



Visual Servoing of Soft Robotic Arms by Binocular

Lizheng Feng, Xiaojiao Chen, and Zheng Wang^(✉)

Department of Mechanical Engineering, The University of Hong Kong,
Pokfulam, Hong Kong SAR, China
zheng.wang@ieee.org

Abstract. Soft robotic arms are complementing traditional rigid arms in many fields due to its multiple degrees of freedom, safety and adaptability to the environment. In recent years, soft robotic arms have become the focus in robotics research and gained increasing attention from scientists and engineers. Despite the rapid progress of its design and manufacturing processes in the past decade, an obstacle restricting the development of soft robotic arms remained unsolved. The suitable sensors for soft robotic arm have not appeared on the market and the integration of sensors into soft robotic arm has been difficult, since most sensors and actuator systems, such as those used in traditional robotic arms, are rigid sensors and rather simple. Therefore, finding a suitable soft robotic arm sensor has become an urgent issue in this field. In this paper, a simple and feasible method with a binocular camera is proposed to control the soft robotic arm. Binocular is employed to detect the spatial target position at first and then coordinates of target point will be transmitted to the soft robot to generate a control signal moving the soft robotic arm, and then the distance from target to the end effector will be measured in real time.

Keywords: Soft robotic arm · Visual servoing · Binocular disparity

1 Introduction

Soft robotic arms are safe, user-friendly and flexible in human-machine and machine-environment interaction [1], and therefore, they have a wide range of real-world applications and can revolutionize status quo with technological innovations [2]. For example, soft robotic arms can be applicable in the sorting of different shapes and fragile objects (fruits, vegetables, biological tissues etc.) [3], in medical and healthcare industry such as rehabilitation and auxiliary devices for stroke patients and assisted surgery [4, 5, 11]. Their physical adaptation to external environment empowers them with excellent capability to deal with uncertainty and disturbance [6, 7], allowing for low cost, safe and pleasurable human-robot interaction [8–10].

However, there is currently no suitable soft robotic arm sensor on the market and most of the soft robotic arms are still controlled by traditional sensors [12]. Although traditional rigid sensors have many applications on soft robotic arms, they are not

always well-matched [13, 14]. Compared with traditional robotic arms, soft robotic arms have no so-called link and joint structure even do not need to drive with an electric motor [16]. In theory, the soft arm has multiple degrees of freedom, without an accurately characterized stiffness and damping [13]. Therefore, many mature sensors used in traditional robotic arms cannot be applied well in soft mechanical arms. These factors have greatly affected the control of the soft robotic arms [14]. To bridge the gap, a novel method to control the robotic arm with vision is employed. Visual control system automatically receives and processes images of real objects through optical and non-contact sensors to obtain the information required for robot motion [15, 18, 19].

Many current visual servoing robotic arms are not well solved in visual positioning problem. The main obstacle is the inability to apply the vision sensor alone to accurately obtain the depth of the target point [20]. For example, it is difficult to solve the depth problem with a monocular camera very accurately [27]. At present, the monocular measurement distance is very popular in machine learning model [28], and the data comes from statistics. Even if the correct rate is continuously improved under the supervised learning model, it is still only a regular information under big data, and there is no physical theory or geometric model to support it. The result analysis has uncontrollable factors. What's more, the monocular camera is fixed in focal length and cannot be zoomed as fast as the human eye, it cannot solve the problem of imaging at different distances accurately. The binocular camera is a good complement to this shortcoming. The binocular camera can cover different ranges of scenes by using two identical cameras, which solves the problem that the monocular camera cannot switch the focal length back and forth and can also solve the problem of recognizing the images sharply at different distances.

By using a binocular camera, depth measurement may also face many issues such as result accuracy, real-time trade-off [21], and difficulty in obtaining accurate corresponding disparity pixel values to the actual distance in a linear model [22–24]. Therefore, in this work, some simple and feasible methods are proposed to improve the measurement efficiency and accuracy of the binocular system, and at the same time, it is well combined with the soft robotic arm control.

There are two main tasks of the binocular camera in this proposed system:

First, measure the depth from the target point to the end effector by binocular disparity relatively, accurately and efficiently [25].

Second, employ the left camera of the binocular camera to establish the imaging geometry model. And the depth information is used to obtain the X, Y coordinates of the target point [26].

