



دانشکده برق و رباتیک

گروه برق و الکترونیک

پایان نامه کارشناسی ارشد

تشخیص زبان اشاره بر اساس مشاهده

فرید آقاجانی درونکلا

استاد راهنما:

دکتر علی سلیمانی

استاد مشاور:

دکتر حسین مروی

مهر ۱۳۸۸

تاریخ : ۱۳۸۸/۰۶/۳۱

ویرایش : - - - -



مدیریت تحصیلات تکمیلی

فرم شماره (۶)

بسمه تعالی

فرم صورتجلسه دفاع پایان نامه تحصیلی دوره کارشناسی ارشد

با تأییدات خداوند متعال و با استعانت از حضرت ولی عصر (عج) جلسه دفاع از پایان نامه کارشناسی ارشد خانم / آقای فرید آقاجانی درونکلا رشته برق گرایش الکترونیک تحت عنوان : تشخیص زبان اشاره بر اساس مشاهده که در تاریخ ۱۳۸۸/۰۶/۳۱ با حضور هیأت محترم داوران در دانشگاه صنعتی شاهرود برگزار گردید به شرح زیر است :

<input type="checkbox"/> مردود	<input type="checkbox"/> دفاع مجدد	<input checked="" type="checkbox"/> قبول (با درجه : <u>بسیار خوب - امتیاز ۱۸</u>)
--------------------------------	------------------------------------	--

۲- بسیار خوب (۱۸/۹۹ - ۱۸)

۱- عالی (۲۰ - ۱۹)

۴- قابل قبول (۱۵/۹۹ - ۱۴)

۳- خوب (۱۷/۹۹ - ۱۶)

عضو هیأت داوران	نام و نام خانوادگی	مرتبه علمی	امضاء
۱- استاد راهنما	سید علی سلیمان	استادیار	
۲- استاد مشاور	سین نوری	استادیار	
۳- نماینده شورای تحصیلات تکمیلی	ایرینا مرزفی	استادیار	
۴- استاد ممتحن	علیرضا اندریزد	استادیار	
۵- استاد ممتحن	منیا رحیمی	استادیار	

تأیید رئیس دانشکده :

تقدیم به

پدر و مادر عزیزم؛

بزرگ‌ترین داری‌های همیشه‌ی زندگیم که هست و نیستم به گرمی نفس‌هاشان بسته است،

پاسکزاری؛

ابتدا از زحمات استاد راهنمای کراتقدر جناب آقای دکتر علی سلیمانی به خاطر راهنمایی‌های بی‌دریغ و حمایت‌های
دلسوزانه‌شان تشکر می‌کنم.

از جناب دکتر حسین مروی که استاد مشاور بنده در این پایان نامه بوده‌اند و جناب دکتر علیرضا احمدی فرد، استاد
کراتقدر که در مقطع مختلف انجام این پایان نامه، راهنمایی‌های خود را از بنده دریغ ننموده‌اند، تشکر می‌نمایم.
صداقانه و خالصانه‌ترین سپاسم را به دوستان عزیزم که راهنمایی‌ها و توصیه‌های بی‌دریغ به بنده داشته‌اند، تقدیم می‌کنم.

تعهد نامه

اینجانب
دانشگاه صنعتی شاهرود نویسنده پایان نامه/ رساله
تحت راهنمایی دکتر علی سلیمان متعهد می شوم .

- تحقیقات در این پایان نامه/ رساله 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/ رساله 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 / رساله تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه/ رساله رعایت می گردد .
- در کلیه مراحل انجام این پایان نامه/ رساله ، 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/ رساله ، 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: ۸۸/۶/۲۰
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 / رساله بدون ذکر مرجع مجاز نمی باشد .

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

چکیده

تشخیص زبان اشاره برای برقراری ارتباط میان افراد ناشنوا و نیمه شنوا با جامعه و یا کامپیوتر در رنج وسیعی کاربرد دارد و انجام این کار به کمک پردازش تصویر در سال‌های اخیر بخاطر اهمیت آن مورد توجه محققین قرار گرفته است. در این تحقیق می‌خواهیم با استفاده از پردازش تصاویر دو بعدی اشاره‌ها و علائم دست‌ها را بدون استفاده از تجهیزات ویژه تشخیص دهیم. یکی از بهترین روش‌ها برای شناسایی علائم و اشارات، مقایسه مسیر حرکت دست و مدل‌سازی شکل دست و نحوه قرارگیری انگشتان در آن می‌باشد.

در این پروژه به معرفی و بررسی تئوری بیشترین راستای لبه (MED) پرداخته و نشان داده می‌شود MED یک ویژگی هندسی است که به ساختار دست بستگی دارد. ویژگی MED به صورت آماری بیانگر بیشترین راستای لبه‌های تصویر است که با دقت خوبی نشان دهنده راستای قرار گرفتن انگشتان در هر اشاره می‌باشد. برای استخراج این ویژگی ابتدا تصویر باینری لبه از تصویر دست گرفته شده و سپس از المان ساختاری پیشنهادی مناسب که تنها روی پیکسل‌های لبه عمل می‌کند استفاده می‌شود. برای مدل‌سازی بهتر شکل دست از ویژگی LMED استفاده می‌شود. به این ترتیب تصویر مورد نظر به نواحی کوچکتر با اندازه معین افزاز و در هر ناحیه ویژگی MED محاسبه می‌گردد. همچنین استفاده از این ویژگی حجم پایگاه داده مورد نیاز را به شدت کاهش داده و موجب افزایش سرعت تشخیص می‌شود.

الگوریتم پیشنهادی برای تشخیص زبان اشاره شامل دو مرحله می‌باشد: ۱- کلاسه‌بندی اولیه ۲- کلاسه‌بندی نهایی. در مرحله اول ابتدا ویژگی‌های مسیر حرکت دست، منحنی زاویه‌ی وتر دست و منحنی مساحت تصویر دست استخراج شده و سپس با کمک الگوریتم DTW و کلاسه‌بندی نزدیک ترین همسایگی، بیشتر کلاس‌های نامطلوب حذف می‌شود. در نهایت سه کلاس باقی می‌ماند به طوری که احتمال حضور کلاس درست در این سه کلاس صد در صد می‌باشد (دقت صد در صد با رنگ سه).

در نتیجه دقت و سرعت تشخیص در زبان اشارات بهبود داده می‌شود. در مرحله دوم ابتدا با استفاده از الگوریتم DTW که تابع آن منحنی مسیر حرکت دست می‌باشد فریم‌های متناظر با هم انتخاب و سپس با کمک دو ویژگی LMED و مسیر حرکت دست، فاصله اشاره ورودی از سه کلاس تعیین شده در مرحله اول محاسبه می‌شود. در این پروژه سه روش برای کلاسه‌بندی زبان اشاره پیشنهاد شده است: ۱- روش مینیمم فاصله با استفاده از ویژگی LMED ۲- روش مینیمم میانگین هندسی ۳- روش مینیمم فاصله اقلیدسی به صورت تطبیقی. کلاسه‌بندی‌های پیشنهادی روی مجموعه‌ای از اشارات ASL صورت گرفت که شامل ۴۷ کلمه متفاوت با ۱۳۷ نمونه می‌باشد. در این آزمایش میزان دقت به مقدار ۹۷.۷٪ رسیده که در بهترین حالت برای اشارات با یک دست ۹۸.۹٪ و برای اشارات با هر دو دست ۹۵.۶٪ بوده است. نتایج بدست آمده کارایی بالای روش‌های پیشنهادی را نشان می‌دهد.

کلمات کلیدی: تشخیص زبان اشاره، الگوریتم DTW، ویژگی MED و LMED.

فهرست مطالب

۱	فصل اول
۱	مقدمه
۲	۱-۱ بیان مسأله
۴	۲-۱ ضرورت انجام تحقیق
۹	۳-۱ اهداف
۱۱	۴-۱ اشاره‌ای به محتوای فصل‌های بعد
۱۳	فصل دوم
۱۳	مطالعات پیشین
۱۴	مقدمه
۱۴	۱-۲ تشخیص علائم قسمت‌بندی شده (کلمات):
۲۱	فصل سوم
۲۱	تئوری‌های مورد استفاده
۲۲	مقدمه
۲۲	۱-۳ دنبال کننده
۲۵	۲-۳ مستطیل محاط بر دست
۲۵	۳-۳ پیدا کردن لبه
۲۶	۱-۳-۳ آشکار ساز لبه با مشتق مرتبه اول
۲۸	۲-۳-۳ عملگر گرادیان در همسایگی کوچک
۳۲	۳-۳-۳ عملگرهای گرادیان در همسایگی بزرگ
۳۵	۴-۳-۳ عملگر گرادیان Canny
۳۹	۴-۳ DTW
۴۰	۱-۴-۳ الگوریتم DTW
۴۲	۲-۴-۳ مثال‌هایی از روش‌های پیدا کردن خط سیر برای اطمینان از پیوستگی محلی
۴۵	۵-۳ راستای ماکزیمم لبه و طول آن در یک مجموعه
۵۰	فصل چهارم
۵۰	پایگاه‌های داده

۵۱ مقدمه
۵۱ RWTH-Boston-104 ۱-۴
۵۴ RWTH-Boston-50 ۲-۴
۵۴ پایگاه داده نهائی
۵۶ ۴-۴ مشخصه‌های تصاویر پایگاه داده
۵۶ ۱-۴-۴ الگوی حرکت دست نسبت به سر
۵۷ ۲-۴-۴ الگوی حرکت دست و شکل دست
۵۸ ۳-۴-۴ الگوی حرکت دو دست نسبت به هم
۶۰ فصل پنجم
۶۰ ویژگی‌های بکار رفته
۶۱ مقدمه
۶۱ ۱-۵ منحنی مسیر حرکت دست نسبت به سر
۶۳ ۲-۵ شیب وتر $1 - Aspect Ratio$ مستطیل محاط بر تصویر دست؛ $\Delta y \Delta x(n)$
۶۴ ۳-۵ مساحت مستطیل محاط بر تصویر دست؛
۶۷ ۴-۵ استخراج تصویر لبه با استفاده از روش Canny
۶۸ ۵-۵ راستای ماکزیمم لبه (MED) و طول آن در کل تصویر دست
۶۸ ۶-۵ راستای ماکزیمم لبه و طول آن به صورت محلی از تصویر (LMED)
۷۰ فصل ششم
۷۰ الگوریتم کلاسه‌بندی
۷۱ مقدمه
۷۱ ۱-۶ کلاسه‌بندی اولیه
۷۳ ۱-۱-۶ کلاسه‌بندی با استفاده از ویژگی منحنی مساحت تصویر دست
۷۳ ۲-۱-۶ کلاسه‌بندی با استفاده از ویژگی منحنی زاویه‌ی وتر تصویر دست
۷۴ ۳-۱-۶ کلاسه‌بندی با استفاده از ویژگی منحنی مسیر حرکت دست
۷۵ ۴-۱-۶ انتخاب کلاس‌های مطلوب با استفاده از نتایج بخش‌های قبلی
۷۶ ۲-۶ کلاسه‌بندی نهایی
۷۸ ۱-۲-۶ روش نزدیکترین همسایگی با استفاده از ویژگی LMED و الگوریتم DTW
۸۰ ۲-۲-۶ روش مینیمم میانگین هندسی به صورت تطبیقی

۳-۲-۶	روش نزدیکترین همسایگی اقلیدسی به صورت تطبیقی	۸۲
	فصل هفتم	۸۵
	نتایج و مقایسه	۸۵
	مقدمه	۸۶
۱-۷	بررسی نتایج الگوریتم پیشنهادی	۸۶
۱-۱-۷	نتیجه کلاسه‌بندی ۱ با استفاده از ویژگی منحنی مسیر حرکت دست	۸۶
۲-۱-۷	نتیجه کلاسه‌بندی با استفاده از ویژگی منحنی مساحت تصویر دست	۸۷
۳-۱-۷	نتیجه کلاسه‌بندی با استفاده از ویژگی منحنی شیب وتر تصویر دست	۸۸
۴-۱-۷	نتایج انتخاب کلاس‌های مطلوب با استفاده از نتایج بخش‌های ۱، ۲ و ۳	۸۹
۵-۱-۷	نتایج کلاسه‌بندی بر اساس ویژگی پیشنهادی LMED (روش اول)	۹۱
۶-۱-۷	نتایج کلاسه‌بندی به روش مینیمم میانگین هندسی (روش دوم)	۹۱
۷-۱-۷	نتایج کلاسه‌بندی به روش نزدیکترین همسایگی اقلیدسی انطباقی (روش سوم)	۹۴
۸-۱-۷	نتیجه کلاسه‌بندی نهایی	۹۴
۲-۷	مقایسه نتایج	۹۶
	مراجع	۹۹

فهرست شکل‌ها

- شکل (۱-۱): نمونه‌هایی از انواع دستکش‌های مجهز ۷
- شکل (۲-۳): مدل‌های یک بعدی لبه و خط در فضای پیوسته ۲۶
- شکل (۳-۳): سازنده گرادیان متعامد ۲۷
- شکل (۴-۳): گرادیان تفاضلی عمودی و افقی از تصویر peppers_mon ۲۹
- شکل (۵-۳): گرادیان Roberts از تصویر peppers_mon ۳۰
- شکل (۶-۳): شماره گذاری قراردادی برای عملگر آشکارساز لبه با ابعاد 3×3 ۳۱
- شکل (۷-۳): گرادیان‌های Sobel, Prewitt و Frei-chen از تصویر peppers_mon ۳۳
- شکل (۸-۳): گرادیان‌های در همسایگی بزرگ از تصویر peppers ۳۶
- شکل (۹-۳): مقایسه توابع پاسخ ضربه Canny و FDOG ۳۸
- شکل (۱۰-۳): مقایسه تطبیق میان دو سیگنال ظاهراً مشابه ۳۹
- شکل (۱۱-۳): شرایط ادامه مسیر: مقادیری که وزن‌های مربوط به مسیرهای مورد نظر را نشان می‌دهند ۴۲
- شکل (۱۲-۳): شرایط ادامه مسیر: مقادیری که وزن‌های مربوط به مسیرهای مورد نظر را نشان می‌دهند ۴۲
- شکل (۱۳-۳): بدست آوردن کوتاه‌ترین مسیر میان دو سیگنال مشابه هم به کمک الگوریتم DTW ۴۴
- شکل (۱۴-۳): الف) ساختار پنجره مناسب برای پیدا کردن راستای لبه ب) نامگذاری شمارنده هر راستا ۴۵
- شکل (۱۵-۳): مراحل محاسبه p و q ۴۷
- شکل (۱۶-۳): ساختار لبه در یک ناحیه‌ی کوچک ۴۷
- شکل (۱-۴): نمونه‌هایی از تصاویر اشارات پایگاه داده‌ی RWTH-Boston-104 ۵۳
- شکل (۲-۴): نمونه‌هایی از تصاویر فریم‌ها که در آنها مکان سر و مختصات مسطتیل محاط بر دست مشخص گشته‌اند ۵۵
- شکل (۳-۴): مثال‌هایی که در آن دست راست با شکل خاص و در مکان مشخصی نسبت به سر قرار می‌گیرد ۵۷
- شکل (۴-۴): مثال‌هایی که در آنها دست راست با شکل و الگوی حرکت خاص از ناحیه‌ی مشخصی نسبت به سر عبور می‌کند ۵۷
- شکل (۵-۴): اشاراتی که در آن دست دارای الگوی حرکت خاصی نسبت به سر می‌باشد ۵۸
- شکل (۶-۴): اشاراتی که در آنها دو دست الگوی حرکت مشخصی نسبت به هم دارند ۵۹
- شکل (۱-۵): نمودارهای منحنی مسیر حرکت دست راست نسبت به سر برای دو کلاس متفاوت ۶۲

- شکل (۲-۵): شیب قطر مستطیل محاط بر تصویر دست ۶۳
- شکل (۳-۵): نمودارهای ویژگی منحنی شیب وتر مستطیل محاط بر تصویر دست راست برای دو کلاس متفاوت . ۶۴
- شکل (۴-۵): مساحت مستطیل محاط بر تصویر دست ۶۵
- شکل (۵-۵): نمودارهای منحنی مساحت مستطیل محاط بر تصویر دست برای دو کلاس مختلف ۶۶
- شکل (۶-۵): تصاویر لبه از شکل دست در یک فریم از دو اشاره‌ی نمونه و قسمت‌بندی آنها برای محاسبه ویژگی LMED ۶۹
- شکل (۱-۶): الگوریتم کلاسه‌بندی در مرحله‌ی اول ۷۲
- شکل (۲-۶): الگوریتم کلی کلاسه‌بندی نهایی ۷۷
- شکل (۳-۶): منحنی‌های نشان دهنده‌ی نمونه‌ها با فاصله‌ی یکسان از اشاره‌ی ورودی در روش دوم ۸۲
- شکل (۴-۶): منحنی‌های نشان دهنده‌ی نمونه‌ها با فاصله‌ی یکسان از اشاره‌ی ورودی در روش سوم ۸۶
- شکل (۱-۷): دقت کلاسه‌بندی با استفاده از ویژگی منحنی مساحت تصویر دست بر حسب رنگ ۸۷
- شکل (۲-۷): دقت کلاسه‌بندی با استفاده از ویژگی منحنی زاویه‌ی وتر تصویر دست بر حسب رنگ برای اشارات با ۱ دست ۸۹
- شکل (۳-۷): نتایج کلاسه‌بندی مرحله‌ی اول با استفاده از نتایج کلاسه‌بندی با رنگ بیشتر از ۱ ۹۰
- شکل (۴-۷): نتایج کلاسه‌بندی بر اساس ویژگی پیشنهادی (LMED) ۹۲
- شکل (۵-۷): نتایج کلاسه‌بندی به روش مینیمم میانگین هندسی با استفاده از هر دو ویژگی منحنی مسیر حرکت دست و LMED ۹۳
- شکل (۶-۷): نتایج کلاسه‌بندی به روش نزدیکترین همسایگی اقلیدسی با استفاده از هر دو ویژگی منحنی مسیر حرکت دست و LMED ۹۵
- شکل (۷-۷): نتیجه کلاسه‌بندی نهایی ۹۶

فهرست جداول

جدول (۱-۳): شماره، نام و زاویه‌ی هر راستا..... ۴۵

جدول (۲-۳): چهار حالت ممکن برای بدست آوردن طول و زاویه‌ی راستای ماکزیموم لبه در ناحیه‌ی a ۴۸

جدول (۱-۷): مقایسه‌ی شرایط و نتیجه‌ی الگوریتم پیشنهادی با چند کار انجام شده در زمینه‌ی تشخیص زبان اشاره

..... ۹۸

فصل اول

مقدمه

۱-۱ بیان مسأله

تشخیص اشاره و علائم مربوط به شکل معنی دار حرکات یک فرد شامل دست ها، بازوها، صورت، سر و بدن است. این مطلب در برقراری ارتباط هوشمند و کارآمد بین انسان و رایانه از بیشترین درجه اهمیت برخوردار می‌باشد. کاربردهای تشخیص اشاره و علائم [1]، از تشخیص زبان اشاره گرفته تا توانبخشی پزشکی، در واقعیت مجازی بسیار گسترده و متفاوت است و می‌تواند به‌عنوان یک روش در این راستا به کار رود. روند کار بر این اصل استوار است که اشارات و علائم ساخته شده به وسیله کاربر توسط دریافت‌کننده تشخیص داده شود. اشارات، علائم و حرکات بدن؛ شامل حرکت فیزیکی انگشتان، دست، بازوها، سر، صورت و یا بدن؛ رسا و معنی‌دار هستند که با اهداف انتقال اطلاعات پرمعنی و ارتباط متقابل با محیط اطراف، شامل یک زیرمجموعه کوچک جالب از حرکات ممکن برای بدن انسان می‌باشند. همچنین یک اشاره ممکن است توسط محیط به‌عنوان یک تکنیک فشرده‌سازی برای انتقال به هر جای دیگر استفاده شده و سپس توسط گیرنده بازسازی شود.

به طور عمومی روش‌های زیادی برای نشان دادن یک مفهوم به وسیله اشاره و بالعکس وجود دارند [2]، از این‌رو اشارات مبهم و نامعین هستند. به طور مثال برای نشان دادن مفهوم "ایست" یک شخص می‌تواند از حالتی مانند بالا بردن دست، کشیدن صورت به جلو و یا تکان دادن سریع دو دست بالای سر استفاده کند. اشارات و علائم نیز مانند صحبت و دست نوشته‌ها میان افراد مختلف و حتی در جملات مجزا از یک فرد متفاوت است. از جمله روش‌های متنوعی که برای تشخیص اشاره و علائم به کار می‌رود می‌توان به مدل‌های ریاضی بر اساس زنجیره HMM¹ [3] و روش‌ها و ابزارهایی بر اساس محاسبات نرم افزاری [4] اشاره کرد.

اشارات و علائم می‌توانند ایستا (کاربر در این حالت وضع و شکل خاصی به خود می‌گیرد) و یا متحرک و پویا (علامت دارای حرکت و تکان خاصی است که مراحل آن از یک زمان معین آغاز و در

¹ Hidden Markov models

یک زمان معین خاتمه می‌یابد) باشند. بعضی اشارات مانند زبان اشاره نیز هر دو قسمت ایستا و پویا را دارا هستند. از طرف دیگر تشخیص خودکار اشارات پیوسته با زمان، نیاز به گسسته کردن و جداسازی تصاویر نسبت به زمان دارد. در این صورت اغلب نیاز است که نقاط شروع و پایان یک اشاره که در غالب فریم‌هایی به حرکت در آمده است هم برحسب زمان و هم بر حسب مکان تعیین شوند. همچنین گاهی یک اشاره از مفهوم اشاره قبلی تأثیر می‌پذیرد و این زمانی اتفاق می‌افتد که اشارات دنبال هم بیایند. علاوه بر این، اشارات مربوط به زبان، فرهنگ و ملیت خاصی هستند که این امر خود تشخیص آن‌ها را دشوارتر می‌سازد. تشخیص حالات شامل زیرمجموعه‌های زیر می‌باشد:

۱- حالت دست و بازو: تشخیص شکل دست و بازو، تشخیص زبان‌های اشاره و کاربردهای آموزشی و سرگرمی (ارتباط کودکان با محیط مجازی).

۲- حالت سر و صورت: مانند تکان دادن و حرکات سر، جهت نگاه چشم، بالا بردن ابروها، باز کردن لب‌ها برای صحبت، چشمک زدن، تغییر شکل بینی و صورت‌های تعجب، خوشحالی، نفرت، ترس، خشم، ناراحتی، تحقیر و غیره.

۳- حالت بدن: شامل حرکت یک بخش و یا کل بدن در دنبال کردن حرکات ناشی از اثر متقابل دو فرد در فضای بیرون، آنالیز حرکات یک رقاص برای ایجاد هماهنگی بین نمایش و موزیک (آهنگ)، تشخیص نحوه راه رفتن افراد برای توان بخشی پزشکی و آموزش‌های ورزشی.

به طور نوعی مفهوم یک اشاره می‌تواند از اطلاعات زیر بدست آید:

- اطلاعات مکانی: موقعیت و مکان اجرا
- اطلاعات مسیر: خط سیر اشاره
- اطلاعات نمادین: شکل و علامتی که می‌سازد
- اطلاعات مؤثر: میزان احساسات

مرحله حالت‌های چهره شامل استخراج ویژگی‌های احساس (مربوط به وضعیت احساسی فرد) از نقاط اختصاصی صورت مانند نواحی اطراف لب، بینی و چشم‌ها در یک تصویر نرمال شده است. اغلب برای بدست آوردن ویژگی‌های مناسب فریم‌های متحرک، این نواحی پیدا و دنبال و جدا می‌شوند. مکان، شدت نور و حرکات قسمت‌های فعال صورت برای درک و تشخیص حالت‌های چهره مهم هستند. علاوه بر این، اندازه‌گیری شدت حالت‌های غیر ارادی و فوری چهره نسبت به حالت‌های ثابت بسیار مشکل‌تر می‌باشد. گاهی نشانه‌هایی از قبیل کشیدگی دستان، کشیدگی ماهیچه‌های سرتاسر بدن، مکان نقاط تماس به خودی و گشادگی مردمک چشم نیز جهت راهنمایی بیشتر به کار می‌روند.

۱-۲ ضرورت انجام تحقیق

به طوری که در مطالب فوق ذکر گردید تشخیص اشارات و علائم در برقراری ارتباط هوشمند و کارآمد بین انسان و رایانه از بیشترین درجه اهمیت برخوردار می‌باشد و در رنج وسیعی کاربرد دارد [1] از جمله آن می‌توان به موارد زیر اشاره کرد:

- ۱- پیشرفت و گسترش کمک به افرادی که دارای مشکل شنوایی هستند
- ۲- قادر ساختن بچه‌های بسیار کوچک به برقراری ارتباط با رایانه
- ۳- تکنیک‌های طراحی برای تشخیص هویت
- ۴- تشخیص زبان اشاره
- ۵- نمایش وضعیت هیجان بیماران و یا سطح استرس آن‌ها از لحاظ طبی
- ۶- پیدا کردن موقعیت
- ۷- هدایت و یا اداره کردن در محیط مجازی
- ۸- برقراری ارتباطات در کنفرانس‌های ویدئویی
- ۹- آموزش و فراگیری از راه دور

علاوه بر جنبه تئوری موضوع، هر گونه پیاده سازی عملی تشخیص اشاره به طور نوعی نیاز به ابزار و وسایل مختلف برای تصویربرداری و دنبال کردن تصاویر دارد. این ابزار شامل دستکش‌های مجهز، لباس‌های مناسب، و نشانه‌هایی برای ردگیری نوری (از نظر میزان روشنایی و یا رنگ) هستند. روش‌های مرسوم دو بعدی مانند استفاده از مداد، کیبرد، ماوس‌های جهت دار به عنوان رابط‌های گرافیکی کاربر اغلب برای کار در محیط حقیقی مناسب نیستند، بلکه ترجیحاً لوازمی که موقعیت و جهت قسمت‌های مختلف بدن (مانند سر و دست)، راستای نگاه کردن، صحبت و صدا، حالت‌های چهره، پاسخ برقی پوست و جنبه‌های دیگر انسان می‌تواند برای مدل کردن ارتباط انسان و محیط اطراف استفاده شود.

برای اینکه همه این نمودها تعیین شوند، نیاز به دریافتن موقعیت، بیکره‌بندی (زوایا و چرخش‌ها) و حرکات (سرعت‌ها) بدن انسان داریم. این کار می‌تواند توسط دستگاه‌های حسگر که به کاربر متصل می‌شوند؛ از قبیل حسگرها و دنبال‌کننده‌های مغناطیسی، دستکش‌های مجهز برای انتقال داده و لباس‌های مناسب و یا با استفاده از دوربین‌ها و تکنولوژی بینایی کامپیوتر انجام پذیرد.

هر تکنولوژی حسگر در راستای چندین بعد از جمله دقت، میزان تفکیک‌پذیری، تأخیر، محدوده کاری، سهولت کار و در نهایت هزینه متفاوت است. در استفاده از واسطه‌های اشارات و علائم که بر اساس دستکش‌ها هستند به صورت نوعی کاربر نیاز به پوشیدن یک دستگاه طاقت فرسا و حمل یک کابل ارتباط برای وصل کردن دستگاه به کامپیوتر دارد. این مساله مانع راحتی کاربر و آسانی عکس العمل متقابل کاربر با کامپیوتر می‌شود. در شکل (۱-۱) نمونه‌هایی از انواع دستکش‌های مجهز نشان داده شده است. شگرد فنی که بر اساس بینایی می‌باشد با این که بر این مشکل غلبه کرده نیاز به ستیز با مشکلات دیگری که وابسته به انسداد بدن انسان است دارد. اگر دستگاه‌های ردیابی‌کننده توانایی آشکارسازی دقیق و پر سرعت حرکات انگشتان را زمانی که دستان کاربر حرکت می‌کند

داشته باشند آنگاه یک دستگاه بر اساس بینایی می‌تواند در بهترین حالت یک درک کلی از نحوه‌ی حرکت انگشتان به ما بدهد. پس دستگاه‌هایی که مبنی بر بینایی هستند می‌توانند ویژگی‌هایی همچون بافت و رنگ را برای موشکافی بیشتر اشارات (با اینکه دستگاه‌های مبنی بر ردیابی نمی‌توانند) بکار گیرند. علاوه بر این، فنون مربوط به بینایی در موارد زیر می‌توانند متغیر باشند:



(الف)



(ب)



(ج)



(د)

شکل (۱۰۱) : نمونه‌هایی از انواع دستکش‌های مجهز: الف) دستکش مجهز به پنج سنسور شتاب سنج، ب) دستکش داده 5DT که پنج درجه آزادی را اندازه گیری می کند، ج) دستکش با ۱۰ تا ۱۶ سنسور خمشی، د) دستکش مجهز به ۱۸ تا ۲۲ سنسور خمشی

۱- تعداد دوربینهایی که به کار گرفته می‌شود

۲- سرعت و تأخیر آنها

۳- ساختار محیط اطراف (مانند میزان تابش و یا سرعت حرکات)

۴- لوازم مورد نیاز برای کاربر (آیا کاربر باید لباس ویژه‌ای بپوشد؟)

۵- ویژگیهای سطح پایین که به کار گرفته می‌شود (لبه‌ها، نواحی، نیمرخ‌ها، زمان‌ها و

هیستوگرام)

۶- آیا نمایش در دو بعد به کار گرفته شده یا سه بعد.

