راهکاری ساده و مؤثر برای تشخیص حروف الفبای فارسی در زبان اشاره حرکات انگشتان

محمدجواد برزگر سخویدی و احمدرضا شرافت

چکیده: در سال‌های اخیر، تشخیص حرکات اشاره (زبان اشاره) مورد توجه پژوهشگران قرار گرفته است. زبان اشاره، ترکیبی از حالات دست، حرکات دست و حالات چهره است. املای انگشتی، یک نمایش برای حروف الفبای کلماتی است که در لغت‌نامه زبان وجود ندارد. در این مقاله یک سامانه املای انگشتی برای تشخیص حروف الفبای فارسی ارائه شده که در آن برای هر حرف الفبا یک شکل دست در نظر گرفته شده است. این سامانه شامل پنج مرحله است: اول، جمع‌آوری داده تصویری؛ دوم، پیش‌پردازش؛ سوم، استخراج و آشکارسازی ویژگی‌های شکل دست؛ چهارم، کاهش اندازه بردار ویژگی و پنجم، پیاده‌سازی تشخیص با استفاده از سه روش نزدیک‌ترین همسایه (معیار فاصله اقلیدسی و معیار فاصله اقلیدسی نرمالیزه) و شبکه عصبی. در این مقاله از تبدیل کسینوسی گسسته (DCT) برای کاهش اندازه بردار ویژگی استفاده شده است که نسبت به روش‌های موجود، نظیر تبدیل فوریه گسسته و ضرایب توصیف‌گر فوریه عملکردی بهتر دارد. نتایج پیاده‌سازی با شبکه عصبی، دقت تشخیص حروف الفبا را ۹۹٫۱٪ نشان داده است که نسبت به عملکرد سامانه‌های موجود بهبود یافته است.

کلید واژه: **آشکارسازی پوست، املای انگشتی فارسی، تشخیص زبان اشاره، شبکه عصبی، نزدیک‌ترین همسایه.**

۱- مقدمه

رایانه‌ها و فن‌آوری‌های هوشمند بسیاری از زوایای زندگی انسان را تحت تأثیر قرار داده‌اند. هر روز شاهد فن‌آوری‌های جدید و هوشمند هستیم. اولین مسئله‌ای که با آن روبه‌رو هستیم، چگونگی تعامل و ارتباط با آنهاست. علم تعامل انسان با رایانه [1](HCI) به این موضوع می‌پردازد. هدف HCI، تقویت تعاملات کاربران و رایانه به‌وسیله کاربردی‌تر کردن رایانه‌ها و مطابقت آنها با نیاز کاربران است. تلاش‌های زیادی در زمینه HCI صورت گرفته تا تعامل انسان و رایانه به‌صورت طبیعی و بدون وسایل جانبی انجام شود، بدین معنی که تعامل انسان با رایانه همانند تعامل بین انسان‌ها باشد. همان‌طور که انسان‌ها با حرکات بدن نظیر حرکات دست و چهره و صدا با یکدیگر تعامل برقرار می‌کنند، می‌توان از این حرکات به‌عنوان ابزاری جهت ارتباط با رایانه استفاده کرد و دیگر نیاز به وسایل جانبی نظیر ماوس و صفحه کلید نباشد. در زبان اشاره به‌جای صدا از ترکیب شکل، جهت و حرکات دست‌ها، بازوها، بدن و حالات چهره به‌عنوان ابزار ارتباط جهت انتقال مفاهیم و تعامل با محیط بیرون استفاده می‌شود. زبان اشاره به‌عنوان یک زبان ارتباطی برای معلولین شنوایی استفاده می‌شود. هدف پژوهش‌های انجام‌شده در سامانه‌های تشخیص زبان اشاره، کمک به ارتباط معلولین شنوایی با افراد سالم و محیط بیرون است [۱].

جهت انتقال مفاهیم در زبان اشاره سه شکل اساسی وجود دارد؛ اول، املای انگشتی که با هجی‌کردن کلمات، حرف به حرف استفاده می‌شود. دوم، حرکات اشاره‌ای نشانگر کلمه که در آن هر کلمه با یک حرکت اشاره بیان می‌گردد و سوم ویژگی‌های غیر دستی نظیر حالات چهره، زبان، لب و موقعیت بدن. املای انگشتی برای کلمه‌هایی که در لغت‌نامه زبان اشاره وجود ندارند، استفاده می‌شود. بدین ترتیب که با تجزیه کلمه مورد نظر به حروف آن، هر حرف آن توسط املای انگشتی (شکل و حالت خاصی از دست و بازو) بیان می‌گردد [۲]. به‌طور مثال کلمه "علی" که در لغت‌نامه زبان اشاره وجود ندارد، با یک حرکت اشاره بیان نمی‌گردد بلکه هر یک از حروف آن "ع"، "ل" و "ی" توسط املای انگشتی با حالت و شکلی از دست بیان می‌گردد.

در بخش بعد به مروری بر روش‌های مختلف جمع‌آوری داده از محیط در سامانه‌های تشخیص حالات دست می‌پردازیم.

۲- پیشینه کار

این بخش به روش‌های مختلف اندازه‌گیری داده در سامانه‌های تشخیص حرکت و حالت دست می‌پردازد. دو روش جهت اندازه‌گیری داده در سامانه‌های تشخیص حالات و حرکات دست وجود دارد؛ اول، استفاده از دستکش‌هایی که مجهز به حسگرهایی جهت اندازه‌گیری زوایای مفاصل دست هستند. دوم، استفاده از سامانه تصویری و اندازه‌گیری داده‌ها با استفاده از پردازش تصاویر.

۲-۱ روش مبتنی بر دستکش الکترونیکی

در برخی از سامانه‌های تشخیص زبان اشاره برای اندازه‌گیری حالات و موقعیت‌های دست از دستکش‌های الکترونیکی استفاده می‌کنند. این دستکش‌ها با استفاده از حسگرهایی که روی آن قرار دارد، زوایای انگشتان را برای تعیین موقعیت و جهت دست‌ها محاسبه می‌کنند. با استفاده از پردازنده متصل به دستکش‌ها، داده‌های به‌دست آمده پردازش شده و خروجی به‌دست می‌آید. این روش نتایج قابل قبولی داشته، اما نسبتاً پرهزینه بوده و کاربر آزادی حرکت مناسبی ندارد [۱] و [۳].

