

تشخیص و ردیابی اشیاء ثابت و متحرک در تصاویر زنده و ضبط شده

تاریخ دریافت: ۱۴۰۱/۰۴/۲۰

تاریخ پذیرش: ۱۴۰۱/۰۵/۲۴

کد مقاله: ۳۵۸۹۴

نیما دلبینا^{۱*}، غزال ملایی^۲

چکیده

تشخیص و ردیابی اشیاء یک وظیفه مهم در نظارت تصویری است؛ بنابراین، در پروژه حاضر، با استفاده از تشخیص و ردیابی توسط دو الگوریتم YOLO و DEEP SORT که به تفصیل مورد بررسی قرار گرفته‌اند، قصد داریم به سرعت در زمان واقعی شناسایی و ردیابی کنیم. در مرحله اول، الگوریتم خوشه‌بندی Kmeans برای بهینه‌سازی مقدار اولیه استفاده می‌شود. بودن در سال‌های گذشته، الگوریتم‌های SORT و YOLO در پروژه‌هایی استفاده می‌شد که سرعت و دقت پایینی داشتند. در این پروژه از DEEP SORT و YOLO-tiny استفاده شده است که چندین برابر سریع‌تر از مدل‌های قبلی خود است و از قابلیت تشخیص و ردیابی و همچنین ردیابی اجسام در زمان واقعی در سال‌های اخیر، DEEP SORT نشان داده است که در مواجهه با نمونه‌های پیچیده بسیار سریع‌تر از الگوریتم‌های MIL، KCF و MOSSE است.

واژگان کلیدی: ردیابی اشیاء، ثابت، متحرک، DEEP SORT، YOLO-tiny

۱- کارشناسی ارشد رشته مهندسی برق (مخابرات-سیستم)، موسسه آموزش عالی علم و فن ارومیه (نویسنده مسئول)
nimadelbina2022@gmail.com

۲- کارشناسی ارشد، رشته کامپیوتر، گرایش نرم افزار، دانشگاه آزاد اسلامی واحد سنندج

۱- مقدمه

ردیابی شی کار مهمی در حوزه بینایی ماشین محسوب می شود. گسترش روز افزون کامپیوترهای قوی، کیفیت بالا و در دسترس بودن دوربین های فیلم برداری ارزان قیمت و نیاز روزافزون برای تجزیه و تحلیل خودکار فیلمها باعث افزایش کاربرد ردیابی اشیاء گشته است. سه گام کلیدی در تجزیه و تحلیل فیلم های ویدیویی وجود دارد (چن و همکاران، ۲۰۱۸):

- شناسایی شی متحرک مورد نظر
 - ردیابی اشیاء فریم به فریم
 - تجزیه و تحلیل حرکت شی برای تشخیص رفتار آن
- ردیابی می تواند شامل تعریف مسیر و حرکت یک شی در آن مسیر باشد. به عبارت دیگر ردیابی یعنی برچسب زدن به شی مورد نظر در فریم های مختلف تصویر، علاوه بر این، ردیابی می تواند اطلاعات دیگری را نیز مانند جهت گیری، مساحت یا شکل جسم در اختیار قرار دهد. ردیابی اشیاء می تواند به دلایل زیر پیچیده باشد:

۱. از دست رفتن بخشی از اطلاعات دنیای واقعی سه بعدی
۲. نویز در تصاویر
۳. حرکات پیچیده شی
۴. ماهیت غیر شبکه ای یا بند بند بودن شی
۵. شکل پیچیده شی
۶. ناقص یا مسدود بودن جسم
۷. تغییرات در نور صحنه
۸. احتیاج به پردازش بلادرنگ

منظور از ردیابی شی، ایجاد مسیری از جابجایی شی در طول زمان در فریم های متوالی است و ممکن است ردیاب در هر لحظه از زمان، منطقه کمی از تصویر را زیر نظر داشته باشد (دلال و همکاران، ۲۰۰۵). تشخیص شی و برقراری ارتباط میان شی نمونه در فریم های مختلف ممکن است به صورت جداگانه با مشترک صورت پذیرد. در حالت جداگانه، محدوده شی در هر فریم به کمک الگوریتم تشخیص شی پیدا و سپس ردیاب در سراسر فریم ها، شی مربوطه را ردیابی می کند (همان). در حالت دوم (مشترک) با استفاده از اطلاعات بدست آمده از فریم های قبلی و به روزرسانی اطلاعات مرتبط باشد در هر لحظه، در هر فریم شی در منطقه جدید تشخیص و ردیابی می شود. در هر دو روش، ردیابی اشیاء با استفاده از شکل یا ظاهر تشخیص داده می شود. سپس براساس شکل جسم و بسته به نوع حرکت و یا تغییر شکل آن، مدلی در نظر گرفته می شود. مثلاً اگر شی به شکل یک نقطه نشان داده شود تنها می توان از مدل انتقالی برای آن استفاده کرد یا اگر از شکل هندسی مانند بیضی برای نمایش استفاده شود از مدل حرکتی affine یا تغییرات تصویری مناسب می توان بهره برد.

