

تخمین عمق نگاه ناظر با استفاده از ردیابی سه بعدی چشم با روش یادگیری عمیق

داریوش فرجی مزرعه خلف^۱، دانشجوی دکترا، میرهادی سیدعربی^۲، استاد، رضا افروزیان^۳، استادیار

۱- دانشکده مهندسی برق و کامپیوتر-دانشگاه تبریز-تبریز-ایران-daryush.faraji@gmail.com

۲- دانشکده مهندسی برق و کامپیوتر-دانشگاه تبریز-تبریز-ایران-seyedarabi@tabrizu.ac.ir

۳- دانشکده فنی و مهندسی میانه-دانشگاه تبریز-میانه-ایران-afrouzian@tabrizu.ac.ir

چکیده: در سال‌های اخیر با رشد فناوری‌های هوش مصنوعی، تحقیقات در زمینه برقراری ارتباط تعاملی بین انسان و کامپیوتر رشد فزاینده‌ای داشته است. روش‌های متعددی برای تعامل بین مغز انسان و ماشین ارائه شده‌اند. ردیابی خیرگی چشم با استفاده از بینایی ماشین و پردازش تصویر به علت مزایایی همچون غیرتهاجمی بودن و عدم استفاده از سیگنال‌های مضر رادیویی و مغناطیسی، ارزان بودن و قابل حمل بودن مورد توجه محققین قرار گرفته است. در این مقاله یک شیوه نوین برای تخمین عمق نگاه ناظر با استفاده از ردیابی سه بعدی چشم با روش‌های رگرسیون خطی و یادگیری عمیق ارائه شده است. برای دستیابی به اهداف تحقیق برای اولین بار یک مجموعه داده تهیه شد و در آن تعداد زیادی تصویر از چشم افراد با شرایط سنی و جنسی متنوع در حال نگاه به یک هدف متحرک در فاصله‌ها و شرایط روشنایی متفاوت ثبت شد. جهت افزایش دقت استخراج ویژگی‌های چشم از یک شبکه عصبی مبتنی بر یادگیری عمیق استفاده شد. با استفاده از ایده ره‌گیری تغییرات فیزیولوژیکی چشم از جمله محل قرارگیری و اندازه مردمک و فاصله مردمک دو چشم و تغییر محل نقاط انعکاسی قرنیه در اثر تغییرات میزان روشنایی محیط و فاصله هدف از چشم و همچنین پدیده‌هایی مانند اختلاف منظر و تطابق که مغز برای تشخیص عمق از آن بهره می‌گیرد، برای تشخیص عمق نگاه ناظر استفاده شد. نتایج نشان دادند که با روش رگرسیون خطی با دقت بالای ۹۰ درصد و با روش یادگیری عمیق با دقت نزدیک به ۹۳ درصد، عمق نگاه ناظر تخمین زده می‌شود.

واژه‌های کلیدی: ردیابی خیرگی چشم، تشخیص عمق نگاه، یادگیری عمیق، شبکه عصبی کانولوشنی، مردمک، پورکینیا.

Depth of Eye Gaze Estimation using 3D Eye Tracking using Deep Learning

Daryoush Faraji Mazraehkhalaf, MSc¹, Hadi Seyedarabi, Professor², Reza Afrouzian, Asisitant Professor

1- Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran, seyedarabi@tabrizu.ac.ir

2- Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran, daryush.faraji@gmail.com

3- Miyaneh Faculty of Engineering, University of Tabriz, Miyaneh, Iran, afrouzian@tabrizu.ac.ir

Abstract: Recently, with the growth of artificial intelligence technologies, research in the field of HCI has grown increasingly. Eye gaze tracking using machine vision and image processing has attracted the attention of researchers due to its advantages such as non-intrusive and no use of harmful radio and magnetic signals, cheapness and portability. In this article, a new method for estimating the observer's gaze depth using 3D eye tracking with linear regression and deep learning methods is presented. A Valuable data set was prepared for the first time, and in it, a large number of images of the eyes of people, with various age and gender in variable lighting conditions looking at a moving target at different distances, were recorded. A deep learning - based neural network was used to increase the accuracy of eye feature extraction. The idea of tracking the physiological changes of the eye, including the location and size of the pupil and the distance between the pupils of the two eyes and the change in the location of the reflective points of the cornea due to changes in the brightness of the environment and the distance of the target from the eye, also phenomena such as perspective difference and adaptation, was used to detect the observer's gaze depth. The results showed that the depth of observer's gaze is estimated with the linear regression method with an accuracy of over 90% and with the deep learning method with an accuracy of nearly 93%.

Keywords: Eye gaze detection, gaze depth Estimation, deep learning, Convolutional neural network

تاریخ ارسال مقاله: ۱۴۰۲/۱۱/۰۱

تاریخ اصلاح مقاله: ۱۴۰۳/۰۴/۱۱

تاریخ پذیرش مقاله: ۱۴۰۳/۰۶/۰۸

نام نویسنده مسئول: میرهادی سیدعربی

نشانی نویسنده مسئول: تبریز - بلوار ۲۹ بهمن - دانشگاه تبریز - دانشکده مهندسی برق و کامپیوتر - گروه مخابرات.



۱- مقدمه

دستگاه بینایی انسان^۱ یکی از تخصصی‌ترین عضوها در ادراک بشر و کسب اطلاعات از محیط اطراف و یکی از مهم‌ترین حواس پنج‌گانه است. استفاده از ردیابی چشم به‌عنوان یک ابزار مناسب برای ارتباط انسان با کامپیوتر محسوب می‌شود. این روش، به‌خصوص برای افرادی که معلول شده‌اند و توانایی حرکت خود را به‌طور کامل از دست داده‌اند، بسیار مفید خواهد بود. برخلاف آنچه که به نظر می‌آید، مشاهده دنیای اطراف ما ساده نیست و درک دیداری، یک فرایند واقعاً پیچیده است. چشم، یک عضو فعال و حساس به محیط اطراف است و در هر لحظه به‌صورت آبی به تغییرات محیط از خود واکنش نشان می‌دهد. جالب‌تر اینکه چشم انسان نه‌تنها از محیط پیرامون اثر می‌پذیرد، بلکه حالات درونی و فعالیت‌های ذهنی انسان نیز بر مشخصات هندسی چشم مؤثر است و موقعیت مردمک چشم، اندازه مردمک چشم و سرعت تغییرات اندازه مردمک، تغییرات اندازه خود چشم، ریتم پلک زدن و برخی حرکات غیرارادی چشم کاملاً از مغز و ذهن انسان تأثیر می‌پذیرند. تشخیص اینکه فردی در یک محیط خاص به کدام نقطه نگاه می‌کند و یا به کدام هدف توجه ویژه‌ای دارد، یک فرم ساده و طبیعی از تعامل انسان و کامپیوتر است. لذا ردیابی حرکات چشم^۲ و توسعه کاربردهای آن به‌عنوان یکی از روش‌های مدرن تعامل انسان - کامپیوتر و مغز - کامپیوتر برای خواندن ذهن انسان مطرح شده است.

باوجوداینکه در سال‌های اخیر الگوریتم‌ها و روش‌های مختلفی برای ردیابی چشم ارائه شده‌اند، اما تحقیقات برای پیدا کردن یک روش دقیق و پایدار همچنان ادامه دارد. شاید بتوان گفت علت اصلی خطاهای سیستم‌های ردیابی چشم، مشکل حساسیت چشم به تغییرات نور محیط و لحاظ نکردن شدت روشنایی در الگوریتم ردیابی چشم است. همان‌طور که می‌دانیم مردمک چشم میزان نور ورودی به چشم را با انبساط و انقباض ماهیچه‌های عنبیه کنترل می‌کند. اندازه مردمک باتوجه‌به شدت نور تنگ و گشاد می‌شود؛ بنابراین تغییرات نور محیط نه‌تنها اندازه مردمک چشم بلکه کل ساختار کره چشم را تحت‌تأثیر قرار می‌دهد و روی کلیه پارامترهای ردیابی چشم با هر روش و الگوریتمی که باشد، تأثیر می‌گذارد [۱].

یکی دیگر از مسائل اساسی در طراحی سیستم‌های ردیابی چشم مطالعه و لحاظ کردن دقیق آناتومی حرکات سر و چشم برای دیدن محیط اطراف است. حرکات چشم را می‌توان از دو منظر دسته‌بندی کرد: دسته‌بندی آناتومیک حرکات چشم و دسته‌بندی رفتاری حرکات چشم. چشم به‌واسطه اتصال به شش عضله، مجموعاً در ده جهت حرکات آناتومی دارد. حرکات آناتومی چشم در سه بخش کلی طبقه‌بندی می‌شوند:

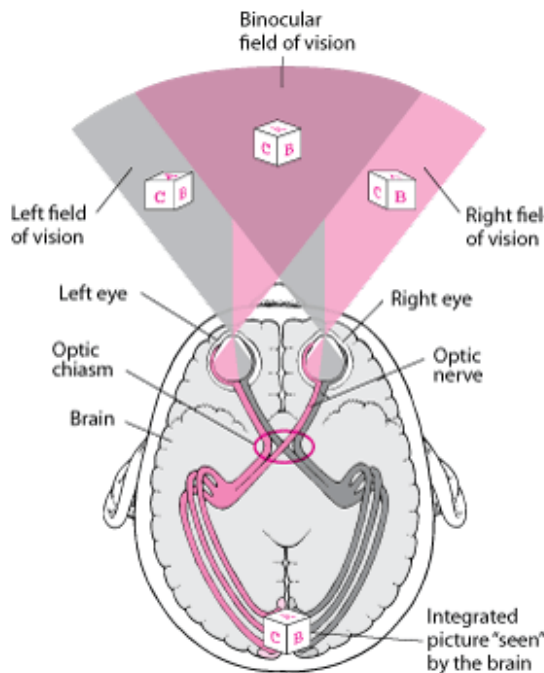
الف) حرکاتی که شامل یک چشم می‌شود. (ب) حرکاتی که هر دو چشم به‌طور موازی انجام می‌دهند و (ج) حرکاتی که دو چشم در خلاف یکدیگر انجام می‌دهند.

