

تأثیر اطلاعات بافتی و استخراج نقاط کلیدی در ردیابی بصری شی مبتنی بر تبدیل شباهت

سولماز عباسی^۱، مهدی رضائیان^{۲*}

^۱ دانشجوی دکترای دانشکده مهندسی کامپیوتر، پردیس فنی و مهندسی، دانشگاه یزد، یزد، ایران

^۲ دانشیار دانشکده مهندسی کامپیوتر، پردیس فنی و مهندسی، دانشگاه یزد، یزد، ایران

چکیده

مقاله پژوهشی

تاریخ دریافت:

۱۴۰۰/۰۸/۱۱

تاریخ پذیرش:

۱۴۰۱/۰۱/۰۷

کلیدواژه‌ها:

ردیابی خودکارشی، تبدیل شباهت، محدودسازی حداقل واریانس، الگویی دودویی محلی رنگی متضاد، ویژگی مقاوم تسریع یافته

نویسنده مسئول:

mrezaeian@yazd.ac.ir

در سال‌های اخیر، ردیابی شی در محیط‌های مختلف با اشیا متنوع، اهمیت قابل توجهی یافته است. یک ویژگی بسیار مهم، ردیابی سریع، بدون نیاز به سخت‌افزار خاص و پیش آموزش است. ردیاب‌های مبتنی بر فیلتر همبستگی متمایزکننده، نتایج مثبتی را از نظر سرعت و دقت ارائه داده‌اند. اگرچه در بسیاری از این ردیاب‌ها، موقعیت شی در هر فریم بر اساس انتقال و مقیاس‌های هرمی تخمین زده می‌شود، در الگوریتم تبدیل شباهت، انتقال، مقیاس و چرخش برای یافتن موقعیت شی برآورد می‌شوند. در این الگوریتم، ویژگی هیستوگرام گرادیان‌های جهت‌دار استخراج شده است. در مقاله‌ی پیش‌رو، دو رویکرد متفاوت جهت استخراج ویژگی در این الگوریتم اتخاذ شده است. روش اول از تصاویر مقیاس بندی شده با استفاده از محدودسازی حداقل واریانس استفاده می‌کند. سپس با روش تغییر مقیاس در ماتریس‌های هم‌رخداد، ویژگی‌ها به سطح دیگری نگاشت می‌شوند. رویکرد دوم، ترکیبی از ویژگی‌های الگوی دودویی محلی رنگی متضاد و ویژگی مقاوم تسریع یافته ارائه می‌دهد. مجموعه داده‌ی مورد ارزیابی OTB-2015 شامل ۱۰۰ دنباله‌ی ویدیویی است. هر دو رویکرد، نتایج کلی مقاله‌ی پایه را تا حدود ۳ درصد بهبود داده‌اند. روش اول در چالش رزولوشن پایین تا ۷ درصد و روش دوم در چالش چرخش تا ۴ درصد، نتایج را افزایش داده‌اند.

۱- مقدمه

مبتنی بر فیلتر همبستگی، از روش ردیابی توسط تشخیص^۳ پیروی می‌کنند. در این تحقیق بر الگوریتم‌های ردیابی با فیلتر همبستگی تمرکز شده است.

فیلترهای همبستگی متمایزکننده به‌عنوان روش‌های مبتنی بر فیلتر همبستگی به موضوع ردیابی شی وارد شده‌اند که دقت و سرعت قابل قبولی را ارائه می‌دهند [۸]. روش‌های مبتنی بر فیلتر همبستگی با استفاده از شیفت چرخشی از منطقه مورد جستجو، قادر هستند نامزدهای مطلوب تولید کنند. بر اساس تئوری کانولوشن، تبدیل محاسبات کانولوشن در حوزه زمان به ضرب در حوزه فوری که به شکل قابل توجهی پیچیدگی زمانی را کاهش می‌دهد، امکان‌پذیر است [۹]. شکل (۱)، چارچوب کلی ردیابی بصری بر اساس فیلتر همبستگی متمایزکننده را نشان می‌دهد. بر اساس روش‌های ردیابی موجود، ابتدا فیلتر همبستگی با قطعه‌ای از تصویر که موقعیتی هدف در اولین فریم است، آموزش داده می‌شود. سپس در فریم‌های بعدی، بخش پیش‌بینی شده در فریم قبلی برای تشخیص و شناسایی شی هدف به کار می‌رود. پس‌ازاین، ویژگی‌های مختلف را می‌توان از تصویر خام استخراج کرد و معمولاً یک پنجره کسینوسی برای هموار کردن تأثیر لبه‌های تصویر به کار می‌رود. در ادامه، عملیات همبستگی مؤثر با جایگزین کردن کانولوشن پیچیده و زمان‌بر با ضرب‌های عنصری با استفاده از تبدیل فوری گسسته انجام می‌شود. به دنبال روش همبستگی، یک نقشه پاسخ^۴ با استفاده از عکس تبدیل فوری به دست می‌آید. حداکثر مقدار در این نقشه به‌عنوان موقعیت جدید هدف تعریف می‌شود. سپس ویژگی‌ها در موقعیت تخمینی برای آموزش و به‌روزرسانی فیلتر همبستگی استخراج می‌شود. از آنجا که تبدیل فوری گسسته فیلتر همبستگی برای تشخیص مورد نیاز است، فرآیند آموزش و به‌روزرسانی در حوزه فرکانس انجام می‌شود.

اکثر الگوریتم‌های فیلتر همبستگی متمایزکننده یک قاب مستطیلی در راستای محور مختصات بدون در نظر گرفتن چرخش

ردیابی خودکارشی نقش مهمی را در جامعه بینایی کامپیوتر ایفا می‌کند. در سال‌های اخیر به دلیل کاربرد گسترده آن در بسیاری از زمینه‌ها از جمله تجزیه و تحلیل حرکت، تشخیص حرکت، راندگی خودکار، نظارت ویدیویی و تعامل انسان و کامپیوتر اهمیت فراوانی یافته است. الگوریتم ردیابی، بر اساس اطلاعات موجود در فریم اول، موقعیت شی را در فریم‌های بعدی تخمین می‌زند [۱]. اگرچه الگوریتم‌های بسیاری در سال‌های اخیر در زمینه ردیابی شی ارائه شده و توسعه یافته‌اند، این موضوع به دلیل محیط‌های پیچیده و ویژگی‌های مختلف در شرایط واقعی مانند چرخش درون صفحه، چرخش برون صفحه، هم‌پوشانی، وضوح کم و تغییرات روشنایی و مقیاس، همچنان یک مشکل چالش‌برانگیز است [۲].

به‌طور کلی روش‌های ردیابی شی را می‌توان در دو گروه طبقه‌بندی کرد: ردیاب‌های مبتنی بر فیلتر همبستگی^۱ و سایر ردیاب‌ها [۳]. بر اساس این تفکیک، برخی از روش‌ها از فیلتر همبستگی برای تخمین موقعیت شی استفاده می‌کنند، درحالی‌که برخی روش‌های مبتنی بر گراف [۴] و ردیاب‌های مبتنی بر تنکی^۲ [۵] می‌توانند در دسته دیگر قرار بگیرند. روش‌هایی که بر اساس گراف کار می‌کنند، پیکسل‌ها، سوپر پیکسل‌ها و بخش‌هایی از شی را به‌عنوان گره در نظر می‌گیرند [۶]. لبه‌ها توسط پیوندهایی بین گره‌ها تعریف می‌شود. این رویکردها به شکل گسترده‌ای در تشخیص چهره و تشخیص شی مورد بررسی قرار می‌گیرد. در موضوع ردیابی، گراف، بین بخش‌هایی از اشیا در فریم‌های مختلف ساخته می‌شود. قسمت‌هایی از اشیا نشان‌دهنده رأس‌ها، لبه‌ها و ساختارهای هندسی داخلی هستند. ردیاب‌هایی که بر اساس تنکی کار می‌کنند می‌توانند به‌عنوان مدل مولد در نظر گرفته شوند [۷]. هدف اصلی در ردیابی تنک، ایجاد یک نمایش بهینه از هدف به‌دست‌آمده مشروط به تنک بودن و به حداقل رساندن خطای بازسازی است. از طرف دیگر، معماری کلی الگوریتم‌های ردیابی

