

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشکده مهندسی کامپیوتر و فناوری اطلاعات

شناسایی زبان اشاره فارسی با استفاده از دوربین های عمقی

زهرا فروتن جهرمی

استاد راهنما:

دکتر حمید حسن پور

پایان نامه ارشد جهت اخذ درجه کارشناسی ارشد

آذر ماه ۱۳۹۲

دانشگاه صنعتی شاهرود

دانشکده مهندسی کامپیوتر و فناوری اطلاعات

پایان نامه کارشناسی ارشد خانم زهرا فروتن جهرمی
تحت عنوان: شناسایی زبان اشاره فارسی با استفاده از دوربین های عمقی

در تاریخ ۱۳۹۲/۰۹/۱۷ توسط کمیته تخصصی زیر جهت اخذ مدرک کارشناسی ارشد مهندسی کامپیوتر،
گرایش هوش مصنوعی مورد ارزیابی و با درجه _____ مورد پذیرش قرار گرفت.

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

امضاء	نماینده تحصیلات تکمیلی	امضاء	اساتید داور
	نام و نام خانوادگی : مهندس فرهادی		نام و نام خانوادگی : دکتر مرتضی زاهدی
			نام و نام خانوادگی : دکتر سعیده فردوسی

تقدیم بہ:

پدر و مادر عزیزم، با فداکاری و مہربانی ہمیشگی شان

ہمسرم، با ہم راہی و ہم یاری ہائش

دخترم بالہنذ زیبا و مگاہ پر امیدش

تقدیر و تشکر

از اساتید کرامت و با فضیلت خویش، جناب آقای دکتر حمید حسن پور، دکتر علی اکبر پویان و دکتر مرتضی زاهدی که در کلیه مراحل تحصیل با حمایت های علمی و معنوی خویش همواره راه 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 کرده جناب آقای طهری زاکری که نا بهواری و دغدگی های این راه صعب را با برتو نور دانش خود و همراهی با اینجانب هموار و روشن نمودند تشکر و قدردانی می نمایم.

تعهد نامه

اینجانب زهرا فروتن جهرمی 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 فارسی با استفاده از دوربین های عمقی تحت راهنمایی دکتر حمید حسن پور 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 اخلاق انسانی رعایت شده است .

تاریخ ۱۳۹۲/۰۹/۱۷

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

شناسایی زبان اشاره توسط رایانه گام اولیه و مهمی در راستای ارتقای تعامل کاربران ناشنوا با رایانه است. با وجود گستردگی کارهای انجام شده در شناسایی زبان اشاره سایر کشورها به خصوص زبان اشاره آمریکایی، تنها هنوز گام های اولیه و آزمایشگاهی در مورد زبان اشاره فارسی انجام پذیرفته است.

در این پایان نامه روشی جهت شناسایی حروف الفبای زبان اشاره فارسی باغچه بان (اولین و پرکاربردترین زبان اشاره فارسی) ارائه شده است که با استفاده از دوربین های عمقی و ویژگی های های هندسی دست، اسکلت و صورت، در محیطی با پس زمینه پیچیده درصد شناسایی بالایی را ارائه می دهد.

تحقیق پیشرو اولین تحقیق کاربردی برای شناسایی زبان اشاره فارسی در محیطی با پس زمینه پیچیده است. هم چنین استخراج ویژگی از اسکلت و دهان فرد نیز تا کنون در شناسایی زبان اشاره فارسی انجام نگرفته است.

کارهای اصلی انجام شده در این تحقیق عبارت اند از تولید اولین پایگاه داده عمقی زبان اشاره فارسی باغچه بان شامل تصاویر کامل بدن و به خصوص صورت و دهان، شناسایی کامل ۳۸ حرف الفبای معمول ناشنوایان با سرعت بلادرنگ، استفاده از ویژگی های اسکلتی و نقاط صورت در تشخیص زبان اشاره فارسی و تحلیل و بررسی نتایج سیستم تا رسیدن به درصد تشخیص بیش از ۹۹٪.

پایگاه داده استفاده شده در این سیستم دارای ۱۵۰ تا ۵۰۰ نمونه متفاوت از ۳۸ حرف الفبای زبان اشاره فارسی باغچه بان با فضایی دارای تداخل رنگی بین دست، صورت، بدن، لباس و پس زمینه

تصویر می باشد که توسط سنسور کینکت ضبط شده است و به ازای هر فریم آن، تصویر رنگی، تصویر عمقی، اسکلت نیم تنه بالای بدن و ۱۲۱ نقطه صورت فرد استخراج شده است.

تحلیل نتایج سیستم نشان دهنده بهترین درصد شناسایی ۹۹,۴٪ برای ۳۸۰۰ نمونه (۱۰۰ نمونه از هر حرف) و ۹۹,۲٪ به ازای ۵۷۰۰ نمونه (۱۵۰ نمونه از هر حرف) است. تاثیر استفاده از ویژگی های هندسی دست و دهان، ویژگی های بر پایه تبدیل موجک استخراج شده از تصویر قطعه بندی شده دهان و ویژگی های اسکلتی مربوط به مکان قرارگیری دست، در این تحقیق مورد ارزیابی قرار گرفته اند. بررسی عملکرد سیستم نیز نشان دهنده سرعت بین ۴۰ تا ۷۰ فریم بر ثانیه در شناسایی داده های پایگاه داده و با استفاده از دسته بند بیز می باشد.

کلمات کلیدی: شناسایی زبان اشاره فارسی، تبدیل موجک، ویژگی های هندسی، کینکت، املاي انگشتی، الفبای باغچه بان، اسکلت.

مقالات مستخرج از پایان نامه

1. Zahra ForootanJahromi, Hamid Hassanpour, Alireza Manashty and Amir Sadeghi, "A Novel Approach towards Finger Spelled Persian Alphabet Recognition using Depth-Based Face, Skeleton and Geometric Feature Extraction", IJTTE Journal, (Submitted for publication).

۲. زهرا فروتن جهرمی، حمید حسن پور و علیرضا مانشتی، "شناسایی زبان اشاره با استفاده

از استخراج ویژگی بر پایه تبدیل موجک و تصاویر عمقی"، یازدهمین کنفرانس سیستم

های هوشمند ایران، ۹ و ۱۰ اسفند ماه ۱۳۹۱، دانشگاه خوارزمی، تهران، ایران

فهرست مطالب

فصل ۱: مقدمه	۱
۱-۱- مقدمه	۲
۲-۱- زبان اشاره	۳
۳-۱- کاربردهای زبان اشاره	۴
۴-۱- انواع ساختارهای زبان اشاره	۵
۵-۱- انواع اشاره	۷
۱-۵-۱- نیازمندی ها و چالش های تشخیص اشاره	۸
۶-۱- انواع الفبای دستی در زبان اشاره فارسی	۸
۱-۶-۱- الفبای گویای باغچه بان	۸
۲-۶-۱- الفبای دستی فارسی جولیا سمیعی	۱۱
۳-۶-۱- الفبای دستی روزیک شهبازیان	۱۲
۴-۶-۱- الفبای دستی محسن لوح موسوی	۱۳
۷-۱- اهمیت تشخیص زبان اشاره	۱۵
۸-۱- تعریف مسئله	۱۶
۹-۱- روش حل مسئله	۱۸
فصل ۲: مروری بر کارهای پیشین	۱۹
۱-۲- بررسی موردی کارهای انجام شده جهت تشخیص زبان اشاره فارسی	۲۰

۲۱	۱-۱-۲- کار کرمی و همکارانش
۲۳	۲-۱-۲- کار برزگر و شرافت
۲۵	۲-۲- طبقه بندی کارهای پیشین بر حسب سامانه دریافت داده های ورودی
۲۵	۱-۲-۲- استفاده از دستکش و سنسور های شتاب سنج
۲۸	۲-۲-۲- کارهای پیشین با استفاده از دوربین های رنگی
۳۰	۳-۲-۲- کارهای پیشین با استفاده از دوربین های استریو
۳۲	۴-۲-۲- کارهای پیشین با استفاده از دوربین های عمقی
۳۵	۵-۲-۲- کارهای پیشین با استفاده از کینکت
۴۰	۳-۲- طبقه بندی کارهای پیشین بر حسب ویژگی های استخراج شده
۴۰	۱-۳-۲- استفاده از ویژگی های هندسی
۴۲	۲-۳-۲- ویژگی هایی بر حسب تبدیل موجک
۴۳	۳-۳-۲- کارهای پیشین بر حسب ویژگی های اسکلتی
۴۴	۴-۲- روش های دسته بندی
۴۴	۱-۴-۲- مدل مخفی مارکوف (HMM)
۴۵	۲-۴-۲- دسته بند بیز
۴۵	۳-۴-۲- شبکه های عصبی
۴۶	۵-۲- استخراج ویژگی های صورت جهت تشخیص اشاره
۴۹	۶-۲- نتیجه گیری
۵۲	فصل ۳: روش پیشنهادی

- ۳-۱- پایگاه داده ۵۳
- ۳-۱-۱- کینکت و ویژگی های آن ۵۳
- ۳-۱-۲- دلیل انتخاب حروف الفبای گویای باغچه بان ۵۷
- ۳-۱-۳- پایگاه داده عمقی الفبای گویای باغچه بان ۵۸
- ۳-۱-۴- جدا سازی حروف ۶۳
- ۳-۱-۵- مشخصات پایگاه داده نهایی ۶۸
- ۳-۱-۶- مقایسه پایگاه داده با پایگاه داده های موجود ۶۹
- ۳-۱-۷- نمونه های پایگاه داده ۷۱
- ۳-۲- پیش پردازش ۷۳
- ۳-۲-۱- برش تصاویر دست ۷۳
- ۳-۲-۲- قطعه بندی دهان ۷۴
- ۳-۲-۳- فرآیند پیش پردازش ۷۹
- ۳-۲-۴- ساخت ماسک عمقی با آستانه گیری ۸۰
- ۳-۲-۵- پیش پردازش بر روی ماسک عمقی ۸۱
- ۳-۲-۶- پیش پردازش بر روی تصویر رنگی ۸۲
- ۳-۳- استخراج ویژگی های هندسی دست ۸۳
- ۳-۳-۱- Moment ها ۸۳
- ۳-۳-۲- ویژگی های بر مبنای کانتور ۸۷
- ۳-۴- استخراج ویژگی های اسکلتی ۹۰

۹۱	۳-۵- استخراج ویژگی بر پایه تبدیل موجک
۹۴	۳-۶- استخراج ویژگی های دهان
۹۵	۳-۷- تحلیل و انتخاب ویژگی ها
۹۶	۳-۸- روش دسته بندی
۹۸	فصل ۴: نتایج بدست آمده
۹۹	۴-۱- تعاریف
۹۹	۴-۱-۱- اعتبار سنجی ضربدری با n -تا کردن
۱۰۰	۴-۱-۲- معیارهای ارزیابی
۱۰۲	۴-۱-۳- تصویر استفاده شده
۱۰۲	۴-۲- نتایج
۱۰۳	۴-۳- تحلیل آزمایشات
۱۰۳	۴-۳-۱- تاثیر تعداد نمونه ها
۱۰۳	۴-۳-۲- تاثیر تصویر انتخاب شده
۱۰۴	۴-۳-۳- تاثیر استفاده از تبدیل موجک
۱۰۵	۴-۴- تحلیل شناسایی هریک از حروف
۱۰۷	۴-۵- مقایسه با روش های پیشین
۱۱۰	فصل ۵: اعتبارسنجی و تحلیل داده ها
۱۱۱	۵-۱- تحلیل انتخاب ویژگی ها با استفاده از LDA
۱۱۲	۵-۲- تحلیل تابع چگالی احتمال ویژگی ها

۵-۳- سرعت پردازش روش پیشنهادی ۱۱۴

فصل ۶: نتیجه گیری و کارهای آینده..... ۱۱۶

۶-۱- نتیجه گیری ۱۱۷

۶-۲- پیشنهادات برای کارهای آینده ۱۱۷

فهرست شکل ها

- شکل ۱-۱- الفبای گویای باغچه بان. تصویر از [۷] ۱۰
- شکل ۲-۱- الفبای دستی جولیا سمیعی. تصویر از [۶] ۱۲
- شکل ۳-۱- الفبای دستی روزیک شهبازیان و شباهت برخی حروف با الفبای دستی آمریکایی. [۶]. ۱۳
- شکل ۴-۱- الفبای دستی محسن لوح موسوی. تصویر از [۶] ۱۴
- شکل ۵-۱- پایگاه داده برای تشخیص زبان اشاره با استفاده از دوربین معمولی (RGB) ۱۷
- شکل ۶-۱- پایگاه داده برای تشخیص زبان اشاره با استفاده از دوربین عمقی ۱۷
- شکل ۱-۲- پایگاه داده مورد استفاده در کار کرمی و همکارانش. تصویر از [۹] ۲۲
- شکل ۲-۲- حروف الفبای استفاده شده در کار برزگر و شرافت. تصویر از [۱۰] ۲۴
- شکل ۳-۲- تصویری از دستکش های رنگی به همراه شتاب سنج و جزئیات شتاب سنج سه بعدی به کار رفته در بازی Copy Cat [۴] ۲۶
- شکل ۴-۲- نمایی از دستکش و شبکه حسگری بکار رفته در تحقیق ساجاکارن و همکارانش. [۱۹] ۲۸
- شکل ۵-۲- 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 [۲۳] ۳۰
- شکل ۶-۲- تصویری از یک دوربین فیلم برداری استریو. ۳۱
- شکل ۷-۲- 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 [۲۴] ۳۲
- شکل ۸-۲- 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 [۲۴] ۳۲
- شکل ۹-۲- نمایی از نحوه کارکرد دوربین های عمقی در تصویربرداری از محیط. ۳۳

- شکل ۲-۱۰- نمایش از سنسور کینکت. ۳۶
- شکل ۲-۱۱- نمایش از حالت ایستاده (راست) و نشسته (چپ) در استفاده از سنسور کینکت در بازی CopyCat. تصویر از [۲۹]. ۳۷
- شکل ۲-۱۲- نمونه پایگاه داده استفاده شده در کار [۳۱]. ۳۹
- شکل ۲-۱۳- مراحل کار جهت مدل کردن چهره به وسیله ی یک مدل ظاهری فعال در [۵۲]. ۴۷
- شکل ۲-۱۴- نقاط برجسته ی استخراج شده از صورت در کار [۵۳]. ۴۸
- شکل ۲-۱۵- ردیابی نقاط برجسته صورت توسط روش پیشنهادی در [۵۴]. نقاط آبی، نقاطی است که با استفاده از سیستم پیشنهادی استخراج و ردیابی شده اند و نقاط قرمز نقاطی است که با استفاده از KLT ردیابی شده اند. ۴۹
- شکل ۳-۱- نمایش 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. ۵۵
- شکل ۳-۲- نمایش از اسکلت استخراج شده توسط سنسور کینکت. ۵۶
- شکل ۳-۳- نمایش از ۱۲۱ نقطه استخراج شده از صورت که به صورت شبکه ای بر روی صورت منطبق شده است. ۵۷
- شکل ۳-۴- نمایش از محیط پیرامون و پس زمین پیچیده در ایجاد پایگاه داده. ۵۹
- شکل ۳-۵- بالا بردن دست برای مشخص کردن آمادگی برای شروع ضبط یک حرف خاص. ۶۰
- شکل ۳-۶- حروف آ تا د از اجرای الفبای گویای باغچه بان در زبان اشاره فارسی. ۶۱
- شکل ۳-۷- حروف ذ تا ک از اجرای الفبای گویای باغچه بان در زبان اشاره فارسی. ۶۲
- شکل ۳-۸- حروف گ تا ء از اجرای الفبای گویای باغچه بان در زبان اشاره فارسی. ۶۳
- شکل ۳-۹- نمونه ای از نرم افزار ضبط پیوسته داده های کینکت، Kinect Studio. ۶۴
- شکل ۳-۱۰- نمایش از نرم افزار نوشته شده برای جداسازی حروف از یکدیگر. ۶۷

شکل ۳-۱۱-نمایی از پوشه های ایجاد شده برای پایگاه داده به ازای حروف مختلف و داده های ذخیره شده هر حرف..... ۶۷

شکل ۳-۱۲-مقایسه پیچیدگی پس زمینه تصاویر در روش پیشنهادی و دیگر روش های تشخیص زبان اشاره فارسی. تصویر بالا مربوط به پایگاه داده روش پیشنهادی، تصاویر پایین مربوط به پایگاه داده دیگر کارهای فارسی (به ترتیب از راست به چپ [۱۰] و [۹]) ۷۰

شکل ۳-۱۳-سه تصویر از حرف ذ نشان دهنده تفاوت درون کلاسی نمونه های پایگاه داده..... ۷۲

شکل ۳-۱۴-شباهت بین کلاسی در حرف ش (سمت راست) و حرف س (سمت چپ). تفاوت دو حرف تنها در شکل دهان است. ۷۲

شکل ۳-۱۵-برش دست با استفاده از مختصات اسکلتی دست به فاصله ۷۰ پیکسل از هر جهت. ... ۷۳

شکل ۳-۱۶-نقاط استخراج شده از دهان..... ۷۵

شکل ۳-۱۷-تصویر ۳۸ دهان استخراج شده از حروف مختلف الفبای زبان اشاره فارسی باغچه بان.. ۷۹

شکل ۳-۱۸-مراحل مختلف پیش پردازش انجام شده بر روی داده های تصویر. نمودار از [۳۱] ۸۰

شکل ۳-۱۹- یک تصویر رنگی از نمونه های پایگاه داده مورد استفاده. ۸۰

شکل ۳-۲۰- اعمال اصلی پیش پردازش بر روی تصویر رنگی شکل ۳-۱۹. ساخت ماسک عمقی (۱) و اعمال ماسک بر روی تصویر رنگی (۲)..... ۸۰

شکل ۳-۲۱-نمایی از بخش تحلیل خطای نرم افزار پیاده سازی شده و مراحل مختلف پیش پردازش و استخراج ویژگی یک نمونه ورودی در آن..... ۸۲

شکل ۳-۲۲-بزرگترین لبه تصویر یا کانتور استخراج شده که نقاط آن در اطراف دست قطعه بندی شده مشخص شده اند. تصویر از [۳۱]..... ۸۷

شکل ۳-۲۳-نمایشی از مقادیر مختلف ویژگی فشردگی (C) بر اساس شکل ظاهری اجسام موجود در تصویر. تصاویر از [۳۱] ۸۸

- شکل ۳-۲۴- تصویری از پوسته محدب استخراج شده از یک نمونه از حرف y از پایگاه داده کار [۳۱].
 ۸۹.....
- شکل ۳-۲۵- تصویر محدوده استخراج شده اطراف دست، خط بالای سر تا دست و خط بالای سر تا
 چانه. ۹۱.....
- شکل ۳-۲۶: خروجی یک تبدیل موجک دوبعدی گسسته در سطح یک. ۹۲.....
- شکل ۳-۲۷: اعمال تبدیل موجک تا سه سطح بر روی تصویر حرف زبان اشاره. سائز تصاویر از چپ به
 راست ۱۶*۱۶ و ۸*۸ و ۴*۴ می باشند. ۹۳.....
- شکل ۳-۲۸- بردار ویژگی های استخراج شده از نمونه های ورودی ۹۵.....
- شکل ۴-۱- نمودار مقایسه تاثیر تغییر تصویر قطعه بندی شده در صحت خروجی های سیستم به ازای
 انتخاب های متنوع از پایگاه داده. ۱۰۴.....
- شکل ۵-۱- نمودار کاهش درصد شناسایی سیستم با کاهش ویژگی ها در حالت بدون موجک و تصویر
 رنگی. ۱۱۲.....
- شکل ۵-۲- تابع چگالی توزیع احتمال برخی از ویژگی های استخراج شده. ۱۱۳.....

فهرست جداول

- جدول ۱-۳- لیست حروف موجود در پایگاه داده و تعداد نمونه از هر حرف. ۶۸.....
- جدول ۲-۳- لیست نقاط استخراج شده از صورت برای جدا سازی دهان از تصویر صورت. ۷۴.....
- جدول ۳-۳- لیست ۲۶ ویژگی استخراج شده از دست. ۸۹.....
- جدول ۴-۳- مقایسه نتایج به دست آمده از روش پیشنهادی با استفاده از تبدیل موجک بر روی پایگاه داده کار [۳۱] و یک کار دیگر. ۹۳.....
- جدول ۱-۴- چهار تعریف مهم بر روی جدول در هم ریختگی. ۱۰۱.....
- جدول ۲-۴- نتایج بدست آمده توسط سیستم پیشنهادی به تفکیک پارامترهای مختلف. ۱۰۳.....
- جدول ۴-۴- ماتریس در هم ریختگی بهترین نتیجه بر روی کل پایگاه داده مورد استفاده بر مبنای درصد. ۱۰۶.....
- جدول ۳-۴- جدول مقایسه روش پیشنهادی با روش های پیشین. ۱۰۸.....
- جدول ۱-۵- زمان پردازش و ۲ بار دسته بندی در سیستم پیشنهادی برای ۳۸۰۰ نمونه و ۴۸ ویژگی به تفکیک نوع پردازش. زمان های محاسبه شده مستقیم به صورت زیر خطی مشخص شده اند... ۱۱۴

فصل ١: مقدمه

۱-۱- مقدمه

در بین انسان ها، افراد مختلف از جمله کودکان وجود دارند که قادر به شنیدن صدا نبوده و قادر به برقراری ارتباط کلامی نمی باشند، در نتیجه زبان اشاره تنها روش ارتباطی آن ها می باشد که در بسیاری از موارد، تنها راه ارتباطی آن ها با دنیای اطراف نیز می باشد. در جدید ترین آمار حدود ۳۶۰ میلیون نفر در دنیا ناشنوا بوده و یا مشکلات شنوایی دارند [۱]. این افراد اغلب در ارتباط با افراد شنوا دچار مشکل می شوند، چرا که اکثر افراد شنوا با زبان اشاره آشنایی نداشته و قادر به برقراری ارتباط با ناشنوایان نمی باشند.

به کار آمدن رایانه و شبکه های جهانی همچنین نیاز این افراد به برقراری ارتباط با رایانه یا دیگر افراد در دنیا، باعث ظهور روش های تشخیص زبان اشاره توسط رایانه در مراکز تحقیقاتی و پژوهشی گردید. تمایل این افراد برای استفاده از این زبان، مخصوصا در کودکان که هنوز زبان نوشتاری را به خوبی فرا نگرفته اند، باعث رشد این حوزه تحقیقاتی در سراسر جهان گردید.

در حال حاضر برای زبان های مختلف در دنیا، زبان اشاره مخصوص آن ها وجود دارد که معروفترین آن ها زبان اشاره امریکایی^۱ می باشد. در ایران نیز زبان اشاره فارسی^۲ مورد نظر برخی از پژوهشگران قرار گرفته است تا بتوانند توسط رایانه آن را تشخیص دهند. ناشنوایان کشورهای مختلف همچون ایران، زبان اشاره مخصوص به خودشان را دارند.

مبحث شناسایی خودکار زبان اشاره^۳ به شناسایی زبان اشاره با استفاده از رایانه می پردازد که در واقع ترکیبی از روش های نرم افزاری شناسایی و دسته بندی بخش های زبان و سخت افزار های دریافت اطلاعات از فرد اجرا کننده آن می باشد [۲]. این سخت افزار ها می توانند از دوربین های معمولی تا دستکش ها و شتاب سنج های سه بعدی را شامل شوند.

^۱ American Sign Language

^۲ Persian Sign Language

^۳ Automatic Sign Language Recognition

در این تحقیق با بررسی روش های پیشین بر حسب ابزارهای سخت افزاری مورد استفاده و روش های شناسایی نرم افزاری، روشی با استفاده از دوربین های عمقی ارائه می شود 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، انواع آن، کاربردها و اهمیت تشخیص زبان اشاره توسط رایانه پرداخته شده است.

