مروری بر روش‌های ردیابی مبتنی بر بینایی؛ ویژگی‌های زمانی و مکانی

محمدحسین بیات[۱]، بهرام تارویردی‌زاده[۲]، محمد شهبازی*[۳]

دانشجوی دکتری مکاترونیک، دانشکدگان علوم و فناوری‌های میان‌رشته‌ای، دانشگاه تهران، تهران، ایران[۱]

دانشیار گروه مکاترونیک، دانشکدگان علوم و فناوری‌های میان‌رشته‌ای، دانشگاه تهران، تهران، ایران[۲]

استادیار گروه ساخت و تولید، دانشکدهٔ مکانیک، دانشگاه علم و صنعت، تهران، ایران[۳*]

چکیده

ردیابی مبتنی بر بینایی که یکی از پرچالش‌ترین زمینه‌های موجود در بینایی ماشین است به معنای دنبال‌کردن یک یا چند هدف در دنباله‌ای از تصاویر است؛ چالش‌هایی نظیر تغییر در ظاهر هدف، پوشانده‌شدن با عوامل محیطی، حرکات سریع و ناگهانی که هر یک زمینهٔ پژوهشی فعالی را به خود اختصاص داده‌اند. دو شاخص مهم در عملکرد یک ردیاب، سرعت اجرا و دقت آن هستند. با ساده‌ترشدن الگوریتم سرعت افزایش می‌یابد؛ اما از دقت کاسته می‌شود و تعامل میان این دو، موضوع مهمی به‌ویژه در پیاده‌سازی‌های عملی است. در این پژوهش به بررسی جامع و پیاده‌سازی الگوریتم‌های ردیابی مختلف پرداخته و روش‌های مناسب با کاربردهای عملی معرفی شده‌است. از سوی دیگر ساختارهای مختلف ردیابی بررسی و بر اساس ویژگی‌های مکانی، زمانی، بصری و حرکتی دسته‌بندی شده‌اند؛ همچنین به‌دلیل توسعهٔ روش‌های یادگیری عمیق و تأثیر آن‌ها در ردیابی، معماری‌های عمیق، پایگاه‌های داده، مجموعه‌های آموزشی، روش‌ها و استانداردهای ارزیابی معرفی شده و افق پیش‌روی این حوزه مورد بحث قرار گرفته‌است. بررسی‌ها نشان می‌دهد که ویژگی‌های زمانی و حرکتی علی‌رغم تأثیر مطلوب بر عملکرد ردیابی کمتر مورد توجه قرار گرفته‌اند. با توسعهٔ شبکه‌های عمیق حافظه‌دار استفاده از این ویژگی‌ها روبه‌افزایش بوده و سهم بیشتری را در ردیابی به خود اختصاص داده‌اند؛ ازاین‌رو ویژگی‌های زمانی و حرکتی با تمرکز بیشتری بررسی شده‌اند.

واژگان کلیدی: ردیابی مبتنی بر بینایی، ویژگی‌های ظاهری، ویژگی‌های حرکتی، یادگیری عمیق، بینایی ماشین.

A Review of Vision-Based Tracking Methods: Temporal and Spatial Features

Mohammad Hosein Bayat[1], Bahram Tarvirdizadeh[2], Mohammad Shahbazi[3*]

PhD Student, Department of Mechatronics Engineering, College of Interdisciplinary Science and Technology, Tehran University, Tehran, Iran[1]

Associated Professor, Department of Mechatronics Engineering, College of Interdisciplinary Science and Technology, Tehran University, Tehran, Iran[2]

Assistant Professor, School of Mechanical Engineering, Iran University of Science and Technology, Tehran, Iran[3*]

Abstract

Vision-based object tracking, as one of the most challenging fields in machine vision, means following the target(s) in the sequence of image frames in the presence of various challenges. In general, tracking algorithms can be classified to the single-target and multi-target based on the number of objects that should be tracked in frames. Trackers use two basic features in tracking: the appearance and motion. The appearance features are extracted from independent images but the motion features are produced through sequence of frames. According to the evaluations, motion models improve the tracking performance and take less process compared to the appearance features. Our investigations show that in contrast of single-target algorithms, the multi-target algorithms consider more contribution for the motion models, and due to the multiplicity of objectives in the scene they focus less on the appearance features.

* نویسندهٔ عهده‌دار مکاتبات

* Corresponding author

• تاریخ ارسال مقاله: ۱۴۰۲/۳/۹ • تاریخ پذیرش: ۱۴۰۳/۹/۱۴ • تاریخ انتشار: ۱۴۰۴/۳/۲۸ • نوع مطالعه: پژوهشی

[DOI: 10.61186/jsdp.22.1.83]

Despite the wide range of methods and significant progress in machine vision, reliable and flawless performance cannot be expected in the use of tracking algorithms with real-time criteria. This will be aggravated if one of the challenges occurs. Challenges such as sudden and fast movements by the target, occlusion by obstacles or other targets in the scene, extreme changes in the appearance and dimensions of the target, as well as entering and exiting the scene, which cause tracking algorithms to fail.

Having a good trade-off between the accuracy and the execution speed is one of the main problems for applied tracking algorithms. Detection algorithms, which are known to detect different targets in an independent image, have shown acceptable accuracy, but it is not possible to use them in every frame for a real-time tracking, because either due to the high processing volume of these algorithms, the execution speed of the detector is limited or they are only able to identify certain classes. But the purpose of a general tracking is to follow an object in a sequence of images regardless of its type and class as well as considering temporal and spatial dependencies among successive frames.

With the development of recurrent neural networks and their great ability to process sequential data such as text, audio and video, their use in tracking algorithms is increasing. The use of these networks has helped to improve the performance of tracking algorithms due to their short-term and long-term memory in maintaining important features during tracking. Different methods of integrating convolutional and recurrent neural networks are presented and showed grate performance in tracking, but the main drawback of most of them is the low execution speed of the algorithms. Our studies show that direct feeding the high-dimensional inputs, such as features extracted from images, to the recurrent networks greatly reduces their processing speed. Therefore, in some methods with the approach of real-time execution, the dimensions of the recurrent networks input are downsampled and reduced to the smaller size, although the accuracy is also slightly reduced.

