

An assistive System to Guide Visually Impaired People Behind Work Desks

Research Article

Amirmohammad Barsalani¹, Hamidreza Mohammadi Daniali² , Arman Mardani³

 [10.22067/jacsm.2025.91402.1310](https://doi.org/10.22067/jacsm.2025.91402.1310)

Abstract-- According to the World Health Organization, over 38 million people are blind, and more than 110 million suffer from severe visual impairments. This research aims to develop a low-cost, accessible assistance system for blind individuals in work environments, utilizing a single camera (e.g., a smartphone camera) and pre-trained deep learning models. By employing computationally efficient image processing techniques, the system optimizes hardware and processing expenses. It detects objects and the user's hand, computes the heights of objects and their distances to the hand, and then provides real-time audio guidance (e.g., verbal directions) to enable safe object retrieval. Compared to expensive alternatives like dual-camera setups or Kinect sensors, the core innovation here lies in dramatically reducing costs while retaining functional accuracy, as millimeter-level precision is unnecessary for workplace navigation. Moreover, the system performance will be validated to achieve a distance estimation error of approximately 5-10 mm; its accuracy is anticipated to match that of comparable stereo vision approaches.

Keywords: Low-cost, Blind Individuals, Workstation Assistance, Object Detection, Audio Guidance, Pre-trained Models

1- Introduction

Visual impairment poses a global challenge, with the World Health Organization (WHO) reporting approximately 38 million individuals living with blindness and over 110 million suffering severe visual disabilities [1]. Assistive technologies for this demographic aim to enhance independence, safety, and quality of life by addressing functional limitations in daily environments. Historically, designing such systems faced significant hurdles due to computational constraints in machine vision, particularly before the advent of deep learning. Most existing solutions—such as ultrasonic sensor-equipped canes [2], stereo-vision glasses [3], or smart gloves [4]—prioritize navigation in open environments. However, workplace settings (e.g., desk-based tasks) remain critically underserved. Current approaches often

rely on expensive hardware like depth sensors (ToF) [4], dual-camera rigs [5], or Kinect [6], incurring high computational and financial costs. This gap necessitates a cost-effective system optimized for object interaction in confined spaces, where one millimeter precision is nonessential.

To address this, we propose a novel smartphone-based framework leveraging single-camera input, pre-trained models (YOLOv8n [7] + MediaPipe [8]), and lightweight image processing. Our innovation eliminates depth sensors and stereo imaging, instead employing mathematical methods to estimate object height/hand distance in real time. Key advantages of our proposed framework include drastic cost reduction through ubiquitous hardware (mobile cameras), computational efficiency enabling real-time guidance, and audio-based guidance ensuring safe object retrieval. The system targets functional accuracy (5–10 mm error), rivaling stereo vision without its hardware overhead. By focusing on desktop workspaces, this research pioneers the use of accessible technology to promote occupational independence among visually impaired individuals.

2- Materials and Methods

The system leverages a single smartphone camera (user-provided) and Python-based processing to deliver low-cost spatial guidance for visually impaired users in a desktop environment. Moreover, camera calibration is performed using OpenCV with a chessboard pattern, deriving intrinsic parameters (focal length, principal point) and extrinsic parameters (camera-to-table pose) to estimate the transformation matrix (T_c). Coordinate conversion from pixels to real-world space is achieved through Equation 1.

$$\begin{aligned} K_{ij} &= \begin{bmatrix} P_x & P_y & 1 \\ X & Y & Z \end{bmatrix} \\ K_{ij}^w &= \begin{bmatrix} X & Y & Z & 1 \end{bmatrix} \end{aligned} \quad (1)$$

* Manuscript received 2024 December 29, Revised 2025 May 12, Accepted 2025 July 20.

¹ Master, Department of Mechanical Engineering, Babol Noshirvani University of Technology, Mazandaran, Iran.

Email: amirmohammadbarsalani@gmail.com

² Corresponding author. Professor, Department of Mechanical Engineering, Babol Noshirvani University of Technology, Mazandaran, Iran, **Email:** mohammadi@nit.ac.ir

³ Assistant Professor, Department of Mechanical Engineering, Babol Noshirvani University of Technology, Mazandaran, Iran, **Email:** armanmardani@nit.ac.ir

$$K_{ij} = T_c \cdot K_{ij}^w$$

In which K_{ij} represents a pixel coordinate with horizontal (P_x) and vertical (P_y) components, while its real-space counterpart K_{ij}^w consists of X, Y, and Z spatial dimensions. This calibration assumes objects rest on the table plane ($Z = 0$), simplifying spatial computations while maintaining target accuracy of 5–10 mm. Object detection utilizes YOLOv8n for its optimal speed-accuracy balance, identifying bounding boxes of target items. For tall objects, the estimated height is the value that, when substituted into Equation 1, yields the pixel coordinate closest to point z in Figure 1; by assuming similar x-y coordinates for top/bottom midpoints in real space.

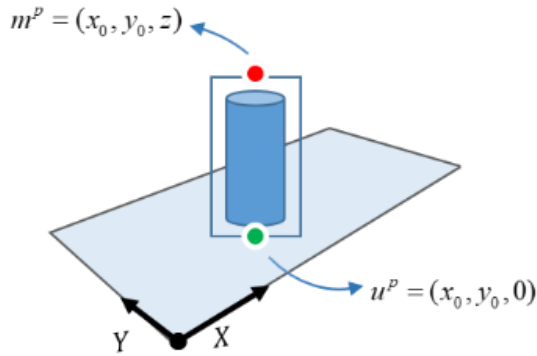


Fig. 1 Two points of interest for estimating the height of objects (The superscript P represents the real-world coordinates for points m and u.)

MediaPipe is employed for hand detection and tracking, utilizing the middle fingertip (Landmark 12 in Fig. 2(a)) as

the primary reference point and the proximal joint (Landmark 11 in Fig. 2(a)) to determine hand orientation. Spatial guidance uses a hand-centric coordinate system: Axis F aligns with the middle finger (Landmarks 11→12), while Axis G is perpendicular to F at the fingertip, dividing the workspace into quadrants. Object positions are translated into clock directions (e.g., "Move 15 cm toward 2 o'clock") based on angular displacement from Axis F and Euclidean 2-norm. Voice interaction via pyttsx3 supports two commands: "Report" (lists detected objects) or a named-object retrieval (triggers guided navigation).

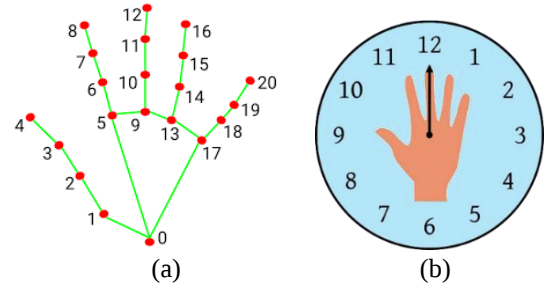


Fig. 2 (a) Mediapipe-detectable landmarks on hand, (b) Clock-based directional Guidance view

The workflow begins with one-time calibration, followed by continuous operation (Fig.3): voice input → image capture → YOLOv8n/MediaPipe inference → coordinate transformation → geometric analysis → audio feedback. Computational efficiency is achieved through model optimizations and mathematical approximations, avoiding depth sensors or stereo vision. Testing assumes static objects and fixed camera height.

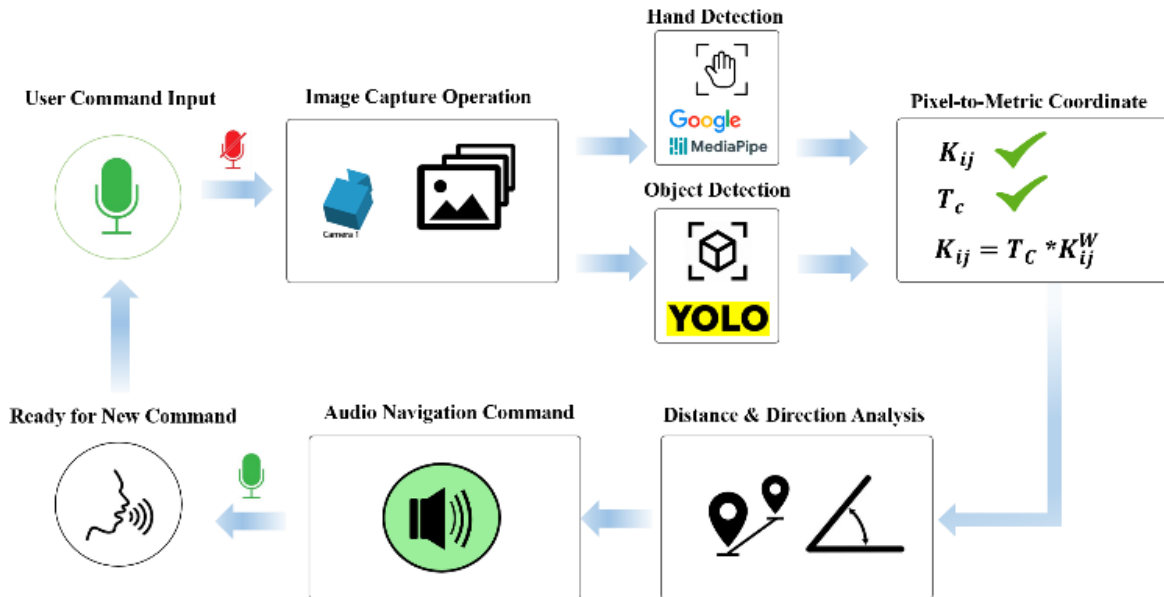


Fig. 3 Overview of how the system works

3- Results

The system demonstrated sub-centimeter spatial accuracy in desktop-scale environments, achieving a mean distance estimation error of 6.1 mm (range: 5–8.5 mm) across geometric validation tests. While the magnitude of the error increased with distance from the camera due to pixel density limitations, performance remained comparable to stereo vision systems (1.22% error reported in prior work [9]). Height estimation for a standardized 150 mm object yielded a mean error of 11 mm (range: 4–21 mm), with optimal accuracy observed in mid-range positions. Computational efficiency was validated on an Intel Core i5-4300M processor, where hand landmark detection (MediaPipe) averaged 0.4 s per image, object recognition (YOLOv8n) required 1.1 s, and audio guidance generation added only 0.03 s latency per command cycle. The total processing time confirms real-time viability for desktop assistance scenarios, with vision tasks dominating the computational load. These results collectively validate the system's core capability to deliver stereo vision-comparable spatial precision using low-cost smartphone hardware while maintaining practical responsiveness.

4- Conclusion

The proposed system achieved spatial accuracy comparable to stereo vision—with mean distance and height estimation errors of 6.1 mm and 11 mm, respectively—while fundamentally overcoming its computational and hardware limitations. By leveraging a single smartphone camera, optimized deep learning frameworks (YOLOv8n for object detection, MediaPipe for hand tracking), and a constrained 2D desktop workspace, the architecture reduces processing load and eliminates the need for specialized dual-camera rigs. This research pioneer assistive technology specifically designed for workplace independence, transforming ubiquitous smartphones into tools for visually impaired users to interact with desktop objects via real-time audio guidance. Future enhancements should explore object segmentation for spatial refinement, language model integration for natural voice interaction, server-based processing acceleration, and haptic interfaces to enrich multimodal feedback—collectively advancing toward assistive systems that empower occupational autonomy through accessible technology.

