



Optimal, Low-Latency, and Robust Vehicle Tracking in Aerial UAV Videos

Mohammad Javad Rezaei^{1*}

1. Ph.D. Electrical and Electronics Engineering, Iran University of Science and Technology, Tehran,
Iran

Key words:

*Tracking
Aerial Image
Background
Foreground*

Abstract

Object tracking in video plays a fundamental role in intelligent analysis systems, video surveillance, and traffic management. In this context, real-time vehicle tracking in aerial video data acquired by unmanned aerial vehicles (UAVs) involves several technical challenges, including low frame rates, limited resolution, motion blur, and continuous camera motion. Conventional tracking methods are generally based on model updating using previous frames, which may face limitations in hardware implementations. In this study, an agile and robust vehicle-tracking method for aerial videos is proposed, enabling real-time implementation on FPGA hardware through algorithm parallelization. The proposed approach improves both tracking accuracy and stability by employing background modeling and integrating it with the object feature-space representation. Experimental evaluations conducted on the standard CLIF 2007 dataset demonstrate the efficiency and effectiveness of the proposed method under practical operating conditions

* Corresponding author's E-mail: mj_rezaei@ssau.ac.ir



ردیابی بهینه با تأخیر کم و پایداری بالای خودرو در تصاویر هوایی پرنده‌های هدایت‌پذیر از راه دور (پهپادها)

محمدجواد رضائی^{*}

۱. دکتری مهندسی برق و الکترونیک، دانشگاه علم و صنعت ایران، تهران، ایران

واژگان کلیدی

ردیابی

تصویر هوایی

پس زمینه

تصویر پیش زمینه

چکیده

ردیابی اشیاء در ویدئو، نقشی اساسی در سامانه‌های تحلیل هوشمند، نظارت تصویری و مدیریت ترافیک ایفا می‌کند. در این میان، ردیابی بلادرنگ خودرو در داده‌های ویدئویی هوایی (پهپاد-پایه) به دلیل چالش‌های ذاتی نظیر نرخ فریم پایین، رزولوشن محدود، تاری حرکت (Motion Blur) و حرکت مداوم دوربین، با پیچیدگی‌های فنی متعددی همراه است. روش‌های مرسوم ردیابی، عمدتاً بر به‌روزرسانی مدل‌ها بر اساس فریم‌های پیشین استوارند که در پیاده‌سازی‌های سخت‌افزاری با محدودیت مواجه می‌شوند. در این پژوهش، یک روش ردیابی چابک و پایدار برای خودروها در ویدئوهای هوایی ارائه شده است که قابلیت پیاده‌سازی بلادرنگ بر روی سخت‌افزارهای FPGA را از طریق موازی‌سازی الگوریتم فراهم می‌آورد. رویکرد پیشنهادی با بهره‌گیری از مدل‌سازی پس‌زمینه و تلفیق آن با ساختار فضای ویژگی‌های، دقت و پایداری ردیابی را به طور همزمان ارتقا می‌دهد. نتایج ارزیابی الگوریتم بر روی مجموعه داده‌گان استاندارد CLIF ۲۰۰۷، حاکی از کارایی و اثربخشی روش پیشنهادی در شرایط عملیاتی است.

۱- مقدمه

با وجود این، اجرای الگوریتم‌های پردازش تصویر و بینایی ماشین برای شناسایی و ردیابی اهداف متحرک، به حجم محاسباتی قابل توجهی نیاز دارد و می‌تواند موجب افزایش زمان پردازش شود. از سوی دیگر، ارسال تصاویر ویدئویی به یک ایستگاه زمینی برای انجام پردازش، همواره راهکاری عملی و قابل اعتماد نیست؛ زیرا کیفیت ویدئوی دریافت‌شده به عواملی مانند پهنای باند، سرعت، تأخیر و پایداری کانال ارتباطی بی‌سیم وابسته است. این محدودیت‌ها در شرایطی مانند عملیات جست‌وجو و نجات یا مأموریت‌های نظارتی، که تصمیم‌گیری سریع اهمیت زیادی دارد، می‌تواند عملکرد سامانه را تحت تأثیر قرار دهد. در چنین شرایطی، استفاده از پهپادهای خودکار مجهز به قابلیت پردازش روی‌برد، راهکاری مطلوب به شمار می‌رود. چنین سامانه‌ای باید بتواند بدون

هواپیماهای بدون سرنشین، موسوم به UAV یا پهپاد، یکی از مهم‌ترین سامانه‌های تصویربرداری هوایی به شمار می‌روند. قابلیت پرواز، جابه‌جایی سریع و تصویربرداری از نواحی مختلف، این سامانه‌ها را به گزینه‌ای مناسب برای پایش و نظارت بر محیط‌های گوناگون تبدیل کرده است. پردازش تصاویر ویدئویی حاصل از پهپادها می‌تواند کاربردهای متعددی داشته باشد که از جمله آن‌ها می‌توان به شناسایی و ردیابی اشیاء، پایش ترافیک، جست‌وجو و نجات، نظارت بر مناطق آسیب‌دیده و کاربردهای نظامی و امنیتی اشاره کرد.

* ایمیل نویسنده پاسخگو: mj_rezaei@ssau.ac.ir

خودروها ممکن است در کنار خیابان، زیر سایه درختان یا ساختمانها حرکت کنند و بخشی از آنها توسط موانع محیطی پوشانده شود. در چنین شرایطی، ویژگیهای ظاهری خودرو به طور کامل در دسترس نیست و ردیاب ممکن است هدف را از دست بدهد یا آن را با شیء دیگری اشتباه بگیرد.

