

ارائه‌ی یک الگوی تجربی حرکت-زمان از رفتار جهت نگاه انسان در تعدادی موقعیت اجتماعی مختلف

محمدحسین مشاقي^۱ - علیرضا طاهري^۲ - سعید بهزادی پور^۳

^۱ دانشجوی کارشناسی ارشد، دانشکده مهندسی مکانیک (دانشگاه صنعتی شریف)

^۲ استادیار، دانشکده مهندسی مکانیک (دانشگاه صنعتی شریف)

^۳ استاد، دانشکده مهندسی مکانیک (دانشگاه صنعتی شریف)

mh.mashaghi@student.sharif.edu

artaheer@sharif.edu

behzadipour@sharif.edu

چکیده

ربات‌های اجتماعی که برای تعامل با انسان ساخته شده‌اند، نیازمند این هستند که رفتار تعاملی (نظیر جهت نگاه اجتماعی) مشابه انسان داشته باشند. هدف از این پژوهش، ارائه‌ی یک الگوی تجربی حرکت-زمان از جهت نگاه انسان در تعدادی موقعیت اجتماعی مختلف است؛ این موقعیت‌ها شامل صحنه‌هایی با حضور ۲ تا ۴ نفر در یک فیلم است که افراد حاضر در صحنه، رفتارهای اجتماعی «صحبت کردن»، «دست تکان دادن»، «اشاره کردن»، «ورود به صحنه» و «خروج از صحنه» را به نحوی سازماندهی شده انجام می‌دهند. ۱۵ فرد بزرگسال به مشاهده این فیلم پرداخته و موقعیت نگاه آن‌ها توسط ردیاب چشم ثبت گشته است. در ادامه، با استفاده از الگوریتم ژنتیک، ضریب اهمیت هر کدام از رفتارهای اجتماعی ذکر شده استخراج گردید. نتایج بازسازی نگاه شرکت‌کنندگان بر روی داده‌های تست، به عملکرد واقعی افراد بسیار شبیه بود. در انتها قابلیت پیاده‌سازی این الگو، با اجرای آن بر روی ربات نانو با موفقیت مورد آزمایش قرار گرفت و عملکرد مثبت آن با استفاده از تحلیل داده‌های پرسشنامه ارائه شده به ۱۰ شرکت‌کننده تایید گردید. در سه مورد از سوالات مطرح شده، تفاوت معنادار آماری الگوها در دو وضعیت مطالعه شده، مشاهده گردید.

واژه‌های کلیدی

ربات اجتماعی، جهت نگاه اجتماعی، ردیابی چشم، الگوی تجربی حرکت-زمان، الگوریتم ژنتیک

Proposing an empirical motion-time pattern for human gaze behavior in different social situations

Mohammad Hossein Mashaghi: M.Sc. student, Faculty of Mechanical Engineering (Sharif University of Technology)

Alireza Taheri: Assistant Professor, Faculty of Mechanical Engineering (Sharif University of Technology)

Saeed Behzadipour: Professor, Faculty of Mechanical Engineering (Sharif University of Technology)

mh.mashaghi@student.sharif.edu
artaheer@sharif.edu
behzadipour@sharif.edu

Abstract

Social robots that are fabricated to interact with humans and to help them in education, healthcare, etc., are required to have an interactive behavior similar to humans. One of the important interactive behaviors of humans is social eye gaze. Eye gaze is significantly more important than other nonverbal signals; it is shown that eyes are special cognitive stimuli with unique hardwired pathways in the brain dedicated to their interpretation. Studying the literature, we found out that in previous researches conducted to control the social robots' gaze behavior, human gaze behavior was investigated in some limited situations such as two- or three-way conversation in order to extract the pattern of this behavior. Therefore, increasing the variety of studied social situations is a way to fill this gap. In order to design a gaze control system for a social robot, it is required to find out details about the human gaze behavior. The purpose of this research is to propose an empirical motion-time pattern for human gaze behavior in a number of different social situations; these situations include scenes with 2 to 4 people in a prepared video where the people in the scene show the social behaviors of "talking", "waving", "pointing", "entering the scene" and "exiting the scene" in a structured way. Fifteen normal adults (mean age: 24 and std: 3,5 years) watched this movie and their gaze position was recorded by an eye tracker system (SR-Research EyeLink 1000 plus). Next, by using the genetic algorithm (which is an optimization process), we were able to extract the relative coefficient of each of the mentioned social behaviors in our proposed model. The results of reconstructing the participants' gaze on the test data are very similar to the real performance of the subjects. **Finally, the ability to implement this model was successfully tested by implementing it on a Nao robot, and its positive performance was confirmed using a survey.** The model showed significant differences between the two studied situations in 3 questions out of the whole survey's 10 questions.

Keywords: social robot, social eye gaze, eye tracking, motion-time pattern, genetic algorithm

۱- مقدمه

۱-۱- اهمیت و کاربرد موضوع

با توجه به روند روبه رشد فناوری‌های رباتیک، ربات‌های اجتماعی^۱ نیز در سال‌های اخیر پیشرفت قابل توجهی داشته اند. ربات‌های اجتماعی، ربات‌هایی هستند که برای تعامل با انسان‌ها به شیوه‌ای طبیعی و بین فردی، و اغلب برای دستیابی به اهدافی در زمینه‌های آموزش، سلامت، سرگرمی و ... طراحی و به کار گرفته می‌شوند [۱]. تعامل با ربات‌ها همواره برای انسان‌ها و مخصوصاً کودکان جذابیت خاصی دارد. بنابراین می‌توان از این جلب توجه و جذابیت برای هدف‌هایی که به آن‌ها اشاره گردید، به خوبی استفاده نمود.

رفتار تعاملی ربات اجتماعی باید طبیعی و شبیه به انسان باشد تا ربات بتواند با جامعه‌ی هدفش ارتباط مؤثری برقرار کند. درحالی که در تعامل انسان با انسان ارتباطات کلامی^۲ دارای نقش اساسی هستند، رفتارهای غیرکلامی^۳ مانند جهت نگاه^۴ یا ژست^۵ می‌توانند حالت ذهنی فرد را منتقل و ارتباط کلامی را تکمیل کنند (یا آن چه را که گفته می‌شود، تقویت نمایند). از دیدگاه روانشناسی، جهت نگاه از اهمیت بیشتری نسبت به سایر رفتارهای غیرکلامی برخوردار است [۲]. عمده پژوهش‌هایی که برای بهبود تعامل ربات‌های اجتماعی با انسان صورت گرفته اند، تنها متمرکز بر رفتارهای کلامی بوده‌اند؛ بنابراین حرکت در راستای بهبود رفتار جهت نگاه ربات‌های اجتماعی حائز اهمیت می باشد (که در ادبیات موضوعی کمتر به آن پرداخته شده است).

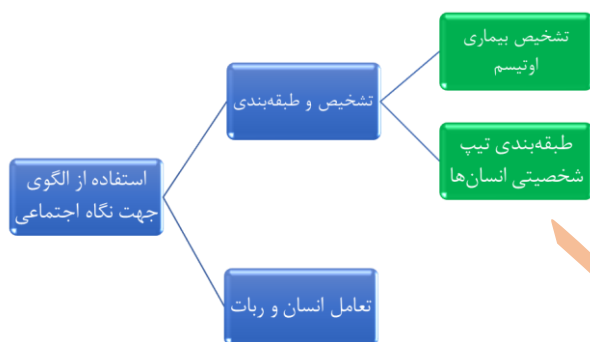
۱-۲- هدف پژوهش

هدف از انجام این پژوهش، استخراج الگوی تجربی حرکت- زمان^۶ چشم انسان (یا الگوی حرکت چشم با در نظر گرفتن زمان) در موقعیت‌های اجتماعی مختلف می‌باشد؛ این مدل نشان‌دهنده‌ی این است که فرد از میان محرک^۷‌های بصری موجود در میدان دید خود، به کدام محرک، توجه می‌کند و نیز با چه سرعتی توجه را از یک محرک به محرک دیگر تغییر می‌دهد. دستیابی به این هدف، نیازمند طراحی آزمایش برای ردیابی جهت نگاه انسان می‌باشد؛ به عبارت دیگر، می‌بایست آزمایشی طراحی شود تا محل نگاه هر فرد را هنگام مواجهه با

موقعیت‌های اجتماعی ضبط و ذخیره نماید. پس از انجام این آزمایش، با استفاده از ابزارهای ریاضی مناسب، یک الگوی تجربی که نماینده‌ی جهت نگاه افراد شرکت کننده در آزمایش باشد، استخراج می‌گردد.

