

صلى الله عليه وسلم

تعهد نامه

اینجانب اسد زهرنی دانشجوی دوره کارشناسی ارشد / دکتری رشته برق دانشگاه صنعتی شاهرود نویسنده پایان نامه / رساله در مباحث سیستم های انرژی در صنایع تحت راهنمایی دکتر سید علی حسینی 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 .
- «عقوق 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: ۱۸/۰۱/۱۹

امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج ، کتاب ، برنامه های رایانه ای ، نرم افزارها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می باشد . این مطلب باید به نحو مقتضی در توثیقات 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 .

شماره: ۰۷۴۶-ت.ب
تاریخ: ۱۳۸۹/۰۱/۱۸
ویرایش: - - - -



مدیریت تحصیلات تکمیلی

فرم شماره (۵)

بسمه تعالی

ارزشیابی نهایی پایان نامه ارشد

(این فرم براساس فرم شماره ۴ تکمیل خواهد شد)

نام و نام خانوادگی دانشجو: ارس درگزی رشته و گرایش تحصیلی: مهندسی برق- الکترونیک

سال ورود: ۳۸۶۱

تاریخ دفاع: ۱۳۸۹/۰۱/۱۸

عنوان پایان نامه: ردیابی حرکت دست در زبان اشاره

ردیف	نمره از ۲۰ (فرم ۱)	پایان نامه	نمره	امضاء
۱- اساتید راهنما ۱- علی سلیمی نی ایزدی ۲-		با مشاور نمره از ۷ بدون مشاور نمره از ۸	۹,۹۱۲	
اساتید مشاور ۱- علیرضا احمدی فرد ۲-		نمره از ۲	۱,۹	
داوران ۱- حمید طهرسیان ۲- حسین مشهوری		با مشاور نمره از ۴/۵ برای هر داور بدون مشاور نمره از ۵ برای هر داور	۴,۱۴۳-۱ ۴,۲۱۹-۲	
نماینده تحصیلات تکمیلی علی دستمان		۱ نمره: نحوه دفاع، طول دوره، گزارشات پیشرفت- حضور در دانشکده	۰/۵	
		(۱ نمره مقالات)	۲	سید سعید محمد زورنالی و نوحه نامه نهایی
نمره کل		عدد: ۱۹,۴۹۶		
		حروف: نوزده و شصت و نه هزار و چهارصد و نود و شش		

تبصره ۱: نحوه نمره دهی مقالات چاپ شده یا دارای پذیرش: (هر مقاله علمی- پژوهشی ۱ نمره) و (هر مقاله کنفرانسی ۰/۵ نمره)

تبصره ۲: چنانچه دانشجویان مقالات علمی و پژوهشی مزاد بر سقف یک نمره داشته باشند به سقف امتیازهای بندهای دیگر می تواند تا یک نمره دیگر اضافه گردد.

امضاء:

مسئول تحصیلات تکمیلی دانشکده: علیرضا احمدی فرد

شماره : ۰۷۴۶/آت.ب
تاریخ : ۱۳۸۹/۰۱/۱۸
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 / آقای ارس درگزی رشته برق گرایش: الکترونیک تحت عنوان: ردیابی حرکت دست در زبان اشاره که در تاریخ ۱۳۸۹/۰۱/۱۸ با حضور هیأت محترم داوران در دانشگاه صنعتی شاهرود برگزار گردید به شرح زیر است:

قبول (با درجه امتیاز ۱۹,۷۴۴) دفاع مجدد مردود

- ۱- عالی (۲۰ - ۱۹)
۲- بسیار خوب (۱۸/۹۹ - ۱۸)
۳- خوب (۱۷/۹۹ - ۱۶)
۴- قابل قبول (۱۵/۹۹ - ۱۴)

امضاء	مرتبه علمی	نام و نام خانوادگی	عضو هیأت داوران
	استاد	علی سلیمی	۱- استاد راهنما
	استاد	علیرضا احمدی زرد	۲- استاد مشاور
	استاد	علی دستمال	۳- نماینده شورای تحصیلات تکمیلی
	استاد	دکتر علی احمدی	۴- استاد ممتحن
	استاد	دکتر خسروی	۵- استاد ممتحن

تأیید رئیس دانشکده:



دانشگاه صنعتی شاهرود

دانشکده مهندسی برق و رباتیک

پایان نامه کارشناسی ارشد

عنوان پایان نامه:

ردیابی دست به روش تلفیقی جابجایی میانگین و فیلتر ذره ای

ارس درگزی

استاد راهنما: دکتر سید علی سلیمانی ایوری

استاد مشاور: دکتر علیرضا احمدی فرد

فروردین ۱۳۸۹

چکیده

فیلتر ذره‌ای و جابجایی میانگین دو روش موفق برای ردیابی اشیاء در رشته‌های تصویری هستند و هر دو روش نقاط ضعف و قوتی دارند. مزیت‌های هر دو روش عبارتند از سرعت ردیابی در روش جابجایی میانگین و دقت ردیابی در روش فیلتر ذره‌ای. در این پایان‌نامه قصد داریم به منظور استفاده از مزیت‌های این دو روش، تلفیقی از هر دو روش فیلتر ذره‌ای و جابجایی میانگین را پیشنهاد کنیم. این روش تلفیقی که با بکار بردن روش جابجایی میانگین در داخل فیلتر ذره‌ای باعث حرکت دادن ذره‌ها به سمت قله‌های محلی موجود در سطح احتمال می‌شود، امکان استفاده از ذره‌های کمتری برای ردیابی دست بصورت دقیق، مقاوم و سریع‌تر را فراهم می‌سازد زیرا جابجایی میانگین بکار رفته در فیلتر ذره‌ای بازده و کارایی نمونه برداری از سطح احتمال را افزایش می‌دهد. نتایج حاصل از ردیابی دست به روش تلفیقی نشان می‌دهد که این روش با استفاده از فقط ۲۰٪ تعداد ذره‌های مورد نیاز در فیلتر ذره‌ای و با دو برابر سرعت نسبت به فیلتر ذره‌ای می‌تواند دست را با همان دقت و مقاومت فیلتر ذره‌ای ردیابی کند. هدف ما در نهایت از ردیابی دست با استفاده از روش تلفیقی، تشخیص اشاره دست در زبان اشاره است که این مهم به کمک استخراج مسیر حرکت دست در رشته‌های تصویری مربوط به دیتابیس زبان اشاره مورد استفاده حاصل می‌شود.

کلید واژه‌ها: ردیابی دست، فیلتر ذره‌ای، جابجایی میانگین، روش تلفیقی، تشخیص اشاره دست، زبان اشاره.

فصل اول: مروری بر ردیابی دست و کارهای انجام شده..... ۱

۱-۱- مقدمه..... ۲

۱-۲- روشهای ردیابی دست و ضرورت آن..... ۳

۱-۳- معرفی سیستم پیشنهادی..... ۶

۱-۴- مختصری بر فصول بعدی..... ۸

فصل دوم: روش های ردیابی اهداف متحرک و انتخاب ویژگی ها در ردیابی..... ۱۰

۱-۲- مقدمه..... ۱۱

۱-۲-۲- روشهای ردیابی اهداف متحرک و دسته‌بندی آنها..... ۱۱

۱-۲-۲-۱- روشهای شار نوری..... ۱۳

۱-۲-۲-۲- ردیابی مبتنی بر مدل..... ۱۴

۱-۲-۲-۳- ردیابی مبتنی بر ناحیه..... ۱۵

۱-۲-۲-۴- ردیابی مبتنی بر کانتور..... ۱۶

۱-۲-۲-۵- ردیابی مبتنی بر ویژگی..... ۱۸

۱-۲-۲-۶- ردیابی با استفاده از دسته بندی کننده..... ۱۹

۱-۲-۳- استخراج مشخصه..... ۲۰

۱-۲-۴- شاخصه لبه..... ۲۰

۱-۲-۴-۱- آشکارسازی لبه..... ۲۱

۱-۲-۴-۲- پیوند لبه..... ۲۴

- ۲-۵- شاخصه رنگ ۲۶
- ۲-۶- توصیف انواع مدل‌های رنگ ۲۶
- ۲-۷- مقایسه برخی از فضای توصیفی ۲۷
- ۲-۸- شاخصه بافت ۲۸
- ۲-۸-۱- کاربردهای بافت ۳۰
- ۲-۸-۲- طبقه بندی مدل های بافت ۳۰
- ۲-۹- تبدیل موجک گسسته دو بعدی ۳۲
- ۲-۱۰- جمع بندی ۳۵
- فصل سوم : ردیابی شیء با استفاده از روش جابجایی میانگین ۳۶**
- ۳-۱- مقدمه ۳۷
- ۳-۲- ارائه هدف ۳۸
- ۳-۲-۱- مدل هدف ۴۱
- ۳-۲-۲- کاندیداهای هدف ۴۳
- ۳-۳- متریک بر اساس ضریب باتاچاریا ۴۴
- ۳-۴- تعیین جایگاه هدف ۴۶
- ۳-۴-۱- مینیمم سازی فاصله ۴۷
- ۳-۵- انتخاب نوع هسته ۴۹
- ۳-۵-۱- استفاده از تبدیل فاصله به عنوان هسته ۵۰
- ۳-۶- انتخاب فضای رنگ ۵۵
- ۳-۷- ویژگی های چند گانه ۵۸

- ۵۸.....۳-۷-۱- مدل سازی شاخصه رنگ.....
- ۵۹.....۳-۷-۲- مدل سازی شاخصه لبه.....
- ۶۰.....۳-۷-۳- مدل سازی شاخصه بافت.....
- ۶۱.....۳-۸-۸- آموزش آماری.....
- ۶۵.....۳-۸-۱- طبقه بندی خطی.....
- ۶۶.....۳-۹-۹- ماشینهای بردار پشتیبان.....
- ۶۶.....۳-۹-۱- حداقل کردن ریسک عملی (ERM).....
- ۶۷.....۳-۹-۲- حداقل کردن ریسک ساختاری (SRM).....
- ۶۸.....۳-۹-۳- ماشینهای بردار پشتیبان خطی در حالت دو کلاسی با الگوی جدایی پذیر.....
- ۶۹.....۳-۹-۴- ماشینهای بردار پشتیبان خطی در حالت دو کلاسی با الگوی جدایی ناپذیر.....
- ۷۰.....۳-۹-۵- ماشینهای بردار پشتیبان غیر خطی و مفهوم هسته.....
- ۷۱.....۳-۱۰-۱- ویژگی های استخراج شده برای آموزش طبقه بند ماشین بردار پشتیبان.....
- ۷۴.....۳-۱۱- آموزش و تست SVM برای آشکارسازی هدف در فریم جدید.....
- ۷۴.....۳-۱۲- الگوریتم تخمین مدل حرکت با استفاده از ویژگی های چند گانه.....
- ۷۵.....۳-۱۳- پیاده سازی الگوریتم جابجایی میانگین در ردیابی اشیاء.....
- ۷۷.....۳-۱۴- نتایج عملی.....
- ۸۴.....۳-۱۴-۱- تخمین مکان هدف با استفاده از ویژگی های متعدد.....
- ۸۶.....۳-۱۴-۲- جداسازی دقیق هدف از پس زمینه.....

۹۰	۱۵-۳- سیستم کلی روش جابجایی میانگین در ردیابی اشیاء
۹۱	۱۶-۳- مروری کلی بر روش جابجایی میانگین و کاربرد آن
۹۱	۱۷-۳- جمع بندی
۹۳	فصل چهارم : ردیابی دست با استفاده از روش پیشنهادی تلفیقی
۹۴	۱-۴- مقدمه
۹۴	۲-۴- الگوریتم تخمین مدل حرکت با استفاده از فیلتر ذره‌ای
۹۷	۳-۴- روش پیشنهادی تلفیقی (تلفیق جابجایی میانگین با فیلتر ذره‌ای)
۱۰۰	۴-۴- ردیابی دست با استفاده از روش پیشنهادی تلفیقی
۱۰۰	۱-۴-۴- مدل دینامیکی
۱۰۱	۲-۴-۴- مدل مشاهداتی
۱۰۵	۳-۴-۴- بهینه سازی ذره توسط روش جابجایی میانگین
۱۰۶	۴-۴-۴- تابع احتمال
	۵-۴- پیاده سازی الگوریتم روش تلفیقی پیشنهادی در تشخیص اشاره
۱۰۸	دست
	۶-۴- نتایج عملی الگوریتم‌های جابجایی میانگین ، فیلتر ذره‌ای و روش
۱۱۲	پیشنهادی تلفیقی
۱۱۳	۱-۶-۴- ردیابی دست
۱۱۳	۱-۱-۶-۴- مطالعات مقایساتی
۱۲۷	۲-۱-۶-۴- نتایج بیشتر ردیابی
۱۲۹	۷-۴- جمع بندی

فصل پنجم: نتیجه گیری و پیشنهادات برای ادامه کار.....	۱۳۰
۱-۵- مقدمه.....	۱۳۱
۲-۵- سیستم کلی روش ارائه شده.....	۱۳۱
۳-۵- جمع بندی و پیشنهادات.....	۱۳۲
مراجع.....	۱۳۴
فهرست مقالات ارائه شده.....	۱۴۰
واژه نامه.....	۱۴۱
پیوست: انواع مدل‌های رنگ.....	۱۴۳

فهرست اشکال

- شکل ۱-۱: بلوک دیاگرام رابط کنترل میانی انسان و ماشین (صندلی چرخدار و یا ماشین مترجم زبان اشاره) ۳
- شکل ۱-۲: طبقه بندی سیستم های ردیابی ۱۲
- شکل ۲-۲: آشکارسازی لبه با عملگرهای مشتق: (الف) نوار روشن روی زمینه تیره. (ب) نوار تیره روی زمینه روشن ۲۲
- شکل ۲-۳: (الف) یک ناحیه تصویری 3×3 (ب) ماسک مورد استفاده برای محاسبه $\frac{\partial I}{\partial x}$ (پ) ماسک مورد استفاده برای محاسبه $\frac{\partial I}{\partial y}$ (عملگرهای سوپل) ۲۳
- شکل ۲-۴: (الف) یک تصویر نمونه (ب) تصویر دو سطحی شده حاصل از اعمال ماسک های سوپل بر روی تصویر "الف". (ج) تصویر مرز نواحی حاصل از پیوند لبه ۲۵
- شکل ۲-۵: خوشه بندی فضای مکعبی جهت مقایسه انواع مدلها ۲۷
- شکل ۲-۶: مقایسه ۴ مدل توصیفی: (الف) مدل XYZ، (ب) مدل YCbCr، (ج) مدل YUV و (د) مدل RGB ۲۸
- شکل ۲-۷: مراحل تجزیه موجک دو بعدی در مطلب ۳۳
- شکل ۲-۸: موجک در سطوح مختلف. اندیس ۱ و ۲ به ترتیب نشان دهنده سطح ۱ و ۲ است. ۳۴
- شکل ۲-۹: ده کانال بدست آمده بعد از اعمال سه مرتبه تبدیل موجک دو بعدی ۳۴
- شکل ۳-۱: بیضی هدف در فریم اول (تصویر مرجع) ۴۱
- شکل ۳-۲: بیضی کاندیدای هدف در فریم دوم از تصویر شکل ۳-۱ ۴۴

- شکل ۳-۳: رویه شباهت بین هدف شکل ۳-۱ با کاندیدای هدف در فریم دوم حول ناحیه مربعی شکل با ابعاد 40×40 در مرکز هدف ۴۶
- شکل ۳-۴: یک تصویر باینری تعریف شده روی شبکه G ۵۱
- شکل ۳-۵: خروجی جسم پس از اسکن اول. (الف) برای همسایگی هشت تایی (ب) برای همسایگی چهارتایی ۵۳
- شکل ۳-۶: خروجی جسم پس از اسکن دوم. (الف) برای همسایگی هشت تایی (ب) برای همسایگی چهارتایی ۵۴
- شکل ۳-۷: رویه شباهت بین هدف شکل ۳-۱ با کاندیدای هدف در فریم دوم حول ناحیه مربعی شکل با ابعاد 40×40 در مرکز هدف با استفاده از تبدیل فاصله ۵۴
- شکل ۳-۸: بیضی با رنگ زرد با ابعاد $\sqrt{2}$ برابر بیضی سیاه رنگ. ناحیه داخل منحنی قرمز رنگ مشخص کننده هدف و ناحیه بین بیضی با رنگ زرد و منحنی قرمز رنگ مشخص کننده ناحیه غیر هدف می باشد ۶۲
- شکل ۳-۹: حاشیه تصمیم برای یک کلاس بندی کننده جدایی پذیر خطی ۶۹
- شکل ۳-۱۰: روش استفاده شده برای بدست آوردن ضرایب موجک ۷۳
- شکل ۳-۱۱: دنباله تصویر عابر پیاده، ردیابی شخص در فریم های ۶۰، ۱۲۰، ۲۴۰، ۳۰۰، ۳۴۰ ۷۸
- شکل ۳-۱۲: تعداد تکرارهای الگوریتم جابجایی میانگین در هر فریم از دنباله تصویر عابر پیاده ۷۹
- شکل ۳-۱۳: مقدار ضریب باتاچاریا بین مدل هدف در فریم اول و کاندیدای نهایی هدف در هر فریم از دنباله تصویر عابر پیاده ۸۰
- شکل ۳-۱۴: دنباله تصویر ماشین، ردیابی ماشین در فریم های ۳۰، ۶۰، ۹۰ ۸۱

- شکل ۳-۱۵: تعداد تکرارهای الگوریتم جابجایی میانگین در هر فریم از دنباله تصویر ماشین ۸۲
- شکل ۳-۱۶: مقدار ضریب باتاچاریا بین مدل هدف در فریم اول و کاندیدای نهایی هدف در هر فریم از دنباله تصویر ماشین ۸۳
- شکل ۳-۱۷: خطای تخمین مکان هدف در هر فریم از دنباله تصویر ماشین ۸۳
- شکل ۳-۱۸: ردیابی دست با استفاده از ویژگی های لبه و رنگ به صورت مستقل و به صورت توام ۸۴
- شکل ۳-۱۹: تعداد تکرارهای الگوریتم جابجایی میانگین در هر فریم از دنباله تصویر دست برای ویژگی رنگ و لبه به صورت مجزا و توام ۸۵
- شکل ۳-۲۰: ردیابی چیتا با استفاده از ویژگی های رنگ و بافت به صورت مستقل و به صورت توام ۸۶
- شکل ۳-۲۱: تصویر مرجع در فریم اول. ناحیه داخل بیضی سیاه رنگ به عنوان هدف در فریم اول در نظر گرفته می شود. ۸۷
- شکل ۳-۲۲: یک مثال از ردیابی ماشین با استفاده از الگوریتم پیشنهادی. الف- ۵ فریم از دنباله تصویر ۲۹۰ فریمی نشان داده شده است. ب- خروجی SVM بعد از اعمال بر روی ناحیه هدف داخل بیضی زرد رنگ. ج- تصویر خروجی پس از اعمال یک سری عملیات شکل شناسی. ۸۸
- شکل ۳-۲۳: یک مثال از ردیابی ماشین با استفاده از الگوریتم پیشنهادی. الف- ۴ فریم از دنباله تصویر ۲۲۲ فریمی نشان داده شده است. ب- خروجی SVM بعد از اعمال بر روی ناحیه هدف داخل بیضی زرد رنگ. ۸۹
- شکل ۳-۲۴: بلوک دیاگرام عملکرد کلی سیستم ۹۰
- شکل ۴-۱- الگوریتم فیلتر ذره ای ۹۵

- شکل ۴-۲- نمایش شماتیکی از الگوریتم فیلتر ذره ای..... ۹۶
- شکل ۴-۳- الگوریتم روش تلفیقی پیشنهادی ۹۹
- شکل ۴-۴- نمایش گرافیکی جابجایی میانگین بکار رفته در روش فیلتر ذره ای به عنوان روش تلفیقی پیشنهادی..... ۱۰۰
- شکل ۴-۵- a فریم ۸۹، b فریم ۹۰، c تصویر توزیع احتمال رنگ فریم ۹۰، d تصویر توزیع رنگ-حرکت فریم ۹۰ ۱۰۴
- شکل ۴-۶- a فریم ۸، b فریم ۹، c تصویر توزیع احتمال رنگ..... ۱۰۴
- شکل ۴-۷- منحنی مشاهداتی توزیع چند مد در دنیای واقعی. سمت راست نقشه رویه مشاهداتی محیط مربع سفید تصویر سمت چپ است..... ۱۰۷
- شکل ۴-۸- منحنی مشاهداتی توزیع چند مد در تصویر دست. سمت راست نقشه رویه مشاهداتی محیط برای مربع سیاه تصویر سمت چپ است..... ۱۰۷
- شکل ۴-۹- نمونه های اشاره حرکت موجی، ساعتگرد و حرکت افقی..... ۱۱۰
- شکل ۴-۱۰- نمونه هایی از زبان اشاره با حرکت دست از بالا به پایین..... ۱۱۰
- شکل ۴-۱۱- تولید TTBT برای اشاره پادساعتگرد..... ۱۱۱
- شکل ۴-۱۲- تولید TTBT برای اشاره دست در زبان اشاره..... ۱۱۱
- شکل ۴-۱۳- TTBT برای پنج اشاره از پیش تعریف شده. از چپ به راست: حرکت موجی، پادساعتگرد، ساعتگرد، افقی، عمودی..... ۱۱۱
- جدول ۴-۱- نتایج ردیابی دست به روش تلفیقی پیشنهادی برای ۵۰ ذره و فیلتر ذره ای برای ۲۰۰ ذره و جابجایی میانگین..... ۱۱۴
- شکل ۴-۱۴- الف- بعضی از نتایج ردیابی دست به روش های جابجایی میانگین، فیلتر ذره ای و روش تلفیقی پیشنهادی برای یک سکانس آزمایشی از فریم ۶۱ تا ۶۴..... ۱۱۵

شکل ۴-۱۴-ب- مقایسه دقت ردیابی برای یک سکانس آزمایشی. چپ: مختصات x و راست: مختصات y. ۱۱۵.....

جدول ۴-۲- نتایج ردیابی دست به روش های جابجایی میانگین و فیلتر ذره ای و روش تلفیقی. ۱۱۶.....

شکل ۴-۱۵- الف- بعضی از نتایج ردیابی دست به روش های جابجایی میانگین , فیلتر ذره ای و روش تلفیقی پیشنهادی برای یک سکانس آزمایشی زبان اشاره از فریم ۱۷ تا ۲۷ . ستون راست: روش تلفیقی پیشنهادی , ستون وسط: فیلتر ذره ای , ستون چپ: جابجایی میانگین. ۱۱۷.....

شکل ۴-۱۵-ب- مقایسه دقت ردیابی دست برای سکانس آزمایشی بالا به روش جابجایی میانگین MS و فیلتر ذره ای PF و روش تلفیقی CMP. ۱۱۸.....

جدول ۴-۳- نتایج ردیابی دست به روش های جابجایی میانگین و فیلتر ذره ای و روش تلفیقی. ۱۱۸.....

شکل ۴-۱۶- الف- بعضی از نتایج ردیابی دست به روش های جابجایی میانگین , فیلتر ذره ای و روش تلفیقی پیشنهادی برای یک سکانس آزمایشی زبان اشاره بر از فریم ۲۱ تا ۳۵. ۱۲۰.....

شکل ۴-۱۶-ب- مقایسه دقت ردیابی الگوریتم های جابجایی میانگین MS و فیلتر ذره ای PF و روش تلفیقی CMP در هر فریم از دنباله تصویر دست در زبان اشاره برای ویژگی رنگ برای سکانس آزمایشی بالا. ۱۲۰.....

جدول ۴-۴- نتایج ردیابی دست به روش های جابجایی میانگین و فیلتر ذره ای و روش تلفیقی. ۱۲۱.....

شکل ۴-۱۷- الف- بعضی از نتایج ردیابی دست به روش های جابجایی میانگین و فیلتر ذره ای و روش تلفیقی پیشنهادی برای یک سکانس آزمایشی زبان اشاره از فریم. ۱۲۳.....

- شکل ۴-۱۷-ب- مقایسه دقت ردیابی الگوریتم های جابجایی میانگین MS و فیلتر ذره ای PF و روش تلفیقی CMP در هر فریم از دنباله تصویر دست در زبان اشاره برای ویژگی رنگ برای سکانس آزمایشی بالا. ۱۲۳
- جدول ۴-۵- نتایج ردیابی دست به روش های جابجایی میانگین و فیلتر ذره ای و روش تلفیقی ۱۲۴
- شکل ۴-۱۸- مقایسه خطای ردیابی دست با توجه به تعداد ذره ها. چپ: روش تلفیقی و راست: روش فیلتر ذره ای ۱۲۵
- شکل ۴-۱۹- مقایسه بین مدل های مشاهدات برای یک سکانس آزمایشی (از فریم ۵ تا ۸). ردیف بالا: مدل مشاهدات بر اساس نشانه های رنگ و حرکت. ردیف پایین: مدل مشاهدات بر اساس فقط نشانه های رنگ. ۱۲۷
- شکل ۴-۲۰- نتایج ردیابی دو سکانس با انحراف رنگ پوست. ردیف بالا: فریمهای ۴۹، ۴۷، ۴۵ و ۵۰. ردیف پایین: فریمهای ۲۲، ۲۱، ۱۹ و ۲۴. ۱۲۸
- شکل ۴-۲۱- نتایج ردیابی یک سکانس با تغییرات روشنایی (فریمهای ۱۷۶، ۱۰۳، ۱۷۷ و ۲۲۹) ۱۲۸
- شکل ۴-۲۲- نتایج ردیابی یک سکانس با حرکت سریع دست (فریمهای ۷۱ تا ۷۴). ۱۲۸

فصل اول

مروری بر ردیابی دست و کارهای

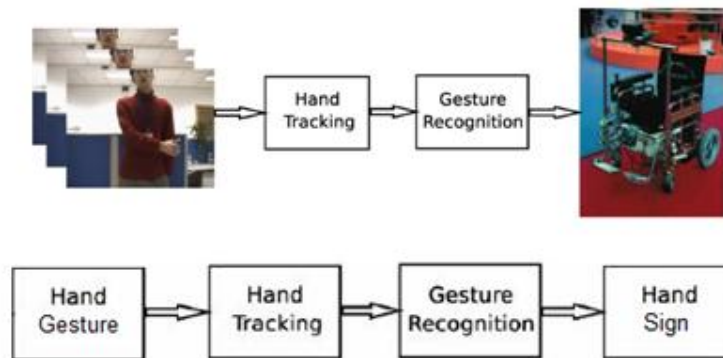
انجام شده

۱-۱- مقدمه

از آنجایی که کامپیوترها در زندگی انسانها فراگیر و نافذ شده اند، مردم احتیاج بسیاری به ارتباط موثر و طبیعی کامپیوتر و انسان^۱ دارند. در حال حاضر محبوبترین و غالبترین حالت HCI هنوز بر پایه وسایلی از قبیل کیبورد و موس است که غیرطبیعی و غیر راحت برای انسانها هستند و همچنین سرعت ارتباط را محدود می کنند. اگر کامپیوترها بتوانند حرکات انسان را گرفته و بفهمند، ارتباط دوستانه تری حاصل می شود. آنالیز بینایی حرکات انسان تحولی شگرف در HCI به پا می کند [۱]. اشاره دست یا حرکات ارادی دستها و بازوها معنی دار هستند. به عنوان یک زبان بدنی جهانی انسانها، اشاره دست یکی از مهم ترین، طبیعی ترین و موثرترین وسایل برای انسانها، برای ارتباط برقرار کردن غیر زبانی و حرفی با سایرین است.

توانایی تشخیص اشارات دست غیر قابل انصراف و بسیار مهم برای یک ارتباط اجتماعی بین فردی موفق است. تشخیص اشاره دست بر پایه بینایی [۲]، که کامپیوترها را قادر به فهم اشارات دست مانند انسان می سازد، یک تکنولوژی مهم برای HCI هوشمند است (شکل ۱-۱). در نتیجه آنالیز بینایی اشارات دست توجهات بسیاری را در این دهه به خود جلب نموده و پیشرفت فراوانی کرده است [۲].

^۱ Human- Computer Interaction



شکل ۱-۱. بلوک دیاگرام رابط کنترلی میان انسان و ماشین (صندلی چرخدار و یا ماشین مترجم زبان اشاره)

۱-۲- روشهای ردیابی دست و ضرورت آن

ردیابی دست به منظور تخمین و محاسبه حرکت پیوسته و غیر منقطع دست در سکانسهای تصویری که همان فریم های مختلف یک فیلم است، یک قدم بسیار چالش آمیز و در عین حال فوق العاده مهم برای تشخیص اشاره دست است. بخاطر همین نقش خطیرش در بینایی ماشین و مشکلات مربوط به آن که مربوط به حرکت دست می شود، ردیابی دست علاقه زیادی را در سالهای اخیر به وجود آورده و افراد زیادی را متوجه خود کرده است [۱۷-۳].

حرکات دست در فضای سه بعدی و در سکانسهای تصویری دوبعدی انجام شده است. پس ردیابی دست می تواند در صفحات و فضای دو بعدی (2d) یا سه بعدی (3d) پیاده شود. روشهای موجود در ردیابی دست می تواند بصورت اصلی با روشهای دو بعدی یا سه بعدی کلاسه بندی شوند. در روشهای دو بعدی، یک دست بوسیله ویژگیهای هندسی آن نمایش داده می شود نظیر کانتورها [۱۸-۳] و برآمدگی انگشتان و نوک انگشتان [۴]، یا ویژگیهای غیرهندسی آن نظیر رنگ [۵]. در روش Isard&Blake [۵۸] منحنی های B-spline بصورت پارامتریک برای مدل کردن کانتورهای دست گرفته شده اند و دستها بوسیله دنبال کردن و ردیابی منحنی های تغییر شکل داده شده ردیابی شده اند. به هر حال از آنجایی که کانتورهای دست به بینایی وابستگی داشته و بصورت دراماتیک و نمایشی

دراثر حرکت طبیعی دست تغییر می کنند، دنبال کننده ها یا ردیابهای بر پایه کانتور معمولاً فضای دید و دیدگاه را محدود کرده و فرض می کنند که دستها چندین شکل از پیش تعیین شده را حفظ می کنند. برای مثال [۳] Mcalister یک سیستم ردیابی دو دستی بر اساس کانتور ارائه داد.

نوک انگشتان ویژگیهای هندسی موثر دیگری برای ردیابی هستند. برای نمونه [۴] Oka et al. پیشنهاد داد که از دنبال کردن نوک انگشت های چند تایی برای تشخیص اشاره دست استفاده شود. رنگ پوست یک ویژگی متمایز دستهاست اگر قسمتهای دیگر بدن معلوم نباشند یا در پیش زمینه اشیاء هم رنگ پوست نباشند. خیلی از ردیابهای رنگی که برای ردیابی حرکات دست استفاده می شوند [۵،۱۹] از این ایده نشأت می گیرند که انواع مختلفی از ویژگیهای تصویری بر اساس مقیاس مشاهده قابل دیدن هستند.

هر کدام از ویژگیها به تنهایی امتیازات و محدودیتهایی دارد. برای مثال رنگ پوست بخاطر تغییر روشنایی قابل اتکا نیست. به جهت چیره شدن بر این ضعف ویژگی رنگ، الگوریتمهای پیشنهادی بسیاری برای استفاده از نشانه های دیگر و یکی کردن نشانه های چند تایی برای ردیابی دست بصورت قوی پیشنهاد شده اند [۶،۱۲،۱۳،۲۰،۲۱]. در مرجع [۲۰] Isard & Blake، یک روش پیاده کردن و گسترش ردیاب کانتور بوسیله ردیابی نقطه ای برای ردیابی دست مطرح شده است. در مرجع [۶] Hung & Reid یک فیلتر Bayesian مشترک را برای ترکیب فیلتر ذره ای و مدل مخفی مارکوف (HMM) که برای ردیابی و تشخیص حرکات بیانگرایانه ی حرکت دست استفاده شده ابداع کرده اند، رنگ و شکل به عنوان وسیله نشان دادن دست استفاده شده بودند و فیلتر ذره ای برای ردیابی محیط رنگی انتخاب شده و همچنین برای کمک به مدل مخفی مارکوف در تحلیل تغییرات شکل دست بکار برده شده است. [۲] Kolch & Turk اخیراً پیشنهاد داده اند که با ترکیب شار نوری و نقاط رنگی با سایر ویژگیها می توان سرعت ردیابی دست را بهبود بخشید.

ردیاب دوبعدی دست برای کاربردهای همزمان بصورت محاسباتی موثر است. بنابراین کاربردهای زیادی بر اساس روشهای دو بعدی موجود هستند [۲۲]. در هر حال، ردیابی های دوبعدی معمولاً فقط می توانند حرکت کلی دست را ردیابی و نمی توانند حرکت بند به بند و تشریحی انگشتان را مشخص کنند. در حالت دیگر و برخلاف آن، در ردیابی دست به صورت سه بعدی، ردیاب ها می توانند دستها را در فضای سه بعدی مکان یابی کرده، مکان سه بعدی را تعیین کنند و اطلاعات گرایشات و جهت گیریها را بصورت مشخص و صریح افشا کنند. اخیراً ردیابی سه بعدی بر پایه معادله سه بعدی یک گستره تحقیقاتی فعال و در حال رشد است [۷، ۱۱، ۱۵، ۱۷، ۲۳]. این رشته همچنین توانایی مقابله و رویارویی با مسدود شدن و خود مسدود کنی را داشته و می تواند اطلاعات حرکت دقیق و ریز شده به همراه جزئیات را بدست آورد که بتواند در شکلها و فرمهای گوناگونی استفاده شود. در ردیابی بر اساس مدل، معمولاً موقعیت یک دست بوسیله انعکاس مدل دست سه بعدی از قبل ذخیره شده به صفحه تصویر و مقایسه آن با ویژگیهای تصویر بدست آورده می شود. در مرجع [۷] یک روش بر اساس مدل را نشان داده است که از یکی کردن نشانه های چندتایی مانند لبه ها، شار نوری و اطلاعات سایه، برای ردیابی حرکت واضح و بند به بند دست استفاده شده است. در مرجع [۱۵] یک روشی برای ردیابی بر اساس مدل پیشنهاد داده است که هر دو اطلاعات انتقال حرکت متوالی و اطلاعات ظاهری را یکی می کند.

در مرجع [۸] مفهوم متحرکهای ویژه و خودی برای مدل کردن دینامیک های حرکت طبیعی دست پیشنهاد شده است. حرکت دست به عنوان یک سیستم متحرک دینامیک خطی تصادفی با درجه بالا (LSD) مدل می شود که شامل پنج LDS با درجه پایین بوده و هر کدام متناظر با یک محرک خودی است. همچنین اطلاعات سه بعدی بدست آمده بوسیله دوربینهای استریو واسکنرها را سه بعدی ردیابی کرد. اگرچه ردیابی سه بعدی نتایج دقیقتری نسبت به ردیابی دو بعدی فراهم می کند ولی روش های سه بعدی هزینه محاسباتی سنگینی را به دنبال دارند.

۱-۳- معرفی سیستم پیشنهادی

در ادامه این گزارش ما یک روش کارآمد برای ردیابی دست ارائه می دهیم. با در نظر گرفتن سادگی و اجرای عملی آن، ما ردیابی دست را در صفحه دو بعدی به انجام می رسانیم. اگر ردیابی بصورت مستقیم برای اشاره دست اعمال شود، آنگاه با دو مشکل اصلی مواجه می شود:

۱- پیش زمینه ممکن است متغیر باشد و با اشیاء مزاحم نویزی شده است.

۲- همچنین روشنایی بصورت قابل توجهی تغییر می کند.

پس ردیابی دست باید در محیط ها و موقعیت های واقعی به اندازه کافی قوی بوده و همچنین موثر و کم هزینه باشد.

وجود اشیاء مزاحم در پیش زمینه، حرکت پیچیده دست و تغییرات روشنایی باعث می شوند که ردیابی دست با مشکل مواجه شود. فیلتر ذره ای که به عنوان الگوریتم تراکم هم شناخته شده است برای حل مشکل مکان یابی یک مد در رویه ی چند مد می باشد. بنابراین از فیلتر ذره ای به عنوان چهارچوب ردیابی ما در این گزارش، بهره گرفته شده است. توجه به این نکته ضروری است که در فیلتر ذره ای، جهت تنظیم قدرت ذره ها، نمونه برداری کافی برای گرفتن تغییرات در فضای حالت صورت گیرد. نمونه برداری فشرده یا چگال، بار محاسباتی سنگینی را به همراه می آورد، که با گرایش به سمت کم هزینه‌گی در تضاد است، در نتیجه نیاز است مدلی ابداع شود که با تعداد ذره های کمتری همان موفقیت را تا حدودی بدست آورد که در این صورت ما موفق به کاهش هزینه محاسبات هم شده ایم. بعضی تکنیکها برای بهبود بازدهی نمونه برداری از فیلتر ذره ای ارائه شده اند [۲۰، ۲۵، ۳۰]. توسط انتخاب توزیع پیشنهادی بهتر کاهش تعداد ذره ها امکانپذیر است [۲۵]. اهمیت نمونه برداری [۲۰] معرفی شده برای بدست آوردن پیشنهاد و روش بهتر، توسط ترکیب پیش بینی بر اساس چیدمان قبلی با دانش اضافی از محاسبات کمکی است. به منظور ردیابی بیش از یک شیء، نمونه برداری بخش بندی شده [۲۶] برای کاهش تعداد ذره های لازم

ارائه شده است. این روش بر اساس یک فرضیه گفته شده است که فضای پیکره بندی شده آن می تواند به چندین قسمت تقسیم شود. راه دیگر برای نمونه برداری بهتر، معرفی یک روش بهینه سازی جدید است که پروسه آن به این ترتیب است که ذره ها و نمونه های انتخابی ما را به قله سطح چگالی یا دانسیته انتقال می دهد [۲۸،۲۷]. در طی پروسه بهینه سازی، ذره ها از توزیع قبلی پیروی نمی کنند، بنابراین این روشها تخمین خوبی برای توزیع پیشین مورد علاقه ما نیستند [۲۹ و ۳۰].

با تلفیق یک روش موثر غیر پارامتری در فیلتر ذره ای، ما یک روش را برای بهبود کارآمدی نمونه برداری فیلتر ذره ای ارائه می دهیم. روش جابجایی میانگین یک روش بهینه سازی بر اساس گرادیان است و بصورت موفقیت آمیزی برای ردیابی همزمان استفاده شده است [۳۲]. از آنجایی که ردیابهای این روش بهترین حالت کلی را نمی توانند تضمین کنند، در نتیجه به راحتی در بیشینه های محلی گیر می افتند. لذا پیشنهاد می کنیم که برای نمونه برداری بهتر، ذره ها را در فیلتر ذره ای به روش جابجایی میانگین به سمت قله مورد نظر حرکت دهیم. روش جابجایی میانگین استفاده شده در فیلتر ذره ای، ایده استفاده از روش تلفیقی را می دهد که ما در آن از تعداد ذره های کمتری برای پیدا کردن یک مد در میان چند مد بر روی یک رویه ی چند مد استفاده می کنیم. ما با استفاده از روش تلفیقی، ردیابی دست را بصورت دقیق و قابل اجرا در زمان واقعی به انجام می رسانیم. روش ردیابی پیشنهاد شده یک روش کلی است که می تواند برای مسائل ردیابی استفاده شده و همچنین روشی ساده برای تشخیص اشاره دینامیکی است. تشخیص اشاره به صورت آنلاین بخوبی عمل می کند و عملکرد بسیار بالایی را از الگوریتم ارائه شده در محیط و پیش زمینه دینامیک نشان می دهد.

مفهوم اصلی در فرمول بندی و توجیه کردن مدل ردیابی بهبود یافته، آزمایشات بیشتر و تشخیص اشاره دست قرار می گیرد. پیامدهای اصلی این تحقیق به صورت زیر خلاصه می شوند:

بوسیله بکار بستن جابجایی میانگین در فیلتر ذره ای، به منظور حرکت دادن ذره ها به سمت مدهای محلی، ما روش تلفیقی را به منظور ردیابی تصویری آورده‌ایم، زیرا ذره های مورد نیاز را به شدت کاهش می دهند. توسط این روش تلفیقی پیشنهادی، ما ردیابی تصویری بصورت همزمان می سازیم و همزمان می توانیم آنرا محقق سازیم. از طریق اتصال مشخصه های رنگ و حرکت، مدل مشاهده ای برای جابجایی میانگین و تابع احتمال و شباهت ارائه می شود.

۱-۴- مختصری بر فصول بعدی

در فصل بعد (فصل دو) "ردیابی اهداف متحرک و انتخاب ویژگی ها" مبحث ردیابی اهداف گوناگون به روش های اولیه‌ی پیشین معرفی و مطرح می شود و مبحث ردیابی اشیاء بصورت مختصر توضیح داده می شود.

در فصل سه ردیابی شیء به روش جابجایی میانگین و نحوه استفاده از این روش سریع برای تخمین مکان شیء را در هر فریم بصورت کامل بیان می کنیم. در این فصل همچنین استفاده از یک تبدیل فاصله^۱ برای ارائه بهتر هدف (شیء) در هر فریم را معرفی می کنیم.

در فصل چهار روش جابجایی میانگین را با فیلتر ذره ای تلفیق می کنیم و الگوریتم تلفیقی حاصل را برای ردیابی دست در زبان اشاره پیشنهاد می کنیم. سپس به منظور مشخص کردن بازدهی روش تلفیقی در مقایسه با دو روش جابجایی میانگین و فیلتر ذره ای، نتایج عملی هر سه روش را در ردیابی دست در رشته های تصویری مورد استفاده مربوط به زبان اشاره (RWTH-Boston) با هم مقایسه می کنیم. سپس در انتهای این فصل، نشان می دهیم که چگونه با استفاده از روش پیشنهادی می توان اشارات دست در این رشته های تصویری مربوط به زبان اشاره را بدست آورد (روشی را برای استخراج اشاره دست بیان می کنیم). در فصل پنجم این پایان نامه در نهایت نتیجه کار خود را بصورت مختصر بیان کرده

^۱ Distance Transform

و پیشنهادات خود را برای ادامه بهینه سازی این روش ردیابی دست ارائه می‌دهیم.