5-References

- [1] World Helath Organization, "Rehabilitation," [Online]. [Accessed: Apr. 22, 2024]. Available: <https://www.who.int/news-room/fact-sheets/detail/rehabilitation>
- [2] T. L. Narayani, M. Sivapalanirajan, B. Keerthika, M. Ananthi, and M. Arunarani, "Design of smart cane with integrated camera module for visually impaired people," in Proceeding 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), 2021, pp. 999–1004. <https://doi.org/10.1109/ICAIS50930.2021.9395840>
- [3] O. Younis, W. Al-Nuaimy, F. Rowe, M. H. Alomari, "A smart context-aware hazard attention system to help people with peripheral vision loss," *Sensors*, vol. 19, no. 7, p. 1630, 2019. <https://doi.org/10.3390/s19071630>
- [4] J. Kilian, A. Neugebauer, L. Scherffig, S. Wahl, "The unfolding space glove: A wearable spatio-visual to haptic sensory substitution device for blind people," *Sensors*, vol. 22, no. 5, p. 1859, 2022. <https://doi.org/10.3390/s22051859>
- [5] M. Aharchi, M. H. A. Kbir, "Localization and Navigation System for Blind Persons Using Stereo Vision and a GIS," in WITS 2020: Proceedings of the 6th International Conference on Wireless Technologies, Embedded, and Intelligent Systems, Springer, 2022, pp. 1–12. https://doi.org/10.1007/978-981-33-6893-4_35
- [6] H. Takizawa, S. Yamaguchi, M. Aoyagi, N. Ezaki, "Kinect cane: An assistive system for the visually impaired based on the concept of object recognition aid." *Personal and Ubiquitous Computing* 19. vol. 5 pp. 955-965. (2015).
- [7] M. Sohan, T. Sai Ram, C.V. Rami Reddy, "A review on YOLOv8 and its advancements," in International Conference on Data Intelligence and Cognitive Informatics, Springer, 2024, pp. 1–8. https://doi.org/10.1007/978-981-99-7962-2_39
- [8] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C. L. Chang, and M. Grundmann, "Mediapipe hands: On-device real-time hand tracking," *arXiv preprint arXiv:2006.10214*, 2020. <https://doi.org/10.48550/arXiv.2006.10214>
- [9] R. A. Setyawan, R. Soenoko, P. Mudjirahardjo, and M. A. Choiron, "Measurement accuracy analysis of distance between cameras in stereo vision," In 2018 Electrical power, electronics, communications, controls and informatics seminar (EECCIS), 2018, pp. 169-172. <https://doi.org/10.1109/EECCIS.2018.8692999>



سیستم کمکی برای هدایت افراد دارای اختلالات بینایی پشت میزهای کار*

مقاله پژوهشی

آرمان مردانی^(۳)

حمیدرضا محمدی دانیالی^(۲)

امیر محمد برسلانی^(۱)

10.22067/jacsm.2025.91402.1310

چکیده بر اساس گزارش سازمان بهداشت جهانی، بیش از ۳۸ میلیون نفر نابینا و بیش از ۱۱۰ میلیون نفر با مشکلات جدی بینایی روبه‌رو هستند. این پژوهش با هدف ارائه سیستمی کم‌هزینه و در دسترس برای کمک به افراد نابینا در محیط‌های کاری، از یک دوربین تک‌گانه (مثلاً دوربین تلفن همراه) و مدل‌های از پیش آموزش دیده بهره می‌برد. با استفاده از روش پردازش تصویر با حجم محاسباتی کم، هزینه‌های سخت‌افزاری و پردازشی بهینه می‌شود. در این سیستم، اشیاء و دست کاربر تشخیص داده می‌شوند و ارتفاع اجسام و فاصله آن‌ها تا دست محاسبه می‌شود. سپس از طریق راهنمایی صوتی، کاربر برای دسترسی ایمن به اشیاء هدایت می‌شود. در مقایسه با روش‌های پرهزینه مانند سیستم‌های دو دوربین یا سنسور کینکت، نوآوری اصلی این پژوهش کاهش چشمگیر هزینه‌ها با حفظ دقت کاربردی است؛ چرا که برای هدایت افراد نابینا در محیط‌های کاری، دقت میلی‌متری ضروری نیست. عملکرد سیستم به‌منظور دستیابی به خطای تخمین فاصله‌ای در حدود ۵ تا ۱۰ میلی‌متر بررسی خواهد شد؛ انتظار می‌رود دقت این سیستم در مقایسه با روش استریو ویژن مشابه باشد.

واژه‌های کلیدی سیستم کم‌هزینه، افراد نابینا، تشخیص اشیاء، راهنمایی صوتی، مدل‌های پیش‌آموزش دیده، دوربین تک‌گانه.

مقدمه

بینایی و عملکرد بینایی در نظر گرفته شود. اولین جنبه‌ای که باید در نظر گرفته شود، ساختار و عملکرد چشم است و سپس توانایی‌های فرد و پیامدهای عملی از دست دادن بینایی [3]. خدمات جامع توانبخشی بینایی و سیستم‌های کمکی به افراد دارای اختلالات بینایی این امکان را می‌دهد که کنترل بیشتری بر محیط خود داشته باشند، این امر به افزایش اعتماد به نفس، کاهش خطر افسردگی و اضطراب، و بهبود کیفیت زندگی منجر می‌شود [4].

در گذشته و پیش از ظهور الگوریتم‌های مبتنی بر یادگیری عمیق، طراحی سیستم‌های کمک‌رسان برای افراد نابینا یا کم‌بینا با

طبق آماری که سازمان بهداشت جهانی ارائه داده است، به طور تخمینی حدود ۳۸ میلیون فرد نابینا در دنیا وجود دارند و بیش از ۱۱۰ میلیون نفر از عارضه‌های بینایی رنج می‌برند که ممکن است برخی از این موارد به نابینایی کامل ختم شوند [1]، این موضوع بیانگر میزان اهمیت و جدی بودن مساله است. هدف از توانبخشی بینایی و ارائه سیستم‌های کمکی بهبود استقلال، ایمنی، کارایی و کیفیت کلی زندگی برای بیمارانی است که با از دست دادن بینایی یا اختلالات در عملکرد بینایی خود مواجه هستند [2]. جهت درک جوانب مختلف موضوع لازم است چهار جنبه اصلی از

* تاریخ دریافت مقاله ۱۴۰۳/۱۰/۹ تاریخ پذیرش آن ۱۴۰۴/۴/۲۹ می‌باشد.

(۱) دانشکده مهندسی مکانیک - دانشگاه صنعتی نوشیروانی بابل - مازندران - بابل - ایران.

(۲) نویسنده مسئول/دانشکده مهندسی مکانیک - دانشگاه صنعتی نوشیروانی بابل - مازندران - بابل - ایران.

(۳) دانشکده مهندسی مکانیک - دانشگاه صنعتی نوشیروانی بابل - مازندران - بابل - ایران.

چالش‌های زیادی همراه بود. دلیل اصلی این مشکل، نیاز به درک داده‌های تصویری توسط ماشین بود؛ داده‌هایی که به دلیل جزئیات فراوان و ویژگی‌های پیچیده، بدون استفاده از شبکه‌های عصبی به‌سختی تحلیل می‌شدند. اما امروزه شبکه‌های عصبی-پایه‌ی اصلی الگوریتم‌های یادگیری عمیق - قادر به درک و پردازش داده‌های بسیار پیچیده هستند. پژوهشگران معماری‌های مختلف شبکه‌های عصبی را در حوزه‌های گوناگونی مانند رباتیک [6,5]، بهینه‌سازی [9-7]، کنترل [11,10] و بسیاری زمینه‌های دیگر به‌کار می‌برند. در حوزه‌ی پردازش تصویر و بینایی ماشین [14-12] نیز معماری‌های متنوعی از شبکه‌های عصبی توسعه یافته‌اند که به‌طور ویژه برای درک محتوای تصاویر طراحی شده‌اند.

در ادامه به بخشی از پژوهش‌های صورت پذیرفته در حوزه سیستم‌های کمکی برای افراد دارای اختلالات بینایی پرداخته شده است. نارایانی و همکاران [15]، به منظور بهبود عملکرد افراد نابینا در راه‌رفتن، با استفاده از دو سنسور اولتراسونیک، یک ماژول دوربین و یک حسگر مادون قرمز غیرفعال، یک عصا طراحی کردند که قادر به شناسایی موانع موجود بر سر راه فرد نابینا می‌باشد و در صورت نزدیک شدن فرد به آن‌ها، هشدارهای صوتی لازم را ارائه می‌دهد. مارکون و اورتوچلی [16]، طی تحقیقی میزان مؤثر بودن شناسایی اجسام به صورت بی‌درنگ را در شرایط کنترل نشده در یک محیط واقعی مورد بررسی قرار دادند در طول این فرآیند عملکرد ورژن‌های مختلف یولو تا ورژن چهارم مورد ارزیابی قرار گرفتند و در نهایت بهترین نتیجه توسط یولو ورژن چهارم حاصل شد که نتیجه آن بیش از ۹۳ درصد موفقیت‌آمیز بود. کیلیان و همکاران [17]، دستکش هوشمندی را طراحی کردند که با استفاده از دوربین زمان پرواز تعبیه شده روی آن، این امکان را داشت تا اجسام، موانع و همینطور فاصله آن‌ها را تشخیص دهد. پس از آنکه جسم یا مانعی تشخیص داده شود با استفاده از ۹ عملگر لرزشی تعبیه شده بر روی دستکش، به فرد محل قرار گیری آن را گزارش می‌دهد.

اهرچی و کبیر [18]، طرحی را پیشنهاد دادند. به این صورت که با استفاده از استریوویژن و سنسور GPS میتوان افراد نابینا را به سمت هدف و مسیری که دارند هدایت کرد. آدپ و همکاران [19]، برای توسعه استقلال افراد نابینا در راه رفتن، سیستم هوشمندی را طراحی کردند. در این پروژه سنسورهای فروسرخ

و دوربین مورد استفاده قرار گرفتند و محل پردازش داده‌ی حاصل شده در یک برد رزبری پای می‌باشد. سپس فرمان‌های صوتی لازم جهت هدایت فرد نابینا صادر می‌شود. زنتو و همکاران [20]، عصای هوشمندی را طراحی کردند که توانایی تشخیص اجسام و موانع تا فاصله ۲ متری را دارد و همینطور در این فاصله طولی قادر به تشخیص ارتفاع اجسام تا ۲ متر نیز می‌باشد. بانو و همکاران [21]، یک سیستم پوششی را ارائه دادند که قابلیت شناسایی موقعیت‌های ترافیکی، چراغ‌های راهنما، دوچرخه‌سواران، پل‌های عابر پیاده، عابرین پیاده، اتومبیل‌ها و برخی موارد دیگر را دارد. این سیستم با بهره‌گیری از استریوویژن به افرادی دارای اختلالات بینایی کمک می‌کند تا در محیط‌های بیرونی و همینطور داخلی بتوانند عملکرد بهتری داشته باشند. یونس و همکاران [22]، عینک هوشمندی را طراحی کرده‌اند که میزان خطر موجود در طول مسیر فردی که بخشی از توان بینایی خود را از دست داده است را بررسی می‌کند. بویمر و همکاران [23]، با استفاده از یک دوربین و یک کمر بند لمسی، سیستمی را پیاده‌سازی کردند که به افراد دارای اختلالات بینایی کمک می‌کند تا حالت چهره مخاطب آن‌ها را شناسایی و گزارش کند. بوترا [24]، سیستمی را جهت تشخیص مسیر امن برای افرادی که با عارضه‌های بینایی روبرو هستند طراحی و پیشنهاد کرده است. تنکون و همکاران [25]، مدلی را طراحی کردند که به افرادی که از اختلالات بینایی رنج می‌برند کمک می‌کند تا با استقلال بیشتری برای خرید به فروشگاه‌ها بروند.