تخمین حرکت، قسمت مهمی از هر سیستم پردازش ویدئویی است. در این فصل، تمرکز روی تخمین حرکت دو بعدی است. تخمین حرکت دو بعدی، علاوه بر این که دامنه وسیعی از کاربردها مثل تغییر نرخ نمونه برداری، فیلتر کردن، فشرده سازی ویدئویی و غیره را شامل می شود، اغلب مرحله ای پیش پردازش و لازم در تخمین حرکت و نیز تخمین ساختار سه بعدی محسوب می شود. بسته به نوع کاربرد روش های تخمین حرکت می توانند بسیار متنوع باشند برای مثال، برای کاربردهای بینایی ماشین که در آن از بردارهای حرکت دو بعدی برای استخراج ساختار و پارامترهای حرکت سه بعدی استفاده شود، مجموعه ای محدود و پراکنده از بردارهای حرکت دو بعدی در نقاط مشخص کفایت می کند. از طرف دیگر، برای کاربردهای فشرده سازی ویدئویی، بردارهای حرکت تخمین زده می شوند و برای پیش گویی حرکت در یک فریم که باید براساس فریم مرجع گذشته قبلی، کد شود به کار می روند. هدف نهایی، به حداقل رسانیدن تعداد بیت های استفاده شده برای کد کردن بردارهای حرکت و پیش بینی خطاها است. چالشی بین دقت تخمین زده شده و تعداد بیت های استفاده شده برای مشخص کردن مشخصات حرکت وجود دارد (لو و همکاران، ۲۰۱۳). حتی ممکن است هنگامی که حرکت تخمین زده شده بیانی دقیق از حرکت واقعی فیزیکی نباشد، منجر به پیش بینی مطلوبی از حرکت شود لذا همچنان یک تخمین خوب در نظر گرفته می شود (همان).

در سال های اخیر، با پیشرفت سریع بینایی رایانه ای و هوش مصنوعی، سیستم های نظارت تصویری هوشمند به تدریج در اشکال مختلف به کار گرفته شده است. محققان بسیاری از روشهای تشخیص سنتی را پیشنهاد کرده اند. این الگوریتم ها دارای سه فرایند مشترک از جمله انتخاب ناحیه، استخراج ویژگی ها و طبقه بندی هستند. انتخاب منطقه به طور کلی از یک پنجره کشویی برای پیمایش استفاده می کند. استخراج ویژگی نقش مهمی در تشخیص شی دارد. پس از انتخاب منطقه، ویژگی ها از هدف استخراج می شوند. بنابراین، ویژگیهای کلاسیک مانند ویژگیهای تغییر مقیاس (SIFT) و هیستوگرامهای شیب که به HOG نیاز دارند باید نشان دادن هدف طراحی شوند. علاوه بر این، استفاده از شبکه های عصبی پیچشی برگشت پذیر برای تشخیص اجسام در این پروژه به عنوان الگوریتم اساسی در این فرایند استفاده شده است.

۲- ردیابی شی

هدف مسیریابی با ردیابی، نشان دادن این موضوع است که شی x دریافت شده در تصویر در زمان o در مکان (x, y) ، همان شیء مشابه y در تصاویر یافته شده در زمان $o+1$ در مکان (x', y', z') است. روش ها و دستاوردهای مختلفی برای ردیابی شیء پیشنهاد شده است که در فصول قبل توضیح دادیم. در این کار، ما الگوریتم مسیریابی را بر اساس مسیریابی حباب پیشنهاد

کرده ایم. امتیاز اصلی این الگوریتم سریع بودن آن است. برای هر شی حاصل از تفریق پس زمینه در سکانس ویدئو که حاصل شد یک حباب در نظر می گیریم. اگر در یک فریم ویدئو حبابی جدید ایجاد شد، مرکز آن حباب را ردیابی می کنیم. از این حباب برای تجزیه و تحلیل حرکت استفاده و سپس تصویر را به فریم اصلی، مطابق شکل زیر که شامل شی متحرک است، اضافه می کنیم. هر حباب نشان دهنده یک شی متحرک است. برای این الگوریتم بیشتر از دو فریم نیاز است (واپنیک و همکاران، ۱۹۹۸).