۲- مرور ادبیات موضوع

عمده‌ی پژوهش‌های صورت گرفته در حوزه جهت نگاه اجتماعی را می‌توان به دو گروه عمده تقسیم‌بندی کرد (شکل ۱).



شکل ۱- طبقه بندی پژوهش‌های صورت گرفته در زمینه‌ی جهت نگاه اجتماعی

پژوهش‌هایی که در گروه نخست، یعنی تشخیص و طبقه‌بندی، صورت گرفته‌اند، اگرچه با هدف این مقاله همخوانی ندارند، اما حاوی روش‌هایی هستند که در انتخاب روش انجام پژوهش در این مقاله موثر بودند؛ به عنوان مثال، Liu و همکاران [۳] با استفاده از داده‌برداری از جهت نگاه کودکان مبتلا به اتیسم^۸ هنگام مشاهده‌ی چهره‌ی افراد و بر مبنای یادگیری ماشین^۹، توانستند روشی برای غربالگری کودکان سالم از کودکان مبتلا به اتیسم ارائه بدهند. در پژوهشی دیگر، Jones و همکارش [۴] با اتکا بر داده‌های عددی میزان توجه کودک به چشم‌های فرد مقابل و تحلیل آماری بر این داده‌ها متوجه شدند این میزان توجه در کودکان مبتلا به اتیسم با گذشت زمان کاهش می‌یابد.

دسته‌ی دوم پژوهش‌ها، تعامل انسان و ربات یا به اختصار HRI^{۱۰} نام دارد. حوزه‌ی تعامل انسان و ربات در تلاش است تا یک تعامل ساده و شهودی را میان انسان‌ها و ربات‌ها فراهم سازد [۵]. Yoshikawa و همکاران [۶] با روبروی هم قرار دادن یک انسان و یک ربات انسان نما^{۱۱} متوجه شدند در شرایطی که ربات

^۷ Stimulus

^۸ Autism

^۹ Machine Learning

^{۱۰} Human-Robot Interaction

^{۱۱} Humanoid robot

^۱ Social robot

^۲ Verbal

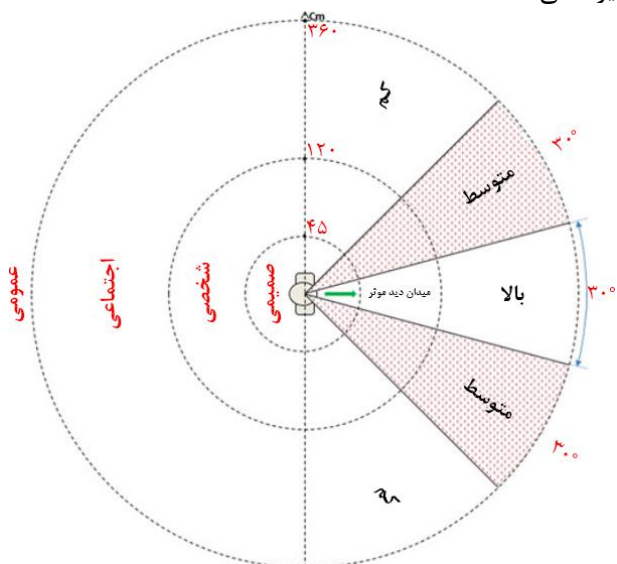
^۳ Nonverbal

^۴ Eye gaze

^۵ Gesture

^۶ Empirical motion-time pattern

نیز مشهود است؛ زیرا انسان هیچ گاه تمام مدت به یک محرک خیره نمی‌ماند.



شکل ۲- مدل ارائه شده توسط Zaraki [۱۰] که نشان‌دهنده‌ی دو عامل فاصله و زاویه دید در تعیین جهت نگاه ربات می‌باشد. میدان دید مؤثر ربات با $FOV_e^{۱۴}$ نشان داده شده است.

در پژوهش Aliasghari و همکاران [۱۱]، اصلاحاتی بر روی مدل ارائه شده توسط Zaraki [۱۰] صورت گرفته است. نویسندگان در این پژوهش پیشنهاد می‌کنند که در امتیاز جلب توجه برای هر محرک، به جای حاصل جمع ویژگی‌های آن محرک، از حاصل-ضرب آن‌ها استفاده شود. همچنین، در این پژوهش قابلیت پلک زدن ربات نیز به مدل [۱۰] اضافه شده است. در [۱۱]، از ۲۳ نفر آزمایش ردیابی نگاه به عمل آمده که از این نظر نسبت به [۱۰] برتری دارد؛ اما داده‌برداری جهت نگاه با دقت پایین‌تری صورت گرفته است. علت این امر استفاده از سنسور کیپتکت برای ردیابی نگاه افراد بوده است. در پژوهش ذکر شده، محرک‌هایی که مورد بررسی قرار گرفتند عبارت‌اند از: صحبت کردن، دست تکان دادن، مورد اشاره قرار گرفتن توسط فرد دیگر، ورود به تصویر و خروج از آن و نگاه کردن فرد به دوربین.

در پژوهش Domingo و همکاران [۱۲]، ایده‌ی بهینه‌سازی ضرایب برای یافتن بهترین رفتار جهت نگاه مطرح شده است. در این پژوهش، ۷ محرک مورد بررسی قرار گرفته‌اند: قرار داشتن فرد در میدان دید ربات، صحبت کردن فرد، نگاه کردن مستقیم فرد به ربات، حرکت کردن فرد، عدم حضور فرد در میدان دید ربات به رغم شنیده شدن صدای او، عدم حضور فرد در میدان دید ربات اما مورد توجه قرار داشتن وی توسط سایر افراد حاضر در میدان دید ربات، و فاصله از ربات. نکته‌ای که در این پژوهش

جهت نگاه فرد مقابلش را دنبال می‌کند، می‌تواند ارتباط چشمی بهتر و مداوم‌تری را از فرد مقابل دریافت کند.

تعدادی از پژوهش‌ها، بدون استفاده از الگوی جهت نگاه انسان، برای کنترل جهت نگاه ربات‌های اجتماعی تلاش کرده‌اند؛ به عنوان مثال Cornia و همکاران [۷]، با آموزش دادن یک شبکه‌ی عصبی بازگشتی از نوع LSTM^{۱۲} توانسته‌اند نقاط برجسته‌ی تصاویر ایستار را استخراج کنند. در پژوهشی دیگر، Lathuilière و همکاران [۸] با استفاده از یادگیری تقویتی، روشی برای کنترل جهت نگاه ربات ارائه کرده‌اند. در این روش، شبکه‌ی عصبی بر اساس داده‌های ورودی تصویر و صدای محیط اطراف، یاد می‌گیرد تا توجه ربات را در جهتی قرار بدهد که تعداد افراد در میدان دیدش بیشینه شود. شبکه‌ی طراحی شده ابتدا با استفاده از ویدیوهای از پیش تهیه شده آموزش داده می‌شود و سپس برای صحت‌سنجی در محیط واقعی قرار می‌گیرد. همچنین در یک پژوهش بسیار جامع، Yoo و همکارش [۹] یک روش بر مبنای انتگرال فازی برای کنترل جهت توجه و نگاه یک سر رباتیک معرفی کرده‌اند.

در دسته دیگر پژوهش‌ها، از الگوی جهت نگاه انسان جهت اصلاح عملکرد سیستم کنترل جهت نگاه ربات استفاده شده است. در پژوهش Zaraki و همکاران [۱۰] یک سیستم کنترل جهت نگاه برای ربات انسان‌نمای FACE ارائه شده است. در این سیستم با در نظر گرفتن عواملی مانند: میزان نزدیک بودن فرد به ربات، نشانه‌های کلامی و غیرکلامی، میدان دید مؤثر و اثر خوگیری^{۱۳}، به هر کدام از محرک‌های محیط یک امتیاز منظور می‌گردد. امتیاز بالاتر یک محرک، بیانگر اولویت ربات برای معطوف کردن جهت نگاهش به آن محرک می‌باشد. رابطه‌ای که مقاله مذکور برای محاسبه‌ی امتیاز هر محرک در نظر گرفته است، به صورت جمع تعدادی از عوامل است. نخستین عامل آن، جمع وزن‌دار ویژگی‌های نرمال شده‌ی محرک است؛ یعنی ویژگی هر محرک در میزان اهمیت آن ویژگی ضرب و مجموع حاصل ضرب‌ها برای تمام ویژگی‌های آن محرک محاسبه می‌گردد. دومین عامل بیانگر میزان نزدیکی یک محرک به ربات است. سومین عامل بیانگر زاویه‌ی قرارگیری محرک نسبت به ربات و آخرین عامل نشان‌دهنده‌ی حافظه‌ی ربات از اتفاقات گذشته‌ی مرتبط با آن محرک است (شکل ۲). نویسندگان [۱۰] اثر خوگیری را به این صورت وارد مسئله کرده‌اند که امتیاز محرکی که برنده شده و جهت نگاه ربات به آن معطوف می‌گردد، ثابت باقی نمی‌ماند و به طور خطی نزول می‌کند. این اثر در واقعیت

^{۱۴} Effective Field of View

^{۱۲} Long Short-Term Memory

^{۱۳} Habituation

جلب توجه می‌کند این است که از ۷ محرک یاد شده، ۶ مورد آن به صورت صفر و یکی در نظر گرفته شده‌اند و هیچ‌گونه طیفی برای محرک‌ها لحاظ نشده است.

