دانشگاه صنعتی امیرکبیر

دانشکده مهندسی کامپیوتر و فناوری اطلاعات

پایان نامه کارشناسی ارشد

گرایش هوش مصنوعی

تشخیص بی‌درنگ حرکات دست در تعامل انسان با کامپیوتر

نگارش

سید ناصر نوراشرف‌الدین

استاد راهنما

دکتر رضا صفابخش

اسفند ۸۶

بسمه تعالی

دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)
معاونت پژوهشی

فرم اطلاعات پایان نامه
کارشناسی ارشد و دکترا

تاریخ: .....

پیوست: .....

نام و نام خانوادگی:	سیدناصر نوراشرف‌الدین	**دانشجوی آزاد** ■	**بورسیه** □	**معادل** □
شماره دانشجوئی:	۸٤۱۳۱۰۳۷	**دانشکده:** مهندسی کامپیوتر	**رشته تحصیلی:** هوش مصنوعی	
نام و نام خانوادگی استاد راهنما:	رضا صفابخش			
عنوان پایان نامه به فارسی:	تشخیص بی‌درنگ حرکات دست در تعامل انسان با کامپیوتر			
عنوان پایان نامه به انگلیسی:	Real-time hand gesture recognition in human-computer interaction			
نوع پروژه:	کارشناسی ارشد ■ / دکترا □	**کاربردی** ■	**بنیادی** □	**توسعه ای** □ **نظری** □
تاریخ شروع:	۸۵/۱۱/۲۰	**تاریخ خاتمه:** ۸۶/۱۲/۲۵	**تعداد واحد:** ٦	
سازمان تأمین کننده اعتبار:				
واژه های کلیدی به فارسی:	تعامل انسان-کامپیوتر، حرکتهای ایستا، حرکتهای پویا، بینایی ماشین، مدل رنگ پوست.			
واژه های کلیدی به انگلیسی:	Human-computer interaction, static gestures, dynamic gestures, computer vision, human skin model.			

نظرها و پیشنهادها به منظور بهبود فعالیت های پژوهشی دانشگاه:

استاد راهنما:

دانشجو:

امضاء استاد راهنما: **تاریخ:**

نسخه ۱: معاونت پژوهشی
نسخه ۲: کتابخانه و به انضمام دو جلد پایان نامه به منظور تسویه حساب با کتابخانه و مرکز اسناد و مدارک علمی

تقدیم به پدر و مادر عزیزم.

تقدیر و تشکر

بدینوسیله از راهنمایی ها و زحمات استاد عزیز و گرامی **دکتر رضا صفابخش** که به طرق مختلف در مسیر انجام این پایان نامه من را یاری نمودند، صمیمانه تشکر و قدردانی می‌نمایم. همچنین از تمامی دوستان عزیزی که در تهیه پایگاه داده این پروژه همکاری داشتند نیز صمیمانه تشکر می‌کنم. این پایان نامه بر اساس قرارداد شماره ۵۰۰/۱۲۳۵۱/ت مورخ ۸۶/۸/۲۰ تحت حمایت مالی مرکز تحقیقات مخابرات ایران انجام شده است.

بدینوسیله اینجانب سیدناصر نوراشرف‌الدین تعهد می‌نمایم که مطالب ارائه شده در این پایان‌نامه حاصل کار پژوهشی و تحقیق اینجانب می‌باشد و قبلا برای احراز مدرک دیگری ارائه نشده است. رجوع به دست‌آوردهای پژوهشی دیگران که در این پایان‌نامه از آنها استفاده شده، مطابق مقررات ارجاع داده شده است. چنانچه در هر شرایطی این موارد بدرستی رعایت نگردد، دانشگاه مجاز به ابطال پایان‌نامه خواهد بود. کلیه حقوق مادی و معنوی این اثر متعلق به دانشکده مهندسی کامپیوتر دانشگاه صنعتی امیرکبیر می‌باشد.

تاریخ: نام و نام‌خانوادگی دانشجو: سیدناصر نوراشرف‌الدین

امضاء:

چکیده

در سالهای اخیر تحقیقات زیادی برای استفاده از ابزارها و روش های جدید برای ارتباط با کامپیوتر انجام شده است. در حال حاضر برای برقراری ارتباط با وسایلی مانند کامپیوتر، رباتها و یا سایر وسایل الکترونیکی نظیر تلویزیون از ابزارهای واسطی مانند موشواره و صفحه کلید، اهرمهای مکانیکی و الکتریکی و یا کنترلهای از راه دور استفاده می‌شود. با پیشرفت فناوری در آینده نزدیک شاهد خواهیم بود که انسانها بسیاری از کارهای روزمره خود را با استفاده از این وسایل انجام می‌دهند و بنابراین اگر بتوانند با این وسایل به همان شیوه ای که انسانها با یکدیگر تعامل دارند، ارتباط برقرار کنند، کارها بسیار راحت‌تر و بدون نیاز به ابزارهای واسط قابل انجام خواهد بود.