۷- آیا زمان نیز در نظر گرفته شده است

با این وجود در نگاشت داده‌ها از فضای سه بعدی به فضای دو بعدی بخشی از اطلاعات به طور ذاتی از بین می‌رود. از طرف دیگر ساختن استادانه و ماهرانه مدل سه بعدی با سختی زیاد همراه است و یک فضای پارامتری بیشتر دارد. همچنین یک دنبال‌کننده نیاز به توانایی حس کردن تغییرات اشکال و اندازه‌های قسمت‌هایی از بدن که اشاره و علائم را به وجود می‌آورند (در بین افراد مختلف متفاوتند)، اشیاء دیگر که در پس زمینه حرکت می‌کنند و نویز دارد.

کار تشخیص اشارات و علائم، یک مثال مناسب از پژوهش چند رشته‌ای است و ابزار زیادی بر اساس روش‌هایی از مدل‌های آماری، بینایی کامپیوتر، تشخیص الگو، پردازش تصویر، سیستم‌های ارتباطی و ... برای تشخیص اشاره وجود دارند.

روش‌های آماری: بیشتر مشکلات توسط مدل‌های آماری از قبیل PCA، HMM [3],[4],[5]، کالمن فیلترینگ^۱ [7]، فیلترهای تکه‌ای پیشرفته‌تر^۲ [8],[9] و الگوریتم‌های خلاصه^۳ [10]-[12],[13] حل شده‌اند.

روش‌های FMS^۴ [14]-[17] نیز در مدل‌سازی حرکات بدن انسان به طور کاملاً مؤثری به کار گرفته شده‌است.

¹ Kalman filtering

² more advanced particle filtering

³ condensation algorithms

⁴ finite-state machine

بینایی کامپیوتر و تشخیص الگو [18]: شامل فنونی از جمله بیرون کشیدن ویژگی‌ها، دنبال کردن اشکال، دسته بندی و کلاسه بندی بوده و به طور موفقیت آمیزی در بسیاری از سیستم‌های تشخیص اشاره به کار گرفته شده‌اند.

پردازش تصویر [19]: فنون این روش نیز شامل کاوش و پیدا کردن اشکال، بافت، حرکات، جریان‌های نوری [20],[21]، بهبودی تصویر، جداسازی و مدل‌سازی نمای کرانه‌ای [22] می‌باشد.

روش‌های ارتباط: سیستم‌های ارتباط [23] نیز از قبیل MPL^۱ [24]، [25]، [26]، [27]^۲ TDNN و [28]^۳ RBFN، [29]، به همان صورت مورد استفاده قرار گرفته‌اند.

با اینکه تشخیص اشارات ایستا (علائم ثابت و بدون حرکت^۴) به طور نوعی توسط تطبیق الگوها، تشخیص الگوی استاندارد و شبکه‌های عصبی انجام می‌پذیرد تشخیص اشارات پویا نیاز به فنون دیگری مانند الگوهای فشرده‌سازی زمان، پیچش زمان به صورت پویا^۵، زنجیره‌ی HMM و TDNN دارد.

۱-۳ اهداف

۱- ساخت یک سیستم تشخیص اشاره با قابلیت گسترش دامنه‌ی لغات: برای رسیدن به این هدف سعی شده است تا از بیشتر ویژگی‌های ممکن در روش تشخیص اشاره بر اساس بینایی استفاده شوند. این ویژگی‌ها عبارتند از منحنی مسیر حرکت دست‌ها نسبت به سر، نسبت

¹ involving multilayer perceptron

² time delay neural network

³ radial basis function network

⁴ static

⁵ Dynamic Time Warping

ارتفاع به طول و مساحت مستطیل محاط بر ناحیه دست، بافت دست از نظر ماکزیمم جهت لبه و طول آن.

۲- ایجاد یک پایگاه داده مناسب از اشارات جدا از هم: برای آزمایش دقیق تر ویژگی های مورد نظر در تشخیص اشاره لازم است یک پایگاه داده حاوی اطلاعات کافی از روی یک پایگاه داده معتبر که در دسترس همه پژوهشگران قرار دارد ساخته شود. این اطلاعات شامل مکان سر، مختصات مستطیل محاط بر دست، تصویر دست در داخل این مستطیل محاط و تصویر باینری لبه ی دست می باشد.

۳- استخراج ویژگی تنها بر اساس مشاهده: هدف از بکارگیری ویژگی هایی که تنها بر اساس مشاهده بدست می آیند این است که بتوان آن ها را در سیستمی بکار برد که مبتنی بر یک دوربین فیلم برداری در یک مکان ثابت هستند و کاربر نیازی به تجهیزات ویژه مانند دستکش های رنگی، سنسورهای ردیابی، لباس های مخصوص و دوربین های قابل حمل برای استخراج ویژگی ندارد. بنابراین، ساختار سیستم کاربردی و کار با آن راحت تر خواهد بود.

۴- انعطاف پذیری در برابر تغییرات اجرا: برای پایداری کلاسه بندی در برابر تغییرات اجراهای متفاوت یک اشاره از فاصله ی اقلیدوسی بین نمونه های دو منحنی ویژگی های مرتبط با هم استفاده می شود، به نحوی که ابتدا توسط یک الگوریتم زمانی بصورت پویا نمونه هایی که مشابه هم هستند بهم متناظر شده و سپس مجموع فاصله ی این نمونه های متناظر محاسبه می شوند.

۵- استخراج ویژگی های هندسی: منحنی مسیر حرکت دست اشاره کننده به همراه شکل هندسی دست مهمترین نقش را در اجرای یک اشاره دارد. بنابراین، مکان دست، محدوده ی دست، شکل قرارگیری انگشتان و جهت آن ها در هر علامت باید استخراج شوند؛ بصورتی که مستقل از فرد اجرا کننده و زمان اجرای آن اشاره باشد. به این ترتیب دقت سیستم کلاسه بندی نسبت به تغییر فرد اجرا کننده و یا اجراهای متفاوت وی دچار افت نمی شود.

۶- افزایش سرعت تشخیص اشاره: اطلاعات درون فریم‌های نزدیک بهم تقریباً مشابه است. بنابراین، با کاهش تعداد فریم‌ها در ثانیه می‌توان بدون اینکه دقت کلاسه‌بندی تغییر چندانی داشته باشد سرعت اجرای تشخیص را افزایش داد. فریم‌ها، همه جزئیاتی که بیشتر مربوط به اشاره ایفا شده توسط شخص خاص و زمان معین هستند را در خود دارند، که استفاده از همه فریم‌ها علاوه بر اینکه جزئیات را افزایش داده و اطلاعات نویزی را در بر خواهند داشت، باعث کاهش دقت و سرعت تشخیص اشاره می‌شود.

۷- دسته‌بندی اشارات: نحوه‌ی ادای یک اشاره بسیار متنوع می‌باشد و حتی یک شخص خاص نیز ممکن است به شکل‌های مختلف مفهوم یک اشاره را برساند. به همین خاطر از همه نمونه‌های یک کلاس برای مقایسه و تشخیص اشاره مورد آزمایش استفاده می‌گردد. برای این کار ابتدا از همه ویژگی‌های منحنی مسیر حرکت دست، منحنی شیب وتر مستطیل محاط و مساحت آن کلاس‌هایی نزدیک‌تر با رنگ مناسب انتخاب و سپس توسط ویژگی منحنی حرکت، کلاسه‌بندی با رنگ کمتر انجام می‌گیرد که در نهایت با توجه به نوع بافت لبه، کلاس اشاره مورد نظر تشخیص داده می‌شود.

۱- ۴ اشاره‌ای به محتوای فصل‌های بعد

در فصول بعد به شرح کارها و تئوری‌های مورد استفاده پرداخته می‌شود به طوری که فصل دوم شامل: مروری بر کارهای انجام شده در زمینه تشخیص زبان اشاره بر اساس کلمات جدا از هم به همراه روش انجام کار، پایگاه داده مورد استفاده و نتایج آن می‌باشد. در فصل سوم تئوری‌های به کار رفته به همراه تئوری پیشنهادی بیشترین راستای لبه بعلاوه ارزش آن در این پروژه معرفی و توضیح داده می‌شوند. فصل چهارم پایگاه داده مورد استفاده و نحوه ساخت پایگاه داده نهائی به همراه مشخصه‌های آن را در بر می‌گیرد. ویژگی‌های به کار رفته در این پروژه و الگوریتم‌های پیشنهادی در فصل پنجم

معرفی و علل استفاده از آنها بیان می‌شود. در فصل ششم به محاسبه و استخراج نتایج پرداخته‌شده‌ها -
اینکه جداول و نمودارهای مربوط به آن نمایش داده می‌شود و در نهایت نتایج به دست آمده مورد
تجزیه و تحلیل قرار گرفته و با روش‌های دیگر مقایسه می‌شود.

فصل دوم

مطالعات پیشین

مقدمه

برای تشخیص سخن، دست‌نوشته و یا زبان اشاره روش‌های بسیاری پیشنهاد شده است که از ترکیب انحنای زمانی به صورت پویا (DTW) یا مدل‌های HMM با کلاسه‌بندی‌هایی با قابلیت تشخیص ساخته شده‌اند. در این بخش سعی بر این می‌شود تا در مورد سابقه تحقیق در رابطه با تشخیص زبان اشاره و روش‌های پیشنهادی آن تا به حال بحث شود.

۲-۱- تشخیص علائم قسمت‌بندی شده:

یکی از اولین سیستم‌هایی که برای تشخیص علائم قسمت‌بندی شده بر اساس مشاهده به کار می‌رود در [30]، آورده شده است. در این سیستم برای آشکارسازی دو دست و مکان انگشتان از دست-کش‌های رنگی استفاده می‌شود و پایگاه داده در این مجموعه شامل ۲۶۲ علامت (Sign language of Netherlands) بوده که توسط دو شخص منفرد، به صورتی که رنگ لباس اجراکننده‌ها با رنگ پس-زمینه برابر است، برای آموزش در این پروژه اجرا شده است. روش کلاسه‌بندی خطی (rule – based)، مکان نشانه‌ها را همراه با محور عمودی بدن پیدا می‌کند و بردار ویژگی شامل شکل دست، جهت دست‌ها، جهت انگشتان دست اشاره‌گر و مکان دست نسبت به شانه و بردار عمودی بدن می‌باشد. برای تشخیص اشارات به صورت مستقل از شخص ارائه‌کننده، از کلاسه‌بندی مدل مارکوف مخفی (HMM) استفاده شد که به ما نرخ خطا را بین ۷٪ تا ۹٪ می‌دهد. اگر از همه فیلم‌های ویدئویی مربوط به دو شخص، هم برای آموزش و هم برای آزمایش استفاده شود، در این صورت نرخ خطا به ۸٪ می‌رسد.

در [31] از یک روش آشکارساز بر اساس دنبال کردن مدل (Model based tracking) استفاده شده که فرض می‌کند شکل دست در فریم‌های پشت سرهم تغییر چندانی نمی‌کند. در این سیستم الگوریتم آشکارساز، نیاز به یک پیام برای شروع کار دارد. به این صورت که در ابتدا باید دست تنها

شیء متحرک درون فیلم باشد و علاوه بر آن فرد اشاره‌گر باید لباس‌های سیاه با آستین بلند پوشیده و در جلو یک پس‌زمینه تیره قرار گیرد. از مکان دست که آشکار شده است برای استخراج ویژگی‌هایی از قبیل مسیر حرکت، جهت دست و تبدیل فوریه شکل دست استفاده می‌شود. ویژگی‌های مورد نظر تنها از فریم‌هایی استخراج می‌شوند که در آنها شکل دست تغییر کرده و کلاسه‌بندی توسط شبکه عصبی سه بعدی (3D Hopfield neural network) با نرخ خطای ۹٪ برای ۱۵ علامت انجام شده است.

در [32] الگوریتم آشکارکننده نیاز به دستکش‌های رنگی برای مکان‌یابی و بدست آوردن مدل سه-بعدی دست دارد و علائم به کار رفته، مکان شروع و پایان معینی برای دست دارند. در این سیستم شکل و مدل دست تخمین زده می‌شود و اطلاعات به دست آمده وارد یک سیستم هوشمند فازی به صورت انطباق‌پذیر^۱ می‌شود که ۲۲ علامت را (Australian sign language mixed with artificial signs) با نرخ خطای ۵٪ کلاسه‌بندی می‌نماید و کار آموزش و آزمایش با یک اجرا برای هر علامت انجام شده است.

در [33] از یک شبکه عصبی تاخیر زمانی^۲ برای کلاسه‌بندی استفاده شده است. این سیستم از منحنی‌های سطوح پیکسلی دست به عنوان ویژگی استفاده می‌کند و این خطوط توسط تکه کردن تصاویر و حرکت به دست می‌آیند. مکان دست و سر توسط یک الگوریتم رنگ پوست و ساختار هندسی پیدا شده و انتخاب می‌گردد. در این حالت سیستم برای ۴۰ علامت (ASL) و تعداد زیادی از نمونه‌ها، دارای نرخ خطای ۴٪ بوده است.

در [34] سیستمی شرح داده شده است که زبان اشاره ژاپنی را تشخیص می‌دهد و در مرحله اول با استفاده از حالت خاص شخص اشاره‌گر مکان دست، سر و آرنج آشکار می‌شود. از این نواحی برای

¹ Adaptive fuzzy expert system

² Time – delay neural network

جدا کردن رنگ، دست، سر، آرنج و لباس‌ها استفاده می‌شود و مهم نیست که پس‌زمینه شلوغ باشد یا نه. دست و صورت با یک پردازش اولیه از روی الگوی ساختار مشخص می‌شوند. ویژگی‌های هندسی شکل دست، مکان دست و جهت حرکت آن نیز از نواحی منطقه‌بندی و جداشده مربوط به دست استخراج شده و از اطلاعات هر دو دست برای آموزش HMM استفاده می‌شود. در این روش برای آزمایش ۶۵ نمونه انتخاب گردید که سیستم طبقه‌بندی برای این نمونه‌ها هیچ خطائی را نشان نداد.

در [35] ویژگی هندسی انگشتان و هر دو دست اشاره‌کننده توسط آشکارسازی مردمک، دستکش-های رنگی، جداسازی پس‌زمینه و قوانین هندسی استخراج شدند. یک مدل HMM بوسیله ۴ نمونه از هر علامت با فرهنگ اشاره‌ی ۴۳۹ علامتی (Chinese sign language) آموزش دیده است. یک نمونه از هر علامت برای تشخیص اشاره مربوط به یک شخص به کار رفته که نتیجه آن نرخ خطائی برابر با ۷٪ بوده است. در ادامه آزمایش‌ها از تراکم ترکیبی بسته‌شده^۱ استفاده گردید که با نرخ خطای ۸٪ سرعت آن دو برابر حالت قبل شده است.

یک بردار ویژگی باینری که بردار ویژگی زبانی^۲ نامیده می‌شود در [36] پیشنهاد شده است و این سیستم از یک کلاسه‌بندی دو بخشی بهره می‌برد. در مرحله اول، مکان حرکت و شکل دست کلاسه‌بندی می‌شوند و نتیجه این کلاسه‌بندی در یک بردار ویژگی باینری ۳۴ بعدی قرار داده می‌شود. در مرحله دوم از آنالیز مولفه‌های مستقل (ICA)^۳ برای نگاشت بردار ویژگی باینری به یک فضای ویژگی اقلیدسی با ابعاد کمتر استفاده می‌شود. کلاسه‌بندی توسط زنجیره HMM صورت می‌گیرد، به نحوی که برای آموزش تنها از یک نمونه در ۴۹ علامت (British sign language) استفاده می‌شود و نرخ خطای آن ۱۶٪ گزارش شده است.

¹ Tied mixture densities

² Linguistic texture

³ Independent component analysis

در [37] سیستمی شرح داده شده که از ویژگی‌های هر دو دست استفاده می‌کند و توسط رنگ پوست و قوانین هندسی، مکان سر و صورت مشخص می‌گردد. زمینه تصویر نیز با کمک فیلتر مدین (Median) در سطوح پیکسلی جدا و حذف می‌شود. ناحیه دست با کمک رنگ پوست، یک مدل زیست - مکانیکی از بدن و فرضیات کمکی دیگر جدا و پی‌گیری می‌شود. مکان دست، ویژگی‌های هندسی و مشتقات آن‌ها (در مجموع ۲۲ ویژگی) با استفاده از مکان و اندازه صورت نرمالیزه می‌شوند. کلاسه‌بندی نرخ خطائی برابر با ۱٪ را (در ۲۲۹ اشاره) برای تنها یک شخص اشاره‌گر ارائه می‌دهد.

آزمایش برای افراد متفاوت از چهار فرد (سه نفر برای آموزش و یک نفر برای آزمایش) نرخ خطائی بین ۶۹٪ و ۵۶٪ به طور وابسته به نوع تعیین گروه آموزشی و شخص اشاره‌گر می‌دهد. تکنیک انطباق با اشاره‌گر که در این سیستم پیاده‌سازی می‌شود، در [38] ارائه شده است و با استفاده از انطباق بازنگری شده و همچنین مستقل از فرد اشاره‌گر برای ۱۵۳ علامت به نرخ خطائی برابر با ۲٪ دست پیدا کرد.

یک سیستم وابسته به اشاره‌گر همراه با استفاده از ویژگی‌های غیر دستی (non- manual) در [39] ارائه شده است. ویژگی‌های دیگری از جمله مکان سر، نگاه چشم و شکل دور لب در این پروژه به کار رفتند. آزمایش‌ها روی پایگاه داده‌ای با ۱۴۵ علامت انجام و تنها برای ویژگی‌های غیر دستی نرخ خطای ۳۶٪ به دست آمد. در حالیکه نرخ خطا تنها برای ویژگی‌های دستی ۳٪ شد و ترکیب هر دو نوع ویژگی باعث بهبود نتیجه و رسیدن به نرخ خطای ۲۲/۰٪ گشته است.

در [40] یک الگوریتم تشخیص زبان اشاره به صورت تغییرناپذیر با جهت دید، پیاده شده که تنها از یک دوربین استفاده می‌کند. روش پیش‌نهادی، کار تشخیص زمانی - مکانی را به کار تطبیق یک سیستم بینائی دوبعدی تبدیل می‌کند و در آن از روش دمپاستر شفر^۱ برای بهبود دقت تشخیص استفاده شده است. روش پیش‌نهادی برای یک فرهنگ اشاره با ۱۰۰ علامت (Chinese sign)

¹ Dempster_ shafer

language) با رنگ ۲ و دقت ۹۲٪ دست پیدا کرده است. علاوه بر آن روش پیاده‌سازی شده می‌تواند برای کارهای مشابه دیگری از جمله تشخیص افراد از روی راه رفتن (Gait) و تشخیص صحبت از روی لب‌ها به کار رود.

۱- برای آشکارسازی و استخراج ویژگی مکان و شکل دست از دست‌کش‌های رنگی استفاده شده است

۲- برای محاسبه فاصله بین منحنی‌ها نیز الگوریتم DTW به کار رفته است

در [41] یک الگوریتم موثر و پر قدرت برای تشخیص اشاره به صورت کلمات جدا شده از هم در علائم انگلیسی (SEE)^۱ ارائه شده که شامل کلاسه‌بندی‌هایی برای شکل دست، حرکت دست و مکان دست می‌باشد. داده‌های مربوط به شکل و حالت SEE با استفاده از دست‌کش‌های مخصوص (Cyber Glove) و آشکارسازهای مغناطیسی به دست آمد. یک کلاسه‌بندی به صورت درخت تصمیم‌گیری خطی با تابع تصمیم‌گیری خطی فیشر (FLD)^۲ برای ۲۷ شکل دست استفاده شد، که منحنی مسیر حرکت دست نیز با استفاده از آنالیز مولفه‌های اصلی کوانتیزه‌برداری (VQPCA)^۳ کلاسه‌بندی گردید. هم علائم تناوبی و هم علائم غیر تناوبی در SEE از مسیرهای سه‌بعدی حرکت دست تشخیص داده شدند. آزمایشات انجام شده برای تشخیص شکل دست از اشاره‌گرهائی که در آموزش نقشی نداشته‌اند، به دقت متوسط ۹۶٪ دست پیدا کرد و تشخیص علائم تناوبی و غیر تناوبی از روی منحنی مسیر حرکت دست توسط VQPDA به ترتیب ۳/۹۷٪ و ۹۷٪ بوده که ترتیب آنها با یک کلاسه‌بندی مکان دست برای تشخیص علامت از روی ۲۸ اشاره به دقت ۸/۹۶٪ رسید.

¹ Signing Exact English

² Fisher's Linear Discriminate

³ Vector quantization principal component analysis

در [42] برای تشخیص سخن، دست‌نوشته و یا زبان اشاره روش‌های بسیاری پیشنهاد شده است که از ترکیب انحنای زمانی به صورت پویا (DTW) یا مدل‌های هیدن‌مارکوف (HMM) با کلاسه‌بندی‌هایی

با قابلیت تشخیص ساخته شده‌اند. هر چند، همه روش‌ها مستقیماً به مدل‌های بیشترین شباهت HMM/ DTW (likelihood) تکیه دارند اما از نظر طراحان این پروژه، انحنای زمانی و کلاسه‌بندی به دلیل تضاد نیازهای مدل‌سازی بیشترین شباهت باید از هم جدا شوند. برای غلبه بر این محدودیت‌ها از انحنای زمانی پویا به صورت آماری (SDTW)^۱ تنها برای انحنای زمانی استفاده شده و ویژگی‌های انحنای یافته توسط دو روش جدید پیشنهادی کلاسه‌بندی می‌شود:

۱- آشکارسازهای ویژگی با قابلیت تمایز به صورت ترکیبی (CDFD)^۲

۲- کلاسه‌بندی درجه دو روی نگاشت فیشر DF (Q_DFFM)^۳

که هر دو از یک مجموعه ویژگی‌های متمایز (DFS) استفاده کرده و مشخص گردیده که این روش‌ها بهتر از HMM و SDTW عمل می‌کنند. ترکیب چندین روش بیشترین تشابه در مرحله دوم کلاسه‌بندی، در حالی که کارایی کلاسه‌بندی پیشنهادی را کاهش می‌دهد باعث افزایش کارایی HMM و SDTW می‌گردد. یک آزمایش ایده‌ای مستدل، که نگاشت DFFM از مدل‌های SDTW چندتایی را با تشابه SDTW ترکیب می‌کند؛ پیشرفت خوبی نسبت به SDTW نشان می‌دهد. اگرچه در اصل تشخیص بر اساس ویژگی‌های سه‌بعدی حرکت دست می‌باشد اما می‌توان از جزئیات بیشتری مانند

¹ Statistical DTW

² Combined discriminative feature detectors

³ Quadratic Classification on DF Fisher Mapping

شکل و حالت دست، بدن و حالت چهره برای استخراج ویژگی استفاده کرد. در نهایت این الگوریتم توانست برای ۱۲۰ اشاره متفاوت^۱(DSL) که توسط ۷۵ نفر اجرا شد، به دقت ۹۷.۵٪ دست پیدا کند.

^۱ Dutch Sign Language

فصل سوم

تئوری‌های مورد استفاده

مقدمه

جهت انجام این تحقیق از تئوری‌ها و روش‌های گوناگونی همچون روش دنبال کردن، مستطیل محاط بر دست، آشکارساز لبه Canny، DTW، LMED و ... استفاده شده است. این روش‌ها به ما کمک می‌کند تا تشخیص زبان اشاره از دقت بالاتری برخوردار باشد. در این فصل به بیان و تشریح تئوری‌های مورد استفاده پرداخته می‌شود.

۳-۱ دنبال کننده

روش‌های دنبال کردن در بسیاری از پروژه‌های مختلف کاربرد دارند. از جمله این موارد می‌توان به تشخیص اشاره، دنبال کردن حرکت انسان، تشخیص صورت، دیده‌بانی هوایی توسط رادار و نظارت بر رفت و آمد وسایل اشاره کرد. وقتی در یک فیلم ویدئویی اطلاعات مفیدی از اشیاء متحرک نیاز باشد از الگوریتم‌های ردیابی استفاده می‌شود، بعنوان مثال الگوریتم ردیابی می‌تواند برای گرفتن دیفرانسیل چندین هدف متحرک و ردیابی مکان حرکت آن‌ها برای استخراج اطلاعات تصویری پرمحتوا و آشکار کردن زمان بهم رسیدن و همپوشانی دو دست ردیابی شده به کار گرفته شود. بعضی از روش‌های دنبال کردن که به طور معمول استفاده می‌شوند در مراجع زیر آمده‌اند؛ بعنوان مثال روش ردیابی چگالش^۱ در [43]، فیلتر کالمن^۲ در [44]، ردیابی مین‌شیفت^۳ در [45] و ردیابی

¹ Condensation tracking

² Kalman filtering

³ Mean shift tracking

کم شیفت^۱ در [46]، ردیابی مستطیل محاط^۲ در [47] و دنبال کردن اشیاء با اندازه ثابت و متغیر به صورت پویا^۳.

از آنجا که در این پروژه کار دنبال کردن و مکان‌یابی دست از اهداف مورد نظر نمی‌باشد در نتیجه به بررسی جامع مزایا و معایب این روش‌های معمول نمی‌پردازیم. یک بررسی کلی و جامع مناسب از برخی روش‌های متداول ردیابی که در تشخیص شکل، حالت و حرکت اشخاص بکار می‌روند در [48] آمده است.

الگوریتم ردیابی چگالش (انتشار چگالی شرطی)^۴ در [43] پیاده‌سازی شده است. این الگوریتم یک روش بر اساس مدل‌سازی می‌باشد و قابلیت ردیابی اشیاء در محیط‌های شلوغ را دارد.

الگوریتم مین‌شیفت که در [45] پیاده‌سازی شده است حتی اشیاء انعطاف‌پذیر را بر اساس ویژگی‌های مبنی بر بینائی از قبیل رنگ و بافت ردیابی می‌کند. توزیع‌های آماری نیز برای مشخص کردن شیئی مورد نظر به کار گرفته می‌شود، این الگوریتم در برابر انسداد جزئی شیئی ردیابی‌شونده، در هم آمیختگی، گردش در عمق تصویر (بصورت افقی) و تغییرات در مکان دوربین فیلم‌برداری به خوبی مقاوم است.

الگوریتم کم شیفت (حالت ادامه‌دار و انطباقی مین شیفت) در واقع شکل توسعه‌یافته الگوریتم مین شیفت می‌باشد که قابلیت رسیدگی کردن به تغییرات توزیع‌های احتمال رنگ به صورت

¹ Continuously adaptive mean shift

² Bounding _ Box tracking

³ Dynamic tracking

⁴ Conditional Density Propagation

پویا را دارد. در سیستمی که در [46] پیاده‌سازی شده است، این الگوریتم برای ردیابی صورت انسان با سرعت واقعی^۱ به کار رفته است.

مکان‌یابی بر اساس مشاهده و ردیابی اجزای بدن برای برقراری ارتباط بینایی که در [49] و [50] بررسی و کنکاش شده است، نیاز به دنبال کردن اشیاء در سیستم‌های آشکارسازی با عملکرد تشخیص بر اساس مشاهده را نشان می‌دهد. در [51] نیز یک روش بر اساس مشاهده برای طبقه‌بندی افراد بصورت سازگار و برای آشکارسازی و ایجاد ارتباط با استفاده از اجزای بدن انسان با کمک یک دوربین فیلم‌برداری پیاده‌سازی شده است.

یک مدل جدیدتر روش ردیابی نسبت به روش‌های قبلی، ردیابی با استفاده از برنامه‌ریزی به صورت پویا^۲ می‌باشد که همراه با کاربرد آن در تشخیص زبان اشاره در [52] به تفصیل بیان شده است؛ این الگوریتم قادر است سر و دست غالب شخص اشاره‌گر را با دقتی مناسب برای استخراج ویژگی‌های مربوط به دست و غیر از دست (حالت چهره و بدن) ردیابی کند. در این روش از برنامه‌ریزی پویا برای شکل‌دهی عملیات ردیابی به صورت یک مساله بهینه‌سازی استفاده می‌شود تا تمام مسیر شیء مورد نظر را به بهترین نحو ممکن تشخیص دهد و نیازی به مقداردهی اولیه برای ردیابی و یا مدل‌های از پیش تعریف شده ندارد [53].

در روش ردیابی مبنی بر بینایی عموماً از یک تصویر کوچک شده و یا تغییر شکل یافته بعنوان بردار ویژگی استفاده می‌گردد. وقتی اندازه یک تصویر کاهش می‌یابد؛ بعنوان مثال یک تصویر 320×240 مبدل به یک تصویر 32×24 می‌شود؛ بسیاری از اطلاعات و جزئیات مربوط به بینایی به خاطر نحوه عملکرد کاهش اندازه از دست می‌رود. به همین دلیل ما به روشی نیاز داریم که تنها بخش مناسب با هدف را از تصاویر جدا کند. نواحی مطلوب پس از استخراج با عنوان بردار ویژگی

¹ Real Time

² Dynamic programing

به کار می‌روند و تغییرات مورد نیاز تنها روی این نواحی صورت می‌گیرد. برای دستیابی به این هدف می‌توان از روش ردیابی مستطیل محاط استفاده نمود. مختصات مستطیل به گونه‌ای تعیین می‌شود تا در حد امکان تنها نواحی مطلوب را در بر بگیرد.