The novelty of this work:

- ① Proposes a simple method to improve the linear model of binocular disparity, making measurement results more accurate, and presents a strategy to improve the calculation speed of binocular disparity.

- ② Provides a new idea for the eye-in-hand model. Compared to the most eye-in-hand model by monocular, the binocular camera is used in this work to obtain accurate depth measurement in real time with a high positioning accuracy.
- ③ Matches the vision system with the soft robotic arm model to control the soft robotic arm moving to the target point.

2 Design

The soft robotic arm system is shown in Fig. 1. The vision system in this paper is a binocular camera with a variable baseline length from 4.2 cm to 17.0 cm, and the parameters of the binocular camera are listed in Table 1. It will be mounted on the end effector of the soft robot arm, moving with the end effector of the arm as the “hand-in-eye” model. And the details of binocular are shown in Fig. 2. Finally, the camera detects the depth of the target to the camera in real time and transmits it back to the base coordinate system and keeps end effector approaching the object to a predetermined distance.

The first part of this work is to use the binocular camera for depth measurement. At the beginning of the measurement, the camera is calibrated to remove the distortion, followed by the camera parameters and distortion coefficient. The intrinsic parameters will be combined with the depth information for the spatial position calculation. In this work, the camera is calibrated using the classic Zhang Zhengyou calibration method. A 50 mm * 50 mm, 10 * 6 calibration plate is chosen in this work. The binocular calibration phasing has special characteristics compared to the monocular calibration, and the calibration plate must appear in the left and right frames at the same time. Making the calibration plate appear in the entire view can effectively improve the accuracy of the calibration. A total of 30 pairs of image data to calibrate the binocular to get the camera parameters are conducted. Then the original frames are rectified with camera parameters and distortion coefficient. Next, the epipolar geometry is used to convert binocular into a standard format. Disparity calculation of two vertically aligned images is then conducted. There are currently three popular methods for calculating disparity: StereoBMState, StereoSGBMState, StereoGCState. The comparison of the three methods is as follows:

- ① Calculation speed: BM method > SGBM method > GC method.
- ② Disparity accuracy: BM method < SGBM method < GC method.

In this work, real-time and measurement accuracy are critical for the visual system, hence, average StereoSGBMState mode is used. Based on this algorithm, a scientific method will be employed to enhance the measurement results.

The spatial target point coordinate information is then transmitted back to the camera coordinate frame, after that from the camera coordinate frame to the end effector coordinate frame, and finally from the end effector coordinate frame to the robotic arm base coordinate frame.

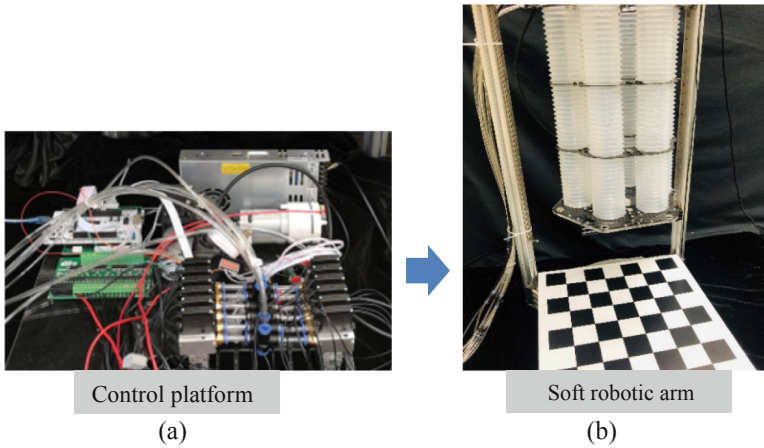


Fig. 1. (a) shows the control platform of the soft robotic arm system. (b) shows the soft robotic arm.