تمرکز این پژوهش بر ردیابی خودرو در ویدئوهای هوایی با نرخ فریم پایین است؛ به ویژه در شرایطی که تصاویر از ارتفاع زیاد ثبت شده اند و اهداف با انسداد، رزولوشن پایین و تغییرات ظاهری قابل توجه مواجه هستند. در این پژوهش، مجموعه داده CLIF ۲۰۰۷ به عنوان یکی از دادگان رایج در زمینه تحلیل ویدئوهای هوایی مورد بررسی قرار می گیرد [۱]. ریلی و همکاران در پژوهشی، مسئله شناسایی و ردیابی خودرو را با استفاده از مجموعه داده CLIF ۲۰۰۶ بررسی کرده اند. در این مجموعه داده، اهداف مورد نظر بسیار کوچک و در حد چند پیکسل هستند؛ از این رو، مدل سازی ظاهر خودرو دشوار بوده و تمرکز اصلی پژوهش بر اتصال خودروهای شناسایی شده در فریم های متوالی قرار گرفته است [۶]. با توجه به وجود انسدادهای متعدد در مجموعه داده CLIF ۲۰۰۷، روش مذکور لزوماً عملکرد مطلوبی در این مجموعه داده نخواهد داشت.

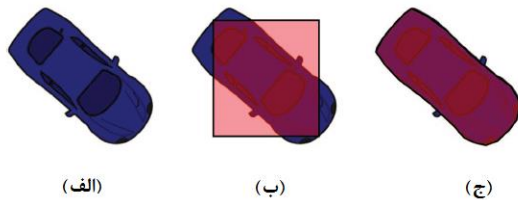
در پژوهشی دیگر، یک ردیاب مبتنی بر شار نوری برای مجموعه داده CLIF ۲۰۰۷ ارائه شده است [۷]. در این روش، ویژگیهای مناسب برای تخمین شار نوری و تطبیق هدف در فریم های متوالی انتخاب می شوند. با این حال، این روش نیز در شرایطی که خودرو دچار انسداد شده یا تنها بخشی از آن قابل مشاهده است، با کاهش عملکرد مواجه می شود و عمدتاً برای ردیابی خودروهای قابل مشاهده در فضاهای باز و بدون انسداد مناسب است. بر این اساس، در این مقاله روشی مبتنی بر ترکیب ردیابی و قطعه بندی برای شناسایی و دنبال کردن خودرو در شرایط چالش برانگیز ارائه می شود. روش پیشنهادی با هدف افزایش پایداری ردیابی در حضور انسداد، تغییرات ظاهری، نرخ فریم پایین و حرکت دوربین طراحی شده است. افزون بر دقت مناسب، ساختار الگوریتم به گونه ای در نظر گرفته شده است که امکان موزایی سازی و پیاده سازی بلادرنگ آن بر روی FPGA وجود داشته باشد. هدف نهایی، طراحی یک سامانه سخت افزاری کارآمد برای شناسایی و ردیابی اهداف متحرک در کاربردهای مبتنی بر پهپاد است. در ادامه، روش پیشنهادی و جزئیات پیاده سازی آن تشریح می شود.

۲- روش کار

اگرچه مسئله اصلی این پژوهش ردیابی خودرو است، مطالعات سال های اخیر نشان داده اند که ترکیب هم زمان **ردیابی و قطعه بندی** می تواند کارایی سامانه های ردیابی را به طور محسوس افزایش دهد. در روش های اولیه،

نیاز به مداخله مستمر انسان، اهداف مورد نظر را شناسایی و ردیابی کند. بنابراین، پهپاد خودکار به یک بستر پردازشی با توان محاسباتی مناسب، مصرف توان پایین، ابعاد کوچک و قابلیت پشتیبانی از وظایف مختلف نیاز دارد. آرایه های دروازه ای قابل برنامه ریزی میدانی یا FPGA، به دلیل برخورداری از قابلیت پردازش موازی، مصرف توان پایین، انعطاف پذیری و امکان مجتمع سازی، بستر مناسبی برای پیاده سازی الگوریتم های بینایی ماشین بلادرنگ در سامانه های متحرک فراهم می کنند. همچنین، ساختار FPGA را می توان متناسب با نوع مأموریت یا الگوریتم مورد استفاده، مجدداً پیکربندی کرد. یکی از مسائل مهم در پیاده سازی الگوریتم های ردیابی بر روی سامانه های سخت افزاری، ایجاد توازن میان سرعت پردازش و دقت ردیابی است. همچنین الزامات محیطی سامانه های هوایی مانند محدوده دمایی، مقاومت در برابر لرزش، قابلیت اطمینان و تأثیر آن بر طراحی الگوریتم نیز باید لحاظ گردد. الگوریتم انتخاب شده باید ضمن برخورداری از پیچیدگی محاسباتی مناسب و قابلیت پیاده سازی سخت افزاری، دقت قابل قبولی در شناسایی و دنبال کردن هدف داشته باشد. در غیر این صورت، افزایش سرعت پردازش ممکن است به قیمت کاهش دقت ردیابی حاصل شود. ردیابی اشیاء در ویدئوهای ثبت شده در محیط های باز و کنترل نشده، به طور کلی مسئله ای چالش برانگیز است. این چالش در تصاویر هوایی، به ویژه تصاویری که از ارتفاع زیاد و با نرخ فریم پایین ثبت می شوند، شدت بیشتری پیدا می کند. در بسیاری از ویدئوهای هوایی، نرخ فریم به یک یا دو فریم بر ثانیه محدود می شود. در نتیجه، هدف ممکن است در فاصله زمانی میان دو فریم جابه جایی قابل توجهی داشته باشد و ظاهر آن از یک فریم به فریم بعد تغییر کند. افزون بر این، رزولوشن پایین، تاری ناشی از حرکت، کوچک بودن اهداف و دشواری جداسازی آنها از پس زمینه، از دیگر محدودیت های مهم این نوع تصاویر هستند. همچنین، به دلیل حرکت مداوم پهپاد، دوربین نیز نسبت به محیط ثابت نیست و پس زمینه تصویر در طول توالی ویدئویی تغییر می کند؛ بنابراین، فرض ثابت بودن پس زمینه که در برخی روش های متداول به کار می رود، در این کاربرد قابل اتکا نیست. ردیابی خودرو در ویدئوهای هوایی ارسال شده از پهپادها، یکی از کاربردهای مهم تحلیل ویدئو در حوزه های نظامی، امنیتی و نظارتی است. با این حال، بخش قابل توجهی از پژوهش های پیشین در زمینه ردیابی ویدئوهای هوایی، بر تصاویر ثبت شده از ارتفاع پایین تمرکز داشته اند. این ویدئوها معمولاً نرخ فریم بالاتر و اهدافی با رزولوشن بیشتر دارند؛ در نتیجه، تغییرات ظاهری خودرو بین فریم های متوالی کمتر بوده و مسئله ردیابی با استفاده از ردیاب های عمومی، از جمله روش های مبتنی بر کرنل، با دشواری کمتری انجام می شود [۱-۵].