پس از جمع‌بندی مرور ادبیات موضوعی نتیجه این شد که با تکیه بر داده‌برداری صورت گرفته از تعدادی از افراد سالم، الگوی تثبیت و حرکت چشم افراد در مجموعه‌ی تا حد امکان غنی از موقعیت‌های اجتماعی اشاره شده بررسی و استخراج گردد. موقعیت‌های مد نظر فقط شامل محرک انسانی هستند؛ اما در این پژوهش اولاً تعداد انسان‌هایی که به عنوان محرک در نظر گرفته می‌شوند، بیش از ۲ نفر خواهد بود؛ و ثانیاً به هر محرک ویژگی‌های گسترده‌تری نسبت داده می‌شود که هر کدام از ویژگی‌ها با وزن مشخصی در اولویت جلب توجه آن محرک تأثیرگذار است. با توجه به این که پژوهش فعلی به نوعی ادامه‌دهنده و تکمیل‌کننده‌ی پژوهش Aliasghari و همکاران [۱۱] است، محرک‌های مورد بررسی در پژوهش فعلی نیز با هدف تکمیل محرک‌های آن پژوهش انتخاب گردیدند.

۳- روش انجام پژوهش

برای رسیدن به هدف این پژوهش باید دو مرحله انجام شود؛ مرحله‌ی نخست جمع‌آوری داده‌ها نام دارد که خود شامل دو زیرمرحله اصلی است؛ طراحی آزمایش و اجرای آزمایش. در مرحله‌ی دوم نیز تحلیل داده‌ها صورت می‌گیرد که استخراج الگوی جهت نگاه در این بخش اتفاق می‌افتد.

۳-۱- جمع‌آوری داده‌ها

نخستین گام از انجام این پژوهش، جمع‌آوری داده‌هاست. همانطور که در عنوان این مقاله مشهود است، هدف استخراج الگویی از رفتار جهت نگاه انسان است؛ بنابراین لازم است تا در ابتدا این رفتار مشاهده و ضبط شود. به منظور ضبط داده‌های مورد نیاز برای این پژوهش، نیاز است تا یک آزمایش مناسب طراحی و اجرا گردد.

۳-۱-۱- طراحی آزمایش

همان‌طور که در مرور ادبیات مطرح شد، انتخاب محرک‌های اجتماعی مورد بررسی با هدف تکمیل پژوهش Aliasghari و همکاران [۱۱] صورت گرفته است؛ بنابراین محرک‌های «صحبت کردن» و «دست تکان دادن» به همان شکل مورد بررسی قرار می‌گیرند. با توجه به این که تعداد افراد حاضر در صحنه در پژوهش فعلی تا ۴ نفر قابل افزایش است، محرک «اشاره کردن» به صورت متفاوتی از پژوهش قبلی مورد بررسی قرار می‌گیرد؛

یعنی به جای آن که افراد مورد اشاره‌ی یکدیگر قرار بگیرند، یک هدف مشترک مورد اشاره‌ی افراد حاضر در آزمایش قرار می‌گیرد. محرک‌های «وارد شدن به تصویر» و «خارج شدن از تصویر» به منظور بهبود دقت بررسی نسبت به پژوهش قبلی، با دو سرعت متفاوت بررسی می‌گردند. «فاصله‌ی افراد از دوربین» و «زاویه‌ی قرارگیری نسبت به دوربین» هم مشابه پژوهش قبلی بررسی می‌شوند.

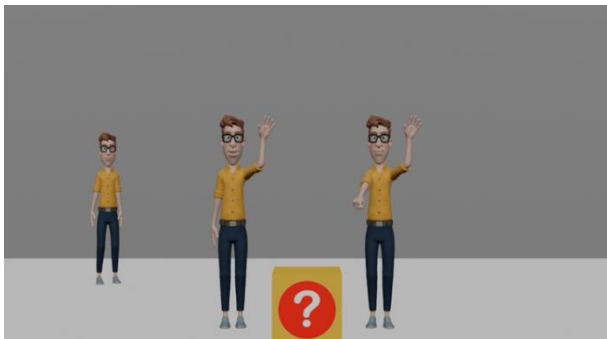
بر اساس توضیحاتی که داده شد، هدف از این آزمایش بررسی اثر محرک‌های اجتماعی زیر بر رفتار جهت نگاه انسان است:

- صحبت کردن
- دست تکان دادن
- اشاره کردن
- وارد شدن به تصویر (با دو سرعت متفاوت)
- خارج شدن از تصویر (با دو سرعت متفاوت)
- فاصله از دوربین (در دو حالت نزدیک و دور)
- زاویه با دوربین (در دو حالت زاویه مستقیم و زاویه بسته)
- تعداد افراد حاضر در صحنه (از ۲ تا ۴ نفر)

همانطور که در لیست اشاره شده مشاهده می‌شود، بررسی هر یک از پارامترهای ذکر شده در یک آزمایش جداگانه، نیازمند اجرای چند ده هزار آزمایش می‌باشد؛ این مسئله با توجه به وابستگی به شرکت‌کننده‌های انسانی در این آزمایش، غیرممکن است؛ بنابراین لازم است تا با روش‌های استاندارد طراحی آزمایش، اثر تمام پارامترهای مذکور در تعدادی آزمایش محدود گنجانده شود؛ بنابراین از روش طراحی آزمایش تاگوجی استفاده گردید. در پایان طراحی آزمایش، یک ویدیو مورد نیاز است که تمام پارامترهای لیست بالا را شامل باشد و اثر متقابل هر کدام را مورد آزمایش قرار بدهد؛ در نتیجه ابتدا لازم است برای این ویدیو یک سناریو طراحی و پس از آن به ساختن ویدیو پرداخته شود. با توجه به این که تعداد افراد حاضر در صحنه از ۲ تا ۴ نفر متغیر است، برای هر کدام از این سه حالت طراحی تاگوجی به صورت جداگانه در نظر گرفته و سپس با کنار هم قرار دادن این سه حالت، یک سناریوی تا حد امکان پیوسته ساخته شد.

در شکل ۳ ضربدرها نشان‌دهنده‌ی محل‌های مجاز برای قرارگیری افراد هستند؛ به عبارت دیگر، هر خط سبز یک امتداد مجاز برای قرارگیری افراد است. همچنین مستطیل رنگی روبروی دوربین محلی است که افراد در صورتی که در حال اشاره کردن باشند، به سمت این مستطیل اشاره می‌کنند.

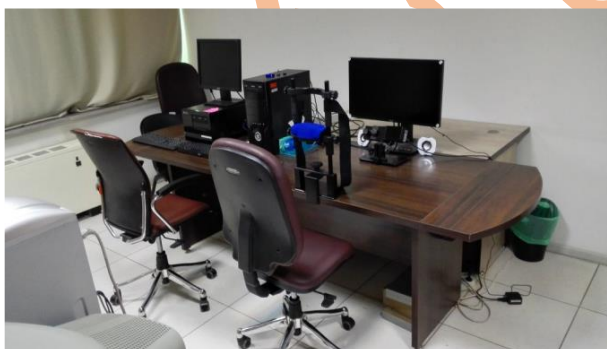
است. شخصیت‌هایی که در شکل ۴ مشاهده می‌شوند، همگی یکسان می‌باشند (این شخصیت Vincent نام دارد که یک شخصیت رایگان در کتابخانه‌ی Blender و با قابلیت‌های حرکتی بسیار زیاد است). همانطور که در شکل ۴ مشاهده می‌شود، اگر شخصیت‌ها در حال اشاره باشند، همگی به جعبه‌ی موجود در مرکز تصویر اشاره می‌کنند. خروجی نرم‌افزار Blender دارای ۱۵۳۴۴ فریم و مدت زمان تقریبی ۶۳۹.۳۳ ثانیه می‌باشد.