شکل ۱- اضافه کردن حباب به تصویر اصلی
(مآخذ: واپنیک و همکاران، ۱۹۹۸)

در این جا از چگونگی طبقه بندی شدن هر وسیله در هر فریم، تعداد اشیای متحرک فعال در هر زمان، سرعت شی در هر زمان، میانگین سرعت شی در طی مدت زمانی که در دید دوربین (آن) منطقه بوده است. سایقه های نگهداری می شود. سرعت شی با استفاده از اطلاعات موجود از محل های ثبت شده از آن به دست می آید. اگر شی X در مکان (X,Y,Z) در هماهنگی های جهانی سه بعدی در تصویر زمانی مشاهده شود و اگر همان شی X که متعلق به مسیریابی مشابهی است (با فرض اینکه مسیریابی موفق آمیز باشد) در مکان (x,y'z') در مسیریابی روش زمانی (t+T) مورد مشاهده قرار گیرد، سرعت آنی به صورت فرمول زیر ارائه می گردد.

$$\sqrt{(x - x_T)^2 + (y - y_T)^2 + (z - z_T)^2} \quad (1)$$

ممکن است سرعت شی با استفاده از اطلاعات مکانی آن در زمان در ردیابی پیدا شود. حال برای سرعت متوسط مجموع جابجایی ها را به کار می بریم و در پایان به مجموع فریم های شامل جابجایی تقسیم می شود.

معادله ردیابی: معادله ردیابی یکجا اصولی ترین معادله برای شناختن اغلب سامانه های ردیابی است.

$$Pr = \frac{PtGtA\sigma}{(4\pi)^2Rt^4} \quad (1)$$

۲-۱- تقسیم بندی روش ها بر اساس چگونگی تخمین خود حرکتی و تشخیص شی متحرک

همان طور که در بخش فیل دیدیم نمی توانیم روش های تشخیص شی متحرک که برای دوربین های ثابت توسعه داده شده اند را به طور مستقیم در کاربردهای دوربین متحرک مورد استفاده قرار دهیم. از این رو انواع تغییر یافته الگوریتم های تفریق پس زمینه و تفاضل زمانی و شار نوری توسعه داده شده اند که در بخش قبل نیز اشاراتی به این روش ها شد. کیم و همکارانش این روش ها را بر اساس رویکرد آنها برای انجام عمل تطبیق در چهار گروه دسته بندی کرده است، که عبارتند از: رویکرد موزائیک پس زمینه، رویکرد شار نوری، رویکرد تطبیق ویژگی و رویکرد مدل هندسی دوربین

۲-۲ فیلتر کالمن

فیلتر کالمن (Kalman Filter) یک تخمین گر است که از تخمین حالت قبل و مشاهده فعلی برای محاسبه تخمین حالت فعلی استفاده می کند و یک ابزار بسیار قوی برای ترکیب اطلاعات در حضور نامعینی ها است. در برخی موارد، توانایی فیلتر کالمن برای استخراج اطلاعات دقیق خیره کننده است. فیلتر کالمن مدت هاست که به عنوان راه حل بهینه برای بسیاری از کارهای ردیابی و پیش بینی داده ها مورد استفاده قرار می گیرد.

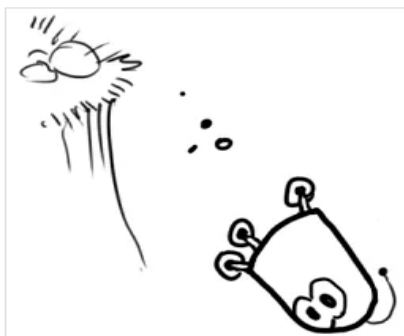


شکل ۲- کارایی فیلتر کالمن (مآخذ، پژوهشگر)

فرض کنید ربات کوچکی ساخته ایم که می تواند آزادانه در یک جنگل حرکت کند. این ربات برای آنکه مسیریابی کند، باید بداند دقیقاً در کجا قرار گرفته است. وضعیت یا حالت (State) ربات را با x_k نشان می دهیم که شامل موقعیت (Position) و سرعت (Velocity) است:

$$x_k = (p, v) \rightarrow x_{k+1} = (p, v)$$

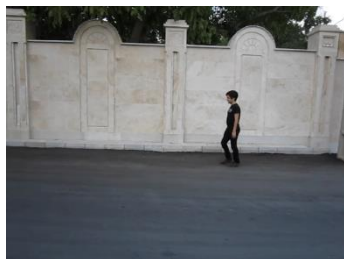
حالت یا بردار حالت، شامل اعدادی است که پیکربندی سیستم را نشان می دهند. در مورد این ربات، حالت شامل موقعیت و سرعت است. برای مثال، حالت یک سیستم می تواند اطلاعاتی درباره مقدار سیال یک مخزن، دمای موتور خودرو، موقعیت انگشت کاربر روی صفحه لمسی یا هر مورد دیگری باشد که ما بسته به نیاز، آن را در نظر می گیریم.