³ Tracking-by-detection

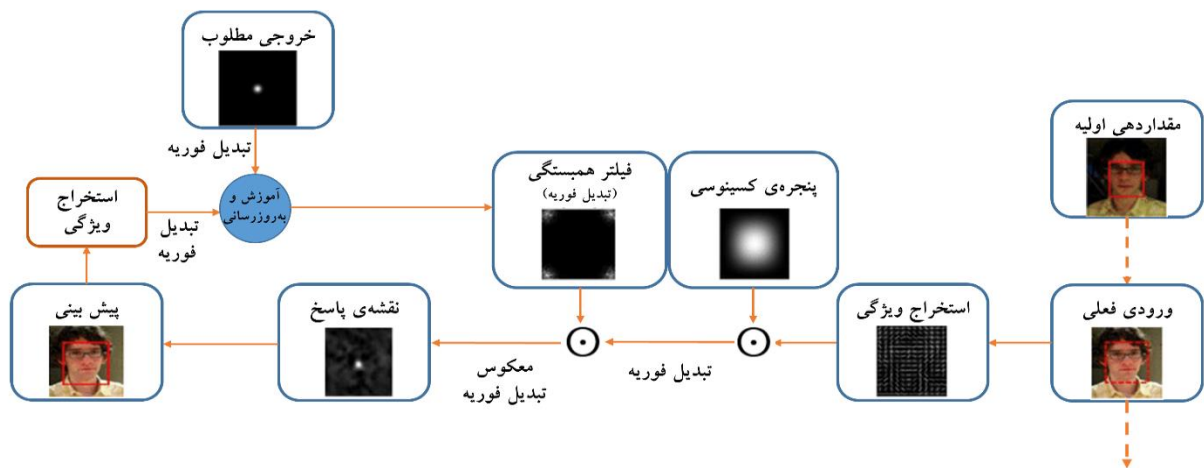
⁴ Response map

¹ Correlation Filter-based Trackers (CFT)

² Sparsity

فضای قطبی تخمین می‌زند. هیستوگرام گرادیان‌های جهت‌دار [۱۱] به‌عنوان یک ویژگی در فضای دکارتی و همچنین قطبی به‌کاررفته است. این روش از سخت‌افزار خاص شامل کارت گرافیکی قوی استفاده نمی‌کند و نیازی به آموزش قبلی ندارد. در تحقیق پیش رو، یک رویکرد سریع و خودکار بدون نیاز به سخت‌افزار خاص و شبکه‌های پیش آموزش دیده ارائه می‌شود.

و مقیاس تخمین می‌زند. در سال ۲۰۱۷، یک مسئله با دو درجه آزادی شامل انتقال در راستای محورهای x و y به مسئله‌ای با سه درجه آزادی شامل نمونه‌گیری در مقیاس‌های هرمی توسعه یافت [۹]. لی و همکاران [۱۰] در سال ۲۰۱۹، روشی با عنوان تخمین جابجایی بزرگ تبدیل همانندی ارائه دادند. در این تحقیق، دو مسئله با چهار درجه آزادی ارائه شد. یک مسئله جابجایی را در فضای دکارتی پیش‌بینی می‌کند و دیگری مقیاس و چرخش را در



شکل (۱): چارچوب کلی روش‌های ردیابی مبتنی بر فیلتر همبستگی. در هر فریم پس از مقداردهی اولیه، یک قطعه از تصویر در موقعیت تخمینی قبلی به‌عنوان ورودی فعلی برش داده می‌شود. سپس برای توصیف بهتر ورودی می‌توان ویژگی‌هایی را استخراج کرد و یک پنجره کسینوسی معمولاً برای هموارسازی لبه‌های قاب تصویر اعمال می‌شود. در مرحله بعد همبستگی ورودی فعلی و فیلتر آموخته‌شده بر اساس تئوری کانولوشن محاسبه می‌شود. در این شکل علامت \odot نشان‌دهنده محاسبات عنصر به عنصر است. پس از محاسبه همبستگی یک نقشه پاسخ از معکوس تبدیل فوری به دست می‌آید که حداکثر مقدار نشان‌دهنده موقعیت جدیدی است. در نهایت ویژگی‌ها در موقعیت جدید استخراج می‌شود و فیلتر همبستگی به‌روزرسانی خواهد شد.

می‌شود [۱۳]. پس‌از این مرحله، از تبدیل دیگری برای استخراج ویژگی استفاده شده است که ۲۵۶ شدت روشنایی را به ۸ سطح نگاهت می‌کند. فرآیند تغییر مقیاس در مرحله دوم، در ماتریس هم رخداد^۲ [۱۴] به کار می‌رود. نتایج نشان می‌دهد تغییر مقیاس می‌تواند به‌عنوان یک ویژگی مؤثر و سریع به بهبودی نتایج بیانجامد. در رویکرد دوم، ترکیبی از ویژگی‌های الگوی دودویی

روش پیشنهادی از تبدیل شباهت و ویژگی‌های معنادار بهره می‌برد. در رویکرد اول از تکنیکی برای تغییر مقیاس شدت روشنایی تصویر در فضای رنگی باهدف ردیابی و به‌عنوان ویژگی استفاده شده است. الگوریتم تغییر مقیاس به روش محدودسازی^۱ عمل می‌کند؛ به طوری که مکعب فضای رنگی قرمز، آبی و سبز را به مجموعه‌ای از مکعب‌های کوچک‌تر تقسیم می‌کند. رنگ‌های موجود در هر مکعب به مقدار رنگ واقع در مرکز آن واحد نگاهت

² Co-occurrence matrix

¹ M Quantization

همسایگی ویژگی استخراج خواهد شد. ترکیب این دو ویژگی ضمن اجرایش سریع قادر است نتایج ردیابی را در برخورد با ویژگی‌های چالش‌برانگیز بهبود بخشد.

ادامه مقاله بدین شرح تنظیم شده است: در بخش ۲ مقالات مرتبط بررسی خواهند شد. بخش ۳ روش‌های پیشنهادی را با ذکر جزئیات شرح می‌دهد. نتایج و آزمایش‌ها در بخش ۴ ارائه شده است. بخش ۵ نتیجه تحقیق را بیان خواهد کرد.

این روش از یک تابع کرنل گوسی برای تمایز بین شی و پس‌زمینه استفاده می‌کند. ویژگی هیستوگرام گرادیان‌های جهت‌دار از تصویر استخراج می‌شود. برخی از ردیاب‌ها می‌توانند به‌عنوان فیلتر همبستگی منظم^[۳] با محدودیت در تشخیص، در نظر گرفته شوند. تحلیل متمایزکننده خطی مبتنی بر نمونه^۷ برای جدا کردن شی از پس‌زمینه با استفاده از آنالیز افتراقی خطی^۸ معرفی شد. در این رویکرد چندین نمونه منفی از پس‌زمینه و یک نمونه مثبت از شی هدف انتخاب و سپس شی هدف و اجزای پس‌زمینه مدل می‌شوند. مدل پس‌زمینه شامل مدل برخط و برون‌خط^۹ است. مدل برون‌خط بر روی تعداد زیادی نمونه منفی از تصاویر واقعی آموزش داده شده و مدل برخط از بخش‌های متفاوت پس‌زمینه آموزش می‌بیند. یانگ لی و همکاران^[۱۰]، یک ردیاب کارآمد و سریع برای تخمین موقعیت هدف با تغییر مقیاس، چرخش و انتقال ارائه دادند. برخلاف روش‌های پیشین که از مجموعه^۶ هر می مقیاس با پیچیدگی زیاد استفاده می‌کردند، این ردیاب را می‌توان در زمان واقعی اجرا کرد. آن‌ها مسئله را به دو زیرمسئله فرعی شامل بهینه‌سازی تغییرات و بهینه‌سازی مقیاس و چرخش تقسیم کردند. این دو زیر مسئله به‌طور مکرر برای رسیدن به نقطه^{۱۰} بهینه اجرا می‌شوند. ویژگی هیستوگرام گرادیان‌های جهت‌دار در هر دو فضای دکارتی و قطبی استخراج شده است.