Our investigations show that the use of motion models in single-target tracking algorithms is less explored compared to the multi-target methods. Meanwhile, the studies show the success of these models in improving tracking performance. Before the introduction of convolutional networks and their remarkable success in extracting deep features from the image, motion models were mostly used, but in recent methods, especially in single-target trackers, appearance features are used more. In single-target algorithms, the presence of only one object in the image and less computational volume compared to multi-target algorithms allows for more free use of appearance features, but this is not possible in multi-target tracking due to the multiplicity of targets so the motion models are more useful in these algorithms. Therefore, in this paper, a more detailed investigation of motion models and their effect on tracking performance is done. The results show that motion models have a profound effect on improving tracking performance while being simple and impose low processing volume.

In this paper, a comprehensive review and implementation of different tracking algorithms is discussed and appropriate methods are introduced for practical implementations. On the other hand, different tracking structures are investigated and categorized based on spatial, temporal, appearance and motion features. Also, due to the development of deep learning methods and their impact on tracking, deep architectures, training datasets and standard evaluation methods are studied and the future horizon of this field is discussed. Our studies show that temporal and motion features have received less attention despite their favorable impact on tracking performance. With the development of deep memory networks, the use of these features is increasing and they have taken a greater portion in tracking.

Keywords: Vision-Based Object Tracking, Appearance Features, Motion Features, Deep Learning, Machine Vision.

۱-مقدمه

بینایی ماشین یکی از زمینه‌های فعال و پرچالش در حوزهٔ هوش مصنوعی است. پیش از مطرح‌شدن روش‌های مبتنی بر یادگیری عمیق در پردازش تصویر، توسعهٔ روش‌های مطرح در این حوزه به وسیلهٔ الگوریتم‌های کلاسیک موجود در یادگیری ماشین صورت گرفته‌است[۱، ۲].

در سال‌های اخیر با پیشرفت چشم‌گیر در یادگیری عمیق و توسعهٔ شبکه‌های عمیق [۳] استفادهٔ روزافزون از این شبکه‌ها در حوزه‌های گوناگون از جمله بینایی ماشین شدت گرفته‌است[۴]؛ به‌گونه‌ای که ساختارهای مبتنی بر شبکه‌های عصبی در پاسخ به نیازهای ادراکی به ابزارهایی ضروری تبدیل شده‌اند. این امر به‌دلیل توانایی بالا و نزدیکی آن به عملکرد انسان در حل چالش‌های مختلف است[۵]. یکی از پرچالش‌ترین حوزه‌های موجود در بینایی ماشین، ردیابی مبتنی بر بینایی اهداف مختلف[1] است.

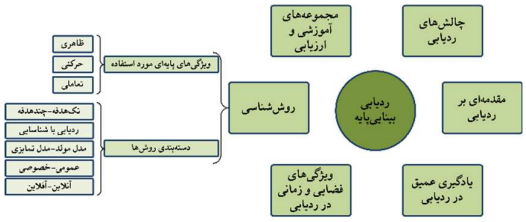
مهم‌ترین اصل در تعریف ردیابی تفاوت آن با شناسایی مبتنی بر بینایی اهداف[2] است. هدف از شناسایی جست‌وجو در تمام یک تصویر برای یافتن هدفی خاص در بین تمامی اهداف موجود بوده و خروجی آن کادر محصورکننده‌ای[3] به دور هدف شناسایی شده‌است[۶].

الگوریتم‌های شناسایی به دو دسته تقسیم می‌شوند: تک‌مرحله‌ای و دومرحله‌ای. ساختار کلی هر دو روش

[1] Vision-Based Generic Object Tracking
[2] Vision-Based Object Detection
[3] Bounding Box

[DOI: 10.61186/jsdp.22.1.83]

یکسان است و در آن ویژگی‌های ظاهری تصویر به‌وسیلهٔ یک معماری عمیق [۷–۹] استخراج می‌شود. در ادامه ویژگی مربوط به بخش‌های مختلف بررسی و اهداف موجود در آن‌ها شناسایی و دسته‌بندی می‌شوند. روش تک‌مرحله‌ای تمامی این مراحل را در یک مرحله انجام داده و با استفاده از سازوکار کادرهای معین اهداف با ابعاد مختلف را شناسایی می‌کنند[۶، ۱۰، ۱۱]؛ اما در روش دومرحله‌ای تصویر به ناحیه‌های کوچک‌تر تقسیم شده و ویژگی‌های ظاهری هر ناحیه جداگانه استخراج می‌شوند. در مرحلهٔ دوم ویژگی‌های به‌دست‌آمده دسته‌بندی شده و اهداف و امتیاز آن‌ها تولید می‌شوند[۱۲، ۱۳]. دقت الگوریتم‌های دومرحله‌ای بیشتر از تک‌مرحله‌ای، اما سرعت اجرای آن‌ها بسیار کمتر است؛ ازاین‌رو در پیاده‌سازی‌های عملی و کاربردی روش‌های تک‌مرحله‌ای مورد استفاده قرار می‌گیرند، اما از سوی دیگر مقصود از ردیابی دریافت کادر محصورکننده‌ای به دور هدف در قاب نخست و در ادامه جست‌وجو و یافتن آن در باقی قاب‌های یک فیلم است. این جست‌وجو و دنبال‌کردن می‌تواند در بخش محدودتری از تمام تصویر انجام شده و در شرایطی نظیر جابه‌جاشدن هدف، تغییر در شکل ظاهری و ابعاد، حرکات دوربین و تغییرات محیطی نظیر تغییر در نور و پس‌زمینه صورت می‌پذیرد. الگوریتم‌های ردیابی برای دنباله‌ای از قاب‌های متوالی از یک فیلم مورد استفاده قرار گرفته و وظیفهٔ آن درنظرگرفتن ارتباطی میان اهداف شناسایی‌شده در این قاب‌ها است.



(شکل-۱): ساختار مقاله در یک نگاه

(Figure-1): The structure of the article at a glance

ردیابی مبتنی بر بینایی کاربردهای گسترده‌ای در حوزه‌های گوناگون دارد؛ از جملهٔ این کاربردها می‌توان به سامانه‌های نظارتی بر اجسام متحرک نظیر وسایل نقلیه و عابران پیاده، استفاده در اتومبیل‌ها و پهپادهای خودران، موقعیت‌یابی و کنترل مبتنی بر بینایی ربات‌های صنعتی و خدماتی، آنالیز مسابقات ورزشی و بررسی حرکات دسته‌جمعی حشرات، آبزیان و یا سلول‌ها در تصاویر میکروسکوپی در حوزهٔ علوم زیستی اشاره کرد[۵، ۱۴، ۱۵].