۱-۲- زبان اشاره

زبان اشاره سیستم قراردادی خاص منظم و پیشرفته زبانی است که برای انتقال معنا، به جای الگوهای صوتی یا نوشتاری در زبان، از الگوهای علایم دیداری (ترکیب همزمان شکل، جهت و حرکت دستان، بازوها یا بدن و حالات صورت برای بیان روان افکار گوینده) در یک کلمه اشاره استفاده می کند [3]. عناصر یک اشاره عبارتند از: شکل دست ها، جهت کف دست، موقعیت حرکت دست و علامت های غیردستی (یا حالت صورت).

زبان های رایج ارتباطی افراد کر و لال در جهان زبان های اشاره هستند. زبان های اشاره دارای ساختارهای ویژه به خود هستند و معمولاً از لحاظ ساختاری و تطابق یک به یک، با زبان های گفتاری کاملاً متفاوت هستند. حداقل ۱۳۰ زبان اشاره شناخته شده در جهان وجود دارد [۴] که هر یک دارای ساختار زبانی خود هستند.

جهت انتقال مفاهیم در زبان اشاره سه روش اساسی وجود دارد:

- ۱- املای انگشتی که با هجی کردن حرف به حرف هر کلمه توسط شکل و حالت انگشتان دست (که در بعضی انواع زبان اشاره شامل حالت دهان نیز می باشد) صورت می گیرد.
- ۲- حرکات اشاره ایی که هر حرکت نشانگر یک کلمه در زبان اشاره می باشد.
- ۳- ویژگی های غیر دستی که شامل حالات چهره، زبان، لب و موقعیت بدن می باشد.

املائی انگشتی یک روش برای نمایش کلماتی است که در لغت نامه ی زبان اشاره، نمادی (اشاره ایی) برای آن وجود ندارد. بدین ترتیب که با تجزیه کلمه مورد نظر به حروف تشکیل دهنده ی آن، هر حرف آن توسط املائی انگشتی (شکل و حالت خاصی از انگشتان دست و حالت دهان) بیان می گردد [۵] که در این تحقیق سعی در شناسایی حروف الفبای زبان اشاره فارسی بدین منظور گردیده است.

۱-۳- کاربردهای زبان اشاره

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.

از دیگر کاربرد های زبان اشاره می توان به استفاده از زبان اشاره برای برقراری ارتباط بین افراد ناشنوا، انتقال فرمان های نظامی، که به دلیل خاموشی رادیویی و گفتاری تنها راه برقراری ارتباط در عملیات ها می باشد، برقراری سریع ارتباط در مسابقات ورزشی و دیگر کاربردهای روزمره که انسان زبان بدن را به زبان گفتاری ترجیح می دهد؛ نام برد.

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 کاربرد های 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.

۴-۱- انواع 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 [۴]:

املائی انگشتی^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^۲ خاصی وجود نداشته باشد، معمولاً از این نوع ساختار برای برقراری ارتباط استفاده می شود.

حالات 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 Finger Spelling

^۲ Gesture

^۳ در این گزارش منظور از زبان اشاره Sign Language و منظور از اشاره Gesture می باشد.

^۴ Hand Gestures

حالات صورت^۱

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 امریکایی، از ابروها برای تغییر نوع سوال در گروهی از اشاره ها استفاده می شود.

حالات بدن^۲

حرکات قسمت بالا تنه بدن^۳ نیز در انتقال مفاهیم به خصوص در تاکید بر روی کلمات ادا شده توسط دست و یا چهره از طریق انبساط دادن^۴ و فشرده کردن^۵ بدن مورد استفاده قرار می گیرند.

استفاده از فضا^۶

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^۷ و یا تصادم^۸ اشیاء می تواند معانی خاصی را بدهد که تنها با استفاده از فضای اطراف امکان پذیر می باشند.

^۱ Face Gestures

^۲ Body Gestures

^۳ Torso

^۴ Expansion

^۵ Compression

^۶ Spatial Use

^۷ Separation

^۸ Collision

۱-۵- انواع اشاره

در یک دسته بندی اشارات را می توان در دو دسته ی ایستا (Static) و پویا (Dynamic) دسته بندی کرد. در اشارات ایستا کاربر یک حالت معین و شکل خاصی از اشاره را در نظر گرفته و آن را نشان می دهد. اشارات پویا همراه با حرکات پیشین و پسین می باشند. برخی از اشارات مانند زبان اشاره ترکیبی از هر دو نوع اشارات پویا و ایستا می باشند.

عموماً اشارات در سه نوع زیر قرار می گیرند:

۱- اشاره های دست و بازو

اشاره هایی که دست و صورت در ادا کردن آن دخیل هستند مانند زبان اشاره، سرگرمی هایی که در محیط مجازی به کاربر امکان تعامل با محیط مجازی را با استفاده از اشاره های دست می دهند.

۲- اشاره های سر و صورت

حالاتی که سر و صورت در نشان داده اشاره دخیل هستند. مانند حرکت سر و صورت به نشانه ی سلام و یا حرکت دادن دست جهت دست دادن، و یا باز کردن دهان به نشانه ی صحبت کردن و یا اشاره با استفاده از چشمک زدن.

۳- اشاره های بدن

اشاره هایی که شامل حرکت کل بدن می باشد. مانند تشخیص حالات بدن انسان جهت مراقبت های درمانی.

۱-۵-۱- نیازمندی ها و چالش های تشخیص اشاره

اصلی ترین نیازمندی تشخیص اشاره تکنولوژی ردیابی اشاره و دریافت تصاویر ورودی و پردازش آن ها می باشد.

یک راه معمول جهت ردیابی حالت دست و تشخیص اشاره ی دست، استفاده از دستکش های مخصوص می باشد. این دستکش ها مجهز به سنسورها و شتاب سنج هایی هستند که اطلاعاتی در مورد حالت دست و مکان قرار گیری دست در فضا و حالت انگشتان را فراهم می کنند. اگرچه این دستکش ها اطلاعات بسیار دقیقی از حالت و نحوه ی حرکت دست در اختیار کاربر قرار می دهند، اما گران قیمت بوده و استفاده از آن ها طبق تمایل افراد نمی باشد.

علوم بینایی ماشین و تشخیص اشاره بر پایه تصویر، تکنیک هایی را برای رفع این محدودیت ها ارائه می دهند.

۱-۶-۱- انواع الفبای دستی در زبان اشاره فارسی

علاقه به ساخت نوعی الفبای دستی که بتواند برای آموزش کودکان ناشنوا مورد استفاده قرار گیرد ، پیشینه ای دراز دارد. از آنجا که الفباهای نوشتاری متفاوتی وجود دارد ، به همان نسبت انواع مشابهی از الفباهای دستی ناشنوایان در جهان نیز پدید آمده است. در ایران ، اولین تلاش برای ساخت الفبای دستی ناشنوایان توسط جبار باغچه بان برای تعلیم به سه کودک ناشنوای باغچه اطفال تبریز آغاز شده است.

۱-۶-۱- الفبای گویای باغچه بان

جبار باغچه بان بعد از اندیشه و تلاش زیاد ، موفق به کشف اسرار صداهای زبان فارسی شد و آن را به دو گروه صداهای تنفسی و حنجره ای و هریک از آنها را به دو دسته صداهای امتداد پذیر و غیر امتداد پذیر تقسیم کرد.

باغچه بان برای اینکه صدا را برای شاگرد ناشنوایش محسوس کند، از حواس بینایی و لامسه اش مدد گرفت و الفبای دستی خود را در سال ۱۳۰۳ شمسی اختراع کرد که در نوع خود، در دنیا کم نظیر است. چون الفباهای دستی کشورهای مختلف ، اغلب مجموعه ای از علائم قرار دادی برای به یاد آوردن شکل حروف هستند. یعنی مخترعین این گونه الفباها کوشیده اند تا حد امکان با شکل انگشتان دست "شکل حروف" مورد نظر را نشان دهند. در حالیکه در الفبای دستی باغچه بان کوشش شده است که "تلفظ هر حرف" با شکل انگشتان و محل گذاشتن دست بر سینه و یا نگاهداشتن آن در مقابل دهان نشان داده شود. به زبان بهتر، الفبای دستی باغچه بان بر اساس "چگونگی اصوات گفتار" بنا شده است، نه به "شکل حرف". در نتیجه، غالبا آن را "الفبای گویای باغچه بان" می نامند. الفبای گویای باغچه بان در تمام مدارس ناشنویان ایران به طور وسیع مورد استفاده قرار گرفت [6]. نمادهای حروف الفبای باغچه بان در شکل ۱-۱ نشان داده شده است.



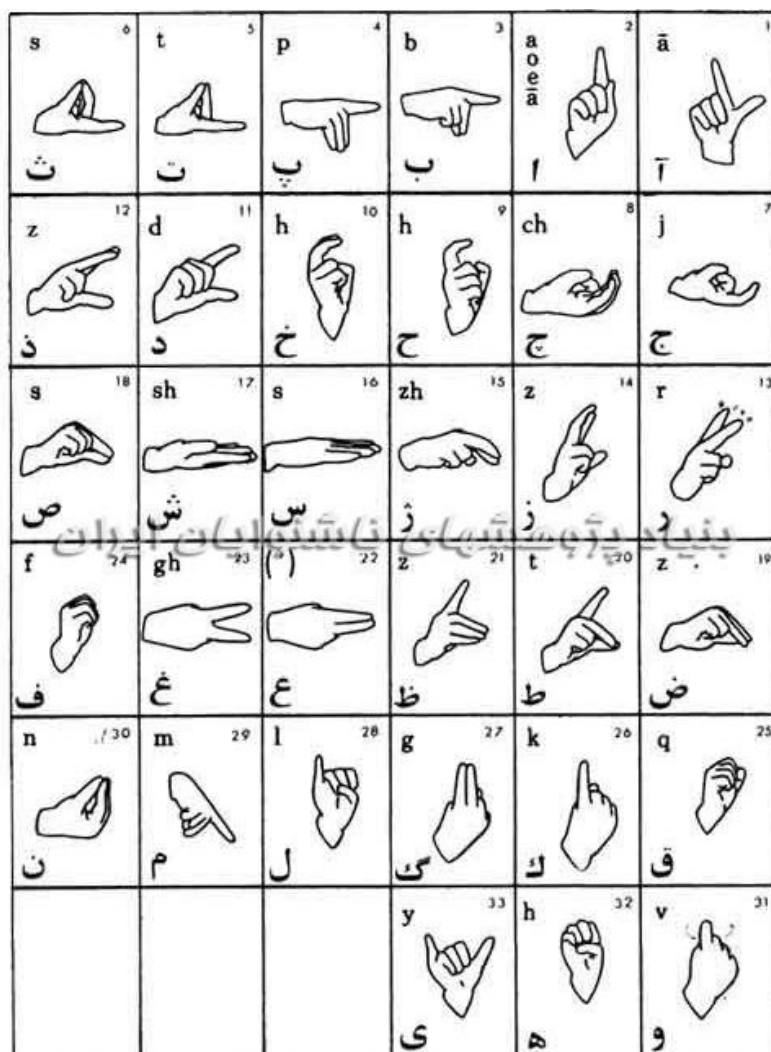
شکل ۱-۱- الفبای گویای باغچه بان. تصویر از [۷]

۱-۶-۲-الفبای دستی فارسی جولیا سمیعی

جولیا سمیعی (Julia Ann Oliver Samii) بعد از تاسیس سازمان ملی رفاه ناشنوایان در سال ۱۳۵۰ شمسی، به استخدام این سازمان در آمد. مهم ترین کار وی در این دوران ابداع "اولین الفبای دستی فارسی" در سال ۱۳۵۴ بود که آن را Persian Hand Alphabet نامید (شکل ۱-۲).

تفاوت الفبای دستی فارسی جولیا سمیعی با الفبای گویای باغچه بان، در عدم استفاده از صورت و لب برای ادای حروف الفبای فارسی می باشد بنابر این با سهولت بیشتری به کار برده می شود. البته الفبای دستی جولیا سمیعی فقط شامل تصویر ۳۳ حرف الفبا می باشد.

الفبای دستی مذکور در دبستان ناشنوایان باغچه بان به آزمایش گذاشته شد که نتایج آن چشمگیر ارزیابی گردیده است اما به علت مرگ نابهنگام جولیا سمیعی مجال گسترش در مراکز آموزشی ناشنوایان ایران را نیافت و به مرور زمان، کم کم به فراموشی سپرده شد.



شکل ۱-۲- الفبای دستی جولیا سمیعی. تصویر از [۶]

۱-۶-۳- الفبای دستی روزیک شهبازیان

این الفبا دارای تصاویر ۳۸ حرف فارسی می باشد که از نظر تعداد، نسبت به الفبای دستی جولیا سمیعی کامل تر به نظر می آید. با این حال، هفده حروف دیده شده در الفبای روزیک شهبازیان، یعنی آ، ب، ت، ج، د، ر، ژ، ف، ق، ک، گ، ل، م، ن، و، ه، ی؛ به صورت کپی ناشیانه از الفبای دستی آمریکایی گرفته شده است (شکل ۱-۳). از آنجا که شکل الفبای نوشتاری زبان فارسی با انگلیسی بسیار متفاوت است، بهره گیری از این نوع الفبای دستی هیچ کمکی به جز سردرگمی به ناشنویان نمی نمود.



شکل ۱-۳- الفبای دستی روزیک شهبازیان و شباهت برخی حروف با الفبای دستی آمریکایی. [۶].

۱-۶-۴- الفبای دستی محسن لوح موسوی

اعضای کمیته پژوهش و توسعه زبان اشاره فارسی دانشگاه علوم بهزیستی و توانبخشی کشور در سال ۱۳۷۵ شمسی چون الفبای دستی روزیک شهبازیان را سودمند به حال ناشنوایان ایران ندیدند و همچنین آن را برای به کارگیری در اشارات فارسی مفید ندانستند، به فکر احیای مجدد الفبای دستی جولیا سمیعی آن هم بعد از بیست سال افتادند و با ایجاد تغییرات جزئی در چهار حروف الفبایی همچون آ، ژ، ش و ف، سومین الفبای دستی فارسی با تصاویر ۳۷ حروف فارسی را تهیه کردند که بعدها در کتاب زبان اشاره فارسی جلد سوم به سال ۱۳۷۸ شمسی چاپ شد و گاهگاه آن را به نام الفبای دستی محسن موسوی می شناسند (شکل ۱-۴). دلیل اصلی تغییرات ایجاد شده در الفبای دستی جولیا سمیعی به افزایش کارایی و سهولت به کار گیری آن در اشارات فارسی به همراه رعایت ضوابط علمی بر می گردد. [۶]



شکل ۱-۴- الفبای دستی محسن لوح موسوی. تصویر از [۶]

۱-۷- اهمیت تشخیص زبان اشاره

تشخیص اشاره^۱ مبحثی در علوم رایانه و تکنولوژی زبان است که هدف آن تفسیر حرکات انسان از طریق الگوریتم‌های ریاضی می‌باشد. منشا حرکات می‌تواند به صورت جابجایی یا تغییر حالت هر عضو بدن باشد، اما معمولاً از دست یا صورت استفاده می‌شود. در حال حاضر تمرکز در این زمینه شامل تشخیص احساسات از طریق چهره و تشخیص حرکات دست می‌باشد.

الگوریتم‌های بسیاری برای استفاده از دوربین و بینایی رایانه‌ای ساخته شده‌اند تا بتوانند زبان اشاره را تفسیر کنند. همچنین شناسایی و تشخیص طرز ایستادن یا نشستن، راه رفتن، همجواری و رفتار انسان نیز بخشی از تکنولوژی تشخیص حرکت است.

تشخیص حرکت می‌تواند به عنوان آغاز راهی برای درک زبان بدن انسان توسط کامپیوترها^۲ باشد. بدین ترتیب می‌تواند شروعی برای ایجاد یک ارتباط غنی‌تر بین ماشین و انسان باشد، که تاکنون از رابط کاربری متنی اولیه و یا حتی واسطه گرافیکی کاربر^۳ استفاده کرده و هنوز هم بیشتر محدود به ورودی‌های صفحه کلید و ماوس است [۸].

تشخیص زبان اشاره نیز از این جهت مورد توجه واقع شده است که بیشتر افراد کر و لالی که در جهان وجود دارند برای یادگیری زبان‌های معمولی گفتاری یا نوشتاری دچار مشکل هستند و زبان‌های اشاره و زبان‌های رایج نیز به راحتی قابل ترجمه به یکدیگر نیستند [۴]. گفته شده است که ۹۵٪ کودکان کر و لال از والدین سالم متولد می‌شوند 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 می‌افتد که مشکلات زیادی را برای این کودکان به دنبال دارد [۴].

^۱ Gesture recognition

^۲ Human Computer Interface (HCI)

^۳ Grafical User Interface (GUI)

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.

۱-۸- تعریف مسئله

با توجه به موارد کاربرد زبان اشاره که در بخش های پیشین 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^۱، ترجمه زبان اشاره و ... 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 بپردازند هنوز در حال توسعه بوده و نیاز به تکامل دارند.

از آنجا که زبان اشاره توسط دست ها انجام می گیرد و برای شناسایی آن بایستی از دوربین های فیلم برداری و تصاویر ویدئویی استفاده نمود. بخش مهم و اولیه کار، تشخیص و جداسازی دست از دیگر اجزای بدن می باشد که همیشه با چالش هایی همراه بوده است. در روشهای جداسازی دست با استفاده از تصاویری که با دوربین های RGB گرفته شده اند، تشخیص دست با استفاده از پیکسلهایی از تصویر که به رنگ پوست می باشند، صورت می گیرد. در این روش ها می بایست دست حتی الامکان در مقابل صورت

^۱ Human Computer Interaction (HCI)

قرار نگیرد و همچنین پس زمینه می بایست بسیار ساده بوده و رنگ پوست در آن دخیل نباشد. در مقابل با استفاده از دوربین های عمقی، چون دست را با استفاده از عمق آن در تصویر جدا می کنند، نه با استفاده از رنجی از پیکسلها، دست می تواند کاملاً در مقابل صورت نیز قرار گیرد و پس زمینه ی تصویر هر قدر هم که پیچیده باشد منجر به ایجاد خطا نمی گردد.

در شکل ۵-۱ و شکل ۶-۱ نمونه ایی از پایگاه داده هایی که با استفاده از دوربین های RGB به شناسایی زبان اشاره پرداخته است را در مقایسه با پایگاه داده ایی که با استفاده از دوربین های عمقی این کار را صورت داده، نشان داده شده است:



شکل ۵-۱- پایگاه داده برای تشخیص زبان اشاره با استفاده از دوربین معمولی (RGB)



شکل ۶-۱- پایگاه داده برای تشخیص زبان اشاره با استفاده از دوربین عمقی

همچنین، تاکنون از مهمترین مدل زبان اشاره فارسی یعنی مدل باغچه بان در شناسایی الفبای فارسی استفاده نشده است. این مدل که هم اکنون پرکاربردترین مدل در بین ناشنوایان فارسی زبان است، هم شامل اشاره دست و هم شامل شکل دهان است. تاکنون پایگاه داده و کار مشخصی بر روی این مدل از زبان اشاره فارسی ارائه نشده است.

۱-۹- روش حل مسئله

در این پایان نامه جهت تشخیص املای انگشتی حروف الفبای زبان اشاره فارسی، با استفاده از سنسور کینکت به عنوان دستگاه ضبط تصاویر عمقی، حروف زبان اشاره تشخیص داده می شوند. همچنین با امکان سنجی استفاده از دوربین های عمقی (به شکل خاص دستگاه کینکت) برای تشخیص زبان اشاره، اهمیت استفاده از عمق و اسکلت در قطعه بندی دست در پس زمینه پیچیده و استخراج ویژگی های صورت نشان داده می شود. با ارائه روش پیشنهادی جهت استفاده از عمق و اسکلت برای قطعه بندی دست در تصویر رنگی، فرآیند تشخیص زبان اشاره بر روی یک پایگاه داده جدید انجام شده و نتایج بدست آمده با روش های مشابه که بدون استفاده از قطعه بندی عمقی می باشند مقایسه می شود.

همچنین برای اولین بار با استفاده از ویژگی های اسکلت صورت و بدن، دهان نیز از صورت استخراج شده و در پرکاربردترین مدل زبان اشاره فارسی (مدل باغچه بان) به شناسایی الفبای گویای فارسی باغچه بان پرداخته می شود.

لازم به ذکر است که به دلیل عدم وجود پایگاه داده عمقی زبان اشاره فارسی باغچه بان، در این تحقیق بایستی یک پایگاه داده جدید نیز ایجاد و معرفی شود.

فصل ۲: مروری بر

کارهای پیشین

تشخیص خودکار اشاره و زبان های اشاره از حدود سالهای ۱۹۹۵ مورد بررسی قرار گرفته است. محققان روش های بسیاری را جهت تشخیص زبان اشاره توسط رایانه استفاده کرده اند. کارهایی که تا بحال در این زمینه صورت گرفته است را می توان بر اساس فاکتورهای مختلفی از قبیل حوزه شناسایی زبان اشاره، ابزارهای مورد استفاده جهت تشخیص دست و ثبت داده ها، روش های استخراج ویژگی و روش های دسته بندی طبقه بندی نمود.

در زمینه تشخیص زبان اشاره فارسی کارهای زیادی تا به حال صورت نگرفته است و در هیچ یک از این تحقیقات از دوربین های عمقی جهت ثبت تصاویر دست استفاده نشده است. به همین دلیل پایگاه داده ایی نیز شامل تصاویر عمقی جهت تشخیص اشاره تا کنون تهیه نگردیده است. به دلیل اندک بودن کارهای پیشین در زمینه تشخیص زبان اشاره فارسی به بررسی موردی کارهای پیشین در حوزه تشخیص زبان اشاره فارسی پرداخته می شود و سایر کارهای انجام شده بر روی تشخیص اشاره زبان های دیگر را بر اساس سه فاکتور ذکر شده فوق دسته بندی می نمایم و در نهایت جایگاه کار فعلی را در این میان بیان می کنیم.

۲-۱- بررسی موردی کارهای انجام شده جهت تشخیص زبان

اشاره فارسی

در حوزه ی تشخیص زبان اشاره فارسی کارهای زیادی تا بحال صورت نگرفته است. در این بخش به بررسی دو مورد از جدیدترین و جامع ترین مقالاتی که در این زمینه منتشر شده است پرداخته می شود. اگر چه در زمینه ی تشخیص حالات دست از جمله شمارش عددی که انگشتان دست را نشان می دهد نیز کارهایی صورت گرفته است اما به دلیل سادگی روش و عدم ارتباط کار با تشخیص حروف زبان اشاره فارسی به بررسی آنها پرداخته نشده است.

۲-۱-۱- کار کرمی و همکارانش

در این مقاله که در سال ۲۰۱۱ در نشریه Elsevier به چاپ رسیده است، با استفاده از تبدیل موجک و شبکه های عصبی به تشخیص ۳۲ حرف از الفبای زبان اشاره فارسی پرداخته شده است [۹]. تصاویر حروف مورد نظر با استفاده از دوربین دیجیتال و با پس زمینه ایی بسیار ساده و تک رنگ (رنگ مشکی) تهیه شده است. روش کار در این تحقیق بدین صورت است که تصاویر رنگی حروف الفبا پس از برش و تغییر سایز به مد خاکستری تبدیل می کنند و سپس با استفاده از تبدیل موجک دو بعدی تا ۹ سطح، ماتریس ضرایب در سطوح هشت و نه را به عنوان ویژگیهای استخراج شده انتخاب می نمایند. در خاتمه جهت تشخیص زبان اشاره با استفاده از این ویژگی های استخراج شده شبکه عصبی پرسپترون چند لایه (MLP) را جهت دسته بندی حروف آموزش می دهند.

در مورد پایگاه داده مورد استفاده، تنها توضیحی که ارائه شده است، تعداد نمونه های هر حرف است. در این پایگاه داده به ازای هر حرف ۲۰ تصویر تهیه شده و در نتیجه جهت تشخیص ۳۲ حرف الفبای زبان اشاره فارسی از ۶۴۰ نمونه تصویر دست که املای انگشتی حروف الفبا را نشان می دهد، استفاده شده است. در این تحقیق متوسط دقت تشخیص در دسته بندی ۳۲ حرف، ۹۴,۰۶ درصد بیان شده است.