با وجود عملکرد مناسب این روش ها در شرایط ساده، بسیاری از آنها در مواجهه با انسداد هدف با محدودیت روبه رو می شوند. در کاربردهای واقعی،



شکل (۱): (الف): تصویر هدف، (ب) قطعه بندی توسط مستطیل محاطی، (ج): قطعه بندی دقیق.

در مجموعه داده استاندارد مورد استفاده در این پژوهش، به منظور ایجاد مدل دقیق‌تر از پس‌زمینه، فرایند کالیبراسیون تصویر با بهره‌گیری از روش‌های متداول انجام می‌شود [۱۲-۲۰].

۳- تولید مدل پس زمینه در توالی ویدئوهای هوایی

بسیاری از الگوریتم‌های ردیابی برای عملکرد صحیح به یک مدل مناسب از پس‌زمینه نیاز دارند. در این بخش، روش تولید مدل پس‌زمینه برای ناحیه‌ای کوچک در اطراف موقعیت موردنظر x_w بررسی می‌شود. برای خودروهایی که با سرعت زیاد در بزرگراه حرکت می‌کنند، می‌توان مدل پس‌زمینه را با استفاده از میانگین‌گیری پیکسلی در سه فریم متوالی تولید کرد [۲۱]. این روش در چنین شرایطی عملکرد مناسبی دارد؛ زیرا خودرو از یک فریم به فریم دیگر جابه‌جا می‌شود و معمولاً حداقل دو فریم از سه فریم انتخاب‌شده، اطلاعاتی از پس‌زمینه در محل هدف در اختیار قرار می‌دهند. باین‌حال، این روش برای خودروهایی که با سرعت پایین حرکت می‌کنند یا برای اهدافی که در تصویر ثابت هستند، کارایی کافی ندارد. برای مثال، ممکن است خودرویی پشت چراغ قرمز متوقف شده باشد. در این حالت، میانگین‌گیری از چند فریم متوالی همچنان شامل پیکسل‌های مربوط به خودرو خواهد بود و مدل حاصل، پس‌زمینه واقعی را نشان نخواهد داد. بنابراین، به روشی نیاز است که بتواند در حضور اهداف ثابت و متحرک، تخمین قابل‌اعتمادی از پس‌زمینه ارائه دهد.

روش پیشنهادی برای تولید مدل پس‌زمینه شامل دو مرحله اصلی است:

۱- استخراج نواحی متناظر از فریم جاری و فریم‌های پیشین، پیش‌بینی موقعیت هدف و ثبت آن

۲- جایگزین کردن پیکسل‌های مربوط به اهداف ثابت و متحرک در فریم جاری با استفاده از اطلاعات پس‌زمینه موجود در فریم جاری

قطعه‌بندی توالی ویدئویی بر مبنای نواحی دارای الگوی حرکتی مشابه انجام می‌شود [۸، ۹]. در مقابل، پژوهش‌های جدیدتر عمدتاً بر قطعه‌بندی مبتنی بر ظاهر و استفاده از مدل پس‌زمینه متمرکز شده‌اند [۱۰، ۱۱].

انجام هم‌زمان ردیابی و قطعه‌بندی دو مزیت اساسی دارد. نخست، استفاده از اطلاعات پس‌زمینه سبب می‌شود سامانه نسبت به تغییرات ظاهری هدف مقاوم‌تر باشد. به بیان دیگر، پیکسل‌هایی که با مدل پس‌زمینه مطابقت ندارند، حتی در صورت تغییر ظاهر هدف، می‌توانند به‌عنوان بخشی از هدف شناسایی شوند. این ویژگی در شرایطی مانند تغییر روشنایی، وجود سایه یا تغییرات ظاهری ناشی از حرکت هدف اهمیت زیادی دارد. برای مثال، سایه می‌تواند ظاهر خودرو را به‌طور قابل‌توجهی تغییر دهد و موجب کاهش عملکرد ردیاب‌های متکی بر مدل ظاهری ثابت شود.

دومین مزیت، امکان استخراج دقیق‌تر ناحیه هدف از پس‌زمینه است. در بسیاری از روش‌های رایج، هدف با یک مستطیل محاطی نمایش داده می‌شود. این نوع نمایش، هرچند ساده و مناسب برای پیاده‌سازی است، معمولاً مرز واقعی هدف را به‌درستی مشخص نمی‌کند. در این حالت، بخشی از پیکسل‌های پس‌زمینه که داخل مستطیل قرار دارند، به اشتباه در مدل هدف وارد می‌شوند؛ در مقابل، ممکن است بخشی از پیکسل‌های متعلق به هدف که خارج از مستطیل قرار گرفته‌اند، نادیده گرفته شوند. استفاده از قطعه‌بندی دقیق‌تر، امکان به‌روزرسانی مناسب‌تر مدل ظاهری هدف را فراهم کرده و دقت ردیابی را افزایش می‌دهد.

شکل ۱ تفاوت میان نمایش هدف با استفاده از مستطیل محاطی و قطعه‌بندی دقیق را نشان می‌دهد. در شکل ۱-الف، تصویر اصلی هدف نمایش داده شده است. شکل ۱-ب نمایش هدف بر اساس مستطیل محاطی و شکل ۱-ج نتیجه قطعه‌بندی دقیق هدف را ارائه می‌کند.