شکل ۴- تصویر یک فریم از انیمیشن مورد استفاده در آزمایش

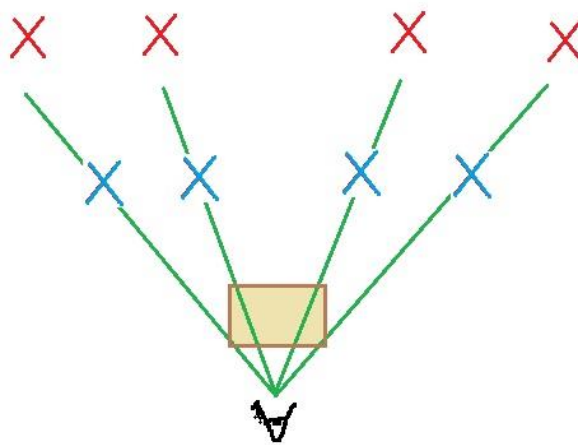
۳-۱-۲- اجرای آزمایش

آزمایش ردیابی چشم توسط دستگاه SR-Research EyeLink ۱۰۰۰ موجود در مرکز توانبخشی عصبی موفقیان صورت گرفت. این دستگاه قابلیت داده‌برداری تا فرکانس ۲۰۰۰ هرتز را دارد و همچنین خطای اندازه‌گیری آن حدوداً ۰.۱۵ درجه می‌باشد. همانطور که در شکل ۵ مشاهده می‌شود، این دستگاه دارای یک پایه‌ی نگهدارنده‌ی سر است تا سر فرد شرکت‌کننده در آزمایش را در یک موقعیت ثابت نگه دارد.



شکل ۵- چیدمان دستگاه EyeLink ۱۰۰۰ plus در مرکز موفقیان

تعداد افراد شرکت‌کننده در این آزمایش ۱۵ فرد بهنجار (شامل ۹ آقا و ۶ خانم با میانگین سنی ۲۵.۴ و انحراف معیار ۳.۳ سال) بودند. داده‌برداری در این آزمایش با فرکانس ۱۰۰۰ هرتز انجام گرفت. حین انجام آزمایش، هر یک از شرکت‌کننده‌ها به ویدئوی طراحی شده نگاه کرده و موقعیت نگاه آن‌ها بر روی تصویر دو بعدی ثبت شده است.



شکل ۳- محل‌های مجاز برای قرارگیری افراد حاضر در ویدئو

با استفاده از روش تاگوچی، ۱۶ سناریو برای زمانی که ۲ نفر در صحنه حضور دارند، ۳۲ سناریو برای زمانی که ۳ نفر در صحنه هستند و ۱۶ سناریو برای زمانی که ۴ نفر در صحنه حاضر هستند، طراحی شد. یعنی در مجموع با ۶۴ سناریو، تمام پارامترهای مدنظر مورد بررسی قرار می‌گیرند.

پس از تشکیل سناریوی کلی ویدئو، به ساختن ویدئو پرداخته شد. برای تولید ویدئو باید به این نکته توجه داشت که باید از وجود هر متغیری جز متغیرهای مورد بررسی در هدف این پژوهش اجتناب کرد؛ با توجه به این مسئله، ویژگی‌های کلی زیر برای ویدئو لحاظ گردید:

- شخصیت‌های موجود در ویدئو یکسان باشند.
 - غیر از محل اشاره‌ی افراد، هیچ‌گونه عامل جالب توجهی وجود نداشته باشد.
 - تمام محرک‌های اجتماعی موجود در سناریوها به روش یکسانی انجام شوند.
 - هر کدام از ردیف‌های سناریوی کلی در مدت زمان یکسانی اجرا شوند.
- با در نظر گرفتن ویژگی‌های فوق تصمیم بر این شد که یک انیمیشن با ویژگی‌های جدول ۱ ساخته شود.

جدول ۱- ویژگی‌های انیمیشن مورد استفاده در آزمایش

اندازه‌ی فریم	۱۹۲۰ در ۱۰۸۰
نرخ پخش فریم	۲۴ فریم بر ثانیه
مدت زمان کل	۶۴۰ ثانیه

در ساخت این انیمیشن از نرم‌افزار Blender استفاده شد که یک نرم‌افزار رایگان و متن‌باز^{۱۵} در زمینه‌ی گرافیک سه‌بعدی

^{۱۵} Open-source

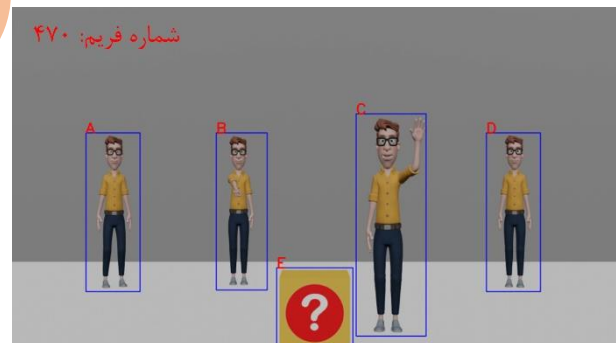
۲-۳- تحلیل داده‌ها

۲-۳-۱- پالایش داده‌های خام

داده‌های خام به دست آمده از آزمایش ردیابی چشم به صورت زوج اعداد (x, y) هستند که بیانگر مختصات جهت نگاه بر روی صفحه نمایش است. با توجه به فرکانس بالای داده‌برداری در این آزمایش، احتمال این که تعدادی از نمونه‌ها از دست بروند، وجود دارد. برای جایگزین کردن نمونه‌های از دست رفته از درونبایی خطی استفاده شد.

همانطور که گفته شد، فرکانس داده‌برداری ۱۰۰۰ هرتز است؛ اما فرکانس پخش فریم‌های ویدیوی آزمایش برابر با ۲۴ هرتز (معادل ۲۴ فریم بر ثانیه) می‌باشد؛ یعنی به ازای هر فریم، حدوداً ۴۰ بار نمونه‌برداری اتفاق افتاده است. برای آن که بتوان خروجی به دست آمده از تحلیل‌ها را مجدداً بر روی ویدیوی آزمایش نمایش داد، لازم است تا تعداد نمونه‌ها با تعداد فریم‌های ویدیو برابر باشد؛ بنابراین در مرحله‌ی بعدی پالایش داده‌های خام، میانگین حسابی هر ۴۰ نمونه به یک فریم اختصاص داده شد.

پالایش داده‌ها در این مرحله ختم نمی‌شود. در مرحله‌ی بعدی از پالایش داده‌ها تعیین می‌شود که فرد شرکت‌کننده در هر فریم به کدام یک از المان‌های موجود در صحنه نگاه کرده است. همانطور که در شکل ۶ مشاهده می‌شود، افراد از چپ به راست به ترتیب A، B، C و D نامگذاری می‌شوند و جعبه‌ی موجود در جلوی تصویر نیز E نام می‌گیرد. سایر نقاط تصویر نیز env نامیده می‌شوند.



شکل ۶- رسم مستطیل دربرگیرنده برای هر کدام از المان‌های موجود در صحنه

در هر فریم از ویدیو، به دور هر المان موجود در تصویر یک مستطیل رسم می‌گردد؛ بنابراین اگر مختصات جهت نگاه درون هر یک از این مستطیل‌ها قرار بگیرد، بیانگر این است که فرد شرکت‌کننده در آن فریم در حال نگاه کردن به المان متناظر با آن مستطیل بوده است. در صورتی که مختصات جهت نگاه درون هیچ کدام از مستطیل‌ها قرار نگیرد، این گونه تفسیر می‌شود که فرد در حال نگاه کردن به محیط بوده است.

مرحله‌ی بعد، تهیه‌ی ۶ سری زمانی A، B، C، D، E و env است؛ هر سری زمانی نشان‌دهنده‌ی این است که در هر فریم، چه درصدی از شرکت‌کنندگان آزمایش به المان متناظر با آن سری زمانی نگاه کرده‌اند.

در مرحله‌ی پایانی پالایش داده‌های خام، لازم است تا نویز موجود در سری‌های زمانی حذف شود. به این منظور از فیلتر Savitzky-Golay با طول پنجره‌ی ۳ و مرتبه‌ی چندجمله‌ای ۵۱ استفاده گردید.

۲-۳-۲- محاسبه میانگین و کوواریانس

پس از پالایش داده‌های خام، میانگین و کوواریانس تمام شرکت‌کنندگان آزمایش روی هر فریم محاسبه می‌شود و روی ویدیوی آزمایش ترسیم می‌گردد. هدف از انجام این کار داشتن معیاری برای مقایسه‌ی الگوی نهایی به دست آمده با رفتار کلی شرکت‌کنندگان در آزمایش است.