بزرگترین نقطه ضعفی که در این کار وجود دارد سادگی و تک رنگ بودن پس زمینه (رنگ سیاه) تصاویر پایگاه داده می باشد که خود محدودیتی بزرگ را برای کاربر ایجاد می نماید همانطور که در شکل ۱-۲ نمونه ی تصاویر پایگاه داده نشان داده شده است پس زمینه ی دست می بایست به رنگ سیاه بوده و شامل هیچ پیچیدگی و اشیای دیگر نباشد.



شکل ۱-۲- پایگاه داده مورد استفاده در کار کرمی و همکارانش. تصویر از [۹]

با بهره گیری از همین محدودیت دست از تصویر اولیه استخراج می گردد. بدین معنی که ابتدا با میانگین گرفتن از پیکسل های تصویر (تصویر خاکستری) یک حد آستانه^۱ تعیین می شود و سپس با صفر کردن پیکسل هایی از تصویر که مقداری کم تر از این حد آستانه دارا می باشند و حذف سطر و ستون هایی از لبه ی تصویر که دارای مقدار صفر می باشند، تصویر دست از تصویر اولیه برش^۲ داده می شود.

^۱ Threshold

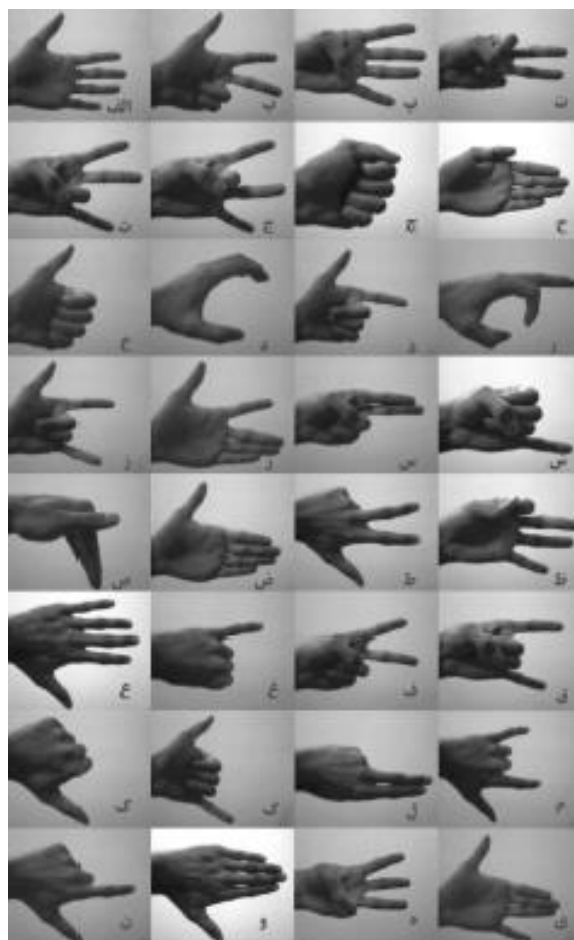
^۲ Crop

درصد تشخیص سیستم برای ۲۲۴ داده تست برابر ۸۳,۰۳ درصد بیان شده و برای کل داده های تست و آموزش که شامل ۶۴۰ نمونه می باشد برابر ۹۴,۰۶ درصد بیان شده است.

۲-۱-۲- کاربرزگر و شرافت

یکی دیگر از کارهایی که در زمینه تشخیص حروف الفبای زبان اشاره فارسی صورت گرفته است، کاربرزگر و شرافت [۱۰] می باشد که در آن از مشخصه رنگ جهت آشکار سازی پوست و استخراج دست استفاده شده و این خود محدودیتی بزرگ برای سیستم در نظر گرفته می شود چرا که رنگ های موجود در پس زمینه نباید شامل رنگ پوست باشند. همچنین جهت استخراج دست از تصاویر اولیه گرفته شده از کاربر بیان شده است که کاربر می بایست از مچ بند رنگی (به غیر از رنگ پوست) استفاده نماید و یا فقط تا مچ دست کاربر در سامانه ایی که تصویر ورودی را تهیه می کند قرار گیرد.

پایگاه داده ی مورد استفاده شامل ۱۶۰۰ نمونه می باشد (به ازای هر یک از ۳۲ حرف الفبا ۵۰ نمونه تصویر وجود دارد). در شکل زیر (شکل ۲-۲) نیز نمونه پایگاه داده مورد استفاده نشان داده شده است:



شکل ۲-۲- حروف الفبای استفاده شده در کار برزگر و شرافت. تصویر از [۱۰].

نتایج حاصل از این سیستم دقت تشخیص ۹۹,۱ درصد را بیان نموده است و زمان لازم برای پردازش تصویر یک حالت از دست ۳۵۰ میلی ثانیه با دسته بند شبکه عصبی می باشد.

یکی از نکات قابل تامل در این کار، عدم تطابق پایگاه داده با هیچ یک از مدل های استاندارد الفبای زبان فارسی و یا انگلیسی است. این به این معنی است که محققان به دلخواه خود چندین کلاس که تفاوت بین کلاسی بارزی نیز با یکدیگر دارند، به عنوان حروف الفبای فارسی در نظر گرفته و نسبت به دسته بندی آن اقدام نموده اند. از آنجا که این حروف با حروف الفبای اشاره فارسی تطابقی ندارند نمی توان از آن ها برای شناسایی زبان اشاره فارسی مستقیما استفاده نمود.

۲-۲- طبقه بندی کارهای پیشین بر حسب سامانه دریافت داده

های ورودی

در این قسمت ابزارهایی که در کارهای مختلف برای ذخیره و دریافت اطلاعات اجرای زبان اشاره مورد استفاده قرار گرفته است بررسی می شود.

برخی از دستگاه ها بر پایه اطلاعات دیداری^۱ مانند انواع دوربین های رنگی، سطح خاکستری^۲ 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 به افزایش کاربری سیستم 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.

۲-۲-۱- استفاده از دستکش و سنسور های شتاب سنج

دستکش ها معمولاً 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^۳ کنند تا با بخش های مختلف بدن مانند صورت و گردن در قطعه بندی اشتباه گرفته نشوند. مورد دیگر استفاده از

^۱ Visual

^۲ Gray-Scale

^۳ Segmentation

دستکش‌ها قرار دادن سنسورهایی بر روی انگشتان می‌باشد که این سنسورها معمولاً از نوع شتاب سنج‌های سه بعدی^۱ هستند. ادغام اطلاعات بدست آمده از تصویر و داده‌های مربوط به حرکت انگشتان و دست در فضا با استفاده از شتاب سنج‌ها می‌تواند ترکیب خوبی از ویژگی‌ها را برای فرآیند شناخت زبان اشاره فراهم سازد. استفاده از شتاب سنج‌ها و سنسورهای قابل نصب بر روی دست و بازو و بدون استفاده از دستکش نیز مانند مورد فوق می‌تواند اطلاعاتی در مورد حرکت دست و یا بازوان به ورودی سیستم بدهد. نمونه‌ای از دستکش‌های شتاب سنج در شکل زیر (شکل ۲-۳) نشان داده شده است.



شکل ۲-۳- تصویری از دستکش‌های رنگی به همراه شتاب سنج و جزئیات شتاب سنج سه بعدی به کار رفته در بازی Copy Cat [۴]

کارهای مختلفی با استفاده از دستکش‌های سنسوری [۱۱، ۱۲] و یا دستکش‌های رنگی [۱۳] انجام شده است. در زیر نیز به بررسی مختصر چند کار با استفاده از سنسور‌ها و دستکش‌ها می‌پردازیم.

یکی از مهمترین پروژه‌ها در راستای شناسایی زبان اشاره، پروژه بازی CopyCat می‌باشد که سال‌ها به عنوان پروژه تحقیقاتی در دانشگاه جورجیا تک^۲ در کشور ایالات متحده آمریکا در حال توسعه می‌باشد [۲، ۱۴-۱۷]. این بازی که در واقع یک بستر^۳ برای جمع‌آوری اطلاعات و داده‌های مورد نیاز تشخیص اشاره

^۱ 3D Accelerometer

^۲ Georgia Tech

^۳ Platform

می باشد که در قالب این بازی به کودکان کر و لال به یادگیری زبان اشاره و همچنین تقویت حافظه کمک می شود.

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.

از دیگر کارهایی که اخیراً با استفاده از سنسور ها انجام شده تحقیق ژانگ^۱ و همکارانش است [۱۸] 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.

وابستگی به استفاده از این سنسور ها 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.

دیگر کاری که جهت تشخیص زبان اشاره و در سال های اخیر از سنسور ها و دستکش ها استفاده کرده است تحقیقی از ساجاکارن^۳ و همکارانش [۱۹] در سال ۲۰۱۱ است. در این کار استفاده از شبکه های حسگری بدن^۴ یا BSN به شناسایی حروف، کلمات و یک جمله زبان اشاره پرداخته شده است. در تصویری که در شکل ۲-۴ مشاهده می کنید، نمایی از دستکش و شبکه حسگری بدن مورد استفاده در این تحقیق

^۱ Xu Zhang

^۲ Electromyogram

^۳ Satjakarn Vutinuntakame

^۴ Body Sensor Networks (BSN)

نمایش داده شده است [۱۹]. 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 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 ۲-۴- نمایش از دستکش و شبکه حسگری بکار رفته در تحقیق ساجاکارن و همکارانش. [۱۹]

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. چرا که 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 [۱۳، ۲۰، ۲۱]. 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 [۲۲] 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 سپس اعمال کردن قطعه بندی با استفاده از یک مقدار آستانه^۵ بر روی رنگ پوست، نتایج خوب و مقاومی^۶ برای دسته بندی اشارات دست به ما ارائه می دهد.

ورودی اصلی کار از یک دوربین فیلم برداری رنگی گرفته شده است و در نهایت نیز اشارات دست دسته بندی شده اند.

از دیگر کارهایی که اخیراً با استفاده از دوربین های رنگی انجام شده است، می توان به کار آشوین^۷ و همکارانش که در سال ۲۰۱۰ در دانشگاه باستن^۸ 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 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،

^۱ Lim Wei Howe

^۲ Hand Segmentation

^۳ Skin Segmentation

^۴ Motion Difference

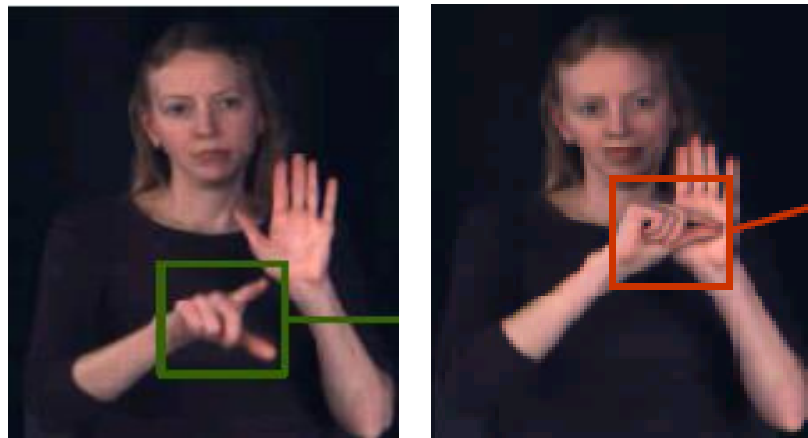
^۵ Threshold

^۶ Robust

^۷ Ashwin Thangali

^۸ Boston University

جستجو بر اساس فریم ابتدایی و انتهایی هر کلمه انجام می شود. به عنوان مثال، در شکل ۲-۵، تصویر چپ مربوط به فریم ابتدایی یک 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 ۲-۵- تصویر ابتدایی (سمت چپ) و تصویر انتهایی (سمت راست) یک کلمه جهت جستجو و دسته بندی کلمه مورد نظر در پایگاه داده در کار آشوبین و همکارانش [۲۳].

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

این دوربین ها عملکرد انسان را در بینایی سه بعدی محیط شبیه سازی می کنند و از دو دوربین که به موازات در کنار هم و رو به جلو قرار گرفته اند تشکیل می شود که به دلیل 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.

^۱ Nearest Neighbor Classifier

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 نمایی از یک دوربین فیلم برداری استریو دیده می شود.



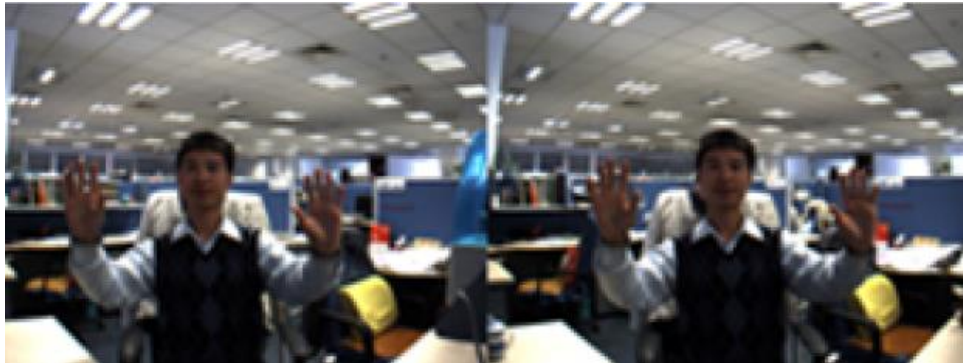
شکل ۲-۶- تصویری از یک دوربین فیلم برداری استریو.^۱

از جدید ترین کارهایی که توسط دوربین های استریو انجام شده است می توان به کار وائگ^۲ و همکارانش اشاره نمود [۲۴]. 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، دستهای وی شناسایی می شوند و سپس برای بهبود الگوریتم از فیلتر رنگ پوست برای جدا سازی نهایی تصویر استفاده می شود. در شکل ۲-۷ 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 هم مشاهده می شود. همانطور که مشخص است، دو تصویر از لحاظ زاویه دید اندکی با یکدیگر اختلاف دارند. در شکل ۲-۸ (چپ)، در بخش اول الگوریتم با شناسایی چهره، بدن فرد موجود در تصویر استریو شکل ۲-۷ استخراج شده است. در شکل ۲-۸ (وسط) مکان دست ها تخمین زده شده اند و در شکل ۲-۸ (راست) با استفاده از فیلتر کردن رنگ پوست قطعه بندی دست 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

^۱ تصویر از www.studiodaily.com

^۲ Qi Wang

بندی بر اساس رنگ پوست، استفاده از آستین های بلند را الزامی می کند که یک محدودیت برای سیستم ارزیابی می شود.



شکل ۲-۷- 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 [۲۴]



شکل ۲-۸- 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 [۲۴].

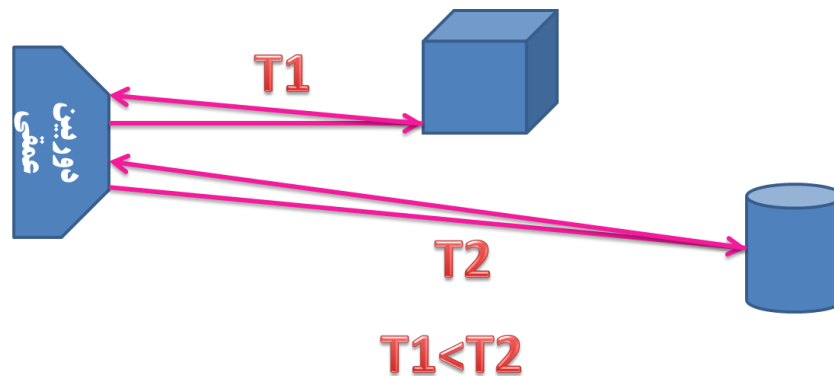
۲-۲-۴- کارهای پیشین با استفاده از دوربین های عمقی

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.

^۱ Time-of-Flight (ToF)

^۲ Infra-Red

در شکل ۹-۲ 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 ۹-۲- 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:

$$distance = c * T/2; \quad (۱-۲)$$

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.

از کارهایی که اخیراً (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 [۲۵]. در این کار با استفاده از یک دوربین زمان پرواز SwissRanger SR2 به بررسی چندین حرکت اشاره از زوایای مختلف 0° ، 45° و -45° پرداخته اند. نتیجه نهایی با استفاده از تفریق فریم های متوالی عمقی 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^۲، بالا بردن دست^۳، دست زدن^۴ و دست تکان دادن^۵ که توسط ۱۰ فرد متفاوت انجام گرفته اند.

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 نمود.

دیگر تحقیقی که در سال ۲۰۰۷ و در آزمایشگاهی در کشور آلمان (آزمایشگاه ZESS^۶) با استفاده از دوربین های عمقی زمان پرواز انجام گرفته است مربوط به کار قبادی^۷ و همکارانش در مورد قطعه بندی دست با دوربین های عمقی و شدت روشنایی^۸ می باشد [۲۶]. در این آزمایش ها، قطعه بندی دست با استفاده از ترکیب دو روش K-Means و EM^۹ انجام گرفته است که به صورت بدون نظارت^{۱۰} انجام گرفته

^۱ M.B. Holte

^۲ Point Right

^۳ Raise Hand

^۴ Clap

^۵ Wave

^۶ Center for Sensor Systems

^۷ S.E. Ghobadi

^۸ Intensity

^۹ Expectation Maximization

^{۱۰} Unsupervised

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.

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 [۲۷]. در این کار قدرت دوربین های عمقی با دوربین های دوبعدی مقایسه شده است. این مقایسه بر روی ۱۲ علامت زبان اشاره و در کل ۳۰۰ نمونه آزمایشی 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

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.

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

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^۴ تشکیل شده است. نمایی از این وسیله را شکل ۲-۱۰ مشاهده می نمایید.

^۱ Arato Sato

^۲ Hidden Markov Model (HMM)

^۳ Resolution

^۴ Tilt Motor



شکل ۲-۱۰- نمایش از سنسور کینکت.

دستگاه کینکت عمق را از فاصله ۰,۸ متر تا ۳,۲ متر با دقت یک میلی متر شناسایی می کند. 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.

در یک پایان نامه کارشناسی که در پاییز ۲۰۱۱ به چاپ رسیده است، لنگ^۱ با استفاده از کینکت و قابلیت تشخیص اسکلت آن به شناسایی کلمات زبان اشاره با استفاده از مدل مخفی مارکوف پرداخته است [۲۸]. 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 بر اهمیت کینکت و دوربین ها عمقی در دریافت تصاویر مادون قرمز یکسان در اکثر شرایط نوری تاکید شده است. این کار مقاوم بودن کینکت در شرایط نوری مختلف و قدرت شناسایی بالا در استفاده از آن را نشان می دهد.

در این قسمت به سه مورد از مهمترین کارهای اخیر در رابطه با زبان اشاره و دستگاه کینکت اشاره می کنیم.

یکی از این کارها، تحقیق زفرولا^۲ و همکارانش است [۲۹]. در این تحقیق که نتایج آن در سال ۲۰۱۱ در یک کنفرانس به چاپ رسیده است، استفاده از کینکت در برابر استفاده از دستکش ها و سنسورهای شتاب سنج در شناسایی کلمات و جملات زبان اشاره مقایسه شده است. همچنین مقایسه بین

^۱ Simon Lang

^۲ Zahur Zafrulla

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 ۲-۱۱ نمایی از حالت ایستاده و نشسته با استفاده از سنسور کینکت را مشاهده می نمایید [۲۹].

با توجه به مقاوم بودن این سنسور به نور محیط، عدم نیاز به دستکش و همچنین درصد شناسایی بالاتر نسبت به استفاده از دستکش و سنسور ها، این تحقیق برتری این سنسور را نسبت به دوربین های رنگی و دستکش ها و شتاب سنج ها نشان می دهد.



شکل ۲-۱۱ - نمایی از حالت ایستاده (راست) و نشسته (چپ) در استفاده از سنسور کینکت در بازی CopyCat. تصویر از [۲۹].

کار مهم دیگری که با استفاده از کینکت انجام شده است، کار پوگولت^۱ و بودن^۲ است [۳۰]. در این کار که نتایج آن در سال ۲۰۱۱ منتشر شده است، ۲۴ حرف از ۲۶ حرف زبان اشاره به شکل هجی انگشتی^۳ دسته بندی شده اند و پایگاه داده ای با بیش از ۴۸۰۰۰ نمونه که از ۴ نفر گرفته شده 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^۴ با استفاده از دوربین عمقی کینکت و مقدار آستانه ای به عمق ۲۰ سانتی متر از تصویر عمقی، بانکی از ۴۰ فیلتر گبور^۵ 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^۶ داده شده است تا عمل دسته بندی با استفاده از آن انجام شود.

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.

یکی دیگر از جدیدترین کارهایی که در زمینه ی تشخیص زبان اشاره با استفاده از سنسور کینکت و تصاویر عمقی صورت گرفته است، تحقیق ارائه شده در [۳۱] است که در قالب پایان نامه جهت اخذ مدرک کارشناسی ارشد تهیه گردیده است. در این تحقیق یک سیستم بلادرنگ تشخیص املای انگشتی حروف الفبای زبان اشاره آمریکایی بر مبنای قطعه بندی عمقی دست از طریق سنسور کینکت ارائه شده است.

^۱ Nicolas Pugeault

^۲ Richard Bowdan

^۳ Finger-spelled

^۴ Hand Extraction

^۵ Gabor

^۶ Random Forest

^۷ Mean Precision

پایگاه داده مورد استفاده در این سیستم، پایگاه داده ای شامل حدود ۶۰۰۰۰ نمونه از ۲۴ حرف الفبای زبان اشاره آمریکایی (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. در واقع این پایگاه داده همان پایگاه داده کار پوگولت و بودن است که در فوق به آن اشاره شد.

نمونه های موجود در پایگاه داده مورد استفاده در این تحقیق در نوع خود پیچیده و منحصر به فرد هستند. که نمونه ایی از آن در شکل ۲-۱۲ نشان داده شده است. کنارهم قرار گرفتن دست، صورت و گردن در تصاویری از پایگاه داده به وضوح مشخص است که همین امر سبب پیچیدگی پس زمینه و لزوم استفاده از عمق جهت قطعه بندی دست شده است.



شکل ۲-۱۲- نمونه پایگاه داده استفاده شده در کار [۳۱]

ویژگی های استخراج شده در این تحقیق شامل ۱۹ ویژگی بر حسب Moment ها و ۷ ویژگی بر حسب لبه بیرونی دست قطعه بندی شده می باشد که در نهایت با استفاده از دسته بندی بیزین به تشخیص ۲۴ حرف الفبای زبان اشاره آمریکایی پرداخته است.

در این تحقیق نتایج مختلفی بر حسب انتخاب های مختلفی از پایگاه داده برای ۲۴ حرف الفبا ارائه شده است. که برای کل پایگاه داده شامل پنج اجرا کننده و ۴۵۰ نمونه از حروف به ازای هر فرد درصد

تشخیص برابر ۷۸,۲۹٪ عنوان شده است و برای یک اجرا کننده با ۱۰۰ نمونه برای هر حرف درصد تشخیص برابر ۹۹,۷۵٪ بیان شده است.

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

استخراج شده

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- استفاده از ویژگی های هندسی

تفاوت اصلی اشارات مختلف بیشتر اوقات در شکل ظاهری این اشارات نمایان است، بنابراین استخراج ویژگی هایی که بر حسب هندسه و شکل ظاهری اشارات باشند می تواند راه حل ساده و مناسبی برای شناسایی زبان اشاره به شمار آید. البته باید این نکته را نیز در نظر گرفت که به دلیل تفاوت در شکل هندسی از زوایای مختلف، ممکن است استخراج صحیح این ویژگی ها و شناسایی صحیح آن 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.

^۱ Geometric Features

^۲ Wavelet Transform

^۳ Edge

^۴ Contour

۲-۳-۱-۱- 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 [۴, ۲۳, ۲۹] اشاره نمود.

