

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشکده مهندسی کامپیوتر و فناوری اطلاعات

شناسایی زبان اشاره با استفاده از دوربین های عمقی

علیرضا مانشتی

استاد راهنما:

دکتر مرتضی زاهدی

استاد مشاور:

دکتر حمید حسن پور

پایان نامه ارشد جهت اخذ درجه کارشناسی ارشد

شهریور ۱۳۹۱

دانشگاه صنعتی شاهرود

دانشکده مهندسی کامپیوتر و فناوری اطلاعات

پایان نامه کارشناسی ارشد آقای علیرضا مانشتی

تحت عنوان: شناسایی زبان اشاره با استفاده از دوربین های عمقی

در تاریخ ۱۳۹۱/۰۶/۲۵ توسط کمیته تخصصی زیر جهت اخذ مدرک کارشناسی ارشد مهندسی کامپیوتر،
گرایش هوش مصنوعی مورد ارزیابی و با درجه عالی مورد پذیرش قرار گرفت.

امضاء	استاد مشاور	امضاء	استاد راهنما
	نام و نام خانوادگی : دکتر حمید حسن پور		نام و نام خانوادگی : دکتر مرتضی زاهدی

امضاء	نماینده تحصیلات تکمیلی	امضاء	اساتید داور
	نام و نام خانوادگی : مهندس علی دوست محمدی		نام و نام خانوادگی : دکتر علی اکبر پویان
			نام و نام خانوادگی : دکتر امید رضا معروضی

تقدیم بہ:

پدر و مادر عزیزوارجمندم،

برادران خوب و بزرگوارم،

ہمسفر راہ زندگی، ہمسر مہربانم،

و کانون شادی و زیبائی ہستی، دتھرم پانیا

تقدیر و تشکر

از اساتید گرانقدر و با فضیلت خویش، جناب آقای دکتر مرتضی زاهدی، دکتر علی اکبر پویان و دکتر حمید حسن پور که در کلیه مراحل تحصیل با حمایت های علمی و معنوی خویش همواره ره گشای مسیر موفقیت اینجانب و راهنمای رسیدن به قله های فضل و دانش بوده اند، صمیمانه تشکر و قدردانی می نمایم.

از کلیه اعضای هیئت علمی دانشکده مهندسی کامپیوتر و فن آوری اطلاعات، کارشناسان محترم این دانشکده به خصوص سرکار خانم سرایی و منشی محترم گروه جناب آقای علیرضا کریمی که ناهمواری ها و تاریکی های این راه صعب را با پرتو نور دانش خود و همراهی با اینجانب هموار و روشن نمودند تشکر و قدردانی می نمایم.

همچنین از کارشناسان محترم تحصیلات تکمیلی، جناب آقای حسینی، سرکار خانم بابامحمدی و جناب آقای یاقوتی نیز جهت راهنمایی ها و همکاری های ایشان کمال تشکر را دارم.

از بنیاد ملی نخبگان نیز جهت ارائه تسهیلات و جوایز تحصیلی در این دوره، شامل کمک هزینه ماهیانه، پژوهانه پایان نامه و اردوهای فرهنگی-تفریحی کمال تشکر و قدردانی را دارم. پر واضح است که این تسهیلات کمک شایانی در پیش برد اهداف تحصیلی این جانب بدون داشتن نگرانی های معیشتی داشته اند.

تعهد نامه

اینجانب علیرضا مانشتی دانشجوی دوره کارشناسی ارشد رشته مهندسی کامپیوتر گرایش هوش مصنوعی دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی شاهرود نویسنده پایان نامه شناسایی زبان اشاره با استفاده از دوربین های عمقی تحت راهنمایی دکتر مرتضی زاهدی متعهد می شوم.

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است .
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است .
- مطالب مندرج در پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « Shahrood University of Technology » به چاپ خواهد رسید .
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان نامه ، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است .
- در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری ، ضوابط و اصول اخلاقی انسانی رعایت شده است .

تاریخ ۱۳۹۱/۶/۲۵

امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج ، کتاب ، برنامه های رایانه ای ، نرم افزار ها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می باشد . این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود .
- استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی باشد.

* متن این صفحه نیز باید در ابتدای نسخه های تکثیر شده پایان نامه وجود داشته باشد .

چکیده فارسی

تشخیص زبان اشاره توسط رایانه در حوزه های مختلف هنوز یک چالش محسوب می شود. یکی از مهمترین مشکلات در این حوزه قطعه بندی دست، به خصوص در پس زمینه های پیچیده است. با ارائه دوربین های عمقی به خصوص سنسور کینکت، عمق پیکسل های تصویر نیز در دسترس می باشد که می تواند در قطعه بندی تصویر دست بسیار مفید باشد. در این تحقیق یک سیستم بلادرنگ تشخیص هجی انگشتی حروف الفبای زبان اشاره آمریکایی بر مبنای قطعه بندی عمقی دست از طریق سنسور کینکت ارائه شده است. تعداد ۲۶ ویژگی هندسی از تصاویر قطعه بندی شده دست پس از چندین مرحله پیش پردازش استخراج و پس از تحلیل ویژگی ها توسط LDA، توسط دسته بند بیز دسته بندی شده اند. ویژگی های استخراج شده شامل ۱۹ ویژگی بر حسب Moment ها و ۷ ویژگی بر حسب لبه بیرونی دست قطعه بندی شده می باشند. مراحل پیش پردازش نیز شامل آستانه گیری تصویر عمقی برای قطعه بندی دست جهت ایجاد ماسک، حذف نویز ماسک بدست آمده، جابجایی و تغییر اندازه آن می باشد. قطعه بندی تصویر رنگی و عمقی توسط اطلاعات عمقی سنسور کینکت و ارائه ویژگی ها و سیستم بلادرنگ مناسبی جهت دسته بندی بهینه داده های بدست آمده دستاورد اصلی این تحقیق می باشد.

نتایج مختلفی با اعتبارسنجی ضربدری ارائه شده است که برتری این روش را نسبت به شش روش مشابه قبلی نشان می دهد. استفاده از یک پایگاه داده استاندارد با بیش از ۵۴۰۰۰ نمونه در پنج اجرا مختلف توسط چهار فرد، اعتبار نتایج بدست آمده را تضمین نموده است. نتایج مختلفی بر حسب انتخاب های مختلفی از پایگاه داده برای ۲۴ حرف الفبا ارائه شده است که از ۷۸.۲۹٪ برای کل پایگاه داده شامل پنج اجرا کننده و ۴۵۰ نمونه از حروف به ازای هر فرد تا ۹۹.۷۵٪ برای یک اجرا کننده با ۱۰۰ نمونه برای هر حرف را شامل می شود.

بررسی سرعت پردازش ورودی ها بر روی یک پردازنده Core i7 لپ تاپ، بلادرنگ بودن سیستم با حداقل ۳۸ فریم بر ثانیه در حالت تک نخه و ۲۴۵ فریم بر ثانیه در حالت چند نخه را نشان می دهد. همچنین در تحلیل و اعتبار سنجی داده ها و نتایج نیز مشخص شده است که استفاده از LDA و ویژگی های ترکیبی حاصل از اعمال ماسک عمقی بر روی تصاویر قطعه بندی شده عمقی و رنگی بهترین نتیجه را در سیستم ارائه نموده و در صورت تصادفی کردن داده ها و اعمال آموزش برای هرفرد جدید قبل از استفاده از سیستم، بهبود چشم گیری در نتایج بدست آمده حاصل می شود.

کلمات کلیدی: شناسایی زبان اشاره، دوربین عمقی، ویژگی هندسی، کینکت، هجی انگشتی.

مقالات مستخرج از پایان نامه

1. Ali Reza Manashty and Morteza Zahedi, "Finger Spelled ASL Alphabet Recognition via Kinect-based Hand Segmentation", IOSR Journal of Engineering (IOSRJEN), ISSN: 2250-3021 (Accepted for publication on August 30th 2012).
2. Morteza Zahedi and Ali Reza Manashty, "Robust Sign Language Recognition using ToF Depth Cameras", World of Computer Science and Information Technology (WCSIT), Vol. 1, No. 3, April 2011, ISSN 2221-0741, Pages 50-55.

فهرست مطالب

فصل ۱: مقدمه	۱
۱-۱- پیش گفتار	۲
۲-۱- زبان اشاره	۳
۱-۲-۱- اهمیت و کاربرد زبان اشاره	۴
۲-۲-۱- ساختارهای زبان اشاره	۵
۳-۲-۱- نوشتار زبان های اشاره	۷
۳-۱- تعریف مسئله	۸
۴-۱- کاربرد ها	۹
۵-۱- فرض ها و محدودیت ها	۱۰
فصل ۲: مروری بر کارهای انجام شده	۱۱
۱-۲- طبقه بندی بر حسب حوزه شناسایی زبان اشاره	۱۲
۱-۱-۲- حروف الفبا	۱۲
۲-۱-۲- کلمات	۱۳
۳-۱-۲- جملات	۱۳
۴-۱-۲- تشخیص اشاره ها	۱۳
۵-۱-۲- قطعه بندی دست	۱۴
۲-۲- طبقه بندی کارهای پیشین بر حسب ابزار مورد استفاده	۱۵

- ۱۶-۲-۱- کارهای پیشین با استفاده از دستکش و سنسور ها..... ۱۶
- ۲۲-۲-۲- کارهای پیشین با استفاده از دوربین های رنگی..... ۲۲
- ۲۶-۲-۳- کارهای پیشین با استفاده از دوربین های استریو..... ۲۶
- ۲۹-۲-۴- کارهای پیشین با استفاده از دوربین های عمقی..... ۲۹
- ۳۵-۲-۵- کارهای پیشین با استفاده از کینکت..... ۳۵
- ۴۲-۳-۳- طبقه بندی کارهای پیشین بر حسب ویژگی های استخراج شده..... ۴۲
- ۴۲-۳-۱- کارهای پیشین با استفاده از ویژگی های هندسی..... ۴۲
- ۴۶-۳-۲- کارهای پیشین با ویژگی هایی بر حسب تبدیل موجک..... ۴۶
- ۴۹-۳-۳- کارهای پیشین بر حسب ویژگی های اسکلتی..... ۴۹
- ۵۳-۳-۴- کارهای پیشین با سایر ویژگی ها..... ۵۳
- ۵۳-۴-۴- طبقه بندی بر حسب روش دسته بندی..... ۵۳
- ۵۳-۴-۱- مدل مخفی مارکوف (HMM)..... ۵۳
- ۵۴-۴-۲- دسته بند بیز..... ۵۴
- ۵۴-۴-۳- شبکه های عصبی..... ۵۴
- ۵۴-۴-۴- سایر دسته بند ها..... ۵۴
- ۵۵-۲-۵- خلاصه و نتیجه گیری..... ۵۵
- ۵۵-۵-۱- بررسی ابزارهای موجود..... ۵۵
- ۵۶-۵-۲- بررسی ویژگی های مورد استفاده..... ۵۶
- ۵۷-۵-۳- بررسی روش های دسته بندی..... ۵۷

۵۹	۲-۵-۴- نتیجه گیری کلی
۶۰	فصل ۳: روش پیشنهادی
۶۱	۳-۱- پایگاه داده
۶۲	۳-۱-۱- نمونه های پایگاه داده
۶۵	۳-۲- بستر پیاده سازی
۶۸	۳-۳- پیش پردازش
۶۹	۳-۳-۱- ساخت ماسک عمقی با آستانه گیری
۷۱	۳-۳-۲- جابجایی ماسک عمقی در صفحه
۷۲	۳-۳-۳- هموارسازی ماسک عمقی
۷۵	۳-۴- تغییر مقیاس ماسک عمقی
۷۷	۳-۵- تشخیص کادر محاطی و برش تصویر رنگی
۷۷	۳-۶- تغییر اندازه تصاویر
۷۸	۳-۴- استخراج ویژگی
۷۹	۳-۴-۱- Moment ها
۸۲	۳-۴-۲- ویژگی های بر مبنای کانتور
۸۶	۳-۵- تحلیل و انتخاب ویژگی ها
۸۷	۳-۶- روش دسته بندی
۸۹	فصل ۴: نتایج آزمایشات
۹۰	۴-۱- تعاریف مورد استفاده در نتایج

- ۹۰ ۴-۱-۱- اعتبار سنجی ضربداری با n -تا کردن
- ۹۱ ۴-۱-۲- تصادفی کردن داده ها
- ۹۲ ۴-۱-۳- معیارهای ارزیابی
- ۹۴ ۴-۱-۴- تصویر قطعه بندی شده
- ۹۵ ۴-۲- نتایج
- ۹۷ ۴-۳- تحلیل نتایج
- ۹۷ ۴-۳-۱- تاثیر تغییر تعداد افراد
- ۹۸ ۴-۳-۲- تاثیر تغییر تعداد نمونه ها
- ۹۹ ۴-۳-۳- تاثیر تغییر تصویر قطعه بندی شده
- ۱۰۰ ۴-۳-۴- تاثیر تصادفی کردن داده ها در نتایج
- ۱۰۲ ۴-۴- تحلیل شناسایی هریک از حروف
- ۱۰۴ ۴-۵- مقایسه با روش های پیشین
- ۱۰۷ فصل ۵: اعتبارسنجی و تحلیل داده ها
- ۱۰۸ ۵-۱- اعتبارسنجی مقاوم بودن استفاده از عمق جهت قطعه بندی دست
- ۱۰۸ ۵-۱-۱- داده های مورد آزمایش
- ۱۰۹ ۵-۱-۲- پیش پردازش
- ۱۱۰ ۵-۱-۳- استخراج و تحلیل ویژگی ها
- ۱۱۲ ۵-۱-۴- دسته بندی داده ها
- ۱۱۲ ۵-۱-۵- نتایج و مقایسه با روش های پیشین

- ۱۱۳ نتیجه گیری ۵-۱-۶
- ۱۱۴ LDA از استفاده ۵-۲
- ۱۱۵ LDA از استفاده با ویژگی ها ۵-۳
- ۱۱۶ تحلیل استقلال بردار ویژگی ها ۵-۴
- ۱۱۸ سرعت پردازش روش پیشنهادی ۵-۵
- ۱۲۰ فصل ۶: نتیجه گیری و کارهای آینده
- ۱۲۱ ۶-۱ نتیجه گیری
- ۱۲۲ ۶-۲ پیشنهادات برای کارهای آینده

فهرست شکل ها

- شکل ۱-۲- تصویر ورودی، تصویر قطعه بندی شده دست و حرکت تشخیص داده شده (چرخش روبات به چپ) در سیستم پیشنهادی [۲۵]. تصویر از [۲۵] ۱۴
- شکل ۲-۲- تصویری از محیط کار و بازی CopyCat. برگرفته از سایت اصلی پروژه در [۱] ۱۷
- شکل ۳-۲- تصویری از یک کیوسک مخصوص بازی CopyCat که توسط دوربین از حرکات دست تصویربرداری می شود. [۲] ۱۸
- شکل ۴-۲- تصویری از دستکش های رنگی به همراه شتاب سنج (بالا) و جزئیات شتاب سنج سه بعدی (پایین) به کار رفته در بازی Copy Cat [۲] ۱۹
- شکل ۵-۲- سنسورهای EMG و شتاب سنج سه بعدی نصب شده بر روی دست در کار ژانگ و همکارانش. تصویر از [۳۵] ۲۰
- شکل ۶-۲- نمای رویی و پشتی دستکش مورد استفاده در تحقیق هوانگ و همکارانش [۳۶] ۲۱
- شکل ۷-۲- نمایی از دستکش و شبکه حسگری بکار رفته در تحقیق ساجاکارن و همکارانش. [۳۷] ۲۲
- شکل ۸-۲- مراحل کلی انجام دسته بندی اشارات دست با استفاده از دوربین فیلم برداری رنگی در کار لیم و همکارانش. تصویر از [۲۳] ۲۴
- شکل ۹-۲- تصویر دو فریم متوالی از حرکات اشاره دست (تصاویر بالا) و نتیجه خروجی الگوریتم تلفیقی پیشنهادی (در پایین) در کار لیم و همکارانش. تصویر از [۲۳] ۲۵
- شکل ۱۰-۲- تصویر ابتدایی (سمت چپ) و تصویر انتهایی (سمت راست) یک کلمه جهت جستجو و دسته بندی کلمه مورد نظر در پایگاه داده در کار آشوین و همکارانش [۴۰]. تصویر از [۴۰] ۲۶
- شکل ۱۱-۲- تصویری از یک دوربین فیلم برداری استریو. ۲۷
- شکل ۱۲-۲- یک تصویر بدست آمده از دوربین استریو در کار وانگ و همکارانش. تصویر راست و چپ به ترتیب بدست آمده از لنز سمت راست و سمت چپ دوربین می باشد. تصویر از [۲۶] ۲۸

شکل ۲-۱۳- مراحل مختلف الگوریتم قطعه بندی دست از دوربین های استریو توسط وانگ و همکارانش. تشخیص فرد (سمت چپ)، شناسایی مکان دست (وسط) و بدست آوردن پیکسل های دست (سمت راست). تصویر از [۲۶]. ۲۸

شکل ۲-۱۴- نمایی از نحوه کارکرد دوربین های عمقی در تصویربرداری از محیط. ۲۹

شکل ۲-۱۵- نمایی از یک دوربین عمقی زمان پرواز مدل MESA SR 3000. تصویر از [41]. ۳۰

شکل ۲-۱۶- (الف) یک فریم از دوربین عمقی، (ب) اختلاف بدست آمده از تفریق فریم های متوالی و (ج) حرکت تشخیص داده شده از اختلاف فریم ها در کار هولت و همکارانش [۲۴]. تصویر از [۲۴]. ۳۲

شکل ۲-۱۷- تصاویر مربوط به قطعه بندی دست در کار قبادی و همکارانش [۲۷]. سطر اول مربوط به تصاویر شدت روشنایی، سطر دوم مربوط به تصاویر دوربین عمقی و سطر سوم نتایج قطعه بندی را نشان می دهد. ۳۳

شکل ۲-۱۸- فریم ابتدایی و انتهایی یکی از علائم اشاره اجرا شده در کار ساتو و همکارانش [۹]. .. ۳۴

شکل ۲-۱۹- نمایی از سنسور کینکت. تصویر از [۴۳]. ۳۵

شکل ۲-۲۰- نمایی از دو تصویر رنگی و عمقی گرفته شده از یک صحنه با استفاده از کینکت به همراه چینش و انواع لنزهای استفاده شده در آن. ۳۶

شکل ۲-۲۱- توالی کلمه PAKET در کار لنگ با استفاده از دوربین عمقی کینکت. [۴۴]. ۳۷

شکل ۲-۲۲- نمایی از انگشتان و کف دست تشخیص داده شده در کار جادیش و همکارانش. تصویر از [۲۲]. ۳۸

شکل ۲-۲۳- نمایی از حالت ایستاده (راست) و نشسته (چپ) در استفاده از سنسور کینکت در بازی CopyCat. تصویر از [۱۳]. ۳۹

شکل ۲-۲۴- نمایی از سیستم پیشنهادی پوگالت و بودن برای تشخیص حروف الفبای زبان اشاره [۸]. تصویر از [۸]. ۴۱

شکل ۲-۲۵- تصویر تعدادی از داده های ورودی، تصاویر پس از حذف پس زمینه، کانتور استخراج شده و نتیجه بدست آمده از الگوریتم در کار مِکالا و همکارانش [۶]. تصویر از [۶] ۴۵

شکل ۲-۲۶- یکی از تصاویر ورودی و خروجی لبه یاب Canny در کار جیتا و همکارانش [۷]. تصویر از [۷] ۴۶

شکل ۲-۲۷- خروجی یک تبدیل موجک دوبعدی گسسته در سطح یک. جزئیات قطری تصویر در سمت پایین راست، جزئیات افقی و عمودی تصویر به ترتیب در سمت بالا راست و پایین چپ و کلیات تصویر در بالا سمت چپ تصویر دیده می شود..... ۴۷

شکل ۲-۲۸- تبدیل موجک یک تصویر تا سطح سه. جزویات تصویر در سطوح مختلف و در نهایت یک تصویر تخمینی در شکل مشاهده می شود. ۴۸

شکل ۲-۲۹ - یکی از تصاویر اصلی مورد استفاده در کار ایزاکس و همکارانش [۱۰]. تصویر از [۱۰] ۴۹

شکل ۲-۳۰- تصویری از دست قطعه بندی شده (تصویر سمت چپ)، نزدیک ترین مدل به دست (تصویر وسط) و مدل قرارداد شده بر روی دست (تصویر سمت راست) در کار [۴۸]. تصویر از [۴۸] ۵۰

شکل ۲-۳۱- تصویری از نمونه اولیه دستگاه حس ششم و سازنده آن. نوارهای رنگی بر روی انگشتان برای تشخیص مکان دقیق انگشتان در اسکلت دست در تصویر مشخص است. تصویر از [۵۰] ۵۱

شکل ۲-۳۲- تصویری از اجرای یک اشاره در کار [۱۳] و اسکلت بدست آمده توسط سنسور کینکت. تصویر از [۱۳] ۵۲

شکل ۳-۱- نمایی از ۲۴ حرف موجود در پایگاه داده به تفکیک ۵ اجرا توسط ۴ فرد مختلف. تصویر از [۴۵] ۶۲

شکل ۳-۲- هشت تصویر از حرف P اجرا شده توسط یکی از افراد، نشان دهنده تفاوت درون کلاسی نمونه های پایگاه داده. ۶۳

شکل ۳-۳- شباهت بین کلاسی در شش کلاس مختلف از نمونه های اجرا شده توسط یکی از افراد. ۶۴

شکل ۳-۴- کنارهم قرار گرفتن دست، صورت و گردن در تصاویری از پایگاه داده که استفاده از آستانه گیری برای قطعه بندی را مشکل می سازد. ۶۴

شکل ۳-۵- نمایی از بخش آموزش و تست نرم افزار پیاده سازی شده در محیط Net. با استفاده از زبان C# ۶۷

شکل ۳-۶- نمایی از بخش تحلیل خطای نرم افزار پیاده سازی شده و مراحل مختلف پیش پردازش و استخراج ویژگی یک نمونه ورودی در آن. ۶۸

شکل ۳-۷- مراحل مختلف پیش پردازش انجام شده بر روی داده های تصویر. ۶۹

شکل ۳-۸- یک تصویر رنگی از نمونه های پایگاه داده مورد استفاده. ۶۹

شکل ۳-۹- اعمال مراحل پیش پردازش ارائه شده در شکل ۳-۷ که بر روی تصویر رنگی شکل ۳-۸ اعمال شده است. ۶۹

شکل ۳-۱۰- نتیجه اعمال جابجایی ماسک عمقی در قطعه بندی دست و انطباق ماسک بر تصویر رنگی. ۷۱

شکل ۳-۱۱- هسته یا عنصر ساختاری مورد استفاده در فیلتر مورفولوژیک اعمال شده در مرحله پیش پردازش تصویر عمقی. مرکز هسته فوق در مختصات (۱، ۱) واقع شده است. ۷۴

شکل ۳-۱۲- نتیجه اعمال فیلتر مورفولوژیک بر روی تصویر عمقی حاصل از مرحله قبل. ۷۴

شکل ۳-۱۳- نتیجه اعمال فیلتر میانه با سایز پنجره ۳ در ۳ بر روی تصویر عمقی. حذف نویز با حفظ جزئیات بین انگشتان انجام پذیرفته است. ۷۵

شکل ۳-۱۴- اعمال تغییر مقیاس با اندازه (۰.۸، ۰.۸) بر روی محور های (x, y) که منجر به از دست رفتن اطلاعات انگشتان شده است. ۷۷

شکل ۳-۱۵- بزرگترین لبه تصویر یا کانتور استخراج شده که نقاط آن در اطراف دست قطعه بندی شده مشخص شده اند. ۸۲

شکل ۳-۱۶- نمایشی از مقادیر مختلف ویژگی فشردگی (C) بر اساس شکل ظاهری اجسام موجود در تصویر. ۸۳.....

شکل ۳-۱۷- تصویری از پوسته محدب استخراج شده از یک نمونه از حرف y از پایگاه داده. ۸۴.....

شکل ۴-۱- نمودار تاثیر افزایش افراد پایگاه داده در صحت خروجی سیستم به ازای تعداد نمونه هر حرف. ۹۷.....

شکل ۴-۲- نمودار تاثیر افزایش تعداد نمونه به ازای هر حرف در صحت خروجی سیستم. ۹۹.....

شکل ۴-۳- نمودار مقایسه تاثیر تغییر تصویر قطعه بندی شده در صحت خروجی های سیستم به ازای انتخاب های متنوع از پایگاه داده. ۱۰۰.....

شکل ۴-۴- تاثیر تصادفی کردن داده ها در صحت سیستم پیشنهادی به ازای نمونه داده های انتخابی مختلف. ۱۰۱.....

شکل ۵-۱- تصویری از فریم ابتدای (الف) و انتهای (ب) حرکت دست به راست پس از پیش پردازش و استخراج مرکز ثقل تصویر. ۱۱۰.....

شکل ۵-۲- نمودار پراکندگی مرکز ثقل فریم های شروع حرکات مختلف پایگاه داده. رنگ های موجود در تصویر قرمز، آبی، زرد و سبزآبی هستند که به ترتیب نشان دهنده کلاس های حرکت به بالا، پایین، راست و چپ می باشند. ۱۱۱.....

شکل ۵-۳- نمودار پراکندگی مرکز ثقل فریم های پایانی حرکات مختلف پایگاه داده. رنگ های موجود در تصویر قرمز، آبی، زرد و سبزآبی هستند که به ترتیب نشان دهنده کلاس های حرکت به بالا، پایین، راست و چپ می باشند. ۱۱۱.....

شکل ۵-۴- نمودار پراکندگی اندازه بردار جابجایی مرکز ثقل فریم های شروع و پایانی حرکات مختلف پایگاه داده. رنگ های موجود در تصویر قرمز، آبی، زرد و سبزآبی هستند که به ترتیب نشان دهنده کلاس های حرکت به بالا، پایین، راست و چپ می باشند. ۱۱۲.....

شکل ۵-۵- نمودار تاثیر استفاده از LDA در بهبود کارایی سیستم به ازای انتخاب های مختلف از

پایگاه داده. ۱۱۴

شکل ۵-۶- نمودار کاهش درصد شناسایی سیستم با کاهش ویژگی ها در حالت ویژگی های ترکیبی.

..... ۱۱۶

فهرست جداول

جدول ۱-۲-جدول مقایسه ابزارهای مورد استفاده در تشخیص زبان اشاره.....	۵۵
جدول ۲-۲-مقایسه ویژگی های مختلف مورد استفاده در تشخیص زبان اشاره.....	۵۷
جدول ۳-۲-مقایسه روش های دسته بندی مختلف در تشخیص زبان اشاره.....	۵۸
جدول ۱-۳-لیست ویژگی های استخراج شده به همراه نام اختصاری آن ها.....	۸۵
جدول ۱-۴-نمونه ای از یک ماتریس در هم ریختگی.....	۹۲
جدول ۲-۴-چهار تعریف مهم بر روی جدول در هم ریختگی.....	۹۳
جدول ۳-۴-نتایج بدست آمده توسط سیستم پیشنهادی به تفکیک پارامترهای مختلف.....	۹۶
جدول ۴-۴-ماتریس در هم ریختگی بهترین نتیجه بر روی کل پایگاه داده مورد استفاده بر مبنای درصد. سطر آخر دقت هر سطر را مشخص می نماید.....	۱۰۳
جدول ۵-۴-مقایسه روش پیشنهادی و روش ارائه شده توسط پوگالت و بُودِن [45].....	۱۰۵
جدول ۶-۴-جدول مقایسه روش پیشنهادی با روش های پیشین.....	۱۰۶
جدول ۱-۵-حرکات شناسایی شده در سیستم پیشنهادی ونمادهای آن ها در سیستم های نوشتاری مختلف.....	۱۰۹
جدول ۲-۵-مقایسه روش ارائه شده و یکی از روش های پیشین.....	۱۱۳
جدول ۳-۵-جدول همبستگی ویژگی های استخراج شده از تصویر قطعه بندی شده رنگی دست.....	۱۱۷
جدول ۴-۵-زمان پردازش سیستم پیشنهادی برای ۲۴۰۰ نمونه و ۲۶ ویژگی به تفکیک زیر بخش ها. زمان های محاسبه شده مستقیم به صورت زیر خطی مشخص و دیگر داده ها به صورت سطری محاسبه شده اند.....	۱۱۹

فصل ١: مقدمه

۱-۱- پیش‌گفتار

از دیرباز انسان‌ها به عنوان موجوداتی اجتماعی، نیاز به تبادل اطلاعات و برقراری ارتباط بین یکدیگر را احساس کرده بودند و جهت این ارتباط زبان را ابداع کردند. تعریف زبان علاوه بر زبان گفتاری، زبان‌هایی را نیز که از طریق اشاره باعث ارتباط می‌شوند را شامل می‌شود. گرچه به نظر می‌رسد که استفاده از زبان اشاره^۱ با توسعه زبان گفتاری رو به فراموشی رفته باشد، با اندکی تامل در می‌یابیم که استفاده از زبان اشاره هنوز کاربردهای خاص خود را از دست نداده است. از کاربردهای زبان اشاره می‌توان به استفاده از زبان اشاره برای برقراری ارتباط بین افراد ناشنوا، انتقال فرمان‌های نظامی، که به دلیل خاموشی رادیویی و گفتاری تنها راه برقراری ارتباط در عملیات‌ها می‌باشد، برقراری سریع ارتباط در مسابقات ورزشی و دیگر کاربردهای روزمره که انسان زبان بدن را به زبان گفتاری ترجیح می‌دهد نام برد.

با توجه به پیشرفت رایانه‌ها و ابزارهای ارتباطی در عصر حاضر، بسترهای لازم و نیاز به تبادل الکترونیکی اطلاعات به وجود آمده است و در کنار نیاز به انتقال گفتار و نوشتار به صورت الکترونیکی، نیاز به تبادل اطلاعات از طریق زبان اشاره نیز، به خصوص برای افراد ناشنوا که معمولاً قادر به یادگیری زبان‌های معمول نیستند، از این طریق ایجاد شده است. از آنجا که انتقال زبان اشاره معمولاً از طریق انتقال ویدئو میسر می‌شود، از این رو نیاز به ابزارهایی جهت شناسایی در مبدأ، مدل کردن به زبان‌های رایانه‌های زبان اشاره و سپس تبدیل به نمایش سه بعدی در مقصد می‌تواند تا حد بسیار زیادی در جهت کاهش ترافیک باند انتقالی مورد نظر میسر باشد. همچنین امکان اصلاحات و حتی ترجمه زبان‌های اشاره نیز از این طریق فراهم می‌شود.

¹ Sign Language

مبحث شناسایی خودکار زبان اشاره^۱ به شناسایی زبان اشاره با استفاده از رایانه می‌پردازد که در واقع ترکیبی از روش‌های نرم افزاری شناسایی و دسته‌بندی بخش‌های زبان و سخت افزارهای دریافت اطلاعات از فرد اجرا کننده آن می‌باشد [۱]. این سخت افزارها می‌توانند از دوربین‌های معمولی تا دستکش‌ها و شتاب‌سنج‌های سه بعدی را شامل شوند.

در این تحقیق با بررسی روش‌های پیشین بر حسب ابزارهای سخت افزاری مورد استفاده و روش‌های شناسایی نرم افزاری، روشی با استفاده از دوربین‌های عمقی ارائه می‌شود که علاوه بر افزایش صحت و دقت بتواند به راحتی امکان استفاده از این سیستم را نیز به کاربران بدهد. در ادامه به توضیحاتی در مورد زبان اشاره، بررسی روش‌های پیشین و ارائه توضیحات مفصل در مورد روش پیشنهادی پرداخته خواهد شد.

۱-۲- زبان اشاره

زبان‌های رایج ارتباطی افراد کر و لال در جهان زبان‌های اشاره هستند. زبان‌های اشاره دارای ساختارهای ویژه به خود هستند و معمولاً از لحاظ ساختاری و تطابق یک به یک، با زبان‌های گفتاری کاملاً متفاوت هستند. حداقل ۱۳۰ زبان اشاره شناخته شده در جهان وجود دارد [۲] که هریک دارای ساختار زبانی خود هستند. این زبان‌ها به دو نوع زبان‌های اصلی^۲ یا نمونه^۳ A مانند زبان اشاره امریکایی^۴ و زبان‌های نمونه^۵ R یا زبان‌های مشتق شده^۶ از دیگر زبان‌های نمونه^۷ A مانند زبان اشاره آلمانی^۷ تقسیم می‌شوند [۳].

^۱ Automatic Sign Language Recognition

^۲ Original

^۳ Prototype A

^۴ American Sign Language (ASL)

^۵ Prototype R

^۶ Derived

^۷ German Sign Language

۱-۲-۱- اهمیت و کاربرد زبان اشاره

زبان اشاره از این جهت مورد توجه واقع شده است که بیشتر افراد کر و لالی که در جهان وجود دارند برای یادگیری زبان های معمولی گفتاری یا نوشتاری دچار مشکل هستند و زبان های اشاره و زبان های رایج نیز به راحتی قابل ترجمه به یکدیگر نیستند [۲]. گفته شده است که ۹۵٪ کودکان کر و لال از والدین سالم متولد می شوند و این والدین برای آموزش این زبان به کودکانشان دچار مشکل بوده و حتی اولین برخورد این کودکان با زبان اشاره ساخت یافته در دبستان اتفاق می افتد که مشکلات زیادی را برای این کودکان به دنبال دارد [۲].

استفاده افراد معلول از این زبان در ارتباطات بین المللی نیز با چالش هایی همراه بوده و علت اصلی آن تفاوت ساختاری زبان های اشاره کشور های مختلف است. برای مثال چنانچه یک فرد معلول ایرانی بخواهد با فردی معلول که در کشوری خارجی زندگی می کند ارتباط برقرار کند، دچار مشکل خواهد شد. همانطور که گفته شد، یادگیری زبان های معمول (مثلا زبان انگلیسی) برای این افراد بسی دشوارتر و غیر معمول تر از یادگیری زبان اشاره است. از این رو تکامل سیستم های کامپیوتری که بتوانند ارتباط افراد معلول را از طریق زبان اشاره با رایانه و یا دیگر افراد ممکن سازند، از اهمیت زیادی برخوردار است. ترجمه زبان های اشاره به یکدیگر نیز چالشی است که در راستای تحقیق و پژوهش بیشتر در این حوزه قابل دستیابی خواهد بود.

علاوه بر اهمیت بالقوه ای که توسعه سیستم های نرم افزاری و سخت افزاری در حوزه زبان اشاره معلولان دارد، توسعه سیستم های شناسایی و ترجمه زبان اشاره در دیگر کاربردهای زبان اشاره نیز حائز اهمیت می باشد. برای مثال در عملیات های نظامی و یا فضایی که ممکن است به هردلیل امکان مکالمه گفتاری و یا رادیویی وجود نداشته باشد، زبان اشاره نقش به سزایی را ایفا می کند و قطعاً با پیشرفت تکنولوژی و بوجود آمدن تجهیزات و یا ربات های هوشمند به جای سربازان و یا فضانوردان، این نوع ارتباطات نیز از اهمیت بسزایی برخوردار خواهند شد.

۱-۲-۲- ساختارهای زبان اشاره

زبان های اشاره از ساختارهای متفاوتی تشکیل می شوند که زبان اشاره توسط آن ها و یا ترکیبی از آن ها اجرا می شود. برخی از مهمترین آن ها در زیر آمده است [۲]:

۱-۲-۲-۱- هجی با انگشتان^۱

در این ساختار برای حروف الفبا (معمولاً حروف الفبای زبان های گفتاری) حرکاتی از زبان اشاره مشخص می شود که توسط آن ها کلمات به صورت حرف به حرف هجی می شوند. در صورتی که برای برخی از کلمات یا اسامی اشاره^۳ خاصی وجود نداشته باشد، معمولاً از این نوع ساختار برای برقراری ارتباط استفاده می شود.

۱-۲-۲-۲- اشاره های دست^۴

به مجموعه ای از اشاره های زبان اشاره اطلاق می شود که از شکل دست ها، مکان و حرکت آن ها برای انتقال معنی استفاده می شود. بیشتر کلمات در زبان های اشاره معمولاً دارای معادلی با اشاره های دست هستند.

۱-۲-۲-۳- اشاره های صورت^۵

اشاره های صورت، شامل حرکات ابرو و یا دهان نیز در ادای مفاهیم و یا کامل کردن مفاهیم کاربرد دارند. برای مثال در زبان اشاره آمریکایی، از ابروها برای تغییر نوع سوال در گروهی از اشاره ها استفاده می شود.

¹ Structure

² Finger Spelling

³ Gesture

⁴ در این گزارش منظور از زبان اشاره همان Sign Language و منظور از اشاره Gesture می باشد.

⁵ Hand Gestures

⁶ Face Gestures

۱-۲-۲-۴- اشاره های بدن^۱

حرکات قسمت بالا تنه بدن^۲ نیز در انتقال مفاهیم به خصوص در تاکید بر روی کلمات ادا شده توسط دست و یا چهره از طریق انبساط دادن^۳ و فشرده کردن^۴ بدن مورد استفاده قرار می گیرند.

۱-۲-۲-۵- خم کردن^۵ها

معنای برخی اشاره ها با خم کردن برخی از اعضای بدن مانند انگشتان تغییر می کند. برای مثال جنباندن انگشتان^۶ در حین انجام یک اشاره با دست، می تواند معنای آن را تغییر دهد.

۱-۲-۲-۶- استفاده از فضا^۷

یکی از راه های انتقال معنا در زبان های اشاره استفاده از فضای اطراف می باشد. استفاده از اشیاء و مکان های آن ها می تواند به رساندن معانی کمک بسزایی نماید. برای مثال قرار دادن دست ها یا اشیاء در زیر، کنار، جلو و یا پشت یکدیگر می توانند معانی خاص خود را بدهد. همچنین رفتارهایی مانند راه رفتن و حرکت دادن و یا انجام کارهایی مانند جدا کردن^۸ و یا تصادم^۹ اشیاء می تواند معانی خاصی را بدهد که تنها با استفاده از فضای اطراف امکان پذیر می باشند.

¹ Body Gestures

² Torso

³ Expansion

⁴ Compression

⁵ Inflections

⁶ Wiggle Fingers

⁷ Spatial Use

⁸ Separation

⁹ Collision

۱-۲-۳-نمادگذاری زبان های اشاره^۱

در نوشتن نیز زبان اشاره تفاوت هایی با زبان های گفتاری معمولی دارد [۴]. دلیل اصلی تفاوت آن در این است که ادای کلمات در زبان اشاره تنها با استفاده از یک دست صورت نمی گیرد. در بسیاری از موارد صورت، هردو دست، بدن و یا سر نیز دخالت دارند و زبان نمادین اشاره نیز بایستی قابلیت ارائه این بخش از زبان را نیز داشته باشد.

نمادگذاری های متفاوتی برای زبان های اشاره در جوامع مختلف مورد استفاده قرار می گیرد ولی در کارهای تحقیقاتی تعدادی خاص از این سیستم ها بیشتر مورد استفاده قرار می گیرند. در زیر به برخی از آن ها اشاره می کنیم:

۱-۲-۳-سیستم Stokoe

این سیستم که به نام سازنده آن، ویلیام استوکو^۲، نام گرفته است از ابتدا از ۵۵ نماد که در سه گروه مکان^۳، شکل دست^۴ و حرکت^۵ جای گرفته اند، تشکیل شده بود [۴]. این سیستم با استفاده گسترده و خاص توسعه یافت ولی یک توسعه استاندارد برای آن در نظر گرفته نشده است و امروزه در واقع مجموعه ای از سیستم های نوشتاری زبان اشاره برپایه سیستم اولیه را شامل می شود.