با وجود پیشرفت‌های چشمگیر در حوزهٔ یادگیری عمیق و استفادهٔ گسترده از آن در بینایی ماشین به منظور استخراج ویژگی‌های عمیق از تصاویر [۱۶]، وضعیت کنونی در ردیابی اهداف موجود در تصویر همچنان بسیار دور از توانایی‌های یک انسان است؛ چرا که ایجاد ارتباط میان فضای زمانی و مکانی با استفاده از ویژگی‌های بصری، حرکتی و تعاملی میان اهداف مختلف امری دشوار برای شبکه‌های عصبی است[۵]. یکی دیگر از این دلایل ایجاد تعاملی میان سرعت و دقت ردیابی است. دست‌یابی به توانایی‌های چندجانبه برای یک شبکهٔ عصبی نیازمند سنگین‌ترشدن معماری آن است و هرچه ساختار الگوریتم ردیابی پیچیده‌تر باشد دقت ردیابی بالا می‌رود، اما در مقابل آن افزایش حجم محاسبات منجر به کاهش سرعت اجرای الگوریتم می‌شود؛ ازاین‌رو به‌دلیل عدم اجرای بلادرنگ استفادهٔ به‌لحظه در کاربردهای عملی ناممکن می‌شود[۱۷].

در این مقاله، روش‌های ردیابی مبتنی بر بینایی مورد بررسی قرار گرفته و ویژگی‌های اصلی مورد استفاده برای ردیابی؛ یعنی ویژگی‌های ظاهری و حرکتی تشریح شده‌اند. برای این منظور مقالات متعددی مطالعه و پیاده‌سازی شده و ساختار آن‌ها معرفی شده‌است. با توجه به اهمیت یادگیری عمیق و توسعهٔ روزافزون آن در حوزه‌های مختلف به‌ویژه بینایی ماشین، تمرکز اصلی این مقاله بر روی روش‌های ارائه‌شده در این زمینه است؛ هرچند که روش‌های کلاسیک نیز مورد توجه بوده‌اند؛ همچنین پایگاه‌های داده، مجموعه‌های آموزشی و روش‌های استاندارد ارزیابی معرفی شده‌اند. نتایج این پژوهش نشان می‌دهد که ویژگی‌های حرکتی در مقایسه با ویژگی‌های ظاهری کمتر مورد توجه قرار گرفته‌اند در حالی که عملکرد قابل توجهی در بهبود عملکرد ردیابی دارند؛ ازاین‌رو توجه ویژه‌ای به این زمینه شده و مدل‌های ارائه‌شده با جزئیات بیشتری بررسی شده‌اند.

در بخش دوم به معرفی چالش‌های ردیابی، مجموعه‌های مطرح برای آموزش الگوریتم‌های عمیق و همچنین ارزیابی ردیاب‌ها پرداخته شده‌است. مطالعهٔ روش‌های ردیابی و دسته‌بندی مقالات در بخش سوم صورت گرفته‌است. در بخش چهارم به بررسی و تأثیر یادگیری عمیق در ردیابی مبتنی بر بینایی پرداخته شده و در بخش پنجم تشریح مقالات صورت گرفته‌است. مدل‌های حرکتی به‌عنوان بخش کمتر توجه شده در زمینهٔ ردیابی در بخش شش بررسی و مقالات مرتبط دسته‌بندی و تشریح شده‌اند. در بخش پایانی هفتم نیز تحلیلی از دریافت‌های این مقاله و افق پیش‌رو برای این حوزه ارائه

[DOI: 10.61186/jsdp.22.1.83]

شده‌است. چکیده‌ای از ساختار مقاله در (شکل-۱) قابل مشاهده است.

۲-چالش‌ها، مجموعه‌های آموزشی و ارزیابی

ردیابی یک شئ در دنباله‌ای از تصاویر دوبعدی موجب بروز چالش‌های گسترده‌ای می‌شود؛ ازاین‌رو به‌دلیل وجود چالش‌های مختلف و از سوی دیگر تعریف ردیابی مبنی بر حفظ وابستگی‌ها[1] در دنباله‌ای از قاب‌ها، عملکرد الگوریتم‌های ردیابی در مقایسه با زمینه‌های دیگری همچون شناسایی مبتنی بر بینایی [۶، ۱۸] ضعیف‌تر است.

از چالش‌های مطرح در زمینهٔ ردیابی می‌توان به انسداد[2] توسط موانع موجود در صحنه، تغییرات ظاهری هدف، تغییر ابعاد و زاویهٔ هدف، تغییرات در محیط پیرامون، حرکات دوربین، تاری تصویر، تغییرات نور محیط، پس‌زمینهٔ شلوغ و مبهم، چرخش هدف در صفحه و خارج از صفحهٔ تصویر و کیفیت پایین تصویر اشاره کرد[۱۹]. در توسعهٔ الگوریتم‌های ردیابی غلبه بر هر یک از چالش‌های مطرح‌شده به تنهایی و یا ترکیبی از آن‌ها مورد توجه قرار دارد.

استفاده از معماری‌های مبتنی بر یادگیری عمیق مستلزم وجود مجموعهٔ آموزشی مناسب، گسترده، با تعداد و تنوع بسیار بالا است؛ هرچه تعداد و تنوع داده‌ها بیشتر باشد ویژگی‌های غنی‌تری استخراج شده و آموزش بهتری را از آن می‌توان به‌دست آورد. با توسعهٔ روزافزون روش‌های یادگیری عمیق در بینایی ماشین و به‌خصوص شناسایی و ردیابی مبتنی بر بینایی، مجموعه‌های آموزشی و ارزیابی مختلف با کاربردهای گوناگون ارائه شده‌است؛ از سوی دیگر برای ارزیابی الگوریتم‌های ردیابی فارغ از نوع ساختار آن داده‌های برچسب‌دار لازم است تا پیش‌بینی‌های الگوریتم با مقدار حقیقی آن مقایسه شود. در ادامه به معرفی این مجموعه‌ها برای ردیابی تک‌هدفه و چندهدفه پرداخته شده‌است.

OTB -۱-۲

این مجموعه از نخستین پایگاه‌های ارائه‌شده در حوزهٔ ردیابی و معیار ارزیابی معتبری برای سنجش عملکرد الگوریتم‌های ردیابی است. این مجموعه در دو نسخه شامل پنجاه فیلم برچسب‌دار تحت عنوان OTB50 [۲۰] و همچنین شامل صد فیلم برچسب‌دار تحت عنوان OTB100 [۲۱] است که ذیل یازده نوع چالش دسته‌بندی شده‌اند. این مجموعه برای ارزیابی عملکرد الگوریتم‌های ردیابی مورد استفاده قرار می‌گیرد، اما به‌دلیل نبود به‌روزرسانی از سال ۲۰۱۵ محبوبیت آن رو به کاستی است.