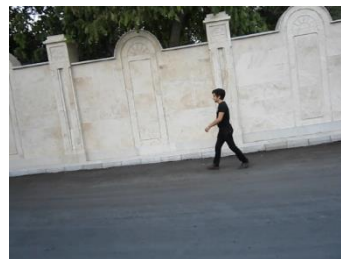
شکل ۳- موقعیت یابی با فیلتر کالمن (مأخذ، پژوهشگر)

رباتی که درباره آن بحث کردیم، یک سنسور GPS با دقت تقریباً ۱۰ متر دارد که مقدار تقریباً مناسبی است. اما لازم است موقعیت ربات به صورت دقیق‌تر و کمتر از ۱۰ متر نیز در اختیار باشد، زیرا در جنگل صخره‌ها و گودال‌هایی وجود دارد و اگر ربات چند قدم اشتباه گام بردارد، ممکن است مثلاً درون گودالی بیفتد و دچار مشکل شود. بنابراین، وجود GPS به تنهایی برای مسیریابی ربات کافی نیست. ممکن است چیزهای دیگری نیز درباره ربات بدانیم. مثلاً می‌دانیم فرمان‌هایی به موتور چرخ‌ها ارسال می‌شود و همچنین، اگر ربات در جهت خاصی حرکت کند و مشکلی برایش پیش نیاید، همان جهت حرکت قبلی را ادامه خواهد داد. اما اطلاعاتی درباره حرکت ربات در دست نیست.

بنابراین، تعداد دور چرخ‌های ربات، نمی‌تواند دقیقاً تعیین کند که چه اندازه حرکت کرده است و در نتیجه، پیش‌بینی کامل نخواهد بود. سنسور GPS اطلاعاتی را درباره حالت ربات به ما می‌دهد که البته غیرمستقیم بوده و صطلاحاً نامعینی‌هایی دارد. پیش‌بینی ما نیز، اطلاعاتی را درباره چگونگی حرکت ربات به دست می‌دهد که غیرمستقیم بوده و با نایقینی یا نامعینی همراه است.

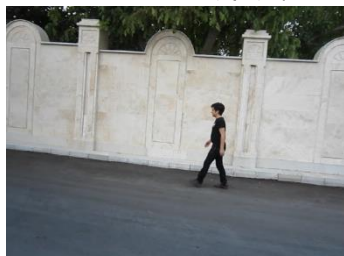


فریم شماره ۴

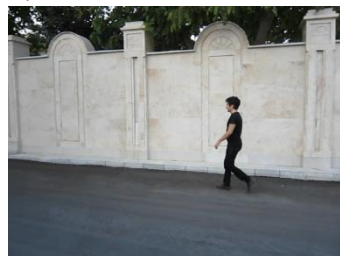


فریم شماره ۵۶

شکل ۳- مقدار خطا با معیار MSE در دنباله فریم دوم



فریم شماره ۷۱

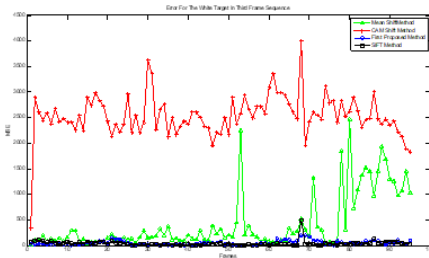


فریم شماره ۹۴

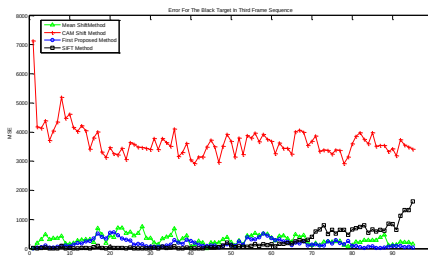
شکل ۴- فریم‌های نمونه از دنباله فریم دوم

همانطور که در جدول ۳ مشاهده می‌شود، ضعیف‌ترین عملکرد در این دنباله مربوط به اولین روش پیشنهادی است. دلیل این امر مشابه دلیلی است که در رابطه با سکانس فریم قبلی ذکر شد. مقایسه فریم‌های ۵۶ و ۷۱ به خوبی تغییر زاویه دوربین را در طول این دنباله فریم نشان می‌دهد. با توجه به اینکه ابعاد تصویر تغییر چندانی نکرد، الگوریتم Mean Shift نیز توانست هدف را به خوبی ردیابی کند. الگوریتم CAM Shift به همان دلیلی که در دنباله قبلی ذکر شد عملکرد نسبتاً ضعیفی داشت. مشابه دنباله قبلی، بهترین عملکرد در این دنباله مربوط به دومین الگوریتم پیشنهادی است. این نشان می‌دهد که قدرت این الگوریتم در ردیابی تصاویر بدست آمده از یک دوربین متحرک چقدر است.