ورود پیکسل‌های پس‌زمینه به مدل هدف، باعث تغییر تدریجی و نامطلوب مدل ظاهری در طول ردیابی می‌شود که از آن با عنوان رانش مدل (Model Drift) یاد می‌شود. این پدیده می‌تواند دقت ردیابی را کاهش داده و حتی به گم‌شدن هدف منجر شود. همچنین، حذف بخشی از پیکسل‌های واقعی هدف از مدل آن، مقاومت ردیاب را در برابر انسدادهای جزئی کاهش می‌دهد. بنابراین، دستیابی به یک قطعه‌بندی دقیق و قابل‌اعتماد، نقشی مهم در حفظ پایداری و افزایش دقت ردیابی دارد.

و فریم های پیشین، بر اساس الگوریتم بازسازی تصویر (Inpainting) [۲۲].

مرحله اول: تعیین پیکسل های نامعتبر پس زمینه

اجرای مرحله نخست نسبتاً ساده است؛ زیرا اشیاء ثابت در تصویر تفاضلی میان دو فریم متوالی ظاهر نمی شوند و عمدتاً اشیاء متحرک در نتیجه تفاضل گیری شناسایی می شوند. برای تعیین پیکسل های مشکوک، ابتدا قدرمطلق اختلاف میان فریم جاری و فریم قبلی محاسبه شده و سپس با اعمال یک آستانه ساده، ماسک اولیه M به دست می آید. از آنجا که ممکن است برخی اهداف در فریم جاری و فریم قبلی تقریباً در یک موقعیت قرار داشته باشند، پیکسل های مربوط به هدف با استفاده از ناحیه بندی هدف در فریم قبلی نیز به ماسک M افزوده می شوند. این مرحله برای شناسایی اهداف ثابت یا اهدافی که با سرعت بسیار کم حرکت می کنند، ضروری است. در واقع، نتایج حاصل از ردیابی و ناحیه بندی هدف برای تکمیل ماسک و تولید مدل دقیق تر از پس زمینه به کار گرفته می شوند. پیکسل هایی که در ماسک M قرار ندارند، به عنوان پیکسل های معتبر پس زمینه در نظر گرفته شده و مقدار آن ها از فریم جاری برای تشکیل مدل پس زمینه استفاده می شود. در مقابل، پیکسل های موجود در M به عنوان پیکسل های مشکوک یا نامعتبر در نظر گرفته می شوند و باید در مرحله بعد طبقه بندی و بازسازی شوند.

مرحله دوم: بازسازی پیکسل های نامعتبر

در مرحله دوم، پیکسل های موجود در ماسک M با استفاده از الگوریتم Inpainting معرفی شده در مرجع [۲۲] طبقه بندی و جایگزین می شوند. در روش اصلی، بازسازی پیکسل ها به صورت تدریجی و بر اساس جهت هموارشدگی و اطلاعات پیکسل های مجاور انجام می شود. این فرایند تا زمانی ادامه می یابد که ناحیه موردنظر به طور کامل با مقادیر تخمینی پس زمینه جایگزین شود. چنین رویکردی، به ویژه در نواحی نسبتاً ثابت مانند سطح جاده، می تواند مدل دقیقی از پس زمینه ایجاد کند. برای هر پیکسل، الگوریتم مقدار تخمینی V_g را محاسبه می کند. این مقدار، تخمینی از مقدار واقعی پیکسل در پس زمینه است و می توان از آن برای تعیین وضعیت پیکسل در فریم جاری و فریم قبلی استفاده کرد. مقادیر V_1 و V_2 به ترتیب نشان دهنده مقدار یک پیکسل متناظر در دو فریم مورد استفاده هستند. با توجه به اطلاعات موجود در این دو فریم، برای هر مختصات پیکسلی چهار حالت ممکن در نظر گرفته می شود:

- ۱- پیکسل متناظر در فریم قبلی متعلق به پس زمینه باشد؛
- ۲- پیکسل متناظر در فریم جاری متعلق به پس زمینه باشد؛

- ۳- هر دو فریم شامل مقدار پس زمینه برای آن پیکسل باشند؛
- ۴- هیچ یک از دو فریم اطلاعات معتبر پس زمینه را در آن مختصات نداشته باشند.

روش پیشنهادی تلاش می کند هر پیکسل موجود در ماسک M را به یکی از این چهار گروه نسبت دهد. سپس، بر اساس گروه تعیین شده و مقایسه مقدار پیکسل با مقدار تخمینی V_g ، مقدار مناسب برای مدل پس زمینه انتخاب می شود. مقدار احتمالی V_i که شامل اطلاعات پس زمینه است، مطابق رابطه (۱) تعریف می شود:

$$p_i = \exp\left\{-\frac{(v_i - v_g)^2}{2\sigma^2}\right\} \times f_i \quad (1)$$

ک در آن f_i یک کسری از پیکسل های اطراف است که به عنوان پیکسل های پس زمینه تعریف شده می باشد و σ یک ثابت است که توسط کاربر تعریف می شود. در این مقاله $\sigma = 20$ در نظر گرفته شده است. بر اساس این مقدار پیکسل های خروجی طبق رابطه ۲ بدست می آیند:

$$v = \begin{cases} v_1, & (p_1 \geq T) \cap (p_2 < T) \\ v_2, & (p_1 < T) \cap (p_2 \geq T) \\ \frac{v_1 p_1 + v_2 p_2}{p_1 + p_2}, & (p_1 \geq T) \cap (p_2 \geq T) \\ v_g, & otherwise \end{cases} \quad (2)$$

که در آن T مقدار آستانه است که احتمال مورد نیاز برای تعلق پیکسل ها به عنوان پیکسل های پس زمینه را کنترل می کند. در این مقاله از $T=1/3$ در پیاده سازی استفاده شده است. به این ترتیب، پیکسل های متعلق به اهداف ثابت یا متحرک از مدل اولیه حذف شده و با مقادیر مناسب پس زمینه جایگزین می شوند. خروجی این فرایند، مدلی استاتیک و قابل اعتماد از ناحیه موردنظر است که می تواند در مراحل بعدی برای بهبود قطعه بندی و افزایش دقت ردیابی به کار گرفته شود. در شکل ۲، روش پیشنهادی در این مقاله با روش ارائه شده در مرجع [۲۲] مقایسه شده است.