در این پایان نامه یک الگوریتم مبتنی بر بینایی ماشین ارائه شده است که می‌تواند بصورت بی‌درنگ حرکتهای دست کاربران را تشخیص داده و معادل دستوری آنها را در کامپیوتر اجرا کند. در ابتدا با استفاده از روش تفاضل پس‌زمینه، تصویر کاربر از زمینه جدا و سپس با استفاده از یک مدل گاسی، پیکسلهایی که به رنگ پوست انسان هستند مشخص می‌شوند. از ویژگی های مهم الگوریتم پیشنهادی که برتری آن نسبت به سایر الگوریتم های ارائه شده نیز بشمار می‌رود، عدم نیاز به داشتن تصویر زمینه از قبل می‌باشد. در این روش الگوریتم خود از روی فریمهای اولیه تصویر زمینه را می‌سازد و در فریمهای بعدی آن را بروزرسانی می‌کند. در مرحله بعد با استفاده از عملگرهای ریخت شناسی نویزهای تصویر و ناحیه‌های آویزان دست حذف می‌شود. پس از تقطیع دستها، کف دست از بقیه قسمتهای آن مانند ساعد و بازوها جدا می‌شود. برای شناسایی دستها با استفاده از گشتاورهای مرکزی نرمال شده و گشتاورهای پیرامون، بردار ویژگی هر دست استخراج شده و این بردارهای ویژگی به عنوان ورودی به شبکه های LVQ داده می‌شوند که این شبکه ها نوع حرکت کاربر را تشخیص می‌دهند. کلیه حرکتهای این سامانه ایستا بوده و کاربران حرکتها را با دو دست انجام می‌دهند. این الگوریتم می‌تواند ۲۵ حرکت مختلف را که توسط پنج وضعیت پایه انجام می‌شود را با سرعت ۲/۵ فریم در ثانیه شناسایی کند. الگوریتم ارائه شده می‌تواند در حالت برون-خط با دقت ۹۷/٦ و در حالت برخط با دقت ۹٤ درصد حرکتهای کاربران را تشخیص دهد. در این الگوریتم مشکل همپوشانی صورت و دستهای کاربر حل شده و برای نور محیط هیچ نورپردازی خاصی در نظر گرفته نشده و سامانه می‌تواند در نورپردازی های مختلف کار کند.

واژگان کلیدی: تعامل انسان-کامپیوتر، حرکتهای ایستا، حرکتهای پویا، بینایی ماشین، مدل رنگ پوست.

فهرست مطالب

فهرست شکلها

فهرست جداول

فصل اول

۱- مقدمه

از ابتدای تولید کامپیوترها یکی از مهمترین فعالیتهای طراحان این بوده است که کاربران بتوانند هر چه ساده‌تر با کامپیوترها ارتباط برقرار کرده و به راحتی از قابلیتهای آن استفاده کنند. در ابتدا کاربران تنها از طریق وسائل مکانیکی خاصی با کامپیوترها تعامل داشتند و بتدریج با پیشرفت فناوری ابزارهای جدیدی نظیر موشواره، صفحه کلید و قلم نوری نیز در اختیار کاربران قرار گرفت. متداول ترین شیوه تعامل با کامپیوترهای فعلی که برای چندین دهه است که تغییرات زیادی نیز نداشته، استفاده از سخت افزارهایی نظیر موشواره و صفحه کلید می‌باشد. با وجود سادگی این وسایل، هنوز هم افراد زیادی هستند که در استفاده از کامپیوترها مشکل داشته و حتی استفاده از این ابزارها برای کاربران حرفه‌ای نیز در بعضی از موارد خسته-کننده می‌باشد. امروزه با پیشرفت فناوری شاهد هستیم که کامپیوترها در هر ابعادی به وسایل جانبی نظیر دوربین، میکروفون و سنسورها مجهز هستند و از طرف دیگر آمار و ارقام نیز نشان می‌دهد که در آینده‌ای نزدیک انسانها بیشتر وقت خود را با کامپیوترها می‌گذرانند تا با سایر انسانها. بنابراین این نیاز احساس خواهد شد که انسانها بتوانند طوری با کامپیوترها ارتباط برقرار کنند که گویی با انسانهای دیگر تعامل دارند.

