

توانبخشی بینایی پشت میزهای کار با به کارگیری بینایی ماشین

امیرمحمد برسلانی^۱، حمیدرضا محمدی دانیالی^۲، آرمان مردانی^۳

^۱ دانشکده مهندسی مکانیک، دانشگاه صنعتی نوشیروانی بابل، مازندران، ایران. ایمیل: amirmohammadbarsalani@gmail.com

^۲ دانشکده مهندسی مکانیک، دانشگاه صنعتی نوشیروانی بابل، مازندران، ایران. ایمیل: mohammadi@nit.ac.ir

^۳ دانشکده مهندسی مکانیک، دانشگاه صنعتی نوشیروانی بابل، مازندران، ایران. ایمیل: armanmardani@nit.ac.ir

طرف مکاتبه با نشریه : حمیدرضا محمدی دانیالی

چکیده

بر اساس آمار سازمان بهداشت جهانی، بیش از ۳۸ میلیون نفر نابینا و بیش از ۱۱۰ میلیون نفر با مشکلات جدی بینایی روبه رو هستند. این پژوهش به طراحی یک سیستم نوآورانه، کم هزینه و در دسترس برای کمک به افراد نابینا در محیط‌های کاری پرداخته است. هدف اصلی این سیستم، توانبخشی بینایی با تمرکز بر شناسایی اجسام، هدایت صوتی و کمک به برداشتن ایمن اجسام توسط فرد نابینا است. این سیستم با استفاده از پردازش تصویر و مدل‌های از پیش آموزش دیده، اجسام و دست انسان را شناسایی کرده و فاصله و ابعاد اجسام را تخمین می‌زند. سپس با صدور فرمان‌های صوتی، فرد نابینا را به گونه‌ای هدایت می‌کند که بتواند با ایمنی و دقت بالا جسم موردنظر را بردارد. تمرکز این پژوهش بر افزایش استقلال و ایمنی افراد نابینا بوده و طراحی آن به گونه‌ای صورت گرفته است که بهره‌گیری از آن ساده، مقرون به صرفه و اثربخش باشد و بتواند نقش مؤثری در بهبود کیفیت زندگی افراد نابینا ایفا کند.

واژه‌های کلیدی: بینایی ماشین، افراد دارای اختلال بینایی، شناسایی اجسام، هدایت صوتی، مدل‌های پیش‌آموزش دیده

۱. مقدمه

توانبخشی بینایی در طیف وسیعی از اختلالات بینایی کاربرد دارد، از جمله نابینایی، کم بینایی، اختلالات میدان دید و اختلالات حرکات چشم. این روش‌ها با هدف بازگرداندن حداکثر توانایی‌های بینایی فرد و کمک به او برای زندگی مستقل و مشارکت فعال در جامعه انجام می‌شوند. طبق آماري 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 و ممکن است برخی از این موارد به نابینایی کامل ختم شوند [۱]، که بیانگر میزان اهمیت و جدی بودن مساله است. هدف از توانبخشی بینایی 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 [۲]. جهت درک جوانب مختلف توانبخشی بینایی لازم است چهار جنبه اصلی از 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 [۳]. خدمات جامع توانبخشی بینایی 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 [۴]. در ادامه به بخشی از پژوهش‌های صورت پذیرفته در حوزه توانبخشی بینایی پرداخته شده است. نارایانی و همکاران [۵]، به منظور بهبود عملکرد افراد نابینا در راه رفتن، با استفاده از دو سنسور اولتراسونیک، یک ماژول دوربین و یک سنسور PIR (Passive infrared sensor)، یک عصا طراحی کردند که قادر به شناسایی موانع موجود بر سر راه فرد نابینا می‌باشد. مارکون و اورتوچلی [۶]، طی تحقیقی میزان موثر بودن شناسایی اجسام به صورت بی‌درنگ (Real-time) را در شرایط کنترل نشده در یک محیط واقعی مورد بررسی قرار دادند. کیلیان و همکاران [۷]، دستکش هوشمندی را طراحی کردند که با استفاده از دوربین ToF (Time of Flight) تعبیه شده روی آن، این امکان را داشت تا اجسام، موانع و همینطور فاصله آن‌ها را تشخیص دهد.