در بررسی رفتاری، حرکات چشم به دو نوع ارادی و غیرارادی تقسیم می‌شوند. حرکات ارادی چشم عبارت‌اند از:

ساکید^۳، که حرکتی تند و پرش‌گونه است و سرعت آن تا حداکثر ۱۰۰ درجه بر ثانیه می‌رسد.
دنبال کردن آرام^۴، که حرکتی است پیچیده و فقط زمانی ممکن است که هدفی برای تعقیب وجود داشته باشد.
خیره شدن^۵، که در این حالت چشم به نقطه‌ای در محیط خیره می‌شود [۲].

در اکثر پژوهش‌های به‌عمل‌آمده در زمینه ردیابی چشم، ردیابی دوبعدی چشم مورد مطالعه قرار گرفته است که تنها دو بعد X-Y دید کاربر را تشخیص می‌دهند و بازخوانی اطلاعات بعد سوم خیرگی یعنی عمق نگاه کاربر انجام نگرفته است. در صورتی که درک ما از محیط اطرافمان به‌صورت سه‌بعدی است و در بسیاری از کاربردها تشخیص عمق دید کاربر از اهمیت بسزایی برخوردار است. این پژوهش معطوف به طراحی یک سیستم ردیاب سه‌بعدی چشم با استفاده از پایش اندازه مردمک و محل قرارگیری آن نسبت به سایر بخش‌های چشم به علت اثر انطباق^۶ چشم برای درک عمق و دید سه‌بعدی محیط و ارائه روشی نوین برای ردیابی چشم با قابلیت تشخیص عمق نگاه کاربر به کمک الگوریتم‌های یادگیری عمیق است.

در این مقاله یک روش نوین برای تخمین عمق نگاه ناظر^۷، ارائه شده است که با استفاده از ردیابی سه‌بعدی چشم بر اساس ردیابی توأمان هر دو چشم با روش‌های پردازش تصویر و یادگیری عمیق، با دقت بالایی می‌تواند عمق نگاه ناظر را تشخیص دهد و مسافت قابل تشخیص عمق دید نسبت به کارهای قبلی افزایش قابل توجهی دارد. برای این کار از تغییرات فیزیولوژیکی چشم از جمله محل قرارگیری مردمک چشم^۸، اندازه مردمک، فاصله مردمک دو چشم، تغییر محل نقاط انعکاسی قرنیه^۹ در اثر تغییرات فاصله هدف از چشم و میزان شدت روشنایی محیط و همچنین پدیده‌هایی مانند اختلاف منظر^{۱۰} و تطابق که مغز برای تشخیص عمق از آن بهره می‌گیرد، برای تشخیص عمق نگاه ناظر استفاده شده است. برای نیل به اهداف مطالعه، برای اولین بار یک مجموعه داده دقیق شامل تعداد زیادی تصویر از چشم افراد با شرایط سنی و جنسی متنوع در حال نگاه به یک هدف متحرک در فاصله‌ها و شرایط روشنایی متفاوت تهیه شد. از روش‌های مختلف پردازش تصویر^{۱۱} و همچنین یادگیری عمیق^{۱۲} برای استخراج نواحی مختلف چشم از جمله مردمک، عنبیه^{۱۳}،



شکل ۱- نحوه تشکیل تصویر سه بعدی در شبکه چشم انسان

کلمات، یک سری حرکات پرشی تند به نام ساکید و نقاط مکث یا فیکسیشن دیده می شود [۴]. در سال ۱۹۰۱ میلادی دوج و کلاین سرعت حرکات چشم را بررسی کردند و یک سیستم ردیاب چشم نسبتاً دقیق و غیرتهاجمی را که بر اساس انعکاس قرینه کار می کرد، پایه ریزی کردند. این سیستم که فتوکرونوگراف نام داشت، تنها حرکات افقی چشم را ثبت می کرد [۵] چهار سال بعد فیزیولوژیست آمریکایی به نام چارلز جود توانست این سیستم را بهبود بخشیده و در هر دو جهت افقی و عمودی حرکات چشم را ثبت کند [۶]. در سال ۱۹۰۸ میلادی ادموند هیوی با استفاده از روشی کاملاً تنهاجمی و با کمک یک لنز تماسی، آزمایش هایی جهت اندازه گیری حرکات چشم انجام داد. در وسط این لنز سوراخی برای مردمک وجود داشت و یک اشاره گر آلومینیومی به آن متصل بود که با حرکت کره چشم این اشاره گر نیز حرکت می کرد [۷]. در دهه ۹۰ با قدرتمند شدن سیستم های کامپیوتری امکان ردیابی بلادرنگ چشم میسر شد و کاربرد سیستم های ویدیویی ردیاب چشم در بحث تعامل انسان- کامپیوتر ممکن شد. از سال ۲۰۰۰ به بعد با متحول شدن فن آوری های ردیابی چشم، کاربردهای آن در بسیاری از حوزه های زندگی مردم مانند سرویس هایی نظیر جستجوی اطلاعات و خرید تحت وب، کنترل سیستم های ناوبری هواپیما یا خودرو، سیستم های کمک رسان به معلولین و ... گسترش یافت.

در سال های اخیر روش های مختلفی برای ردیابی چشم بر اساس بینایی ماشین و پردازش تصویر مطرح شده است. در این روش معمولاً یک دوربین برای فوکوس بر روی یک یا دو چشم تنظیم می شود و حرکات چشم را ثبت می کند. یکی از روش های ردیابی چشم، ردیابی چشم با دوربین نصب شده روی سر است که در آن دوربین بر روی یک عینک نصب شده و تصاویر ویدیویی چشم را به سیستم ردیابی چشم ارسال می

صلیبه^{۱۴} و نقاط انعکاسی قرینه استفاده گردید. در ادامه از دو روش عددی و استفاده از شبکه عصبی^{۱۵} مبتنی بر یادگیری عمیق برای تخمین عمق نگاه ناظر استفاده گردید. نتایج حاصله نشان داد که ایده به کار رفته و همچنین تکنیک های مورد استفاده بسیار مؤثر بوده و در هر دو روش، هم دقت تشخیص عمق و هم حداکثر عمق قابل تشخیص، نسبت به مطالعات قبلی به مراتب افزایش یافت.

۲- مفاهیم اولیه

۲-۱- ردیابی چشم

مفهوم ردیابی چشم که در این پژوهش استفاده شده به معنی تخمین زدن جهت و محل نگاه کردن شخص (خیرگی یا زل زدن) کاربر است. در بیشتر موارد تشخیص جهت خیرگی به معنی تعیین شیئی است که نگاه شخص روی آن خیره شده است. به بیان دیگر ردیابی چشم فرایند اندازه گیری نقطه خیره شدن یا حرکت چشم نسبت به سر است که اطلاعات جمع آوری شده نظیر موقعیت چشم و حرکت چشم، اندازه مردمک چشم و سرعت تغییرات اندازه مردمک، تغییرات اندازه خود چشم و ... می تواند الگوی حرکتی چشم را به ما بدهد.

۲-۲- اختلاف شبکه ای

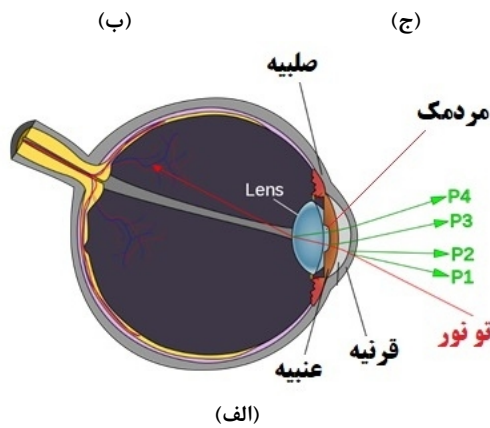
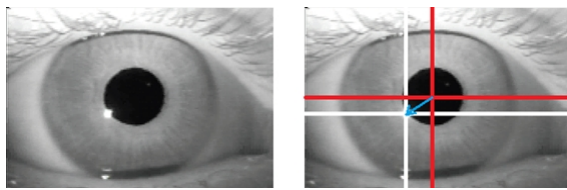
محیط فیزیکی اطراف ما، یک محیط سه بعدی است. اما مغز انسان چطور قادر به درک محیط سه بعدی اطراف خود است؟ مرکز چشم های هر انسان بالغ در حدود ۶/۵ سانتیمتر از یکدیگر فاصله دارند، بنابراین به محیط اطراف از دو زاویه افقی کمی متفاوت نگاه می کند. هنگامی که ما به اطراف نگاه می کنیم، به علت تفاوت در زاویه هر چشم با شیء مورد مشاهده، در تصویر حاصله در شبکه هر چشم با چشم دیگر تفاوت جزئی وجود دارد که به آن اختلاف شبکه ای^{۱۶} گفته می شود. شبکه هر چشم تصویری دوبعدی از محیط اطراف ضبط و برای مغز ارسال می کند. مغز، این دو تصویر دوبعدی را طی فرآیندی ادغام و محیط اطراف را به شکل سه بعدی درک می کند (شکل ۱). بنابراین انسان دارای قابلیت تشخیص عمق می شود؛ قابلیتی که به ما اجازه می دهد اجسام دور را از اجسام نزدیک تشخیص دهیم و درباره فواصل اشیاء قضاوت کنیم [۳].

۲-۳- تطابق

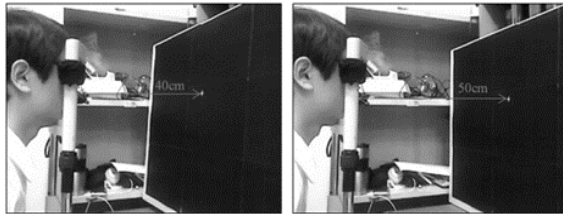
هنگامی که یک فرد نگاه خود را از شیئی که در نزدیکی وی قرار دارد متوجه چیزی که دورتر است می کند، مردمک چشم گشاد می شود و در زمان تماشای اهداف نزدیک تر، مردمک چشم تنگ تر می شود که این عمل را تطابق می گویند. درواقع، عدسی چشم به کمک انقباض و انبساط جسم مژگانی تغییر شکل می دهد و تحذب خود را به میزان مورد نیاز کم و زیاد کرده و این تغییر شکل عدسی باعث می شود که تصویر اجسام باوجود حرکت آن ها و دور شدن از چشم یا نزدیک شدن به آن، روی شبکه چشم بیافتد. این تصویر موجب تحریک یاخسته های بینایی می شود و موج عصبی از طریق عصب بینایی به مغز می رسد و در آنجا درک می شود [۳].