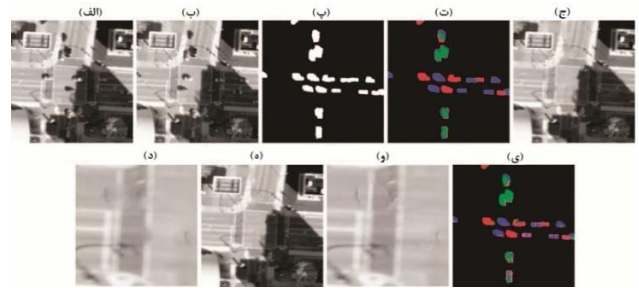
۴- ترکیب ردیابی و ناحیه بندی

راهبرد اصلی ردیابی در این پژوهش بر تلفیق فرایندهای ردیابی و ناحیه‌بندی استوار است که چارچوب آن در مرجع [۱۰] ارائه شده است. با در اختیار داشتن مجموعه‌ای از موقعیت‌های احتمالی اهداف، ابتدا برای هر پیکسل تصویر، یک تخصیص نرم به یک یا چند هدف و نیز پس‌زمینه انجام می‌شود. سپس، نسبت درست‌نمایی مشاهده هر پیکسل تحت مدل ظاهری هر هدف و مدل پس‌زمینه محاسبه می‌گردد. مقادیر حاصل برای تمام پیکسل‌ها ترکیب شده و یک امتیاز کلی برای هر مجموعه از موقعیت‌های احتمالی اهداف به دست می‌آید. این فرایند برای چندین فرضیه از موقعیت اهداف تکرار می‌شود. در نهایت، مجموعه موقعیت‌هایی که بیشترین مقدار درست‌نمایی یا امتیاز کلی را داشته باشد، به‌عنوان پاسخ نهایی ردیابی انتخاب می‌شود. نخستین مزیت تلفیق ردیابی و ناحیه‌بندی، فراهم کردن روشی اصولی برای ادغام هم‌زمان اطلاعات مربوط به هدف، پس‌زمینه و سایر اهداف موجود در صحنه است. برای مثال، سایه‌های ناشی از درختان، ساختمان‌ها یا سایر اجسام می‌توانند ظاهر هدف را به‌طور قابل توجهی تغییر دهند. در چنین شرایطی، مدل ظاهری مناسب برای هدف باید بتواند پیکسل‌های متعلق به پس‌زمینه یا اهداف دیگر را از پیکسل‌های واقعی هدف تفکیک کند. استفاده از اطلاعات ناحیه‌بندی در کنار ردیابی، این قابلیت را فراهم می‌سازد و از ورود اطلاعات نامرتبط به مدل ظاهری هدف جلوگیری می‌کند. مزیت دیگر این رویکرد، انعطاف‌پذیری بیشتر در نمایش شکل هدف نسبت به استفاده از جعبه‌های محاطی مستطیلی استاندارد است. نمایش مستطیلی معمولاً بخشی از پیکسل‌های پس‌زمینه را نیز دربر می‌گیرد و ممکن است تمام پیکسل‌های متعلق به هدف را پوشش ندهد. در مقابل، ناحیه‌بندی امکان تعیین دقیق‌تر مرزهای هدف را فراهم می‌کند. در نتیجه، پیکسل‌های پس‌زمینه با دقت بیشتری از مدل هدف حذف شده، خطای به‌روزرسانی مدل کاهش می‌یابد و پایداری ردیابی افزایش پیدا می‌کند.

از دیدگاه محاسباتی، هدف الگوریتم ردیابی، حل مسئله بهینه‌سازی زیر برای تعیین محتمل‌ترین موقعیت اهداف در هر فریم است:

$$X^{(k)} = \arg \max_{X \in F} P\{I|X\}P\{X|X^{(k-1)}\} \quad (3)$$

که x^k نشان گر موقعیت همه اهداف در فریم k است، I نشانگر تصویر ورودی و F مجموعه موقعیت‌های امکان پذیر هدف است. از F برای تحمیل نمودن راه حل به منظور معنا دار بودن فیزیکی استفاده می شود. یعنی هیچ هدفی نمی تواند با یکدیگر همپوشانی داشته باشند و بازه‌ی هر هدف محدود است.



شکل (۲): نتیجه مدلسازی پس زمینه با روش پیشنهادی در این مقاله و روش ارائه شده در مرجع [۲۲]. (الف) فریم جاری، (ب) فریم قبلی هم‌تراز شده، (پ) نیز ماسک اولیه M ، (ت) طبقه‌بندی نهایی پیکسل‌های جمع‌آوری شده، (ج) نتیجه بازسازی پس‌زمینه با استفاده از الگوریتم Inpainting، (د) جهت مقایسه جزئیات کوچک و بررسی دقت بازسازی در دو روش، (ه) نتیجه مدل پس‌زمینه حاصل از اعمال الگوریتم پیشنهادی، (و) و (ی) کلاس هر یک از پیکسل‌های کاندید در مجموعه T .

شکل ۲(ه) نتیجه مدل پس‌زمینه حاصل از اعمال الگوریتم پیشنهادی با استفاده از فریم جاری و فریم قبلی را نشان می‌دهد. شکل ۲(الف) فریم جاری و شکل ۲(ب) فریم قبلی هم‌تراز شده را نمایش می‌دهند. شکل ۲(پ) نیز ماسک اولیه M را نشان می‌دهد که بر اساس اطلاعات فریم‌های قبلی استخراج شده است. همچنین شکل ۲(ی) کلاس هر یک از پیکسل‌های کاندید در مجموعه T را مطابق رابطه (۲) مشخص می‌کند. در این نمایش، پیکسل‌های قرمز از فریم قبلی کپی شده‌اند، پیکسل‌های آبی از فریم جاری انتخاب شده‌اند، پیکسل‌های سبز مقادیر حفظ‌شده از الگوریتم Inpainting مرجع [۲۲] هستند و پیکسل‌های صورتی از میانگین وزن‌دار مقادیر متناظر در فریم‌های قبلی و جاری به‌دست آمده‌اند. در شکل ۲(ت)، طبقه‌بندی نهایی پیکسل‌های جمع‌آوری شده نمایش داده شده است. این طبقه‌بندی بر مبنای بیشترین میزان تطابق انجام می‌شود و مبنای بازسازی پس‌زمینه در روش پیشنهادی را تشکیل می‌دهد. به این ترتیب، پیکسل‌های متعلق به پس‌زمینه با دقت مناسبی از میان اطلاعات موجود در هر دو فریم شناسایی و انتخاب می‌شوند. برای ارزیابی کیفی عملکرد روش، شکل ۲(ج) نتیجه بازسازی پس‌زمینه با استفاده از الگوریتم Inpainting مرجع [۲۲] را نشان می‌دهد، در حالی که شکل ۲(ه) خروجی روش پیشنهادی را ارائه می‌کند. همچنین شکل‌های ۲(د) و ۲(و) به‌منظور مقایسه جزئیات کوچک و بررسی دقت بازسازی در دو روش آورده شده‌اند.