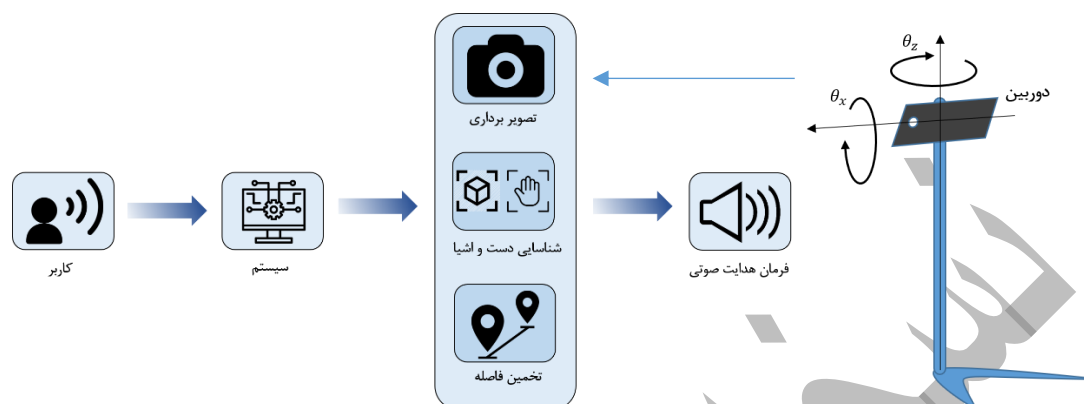
اهرچی و کبیر [۸]، طرحی را پیشنهاد دادند. به این صورت که با استفاده از استریوویژن (Stereovision) و سنسور GPS میتوان افراد نابینا را به سمت هدف و مسیری که دارند هدایت کرد. آدپ و همکاران [۹]، برای توسعه استقلال افراد نابینا در راه رفتن، سیستم هوشمندی را طراحی کردند. در این پروژه سنسورهای فروسرخ و دوربین مورد استفاده قرار گرفتند و محل پردازش داده‌ی حاصل شده در یک برد رزبری پای (Raspberry Pi) می‌باشد. سپس فرمان‌های صوتی لازم جهت هدایت فرد نابینا صادر می‌شود. زنتو و همکاران [۱۰]، 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 [۱۱]، یک سیستم پوششی را ارائه دادند که قابلیت شناسایی موقعیت‌های ترافیکی، چراغ‌های راهنما، دوچرخه سواران، پل‌های عابر پیاده، عابرین پیاده، اتومبیل‌ها و برخی موارد دیگر را دارد. یونس و همکاران [۱۲]، 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 [۱۳]، با استفاده از یک دوربین و یک کمربند لمسی، سیستمی را پیاده سازی کردند که به افراد دارای اختلالات بینایی کمک می‌کند تا حالت چهره مخاطب آن‌ها را شناسایی و گزارش کند. بوترا [۱۴]، سیستمی را جهت تشخیص مسیر امن برای افرادی که با عارضه‌های بینایی روبرو هستند طراحی و پیشنهاد کرده است. تنکون و همکاران [۱۵]، مدلی را طراحی کردند که به افرادی که از اختلالات بینایی رنج می‌برند کمک می‌کند تا با استقلال بیشتری برای خرید به فروشگاه‌ها بروند.

در پژوهش‌های اخیر تمرکز عموماً بر روی مساله جابجایی افراد توانخواه معطوف بوده است در حالیکه این افراد با چالش‌های بسیاری رو به رو هستند. با توجه به سبک زندگی امروزه که بخش قابل توجهی از زمان افراد پشت میزهای کار گذرانده می‌شود این نیاز حس می‌شود که افراد توانخواه نیز این امکان را داشته باشند تا بتوانند در فعالیت‌های خود استقلال بیشتری کسب کنند. این پژوهش به مرتفع سازی این چالش پرداخته و سعی شده است تا گامی در راستای بهبود وضعیت استقلال این افراد بردارد.

۲. مواد و روش‌ها

پژوهش پیش‌رو ایده و طرح اولیه از سیستمی را ارائه می‌دهد که قادر به شناسایی اجسام و دست انسان می‌باشد و پس از به انجام رساندن این مرحله، به تخمین فاصله اجسام از دست انسان که روی میز قرار گرفته است می‌پردازد. با بهره‌گیری از سیستم تعامل صوتی تعبیه شده در برنامه سیستم، پیام صوتی مناسب، جهت راهنمایی فرد به منظور دسترسی به جسم مورد نظر ارائه می‌شود. لازم به ذکر است، زبان برنامه نویسی پایتون در این پروژه مورد 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.



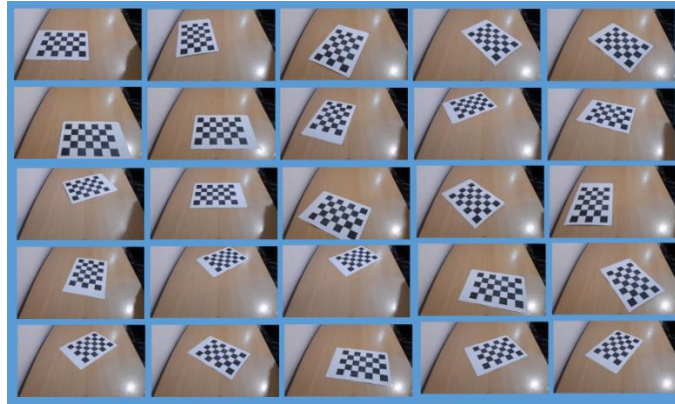
شکل ۱: نمایی کلی از نحوه عملکرد سیستم

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 قرار گیری دوربین 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 فرایند به درستی انجام خواهد گرفت.

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 فرایندی جهت کالیبراسیون دوربین‌های مختلف نسبت به مشخصه‌های هر دوربین و موقعیت قرار گیری دوربین نسبت به محیط ضروری است. فرایند کالیبراسیون با هدف محاسبه ماتریس تبدیل T_c که در برگیرنده مشخصه‌های ذاتی دوربین و موقعیت قرار گیری آن نسب به سطح مورد بررسی است، صورت می‌پذیرد.

$$T_c = T_I T_E \quad (۱)$$

در این پژوهش فرایند کالیبراسیون با هدف محاسبه ماتریس T_c با استفاده از صفحه شطرنجی و کتابخانه OpenCV صورت پذیرفته است. در شکل زیر نمونه تصاویر مربوط به فرایند کالیبراسیون ارائه شده است.