۲-۳-۱-۲- ویژگی های برحسب لبه و کانتور

لبه ها در یک تصویر نقاطی هستند که در آن ها تغییرات روشنایی و یا رنگی تصویر شدید است. 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 جزء تصویر، به خصوص اشاره های دست، به طور معمول قابل تشخیص از روی کانتور آن شکل خواهد بود؛ به ویژه زمانی که بافت شیء اهمیت چندانی نداشته باشد.

کارهای پیشین بسیاری از کانتور در تحقیق خود استفاده نموده اند [۳۲, ۳۳]. از روش هایی که در استخراج ویژگی از کانتور استفاده نموده اند می توان به کار مکالا^۳ و همکارانش در سال ۲۰۱۱ اشاره نمود [۳۴]. در این کار تعداد نامشخصی از تصاویر با پس زمینه مشکی شامل ۲۶ حرف الفبای زبان اشاره امریکایی، داده های اولیه تحقیق را تشکیل می دهند. حروف مورد نظر 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.

^۱ Gray-Scale

^۲ Black & White

^۳ Pryanka Mekala

کار دیگری که از لبه ها در تشخیص زبان اشاره استفاده کرده است تحقیق جیتا^۱ و همکارانش [۳۵] در سال ۲۰۱۱ است. در این تحقیق 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. ساده بودن پس زمینه و مشخص نبودن داده های ورودی و برخی تناقضات در تصاویر موجود از ضعف های اصلی این مقاله است.

۲-۳-۲- ویژگی های بر حسب تبدیل موجک

تبدیل موجک کاربردهای بسیاری در حوزه پردازش تصویر سیگنال دارد. در تبدیل موجک دو بعدی گسسته^۲ که در پردازش تصویر کاربرد دارد، یک موجک مادر^۳ 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^۴ اعمال این تبدیل بر روی تصویر، چهار سری از ضرایب^۵ جدید حاصل می شود که هریک به اندازه یک چهارم اندازه تصویر اولیه است. در روش پیشنهادی به صورت کامل در مورد این نوع تبدیل صحبت خواهد شد.

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.

^۱ Geetha M.

^۲ Discrete Wavelet Transform

^۳ Mother Wavelet

^۴ Level

^۵ Coefficient

در کار مِکالا و همکارانش [۳۴] 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^۱ و همکارانش [۳۶] 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 عصبی نیز جهت دسته بندی حروف استفاده شده است. جزئیاتی در مورد تعداد داده های مورد نظر و کیفیت تصاویر داده نشده است. 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. همچنین کار [۹] نیز در شناسایی زبان اشاره فارسی از این تبدیل برای استخراج ویژگی استفاده نموده است که در مورد آن صحبت شده است.

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

به دلیل تغییرات شکل ظاهری دست و بدن و در نتیجه تغییرات محل قرارگیری و زاویه اسکلت آن ها، استفاده از اسکلت دست در شناسایی زبان اشاره می تواند بسیار کاربردی باشد.

از کارهایی که اخیرا (سال ۲۰۱۱) در راستای استخراج اسکلت دست صورت گرفته است می توان به تحقیق ارائه شده در [۳۷] 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.

^۱ Jason Issacs

در تحقیق دیگری که توسط میستری^۱ در دانشگاه MIT انجام گرفته است [۳۸] و در آن دستگاهی به نام حس ششم^۲ ساخته شده است، از نوارهای رنگی که بر روی انگشتان قرار گرفته است استفاده شده است تا مکان دقیق هریک از انگشتان در اسکلت دست و همچنین زاویه آن ها نسبت به یکدیگر تشخیص داده شود. اطلاعات مورد نظر در رابطه با دنیای اطراف در نزدیک ترین سطح خارجی توسط دوربین و ویدئو پروژکتوری که بر روی بدن نصب می شود، در جلوی فرد نمایش داده می شود.

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.

۲-۴- روش های دسته بندی

کارهای مختلف از دسته بند های متفاوتی برای شناسایی داده ها استفاده نموده اند که برخی از آن ها را در زیر بررسی خواهیم کرد:

۲-۴-۱- مدل مخفی مارکوف (HMM)

مدل مخفی مارکوف با ارائه مدلی از مسائل به صورت زنجیره ای از حالت ها به شکلی که هر حالت می تواند با احتمالی خاص به حالتی دیگر برود و یا با احتمالی دیگر یک خروجی تولید کند، در بسیاری از مسائل 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. [۳۱]

^۱ Pranav Mistry

^۲ Sixth Sense

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 [۴، ۲۹، ۳۹-۴۷] 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 بهتر به دسته بند، می توان از قدرت بالای تحلیل 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 [۲۳] 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، [۹، ۳۴، ۳۶، ۴۸-۵۱] 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. [۳۱]

^۱ Bayes Classifier

^۲ A Priori Probabilities

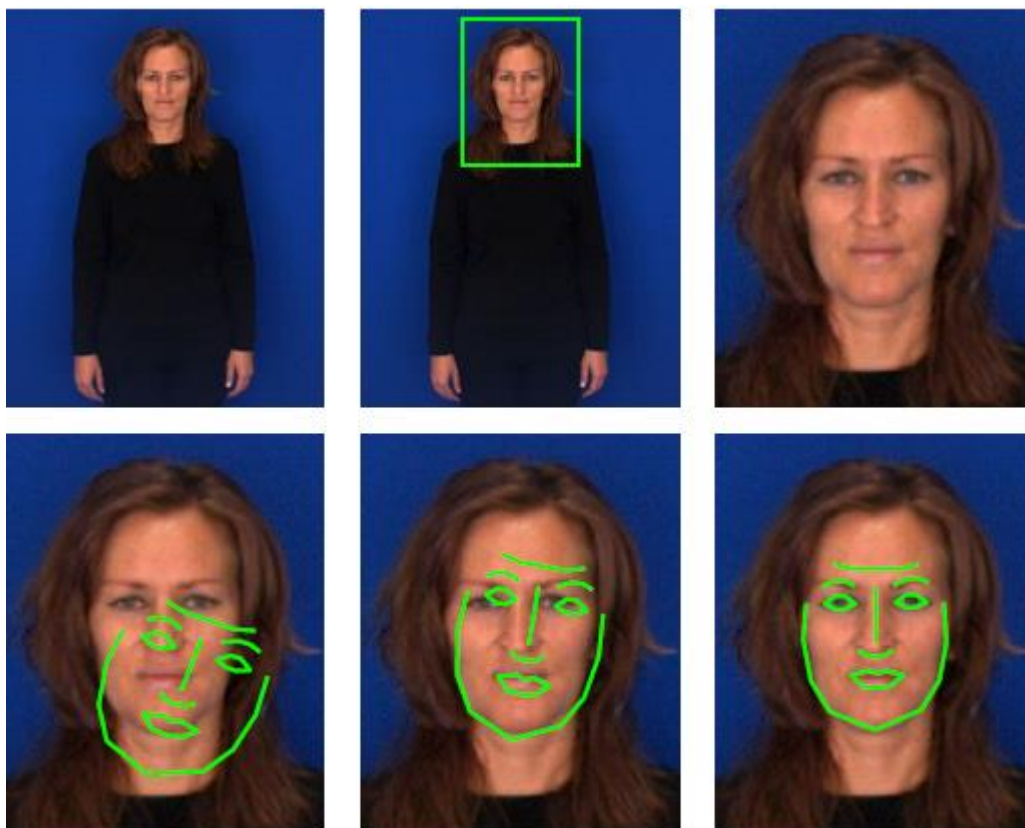
^۳ Black Box

۲-۵- استخراج ویژگی های صورت جهت تشخیص اشاره

از آنجا که در تحقیق پیش رو جهت تشخیص حروف زبان اشاره فارسی از الفبای گویای باغچه بان استفاده شده است و نماد اشاره مربوط به این حروف علاوه بر حالت دست شامل حالت دهان نیز می باشد؛ جهت تشخیص این حروف علاوه بر استخراج دست به تشخیص صورت و استخراج ویژگی هایی از دهان نیز پرداخته شده است. در این بخش مروری بر کارهایی که اخیراً در این رابطه بر روی صورت و روش های استخراج ویژگی از حالات صورت جهت تشخیص اشاره انجام شده اند صورت گرفته است.

در [۵۲] به اهمیت ویژگی های صورت در تشخیص خودکار زبان اشاره پرداخته شده است. در این تحقیق با استفاده از ویژگی های دست و صورت به تشخیص زبان اشاره پرداخته است. به این صورت که جهت اعمال حالات چهره در تشخیص اشاره مورد نظر، ابتدا قسمت هایی از چهره که بیشتر حالات مختلف چهره را نشان می دهند (مانند چشم ها، ابروها و دهان) استخراج شده اند. این کار با استفاده از مدل کردن چهره به وسیله ی یک مدل ظاهری فعال^۱ صورت می گیرد. که مراحل کار در شکل ۲-۱۳ نشان داده شده است.

^۱ Active appearance model

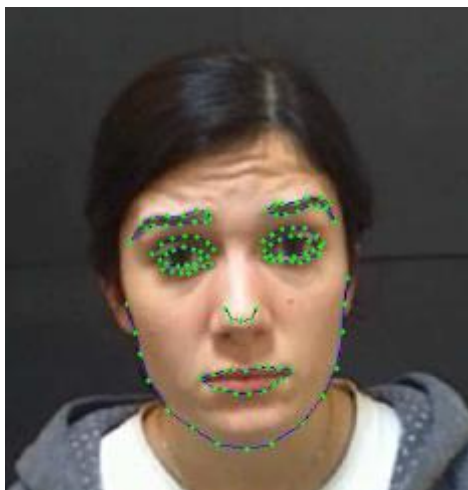


شکل ۲-۱۳- مراحل کار جهت مدل کردن چهره به وسیله ی یک مدل ظاهری فعال در [۵۲]

نتایج حاصل از این تحقیق نشان می دهد که هنگامی که تنها از حالات دست جهت تشخیص کلمات مورد نظر استفاده می شود، درصد تشخیص ۸۶,۷ بیان شده است و هنگامی که از هر دو حالات چهره و دست استفاده شده است درصد تشخیص برابر ۸۸,۳ بوده است.

در کار [۵۳] نیز با استفاده از استخراج ویژگی های صورت به تشخیص زبان اشاره پرداخته است که به ازای تشخیص ۷ علامت اشاره زبان ترکی (که تنها شامل حالات صورت می باشد) به درصد تشخیص بالای ۹۰ درصد رسیده است. هدف از این تحقیق تشخیص نقاط برجسته ایی از چهره^۱ می باشد که بیانگر حالات چهره می باشند. با استخراج این نقاط که در شکل ۲-۱۴ نشان داده شده است و دنبال کردن آنها به تشخیص اشاره ی مورد نظر که با استفاده از حالات صورت نشان داده شده است، پرداخته است.

^۱ Facial feature points



شکل ۲-۱۴- نقاط برجسته ی استخراج شده از صورت در کار [۵۳]

بردار ویژگی ها با استفاده از حداکثر فاصله ی نقاط برجسته ی صورت (landmark) تشکیل شده است و سپس با دادن این بردار ویژگی ها به دسته بند ماشین بردار پشتیبان (^۱SVM) اشارات مورد نظر ، دسته بندی شده اند. در نهایت نتیجه ایی که از این تحقیق حاصل شده است نشان می دهد که ترکیب حالات دست و صورت منجر به رنج تشخیص بالاتری در زبان اشاره گردیده است.

در کار [۵۴] جهت تشخیص چهار حالت رایج چهره در زبان اشاره آمریکایی یک شمای ردیابی مناسب برای نقاط برجسته صورت در ویدیوهای زبان اشاره ارائه شده است. دستاورد اصلی این کار توانایی در ردیابی نقاط برجسته صورت در مواقعی که تداخل^۲ چهره با دست به وجود می آید، بیان شده است. در این تحقیق از ردیاب^۳ KLT به عنوان ردیاب اولیه نقاط برجسته صورت استفاده شده است اما چون این ردیاب به تنهایی نمی تواند در مقابل حرکات سریع دست و تداخلات دست و صورت به درستی نقاط برجسته صورت را ردیابی نماید، مکانیزمی ارائه شده که این مشکل را برطرف نموده است. همانطور که در شکل زیر نشان داده شده است هنگامی که دست نیز در مقابل صورت قرار می گیرد این سیستم توانسته

^۱ Support Vector Machine

^۲ Occlusion

^۳ Kanade-Lucas-Tomasi Feature Tracker

است نقاط برجسته ی صورت را استخراج و ردیابی نماید. نقاط آبی، نقاطی است که با استفاده از سیستم پیشنهادی استخراج و ردیابی شده اند و نقاط قرمز نقاطی است که با استفاده از KLT ردیابی شده اند.



شکل ۲-۱۵- ردیابی نقاط برجسته صورت توسط روش پیشنهادی در [۵۴]. نقاط آبی، نقاطی است که با استفاده از سیستم پیشنهادی استخراج و ردیابی شده اند و نقاط قرمز نقاطی است که با استفاده از KLT ردیابی شده اند. اگرچه روش پیشنهاد شده در این کار توانسته است نقاط برجسته ی صورت را استخراج و ردیابی نماید اما پیچیدگی زیاد روش بیان شده و محاسبات انجام شده جهت رسیدن به این هدف از معایب این کار می باشد. در صورتی که جدا سازی تداخلات دست و صورت با استفاده از عمق و استفاده از ویژگی های اسکلتی بدن و صورت چنین پیچیدگی و سرباری را ندارند.^۱

۲-۶- نتیجه گیری

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 سامانه ورودی جهت دریافت داده ها در تشخیص زبان اشاره، سنسور کینکت یا ابزار های مشابه می باشد که علاوه بر ارائه سیستم رابط کاربری طبیعی^۲ 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.

^۱ سربار این قسمت بر روی سخت افزار کینکت قرار داده شده است و ورودی نرم افزار با سرعت ۳۰ فریم بر ثانیه می تواند تصاویر را دریافت نماید.

^۲ Natural User Interface (NUI)

جهت تشخیص و استخراج دست، استفاده از ویژگی های اسکلتی حاصل از کینکت می تواند بسیار موثر باشد. همچنین استخراج دهان و ویژگی های لب از صورت در تشخیص زبان اشاره فارسی باغچه بان با استفاده از نقاط سه بعدی استخراج شده از تصویر عمقی صورت ویژگی های مناسبی جهت تشخیص زبان اشاره می باشد. 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.

روش های احتمالاتی به ویژه دسته بند بیز نیز جزء بهترین دسته بندها برای تشخیص زبان اشاره می باشد چرا که علاوه بر سادگی در پیاده سازی، 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.

فصل ۳: روش پیشنهادی

همانطور که با بررسی کارهای پیشین می توان نتیجه گرفت، استفاده از سنسور کینکت می تواند امکانات خوبی برای جدا سازی تصویر از پس زمینه و همچنین استخراج اسکلت دست و صورت به ما ارائه دهد. در این بخش روش پیشنهادی بر پایه پایگاه داده ایجاد شده برای حروف الفبای زبان اشاره باغچه بان، استفاده از داده های عمقی، استخراج ویژگی های هندسی، اسکلتی و بر پایه تبدیل موجک و در نهایت بر پایه دسته بندی با دسته بند بیز توضیح داده می شود.

۳-۱- پایگاه داده

برای تست نتایج بدست آمده از روش پیشنهادی، بایستی روش پیشنهادی بر روی یک پایگاه داده استاندارد تست شود. متأسفانه هنوز پایگاه داده استاندارد برای زبان اشاره فارسی ایجاد نشده است. کارهای پیشین نیز همگی از پایگاه داده های نامعلوم با تعداد نمونه های نامعلوم و بدون ارجاع به مکانی برای ارائه پایگاه داده استفاده نموده اند که هیچ کدام قابل استفاده مجدد و دسترسی نمی باشند. به همین دلیل در این بخش به ساخت یک پایگاه داده بر پایه سنسور عمقی کینکت و همچنین الفبای باغچه بان به عنوان اولین و پرکاربرد ترین زبان اشاره فارسی پرداخته می شود.

تاکنون و قبل از انجام این پایان نامه پایگاه داده ای که به صورت کامل از ۳۸ حرف الفبای زبان اشاره فارسی باغچه بان تولید شده باشد ساخته نشده است. همچنین از دوربین های عمقی نیز برای این منظور استفاده نشده است که در نهایت تصمیم به ساخت چنین پایگاه داده ای با این مشخصات توسط سنسور کینکت گرفته شد.

۳-۱-۱- کینکت و ویژگی های آن

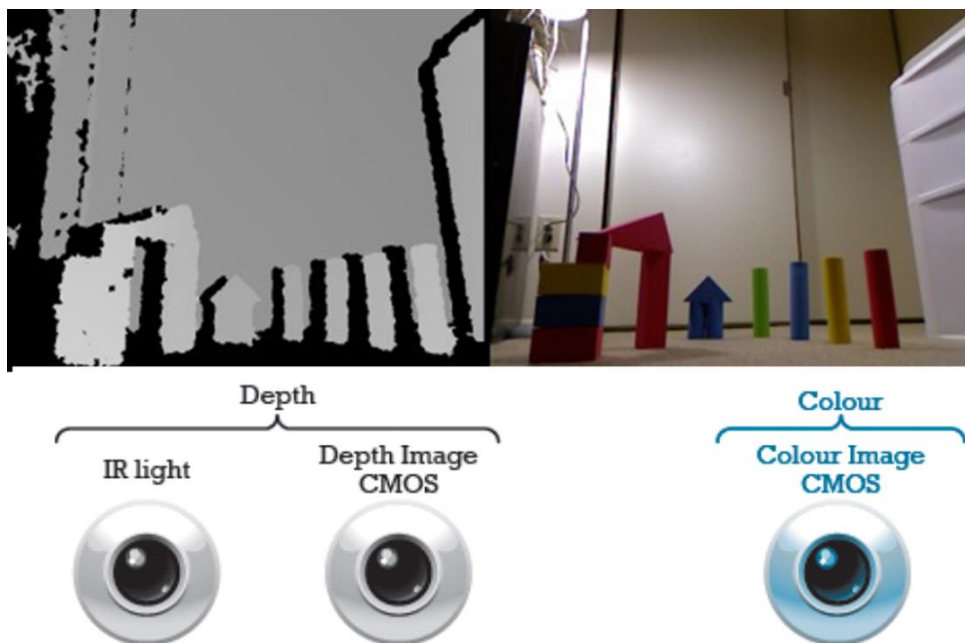
همانطور که گفته شد، دستگاه کینکت عمق را از فاصله ۰٫۸ متر تا ۳٫۲ متر با دقت یک میلی متر شناسایی می کند و قابلیت تشخیص ۲۰ نقطه از اسکلت بدن را دارا می باشد.

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 [۵۵] با قیمت ۱۴۹ دلار وارد بازار گردید و با استقبال فراوانی مواجه و طولی نکشید که ۸ میلیون دستگاه از این وسیله فروخته شد.

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 گردیده است. در شکل ۱-۳ 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.

^۱ Project Natal

^۲ Gamepad/Joystick



شکل ۳-۱- نمایش 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.

۳-۱-۱-۱- ویژگی‌های اسکلتی

با استفاده از SDK که شرکت Microsoft برای این ابزار ارائه نموده است، می‌توان ۲۰ نقطه بدن هر اسکلت را با مختصات ۳ بعدی استفاده نمود. در هر فریم، لیست حداکثر ۶ اسکلت (فرد) موجود و در صورت وجود هر یک، ۲۰ نقطه از اسکلت و وضعیت ردیابی آن‌ها (ردیابی شده، ردیابی نشده و تخمین زده شده) در اختیار برنامه نویس قرار می‌گیرد. در حالت نشسته نیز ۱۰ نقطه اسکلت داده می‌شود. در شکل ۳-۲، نمایش از نقاط استخراج شده مشاهده می‌شود. به ازای هر نقطه (که نام آن نیز مشخص است)، مختصات دو بعدی در تصویر عمقی (X, Y)، عمق (Z)، و وضعیت شناسایی آن ارائه می‌شود. در این تحقیق از مختصات Hand Right برای پیدا کردن نقطه دست راست استفاده شده است. قابل توجه است که تصویر ضبط شده آینه‌ای^۱ شده است و به نظر می‌رسد که دست چپ و راست جابجا شده‌اند که در واقعیت اینطور نیست.

^۱ Mirrored



شکل ۳-۲- نمایشی از اسکلت استخراج شده توسط سنسور کینکت.

۳-۱-۱- ویژگی های صورت

در نسخه های ۱,۵ به بعد از SDK شرکت Microsoft برای Kinect، امکان استخراج ۱۲۱ نقطه از صورت به صورت سه بعدی فراهم شده است. در این حالت، پس از تشخیص اسکلت، این ۱۲۱ نقطه نیز برای استفاده از نقاط صورت قابل استفاده است. در شکل ۳-۳ نقاط استخراج شده از صورت را مشاهده می نمایید. به ازای هر نقطه، مختصات دو بعدی در تصویر عمقی (X, Y) و عمق (Z) ارائه می شود.



شکل ۳-۳- نمایش از ۱۲۱ نقطه استخراج شده از صورت که به صورت شبکه ای بر روی صورت منطبق شده است.

۳-۱-۲- دلیل انتخاب حروف الفبای گویای باغچه بان

در این بخش به توضیح در مورد دلایل ایجاد پایگاه داده حروف زبان اشاره فارسی بر اساس الفبای گویای باغچه بان پرداخته می شود.

طبق تاریخچه ایجاد زبان اشاره فارسی مدل های متفاوت از الفبای زبان اشاره فارسی توسط افراد مختلف ایجاد شده است [۶]. با توجه به توضیحات مربوط به تاریخچه ی ایجاد الفبای زبان اشاره فارسی که در مقدمه به آن اشاره شد، در نهایت دلایل تهیه پایگاه داده بر اساس مدل الفبای گویای باغچه بان در ذیل عنوان شده است.

در کارهای قبلی صورت گرفته [۹، ۱۰، ۵۶] جهت تشخیص حروف الفبای زبان اشاره فارسی از مدل های دیگر الفبای زبان اشاره فارسی استفاده شده که شامل نمای دهان و صورت در املای این حروف نمی باشند. این در صورتی است که اصلی ترین و رایج ترین مدل زبان اشاره که توسط ناشنوایان در ایران استفاده

می گردد الفبای گویای باغچه بان می باشد. اینکه هیچ یک از سه الفبای جدید در کشور ما کاربردی نشده اند را می توان از تایخچه و سایت بنیاد پژوهش های ناشنویان ایران [۶] استنباط نمود. این مهم نیز توسط اینجانب از جانب یکی از ناشنویان در اداره ی بهزیستی شهرستان شاهرود مورد تحقیق قرار گرفت. بر اساس نتایج بدست آمده از گفتگو ها و تحقیقات انجام شده، کل ناشنویان این شهرستان و ایران از زبان اشاره ابداع شده توسط جبار باغچه بان استفاده می کنند. یکی از کتاب های اصلی در این زمینه نیز که توسط انتشارات بهزیستی منتشر شده است (شامل ۴ جلد) نیز بر پایه همین الفبا می باشد. [۷]

از دیگر دلایل استفاده از مدل الفبای گویای باغچه بان برای ایجاد پایگاه داده حروف زبان اشاره، دخیل بودن صورت و حالات دهان علاوه بر حالات دست می باشد. همانطور که گفته شد، در کلمات زبان اشاره فارسی، علاوه بر اشاره دست، حالات صورت نیز حائز اهمیت بوده و نقش مهمی در معنای کلمه ادا شده ایفا می کنند. به این صورت که خیلی از حروف و کلمات این زبان، فقط در شکل دهان تفاوت دارند و شکل دست آن ها یکی می باشد. توجه به این نکته باعث می شود که جهت ادامه کار فعلی برای تشخیص کلمات زبان اشاره فارسی، به راحتی بتوان سیستم فعلی را بسط داده و برای کاربری کردن نهایی سیستم، پایگاه داده و سیستم تشخیص کلمات را با کمترین تغییر بتوان پیاده سازی نمود. با وجود مهم بودن این جمله، کارهای مهم در زمینه تشخیص زبان اشاره فارسی، تنها مدل هایی ها انتخاب کرده اند که دهان در آن ها تداخل نداشته و در نتیجه پیچیدگی کمتری در عین کاربر کمتر برای سیستم های پیشنهادی مورد نیاز بوده است. [۹، ۱۰، ۱۵۶].