۵- نتایج تجربی

می گردد. در روش سوم، الگوریتم پایه Inpainting معرفی شده در مرجع [۲۲] ارزیابی می شود و در نهایت، عملکرد روش پیشنهادی با این روش ها مقایسه می گردد. جدول ۱ میانگین خطای مطلق پیکسل ها را بین پس زمینه تولید شده و مرجع ارزیابی، برای هر چهار روش نشان می دهد. در محاسبه این معیار، تنها پیکسل هایی لحاظ شده اند که در فریم جاری یا فریم قبلی به خودرو تعلق دارند؛ زیرا این پیکسل ها نواحی چالش برانگیز برای بازسازی پس زمینه محسوب می شوند. ستون دوم جدول، نتایج مربوط به تمام پیکسل های متعلق به خودرو در فریم جاری و فریم قبلی را ارائه می دهد. در ستون سوم، محدودیت بیشتری اعمال شده است و تنها پیکسل هایی در نظر گرفته می شوند که در یکی از دو فریم، مقدار معتبر پس زمینه برای آن ها در دسترس باشد. این محدودیت، امکان تمرکز بر ارزیابی خودروهای در حال حرکت را فراهم می سازد و اثر خودروهای متوقف را از تحلیل جدا می کند. نتایج در هر دو حالت نشان می دهد که روش پیشنهادی، نسبت به سایر روش ها، کمترین میانگین خطای مطلق را به دست می آورد.

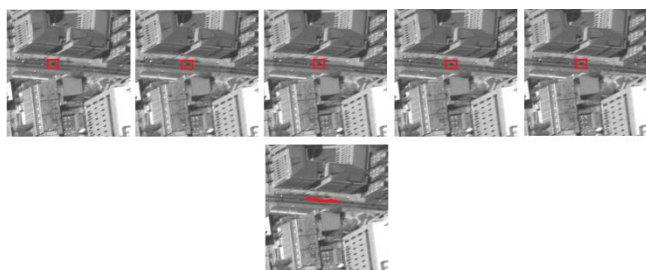
۶- ردیابی

در این بخش، نتایج کیفی و کمی الگوریتم ردیابی ارائه شده در بخش قبل بررسی می شود. شکل ۴ نمونه ای از عملکرد روش پیشنهادی را در ردیابی یک خودرو طی ۱۵۱ فریم، دریافت شده از پنج دوربین متوالی، تا پیش از خروج خودرو از میدان دید نشان می دهد. ردیابی این خودرو با چالش های متعددی همراه است؛ زیرا هدف در بخش هایی از مسیر به وسیله درختان و سایه های ایجاد شده دچار انسداد جزئی می شود. علاوه بر این، خودرو در دو تقاطع تغییر مسیر می دهد که موجب تغییرات قابل توجه در جهت و الگوی حرکت آن می شود. وجود انسداد، سایه و تغییرات ناگهانی جهت حرکت، استفاده از ردیاب های مبتنی بر مدل ظاهری ثلثت را با محدودیت مواجه می سازد. با این حال، روش پیشنهادی با بهره گیری همزمان از اطلاعات ظاهری هدف و اطلاعات پس زمینه، قادر است خودرو را با وجود این شرایط چالش برانگیز به طور پایدار ردیابی کند.

به منظور ارزیابی کمی عملکرد ردیابی، دو معیار متداول برای سنجش کارایی الگوریتم های ردیابی چندهدفه محاسبه شده است. نخستین معیار، نسخه ساده شده ای از معیار دقت ردیابی چندشیئی (Multiple Object Tracking Accuracy یا MOTA) است که مطابق رابطه زیر تعریف می شود:

در این بخش، نتایج حاصل از روش پیشنهادی برای تولید مدل پس زمینه و همچنین تلفیق الگوریتم های ردیابی و ناحیه بندی ارائه شده است. ارزیابی های صورت گرفته بر اساس مجموعه دادگان استاندارد CLIF ۲۰۰۷ گزارش می گردد. ارزیابی دقیق الگوریتم های تولید پس زمینه استاتیک در تصاویر هوایی به دلیل نبود داده های «مرجع صحیح» (Ground Truth) برای پس زمینه، چالش برانگیز است. اگرچه بخش عمده ای از تفاوت میان فریم های متوالی ناشی از حرکت اشیاء (خودروها) است، اما وجود ساختمان های بلند منجر به ایجاد پدیده «اختلاف منظر» (Parallax) می شود؛ این پدیده باعث جابه جایی ظاهری لبه های ساختمان ها تا چندین متر در هر فریم شده و تعیین دقیق پس زمینه واقعی را دشوار می سازد. علاوه بر این، تغییرات در میزان تاری (Blurring) و کنتراست فریم ها، پیچیدگی استخراج مدل پس زمینه را دوچندان می کند. به منظور سنجش عملکرد روش پیشنهادی، ۳۰ قطعه تصویر (Patch) متوالی و هم تراز شده با ابعاد ۵۰۰ در ۴۰۰ پیکسل از یک تقاطع شلوغ استخراج شد. سپس برای ایجاد یک مرجع ارزیابی، خودروهای موجود در مسیرها در هر فریم به صورت دستی ناحیه بندی شدند. در این فرایند، خودروهای پارک شده به عنوان بخشی از پس زمینه در نظر گرفته شدند، اما خودروهایی که به طور موقت (مثلاً پشت چراغ قرمز) متوقف شده بودند، به عنوان هدف علامت گذاری شدند. این نواحی قطعه بندی شده، نشان دهنده بخش هایی از تصویر هستند که نباید در مدل پس زمینه استاتیک ظاهر شوند. با مقایسه مدل پس زمینه تولیدی توسط الگوریتم با این نواحی مرجع، می توان میزان موفقیت روش را در حذف اهداف متحرک و ساکن و بازسازی دقیق پس زمینه زیر آن ها ارزیابی کرد. فرض می شود سایر نواحی تصویر، تخمین مناسبی از پس زمینه واقعی ارائه می دهند. شکل ۳ نمونه ای از تصاویر انتخاب شده به همراه ماسک خودروهای تولید شده به صورت دستی را نشان می دهد. سپس، هر جفت فریم متوالی از مجموعه ۳۰ تایی در نظر گرفته شده و عملکرد روش های مختلف تولید پس زمینه ارزیابی می شود. برای ایجاد مرجع ارزیابی، مقدار صحیح هر پیکسل از میان فریم های اخیر و در شرایطی انتخاب می شود که آن پیکسل در هیچ یک از آن فریم ها به عنوان پیکسل متعلق به خودرو برچسب گذاری نشده باشد. در این ارزیابی، چهار روش مختلف بررسی شده است. روش نخست، استفاده از میانگین دو فریم متوالی به عنوان مدل پس زمینه است. این روش یک راهکار ساده و ابتدایی محسوب می شود و به عنوان خط مبنای مقایسه، بیانگر حالت با عملکرد ضعیف تر است. در روش دوم، الگوریتم ارائه شده در مرجع [۲۱] به کار گرفته می شود که در آن مدل پس زمینه از میانگین گیری سه فریم متوالی، شامل فریم قبلی، تشکیل

نتایج ارزیابی نشان می‌دهد که در صورت حذف ناحیه‌بندی دقیق و استفاده صرف از جعبه محاطی مستطیلی، مقدار معیار دقت ردیابی چندشئی، یعنی MOTA، به ۴۹ درصد کاهش می‌یابد.



شکل (۴): ردیابی یک خودروی مشخص در کناره‌ی خیابان با انسداد و سایه‌های قابل توجه.

۷- نتیجه گیری

در این مقاله، چارچوبی جامع برای ردیابی خودروها در ویدئوهای هوایی پوشش‌دهنده نواحی گسترده ارائه شد. روش پیشنهادی بر تلفیق ردیابی و قطعه‌بندی استوار است و با هدف دستیابی به ردیابی مقاوم خودروها در شرایط چالش‌برانگیز طراحی شده است. همچنین، با توجه به قابلیت موازی‌سازی مراحل الگوریتم، این چارچوب می‌تواند برای پیاده‌سازی بلادرنگ روی سخت‌افزار FPGA و طراحی سامانه‌های تشخیص و ردیابی اهداف متحرک در کاربردهای پهپادی مناسب باشد. بیشتر پژوهش‌های پیشین در این حوزه بر ردیابی خودرو در محیط‌های باز و مسیرهای نسبتاً ساده، مانند بزرگراه‌ها، متمرکز بوده‌اند. در مقابل، این پژوهش بر خیابان‌های شهری کوچک‌تر تمرکز دارد که چالش‌های بیشتری از جمله سایه‌های ناشی از ساختمان‌ها، انسداد هدف به وسیله درختان و تغییرات قابل توجه در ظاهر خودرو را در بر می‌گیرند. به منظور مواجهه با این چالش‌ها، روشی برای قطعه‌بندی و ردیابی هم‌زمان خودروها ارائه شد. از آنجا که دقت قطعه‌بندی به وجود یک مدل پس‌زمینه مناسب وابسته است، مدل پس‌زمینه با استفاده از یک روش مبتنی بر الگوریتم Inpainting تولید شد. تلفیق اطلاعات قطعه‌بندی با فرایند ردیابی، امکان تفکیک بهتر هدف از پس‌زمینه و حفظ پایداری ردیابی را، حتی در شرایطی که ظاهر هدف به‌طور محسوسی تغییر می‌کند، فراهم می‌سازد.

جدول (۱): میانگین خطای پیکسل برای کل تصویر در ۴ روش مدل‌سازی پس زمینه

مدل	پیکسل همه خودروها	پیکسل های خودرو محدود
میانگین	۳۲/۴۲	۳۰/۱۸
میان [۲۱]	۱۵/۸۸	۱۰/۲۵
روش [۲۲]	۱۳/۰۷	۱۱/۵۳
روش پیشنهادی	۹/۷۶	۸/۳۱

$$MOTA = 1 - \frac{\sum_{t=1}^n \sum_{i=1}^{N_t} [\text{overlap}(T_{i,t}, G_{i,t}) < 0.25]}{\sum_{t=1}^n N_t} \quad (4)$$

که در آن N تعداد فریم‌ها، N_t تعداد اهداف مشخص شده از توصیفات مرجع برای T امین فریم، $T_{i,t}$ و $G_{i,t}$ جعبه‌های محاط مبتنی بر ردیاب و توصیفات واقعی برای I امین هدف در T امین فریم است. Overlap در واقع همپوشانی بین این جعبه‌های محاط را طبق رابطه زیر محاسبه می‌کند:

$$\text{overlap}(T_{i,t}, G_{i,t}) = \frac{|T_{i,t} \cap G_{i,t}|}{|T_{i,t} \cup G_{i,t}|} \quad (5)$$

برای یک هدف نشان داده شده در شکل ۴، MOTA یک محاسبه شده است چرا که خودروی موردنظر به خوبی در هر فریمی که قبل مشاهده است، ردیابی می‌شود.