۱-۲-۳-سیستم HamNoSys^۶

این سیستم نوشتاری که یک شکل کلی تر از سیستم Stokoe می باشد در سال ۱۹۸۹ ارائه گردید [۵]. در این سیستم ۲۰۰ نماد وجود دارد که علاوه بر سه گروه نماد های سیستم Stokoe،

^۱ Sign Language Notation

^۲ William Stokoe

^۳ Location

^۴ Hand Shape

^۵ Movement

^۶ The Hamburg Sign Language Notation System

گروه پیکربندی دست^۱ نیز قرار داده شده است. علاوه بر آن، در این سیستم ثبت حرکات صورت^۲ نیز پیش بینی شده است.

۱-۲-۳-سیستم SiGML^۳

این سیستم که بر پایه سیستم HamNoSys بنا شده است، با ترکیب نماد های آن سیستم و زبان XML^۴ ترکیبی انعطاف پذیر را پدید آورده است که در آن اضافه کردن زیر ساختار ها نیز ممکن است و علاوه بر قابلیت توسعه آسان، قابلیت استفاده آن در سیستم های اطلاعاتی و رایانه ای دیگر نیز ممکن می باشد.

۱-۳-تعریف مسئله

با توجه به موارد کاربرد زبان اشاره که در بالا اشاره شد، توسعه و ارتقاء سیستم های رایانه ای که بتواند زبان اشاره را تشخیص دهد و در ارائه سرویس های تعامل انسان با رایانه^۵، ترجمه زبان اشاره و ... به افراد کر و لال و سایر افراد ذینفع بتواند یاری دهنده باشد، بسیار حائز اهمیت است. سیستم هایی که بتوانند در محیط های معمول با پس زمینه پیچیده و تغییرات روشنایی و همچنین در تعامل با افراد متفاوت به شناسایی زبان اشاره پردازند هنوز در حال توسعه بوده و نیاز به تکامل دارند. علاوه بر الگوریتم ها و روش های پیشرفته ای که در این زمینه پیشنهاد شده است، ابزارهای متفاوتی نیز برای دریافت اطلاعات از محیط توسعه یافته اند و هنوز نیز در حال توسعه اند.

در این تحقیق قصد داریم کاربرد یکی از جدید ترین ابزارها در تشخیص اشاره جهت بازی های ویدئویی را در تشخیص زبان اشاره توسط رایانه بررسی نماییم. سنسور کینکت^۶ که در اواخر سال

^۱ Hand Configuration

^۲ Facial Expression

^۳ The Sign Gesture Meta Language

^۴ Extensible Mark-up Language

^۵ Human Computer Interaction (HCI)

^۶ Kinect Sensor

۲۰۱۰ توسط شرکت مایکروسافت^۱ برای بازی کردن بدون دسته با دستگاه بازی^۲ XBOX 360 ارائه شد، با در کنار هم قرار دادن یک دوربین عمقی و یک دوربین رنگی و قیمتی مناسب، امکان جدیدی فراهم آورد تا علاوه بر کاربرد سرگرمی، برای کاربردهای دیگر مانند تشخیص زبان اشاره نیز بتوان از این دستگاه بهره برد.

در این پایان نامه با امکان سنجی استفاده از دوربین های عمقی (به شکل خاص دستگاه کینکت) برای تشخیص زبان اشاره، اهمیت استفاده از عمق در قطعه بندی دست نشان داده می شود و با ارائه روش پیشنهادی جهت استفاده از عمق برای قطعه بندی دست در تصویر رنگی، فرآیند تشخیص زبان اشاره بر روی یک پایگاه داده استاندارد انجام شده و نتایج بدست آمده با روش های مشابه که بدون استفاده از قطعه بندی عمقی می باشند مقایسه می شود. به طور خلاصه در مسئله بهبود شناسایی زبان اشاره، روش قطعه بندی دست در تصویر رنگی با استفاده از اطلاعات عمقی ارائه شده و صحت این پیشنهاد نیز با پیاده سازی سیستمی جهت آزمایش این روش نشان داده شده است.

۱-۴- کاربردها

مهمترین کاربرد شناسایی زبان اشاره، توسعه سیستم های کامپیوتری جهت بهبود ارتباطات و سطح زندگی ناشنوایان است. کاربردهای دیگری از جمله انتقال مفاهیم در خانه های هوشمند و همچنین ارتقاء سیستم های تعامل با ربات ها و یا ابزارآلات پیشرفته نظامی نیز می تواند از کاربرد های زبان اشاره باشد. هم اکنون استفاده از زبان اشاره و اشارات بدن برای تعامل با سیستم های کامپیوتری مخصوصاً بازی های ویدئویی به شدت رونق یافته است و ارتقاء سطح شناسایی سیستم های موجود در زمینه شناسایی زبان اشاره نقش بسزایی در افزایش کارایی و تاثیر پذیری اینگونه تفریحات کامپیوتری-بدنی نیز دارد.

¹ Microsoft

² Gaming Console

۱-۵- فرض ها و محدودیت ها

نتایج این تحقیق با فرض استفاده از اطلاعات عمقی و رنگی سنسور کینکت بیان شده است و جهت استفاده و یا تکرار نتایج حاصله بایستی از این سنسور و یا دستگاهی مشابه استفاده شود. همچنین داده های بخش اصلی این تحقیق با استفاده از تصاویر پایگاه داده ای می باشد که اطلاعات دریافتی آن از سنسور کینکت با مقداری برش اولیه همراه بوده و بخش اصلی سیستم پیشنهادی مستقیماً بر داده های اولیه سنسور کینکت بنا نشده است ولی قطعاً بر داده های اولیه سنسور کینکت نیز قابل انطباق می باشد.

فصل ۲: مروری بر

کارهای انجام شده

در این بخش به بررسی کارهای پیشین در حوزه زبان اشاره خواهیم پرداخت. کارهای پیشین را از دیدگاه های مختلف بررسی و دسته بندی خواهیم کرد و در نهایت جایگاه کار فعلی را در این میان بیان می کنیم. دسته های مختلفی که کارهای پیشین در آن ها قرار خواهند گرفت، به شرح زیر می باشند:

۲-۱- طبقه بندی بر حسب حوزه شناسایی زبان اشاره

کارهای پیشین در حوزه زبان اشاره، بسته به نوع کاربرد، معمولاً بخش خاصی از گستره وسیع زبان اشاره را به عنوان دامنه مورد شناسایی هدف گرفته اند. برخی در حوزه قطعه بندی دست، برخی در حوزه شناسایی حروف الفبا، برخی کلمات، برخی جملات و برخی نیز به تشخیص اشاره ها پرداخته اند. می توان گفت که این حوزه ها جدا از یکدیگر نیستند و برخی پیش زمینه ی دیگری و برخی در هم آمیخته با حوزه دیگر هستند. برای مثال قطعه بندی دست پیش نیاز مراحل دیگر است و تشخیص حروف الفبا به درستی و در محیط های پیچیده می تواند پایه خوبی برای تشخیص کلمات و جملات باشند. جملات نیز معمولاً از کلمات تشکیل شده اند و در واقع بدون تشخیص کلمات، تشخیص جملات ممکن نمی باشد. در زیر کارهای پیشین مربوط به هر حوزه را بررسی می نمایم:

۲-۱-۱- حروف الفبا

حروف الفبای زبان های اشاره معمولاً با اشاره های ایستای انگشتان دست (بدون حرکت) مشخص می شوند. در زبان اشاره آمریکایی، از ۲۶ حرف الفبا، ۲۴ حرف آن بدون حرکت دست انجام می پذیرد و حروف J و Z نیز با حرکت دست همراه هستند. از این رو معمولاً در اولین حرکت از شناسایی زبان اشاره، به شناسایی ۲۴ حرف ایستای این زبان پرداخته می شود. از کارهای پیشین در این حوزه می توان به کارهای [۶-۱۱] اشاره کرد.

۲-۱-۲- کلمات

کلمات در زبان های اشاره معمولاً از حروف الفبا ساخته نمی شوند و هریک دنباله ای از حرکات مخصوص به خود را دارند. شناسایی کلمات معمولاً شامل شناسایی سلسله حرکات آن ها از روی ویدئو و تشخیص شکل ظاهری دست ها می باشد. از کارهایی که در حوزه تشخیص کلمات زبان اشاره انجام گرفته است می توان کارهای [۲، ۱۲-۲۱] را نام برد.

۲-۱-۳- جملات

تشخیص جملات در زبان اشاره علاوه بر شامل شدن تشخیص کلمات، حوزه هایی مربوط به شناخت دستور زبان^۱ این زبان ها را نیز شامل می شود و با بررسی احتمال جای گرفتن کلمات در الگو هایی از جملات همراه است. [۲، ۱۳] از کارهایی هستند که به شناسایی جملات زبان اشاره پرداخته اند.

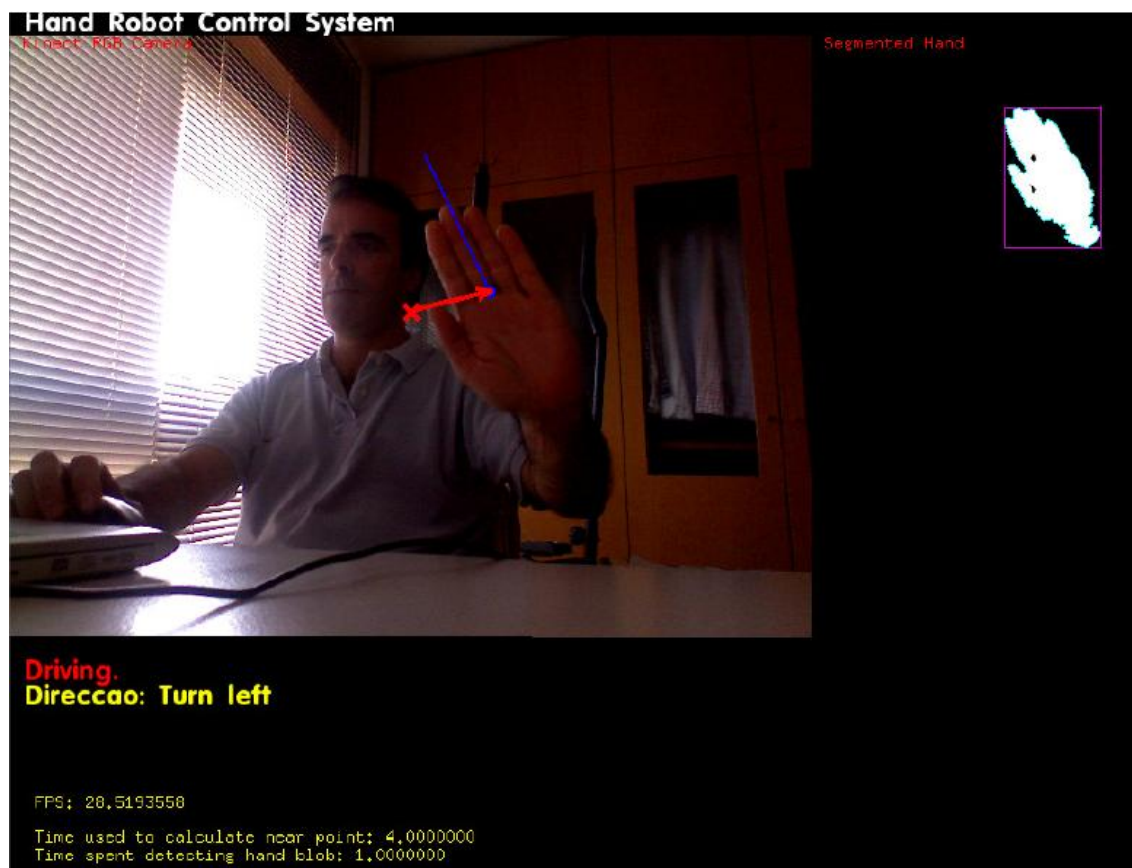
۲-۱-۴- تشخیص اشاره ها

تشخیص اشاراتی به جز اشاره های موجود در زبان اشاره نیز به نحوی پایه شناسایی زبان اشاره محسوب شده و در بسیاری از کارهای پیشین مورد توجه قرار گرفته است. این اشارات نیز ممکن است اشارات مربوط به حالت ایستای دست از روی تصاویر اشارات باشد و یا حرکت این اشاره ها را مورد شناسایی قرار دهند.

کار انجام شده در [۲۲] در سال ۲۰۱۱ مربوط به شناسایی مکان انگشتان و مرکز دست، تحقیق [۲۳] در مورد روش های شناسایی شکل ظاهری دستان از کارهای انجام شده در این حوزه می باشد. همچنین کار مربوط به [۲۴] نیز در راستای شناسایی ۴ حرکت اشاره می باشد که با

¹ Grammar

استفاده از دوربین های عمقی انجام گرفته است. همچنین در [۲۵] از کینکت برای تشخیص حرکات اشاراتی برای کنترل یک روبات استفاده شده است. در شکل ۱-۲ نمایی از تصویر ورودی، تصویر قطعه بندی شده دست و حرکت تشخیص داده شده (چرخش روبات به چپ) در سیستم پیشنهادی [۲۵] نمایش داده شده است.



شکل ۱-۲- تصویر ورودی، تصویر قطعه بندی شده دست و حرکت تشخیص داده شده (چرخش روبات به چپ) در سیستم پیشنهادی [۲۵]. تصویر از [۲۵].

۲-۱-۵- قطعه بندی دست

کارهای بسیاری نیز صرفاً به قطعه بندی دست پرداخته اند. همانطور که گفته شد قطعه بندی دست از اولین و پایه ای ترین مراحل هر سیستم تشخیص زبان اشاره محسوب می باشد و راه حل های موجود برای تشخیص کامل و صحیح دست در تمامی محیط ها و شرایط نوری در حال توسعه است. کارهای [۲۶, ۲۷] از کارهایی هستند که به قطعه بندی دست پرداخته اند.

۲-۲- طبقه بندی کارهای پیشین بر حسب ابزار مورد

استفاده

در این قسمت ابزارهایی که در کارهای مختلف برای ذخیره و دریافت اطلاعات اجرای زبان اشاره مورد استفاده قرار گرفته است بررسی می شود. از آنجا که اطلاعات اجرای زبان اشاره بایستی توسط دستگاه هایی ضبط شده و به عنوان داده ورودی به سیستم داده شود، به دلیل تنوع ویژگی^۱ هایی که با تغییر دستگاه ها به ورودی سیستم اعمال می گردد، انواع مختلفی از دستگاه ها نیز مورد استفاده قرار گرفته اند. برخی از دستگاه ها بر پایه اطلاعات دیداری^۲ مانند انواع دوربین های رنگی، سطح خاکستری^۳ و یا انواع دوربین های سه بعدی نسبت به جمع آوری اطلاعات از محیط اقدام می-کنند و برخی دیگر نیز با استفاده از ابزارهای متصل به دست مانند دستکش های مخصوص و مجهز به سنسور هایی نظیر شتاب سنج های سه بعدی به دریافت اطلاعات حرکتی می پردازند.

استفاده از دوربین های معمولی و ارزان قیمت در افزایش کاربری سیستم در بین افراد کمک خواهد کرد ولی به دلیل اینکه دوربین های سه بعدی و سنسور هایی که به دست متصل می شوند، اطلاعات بیشتری را می توانند دریافت کنند، دقت بالاتری را نیز به سیستم ارائه می دهند. استفاده از دوربین های سه بعدی، تا قبل از ظهور کینکت هزینه های بسیار زیادی را برای سیستم می طلبیده است و استفاده از سنسورهای متصل به دست نیز راحتی استفاده از سیستم را از کاربران می گرفته است.

¹ Feature

² Visual

³ Gray-Scale

۲-۱-۲- کارهای پیشین با استفاده از دستکش و سنسور ها

استفاده از دستکش ها و سنسور ها (به شکل خاص شتاب سنج ها) به چند دلیل انجام می - شود. دستکش ها معمولا به دو صورت مورد استفاده قرار می گیرند. حالت اول از دستکش های رنگی استفاده می شود که در هنگام کار با دوربین های رنگی بتوانند دست ها را راحت تر قطعه بندی^۱ کنند تا با بخش های مختلف بدن مانند صورت و گردن در قطعه بندی اشتباه گرفته نشوند. مورد دیگر استفاده از دستکش ها قراردادن سنسورهایی بر روی انگشتان می باشد که این سنسورها معمولا از نوع شتاب سنج های سه بعدی^۲ هستند. ادغام اطلاعات بدست آمده از تصویر و داده های مربوط به حرکت انگشتان و دست در فضا با استفاده از شتاب سنج ها می تواند ترکیب خوبی از ویژگی ها را برای فرآیند شناخت زبان اشاره فراهم سازد. استفاده از شتاب سنج ها و سنسورهای قابل نصب بر روی دست و بازو و بدون استفاده از دستکش نیز مانند مورد فوق می تواند اطلاعاتی در مورد حرکت دست و یا بازوان به ورودی سیستم بدهد. کارهای مختلفی با استفاده از دستکش های سنسوری [۲۸، ۲۹] و یا دستکش های رنگی [۳۰] انجام شده است. در زیر نیز به بررسی مختصر چند کار با استفاده از سنسور ها و دستکش ها می پردازیم.

یکی از مهمترین پروژه ها در راستای شناسایی زبان اشاره، پروژه بازی CopyCat می باشد که سال ها به عنوان پروژه تحقیقاتی در دانشگاه جورجیا تک^۳ در کشور ایالات متحده امریکا در حال توسعه می باشد [۱، ۳۱-۳۴]. این بازی که در واقع یک بستر^۴ برای جمع آوری اطلاعات و داده های تشخیص اشاره نیز می باشد که در قالب این بازی به کودکان کر و لال به یادگیری زبان اشاره و همچنین تقویت حافظه کمک می شود. در شکل ۲-۲ تصویری از محیط کار و خود بازی CopyCat دیده می شود که در آن کودکی خردسال در حال تعامل با بازی می باشد [۱]. در شکل وجود دوربین

¹ Segmentation

² 3D Accelerometer

³ Georgia Tech

⁴ Platform

رنگی در بالای صفحه نمایشگر، همچنین دستکش های رنگی در دست کودک و همچنین سنسور شتاب سنج بر روی دست وی دیده می شود.



شکل ۲-۲- تصویری از محیط کار و بازی CopyCat. برگرفته از سایت اصلی پروژه در [۱]

پروفسور تاد اشتارنر^۱ از این دانشگاه مسئول این پروژه تحقیقاتی می باشد که از سال ۲۰۰۵ تاکنون دانشجویان و رساله های زیادی را در این راستا رهبری نموده و مقالات زیادی از این پروژه به چاپ رسانده است. جدید ترین رساله دکتری که در این راستا به چاپ رسیده است رساله خانم هلن براشیر^۲ [۲] می باشد که در سال ۲۰۱۰ به چاپ رسیده است.

در این رساله دکتری که بر روی بهبود ابزارهای تمرین زبان اشاره کار شده است [۲] سعی شده است تا با وارد کردن معیارهای زبانی به سیستم تشخیص زبان اشاره به خصوص داده های بازی CopyCat در بهبود قدرت تشخیص این سیستم ها تلاش شود. ورودی های مورد استفاده در سیستم این بازی به صورت ترکیبی از داده های تصویری و داده های سنسورهای شتاب سنج ها تشریح شده

¹ Thad Starner

² Helene Brashear

است. هر فرد در یک اصطلاحاً کیوسک^۱ قرار می گیرد که در آن حرکات فرد توسط دوربین تحت کنترل قرار می گیرد. در شکل ۳-۲ تصویری از یک کیوسک دیده می شود [۲].



شکل ۳-۲- تصویری از یک کیوسک مخصوص بازی CopyCat که توسط دوربین از حرکات دست تصویربرداری می شود. [۲]

برای افزایش کارایی در این سیستم از دستکش های رنگی با رنگ های بنفش و صورتی برای بخش بینایی ماشین و از سنسور های شتاب سنج سه بعدی برای افزایش ویژگی های ورودی استفاده شده است. در بالای شکل ۴-۲ تصویر این دستکش ها به همراه شتاب سنج های نصب شده بر روی آن و در پایین این شکل تصویر شتاب سنج سه بعدی و جزئیات آن دیده می شود [۲]. مشکل اصلی در این کار استفاده از دستکش ها و سنسورهای متصل به دست کودکان می باشد که برای آن ها نمی تواند خوشایند باشد و قطعاً تجربه ای بدون نیاز به این دستگاه های اضافی برای آنان خوشایند خواهد بود.

¹ Kiosk



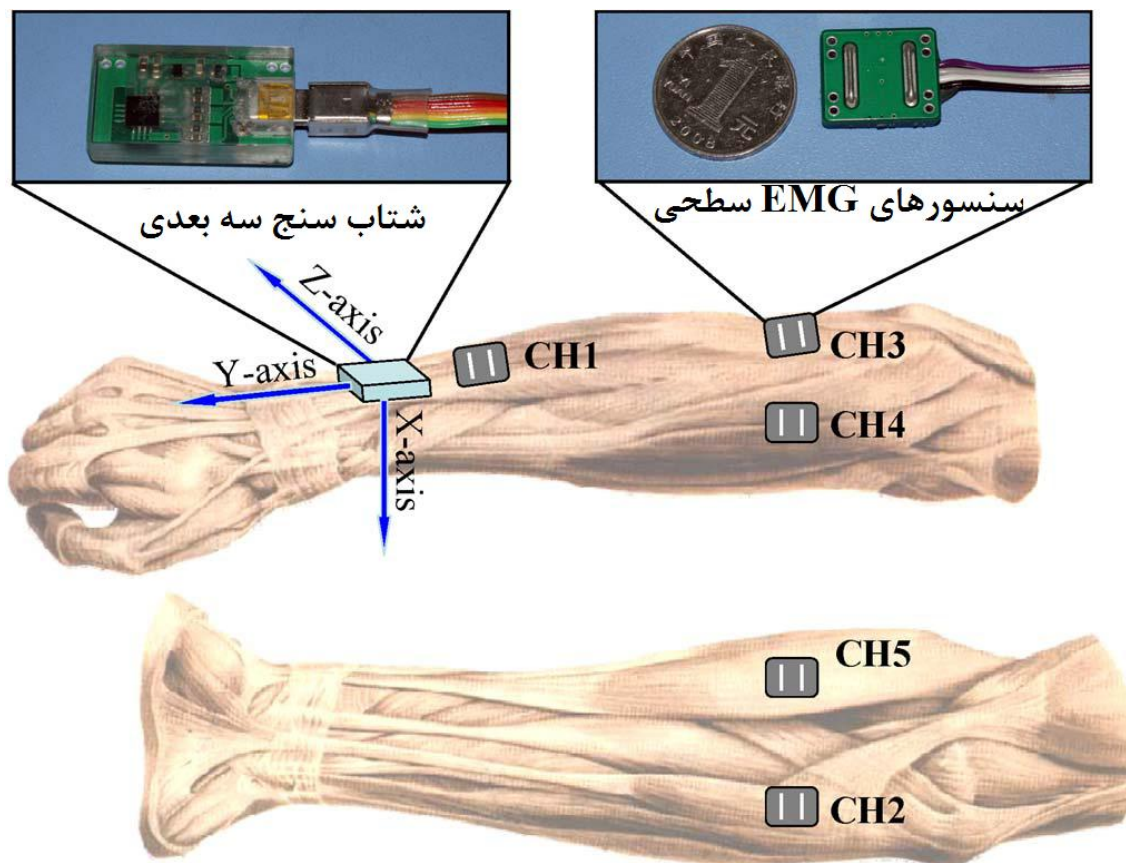
شکل ۲-۴- تصویری از دستکش های رنگی به همراه شتاب سنج (بالا) و جزئیات شتاب سنج سه بعدی (پایین) به کار رفته در بازی Copy Cat [۲]

از دیگر کارهایی که اخیرا با استفاده از سنسور ها انجام شده کاری توسط ژانگ^۱ و همکارانش است [۳۵] که در سال ۲۰۱۱ در ژورنالی از IEEE به چاپ رسیده است. در این کار ۷۲ کلمه زبان اشاره چینی با استفاده از سنسور های شتاب سنج و حسگر های ماهیچه ای EMG^۲ که بر روی دست نصب شده اند مورد شناسایی قرار گرفته اند و شناسایی با دقت ۹۵.۳٪ و ۹۶.۳٪ بر روی دو فرد متفاوت بدست آمده است. سنسور های بکار رفته در این آزمایش از یک سنسور شتاب سنج سه بعدی و ۵ کانال سنسور EMG تشکیل شده است که همگی بر روی دست و بازوی فرد متصل می شوند. در شکل ۲-۵ تصویری از این سنسور ها و مکان قرارگیری آن ها بر روی ماهیچه مشاهده می شود. بستر

¹ Xu Zhang

² Electromyogram

ایجاد شده در این کار علاوه بر کارایی در زبان اشاره، در بررسی و تحلیل حرکات دست در بازی مکعب روبیک^۱ نیز کارایی خود را نشان داده است [۳۵]. وابستگی به استفاده از این سنسورها به خصوص عدم طبیعی بودن نصب سنسورها بر روی ماهیچه ها برای کار با سیستم از مشکلات استفاده از این روش می باشد. البته قدرت بالای شناسایی در این روش را نیز بایستی در نظر گرفت.



شکل ۲-۵- سنسورهای EMG و شتاب سنج سه بعدی نصب شده بر روی دست در کار ژانگ و همکارانش. تصویر از [۳۵].

هوانگ^۲ و همکارانش [۳۶] نیز در مقاله ای در سال ۲۰۱۱ با استفاده از یک دستکش با ۵ سنسور متصل به انگشتان به بررسی اشارات پرداخته اند. در این مقاله به بررسی تشخیص باز و یا بسته بودن انگشتان پرداخته شده است و با ۲۰ بار تکرار هر حالت از باز یا بسته بودن انگشتان (جمعا ۳^۵ معادل ۳۲ حالت) پایگاه داده ای با ۶۴۰ نمونه ایجاد شده است. بالاترین دقت با روش پیشنهادی

¹ Rubik's Cube

² Yu Huang

۷۲.۲٪ بیان شده است. در شکل ۶-۲ تصویری از دستکش داده ای مورد استفاده در این کار ارائه شده است [۳۶]. علاوه بر مشکل استفاده از دستکش ها در این روش، درصد پایین شناسایی نیز از معایب این کار می باشد.



شکل ۶-۲-نمای رویی و پشتی دستکش مورد استفاده در تحقیق هوانگ و همکارانش [۳۶]

دیگر کاری که جهت تشخیص زبان اشاره و در سال های اخیر از سنسور ها و دستکش ها استفاده کرده است تحقیقی از ساجاکارن^۱ و همکارانش [۳۷] در سال ۲۰۱۱ است که با استفاده از شبکه های حسگری بدن^۲ یا BSN به شناسایی حروف، کلمات و یک جمله زبان اشاره پرداخته شده است. در تصویری که در شکل ۷-۲ مشاهده می کنید، نمایی از دستکش و شبکه حسگری بدن مورد استفاده در این تحقیق نمایش داده شده است [۳۷]. دستکش فوق علاوه بر ۵ سنسور جهت انگشتان از یک شتاب سنج سه بعدی نیز بهره برده است. داده های مورد استفاده در این سیستم از ۱۰ فرد جوان و هر فرد ۲۸ اشاره با ۵ تکرار گرفته شده است و ۱۱ ویژگی نیز برای این کار استخراج شده

¹ Satjakarn Vutinuntakasame

² Body Sensor Networks (BSN)

است. بالاترین دقت بدست آمده در این کار ۷۳.۶٪ بوده است که با توجه به استفاده از دستکش ها درصد بالایی محسوب نمی شود. البته تعداد متفاوت افراد در شناسایی این ۲۸ اشاره نقطه قوت این کار به شمار می آید.



شکل ۲-۷-نمایی از دستکش و شبکه حسگری بکار رفته در تحقیق ساجاکارن و همکارانش. [۳۷]

۲-۲-۲- کارهای پیشین با استفاده از دوربین های رنگی

از مهمترین و پایه ای ترین ابزارهای شناسایی زبان اشاره می توان به دوربین های رنگی معمولی اشاره نمود. این دوربین ها که معمولا تصاویر رنگی را در سه کانال رنگی قرمز، آبی و سبز ذخیره می کنند، به دوربین های RGB^۱ نیز معروفند. با بررسی در کارهای انجام شده در چند دهه

^۱ Red-Green-Blue

اخیر به جرات می توان گفت که اکثر کارها از دوربین های رنگی به عنوان ابزار اصلی استفاده کرده اند و ابزارهای دیگر در اغلب اوقات به عنوان ابزار کمکی به حساب می آمده اند. این امر نیز قابل پیش بینی می باشد زیرا زبان اشاره نیز به صورت پیشفرض در قالب ویدئو و تصاویر رنگی برای انسان قابل فهم می باشد و اکثر افراد نیز برای ضبط داده های خود از این دوربین ها استفاده می کنند. حتی اخبارهای مربوط به ناشنوایان نیز با استفاده از دوربین های رنگی ضبط شده و قابل دسترس می باشند و دسترسی به ابزارهای مختلف از جمله دستکش ها، شتاب سنج ها و دوربین های خاص منظوره برای همه افراد میسر نمی باشد. افراد کر و لال به راحتی از طریق برقراری ارتباط ویدئویی با استفاده از دوربین های رنگی معمولی از طریق سیستم های فعلی تلفن همراه می توانند با یکدیگر از راه دور ارتباط برقرار کنند درحالی که به دیگر ابزارهای خاص براحتی دسترسی نخواهند یافت.

اکثر کارهای انجام شده در حوزه زبان اشاره حداقل در بخش های استخراج ویژگی و یا قطعه بندی تصویر از دوربین های رنگی استفاده کرده اند [۳۰، ۳۸، ۳۹]. در بازی Copy Cat نیز که قبلاً توضیح داده شد، از دوربین رنگی برای قطعه بندی تصویر و استخراج ویژگی های مورد نظر استفاده شده است. چند مورد از کارهای سال های اخیر که صرفاً با استفاده از دوربین های رنگی معمولی بوده است در زیر مختصراً شرح داده می شود.

لیم^۱ و همکارانش [۲۳] در سال ۲۰۰۸ جهت قطعه بندی دست^۲ با استفاده از دوربین های رنگی تلفیق دو روش قطعه بندی با استفاده از رنگ پوست^۳ و استفاده از اختلاف حرکت^۴ در فریم های متوالی تصویر را مورد ارزیابی قرار داده اند. در این تحقیق تلاش شده است که نشان داده شود تلفیق استفاده از اختلاف فریم حرکتی برای بدست آوردن مکان های محتمل برای وجود دست و

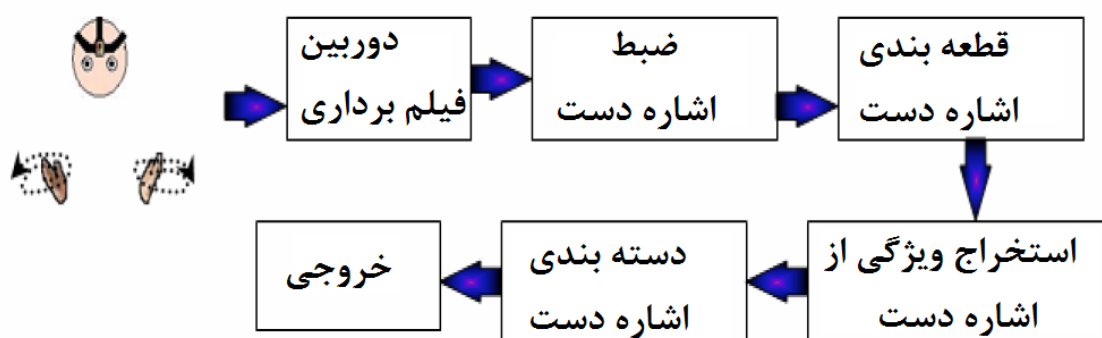
^۱ Lim Wei Howe

^۲ Hand Segmentation

^۳ Skin Segmentation

^۴ Motion Difference

سپس اعمال کردن قطعه بندی با استفاده از یک مقدار آستانه^۱ بر روی رنگ پوست، نتایج خوب و مقاومی^۲ برای دسته بندی اشارات دست به ما ارائه می دهد. شمای^۳ کلی کاری که در این تحقیق انجام شده است در شکل ۲-۸ آمده است [۲۳]. همانطور که در این شکل دیده می شود ورودی اصلی کار از یک دوربین فیلم برداری رنگی گرفته شده است و در نهایت نیز اشارات دست دسته بندی شده اند. نمونه ای از نتایج کار انجام شده در این تحقیق در شکل ۲-۹ آمده است. در قسمت بالای شکل تصویر دو فریم متوالی از یک حرکت دست نمایش داده شده است و در قسمت پایین شکل فوق نتیجه خروجی بدست آمده پس از اعمال الگوریتم بدست آوردن پیش زمینه از اختلاف فریم و قطعه بندی با استفاده از رنگ پوست نمایش داده شده است.



شکل ۲-۸-مراحل کلی انجام دسته بندی اشارات دست با استفاده از دوربین فیلم برداری رنگی در کار لیم و همکارانش. تصویر از [۲۳].

همانطور که مشاهده می شود، قطعه بندی به خوبی انجام گرفته است و پس زمینه پیچیده نیز مانع انجام این کار نشده است. البته از لباس پوشیده ای که فرد در تصویر پوشیده است مخصوصاً آستین متصل به مچ دست و یقه های بسته، محدودیت استفاده از آستانه گیری رنگی مشخص می-شود. این محدودیت گرچه کم اهمیت تر از دستکش ها می باشد ولی قابل نادیده گرفتن نیز نخواهد بود.

¹ Threshold

² Robust

³ Schema



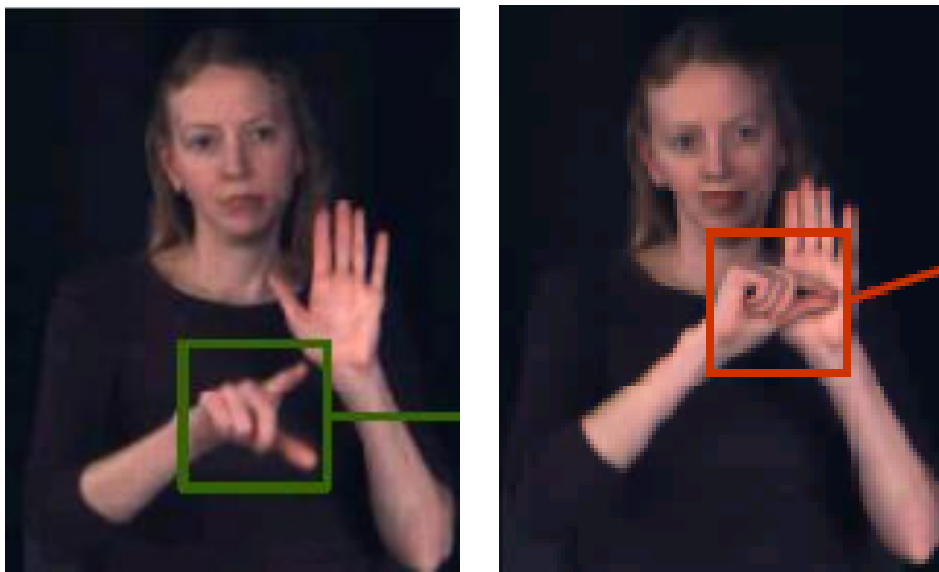
شکل ۲-۹- تصویر دو فریم متوالی از حرکات اشاره دست (تصاویر بالا) و نتیجه خروجی الگوریتم تلفیقی پیشنهادی (در پایین) در کار لیم و همکارانش. تصویر از [۲۳].

از دیگر کارهایی که اخیرا با استفاده از دوربین های رنگی انجام شده است، می توان به کار آشوین^۱ و همکارانش که در سال ۲۰۱۰ در دانشگاه باسِتن^۲ ایالات متحده انجام و ارائه شده است اشاره نمود. در این تحقیق نیز جهت انجام شناسایی زبان اشاره از ویدئو های دریافتی از یک دوربین رنگی استفاده شده است. در این کار شناسایی ۱۵۰۰ لغت زبان اشاره (که توسط ۳ نفر نیز اجرا شده

^۱ Ashwin Thangali

^۲ Boston University

اند) توسط یک مدل سازی بر اساس شبکه بیزین و بر پایه فریم ابتدایی و انتهایی اجرای هر کلمه ارائه شده است. در روش ارائه شده در این کار، جستجو بر اساس فریم ابتدایی و انتهایی هر کلمه انجام می شود. به عنوان مثال، در شکل ۲-۱۰، تصویر چپ مربوط به فریم ابتدایی یکی کلمه و تصویر سمت راست مربوط به فریم انتهایی همان کلمه می باشد که در نهایت با استفاده از دسته بند نزدیک ترین همسایه^۱ و جستجو در پایگاه داده دسته بندی خواهد شد. در روش پیشنهادی در این تحقیق، در بهترین حالت کارایی ۸۹.۶٪ برای پایگاه داده مورد نظر از کلمات گزارش شده است.



شکل ۲-۱۰- تصویر ابتدایی (سمت چپ) و تصویر انتهایی (سمت راست) یک کلمه جهت جستجو و دسته بندی کلمه مورد نظر در پایگاه داده در کار آشوب و همکارانش [۴۰]. تصویر از [۴۰].

۲-۲-۳- کارهای پیشین با استفاده از دوربین های استریو

از اولین نشانه های روی آوردن به عمق و تصاویر سه بعدی در کارهای پردازش تصویر و بینایی ماشین، می توان استفاده از دوربین های استریو^۲ در مرحله ضبط تصاویر را نام برد. این دوربین ها که به نوعی عملکرد انسان را در بینایی سه بعدی محیط شبیه سازی می کنند، از دو دوربین که به موازات در کنار هم قرار گرفته اند و هردو رو به جلو قرار گرفته اند تشکیل می شود که به دلیل

^۱ Nearest Neighbor Classifier

^۲ Stereo

اختلاف اندکی که در تصاویر بدست آمده از یکدیگر خواهند داشت، بسته به عمق اشیاء، تصاویری با اندکی جابجایی در نقاط به دست خواهند داد. این جابجایی با فاصله جسم از دوربین رابطه دارد و با محاسبه میزان اختلاف هر پیکسل در تصاویر می توان به نوعی به عمق این پیکسل ها در تصاویر پی برد. از آنجا که معمولا دست نزدیک تر از سر و بدن به دوربین می باشد، این ایده که بتوان از طریق عمق بدست آمده از این دوربین ها به شکل ساده تری عمل قطعه بندی دست را انجام داد شکل گرفت و در کارهایی که نیاز به قطعه بندی دست احساس می شد، به این روش نیز به عنوان روشی مناسب جهت این کار نگریده شد. در شکل ۱۱-۲ نمایی از یک دوربین فیلم برداری استریو دیده می-شود.