۳ ۲ مستطیل محاط بر دست^۱

می‌توان با پیدا کردن ناحیه‌ی مربوط به شکل دست و تبدیل آن به یک تصویر باینری توسط یک حد آستانه‌ی مناسب و لیبل‌بندی^۲ ناحیه‌ی مربوط به دست، مختصات مستطیل محاط بر دست را بدست آورد.

۳ ۳ پیدا کردن لبه

تغییرات و گسستگی‌های موجود در ویژگی دامنه روشنایی^۳ یا مقادیر نسبی مرتبط با رنگ‌های اصلی^۴ که برای ایجاد رنگ‌های دیگر با هم ترکیب می‌شوند، اساساً مشخصات اصلی یک تصویر هستند و این تغییرات و گسستگی‌ها اغلب نشانه‌ای از اندازه فیزیکی اشیاء داخل تصویر به ما می‌دهند. در این بخش به تشریح لبه روشنایی در گسستگی دامنه تصویر تنها بین نواحی تصویر که به صورت محلی هموار^۵ شده‌اند می‌پردازیم. گسستگی محلی در روشنایی تصویر از یک سطح به سطح دیگر لبه روشنایی^۶ نامیده می‌شود که برخی از انواع لبه در شکل (۲-۳) آورده شده است.

¹ Bounding box

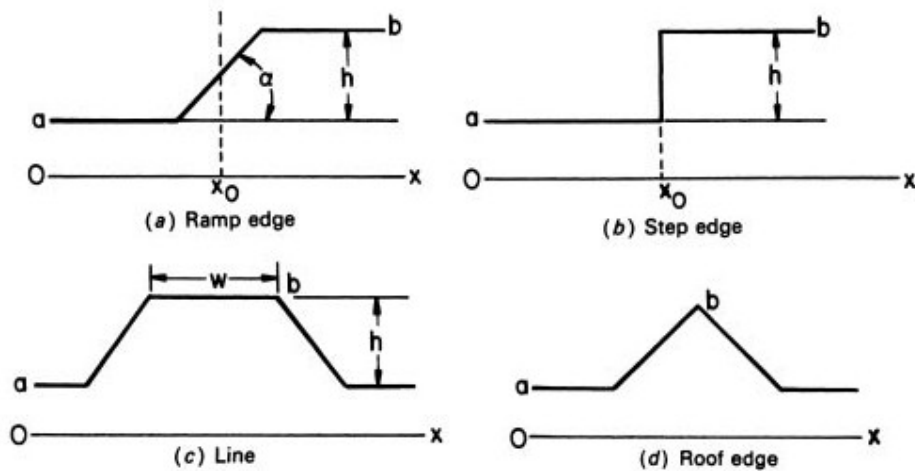
² Labeling

³ Luminance

⁴ Tristimulus value

⁵ Smooth

⁶ Luminance Edge



شکل (۲-۳): مدل‌های یک بعدی لبه و خط در فضای پیوسته

۳-۳-۱ آشکار ساز لبه با مشتق مرتبه اول

دو راه اساسی برای به دست آوردن گرادیان لبه با مشتق مرتبه اول وجود دارد؛ روش اول شامل به دست آوردن گرادیان در دو جهت متعامد، در داخل تصویر می‌باشد و روش دوم از مجموعه‌ای از مشتق‌های جهت‌دار استفاده می‌کند. یک لبه را که در یک ناحیه $F(x, y)$ (ناحیه‌ای از فضای پیوسته) قرار دارد می‌توان با شکل‌دهی یک گرادیان یک بعدی $G(x, y)$ در راستای خط عمود بر شیب لبه به دست آورد. اگر گرادیان به اندازه کافی بزرگ باشد (بیشتر از یک حد آستانه) فرض می‌شود که در آن ناحیه یک لبه وجود دارد. گرادیان در راستای خط عمود بر شیب لبه می‌تواند توسط دو مولفه مشتق عمود بر هم در راستاهای x و y به صورتی که در زیر آمده است محاسبه شود.

$$G(x, y) = \frac{\partial F(x, y)}{\partial x} \cos \theta + \frac{\partial F(x, y)}{\partial y} \sin \theta \quad (۱-۳)$$

تصویر (۳-۳) نحوه تولید یک گرادیان لبه $G(x, y)$ را در فضای گسسته نشان می‌دهد که مولفه سطری آن $G_R(j, k)$ و مولفه ستونی آن $G_C(j, k)$ می‌باشد^۱. دامنه گرادیان دو بعدی با فرمول زیر به دست می‌آید:

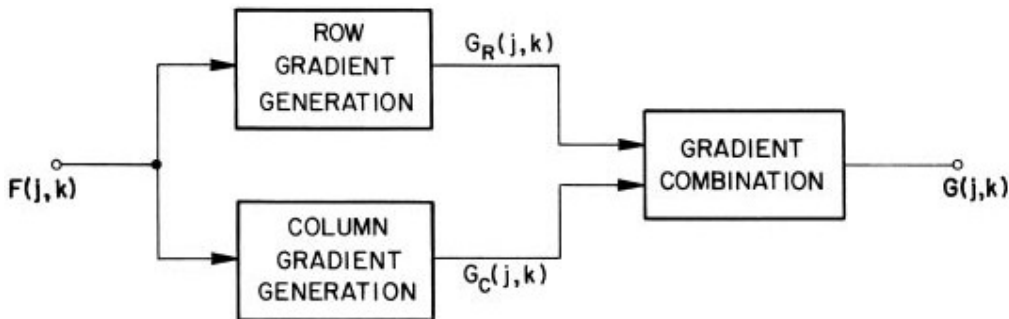
$$G(j, k) = [[G_R(j, k)]^2 + [G_C(j, k)]^2]^{1/2} \quad (۲-۳)$$

گاهی برای افزایش راندمان محاسباتی، دامنه گرادیان با ترکیب قدر مطلق دامنه مولفه‌های عمود بر هم تخمین زده می‌شود:

$$G(j, k) = |G_R(j, k)| + |G_C(j, k)| \quad (۳-۳)$$

و زاویه گرادیان دو بعدی نسبت به بردار سطر به صورت زیر به دست می‌آید:

$$\theta(j, k) = \tan^{-1} \left\{ \frac{G_C(j, k)}{G_R(j, k)} \right\} \quad (۴-۳)$$



شکل (۳-۳): سازنده گرادیان متعامد

نکته باقی‌مانده برای به دست آوردن گرادیان متعامد در فضای گسسته انتخاب یک تقریب گسسته مناسب از مشتق پیوسته رابطه (۱-۳) می‌باشد.

^۱ نامگذاری آرایه‌های مختصات به این صورت است که مبدأ در گوشه سمت چپ و بالای تصویر قرار می‌گیرد در این حالت z به صورت افقی و k به صورت عمودی افزایش می‌یابند.

۳-۳-۲ عملگر گرادیان در همسایگی کوچک

راحت‌ترین راه به دست آوردن گرادیان گسسته، اندازه‌گیری اختلاف سطح پیکسل‌ها در راستای سطرها و ستون‌های تصویر می‌باشد. گرادیان سطری به صورت زیر تعریف می‌شود

$$G_R(j, k) = F(j, k) - F(j - 1, k) \quad (۵-۳)$$

و گرادیان ستونی نیز از رابطه زیر به دست می‌آید^۱:

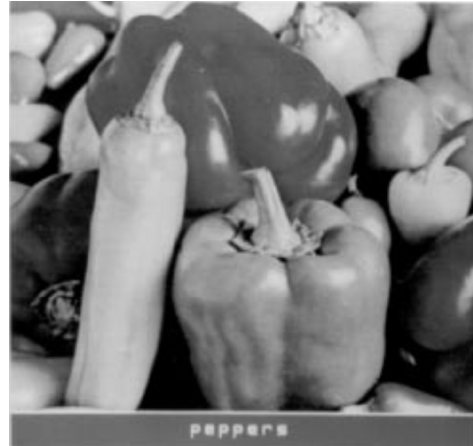
$$G_C(j, k) = F(j, k) - F(j, k + 1) \quad (۶-۳)$$

این آشکارساز لبه نمی‌تواند برای لبه‌های شیبدار^۲ مکان لبه را به پیکسل مناسب نسبت دهد. شکل (۳-۴) تصاویری از گرادیان تفاضلی افقی و عمودی تصویر سیاه و سفید Peppers به همراه عکس اصلی نشان می‌دهد.

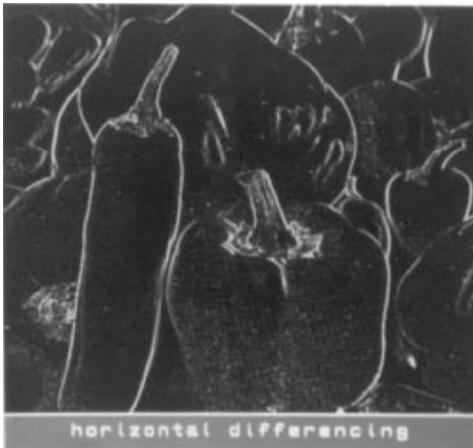
در این شکل، تصاویر حاصل از گرادیان با حداکثر کنتراست تصویر به نمایش درآمده است. از تصاویر نشان داده شده کاملاً مشخص است که فن آشکارسازی لبه با گرادیان تفاضلی به شدت به نوسانات کوچک روشنایی در تصویر حساس بوده و همچنین مرزهای بین اشیاء به خوبی ترسیم نشده‌اند.

۱ این تعاریف از گرادیان سطری و ستونی و همچنین بقیه موارد به این ترتیب صورت گرفته است تا G_R و G_C برای یک لبه که در آن دامنه روشنایی از چپ به راست و از پائین به بالا افزایش پیدا می‌کند، مثبت شود.

^۲ Ramp Edge



(الف)



(ج)



(ب)

شکل (۳-۴): گرادیان تفاضلی عمودی و افقی از تصویر peppers_mon (الف) تصویر اصلی، (ب) گرادیان عمودی، (ج) گرادیان افقی

گرادیان لبه‌های اریب را می‌توان با بکارگیری اختلاف سطح جفت پیکسل‌های قطری به دست آورد. این اساس Roberts [54] در جهت عملگر تفاضلی می‌باشد که حالت قدرمطلق آن به صورت زیر تعریف می‌شود:

$$G(j, k) = |G_1(j, k)| + |G_2(j, k)| \quad (۳-۷)$$

و حالت رادیکال مجموع مربعات آن به شکل زیر است:

$$G(j, k) = [[G_1(j, k)]^2 + [G_2(j, k)]^2]^{1/2} \quad (۳-۸)$$

که در این دو داریم:

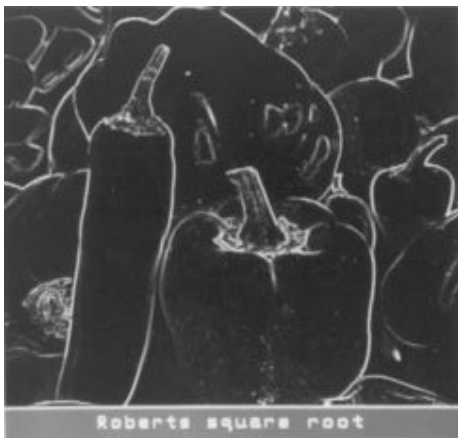
$$G_1(j, k) = F(j, k) - F(j - 1, k + 1) \quad (9-3)$$

$$G_2(j, k) = F(j, k) - F(j + 1, k + 1) \quad (10-3)$$

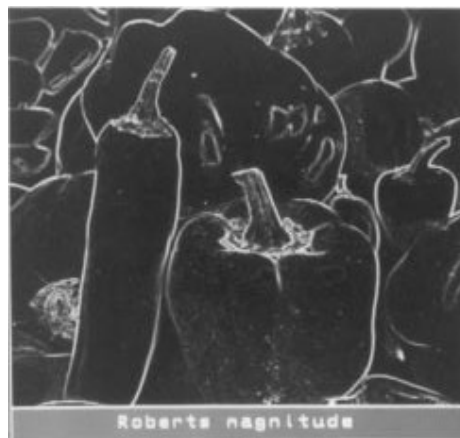
و زاویه لبه نسبت به بردار سطر از رابطه زیر به دست می‌آید:

$$\theta(j, k) = \frac{\pi}{4} + \tan^{-1} \left\{ \frac{G_2(j, k)}{G_1(j, k)} \right\} \quad (11-3)$$

در شکل (۳-۵) گرادیان تصویر Peppers_mon برای عملگر Roberts نمایش داده شده است. با نگاه به این تصاویر می‌توان دریافت که اشیاء در تصویر گرادیان لبه حاصل از رادیکال مجموع مربعات Roberts نسبت به قدرمطلق آن کمی بهتر مشخص و متمایز شده‌اند.



(ب)



(الف)

شکل (۳-۵): گرادیان Roberts از تصویر peppers_mon (الف) قدرمطلق، (ب) رادیکال مجموع مربعات

برای به دست آوردن مکان مرکز لبه از نوع شیب‌دار می‌توان اختلاف پیکسل‌ها را که در دو طرف پیکسل مرکزی قرار دارند محاسبه کرد. با این کار گرادیان‌های سطری و ستونی به شکل زیر در می‌آیند:

$$G_R(j, k) = F(j + 1, k) - F(j - 1, k) \quad (۱۲-۳)$$

$$G_C(j, k) = F(j, k - 1) - F(j, k + 1) \quad (۱۳-۳)$$

هر چند لبه شیب‌دار به خوبی مکان‌یابی شده اما روش گرادیان تفاضلی پیکسل‌های جدا از هم نیز به شدت به نوسانات روشنایی کم در تصویر حساسیت نشان می‌دهد. برای اینکه مشکل مذکور تا اندازه‌ای حل شود می‌توان از یک عملگر مولد گرادیان دو بعدی استفاده کرد؛ به نحوی که در یک جهت مختصات مقدار اختلاف میان پیکسل‌ها را به دست آورده و همزمان در جهت عمود بر آن مقدار متوسط پیکسل‌ها را محاسبه کند. این کار توسط یک عملگر گرادیان لبه با ابعاد 3×3 پیکسل که در شکل (۶-۳) نشان داده شده است انجام می‌گیرد.

A_0	A_1	A_2
A_7	$F(j, k)$	A_3
A_6	A_5	A_4

شکل (۶-۳): شماره گذاری قراردادی برای عملگر آشکارساز لبه با ابعاد 3×3

در این عملگر، گرادیان سطری و ستونی لبه را روابط زیر به دست می‌آید:

$$G_R(j, k) = \frac{1}{K+2} [(A_2 + KA_3 + A_4) - (A_0 + KA_7 + A_6)] \quad (۱۴-۳)$$

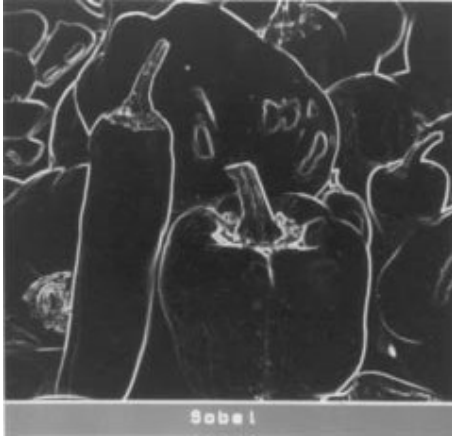
$$G_C(j, k) = \frac{1}{K+2} [(A_0 + KA_1 + A_2) - (A_6 + KA_5 + A_4)] \quad (۱۵-۳)$$

در این تعاریف گرادیان‌های سطری و ستونی، جهت به دست آوردن بهره واحد برای میانگین تخمین توازی در دو طرف مکان لبه، با تقسیم بر $K + 2$ نرمالیزه می‌شوند.

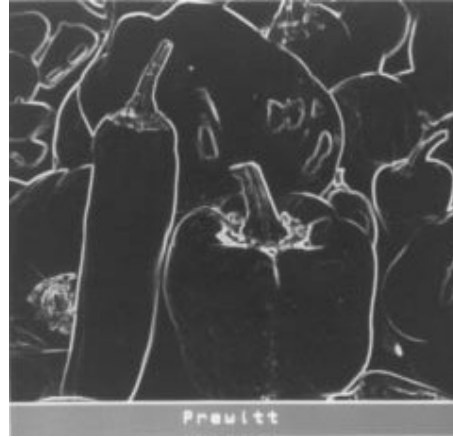
چنانچه داشته باشیم $K=1$ در اینصورت آشکارساز لبه، عملگر Prewitt ([55].P.108) نامیده می‌شود. در آشکارساز لبه Sobel ([56].P.271) از همان عملگر استفاده می‌شود، با این تفاوت که مقدار K برابر با ۲ می‌باشد. انگیزه این وزن‌دهی برای K ، دادن اهمیت متمایز به هر یک از پیکسل‌ها به عنوان سهم آن در گرادیان دو بعدی در آن ناحیه می‌باشد. در روش آشکارساز لبه Frei- Chen [57] مقدار K برابر با $\sqrt{2}$ در نظر گرفته شده تا گرادیان برای لبه‌های افقی، عمودی و قطری به صورت یکسان عمل کند. در این روش گرادیان لبه در مرکز لبه برای مدل‌های لبه تک پیکسلی و شیب‌دار یکسان به دست می‌آید. در حالیکه گرادیان Prewitt برای لبه قطری ۹۴/۰ برابر آن در لبه عمودی و گرادیان Sobel برای لبه قطری ۰۶/۱ برابر آن در لبه عمودی می‌باشد. از این رو مدل Prewitt حساسیت بیشتری نسبت به لبه‌های عمودی و افقی نشان می‌دهد و برعکس عملگر Sobel حساسیت بیشتری نسبت به لبه‌های قطری دارد. گرادیان‌های روی یک قطر در راستای یک لبه از نوع شیب‌دار قطری با لبه‌های افقی یا قطری در هر سه روش آشکارساز لبه 3×3 متفاوت است و هیچکدام از این آشکارسازها قابلیت نسبت دادن ناحیه لبه به یک پیکسل را ندارند. شکل (۳-۷) مثال‌هایی از گرادیان‌های Prewitt، Sobel و Frei- Chen را نشان می‌دهد. با دیدن این مثال‌ها می‌توان دریافت که روش‌های ذکر شده نسبت به عملگر Robert لبه‌های اشیاء را بهتر ترسیم می‌کنند. یکی از مهمترین دلایل این امر بزرگ‌تر بودن اندازه پنجره می‌باشد که با یک متوسط‌گیری در جهت عمود راستای گرادیان اثر نوسانات کوچک روشنایی را کاهش می‌دهد.

۳-۳-۳ عملگرهای گرادیان در همسایگی بزرگ

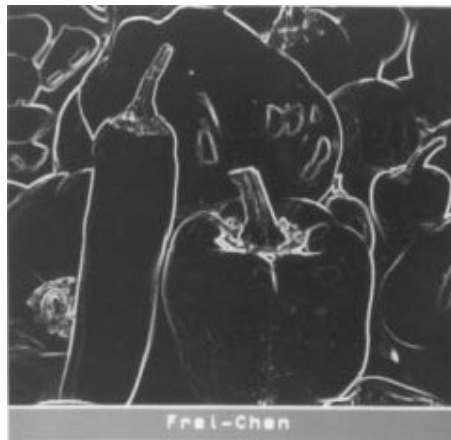
عملگرهای گرادیانی ذکر شده در بخش قبل در آشکارسازی لبه‌های تصویر همراه با نویز زیاد توانایی کمی دارند که این مشکل با افزایش ابعاد پنجره‌ای که در آن گرادیان تفاضلی محاسبه می‌شود به اندازه مناسب تا حدودی حل می‌شود.



(ب)



(الف)



(ج)

شکل (۷-۳): گرادیان‌های Sobel، Prewitt و Frei-chen از تصویر peppers_mon (الف: Prewitt، ب: Sobel، ج: Frei-chen)

به عنوان چند مثال عملگر Prewitt دارای یک پاسخ گرادیانی سطری با ابعاد 7×7 به فرم زیر است:

$$H_R = \frac{1}{21} \begin{bmatrix} 1 & 1 & 1 & 0 & -1 & -1 & -1 \\ 1 & 1 & 1 & 0 & -1 & -1 & -1 \\ 1 & 1 & 1 & 0 & -1 & -1 & -1 \\ 1 & 1 & 1 & 0 & -1 & -1 & -1 \\ 1 & 1 & 1 & 0 & -1 & -1 & -1 \\ 1 & 1 & 1 & 0 & -1 & -1 & -1 \\ 1 & 1 & 1 & 0 & -1 & -1 & -1 \end{bmatrix} \quad (۱۶-۳)$$

که این عملگر boxcar نامیده می‌شود.

عملگر Truncated pyramid [58] دارای اوزانی می‌باشد که با فاصله گرفتن از مرکز لبه به صورت

خطی کاهش می‌یابد و پاسخ گرادیانی سطری سطری 7×7 آن به شکل زیر می‌باشد:

$$H_R = \frac{1}{34} \begin{bmatrix} 1 & 1 & 1 & 0 & -1 & -1 & -1 \\ 1 & 2 & 2 & 0 & -2 & -2 & -1 \\ 1 & 2 & 3 & 0 & -3 & -2 & -1 \\ 1 & 2 & 3 & 0 & -3 & -2 & -1 \\ 1 & 2 & 3 & 0 & -3 & -2 & -1 \\ 1 & 2 & 2 & 0 & -2 & -2 & -1 \\ 1 & 1 & 1 & 0 & -1 & -1 & -1 \end{bmatrix} \quad (17-3)$$

عملگرهای دیگری نیز وجود دارند که برای مقابله با نویز از توابع وزن‌دهی به فرم گوسی و با همسایگی بزرگ استفاده کرده‌اند. از آن جمله می‌توان به Argyle [59]، Maclead [60,61] و FDOG [62] اشاره کرد.

فرض کنید تابع:

$$g(x, s) = [2\pi s^2]^{-1/2} \exp\{-1/2(x/s)^2\} \quad (18-3)$$

یک تابع گوسی در فضای پیوسته با انحراف معیار s باشد، در اینصورت پاسخ ضربه¹ عملگر Argyle

در راستای افق و در فضای پیوسته به صورت زیر تعریف می‌شود:

$$H_R(j, k) = \begin{cases} -2g(x, s)g(y, t) & \text{for } x \geq 0 \\ 2g(x, s)g(y, t) & \text{for } x < 0 \end{cases} \quad (19-3)$$

و همین طور تابع پاسخ ضربه گرادیان افقی عملگر Maclead در فضای پیوسته به شکل:

$$H_R(j, k) = [g(x + s, s) - g(x - s, s)]g(y, t) \quad (20-3)$$

می‌باشد. عملگر FDOG¹ مشتق یک تابع گوسی در راستای مورد نظر است که پاسخ ضربه افقی

آن در فضای پیوسته از رابطه:

¹ Impulse response

$$H_R(j, k) = \frac{-\partial(g(x, s)g(y, t))}{\partial x} \quad (۲۱-۳)$$

به دست می‌آید که بعد از مشتق‌گیری داریم:

$$H_R(j, k) = \frac{-xg(x, s)g(y, t)}{s^2} \quad (۲۲-۳)$$

در شکل (۸-۳) گرادیان‌های Boxcar، Truncated pyramid، Argyle، Macleod و FDOG از تصویر Peppers_mon با هم مقایسه شده‌اند.

۳-۳-۴ عملگر گرادیان Canny

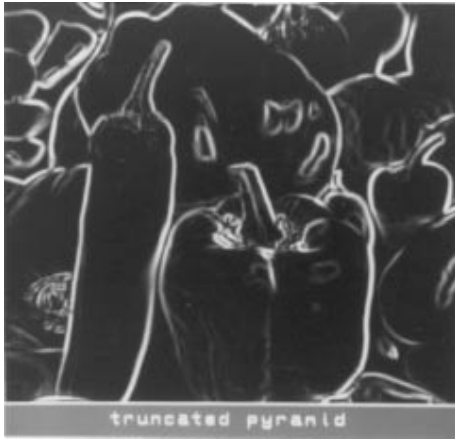
Canny [62] از یک روش پردازش با کمک طرح عملگرهایی که در بخش قبل ذکر شد استفاده می‌کند. اساس پیشرفت Canny بر این استوار بوده است که بتواند یک لبه پله‌ای با دامنه s به همراه یک نویز گوسی با انحراف معیار σ_n را آشکار نماید.

فرض کنید آشکارسازی لبه توسط کانال^۲ یک سیگنال لبه نویزی $f(x)$ با یک تابع پاسخ ضربه نا متقارن $h(x)$ شکل بگیرد به نحوی که دامنه آن در خارج از رنج $[-W, W]$ صفر باشد. در این حالت ماکزیموم محلی گرادیان کانالو $f(x) \otimes h(x)$ به عنوان لبه شناخته می‌شود. عملگر Canny یعنی $h(x)$ به صورتی انتخاب می‌شود که شرایط زیر را برقرار نماید:

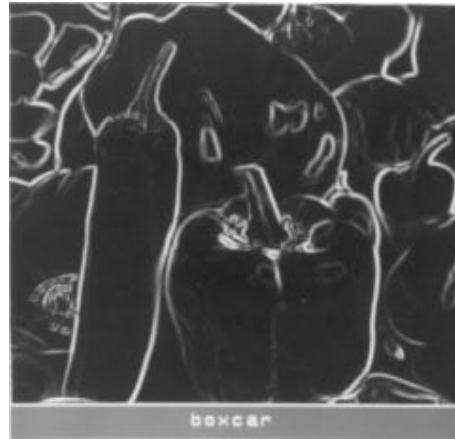
۰۱- آشکارسازی خوب: دامنه سیگنال به نویز (SNR) به بیشترین مقدار برسد تا احتمال از دست رفتن مکان لبه‌های مناسب و یا مکان‌های اشتباه بعنوان لبه کاهش یابد. مقدار (SNR) برای مدل گفته شده [62] به حالت زیر می‌باشد:

¹ First derivative of Gaussians

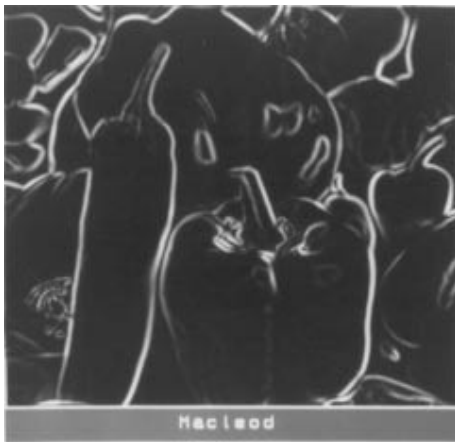
² Convolve



(ب)



(الف)



(د)



(ج)



(ه)

شکل (۳-۸): گرادیان‌های در همسایگی بزرگ از تصویر peppers (الف) 7×7 Boxcar، (ب) 9×9 Truncated pyramid، (ج) Argyle 11×11 ، $s = 2.0$ ، (د) Macleod 11×11 ، $s = 2.0$ ، (ه) FDOG 11×11 ، $s = 2.0$

$$SNR = \frac{\left| \int_{-W}^W h(x)f(-x)dx \right|}{\sigma_n \sqrt{\int_{-W}^W [h(x)]^2 dx}} \quad (23-3)$$

که به صورت زیر ساده می‌شود:

$$SNR = \frac{s \int_{-W}^0 h(x)dx}{\sigma_n \int_{-W}^W [h(x)]^2 dx} \quad (24-3)$$

۲- مکان‌یابی خوب: نقاطی که توسط عملگر بعنوان لبه مشخص می‌شوند باید تا جائی که ممکن

است به مرکز لبه نزدیک باشند. ضریب مکان‌یابی [62] به صورت زیر تعریف می‌شود:

$$LOC = \frac{\left| \int_{-W}^W h'(x)f'(-x)dx \right|}{\sigma_n \sqrt{\int_{-W}^W [h'(x)]^2 dx}} \quad (25-3)$$

که به فرم [63]:

$$LOC = \frac{s|h'(0)|}{\sigma_n \sqrt{\int_{-W}^W [h'(x)]^2 dx}} \quad (26-3)$$

درمی‌آید و در آن $h'(x)$ برابر است با مشتق $h(x)$.