محلی رنگی متضاد^۱ [۱۵] و ویژگی مقاوم تسریع یافته^۲ [۱۶] استخراج شده‌اند. برخلاف برخی نسخه‌های توسعه‌یافته الگوی دودویی محلی که ویژگی‌ها را از تصاویر سطح خاکستری استخراج می‌کنند، این الگوریتم از اطلاعات فضای BG صورتی که ویژگی‌ها از صفحات سبز، قرمز و آبی جداگانه و همچنین ویژگی‌های ترکیبی با مرکز قرار دادن پیکسل از یک صفحه و همسایه‌ها از صفحات دیگر استخراج می‌شوند. علاوه بر این، توصیفگر مقاوم تسریع یافته، نقاط کلیدی را شناسایی کرده و از این نقاط با در نظر گرفتن

۲- مروری بر کارهای پیشین

ردیابی بصری برخط^۳ می‌تواند یک مسئله چالش‌برانگیز در پردازش ویدیو و بینایی کامپیوتر باشد. پیچیدگی زمان و دقت دو اصل مهم در ردیابی هستند. با وجود تلاش‌های مکرر در زمینه ردیابی، ردیاب‌های فعلی هنوز با مشکلات زیادی روبرو هستند. علاوه بر این، علی‌رغم دستاوردهای قابل توجه در زمینه ردیابی اشیاء خاص مانند چهره، اتومبیل، انسان و غیره، توسعه الگوریتم کارآمد برای هر شی دلخواه هنوز مسئله‌ای چالش‌برانگیز است.

برای مرور جامع روش‌های ردیابی موجود، خوانندگان را به منابع [۲-۳] و [۱۷] ارجاع می‌دهیم. در این بخش مقالات مرتبط و روش‌های پیشرفته مورد بررسی قرار می‌گیرند. به‌طور کلی، ردیاب‌های مبتنی بر فیلتر همبستگی با اهداف مقاوم و کارا بودن توسعه یافته‌اند. در تحقیقی [۸]، حداقل خروجی مجموع مربعات خطا برای آموزش تطبیقی سریع معرفی شد، مشروط به اینکه بتواند مجموع مربعات خطا را بین خروجی مطلوب و خروجی واقعی در حوزه فوری حداقل کند. دانلجان و همکاران [۱۸]، ویژگی‌های رنگ را در چارچوب ردیابی با تشخیص اعمال کردند. آن‌ها همچنین ابعاد کم و تطبیقی ویژگی‌های رنگی را ارائه دادند. در تحقیقی دیگر [۱۹]، برای رسیدن به فیلتر همبستگی دقیق‌تر، فیلترهای همبستگی چند نمایه^۴ معرفی شدند. هنریکس و همکاران [۲۰]، ردیابی بر اساس فیلتر همبستگی کرنلی^۵ را پیشنهاد کردند.

⁶ Regularized Correlation Filter

⁷ Exemplar-based Linear Discriminant Analysis

⁸ Linear Discriminant Analysis

⁹ Offline

¹ Opposite Color Local Binary Patterns

² Speeded-Up Robust Features (SURF)

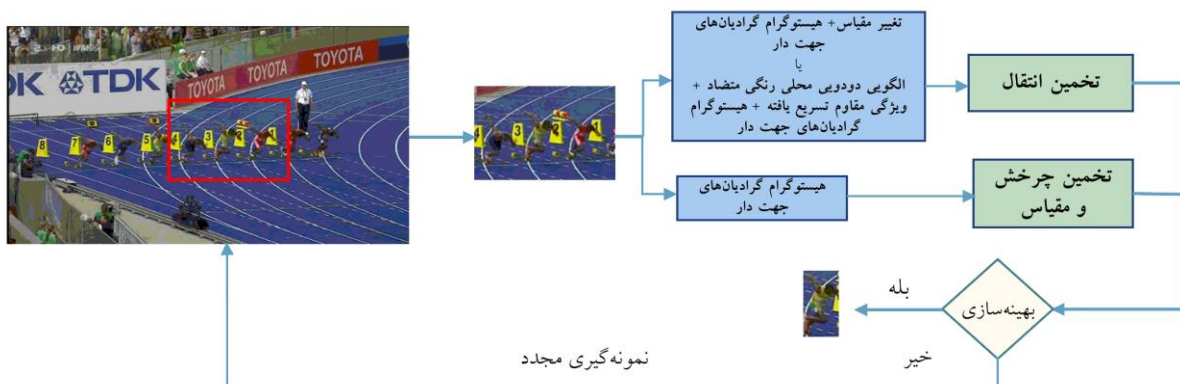
³ Online

⁴ Multi-view

⁵ Kernelized Correlation Filters

توجه به بهترین نامزدها معرفی شد [۲۳]. چائوما و همکاران [۲۴]، فیلترهای همبستگی چندگانه را به منظور دستیابی به حافظه بلندمدت و کوتاه‌مدت برای افزایش انطباق‌پذیری ترکیب کردند.

علاوه بر روش‌های فوق، الگوریتم‌های زیادی بر مبنای شبکه‌های عصبی عمیق توسعه یافته‌اند. در تحقیقی [۲۲]، یک روش جدید مبتنی بر شبکه عصبی سیامی^۱ ارائه شد. شبکه فیلتر همبستگی با مکانیزم توجه^۲ برای افزایش کارایی با معطوف کردن



شکل (۲): مسیر کلی روش پیشنهادی. در رویکرد اول، ترکیبی از هیستوگرام گرادینان‌های و تصویر مقیاس شده برای تخمین انتقال در نظر گرفته شده است. در رویکرد دوم ترکیبی از هیستوگرام گرادینان‌های جهت‌دار، الگوی دودویی محلی رنگی متضاد و ویژگی مقاوم تسریع یافته برای تخمین انتقال اعمال می‌شود. در هر دو رویکرد، ویژگی هیستوگرام گرادینان‌های جهت‌دار برای تخمین مقیاس و چرخش در فضای قطبی استخراج می‌شود. این دو زیر مسئله به طور مکرر برای رسیدن به نقطه بهینه به روزرسانی می‌شوند.

باززش، سریع و هدفمند در روش‌های ردیابی مبتنی بر فیلتر همبستگی بسیار مهم است. ما دو رویکرد متفاوت را برای استخراج ویژگی ارائه می‌کنیم و آن‌ها را در تبدیل شباهت به کار می‌گیریم. روند کلی الگوریتم پیشنهادی در شکل (۲) نشان داده شده است. روش اول با تغییر مقیاس یک تصویر رنگی ویژگی جدیدی استخراج می‌کند. این ویژگی با هیستوگرام گرادینان‌های جهت‌دار ترکیب می‌شود. رویکرد دوم برای دستیابی به یک ردیابی مؤثرتر، الگوهای دودویی محلی رنگی متضاد [۲۶]، ویژگی‌های مقاوم تسریع یافته و هیستوگرام گرادینان‌های جهت‌دار را باهم ترکیب می‌کند. ویژگی‌های استخراج شده در فضای دکارتی برای تخمین انتقال به کار می‌روند. ویژگی هیستوگرام گرادینان‌های جهت‌دار در

در تحقیقی [۲۵]، شبکه فیلترهای همبستگی متمایزکننده با استفاده از شبکه عصبی کانولوشن ارائه شد. این ردیاب از پس انتشار^۳، برای تطبیق لایه‌ها استفاده می‌کرد. در جدول (۱) خلاصه‌ای از روش‌های پیشین ارائه شده است.

۳- روش پیشنهادی

در این تحقیق دو چارچوب معرفی شده است که بر اساس فیلتر همبستگی عمل می‌کنند. این مطالعه بر اساس ردیاب تبدیل شباهت انجام شده است که برای اولین بار موفق به انجام ردیابی از طریق محاسبه تغییرات انتقال، مقیاس و چرخش شد. استخراج اطلاعات

² Attention

³ Back propagation

¹ Siamese network

مکعب رنگی قرمز، آبی و سبز، آرایه‌ای سه‌بعدی شامل رنگ‌هاست؛ ۲۵۶ سطح شدت روشنایی برای هر رنگ تعریف می‌شود. محدودسازی شامل بخش‌بندی کردن مکعب به قسمت‌های کوچک‌تر و سپس مشخص کردن رنگ آن بخش با مقدار رنگ در مرکز آن بخش است. محدودسازی را می‌توان به دو روش مختلف محاسبه کرد: محدودسازی یکنواخت و محدودسازی حداقل واریانس. این دو روش بسته به رویکردی که برای محدودسازی انتخاب می‌شود متفاوت هستند. در محدودسازی یکنواخت، مکعب رنگی به جعبه‌های کوچک و مساوی تقسیم می‌شود. در محدودسازی حداقل واریانس، مکعب به بخش‌های کوچک با اندازه‌های متفاوت و نه لزوماً مکعب تقسیم می‌شود. توزیع رنگ‌ها در تصویر، اندازه بخش‌ها را تعیین می‌کند. به‌عنوان مثال، مجموعه‌ای از پیکسل‌های آبی ممکن است باهم گروه‌بندی شوند زیرا دارای واریانس کمی نسبت به پیکسل مرکزی هستند. شکل (۳) این تقسیم‌بندی را با الگوریتم محدودسازی با حداقل واریانس نشان می‌دهد.

