

Object tracking in video with correlation filter and using histogram of gradient feature

Mehrdad Ruhani¹, Hasan Farsi² and Sajad Mohamadzadeh^{3*}

1- Department of Electrical and Computer Engineering, University of Birjand, Birjand, Iran.

2- Department of Electrical and Computer Engineering, University of Birjand, Birjand, Iran.

3*- Department of Electrical and Computer Engineering, University of Birjand, Birjand, Iran.

¹m.ruhani@birjand.ac.ir, ²hfarsi@birjand.ac.ir, and ^{3*}s.mohamadzadehr@birjand.ac.ir

Corresponding author address: Sajad Mohamadzadeh, Department of Electrical and Computer Engineering, University of Birjand, Birjand, Iran. Postal Code: 9717434765.

Abstract- Nowadays, one of the most principal processes of perceiving the content of videos is the moving object tracking, in which the tracking process of a moving object is implemented in each video frame. The use of filters in this field has been increased during the last decade. Correlation filters are one of the most widely used filters in this field, however, using this filter, as usual, may cause the problem of target drifting. The present study proposes a novel method to improve the performance of the correlation filter. The advanced searching strategy can greatly reduce the problem of target drifting with examining a threshold level by calculating the average and variance in each frame. In this regard, if the level of the threshold is reduced, a mechanism will be activated to search in the current frame to recover the target. In order to describe the target, the histogram of the oriented gradients feature has been used because this feature shows the changes in illumination variation better than other features. The proposed method has been examined in single-camera mode on the TB50 and TB100 datasets. To evaluate the proposed method, several criteria including precision, correct detection rate (CDR), average location error (ALE), and frame per second were used. The results on the TB50 show that the proposed method, compared to the KCF method, has achieved an approximate improvement of 9% in the precision criterion, improvement 6% in the success rate criterion and a 50% reduction in the ALE criterion. Also, the results on TB100 has increased the precision criterion by approximately 15%, the success rate by 12%, and a favorable reduction of 50% on the ALE.

Keywords- Target drifting, Advance search, Object tracking, Correlation filter, Precision criterion.

ردیابی حرکت در ویدئو با استفاده از فیلتر همبستگی و ویژگی هیستوگرام گرادیان جهتدار

مهرداد روحانی^۱، حسن فرسی^۲، سجاد محمدزاده^{۳*}

۱- دانشکده مهندسی برق و کامپیوتر، دانشگاه بیرجند، بیرجند، ایران.

۲- دانشکده مهندسی برق و کامپیوتر، دانشگاه بیرجند، بیرجند، ایران.

۳* - دانشکده مهندسی برق و کامپیوتر، دانشگاه بیرجند، بیرجند، ایران.

¹m.ruhani@birjand.ac.ir, ²hfarsi@birjand.ac.ir, and ^{3*}s.mohamadzadehr@birjand.ac.ir

* نشانی نویسنده مسئول: سجاد محمدزاده، خراسان جنوبی، بیرجند، دانشگاه بیرجند، دانشکده مهندسی برق و کامپیوتر، کد پستی: ۹۷۱۷۴۳۴۷۶۵

چکیده- امروزه یکی از اساسی‌ترین فرآیندها به منظور ادراک محتوای ویدئو، ردیابی اشیاء متحرک است که در آن فرآیند موقعیت‌یابی یک شیء متحرک، در هر فریم ویدئو صورت می‌گیرد. استفاده از فیلترها در این حوزه در دهه‌ی گذشته مورد استقبال قرار گرفته است. فیلتر همبستگی یکی از فیلترهای پرکاربرد در این حوزه است اما، استفاده از این فیلتر به صورت معمول ممکن است سبب بروز مشکل از دست دادن هدف شود. در این مقاله روشی جهت بهبود عملکرد فیلتر همبستگی پیشنهاد شده است. روش جستجوی پیشرفته با بررسی یک سطح آستانه توسط محاسبه میانگین و واریانس هدف، در هر فریم می‌تواند مشکل از دست دادن هدف را تا حد زیادی کاهش دهد. به این ترتیب در صورت کاهش، از سطح آستانه مکانیزمی جهت جستجو در فریم جاری فعال می‌شود تا هدف را بازیابی نماید. به منظور توصیف هدف از ویژگی هیستوگرام گرادیان جهتدار استفاده شده است زیرا، این ویژگی تغییرات شدت روشنایی را نسبت به سایر ویژگی‌ها به خوبی نشان می‌دهد. روش پیشنهادی در حالت تک دوربینه، روی پایگاه داده‌های TB100 و TB50 بررسی شده است. معیارهای ارزیابی دقت، نرخ موفقیت (CDR)، میانگین خطای موقعیت مرکزی (ALE) و فریم برثانه به منظور ارزیابی استفاده شده است. نتایج بدست آمده روی پایگاه داده TB50 نشان می‌دهد، روش پیشنهادی در مقایسه با روش پایه KCF، با پیشرفت تقریبی ۹ درصد در معیار دقت، بهبود ۶ درصد در معیار نرخ موفقیت و کاهش ۵۰ درصدی معیار ALE همراه بوده است. همچنین روش پیشنهادی روی پایگاه داده TB100، معیار دقت را به میزان تقریبی ۱۵ درصد و نرخ موفقیت را به میزان ۱۲ درصد افزایش و معیار ALE را به میزان تقریبی ۵۰ درصد کاهش داده است.

واژه‌های کلیدی: از دست دادن هدف، جستجوی پیشرفته، ردیابی، فیلتر همبستگی، معیار دقت.

۱- مقدمه

حالت نامرئی از سنسورهای فعال مانند رادار استفاده می‌گردد [۴]. در فرآیند ردیابی، سیستم دنبال‌کننده، ورودی را از دوربین با سنسورهای مختلف (سنسورهای اپتیکی، سنسورهای راداری، سنسورهای مادون قرمز و غیره) می‌گیرد و حالت شیء را در هر نقطه اعم از زمان، مکان، سرعت، ابعاد و دیگر ویژگی‌های آن تخمین می‌زند. اشیایی که ردیابی می‌شوند از جنس‌های متفاوتی مانند انسان، حیوان، ربات، وسایل نقلیه، موشک، هواپیما و غیره هستند. اگرچه بسیاری از پردازش‌ها ردیابی اشیاء را مسأله‌ای حل شده

ردیابی اشیاء مسأله‌ای اساسی در بینایی ماشین است و در موارد مختلفی مانند رباتیک، نظارت بر سیستم نقلیه، ردیابی اهداف متحرک هوایی، کاربردهای نظامی و غیره استفاده می‌شود [۱-۳]. موضوع ردیابی به طور کلی شامل دو بخش است که بخش اول آن در طیف مرئی و بخش دوم آن در طیف نامرئی توسعه یافته است. در طیف مرئی اغلب از دوربین‌های ویدئویی استفاده می‌شود اما در

الگوریتم را از رخداد خطا آگاه می‌سازد. در نتیجه، خطا تصحیح می‌گردد و از انتشار خطا به کل ویدئو جلوگیری می‌شود.

ساختار این تحقیق بدین صورت است که در بخش دوم مروری بر مطالعات گذشته صورت می‌گیرد. در بخش ۳، روش پیشنهادی ارائه می‌شود و در این قسمت زیربخش‌های نحوه‌ی تعیین هدف ردیابی، استخراج ویژگی و جستجوی شی متحرک مطرح می‌گردد. پس از آن در بخش چهارم به ارزیابی روش پیشنهادی، معرفی پایگاه داده، معیارهای ارزیابی و نتایج حاصل از شبیه‌سازی پرداخته می‌شود. در بخش انتهایی نتیجه‌گیری و پیشنهادات بیان شده و در نهایت مراجع ذکر می‌شود.

۲- مروری بر مطالعات گذشته

ردیابی در ویدئو به معنای شناسایی پیوسته موقعیت یک شی متحرک و بروزسانی مداوم موقعیت آن در زمانی است که هدف یا دوربین در حال حرکت باشند.

همانطور که گفته شد، استفاده از فیلتر همبستگی در ردیابی طی سال‌های اخیر به عنوان یک روش بلادرنگ مورد ارزیابی قرار گرفته است. اساس کار این فیلتر بدین صورت است که کانولوشن^۶ دو مسیر (قطعه از تصویر) برابر با حاصل ضرب نقطه‌ای در حوزه‌ی تبدیل فوریه سریع^۷ قرار می‌گیرد و نقشه‌هایی، تحت عنوان نقشه‌های اطمینان^۸ ایجاد می‌شود. در این نقشه‌ها محل قرارگیری هدف دارای پیک بلند^۹ است و مکان‌های شبیه به آن دارای پیک‌های جانبی^{۱۰} با اندازه کوچکتر قرار دارد [۱۲]. اگرچه این روش نقاط قوت بسیاری دارد اما مواردی مثل پاسخ‌های همبستگی نامنظم و از دست دادن هدف نیز در آن مشاهده شده است. این موارد بیشتر زمان بروزسانی ظاهر هدف، در یک ردیابی طولانی همراه با انسداد، لرزش دوربین و تغییرات ظاهری بزرگ و سریع ایجاد می‌شود. در این روش نمونه نویزهایی که برای فیلتر تعریف شده‌اند قادرند تا ردیاب را منحرف کنند [۱۲]. برای کاهش خطر از دست دادن هدف، ردیاب باید پاسخ همبستگی را به‌طور مناسب استخراج کند تا نمونه‌های نویزی، کمتر شوند و ردیابی پایداری فراهم شود.

در سال ۲۰۱۰ بولم و همکارانش با تکیه بر مزایای فیلتر همبستگی، طرحی را به‌منظور حداقل کردن مربع خطای خروجی فیلتر^{۱۱} همبستگی برای تصاویر خاکستری پیشنهاد دادند. مزیت این روش شناسایی انسداد بود. در این طرح رخداد انسداد از طریق نقشه‌های اطمینان و نسبت پیک اصلی به پیک‌های فرعی شناسایی می‌شد و الگوریتم توسط آن متوقف می‌گردید. ایراد اصلی این روش نیاز داشتن به بازنشانی هدف توسط کاربر بود تا فرآیند ردیابی از

می‌داند اما ردیابی بلادرنگ اشیا در ویدئو همچنان با مشکلات و چالش‌هایی نظیر پایین بودن کیفیت تصاویر دوربین‌ها، تغییرات روشنایی، انسداد و هم‌پوشانی، تغییرات ناگهانی حرکت، ماتی حرکتی و تغییرات بافت شی متحرک روبرو است [۵]. بنابراین، تمامی این عوامل سبب پیچیدگی و دشواری مسأله ردیابی گردیده است و همچنان تحقیقات بیشتر و ارائه راهکارهای نوین در این حوزه را طلب می‌کند.