VOT Challenge-۲-۲

این مجموعه [۲۲] برای آموزش و ارزیابی عملکرد الگوریتم‌های ردیابی مورد استفاده قرار می‌گیرد و هر ساله چالشی را در زمینه‌های مختلف برگزار می‌کند. به‌روزرسانی فیلم‌ها و زمینه‌های ارزیابی موجب محبوبیت و اقبال گسترده‌ای به این مجموعه شده‌است. در سال‌های اخیر ارزیابی بر روی تصاویر حرارتی و سه‌بعدی نیز به این مجموعه اضافه شده‌است.

GOT-10K-۳-۲

مجموعه‌ای بزرگ [۲۳] با تصاویر متنوع برای ارزیابی است که در سال‌های اخیر مورد توجه قرار گرفته‌است. این مجموعه دارای بیش از ده هزار فیلم برچسب‌دار است که ۵۶۰ نوع هدف با طبقه مختلف در آن قرار دارد.

LaSOT-۴-۲

این مجموعه [۲۴] بسیار بزرگ و شامل ۱۴۰۰ فیلم برچسب‌دار با کیفیت تصاویر بسیار بالا است. ویژگی این مجموعه برچسب‌گذاری با دقت بالا است.

TrackingNet-۵-۲

این مجموعه [۲۵] شامل بیش از سی هزار فیلم برچسب‌دار است که از محیط‌های مختلف و اهداف گوناگون تهیه شده‌است. این مجموعه هم برای آموزش و هم ارزیابی مورد استفاده قرار می‌گیرد.

NFS-۶-۲

این مجموعه [۲۶] برای ردیابی با اشیای با سرعت بسیار بالا تهیه شده و شامل صد فیلم با نرخ تصویربرداری ۲۴۰ قاب بر ثانیه است.

UAV123-۷-۲

با گسترش استفاده از ربات‌های پرنده و تصویربرداری هوایی، مجموعه‌های آموزشی برای شناسایی و ردیابی در این فضای کاری نیز توسعه یافته‌اند. این مجموعه [۲۷] جزو نخستین‌هاست و در آن ۱۲۳ فیلم برای ردیابی تنها یک هدف برچسب‌گذاری شده‌است.

Visdrone-۸-۲

این مجموعه [۲۸، ۲۹] نیز شامل تصاویر هوایی است و بخش‌های مختلفی برای شناسایی و ردیابی و همچنین شمارش اهداف در فضای شلوغ دارد. در تصاویر این مجموعه تمامی اهداف ازطبقه‌های مختلف برچسب‌گذاری شده‌اند. تصاویر این مجموعه هم به‌صورت قاب‌های متوالی برای ردیابی و هم به‌صورت تصاویر مجزا برای شناسایی

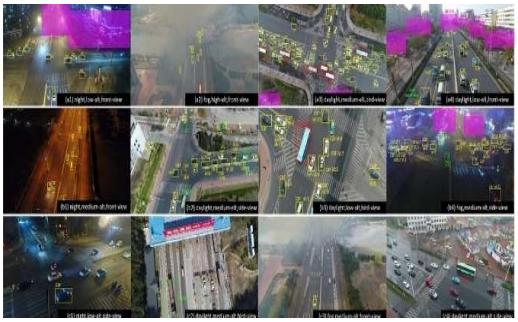
[1] Dependencies
[2] Occlusion

[DOI: 10.61186/jsdp.22.1.83]

است؛ همچنین بخش‌های جدیدی برای دیگر حوزه‌ها نظیر شمارش در فضای شلوغ و شناسایی در تصاویر حرارتی به آن اضافه شده‌است[۳۰].

UAVDT-۹-۲

این مجموعه [۳۱] شامل صد فیلم برای شناسایی و ردیابی در تصاویر هوایی است و در محیط شهری مختلف و آب‌وهوای متنوع تهیه شده‌است و تمامی اهداف از کلاس‌های مختلف برچسب‌گذاری شده‌اند. ویژگی مهم این مجموعه گستردگی شرایط آب‌وهوایی، روشنایی و ارتفاع‌های مختلف تصویربرداری است. نمونه‌ای از تصاویر این مجموعه در (شکل-۲) قرار گرفته‌است.



(شکل-۲): نمونه‌ای از تصاویر مجموعه [۳۱] در شرایط مختلف
(Figure-2): Example of [31] images in different conditions

MOT-۱۰-۲

این مجموعه [۳۲] برای آموزش و ارزیابی ردیاب‌های چندهدفه ارائه شده‌است در آن تمامی اهداف موجود در صحنه با دقت بالایی برچسب‌گذاری شده‌اند. این مجموعه برای هر هدف شناسهٔ مستقلی تخصیص داده و تا زمان حضور آن در دنبالهٔ تصاویر برچسب‌گذاری را انجام داده‌است. ویژگی این مجموعه تعداد زیاد و متنوع اهداف در تصاویر شلوغ و چالش‌برانگیز بوده و به‌عنوان اصلی‌ترین معیار برای ارزیابی ردیاب‌های چندهدفه مطرح است. این مجموعه شامل چالش‌ها و ویژگی‌های متعددی نظیر ردیابی سه‌بعدی و بخش‌بندی تصاویر[1] نیز است.

TAO-۱۱-۲

این مجموعه [۳۳] شامل فیلم‌های با کیفیت از محیط‌های متنوع بوده و تعداد فیلم‌های آن نزدیک به سه هزار مورد است. اهداف موجود در این تصاویر به‌طورکامل برچسب‌گذاری شده و به‌دلیل تنوع و گستردگی فیلم‌ها به یکی از مجموعه‌های مورد توجه در حوزهٔ ردیابی چندهدفه تبدیل شده‌است.

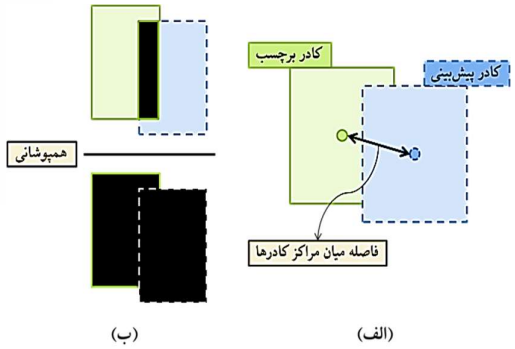
HiEve-۱۲-۲

بیشتر الگوریتم‌های ردیابی چندهدفه به‌دلیل استفاده از الگوریتم شناسایی در ساختار خود تنها برای دنبال‌کردن هدفی خاص طراحی می‌شوند. این هدف عموماً انسان بوده و ازاین‌رو برخی معیارهای ارزیابی توجه زیادی به این نوع هدف دارند؛ ازاین‌رو [۳۴] مجموعه‌ای بزرگ با کاربردهای مختلف نظیر ردیابی، پیش‌بینی حرکت، پیش‌بینی پیکربندی، تشخیص فعالیت، ردیابی عابران پیاده در مکان‌های شلوغ و شناسایی مجدد را برای انسان ارائه می‌کند.