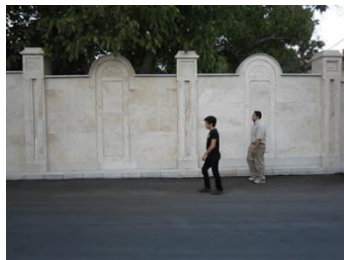
دنباله فریم سوم: در این دنباله فریم، از دو انسان به عنوان هدف استفاده می‌شود. هر دو هدف در یک جهت حرکت می‌کنند. علاوه بر این، به دلیل سرعت اهداف، پدیده تار شدن یا قرار دادن اهداف در یک موقعیت مشابه رخ نمی‌دهد. در نتیجه، ردیابی تقریباً همان ردیابی یک هدف است. تفاوت این است که در اینجا دو مجموعه سیستم ردیابی باید به طور همزمان حرکت اهداف را دنبال کنند. این دنباله شامل ۹۵ فریم است.



شکل ۶- مقدار خطا هدف شماره ۲ با معیار MSE در دنباله فریم سوم



شکل ۵- مقدار خطا هدف شماره ۱ با معیار MSE در دنباله فریم سوم



فریم شماره ۱۰



فریم شماره ۵۰



فریم شماره ۸۳

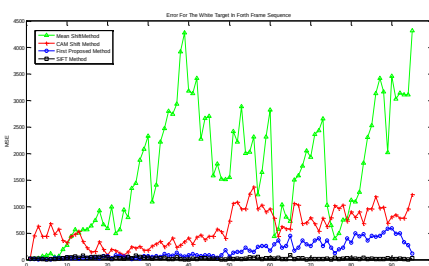


فریم شماره ۹۵

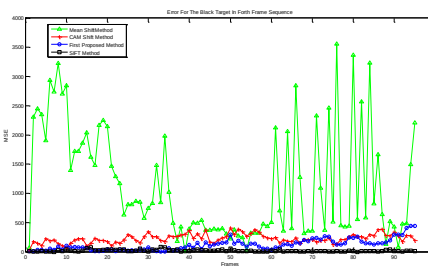
شکل ۷- فریم های نمونه از دنباله فریم سوم

با توجه به حرکت دوربین و همچنین نحوه حرکت اهداف، می توان دریافت که بهترین عملکرد در بین روشهای موجود مربوط به الگوریتم های پیشنهادی اول و دوم است. روش CAM Shift به دلایلی که در قسمت های قبل ذکر شد، کارایی ضعیفی دارد. علاوه بر این، همانطور که قبلاً بحث کردیم، روش Shift میانگین کاملاً بر اساس فرکانس رنگ است. در نتیجه، خطایی در ردیابی هدف ۲ وجود دارد که دارای رنگ روشن مشابه پس زمینه است. این در شکل ۵-۶ مشهود است. همچنین، طبق شکل ۵-۵، این خطا در ردیابی هدف ۱، که تاریک تر است، دیده نمی شود.

دنباله فریم چهارم: در این دنباله فریم، از دو خودرو به عنوان هدف استفاده می شود. اهداف در فریم های قبلی انسان بودند. به همین دلیل، آنها در طول فریم های مختلف تغییر شکل زیادی داشتند. در نتیجه، در این دنباله فریم، با توجه به ساختار سفت و سخت اهداف، الگوریتم های پیشنهادی باید عملکرد بسیار خوبی داشته باشند که با نگاه به جدول، این امر تأیید می شود. حرکات اهداف در این دنباله نیز به عنوان دنباله فریم سوم در نظر گرفته می شود. به طور مشابه، تعداد فریم های این دنباله ۹۵ است.