از طرف دیگر با توجه به افزایش روز افزون نقش رباتها در زندگی مدرن امروزی و گسترش زمینه‌های کاربرد آنها به منظور تسهیل شرایط زندگی، ایجاد زمینه و راهکارهای مناسب برای برقراری ارتباط با آنها به بهترین و راحت‌ترین روش ممکن از جمله چالش های مهم در فناوریهای امروزی است. نیازهای بشر امروزی با توجه به افزایش روزافزون فناوری، انتظارات بیشتری را در این شیوه ارتباطی ایجاد کرده است. بنابراین یک فناوری پیشرفته در طراحی یک ربات شدیداً تحت تاثیر ارتباطی است که محصول مورد نظر، به عنوان یک وسیله هوشمند با کاربر خود برقرار خواهد کرد. هر چقدر شیوه برقراری این ارتباط طبیعی‌تر بوده و به نحوه ارتباط مابین انسان ها نزدیک‌تر باشد، ارتباط مورد نظر مناسب‌تر و راحت‌تر ارزیابی می‌شود و کاربران تمایل بیشتری برای استفاده از این محصول خواهند داشت. برای تحقق این موضوع لازم است که کامپیوترها (یا رباتها) مشابه انسانها عمل کرده و کارهای پایه‌ای نظیر دیدن، تشخیص صدا، صحبت کردن و لمس کردن را یاد بگیرند. همچنین همانطور که ما انسانها در برخورد با رفتارهای متفاوت به شیوه-های مختلفی عمل می‌کنیم، لازم است که کامپیوترها نیز بتوانند کاربران متفاوت خود را مدل کرده و با هر کاربر مطابق نیازهای آن رفتار کنند.

امکان بینایی برای کامپیوترها باعث می‌شود که این دستگاهها بتوانند جهان اطراف خود را ببینند و تغییرات آن را درک کنند. این قابلیت کامپیوترها را قادر خواهد کرد که حرکتهای کاربران خود را حس کنند و پس از تجزیه و تحلیل تصاویر، حرکتهای کاربران را فهمیده و دستورها و پیغامهای موجود در تصاویر را استخراج و اجرا کنند. برای تولید اینگونه واسطهای[1] مبتنی بر بینایی، آشنایی با مباحث بینایی ماشین، پردازش تصویر، گرافیک کامپیوتری، علوم زبان شناسی و روانشناسی مورد نیاز است. با وجود اینکه پیشرفتهای زیادی در این مباحث و علوم صورت گرفته، اما تعامل مبتنی بر بینایی[2] پیشرفتهای زیادی نداشته است و مشکلات زیادی برای تولید یک واسط کاربری مناسب وجود دارد. یکی از مهمترین مشکلاتی که برای تولید این واسطها وجود دارد این است که حرکت انسانها دارای پیچیدگی‌های زیادی است و الگوریتم جامعی برای مدل کردن تمامی حرکتها وجود ندارد. همچنین میزان تغییرات حرکتها در انسانهای مختلف زیاد می‌باشد که این امر نیز مدل کردن حرکتها را دشوار ساخته است. درنتیجه تولید واسطهایی که بتوانند حرکتهای زیادی را با هر درجه پیچیدگی بصورت بی‌درنگ[3] فهمیده و دستورات معادل آنها را اجرا کنند، کاری دشوار است. چند نمونه از حرکتهایی که برای آنها مدلها و الگوریتمهایی ارائه شده است عبارتند از: حرکت دست، مدل کردن حرکت سر و صورت، حرکت اندامهای صورت مانند لبها، ابروها و چشمها که در ادامه استفاده از این حرکتها را بطور مختصر بررسی می‌کنیم.

استفاده از حرکات دست بعد از صحبت کردن بیشترین کاربرد را در برقراری ارتباط بین انسانها دارد و از گذشته‌های دور به کسانی که توانایی سخن گفتن را نداشتند حرکات ایما و اشاره‌ای آموزش داده می‌شد که بتوانند با حرکتهای مختلفی که با دست انجام می‌دهند، منظور خود را برسانند. حتی انسانهایی که توانایی سخن گفتن را نیز دارند در هنگام صحبت کردن و تعامل با سایر انسانها از حرکات دست استفاده می‌کنند. بطور مثال ما برای خواستن چند شئ، چند انگشت دست خود را نشان می‌دهیم و یا برای اینکه توجه شخصی را به چیزی جلب کنیم، با انگشت دست خود به آن اشاره می‌کنیم. در حقیقت اشاره کردن با دست آنقدر در بین انسان ها معمول است که بعضی از آنها حتی وقتی با تلفن صحبت می‌کنند نیز با دست خود حرکات اشاره ای انجام می‌دهند. در یک واسط بینایی مبتنی بر حرکت دست، معمولا سه دسته اطلاعات از تصاویر استخراج می‌شود. اول اینکه در هر تصویر مکان دست در تصویر تشخیص داده می‌شود و سپس از روی مکان دست در فریمهای متوالی تغییرات مکانی دست مشخص می‌شود. همچنین در