به‌منظور تأیید مزیت تلفیق ناحیه‌بندی و ردیابی، همان خودرو با استفاده از نمایش مستطیلی هدف و بدون بهره‌گیری از ناحیه‌بندی پیکسل‌محور ردیابی شد. این روش در شرایطی که تغییرات ظاهری هدف محدود است، می‌تواند موقعیت خودرو را با دقت قابل‌قبولی تخمین بزند. با این حال، هنگامی که خودرو به سمت حاشیه جاده حرکت کرده و وارد نواحی سایه‌دار ناشی از درختان می‌شود، شباهت ظاهری میان هدف و پس‌زمینه افزایش می‌یابد. در نتیجه، ردیاب قادر به تفکیک صحیح خودرو از پس‌زمینه نبوده و هدف را از دست می‌دهد؛ نمونه‌ای از این وضعیت در شکل ۱ نشان داده شده است.

۸- پیشنهادات برای ادامه پژوهش

کارهای آینده می‌تواند در دو راستای اصلی دنبال شود. نخست، توسعه روش برای تشخیص و آغاز خودکار ردیابی خودروهای متحرک، بدون نیاز به دخالت کاربر یا تعیین دستی موقعیت اولیه هدف است. انتظار می‌رود با اعمال تغییرات مناسب در روش تولید مدل پس‌زمینه و سازوکار ایجاد فرضیه‌های اولیه هدف، بتوان به این قابلیت دست یافت. در این مسیر، باید تدابیر ویژه‌ای برای کاهش اثر اختلاف منظر ناشی از ساختمان‌های بلند و جلوگیری از تشخیص‌های کاذب در نظر گرفته شود. دوم، بهبود کارایی محاسباتی روش پیشنهادی است که برای کاربرد آن در ردیابی هم‌زمان تعداد زیادی خودرو در ویدئوهای هوایی، به‌ویژه در سامانه‌های بلادرنج مبتنی بر FPGA، اهمیت اساسی دارد.

مراجع

- Remote Sensing, vol. 13, no. 24, Art. no. 5128, 2021, doi: 10.3390/rs13245128.
- [10] "Feature based remote sensing image registration techniques: A comprehensive and comparative review," International Journal of Remote Sensing, 2022, doi: 10.1080/01431161.2022.2114112.
- [11] "Computational intelligence in remote sensing image registration: A survey," International Journal of Automation and Computing, 2020, doi: 10.1007/s11633-020-1248-x.
- [12] "Self-supervised keypoint detection and cross-fusion matching networks for multimodal remote sensing image registration," Remote Sensing, vol. 14, no. 15, Art. no. 3599, 2022.
- [13] G. Elharrouss, N. Almaadeed, S. Al-Maadeed, and Y. Akbari, "Image inpainting: A review," Neural Processing Letters, vol. 51, pp. 2007–2028, 2020.
- [14] J. Wang, X. Wu, and others, "Deep learning for image inpainting: A survey," Pattern Recognition, vol. 134, Art. no. 109046, 2023, doi: 10.1016/j.patcog.2022.109046.
- [15] "Deep learning-based image and video inpainting: A survey," arXiv preprint arXiv:2401.03395, 2024. [Online]. Available: <https://arxiv.org/abs/2401.03395>
- [16] "Transformer-based image and video inpainting: Current challenges and future directions," arXiv preprint arXiv:2407.00226, 2024. [Online]. Available: <https://arxiv.org/abs/2407.00226>.
- [17] Simple Sparse Bundle Adjustment (SSBA). Available: <http://www.inf.ethz.ch/personal/chzach/opensource.html>.
- [18] Bay, H., Tuytelaars, T., and Van Gool, L. Surf: speeded up robust features. Computer Vision–ECCV 2006, (2006), 404–417.
- [19] Hartley, R., and Zisserman, A. Multiple View Geometry in Computer Vision, Vol. 2. Cambridge, United Kingdom: Cambridge Univ. Press, 2000.
- [20] Fischler, M., and Bolles, R. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24, 6 (1981), 381–395.
- [21] Xiao, J., Cheng, H., Sawhney, H., and Han, F. Vehicle detection and tracking in wide field-of-view aerial video. In Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, 2010, 679–684.
- [22] Telea, A. An image inpainting technique based on the fast marching method. Journal of Graphics Tools, 9, 1 (2004), 23–34.
- [1] Air Force Research Laboratory, "Columbus Large Image Format (CLIF) 2007 Dataset Overview," 2007. [Online]. Available: <https://www.sdms.afrl.af.mil/index.php?collection=clif2007> [Accessed: Aug. 2, 2026].
- [2] X. Wu, W. Li, D. Hong, R. Tao, and Q. Du, "Deep learning for UAV-based object detection and tracking: A survey," IEEE Geoscience and Remote Sensing Magazine, vol. 10, no. 1, pp. 91–124, 2022, doi: 10.1109/MGRS.2021.3119745.
- [3] D. Du, P. Zhu, L. Wen, X. Bian, H. Ling, Q. Hu, T. Peng, J. Zheng, X. Wang, Y. Zhang, and others, "The unmanned aerial vehicle benchmark: Object detection and tracking," in Proc. European Conference on Computer Vision (ECCV), 2018, pp. 370–386.
- [4] P. Zhu, L. Wen, D. Du, X. Bian, H. Fan, Q. Hu, and H. Ling, "Detection and tracking meet drones challenge," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 11, pp. 7380–7399, 2022.
- [5] Y. Du et al., "VETRA: A dataset for vehicle tracking in aerial imagery—New challenges for multi-object tracking," in Proc. European Conference on Computer Vision (ECCV), 2024. [Online]. Available: https://www.ecva.net/papers/eccv_2024/papers_ECCV/papers/11352.pdf
- [6] A. Singh, M. V. Joshi, and others, "A survey on moving object detection for wide area motion imagery," IEEE Geoscience and Remote Sensing Magazine, 2016.
- [7] "Wide-area motion imagery vehicle detection in adverse conditions," In Proceedings of the 2023 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), St. Louis MO, 2023. doi: 10.1109/AIPR60534.2023.10440693.
- [8] J. Li et al., "Recent advances for aerial object detection: A survey," ACM Computing Surveys, 2024, doi: 10.1145/3664598.
- [9] Y. Ma, X. Zhang, and others, "Multimodal remote sensing image registration methods and advancements: A survey,"