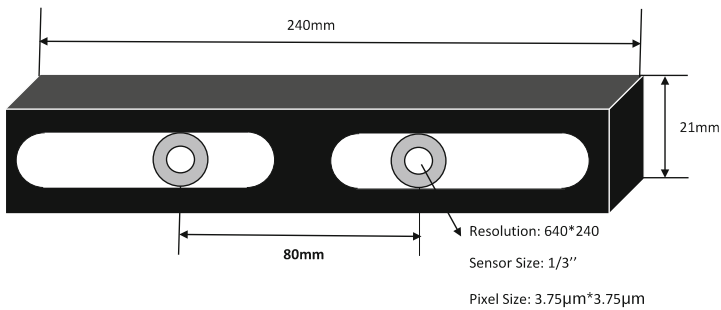


Fig. 2. Details of binocular

Table 1. Parameters of binocular

Image resolution	Baseline range	Sensor size	Pixel size	Image compression	Active array
640 × 320	42 mm–170 mm	1/3"	3.75 μm × 3.75 μm	MJPG YUY2	MJPG (60FPS) YUY2 (10FPS)

3 Modeling

The soft robotic arm is different from the traditional rigid body arm, since its end-effector position is characterized by bending angle, rotation angle, and length. Therefore, in this paper, an algorithm transforming the software robot coordinate system into

a Cartesian coordinate system is proposed to characterize the position of the end effector. The block diagram of the entire work is as shown in Fig. 3.

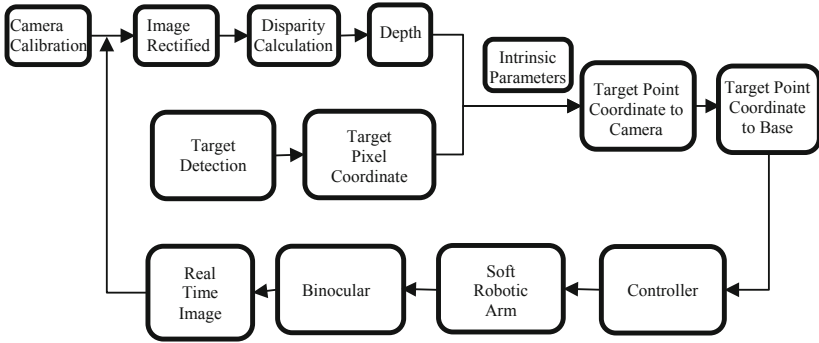


Fig. 3. Block diagram of visual servoing control by binocular vision

3.1 Modeling of Binocular and Monocular

The 3D coordinates of an object in real life can be determined with binocular stereo vision technology. Figure 4 below demonstrates the principle of binocular stereo vision. In Fig. 4, O_L and O_R are optical centers of the left and right cameras. Suppose two cameras have identical intrinsic and external parameters, which include f (focal length), B (the distance between optical centers), two cameras being on the same plane, and equal Y coordinates of their projection centers. The point P in space has imaging points in two cameras as P_{left} and P_{right} .

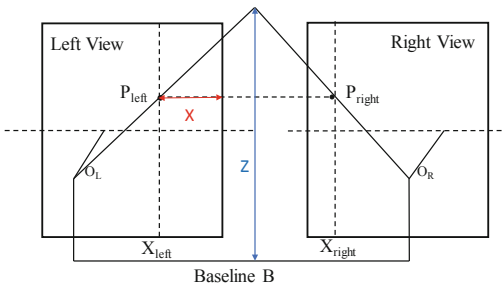


Fig. 4. The principle of binocular stereo vision

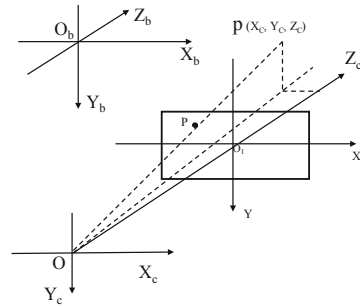


Fig. 5. Principle of camera projection

From trigonometry,

$$X_{left} = f \frac{x}{z} \tag{1}$$

$$X_{right} = f \frac{(x - B)}{z} \quad (2)$$

$$D = X_{left} - X_{right} \quad (3)$$

where X_{left} , X_{right} are the values of the target point in X direction of the left and right imaging planes, and x is the length of target point to the left camera in X direction. z is the depth from the target point to the binocular center. D is the disparity.

Solve (1), (2), (3) simultaneously to calculate depth as well. Consequently, it can be derived that:

$$z = \frac{Bf}{D} \quad (4)$$

During the imaging process of the camera, there exist 4 coordinate frames, which are pixel, image, camera and world coordinate systems. Figure 5 demonstrates the principle of camera imaging.