پرادا و همکاران [26]، با استفاده از دو سنسور اولتراسونیک و یک سنسور فروسرخ، کمر بند هوشمندی را طراحی کردند که قابلیت تشخیص موانع با ارتفاعات مختلف را داشت. این ایده توسط جمعی از افرادی که اختلالات بینایی داشتند مطرح و ارائه شد. نمونه اولیه این طرح عملکرد خوبی در محیط‌های مختلف داشت اما قادر به تشخیص جهت موانع موجود بر سر راه افراد نبود. کانز و همکاران [27]، الگوریتمی را ارائه دادند که بر اساس آن زمان تخمینی برخورد فرد نابینا با موانع موجود بر سر راه محاسبه می‌شود. این طرح از متد جریان اپتیکال، مدل دوربین بین‌هول به منظور یافتن معادلات برخورد و تکنیک‌های پردازش تصویر برای کاهش خطا و نویز استفاده می‌کند. بویمر و همکاران [28]، با استفاده از یک دوربین و یک کمر بند لمسی، سیستمی را پیاده‌سازی کردند که به افراد دارای اختلالات بینایی کمک می‌کند

به رو هستند. با توجه به سبک زندگی امروزه که بخش قابل توجهی از زمان افراد پشت میزهای کار گذرانده می شود این نیاز حس می شود که افراد توانخواه نیز این امکان را داشته باشند تا بتوانند در فعالیت های خود استقلال بیشتری کسب کنند. نوآوری اصلی این پژوهش در طراحی سیستمی تک دوربینه و کم هزینه مبتنی بر تلفن همراه است که با هدف افزایش استقلال نابینایان در تعامل با اشیاء محیط های کاری ثابت (به ویژه فضای میز کار) توسعه یافته است. در این راستا، نخستین بار تعامل نابینایان با اشیاء در محیط بسته پشت میز به صورت هدفمند مورد توجه قرار گرفته است. سیستم پیشنهادی با به کارگیری ترکیبی از مدل یولو نسل هشتم سری نانو برای شناسایی اشیاء و مدل مدیاپایپ برای موقعیت یابی دست در فضای حقیقی، روشی ساده و ریاضی محور جهت تخمین ارتفاع و موقعیت اجسام بدون نیاز به سنسور عمق ارائه می دهد. همچنین روش پیشنهادی دقتی معادل روش استریوویژن - اما با دو مزیت بنیادین کاهش چشمگیر حجم محاسبات و حذف نیاز به سخت افزار اضافی (استفاده از یک دوربین به جای دو دوربین) - محقق ساخته است. این رویکرد کم هزینه با حداقل پردازش، گامی مؤثر در جهت کاربردهای عملیاتی محیط های واقعی است. در بخش دوم مقاله (مواد و روش ها)، معماری سیستم پیشنهادی شامل مراحل کالیبراسیون دوربین، شناسایی اشیاء و دست با به کارگیری مدل های یولو و مدیاپایپ، تخمین ارتفاع اجسام و شیوه هدایت دست تشریح می شود. بخش سوم (یافته ها) به ارزیابی عملکرد سیستم در سه محور خطای تخمین فاصله، دقت سنجش ارتفاع اجسام و هزینه زمانی پردازش می پردازد. در ادامه، محدودیت های فنی سیستم در بخش چهارم و نتیجه گیری همراه با پیشنهادات توسعه آینده در بخش پنجم ارائه می گردد.

مواد و روش ها

پژوهش پیش رو ایده و طرح اولیه از سیستمی را ارائه می دهد که قادر به شناسایی اجسام و دست انسان می باشد و پس از به انجام رساندن این مرحله، به تخمین فاصله اجسام از دست انسان که روی میز قرار گرفته است می پردازد. با بهره گیری از سیستم تعامل صوتی تعبیه شده در برنامه سیستم، پیام صوتی مناسب، جهت راهنمایی فرد به منظور دسترسی به جسم مورد نظر ارائه می شود. لازم به ذکر است، زبان برنامه نویسی پایتون در این پروژه مورد

تا حالت چهره مخاطب آن ها را شناسایی و گزارش کند. این سیستم به افراد کمک می کند در تعاملات اجتماعی عملکرد مناسبی داشته باشند و بتوانند کیفیت روابط خود را ارتقا بخشند. عملکرد سیستم ارائه شده، بهبود ۴۴/۴ درصدی در تشخیص عواطف مخاطبین توسط افراد نابینا داشته است. لین و همکاران [29]، جهت توسعه درک بصری افرادی که از اختلالات بینایی رنج می برند، سیستمی را طراحی کردند که با استفاده از یک دوربین رنگی-عمقی و الگوهای یادگیری عمیق قادر به تشخیص اجسام و شناسایی مسیر پیش روی فرد می باشد. این سیستم پس از تحلیل تصاویر، اجسام شناسایی شده را به صورت صوتی به کاربر گزارش می دهد و سپس مطابق خواسته او فرمان های هدایتی صادر می شود.

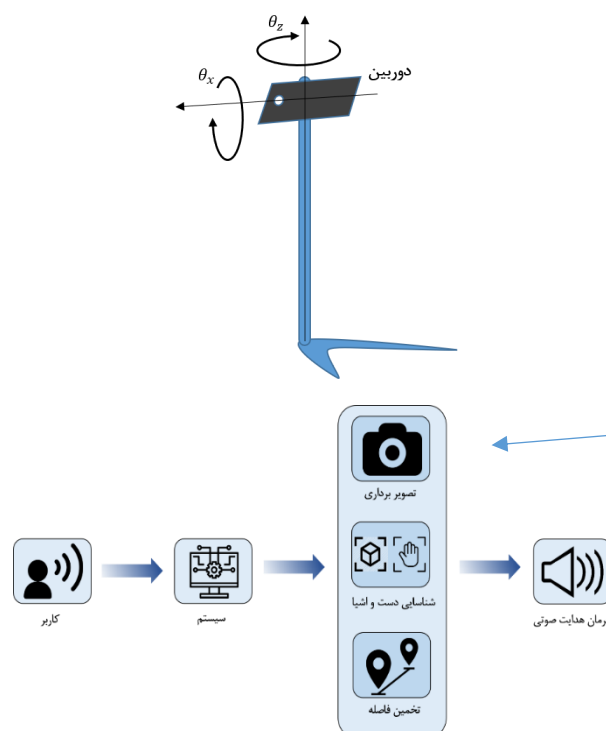
کوریاکوزه و همکاران [30] در یک مطالعه آزمایشگاهی، بهینه بودن دستورالعمل های صوتی و لرزشی (به دو شکل گام به گام کوتاه و شرح جامع مسیر) را برای ناوبری شرکت کنندگان کم بینا و نابینا ارزیابی کردند. عملکرد افراد با اندازه گیری خطای انحراف از مسیر و زمان تکمیل مسیر تحلیل شد که تفاوت معناداری بین کارایی دو روش دستورالعمل (کوتاه در مقابل مفصل) نشان نداد. با این حال، نظرسنجی ها بیانگر آن بود که ترجیح بین این روش ها پارامتری نسبی و وابسته به سلیقه فردی استاسدحسین و همکاران [31] نمونه اولیه ای از یک عصای هوشمند مبتنی بر میکروکنترلر ارائه کردند که با بهره گیری از حسگرهای تشخیص آب، اولتراسونیک، و مژول های جی پی اس، موانع را برای کاربران نابینا شناسایی می کند. طراحی تاشونده ای این سیستم، قابلیت حمل را بهبود بخشیده و یک سیستم ناوبری هوشمند برای فضاهای باز مبتنی بر آن پیاده سازی شده است که آزمایش های میدانی، اثربخشی آن در کمک به مسیریابی افراد دارای اختلال بینایی را تأیید می کند. علیرغم این مزایا، وزن نسبتاً بالای دستگاه، چالشی برای حمل آسان محسوب می شود. ماتاچیا و همکاران [32] عینکی طراحی کردند که در آن دوربین های استریوویژن تعبیه شده بود و قابلیت تشخیص اجسام و فاصله آن ها را داشت. این سیستم با استفاده از عملگرهای لرزشی و هشدارهای صوتی، فرد دارای اختلالات بینایی را در مواجهه با اجسام و محیط های مختلف هدایت و راهنمایی می کند. در پژوهش های اخیر تمرکز عموماً بر روی مساله جابجایی افراد معطوف بوده است در حالیکه این افراد با چالش های بسیاری رو

مشخص و در آن ارتفاع و زاویه نسبت به سطح میز در شرایط ثابتی قرار گیرد و سپس در همان شرایط کالیبره شده و در همان موقعیت باقی بماند. فرض شده است اجسام روی میز به صورت ثابت قرار گرفته‌اند و حرکتی به خودی خود ندارند اما در صورت جابجایی آن‌ها به واسطه دست افراد، به روز رسانی جایگاه آن‌ها در طی فرآیند به درستی انجام خواهد گرفت.

کالیبراسیون

فرآیند کالیبراسیون با هدف تخمین مشخصه‌های ذاتی دوربین و موقعیت قرارگیری آن نسبت به فضای کاری سیستم صورت می‌پذیرد. در رابطه (۱)، ماتریس T_c بیانگر تبدیل کلی از فضای سه‌بعدی جهان واقعی به تصویر دوبعدی ثبت‌شده توسط دوربین است. این تبدیل از ترکیب دو بخش اصلی تشکیل می‌شود: ماتریس ذاتی دوربین و ماتریس بیرونی. ماتریس ذاتی دوربین یک ماتریس 3×3 است که ویژگی‌های داخلی دوربین را در فرآیند تصویربرداری مدل‌سازی می‌کند. این ماتریس شامل پارامترهایی مانند فاصله کانونی، نقطه مرکزی تصویر و ضریب انحراف تصویر است. در رابطه (۱)، f_x و f_y طول‌های کانونی هستند، c_x و c_y مختصات نقطه مرکزی در صفحه تصویر هستند و s ضریب انحراف است. تعیین دقیق پارامترهای ماتریس ذاتی (T_I) دوربین نیاز به تکرارهای متعدد با تصاویر مختلف، جهت حل معادلات پیچیده بهینه‌سازی دارد. ماتریس بیرونی (T_E) در بینایی ماشین، ماتریسی است که موقعیت و جهت‌گیری دوربین را در فضای سه‌بعدی نسبت به یک سیستم مختصات مرجع توصیف می‌کند. این ماتریس شامل دو بخش اصلی است: یک ماتریس چرخش که جهت‌گیری دوربین را نشان می‌دهد و یک بردار انتقال که موقعیت دوربین را مشخص می‌کند. ماتریس بیرونی برای تبدیل مختصات نقاط سه‌بعدی از سیستم مختصات حقیقی به سیستم مختصات دوربین استفاده می‌شود و نقش مهمی در فرآیند کالیبراسیون دوربین ایفا می‌کند. از آنجایی که سیستم به گونه‌ای طراحی شده است که افراد می‌توانند تنها با استفاده از دوربین تلفن همراه خود امکان استفاده از آن را داشته باشند، انجام فرآیند کالیبراسیون دوربین‌های مختلف نسبت به مشخصه‌های هر دوربین و موقعیت قرارگیری دوربین نسبت به محیط (ارتفاع و زاویه قرارگیری دوربین نسبت به سطح میز)

استفاده قرار گرفته است. توجه داشته باشید در این پژوهش سعی شده است تا با کمترین میزان هزینه، سیستمی کاربردی پیشنهاد و ارائه گردد و به همین سبب سیستم مورد نظر از دوربین تلفن همراه افراد برای تصویر برداری استفاده می‌کند که این موضوع موجب دسترسی راحت‌تر و مقرون به صرفه بودن آن می‌گردد، علاوه بر این در بخش استفاده از داده‌های تصویری سعی شده است روشی بهینه جهت حصول کمترین میزان حجم پردازشی پیاده‌سازی شود. شکل (۱) نمای کلی از عملکرد سیستم را نشان می‌دهد. همان‌طور که مشاهده می‌شود، ابتدا سیستم از طریق گیرنده صوتی فرمان کاربر را دریافت می‌کند. سپس دستور تصویربرداری صادر شده و تصویر ثبت می‌شود. در مرحله بعد، تصویر برای شناسایی دست و اشیاء پردازش می‌گردد. پس از آن، با انجام محاسبات لازم، فاصله و جهت حرکت دست نسبت به شیء مورد نظر تخمین زده می‌شود و در نهایت، این اطلاعات به صورت فرمان صوتی هدایت‌کننده به کاربر ارائه می‌شود.