فصل دوم

روش های ردیابی اهداف متحرک و
انتخاب ویژگی ها در ردیابی

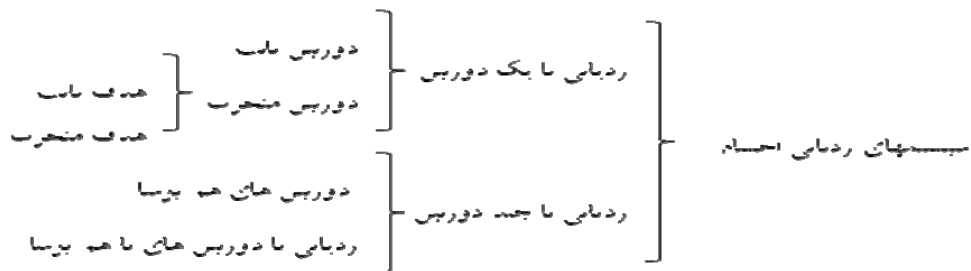
۲-۱- مقدمه

تخمین مسیر اجسامی که در میدان دید یک دوربین حرکت می کنند تحت نام ردیابی شناخته شده و به یکی از مهمترین شاخصه های علم بینایی ماشین تبدیل شده است. به عبارت دیگر می توان ردیابی را نوعی ایجاد روابط منسجم بین ویژگیهای اشیاء تقطیع شده ی تصویر مثل موقعیت، سرعت، شکل، بافت، رنگ و غیره در فریم های متوالی دانست.

سامانه های ردیابی را براساس تعداد دوربین های بکار رفته در آن می توان به چند دسته تقسیم کرد که در قسمت بعدی توضیح داده شده است. در اکثر سامانه های قدیمی ردیابی توسط یک دوربین صورت می گرفت که این دوربین تنها منطقه محدودی را تحت پوشش قرار می داد. اما برای طراحی یک سامانه نظارت خودکار تصویری نمی توان تمام منطقه نظارتی را با یک دوربین تحت پوشش قرار داد. بنابراین یک سامانه نظارتی عملی و کاربردی که بتواند اشیاء و افراد را در یک منطقه نظارتی وسیع ردیابی کند به چندین دوربین نیاز دارد. بعلاوه در اکثر مواقع نمی توان تمام منطقه نظارتی را توسط دوربینهای همپوشا تحت پوشش قرار داد زیرا این امر نیازمند استفاده از تعداد زیادی از دوربینها در یک شبکه است که هم به لحاظ اقتصادی و هم به لحاظ بار محاسباتی بسیار سنگین خواهد بود. در ادامه این بخش ابتدا به انواع روش های ردیابی پرداخته و سپس مثال هایی از ردیابی با یک دوربین را ارائه می کنیم.

۲-۲- روشهای ردیابی اهداف متحرک و دسته بندی آنها

سامانه های ردیابی را براساس تعداد دوربین های بکار رفته در آن می توان به چند دسته تقسیم کرد که در شکل (۲-۱) آمده است. لازم به ذکر است که در محیط های واقعی برای ردیابی اشیاء و افراد در یک منطقه نظارتی از چیدمان دوربین های ناهمپوشا استفاده می شود. با توجه به مطالب بیان شده انواع سامانه های ردیابی براساس تعداد و نحوه چینش دوربین ها نسبت به یکدیگر در یک طبقه بندی کلی بصورت شکل (۲-۱) ارائه می شود.



شکل ۲-۱: طبقه بندی سیستم های ردیابی

در یک منظره‌ای که توسط دوربین دیجیتال تصویربرداری می‌گردد، حرکت اشیا و همچنین حرکت خود دوربین باعث تغییرات شدت روشنایی در نقاط مختلف تصاویر می‌گردد. براساس این تغییرات الگوریتم‌های ردیابی هدف متحرک، حرکت واقعی هدف یا حرکت دوربین را تخمین می‌زنند. در ردیابی با یک دوربین در مواردی که دوربین ثابت فرض گردد، استخراج اطلاعات حرکتی هدف که هم می‌تواند برای آشکارسازی و هم برای ردیابی مورد استفاده قرار گیرد، ساده بوده و می‌تواند با استفاده از روش تفاضل دو تک تصویر متوالی صورت پذیرد. اما در حالت کلی‌تر، در مواردی که هم دوربین و هم هدف دارای حرکت باشند پیچیدگی الگوریتم‌ها افزایش می‌یابد.

با توجه به کاربردهای زیاد آشکارسازی و ردیابی هدف متحرک، روش‌های گوناگونی برای این منظور معرفی شده است با این وجود نمی‌توان روشی را پیدا کرد که برای همه حالات و همه حرکت‌ها جوابگو باشد. زیرا هر یک از این تکنیک‌ها و روش‌ها براساس فرضیاتی استوار شده‌اند و برای محیط‌هایی که این فرضیات صدق می‌کنند، کاربرد دارند. اطلاعات صحنه، نحوه حرکت دوربین و سرعت حرکت آن و نرخ تک تصویرها در ثانیه، اطلاعاتی هستند که می‌توانند در کاهش پیچیدگی‌های ردیابی هدف متحرک و انتخاب نوع الگوریتم مفید باشند.

الگوریتم‌های ردیابی در حین پردازش معمولاً فصل مشترک زیادی با بخش آشکارسازی حرکت دارند. ردیابی معمولاً شامل تطبیق اشیا تقطیع شده در فریم‌های متوالی با استفاده از ویژگی‌هایی از قبیل نقاط، خطوط یا لکه‌ها است. ابزارهای ریاضی مفیدی نیز برای ردیابی اشیا وجود دارند که در میان آنها می‌توان

از مدل مخفی مارکوف^۱، فیلتر کالمن^۲، الگوریتم تراکم^۳ و شبکه بییزی پویا^۴ نام برد [۳۳]. روشهای ردیابی را می توان در شش دسته اصلی طبقه بندی کرد که عبارتند از: روشهای شار نوری^۵، ردیابی مبتنی بر مدل^۶، ردیابی مبتنی بر ناحیه^۷، ردیابی مبتنی بر کانتور فعال^۸، ردیابی مبتنی بر ویژگی^۹، ردیابی با استفاده از طبقه بندی کننده ها^{۱۰}. البته باید توجه داشت که این طبقه بندی تمامی روشها را نمی پوشاند و روشهایی وجود دارند که از ترکیب روشهای فوق ایجاد شده اند [۳۳].

۲-۲-۱- روشهای شار نوری

هدف روشهای شار نوری تعیین بردار حرکت ظاهری نواحی مختلف یک تصویر می باشد. به کارگیری روشهای شار نوری شامل اجرای دو مرحله می باشد [۳۴].

• در مرحله اول با استفاده از اطلاعات نقاط مجاور، بردارهای حرکت برای نقاط مختلف تصویر محاسبه می گردد. در این مرحله بردارهای حرکت محلی نقاط تصویر به حرکت کلی اشیاء ربط داده نمی شود.

^۱ Hidden Markov Model

^۲ Kalman filter

^۳ Condensation algorithm

^۴ Dynamic Bayesian network

^۵ Optical Flow

^۶ Model-based tracking

^۷ Region-based tracking

^۸ Active-contour-based tracking

^۹ Feature-based tracking

^{۱۰} Classifier-based tracking

• بعد از اینکه بردارهای حرکت برای تک تک نقاط تصویر محاسبه گردید، با استفاده از گسستگی‌های موجود در بردارهای حرکت، تصویر به ناحیه‌های مختلفی که دارای حرکت‌های متفاوت هستند، تقسیم‌بندی می‌گردد که معمولاً کار پیچیده‌ای است.

با توجه به اینکه بردار حرکت یک نقطه تصویر یک بردار دو بعدی بوده و شامل دو مؤلفه می‌باشد، نمی‌توان تنها از اطلاعات تغییر شدت روشنایی آن را بدست آورد. بنابراین محدودیت‌های دیگری نیز لازم است. برحسب اینکه چه محدودیت‌های دیگر قائل شویم، روش‌های متفاوتی پیشنهاد شده است.

۲-۲-۲- ردیابی مبتنی بر مدل

الگوریتم‌های ردیابی مبتنی بر مدل، ردیابی اشیاء را با تطبیق مدل‌های آنها (که با استفاده از دانش قبلی بدست آمده‌اند) با داده‌های تصویر انجام می‌دهند. این مدل‌ها معمولاً بصورت غیربرخط و با اندازه‌گیری دستی و یا روش‌های بینایی ماشین بدست می‌آیند.

با توجه به اینکه مدل‌های مربوط به اشیاء صلب^۱ (مثل وسایل نقلیه) از مدل‌های مربوط به اشیاء مفصل‌دار^۲ (مثل انسان) بسیار متفاوت هستند لذا ما تنها به بررسی مدل‌های اشیاء مفصل‌دار می‌پردازیم.

روش کلی در ردیابی انسان با استفاده از مدل با نام تحلیل با ترکیب^۳ شناخته می‌شود که به صورت پیشگویی-تطبیق-به‌روزرسانی^۴ بکار برده می‌شود. در ابتدا وضعیت مدل برای فریم بعدی با استفاده از دانش قبلی و سابقه ردیابی پیش‌بینی می‌شود. سپس مدل پیش‌بینی شده، ترکیب شده و برای مقایسه با داده تصویر بر

^۱ Rigid

^۲ Articulated

^۳ Analysis-by-synthesis

^۴ Predict-match-update

روی صفحه تصویر افکنده^۱ می شود. سپس برای اندازه گیری شباهت بین مدل افکنده شده و داده تصویر از یک تابع ارزیابی وضعیت استفاده می شود. در انتها نیز با استفاده از روشهای جستجو، وضعیت صحیح بدست آمده و برای به روزرسانی مدل از آن استفاده می شود [۳۳].

۲-۲-۳- ردیابی مبتنی بر ناحیه

الگوریتمهای ردیابی مبتنی بر ناحیه، اشیاء را با استفاده از تغییرات نواحی تصویر که متناظر با اشیاء متحرک هستند ردیابی می کنند. در این الگوریتمها تصویر زمینه بصورت پویا نگهداری شده و نواحی متحرک با استفاده از تفریق زمینه از تصویر جاری بدست می آیند. ایده اصلی در این روش، تعیین ناحیه ی متصل مربوط به هر شیء متحرک در تصویر و سپس ردیابی آن در طول زمان با استفاده از معیار همبستگی متقابل^۲ است.

ردیابی مبتنی بر ناحیه در سالهای اخیر کاربرد بسیار زیادی پیدا کرده است. به عنوان مثال Warn و همکارانش [۳۵] از ویژگیهای لکه های کوچک برای ردیابی یک شخص در محیط بسته استفاده می کنند. [Warn] در مطالعه آنها هر شیء به شکل ترکیبی از لکه های متناظر در نظر گرفته می شود. ضمناً شیء مورد نظر و زمینه هر دو توسط توزیعهای گوسی مدل می شوند. سپس پیکسلهای متعلق به اجزای مختلف شیء با استفاده از معیار log-likelihood به اجزای شیء نسبت داده می شوند. بنابراین با ردیابی هر لکه کوچک، شیء متحرک موجود در تصویر قابل ردیابی هستند.

در [۳۶] Mckenna و همکارانش یک روش تفریق زمینه و فقی ارائه می کنند که در آن از ترکیب رنگ و اطلاعات گرادیان برای مقابله با سایه ها و رنگهای غیرقابل اطمینان در تقطیع حرکت استفاده می شود. سپس روال ردیابی در سه

^۱ Projection

^۲ Cross-Correlation

سطح ناحیه، افراد و گروه‌ها صورت می‌گیرد. هر ناحیه یک مستطیل محاط کننده مخصوص دارد که می‌تواند با نواحی دیگر ترکیب شده و یا به نواحی مجزا تقسیم شود. هر شخص با ترکیب تعدادی از نواحی که محدودیتهای مربوط به ساختار هندسی بدن انسان را برآورده می‌کنند، تشخیص داده می‌شود و هر گروه نیز با ترکیب تعدادی از افراد تشکیل می‌شود. بنابراین آنها توانستند با استفاده از ردیابی نواحی و استفاده از مدل ظاهری رنگ چندین نفر را حتی در شرایطی که همپوشانی دارند ردیابی کنند. ردیابی مبتنی بر ناحیه نسبتاً خوب عمل می‌کند ولی در دو وضعیت مشکل رخ می‌دهد. وضعیت اول زمانی است که سایه برای مدت طولانی در تصویر وجود داشته باشد که چنین شرایطی ممکن است منجر به ترکیب نواحی مربوط به افراد مختلف شود. البته این مشکل با استفاده از رنگ و توجه به این نکته که نواحی سایه فاقد بافت هستند تا حدی قابل برطرف شدن می‌باشد. مشکل جدی‌تر زمانی رخ می‌دهد که افراد در محیط بصورت متراکم حضور داشته باشند. در چنین شرایطی که افراد بصورت جزئی با یکدیگر همپوشانی دارند، تقطیع افراد حاضر در صحنه بسیار دشوار می‌شود. برای برطرف کردن این مشکل باید از چند دوربین استفاده کرد.

۲-۲-۴- ردیابی مبتنی بر کانتور

ردیابی کانتور اهداف متحرک در تصاویر ویدئویی یکی از زمینه‌های پرکاربرد در بینایی ماشین می‌باشد. در اکثر روشهای موجود از کانتور فعال به عنوان تکنیکی با انعطاف پذیری بالا استفاده شده است. کانتور فعال [۳۷] یک منحنی فرم پذیر می‌باشد که با کمینه سازی انرژی تعریف شده برای آن یک مشخصه را در تصویر که معمولاً مرز هموار یک ناحیه می‌باشد، آشکارسازی می‌نماید. کانتور فعال به تنهایی قادر به ردیابی اهداف متحرک در تصاویر ویدئویی می‌باشد، ولی عواملی چون حرکت‌های سریع هدف، پس زمینه شلوغ، وجود لبه‌های قوی در پس زمینه و لرزش دوربین در حین تصویربرداری، به شدت بر روی عملکرد کانتور فعال در ردیابی تاثیر گذاشته و موجب ناموفق بودن آن می‌شوند. علت عدم کارایی کانتور فعال در حضور شرایط ذکر شده، ناکافی بودن دقت حالت اولیه کانتور فعال

در هر تک تصویر می باشد. به این صورت که حرکت‌های سریع هدف و یا لرزش دوربین بین دو تک تصویر متوالی از ویدئو موجب دور افتادن حالت اولیه کانتور فعال از مرزهای هدف در تک تصویر دوم می شود و وجود پس زمینه شلوغ و لبه های قوی در پس زمینه موجب می شود کانتور فعال در کمینه های محلی^۱ انرژی دیگری (ناشی از لبه های پراکنده موجود در پس زمینه) به تله افتاده و به تدریج هدف را گم کند.

با توجه به مشکل ذکر شده ، معمولاً کانتور فعال را به همراه یک تکنیک کمکی برای ردیابی به کار می برند [۳۸-۴۱].

از جمله petrfreund [۳۸] از فیلتر کالمن برای کمک به عملکرد بهتر کانتور فعال در ردیابی استفاده نموده است. [۳۹] از اطلاعات شار نوری و سرعت حرکت کانتور در تعیین یک محدوده برای فعالیت کانتور فعال در هر تک تصویر سود برده است. [۴۰] نیز از کانتور فعال به منظور تصحیح نواحی بدست آمده از بخش بندی رنگ و حرکت در هر تک تصویر استفاده کرده است. جاماسبی [۴۱] نیز از تبدیل مستوی^۲ برای تخمین مدل حرکتی^۳ کانتور فعال استفاده کرده است. در این روش برای شروع ردیابی، ابتدا کانتور هدف در اولین تک تصویر به صورت تقریبی توسط کاربر ترسیم می شود. سپس کانتور تقریبی به عنوان حالت اولیه یک کانتور فعال ساده بکار می رود. با کمینه سازی انرژی کانتور فعال، مرزهای دقیق هدف در اولین تصویر مشخص می شوند. در ادامه، ردیابی شامل پیشگویی کانتور و کمینه سازی انرژی کانتور فعال در هر تک تصویر از دنباله تصاویر ویدئویی می شود. برای پیشگویی کانتور هدف در یک تک تصویر، نخست مدل حرکت مستوی هدف بین تصویر قبلی تا تصویر کنونی محاسبه شده و سپس کانتور نهایی هدف در تک تصویر قبلی توسط این مدل تبدیل می گردد. کانتور بدست آمده که پیشگویی

^۱ Local Minimum

^۲ Affine Transform

^۳ Motion Estimation Model

کانتور هدف در تک تصویر کنونی می باشد، بعد از انجام بعضی اصلاحات به عنوان حالت اولیه کانتور فعال در تصویر کنونی بکار برده می شود. کانتور فعال نیز با کمینه سازی انرژی مرزهای هدف را یافته و کانتور دقیق هدف در تک تصویر کنونی بدست می آید. نقش کانتور فعال در این روش، تصحیح خطای پیشگویی کانتور هدف با استفاده از اطلاعات گرادیان تصویر می باشد. نقش جانبی آن نیز مشخص نمودن تغییر شکل تدریجی کانتور هدف است، که معمولاً به علت تغییر نمود هدف اتفاق می افتد و تبدیل مستوی قادر به مدل کردن آن نمی باشد.

۲-۲-۵- ردیابی مبتنی بر ویژگی

الگوریتمهای ردیابی مبتنی بر ویژگی، ردیابی اشیاء را با استخراج عناصر، دسته بندی آنها در ویژگی های سطح بالاتر و سپس تطبیق ویژگیها بین تصاویر انجام می دهد. این الگوریتمها را با توجه به نوع ویژگیهای انتخاب شده می توان در سه دسته اصلی دسته بندی کرد: الگوریتمهای مبتنی بر ویژگی های سراسری^۱، الگوریتمهای مبتنی بر ویژگیهای محلی^۲ و الگوریتمهای مبتنی بر گراف وابستگی^۳.

ویژگیهای مورد استفاده در الگوریتم های مبتنی بر ویژگیهای سراسری شامل مرکز ثقل^۴، محیط، مساحت، رنگ و ویژگیهایی از این دست هستند. ویژگیهای مورد استفاده در الگوریتمهای مبتنی بر ویژگیهای محلی شامل قطعه خطها، قطعه منحنیها، رئوس زوایا و ویژگیهایی مشابه آنها و ویژگی های مورد استفاده در الگوریتم مبتنی بر گراف وابستگی شامل مجموعه هایی از فواصل و روابط هندسی بین ویژگی ها هستند.

^۱ Global feature-based

^۲ Local feature-based

^۳ Dependence-graph-based

^۴ Centroid

ردیابی مبتنی بر ویژگی شامل استخراج و تطبیق ویژگیها است. ویژگیهای سطح پایین از قبیل نقاط برای استخراج ساده تر هستند. در مقابل ردیابی ویژگیهای سطح بالاتر مثل خطوط و لکهها مشکل تر است. لذا بین پیچیدگی ویژگیها و کارایی ردیابی یک مصالحه^۱ است [۴۲]. مزیت استفاده از ویژگیهای سطح بالاتر این است که حتی زمانی که همپوشانی وجود دارد، تعدادی از ویژگیهای اشیاء مورد ردیابی قابل رویت هستند. در یکی از روشهایی که از نقطه ویژگی برای ردیابی استفاده می شود، هر شخص توسط یک مستطیل محاط شده است و مرکز این مستطیل بعنوان نقطه ویژگی برای ردیابی بکار می رود. حتی در شرایطی که همپوشانی بین اشیاء متفاوت وجود داشته باشد، تا زمانی که سرعت نقاط مراکز مستطیلها قابل تشخیص باشد، ردیابی بخوبی انجام می شود. در روش دیگری برای ردیابی از نقاط گوشه‌ی سایه‌های مربوط به افراد متحرک بعنوان نقاط ویژگی برای ردیابی استفاده می شود و این نقاط ویژگی با استفاده از یک معیار فاصله بین موقعیت منحنی‌های موجود در فریم‌های متوالی تطبیق داده می شوند.

۲-۲-۶- ردیابی با استفاده از دسته بندی کننده

اگر مسئله ردیابی اشیاء متحرک به عنوان یک مسئله طبقه بندی در نظر گرفته شود آنگاه با انتخاب یک طبقه بندی کننده مناسب می توان جسم را از پس زمینه در فریمهای متوالی جداسازی کرد [۴۳-۴۵]. بعنوان مثال Avidan در [۴۵] مسئله ردیابی را مانند یک مسئله دسته بندی باینری در نظر گرفته است. مجموعه‌ای از دسته بندی کننده‌های ضعیف بصورت برخط برای تمایز میان اشیاء پیش زمینه و زمینه آموزش داده می شوند. سپس با الگوریتم آبخاری^۲ از ترکیب این مجموعه از دسته بندی کننده‌ها برای ایجاد یک دسته بندی کننده ی قوی تر استفاده می گردد. سپس در فریم بعدی از این دسته بندی کننده استفاده

^۱ Trade-off

^۲ Adaboost

می شود تا پیکسل های اشیاء زمينه و پيش زمينه را در دو دسته ی متفاوت قرار دهيم. ضمناً در حين روال ردیابی دسته بندی کننده های ضعیف بصورت برخط آموزش می بینند و در ایجاد دسته بندی کننده ی قوی تر شرکت می کنند.

اخيراً J.Ahmed در [۴۶] با استفاده از یک شبکه عصبی و آموزش آن با استفاده از ویژگی های استخراج شده توسط تبدیل موجک موفق به ردیابی یک هواپیمای مشخص شده است.

۲-۳- استخراج مشخصه

یکی از مراحل اساسی در پردازش تصاویر رقمی که به عنوان یک مرحله اولیه در بینایی ماشین مطرح می باشد استخراج مشخصه است. مشخصه هایی چون لبه ها، گوشه ها و نواحی، رنگ هدف، نوع بافت آن و ... اطلاعات موجود در تصویر را به صورت خلاصه تر و مناسب تر برای کاربردهای مطرح در بینایی ماشین توصیف می نمایند. از جمله در اکثر روشهای ردیابی اهداف متحرک لازم است در ابتدا مشخصه هایی از هدف استخراج شوند تا با ردیابی این مشخصه ها بتوانیم هدف را در طول تصاویر ویدئویی ردیابی نماییم. در گروهی از روشهای ردیابی نیز مشخصه ها در هر تک تصویر استخراج شده و سپس مشخصه های مربوط به هدف شناسایی می گردند. در این بخش مروری بر شاخصه های به کار رفته در الگوریتم پیشنهادی شده و در مورد نحوه محاسبه هر کدام از این شاخصه ها بحث خواهد شد.

۲-۴- شاخصه لبه

نقاطی از تصویر را که مرز نواحی با شدت روشنایی نسبتاً متفاوت را مشخص می کنند، لبه می نامیم. در اینجا به شرح روشهای کلاسیک لبه یابی می پردازیم که معمولاً شامل دو مرحله هستند: آشکارسازی لبه و پیوند لبه.

۲-۴-۱- آشکارسازی لبه^۱

نظریه پایه در بیشتر روشهای آشکارسازی لبه، محاسبه یک عملگر مشتق محلی است. در شکل (۲-۲) یک تصویر و مشتقات آن را ملاحظه می‌کنید. از بزرگی مشتق اول می‌توان برای تعیین اینکه آیا پیکسل بر روی لبه قرار دارد یا خیر و از علامت مشتق دوم برای تعیین اینکه آیا پیکسل لبه مورد نظر روی طرف تیره یا روشن لبه قرار دارد، استفاده کرد. مشتق دوم در نقطه وسط هر گذر سطح خاکستری یک عبور از صفر دارد. عبور از صفرها راهی قوی برای تعیین محل لبه‌های تصویر فراهم می‌آورند. توضیحات مشابهی برای یک لبه در تصویر با هر جهت دلخواه نیز صدق می‌کند. اندازه مشتق اول تصویر در هر نقطه برابر بزرگی گرادیان است. مشتق دوم نیز با استفاده از لاپلاسین بدست می‌آید.

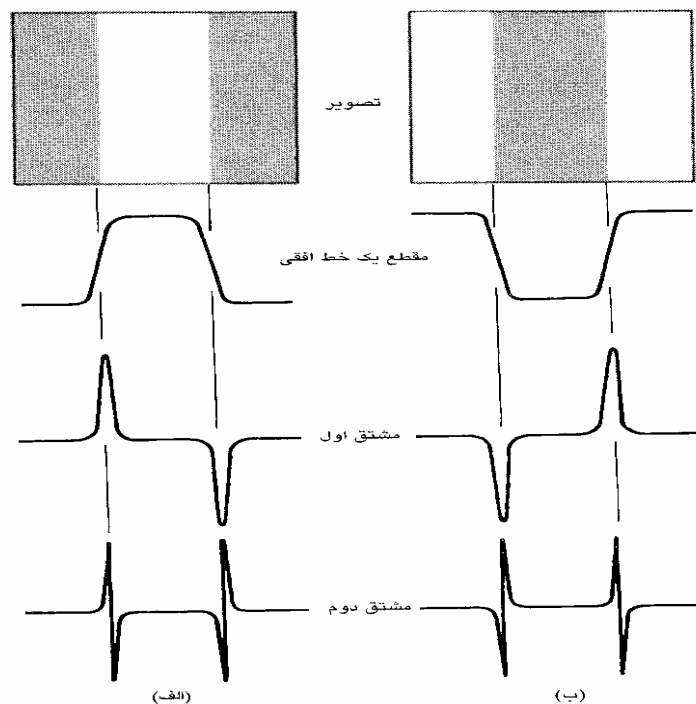
گرادیان تصویر $I(x,y)$ در نقطه (x,y) ، همچنین اندازه و زاویه بردار گرادیان با روابط زیر تعریف می‌شوند:

$$\nabla I = \begin{bmatrix} \frac{\partial I}{\partial x} \\ \frac{\partial I}{\partial y} \end{bmatrix} \quad (1-2)$$

$$|\nabla I| = \sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2} \quad (2-2)$$

$$\alpha(x, y) = \tan^{-1} \left(\frac{\partial I / \partial y}{\partial I / \partial x} \right) \quad (3-2)$$

^۱ Edge Detection



شکل ۲-۲: آشکارسازی لبه با عملگرهای مشتق: (الف) نوار روشن روی زمینه تیره. (ب) نوار تیره روی زمینه روشن.

جهت بردار گرادیان جهتی را (که آهنگ تغییر I) بیشینه باشد نشان می‌دهد. اندازه گرادیان نیز شدت تغییرات را نشان می‌دهد. محاسبه گرادیان یک تصویر مبتنی بر بدست آوردن مشتقات جزئی $\frac{\partial I}{\partial x}$ و $\frac{\partial I}{\partial y}$ در محل هر پیکسل است. به چند طریق می‌توان مشتقات را به شکل رقمی پیاده‌سازی کرد. معمولاً از عملگرهای "سobel" ^۱ به دلیل خاصیت آرام‌سازی که دارند بیشتر استفاده می‌شود. مشتقات مبتنی بر ماسک‌های سوبل عبارتند از:

$$\frac{\partial I}{\partial x} \approx (z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3) \quad (۴-۲)$$

^۱ Sobel

$$\frac{\partial I}{\partial y} \approx (z_3 + 2z_6 + z_9) - (z_1 + 2z_4 + z_7) \quad (۵-۲)$$

که در آن z ها سطوح خاکستری پیکسل‌های زیرماسک‌ها در هر نقطه تصویر هستند. (شکل ۲-۳)

z_1	z_2	z_3
z_4	z_5	z_6
z_7	z_8	z_9

(الف)

-1	-2	-1
0	0	0
1	2	1

(ب)

-1	0	i
-2	0	2
-1	0	1

(پ)

شکل ۲-۳: الف) یک ناحیه تصویری 3×3 (ب) ماسک مورد استفاده برای محاسبه $\frac{\partial I}{\partial x}$

(پ) ماسک مورد استفاده برای محاسبه $\frac{\partial I}{\partial y}$ (عملگرهای سوبل)

مقدار گرادیان در مرکز ماسک با استفاده از رابطه اندازه گرادیان (۱-۵) محاسبه می‌شود. برای بدست آوردن مقدار بعدی گرادیان، این ماسک‌ها به محل پیکسل‌های بعدی جابجا می‌شوند و روال قبل تکرار می‌شود. پس از اجرای روال فوق برای تمام نقاط تصویر یک تصویر گرادیان با همان ابعاد تصویر اولیه بدست می‌آید. با اعمال یک مقدار آستانه به تصویر گرادیان بدست آمده، لبه‌های قوی تر آشکارسازی می‌شوند. نتیجه اعمال این روش در شکل (۲-۴) نمایش داده شده است.

❖ آشکارسازی لبه‌های متحرک

در تصاویر ویدئویی دوربین ثابت چنانچه هدف از ناحیه بندی، آشکارسازی نواحی متحرک تصویر باشد، تنها لبه‌های متحرک مورد توجه هستند. به منظور آشکارسازی لبه‌های متحرک می‌توان از ملاک زیر استفاده نمود:

$$E = |\nabla I| \cdot \left| \frac{dI}{dt} \right| \quad (۶-۲)$$

سپس با اعمال یک مقدار آستانه به تصویر بدست آمده، لبه‌های متحرک آشکارسازی می‌شوند. توجه نمایید که روش فوق، لبه‌های قوی با حرکت ناچیز و لبه‌های ضعیف با حرکت سریع را نیز آشکارسازی می‌کند.

۲-۴-۲- پیوند لبه^۱

حالت مطلوب در روشهای آشکارسازی لبه این است که کلیه پیکسل‌های روی مرز نواحی آشکار شوند. در عمل به عللی مانند وجود نویز، انفصالات مرز که ناشی از روشنایی غیریکنواخت است، و سایر عواملی که ناپیوستگی‌های شدت روشنایی را ایجاد می‌کنند، به ندرت این پیکسل‌ها مرز کاملی را مشخص می‌سازند. این اثر در شکل (۲-۴) مشاهده می‌شود. بنابراین معمولاً به دنبال اجرای الگوریتم‌های آشکارسازی لبه، الگوریتم‌های پیوند و الگوریتم‌های آشکارسازی سایر مرزها نیز اجرا می‌شوند تا در نهایت مرزهای مورد نظر لبه یابی به دست آید. در اینجا روش پردازش محلی را بررسی می‌کنیم.

❖ روش پردازش محلی

در تصویری که لبه‌های آن آشکار شده است یکی از ساده‌ترین راهها برای پیوند نقاط لبه، تحلیل مشخصات پیکسل‌ها در یک همسایگی کوچک (مثلاً 3×3 یا 5×5) حول هر نقطه (x,y) است. تمام نقاط مشابه در همسایگی به هم پیوند می‌خورند، و بدین ترتیب مرزی از پیکسل‌هایی که چند خاصیت مشترک دارند، تشکیل می‌شود.

در این تحلیل دو خاصیت اصلی مورد استفاده برای تعیین شباهت پیکسل‌های لبه عبارتند از:

^۱Edge Linking

۱- بزرگی پاسخ عملگر گرادیان به کار رفته برای آشکارسازی لبه،

۲- جهت گرادیان.

همانطور که در معادله (۲-۲) تعریف شد، اولین خاصیت با مقدار $|\nabla I|$ تعیین می‌شود. بنابراین یک پیکسل لبه با مختصات (x', y') و در همسایگی از پیش تعریف شده (x, y) بزرگی گرادیان مشابهی با پیکسل (x, y) دارد اگر:

$$|\nabla I(x, y) - \nabla I(x', y')| \leq T \quad (۷-۲)$$

در رابطه فوق T یک آستانه مثبت است. جهت بردار گرادیان با معادله (۳-۲) تعیین می‌شود. بنابراین یک پیکسل لبه در (x', y') و در همسایگی پیش‌تعریف شده (x, y) زاویه مشابهی با پیکسل (x, y) دارد اگر:

$$|\alpha'(x, y) - \alpha'(x', y')| < A \quad (۸-۲)$$

که A یک آستانه مثبت است. نقطه (x', y') به (x, y) پیوند می‌خورد اگر هر دو معیار بزرگی و جهت برآورده شود.



شکل ۲-۴: (الف) یک تصویر نمونه (ب) تصویر دو سطحی شده حاصل از اعمال ماسک های سوبل بر روی تصویر "الف". (ج) تصویر مرز نواحی حاصل از پیوند لبه

۲-۵- شاخصه رنگ

در پردازش تصاویر ویدئویی انتخاب مناسب فضای رنگ یک کار مقدماتی خصوصاً برای بخش بندی تصاویر ویدئویی است [۴۷]. ما به طور گسترده خواص فضاهای رنگی را برای تعیین یک فضای رنگی مناسب که در بخش بندی تصاویر ویدئویی به کار می رود مورد جستجو قرار داده و سپس با روشی که در فصل سوم ذکر خواهد شد و با محاسبه معیار شباهت بین جسم مورد ردیابی و پس زمینه برای هر یک از مؤلفه های فضاهای رنگ به انتخاب مناسب فضای رنگ دست خواهیم زد.

۲-۶- توصیف انواع مدل‌های رنگ

برای بیان فضای رنگ، مدل‌های متفاوتی ارائه گردیده است. در سال ۱۹۳۱، مؤسسه‌ای جهانی با نام CIE، طیفی از رنگها را به عنوان رنگهای مبنا معرفی کرد. وجود این رنگها حالت واقعی نداشت بلکه به صورت تصویری بود و بیان آنها تنها از طریق توصیف معادلات ریاضی امکان‌پذیر بود. بعد از آن مدل‌های زیادی برای رنگها، بصورت معادلات ریاضی توصیف گردید. تعداد این مدلها را می‌توان وجود شرایط مختلف در تصویر از قبیل، میزان نویز پذیری تصویر، میزان روشنایی و ... جستجو کرد. انواع مدل‌های رنگ عبارتند از:

۱- مدل Huncell

۲- مدل RGB (قرمز، سبز، زرد)

۳- مدل نرمالیزه شده RGB

۴- مدل CMY

۵- مدل XYZ

۶- مدل xyz

۷- مدل Lab

۸- مدل YUV

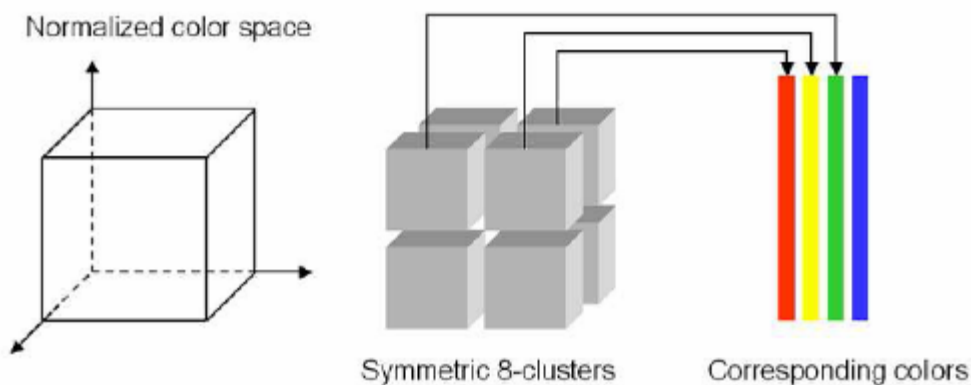
۹- مدل HSI

۱۰- مدل YCbCr

توصیف مدل‌های کاربردی رنگ فوق در تصاویر رنگی در پیوست گزارش آمده است.

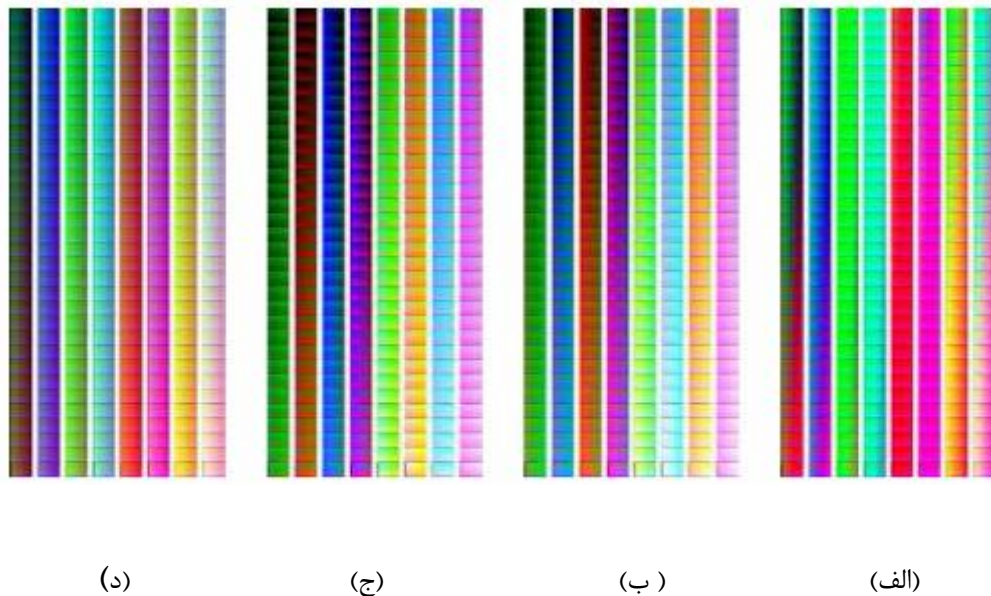
۲-۷- مقایسه برخی از فضای توصیفی

برای مقایسه بهتر برخی از مدل‌های رنگ ارائه شده، فضای متعامد مکعبی دکارتی را مانند شکل زیر به ۸ قسمت تقسیم کرده و با اعمال ۴ مدل RGB، XYZ، YUV و YCbCr به کل فضا، برای هر مدل، هشت کلاستر به وجود خواهد آمد.



شکل ۲-۵: خوشه بندی فضای مکعبی جهت مقایسه انواع مدلها

همانطور که از نتایج بدست آمده از شکل (۲-۶) مشاهده می‌شود، مدل‌های XYZ و RGB در میزان روشنایی تفاوت عمده‌ای با هم دارند. در ضمن مدل XYZ، بیشترین ناهمگنی را نسبت به مدل‌های دیگر دارد. از دیگر نتایج اینکه، مدل YUV مانند مدل RGB، حساسیت مدل به پارامتر روشنایی را بیان نمی‌کند، هرچند که جداسازی محورها در مدل RGB بهتر مشاهده می‌گردد.



شکل ۲-۶: مقایسه ۴ مدل توصیفی: (الف) مدل XYZ، (ب) مدل YCbCr، (ج) مدل YUV و (د) مدل RGB

۲-۸- شاخصه بافت

بافت در تصویر یک مقیاس موثر در تخمین ساختار، جهت، سختی، نرمی و یا تفاوت های منظم در نواحی گوناگون در یک منظره از تصویر است [۴۸]. تشخیص کلاس های بافت متفاوت در مناظر جهان واقعی یا یک تصویر اغلب برای سیستم بصری بشر یک کار جزئی می باشد در حالیکه این موضوع یک مسئله مهم در حوزه بینایی ماشینی^۱ و پردازش تصویر می باشد. با افزایش محبوبیت کتابخانه های دیجیتال^۲، پایگاههای داده چند رسانه ای^۳ و تصویری، تحلیل بافت به یک حوزه مرکزی در تحقیقات در مورد نواحی تصویر (که با توجه به خصوصیات بافت آن توصیف می شوند) تبدیل شده است. همچنین به طور

^۱ Machine Vision

^۲ Digital Libraries

^۳ Multimedia Databases

فوق العاده مفید در بسیاری از حوزه های مربوط دیگر مانند بازیابی تصویر براساس محتوی، شرح و تقسیم بندی های تصویر، تشخیص اشیا، کنترل از راه دور^۱ در صنعت کاربرد دارد.

بافت می تواند به صورت الگوهای گام به گام از تغییرات محلی شدت پیکسل ها بررسی شود. برای مثال، تصویر یک سطح چوبی یکنواخت نیست بلکه تغییرات شدتی که از الگوهای تکرار شده معینی که بافت بصری نامیده می شود را در بر می گیرد. الگوها می توانند در نتیجه ویژگی های سطح فیزیکی از قبیل استانداردهای جهتی یا ناهمواری (که اغلب کیفیت لامسه ای دارد و یا در نتیجه تفاوت انعکاس های حاصل از رنگ یک سطح است) باشد. ممکن است بافت را در یک ناحیه ماکروسکوپی^۲ در نظر بگیریم، در اینصورت ساختار بافت به سادگی با الگوهای گام به گام که در آن عناصر بر طبق یک قانون مکانی مرتب شده اند، توصیف می شود.

یک ناحیه از تصویر، یک بافت ثابت دارد اگر یک مجموعه از خواص آماری محلی یا خواص دیگری از تابع تصویر باشد که به تدریج تغییر کند و یا تقریباً پیرودیک باشد. ایده بافت به سه عامل وابسته است: (۱) نظم محلی که در یک ناحیه بزرگ در مقایسه با اندازه آن نظم تکرار شود. (۲) نظمی که نظم غیر تصادفی بخش های اولیه را شامل می شود. (۳) بخش هایی که به طور یکنواخت ناهموارند و تقریباً ابعاد مساوی با ابعاد ناحیه بافت دارند. این مجموعه از تعاریف بر این مطلب دلالت می کند که تعریفی که از بافت توسط افراد مختلف ارائه می شود وابسته به کاربرد های خاص است و موافقت کلی با این تعاریف وجود ندارد. بعضی از تعاریف بصورت مفهومی بیان شده است و بعضی به طور کامل از کاربردهایی که از بافت دیده شده، نتیجه می شود.

^۱ Remote sensing

^۲ Macroscopic

بافت در تصویر به عنوان یک تابع از تغییرات فضایی در شدت پیکسل ها (مقادیر سطوح خاکستری) که برای بسیاری از کاربرد های متفاوت که هدف مطالعه بسیاری از متخصصان است، مفید می باشد. یک کاربرد ضروری از بافت در تصویر، تشخیص نواحی تصویر با استفاده از خواص بافت می باشد.

۲-۸-۱- کاربردهای بافت

تحلیل بافت یک حوزه مفید و مهم در مطالعات بینایی ماشین می باشد. بیشتر سطوح طبیعی شامل بافت می باشند و یک سیستم بصری موفق باید توانایی این را داشته باشد که با دنیای دارای بافت اطراف خود در ارتباط باشد. روش های تحلیل بافت در حوزه های کاربردی متفاوتی استفاده می شوند. در بعضی از این حوزه های کامل شده از قبیل کنترل از راه دور بافت از پیش نقش مهمی را ایفا می کند. به عنوان مثال بافت برای بررسی یکنواختی مناظر محلی تعریف می شود و این ویژگی برای کلاس بندی زمین براساس مسائلی از قبیل وضعیت آب ، نواحی کشاورزی و غیره مفید می باشد. به علاوه تحلیل بافت در بسیاری از تصاویر پزشکی کاربرد دارد چنانکه در تصاویر فراصوتی^۱ قلب در بافت به عنوان مقداری تصادفی (که مقداری کمتر از حفره پر شده خون در همسایگی مرز میان حفره قلب و دیوار داخلی آن می باشد) تعریف شده است. از این حقیقت می توان برای انجام تقسیم بندی و آشکارسازی کرانه ها با استفاده از روش های تحلیل بافت استفاده کرد.