[1] Interface
[2] VBI (vision-based interaction)
[3] Real-time

بسیاری از این سامانه ها (سیستم ها) هر وضعیت[1] دست برای خود یک معنا و مفهوم جداگانه ای دارد و بنابراین تشخیص نوع حرکت نیز یکی از عملیاتهای مورد نیاز برای این واسط ها می‌باشد. در بعضی از الگوریتمها یک سری اطلاعات پارامتریک دیگر مانند جهتی که انگشت دست اشاره می‌کند نیز استخراج می‌شود.

در استفاده از حرکت لبها در ابتدا باید به این موضوع اشاره کنیم که تشخیص گفتار[2] یکی دیگر از روشهای تعامل انسان با کامپیوتر (HCI[3]) می‌باشد که با استفاده از این روش، کامپیوتر قادر خواهد بود که صحبتهای کاربران را شنیده و اطلاعات گفتاری آنها را استخراج کند. در حال حاضر بسیاری از سامانه های تشخیص گفتار موجود، تنها در محیط آزمایشگاهی عملکرد بسیار خوبی دارند و در محیطهای واقعی بدلیل نویز بالای محیط، این سامانه ها بخوبی عمل نکرده و کارایی آنها کاسته می‌شود. در مقابل این سامانه ها، انسانها قادرند که با ترکیب سیگنالهای گفتار و اطلاعات بصری دیگر مانند حرکت بدن و صورت، در محیطهای نویزی نیز تشخیص گفتار را بصورت قابل قبولی انجام دهند که این امر در بین افرادی که توانایی شنیدن را ندارند بوضوح دیده می‌شود. با الهام گرفتن از نحوه عملکرد انسان در تشخیص گفتار و اینکه اطلاعات بصری گفتار در محیطهای نویزی تغییر نکرده و حساس به نویزهای صوتی نیستند، این انگیزه در بین محققان بوجود آمد که از اطلاعات بصری نیز در سامانه های تشخیص گفتار استفاده کنند. از آنجا که سیگنالهای صوتی از مجرای گفتار انسان خارج می‌شوند، در تمامی سامانه های تشخیص گفتار مبتنی بر اطلاعات بصری (سامانه های صوتی- بصری[4]) سعی شده است که حرکت دهان و لبها مدل شود و اطلاعات استخراج شده از این حرکتها با اطلاعات صوتی ترکیب شده و یک سامانه مقاوم در مقابل نویز ایجاد شود. در واقع در اینگونه سامانه ها، تعامل بصورت ترکیبی از بینایی ماشین و سیگنالهای صوتی می-باشد.

یکی دیگر از حرکتهایی که در واسطهای انسان – کامپیوتر مورد استفاده قرار گرفته است، حرکت اندامهای صورت بخصوص چشمها و ابروها می‌باشد. اگر چه استفاده از حرکات دست بیشترین کاربرد را در واسطهای مبتنی بر بینایی دارد، اما اخیرا محققان در پی این بوده‌اند که با ابداع روشهای جدید، واسطهای موجود را طبیعی‌تر کرده و هر چه بیشتر به شیوه تعامل انسان- انسان نزدیک‌تر کنند. یکی از این روشها استفاده از حرکت چشم برای انتخاب یک منو بر روی صفحه نمایش می‌باشد. انسانها برای اینکه کاری را

[1] Posture
[2] Speech Recognition
[3] Human-Computer Interaction (HCI)
[4] Audio-Visual