The relation between the pixel coordinate frame and the camera coordinate frame can be expressed with homogeneous matrices:

$$\begin{bmatrix} u \\ v \\ 1 \\ 1 \end{bmatrix} = \frac{1}{z_c} \begin{bmatrix} \frac{1}{d_x} & 0 & u_0 & 0 \\ 0 & \frac{1}{d_y} & v_0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \quad (5)$$

where u, v represent the column and row numbers of the pixel in the image, u_0, v_0 represent pixel coordinate of the principle point, and d_x, d_y are the physical measurements of the unit pixel on the horizontal and vertical axes. Because depth will be considered in this issue, a 4×4 matrix format is used where x_c, y_c, z_c indicate the point in the camera coordinate frame, f represents the focal length of the camera. $\frac{f}{d_x}, \frac{f}{d_y}$ are abbreviated into f_x, f_y . Camera intrinsic parameters obtained from camera calibration can be shown below:

$$f_x = 2.3643976238380526 \times 10^2$$

$$u_0 = 1.5168679331906756 \times 10^2$$

$$f_y = 2.3486572411007802 \times 10^2$$

$$v_0 = 1.2106158962347398 \times 10^2$$

In this work, since the camera is mounted on the end effector, it is assumed that the camera coordinate frame and the robot end effector coordinate frame are the same.

3.2 Modeling of Soft Robotic Arm

The soft robotic arm studied in this paper is a light-weight backboneless soft robotic arm, consisting of 6 long elastic bellows installed circularly, and its end-effector position is characterized by bending angle α , rotation angle β , and length l . The geometry model of the soft arm is shown as follows in Fig. 6, and the frame transfer relation between the base and the end effector is shown in Fig. 7.

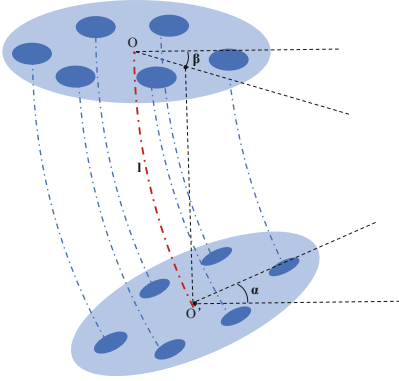


Fig. 6. Geometric model of the soft arm

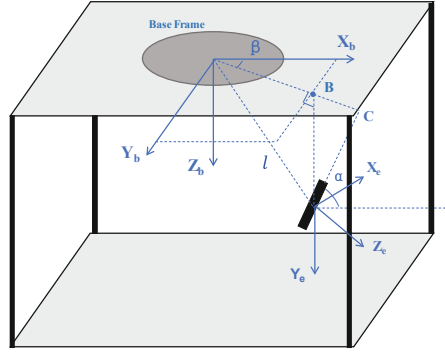


Fig. 7. Frame transfer relation

The homogeneous transformation from base to end effector coordinate frame is shown as follows:

$${}^eT_b = \begin{bmatrix} c^2\beta(c\alpha - 1) + 1 & s\beta c\beta(c\alpha - 1) & c\beta s\alpha & -\frac{1}{\alpha}c\beta(c\alpha - 1) \\ c\beta s\beta(c\alpha - 1) & s^2\beta(c\alpha - 1) + 1 & s\beta s\alpha & -\frac{1}{\alpha}s\beta(c\alpha - 1) \\ -c\beta s\alpha & -s\beta s\alpha & c\alpha & \frac{1}{\alpha}s\alpha \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (6)$$

Finally, (5) and (6) can be solved to obtain the result:

$$\begin{bmatrix} u \\ v \\ 1 \\ 1 \end{bmatrix} = \frac{1}{Z(t)} \Omega_L {}^eT_b(t) \begin{bmatrix} x_b \\ y_b \\ z_b \\ 1 \end{bmatrix} \quad (7)$$

in which Ω_L is the intrinsic parameter of the left camera.

Moreover, from calculation, it can be obtained that the target α, β, l are:

$$\alpha = 2 \arctan\left(\frac{\sqrt{x^2 + y^2}}{z}\right) \quad (8)$$

$$\beta = \arctan \frac{y}{x} \quad (9)$$

$$l = \frac{\sqrt{x^2 + y^2 + z^2}}{\sin\left(\arctan \frac{\sqrt{x^2 + y^2}}{z}\right)} \cdot \arctan \frac{\sqrt{x^2 + y^2}}{z} \quad (10)$$

Plug the coordinates of the target point in the base coordinate frame into (8), (9), (10) and obtain the target soft robotic arm parameters α , β , l to control the arm close to the target point.