۲-۸-۲- طبقه بندی مدل های بافت

تعیین کیفیت های مشاهده شده از بافت در یک تصویر یک مرحله ابتدایی مهم در ساختن مدل های ریاضی برای بافت می باشد. تغییرات شدت در یک تصویر که بافت را مشخص می کند ، عموماً مربوط به بعضی تغییرات فیزیکی در منظره می باشد (از قبیل سنگریزه های روی ساحل یا امواج دریا). مدل سازی

^۱ Ultrasound

این تغییرات فیزیکی بسیار مشکل است و بنابراین بافت معمولاً با تغییرات دو بعدی در شدت های نشان داده شده در تصویر مشخص می شود. بافت در تصویر تعدادی کیفیت قابل مشاهده دارد که نقش مهمی را در شرح بافت بازی می کند. مانند: یکنواختی، تراکم، نرمی، زبری، نظم، خطی بودن، جهت، فرکانس و فاز. بعضی از این کیفیت های مشاهده شده مستقل نیستند. برای مثال فرکانس و چگالی مستقل نیستند و خاصیت جهتی تنها برای بافت های جهت دار بکار می رود. این حقیقت که فهم بافت ابعاد متفاوت زیادی دارد به علت آن است که یک روش واحد برای نمایش بافت که مناسب برای انواع مختلفی از بافت می باشد، وجود ندارد. به طور کلی روش های زیر برای تحلیل بافت به کار می روند:

۱- روش های آماری که از خواص آماری موجود در تصویر برای تحلیل بافت استفاده می کنند.

۲- روش های هندسی که به وسیله تعریفشان از بافت که از ترکیب عناصر بافت مشخص می گردند. روش تحلیل معمولاً وابسته به خواص هندسی عناصر بافت می باشد.

۳- روش های بر پایه مدل که مبتنی بر ساختن مدل یک تصویر (که می تواند نه فقط برای شرح بافت بلکه برای ترکیب آنها هم استفاده شود) می باشد. پارامترهای مدل از کیفیت های مشاهده شده اصلی در بافت بدست می آید.

۴- روش های پردازش سیگنال که سعی می کنند تا ویژگی های معینی از تصاویر فیلتر شده (که می توانند در کاربرد های طبقه بندی یا تقسیم بندی بکار روند) را محاسبه کنند.

انواع زیادی از روش های استخراج ویژگی بر پایه پردازش سیگنال وجود دارد. فیلترینگ موجک یکی از این روش هاست که می تواند با موفقیت عمل استخراج ویژگی در بافت را انجام دهد. ما از تبدیل موجک گسسته دو بعدی^۱ برای استخراج ویژگی های بافت تصویر استفاده شده است که در قسمت بعدی به آن می پردازیم.

^۱ Two Dimensional Discrete Wavelet Transform

۲-۹- تبدیل موجک گسسته دو بعدی

اگر چه تبدیل فوریه از اواخر ۱۹۵۰ تا کنون تکیه گاه اصلی در پردازش تصویر بوده ولی با استفاده از تبدیل موجک امکان آنالیز قوی تر تصاویر فراهم گشته است [۴۹]. برخلاف تبدیل فوریه که در آن مینا، یک تابع سینوسی نامحدود در زمان می باشد، در تبدیل موجک تابع مینا، بر فواصل زمانی کوتاه در فرکانس های مختلف تعریف می شود. چنین ویژگی توابع مینا در تبدیل موجک، هم امکان مشاهده اطلاعات فرکانس و هم اطلاعات زمان (مکان) را فراهم می سازد. این ویژگی برخلاف تبدیل فوریه است که در حوزه فرکانس اطلاعات مربوط به زمان و در حوزه زمان اطلاعات مربوط به فرکانس را نداشتیم. موجک را می توان در سطوح مختلف به کار برد. برای سیگنالهای دو بعدی مثل تصویر، موجک در سه جهت مختلف افقی^۱، عمودی^۲ و قطری^۳ روی تصویر اعمال می شود. به این ترتیب امکان آنالیز تصویر در سه راستای مختلف بدست می آید. در هر سطح از تجزیه به کمک موجک، ابعاد تصویر در هر جهت نصف شده و جزئیات بیشتر با وضوح^۴ کمتر در اختیار ما قرار می گیرد. در جعبه ابزار موجک^۵ در نرم افزار MATLAB هنگام تجزیه هر تصویر در هر سطح، تصویر نرم شده^۶، جزئیات در جهت افقی، جزئیات در جهت عمودی و جزئیات در جهت قطری داده می شود که اگر این اعمال به صورت سلسله مراتبی اعمال شود، تصویر نرم شده مرحله قبل مجدداً به چهار جزء، یعنی تصویر نرم شده مرحله بعد، جزئیات در جهت افقی، جزئیات در جهت عمودی و جزئیات در جهت قطری تجزیه می شود. این

^۱ Horizontal

^۲ Vertical

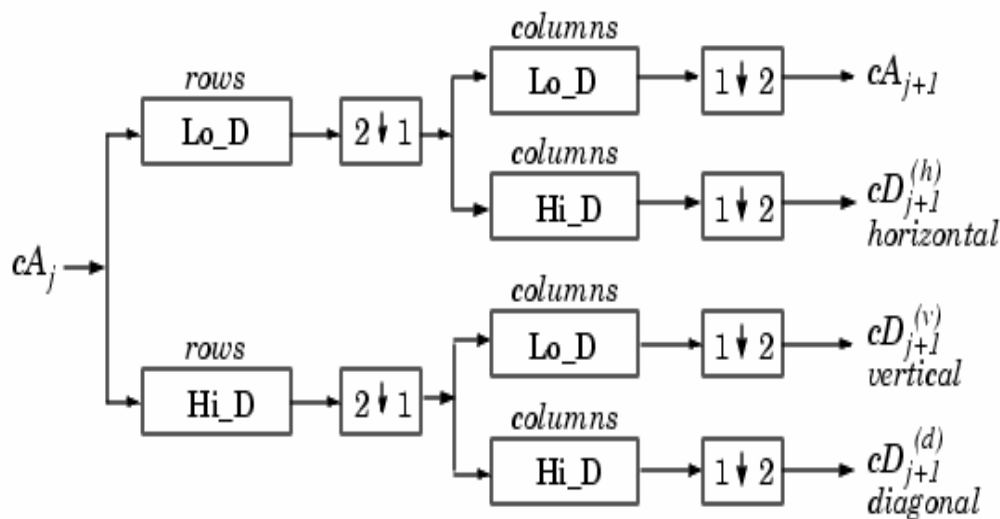
^۳ Diagonal

^۴ Resolution

^۵ Wavelet toolbox

^۶ Approximated image

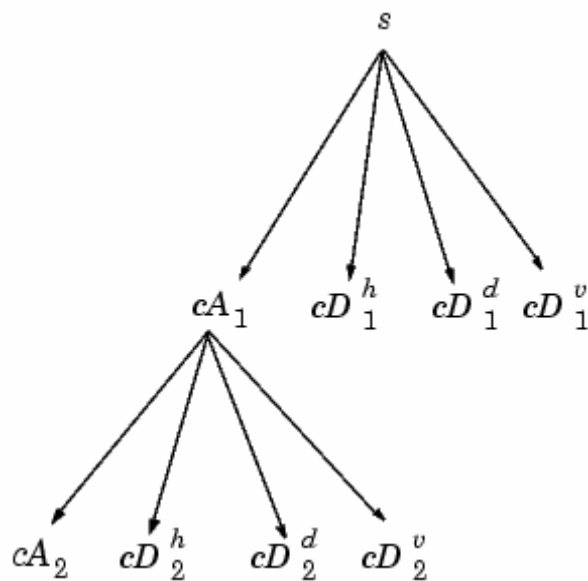
عمل تا هر سطحی که مورد نظر آنالیز کننده باشد قابل پیاده سازی است. شکل (۷-۲) مراحل تجزیه موجک دو بعدی را در MATLAB را نشان می دهد.



شکل ۷-۲: مراحل تجزیه موجک دو بعدی در MATLAB

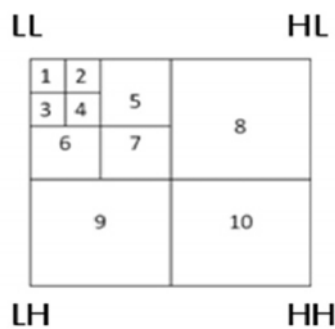
بلوک $2 \downarrow 1$ وظیفه کم کردن وضوح سطرها و بلوک $1 \downarrow 2$ وظیفه کم کردن وضوح ستون ها را بر عهده دارد. در هر سطح از تجزیه، دو فیلتر پایین گذر و بالا گذر روی تصویر اعمال می شود. عمل فیلتر کردن در دو مرحله انجام می شود، اول روی سطرهای تصویر و در مرحله بعد روی ستون ها.

نتیجه این تجزیه، تصویر نرم شده (LL)، جزئیات در جهت افقی (LH)، جزئیات در جهت عمودی (HL) و قطری (HH) است. $cA_{j+1}^{(h)}$ ، $cD_{j+1}^{(v)}$ ، $cD_{j+1}^{(d)}$ و $cD_{j+1}^{(h)}$ در شکل (۷-۲) به ترتیب نشان دهنده ضرایب تصویر نرم شده و جزئیات در جهت های افقی، عمودی و قطری است. شکل (۸-۲) موجک را در سطوح مختلف نشان می دهد.



شکل ۲-۸: موجک در سطوح مختلف. اندیس ۱ و ۲ به ترتیب نشان دهنده سطح ۱ و ۲ است.

در استخراج ویژگی های بافت تصویر از تبدیل موجک دو مرحله ای نشان داده شده در شکل (۲-۸) استفاده می شود. با انجام سه باره این تجزیه دو مرحله ای بر روی پنجره های R ، G و B برای هر پنجره رنگی ۱۰ زیر تصویر همانند شکل (۲-۹) بدست می آید که از آن برای محاسبه هیستوگرام بافت استفاده می کنیم.



شکل ۲-۹: ده کانال بدست آمده بعد از اعمال سه مرتبه تبدیل موجک دو بعدی

۲-۱۰- جمع بندی

در این فصل به معرفی کلی مبحث ردیابی شیء پرداختیم. سپس روش های پیشین ردیابی اشیاء را مطرح کردیم. در ادامه ویژگی های گوناگون اشیاء مورد استفاده در روش های ردیابی شیء را ارائه دادیم. حال در فصل بعد یک چهارچوب جدید برای ردیابی مؤثر اشیاء (انواع متفاوتی از اهداف متحرک اعم از صلب و غیر صلب) را بیان می کنیم. این چهارچوب یک الگوریتم دو مرحله ای سریع برای ردیابی شیء است که با بهره گیری از ویژگی های بدست آمده از هدف، بالاحص ویژگی های موجود در رنک هدف، به بهبود ردیابی کمک می کند.

فصل سوم

ردیابی شیء با استفاده از روش
جابجایی میانگین

۳-۱- مقدمه

در این فصل به ارائه یک روش سریع ردیابی شیء برای تخمین مکان هدف می پردازیم. این روش ردیابی سریع روش جابجایی میانگین^۱ نامیده می شود که توسط Meer در [۵۰] بیان شده مکان تقریبی جسم در هر فریم تخمین میزند. در این روش ابتدا هدف را در فریم اول به وسیله یک هیستوگرام وزن دار^۲ (که از اطلاعات رنگ هدف و موقعیت پیکسلهای درون هدف استفاده می کند) مدل می کنیم. اگر ناحیه هدف را به صورت یک بیضی در نظر بگیریم، نقاط داخل این بیضی را درون دایره واحد به مرکز صفر نرمالیزه می کنیم. از آنجایی که پیکسلهای حاشیه ای بیشتر تحت تاثیر عواملی چون پوشیدگی و تداخل با زمینه قرار دارند نسبت به پیکسلهای مرکزی دارای اهمیت کمتری هستند. برای منظور نمودن این اهمیت بر اساس فاصله از مرکز، وزنی به هر پیکسل اختصاص داده می شود. البته برای بالا بردن دقت این الگوریتم می توان از یک تبدیل فاصله ای که در پایان این فصل شرح داده می شود بهره برد. با استفاده از این تبدیل فاصله پیکسلهای متعلق به هدف بر حسب فاصله از مرکز جرم آن وزن دهی می شوند.

در فریم جدید به دنبال نقطه ای هستیم که اگر مرکز بیضی را روی آن نقطه قرار دهیم و هیستوگرام وزن دار را محاسبه کنیم، بیشترین شباهت را با هیستوگرام هدف داشته باشد. از آنجاییکه جابجایی هدف در دو فریم متوالی زیاد نیست، می توان جستجو را در فریم جدید حول مکان هدف در فریم قبلی انجام داد و بیشترین شباهت را به عنوان مکان جدید جسم تخمین زد [۵۱]. این جستجوی کلی بار محاسباتی بالایی داشته، از این رو استفاده از الگوریتم های تکراری مناسب با بار محاسباتی کمتر توصیه می شود. معیار شباهت بین هدف و

^۱ Mean Shift

^۲ Weighting Histogram

کاندیداهای در نظر گرفته شده برای هدف در فریم بعدی با استفاده از یک معیار^۱ که از ضریب باتاچاریا^۲ بدست می‌آید، اندازه گرفته می‌شود. در مورد کار ما، ضریب باتاچاریا به منزله مفهوم یک درجه ای از همبستگی^۳ (ارتباط) است. در این روش تعیین جایگاه هدف می‌تواند با روشهای مختلف فیلترینگ و تکنیک های متحدسازی داده ای^۴ ترکیب شود.

۳-۲-۳- ارائه هدف

در ابتدا یک فضای ویژگی برای مشخص کردن هدف انتخاب می‌شود. مدل هدف مرجع به وسیله تابع چگالی احتمالی q آن در فضای ویژگی مشخص می‌شود. برای مثال، مدل مرجع می‌تواند به عنوان تابع چگالی احتمال رنگ هدف انتخاب شود. مرکز مدل هدف^۵ را در نقطه صفر انتخاب می‌کنیم. در فریم بعدی، یک کاندیدای هدف^۶ در جایگاه y تعریف شده است و به وسیله تابع چگالی احتمال $P(y)$ مشخص می‌شود. هر دو تابع چگالی احتمال (pdf ها) از داده ها تخمین زده می‌شوند. برای نشان دادن هزینه محاسباتی پایین به وسیله چگالی های گسسته و پردازش Real-Time، هیستوگرام هایی با m بین (bin) (تعداد مؤلفه های هیستوگرام) باید مورد استفاده قرار گیرند. بنابراین خواهیم داشت:

$$\hat{q} = \{\hat{q}_u\}_{u=1\dots m} \quad \sum_{u=1}^m \hat{q}_u = 1 \quad \text{مدل هدف}$$

^۱ Metric

^۲ Bhattacharyya Coefficient

^۳ Correlation

^۴ Data association techniques

^۵ Target Model

^۶ Target Candidate

$$\hat{p}(y) = \{\hat{p}_u(y)\}_{u=1\dots m} \quad \sum_{u=1}^m \hat{p}_u = 1 \quad \text{کاندیدای هدف}$$

شایان ذکر است که هیستوگرام، بهترین تخمین چگالی غیر پارامتریک نیست ولی از آنجایی که محاسبه آن آسان است از آن استفاده می کنیم.

ما به وسیله :

$$\hat{\rho}(y) \equiv \rho \left[\hat{p}(y), \hat{q} \right] \quad (1-3)$$

یک تابع شباهت میان \hat{p} و \hat{q} مشخص می کنیم. تابع $\hat{\rho}(y)$ نقش یک تابع احتمال کلی^۱ را بازی می کند و بیشینه های محلی آن در تصویر معرف حضور جسم در آن بیشینه ها است.

در این الگوریتم با اعمال یک ماسک مکانی بر روی هدف که دارای یک هسته ایزوتروپیک^۲ می باشد و سپس تعریف یک تابع شباهت بین هدف در فریم فعلی و کاندیداهای هدف در فریم بعدی؛ می توان به جستجوی مؤثر هدف در فریم بعدی دست زد. در حقیقت با این روش به جای جستجوی فراگیر هدف در فریم بعدی با استفاده از یک روش جستجوی مؤثر که زمان کمتری را نیاز دارد مکان هدف در فریم بعدی را تخمین می زنیم.

معیار شباهت هدف در فریم جاری و کاندیداهای هدف در فریم بعدی با استفاده از معیار باتاچاریا^۳ محاسبه می شود. در اینجا در حقیقت ضریب باتاچاریا

^۱ Maximum Likelihood Function

^۲ Isotropic

^۳ Bhattacharya measure

درجه همبستگی^۱ هدف و کاندیداهای آن را مشخص می کند. حال بعد از تخمین جایگاه هدف در فریم بعدی می توانیم همچنین با استفاده از یک طبقه بندی کننده^۲ دو کلاسه که در این جا ماشین های بردار پشتیبان (SVM) می باشد مرزهای دقیق هدف را بدست می آوریم. در این مرحله یک بردار ویژگی از پیکسلهای موجود در تصویر مرجع^۳ تشکیل داده و طبقه بندی کننده را با استفاده از این بردار به صورتی که پیکسلهای متعلق به هدف در یک کلاس و پیکسلهای متعلق به پس زمینه در کلاس دیگر قرار گیرند، آموزش می دهیم. سپس در فریم جدید، به کمک این طبقه بندی کننده پیکسلهای موجود در داخل بیضی با ابعاد $\sqrt{2}$ برابر بیضی شامل هدف که از الگوریتم تخمین مدل حرکتی بدست آمده است را تست می نماییم. با این روش خطاهای ناشی از مرحله اول تصحیح شده و مرزهای دقیق هدف در هر فریم آشکار خواهد شد. به علاوه بعد از هر چند فریم می توان ورودی طبقه بندی کننده را به روز کرده و دقت الگوریتم ردیابی را افزایش داد. ما در آزمایش ها نشان خواهیم داد که الگوریتم جابجایی میانگین در مقابل حرکت دوربین^۴، پوشیدگی جزئی هدف^۵، شلوغی پس زمینه^۶ و تغییرات در ابعاد^۷ و ظاهر هدف^۸ توانمند است و موفق به ردیابی مؤثر هدف می شود.

^۱ Correlation

^۲ Classifier

^۳ Reference image

^۴ Camera motion

^۵ Partial Occlusion

^۶ Background clutter

^۷ Scale changes

^۸ Appearance changes

اگر فقط از اطلاعات مکانی برای مشخص کردن هدف استفاده شود، تابع شباهت می‌تواند متغیرهای زیادی برای جایگاههای مجاور روی شبکه تصویر داشته باشد و این کار ما را مشکل می‌کند. برای پیدا کردن بیشینه چنین توابعی، بکاربردن روشهای بهینه سازی مبتنی بر گرادیان مشکل است و فقط می‌توان از روشهای جستجوی کلی که زمان زیادی را صرف می‌کند استفاده شود. ما تابع شباهت را به وسیله فیلتر کردن اشیاء با یک هسته ایزوتروپیک در یک حوزه مکانی قانونمند می‌کنیم [۵۲]. از آنجایی که وزنه‌های هسته حامل اطلاعات مکانی پیوسته هستند و در تعریف فضای ویژگی مورد استفاده قرار می‌گیرند، $\hat{p}(y)$ یک تابع هموار در y می‌شود.

۳-۲-۱- مدل هدف

برای ردیابی مؤثر یک هدف مشخص باید تعریف دقیقی از آن داشت، بنابراین با انتخاب یک مدل مناسب از هدف در فریم اول می‌توان توصیف مناسبی از آن شکل بدست آورد. برای این منظور هدف به صورت یک ناحیه بیضی شکل همانطور که در تصویر (۳-۱) مشاهده می‌شود مشخص شده است.



شکل ۳-۱: بیضی هدف در فریم اول (تصویر مرجع)

برای مستقل کردن الگوریتم جابجایی میانگین از تغییرات در ابعاد هدف، همه اهداف در ابتدا به دایره واحد نرمالیزه می شوند. این عمل با تغییر مقیاس ابعاد ردیف و ستون هدف صورت می گیرد. فرض می کنیم $\{x_i^*\}_{i=1 \dots n}$ مختصات های نرمالیزه شده پیکسل در ناحیه تعریف شده به عنوان مدل هدف باشد. نقطه صفر را به عنوان مرکز ناحیه در نظر می گیریم. یک هسته ایزوتروپیک، با یک نمایش $k(x)$ محدب (کوژ)^۱ و کاهشی یکنواخت^۲، وزنه های کمتر را به پیکسل های که دورتر از هدف هستند اختصاص می دهد. توابعی مانند e^{-x} ، $1-x$ از این نوع هستند. به کمک این وزن ها توانمندی در تخمین چگالی افزایش می یابد، زیرا به پیکسل های محیطی^۳ (دور از هدف) که کمتر مورد اعتماد هستند و اغلب تحت تأثیر همپوشانی با اشیاء دیگر یا اختلالات پس زمینه هستند وزن کمتری اختصاص می یابد. تابع $b: R^2 \rightarrow \{1 \dots m\}$ مربوط به پیکسل در جایگاه x_i^* است و مؤلفه هیستوگرام آن در فضای ویژگی کوانتیزه^۴ شده با $b(x_i^*)$ مشخص شده است. احتمال ویژگی $u = 1 \dots m$ در مدل هدف سپس به صورت زیر محاسبه می شود:

$$\hat{q}_u = C \sum_{i=1}^n k\left(\|x_i^*\|^2\right) \delta[b(x_i^*) - u] \quad (2-3)$$

تابع $k(x)$ طوری است که با افزایش فاصله از مرکز کاهش می یابد.

^۱ Convex

^۲ Monotonic decreasing

^۳ Peripheral pixels

^۴ Quantized Feature Space

جاییکه δ تابع دلتا کرونکر^۱ است. ثابت نرمالیزاسیون برای اینکه شرط $\sum_{u=1}^m \hat{q}_u = 1$ را برآورده سازد به صورت:

$$c = \frac{1}{\sum_{i=1}^n k(\|x_i^*\|^2)} \quad (3-3)$$

بدست می آید، زیرا که جمع تابع دلتا برای $u = 1 \dots m$ برابر یک است.

۳-۲-۲- کاندیداهای هدف

فرض می کنیم $\{x_i\}_{i=1 \dots n_h}$ جایگاههای نرمالیزه شده پیکسلهای کاندیدای هدف (بیضی شامل هدف) باشد و y مرکز آن در فریم فعلی باشد (شکل ۳-۲).

با استفاده از همان نمایش $k(x)$ از هسته، اما با پهنای باند h احتمال ویژگی $u = 1 \dots m$ در کاندیدای هدف به وسیله فرمول زیر بیان می شود:

$$\hat{p}_u(y) = C_h \sum_{i=1}^{n_h} k\left(\left\|\frac{y - x_i}{h}\right\|^2\right) \delta[b(x_i) - u] \quad (4-3)$$

جاییکه

$$C_h = \frac{1}{\sum_{i=1}^{n_h} k\left(\left\|\frac{y - x_i}{h}\right\|^2\right)} \quad (5-3)$$

ثابت نرمالیزاسیون است.

^۱ Kronecker Delta Function



شکل ۳-۲: بیضی کاندیدای هدف در فریم دوم از تصویر شکل ۳-۱

از آنجایی که جایگاههای x_i پیکسل در یک شبکه عادی سازماندهی شده اند و y یکی از نودهای شبکه است در نتیجه C_h وابسته به y نیست، بنابراین C_h برای یک هسته داده شده و مقادیر متفاوت از h به طور مقدماتی می تواند محاسبه شود. پهنای باند h پارامتر تعیین مقیاس هدف در فریم جدید بوده و تعداد پیکسلهای مورد توجه قرار گرفته در فرآیند تعیین جایگاه هدف را مشخص می کند.

۳-۳- متریک بر اساس ضریب باتاچاریا

تابع شباهت در حقیقت مبین فاصله میان مدل هدف و کاندیداهای هدف است. برای مقایسه میان اهداف متفاوت، این فاصله باید یک ساختار متریک داشته باشد. ما فاصله میان دو توزیع گسسته را به صورت زیر تعریف می کنیم:

$$d(y) = \sqrt{1 - \rho[\hat{p}(y), \hat{q}]} \quad (۶-۳)$$

در عبارت فوق $\hat{\rho}(y)$ را به صورت زیر:

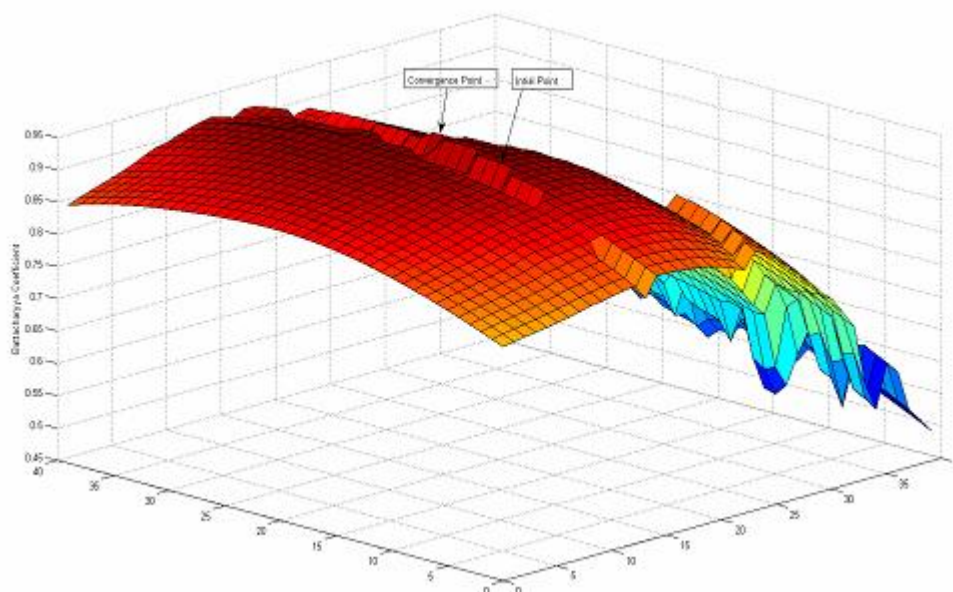
$$\hat{\rho}(y) \equiv \rho[\hat{p}(y), \hat{q}] = \sum_{u=1}^m \sqrt{\hat{p}_u(y) \hat{q}_u} \quad (۷-۳)$$

به عنوان تخمین نمونه از ضریب باتاچاریا میان p و q انتخاب می‌کنیم [۵۳].

در شکل (۳-۳) ضریب باتاچاریا برای یک ناحیه مربعی شکل با ابعاد ۴۰×۴۰ حول مرکز هدف شکل (۱-۳) با کاندیدای هدف در فریم دوم محاسبه شده است و رویه شباهت بدست آمده رسم شده است. نقطه ای از رویه که مقدار بیشینه را دارد مکان هدف در فریم جدید را نشان می‌دهد.

ضریب باتاچاریا یک اندازه از نوع دیورژانس است که یک ارائه هندسی ساده

دارد. این برابر با کسینوس زاویه میان بردارهای واحد m بعدی $(\sqrt{\hat{q}_1}, \dots, \sqrt{\hat{q}_m})^T$ و $(\sqrt{\hat{p}_1}, \dots, \sqrt{\hat{p}_m})^T$ است [۵۴]. ما می‌توانیم (۷-۳) را به عنوان همبستگی نرمالیزه شده میان بردارهای $(\sqrt{\hat{q}_1}, \dots, \sqrt{\hat{q}_m})^T$ و $(\sqrt{\hat{p}_1}, \dots, \sqrt{\hat{p}_m})^T$ بیان کنیم.



شکل ۳-۳: رویه شباهت بین هدف شکل ۳-۱ با کاندیدای هدف در فریم دوم حول ناحیه مربعی شکل با ابعاد 40×40 در مرکز هدف

مقیاس آماری (۳-۶) چندین خاصیت به شرح زیر دارد:

۱- این یک ساختار متریک را وضع می‌کند که می‌تواند به صورت فاصله دو بردار مطرح شود.

۲- این یک بیان هندسی واضح دارد. توجه کنید که متریک های هیستوگرام

$$\text{شرایط } \sum_{u=1}^m \hat{p}_u = 1 \text{ و } \sum_{u=1}^m \hat{q}_u = 1 \text{ را برآورده نمی‌سازند.}$$

۳- از آنجایی که از چگالی های گسسته استفاده می‌کند مستقل از تغییر مقیاس هدف است.

۳-۴- تعیین جایگاه هدف

برای تعیین جایگاه هدف در فریم جاری، فاصله ذکر شده در رابطه (۳-۶) باید به عنوان تابعی از γ مینیمم شود. فرآیند تعیین جایگاه هدف از جایگاه هدف

در فریم قبلی (مدل هدف) شروع شده و همسایگی های هدف را مورد بررسی قرار می دهد [۵۵]. در این قسمت از اطلاعات رنگ به عنوان ویژگی هدف استفاده می شود با این حال در ادامه این فصل، الگوریتم جابجایی میانگین از شاخصه های چندگانه^۱ در تعیین ویژگی ها بهره خواهد گرفت.

در این مرحله فرض می شود اطلاعات زیر موجود است:

(۱) جایگاه هدف در فریم اولیه کاملاً مشخص است.

(۲) ویژگی های هدف استخراج شده است و مدل هدف از تصویر مرجع تشکیل گردیده است.

۳-۴-۱- مینیم سازی فاصله

مینیم سازی فاصله در رابطه (۳-۶) معادل با بیشینه سازی ضریب پاتاچاریا $P(y^{\wedge})$ است. جستجو برای جایگاه جدید هدف در فریم جاری از محل y_0^{\wedge} که مکان هدف در فریم قبلی است شروع می شود.

بنابراین در ابتدا احتمالات $\{P_u^{\wedge}(y_0^{\wedge})\}_{u=1, \dots, m}$ کاندیدهای هدف در جایگاه y_0^{\wedge} فریم جاری باید محاسبه شوند. با استفاده از بسط تیلور حول مقادیر $P_u(y_0^{\wedge})$ ، تقریب خطی از ضریب پاتاچاریا (۳-۷) با کمی تغییرات در آن به صورت زیر به دست آمده است:

$$\rho(\hat{p}(y), \hat{q}) \cong \frac{1}{2} \sum_{u=1}^m \sqrt{p_u^{\wedge}(y_0^{\wedge}) q_u^{\wedge}} + \frac{1}{2} \sum_{u=1}^m p_u^{\wedge}(y) \sqrt{\frac{q_u^{\wedge}}{p_u^{\wedge}(y_0^{\wedge})}} \quad (۳-۸)$$

^۱ Multiple Features

تقریب زمانی قابل قبول است که کاندیدای هدف $\{P_u^\wedge(y)\}_{u=1,\dots,m}$ تغییر شدیدی نسبت به حالت اولیه $\{P_u^\wedge(y_0^\wedge)\}_{u=1,\dots,m}$ نداشته باشد که در بیشتر موارد این یک فرض قابل قبول میان فریم های متوالی است. با فراخوانی رابطه (۳-۴) خواهیم داشت:

$$\rho[\hat{P}(y), \hat{q}] \approx \frac{1}{2} \sum_{u=1}^m \sqrt{p_u^\wedge(y_0^\wedge) q_u^\wedge} + \frac{c_h}{2} \sum_{i=1}^{x_h} w_i k\left(\left\|\frac{y-x_i}{h}\right\|^2\right) \quad (۹-۳)$$

در رابطه فوق w_i برابر است با:

$$w_i = \sum_{u=1}^m \sqrt{\frac{q_u^\wedge}{p_u^\wedge(y_0^\wedge)}} \delta[b(x_i) - u] \quad (۱۰-۳)$$

است.

بنابراین، برای مینیمم سازی فاصله (۳-۶) جمله دوم در (۳-۹) باید بیشینه شود، جمله اول مستقل از y است. جمله دوم تخمین چگالی محاسبه شده با نمایش هسته (کرنل) $k(x)$ در نقطه y در فریم جاری با استفاده از داده های وزن دهی شده به وسیله w_i در رابطه (۳-۱۰) را ارائه می دهد. محاسبه این چگالی در همسایگی محلی با استفاده از روش جابجایی میانگین انجام می گیرد. در این روش هسته به طور برگشتی از جایگاه فعلی y_0 به جایگاه y_1^\wedge بر حسب رابطه:

$$y_1^\wedge = \frac{\sum_{i=1}^{n_h} x_i w_i g\left(\left\|\frac{y_0^\wedge - x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n_h} w_i g\left(\left\|\frac{y_0^\wedge - x_i}{h}\right\|^2\right)} \quad (۱۱-۳)$$

از جایگاه قبلی خود به جایگاه فعلی حرکت می کند.

جاییکه $g(x) = -k'(x)$ به عنوان مشتق $k(x)$ فرض شده و برای همه $x \in [0, \infty]$ به جز برای مجموعه محدودی از نقاط وجود دارد.

فرض های اولیه:

$$\rho[\hat{p}(y), \hat{q}] \text{ بیشینه سازی ضریب باتاچاریا}$$

مدل هدف $\{q_u^{\wedge}\}_{u=1, \dots, m}$ و جایگاه آن y_o^{\wedge} در فریم قبلی داده شده است.

* الگوریتم کامل ردیاب جابجایی میانگین برای تعیین جایگاه هدف

(شیء) به صورت زیر معرفی شده است :

۱- جایگاه اولیه هدف در فریم جاری را y_o^{\wedge} در نظر گرفته شده، سپس

$$\rho[\hat{p}(y_o^{\wedge}), \hat{q}] = \sum_{u=1}^m \sqrt{p_u^{\wedge}(y_o^{\wedge}) q_u^{\wedge}} \text{ را محاسبه کرده و رابطه } \{p_u^{\wedge}(y_o^{\wedge})\}_{u=1, \dots, m}$$

ارزیابی می شود.

۲- وزنها $\{w_i\}_{i=1, \dots, n_h}$ بر حسب رابطه (۳-۱۰) راه اندازی می گردند.

۳- جایگاه بعدی کاندیدای هدف بر حسب (۳-۱۱) پیدا می شود.

$$\rho[\hat{p}(y_1^{\wedge}), \hat{q}] = \sum_{u=1}^m \sqrt{p_u^{\wedge}(y_1^{\wedge}) q_u^{\wedge}} \text{ را محاسبه کرده و } \{p_u^{\wedge}(y_1^{\wedge})\}_{u=1, \dots, m}$$

ارزیابی خواهند شد.

۵- زمانیکه $\rho[\hat{p}(y_1^{\wedge}), \hat{q}] < \rho[\hat{p}(y_o^{\wedge}), \hat{q}]$ آنگاه $\frac{1}{2}(y_o^{\wedge} + y_1^{\wedge})$ و y_1^{\wedge}

سپس $\rho[\hat{p}(y_1^{\wedge}), \hat{q}]$ دوباره ارزیابی می شود.

۶- اگر $\|y_1^{\wedge} - y_o^{\wedge}\| < \varepsilon$ آنگاه می ایستیم وگرنه $y_1^{\wedge} \leftarrow y_o^{\wedge}$ و به مرحله ۲ بر

می گردیم.

۳-۵- انتخاب نوع هسته

در پیاده سازی الگوریتم تخمین مکان هدف کرنل اپانچینکف^۱ به صورت:

$$k(x) = \begin{cases} \frac{1}{2} c_d^{-1} (d+2)(1-x) & \text{if } x \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (12-3)$$

برای استفاده پیشنهاد شده است.

با استفاده از این نمایش کرنل، تابع $g(x)$ یک عدد ثابت شده و رابطه (۳-۳) (۱۱) به رابطه زیر خلاصه می شود:

$$y_1^{\wedge} = \frac{\sum_{i=1}^{n_h} x_i w_i}{\sum_{i=1}^{n_h} w_i} \quad (13-3)$$

که یک میانگین وزنی ساده است.

۳-۵-۱- استفاده از تبدیل فاصله به عنوان هسته

از آنجایی که تمام اهداف مورد ردیابی شکل هندسی منظمی ندارند در این صورت مدل سازی تمام آنها با یک کرنل ثابت مانند $1-x$ مطلوب نیست. اینجا به معرفی یک تبدیل فاصله می پردازیم که با استفاده از آن می توان اشکال گوناگون را مدل کرد [۵۶]. به کمک این تبدیل پیکسلهای شکل مورد ردیابی بر اساس فاصله از مرکز جرم آن به صورت مناسبی وزن دهی می شوند که این امر به بهبود عملکرد ردیابی در فریم های متوالی کمک می کند.

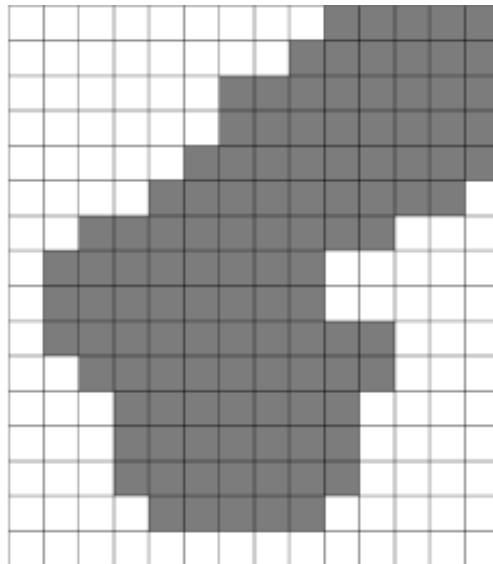
فرض کنید P مطابق شکل (۳-۴) یک تصویر باینری تعریف شده روی شبکه G است به طوری که:

^۱ Epanechnikov

$$\langle \bar{P} \rangle = \{p: p \in G \wedge P(p) = 0\}$$

$$\langle P \rangle = \{p: p \in G \wedge P(p) = 1\}$$

زیر مجموعه های در نظر گرفته شده بر روی شبکه G هستند. برای هر معیار فاصله d_x ، تبدیل فاصله d_x از P برابر است با هر پیکسل p متعلق به $\langle P \rangle$ که در آن d_x فاصله p تا $\langle \bar{P} \rangle$ است.



شکل ۳-۴: یک تصویر باینری تعریف شده روی شبکه G

الگوریتم دو مرحله محاسبه تبدیل های فاصله d_4 و d_8

تبدیل های فاصله d_4 و d_8 مربوط به جسم P که از آنها استفاده می کنیم با به کار گیری یک سری عملیات محلی و اسکن دو مرحله ای شبکه G قابل محاسبه هستند. هر عملیات محلی به هر پیکسل p یک مقدار جدید که وابسته به مقادیر پیشین همسایگی های آن است نسبت می دهد. برای هر $p \in G$ فرض می کنیم $B(p)$ مجموعه ای از پیکسلهای موجود در همسایگی چهار تایی یا هشت تایی p است که هنگام اسکن شبکه G پیش از p قرار می گیرند. بنابراین

اگر p دارای مختصات (x, y) باشد، B پیکسلهایی با مختصات $(x, y - 1)$ و $(x - 1, y)$ برای همسایگی چهارتایی و پیکسلهایی با مختصات $(x - 1, y + 1)$ ، $(x - 1, y)$ ، $(x - 1, y - 1)$ و $(x, y - 1)$ را به همسایگی هشت تایی در بر می گیرد.

*اسکن اول

الگوریتم محاسبه مقادیر هر پیکسل مربوط به جسم P در اسکن اول به صورت زیر است:

$$f_1(p) = \begin{cases} 0 & \text{if } p \in (\bar{P}) \\ \min \{f_1(q) + 1 : q \in B(p)\} & \text{if } p \in (P) \end{cases} \quad (۱۴-۳)$$

توجه کنید که برای هر p ، f_1 برای تمام q های متعلق به $B(p)$ محاسبه شده است. دقت شود در مواردی که p روی ردیف بالا یا در سمت چپ G قرار دارد، بعضی از این q ها بیرون G با مقدار $f_1 = 0$ هستند. در شکل (۳-۵) وزنه‌های اختصاص داده شده به جسم برای همسایگی های چهارتایی و هشت تایی را پس از اسکن اول مشاهده می نمایید.

اگر چه وزن دهی پیکسل های جسم با استفاده از تبدیل فاصله می تواند به بهبود عملکرد الگوریتم تعیین مکان هدف کمک کند، اما دو مشکل عمده باعث شده است که ما در پیاده سازی الگوریتم از آن استفاده نکنیم:

- ۱- چون شیء مورد ردیابی در دنباله تصویر ویدئویی تغییر نمود دارد بنابراین در هر فریم نیاز به تعریف دوباره جسم با استفاده از تبدیل فاصله داشته که این امر بار محاسباتی بالایی داشته و زمان اجرای الگوریتم بسیار بالا می رود.
- ۲- اکثر اشیائی که ما در این پروژه قصد ردیابی آنها را داریم مانند دست، ماشین و اشخاص ماهیتی بیضی شکل داشته و در نتیجه به راحتی با استفاده از بیضی قابل ارائه هستند.

۳-۶- انتخاب فضای رنگ

از آنجا که در فرایند ردیابی با اجسامی با رنگ های متفاوت روبرو هستیم که آنها نیز به نوبه خود در پس زمینه هایی با رنگ های متفاوت سیر می کنند، انتخاب یک فضای رنگ بهینه برای مدل مرجع می تواند در بهبود عملکرد الگوریتم ردیابی مؤثر باشد [۵۷]. در این قسمت از ضریب باتاچاریا که در بخش (۳-۳) به تفصیل مورد بررسی قرار گرفته استفاده می شود. به این صورت که مدل هدف مرجع q را برای هر یک از مؤلفه های رنگی فضاها رنگی ذکر شده در فصل ۲ محاسبه می نماییم. همچنین مدل پس زمینه را که شامل تمام پیکسل های غیر هدف است و با b نمایش داده می شود را نیز برای هر یک از مؤلفه های رنگی فضاها رنگی ذکر شده در فصل ۲ محاسبه می نماییم. ما فاصله میان دو توزیع گسسته را به صورت زیر تعریف می کنیم:

$$d = \sqrt{1 - \rho[\hat{q}, \hat{b}]} \quad (۱۶-۳)$$

در عبارت فوق:

$$\hat{\rho} \equiv \rho \left[\hat{q}, \hat{b} \right] = \sum_{u=1}^m \sqrt{\hat{q}_u, \hat{b}_u} \quad (17-3)$$

را به عنوان تخمین نمونه از ضریب باتاچاریا میان q و b انتخاب می‌کنیم.