۲-۳-۳- مدل‌سازی و استخراج الگو

در پژوهش [۱۰] یک مدل ریاضی برای امتیازدهی به افراد موجود در صحنه معرفی گردیده است (رابطه‌ی ۱).

$$EA = \left(\sum w_i \right) + P(r) + \Theta(\theta) \quad (1)$$

در رابطه‌ی ۱، عبارت $\sum w_i$ بیانگر محرک‌های اجتماعی است؛ یعنی هر محرک اجتماعی که فرد در حال انجام دادن آن است، دارای یک وزن است که با w_i نشان داده می‌شود. مجموع این وزن‌ها در میزان جلب توجه فرد تاثیر مستقیم دارند. عبارت‌های $P(r)$ و $\Theta(\theta)$ نیز به ترتیب اثر فاصله و زاویه را در میزان جلب توجه نشان می‌دهند. در [۱۰]، فاکتور خوگیری نیز که به صورت زیر تعریف می‌شود، در بالاترین امتیاز موجود ضرب می‌شود (رابطه‌ی ۲).

$$HF(\Delta t) = Peak \cdot \max \left(0, 1 - \frac{\Delta t}{\tau} \right) \quad (2)$$

در رابطه‌ی ۲، عبارت Δt بیانگر مدت زمانی است که یک فرد مشخص دارای بالاترین امتیاز بوده است.

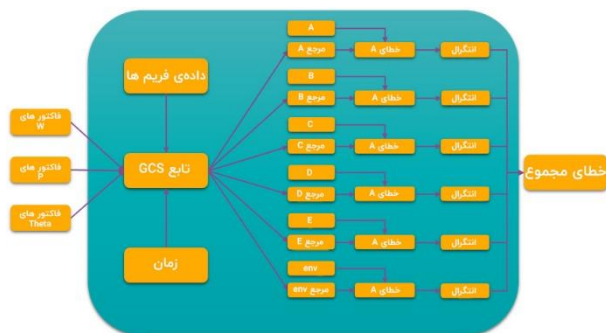
در پژوهش Aliasghari و همکاران [۱۱]، مدل امتیازدهی به افراد مشابهت زیادی (البته با کمی تفاوت) با پژوهش [۱۰] دارد (رابطه‌ی ۳).

$$EA = \left(\sum w_i \right) P(r) \Theta(\theta) \quad (3)$$

همانطور که در رابطه‌ی ۳ مشاهده می‌شود، تنها تفاوت مدل امتیازدهی در [۱۱] با پژوهش [۱۰] این است که تمام عوامل موثر در جلب توجه در یکدیگر ضرب می‌شوند.

داشته باشد، لازم است که خطای مدل از داده‌ی واقعی اثر مستقیمی روی تابع هزینه داشته باشد.

همانطور که در شکل ۷ مشاهده می‌شود، ورودی تابع هزینه همان پارامترهای مجهول هستند. خروجی تابع GCS^{۱۶} شش سری زمانی است که اندازه‌ی آن برابر با اندازه‌ی سری‌های زمانی به دست آمده از پالایش داده‌های خام آزمایش است. ابتدا قدر مطلق تفاضل هر سری زمانی با سری زمانی خروجی از آزمایش محاسبه می‌گردد و سپس از سری‌های زمانی حاصل انتگرال گرفته می‌شود. مجموع این ۶ انتگرال برابر با خروجی تابع هزینه می‌باشد.



شکل ۷- نحوه‌ی عملکرد تابع هزینه برای الگوریتم ژنتیک

با داشتن تابع هزینه، می‌توان الگوریتم ژنتیک را برای بهینه‌سازی تابع هزینه اجرا کرد (جدول ۳).

جدول ۳- پارامترهای الگوریتم ژنتیک برای بهینه‌سازی مدل

۳۲	جمعیت اولیه
۱۶	جمعیت
۸	جمعیت خوب
۱۰ درصد	نرخ جهش
۱۰۰۰۰	حداکثر تعداد نسل
۰.۱ تا ۵	بازه‌ی مجاز w_i
۱ تا ۵	بازه‌ی مجاز P
۱ تا ۵	بازه‌ی مجاز Θ

همانطور که در گذشته نیز بیان شد، ویدیوی آزمایش فاز دوم دارای ۱۵۳۴۴ فریم است. برای آن که بتوان از داده‌های به دست آمده از آزمایش هم برای آموزش مدل و هم برای ارزیابی آن استفاده کرد، ۱۲۰۰۰ فریم آغازین در الگوریتم ژنتیک شرکت داده می‌شوند و سایر فریم‌ها نیز برای ارزیابی نتیجه مورد استفاده قرار می‌گیرند.

۳-۲-۴- نحوه‌ی انتخاب هدف نگاه

مدل پیشنهادی این پژوهش نیز بسیار مشابه دو روش قبل انتخاب شده است (رابطه‌ی ۴).

$$EA = \left(\prod w_i \right) P(r) \Theta(\theta) \quad (4)$$

همانطور که در رابطه‌ی ۴ مشاهده می‌شود، تنها تفاوت مدل پیشنهادی این پژوهش با [۱۱] این است که وزن محرک‌های اجتماعی به جای جمع شدن با یکدیگر، در هم ضرب می‌شوند. همچنین علاوه بر فاکتور خوگیری، تصمیم گرفته شد که امتیاز نهایی هر فرد عددی تصادفی در یک توزیع نرمال حول خروجی رابطه‌ی ۴ باشد؛ با این روش می‌توان حرکت‌های سریع و تصادفی چشم را نیز مدل کرد.

همانطور که در جدول ۲ مشاهده می‌شود، برای این مدل ۱۴ پارامتر مجهول وجود دارد.

جدول ۲- پارامترهای مدل پیشنهادی این پژوهش

Θ	P	w_i	
مستقیم (مجهول) زاویه‌ی بسته (که برابر با یک است)	نزدیک (مجهول) دور (که برابر با یک است)	صحبت کردن (مجهول) دست تکان دادن (مجهول) اشاره کردن (مجهول) ورود سریع (مجهول) ورود آهسته (مجهول) خروج سریع (مجهول) خروج آهسته (مجهول) بی حرکت ماندن (مجهول) بیرون از صحنه بودن (که برابر با صفر است)	برای افراد
-	-	اشاره شدن توسط فرد نزدیک و مستقیم (مجهول) اشاره شدن توسط فرد دور و مستقیم (مجهول) اشاره شدن توسط فرد نزدیک و زاویه بسته (مجهول) اشاره شدن توسط فرد دور و زاویه بسته (مجهول)	برای جعبه

برای آن که بتوان از داده‌های آزمایش ردیابی چشم برای یافتن مقدار بهینه‌ی پارامترهای مجهول استفاده کرد، نیاز است که از یک روش مناسب بهینه‌سازی استفاده نمود. با توجه به تعداد بالای پارامترهای مجهول و پیچیده و غیرخطی بودن تابع مدل، تصمیم گرفته شد که از روش‌های فراابتکاری استفاده شود؛ بدین منظور، الگوریتم ژنتیک برای این کار انتخاب شد.

برای استفاده از الگوریتم ژنتیک نیاز است تا یک تابع هزینه مناسب تعریف گردد. با توجه به این که ضروری است که خروجی مدل بیشترین شباهت را به داده‌های به دست آمده از آزمایش

^{۱۶} Gaze Control System



شکل ۸- ربات نائو در مسابقات روبوکاپ ۲۰۱۶ [۱۶]

۳-۳-۲- پیاده سازی الگو بر روی ربات نائو

ربات نائو تمام تجهیزات سخت‌افزاری برای اندازه‌گیری پارامترهای مورد نیاز در این پژوهش را به جز اندازه‌گیری فاصله‌ی اجسام از ربات را دارد؛ بنابراین برای برطرف کردن این کمبود از دوربین کینکت استفاده شد. شرکت سازنده‌ی ربات نائو، به منظور تسهیل ارسال فرامین به ربات مجموعه‌های نرم‌افزاری را ارائه نموده است. همچنین برای استفاده از قابلیت‌های دوربین کینکت، نیاز به برنامه‌نویسی به زبان C# می‌باشد. نحوه‌ی انتخاب هدف که پیش‌تر توضیح داده شد نیز نیازمند برنامه‌نویسی است. که در این پژوهش با استفاده از زبان پایتون ۳ انجام گرفته است. با استفاده از برنامه نویسی C#، قابلیت تشخیص موقعیت قرارگیری نسبت به دوربین (فاصله و زاویه) برای دوربین کینکت تعریف گردید. در قسمت مرتبط با ربات نیز ابتدا قابلیت تشخیص اسکلت هر فردی که مقابل دوربین قرار می‌گیرد به ربات افزوده شد. سپس با تحلیل هر اسکلت، قابلیت تشخیص حرکت‌های ورود به تصویر و خروج از آن، دست تکان دادن و اشاره کردن فراهم گردید. به منظور تشخیص صحبت کردن افراد نیز از آرایه میکروفونی سنسور کینکت استفاده شد. در نهایت با اعمال ضرابی که از خروجی الگوریتم ژنتیک به دست آمده بود، امکان افزودن قابلیت کنترل جهت نگاه به ربات نائو فراهم گردید.