روی یک شی انجام دهند ابتدا به شی نگاه کرده و سپس با استفاده از حرکت دست، کاری را روی شی انجام می‌دهند. همچنین هنگامی که یک کاربر از کامپیوتر استفاده می‌کند، معمولا ابتدا به قسمت مشخصی از صفحه نمایش نگاه کرده و سپس کار خود را انجام می‌دهد. در بعضی از واسطهای مبتنی بر حرکت دست، کاربر ابتدا با نگاه کردن شی مورد نظر خود را پیدا کرده و سپس با حرکت دست کاری را روی آن انجام می‌دهد و نگاه کردن به شی قبل از و سریعتر از حرکت دست انجام می‌شود، مانند محیطهایی که برای واقعیتهای مجازی[۱] طراحی شده‌اند. همچنین برای معلولینی که توانایی حرکت دادن دستهای خود را ندارند نگاه کردن راحت‌ترین کار ممکن برای استفاده از کامپیوتر می‌باشد. روش کار با واسطهای مبتنی بر حرکت چشم بدین صورت است که روی صفحه نمایش منوهای سامانه گنجانده می‌شود و کاربر در زمانی حدود چند ثانیه به این منوها خیره شده و سپس سامانه منوی مورد نظر کاربر را انتخاب می‌کند. همچنین چند الگوریتم ترکیبی نیز ارائه شده است که در این روشها با استفاده از نگاه کردن و یک ورودی دیگر مانند حرکت دست، حرکت ابروها و یا استفاده از موشواره، کاربر سریعتر کارهای خود را انجام می‌دهد. در چند نمونه از این سامانه ها محققان با استفاده از روش برق نگاری ماهیچه (EMG[۲]) تغییر سیگنالهای الکتریکی حرکات صورت را اندازه گرفته و با تلفیق این روش با حرکت چشم، واسطهایی را طراحی کردند. ایده این موضوع از روی تعاملهای انسان با انسان گرفته شده است، بدین صورت که در بسیاری از موارد انسانها با استفاده از حرکتهای قسمتهای مختلف صورت مطلبی را بیان کرده و یا حالتی را نشان می‌دهند. در یکی از این واسطها با استفاده از روش EMG تغییر سیگنالهای الکتریکی ناشی از حرکت ابرو با داده های خیره شدن چشم ترکیب شده و کاربر بعد از اینکه به یک شی خاص خیره می‌شود با حرکت دادن ابروهای خود آن شی را انتخاب می‌کند[۱].

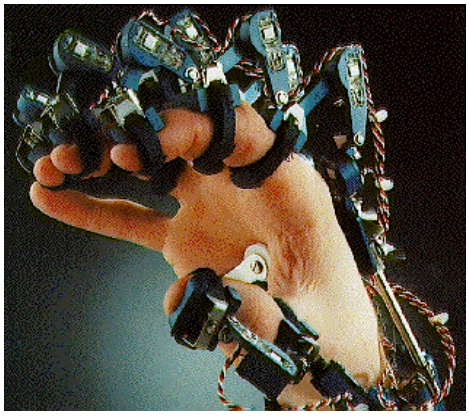
از حرکتهای سر و صورت نیز در سامانه های HCI استفاده شده است. در این سامانه ها کاربر می‌تواند با حرکت دادن سر و صورت در جهتهای مختلف با سامانه ارتباط برقرار کند و دستوراتی را به سامانه منتقل نماید. یکی از مشکلات این سامانه ها این است که تعداد حرکتهای مختلفی که می‌توان با سر و صورت انجام داد محدود است و کاربران نیز تمایل زیادی به استفاده از حرکت سر وصورت برای تعامل ندارند، ولی می‌توان این حرکتها را بصورت ترکیبی با سایر حرکتها بکار برد.

از بین تمامی حرکتهایی که در این بخش به آن اشاره شد، حرکت دست بیشترین کاربرد را در واسطهای انسان – کامپیوتر داشته است و بیشترین الگوریتمها نیز در مورد حرکت دست ارائه شده است.

[۱] Virtual Reality
[۲] Electromyography (EMG)

علت اینکه حرکت دست بیشترین کاربرد را دارد یکی این است که تعداد حرکتهایی را که می‌توان با دست انجام داد از اندامهای دیگر بدن بیشتر است و این تنوع حرکتها استفاده از سامانه ها را آسانتر می‌کند و اجازه می‌دهد که قابلیتهای بیشتری در سامانه گنجانده شود. دلیل دوم این است که یادگیری حرکتهای دست برای کاربران راحت‌تر بوده و حتی کاربران غیر حرفه‌ای نیز می‌توانند در زمان کوتاهی تمامی حرکتها را یاد بگیرند که این موضوع باعث شده است که کاربران تمایل بیشتری برای استفاده از حرکتهای دست نشان دهند. در ادامه استفاده از حرکتهای دست در سامانه های HCI را بررسی می‌کنیم.