در ابتدا مدل هدف از تصویر مرجع مثلاً شکل (۱-۳) را تشکیل داده و هیستوگرام های ۱۶ تایی (تعداد مؤلفه های هیستوگرام) را برای پیکسلهای داخل بیضی برای هر یک از اجزای فضاهاى رنگی محاسبه کرده و همچنین هیستوگرام های ۱۶ تایی را برای همان اجزا از پس زمینه محاسبه می‌کنیم. سپس فاصله میان q و b را بدست می‌آوریم. مؤلفه رنگی که بیشترین فاصله را به ما بدهد را به عنوان اولین مؤلفه فضای رنگی انتخاب می‌کنیم. به عنوان مثال برای تصویر شکل (۱-۳) مؤلفه رنگی R از فضای رنگی RGB را که بیشترین فاصله را بین مدل مرجع و پس زمینه می‌دهد به عنوان اولین مؤلفه انتخاب می‌کنیم. حال مؤلفه R با هریک از مؤلفه های دیگر فضای رنگ در نظر گرفته و هیستوگرام های ۱۶×۱۶ تایی (تعداد مؤلفه های هیستوگرام) را برای مدل هدف و پس زمینه محاسبه می‌نماییم. در اینجا هیستوگرام دو بعدی R با H از فضای رنگ HSV بیشترین فاصله را بین هدف و پس زمینه می‌دهد بنابراین H را به عنوان دومین مؤلفه هیستوگرام در نظر می‌گیریم. حال مؤلفه های R و H با هریک از مؤلفه های دیگر فضای رنگ در نظر گرفته و هیستوگرام های ۱۶×۱۶×۱۶ تایی (تعداد مؤلفه های هیستوگرام) را برای مدل هدف و پس زمینه محاسبه می‌کنیم. در اینجا هیستوگرام سه بعدی R، H با X از فضای رنگ XYZ بیشترین فاصله را بین هدف و پس زمینه می‌دهد بنابراین X را به عنوان سومین مؤلفه هیستوگرام در نظر می‌گیریم. فضای رنگ RHX به عنوان فضای رنگ بهینه^۱ برای اعمال الگوریتم ردیابی در نظر گرفته می‌شود. جدول ۱-۳ فاصله میان هدف و پس زمینه را پس از محاسبه هیستوگرام های یک بعدی و دو بعدی و سه بعدی، ۱۸ مؤلفه رنگی ذکر شده جدول را نشان می‌دهد. کل تعداد دفعات

^۱ Optimum

محاسبه هیستوگرام برای این ۱۸ مؤلفه رنگی برابر $۱۸+۱۷+۱۶=۵۱$ می باشد. همانطور که گفته شد روش ذکر شده یک روش کاملاً بهینه نیست در حالیکه اگر می خواستیم از روش کاملاً بهینه استفاده کنیم تعداد دفعات محاسبات هیستوگرام برای انتخاب سه مؤلفه رنگی از مجموع ۱۸ مؤلفه برابر بود با $\binom{18}{3} = 816$ که بار محاسباتی آن بسیار زیاد می شد. ضمناً نتایج عملی نشان می دهد که تفاوت چندانی در نتایج این دو روش وجود ندارد.

مرحله اول	فاصله میان b و q	مرحله دوم	فاصله میان q و b	مرحله سوم	فاصله میان q و b
R	.72				
G	.55	RG	.75	RHG	.82
B	.63	RB	.74	RHB	.84
H	.66	RH	.81		
S	.44	RS	.72	RHS	.87
V	.51	RV	.76	RHV	.80
L	.7	RL	.71	RHL	.84
a	.43	Ra	.71	RHa	.81
b	.54	Rb	.72	RHb	.88
Y	.61	RY	.77	RHY	.80
U	.43	RU	.73	RHU	.81
V	.56	RV	.75	RHV	.86
Y	.71	RY	.77	RHY	.88
Cb	.33	RCb	.74	RHCb	.81
Cr	.37	RCr	.73	RHCr	.84
X	.68	RX	.79	RHX	.93
Y	.56	RY	.76	RHY	.86
Z	.34	RZ	.71	RHZ	.82

جدول ۳-۱ فاصله میان هدف و پس زمینه پس از محاسبه هیستوگرام های یک بعدی و دو بعدی و سه بعدی

۱۸ مؤلفه رنگی

۳-۷- ویژگی های چند گانه

در این قسمت چگونگی مدلسازی ویژگی های مجزا (ذکر شده در فصل ۲) را شرح داده و روش مورد نظر برای استفاده از این ویژگی های چند گانه را برای بهبود تخمین مدل حرکت ذکر شده در قسمت قبل را شرح می دهیم [۵۸-۵۹].

از آنجایی که تمام این مدلها مبتنی بر هیستوگرام بوده می توان آنها را با یکدیگر ترکیب نمود و یک مدل کلی برای توصیف ناحیه داخل بیضی تعریف کرد.

۳-۷-۱- مدل سازی شاخصه رنگ

از آنجایی که هیستوگرام های رنگی در مقابل پوشیدگی جزئی، تغییر اندازه هدف و تغییر جهت هدف مقاوم هستند از این رو الگوریتم های مبتنی بر آن می توانند به طور مؤثر و دقیقی از عهده تغییر در ظاهر اجسام غیر صلب و پیچیدگی پس زمینه بر آیند. با این حال در مواقعی که چندین جسم هم رنگ در پس زمینه وجود دارند این الگوریتم ها از عهده ردیابی مؤثر جسم بر نمی آیند.

همان طور که در بخش (۳-۲-۱) گفته شده است مدل مرجع را برای

شاخصه رنگ تشکیل داده و هیستوگرام وزن دار \hat{q}_{u_c} را برای آن به صورت زیر محاسبه می کنیم:

$$\hat{q}_{u_c} = C_c \sum_{i=1}^n k \left(\|x_i^*\|^2 \right) \delta [b_c(x_i^*) - u_c] \quad (۳-۱۸)$$

که در آن $\{x_i^*\}_{i=1, \dots, n}$ مکان نرمالیزه شده پیکسلهای های داخل ناحیه هدف می باشد.

ضریب C_c به منظور اینک که $\sum_{u_c=1}^{m_c} \hat{q}_{u_c} = 1$ تعریف شده است. تابع

$b_c: R^2 \rightarrow \{1 \dots m_c\}$ مربوط به پیکسل در جایگاه x_i^* است و مؤلفه

هیستوگرام آن در فضای ویژگی کوانتیزه شده $b_c(x_i^*)$ مشخص شده است. با استفاده از همان نمایش هسته $k(x)$ هیستوگرام وزن دار $\hat{p}_{u_c}(y)$ برای فضای ویژگی $u_c = 1 \dots m_c$ برای کاندیدای هدف به صورت زیر محاسبه می شود:

$$\hat{p}_{u_c}(y) = C_c \sum_{i=1}^{n_h} k\left(\left\|\frac{y-x_i}{h}\right\|^2\right) \delta[b_c(x_i) - u_c] \quad (19-3)$$

انتخاب فضای رنگ با استفاده از روش ذکر شده در قسمت قبل بدست می آید.

۳-۷-۲- مدل سازی شاخصه لبه

ویژگی های لبه ابزار مفیدی برای مدل سازی شیء مورد ردیابی هستند. لبه ها با استفاده از یک هیستوگرام مبتنی بر جهت بدست آمده از آن توصیف می شوند. یک پیکسل در ناحیه بیضی شکل تصویر مرجع با شدت I در نظر گرفته شده است. در این صورت تصویر لبه با محاسبه گرادیان $\partial I / \partial x$ و $\partial I / \partial y$ در جهات x و y قابل محاسبه است. در این حالت جهات لبه θ به صورت زیر بدست می آیند:

$$\theta_{u_e} = \tan^{-1}\left(\frac{\partial I / \partial y}{\partial I / \partial x}\right) \quad (20-3)$$

بنابراین هیستوگرام وزن دار هدف \hat{q}_{u_e} برای مدل هدف به صورت زیر بدست می آید:

$$\hat{q}_{u_e} = C_e \sum_{i=1}^n k\left(\|x_i^*\|^2\right) \delta[b_e(x_i^*) - u_e] \quad (21-3)$$

که در آن $\{x_i^*\}_{i=1, \dots, n}$ مکان نرمالیزه شده پیکسلهای های داخل ناحیه هدف می باشد. ضریب C_e به منظور اینکه $\sum_{u_e=1}^{m_e} \hat{q}_{u_e} = 1$ تعریف شده است. تابع $b_e: R^2 \rightarrow \{1 \dots m_e\}$ مربوط به پیکسل در جایگاه x_i^* است و مؤلفه

هیستوگرام آن در فضای ویژگی کوانتیزه شده $b_e(x_i^*)$ مشخص شده است. با استفاده از همان نمایش هسته $k(x)$ هیستوگرام وزن دار $\hat{p}_{u_e}(y)$ برای فضای ویژگی $u_e = 1 \dots m_e$ برای کاندیدای هدف به صورت زیر محاسبه می شود:

$$\hat{p}_{u_e}(y) = C_e \sum_{i=1}^{n_h} k\left(\left\|\frac{y-x_i}{h}\right\|^2\right) \delta[b_e(x_i) - u_e] \quad (22-3)$$

ما از یک هیستوگرام ۸ مؤلفه ای یک بعدی برای جهت لبه $0 < \theta < 180$ استفاده می کنیم.

۳-۷-۳- مدل سازی شاخصه بافت

اگر چه تعریف واحدی از بافت وجود ندارد، عموماً بافت ترتیب مکانی از سطوح پیکسل در یک تصویر را بیان می کند که کمیتی تصادفی یا پرپودیک یا شامل هر دو می باشد. روش به کار رفته برای استخراج ویژگی های بافت تصویر روش تبدیل موجک دو بعدی گسسته می باشد که به تفصیل در فصل ۲ توضیح داده شده است. در اینجا فرایند تولید یک بردار ویژگی برای ارائه بافت یک ناحیه از تصویر شرح داده می شود. در ابتدا یک ناحیه مستطیل شکل با طول و عرض برابر با قطر بزرگ و کوچک هدف در مرکز بیضی در نظر گرفته و سپس تبدیل موجک سه مرحله ای برای تولید ۱۰ کانال برای هر پنجره رنگی R,G,B تصویر مرجع به کار می رود. بافت مورد نظر برای m^{th} کانال به صورت زیر محاسبه می شود:

$$t_m = 1/LW \sum_{i=1}^L \sum_{j=1}^W |\varpi(i, j)| \quad (23-3)$$

که در آن L و W ابعاد کانال بوده و $\varpi(i, j)$ ضریب موجک در جایگاه i و j است. با این روش بافت ناحیه برای هر پنجره رنگی R, G, B به وسیله یک بردار شامل ۱۰ عنصر تعریف می شود، برای مثال برای کانال قرمز به صورت $\hat{q}_{u_i^R} = \{t_1^R, \dots, t_{10}^R\}$ بیان می شود. در نهایت یک بردار ویژگی یا به عبارت دیگر

هیستوگرام یک بعدی شامل ۳۰ مؤلفه به صورت

$$\hat{q}_{u_i} = C_t \{t_1^R, \dots, t_{10}^R, t_1^G, \dots, t_{10}^G, t_1^B, \dots, t_{10}^B\}$$

برای همه پنجره های رنگی G, B ناحیه مورد نظر از تصویر مرجع به دست می آید. ثابت C_t به منظور اینکه $\sum_{u_i=1}^{m_i} \hat{q}_{u_i} = 1$ تعریف شده است جایی که $m_i = 30$ است.

به طور مشابه هیستوگرام $p_{u_i}(y)$ برای کاندیدای هدف به صورت

$$\hat{p}_{u_i}(y) = C_t \{t_1^R(y), \dots, t_{10}^R(y), t_1^G(y), \dots, t_{10}^G(y), t_1^B(y), \dots, t_{10}^B(y)\}$$

محاسبه می شود.

۳-۸- آموزش آماری

در این قسمت به معرفی یک طبقه بند (classifier) برای تعیین مرزهای دقیق هدف بعد از تخمین مکان آن که با استفاده از روش ذکر شده (جابجایی میانگین) در بخش قبل از این فصل بدست آمده است، می پردازیم. چنانچه در شکل (۳-۸) دیده می شود در فریم اول توسط اپراتور دور هدف را مشخص کرده و پیکسلهای آن را به عنوان پیکسلهای شیء مورد ردیابی در نظر می گیریم. حال بیضی با ابعاد تقریبی $\sqrt{2}$ را برابر بیضی هدف در نظر گرفته و پیکسلهای داخل این بیضی که متعلق به هدف نمی باشد را به عنوان پیکسلهای غیر هدف می گیریم. در واقع در اینجا با دو نوع کلاس روبرو هستیم:

۱- کلاس اول که همان هدف است.

۲- کلاس دوم که همان پیکسلهای غیر هدف است.

در ابتدا یک طبقه بندی کننده دو کلاسه با استفاده از ویژگی های استخراجی از فریم اول آموزش داده شده و سپس پیکسلهای داخل بیضی بدست آمده با استفاده از روش تخمین مدل حرکتی ذکر شده در بخش قبل در فریم دوم با استفاده از این طبقه بندی کننده تست می شوند. به طوری که پیکسلهای

داخل بیضی را به دو گروه هدف و غیر هدف تقسیم کرده و مرزهای دقیق شیء را به ما می دهد. مزیت این روش این است که اگر خطای روش تخمین مدل حرکتی زیاد باشد به کمک طبقه بندی کننده موفق به تصحیح این خطا شده و از آنجایی که مرز دقیق هدف در فریم های متعدد بدست آمده است امکان بهره گیری از تبدیل فاصله ذکر شده در فصل قبل را برای ما فراهم آورده که این امر منجر به ردیابی دقیق تر شیء مورد نظر می گردد.



شکل ۳-۸: بیضی با رنگ زرد با ابعاد $\sqrt{2}$ برابر بیضی سیاه رنگ. ناحیه داخل منحنی قرمز رنگ مشخص کننده هدف و ناحیه بین بیضی با رنگ زرد و منحنی قرمز رنگ مشخص کننده ناحیه غیر هدف می باشد.

ذکر این نکته ضروری است که اگرچه ما از ماشین های بردار پشتیبان به عنوان طبقه بندی کننده استفاده کرده ایم ولی نوع طبقه بندی کننده در الگوریتم جابجایی میانگین چندان مهم نیست و در دقت این روش ردیابی چندان تأثیری ندارد.

آموزش آماری را می‌توان بصورت یک مساله تقریب تابع با استفاده از کمینه‌سازی ریسک، تعریف نمود [۶۷]. به این ترتیب می‌خواهیم تابعی مانند $f(x, a)$ پیدا کنیم که به ازای بردارهای مشاهده ورودی x ، خروجی یا برچسب مناسب y ، را به ما بدهد. در رابطه $f(x, a)$ ، a مجموعه‌ای از پارامترهاست که در حقیقت هدف آموزش پیدا کردن مقادیر مناسب برای این پارامترها، با استفاده از یک مجموعه بردارهای مشاهده برچسب زده شده محدود که در اختیار داریم $\{(x_1, y_1), (x_2, y_2), \dots, (x_M, y_M)\}$ می‌باشد. در این حالت به عمل تخمین تابع، آموزش تحت نظارت^۱ گفته می‌شود. برای این منظور یک تابع ریسک به شکل زیر، کمینه سازی می‌شود:

$$R(a) = \int L(y, f(x, a)) dP(x, y) \quad (۲۴-۳)$$

که در رابطه فوق $L(y, f(x, a))$ یک تابع تلف^۲ می‌باشد که مبین میزان تفاوت موجود بین y و تقریب بدست آمده برای y از تابع $f(x, a)$ می‌باشد. یک تابع تلف متداول، مربع تفاضلات بین مقادیر واقعی و تخمین زده شده می‌باشد (رابطه ۳-۲۵). در واقع ریسک عبارتست از مجموع مقادیر تلفات برای تمام مقادیر ممکن x و y ، که چون این مقادیر بصورت متغیرهای تصادفی هستند، می‌بایست مقادیر تلفات روی تابع چگالی احتمال توأم این متغیرها انتگرال گرفته شود.

$$L(y, f(x, a)) = [y - f(x, a)]^2 \quad (۲۵-۳)$$

همانطور که گفته شد برای بدست آوردن تخمینی از توابع طبقه بندی، به یکسری از مشاهدات برچسب زده شده نیاز داریم، که در این حالت به تخمین تابع، آموزش تحت نظارت گفته می‌شود. در مواردی که برچسبها مشخص

^۱ Supervised Learning

^۲ Loss Function

نیاشند و یا تمام مشاهدات در واقع دارای یک برچسب باشند، از آموزش بدون نظارت^۱ استفاده می‌شود. در این موارد، هدف خوشه بندی مشاهدات و یا تخمین تابع چگالی احتمال مشاهدات می‌باشد. برای تخمین تابع چگالی احتمال معمولاً تابع تلف بصورت زیر تعریف می‌شود:

$$L(y, f(x, a)) = -\log(y - f(x, a)) \quad (۲۶-۳)$$

برای بوجود آوردن یک سیستم که بصورت آماری آموزش می‌بیند، موارد زیر را می‌توان مد نظر قرار داد:

۱- نوع تابع مورد استفاده، که معمولاً یک تابع طبقه بندی یا تابع چگالی احتمال می‌باشد.

نحوه پیاده سازی تابع مربوطه، از قبیل شبکه‌های عصبی چند لایه^۲، شبکه‌های RBF^۳، توابع تمایز خطی و غیر خطی و یا توابع توزیع گوسی.

۲- روش یا الگوریتم آموزش، از قبیل روش‌های مختلف بهینه سازی که برای کمینه سازی یک تابع تخمین تا حداقل ریسک ممکن، از آنها استفاده می‌شود. معمولاً این بهینه سازی یاروی تابع ریسک تجربی و یا یک تابع ریسک ساختاری انجام می‌شود که در ادامه در مورد آن صحبت می‌شود. انتخاب نوع تابع مناسب برای پیاده سازی به موارد زیر بستگی دارد:

الف) طبیعت توابع چگالی احتمال مشاهدات

ب) میزان داده‌های آموزشی موجود

^۱ Unsupervised

^۲ Multi-Layer Perceptrons

^۳ Radial Basis Function

بطور کلی، توابع با پارامترهای کم و ساده، بایاس زیادی در تقریب خود خواهند داشت در مقابل توابعی که دارای پارامترهای زیادی هستند می‌توانند محدوده وسیع‌تری از توابع را تقریب بزنند ولی دارای واریانس تقریب بیشتری خواهند بود. لذا انتخاب تابع مناسب که هم بایاس کم و هم واریانس قابل قبول داشته باشد نیاز به نوعی موازنه خواهد داشت [۶۸]. به طور دقیق‌تر نتایج تئوری آموزش آماری [۶۷]، نشان می‌دهند که آنچه که باید مد نظر قرار گیرد، تعداد پارامترهای تابع نیست بلکه ظرفیت تابع است. ظرفیت تابع نشان می‌دهد که چه مقدار داده برای آموزش یک تابع مورد نیاز است. بنابر این با توجه به حجم داده آموزشی که در اختیار داریم باید توابعی با ظرفیت مناسب را اختیار کنیم.

۳-۸-۱- طبقه بندی خطی^۱

اگر تابعی خطی وجود داشته باشد که بتواند اطلاعات را بدون خطا طبقه بندی کند (همه نقاط دیتا که متعلق به یک کلاس هستند در یک طرف خط یا صفحه یا ابر صفحه قرار بگیرند)، اصطلاحاً گفته می‌شود که آن اطلاعات بصورت خطی قابل جداسازی هستند. در مواردی که اطلاعات بصورت خطی قابل جداسازی باشند یا با تقریب خوبی بتوان آنها را بصورت خطی جداسازی کرد، از طبقه بندی کننده‌های خطی استفاده می‌شود، حسن این طبقه بندی کننده‌ها در سادگی آنهاست و به راحتی قابل پیاده سازی هستند این طبقه بندی کننده‌ها را به صورت زیر می‌توان نمایش داد:

$$f(x) = (\mathbf{w} \cdot x) + b, \quad \mathbf{w} \in \mathbb{R}^N, \quad b \in \mathbb{R} \quad (۲۷-۳)$$

که در رابطه فوق \mathbf{w} یک بردار N بعدی می‌باشد. هنگامیکه ابعاد اطلاعات $N = 2$ باشد، رابطه فوق با صفر قرار دادن طرف چپ معادله، بیانگر یک صفحه و

^۱ Linear Classification

هنگامیکه ابعاد اطلاعات $N > 2$ باشد، بیانگر یک ابر صفحه^۱ می باشد. موقعیت و جهت این ابر صفحه با تنظیم مجموعه پارامترهای $a = [w \ b]$ ، مشخص می شود. در رابطه (۳-۳۹)، w در واقع بردار نرمال ابر صفحه مذکور می باشد.

۳-۹- ماشینهای بردار پشتیبان

روشهای کلاسیک متداول برای آموزش مانند الگوریتمهای آموزشی در شبکه های عصبی عموماً بر پایه حداقل کردن خطا بر روی مجموعه داده های آموزشی می باشد که این روشها به حداقل کردن ریسک عملی^۲ موسوم هستند. برخلاف این روشها، ماشینهای بردار پشتیبان مبتنی بر حداقل کردن خطای ساختاری^۳ است [۳۰]. ماشینهای بردار پشتیبان دارای قدرت تعمیم پذیری بالاتری نسبت به بقیه روشهای یادگیری برای داده های غیر آموزشی (داده های تست) می باشد و با حداقل کردن یک حد بالا که از مجموع خطاهای آموزشی بعلاوه کمیتی که وابسته به ابعاد VC ^۴ مربوط به خطای تعمیم می باشد بدست می آید. تئوری VC یا همان تئوری یادگیری آماری بیان می کند که به ازای هر رده از توابعی که در ماشین یادگیری استفاده می شود، یک ظرفیت محدود و مشخص از حجم داده های آموزشی را می توان آموزش داد. این حجم از داده های آموزشی به ظرفیت VC موسوم است.

۳-۹-۱- حداقل کردن ریسک عملی (ERM)

فرض کنید که یک ماشین یادگیری با پارامتر قابل تنظیم λ داریم. در زیر به توضیح این موضوع می پردازیم که چگونه این ماشین پارامتر λ را تنظیم می

^۱ Hyper-Plane

^۲ Empirical Risk Minimization

^۳ Structural Risk Minimization

^۴ Vapnic-Cher Theorem

کند تا سیستم تناظر $x \rightarrow y$ را آموزش ببیند. این تناظر که به صورت $x \rightarrow f(x, \lambda)$ نمایش داده می شود به تعریف یک ماشین یادگیری خاص منجر می شود. کارایی این ماشین به وسیله امید ریاضی از خطای تست، به صورت نشان داده شده در معادله (۳-۲۴) قابل اندازه گیری می باشد.

$$R(\lambda) = \int E(y, f(x, \lambda)) dP(x, y) \quad (۳-۲۸)$$

این مقدار به نام ریسک واقعی^۱ شناخته می شود. این عبارت حداقل به یک تخمین از $P(x, y)$ نیاز دارد که در بسیاری از کاربردهای کلاس بندی قابل دسترسی نیست. بنابراین مجبور به محاسبه مقدار ریسک عملی به وسیله معادله (۳-۲۹) می شویم. این عبارت تنها در محاسبه میانگین خطا بر روی داده های تست قابل دسترسی می باشد.

$$R_{\text{emp}}(\lambda) = \frac{1}{L} \sum_{i=1}^L E(y, f(x, \lambda)) \quad (۳-۲۹)$$

اکثر روشهای متداول برای آموزش عموماً بر پایه حداقل کردن ریسک عملی با استفاده از تخمین ML^۲ برای پارامتر λ پایه گذاری شده اند. الگوریتم ممکن است به حالت آموزش بیش از اندازه بیانجامد.

۳-۹-۲- حداقل کردن ریسک ساختاری (SRM)

در مقابل ERM، هدف حداقل کردن ریسک ساختاری پیدا کردن یک ماشین یادگیری است که یک مصالحه میان مقدار کم ریسک عملی و ظرفیت به دست دهد. برای به دست آوردن این هدف دو مشکل اساسی در مقابل داریم. حداقل کردن ریسک ساختاری به یک کمیت قابل اندازه گیری برای ظرفیت ماشین یادگیری و یا حداقل یک باند بالا برای آن نیاز دارد. یک الگوریتم برای

^۱ Actual Risk

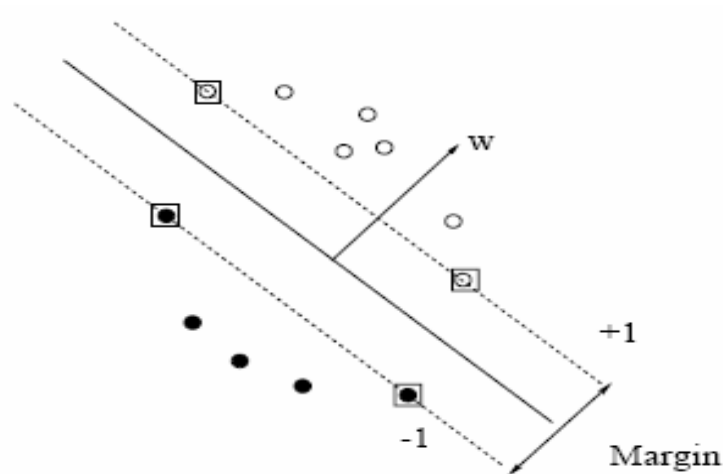
^۲ Maximum Likelihood

انتخاب ماشین یادگیری مطلوب منطبق با حداقل کردن ریسک ساختاری مورد نیاز است. در اینجا یک راه حل کلی برای این مساله وجود دارد که یک ماشین یادگیری با ظرفیت متغیر و یک الگوریتم یادگیری مربوط به آن را در نظر می-گیریم و سپس ظرفیت ماشین و خطای عملی را به صورت همزمان حداقل می-کنیم.

۳-۹-۳- ماشینهای بردار پشتیبان خطی در حالت دو کلاسی با الگوی

جدایی پذیر

ایده اصلی در ماشینهای بردار پشتیبان تشکیل یک صفحه مجازی به عنوان صفحه جداکننده می باشد که کلاسهای مثبت $\{+1\}$ و منفی $\{-1\}$ را با بیشترین حاشیه ممکن جدا می سازد [۷۰]. یک مسئله طبقه بندی باینری که در آن استخراج ویژگی قبلاً صورت گرفته را در نظر می گیریم. فرض می-کنیم، $x_i \in R^d$ بردارهای ورودی و $Y_i = \{-1, +1\}$ برچسب های خروجی باشند که در اینجا $i = 1, \dots, N$ و N تعداد داده های آموزشی و d ابعاد فضای ورودی می باشد. وقتی دو کلاس در فضای R^d به صورت خطی جداپذیر باشند. هدف ما یافتن یک صفحه جداساز است به نحوی که کمترین خطای تعمیم را در بین مجموعه نامحدود صفحات جداساز ایجاد کند. این صفحه ای است که حاشیه جداسازی بین دو کلاس را حداکثر می کند. این حاشیه عبارت است از مجموع فواصل از یک صفحه تا نزدیک ترین نقاط از فضای ورودی به این صفحه برای هر یک از دو کلاس. در شکل (۳-۹) خط ساده نمایانگر این صفحه بهینه برای مسئله ای با ابعاد ورودی ۲ می باشد که به صورت خطی جداپذیرند.



شکل ۳-۹: حاشیه تصمیم برای یک کلاس بندی کننده جدایی پذیر خطی

حال فرض کنید داده های d بعدی به صورت کامل با صفحه ای که با معادله زیر توصیف می شود، جدا شده باشد. تابع تصمیم گیری کلاس بندی کننده به صورت $f(x, \lambda) = \text{sgn}(w \cdot x + b)$ می باشد که در اینجا $x, w \in R^d$ است.

اگر دو کلاس به صورت خطی جداپذیر باشند، یک زوج از $\{b, w\}$ را می توان پیدا کرد که در معادله (۳-۲۶) صدق کند. باید توجه داشت که یک ابهام در تعیین دامنه w و b وجود دارد.

$$y_i (x_i \cdot w + b) - 1 > 0 \quad \forall i \quad (۳-۳۰)$$

۳-۹-۴- ماشینهای بردار پشتیبان خطی در حالت دو کلاسی با الگوی

جدایی ناپذیر

در مسائل واقعی و عملی دو کلاس معمولاً به صورت خطی جدا پذیر نیستند ولی با این حال می توان صفحه ای را پیدا کرد که همراه با حداقل کردن یک کمیت متناسب با خطای طبقه بندی، حاشیه خطای تعمیم را حداکثر کند. برای این منظور در شرط محدود کننده (۳-۲۶) یک متغیر λ را وارد می کنیم. در این حالت معادله (۳-۳۰) به صورت زیر در می آید:

$$y_i (x_i w + b) > 1 - \zeta_i \quad \forall i \quad (31-3)$$

در اینجا وقتی یک خطای طبقه بندی روی می دهد، ζ_i برابر ۱ خواهد شد. بنابراین $\sum \zeta_i$ حد بالای تعداد خطاهای طبقه بندی می باشد. بنابراین تابع هدف معادله (۳-۲۹) به معادله زیر تبدیل می شود.

$$\tau(w, \zeta) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \zeta_i \quad (32-3)$$

که در اینجا C پارامتری است که توسط کاربر انتخاب می شود و اهمیت یکی از دو هدف حداکثر کردن حاشیه خطا و یا خطای طبقه بندی را نسبت به دیگری تعیین می کند. با انتخاب مقادیر بزرگ C حل مسئله بیشتر به سمت کمتر کردن خطای طبقه بندی منجر خواهد شد و با انتخاب مقادیر کمتر C حل مسئله بیشتر به سمت حداکثر کردن حاشیه خطا، اهمیت خواهد یافت. حداقل کردن معادله (۳-۲۸) تحت شرایط معادله (۳-۲۷) منجر به پیدا کردن صفحه جدا ساز تعمیم یافته می شود.

۳-۹-۵- ماشینهای بردار پشتیبان غیر خطی و مفهوم هسته^۱

از آنجایی که برای مسائل واقعی یک طبقه بندی کننده خطی عموماً جوابگو نیست. ضروری به نظر می رسد که بتوانیم سطوح تصمیم گیری خطی را به صورت غیر خطی تعمیم دهیم. از آنجایی که تابع تمایز، تابعی خطی از ورودی های ما نیست، داده ها توسط یک تبدیل غیر خطی از فضای ورودی به یک فضای ورودی با ابعاد بالا تصویر می شوند. در این فضای ویژگیها همانطور که در شکل (۳-۹) مشاهده می شود سطوح جدا سازی تعمیم یافته تشکیل خواهد شد. بر اساس تئوری پوشش اگر تبدیل ما غیر خطی باشد و ابعاد فضای تبدیل به اندازه کافی بزرگ باشد، فضای ورودی ها توسط این تبدیل به صورتی به فضای جدید

^۱ Kernel

تبدیل می گردد که در آنجا به صورت خطی جدایی پذیر باشد. این تبدیل غیرخطی در ماشینهای بردار پشتیبان به صورت ضمنی توسط توابع هسته پیاده سازی می گردد.

جدول ۳-۲ انواع توابع هسته رایج

تابع هسته	نوع طبقه بندی کننده
$k(x, y) = \exp\left(-\frac{\ x - y\ ^2}{2\sigma^2}\right)$	RBF گوسی
$K(x, y) = (x \cdot y + 1)^d$	چند جمله ای از درجه d
$K(x, y) = \tanh(x \cdot y - \Theta)$	شبکه عصبی چند لایه

بنابراین روابط مربوطه به صورت زیر بازنویسی می شوند.

$$f(x) = \text{sign}(\langle x, \phi(x) \rangle + b) \quad (3-33)$$

$$G(W, \zeta) = 1/2 \|w\|^2 + C \sum_{i=1}^N \zeta_i \quad (3-34)$$

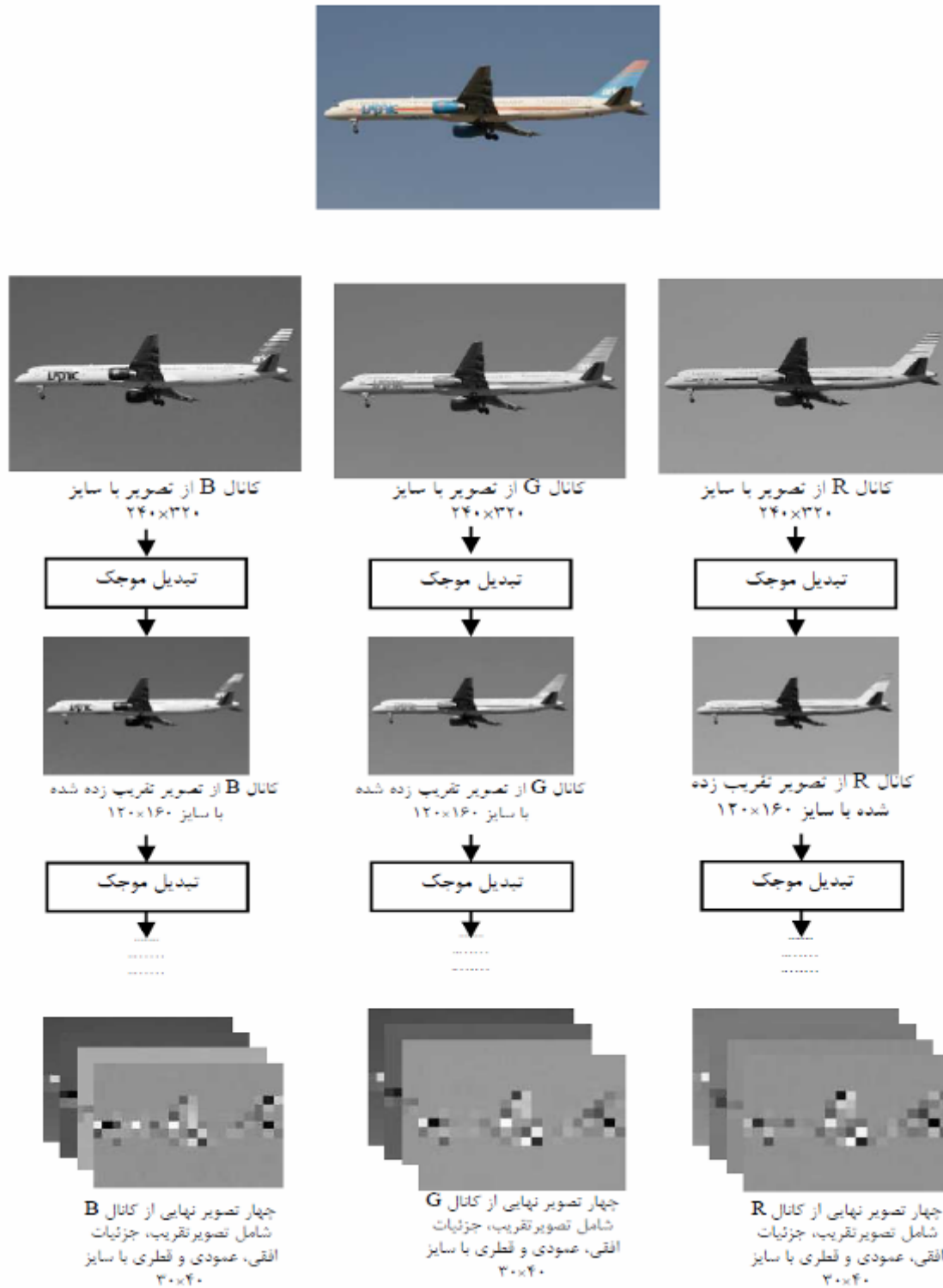
$$y_i (\langle W, \phi(x_i) \rangle + b) \geq 1 - \zeta_i \quad (3-35)$$

۳-۱۰- ویژگی های استخراج شده برای آموزش طبقه بند ماشین بردار پشتیبان

به عنوان اولین گام برای استخراج مشخصه ها، هر تصویر به ابعاد 240×320 پیکسل تبدیل می شود، سپس هر تصویر به هر سه کانال RGB آن تقسیم می شود. در اولین سطح، تبدیل موجک دو بعدی شرح داده شده در فصل دوم بر روی هر یک از کانال های RGB به کار برده می شود تا تصویری در ابعاد 120×160 حاصل شود. هر تصویر شامل تقریب، جزئیات افقی، جزئیات عمودی و

جزئیات قطری می باشد. یک روند مشابه روی تصاویر تقریب^۱ 120×160 اعمال می شود تا تصاویر بعدی در ابعاد 60×80 بدست آید این روند به صورت سطح به سطح ادامه می یابد تا تصاویر با سایزهای 30×40 و حاصل شود. شکل ۳-۱۰ چگونگی بدست آمدن چهار تصویر نهایی با سایز 30×40 که از هر یک از کانالهای RGB تصویر اصلی بدست می آید را نشان می دهد. با انجام روند فوق برای یک تصویر، ۱۲ زیر تصویر با سایز 30×40 بدست می آید. به عبارت دیگر هر پیکسل در تصویر جدید با ابعاد 30×40 ، با یک بردار ۱۲ تایی که همان ضرایب تبدیل موجک هستند ارائه می شود.

^۱ Approximation image



شکل ۳-۱۰: روش استفاده شده برای بدست آوردن ضرایب موجک

۳-۱۱- آموزش و تست SVM برای آشکارسازی هدف در فریم جدید

بعد از به دست آوردن ۱۲ زیر تصویر با اندازه 40×30 از تصویر اصلی یک بردار ویژگی شامل ۱۲ عضو برای هر پیکسل در این تصویر 40×30 تشکیل می دهیم. حال در این تصویر تغییر اندازه داده شده پیکسلهای داخل منحنی قرمز رنگ شکل (۳-۸) را که متعلق به هدف می باشند با مقدار ابرچسب گذاری نموده و پیکسلهای بین بیضی زرد رنگ و این منحنی قرمز رنگ را با کلاس ۱- برچسب می زنیم. بردار ویژگی تشکیل داده شده از تصویر مرجع که ابعاد آن برابر است با تعداد پیکسل هایی (که در داخل ناحیه زرد رنگ قرار گرفتند و هر کدام دارای ۱۲ مؤلفه هستند) به ورودی SVM داده و آن را با این بردار ویژگی آموزش می دهیم. با آمدن یک تصویر جدید در فریم بعدی بعد از آنکه بیضی هدف با روش ذکر شده در فصل قبل تخمین زده شد، بردار ویژگی بدست آمده از پیکسل های داخل ناحیه بیضی را به ورودی SVM داده و آن را تست می کنیم. در این حالت موفق به آشکار سازی مرز دقیق هدف خواهیم شد. پیاده سازی و نتایج عملی روش جابجایی میانگین با آزمایش بر روی تصاویر متعدد در انتهای همین فصل آورده شده است.

۳-۱۲- الگوریتم تخمین مدل حرکت با استفاده از ویژگی های چند گانه

در حالتی که از سه ویژگی رنگ، بافت و لبه استفاده می شود ضریب باتاچاریا ذکر شده در رابطه (۳-۷) برای هر یک از سه ویژگی به طور مستقل به صورت زیر محاسبه می شود:

$$\hat{\rho}_c(y) \equiv \rho \left[\hat{p}(y), \hat{q} \right] = \sum_{u_c=1}^{m_c} \sqrt{\hat{p}_{u_c}(y) \hat{q}_{u_c}} \quad (3-36 \text{ الف})$$

$$\hat{\rho}_e(y) \equiv \rho \left[\hat{p}(y), \hat{q} \right] = \sum_{u_e=1}^{m_e} \sqrt{\hat{p}_{u_e}(y) \hat{q}_{u_e}} \quad (3-36 \text{ ب})$$

$$\hat{\rho}_t(y) \equiv \rho \left[\hat{p}(y), \hat{q} \right] = \sum_{u_t=1}^{m_t} \sqrt{\hat{p}_{u_t}(y) \hat{q}_{u_t}} \quad (3-36 \text{ ج})$$

سپس با تکرار روابط (۳-۸) الی (۳-۱۳) به صورت مستقل برای هر کدام از شاخصه ها \hat{y}_1 را که مکان هدف در فریم جدید است برای هر شاخصه بدست می آوریم. به کمک کوچک ترین معیار فاصله $\hat{d}(y)$ برای هر شاخصه f وزنه‌های اختصاص داده شده به هر یک از این شاخصه در الگوریتم چندگانه برابر است با:

$$f = 1, \dots, 3 \quad \hat{\varepsilon}_f = \frac{1}{\hat{d}_{f_{\min}}} \quad (۳۷-۳)$$

وزنه‌ها سپس به صورت $\sum_{f=1}^3 \hat{\varepsilon}_f = 1$ نرمالیزه می شوند به طوری که:

$$\varepsilon_f = \frac{\hat{\varepsilon}_f}{\sum_{f=1}^3 \hat{\varepsilon}_f} \quad (۳۸-۳)$$

با این روش جایگاه جدید هدف در هر تکرار^۱ برای هر یک از شاخصه ها به طور مستقل با استفاده از رابطه (۳-۱۳) بدست آمده و جایگاه نهایی هدف در آن تکرار به صورت زیر داده می شود:

$$\hat{y}_{1_{\text{multiple}}} = \varepsilon_c \hat{y}_{1c} + \varepsilon_e \hat{y}_{1e} + \varepsilon_t \hat{y}_{1t} \quad (۳۹-۳)$$

پیاده سازی و نتایج عملی الگوریتم جابجایی میانگین با استفاده از ویژگی های چند گانه رنگ، بافت و لبه و مقایسه آن با الگوریتم های مبتنی بر هر یک از این ویژگی ها (به صورت تنها) در انتهای این فصل آورده شده است.

۳-۱۳- پیاده سازی الگوریتم جابجایی میانگین در ردیابی اشیاء

آستانه ε برای خاتمه الگوریتم به وسیله مقایسه بردارهای \hat{y}_0 و \hat{y}_1 که در حدود همان پیکسلهای مختصات تصویر اصلی هستند بدست می آید. انتخاب یک

^۱ Iteration

آستانه کوچکتر دقت الگوریتم را افزایش می دهد. در آزمایش ها مقدار آستانه را برابر ۰/۱ در نظر می گیریم.

در یک پیاده سازی زمان واقعی (real-time) تعداد تکرارهای روش جابجایی میانگین را که همان N_{max} می باشد را به ۲۰ محدود می کنیم. در عمل میانگین تکرارها به مقدار قابل توجهی کمتر از این مقدار بیشینه می باشد. (در حدود (۴) است). پیاده سازی الگوریتم ردیابی بسیار ساده تر از آنچه در انتهای بخش ۳-۴-۱ معرفی شد، می باشد. مرحله پنجم از الگوریتم تنها برای اجتناب از مشکلات عددی که ممکن است در حین اجرای برنامه برای بیشینه سازی جابجایی میانگین پیش آید، در نظر گرفته شده است. این مشکلات بر اثر تقریب های خطی ضریب باتاچاریا به وجود می آیند. با این حال، مجموعه بزرگی از نتایج عملی الگوریتم ردیابی بر روی اشیاء متفاوت در یک دوره زمانی طولانی نشان می دهد که فقط در حدود ۰/۱ درصد از موارد ضریب باتاچار افزایش نمی یابد. بنابراین مرحله پنجم در عمل مورد استفاده نیست. در نتیجه نیازی به ارزیابی مجدد ضریب باتاچاریا در مراحل دوم و چهارم نمی باشد.