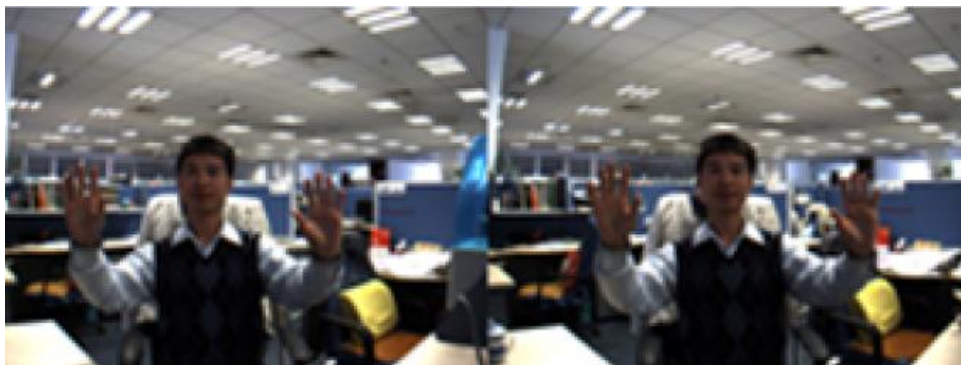
شکل ۱۱-۲- تصویری از یک دوربین فیلم برداری استریو.^۱

از جدید ترین کارهایی که توسط دوربین های استریو انجام شده است می توان به کار وانگ^۲ و همکارانش اشاره نمود[۲۶]. در این کار جهت قطعه بندی دست ها از یک دوربین استریو در چند مرحله این کار استفاده شده است. در اولین مرحله پس از استخراج فرد مورد نظر از پس زمینه، دست های وی شناسایی می شوند و سپس برای بهبود الگوریتم از فیلتر رنگ پوست برای جدا سازی نهایی تصویر استفاده می شود. در شکل ۱۲-۲ یک تصویر بدست آمده از دوربین استریو در کار فوق مشاهده می شود که تصویر دوربین سمت راست و تصویر بدست آمده از دوربین سمت چپ دوربین در کنار

^۱ تصویر از www.studiodaily.com

^۲ Qi Wang

هم مشاهده می شود. همانطور که مشخص است، دو تصویر از لحاظ زاویه دید اندکی با یکدیگر اختلاف دارند. در شکل ۱۳-۲ (چپ)، در بخش اول الگوریتم با شناسایی چهره، بدن فرد موجود در تصویر استریو شکل ۱۲-۲ استخراج شده است. در شکل ۱۳-۲ (وسط) مکان دست ها تخمین زده شده اند و در شکل ۱۳-۲ (راست) با استفاده از فیلتر کردن رنگ پوست قطعه بندی ها کامل شده است. در روش فوق ۸۳.۸٪ دقت در قطعه بندی دست گزارش شده است که نسبت به موارد مشابه ذکر شده درصد بسیار بهتری ارزیابی شده است. وابسته بودن روش بر قطعه بندی بر اساس رنگ پوست، استفاده از آستین های بلند را الزامی می کند که یک محدودیت برای سیستم ارزیابی می شود.



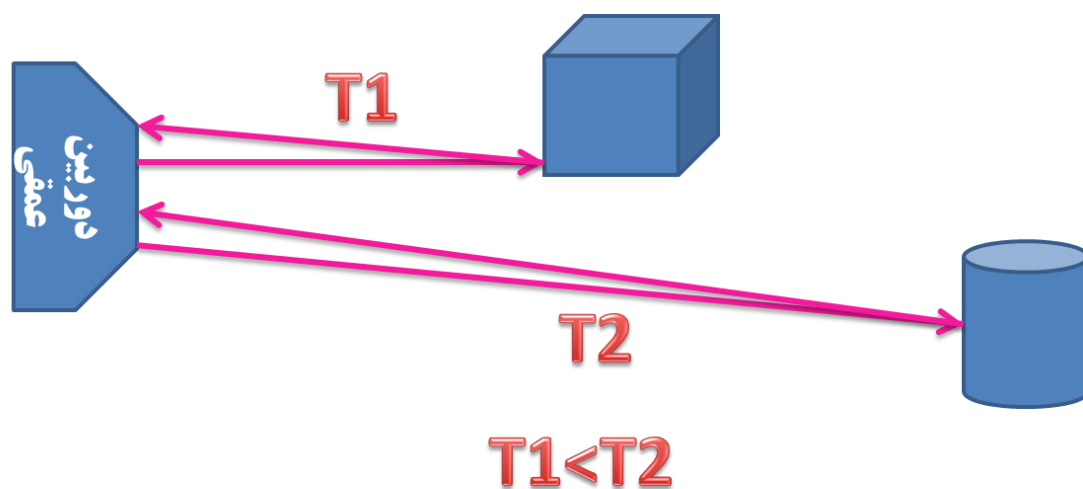
شکل ۱۲-۲- یک تصویر بدست آمده از دوربین استریو در کار وانگ و همکارانش. تصویر راست و چپ به ترتیب بدست آمده از لنز سمت راست و سمت چپ دوربین می باشد. تصویر از [۲۶]



شکل ۱۳-۲- مراحل مختلف الگوریتم قطعه بندی دست از دوربین های استریو توسط وانگ و همکارانش. تشخیص فرد (سمت چپ)، شناسایی مکان دست (وسط) و بدست آوردن پیکسل های دست (سمت راست). تصویر از [۲۶].

۲-۲-۴- کارهای پیشین با استفاده از دوربین های عمقی

دوربین های عمقی یا دوربین های زمان پرواز^۱ دوربین های هستند که با ارسال اشعه های مادون قرمز^۲ به محیط به صورت ماتریسی و محاسبه زمان برگشت هر یک از آن ها، فاصله اشیاء محیط را از دوربین محاسبه و ماتریسی دوبعدی از فواصل هر یک از نقاط موجود در جلوی دوربین ایجاد می کنند. این ماتریس ها به تصاویری از پیکسل های سطح خاکستری که طیفی از نزدیکترین پیکسل تا دورترین پیکسل را ایجاد می کنند تبدیل شده که به تصاویر عمقی معروف هستند. در شکل ۲-۱۴ نمایی از کارکرد این نوع دوربین ها را مشاهده می نمایید. در این شکل دو اشعه مادون قرمز به صورت نمادین در برخورد به دو شیء در محیط نشان داده شده است. زمان پرواز این دو اشعه با T_1 و T_2 مشخص شده است که به دلیل نزدیکی مکعب به دوربین، $T_1 < T_2$ می باشد و در نتیجه فاصله آن از استوانه متمایز می شود.



شکل ۲-۱۴- نمایی از نحوه کارکرد دوربین های عمقی در تصویربرداری از محیط.

به طور کلی ایده بدست آوردن فاصله هر نقطه از محیط توسط دوربین های عمقی توسط

رابطه (۲-۱) مشخص می شود:

¹ Time-of-Flight (ToF)

² Infra-Red

$$distance = c * T/2; \quad (1-2)$$

که در آن distance فاصله هر پیکسل از دوربین، c سرعت موج مورد نظر (در اینجا مادون قرمز) و T زمان پرواز اشعه از دوربین به شیء و برگشت دوباره آن به دوربین است. باید توجه داشت که فرمول دقیق بدست آوردن فاصله معمولا بسیار پیچیده تر و پارامترهای زیادی در آن ها دخیل است. به دلیل سرعت بالای این امواج، تنها با پیشرفت پردازنده ها در دهه اخیر امکان ساخت این نوع دوربین ها فراهم گردیده است و استفاده از آن در کاربرد های مختلف نیز تاکنون با موفقیت صورت گرفته است. در شکل ۲-۱۵ نمایی از یک دوربین عمقی مدل MESA SR 3000 نمایش داده شده است که از رده های معمول این نوع دوربین ها می باشد.

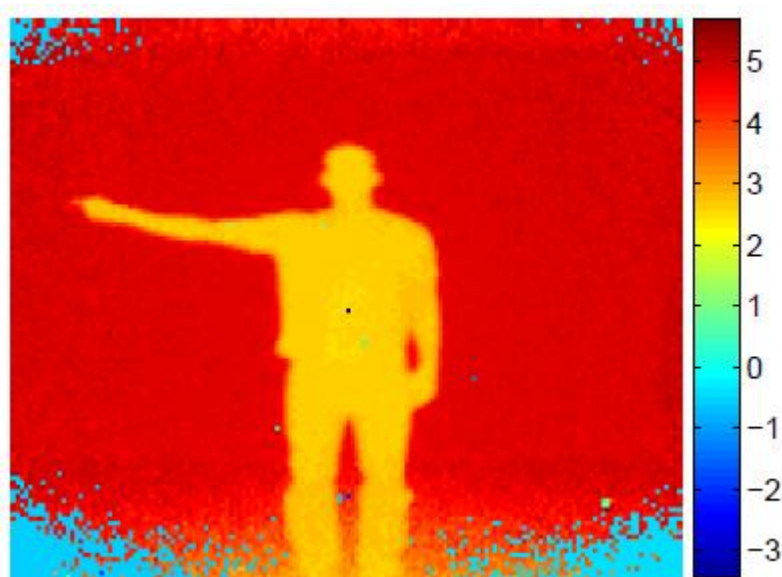


شکل ۲-۱۵-نمایی از یک دوربین عمقی زمان پرواز مدل MESA SR 3000. تصویر از [41].

از کارهایی که اخیرا (سال ۲۰۰۷) در راستای تشخیص اشاره با استفاده از دوربین های عمقی زمان پرواز انجام گرفته است، می توان به کار هولت^۱ و همکارانش اشاره کرد [۲۴]. در این کار با استفاده از یک دوربین زمان پرواز SwissRanger SR2 به بررسی چندین حرکت اشاره از زوایای مختلف ۰°، ۴۵° و ۴۵°- پرداخته اند. نتیجه نهایی با استفاده از تفریق فریم های متوالی عمقی

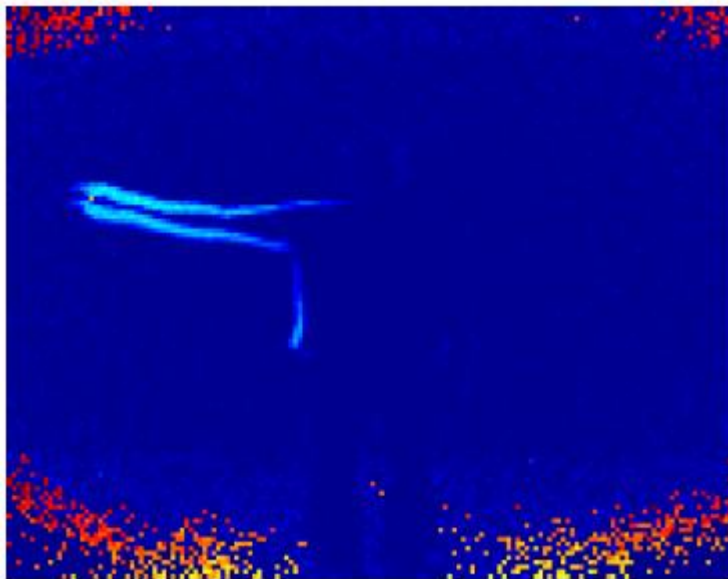
¹ M.B. Holte

بدست آمده است و درصد شناسایی ۹۳.۷۵٪ گزارش شده است. این درصد شناسایی بر روی ۳۸۴ نمونه آزمایشی که شامل ۴ حرکت بوده است گزارش شده است. این چهار حرکت عبارت اند از حرکت دست به راست^۱، بالا بردن دست^۲، دست زدن^۳ و دست تکان دادن^۴ که توسط ۱۰ فرد متفاوت انجام گرفته اند. در شکل ۱۶-۲ (الف) تصویری از یک فریم از داده های دریافتی از دوربین زمان پرواز، در شکل ۱۶-۲ (ب) تصویر اختلاف بدست آمده از فریم های متوالی از تصاویر عمقی و در شکل ۱۶-۲ (ج) حرکت تشخیص داده شده نشان داده شده است [۲۴]. درصد شناسایی ارائه شده در این کار توسط این نوع دوربین، قدرت استفاده از این دستگاه ها را نمایان می سازد. اگرچه به دلیل محدودیت تعداد اشاره ها در این سیستم نمی توان نتیجه گیری کلی از درصد شناسایی آن برای تشخیص تمامی زیرشاخه های زبان اشاره گرفت.

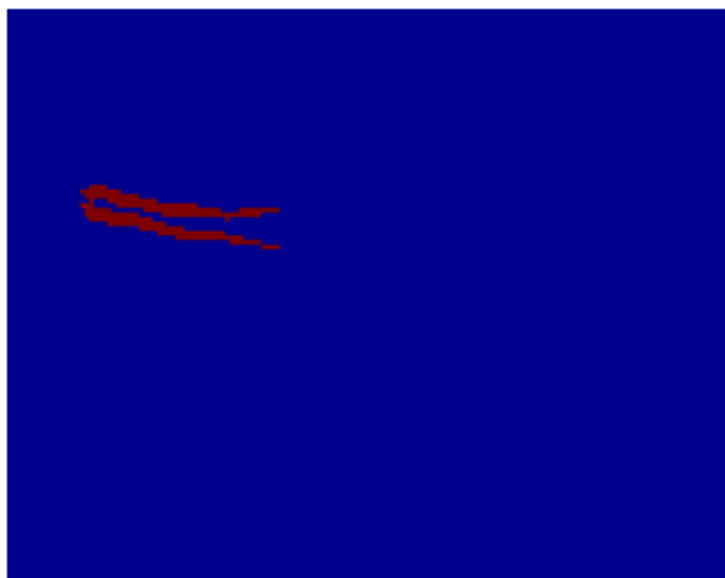


(الف)

-
- ¹ Point Right
 - ² Raise Hand
 - ³ Clap
 - ⁴ Wave



(ب)



(ج)

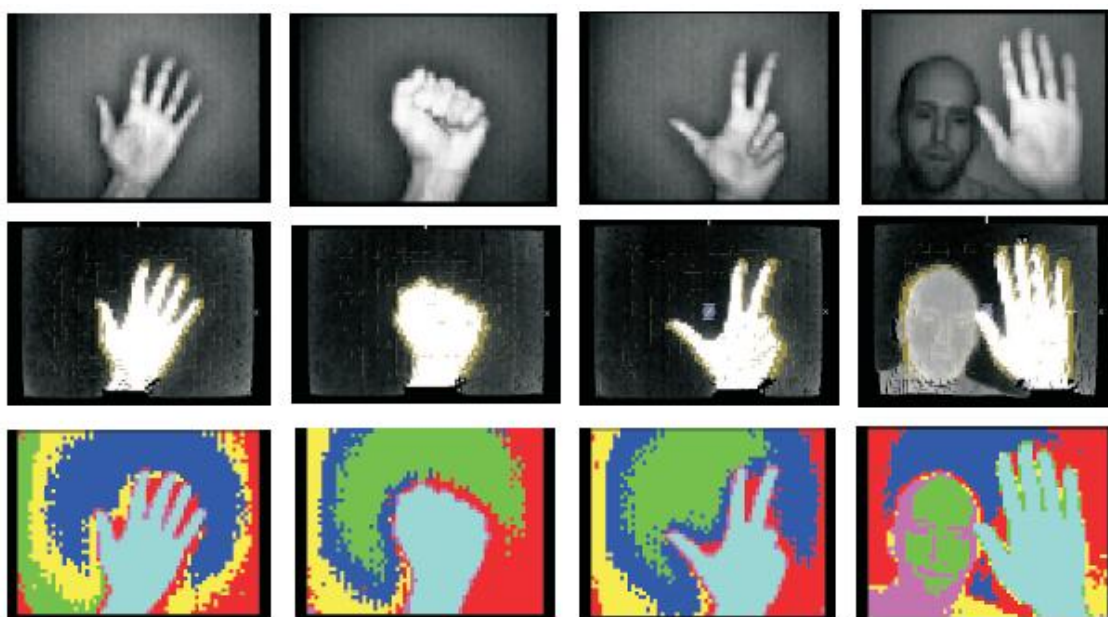
شکل ۲-۱۶- (الف) یک فریم از دوربین عمقی، (ب) اختلاف بدست آمده از تفریق فریم های متوالی و (ج) حرکت تشخیص داده شده از اختلاف فریم ها در کار هولت و همکارانش [۲۴]. تصویر از [۲۴].

دیگر تحقیقی که در همان سال ۲۰۰۷ و در آزمایشگاهی در کشور آلمان (آزمایشگاه ZESS)^۱ با استفاده از دوربین های عمقی زمان پرواز انجام گرفته است مربوط به کار قبادی^۲ و همکارانش در

^۱ Center for Sensor Systems

^۲ S.E. Ghobadi

مورد قطعه بندی دست با دوربین های عمقی و شدت روشنایی^۱ می باشد [۲۷]. در این آزمایش ها، قطعه بندی دست با استفاده از ترکیب دو روش K-Means و EM^۲ انجام گرفته است که به صورت بدون نظارت^۳ انجام گرفته است. نتایج قطعه بندی به گفته محققان رضایت بخش بوده است ولی با این حال در برخی از موارد پیچیده تر، قطعه بندی دست با بخش های دیگر بدن مانند سر و گردن نیز ادغام شده است. نمونه ای از نتایج را در شکل ۲-۱۷ می بینید [۲۷]. سطر اول تصاویر شدت روشنایی، سطر دوم تصاویر دوربین عمقی و سطر سوم تصاویر مربوط به خروجی قطعه بندی نمایش داده شده است. همانطور که مشخص است، دست با استفاده از رنگی متمایز از دیگر قسمت های تصویر جدا شده است.



شکل ۲-۱۷- تصاویر مربوط به قطعه بندی دست در کار قبادی و همکارانش [۲۷]. سطر اول مربوط به تصاویر شدت روشنایی، سطر دوم مربوط به تصاویر دوربین عمقی و سطر سوم نتایج قطعه بندی را نشان می دهد.

¹ Intensity

² Expectation Maximization

³ Unsupervised

در مقاله دیگری که در سال ۲۰۱۰ در رابطه با شناسایی زبان اشاره با استفاده از دوربین های عمقی انجام شده است می توان به کار ساتو^۱ و همکارانش در ژاپن اشاره کرد [۹]. در این کار قدرت دوربین های عمقی با دوربین های دوبعدی مقایسه شده و در ۱۲ علامت زبان اشاره و در کل ۳۰۰ نمونه آزمایشی این مقایسه انجام گرفته و نتیجه نهایی ۵۸.۳٪ صحت برای دوربین های دو بعدی و ۹۱.۳٪ برای دوربین های عمقی زمان پرواز گزارش شده است. روش دسته بندی استفاده شده مدل مخفی مارکوف^۲ و دوربین عمقی مورد استفاده دوربین MESA SR4000 با کیفیت تفکیک پذیری^۳ ۱۷۶ در ۱۴۴ پیکسل عمقی بوده است. نمونه ای از یکی از علامات زبان اشاره که در این مقاله اجرا شده است را در شکل ۲-۱۸ مشاهده می نمایید. در این شکل دو فریم ابتدایی و انتهایی یکی از حرکات نشان داده است. تاثیر استفاده از دوربین های عمقی در بهبود نتایج سیستم در این کار به خوبی مشخص شده است.



شکل ۲-۱۸- فریم ابتدایی و انتهایی یکی از علائم اشاره اجرا شده در کار ساتو و همکارانش [۹].

همانطور که با مطالعه در کارهای انجام شده در سال های اخیر می توان اشاره نمود، استفاده از دوربین های عمقی زمان پرواز از محبوبیت و دقت بالاتری نسب به دوربین های معمولی برخوردار هستند و نتایج بهتری با استفاده از این دوربین ها حاصل می شود.

¹ Arato Sato

² Hidden Markov Model (HMM)

³ Resolution

۲-۲-۵- کارهای پیشین با استفاده از کینکت

کینکت در سال ۲۰۰۹ توسط شرکت مایکروسافت با نام پروژه ناتال^۱ برای دستگاه بازی Xbox 360 معرفی شد. هدف اصلی این دستگاه ایجاد امکان انجام بازی های اکشن، ورزشی و ... بدون استفاده از دسته های بازی معمول^۲ بود. نهایتاً این دستگاه در نوامبر سال ۲۰۱۰ با نام سنسور کینکت [۴۲] با قیمت ۱۴۹ دلار وارد بازار گردید و با استقبال فراوانی مواجه و طولی نکشید که ۸ میلیون دستگاه از این وسیله فروخته شد.

کینکت از یک دوربین رنگی، یک دوربین عمقی زمان پرواز همتراز با دوربین رنگی، یک شتاب سنج سه بعدی، آرایه ای از میکروفون ها و یک موتور شیبی^۳ تشکیل شده است. نمایی از این وسیله را در شکل ۲-۱۹ مشاهده می نمایید.



شکل ۲-۱۹- نمایی از سنسور کینکت. تصویر از [۴۳]

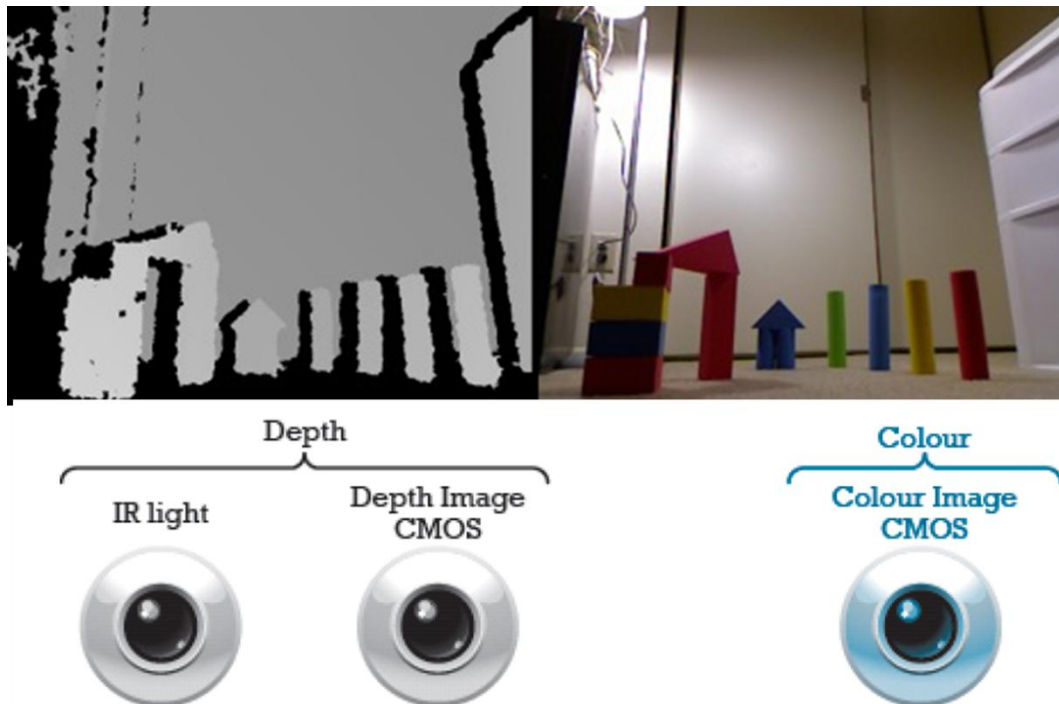
برخلاف دوربین های عمقی زمان پرواز که معمولاً وضوح تصویر عمقی حداکثر تا ۱۷۶ در ۱۴۴ پیکسل داشته اند، سنسور کینکت علاوه بر دوربین عمقی با وضوح تصویر ۶۴۰ در ۴۸۰ با سرعت ۳۰ فریم بر ثانیه، تصویر رنگی با همین وضوح و سرعت را نیز ارائه کرده است که به دلیل موازی بودن و قابلیت انطباق بر یکدیگر، قابلیت ویژه و منحصر به فردی تا زمان خود محسوب می-

^۱ Project Natal

^۲ Gamepad/Joystick

^۳ Tilt Motor

گردیده است. در شکل ۲-۲۰ نمایی از این انطباق به صورت دو تصویر رنگی و عمقی گرفته شده از یک صحنه با استفاده از کینکت به همراه چینش و انواع لنزهای استفاده شده در آن را مشاهده می-نمایید.



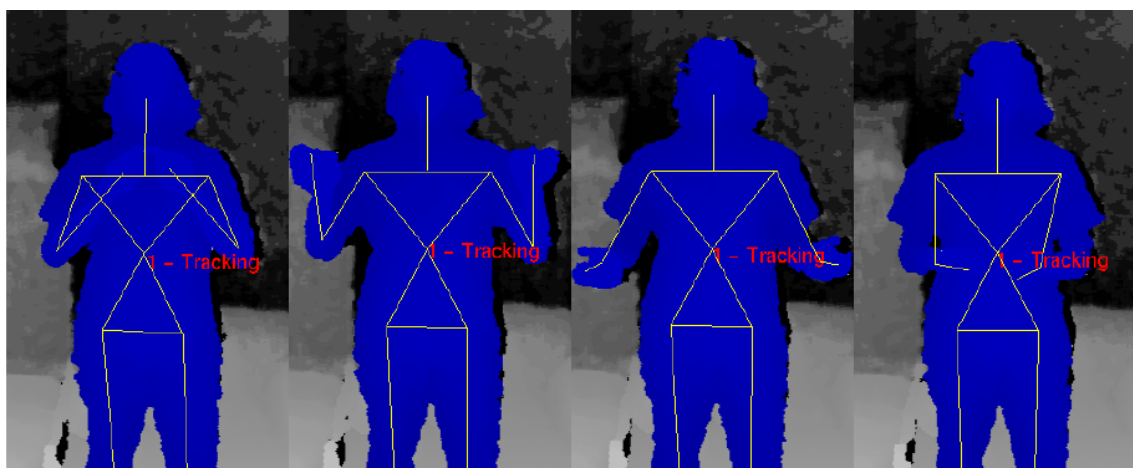
شکل ۲-۲۰-نمایی از دو تصویر رنگی و عمقی گرفته شده از یک صحنه با استفاده از کینکت به همراه چینش و انواع لنزهای استفاده شده در آن.

دستگاه کینکت عمق را از فاصله ۰.۸ متر تا ۳.۲ متر با دقت یک میلی متر شناسایی می کند. در ویرایش جدید این دستگاه که در سال ۲۰۱۲ و منحصر برای رایانه تولید شده است، امکان تصویر برداری در حالت نزدیک^۱ نیز ممکن شده است که در این صورت عمق از فاصله ۰.۴ متری نیز قابل شناسایی می باشد. پیکسل هایی که در فاصله مجاز قرار نگیرند، توسط بیت هایی خاص در تصاویر ورودی مشخص می شوند. این دستگاه همچنین قابلیت تشخیص ۲۰ نقطه از اسکلت بدن را دارا می-باشد و به طور همزمان تا ۶ بازیکن را پشتیبانی می نماید. با استفاده از این ۲۰ نقطه براحتی می توان نقطه قرارگیری دست را شناسایی نمود و عمل قطعه بندی را به صورت ساده تری پی گرفت. البته

¹ Near Mode

همیشه این شناسایی دقیق ممکن نیست که در این صورت توسط بیت هایی می توان صحت تشخیص هر یک از نقاط را توسط رابط نرم افزاری ارائه شده مشخص نمود. اخیراً در سال ۲۰۱۲ نیز سنسورهای شبیه به کینکت مانند دستگاه Xtion Pro Live^۱ شرکت Asus ارائه شده اند که در نتیجه به طور کلی به این نوع سنسور ها دوربین های RGB-D^۲ اطلاق می شود. در ادامه به کارهای پیشین با استفاده از کینکت پرداخته خواهد شد.

در یک پایان نامه کارشناسی که در پاییز ۲۰۱۱ به چاپ رسیده است، لنگ^۳ با استفاده از کینکت و قابلیت تشخیص اسکلت آن به شناسایی کلمات زبان اشاره با استفاده از مدل مخفی مارکوف پرداخته است [۴۴]. در این تحقیق نه کلمه مورد شناسایی قرار گرفت که هشت کلمه آن با دقت ۹۷.۷٪ شناسایی و کلمه دیگر شناسایی نگردید. در این کار با تاکید بر اهمیت کینکت و دوربین ها عمقی در دریافت تصاویر مادون قرمز یکسان در اکثر شرایط نوری تاکید شده است. در شکل ۲-۲۱ تصویری از توالی^۴ کلمه PAKET^۵ در تحقیق فوق نشان داده شده است [۴۴]. این کار مقاوم بودن کینکت در شرایط نوری مختلف و قدرت شناسایی بالا در استفاده از آن را نشان می دهد.



شکل ۲-۲۱- توالی کلمه PAKET در کار لنگ با استفاده از دوربین عمقی کینکت. [۴۴].

^۱ http://www.asus.com/Multimedia/Motion_Sensor/Xtion_PRO_LIVE/

^۲ Red-Green-Blue-Depth

^۳ Simon Lang

^۴ Sequence

^۵ برگردان از آلمانی به فارسی به معنای بسته.

کاری دیگری که با استفاده از کینکت در راستای شناسایی زبان اشاره انجام شده است، کار جادیش^۱ و همکارانش است که به بررسی استفاده از این ابزار در تشخیص انگشتان و کف دست پرداخته است [۲۲]. در این کار با استفاده از یک روش بدیع ابتدا دست توسط سنسور عمقی جدا شده و سپس با تخمین کف دست و تشکیل نمودار فاصله نقاط از آن، اقدام به شناسایی انگشتان در تصویر شده است. در شکل ۲-۲۲ نمایی از انگشتان و کف دست تشخیص داده شده به صورت بلادرنگ^۲ نشان داده شده است [۲۲].



شکل ۲-۲۲- نمایی از انگشتان و کف دست تشخیص داده شده در کار جادیش و همکارانش. تصویر از [۲۲].

در این قسمت به دو مورد از مهمترین کارهای اخیر در رابطه با زبان اشاره و دستگاه کینکت اشاره می کنیم. یکی از این کارها، تحقیق زفرولا^۳ و همکارانش است که در جدیدترین اقدام خود استفاده از دستگاه کینکت را برای بازی CopyCat در نظر گرفته اند [۱۳]. در این تحقیق که نتایج آن در سال ۲۰۱۱ در یک کنفرانس به چاپ رسیده است، استفاده از کینکت در برابر استفاده از

^۱ Jagdish L. Raheja

^۲ Real-time

^۳ Zahur Zafrulla

دستکش ها و سنسورهای شتاب سنج در شناسایی کلمات و جملات زبان اشاره مقایسه شده است. همچنین مقایسه بین ایستادن و نشستن نیز در شناسایی با استفاده از سنسور کینکت مورد ارزیابی قرار گرفته است و در کل ۱۰۰۰ نمونه برای سه حالت فوق ایجاد شده است. نتایج حاصله نشان داده است که در حالت ایستاده، کینکت با شناسایی ۷۶.۱۲٪ نسب به شناسایی سیستم دستکش و سنسور شتاب سنج سه بعدی که ۷۴.۸۲٪ بوده است برتری داشته است و این درحالی است که استفاده از این سنسور، نیاز به هیچگونه وسیله اضافی بر روی دست های افراد استفاده کننده ندارد. البته در حالت نشسته درصد شناسایی ۵۱.۵٪ حاصل شده است که به دلیل خطای بالای دستگاه نیز بوده است که با ارتقای سیستم تشخیص اسکلت کینکت در حالت نشسته در نسخه های جدید نرم افزاری، احتمال بهبود قابل توجه این درصد شناسایی نیز وجود دارد. در شکل ۲-۲۳ نمایی از حالت ایستاده و نشسته با استفاده از سنسور کینکت را مشاهده می نمایید [۱۳]. با توجه به مقاوم بودن این سنسور به نور محیط، عدم نیاز به دستکش و همچنین درصد شناسایی بالاتر از دستکش و سنسور های، این تحقیق برتری این سنسور را نسبت به دوربین های رنگی و دستکش ها و شتاب سنج ها نشان می دهد.



شکل ۲-۲۳- نمایی از حالت ایستاده (راست) و نشسته (چپ) در استفاده از سنسور کینکت در بازی CopyCat. تصویر از [۱۳].

کار مهم دیگری که با استفاده از کینکت انجام شده است، کار پوگولت^۱ و بودن^۲ است [۸]. در این کار که نتایج آن در سال ۲۰۱۱ منتشر شده است، ۲۴ حرف از ۲۶ حرف زبان اشاره به شکل هجی

^۱ Nicolas Pugeault

^۲ Richard Bowdan

انگشتی^۱ دسته بندی شده اند و پایگاه داده ای با بیش از ۴۸۰۰۰ نمونه که از ۴ نفر گرفته شده^۲ معرفی شده است. ویژگی های مهمی که در این پایگاه داده وجود دارد، پیچیده بودن پس زمینه، تغییر زوایای دست در تمامی جهات و تعداد نمونه های بالاست که در نوع خود کم نظیر می باشد. در این تحقیق پس از استخراج دست^۳ با استفاده از دوربین عمقی کینکت و مقدار آستانه ای به عمق ۲۰ سانتی متر از تصویر عمقی، بانکی از ۴۰ فیلتر گبور^۴ تهیه و بر تصاویر عمقی و رنگی اعمال شده است. در نهایت ویژگی های استخراج شده از خروجی های فیلترهای گبور بر تصاویر به یک جنگل تصادفی^۵ داده شده است تا عمل دسته بندی با استفاده از آن انجام شود.

با توجه به تناقضات پاسخ داده نشده ای که در این کار وجود دارد^۶، درصد دقیق خروجی کار به درستی مشخص نیست ولی خروجی نهایی که در گزارش بیان شده است، ۷۵٪ دقت میانگین^۷ با استفاده از هردو داده عمقی و دوربین رنگی و همچنین ۷۳٪ و ۶۹٪ دقت میانگین به ترتیب برای داده های تصاویر رنگی و تصاویر دوربین عمقی می باشد. اعلام این درصد ها از آنجا حائز اهمیت است که ماتریس درهم ریختگی که در مقاله آورده شده است دقت میانگین را ۵۴.۴۷٪ و میزان یادآوری میانگین^۸ را تنها ۵۲.۹۶٪ نشان می دهد. از طرفی در بخشی از گزارش تعداد ۱۰۰ نمونه برای هر حرف و هر نفر گزارش شده است (جمعا ۱۲۰۰۰ نمونه) و در جایی دیگر ۵۰۰ نمونه (جمعا ۴۸۰۰۰ نمونه). در نهایت درصد هایی برای پایگاه داده ارائه شده در اینترنت (با ۶۰۰۰۰ داده) گزارش شد که در آن درصد شناسایی برای حالت عمقی ۳۵٪، حالت تصویر رنگی ۴۷٪ و ترکیب هردو ۴۹٪ اعلام

^۱ Finger-spelled

^۲ پایگاه داده ذکر شده در این کار با پایگاه داده ارائه شده در وبسایت که ۶۰۰۰۰ داده دارد از لحاظ تعداد اجرا ها که در مقاله ۴ اجرا بوده و در وبسایت ۵ اجرا می باشد، متفاوت است. ولی در هر دو حالت ۴ نفر اجرا ها را انجام داده اند (یک نفر دو اجرا داشته است).

^۳ Hand Extraction

^۴ Gabor

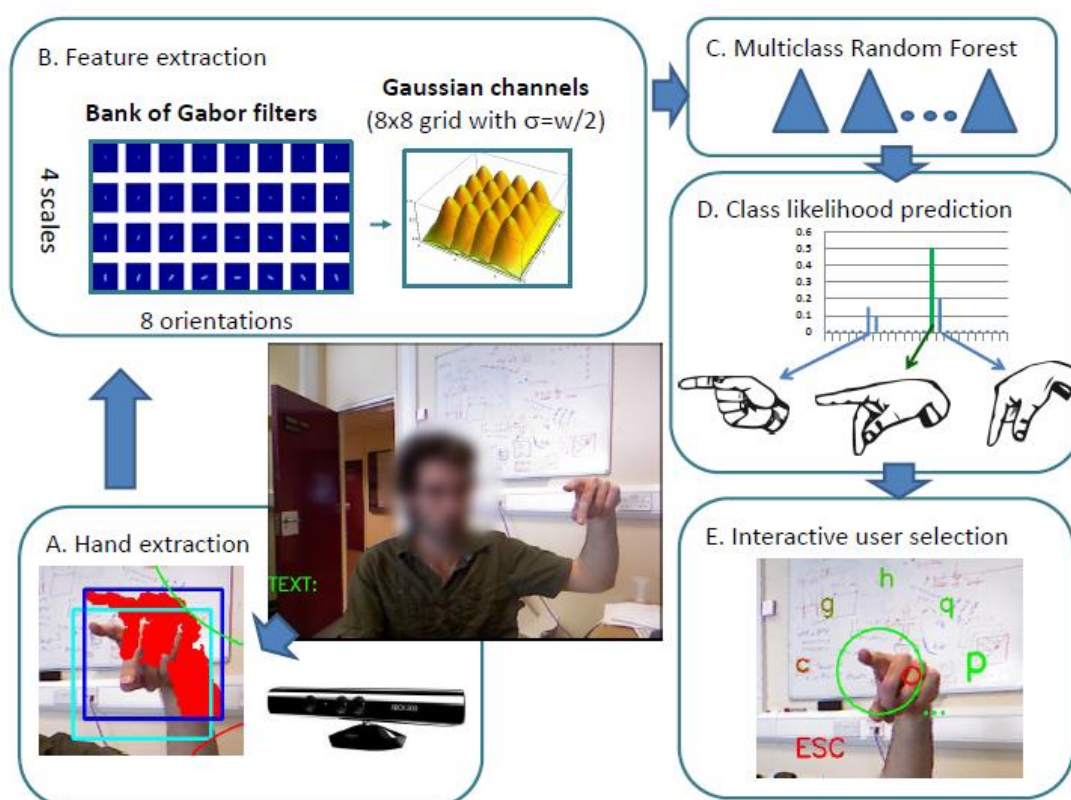
^۵ Random Forest

^۶ تناقضات چندین بار برای نویسندگان ارسال شد ولی پاسخی دریافت نشد و این درحالی بود که مکاتباتی با نویسندگان قبلا برقرار بوده است.

^۷ Mean Precision

^۸ Mean Recall

شد [۴۵]. شکل ۲-۲۴ تصویری از سیستم پیاده سازی شده در این کار نشان داده شده است [۸]. همانطور که مشخص است، پس از انجام مراحل مختلف شامل استخراج دست، استخراج ویژگی، استفاده از دسته بند جنگل تصادفی و تشخیص بر اساس بیشترین شباهت^۱، یک نرم افزار کاربری جهت تعامل با سیستم نیز طراحی شده است که در آن با پیشنهاد دادن حروف و کلمات و ارائه گزینه هایی مانند حذف آخرین حرف هجی شده، امکان استفاده کاربردی از روش پیشنهادی نیز ممکن شده است.



شکل ۲-۲۴-نمایی از سیستم پیشنهادی یوگالت و بودن برای تشخیص حروف الفبای زبان اشاره [۸]. تصویر از [۸].

¹ Likelihood

۲-۳- طبقه بندی کارهای پیشین بر حسب ویژگی های

استخراج شده

استخراج ویژگی یکی از بخش های بسیار مهم یک سیستم تشخیص زبان اشاره می باشد. از آنجا که در این زبان ها با سطوح متفاوتی از داده ها، شامل تصاویر دست، تصاویر صورت، حرکت بدن و دست در فضا، چرخش دست ها و توالی چندین یک از این حرکات برای ساختن کلمه یا جمله روبرو هستیم، اهمیت بخش استخراج ویژگی بیشتر نمایان می شود. سیستم هایی که بتوانند به خوبی ویژگی های مورد نظر خود را از بین سطوح داده ای فوق استخراج کنند، درواقع توانسته اند بخش مهمی از فرآیند شناسایی زبان اشاره را انجام دهند.

روش های مختلف و از دیدگاه های متفاوتی در مورد استخراج ویژگی ها وجود دارد. برخی از آن ها شامل ویژگی های هندسی^۱، تبدیل موجک^۲، ویژگی های برحسب لبه^۳ و کانتور^۴، ویژگی های اسکلت بدن و دست و سایر روش های ابداعی می شوند. در این بخش به بررسی کارهای پیشین بر حسب ویژگی هایی که از داده های ورودی استخراج کرده اند می پردازیم.

۲-۳-۱- کارهای پیشین با استفاده از ویژگی های هندسی

از آنجا که راه اصلی شناسایی زبان اشاره از ورودی های تصویری حاصل از دوربین ها می باشد و از طرفی تفاوت اصلی اشارات مختلف بیشتر اوقات در شکل ظاهری آن اشارات نمایان است، استخراج ویژگی هایی که بر حسب هندسه و شکل ظاهری اشارات باشند می تواند راه حل ساده و مناسبی برای این کار به شمار آیند. البته باید این نکته را نیز در نظر گرفت که به دلیل تفاوت در

¹ Geometric Features

² Wavelet Transform

³ Edge

⁴ Contour

شکل هندسی از زوایای مختلف، ممکن است استخراج صحیح این ویژگی ها و شناسایی صحیح آن ها بر اساس این ویژگی ها با مشکل روبرو شود. از این روست که برای افزایش پایداری سیستم ها، علاوه بر این روش ها از روش های غیر ظاهری نیز استفاده می کنند (مانند استفاده از شتاب سنج ها و ...). دو نوع از ویژگی های هندسی را در زیر بررسی می نماییم.