۱۳-۲-معیارهای ارزیابی

معیارهای مختلفی برای ارزیابی عملکرد الگوریتم‌های ردیابی مورد استفاده قرار می‌گیرد که از جمله مهم‌ترین آن‌ها دقت و موفقیت[2] ردیابی هستند. این دو معیار توسط شیوه‌نامه‌های از پیش تعیین‌شده‌ای قابل محاسبه است [۲۰] و در قالب نمودارهای مشخصی نمایش داده می‌شوند.

در محاسبهٔ دقت ردیابی از فاصلهٔ اقلیدسی میان مراکز کادرها استفاده می‌شود. این فاصله میان مرکز کادر پیش‌بینی‌شده به‌وسیلهٔ ردیاب و مرکز کادر حقیقی در مختصات صفحهٔ تصویر محاسبه می‌شود (شکل-۳-الف). در تعیین موفقیت ردیابی از درصد هم‌پوشانی کادرها بهره‌گیری می‌شود (شکل-۳-ب).



(شکل-۳): (الف) فاصلهٔ میان مراکز کادرهای برچسب و پیش‌بینی معیاری برای دقت ردیابی، (ب) هم‌پوشانی میان کادرها معیاری برای موفقیت ردیابی
(Figure-3): (Right) Distance between the center of groundtruth and prediction bounding boxes as the tracking precision, (Left) Intersection over union as the tracking success

در این روش نسبت اشتراک کادرهای برچسب و کادر پیش‌بینی‌شده به اجتماع آن‌ها محاسبه و به‌عنوان درصد هم‌پوشانی کادرها درنظر گرفته می‌شود.

ترسیم نمودارهای دقت و موفقیت روند پیچیده‌ای دارد. برای این منظور آستانه‌هایی درنظر گرفته شده‌اند و تعداد قاب‌هایی که دقت یا موفقیت بیشتر از آستانه باشد شمارش می‌شود. این روند برای تمامی آستانه‌ها انجام و نمودار مربوطه ترسیم می‌شود[۲۰]. مجموعه‌های ارزیابی روش‌های ابتکاری مختلفی برای استفاده از نتایج «فاصلهٔ کادرها» و «هم‌پوشانی» آن‌ها ارائه کرده‌اند و به‌طور معمول به‌دلیل

[1] Image Segmentation

[2] Precision and Success

[DOI: 10.61186/jsdp.22.1.83]

پیچیدگی روش‌ها، ابزار مربوطه برای ارزیابی را نیز ارائه می‌کنند؛ هرچند که مبنای تمامی این روش‌ها محاسبهٔ دو مقدار فاصلهٔ مراکز و هم‌پوشانی است[۳۵].

همان‌طور که پیش‌تر بیان شد، ردیابی مبتنی بر بینایی یکی از پرچالش‌ترین زمینه‌های پژوهشی در حوزهٔ بینایی ماشین است؛ ازاین‌رو سهم قابل توجهی از ارزیابی‌ها بر روی این چالش‌ها صورت می‌گیرد. به این صورت که چالش‌های موجود در تمامی فیلم‌ها مشخص شده و عملکرد الگوریتم ردیابی در مقابله با این چالش‌ها سنجیده می‌شود. این ارزیابی با محاسبهٔ درصد دقت و موفقیت ردیابی قابل ارائه است. نمونه‌ای از چالش‌های ردیابی و همچنین شیوه‌نامهٔ مربوطه در (جدول-۱) مشاهده می‌شود[۳۶].

(جدول-۱): چالش‌های ردیابی و شیوه‌نامه تعیین آن در تصویر
(Table-1): Object tracking challenges

چالش	دستورالعمل [۳۶]
تغییرات روشنایی	نور موجود در محدودهٔ حضور هدف به صورت ناگهانی تغییر می‌کند.
تغییرات ابعاد	ابعاد هدف در میان قاب‌های مختلف بیش از دو برابر تغییر می‌کند.
انسداد	بخشی از هدف ردیابی یا تمام آن به وسیلهٔ عوامل محیطی پوشیده می‌شود.
تغییر شکل	تغییر شکل در هدف‌های غیر صلب صورت می‌گیرد.
تاری ناشی از حرکت	محدودهٔ حضور هدف به‌دلیل حرکت آن و یا حرکت دوربین تار می‌شود.
حرکت سریع	حرکت هدف با سرعتی بیش از بیست پیکسل در میان قاب‌های متوالی صورت گرفته‌است.
تغییرات در صفحهٔ تصویر	هدف در صفحهٔ تصویر چرخش انجام می‌دهد.
تغییرات خارج از صفحهٔ تصویر	چرخش در خارج از صفحهٔ تصویر؛ هدف در صفحه‌ای خارج از صفحهٔ تصویر چرخش انجام می‌دهد.
خارج از دید	خارج از دید؛ بخشی از هدف یا تمام آن از صفحهٔ تصویر خارج می‌شود.
پس‌زمینهٔ مبهم	پس‌زمینه نزدیک به هدف از لحاظ رنگ و ساختار شبیه به هدف است.
کیفیت پایین	تعداد پیکسل‌های موجود در کادر برچسب کمتر از چهارصد پیکسل است.
تغییرات نور	نور موجود در محدودهٔ حضور هدف ناگهانی تغییر می‌کند

۳- روش‌شناسی

اساس کار الگوریتم‌های ردیابی بر استفاده از دو نوع داده استوار است: مکانی[1] و زمانی[2]. داده‌های مکانی مبتنی بر ویژگی‌های ظاهری[3] استخراج‌شده از پیکسل‌های یک تصویر بوده و مستقل از تصاویر قبل و بعد آن هستند؛ اما داده‌های زمانی هنگامی مطرح می‌شوند که دنباله‌ای از تصاویر موجود بوده و مقصود دنبال‌کردن یک هدف یا اهداف مختلف در میان این تصاویر متوالی باشد. این امر مهم‌ترین تفاوت میان الگوریتم ردیابی با شناسایی است. در ردیابی دنباله‌ای از قاب‌ها و ایجاد ارتباط میان آن‌ها مطرح است، اما در شناسایی تنها یک تصویر موردنظر است.