4 Improved Binocular Measurement Performance

4.1 Measurement Accuracy Optimization

Conversion of disparity pixel values to real distance. Since the soft robotic arm will combined with a gripper in the end and the length of gripper is 60 cm. The experiment is designed based on the actual range of 0–120 cm, with a gradient of 1 cm. And for each group, 10 sets of continuous pixel value output are recorded, which amounts to be a total of 1600 sets of data, as shown below in Fig. 8.

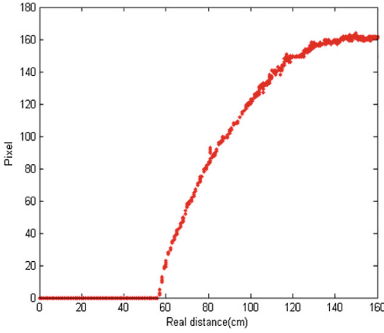


Fig. 8. Relationship between pixel and real distance

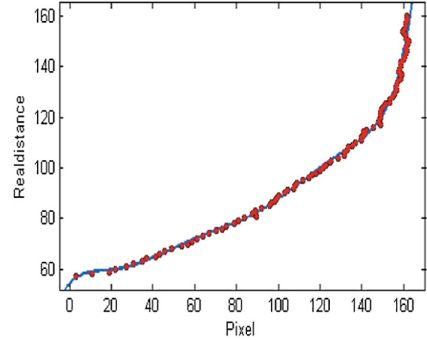


Fig. 9. Curving fitting result

Two regions are omitted in the disparity map during the experiment, including blind spot region when the distance is very close, and the undesirable low-resolution region when the distance is very large. As a result, a clear pattern between the pixel values and the real distances can be discovered. The data is then fitted with a high degree polynomial to obtain a more accurate fit curve, and the result is shown in Fig. 9. The curving fitting formula (11) is used to calculate the actual distance.

$$\text{Real Distance} = p_1x^7 + p_2x^6 + p_3x^5 + p_4x^4 + p_5x^3 + p_6x^2 + p_7x + p_8 \quad (11)$$

where x is the value of pixel, and $p_1 = -4.847 \times 10^{-12}$, $p_2 = 2.856 \times 10^{-9}$, $p_3 = -5.756 \times 10^{-7}$, $p_4 = 2.668 \times 10^{-5}$, $p_5 = 0.006586$, $p_6 = -1.101$, $p_7 = 67.91$, $p_8 = -1532$.

4.2 Measurement Time Improvement

One challenge with binocular disparity is its overly long calculation time, which considerably restrains its application in real-time measurement. The algorithm of the binocular disparity map mainly calculates the difference of the corresponding points according to the matching of each pixel in the left and right images, hence, if a more accurate disparity map is to be obtained, a large amount of time will be consumed in the disparity calculation of left and right images. For example, during the initial experiment, the disparity calculation time for big-scale images was as long as 5 s, which certainly did not meet the time requirements for real-time measurements. To solve this problem, the proposed solution is to reduce the size of the target measurement area as shown in Fig. 10. The specific method is to take a square area of a certain size around the target point, and only perform disparity calculation on the pixels of this selected area. The relationship between the size of the region and the disparity calculation time by this method, and the relative error at each size is shown in Fig. 11.

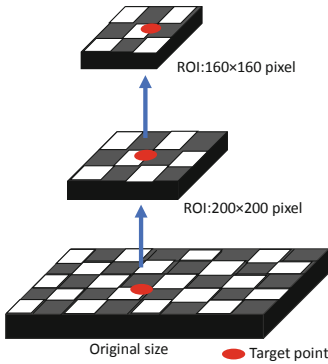


Fig. 10. Method of speeding up calculation time

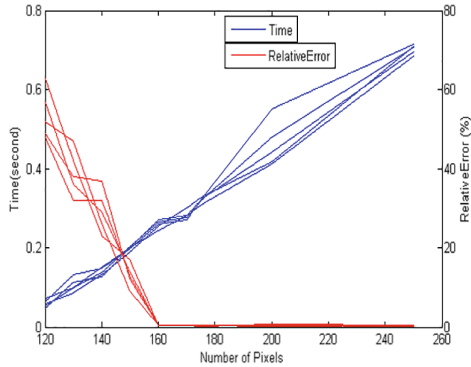


Fig. 11. Relation between time/error and region

From the relationship in Fig. 11, the calculation time is relatively short when the size of the region is 160×160 pixels with a high accuracy, for which the single running time of the vision system is about 0.2 s. Therefore, in the final calculation process, in order to maximize the real-time performance, a pixel area of 160×160 is selected.