قبل از این که بتوان چنین الگویی را بر روی یک ربات اجتماعی پیاده‌سازی کرد، لازم است تا تعیین شود این الگو، انتخاب هدف نگاه خود را به چه نحوی انجام می‌دهد. دو روش پیشنهادی حریصانه و تصادفی وزن‌دار مطرح هستند. در روش اول، همواره المانی که دارای بیشترین امتیاز در تصویر باشد، به عنوان هدف برگزیده خواهد شد؛ اما در روش دوم، تمام المان‌های موجود در تصویر شانس برگزیده شدن را دارند؛ به این صورت که المان با امتیاز بیشتر، از شانس بیشتری برای برگزیده شدن برخوردار است.

مسئله‌ی بعدی در انتخاب هدف نگاه کردن این است که پس از برگزیده شدن هدف، به کدام نقطه از هدف نگاه شود. راه‌کار پیشنهادی برای این مسئله این است که برای هر هدف، یک نقطه‌ی اصلی تعریف شود؛ سپس یک نقطه‌ی تصادفی توزیع نرمال حول نقطه‌ی اصلی و درون مستطیل محیطی هدف، به عنوان مختصات نهایی جهت نگاه برگزیده می‌شود.

۳-۳-۳- پیاده سازی الگو بر روی ربات

۱-۳-۳- انتخاب ربات برای پیاده‌سازی

سه گزینه برای پیاده‌سازی این الگو مورد بررسی قرار گرفتند: (۱) ربات آپو: یک ربات اجتماعی که با هدف تقویت مهارت لب‌خوانی افراد دچار نقص شنوایی در آزمایشگاه رباتیک اجتماعی دانشگاه صنعتی شریف ساخته شده است [۱۳].

(۲) ربات رسا: یکی دیگر از ساخته‌های آزمایشگاه رباتیک اجتماعی دانشگاه صنعتی شریف که به صورت ویژه در پژوهش Basiri و همکاران [۱۴] برای پیاده‌سازی زبان اشاره‌ی فارسی مورد استفاده قرار گرفته است.

(۳) ربات نائو: ربات نائو یک ربات انسان‌نما و محصول شرکت آلدبارن فرانسه و هولدینگ سافت بنک رباتیکس می‌باشد. این ربات معروف در سال ۲۰۰۸ میلادی اولین در انظار عمومی حضور یافت و پس از آن تا کنون در امور بسیاری از جمله دستیاری در شرکت‌ها و مراکز درمانی به منظور خوش‌آمدگویی، اطلاع‌رسانی و سرگرمی مراجعان مورد استفاده قرار گرفته است [۱۶]، [۱۵].

با توجه به امکاناتی که در زمان انجام این پژوهش در دانشگاه صنعتی شریف در دسترس قرار داشتند، ربات نائو^{۱۷} برای پیاده‌سازی ابتدایی این الگو انتخاب گردید.

^{۱۷} Nao Robot

پس از پیاده‌سازی مدل به دست آمده بر روی ربات نائو، با تنظیم یک پرسشنامه از ۱۰ نفر شرکت‌کننده (متفاوت از افراد شرکت‌کننده در بخش آزمایش ردیاب چشم) خواسته شد که با مشاهده‌ی دو ویدیو به تعدادی سوال پاسخ دهند. در هر دو ویدیو ربات نائو در حال تعامل (پارامترهای دست تکان دادن و اشاره کردن هم در این تعامل گنجانده شده است) با ۲ نفر حاضر در صحنه است. در ویدیوی اول ربات نائو از روش انتخاب هدف حریصانه استفاده می‌کند؛ اما در دومین ویدیو از روش تصادفی وزن‌دار برای انتخاب هدف بهره می‌برد.

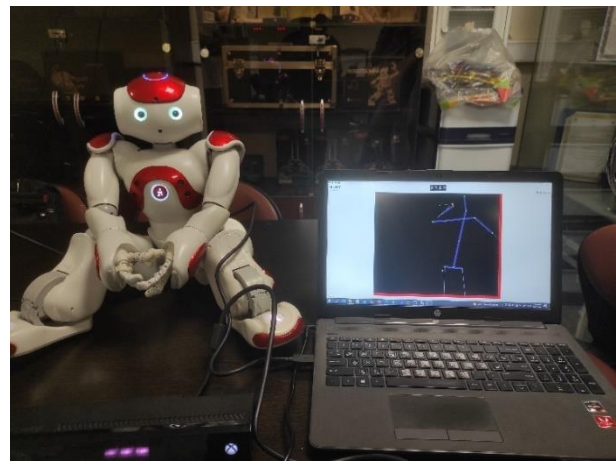
سوالاتی که در پرسشنامه مطرح شدند و می‌بایست با پنج سطح «۱: کاملاً مخالفم»، «۲: مخالفم»، «۳: بی‌نظرم»، «۴: موافقم» و «۵: کاملاً موافقم» پاسخ داده شوند، به شرح زیر هستند:

- ۱) از تعامل ربات با افراد حاضر در ویدیو احساس رضایت دارم.
 - ۲) از هماهنگی ربات با حرکات افراد ویدیو احساس رضایت دارم.
 - ۳) از نظر من ربات در ویدیو، افراد حاضر در ویدیو را درک می‌کند.
 - ۴) حس کردم تعامل ربات در ویدیو با افراد حاضر در ویدیو مانند یک آدم واقعی بود.
 - ۵) به نظر من ربات در ویدیو هم‌نشین اجتماعی خوبی است.
 - ۶) می‌توانم تصور کنم ربات در ویدیو موجودی زنده است.
 - ۷) به نظر می‌رسد ربات در ویدیو توجه خوبی به اطرافیان دارد.
 - ۸) ربات در ویدیو باهوش رفتار کرد.
 - ۹) به نظر من ربات در ویدیو به اقدامات افراد حاضر در ویدیو خوب واکنش نشان داد.
 - ۱۰) به نظر من ربات در ویدیو درک خوبی از اقدامات افراد حاضر در ویدیو داشت.
- داده‌های به دست آمده از طریق پرسشنامه با استفاده از روش paired t-test تحلیل گردیدند.

۴- نتایج

۴-۱- نمایش میانگین و کوواریانس

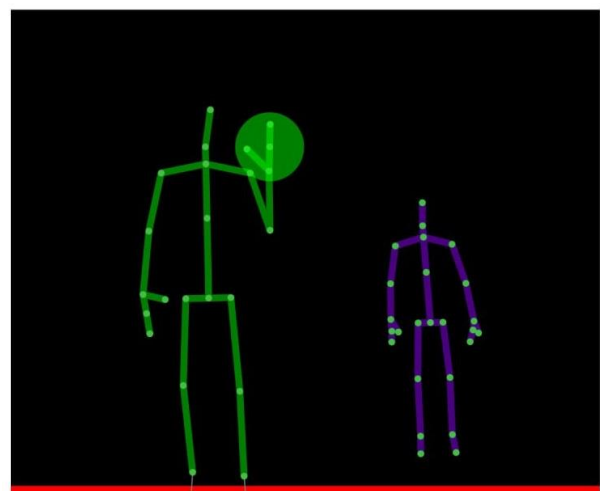
همانطور که در شکل ۱۲ مشاهده می‌شود، داده‌های به دست آمده از محاسبه‌ی میانگین و کوواریانس داده‌های اولیه بر ویدیوی اصلی آزمایش ترسیم شده‌اند.



شکل ۹- پیکربندی ربات نائو و دوربین کینکت برای پیاده‌سازی سیستم کنترل جهت نگاه



شکل ۱۰- اجرای یک سناریو برای ارزیابی عملکرد سیستم کنترل جهت نگاه ربات نائو



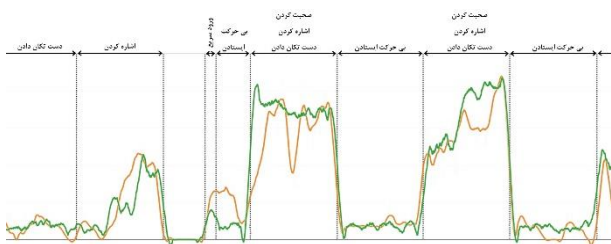
شکل ۱۱- تصویر مشاهده شده توسط ربات در لحظه‌ی اجرای سناریو

همان‌طور که در شکل ۱۰ مشاهده می‌شود، ربات فرد نزدیک‌تر و در حال دست تکان دادن را برای نگاه کردن انتخاب کرده است. شکل ۱۱ تصویر مشاهده شده توسط ربات در لحظه‌ی اجرای سناریو را نمایش می‌دهد.