پس از این تغییر مقیاس، مقیاس دیگری برای تبدیل تصویر به ۸ سطح خاکستری اعمال شده است. این تغییر مقیاس در ماتریس هم‌رخداد کاربرد دارد. شکل (۴)، تصویر اولیه و تصویر مقیاس شده پس از دو مرحله را نشان می‌دهد. به نظر می‌رسد این تغییر مقیاس توانسته است اطلاعات بافتی را از تصویر استخراج کرده و بین اشیا تمایز ایجاد کند.

هیستوگرام گرادین‌های جهت‌دار یک توصیفگر ویژگی است که در بینایی کامپیوتر و پردازش تصویر به منظور تشخیصی به کار می‌رود. این روش، تعداد رخداد‌های جهت گرادین در سلول‌ها را شمارش می‌کند و بر روی شبکه‌ای مترکم از سلول‌ها بافاصله یکنواخت محاسبه می‌شود، همچنین از نرمال‌سازی برای افزایش دقت استفاده می‌کند.

فضای قطبی برای تخمین مقیاس و چرخش اعمال می‌شود. این دو تکنیک مختلف استخراج ویژگی به ترتیب در بخش ۳-۱ شرح داده شده‌اند. تبدیل شباهت در بخش ۳-۲ ارائه شده است.

جدول (۱): خلاصه‌ای از ردیاب‌های پیشین

| نوع | ردیاب | مقیاس | چرخش | پیش آموزش | GPU |
|------------------------|---------------------------|-------|------|-----------|-----|
| مبتنی بر فیلتر همبستگی | MOSSE [۸] | x | x | x | x |
| | ویژگی‌های رنگ تطبیقی [۱۸] | x | x | x | x |
| | مدل چند نمایی [۱۹] | ✓ | x | x | x |
| | KCF [۲۰] | ✓ | x | x | x |
| | OTTC [۳] | ✓ | x | ✓ | ✓ |
| | LDDES [۱۰] | ✓ | ✓ | x | x |
| مبتنی بر یادگیری عمیق | Siamfc [۲۲] | ✓ | x | ✓ | ✓ |
| | LCTdeep [۲۴] | ✓ | x | ✓ | ✓ |
| | ACF [۲۳] | ✓ | x | ✓ | ✓ |
| | DCFNet [۲۵] | ✓ | x | ✓ | ✓ |

۱-۳ استخراج ویژگی

استخراج ویژگی‌های سریع، ارزشمند و معنادار تأثیر بسزایی در مسائل بینایی کامپیوتر دارند. ویژگی‌های قوی به‌طور مستقیم بر نتایج ردیابی به‌ویژه در محیط‌های متفاوت و اشیا دلخواه تأثیر می‌گذارند. ضعف ردیاب در تشخیص و شناسایی شی در یک تصویر ممکن است منجر به منحرف شدن ردیاب از مسیر ردیابی و گم کردن و از دست دادن شیء شود.

- تصویر مقیاس شده و هیستوگرام گرادین‌های جهت‌دار

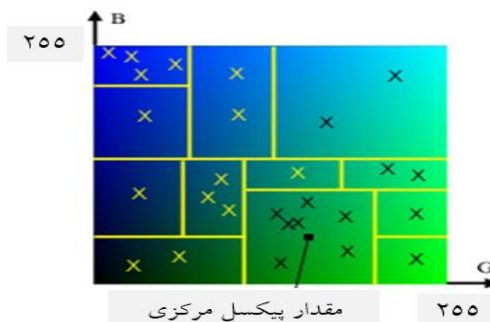
| | | | | | | | | | | |
|-----|-----|----|-----------|---|---|---|---|-----|----|----|
| 112 | 103 | 95 | Threshold | 0 | 0 | 0 | × | 1 | 2 | 4 |
| 75 | 72 | 24 | | 0 | | 1 | | 128 | | 8 |
| 41 | 76 | 20 | | 1 | 0 | 1 | | 64 | 32 | 16 |

الگوی: 01011000 or 88

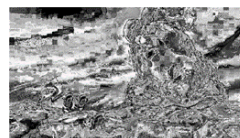
شکل (۵): الگوی دودویی محلی [۲۸]

الگوی دودویی محلی رنگی متضاد برای استخراج ویژگی از تصاویر رنگی اعمال شده است. همان‌طور که در شکل (۶) نشان داده شده است، ابتدا ویژگی‌ها به‌طور جداگانه از صفحات قرمز، آبی و سبز استخراج شده است. سه ویژگی دیگر هرکدام از دو سطح مرتبط استخراج می‌شود؛ به این صورت که پیکسل مرکزی از یکی از صفحات در نظر گرفته شده و پیکسل‌های همسایه از صفحه دیگر است. در سه تصویر اول از شکل (۶) اطلاعات بافتی استخراج می‌شوند، در حالی که در سه تصویر بعد به دلیل تنوع موجود در رنگ‌ها، تفاوت بین شی و پس‌زمینه ایجاد شده است.

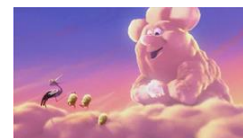
علاوه بر الگوی دودویی محلی رنگی متضاد و هیستوگرام گرادینان‌های محلی، ویژگی مقاوم تسریع یافته نیز استخراج شده است. ویژگی مقاوم تسریع یافته یک الگوریتم قوی و سریع برای نمایش ویژگی‌های محلی و مقایسه تصاویر استفاده می‌شود. مزیت اصلی این روش محاسبات سریع است که کاربردهای زمان واقعی مثل ردیابی و تشخیص اشیا را ممکن می‌سازد. برای تعیین نقاط موردعلاقه، این الگوریتم از تقریب صحیح آشکارساز حباب هسین^۱ استفاده می‌کند. توصیفگر این ویژگی مبتنی بر مجموع پاسخ‌های هار و بولت^۲ در اطراف نقطه موردعلاقه است [۲۹]. شکل (۷)، ده نقطه موردعلاقه در توصیفگر ویژگی مقاوم تسریع یافته را نشان می‌دهد. ردیاب از طریق استخراج نقاط موردعلاقه به‌عنوان ویژگی در روش مبتنی بر فیلتر همبستگی، شی را در فریم‌های متوالی دنبال می‌کند.



شکل (۳): محدودسازی با استفاده از حداقل واریانس. مکعب به بخش‌های کوچک‌تر با اندازه‌های متفاوت بر اساس واریانس بین پیکسل‌ها تقسیم می‌شود



(ب)



(الف)

شکل (۴): تغییر مقیاس شدت روشنایی فریم‌ها. الف) تصویر (ب) فریم پس از مقیاس "Bird 2" اصلی از مجموعه داده

الگوی دودویی محلی رنگی متضاد، ویژگی مقاوم تسریع یافته و هیستوگرام گرادینان‌های جهت‌دار

الگوی دودویی محلی رنگی متضاد رابطه بین پیکسل‌ها و پیکسل‌های همسایه را محاسبه کرده و این رابطه را در یک کد باینری رمزگذاری می‌کند. برخلاف الگوی دودویی محلی معمولی، این تکنیک از فضای رنگی پیکسل‌ها بهره می‌برد و می‌تواند فضای ویژگی غنی‌تری را به دست آورد. همان‌طور که در شکل (۵) نشان داده شده است، در الگوی دودویی محلی، پیکسل مرکزی فقط با همسایگان خود مورد مقایسه قرار می‌گیرد. اگر پیکسل همسایه کوچک‌تر از پیکسل مرکزی باشد مقدار یک در غیر این صورت مقدار صفر می‌گیرد. بدین صورت برای هر پیکسل یک کد دودویی تولید می‌شود. مقدار محاسبه شده از ضرب کد دودویی در وزن‌های متناظرش، به پیکسل مرکزی اختصاص می‌یابد [۲۷].