۳-۱-۳- پایگاه داده عمقی الفبای گویای باغچه بان

برای تهیه پایگاه از یک دستگاه کینکت مخصوص Xbox 360 استفاده شده است. برای تصویر رنگی از Resolution640X48030FPS و برای تصویر عمقی Resolution640X48030FPS استفاده شده است که تصویر رنگی و عمقی با وضوح تصویر 640 در 480 و با سرعت ۳۰ فریم بر ثانیه ارائه می دهد. از Microsoft SDK نسخه ۱,۸ استفاده شده است. برای ذخیره پایگاه داده، به ازای هر ۱۰ حرف، یک فایل مجزای xed

توسط نرم افزار KinectStudio ذخیره شده است. برای هر حرف به طول متوسط حدود ۱۰۰۰ مگابایت و برای کل ۳۸ حرف الفبا به ازای یک نفر ۳۵ گیگا بایت از فضای دیسک اشغال شده است.

سوژه در فاصله حدود یک متری بر روی یک صندلی قرار گرفته است که سعی شده است که پس زمینه با رنگ های مختلف و با پیچیده گی مناسب ایجاد شود. همچنین از پیراهن با آستین کوتاه و رنگ مشابه دست برای حداکثر پیچیدگی استفاده شده است (شکل ۳-۴).



شکل ۳-۴ نمایی از محیط پیرامون و پس زمینه پیچیده در ایجاد پایگاه داده.

برای امکان جداسازی هر حرف، در هنگام شروع اجرای یک حرف توسط دست راست، دست چپ بالا برده می شود و پس از پایان اجرای هر حرف نیز دست چپ مجدداً بالا برده شده و سپس پایین آورده می شود. همچنین سعی شده است که نورپردازی مناسب بر روی صورت شکل بگیرد که به دلیل ضبط در محیط اتاق سایه کمتری بر روی صورت به دلیل نور های سقفی ایجاد شود. برای این منظور یک لامپ سقفی نیز در فاصله حدود ۵ متری از سوژه قرار گرفته که نور مایل تری به صورت تابیده شود (شکل ۳-۵).



شکل ۳-۵- بالا بردن دست برای مشخص کردن آمادگی برای شروع ضبط یک حرف خاص.

برای اجرای حروف الفبای گویای باغچه بان، شکل های اجرا شده زیر به عنوان الگو به فرد اجرا کننده نشان داده شده است. این اجرا ها از سایت [۵۷] گرفته شده است که توسط ربکا رز اُرتن^۱ در سال ۲۰۱۲ گرفته شده و یک ماه بعد نیز مورد بازبینی قرار گرفته است. (شکل ۳-۶، شکل ۳-۷ و شکل ۳-۸).

^۱ Rebecca Rose Orton



آ



ا



ی



اِ



او



ای



ب



پ



ت



ث



ج



چ



ح



خ



د

شکل ۳-۶- حروف آ تا د از اجرای الفبای گویای باغچه بان در زبان اشاره فارسی



ذ



ر



ز



ژ



س



ش



ص



ض



ط



ظ



ع



غ



ف



ق



ک

شکل ۳-۷- حروف ذ تا ک از اجرای الفبای گویای باغچه بان در زبان اشاره فارسی

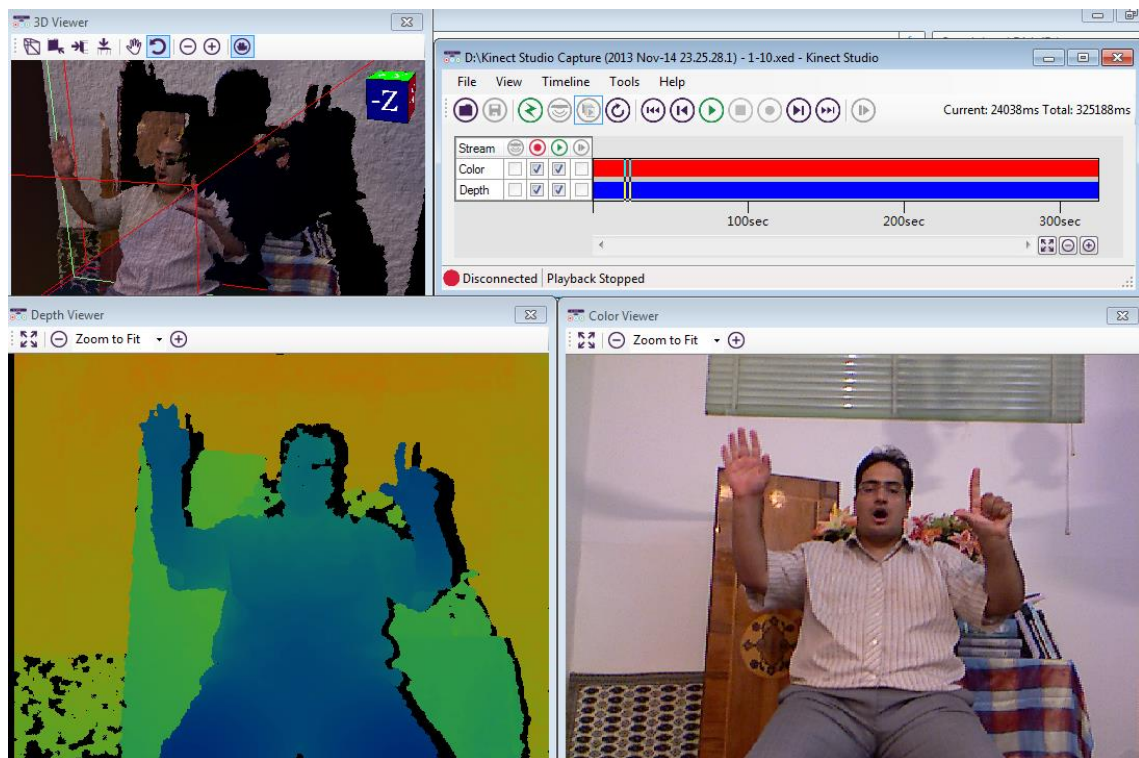


شکل ۳-۸- حروف گ تا ء از اجرای الفبای گویای باغچه بان در زبان اشاره فارسی

۳-۱-۴- جدا سازی حروف

برای جداسازی حروف از فایل های ذخیره شده xed، ابتدا این فایل های توسط نرم افزار Kinect Studio باز می شوند. نمونه ای از نمایش و باز شدن فایل ها را در شکل ۳-۹ مشاهده می نمایید. این نرم افزار از چهار بخش نوار زمان^۱، فریم رنگی، فریم عمقی (که برای نمایش بهتر عمق رنگ آمیزی شده است) و محیط سه بعدی حاصل از ترکیب تصویر عمقی و تصویر رنگی تشکیل شده است. سه بخش آخر برای نمایش و بخش اولی برای پخش داده ها در نرم افزار های دیگر مورد استفاده قرار می گیرد.

^۱ Timeline



شکل ۳-۹- نمونه ای از نرم افزار ضبط پیوسته داده های کینکت، Kinect Studio.

پس از باز کردن فایل های پایگاه داده در نرم افزار، نرم افزار جدا کننده پایگاه داده که به این منظور نوشته شده است را اجرا و مکان ذخیره فایل های پایگاه داده به صورت مجزا را مشخص می کنیم. این نرم افزار واسط ویژگی های زیر را دارا می باشد:

۱. نرم افزار با نمایش شکل صحیح کلمات، و لیست مرتب حروف الفبا، امکان ذخیره یک به یک حروف را فراهم می آورد.

۲. نرم افزار به ازای هر فرد، به ازای هر حرف الفبا، پوشه ای با شماره مورد نظر در پوشه وارد شده ایجاد می کند. در هر پوشه، ۴ فولدر به نام های زیر ایجاد می شود:

- a. Color
- b. Depth
- c. Skeleton

d. Face

به ازای هر فریم دریافت شده، در هریک از این پوشه ها، یک شماره خاص برای هر فایل ایجاد می شود. در پوشه Color تصویر PNG فریم فعلی، در پوشه Depth تصویر عمقی با عمق ۱۶ بیتی و فرمت ushort به صورت فایل PNG، در پوشه Skeleton، لیست تمام نقاط اسکلت بدن به صورت نام نقاط بدن و سه نقطه X، Y و Z هریک در یک خط مجزا آورده شده است. در پوشه Face نیز به ازای هر فریم، ۱۲۱ نقطه صورت به صورت مختصات سه تایی (X, Y, Z) ذخیره شده است.

۳. نرم افزار برای دقت بیشتر، تشخیص اسکلت و نقاط صورت را به فرد نشان می دهد تا فرد از ذخیره قسمت های معیوب پایگاه داده آگاه شود. صورت قرمز رنگ مربوط به صورت، هنگامی که صورت تشخیص داده شود به رنگ سبز در می آید.

۴. به دلیل اینکه حجم فایل های ذخیره شده بسیار بالا می باشد و در صورت عدم استفاده از هارد با سرعت بالا به صورت مجزا سرعت ذخیره هارد پایین می آید، مسئله تولید کننده و مصرف کننده در این نرم افزار حاکم می شود. به این صورت که نرم افزار Kinect Studio سریعتر از سرعتی که نرم افزار ذخیره پایگاه داده فایل ها را ذخیره می کند، فریم تولید می کند. البته این موضوع در حالت استفاده مستقیم از Kinect صادق نمی باشد به دلیل اینکه اطلاعات کینکت از طریق کابل USB به سیستم انتقال می یابد و هارد دیسک آزاد می باشد. از طرفی اطلاعات فقط در حافظه ذخیره می شود و در جایی به صورت دائم ذخیره نمی شود. برای رفع این مشکل در این نرم افزار، از یک صف همروند $\text{ConcurrentQueue}<T>$ استفاده شده است. این صف قابلیت این را فراهم می کند که در چند ترد مجزا، به صف به صورت Safe بتوان دسترسی پیدا نمود بدون اینکه متغیر های آن دچار اختلالات دسترسی چند نخه شود. در این صف به ازای هر فریم یک چند تایی (Tuple) به شکل زیر ذخیره می شود:

```
Tuple<EnumIndexableCollection<FeaturePoint, PointF>, Skeleton, Image<Gray, ushort>, byte[], string, int>
```

این Tuple یک شش تایی است که به ترتیب از چپ به راست نقاط صورت، جلوترین اسکلت

تشخیص داده شده، تصویر عمقی، بایت های تصویر رنگی، آدرس فعلی پوشه حرف الفبا و

شماره فعلی فریم ذخیره شده را در هر فریم در خود نگه می دارد.

استفاده از این صف موجب شده است که فریم های تولیدی از دست نروند و مشکل تداخل در

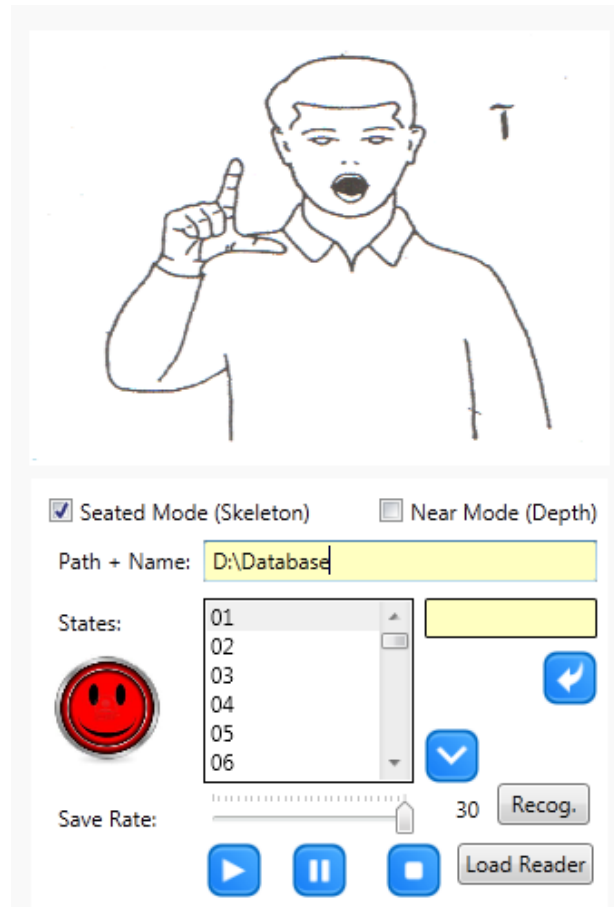
اطلاعات ذخیره شده نیز پیش نیاید و یک چهارتایی از اطلاعات در مکان درست ذخیره شود.

در شکل ۳-۱۰، نمایی از بخشی از نرم افزار را مشاهده می نمایید و در شکل ۳-۱۱ نمایی از سلسله

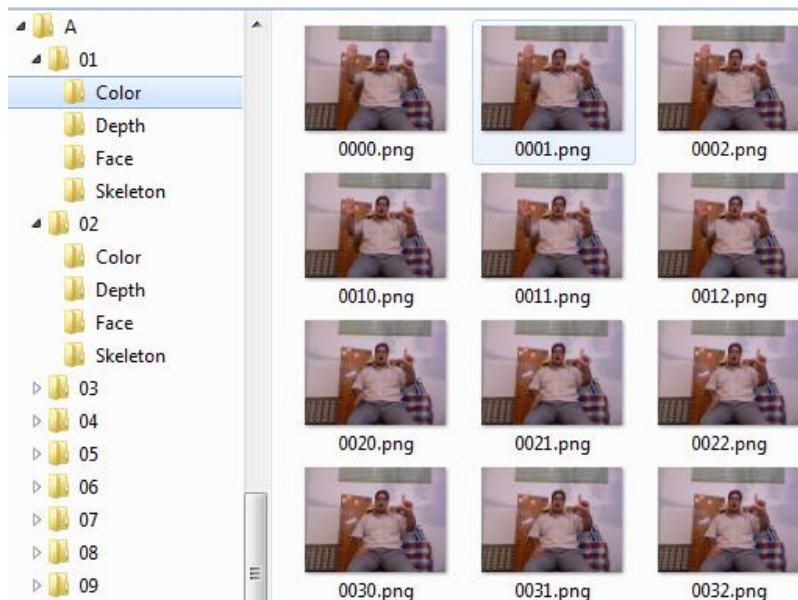
مراتب پوشه های ایجاد شده پایگاه داده جدا شده به ازای هر حرف. پس از شروع فرآیند ذخیره سازی، با

انتخاب پوشه ذخیره پایگاه داده و انتخاب حرف فعلی، در پوشه مورد نظر ۴ پوشه به ازای هر حرف ایجاد می

شود که نمایی از خروجی آن را در تصاویر مشاهده می نمایید.



شکل ۳-۱۰- نمایش از نرم افزار نوشته شده برای جداسازی حروف از یکدیگر.



شکل ۳-۱۱- نمایش از پوشه های ایجاد شده برای پایگاه داده به ازای حروف مختلف و داده های ذخیره شده هر حرف.

۳-۱-۵- مشخصات پایگاه داده نهایی

پایگاه داده نهایی ایجاد شده به دلیل تفاوت زمانی در ضبط هر حرف (تلاش شد هر حرف ۱۰ تا ۱۵ ثانیه اجرا شود)، تفاوت نرخ ضبط اولیه (فریم بر ثانیه) در هنگام ذخیره و مشکلات حافظه ای احتمالی در زمان تبدیل پایگاه داده دارای تعداد نمونه های متفاوتی از هر حرف می باشد. دو مورد آخر مربوط به محدودیت سخت افزاری می باشد و در آخرین مورد فوق با استفاده از یک سیستم سریعتر می توان به تعداد داده های بیشتر بر روی همین پایگاه داده دست یافت. در نهایت تعداد نمونه های هر حرف و شماره آن ها در جدول ۳-۱ آمده است. حداقل تعداد نمونه ۱۵۹ و حداکثر تعداد نمونه برای یک حرف ۵۶۲ نمونه می باشد. پر واضح است که حداکثر تعداد نمونه قابل استفاده از هر حرف برای شناسایی استاندارد، حداقل تعداد نمونه حروف، یعنی ۱۵۹ نمونه از هر حرف الفبا می باشد. تعداد کل نمونه ها، ۱۱۹۹۰ نمونه می باشد که به طور متوسط ۳۱۵ نمونه به ازای هر حرف الفبا می باشد. در حالت حداقلی، پایگاه داده ۶۰۴۲ نمونه با تعداد مساوی برای هر حرف را شامل می شود.

جدول ۳-۱- لیست حروف موجود در پایگاه داده و تعداد نمونه از هر حرف.

شماره حرف	حرف	تعداد نمونه	شماره حرف	حرف	تعداد نمونه
۰۱	آ	۳۴۳	۲۰	م	۳۷۹
۰۲	ا	۲۴۶	۲۱	ن	۵۶۲
۰۳	ع	۳۳۵	۲۲	اُ	۳۱۰
۰۴	ب	۳۱۵	۲۳	او	۳۶۰
۰۵	چ	۲۵۸	۲۴	پ	۴۲۲
۰۶	د	۲۲۹	۲۵	ر	۳۶۰

۲۲۵	ث	۲۶	۱۵۹	اِ	۰۷
۳۳۶	ص	۲۷	۲۱۶	ای	۰۸
۵۱۰	ش	۲۸	۲۱۷	ف	۰۹
۳۸۳	س	۲۹	۱۷۰	گ	۱۰
۳۴۳	ت	۳۰	۴۷۲	ق	۱۱
۴۴۸	ط	۳۱	۲۸۷	غ	۱۲
۴۲۸	و	۳۲	۲۹۲	ح	۱۳
۲۷۹	ی	۳۳	۲۷۶	ه	۱۴
۲۱۹	ز	۳۴	۲۰۹	ج	۱۵
۲۷۰	ظ	۳۵	۳۱۹	ژ	۱۶
۴۰۱	ض	۳۶	۱۷۱	ک	۱۷
۴۱۱	ذ	۳۷	۲۷۰	خ	۱۸
۲۷۹	ء	۳۸	۲۸۱	ل	۱۹

۳-۱-۶- مقایسه پایگاه داده با پایگاه داده های موجود

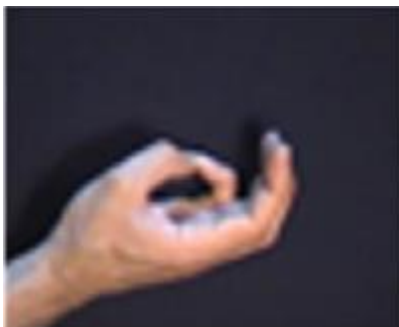
پایگاه داده ایجاد شده در این تحقیق از جهات بسیاری با پایگاه داده های معرفی شده در کارهای پیشین در تشخیص حروف الفبای زبان فارسی برتری داشته که در زیر به آن ها اشاره می شود:

۱. در تمامی کارهای قبلی که در این زمینه به تشخیص حروف الفبای زبان اشاره فارسی پرداخته

اند [۹، ۱۰، ۵۶]، تصاویر حروف زبان اشاره با پس زمینه بسیار ساده و با رنگ متفاوت با رنگ

پوست بوده است. این عدم پیچیدگی با تصویر برداری مجزا، دقیق و یک به یک از تصاویر دست

همراه بوده است. در صورتی که در پایگاه داده ایجاد شده در این پروژه، علاوه بر پیچیدگی پس زمینه، تداخل دست با صورت، بدن، لباس با تشابه رنگی و پس زمینه پیچیده نیز لحاظ شده است که در یک محیط نزدیک به محیط واقعی از کاربر تصویر برداری شده است. در شکل ۱۲-۳ تفاوت پیچیدگی پس زمینه را در کار پیشنهادی و دیگر کارهای فارسی مشاهده می نمایید.



شکل ۱۲-۳- مقایسه پیچیدگی پس زمینه تصاویر در روش پیشنهادی و دیگر روش های تشخیص زبان اشاره فارسی. تصویر بالا مربوط به پایگاه داده روش پیشنهادی، تصاویر پایین مربوط به پایگاه داده دیگر کارهای فارسی (به ترتیب از راست به چپ [۱۰] و [۹])

۲. در هیچ یک از کارهای قبلی توضیحی در مورد اینکه پایگاه داده از قبل وجود داشته است و یا خود افراد اقدام به تهیه آن نموده اند داده نشده است و همچنین نسبت به درج جزئیات بیشتر، کل پایگاه داده و همچنین لینکی برای دانلود پایگاه داده اجتناب نموده اند. به همین دلیل، مخصوصا در برخی از کارهای پیشنهادی، حتی امکان مقایسه تعداد نمونه ها به دلیل عدم ذکر آن ها وجود نداشته و تعداد افراد دخیل در ساخت پایگاه داده مشخص نشده است. این در حالی

است که پایگاه داده فعلی با جزئیات دقیق و تحلیل بخش ها مختلف آن ارائه گردیده است و در صورت نیاز محقق در اختیار آن ها قرار می گیرد. پر واضح است که در تماس با نویسندگان برای ارسال پایگاه داده پاسخی دریافت نشده است.

۳. یکی از بزرگترین نقایص پایگاه داده های استفاده شده در کارهای قبلی این است که هیچ یک از آن ها، از اصلی ترین، اولین و رایج ترین مدل زبان اشاره (الفبای گویای باغچه بان) استفاده ننموده و شاید به دلیل پیچیدگی کمتر در دیگر مدل ها و عدم وجود صورت و دهان در تشخیص، به سادگی از کنار بحث کاربری بودن تحقیق برای ناشوایان عزیز در آینده گذشته اند. این درحالی است که در تحقیق فعلی، از تمامی قسمت های بدن و به ویژه صورت و دهان، تصاویر و ویژگی های مورد نظر استخراج و طبق الفبای گویای باغچه بان پایگاه داده تهیه شده است.

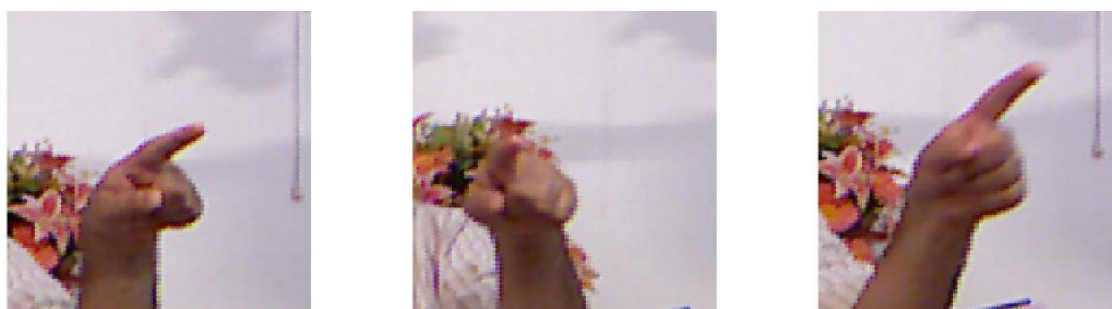
۴. پایگاه داده فعلی تمامی حروف الفبای فارسی که ۳۸ حرف می باشد را پوشش داده است در حالی که دیگر کارها تنها ۳۲ حرف آن را پوشش داده اند. [۹، ۱۰، ۵۶]

۵. تعداد نمونه های پایگاه داده جمع آوری شده، از تعداد نمونه های دیگر پایگاه داده ها بیشتر است. تنها یکی از کارهای پیشین تعداد نمونه های پایگاه داده مورد استفاده خود را اعلام نموده است [۹] که شامل تنها ۲۰ نمونه از هر حرف و در کل ۶۴۰ نمونه برای ۳۲ حرف می باشد. این در حالی است که پایگاه داده فعلی در حداقل حالت ۶۰۴۲ نمونه در کل و ۱۵۹ نمونه از هر حرف را دارا می باشد.