در [۴] جهت تشخیص زبان اشاره و تفسیر آن به متن یا کلام از دستکش الکترونیکی استفاده شده است. در این دستکش با استفاده از حسگرهایی که حساس به خمیدگی و حرکت انگشتان دست هستند، زوایای مفاصل اندازه‌گیری می‌شود. این حسگرها زوایای مفاصل را اندازه گرفته و با توجه به زاویه، یک مقدار خروجی با اندازه بین ۰ و ۲۵۵ ارائه می‌دهند. داده‌های به‌دست آمده از حسگرها به یک شبکه عصبی پرسپترون آموزش‌دیده داده می‌شود. ورودی شبکه عصبی پرسپترون، یک

این مقاله در تاریخ ۶ دی ماه ۱۳۸۷ دریافت و در تاریخ ۲۳ فروردین ماه ۱۳۸۸ بازنگری شد.

محمدجواد برزگر سخویدی، دانشکده مهندسی برق و کامپیوتر، دانشگاه تربیت مدرس، تهران (email: mj_barzegar@yahoo.com).

احمدرضا شرافت، دانشکده مهندسی برق و کامپیوتر، دانشگاه تربیت مدرس، تهران (email: ahmad.sharafat@gmail.com).

1. Human Computer Interaction

ماتریس ۲۴×۱۸ است که ستون‌ها بیانگر تعداد حروف الفبا و سطرها بیانگر مقادیر خروجی به‌دست آمده از حسگرها هستند. خروجی این شبکه یک ماتریس یکه ۲۴×۲۴ است. دقت تشخیص در آن پژوهش ۹۰٪ گزارش شده است. مشکل اصلی در آن پژوهش، استفاده هم‌زمان از چندین نرم‌افزار (Labview, Matlab) در کنار یکدیگر است که باعث کندی پردازش‌ها می‌شود.

۲-۲ روش تصویری

در برخی دیگر از سامانه‌های تشخیص زبان اشاره، از روش‌های تصویری استفاده می‌شود. در این روش تصاویر مربوط به فرد توسط یک یا چند سامانه ورودی تصویر گرفته می‌شود. با استفاده از سامانه‌های ورودی از کاربر، فریم‌هایی (تصاویری) ضبط و سپس این فریم‌ها به بخش پردازش، ارسال شده تا حالات دست تشخیص داده شوند. از روش‌های پردازش تصویر می‌توان به روش‌های بازشناسی الگو نظیر مدل پنهان مارکوف (HMM) اشاره کرد. این روش در حرکت فرد محدودیت ایجاد نمی‌کند و در حیطه کاربردهای بی‌درنگ موفق بوده است [۳] و [۵].

سامانه‌های تشخیص زبان اشاره مبتنی بر بینایی ماشین، همانند سامانه‌های تشخیص حرکت از ویژگی‌هایی نظیر ساختار هندسی، شکل و ساختار ظاهری دست استفاده می‌کنند. این ویژگی‌ها به‌عنوان نمادهای اشاره طبقه‌بندی می‌گردند. در برخی از سامانه‌ها، حرکات اشاره دست با استفاده از مدل مخفی مارکوف طبقه‌بندی می‌گردند [۶] و [۷] و در برخی دیگر از سامانه‌ها، حالت و شکل دست با استفاده از الگوریتم‌های ساده‌تر در تشخیص الگو طبقه‌بندی می‌گردد [۱] و [۳].

در [۱] جهت تشخیص زبان اشاره حرکات انگشتان، از ترکیب یک شبکه عصبی و مدل تشخیص الگو استفاده شده است. هدف آن پژوهش طراحی الگوریتم جدیدی جهت تشخیص حرکت انگشتان با استفاده از شبکه‌های عصبی و مدل تشخیص الگو برای ایجاد ارتباط بین معلولین شنوایی و افراد سالم بود. برای این کار ابتدا ناحیه مربوط به حالت دست در فریم‌های ورودی شناسایی و ویژگی‌های مربوط به آن استخراج و این ویژگی‌ها به یک مجموعه ترکیبی از شبکه عصبی و مدل تشخیص الگو داده می‌شود. سپس داده‌های به‌دست آمده از مرحله قبل برای طبقه‌بندی به یک شبکه عصبی وارد می‌شود و خروجی به‌دست می‌آید. جهت بررسی عملکرد سیستم، شبکه عصبی با ۴۶ حرکت دست مربوط به زبان اشاره ژاپنی آموزش داده شده است. داده‌های آزمون این شبکه از ۵ نفر تهیه شد که نتایج به‌دست آمده از پژوهش، دقت تشخیص ۸۴/۸٪ را نشان داده است.

سامانه طراحی‌شده در [۸] جهت تشخیص ۹ حالت مختلف دست به‌کار رفته است. آن سامانه دارای دو مرحله آموزش و تشخیص است. در مرحله آموزش، چندین شکل مختلف از یک حالت دست کاربر جمع‌آوری می‌شود. سپس ناحیه مربوط به دست از تصاویر ورودی شناسایی شده و تصویر دودویی آن به‌دست می‌آید. جهت استخراج ویژگی‌ها از ضرایب توصیف‌گر فوریه استفاده شده است. ضرایب توصیف‌گر فوریه از منحنی پیرامونی تصاویر دودویی دست محاسبه و در پایگاه داده آموزشی ذخیره می‌شود. در مرحله تشخیص، برای تصویر ورودی دست بردارهای ویژگی استخراج شده و توسط روش نزدیک‌ترین همسایه، تشخیص حالات دست انجام می‌گیرد. نتایج آن پژوهش، بر روی مجموعه آزمون شامل ۱۴۰ تصویر دست از سه کاربر مختلف، دقت تشخیص ۹۵٪–۹۲٪ را نشان داده است.