بطور کلی ردیابی به سه دسته‌ی کلی ردیابی نقطه‌ای، ردیابی توسط یک کرنل^۱ و ردیابی توسط کانتور تقسیم‌بندی می‌شود [۶]. در ردیابی نقطه‌ای، حرکت اشیا توسط چند نقطه نمایش داده می‌شود. این دسته از روش‌ها در صورت بروز چالش انسداد و یا هم‌پوشانی با مشکل مواجه می‌شوند اما برای ردیابی اهداف بسیار کوچک مفید هستند. این روش‌ها از فیلترهای معروف نظیر فیلتر کالمن و فیلتر ذره‌ای استفاده می‌کنند [۷، ۸].

ردیابی توسط کرنل یکی از پرکاربردترین دسته‌های ردیابی محسوب می‌شود. این دسته از روش‌ها هدف را توسط یک شکل هندسی مانند مستطیل و یا بیضی نمایش می‌دهند. در این حالت ممکن است بخشی از هدف، خارج از محدوده شکل هندسی قرار بگیرد و یا بخشی از پس‌زمینه وارد آن شود. این روش‌ها قادرند هر نوع هدفی را اعم از صلب و غیرصلب ردیابی نمایند. برای نمونه می‌توان به روش‌های ردیابی توسط فیلتر همبستگی، ماشین بردار پشتیبان^۲ و روش شیفت میانگین^۳ اشاره کرد [۹]. ردیابی توسط کانتور سومین دسته از روش‌های ردیابی است. علت معرفی این روش‌ها، دشواری توصیف برخی از اهداف متحرک است زیرا بعضی از اهداف مانند سر انسان، دست‌ها و دیگر اجزا به علت داشتن اشکال کامپوزیتی، با استفاده از اشکال هندسی مانند مستطیل و بیضی قابل توصیف نیستند. از جمله این روش‌ها می‌توان به ردیابی توسط کانتور^۴ و تطبیق شکل^۵ اشاره کرد [۱۰].

فیلتر همبستگی در سال‌های گذشته به عنوان یکی از روش‌های مبتنی بر کرنل در ردیابی مطرح شده است. انواع مختلفی از این نوع فیلتر در الگوریتم‌های ردیابی معرفی شده است که هر کدام از آن‌ها دارای نقاط قوت و ضعف می‌باشند. این فیلتر در الگوریتم‌های ردیابی عملکرد سریعی از خود نشان داده است اما در هنگام مواجهه با چالش‌هایی نظیر هم‌پوشانی و انسداد و همچنین چرخش‌های سطحی و سریع، هدف خود را از دست می‌دهد و قادر به بازیابی آن نیست [۱۱]. در این مقاله سعی شده است با بهبود این فیلتر، یک الگوریتم با بازدهی و پایداری بالا طراحی شود. به این منظور روش جستجوی پیشرفته، پیشنهاد شده است. این سازوکار با بررسی شرط تعریف شده توسط معیارهای میانگین و واریانس در صورت بروز خطا

ردیاب متمایز کننده‌ی مقیاس مکانی^{۱۳} بر پایه‌ی آموزش فیلتر همبستگی مکانی ارائه کرد. در این روش پس از اعمال فیلتر به تصویر و محاسبه امتیاز همبستگی در مقیاس‌های مختلف، فرآیند تصمیم‌گیری صورت می‌گرفت. از معایب این روش، عملکرد نامطلوب آن در برابر تصاویر با کیفیت پایین بود [۱۹]. در سال ۲۰۱۸، الگوریتم تطبیق یافته‌ای از فیلتر همبستگی توسط چاو و همکارانش معرفی شد [۲۰]. در این الگوریتم برخلاف روش‌های معمول از تعداد سه فیلتر به جای استفاده از یک فیلتر استفاده شده است. علاوه بر فیلتر همبستگی معمول، دو فیلتر دیگر تحت عنوان‌های فیلتر مقیاس و فیلتر بلند مدت^{۱۴} به کار گرفته شده تا دو ضریب مقیاس و حافظه بلند^{۱۵} را آموزش دهد. در این روش علاوه بر استفاده از ویژگی هیستوگرام گرادیان از ویژگی هیستوگرام محلی نیز بهره گرفته شده است. از مزایای این روش دقت بالای آن است اما این روش از لحاظ زمانی به علت بار محاسباتی بالا، سرعت مطلوبی از خود نشان نداده است [۲۰].

در سال ۲۰۱۹ روشی برای بهبود روش هنریکس و همکاران [۱۱] توسط ژانگ و همکارانش ارائه گردید. هدف آن‌ها افزایش مقاومت روش هنریکس بود. آن‌ها با افزودن فرآیند بروزرسانی مدل هدف، توانستند به نتایج بهتری دست یابند [۲۱]. اما این بهبود تنها، در برخی از چالش‌ها منجر به ثبت نتایج بهتر گردید. در سال ۲۰۲۰ یانگ و همکارانش با استفاده از فیلتر همبستگی و با تکیه بر روش هنریکس [۱۱]، الگوریتم بهبود یافته‌ای ارائه کردند. آن‌ها در این روش از چندین ویژگی به‌منظور توصیف هدف بهره گرفتند. مزیت این روش دقت بالای آن و عیب آن افزایش قابل توجه زمان اجرای الگوریتم به دلیل بار محاسباتی بالای آن بود [۲۲].

اگرچه مطالعات اخیر موفقیت‌های زیادی از خود در زمینه‌ی ردیابی اشیای متحرک نشان داده‌اند اما این حوزه همچنان نیاز به پیشرفت و ارائه روش‌های جدید و کارآمدتری دارد. برای رفع مشکل از دست دادن هدف و همچنین بازیابی هدف، در این مقاله روشی پیشنهاد شده است تا بتواند با کنترل پاسخ خروجی از گم کردن هدف جلوگیری کند و یا پس از گم شدن هدف بتواند در کمترین زمان ممکن هدف را بازیابی نماید. این روش نه تنها ضعف در برابر نمونه‌های نویزی را کاهش داده بلکه پایداری را در برابر تغییرات، افزایش می‌دهد. به این منظور در این مقاله روش جستجوی پیشرفته‌ای برای فیلتر همبستگی ارائه شده است که با استفاده از ویژگی هیستوگرام گرادیان جهتدار^{۱۶} هدف متحرک را ردیابی نموده و مقاومت روش را در برابر چالش‌هایی نظیر تغییرات شدید روشنایی، حرکات سریع، وضوح پایین و دیگر چالش‌ها افزایش می‌دهد.

سرگرفته شود [۱۳]. در این راستا، سال ۲۰۱۲ هنریکس و همکارانش از فیلترهای همبستگی برمبنای استفاده از ساختار چرخشی ردیابی به‌وسیله‌ی تشخیص^{۱۲} طرحی را پیشنهاد دادند که توسط آن به بالاترین سرعت ممکن در حوزه فوریه رسیدند. آن‌ها با تولید نمونه‌های مختلفی از هدف، یک طبقه‌بند را آموزش دادند که سبب کاهش خطای ردیابی می‌شد. مزیت دو روش ذکر شده در سرعت عملکرد آن‌ها بود. این روش‌ها از پیکسل‌های خام به‌عنوان ویژگی استفاده می‌کردند. این ویژگی در تصاویر شلوغ و یا اهداف با ابعاد یکسان و دارای مشابهت رنگی سبب از دست رفتن هدف و خطای الگوریتم می‌گردید [۱۴].

در سال ۲۰۱۴ دنیلجان و همکارانش با استخراج خصوصیات رنگ هدف، نوعی فیلتر همبستگی را ارائه کردند که علاوه بر محاسبه همبستگی نیاز بود تا مقیاس هدف نیز در هر مرحله بروزرسانی شود [۱۵]. در همین سال هنریکس و همکاران، مدل جدیدی از فیلتر همبستگی معرفی کردند که با استفاده از کرنل‌های مختلف، فرآیند جستجو در هر فریم صورت می‌گرفت. آن‌ها با بهره‌گیری از ویژگی هیستوگرام گرادیان، اثرات شدت روشنایی را تا حد زیادی کاهش دادند. همچنین مزیت استفاده از کرنل خطی در این الگوریتم سبب شد تا سرعت اجرای الگوریتم به‌صورت چشم‌گیری افزایش یابد. در دنباله‌هایی که چالش‌های موجود دارای شدت کمتری بود، الگوریتم توانست مقادیر بالایی از معیار دقت را فراهم آورد [۱۱]. عیب این روش سرعت بالا و عدم توانایی در بازیابی هدف در برخی موارد بود که باعث بروز خطا می‌گردید [۱۱].

در سال ۲۰۱۵، Ma و همکارانش با استفاده از اطلاعات زمینه، مدلی را پیشنهاد کردند. این مدل در هر فریم قادر به تخمین مقیاس هدف بود [۱۶]. در سال ۲۰۱۷، هان و همکارانش نیز در همین راستا اطلاعات زمینه را با فیلتر تلفیق کردند و تغییرات مقیاس را براساس پاسخ‌های همبستگی متوالی مدل‌سازی کردند. مزیت این دو روش دقت بالا و عیب این دو روش سرعت بسیار پایین آن‌ها بود [۱۷]. در همان سال مولر و همکارانش در راستای استفاده از اطلاعات زمینه، مدلی را برای بهبود فیلتر همبستگی معرفی کردند [۱۸]. در این روش به جای تمرکز بر تغییرات مقیاس، علاوه بر ویژگی‌های هدف ویژگی‌های چند نمونه از زمینه اطراف هدف نیز استخراج می‌شد. به این ترتیب نقشه‌های اطمینان استخراج شده شامل زمینه اطراف هدف در هر فریم می‌شد و باعث دستیابی به اطلاعات دقیق‌تری از محل هدف در هر فریم می‌گردید. این روش بار محاسباتی را زمانی که چالش‌های موجود کم بودند با افزایش روبرو می‌کرد [۱۸].

