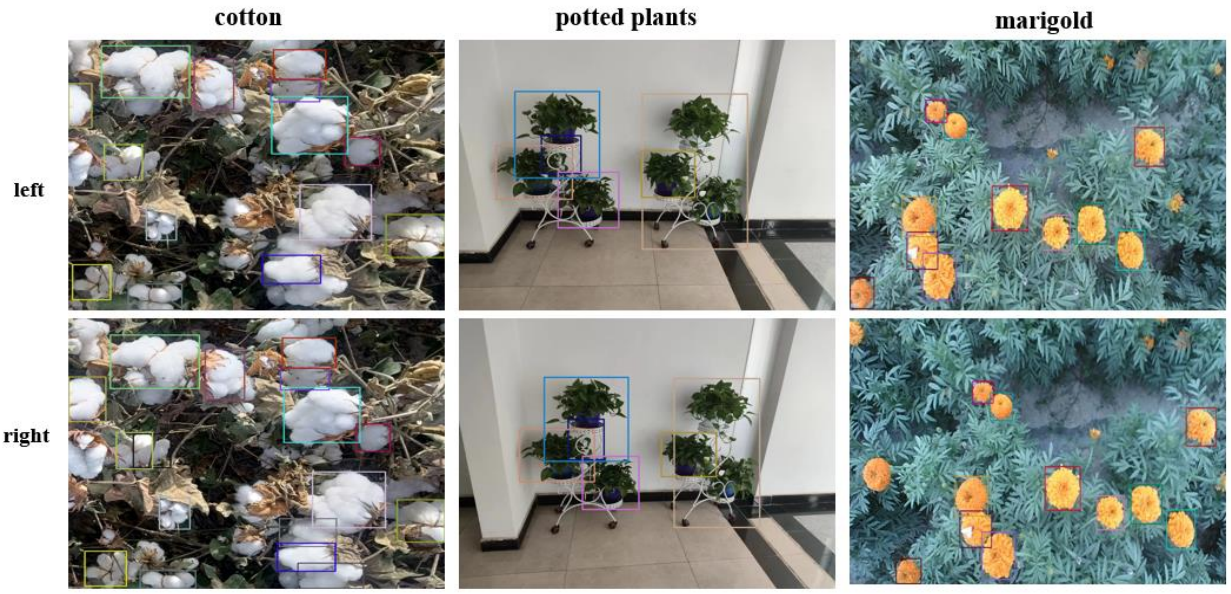


Figure 6. Matching effect of SFEMM.

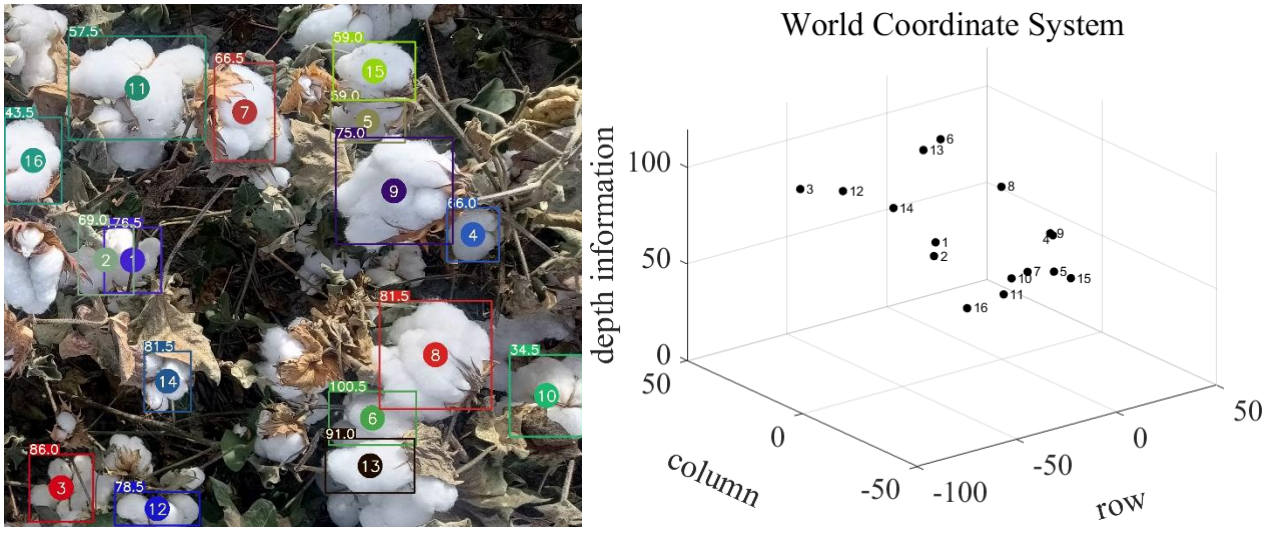
3.2. Selection of target detection model