شکل ۹- مقدار خطا هدف شماره ۱ با معیار MSE در دنباله فریم چهارم



شکل ۸- مقدار خطا هدف شماره ۱ با معیار MSE در دنباله فریم چهارم



فریم شماره ۱۰



فریم شماره ۲۱

شکل ۹- مقدار خطا هدف شماره ۲ با معیار MSE در دنباله فریم چهارم



فریم شماره ۵۳

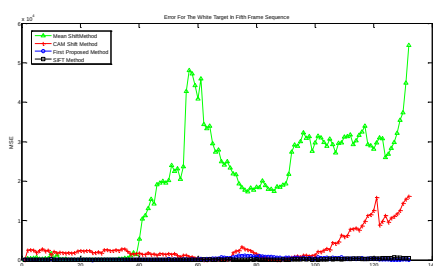


فریم شماره ۹۵

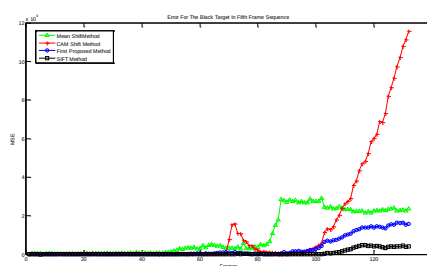
شکل ۱۰- فریم های نمونه از دنباله فریم چهارم

با توجه به نمودارها و جداول این مرحله، مشاهده می شود که الگوریتم های پیشنهادی اول و دوم بهترین عملکرد را دارند. الگوریتم CAM Shift به دلیل افزایش تنوع در فضای رنگ HSV در این دنباله فریم بهتر از قبل عمل کرده است. الگوریتم Mean Shift عملکرد نسبتاً ضعیفی در این زمینه دارد که دلیل اصلی آن را می توان به تغییر در فضای رنگ محیط نسبت داد. وی این موضوع را با توجه به مقدار خطای این الگوریتم برای دو منظور متفاوت مورد بررسی قرار داد. تفاوت بین رنگ هدف تیره تر و رنگ محیط بسیار بیشتر از هدف روشن تر است. به عبارت دیگر، رنگهای پس زمینه این دنباله فریم به طور کلی روشن هستند.

دنباله فریم پنجم: در این سکانس فریم، از دو خودرو به عنوان هدف استفاده شده است. اتومبیل ها در جهت مخالف حرکت می کنند. بنابراین، در طول توالی فریم ها، هم با پدیده محو شدن و هم قرار دادن اهداف در موقعیت مشابه روبرو هستیم. پدیده محو شدن از قاب ۵۴ شروع می شود. این قاب نقطه شروع همپوشانی دو هدف است. با توجه به این که اهداف در جهات مختلف حرکت می کنند، نقش فیلتر کالمن در این توالی بسیار مهم است. به این ترتیب، با پیش بینی صحیح موقعیت هدف در قاب آینده، از خطاهای سیستم ردیابی جلوگیری می شود. نتایج شبیه سازی در شکل ها و جدول زیر نشان داده شده است. همچنین لازم به ذکر است که این دنباله فریم شامل ۱۳۲ فریم است.



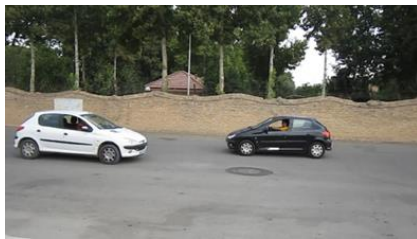
شکل ۱۱- مقدار خطا هدف شماره ۲ با معیار MSE در دنباله فریم پنجم



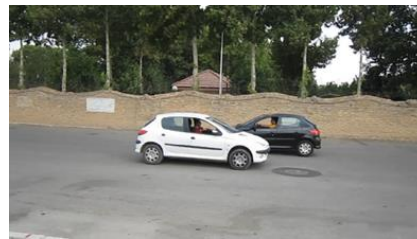
شکل ۱۱- مقدار خطا هدف شماره ۱ با معیار MSE در دنباله فریم پنجم

با توجه به جداول و نمودارهای بدست آمده، مشاهده می شود که روشهای پیشنهادی در ردیابی اهداف در این چارچوب بسیار موفقتر از سایر روشها بوده است. با توجه به ترتیب فریم ها، می توان دریافت که پدیده محو شدن از قاب شماره ۵۴ شروع می شود. در قاب ۱۰۴، هدف ۲ تقریباً به طور کامل هدف ۱ را پوشش می دهد. این بدان معناست که برای چندین فریم متوالی، تقریباً هیچ اثری از هدف ۱ دیده نمی شود به روشهای پیشنهادی با استفاده از پیش بینی موقعیت مورد نظر در قابهای آینده، تا حدودی مشکل را حل کرده و حتی در شرایط ناپدید شدن کامل، ردیابی را ادامه می دهد. ناحیه مورد نظر در الگوریتم Shift میانگین به سمت مرکز جرم پیکسل های منطقه حرکت می کند. بنابراین، در صورت محو شدن، منطقه مورد نظر نیز تغییر می کند. در صورت افزایش میزان تاری، ناحیه هدف در این روش بسیار تغییر می کند و حتی گاهی اوقات ممکن است به اشتباه یک هدف دیگر را ردیابی کند. همچنین، الگوریتم CAM Shift در ردیابی اهداف در این دنباله فریم موفق نبوده است. این شکست همچنین می تواند ناشی از افزایش محو شدن باشد. با توجه به شکل ۵-۱۱، می توان فهمید که عملکرد این روش حتی در صورت تار شدن

جزئی قابل قبول بوده است. اما با افزایش این تاری، الگوریتم توانایی سازگاری خود را از دست می دهد و سیستم ردیابی دچار خطاهای زیادی می شود.