در سال ۲۰۱۷، دنیلجان روش جدیدی را مبتنی بر استفاده از

۳- روش پیشنهادی

داده‌ها موجود است). بنابراین، ابتدا موقعیت شی متحرک به منظور ردیابی مشخص می‌گردد. در این مرحله دو حالت وجود دارد: در حالت اول، شی متحرک به‌طور کامل درون تصویر جاری قرار گرفته باشد یعنی، نقطه‌ی مختصات مرکزی به‌همراه ارتفاع و عرض شی متحرک کاملاً قابل رؤیت باشند، در این حالت مشکلی وجود نخواهد داشت. حالت دوم، اگر شی متحرک به‌طور کامل درون تصویر جاری نباشد ابتدا مختصات مرکز آن و سپس عرض و ارتفاع آن نیز تعیین می‌شود اما برای نقاطی از شی متحرک که قابل رؤیت نیستند، مقادیر مرزهای تصویر تکرار خواهد شد.

۳-۲- استخراج ویژگی

مدل مشاهده^{۱۷}، شیوه‌ی یافتن اطلاعات اندازه‌گیری شده در هر فریم ویدئو را نمایش می‌دهد. این بدان معناست که، با استفاده از اطلاعات اندازه‌گیری شده در هر فریم، مکان شی متحرک تخمین زده می‌شود. یکی از متداول‌ترین مدل‌های مشاهده، استفاده از اطلاعات بافت شی متحرک است. نمونه‌ای از این روش استفاده از اطلاعات هیستوگرام گرادیان^{۱۸} شی متحرک است [۲۳]. هیستوگرام گرادیان، تغییرات شدت روشنایی را نسبت به سایر ویژگی‌ها از قبیل هیستوگرام معمولی تصویر به خوبی نشان می‌دهد به همین دلیل، از آن برای مدل مشاهده می‌توان استفاده کرد. هیستوگرام گرادیان جهتدار، گرادیان را روی همه‌ی پیکسل‌ها به همراه جهت بردارهای گرادیان محاسبه می‌نماید. ایده اصلی این روش بدین صورت است که اگر اطلاعاتی از موقعیت دقیق گرادیان و یا لبه‌های متناظر موجود نباشد، توزیع گرادیان‌های محلی یا جهت‌های لبه می‌تواند به خوبی شکل را توصیف کند.

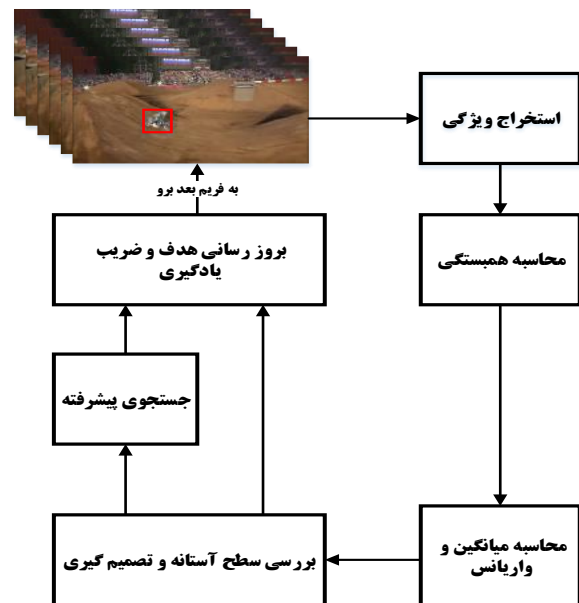
هیستوگرام گرادیان را نمی‌توان به یکباره و طی یک مرحله برای یک تصویر محاسبه نمود. برای محاسبه‌ی آن از یک فیلتر مکانی که لبه‌ها را آشکارسازی می‌کند، استفاده می‌شود. به این منظور ابتدا تصویر در جهت افقی و سپس در جهت عمودی فیلتر می‌شود. فیلتر پیشنهادی برای این کار فیلتر سوبل^{۱۹} می‌باشد که یک نوع آشکارساز لبه است و تغییرات را برجسته می‌کند. بنابراین، می‌توان آن را یک فیلتر بالاگذر نامید. میزان دامنه و جهت گرادیان با استفاده از روابط زیر محاسبه می‌شود [۲۳].

$$g = \sqrt{g_x^2 + g_y^2} \quad (1)$$

$$\theta_g = \arctan \frac{g_y}{g_x} \quad (2)$$

$$\hat{\theta}_g = \begin{cases} \theta_g & ; 0 \leq \theta_g \leq 180 \\ \theta_g - 180 & ; 180 \leq \theta_g \leq 360 \end{cases} \quad (3)$$

سه مرحله‌ی اساسی در روش‌های ردیابی شی در ویدئو وجود دارد که عبارتند از: (۱) مشخص کردن هدف، (۲) توصیف هدف توسط یک ویژگی، (۳) انتخاب یک مکانیزم جستجو در هر فریم از ویدئو. این مراحل شرایط لازم برای یک الگوریتم ردیابی است اما لازم است زمانی که الگوریتم ردیابی دچار خطا شده و هدف ردیابی را از دست می‌دهد، الگوریتم قادر به تصحیح خطای خود بوده و هدف را دوباره بازیابی نماید. روش پیشنهادی با معرفی مکانیزمی تحت عنوان جستجوی پیشرفته سعی بر آن داشته است تا خطای رخ داده را تصحیح و الگوریتم را در برابر چالش‌ها مقاوم کند. روش پیشنهادی با فرض وجود یک دوربین تصویربرداری ارزیابی شده است، اما در حالت تک دوربین فرض شده است که دوربین هم می‌تواند متحرک باشد، یعنی علاوه بر شی متحرک، دوربین نیز می‌تواند در فواصل مختلف حرکت نماید. شکل ۱ بلوک دیاگرام روش پیشنهادی را نمایش می‌دهد. در ادامه هر یک از بلوک‌ها مورد بررسی قرار می‌گیرند.



شکل ۱: بلوک دیاگرام روش پیشنهادی.

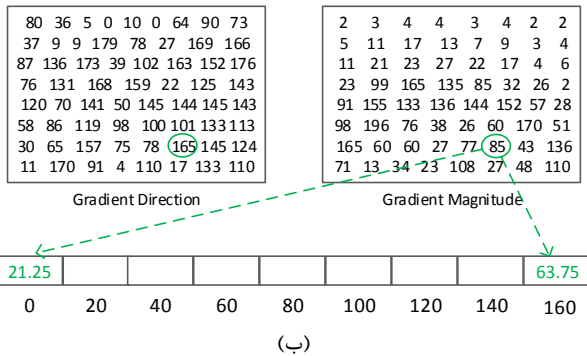
۳-۱- قرار دادن هدف در قاب ردیابی

اولین مرحله در فرآیند ردیابی تعیین موقعیت اولیه شی متحرک است. ابتدا شی مورد نظر برای ردیابی (هر نوع شی مشخصی مانند انسان، اتومبیل، شی صلب، شی غیر صلب و ...) در فریم اول توسط یک جعبه‌ی محاطی بوسیله‌ی کاربر انتخاب می‌گردد. در این مقاله مختصات اولیه شی هدف در اولین فریم استخراج شده و به‌عنوان ورودی الگوریتم در نظر گرفته می‌شود (این مختصات در پایگاه

است. در روش پیشنهادی برای استخراج ویژگی (به منظور محاسبه‌ی همبستگی تصویر شی متحرک و بخش‌های مختلف تصویر) به جای استفاده‌ی مستقیم از پیکسل‌های خام از ویژگی هیستوگرام گرادیان جهت‌دار استفاده می‌شود.



(الف)



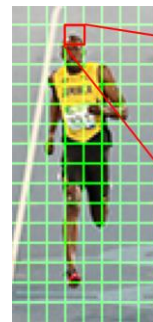
شکل ۳: الف- اعمال گرادیان به تصویر هدف متحرک. ب- محاسبه بردار ویژگی با استفاده از اندازه و جهت گرادیان.

۳-۳- محاسبه پاسخ همبستگی

این مرحله یکی از مهم‌ترین بخش‌های فرآیند ردیابی است که مرحله جستجو نامیده می‌شود. در این مرحله هدف، تعیین و یافتن ناحیه‌ای از تصویر است که شی متحرک (هدف فرآیند ردیابی)، در آن ناحیه قرار دارد. در این بخش به منظور انجام عملیات جستجو از فیلتر همبستگی استفاده شده است. شیوه‌ی تعیین همبستگی میان یک تصویر کوچک و یک بخش از فریم جاری از دنباله ویدئو، به این صورت است که باید میزان همبستگی میان تصویر شی متحرک با کلیه‌ی بخش‌های فریم ورودی سنجیده شود. از آن جایی که در فریم ابتدایی هدف متحرک درون یک قاب در مرحله اول تعیین شده است، در این مرحله همبستگی این قاب با تمامی بخش‌های ممکن فریم ورودی، محاسبه می‌شود و بخشی که دارای بیشینه‌ی میزان همبستگی باشد به عنوان بخشی از تصویر در فریم جاری که شی متحرک، درون آن قرار دارد برگزیده می‌شود.

فیلتر همبستگی با استفاده از بردار ویژگی محاسبه شده از مرحله قبل، عملیات جستجو را انجام می‌دهد. به این ترتیب که، میزان همبستگی هر یک از بردارهای ویژگی بخش‌های تصویر ورودی با

در رابطه (۱) نماد g_x گرادیان در جهت افقی و g_y بیانگر گرادیان در جهت عمودی می‌باشد. g بیانگر دامنه‌ی گرادیان و θ_g بیانگر جهت گرادیان است. در رابطه (۳)، θ_g نشان‌دهنده جهت گرادیان تبدیل شده به زاویه صفر تا ۱۸۰ می‌باشد. شکل ۲ نمایشی از هیستوگرام گرادیان جهت‌دار را نشان می‌دهد که در آن تصویر به سلول‌های 8×8 تقسیم شده است، در روش پیشنهادی از این ویژگی به صورت بلوک‌های 4×4 استفاده شده و اندازه هر سلول 8×8 پیکسل در نظر گرفته شده است. بردار ویژگی برای هر بلوک 4×4 لغزان (با هم‌پوشانی ۵۰ درصد که کل تصویر مورد نظر را در جهت افقی و چرخشی و عمودی پیمایش می‌کند) با پشت سر هم گذاشته شدن بردار ویژگی‌های هر سلول و سپس هر بلوک، محاسبه می‌گردد.



