بسم الله الرحمن الرحيم

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.

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

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

تقدیم به:

پدر و مادر عزیزم، با فداکاری ها و مهربانی همیشگی شان

همسرم، با هم راهی ها و هم یاری هایش

دخترم با لبخند زیبا و نگاه پر امیدش

تقدیر و تشکر

از اساتید گرانقدر 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 رهگشای مسیر موفقیت اینجانب و راهنمای رسیدن به قله های فضل و دانش بوده اند صمیمانه تشکر و قدردانی می نمایم.

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

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 راهنمائی 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 کشور ها به خصوص زبان اشاره امریکایی، 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، املای 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

۱-۲- زبان اشاره ۳

۱-۳- کاربردهای زبان اشاره ۴

۱-۴- انواع ساختارهای زبان اشاره ۵

۱-۵- انواع اشاره ۷

۱-۵-۱- نیازمندی ها و چالش های تشخیص اشاره ۸

۱-۶- انواع الفبای دستی در زبان اشاره فارسی ۸

۱-۶-۱- الفبای گویای باغچه بان ۸

۱-۶-۲- الفبای دستی فارسی جولیا سمیعی ۱۱

۱-۶-۳- الفبای دستی روزیک شهبازیان ۱۲

۱-۶-۴- الفبای دستی محسن لوح موسوی ۱۳

۱-۷- اهمیت تشخیص زبان اشاره ۱۵

۱-۸- تعریف مسئله ۱۶

۱-۹- روش حل مسئله ۱۸

فصل ۲: مروری بر کارهای پیشین ۱۹

۲-۱- بررسی موردی کارهای انجام شده جهت تشخیص زبان اشاره فارسی ۲۰

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

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

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

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 ۱-۲- الفبای دستی جولیا سمیعی. تصویر از [۶] ۱۲
شکل ۱-۳- الفبای دستی روزیک شهبازیان و شباهت برخی حروف با الفبای دستی آمریکایی. [۶]. ۱۳
شکل ۱-۴- الفبای دستی محسن لوح موسوی. تصویر از [۶] ۱۴
شکل ۱-۵ -پایگاه داده برای تشخیص زبان اشاره با استفاده از دوربین معمولی (RGB) ۱۷
شکل ۱-۶ -پایگاه داده برای تشخیص زبان اشاره با استفاده از دوربین عمقی ۱۷
شکل ۲-۱- 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 ۲-۳- 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 ۲-۴- نمایی از دستکش و شبکه حسگری بکار رفته در تحقیق ساجاکارن و همکارانش. [۱۹] ۲۸
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 آشوین و همکارانش [۲۳]. ۳۰
شکل ۲-۶- تصویری از یک دوربین فیلم برداری استریو. ۳۱
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] ۳۲
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]..... ۳۲
شکل ۲-۹- نمایی از نحوه کارکرد دوربین های عمقی در تصویربرداری از محیط..... ۳۳

شکل ۲-۱۰- نمایی از سنسور کینکت. ۳۶
شکل ۲-۱۱ - نمایی 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-نمایی از دو تصویر رنگی و عمقی گرفته شده از یک صحنه با استفاده از کینکت به همراه چینش و انواع لنزهای استفاده شده در آن. ۵۵
شکل ۳-۲-نمایی از اسکلت استخراج شده توسط سنسور کینکت. ۵۶
شکل ۳-۳-نمایی از ۱۲۱ نقطه استخراج شده از صورت که به صورت شبکه ای بر روی صورت منطبق شده است. ۵۷
شکل ۳-۴ نمایی 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-نمایی 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. سایز تصاویر از چپ به راست ۱۶*۱۶ و ۸*۸و 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 ۳-۱- لیست حروف موجود در پایگاه داده و تعداد نمونه از هر حرف. ۶۸
جدول ۳-۲-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 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 ۹۳
جدول ۴-۱-چهار تعریف مهم بر روی جدول در هم ریختگی. ۱۰۱
جدول ۴-۲-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 ۴-۳-جدول مقایسه روش پیشنهادی با روش های پیشین. ۱۰۸
جدول ۵-۱- 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 ۱:مقدمه