در الگوریتم کاربردی، فقط محاسبه وزنها را در مرحله دوم تکرار می نماییم، جایگاه جدید هدف در مرحله ۳ را بدست آورده و این جایگاه جدید را در مرحله ششم تست می نماییم. ضریب باتاچاریا تنها بعد از کامل کردن الگوریتم برای ارزیابی تابع شباهت میان مدل هدف و کاندیدای انتخاب شده، محاسبه شده است. در ضمن از کرنل اپاچینکف برای محاسبه هیستوگرام وزن دار استفاده می شود که پیشتر در همین فصل توضیح داده شده است.

بیشینه کردن ضریب باتاچاریا به عنوان یک روش تطبیق فیلتری بیان می شود. در حقیقت رابطه (۳-۷) ضریب همبستگی میان بردارهای واحد $\sqrt{p^{\wedge}}(y)$ و $\sqrt{q^{\wedge}}$ ، ارائه مدل هدف و کاندیدای هدف، را بیان می کند. روش جابجایی میانگین، بیشینه محلی ضرایب همبستگی را پیدا می کند. اگر در فریم جاری، مرکز هدف در منطقه تصویر مدل هدف در فریم قبلی قرار گرفته باشد، بیشینه محلی بدست آمده از ضریب باتاچاریا یک شاخص قابل اعتماد برای

جایگاه جدید هدف است. روش جابجایی میانگین یک ریشه از مشتق را به عنوان تابع جایگاه هدف پیدا می کند، که همچنین می توان آن را به عنوان یک نقطه که به صورت (زین شده) در نمایش رویه می باشد در نظر گرفت (چون در جایی که مشتق صفر می شود بیشینه های محلی ما هستند). نقاط زینی راه حل های غیر پایداری هستند و از آنجایی که نویز موجود در تصویر به عنوان یک فاکتور مزاحم مستقل در میان فریم های متوالی عمل می کند، آنها می توانند عملکرد الگوریتم ردیابی را در یک دنباله از تصاویر تحت تاثیر قرار دهند.

۳-۱۴- نتایج عملی

در شکل (۳-۱۱) یک دنباله تصویر ۳۴۸ فریمی با ابعاد 240×320 از یک عابر پیاده (که خیابان در حال حرکت می باشد) دیده می شود. این دنباله تصویر ویدئویی با استفاده از دوربین متحرک ۵ مگاپیکسل گرفته شده است. در این آزمایش فضای RHX کوانتیزه شده به مؤلفه های هیستوگرام $16 \times 16 \times 16$ به عنوان فضای ویژگی انتخاب شده است. الگوریتم جابجایی میانگین مطابق بخش ۳-۱۳ پیاده شده است. کرنل اپاچینکف برای محاسبه هیستوگرام وزن دار استفاده شده و بیشینه تعداد تکرارها در هر فریم به عدد ۲۰ محدود شده است. مرکز بیضی اولیه در فریم اول را نقطه [۱۳۲، ۱۴۳] در نظر می گیریم. قطر کوچک و بزرگ بیضی را نیز به ترتیب برابر ۱۰ و ۳۰ پیکسل در نظر می گیریم.



فریم ۵



فریم ۶۰



فریم ۱۲۰



فریم ۲۴۰



فریم ۳۰۰

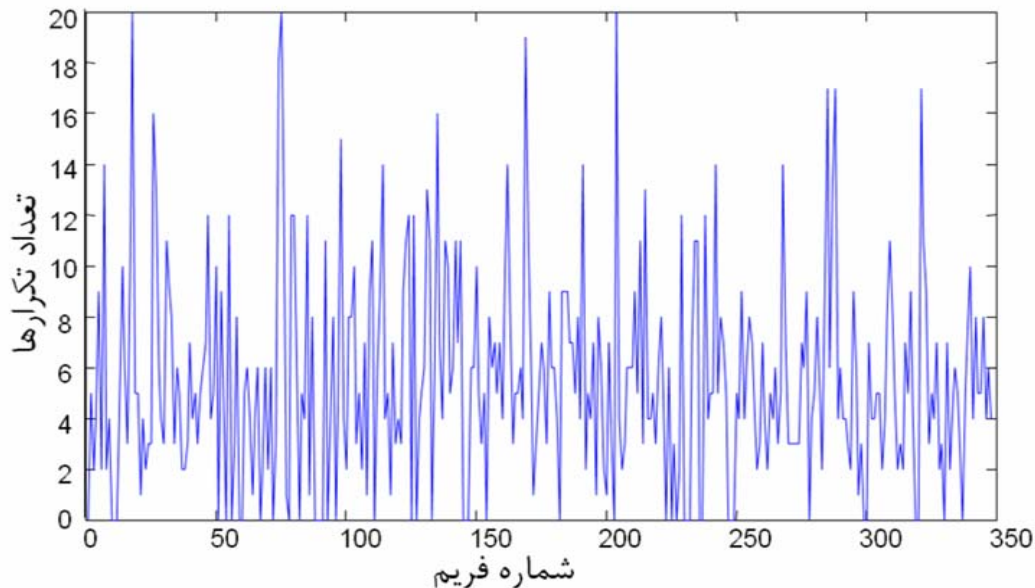


فریم ۳۴۰

شکل ۳-۱۱: دنباله تصویر عابر پیاده، ردیابی شخص در فریم های ۶۰، ۱۲۰، ۲۴۰، ۳۰۰، ۳۴۰

همانطور که در تصاویر دیده می شود، ردیاب^۱ جابجایی میانگین در اکثر فریمها به جز فریم ۲۴۰ و ۳۰۰ موفق به ردیابی و تخمین مؤثر هدف شده است و در اثر تغییر اندازه جسم (فریم ۱۲۰، ۲۴۰) هدف گم نشده است در نتیجه ردیاب جابجایی میانگین در مقابل تغییر اندازه کامل هدف (فریم ۳۴۰) مقاوم است. الگوریتم جابجایی میانگین با استفاده از نرم افزار MATLAB و بر روی یک کامپیوتر با هسته 2.16 GHz Core 2 Due پیاده شده است. زمان اجرای برنامه بر روی این دنباله ویدئویی ۳۴/۰۷ ثانیه ای برابر ۶۷/۴۲ ثانیه است. تعداد تکرار های الگوریتم جابجایی میانگین در هر فریم در شکل (۳-۱۲) نشان داده شده است.

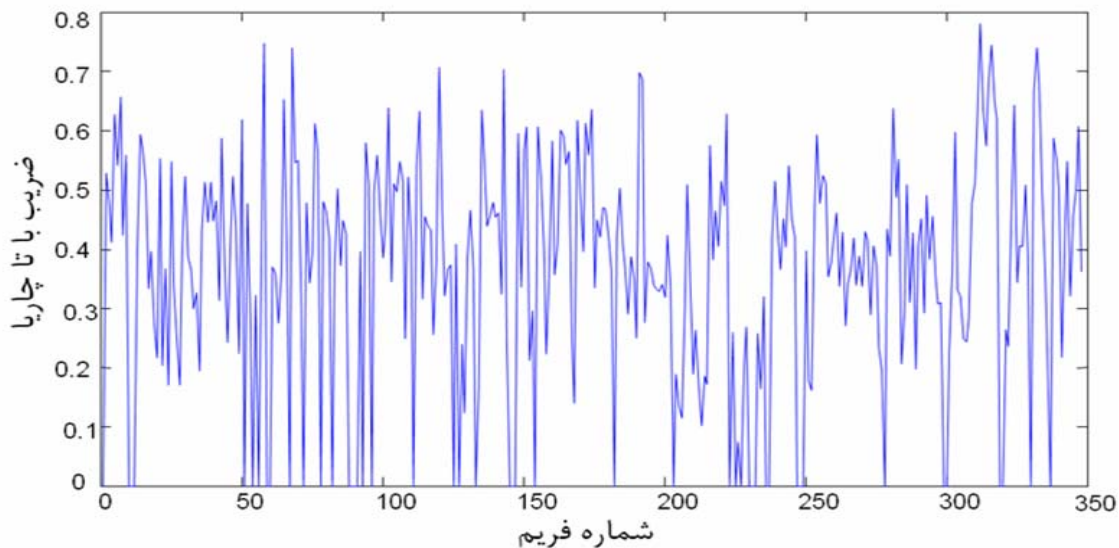
^۱ Tracker



شکل ۳-۱۲: تعداد تکرارهای الگوریتم جابجایی میانگین در هر فریم از دنباله تصویر عابر پیاده

همانطور که از شکل بر می آید اگرچه بیشینه تعداد تکرارها در هر فریم را به عدد ۲۰ محدود کرده ایم ولی میانگین تعداد تکرارها در ۳۴۸ فریم برابر $6/12$ است. ما مقدار ضریب باتاچاریا بین مدل هدف در فریم اول و کاندیدای نهایی هدف در هر فریم را که از رابطه (۳-۷) بدست می آید محاسبه کرده و در شکل (۳-۱۳) نمایش داده ایم.

فریم هایی که ضریب باتاچاریا به صفر نزدیک می شود، نشان دهنده آن است الگوریتم موفق به تخمین دقیق مکان جسم نشده است. همانطور که در شکل (۳-۱۳) دیده می شود در فریمهای ۲۱۰ تا ۲۴۰ مقدار ضریب باتاچاریا زیر $0/3$ است که حاکی از ردیابی نا کارآمد هدف می باشد.



شکل ۳-۱۳: مقدار ضریب باتاچاریا بین مدل هدف در فریم اول و کاندیدای نهایی هدف در هر فریم از دنباله تصویر عابر پیاده

* تغییر اندازه وقتی بیضی هدف

در مواردی که اندازه هدف در فریم های متوالی تغییر می کند، نیاز است که مقدار پهنای باند بیضی هدف یا به عبارت دیگر پارامتر h ، که نشان دهنده اندازه بیضی هدف است، تغییر کند. پهنای باند بهینه (h_{opt}) در فریم جاری با محاسبه سه باره الگوریتم تعیین جایگاه هدف می توانیم با استفاده از رابطه زیر بدست آوریم (این رابطه تقریبی است):

$$h_{new} = (.1h_s + .9)h_{prev} \quad (۴۰-۳)$$

که در آن h_s مقادیر $1/9$ ، 1 ، $1/1$ را می گیرد. برای سه مقدار h_{new} بدست آمده فاصله بین مدل هدف و کاندیدای آن را در فریم فعلی محاسبه کرده، h_{new} که کمترین میزان فاصله را بدهد به عنوان پهنای باند بهینه (h_{opt}) در نظر می گیریم.

^۱ Adaptive

در شکل (۳-۱۴) یک دنباله تصویر ۹۴ فریمی با ابعاد 428×344 از یک ماشین خاکستری رنگ (که در خیابان در حال حرکت می باشد) دیده می شود. این دنباله تصویر ویدئویی با استفاده از دوربین متحرک ۵ مگاپیکسل گرفته شده است. در این آزمایش فضای RGB کوانتیزه شده به مؤلفه های هیستوگرام $8 \times 8 \times 8$ به عنوان فضای ویژگی انتخاب شده است. همانطور که مطابق بخش (۳-۱۳) پیاده شده است، الگوریتم مبتنی بر کرنل اپاچینکف برای محاسبه هیستوگرام وزن دار استفاده شده است و بیشینه تعداد تکرارها در هر فریم به عدد ۲۰ محدود شده است. مرکز بیضی اولیه در فریم اول نقطه $[164, 134]$ در نظر می گیریم. قطر کوچک و بزرگ بیضی را به ترتیب برابر ۲۵ و ۵۲ پیکسل در نظر می گیریم.



فریم ۱



فریم ۳۰



فریم ۶۰

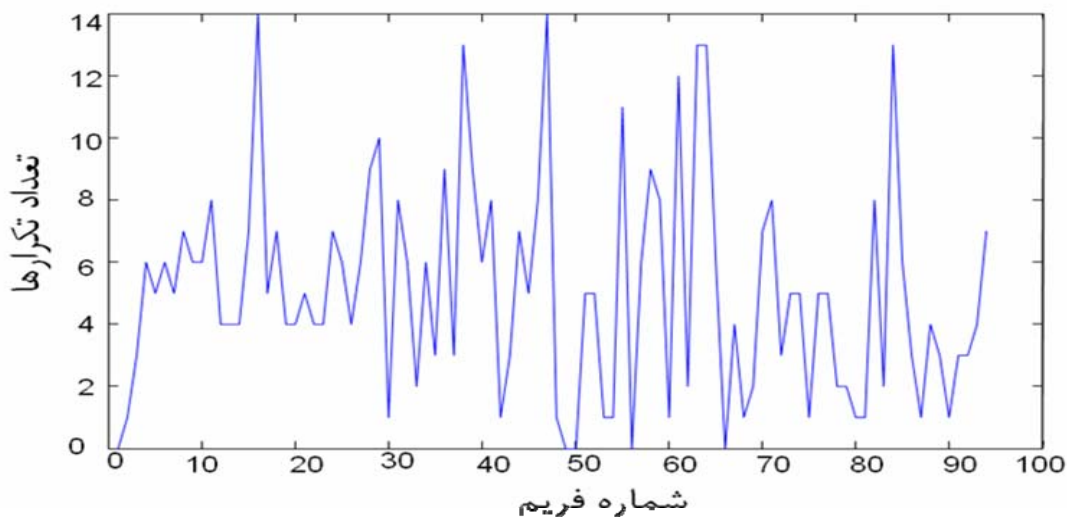


فریم ۹۰

شکل ۳-۱۴: دنباله تصویر ماشین، ردیابی ماشین در فریم های ۳۰، ۶۰، ۹۰

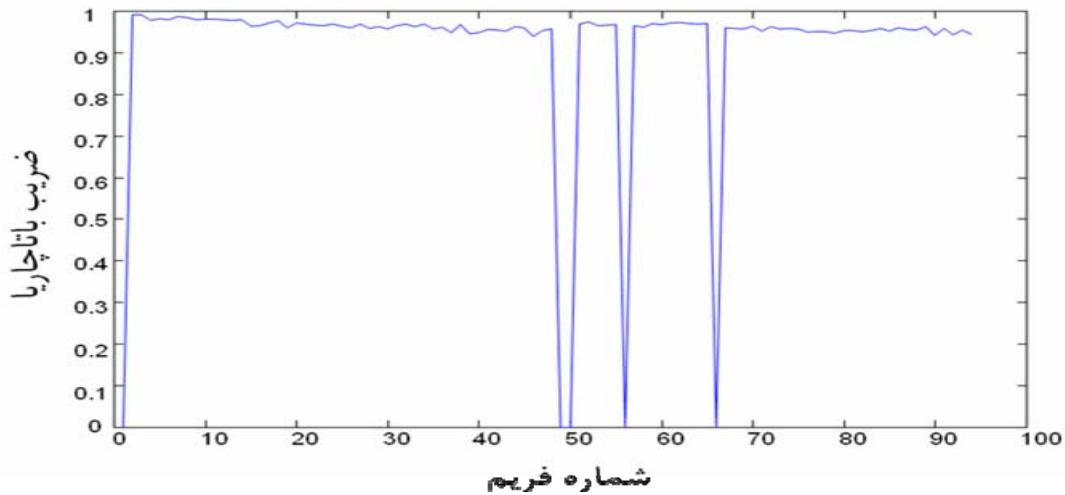
همانطور که در تصاویر دیده می شود، ردیاب جابجایی میانگین در مقابل بزرگ نمایی جسم (فریم ۳۰، ۶۰) و تغییر نمود کامل هدف (فریم ۶۰، ۹۰)

مقاوم است. در این مثال پارامتر به صورت افقی در هر فریم تغییر می کند و ابعاد بیضی هدف را نیز تغییر می دهد به طوری که در فریم ۳۰ و ۶۰ اندازه بیضی هدف بزرگ شده است. الگوریتم جابجایی میانگین با استفاده از نرم افزار MATLAB و بر روی یک کامپیوتر با هسته 2 GHz Core 2 Due پیاده شده است. زمان اجرای برنامه بر روی این دنباله ویدئویی ۹ ثانیه ای برابر ۹۱/۸ ثانیه است. تعداد تکرارهای الگوریتم جابجایی میانگین در هر فریم در شکل (۳-۱۵) نشان داده شده است.



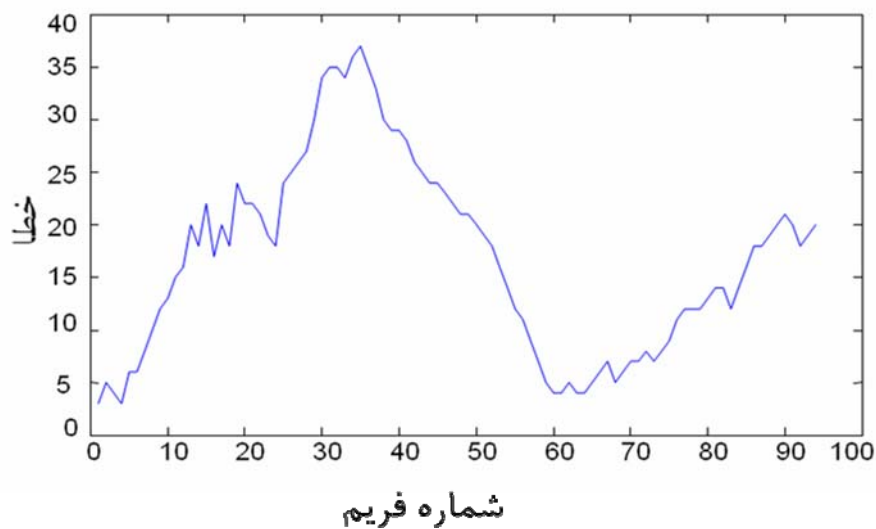
شکل ۳-۱۵: تعداد تکرارهای الگوریتم جابجایی میانگین در هر فریم از دنباله تصویر ماشین

ما مقدار ضریب باتاچاریا بین مدل هدف در فریم اول و کاندیدای نهایی هدف در هر فریم را که از رابطه (۳-۷) بدست می آید محاسبه کرده و در شکل (۳-۱۶) نمایش داده ایم. همانطور که دیده می شود در اکثر موارد ضریب باتاچاریا بیشتر از ۰/۹ است که نشان دهنده ردیابی مؤثر هدف می باشد.



شکل ۳-۱۶: مقدار ضریب باتاچاریا بین مدل هدف در فریم اول و کاندیدای نهایی هدف در هر فریم از دنباله تصویر ماشین

فاصله مرکز بیضی بدست آمده در هر فریم را با مرکز ثقل هدف در همان فریم بر حسب جزء صحیح پیکسلها محاسبه کرده و آن را به عنوان خطای الگوریتم تخمین مکان هدف در شکل (۳-۱۷) نمایش داده ایم. همانطور که دیده می شود مقدار خطا از ۳۸ پیکسل فراتر نمی رود.



شکل ۳-۱۷: خطای تخمین مکان هدف در هر فریم از دنباله تصویر ماشین

۳-۱۴-۱- تخمین مکان هدف با استفاده از ویژگی های متعدد

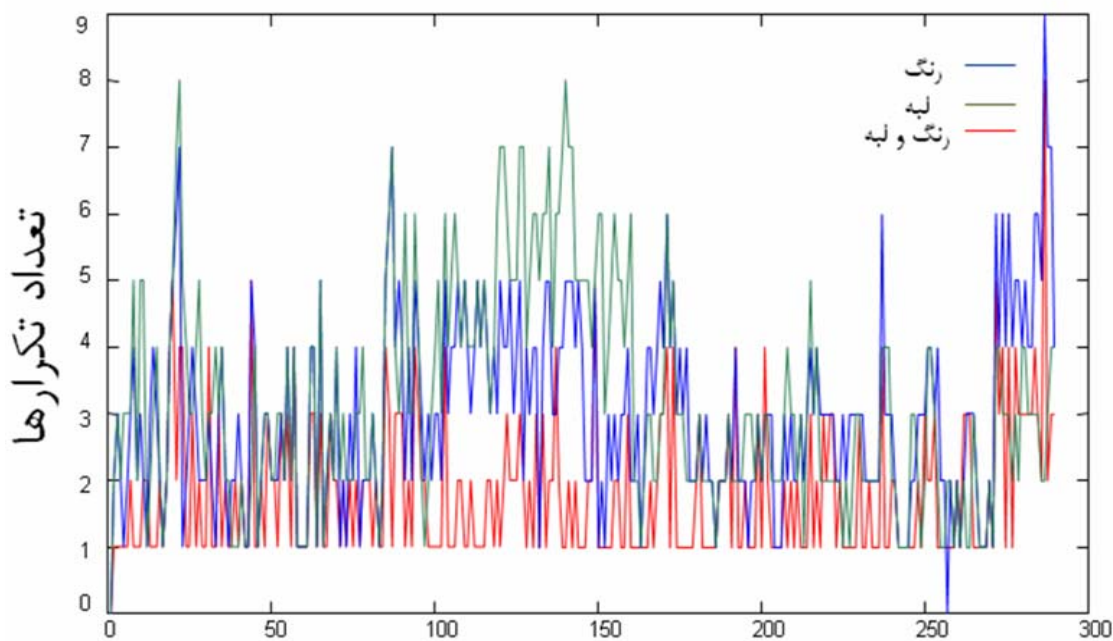
در این قسمت نتایج عملی الگوریتم جابجایی میانگین با استفاده از ویژگی های چند گانه رنگ، بافت و لبه را نشان می‌دهیم و آن را با این الگوریتم مبتنی بر فقط یکی از این ویژگی ها (به صورت تنها) مقایسه می‌کنیم. اولین آزمایش تخمین مکان دست را در یک دنباله تصویر ۳۹۴ فریمی که ابعاد هر فریم در آن برابر 144×176 پیکسل است شامل می‌شود. در این آزمایش الگوریتم جابجایی میانگین یک بار برای ویژگی های لبه و رنگ به صورت مستقل و بار دیگر به صورت توأم به کار گرفته شده است.



شکل ۳-۱۸: ردیابی دست با استفاده از ویژگی های لبه و رنگ به صورت مستقل و به صورت توأم

همانطور که در شکل (۳-۱۸) دیده می‌شود الگوریتم جابجایی میانگین تنها مبتنی بر ویژگی لبه در فریم ۱۴۰ از تخمین دقیق مکان دست ناتوان است در حالیکه این الگوریتم با استفاده از ویژگی رنگ در آن فریم موفق عمل می‌کند. از سوی دیگر الگوریتم جابجایی میانگین تنها مبتنی بر ویژگی رنگ در زمانی که

دست از جلوی صورت می گذرد (فریم ۲۴۰) به علت رنگ مشابه صورت و دست ناموفق عمل کرده، در حالی که این الگوریتم با ترکیب ویژگی های رنگ و لبه موفق به تخمین دقیق مکان جسم در تمام فریم ها شده است. زمان اجرای این الگوریتم این دنباله تصویر ۳۸ ثانیه ای برابر $27/8$ ثانیه می باشد. در شکل (۳-۱۹) تعداد تکرارها در هر فریم برای ویژگی رنگ و لبه به صورت مجزا و توأم نشان داده شده است.



شماره فریم

شکل ۳-۱۹: تعداد تکرارهای الگوریتم جابجایی میانگین در هر فریم از دنباله تصویر دست برای ویژگی رنگ و لبه به صورت مجزا و توأم

از شکل (۳-۱۹) به روشنی در می یابیم که با بهره گیری از ویژگی های رنگ و لبه به صورت توأم تعداد تکرارها در هر فریم کاهش می یابد.

در آزمایش بعدی فریم های متوالی از یک چیتای در حال حرکت در یک بیشه را مشاهده می کنید. همانطور که در شکل (۳-۲۰) دیده می شود رنگ هدف مشابه پس زمینه می باشد به همین دلیل تخمین مکان هدف تنها با

استفاده از ویژگی رنگ در فریم ۲۰۰ با شکست مواجه می شود، در حالیکه ردیابی با استفاده از ویژگی رنگ و بافت توام در تمامی فریم ها موفق عمل می کند.



شکل ۳-۲۰: ردیابی چیتا با استفاده از ویژگی های رنگ و بافت به صورت مستقل و به صورت توام

۳-۱۴-۲- جداسازی دقیق هدف از پس زمینه

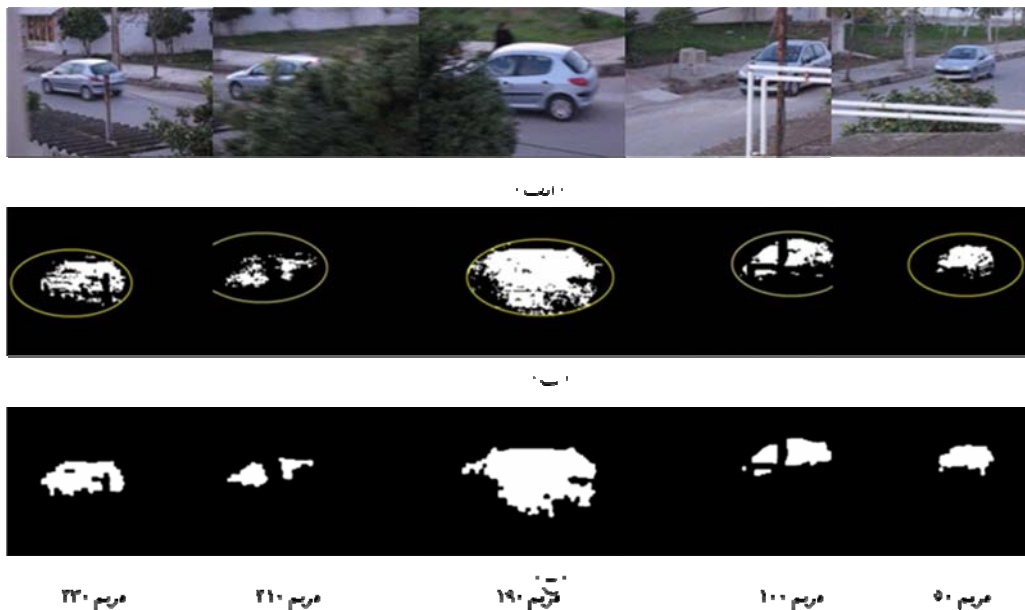
در این قسمت، نتایج عملی بدست آمده از روش کلاسه بندی (SVM) برای تعیین مرزهای دقیق هدف (شیء)، که مکان این شیء با استفاده از روش ذکر شده (جابجایی میانگین) در بخش قبل تخمین زده شد، بیان می شود. شکل (۳-۲۱) یک دنباله تصویر ۲۹۱ فریمی با ابعاد ۲۴۰×۳۲۰ از یک ماشین پژو ۲۰۶ می باشد. این دنباله تصویر ویدئویی با استفاده از دوربین متحرک ۵ مگاپیکسل گرفته

شده است. در این آزمایش فضای RGB کوانتیزه شده به مؤلفه های هیستوگرام $8 \times 8 \times 8$ به عنوان فضای ویژگی انتخاب شده است. در مرحله نخست الگوریتم جابجایی میانگین را بر روی بیضی سیاه رنگ که در شکل (۳-۲۱) نمایش داده شده است به کار گرفته و بیضی هدف در هر فریم را با استفاده از الگوریتم تخمین مکان هدف (جابجایی میانگین) استخراج می کنیم.



شکل ۳-۲۱: تصویر مرجع در فریم اول. ناحیه داخل بیضی سیاه رنگ به عنوان هدف در فریم اول در نظر گرفته می شود.

بعد از بدست آوردن بیضی هدف در فریم جدید پیکسلهای ناحیه ای با ابعاد $\sqrt{2}$ برابر ناحیه هدف را (همانطور که در ۳-۱۱ گفته شده است) با SVM تست می کنیم. خروجی SVM در هر فریم پیکسلهای متعلق به جسم را به ما می دهد. شکل (۳-۲۲-ب) خروجی SVM را نشان می دهد. پیکسلهای با رنگ سفید مشخص کننده هدف است.



شکل ۳-۲۲: یک مثال از ردیابی ماشین با استفاده از الگوریتم جابجایی میانگین. الف- ۵ فریم از دنباله تصویر ۲۹۰ فریمی نشان داده شده است. ب- خروجی SVM بعد از اعمال بر روی ناحیه هدف داخل بیضی زرد رنگ. ج- تصویر خروجی پس از اعمال یک سری عملیات شکل شناسی.

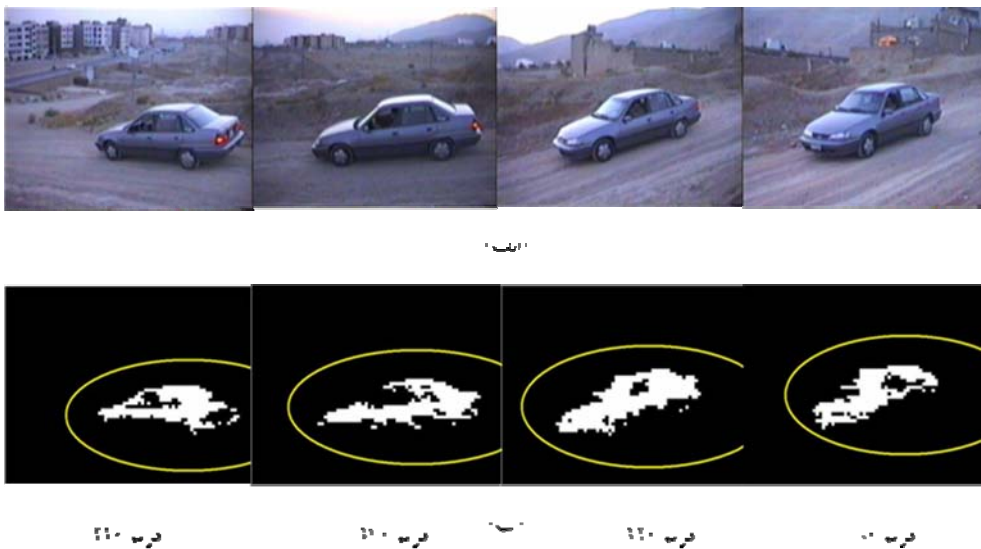
همانطور که در شکل دیده می شود این الگوریتم (جابجایی میانگین) در مقابل انواع مختلفی از پوشیدگی ها مقاوم است. چنانکه در فریم ۱۰۰ دیده می شود، در حالیکه ماشین به صورت جزئی با میله سفید رنگ پوشیده می شود این الگوریتم قادر به آشکار سازی قسمت های قابل مشاهده جسم می باشد. در حالت دیگری از پوشیدگی وقتی ماشین در فریم های ۱۹۰ تا ۲۱۰ کاملاً ناپدید می شود، سیستم ردیاب جابجایی میانگین موفق به از سرگیری ردیابی جسم می گردد. به علاوه در این دنباله تصویر با اینکه در فریم ۲۳۰ تغییر نمود شدیدی در هدف صورت می گیرد به علت ماهیت این روش که اطلاعات مکانی در آن در نظر گرفته نمی شود، باز هم موفق به ردیابی قابل قبول جسم هستیم. در انتها نیز یک سری عملیات شکل شناسی^۱ برای حذف نویز بر روی تصویر خروجی انجام می گیرد که نتایج آن در شکل (۳-۲۲-ج) دیده می شود. در آزمایش فوق از

^۱ Morphological approaches

SVM با هسته گوسی (RBF kernel) استفاده شده و پارامترهای آن برابر مقادیر زیر انتخاب شده است:

$$\text{kernel parameter} = 5 \quad c = 50$$

در شکل (۳-۲۳) یک دنباله تصویر ۲۲۲ فریمی با ابعاد 240×320 از ماشین در حال حرکت دیده می شود. این دنباله تصویر ویدئویی با استفاده از دوربین متحرک ۵ مگاپیکسل گرفته شده است. در این آزمایش فضای RGB کوانتیزه شده به مؤلفه های هیستوگرام $8 \times 8 \times 8$ به عنوان فضای ویژگی انتخاب شده است. بعد از بدست آوردن بیضی هدف در فریم جدید پیکسلهای ناحیه ای با ابعاد $\sqrt{2}$ برابر ناحیه هدف را همانطور که در بخش ۳-۱۱ گفته شده است با SVM تست می کنیم. خروجی SVM در هر فریم پیکسلهای متعلق به جسم را به ما می دهد. شکل (۳-۲۳-ب) خروجی SVM را نشان می دهد. پیکسلهای با رنگ سفید مشخص کننده هدف است.



شکل ۳-۲۳: یک مثال از ردیابی ماشین با استفاده از الگوریتم جابجایی میانگین. الف- ۴ فریم از دنباله تصویر ۲۲۲ فریمی نشان داده شده است. ب- خروجی SVM بعد از اعمال بر روی ناحیه هدف داخل بیضی زرد رنگ.

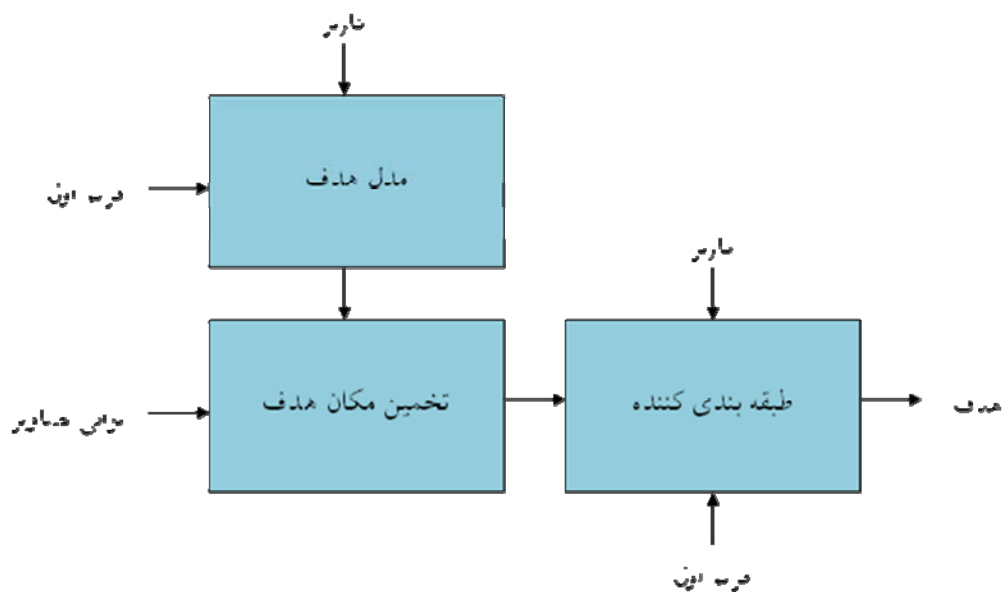
همانطور که در شکل دیده می شود الگوریتم جابجایی میانگین در مقابل تغییر نمود شدید در فریم ۲۱۰ مقاوم است و به طور مؤثری جسم را آشکار می

سازد. در آزمایش فوق از SVM با هسته گوسی (RBF kernel) استفاده شده و پارامترهای آن برابر مقادیر زیر انتخاب شده است:

$$\text{kernel parameter} = .01 \quad c = 50$$

۳-۱۵- سیستم کلی روش جابجایی میانگین در ردیابی اشیاء

سیستم کلی روش جابجایی میانگین ارائه شده به صورتی است که با ورود فریم اول و به کمک کاربر مدل هدف تشکیل داده می شود. با ورود فریم های جدید مکان هدف با استفاده از روش ذکر شده در بالا تخمین زده می شود. پس از تخمین مکان جسم با روش جابجایی میانگین در بالا، به کمک طبقه بند ماشین بردار پشتیبان، هدف (شی) را در هر فریم استخراج می کنیم. بلوک دیاگرام عملکرد کلی سیستم در شکل (۳-۲۴) نشان داده شده است.



شکل ۳-۲۴: بلوک دیاگرام عملکرد کلی سیستم

۳-۱۶- مروری کلی بر روش جابجایی میانگین و کاربرد آن

الگوریتم جابجایی میانگین از روشهای سریع و موفق بکار رفته در ردیابی تصویری است که اخیراً بخاطر سادگی و موثر بودنش بسیار محبوب شده است [۶۱، ۶۰، ۳۲]. در ردیاب های جابجایی میانگین، هیستوگرام رنگ برای توصیف محیط هدف استفاده شده است [۳۲]. ضریب باتاچاریا یا معیارهای دیگر اندازه گیری شباهت بین محیط الگو یا همان مدل و محیط هدف بکار برده می شوند. ردیابی توسط پیدا کردن بیشینه های محلی توابع به انجام می رسد.

به عنوان یک روش بهینه سازی بسیار موثر، جابجایی میانگین دارای همگرایی سریع و هزینه محاسباتی کم می باشد. بعلاوه اینکه یک روش غیر پارامتری است، همچنین یک راه حل بهینه سازی کلی غیر وابسته به ویژگی های هدف را فراهم می کند. در هر حال در جابجایی میانگین، به علت وجود اشیاء مزاحم پیش زمینه، این روش بهینه سازی کلی نمی تواند پیدا کردن بیشینه های بیشینه ها را تضمین کند و ممکن است به راحتی در بیشینه های محلی گیر کند. همانطور که در مرجع [۶۲] نشان داده شده است، ردیابهای جابجایی میانگین به آسانی در ردیابی اشیاء سریع ناکام می مانند و نمی توانند خود را از این ناکامی مصون بدارند و رد شی را به راحتی گم می کنند. در نتیجه دقت آنها با زمان ردیابی شان حالت معکوس دارد. فیلترکالمن (Kalman) برای بهبود همین مقاومت ردیابی در ردیابی جابجایی میانگین استفاده می شود [۶۳]. به هر حال، از آنجاییکه آنها بر روی الگوریتم های مشخصی پایه ریزی شده اند، به همین دلیل ردیابهای جابجایی میانگین نمی توانند به مسئله رویه های چند مد رسیدگی کنند.

۳-۱۷- جمع بندی

در این فصل روش جابجایی میانگین به عنوان یک ردیاب سریع برای تخمین مکان هدف در هر فریم معرفی شده است. در روش جابجایی میانگین می توان مکان هدف در هر فریم را تخمین زد. این روش مبتنی بر ارائه مدل مناسب از

هدف در فریم اول و کاندیدای آن در فریم های بعدی می باشد. مدل با استفاده از هیستوگرام های وزن دار از هدف محاسبه می شود. مزیت استفاده از هیستوگرام این است که به سادگی محاسبه شده و همانطور که در انتهای فصل توضیح داده شد امکان بهره گیری از ویژگی های متعدد از هدف و ترکیب آنها را برای ما فراهم می آورد. وزن دهی هیستوگرام با اعمال یک کرنل خاص بر روی هدف صورت می گیرد. همچنین از تبدیل فاصله نیز می توان برای وزن دهی پیکسلهای هدف استفاده کرد. در ادامه بصورت کلی مروری بر این روش در ردیابی اشیا بویژه دست داشتیم و مزایا و معایب آن را نیز تشریح کردیم. سیستم ارائه شده برای تعیین مرز هدف نیز شرح داده شد. ابتدا در فریم اول پیکسلهای موجود در آن ناحیه را به دو گروه هدف و غیر هدف تقسیم کرده و طبقه بند ماشین بردار پشتیبان را با آن آموزش دادیم. در ادامه پس از تخمین بیضی هدف در فریم های بعدی پیکسلهای حول بیضی هدف را با طبقه بند ذکر شده (SVM) تست کرده و هدف را بصورت دقیقتر استخراج می کنیم. در فصل چهارم از این روش (جابجایی میانگین) برای کاهش زمان ردیابی در ردیاب فیلتر ذره ای استفاده می کنیم که منجر به روش تلفیقی پیشنهادی ما می شود.

فصل چهارم

ردیابی دست با استفاده از روش
پیشنهادی تلفیقی

۴-۱- مقدمه

در این فصل به روش فیلتر ذره ای (ردیاب دقیق) می‌پردازیم و سپس و روش جابجایی میانگین (ردیاب سریع) و فیلتر ذره ای را تلفیق کرده و روش تلفیقی را پیشنهاد می‌کنیم که نتایج حاصل از اعمال این الگوریتم های ردیابی ارائه شده بر روی دنباله تصاویر ویدئویی گوناگون زبان اشاره را در انتها مشاهده می‌کنیم.

۴-۲- الگوریتم تخمین مدل حرکت با استفاده از فیلتر ذره ای

ردیابی موثر اشیاء در محیط های شلوغ در مبحث بینایی ماشین بسیار چالش آمیز است. ردیابی تصویری می تواند به عنوان تخمینی از مکان یک شیء در یک سکانس محسوب شود. هدف ردیابی به روش فیلتر ذره ای این است که بصورت برگشتی چگالی پیشین $p(\mathbf{x}_t | \mathbf{z}_{1:t})$ را برای وضعیت شیء حاضر \mathbf{x}_t که بر طبق مشاهدات $\mathbf{z}_t = (\mathbf{z}_1, \dots, \mathbf{z}_t)$ تا لحظه t عوض شده است محاسبه کند. تابع چگالی احتمال (pdf) $p(\mathbf{x}_t | \mathbf{z}_{1:t})$ می تواند در دو مرحله بصورت بازگشتی بدست آورده شود: پیش بینی و به روز رسانی. اگر موقعیت \mathbf{x}_t که با زمان تغییر می کند به عنوان یک پروسه درجه اول مارکوف مدل شود، pdf $p(\mathbf{x}_t | \mathbf{z}_{1:t})$ بصورت زیر نتیجه گیری می شود:

$$p(\mathbf{x}_t | \mathbf{z}_{1:t}) = \kappa p(\mathbf{z}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{z}_{1:t-1}) \quad (1-4)$$

$$p(\mathbf{x}_t | \mathbf{z}_{1:t-1}) = \int p(\mathbf{x}_t | \mathbf{x}_{t-1}) p(\mathbf{x}_{t-1} | \mathbf{z}_{1:t-1}) dx_{t-1} \quad (2-4)$$

جایی که κ یک ثابت نرمالایز کننده غیر وابسته به \mathbf{x}_t است، $p(\mathbf{z}_t | \mathbf{x}_t)$ تابع شباهت است، $p(\mathbf{x}_t | \mathbf{x}_{t-1})$ مدل دینامیکی است و $p(\mathbf{x}_t | \mathbf{z}_{1:t-1})$ اولویت گذرا بر روی \mathbf{x}_t است که در مشاهده قبلی داده شده است. $p(\mathbf{x}_t | \mathbf{x}_{t-1})$ و $p(\mathbf{z}_t | \mathbf{x}_t)$ در اینجا نیاز نیست که گوسی باشند.