۳- مروری بر کارهای پیشین

در سال ۱۸۷۹ میلادی لوئیس جاوال، چشم پزشک فرانسوی، دریافت که در فرایند خواندن یک متن، در چشم بجای یک حرکت نرم روی



شکل ۲- الف- بازتاب نور از کره چشم ب- تشکیل نقطه پورکینیای اصلی بر روی قرنیه چشم ج- تعیین بردار مرکز قرنیه-پورکینیا [۱۶]



شکل ۳- نحوه نگاه ناظر به اهداف در مانیتور [۱۷]

ارتباط بین تغییرات قطر مردمک و شدت روشنایی در حالت مشاهده دوبعدی به صورت گراف ارائه شده است (شکل ۴-الف). در ادامه، ارتباط بین تغییرات قطر مردمک (d_a) و تغییرات عمق دید در هنگام تماشای همان فیلم در حالت سه بعدی بررسی شده و یک تقریب خطی بین آن‌ها زده شده است. برای به دست آوردن تأثیر تغییرات عمق دید بر روی قطر مردمک و حذف اثر تغییرات شدت روشنایی، تغییرات قطر مردمک در اثر تغییرات روشنایی که در حالت تماشای دوبعدی فیلم به دست آمده بود از نتایج جدید کسر شده و با رگرسیون خطی تنها رابطه بین قطر مردمک و عمق دید به دست آمده است. (شکل ۴-ب)

جی وو لی و همکارانش در [۱۷] روش جالبی را برای تعیین عمق دید ناظر پیشنهاد داده‌اند. آن‌ها علاوه بر استفاده از اصل تطابق چشم و تغییر اندازه مردمک چشم هنگام تغییر عمق دید از این تئوری بهره برده‌اند که به هنگام دیدن اهداف دور، به علت تغییرات ضخامت لنز چشم، فاصله پورکینیای اول و پورکینیای چهارم زیاد شده و برای اهداف نزدیک‌تر این فاصله کم می‌شود. درواقع برای اهداف نزدیک، ضخامت لنز چشم افزایش یافته و باعث کاهش فاصله پورکینیای اول و چهارم می‌شود. در شکل ۶ نحوه تشکیل نقاط انعکاسی و تغییرات فاصله پورکینیای اول و چهارم در اثر تغییر عمق دید نشان داده شده است. در این مطالعه از پانزده نفر

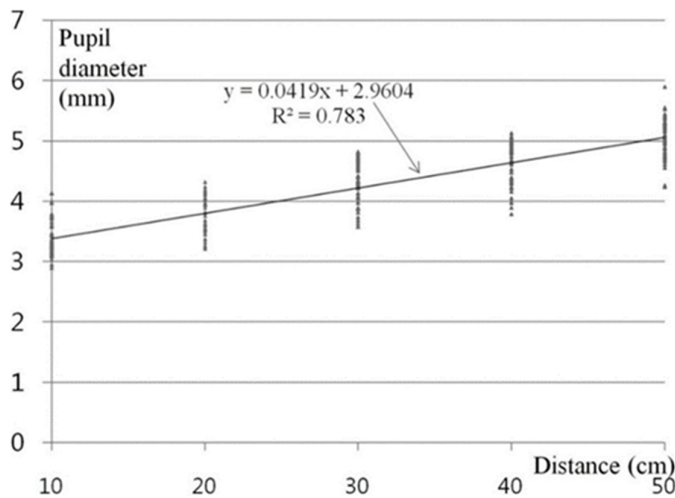
کند. در مرجع [۸] برای جداسازی تصویر چهره از کل صحنه از تبدیل هاف استفاده شده است. نویسندگان در [۹] برای تشخیص چشم از الگوی شبه هار استفاده کرده‌اند. در [۱۰] برای بخش‌بندی تصویر چشم از الگوریتم انتقال به میانگین استفاده شده و برای تشخیص موقعیت مردمک از فیلتر کالمن استفاده شده است. چن و کوبو در [۱۱] روشی ارائه کرده‌اند که در آن دنباله‌ای از یک روش تشخیص چهره و فیلترهای گابور برای ردیابی چشم استفاده شده است. نواحی با پتانسیل وجود چهره بر اساس رنگ پوست آشکارسازی شده‌اند. سپس ناحیه کاندید چشم بر اساس ساختار شکل چهره تعیین شده است. در ادامه چهار فیلتر گابور در جهت‌های مختلف $(0, \pi/4, \pi/2, 3\pi/4)$ به ناحیه موردنظر اعمال می‌شوند و با استفاده از ترکیب پاسخ چهار فیلتر گابور مردمک چشم آشکارسازی می‌شود. کوزجکو و همکارانش در [۱۲] الگوریتم آشکارسازی محل تلاقی طولانی‌ترین خط افقی و عمودی را برای تعیین موقعیت مردمک چشم پیشنهاد کرده‌اند. این الگوریتم با فرض دایروی بودن مردمک چشم، محل تقاطع طولانی‌ترین محور افقی و عمودی مردمک را به عنوان مرکز مردمک چشم شناسایی می‌کند. خیرالفیض و نورعینی در [۱۳] سیستم ردیاب چشم بر اساس مدل ریاضی تبدیل هاف دایروی ارائه کرده‌اند. پرانیس و اسریکانس در [۱۴] روشی را ارائه کرده‌اند که مرز مردمک و قرنیه را با استفاده از تبدیل هاف دایروی تشخیص می‌دهد.

در [۱۵] نویسندگان با مدل‌های ریاضی روابط بین مرکز مردمک ممکن و کلیه گرادیان‌های تصویر را ارائه کرده‌اند. آن‌ها برای پیدا کردن گرادیان‌های تصویر از مشتقات جزئی استفاده نموده‌اند. ردیابی چشم بر اساس نقاط انعکاس قرنیه از روش‌های جدید رهگیری مردمک چشم است. نور بازتابیده از شبکه چشم بر روی قرنیه چشم نقاط روشنی ایجاد می‌کند که در بسیاری از سیستم‌های ردیابی چشم از این نقاط انعکاسی که به آن‌ها تصاویر پورکینیای^{۱۸} گفته می‌شود، برای تشخیص نقطه خیرگی استفاده می‌شود. در ساده‌ترین حالت، از برداری که نقطه مرکز مردمک را به پورکینیای اصلی وصل می‌کند جهت تعیین مسیر نگاه چشم استفاده می‌شود (شکل ۲). با توجه به ساختار چشم، در اثر انعکاس نور در قرنیه و عدسی چشم در حقیقت چهار تصویر پورکینیای ایجاد می‌شود که در حالت پیشرفته‌تر از موقعیت نسبی این تصاویر انعکاسی نسبت به مرکز مردمک و فاصله آن‌ها نسبت به هم برای تعیین مسیر خیرگی و تشخیص عمق نگاه ناظر استفاده می‌شود (شکل ۵)

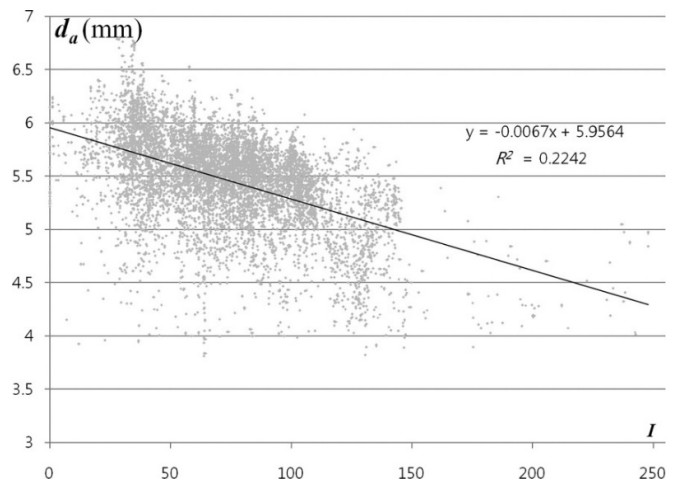
در [۱۶] اوبی چول لی و همکارانش به مطالعه دقیق تأثیر تغییرات روشنایی محیط و عمق دید بر روی اندازه مردمک پرداخته‌اند. روش کار آن‌ها به این شکل بوده است که از تعداد ۲۴ نفر داوطلب (۱۲ مرد و ۱۲ زن با میانگین سنی ۲۶ سال) خواسته شده که یک فیلم ویدئویی سه بعدی را که از طریق یک مانیتور پخش می‌شود، یک بار در حالت دو دوبعدی و یک بار در حالت سه بعدی تماشا کنند. آزمایش‌ها در یک اتاق بسته و با روشنایی ثابت انجام شده است. هنگام تماشای فیلم در حالت دوبعدی که تنها میزان روشنایی تصویر تغییر می‌کند تغییرات اندازه مردمک چشم اندازه‌گیری و ثبت شده است. (شکل ۳)

در این تحقیق بسیار نزدیک به واقعیت حالت چشم در زوایای مختلف خیرگی هست. چرا که عنبیه و مردمک چشم در وضعیت‌های مختلف، شکل دایروی خود را از دست می‌دهند و همچنین بخشی از عنبیه و گاهی مردمک در زیر پلک و مژه‌ها مخفی می‌شوند و روش‌هایی که با فرض دایروی بودن عنبیه و مردمک چشم، از فیلترهایی مثل لبه‌یابی و گابور و تبدیل هاف برای شناسایی عنبیه و مردمک استفاده می‌کنند قطعاً با خطا مواجه خواهند بود. (شکل ۷).