(ج)

2	3	4	4	3	4	2	2
5	11	17	13	7	9	3	4
11	21	23	27	22	17	4	6
23	99	165	135	85	32	26	2
91	155	133	136	144	152	57	28
98	196	76	38	26	60	170	51
165	60	60	27	77	85	43	136
71	13	34	23	108	27	48	110

(الف)

80	36	5	0	10	0	64	90	73
37	9	9	179	78	27	169	166	
87	136	173	39	102	163	152	176	
76	131	168	159	22	125	143		
120	70	141	50	145	144	145	143	
58	86	119	98	100	101	133	113	
30	65	157	75	78	163	145	124	
11	170	91	4	110	17	133	110	

(ب)

شکل ۲: نمایشی از اندازه و جهت گرادیان، الف) اندازه گرادیان، ب) جهت گرادیان، ج) تصویر برش خورده هدف متحرک.

در تصاویر رنگی، گرادیان برای هر کانال رنگی به صورت مجزا محاسبه می‌گردد. جهت گرادیان در این حالت بین صفر تا ۱۸۰ درجه در نظر گرفته می‌شود. به این صورت که فاصله بین صفر تا ۱۸۰ به n طرف ۲۰ تقسیم می‌شود. n نشانگر تعداد کانال هیستوگرام است. مقدار n در روش پیشنهادی برابر با ۹ در نظر گرفته شده است. طبق رابطه (۳) محدوده‌ی جهت، از ۳۶۰ درجه به ۱۸۰ درجه کاهش پیدا می‌کند و علت استفاده از محدوده‌ی صفر تا ۱۸۰ به جای ۳۶۰ درجه بخاطر این است که محدوده صفر تا ۳۶۰ به تعداد ظرف بیشتری برای استخراج اطلاعات نیاز دارد که سبب بالا رفتن حجم محاسبات و میزان داده و در نتیجه افزایش زمان اجرای روش پیشنهادی می‌شود.

این ویژگی همان‌طور که در شکل ۳-الف، نمایش داده شده است به خوبی می‌تواند لبه‌های شی موجود در تصویر را مشخص کند و عملکرد فرآیند ردیابی را به خوبی بهبود ببخشد. در شکل ۳-ب نحوه‌ی محاسبه بردار ویژگی برای یک سلول 8×8 قابل مشاهده

رابطه (۷) تعیین می‌شود [۱۱].

$$y_n = \max (F^{-1}(F(y))) \quad (۷)$$

۳-۴- محاسبه میانگین و واریانس

پاسخ همبستگی شی متحرک براساس رابطه (۸) بیان می‌شود که از توزیع گوسی پیروی می‌کند. هدف فرآیند ردیابی، کمینه کردن فاصله‌ی موقعیت تخمین زده شده‌ی شی متحرک، با مکان اصلی شی متحرک در فریم جاری است. بنابراین y^t همان پاسخ خروجی در فریم t ام است که دارای میانگین (μ^t) و واریانس $(\sigma^{2,t})$ می‌باشد که به‌صورت رابطه (۴) بیان شده است.

$$y^t \sim N(\mu^t, \sigma^{2,t}) \quad (۸)$$

ضریب یادگیری استفاده شده به‌منظور افزایش دقت و کارایی در هر فریم تصحیح می‌شود. این تصحیح توسط نرخ یادگیری η صورت می‌پذیرد و با $F(\alpha^t)$ نمایش داده می‌شود. نماد $F(\alpha^{t-1})$ ضریب یادگیری در فریم قبل می‌باشد. $F(\alpha^t)$ با استفاده از رابطه (۹) بروزرسانی می‌شود که در آن نرخ یادگیری η از رابطه (۱۰) قابل محاسبه است. در این رابطه λ و s مقادیر ثابت هستند. پس از انتخاب و یافتن مسیر جدید در فریم جاری برای هدف متحرک، ظاهر هدف طبق رابطه (۱۱) بروزرسانی می‌گردد. در این رابطه \hat{x}^t هدف بروزرسانی شده برای استفاده در فریم بعد است. x^t هدف محاسبه شده در فریم جاری و \hat{x}^{t-1} هدف بروز رسانی شده از فریم قبل می‌باشد [۱۱].

$$F(\alpha^t) = (1 - \eta)F(\alpha^{t-1}) + \eta F(\alpha) \quad (۹)$$

$$\eta = \frac{F(K^{xx}) + \lambda}{F(K^{xx}) + \lambda + 4\lambda s} \quad (۱۰)$$

$$\hat{x}^t = (1 - \eta)\hat{x}^{t-1} + \eta x^t \quad (۱۱)$$

۳-۵- جست‌وجوی پیشرفته

به‌علت تغییرات ظاهری هدف، اعم از چرخش، تغییر ابعاد و رخداد انسداد ممکن است طی فریم‌های متوالی، شی متحرک در حال ردیابی گم شود و به عبارت دیگر هدف از دست برود. اگر هدف از دست برود و ردیاب قادر به بازیابی آن نباشد، فرآیند ردیابی در این حالت با شکست مواجه می‌شود. برای حل مشکل از دست دادن هدف، در روش پیشنهادی سازوکار جستجوی پیشرفته در نظر گرفته شده است.

بردار ویژگی شی متحرک آخرین فریم مقایسه می‌شود و قسمتی از تصویر که دارای بالاترین مقدار است، به‌عنوان قطعه‌ی معتبر تصویر انتخاب می‌گردد. رابطه (۴) نحوه محاسبه همبستگی را بین دو بخش از تصویر بیان می‌کند [۱۱].

$$K^{xx'} = \exp\left(-\frac{1}{\sigma^2}(\|x\|^2 + \|x'\|^2 - 2F^{-1}(\sum \hat{x}^* \odot \hat{x}'))\right) \quad (۴)$$

رابطه (۴) فرم کرنل گوسی فیلتر را نمایش می‌دهد. σ^2 نشان دهنده واریانس و x و x' دو بردار هستند که نرم آن‌ها با $\|x\|$ و $\|x'\|$ نمایش داده می‌شود و مقادیر ثابتی هستند. بنابراین، تنها حاصل $\hat{x}^* \odot \hat{x}'$ در حوزه‌ی فوریه محاسبه می‌شود.

رابطه (۵) نحوه میزان همبستگی را برای شی متحرک x و مسیر انتخابی z نمایش می‌دهد در این رابطه z بیانگر مسیرهای انتخاب شده و x بیانگر هدف ردیابی در هر فریم است. در این رابطه علاوه بر محاسبه همبستگی یک ضریب یادگیری نیز افزوده می‌شود [۱۱].

$$F(y) = F(K^{zx}) \odot F(\alpha) \quad (۵)$$

$$F(\alpha) = \frac{F(y)}{F(K^{xx}) + \lambda} \quad (۶)$$

در رابطه (۵)، $F(y)$ پاسخ خروجی برای تمامی مسیرهای امتحان شده در دامنه فرکانس است و $F(\alpha)$ برابر با تبدیل فوریه سریع α است. ضریب یادگیری است که در تصویر ابتدایی ورودی براساس تصویر شی متحرک و سایر مسیرهای موجود در فریم اولیه ورودی آموزش می‌بیند. این ضریب در هر مرحله بروزرسانی می‌شود و در پاسخ همبستگی محاسبه شده برای هر مسیر ضرب می‌شود. به این ترتیب چون در ویدئو جابجایی بین دو فریم کم است، در عملیات ردیابی علاوه بر استفاده از فیلتر همبستگی از یک ضریب یادگیری به‌طور موازی استفاده می‌شود. K^{zx} ماتریس همبستگی محاسبه شده برای شی متحرک مرحله‌ی قبل و تمامی مسیرهای موجود در فریم ورودی جدید است. عملگر \odot ضرب نقطه به نقطه را روی دو بردار هم بعد انجام می‌دهد.

نماد K^{xx} در رابطه (۶) همبستگی تصویر شی متحرک با خودش می‌باشد و به عبارتی خودهمبستگی مسیر x را بیان می‌کند. در این رابطه λ یک مقدار ثابت است. در نهایت تمامی مقادیر y_n برای تمامی مسیرهای تصویر ورودی محاسبه شده و مورد بررسی قرار می‌گیرد تا بیشترین مقدار آن انتخاب شده و مسیر مربوط به آن مشخص شود. پس از آن که مسیر مورد نظر در فریم جاری انتخاب شد آن‌گاه، موقعیت شی متحرک در فریم جدید مشخص می‌شود و برای استفاده در تعیین موقعیت شی متحرک در فریم بعدی ذخیره می‌گردد. مقدار y_n از طریق محاسبه عکس تبدیل فوریه سریع طبق

جستجوی پیشرفته براساس میانگین و واریانس محاسبه شده در هر فریم انجام می‌شود. اگر حاصل رابطه (۱۲) از یک حد آستانه که با T نمایش داده می‌شود کمتر باشد فرآیند ردیابی به درستی انجام شده و در فریم‌های آتی ادامه می‌یابد. اگر رابطه (۱۲) از یک حد آستانه بیشتر شود در این حالت پاسخ همبستگی برای مسیر انتخاب شده (بیشترین پاسخ همبستگی)، از توزیع گوسی پیروی نمی‌کند. برای رفع این مشکل و انجام جستجوی پیشرفته، ابتدا باید مناطقی با بیشترین احتمال نزدیک به مرکز موقعیت شی متحرک در فریم قبلی بررسی شود. چون در این نواحی احتمال ظاهر شدن شی متحرک دارای بیشینه احتمال است. پس جستجو در اطراف مرکز آخرین موقعیت شی متحرک (x, y) انجام می‌شود. مختصات مراکز نقاط اطراف آخرین موقعیت شی متحرک براساس روابط (۱۳) و (۱۴) محاسبه می‌شود و به‌عنوان موقعیت جدید شی متحرک در نظر گرفته می‌شود.