In this study, a variety of target detection models were used to train the cotton dataset. The dataset used for model training is the cotton dataset produced in this article, with 3000 images. The training set, test set, and validation set were randomly divided in an 8:1:1 ratio. The performance of the trained model is shown in Table 3. In addition to comparing the commonly used precision, recall, and mAP, precision of the model for cotton partially covered by leaves and cotton of small size was also listed. It can be seen from Table 3 that the mAP of YOLOv7x was 5.98, 8.76, 4.73, 3.4, 3.26 and 1.55% higher than that of Faster-RCNN, YOLOv5s, YOLOv5m, YOLOX, YOLOv7, and YOLOv8 respectively. From precision of each model for cotton partially covered by leaves and cotton of small size, YOLOv7x was better than other models. Therefore, this paper finally selectsYOLOv7x as the target detection model for cotton detection.

Table 3. Comparison of running time of several algorithms.

Model	P (%)	R (%)	mAP (%)	AP (%)	
				Partially covered	Small size
YOLOv7x	**87.14**	**84.73**	**92.51**	**91.53**	**92.62**
Faster-RCNN	80.73	78.32	86.53	82.89	84.37
YOLOv5s	78.92	76.11	83.75	79.66	80.13
YOLOv5m	81.35	77.92	87.78	83.75	85.07
YOLOX	84.16	79.28	89.11	85.91	87.41
YOLOv7	83.62	80.84	89.25	86.24	87.97
YOLOv8	86.47	84.01	90.96	89.02	90.14

3.3. SFEMM computing centroid parallax and locating



(a) Centroid parallax of cotton ball **(b) Spatial location of cotton ball**

Figure 7. Centroid parallax and spatial location of cotton.

SFEMM could accurately match the objects according to their own characteristic information and obtain the parallax as the material center. Then, the depth information of crops could be calculated according to parallax. Finally, the spatial positioning was carried out. Figure 7 shows the centroid parallax map and spatial location map of cotton ball.

4. Conclusions

A spatial positioning method is presented in this paper by using the target detection model based on YOLOv7x and binocular matching algorithm using SFEMM, then applied it to the space location of cotton ball and other types of objects, such as potted plants and marigold. As for the MP, SFEMM are 54.11, 6.15, 12.21, and 45.37% higher and its speeds are also 4.8, 31.5, 70, and 151.5 times faster than BM, PSMNet, GC-net, and SGBM, respectively. In addition to the above work, this study carried out other necessary work such as collecting cotton images in farmland and making cotton dataset. The YOLOv7x model for cotton detection was trained, and the mAP of this model was 5.98, 8.76, 4.73, 3.4, 3.26, and 1.55% higher than other models, respectively. It is gives good results in extracting the space position and shape of cotton binocular matching images. This also provides a feasible algorithm for the benchmark picking of cotton-picking robots.

Highlights

1) A precise and fast binocular matching algorithm has been proposed;
2) This algorithm has better matching accuracy than other similar algorithms;
3) It can be applied to spatial positioning of cotton;
4) Created a dataset on cotton in outdoor farmland environments.

Acknowledgements

The research was supported by the National Natural Science Foundation of China (Project No. 12162031) and State Key Laboratory for Manufacturing Systems Engineering of Xi'an Jiaotong University (Project No. sklms2022022).

Use of AI tools declaration

The authors declare that they have not used Artificial Intelligence tools in the creation of this article.

Conflict of interest

The authors declare that there are no conflicts of interest.

References

1. *Announcement of The National Bureau of Statistics on Cotton Production In 2022*, 2022. Available from: http://www.stats.gov.cn/sj/zxfb/202302/t20230203_1901689.html
2. J. Tian, *Change of Fiber Quality in Machine-harvested Cotton in the Xinjiang and Further Survey of Promising Approaches for Improving*, MS thesis, Shihezi University, 2018.