داوطلب با سن و جنس مختلف، خواسته شده به هدفی بر روی مانیتوری که در فواصل ۱۰، ۲۰، ۳۰، ۴۰ و ۵۰ سانتی‌متری از چشم آن‌ها قرار می‌گیرد، نگاه کرده و روی هدف خیره شوند. نتایج به‌دست‌آمده حاکی از خطای متوسط ۸ سانتی‌متر در تشخیص عمق نگاه ناظرین است و هرچه هدف دورتر می‌شود این خطا بیشتر و بیشتر می‌شود. نیکولا پوپوویچ و همکارانش در [۱۸] از طریق ماسک کردن و بخش‌بندی اجزای مختلف چشم از یک مجموعه داده و با استفاده از یک شبکه عصبی کانولوشنی^{۱۷} چشم استفاده کرده‌اند و به نتایج دقیقی دست یافته‌اند. روش به‌کاررفته

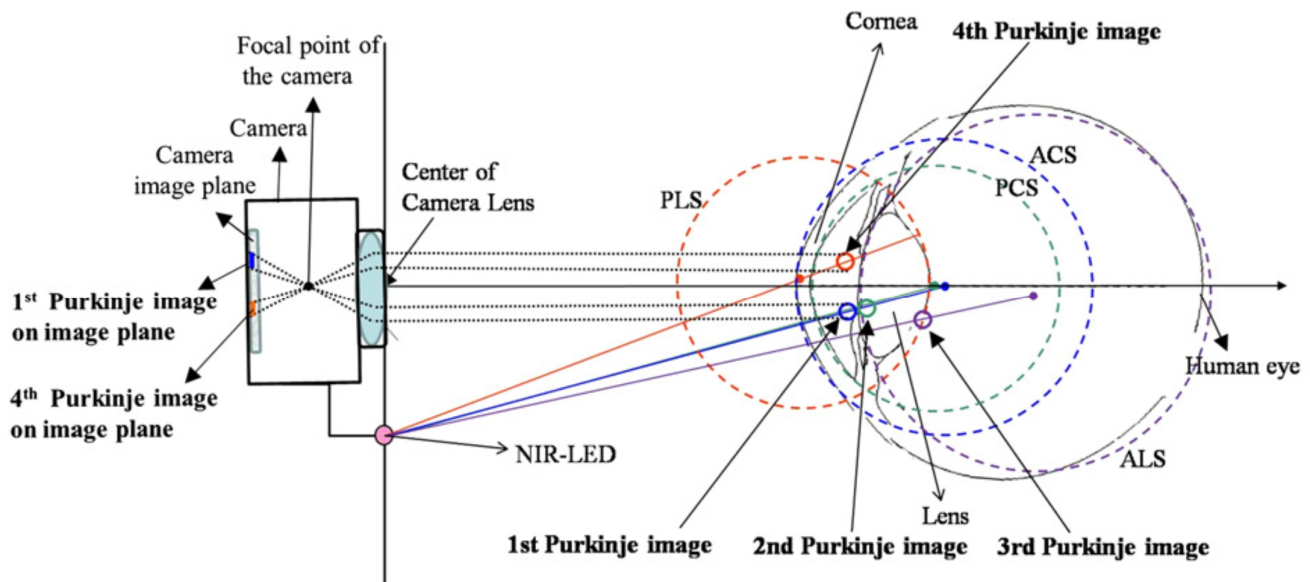


(ب)

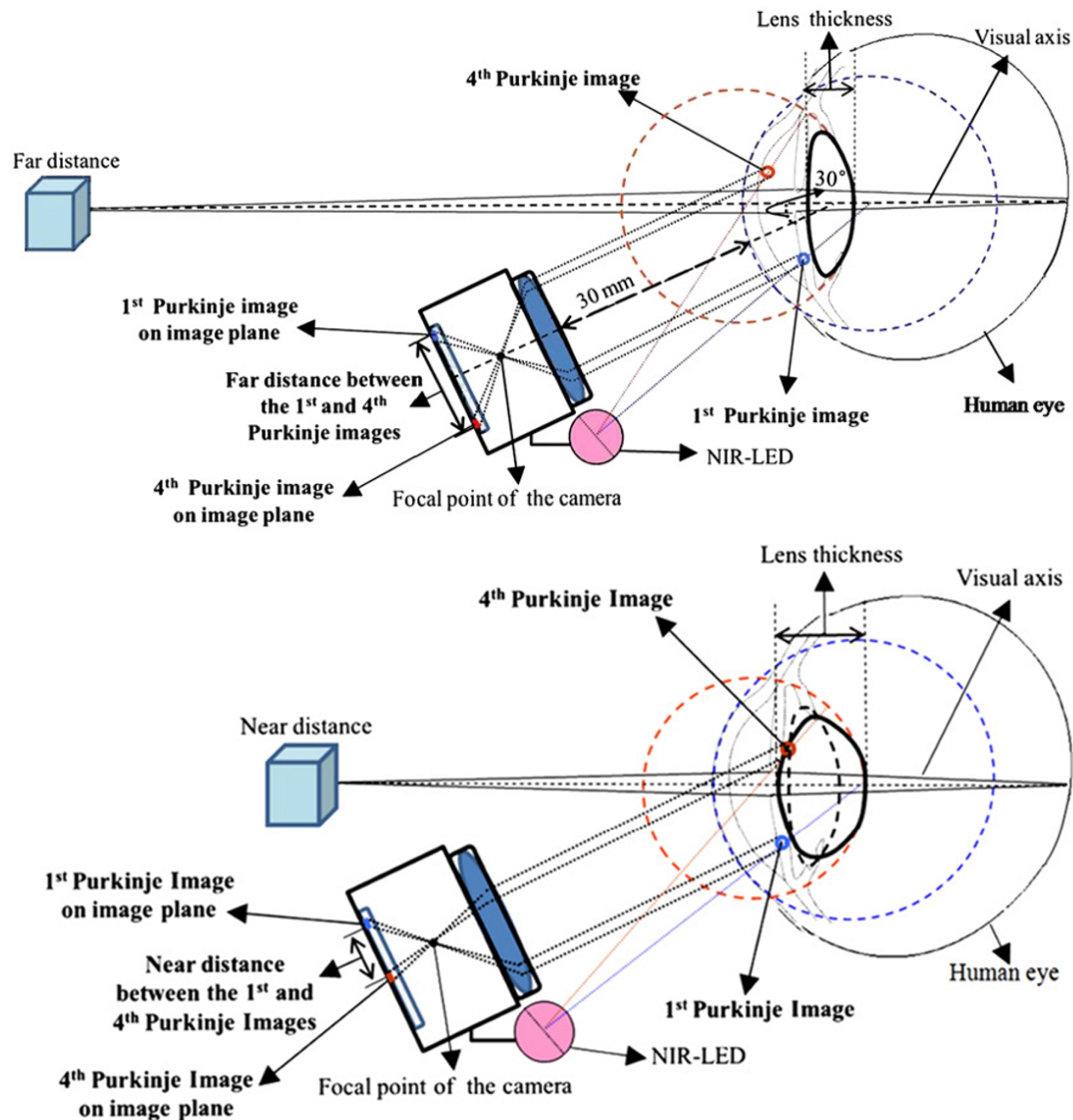


(الف)

شکل ۴-الف- گراف و رگرسیون خطی ارتباط بین اندازه قطر مردمک (d_a) و شدت روشنایی متوسط تصویر (I) در حالت تصویر دوبعدی
ب- گراف رابطه بین تغییرات قطر مردمک و فاصله چشم از هدف [۱۷]



شکل ۵- نحوه تشکیل نقاط انعکاسی اول، دوم، سوم و چهارم قرنیه (پورکینیا)، در اثر بازتابش نور [۱۶]



شکل ۶- تغییر فاصله پورکینیای اول و چهارم در اثر تغییر عمق نگاه (تصویر بالا برای جسم دور و تصویر پایین برای جسم نزدیک)

۴- مواد و روش‌ها

۴-۱- مجموعه داده

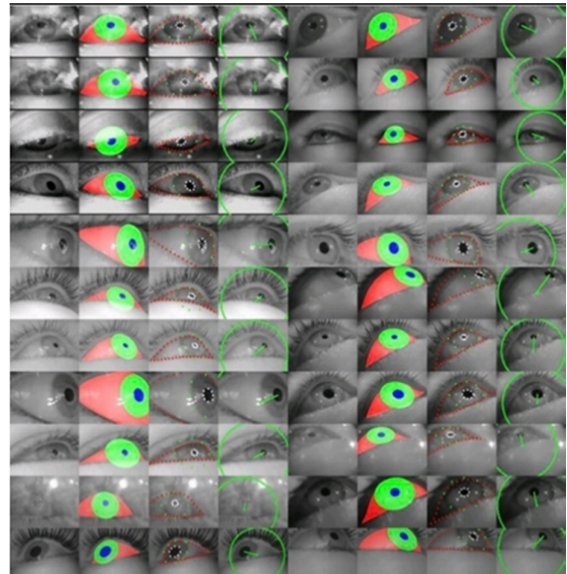
و در آن تعداد زیادی تصویر از چشم افراد با شرایط سنی و جنسی متنوع در حال نگاه به یک هدف متحرک در فاصله‌ها و شرایط روشنایی متفاوت ثبت شد. تهیه این مجموعه داده یکی از چالش‌برانگیزترین مراحل تحقیق بود. برای تصویربرداری از چشم افراد، از یک عینک موجود در پژوهشگاه علوم شناختی آزمایشگاه مرکزی دانشگاه تبریز که دارای دو دوربین با قابلیت تصویربرداری دوچشمی است، استفاده شد. سنسور این دوربین‌ها با نور نزدیک به مادون قرمز تحریک می‌شوند و تفکیک‌پذیری ویدئوی تصویربرداری شده توسط این دوربین‌ها 1080×1920 پیکسل با نرخ ۲۵ فریم در ثانیه است. مشخصات ظاهری دوربین مذکور و نحوه اتصال آن به سر در شکل ۸ نشان داده شده است.