۳- پاسخ یکتا: باید تنها یک پاسخ برای یک لبه مناسب داشته باشیم. حداقل فاصله بین

ماکزیموم‌های محلی گرادیان؛ x_m ؛ زمانی که تنها نویز داشته باشیم برابر با کسری از ضریب W می-

باشد. به این ترتیب داریم:

$$x_m = kW \quad (27-3)$$

بنابراین Canny سه معیار سنجش را همزمان با هم ترکیب کرده است؛ حداکثر نمودن مقدار

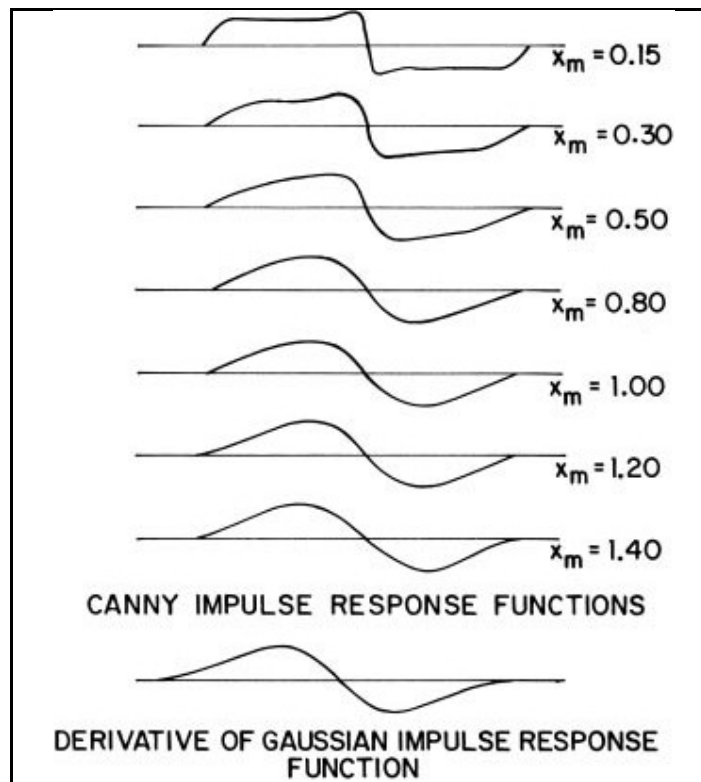
سیگنال به نویز، حداکثر نمودن دقت مکان‌یابی و پاسخ یکتا در ناحیه معین. به دلیل پیچیدگی روابط

داده شده تا کنون هیچ راهی به صورت تئوری ارائه نشده اما، یک روش محاسباتی به صورت متغیر بر

حسب x_m ایجاد شده است. شکل (۹-۳) شامل نمودارهایی از تابع پاسخ ضربه Canny می‌باشد که بر

حسب مقادیر مختلف x_m رسم شده‌اند. همان‌گونه که از نمودارها بر می‌آید برای مقادیر کم x_m تابع

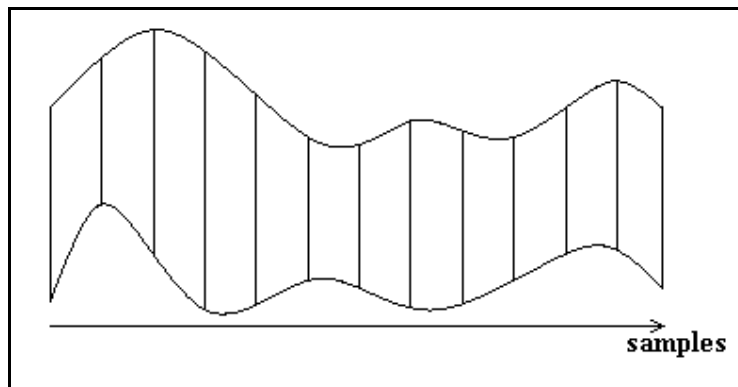
Canny مشابه یک تابع Boxcar عمل می‌کند. در حالیکه برای مقادیر بزرگ x_m رفتار تابع Canny بیشتر شبیه تخمین FDOG می‌شود. می‌توان عملگرهای بزرگ در فضای گسسته را با نمونه‌گیری از پاسخ ضربه عملگرهای بزرگ در فضای پیوسته در یک محدوده $W \times W$ به دست آورد. ابعاد این پنجره باید به اندازه‌ای انتخاب شود تا پله‌های ایجاد شده با نمونه‌گیری از تابع پاسخ ضربه موجب تاثیرات فرکانس بالا نشود [64].



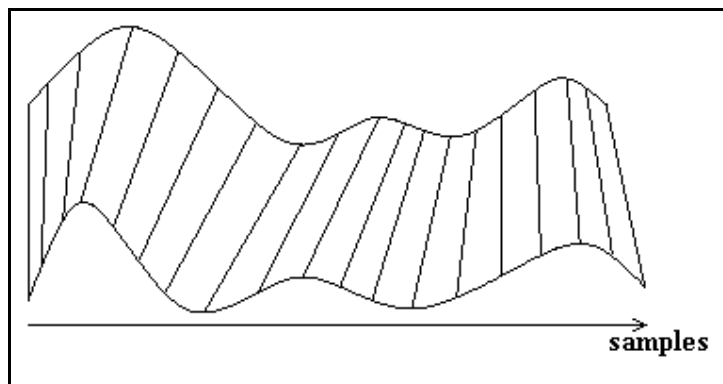
شکل (۹-۳): مقایسه توابع پاسخ ضربه Canny و FDOG

۳ ۴ ۱ DTW

در پردازش سیگنال و به ویژه در زمینه پردازش صوت سری‌های مختلف اندازه‌گیری شده که ظاهراً شرایط یکسانی را در بر دارند (یک کلمه یا یک اشاره)، می‌توانند پیام‌هایی را فراهم کنند که در واقع تفاوت‌های معنی‌دار و قابل توجهی در دامنه دوره، سرعت ادا و ... نشان می‌دهند. الگوریتم DTW [65] به صورت مستقل از تفاوت‌های ممکن در دامنه، دوره و سرعت ادا، یک تطبیق زمانی میان دو رشته مشاهده شده انجام می‌دهد. همان‌طور که در شکل (۳-۱۰) مشاهده می‌شود، دو سیگنال که حاوی پیام مشابهی هستند یک بار به صورت خطی و یک بار با استفاده از الگوریتم DTW به هم تطبیق داده شده‌اند.



(الف)



(ب)

شکل (۳-۱۰): مقایسه تطبیق میان دو سیگنال ظاهراً مشابه (الف) تناظر هندسی، (ب) تناظر به کمک الگوریتم DTW

¹ Dynamic time warping

در موارد بخصوص تشخیص کلمات جدا شده از هم با استفاده از یک واژه‌نامه، شامل نحوه ادای کلمات ضبط شده است که باید تشخیص داده شوند و در حین عملیات تشخیص، سیستم کلاسه‌بندی باید بر اساس سیگنال مشاهده شده تصمیم بگیرد که کدام کلمه ادا شده است. اگر سیگنال صدا یا تصویر مربوط به یک کلمه بوده و بطور کامل ادا شده باشد ما براحتی می‌توانیم هر نمونه از سیگنال ورودی را با یک نمونه از سیگنال مرجع متناظر کنیم. برای اجرای مقایسه و تشخیص، می‌خواهیم بهترین ارتباط ممکن بین سیگنال‌های موجود در مرجع و سیگنالی که باید شناسائی شود را به دست آوریم و بهترین مقدار مناسب را برای ناهمخوانی پیدا کنیم. DTW هر دو عمل ذکر شده را اجرا می‌کند.

۳-۴-۱ الگوریتم DTW

دو دنباله از بردارهایی به صورت $\{x_1, x_2, \dots, x_I\}$ و $\{y_1, y_2, \dots, y_J\}$ در نظر بگیرید به گونه‌ای که هر مؤلفه آن شامل L بعد باشد، فرض کنید d

$$d(i, j) = \sqrt{\sum_{l=1}^d [x(i, l) - y(j, l)]^2} \quad (28-3)$$

فاصله بین دو بردار x_i و y_j باشد به طوری که $1 \leq i \leq I$ و $1 \leq j \leq J$. در این صورت یک راه حل برای تطبیق زمان، شامل گرفتن یک به یک اندیس‌های رشته‌های $\{x_i\}$ و y_j و استفاده از یک جفت تابع‌ها به صورت $\Phi = (\Phi_x, \Phi_y)$ که روی بازه $\{1, \dots, T\}$ تعریف شده در حداصل $\{1, \dots, I\}$ و $\{1, \dots, J\}$ می‌باشد. در این حالت برای محاسبه، حاصل جمع فاصله‌های مربوط به Φ_T داریم:

$$d\Phi(I, J) = M_\Phi^{-1} \sum_{K=1}^T m_k d(\Phi_x(k), \Phi_y(k)) \quad (29-3)$$

که در آن m_k ضرایب اوزان مثبت و M_\emptyset ثابت نرمالیزه کردن می‌باشد و از رابطه زیر به دست می‌آید:

$$M_\emptyset = \sum_{K=1}^T m_k \quad (30-3)$$

ظاهراً به نظر می‌رسد که M_\emptyset باید به انتخاب \emptyset ربط داشته باشد. در عمل مقادیر m_k طوری

انتخاب می‌شوند که M_\emptyset مستقل از \emptyset باشد. به طور مثال با انتخاب m_k به صورت:

$$m_k = \phi_x(k) - \phi_x(k-1) + \phi_y(k) - \phi_y(k-1) \quad (31-3)$$

یک تابع \emptyset موارد زیر را فراهم می‌کند:

۱- مقادیر ابتدا و انتهائی \emptyset یعنی:

$$\begin{cases} \phi_x(1) = 1, \phi_x(T) = I \\ \phi_y(1) = 1, \phi_y(T) = J \end{cases} \quad (32-3)$$

۲- محاسبه یکنواخت و پیوسته:

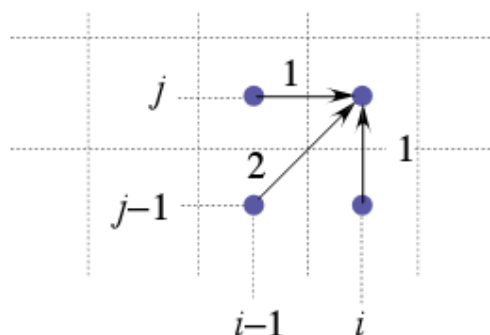
$$\begin{cases} 0 \leq \phi_x(k) - \phi_x(k-1) \leq 1 \\ 0 \leq \phi_y(k) - \phi_y(k-1) \leq 1 \end{cases} \quad (33-3)$$

۳- پیوستگی محلی: در نهایت جفت تابع $\phi_x(k)$ و $\phi_y(k)$ ادامه مسیر پیوسته را مشخص می-

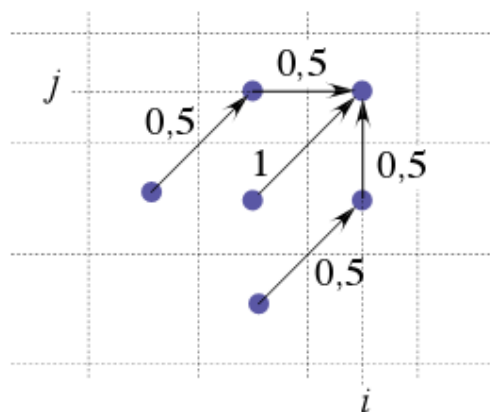
کند.

۳-۴-۲ مثال‌هایی از روش‌های پیدا کردن خط سیر برای اطمینان از پیوستگی محلی

یک رشته از مسیرهای ممکن توسط یک گراف، مشابه آنچه در شکل‌های (۱۱-۳) و (۱۲-۳) نشان داده شده است، تعریف می‌شوند. در این تصاویر پیکان‌ها نشان‌دهنده تنها مسیر ممکن هستند که از هر نقطه می‌توان به نقطه نهایی رسم کرد.



شکل (۱۱-۳): شرایط ادامه مسیر: مقادیری که وزن‌های مربوط به مسیرهای مورد نظر را نشان می‌دهند



شکل (۱۲-۳): شرایط ادامه مسیر: مقادیری که وزن‌های مربوط به مسیرهای مورد نظر را نشان می‌دهند

هدف از الگوریتم DTW آن است که کمترین مقدار $d\emptyset(I, J)$ را به دست آوریم:

(۳۴-۳)

$$D(I, J) = \min_{\emptyset \in \Phi} d\emptyset(I, J)$$

که در آن Φ مجموعه‌ای توابع است که شرایط لازم را فراهم کند

برنامه‌ریزی پویا یک روش بازیافتی است که اجازه می‌دهد معیار $D(I, J)$ با استفاده از مشخصات

زیر به کمترین مقدار برسد.

دو رشته $\{x_1, x_2, \dots, x_I\}$ و $\{y_1, y_2, \dots, y_J\}$ را در نظر بگیرید. فرض کنید $C(I, J)$ یک مسیر با کمترین طول حاصل از DTW باشد که بین این دو رشته کشیده شده است. در این حالت جزئی از مسیر $C(I, J)$ که به نقطه‌ای با مختصات (x_i, y_j) می‌رسد. کوچکترین مسیر برای دو دنباله $\{x_1, x_2, \dots, x_I\}$ و $\{y_1, y_2, \dots, y_J\}$ خواهد بود. دلیل این امر آن است که بین همه مسیرهایی که از ابتدا به نقطه با ابعاد (i, j) می‌رسند، کوتاه‌ترین مسیرها انتخاب شده‌اند. از این رو داریم:

(۳۵-۳)

$$D(i, j) = \min_{\emptyset \in \Phi} d\emptyset(I, J)$$

بنابراین برای گراف فرضی که در شکل (۱۱-۳) نشان داده شده است داریم:

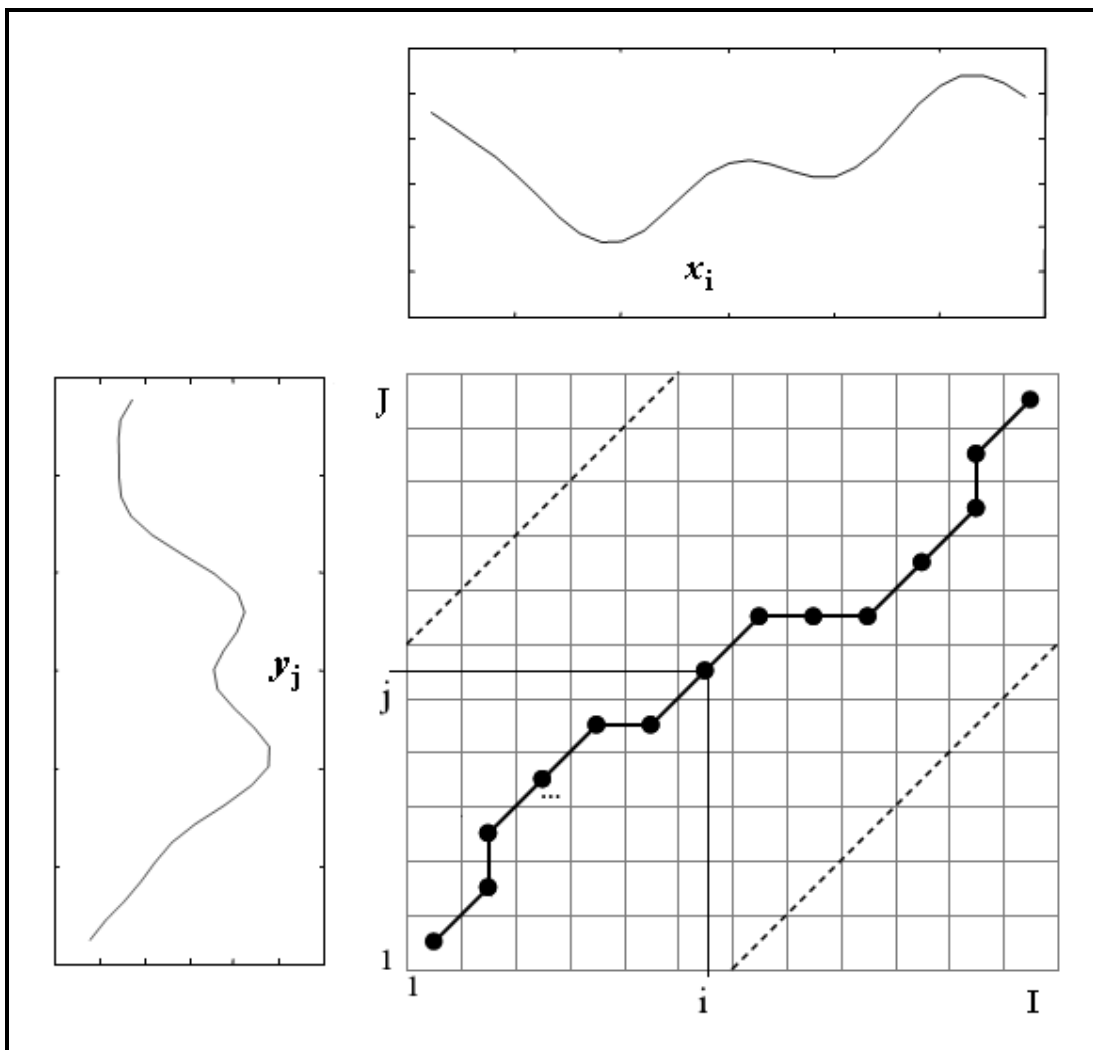
$$D(i, j) = \min \begin{cases} D(i-1, j) + d(i, j) \\ D(i-1, j-1) + 2d(i, j) \\ D(i, j-1) + d(i, j) \end{cases} \quad (۳۶-۳)$$

به همین ترتیب برای گراف شرطی شکل (۱۲-۳) داریم:

$$D(i, j) = \min \begin{cases} D(i-2, j-1) + 0.5d(i-1, j) + 0.5d(i, j) \\ D(i-1, j-1) + d(i, j) \\ D(i-1, j-2) + 0.5d(i, j-1) + 0.5d(i, j) \end{cases} \quad (۳۷-۳)$$

در شکل (۳-۱۳) مثالی از کوتاه‌ترین مسیر میان دو سیگنال x_i و y_j توسط الگوریتم DTW نشان داده شده است. شایان ذکر است که در این پژوهش از رابطه (۳-۳۸) برای تطبیق میان سیگنال‌های ورودی و پایگاه داده استفاده شده است. دلیل این انتخاب سادگی آن، کوتاه بودن طول سیگنال‌ها و آزادی عمل نسبتاً بالای دست‌ها برای ادای یک اشاره می‌باشد.

$$D(i, j) = d(i, j) + \min \begin{cases} D(i-1, j) \\ D(i-1, j-1) \\ D(i, j-1) \end{cases} \quad (۳-۳۸)$$

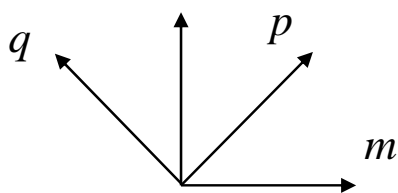


شکل (۳-۱۳): بدست آوردن کوتاه‌ترین مسیر میان دو سیگنال مشابه هم به کمک الگوریتم DTW

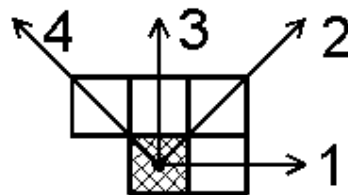
۳ ۵ راستای ماکزیمم لبه و طول آن در یک مجموعه (تئوری پیشنهادی)

برای بدست آوردن راستای ماکزیمم لبه در یک ناحیه تصویر از الگوریتم پیشنهادی زیر استفاده می‌کنیم:

- ۱- تصویر لبه شکل دست توسط یک روش مناسب مانند Canny گرفته می‌شود
- ۲- مختصات همه‌ی نقاط لبه بر حسب X و Y در تصویر لبه بدست آمده و در یک ماتریس $k \times 2$ ذخیره می‌شود که k تعداد نقاط مربوط به لبه است
- ۳- توسط ساختار الف شکل (۳-۱۴) در هر نقطه از لبه بررسی می‌گردد که آیا نقطه دیگری در مجاورت آن قرار دارد و اگر پاسخ مثبت است طبق جدول (۳-۱) در چه راستا و چه زاویه‌ای قرار دارد:



(ب)



(الف)

شکل (۳-۱۴): الف) ساختار پنجره مناسب برای پیدا کردن راستای لبه ب) نامگذاری شمارنده هر راستا

جدول (۳-۱): شماره، نام و زاویه‌ی هر راستا

جهت	شمارنده	راستا
1	m	$0^\circ, 180^\circ$
2	p	$45^\circ, 225^\circ$
3	n	$90^\circ, 270^\circ$
4	q	$135^\circ, 315^\circ$

۴- یک ماتریس E_a با اندازه‌ی 2×2 به صورت مطابق رابطه‌ی (۳۹-۳) ایجاد می‌گردد که a مربوط به ناحیه‌ای است که می‌خواهیم در آن میزان لبه‌ها را به دست آوریم.

$$E_a = \begin{bmatrix} m & n \\ p & q \end{bmatrix} \quad (۳۹-۳)$$

به طوری که m, n, p و q به صورت زیر تعریف می‌شوند:

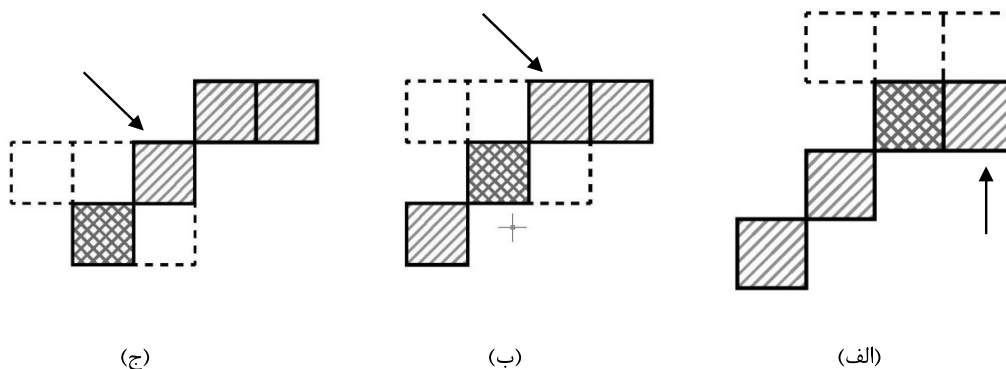
m : تعداد پیکسل‌های لبه در ناحیه a می‌باشد که یک پیکسل لبه در سمت راست آن (زاویه صفر درجه، ۱۸۰ درجه) قرار گرفته است

p : تعداد پیکسل‌های لبه در ناحیه a ، که یک پیکسل لبه در سمت راست و بالای آن (زاویه ۴۵ درجه، ۲۲۵ درجه) وجود دارد

n : تعداد پیکسل‌های لبه در ناحیه a ، که یک پیکسل لبه در بالای آن (زاویه ۹۰ درجه، ۲۷۰ درجه) قرار می‌گیرد

q : تعداد پیکسل‌های لبه در ناحیه a ، که یک پیکسل لبه در سمت چپ و بالای آن (زاویه ۱۳۵ درجه، ۳۱۵ درجه) قرار گرفته است.

که در اینصورت به نوعی ویژگی از ساختار لبه در ناحیه a دست پیدا می‌کنیم. بعنوان مثال برای تصویر باینری در شکل (۳-۱۵) که لبه‌ی یک تصویر فرضی می‌باشد با استفاده از ساختار ارائه شده پس از سه مرحله، m برابر با یک، p برابر با دو و n و q هر دو برابر با صفر می‌باشند. به این ترتیب ماتریس E_a یک نماینده جدید برای ویژگی لبه در ناحیه‌ی a است که به طور تقریبی میزان لبه را در چهار راستا تعیین می‌کند. یک پیکسل لبه می‌تواند همزمان در دو جهت یا سه جهت دیگر در مجاورت یک پیکسل لبه‌ی دیگر قرار گیرد، که در این صورت به متغیرهای مربوط به هر جهت یک واحد اضافه می‌شود.



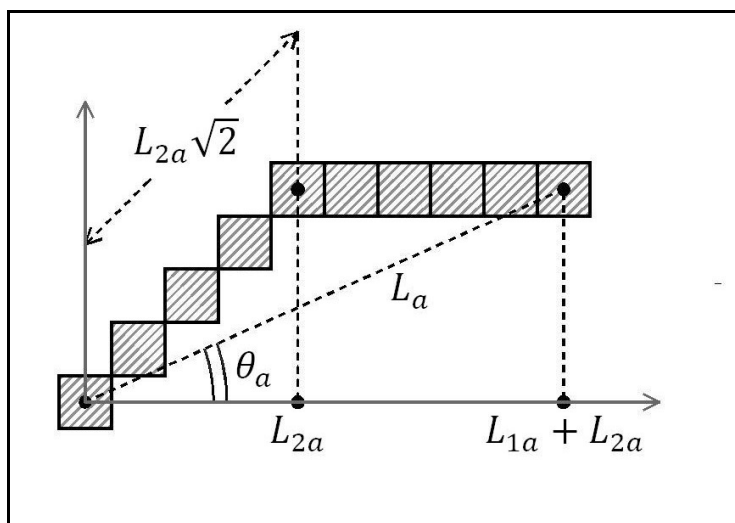
شکل (۱۵-۳): مراحل محاسبه‌ی n, m, p و q (الف) $m = 1, m = m + 1$ (ب) $p = 1, p = p + 1$ (ج)
 $p = 2, p = p + 1$

۵- برای بدست آوردن طول و زاویه‌ی راستای ماکزیموم لبه یعنی L_a و θ_a در ناحیه‌ی a بسته

به شرایط n, m, p و q مطابق جدول (۲-۳) چهار حالت پیش می‌آید.

بطور مثال اگر ساختار لبه در یک ناحیه به صورت شکل (۱۶-۳) باشد در این حالت برای مقادیر

n, m, p و q مطابق با بند چهار داریم:



۶- شکل (۱۶-۳): ساختار لبه در یک ناحیه کوچک

جدول (۲-۳): چهار حالت ممکن برای بدست آوردن طول و زاویه‌ی راستای ماکزیموم لبه در ناحیه‌ی a

$\begin{cases} m > n \\ p > q \end{cases} \Rightarrow \begin{cases} L_{1a} = m - n \\ L_{2a} = p - q \end{cases}$ $\Rightarrow \begin{cases} L_a = L_{1a} + L_{2a}\sqrt{2} \\ \theta_a = \tan^{-1}\left(\frac{L_{2a}}{L_{1a} + L_{2a}}\right) \end{cases}$	1
$\begin{cases} m > n \\ p < q \end{cases} \Rightarrow \begin{cases} L_{1a} = m - n \\ L_{2a} = q - p \end{cases}$ $\Rightarrow \begin{cases} L_a = L_{1a} + L_{2a}\sqrt{2} \\ \theta_a = \pi - \tan^{-1}\left(\frac{L_{2a}}{L_{1a} + L_{2a}}\right) \end{cases}$	2
$\begin{cases} m < n \\ p > q \end{cases} \Rightarrow \begin{cases} L_{1a} = n - m \\ L_{2a} = p - q \end{cases}$ $\Rightarrow \begin{cases} L_a = L_{1a} + L_{2a}\sqrt{2} \\ \theta_a = \frac{\pi}{2} - \tan^{-1}\left(\frac{L_{2a}}{L_{1a} + L_{2a}}\right) \end{cases}$	3
$\begin{cases} m < n \\ p < q \end{cases} \Rightarrow \begin{cases} L_{1a} = n - m \\ L_{2a} = q - p \end{cases}$ $\Rightarrow \begin{cases} L_a = L_{1a} + L_{2a}\sqrt{2} \\ \theta_a = \frac{\pi}{2} + \tan^{-1}\left(\frac{L_{2a}}{L_{1a} + L_{2a}}\right) \end{cases}$	4

$$\begin{cases} m = 5 \\ p = 4 \\ n = 0 \\ q = 0 \end{cases} \quad (۴۰-۳)$$

که نتیجه می‌دهد: $m > n$ و $p > q$. بنابراین طبق جدول (۲-۳) با کمک روابط مربوط به

حالت اول، مقادیر L_a و θ_a به صورت زیر محاسبه می‌شوند:

$$L_{1a} = 5 \quad L_{2a} = 4 \quad (41-3)$$

$$L_a = L_{1a} + L_{2a}\sqrt{2} \approx 10.66 \quad (42-3)$$

$$\theta_a = \tan^{-1}\left(\frac{L_{2a}}{L_{1a} + L_{2a}}\right) \approx 23.96^\circ \quad (43-3)$$

یک مزیت این ویژگی سرعت بالای بدست آوردن آن می‌باشد. زیرا پردازش تنها روی پیکسل‌هایی صورت می‌گیرد که جزو لبه باشند. البته یک پردازش کلی برای آشکار سازی لبه‌های مناسب لازم است. مزیت دیگر این ویژگی در انطباق تصاویری است که دارای لبه‌های ساختاری یکپارچه با تغییرات راستای کم باشند مانند تصویر لبه‌ی انگشتان دست در هنگام ادای یک اشاره. به این ترتیب که اگر تصاویر دارای ساختار مشابهی باشند با این ویژگی به خوبی مشخص می‌گردد.

فصل چهارم

پایگاه‌های داده^۱

^۱ databases

مقدمه

پایگاه‌های داده که برای سنجش و ارزشیابی سیستم‌های تشخیص زبان اشاره بر اساس مشاهده وجود دارند شامل فیلم‌هایی هستند که یک شخص را در حال اجرای یک علامت یا یک جمله از زبان اشاره نشان می‌دهند. هر فیلم توسط یک داده و یا یک توضیح (حاشیه نویسی)^۱ مربوط به اشاره اجرا شده نمایانده می‌شود. همه‌ی گروه‌های پژوهشی از پایگاه‌های مختلف برای سنجش سیستم‌های طراحی شده استفاده می‌کنند. این پایگاه‌های داده از بسیاری لحاظ با هم تفاوت دارند به طور مثال این اختلافات می‌تواند شامل طول زمان اجرا، فهرست کلمات به کار رفته، تعداد افراد اشاره‌گر و نوع تلفظ (نحوی ادای اشاره) باشد.