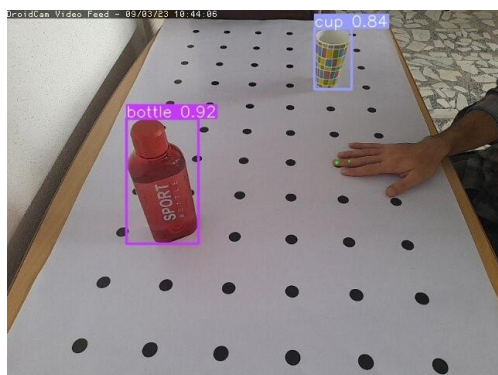
شکل ۲: نمونه تصاویر ضبط شده جهت فرایند کالیبراسیون

با استفاده از رابطه زیر می‌توان با داشتن ماتریس T_c و مختصات پیکسلی هر نقطه از تصویر به مختصات آن در فضای حقیقی دست یافت.

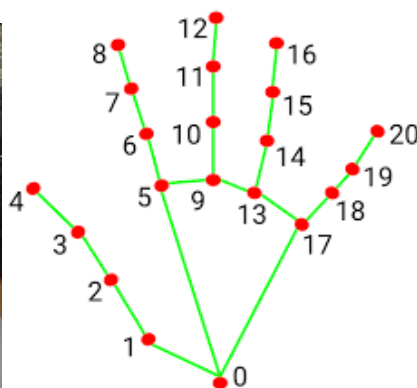
$$K_{ij} = T_c K_{ij}^w \quad (۲)$$

مدل YOLOv8n به واسطه کمترین میزان پارامتر یادگیری آن، سریع‌ترین نسخه و البته کمترین میزان دقت را در مقایسه با سایر مدل‌ها دارد و YOLOv8x 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 نانو دقت مناسبی در کاربرد مورد نظر دارد از این نسخه (YOLOv8n) 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 فرایند را نیز بهبود می‌بخشد. مدل MediaPipe یکی از ابزارهای پیشرفته در زمینه شناسایی نقاط مختلف بدن و تحلیل حرکات است که توسط 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، از 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. مدیاپایپ به دلیل دقت و سرعت بالا در شناسایی نقاط بدن، به یک ابزار قدرتمند در بسیاری از برنامه‌ها تبدیل شده است.



(ب)



(الف)

شکل ۳: (الف) نقاط کلیدی قابل شناسایی توسط مدل مدیاپایپ بر روی دست انسان. (ب) نتیجه مدل های yolo و مدیاپایپ در شناسایی اجسام و نقاط کلیدی دست انسان

همانطور که در شکل ۳ نمایش داده شده است، مدیاپایپ قادر به شناسایی ۲۰ نقطه کلیدی بر روی دست انسان می باشد. در این پژوهش از مدیاپایپ جهت شناسایی دست استفاده شده و به عنوان مرجع انتهای انگشت میانی (نقطه شماره ۱۲) انتخاب شده است. در ادامه با داشتن مختصات مربوط به اجسام و دست به صورت پیکسلی، به محاسبات تعیین موقعیت آن ها در فضای حقیقی پرداخته شده است. با استفاده از ماتریس تبدیل سیستم که در بخش کالیبراسیون محاسبه گردید و رابطه (۲) از مختصات پیکسلی به مختصات حقیقی نقاط می توان دست یافت.

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

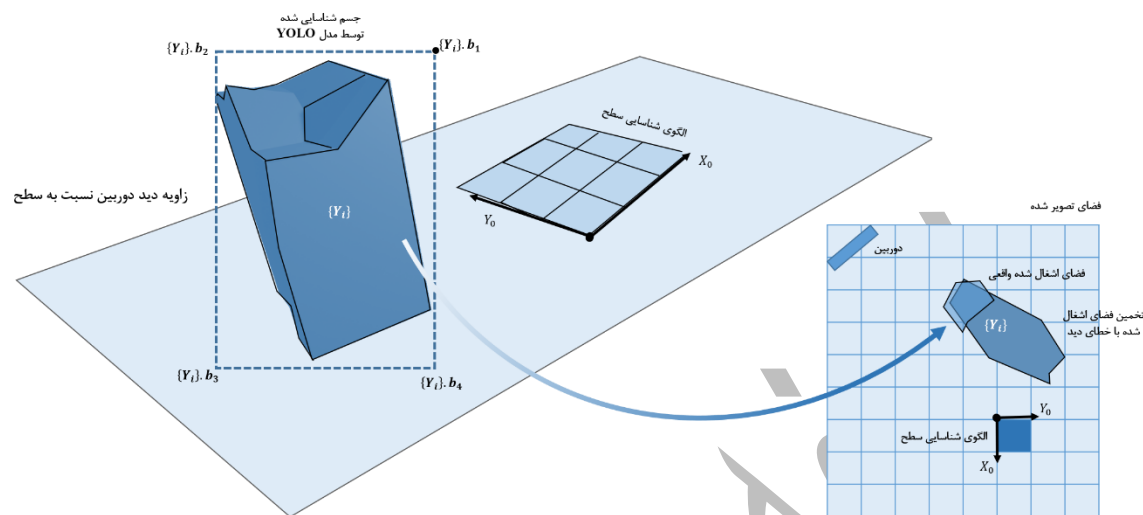
در رابطه (۳)، K_{ij} معرف مختصات پیکسلی و K_{ij}^w معرف مختصات در فضای حقیقی می باشد. از آنجایی که اجسام روی سطح میز قرار داده شده اند و نقطه تماس آن ها با سطح میز مورد نظر است، پارامتر مربوط به ارتفاع عمودی $(\{Y_i\}.z^p)$ برابر با صفر قرار داده شده است. این فرضیه به این دلیل در نظر گرفته می شود که تمامی نقاط تماس اجسام با سطح میز در یک صفحه افقی قرار دارند. در نتیجه، مطابق رابطه (۳) پارامترهای دیگر مانند مختصات $(\{Y_i\}.x_b^p, \{Y_i\}.y_b^p)$ از معادلات حاصل می شوند و موقعیت دقیق نقاط تماس را مشخص می کنند. این رویکرد به ساده سازی معادلات و کاهش پیچیدگی محاسبات کمک می کند، چرا که تنها نیاز به بررسی مختصات دوبعدی اجسام بر روی سطح میز داریم.