بررسی تحقیقات گذشته نشان داد که باتوجه به کثرت پارامترهای تأثیرگذار در تشخیص عمق نگاه ناظر و همچنین متنوع بودن شرایط تهیه مجموعه داده از جمله نور محیط، سن، جنس، رنگ چشم و شرایط روحی و روانی ناظر، برای دستیابی به نتایج دقیق و پایدار ردیابی چشم با استفاده از تکنیک‌های مناسب پردازش تصویر و یادگیری عمیق، چنانکه بتوان از آن در محیط‌های واقعی استفاده کرد، نیاز به یک مجموعه داده با کیفیت داریم. باتوجه به اینکه تاکنون مجموعه داده‌ای از تصاویر چشم که برای تخمین عمق نگاه ناظر تهیه شده باشد و علاوه بر آن همزمان اطلاعات شدت روشنایی محیط نیز بر روی تصویر ثبت شده باشد، تهیه نشده بود، برای نیل به اهداف این تحقیق برای اولین بار یک مجموعه داده تهیه شد

متری هدف تا چشم، در هر ۵ سانتی متر، از ۵۰ سانتی متر تا ۱ متر در هر ۱۰ سانتی متر و از فاصله ۱ متر تا ۵ متر در هر ۵۰ سانتی متر از چشم افراد تصویربرداری انجام گرفت و در مجموع حدود ۲۵۰۰ تصویر از چشم داوطلبین تهیه شد. یکی از چالش‌های اساسی در تهیه مجموعه داده در مراحل اولیه، کسب اطمینان از این موضوع بود که آیا ناظر دقیقاً به هدف مورد نظر نگاه می‌کند؟ و یا در لحظه تصویربرداری به نقطه دیگری خیره می‌شود؟ متوجه شدیم که در مواردی به علت خستگی ناشی از نمونه‌برداری چندانسته و یا پرش‌های غیرارادی چشم، فرد در لحظه تصویربرداری به نقطه‌ای غیر از هدف نگاه می‌کند و تصویر اشتباهی ثبت می‌شود. برای رفع این مشکل از دو تکنیک مؤثر برای افزایش سرعت و دقت تصویربرداری استفاده شد. برای افزایش دقت نمونه‌برداری یک برگه کوچک دارای متن بر روی یک توری مشبک که جهت افزایش تمرکز و کاهش حرکات غیرارادی چشم استفاده می‌شود، نصب شد و در حین نمونه‌برداری از افراد خواسته شد که متن روی برگه را بخوانند و در حین خواندن متن عمل تصویربرداری انجام شد. علاوه بر آن لازم بود که برای استفاده از تصاویر در مراحل بعدی تحقیق، اطلاعات دقیق فرد مورد مطالعه، شدت روشنایی محیط و فاصله هدف تا چشم فرد ذخیره می‌شد. در حالت عادی بعد از ثبت هر تصویر این اطلاعات در نام تصویر ثبت و ذخیره می‌شد که این کار زمان‌بر می‌شد. برای رفع این چالش یک نرم‌افزار GUI با زبان برنامه‌نویسی پایتون تهیه شد که از طریق کامپیوتر به دوربین وصل شده و به صورت بلادرنگ تصویر دوربین را در کامپیوتر نشان داده و اطلاعات نام داوطلب، جنس، سن، تاریخ نمونه‌برداری، شدت روشنایی و فاصله هدف تا چشم در آن ثبت می‌شد و تنها با کلیک بر روی موس عملیات تصویربرداری انجام شده و هم‌زمان تصویر با نام با اطلاعات فوق در دایرکتوری مربوط به فرد ذخیره می‌شد. در یک شدت روشنایی ثابت در هر بار تصویربرداری فقط اطلاعات فاصله هدف در GUI تغییر داده می‌شد و کار تصویربرداری و ذخیره آن با دقت و سرعت زیادی انجام می‌شد. (شکل ۱۰)



شکل ۹- نحوه تصویربرداری از چشم افراد و تهیه مجموعه داده



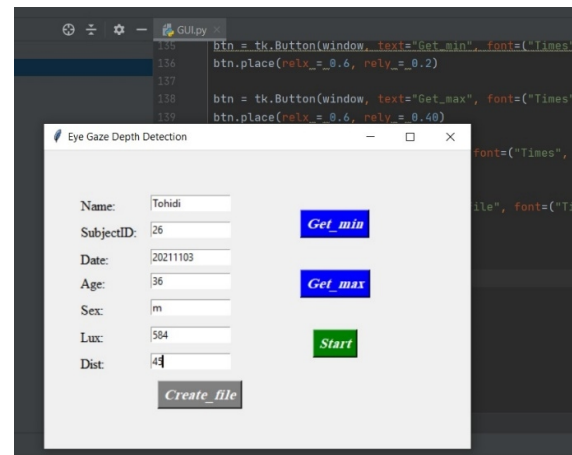
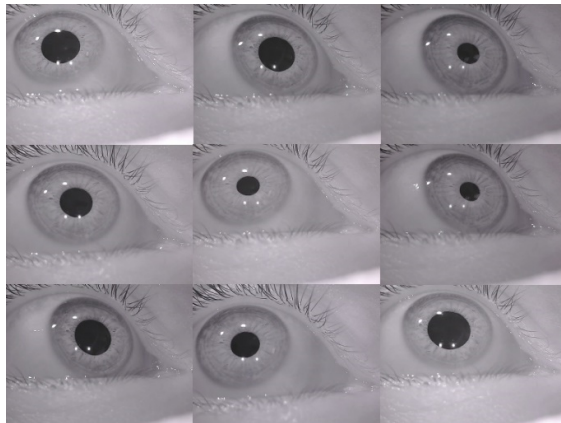
شکل ۷- بخش بندی اجزای چشم و تخمین سه بعدی خیرگی [۱۸]



شکل ۸- شکل ظاهری دوربین مورد استفاده برای تهیه مجموعه داده و نحوه اتصال به سر ناظر

برای تهیه مجموعه داده از چشم‌های ۲۰ نفر زن و مرد با سنین ۱۸ تا ۴۵ سال توسط دوربین‌ها تصویربرداری انجام شد. در حین تصویربرداری از یک دستگاه لوکس‌متر برای اندازه‌گیری شدت روشنایی رسیده به چشم، استفاده شد. همان‌طور که در شکل ۹ نشان داده شده است، دستگاه لوکس-متر در جلوی سینه افراد قرار گرفته و شدت روشنایی محیط در لحظه تصویربرداری از چشم‌های فرد ثبت گردید. در هنگام تصویربرداری از هر فرد با روشن و خاموش کردن چراغ‌های اتاق و باز و بسته کردن پرده‌های ضخیم پنجره‌ها، ۶ ناحیه روشنایی از محدوده ۵۰ لوکس تا ۲۰۰۰ لوکس ایجاد شد. به علت تفاوت زمان‌های نمونه‌برداری در طول روز و همچنین ابری یا غیر ابری بودن آسمان، سطوح روشنایی تصویربرداری افراد تغییرات اندکی با همدیگر داشتند که در محاسبات و تحلیل نهایی شدت روشنایی نرمالیزه شده نسبت به روشنایی بیشینه در نظر گرفته شد و این تغییرات جزئی لحاظ شدند. در هر شدت روشنایی یک هدف از فاصله ۱۰ سانتی متری چشم تا ۵ متری آن حرکت داده شده و با فوکوس کردن نگاه فرد به هدف، از چشم‌های وی تصویربرداری صورت گرفت. تا فاصله ۵۰ سانتی

پیش‌پردازش‌های اولیه روی تصاویر شامل grayscale کردن تصویر و تغییر و تنظیم شدت روشنایی تصویر (imadjust)، اعمال فیلتر median و histogram echolization و در نهایت اعمال تابع contour، ۱۰ لایه و اعمال معیارهایی جهت استخراج مردمک چشم از بین کاندیدهای حاصل از خروجی تابع contour انجام می‌شود. نواحی چشم و عنبیه و مردمک چشم استخراج می‌شوند و قطر مردمک و فاصله دو مردمک اندازه‌گیری شده و با توجه به تاثیر میزان شدت روشنایی محیط در اندازه مردمک، در هر تصویر اطلاعات روشنایی نیز ثبت می‌شود و از طریق تحلیل عددی مجموعه اطلاعات به دست‌آمده، تخمین عمق نگاه ناظر انجام می‌گیرد که در ادامه جزئیات این روش ارائه شده است.

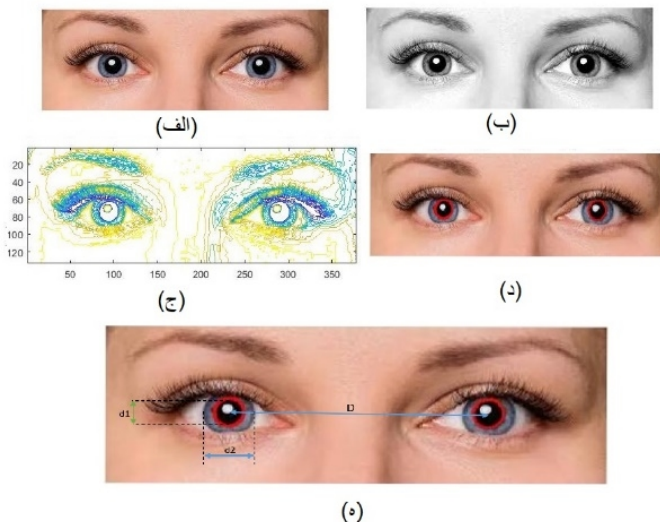


شکل ۱۰- GUI تهیه شده برای عملیات تصویر برداری از چشم

۴-۲- تشخیص عمق نگاه ناظر با استفاده از تحلیل عددی نتایج حاصل از استخراج ویژگی‌های چشم