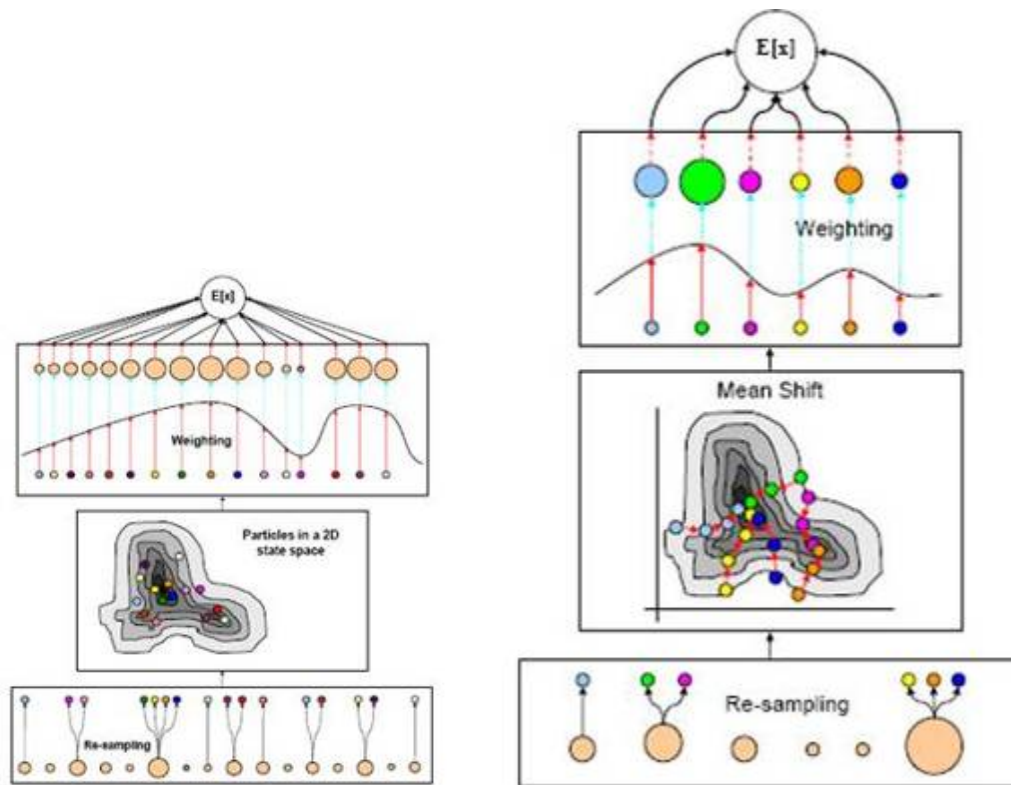
فیلتر ذره ای [۶۴، ۱۸] تکنیکی برای پیاده سازی یک فیلتر Bayesian برگشتی بوسیله شبیه سازیهای Monte Carlo است. در فیلتر ذره ای چگالی پیشین مورد نیاز بوسیله یک مجموعه ذره ای وزن دار شده $\{(s_t^{(n)}, \pi_t^{(n)})\}_{n=1}^N$ در هر زمان t تخمین زده می شود. هر ذره $s_t^{(n)}$ یک موقعیت فرضی شیء را نشان می دهد و بوسیله یک احتمال نمونه برداری گسسته وزن دهی شده است $\pi_t^{(n)} = p(\mathbf{z}_t | \mathbf{x}_t = s_t^{(n)})$ ، که احتمال آنکه مشاهدات کنونی توسط یک احتمال فرضی ایجاد شده باشند. موقعیت در هر لحظه t می تواند در نهایت توسط این ذره ها و وزن ها تخمین زده شود. مجموعه ذره ها توسط مدل دینامیکی سیستم بر روی زمان منتشر می شوند.

ما الگوریتم فیلتر ذره ای را در شکل (۴-۱) خلاصه کرده و در شکل (۴-۲) بصورت شماتیک نشان می دهیم.

با مجموعه ذره های داده شده در زمان $t - 1$ ، مراحل زیر را انجام می دهیم:

۱. نمونه برداری دوباره N ذره از مجموعه $\{(s_{t-1}^{(n)}, \pi_{t-1}^{(n)})\}_{n=1}^N$ برای تولید $\{(s_{t-1}'^{(n)}, \frac{1}{N})\}_{n=1}^N$.
۲. انتشار هر ذره بوسیله مدل دینامیکی $s_t^{(n)} \sim p(\mathbf{x}_t | \mathbf{x}_{t-1} = s_{t-1}'^{(n)})$ برای تولید $\{(s_t^{(n)}, \frac{1}{N})\}_{n=1}^N$.
۳. وزن دهی ذره ها به عنوان $\pi_t^{(n)} \propto p(\mathbf{z}_t | \mathbf{x}_t = s_t^{(n)})$ برای تولید $\{(s_t^{(n)}, \pi_t^{(n)})\}_{n=1}^N$ ، جایگزین $\pi_t^{(n)}$ نرمالایز شده است در نتیجه $\sum_{n=1}^N \pi_t^{(n)} = 1$.
۴. محاسبه نتیجه ردیابی $E[\mathbf{x}_t] = \sum_{n=1}^N \pi_t^{(n)} s_t^{(n)}$.

شکل ۴-۱- الگوریتم فیلتر ذره ای



شکل ۴-۲- نمایش شماتیکی از الگوریتم فیلتر ذره ای

در بحث ردیابی تصویری قابلیت مدیریت رویه چند مدی به فیلتر ذره ای اجازه می دهد که اشیاء مزاحم پیش زمینه را برطرف کرده و از ناکامی هایی که ردیاب جابجایی میانگین در پیدا کردن مد اصلی داشت و در مد های محلی گیر می افتاد بهبود یافته و باز گردد. فیلتر ذره ای به لحاظ مقاومت و دقت بالایش نظرات بسیاری را در حوزه بینایی ماشین و کامپیوتر به سمت خود جلب کرده است [۶۶،۶۵،۲۰] و اخیراً ردیاب های فیلترهای ذره ای بر پایه رنگ نیز معرفی شده اند [۶۵،۶۲].

موفقیت فیلتر ذره ای به توانایی آن برای حفظ یک تخمین خوب از احتمال قبلی موقعیت مکانی شیء بستگی دارد. برای گرفتن تغییرات در فضای مکانی، به تعداد مشخصی از ذره ها برای تضمین نمونه برداری کافی نیاز است و این تعداد بصورت نمایی با بعد مکان افزایش می یابد. هزینه محاسباتی بالا که بخاطر تعداد

بیش از اندازه ذره ها ایجاد می شود باعث می شود که فیلتر ذره ای در حالت زمان واقعی (Real time) را غیر قابل پیاده سازی می کند.

۴-۳- روش پیشنهادی تلفیقی (تلفیق جابجایی میانگین با فیلتر ذره ای)

هر دو روش جابجایی میانگین و فیلتر ذره ای دارای نقاط ضعف و قوت هستند. لذا پیشنهاد می کنیم که هر دو روش را برای بهبود ردیابی از لحاظ سرعت و دقت با هم ترکیب کنیم. با بکار بردن روش جابجایی میانگین در فیلتر ذره ای، روشی را برای بهبود، سادگی و بازدهی بیشتر نمونه برداری فیلتر ذره ای پیشنهاد می کنیم. در مقایسه با فیلتر ذره ای، تعداد ذره های مورد نیاز بشدت کاهش می یابد. روش ارائه شده شبیه به مرجع [۲۳] می باشد به حالتی که ذره ها در تابع احتمال (رویه شباهت) بوسیله روش جابجایی میانگین به سمت پیشینه های محلی (قله ها) کشیده می شوند. در هر حال ما روش ردیابی را بدین صورت برگزیده ایم که روش غیر پارامتری تاثیر گذارتری در مقایسه با جابجایی میانگین و فیلتر ذره ای است.

بعد از پخش و منتشر کردن ذره ها از طریق مدل دینامیکی $p(\mathbf{x}_t | \mathbf{x}_{t-1})$ ، بهینه سازی جابجایی میانگین برای هر ذره اجرا می شود. ذره ها در جهت صعودی گرادیان تابع احتمال حرکت داده می شوند تا زمانی که به قله های همسایگی شان همگرا شوند. مجموعه ذره های حاصل نشان دهنده مناسبی از مدهای توزیع شده بر روی یک رویه هستند که مشخصه های مطلوبی از تابع احتمال (رویه شباهت) را فراهم می کنند. همانطور که همیشه پیشینه های محلی به خوبی توسط مجموعه ذره ها نشان داده می شوند، ذره های بسیار کمتری نیز برای نشان دادن توزیع چند مدی مورد نیاز است.

نمونه برداری اصلی وقتی اعمال می شود که دانش کمکی به فرم یک تابع اهمیت $g_t(\mathbf{x}_t)$ در دسترس باشد و این تابع مشخص می کند که کدام نواحی فضای مکانی بیشترین اطلاعات را درباره احتمال پیشین موقعیت مکانی شیء در بر دارد. در نهایت با اجرای جابجایی میانگین بر روی ذره ها باعث حرکت ذره ها

بسوی قله ها (نواحی مهم رویه) می‌شویم. بعد از اجرای بهینه سازی جابجایی میانگین بر روی ذره ها می‌توانیم از ذره ها برای نمونه برداری از تابع اهمیت $g_t(\mathbf{x}_t)$ استفاده کنیم. برای جبران توزیع ناهموار و نامتعادل ذره ها یک فرم تصحیح f/g باید به وزن ذره ها اضافه شود بصورت زیر:

$$\pi_t^{(n)} = \frac{f_t(\mathbf{s}_t^{(n)})}{g_t(\mathbf{s}_t^{(n)})} p(\mathbf{z}_t | \mathbf{x}_t = \mathbf{s}_t^{(n)}) \quad (3-4)$$

$$f_t(\mathbf{s}_t^{(n)}) = p(\mathbf{x}_t = \mathbf{s}_t^{(n)} | \mathbf{z}_{1:t-1})$$

چگالی پیش بینی $p(\mathbf{x}_t | \mathbf{z}_{1:t-1})$ مخلوطی از کرنل های چگالی پیوستهء شکل داده شده است که بوسیله مدل دینامیکی $p(\mathbf{x}_t | \mathbf{x}_{t-1})$ تخمین زده می‌شود، پس اثر نسبت تصحیح به این خاطر است که اطلاعات پیوستگی حرکت شیء را که در مدل دینامیکی موجود است حفظ کند. اگرچه ذره ها بر اساس $g_t(\mathbf{x}_t)$ تولید شده اند، توزیع تخمین زده شده بوسیله $\{(s_t^{(n)}, \pi_t^{(n)})\}_{n=1}^N$ هنوز $p(\mathbf{x}_t | \mathbf{z}_{1:t})$ را تولید می‌کند. به عبارت دیگر تلفیق فیلتر ذره ای با جابجایی میانگین مدل تابع احتمالی (رویه شباهت) را تغییر نمی‌دهد، در صورتیکه بازه نمونه برداری توسط ذره ها را بسیار بهبود می‌بخشد.

در عمل تابع اهمیت $g_t(\mathbf{x}_t)$ ممکن است بعضی از قله های محتمل را هم بوسیله ذره ها حفظ کند، بنابراین ما ذره های تولید شده بوسیله نمونه برداری استاندارد از $p(\mathbf{x}_t | \mathbf{z}_{1:t-1})$ و ذره های تولید شده توسط نمونه برداری با تابع اهمیت (با استفاده از $g_t(\mathbf{x}_t)$) را به عنوان مجموعه ذره های فیلتر ذره ای با هم ترکیب می‌کنیم. در نتیجه تا آنجاییکه $g_t(\mathbf{x}_t)$ و $p(\mathbf{x}_t | \mathbf{z}_{1:t-1})$ همزمان و باهم در پیش بینی موقعیت شیء شکست نخورده اند، ردیابی موفقیت آمیز خواهد بود. توابع $f_t(\mathbf{x}_t)$ و $g_t(\mathbf{x}_t)$ بصورت مشخصی قابل دسترس نیستند، اما می‌توانند به عنوان یک مخلوطی از گوسی تخمین زده شوند.

$$f_t(\mathbf{x}_t) = \frac{1}{N} \sum_{n=1}^N G(s_t^{(n)}, \Sigma)(\mathbf{x}_t), \quad (4-4)$$

$$g_t(\mathbf{x}_t) = \frac{1}{2} \left(\frac{1}{N} \sum_{n=1}^N G(s_t^{(n)}, \Sigma)(\mathbf{x}_t) + \frac{1}{N} \sum_{n=1}^N G(s_t^{*(n)}, \Sigma)(\mathbf{x}_t) \right), \quad (5-4)$$

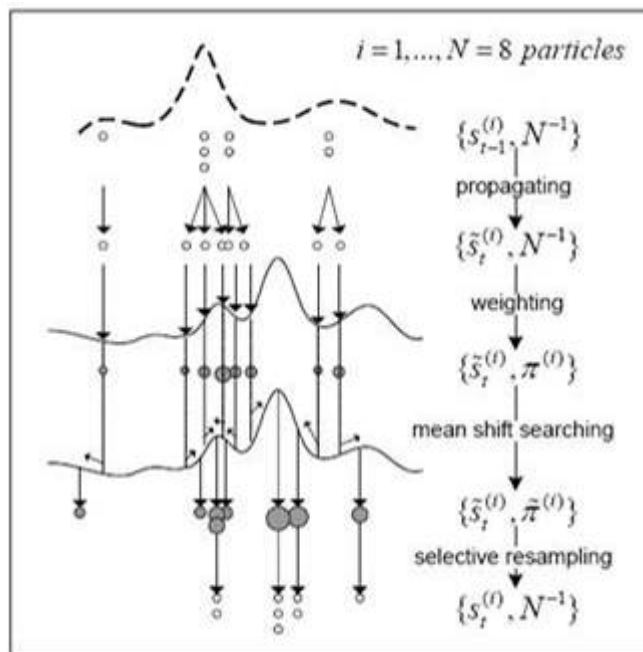
تابع فوق $s_t^{*(n)}$ ذره های جدید را بعد از اجرای روش جابجایی میانگین را بر روی ذره های قبلی ایجاد می کنند.

ما الگوریتم روش تلفیقی پیشنهادی را در شکل (۴-۳) خلاصه کرده و در شکل (۴-۴) بصورت شماتیک نشان می دهیم .

مجموعه ذره های داده شده در زمان $t-1$ ، مراحل زیر را طی می کنند:

۱. نمونه برداری دوباره از N ذره از مجموعه $\{(s_{t-1}^{(m)}, \pi_{t-1}^{(m)})\}_{m=1}^{2N}$ برای تولید $\{(s_{t-1}^{(n)}, \frac{1}{N})\}_{n=1}^N$.
۲. انتشار هر ذره بوسیله مدل دینامیکی $s_t^{(n)} \sim p(\mathbf{x}_t | \mathbf{x}_{t-1} = s_{t-1}^{(n)})$ برای تولید $\{(s_t^{(n)}, \frac{1}{N})\}_{n=1}^N$. (در این دو مرحله ذره ها را بر روی مکان ردیابی شده دست در فریم قبلی پخش می کنیم و سپس آنها را بر اساس مکان شان با کمک رنگ وزن دار می کنیم.)
۳. بهینه سازی هر ذره با استفاده از بهینه سازی جابجایی میانگین: $s_t^{(n)} \sim \text{Mean-Shift}(s_t^{(n)})$.
۴. ترکیب کردن ذره های تولید شده بوسیله نمونه برداری استاندارد و بهینه سازی آنها بوسیله روش جابجایی میانگین، $\{(s_t^{(m)}, \frac{1}{2N})\}_{m=1}^{2N} = \{(s_t^{(n)}, \frac{1}{2N})\}_{n=1}^N \cup \{(s_t^{(m)}, \frac{1}{2N})\}_{m=1}^N$ و محاسبه ضریب تصحیح $\lambda_t^{(m)} = f_t(s_t^{(m)}) / g_t(s_t^{(m)})$, $1 \leq m \leq 2N$.
۵. وزن دهی ذره ها به عنوان $\pi_t^{(m)} \propto \lambda_t^{(m)} p(\mathbf{z}_t | \mathbf{x}_t = s_t^{(m)})$ برای تولید $\{(s_t^{(m)}, \pi_t^{(m)})\}_{m=1}^{2N}$ ، جایکه $\sum_{m=1}^{2N} \pi_t^{(m)} = 1$ (در این سه مرحله ذره های وزن دار شده را بوسیله روش جابجایی میانگین به سوی قله های محلی حرکت داده و موقعیت بهینه شده شان را حساب می کنیم).
۶. محاسبه نتایج ردیابی (مکان احتمالی دست) $E[x_t] = \sum_{m=1}^{2N} \pi_t^{(m)} s_t^{(m)}$.

شکل ۴-۳- الگوریتم روش تلفیقی پیشنهادی



شکل ۴-۴- نمایش گرافیکی جابجایی میانگین بکار رفته در روش فیلتر ذره ای به عنوان روش تلفیقی پیشنهادی

۴-۴-۴- ردیابی دست با استفاده از روش پیشنهادی تلفیقی

این الگوریتم به همه ردیابی های تصویری می تواند اعمال شود. در این بخش، ما نحوه پیاده سازی روش تلفیقی پیشنهادی برای ردیابی دست را ارائه می دهیم.

۴-۴-۴-۱- مدل دینامیکی

در بسیاری از روشهای ردیابی دست [۱۰]، ناحیه ردیابی را بوسیله مربعی که دست (هدف) را در بر می گیرد ارائه داده می شود، که می توانیم به جای مربع از اشکال دیگر نظیر دایره و بیضی استفاده کنیم که این شکلها در واقع نسبت به شکل خود دست مستقل می باشند. فرض کنید که سایز مربع ها را ثابت در نظر بگیریم، در آن صورت موقعیت دست را بصورت زیر نشان می دهیم:

$$\mathbf{x} = \{x, y\}$$

جایی که x موقعیت مکانی است، (x, y) هم مختصات مرکز مربع است. دینامیک های حرکت دست بصورت زیر مدل می شوند:

$$\mathbf{X}_t - \mathbf{X}_{t-1} = \mathbf{X}_{t-1} - \mathbf{X}_{t-2} + w_{t-1} \quad (۴-۶)$$

معادله فوق درجایی است که w_{t-1} جزء تصادفی گوسی با میانگین صفر است. مجموعه ذره ها بوسیله این مدل سرعت ثابت ساده پخش می شوند (مرحله ۲، در شکل ۴-۳).

همانطور که در مرجع [۷۱] توصیف شده است، مدل دینامیکی از طریق یکسری از سکانسها می تواند آموزش داده شود. با وجود اینکه مدل فوق می تواند حدس و پیش بینی بسیار خوبی از انتشار ذره بزند اما مدل دینامیکی آموزش داده شده مدل جامعی نبوده و نمی تواند پیش بینی موثری از سکانسهایی که در مجموعه آموزشی نیستند داشته باشد. در روش تلفیقی و فیلتر ذره ای بسیار سخت است که یک مدل دینامیکی قوی را برای همه حرکات دست آموزش دهیم، در نتیجه ما مدل سرعت ضعیفی را انتخاب می کنیم که برای حرکت دست، طبیعی و جامع بوده و توان پیش بینی آن محدود است.

وقتی از مدل دینامیکی ضعیفی برای فیلتر ذره ای استفاده می شود، تعداد زیادی ذره نیاز خواهد بود. به هر حال در روش تلفیقی پیشنهادی، ذره ها پس از اینکه بوسیله مدل دینامیکی منتشر شدند، بصورت فعال به مدهای محلی (قله های محلی) انتقال داده می شوند که این می تواند به عنوان جبران ساز مدل دینامیکی با مشاهده جاری مد نظر قرار گیرد. پس روش تلفیقی پیشنهادی می تواند تا حدودی از عهده مدلهای دینامیکی ضعیف هم برآمده و امکان ردیابی بهتر با تعداد ذره های کمتری را فراهم کند.

۴-۲-۴-۴ مدل مشاهداتی

به منظور بهینه سازی وزن دهی به ذره ها به روش جابجایی میانگین (حرکت دادن ذره ها بسوی مدها)، لازمست که یک مدل مشاهداتی قابل اتکا

استخراج کنیم. هنگامیکه کاربرها اشارات دست را در زبان اشاره انجام می دهند، از آنجاییکه سرعت حرکت دست زیاد بوده و شکل دستها دائما در سکانسهای تصویری در حال تغییر است، بنابراین سخت است که از ویژگیهای هندسی دست نظیر کانتور دست بصورت قابل قبول استفاده کنیم. در اینجا ما از یکی از ویژگیهای غیر هندسی دست نظیر رنگ استفاده می کنیم همانطور که در فصل قبل هم برای ردیاب جابجایی میانگین استفاده شد. در صورتیکه نقاط و قسمتهای دیگری از بدن مشخص نباشند و یا اشیاء هم‌رنگ پوست در پیش زمینه وجود نداشته باشند، رنگ دست ویژگی متمایز دستها است، به همین دلیل نشانه های رنگ پوست بصورت فراوان در ردیابی دست مورد استفاده قرار می گیرند [۲۱-۱۹، ۱۲، ۶، ۵]. ویژگیهای رنگی دارای مزایایی هستند: از قبیل نامتغیر بودن (مستقل بودن) نسبت به چرخش، نسبت به تغییرات مقیاس و مقاومت در برابر مسدودیت بخشی (partial occlusion). مطابق [36] Bradski، ما ابتدا یک مدل رنگ پوست بر اساس هیستوگرام آموزش داده، سپس تصویر ورودی را به تصویر توزیع احتمال رنگ از طریق هیستوگرام رنگ تبدیل می کنیم. شکل (۴-۵-۵) (b) یک سکانس از دیتابیس زبان اشاره مورد استفاده در این پایان نامه است که دارای پیش زمینه ثابت می باشد و شکل (۴-۵-۴) (c) تصویر توزیع احتمال رنگ متناظر آن است. آشکار است که توزیع رنگ پوست ممکن است بخاطر تغییرات روشنایی عوض شود. در نتیجه ما مدل رنگ پوست را طوری تعریف می کنیم که فریم به فریم با نشانه های رنگی پوست به روز می شوند. به این ترتیب روش ما تا حدودی می تواند از عهده تغییرات روشنایی برآید البته در دیتابیس زبان اشاره مورد استفاده این تغییرات روشنایی محسوس نیست.

در یک محیط واقعی ارتباط بین انسان و ماشین (HCI)، اشیاء هم‌رنگ پوست نظیر صورت آدمها وجود دارند. برای حذف انحرافات که در اثر اشیاء هم‌رنگ پوست ایجاد می شود ما می توانیم نشانه های حرکتی را نیز در مدل مشاهداتی در نظر بگیریم. همانطور که دستها حرکت می کنند ما فرض می کنیم که اشیاء هم‌رنگ پوست نظیر صورت در پیش زمینه با سرعت کمتری نسبت به دست حرکت می کند. از آنجاییکه پیش زمینه در این دیتابیس معمولا ثابت است

ولی با حرکت دوربین می‌تواند تغییر کند (پیش زمینه غیر ثابت)، در نتیجه با تفاوت گیری موضعی دو فریم، میتوان حرکت را در پیش زمینه آشکار کرد. همانطور که در شکل ۴-۵ بصورت کلی مشاهده می‌شود ما از تفاوت گیری موضعی (تفریق تصاویر پیاپی) توصیف شده در مرجع [۷۲] استفاده می‌کنیم، که تفاوتها را در همسایگی اطراف محل دست در هر پیکسل، بصورت صفر و یک (سیاه و سفید) به ما نشان می‌دهد. با استفاده از روش آستانه گیری، آستانه ای را تعیین می‌کنیم تا بتوانیم پیکسل های دست را به محیط متحرک دست نسبت دهیم. بدین صورت فقط ذره های قرار گرفته در محیط های به رنگ پوست (که همان دست است) برای وزن دهی استفاده می‌شوند. عملگر AND منطقی را بین تصویر توزیع احتمال رنگ و تصویر اختلاف اعمال می‌کنیم تا تصویر توزیع احتمال رنگ و حرکت بدست آید. در نتیجه محیط های به رنگ پوست در حال حرکت را نشان می‌دهد. شکل (۴-۵-۱) و (۴-۵-۲) از یک سکانس با دو فریم پیاپی هستند، شکل (۴-۵-۳) تصویر توزیع احتمال رنگ و حرکت متناظر است.

نشانه حرکت و نشانه رنگ مانند زیر محاسبه می‌شوند:

$$\mathbf{M}_c = \sum_x \sum_y I_c(x, y), \quad \mathbf{M}_m = \sum_x \sum_y I_m(x, y) \quad (7-4)$$

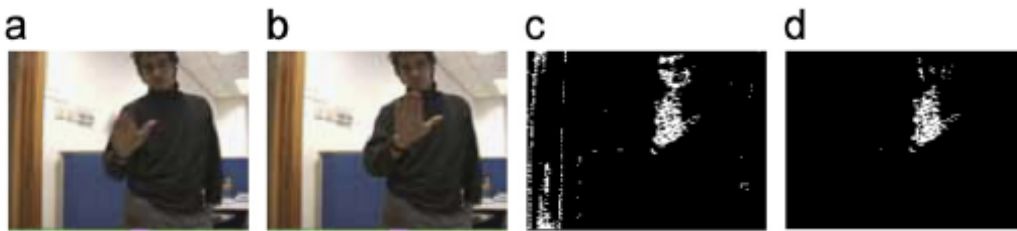
جاییکه M_c درجه یا گشتاور صفرم از مربع در تصویر توزیع احتمال رنگ است، جایی است که M_m گشتاور صفرم تصویر توزیع احتمال رنگ و حرکت است. I_c و I_m مقادیر پیکسل ها در مختصات (x, y) در تصویر توزیع احتمال رنگ و احتمال رنگ- حرکت است و x, y فاصله در روی مربع هستند. در حالت دیگر وقتی که یک دست خیلی آرام حرکت می‌کند، ذره های واقع شده در محیط دست (غریبال شده در آستانه گیری رنگ) در تصویر توزیع احتمال رنگ به وسیله ویژگی رنگ وزن دهی می‌شوند. در نتیجه ما می‌توانیم ترکیب خطی نشانه رنگ و نشانه حرکت را بصورت کلی بگیریم:

$$\begin{aligned} \mathbf{M} &= (1-\alpha)\mathbf{M}_c + \alpha\mathbf{M}_m \\ &= \sum_x \sum_y ((1-\alpha)I_c(x,y) + \alpha I_m(x,y)) \end{aligned} \quad (۸-۴)$$

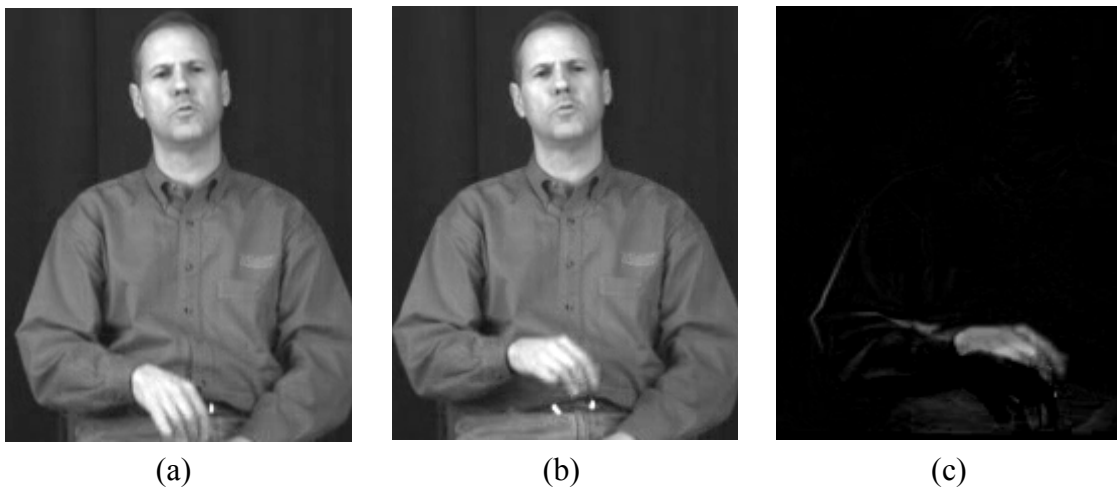
در معادله فوق که \mathbf{M} حاصل تلفیق گشتاورهای صفرم است، زمانیکه سرعت دست کمتر از یک مقدار آستانه است، ضریب α متناسب با سرعت دست است، در غیر این صورت:

$$\alpha = 1, i.e., \alpha = \min\{k\sqrt{\dot{x}^2 + \dot{y}^2}, 1\}$$

که در اینجا \dot{x} و \dot{y} حرکت x و y ، $\sqrt{\dot{x}^2 + \dot{y}^2}$ سرعت دست و k یک ضریب عملی و تجربی است که ما آنرا 0.04 بصورت نمونه قرار می دهیم.



شکل ۴-۵- a فریم ۸۹، b فریم ۹۰، c تصویر توزیع احتمال رنگ فریم ۹۰، d تصویر توزیع رنگ-حرکت فریم ۹۰ [۷۳].



شکل ۴-۶- a فریم ۸، b فریم ۹، c تصویر توزیع احتمال رنگ.

۴-۳-۴- بهینه سازی ذره توسط روش جابجایی میانگین

بر پایه مدل مشاهداتی (بر این اساس رویه شباهت بدست می آید)، ما از روش جابجایی میانگین برای کشاندن ذره ها به قله های محلی (مد ها) در رویه احتمال (رویه شباهت) بهره میگیریم (مرحله ۳ در شکل ۴-۳).

یک ذره داده شده $S^{(i)}$ با موقعیت $C_0(x_0, y_0)$ مطابقت دارد، ما $C_0(x_0, y_0)$ را به عنوان مکان اولیه انتخاب می کنیم و تعداد دفعات تکرار (iteration) را مقدار دهی اولیه می کنیم به عنوان مثال $itn = 0$.

مرحله ۱: نشانه های رنگ (می توان از نشانه حرکت هم در پیش زمینه غیر ثابت بهره برد) داده شده در داخل مستطیل را محاسبه می کنیم:

$$M_{00} = M = \sum_x \sum_y ((1-\alpha)I_C(x, y) + \alpha I_m(x, y)),$$

$$M_{10} = \sum_x \sum_y x \cdot ((1-\alpha)I_C(x, y) + \alpha I_m(x, y)), \quad (9-4)$$

$$M_{01} = \sum_x \sum_y y \cdot ((1-\alpha)I_C(x, y) + \alpha I_m(x, y)),$$

جاییکه M_{00} گشتاور صفرم است، M_{10} گشتاور مرتبه اول برای X و M_{01} گشتاور مرتبه اول برای Y است. α ضریبی است که در معادله ۴-۸ توصیف شده است.

مرحله ۲: مکان میانگین (نقطه ثقل) مستطیل بر اساس گشتاور صفرم و اول را حساب می کنیم (در روش جابجایی میانگین منظور جابجایی ذره بسوی همین میانگین در هر تکرار است):

$$x_l = \frac{M_{10}}{M_{00}}; \quad y_l = \frac{M_{01}}{M_{00}} \quad (10-4)$$

و سپس مرکز مستطیل در مکان میانگین $C(x_1, y_1)$ قرار می گیرد. همچنین تعداد تکرار را افزایش می دهیم: $itn = itn + 1$.

مرحله ۳: اگر $itn > itn_0$ یا $\|C_1 - C_0\| < \varepsilon$ تکرار متوقف شده و ذره $S^{(i)}$ با مکان C_1 به روز شود. در غیر اینصورت $C_0 = C_1$ قرار داده شده و به مرحله ۱ بازگردد.

شرط اول توقف تکرار این است که موقعیت میانگین کمتر از یک آستانه از پیش تعیین شده ε باشد (این آستانه بمعنی جابجایی ناچیز ذره است) و دیگری این است که تعداد تکرار حلقه بیشتر و بزرگتر از یک آستانه از پیش تعیین شده itn_0 باشد و به این خاطر است که از هزینه محاسباتی زیاد بدلیل تکرار بیش از حد حلقه خودداری می شود.

۴-۴-۴- تابع احتمال

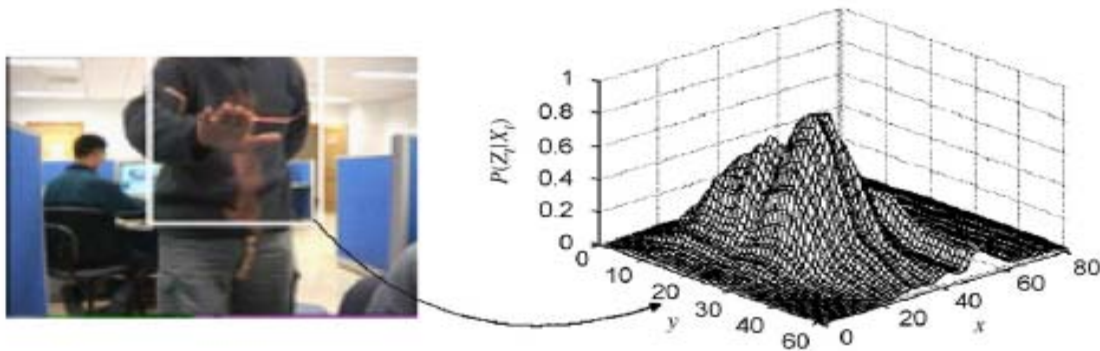
در این بخش تابع احتمال (تابع شباهت هم نامیده می شود) را که در وزن دهی ذره ها استفاده می شود معرفی می کند (مرحله ۵ در شکل ۴-۳). احتمال $p(\mathbf{z}_t | \mathbf{x}_t)$ ، مشاهده \mathbf{z}_t در هر تصویر (فریم) را به مکان \mathbf{x}_t مربوط می کند. همانطور که در معادله ۴-۸ معرفی شده است برای یک ذره $S^{(i)}$ ، مشاهده نشانه های (ویژگیهای) رنگ و حرکت (در پیش زمینه ثابت معمولا ویژگی حرکت استفاده نمی شود بدلیل ثبات پیش زمینه) در مستطیل متناظر است. برای محاسبه احتمال وجود هر ذره در یک موقعیت (که همان وزن ذره محسوب میشود)، ما از یک تابع شباهت استفاده می کنیم که فاصله بین هدف و کاندید را بصورت زیر مشخص می کند:

$$Dis = \sqrt{I - \mathbf{M}/\mathbf{M}_0}, \quad (11-4)$$

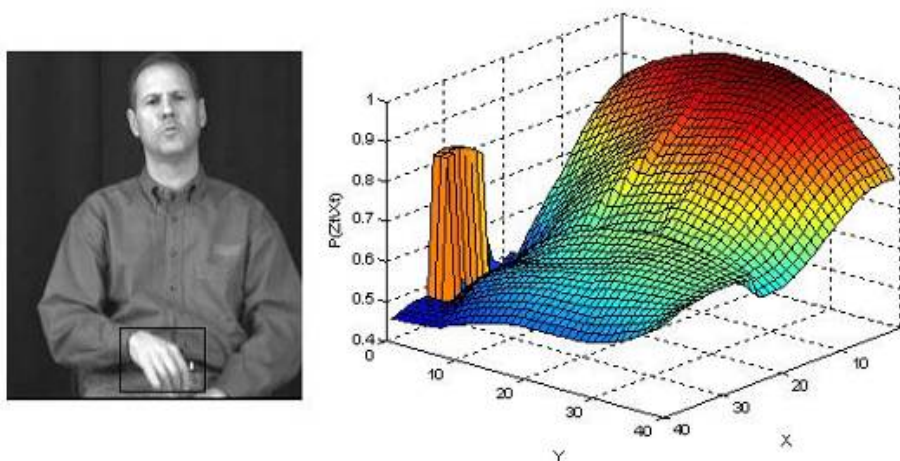
که در این جا \mathbf{M}_0 تعداد پیکسلهای موجود در مستطیل است. پس احتمال ذره $S^{(i)}$ بصورت زیر بوسیله یک تابع گوسی با واریانس سیگما (σ) مشخص شده است:

$$\pi^{(i)} = \frac{-I}{\sqrt{2\pi\sigma}} e^{-Dis^2/2\sigma^2} \quad (12-4)$$

که سیگما (σ) یک ثابت عملی و تجربی انتخاب شده با مشاهده اجرای ردیابی در اهداف متفاوت بویژه دست می باشد. برای نشان دادن کارآمدی تابع احتمال (همان تابع شباهت است که در فصل قبلی این تابع شباهت باتاچاریا بود که برای ردیابی شی به روش جابجایی میانگین استفاده شد)، شکل (۴-۷) توزیع مشاهده (رویه احتمال یا شباهت) در یک فریم را نشان می دهد: سمت راست منحنی مشاهده سطح اطراف دست در تصویر چپ را ترسیم می کند. بخاطر انعکاس نور آفتاب در پیش زمینه، توزیع مشاهده در قسمت بالایی دارای چند مد است (چندین قله دارد).



شکل ۴-۷- منحنی مشاهداتی توزیع چند مد در دنیای واقعی. سمت راست نقشه رویه مشاهداتی محیط مربع سفید تصویر سمت چپ است [۷۳].



شکل ۴-۸- منحنی مشاهداتی توزیع چند مد در تصویر دست. سمت راست نقشه رویه مشاهداتی محیط برای مربع سیاه تصویر سمت چپ است.

۴-۵- پیاده سازی الگوریتم روش تلفیقی پیشنهادی در تشخیص اشاره دست

تشخیص اشاره دست در تصاویر با پیش زمینه ثابت و متحرک گروه بندی شده است. در این بخش ما یک روش ساده برای تشخیص اشاره دست ارائه می‌دهیم. اشاره های دست، توسط ساختار گذرای فضایی الگوهای حرکتی شان (مسیر حرکت دست) مشخص می‌شوند. پس این اشارات معمولاً بوسیله تحلیل مسیرهای اشاره (حرکت دست) تشخیص داده می‌شوند که از طریق ردیابی دست در یک سکانس تصویری به دست می‌آید. مسیرهای اشاره می‌توانند در فضای حالت به صورت آماری با استفاده از الگوریتم‌هایی نظیر [۳۲] HMM و [۲۱] TDNN مدل شوند. در مرجع‌های [۷۵،۷۴] اشاره‌ها به عنوان مسیر الگو (Template Trajectory) نشان داده شده‌اند و مسیر ورودی (اشاره ورودی) با الگوهای ذخیره شده مقایسه و تطبیق (match) داده می‌شوند. در مرجع [۷۶] یک روش بدون مسیر براساس الگوهای گذرا، برای تشخیص اشاره دست ارائه داده شده است. خصوصاً اینکه آنها از تصویر انرژی- حرکت^۱ (MEI) و تصویر تاریخ (سابقه)- حرکت^۲ (MHI) برای تبدیل حرکت از حالت زمانی- فضایی به حالت تصویر ساکن (استاتیک) استفاده نموده‌اند. با تبدیل حرکت گذرا- فضایی به تصویر ساکن (استاتیک)، روش مورد نظر از تحلیل (آنالیز) حرکت گذرای دست برای تطبیق (match) کردن سکانسها پرهیز می‌کند. به علاوه اینکه این روش (متد الگوی گذرای بدون مسیر حرکت دست) به صورت محاسباتی بسیار کارآمد و به صرفه است و به آموزش پیچیده‌ای برای تشخیص اشاره دست در زبان اشاره احتیاج ندارد. به هر حال دو محدودیت در این روش بدون مسیر حرکت دست (منحنی) وجود دارد: اول اینکه تشخیص هر حرکت دست فقط و فقط بر پایه تحلیل شکل کلی ناحیه حرکت دست است، پس ویژگیهای گذرای- فضایی الگوی حرکت دست گم شده‌اند و دوم اینکه MHI و MEI بوسیله تفاوت

^۱ motion-energy image

^۲ motion-history image

گیری تصویر تولید می شوند که برای پیش زمینه غیر ثابت (دینامیک و متحرک) نامناسب است. برای حل این مشکلات و آمادگی در برابر آنها، از طریق معرفی مسیر (منحنی) حرکت دست بر روی الگوهای گذرا، ما یک روش ساده برای تشخیص اشاره دست ارائه می دهیم.

برای سکانس تصویری اشاره دست $I(x, y, t)$ ، ابتدا نواحی حرکت در هر تصویر را بوسیله ردیابی دست بدست می آوریم. $D(x, y, t)$ سکانس تصویری باینری (دوتایی) است که نواحی حرکت دست را نشان می دهد. مدل اشاره دست بصورت زیر تعریف شده است:

$$H_{\tau}(x, y, t) = \begin{cases} \tau & \text{if } D(x, y, t) = 1 \\ \max(0, H_{\tau}(x, y, t-1) - 1) & \text{otherwise} \end{cases}$$

که τ بقای الگو بر اساس [۷۶] است. در این روش مسیرهای حرکت دست که به صورت گذرا و فضایی هستند، اشاره های دست را در تصاویر تولید می کنند. الگوهای گذرا بر اساس مسیر حرکت دست که اشارات دست را ایجاد می کنند منحنی های TTBT^۱ نامیده می شوند. در مقایسه با الگوهای گذرای اولیه، TTBT اطلاعات (دیتاهای) گذرای- فضایی واضحی از مسیر حرکت دست را در بردارند که با استفاده از این اطلاعات می توان بهتر اشارات دست را تشخیص داد. این منحنی ها الزاماً به پیش زمینه ساکن (ثابت) احتیاجی ندارند و می توانند در پیش زمینه های متغیر (محیط های دینامیک و متحرک) برای تشخیص اشارات دست عمل کنند. ما پنج اشاره دینامیکی را تعریف می کنیم، موجی (WV)^۲، حرکت افقی (HR)^۳، حرکت عمودی (VT)^۴، حرکت در جهت همراه عقربه های

^۱ temporal template based trajectories

^۲ wave

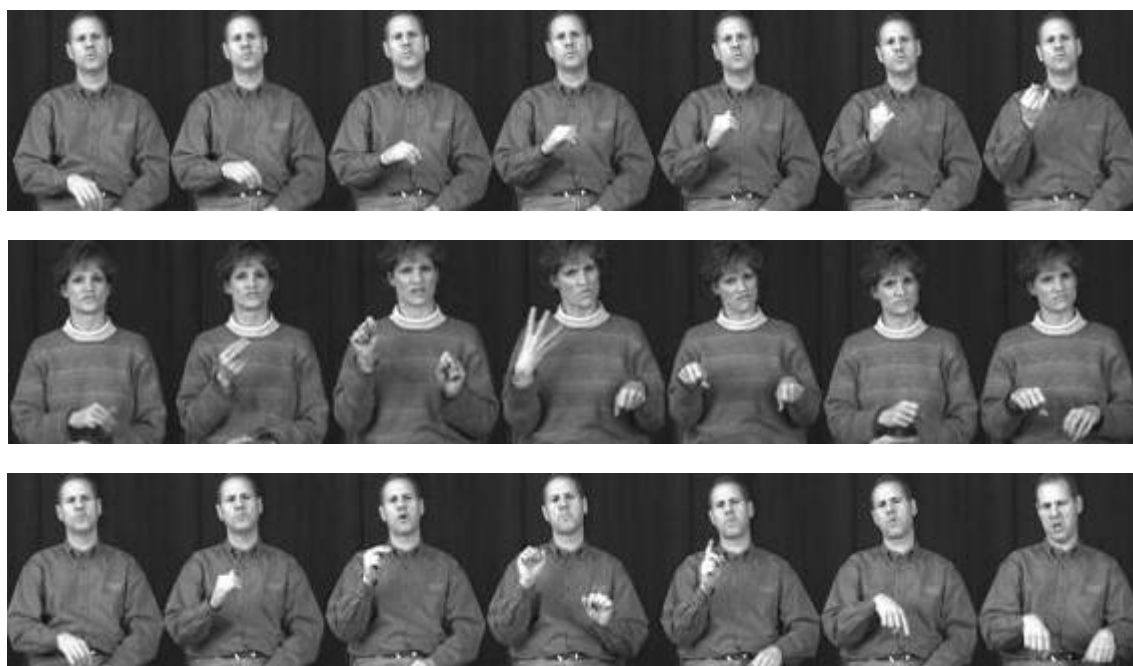
^۳ move horizontally

^۴ move vertically

ساعت (CW)^۱ و در جهت خلاف عقربه های ساعت (ACW)^۲. مثالهایی از این اشارات دست در زبان اشاره به ترتیب در شکل (۹-۴) و (۱۰-۴) نشان داده شده است.