فریم شماره ۱۴



فریم شماره ۷۸



فریم شماره ۱۰۲



فریم شماره ۱۳۱

شکل ۱۳- فریم های نمونه از دنباله فریم پنجم

نتیجه گیری

با توجه به اطلاعات این پایان نامه، مشاهده می شود که حرکت دوربین تأثیر بسزایی در کارایی روش های ردیابی دارد. این امر با بررسی نمودارهای خطا در فصل قبل به وضوح قابل درک است. با توجه به اینکه تاکنون هیچ روش خاصی برای ردیابی در شرایط دوربین تلفن همراه ایجاد نشده است، در این پایان نامه روش هایی را برای دستیابی به این هدف پیشنهاد کرده ایم. علاوه بر این، ما عملکرد آنها را با استفاده از توالی فریم های مختلف ارزیابی کردیم. با توجه به داده های به دست آمده، می توان نتیجه گرفت که در وهله اول، هر دو روش پیشنهادی قادر به ردیابی در شرایط دوربین های متحرک و ثابت هستند. اولین روش پیشنهادی مقاومت بالایی در برابر چرخش ندارد. به عبارت دیگر، اگر دوربین و هدف در جهت هریک از سه محور نسبت به یکدیگر بچرخند، عملکرد این روش به شدت تحت تأثیر قرار می گیرد. نتایج به دست آمده از ردیابی دنباله فریم اول و دوم به خوبی این را نشان می دهد. در شرایط دیگر که حرکت دوربین و هدف نسبت به یکدیگر گذرا است، این روش قابلیت ردیابی با دقت نسبتاً بالایی را دارد. همچنین استفاده از فیلتر کلمن و پیش بینی موقعیت، کارایی این روش را بسیار افزایش داده است. دومین الگوریتم پیشنهادی عمدتاً بر اساس الگوریتم تطبیق نقطه SIFT است. با مشاهده توالی فریم ها می توان دریافت که این الگوریتم قابلیت ردیابی بسیار دقیق در موقعیت های مختلف را دارد. همانطور که در توضیح این روش اشاره کردیم، مقاومت زیادی در برابر چرخش و تغییر اندازه هدف دارد.

به منظور حل مشکل تشخیص شیء، تشخیص و ردیابی شیء بر اساس تشخیص بهتر YOLO و الگوریتم های مرتب سازی عمیق برای تشخیص و ردیابی پیشنهاد شد. بهبود الگوریتم تشخیص YOLO و سپس استفاده از DEEPSORT آن را بسیار سریع و در زمان واقعی انجام داد. اول، بر اساس ویژگی های شکل کشتی، از الگوریتم K-means برای بهینه سازی مقدار اولیه استفاده شد، که زمان آموزش را کوتاه کرد. از الگوریتم ردیابی عمیق برای ردیابی اشیاء استفاده شد. مرتب سازی عمیق از الگوریتم تشخیص YOLO برای ردیابی اشیاء با مقایسه تجربیات مرتب سازی عمیق پیشرفته استفاده می کند و سایر الگوریتم های جستجوی کلاسیک مانند KCF، MIL، MOSSE، Median، TLD در چندین مورد تجزیه و تحلیل قرار گرفت. بر اساس تجزیه و تحلیل، ما می بینیم که الگوریتم مرتب سازی عمیق مزایای عملکرد بیشتری در صحنه های پیچیده ارائه می دهد و در برابر تداخل مانند انسداد و حرکت دوربین مقاوم است. بنابراین می توان از آن در عملیات زمان واقعی استفاده کرد.