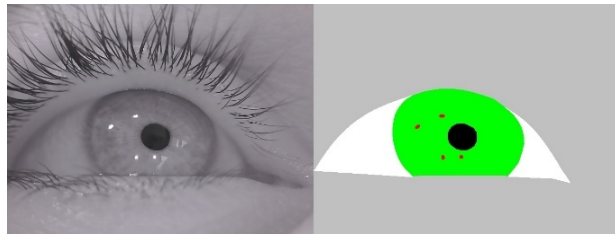
مغز و چشم انسان طی یک فرایند طبیعی برای زوم کردن و تشخیص عمق هدف از تغییر فاصله مردمک دو چشم استفاده می‌کنند، به‌طوری که هنگامی که شخص به یک هدف در فاصله بی‌نهایت نگاه می‌کند، مرکز هر دو مردمک چشم در مراکز کره چشم قرار می‌گیرند و با نزدیک‌تر شدن هدف به چشم، مردمک دو چشم به یکدیگر نزدیک و نزدیک‌تر می‌شوند و به این طریق تشخیص عمق و مشاهده سه‌بعدی میسر می‌شود. بی‌نهایت موردنظر در سیستم بینایی انسان در حدود ۵ متر است. علاوه بر آن هنگامی که فرد نگاه خود را از شیئی که در نزدیکی آن قرار دارد متوجه چیزی که دورتر است می‌کند، مردمک چشم گشاد می‌شود و در زمان تماشای اهداف نزدیک‌تر، مردمک چشم تنگ‌تر می‌شود. در این تحقیق از همین ویژگی‌های ساختار فیزیولوژیکی چشم و مغز انسان برای تشخیص عمق و دید سه‌بعدی استفاده شده است. به‌طور دقیق‌تر با ره‌گیری برخط تغییرات اندازه مردمک چشم و فاصله مردمک دو چشم با دوربین، عمق دید ناظر تخمین زده می‌شود. تا به حال مطالعات بسیار محدودی برای تشخیص عمق نگاه با استفاده از ردیابی چشم صورت گرفته است که اولاً در اکثر آن‌ها فقط تغییرات اندازه مردمک موردتوجه قرار گرفته است و ثانیاً به دلیل عدم لحاظ تأثیر تغییرات شدت روشنایی محیط و سایر عوامل ایجاد خطا معمولاً از دقت پایینی برخوردار بوده‌اند. ایده استفاده از افزودن ره‌گیری تغییرات فاصله مردمک دو چشم برای تشخیص عمق نگاه، برای اولین بار در این تحقیق به کار رفته است. در شکل ۱۱ تعدادی از تصاویر چشم یک فرد در حین نگاه به یک هدف متحرک در راستای مستقیم چشم نشان داده شده است. همان‌طور که از تصاویر پیداست اندازه و محل قرارگیری مردمک چشم کاملاً با تغییرات عمق نگاه ناظر ارتباط دارد. روند کلی کار در روش اول به این شرح است که مطابق شکل ۱۲ با تصویر برداری از دو چشم ناظرین در حال خیره شدن به اهداف در فواصل مختلف، پس از ترکیب تصاویر دو دوربین و

شکل ۱۱- تعدادی از تصاویر چشم یک زن ۲۵ ساله در حال نگاه به عمق‌های مختلف در راستای مستقیم



شکل ۱۲- (الف) تصویر اصلی (ب) تصویر سیاه و سفید (ج) پیش پردازش تصویر (د) تشخیص مردمک (ه) استخراج اندازه مردمک و عنبیه و فاصله مردمک دو چشم

انجام شد و این مجموعه داده با نام گذاری حاوی اطلاعات، برای مطالعات بعدی قابل استفاده خواهد بود. بخشی از مجموعه داده در شکل ۱۵ نشان داده شده است.



شکل ۱۴- نمونه ای از بخش بندی و ماسک گذاری چشم



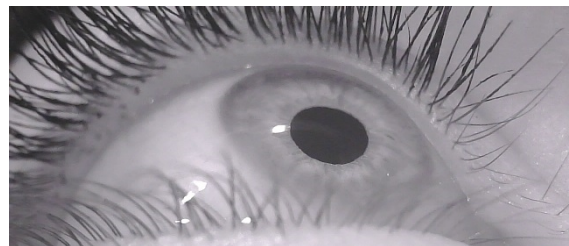
شکل ۱۵- بخشی از مجموعه داده اصلی و ماسک گذاری شده

برای شناسایی نواحی پنج گانه چشم از یک شبکه یادگیری عمیق مبتنی بر الگوریتم U-Net دوبعدی استفاده شد. در میان الگوریتم های یادگیری عمیق، الگوریتم U-Net به دلیل دقت بالا، پاسخ دقیق، سرعت بالای یادگیری و پردازش، و عدم نیاز به سخت افزارهای پیچیده و گران قیمت، به یکی از الگوریتم های قدرتمند شناسایی اجزای تصویر و بخش بندی آنها در پردازش تصاویر پزشکی بدل شده است. معماری U-Net یک شبکه عصبی کانولوشنی است. این شبکه از دو بخش تشکیل شده است؛ بخش اول مسیر فشرده سازی و رمز گذاری و بخش دوم، مسیر رمز گشایی و گسترده سازی است. هدف مسیر فشرده سازی درک مضمون تصویر و نقش مسیر گسترده سازی کمک کردن به فرآیند بخش بندی دقیق اجزای تصویر است.

در این تحقیق حدود ۸۰ درصد از تصاویر ماسک شده (۲۰۰۰ تصویر) برای آموزش شبکه و ۲۰ درصد آن (۵۰۰ تصویر) برای مرحله تست استفاده شدند. ۱۰ درصد از تصاویر مرحله آموزش (۲۰۰ تصویر) برای اعتبارسنجی مدل در نظر گرفته شدند. تابع Binary_crossentropy به عنوان تابع زیان مدل در نظر گرفته شد. با استفاده از زبان برنامه نویسی پایتون مدل شبکه شبیه سازی شد و با ۱۰۰ دوره (epoch) شبکه به آموزش گذاشته شد. تصاویر با ابعاد 256×256 وارد مدل شد و ۵ ناحیه صلیبه، عنیبه، مردمک، نقاط انعکاسی قرینه و پشت زمینه به عنوان خروجی شبکه در نظر گرفته شدند.

۱-۲-۴- استخراج مردمک چشم و مرکز آن و اندازه گیری فاصله بین دو مردمک

تاکنون روش های گوناگونی برای شناسایی و استخراج عنیبه و مردمک چشم معرفی شده است که در اکثر آنها فرض بر این است که عنیبه و مردمک چشم یک شکل دایروی دارند و با استفاده از تکنیک های پردازش تصویر مانند لبه یابی و یا اعمال فیلترهای مختلف، این دایره ها استخراج و به عنوان عنیبه و مردمک معرفی می شوند. اما در عمل شکل مردمک و عنیبه به شکل دایره کامل نیستند و مخصوصاً هنگام نگاه فرد به طرفین و بالا و پایین، عنیبه و مردمک کشیده شده و شکلی مشابه بیضی پیدا می کنند. همچنین در مواقعی بخشی از عنیبه زیر پلک ها می مانند و شکلی خاص به خود می گیرند در شکل ۱۳ مشاهده می شود که این موضوع در مورد تصاویر واقعی از چشم و در شرایط طبیعی صادق است.

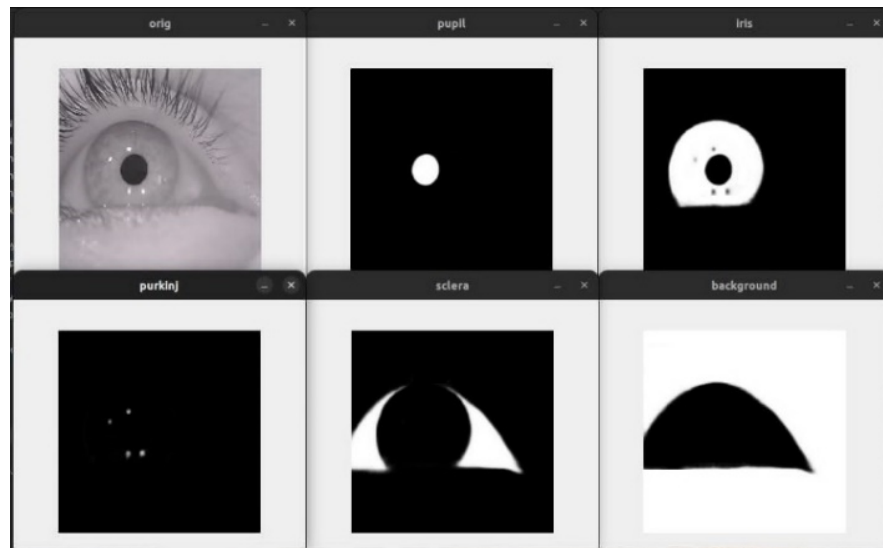


شکل ۱۳- تصویر چشمی که در حال نگاه به هدفی در فاصله نزدیک است و عنیبه و مردمک کشیده شده اند.

در این مطالعه با توجه به لزوم دقت بسیار بالا در تشخیص محل دقیق مردمک و اندازه آن برای تخمین عمق نگاه ناظر از یک شبکه یادگیری عمیق برای تشخیص نواحی مختلف چشم از جمله صلیبه (ناحیه سفید چشم)، عنیبه، مردمک و چهار نقطه انعکاسی (پورکینیا) استفاده شد. در مرحله آموزش شبکه، تمام تصاویر تهیه شده ماسک گذاری شد و به طور دقیق نواحی مذکور با رنگ های منحصر به فردی بخش بندی شدند. در این مرحله یک روش خلاقانه برای لحاظ کردن شدت روشنایی در لحظه تصویربرداری بر روی عکس هایی که به شبکه وارد می شوند، به کار گرفته شد. با توجه به ثبت اطلاعات میزان شدت روشنایی محیط بر روی نام تصویر ذخیره شده، عدد مربوط به شدت روشنایی نسبت به شدت روشنایی پیشینه یعنی ۲۰۰۰ لوکس، نرمالیزه شده و با نسبت به دست آمده، ناحیه بیرونی چشم در تصویر به رنگ خاکستری با میزان روشنایی ۰ تا ۲۵۵ ماسک گذاری شد. در شکل ۱۴ یک نمونه از بخش بندی و ماسک گذاری چشم نشان داده شده است. در این بخش بندی ناحیه صلیبه به رنگ سفید، ناحیه عنیبه به رنگ سبز، ناحیه مردمک به رنگ سیاه، نواحی نقاط انعکاسی به رنگ قرمز و پشت زمینه چشم به رنگ خاکستری متناسب با شدت روشنایی محیط در لحظه تصویربرداری درآمده است. در روش اول تشخیص عمق تنها اطلاعات فاصله دو مردمک و اندازه آن مورد نیاز است و سایر نواحی در روش دوم مورد استفاده قرار گرفته است. برای کل مجموعه داده عملیات بخش بندی و ماسک گذاری

شبکه وارد کنیم، به علت تفاوت ساختار چشم افراد از جمله اندازه مردمک و فاصله دو چشم، ابتدا شعاع مردمک، فاصله بین دو مردمک و همچنین شدت روشنایی، نسبت به بیشترین مقدار این پارامترها برای هر فرد که در زمان تهیه مجموعه داده اندازه گیری شده بودند، نرمالیزه شدند و پس از آن تمامی اطلاعات در یک جدول اکسل کلی جمع آوری شدند. در جدول ۱ بخشی از این جدول نمایش داده شده است. نتایج نشان داد که در حین خیرگی چشم برای تشخیص عمق، شعاع مردمک می تواند تا ۳۰ درصد شعاع بیشینه و فاصله بین دو مردمک تا ۶۷ درصد فاصله بیشینه کاهش یابد که این نتایج یکی از دست آوردهای مهم این تحقیق هستند.