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

۲-۱ 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 می شود که سطح اشغال شده توسط جسم روی میز بسیار بیشتر از میزان واقعی به نظر می رسد. 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.



شکل ۴: نمایی از خطای سیستم در صورت تعیین نکردن محدوده سطح اشغالی اجسام

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 YOLO، 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 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.

$$[\{Y_i\}.m]_{2 \times 1} = \text{mean}(\{Y_i\}.b_3, \{Y_i\}.b_4)$$

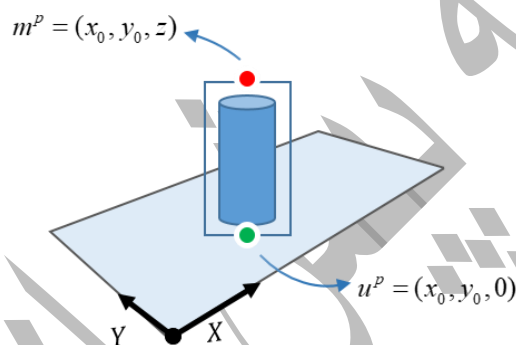
$$[\{Y_i\}.u]_{2 \times 1} = \text{mean}(\{Y_i\}.b_1, \{Y_i\}.b_2)$$

(۵)

در رابطه (۵)، با داشتن مختصات پیکسلی نقطه میانی ضلع پایین از جعبه مرزی ($\{Y_i\} \cdot \mathbf{u}$)، به محاسبه مختصات فضای حقیقی آن ($\{Y_i\} \cdot \mathbf{u}[1]^p$ و $\{Y_i\} \cdot \mathbf{u}[2]^p$) پرداخته شده است. 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.

$$\begin{bmatrix} \{Y_i\} \cdot \mathbf{u}[1] \\ \{Y_i\} \cdot \mathbf{u}[2] \\ 1 \end{bmatrix} = \mathbf{T}_C \begin{bmatrix} \{Y_i\} \cdot \mathbf{u}[1]^p \\ \{Y_i\} \cdot \mathbf{u}[2]^p \\ 0 \\ 1 \end{bmatrix} \quad (۶)$$

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 مولفه مشترک و همسان دارند زیرا در راستای عمود بر سطح میز قرار گرفته‌اند.



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

از آنجایی که مولفه‌های 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 مولفه مختصات فضای حقیقی نقطه پایینی نوشت. در این صورت تنها مولفه مجهول، پارامتر مربوط به محور عمود بر سطح ($\{Y_i\} \cdot m_z^p$) می‌باشد که متعلق به نقطه بالایی از جعبه مرزی بوده و می‌توان آن را معادل ارتفاع اشیاء در نظر گرفت.

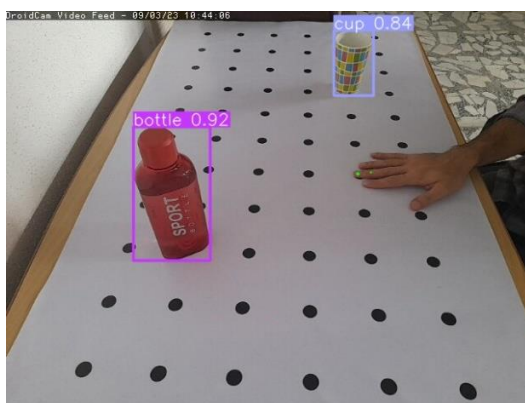
$$\begin{bmatrix} \{Y_i\} \cdot \mathbf{m}[1] \\ \{Y_i\} \cdot \mathbf{m}[2] \\ 1 \end{bmatrix} = \mathbf{T}_C \begin{bmatrix} \{Y_i\} \cdot \mathbf{u}[1]^p \\ \{Y_i\} \cdot \mathbf{u}[2]^p \\ \{Y_i\} \cdot m_z^p \\ 1 \end{bmatrix} \quad (۷)$$

به منظور تخمین پارامتر $\{Y_i\} \cdot m_z^p$ ، 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.

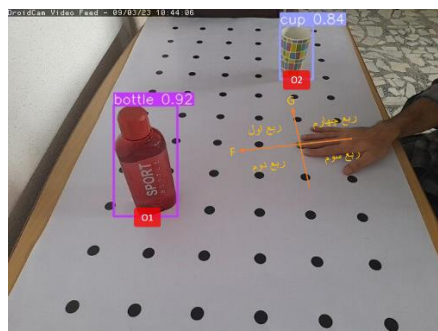
۲-۲ هدایت دست

در مراحل پیشین، شناسایی دست و اشیاء با استفاده از مدل‌های YOLO 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. نقطه شماره ۱۱ (شکل ۳) نیز به دلیل مطرح شده مورد شناسایی قرار گرفته است.