۲-۳-۱-۱- کارهای پیشین با استفاده از Moment ها

مهمترین نوع از ویژگی های هندسی، Moment ها هستند که با درجات مختلف بر روی تصویر سطح خاکستری^۱ و یا سیاه و سفید^۲ قابل اعمال هستند و در این پایان نامه در فصل استخراج ویژگی ها به تفصیل به آن ها پرداخته می شود. از کارهایی که قبلاً بررسی شده است و از ویژگی های هندسی به خصوص Moment ها در استخراج ویژگی استفاده نموده اند می توان به [۲، ۱۳، ۴۰] اشاره نمود.

۲-۳-۱-۲- کارهای پیشین با ویژگی هایی برحسب لبه و کانتور

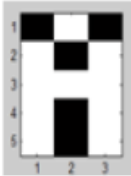
به طور ساده، لبه ها در یک تصویر نقاطی هستند در آن ها تغییرات روشنایی و یا رنگی تصویر شدید بوده و در نتیجه لبه های شکل یا اشیاء موجود در شکل را مشخص می کند. همچنین کانتور یک شیء در تصویر در واقع نقاطی از تصویر یا لبه های آن هستند که مرز بیرونی آن شیء را تشکیل می دهند. شکل ظاهری یک شکل، به خصوص اشاره های دست، به طور معمول قابل تشخیص کلی از روی کانتور آن شکل خواهد بود؛ به ویژه زمانی که بافت شیء اهمیت چندانی نداشته باشد.

^۱ Gray-Scale

^۲ Black & White

کارهای پیشین بسیاری از کانتور در کار خود استفاده نموده اند [۴۶, ۴۷]. از روش هایی که در استخراج ویژگی از کانتور استفاده نموده اند می تواند به کار مکالا^۱ و همکارانش در سال ۲۰۱۱ اشاره نمود [۶]. در این کار تعداد نامشخصی از تصاویر با پس زمینه مشکی که گفته شده شامل ۲۶ حرف الفبای زبان اشاره امریکایی می شوند، داده های اولیه تحقیق را تشکیل می دهند که با استفاده از ۵۵ ویژگی استخراج شده و شبکه عصبی مورد شناسایی قرار گرفته اند. رقم شناسایی ۱۰۰٪ گزارش شده است ولی هیچ اطلاعات دیگری در مورد داده های ورودی، انتخاب داده ها برای آموزش و تست و نحوه اعتبارسنجی مشخص نشده است. در بخشی از این کار بر روی داده های ورودی فیلترهای لبه جهت استخراج کانتور به کار رفته است. در شکل ۲-۲۵ تصویری از داده های ورودی، تصویر با حذف پس زمینه، تصویر کانتور بدست آمده و خروجی الگوریتم پیشنهادی در این تحقیق نشان داده شده است.

¹ Pryanka Mekala

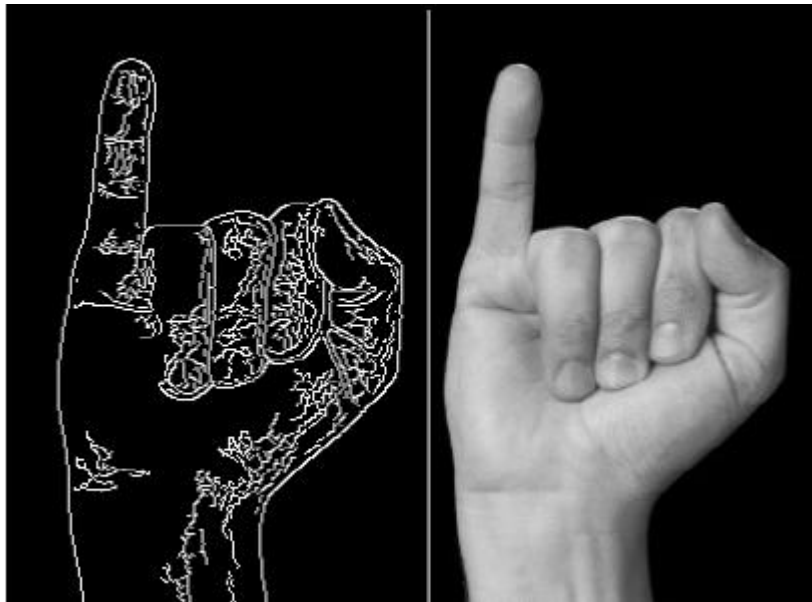
INPUT IMAGE	BACKGROUND SUBTRACTED IMAGE	EDGE OPERATOR	VIRTUAL LED DISPLAY IN MATLAB	ALPHABET DETECTED
				A
				G
				P
				Y

شکل ۲-۲۵- تصویر تعدادی از داده های ورودی، تصاویر پس از حذف پس زمینه، کانتور استخراج شده و نتیجه بدست آمده از الگوریتم در کار میکالا و همکارانش [۶]. تصویر از [۶].

کار دیگری نیز که از لبه ها در تشخیص زبان اشاره استفاده کرده است تحقیق جیتا^۱ و همکارانش [۷] در سال ۲۰۱۱ است. در این کار در مرحله پیش پردازش از الگوریتم لبه یابی Canny استفاده شده است و در نهایت با تخمین چندضلعی تشکیل دهنده دست ها، اقدام به شناسایی ۲۴ حرف الفبای زبان اشاره شده است. تعداد داده های ورودی مشخص نشده است ولی ادعا شده است که به غیر از حروف A و M بقیه حروف به درستی شناسایی شده اند. پس زمینه تصاویر مورد استفاده همگی ساده و به رنگ مشکی می باشد. در شکل ۲-۲۶ سمت راست، تصویر یکی از داده های اصلی مورد استفاده در شکل ۲-۲۶ سمت چپ و نتیجه اعمال الگوریتم لبه یاب Canny در مرحله پیش

¹ Geetha M.

پردازش نمایش داده شده است. ساده بودن پس زمینه و مشخص نبودن داده های ورودی و برخی تناقضات در تصاویر موجود از ضعف های اصلی این مقاله است.



شکل ۲-۲۶- یکی از تصاویر ورودی و خروجی لبه یاب Canny در کار جیتا و همکارانش [۷]. تصویر از [۷].

۲-۳-۲- کارهای پیشین با ویژگی هایی بر حسب تبدیل موجک

تبدیل موجک کاربردهای بسیاری در حوزه پردازش تصویر سیگنال دارد. در تبدیل موجک دو بعدی گسسته^۱ که در پردازش تصویر کاربرد دارد، یک موجک مادر^۲ انتخاب و بر روی سطر ها و ستون های تصویر (که اطلاعات هریک به عنوان یک سیگنال در نظر گرفته می شود) اعمال شده و به ازای هر سطح^۳ اعمال این تبدیل بر روی تصویر، چهار سری از ضرایب^۴ جدید حاصل می شود که هریک به اندازه یک چهارم اندازه تصویر اولیه است. نمایی از این چهار تصاویر خروجی در سطح اول از یک تبدیل موجک در شکل ۲-۲۷ آمده است. در تصویر بالا سمت چپ کلیات تصویر اولیه دیده می - شود که به آن تصویر تخمینی^۵ می گویند. از طرفی جزئیات^۶ تصویر نیز در قسمت پایین سمت چپ

¹ Discrete Wavelet Transform

² Mother Wavelet

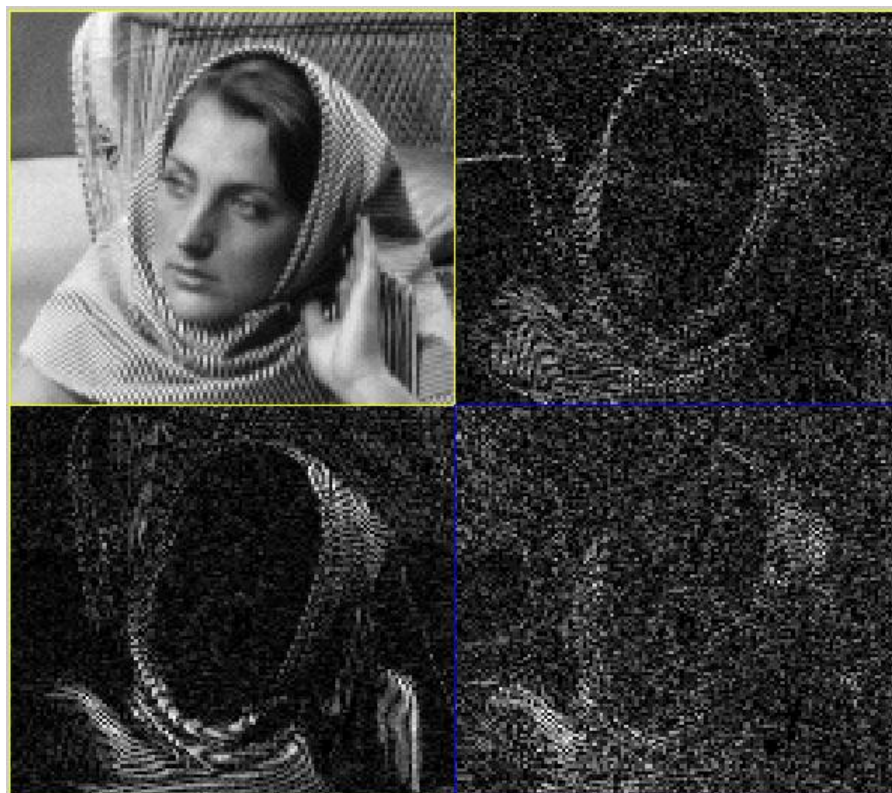
³ Level

⁴ Coefficient

⁵ Approximate

⁶ Details

دیده می شود که شامل نویز ها نیز می شود. ضرایب یا جزئیات دیگر افقی^۱ و عمودی^۲ نیز به ترتیب در دوتصویر بالا سمت راست و پایین سمت چپ مشاهده می شود که جهت انجام معکوس تبدیل موجک^۳ کاربرد دارند.



شکل ۲-۲۷- خروجی یک تبدیل موجک دوبعدی گسسته در سطح یک. جزئیات قطری تصویر در سمت پایین راست، جزئیات افقی و عمودی تصویر به ترتیب در سمت بالا راست و پایین چپ و کلیات تصویر در بالا سمت چپ تصویر دیده می شود.^۴

این تبدیل در سطوح مختلف بر حسب کاربرد می تواند بر روی تصویر تخمینی تکرار شود و تصویری کلی تر و یا جزئیاتی در سطوح پایین تر تولید نماید که در الگوریتم ها و روش های پیشنهادی مختلف کاربرد خاص خود را پیدا کرده اند. در شکل ۲-۲۸ نمایی از یک تبدیل موجک دوبعدی گسسته را تا سطح سوم مشاهده می نمایید. جزئیات موجود در تصویر می تواند در سطوح

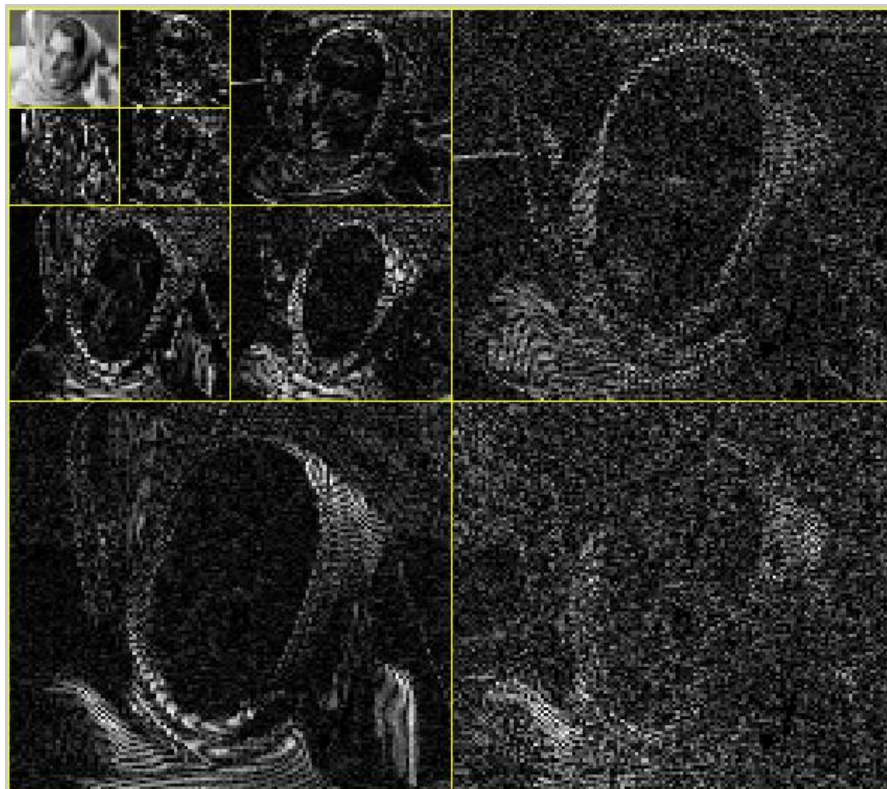
^۱ Horizontal Details

^۲ Vertical Details

^۳ Inverse Wavelet Transform

^۴ تصویر از نرم افزار Matlab 7.4.

مختلف مورد پردازش قرار گیرد و در نهایت بخش تخمینی آن در گوشه ی سمت چپ بالا می تواند جهت استخراج ویژگی مورد استفاده قرار گیرد. معمولاً با انجام پردازش هایی بر روی تصویر جزئیات، عمل هایی مانند حذف نویز را می توان بر روی تصویر اعمال نمود. از طرفی با استخراج ویژگی از بخش تخمینی تصویر نیز می توان به ویژگی هایی پایدار دست یافت که عاری از جزئیات هر تصویر و مربوط به کلیات شکل موجود در تصویر می شوند.



شکل ۲-۲۸- تبدیل موجک یک تصویر تا سطح سه. جزویات تصویر در سطوح مختلف و در نهایت یک تصویر تخمینی در شکل مشاهده می شود.^۱

در کار مکالا و همکارانش [۶] از تبدیل موجک برای استخراج ۴۰ ویژگی استفاده شده است. از دیگر کارهایی که در حوزه تشخیص زبان اشاره با استفاده از تبدیل موجک انجام شده است، کار ایزاکس^۲ و همکارانش [۱۰] می باشد که در سال ۲۰۰۴ ارائه داده اند. در این تحقیق جهت تشخیص ۲۴ حرف الفبای زبان اشاره آمریکایی، از تبدیل موجک جهت استخراج ویژگی استفاده و از شبکه

^۱ تصویر از نرم افزار Matlab 7.4.

^۲ Jason Issacs

عصبی نیز جهت دسته بندی حروف استفاده شده است. جزئیاتی در مورد تعداد داده های مورد نظر و کیفیت تصاویر داده نشده است ولی از نمونه نشان داده شده در شکل ۲-۲۹ به نظر می رسد که تصاویر دست با پس زمینه ساده تهیه شده اند. بدون ذکر کردن تعداد داده های مورد آزمایش و تعداد خطاها و یا ماتریس درهم ریختگی نرخ شناسایی ۹۹.۹٪ گزارش شده است. ساده بودن پس زمینه، مشخص نبودن تعداد داده و نحوه بدست آمدن درصد شناسایی گزارش شده از محدودیت های این مقاله می باشد.



شکل ۲-۲۹ - یکی از تصاویر اصلی مورد استفاده در کار ایزاکس و همکارانش [۱۰]. تصویر از [۱۰].

۲-۳-۳- کارهای پیشین بر حسب ویژگی های اسکلتی

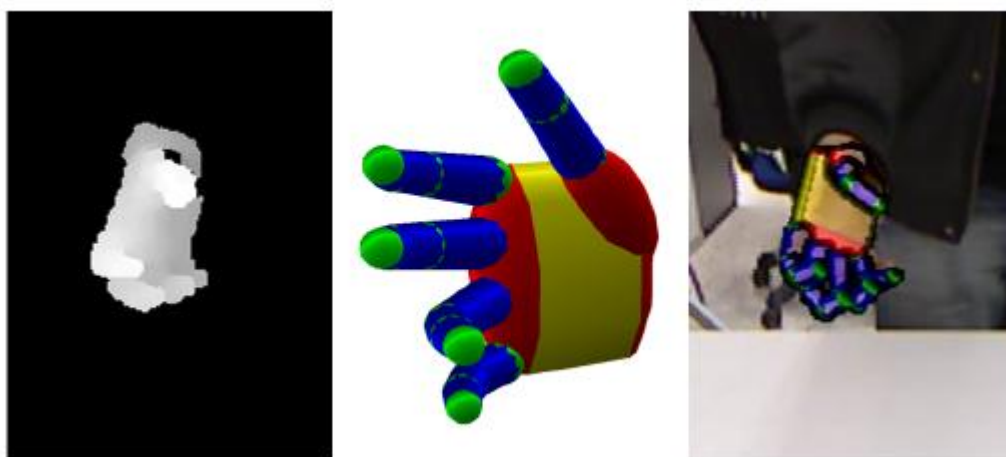
استفاده از اسکلت دست و بدن در شناسایی زبان اشاره می تواند بسیار کاربردی باشد و دلیل آن هم تغییرات شکل ظاهری دست و بدن و در نتیجه تغییرات محل قرارگیری و زاویه اسکلت آن ها می باشد. در کارها به دوشکل از اسکلت ها استفاده شده است: استخراج و استفاده از اسکلت دست ها و یا اسکلت کل بدن. در زیر به هریک از آن ها خواهیم پرداخت:

۲-۳-۳-۱- استفاده از اسکلت دست

اسکلت دست در اشاره های مختلف زبان اشاره بسیار تعیین کننده است ولی استخراج آن بسیار سخت می باشد. علت این سختی را می توان در کوچک و بسته بودن دست دانست و اینکه جلو و عقب شدن هر یک از انگشتان و یا مخفی شدن آن ها در بین یکدیگر شناسایی آن ها را از طریق

روش های ظاهری^۱ مشکل می سازد. استفاده از دوربین های عمقی نظیر کینکت، دستکش های رنگی یا مجهز به سنسور ها می تواند کمک بسزایی در تشخیص انگشتان و استفاده از اسکلت آن ها برای شناسایی زبان اشاره انجام دهد.

از کارهایی که اخیرا (سال ۲۰۱۱) در راستای استخراج اسکلت دست صورت گرفته است می-توان به [۴۸] اشاره نمود. در این کار از سنسور کینکت برای تطبیق دادن اسکلت دست با تعدادی از مدل های از پیش تعیین شده استفاده شده است و کاربرد آن توسط تمامی دوربین های عمقی و استریو پیش بینی شده است. از نتایج این تحقیق می توان در راستای استخراج ویژگی های مناسب جهت شناسایی زبان اشاره استفاده نمود. در شکل ۲-۳۰ تصویری از دست قطعه بندی شده (تصویر سمت چپ)، نزدیک ترین مدل به دست (تصویر وسط) و مدل قرارداده شده بر روی دست (تصویر سمت راست) را در کار [۴۸] مشاهده می کنید.



شکل ۲-۳۰- تصویری از دست قطعه بندی شده (تصویر سمت چپ)، نزدیک ترین مدل به دست (تصویر وسط) و مدل قرارداده شده بر روی دست (تصویر سمت راست) در کار [۴۸]. تصویر از [۴۸].

در تحقیق دیگری که توسط میستری^۲ در دانشگاه MIT انجام گرفته است [۴۹] و در آن دستگاهی به نام حس ششم^۳ ساخته شده است، از نوارهای رنگی که بر روی انگشتان قرار گرفته است

¹ Appearance Based

² Pranav Mistry

³ Sixth Sense

استفاده شده است تا مکان دقیق هریک از انگشتان در اسکلت دست و همچنین زاویه آن ها نسبت به یکدیگر تشخیص داده شود تا توسط دوربین و ویدئو پروژکتوری که بر روی بدن نصب می شود، اطلاعات مورد نظر در رابطه با دنیای اطراف در نزدیک ترین سطح خارجی در جلوی فرد نمایش داده شود. تصویری از نمونه اولیه این سیستم و نوارهای رنگی موجود بر روی انگشتان محقق آن، میستری، در شکل ۳۱-۲ آمده است. قابل توجه است که برای انگشتان هر دست از رنگ های متفاوتی استفاده شده است تا در صورت تداخل دستان، انگشتان هر دست به درستی از یکدیگر تفکیک شوند. محدودیت استفاده همگانی در پوشیدن بند های رنگی انگشتان از ضعف های این روش برای استفاده از اسکلت دست می باشد.



شکل ۳۱-۲- تصویری از نمونه اولیه دستگاه حس ششم و سازنده آن. نوارهای رنگی بر روی انگشتان برای تشخیص مکان دقیق انگشتان در اسکلت دست در تصویر مشخص است. تصویر از [۵۰].^۱

^۱ عکس از Sam Odgen.

۲-۳-۳-۲- استفاده از اسکلت بدن

در شناسایی زبان اشاره به خصوص کلمات و جمله های این زبان ها که با حرکت دست ها، مچ ها و بازوان همراه است دریافت اطلاعات اسکلت بالا تنه بدن می تواند برای سیستم های شناسایی زبان اشاره بسیار مفید باشد. از این رو در کارهای بسیاری از اطلاعات اسکلت بدن در شناسایی زبان اشاره بهره برده شده است که تعدادی از آن ها نیز توسط سنسور کینکت بوده است. کارهای [۱۲]- [۱۴] از این قبیل می باشند.

برای مثال در کار [۱۳] که از سنسور کینکت برای جمع آوری اطلاعات استفاده کرده است، از ۲۰ ویژگی اسکلتی در بردار ویژگی ها^۱ بهره برده شده است. در شکل ۲-۳۲ نمایی از اسکلت بدست آمده توسط کینکت و نقاط استفاده شده در این تحقیق مشخص شده است. بردار نقاط، زاویه بین نقاط و فاصله بین آن ها تشکیل دهنده این ۲۰ ویژگی بوده اند.



شکل ۲-۳۲- تصویری از اجرای یک اشاره در کار [۱۳] و اسکلت بدست آمده توسط سنسور کینکت. تصویر از [۱۳].

¹ Features Vector

۲-۳-۴- کارهای پیشین با سایر ویژگی ها

به غیر از ویژگی های دسته بندی شده در بخش های پیش، ویژگی های خاص دیگری نیز در برخی از کارها مورد استفاده قرار گرفته اند. برای مثال در کار [۸] از بانکی از فیلترهای گبور و در کار [۲۲] نیز از نمودار فاصله مرکز دست از نقاط اطراف جهت پیدا کردن انگشتان بهره برده شده است.

۲-۴- طبقه بندی بر حسب روش دسته بندی

یکی دیگر از بخش های مهم هر سیستم تشخیص و دسته بندی زبان اشاره، بخش دسته بندی و الگوریتم مربوط به آن می باشد. دسته بند های مختلفی در کارهای مختلف استفاده شده اند که کارهای مربوط به برخی از آن ها را در زیر بررسی نموده ایم:

۲-۴-۱- مدل مخفی مارکوف (HMM)

مدل مخفی مارکوف با ارائه مدلی از مسائل به صورت زنجیره ای از حالت ها که هر حالت می تواند با احتمالی خاص به حالتی دیگر برود و یا با احتمالی دیگر یک خروجی تولید کند، در بسیاری از مسائلی که دنباله ها در آن ها دخیل هستند کاربرد دارد. در مدل کردن حرکات زنجیره ای دست و اعضا بدن در هنگام اجرای یک کلمه در ویدئو ای از زبان اشاره و یا شناسایی جملات زبان اشاره معمولاً استفاده از این مدل می تواند بسیار کارساز باشد.

همانطور که گفته شد، استفاده از مدل مخفی مارکوف بیشتر در مواردی که کلمات و یا جملات در ویدئو زبان اشاره بایستی شناسایی شود مورد استفاده قرار می گیرد. از کارهایی که از این مدل برای شناسایی زبان اشاره استفاده نموده اند، می توان به [۲، ۴، ۱۲، ۱۳، ۵۱-۵۷] اشاره نمود. در تمامی این کارها، هدف شناسایی کلمات و جملات زبان اشاره بوده است.

۲-۴-۲- دسته بند بیز^۱

با استفاده از مدل های احتمالی مانند دسته بند بیز در کنار دید بهتری که به دسته بند می-توان داشت، می توان از قدرت بالای احتمالی آن نیز بهره برد و هم به سادگی امکان اعمال احتمالات پیشین^۲ را می توان فراهم نمود. از نمونه کارهای بررسی شده که از این دسته بند استفاده نموده است، می توان به [۴۰] اشاره نمود.

۲-۴-۳- شبکه های عصبی

استفاده از شبکه های عصبی جهت دسته بندی در کارهای بسیاری از افراد مخصوصاً در گذشته دیده می شود. از کارهایی که در زمینه شناسایی زبان اشاره با استفاده از شبکه های عصبی می توان به آن ها اشاره نمود، [۶، ۱۰، ۵۸-۶۲] می باشند. از مشکلات شبکه عصبی می توان به پارامترهای بسیار زیاد جهت بدست آوردن شبکه بهینه، آشکار نبودن داده های موجود در شبکه و به اصلاح جعبه سیاه^۳ بودن اطلاعات لایه های آن و طولانی بودن فرآیند یادگیری آن اشاره نمود.

۲-۴-۴- سایر دسته بند ها

سایر دسته بند ها نیز در کارهای مختلف مورد استفاده قرار گرفته اند. برای مثال [۸] از جنگل تصادفی، [۳۶] از بردار ماشین پشتیبان^۴ و [۱۴] از K-نزدیکترین همسایه^۵ استفاده کرده اند. هریک از این دسته بند ها مزایا و معایب خود را دارند. برای مثال قدرت جنگل تصادفی در یادگیری و آزمایش همزمان، بردار ماشین پشتیبان در قدرت تفکیک پذیری و نویز پذیری بالا و K-نزدیکترین همسایه در راحتی استفاده و پیاده سازی آن است. البته معایبی هم برای این روش ها در کاربردهای

^۱ Bayes Classifier

^۲ A Priori Probabilities

^۳ Black Box

^۴ Support Vector Machine

^۵ K-Nearest Neighbor

مختلف می توان نام برد و برای مثال سرعت پایین K-نزدیکترین همسایه در مواجهه با پایگاه داده های بسیار بزرگ با بردار ویژگی های وسیع را می توان از معایب آن دانست.

۲-۵- خلاصه و نتیجه گیری

در این فصل به دسته بندی و توضیح مفصل کارهای پیشین در حوزه های مختلف پرداخته شد. هریک از روش های پیش گرفته شده در کارهای گذشته دارای نقاط قوت و ضعفی بوده اند که بعضاً در کارها به آن ها اشاره شد. در ذیل به طور خلاصه به مقایسه روش ها و نتیجه گیری از مطالب گفته شده می پردازیم.

۲-۵-۱- بررسی ابزارهای موجود

نتیجه گیری بررسی کارهای پیشین در بخش ابزارهای مورد استفاده به طور خلاصه در جدول ۱-۲ آمده است. سنسور کینکت به دلیل مزایایی که ذکر شده است به عنوان انتخابی مناسب جهت شناسایی زبان اشاره به نظر می رسد و به همین دلیل به عنوان ابزار مناسب در این تحقیق مورد استفاده قرار خواهد گرفت.

جدول ۱-۲-جدول مقایسه ابزارهای مورد استفاده در تشخیص زبان اشاره

ردیف	نام ابزار	مزایا	معایب
۱	کینکت	- داشتن دوربین عمقی زمان پرواز - منطبق بودن دوربین عمقی و دوربین رنگی ۱. قطعه بندی بهتر ۲. حساس نبودن به پس زمینه	-نیازمند هزینه اضافی جهت خرید

		- قیمت مناسب - دارا بودن دیگر سنسور ها مانند شتاب سنج ۳ بعدی و آرایه ای از میکروفون ها	
۲	دوربین های زمان پرواز	- داشتن اطلاعات عمقی ۱. قطعه بندی بهتر ۲. عدم حساسیت به پس زمینه	- قیمت بسیار بالا - تفکیک پذیری پایین - در دسترس نبودن و عدم عمومیت
۳	دوربین های رنگی	- استفاده همگانی - قیمت ارزان	- مشکل در قطعه بندی بر اساس رنگ پوست
۴	دستکش ها و سنسور ها	- دقت بسیار بالا	- عدم دسترسی عمومی - ظاهر نامتعارف و عدم تمایل کاربران - تنوع مدل ها

۲-۵-۲- بررسی ویژگی های مورد استفاده

ویژگی های مختلف و از دیدگاه های متفاوتی در کاربرد های مختلف شناسایی زبان اشاره وجود دارد که در بخش قبلی بررسی شدند. در جدول ۲-۲ مقایسه ای بین مزایا و معایب هر یک از این ویژگی ها مشاهده می شود. به نظر می رسد که ویژگی های هندسی در زمانی که قطعه بندی نیز بر روی تصویر انجام شده است، هم از لحاظ پیچیدگی محاسباتی و هم از لحاظ مفهومی و مهمتر از همه کارایی، بهتر از سایر ویژگی ها، به خصوص در کارهای بلادرنگ، باشد.

جدول ۲-۲- مقایسه ویژگی های مختلف مورد استفاده در تشخیص زبان اشاره

ردیف	نام ویژگی	مزایا	معایب	کاربرد
۱	ویژگی های هندسی	- پیاده سازی ساده - سرعت بالا - عدم حساسیت به اندازه و سایر اعوجاجات	- احتمال عدم استقلال ویژگی ها در تعداد بالای ویژگی های استخراج شده	- داده های قطعه بندی شده - هجی با انگشتان
۲	تبدیل موجک	- تخمین خوب از تصاویر	- پیچیدگی در پیاده سازی - پیچیدگی محاسباتی بالا (سرعت پایین)	- داده های استاتیک و تصاویر
۳	ویژگی های اسکلتی	- تخمین دقیق حرکات اعضای بدن	- نیاز به دوربین های عمقی	- استفاده در ویدئو - شناخت کلمات و جملات

۲-۵-۳- بررسی روش های دسته بندی

روش های دسته بندی مورد استفاده در کارهای مختلف نیز در بخش پیش بررسی شدند.

مزیت ها و معایب هر یک را در جدول ۲-۳ مشاهده می نمایید. استفاده از مدل های احتمالی به دلیل

واضح بودن دسته بند، نیاز به تنها یکبار آموزش داده ها و سرعت بالا به عنوان انتخابی مناسب برای

تحقیق فعلی در نظر گرفته شده اند.

جدول ۳-۲- مقایسه روش های دسته بندی مختلف در تشخیص زبان اشاره

ردیف	نام دسته بند	مزایا	معایب	کاربرد
۱	بیز	سادگی در پیاده سازی- دارای مدل احتمالی- آموزش و آزمایش سریع- پارامترهای محدود- آموزش و آزمایش همزمان- قابلیت دخالت احتمالات پیشین	-	ویژگی های جداکننده- پایگاه داده های بزرگ
۲	شبکه عصبی	قدرت تخمین داده های غیرخطی	بسته بودن سیستم- پارامترهای زیاد- سرعت آموزش پایین	داده های غیرقابل تحلیل
۳	جنگل تصادفی	دارای مدل احتمالی- آموزش و آزمایش همزمان	پارامتر های بسیار زیاد- سرعت آموزش پایین	پایگاه داده های متوسط
۴	مدل مخفی مارکوف	سادگی مدل- دارای مدل احتمالی- حل سه نوع مسئله متفاوت- سرعت بالا	-	شناسایی زنجیره های پیوسته- شناسایی کلمات و جملات

۲-۵-۴- نتیجه گیری کلی

با توجه به بررسی های انجام گرفته در این فصل و مقایسه ابزارهای موجود، ویژگی ها و روش های مختلف دسته بندی، به نظر می رسد که بهترین ترکیب از یک روش شناسایی زبان اشاره به شرح زیر باشد.

بهترین ابزار مورد استفاده در تشخیص زبان اشاره، سنسور کینکت یا ابزار های مشابه می باشد که علاوه بر ارائه سیستم رابط کاربری طبیعی^۱ بدون نیاز به پوشیدن ابزارهای اضافی، از لحاظ توسعه نرم افزاری نیز با ارائه تصویر عمقی و رنگی موازی از محیط شرایط لازم برای قطعه بندی مناسب دست را فراهم می نماید.

مناسب ترین ویژگی های مورد استفاده برای دست قطعه بندی شده، ویژگی های هندسی می باشد که علاوه بر سرعت بالا و سادگی، ویژگی های قدرتمندی برای استخراج از دست محسوب می شوند. در حالتی که قطعه بندی مناسب ممکن است صورت نگیرد، استفاده از تبدیلات مختلف مانند تبدیل موجک و یا پاسخ فیلترهای مختلف از جمله فیلتر گبور مناسب می باشند و در حالتی که شناسایی کلمات و جملات زبان اشاره در ویدئو بایستی انجام گیرد، استفاده از ویژگی های اسکلتی حاصل از کینکت می تواند بسیار موثر باشد.

بهترین دسته بندها برای تشخیص زبان اشاره روش های احتمالاتی به ویژه دسته بند بیز می باشد که علاوه بر سادگی در پیاده سازی، قدرت استفاده از احتمالات پیشین در حالتی که اعمال دستور زبان نیز می تواند در دسته بندی موثر باشد را فراهم می کنند. علاوه بر این، دسته بندی با این روش ها، برای استفاده از پایگاه داده های بزرگ نیز به جهت نیاز به یک بار دسترسی به داده های آموزشی، بسیار مناسب می باشند و سرعت بالای آزمایش را تضمین می نمایند.

¹ Natural User Interface (NUI)

فصل ۳: روش پیشنهادی

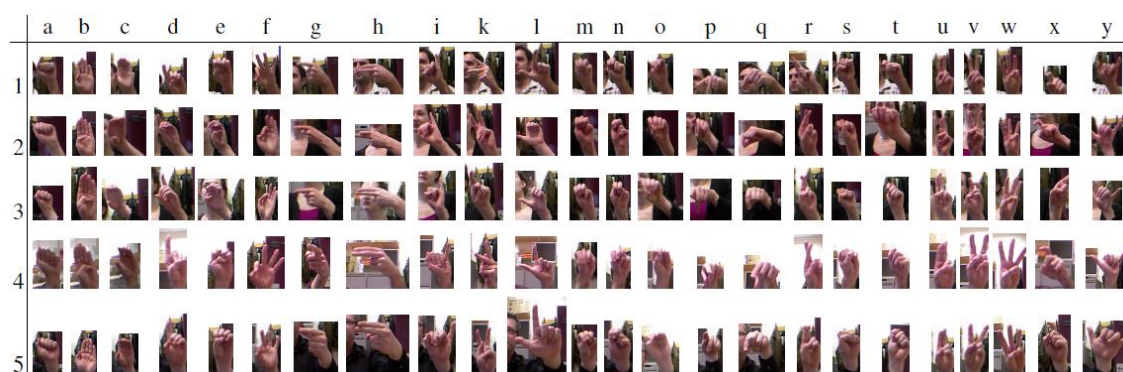
با توجه به نتیجه گیری های انجام شده در بخش بررسی کارهای پیشین، جایگاه روش پیشنهادی نیز با استفاده از بهترین ترکیب از روش های موجود، یعنی استفاده از سنسور کینکت به عنوان ابزار، استفاده از ویژگی های هندسی به عنوان ویژگی های اصلی مورد استفاده بر روی داده های قطعه بندی شده و استفاده از دسته بند بیز به عنوان روش دسته بندی داده ها پیشنهاد می شود. در این فصل به بررسی و توضیح روش پیشنهادی بر مبنای قطعه بندی عمقی دست با کینکت می پردازیم.

۳-۱- پایگاه داده

به منظور اینکه کار انجام شده در این تحقیق قابل مقایسه با کارهای دیگران در سطح جهانی باشد و همچنین جهت اطمینان از اثر نتایج تحقیق انجام شده در پیش برد لبه علم، روشی که در ادامه به توضیح آن خواهیم پرداخت بر اساس یک پایگاه داده استاندارد ارائه شده در سطح جهانی تست خواهد شد.

پایگاه داده مورد استفاده در این سیستم، پایگاه داده ای شامل حدود ۶۰۰۰۰ نمونه از ۲۴ حرف الفبای زبان اشاره امریکایی (به جز حروف J و Z که دارای حرکت می باشند)، ضبط شده توسط سنسور کینکت و در ۵ اجرا از ۴ فرد متفاوت می باشد که در ابتدا در [۸] معرفی شد و با اندکی تغییرات در [۴۵] در دسترس قرار گرفت. پایگاه داده ارائه شده در [۸] شامل ۴۸۰۰۰ داده از ۴ نفر معرفی شده است که به ازای هر نفر و هر حرف ۵۰۰ نمونه دارد. پایگاه داده [۴۵] در واقع کامل شده نمونه قبلی ضبط شده در ۵ اجرا از ۴ نفر قبلی است که با مشکلاتی نیز همراه است. اولین مشکل پایگاه داده ارائه شده این است که تعداد پوشه های افراد (که ۵ پوشه می باشد) در واقع از ۴ نفر ضبط شده است. این بدین معنی است که دو مجموعه از داده ها از یکی از افراد ضبط شده است. نکته دیگر در این پایگاه داده این است که تعداد نمونه های ضبط شده در تمامی فایل ها و افراد یکسان نیست و

بعضی از آنها کمتر از ۵۰۰ نمونه به ازای هر حرف و نفر دارد. برای مثال یکی از حروف دارای ۴۵۰ نمونه است. به همین دلیل، امکان انتخاب ۵۰۰ نمونه به صورت استاندارد از همه داده ها وجود ندارد. به دلیل اینکه مشخص نیست که در کار ارائه شده در [۸] از کدام ۴ پوشه از داده ها استفاده شده است و اینکه چگونه با مشکل کمبود نمونه ها برخورد شده است، برای پایگاه داده مورد استفاده در این سیستم، از تمامی ۵ اجرا افراد پایگاه داده [۴۵] و ۴۵۰ نمونه به ازای هر نفر و هر حرف که در نهایت پایگاه داده ی ۵۴۰۰۰ نمونه ای را تولید می کند استفاده شده است. در شکل ۱-۳ نمایی از هر حرف به ازای افراد مختلف پایگاه داده دیده می شود.



شکل ۱-۳- نمایی از ۲۴ حرف موجود در پایگاه داده به تفکیک ۵ اجرا توسط ۴ فرد مختلف. تصویر از [۴۵].