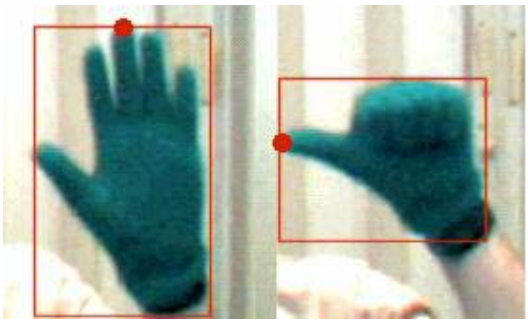
همانطور که پیشتر نیز اشاره شد، در چند دهه اخیر استفاده از حرکتهای دست در سامانه های HCI بسیار مورد توجه قرار گرفته است. در ابتدا برای اینکه کامپیوتر (ماشین) بتواند حرکتهای دست را تفسیر کند از یک سری دستگاههای الکترومغناطیسی استفاده می‌شد که می‌توانستند مکان دست و بازوها و زوایای بین آنها و حتی زوایای بین مفصلهای انگشتان را اندازه بگیرند. به این گروه از دستگاهها، دستکشهای داده‌ای[1] و به این روش برقراری ارتباط بین انسان و کامپیوتر، روش مبتنی بر دستکش[2] گفته می‌شود. دلیل این نامگذاری این است که اینگونه دستگاهها بصورت دستکشهایی هستند که سنسورهایی روی آن قرار دارد و با تعداد زیادی سیم به کامپیوتر متصل می‌شود و کاربران برای استفاده از سامانه باید آن را مانند یک دستکش بپوشند. استفاده از اینگونه دستکشها چندان مورد استقبال کاربران قرار نگرفت، چرا که استفاده از ابزارهایی نظیر موشواره و صفحه کلید بسیار ساده‌تر از این دستکشها بود و در حال حاضر از این دستکش‌ها تنها برای کارهای خاص و کنترل در بعضی محیط ها استفاده می‌شود. مثلاً چنین روشی را برای شبیه سازی جراحی در یک محیط مجازی می توان به کار برد. در شکل ۱-۱ یک نمونه از این دستکشها نشان داده شده است.



شکل ۱-۱ یک نمونه دستکش داده ای

[1] Data Gloves
[2] Glove-Based

برای برطرف کردن محدودیتهای دستکشهای داده‌ای، استفاده از بینایی ماشین برای تشخیص حرکتهای دست پیشنهاد شد. در این رویکرد پیشنهاد شد که با استفاده از دوربینهای ویدئویی و روشهای بینایی ماشین و پردازش تصویر، حرکتهای دست شناسایی و تفسیر شود. در ابتدا به جای استفاده از دستکشهای داده‌ای، پیشنهاد شد که برای متمایز کردن دستها از پس زمینه[۱] از دستکشهای رنگی استفاده شود و یا علامت‌های رنگی بر روی دست کاربران قرار داده شود. یک نمونه از این دستکشها در شکل ۱-۲ نشان داده شده است.



شکل ۱-۲ استفاده از دستکشهای رنگی به جای دستکشهای داده ای

استفاده از این دستکشها در مقایسه با دستکشهای داده‌ای بسیار مناسب‌تر می‌باشد ولی همین دستکشها نیز یک محدودیت برای سامانه‌های HCI به شمار می‌رود، چرا که کاربر برای استفاده از سامانه همیشه مجبور است از این دستکشها استفاده کند. برای حل این مشکل طراحان از سامانه بینایی انسان الهام گرفتند. ما انسانها می‌توانیم تنها با استفاده از شکل و رنگ دستها، مکان دست در تصویر را مشخص کنیم. این موضوع این ایده را در ذهن طراحان بوجود آورد که به جای استفاده از دستکشهای داده‌ای و رنگی، تنها از شکل دستها و رنگ پوست برای تشخیص مکان دست در تصویرها استفاده کنند.

در تحقیقات انجام شده از دو نوع حرکت دست در تعامل انسان با کامپیوتر استفاده می‌شود. در بعضی از سامانه ها ترکیب چند وضعیت[۲] دست در فریمهای متوالی معادل یک دستور در نظر گرفته می-شود، حال آنکه در دیگر سامانه ها، وضعیت دست در هر فریم معنا و مفهوم مستقلی دارد. به حرکت های نوع اول حرکتهای پویا[۳] و به نوع دوم حرکتهای ایستا[۴] گفته می‌شود. در حرکتهای ایستا تنها باید شکل دستها

[۱] Background
[۲] Posture
[۳] Dynamic Gesture
[۴] Static Gesture

در هر فریم شناسایی شود، حال آنکه برای حرکتهای پویا باید ردیابی دستها نیز انجام شود تا علاوه بر شکل دستها در فریم های متوالی، نحوه جابجایی آنها نیز مشخص شود.

تاکنون تحقیقات زیادی برای استفاده از حرکتهای دست انجام شده است. در[۲] آقای گوپتا الگوریتمی برای تشخیص ۱۰ حرف از زبان اشاره‌ای آمریکایی[۱] ارائه داده است. این زبان یکی از قویترین زبانهای اشاره ای می‌باشد که افراد ناشنوا می‌توانند برای بیان کلمات و جملات انگلیسی از آن استفاده کنند. در این الگوریتم از دست کاربران در زمینه سیاه تصویر گرفته می‌شود. پس از تصویر برداری، با استفاده از روش آستانه سازی هیستوگرام، تصویر دست از بقیه قسمتهای زمینه تقطیع می‌شود. آنگاه با استفاده از روش توالی پیرامونی محلی، بردارهای ویژگی از پیرامون دستها استخراج شده و با روش تطبیق خطی، وضعیت-های دست شناسایی می‌شود. در این الگوریتم تصویر برداری از حرکات دست در محیط آزمایشگاهی با زمینه ثابت و نور فلورسنت از بالا انجام شده است و شناسایی هر حرکت حدود ۲۰ ثانیه طول می‌کشد. در این الگوریتم تنها از حرکتهای ایستا استفاده شده است.