شکل ۶: شناسایی دو نقطه کلیدی بر روی دست و اجسام موجود

اولین گام تشخیص موقعیت هر جسم نسبت به دست انسان و تقسیم بندی آن به چهار بخش که در تصویر ۷ نمایش داده شده است می‌باشد. برای مثال در تصویر زیر موقعیت جسم شماره ۲ (CUP) نسبت به دست در ربع چهارم و بطری در ربع دوم قرار گرفته است. محور F راستای انگشت میانی 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 ۷: تقسیم بندی چهارگانه نواحی اطراف دست

معادلات مربوط به محورهای F و G مطابق روابط (۸) قابل محاسبه می‌باشد.

$$M_F = \frac{P_{h_1}^w[2] - P_{h_2}^w[2]}{P_{h_1}^w[1] - P_{h_2}^w[1]}$$

$$F(x) = M_F \times (x - P_{h_2}^w[1]) + P_{h_2}^w[2]$$

$$G(x) = \frac{-1}{M_F} \times (x - P_{h_2}^w[1]) + P_{h_2}^w[2]$$

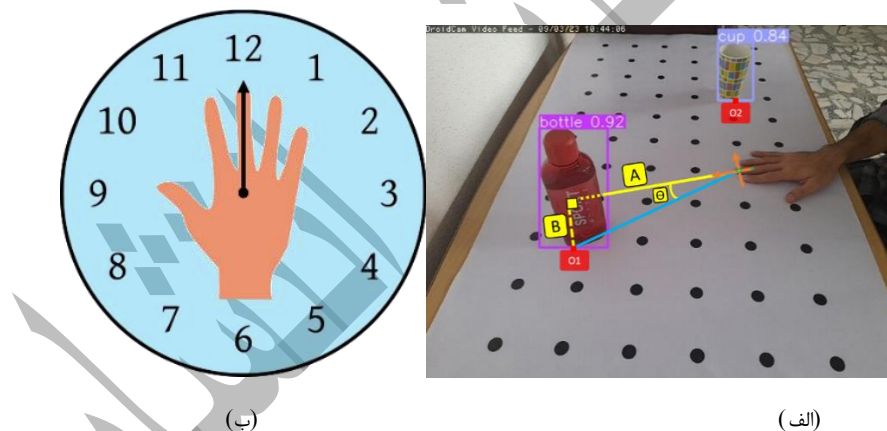
(۸)

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 ۴ ربع شناسایی کرد.

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

$o^w[2] > F(o^w[1])$, $o^w[2] > G(o^w[1])$	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

گام بعدی تشخیص زاویه نقاط کلیدی اجسام نسبت به دست انسان می‌باشد. مطابق شکل ۸ زوایای نقاط کلیدی (نامگذاری شده با حرف **O**) نسبت به محورهای **F** و **G** محاسبه شده و مقدار θ برگردانده می‌شود.



شکل ۸: (الف) نمایی از روش تخمین زاویه جسم نسبت به دست (ب) همپوشانی راستای دست و عقربه‌های ساعت

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 (۸) 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.

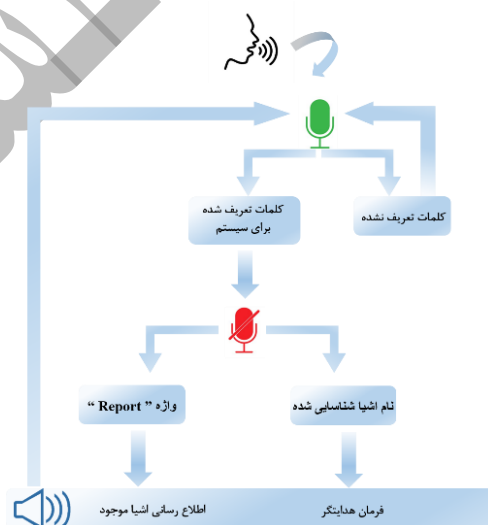
جدول ۲: تقسیم بندی جهت‌های دست با استفاده از جهت عقربه‌های ساعت

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

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 pytsx3، تعدادی از واژگان معنادار برای سیستم تعریف شده‌اند که شامل لیست نام اجسام قابل شناسایی توسط YOLO و واژه‌ی "Report" می‌شود. زمانی که سیستم دستور "Report" را دریافت می‌کند، ابتدا تصویری توسط دوربین ثبت می‌شود و با استفاده از مدل YOLO، اجسام موجود در تصویر شناسایی می‌شوند. سپس نام این اجسام در لیستی ذخیره می‌شود و جایگاه دست کاربر نیز با استفاده از مدل مدیاپایپ 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 pytsx3 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.