۳-۱-۱- نمونه های پایگاه داده

نمونه های موجود در پایگاه داده مورد استفاده در این تحقیق در نوع خود پیچیده و منحصر به فرد هستند. در هنگام ضبط داده ها، از کاربران خواسته شده است که دست خود را در فضا در جهات مختلف حرکت داده، چرخانده و انگشتان خود را نیز اندکی حرکت دهند. این کار پایگاه داده را با تعداد بسیار زیادی نمونه متفاوت با پس زمینه های کاملاً پیچیده شکل داده است که تا کنون در پایگاه داده های قبلی سابقه نداشته است. برای مثال در شکل ۲-۳ هشت نمونه از حرف P نشان داده شده توسط تنها یک نفر از افراد نشان داده شده است که تفاوت درون کلاسی^۱ در این پایگاه داده را

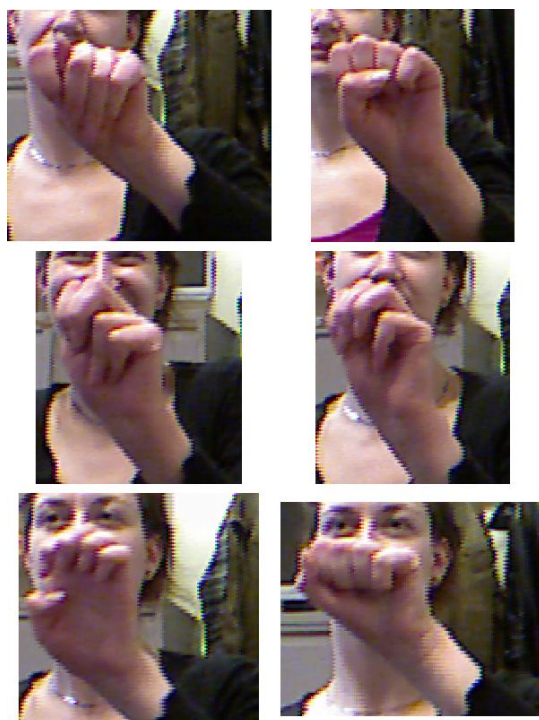
¹ Inner-class

نشان می دهد. همچنین در شکل ۳-۳ شباهت بین کلاسی^۱ چند نمونه نشان داده شده است. در این شکل هر یک از تصاویر از حروف الفبای متفاوتی هستند که از لحاظ ظاهری بسیار به یکدیگر شباهت دارند و کار دسته بندی را بسیار مشکل می کند. همچنین پیچیده بودن پس زمینه و قرارگرفتن دست در کنار صورت یا گردن که از لحاظ رنگ پوست بسیار به یکدیگر شباهت دارند نیز در شکل ۴-۳ نمایش داده شده است. نزدیک بودن رنگ پس زمینه به رنگ پوست نیز در برخی از تصاویر این پایگاه داده دیده می شود. نزدیک بودن شرایط محیط به شرایط طبیعی و پیچیدگی های گفته شده در بالا استفاده از آستانه گیری رنگ پوست جهت قطعه بندی دست را برای این پایگاه داده سخت نموده است و روش های دیگری بایستی برای قطعه بندی دست در این پایگاه داده اتخاذ گردد.

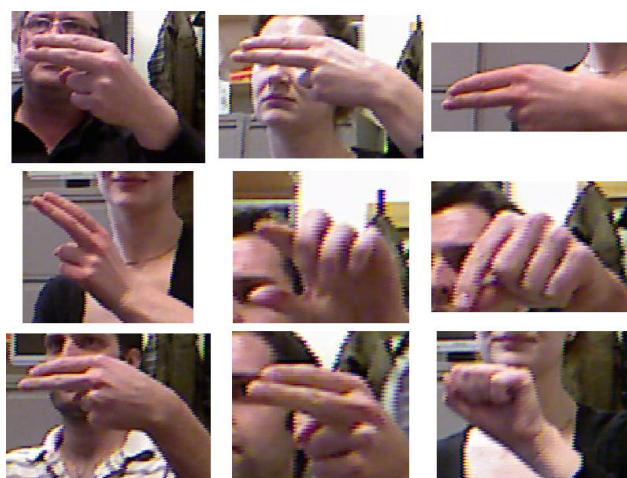


شکل ۳-۲- هشت تصویر از حرف P اجرا شده توسط یکی از افراد، نشان دهنده تفاوت درون کلاسی نمونه های پایگاه داده.

¹ Between-class



شکل ۳-۳- شباهت بین کلاسی در شش کلاس مختلف از نمونه های اجرا شده توسط یکی از افراد.



شکل ۳-۴- کنارهم قرار گرفتن دست، صورت و گردن در تصاویری از پایگاه داده که استفاده از آستانه گیری برای قطعه بندی را مشکل می سازد.

هر نمونه در پایگاه داده دارای دو فایل مجزاست: تصویر رنگی RGB و تصویر عمقی تصویر رنگی مورد نظر که هر دو توسط سنسور کینکت ضبط شده اند. پس از دریافت تصاویر از سنسور کینکت، مختصات دست توسط راه انداز^۱ کینکت OpenNI [۶۳] و چارچوب کاری^۲ NITE شناسایی

^۱ Driver

^۲ Framework

شده است و برای برش^۱ تصاویر مورد استفاده قرار گرفته است. سپس با فرض اینکه معمولاً دست تا عمق ۲۰ سانتیمتر از فضا قرار گرفته است، تصاویر در اطراف دست برش خورده اند. تمامی تصاویر پایگاه داده به صورت برش داده شده ارائه شده اند و دسترسی به تصاویر اصلی ممکن نیست. برخی از تصاویر نیز به نظر می رسد که عمقی کمتر و یا بیشتر از ۲۰ سانتی متر داشته باشند. اگر دسترسی به تصاویر اصلی نیز ممکن بود، امکان کار بیشتر بر روی قطعه بندی تصاویر نیز ممکن بود.

تصاویر عمقی به صورت پیکسل هایی ۱۶ بیتی unsigned short هستند که عمق را در فاصله حدود ۸۰۰ میلی متر (۰.۸ متر) تا ۳۲۰۰ میلی متر (۳.۲ متر) به ازای هر پیکسل نشان می دهند. البته در پایگاه داده ارائه شده، برخی از پیکسل ها مقادیری کمتر از ۸۰۰ میلی متر تا حدود ۷۰۰ میلی متر را نیز دارا هستند.

۳-۲-بستر پیاده سازی

پیاده سازی اصلی روش پیشنهادی با استفاده از زبان C#.Net و محیط ویژوال استودیو^۲ و در چارچوب Net Framework 4.0 انجام گرفته است. در این پیاده سازی برخی از عناصر مربوط به کینکت نیز که بایستی تحت WPF^۳ مورد استفاده قرار می گرفتند نیز وجود داشت که به این دلیل برخی از ساختار های نرم افزار پیاده شده نیز تحت این تکنولوژی پیاده سازی شده اند. علت استفاده از این زبان، قابلیت اتصال بسیار آسان به دستگاه کینکت و دریافت مستقیم اطلاعات تصویری از آن، قابلیت توسعه، اتصال و ارسال خروجی های سیستم به نرم افزار ها و دستگاه های مورد نیاز و همچنین سرعت بالای این محیط نسب به محیط هایی که بیشتر جهت شبیه سازی مورد استفاده قرار می گیرند (مانند MATLAB) می باشد.

^۱ Crop

^۲ Microsoft Visual Studio

^۳ Windows Presentation Foundation

در راستای استفاده از کتابخانه های الگوریتم ها و روش های مورد نظر نیز، از کتابخانه های Emgu CV^۱ که در واقع یک Wrapper مخصوص بستر Net. برای کتابخانه استاندارد OpenCV^۲ می باشد و همچنین ALGLIB^۳ که کتابخانه ای شامل الگوریتم های پیاده سازی شده ی تحلیل عددی و پردازش داده ها می باشد، استفاده شده است.

سیستم پیاده سازی شده از دو بخش تشکیل شده است: بخش آموزش و تست و بخش عیب یابی. در بخش آموزش و تست، پارامترهای مورد نظر به سیستم داده می شود و پس از تعیین پارامترهای مورد نظر مربوط به انتخاب داده ها، پیش پردازش، استخراج ویژگی ها، انتخاب ویژگی، دسته بندی، اعتبارسنجی و ... سیستم مرحله آموزش و تست را آغاز نموده و پس از اتمام مراحل کار گزارش های لازم را اعم از نتایج دسته بندی و ماتریس های درهم ریختگی^۴ را تولید می نماید. لازم به ذکر است سیستم به نحوی پیاده سازی شده است که جهت افزایش موازی سازی، از جدیدترین پردازنده های چند هسته ای^۵ و چند نخه^۶ سخت افزاری نظیر Core i7 و Core i5 پشتیبانی می نماید و کارها را در هشت نخ^۷ موازی تقسیم کرده (از ۲۴ حرف الفبا، هر سه حرف به یک نخ داده می شوند) و علاوه بر نمایش پیشرفت هر نخ سخت افزاری، باعث افزایش سرعت اتمام کار حداکثر تا هشت برابر سرعت معمولی (در پردازنده های Core i7 با چهار هسته سخت افزاری که هریک دو نخ را به صورت سخت افزاری پشتیبانی می کنند) می شود^۸. نمایی از بخش آموزش و تست نرم افزار را در شکل ۳-۵

^۱ <http://www.emgu.com/>

^۲ Intel's Open Source Computer Vision Library (<http://www.opencv.org/>)

^۳ <http://www.alglib.net/>

^۴ Confusion Matrix

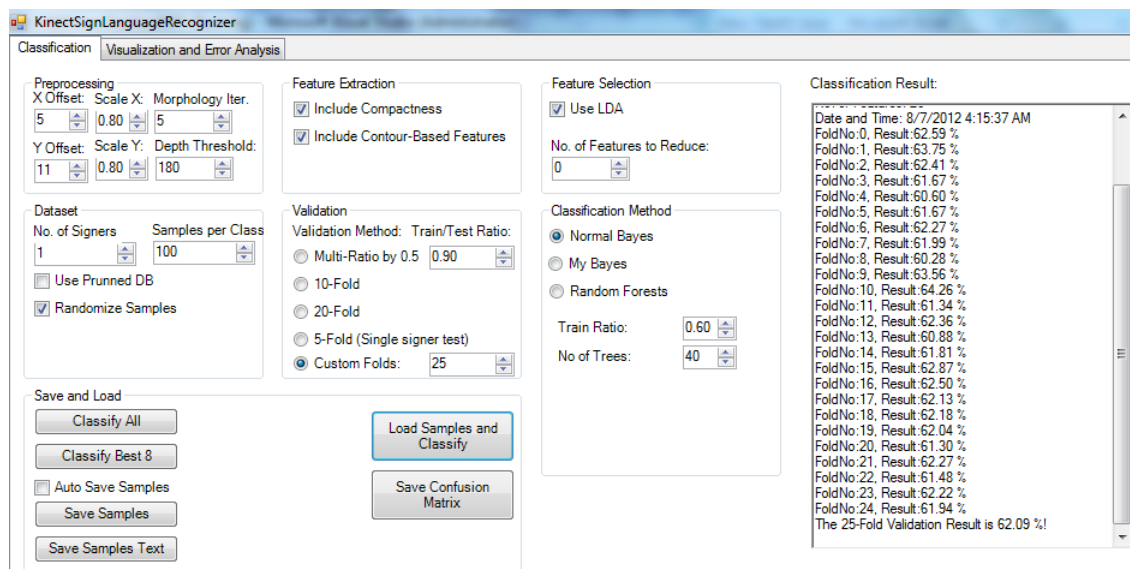
^۵ Multi-Core

^۶ Multi-Threaded

^۷ Thread

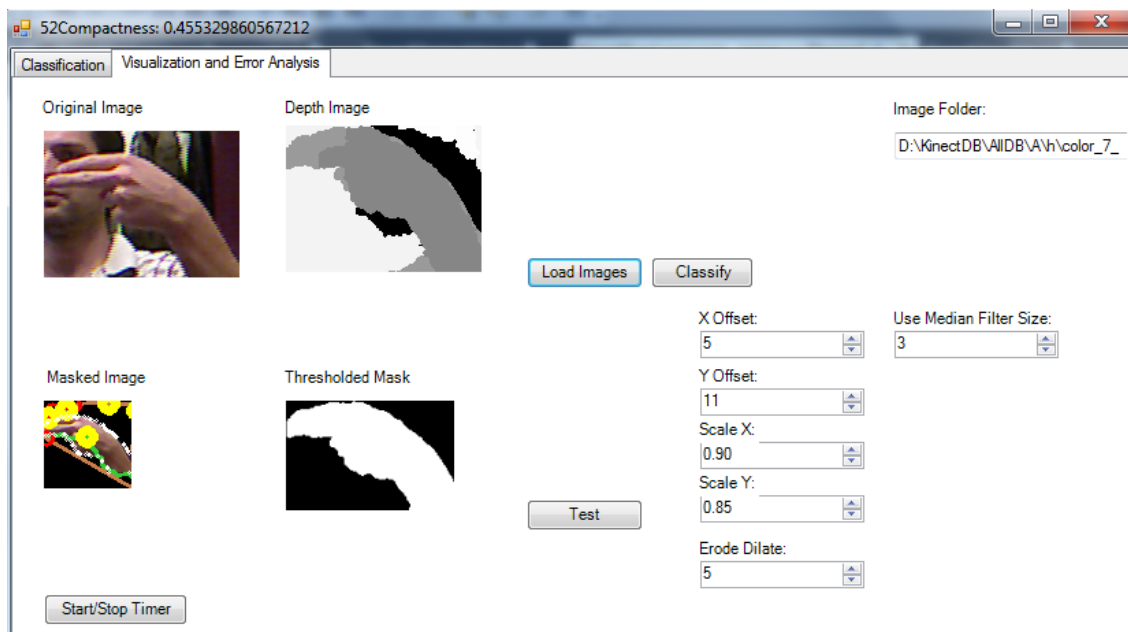
^۸ از آنجا که در پردازنده های جدید، یک وظیفه به تنها به یکی از نخ ها یا هسته های سخت افزاری واگذار می شود، عملاً بقیه هسته ها برای نرم افزارهای دیگر آزاد می باشد. موازی سازی باعث می شود که از ۱۰۰٪ توان پردازنده استفاده شود و عملاً از اتلاف توان پردازشی آزاد جلوگیری شود.

مشاهده می نمایید. همانطور که مشاهده می شود، گزینه های مختلف مربوط به سیستم در نظر گرفته شده است تا کار اعتبارسنجی نیز بتواند به درستی انجام پذیرد.



شکل ۳-۵-نمایی از بخش آموزش و تست نرم افزار پیاده سازی شده در محیط Net. با استفاده از زبان C#.

در بخش تحلیل خطای سیستم، مراحل پیش پردازش و ویژگی های استخراج شده نمایان می شوند و با توجه به نتایج بدست آمده از ماتریس درهم ریختگی، مقادیر کمی و کیفی ویژگی های استخراج شده هر یک از نمونه های یک حرف خاص مورد بررسی قرار گرفته تا تحلیل لازم جهت بهبود یا رفع نقایص سیستم صورت گیرد. در شکل ۳-۶-نمایی از این بخش نرم افزار را مشاهده می نمایید.



شکل ۳-۶- نمایشی از بخش تحلیل خطای نرم افزار پیاده سازی شده و مراحل مختلف پیش پردازش و استخراج ویژگی یک نمونه ورودی در آن.

۳-۳- پیش پردازش

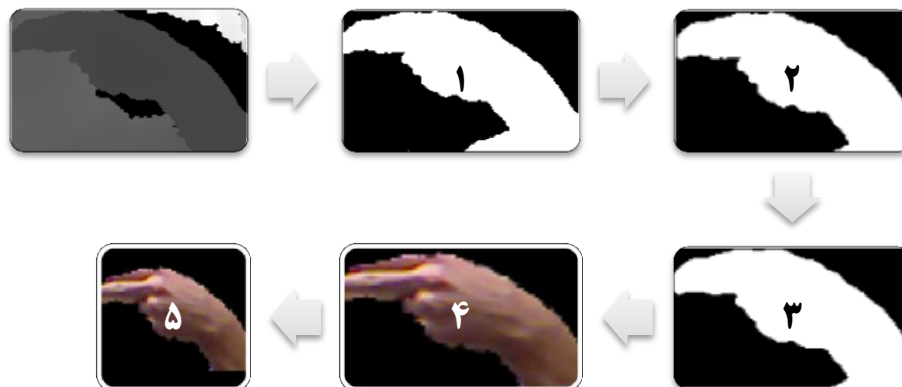
جهت آماده سازی تصاویر برای استخراج ویژگی، چند مرحله برای پیش پردازش بر روی تصویر انجام شده است. در شکل ۳-۷ این مراحل نمایش داده شده اند. به عنوان مثالی از ورودی و خروجی پیش پردازش، یک تصویر رنگی از پایگاه داده انتخاب شده است (شکل ۳-۸) و به ازای تمامی مراحل پیش پردازش و نگاشت یک به یک با نمودار شکل ۳-۷، مراحل مختلف قطعه بندی دست از تصویر عمقی و رنگی شکل ۳-۸، در شکل ۳-۹ نمایش داده شده است. در ادامه هر یک از این مراحل به تفصیل شرح داده می شود.



شکل ۳-۷- مراحل مختلف پیش پردازش انجام شده بر روی داده های تصویر.



شکل ۳-۸- یک تصویر رنگی از نمونه های پایگاه داده مورد استفاده.



شکل ۳-۹ - اعمال مراحل پیش پردازش ارائه شده در شکل ۳-۷ که بر روی تصویر رنگی شکل ۳-۸ اعمال شده است.

۳-۳-۱- ساخت ماسک عمق^۱ با آستانه گیری^۲

همانطور که قبلا گفته شده است، پیکسل های تصاویر عمق دارای مقادیری بین ۰ و ۳۲۰۰ هستند که مشخص کننده عمق پیکسل های تصویر به میلیمتر می باشد. در اینجا با ایجاد دو فرض که در این پایگاه داده درست به نظر می رسد، کار را ادامه می دهیم:

^۱ Depth Mask

^۲ Thresholding

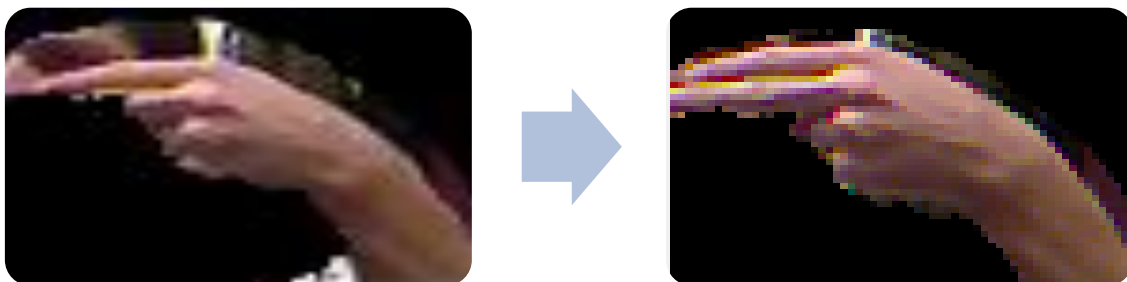
- فرض اول این است که در تصاویر ضبط شده، نزدیکترین پیکسل در تصویر به دست افراد تعلق دارد و دست آن ها در کنار یا پشت سر یا بالاتنه آن ها قرار نگرفته است.
- فرض دوم این است که طبق گفته سازنده پایگاه داده، اندازه دست در حدود ۰.۲ متر می باشد.

با توجه به فرض های ذکر شده در بالا، اگر بتوان از نزدیکترین پیکسل موجود در تصویر تا حدود ۲۰۰ میلی متر در عمق پیش رفت و سپس با نگاه داشتن پیکسل های این دامنه و حذف باقی پیکسل های تصویر عمقی، تصویر باقی مانده را ذخیره کرد، به تصویر قطعه بندی شده دست می توان دسترسی پیدا نمود. برای همین کار، در این بخش ابتدا مقدار عددی پیکسلی را که کمترین فاصله را در تصویر عمقی بعد از صفر (که نشان دهنده نا معتبر بودن عمق تصویر به دلیل فاصله خارج از محدوده کارکرد کینکت می باشد) دارا می باشد را انتخاب، و عدد ۲۵۰ را به نشانه ۲۵۰ میلی متر به آن اضافه نموده و به عنوان مقدار آستانه عمقی در نظر می گیریم. علت انتخاب عدد ۲۵۰ به جای ۲۰۰، انتخاب آسان گیرانه آستانه، جهت نگه داشتن اطلاعات بیشتری از دست می باشد.

پس از انتخاب مقدار آستانه در قسمت قبل، پیکسل هایی از تصویر عمقی را با این مقدار آستانه مقایسه کرده و در صورتی که پیکسلی، مقداری بیشتر از این آستانه (پیکسل هایی با فاصله بیشتری از دوربین) داشت، آن را حذف نموده و سپس بقیه پیکسل های غیر صفر را نگاه می داریم. پس از این آستانه گیری، تصویر مورد نظر را به صورت دو دویی در آورده تا یک ماسک دودویی بدست آید. در مرحله ۱ شکل ۳-۹ خروجی این بخش را مشاهده می نمایید.

۳-۲-۳- جابجایی^۱ ماسک عمقی در صفحه

دوربین رنگی و عمقی کینکت در یک راستا و به صورت موازی با یکدیگر قرار گرفته اند ولی برای اینکه بر روی یکدیگر هم تراز^۲ شوند نیاز به کالیبره شدن دارند. این کار توسط رابط های نرم افزاری معمولاً تا حدی انجام می شود ولی در پایگاه داده فعلی به علت اینکه تصاویر برش خرده اند، نمی توان از فرمول های معمول برای اینکار به راحتی استفاده نمود. از آنجا که ماسک عمقی نهایتاً بایستی بر روی تصویر رنگی اعمال شود، نیاز به یک جابجایی در صفحه دارد. مقدار این جابجایی با آزمایش ظاهری در راستای محورهای مختصات به صورت جفت (۱۱، ۵) قابل اعمال در بردار جابجایی (x, y) تعیین شده است. با جابجایی تمام پیکسل های ماسک عمقی با این بردار، ماسک عمقی تا حد بسیار زیادی بر روی تصویر رنگی منطبق می شود. در شکل ۳-۱۰ تصویر قطعه بندی دست را قبل از اعمال این جابجایی و پس از اعمال این جابجایی مشاهده می نمایید. همانطور که مشخص است، با اعمال جابجایی فوق، بخش بیشتری از تصویر دست پوشانده شده و بخش کمتری از پس زمینه در تصویر مشخص می باشد. در مرحله دوم شکل ۳-۹ نیز نتیجه اعمال این جابجایی را در ماسک عمقی مشاهده می نمایید.



شکل ۳-۱۰- نتیجه اعمال جابجایی ماسک عمقی در قطعه بندی دست و انطباق ماسک بر تصویر رنگی.

¹ Translating

² Align

۳-۳-۳- هموارسازی^۱ ماسک عمقی

همانطور که از تصاویر عمقی بدست آمده در مرحله ۲ شکل ۳-۹ مشخص است، تصاویر عمقی دارای دندان‌های^۲ در لبه تصویر هستند که با توجه به نحوه کارکرد دوربین عمقی کینکت و نویز موجود در تصاویر حاصل، انتظار این نوع نویز در لبه تصویر می‌رود. وجود این نوع نویز باعث اختلاف در تصاویر مشابه شده و به همین دلیل بایستی برای کاهش آن اقدام نمود. برای از بین بردن این نوع نویز بایستی از روش‌های هموارسازی استفاده نمود. فیلترهای مورفولوژیک^۳ و فیلترهای معدل‌گیری^۴ (به خصوص فیلتر میانه^۵) از فیلترهای موثر در این حوزه هستند. از آنجا که فیلترهای مورفولوژیک که در بستن حفره‌ها نیز موثر هستند علاوه بر حذف نویز دندان‌های اطراف تصویر حفره‌های موجود در شکل را نیز پر می‌کنند، این احتمال وجود دارد که علاوه بر پرکردن نویز‌هایی در تصویر عمقی که به صورت حفره‌های کوچکی ممکن است در بیابند، بخش‌هایی از تصویر دست در قسمت بین انگشتان را نیز از بین ببرد. قسمت‌های میان انگشتان در بسیاری از حروف اشاره دست، به دلیل این که حد فاصل بین دو انگشت هستند و در تعداد انگشتان تشخیصی موثر هستند، بایستی تا جایی که ممکن است پس از عمل حذف نویز باقی بمانند. از این رو بایستی روش حذف نویز و دندان‌ها علاوه بر حذف نویز بهینه، این ویژگی‌ها را نیز حفظ نماید. در این قسمت دو فیلتر مورفولوژیک و فیلتر میانه را در این موارد مورد بررسی قرار داده و پس از مشاهده نتایج انتخاب دقیق‌تری در مورد فیلتر حذف نویز انجام خواهیم داد.

¹ Smoothing

² Jaggedness

³ Morphological Filtering

⁴ Averaging Filters

⁵ Median Filter

۳-۳-۱- فیلترهای مورفولوژیک

مورفولوژی شاخه ای از پردازش تصویر است که در تحلیل شکل ها در تصاویر می تواند بسیار کاربردی باشد [۶۴]. بر اساس تئوری های موجود در این شاخه، عملیات ها و فیلتر های مختلفی برای اعمال بر روی تصاویر بوجود آمده اند. از مهمترین عملیات های مورفولوژیک می توان به سایش^۱ و افزایش^۲ اشاره نمود. عملیات های دیگری بر پایه این دو عملیات شکل گرفته اند که opening و closing از این دسته هستند. opening در واقع افزایشی است که پس از یک سایش بر روی تصویر انجام گرفته است و closing نیز سایشی است که پس یک افزایش بر روی تصویر قرار گرفته است. از opening برای حذف نویز های نقطه ای در تصویر استفاده می شود (که در مرحله سایش حذف می-شوند ولی در مرحله افزایش به وجود نمی آیند) و از closing برای پر کردن حفره های موجود در تصویر استفاده می شود (که در مرحله افزایش پر شده ولی در مرحله سایش دوباره به حفره تبدیل نمی شوند). با ترکیب یک مرحله سایش، دو مرحله افزایش و سپس یک مرحله سایش دیگر، می توان تاحد زیادی حذف نویز نقطه ای، پرشدن حفره ها و سپس تبدیل تصویر به اندازه اولیه را تضمین نمود. به این چهار عملیات که درواقع یک عمل closing پس از یک عمل opening می باشد، فیلتر مورفولوژی می گویند که بر روی تصاویر سیاه و سفید و یا سطح خاکستری قابل اعمال هستند. هر فیلتر مورفولوژیک دارای یک هسته^۳ یا عنصر ساختاری^۴ نیز هست که بر حسب کاربردهای مختلف متفاوت است.

برای حذف نویز در ماسک عمقی، از فیلتر مورفولوژیک با استفاده از هسته نشان داده شده در شکل ۳-۱۱ استفاده شده است. اندازه پنجره این عنصر ۳ در ۳ می باشد و علت استفاده از این عنصر

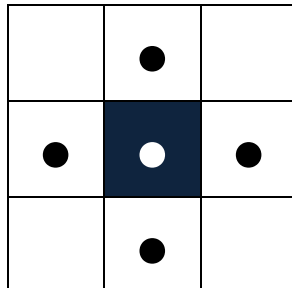
¹ Erosion

² Dilation

³ Kernel

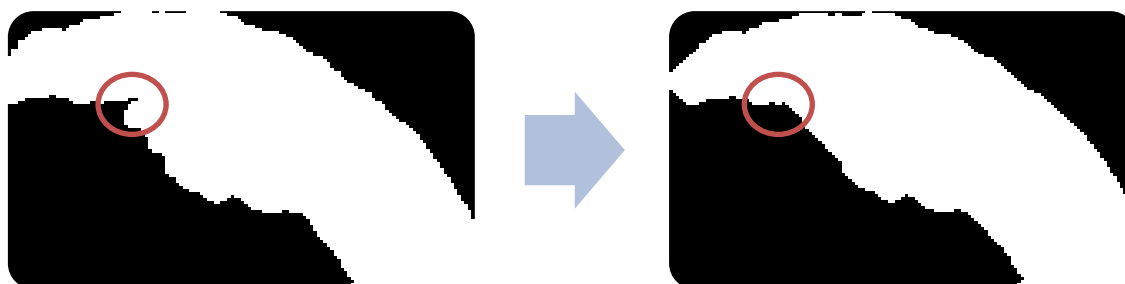
⁴ Structuring Element

ساختاری با مرکز (۱، ۱) قابلیت این عنصر در پر کردن سوراخ های بسیار ریز تصویر و در عین حال از بین بردن نویز ها و پیکسل های زائد اطراف تصویر می باشد.



شکل ۳-۱۱- هسته یا عنصر ساختاری مورد استفاده در فیلتر مورفولوژیک اعمال شده در مرحله پیش پردازش تصویر عمقی. مرکز هسته فوق در مختصات (۱، ۱) واقع شده است.

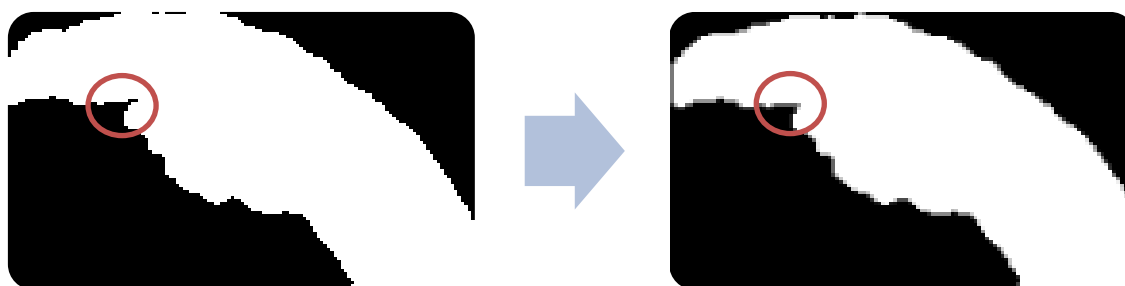
نتیجه اعمال فیلتر فوق بر تصویر عمقی نتیجه شده از مرحله قبل در شکل ۳-۱۲ نشان داده شده است. همانطور که دیده می شود علاوه بر حذف نویز موفقیت آمیز در اطراف تصویر دست، فاصله بین انگشتان نیز به واسطه این فیلتر تا حدودی پر شده است. همانطور که قبلا اشاره شد، از آنجا که این فاصله می تواند اطلاعاتی راجع به نوع الفبای نشان داده شده توسط فرد را برای ما داشته باشد، حذف این فاصله حذف اطلاعات مفیدی محسوب خواهد شد که در دسته بندی می تواند مورد استفاده قرار بگیرد. به همین دلیل استفاده از این فیلتر برای حذف نویز در این کاربرد پیشنهاد نمی-شود.



شکل ۳-۱۲- نتیجه اعمال فیلتر مورفولوژیک بر روی تصویر عمقی حاصل از مرحله قبل.

۳-۳-۲- فیلتر میانه

یکی دیگر از فیلترهای معمول در کاهش نویز اطراف تصویر، فیلتر میانه است. برای اعمال فیلتر میانه بر روی تصویر عمقی، از یک فیلتر با پنجره ۳ در ۳ استفاده شده است. نتیجه اعمال این فیلتر را در شکل ۳-۱۳ مشاهده می کنید. همانطور که مشخص است، تصویر عمقی علاوه بر اینکه با کاهش نویز در اطراف دست مواجه شده است، جزئیات مهم میانه انگشتان نیز در آن تا حد بسیار زیادی حفظ شده است. این ویژگی این فیلتر را از فیلتر مورفولوژیک که در شکل ۳-۱۲ نتیجه آن نشان داده شده است می توان به خوبی تمیز داد و عدم پر شدن فاصله خالی میان انگشتان را در فیلتر میانه مشاهده نمود. از این رو برای بخش حذف نویز اطراف تصویر عمقی، از فیلتر میانه با پنجره ۳ در ۳ استفاده خواهد شد که هم کاهش نویز قابل قبولی داشته باشد و هم فاصله میان انگشتان را تا حد ممکن حفظ نماید. مرحله ۳ از شکل ۳-۹ نمایی از اعمال فیلتر میانه بر روی تصویر بدست آمده از مرحله ۲ همان شکل را نشان می دهد.



شکل ۳-۱۳- نتیجه اعمال فیلتر میانه با سایز پنجره ۳ در ۳ بر روی تصویر عمقی. حذف نویز با حفظ جزئیات بین انگشتان انجام پذیرفته است.

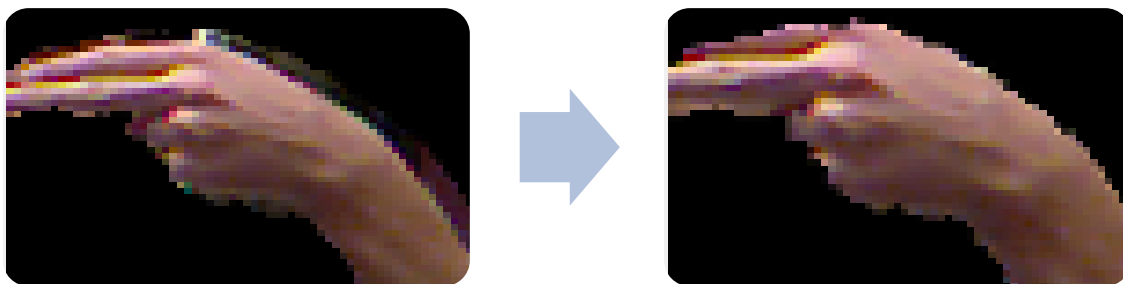
۳-۳-۴- تغییر مقیاس^۱ ماسک عمقی

به علت تفاوت در عمق میدان^۲ دوربین عمقی و رنگی کینکت، تصاویر دو دوربین هنگامی که بر هم منطبق می شوند از لحاظ اندازه نیز با هم اختلاف دارند. این اختلاف باعث می شود که تصویر

¹ Scaling

² Field of View (FOV)

عمقی بدست آمده اندکی بزرگتر از تصویر رنگی کینکت باشد. با آزمایش ظاهری و تجربی چند مقدار برای تغییر مقیاس در راستای محور ها و به صورت بردار (x, y) ، مانند $(0.8, 0.8)$ ، $(0.9, 0.9)$ و $(1.0, 1.0)$ (بدون تغییر مقیاس)، مشخص گردید که پارامتری که بتواند جواب بهتری نسبت به دیگر پارامتر ها به ما بدهد به راحتی به دست نمی آید و به همین دلیل جهت جلوگیری از حذف بخش های از تصویر، نهایتاً از مقیاس $(1.0, 1.0)$ استفاده شده است. بایستی در نظر گرفت که در صورتی که از قبل چنین مقیاسی توسط رابط نرم افزاری بر روی تصاویر اعمال نشده باشد، بایستی با توجه به اطلاعات هم ترازی دوربین عمقی و رنگی کینکت، اقدام به هم تراز کردن مقیاسی تصویر عمقی و تصویر رنگی صورت گیرد. نمونه ای از دست رفتن اطلاعات انگشتان پس از اعمال مقیاس $(0.8, 0.8)$ بر روی ماسک عمقی و سپس اعمال آن بر روی تصویر رنگی را در شکل ۳-۱۴ می بینید. همانطور که مشخص است، تصویر اولیه، بخشی از پس زمینه را نیز شامل شده است که نسب به از دست رفتن اطلاعات دست چشم پوشی شده است. مرحله ۳ از شکل ۳-۹ در واقع نمایی از اعمال فیلتر میانه بدون اعمال عملیات تغییر مقیاس را نشان می دهد. علت عدم حذف این مرحله پس از رسیدن به این نتیجه که عدم اعمال چنین مرحله ای مناسب تر است، این است که در صورتی که پایگاه داده ای با داده های برش داده نشده کینکت در اختیار باشد، این مرحله بایستی به شکل دقیق انجام گردد تا نتایج مطلوب تری بدست آید. همچنین آزمایشات انجام گرفته می تواند برای روش هایی که قصد بهبود این کار را دارند مفید باشد تا دید بهتری از کارهای قبلا انجام شده و قابل انجام برای بهبود این کار به آن ها ارائه شود.



شکل ۳-۱۴- اعمال تغییر مقیاس با اندازه (۰.۸، ۰.۸) بر روی محور های (x, y) که منجر به از دست رفتن اطلاعات انگشتان شده است.

۳-۳-۵- تشخیص کادر محاطی^۱ و برش تصویر رنگی

پس از انجام پیش پردازش های مراحل قبلی، ماسک عمقی بدست آمده بر روی تصویر اعمال می شود (مرحله ۴ از شکل ۳-۹). در این قسمت مشاهده شد که برخی از تصاویر تمام صفحه را به طور کامل نپوشانده و در بخشی از تصاویر تا لبه تصویر تنها پیکسل های سیاه رنگ مشاهده می شود. به همین دلیل، در این مرحله مستطیلی که محاط بر دست قطعه بندی شده می باشد تشخیص داده شده و بقیه قسمت های تصویر رنگی که خارج از کادر محاطی دست می باشد حذف شده است. تشخیص کادر محاطی نیز به سادگی توسط پیدا کردن اولین و آخرین پیکسل غیر سیاه در هریک از جهت های محورهای مختصات انجام گرفته است.

۳-۳-۶- تغییر اندازه^۲ تصاویر

قبل از شروع مرحله استخراج ویژگی ها، بهتر است سایز تمامی تصاویر به شکل یکسانی در آید. با توجه به سایز تصاویر موجود در پایگاه داده که معمولاً از ۲۰۰ در ۲۰۰ پیکسل تجاوز نمی کنند و گاهی تا حدود ۶۰ پیکسل در طول یا عرض می رسند، بایستی اندازه ای انتخاب شود که مناسب ترین اندازه برای مرحله استخراج ویژگی باشد. انتخاب تصاویر با سایز بالاتر، به دلیل اینکه کلی بودن^۳

^۱ Bounding Box

^۲ Resizing

^۳ Approximation

تصاویر را ممکن است از بین ببرد و تاثیر نويز را در تصاویر بالاتر ببرد، به نظر مناسب نمی آید. از طرفی بسیار کوچک بودن تصاویر نیز قابلیت استخراج مناسب ویژگی ها را خدشه دار می سازد. با توجه به اعتبارسنجی های انجام شده در [۴] که اندازه های مختلف تصویر نهایی را جهت انجام تشخیص زبان اشاره با یکدیگر بررسی نموده است، بهترین نتایج با سایز ۳۲ در ۳۲ پیکسل بدست آمده است. در این فرآیند نیز این اندازه جهت تغییر اندازه تمامی نمونه ها قبل از ورود به مرحله استخراج ویژگی انتخاب شده و بر روی آن ها اعمال گردیده است. لازم به ذکر است که این تغییر اندازه با حفظ نسبت تصویر^۱ صورت گرفته است. نتیجه اعمال تشخیص کادر محاطی، برش تصویر و تغییر اندازه آن در مرحله ۵ از شکل ۳-۹ نمایش داده شده است. تصویر بدست آمده در این قسمت پس از تبدیل به فرمت سطوح خاکستری، مستقیماً در مرحله استخراج ویژگی مورد استفاده قرار می-گیرد.

۳-۴- استخراج ویژگی

یکی از مراحل مهم در شناسایی الگو استخراج ویژگی می باشد. در تحقیق فعلی، از آنجا که در مرحله پیش پردازش شکل دست از پس زمینه جدا شده است، یکی از بهترین مجموعه از ویژگی هایی که می توان از تصویر استخراج نمود، ویژگی های هندسی^۲ می باشد [۴]. ویژگی هایی که در این مرحله استخراج شده است در زیر شرح داده شده است.

۳-۴-۱ Moment ها

در واقع Moment ها جهت ارائه مقادیر عددی جهت توصیف یک توزیع^۳ از یک نقطه مرکزی و یا محور خاص بیان شده اند [۶۵]. اولین بار توسط Hu^۱ در سال ۱۹۶۲ برای استفاده در تصاویر به

^۱ Aspect Ratio

^۲ Geometric Features

^۳ Distribution

کار رفت که تصاویر در آن یک توزیع دو بعدی فرض شد که تمامی مرتبه^۲ های Moment ها نیز برای آن قابل محاسبه است. به طور کلی برای یک تصویر سیاه و سفید و یا خاکستری، Moment دو بعدی از مرتبه $p + q$ بر روی تابع توزیع چگالی^۳ $f(x, y)$ (که در اینجا پیکسل های یک تصویر $N \times M$ می باشد) به صورت زیر بدست می آید [۶۵]:

$$m_{pq} = \sum_{y=0}^{M-1} \sum_{x=0}^{N-1} x^p y^q f(x, y) \quad (\text{فصل: ۱-۳۳})$$

برخی از Moment ها در تصویر معنای خاصی در رابطه با قطعه^۴ موجود در تصویر دارند. برای مثال m_{00} به معنای مساحت تصویر یا قطعه خاص موجود می باشد و m_{10} و m_{01} به ترتیب جهت یافتن مرکز ثقل تصویر در راستای محور x ها و y ها استفاده می شوند که ترکیب $(\bar{x} = m_{10}/m_{00})$ $m_{00}, \bar{y} = m_{01}/m_{00}$ مرکز ثقل قطعه موجود در تصویر را مشخص می کند. مرتبه دوم از Moment ها نیز جهت محور توزیع شکل را مشخص می کنند [۴].