تقریباً بجز پایگاه داده‌ی 16-BOSTON201 هیچ پایگاه داده‌ی دیگری که شامل زبان اشاره‌ی پیوسته باشد و در دسترس همه‌ی گروه‌های پژوهشی برای تشخیص زبان اشاره قرار گیرد، وجود ندارد. پایگاه داده‌ی RWTH-Boston-104 [66] که بر گرفته از پایگاه داده‌ی 16-BOSTON201 می‌باشد نیز برای ارزیابی روش‌های پیاده‌سازی شده، در تشخیص زبان اشاره به کار گرفته شده است. در ادامه این پایگاه داده معرفی و ساختار آن توضیح داده می‌شود.

۴-۱ RWTH-Boston-104

پایگاه داده‌ی RWTH-Boston-104 یک پایگاه داده بر اساس زبان اشاره می‌باشد که توسط مرکز ملی تشخیص حالت و زبان اشاره‌ی دانشگاه بوستون^۲ انتشار یافته است. این پایگاه داده اساساً برای پژوهش بر روی ساختار صرف و نحو زبان اشاره‌ی آمریکا (ASL)^۳ [66] ثبت شده است. از این رو داده‌ها

¹ Annotated

² <http://www.bu.edu/asllrp/ncslgr.html>

³ American sign language

برای کار تشخیص بهینه‌سازی نشده و نسبت به پایگاه‌های داده که برای اهداف تشخیص زبان اشاره ساخته شده‌اند (به طور مثال پایگاه داده‌ی به کار گرفته شده در [67]) بیشتر واقع‌گرایانه می‌باشد. داده‌های راهنما در فریم‌ها بوسیله‌ی دانشگاه بوستون با استفاده از *SignStreamTM* ایجاد شده است و شامل جنبه‌های زیادی از زبان اشارات می‌باشد [68].

مدل قبلی این پایگاه داده که "16-BOSTON201" نامیده می‌شود در [69] و در دسترس عموم قرار دارد^۱. پایگاه داده‌ی RWTH-Boston-104 شامل ۲۰۱ فیلم حاشیه‌نویسی شده از جملات ASL می‌باشد. این جملات توسط سه نفر (یک مرد و دو زن) اجرا و فیلم‌ها به طور همزمان توسط چهار دوربین گرفته شده‌اند، دو تا از دوربین‌ها به صورت جفت در جلو قرار گرفته و نمای روبروئی فرد اشاره‌کننده را نشان می‌دهد. یک دوربین دیگر در قسمت کناری شخص قرار گرفته و دوربین چهارم تنها تصویر چهره را ضبط می‌کند. فیلم‌ها با سرعت ۳۰ فریم در ثانیه و با وضوح 312×242 پیکسل گرفته می‌شوند. فیلم‌های تهیه شده بوسیله‌ی همه‌ی دوربین‌ها بجز دوربین مربوط به صورت از نوع سیاه و سفید^۳ هستند. نمونه‌هایی از تصویر فیلم در شکل (۴-۱) نشان داده شده است.

¹ <http://www-i6.informatik.rwth-aachen.de/~zahedi/databaseBOSTON201.html>

² Resolution

³ Gray-Scale



مبارت پایگاه

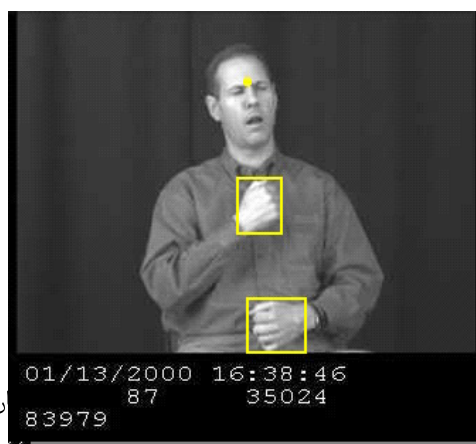
RWTH-Boston-50 ۲- ۴

این پایگاه داده [70,71] شامل ۴۸۳ اشاره از ۵۰ کلمه به صورت کلمه‌های جدا شده از جملات زبان اشاره است به صورتی که، کلمه‌های مشابه تکرار شده در جملات مختلف، جدا شده و در یک پایگاه داده جدید دسته‌بندی شده‌اند. در ساخت این مجموعه از جملات زبان اشاره موجود در پایگاه داده اصلی استفاده شده و برای هر علامت یا اشاره چند نمونه که توسط یک، دو و یا هر سه شخص اشاره-گر اجرا شده است، وجود دارد.

۳- ۴ پایگاه داده نهائی

در این پروژه از یک مجموعه‌ی داده ایجاد شده با کمک پایگاه داده‌ی **RWTH-Boston-50** برای ارزیابی سیستم تشخیص زبان اشاره بر اساس مشاهده استفاده شده است. در این پایگاه ابتدا ۴۷ لغت خاص و متفاوت به صورت تصادفی انتخاب و سپس چند نمونه از هر لغت (در مجموع ۱۳۸ نمونه) بعنوان یک کلاس مخصوص آن لغت برای آموزش و آزمایش الگوریتم تشخیص ارائه شده در این پروژه به کار گرفته شده‌اند.

هر نمونه از لغت بسته به طول زمان اجرای آن دارای تعدادی فریم می‌باشد و بسته به نوع کلمه؛ اینکه دست چپ در ایفای کلمه نقش دارد یا نه؛ باید از هر فریم مکان دست راست یا هر دو دست مشخص گردد. برای افزایش دقت در نتیجه‌ی آزمایش ابتدا مختصات مستطیل محاط بر دست (یا دست‌ها) و مکان سر استخراج و در فایل *hc* با فرمت *mat* ذخیره شد. در شکل (۲-۴) نمونه‌هایی از تصاویر فریم‌ها که در آنها مکان سر و مختصات مستطیل محاط بر دست به صورت دستی مشخص گشته‌اند نشان داده شده است.



ن سرو م

سپس تصاویر دست با کمک ناحیه‌ی داخل مستطیل محاط (از مچ دست تا انگشتان) از تصاویر اصلی جدا و در پایگاه داده‌ی جدید قرار داده شده‌اند. همچنین، سایر ویژگی‌ها از جمله مکان دست نسبت به سر، شیب وتر مستطیل محاط، مساحت آن، تصاویر باینری بدست آمده توسط یک حد آستانه‌ی مناسب، تصاویر لبه‌ی Canny با یک حد آستانه و زیگمای (σ) مناسب و در نهایت ویژگی‌های پیشنهادی راستای ماکزیموم لبه و طول آن در هر بخش تصویر قسمت بندی شده از همه‌ی فریم‌های هر نمونه گرفته و به پایگاه داده اضافه شدند. به این ترتیب یک پایگاه داده شامل کلمات مجزا از پایگاه قبلی ساخته می‌شود تا توسط آن سیستم‌هایی را که برای تشخیص جملات اشاره به صورت پیوسته با زمان طراحی شده‌اند؛ بطوری که سیستم با پیدا کردن کلمات جمله مورد نظر، اشاره‌ی اجرا شده را کلاسه‌بندی نماید؛ بتوان ارزیابی کرد.

۴-۴ مشخصه‌های تصاویر پایگاه داده

۴-۴-۱ الگوی حرکت دست نسبت به سر

در اشارات با یک دست که مکان دست راست جای مشخصی از سر می‌باشد شناسائی اشاره از روی منحنی مسیر حرکت دست آسان می‌شود. زمانی که دست به نقطه خاصی از سر اشاره کند، از آنجا که مکان دست نسبت به مکان سر سنجیده می‌شود، منحنی حرکت دست ویژگی بسیار خوبی برای شناسائی اشاره خواهد بود بطور مثال در اشاراتی مانند فکر کردن، دانستن و یا اشاره به کلمات ذرت، زن و...، همان طور که در شکل‌های (۳-۴) و (۴-۴) مشاهده می‌شود، دست راست در مکانی مشخص از سر قرار می‌گیرد و یا از ناحیه‌ی معینی از سر عبور می‌کند. بنابراین، مکان و منحنی حرکت مخصوص به خود را دارد.



شکل (۳-۴): مثال‌هایی که در آن دست راست با شکل خاص و در مکان مشخصی نسبت به سر قرار می‌گیرد: الف) دانستن، ب) فکر کردن



شکل (۳-۱): مثال‌هایی که در آنها دست راست با شکل و الگوی حرکت خاص از ناحیه مشخصی نسبت به سر عبور می‌کند: الف) زن، ب) ذرت

۴-۴-۲ الگوی حرکت دست و شکل دست

در اشاراتی که مکان دست جای کاملا مشخصی ندارد و می‌تواند در محدوده‌ی وسیعی اجرا شود منحنی حرکت دست ویژگی مناسبی برای شناسائی اشاره‌ی ادا شده نخواهد بود. در این حالت باید از

ویژگی‌های دیگری مانند الگوی حرکت دست، شکل دست، حالت قرارگیری انگشتان، تصویر لبه و ... استفاده نمود. در اشاراتی مانند شخص، رسیدن، گرفتن، خوب و ... با اینکه دست از نزدیکی سر عبور می‌کند اما از مکان مشخصی نسبت به سر نمی‌گذرد، در عوض دارای شکل خاصی از حالت دست و قرارگیری انگشتان می‌باشد. بطور مثال در اشاره‌ی کلمه‌ی شخص که در شکل (۴ ۵) نشان داده شده‌است همه‌ی انگشتان جمع و تنها انگشت اشاره به سمت بالا اشاره می‌کند و یا اشاره کلمه رسیدن که در آن دست باز و به سمت پایین حرکت می‌کند. بنابراین، ویژگی‌هایی مانند منحنی شیب وتر مستطیل محاط، منحنی مساحت این مستطیل، منحنی ماکزیموم جهت لبه و طول آن و ... که مربوط به شکل و حالت دست هستند انتخابی مناسب‌تر خواهد بود.



۴-۳ الگوی حرکت دو دست نسبت به هم

همچنین در اشاره کلماتی مانند کتاب، ماشین، خریدن، جعبه و... دست مکان و شکل کاملاً مشخصی ندارد. در اشارات از این نوع عموماً هر دو دست در ایفای اشاره نقش دارند و دارای الگوی حرکت خاصی نسبت به هم هستند. به عنوان نمونه در اشاراتی از جمله کتاب و خریدن که در شکل (

۶-۴) نشان داده شده‌اند هنگام ادای اشاره، دست‌ها مدت کوتاهی به هم می‌خورند. در این حالت می‌توان مسیر حرکت هر دو دست نسبت به یک مرجع یکسان؛ به عنوان مثال مکان سر مقایسه کرد.



شکل (۶-۴): اشاراتی که در آنها دو دست الگوی حرکت مشخصی نسبت به هم دارند: الف) کتاب، ب) خریدن

فصل پنجم

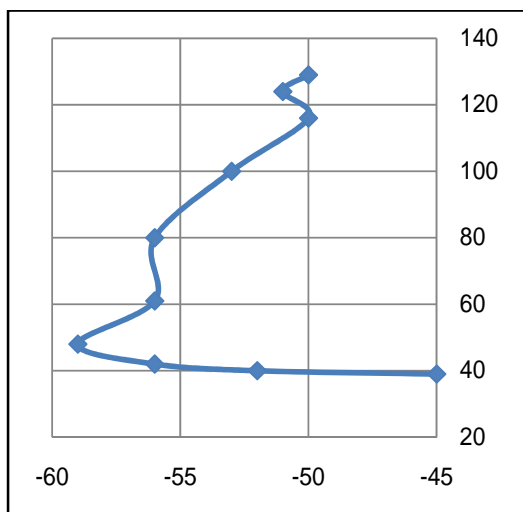
ویژگیهای بکار رفته

مقدمه

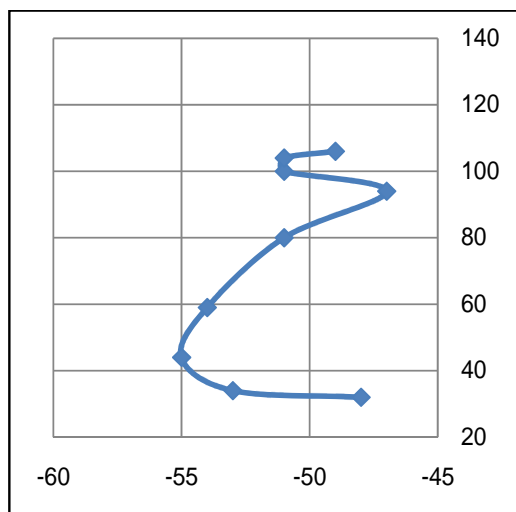
یک سیستم تشخیص زبان اشاره تنها بر اساس مشاهده نیازمند به کارگیری ویژگی‌هایی می‌باشد که تنها به کمک تصاویر دریافتی از یک دوربین فیلمبرداری به دست می‌آیند. همچنین در این پروژه سعی بر این بوده است تا از تجهیزات اضافی دیگر مانند دستکش‌های رنگی، تعداد دوربین‌های بیشتر، پروژکتورهای مخصوص و غیره استفاده نشود. برای این کار باید یک پردازش تصویر قدرتمند روی تصاویر شخص اشاره‌کننده صورت پذیرد تا ویژگی‌های مربوط به شکل، حالت و مسیر حرکت دست استخراج شوند. از طرف دیگر برای اینکه پاسخ سیستم تشخیص نسبت به زمان اجرا و فرد اشاره‌کننده مستقل باشد باید تا جایی که امکان دارد ویژگی‌های هندسی که نشان دهنده ساختار و مدل دست در هنگام اجرای اشاره هستند مدنظر قرار گیرند. در این فصل به معرفی ویژگی‌های به کار رفته در این تحقیق و بررسی دلایل استفاده از آنها پرداخته می‌شود.

۵-۱- منحنی مسیر حرکت دست نسبت به سر

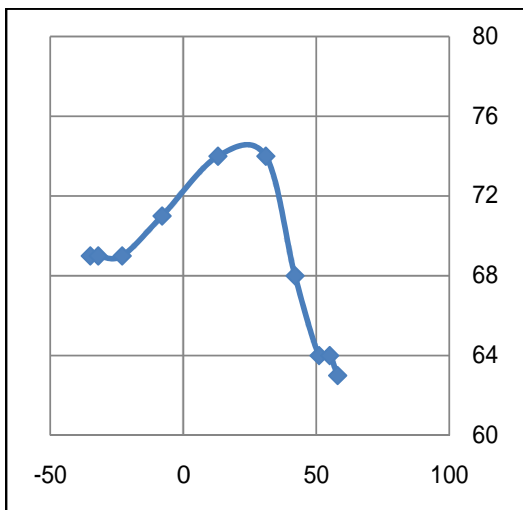
یکی از مهمترین ویژگی‌های مورد استفاده منحنی مسیر حرکت دست می‌باشد که در اکثر اشاره‌ها همراه با شکل و حالت دست، پیام اشاره‌ی مورد نظر را به مخاطب می‌رساند. در شکل (۱-۵) نمودار مسیر حرکت دست نسبت به سر از دو کلاس متفاوت و برای هر کلاس دو نمونه نشان داده شده است. همانطور که در این شکل دیده می‌شود مسیر حرکت دست برای اشاره‌های یکسان مشابه هم و برای اشاره‌های غیر یکسان متفاوت می‌باشد. این ویژگی یک ویژگی هندسی بوده و تقریباً مستقل از زمان اجرا و شخص ادا کننده از نظر سن و جنس می‌باشد اما از نظر اندازه و قد بدن باید نرمالیزه شود. به طور مثال به یک معیار هندسی از بدن شخص اشاره‌گر مانند اندازه سر و یا اندازه دست تقسیم شود.



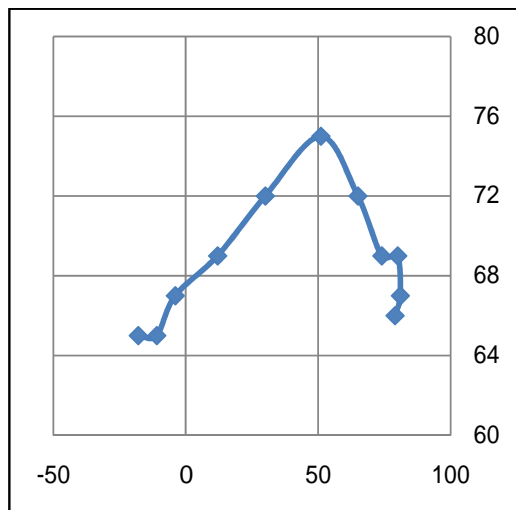
(ب)



(الف)



(د)



(ج)

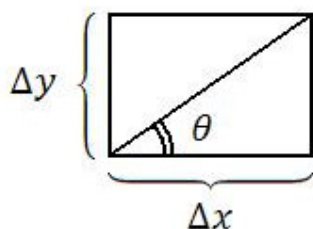
شکل (۱۰۵): نمودارهای منحنی مسیر حرکت دست راست نسبت به سر برای دو کلاس متفاوت: الف و ب مربوط به اشاره‌ی رسیدن، ج و د مربوط به اشاره‌ی جعبه

از آنجا که منحنی مسیر حرکت دست نسبت به سر محاسبه می‌شود استفاده از این ویژگی مشکل حرکت شخص نسبت به دوربین و جابجایی کم دوربین را تا اندازه‌ای حل می‌کند. ویژگی منحنی مسیر حرکت دست با استفاده از دنبال کردن دست و پیدا کردن مرکز تصویر دست صورت می‌گیرد.

۵-۲ شیب وتر ($Aspect\ Ratio^{-1}$) مستطیل محاط بر تصویر دست؛

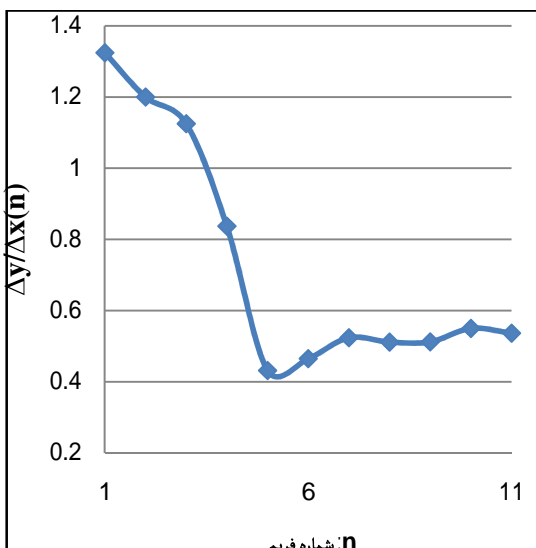
$$\frac{\Delta y}{\Delta x} (n)$$

در این تابع همانطور که در شکل (۵-۲) ملاحظه می‌شود Δy ارتفاع و Δx طول مستطیل محاط بر دست اشاره‌گر و n شماره‌ی فریم در کلاس مربوط می‌باشد. شیب وتر مستطیل محاط بر تصویر دست برای هر فریم کمیتی است که بیشتر به ساختار دست افراد بستگی دارد تا به اندازه‌ی آن زیرا برای محاسبه این ویژگی مقدار Δy تقسیم بر Δx می‌شود بنابراین نتیجه نسبت به اندازه دست اشاره-گر مستقل می‌شود و می‌توان انتظار داشت که مشکل اندازه‌های مختلف را تا اندازه‌ای حل می‌کند و از طرف دیگر رابطه‌ی چندانی با سن و رنگ افراد ندارد.

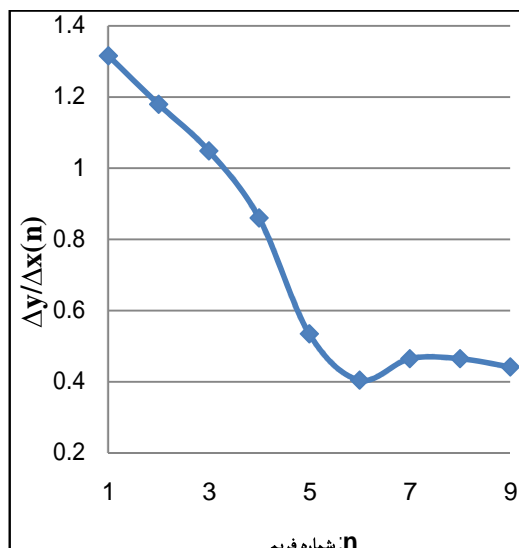


شکل (۵-۲): شیب قطر مستطیل محاط بر تصویر دست

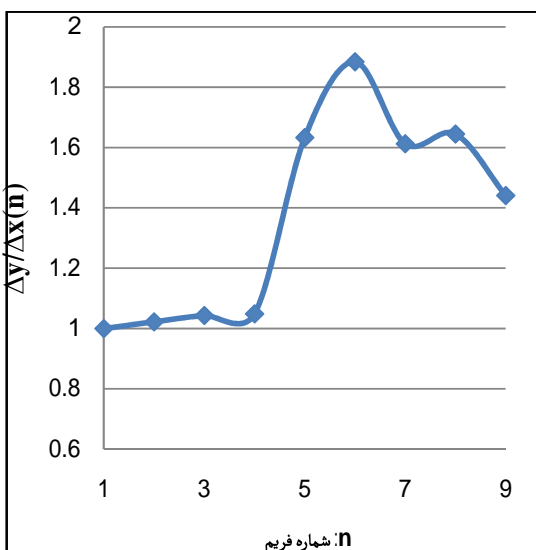
در شکل (۵-۳) مثال‌هایی از نمودار ویژگی منحنی شیب وتر مستطیل محاط بر تصویر دست برای دو کلاس متفاوت نشان داده شده است. همانطوری که در این نمودارها مشاهده می‌شود الگوی تغییرات و یا مقدار متوسط محدوده‌ی این ویژگی به خوبی دسته‌ای از کلاس‌ها را نسبت به کلاس‌های دیگر متمایز می‌نماید. به عنوان نمونه نمودارهای مربوط به کلاس اول به طور متوسط سیر نزولی دارند در حالیکه نمودارهای مربوط به کلاس دوم به طور متوسط دارای سیر صعودی هستند. همچنین مقدار متوسط شیب وتر در کلاس اول کمتر از یک و در کلاس دوم بیشتر از یک می‌باشد.



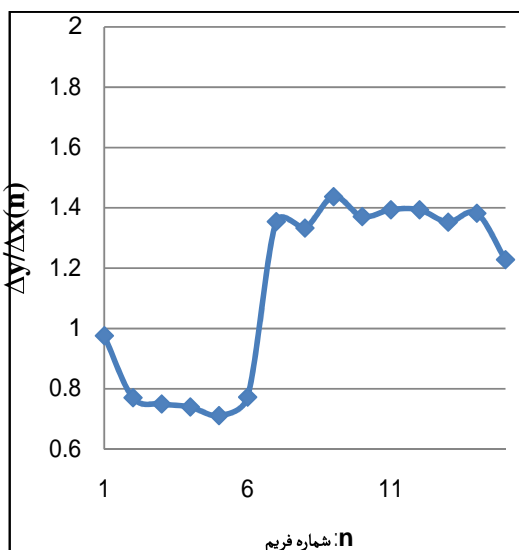
(ب)



(الف)



(د)



(ج)

شکل (۳ ۵): نمودارهای ویژگی منحنی شیب وتر مستطیل محاط بر تصویر دست راست برای دو کلاس متفاوت: الف و ب برای اشاره‌ی فعل رسیدن، ج و د برای اشاره‌ی کلمه‌ی گروه

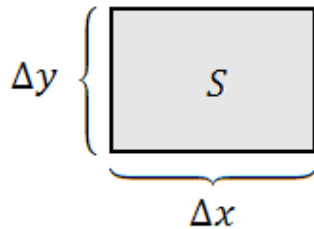
۵-۳ مساحت مستطیل محاط بر تصویر دست؛ $\Delta y \cdot \Delta x(n)$

این ویژگی به طور عمده نشان دهنده‌ی باز و یا جمع بودن دست می‌باشد به عبارت دیگر در دو تصویر دست اگر مقدار شیب وتر مستطیل محاط یکسان باشد مقدار مساحت آن تعیین کننده خواهد

بود و چنانچه دست به صورت کاملا باز رو به دوربین قرار گیرد اندازه‌ی آن به بیشترین مقدار می‌رسد.

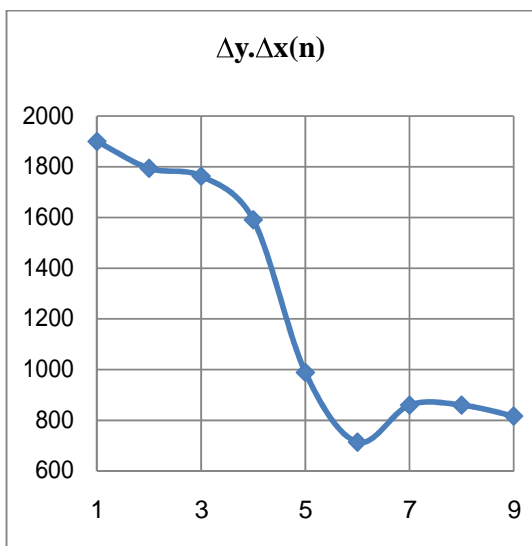
برخلاف تابع $\frac{\Delta y}{\Delta x}(n)$ ، تابع $\Delta y \cdot \Delta x(n)$ بر حسب n (شماره‌ی فریم) همانطور که در شکل (۴-۵) قابل

مشاهده می‌باشد به اندازه‌ی دست شخص اشاره‌گر نیز بستگی دارد.

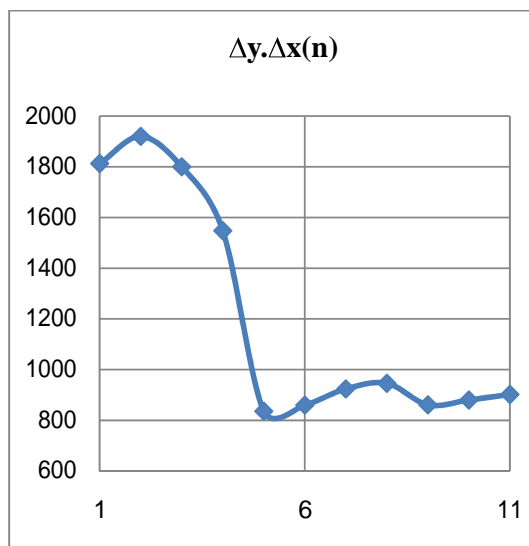


شکل (۴-۵): مساحت مستطیل محاط بر تصویر دست

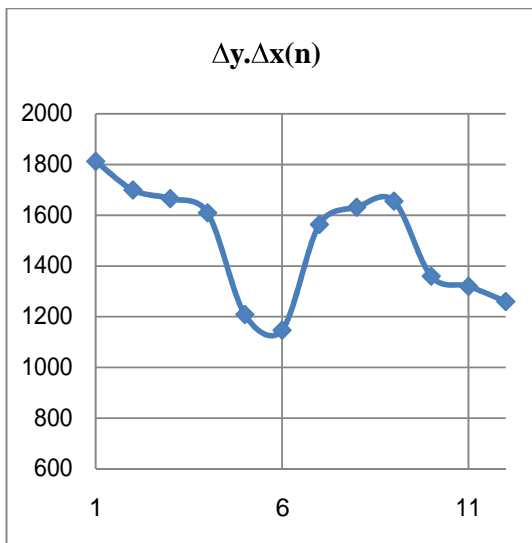
به این ترتیب که دست بزرگتر دارای مشتی بزرگتر و همچنین انگشتانی بلندتر و کشیده‌تر است و حتی تابع تغییرات $(\Delta y \cdot \Delta x)$ نیز دارای دامنه‌ی تغییرات بیشتری برای دست بزرگتر می‌باشد. بنابراین، برای استفاده از این ویژگی به صورتی که مستقل از سن افراد (و اندازه‌ی آنها) باشد باید تابع را نرمالیزه کرد. نکته‌ی مثبت در مورد منحنی $(\Delta y \cdot \Delta x)$ این است که به هر حال به ساختار دست نیز بستگی دارد و در اعمال اشاره، اطلاعات با ارزشی به ما می‌دهد. در شکل (۵-۵) چند نمونه از نمودار از نمودار ویژگی مساحت مستطیل محاط بر دست برای دو اشاره‌گر غیر یکسان آورده شده است. همانگونه که در این شکل ملاحظه می‌شود نمودارهای یک کلاس به طور مشخص نسبت به هم مشابه و نسبت به نمودارهای کلاس دیگر متفاوت می‌باشند بنابراین با کمک این ویژگی نیز می‌توان گروهی از اشارات مشابه هم را از سایر کلاس‌ها جدا نمود.



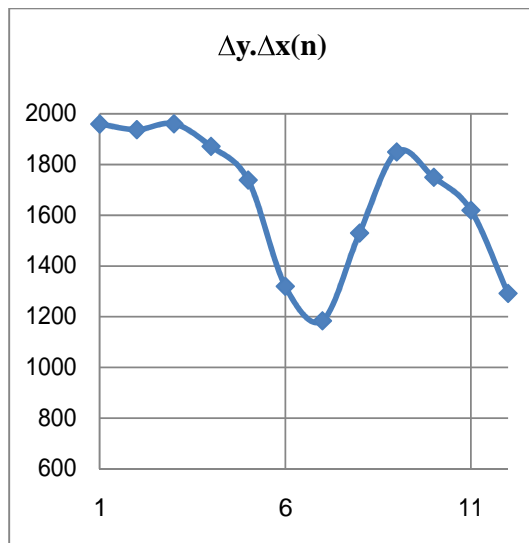
(ب)



(الف)



(د)



(ج)

شکل (۵-۵): نمودارهای منحنی مساحت مستطیل محاط بر تصویر دست برای دو کلاس مختلف: الف و ب برای اشاره‌ی فعل رسیدن، ج و د برای اشاره‌ی کلمه‌ی نبات

در صورت استفاده‌ی همزمان از $\frac{\Delta y}{\Delta x}$ و $\Delta y \cdot \Delta x$ برای کلاسه بندی می‌توان از یکی از دو عبارت، Δx

و یا Δy را حذف کرد زیرا عبارت مورد نظر از مقدار Δx یا Δy و عبارت اول بدست می‌آید.