شکل ۱. نمای کلی از نحوه عملکرد سیستم

سیستم پیشنهادی جهت استفاده روی میز طراحی شده است و به توانخواهان کمک می‌کند پشت میزهای خود بتوانند از وجود اجسام مختلف آگاهی پیدا کنند و در صورت نیاز به آن‌ها دسترسی داشته باشند. ابتدا لازم است محل قرارگیری دوربین

ضروری است.

$$T_c = T_I T_E$$

$$T_I = \begin{bmatrix} 1 & 0 & x_0 \\ 0 & 1 & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_x & 0 & 0 \\ 0 & f_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{s}{f_x} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$T_E = [I|P] \times \begin{bmatrix} R & 0 \\ 0 & 1 \end{bmatrix}$$

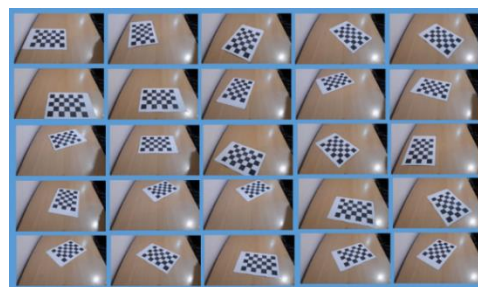
$$P = [x_c \ y_c \ z_c]^T$$

$$R = (R_x R_z)$$

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & C\theta_x & -S\theta_x \\ 0 & S\theta_x & C\theta_x \end{bmatrix}, R_z = \begin{bmatrix} C\theta_z & -S\theta_z & 0 \\ S\theta_z & C\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(۱)

در این پژوهش فرآیند کالیبراسیون با هدف محاسبه ماتریس تبدیل T_c که در برگرنده مشخصه‌های ذاتی دوربین و موقعیت قرار گیری آن نسب به سطح مورد بررسی است، با استفاده از صفحه شطرنجی و کتابخانه OpenCV صورت پذیرفته است. در طی این فرآیند لازم است تا صفحه شطرنجی در زوایا و موقعیت‌های مختلفی نسبت به دوربین قرار داده شود و از آن تصویر برداری شود. در ادامه این تصاویر برای تخمین پارامترهای ماتریس T_I و T_E مورد استفاده قرار می‌گیرند. در شکل (۲)، بخشی از این تصاویر نمایش داده شده‌اند.



شکل ۲ نمونه تصاویر ضبط شده جهت فرآیند کالیبراسیون

رابطه‌ی (۲) بیانگر فرآیند تبدیل مختصات یک نقطه از فضای واقعی (جهان) به صفحه‌ی تصویر دوربین است. در این معادله، مختصات متریک نقطه مورد نظر در فضای سه‌بعدی واقعی (سیستم مختصات جهانی) را نشان می‌دهد، و K_{ij} مختصات پیکسلی همان نقطه را در تصویر دوبعدی ثبت‌شده توسط دوربین نمایش می‌دهد. ماتریس T_c نیز ماتریس تبدیل دوربین است که ترکیبی از پارامترهای ذاتی و موقعیتی دوربین بوده و مسئول انتقال صحیح اطلاعات از فضای سه‌بعدی به تصویر دوبعدی می‌باشد.

این تبدیل نقش کلیدی در بازسازی هندسی صحنه و تحلیل موقعیت اشیاء در تصویر ایفا می‌کند. می‌توان با استفاده از رابطه (۲)، معلوم بودن ماتریس T_c و مختصات پیکسلی هر نقطه از تصویر به مختصات آن در فضای حقیقی دست یافت.

$$\begin{aligned} K_{ij} &= T_c K_{ij}^w \\ K_{ij} &= [P_x \ P_y \ 1]^T, \\ K_{ij}^w &= [X \ Y \ Z \ 1]^T \end{aligned} \quad (۲)$$

پارامترهای مربوط به دو ماتریس K_{ij} و K_{ij}^w در رابطه (۲) ارائه شده است. P_x و P_y به ترتیب معرف مختصات پیکسلی افقی و عمودی در تصویر بوده و X ، Y و Z بیانگر پارامترهای مختصاتی در فضای سه بعدی (فضای حقیقی) به صورت متریک می‌باشند.

شناسایی اشیاء و دست

یولو نسل هشتم [34,33] یک مدل یادگیری عمیق در حوزه تشخیص اشیاء است که به دلیل دقت بالا و سرعت مناسب در پردازش تصاویر، به‌ویژه برای کاربردهای بلادرنگ، در بسیاری از پروژه‌های بینایی ماشین و سیستم‌های هوشمند مورد استفاده قرار می‌گیرد. این مدل دارای طراحی بهینه‌شده‌تری نسبت به نسخه‌های پیشین خود است و از ساختار ماژولار و پیاده‌سازی ساده‌تری بهره می‌برد که آن را برای آموزش و استفاده در کاربردهای پژوهشی و صنعتی مناسب می‌سازد. یولو نسل هشتم در نسخه‌های مختلفی مانند s ، m ، x و l ارائه شده که هر کدام توانایی میان دقت و سرعت پردازش را فراهم می‌کنند. نسخه n به واسطه کمترین میزان پارامتر یادگیری آن، سریع‌ترین نسخه و البته کمترین میزان دقت را در مقایسه با سایر مدل‌ها دارد و نسخه x بیشترین زمان انجام عملیات و بیشترین دقت را به خود اختصاص داده است. از آنجایی که در این پژوهش، زمان از اهمیت بسیار زیادی برخوردار است و در طی تست‌های گرفته شده مشخص شد نسخه n دقت مناسبی در کاربرد مورد نظر دارد از این نسخه جهت پیاده‌سازی شناسایی اجسام استفاده شده است. سناریو مورد نظر به این صورت است که ابتدا اجسام شناسایی می‌شوند و در ادامه لازم است دست شناسایی شده و یک نقطه مرجع روی آن تعیین شود تا با دقت بیشتری موقعیت آن بررسی شود و سپس نقطه تعیین شده دنبال شود. دنبال کردن نقطه تعیین شده علاوه بر اینکه توان پردازشی کمتری نسبت به شناسایی مجدد آن صرف می‌کند، سرعت فرآیند را نیز بهبود می‌بخشد. مدل مدیاپایپ [35] یکی از ابزارهای پیشرفته در زمینه شناسایی نقاط مختلف بدن و تحلیل حرکات است که توسط

پیکسلی به مختصات حقیقی نقاط می‌توان دست یافت.

$$K_{ij} = [\{Y_i\}.x_b \quad \{Y_i\}.y_b \quad 1]^T, \\ K_{ij}^w = [\{Y_i\}.x_b^p \quad \{Y_i\}.y_b^p \quad \{Y_i\}.z^p \quad 1]^T \quad (3)$$

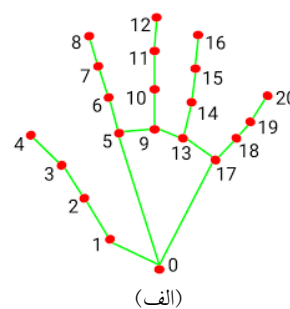
همانطور که پیش‌تر اشاره شد، K_{ij} معرف مختصات پیکسلی و K_{ij}^w معرف مختصات فضای حقیقی در رابطه (۳) می‌باشد. لازم به ذکر است $\{Y_i\}$ بیانگر شیء شناسایی شده یا نقاط کلیدی تشخیص داده شده بر روی دست انسان می‌باشد. $\{Y_i\}.x_b$ و $\{Y_i\}.y_b$ به ترتیب مختصات موقعیت نقطه‌ای موارد شناسایی شده در فضای پیکسلی تصویر می‌باشد. از سوی دیگر $\{Y_i\}.x_b^p$ ، $\{Y_i\}.y_b^p$ و $\{Y_i\}.z^p$ مختصات نقطه‌ای در فضای سه بعدی را به صورت متریک نشان می‌دهند. از آنجایی که اجسام روی سطح میز قرار داده شده‌اند و نقطه تماس آن‌ها با سطح میز مورد نظر است، پارامتر مربوط به ارتفاع عمودی ($\{Y_i\}.z^p$) برابر با صفر قرار داده شده است (مطابق رابطه ۴). این فرضیه به این دلیل در نظر گرفته می‌شود که تمامی نقاط تماس اجسام با سطح میز در یک صفحه افقی قرار دارند. در نتیجه، مطابق رابطه (۳) پارامترهای دیگر مختصات ($\{Y_i\}.x_b^p$ ، $\{Y_i\}.y_b^p$) از معادلات حاصل می‌شوند و موقعیت دقیق نقاط تماس را مشخص می‌کنند. این رویکرد به ساده‌سازی معادلات و کاهش پیچیدگی محاسبات کمک می‌کند، چرا که تنها نیاز به بررسی مختصات دوبعدی اجسام بر روی سطح میز داریم. در نتیجه این موضوع فضای کاری سیستم تنها فضای ۲ بعدی سطح میز می‌باشد.

$$\begin{bmatrix} \{Y_i\}.x_b \\ \{Y_i\}.y_b \\ 1 \end{bmatrix} = T_c \begin{bmatrix} \{Y_i\}.x_b^p \\ \{Y_i\}.y_b^p \\ 0 \\ 1 \end{bmatrix} \quad (4)$$

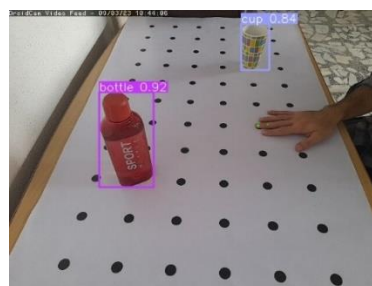
تخمین ارتفاع اجسام و محدوده برخورد

با توجه به شکل (۴)، پس از آنکه تصویر ضبط شده توسط ماتریس تبدیل سیستم در فضای حقیقی تصویر شده است، مشاهده می‌شود که سطح اشغال شده توسط جسم روی میز بسیار بیشتر از میزان واقعی به نظر می‌رسد. دلیل این موضوع، در نظر نگرفتن پارامترهای عمق و ارتفاع اجسام است. عدم توجه به این پارامترها باعث می‌شود تا ابعاد و موقعیت جسم در تصویر تحریف شده و به اشتباه بزرگ‌تر از اندازه واقعی خود نمایش داده شود، که می‌تواند منجر به تحلیل نادرست از فضای اشغال‌شده توسط جسم گردد.

گوگل توسعه یافته است. این مدل به ویژه در زمینه‌هایی مانند واقعیت افزوده و ردیابی حرکات بدن کاربرد دارد و توانایی شناسایی و دنبال کردن نقاط کلیدی بدن، از جمله دست‌ها، صورت، و بدن کامل را داراست. برای آموزش این مدل، مجموعه داده‌های بسیار وسیع و متنوعی استفاده شده است که شامل میلیون‌ها تصویر و ویدئو از افراد در موقعیت‌ها و حالات مختلف است. این داده‌ها به مدل کمک می‌کند تا با دقت بالا نقاط مختلف بدن را شناسایی و دنبال کند، حتی در شرایط نوری و زاویه‌های مختلف. مدیapایپ به دلیل دقت و سرعت بالا در شناسایی نقاط بدن، به یک ابزار قدرتمند در بسیاری از برنامه‌ها تبدیل شده است. همانطور که در شکل (۳-الف) نمایش داده شده است، مدیapایپ قادر به شناسایی ۲۰ نقطه کلیدی بر روی دست انسان می‌باشد. در این پژوهش از مدیapایپ جهت شناسایی دست استفاده شده و به عنوان مرجع انتهای انگشت میانی (نقطه شماره ۱۲) انتخاب شده است.