1. Chen, J.; Chen, D.; Meng, S. A Novel Region Selection Algorithm for Auto-Focusing Method Based on Depth from Focus. In Proceedings of the Fourth Euro-China Conference on Intelligent Data Analysis and Applications, Cham, Switzerland, 1 October 2018; pp. 101–108.
2. Tiwari, A.; Goswami, A.K.; Saraswat, M. Feature Extraction for Object Recognition and Image Classification. *Int. J. Eng. Res.* 2013, 2, 9.
3. Du, C.-J.; He, H.-J.; Sun, D.-W. Chapter Object Classification Methods. In *Computer Vision Technology for Food Quality Evaluation*, 2nd ed.; Sun, D.-W., Ed.; Academic Press: San Diego, CA, USA, 2016; pp. 87–110, ISBN 978-0-12-802232-0.
4. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* 2004, 60, 91–110. [CrossRef]
5. Dalal, N.; Triggs, B. Histograms of Oriented Gradients for Human Detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 886–893.
6. Dalal, N.; Triggs, B. Histograms of Oriented Gradients for Human Detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 886–893.
7. Vapnik, V. The Support Vector Method of Function Estimation. In *Nonlinear Modeling: Advanced Black-Box Techniques*; Suykens, J.A.K., Vandewalle, J., Eds.; Springer: Boston, MA, USA, 1998; pp. 55–85, ISBN 978-1-4615-5703-6.
8. Schapire, R.E. Explaining AdaBoost. In *Empirical Inference*; Schölkopf, B., Luo, Z., Vovk, V., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 37–52, ISBN 978-3-642-41135-9.
1. S. Avidan, "Ensemble tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, no., pp. 791–71,
2. Feb...V. VR T. Collins and Y. Liu, "On-line selection of discriminative trackingfeatures," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit*, Madison, WI, Jun.T.. Pp.
3. TES TAY [T]J. Ho, K. Lee, M.-H. Yang, and D. J. Kriegman, "Visual tracking using learned linear subspaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit*, Washington D.C., Jun. Y., pp. VAY-VAS.
4. M. Andriluka, S. Roth, and B. Schiele, "People-tracking-by-detectionand people-detection-bytracking," in *Proc. IEEE Conf. Comput. Vis Pattern Recognit.*, Anchorage, AK, Jun. T. A. pp. 1-A, (A) B.
5. Leibe, K. Schindler, and L. Van Gool, "Coupled detection and trajectory estimation for multiobject tracking" in *Proc. IEEE 1th IntConf. Comput. Vis.*, Rio de Janeiro, Brazil, Oct..., pp. 1-A TAIK. Okuma, A. Taleghani, O. Freitas, J. Ittle, and D. Lowe, "A boostedparticle filter Multitarget detection and tracking," in *Proc. Eur. Conf. Comput. Vis.*, Prague, Czech Republic, May ..., pp. YA
6. R. Serrano-Gotarredona, M. Oster, P. Lichtstelter, A. Linares-Barranco R. Paz-Vicente, F. GomezRodriguez, L. Camunas-Mesa, R. Berner, M Rivas-Perez, T. Delbruck, S.-C. Liu, R. Douglas, P. Hafliger, G. Jimenez Moreno, A. C. Ballcells, T. Serrano-Gotarredona, A. J. Acosta-Jimenezand B. Linares-Barranco, "CAVIAR: A 45k neuron, 5M synapse,
7. Gconnects/s AER hardware sensoryprocessing-learning-actuating systemfor high-speed visual object recognition and tracking," *IEEE Trans Neural Netw.*, vol. T., no. 1, pp. IFTY-TA, Sep. T... (AJO. Tuzel, F. Porikli, and P. Meer, "Leaming on lie groups for invariantdetection and tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Anchorage, AK, Jun...pp. -
8. Y. Wu and T. Yu, "A field model for human detection and tracking" *JEE E Trans. Pattern Anal. Mach. Intell.*, vol. TA, NO. 2, pp.YAT-Y40, May 2006
9. B. Wu and R. Nevatia, "Detection and tracking of multiple, partiallyoccluded humans by Bayesian combination of edgelet based part detectors *Int. J. Comput. Vis.*, vol. va, no. 1, pp. TFY-TFF, Nov 2007
- A. Ess, B. Leibe, K. Schindler, and L. V. Gool, "A mobile vision systemfor robust multi-person tracking." In *Proc. IEEE Conf. Comput. Vis Pattern Recognit.*, Anchorage, AK, Jun. T. A. pp. 1-8
10. D. Ramanan, D. A. Forsyth, and A. Zisserman, "Tracking people by learning their appearance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol., no. 1, pp. 8-11, Jan 2007
11. S. Avidan, "Support vector tracking." *IEEE Trans. Pattern Anal. Mach Intell.* vol. 15, no. 1, pp. 1.1.VY, Aug 2004
12. H. Grabner and H. Bischof, "On-line boosting and vision," in *ProcIEEE Conf. Comput. Vis. Pattern Recognit*, New York, Jun. T... Pp. 260-267