۵-۴ استخراج تصویر لبه با استفاده از روش Canny

شکل دست و نحوه‌ی قرارگیری انگشتان در هر اشاره مانند مسیر حرکت دست در ایفای اشاره سهیم است. برای تشخیص بهتر زبان اشاره لازم است این ویژگی استخراج و از آن برای کلاسه‌بندی استفاده شود. حالات و نحوه‌ی قرارگیری انگشتان در هر اشاره و علامت متفاوت و متغیر با زمان است. به همین خاطر براحتی نمی‌توان به دنبال شکل خاصی برای تطبیق الگوی شکل دست در هر اشاره بود. برای حل این مشکل در این پروژه از ویژگی بافت شکل دست استفاده می‌شود؛ به این ترتیب که تصویر لبه از هر فریم به صورت بافتی خاص در نظر گرفته می‌شود، بطوری که هر علامت از نظر اندازه و جهت، نوع لبه‌ی مخصوص به خود را دارد.

برای بدست آوردن تصویر باینری لبه‌ی دست، از الگوریتم آشکار سازی لبه Canny؛ با دو حد آستانه و پنجره‌ی همسایگی با ابعاد مناسب استفاده می‌شود تا مرز میان انگشتان و دست به خوبی مشخص گردد. از آنجایی که تصویر لبه‌ی دست یک ویژگی هندسی از تصویر دست اشاره‌گر می‌باشد، تا حد زیادی نسبت به تغییرات نور، شخص اجرا کننده و زمان اجرا مقاوم است. همان گونه که در بخش آشکار ساز Canny گفته شد این عملگر با استفاده از دو حد آستانه به خوبی لبه‌های مناسب را پیدا و لبه‌های نویزی را حذف می‌نماید و تصویر باینری حاصل به گونه‌ای است که در هر ناحیه‌ی کوچک از تصویر، در هر راستا تنها یک لبه وجود دارد به عبارت دیگر در یک ناحیه کوچک دو لبه به موازات هم که به یکدیگر چسبیده باشند نداریم. از این خصوصیت برای بدست آوردن ویژگی راستای ماکزیمم لبه (MED^1) استفاده می‌کنیم.

¹ Maximum Edge Direction

۵-۵ راستای ماکزیمم لبه (MED) و طول آن در کل تصویر دست

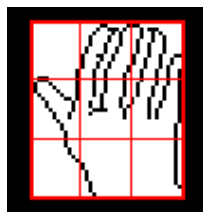
راستای ماکزیمم لبه یک ویژگی پیشنهادی می‌باشد که با دقت خوبی سمت جهت گیری کلی انگشتان را نشان می‌دهد در حالی که ممکن است راستای بیشترین توزیع لبه در جهت دیگری باشد، همچنین ویژگی طول راستای ماکزیمم لبه به میزان ارزش راستای بدست آمده اشاره دارد و در صورتی که مقدار آن نزدیک به صفر باشد به این معنی است که لبه دارای جهت خاصی نمی‌باشد. صورت محلی این ویژگی در بخش ۵-۶ به طور کامل‌تر بررسی می‌گردد.

۵-۶ راستای ماکزیمم لبه و طول آن به صورت محلی از تصویر (LMED)^۱

برای این که بتوان شکل و نحوه‌ی قرار گیری انگشتان را نسبت به کف دست مدل سازی کرد تصویر دست را به اندازه‌های مناسب تقسیم‌بندی و در هر ناحیه ویژگی راستای ماکزیمم لبه و طول آن محاسبه می‌گردد. چنان ابعاد پنجره‌ای که در آن ویژگی‌های مربوط به راستای ماکزیمم لبه باید استخراج شوند، خیلی بزرگ باشند مقدار لبه‌ها در هر ناحیه زیاد شده و در بیشتر موارد راستای بدست آمده معنی‌دار نخواهد بود. به همین ترتیب اگر ابعاد پنجره خیلی کوچک انتخاب شود در اثر جابجایی و انحنای کم دست و انگشتان اطلاعات مربوط به هر پنجره (راستای ماکزیمم لبه و طول آن) برای یک کلاس مشابه به شدت تغییر می‌کند. در این پروژه با در نظر گرفتن شرایط ذکر شده تصویر دست به نه قسمت تقسیم می‌شود به عبارت دیگر اندازه طول و عرض پنجره مورد نظر برابر با یک سوم طول و عرض تصویر دست انتخاب می‌گردد. هر ناحیه از یک تا نه شماره گذاری می‌شود. به این ترتیب اگر در هر اشاره یک یا چند انگشت جهت و یا حالت خاصی بگیرند با احتمال خوبی ویژگی

¹ Local Maximum Edge Direction

راستای ماکزیمم لبه در نواحی مشخص شده الگوی حالت دست را نشان می‌دهد. از آنجا که برای بدست آوردن این ویژگی از آشکار ساز لبه Canny استفاده شده و همچنین تصاویر دست به صورت مشابه تقسیم بندی شده‌اند نتیجه نهایی یک ویژگی هندسی خواهد بود که به ساختار، مدل، نحوه‌ی شکل گیری و جهت گیری دست و انگشتان اشاره دارد. به این ترتیب ویژگی پیشنهادی تا حد زیادی مستقل از شخص اشاره کننده و زمان اجرای اشاره می‌باشد. در شکل (۵-۶) دو نمونه از تصویر دست که دارای الگوی متفاوتی هستند نشان داده شده است. همانگونه که مشاهده می‌شود دو تصویر از نظر ویژگی راستای ماکزیمم لبه و طول آن به صورت محلی، میان نواحی متناظر با هم، کاملاً از هم متمایز می‌شوند.



(ب)



(الف)



(د)



(ج)

شکل (۵-۶): تصاویر لبه از شکل دست در یک فریم از دو اشاره‌ی نمونه و قسمت‌بندی آنها برای محاسبه ویژگی LMED: (الف) تصویر لبه از شکل دست در اشاره فعل رفتن، (ب) تصویر لبه از شکل دست در اشاره فعل رسیدن، (ج) نواحی جدا شده از تصویر الف، (د) نواحی جدا شده از تصویر ب

فصل ششم

الگوریتم کلاسه بندی

مقدمه

روش پیشنهادی برای تشخیص زبان اشاره شامل دو مرحله می‌باشد: ۱- کلاسه‌بندی اولیه برای حذف بیشتر کلاسهای نامطلوب ۰۲- کلاسه‌بندی نهایی با استفاده از ویژگی LMED. در ادامه به بیان و تفصیل هریک از این مراحل می‌پردازیم.

۶-۱- کلاسه‌بندی اولیه

در مرحله اول ویژگی‌های مسیر حرکت دست، منحنی زاویه‌ی وتر تصویر دست و منحنی مساحت تصویر دست استخراج شده و با کمک الگوریتم DTW و کلاسه‌بندی نزدیکترین همسایگی^۱؛ همانگونه که در شکل (۱-۶) نشان داده شده است، بیشتر کلاسهای نامطلوب حذف می‌شود. در این مرحله هدف اصلی کاهش تعداد کلاس‌ها به منظور افزایش سرعت و دقت کار در مرحله‌ی دوم می‌باشد. به این ترتیب برای انتخاب کلاس‌هایی که با احتمال بیشتری ممکن است کلاس اشاره مورد آزمایش باشند از روش‌هایی استفاده می‌شود که دارای سرعت نسبتاً بالایی در پردازش همه نمونه‌های موجود در پایگاه داده بوده و در کلاسه‌بندی با رنک مناسب کلاس اصلی را از دست ندهند. روند کار به چهار بخش تقسیم می‌گردد:

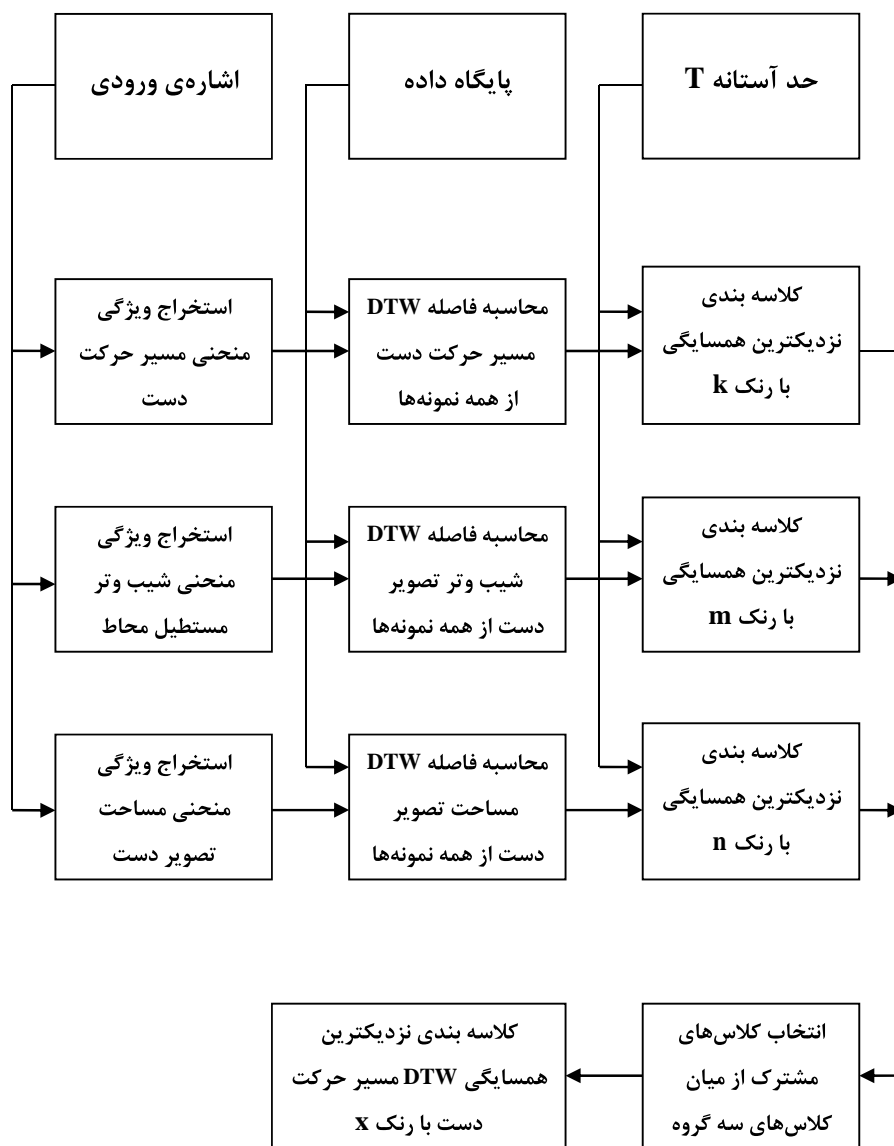
۰۱- کلاسه‌بندی با استفاده از ویژگی منحنی مساحت تصویر دست

۰۲- کلاسه‌بندی با استفاده از ویژگی منحنی زاویه‌ی وتر تصویر دست

۰۳- کلاسه‌بندی با استفاده از ویژگی منحنی مسیر حرکت دست

۰۴- انتخاب کلاس‌های مطلوب با استفاده از نتایج بخش‌های ۱، ۲ و ۳.

¹ Nearest neighbor



شکل (۱-۶): الگوریتم کلاسه‌بندی در مرحله‌ی اول

در ادامه به بررسی نحوه‌ی عملکرد هر یک از این بخش‌ها پرداخته می‌شود.

۶-۱-۱- کلاسه‌بندی با استفاده از ویژگی منحنی مساحت تصویر دست

در این بخش ابتدا با کمک گرفتن از الگوریتم DTW فاصله منحنی مساحت تصویر دست مربوط به اشاره ورودی از همه نمونه‌های موجود در پایگاه داده محاسبه می‌شود. سپس با در نظر گرفتن یک حد آستانه مناسب نزدیکترین نمونه‌ها به اشاره ورودی انتخاب می‌گردد. در نهایت کلاس‌های نمونه‌های انتخاب شده تعیین و با توجه به میانگین فاصله نمونه‌های هر کلاس از اشاره ورودی رده‌بندی می‌شوند. این روش کلاسه‌بندی بخاطر استفاده از ویژگی منحنی مساحت تصویر دست دقت نسبتاً پایینی دارد اما دارای چند مزیت می‌باشد. اول آنکه اگر تعداد کلاس‌های نزدیک به یک اشاره زیاد باشد کلاس‌های بیشتری انتخاب می‌شوند در نتیجه کلاس اصلی از دست نمی‌رود. دوم آنکه دقت کلاسه‌بندی با رنگ ۱۵ برای اشارات دو دستی و کلاسه‌بندی با رنگ ۲۰ برای اشارات یک دستی به صد در صد می‌رسد. سوم آنکه دارای پیچیدگی محاسبات کم و در نتیجه سرعت بسیار بالایی می‌باشد. این مزایا موجب می‌شود تا از ویژگی منحنی مساحت تصویر دست در مرحله اول استفاده شود.

۶-۱-۲- کلاسه‌بندی با استفاده از ویژگی منحنی زاویه‌ی وتر تصویر دست

در این قسمت نیز ابتدا فاصله DTW منحنی زاویه‌ی وتر تصویر دست در فریم‌های اشاره میان نمونه ورودی و همه نمونه‌های پایگاه داده بدست می‌آید. سپس با در نظر گرفتن یک حد آستانه مناسب، نمونه‌های نزدیک به اشاره ورودی مشخص می‌شوند. کلاس نمونه‌های انتخاب شده، خروجی این قسمت خواهد بود. در واقع کلاس‌هایی انتخاب می‌شوند که حداقل یک نمونه از آنها در فاصله‌ای کمتر از مقدار تعیین شده نسبت به اشاره ورودی قرار داشته باشد. کلاسه‌بندی در این بخش دقت نسبتاً خوبی دارد به طوری که با رنگ ۵ برای اشارات دو دستی و رنگ ۲۰ برای اشارات یک دستی به صد در صد می‌رسد. علاوه بر آن به دلیل داشتن پیچیدگی محاسبات کم و سرعت عمل بالا در مرحله‌ی اول به کار گرفته می‌شود.

۶-۱-۳ کلاسه‌بندی با استفاده از ویژگی منحنی مسیر حرکت دست

کلاسه‌بندی در این بخش مشابه بخش‌های اول و دوم می‌باشد با این تفاوت که در این قسمت ویژگی منحنی دو بعدی مسیر حرکت دست نسبت به سر برای حذف کلاس‌های نامطلوب بکار گرفته می‌شود. برای این که بتوان از الگوریتم یک بعدی DTW در محاسبه فاصله منحنی‌های دوبعدی اشاره ورودی و نمونه‌های موجود در پایگاه داده استفاده کرد ابتدا این منحنی تبدیل به دو منحنی یک بعدی می‌شود؛ یکی مسیر حرکت دست نسبت به سر در راستای افق (در جهت x) برحسب زمان (شماره فریم) و دیگری مسیر حرکت دست نسبت به سر در راستای عمود (در جهت y) نسبت به زمان. سپس فاصله‌ی DTW میان منحنی‌های x و y اشاره مورد نظر از همه نمونه‌های پایگاه داده محاسبه شده و با هم جمع می‌شوند. برای انتخاب نمونه‌هایی که ممکن است جزو کلاس اشاره‌ی ورودی باشند یک حد آستانه مناسب در نظر گرفته شده و نمونه‌هایی که حاصل جمع فواصل DTW منحنی‌های x و y آنها نسبت به اشاره مورد نظر کمتر از مقدار تعیین شده باشد گرفته و بقیه حذف می‌شوند. در نهایت کلاس‌های نمونه‌های انتخاب شده به خروجی ارسال می‌شوند. اختلاف این روش با روش استفاده از الگوریتم DTW دو بعدی در نحوه‌ی تناظر فریم‌های مشابه به هم می‌باشد به این صورت که در روش اول مسیر حرکت دست در راستای x و y به طور جداگانه مورد بررسی قرار می‌گیرد و فریم‌های متناظر شده به هم در راستای x ممکن است همان فریم‌های متناظر شده در راستای y نباشند اما در روش دوم فریم‌ها تنها به یک صورت با هم متناظر می‌شوند. از آنجا که در بسیاری از اشارات یکسان نحوه‌ی انحنای زمانی مسیر حرکت دست در یک راستا با راستای دیگر تفاوت دارد برای سنجش نمونه‌ها از الگوریتم یک بعدی DTW استفاده شده است. برای مقایسه نتایج این قسمت با بخش‌های ۱ و ۲، میانگین فواصل نمونه‌های کلاس‌های تعیین شده محاسبه و کلاس‌ها بر اساس میزان فاصله بدست آمده طبقه‌بندی می‌شوند (هرچند ممکن است این نمونه‌ها در گروه نمونه‌های انتخابی اولیه نباشند). در این روش کلاسه‌بندی میزان دقت با رنگ ۴ برای اشارات دو دستی و رنگ ۵ برای اشارات یک دستی به صد در صد می‌رسد. علاوه بر آن این روش سرعت اجرای بسیار بالایی دارد.

این مزایا موجب می‌شود تا از ویژگی منحنی مسیر حرکت دست، هم در مرحله‌ی اول و هم در مرحله‌ی دوم استفاده شود.

۶-۱-۴ انتخاب کلاس‌های مطلوب با استفاده از نتایج بخش‌های قبلی

همانطور که ملاحظه گردید در هر یک از بخش‌های قبلی تعدادی از کلاس‌ها به عنوان کلاس اشاره ورودی انتخاب شده‌اند به نحوی که هیچکدام از آنها کلاس درست را از دست نداده‌اند. در این قسمت با به اشتراک گذاشتن نتایج بخش‌های قبل بیشتر از کلاس‌های نامطلوب حذف می‌شوند و تنها تعداد کمی از کلاس‌ها به عنوان پاسخ‌های ممکن باقی می‌ماند. در ادامه اگر تنها یک کلاس بین هر سه مجموعه مشترک باشد همان کلاس به عنوان نتیجه نهایی انتخاب می‌شود و اگر تعداد کلاس‌های مشترک در اشارات دو دستی بیشتر از دو و در اشارات یک دستی بیشتر از سه شود توسط کلاسه-بندی بخش سوم؛ یعنی حذف کلاس‌ها با استفاده از روش نزدیکترین همسایگی به کمک فاصله‌ی DTW منحنی مسیر حرکت دست، بقیه کلاس‌ها که فاصله بیشتری دارند حذف می‌شوند. در نهایت برای اشارات دو دستی حداکثر دو کلاس و برای اشارات یک دستی حداکثر سه کلاس برای مرحله‌ی بعدی انتخاب می‌شوند به نحوی که کلاس اصلی در آنها حضور دارد و از دست نرفته است. به این ترتیب نه تنها میزان محاسبات در مرحله بعد کاهش چشمگیری خواهد یافت بلکه بخاطر حذف بیشتر کلاس‌های نامطلوب، دقت آن افزایش پیدا خواهد کرد.

نکته‌ی مورد توجه در کلاس‌بندی این است که در استفاده از فاصله بین منحنی‌های حرکت، فاصله بین منحنی شیب وتر مستطیل محاط بر دست و منحنی مساحت آن برای هر ویژگی به طور مستقل عمل شده است. به این معنی که فریم‌های متناظر شده با هم در منحنی حرکت دست ممکن است همان فریم‌های متناظر شده در منحنی شیب و یا منحنی مساحت مستطیل محاط بر دست نباشند. مزیت این کار در آن است که اگر در یک کلاس آزادی زیادی نسبت به نوع حرکت دست و مسیر آن

باشد این اختلاف توسط ویژگی‌های دیگر جبران شود. عیب این عملکرد زمانی مشخص می‌شود که دو کلاس قابل مقایسه از نظر منحنی مسیر حرکت، منحنی شیب وتر مستطیل محاط و منحنی مساحت آن تا حدودی مشابه هم باشند؛ به‌گونه‌ای که کلاسه‌بندی از طریق به دست آوردن فاصله توسط الگوریتم انحنای زمانی به صورت پویا دچار خطا شود.

با تناظر فریم‌های یک نمونه به نمونه دیگر تنها با کمک یکی از ویژگی‌ها؛ به طور مثال منحنی حرکت دست و استفاده از این تناظر برای به دست آوردن فاصله مربوط به سایر ویژگی‌ها می‌توان این مشکل را در مورد کلاس‌هایی که دارای نمونه‌های دقیق هستند و تنها فرق جزئی دارند تا حدودی حل کرد اما با این کار الگوریتم کلاسه‌بندی ممکن است نتواند کلاس‌هایی که دارای نمونه‌هایی با آزادی عملکرد بالا هستند را به خوبی شناسایی نماید.

۶ ۴ کلاسه‌بندی نهایی

در مرحله دوم به بررسی و سنجش دقیق‌تر کلاس‌های انتخاب شده در مرحله اول پرداخته می‌شود؛ به نحوی که از میان این کلاس‌ها، کلاس اصلی به دست آید. این کار در حالی باید انجام پذیرد که این کلاس‌ها از نظر ویژگی‌های به کار رفته در مرحله قبلی تقریباً مشابه هم هستند، بنابراین در این مرحله از ویژگی‌هایی که دقت بالایی ندارند و در مرحله قبل به کار گرفته شده‌اند استفاده نمی‌شود و از میان آن‌ها تنها از ویژگی منحنی مسیر حرکت دست نسبت به سر که از دقت نهایی خوبی برخوردار است کمک گرفته می‌شود.

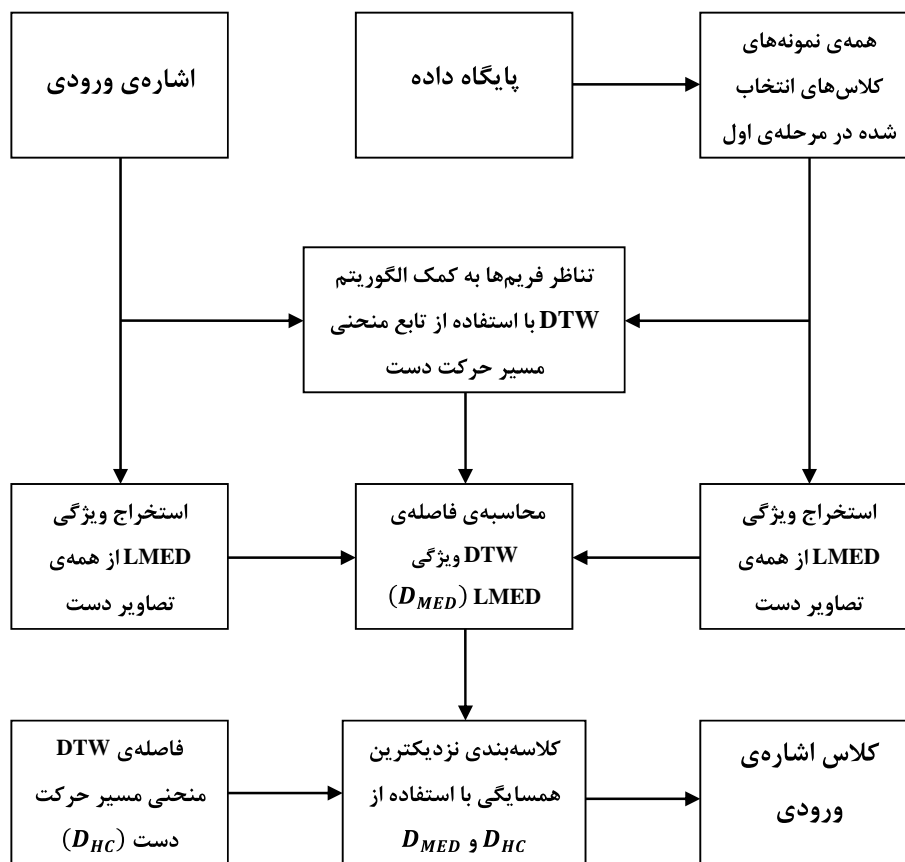
ویژگی مهم دیگری که در این مرحله از آن استفاده می‌شود، ویژگی LMED (بیشترین راستای لبه به صورت محلی) می‌باشد. این ویژگی همانطور که در فصل پنجم به تفصیل شرح داده شده است، حاوی اطلاعات با ارزشی از نحوه‌ی قرارگیری و جهت خم شدن انگشتان دست بوده و بر اساس آزمایش‌های انجام شده اشاراتی را که در آنها حالت شکل دست و نحوه‌ی قرارگیری انگشتان متفاوت

است را با دقت خوبی از هم تمیز می‌دهد. الگوریتم کلی استفاده از ویژگی LMED و نتایج حاصل از مرحله‌ی اول در شکل (۲-۶) نشان داده شده است. در این مرحله سه روش برای کلاسه‌بندی زبان اشاره پیشنهاد شده که عبارت است از:

۱ - روش نزدیکترین همسایگی با استفاده از ویژگی LMED و الگوریتم DTW

۲ - روش نزدیکترین همسایگی هندسی به صورت تطبیقی^۱

۳ - روش نزدیکترین همسایگی اقلیدسی به صورت تطبیقی^۲



شکل (۲-۶): الگوریتم کلی کلاسه‌بندی نهایی

¹ Adaptive Geometric Distance

² Adaptive Euclidean Distance

۶-۲-۱ روش نزدیکترین همسایگی با استفاده از ویژگی LMED و الگوریتم

DTW

در این روش با استفاده از ویژگی LMED کلاس‌های باقی‌مانده از مرحله اول طبقه‌بندی می‌شوند، به طوری که فریم‌های متناظر شده در مرحله اول توسط الگوریتم DTW با کمک تابع منحنی مسیر حرکت دست از نظر اختلاف و یا تشابه ویژگی LMED با هم مقایسه می‌شوند. این کار موجب می‌شود تا تنها تصاویری که از نظر زمان و مکان، موقعیت مشابهی نسبت به هم دارند و با احتمال خوبی دارای شکل و حالت مشابهی هستند مورد قیاس قرار گیرند که در نتیجه دقت و سرعت مقایسه افزایش چشمگیری پیدا می‌کند. حال با کمک ویژگی LMED فاصله اشاره ورودی از تمام نمونه‌های کلاس-های تعیین شده در مرحله اول محاسبه می‌شود. به طوری که در فصل پنجم نیز بدان اشاره شد، هر فریم حاوی تصویر دست به ۹ تصویر با ابعاد مساوی افراز می‌شود. در هر تصویر ویژگی بیشترین راستای لبه (θ_n) به همراه طول آن (l_n) استخراج و یک تابع فاصله میان هر دو تصویر از ناحیه‌ی یکسان فریم‌های متناظر شده با هم به صورت زیر تعریف می‌گردد:

$$d_{LMED}(n, i, j) = \frac{(\Delta\theta_n + \theta_0) \times (\Delta l_n + l_0)}{l_{ni} + l_{nj} + L_0} \quad (۱-۶)$$

در این رابطه:

i و j : اندیس فریم‌های متناظر شده به هم از هر اشاره

n : شماره تصویر افراز شده از هر فریم

l_n : طول (ارزش) راستا تعیین شده در ناحیه n ام فریم

$\Delta\theta_n$: اختلاف راستا میان دو ناحیه n ام فریم‌های متناظر شده با هم

Δl_n : اختلاف طول بیشترین راستای لبه میان دو ناحیه n ام فریم‌های متناظر شده با هم

θ_0 ، l_0 و L_0 : ثابت‌های جبران‌ساز. می‌باشند

محدوده راستای لبه در بازه‌ی $0^\circ \leq \theta_n \leq 180^\circ$ قرار گرفته است در حالی که بازه‌ی تغییرات زاویه لبه برابر با 360° می‌باشد. این بدان معنا است که دو راستا ممکن است نزدیک به هم بوده اما از نظر میزان زاویه لبه نزدیک به 180° اختلاف داشته باشند. از این رو رابطه (۲-۶) برای به دست آوردن اختلاف راستا میان دو ناحیه یکسان فریم‌های متناظر شده با هم به کار رفته است. برای محاسبه Δl_n نیز از رابطه‌ی (۳-۶) استفاده می‌شود.

$$\Delta\theta_n = \min\{|\theta_{ni} - \theta_{nj}|, \pi - |\theta_{ni} - \theta_{nj}|\} \quad (۲-۶)$$

$$\Delta l_n = |l_{ni} - l_{nj}| \quad (۳-۶)$$

مقادیر θ_0 و l_0 ثابت‌هایی هستند که برای جلوگیری از صفر شدن و همچنین کاهش بیش از حد تابع فاصله در زمانی که دو ناحیه‌ی تصویر مورد مقایسه از نظر یکی از دو ویژگی θ_n و l_n تشابه زیاد و از نظر ویژگی دیگر با هم اختلاف قابل ملاحظه‌ای داشته باشند به کار می‌روند. از آنجا که هرچه ارزش l_{ni} و l_{nj} بیشتر باشد، ارزش تشابه نواحی n ام از فریم‌های i و j بیشتر خواهد بود، بنابراین تابع فاصله متناسب با معکوس مجموع ارزش‌های l_{ni} و l_{nj} قرار داده می‌شود. مقدار ثابت L_0 نیز زمانی که ارزش‌های l_{ni} و l_{nj} به سمت صفر میل کنند مانع افزایش بیش از اندازه‌ی تابع فاصله می‌گردد. برای بدست آوردن فاصله‌ی میان هر فریم با فریم متناظر آن از رابطه‌ی (۴-۶) و در نهایت برای محاسبه‌ی فاصله دو اشاره از رابطه‌ی (۵-۶) استفاده می‌شود که در آن I و J تعداد فریم‌های اشارات r و s می‌باشد. به این ترتیب مقدار فاصله‌ی دو اشاره‌ی مورد مقایسه، بر حسب مدت زمان اجرای آنها نرمالیزه می‌شود.

$$d_{MED}(i, j) = \sum_{n=1}^9 d_{LMED}(n) \quad (۴-۶)$$

$$D_{MED}(r, s) = \frac{1}{I+J} \sum_{(i,j) \in \Phi} d_{MED}(i, j) \quad (5-6)$$

این روش برای اجرای کلاسه‌بندی اشارات دارای سرعت و دقت بالایی بوده و نتایج آزمایش توانایی ویژگی LMED در مدل‌سازی شکل و حالت دست را به خوبی اثبات می‌کند.