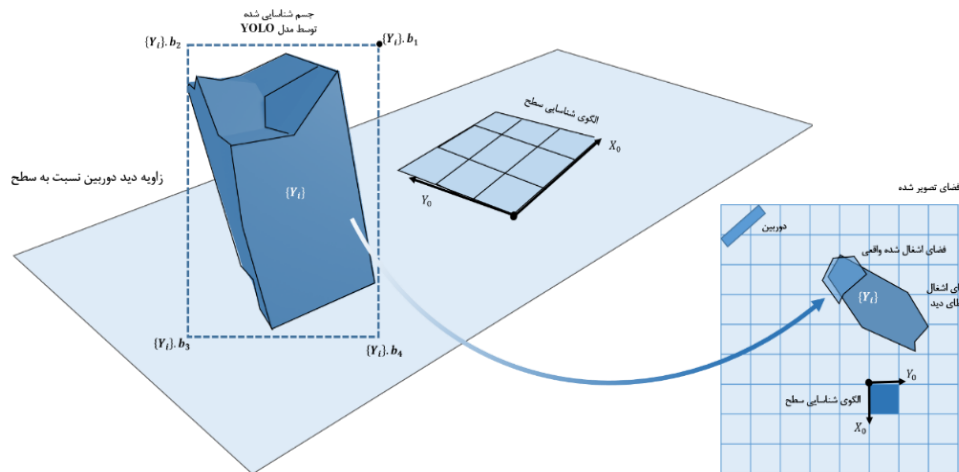
(الف)



(ب)

شکل ۳ الف) نقاط کلیدی قابل شناسایی توسط مدل مدیapایپ بر روی دست انسان. ب) نتیجه مدل‌های yolo و مدیapایپ در شناسایی اجسام و نقاط کلیدی دست انسان

در ادامه با داشتن مختصات مربوط به اجسام و دست به صورت پیکسلی، به محاسبات تعیین موقعیت آن‌ها در فضای حقیقی پرداخته شده است. با استفاده از ماتریس تبدیل سیستم که در بخش کالیبراسیون محاسبه گردید و رابطه (۲) از مختصات



شکل ۴. نمایشی از خطای سیستم در صورت تعیین نکردن محدوده سطح اشغالی اجسام

نقطه میانی ضلع بالا و پایین قابل محاسبه می‌باشد. در ادامه مختصات پیکسلی این دو نقطه در معادلات مربوط به محاسبه ارتفاع اشیاء مورد استفاده قرار گرفته می‌شود. در رابطه (۵)، با داشتن مختصات پیکسلی نقطه میانی ضلع پایین از جعبه مرزی $\{Y_i\}.u$ ، به محاسبه مختصات فضای حقیقی آن $\{Y_i\}.u[1]^p$ و $\{Y_i\}.u[1]^p$ پرداخته شده است.

$$\begin{aligned} \{Y_i\}.m_{2 \times 1} &= \text{mean}(\{Y_i\}.b_3, \{Y_i\}.b_4) \\ \{Y_i\}.u_{2 \times 1} &= \text{mean}(\{Y_i\}.b_1, \{Y_i\}.b_2) \end{aligned} \quad (5)$$

از آنجایی که جسم شناسایی شده بر روی سطح میز قرار دارد، تمام نقاط ضلع پایینی جعبه مرزی نیز روی میز فرض شده و به همین سبب مقادیر مربوط به پارامتر ارتفاع (مشخصه z) برای تمامی این نقاط صفر در نظر گرفته شده است.

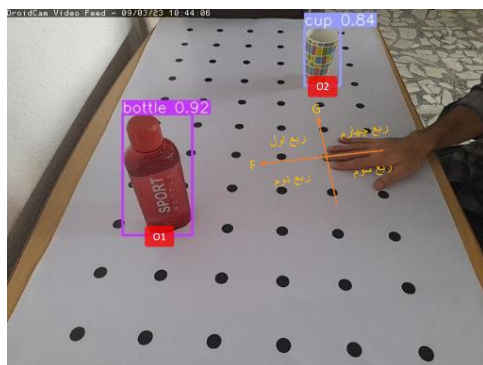
$$\begin{bmatrix} \{Y_i\}.u[1] \\ \{Y_i\}.u[2] \\ 1 \end{bmatrix} = T_c \begin{bmatrix} \{Y_i\}.u[1]^p \\ \{Y_i\}.u[2]^p \\ 0 \\ 1 \end{bmatrix} \quad (6)$$

جهت محاسبه ارتفاع جسم، لازم است مختصه ارتفاع در فضای حقیقی، در نقطه میانی ضلع بالایی جعبه مرزی تخمین زده شود. همانطور که در شکل (۵) قابل مشاهده است، دو نقطه میانی بالا و پایین جعبه مرزی در فضای حقیقی، دو مؤلفه مشترک و همسان دارند زیرا در راستای عمود بر سطح میز قرار گرفته‌اند.

در این پژوهش، به لیست اجسام قابل شناسایی توسط سیستم، پارامتری برای تفکیک اجسام مرتفع و غیرمرتفع اضافه شده است. این پارامتر به سیستم کمک می‌کند تا بین اجسامی که ارتفاع قابل توجهی دارند و آنهایی که ارتفاع کمتری دارند، تمایز قائل شود. لازم به ذکر است که در این پژوهش، برای تخمین سطح اشغالی روی میز توسط اجسام مرتفع، پارامتر عمق آن‌ها برابر با ضلع کوچکتر جعبه مرزی در نظر گرفته شده است. این فرض به این معنی است که عمق جسم مرتفع به اندازه ضلع کوچکتر جعبه مرزی که دور آن جسم رسم شده، در نظر گرفته می‌شود. از سویی دیگر، برای اجسام غیرمرتفع، فرض شده است که فضای اشغالی آن‌ها مطابق با خروجی ماتریس تبدیل است. این به این معنی است که اجسام غیرمرتفع بدون نیاز به در نظر گرفتن عمق اضافی، تنها با استفاده از ابعاد دو بعدی تصویر تحلیل می‌شوند. پس از شناسایی اجسام توسط مدل یولو، و دریافت مشخصات جعبه مرزی آن‌ها به عنوان خروجی، مرحله بعدی استفاده از این داده‌ها برای تخمین دقیق‌تر ابعاد اجسام است.

برای تخمین میزان ارتفاع اجسام، روشی ارائه شده است که بر اساس مختصات پیکسلی نقاط میانی ضلع بالایی و پایینی جعبه مرزی عمل می‌کند. این روش با استفاده از این مختصات، قادر است ارتفاع واقعی اجسام را با دقت مناسبی تخمین بزند. این فرآیند شامل تحلیل داده‌های پیکسلی و تبدیل آن‌ها به فضای سه‌بعدی است که به نوبه خود به بهبود دقت تحلیل فضای اشغالی و ارائه اطلاعات مفیدتر در مورد ابعاد اجسام کمک می‌کند. با استفاده از روابط (۵) مختصات دو نقطه مورد نظر از جعبه مرزی،

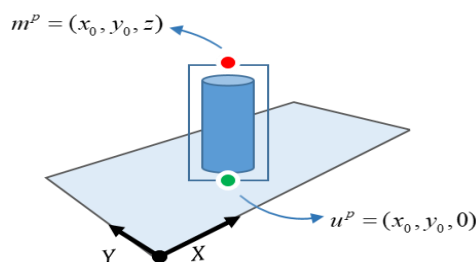
نقطه شماره ۱۱ (شکل ۳) نیز به دلیل مطرح شده مورد شناسایی قرار گرفته است. اولین گام تشخیص موقعیت هر جسم نسبت به دست انسان و تقسیم بندی آن به چهار بخش که در تصویر (۶) نمایش داده شده است میباشد. برای مثال در شکل (۶) موقعیت جسم شماره ۲ (O2) نسبت به دست در ربع چهارم و بطری در ربع دوم قرار گرفته است. محور F راستای انگشت میانی (محور گذرا از دو نقطه با شماره‌های ۱۱ و ۱۲ بر روی دست) است که با استفاده از دو نقطه تشخیص داده شده توسط مدل مدیاپایپ بر روی این انگشت ترسیم شده است. محور G، عمود بر محور F در نقطه انتهایی انگشت میانی می‌باشد، و به این ترتیب ۴ ناحیه توسط این دو محور ایجاد شده است.



شکل ۶ تقسیم بندی چهارگانه نواحی اطراف دست

معادلات محاسبه محورهای F و G طبق روابط شماره (۸) قابل محاسبه می‌باشد. P_{h11}^w و P_{h12}^w به ترتیب بیانگر مختصات نقاط کلیدی دست با شماره‌های ۱۱ و ۱۲ (مطابق شکل ۳-الف) می‌باشند که در صفحه میز بوده و در واحد متریک بیان شده‌اند و M_F شیب خط گذرا از دو نقطه ۱۱ و ۱۲ (شیب محور F) می‌باشد. از آنجایی که سیستم مختصات را تنها بر روی صفحه مرجع (سطح میز و به صورت ۲ بعدی) در نظر می‌گیرد، پارامتر مرتبط به ارتفاع آن‌ها صفر در نظر گرفته شده است.

$$\begin{aligned} P_{h11}^w &= (P_{h11}^w[1], P_{h11}^w[2]) \\ P_{h12}^w &= (P_{h12}^w[1], P_{h12}^w[2]) \\ M_F &= \frac{P_{h11}^w[2] - P_{h12}^w[2]}{P_{h11}^w[1] - P_{h12}^w[1]} \\ F(x) &= M_F \times (x - P_{h12}^w[1]) + P_{h12}^w[2] \\ G(x) &= \frac{-1}{M_F} \times (x - P_{h12}^w[1]) + P_{h12}^w[2] \end{aligned} \quad (8)$$



شکل ۵ نمایشی از دو نقطه میانی بالا و پایین جعبه مرزی و مختصات فضای حقیقی آن‌ها

از آنجایی که مؤلفه‌های فضای حقیقی نقطه پایینی با استفاده از رابطه (۶) قابل محاسبه است، می‌توان از این مقادیر استفاده کرد و رابطه (۷) را با استفاده از جایگذاری مختصات پیکسلی نقطه بالایی و دو مؤلفه مختصات فضای حقیقی نقطه پایینی نوشت. در این صورت تنها مؤلفه مجهول، پارامتر مربوط به محور عمود بر سطح $\{Y_i\}, m_z^p$ می‌باشد که متعلق به نقطه بالایی از جعبه مرزی بوده و می‌توان آن را معادل ارتفاع اشیاء در نظر گرفت.

$$\begin{bmatrix} \{Y_i\}, m[1] \\ \{Y_i\}, m[2] \\ 1 \end{bmatrix} = T_C \begin{bmatrix} \{Y_i\}, u[1]^p \\ \{Y_i\}, u[2]^p \\ \{Y_i\}, m_z^p \\ 1 \end{bmatrix} \quad (7)$$

به منظور تخمین پارامتر $\{Y_i\}, m_z^p$ بازه‌ای از اعداد محتمل در طی برنامه تولید شده و بجای آن جایگذاری می‌شود، سپس سمت راست معادله (۷) محاسبه می‌شود. نتیجه این عملیات، ایجاد یک مختصات پیکسلی برای هر مقداری است که جایگذاری شده است. از میان مختصات پیکسلی ایجاد شده نزدیک‌ترین نتیجه به نقطه $\{Y_i\}, m$ گزینش شده و مقدار جایگذاری شده برای پارامتر ارتفاع آن به عنوان ارتفاع تخمین زده شده گزارش می‌شود.