² Haar wavelet

¹ Hessian blob detector

$\rho = \{\theta, s\}$ و $\{t_x, t_y\}$ برای تخمین شباهت تعریف شده است. رابطه (۱) یک تابع هدف که ترکیب خطی سه بخش است را نشان می‌دهد:

$$f_i(\tau; h_{i-1}) = \eta f_t(t; h_t) + (1 - \eta) f_\rho(\rho; h_\rho) + g(t, \rho) \quad (1)$$

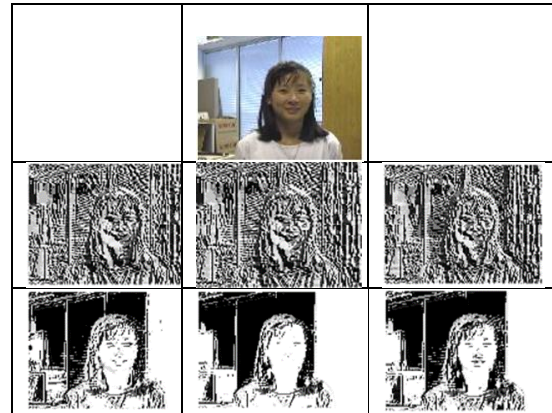
در این رابطه η ضریب درونیابی را نشان می‌دهد و f_ρ و f_t به ترتیب توابع هدف انتقال و چرخش-مقیاس را نشان می‌دهند. مدل آموزش دیده از مرحله قبل با h مشخص شده است. تابع g که مدل حرکتی است در رابطه (۲) نشان داده شده است و ترجیح می‌دهد موقعیتی در فریم فعلی نزدیک موقعیت آن در فریم قبلی باشد.

$$g(t, \rho) = \exp(|t - \tau_{i-1}|_2)^{-1} \quad (2)$$

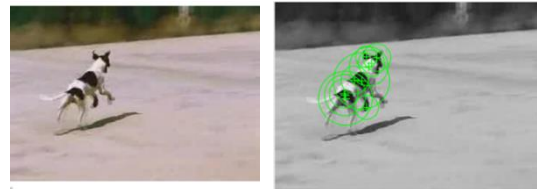
که پانویس $i - 1$ مشخص کننده موقعیتی در فریم قبلی است. بر اساس رابطه (۱)، دو زیر مسئله با روابط (۳) و (۴) تعریف و مکرراً به روزرسانی می‌شوند تا زمانی که تابع هدف افزایش پیدا می‌کند و یا به حداکثر تعداد تکرارها برسد. پس از بهینه‌سازی، فیلتر همبستگی به روزرسانی خواهد شد.

$$\operatorname{argmax}_t g(t, \rho^*) + \eta f_t(t) \quad (3)$$

$$\operatorname{argmax}_\rho g(t^*, \rho) + (1 - \eta) f_\rho(\rho) \quad (4)$$



شکل (۶): الگوی دودویی محلی رنگی متضاد. اولین تصویر، فریم اولیه است که از مجموعه داده Girl انتخاب شده است. سطر دوم، استخراج ویژگی از هر صفحه را به صورت جداگانه نشان می‌دهد. سطر سوم ویژگی‌های استخراج شده از دو صفحه مرتبط را به تصویر می‌کشد.



شکل (۷): ویژگی مقاوم تسریع یافته. تصویر سمت چپ، تصویر اصلی و تصویر سمت راست استخراج ده نقطه موردهلاقه از تصویر را نشان می‌دهد.

۲-۳ الگوریتم تبدیل شباهت

برای تخمین شباهت در هر فریم، فرآیند تبدیل با چهار درجه آزادی با پارامتر (τ) شامل انتقال (t_x, t_y) ، چرخش (θ) ، مقیاس (s) ، با در نظر گرفتن فریم قبلی باید پیش‌بینی شود. همان‌طور که در [۱۰] ذکر شد، یک مسئله با چهار درجه آزادی به دو زیر مسئله با دو درجه آزادی تفکیک شده است و دو بلوک شامل $t =$

جدول (۲): توصیف ویژگی‌های مجموعه داده OTB-۲۰۱۵

| IV | تغییر روشنایی | روشنایی شی هدف به دلیل موقعیتی و یا موقعیت دوربین تغییر می‌کند. |
|-----|---------------|--|
| SV | تغییر مقیاس | نسبت قاب اطرفشی در اولین فریم به فریم فعلی خارج از محدوده آستانه مقیاس قرار بگیرد. (حد آستانه مقیاس = ۲) |
| OCC | هم‌پوشانی | شی هدف به صورت کلی و یا جزئی پوشانده شود. |
| DEF | تغییر شکل | تغییر شکل اجسام غیر صلب |



| | | |
|------------|-------------------|--|
| MB | تاری حرکت | ناحیه هدف در نتیجه حرکتی هدف یا دوربین تار می‌شود. |
| FM | حرکت سریع | حرکتی بیش از پیکسل‌های حد آستانه حرکت (حد آستانه حرکت = ۲۰) |
| IPR | چرخش در صفحه | چرخشی در راستای صفحه رخ می‌دهد. |
| OPR | چرخش خارج از صفحه | چرخشی خارج از صفحه رخ می‌دهد. |
| OV | خارج از دید | بخشی از شی خارج از دید دوربین قرار می‌گیرد |
| BC | شلوغی پس‌زمینه | پس‌زمینه نزدیک به شی هدف دارای رنگ و یا بافت مشابه به هدف است. |
| LR | رزولوشن پایین | تعداد پیکسل‌هایی که در قاب هدف قرار گرفته‌اند کمتر از حد آستانه رزولوشن است. (حد آستانه رزولوشن = ۴۰۰) |

جدول (۳): دقت با حد آستانه ۲۰ پیکسل. بهترین نتایج قرمز رنگ

و دومین بهترین زیر خط دار هستند.

| | رو ش اول | LD ES [2] | Siam fc [3] | LCTd eep [4] | ACF N [5] | DCF Net [۲۵] |
|-----------------|------------------|------------------|-------------------|--------------------|-----------------|--------------------|
| IV | $\frac{.71}{16}$ | $\frac{0.70}{7}$ | ۲۶.۶۹ | ۵۵.۷۳ | ۹۵.۶۹ | ۳۶.۶۷ |
| SV | $\frac{.74}{07}$ | $\frac{3.71}{9}$ | ۵۰.۶۹ | ۸۴.۶۸ | ۳۰.۶۸ | ۸۰.۶۷ |
| OC C | $\frac{.73}{71}$ | $\frac{1.71}{8}$ | ۸۹.۶۶ | ۸۴.۶۶ | ۴۳.۶۶ | ۸۷.۷۰ |
| DE F | $\frac{.65}{76}$ | ۴.۵۸ | ۵۶.۶۴ | ۶۵.۷۰ | ۳۱.۶۵ | ۱۸.۶۳ |
| MB | $\frac{.66}{27}$ | $\frac{0.65}{9}$ | ۱۱.۶۴ | ۲۴.۶۸ | ۱۰.۶۴ | ۱۸.۶۰ |
| FM | $\frac{.76}{32}$ | $\frac{4.75}{4}$ | ۸۳.۷۰ | ۲۳.۷۰ | ۱۳.۶۴ | ۰۲.۶۲ |
| IP R | $\frac{.73}{26}$ | $\frac{0.71}{3}$ | ۳۴.۶۸ | ۵۳.۷۱ | ۷۵.۶۷ | ۲۹.۶۴ |
| OP R | $\frac{.75}{44}$ | $\frac{1.71}{2}$ | ۳۵.۷۰ | ۴۹.۷۳ | ۷۰.۶۷ | ۳۱.۷۰ |
| OV | $\frac{.71}{94}$ | $\frac{2.69}{9}$ | ۵۷.۶۱ | ۱۰.۵۱ | ۳۶.۵۵ | ۲۲.۶۰ |
| BC | $\frac{.77}{35}$ | $\frac{0.75}{9}$ | ۴.۶۸ | ۰۷.۷۴ | ۱۵.۷۳ | ۲۷.۷۱ |

۴- نتایج و آزمایش‌ها

روش‌های پیشنهادی در متلب ۲۰۱۹ پیاده‌سازی شده و آزمایش‌ها بر روی رایانه معمولی با پردازنده ۸ هسته‌ای و ۳۲ گیگابایت RAM انجام شده است. مجموعه داده مورد ارزیابی OTB-2015 شامل ۱۰۰ دنباله ویدیویی با ویژگی‌های چالش‌برانگیز متفاوت بوده که تفسیر این ویژگی‌ها در جدول (۲) آورده شده است.