۴- نتایج شبیه‌سازی

۴-۱- پایگاه داده

پایگاه داده‌های استفاده شده در این مقاله، پایگاه داده TB50 [۲۴] و پایگاه داده TB100 [۲۵] نام دارند که از جمله پایگاه داده‌های معروف و متداول در فرآیند ردیابی می‌باشند. این پایگاه داده‌ها به ترتیب شامل ۵۰ و ۱۰۰ دنباله تصویر ویدئویی هستند که هر یک از این ویدئوها یک و یا چند چالش را در بردارند. شکل ۴ نمونه‌هایی از تصاویر این دو پایگاه داده را نمایش می‌دهد.

۴-۲- معیارهای ارزیابی

۴-۲-۱- معیار دقت (Precision)

اگر هدف در حال ردیابی در فریم t ام با T_i و تخمین ردیابی با E_i نمایش داده شود آنگاه معیار دقت به‌صورت رابطه (۱۷) بیان می‌گردد.

$$Precision = \frac{1}{M} \sum_{i=1}^M 1(|E_i - T_i| \leq \tau) \quad (17)$$

در این رابطه عبارت (.) نشان دهنده تابعی است که اگر شرط درون آن برآورده شود مقدار آن برابر یک و در غیر این صورت صفر خواهد شد. τ مقدار آستانه خطای قابل تعیین برای محاسبه میزان دقت است که برحسب پیکسل بیان می‌گردد. همچنین $|.$ بیانگر فاصله میان مرکز هدف و مرکز تخمین می‌باشد [۱۱].

۴-۲-۲- معیار میانگین خطای موقعیت (ALE²¹)

این معیار بیانگر میانگین فاصله اقلیدسی بین مرکز هدف متحرک (T_i) و مرکز تخمین (E_i) است و طبق رابطه (۱۸) بیان می‌شود.

$$ALE = \frac{1}{M} \sum_{i=1}^M \|E_i - T_i\| \quad (18)$$

در این رابطه $\|.$ بیانگر فاصله اقلیدسی و M تعداد فریم‌های یک دنباله را نشان می‌دهد. در این معیار خط سیر حرکتی هدف برحسب پیکسل مورد ارزیابی قرار می‌گیرد [۲۶].

جستجوی پیشرفته براساس میانگین و واریانس محاسبه شده در هر فریم انجام می‌شود. اگر حاصل رابطه (۱۲) از یک حد آستانه که با T نمایش داده می‌شود کمتر باشد فرآیند ردیابی به درستی انجام شده و در فریم‌های آتی ادامه می‌یابد. اگر رابطه (۱۲) از یک حد آستانه بیشتر شود در این حالت پاسخ همبستگی برای مسیر انتخاب شده (بیشترین پاسخ همبستگی)، از توزیع گوسی پیروی نمی‌کند. برای رفع این مشکل و انجام جستجوی پیشرفته، ابتدا باید مناطقی با بیشترین احتمال نزدیک به مرکز موقعیت شی متحرک در فریم قبلی بررسی شود. چون در این نواحی احتمال ظاهر شدن شی متحرک دارای بیشینه احتمال است. پس جستجو در اطراف مرکز آخرین موقعیت شی متحرک (x, y) انجام می‌شود. مختصات مراکز نقاط اطراف آخرین موقعیت شی متحرک براساس روابط (۱۳) و (۱۴) محاسبه می‌شود و به‌عنوان موقعیت جدید شی متحرک در نظر گرفته می‌شود.

$$\left| \frac{y^t - \mu^t}{\sigma^t} \right| \quad (12)$$

$$P_x = \begin{cases} x_0 + i_r \times r_s \times \cos(i_t \times t_s) & ; i_t: even \\ x_0 + i_r \times r_s \times \cos\left(i_t \times t_s + \left(\frac{t_s}{2}\right)\right) & ; i_t: odd \end{cases} \quad (13)$$

$$P_y = \begin{cases} y_0 + i_r \times r_s \times \sin(i_t \times t_s) & ; i_t: even \\ y_0 + i_r \times r_s \times \sin\left(i_t \times t_s + \left(\frac{t_s}{2}\right)\right) & ; i_t: odd \end{cases} \quad (14)$$

شرط گوسی، طبق رابطه (۱۲)، به گونه‌ای انتخاب می‌شود که پاسخ خروجی از توزیع گوسی پیروی کند به این معنی که، یک نمونه زمانی انتخاب می‌شود که پاسخ خروجی توزیع گوسی داشته باشد. در رابطه‌های (۱۳) و (۱۴) مقادیر (x_0, y_0) همان مختصات نقطه‌ی موقعیت شی متحرک، در فریم قبلی است (آخرین موقعیت ثبت شده‌ی شی متحرک). P_x و P_y مختصات هدف در فریم جاری است. همچنین مقادیر $i_r = 1, 2, \dots, 5$ و $i_t = 1, 2, \dots, 16$ تعیین شده‌اند. انتخاب ابعاد و مقادیر بزرگتر برای جستجو باعث افزایش دقت و همچنین افزایش میزان محاسبات خواهد شد که نتیجه‌ی آن، کاهش عملکرد زمان واقعی برنامه است. مقدار $t_s = \frac{\pi}{8}$ در نظر گرفته شده است. مسیرهای (Z_i) در نظر گرفته شده برای جستجوی پیشرفته برابر $5 \times 16 = 80$ است که از اطراف آخرین مرکز شی متحرک استخراج شده‌اند و به‌صورت رابطه (۱۵) نمایش داده می‌شود.

$$Z = \{Z_1, Z_2, \dots, Z_{80}\} \quad (15)$$

$$R_i = \max (F^{-1}(F(Z_i) = F(K^{Z_i X}) \odot F(\alpha))) \quad (16)$$

در روش جستجوی پیشرفته بیشترین مقدار همبستگی به‌ازای



شکل ۴: نمونه‌هایی از تصاویر پایگاه داده‌های TB50 (۸-۱) و TB100 (۹-۱۶).

ثانیه است.

۴-۲-۳- معیار نرخ موفقیت (CDR^{22})

معیار نرخ تشخیص درست یا نرخ موفقیت، میانگین تشخیص‌های صحیح الگوریتم را گزارش می‌کند که توسط رابطه (۱۹) بیان می‌گردد.

$$CDR = \frac{1}{M} \sum_{i=1}^M 1_{\left(\frac{|E_i \cap T_i|}{|E_i \cup T_i|} > 0.5\right)} \quad (19)$$

در این رابطه $E_i \cap T_i$ بیانگر میزان همپوشانی تخمین و مکان واقعی هدف است و $E_i \cup T_i$ نشان دهنده اجتماع تخمین E_i و هدف T_i می‌باشد. در این رابطه عبارت $1_{(\cdot)}$ نشان دهنده تابعی است که اگر شرط درون آن برآورده شود مقدار آن برابر یک و در غیر این صورت صفر خواهد شد. اگر حاصل عبارت $\frac{|E_i \cap T_i|}{|E_i \cup T_i|}$ در یک فریم از $0/5$ بیشتر شود، در آن فریم ردیابی صحیح، صورت گرفته است. این عبارت برای تمامی فریم‌ها محاسبه و متوسط‌گیری می‌شود بنابراین، این معیار، میانگین تشخیص‌های درست را گزارش می‌کند [۲۳].

۴-۲-۴- متوسط نرخ فریم برثانیه (Fps^{23})

این معیار بیانگر متوسط تعداد فریم‌هایی است که در یک ثانیه پردازش می‌شوند و برحسب فریم بر ثانیه بیان می‌شود. به عبارت دیگر، متوسط نرخ فریم برثانیه برابر با متوسط حاصل تقسیم تعداد فریم‌های یک دنباله، بر زمان اجرای الگوریتم روی آن دنباله برحسب

۴-۳- تعیین پارامترهای مسأله

ثابت‌های استفاده شده در مسأله با بررسی معیار دقت و فریم برثانیه، انتخاب شده‌اند. شکل ۵ نشان دهنده معیار دقت برای ۱۱ دنباله ویدئو می‌باشد. در این شکل معیار دقت برحسب مقادیر آستانه T در رابطه (۱۲)، محاسبه و ترسیم شده است. شکل ۶ نشان دهنده معیار فریم برثانیه برحسب مقادیر مختلف آستانه است. انتخاب مقادیر بزرگ برای سطح آستانه، سبب کاهش دقت و افزایش میزان فریم برثانیه برای هر ویدئو می‌گردد و انتخاب مقادیر کوچک سبب کاهش عملکرد زمان واقعی الگوریتم می‌شود. معیار دقت در شکل ۵ برای مقدار $1/6$ سطح آستانه، دارای بالاترین میانگین است. همچنین در شکل ۶ پس از مقدار $1/6$ ، سیر صعودی نمودارهای فریم برثانیه کاهش یافته و نرم می‌شود.

۴-۴- نتایج

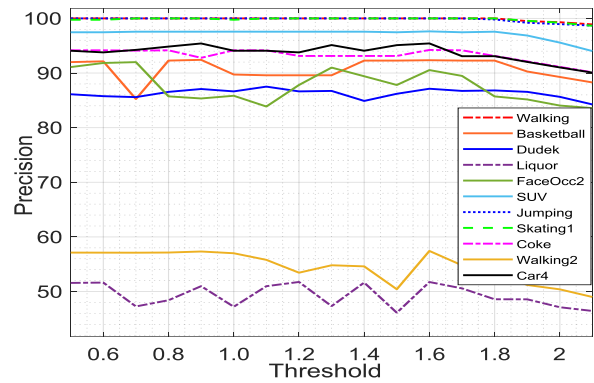
روش پیشنهادی برای تمامی دنباله‌های پایگاه داده TB100 و TB50 اجرا شده است. هر یک از دنباله‌ها شامل تعداد فریم‌های متفاوت است. در جدول ۱ از دنباله‌هایی استفاده شده است که علاوه بر بالا بودن تعداد فریم، تعداد چالش رخ داده در آن‌ها بیش‌تر باشد و هم‌چنین، چالش، با شدت بالاتری رخ داده باشد. معیارهای دقت و CDR در جدول ۱ برحسب درصد بیان شده‌اند. معیار ALE بیانگر فاصله برحسب پیکسل و معیار Fps برحسب فریم بر ثانیه است.