۶-۲-۲ روش مینیمم میانگین هندسی به صورت تطبیقی

هر اشاره شامل دو نوع اطلاعات مربوط به دست می‌باشد؛ یکی شکل و حالت دست در حال ادای اشاره و دیگری مسیر و نحوه‌ی حرکت دست نسبت به بدن. به دلیل اینکه ویژگی LMED که مربوط به شکل و حالت دست می‌باشد به تنهایی نمی‌تواند معنی و مفهوم همه علائم زبان اشاره را تشخیص دهد، بنابراین در این روش از فاصله‌ی DTW منحنی مسیر حرکت دست نیز برای کلاسه‌بندی استفاده می‌شود. ویژگی منحنی مسیر حرکت دست در مرحله اول به عنوان تابع الگوریتم DTW برای متناظر کردن فریم‌های دو اشاره به هم و انتخاب کلاس‌های نزدیک‌تر به کار رفته است. در نتیجه، کلاس‌های انتخاب شده از این لحاظ نزدیک به هم هستند و چنانچه اشارات متفاوتی دارای مسیر حرکت دست مشابه هم باشند ممکن است استفاده از این ویژگی از دقت تشخیص نهائی بکاهد. اما همانطور که از پایگاه داده برداشت می‌شود، اشاراتی که دارای مسیرهای مشابهی با هم هستند عموماً از نظر شکل و فرم قرارگیری انگشتان با هم تفاوت زیادی دارند تا به بیننده پیام شاخصی را القا نمایند. بنابراین در این نوع از اشارات اختلاف ویژگی LMED تقریباً به اندازه‌ای است که تشابه ویژگی منحنی مسیر حرکت دست را جبران نماید. از طرف دیگر چنانچه دو اشاره با معنای متفاوت دارای ویژگی LMED نزدیک به هم باشند با احتمال خوبی از نظر ویژگی منحنی مسیر حرکت دست اختلاف قابل توجهی دارند. به این ترتیب در روش‌های ۲ و ۳ از هر دو ویژگی ذکر شده در کلاسه-

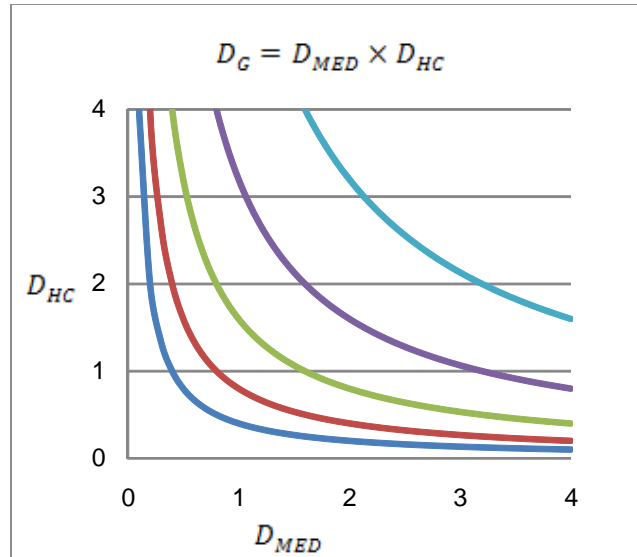
بندی نهائی بهره می‌جوئیم، در این روش تابع میانگین هندسی دو اشاره از هم به صورت زیر تعریف می‌شود:

$$D_G = \sqrt{D_{MED} \times D_{HC}} \quad (6-6)$$

که در آن D_{HC} فاصله DTW منحنی مسیر حرکت دست نسبت به سر و D_{MED} فاصله به دست آمده از ویژگی LMED میان دو اشاره می‌باشد. سپس فاصله اشاره ورودی از همه نمونه‌های کلاس‌های انتخاب شده در مرحله اول که در پایگاه داده ذخیره شده‌اند، محاسبه شده و کلاس نمونه با کمترین فاصله به عنوان کلاس اشاره ورودی در نظر گرفته می‌شود. از آنجائی که به طور مستقیم نمونه‌ای که کمترین فاصله را نسبت به اشاره مورد آزمایش دارد کلاس آن را تعیین می‌کند، می‌توان رابطه (6-6) را به صورت زیر ساده نمود:

$$D_G = D_{MED} \times D_{HC} \quad (7-6)$$

در واقع می‌توان رابطه بیان شده را به صورت نموداری که در شکل (6-3) نشان داده شده است مدل کرد که هر منحنی در آن نماینده نمونه‌هائی با فاصله یکسان از اشاره ورودی می‌باشد. بنابراین چنانچه هر نمونه از لحاظ یک ویژگی فاصله بسیار کمی داشته باشد، آن ویژگی نسبت به ویژگی دیگر اهمیت بیشتری پیدا می‌کند.



شکل (۳-۶): روش سوم

۳-۲-۶ روش نزدیکترین همسایگی اقلیدسی به صورت تطبیقی

در روش سوم نیز همه نمونه‌های کلاس‌های انتخاب شده مورد آزمایش قرار می‌گیرند با این تفاوت

که تابع فاصله این بار به صورت زیر تعریف می‌شود:

$$D_E = \sqrt{\alpha^2 D_{MED}^2 + \beta^2 D_{HC}^2} \quad (۸-۶)$$

که در آن نسبت $\frac{\alpha}{\beta}$ نیز از رابطه زیر به دست می‌آید:

$$\frac{\alpha}{\beta} = \frac{\bar{\mu}_{HC}}{\bar{\mu}_{MED}} \times \frac{\bar{A}_{MED}}{\bar{A}_{HC}} \quad (۹-۶)$$

که در این رابطه $\bar{\mu}$ میانگین تخمینی و \bar{A} میزان دقت تخمینی هر یک از ویژگی‌ها می‌باشد. رابطه

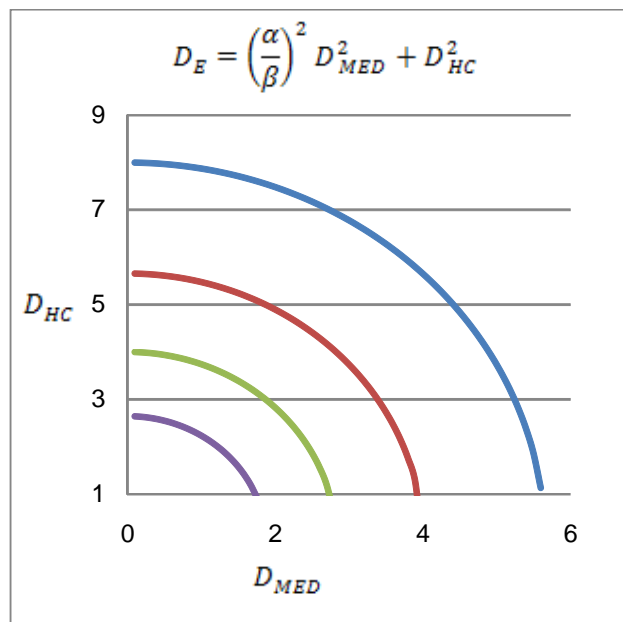
ضرایب تطبیقی α و β بیانگر این مطلب می‌باشد که فواصل D_{MED} و D_{HC} ممکن است دارای میانگین

یکسانی نباشند. همچنین با توجه به نتایج آزمایشات، دقت کلاسه‌بندی بوسیله ویژگی LMED با کلاسه‌بندی توسط ویژگی منحنی مسیر حرکت دست با یکدیگر متفاوت است. به این ترتیب با محاسبه ضرایب تطبیقی فوق فواصل D_{HC} و D_{MED} دارای مقیاس تقریباً یکسانی می‌شوند و چنانچه دقت کلاسه‌بندی با کمک یک ویژگی نسبت به دیگری بالاتر باشد، ضریب تطبیقی فاصله مربوط به آن ویژگی بزرگتر می‌شود تا آن فاصله اهمیت بیشتری پیدا کند.

در این روش نیز مانند روش اول نمونه‌ای که کمترین فاصله را از اشاره ورودی داشته باشد، تعیین-کننده کلاس آن خواهد بود. به همین دلیل می‌توان رادیکال را از رابطه تابع فاصله حذف نمود و آن را به صورت زیر ساده کرد:

$$D_E = \left(\frac{\alpha}{\beta}\right)^2 D_{MED}^2 + D_{HC}^2 \quad (۱۰-۶)$$

منحنی‌های کلاس‌هائی با فاصله یکسان، به صورت بیضی‌های متحدالمرکز در می‌آید که در شکل (۶-۴) نشان داده شده است. همانطور که ملاحظه می‌شود در روش سوم برعکس روش قبلی چنانچه نمونه‌ای از لحاظ یک ویژگی فاصله بسیار کمی پیدا کند، اهمیت آن برای تعیین فاصله‌ی کلی کاهش می‌یابد.



شکل (۴-۶): منحنی‌های نشان دهنده‌ی نمونه‌ها با فاصله‌ی یکسان از اشاره‌ی ورودی در روش سوم

فصل هفتم

نتایج و مقایسه

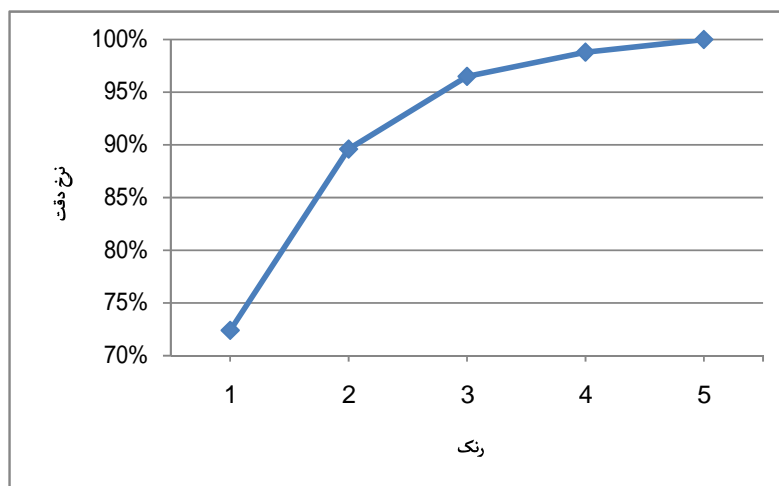
مقدمه

این فصل به دو قسمت تقسیم‌بندی می‌شود که قسمت اول شامل بررسی نتایج بدست آمده در مرحله‌ی اول برای حذف کلاس‌های نامطلوب و نتایج روش‌های پیشنهادی در مرحله‌ی دوم برای کلاسه‌بندی نهایی به همراه تجزیه و تحلیل هر یک می‌باشد. در قسمت دوم به مقایسه الگوریتم پیشنهادی از نظر تجهیزات به کار رفته، تئوری‌ها و ویژگی‌های مورد استفاده، نوع پایگاه داده، تعداد اشارات، روش پیاده سازی شده و نتایج بدست آمده با چند پروژه‌ی دیگر که در زمینه‌ی تشخیص زبان اشاره به صورت کلمات جدا از هم صورت گرفته است پرداخته می‌شود.

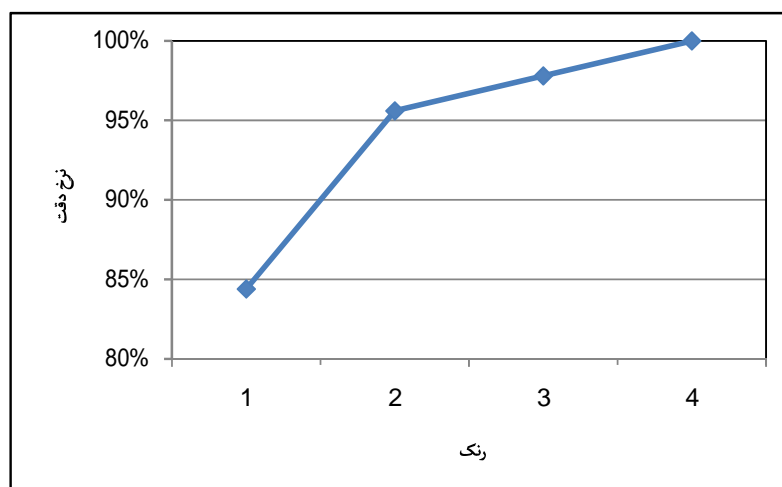
۷-۱ بررسی نتایج الگوریتم پیشنهادی

۷-۱-۱ نتیجه کلاسه‌بندی با استفاده از ویژگی منحنی مسیر حرکت دست

کلاسه‌بندی بر اساس کمترین فاصله DTW میان منحنی‌های حرکت دست نسبت به سر از دقت نسبتاً خوبی برخوردار است و مقدار آن، همانطور که در نمودارهای شکل (۷-۱) مشاهده می‌شود برای اشارات یک دستی با رنگ ۵ و اشارات دو دستی با رنگ ۴ به صد در صد می‌رسد. علاوه بر آن، این روش به علت کاهش تعداد فریم‌ها از سرعت بالایی برخوردار می‌باشد. نکته‌ی دیگری که از نمودارها بر می‌آید آن است که دقت کلاسه‌بندی برای اشارات دو دستی بیشتر از اشارات یک دستی می‌باشد. دلیل این امر نیز آن است که در اشارات با دو دست، از منحنی مسیر حرکت هر دو دست برای محاسبه‌ی فاصله‌ی DTW میان اشارات بهره گرفته می‌شود و در نتیجه دقت تشخیص، بالاتر می‌رود.



(الف)



(ب)

شکل (۱۰۷): دقت کلاسه‌بندی با استفاده از ویژگی منحنی مساحت تصویر دست بر حسب رنگ، لف) برای اشارات با ۱ دست، ب) برای اشارات با ۲ دست

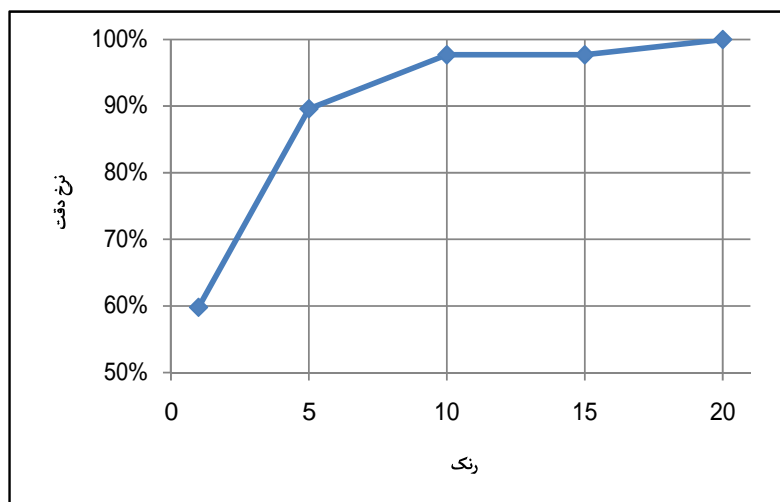
۷-۱ نتیجه کلاسه‌بندی با استفاده از ویژگی منحنی مساحت تصویر دست

میزان دقت در کلاسه‌بندی با رنگ بیشتر از ۱ بر اساس کمترین فاصله DTW میان منحنی مساحت مستطیل محاط بردست برای اشارات یک دستی و با رنگ ۲۰ و برای اشارات دو دستی، با رنگ ۱۵ به ۱۰۰٪ می‌رسد. همچنین بخاطر کاهش تعداد فریم‌ها، این بخش همانند بخش اول از

سرعت بالایی برخوردار بوده و در نتیجه برای استفاده در مرحله‌ی حذف کلاس‌های نامطلوب انتخاب شده است. علاوه بر این کلاسه‌بندی برای اشارات با ۲ دست نسبت به اشارات با ۱ دست از دقت بیشتری برخوردار است که دلیل آن را می‌توان استفاده از ویژگی منحنی مساحت تصویر هر دو دست عنوان نمود.

۷-۱-۳ نتیجه کلاسه‌بندی با استفاده از ویژگی منحنی شیب وتر تصویر دست

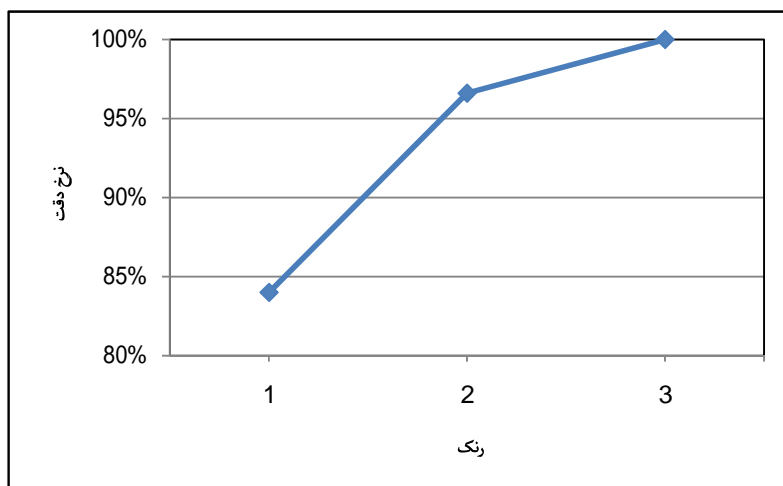
نتیجه کلاسه‌بندی بر اساس فاصله DTW میان منحنی شیب وتر مستطیل محاط بر دست با رنگ بیشتر از ۱_ برای اشارات یک دستی در نمودار شکل (۷-۳) نشان داده شده است. همانطور که در شکل (۷-۲) دیده می‌شود میزان دقت برای کلاسه‌بندی این دسته از اشارات با رنگ ۲۰ به ۱۰۰٪ می‌رسد. در این بخش میزان دقت برای اشارات دو دستی با رنگ ۵ به ۱۰۰٪ می‌رسد. دلیل این اختلاف زیاد را می‌توان علاوه بر بیشتر بودن تعداد نمونه‌ها برای اشارات یک دستی در پایگاه داده مورد استفاده و میزان اطلاعات کمتر به خاطر استفاده از یک دست در این گروه از اشارات، نوع خاص ویژگی شیب وتر مستطیل محاط بر دست از نظر غیر خطی بودن و نیز تشابه زیاد آن بین اشارات به دلیل محدودی تغییرات کم بیان نمود. این روش با توجه به مطالب ذکر شده و همچنین داشتن سرعت بالا در مرحله‌ی اول بکار گرفته شده است.



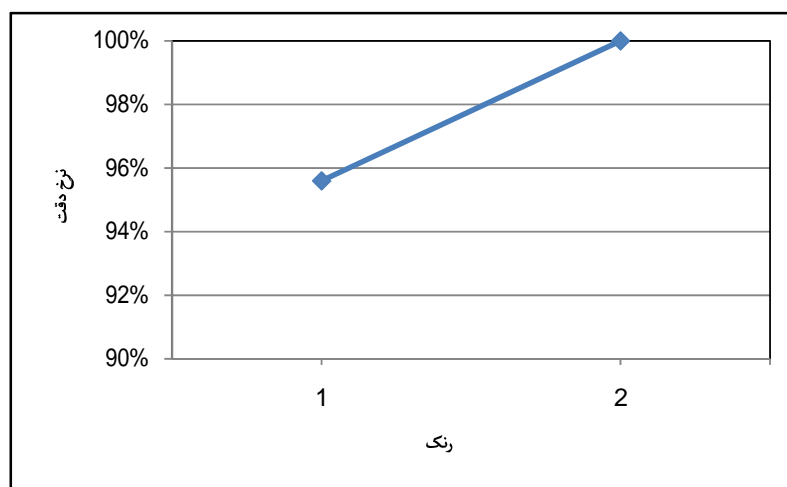
شکل (۲۰۷): دقت کلاسه‌بندی با استفاده از ویژگی منحنی زاویه‌ی وتر تصویر دست بر حسب رنک برای اشارات با ۱ دست

۱-۷ ۴ نتایج انتخاب کلاس‌های مطلوب با استفاده از نتایج بخش‌های ۱، ۲ و ۳.

در این بخش به بررسی نتایج حاصل از انتخاب کلاس‌های مطلوب در مرحله‌ی اول پرداخته می‌شود. همانگونه که در بخش‌های پیشین ملاحظه گردید به طور متوسط دقت کلاسه‌بندی کمترین فاصله DTW با استفاده از منحنی مسیر حرکت دست نسبت به استفاده از دو ویژگی دیگر بیشتر بوده است و به همین خاطر در این بخش نیز مجدداً برای کلاسه‌بندی میان کلاس‌های مشترک در سه بخش قبلی برای حذف تعداد کلاس بیشتر به کار گرفته شده است. نمودارهای شکل (۳-۷) میزان دقت کلاسه‌بندی در این مرحله را بر حسب رنک نشان می‌دهد. همانطور که در این نمودارها دیده می‌شود دقت کلاسه‌بندی با رنک یکسان برای اشارات دو دستی بیشتر از اشارات یک دستی می‌باشد و در نهایت این مقدار برای اشارات دو دستی با رنک ۲ و برای اشارات یک دستی با رنک ۳ به ۱۰۰٪ می‌رسد.



(الف)



(ب)

شکل (۳-۷): نتایج کلاسه‌بندی مرحله‌ی اول با استفاده از نتایج کلاسه‌بندی با رنگ بیشتر از ۱ (الف) برای اشارات با ۱ دست، (ب) برای اشارات با ۲ دست

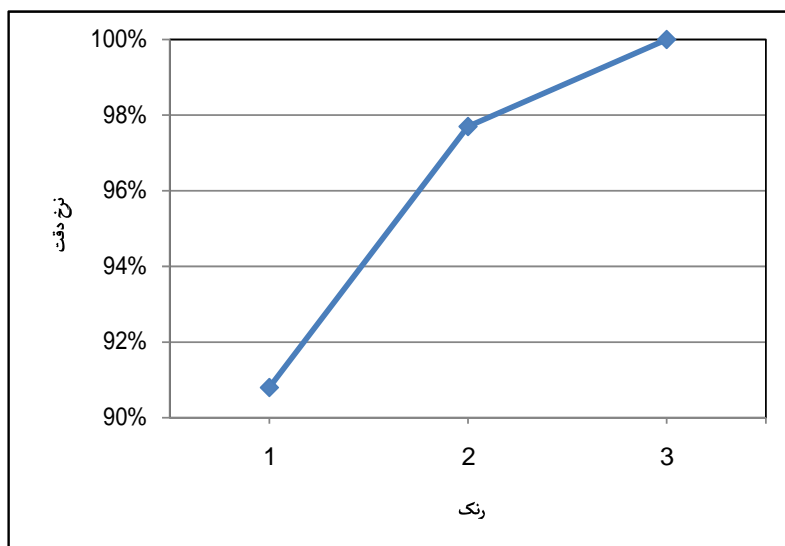
نکته‌ی دیگری که در این مرحله مطرح می‌شود دقت نسبتاً پایین کلاسه‌بندی در آن می‌باشد. دلیل این امر انتخاب حدود آستانه در بخش‌های اول به اندازه‌ای است تا در هیچ یک از اشارات، کلاس اصلی از دست نرود. در صورتی که حدود آستانه به نحوی تغییر کنند تا تعداد کلاس‌های کمتری در هر بخش انتخاب شود دقت مرحله‌ی اول افزایش پیدا خواهد کرد اما به خاطر از دست رفتن کلاس‌های اصلی در برخی از اشارات در ازای هیچ رنگ به ۱۰۰٪ نخواهد رسید.

۱-۷ ۵ نتایج کلاسه‌بندی بر اساس ویژگی پیشنهادی LMED (روش اول)

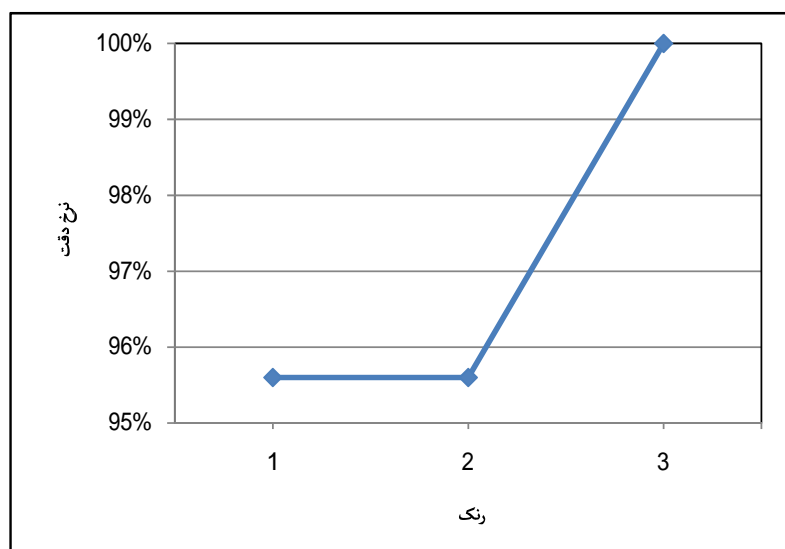
شکل (۴-۷) نمودارهای دقت کلاسه‌بندی با استفاده از ویژگی پیشنهادی LMED که روی کلاس‌های انتخاب شده در مرحله‌ی اول صورت گرفته است را بر حسب رنک نشان می‌دهد. همانطور که در این شکل مشاهده می‌شود کلاسه‌بندی توسط این ویژگی از دقت خوبی برخوردار و با کلاسه‌بندی که از ویژگی منحنی مسیر حرکت دست به تنهایی استفاده کرده است از لحاظ دقت قابل مقایسه می‌باشد چنانچه دقت آن در کلاسه‌بندی کلاس‌های انتخاب شده برای اشارات با دو دست به مقدار جزئی کمتر و برای اشارات با یک دست بیشتر است. از آنجا که ویژگی LMED به شکل دست و نحوه‌ی قرارگیری انگشتان بستگی دارد این نتیجه از لحاظ مقایسه و تشخیص زبان اشاره به کمک شکل دست بسیار حائز اهمیت می‌باشد.

۱-۷-۶ نتایج کلاسه‌بندی به روش مینیمم میانگین هندسی (روش دوم)

استفاده از هر دو ویژگی منحنی مسیر حرکت دست و LMED که به نوعی تعیین کننده شکل دست و حالت قرار گیری انگشتان می‌باشد موجب شده است تا در این روش دقت برای اشارات با یک دست افزایش چشمگیری داشته باشد. همانطور که در شکل (۵-۷) نیز نشان داده شده است دقت روش کلاسه‌بندی پیشنهادی برای این دسته از اشارات به ۹۸.۸٪ رسیده است.