در [۳] آقای الجراح برای شناسایی ۳۰ حرف زبان اشاره‌ای عربی از شبکه‌های عصبی- فازی استفاده کرده است. در این الگوریتم نیز از تصاویر سطح خاکستری در زمینه ثابت استفاده شده است. برای تقطیع دستها از یک الگوریتم آستانه‌سازی تکراری استفاده می‌شود. برای استخراج ویژگی ابتدا مرکز ثقل دستها محاسبه شده و سپس بردارهایی از این نقطه به نقاط پیرامون وصل می‌شود. طول این بردارها مولفه-های بردار ویژگی را تشکیل می‌دهند. بردارهای ویژگی به ۳۰ شبکه داده شده و شبکه‌ای که خروجی بزرگتری تولید می‌کند، نوع حرکت دست را مشخص می‌نماید. در این الگوریتم نیز تنها از حرکتهای ایستا استفاده شده است.

در [٤] آقای تِنگ الگوریتمی بلادرنگ برای شناسایی ۳۰ حرف زبان اشاره ای چینی ارائه کرده است. در این سیستم از هیچ نشانه ای بر روی دستها استفاده نمی‌شود و زمینه تصاویر نیز ساده نیست و همچنین هیچ نورپردازی خاصی برای سیستم در نظر گرفته نشده است. برای تقطیع تصاویر دست، از روش تفاضل فریمهای متوالی و یک مدل رنگ پوست استفاده شده است. در این الگوریتم کل تصویر دست به عنوان بردار ویژگی در نظر گرفته می‌شود و برای شناسایی حرکتها از روش نزدیکترین همسایگی استفاده شده است. در این الگوریتم نیز تنها از حرکتهای ایستا استفاده شده است.

[۱] American Sign Language (ASL)

در [۵] آقای لیکسار و همکارش یک الگوریتم تطبیقی برای شناسایی ۹ وضعیت ایستا ارائه داده است. تاکید نویسندگان این مقاله بر ارائه سیستمی بوده است که بتواند حتی با تعداد کم نمونه های آموزشی نیز قدرت تعمیم خوبی داشته باشد و با دقت مناسبی حرکات کاربران جدید را تشخیص دهد. الگوریتم ارائه شده در این مقاله این قابلیت را دارد که بصورت برخط توسط کاربران آموزش مجدد داده شود. این آموزش می‌تواند در دو حالت با نظارت و بی نظارت انجام شود. حالت با نظارت زمانی فعال می‌شود که حرکتی اشتباه تشخیص داده شود، ولی زمانیکه سیستم حرکتها را درست تشخیص می‌دهد نیز آموزش انجام می‌شود (یادگیری بی نظارت). در حالت بی نظارت، اگر سیستم حرکتهای کاربر را درست تشخیص دهد، پارامترهای این حرکتها را جایگزین نمونه های ذخیره شده می‌کند. آزمایشات انجام شده در مقاله بدین صورت است که یک تصویر با استفاده از ویدئو پروژکتور بر روی دیوار نشان داده می‌شود و کاربر بصورت بلادرنگ می‌تواند اسلاید قبلی یا بعدی را نمایش دهد و یا اینکه بر روی اسلاید جاری تغییراتی اعمال کند. برای تقطیع دستها از روش تفاضل پس زمینه استفاده شده و تصویر پس زمینه همان تصویری است که ویدئو پروژکتور بر روی دیوار منعکس می‌کند. برای بردارهای ویژگی از ٦ مولفه اول توصیفگرهای فوریه استفاده شده و دسته بندی حرکتها با استفاده از روش نزدیکترین همسایگی انجام می‌شود.