شکل ۴-۹- نمونه های اشاره حرکت موجی، ساعتگرد و حرکت افقی. از بالا به پایین [۷۳].

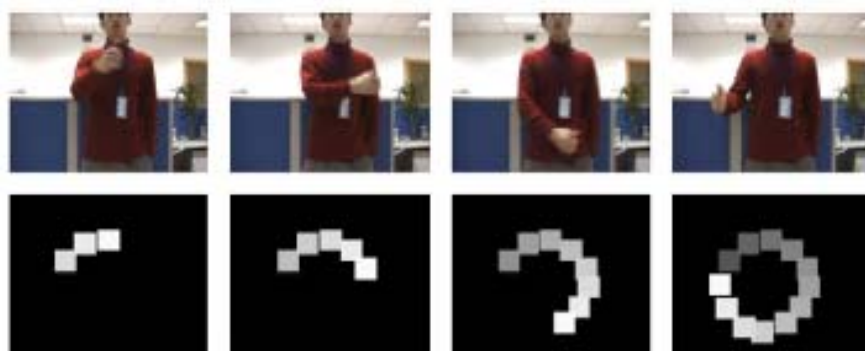


شکل ۴-۱۰- نمونه هایی از زبان اشاره با حرکت دست از بالا به پایین.

^۱ clockwise

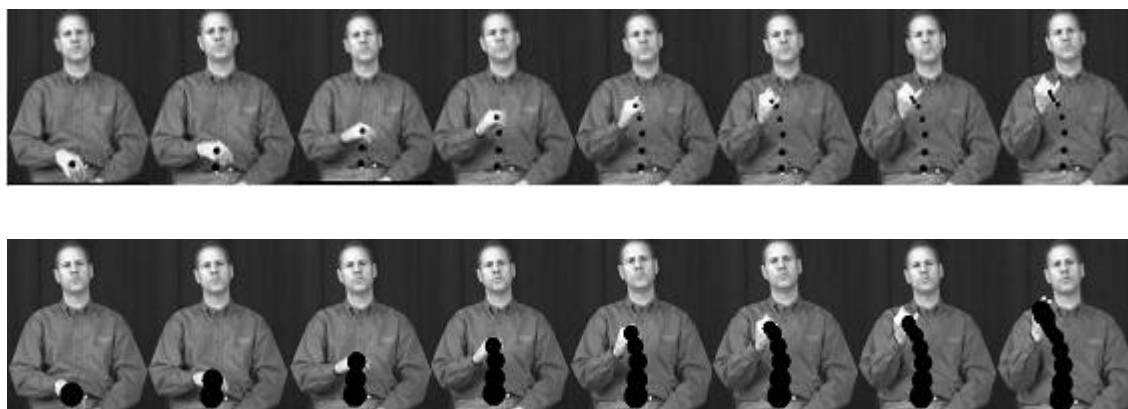
^۲ anti-clockwise

تولید منحنی های TTBT (الگوهای گذرای مبتنی بر مسیر حرکت دست) برای اشاره خلاف عقربه ساعت ACW در شکل (۴-۱۱) نشان داده شده است.



شکل ۴-۱۱- تولید TTBT برای اشاره پادساعتگرد [۷۳].

تولید منحنی TTBT برای اشاره دست (اشاره های معنی دار) در شکل (۴-۱۲) و (۱۲-۴) نشان داده شده است.



شکل ۴-۱۲- تولید TTBT برای اشاره دست در زبان اشاره.



شکل ۴-۱۳- TTBT برای پنج اشاره از پیش تعریف شده. از چپ به راست: حرکت موجی (wave)، پادساعتگرد، ساعتگرد، افقی، عمودی [۷۳].

ما تشخیص اشاره دست را بر اساس تحلیل (آنالیز) حرکت منحنی TTBT برای یک دست انجام می دهیم. از ویژگی های بارز شکل منحنی TTBT مربوط به اشارات دست در زبان اشاره می توان به نامتغیر بودن این منحنی ها به جابجایی و مقیاس اشاره کرد. پس از استخراج این اشارات توسط منحنی های فوق، ما میتوانیم از ماتریس فاصله ماهالانوبیس^۱ به عنوان یک ویژگی خوب برای کلاسه بندی این اشارات استفاده می کنیم. منحنی TTBT برای اشاره افقی (HR) و عمودی (VT) شکل مشابهی دارند، همانطوریکه اشاره موافق و مخالف عقربه ساعت CW و ACW اشکال مشابهی دارند. پس ما در ادامه اطلاعات (دیتا) گرایش و جهت حرکت منحنی TTBT را برای تشخیص اشارات دست در زبان اشاره استفاده می کنیم زیرا TTBT برای اشارات مشابه شکل یکسانی دارد.

همانطور که اشارات دست ممکن است با سرعت های متغیر انجام داده شود، در قسمت آموزش، ما بیشینه (ماکزیمم) و کمینه (مینیمم) طول اشارات دست را اندازه گیری می کنیم $(\tau_{\max}$ و $\tau_{\min})$. در قسمت تشخیص اشاره دست، ما $\Delta\tau$ را $((\tau_{\max} - \tau_{\min}) / (n - 1))$ انتخاب می کنیم، که در آن n تعداد پنجره های گذرایی است که برای نشان دادن منحنی TTBT مد نظر قرار می گیرد، سپس در هر مرحله ما می توانیم به n مقدار متفاوت τ ، بوسیله فرمول زیر دست پیدا کنیم:

$$\tau_i = \tau_{\max} - \Delta\tau \cdot (i - 1), \quad i = 1, \dots, n$$

در نتیجه، n مقدار پنجره استخراج شده برای تولید منحنی TTBT است که برای انجام تشخیص اشاره دست در زبان اشاره به کار می رود.

۴-۶- نتایج عملی الگوریتم های جابجایی میانگین، فیلتر ذره ای

و روش پیشنهادی تلفیقی

^۱ Mahalanobis

در این بخش ما نتایج ردیابی دست در دیتابیس زبان اشاره با پیش زمینه ثابت و غیر ثابت (متحرک) را ارائه می دهیم. سپس برای اینکه کارایی و کارآمدی الگوریتم روش تلفیقی پیشنهادی در ردیابی دست را بصورت عملی نشان دهیم، نتایج عملی آن را با دو الگوریتم فیلتر ذره ای و جابجایی میانگین مقایسه کرده و از این روش پیشنهادی در نهایت برای ردیابی دست به منظور تشخیص اشاره دست در زبان اشاره استفاده می کنیم.

۴-۶-۱- ردیابی دست

آزمایشات عملی ردیابی دست ابتدا بر روی بیش از ۳۰ سکانس ویدئویی از اشارات دست در زبان اشاره با پیش زمینه متحرک انجام شده اند. این سکانس های ویدئویی در نرخ نمونه برداری 12fps (فریم در هر ثانیه) با استفاده از یک دوربین Sony، در محیط های عادی (معمولی) اتاق کار گرفته شده اند. هر سکانس حدوداً ۴۰۰ فریم را در بر دارد و وضوح تصویر 240x180 پیکسل است، هیچ محدودیتی در خصوص روشنایی و پیش زمینه در تنظیمات تصویربرداری گذاشته نشده است، بنابراین روشنایی متغیر و اشیاء مزاحم و پارازیت در پیش زمینه بعضی از سکانس ها کاملاً وجود دارد. سپس خودمان ردیابی دست به هر سه روش را بر روی دیتابیس زبان اشاره با نام RWTH-Boston اعمال می کنیم. هر سه الگوریتم ردیابی بر روی یک کامپیوتر نوت بوک (Asus Pentium 2.4G, Windows Vista با Matlab 8 بر روی دیتابیس زبان اشاره ذکر شده پیاده سازی شده اند.

۴-۶-۱-۱- مطالعات مقایساتی

ما در ابتدا کارایی اجرای الگوریتم روش تلفیقی پیشنهادی را در مقایسه با ردیاب جابجایی میانگین و فیلتر ذره ای بررسی می کنیم. به خاطر مقایسه عادلانه، در ردیاب فیلتر ذره ای از مدل دینامیکی و مدل مشاهداتی (تابع احتمال) یکسان با ردیاب تلفیقی پیشنهادی استفاده می کنیم و همچنین ردیاب جابجایی میانگین را همانطوریکه در فصل قبل در ردیابی اشیا توضیح داده شد

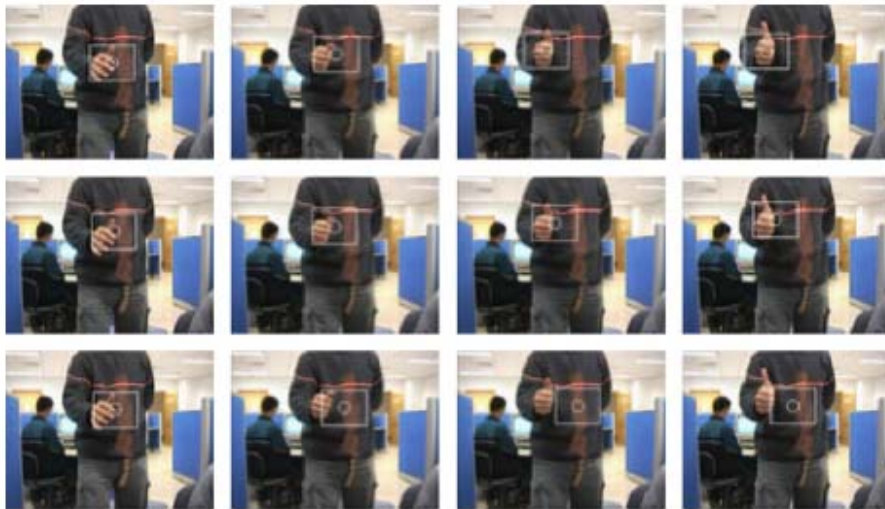
برای ردیابی دست با تابع احتمال (شباهت) یکسان با دو ردیاب دیگر پیاده سازی می کنیم [۶۰]. در ردیابی دست به روش جابجایی میانگین اطلاعات رنگ را همانند روش تلفیقی و فیلتر ذره ای استفاده می کنیم.

زمانی که مرکز حقیقی ناحیه دست (Ground Truth) در مستطیل ردیابی دست مورد نظر وجود نداشته باشد شکست در ردیابی دست را اعلام می کنیم. در آن زمان ردیابی برای سکانس در آن لحظه متوقف می شود، حتی با اینکه ممکن است بصورت تصادفی ردیاب دوباره دست را در مسیر ردیابی شده پیدا کند. حقیقت زمینه (Ground Truth) که همان مرکز محاسبه شده حقیقی دست است، بصورت دستی نام گذاری می شود. جدول ۴-۱ کارایی و کارآمدی هر سه ردیاب را باهم نشان می دهد.

جدول ۴-۱- نتایج کلی ردیابی دست به روش های تلفیقی پیشنهادی برای ۵۰ ذره و فیلتر ذره ای برای ۲۰۰ ذره و جابجایی میانگین.

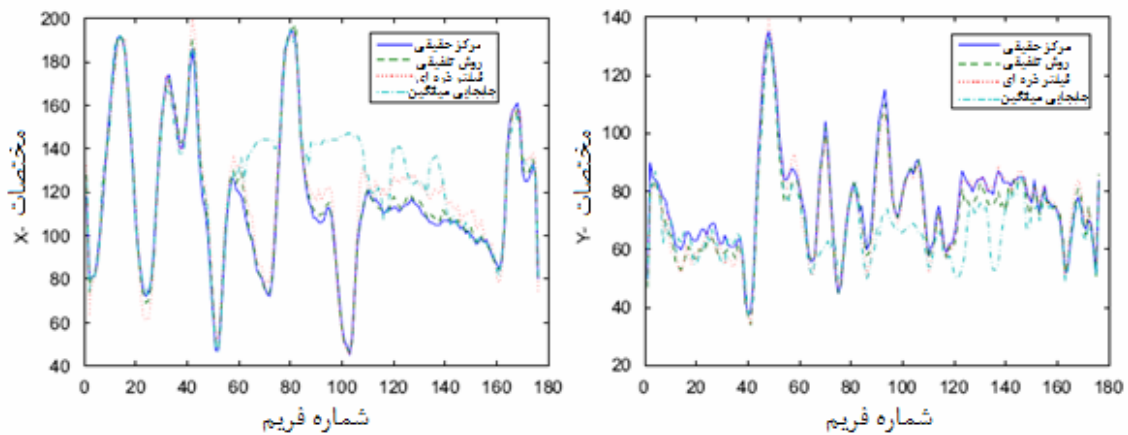
روش جابجایی میانگین	روش فیلتر ذره ای	روش تلفیقی پیشنهادی	نرخ های صحیح ردیابی
73.3%	100%	100%	

ما می توانیم مشاهده کنیم که روش تلفیقی پیشنهادی (۵۰ذره) و فیلتر ذره ای (۲۰۰ ذره) کارآمدی و کارایی بسیار بهتری را نسبت به ردیاب جابجایی میانگین نشان می دهند. بعضی از نتایج ردیابی دست در یک سکانس آزمایشی با پیش زمینه متحرک و در محیط RGB در شکل ۴-۱۴ نشان داده شده اند. در این سکانس، حرکت های ممتد بدن باعث می شود که بر روی بخش بالایی بدن روشنایی هم رنگ پوست به وجود بیاید که باعث چند مد شدن رویه شباهت می شود که در نتیجه ردیابی دست را با مشکل مواجه می کند. پس به این ترتیب سطح احتمال (رویه شباهت) در فریم های زیادی چندین مد دارد. دقت ردیابی هر سه الگوریتم برای بخشی از این سکانس (بالا) در شکل ۴-۱۴ و جدول ۴-۲ رسم شده است.



شکل ۴-۱۴-الف- بعضی از نتایج ردیابی دست به روش های جابجایی میانگین، فیلتر ذره ای و روش تلفیقی پیشنهادی برای یک سکانس آزمایشی با پیش زمینه متحرک از فریم ۶۱ تا ۶۴ [۷۳].

ردیف بالا: روش تلفیقی پیشنهادی، ردیف وسط: فیلتر ذره ای، ردیف پایین: جابجایی میانگین



شکل ۴-۱۴-ب- مقایسه دقت ردیابی برای سکانس آزمایشی بالا. چپ: مختصات x و راست: مختصات y [۷۳].

جدول ۴-۲- نتایج ردیابی دست به روش جابجایی میانگین و فیلتر ذره ای و روش تلفیقی در سکانس بالا [73].

الگوریتم	ذره های مورد نیاز	میانگین زمان در هر فریم میلی ثانیه (ms)	میانگین خطای مکانی در پیکسل ها
روش تلفیقی	۲۰	۲۸	۵/۸
روش فیلتر ذره ای	۱۵۰	۶۳	۷/۸
روش جابجایی میانگین	۱	۲۰	بعضی مواقع رد گم می شود

روش تلفیقی پیشنهادی

روش فیلتر ذره ای

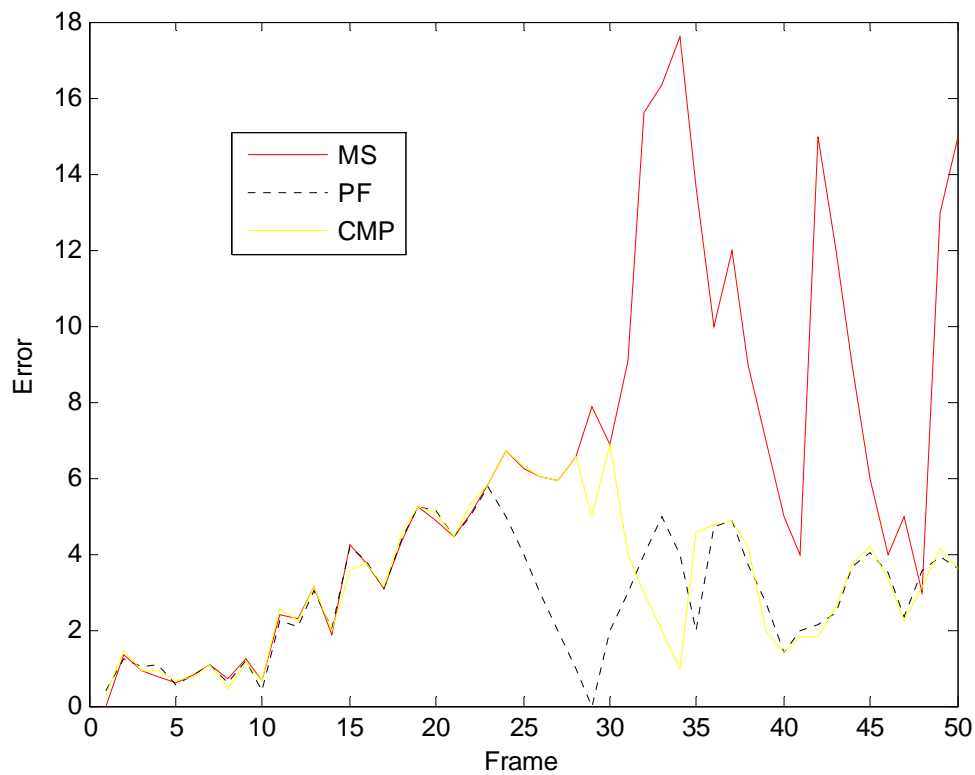
روش جابجایی میانگین





شکل ۴-۱۵-الف- بعضی از نتایج ردیابی دست به روش های جابجایی میانگین ، فیلتر ذره ای و روش تلفیقی پیشنهادی برای یک سکانس آزمایشی زبان اشاره از فریم 17 تا 27.

ستون راست: روش تلفیقی پیشنهادی ، ستون وسط: فیلتر ذره ای ، ستون چپ: جابجایی میانگین.



شکل ۴-۱۵-ب- مقایسه دقت ردیابی دست برای سکانس آزمایشی بالا به روش جابجایی میانگین MS و فیلتر ذره ای PF و روش تلفیقی CMP.

جدول ۴-۳- نتایج ردیابی دست به روش جابجایی میانگین و فیلتر ذره ای و روش تلفیقی در سکانس بالا.

میانگین زمان ردیابی دست در هر فریم (ثانیه)(s)	میانگین خطای مکانی در پیکسل ها	ذره های مورد نیاز	الگوریتم
۱,۵	۳/۲	۲۰	روش تلفیقی
۵,۴	۲/۱	۱۵۰	روش فیلتر ذره ای
۰,۷	۹/۸	۱	روش جابجایی میانگین

بعضی دیگر از نتایج ردیابی دست در یک سکانس آزمایشی زبان اشاره، در شکل ۴-۱۵ و ۴-۱۶ و ۴-۱۷ نشان داده شده اند. تخمین مکان دست را در یک دنباله تصویر ۳۹۴ فریمی که ابعاد هر فریم در آن برابر 176×144 پیکسل است صورت می گیرد. در این آزمایش الگوریتم جابجایی میانگین، فیلتر ذره ای و روش تلفیقی پیشنهادی یک بار هر کدام برای ویژگی رنگ به کار گرفته شده اند.

جابجایی میانگین

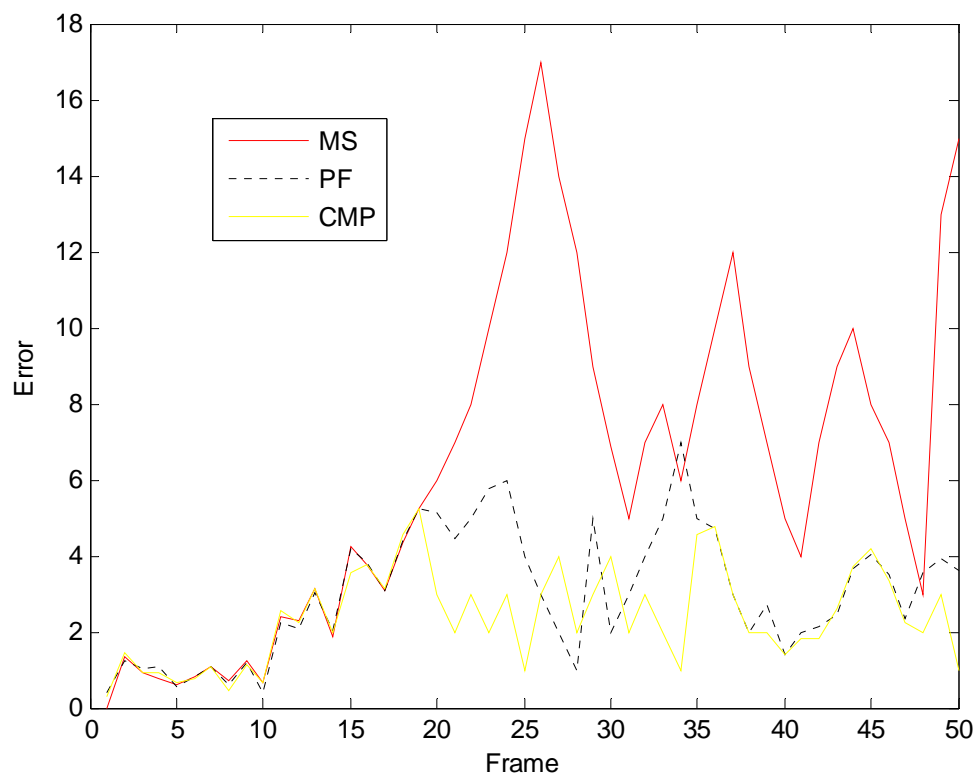
روش فیلتر ذره ای

روش تلفیقی





شکل ۴-۱۶-الف- بعضی از نتایج ردیابی به روش های جابجایی میانگین، فیلتر ذره ای و روش تلفیقی پیشنهادی برای یک سکانس آزمایشی زبان اشاره از فریم ۲۱ تا ۳۵.



شکل ۴-۱۶-ب- مقایسه دقت ردیابی الگوریتم های جابجایی میانگین MS و فیلتر ذره ای PF و روش تلفیقی CMP در هر فریم از دنباله تصویر دست در زبان اشاره برای ویژگی رنگ برای سکانس آزمایشی بالا.

جدول ۴-۴- نتایج ردیابی دست به روش جابجایی میانگین و فیلتر ذره ای و روش تلفیقی در سکانس بالا.

الگوریتم	ذره های مورد نیاز	میانگین زمان در هر فریم ثانیه (s)	میانگین خطای مکانی در پیکسل ها
روش تلفیقی	۲۰	۲,۲	۳
روش فیلتر ذره ای	۱۵۰	۵,۰	۳/۷
روش جابجایی میانگین	۱	۱,۰	۱۲

همانطور که در شکل (۴-۱۶) و جدول ۴-۴ دیده می شود الگوریتم جابجایی میانگین تنهایی مبتنی بر ویژگی رنگ در بعضی فریم ها از تخمین دقیق مکان دست ناتوان است در حالیکه الگوریتم فیلتر ذره ای و روش تلفیقی با استفاده از همان ویژگی رنگ در آن فریم ها موفق عمل می کنند. از سوی دیگر الگوریتم جابجایی میانگین مبتنی بر ویژگی رنگ در زمانی که دست از نزدیکی صورت می گذرد به علت رنگ مشابه صورت و دست ناموفق عمل کرده، در حالی که الگوریتم فیلتر ذره ای و روش تلفیقی پیشنهادی با ویژگی رنگ موفق به تخمین دقیق مکان دست در این سکانس در تمام فریم ها شده است. زمان اجرای این الگوریتم ها در این دنباله تصویر ۳۸ ثانیه ای برابر با جدول آن می باشد. در شکل (۴-۱۷) و جدول ۴-۵ هم نتایج ردیابی در هر فریم برای هر کدام از الگوریتم ها با استفاده از ویژگی رنگ را نشان داده شده است.

جابجایی میانگین

فیلتر ذره ای

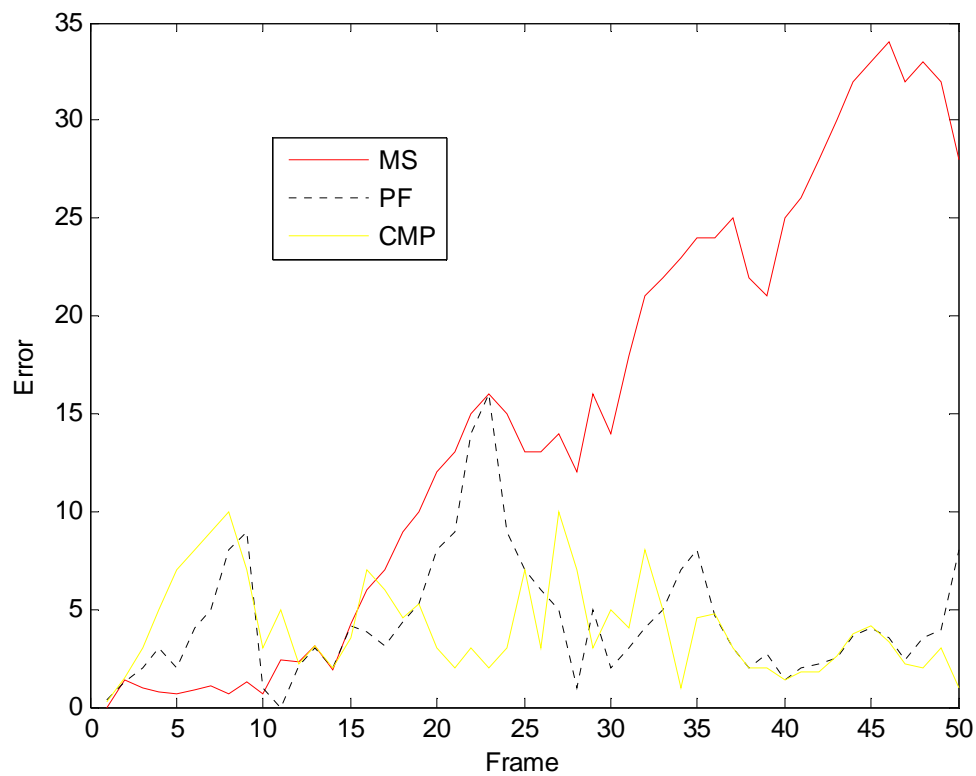
روش تلفیقی







شکل ۴-۱۷-الف- بعضی از نتایج ردیابی دست به روش های جابجایی میانگین و فیلتر ذره ای و روش تلفیقی پیشنهادی برای یک سکانس آزمایشی زبان اشاره.



شکل ۴-۱۷-ب- مقایسه دقت ردیابی الگوریتم های جابجایی میانگین MS و فیلتر ذره ای PF و روش تلفیقی CMP در هر فریم از دنباله تصویر دست در زبان اشاره برای ویژگی رنگ برای سکانس آزمایشی بالا.

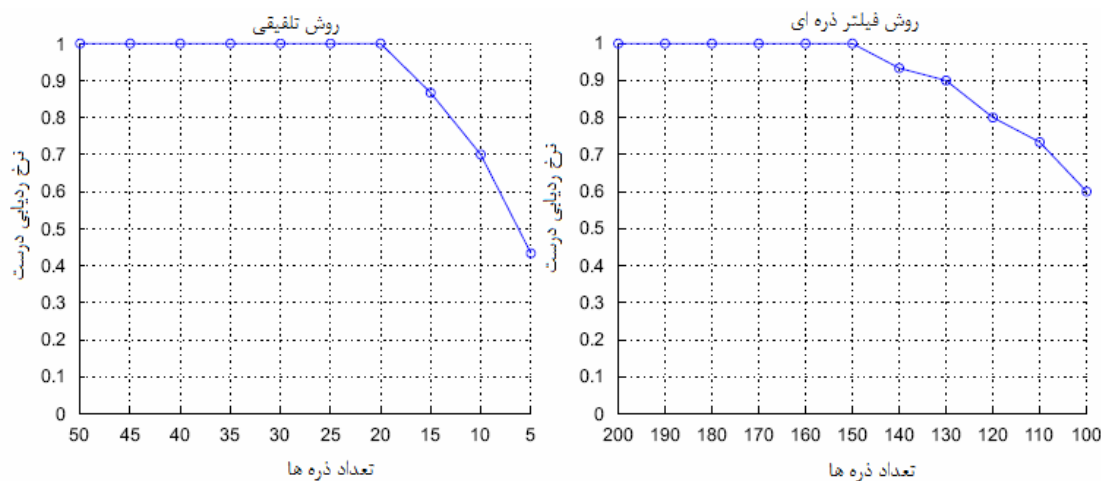
جدول ۴-۵- نتایج ردیابی دست به روش جابجایی میانگین و فیلتر ذره ای و روش تلفیقی در سکانس بالا.

الگوریتم	ذره های مورد نیاز	میانگین زمان در هر فریم ثانیه (s)	میانگین خطای مکانی در پیکسل ها
روش تلفیقی	۲۰	۱,۸	۵
روش فیلتر ذره ای	۱۵۰	۳,۵	۸
روش جابجایی میانگین	۱	۱,۰	رد دست گم می شود

از شکل (۴-۱۷) به روشنی در می یابیم که با بهره گیری از الگوریتم های جابجایی میانگین، فیلتر ذره ای و روش تلفیقی با ویژگی رنگ نتایج متفاوتی در هر فریم خواهیم داشت.

همانطور که پیداست در شکل‌های ۴-۱۴ و ۴-۱۵ و ۴-۱۶ و ۴-۱۷، روش تلفیقی و فیلتر ذره ای بصورت موفق دست را ردیابی می کنند ولی با این تفاوت که زمان اجرای الگوریتم روش تلفیقی پیشنهادی خیلی کمتر از روش فیلتر ذره ای است (تقریباً کمتر از نصف). در صورتیکه روش جابجایی میانگین در بعضی از مواقع دست را بطور کامل گم می کند و بطور کلی در ردیابی دست دچار مشکل می شود، برای مثال، در ۴-۱۷-الف و بخاطر آشفتگی ایجاد شده بوسیله روشنایی هم‌رنگ پوست دست و صورت، دست کامل گم می شود. نهایتاً به این نتیجه می‌رسیم که روش جابجایی میانگین نمی تواند به سطوح احتمال چند مدی رسیدگی کند، پس به آسانی به یک بیشینه (ماکزیمم) محلی همگرا می شود و در صورت وجود شیء هم‌رنگ دست، ردیابی دست را نمی تواند با موفقیت ادامه دهد. بالعکس فیلتر ذره‌ای و روش تلفیقی می توانند بصورت مقاوم دست را در بین پارازیتها و نویزهای موجود در پیش زمینه ردیابی کنند. در نتیجه این دو روش قابلیت رسیدگی به سطح چند مد را دارند. برای اینکه تحقیق کنیم که جابجایی میانگین بکاررفته چقدر توانسته است کارایی نمونه برداری را در روش

فیلتر ذره ای برای ردیابی دست بهبود دهد، ما مطالعه مقایساتی را نیز بر روی تعداد ذره های مورد نیاز روش تلفیقی پیشنهادی و فیلتر ذره ای برای داشتن خطای ردیابی دست یکسان هم بررسی کردیم. شکل ۴-۱۸ بصورت کلی نتایج ردیابی را هنگامیکه تعداد ذره ها تغییر می کنند نشان می دهد (برای مقایسه بهتر روش تلفیقی پیشنهادی با فیلتر ذره ای).



شکل ۴-۱۸- مقایسه خطای ردیابی دست با توجه به تعداد ذره ها. چپ: روش تلفیقی و راست: روش فیلتر ذره ای

همچنین مشاهده شده است که با حرکت و تکانهای زیاد دست و به کار بردن یک مدل دینامیکی ضعیف، فیلتر ذره ای حداقل به ۱۵۰ ذره برای ردیابی پایدار و مقاوم برای سکانسهای آزمایشی (تستی) احتیاج دارد، در صورتیکه روش تلفیقی پیشنهادی فقط به ۲۰ ذره احتیاج دارد. با توجه به زمان محاسبات، فیلتر ذره ای (۱۵۰ ذره)، ۳,۵ ثانیه بصورت میانگین برای هر فریم زمان مصرف می کند، در حالیکه برای روش تلفیقی پیشنهادی (۲۰ ذره) با وجود اینکه تکررهای روش جابجایی میانگین اضافه می شود، در هر فریم فقط یک متوسط زمانی ۱,۸ ثانیه را برای ردیابی دست لازم دارد زیرا تعداد ذره ها به مقدار زیادی کاهش می یابد. در نتیجه روش تلفیقی ۸۵٪ تعداد ذره های لازم برای ردیابی دست را نسبت به ردیاب فیلتر ذره ای کاهش می دهد و به همین دلیل خیلی سریعتر از فیلتر ذره ای دست را ردیابی می کند. جدول ۴-۲ و ۴-۳ و ۴-۴ و ۴-۵ اجرای ردیابی دست توسط روش تلفیقی پیشنهادی را برای سکانسهای آزمایشی نشان می دهد.

ما مشاهده می کنیم که روش تلفیقی نتایج بهتر ردیابی نسبت به جابجایی میانگین و ردیابی سریعتر نسبت به فیلتر ذره ای را ارائه می دهد و به طور کلی نتایج قابل قبول تری نسبت به هر دو روش فیلتر ذره ای و جابجایی میانگین بدست می آورد. با اینکه جابجایی میانگین خیلی سریعتر از روش تلفیقی است، جابجایی میانگین نمی تواند ردیابی پایدار و مقاومی را انجام دهد این بدین معنی است که بخاطر عدم دقت کافی و داشتن خطای قابل ملاحظه با مرکز واقعی دست، بعد از چندین فریم این خطا باعث می شود که ردیاب جابجایی میانگین رد دست را بکلی گم کند. مشخصاً و آشکارا بهبود بازدهی روش تلفیقی نسبت به فیلتر ذره ای به این موضوع بستگی دارد که چقدر خوب مکانهای ذره ها در فیلتر ذره ای به روز می شوند که این موضوع را تکرر جابجایی میانگین می تواند فراهم کند در نتیجه بصورت کلی نمی توانیم بگوییم که چقدر تعداد ذره ها دقیقاً می تواند در روش تلفیقی کاهش یابد. چیزی که می توانیم نتیجه بگیریم این است که به کاربردن روش جابجایی میانگین در روش فیلتر ذره ای کارایی و بازدهی روش فیلتر ذره ای را بهبود می بخشد. در این پایان نامه، ما ذره ها را در فیلتر ذره ای بوسیله روش جابجایی میانگین بهینه می کنیم که ما این پروسه را در روش تلفیقی پیشنهادی ارائه دادیم. در مصارف موجود معمولاً تعداد ذره ها را به کمک معیار ad hoc (سعی و خطا) یا به کمک متدهای آماری نظریه شبیه سازی monte carlo یا بعضی از مرزها و محدودیت های آماری [۷۷] انتخاب می کنند. برای مثال Boer [۷۷] روشی را مطرح کرد برای تخمین تعداد ذره های لازم بوسیله pdf $p(\mathbf{x}_t | \mathbf{z}_{1:t})$ که توزیع گسسته احتمال مشاهداتی است. اکثر فیلتر های ذره ای موجود تعداد ثابتی از ذره ها را انتخاب می کنند. به هر حال این امر ضروری نبوده و می تواند بسیار ناکارآمد باشد. برای مثال در ابتدا تعداد زیادی از ذره ها مورد نیاز است ولی همینطور که زمان جلو می رود و پراکندگی (واریانس) کاهش می یابد تعداد کمی از ذره ها برای انجام تخمین خوب کفایت می کند.

۴-۱-۶-۲- نتایج بیشتر ردیابی

برای تأیید بازدهی بهتر الگوریتم روش تلفیقی با استفاده از ویژگی های رنگ و حرکت با هم، ما الگوریتم روش تلفیقی را که روی بعضی از سکانسها با پیش زمینه متحرک پیاده شده است و حاوی انحراف ایجاد شده بوسیله اشیاء هم رنگ پوست در پیش زمینه می باشد را نشان می دهیم. شکل ۴-۱۹ بعضی نتایج ردیابی یک سکانس آزمایشی را با دو مدل مشاهداتی متفاوت نشان می دهد.



شکل ۴-۱۹- مقایسه بین مدل های مشاهداتی برای یک سکانس آزمایشی (از فریم ۵ تا ۸). ردیف بالا: مدل مشاهداتی بر اساس ویژگی های (نشانه های) رنگ و حرکت. ردیف پایین: مدل مشاهداتی بر اساس فقط نشانه های رنگ [۷۳].

در شکل ۴-۱۹ مشاهده شده است که ردیاب بر اساس فقط ویژگی رنگ، دست را بخاطر تداخل با صورت گم می کند. در صورتیکه ردیاب مجهز به مدل مشاهداتی پیشنهادی (رنگ و حرکت)، بصورت موفقیت آمیز از انحرافات مصون می ماند. پس ردیابی بر اساس اتصال ویژگی های رنگ و حرکت خیلی مطمئن تر از ردیابی توسط ویژگی رنگ می باشد. نتایج بیشتر ردیابی در رابطه با کنترل انحراف ایجاد شده توسط شیء هم رنگ پوست در شکل ۴-۲۰ نشان داده شده است.



شکل ۴-۲۰- نتایج ردیابی دو سکانس با انحراف رنگ پوست. ردیف بالا: فریمهای ۴۹، ۴۷، ۴۵ و ۵۰. ردیف پایین: فریمهای ۲۲، ۲۱، ۱۹ و ۲۴ [۷۳]

بعضی از نتایج ردیابی یک سکانس که شامل تغییرات روشنایی است در شکل ۴-۲۱ نشان داده شده است.



شکل ۴-۲۱- نتایج ردیابی یک سکانس با تغییرات روشنایی (فریمهای ۱۷۶، ۱۰۳، ۷۷ و ۲۲۹) [۷۳]

در شکل ۴-۲۱ مشاهده می شود که الگوریتم روش تلفیقی پیشنهادی می تواند تغییر رنگهای پوست را بخاطر تغییرات روشنایی تا حدودی در ردیابی دست خنثی کند. شکل ۴-۲۲ نشان می دهد که الگوریتم تلفیقی با موفقیت حرکت سریع دست را ردیابی می کند حتی اگر محوطه های جستجو در فریمهای متوالی همپوشانی نداشته باشد.



شکل ۴-۲۲- نتایج ردیابی یک سکانس با حرکت سریع دست (فریمهای ۷۱ تا ۷۴) [۷۳].

در صورتیکه در شکل ۴-۲۲ ردیاب جابجایی میانگین هدف را بدلیل حرکت سریع دست در این موقعیت کاملاً گم می کند. آزمایشات بالا نشان می دهد که الگوریتم تلفیقی ردیابی دست را به صورت مطمئن و کارآمد انجام می دهد.

۴-۷- جمع بندی

در این فصل روش فیلتر ذره ای را به عنوان یک ردیاب دقیق و مقاوم معرفی کردیم. سپس برای کاهش زمان ردیابی آن، سعی کردیم تعداد ذره های مورد نیاز در این روش را بوسیله روش جابجایی میانگین کاهش دهیم که بتوانیم حداقل زمان ردیابی دست بوسیله این روش (فیلتر ذره ای) را نصف کنیم. این پروسه منجر به ایجاد روشی در اثر تلفیق این دو روش شد که ما آن را روش تلفیقی نامیدیم. سپس نشان دادیم که روش تلفیقی پیشنهادی از لحاظ دقت در ردیابی دست با روش فیلتر ذره ای برابری می کند و لی زمان ردیابی دست را بصورت قابل ملاحظه ای کاهش می دهد. پس از معرفی این روش خوب برای ردیابی دست، روشی را برای تشخیص اشاره دست با استفاده از این روش ردیابی بیان کردیم که در آن ما با استفاده از مکان های ردیابی شده دست بوسیله روش تلفیقی پیشنهادی میتوانیم منحنی هایی را بصورت الگوهای گذرای مبتنی بر مسیر حرکت دست (TTBT) تولید کنیم. با بهره گیری از این منحنی ها می توانیم اشارات دست را در زبان اشاره براحتی کلاسه بندی کنیم.