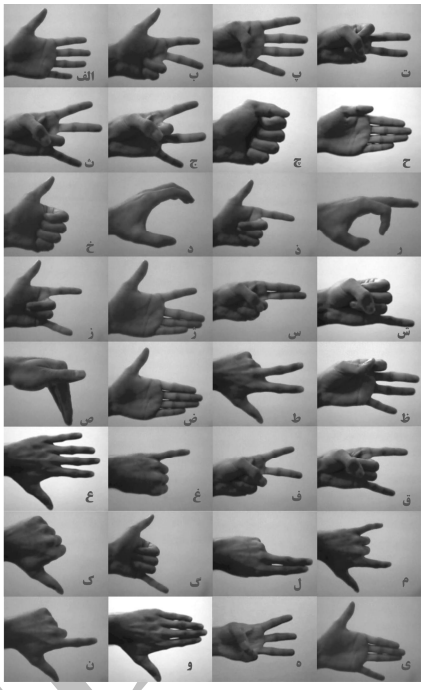
در [۹] سامانه‌ای طراحی شده است که می‌تواند ۱۰ حالت مختلف دست را هنگام تعامل انسان با رایانه تشخیص دهد. جهت شناسایی ناحیه دست، روشنایی محیط ثابت و رنگ پس‌زمینه تیره در نظر گرفته شده است و ناحیه دست با تعیین یک حد آستانه با توجه به هیستوگرام تصویر ورودی از پس‌زمینه جدا می‌شود. سه ویژگی مختلف جهت مقایسه دقت تشخیص حالات دست از تصویر دودویی دست استخراج شده است. این ویژگی‌ها شامل گشتاورهای Hu، ضرایب ویژه و توصیف‌گرهای فوریه بودند. جهت تشخیص حالات دست از روش نزدیک‌ترین همسایه به‌عنوان دسته‌بند استفاده شده است. جهت ارزیابی آن سامانه، مجموعه داده‌ها از ۸ کاربر مختلف تهیه شد که شامل ۹۶۶ تصویر بود. دقت تشخیص حالات دست با انتخاب هر یک از این ویژگی‌ها ۹۸/۶٪–۹۶/۲٪ و هنگام ترکیب این ویژگی‌ها، ۹۹/۵٪ گزارش شده است. زمان پردازش برای تشخیص برای هر فریم ورودی ۱۵–۱۷ میلی‌ثانیه بوده است.

در [۱۰] سامانه‌ای ارائه شده است که از ۶ حالت مختلف دست جهت تعامل با رایانه استفاده می‌کند. جهت شناسایی ناحیه دست از پس‌زمینه ثابت، با توجه به تغییر حالت دست، تصویر موجود با تصویر قبل آن مقایسه شده و با محاسبه آنتروپی، ناحیه دست جدا می‌شود. بعد از به‌دست آوردن منحنی پیرامونی تصویر دودویی دست، جهت استخراج ویژگی از تابع فاصله از گرانیگاه استفاده شده است. داده‌های آزمون از ۶ کاربر مختلف شامل ۷۲۰ تصویر (هر حالت دست ۲۰ تصویر) تهیه شد که نتایج به‌دست آمده از آن پژوهش، دقت تشخیص ۹۵٪–۹۹٪ را نشان داده است. زمان پردازش برای تشخیص برای هر فریم ورودی ۲۰۰ میلی‌ثانیه بوده است.

در [۱۱] سامانه‌ای طراحی شده است که از ۲۶ حالت مختلف دست جهت تعامل با ماشین استفاده می‌کند. ناحیه دست بر اساس مشخصه رنگ از فریم‌ها جدا می‌شود. در ادامه منحنی پیرامونی هر حالت دست به‌دست می‌آید. برای مجموعه آموزشی مربوط به هر حالت از دست، تعدادی تصویر کانتور از آن حالت دست در یک پایگاه تصویری ذخیره می‌شود. تشخیص حالات دست با استفاده از معیار فاصله هاسدورف[1] [۱۲]، بین حالت‌های دست ناشناخته و تصاویر ذخیره‌شده انجام می‌شود. دقت تشخیص در آن پژوهش ۹۰٪ گزارش شده است. در آن روش، مرحله کاهش ویژگی وجود ندارد و مرحله شناسایی حالات دست محاسبات نیاز به محاسبات طولانی داشته و زمان‌بر است و مجموعه آموزشی که یک پایگاه تصویری است، به حافظه بزرگ نیاز دارد. از دیگر مشکلات آن روش، حساس‌بودن به چرخش دست است. به همین دلیل نتایج تشخیص حالات دست دقت مناسبی را ندارد.

در [۳] سامانه‌ای طراحی شده است که بتواند به‌صورت موفقیت‌آمیزی زبان اشاره انگلیسی را به یک متن تبدیل کند. برای این کار دو مرحله جمع‌آوری و پردازش داده در نظر گرفته می‌شوند. در مرحله جمع‌آوری داده، تصاویر از یک یا چندین سامانه ورودی گرفته می‌شود. مرحله پردازش خود شامل ۳ قسمت است؛ اول استخراج فریم (تصویر) از سامانه ورودی، دوم پردازش تصویر و سوم ذخیره نتایج. در مرحله پردازش، ابتدا ناحیه مربوط به دست از فریم انتخابی شناسایی شده و ناحیه تصویری مربوطه به‌صورت تصویر دودویی و به مقیاس کوچک‌تری تبدیل می‌شود. این تصاویر کوچک برای هر حرف الفبای انگلیسی در یک پایگاه داده تصویری ذخیره می‌شود. در مرحله آموزش، برای هر حرف الفبا چندین تصویر در پایگاه داده تصویری ذخیره می‌شود. در مرحله تشخیص، فریم‌های مربوط به هر حرف از فریم‌های ورودی با استفاده از یک میزان خطای در نظر گرفته شده، تعیین و حرف الفبای مربوط به فریم تشخیص

1. Hausdorff



شکل ۱: حالات دست که بیانگر ۳۲ حرف الفبای فارسی هستند.

داده می‌شود. دقت تشخیص آن پژوهش، ۷۹٫۴٪ گزارش شده است. انتخاب نامناسب ویژگی‌ها از دلایل پایین‌بودن دقت آن سامانه به شمار می‌رود. نتایج آن پژوهش نشان داد که اگر سرعت پردازش و ذخیره داده‌ها با استفاده از روش‌های فشرده‌سازی و استخراج ویژگی افزایش یابد، دقت تشخیص سامانه افزایش خواهد یافت.

در [۵] سامانه‌ای طراحی شده است که از ۸ حالت مختلف دست جهت تعامل با روبات استفاده می‌کند. جهت شناسایی ناحیه دست در تصاویر از روش مبتنی بر رنگ استفاده شده است که در آن یک شبکه عصبی آموزش‌دیده، پیکسل‌های مشابه رنگ پوست را از تصویر جدا می‌کند. خروجی شبکه عصبی یک تصویر دودویی شامل دست است. جهت استخراج ویژگی‌های هر حالت از دست از منحنی پیرامونی تصویر دودویی آن استفاده شده است. این ویژگی‌ها شامل موقعیت و تعداد انگشتان دست در تصاویر است. تشخیص حالت دست با استفاده از تجزیه و تحلیل این ویژگی‌ها انجام می‌شود. دقت تشخیص آن سامانه ۹۵٪ گزارش شده است.