۴-۳- ارزیابی خروجی پیاده‌سازی

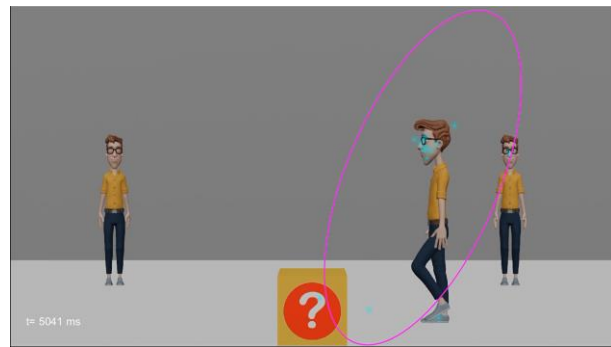
حال با محاسبه و در ادامه بهره‌گیری از این ضرایب می‌توان خروجی‌های تابع GCS موجود در تابع هزینه را به صورت جداگانه رسم کرد.

در نمودار شکل ۱۳، منحنی نارنجی نشان‌دهنده داده‌های اصلی آزمایش (استخراج شده از میانگین نگاه شرکت‌کنندگان) پس از عبور از فیلتر Savitzky-Golay است. منحنی سبز نیز بیانگر خروجی تابع مدل‌سازی شده پس از عبور از فیلتر Savitzky-Golay می‌باشد. پیکان‌های دوطرفه در بالای نمودار نیز بیانگر این هستند که در آن مدت زمان مشخص، شخصیت A در حال اجرای چه محرک اجتماعی بوده است.



شکل ۱۳- بخشی از سری زمانی متناظر با شخصیت A

همانطور که در نمودار شکل ۱۳ قابل مشاهده است (و همچنین در سایر داده‌های تست نیز مشاهده شده است)، منحنی خروجی مدل با تقریب بسیار خوبی، منحنی داده‌های آزمایش را دنبال می‌کند. با توجه به این که مقیاس هر دو نمودار یکسان است، با



شکل ۱۲- نمایش میانگین و کوواریانس داده‌های به دست آمده از آزمایش فاز دوم بر روی ویدیوی آزمایش

۲-۴- نتایج مدل‌سازی و استخراج الگو

با اجرای ۱۰ هزار نسل از الگوریتم ژنتیک، پارامترهای تابع هزینه مطابق جدول ۴ تنظیم شدند. با داشتن این ضرایب می‌توان خروجی‌های تابع GCS موجود در تابع هزینه را به صورت جداگانه رسم نمود. در مقایسه با مطالعات [۱۰] و [۱۱]، پارامترهای به دست آمده در این تحقیق، از تعداد غنی‌تری از سناریوهای اجتماعی و با بهره‌گیری از روش‌های غیرخطی و پیچیده‌تری محاسبه شده‌اند.

جدول ۴- نتایج بهینه‌سازی با الگوریتم ژنتیک (تعیین ضرایب مجهول)

W_i	صحبت کردن	۴.۳۵
	دست تکان دادن	۴.۰۴
	اشاره کردن	۴.۷۸
	ورود سریع	۱.۴۸
	ورود آهسته	۱.۹۸
	خروج سریع	۰.۳۰
	خروج آهسته	۰.۵۹
	بی حرکت ماندن	۱.۰۵
	بیرون از صحنه بودن	۰
	اشاره شدن توسط فرد نزدیک و مستقیم	۱.۶۰
	اشاره شدن توسط فرد دور و مستقیم	۱.۷۶
	اشاره شدن توسط فرد نزدیک و زاویه بسته	۰.۲۹
	اشاره شدن توسط فرد دور و زاویه بسته	۰.۳۱
	P	نزدیک
دور		۱
Θ	مستقیم	۴.۹۸
	زاویه بسته	۱

جدول ۵- نتایج آزمون paired t-test بر پرسشنامه‌ی خروجی پیاده‌سازی

شماره سوالات پرسشنامه	ویدیوی اول		ویدیوی دوم		p-value
	میانگین	انحراف معیار	میانگین	انحراف معیار	
۱	۳.۶	۰.۸۴	۴	۱.۰۵	۰.۳۱
۲	۳.۵	۰.۹۷	۴.۵	۰.۷۱	۰.۰۱
۳	۳.۵	۱.۰۸	۴.۲	۰.۹۲	۰.۰۷
۴	۲.۹	۱.۲۰	۳.۵	۰.۹۷	۰.۰۵
۵	۳	۱.۰۵	۲.۹	۱.۲۹	۰.۸۷
۶	۲.۶	۱.۲۶	۲.۴	۱.۰۷	۰.۶۸
۷	۳.۳	۱.۱۶	۳.۹	۱.۴۵	۰.۳۶
۸	۳.۳	۱.۰۶	۳.۵	۰.۹۷	۰.۴۴
۹	۳.۳	۱.۰۶	۳.۷	۱.۳۴	۰.۴۶
۱۰	۳.۵	۱.۱۸	۳.۴	۱.۴۳	۰.۸۲

همان‌طور که در جدول ۵ مشاهده می‌شود، در خصوص سوالات ۲، ۴ (و تا حدودی ۳) با توجه به این که میانگین امتیاز افراد به ویدیوی اول با میانگین امتیاز افراد به ویدیوی دوم تفاوت قابل توجهی دارد و مقدار p-value برای این سوالات کمتر (و یا در نزدیکی) ۰.۰۵ می‌باشد، می‌توان نتیجه گرفت که روش تصادفی وزن‌دار در این موارد موفق عمل کرده و توانسته تفاوت معنادار رقم بزند. به عبارت دیگر، از نظر شرکت‌کنندگان پرسشنامه، روش تصادفی وزن‌دار باعث (۱) هماهنگی بهتر ربات با افراد ویدیو، (۲) درک بهتر افراد حاضر در ویدیو و (۳) تعاملی که شباهت بیشتری به یک انسان واقعی دارد، شده است.

۵- جمع‌بندی

در این پژوهش، هدف استخراج یک الگوی حرکت-زمان از رفتار جهت نگاه انسان در موقعیت‌های اجتماعی مختلف بود. با اجرای آزمایش این امکان فراهم شد تا با بهبود ویژگی‌های آزمایش و تغییر روش مدل‌سازی موجود در ادبیات موضوع، یک الگوی واقع‌گرایانه‌تر نسبت به پژوهش‌های گذشته ارائه شود. ویژگی اصلی این الگو مشخص بودن رابطه‌ی آن با محرک‌های اجتماعی حاضر در تصویر بود که باعث کاربردی شدن این الگو گردید. نتایج این پژوهش، قابلیت پیاده‌سازی بر روی ربات‌های اجتماعی به عنوان سیستم کنترل جهت نگاه مشابه انسان‌ها را دارد.

با پیاده‌سازی الگوی به دست آمده بر روی ربات نائو و ارزیابی خروجی آن توسط پرسشنامه ارائه شده، اثرگذاری قابل

محاسبه ضریب همبستگی پیرسون^{۱۸} می‌توان مقدار تطبیق دو سری داده را به صورت کمی مشاهده کرد. مقدار این ضریب برای این دو سری داده حدود ۰.۸۳ به دست آمد که نشان‌دهنده‌ی همبستگی و به تبع آن تطبیق بالای این دو سری داده است. برای مقایسه با پژوهش Aliasghari و همکاران [۱۱]، ضرایب به دست آمده در جدول ۴ در رابطه‌ی ۱ قرار داده شد و مجدداً ضریب همبستگی پیرسون محاسبه گردید که این بار عدد ۰.۶۷ به دست آمد که نشان از بهبود نسبی مدل پیشنهادی پژوهش فعلی دارد.

۴-۳- نتایج نحوه‌ی انتخاب هدف نگاه

هر دو روش انتخاب هدف مذکور بر روی ویدیوی آزمایش اجرا گردید. همانطور که از مقایسه‌ی شکل‌های ۱۴ و ۱۵ قابل برداشت است، فردی که در روش حریم‌بانه وارد تصویر شده است، نتوانسته توجهی را به خودش جلب کند؛ در صورتی که این اتفاق در روش تصادفی وزن‌دار افتاده است؛ به عبارتی این مشاهده حاصل شد که روش تصادفی وزن‌دار در عمل واقع‌گرایانه‌تر رفتار می‌کند.