۳-۴-۱-۱ Moment های مرکزی^۵

این Moment ها با محاسبه مقادیر نسبی از مرکز ثقل تصویر محاسبه می شوند و استفاده از آن ها به جای Moment ها اولیه، ویژگی هایی نامتغیر^۶ نسب به جابجایی را تولید می کنند. این Moment ها که با μ_{pq} نشان داده می شوند با استفاده از فرمول زیر محاسبه می شوند:

$$\mu_{pq} = \sum_{y=0}^{M-1} \sum_{x=0}^{N-1} (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad (۲-۳)$$

که $\bar{x}$ و $\bar{y}$ به ترتیب مرکز ثقل در راستای محور x ها و y ها می باشند.

^۱ M.K. Hu

^۲ Moment Order

^۳ Density Distribution Function

^۴ Segment

^۵ Central Moments

^۶ Invariant

۳-۴-۱-۲ Hu-Moment ها

با ترکیب Moment های مراتب مختلف با یکدیگر نیز می توان ویژگی های هندسی مناسبی از تصویر مورد نظر استخراج نمود. تعداد هفت ویژگی ترکیبی از Moment های مرکزی تا مرتبه سه توسط Hu ارائه شده اند که از نظر هندسی نسبت به تغییر مقیاس، مکان و جهت^۱ شکل نامتغیر هستند [۶۵]. این ویژگی ها در زیر آمده است. ویژگی هفتم علاوه به موارد ذکر شده فوق، نسبت به کجی^۲ نیز نامتغیر می باشد.

$$M_1 = \mu_{20} + \mu_{02} \quad (۳-۳)$$

$$M_2 = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2 \quad (۴-۳)$$

$$M_3 = (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2 \quad (۵-۳)$$

$$M_4 = (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2 \quad (۶-۳)$$

$$M_5 = (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] \\ + (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \quad (۷-۳)$$

$$M_6 = (\mu_{20} - \mu_{02})[(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \\ + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03}) \quad (۸-۳)$$

$$M_7 = (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] \\ - (\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \quad (۹-۳)$$

¹ Orientation

² Skew

۳-۴-۱-۳- ویژگی‌هایی بر حسب Moment ها

تعداد دیگری از ویژگی‌های نامتغیر نیز می‌توان با استفاده از Moment ها بدست آورد. چند ویژگی‌هایی که در [۴] جهت تشخیص زبان اشاره استفاده شده است نیز در اینجا مورد استفاده قرار گرفته است.

اولین ویژگی بر حسب Moment ها، گریز از مرکز یا بی‌قاعدگی^۱ است که با فرمول زیر از Moment های مرتبه دوم بدست می‌آید و با E نشان داده شده است:

$$E = \frac{(m_{20}-m_{02})^2+4m_{11}^2}{(m_{20}+m_{02})^2} \quad (۱۰-۳)$$

ویژگی‌های دیگر J_1 و J_2 می‌باشد که به ترتیب اینرسی^۲ موازی با محور اصلی و اینرسی عمود بر محور اصلی می‌باشند که نسبت به جابجایی، چرخش و قرینه شدن^۳ نامتغیر هستند و در زیر مشخص شده اند:

$$J_1 = \frac{m_{00}}{2} * (m_{20}+m_{02} + \sqrt{(m_{20} - m_{02})^2 + 4m_{11}^2}) \quad (۱۱-۳)$$

$$J_2 = \frac{m_{00}}{2} * (m_{20}+m_{02} - \sqrt{(m_{20} - m_{02})^2 + 4m_{11}^2}) \quad (۱۲-۳)$$

ویژگی دیگر جهت محور اصلی می‌باشد که با O نشان داده می‌شود و نسبت به جابجایی و تغییر مقیاس نامتغیر می‌باشد و به صورت زیر محاسبه می‌شود:

$$O = \frac{180}{2\pi} \arctan\left(\frac{2m_{11}}{m_{20}-m_{02}}\right) \quad (۱۳-۳)$$

چهار ویژگی فوق از تصاویر استخراج شده و به بردار ویژگی‌ها اضافه شده اند. جمع کل تعداد ویژگی‌های مورد استفاده بر پایه Moment ها ۱۹ عدد می‌باشد.

¹ Eccentricity

² Inertia

³ Flip

۳-۴-۲- ویژگی های بر مبنای کانتور

تعدادی از ویژگی ها نیز برپایه استخراج لبه بیرونی تصویر یا کانتور آن استخراج شده اند. برای این کار ابتدا تمامی لبه های بیرونی تصویر (لبه جسم^۱ های موجود در تصویر) استخراج و سپس لبه بزرگترین جسم تصویر به عنوان لبه بیرونی یا کانتور شکل استخراج گردیده است. شکل ۳-۱۵ کانتور استخراج شده از یک نمونه دست قطعه بندی شده را نشان می دهد که به صورت آرایه ای از نقاط سفید رنگ نشان داده شده است.



شکل ۳-۱۵- بزرگترین لبه تصویر یا کانتور استخراج شده که نقاط آن در اطراف دست قطعه بندی شده مشخص شده اند.

یکی از ویژگی های مهمی که می توان از کانتور استخراج نمود، ویژگی فشردگی^۲ است که به صورت زیر محاسبه می شود [۴]:

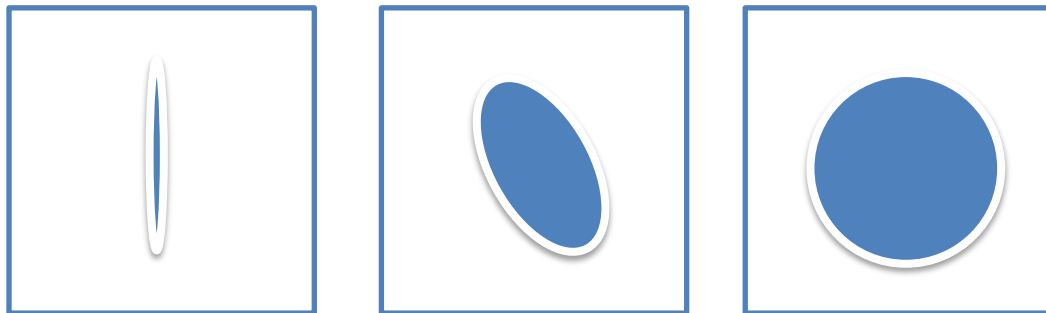
$$C = \frac{4 * \pi * A}{B^2} \quad (3-14)$$

که در آن C نماد فشردگی، A مساحت جسم (یا همان m_{00}) و B محیط جسم است که توسط طول کانتور جسم بدست می آید. در شکل ۳-۱۶ می توانید نمایی از اشکالی که مقادیر مختلف فشردگی را دارند را مشاهده کنید. همانطور که از شکل و نیز فرمول فشردگی مشخص است، این ویژگی برای جسمی شبیه به یک خط با مساحت کم و محیط زیاد عددی نزدیک به صفر دارد (تصویر

¹ Blob

² Compactness

ج) و برای یک دایره کامل مقدار عددی یک را به عنوان ویژگی به دست می دهد (تصویر الف). این ویژگی می تواند معیار بسیار خوبی برای جداسازی دست قطعه بندی شده باشد.



الف- $C = 1$ ب- $0 < C < 1$ ج- $0 \cong C \ll 1$

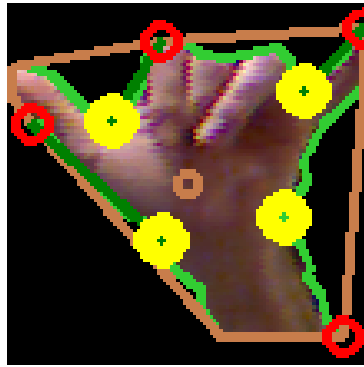
شکل ۱۶-۳- نمایشی از مقادیر مختلف ویژگی فشردگی (C) بر اساس شکل ظاهری اجسام موجود در تصویر.

۳-۴-۲-۱- ویژگی های بر حسب پوسته محدب^۱

ویژگی های دیگری که بر اساس کانتور شکل می توان بدست آورد، ویژگی هایی بر حسب پوسته محدب شکل می باشد. پوسته محدب شکل در واقع مانند نواری است که بر دور شکل کشیده می شود ولی بر خلاف کانتور شکل، از روی چاله^۲ های مقعر شکل مستقیماً عبور کرده تا شکل تحدب خود را از دست ندهد. در شکل ۱۷-۳ تصویری از پوسته محدب دور دست به رنگ قهوه ای، کانتور شکل به رنگ سبز، نقاط شروع چاله های پوسته محدب به شکل دایره های توخالی قرمز و نقاط پایانی چاله های پوسته که در عمق چاله ها واقع شده اند را به شکل دایره های توپُر زرد رنگ سوراخ شده مشاهده می نمایید. چهار چاله در این شکل پیدا شده است.

¹ Convex Hull

² Defects



شکل ۳-۱۷- تصویری از پوسته محدب استخراج شده از یک نمونه از حرف y از پایگاه داده.

با توجه به توضیحات فوق، چندین ویژگی بر اساس پوسته محدب استخراج شده است. اولین ویژگی، تعداد انگشتان است. انگشتان با چند شرط از نقاط شروع و پایان چاله ها محاسبه شده است. اگر ارتفاع نقطه شروع و یا نقطه پایان یک چاله از چهار ششم ارتفاع پوسته دست بیشتر باشد و همزمان ارتفاع نقطه شروع از نقطه پایان بالاتر باشد و همزمان عمق چاله از یک ششم ارتفاع پوسته دست بیشتر باشد، آن چاله به عنوان محل فرورفتگی کنار یک انگشت محاسبه می شود. این شرط با آزمایش تجربی بدست آمده است. تعداد انگشتان دست، به عنوان اولین ویژگی بر مبنای پوسته، به بردار ویژگی ها اضافه شده است.

دیگر ویژگی، تعداد چاله ها بدون توجه به تعداد انگشتان است و دو ویژگی بعدی به صورت میانگین طول بردار نقاط شروع تا عمق چاله ها و میانگین طول بردار عمق تا نقاط انتهایی چاله ها محاسبه شده است. میانگین عمق چاله ها نیز به عنوان ویژگی دیگر استخراج و در نهایت نسبت طول کل بردارهای نقاط شروع تا عمق به طول کل بردارهای عمق تا پایانی نیز به عنوان ویژگی دیگر تعیین و به بردار ویژگی ها اضافه گردیده است.

در نهایت هفت ویژگی بر حسب لبه بیرونی یا کانتور استخراج و تعداد کل بردار ویژگی ها به ۲۶ عدد رسیده است. این ۲۶ ویژگی به طور خلاصه به همراه توصیف لاتین و نام اختصاری آن ها در جدول ۳-۱ آمده است. نام های اختصاری در جدول وابستگی ویژگی ها در فصل اعتبارسنجی مورد استفاده قرار خواهند گرفت.

جدول ۱-۳- لیست ویژگی های استخراج شده به همراه نام اختصاری آن ها.

ردیف	نام ویژگی	اختصار	ردیف	نام ویژگی	اختصار
۱	Hu Moment 1	Hu1	۱۴	μ_{12}	u12
۲	Hu Moment 2	Hu2	۱۵	μ_{11}	u11
۳	Hu Moment 3	Hu3	۱۶	Eccentricity	E
۴	Hu Moment 4	Hu4	۱۷	Inertia Parallel to Main Axis	J1
۵	Hu Moment 5	Hu5	۱۸	Inertia Orthogonal to Main Axis	J2
۶	Hu Moment 6	Hu6	۱۹	Orientation of Main Axis	O
۷	Hu Moment 7	Hu7	۲۰	Compactness	C
۸	1/sqrt(m ₀₀)	s00	۲۱	No. of Fingers	NF
۹	μ_{20}	u20	۲۲	No. of Defects	ND
۱۰	μ_{02}	u02	۲۳	Mean Length of Defects Start to Depth Vector	DLS
۱۱	μ_{30}	u30	۲۴	Mean Length of Defects Depth to End Vector	DLE
۱۲	μ_{03}	u03	۲۵	Defects Total Start-Depth Vectors Length to Total Depth-End Vectors Length Ratio	DLR
۱۳	μ_{21}	u21	۲۶	Mean of Defects Depth Height	DD

۳-۵- تحلیل و انتخاب ویژگی ها

در قسمت استخراج ویژگی ها ۲۶ ویژگی از تصویر دست قطعه بندی شده استخراج شده است که قبل از ورود به مرحله دسته بندی بایستی تحلیل شوند و در صورت نیاز تعدادی از آن ها که لازم هستند، انتخاب شوند. برای تحلیل ویژگی ها از LDA^۱ استفاده شده است. ورودی LDA بردار ویژگی های داده آموزشی و کلاس بندی داده ها می باشد. بر خلاف PCA^۲ که تحلیل بردار ویژگی ها را بدون استفاده از اطلاعات کلاس بندی داده ها انجام می دهد، LDA با دریافت اطلاعات کلاسی، بردار ویژگی ها را به شکلی تغییر می دهد که بردار ویژگی ها بتوانند بهترین جداسازی را برای داده ها مهیا سازند. این مهم با کمینه کردن فاصله درون کلاسی ویژگی ها و بیشینه کردن فاصله بین کلاسی آن ها انجام می شود.

در تحقیق فعلی از LDA برای بدست آوردن ضریب هایی که تبدیل بردار ویژگی های داده های آموزش و آزمایش را مهیا می سازند استفاده شده است. ماتریس ضرایب که یک ماتریس مربعی با اندازه طول بردار ویژگی ها (۲۶) است، همچنین به شکلی مرتب شده است که ویژگی هایی با بیشترین تاثیر در جداسازی داده ها در ابتدای آن و ویژگی هایی با کمترین تاثیر در جداسازی داده ها در انتهای آن قرار گیرد. به این ترتیب با ضرب این ماتریس ضرایب در داده ها و حذف تعدادی از ویژگی ها از انتهای بردار ویژگی ها، می توان بار محاسباتی را تا جایی که لطمه شدیدی به درصد شناسایی نخورده باشد، کاهش داد.

تاثیر کاهش ویژگی ها توسط LDA و بحث های مربوط به استقلال ویژگی ها در فصل مربوط به اعتبار سنجی (فصل ۵) مورد بررسی قرار خواهند گرفت. به دلیل اینکه تعداد ویژگی های مورد استفاده که ۲۶ ویژگی می باشد، سربار زیادی بر سیستم نخواهد داشت (در فصل ۵ بحث خواهد شد)

^۱ Linear Discriminant Analysis

^۲ Principle Component Analysis

و همه ویژگی ها با دلیل انتخاب شده اند، در دسته بندی از تعداد ویژگی ها کاسته نشده است و تمامی آن ها برای آزمایش سیستم مورد استفاده قرار گرفته اند. بدیهی است که اطلاعات مربوط به تاثیر کاهش ویژگی ها در فصل ۵ ارائه خواهد شد.

۳-۶- روش دسته بندی

برای دسته بندی داده ها با استفاده از دسته بند روش های مختلفی وجود دارد. با توجه به ماهیت داده ها و پایگاه داده که تعداد زیادی ویژگی به همراه تعداد زیادی داده را به همراه دارد، استفاده از دسته بند هایی مناسب خواهد بود که ترجیحاً مرحله آموزش آن ها یک مرحله ای باشد و برای هر داده تکرار نشود. برای مثال دسته بند نزدیکترین همسایه یا KNN^۱ برای این کار مناسب نمی باشد زیرا در مرحله از تست بایستی داده های بسیار زیادی با داده ورودی مقایسه شده و سربار بسیار زیادی دربرخواهد داشت. با توجه به این که طول بردار ویژگی های انتخابی ۲۶ می باشد و همگی با توجه به ویژگی های دست قطعه بندی انتخاب شده و به نظر جداکننده^۲ می آیند، دسته بند های احتمالی مانند دسته بند بیز و جنگل تصادفی (که در کار [۸] استفاده شده بود) که براحتی نیز بتوانند نتیجه دسته بندی بر حسب هر ویژگی را با یکدیگر ترکیب نمایند، مناسب تر به نظر می رسند. در هر دو این روش ها، پس از پایان مرحله آموزش، محله آزمایش به سرعت انجام می پذیرد و سربار بسیار کمی را برای سیستم خواهند داشت.

در نهایت جهت سیستم پیشنهادی، دسته بند بیز ساده^۳ به دلیل پیچیدگی پایین تر و سرعت بالاتر و قدرت احتمالی که اجازه اعمال احتمالات پیشین را نیز در آینده میسر می کند، انتخاب شده است و در دسته بندی ویژگی های بدست آمده از مرحله تحلیل ویژگی ها استفاده گردیده است.

^۱ K-Nearest Neighbor

^۲ Discriminant

^۳ Naïve Bayes

در این دسته بند، نمونه S_k با بردار ویژگی های $(F_1 = v_1, F_2 = v_2, \dots, F_m = v_m)$ به کلاس C_i اختصاص می یابد به صورتی که به ازای تمامی مقادیر i ، مقدار احتمال شرطی $Prob(C_i | (v_1, v_2, \dots, v_m))$ بیشینه شود. در فرمول زیر نحوه بدست آوردن این احتمال نشان داده شده است:

$$Prob(C_i | (v_1, v_2, \dots, v_m)) = \frac{Prob((v_1, v_2, \dots, v_m) | C_i) P(C_i)}{Prob((v_1, v_2, \dots, v_m))} \quad (۱۵-۳)$$

که در آن $P(C_i)$ احتمال وقوع هر کلاس است که به احتمال پیشین معروف است. در حالتی که تعداد نمونه های کلاس ها متفاوت باشند از این احتمال می توان استفاده نمود. همانطور که گفته شد، یکی از دلایل استفاده از این دسته بند، وجود احتمال پیشین است به صورتی که اگر در ادامه بر روی تشخیص کلمات و جملات زبان اشاره کار شود، می توان از قواعد ساختاری و احتمالی کنار هم قرار گرفتن کلمات به صورت احتمالی استفاده نمود و براحتی به جای این احتمال پیشین قرار داد.

فصل ۴: نتایج آزمایشات

در این فصل به بررسی نتایج بدست آمده از آزمایش های انجام شده بر روی داده های پایگاه داده مورد استفاده تحت انتخاب پارامترهای مختلف و مجموعه داده های مختلف از این پایگاه داده می پردازیم و سپس نتایج بدست آمده را با کارهای پیشین در همین حوزه مقایسه خواهیم نمود.

۴-۱- تعاریف مورد استفاده در نتایج

در این قسمت برخی از تعاریف مورد استفاده در تحلیل نتایج را بیان نموده و سپس نتایج را بر این اساس ارائه خواهیم نمود.

۴-۱-۱- اعتبار سنجی ضربدری^۱ با n -تا کردن^۲

برای اطمینان از اعتبار نتایج ارائه شده و اینکه نتایج بر اساس بخش خاصی از داده ها که به عنوان داده های آزمایشی^۳ جدا شده اند نباشد، از اعتبار سنجی ضربدری با n -تا کردن استفاده شده است. در این روش بر اساس مقدار n ، کل پایگاه داده را به n قسمت تقسیم نموده و $n-1$ بخش آن را جهت آموزش^۴ سیستم و یک بخش آن را جهت آزمایش سیستم استفاده می کنیم. جهت اطمینان از نتیجه حاصله در تمامی حالات، این عمل n بار تکرار شده و در هر مرتبه بخش آزمایش سیستم یکی از n قسمت خواهد شد به طوری که تمامی قسمت ها در بخش آزمایش سیستم بررسی شوند. نتیجه نهایی این بخش به صورت میانگین نتیجه این n قسمت محاسبه و اعلام شده است.

برای پایگاه داده فعلی، از ۵-تا کردن جهت بررسی روش به صورتی استفاده می شود که در هر مرتبه ۴ اجرا از ۵ اجرا پایگاه داده مورد بررسی قرار گیرد و اجرا باقی مانده به عنوان داده تست به سیستم داده شود. از ۱۰-تا کردن نیز به عنوان یک الگوی معمول در مقالات جهت بررسی عملکرد

^۱ Cross Validation

^۲ n-Fold

^۳ Test

^۴ Train

سیستم استفاده شده است و از ۲۵- تا کردن برای تست سیستم با دقت فوق العاده بالا و متناسب با ساینز پایگاه داده استفاده شده است. درصد آموزش و آزمایش نیز با توجه به n تعیین می شود به این صورت که درصد داده های آموزشی و آزمایشی از فرمول های زیر تعیین می گردد:

$$Train Ratio = \frac{(n - 1)}{n} * 100 \quad (۱-۴)$$

$$Test Ratio = \frac{1}{n} * 100 \quad (۲-۴)$$

که $Train Ratio$ درصد داده های آموزشی نسبت به کل داده ها و $Test Ratio$ درصد داده های آزمایشی نسبت به کل داده ها را تعیین می نماید که در حالت ۵- تا کردن نسبت داده های آزمایشی به آموزشی ۲۰ به ۸۰، برای حالت ۱۰- تا کردن این نسب ۱۰ به ۹۰ و برای حالت ۲۵- تا کردن ۴ به ۹۶ می باشد. با ضرب این مقادیر در کل داده های مورد استفاده در سیستم، تعداد نمونه های آموزشی و آزمایشی سیستم بدست می آید. برای مثال در حالتی که اعتبارسنجی ضربدری با ۲۵- تا کردن استفاده می شود و کل پایگاه داده مورد استفاده قرار می گیرد، ۲۱۶۰ نمونه به عنوان داده آزمایشی و ۵۱۸۴۰ نمونه به عنوان داده آموزشی در هر بار تا کردن به سیستم داده شده است.

۴-۱-۲- تصادفی کردن^۱ داده ها

نمونه های مورد استفاده در سیستم از اولین حرف الفبای فرد اول شروع و پس از پایان نمونه های فرد اول، نمونه های دیگر افراد به ترتیب حروف الفبا ادامه می یابد و با حرف y از اجرا پنجم پایان می یابد (همانطور که قبلا گفته شد، حروف z و z به دلیل حرکتی بودن در این پایگاه داده وجود ندارند). اگر داده ها به ترتیب در سیستم قرار گیرند، در حالتی که ۵- تا کردن اتفاق می افتد، همیشه در داده های آموزشی ۴ نفر حضور دارند و در داده های آزمایشی نفر پنجمی حضور دارد که هیچ یک

¹ Randomization

از داده های آن در آموزشی حضور نداشته است. در ۱۰- تا کردن و ۲۵- تا کردن چنین وضعیتی تخفیف می یابد و کمرنگ تر می شود. برای اینکه در حالات مختلف مخصوصاً زمانی که برای یک نفر سیستم آزمایش می شود، نمونه ها به شکل نسبتاً یکنواختی در تمامی بخش های آموزش و آزمایش توزیع شده باشند، داده ها را به صورت تصادفی در پایگاه داده قرار می دهیم. گزینه ای به نام تصادفی کردن در جدول نتایج وجود دارد که در صورت مثبت بودن آن، این عمل انجام شده است و در غیر اینصورت داده ها به همان شکل اولیه وجود خواهند داشت.

۴-۱-۳- معیارهای ارزیابی

برای بیان کمی چگونگی عملکرد سیستم، بایستی از معیار های اندازه گیری درستی عملکرد دسته بندی استفاده نمود. برای بیان این معیار ها، ابتدا ماتریس در هم ریختگی توضیح داده می شود. در ماتریس در هم ریختگی، که یک ماتریس مربعی با اندازه تعداد کلاس های دسته بندی مورد نظر می باشد، مشخص می شود که چه تعداد از کلاس های مورد نظر در سیستم، به درستی در کلاس خود دسته بندی شده اند. استفاده از این ماتریس همچنین مشخص می کند که کدامین کلاس ها اشتباهاً در کدامین کلاس های دیگر دسته بندی شده اند که بسیار در بحث تحلیل خطای سیستم می تواند موثر باشد. در جدول ۴-۱ نمونه ای از یک ماتریس در هم ریختگی ساده شده برای حروف a، b و c و ۱۰ داده آزمایشی به ازای هر حرف مشاهده می شود.

جدول ۴-۱- نمونه ای از یک ماتریس در هم ریختگی

خروجی\هدف	a	b	c
a	۹	۰	۱
b	۲	۷	۱
c	۰	۰	۱۰

برای مثال در این جدول مشخص شده است که از ۱۰ نمونه حرف a که به سیستم برای آزمایش داده شده است، ۹ نمونه به درستی a و یک نمونه به اشتباه c تشخیص داده شده است. همچنین مشخص است که تمامی نمونه های حرف c به درستی تشخیص داده شده اند در حالی که برای حرف b کمترین تعداد شناسایی صحیح نمونه های آزمایشی با ۷ شناسایی صحیح گزارش شده است.

چهار تعریف مهم بر روی ماتریس درهم ریختگی بوجود می آید. این معیارها در جدول ۲-۴ مشخص شده اند. برای مثال معیار FN تعداد نمونه هایی را مشخص می کند که بایستی در کلاس مشخص شده قرار می گرفته اند، ولی در کلاس دیگری قرار گرفته اند و معیار FP تعداد نمونه هایی را مشخص می کند که نبایستی در یک کلاس قرار می گرفته اند ولی به اشتباه در آن کلاس قرار گرفته اند [۶۶].

جدول ۲-۴- چهار تعریف مهم بر روی جدول درهم ریختگی.

		خروجی سیستم	
		درست	غلط
هدف شناسایی	درست	True Positive (TP)	False Negative (FN)
	غلط	False Positive (FP)	True Negative (TN)

معیار های ارزیابی مختلفی با توجه به جدول فوق و بر اساس تعاریف TP، FN، FP و TN مشخص شده اند که ارزیابی نهایی سیستم بر اساس آن ها مشخص می شود. مهمترین این معیار ها دقت^۱، یادآوری^۲ و صحت^۳ می باشد که با فرمول های زیر مشخص شده اند [۶۶]:

$$Precision = TP / (TP + FP) \quad (۳-۴)$$

^۱ Precision

^۲ Recall

^۳ Accuracy

$$Recall = TP / (TP + FN) \quad (۴-۴)$$

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (۵-۴)$$

که در آن Precision دقت، Recall یادآوری و Accuracy صحت خروجی های دسته بند را مشخص می کند. به بیان دیگر دقت مشخص می کند که چند درصد از نمونه هایی که در یک کلاس قرار داده شده اند، در واقع به آن کلاس تعلق داشته اند. همچنین یادآوری مشخص می کند که چند درصد از نمونه های آزمایشی یک کلاس به درستی در کلاس خود قرار داده شده اند. در نهایت صحت به عنوان یک معیار اصلی، نشان می دهد که چند درصد از کل نمونه های ورودی یک سیستم، به درستی در کلاس های خود قرار داده شده اند (یا به عبارتی میانگین یادآوری کلاس ها می باشد). برای مثال برای حرف b در جدول ۴-۱ دقت ۱۰۰٪ و یادآوری ۷۰٪ می باشد در حالی که برای حرف c دقت ۸۳٪ و یادآوری ۱۰۰٪ می باشد. صحت کل سیستم فرضی این جدول نیز ۸۶.۷٪ می باشد. این نشان می دهد که هر دو معیار دقت و یادآوری برای ارزیابی سیستم بایستی ارائه شود تا بتوان توانایی و دقت یک سیستم دسته بندی را ارائه نمود. در جداول نتیجه این تحقیق، هنگامی که از درصد شناسایی سیستم نام برده شود، منظور صحت شناسایی سیستم است و هنگامی که از دقت سیستم یاد شود، منظور میانگین دقت کلاس های مختلف یک سیستم می باشد.

۴-۱-۴- تصویر قطعه بندی شده

در نتایج ارائه شده، در قسمت تصویر قطعه بندی شده، سه نوع داده قرار داده شده است: تصویر رنگی، تصویر عمقی و ترکیب ویژگی ها. این سه نوع متغیر به این معنی می باشد که پس از ایجاد ماسک عمقی در بخش پیش پردازش، این ماسک بر تصویر رنگی اعمال شود یا بر تصویر عمقی و یا ترکیبی از ویژگی های این دو باشد. اعمال ماسک بر تصویر عمقی نیز به دلیل وجود جزئیات عمقی دست در این تصویر می تواند مفید باشد. در حالت ترکیبی، بردار ویژگی حاصل از پردازش

نهایی دو حالت فوق (هر حالت ۲۶ ویژگی) با یکدیگر ترکیب شده و ۵۲ ویژگی حاصل، پس از تحلیل ویژگی ها به دسته بند داده می شود تا با اطلاعاتی مضاعف دسته بندی انجام شود.

۴-۲-نتایج

نتیجه دسته بندی در این سیستم با پارامترهای مختلف و با انتخاب های مختلف از پایگاه داده در جدول ۳-۴ آمده است. در این جدول ۲۴ حرف الفبای زبان اشاره آمریکایی که در پایگاه داده استفاده شده موجود بوده است مورد دسته بندی قرار گرفته است. ترکیبات مختلف از تعداد افراد انتخاب شده از پایگاه داده، تعداد نمونه های انتخاب شده از هر حرف به ازای هر نفر، تصویر مورد استفاده در قطعه بندی، استفاده از داده های تصادفی و n -تاکردن های متفاوت در جدول ۳-۴ قرار داده شده است و نتیجه دسته بندی به دو صورت صحت و دقت میانگین ارائه شده است. در هر بخش نیز به ازای تعداد افراد و تعداد نمونه های به ازای هر حرف و هر نفر، بهترین جواب بدست آمده نشان داده شده است. نتایج نشان می دهد که دسته بندی روش پیشنهادی بر روی کل پایگاه داده با ۴۵۰ نمونه به ازای هر حرف و نفر و تعداد ۵۴۰۰۰ نمونه کل در بهترین حالت با ۷۸.۲۹٪ صحت و ۸۰.۲۳٪ دقت انجام گرفته است و در حالتی که دسته بندی بر روی اولین نفر موجود در پایگاه داده با ۱۰۰ نمونه اول به ازای هر حرف و ۲۴۰۰ نمونه کل انجام گرفته، ۹۹.۷۵٪ صحت و ۹۹.۷۶٪ دقت در بهترین حالت بدست آمده است. در ادامه نتایج بدست آمده تحلیل و با روش های پیشین مقایسه خواهد شد.

جدول ۳-۴- نتایج بدست آمده توسط سیستم پیشنهادی به تفکیک پارامترهای مختلف.

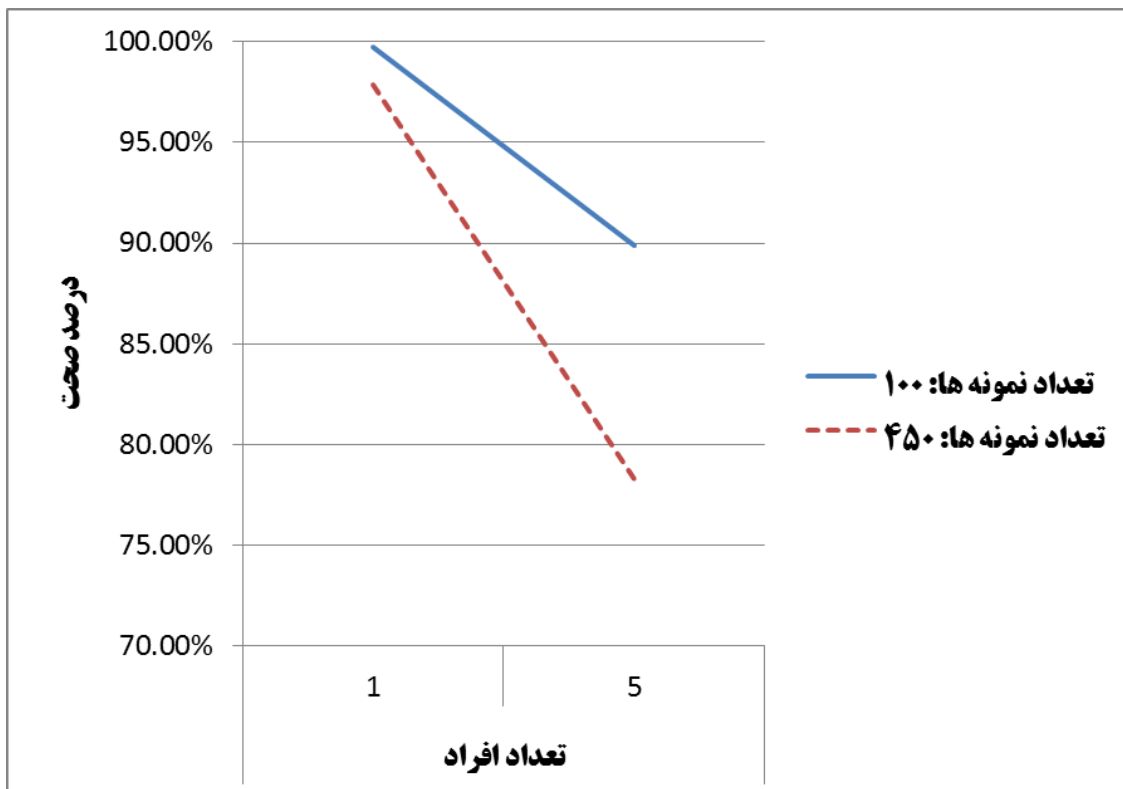
ردیف	تعداد افراد	نمونه از هر حرف	کل نمونه ها	تصویر قطعه بندی شده		صحت	دقت	صحت	دقت	صحت	دقت	صحت	دقت
					تصادفی کردن	۵- تا کردن	۱۰- تا کردن	۲۵- تا کردن	بالاترین درصد				
۱	۱	۱۰۰	۲۴۰۰	رنگی	خیر	89.58	91.92	95.75	96.07	98.04	98.11	99.76	99.75
۲					بله	99.12	99.14	99.17	99.18	99.21	99.22		
۳				عمقی	خیر	90.13	91.94	94.75	95.24	98.37	98.44		
۴					بله	99.13	99.15	99.25	99.27	99.21	99.23		
۵				ترکیبی	خیر	91.42	93.42	96.04	96.45	99.21	99.25		
۶					بله	99.75	99.75	99.71	99.72	99.75	99.76		
۷	۱	۴۵۰	۱۰۸۰۰	رنگی	خیر	89.03	89.8	90.54	91.21	92.21	92.72	98.14	97.86
۸					بله	95.6	95.99	95.63	96.02	95.67	96.05		
۹				عمقی	خیر	86.52	87.37	88.33	89.01	90.59	91.13		
۱۰					بله	94.63	95.01	94.65	95.02	94.63	95.01		
۱۱				ترکیبی	خیر	90.83	91.76	92.62	93.24	94.57	94.97		
۱۲					بله	97.85	98.11	97.81	98.08	97.86	98.14		
۱۳	۵	۱۰۰	۱۲۰۰۰	رنگی	خیر	47.38	48.77	66.53	66.11	72.68	72.72	90.24	89.87
۱۴					بله	80.76	81.45	80.69	81.43	80.7	81.43		
۱۵				عمقی	خیر	45.34	45.79	60.58	60.79	66.83	67.36		
۱۶					بله	74.92	76.6	74.88	76.41	74.87	76.45		
۱۷				ترکیبی	خیر	46.98	51.75	74.18	74.79	80.67	81.03		
۱۸					بله	89.71	90.07	89.87	90.24	89.73	90.1		
۱۹	۵	۴۵۰	۵۴۰۰۰	رنگی	خیر	50.02	50.11	61.8	62.23	66.32	67.24	80.23	78.29
۲۰					بله	70.71	72.24	70.76	72.33	70.81	72.37		
۲۱				عمقی	خیر	47.36	46.8	56.75	57.89	60.57	62.36		
۲۲					بله	64.59	66.91	64.58	66.91	64.53	66.85		
۲۳				ترکیبی	خیر	53.13	54.65	68.19	69.46	72.66	74.16		
۲۴					بله	78.27	80.09	78.29	80.22	78.29	80.23		

۴-۳-تحلیل نتایج

در این بخش نتایج بدست آمده از دیدگاه های مختلف تحلیل شده و دستاورد های آن شرح داده می شود. همچنین تاثیر تغییر پارامترهای مختلف که در نتایج جدول ۴-۳ آورده شده است بررسی شده و نمودارهای مربوط به آن ها ارائه می شود.

۴-۳-۱-تاثیر تغییر تعداد افراد

با توجه به نتایج نشان داده شده در جدول ۴-۳، مشخص شده است که با افزایش تعداد افراد موجود در پایگاه داده، قدرت شناسایی سیستم نیز کاهش می یابد. در نمودار شکل ۴-۱ نیز این کاهش صحت به ازای افزایش افراد از یک به پنج نفر و به ازای تعداد نمونه های مختلف نشان داده شده است. درصد های ارائه شده در شکل به در حالت اعتبارسنجی با ۲۵- تا کردن و تصادفی کردن داده ها بدست آمده است.



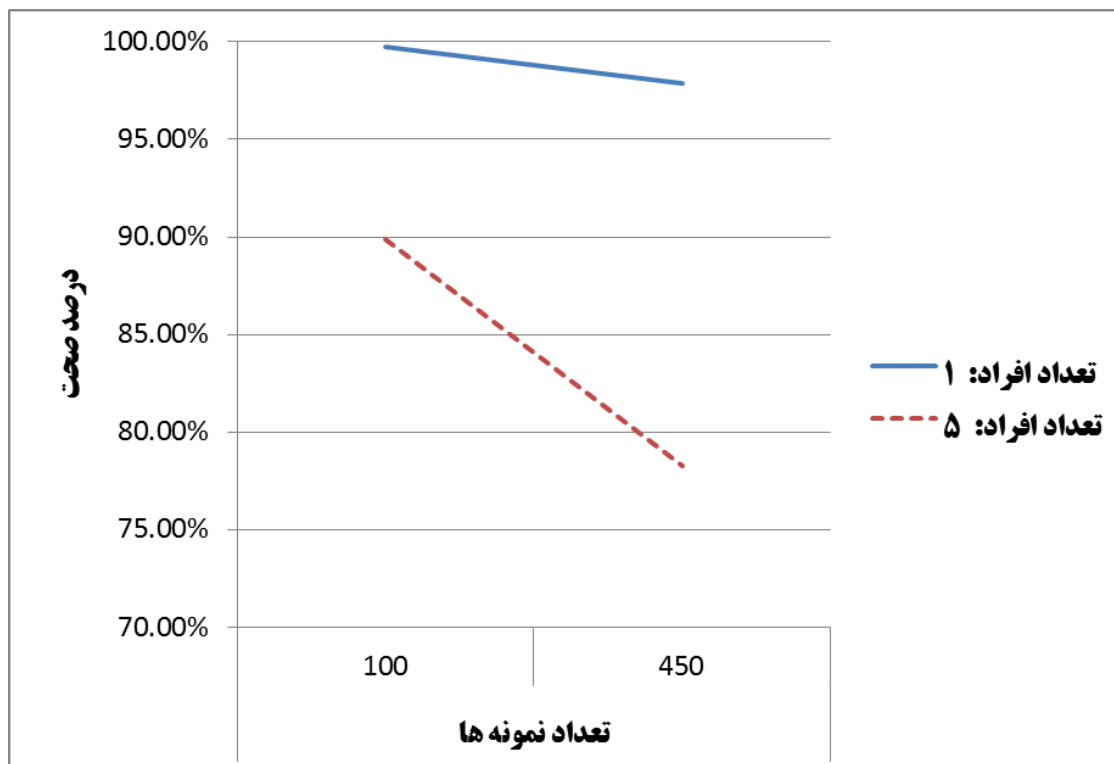
شکل ۴-۱-نمودار تاثیر افزایش افراد پایگاه داده در صحت خروجی سیستم به ازای تعداد نمونه هر حرف.