جدول ۱: نمونه‌ای از دنباله‌های منتخب به همراه معیارهای ارزیابی.

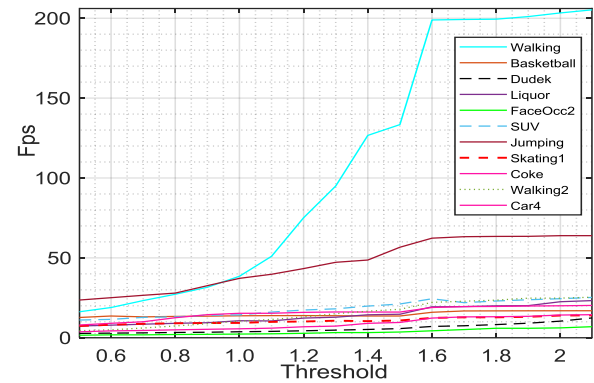
چالش	دنباله	دقت	ALE	CDR	Fps
تغییرات روشنایی	Car4	۹۵/۴۱	۹/۷۳	۸۷/۰۸	۱۲/۳۲
تغییرات مقیاس	Walking2	۵۷/۴۳	۲۰/۴۹	۳۹/۶۴	۲۲/۱۳
انسداد	Coke	۹۴/۲۲	۱۵/۳۵	۸۲/۲۸	۱۸/۹۳
تغییر فرم	Skating1	۱۰۰	۷/۲۸	۷۸/۱۶	۱۲/۴۲
ماتی حرکتی	Jumping	۱۰۰	۳/۵۴	۹۷/۱۲	۶۲/۳۲
حرکت سریع	SUV	۹۷/۶۲	۷/۹۵	۸۵/۱۴	۲۴/۳۹
چرخش سطحی	FaceOcc2	۹۰/۵۳	۷/۳۷	۶۴/۱۸	۰۴/۳۴
پس‌زمینه شلوغ	Liquor	۵۱/۷۴	۲۳/۷۷	۳۵/۲۲	۱۹/۳۵
تغییر سطح	Dudek	۸۷/۱۲	۱۱/۱۵	۷۹/۲۳	۰۷/۱۷
خارج از دید	Basketball	۹۲/۳۴	۶/۰۴	۸۳/۰۴	۱۵/۹۳
وضوح پایین	Walking	۱۰۰	۴/۱۵	۹۸/۱۱	۱۹۸/۹۰

شکل ۹ نمودارهای میزان دقت برحسب آستانه‌های متفاوت را نشان می‌دهد. نمودارها برای دنباله‌های منتخب رسم شده‌اند و نام این دنباله‌ها به صورت برچسب در پایین نمودار آورده شده است. همچنین، نمودار هر دنباله با رنگ متمایزی مشخص شده است. در این شکل با توجه به پایین آمدن میزان آستانه خطا، معیار دقت دچار افت می‌شود و بالعکس. این به آن معناست که برای داشتن معیار دقت قابل اطمینان، نیاز به تعیین یک آستانه‌ی خطای قابل قبول است. به این منظور برای محاسبه میزان دقت، مقدار آستانه τ برابر ۲۰ پیکسل را می‌توان در نظر گرفت. به عبارت دیگر درصد فریم‌هایی از یک دنباله که موقعیت تخمین زده شده آن‌ها نسبت به موقعیت واقعی هدف متحرک، کمتر از ۲۰ پیکسل باشد، مقادیر قابل اطمینانی را به دست می‌دهند [۱۱].

شکل ۱۰ شامل نمودارهایی است که هر یک میزان انحراف مرکز محاسبه شده را برای شی متحرک در حال ردیابی، در هر فریم نشان می‌دهد. به منظور یکپارچه‌سازی نمودارها محور افقی (شماره فریم) به‌ازای تمامی دنباله‌ها تا چهارصد فریم و محور عمودی تا صد

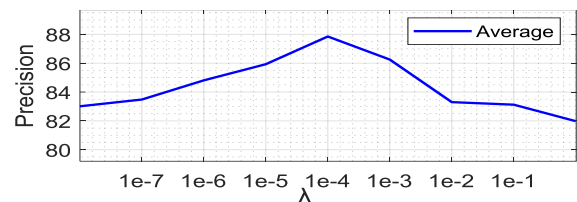


شکل ۵: معیار دقت برحسب میزان آستانه‌های مختلف.

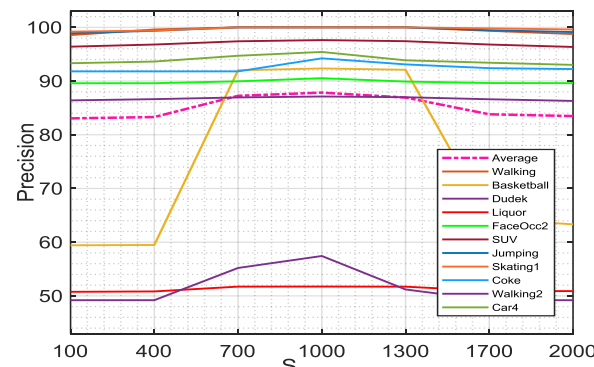


شکل ۶: نمودار معیار فریم بر ثانیه برحسب آستانه‌های مختلف.

شکل ۷ دقت متوسط را برای ۱۱ ویدئو ذکر شده در جدول ۱، برحسب λ نمایش می‌دهد. در این شکل متوسط دقت برای λ برابر 10^{-4} ، بالاترین مقدار را دارد. شکل ۸ میزان دقت را برحسب s برای ۱۱ ویدئو نمایش می‌دهد. مقدار متوسط دقت برای s برابر ۱۰۰۰، بالاترین مقدار را کسب کرده است.



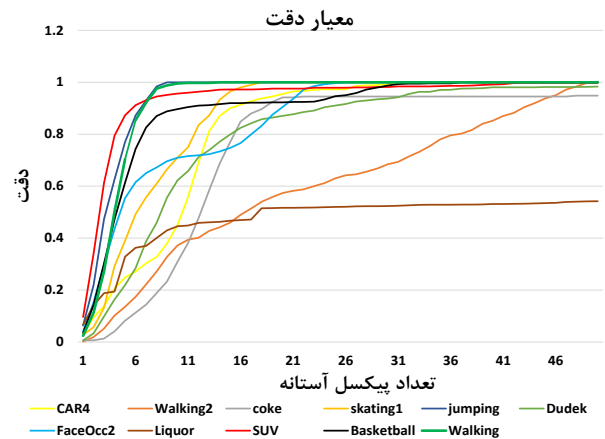
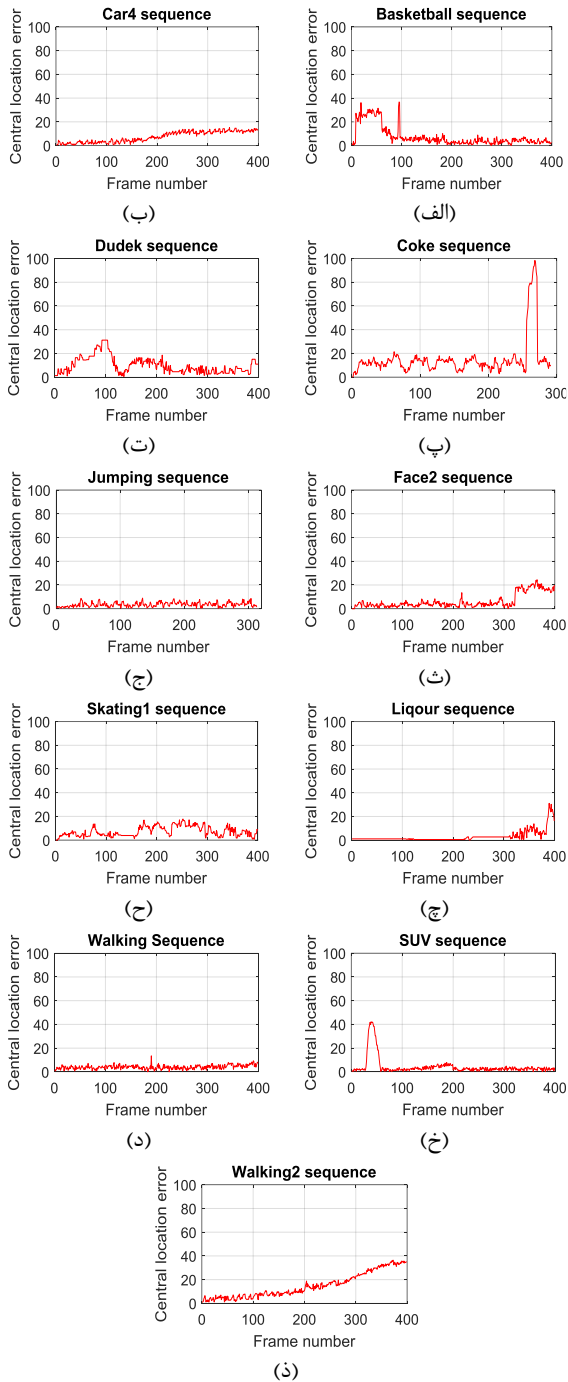
شکل ۷: میزان دقت به ازای مقادیر مختلف λ .



شکل ۸: میزان دقت به ازای مقادیر مختلف s .

پیکسل رسم شده است.

خط کشیده شده است.



این نمودارها میزان پایداری را برای الگوریتم پیشنهادی نمایش می‌دهند. به‌عنوان مثال در دنباله Coke در شکل ۱۰-پ پس از گم شدن هدف در فریم ۲۵۷ نمودار دچار پرش شده و به اندازه ۴۷ پیکسل مرکز شی تخمین زده شده با مرکز واقعی شی متحرک فاصله گرفته است. در فریم ۲۶۹ بیشترین فاصله بین تخمین و موقعیت اصلی هدف به میزان ۹۸ پیکسل رسیده است. پس از بکارگیری روش جستجوی پیشرفته، بعد از ۱۶ فریم سرانجام در فریم ۲۷۳ هدف بازیابی شده و نمودار پایدار گردیده است، همچنین فاصله به ۱۳ پیکسل کاهش یافته است. همانطور که در شکل ۹ بیان شد، فواصل کمتر از آستانه ۲۰ پیکسل به عنوان ردیابی موفق هدف در نظر گرفته شده است. در شکل ۱۰ مشاهده می‌شود در بیشتر نمودارها این فاصله کمتر از میزان آستانه خطای ۲۰ پیکسل قرار دارد و ردیابی موفقیت آمیزی حاصل شده است.