۳-۱-ویژگی‌های پایه‌ای مورد استفاده برای ردیابی

در یک الگوریتم ردیابی استفاده از اطلاعات زمانی در کنار ویژگی‌های مکانی بسیار حائز اهمیت است. اطلاعات مکانی تنها با تکیه‌بر ویژگی ظاهری به‌دست‌آمده است؛ اما برای استخراج اطلاعات زمانی، ویژگی‌های مختلفی همچون ظاهری، حرکتی و تعاملی را در طول زمان می‌توان درنظرگرفت. در ادامه به بررسی دقیق‌تر این موارد پرداخته شده‌است.

۳-۱-۱-ویژگی‌های ظاهری

ویژگی‌های ظاهری در حین ردیابی دست‌خوش تغییرات گوناگون می‌شوند. این تغییرات هنگامی که هدف ردیابی به‌طورکامل صلب نبوده و یا متناسب با نور و محیط اطراف آن دچار تغییر در رنگ و ظاهر شود، بسیار مشهود خواهد بود. به‌خاطرسپردن این موارد توسط الگوریتم ردیابی و درنظرگرفتن تمهیدات لازم به‌منظور تطبیق با این تغییرات امری مهم و مؤثر در عملکرد آن خواهد بود؛ ازاین‌رو یکی از ویژگی‌های زمانی مهمی که توسط الگوریتم‌های ردیابی درنظر گرفته می‌شود، تغییرات ظاهری هدف در طول ردیابی است.

۳-۱-۲-ویژگی‌های حرکتی

اصلی‌ترین ویژگی مبتنی بر زمان را می‌توان ویژگی حرکتی دانست. این امر از آنجا ناشی می‌شود که توجه به شیوهٔ حرکت هدف می‌تواند باعث تمایز آن از دیگر اهداف موجود در تصویر شده و همچنین با دانستن جهت و سرعت حرکت هدف در قاب‌های پیشین می‌توان محل حضور آن در قاب بعدی را تا حد مطلوبی پیش‌بینی کرد. این امر موجب کوچک‌ترشدن محدودهٔ جست‌وجو و کاهش حجم پردازش الگوریتم شده و از سوی دیگر احتمال یافتن هدف را افزایش می‌دهد. این ویژگی در ردیابی چندهدفه به‌دلیل تعدد اهداف هم‌نوع و همچنین مشابهت آن‌ها با یکدیگر در مقایسه با ردیابی تک‌هدفه مورد توجه بیشتری قرار گرفته‌است.

۳-۱-۳-ویژگی‌های تعاملی

حرکت دسته‌جمعی اهداف نیز یکی دیگر از ویژگی‌های زمانی مورد توجه در ردیابی است. در کل اهدافی از یک نوع که در کنار همدیگر حرکت می‌کنند را می‌توان دارای رفتار مشابه درنظرگرفت و شیوهٔ حرکت آن‌ها را شبیه به همدیگر دانست؛ از سوی دیگر توجه اهداف به محیط پیرامون خود نیز راهی را

[1] Spatial
[2] Temporal
[3] Appearance Features

[DOI: 10.61186/jsdp.22.1.83]

برای پیش‌بینی حرکت بعدی آن‌ها باز می‌کند؛ برای نمونه انسان‌ها در هنگام حرکت خود برای عبور از موانع موجود در مسیر و عدم برخورد با دیگر انسان‌ها، مسیرهای ممکن را بررسی کرده و مناسب‌ترین آن‌ها را انتخاب می‌کنند؛ ازاین‌رو یکی از ویژگی‌های مورد استفاده در ردیابی، توجه به نحوهٔ برخورد و تعامل اهداف گوناگون با موانع و پیشامدهای موجود در مسیر حرکت است. پیچیدگی این امر و همچنین تحمیل حجم زیاد محاسبات به الگوریتم ردیابی استفاده از این نوع ویژگی را محدود کرده‌است.

برای یک ردیاب ایده‌آل استفاده از ویژگی‌های مختلف، متناسب با تغییرات ظاهری، ابعادی و نحوهٔ حرکت هدف لازم است و این ردیاب باید قادر باشد تا مدل خود را در طول ردیابی اصلاح کرده و با مشاهدات جدید منطبق کند[۳۷].

۳-۲-دسته‌بندی روش‌های ردیابی

به‌دلیل گستردگی روش‌های موجود در ردیابی دسته‌بندی‌های مختلفی درنظر گرفته می‌شود. کلی‌ترین دسته‌بندی را می‌توان تعداد اهداف موردنظر برای ردیابی دانست: ردیابی تک‌هدفه و ردیابی چندهدفه، اما دسته‌بندی‌های مختلف دیگری نیز مورد توجه است که مبتنی بر روش و معماری مورد استفاده برای ردیاب هستند. در ادامه به معرفی این موارد و تشریح جزئیات مربوطه پرداخته شده‌است.

۳-۲-۱-تک‌هدفه یا چندهدفه

کلی‌ترین دسته‌بندی را می‌توان تقسیم ردیاب‌ها به چندهدفه و تک‌هدفه دانست. در ردیابی تک‌هدفه تنها یک هدف در هر قاب مورد توجه است؛ درحالی‌که ردیابی چندهدفه به دنبال‌کردن اهداف متعددی در هر قاب می‌پردازد.

در ردیابی تک‌هدفه بیشتر از ویژگی‌های ظاهری استفاده شده‌است، اما در ردیابی چندهدفه مدل‌های حرکتی نیز مورد توجه‌اند. این امر از آنجا ناشی می‌شود که حجم پردازش بسیار زیادی برای استخراج ویژگی‌های ظاهری صرف شده است و در الگوریتم‌های چندهدفه به‌دلیل تعدد اهداف، استفاده از ویژگی ظاهری موجب افزایش شدید حجم محاسبات می‌شود؛ همچنین در ردیابی چندهدفه، برخلاف تک‌هدفه، برای هر هدف شماره شناسهٔ منحصربه‌فردی درنظر گرفته شده‌است تا در طول ردیابی از دیگر اهداف متمایز شود[۳۸]؛ ازاین‌رو کاهش محاسبات پردازشی یکی از مهم‌ترین مسائل در ردیابی چندهدفه است، اما در ردیابی تک‌هدفه به‌دلیل وجود تنها یک هدف، در استفاده از ویژگی‌های ظاهری محدودیت کمتری وجود دارد؛ ازاین‌رو روش‌های موجود در ردیابی تک‌هدفه، بیشتر به تقویت و توسعهٔ استخراج ویژگی‌های ظاهری و شیوهٔ استفاده از آن‌ها پرداخته‌اند و توجه کمتری به ویژگی‌های حرکتی داشته‌اند.