علت این کاهش در صحت با افزایش تعداد افراد به دلیل این است که با تنوع افراد، تنوع نمونه ها نیز زیاد شده و ممکن است سبک اجرای حروف زبان اشاره از فردی با فرد دیگر متفاوت باشد. از طرفی اندازه دست و انگشتان در افراد مختلف نیز با یکدیگر متفاوت بوده و در نتیجه سیستم در حالتی که بایستی برای پنج نفر دسته بندی را انجام دهد، نسبت به زمانی که تنها برای یک نفر این کار را انجام می دهد، با تنوع داده های بیشتری مواجه است و دسته بندی برای آن مشکل تر می شود.

با توجه به این نتایج، پیشنهاد می شود که در محیط های فردی و خانگی، سیستم تنها برای فرد مورد نظر آموزش ببیند تا بالاترین نتیجه ممکن حاصل شود و در سیستم های چند کاربره (با توجه به نتایجی که در مورد تصادفی کردن در ادامه ارائه خواهد شد)، نیز با آموزش دادن فرد تازه وارد به سیستم (تصادفی کردن) می توان درصد بسیار بالایی از شناسایی را تضمین نمود.

۴-۳-۲-تاثیر تغییر تعداد نمونه ها

همانند حالت قبلی، تاثیر افزایش تعداد نمونه ها در صحت دسته بند سیستم توسط یک نمودار در شکل ۴-۲ نمایش داده شده است. درصد های ارائه شده در شکل در حالت اعتبارسنجی با ۲۵- تا کردن و تصادفی کردن داده ها بدست آمده است. همانطور که مشاهده می شود، با افزایش تعداد نمونه ها، درصد صحت خروجی های سیستم نیز کاهش می یابد. با افزایش تعداد داده ها، به دلیل افزایش نمونه های آموزشی، این دید به وجود می آید که درصد صحت نیز بایستی افزایش یابد ولی در واقع با افزایش نمونه های آموزشی و آزمایشی، داده های متنوع تری به سیستم اعمال می شوند که پوشش این تنوع جهت شناسایی کاری دشوارتر از حالتی است که تنها نمونه های محدودی به سیستم وارد می شوند.



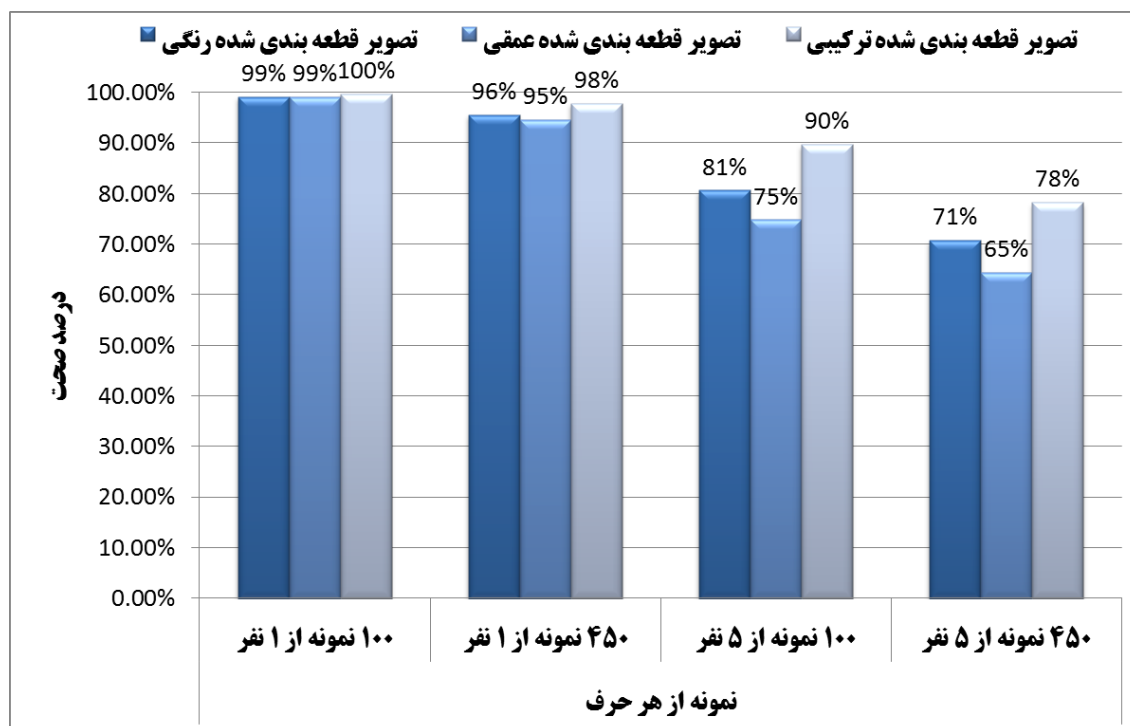
شکل ۲-۴- نمودار تاثیر افزایش تعداد نمونه به ازای هر حرف در صحت خروجی سیستم.

۳-۳-۴- تاثیر تغییر تصویر قطعه بندی شده

همانطور که قبلا نیز گفته شد، در هنگام اعمال ماسک عمقی بر روی تصویر رنگی، در صورت تعویض تصویر رنگی با تصویر عمقی اولیه، و یا ادغام ویژگی های بدست آمده از آن ها، نتایج متفاوتی حاصل می شود که این نتایج در جدول ۳-۴ قرار داده شده اند. در این قسمت به بررسی تاثیر تعویض تصویر قطعه بندی شده می پردازیم.

اولین مورد قابل توجه این است که تمامی بهترین نتایج بدست آمده به ازای تنوع افراد و نمونه های اولیه، از ترکیب ویژگی های رنگی و عمقی بدست آمده است که نشان دهنده قدرتمند تر شدن دسته بند با استفاده از ترکیب ویژگی ها می باشد. البته بایستی کاهش کارایی نسبی سیستم به دلیل پردازش ۵۲ ویژگی در برابر ۲۶ ویژگی مخصوصا در هنگام محاسبه LDA را نیز در نظر گرفت. در نمودار شکل ۳-۴ می توان دید بهتری نسبت به تاثیر تغییر تصویر قطعه بندی شده پیدا نمود. درصد های ارائه شده در شکل به صورت گرد شده و در حالت اعتبارسنجی با ۲۵- تا کردن و تصادفی

کردن داده ها بدست آمده است. همانطور که در شکل ۳-۴ مشخص است، تقریباً به ازای تمامی انتخاب ها از پایگاه داده، استفاده از تصویر قطعه بندی شده رنگی نتیجه بهتری از تصویر قطعه بندی شده عمقی به ما می دهد و همچنین در تمامی حالات استفاده از ترکیب ویژگی های تصویر قطعه-بندی شده رنگی و عمقی، بهترین نتیجه را ارائه نموده است.

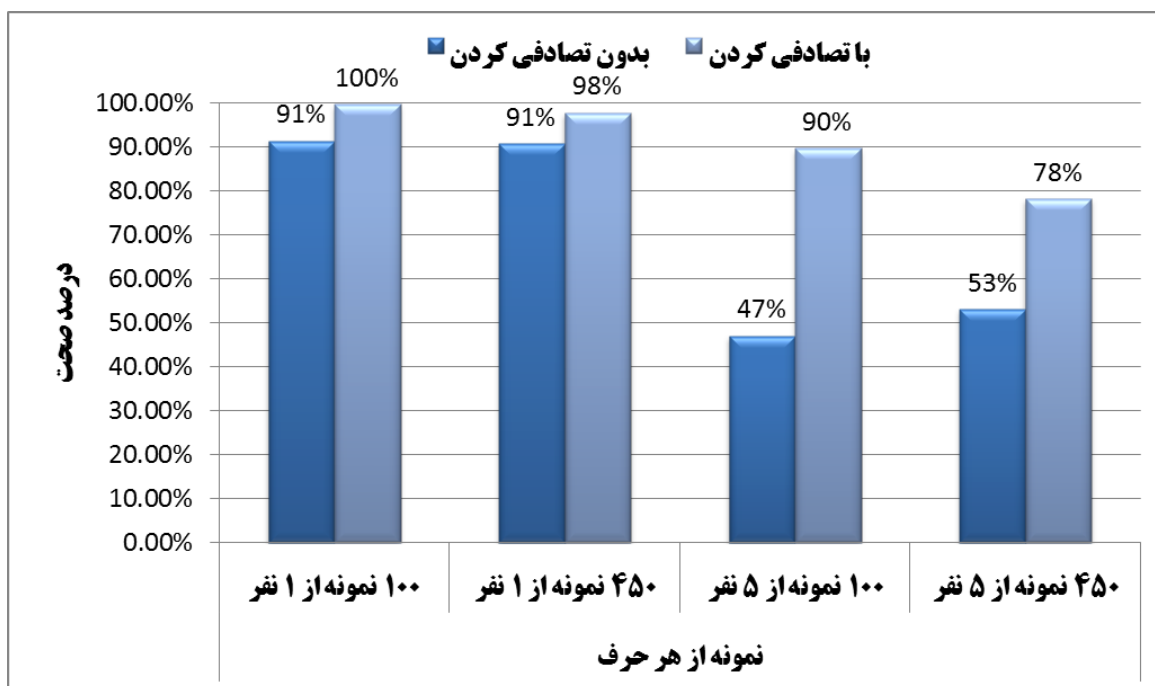


شکل ۳-۴- نمودار مقایسه تاثیر تغییر تصویر قطعه بندی شده در صحت خروجی های سیستم به ازای انتخاب های متنوع از پایگاه داده.

۴-۳-۴- تاثیر تصادفی کردن داده ها در نتایج

همانطور که انتظار می رفت، با تصادفی کردن داده ها به دلیل توزیع یکنواخت داده های آموزشی و آزمایشی در کل داده ها، نتایج بهتری حاصل شده است. در تمامی حالت ها این بهبود مشاهده می شود و در شرایطی که تعداد افراد بیشتری در داده ها حضور دارند، این بهبود به حداکثر خود می رسد به شکلی که مثلاً در ردیف ۱۷ و ۱۸ جدول ۳-۴ در حالت ۵-تا کردن و ویژگی های ترکیبی، درصد ۴۶.۹۸٪ صحت بدون تصادفی کردن به ۸۹.۷۱٪ با تصادفی کردن داده ها رسیده است که افزایش ۹۱ درصدی در صحت خروجی های سیستم را سبب شده است. در شکل ۴-۴ این بهبود

را برای حالت ۵-تا کردن و ویژگی های ترکیبی به ازای ترکیب های مختلف داده های انتخاب شده مشاهده می شوند که بیانگر تاثیر تصادفی کردن در بهبود درصد شناسایی سیستم است. علت این افزایش مخصوصا در حالت ۵-تا کردن به این دلیل است که در این حالت در هر یک از ۵ بار تا کردن داده ها، دقیقا فردی که داده های آن در آموزش وجود نداشته است به عنوان داده آزمایشی به سیستم داده می شود. این بدین معناست که در صورتی که بتوان قبل از استفاده از سیستم برای افراد جدید، یک سری آموزشی از فرد مورد نظر ضبط شده و در داده ها تاثیر داده شود، تاثیر فوق العاده زیاد و غیر قابل نادیده گرفتن در صحت سیستم خواهد داشت. با نگاهی به جدول ۴-۳ همچنین مشخص می شود که با افزایش تا کردن ها از ۵ به ۱۰ و ۲۵ و بدون تصادفی کردن، درصد بهبود می-یابد که به دلیل وارد شدن تعدادی از داده های فرد مورد آزمایش در داده های آموزشی است که توزیع داده بهتری را حاصل نموده است.



شکل ۴-۴- تاثیر تصادفی کردن داده ها در صحت سیستم پیشنهادی به ازای نمونه داده های انتخابی مختلف.

۴-۴- تحلیل شناسایی هریک از حروف

درصد های ارائه شده همگی درصد های میانگین تمامی حروف می باشند. برای بررسی درصد شناسایی تک تک حروف، از ماتریس درهم ریختگی استفاده می کنیم. برای بهترین درصد شناسایی بر روی کل پایگاه داده که با ویژگی های ترکیبی، تصادفی کردن و ۲۵-تا کردن بدست آمده است، در جدول ۴-۴ ماتریس در هم ریختگی به تفکیک حروف نشان داده شده است. نتایج نشان می دهد که ۲۹٪ از نمونه های حرف g، اشتباهات h و ۲۴٪ از نمونه های حرف m، اشتباهات n تشخیص داده شده است. به همین ترتیب ۲۷٪ از حرف o، اشتباهات e، ۳۳٪ از حرف r، اشتباهات u، ۳۸٪ از حرف t، اشتباهات n و ۳۰٪ از حرف x، اشتباهات d تشخیص داده شده است.

از درصد دقت حروف نیز مشخص شده است که تنها ۴۶٪ از حروفی که به عنوان حرف n شناسایی شده اند، در واقع خود n بوده اند. به همین ترتیب تعداد زیادی از حروف اشتباهاً e، d و m تشخیص داده شده اند.

کمترین درصد یادآوری برای حروف s، t و u با درصد یادآوری ۴۰٪، ۴۴٪ و ۴۰٪ بوده است در حالی که حروف y، u، l، i، h، f، d، c، b و a با درصد یادآوری بالای ۹۰٪ به خوبی شناسایی شده اند که این شناسایی برای حرف y به بالای ۹۷٪ و برای حروف l و b به بالای ۹۵٪ نیز رسیده است.

با بررسی ویژگی های حروفی که در شناسایی به یکدیگر نسبت داده شده اند و تولید ویژگی هایی که بتواند این حروف را از یکدیگر جدا کند، می توان درصد شناسایی سیستم را به خوبی افزایش داد.

جدول ۴-۴- ماتریس در هم ریختگی بهترین نتیجه بر روی کل پایگاه داده مورد استفاده بر مبنای درصد. سطر آخر دقت هر سطر را مشخص می نماید.

کلاس شناسایی شده		a	b	c	d	e	f	g	h	i	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y
کلاس اصلی داده ها	a	90%	0%	0%	0%	0%	0%	1%	1%	0%	0%	0%	0%	3%	1%	0%	0%	0%	2%	0%	0%	0%	0%	0%	0%
	b	0%	95%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	2%	0%	0%	0%	0%
	c	0%	0%	92%	0%	2%	0%	1%	0%	0%	0%	0%	0%	1%	2%	1%	0%	0%	1%	0%	0%	0%	0%	0%	0%
	d	0%	0%	0%	92%	2%	0%	1%	0%	0%	0%	1%	0%	0%	1%	0%	0%	1%	0%	0%	0%	0%	0%	1%	0%
	e	0%	0%	2%	0%	86%	0%	0%	0%	0%	0%	0%	3%	4%	2%	1%	0%	0%	1%	0%	0%	0%	0%	0%	0%
	f	0%	3%	0%	0%	0%	91%	0%	0%	1%	0%	0%	0%	0%	0%	1%	0%	1%	0%	0%	0%	0%	2%	0%	0%
	g	0%	0%	0%	1%	0%	0%	65%	29%	0%	1%	0%	0%	0%	0%	1%	0%	0%	0%	1%	0%	0%	0%	0%	0%
	h	0%	0%	0%	0%	0%	0%	6%	91%	0%	0%	0%	0%	0%	0%	2%	1%	0%	0%	0%	0%	0%	0%	0%	0%
	i	0%	0%	0%	2%	1%	1%	0%	0%	92%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	1%	1%
	k	0%	0%	0%	10%	0%	1%	0%	1%	0%	65%	0%	0%	0%	0%	2%	0%	3%	0%	0%	6%	10%	0%	1%	0%
	l	0%	0%	0%	2%	0%	0%	0%	0%	0%	1%	95%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	m	0%	0%	0%	0%	3%	0%	0%	0%	0%	0%	0%	65%	24%	1%	1%	0%	0%	3%	2%	0%	0%	0%	0%	0%
	n	1%	0%	1%	1%	1%	0%	0%	0%	0%	0%	0%	5%	83%	1%	1%	0%	0%	1%	3%	0%	0%	0%	1%	0%
	o	3%	0%	3%	1%	27%	0%	2%	0%	0%	0%	0%	5%	11%	40%	2%	1%	0%	4%	2%	0%	0%	0%	0%	0%
	p	0%	0%	1%	0%	0%	0%	2%	0%	0%	0%	0%	0%	0%	2%	86%	6%	0%	0%	2%	0%	0%	0%	2%	0%
	q	0%	0%	0%	0%	0%	0%	0%	2%	0%	0%	0%	2%	1%	0%	4%	89%	0%	0%	0%	0%	0%	0%	0%	0%
	r	0%	0%	0%	4%	0%	0%	1%	1%	0%	0%	0%	0%	0%	0%	0%	0%	57%	0%	0%	33%	1%	1%	1%	0%
	s	3%	0%	0%	0%	12%	0%	0%	0%	0%	0%	0%	19%	16%	2%	1%	0%	0%	40%	6%	0%	0%	0%	0%	0%
	t	3%	0%	1%	1%	2%	0%	1%	0%	0%	0%	0%	3%	38%	1%	1%	0%	0%	5%	44%	0%	0%	0%	0%	0%
	u	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	94%	0%	0%	1%	0%
	v	0%	0%	0%	0%	0%	1%	0%	0%	0%	2%	0%	0%	0%	0%	1%	0%	2%	0%	0%	2%	85%	4%	0%	1%
	w	0%	1%	0%	1%	0%	2%	0%	0%	0%	2%	0%	0%	0%	0%	1%	0%	0%	0%	0%	1%	4%	87%	0%	1%
	x	0%	0%	0%	30%	0%	1%	1%	0%	0%	1%	0%	0%	1%	0%	3%	0%	2%	0%	1%	1%	0%	0%	60%	0%
	y	0%	0%	0%	0%	0%	0%	1%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	97%
دقت شناسایی حروف		89%	95%	91%	63%	63%	93%	78%	71%	96%	90%	98%	63%	46%	74%	78%	90%	84%	68%	72%	68%	84%	91%	85%	95%

۴-۵- مقایسه با روش های پیشین

در این قسمت روش پیشنهادی را با تعدادی از روش های پیشین مقایسه خواهیم نمود. یکی از مهمترین کارهایی که این روش بایستی با آن مقایسه شود، کار [۸] است که در ابتدا نتایجی را بر روی این پایگاه داده ارائه نمود. مقایسه کار با نتایج ارائه شده همزمان با ارائه اینترنتی پایگاه داده فوق در [۴۵] به چند دلیل مناسب تر است. اولین دلیل این است که مقاله اول بر روی ۴ اجرا از ۵ اجرا پایگاه داده انجام گرفته است ولی نتیجه ارائه شده در وب سایت دانشگاهی نویسنده بر روی تمامی ۵ اجرا صورت گرفته است. همچنین در مقاله ابهاماتی وجود دارد که پاسخ داده نشده است، نظیر عدم تطابق نتایج ارائه شده با نتایج بدست آمده از ماتریس در هم ریختگی مقاله و همچنین ابهام در تعداد نمونه های استفاده شده که یک بار ۵۰۰ و در جایی دیگر ۱۰۰ عدد عنوان شده است. از طرف دیگر نتایج موجود در وبسایت جدید تر از نتایج مقاله و مستقیماً به پایگاه داده استفاده شده مربوط است. نتایج ارائه شده در [۴۵] بر روی داده های قطعه بندی شده تصویر رنگی، تصویر عمقی و ترکیب ویژگی های این دو و به صورت اعتبارسنجی با ۵-تاکردن بدون تصادفی کردن ارائه شده است و تحلیل بیشتری بر روی دیگر ترکیبات از پایگاه داده و یا اعتبارسنجی هایی با n-تاکردن های متفاوت داده نشده است. نتایج روش فوق در ردیف ۱۹، ۲۱ و ۲۳ جدول ۴-۳ در بخش اعتبارسنجی با ۵-تاکردن، بدون استفاده از تصادفی کردن قابل مشاهده است. نتایج مقایسه روش فعلی و روش ارائه شده در [۴۵] در جدول ۴-۵ آمده است. همانطور که مشاهده می شود، روش پیشنهادی در استفاده از تصویر قطعه بندی شده رنگی و استفاده از ویژگی های ترکیبی با ۵۳.۱۳٪ صحت بهتر از روش ارائه شده در [۴۵] عمل کرده است و تنها در درصد صحت بدست آمده از تصویر قطعه بندی شده عمقی نتیجه کمتری حاصل شده است. تفاوت روش پیشنهادی با روش فوق، قطعه بندی دست با استفاده از داده های عمقی و سپس استخراج ویژگی متناسب با آن بوده است در حالی که در روش فوق تصاویر بدون قطعه بندی به مرحله استخراج ویژگی ارسال شده بودند. نتایج این جدول همچنین نشان دهنده

این است که اگر داده ها با تصادفی کردن همراه باشد و یا به عبارتی سیستم قبل از استفاده توسط اشخاص جدید، مقداری توسط آن ها آموزش ببیند، نتایج با جهشی غیر قابل چشم پوشی بهبود می- یابد و به صحت ۷۸.۲۷٪ می توان دست پیدا نمود. اگر چه ممکن است به نظر رسد که در حالت مقایسه مستقیم، حدود ۴ درصد بهبود روش پیشین، افزایش قابل توجهی نباشد ولی این افزایش در واقع به معنای شناسایی صحیح به طور میانگین ۴۴۶ نمونه بیشتر از بین ۱۰۸۰۰ نمونه آزمایشی می- باشد و از طرفی نیز ارائه بهبود نتیجه سیستم در استفاده از تصادفی کردن داده ها توانسته است دید مثبتی برای استفاده کاربردی این سیستم به صورت مطمئن تر ارائه دهد.

جدول ۴-۵-مقایسه روش پیشنهادی و روش ارائه شده توسط پوگالت و بُودِن [45].

روش/تصویر قطعه بندی	تصویر رنگی	تصویر عمقی	ترکیب ویژگی ها	بهترین نتیجه
پوگالت و بُودِن [45]	۳۵٪	۴۹٪	۴۷٪	۴۹٪
روش پیشنهادی (بدون تصادفی کردن)	۵۰.۰۲٪	۴۷.۳۶٪	۵۳.۱۳٪	۵۳.۱۳٪
روش پیشنهادی (با تصادفی کردن)	۷۰.۷۱٪	۶۴.۴۹٪	۷۸.۲۷٪	۷۸.۲۷٪

علاوه بر روش فوق، روش های دیگری نیز هستند که روش فعلی را می توان با آن ها مقایسه نمود. به دلیل اینکه این روش ها از لحاظ پایگاه داده، نمونه های بدست آمده و تعداد داده های مورد استفاده ممکن است متفاوت باشند، تعدادی از داده متناسب با تعداد داده مورد استفاده در این سیستم ها انتخاب و خروجی سیستم برای آن ها نیز محاسبه گردید. نتیجه این مقایسه ها در جدول ۴-۶ آمده است. همانطور که در این جدول نیز مشخص است، روش پیشنهادی حتی با داشتن پایگاه داده ای پیچیده تر با تنوع داده ها و پس زمینه ها، از تمامی روش های دیگر نتیجه بهتری را کسب کرده است.

جدول ۴-۶- جدول مقایسه روش پیشنهادی با روش های پیشین.

روش	ابزار مورد استفاده	دامنه شناسایی	تعداد کلاس ها	تعداد افراد	نمونه هر کلاس- فرد	کل نمونه ها	بهترین درصد شناسایی	مزیت ها	معایب
[۳۷]	دستکش سنسوری	هجی انگشتی حروف الفبای ASL	۲۸	۱۰	۵	۱۴۰۰	۸۲.۹۳٪	تعداد بیشتر کلاس و افراد	استفاده از دستکش
[۶۷]	دوربین رنگی	اشاره های دست	۱۲	N/A	N/A	۱۰۰۰ (۶۰۴ پس زمینه ساده، ۳۳۸ پس زمینه پیچیده)	۹۲.۹٪ پس زمینه ساده و ۸۵.۵٪ پس زمینه پیچیده	پس زمینه های پیچیده، عدم نیاز به دستکش	تعداد کلاس های محدود
[۶۸]	دستکش سنسوری	اشاره های انگشتی	۳۲	N/A	۲۰	۶۴۰	۷۲.۲٪	تعداد بالای کلاس ها	استفاده از دستکش
[۶۹]	دستکش رنگی	حروف هجی انگشتی اسپانیایی	۱۹	N/A	N/A	نامشخص	۹۰.۵۷٪	-	استفاده از دستکش و حذف دو کلاس با شباهت ظاهری از نتایج
[۷۰]	دوربین رنگی	حروف هجی انگشتی کرواسی	۲۱	N/A	۶۰	۱۲۶۰	۹۳.۱۴٪	استفاده از یک ویژگی، عدم نیاز به دستکش	پس زمینه بسیار ساده
روش پیشنهادی	کینکت	حروف هجی انگشتی ASL	۲۴	۵	۱۲	۱۴۴۰	۹۷.۵۷٪ (بردار ترکیبی)	پس زمینه پیچیده، تنوع نمونه ها، عدم نیاز به دستکش	استفاده از کینکت

فصل ۵: اعتبار سنجی^۱

و تحلیل داده ها

¹ Validation

در این فصل اعتبار برخی از فرضیات و پارامترهای موجود و همچنین داده های استفاده شده در سیستم مورد ارزیابی قرار می گیرد. از آنجا که بررسی تمامی فرضیات و پارامترها و بیان نتایج آن ها در این گزارش نمی گنجد، برخی از مهمترین آن ها مورد بررسی قرار خواهند گرفت.

۵-۱-۱- اعتبارسنجی مقاوم بودن^۱ استفاده از عمق جهت قطعه

بندی دست

در این گزارش فرض شده است که یکی از بهترین روش های قطعه بندی دست، استفاده از دوربین عمقی سنسور کینکت می باشد. این فرض را با آزمایشی بر روی مجموعه ای محدود از اشاره ها که توسط سنسور کینکت جمع آوری شده اند بررسی می نماییم و سپس با تحقیقی مشابه در این زمینه مقایسه می نماییم.

۵-۱-۱-۱- داده های مورد آزمایش

برای تست سیستم مورد نظر، چهار اشاره اولیه معمول در سیستم های مختلف نوشتاری زبان اشاره مورد استفاده قرار گرفته است. این اشارات با نماد معادل آن ها در سیستم SiGML و HamNoSys و فونت یونیکد HamNoSys در جدول ۵-۱ قرار داده شده اند. برای تست سیستم، از یک دستگاه کینکت استفاده شده و هریک از چهار اشاره جدول ۵-۱، نه بار در مکان های متفاوتی از صحنه اجرا و در قالب ویدئوهایی از سنسور عمقی کینکت، با تفکیک پذیری ۶۴۰ در ۴۸۰ پیکسل و سرعت ۳۰ فریم بر ثانیه ضبط گردیده است. این نه (۹) مجموعه داده تمامی حالات حرکت های فوق در قسمت های مختلف صفحه را شامل شده تا پایگاه داده جامعی حاصل شود. کل نمونه های پایگاه داده حاصل، شامل ۳۶ ویدئو می باشد.

¹ Robustness

جدول ۱-۵- حرکات شناسایی شده در سیستم پیشنهادی و نمادهای آن ها در سیستم های نوشتاری مختلف.

نماد های معادل			اشاره
سیستم SiGML	یونیکد HamNoSys	نماد HamNoSys	
handmover	E082	→	دست به راست
handmovel	E086	←	دست به چپ
Handmoveu	E080	↑	دست به بالا
handmoved	E084	↓	دست به پایین

۵-۱-۲- پیش پردازش

برای هریک از ۳۶ ویدئو بدست آمده در مرحله قبل، فریم ابتدایی و فریم انتهایی آن استخراج شده است. سپس با استفاده از آستانه گیری عمقی در بازه بین ۱۱۲ و ۱۲۸ (۱۶ پیکسل) بر روی تصاویر بدست آمده، دست ها از تصویر عمقی جدا و دیگر قسمت های تصویر به رنگ سیاه در آمده اند. این آستانه گیری به این دلیل بوده است که دست ها تقریباً در مکان ثابتی از دوربین قرار گرفته بوده اند. در شکل ۱-۵، تصویری از فریم ابتدایی (تصویر الف) و فریم انتهایی (تصویر ب) حرکت دست به راست را پس از پیش پردازش و بدست آوردن مرکز ثقل تصویر مشاهده می نمایید.



شکل ۵-۱- تصویری از فریم ابتدای (الف) و انتهایی (ب) حرکت دست به راست پس از پیش پردازش و استخراج مرکز ثقل تصویر.

۵-۱-۳- استخراج و تحلیل ویژگی ها

در این قسمت از تصاویر بدست آمده از قبل، با استفاده از Moment ها ویژگی های مورد نیاز برای دسته بندی استخراج گردیده است. با استخراج $\bar{x}$ و $\bar{y}$ از تصاویر بدست آمده از مرحله پیش پردازش، مرکز ثقل تصویر محاسبه گردیده است. به ازای هر نمونه از ویدئو ها، این عمل دو بار برای فریم های ابتدایی و انتهایی انجام گرفته است.

همانطور که گفته شد، نه مجموعه مختلف از داده ها با نقاط شروع و پایان متفاوت در پایگاه داده وجود دارد. در شکل ۵-۲ نمودار پراکندگی^۱ نقاط شروع ۳۶ نمونه از داده ها و در شکل ۵-۳ نمودار پراکندگی نقاط پایانی حرکت این نمونه ها را مشاهده می نمایید. همانطور که مشخص است به دلیل متفاوت بودن نقاط شروع و پایان، جداسازی این داده ها به راحتی میسر نمی باشد. به همین دلیل ویژگی جدیدی از داده های فوق به صورت زیر استخراج شده است:

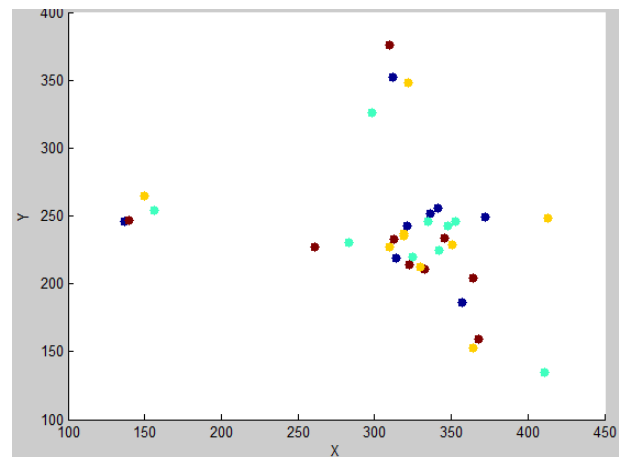
$$(X, Y)_{CoG-Movement Vector} = (X_{CoG-End} - X_{CoG-Start}, Y_{CoG-End} - Y_{CoG-Start}) \quad (۵-۱)$$

که در آن $X_{CoG-Start}$ و $Y_{CoG-Start}$ به ترتیب نقطه x و y مختصات مرکز ثقل فریم ابتدایی،

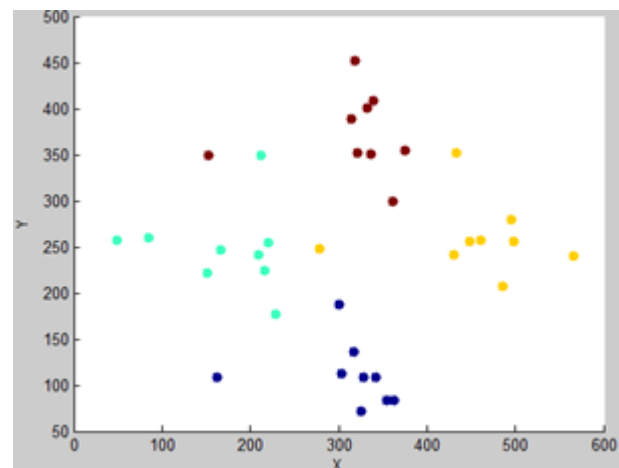
$X_{CoG-End}$ و $Y_{CoG-End}$ به ترتیب نقطه x و y مختصات مرکز ثقل فریم انتهایی و $(X, Y)_{CoG-Movement Vector}$

¹ Scatter Diagram

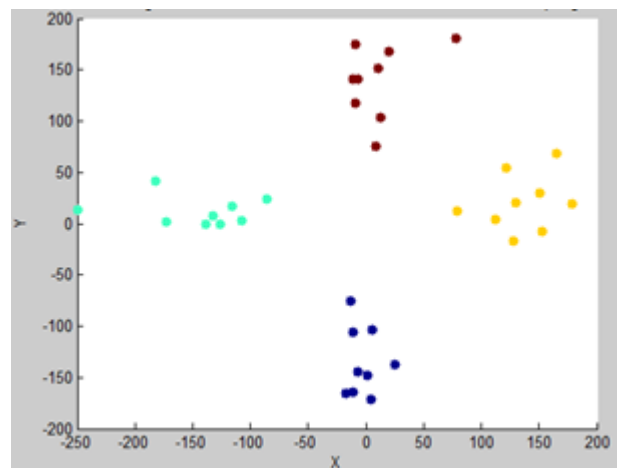
نشان دهنده اندازه حرکت دوبعدی نقطه مرکز ثقل تصویر از فریم ابتدایی تا فریم انتهایی است. در شکل ۴-۵ نمودار پراکندگی ویژگی جدید یا بردار حرکت مرکز ثقل را مشاهده می نمایید. همانطور که مشاهده می شود، داده های حرکات مختلف به صورت ایده آل از یکدیگر جدا شده و در بخش های خاصی از صفحه در خوشه هایی قرار گرفته اند که برای دسته بندی می توانند استفاده شوند.



شکل ۵-۲-نمودار پراکندگی مرکز ثقل فریم های شروع حرکات مختلف پایگاه داده. رنگ های موجود در تصویر قرمز، آبی، زرد و سبز آبی هستند که به ترتیب نشان دهنده کلاس های حرکت به بالا، پایین، راست و چپ می باشند.



شکل ۵-۳-نمودار پراکندگی مرکز ثقل فریم های پایانی حرکات مختلف پایگاه داده. رنگ های موجود در تصویر قرمز، آبی، زرد و سبز آبی هستند که به ترتیب نشان دهنده کلاس های حرکت به بالا، پایین، راست و چپ می باشند.



شکل ۴-۵- نمودار پراکندگی اندازه بردار جابجایی مرکز ثقل فریم های شروع و پایانی حرکات مختلف پایگاه داده. رنگ های موجود در تصویر قرمز، آبی، زرد و سبز آبی هستند که به ترتیب نشان دهنده کلاس های حرکت به بالا، پایین، راست و چپ می باشند.

۵-۱-۴- دسته بندی داده ها

با توجه به نمودار پراکندگی شکل ۴-۵، به علت مجزا شدن تمامی کلاس ها از یکدیگر، یکی از ساده ترین و درعین حال کارآمدترین دسته بندها یعنی k -نزدیکترین همسایه می تواند مورد استفاده قرار گیرد. در این قسمت با قرار دادن پنج مجموعه از نه مجموعه (۲۰ نمونه) برای آموزش سیستم و ۴ مجموعه از نه مجموعه (۱۶ نمونه) برای آزمایش سیستم، از ۱-نزدیکترین همسایه برای دسته بندی داده ها استفاده شده است.

۵-۱-۵- نتایج و مقایسه با روش های پیشین

آزمایش سیستم با استفاده از تمامی ترکیب های مختلف از داده های آموزشی و آزمایشی صورت گرفته است. جایگشت^۱ انتخاب ۵ مجموعه آموزشی از ۹ مجموعه آموزشی برابر ۱۲۶ حالت است. تمامی این ۱۲۶ حالت برای دسته بندی سیستم آزمایش شده است و در تمامی حالت ها، همانطور که از نمودار پراکندگی انتظار می رفت، داده ها با ۱۰۰٪ صحت شناسایی شده اند.

¹ Permutation

نتایج این تحقیق در مقایسه با تحقیقی مشابه [۷۱] که با دوربین رنگی اقدام به شناسایی چهار اشاره دست در جهت های بالا، پایین، راست و چپ نموده است و در سال ۲۰۱۱ ارائه شده است در جدول ۲-۵ با کار فعلی مقایسه شده است. نتایج نشان دهنده این است که روش پیشنهادی با استفاده قطعه بندی عمقی دست در تمامی قسمت ها بهتر از روش [۷۱] عمل نموده است. لازم به ذکر است که روش پیشنهادی حرکت اشاره ها و روش مقایسه شده شکل ایستای این اشاره ها را شناسایی نموده است.

جدول ۲-۵-مقایسه روش ارائه شده و یکی از روش های پیشین.

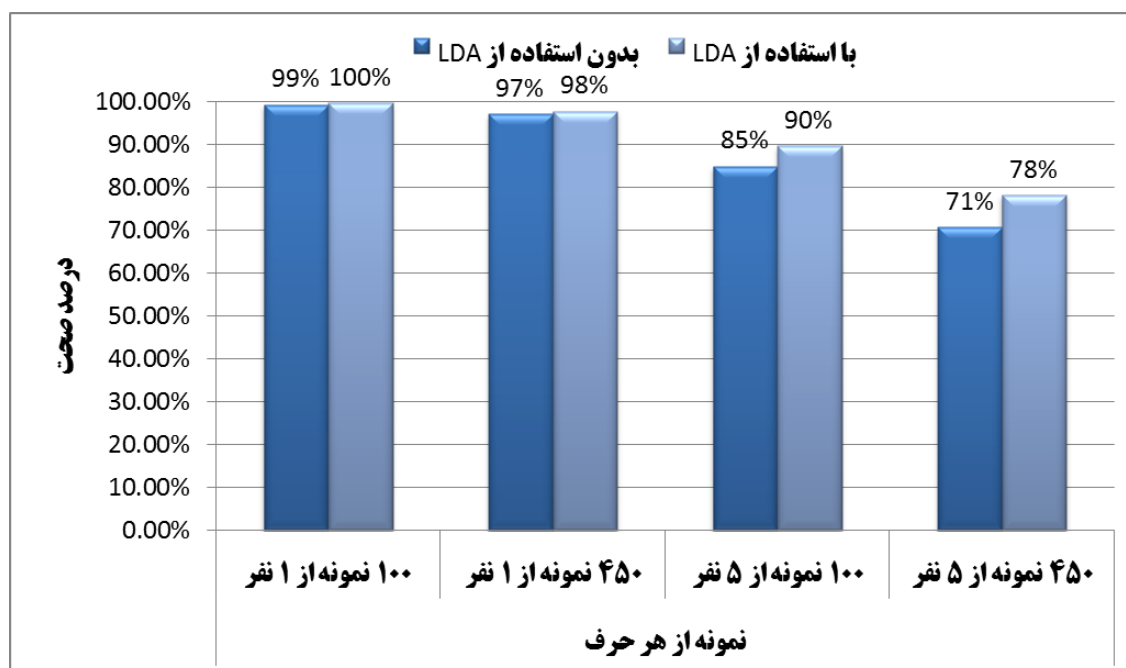
درصد یادآوری شناسایی حرکات		اشاره
روش [۷۱]	روش پیشنهادی	
٪۹۴	٪۱۰۰	دست به راست
٪۹۲	٪۱۰۰	دست به چپ
٪۹۲	٪۱۰۰	دست به بالا
٪۹۵	٪۱۰۰	دست به پایین

۵-۱-۶-نتیجه گیری

نتایج حاصل از این سیستم نشان داده است که استفاده از قطعه بندی عمقی اشاره های دست می تواند نتایج بسیار مقاومی را ایجاد نماید. این روش در مقایسه با روش مشابهی که از دوربین رنگی برای شناسایی این حرکات استفاده نموده بود برتری خود را نشان داده و به نظر می رسد که استفاده از قطعه بندی عمقی دست با استفاده از سنسور کینکت گزینه بسیار مناسبی برای پایه روش پیشنهادی کلی این سیستم بوده است.