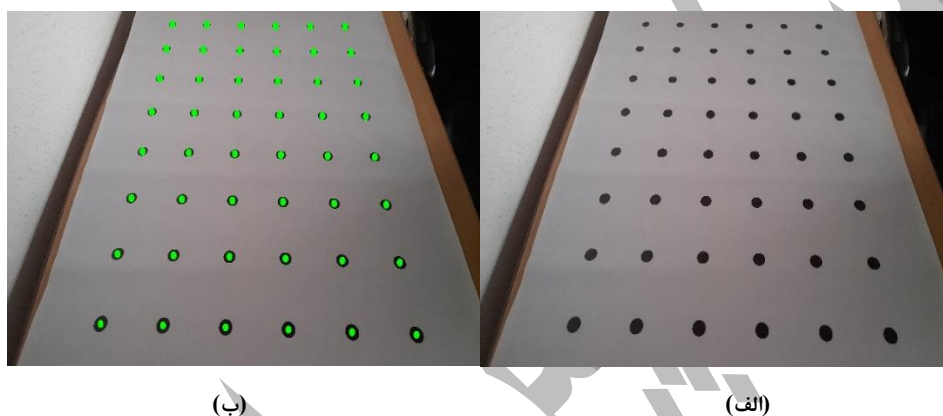
پس از ارائه‌ی گزارش، برنامه در حالت انتظار باقی می‌ماند تا دستور جدیدی دریافت کند. این دستور می‌تواند فراخواندن نام یکی از اجسام شناسایی شده باشد. در این حالت، سیستم با دریافت نام جسم مورد نظر، کاربر را به صورت صوتی هدایت می‌کند و مشخص می‌شود که دست خود را به چه سمتی (بر اساس عقربه‌های ساعت) و به چه طولی (در واحد سانتی‌متر) جابجا کند تا بتواند به جسم مورد نظر دسترسی پیدا کند.



شکل ۹: نحوه عملکرد سیستم تعامل صوتی تعبیه شده در سیستم

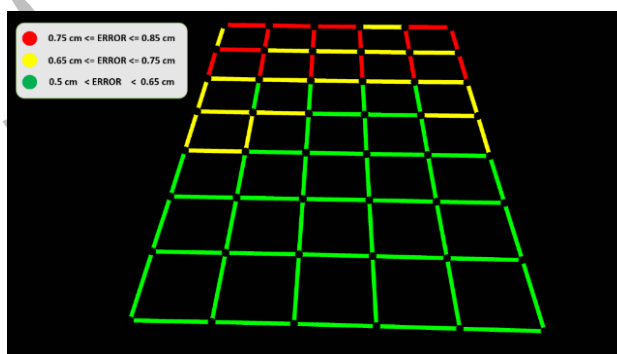
۳. یافته‌ها

در این بخش، سیستم از جهات مختلفی مورد ارزیابی و آزمایش قرار گرفته است. این ارزیابی شامل دقت تخمین فواصل، اندازه‌گیری ارتفاع اجسام، و شناسایی حرکت 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. به منظور بررسی و ارزیابی خطای سیستم در تشخیص فواصل مختلف در محدوده کاری آن، بنری مشابه با تصویر ۱۰ طراحی و آماده شده است. این بنر 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 شده و گزارش شده است تا قابلیت‌های سیستم در اندازه‌گیری فواصل مختلف به‌طور جامع تحلیل گردد.



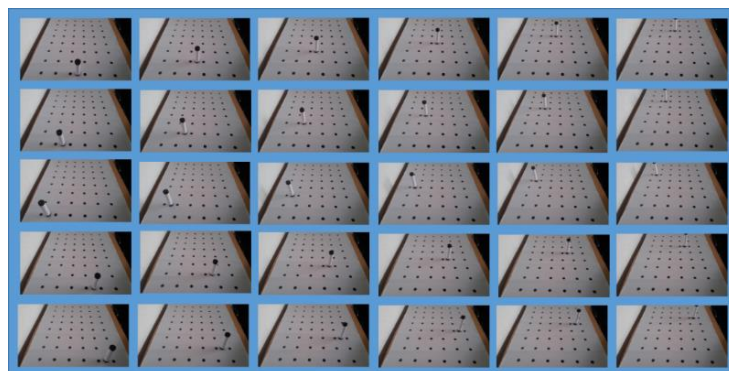
شکل ۱۰: (الف) بنر طراحی شده جهت آزمایش شناسایی فواصل (ب) شناسایی دایره‌های سیاه رنگ توسط برنامه

در این آزمایش، ابتدا دایره‌های سیاه رنگ روی بنر توسط برنامه نوشته شده شناسایی شده و گزارش گردیدند. 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 ۰٫۶۱ سانتی‌متر و بیشترین و کمترین میزان خطای گزارش شده به ترتیب ۰٫۸۵ و ۰٫۵۲ سانتی‌متر بوده است.



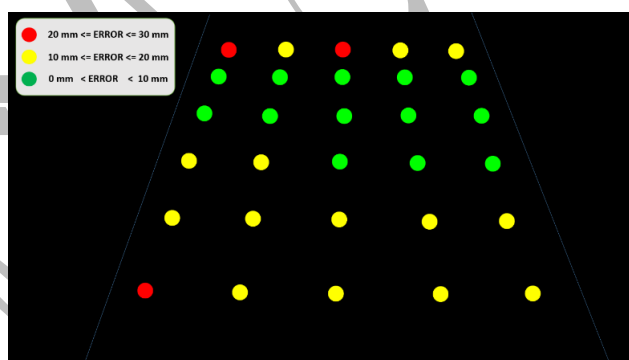
شکل ۱۱: نمایش خطای تخمین فواصل توسط سیستم

همانطور که در تصویر ۱۱ مشاهده می‌شود، هر چه فاصله از دوربین بیشتر شده است، خطای تخمین گزارش شده نیز افزایش یافته است. این موضوع به سبب عاملی همچون تراکم پیکسلی کمتر و معرف فضای واقعی گسترده‌تر نسبت به محدوده‌های نزدیک‌تر به دوربین اتفاق افتاده است. از دیگر عوامل موثر نیز می‌توان به نوع و مشخصه‌های ذاتی دوربین همچون تراکم پیکسلی آن اشاره کرد. همانطور که اشاره شد، سیستم قابلیت شناسایی ارتفاع اجسام را داشته 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.