در این تحقیق، معیارهای دقت^۲ و موفقیت^۳ برای مقایسه نتایج ما با ردیاب‌های قبلی اعمال شده است. دقت یکی از معیارهای رایج در ردیابی بصری است. دقت، متوسط فاصله اقلیدسی بین مکان پیش‌بینی شده و مقدار واقعی شی هدف را محاسبه می‌کند. این فاصله در تمام فریم‌های هر دنباله ویدیویی محاسبه می‌شود تا عملکرد کلی آن دنباله برآورد گردد. به‌طور کلی، نمودار دقت برای ارزیابی عملکرد ردیاب بررسی می‌شود. نمودار دقت، درصد فریم‌هایی که فاصله با مکان واقعی در محدوده حد آستانه است را نمایش می‌دهد. برای ارزیابی دقت هر ردیاب، مقدار حد آستانه، ۲۰ پیکسل در جداول (۲ و ۳) نشان داده شده‌اند. در جدول (۳)، دقت ردیاب پیشنهادی با ترکیب ویژگی روش اول با ۵ ردیاب پیشرفته مقایسه شده است.

³ Success

¹http://cvlab.hanyang.ac.kr/tracker_benchmark/datasets.htm

² Precision

گزارش شده است. در این شکل‌ها سطح زیر منحنی برای نمودار موفقیت و دقت با آستانه ۲۰ پیکسل نشان داده شده است.

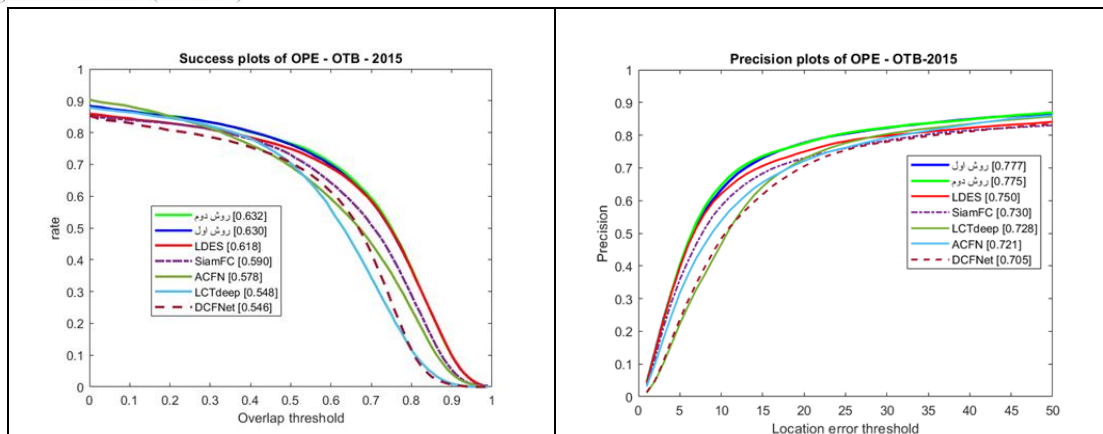
جدول (۴): دقت با حد آستانه ۲۰ پیکسل. بهترین نتایج قرمز رنگ و دومین بهترین زیرخط دار هستند.

| | روش دوم | LD ES [۱۰] | Siam fc [۲۲] | LCTd eep [۲۴] | ACF N [۲۳] | DCF Net [۲۵] |
|----------------|-----------------|------------------|-------------------|-------------------|------------|--------------|
| IV | $\frac{۷۱}{۶۶}$ | $\frac{۷۰}{۷}$ | ۶۹.۲۶ | ۷۳.۵۵ | ۶۹.۹۵ | ۶۷.۳۶ |
| SV | $\frac{۷۳}{۱۸}$ | $\frac{۷۱.۳}{۹}$ | ۶۹.۵۰ | ۶۸.۸۴ | ۶۸.۳۰ | ۶۷.۸۰ |
| OC C | $\frac{۷۳}{۵۶}$ | $\frac{۷۱.۱}{۸}$ | ۶۶.۸۹ | ۶۶.۸۴ | ۶۶.۴۳ | ۷۰.۸۷ |
| DEF | $\frac{۶۴}{۴۰}$ | ۵۸.۴ | $\frac{۶۴.۵۶}{۸}$ | ۷۰.۶۵ | ۶۵.۳۱ | ۶۳.۱۸ |
| MB | $\frac{۶۶}{۶۲}$ | $\frac{۶۵.۵}{۹}$ | ۶۴.۱۱ | ۶۸.۲۴ | ۶۴.۱۰ | ۶۰.۱۸ |
| FM | $\frac{۷۶}{۳۷}$ | $\frac{۷۵.۴}{۴}$ | ۷۰.۸۳ | ۷۰.۲۳ | ۶۴.۱۳ | ۶۲.۰۲ |
| IPR | $\frac{۷۴}{۱۹}$ | $\frac{۷۱.۰}{۳}$ | ۶۸.۳۴ | $\frac{۷۱.۵۳}{۳}$ | ۶۷.۷۵ | ۶۴.۲۹ |
| OP R | $\frac{۷۵}{۳۶}$ | $\frac{۷۱.۱}{۲}$ | ۷۰.۳۵ | $\frac{۷۳.۴۹}{۲}$ | ۶۷.۷۰ | ۷۰.۳۱ |
| OV | $\frac{۷۰}{۴۸}$ | $\frac{۶۹.۲}{۹}$ | ۶۱.۵۷ | ۵۱.۱۰ | ۵۵.۳۶ | ۶۰.۲۲ |
| BC | $\frac{۷۵}{۱۲}$ | $\frac{۷۵.۰}{۹}$ | ۶۸.۴ | ۷۴.۰۷ | ۷۳.۱۵ | ۷۱.۲۷ |
| LR | $\frac{۸۷}{۴۷}$ | $\frac{۸۳.۷}{۵}$ | ۹۰.۰۰ | ۸۱.۶۴ | ۷۹.۸۱ | ۶۹.۹۲ |
| تمام دنباله ها | $\frac{۷۷}{۵۴}$ | $\frac{۷۴.۹}{۶}$ | ۷۲.۹۹ | ۷۲.۸۴ | ۷۲.۱۴ | ۷۰.۵۰ |

| | | | | | | |
|----------------|-----------------|------------------|-------------------|-------|-------|-------|
| LR | $\frac{۹۰}{۶۲}$ | $\frac{۷۸۳}{۵}$ | $\frac{۰۰.۹۰}{۰}$ | ۶۴.۸۱ | ۸۱.۷۹ | ۹۲.۶۹ |
| تمام دنباله ها | $\frac{۷۷}{۶۶}$ | $\frac{۹.۷۴}{۶}$ | ۹۹.۷۲ | ۸۴.۷۲ | ۱۴.۷۲ | ۵۰.۷۰ |

در ۸ ویژگی از ۱۱ ویژگی نتیجه ردیاب پیشنهادی بهتر از سایر ردیاب‌ها بوده است و دقت مقاله پایه را در ویژگی رزولوشن پایین حدود ۷ درصد بهبود یافته است. در جدول (۴)، نتایج ردیاب پیشنهادی با ترکیب ویژگی‌های روش دوم ارائه شده است. بهترین نتایج بارنگ قرمز، دومین بهترین نتایج دارای زیرخط هستند. در ۷ ویژگی از ۱۱ ویژگی، نتیجه این تحقیق بهتر از سایر ردیاب‌ها است. با توجه به نتایج، ترکیب اطلاعات بافتی و نقاط کلیدی، ویژگی‌های ارزشمندی را در برابر چالش‌های چرخش در صفحه و حرکت سریع ارائه داده است. تغییر مقیاس تصویر قادر است نتایج را در مواجهه با هم‌پوشانی و خارج از دید بهبود بخشد. جدول (۵)، معیار دقت را برای هر دنباله ویدیویی دارای چالش خارج از دید به تفکیک نشان می‌دهد. این دنباله‌ها دارای چالش‌های دیگری علاوه بر خارج از دید هستند ولی به‌طور مشترک نیز این چالش را دارند. بهترین نتایج بارنگ قرمز و دومین بهترین با زیرخط مشخص شده‌اند. یکی دیگر از معیارهای ارزیابی که نمودار موفقیت نامیده می‌شود، هم‌پوشانی قاب است. فرض می‌شود قاب اصلی شی r_a است و قاب تخمین زده شده است. معیار هم‌پوشانی با استفاده از رابطه $IoU = \frac{|r_t \cap r_a|}{|r_t \cup r_a|}$ تعریف می‌شود که \cap و \cup به ترتیب اشتراک و اجتماع قاب‌های تصویر را نشان می‌دهند و $| \cdot |$ به معنی هم‌پوشانی بیش‌ازحد آستانه داده شده باشد، محاسبه می‌شود و کارایی دنباله ویدیویی ارزیابی می‌شود. نمودار موفقیت نسبت فریم‌های موفق را در آستانه‌های بین ۰ تا ۱ نشان می‌دهد.