آقای چانوا و همکارش در سال ۲۰۰۲ یک واسط گرافیکی ساده پیاده سازی کردند که کاربر می-تواند پنجره‌ها و اشیاء موجود در آن را تنها با استفاده از حرکات دست انتخاب کرده، جابجا کند و یا اندازه آنها را تغییر دهد[٦]. همچنین کاربر می‌تواند پنجره‌ها را باز و بسته نماید و یا کارهای انجام شده را به حالت اول بر گرداند. برای این واسط گرافیگی ١٤ حرکت در نظر گرفته شده که از بین این حرکتها، چهار حرکت ایستا و مابقی پویا هستند. همچنین شش حرکت این واسط تنها با یک دست انجام داده می‌شوند و برای مابقی، کاربر باید از هر دو دست خود استفاده کند. برای تقطیع تصویر دستها از پس‌زمینه، دو روش تفاضل پس‌زمینه و مدل رنگ پوست استفاده شده است. در این واسط محدودیتی برای پوشش کاربر در نظر گرفته نشده و از آنجا که کاربر می‌تواند از پیراهن های آستین کوتاه نیز استفاده کند، طراحان سیستم برای حذف قسمتهای ناخواسته دست از روشهای برش مچ استفاده کردند. برای استخراج ویژگیها از پیرامون دستها و تبدیل فوریه گسسته استفاده شده است. شناسایی دستها با کمک شبکه های عصبیRBF و مدل مخفی مارکوف انجام می‌شود و این واسط می‌تواند با دقت ۹۲ درصد و بصورت بی‌درنگ حرکتها را تشخیص داده و معادل دستوری آنها را اجرا کند.

در[۷] آقای یوانسین ژو و همکارش یک مرورگر نقشه[۱] پیاده سازی کردند که کاربر می‌تواند با استفاده از ۱۲ حرکت پویا تصاویر را به بالا، پایین، چپ و راست جابجا کند، بزرگ نمایی یا کوچک نمایی کرده و حتی در زوایای مختلف دوران دهد. کلیه حرکتهای این واسط با یک دست انجام می‌شود. برای تقطیع تصاویر دست از روش تفاضل فریمهای متوالی و یک جدول جستجو[۲] برای تشخیص نواحی به رنگ پوست استفاده شده است. از آنجا که حرکتهای این مرورگر نقشه حرکتهای پویا هستند، لازم است برای تشخیص حرکات هم اطلاعات مربوط به شکل دستها و هم اطلاعات مربوط به جابجایی دستها استخراج شود. برای استخراج ویژگیهای حرکتی از روش رگرسیون حرکت تصویر[۳] استفاده شده است. با استفاده از این روش، ویژگیهای مربوط به حرکت های دست مشخص می‌شود. برای استخراج ویژگیهای مربوط به وضعیت دستها از گشتاورهای هندسی مرتبه اول و دوم استفاده شده است. ترکیب ویژگیهای حرکتی و ویژگیهای مربوط به وضعیت دستها در هر فریم، بردارهای ویژگی را تشکیل می‌دهند. برای شناسایی حرکتها از روشی مشابه روش نزدیکترین همسایگی استفاده شده است. این واسط می‌تواند در حالت برخط با دقتی بین ۸۰ تا ۹۰ درصد و در حالت برون‌خط با دقت ۹۰/۸۳ درصد حرکتها را شناسایی کند. کل زمان لازم برای تقطیع تصاویر، استخراج ویژگی و شناسایی حرکتها حدود ۲ ثانیه می‌باشد.

در [۸] آقای استارنر و همکارانش دو سامانه برای شناسایی ٤٠ علامت زبان اشاره ای آمریکایی پیاده سازی کرده اند. در سامانه اول دوربین بر روی یک میز قرار دارد و کاربر در مقابل دوربین حرکتهای دست را انجام می‌دهد. در سامانه دوم یک دوربین بر روی کلاه کاربر نصب شده است که از بالا حرکتهای دست کاربر را تصویربرداری می‌کند. برای تقطیع تصاویر دست از مدل رنگ پوست و روش رشد دادن نواحی استفاده شده است. پس از تقطیع تصاویر دست ها، از گشتاورهای درجه دوم برای استخراج ویژگی استفاده شده و ١٦ ویژگی از هر دست استخراج می‌شود. در حالتی که همپوشانی دستها رخ می‌دهد تنها یک ناحیه به رنگ پوست استخراج می‌شود. دراین حالت ویژگیهای استخراج شده این ناحیه به هر دو دست تعلق می‌گیرد. علاوه بر این در نمونه های آموزشی نیز چند مورد همپوشانی در نظر گرفته شده است تا دقت سامانه ها بخاطر همپوشانی دستها کاهش نیابد. برای شناسایی حرکتها نیز از مدل مخفی مارکوف استفاده شده است. برای سامانه اول که دوربین بر روی میز قرار داده شده است، دقت ۹۲ درصد و برای سامانه دوم که دوربین بر روی کلاه کاربر تعبیه شده است، دقت ۹۷/۸ درصد می‌باشد. دلیل این برتری در این است که در سامانه دوم مشکل همپوشانی دستها کمتر وجود دارد.

[۱] Map Browser
[۲] Look up Table
[۳] Image Motion Regretion