شکل ۱۴- پیاده‌سازی روش انتخاب هدف تصادفی وزن‌دار بر روی ویدیوی آزمایش



شکل ۱۵- پیاده‌سازی روش انتخاب هدف حریم‌بانه بر روی ویدیوی آزمایش

۴-۴- نتایج پرسشنامه

^{۱۸} Pearson Correlation Coefficient

ملاحظه‌ی الگوی پیاده شده در ۳ مورد از ۱۰ مورد پرسشنامه تایید گردید.

عصبی موفقیان برای همکاری با پژوهشگران این مطالعه جهت استفاده از دستگاه ردیاب مردمک چشم قدردانی می‌گردد.

۶- محدودیت‌ها و کارهای آینده

وجود محدودیت در هر پژوهش امری اجتناب ناپذیر است و این پژوهش هم از این واقعیت مستثنی نیست. در این قسمت به طور خلاصه به محدودیت‌هایی که در این پژوهش وجود داشت، اشاره می‌گردد:

- تعداد محرک‌های اجتماعی بررسی شده؛ محرک‌های اجتماعی بررسی شده در این پژوهش عبارت‌اند از: صحبت کردن، اشاره کردن، دست تکان دادن، ورود به صحنه و خروج از آن (در دو سرعت مختلف). این محرک‌ها تعداد زیادی از رفتارهای اساسی انسانی در موقعیت‌های اجتماعی را شامل می‌شوند؛ اما باید دقت داشت که خروجی این پژوهش برای سایر محرک‌های اجتماعی (پیش بینی شده و نشده) قابل تعمیم نیست.
- بازه‌ی سنی شرکت‌کنندگان در آزمایش؛ میانگین سنی شرکت‌کنندگان در آزمایش برابر با ۲۵.۴ سال و انحراف معیار شرکت‌کنندگان برابر با ۳.۳ سال است. مسن‌ترین شرکت‌کننده ۳۴ سال و جوان‌ترین شرکت‌کننده ۲۱ سال دارد. بنابراین ممکن است نتوان نتایج این پژوهش را به افراد تمام سنین تعمیم داد.

موارد زیر جهت ادامه این پژوهش پیشنهاد می‌گردد:

- بهبود آزمایش ردیابی چشم با استفاده از ردیاب چشم قابل جابجایی، به جای ردیاب چشم ثابت
- ساخت یک ویدیوی واقع‌گرایانه‌تر (غیر کارتونی) برای آزمایش ردیابی چشم
- در نظر گرفتن محرک‌های اجتماعی غیر انسانی، به جای پرداختن مجدد به محرک‌های انسانی
- معرفی روش‌های جدید مدلسازی و بهره‌گیری از الگوریتم‌های مبتنی بر یادگیری عمیق برای محاسبه‌ی امتیاز جلب توجه

تقدیر و تشکر:

این پژوهش با حمایت گرنت اخذ شده از معاونت پژوهشی دانشگاه صنعتی شریف (به شماره G۹۸۰۵۱۷) و همچنین گرنت یادبود دکتر علی اکبر سیاسی انجام شده است. از مرکز تحقیقات

منابع و مراجع:

- [۱] C. Breazeal, K. Dautenhahn, and T. Kanda, "Social Robotics," *Springer Handbook of Robotics*. Springer International Publishing, pp. ۱۹۳۵-۱۹۷۲, ۲۰۱۶. doi: ۱۰.۱۰۰۷/۹۷۸-۳-۳۱۹-۳۲۵۵۲-۱_۷۲.
- [۲] H. Admoni and B. Scassellati, "Social Eye Gaze in Human-Robot Interaction: A Review," *J. Human-Robot Interact.*, vol. ۶, no. ۱, p. ۲۵, ۲۰۱۷, doi: ۱۰.۵۸۹۸/jhri.۶.۱.admoni.
- [۳] W. Liu, M. Li, and L. Yi, "Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework," *Autism Res.*, vol. ۹, no. ۸, pp. ۸۸۸-۸۹۸, ۲۰۱۶, doi: ۱۰.۱۰۰۲/aur.۱۶۱۵.
- [۴] W. Jones and A. Klin, "Attention to eyes is present but in decline in ۲-۶-month-old infants later diagnosed with autism," *Nature*, vol. ۵۰۴, no. ۷۴۸۰, pp. ۴۲۷-۴۳۱, ۲۰۱۳, doi: ۱۰.۱۰۳۸/nature۱۲۷۱۵.
- [۵] H. Admoni and B. Scassellati, "Social Eye Gaze in Human-Robot Interaction: A Review," *J. Human-Robot Interact.*, vol. ۶, no. ۱, p. ۲۵, ۲۰۱۷, doi: ۱۰.۵۸۹۸/jhri.۶.۱.admoni.
- [۶] Y. Yoshikawa, K. Shinozawa, H. Ishiguro, N. Hagita, and T. Miyamoto, "Responsive robot gaze to interaction partner.," in *Robotics: Science and systems*, ۲۰۰۶, pp. ۳۷-۴۳.
- [۷] M. Cornia, L. Baraldi, G. Serra, and R. Cucchiara, "Predicting human eye fixations via an LSTM-Based saliency attentive model," *IEEE Trans. Image Process.*, vol. ۲۷, no. ۱۰, pp. ۵۱۴۲-۵۱۵۴, ۲۰۱۸, doi: ۱۰.۱۱۰۹/TIP.۲۰۱۸.۲۸۵۱۶۷۲.
- [۸] S. Lathuilière, B. Massé, P. Mesejo, and R. Horaud, "Neural network based reinforcement learning for audio-visual gaze control in human-robot interaction," *Pattern Recognit. Lett.*, vol. ۱۱۸, pp. ۶۱-۷۱, ۲۰۱۹, doi: ۱۰.۱۰۱۶/j.patrec.۲۰۱۸.۰۵.۰۲۳.
- [۹] B. Yoo and J. Kim, "Fuzzy Integral-Based Gaze Control of a Robotic Head for Human Robot Interaction," *IEEE Trans. Cybern.*, vol. ۴۵, no. ۹, pp. ۱۷۶۹-۱۷۸۳, ۲۰۱۵, doi: ۱۰.۱۱۰۹/TCYB.۲۰۱۴.۲۳۶۰۲۰۵.
- [۱۰] A. Zaraki, D. Mazzei, M. Giuliani, and D. De Rossi, "Designing and Evaluating a Social Gaze-Control System for a Humanoid Robot," *IEEE Trans. Human-Machine Syst.*, vol. ۴۴, no. ۲, pp. ۱۵۷-۱۶۸, ۲۰۱۴.
- [۱۱] P. Aliasghari, A. Taheri, A. Meghdari, and E. Maghsoodi, "Implementing a gaze control system on a social robot in multi-person interactions," *SN Appl. Sci.*, vol. ۲, no. ۶, ۲۰۲۰, doi: ۱۰.۱۰۰۷/s۴۲۴۵۲-۰۲۰-۲۹۱۱-۰.
- [۱۲] J. D. Domingo, J. Gómez-García-Bermejo, and

- E. Zalama, "Optimization and improvement of a robotics gaze control system using LSTM networks," *Multimed. Tools Appl.*, vol. 81, no. 3, pp. 3351-3368, 2022, doi: 10.1007/s11042-021-11112-7.
- [13] A. Esfandbod, A. Nourbala, Z. Rokhi, A. F. Meghdari, A. Taheri, and M. Alemi, "Design, Manufacture, and Acceptance Evaluation of APO: A Lip-syncing Social Robot Developed for Lip-reading Training Programs," *Int. J. Soc. Robot.*, 2022, doi: 10.1007/s12369-022-00933-7.
- [14] S. Basiri, A. Taheri, A. Meghdari, and M. Alemi, "Design and Implementation of a Robotic Architecture for Adaptive Teaching: a Case Study on Iranian Sign Language," *J. Intell. Robot. Syst.*, vol. 102, no. 2, p. 48, 2021, doi: 10.1007/s10846-021-01413-2.
- [15] "NAO the humanoid and programmable robot | Aldebaran." <https://www.aldebaran.com/en/nao> (accessed May 22, 2023).
- [16] "Nao (robot) - Wikipedia." [https://en.wikipedia.org/wiki/Nao_\(robot\)](https://en.wikipedia.org/wiki/Nao_(robot)) (accessed May 22, 2023).

موسسه
پژوهشی
فصلنامه