در این مقاله تشخیص حالات اشاره دست توسط املای انگشتی برای حروف الفبای فارسی با استفاده از روش تصویری مورد تحقیق و بررسی قرار می‌گیرد که نمونه مشابه آن برای زبان فارسی گزارش نشده است. تعداد حروف الفبای فارسی ۳۲ عدد است که برای هر حرف آن یک حالت دست در نظر گرفته می‌شود. تصویر یا دنباله‌ای از تصاویر از حالت دست کاربر که بیانگر حرف الفبا است به ورودی سامانه اعمال می‌شود. بعد از جداسازی ناحیه دست از تصاویر، منحنی پیرامونی آن به‌دست می‌آید. جهت استخراج ویژگی از تابع فاصله از گرانیگاه استفاده شده است. کاهش بردار ویژگی با استفاده از تبدیل کسینوسی گسسته روی تابع فاصله از گرانیگاه انجام می‌شود که برای اولین بار است، جهت کاهش بردار ویژگی‌ها در سامانه‌های تشخیص حالات دست از آن استفاده می‌شود. این تبدیل نسبت به تبدیلات دیگر نظیر تبدیل فوریه گسسته یا توصیف‌گرهای فوریه از کارایی بالاتری برخوردار است. تشخیص حالات دست نیز با استفاده از سه روش نزدیک‌ترین همسایه (معیار فاصله اقلیدسی- معیار فاصله اقلیدسی نرمالیزه) و شبکه عصبی انجام شده است و در آخر با توجه به حالت دست و پردازش‌های انجام‌شده، متن خروجی به‌دست می‌آید.

۳- پیاده‌سازی

در این مقاله جهت پیاده‌سازی سامانه تشخیص حروف الفبای فارسی، پنج بخش اصلی در نظر گرفته شده است که عبارتند از: جمع‌آوری داده‌های تصویری؛ پیش‌پردازش؛ استخراج بردار ویژگی؛ کاهش اندازه بردار ویژگی و تشخیص که در بخش‌های بعد توضیح داده می‌شود.

۳-۱ جمع‌آوری داده‌های تصویری

جهت گرفتن تصاویر دست از وب‌کم[1] به‌عنوان سامانه تصویر ورودی استفاده می‌شود. درجه وضوح تصاویر ورودی ۳۲۰×۲۴۰ پیکسل است که به‌صورت بی‌درنگ به بخش پیش‌پردازش ارسال می‌شود. رنگ روشنایی محیط ثابت در نظر گرفته می‌شود. به‌علت این که از مشخصه رنگ جهت آشکارسازی پوست استفاده می‌شود، رنگ‌های موجود در پس‌زمینه نباید شامل رنگ پوست باشند. با توجه به این که در تشخیص حروف الفبا از حالات ممکن یک دست استفاده می‌شود، بنابراین لازم است کاربر از مچ‌بند رنگی (به‌غیر از رنگ پوست) استفاده کند. راه حل دیگر آن است که دست کاربر تا مچ در تصاویری که سامانه ورودی تهیه می‌کند، قرار گیرد. این راهکار نیاز به دقت کاربر دارد. فاصله دست کاربر تا سامانه ورودی تصویر نیز باید بین ۳۰ الی ۷۰ سانتی‌متر باشد. در فاصله کمتر از ۳۰ سانتی‌متر، به‌علت نزدیک‌بودن دست کاربر تا سامانه ورودی، تصویر دست ناقص و در فاصله بیشتر از ۷۰ سانتی‌متر، تصویر حالت دست از کیفیت پایینی برخوردار است. حالت‌های دست که بیانگر حروف الفبای فارسی هستند در شکل ۱ آمده است.

۳-۲ پیش‌پردازش

برای کاهش محاسبات در تشخیص حالات دست، ابتدا باید ویژگی‌های مناسب دست استخراج گردد. برای این کار ابتدا ناحیه مربوط به دست از پس‌زمینه استخراج و نویزهای آن حذف می‌گردد. پیش‌پردازش شامل جداسازی دست و پاکسازی نقاط غیر مطلوب و نرمال‌سازی اندازه دست است که در بخش‌های بعد توضیح داده می‌شود.

۳-۲-۱ جداسازی دست و پاک‌سازی نقاط غیر مطلوب

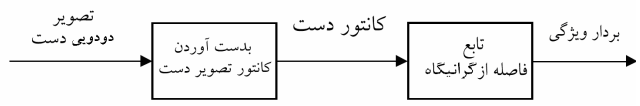
برای جداسازی ناحیه دست در فریم‌ها از ویژگی رنگ پوست استفاده می‌کنیم. در این روش با توجه به رنگ پوست و فرض روشنایی ثابت محیط، ناحیه دست از فریم‌ها جدا می‌شود. این روش نسبت به جداسازی مبتنی بر حرکت مزایایی دارد. به‌طور مثال اگر جسم متحرکی به غیر از دست در فریم‌ها وجود داشته باشد روش مبتنی بر حرکت جهت جداسازی دست، آن جسم را نیز از فریم جدا می‌کند که باعث خطا خواهد شد.

جهت آشکارسازی پوست ابتدا فریمی را که رنگ روشنایی آن تصحیح شده است از مبنای رنگ RGB به مبنای رنگ HSV (اصل رنگ، اشباع و مقدار) نگاشت می‌دهیم. سپس با استفاده از مقادیر آستانه‌ای که توسط (۱) مشخص شده است، رنگ پوست آشکارسازی می‌گردد. خروجی آشکارساز پوست، تصویر دودویی است که پیکسل‌های محدوده رنگ پوست، دارای مقدار یک و پیکسل‌های بقیه نقاط تصویر دارای مقدار صفر هستند

$$۰٫۰۲ < H < ۰٫۱۳ \quad , \quad ۰٫۲ < S < ۰٫۶۵ \quad , \quad ۰٫۲۵ < V \qquad (۱)$$

که V، S و H مؤلفه‌های مدل رنگی HSV هستند. تصویر حاصله، تصویری دودویی است که ناحیه دست و نواحی دیگری که شبیه به رنگ

1. Webcam



شکل ۲: مراحل مختلف در بخش استخراج ویژگی.