۳-۱-۷- نمونه های پایگاه داده

نمونه های استخراج شده از پایگاه داده با توجه به حرکت دست و سر در هنگام ضبط دارای نمونه های متفاوت مکانی و زاویه ای هستند. همچنین همانطور که از الفبا مشخص است، تعدادی از حروف فقط

در شکل ظاهری دهان تفاوت داند و شکل دست در آن ها کاملاً یکسان است. برای مثال در شکل ۱۳-۳ سه نمونه از حرف "ذ" نشان داده شده است که تفاوت درون کلاسی^۱ در این پایگاه داده را نشان می دهد. شکل ۱۴-۳ شباهت بین کلاسی^۲ نمونه های حرف س و ش را نشان می دهد که تنها در شکل دهان با یکدیگر اختلاف دارند. و همچنین همانطور که از تصاویر نیز مشخص است، پس زمینه با دست و صورت تداخل رنگی دارد.



شکل ۱۳-۳- سه تصویر از حرف ذ نشان دهنده تفاوت درون کلاسی نمونه های پایگاه داده.



شکل ۱۴-۳- شباهت بین کلاسی در حرف ش (سمت راست) و حرف س (سمت چپ). تفاوت دو حرف تنها در شکل دهان است.

^۱ Inner-class

^۲ Between-class

۳-۲-پیش پردازش

جهت آماده سازی تصاویر برای استخراج ویژگی، چند مرحله برای پیش پردازش بر روی تصویر انجام شده است. این مراحل به تفصیل در زیر شرح داده شده اند.

۳-۲-۱-برش تصاویر دست

پس از ذخیره پایگاه داده، اولین مرحله برش تصویر دست است. برای اینکار، از آنجا که مختصات دست سمت راست قبلاً توسط سیستم تشخیص اسکلت شناسایی شده است، به میزان ۷۰ پیکسل در چهار جهت اطراف دست به عنوان تصویر دست قرار می دهیم و دست را جدا می کنیم. این کار هم بر روی تصویر عمقی و هم بر روی تصویر رنگی اعمال می شود. در شکل ۳-۱۵ تصویری از دست برش داده شده از یک فریم حرف آ مشاهده می نمایید.



شکل ۳-۱۵-برش دست با استفاده از مختصات اسکلتی دست به فاصله ۷۰ پیکسل از هر جهت.

۳-۲-۲- قطعه بندی دهان

یکی از بخش های مهم در شناسایی الفبای زبان اشاره فارسی، شکل ظاهری دهان می باشد. چندین حرف الفبای زبان اشاره فارسی فقط در شکل دهان متفاوت هستند و راه دیگری برای تمیز دادن آن ها از یکدیگر وجود ندارد. به همین دلیل بایستی دهان را نیز در بین ویژگی های سیستم در نظر گرفت. برای این منظور، با استفاده از نقاط صورت، نقاط اطراف دهان را تشخیص و دهان را از تصاویر جدا می کنیم.

از بین ۱۲۱ نقطه صورت، نقاط لیست شده در جدول ۳-۲ که مربوط به دهان فرد می شوند استخراج شده اند. در شکل ۳-۱۶، نقاط استخراج شده فوق را بر روی دهان فرد مشاهده می نمایید.

جدول ۳-۲- لیست نقاط استخراج شده از صورت برای جدا سازی دهان از تصویر صورت.

ردیف	شماره نقطه	مکان نقطه	لب	مکان نقطه در لب
۱	۳۱	نقطه چپ	هر دو	بیرونی
۲	۷۹	نقطه بین چپ و بالا	بالا	بیرونی
۳	۶	نقطه بالا	بالا	بیرونی
۴	۸۰	نقطه بین بالا و راست	بالا	بیرونی
۵	۶۴	نقطه راست	هر دو	بیرونی
۶	۸۶	نقطه بین راست و پایین	پایین	بیرونی
۷	۹	نقطه پایین	پایین	بیرونی
۸	۸۵	نقطه بین پایین و چپ	پایین	بیرونی



شکل ۳-۱۶- نقاط استخراج شده از دهان.

پس از اینکه این نقاط از نقاط صورت استخراج شد، نوبت به این می رسد که تصویر دهان از تصویر اصلی استخراج شود. برای این کار از الگوریتم زیر استفاده شده است:

۱. نقاط مورد نظر دهان استخراج شود.

۲. نقاط به ترتیب در جهت عقربه های ساعت، در آرایه ای قرار میگیرند که تشکیل چند ضلعی اطراف دهان را می دهد.

۳. تصویر دهان با مختصات زیر از تصویر اصلی استخراج می شود:

Rectangle

```
{
    Left=mouthLeftPoint.X;// جدول ۱ ردیف
    Top=mouthTopPoint.Y;// جدول ۳ ردیف
    Width=mouthRightPoint.X-mouthLeftPoint.Y;
    // از چپ به راست، نقطه ردیف ۵ و ۱ جدول //
    Height=mouthBottomPoint.Y-mouthTopPoint.Y;
    // از چپ به راست، نقطه ردیف ۷ و ۳ جدول //
}
```

۴. به ازای تمام پیکسل های تصویر استخراج شده دهان:

الف. مکان پیکسل به همراه مکان آرایه نقاط اطراف دهان (چند ضلعی دهان) به الگوریتم PIP داده می شود.

ب. اگر خروجی الگوریتم نشان دهنده این بود که نقطه در چند ضلعی واقع شده بود، به نقطه بعد برو. در غیر این صورت مقدار پیکسل در آن نقطه را صفر (رنگ سیاه) قرار بده.

۵. تصویر حاصل را به عنوان تصویر قطعه بندی شده دهان ذخیره کن.

پس از اعمال این الگوریتم بر روی تصویر عمقی، تصویر رنگی دهان نیز به همین شکل از روی تصویر عمقی بدست می آید.

۳-۲-۱- الگوریتم نقطه در چند ضلعی (PIP')

در قسمتی از الگوریتم استخراج دهان، به الگوریتمی نیاز است که بتواند مشخص کند که آیا یک نقطه در درون یک چند ضلعی وجود دارد و یا خیر. به این الگوریتم، الگوریتم نقطه در چند ضلعی می گویند. این الگوریتم مشخص می کند که چه پیکسل هایی در درون دهان و چه پیکسل های در خارج از دهان فرد واقع شده است.

یکی از راه حل های مشهور این الگوریتم، راه حل تابش اشعه^۲ می باشد که از سال ۱۹۶۲ استفاده می شده است. [۵۸]

تعریف اشعه^۳:

در ریاضیات، اشعه به صورت زیر تعریف می شود. اگر یک خط وجود داشته باشد و نقطه A روی خط باشد، A خط را به دو نیمه تقسیم می کند. هر یکی از این طرفین یک **اشعه** نامیده می شود و A نیز نقطه آغازی^۴ آن است. نقطه A بخشی از اشعه نیز به حساب می آید. یک اشعه از نقطه آغازی A به یک

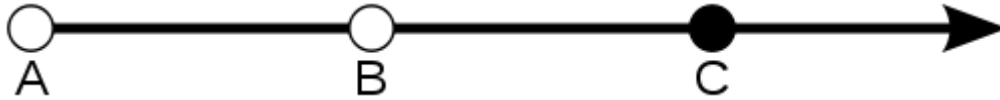
^۱ Point in Polygon

^۲ Ray-Casting

^۳ Ray

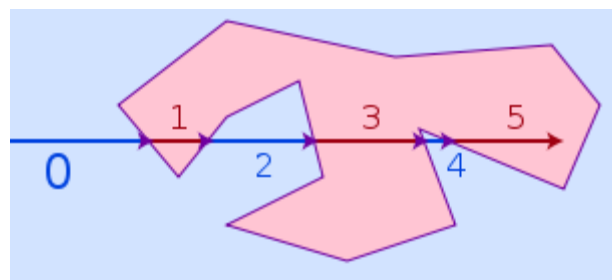
^۴ Initial Point

جهت (که با یک نقطه دیگر مانند B مشخص می شود) تا بی نهایت ادامه دارد. در شکل زیر یکی اشعه را با نقطه آغازی A مشاهده می نمایید [۵۹]



الگوریتم تابش اشعه:

در این الگوریتم گفته می شود که اگر از نقطه مورد نظر در درون یک چند ضلعی، یک اشعه به هرجهتی ایجاد نماییم، اگر آن اشعه اضلاع چند ضلعی را به تعداد زوج قطع نمود، نقطه در خارج از چند ضلعی است و اگر به تعداد فرد اضلاع چند ضلعی را قطع نمود، نقطه در داخل چند ضلعی واقع شده است. در شکل زیر منظور از قطع کردن توسط یک اشعه نشان داده شده است.



پیاده سازی:

برای پیاده سازی این الگوریتم، معمولاً از پیاده سازی W. Randolph Franklin استفاده می شود [۶۰]. در این روش فرض می شود که اشعه خارج شده از نقطه دارای عرض از مبدا صفر ($y=0$) نسبت به نقطه می باشد تا معادلات ساده تر شوند. سپس با استفاده از معادلات خطوط ایجاد شده از نقاط و بررسی قطع شدن در اشعه حاصل از نقطه، یک پرچم^۱ را معکوس^۲ می نماید تا در نهایت جواب نهایی در آن قرار گیرد. کد نهایی این روش در زیر آمده است:

```
private bool IsInPolygon(PointF[] points, PointF test)
```

^۱ Flag

^۲ Flip

```

{
    int i;
    int j;
    bool result = false;
    for (i = 0, j = points.Length - 1; i < points.Length; j = i++)
    {
        if ((points[i].Y > test.Y) != (points[j].Y > test.Y) &&
            (test.X < (points[j].X - points[i].X) * (test.Y - points[i].Y) /
            (points[j].Y - points[i].Y) + points[i].X))
        {
            result = !result;
        }
    }
    return result;
}

```

۳-۲-۲-۲- نتیجه استخراج

پس از اعمال الگوریتم های فوق، دهان فرد به صورت کامل از صورت جدا می شود و اضافات اطراف

آن نیز به رنگ سیاه در می آید. کد مورد استفاده در این بخش در زیر آمده است:

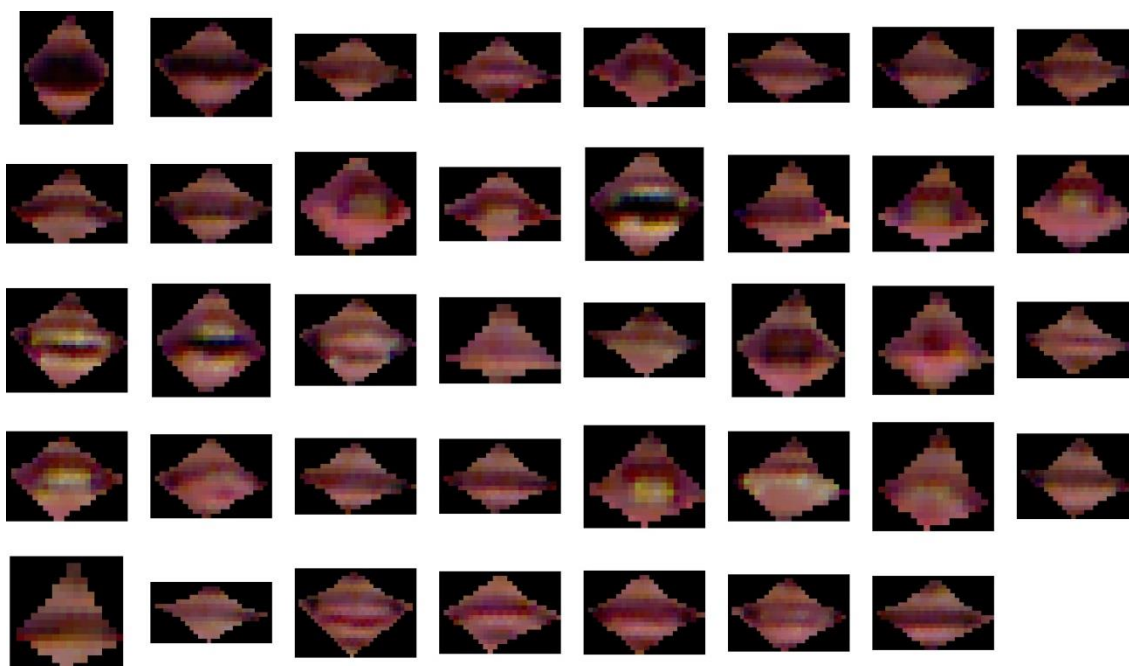
```

for (int i = 0; i < lipColor.Width; i++)
{
    for (int j = 0; j < lipColor.Height; j++)
    {
        if (!IsInPolygon(lipPolygonPoints, new PointF(i, j)))
        {
            lipColor[new System.Drawing.Point(i, j)] = new
            Rgb(System.Drawing.Color.Black);
            lipDepth[new System.Drawing.Point(i, j)] = new Gray(0); //Black
        }
    }
}

```

در شکل ۳-۱۷ تصویر تمامی دهان های استخراج شده از ۳۸ حرف پایگاه داده نشان داده شده

است. از بالا به پایین و از سمت چپ به راست به ترتیب طبق جدول حروف الفبای تهیه شده قرار گرفته اند:



شکل ۳-۱۷- تصویر ۳۸ دهان استخراج شده از حروف مختلف الفبای زبان اشاره فارسی باغچه بان.

۳-۲-۳- فرآیند پیش پردازش

مراحل پیش پردازش انجام شده در این تحقیق در شکل ۳-۱۸ نمایش داده شده اند. در این مراحل از پیش پردازش برای تصاویر دست و دهان، از مراحل پیش پردازش دست در تحقیق [۳۱] الهام گرفته شده است. یکی از مراحل که در تحقیق فعلی باید به دیگر مراحل اضافه شود، استخراج نقطه دست از تصاویر و برش دست است که در پایگاه داده کار قبلی از قبل انجام شده بوده و در این کار در قسمت های بالا شرح داده شد. شکل ۳-۱۹ یک نمونه تصویر استخراج شده می باشد و به ازای تمامی مراحل پیش پردازش و نگاشت یک به یک با نمودار شکل ۳-۱۸، مراحل مختلف قطعه بندی دست از تصویر عمقی و رنگی شکل ۳-۱۹، در شکل ۳-۲۰ نمایش داده شده است.



شکل ۳-۱۸- مراحل مختلف پیش پردازش انجام شده بر روی داده های تصویر. نمودار از [۳۱]



شکل ۳-۱۹- یک تصویر رنگی از نمونه های پایگاه داده مورد استفاده.



شکل ۳-۲۰- اعمال اصلی پیش پردازش بر روی تصویر رنگی شکل ۳-۱۹. ساخت ماسک عمقی (۱) و اعمال ماسک بر روی تصویر رنگی (۲).

۳-۲-۴- ساخت ماسک عمقی^۱ با آستانه گیری^۲

پیکسل های تصاویر عمقی دارای مقادیری بین ۰ و ۳۲۰۰ هستند که مشخص کننده عمق پیکسل های تصویر به میلیمتر می باشد. برای قطعه بندی دست، از تصویر برش خورده دست که در مراحل قبل به دست آمده است، به صورت زیر ماسک عمقی ایجاد می شود:

^۱ Depth Mask

^۲ Thresholding

۱. نزدیک ترین نقطه تصویر عمقی را بدست بیاور.

۲. به ازای هر نقطه از تصویر عمقی:

a. اگر پیکسل مورد نظر در بازه بین نزدیک ترین نقطه و نزدیک ترین نقطه + ۲۵۰

میلیمتر قرار داشت، پیکسل را نگه دار.

b. در غیر اینصورت پیکسل را سیاه کن.

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 ۳-۲۰ خروجی این بخش را مشاهده می نمایید.

۳-۲-۵-پیش پردازش بر روی ماسک عمقی

پس از بدست آوردن ماسک عمقی، چندین عمل پیش پردازش به ترتیب زیر بر روی ماسک عمقی انجام گرفته است که مشابه کار [۳۱] می باشد و جزئیات هر بخش آن در این کار مفصل شرح داده شده است.

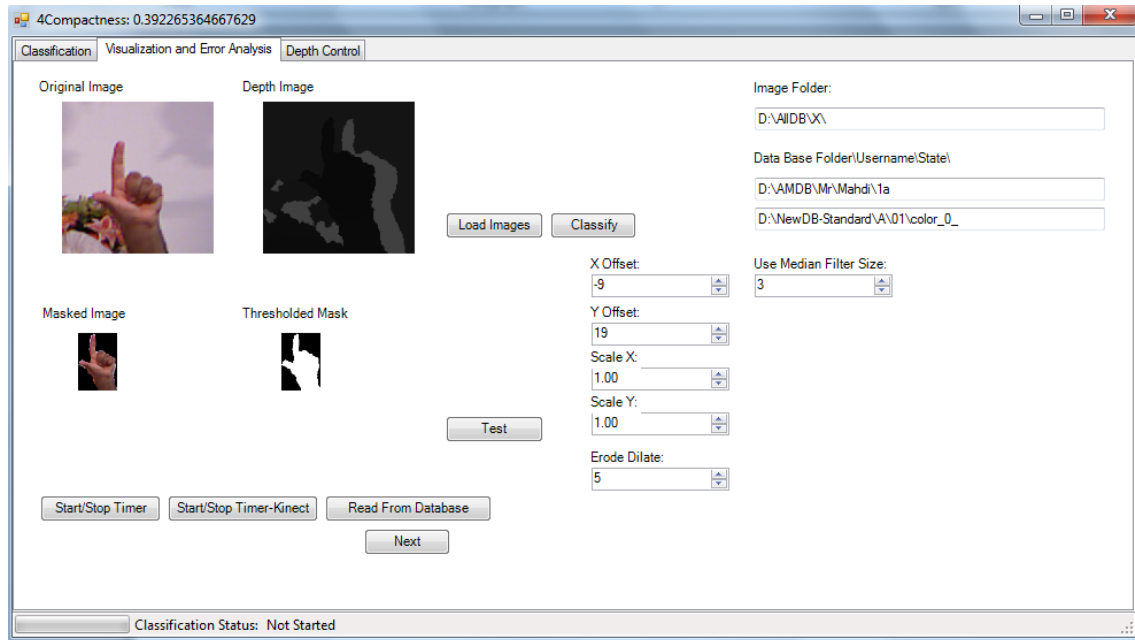
۱. جابجایی^۱ تصویر در صفحه با بردار (9,19-) به دلیل عدم تطابق دوربین رنگی و عمقی کینکت

ماسک با بردار (9,19-) در تصویر جابجا می شود تا تصویر رنگی و عمقی بر هم منطبق شوند.

این عدد به صورت تجربی با استفاده از بخش خطایابی سیستم پیاده سازی شده (شکل ۳-۲۱) بدست آمده است.

^۱ Translation

۲. هموار سازی^۱ ماسک با فیلتر میانه^۲ با اندازه پنجره ۳ در ۳



شکل ۳-۲۱- نمایش از بخش تحلیل خطای نرم افزار پیاده سازی شده و مراحل مختلف پیش پردازش و استخراج ویژگی یک نمونه ورودی در آن.

۳-۲-۶- پیش پردازش بر روی تصویر رنگی

در مرحله ۲ از شکل ۳-۲۰ ماسک عمقی بر روی تصویر رنگی اعمال شده و تصویر دست قطعه بندی شده را به ما می دهد. قبل از استخراج ویژگی بر روی این تصویر، دو مرحله پیش پردازش نیز بر روی این تصویر اعمال می شود که دلایل آن در کار [۳۱] گفته شده است:

۱. تشخیص کادر محاطی^۳ و برش تصویر رنگی

در این قسمت بخش های اضافی اطراف تصویر (پیکسل های سیاه) حذف شده و تصویر دست برش خورده به دست می آید.

^۱ Smoothing

^۲ Median Filter

^۳ Bounding Box

۲. تغییر اندازه^۱ تصاویر به ۳۲ در ۳۲ با حفظ نسبت تصویر^۲

در کار [۴۰] و [۳۱] نشان داده شده است که این بهترین اندازه برای تغییر اندازه نهایی می باشد.

۳-۳- استخراج ویژگی های هندسی^۳ دست

جهت شناسایی نمونه ها بایستی ویژگی هایی از آن ها استخراج و در اختیار دسته بند قرار داد. کارهایی که در آن ها نمی توان تصویر دست را به درستی استخراج نمود [۳۰] نمی توان از ویژگی های هندسی استفاده کرد. ولی در این تحقیق، از آنجا که شکل دست از پس زمینه جدا شده است، ویژگی های هندسی می تواند یکی از بهترین مجموعه از ویژگی هایی باشد که می توان از داده ها استخراج نمود [۴۰]. ویژگی هایی که در این مرحله استخراج شده است، ترکیبی از بهترین ویژگی های هندسی است که کاربرد آن ها در کار هایی مانند [۴۰] و [۳۱] برای استخراج ویژگی دست نشان داده شده است.

۳-۳-۱- Moment ها

در واقع Moment ها جهت ارائه مقادیر عددی جهت توصیف یک توزیع^۴ از یک نقطه مرکزی و یا محور خاص بیان شده اند [۶۱]. اولین بار توسط Hu^۵ در سال ۱۹۶۲ برای استفاده در تصاویر به کار رفت که تصاویر در آن یک توزیع دو بعدی فرض شد. تمامی مرتبه^۶ های Moment ها نیز برای تصویر تحت بررسی قابل محاسبه است. به طور کلی برای یک تصویر سیاه و سفید و یا خاکستری، Moment دو بعدی از مرتبه

^۱ Resizing

^۲ Aspect Ratio

^۳ Geometric Features

^۴ Distribution

^۵ M.K. Hu

^۶ Moment Order

$p + q$ بر روی تابع توزیع چگالی^۱ $f(x, y)$ (که در اینجا پیکسل های یک تصویر $N \times M$ می باشد) به صورت زیر بدست می آید [۶۱]:

$$m_{pq} = \sum_{y=0}^{M-1} \sum_{x=0}^{N-1} x^p y^q f(x, y) \quad (۱-۳)$$

برخی از Moment ها در تصویر معنای خاصی در رابطه با قطعه^۲ 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 $(\bar{x} = m_{10}/m_{00}, \bar{y} = m_{01}/m_{00})$ مرکز ثقل قطعه موجود در تصویر را مشخص می کند. مرتبه دوم از Moment ها نیز جهت محور توزیع شکل را مشخص می کنند [۴۰].

۳-۳-۱-۱ Moment های مرکزی^۳

این Moment ها با محاسبه مقادیر نسبی از مرکز ثقل تصویر محاسبه می شوند و استفاده از آن ها به جای Moment های اولیه، ویژگی هایی نامتغیر^۴ نسبت 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 [۳۱]:

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

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

^۱ Density Distribution Function

^۲ Segment

^۳ Central Moments

^۴ Invariant

۳-۱-۲-۳ 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 [۶۱]. این ویژگی ها در زیر آمده است. ویژگی هفتم علاوه به موارد ذکر شده فوق، نسبت به کجی^۲ نیز نامتغیر می باشد.

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

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

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

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

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

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

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

^۱ Orientation

^۲ Skew

۳-۳-۱-۳-ویژگی‌هایی بر حسب Moment ها

تعداد دیگری از ویژگی‌های نامتغیر نیز می‌توان با استفاده از Moment ها بدست آورد. چند ویژگی دیگر که در [۴۰] و [۳۱] 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:

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

^۱ Eccentricity

^۲ Inertia

^۳ Flip

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

۳-۳-۲- 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 [۳۱]. شکل ۳-۲۲ 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 ۳-۲۲- بزرگترین لبه تصویر یا کانتور استخراج شده که نقاط آن در اطراف دست قطعه بندی شده مشخص شده اند. تصویر از [۳۱]

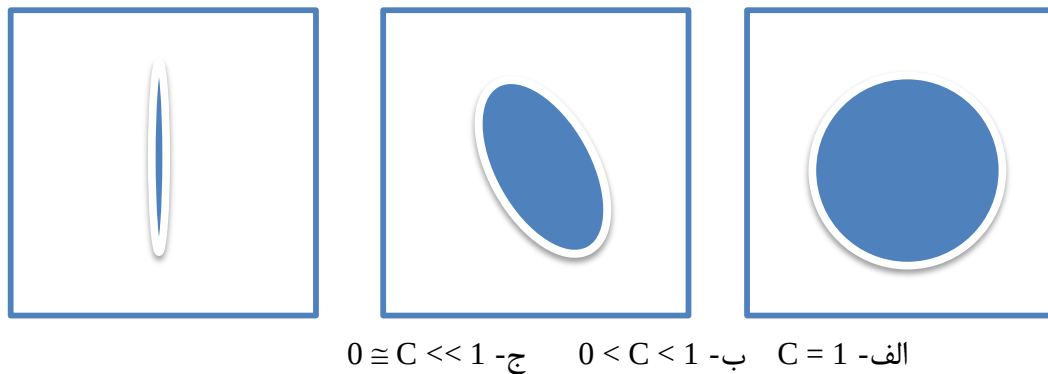
یکی از ویژگی های مهمی که می توان از کانتور استخراج نمود، ویژگی فشردگی^۲ است که به صورت زیر محاسبه می شود [۴۰]:

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

^۱ Blob

^۲ Compactness

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 ۳-۲۳ 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 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.