۳-۲-۲-ردیابی با شناسایی

یکی از روش‌های ردیابی، استفاده از الگوریتم‌های شناسایی در تمامی قاب‌ها است که به‌عنوان روش استاندارد ردیابی شناخته می‌شود[۳۸]. این روش بیشتر برای دنبال‌کردن اهداف خاص و هم‌نوع به کار می‌رود؛ چرا که دسته‌ای از ردیاب‌ها برای دست‌یابی به دقت بالاتر تنها برای اهداف خاصی طراحی می‌شوند[۳۹، ۴۰]. این ردیاب‌ها برای شناسایی هدف مورد نظر از روش‌های مختلفی همچون ویژگی‌های دست‌چین‌شده[1] و یا از ویژگی‌عمیق[2] استفاده می‌کنند.

روش نخست مبتنی بر استخراج ویژگی‌های ظاهری از پیکسل‌های تصویر است، اما روش دوم با اعمال پردازش بیشتر ویژگی‌های معنایی و عمیق‌تری را از تصویر استخراج می‌کند. این امر مستلزم طی‌شدن فرایند یادگیری بر روی داده‌های آموزشی مرتبط برای یک معماری عمیق است. روش کار این نوع ردیاب‌ها به این صورت است که ابتدا الگوریتم شناسایی تمامی اهداف هم‌نوع موجود در تصویر را شناسایی کرده و برقراری ارتباط میان آن‌ها با نتایج اطلاعات مربوط به قاب‌های پیشین به الگوریتم ردیابی سپرده می‌شود. این روش کاربرد بسیاری در ردیابی چندهدفه و دنبال‌کردن اهداف هم‌نوع دارد.

۳-۲-۳-ردیابی با مدل مولد یا مدل تمایزی

این دسته‌بندی را می‌توان به‌عنوان شاخص‌ترین روش تقسیم‌بندی الگوریتم‌های ردیابی دانست. در این دسته‌بندی الگوریتم‌های ردیابی را می‌توان به دو دسته تمایزی[3] و مولد[4] تقسیم کرد[۱۹، ۴۱].

روش تمایزی، ردیابی را به‌عنوان مسئلهٔ دسته‌بندی[5] درنظر می‌گیرد؛ در این روش چندین ناحیه را در اطراف محل حضور هدف در قاب پیشین درنظر گرفته و با استفاده از یک دسته‌بند از پیش آموزش‌دیده، نواحی مربوط به هدف را از نواحی مربوط به پس‌زمینه متمایز می‌کند. این دسته‌بند می‌تواند در طول ردیابی مدل خود را به‌روزرسانی کرده و با تغییرات ظاهری هدف تطبیق پیدا کند. به‌طورمعمول این امر مستلزم تغییر در پارامترهای الگوریتم بوده و حجم پردازش بالایی را طلب می‌کند، اما از طرف دیگر به دقت عملکردی بالایی در ردیابی منجر می‌شود.

[1] Hand-Crafted Features
[2] Deep Features
[3] Discriminative Model
[4] Generative Model
[5] Classification

[DOI: 10.61186/jsdp.22.1.83]

عملکرد مطلوب در عین سرعت اجرای بالا قابلیت اصلی یک الگوریتم ردیابی است؛ در این راستا به‌منظور ساده‌سازی معماری و عدم استفاده از الگوریتم شناسایی، ردیابی با استفاده از مدل مولد ارائه شده‌است. این روش به‌دلیل دست‌یابی به سرعت بالا و از طرفی عملکرد مطلوب یکی از روش‌های پرطرف‌دار در ردیابی است. در این روش که به آن تطبیق الگو[1] نیز گفته می‌شود، در ابتدا کادری به‌عنوان الگو و هدف ردیابی تعیین شده و در ادامه با ایجاد مدلی به منظور مقایسه میان اهداف موجود در قاب کنونی با الگوی تعیین‌شده ردیابی صورت می‌پذیرد. به‌منظور کاهش حجم پردازش جست‌وجو در بخش محدودی از تصویر و البته در نزدیکی محل حضور هدف در قاب پیشین صورت می‌گیرد. بسته به ساختار الگوریتم الگوی تعیین‌شده می‌تواند در حین ردیابی به‌روزرسانی شود.

برخلاف روش تمایزی، تمرکز روش تطبیق الگو بر روی ویژگی‌های ظاهری خود هدف بوده و پس‌زمینهٔ تصویر درنظر گرفته نمی‌شود؛ ازاین‌رو در صورت تغییرات اساسی در ظاهر هدف و یا پوشانده‌شدن آن به‌وسیلهٔ موانع موجود در صحنه، عملکرد الگوریتم با شکست مواجه می‌شود[۴۲]؛ از سوی دیگر به‌دلیل عدم به‌روزرسانی و آموزش برخط در حین ردیابی، روش تطبیق الگو راه‌های مؤثری برای انطباق با تغییرات ایجادشده در ظاهر هدف را نداشته و این امر موجب کاهش دقت ردیابی می‌شود، اما از طرفی این روش به‌دلیل پرهیز از آموزش برخط سرعت اجرای بالایی دارد؛ ازاین‌رو در پیاده‌سازی‌های عملی محبوبیت بیشتری نسبت به ردیاب‌های مبتنی بر مدل تمایزی دارد[۱۶، ۴۳].

۳-۲-۴-ردیابی عمومی یا خصوصی

در کل هدف ردیابی از پیش تعیین نشده و به تعبیر دیگر عمومی[2] است[۱۷، ۴۴] و تنها با تعیین آن در قاب نخست، ردیاب باید قادر باشد به‌سرعت مدل مختص آن را تولید کرده و در خلال ردیابی متناسب با تغییرات به وجود آمده خود را به‌روزرسانی کند؛ از طرفی مطلوب است تا ردیاب قادر باشد تنها ویژگی‌های مهم هدف را درنظر گرفته و تغییرات آن‌ها را به خاطر بسپارد. این امر به کاهش محاسبات و افزایش سرعت پردازش کمک شایانی می‌کند[۳۷]. این رویکرد در ردیابی بیشتر در روش تک‌هدفه در نظر گرفته می‌شود، اما دستهٔ دیگری از ردیاب‌ها تنها برای هدفی از یک نوع خاص (برای مثال انسان) طراحی می‌شوند[۳۹، ۴۰] و برای اهداف دیگر قابل استفاده نیستند. این ردیاب‌ها کاربرد کمتری نسبت به ردیاب‌های عمومی دارند.