4.3 Distance Measurement Error

According to the improvement method above, the distance measurement experiment was conducted. Experiments were conducted on the binocular measurement system, with a distance ranging from 57 cm to 120 cm. The measurement was recorded 10 times at each actual distance by two binocular disparity models which are linear and curve fitting. And the error of the experiment was calculated by the following expression

$$\text{relative error} = \frac{\text{error}}{\text{real depth}} \quad (12)$$

The error analysis chart is shown as follows in Fig. 12.

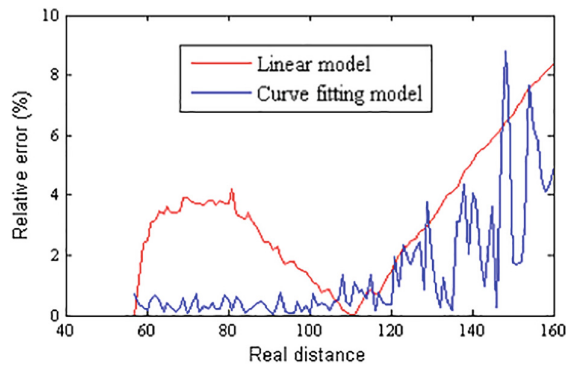


Fig. 12. Relative average error by real distance

Because the fitting model has a significantly higher measurement accuracy in the working area than the linear model. It can be seen from Fig. 12 that the binocular measurement system meets the required error in the working range well.

5 Experiment

5.1 Visual Servoing Experiment

Binocular is mounted on the center of the end effector and put the mark point at any position in the camera view. In experiment 1, the visual servoing control of robotic arm is set to the target point at a fixed distance. The initial distance is set to be 60 cm, then the robotic arm moves away from the target robotic arm, and finally, it comes back to 60 cm. The experiment 1 configuration is shown in Fig. 13. Experiment 2 is a real-time tracking of the target object. When the target moves in the view of the camera, the camera tracks the target and records the pixel coordinates. The experiment 2 configuration is shown in Fig. 14.

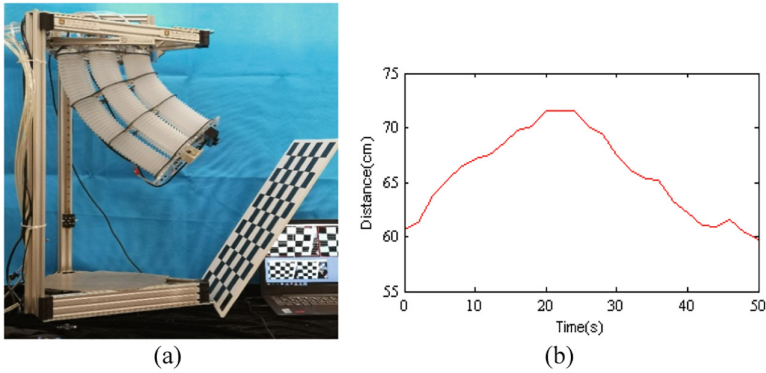


Fig. 13. (a) shows the visual servoing of the robot arm under bending condition to the fixed distance from the target point and (b) shows the depth change over time in experiment 1.

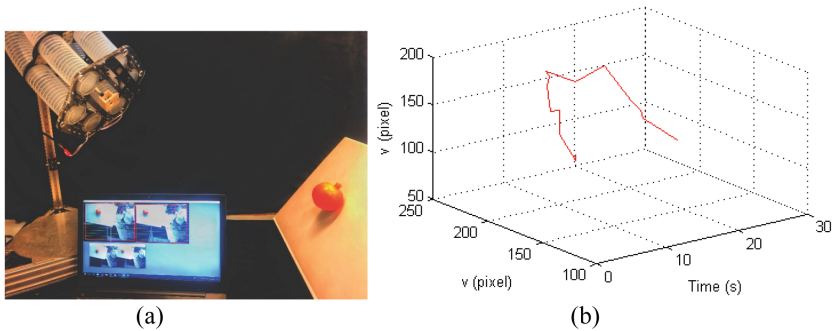


Fig. 14. (a) shows the real-time tracking of the target object. (b) shows the pixel coordinate change in the tracking task.