شکل ۱۲: تصاویر آزمایش تخمین ارتفاع

همانطور که در شکل ۱۲ قابل مشاهده است، جسم هدف در ۳۰ موقعیت مختلف جایگذاری شده و در هر نقطه ارتفاع آن تخمین زده شده است.



شکل ۱۳: نمایی از خطای سیستم طی 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 قرارگیری 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 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 فرایند تخمین ارتفاع صورت می‌پذیرد.

۴. بحث

طی پژوهش‌های صورت گرفته، سیستم‌های توانبخشی طراحی شده برای افرادی که با اختلالات بینایی مواجه هستند، عموماً در زمینه جابجایی آنها از محلی به محل دیگر متمرکز شده‌اند. با توجه به سبک زندگی امروزی، بسیاری از افراد به صورت اجتناب‌ناپذیری بخش قابل توجهی از زمان خود را پشت میزها می‌گذرانند و این موضوع یکی از چالش‌هایی است که افراد توانخواه در حوزه بینایی با آن مواجه هستند. در این پژوهش تلاش شده است تا گامی در جهت توسعه سیستم‌های توانبخشی در این زمینه و به رفع این چالش برداشته شود. در حوزه بینایی ماشین، تخمین فواصل معمولاً با استفاده از دو لنز دوربین، کینکت، ترکیب دوربین و لیزر و ... صورت می‌پذیرد. 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 به صرفه‌تر و در دسترس‌تر باشد. از سوی دیگر، تکنولوژی‌های به‌روزی همچون YOLO و مدیاپایپ در پیاده‌سازی این سیستم به کار گرفته شده‌اند که عملکرد بهینه‌ای را برای سیستم به ارمغان می‌آورند. ایده‌های مطرح شده با این هدف ارائه شده‌اند که فرد توانخواه با کمترین هزینه و بالاترین دسترسی بتواند از این سیستم بهره‌مند شود. طراحی این سیستم می‌تواند توسعه یابد و به صورت یک نرم‌افزار قابل نصب بر روی تلفن همراه افراد درآید، به‌طوری‌که توانخواهان تنها با نصب آن بتوانند از امکانات آن استفاده کنند.

۵. محدودیت‌ها

سیستم پیشنهاد شده فضای اشغالی توسط اجسام را با دقت قابل قبولی تخمین می‌زند، و امکان محاسبه دقیق ابعاد آن‌ها فراهم نشده. سیستم نسبت به لنز دوربین‌های مختلف انعطاف پذیر بوده و تنها محدودیت آن در این زمینه است که پس از فرایند کالیبراسیون، ارتفاع آن از سطح و زاویه‌ای که نسبت به میز دارد باید ثابت و بدون تغییر باقی بماند.

۶. پیشنهادها

برای گسترش و پیشبرد تحقیق ارائه‌شده، مسیرهای گوناگونی قابل پیگیری است که می‌تواند به بهبود کارایی و دقت سیستم کمک کند. از جمله این مسیرها می‌توان به بهره‌گیری از سخت‌افزارهای پیشرفته‌تر اشاره کرد، که تأثیر چشمگیری بر کارایی کلی سیستم خواهد داشت. همچنین، انتقال بخشی از پردازش‌ها به سرور و طراحی برنامه‌ای برای تعامل با آن، امکان کاهش بار پردازشی دستگاه‌های کاربران و افزایش کارایی سیستم را فراهم می‌کند. به‌علاوه، کاهش ابعاد تصاویر در مرحله پیش‌پردازش، می‌تواند زمان پردازش را کاهش داده و دقت تحلیل‌ها را بهبود بخشد. یکی دیگر از اقدامات مؤثر، استفاده از پردازنده‌های گرافیکی به جای پردازنده‌های معمولی است؛ این پردازنده‌ها با ساختار ویژه‌ای که برای پردازش چندرخی دارند، می‌توانند به افزایش سرعت سیستم کمک کنند. در کنار این موارد، استفاده از سیستم تعامل صوتی گوگل به شرط دسترسی به اینترنت، می‌تواند کیفیت رابط صوتی را ارتقا دهد. از سوی دیگر، با افزایش تعداد اجسام قابل‌شناسایی در شبکه YOLO و آموزش آن با داده‌های جدید، امکان شناسایی تنوع بیشتری از اجسام فراهم می‌شود. کاربرد الگوریتم‌های یادگیری ماشین برای کاهش خطاهای عددی نیز می‌تواند به دقت بیشتر سیستم منجر شود. همچنین، شناسایی دو دست با استفاده از مدل مدیاپایپ، قابلیت‌های بیشتری در تعامل با کاربر ایجاد می‌کند، به‌ویژه برای هدایت دست نزدیک‌تر به جسم هدف. استفاده از قابلیت Segmentation در YOLO به جای جعبه مرزی نیز امکان تشخیص دقیق‌تر و بهینه‌تر فضای اشغالی جسم را مهیا می‌سازد. افزودن بعد سوم به فضای کاری سیستم از طریق نوردهی فلش‌ری، درک عمیق‌تری از موقعیت اجسام فراهم می‌کند و به نوعی عمق را به سیستم دوبعدی موجود اضافه می‌کند. علاوه بر این، برای