هدایت دست

در مراحل پیشین، شناسایی دست و اشیاء با استفاده از مدل‌های یولو و مدیاپایپ صورت گرفته است. و همانطور که گفته شد یک نقطه مشخص و مرجع جهت تعیین مختصات در فضای حقیقی مورد نیاز است. نقطه انتهایی انگشت میانی (نقطه شماره ۱۲ در شکل ۳) به عنوان نقطه مرجع روی دست مد نظر قرار گرفته است. با توجه به این موضوع که جهت قرار گیری دست نسبت به اجسام نیز حائز اهمیت می‌باشد، نیاز است نقطه‌ای دیگر نیز به منظور تشخیص جهت قرار گیری دست شناسایی شود.

پس از آنکه موقعیت و زاویه قرارگیری جسم نسبت به دست انسان طی مراحل قبل محاسبه شد، به منظور هدایت دست به سمت جسم مورد نظر از روش جهت عقربه‌های ساعت استفاده شده است. همانطور که در شکل (۷) مشاهده می‌شود راستای انگشت میانی، جهت ساعت ۱۲ را نشان می‌دهد و سایر موقعیت‌ها متناسب با زاویه آن‌ها و اینکه در کدام ربع نسبت به دست قرار گرفته اند در جهت سایر اعداد ساعت تقسیم بندی می‌شوند.

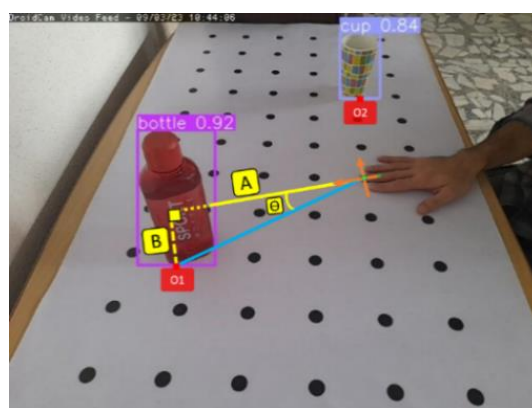
سیستم به گونه‌ای طراحی شده است که قادر است دستورات را به صورت صوتی دریافت کرده و متناسب با درخواست کاربر، پیام و گزارش مناسب را نیز به صورت صوتی ارائه دهد. با استفاده از کتابخانه‌ی تبدیل متن به صوت در پایتون، تعدادی از واژگان معنادار برای سیستم تعریف شده‌اند که شامل لیست نام اجسام قابل شناسایی توسط یولو و واژه‌ی "ریپورت" می‌شود. زمانی که سیستم دستور "ریپورت" را دریافت می‌کند، ابتدا تصویری توسط دوربین ثبت می‌شود و با استفاده از مدل یولو، اجسام موجود در تصویر شناسایی می‌شوند. سپس نام این اجسام در لیستی ذخیره می‌شود و جایگاه دست کاربر نیز با استفاده از مدل مدیایپایپ تعیین می‌گردد. در نهایت، برنامه متنی را آماده می‌کند که شامل توضیحاتی درباره‌ی اجسام شناسایی شده و موجود در تصویر است. این متن به کتابخانه تبدیل متن به صوت ارجاع داده شده و سیستم آن را به یک فایل صوتی تبدیل می‌کند و سپس پخش می‌نماید، به این ترتیب کاربر اطلاعات لازم را در ارتباط با اجسام قابل دسترسی و موجود دریافت می‌کند. پس از ارائه‌ی گزارش، برنامه در حالت انتظار باقی می‌ماند تا دستور جدیدی دریافت کند. این دستور می‌تواند فراخواندن نام یکی از اجسام شناسایی شده باشد. در این حالت، سیستم با دریافت نام جسم مورد نظر، کاربر را به صورت صوتی هدایت می‌کند و مشخص می‌شود که دست خود را به چه سمتی (بر اساس عقربه‌های ساعت) و به چه طولی (در واحد سانتی‌متر) جابجا کند تا بتواند به جسم مورد نظر دسترسی پیدا کند. شکل (۸)، نحوه تعامل صوتی سیستم با کاربر را نمایش می‌دهد. همچنین جدول ۲ زوایا و یک چهارم‌های تعامل را بررسی می‌کند.

در ادامه با داشتن معادلات دو محور F و G می‌توان موقعیت هر نقطه بر روی میز را نسبت به دست با استفاده از جدول (۱) در چهار ربع شناسایی کرد.

جدول ۱ شناسایی موقعیت قرارگیری هر نقطه نسبت به دستگاه مختصات تشکیل داده شده روی دست

ربع اول	$o^w[2] > F(o^w[1]), o^w[2] > G(o^w[1])$
ربع دوم	$o^w[2] < F(o^w[1]), o^w[2] > G(o^w[1])$
ربع سوم	$o^w[2] < F(o^w[1]), o^w[2] < G(o^w[1])$
ربع چهارم	$o^w[2] > F(o^w[1]), o^w[2] < G(o^w[1])$

گام بعدی تشخیص زاویه نقاط کلیدی اجسام نسبت به دست انسان می‌باشد. مطابق شکل (۷) زوایای نقاط کلیدی (نامگذاری شده با حرف O) نسبت به محورهای F و G محاسبه شده و مقدار θ برگردانده می‌شود.



(الف)

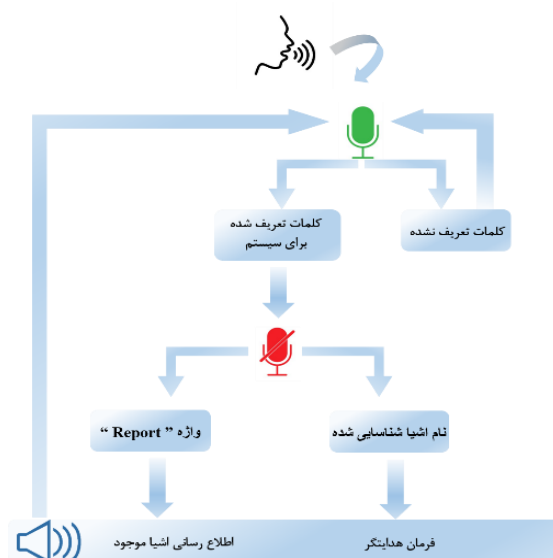


(ب)

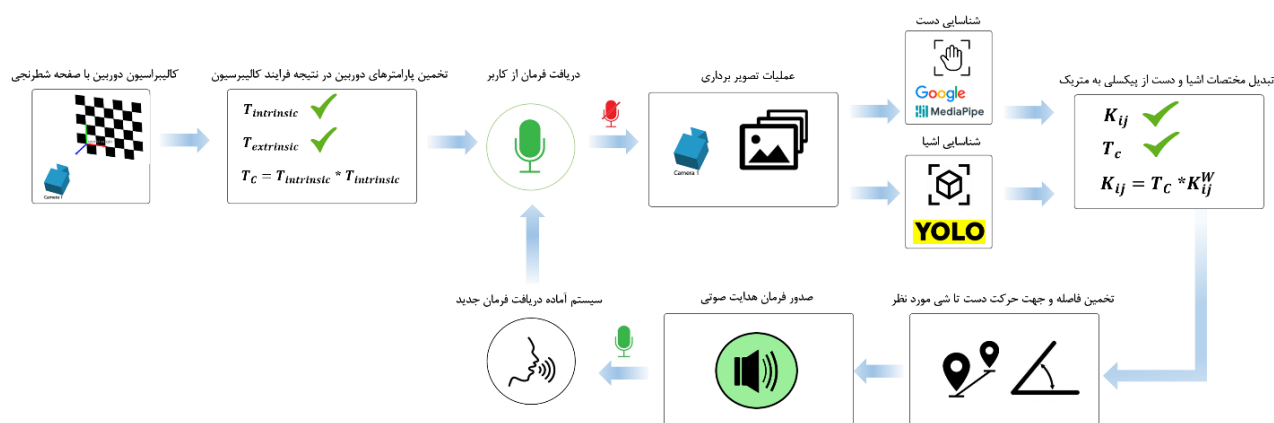
شکل ۷ (الف) نمایی از روش تخمین زاویه جسم نسبت به دست (ب) همپوشانی راستای دست و عقربه‌های ساعت

جدول ۲ تقسیم بندی جهت های دست با استفاده از جهت عقربه های ساعت

ناحیه قرار گیری جسم نسبت به دست	$0 < \theta < 15$	$15 < \theta < 45$	$45 < \theta < 75$	$75 < \theta < 90$
ربع اول	۱۲ ساعت	۱ ساعت	۲ ساعت	۳ ساعت
ربع دوم	۱۲ ساعت	۱۱ ساعت	۱۰ ساعت	۹ ساعت
ربع سوم	۶ ساعت	۷ ساعت	۸ ساعت	۹ ساعت
ربع چهارم	۶ ساعت	۵ ساعت	۴ ساعت	۳ ساعت



شکل ۸ نحوه عملکرد سیستم تعامل صوتی تعبیه شده در سیستم



شکل ۹ نمایی از نحوه عملکرد سیستم از مراحل آغازین تا انجام عملیات هدایت کاربر

(برای مشاهده دمو ویدیویی عملکرد سیستم به لینک مقابل مراجعه کنید <https://barsalani.com/AssistiveSystem.mp4>)

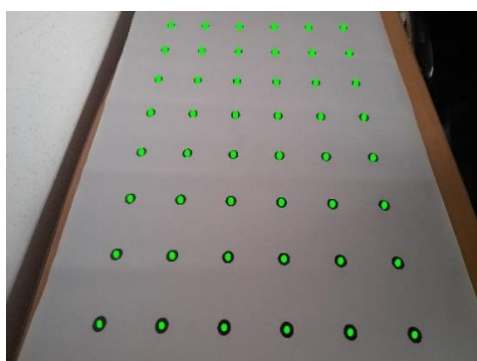
فرمان کاربر را دریافت می کند: این فرمان یا درخواست شناسایی اشیاء موجود است یا نام شیء خاصی را هدف قرار می دهد. در حالت اول، سیستم با تصویربرداری و شناسایی اشیاء، نتیجه را به صورت صوتی گزارش کرده و به حالت انتظار بازمی گردد. در

شکل (۹) نمای کلی جریان مرحله به مرحله سیستم را نشان می دهد که در آن هر گام پیش شرط گام بعدی است. عملکرد سیستم با کالیبراسیون اولیه آغاز می شود (که در صورت ثابت ماندن ارتفاع و زاویه دوربین، تکرار نمی شود). سپس سیستم

گزارش شده نیز افزایش یافته است. این موضوع به سبب عواملی همچون تراکم پیکسلی کمتر که معرف فضای واقعی گسترده‌تر نسبت به محدوده‌های نزدیک‌تر به دوربین بوده اتفاق افتاده است. از میان سایر عوامل مؤثر نیز می‌توان به نوع و مشخصه‌های ذاتی دوربین همچون تراکم پیکسلی آن اشاره کرد.

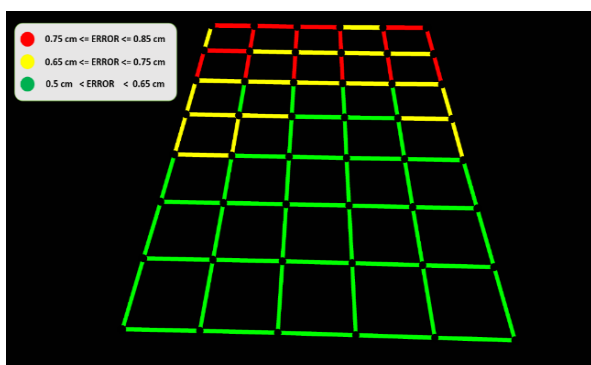


(الف)



(ب)

شکل ۱۰ (الف) بئر طراحی شده جهت آزمایش شناسایی فواصل (ب) شناسایی دایره‌های سیاه رنگ توسط برنامه



شکل ۱۱ نمایش خطای تخمین فواصل توسط سیستم

در طی پژوهشی [36] خطای عملکرد روش استریوویژن (با استفاده از دو دوربین ۲ مگاپیکسلی لاجیتک) مورد بررسی قرار

حالت دوم، سیستم پس از تصویربرداری، علاوه بر اشیاء، دست کاربر و موقعیت‌ها را تشخیص می‌دهد، سپس مختصات را پردازش نموده، فاصله و جهت حرکت دست را محاسبه می‌کند و نتیجه نهایی را به صورت فرمان صوتی راهنما اعلام می‌نماید. پس از اجرای هر فرمان، سیستم به حالت دریافت دستور بازگشته و چرخه به طور پیوسته قابل تکرار است.