همان طور که در شکل ۱۶ و ۱۷ مشاهده می شود، شبکه با دقت بسیار زیادی موفق به بخش بندی نواحی مورد نظر چشم شده است. در حالت بخش بندی هم زمان ۵ ناحیه، دقت ۹۸ درصد ثبت شد و در حالت تشخیص فقط ناحیه مردمک شبکه با دقت نزدیک به ۱۰۰ درصد موفق به استخراج ناحیه مردمک شد. در ادامه بر اساس اطلاعات به دست آمده از شبکه یادگیری عمیق، شعاع مردمک و فاصله دو مردمک چشم به دست آمد. کلیه اطلاعات تصاویر هر کدام از افراد در یک جدول اکسل، شامل اطلاعات شدت روشنایی، فاصله هدف از چشم، شعاع مردمک و فاصله دو مردمک هر تصویر و شعاع بیشینه مردمک و فاصله بیشینه دو مردمک هر فرد ثبت شد. برای آنکه بتوانیم اطلاعات تمام افراد را در یک



شکل ۱۶- خروجی شبکه عصبی و بخش بندی نواحی ۵ گانه چشم (صلیبیه، عنبیه، مردمک، نقاط انعکاسی قرنیه و پشت زمینه)



شکل ۱۷- روند افزایش دقت و کاهش خطای شبکه عصبی در طول یادگیری و تشخیص نواحی ۵ گانه چشم

جدول ۱- بخشی از جدول اطلاعات به دست آمده از شبکه یادگیری عمیق شامل: اطلاعات شدت روشنایی (Light)، شعاع مردمک (R)، شعاع مردمک نرمالیزه شده، فاصله دو مردمک هر تصویر (L)، فاصله دو مردمک نرمالیزه شده و فاصله هدف از چشم (distance)

Light (Lux)	Light (normalized)	R (mm)	R (normalized)	L (mm)	L (normalized)	Distance (cm)
800	0.333333333	39	0.302325581	294	0.790322581	15
800	0.333333333	40	0.310077519	281	0.755376344	10
600	0.25	44	0.341085271	338	0.908602151	35
330	0.1375	46	0.356589147	351	0.943548387	50
900	0.375	41	0.359649123	252	0.692307692	10
535	0.222916667	49	0.379844961	308	0.827956989	20
600	0.25	49	0.379844961	314	0.844086022	25
992	0.413333333	56	0.380952381	348	0.764835165	10
640	0.266666667	42	0.385321101	294	0.753658537	10
600	0.25	50	0.387596899	339	0.911290323	50
900	0.375	50	0.387596899	341	0.916666667	500
900	0.375	45	0.394736842	268	0.736263736	15
330	0.1375	51	0.395348837	272	0.846774194	25
968	0.403333333	34	0.404761905	321	0.810606061	10
800	0.333333333	53	0.410852713	328	0.88172043	200
992	0.413333333	61	0.414965986	376	0.826373626	20
450	0.1875	54	0.418604651	290	0.779569892	15
600	0.25	54	0.418604651	331	0.889784946	30
330	0.1375	55	0.426356589	328	0.88172043	35
399	0.16625	55	0.426356589	338	0.908602151	500
864	0.36	36	0.428571429	306	0.787878788	10
1000	0.416666667	49	0.4375	398	0.872844828	30
580	0.241666667	48	0.440366972	321	0.782926829	15
420	0.175	48	0.440366972	307	0.768292683	15
450	0.1875	57	0.441860465	282	0.758064516	10
773	0.322083333	65	0.442176871	302	0.843575419	10
1072	0.446666667	65	0.442176871	390	0.89010989	30
1028	0.428333333	65	0.442176871	444	0.975824176	400
1040	0.433333333	65	0.442176871	448	0.984615385	500
980	0.408333333	66	0.448979592	444	0.975824176	300
1000	0.416666667	51	0.455357143	394	0.829741379	10
980	0.408333333	67	0.455782313	444	0.975824176	350
1010	0.420833333	67	0.455782313	442	0.971428571	450
660	0.275	52	0.456140351	259	0.711538462	10
2000	0.916666667	32	0.457142857	249	0.754545455	10
300	0.125	67	0.632075472	206	0.679867987	10
310	0.129166667	50	0.458715596	314	0.765853659	10
662	0.275833333	68	0.462585034	316	0.882681564	15
1000	0.416666667	52	0.464285714	385	0.829741379	15

۵- شبیه سازی و نتایج:

۵-۱- روش اول: تخمین عمق نگاه ناظر با استفاده از

شبکه عصبی RBF

در روش اول این مطالعه برای تخمین عمق نگاه ناظر، از الگوریتم طبقه‌بندی شبکه عصبی پایه شعاعی (RBF^{۱۹}) استفاده شد. شبکه عصبی RBF در حوزه مدل‌سازی ریاضی از توابع پایه‌ای به‌عنوان تابع فعالیت استفاده می‌کند. خروجی این شبکه یک ترکیب خطی از توابع پایه شعاعی برای پارامترهای ورودی و نورون‌ها است. این شبکه‌ها در توابع تقریب، پیش‌بینی سری‌های زمانی، کلاس‌بندی، خوشه‌بندی، تشخیص الگو و تخمین و بهینه‌سازی مورد استفاده قرار می‌گیرند. شبکه‌های RBF معمولاً شامل سه لایه هستند: لایه ورودی، لایه مخفی با یک تابع فعالیت پایه شعاعی غیرخطی و لایه خروجی. ورودی می‌تواند به عنوان یک بردار از اعداد حقیقی مدل شود و خروجی شبکه می‌تواند یک تابع از بردار ورودی باشد که با معادله (۱) محاسبه می‌شود. در این معادله، X بردار ورودی، n تعداد نورون‌های لایه مخفی، G_i بردار مرکز نورون i و W_i وزن نورون i در نورون خروجی و ϕ یک تابع شعاعی غیر خطی مانند تابع گاوسی هست.

$$f(x) = \sum_{i=1}^n W_i \phi(r), \quad r = \|X - G_i\| \quad (1)$$

در این تحقیق از شبکه عصبی RBF برای مدل‌سازی و درون‌یابی خطی از اطلاعات استخراج‌شده از چشم، برای تخمین عمق نگاه ناظر استفاده شد. داده‌های خروجی شبکه U-Net شامل اطلاعات شعاع مردمک، فاصله دو مردمک هر تصویر، شعاع بیشینه مردمک و فاصله بیشینه دو مردمک هر فرد به همراه اطلاعات روشنایی محیط به عنوان ورودی شبکه و فاصله هدف از چشم، بر اساس فاصله‌های مختلف هدف از چشم ناظرین که در حین تهیه مجموعه داده به تعداد ۲۲ فاصله (از ۱۰ سانتی متر تا ۵ متر) در نظر گرفته شده بود، اعداد از ۱ تا ۲۲ به آن اعداد برچسب گذاری شد و به عنوان خروجی شبکه در نظر گرفته شد. ۸۰ درصد داده (اطلاعات ۲۰۰۰ تصویر) برای آموزش شبکه و ۲۰ درصد آن (۵۰۰ تصویر) برای تست شبکه اختصاص داده شد. ۱۰ درصد از تصاویر مرحله آموزش (۲۰۰ تصویر) برای اعتبارسنجی مدل و جلوگیری از بیش‌برازش شبکه عصبی در نظر گرفته شدند. مطابق نمودار شکل ۱۸، شبکه عصبی پس از ۲۰۰ epoch با دقت ۹۰/۴۵ درصد موفق به تخمین عمق نگاه ناظر شد که با توجه به افزایش عمق قابل تشخیص نسبت به کارهای قبلی و همچنین انجام ردیابی چشم در محیط واقعی و با تغییرات روشنایی محیط، نتیجه قابل قبولی می‌تواند باشد و در صورت افزایش تعداد داده ها دقت تشخیص بالاتر خواهد رفت.

۵-۲- روش دوم: تشخیص عمق نگاه ناظر با استفاده از تحلیل

تصاویر مجموعه داده با استفاده از یک شبکه یادگیری عمیق

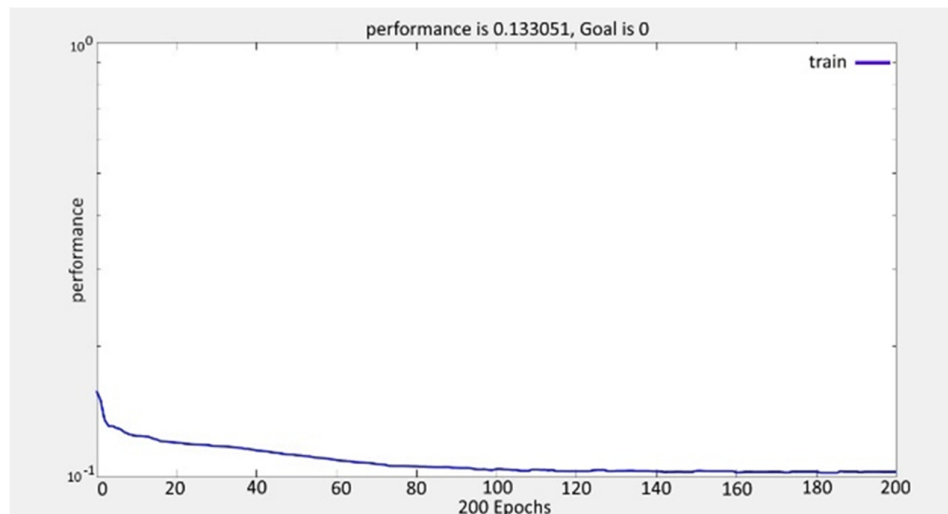
در روش دوم تشخیص عمق نگاه ناظر، کل مراحل تشخیص عمق با استفاده از یک شبکه یادگیری عمیق صورت گرفت. روش کار بدین صورت بود که ابتدا بر روی کلیه تصاویر مجموعه داده، با استفاده از شبکه U-net طبق