برای رتبه‌بندی ردیاب‌ها، ناحیه زیر منحنی (AUC) هر نمودار موفقیت محاسبه می‌شود. در شکل (۸)، نمودارهای دقت و موفقیت



شکل (۸): نمودارهای دقت و موفقیت کل مجموعه داده

جدول (۵): دقت با حد آستانه ۲۰ پیکسل در دنباله‌های ویدیویی دارای چالش خارج از دید.

بهترین نتایج قرمز رنگ و دومین بهترین زیرخط دار هستند.

| | روش اول | روش دوم | LDES [۱۰] | Siamfc [۲۲] | LCTdeep [۲۴] | ACFN [۲۳] | DCFNet [۲۵] |
|---------------|---------|---------|-----------|-------------|--------------|-----------|-------------|
| Biker | ۵۰.۷۰ | ۵۰ | ۵۳.۵۲ | ۱۰۰ | ۴۵.۷۷ | ۴۸.۵۹ | ۴۴.۳۷ |
| Bird1 | ۵۵.۸ | ۳۶.۷۶ | ۳۴.۵۶ | ۳۷.۵ | ۳۲.۸۴ | ۸.۸۲ | ۶.۱۳ |
| Board | ۷۶.۰۷ | ۶۷.۴۸ | ۶۷.۶۲ | ۹.۶۰ | ۳۱.۸۱ | ۴۰.۵۴ | ۶۵.۹۰ |
| Box | ۴۱.۰۶ | ۴۱.۱۷ | ۳۹.۱۰ | ۲۶.۳۶ | ۲۱.۲۷ | ۳۷.۶۴ | ۳۹.۱۰ |
| ClifBar | ۹۹.۳۶ | ۱۰۰ | ۹۹.۷۹ | ۴۱.۵۳ | ۵۱.۲۷ | ۹۵.۱۴ | ۳۸.۷۷ |
| DragonBaby | ۸۹.۳۸ | ۹۲.۹۲ | ۹۲.۹۲ | ۲۹.۲۰ | ۷۸.۷۶ | ۴۳.۳۶ | ۶۳.۷۲ |
| Dudek | ۹۱.۱۸ | ۹۱ | ۹۱.۴۴ | ۹۶.۷۷ | ۶۹.۲۶ | ۷۴.۰۶ | ۸۱.۸۳ |
| Human6 | ۹۹.۱۲ | ۹۹.۱۲ | ۹۸.۹۹ | ۹۵.۳۳ | ۲۸.۶۶ | ۳۳.۸۴ | ۴۳.۸۱ |
| Ironman | ۲.۴۱ | ۹.۰۴ | ۷.۲۳ | ۶۰.۸۴ | ۶۳.۸۶ | ۶.۶۳ | ۶۰.۲۴ |
| Lemming | ۹۸.۸۸ | ۹۹.۳۳ | ۹۸.۹۵ | ۸۵.۴۰ | ۲۱.۷۱ | ۸۱.۹۶ | ۸۴.۶۶ |
| Liquor | ۶۶.۴۶ | ۶۸.۶۴ | ۶۸.۴۱ | ۲۵.۱۰ | ۷۷.۳۷ | ۸۹.۶۰ | ۸۱.۷۹ |
| Panda | ۸۲.۲۰ | ۸۰.۳۰ | ۶۱.۱۰ | ۱۰۰ | ۸۰.۳۰ | ۵۸.۵۰ | ۹۸.۱۰ |
| Suv | ۹۷.۷۸ | ۹۷.۷۸ | ۹۷.۷۸ | ۹۲.۳۸ | ۴۵.۶۱ | ۹۷.۴۶ | ۸۳.۷۰ |
| Tiger2 | ۵۶.۱۶ | ۵۳.۱۵ | ۵۸.۶۳ | ۶۱.۹۲ | ۶۶.۸۵ | ۵۸.۹۰ | ۵۰.۹۶ |
| همه دنباله‌ها | ۷۱.۹۴ | ۷۰.۴۸ | ۶۹.۲۹ | ۶۱.۵۷ | ۵۱.۱۰ | ۵۵.۳۶ | ۶۰.۲۲ |

حداقل واریانس مقیاس بندی شدند. سپس این ویژگی به ۸ سطح خاکستری نگاشت شد. به همراه هیستوگرام گرادین‌های جهت‌دار در ردیابی استفاده شد. نتایج تجربی بهبود را در چالش‌ها به‌ویژه در چالش با وضوح پایین نشان می‌دهد. در رویکرد دوم، الگوی

۵- نتیجه‌گیری و کارهای آینده

در این مقاله، دو روش مختلف استخراج ویژگی در فضای دکارتی برای ردیابی شی بصری بر اساس تبدیل شباهت استخراج شد. در رویکرد اول، بخش‌های تصویر با استفاده از روش محدودسازی

دلیل ایجاد تمایز در اشیاء مختلف در تصویر به جهت فاصله‌های متفاوت از دوربین، به ردیابی موفق‌تر منجر شود. در رویکردی دیگر می‌خواهیم از نقاط کلیدی استخراج‌شده در تصویر الگو و بخش مورد جستجو استفاده کنیم و معیار شباهت متفاوتی را برای تخمین موقعیت‌های تعریف کنیم.

دودویی رنگی محلی متضاد، ویژگی مقاوم تسریع یافته و هیستوگرام گرادینان‌های جهت‌دار ترکیب شدند. نتایج نشان می‌دهد که اطلاعات بافتی تصاویر رنگی مربوط به صفحات، نقاط کلیدی مهم و اطلاعات لبه‌ها می‌تواند عملکرد ردیابی را به‌ویژه در ویژگی‌های انسداد و خارج از دید افزایش دهد. این رویکردها بر روی یک سیستم معمولی و بدون نیاز به شبکه‌های از پیش آموزش‌دیده آزمایش شدند.

در کارهای آینده قصد داریم بر ویژگی‌های استخراج‌شده از تخمین عمق استفاده می‌کنیم. به نظر می‌رسد ویژگی‌های تخمین عمق به

- Consistent Low-Rank Sparse Learning,” *Int. J. Comput. Vis.*, vol. 111, no. 2, pp. 171–190, 2015, doi: 10.1007/s11263-014-0738-0.
- [8] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, “Visual object tracking using adaptive correlation filters,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2544–2550, doi: 10.1109/CVPR.2010.5539960.
- [9] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, “Discriminative Scale Space Tracking,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 8, pp. 1561–1575, Aug. 2017, doi: 10.1109/TPAMI.2016.2609928.
- [10] Y. Li, J. Zhu, W. Song, Z. Wang, H. Liu, and S. C. H. Hoi, “Robust Estimation of Similarity Transformation for Visual Object Tracking with Correlation Filters,” *Aaai*, 2019, [Online]. Available: <http://arxiv.org/abs/1712.05231>.
- [11] T. Surasak, I. Takahiro, C. H. Cheng, C. E. Wang, and P. Y. Sheng, “Histogram of oriented gradients for human detection in video,” *Proc. 2018 5th Int. Conf. Bus. Ind. Res. Smart Technol. Next Gener. Information, Eng. Bus. Soc. Sci. ICBIR 2018*, pp. 172–176, Jun. 2018, doi: 10.1109/ICBIR.2018.8391187.
- [12] S. W. Thomas, “Efficient inverse color map computation,” *Graph. Gems II*, pp. 116–125, Jan. 1991, doi: 10.1016/B978-0-08-050754-5.50034-7.
- [13] M. Milosevic, D. Jankovic, and A. Peulic, “Thermography based breast cancer detection using texture features and minimum variance quantization,” *EXCLI J.*, vol. 13, pp. 1204–1215, 2014, doi: 10.17877/DE290R-7338.