یافته‌ها

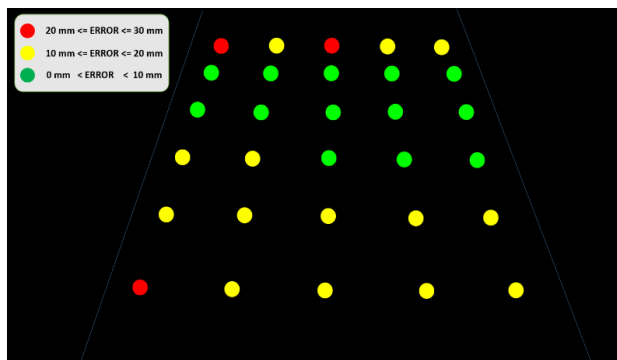
در این بخش، سیستم از جهات مختلفی مورد ارزیابی و آزمایش قرار گرفته است. این ارزیابی شامل دقت تخمین فواصل، اندازه‌گیری ارتفاع اجسام و هزینه زمان مصرفی سیستم بوده و نتایج حاصل از این آزمایش‌ها به تفصیل گزارش شده است. نتایج به دست آمده، میزان کارایی و دقت سیستم را در انجام این وظایف مشخص می‌کند.

خطای تخمین فاصله

به منظور بررسی و ارزیابی خطای سیستم در تشخیص فواصل مختلف در محدوده کاری آن، بئر مشابه با شکل (۱۰) طراحی و آماده شده است. این بئر شامل دایره‌های مشکی رنگی است که مراکز آن‌ها با فاصله ۱۰ سانتی‌متر از یکدیگر قرار گرفته‌اند. در طول این آزمایش، سیستم به تخمین فاصله بین دایره‌ها پرداخته و دقت آن مورد ارزیابی قرار گرفته است. پس از تخمین فاصله‌ها، میانگین خطا و خطای اندازه‌گیری در محدوده‌های مختلف به دقت محاسبه و گزارش شده است تا خطای سیستم در اندازه‌گیری فواصل مختلف به طور جامع تحلیل گردد.

در این آزمایش، ابتدا دایره‌های سیاه رنگ موجود بر روی بئر توسط برنامه شناسایی شده‌اند و سپس، موقعیت هر یک از این نقاط در فضای واقعی با استفاده از رابطه (۲) استخراج شد. این فرآیند شامل تحلیل دقیق داده‌های شناسایی شده و تبدیل آن‌ها به مختصات فضای واقعی است تا دقت سیستم در تعیین موقعیت نقاط را ارزیابی کند. با داشتن موقعیت واقعی هر کدام از دایره‌ها می‌توان فواصل مورد نظر را محاسبه و گزارش کرد. طی بررسی نتایج حاصل شده، میانگین خطای تخمین فواصل ۶/۱ میلی‌متر و بیشترین و کمترین میزان خطای گزارش شده به ترتیب ۸/۵ و ۵/۲ میلی‌متر بوده است. همانطور که در شکل (۱۱) مشاهده می‌شود، هر چه فاصله از دوربین بیشتر شده است، خطای تخمین

بسیار نزدیک بودن جسم نسبت به دوربین نیز موجب افزایش خطای تخمین شده، زیرا دوربین نسبت به بعد ارتفاع جسم دید محدودتری داشته و به همین سبب با احتمال خطای بیشتری فرآیند تخمین ارتفاع صورت می‌پذیرد.



شکل ۱۳: نمایی از خطای سیستم طی آزمایش تخمین ارتفاع

هزینه زمانی سیستم

لازم به ذکر است که کلیه پردازش‌ها بر روی سیستمی با پردازنده‌ی Intel Core i5-4300M انجام شده است. میانگین زمان اختصاص‌یافته به هر تصویر برای شناسایی نقاط کلیدی دست توسط مدیاپایپ، ۰/۴ ثانیه، برای تشخیص اشیاء با یولو نسل هشتم سری نانو، ۱/۱ ثانیه و برای صدور فرمان‌های هدایت صوتی پس از محاسبات ناوبری (تخمین فاصله و جهت حرکت)، تنها ۰/۰۳ ثانیه ثبت گردید. با توجه به ناچیز بودن زمان فرآیند هدایت صوتی نسبت به دو مرحله پیشین، سهم آن در زمان کل پردازش ناچیز محسوب می‌شود.

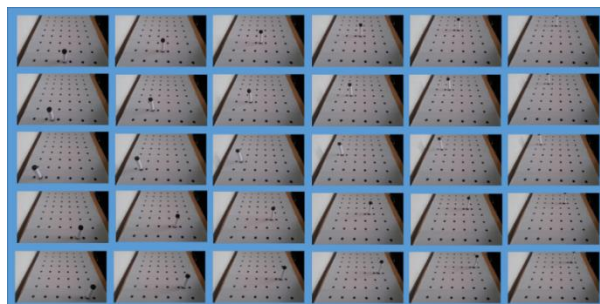
محدودیت‌ها

این سیستم دارای محدودیت‌هایی در زمینه کالیبراسیون است که نیازمند توجه کاربران می‌باشد. برای عملکرد دقیق، لازم است دستگاه در ارتفاع و زاویه مشخصی نسبت به سطح میز (به عنوان موقعیت مرجع) کالیبره شود. در صورتی که ارتفاع دستگاه یا زاویه آن نسبت به سطح میز تغییر کند، انجام فرآیند کالیبراسیون مجدد ضروری خواهد بود. این موضوع به دلیل وابستگی الگوریتم‌های پردازشی به پارامترهای موقعیتی اولیه است که تغییر آن‌ها می‌تواند بر دقت خروجی تأثیر گذار باشد. با این حال، سیستم در برابر جابه‌جایی‌های افقی میز یا چرخش در صفحه افقی (بدون تغییر ارتفاع یا زاویه) مقاوم است و در این

گرفت، میانگین خطای تخمین فاصله و موقعیت نقاط در فضای کاری مشابه سیستم پیشنهادی، معادل ۱/۲۲ درصد فاصله دوربین تا نقاط اندازه‌گیری گردید. نتایج این مطالعه نشان‌دهنده آن است که سیستم پیشنهادی حاضر با دستیابی به دقتی در همین حدود، عملکردی قابل‌قیاس با روش استریوویژن دارد و از قابلیت اطمینان مطلوبی برخوردار است.

خطای تخمین ارتفاع

یکی دیگر از قابلیت‌های سیستم، توانایی تخمین ارتفاع اجسام را بوده و در این بخش آزمایشی جهت بررسی دقت و صحت عملکرد آن در این خصوص طراحی شده است که در ادامه به آن پرداخته می‌شود. جهت بررسی دقت سیستم در تخمین ارتفاع جسمی با ارتفاع ۱۵ سانتی‌متر در نظر گرفته شده و در موقعیت‌های مختلف قرار داده شده است. مطابق روش ارائه شده ارتفاع تخمین زده شده و میزان دقت مقادیر گزارش شده است و در ادامه مورد ارزیابی قرار گرفته است. مطابق شکل (۱۲)، جسم هدف در ۳۰ موقعیت مختلف جایگذاری شده و در هر نقطه ارتفاع آن تخمین زده شده است.



شکل ۱۲: تصاویر آزمایش تخمین ارتفاع

همانطور که در تصویر (۱۳) مشخص شده است، در نزدیک‌ترین و دورترین نواحی سیستم در تشخیص ارتفاع جسم دقت کمتری در مقایسه با ناحیه میانی داشته است. طبق نتایج حاصل شده، میانگین خطای کلی ۱۱ میلی‌متر گزارش شده و همچنین بیشترین و کمترین میزان خطا به ترتیب ۲۱ و ۴ میلی‌متر بوده است. از جمله عوامل ایجاد خطا در این بخش می‌توان به زاویه قرار گیری اجسام نسبت به دوربین اشاره نمود. هر چه جسم از دوربین دورتر باشد به دلیل تراکم پیکسلی کمتر و فضای واقعی گسترده‌تر، دقت تخمین کاهش یافته است. از سویی دیگر

تکنولوژی‌های به‌روزی همچون یولو و مدیایپ در پیاده‌سازی این سیستم به کار گرفته شده‌اند که عملکرد بهینه‌ای را برای سیستم فراهم می‌کند. این پژوهش با تلفیق یولو و فریمورک مدیایپ در تشخیص اشیا و دست، و همچنین محدودسازی فضای کاری سیستم به سطح میز در قالب یک فضای دوبعدی، موفق شد حجم پردازشی سیستم را کاهش دهد زیرا با بهره‌گیری از روش پیشنهادی سیستم نیازی به پردازش سنگین محاسبات همپوشانی برای تخمین فواصل با استفاده از دو دوربین و یا سایر سنسورها را نداشته و در عین حال دقت قابل قبولی برای کاربرد مورد نظر فراهم کند. برای گسترش و پیشبرد تحقیق ارائه‌شده، مسیرهای گوناگونی قابل پیگیری است که می‌تواند به بهبود کارایی و دقت سیستم کمک کند. از جمله پیشنهادات پژوهشی می‌توان به مواردی همچون بهبود تعامل صوتی با بهره‌گیری از مدل‌های زبانی، تخصیص واحد پردازش به سرور و پیاده‌سازی رابط لغزشی برای تعامل مؤثر تر اشاره کرد.

واژه نامه

پیشنهادی نویسنده	واژه لاتین	پیشنهادی ویراستار
حسگر مادون قرمز غیرفعال	Passive infrared sensor	
یولو	YOLO	
دوربین زمان پرواز	Time of Flight Camera	
استریوویژن	Stereovision	بینایی دوچشمی؛ بینایی استریو
دوربین رنگی - عمقی	RGBD camera	
مدیایپ	MediaPipe	

شرایط نیازی به فرآیند کالیبراسیون مجدد ندارد. علاوه بر این، با توجه به استفاده سیستم از سنسور دوربین، در محیط‌های با نور کم ممکن است دقت عملکرد نسبت به شرایط نوری مطلوب کاهش یابد. این محدودیت در سیستم‌های مبتنی بر پردازش تصویر امری شناخته‌شده است که در نسخه‌های آتی می‌توان با بهینه‌سازی الگوریتم‌ها یا افزودن نور کمکی، آن را بهبود بخشید. با توجه به اینکه سیستم پیشنهادی در محدوده‌ی مشخصی از فضای کاری که هم‌سطح میز و با ارتفاع و زاویه‌ی ثابتی نسبت به آن تنظیم شده عمل می‌کند، در صورتی که دست یا اجسام در ارتفاعی بالاتر یا پایین‌تر از سطح میز قرار گیرند، دقت سیستم در تخمین موقعیت کاهش یافته و احتمال بروز خطا در موقعیت‌یابی افزایش می‌یابد.