شکل ۳-۲۳- نمایشی از مقادیر مختلف ویژگی فشردگی (C) بر اساس شکل ظاهری اجسام موجود در تصویر. تصاویر از [۳۱]

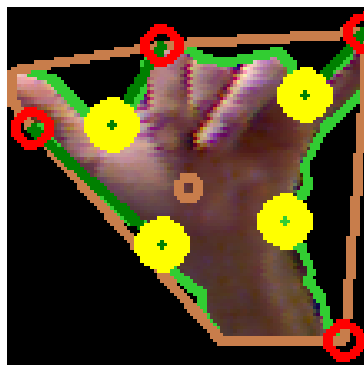
۳-۳-۲-۱- ویژگی های بر حسب پوسته محدب^۱

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 [۳۱]. در شکل ۳-۲۴ 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

^۱ Convex Hull

^۲ Defects

عمق چاله ها واقع شده اند را به شکل دایره های تو پر زرد رنگِ سوراخ شده مشاهده می نمایید. چهار چاله در این شکل پیدا شده است.



شکل ۳-۲۴- تصویری از پوسته محدب استخراج شده از یک نمونه از حرف y از پایگاه داده کار [۳۱].

با توجه به پوسته محدب، در کار [۳۱] ۶ ویژگی استخراج شده است. اولین ویژگی، تعداد انگشتان است. انگشتان با چند شرط از نقاط شروع و پایان چاله ها محاسبه شده است [۳۱]. تعداد چاله ها و دو ویژگی 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 فشردگی به عنوان یک ویژگی و شش ویژگی بر حسب پوسته محدب استخراج و تعداد کل بردار ویژگی ها به ۲۶ عدد رسیده است. نام این ۲۶ ویژگی به طور خلاصه در جدول ۳-۳ آمده است.

جدول ۳-۳- لیست ۲۶ ویژگی استخراج شده از دست.

ردیف	نام ویژگی	ردیف	نام ویژگی
۱	Hu Moment 1	۱۴	μ_{12}
۲	Hu Moment 2	۱۵	μ_{11}

Eccentricity	۱۶	Hu Moment 3	۳
Inertia Parallel to Main Axis	۱۷	Hu Moment 4	۴
Inertia Orthogonal to Main Axis	۱۸	Hu Moment 5	۵
Orientation of Main Axis	۱۹	Hu Moment 6	۶
Compactness	۲۰	Hu Moment 7	۷
No. of Fingers	۲۱	1/sqrt(m ₀₀)	۸
No. of Defects	۲۲	μ ₂₀	۹
Mean Length of Defects Start to Depth Vector	۲۳	μ ₀₂	۱۰
Mean Length of Defects Depth to End Vector	۲۴	μ ₃₀	۱۱
Defects Total Start-Depth Vectors Length to Total Depth-End Vectors Length Ratio	۲۵	μ ₀₃	۱۲
Mean of Defects Depth Height	۲۶	μ ₂₁	۱۳

۳-۴- استخراج ویژگی های اسکلتی

یکی از موارد قابل توجه در تشخیص حروف الفبای زبان اشاره فارسی، ارتفاع قرار گیری دست می باشد. از آنجا که از ویژگی های اسکلتی بدن که توسط کینکت استخراج شده است، مکان نقاط مختلف بدن و صورت در اختیار می باشد، نسبت فاصله دست از بالای سر فرد به ارتفاع صورت (از بالای صورت تا چانه) نیز به عنوان یک ویژگی اسکلتی به ویژگی ها اضافه شده است:

$$\text{headTop.Distance(handPoint)} / \text{headTop.Distance(cheekPoint)}$$

علت انتخاب این کسر برای این فاصله، نرمال سازی ویژگی بر حسب فاصله افراد از دوربین و همچنین اندازه بدن فرد می باشد. نرمال سازی باعث می شود که فاصله دست از سر نسبت به اندازه فرد و فاصله فرد از دوربین نامتغیر باشد. این کسر مقداری بزرگتر از صفر دارد که ممکن است عددی بیش از ۱ نیز بدهد که بعداً تمامی مقادیر نرمال شده و بین صفر و یک قرار می گیرند. در شکل ۳-۲۵ نقاط متصل شده به این نقاط و فاصله آن ها نشان داده شده است:



شکل ۳-۲۵ - تصویر محدوده استخراج شده اطراف دست، خط بالای سر تا دست و خط بالای سر تا چانه.

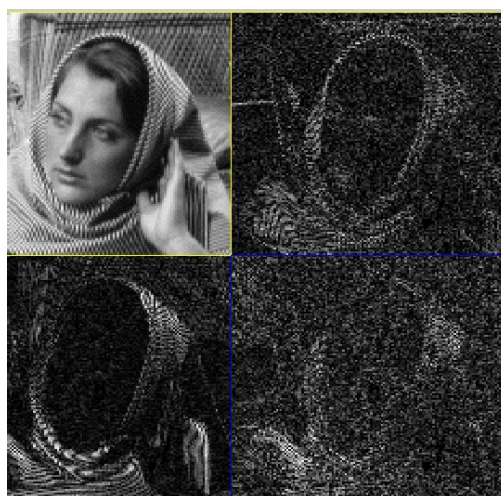
۳-۵- استخراج ویژگی بر پایه تبدیل موجک

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 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 اعمال این تبدیل بر روی تصویر، چهار سری از ضرایب^۲ جدید حاصل می شود که هریک به اندازه یک چهارم اندازه

^۱ Discrete Wavelet Transform

^۲ Coefficient

تصویر اولیه است. نمایی از این چهار تصویر خروجی در سطح اول از یک تبدیل موجک در شکل ۳-۲۶ آمده است. در تصویر بالا سمت چپ کلیات تصویر اولیه دیده می‌شود که به آن تصویر تخمینی^۱ می‌گویند. از طرفی جزئیات تصویر نیز در قسمت پایین سمت چپ دیده می‌شود که شامل نویزها نیز می‌شود. ضرایب یا جزئیات دیگر افقی^۲ و عمودی^۳ نیز به ترتیب در دو تصویر بالا سمت راست و پایین سمت چپ مشاهده می‌شود که جهت انجام معکوس تبدیل موجک کاربرد دارند.



شکل ۳-۲۶: خروجی یک تبدیل موجک دوبعدی گسسته در سطح یک.

این تبدیل در سطوح مختلف بر حسب کاربرد می‌تواند بر روی تصویر تخمینی تکرار شود و تصویری کلی تر و یا جزئیاتی در سطوح پایین تر تولید نماید که در الگوریتم‌ها و روش‌های پیشنهادی مختلف کاربرد خاص خود را پیدا کرده‌اند.

برای اینکه مشخص شود که استفاده از تبدیل موجک می‌تواند منجر به بهبود در شناسایی نمونه‌های زبان اشاره گردد، به جای ویژگی‌های معمول در کار [۳۱] از تبدیل موجک تا ۳ سطح به عنوان بردار ویژگی استفاده شده است. نتایج این تحقیق که در مقاله ی [۶۲] ارائه شده و به چاپ رسیده است نشان می‌دهد که استفاده از تبدیل موجک می‌تواند شناسایی زبان اشاره را بهبود بخشد. (جدول ۳-۴)

^۱ Approximate

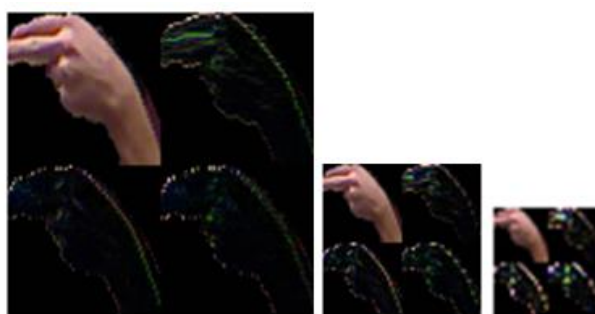
^۲ Horizontal Details

^۳ Vertical Details

جدول ۳-۴- مقایسه نتایج به دست آمده از روش پیشنهادی با استفاده از تبدیل موجک بر روی پایگاه داده کار [۳۱] و یک کار دیگر

کارهای قبلی	تعداد ویژگی ها	روش استخراج ویژگی	نوع تصاویر	تعداد نمونه ها	دسته بند	درصد تشخیص
[۳۱]	۲۶	Geometric Features	تصاویر عمقی و دیجیتال	۱۲۰۰۰	بیز	۸۰,۷۶٪
[۹]	۹۲	2D-DWT	تصاویر دیجیتال	۶۴۰	شبکه عصبی	۹۴,۰۶٪
روش پیشنهادی	۶۴	2D-DWT	تصاویر عمقی و دیجیتال	۱۲۰۰۰ ۷۲۰(۲۴*۶*۵)	بیز	۸۷,۹۶٪ ۹۶,۳۶٪

در این روش، تبدیل موجک گسسته با استفاده از تابع پایه Haar تا سه سطح بر روی تصاویر حروف زبان اشاره با سایز $۳۲*۳۲$ اعمال شده است. این عمل در شکل ۳-۲۷ نشان داده شده است.



شکل ۳-۲۷: اعمال تبدیل موجک تا سه سطح بر روی تصویر حرف زبان اشاره. سایز تصاویر از چپ به راست $۱۶*۱۶$ و $۸*۸$ و $۴*۴$ می باشند.

در نهایت هر چهار تصویر یا همان ماتریس ضرایب (شامل ضرایب عمودی^۱، ضرایب افقی^۱، ضرایب جزییات^۲، ضرایب تخمین^۳) حاصل از اعمال تبدیل موجک را به عنوان ویژگی های استخراج شده در نظر می

^۱ Coefficient Vertical

گیریم. پس از اعمال تبدیل موجک تا ۳ سطح بر روی تصاویر 32×32 به تصاویر 4×4 می‌رسیم که تمامی پیکسل‌های این ۴ تصویر را که معادل ۶۴ پیکسل می‌شوند به عنوان بردار ویژگی در نظر می‌گیریم.

در این تحقیق نیز همانند روش فوق، ولی اینبار فقط از تصویر استخراج شده دهان این ویژگی‌ها استخراج می‌شود. دلیل آن هم این است که تعداد ویژگی‌هایی بالایی به سیستم اضافه خواهد شد که باعث کند شدن سیستم و حتی مشکل در همگرایی LDA ممکن است به وجود بیاید و همچنین به دلیل تشابه بیشتر دهان افراد به هم تا دست آن‌ها در اجرای زبان اشاره، استفاده از تبدیل موجک بر روی دهان منطقی‌تر به نظر می‌رسد.

۳-۶- استخراج ویژگی‌های دهان

از تصویر بدست آمده از دهان، ۲۰ ویژگی از ۲۶ ویژگی استخراج شده از دست (ویژگی‌های هندسی) استخراج می‌شود. لازم به ذکر است که هفت ویژگی پوسته محدب که به منظور تشخیص تعداد انگشتان دست استفاده شد، برای دهان استخراج نمی‌شود.

همچنین از نقاط بدست آمده و استخراج شده از صورت فرد، نسبت طول به عرض دهان نیز به عنوان یکی ویژگی برای دهان که میزان باز بودن دهان را مشخص می‌کند در نظر گرفته شده است.

استخراج ویژگی تبدیل موجک از دهان نیز در سه سطح انجام شده است. تصویر به سائز ۳۲ در ۳۲ در آمد و سپس کل ویژگی‌های سطح سوم مورد استفاده قرار گرفت. در هر زیر بخش سطح سوم اندازه تصویر ۴ در ۴ و کل زیر بخش‌ها ۸ در ۸ می‌باشد که مجموعاً ۶۴ ویژگی از کنار هم قرار گرفتن آن‌ها استخراج گردیده است.

^۱ Coefficient Horizontal

^۲ Coefficient Detaile

^۳ Coefficient Approximate

۳-۷- تحلیل و انتخاب ویژگی ها

بردار کل ویژگی های استخراج شده را می توانید در شکل ۳-۲۸ زیر ببینید.

۶۴ ویژگی تبدیل ویژگی دهان	نسبت فاصله سر از دست به سر از چانه	نسبت طول به عرض دهان	۲۰ ویژگی دهان	۲۶ ویژگی دست
---------------------------------	--	-------------------------	---------------	--------------

شکل ۳-۲۸- بردار ویژگی های استخراج شده از نمونه های ورودی

در قسمت استخراج ویژگی ها ۴۸ (بدون ۶۴ ویژگی تبدیل موجک) 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 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 [۳۱].

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 (۴۸) 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.

^۱ Linear Discriminant Analysis

^۲ Principle Component Analysis

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 ویژگی ها در فصل مربوط به اعتبار سنجی (فصل ۵) مورد بررسی قرار خواهند گرفت. به دلیل اینکه تعداد ویژگی های مورد استفاده که ۴۸ ویژگی می باشد، سربار زیادی بر سیستم نخواهد داشت (در فصل ۵ بحث خواهد شد) 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.

۳-۸- روش دسته بندی

همانطور که توضیح داده شد، برای دسته بندی داده ها 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 بندی مناسب خواهد بود که ترجیحاً مرحله آموزش آن ها یک مرحله ای باشد و برای هر داده تکرار نشود. برای مثال دسته بندی نزدیکترین همسایه یا KNN^۱ برای این کار مناسب نمی باشد زیرا در هر 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 ۴۸ می باشد و همگی با توجه به ویژگی های دست قطعه بندی شده انتخاب شده و به نظر جداکننده^۲ می آیند، دسته بندی های احتمالی مانند دسته بندی و جنگل تصادفی (که در کار [۳۰] استفاده شده است) می تواند مناسب باشد. در هردوی این روش ها، پس از پایان مرحله آموزش، مرحله آزمایش به سرعت انجام می پذیرد و سربار بسیار کمی را برای سیستم خواهند داشت. استفاده از شبکه عصبی نیز به دلیل سرعت نسبتاً پایین در یادگیری و مناسب بودن برای داده های با

^۱ K-Nearest Neighbor

^۲ Discriminant

مشخصات نامعلوم (مانند پیکسل ها تصاویر) و عدم قابلیت ادغام سریع با داده های آزمایشی متفاوت (مثلا ادغام دو دسته بند) برای این کار انتخاب نشده است.

در نهایت جهت سیستم پیشنهادی، دسته بند بیز ساده^۱ 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.

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 [۳۱]. 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 براحتمالی به جای این احتمال پیشین قرار داد [۳۱].

^۱ Naïve Bayes

فصل ۴: نتایج بدست آمده

نتایج به دست آمده از فرایند دسته بندی توسط روش پیشنهادی در این قسمت به بحث گذاشته خواهد شد. در این فصل همچنین روش پیشنهادی با نتایج روش های پیشین مقایسه و نتایج تحلیل می شوند.

۴-۱- تعاریف

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 ضربدري^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 ضربدري با n -تا کردن استفاده می شود [۳۱]. 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.

از ۲۵-تا کردن برای تست سیستم با دقت فوق العاده بالا و متناسب با سایز پایگاه داده استفاده می شود. 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 [۳۱]:

^۱ Cross Validation

^۲ n-Fold

^۳ Test

^۴ Train

$$Train\ Ratio = (n - 1)/n * 100 \quad (1-4)$$

$$Test\ Ratio = 1/n * 100 \quad (2-4)$$

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 برای حالت ۱۰- تا کردن این نسبت 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 با ۵۷۰۰ نمونه مورد استفاده قرار می گیرد، ۲۲۸ نمونه به عنوان داده آزمایشی و ۵۴۷۲ نمونه به عنوان داده آموزشی در هر بار تا کردن به سیستم داده می شود. در نهایت، می توان گفت که سیستم با ۵۷۰۰ نمونه آموزش دیده است و با ۵۷۰۰ نمونه تست شده است ولی با این تفاوت که در هر بار تست، اشتراک مجموعه نمونه های آموزشی و آزمایشی تهی بوده است و هیچ نمونه آموزشی در آزمایش تکرار نشده است.

۴-۱-۲- معیارهای ارزیابی

برای بیان کمی چگونگی عملکرد سیستم، بایستی از معیار های اندازه گیری درستی عملکرد دسته بندی استفاده نمود [۳۱]. 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 [۳۱].

چهار تعریف مهم بر روی ماتریس درهم ریختگی بوجود می آید. این معیارها در جدول ۱-۴ 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 [۶۳].

جدول ۱-۴-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. بهترین این معیار ها دقت^۱، یادآوری^۲ و صحت^۳ می باشد که با فرمول های زیر مشخص شده اند [۶۳]:

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

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

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

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

^۱ Precision

^۲ Recall

^۳ Accuracy

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). 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.

۴-۱-۳- تصویر استفاده شده

در نتایج ارائه شده در جدول ۴-۲، در قسمت تصویر استفاده 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 ۴۸ ویژگی (۱۱۲ ویژگی با تبدیل موجک)) با یکدیگر ترکیب شده و ۹۶ ویژگی حاصل (۲۲۲ ویژگی با تبدیل موجک)، 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 ۴-۲ آمده است. همانطور که مشخص است، سیستم در حالت رنگی و بدون اضافه کردن ویژگی های تبدیل موجک توانسته است با ۹۹,۴٪ با ۱۰۰ نمونه به ازای هر حرف و با ۹۹,۲٪ با ۱۵۰ نمونه به ازای هر حرف به بهترین شکل نمونه های آزمایشی را با اعتبار سنجی ۲۵ تا کردن شناسایی کند. در ادامه نتایج بدست آمده تحلیل و با روش های پیشین مقایسه خواهد شد.

جدول ۲-۴- نتایج بدست آمده توسط سیستم پیشنهادی به تفکیک پارامترهای مختلف.

ردیف	نمونه از هر حرف	کل نمونه ها	تصویر استفاده شده	صحت	دقت	صحت	دقت	صحت	دقت
				با تبدیل موجک	بدون تبدیل موجک	بالاترین درصد			
۱	۱۰۰	۳۸۰۰	رنگی	۹۷,۹۷	۹۸,۰۹	۹۹,۳۹	۹۹,۴۰	%۹۹,۴	%۹۹,۴
۲			عمقی	۹۷,۶۱	۹۷,۶۶	۹۹,۰۸	۹۹,۰۹		
۳			ترکیبی	۹۹,۱۱	۹۹,۱۲	۹۸,۴۷	۹۸,۵۰		
۴	۱۵۰	۵۷۰۰	رنگی	۹۸,۷۷	۹۸,۸۱	۹۹,۱۶	۹۹,۱۷	%۹۹,۲	%۹۹,۲
۵			عمقی	۹۸,۰۲	۹۸,۰۹	۹۸,۸۹	۹۸,۹۱		
۶			ترکیبی	۹۷,۵۸	۹۷,۶۴	۹۹,۱۴	۹۹,۱۵		

۳-۴- تحلیل آزمایشات

جهت مشخص شدن نتیجه این تحقیق، در این قسمت به تحلیل نتایج آزمایشات پرداخته خواهد شد که در آن نتیجه گیری مربوط به تغییر پارامترهای مختلف بحث خواهد شد.

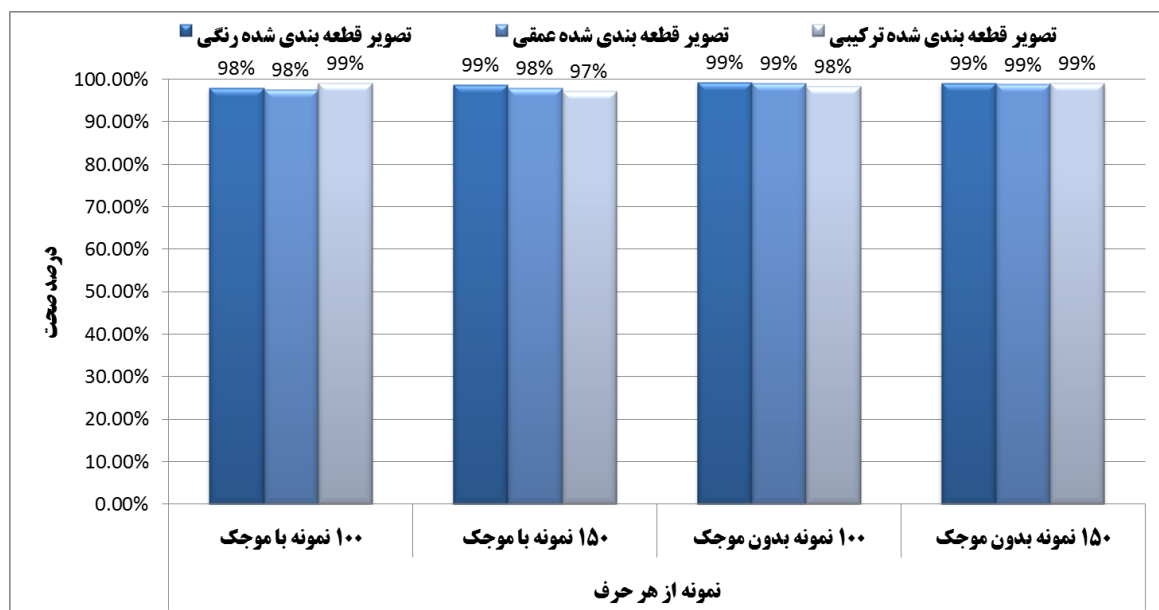
۴-۳-۱- تاثیر تعداد نمونه ها

همانطور که از جدول نتایج مشخص است، با افزایش تعداد نمونه های پایگاه داده، درصد تشخیص نهایی تا حدی کاهش یافته است. علت این امر نیز تنوع بیشتر داده های سیستم در حالت نمونه های بیشتر و از طرفی تعداد بیشتر داده های آزمایشی می باشد. هرچه که نمونه ها بیشتر شوند، سیستم باید با مسئله بزرگتری روبرو شود که کار تشخیص را مشکل تر می نماید.

۴-۳-۲- تاثیر تصویر انتخاب شده

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 شده است.

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 نیامده است (برخلاف کار [۳۱]) که نشان دهنده عدم برتری از ترکیب ویژگی ها می باشد. علت آن نیز می تواند افزایش بی رویه تعداد ویژگی ها باشد که بایستی با تحلیل LDA از تعداد آن ها کاسته شود. تنها موردی که با اطمینان می توان از آن سخن گفت برتری نسبی استفاده از تصاویر رنگی نسبت به تصاویر عمقی است. در نمودار 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 داده ها بدست آمده است. همانطور که شکل ۴-۱ 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.



شکل ۴-۱- نمودار مقایسه تاثیر تغییر تصویر قطعه بندی شده در صحت خروجی های سیستم به ازای انتخاب های متنوع از پایگاه داده.

۴-۳-۳- تاثیر استفاده از تبدیل موجک

استفاده از تبدیل موجک برای استخراج ویژگی دهان که منطقی نیز به نظر می رسد، نتوانسته است با وجود اضافه کردن ویژگی های بسیار زیاد به سیستم و افزایش قابل توجه بار محاسباتی چه در زمان

استخراج ویژگی و چه در زمان تحلیل ویژگی ها و دسته بندی، صحت و دقت شناسایی سیستم را بهبود بخشد. همانطور که از شکل بر می آید، تقریباً در تمامی مواردی که تبدیل موجه به سیستم اضافه شده است، درصد شناسایی سیستم کاهش پیدا کرده است. به همین دلیل استفاده از این ویژگی در شناسایی برای یک فرد توصیه نمی شود.