References

- [1] D. Yuan, X. Li, Z. He, Q. Liu, S. L.-K.-B. Systems, and undefined 2020, “Visual object tracking with adaptive structural convolutional network,” Elsevier, Accessed: Jul. 19, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0950705120300472>.
- [2] X. Li, W. Hu, C. Shen, Z. Zhang, A. Dick, and A. Van Den Hengel, “A survey of appearance models in visual object tracking,” *ACM Trans. Intell. Syst. Technol.*, vol. 4, no. 4, pp. 1–48, Sep. 2013, doi: 10.1145/2508037.2508039.
- [3] M. Fiaz, A. Mahmood, S. Javed, and S. K. Jung, “Handcrafted and deep trackers: Recent visual object tracking approaches and trends,” *ACM Comput. Surv.*, vol. 52, no. 2, 2019, doi: 10.1145/3309665.
- [4] T. Wang and H. Ling, “Gracker: A Graph-Based Planar Object Tracker,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 6, pp. 1494–1501, Jun. 2018, doi: 10.1109/TPAMI.2017.2716350.
- [5] T. Zhang et al., “Structural Sparse Tracking,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 150–158, 2015, doi: 10.1109/CVPR.2015.7298610.
- [6] D. Du, H. Qi, W. Li, L. Wen, Q. Huang, and S. Lyu, “Online Deformable Object Tracking Based on Structure-Aware Hyper-Graph,” *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3572–3584, 2016, doi: 10.1109/TIP.2016.2570556.
- [7] T. Zhang, S. Liu, N. Ahuja, M. H. Yang, and B. Ghanem, “Robust Visual Tracking Via



- Comput. Vis.*, vol. 126, no. 8, pp. 771–796, 2018, doi: 10.1007/s11263-018-1076-4.
- [25] Q. Wang, J. Gao, J. Xing, M. Zhang, and W. Hu, “DCFNET: Discriminant correlation filters network for visual tracking,” arXiv. 2017, Accessed: Jun. 22, 2020. [Online]. Available: <https://github.com/foolwood/DCFNet>.
- [26] “Opposite Color Local Binary Patterns (OC-LBP) - File Exchange - MATLAB Central.” <https://de.mathworks.com/matlabcentral/fileexchange/44284-opposite-color-local-binary-patterns-oc-lbp> (accessed Aug. 20, 2021).
- [27] M. Sotoodeh, M. R. Moosavi, and R. Boostani, “A structural based feature extraction for detecting the relation of hidden substructures in coral reef images,” *Multimed. Tools Appl.*, vol. 78, no. 24, pp. 34513–34539, 2019, doi: 10.1007/s11042-019-08050-w.
- [28] M. Sotoodeh, M. R. Moosavi, and R. Boostani, “A novel adaptive LBP-based descriptor for color image retrieval,” *Expert Syst. Appl.*, vol. 127, pp. 342–352, Aug. 2019, doi: 10.1016/j.eswa.2019.03.020.
- [29] S. K. Jena, S. Chakraverty, and M. Malikan, “Implementation of Haar wavelet, higher order Haar wavelet, and differential quadrature methods on buckling response of strain gradient nonlocal beam embedded in an elastic medium,” *Eng. Comput.*, vol. 37, no. 2, pp. 1251–1264, 2021, doi: 10.1007/s00366-019-00883-1.
- [14] N. Varish and A. K. Pal, “A novel image retrieval scheme using gray level co-occurrence matrix descriptors of discrete cosine transform based residual image,” *Appl. Intell.*, vol. 48, no. 9, pp. 2930–2953, Jan. 2018, doi: 10.1007/s10489-017-1125-7.
- [15] T. Mäenpää, M. Pietikäinen, and J. Viertola, “Separating color and pattern information for color texture discrimination,” *Proc. - Int. Conf. Pattern Recognit.*, vol. 16, no. 1, pp. 668–671, 2002, doi: 10.1109/icpr.2002.1044840.
- [16] J. Koh, E. Ng, S. Bhandary, ... A. L.-A., and undefined 2018, “Automated detection of retinal health using PHOG and SURF features extracted from fundus images,” Springer, vol. 48, no. 5, pp. 1379–1393, May 2018, doi: 10.1007/s10489-017-1048-3.
- [17] S. You, H. Zhu, M. Li, and Y. Li, “A Review of Visual Trackers and Analysis of its Application to Mobile Robot,” 2019, [Online]. Available: <http://arxiv.org/abs/1910.09761>.
- [18] M. Danelljan, F. S. Khan, M. Felsberg, and J. Van De Weijer, “Adaptive color attributes for real-time visual tracking,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1090–1097, doi: 10.1109/CVPR.2014.143.
- [19] X. Li, Q. Liu, Z. He, H. Wang, C. Zhang, and W. S. Chen, “A multi-view model for visual tracking via correlation filters,” *Knowledge-Based Syst.*, vol. 113, pp. 88–99, 2016, doi: 10.1016/j.knsys.2016.09.014.
- [20] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, “High-speed tracking with kernelized correlation filters,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 583–596, 2015, doi: 10.1109/TPAMI.2014.2345390.
- [21] C. Gao, F. Chen, J. G. Yu, R. Huang, and N. Sang, “Robust Visual Tracking Using Exemplar-Based Detectors,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 2, pp. 300–312, 2017, doi: 10.1109/TCSVT.2015.2513700.
- [22] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. S. Torr, “Fully-convolutional siamese networks for object tracking,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 9914 LNCS, pp. 850–865, doi: 10.1007/978-3-319-48881-3_56.
- [23] J. Choi, H. J. Chang, S. Yun, T. Fischer, Y. Demiris, and J. Y. Choi, “Attentional correlation filter network for adaptive visual tracking,” in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-Janua, pp. 4828–4837, doi: 10.1109/CVPR.2017.513.
- [24] C. Ma, J. Bin Huang, X. Yang, and M. H. Yang, “Adaptive Correlation Filters with Long-Term and Short-Term Memory for Object Tracking,” *Int. J.*

The Effects of Textural Information and Key Points Extraction on Visual Object Tracking, Based on Similarity Transformation

Solmaz Abbasi¹, Mehdi Rezaeian^{2*}

¹ Ph.D Candidate, Computer Engineering Department, Yazd University, Yazd

² Associate Professor, Computer Engineering Department, Yazd University, Yazd

Article Information

Original Research Paper

Received:

25 October 2021

Accepted:

28 March 2022

Keywords:

Visual Object Tracking,
Similarity Transformation,
Minimum Variance
Quantization, Opposite Color
Local Binary Patterns, Speeded-
Up Robust Features

Corresponding Author*:

mrezaeian@yazd.ac.ir

Abstract

Visual object tracking in arbitrary environments with arbitrary objects has gained considerable importance in recent years. A very significant feature, which makes a tracker useful, is real time tracking without needing GPU and pre-train algorithms. In the recent decade, the trackers, which function on the basis of discriminative correlation filters, have promised positive results in terms of both speed and accuracy. Although, in most of such methods, the estimation of the position of the object in each frame is computed based on transformation and pyramid scales, in Large Displacement Estimation of Similarity transformation algorithm, translation, scale and rotation are estimated in each frame. In this paper, the Histogram of Oriented Gradient is considered as feature extraction. Here, we adopt two different approaches. The first approach uses scaled images as a feature matrix by applying minimum variance quantization. The second approach, uses a combination of opposite color local binary patterns and Speeded-Up Robust Features. By using these two methods, we are able to extract helpful and fast features, and therefore improve the results of tracking against challenging attributes. The OTB-2015 dataset is utilized for evaluating tracker. The results show precision of trackers improve 3%. Additionally, the first tracker increase the result about 7% against low resolution and the second one can be helpful about 4% in facing rotation challenge.