پوست هستند، دارای مقدار یک و بقیه نقاط مقدار صفر دارند. فرض بر این است که بزرگ‌ترین ناحیه، مربوط به دست است. بنابراین با انتخاب یک مقدار آستانه، نواحی غیر مطلوب حذف می‌شوند و مقدار آنها صفر می‌شود. فقط بزرگ‌ترین ناحیه که مربوط به دست است باقی می‌ماند. در ادامه این کار، تصویر دودویی کوچک می‌شود به‌طوری که فقط شامل ناحیه دست باشد.

۳-۲-۲ نرمال‌سازی اندازه دست

با توجه به فاصله دست از سامانه تصویری که باعث کوچکی یا بزرگی تصویر دست می‌گردد و نیز اندازه‌های مختلف دست انسان، لازم است که ناحیه دست در تصاویر نرمال‌سازی شود. روش‌های مختلفی برای نرمال‌سازی استفاده شده است [۱۳]. روشی که در این مقاله جهت نرمال‌سازی دست استفاده کردیم، نیاز به محاسبات پیچیده نداشته و به‌راحتی انجام‌پذیر است.

در این روش ابعاد تصویر دودویی دست که از مرحله قبل به‌دست آمده است نرمال‌سازی می‌شود، به‌طوری که بعد بزرگ‌تر تصویر جدید به ۱۰۰ پیکسل و بعد کوچک‌تر تصویر جدید از تقسیم بعد کوچک‌تر تصویر قبل بر بعد بزرگ‌تر تصویر قبل در عدد ۱۰۰ به‌دست می‌آید، یعنی

$$y_{new} = ۱۰۰$$
$$x_{new} = \frac{x_{old}}{y_{old}} \times ۱۰۰ \qquad (۲)$$

که در آن x_{old}، y_{old}، x_{new} و y_{new} به‌ترتیب بعد کوچک‌تر و بزرگ‌تر تصویر قبل از نرمال‌سازی و بعد کوچک‌تر و بزرگ‌تر تصویر بعد از نرمال‌سازی هستند. با اعمال این روش، اندازه‌های بزرگ‌تر و کوچک‌تر دست به اندازه‌هایی استاندارد و متناسب با همان اندازه‌های اولیه، نرمال‌سازی می‌شوند.

۳-۳ استخراج ویژگی

در این مرحله ویژگی‌های تصویر دودویی حالت دست که خروجی مرحله پیش‌پردازش است، استخراج می‌گردد. در این پژوهش از کانتور یا منحنی پیرامونی جهت توصیف حالت دست استفاده کرده‌ایم. بنابراین لازم است کانتور شکل از تصویر دودویی دست استخراج گردد. برای این که ویژگی‌های استخراجی از کانتور دست در برابر دوران و تغییر اندازه مقاوم باشند، از تابع فاصله از گرانیگاه همراه با نرمال‌سازی استفاده کردیم.

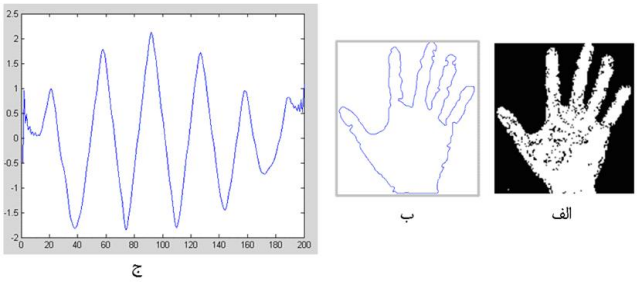
تابع فاصله از گرانیگاه، نشان‌دهنده فاصله هر نقطه از مرز شکل با گرانیگاه آن است. جابه‌جایی شکل بر روی این تابع اثری ندارد. تابع فاصله از گرانیگاه طبق (۳) محاسبه می‌شود [۱۰]

$$r = \sqrt{(x(t) - x_c)^۲ + (y(t) - y_c)^۲} \qquad (۳)$$

که $(x(t), y(t))$ مختصات نقاط مرزی شکل و (x_c, y_c) مختصات گرانیگاه شکل است که از (۴) و (۵) به‌دست می‌آید

$$x_c = \frac{۱}{L}\sum_{t=۰}^{L-۱} x(t) \qquad (۴)$$

$$y_c = \frac{۱}{L}\sum_{t=۰}^{L-۱} y(t) \qquad (۵)$$



شکل ۳: (الف) تصویر دودویی دست، (ب) کانتور دست و (ج) تابع فاصله از گرانیگاه.

سپس با استفاده از این تابع، بردار ویژگی مربوط به کانتور حالت دست استخراج می‌گردد. به‌ازای هر حالت دست، بردار ویژگی به ابعاد ۱×۲۰۰ محاسبه می‌گردد. شکل ۲ مراحل استخراج ویژگی و شکل ۳ تصاویر مربوط به استخراج ویژگی را نشان می‌دهد.

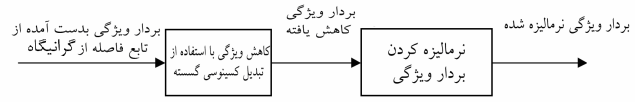
۳-٤ کاهش ابعاد بردار ویژگی (روش پیشنهادی)

بزرگ‌بودن اندازه بردار ویژگی (۱×۲۰۰) مستلزم انجام محاسبات زیادی است. برای کاهش میزان محاسبات بایستی ابعاد بردار ویژگی کاهش یابد. در این پژوهش از تبدیل کسینوسی گسسته (DCT) جهت کاهش بردار ویژگی استفاده کرده‌ایم. در DCT انرژی سیگنال در مؤلفه‌های اولیه این تبدیل قرار دارد، در صورتی که در تبدیل فوریه گسسته (DFT) انرژی سیگنال در طیف بزرگ‌تری از مؤلفه‌های آن نسبت به تبدیل کسینوسی گسسته قرار دارد. مزیت DCT نسبت به تبدیلات گسسته دیگر در همین ویژگی است. اگر مؤلفه‌های اولیه تبدیلات گسسته را به‌عنوان ویژگی انتخاب کنیم، به‌علت این که مؤلفه‌های DCT نسبت به تبدیلات گسسته دیگر فشرده‌تر هستند، ابعاد بردار ویژگی در DCT نسبت به تبدیلات دیگر کوچک‌تر خواهد بود. از کاربردهای تبدیل کسینوسی گسسته، فشرده‌سازی تصاویر است [۱۴]. در این مقاله از DCT برای اولین بار جهت کاهش بردار ویژگی‌ها بر روی تابع فاصله از گرانیگاه در سامانه‌های تشخیص حالات دست استفاده شده است. در [۱۵] روش‌های مختلف تشخیص شکل بررسی شده است. نتایج آن پژوهش نشان داد که اعمال تبدیل فوریه گسسته (DFT) روی تابع فاصله از گرانیگاه شکل‌ها نسبت به روش‌های دیگر نظیر ضرایب توصیف‌گر فوریه از دقت تشخیص بالاتری برخوردار است. ما در این مقاله با اعمال DCT روی تابع فاصله از گرانیگاه که نسبت به DFT فشردگی بیشتری دارد، ویژگی‌ها را که همان ضرایب اولیه DCT هستند، استخراج کرده‌ایم.