Based on the two experiments above, a third experiment to detect the object spatial position is operated. A control signal moving the soft robotic arm is generated so that the end of the arm points at the target and moves to a fixed distance from the target object. The experiment configuration is shown in Fig. 15. The target pixel coordinate in the image is tracked to the middle of the camera view. And the distance information is measured in real-time to reach the fixed distance in the task. The change of target pixel and distance are recorded and shown in Fig. 16.

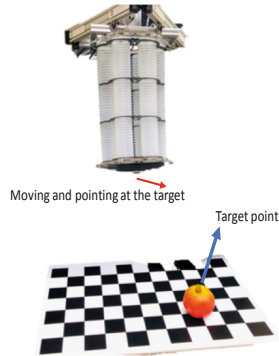


Fig. 15. Experiment 3 configuration

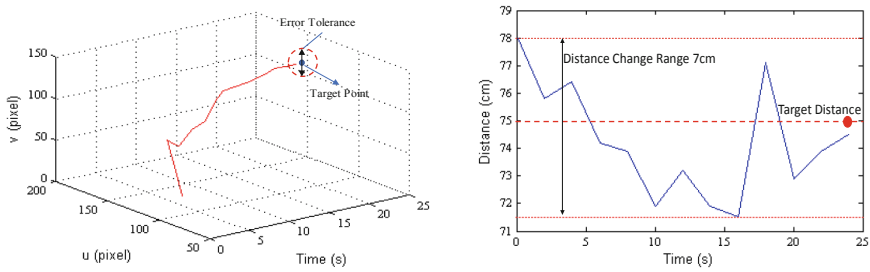


Fig. 16. The change of pixel coordinate and distance

6 Conclusion

In this work, the binocular vision system can rather accurately control the distance from the end effector to the target during the motion of the soft robotic arm so that the end-effector can perform the task within the acceptable error range.

This simple and feasible binocular servoing control method will provide new ideas and inspirations for solving the difficulties in soft robots controlling and sensing fields.

In the future, the aim is to achieve a more accurate estimation of the end position of the arm, enabling the motion control to be more precise, and to equip the soft robotic arms with better capabilities of fulfilling more tasks.

References

1. Zhou, J., Chen, X., Chang, U., Pan, J., Wang, W., Wang, Z.: Intuitive control of humanoid soft-robotic hand BCL-13. In: 2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids), pp. 314–319 (2018)
2. Chen, X., Hu, T., Song, C., Wang, Z.: Analytical solution to global dynamic balance control of the Acrobot. In: 2018 IEEE International Conference on Real-time Computing and Robotics, pp. 405–410 (2019)