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 ۱۵۰ نمونه از پایگاه داده بدون تبدیل موجه، در جدول ۳-۴ ماتریس در هم ریختگی به تفکیک حروف نشان داده شده است. نتایج نشان می دهد که بیشترین تشخیص نادرست مربوط به تشخیص حرف س به جای حرف ش و بلعکس بوده است که به علت تفاوت ظاهری اندک در اجرای آن ها می باشد. همچنین تعداد زیادی از حروف به اشتباه ژ تشخیص داده شده اند. باقی حروف یا به درستی تشخیص داده شده اند و یا با خطای بسیار کمی همراه بوده اند.

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 ۳-۴- ماتریس در هم ریختگی بهترین نتیجه بر روی کل پایگاه داده مورد استفاده بر مبنای درصد.

کلاس شناسایی شده																														ر	پ	او	أ	ن	م	ل	خ	ک	ژ	ج	ه	ح	غ	ق	گ	ف	ای	اِ	د	چ	ب	ع	آ	آ
کلاس اصلی داده ها	ا	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																				
	ا	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																				
	ع	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																				
	ب	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																				
	چ	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																				
	د	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																				
	اِ	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																				
	ای	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																				
	ف	0	0	0	0	0	0	0	0	99	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																				
	ی	0	0	0	0	0	0	0	0	0	99	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																				
	ق	0	0	0	0	1	0	0	0	0	0	99	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																				
	غ	0	0	0	0	0	0	0	0	0	0	0	99	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																			
	ح	0	0	0	0	0	0	0	0	0	0	0	0	99	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																			
	ه	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																			
	ج	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0																			
	ژ	0	0	0	0	1	0	ا	0	0	0	0	0	0	0	0	0	99	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																			
	ک	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0																			
	خ	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	99	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																		
	ل	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																		
	م	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0																		
	ن	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0																		
	ا	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0																		
	او	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	99	0	0	0	0	0	0	0	0	0	0	1																		
	پ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0																		
	ر	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0																		
	ث	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	97	1	1	0	0	0	0	0	0	0																		
	ص	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0																		
	ش	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	91	6	0	0	0	0	0	0	0																		
	س	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	95	0	0	0	0	0	0	0																		
	ت	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0																		
ط	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	0	0	0	0	0	0																			
و	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	0	0	0	0	0																			
ی	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	0	1	0	0																			
ز	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98	0	0	0																			
ظ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0																			
ض	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	1	0																			
ذ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	99	0																			
ء	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100																			

کلاس اصلی داده ها

۴-۵- مقایسه با روش های پیشین

روش های اصلی که نتایج این سیستم باید با آن ها مقایسه شود، کارهایی است که بر روی زبان اشاره فارسی انجام گرفته است. 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 ۴-۴ 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 ۴-۴- جدول مقایسه روش پیشنهادی با روش های پیشین.

روش	ابزار مورد استفاده	دامنه شناسایی	تعداد کلاس ها	نمونه هر حرف	کل نمونه ها	بهترین درصد شناسایی	دسته بند	مزیت ها	معایب
کرمی و همکاران	دوربین رنگی	حروف هجی انگشتی زبان اشاره لوح موسوی	۳۲	۲۰	۶۴۰	۸۳,۰۳٪ تست ۹۴,۰۶٪ تست و آموزش	شبکه عصبی	تطابق با یک مدل زبان اشاره فارسی استاندارد	پایگاه داده آزمایشگاهی پس زمینه ساده انتخاب زبان غیر رایج زبان اشاره تعداد نمونه های کم
برزگر و شرافت	دوربین رنگی	حروف هجی انگشتی زبان اشاره با عدم تطابق با هیچ یک از مدل های موجود	۳۲	۵۰	۱۶۰۰	۹۹,۱٪	شبکه عصبی	اولین کار بر روز زبان اشاره فارسی درصد شناسایی بالا	کلاس های نامرتبط با زبان اشاره فارسی عدم کاربرد در زبان اشاره فارسی پایگاه داده آزمایشگاهی پس زمینه ساده
روش پیشنهادی	کینکت	حروف هجی انگشتی زبان اشاره باغچه بان	۳۸	۱۵۰ ۱۰۰	۵۷۰۰ ۳۸۰۰	۹۹,۲٪ (بردار ترکیبی) ۹۹,۴٪ (رنگی)	بیز	پس زمینه پیچیده، تنوع نمونه ها، عدم نیاز به دستکش	استفاده از کینکت

فصل ۵: اعتبار سنجی^۱

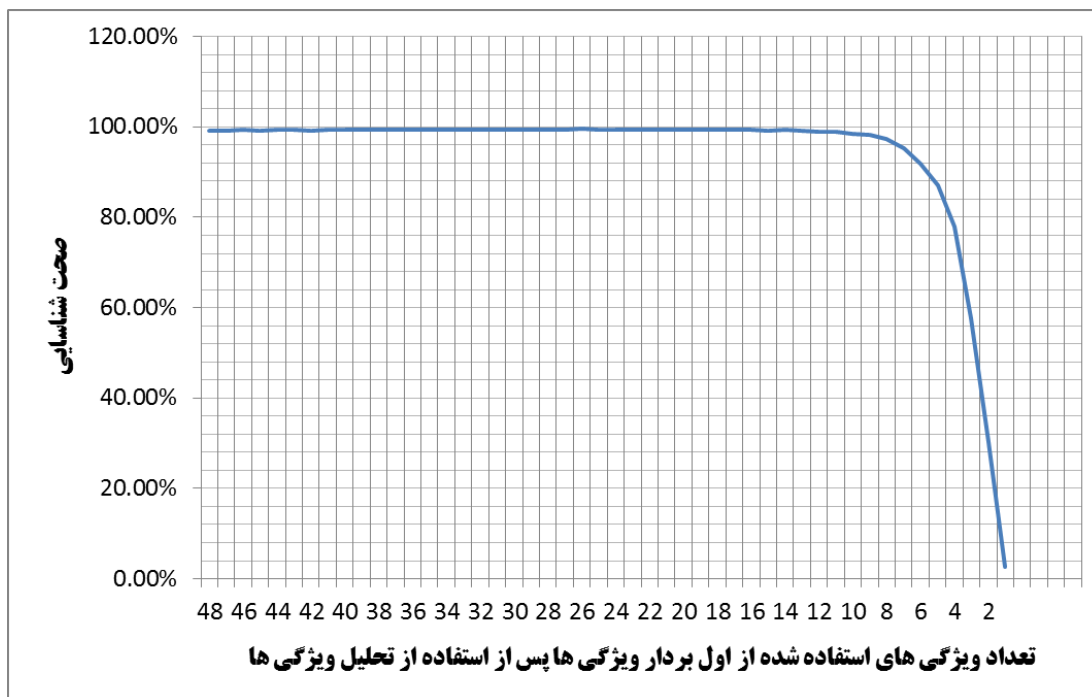
و تحلیل داده ها

^۱ 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- تحلیل انتخاب ویژگی ها با استفاده از LDA

ویژگی های استخراج شده در این تحقیق ۴۸ ویژگی برای حالت تصویر قطعه بندی شده رنگی و بدون استفاده از تبدیل موجک می باشد. 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 بهترین حالت یعنی استفاده از تصویر رنگی با ۴۸ ویژگی بدون استفاده از ویژگی های تبدیل موجک تحلیل کاهش تعداد ویژگی ها صورت خواهد گرفت. درصد صحت خروجی سیستم به ازای کم کردن صفر ویژگی (بردار اولیه ویژگی ها) تا کم کردن ۴۷ ویژگی (تست سیستم با تنها یک ویژگی) در نمودار شکل ۵-۱ نشان داده شده است. همانطور که از نمودار نیز مشخص است، در صورت کاهش نیمی از ویژگی از انتهای بردار مرتب شده ویژگی های حاصل از تحلیل LDA، سیستم بالاترین درصد شناسایی را نشان می دهد که می تواند در کاهش بار محاسباتی و حتی افزایش دقت سیستم موثر باشد.



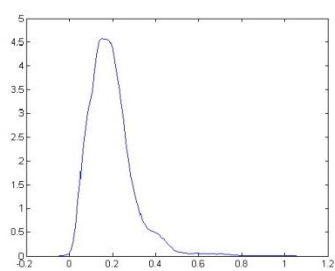
شکل ۵-۱- نمودار کاهش درصد شناسایی سیستم با کاهش ویژگی‌ها در حالت بدون مویک و تصویر رنگی.

۵-۲- تحلیل تابع چگالی احتمال ویژگی‌ها

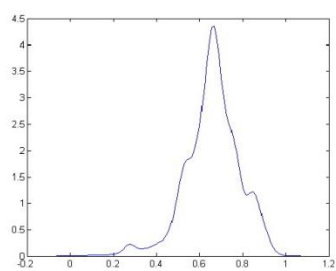
ویژگی‌های استخراج شده در این سیستم با این فرض به دسته بند بیز داده شده است که تابع توزیع چگالی آن‌ها دارای توزیع نرمال می‌باشد. برای اینکه به صحت این امر پی برده شود، در این قسمت به برآورد چگالی^۱ پرداخته می‌شود. یکی از روش‌های تخمین تابع چگالی احتمال، روش هموار کردن به کمک هسته‌ها می‌باشد^۲. این روش در MATLAB به صورت تابع `ksdensity` در دسترس می‌باشد. با استفاده از این تابع، توزیع تابع چگالی احتمال برخی از ویژگی‌ها در شکل ۵-۲ آورده شده است که نشان می‌دهد توابع توزیع نرمالی دارند و استفاده آن‌ها برای این دسته بند به درستی انجام گرفته است.

^۱ Density Estimation

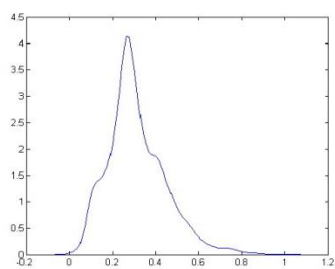
^۲ Kernel Smoothing



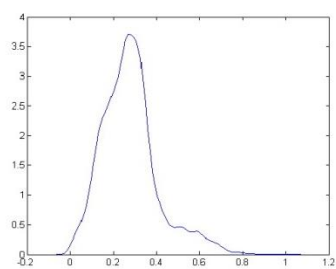
۱



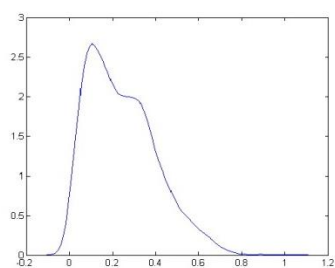
۱۵



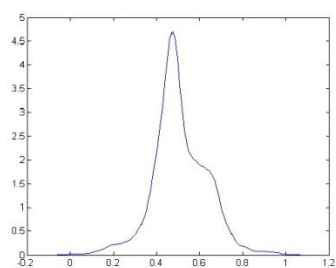
۸



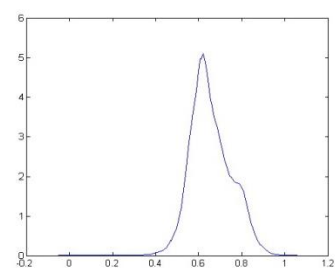
۹



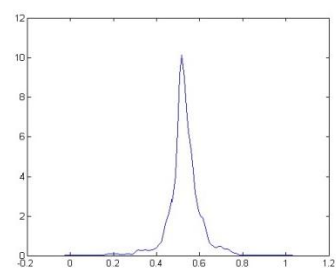
۱۰



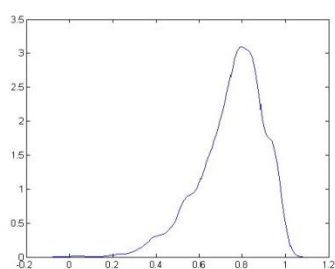
۱۱



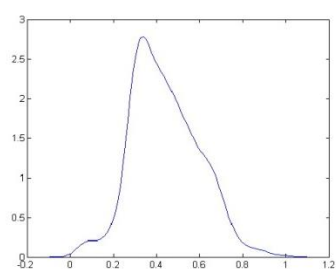
۱۳



۱۴



۱۶



۲۰

شکل ۵-۲- تابع چگالی توزیع احتمال برخی از ویژگی های استخراج شده.

۵-۳- 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 پایگاه داده با ۱۰۰ نمونه از هر حرف و در کل با ۳۸۰۰ نمونه بر روی تصویر رنگی انجام گرفت، نتایجی بدست آمد که به طور خلاصه در جدول ۱-۵ 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 ۱۹ نخ نرم افزاری سیستم استفاده شده است. زمان هایی در جدول ۱-۵ 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 ۱-۵- زمان پردازش و ۲ بار دسته بندی در سیستم پیشنهادی برای ۳۸۰۰ نمونه و ۴۸ ویژگی به تفکیک نوع پردازش. زمان های محاسبه شده مستقیم به صورت زیر خطی مشخص شده اند.

تعداد نخ/پارامتر	کل زمان پردازش	پردازش هر نمونه	تعداد فریم بر ثانیه
تک نخه	<u>۹۲ ثانیه</u>	۲۴,۲ میلی ثانیه	۴۱,۳ فریم بر ثانیه
چند نخه	<u>۵۵ ثانیه</u>	۱۴,۵ میلی ثانیه	۶۹ 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.

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 به منظور افزایش سرعت نیز نمی باشد ولی برای افزایش دقت می توان از LDA استفاده نمود.

فصل ۶: نتیجه گیری

و کارهای آینده

۶-۱- نتیجه گیری

در این تحقیق یک سیستم تشخیص هجی انگشتی حروف الفبای زبان اشاره فارسی بر مبنای ویژگی های استخراج شده از داده های عمقی سنسور کینکت ارائه شده است. تعداد ۴۸ الی ۱۱۲ ویژگی از پایگاه داده ای که بر مبنای مدل الفبای باغچه بان ایجاد شده، استخراج، و پس از آنالیز و دسته بندی، با دقت ۹۹,۲ درصد بر روی ۵۷۰۰ نمونه از ۳۸ حرف الفبا شناسایی شده اند.

این تحقیق اولین تحقیقی است که در آن از متداول ترین الفبای مورد استفاده در کشور، یعنی الفبای گویای باغچه بان، استفاده شده است. همچنین شناسایی تصاویر با پس زمینه پیچیده در یک محیط طبیعی نیز دستاورد دیگر این تحقیق است. تعداد نمونه های بیشتر نسبت به دیگر کارها و درصد شناسایی قابل قبول با سرعت بلادرنگ نتایجی است که این تحقیق را هرچه بیشتر برای استفاده کاربردی از دیگر کارها تمیز می دهد.

یکی از ویژگی های این تحقیق بررسی استفاده از تبدیل موجک بر روی تصاویر قطعه بندی شده دهان می باشد که نتایج بهبود چشم گیری را در استفاده از این ویژگی بروز نداده اند. البته ۲۰ ویژگی هندسی از دهان نیز به عنوان بخش جدا نشدنی از زبان اشاره نیز این کار را گام اولیه مهمی در شروع شناسایی کلمات زبان اشاره فارسی محسوب می نماید.

۶-۲- پیشنهادات برای کارهای آینده

در ادامه کار این تحقیق، پیشنهاد می شود که نتایج بر روی افراد متعدد نیز تحقیق شوند. همچنین شروع کار بر روی کلمات پرکاربرد و سپس همه کلمات زبان اشاره فارسی نیز می تواند گام مهمی در جهت ادامه این تحقیق باشد.

تحلیل ویژگی ها با استفاده از LDA نشان می دهد که ترکیبات بهتری برای استخراج ویژگی می تواند وجود داشته باشد که بایستی آزمایش و بهترین ترکیب یافت شود.

همچنین بهتر است سیستم مستقیماً توسط ناشنوایان عزیز تست شده و تغییرات بر حسب کاربرد آن ها از نرم افزار به سیستم پیشنهادی اعمال شود. تحلیل های دیگری مانند تاثیر دیگر دسته بند ها نیز در صورت لزوم می توان کارساز باشد.

- [1] "Kinect Possibilities for Deaf", [cited 11/24/2013]; Available from: <http://phys.org/news/2013-11-kinect-language-possibilities-deaf.html>.
- [2] "Copycat", [cited August 2012]; Available from: <http://www.cats.gatech.edu/content/copycat>.
- [3] "زبان اشاره", [cited 11/24/2013]; Available from: <http://fa.wikipedia.org/wiki/%D8%B2%D8%A8%D8%A7%D9%86%D8%A7%D8%B4%D8%A7%D8%B1%D9%87>.
- [4] Brashear H. M. (2010), "Improving the Efficacy of Automated Sign Language Practice Tools", PhD Thesis, Georgia Institute of Technology.
- [5] Dreuw P. (2005), "Appearance - Based Gesture Recognition", PhD Thesis, RWTH Aachen University of Technology.
- [6] "The Iranian Deaf Study Foundation", [cited 11/24/2013]; Available from: <http://idsf.ir/>.
- [7] بهادری ا., موسوی م., ل., محمودی ر., پیروزی م., زاده ح., ت., شاد ا. ق., "زبان اشاره فارسی (جلد اول)", زبان اشاره فارسی, Vol. 1, دانشگاه علوم بهزیستی و توان بخشی, ۲۳۲, ۱۳۸۹.
- [8] "Gesture Recognition", [cited 11/24/2013]; Available from: http://fa.wikipedia.org/wiki/%D8%AA%D8%B4%D8%AE%DB%8C%D8%B5_%D8%AD%D8%B1%DA%A9%D8%AA.
- [9] 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.

[10]برزگرسخویدی م., شرافت ا. (۱۳۸۸), "راهکاری ساده و موثر برای تشخیص حروف الفبای

فارسی در زبان اشاره حرکات انگشتان", نشریه مهندسی برق و کامپیوتر ایران, ۴(۷), pp. 323-328.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

- [17] 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.
- [18] 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).
- [19] 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*.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] 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.

- [24] 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.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] Zafrulla Z., Brashear H., Starner T., Hamilton H., Presti P. (2011), "American Sign Language Recognition with the Kinect", ICMI'11, Alicante, Spain.
- [30] 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.
- [31] Manashty A. R. (2012), "Sign Language Recognition Using Tof Cameras", Master of Science Thesis, Shahrood University of Technology.

- [32] 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.
- [33] 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.
- [34] 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.
- [35] 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.
- [36] 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.
- [37] 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.
- [38] Mistry P., Maes P. (2009), "Sixthsense – a Wearable Gestural Interface", Proceedings of SIGGRAPH Asia 2009, Sketch. Yokohama, Japan.
- [39] Lang S. (2011), "Sign Language Recognition with Kinect", B.Sc. Thesis, Freie Universität.
- [40] Zahedi M. (2007), "Robust Appearance-Based Sign Language Recognition", PhD Thesis, RWTH.

- [41] Gao W., Fang G., Zhao D., Chen Y. (2004), "A Chinese Sign Language Recognition System Based on Sofm/Srn/Hmm", **Pattern Recognition**, 37(12), pp. 2389-2402.
- [42] 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.
- [43] Nguyen T. D., Ranganath S. (2012), "Facial Expressions in American Sign Language: Tracking and Recognition", **Pattern Recognition**, 45(5), pp. 1877-1891.
- [44] 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.
- [45] 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.
- [46] 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.
- [47] 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.
- [48] Vamplew P. W. (1996), "Recognition of Sign Language Using Neural Networks", PhD Thesis, University of Tasmania.
- [49] 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.

- [50] 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.
- [51] 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.
- [52] von Agris U., Knorr M., Kraiss K. F. (2008), "The Significance of Facial Features for Automatic Sign Language Recognition", Automatic Face & Gesture Recognition, 2008. FG '08. 8th IEEE International Conference on, pp. 1-6.
- [53] Ari I., Akarun L. (2009), "Facial Feature Tracking and Expression Recognition for Sign Language", Signal Processing and Communications Applications Conference, 2009. SIU 2009. IEEE 17th, pp. 229-232.
- [54] Tan Dat N., Surendra R. (2008), "Tracking Facial Features under Occlusions and Recognizing Facial Expressions in Sign Language", Automatic Face & Gesture Recognition, 2008. FG '08. 8th IEEE International Conference on, pp. 1-7.
- [55] Microsoft-Corporation (2010)."Gesture Keyboarding", Patent US 2010/0 199 228 A1.
- [56] Kiani Sarkaleh A., Poorahangaryan F., Zanj B., Karami A. (2009), "A Neural Network Based System for Persian Sign Language Recognition", Signal and Image Processing Applications (ICSIPA), 2009 IEEE International Conference on, pp. 145-149.
- [57] Orton R. R. (2012), "Jabbar Baghcheban Phonetic Hand Alphabet for Persian Sign Language", 2012 [cited 11/25/2013]; Available from:

<http://www.geocities.ws/rebaorton/Jabbar/BaghchebanAlphabet.html>

- .
- [58] "Ray Casting Algorithm", [cited 11/25/2013]; Available from: http://en.wikipedia.org/wiki/Point_in_polygon#Ray_casting_algorithm.
- [59] "Ray (Mathematics)", [cited 11/25/2013]; Available from: [http://en.wikipedia.org/wiki/Ray_\(mathematics\)](http://en.wikipedia.org/wiki/Ray_(mathematics)).
- [60] Franklin W. R. "Point in Polygen Algorithm Implementation", [cited 11/25/2013]; Available from: http://www.ecse.rpi.edu/Homepages/wrf/Research/Short_Notes/pnpoly.html.
- [61] Hu M. K. (1962), "Visual Pattern Recognition by Moment Invariants", **IRE Transactions on Information Theory**, IT-8, pp. 179-187.
- [62] ForootanJahromi Z., Hassanpour H., Manashty A. R. (2013), "Sign Language Recognition Using Wavelet Transform Feature Extraction and Depth Cameras", 11th Iranian Conference on Intelligence Systems (ICIS2013), Kharazmi University, Tehran, Iran.
- [63] Eisner (2012), "Classifier Evaluation (Measuring Prediction Performance)", 2012 [cited August 2012]; Available from: webdocs.cs.ualberta.ca/.

Abstract

Sign language recognition is the first step towards interaction between deaf community and computer systems. Although there are many previous works on American Sign Language and languages other than Persian, there are only limited efforts on Persian sign language (PSL) recognition.

In this research, a robust Persian sign language recognition system based on Baghcheban notation (the first and most practical PSL used) is presented. Depth cameras, geometric features of the hand, skeleton and face are extracted and the results show a high rate of recognition in an environment with complex background. This is the first research on PSL based on complex backgrounds. Also this is the first time that skeleton and mouth features are used in PSL recognition.

Main contributions of the proposed system include creating the first depth-based Baghcheban notation PSL database with full body and facial images, recognition of all PSL alphabet containing 38 letters, presenting a real-time recognition system, using skeleton and facial features in recognizing PSL and further analysis of results to achieve the maximum recognition rate of %99.

The database created in this research consists of 150 to 500 samples per 38 alphabet letters in a complex scene with occlusion between hand, face, body, cloth and background. Kinect sensor is used to record the RGB image, depth image, upper body skeleton and 121 facial points of the user.

The results indicate a %99.4 recognition rate based on geometric features of the hand and mouth, and skeleton features of the relational placement of the hand on 3800 samples and %99.2 on 5700 samples. The performance analysis on the proposed system shows a real-time speed of 40-60 fps using normal Bayes classifier.

Keywords: Persian Sign Language Recognition, Baghcheban notation, Geometric Features, Wavelet Transform, Kinect, Depth Based, Fingerspelled



Shahrood University of Technology

Faculty of Computer Engineering and IT

Persian Sign Language Recognition using Depth Cameras

Zahra ForootanJahromi

Supervisor:

Dr. Hamid Hassanpour

Date: December 2013