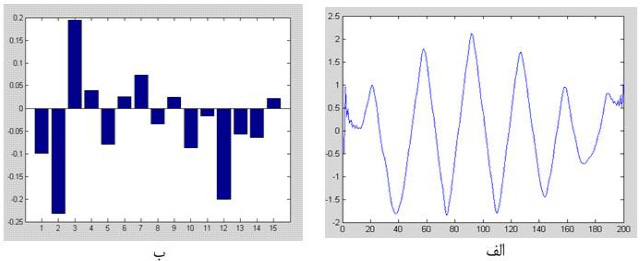
با اعمال DCT روی بردار ویژگی حالات مختلف دست، مشاهده می‌کنیم که جهت کاهش بردار ویژگی می‌توان از مؤلفه‌های بزرگ‌تر از ۱۶ صرف نظر کرد. سپس بردار کاهش‌یافته ویژگی، نرمال‌سازی می‌شود. جهت نرمال‌سازی بایستی تک‌تک مؤلفه‌های کسینوسی گسسته بر اولین مؤلفه تقسیم گردند. بردار کاهش‌یافته ویژگی شامل اولین مؤلفه نخواهد بود زیرا در تمامی بردارهای نرمالیزه‌شده این مقدار برابر یک است و بردار ویژگی از دومین مؤلفه تا شانزدهمین مؤلفه خواهد بود. بنابراین با استفاده از DCT و سپس نرمال‌سازی آن، اندازه بردار ویژگی از ۱×۲۰۰ به ۱×۱۵ کاهش می‌یابد. شکل ۴ مراحل کاهش ابعاد ویژگی و شکل ۵ تصاویر مربوط به کاهش ویژگی را نشان می‌دهد.

۳-۵ تشخیص

در این مرحله از سه دسته‌بند مختلف شامل نزدیک‌ترین همسایه با معیارهای مختلف فاصله (فاصله اقلیدسی، فاصله اقلیدسی نرمالیزه‌شده) و شبکه عصبی پس‌انتشار خطا جهت طبقه‌بندی و تشخیص حالات دست



شکل ۴: مراحل کاهش بردار ویژگی با استفاده از DCT.



شکل ۵: (الف) تابع فاصله از گرانیگاه و (ب) بردار کاهش‌یافته ویژگی.

استفاده کردیم. در روش نزدیک‌ترین همسایه بردار کاهش‌یافته ویژگی که از مرحله قبل به‌دست آمده است با داده‌های مجموعه آموزشی مقایسه شده و طبقه‌بندی صورت می‌گیرد. جهت مقایسه از دو معیار کم‌ترین فاصله اقلیدسی (۶) و کم‌ترین فاصله اقلیدسی نرمالیزه‌شده (۷) استفاده می‌کنیم

$$d_{ij} = \sqrt{\sum_{k=1}^{m} (x_m - y_m)^2} \qquad (۶)$$

$$d'_{ij} = \sqrt{\sum_{k=1}^{m} \left(\frac{x_m - y_m}{\sigma_m}\right)^2} \qquad (۷)$$

که x_m ، y_m و σ_m به‌ترتیب mامین درایه از بردار ویژگی آزمون، بردار ویژگی آموزش و انحراف معیار mامین درایه بردار ویژگی است. در نزدیک‌ترین همسایه، بردار کاهش‌یافته ویژگی با مجموعه آموزش که شامل بردار کاهش‌یافته ویژگی تمام حالات دست است، مقایسه شده و حاصل این مقایسه، برداری از اعداد خواهد بود. بردار کاهش‌یافته ویژگی حاضر به حالتی از دست تعلق دارد که نزدیک‌ترین مقدار به صفر را دارد و بدین ترتیب خروجی به‌دست می‌آید.

شبکه عصبی پس‌انتشار خطا دسته‌بند دیگری بود که جهت تشخیص حالات دست از آن استفاده کردیم. با توجه به تعداد حالات دست، در لایه خروجی شبکه عصبی ۳۲ گره قرار داده شده است. از ۱۵ ویژگی که همان بردار ویژگی کاهش‌یافته هستند برای تشخیص استفاده شده است. لایه ورودی این شبکه نیز ۱۵ گره دارد. تعداد گره‌های لایه میانی مورد بررسی قرار گرفت و مشخص شد که با تعداد ۲۷ تا ۳۳ گره، شبکه عصبی در بهترین وضعیت یادگیری قرار می‌گیرد. در لایه میانی ۳۰ گره قرار داده شده است. پارامتر یادگیری این شبکه ۰/۰۱ بود.

٤- نتایج پیاده‌سازی

در این بخش نتایج پیاده‌سازی را بررسی می‌کنیم. سامانه ورودی تصویر شامل یک عدد وب‌کم جنیوس، پردازنده سیستم ۱/۷ گیگاهرتز پنتیوم ۳ و مقدار حافظه رایانه ۲۵۶ مگابایت است. تمام مراحل این مقاله در محیط MATLAB پیاده‌سازی شده است. جهت آموزش دسته‌بند نیاز به یک مجموعه آموزش جهت مقایسه با داده‌های آزمون داریم. در مرحله آموزش ابتدا برای هر حرف الفبای فارسی، ۴۰ نمونه آموزشی که شامل تغییرات وضعیت مختلف دست در هر حالت بود در نظر گرفتیم. با توجه به تعداد حروف الفبای فارسی که ۳۲ حرف است، ۱۲۸۰ نمونه آموزشی به‌دست می‌آید. نمونه‌های آموزش بعد از استخراج و کاهش اندازه بردار

جدول ۱: دقت تشخیص حالات اشاره دست.

دسته‌بند	نزدیک‌ترین همسایه (معیار فاصله اقلیدسی)	نزدیک‌ترین همسایه (معیار فاصله اقلیدسی نرمالیزه)	شبکه عصبی
دقت تشخیص	٪۹۵/۲۵	٪۹۶/۸	٪۹۹/۱

جدول ۲: مقایسه دقت تشخیص حالات اشاره دست.