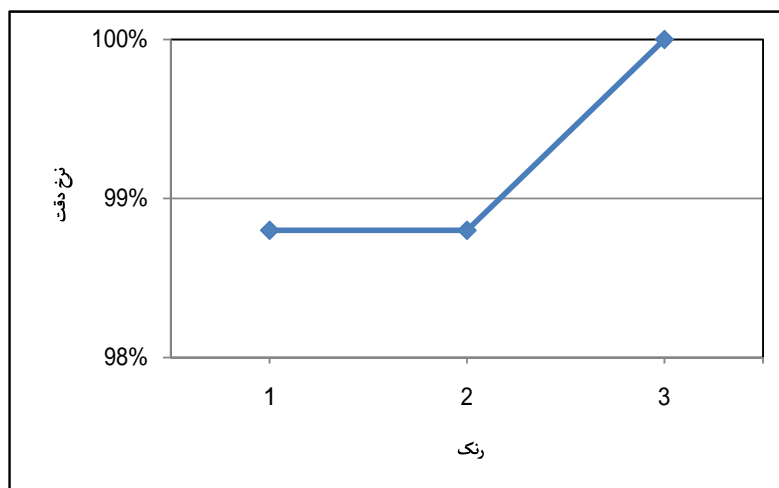


(الف)

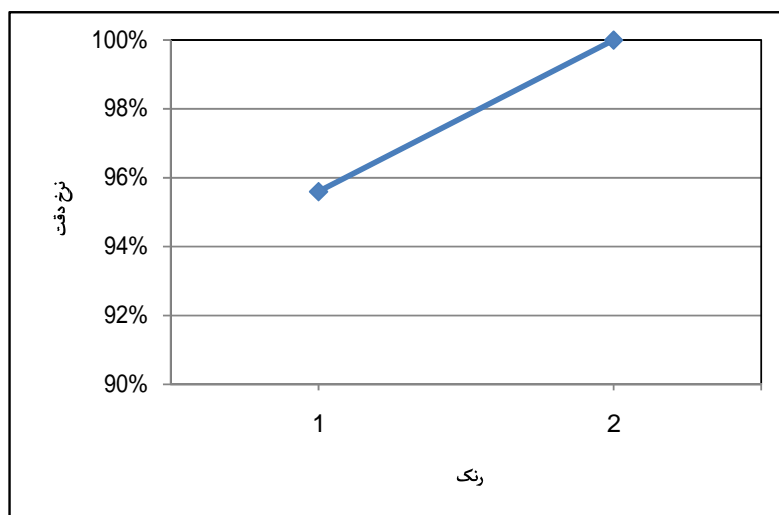


(ب)

شکل (۴-۷): نتایج کلاسه‌بندی بر اساس ویژگی‌های پیشنهادی (LMED)، الف) برای اشارات با ۱ دست، ب) برای اشارات با ۲ دست



(الف)



(ب)

شکل (۷-۵): نتایج کلاسه‌بندی به روش مینیمم میانگین هندسی با استفاده از هر دو ویژگی منحنی مسیر حرکت دست و LMED، الف) برای اشارات با ۱ دست، ب) برای اشارات با ۲ دست

همچنین برای اشارات با دو دست نیز مقدار مناسب قبلی حفظ شده است. این نتایج ارزش و اهمیت شکل و حالت دست را در زمان اجرای اشارات به خصوص اشارات یک دستی نشان می‌دهد و علاوه بر آن قدرت و توانایی ویژگی LMED را در مدل‌سازی شکل دست به خوبی اثبات می‌کند.

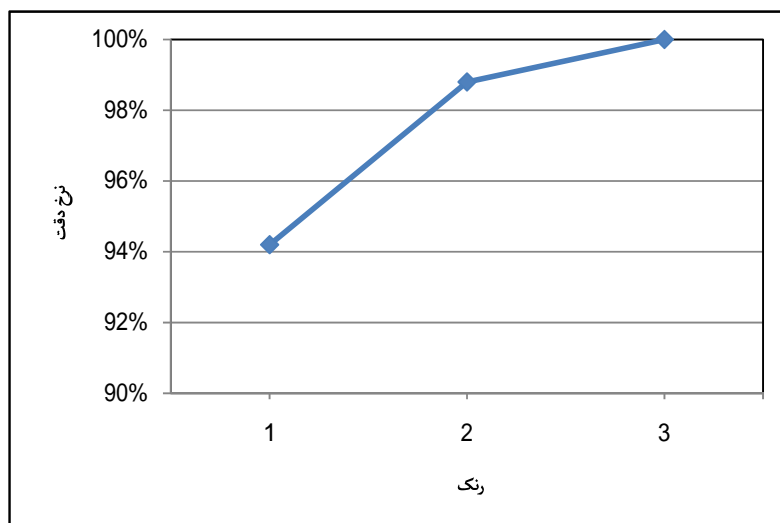
۷-۱ نتایج کلاسه‌بندی به روش نزدیکترین همسایگی اقلیدسی

انطباقی (روش سوم)

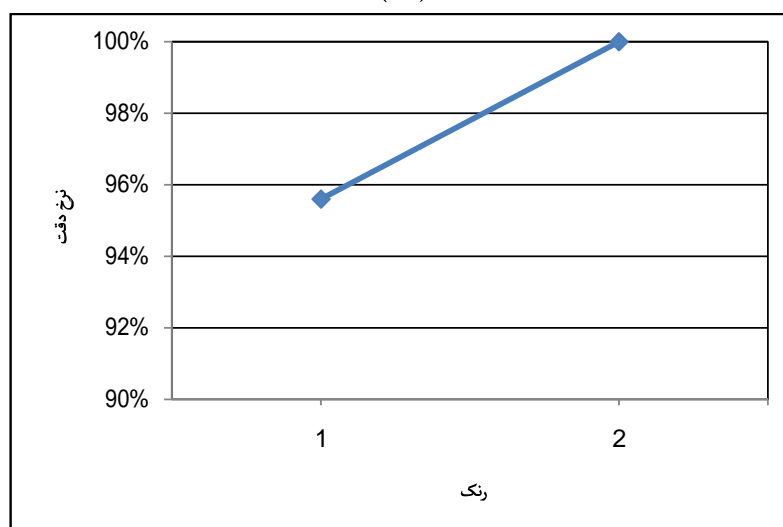
این روش نیز همانطور که در شکل (۶-۷) نشان داده شده است دقت تشخیص را برای اشارات یک دستی نسبت به روش پیشنهادی اول بهبود داده و برای اشارات دو دستی، دقت قبلی را حفظ کرده است.

۷-۱ نتیجه کلاسه‌بندی نهایی

در این بخش بهترین نتیجه‌ی بدست آمده برای همه‌ی اشارات محاسبه می‌گردد. به این ترتیب که برای اشارات با یک دست از روش دوم که بیشترین دقت را در میان روش‌های پیشنهادی داشته است استفاده می‌شود و برای اشارات با دو دست نیز یکی از سه روش پیشنهادی بکار گرفته می‌شود زیرا هر سه دارای دقت یکسانی بوده‌اند. نمودار شکل (۷-۷) نمودار نتیجه‌ی کلاسه‌بندی نهایی برای همه‌ی اشارات را نشان می‌دهد. همانطور که در این نمودار دیده می‌شود میزان دقت نهایی برای الگوریتم پیشنهادی به ۹۷.۷٪ رسیده است.

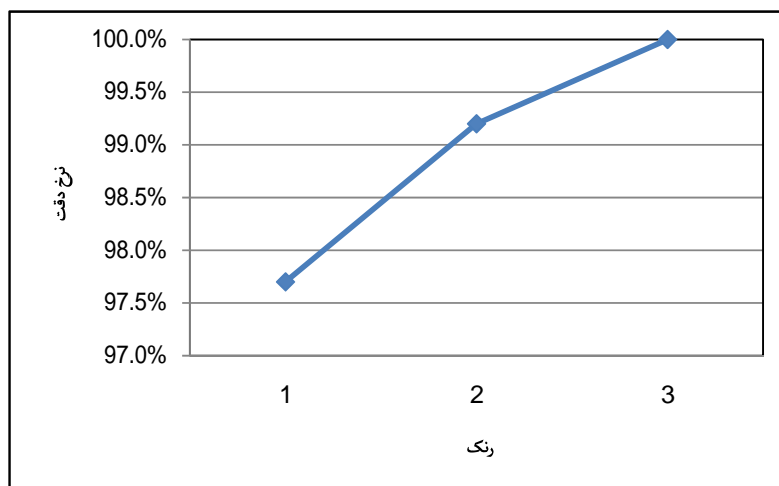


(الف)



(ب)

شکل (۶-۷): نتایج کلاسه‌بندی به روش نزدیکترین همسایگی اقلیدسی با استفاده از هر دو ویژگی منحنی مسیر حرکت دست و LMED، الف) برای اشارات با ۱ دست، ب) برای اشارات با ۲ دست



شکل (۷-۷): نتیجه کلاسه‌بندی نهایی

۷-۲ مقایسه نتایج

در این قسمت به مقایسه شرایط و نتایج این پروژه با چند کار دیگر که در زمینه‌ی تشخیص زبان اشاره به صورت کلمات جدا از هم انجام گرفته است پرداخته می‌شود. برای این کار ابتدا به بررسی شرایط و شرح مختصر دلائل استفاده از تئوری‌ها و روش‌های پیشنهادی در این پروژه می‌پردازیم. همانگونه که در فصل اول نیز بدان اشاره گردید الگوریتم پیشنهادی برای تشخیص زبان اشاره تنها بر اساس مشاهده بوده و به همین دلیل در آن از تجهیزات ویژه‌ی دنبال کننده همچون ردیاب مافوق صوت و آشکارسازهای مغناطیسی استفاده نشده است. علاوه بر آن به خاطر راحتی کاربر و عملی ساختن آن از هیچ نوع تجهیزات اضافی مانند پروژکتورهای مخصوص و دستکش‌های رنگی کمک گرفته نشده و ویژگی‌های مورد نیاز تنها از تصاویری که توسط یک دوربین در مکان ثابت گرفته شده-ند استخراج می‌شود. این کار و همچنین پیاده سازی روش‌های کلاسه‌بندی پیشنهادی توسط برنامه نویسی در نرم‌افزار مطلب صورت گرفته است.

در الگوریتم کلاسه‌بندی از تئوری DTW برای انطباق زمانی میان اشارات و از تئوری LMED برای مدل‌سازی شکل دست بهره گرفته می‌شود. همچنین برای افزایش سرعت و دقت تشخیص اشاره، الگوریتم پیشنهادی به دو مرحله تقسیم‌بندی می‌شود؛ مرحله‌ی اول حذف بسیاری از کلاس‌های نامطلوب به کمک روش‌های پر سرعت و مرحله‌ی دوم کلاسه‌بندی نهایی با استفاده از پردازش تصویر دست با دقت مناسب. در نهایت برای آزمایش الگوریتم پیشنهادی از یک پایگاه داده حاوی ۱۳۷ نمونه برای ۴۷ اشاره متفاوت استفاده می‌شود. بنابراین به طور متوسط در هر کلاس کمتر از سه نمونه برای آزمایش وجود خواهد داشت که این مطلب کار کلاسه‌بندی را برای تعداد ۴۷ اشاره با دشواری مواجه کرده و در نتیجه الگوریتم پر قدرتی را برای این کار طلب می‌نماید.

در جدول (۷-۱) این پروژه و چند پروژه‌ی دیگر که اخیراً در زمینه‌ی تشخیص زبان اشاره انجام گرفته از لحاظ شرایط اجرا و نتیجه‌ی نهایی با هم مقایسه شده‌اند. همانطور که در این جدول ملاحظه می‌شود بسیاری از روش‌ها از تجهیزات اضافی استفاده کرده‌اند و عموماً از پایگاه داده متفاوت با تعداد نمونه‌های زیاد نسبت به تعداد اشارات به کار رفته بهره برده‌اند. از آنجا که پایگاه داده‌ی بیشتر این پروژه‌ها در دسترس نبوده شرایط مقایسه‌ی دقیق میان آنها و الگوریتم پیشنهادی نیز فراهم نشده است اما این الگوریتم با توجه به شرایط اجرا و برخورداری از سرعت بالا نسبت به روش‌های دیگر از دقت بسیار خوبی برخوردار می‌باشد.

جدول (۱۰۷) : مقایسه‌ی شرایط و نتیجه‌ی الگوریتم پیشنهادی با چند کار انجام شده در زمینه‌ی تشخیص زبان اشاره

نرخ دقت	روش کلاسه‌بندی	تئوری‌ها	تجهیزات اضافی	تعداد اشارات	زبان اشاره
۷۵٪	CDFD, Q_DFFM	SDTW	استفاده از منحنی مسیر سه‌بعدی	۱۲۰	DSL
۶۸٪	FLD	VQPCA	Cyber Glove و آشکارسازهای مغناطیسی	۲۸	SEE
۹۲٪	DTW با رنگ ۲	Dempster_shafer	دست‌کش‌های رنگی	۱۰۰	CSL
۵۴٪	HMM	ICA	—	۴۹	BSL
۸۲٪	HMM	Tied mixture densities	دستکش‌های رنگی	۴۳۹	CSL
۸۶٪	Time – delay neural network	Pixel level	—	۴۰	ASL
۵۵٪	Adaptive fuzzy expert system	—	دستکش‌های رنگی	۲۲	Australian sign language mixed with artificial signs
۷۷٪	Minimum Geometric Mean Distance	DTW, LMED	—	۴۷	16-BOSTON201

- [1] C. L. Lisetti and D. J. Schiano, (2000), “**Automatic classification of single facial images**,” Pragmatics Cogn., vol. 8, pp. 185–235.
- [2] V. I. Pavlovic, R. Sharma, and T. S. Huang, (Jul. 1997), “**Visual interpretation of hand gestures for human computer interaction**,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 7, pp. 677–695.
- [3] L. R. Rabiner, (Feb. 1989), “**A tutorial on hidden Markov models and selected applications in speech recognition**,” Proc. IEEE, vol. 77, no. 2, pp. 257–285.
- [4] S. Mitra and T. Acharya, (2003), Data Mining: “**Multimedia, Soft Computing, and Bioinformatics**”. New York: Wiley.
- [5] J. Yamato, J. Ohya, and K. Ishii, (1992), “**Recognizing human action in timesequential images using hidden Markov model**,” in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogn., Champaign, IL, pp. 379–385.
- [6] F. Samaria and S. Young, (1994), “**HMM-based architecture for face identification**”, Image Vis. Comput., vol. 12, pp. 537–543.
- [7] G. Welch and G. Bishop, (2000), “**An introduction to the Kalman filter**,” Dept. Comput. Sci., Univ. North Carolina, Chapel Hill, Tech. Rep. TR95041.
- [8] S. Arulapalam, S. Maskell, N. Gordon, and T. Clapp, (Feb. 2001) “**A tutorial on particle filters for on-line nonlinear/non-Gaussian Bayesian tracking**,” IEEE Trans. Signal Process., vol. 50, no. 2, pp. 174–188.
- [9] C. Kwok, D. Fox, and M. Meila, (Mar. 2004), “**Real-time particle filters**,” Proc. IEEE, vol. 92, no. 3, pp. 469–484.
- [10] M. Isard and A. Blake, (1996), “**Contour tracking by stochastic propagation of conditional density**,” in Proc. Eur. Conf. Comput. Vis., Cambridge, U.K , pp. 343–356.
- [11] M. Isard and A. Blake, “**CONDENSATION—Conditional density propagation for visual tracking**,” Int. J. Comput. Vis., vol. 1, pp. 5–28, 1998.
- [12] A. Doucet, N. de Freitas, and N. Gordon, Eds.,(2001), “**Sequential Monte Carlo in Practice**”. New York: Springer-Verlag.
- [13] M. Isard and A. Blake, (1998), “**A mixed-state condensation tracker with automatic model-switching**,” in Proc. 6th Int. Conf. Comput. Vis., Mumbai, India, pp. 107–112.
- [14] J. Davis and M. Shah, (1994), “**Visual gesture recognition**,” Vis., Image Signal Process., vol. 141, pp. 101–106.
- [15] A. F. Bobick and A. D. Wilson, (Dec. 1997), “**A state-based approach to the representation and recognition of gesture**,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 12, pp. 1235–1337.

- [16] M. Yeasin and S. Chaudhuri, (2000), “**Visual understanding of dynamic hand gestures**,” Pattern Recogn., vol. 33, pp. 1805–1817.
- [17] P. Hong, M. Turk, and T. S. Huang, (Mar. 2000), “**Gesture modeling and recognition using finite state machines**,” in Proc. 4th IEEE Int. Conf. Autom. Face Gesture Recogn., Grenoble, France, pp. 410–415.
- [18] R. O. Duda and P. E. Hart, (1973), “**Pattern Classification and Scene Analysis**”. New York: Wiley.
- [19] R. C. Gonzalez and R. E. Woods, (1992), “**Digital Image Processing**”. Reading, MA: Addison-Wesley.
- [20] K. Mase, (1991), “**Recognition of facial expression from optical flow**,” IEICE Trans., vol. E 74, pp. 3474–3483.
- [21] B. K. P. Horn and B. G. Schunck, (1981), “**Determining optical flow**,” Artif. Intell., vol. 17, pp. 185–203.
- [22] M. Kass, A. Witkin, and D. Terzopoulos, (1988), “**SNAKE: Active contour models**”, Int. J. Comput. Vis., pp. 321–331.
- [23] S. Haykin, (1994), Neural Networks: “**A Comprehensive Foundation**”. New York: Macmillan.
- [24] K. K. Sung and T. Poggio, (1995), “**Learning human face detection in cluttered scenes**,” in Proc. IEEE 6th Int. Conf. CAIP’95, Prague, Czech Republic, pp. 432–439.
- [25] H. Rowley, S. Baluja, and T. Kanade, (Jan. 1998), “**Neural network-based face detection**”, IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 1, pp. 23–38.
- [26] Y. L. Tian, T. Kanade, and J. F. Cohn, (Feb. 2001), “**Recognizing action units for facial expression analysis**,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 2, pp. 97–115.
- [27] M. S. Yang and N. Ahuja, (Jun. 1998), “**Recognizing hand gesture using motion trajectories**”, in Proc. IEEE CS Conf. Comput. Vis. Pattern Recogn., vol. 1, Fort Collins, CO, pp. 466–472.
- [28] S. Gutta and H. Wechsler, (1997), “**Face recognition using hybrid classifiers**,” Pattern Recogn., vol. 30, pp. 539–553.
- [29] M. Rosenblum, Y. Yacoob, and L. S. Davis, (Sep. 1996), “**Human expression recognition from motion using a radial basis function network architecture**”, IEEE Trans. Neural Netw., vol. 7, no. 5, pp. 1121–1138.
- [30] M. Assan, K. Grobel, (Oct. 1997), “**Isolated Sign Language Recognition using Hidden Markov Models**”. In IEEE International Conference on Systems, Man and Cybernetics, Vol. 1, pp. 162–167, Orlando, FL, USA.
- [31] C.L. Huang, W.Y. Huang, (April 1998), “**Sign Language Recognition using Model-based Tracking and a 3D Hopfield Neural Network**”. Machine Vision and Applications, Vol. 10, No. 5-6, pp. 292–307.

- [32] E.J. Holden, R. Owens, (March 2000), “**Visual Sign Language Recognition**”. In 10th International Workshop on Theoretical Foundations of Computer Vision, Vol. 2032 of Lecture Notes in Computer Science, pp. 270–288, Dagstuhl Castle, Germany. Springer.
- [33] M.H. Yang, N. Ahuja, M. Tabb, (Aug. 2002), “**Extraction of 2D Motion Trajectories and Its Application to Hand Gesture Recognition**”. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 8, pp. 1061–1074.
- [34] N. Tanibata, N. Shimada, Y. Shirai, (May 2002), “**Extraction of Hand Features for Recognition of Sign Language Words**”. In 15th International Conference on Vision Interface (VI 2002), pp. 391–398, Calgary, Canada.
- [35] L.G. Zhang, Y. Chen, G. Fang, X. Chen, W. Gao, (2004), “**A Vision-Based Sign Language Recognition System using Tied-Mixture Density HMM**”. In 6th International Conference on Multimodal Interfaces (ICMI '04), pp. 198–204, New York, NY, USA. ACM Press.
- [36] R. Bowden, D. Windridge, T. Kadir, A. Zisserman, M. Brady, (May 2004), “**A Linguistic Feature Vector for the Visual Interpretation of Sign Language**”. In 8th European Conference on Computer Vision (ECCV), Part IV, Vol. 3024 of Lecture Notes in Computer Science, pp. 390–401, Prague, Czech Republic. Springer.
- [37] J. Zieren, K.F. Kraiss, (June 2005), “**Robust Person-Independent Visual Sign Language Recognition**”. In 2nd Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA 2005), Vol. 3522 of Lecture Notes in Computer Science, pp. 520–528, Estoril, Portugal. Springer.
- [38] U. von Agris, D. Schneider, J. Zieren, K.F. Kraiss, (June 2006), “**Signer Adaptation for Isolated Sign Language Recognition**”. In IEEE Workshop on Vision for Human Computer Interaction (V4HCI), New York, NY, USA.
- [39] Canzler U., (2005), Ph.D. thesis, “**Nicht-intrusive Mimikanalyse**”. RWTH Aachen University, Lehrstuhl für technische Informatik, Aachen, Germany.
- [40] Qi Wang a, Xilin Chen b, Liang-Guo Zhang b, Chunli Wang b, Wen Gao a,b, (2007), “**Viewpoint invariant sign language recognition**” Computer Vision and Image Understanding 108, 87–97
- [41] W.W. Kong, Surendra Ranganath, (2008), “**Signing Exact English (SEE): Modeling and recognition**” Pattern Recognition 41 1638 – 1652
- [42] Jeroen F. Lichtenauer, Emile A. Hendriks, and Marcel J.T. Reinders, (11 NOVEMBER 2008), “**Sign Language Recognition by Combining Statistical DTW and Independent Classification**” IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 30, NO..
- [43] M. Isard, A. Blake, (August 1998), “**CONDENSATION – conditional density propagation for visual tracking**”. International Journal of Computer Vision, Vol. 29, No. 1, pp. 5–28.
- [44] R.E. Kalman, (March 1960), “**A New Approach to Linear Filtering and Prediction Problems**”. Transaction of the ASME - Journal of Basic Engineering, Vol. 82, pp. 35–45.

- [45] D. Comaniciu, V. Ramesh, P. Meer, (June 2000), “**Real-Time Tracking of Non-Rigid Objects using Mean Shift**”. In IEEE Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 142–151, Hilton Head Island, South Carolina, USA.
- [46] G.R. Bradski, (1998), “**Computer Vision Face Tracking For Use in a Perceptual User Interface**”. Intel Technology Journal, Vol. Q2, pp. 15–26.
- [47] Philippe Dreuw, (Januar 2005), master-thesis, “**Appearance-Based Gesture Recognition**”, Diplomarbeit im Fach Informatik Rheinisch-Westfälische Technische Hochschule Aachen Lehrstuhl für Informatik VI Prof. Dr.-Ing. H. Ney.
- [48] D.M. Gavrila, (February 1999), “**The Visual Analysis of Human Movement: A Survey**”. Computer Vision and Image Understanding, Vol. 73, No. 1, pp. 82–98.
- [49] A.Micilotta, R. Bowden, (September 2004), “**View-based Location and Tracking of Body Parts for Visual Interaction**”. In BMVC, Vol. 2, pp. 849–858, Kingston UK.
- [50] D.M. Gavrila, J. Giebel, S. Munder, (June 2004), “**Vision-based Pedestrian Detection: the PROTECTOR+ System**”. In Proceedings of the IEEE Intelligent Vehicles Symposium, pp. 13– 18, Parma, Italy.
- [51] M. Balcells, D. DeMenthon, D. Doermann, (2004), “**An appearance-based approach for consistent labeling of humans and objects in video**”. Pattern Analysis and Applications, Vol., pp. in press.
- [52] P. Dreuw, T. Deselaers, D. Rybach, D. Keysers, H. Ney, (April 2006), “**Tracking Using Dynamic Programming for Appearance-Based Sign Language Recognition**”. In 7th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2006), pp. 293–298, Southampton, UK.
- [53] David Rybach, (2006), “**Appearance-Based Features for Automatic Continuous Sign Language Recognition**” Diplomarbeit im Fach Informatik Rheinisch-Westfälische Technische Hochschule Aachen Lehrstuhl für Informatik 6 Prof. Dr.-Ing. H. Ney.
- [54] L. G. Roberts, (1965), “**Machine Perception of Three-Dimensional Solids,**” in Optical and Electro-Optical Information Processing, J. T. Tippett et al., Eds., MIT Press, Cambridge, MA, 159–197.
- [55] J. M. S. Prewitt, (1970), “**Object Enhancement and Extraction,**” in Picture Processing and Psychopictorics, B. S. Lipkin and A. Rosenfeld, Eds., Academic Press, New York.
- [56] R. O. Duda and P. E. Hart, (1973), “**Pattern Classification and Scene Analysis**”, Wiley, New York.
- [57] W. Frei and C. Chen, (October 1977) “**Fast Boundary Detection: A Generalization and a New Algorithm,**” IEEE Trans. Computers, C-26, 10, 988–998.
- [58] I. Abdou, (1973), “**Quantitative Methods of Edge Detection,**” USCIP Report 830, Image Processing Institute, University of Southern California, Los Angeles.
- [59] E. Argyle, (February 1971), “**Techniques for Edge Detection,**” Proc. IEEE, 59, 2, 285–287.

- [60] I. D. G. Macleod, (1970), “**On Finding Structure in Pictures,**” in *Picture Processing and Psychopictorics*, B. S. Lipkin and A. Rosenfeld, Eds., Academic Press, New York.
- [61] I. D. G. Macleod, (March 1972), “**Comments on Techniques for Edge Detection,**” *Proc. IEEE*, 60, 3, 344.
- [62] J. Canny, (November 1986), “**A Computational Approach to Edge Detection,**” *IEEE Trans. Pattern Analysis and Machine Intelligence*, PAMI-8, 6, 679–698.
- [63] D. Demigny and T. Kamie, (November 1997), “**A Discrete Expression of Canny’s Criteria for Step Edge Detector Performances Evaluation,**” *IEEE Trans. Pattern Analysis and Machine Intelligence*, PAMI-19, 11, 1199–1211.
- [64] WILLIAM K. PRATT, (2007), “**DIGITAL IMAGE PROCESSING**” , PIKS Scientific Inside , Copyright © 2007 by John Wiley & Sons, ISBN: 978-0-471-76777-0
- [65] Gérard Blanchet Maurice Charbit, (2006), “**Digital Signal and Image Processing using MATLAB®**”, © HERMES Science Europe Ltd, 2001, © ISTE Ltd, 2006, ISBN-13: 978-1-905209-13-2, ISBN-10: 1-905209-13-4
- [66] C. Neidle, J. Kegl, D. MacLaughlin, B. Bahan, R. Lee, (Dec. 1999), “**The Syntax of American Sign Language: Functional Categories and Hierarchical Structure**”. MIT Press, Cambridge, MA, USA.
- [67] B. Bauer, H. Hienz, K.F. Kraiss, (Sept. 2000), “**Video-Based Continuous Sign Language Recognition Using Statistical Methods**”. In *Proceedings of the International Conference on Pattern Recognition*, Vol. 2, pp. 463–466, Barcelona, Spain.
- [68] C. Neidle, (2001), “**SignStreamTM: A Database Tool for Research on Visual-Gestural Language**”. *Sign Language and Linguistics*, Vol. 4, No. 1/2, pp. 203–214.
- [69] M. Zahedi, P. Dreuw, D. Rybach, T. Deselaers, H. Ney, (May 2006), “**Continuous Sign Language Recognition - Approaches from Speech Recognition and Available Data Resources**”. In *2nd Workshop on the Representation and Processing of Sign Languages: Lexicographic Matters and Didactic Scenarios*, Genoa, Italy.
- [70] M. Zahedi, D. Keysers, H. Ney, (May 2005), “**Pronunciation Clustering and Modelling of Variability for Appearance-Based Sign Language Recognition**”. In *Proceedings of GW 2005, 6th Interactional Gesture Workshop*, Vol. 3881 of LNAI, pp. 68–79, Berder Island, France. Springer Verlag.
- [71] M. Zahedi, D. Keysers, T. Deselaers, H. Ney, (2005), “**Combination of Tangent Distance and an Image Distortion Model for Appearance-Based Sign Language Recognition**”. In *Proceedings of DAGM 2005, 27th Annual meeting of the German Association for Pattern Recognition*, Vol. 3663 of LNCS, pp. 401–408, Vienna, Austria. Springer Verlag.

Abstract

Recognizing sign languages are widely used for communicating deaf people and people who hear hard with other people or with computer. Nowadays, recognizing by using images processing tools have been interested by researchers, because of its importance. In this research we want to recognize hints and hand signs by using 2-dimensional image processing without using any special equipment. One of the best methods for identifying hints and signs is to comparing hand motion path and modeling hand shape and how of locating its fingers.

We introduced and investigated maximum edge direction (MED) in our research and indicated that MED is a geometric feature that is related to hand structure. MED feature, statistically, indicate maximum direction of image edges and indicate direction of locating fingers in each hint, accurately. For extracting this feature, first we take a binary image of hand image edges and then appropriate proposed structural element, which works only on edge pixel, are used. Also for a better modeling of hand shape, LMED feature are used. Then target images are partitioned to smaller region with definite size and MED of each region are calculated. Using this feature strongly decrease the amount of needed database and increase recognizing rate.

Proposed algorithm for recognizing sign language contains 2 main steps: 1-initial classification 2- final classification. In first step, features like: hand motion path and curves of hand image's area and hand sinew angle are extracted. Then undesirable classes are removed by using DTW algorithm and nearest neighbor classifier. Finally 3 classes remain that right class is certainly between these classes. Consequently speed and accuracy of sign language recognition are improved. In second step ,first by using DTW algorithm, that curve of hand motion path is it's function, same frames are selected and then distance between input sign and 3 classes ,in first step, are calculated by using two features :LMED and hand motion path.

In this project 3 methods are proposed for sign language classification:

- 1- Minimum distance method by using LMED feature.
- 2- Minimum geometrical average method
- 3- Minimum Euclidean distance method adaptively.

Proposed classifiers are tested on a collation of ASL sign that contain 47 different words with 137 samples. In this experiment we reached to accuracy rate of %97.7 that are about %98.9 in best situation for hinting with one hand and about %95.6 for hinting with 2 hands. Results indicate that proposed method has a high efficiency.

Key words:

Sign language recognition, DTW algorithm, MED and LMED features.