هدایت دست کاربر توانخواه، می‌توان به‌جای تعامل صوتی از عملگرهای لغزشی بهره برد که به کاهش آلودگی صوتی و تمرکز بیشتر منجر می‌شود. در نهایت، به‌کارگیری مدل‌های زبانی مانند ChatGPT از طریق API نیز می‌تواند سیستم تعامل صوتی را به‌طرز موثرتری ارتقا دهد و تجربه کاربری بهتری ایجاد کند.

۷. نتیجه‌گیری

سیستم پیشنهادی به دلیل دسترسی بالا و دقت قابل قبولی که در تست‌های انجام‌شده از خود نشان داده است، پتانسیل بالایی برای استفاده مؤثر در حوزه توانبخشی بینایی دارد و می‌تواند به بهبود کیفیت زندگی افراد کمک شایانی کند.

۸. مراجع

- [1] Organization, W.H. Rehabilitation. 2024; Available from: <https://www.who.int/news-room/fact-sheets/detail/rehabilitation>.
- [2] Edwards, J. D., Valdés, E. G., Peronto, C., Castora-Binkley, M., Alwerdt, J., Andel, R., & Lister, J. J., 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*, 2015. 70(3): p. 417-422.
- [3] Colenbrander, A., My Perspective on Vision and Vision Rehabilitation. *Ophthalmology Science*, 2024: p. 100532.
- [4] Wilkinson, M.E. and K.S. Shahid, Low vision rehabilitation: An update. *Saudi Journal of Ophthalmology*, 2018. 32(2): p. 134-138.
- [5] Narayani, T. L., Sivapalanirajan, M., Keerthika, B., Ananthi, M., & Arunarani, M. Design of Smart Cane with integrated camera module for visually impaired people. in 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS). 2021. IEEE.
- [6] Marcon, M. and A. Ortoncelli, Object Recognition to Support Navigation Systems for Blind in Uncontrolled Environments. *Anais do Computer on the Beach*, 2022. 13: p. 274-281.
- [7] Kilian, J., Neugebauer, A., Scherffig, L., & Wahl, S., The unfolding space glove: A wearable spatio-visual to haptic sensory substitution device for blind people. *Sensors*, 2022. 22(5): p. 1859.
- [8] Aharchi, M. and 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. 2022. Springer.
- [9] Adep, T., Nikam, R., Wanewe, S., & Naik, D. K. B., Visual assistant for blind people using raspberry pi. *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol*, 2021. 3307: p. 671-675.
- [10] Zantou, P., A.M. Mousse, and B. Atohou, An intelligent based system for blind people monitoring in a smart home. 2020.

- [11] Banu, K. T., Supriya, K., Sony, K., Chandana, M., Bhavana, M., & Ali, M. B. F, OUTDOOR & INDOOR VISION BASED LOCALIZATION FOR BLIND PEDESTRIAN NAVIGATION ASSISTANCE.
- [12] Younis, O., Al-Nuaimy, W., Rowe, F., & Alomari, M. H., A smart context-aware hazard attention system to help people with peripheral vision loss. *Sensors*, 2019. 19(7): p. 1630.
- [13] Buimer, H. P., Bittner, M., Kostelijk, T., Van Der Geest, T. M., Nemri, A., Van Wezel, R. J., & Zhao, Y., Conveying facial expressions to blind and visually impaired persons through a wearable vibrotactile device. *PloS one*, 2018. 13(3): p. e0194737.
- [14] Bouteraa, Y., Smart real time wearable navigation support system for BVIP. *Alexandria Engineering Journal*, 2023. 62: p. 223-235.
- [15] Tennekoon, S., N. Abhayasinghe, and N. Wedasingha, Object Recognition and Assistance System for Visually Impaired Shoppers. 2023.

Vision Rehabilitation at Workstations Using Machine vision

Amirmohammad Barsalani ¹, Hamidreza Mohammadi Daniali ², Arman Mardani ³

¹ Department of Mechanical Engineering, Babol Noshirvani University of Technology, Mazandaran, Iran. Email: amirmohammadbarsalani@gmail.com

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

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

Abstract— According to the World Health Organization, over 38 million people are blind, and more than 110 million face severe visual impairments. This research focuses on designing an innovative, cost-effective, and accessible system to assist visually impaired individuals in workstation environments. The primary objective of this system is to provide assistive vision by identifying objects, issuing audio guidance, and aiding in the safe handling of objects by blind individuals. Utilizing image processing techniques and pre-trained models, the system identifies objects and human hands while estimating the distance and dimensions of the objects. Subsequently, audio commands guide the visually impaired user, ensuring they can pick up the desired object safely and efficiently. The system emphasizes enhancing independence and safety for visually impaired individuals in performing daily tasks. Its design ensures ease of use, affordability, and effectiveness, aiming to significantly improve the quality of life for the visually impaired by facilitating their interaction with their environment.

Keywords: Machine Vision, Visually Impaired, Workstation Assistance, Object Detection, Audio Guidance, Pre-trained Models

نسخه چاپی
از نشر
گستر