کاربران	کاربر ۱	کاربر ۲	کاربر ۳
دقت تشخیص	٪۹۶/۳	٪۹۹/۱	٪۹۸/۴

ویژگی در مجموعه آموزش ذخیره می‌شوند تا برای آموزش دسته‌بندهای مختلف استفاده شوند.

جهت ارزیابی کارایی سامانه تشخیص حروف الفبا، مجموعه آزمون ایجاد کردیم که شامل ۱۶۰۰ تصویر (به‌ازای هر حالت دست، ۵۰ تصویر) است. نتایج طبقه‌بندی نمونه‌های آموزش دقت تشخیص سیستم در نور معمول اتاق در جدول ۱ آمده است.

همان‌طور که در جدول ۱ ملاحظه می‌شود، شبکه عصبی بالاترین دقت تشخیص را نسبت به دو دسته‌بند دیگر داراست. زمان لازم برای پردازش تصویر یک حالت از دست ۳۵۰ میلی‌ثانیه با دسته‌بند شبکه عصبی برآورد شد.

جهت ارزیابی دقت تشخیص سامانه نسبت به کاربران مختلف، مجموعه آزمون از سه کاربر شامل ۴۸۰۰ تصویر (۱۶۰۰ تصویر به‌ازای هر فرد و به ازای هر حالت دست، ۵۰ تصویر) ایجاد کردیم. از شبکه عصبی به‌عنوان دسته‌بند استفاده گردید. نتایج طبقه‌بندی مجموعه آزمون برای سه کاربر در جدول ۲ آمده است. همان‌طور که ملاحظه می‌شود، دقت تشخیص این سامانه مستقل از کاربران است و به کاربران مختلف بستگی ندارد.

بعضی از حالات دست نظیر (ک، ض) به‌جای هم تشخیص داده می‌شوند و علت این است که توابع فاصله از گرانیگاه هر دو شکل متشابه می‌شود. برای این که دقت تشخیص بالاتر رود می‌توان با نمونه‌های بیشتر شبکه عصبی را آموزش داد یا از دو دسته‌بند متوالی استفاده کرد و یا این که علاوه بر ویژگی فاصله از گرانیگاه از ویژگی‌های دیگر نظیر موقعیت نوک انگشتان بهره جست.

یکی از دلایل ایجاد خطا در این سامانه، تشکیل سایه روی حالات دست و انعکاس نور از پوست به دلیل توزیع نامناسب نور است که باعث می‌شود تا آشکارسازی پوست به‌دلیل سیاهی بیش از حد و سفیدی حاصل از سایه و انعکاس نور، به درستی صورت نگیرد و در نتیجه بر دقت تشخیص سامانه تأثیرگذار باشد. جهت حل این مشکل می‌توان از مشخصه حرکت نیز جهت استخراج دست از تصاویر استفاده کرد که باعث افزایش زمان پردازش خواهد شد.

برای این که مقایسه‌ای بین روش ارائه‌شده و روش‌های موجود در مراجع صورت پذیرد، روش‌های استفاده‌شده در برخی از مراجع بخش پیشینه کار و روش پیشنهادی شبیه‌سازی شده و با یکدیگر مقایسه کردیم. لازم به توضیح است که شرایط تصاویر ورودی در هر مقایسه یکسان است. تعداد تصاویر برای هر حالت از دست در مجموعه آموزش ۴۰ تصویر و مجموعه آزمون ۵۰ تصویر انتخاب شد. از شبکه عصبی جهت طبقه‌بندی حالات دست استفاده شده است.

نتایج روش پیشنهادی و مقایسه آن با نتایج مراجع دیگر (جدول ۳) حاکی از آن است که روش پیشنهادی از دقت تشخیص بالاتری برخوردار است. روش پیشنهادی در مقایسه با [۵] و [۸] نیاز به زمان پردازش دارد

جدول ۳: مقایسه دقت تشخیص روش پیشنهادی و برخی تحقیقات موجود.

پژوهش	ویژگی	تعداد حالات	دقت	زمان (ms)	دقت روش پیشنهادی	زمان روش پیشنهادی
[۸]	توصیف‌گر فوریه	۹	٪۹۴	۱۵۰	٪۹۹٫۹	۱۶۰
[۹]	توصیف‌گر فوریه + گشتاورهای Hu	۱۰	٪۹۸	۲۴۰	٪۹۹٫۸	۱۸۰
[۱۰]	تابع فاصله از گرانیگاه	۶	٪۹۶	۲۰۰	٪۹۹٫۹	۱۷۰
[۵]	موقعیت و تعداد انگشتان	۸	٪۹۴٫۲	۱۶۰	٪۹۹٫۹	۱۷۰
[۳]	تصویر دودویی دست	۲۶	٪۷۵	۲۲۰۰	٪۹۸	۲۵۰

بیشتری اما با توجه به سادگی محاسباتی، پیاده‌سازی آن به‌صورت بی‌درنگ میسر است.

در [۳] و [۱۰] ویژگی‌های استخراجی کاهش نیافته و از نزدیک‌ترین همسایه جهت طبقه‌بندی حالات دست استفاده شده که باعث کاهش دقت تشخیص شده است. در [۵] از موقعیت و تعداد انگشتان به‌عنوان ویژگی استفاده شده است که با افزایش تعداد حالات دست، تشخیص حالات دشوار شده و دقت تشخیص کاهش می‌یابد.

در مواقعی که تصویر ورودی دارای نویز باشد، می‌توان با روش‌های پردازش تصویر تا حدی نویزهای تصویر را در مرحله پیش‌پردازش حذف کرد و ناحیه دست و منحنی پیرامون آن را به دقت استخراج کرد. جهت ارزیابی سامانه پیشنهادی، تصویر ورودی با نویز گاوسی آمیخته شد. نتایج نشان داد که با افزایش مقدار نویز، دقت تشخیص حالات دست کاهش می‌یابد. جهت حل این مشکل می‌توان در مرحله پیش‌پردازش از راهکارهای قوی‌تری جهت استخراج دست از تصاویر ورودی استفاده کرد که این خود باعث افزایش زمان پردازش خواهد شد.