3. L. Wang, S. Liu, W. Lu, B. Gu, R. Zhu, H. Zhu, Laser detection method for cotton orientation in robotic cotton picking, *Trans. Chinese Soc. Agric. Eng.*, **30** (2014), 42–48.
4. J. Liu, D. Zhou, Y. Li, D. Li, Y. Li, R. Rubel, Monocular distance measurement algorithm for pomelo fruit based on target pixels change, *Trans. Chinese Soc. Agric. Eng.*, **37** (2021), 183–191.
5. R. Girshick, Fast R-CNN, in *2015 IEEE International Conference on Computer Vision (ICCV)*, (2015), 1440–1448. https://doi.org/10.1109/ICCV.2015.169
6. S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, *IEEE Trans. Pattern Anal. Mach. Intell.*, **9** (2017), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031
7. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 779–788. https://doi.org/10.1109/CVPR.2016.91
8. J. Redmon, A. Farhadi, YOLO9000: Better, faster, stronger, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017), 6517–6525. https://doi.org/10.1109/CVPR.2017.690
9. J. Redmon, A. Farhadi, YOLOv3: An incremental improvement, preprint, arXiv:1804.02767.
10. A. Bochkovskiy, C. Y. Wang, H. Y. M. Liao, YOLOv4: Optimal speed and accuracy of object detection, preprint, arXiv:2004.10934. http://arxiv.org/abs/2004.10934
11. D. T. Fasiolo, L. Scalera, E. Maset, A. Gasparetto, Towards autonomous mapping in agriculture: A review of supportive technologies for ground robotics, *Rob. Auton. Syst.*, **169** (2023), 104514. https://doi.org/10.1016/j.robot.2023.104514
12. H. Gharakhani, J. A. Thomasson, Y. Lu, Integration and preliminary evaluation of a robotic cotton harvester prototype, *Comput. Electron. Agr.*, **211** (2023), 107943. https://doi.org/10.1016/j.compag.2023.107943.
13. S Liu, M Liu, Y Chai, S Li, H Miao, Recognition and location of pepper picking based on improved YOLOv5s and depth camera, *Appl. Eng. Agr.*, **39** (2023), 179–185. https://doi.org/10.13031/aea.15347
14. G. P. Stein, O. Mano, A. Shashua, Vision-based ACC with a single camera: Bounds on range and range rate accuracy, in *IEEE IV2003 Intelligent Vehicles Symposium*, (2003), 120–125. https://doi.org/10.1109/IVS.2003.1212895
15. H. Hirschmuller, Stereo processing by semiglobal matching and mutual information, *IEEE Trans. Pattern Anal. Mach. Intell.*, **30** (2008), 328–341. https://doi.org/10.1109/TPAMI.2007.1166
16. Y. Liu, G. Mamtimin, T. Gulbahar, M. Julaiti, An improved SM spectral clustering algorithm for accurate cotton segmentation in cotton fields, *J. Agr. Mech.*, **44** (2022).
17. A. Kendall, H. Martirosyan, S. Dasgupta, P. Henry, R. Kennedy, A. Bachrach, et al., End-to-end learning of geometry and context for deep stereo regression, in *2017 IEEE International Conference on Computer Vision (ICCV)*, (2017), 66–75. https://doi.org/10.1109/ICCV.2017.17
18. J. R. Chang, Y. S. Chen, Pyramid stereo matching network, preprint, arXiv:1803.08669. http://arxiv.org/abs/1803.08669
19. E. Goldman, R. Herzig, A. Eisenschtat, J. Goldberger, T. Hassner, Precise detection in densely packed scenes, preprint, arXiv:1904.00853. http://arxiv.org/abs/1904.00853
20. T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, preprint, arXiv:1612.03144. http://arxiv.org/abs/1612.03144

21. C. Wang, A. Bochkovskiy, H. M. Liao, YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2023), 7464–7475. https://doi.org/10.1109/CVPR52729.2023.00721
22. C. Huang, X. Pan, J. Cheng, J. Song, Deep image registration with depth-aware homography estimation, *IEEE Signal Process. Lett.*, **30** (2023), 6–10. https://doi.org/10.1109/LSP.2023.3238274
23. X. Peng, J. Zhou, Y. Xu, G. Xi, Cotton top bud recognition method based on YOLOv5-CPP in complex environment, *Trans. Chinese Soc. Agr. Eng.*, **39** (2023), 191–197.

AIMS Press