3. Chen, X., Yi, J., Li, J., Zhou, J., Wang, Z.: Soft-actuator-based robotic joint for safe and forceful interaction with controllable impact response. *IEEE Robot. Autom. Lett.* **3**(4), 3505–3512 (2018)
4. Polygerinos, P., et al.: Modeling of soft fiber-reinforced bending actuators. *IEEE Trans. Robot.* **31**(3), 778–789 (2015)
5. Wang, Z., Peer, A., Buss, M.: Fast online impedance estimation for robot control. In: *IEEE 2009 International Conference on Mechatronics, ICM 2009*, pp. 1–6 (2009)
6. Yi, J., et al.: Customizable three-dimensional-printed origami soft robotic joint with effective behavior shaping for safe interactions. *IEEE Trans. Robot.* (2018, accepted)
7. Zhou, J., Chen, X., Li, J., Tian, Y., Wang, Z.: A soft robotic approach to robust and dexterous grasping. In: *2018 IEEE International Conference on Soft Robotics, RoboSoft 2018*, no. 200, pp. 412–417 (2018)
8. Wang, Z., Peer, A., Buss, M.: An HMM approach to realistic haptic human-robot interaction. In: *Proceedings - 3rd Joint EuroHaptics Conference and Symposium on Haptic Interfaces Virtual Environment Teleoperator Systems World Haptics 2009*, pp. 374–379 (2009)
9. Wang, Z., Sun, Z., Phee, S.J.: Haptic feedback and control of a flexible surgical endoscopic robot. *Comput. Methods Programs Biomed.* **112**(2), 260–271 (2013)
10. Wang, Z., Polygerinos, P., Overvelde, J.T.B., Galloway, K.C., Bertoldi, K., Walsh, C.J.: Interaction forces of soft fiber reinforced bending actuators. *IEEE/ASME Trans. Mechatron.* **22**(2), 717–727 (2017)
11. Viry, L., et al.: Flexible three-axial force sensor for soft and highly sensitive artificial touch. *Adv. Mater.* **26**, 2659–2664 (2014). <https://doi.org/10.1002/adma.201305064>
12. Bauer, S., Bauer-Gogonea, S., Graz, I., Kaltenbrunner, M., Keplinger, C., Schwödiauer, R.: 25th anniversary article: a soft future: from robots and sensor skin to energy harvesters. *Adv. Mater.* **26**, 149–162 (2014). <https://doi.org/10.1002/adma.201303349>
13. Ilievski, F., Mazzeo, A.D., Shepherd, R.F., Chen, X., Whitesides, G.M.: Soft robotics for chemists. *Angew. Chem. Int. Ed.* **50**, 1890–1895 (2011). <https://doi.org/10.1002/anie.201006464b>
14. Espiau, B., Chaumette, F., Rives, P.: A new approach to visual servoing in robotics. *IEEE Trans. Robot. Autom.* **8**(3), 313–326 (1992)
15. Yi, J., Chen, X., Song, C., Wang, Z.: Fiber-reinforced origamic robotic actuator. *Soft Robot.* **5**(1), 81–92 (2017)
16. Wilson, W.J., Hulls, C.W., Bell, G.S.: Relative end-effector control using Cartesian position based visual servoing. *IEEE Trans. Robot. Autom.* **12**(5), 684–696 (1996)
17. Allen, P.K., Yoshimi, B., Timcenko, A.: Real-time visual servoing. In: *Proceedings. 1991 IEEE International Conference on Robotics and Automation, Sacramento, CA, USA*, vol. 1, pp. 851–856 (1991)
18. Jagersand, M., Fuentes, O., Nelson, R.: Experimental evaluation of uncalibrated visual servoing for precision manipulation. In: *Proceedings of International Conference on Robotics and Automation, Albuquerque, NM, USA*, vol. 4, pp. 2874–2880 (1997)
19. Vahrenkamp, N., Wieland, S., Azad, P., Gonzalez, D., Asfour, T., Dillmann, R.: Visual servoing for humanoid grasping and manipulation tasks. In: *Humanoids 2008 - 8th IEEE-RAS International Conference on Humanoid Robots, Daejeon*, pp. 406–412 (2008)
20. De Luca, A., Oriolo, G., Robuffo Giordano, P.: Feature depth observation for image-based visual servoing: theory and experiments. *Int. J. Robot. Res.* **27**(10), 1093–1116 (2008). <https://doi.org/10.1177/0278364908096706>
21. Malis, E.: Visual servoing invariant to changes in camera-intrinsic parameters. *IEEE Trans. Robot. Autom.* **20**(1), 72–81 (2004)

22. De Luca, A., Oriolo, G., Giordano, P.R.: On-line estimation of feature depth for image-based visual servoing schemes. In: *Proceedings 2007 IEEE International Conference on Robotics and Automation*, Roma, pp. 2823–2828 (2007)
23. Fujimoto, H.: Visual servoing of 6 DOF manipulator by multirate control with depth identification. In: *42nd IEEE International Conference on Decision and Control (IEEE Cat. No. 03CH37475)*, Maui, HI, vol. 5, pp. 5408–5413 (2003)
24. Zhang, Z.: A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(11), 1330–1334 (2000)
25. Zhang, Z.: Determining the epipolar geometry and its uncertainty: a review. *Int. J. Comput. Vis.* **27**(2), 161–195 (1998)
26. Davison, A.: Real-time simultaneous localization and mapping with a single camera. In: *Proceedings of International Conference Computer Vision*, pp. 1403–1410, October 2003
27. Godard, C., Mac Aodha, O., Brostow, G.J.: Unsupervised monocular depth estimation with left-right consistency. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 270–279 (2017)
28. Kuznietsov, Y., Stuckler, J., Leibe, B.: Semi-supervised deep learning for monocular depth map prediction. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6647–6655 (2017)