نتیجه‌گیری

میانگین خطای سیستم پیشنهادی در تخمین فاصله و ارتفاع اشیا به ترتیب ۶/۱ و ۱۱ میلی‌متر اندازه‌گیری شد که نشان‌دهنده دقتی مشابه روش استریوویژن در تعیین موقعیت و فواصل است. با این حال، سیستم پیشنهادی با وجود دستیابی به این سطح دقت، دو مزیت بنیادین نسبت به استریوویژن دارد: نخست آنکه حجم محاسبات مورد نیاز به‌طور چشمگیری کاهش یافته است، چراکه روش استریوویژن ذاتاً مستلزم پردازش‌های حجیم محاسباتی است. دوم آنکه در معماری سخت‌افزاری، نیاز به دو دوربین در روش استریوویژن به استفاده از تنها یک دوربین در سیستم پیشنهادی تقلیل یافته است. بنابراین پژوهش حاضر با به‌کارگیری حداقل منابع پردازشی و سخت‌افزاری، همان دقت عملکردی را محقق ساخته است. از سوی دیگر، طی پژوهش‌های صورت گرفته، سیستم‌های کمکی طراحی شده برای افرادی که با اختلالات بینایی مواجه هستند، عموماً در زمینه جابجایی آن‌ها از محلی به محل دیگر متمرکز شده‌اند و توجهی به کمک به نابینایان در انجام کارهای پشت میز را ندارند. در این پژوهش تلاش شده است تا گامی در جهت توسعه سیستم‌های کمکی تصویری اختصاصی ارزان قیمت در زمینه هدایت نابینایان در خصوص فعالیت‌های پشت میز برداشته شود. در سیستم ارائه شده، تخمین فواصل تنها با استفاده از یک دوربین، که می‌تواند دوربین تلفن همراه افراد باشد، انجام می‌شود. این امر باعث می‌شود که سیستم مقرون به‌صرفه‌تر و در دسترس‌تر باشد.

مراجع

- [1] World Helath Organization, "Rehabilitation," [Online]. [Accessed: Apr. 22, 2024]. Available: <https://www.who.int/news-room/fact-sheets/detail/rehabilitation>
- [2] J. D. Edwards, E. G. Valdés, C. Peronto, M. Castora-Binkley, J. Alwerdt, R. Andel, J. J. Lister, "The efficacy of InSight cognitive training to improve useful field of view performance: A brief report," *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, vol. 70, no. 3, pp. 417–422, 2015. <https://doi.org/10.1093/geronb/gbt113>
- [3] A. Colenbrander, "My Perspective on Vision and Vision Rehabilitation," *Ophthalmology Science*, p. 100532, 2024. <https://doi.org/10.1016/j.xops.2024.100532>
- [4] M. E. Wilkinson, K. S. Shahid, "Low vision rehabilitation: An update," *Saudi Journal of Ophthalmology*, vol. 32, no. 2, pp. 134–138, 2018. <https://doi.org/10.1016/j.sjopt.2017.10.005>
- [5] I. Kardan, A. Akbarzadeh, "A New Hybrid Strategy for Forward Kinematic Analysis of Parallel Robots (Research Note)," *The Journal of Applied and Computational Sciences in Mechanics*, pp. 163–172, 2016. <http://dx.doi.org/10.1080/01691864.2014.994034>
- [6] M. Nizamuddin, A. Kamisetty, J. C. S Gummadi, and R. R Talla, "Integrating Neural Networks with Robotics: Towards Smarter Autonomous Systems and Human-Robot Interaction," *Robotics Xplore: USA Tech Digest* 1, no. 1, pp. 157-169, 2024.
- [7] A. Sadallah, J. Riazat, "Parameter Optimization of Various Solar Cell Models by Neural Network Algorithm," *Journal of Applied & Computational Sciences in Mechanics*, vol. 36, no. 1, pp. 123-142, 2024. <https://doi.org/10.22067/jacsm.2023.81702.1176>
- [8] V. Modanloo, A. Mashayekhi, and B. Akhoundi, "Comparison of regression, bee colony, and artificial neural network for predicting die filling in stamping of bipolar plates fuel cell," *Journal of Applied and Computational Sciences in Mechanics*, vol. 36, no. 2, pp. 47–58, 2024. <https://doi.org/10.22067/jacsm.2023.84567.1205>
- [9] S. A. Bagherzadeh, S. Emami Koopaei, and A. Salehi, "Optimal design of divergent part of an over-expanded nozzle profile using artificial intelligence," *Journal of Applied and Computational Sciences in Mechanics*, vol. 35, no. 1, pp. 19–36, 2023. <https://doi.org/10.22067/jacsm.2022.78011.1134>
- [10] M. Taghi Zadeh, M. Mokhtari, and M. Mazare, "Fixed-time fault tolerant control based on neural network and adaptive sliding mode controller of autonomous underwater vehicle with actuator saturation," *Journal of Applied and Computational Sciences in Mechanics*, vol. 36, no. 4, pp. 35–56, 2024. <https://doi.org/10.22067/jacsm.2024.85527.1219>
- [11] A. Liu, H. Zhao, T. Song, Z. Liu, H. Wang, and D. Sun, "Adaptive control of manipulator based on neural network," *Neural Computing and Applications*, vol 33, no. 9, pp. 4077-4085, 2021. <https://doi.org/10.1007/s00521-020-05515-0>
- [12] S. Targ, D. Almeida, K. Lyman, "Resnet in resnet: Generalizing residual architectures," *arXiv preprint arXiv:1603.08029*, 2016. <https://doi.org/10.48550/arXiv.1603.08029>

- [13] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014. <https://doi.org/10.48550/arXiv.1409.1556>
- [14] P. Jiang, D. Ergu, F. Liu, Y. Cai, and Bo Ma, "A Review of Yolo algorithm developments," *Procedia Computer Science*, vol. 199, pp. 1066-1073, 2022. <https://doi.org/10.1016/j.procs.2022.01.135>
- [15] T. L. Narayani, M. Sivapalanirajan, B. Keerthika, M. Ananthi, and M. Arunarani, "Design of smart cane with integrated camera module for visually impaired people," in *Proceeding 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, 2021, pp. 999–1004. <https://doi.org/10.1109/ICAIS50930.2021.9395840>
- [16] M. Marcon, A. Ortoncelli, "Object Recognition to Support Navigation Systems for Blind in Uncontrolled Environments," *Anais do Computer on the Beach*, vol. 13, pp. 274–281, 2022. <https://doi.org/10.14210/cotb.v13.p274-281>
- [17] J. Kilian, A. Neugebauer, L. Scherffig, S. Wahl, "The unfolding space glove: A wearable spatio-visual to haptic sensory substitution device for blind people," *Sensors*, vol. 22, no. 5, p. 1859, 2022. <https://doi.org/10.3390/s22051859>
- [18] M. Aharchi, M. H. A. Kbir, "Localization and Navigation System for Blind Persons Using Stereo Vision and a GIS," in *WITS 2020: Proceedings of the 6th International Conference on Wireless Technologies, Embedded, and Intelligent Systems*, Springer, 2022, pp. 1–12. https://doi.org/10.1007/978-981-33-6893-4_35
- [19] T. Adep, R. Nikam, S. Wanewe, and D. K. B. Naik, "Visual assistant for blind people using Raspberry Pi," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 3307, pp. 671–675, 2021. <http://dx.doi.org/10.32628/CSEIT2173142>
- [20] P. Zantou, A. M. Mousse, B. Atohou, "An Intelligent based System for Blind People Monitoring in a Smart Home." *Computer Science & Information Technology (CS & IT)*, 2020. <https://doi.org/10.5121/csit.2020.101910>
- [21] K. T. Banu, K. Supriya, K. Sony, M. Chandana, M. Bhavana, M. B. F. Ali, "Outdoor & Indoor Vision Based Localization for Blind Pedestrian Navigation Assistance," *International Journal of Engineering Applied Sciences and Technology*, vol. 4, no. 12, pp. 715-720, 2020. <http://dx.doi.org/10.33564/IJEAST.2020.v04i12.127>
- [22] O. Younis, W. Al-Nuaimy, F. Rowe, M. H. Alomari, "A smart context-aware hazard attention system to help people with peripheral vision loss," *Sensors*, vol. 19, no. 7, p. 1630, 2019. <https://doi.org/10.3390/s19071630>
- [23] H. P. Buimer, M. Bittner, T. Kosteljik, T. M. Van Der Geest, A. Nemri, R. J. Van Wezel, Y. Zhao, "Conveying facial expressions to blind and visually impaired persons through a wearable vibrotactile device," *PloS One*, vol. 13, no. 3, p. 0194737, 2018. <https://doi.org/10.1371/journal.pone.0194737>
- [24] Y. Bouteraa, "Smart real time wearable navigation support system for BVIP," *Alexandria Engineering Journal*, vol. 62, pp. 223–235, 2023. <https://doi.org/10.1016/j.aej.2022.06.060>
- [25] S. Tennekoon, N. Abhayasinghe, and N. Wedasingha. "Object Recognition and Assistance System for Visually Impaired Shoppers," *Proceedings of the SLIIT International Conference On Engineering and Technology*, vol. 2, 2023. <https://doi.org/10.54389/RRFZ2311>

- [26] E. J. Argüello Prada, and L. M. Santacruz Forero, "A belt-like assistive device for visually impaired people: Toward a more collaborative approach," *Cogent Engineering*, vol. 9, no. 1, p. 2048440, 2022. <https://doi.org/10.1080/23311916.2022.2048440>
- [27] A.V. Canez, J. Sartori, R. Barwaldt, and R. N. Rodrigues, "Collision detection with monocular vision for assisting in mobility of visually impaired people." In 2019 8th Brazilian Conference on Intelligent Systems (*BRACIS*), 2019, pp. 269-274. <http://dx.doi.org/10.1109/BRACIS.2019.00055>
- [28] H. P. Buimer, M. Bittner, T. Kostelijk, T. M. Van Der Geest, A. Nemri, R. J. A. Van Wezel, and Y. Zhao. "Conveying facial expressions to blind and visually impaired persons through a wearable vibrotactile device." *PloS one* 13, no. 3 , p. 0194737, 2018. <https://doi.org/10.1371/journal.pone.0194737>
- [29] Y. Lin, K. Wang, W. Yi, and S. Lian, "Deep learning based wearable assistive system for visually impaired people," In Proceedings of the IEEE/CVF international conference on computer vision workshops, 2019. <https://doi.org/10.48550/arXiv.1908.03364>
- [30] B. Kuriakose, I. M. Ness, M. Å. skov Tengstedt, J. M. Svendsen, T. Bjørseth, B. L. Pradhan, and R. Shrestha, "Turn Left Turn Right-Delving type and modality of instructions in navigation assistant systems for people with visual impairments," *International Journal of Human-Computer Studies*, vol. 179, p. 103098, 2023. <https://doi.org/10.1016/j.ijhcs.2023.103098>
- [31] M. A. Hussain, M. G. Ullah, A. Fareed, and B. Sohail, "The smartcane for blind people an electronically smart stick to aid mobility," *International Journal of Computer Science and Information Security*, vol. 14, no. 4, p. 276, 2016.
- [32] S. Mattoccia and P. Macri, "3D glasses as mobility aid for visually impaired people," in Computer Vision – ECCV 2014 Workshops, Zurich, Switzerland, Sept. 6–7 and 12, 2014, Proceedings, Part III, vol. 13, pp. 643–658, 2015. https://doi.org/10.1007/978-3-319-16199-0_38
- [33] M. Sohan, T. Sai Ram, C.V. Rami Reddy, "A review on YOLOv8 and its advancements," in International Conference on Data Intelligence and Cognitive Informatics, Springer, 2024, pp. 1–8. https://doi.org/10.1007/978-981-99-7962-2_39
- [34] Ultralytics, "YOLOv8 Documentation," [Online]. Available: <https://docs.ultralytics.com/models/yolov8>, [Accessed: Nov.12, 2023].
- [35] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C. L. Chang, and M. Grundmann, "Mediapipe hands: On-device real-time hand tracking," *arXiv preprint arXiv:2006.10214*, 2020. <https://doi.org/10.48550/arXiv.2006.10214>
- [36] R. A. Setyawan, R. Soenoko, P. Mudjirahardjo, and M. A Choiron, "Measurement accuracy analysis of distance between cameras in stereo vision," In 2018 Electrical power, electronics, communications, controls and informatics seminar (EECCIS), 2018, pp. 169-172. <https://doi.org/10.1109/EECCIS.2018.8692999>