روشی که در بخش قبلی توضیح داده شد، عملیات بخش‌بندی اعمال شده و تصاویر برچسب‌گذاری شده با نواحی صلبیه، عنبیه، مردمک، چهار نقطه پورکینیا و پشت زمینه تصویر چشم به یک شبکه عصبی دیگر با مدل MoblieNet-V2 وارد شدند. شبکه MobileNet یکی از مدل‌های شبکه عمیق پیش آموزش‌دیده است که به‌خاطر تعداد پایین پارامترهای شبکه کانولوشنی از محبوبیت بالایی برای کاربردهای یادگیری عمیق برخوردار است. نحوه اعمال تصاویر بخش‌بندی شده به شبکه MobileNet به این صورت است که در لایه خروجی شبکه U-Net، ۵ تصویر مختلف شامل تصویر پشت زمینه، صلبیه (بخش سفید چشم)، عنبیه، مردمک، و پورکینیاها، ابتدا با هم ترکیب شده و شکلی مانند شکل ۱۴ به دست می‌آید و سپس این تصاویر به شبکه MobileNet وارد می‌شود. در واقع ورودی شبکه MobileNet مجموعه‌ای از تصاویر بازسازی شده مانند شکل ۱۴ است. علت اینکه تصاویر اصلی چشم به شبکه وارد نشدند این است که در تصاویر خام چشم، اطلاعات اضافی زیادی از جمله پلک، اشکال و خطوط داخل عنبیه، میزان تیرگی متفاوت عنبیه افراد مختلف و انعکاس روشنایی پنجره و... وجود دارد که هیچ تأثیری از تغییرات عمق نگاه نداشته و علاوه‌برآن موجب پیچیدگی تحلیل شبکه می‌شدند. با توجه به فواصل مختلف هدف از چشم، تصاویر دسته بندی و لیبل زده شدند. به طوریکه هر پوزه دارای یک لیبل حاوی اطلاعات فاصله هدف و شامل تصاویر مربوط به آن فاصله شد. با استفاده از کتابخانه keras در محیط برنامه‌نویسی پایتون شبکه موردنظر مدل‌سازی شد. در شبکه MobileNet خروجی یک ماتریس کلاس بندی شده است با ۲۲ پارامتر که هر کدام دلالت بر یک فاصله هدف (عمق نگاه ناظر) دارد. درواقع ۲۲ دسته تصویر که در هر دسته تصاویر مربوط به یک فاصله هدف از چشم (از ۱۰ سانتی متر تا ۵ متر) قرار داده شده بودند، به‌عنوان خروجی هدف شبکه عمیق در نظر گرفته شده اند. مطابق شکل ۱۹ پس از طی ۱۳۰ epoch آموزش شبکه، دقت شبکه در تشخیص عمق نگاه ناظرین برای تصاویر مورد تست به عدد ۹۳ درصد رسید. در جدول ۲ مقایسه نتایج دو روش مورد استفاده در این مطالعه با نتایج مطالعات قبلی در این زمینه ارائه شده است.

۶- نتیجه گیری و پیشنهادات

در این مقاله از یک ایده نوین برای تشخیص عمق نگاه ناظر از طریق ردیابی چشم استفاده شد. تغییرات برخی مشخصات ظاهری چشم از جمله تغییرات شکل و اندازه مردم و عنبیه چشم، تغییرات فاصله مردمک چشم و تغییر محل نقاط انعکاسی قرنیه در اثر تغییر عمق نگاه موردتوجه قرار گرفت. یک مجموعه داده ۲۵۰۰ تصویری، از تصاویر چشم ۲۰ نفر در حال نگاه به یک هدف متحرک تهیه شد. اطلاعات شدت روشنایی محیط به عنوان یکی از تاثیرگذارترین پارامترها بر روی چشم در حین تهیه تصاویر ثبت و در روی تصاویر ذخیره شد. دو روش برای تخمین عمق نگاه ناظر به کار گرفته شد. در روش اول از تلفیق پردازش تصویر و یادگیری عمیق با تحلیل عددی نتایج حاصل از استخراج ویژگی‌های چشم و شدت روشنایی محیط، برای تشخیص

عمق نگاه ناظر استفاده شد. با استفاده از یک شبکه عصبی با مدل U-Net با دقت ۹۰/۴۵ عمق نگاه ناظر تخمین زده شد. در روش دوم کل مراحل تشخیص عمق نگاه ناظر با استفاده از یادگیری عمیق صورت گرفت. در این روش تصاویر بخش‌بندی‌شده چشم با شبکه U-Net، وارد یک شبکه عصبی دیگری با مدل MoblieNet شد. در این روش نیز با دقت نزدیک به ۹۳ درصد عمق نگاه ناظر تخمین زده شد.



شکل ۱۸- نمودار دقت خروجی شبکه RBF در تشخیص عمق نگاه ناظر



شکل ۱۹- نمودار دقت و خطای خروجی مدل MoblieNet-V2 در تشخیص عمق نگاه ناظر

جدول ۲- مقایسه نتایج دقت روش‌های پیشنهادی مقاله با مطالعات قبلی در تخمین عمق نگاه ناظر

ردیف	روش ارائه شده	حداکثر عمق قابل تشخیص	محیط آزمایش	روشنایی محیط	دقت تشخیص متوسط عمق
۱	Ji Woo lee[16]	۵۰ سانتی متر	مجازی	ثابت	۸۴ درصد
۲	Eui Chul [17]	۵۰ سانتی متر	مجازی و واقعی	متغیر	۸۸ درصد
۳	روش پیشنهادی ۱	۵۰۰ سانتی متر	واقعی	متغیر	۹۰ درصد
۴	روش پیشنهادی ۲	۵۰۰ سانتی متر	واقعی	متغیر	۹۳ درصد

مراجع

- [10] Alex Poole and Linden J. Ball, Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects, IEEE April 2011
- [11] Y. W. Chen and K. Kubo, "A robust eye detection and tracking technique using gabor filters," in Proc. 3rd Int. Conf. IEEE Intell. Inf. Hiding Multimedia Signal Process, vol. 1. Nov. 2007, pp. 109–112.
- [12] T. Koceljko, A. Bujnowski, and J. Wtorek, "Eye mouse for disabled," in Proc. IEEE Conf. Human Syst. Interact., May 2008, pp. 199–202.
- [13] W. Khairoufaizal and A. Nor'aini, "Eye detection in facial images using circular Hough transform," in Proc. 5th CSPA, Mar. 2009, pp. 238–242
- [14] A. Pranitha and C. R. Srikanth, "Iris recognition using corner detection," in Proc. 2nd ICISE, 2010, pp. 2151–2154.
- [15] Fabian Timm and Erhardt Barth, "Accurate Eye Centre Localisation By Means Of Gradients," Institute for Neuro- and Bioinformatics, University of Lubeck, Ratzeburger Allee 160, D-23538 Lubeck, Germany
- [16] Eui Chul Lee; Ji Woo Lee; Kang Ryoung Park, "Experimental Investigations of Pupil Accommodation Factors," Investigative Ophthalmology & Visual Science August 2011, Vol. 52, 6478–6485. doi:10.1167/iovs.10-6423.
- [17] Ji Woo Lee, Chul Woo Cho, Kwang Yong Shin, Kang Ryoung Park, "3D Gaze Tracking Method Using Purkinje Images on Eye Optical Model and Pupil," Elsevier, Optic and Laser Engineering, 50 (2012) 736–751
- [18] Nikola Popovic, Dimitrios Christodoulou, Danda Pani Paudel, Xi Wang, Luc Van Gool, Model-aware 3D Eye Gaze from Weak and Few-shot Supervisions, 2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct
- [1] Kenneth W. Wright, "Anatomy and Physiology of Eye Movements". Pediatric Ophthalmology and Strabismus pp 125 - 143
- [2] Adler, FH; Fliegelman (1934). "Influence of fixation on the visual acuity". Arch. Ophthalmology. 12: 475–483
- [3] Godwin Ovenseri-Ogbomo, Mechanism of accommodation: A review of theoretical propositions, African Vision and Eye Health 74(1), March 2015
- [4] Adler, FH; Fliegelman (1934). "Influence of fixation on the visual acuity". Arch. Ophthalmology. 12: 475–483
- [5] Jacob R.J.K., Karn K.S, Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises (Section Commentary), in The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research. Ed. by J. Hyona, R. Radach, and H. Deubel, 573–605, Amsterdam, Elsevier Science (2003).
- [6] Shahzad M.I., Saqib Mehmood, Control of Articulated Robot Arm by Eye Tracking. Master Thesis no: MCS-2010-33, School of Computing Blekinge Institute of Technology, Sweden, september, 2010
- [7] Buswell G.T. (1935). How People Look at Pictures. Chicago: Univ. Chicago Press 137–55. Hillsdale, NJ: Erlbaum
- [8] Jianbin Xiong, Member, IEEE, Weichao Xu, Member, IEEE, Wei Liao, Qinruo Wang, Member, IEEE, Jianqi Liu, and Qiong Liang, Eye Control System Base on Ameliorated Hough Transform Algorithm, IEEE Sensors Journal, Vol. 13, No. 9, September 2013
- [9] Vandna Singh¹, Dr. Vinod Shokeen², Face Detection By Haar Cascade Classifier With Simple And Complex Backgrounds Images Using Opencv Implementation, December 2013
- [19]
 - 1- Human Vision System
 - 2- Eye Gaze Tracking
 - 3- Saccade
 - 4- Smooth Pursuit
 - 5- Fixation
 - 6- Accommodation
 - 7- Depth of Eye Gaze Estimation
 - 8- Pupil
 - 9- Cornea Purkinje
 - 10- parallax
 - 11- Image Processing
 - 12- Deep Learning
 - 13- Iris sclera
 - 14- Sclera
 - 15- Neural Network
 - 16- Retinal Difference
 - 17- Convolutional Neural Network
 - 18- Purkinje
 - 19- Radial Basis Function