شکل ۱۱ نمایشی از دنباله‌ی Subway است. علت انتخاب این دنباله، کیفیت پایین تصاویر، بالا بودن تعداد اشیاء، تشابه زیاد بین اشیاء، فاصله نسبتاً زیاد با دوربین و کوچک بودن هدف در آن است. در نهایت روش پیشنهادی توانسته است با وجود چالش‌های ذکر شده در این دنباله، هدف (شخص) را با دقت ۹۷ درصد ردیابی کند.

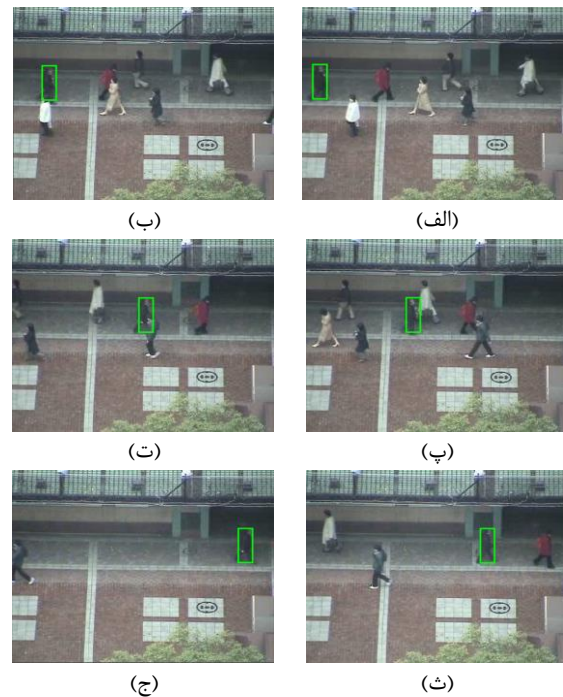
جدول ۲ به‌منظور ارزیابی روش پیشنهادی ارائه شده است. در این جدول اعداد در ستون‌های دقت و CDR برحسب درصد ثبت شده است، معیارهای ALE برحسب پیکسل و Fps برحسب فریم بر ثانیه محاسبه شده‌اند. در جدول ۲، روش پیشنهادی در مقایسه با روش ردیابی که بر پایه فیلتر هم‌بستگی مرجع [۱۱] هستند، آورده شده است. این جدول دربردارنده نتایج روی پایگاه داده‌های TB50 و TB100 است. مطلوبترین مقادیر روی پایگاه داده TB100 به‌صورت پر رنگ درآمده و زیر مطلوبترین مقادیر روی پایگاه داده TB50

شکل ۱۰: میزان انحراف تخمین از مختصات واقعی موقعیت شی متحرک در ۱۱ دنباله. (الف) دنباله Basketball، (ب) دنباله Car4، (پ) دنباله Coke، (ت) دنباله Dudek، (ث) دنباله FaceOcc2، (ج) دنباله Jumping، (چ) دنباله Liquor، (ح) دنباله Skating1، (خ) دنباله Walking، (ذ) دنباله Walking2.

معیار دقت روی TB100 نسبت به TB50 بر خلاف سایر روش‌ها در روش KCF کاهش می‌یابد. علت این پدیده وجود چالش حرکت سریع و وضوح پایین‌تر ویدئوهای TB100 نسبت به TB50 است. روش KCF دارای بالاترین سرعت اجرا است که با ۱۸۶ فریم برثانیه روی TB50 و ۱۹۶ فریم برثانیه روی TB100، قادر است هدف خود را ردیابی نماید. پس از آن روش پیشنهادی با مقدار ۵۹ و ۶۴ فریم بر ثانیه در جایگاه دوم قرار دارد. دلیل آن سادگی محاسباتی روش پیشنهادی نسبت به سایر روش‌ها است. علت بالا بودن سرعت در روش KCF عدم وجود روش جستجوی پیشرفته می‌باشد، اما این روش در مواجهه با چالش‌هایی مانند انسداد و حرکات سریع، توانایی بازیابی هدف را ندارد در حالیکه، روش پیشنهادی در برابر این چالش‌ها مقاومت بیشتری از خود نشان می‌دهد و قادر است بسیاری از دنباله‌ها را با دقت قابل قبولی ردیابی نماید.

در معیار ALE روش STCCF با ثابت مقدار ۱۲/۴۳ و ۱۱/۶۹ پیکسل روی هر دو پایگاه داده دارای بهترین مقدار و روش پیشنهادی با مقدار ۱۳/۱۲ پیکسل روی TB50 و ۱۲/۰۶ پیکسل روی TB100 پیکسل در جایگاه دوم قرار دارد. در معیار نرخ موفقیت روش STCCF با ثابت مقادیر ۶۳/۸۰ درصد روی TB50 و ۶۴/۳۱ درصد روی TB100 توانسته است نتایج بهتری نسبت به سایر الگوریتم‌ها کسب کند. پس از آن روش‌های CACF با ۵۹/۸۰ درصد روی TB50 و LSCF با ۶۲/۸۳ درصد روی TB100 قرار دارند. در نهایت روش پیشنهادی با کسب مقدار ۵۷/۴۸ درصد روی TB50 پس از روش FDSST در جایگاه چهارم و با ۶۰/۴۱ درصد روی TB100 در جایگاه سوم قرار دارد. علت پایین بودن این معیار، انتخاب قاب ردیابی کوچکتر در روش پیشنهادی نسبت به سایر روش‌ها است. از آنجایی که معیار نرخ تشخیص درست بر مبنای اشتراک تخمین و هدف متحرک است، با انتخاب قاب ردیابی بزرگتر می‌توان به مقادیر بالاتری در این معیار دست یافت، این در حالی است که، انتخاب قاب بزرگتر سبب کاهش دقت بصری در ردیابی برای کاربران می‌گردد.

جدول ۳ متوسط معیار دقت را برای ۱۱ چالش، روی پایگاه داده TB50 نشان می‌دهد. این جدول نقاط ضعف و قوت هر روش را در برابر هر چالش مشخص می‌کند. مطلوبترین مقدار به صورت پررنگ درآمده و پس از آن، زیر مقدار مطلوب در جایگاه دوم، خط کشیده شده است. نتایج محاسبه شده نشان می‌دهد روش پیشنهادی در چالش‌های انسداد، تغییر فرم، ماتی حرکتی، حرکت سریع، خارج از دید و وضوح پایین توانسته است بیشترین مقادیر را برای معیار دقت ثبت نماید. پس از آن روش LSCF توانسته است در چالش‌های تغییر روشنایی، تغییر مقیاس، پس‌زمینه شلوغ، چرخش سطحی و تغییر سطح بالاترین مقادیر را کسب نماید. با این حال، روش



شکل ۱۱: نمونه‌ای از فرآیند ردیابی در دنباله Subway، الف) فریم ۱، ب) فریم ۴۳، پ) فریم ۷۴، ت) فریم ۹۵، ث) فریم ۱۳۷، ج) فریم ۱۷۵.

جدول ۲: مقایسه روی پایگاه داده‌های TB50(I) و TB100(II)

روش‌های ردیابی		معیارهای ارزیابی			
		دقت	Fps	CDR	ALE
KCF [11]	I	۷۴/۲۲	۱۸۶	۵۱/۷۳	۲۴/۸۹
	II	۷۱/۲۵	۱۹۶	۴۹/۰۲	۲۶/۵۲
FDSST [19]	I	۷۳/۷۴	۵۴/۳۰	۵۸/۴۶	۱۵/۰۳
	II	۷۶/۴۵	۵۷/۶۲	۵۹/۹۸	۱۴/۲۶
STCCF [17]	I	۸۲/۱۰	۱۲/۰۰	۶۳/۸۰	۱۲/۴۳
	II	۸۳/۸۲	۱۳/۵۰	۶۴/۳۱	۱۱/۶۹
CACF [18]	I	۸۱/۱۰	۳۵/۲۰	۵۹/۸۰	۱۴/۳۲
	II	۸۳/۳۳	۳۷/۲۲	۶۱/۱۳	۱۳/۶۱
PAGKCF [21]	I	۷۸/۱۰	۴۸/۳۱	۵۵/۱۰	۱۴/۹۷
	II	۸۰/۲۵	۵۰/۶۱	۵۷/۹۱	۱۳/۳۱
LSCF [22]	I	۸۱/۰۰	۲۴/۰۰	۵۶/۳۱	۱۴/۱۷
	II	۸۵/۲۱	۲۵/۶۰	۶۲/۸۳	۱۲/۸۵
روش پیشنهادی	I	۸۳/۴۱	۵۹/۰۰	۵۷/۴۸	۱۳/۱۲
	II	۸۶/۰۳	۶۴/۰۰	۶۰/۴۱	۱۲/۰۶

همان‌طور که مشاهده می‌شود روش پیشنهادی توانسته است در معیار دقت بالاترین مقدار ۸۳/۴۱ درصد را روی پایگاه داده TB50 و ۸۶/۰۳ درصد روی پایگاه داده TB100 کسب کند و پس از روش STCCF با مقدار ۸۲/۱۰ درصد روی TB50 و LSCF با مقدار ۸۵/۲۱ درصد روی TB100 در جایگاه دوم قرار گرفته‌اند. متوسط

جدول ۳: معیار دقت به تفکیک چالش روی پایگاه داده TB50.