نتایج پیاده‌سازی با روش پیشنهادی به‌خوبی مزیت این روش را نسبت به روش‌های قبلی نشان می‌دهد. در روش پیشنهادی، تبدیل کسینوسی گسسته، ویژگی‌های بهینه را از تابع فاصله از گرانیگاه هر حالت از دست استخراج کرده و باعث کاهش ویژگی‌ها می‌شود. تعداد کم ویژگی‌ها، محاسبات را کاهش داده و مجموعه آزمون نیز حجم کمی از حافظه سیستم را اشغال می‌کند و دقت تشخیص بالاتر می رود.

۵- نتیجه‌گیری

در این مقاله یک سامانه جدید برای تشخیص حروف الفبای فارسی در زبان اشاره حرکت انگشتان ارائه کردیم. پس از جداسازی ناحیه دست از فریم‌ها و اعمال مراحل پیش‌پردازش، کانتور شکل دست را به‌دست آوردیم. ویژگی‌ها را با استفاده از تابع فاصله از گرانیگاه استخراج و سپس با اعمال DCT روی آن کاهش دادیم. جهت تشخیص حالات دست از سه روش نزدیک‌ترین همسایه (معیار فاصله اقلیدسی- معیار فاصله اقلیدسی نرمالیزه) و شبکه عصبی استفاده کردیم. نتایج پیاده‌سازی، دقت تشخیص شبکه عصبی را ٪۹۹٫۱ نشان داد که نسبت به سامانه‌های موجود بهبود عملکرد داشته است. به علاوه این سامانه نسبت به اندازه‌های مختلف دست، موقعیت دست نسبت به سامانه تصویری و چرخش دست، حساسیت بسیار ناچیزی دارد. با توجه به سادگی محاسباتی روش پیشنهادی می‌توان آن را به‌صورت بی‌درنگ پیاده‌سازی نمود و جهت تعامل با سامانه‌های هوشمند نظیر رایانه و تلفن همراه به‌کار برد.

مراجع

[1] M. Shimada, S. Iwasaki, and T. Asakura, "Finger spelling recognition using neural network with pattern recognition model," in *Proc. SICE Annual Conf.*, vol. 3, pp. 2458-2463, Fukui, Japan, 4-6 Aug. 2003.

[2] P. Dreuw, *Appearance - Based Gesture Recognition*, Ph.D. Thesis, Faculty of Engineering and Science, RWTH Aachen University of Technology, Aachen, Germany, 2005.

[3] C. M. Glenn, D. Mandloi, K. Sarella, and M. Lonon, "An image processing technique for the translation of ASL finger-spelling to digital audio and text," in *Proc. Int. Symp. National Technical Institute for the Deaf*, pp. 1-7, Rochester, New York, Jun. 2005.

[4] M. Jerome, K. Pierre, and R. Foulds, "American sign language finger spelling recognition system," in *Proc. IEEE 29th Annual Northeast Bioengineering Conf.*, pp. 285-286, 22-23 Mar. 2003.

[5] X. Yin and M. Xie, "Finger identification and hand posture recognition for human - robot interaction," *Image and Vision Computing*, vol. 25, no. 8, pp. 1291-1300, Aug. 2007.

[6] T. Starner and A. Pentland, "Real-time American sign language recognition from video using hidden Markov models," in *Proc. Int. Symposium on Computer Vision*, pp. 265-270, Florida, US, 21-23 Nov. 1995.

[7] T. Starner and A. Pentland, "Visual recognition of American sign language using hidden Markov models," in *Proc. of the Int. Workshop on Automatic Face and Gesture Recognition*, pp. 189-194, Zurich, Switzerland, 26-28 Jun. 1995.

[8] A. Licsar and T. Sziranyi, "Supervised training based hand gesture recognition system," in *Proc. IEEE 16th Pattern Recognition Conf., Quebec, Canada*, vol. 3, pp. 999-1002, 11-15 Aug. 2002.

[9] S. Funck, *Video-Based Handsign Recognition for Intuitive Human Computer Interaction*, Lecture Notes in Computer Science, Berlin/Heidelberg: Springer, vol. 2449, 2002.

[10] J. H. Shin, et al., "Hand region extraction and gesture recognition using entropy analysis," *Int. J. of Computer Science and Network Security*, vol. 6, no. 2, pp. 216-222, Feb. 2006.

[11] E. Sunchez-Nielsen, L. Anton-Canalis, and M. Hernandez-Tejera, "Hand gesture recognition for human-machine interaction," *J. of WSGC 2004*, vol. 12, no. 1-2, pp. 395-402, Feb. 2003.

[12] W. Ruckelidge, "Efficient Visual Recognition Using the Hausdorff Distance," Lecture Notes in Computer Science, Berlin / Heidelberg: Springer, vol. 1173, 1996.

[13] H. Jag, J. H. Do, J. Jung, K. H. Park, and Z. Bien, "View invariant hand posture recognition method for soft-remocon-system," in *Proc. IEEE Int. Conf. on Intelligent Robots and Systems*, vol. 1, pp. 295-300, Sendai, Japan, 28 Sep.-2 Oct. 2004.

[14] G. Strang, "The discrete cosine transform," *Society for Industrial and Applied Mathematical*, vol. 41, no. 1, pp. 135-147, Mar. 1999.

[15] D. Zhang and G. Lu, "Study and evaluation of different Fourier methods for image retrieval," *Image and Vision Computing*, vol. 23, no. 1, pp. 33-49, 1 Jan. 2004.

محمدجواد برزگرسخویدی در سال ۱۳۸۳ مدرک کارشناسی مهندسی الکترونیک خود را از دانشگاه یزد و در سال ۱۳۸۷ مدرک کارشناسی ارشد مهندسی پزشکی خود را از دانشگاه تربیت مدرس دریافت نمود. زمینه‌های علمی مورد علاقه او پردازش سیگنال، شبکه‌های عصبی و پردازش تصویر است.

احمدرضا شرافت در سال ۱۳۵۴ مدرک کارشناسی مهندسی برق خود را از دانشگاه صنعتی شریف و در سال های ۱۳۵۵ و ۱۳۶۰ مدارک کارشناسی ارشد و دکترای خود را در رشته مهندسی برق از دانشگاه استنفورد در آمریکا دریافت نمود. وی اِینک در دانشکده مهندسی برق و کامپیوتر دانشگاه تربیت مدرس استاد تمام است. موضوعات پژوهشی مورد علاقه ایشان عبارتند از: روشهای پیشرفته پردازش اطلاعات و سیگنال‌ها، و همچنین سیستم‌ها و شبکه‌های مخابراتی.