فصل پنجم

نتیجه گیری و پیشنهادات
برای ادامه کار

۵-۱- مقدمه

در این پایان نامه ابتدا ما ضرورت ردیابی دست را بصورت کلی بیان کردیم. سپس مبحث ردیابی اشیا را توضیح دادیم و چندین روش پیشین موفق در ردیابی شیء را معرفی کرده و به منظور آشنایی بیشتر با این مبحث بصورت عملی از معرفی ویژگی های متفاوت اشیاء مورد استفاده در ردیابی شروع کردیم. پس از معرفی ویژگی های متفاوت اشیاء، ردیابی را با یک روش سریع ردیابی شروع کردیم. انواع گوناگونی از اجسام متحرک با دوربین متحرک ارائه کرده ایم. ردیاب جابجایی میانگین را بر ویژگی های استخراج شده از هدف بالاخص رنگ شیء بنا کردیم. به عبارت دیگر توضیح دادیم که عملکرد ردیاب جابجایی میانگین وابستگی زیادی به استخراج درست ویژگی های متمایز موجود در هدف دارد. در مرحله بعد یک الگوریتم دیگر برای ردیابی دقیق و مقاوم اشیاء را به نام روش فیلتر ذره ای معرفی کردیم و سعی کردیم که با کمک روش جابجایی میانگین تعداد ذره های مورد نیاز برای ردیابی به روش فیلتر ذره ای را کاهش دهیم که نهایتاً بتوانیم ردیابی با دقت و مقاومت فیلتر ذره ای داشته باشیم که دست را در زمان کمتری (نصف زمان قبلی) ردیابی کند. این روش را که حاصل تلفیق روش جابجایی میانگین و فیلتر ذره ای بود روش تلفیقی نامیدیم و از آن را به عنوان یک ردیاب سریع و مقاوم بهره بردیم. در پایان یک روش ساده ولی کارآمد را برای تشخیص اشاره دست با استفاده از ردیاب پیشنهادی ارائه کردیم که در این روش اشارات دست را بر اساس مسیر حرکت دست بدست آمده (TTBT) از ردیابی دست مدل می کنیم.

۵-۲- سیستم کلی روش ارائه شده

در این قسمت روش تلفیقی پیشنهادی را بصورت مختصر ارائه می دهیم. در این روش پیشنهادی با بکاربردن روش جابجایی میانگین در فیلتر ذره ای، سعی میکنیم که ذره ها را به قله های محلی در سطح احتمال (رویه شباهت) حرکت دهیم، بدین ترتیب فیلتر ذره ای به ذره های کمتری برای پیدا کردن مد اصلی (دست) نیاز دارد. در نتیجه با داشتن ذره های کمتر می توانیم زمان لازم برای

ردیابی را به حداقل نصف کاهش دهیم. در روش تلفیقی از ویژگی رنگ به عنوان ویژگی متمایز دست استفاده می‌کنیم. سیستم کلی روش تلفیقی بدین صورت است که به کمک کاربر مدل دست به عنوان هدف اصلی ردیابی در فریم اول مشخص می‌کنیم. سپس از این فضای مشخص شده برای دست ویژگی رنگ را استخراج کرده و ذره‌ها را در این فضا پخش می‌کنیم. سپس با یک عملیات آستانه‌گذاری ذره‌هایی که ارزش رنگی آنها کمتر از رنگ دست است را حذف کرده و به باقی ذره‌ها متناسب با فاصله شان به مرکز دست وزن می‌دهیم (هر چه نزدیکتر، وزن بیشتر). در نهایت با استفاده از روش جابجایی میانگین هر کدام از ذره‌ها را تا جای ممکن به مرکز دست نزدیکتر می‌کنیم تا بالا ترین وزن ممکن را داشته باشد. در آخر میانگین ذره‌ها را با در نظر گرفتن وزن آنها می‌گیریم که منجر به مختصات مرکز دست می‌شود. با ورود فریم‌های جدید مکان هدف با استفاده از روش تلفیقی پیشنهادی به همین ترتیب دوباره مشخص می‌شود. پس از تخمین مکان دست با روش تلفیقی پیشنهادی، مکان دست را در هر فریم استخراج می‌کنیم.

۵-۳- جمع بندی و پیشنهادات

در این تحقیق، روش تلفیقی به عنوان یک روش سریع، دقیق و مقاوم در برابر اشیا پیش زمینه معرفی شد. بطور کلی این روش توانایی پیدا کردن یک مد مشخص را در یک رویه چند مد دارا است. ما این روش را برای خیلی از مشکلات ردیابی اشیا پیشنهاد می‌کنیم. در ادامه این کار می‌توان ویژگیهای دیگری مثل بافت دستها را برای ردیابی قوی تر و مقاوم تر استفاده کنیم. اگر چه روش تلفیقی پیشنهادی برای ردیابی دست موثر و کارآمد است، تحقیق بعدی احتمالی باید برای رسیدگی به کارآمدی این روش در ردیابی اشیا باشد همانطوریکه در ابتدای این پایان نامه کار آمدی روش جابجایی میانگین در ردیابی اشیا بررسی شد، در نتیجه در امتداد این تحقیق مسائلی از قبیل: ردیابی اشیا در فضای سه بعدی می‌تواند مورد بررسی قرار بگیرد. همچنین تا آنجایی که هدف ما تشخیص اشاره دست است، مدل‌های الگوهای گذرا مبتنی بر مسیرهای دست از طریق فقط یک زاویه دید در لحظه آموزش داده

می شوند. ما می توانیم چندین زاویه دید را در آموزش ترکیب کنیم، که بازدهی عملکرد تشخیص اشاره دست را بهبود بخشد. بطور دقیق تر پیشنهادهایی که برای بهبود الگوریتم روش تلفیقی می توان مطرح کرد به شرح زیر می باشد:

۱- در الگوریتم روش تلفیقی ارائه شده برای تخمین مکان دست نقطه شروع را برای اعمال الگوریتم جابجایی میانگین بر روی ذره ها در فریم جدید، بصورت تصادفی اطراف نزدیک مکان دست در فریم قبلی در نظر گرفتیم، حال آنکه اگر با روشی مدل حرکت دست را در هر فریم بدست آورده و نقطه شروع ذره ها را تحت این مدل جابجا کرد با این امر می توانیم تعداد تکرارها را در هر فریم کاهش داده و همچنین دقت الگوریتم ردیابی دست را بالا ببریم.

۲- پس از آنکه مرزهای دست در هر فریم بدست آمد به کمک تبدیل فاصله ذکر شده می توان پیکسلهای متعلق به دست را به طور مناسب تری وزن دهی کرده و به رویه هموارتری از معیار شباهت دست در دو فریم متوالی دست یافته و در نتیجه ردیابی دست را سریع تر انجام دهیم.

۳- معیار شباهت معرفی شده مبتنی بر هیستوگرام ویژگی های رنگی بدست آمده از دست بوده است. می توان با بهره گیری از ویژگی های دیگری به غیر از هیستوگرام رنگ، مانند ماتریس کوواریانس هدف و ارائه یک معیار شباهت جدید برای دست به بهبود تخمین مکان دست (کاهش خطای ردیابی) کمک کرد.

- [1] Y. Wu, "Vision and Learning for Intelligent Human-Computer Interaction", Ph.D. Thesis, University of Illinois at Urbana-Champaign, 2001.
- [2] V. Pavlovic, R. Sharma, T. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction", IEEE Trans. Pattern Anal. Mach. Intell., 1997.
- [3] G. McAllister, S. McKenna, I. Ricketts, "Hand Tracking for Behaviour Understanding", Image Vision Comput., 2002.
- [4] K. Oka, Y. Sato, H. Koike, "Real-Time Tracking of Multiple Fingertips and Gesture Recognition for Augmented Desk Interface Systems", Proceedings of IEEE International Conference on Automated Face and Gesture Recognition (FG'02), 2002.
- [5] I. Laptev, T. Lindeberg, "Tracking of Multi-State Hand Models Using Particle Filtering and a Hierarchy of Multi-Scale Image Features", Proceedings of IEEE Workshop on Scale-Space and Morphology, 2001.
- [6] H. Fei, I. Reid, "Probabilistic Tracking and Recognition of non-Rigid Hand Motion", Proceedings of IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG'03), 2003.
- [7] S. Lu, D. Metaxas, D. Samaras, J. Oliensis, "Using Multiple Cues for Hand Tracking and Model Refinement", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'03), 2003.
- [8] H. Zhou, T. Huang, "Tracking Articulated Hand Motion with Eigen Dynamics Analysis", Proceedings of IEEE International Conference on Computer Vision (ICCV'03), 2003.
- [9] L. Tsap, "Gesture-Tracking in Real Time with Dynamic Regional Range Computation", Real Time Imaging, 2002.
- [10] A. Shamaie, A. Sutherland, "A Dynamic Model for Real-Time Tracking of Hands in Bimanual Movements", Gesture-Based Communication in Human Computer Interaction, 2003.
- [11] E. Sudderth, M. Mandel, W. Freeman, A. Willsky, "Visual Hand Tracking Using Nonparametric Belief Propagation", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04), 2004.
- [12] M. Kolsch, M. Turk, "Fast 2d Hand Tracking with Flocks of Features and Multi-Cue Integration", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04), 2004.
- [13] Q. Yuan, S. Sclaroff, V. Athitsos, "Automatic 2d Hand Tracking in Video Sequences", Proceedings of IEEE Workshop on Applications of Computer Vision (WACV'05), 2005.
- [14] T. Inaguma, H. Saji, H. Nakatani, "Hand Motion Tracking Based on a Constraint of Three-Dimensional Continuity", J. Electron. Imaging, 2005.

- [15] W. Chang, C. Chen, Y. Hung, "Appearance-Guided Particle Filtering for Articulated Hand Tracking", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005.
- [16] S. Wu, L. Hong, "Hand Tracking in a Natural Conversational Environment by the Interacting Multiple Model and Probabilistic Data Association (IMM-PDA) Algorithm, Pattern Recognition, 2005.
- [17] B. Stenger, A. Thayananthan, P. Torr, R. Cipolla, "Model-Based Hand Tracking Using a Hierarchical Bayesian Filter", IEEE Trans. Pattern Anal. Mach. Intell., 2006.
- [18] M. Isard, A. Blake, "Condensation-Conditional Density Propagation for Visual Tracking", Int. J. Comput. Vision, 1998.
- [19] C. Ng, S. Ranganath, "Real-Time Gesture Recognition System and Application", Image Vision Comput., 2002.
- [20] M. Isard, A. Blake, "ICONDENSATION: Unifying Low-Level Tracking in a Stochastic Framework", Proceedings of European Conference on Computer Vision (ECCV'98), 1998.
- [21] M. Yang, N. Ahuja, M. Tabb, "Extraction of 2d Motion Trajectories and its Application to Hand Gesture Recognition", IEEE Trans. Pattern Anal. Mach. Intell., 2002.
- [22] "Vision Based Hand Gesture Recognition Systems", URL_ <http://ls7-www.cs.unidortmund.de/research/gesture/vbgrable.html>_.
- [23] M. Bray, E. Koller-Meier, L. Van Gool, "Smart Particle Filtering for 3d Hand Tracking", Proceedings of IEEE International Conference on Automated Face and Gesture Recognition (FG'04), 2004.
- [24] M. Arulampalam, S. Maskell, N. Gordon, T. Clapp, "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking", IEEE Trans. Signal Process., 2002.
- [25] Y. Rui, Y. Chen, "Better Proposal Distributions: Object Tracking Using the Unscented Particle Filter", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'01), 2001.
- [26] J. MacCormick, M. Isard, "Partitioned Sampling, Articulated Objects, and Interface-Quality Hand Tracking", Proceedings of European Conference on Computer Vision (ECCV'00), 2000.
- [27] T. Cham, J.M. Rehg, "A Multiple Hypothesis Approach to Figure Tracking", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'99), 1999.
- [28] J. Deutscher, A. Blake, I. Reid, "Articulated Body Motion Capture by Annealed Particle Filtering", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'00), 2000.
- [29] K. Choo, D.J. Fleet, "People Tracking Using Hybrid Monte Carlo Filtering", Proceedings of IEEE International Conference on Computer Vision (ICCV'01), 2001.

- [30] C. Chang, R. Ansari, "Kernel Particle Filter: Iterative Sampling for Efficient Visual Tracking", Proceedings of IEEE International Conference on Image Processing (ICIP'03), 2003.
- [31] B. Han, D. Comaniciu, Y. Zhu, L. Davis, "Incremental Density Approximation and kernel-based Bayesian Filtering for Object Tracking", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'04), 2004.
- [32] D. Comaniciu, V. Ramesh, P. Meer, "Real-Time Tracking of Non-Rigid Objects Using Mean Shift", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'00), 2000.
- [33] L. Wang, W. Hu, and T. Tan, "Recent Developments in Human Motion Analysis", Pattern Recognition, 2003.
- [34] D. DeCarlo and D. Metaxas, "Optical Flow Constraints on Deformable Models with Applications to Face Tracking," Int'l J. Computer Vision, 2000.
- [35] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: Real-Time Tracking of the Human Body", IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997.
- [36] S.J. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler, "Tracking Groups of People", Computer Vision and Image Understanding, 2000.
- [37] M. Kass, A. Witkin, D. Terzopoulos, "Snakes: Active Contour Models," Proceedings of the International Conference on Computer Vision, 1987.
- [38] N. Peterfreund, "Robust tracking of position and velocity with Kalman snakes," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 21, No. 6, June 1999.
- [39] G. Tsechpenakis, K. Rapantzikos, N. Tsapatsoulis and S. Kollias, "A Snake Model for Object Tracking in Natural Sequences" , Signal Processing: Image Communication, 2004.
- [40] Lili Qui, LiLi, "Contour Extraction of Moving Objects", Proceedings of 14th International Conference on Pattern Recognition, 1998.
- [41] B. Jamasbi, A. Motamedi and A. Behrad, "Tracking Vehicle Targets with Large Aspect Change", IEEE Workshop on Motion and Video Computing, 2007.
- [42] W. Hu, T. Tan, L. Wang, and S. Maybank, "A Survey on Visual Surveillance of Object Motion and Behaviors" , IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews, 2004.
- [43] C.Fu, C.Huang, and Y.Chen, "Vision-Based Preceding Vehicle Detection and Tracking", IEEE Int'l Conf. Computer Vision, 2006.
- [44] S.Avidan, "Support Vector Tracking", IEEE Trans. Pattern Analysis and Machine Intelligence, 2004.
- [45] S.Avidan, "Ensemble Tracking", IEEE Trans. Pattern Analysis and Machine Intelligence, 2007.
- [46] J.Ahmed, M.N. Jafri, and J. Ahmad, "Target Tracking in an Image Sequence Using Wavelet Features and a Neural Network", IEEE conf. TENCON, 2005.

- [47] F. Porikli, "Video Object Segmentation", Ph.D. Thesis Polytechnic University, 2001.
- [48] R. Porter, "Texture Classification and Segmentation," Ph.D thesis, University of Bristol, 1997.
- [49] L. Shapiro, G. Stockman, "3D Model & Matching", Prentice Hall, 2001.
- [50] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach toward Feature Space Analysis", IEEE Trans. Pattern Anal. Mach. Intell., 2002.
- [51] D. Comaniciu, V. Ramesh, and P. Meer, "Real-Time Tracking of Non-Rigid Objects Using Mean Shift", US patent pending, 2000.
- [52] J. Puzicha, Y. Rubner, C. Tomasi, and J. Buhmann, "Empirical Evaluation of Dissimilarity Measures for Color and Texture", Proc. Seventh Int'l Conf. Computer Vision, 1999.
- [53] T. Kailath, "The Divergence and Bhattacharyya Distance Measures in Signal Selection", IEEE Trans. Comm. Technology, 1967.
- [54] A. Djouadi, O. Snorrason, and F. Garber, "The Quality of Training- Sample Estimates of the Bhattacharyya Coefficient", IEEE Trans. Pattern Analysis and Machine Intelligence, 1990.
- [55] S. McKenna, Y. Raja, and S. Gong, "Tracking Colour Objects Using Adaptive Mixture Models", Image and Vision Comp. Journal, 1999.
- [56] Gunilla Borgefors, "Hierarchical Chamfer Matching: A Parametric Edge Matching Algorithm", IEEE Transaction on PAMI, 1988.
- [57] P. Pe'rez, J. Vermaak, A. Blake, "Data Fusion for Tracking with Particles", Proceedings of the IEEE , 2004.
- [58] E. Ozyildiz, N. Krahnstoever, R. Sharma, "Adaptive Texture and Color Segmentation for Tracking Moving Objects", Pattern Recognition ,2002.
- [59] P. Brasnett, L. Mihaylova, N. Canagarajah, D. Bull, "Particle Filtering with Multiple Cues for Object Tracking in Video Sequences", Proceedings of SPIE's 17th Annual Symposium on Electronic Imaging, Science and Technology, 2005.
- [60] G. Bradski, "Computer Vision Face Tracking for Use in a Perceptual User Interface", Intel Technol. J. Q2.
- [61] C. Yang, R. Duraiswami, L. Davis, "Efficient Mean-Shift Tracking via a New Similarity Measure", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005.
- [62] K. Nummiaro, E. Koller-Meier, L. Van Gool, "An Adaptive Color-Based Particle Filter", Image Vision Computing, 2003.

- [63] D. Comaniciu, V. Ramesh, "Mean Shift and Optimal Prediction for Efficient Object Tracking", Proceedings of IEEE Conference on Image Processing (ICIP'00), 2000.
- [64] T. Starner, J. Weaver, A. Pentland, "Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video", IEEE Trans. Pattern Anal. Mach. Intell., 1998.
- [65] P. Perez, C. Hue, J. Vermaak, M. Gangnet, "Color-Based Probabilistic Tracking, Proceedings of European Conference on Computer Vision (ECCV'02), 2002.
- [66] K. Nummiaro, E. Koller-Meier, L. Van Gool, "An Adaptive Color-Based Particle Filter", Image Vision Comput., 2003.
- [67] V. Vapnik, "The Nature of Statistical Learning Theory", Springer Verlag, 1995.
- [68] S. Geman, E. Bienenstock, and R. Doursat, "Neural Networks and the Bias/Variance Dilemma", Neural Computation, 1992.
- [69] J. Platt, "Probabilities for SV Machines, Advances in Large Margin Classifiers", MIT Press, 1999.
- [70] A. Bayesteh Tashk and P. Mowlae, "Pattern Classification Using SVM with GMM Data Selection Training Method", IEEE International Conference on Signal Processing and Communications (ICSPC 2007), 2007.
- [71] A. Blake, M. Isard, "Active Contours", Springer, 1998.
- [72] B. Han, D. Comaniciu, Y. Zhu, L. Davis, "Incremental Density Approximation and Kernel-Based Bayesian Filtering for Object Tracking", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'04), 2004.
- [73] C. Shan, T. Tanb and Y. Weib, "Real-Time Hand Tracking Using a Mean Shift Embedded Particle Filtering", Pattern Recognition Society, 2008.
- [74] M.J. Black, A.D. Jepson, "Recognition Temporal Trajectories Using the CONDENSATION Algorithm", Proceedings of IEEE International Conference on Automated Face and Gesture Recognition (FG'98), 1998.
- [75] A. Psarrou, S. Gong, M. Walter, "Recognition of Human Gestures and Behaviour Based on Motion Trajectories", Image Vision Comput., 2002.
- [76] A. Bobick, J. Davis, "The Recognition of Human Movement Using Temporal Templates", IEEE Trans. Pattern Anal. Mach. Intell., 2001.
- [77] Y. Boers, "On the Number of Samples to Be Drawn in Particle Filtering", IEE Colloquium on Target Tracking: Algorithms and Applications, 1999.
- [78] D. Koeller, R. Fratkina, "Using Learning for Approximation in Stochastic Processes", Proceedings of International Conference on Machine Learning (ICML'98), 1998.

- [79] D. Fox, W. Burgard, F. Dellaert, S. Thrun, "Monte Carlo Localization: Efficient Position Estimation for Mobile Robots", Proceedings of National Conference on Artificial Intelligence (AAAI), 1999.
- [80] D. Fox, "Kld-Sampling: Adaptive Particle Filters", Advances in Neural Information Processing Systems (NIPS), 2001.
- [81] A. Soto, "Self Adaptive Particle Filter", Proceedings of International Joint Conference on Artificial Intelligence (IJCAI'05), 2005.
- [82] C. Shan, Y. Wei, T. Tan, F. Ojardias, "Real Time Hand Tracking by Combining Particle Filtering and Mean Shift", Proceedings of IEEE International Conference on Automated Face and Gesture Recognition (FG'04), 2004.
- [83] M. Soriano, B. Martinkauppi, S. Huovinen, M. Laaksonen, "Skin Detection in Video under Changing Illumination Conditions", Proceedings of International Conference on Pattern Recognition (ICPR'00), 2000.
- [84] H.P. Graf, E. Cosatto, D. Gibbon, M. Kocheisen, E. Petajan, "Multi-Modal System for Locating Heads and Faces", Proceedings of International Conference on Automatic Face and Gesture Recognition (FG'96), 1996.
- [85] A. Dargazany, A. Soleimani, "Hand Tracking Using Kernel Density Approximator", Australian Journal of Basic and Scientific research (ISI), 2010.
- [86] M. Lotfi, A. Soleimani, A. Dargazany, H. Afzal, M. Bandarabadi, "Combining Wavelet Transforms and Neural Networks for Image Classification", 41st IEEE Southeastern Symposium on System Theory, 2008.
- [87] A. Dargazany, A. Soleimani, "A New Kernel Density Modeling", Australian Journal of Basic and Scientific research (ISI), 2010.
- [88] A. Dargazany, A. Soleimani, "Hand Tracking Using Kernel Density Approximation", IEEECS International Conference on Artificial Intelligence and Computational Intelligence, 2010.
- [89] A. Dargazany, A. Soleimani, A. Ahmadyfard, "Multibandwidth Kernel-Based Object Tracking", Hindawi Publication Service: Advances in Artificial Intelligence (AAI), 2010.

[۹۰] بهراد علیرضا، " آشکار سازی و ردیابی هدف متحرک با دوربین ویدئوی متحرک"، پایان نامه دکتری، دانشگاه صنعتی امیر کبیر، سال ۱۳۸۳.

[۹۱] جاماسبی بهزاد، " تصحیح خطاهای ناشی از تغییر نمود در ردیابی به وسیله آشکار سازی کانتور هدف"، پایان نامه کارشناسی ارشد، دانشگاه امیر کبیر، ۱۳۸۵.

[۹۲] نظری محمد، " بهبود بازشناسی واج های مصوت در زبان فارسی با استفاده از روش ترکیب SVM و GMM"، پایان نامه کارشناسی ارشد، دانشگاه صنعتی امیر کبیر، سال ۱۳۸۷.

[۹۳] وحدت علیرضا، "ردیابی اشخاص با استفاده از دوربینهای متعدد"، پایان نامه کارشناسی ارشد، دانشگاه صنعتی امیر کبیر، سال ۱۳۸۷.

فهرست مقالات ارائه شده

- [1] A. Dargazany, A. Soleimani, "Hand Tracking Using Kernel Density Approximator", Australian Journal of Basic and Scientific research (ISI), 2010.
- [2] M. Lotfi, A. Soleimani, A. Dargazany, H. Afzal, M. Bandarabadi, "Combining wavelet transforms and neural networks for image classification", 41st IEEE Southeastern Symposium on System Theory, Tullahoma March-2008, USA.
- [3] A. Dargazany, A. Soleimani, "A New Kernel Density Modeling", Australian Journal of Basic and Scientific research (ISI), 2010.
- [4] A. Dargazany, A. Soleimani, "Hand Tracking Using Kernel Density Approximation", IEEECS International Conference on Artificial Intelligence and Computational Intelligence, Shanghai Nov-2009, China.
- [5] A. Dargazany, A. Soleimani, A. Ahmadyfard, "Multibandwidth Kernel-Based Object Tracking", Hindawi Publication Service: Advances in Artificial Intelligence (AAI), 2010.

[6] درگزنی ارس، سلیمانی علی، احمدی فرد علیرضا، "ردیابی اشیا با استفاده از جابجایی میانگین در فیلتر ذره ای"، کنفرانس بین المللی کنترل اتوماسیون و ابزار دقیق، دانشگاه تربیت مدرس، سال

واژه نامه

Hidden Markov Model	مدل مخفی مارکوف
Articulated	مفصل دار
Analysis-by-synthesis	تحلیل با ترکیب
Predict-match-update	پیشگویی-تطبیق-به روزرسانی
Cross-Correlation	همبستگی متقابل
Affine Transform	تبدیل مستوی
Motion Estimation Model	تخمین مدل حرکتی
Trade-off	مصالحه
Adaboost	آبشاری
Bhattacharya measure	معیار باتاچاریا
Reference image	تصویر مرجع
Camera motion	حرکت دوربین
Partial Occlusion	پوشیدگی جزئی
Background clutter	شلوغی پس زمینه
Distance Transform	تبدیل فاصله
Approximation image	تصویر نرم شده
Mean Shift	جابجایی میانگین
Weighting Histogram	هیستوگرام وزن دار
Bhattacharyya Coefficient	ضریب باتاچاریا
Aspect Change	تغییر نمود
Target Model	مدل هدف
Target Candidate	کاندیدای هدف
Maximum Likelihood Function	تابع احتمال کلی
Convex	محدب
Monotonic decreasing	کاهشی یکنواخت
Peripheral pixels	پیکسل‌های محیطی

Kronecker Delta Function	تابع دلتا کرونکر
Multiple Features	شاخصه های چند گانه
Epanechnikov	اپانچینکف
Supervised Learning	آموزش تحت نظارت
Loss Function	تابع تلف
Multi-Layer Perceptrons	شبکه های عصبی چند لایه
Support Vector Machine	ماشین بردار پشتیبان
Radial Basis Function	شبکه های RBF
Hyper-Plane	ابر صفحه
Empirical Risk Minimization	حداقل کردن ریسک عملی
Structural Risk Minimization	حداقل کردن خطای ساختاری
Kernel	هسته
Adaptive	وفاقی
Morphological approaches	عملیات شکل شناسی

پیوست: انواع مدل‌های رنگ

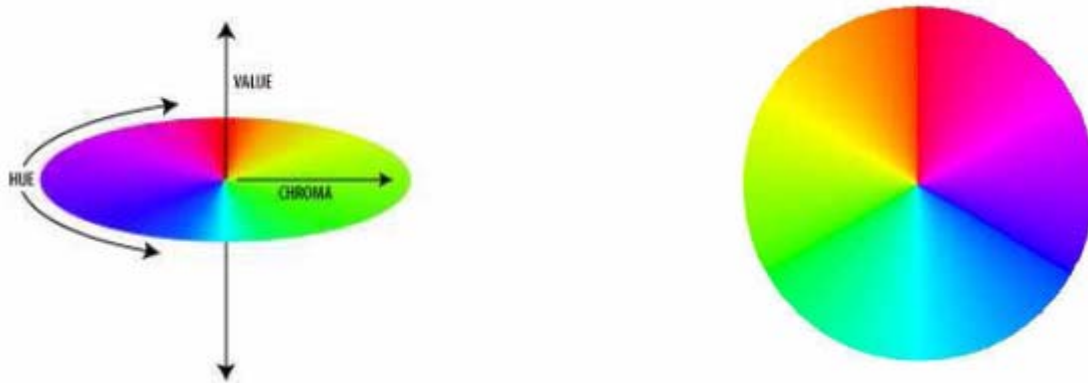
۱- مدل Huncell

یکی از مؤثرترین مدل‌های پیشنهادی اولیه توسط شخصی به نام A.Munsell، ارائه گردید. او برای هر رنگ سه مشخصه در نظر گرفت که این سه مولفه Hue، Chromine، Value نام گرفت.

Hue: مولفه‌ای که برای تعیین جهت رنگ بکار گرفته شد. در این حالت با کنار هم قرار دادن و تنظیم کردن طیف رنگها، می‌توان تغییرات همسان و همگنی در هر جهت را مشاهده کرد.

Chrominance: دومین مولفه برای تشخیص یک رنگ، که به عنوان مشخصه دیگری از رنگ در نظر گرفته شد. این مولفه در واقع معیاری برای میزان فاصله از رنگ سفید است. به بیان دیگر، درجه خالص بودن رنگ را نسبت به رنگ سفید توصیف می‌کند.

Value: این مولفه، میزان روشنایی را بیان می‌کند، میزان انرژی تابیده توسط هر رنگ، معرف دیگری از این پارامتر می‌باشد. هرچقدر میزان این تابش بیشتر باشد، رنگ روشن‌تر به نظر می‌رسد. این پارامتر از منظر Munseu، معیاری برای بیان میزان درجه روشنایی و تاریکی، نامیده شد در شکل زیر، محور Value، در واقع اشاره به مقدار Grey Level دارد:

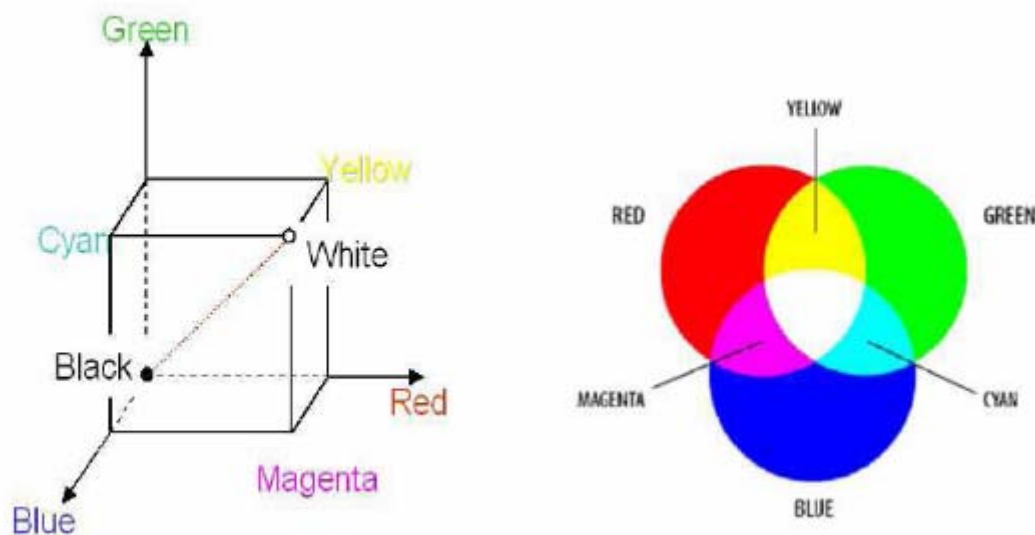


شکل پ-۱: توصیف مدل Huncell و دیاگرام Hue

۲- مدل RGB (قرمز، سبز، زرد)

در مدل RGB هر رنگ از سه مؤلفه قرمز، سبز و آبی تشکیل شده است. این مدل براساس دستگاه مختصات کارتیزین می‌باشد. فضای رنگ بصورت زیر فضای مکعبی در شکل نشان داده شده است که در آن سبز و قرمز و آبی سه گوشه و زرد و فیروزه‌ای و کهربایی در سه گوشه دیگر آن قرار دارند. سیاه در مبدأ مختصات است و سفید در دورترین گوشه از آن قرار دارد.

در این مدل سطوح خاکستری روی قطری از مکعب قرار دارند که نقطه سیاه را به سفید وصل می‌کند. رنگها با بردارهایی که مرکز را به نقاط داخل مکعب وصل می‌کنند معرفی می‌شوند. در نرم‌افزار بکار رفته برای هر پیکسل سه بایت برای سه رنگ متناظر قرمز، سبز و آبی بکار می‌رود. یعنی برای هر رنگ اصلی ۲۵۶ حالت می‌تواند وجود داشته باشد. در این مدل بین سه مؤلفه رنگ وابستگی زیادی وجود دارد و همچنین اطلاعات رنگ و روشنایی تفکیک نشده است و به همین دلیل، استفاده از این مدل رنگ برای پیاده‌سازی فیلترهای رنگ مناسب نیست.



شکل پ-۲: نمایش مدل رنگی RGB بصورت مکعبی

۳- مدل نرمالیزه شده RGB

در این مدل مؤلفه‌های مدل RGB نرمالیزه شده‌اند:

$$r = \frac{R}{R+G+B}, \quad g = \frac{G}{R+G+B}, \quad b = \frac{B}{R+G+B}$$

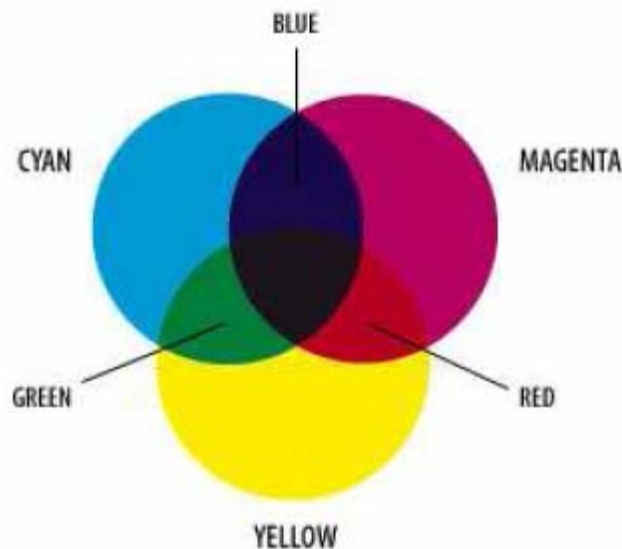
حاصل جمع این سه مؤلفه برابر یک است ($r + g + b = 1$) و شناسایی هر رنگ با تعیین دو مؤلفه از سه مؤلفه فوق صورت می‌گیرد.

۴- مدل CMY

مدلی است که از تفریق متغیرهای R, G و B بدست می‌آید. از این مدل از رنگ معمولاً در پرینترها استفاده می‌گردد و به صورت زیر توصیف می‌گردد:

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = 1 - \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

این مدل برخلاف مدل RGB، که می‌توانست ترکیبی از رنگهای متفاوتی را به وجود آورد، رنگهای روشن و درخشان را نمی‌توان تولید کند.



شکل پ-۳: نمایش فضای CMY

۵- مدل XYZ

این مدل یکی دیگر از مدل‌های ارائه شده است که با استفاده از متغیرهای R، G و B شکل می‌گیرد. مهمترین مشخصه این مدل آن است که با تنظیم درست مقادیر متغیرها، می‌توان به هر طول موجی از رنگ دست یافت. مدل ریاضی ارائه شده برای توصیف این مدل به شرح زیر است:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.618 & 0.177 & 0.205 \\ 0.299 & 0.587 & 0.116 \\ 0.000 & 0.056 & 0.996 \end{bmatrix} = \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

۶- مدل xyz

این مدل با نرمالیزه کردن مؤلفه‌های مدل XYZ بدست می‌آید. بدین صورت که هر کدام از مشخصه‌های X، Y و Z بر مجموع این سه مؤلفه، تقسیم می‌شود. مدل ریاضی زیر مبین توضیحات بالا است:

$$x = \frac{X}{X+Y+Z}, \quad y = \frac{Y}{X+Y+Z}, \quad z = \frac{Z}{X+Y+Z}$$

از آنجایی که مجموع x، y و z برابر یک می‌باشد. دو مقدار از سه مقدار بدست آمده، برای توصیف رنگ کافی می‌باشد.

۷- مدل Lab

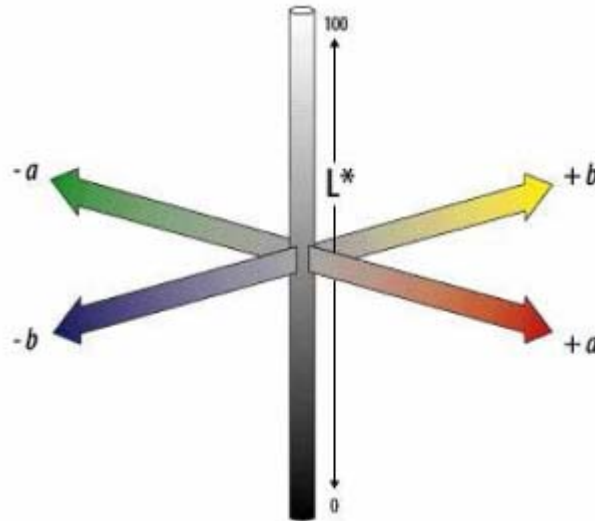
در سال ۱۹۷۶، در CIE مدلی بهبودیافته از مدل XYZ به وجود آمد که Lab نامیده شد. L، مقدار روشنایی، برای a، (مقادیر -a (سبز) و +a (قرمز)) شکل می‌گیرد و برای متغیر b، (مقادیر -b (آبی) و +b (زرد))، شکل گرفت. در این مدل، تمامی طیف مرئی پوشش داده می‌شود. در برخی از متون، نام VIQ بر این مدل گذارده شد. X_0 ، Y_0 و Z_0 به عنوان متغیرهای اولیه این مدل بصورت زیر تعریف می‌گردند:

$$X_0 = \frac{X}{\max(X)}, \quad Y_0 = \frac{Y}{\max(Y)}, \quad Z_0 = \frac{Z}{\max(Z)}$$

در آنصورت بیان ریاضی مدل Lab به صورت زیر خواهد بود:

$$L = \begin{cases} 116(Y_0)^{\frac{1}{3}} - 16 & Y_0 > 0.008856 \\ 903.3(Y_0) & Y_0 \leq 0.008856 \end{cases}$$

$$\begin{cases} a = 500 \left[(X_0)^{\frac{1}{3}} - (Y_0)^{\frac{1}{3}} \right] \\ b = 200 \left[(Y_0)^{\frac{1}{3}} - (Z_0)^{\frac{1}{3}} \right] \end{cases}$$



شکل پ-۴: توصیف فضای CIE_Lab

۸- مدل YUV

این مدل از رنگ، به عنوان یک مدل پایه در تصاویر استاندارد ویدیویی رنگی مورد استفاده قرار می‌گیرد. مدل ریاضی ارائه شده آن در زیر بیان شده است. متغیرهای R، G و B به عنوان مقادیر اولیه آن لحاظ می‌گردند.

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.500 \\ 0.500 & -0.419 & -0.081 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

نکته قابل توجه در این مدل اینکه، مدل VCBcR که به عنوان فیلتر رنگ در موارد زیادی کاربرد دارد، در واقع نسخه‌ای از مدل YUV می‌باشد، در مدل YUV نیز مقدار Y از مقادیر U و V جدا گردیده است.

۹- مدل HSI

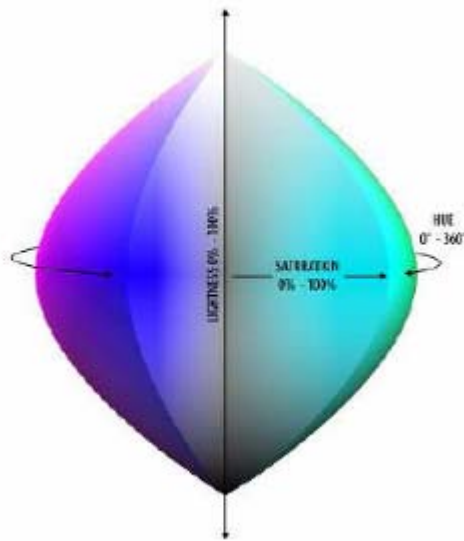
با تغییر محورهای مختصات YUV به مختصات قطبی، این مدل به وجود خواهد آمد. جهت (hue) به پارامتر زاویه (H) و میزان اشباع بودن رنگ به مؤلفه شعاع (S) تبدیل می‌گردد مؤلفه میزان شدت روشنایی بر محور uv عمود است. مقدار H بین 0° تا 360° را می‌تواند شامل گردد و زیر 60° جهت جدیدی تعریف می‌گردد. پارامتر S نیز مقداری بین 0 تا 1 را شامل می‌شود، و مبین اندازه پارامتر اشباع به ماکزیمم مقدارش است. معادلات ریاضی توصیفی زیر، برای توصیف این مدل بکار می‌رود.

$$H = \text{Arc tan}\left(\frac{\sqrt{3}(G - B)}{2R - G - B}\right)$$

$$S = 1 - 3 \min(r, g, b)$$

$$I = \frac{R + G + B}{3}$$

چگونگی نسبت‌دهی مقادیر به رنگ‌های مختلف در فضای HSI در شکل زیر بصورت یک نمودار مخروطی آمده است.



شکل پ-۵: نمایش مدل HSI

مشکل مدل HSI، پارامتر H است، این پارامتر می‌تواند وقتی که S به صفر نزدیک شود، به حالت ناپایدار درآید.

۱۰- مدل YCbCr

این مدل بیانی دیگر از مدل RGB است که در فشرده‌سازی تصاویر بکار می‌رود. انتقال رنگ از فضای RGB به این فضا به سادگی و با استفاده از رابطه زیر صورت می‌گیرد که در آن مولفه‌های رنگ و روشنایی جدا شده‌اند. این مدل در بسیاری از سیستم‌های بینایی ماشین برای شناسایی هدف در تصویر بکار رفته است.

$$Y = 0/299R + 0/587G + 0/114B$$

$$Cb = B - Y$$

$$Cr = R - Y$$

از مزایای مدل رنگی YCbCr این است که این مدل در تست‌ها و در شرایط مختلف نتایج بهتری را در فیلتر کردن رنگ آبی نشان داده است.

Abstract

In this contribution, a new framework for effective tracking of moving objects is introduced. In order to apply this object tracking method to hand gesture recognition in sign language, more importance is given to hand tracking.

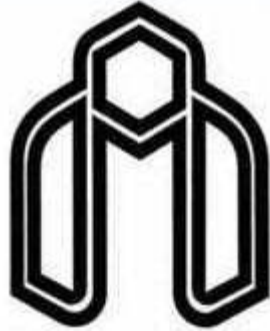
Through developing a two-step approach of Mean-Shift algorithm and corresponding extracted features namely colors, reasonable detection ability is obtained which yields desired fast tracking results feasible for real-time applications. Particle filtering is also another well-known approach in object tracking which is in fact more robust than Mean-Shift tracker but much slower.

Both particle filtering (PF) and mean shift (MS) are two successful approaches to visual tracking. Both have their respective strengths and weaknesses. Mean Shift is a fast tracker and Particle filter is a robust tracker. In this Study, it is proposed to integrate advantages of these two algorithms into one procedure in order to improve hand tracking results by incorporating the MS optimization into particle filtering to move particles to local peaks in the likelihood, so that the proposed method improves the sampling efficiency considerably.

The experimental results demonstrate that the proposed method, called CMP (Combination of Mean shift and Particle filter), can track hands two time faster than PF. CMP also tracks hand as robust as PF using just 20% number of total required particles in PF. Conclusively, Extensive experimental results illustrate that CMP outperforms MS and PF in hand tracking.

Having proposed a capable hand tracker (CMP), a simple procedure for extracting hand trajectory from a sequence of images in the utilized sign language database lead us to hand gesture recognition reasonably and acceptably.

Keyword: Hand tracking; Particle filter; Mean shift; Combination of Mean shift and Particle filter; Hand gesture recognition; sign language; Human–computer interaction.



Shahrood University of Technology

DEPARTMENT: Department of Electrical and Robotics
Eng., Shahrood University of Technology

Master Thesis

TITLE:

Hand Tracking Using Combination of Mean Shift and
Particle Filter

AUTHOR: Aras Dargazany

SUPERVISOR: Dr. Ali Soleimani, Ph.D.

ADVISOR: Dr. Alireza Ahmadyfard, Ph.D.

THE YEAR OF PRESENTATION: APRIL 2010