۵-۲- تحلیل استفاده از LDA

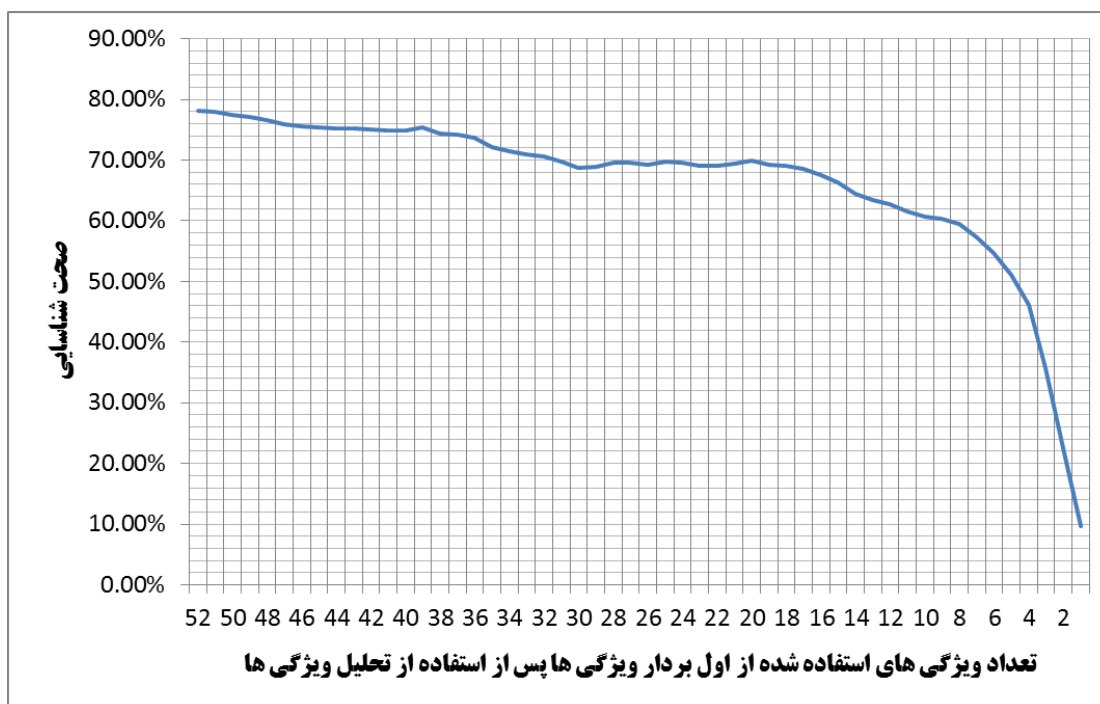
پس از استخراج ویژگی ها و قبل از دسته بندی داده ها، از LDA برای تحلیل بردار ویژگی ها استفاده شده است ولی پس از اعمال ضرایب بدست آمده از LDA، هیچ ویژگی ای از سیستم حذف نشده است تا حداکثر توان شناسایی سیستم برآورد شود. در این قسمت تاثیر استفاده و یا عدم استفاده از LDA را در سیستم بررسی می کنیم. جهت این کار، برای حالت های ترکیبی و با تصادفی کردن سیستم و به ازای انتخاب های مختلف از پایگاه داده، نتایج صحت شناسایی سیستم با استفاده از ۲۵- تا کردن در نمودار شکل ۵-۵ نشان داده شده است. همانطور که مشاهده می شود، استفاده از LDA برای تحلیل بردار ویژگی ها، در تمامی حالات و انتخاب های سیستم، نتایج بهتری را حاصل نموده است. پس نتیجه می گیریم که انتخاب تحلیل ویژگی ها با استفاده از LDA توانسته است کارایی سیستم را بهبود دهد.



شکل ۵-۵- نمودار تاثیر استفاده از LDA در بهبود کارایی سیستم به ازای انتخاب های مختلف از پایگاه داده.

۵-۳- تحلیل انتخاب ویژگی ها با استفاده از LDA

ویژگی های استخراج شده در این تحقیق ۲۶ ویژگی برای حالت های تصویر قطعه بندی شده رنگی و یا عمقی و ۵۲ ویژگی در حالت ترکیبی بوده است. همانطور که قبلاً اشاره شد، از LDA برای تحلیل ویژگی ها و استخراج ماتریس ضرایب برای جداسازی داده ها در فضای ویژگی ها استفاده شده است ولی جهت استفاده از حداکثر توان سیستم، ویژگی هایی از سیستم کاسته نشده است. در این قسمت جهت تحلیل بیشتر و استفاده در آینده، در بهترین حالت رسیدن به جواب بر روی کل پایگاه داده یعنی استفاده از تمامی ۵ اجرا کننده و ۴۵۰ نمونه به ازای هر حرف و فرد و در کل ۵۴۰۰۰ داده با ۵۲ ویژگی ترکیبی به همراه تصادفی کردن، تحلیل کاهش تعداد ویژگی ها صورت خواهد گرفت. از ۵-تاکردن برای این بخش استفاده شده است و درصد صحت خروجی سیستم به ازای کم کردن صفر ویژگی (بردار اولیه ویژگی ها) تا کم کردن ۵۱ ویژگی (تست سیستم با تنها یک ویژگی) در نمودار شکل ۵-۶ نشان داده شده است. همانطور که از نمودار نیز مشخص است، در حالت ترکیبی، در صورت کاهش ۱۳ ویژگی از انتهای بردار مرتب شده ویژگی های حاصل از تحلیل LDA، تغییر قابل ملاحظه ای در درصد شناسایی سیستم رخ نمی دهد (تغییر ۷۸.۲۷٪ به ۷۵.۳۴٪ صحت با ۳۹ ویژگی) و در صورت محدودیت توان محاسباتی سیستم کامپیوتری، می توان این تعداد ویژگی ها را از انتهای بردار حذف نمود.



شکل ۵-۶- نمودار کاهش درصد شناسایی سیستم با کاهش ویژگی‌ها در حالت ویژگی‌های ترکیبی.

۵-۴- تحلیل استقلال بردار ویژگی‌ها

ویژگی‌های مورد استفاده در این سیستم ممکن است به طور خطی از یکدیگر مستقل نباشند و به شکلی به یکدیگر وابسته بوده و در نتیجه استفاده از چند ویژگی وابسته به یکدیگر ممکن است بار محاسباتی سیستم را بی جهت زیاد نموده و کارایی سیستم کاهش یابد. جهت تحلیل بیشتر و استفاده در کارهای آینده، در این قسمت همبستگی^۱ ویژگی‌های استفاده شده بدست آمده از تصویر قطعه بندی شده رنگی و کل پایگاه داده (ردیف ۲۱ و ۲۲ جدول ۴-۳) جهت تحلیل وابستگی مورد تحلیل قرار گرفته و جدول همبستگی آن‌ها نسبت به یکدیگر بدست آمده است. جدول ۵-۳ همبستگی ویژگی‌های مورد استفاده در سیستم را نشان می‌دهد که با استفاده از دستور corr در نرم افزار Matlab بر روی بردار ویژگی‌ها بدست آمده است. اعدادی که در جدول به ۱ و یا ۱- نزدیکتر باشند (در جدول قدر مطلق اعداد نوشته شده است)، همبستگی بیشتر داده‌ها را نشان می‌دهند. اعداد نزدیک به یک با رنگ تیره تر نشان داده شده‌اند.

¹ Correlation

جدول ۵-۳- جدول همبستگی ویژگی های استخراج شده از تصویر قطعه بندی شده رنگی دست.

ویژگی	Hu1	Hu2	Hu3	Hu4	Hu5	Hu6	Hu7	s00	u20	u02	u30	u03	u21	u12	u11	E	J1	J2	O	C	NF	ND	DLS	DLE	DLR	DD
Hu1	1.00	0.83	0.14	0.07	0.04	0.05	0.04	0.75	0.32	0.44	0.00	0.04	0.05	0.20	0.35	0.40	0.68	0.54	0.00	0.12	0.13	0.02	0.04	0.09	0.07	0.03
Hu2	0.83	1.00	0.14	0.13	0.11	0.12	0.11	0.65	0.26	0.40	0.01	0.04	0.13	0.11	0.39	0.56	0.53	0.45	0.01	0.09	0.23	0.12	0.13	0.04	0.10	0.18
Hu3	0.14	0.14	1.00	0.98	0.98	0.98	0.98	0.11	0.01	0.07	0.01	0.02	0.01	0.06	0.02	0.01	0.06	0.04	0.04	0.04	0.01	0.05	0.05	0.06	0.01	0.04
Hu4	0.07	0.13	0.98	1.00	1.00	1.00	1.00	0.07	0.01	0.02	0.02	0.02	0.02	0.02	0.01	0.00	0.03	0.02	0.01	0.01	0.02	0.01	0.01	0.01	0.00	0.01
Hu5	0.04	0.11	0.98	1.00	1.00	1.00	1.00	0.05	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.01
Hu6	0.05	0.12	0.98	1.00	1.00	1.00	1.00	0.06	0.01	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.01
Hu7	0.04	0.11	0.98	1.00	1.00	1.00	1.00	0.05	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.00
s00	0.75	0.65	0.14	0.07	0.05	0.06	0.05	1.00	0.64	0.80	0.06	0.26	0.05	0.02	0.00	0.51	0.86	0.71	0.10	0.28	0.07	0.04	0.09	0.13	0.06	0.04
u20	0.32	0.26	0.14	0.13	0.11	0.12	0.01	0.64	1.00	0.36	0.21	0.29	0.30	0.05	0.13	0.30	0.71	0.70	0.35	0.04	0.00	0.06	0.03	0.04	0.03	0.07
u02	0.44	0.40	0.14	0.07	0.04	0.05	0.01	0.80	0.36	1.00	0.03	0.33	0.01	0.03	0.02	0.52	0.75	0.69	0.08	0.40	0.00	0.13	0.18	0.23	0.09	0.17
u30	0.00	0.00	0.01	0.02	0.01	0.02	0.01	0.03	0.21	0.03	1.00	0.06	0.48	0.29	0.12	0.30	0.01	0.19	0.17	0.09	0.16	0.00	0.06	0.11	0.05	0.04
u03	0.04	0.11	0.01	0.02	0.00	0.01	0.00	0.26	0.29	0.33	0.06	1.00	0.23	0.37	0.25	0.35	0.12	0.24	0.52	0.24	0.05	0.10	0.10	0.09	0.01	0.15
u21	0.05	0.13	0.01	0.02	0.00	0.01	0.00	0.48	0.21	0.03	0.23	0.06	1.00	0.05	0.07	0.07	0.09	0.12	0.13	0.07	0.28	0.11	0.19	0.01	0.15	0.10
u12	0.20	0.11	0.06	0.02	0.05	0.04	0.00	0.05	0.03	0.33	0.03	0.37	0.05	1.00	0.46	0.18	0.10	0.19	0.23	0.05	0.03	0.13	0.03	0.03	0.01	0.10
u11	0.35	0.39	0.02	0.01	0.04	0.03	0.00	0.46	0.14	0.25	0.12	0.25	0.14	0.46	1.00	0.58	0.15	0.38	0.16	0.05	0.20	0.12	0.10	0.03	0.07	0.13
E	0.40	0.56	0.01	0.00	0.01	0.01	0.00	0.58	0.35	0.52	0.30	0.35	0.07	0.18	0.58	1.00	0.44	0.63	0.09	0.26	0.11	0.04	0.01	0.10	0.11	0.02
J1	0.68	0.53	0.06	0.03	0.02	0.08	0.08	0.44	0.12	0.75	0.01	0.12	0.09	0.10	0.15	0.44	1.00	0.89	0.08	0.20	0.08	0.02	0.08	0.10	0.04	0.02
J2	0.54	0.45	0.04	0.01	0.01	0.19	0.11	0.63	0.24	0.69	0.19	0.24	0.12	0.19	0.38	0.63	0.89	1.00	0.05	0.10	0.07	0.06	0.10	0.05	0.02	0.02
O	0.00	0.01	0.04	0.04	0.01	0.03	1.00	0.11	0.08	0.09	0.16	0.13	0.52	0.23	0.16	0.09	0.08	0.35	0.10	0.03	0.03	0.03	0.00	0.01	0.02	0.02
C	0.12	0.09	0.04	0.01	0.05	0.03	0.03	0.26	0.04	0.40	0.09	0.24	0.07	0.05	0.05	0.26	0.20	0.19	0.03	1.00	0.42	0.43	0.63	0.65	0.10	0.76
NF	0.13	0.23	0.04	0.12	0.11	0.07	0.03	0.07	0.01	0.00	0.16	0.05	0.28	0.03	0.20	0.11	0.08	0.07	0.03	0.42	1.00	0.28	0.38	0.42	0.14	0.50
ND	0.02	0.12	1.00	0.98	0.98	0.98	0.98	0.04	0.06	0.13	0.00	0.10	0.11	0.13	0.12	0.04	0.02	0.03	0.03	0.43	0.28	1.00	0.68	0.67	0.12	0.63
DLS	0.04	0.13	0.05	0.01	0.01	0.01	0.01	0.09	0.03	0.18	0.06	0.10	0.19	0.03	0.10	0.01	0.08	0.01	0.00	0.63	0.38	0.68	1.00	0.73	0.21	0.85
DLE	0.09	0.04	0.06	0.01	0.01	0.01	0.01	0.13	0.04	0.23	0.11	0.09	0.18	0.03	0.03	0.10	0.10	0.10	0.01	0.65	0.42	0.67	0.73	1.00	0.43	0.84
DLR	0.07	0.10	0.00	0.00	0.00	0.00	0.00	0.06	0.03	0.09	0.05	0.01	0.01	0.01	0.07	0.11	0.04	0.05	0.02	0.10	0.14	0.12	0.21	1.00	0.08	1.00
DD	0.03	0.18	0.04	0.07	0.04	0.05	0.04	0.17	0.04	0.17	0.04	0.15	0.15	0.10	0.13	0.02	0.02	0.02	0.02	0.76	0.50	0.63	0.85	0.84	0.08	1.00

همانطور که مشاهده می شود، بیشترین وابستگی ویژگی ها به هم، مربوط به ویژگی های Hu می باشد. ویژگی های Hu1 و Hu2 با یکدیگر و ویژگی های Hu3 تا Hu7 به شدت به یکدیگر وابسته هستند. همچنین میانگین عمق چاله ها به تعداد انگشتان و میانگین طول بردار ورودی و خروجی چاله ها وابستگی شدید دارد. J1 و J2 نیز به Moment های $1/\sqrt{m00}$ و u02 و u20 وابستگی نسبی دارند. این وابستگی ها نشان می دهد که تاثیر این ویژگی ها در کنار یکدیگر ممکن است زیاد نباشد و استفاده از ویژگی هایی که از یکدیگر مستقل تر هستند و جداکننده ی مشابه آن ها در بردار ویژگی ها وجود ندارد، اولویت بیشتری دارد.

۵-۵- سرعت پردازش روش پیشنهادی

زمان لازم برای انجام پردازش و دسته بندی سیستم در این بخش بررسی می شود. در آزمایشی که در بخشی از سیستم بر روی کل پایگاه داده با ۴۵۰ نمونه از پنج نفر با مجموع ۵۴،۰۰۰ داده بر روی تصویر رنگی انجام گرفت، نتایجی بدست آمد که به طور خلاصه در جدول ۴-۵ مشاهده می نماید. توجه نمایید که در هر دو حالت تک نخه و چند نخه از یک پردازنده لپ تاپ Intel Corei7 2630QM با سرعت 2 GHz که دارای ۴ هسته سخت افزاری که هر هسته ۲ نخ سخت افزاری دارد و ۶ گیگابایت رم استفاده شده است. در حالت یک نخه از یکی از ۸ نخ سخت افزاری این پردازنده و در حالت چند نخه، از هر ۸ نخ سخت افزاری سیستم استفاده شده است. زمان هایی در جدول ۴-۵ به صورت پررنگ و با خط زیرین مشخص شده است، مستقیماً توسط سیستم محاسبه شده اند و بقیه زمان ها بدون خط زیرین، توسط داده های هر سطر محاسبه شده اند و محاسباتی به صورت ستونی انجام نگرفته است. همان طور که مشاهده می شود، در حالت توزیع پردازش بر روی نخ های مختلف، سرعت حداکثر تا ۸ برابر (در اینجا حدود ۶.۴ برابر) افزایش یافته است که به دلیل توزیع پردازش بر روی ۸ نخ سخت افزاری و استفاده از ۱۰۰٪ توان پردازشگر می باشد.

جدول ۴-۵- زمان پردازش و ۲ بار دسته بندی در سیستم پیشنهادی برای ۵۴,۰۰۰ نمونه و ۲۶ ویژگی به تفکیک نوع پردازش. زمان های محاسبه شده مستقیم به صورت زیر خطی مشخص و دیگر داده ها به صورت سطری محاسبه شده اند.

تعداد نخ/پارامتر	کل زمان پردازش	پردازش هر نمونه	تعداد فریم بر ثانیه
تک نخه	<u>۱۴۱۴.۸ ثانیه</u>	۲۶.۲ میلی ثانیه	۳۸ فریم بر ثانیه
چند نخه	۲۱۲ ثانیه	<u>۳.۹ میلی ثانیه</u>	۲۴۵ فریم بر ثانیه

همان طور که از داده های بدست آمده مشخص شده است، سیستم توانسته است به صورت بلادرنگ عمل پردازش و دسته بندی را انجام دهد. خواندن اطلاعات از حافظه فیزیکی نیز جزئی از این پردازش بوده است و می توان پیش بینی نمود که با دریافت اطلاعات سنسور کینکت با ۳۰ فریم بر ثانیه، می توان از این سیستم در حالت تک نخه به راحتی استفاده نمود. همچنین در صورت موازی سازی نیز می توان تا ۶ سنسور کینکت را توسط سیستم ارائه شده پوشش داد که نشان از توانایی بالای سیستم ارائه شده می باشد. استفاده از ۲۴۵ فریم بر ثانیه نیز به سادگی با اختصاص دادن نوبتی فریم ها به نخ های سخت افزاری سیستم امکان پذیر است.

نتایج نشان دهنده این است که استخراج ویژگی ها یا دسته بندی با سرعت بسیار زیادی انجام می شود و نیازی به کاهش ویژگی های حاصل از LDA نیز نمی باشد.

فصل ۶: نتیجه گیری

و کارهای آینده

۶-۱- نتیجه گیری

در این تحقیق یک سیستم تشخیص هجی انگشتی حروف الفبای زبان اشاره آمریکایی بر مبنای قطعه بندی عمقی دست از طریق سنسور کینکت ارائه شده است. تعداد ۲۶ ویژگی هندسی از تصاویر قطعه بندی شده دست پس از چندین مرحله پیش پردازش استخراج و پس از تحلیل ویژگی ها توسط LDA، توسط دسته بند بیز دسته بندی شده اند.

ویژگی های استخراج شده شامل ۱۹ ویژگی بر حسب Moment ها و هفت ویژگی بر حسب لبه بیرونی دست قطعه بندی شده می باشند. مراحل پیش پردازش نیز شامل آستانه گیری تصویر عمقی برای قطعه بندی دست و ایجاد ماسک، حذف نویز ماسک بدست آمده و جابجایی و تغییر اندازه آن می باشد.

قطعه بندی تصویر رنگی و عمقی توسط سنسور کینکت و ارائه ویژگی ها و سیستم مناسبی جهت دسته بندی بهینه داده های بدست آمده دستاورد اصلی این تحقیق می باشد. نتایج مختلفی با اعتبارسنجی ضربدری ارائه شده است که برتری این روش را نسبت به شش روش مشابه قبلی نشان می دهد. استفاده از یک پایگاه داده استاندارد با بیش از ۵۴۰۰۰ نمونه در پنج اجرا مختلف توسط چهار فرد، اعتبار نتایج بدست آمده را تضمین نموده است.

نتایج مختلفی بر حسب انتخاب های مختلفی از پایگاه داده برای ۲۴ حرف الفبا ارائه شده است که از ۷۸.۲۹٪ برای کل پایگاه داده شامل پنج اجرا کننده و ۴۵۰ نمونه از حروف به ازای هر فرد تا ۹۹.۷۵٪ برای یک اجرا کننده با ۱۰۰ نمونه برای هر حرف را شامل می شود. بررسی سرعت پردازش ورودی ها بر روی یک پردازنده Core i7 لپ تاپ بلادرنگ بودن سیستم را با بیش از ۳۰ فریم بر ثانیه نشان می دهد.

در تحلیل و اعتبار سنجی داده ها و نتایج نیز مشخص شده است که استفاده از LDA و ویژگی های ترکیبی حاصل از اعمال ماسک عمقی بر روی تصاویر قطعه بندی شده عمقی و رنگی بهترین نتیجه را در سیستم ارائه نموده و در صورت تصادفی کردن داده ها و اعمال آموزش برای هر فرد جدید قبل از استفاده از سیستم، بهبود چشم گیری در نتایج بدست آمده حاصل می شود.

۶-۲-پیشنهادهای برای کارهای آینده

کارهای آینده ای که برای این تحقیق می توان متصور شد، در اول تحلیل خطا برای حروف مختلف جهت بهبود این روش می باشد. با بررسی ویژگی های حروفی که در شناسایی به یکدیگر نسبت داده شده اند و تولید ویژگی هایی که بتواند این حروف را از یکدیگر جدا کند، می توان درصد شناسایی سیستم را به خوبی افزایش داد.

در ادامه استفاده از این روش برای شناسایی کلمات و جملات زبان اشاره می باشد. شناسایی کلمات معمولاً از دو بخش استخراج ویژگی های دست و شناسایی زنجیره ای است که حرکت دست می تواند در آن مدل شود. با استفاده از نتایج بدست آمده در این روش برای استخراج ویژگی های دست و استفاده از اسکلت بدست آمده از کینکت برای مدل سازی زنجیره حرکتی دست ها، می توان نتایج مهمی را در استفاده از این تحقیق متصور شد.

با ارتقاء و پیاده سازی سیستم جهت یک استفاده کاربردی در رایانه، تلفن های همراه و یا سامانه های هوشمند می توان از نتایج این تحقیق برای کارکرد های مختلف تشخیص زبان اشاره و یا تشخیص اشاره ها استفاده نمود تا متقاضیان اصلی این تحقیق که افراد ناشنوا هستند بتوانند از تحقیقات بدست آمده استفاده واقعی را ببرند.

همچنین ارتقاء سیستم جهت کارکرد بر روی زبان های اشاره دیگر و ترجمه زبان های اشاره به یکدیگر نیز می تواند در آینده بر اساس سیستم پیشنهادی انجام شود. تست سیستم برای شناسایی

دیگر زبان های اشاره مانند زبان اشاره فارسی می توانند پلی جهت ارتباط جهانی ناشنویان با یکدیگر باشد.

در صورتی که روش قطعه بندی مناسب تری نیز برای دوربین های رنگی ارائه شده باشد، این روش نیز می تواند بر روی این نوع سیستم ها آزمایش شود تا علاوه بر کارکرد بالا، در محیط هایی که سنسور کینکت و یا دیگر وسائل مشابه ممکن است وجود نداشته باشند، استفاده و عمومیت بیشتری نیز پیدا کند.

با توجه به تحلیل سرعت سیستم نیز، می توان برای این نتیجه گیری استفاده نمود که استخراج ویژگی های بیشتر در صورتی که بتواند درصد شناسایی سیستم را بهبود دهد، می تواند به سیستم اضافه گردد بدون اینکه تاثیر چندانی در سرعت نهایی سیستم داشته باشد.

-
- [1] "Copycat", [cited August 2012]; Available from: <http://www.cats.gatech.edu/content/copycat>.
 - [2] Brashear H. M. (2010), "Improving the Efficacy of Automated Sign Language Practice Tools", PhD Thesis, Georgia Institute of Technology.
 - [3] "Sign Language", [cited August 2012]; Available from: http://en.wikipedia.org/wiki/Sign_language.
 - [4] Zahedi M. (2007), "Robust Appearance-Based Sign Language Recognition", PhD Thesis, RWTH.
 - [5] Prillwitz S. (1989), "Hamnosys. Version 2.0; Hamburg Notation System for Sign Language. An Introduction Guide": Signum Verlag.
 - [6] Mekala P., Gao Y., Fan J., Davari A. (2011), "Real-Time Sign Language Recognition Based Onneural Network Architecture", System Theory (SSST), 2011 IEEE 43rd Southeastern Symposium on, pp. 195-199.
 - [7] M G., Menon R., Jayan S., James R., G.V.V J. (2011), "Gesture Recognition for American Sign Language with Polygon Approximation", IEEE International Conference on Technology for Education.
 - [8] Pugeault N., Bowden R. (2011), "Spelling It Out: Real-Time Asl Fingerspelling Recognition", Proceedings of the 1st IEEE Workshop on Consumer Depth Cameras for Computer Vision, in conjunction with ICCV'2011, Barcelona, Spain.

- [9] Sato A., Shinoda K., Furui S. (2010), "Sign Language Recognition Using Time-of-Flight Camera", Proc. Meeting on Image Recognition & Understanding 2010 (MIRU2010), pp. 1861-1868.
- [10] Isaacs J., Foo S. (2004), "Hand Pose Estimation for American Sign Language Recognition", System Theory, 2004. Proceedings of the Thirty-Sixth Southeastern Symposium on, pp. 132-136.
- [11] Rajam P. S., Balakrishnan G. (2012), "Recognition of Tamil Sign Language Alphabet Using Image Processing to Aid Deaf-Dumb People", **Procedia Engineering**, 30(0), pp. 861-868.
- [12] Lang S. (2011), "Sign Language Recognition with Kinect", B.Sc. Thesis, Freie Universität.
- [13] Zafrulla Z., Brashear H., Starner T., Hamilton H., Presti P. (2011), "American Sign Language Recognition with the Kinect", ICMI'11, Alicante, Spain.
- [14] Li K. F., Lothrop K., Gill E., Lau S. (2011), "A Web-Based Sign Language Translator Using 3d Video Processing", International Conference on Network-Based Information Systems.
- [15] Han J., Awad G., Sutherland A. (2009), "Modelling and Segmenting Subunits for Sign Language Recognition Based on Hand Motion Analysis", **Pattern Recognition Letters**, 30(6), pp. 623-633.
- [16] Shanableh T., Assaleh K. (2011), "User-Independent Recognition of Arabic Sign Language for Facilitating Communication with the Deaf Community", **Digital Signal Processing**, 21(4), pp. 535-542.
- [17] Derpanis K. G., Wildes R. P., Tsotsos J. K. (2008), "Definition and Recovery of Kinematic Features for Recognition of American Sign Language Movements", **Image and Vision Computing**, 26(12), pp. 1650-1662.

- [18] Yang R., Sarkar S. (2009), "Coupled Grouping and Matching for Sign and Gesture Recognition", **Computer Vision and Image Understanding**, 113(6), pp. 663-681.
- [19] Orfanidou E., Adam R., Morgan G., McQueen J. M. (2010), "Recognition of Signed and Spoken Language: Different Sensory Inputs, the Same Segmentation Procedure", **Journal of Memory and Language**, 62(3), pp. 272-283.
- [20] Kelly D., McDonald J., Markham C. (2010), "A Person Independent System for Recognition of Hand Postures Used in Sign Language", **Pattern Recognition Letters**, 31(11), pp. 1359-1368.
- [21] Bowden R., Sarhadi M. (2002), "A Non-Linear Model of Shape and Motion for Tracking Finger Spelt American Sign Language", **Image and Vision Computing**, 20(9–10), pp. 597-607.
- [22] Raheja J. L., Chaudhary A., Singal K. (2011), "Tracking of Fingertips and Centres of Palm Using Kinect", Third International Conference on Computational Intelligence, Modelling & Simulation.
- [23] Lim Wei H., Wong F., Chekima A. (2008), "Comparison of Hand Segmentation Methodologies for Hand Gesture Recognition", Information Technology, 2008. ITSIM 2008. International Symposium on, pp. 1-7.
- [24] Holte M. B., Moeslund T. B., Fihl P. (2008), "View Invariant Gesture Recognition Using the Csemswissranger Sr-2 Camera", **International Journal of Intelligent Systems Technologies and Applications**, 5(3/4), pp. 295-303.
- [25] Trigueiros P., Ribeiro F., Lopes G. "Vision-Based Hand Segmentation Techniques for Human-Robot Interaction for Real-Time Applications", VIPIIMAGE, III Eccomas Thematic Conference

on Computational Vision and Medical Image Processing, Olhão, Portugal.

- [26] Wang Q., Chen X., Gao W. (2010), "Skin Color Weighted Disparity Competition for Hand Segmentation from Stereo Camera", Proceedings of British Machine Vision Conference (2010), pp. 1-11.
- [27] Ghobadi S. E., Loepprich O. E., Hartmann K., Loffeld O. (2007), "Hand Segmentation Using 2d/3d Images", Proceedings of Image and Vision Computing, Hamilton, New Zealand, pp. 64–69.
- [28] Oz C., Leu M. C. (2011), "American Sign Language Word Recognition with a Sensory Glove Using Artificial Neural Networks", **Engineering Applications of Artificial Intelligence**, 24(7), pp. 1204-1213.
- [29] Ibarguren A., Murtua I., Sierra B. (2010), "Layered Architecture for Real Time Sign Recognition: Hand Gesture and Movement", **Engineering Applications of Artificial Intelligence**, 23(7), pp. 1216-1228.
- [30] Wang Q., Chen X., Zhang L.-G., Wang C., Gao W. (2007), "Viewpoint Invariant Sign Language Recognition", **Computer Vision and Image Understanding**, 108(1–2), pp. 87-97.
- [31] Zafrulla Z., Brashear H., Hamilton H., Starner T. (2010), "A Novel Approach to American Sign Language (Asl) Phrase Verification Using Reversed Signing", Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, pp. 48-55.
- [32] Weaver K. A., Hamilton H., Zafrulla Z., Brashear H., Starner T., Presti P., Bruckman A. (2010), "Improving the Language Ability of Deaf Signing Children through an Interactive American Sign Language-Based Video Game", Proceedings of the 9th International

- Conference of the Learning Sciences - Volume 2, Chicago, Illinois, International Society of the Learning Sciences, pp. 306-307.
- [33] Zafrulla Z., Brashear H., Pei Y., Presti P., Starner T., Hamilton H. (2010), "American Sign Language Phrase Verification in an Educational Game for Deaf Children", Pattern Recognition (ICPR), 2010 20th International Conference on, pp. 3846-3849.
 - [34] Pei Y., Essa I., Starner T., Rehg J. M. (2008), "Discriminative Feature Selection for Hidden Markov Models Using Segmental Boosting", Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on, pp. 2001-2004.
 - [35] Zhang X., Chen X., Li Y., Lantz V., Wang K., Yang J. (November 2011), "A Framework for Hand Gesture Recognition Based on Accelerometer and Emg Sensors", **IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans**, 41(6).
 - [36] Huang Y., Monekosso D., Wang H., Augusto J. C. (2011), "A Concept Grounding Approach for Glove-Based Gesture Recognition", Seventh International Conference on Intelligent Environments.
 - [37] Vutinuntakasame S., Jaijongrak V.-r., Thiemjarus S. (2011), "An Assistive Body Sensor Network Glove for Speech- and Hearing-Impaired Disabilities", International Conference on Body Sensor Networks.
 - [38] Starner T., Weaver J., Pentland A. (1998), "Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video", **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, 20(12), pp. 1371-1375.

- [39] Ding L., Martinez A. M. (2009), "Modelling and Recognition of the Linguistic Components in American Sign Language", **Image and Vision Computing**, 27(12), pp. 1826-1844.
- [40] Thangali A., Nash J. P., Sclaroff S., Neidle C. (2011), "Exploiting Phonological Constraints for Handshape Inference in Asl Video", Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, pp. 521-528.
- [41] Scaramuzza R. S. D. (2011), "Asl Lecture Notes - 4a - Perception - Sensors": ETH Zurich.
- [42] Corporation M. (2010). "Gesture Keyboarding", Patent US 2010/0199 228 A1.
- [43] Website K. f. W. (2012), "Kinect Sensor", 2012 [cited March 1st 2012]; Available from: <http://www.kinectforwindows.org>.
- [44] Simon C., Meessen J., Tzovaras D., De Vleeschouwer C. (2007), "Using Decision Trees for Knowledge-Assisted Topologically Structured Data Analysis", Image Analysis for Multimedia Interactive Services, 2007. WIAMIS '07. Eighth International Workshop on, pp. 2-2.
- [45] Pugeault N., Bowden R. (2012), "Database Page at Nicolas Pugeault Homepage", 2012 [cited August 10 2012]; Available from: <http://personal.ee.surrey.ac.uk/Personal/N.Pugeault/index.php?key=FingerSpellingDataset>.
- [46] Flasiński M., Myśliński S. (2010), "On the Use of Graph Parsing for Recognition of Isolated Hand Postures of Polish Sign Language", **Pattern Recognition**, 43(6), pp. 2249-2264.
- [47] Al-Jarrah O., Halawani A. (2001), "Recognition of Gestures in Arabic Sign Language Using Neuro-Fuzzy Systems", **Artificial Intelligence**, 133(1-2), pp. 117-138.

- [48] Oikonomidis L., Kyriazis N., Argyros A. A. (2011), "Efficient Model-Based 3d Tracking of Hand Articulations Using Kinect", Proceedings of the 22nd British Machine Vision Conference, BMVC'2011, University of Dundee, UK.
- [49] Mistry P., Maes P. (2009), "Sixthsense – a Wearable Gestural Interface", Proceedings of SIGGRAPH Asia 2009, Sketch. Yokohama, Japan.
- [50] "Sixth Sense", [cited August 2012]; Available from: <http://www.pranavmistry.com/projects/sixthsense/>.
- [51] Gao W., Fang G., Zhao D., Chen Y. (2004), "A Chinese Sign Language Recognition System Based on Softmax/SVM/HMM", **Pattern Recognition**, 37(12), pp. 2389-2402.
- [52] Mohandes M., Deriche M., Johar U., Ilyas S. (2012), "A Signer-Independent Arabic Sign Language Recognition System Using Face Detection, Geometric Features, and a Hidden Markov Model", **Computers & Electrical Engineering**, 38(2), pp. 422-433.
- [53] Nguyen T. D., Ranganath S. (2012), "Facial Expressions in American Sign Language: Tracking and Recognition", **Pattern Recognition**, 45(5), pp. 1877-1891.
- [54] Zaki M. M., Shaheen S. I. (2011), "Sign Language Recognition Using a Combination of New Vision Based Features", **Pattern Recognition Letters**, 32(4), pp. 572-577.
- [55] Aran O., Akarun L. (2010), "A Multi-Class Classification Strategy for Fisher Scores: Application to Signer Independent Sign Language Recognition", **Pattern Recognition**, 43(5), pp. 1776-1788.
- [56] Jiang F., Gao W., Yao H., Zhao D., Chen X. (2009), "Synthetic Data Generation Technique in Signer-Independent Sign Language Recognition", **Pattern Recognition Letters**, 30(5), pp. 513-524.

- [57] Al-Rousan M., Assaleh K., Tala'a A. (2009), "Video-Based Signer-Independent Arabic Sign Language Recognition Using Hidden Markov Models", **Applied Soft Computing**, 9(3), pp. 990-999.
- [58] Vamplew P. W. (1996), "Recognition of Sign Language Using Neural Networks", PhD Thesis, University of Tasmania.
- [59] Oz C., Leu M. C. (2007), "Linguistic Properties Based on American Sign Language Isolated Word Recognition with Artificial Neural Networks Using a Sensory Glove and Motion Tracker", **Neurocomputing**, 70(16–18), pp. 2891-2901.
- [60] Munib Q., Habeeb M., Takruri B., Al-Malik H. A. (2007), "American Sign Language (Asl) Recognition Based on Hough Transform and Neural Networks", **Expert Systems with Applications**, 32(1), pp. 24-37.
- [61] Karami A., Zanj B., Sarkaleh A. K. (2011), "Persian Sign Language (Psl) Recognition Using Wavelet Transform and Neural Networks", **Expert Systems with Applications**, 38(3), pp. 2661-2667.
- [62] Lee Y.-H., Tsai C.-Y. (2009), "Taiwan Sign Language (Tsl) Recognition Based on 3d Data and Neural Networks", **Expert Systems with Applications**, 36(2, Part 1), pp. 1123-1128.
- [63] OpenNI [cited August 23th 2012]; Available from: <http://www.openni.org>.
- [64] McAndrew A., "An Introduction to Digital Image Processing with Matlab", Victoria University of Technology, 2004.
- [65] Hu M. K. (1962), "Visual Pattern Recognition by Moment Invariants", **IRE Transactions on Information Theory**, IT-8, pp. 179-187.

- [66] Eisner (2012), "Classifier Evaluation (Measuring Prediction Performance)", 2012 [cited August 2012]; Available from: webdocs.cs.ualberta.ca/.
- [67] J. Triesch C. v. d. M. (2001), "A System for Person-Independent Hand Posture Recognition against Complex Backgrounds", **IEEE Trans. Pattern Analysis and Machine Intelligence**, 23(12), pp. 1449–1453.
- [68] Amin M. A., Hong Y. (2007), "Sign Language Finger Alphabet Recognition from Gabor-Pca Representation of Hand Gestures", Machine Learning and Cybernetics, 2007 International Conference on, pp. 2218-2223.
- [69] Incertis I. G., García-Bermejo J. G., Casanova E. Z. (2006), "Hand Gesture Recognition for Deaf People Interfacing", Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, pp. 100-103.
- [70] Kuzmanic A., Zanchi V. (2007), "Hand Shape Classification Using Dtw and Lcss as Similarity Measures for Vision-Based Gesture Recognition System", EUROCON, 2007. The International Conference on Computer as a Tool, pp. 264-269.
- [71] Chaki A., Jain P., Gupta R. K. (2011), "Information Measure Ratio Based Real Time Approach for Hand Region Segmentation with a Focus on Gesture Recognition", Intelligent Systems, Modelling and Simulation (ISMS), 2011 Second International Conference on, pp. 172-176.

Abstract

Sign language recognition is an important challenge yet to be conquered in many areas of its applications. One of the most important problems in sign language recognition is hand segmentation from a scene with complex background. With the introduction of Microsoft Kinect, a depth image is now available along with the RGB image, easing the segmentation process if used correctly. In this research, a real-time finger-spelled alphabet recognition scheme is proposed based on geometric features extracted from the depth-segmented RGB images. A total of 26 geometric features, including 19 Moment-based and seven contour-based features, are extracted from the depth-segmented images after several preprocessing steps. The preprocessing steps include thresholding the depth image to create a mask for hand segmentation, smoothing the mask, translating and resizing it. Bayes classifier is then applied on the feature vector after an analysis using linear discriminant analysis (LDA). Kinect depth-based segmentation of RGB and depth images and proposing a real-time system for accurate classification of the resulting data are considered as the main contributions of this thesis.

The proposed method is tested on a vast database with more than 54,000 samples of static finger-spelled American Sign Language (ASL) alphabets covering different variations of hand gestures and complicated backgrounds obtained from Kinect. Several cross-validated recognition results are demonstrated based on different selections of the database, ranging from 78.29% on the whole database including 5 signers with 450 samples per 24 alphabets for each of signers (whole database) and 99.75% on a single signer with 100 samples per alphabet. Performance tests run on the system indicate a real-time processing speed with 38 fps (frames per second) using a single thread and 245 fps using a multi-thread architecture, both on laptop with a 2GHz Core i7 CPU. It is also manifested in results analysis and validation section that using LDA, a combination of RGB and depth features and samples randomization yields the best recognition results, with the latter one approving the necessity of a training phase for new users just before using the system.

Keywords: Sign Language, Depth, Kinect, Hand Segmentation, Finger spelled.



Shahrood University of Technology

Faculty of Computer Engineering and IT

Sign Language Recognition using ToF Depth Cameras

Ali Reza Manashty

Supervisor:

Dr. Morteza Zahedi

Advisor:

Dr. Hamid Hassanpour

Date: September 2012