روش‌های ردیابی	چالش										
	تغییرات روشنایی	تغییرات مقیاس	انسداد	تغییر فرم	ماتی حرکتی	حرکت سریع	چرخش سطحی	پس‌زمینه شلوغ	تغییر سطح	خارج از دید	وضوح پایین
KCF [11]	۷۳/۳۵	۶۷/۹۱	۷۵/۳۶	۷۴/۷۱	۶۶/۰۴	۶۱/۰۳	۷۲/۹۳	۷۵/۳۲	۷۳/۲۲	۶۵/۰۵	۳۸/۱۳
FDSST [19]	۷۲/۸۱	۶۸/۱۰	۷۵/۱۲	۷۶/۹۰	۶۵/۸۹	۶۲/۹۲	۷۲/۳۰	۷۴/۴۰	۷۱/۲۹	۶۴/۳۱	۳۴/۹۰
STCCF [17]	۸۲/۳۱	۷۹/۳۴	۷۸/۵۶	۷۶/۲۵	۸۰/۲۲	۷۷/۶۴	۸۲/۲۱	۸۲/۰۹	۸۰/۷۱	۷۶/۹۸	۷۸/۳۲
CACF [18]	۸۴/۵۶	۸۰/۴۴	۷۸/۳۲	۷۷/۶۵	۸۱/۳۱	۷۹/۵۲	۸۳/۴۳	۸۳/۴۸	۸۱/۲۱	۷۷/۲۱	۷۶/۵۴
PAGKCF [21]	۷۸/۶۵	۷۵/۵۳	۷۹/۳۲	۷۸/۳۲	۷۷/۵۱	۷۶/۴۱	۷۸/۶۲	۸۰/۳۱	۸۰/۲۵	۷۶/۴۲	۷۳/۲۱
LSCF [22]	۸۵/۵۱	۸۱/۶۵	۷۸/۲۲	۷۹/۳۴	۸۲/۶۳	۸۰/۸۳	۸۷/۱۲	۸۶/۴۰	۸۳/۷۶	۷۸/۱۳	۸۲/۴۱
روش پیشنهادی	۸۴/۶۲	۷۶/۰۳	۸۰/۰۶	۸۵/۳۱	۸۳/۲۹	۸۱/۸۳	۸۶/۱۲	۸۲/۸۰	۸۲/۹۰	۸۰/۰۵	۸۲/۹۳

بهبود ۶ درصد و کاهش مطلوب ۵۰ درصدی معیار ALE کسب کند. این در حالی است که، روش پیشنهادی روی پایگاه داده TB100 توانسته است معیار دقت را به میزان تقریبی ۱۵ درصد و نرخ موفقیت را به میزان ۱۲ درصد افزایش دهد. همچنین معیار ALE نیز، کاهش تقریبی ۵۰ درصد را تجربه کرده است.

با بررسی نتایج می‌توان دریافت که روش پیشنهادی توانسته است عملکرد روش پایه KCF را به میزان تقریبی ۲۰ درصد در چالش حرکت سریع، و ۴۴ درصد در چالش وضوح پایین، بهبود ببخشد. هر چند این روش هنوز نیاز به بررسی در زمینه سرعت عملکرد دارد، با این حال، می‌توان ادعا کرد روش پیشنهادی ارائه شده در این مقاله می‌تواند اشیا متحرک را در یک دنباله ویدئویی به‌صورت بلادرنگ به‌طور مطلوبی ردیابی کند. در کارهای آتی به‌منظور بهبود روش پیشنهادی می‌توان از ویژگی‌های بیشتر و جدیدتری برای توصیف هدف استفاده کرد. علاوه بر این به‌منظور تعیین سطح آستانه می‌توان به جای میانگین و واریانس از معیارهای دیگری بهره گرفت.

مراجع

- [1] M. Ullah and F. A. Cheikh, "Deep feature based end-to-end transportation network for multi-target tracking", in *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 3738-3742, 2018.
- [2] H. Farsi and S. Mohamadzadeh, "Combining Hadamard matrix, discrete wavelet transform and DCT features based on PCA and KNN for image retrieval", *Majlesi Journal of Electrical Engineering*, Vol. 7, No. 1, pp. 9-15, 2013.
- [3] L. Wu, Z. Li, and X. Ma, "Video object tracking in traffic monitoring", ed: Google Patents, 2018.
- [4] M. Chiani, A. Giorgetti, and E. Paolini, "Sensor radar for object tracking", *Proceedings of the IEEE*, vol. 106, no. 6, pp. 1022-1041, 2018.

پیشنهادی در چالش‌های تغییرات روشنایی و چرخش سطحی توانسته است مقادیر بسیار نزدیکی را به روش LSCF ثبت نماید. از طرفی روش پایه KCF پایین‌ترین مقادیر را برای چالش‌های وضوح پایین با ۳۸/۱۳ درصد و حرکت سریع با ۶۱/۰۳ درصد کسب کرده است.

۵- نتیجه‌گیری

در یک فرآیند ردیابی یکی از چالش‌های موجود این است که اشیا متحرک در هنگام حرکت از مدل خاصی پیروی نمی‌کنند و در ویدئو گاهی دچار حرکات ناگهانی و سریع شده و گاهی نیز جهت و سرعت خود را تغییر می‌دهند. این مسأله از جمله مسائلی است که فرآیند ردیابی را با مشکل روبرو کرده و آن را دشوار می‌کند. در این تحقیق روشی به‌منظور بهبود فیلتر همبستگی برای ردیابی اشیا متحرک پیشنهاد شده است. این روش چارچوب جدیدی برای بهبود عملکرد ردیابی براساس روش محاسبه‌ی همبستگی است. به‌منظور بهبود بخشیدن به پایداری فیلتر همبستگی برای تغییرات هدف متحرک، استفاده ایده‌ی جستجوی پیشرفته مطرح شده است. این مسأله کارایی روش پیشنهادی را بالا می‌برد و پیچیدگی محاسباتی را نیز کاهش می‌دهد. همچنین ارزیابی‌های انجام شده نشان داد که می‌توان از روش پیشنهادی در کاربردهای بلادرنگ نیز استفاده کرد. نتایج تجربی روش پیشنهادی نشان دهنده آن است که این روش توانایی بالایی در ردیابی اشیا متحرک دارد. روش پیشنهادی در مقایسه با روش پایه خود [۱۱]، توانسته است در ۵۰ دنباله ویدئو شامل چالش‌های متفاوت از پایگاه داده TB50 بهترین نتایج را در معیارهای دقت با پیشرفت تقریبی ۹ درصد، معیار نرخ موفقیت با

- [16] C. Ma, X. Yang, C. Zhang, and M.-H. Yang, "Long-term correlation tracking", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5388-5396, 2015.
- [17] H. Lou, D. Wang, Z. Jiang, A. Men, and Y. Zhou, "Learning spatial-temporal consistent correlation filter for visual tracking", in *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 2017, pp. 501-506, 2017.
- [18] M. Mueller, N. Smith, and B. Ghanem, "Context-aware correlation filter tracking", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1396-1404, 2017.
- [19] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, "Discriminative scale space tracking", *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 8, pp. 1561-1575, 2017.
- [20] C. Ma, J.-B. Huang, X. Yang, and M.-H. Yang, "Adaptive correlation filters with long-term and short-term memory for object tracking", *International Journal of Computer Vision*, vol. 126, no. 8, pp. 771-796, 2018.
- [21] Z. He, Q. Li, M. Chang, H. Feng, and Z. Xu, "Adaptive model update strategy for correlation filter trackers", *IEEE Access*, vol. 7, pp. 151493-151505, 2019.
- [22] Y. Yang *et al.*, "Visual tracking with long-short term based correlation filter", *IEEE Access*, vol. 8, pp. 20257-20269, 2020.
- [23] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool, "Robust tracking-by-detection using a detector confidence particle filter", in *Computer Vision, 2009 IEEE 12th International Conference on*, pp. 1515-1522, 2009.
- [24] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2411-2418, 2013.
- [25] Y. Wu, J. Lim, and M.-H. Yang, "Object tracking benchmark", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1834-1848, 2015.
- [26] L. Zhang and L. Van Der Maaten, "Preserving structure in model-free tracking", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 4, pp. 756-769, 2013.
- [5] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey", *Acm computing surveys (CSUR)*, vol. 38, no. 4, p. 13, 2006.
- [6] R. B. Hatwar, S. D. Kamble, N. V. Thakur, and S. Kakde, "A Review on Moving Object Detection and Tracking Methods in Video", *International Journal of Pure and Applied Mathematics*, vol. 118, no. 16, pp. 511-526, 2018.
- [7] P. R. Gunjal, B. R. Gunjal, H. A. Shinde, S. M. Vanam, and S. S. Aher, "Moving Object Tracking Using Kalman Filter", in *2018 International Conference On Advances In Communication and Computing Technology (ICACCT)*, pp. 544-547, 2018.
- [8] V. Monteleone, L. L. Presti, and M. La Cascia, "Particle filtering for tracking in 360 degrees videos using virtual PTZ cameras", in *International Conference on Image Analysis and Processing*, pp. 71-81, 2019.
- [9] S. Zhang, X. Yu, Y. Sui, S. Zhao, and L. Zhang, "Object tracking with multi-view support vector machines", *IEEE Transactions on Multimedia*, vol. 17, no. 3, pp. 265-278, 2015.
- [10] J. W. Choi, T. K. Whangbo, and C. G. Kim, "A contour tracking method of large motion object using optical flow and active contour model", *Multimedia Tools and Applications*, vol. 74, no. 1, pp. 199-210, 2015.
- [11] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters", *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 3, pp. 583-596, 2014.
- [12] K. Zhang, L. Zhang, Q. Liu, D. Zhang, and M.-H. Yang, "Fast visual tracking via dense spatio-temporal context learning", in *European conference on computer vision*, pp. 127-141, 2014.
- [13] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters", in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2544-2550, 2010.
- [14] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels", in *European conference on computer vision*, pp. 702-715, 2012.
- [15] M. Danelljan, G. Häger, F. Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking", in *British Machine Vision Conference, Nottingham, September 1-5, 2014*.

زیر نویس ها:

¹³ Discriminative Scale Space Tracker (DSST)

¹⁴ Long term filter

¹⁵ Long memory

¹⁶ Histogram of oriented gradient

¹⁷ Observation model

¹⁸ Histogram of Gradient (HOG)

¹⁹ Sobel

²⁰ Bin

²¹ Average Location Error (ALE)

²² Correct Detection Rate (CDR)

²³ Frames per second

¹ Kernel

² Support Vector Machine

³ Mean shift

⁴ Contour tracking

⁵ Shape matching

⁶ Convolution

⁷ Fast Fourier Transform

⁸ Confidence maps

⁹ Main peak

¹⁰ Side-lobes

¹¹ Minimum Output Sum of Squared Error (MOSSE) filter

¹² Circulate Structure of Tracking-by-Detection with Kernel