[1] Template Matching
[2] Generic

۳-۲-۵-ردیابی برخط یا برون‌خط

در ردیابی با مدل برخط ساختار ردیاب در حین استفاده آموزش‌دیده و خود را با تغییرات هدف منطبق می‌سازد؛ هر چند که این نوع رویکرد موجب قدرتمندشدن ردیاب در مقابل تغییرات ظاهری هدف و افزایش دقت ردیابی می‌شود، اما به‌دلیل هزینهٔ محاسباتی بالا در آموزش مدل به‌خصوص در الگوریتم‌های مبتنی بر یادگیری عمیق، سرعت اجرای این الگوریتم‌ها بسیار پایین بوده و استفادهٔ بلادرنگ از آن‌ها در پیاده‌سازی‌های عملی ناممکن است.

برخلاف آن، در ردیابی به شیوهٔ برون‌خط ابتدا تمامی تغییرات و آموزش‌های مربوط به ساختار الگوریتم ردیابی انجام‌شده و در هنگام استفاده به‌طور کامل ثابت است و بدون تغییر باقی می‌ماند. این روش هر چند دقت کمتری نسبت به روش برخط دارد، اما افزایش چشمگیر سرعت اجرای الگوریتم ویژگی بسیار مطلوبی برای آن است. این امر در پیاده‌سازی‌های عملی و در کاربری‌های با سرعت بالا بسیار مهم است.

گفتنی است که در ادبیات موضوع برداشت دیگری نیز از عبارت برخط وجود دارد که منظور از آن، اجرای الگوریتم ردیابی بر روی قاب‌ها به صورت متوالی است. در مقابل آن روشی تحت عنوان پردازش دسته‌ای وجود داشته که در آن تمامی قاب‌ها به یک‌باره به الگوریتم داده می‌شوند. در این حالت ردیابی با استفاده از اطلاعات موجود در تمامی قاب‌ها و بدون درنظرگرفتن ترتیب میان قاب‌های متوالی صورت می‌گیرد. در این روش با بهره‌گیری از اطلاعات موجود در قاب‌های آتی ردیابی با دقت بیشتری صورت می‌پذیرد[۳۸]. روش مورد توجه در این پژوهش ردیابی به شیوهٔ برخط به منظور استفاده در پیاده‌سازی‌های عملی، بلادرنگ و کاربردی است.

۴-یادگیری عمیق در ردیابی مبتنی بر بینایی

با روی کارآمدن شبکه‌های عصبی پیچشی[3] و توفیق آن‌ها در استخراج ویژگی‌های عمیق از تصاویر [۴۵] اقبال گسترده و روزافزونی به استفاده از یادگیری عمیق در بینایی ماشین و به‌ویژه ردیابی مبتنی بر بینایی صورت گرفته‌است[۴۶]. رویکرد اصلی الگوریتم‌های مبتنی بر بینایی، مبتنی بر یادگیری عمیق، آموزش دنبال‌کردن اهداف مختلف به یک شبکهٔ عصبی با استفاده از مجموعهٔ بزرگی از داده‌های آموزشی است. از طرفی به‌دلیل طاقت‌فرسابودن روند تولید داده‌های آموزشی و

[3] Convolutional Neural Networks

[DOI: 10.61186/jsdp.22.1.83]

برچسب‌گذاری تصاویر مجزا و یا قاب‌های متوالی یک فیلم، مجموعه‌های محدودی برای آموزش این الگوریتم‌ها موجود است که یکی از محدودیت‌های اصلی در ردیابی با روش یادگیری عمیق محسوب می‌شود.

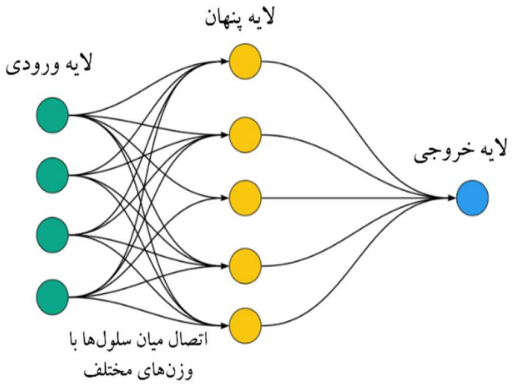
به‌منظور غلبه بر این مشکل گروهی از مقالات روش آموزش در حین ردیابی را پیشنهاد کرده‌اند[۴۷–۴۹]. در این روش، تحت عنوان ردیابی برخط، شبکه تنها در حین اجرا و بر روی قاب‌های فیلم در حال ردیابی آموزش می‌بینند. در رویکردی دیگر از این روش ابتدا الگوریتم ردیابی بر روی مجموعهٔ آموزشی محدودی آموزش‌دیده و در حین ردیابی بازتنظیم[1] می‌شود[۴۲]. این روش به‌دلیل حجم محاسبات بسیار بالا در آموزش شبکه‌های عصبی سرعت اجرای بسیار پایینی در حین ردیابی داشته، به‌طوری‌که امکان ردیابی بلادرنگ در پیاده‌سازی‌های عملی با استفاده از این الگوریتم‌ها وجود نخواهد داشت؛ همچنین به‌دلیل محدودبودن مجموعه‌های آموزشی و نبود داده‌های مختلف و متمایز در میان آن‌ها، شبکه تنها حالت‌های ساده‌ای را آموزش‌دیده و این امر عملکرد آن را تضعیف می‌کند[۱۷].

با استفاده از روش‌های مختلف در تقویت و افزایش داده‌های آموزشی و همچنین معرفی مجموعه‌های آموزشی جدید در حوزهٔ ردیابی مبتنی بر بینایی [۵۰–۵۲] دستهٔ دیگری از مقالات به توسعهٔ الگوریتم‌هایی با آموزش برون‌خط و ردیابی به صورت پایان‌به‌پایان[2] پرداخته‌اند. در این رویکرد شبکه در مرحلهٔ آموزش رابطه‌ای را میان ویژگی‌های مختلف نظیر شیوهٔ حرکت و تغییر ظاهر هدف آموخته و در هنگام ردیابی بدون هیچگونه آموزش مجددی مورد استفاده قرار می‌گیرد[۱۷، ۳۷، ۴۴]؛ هرچند دقت عملکرد این الگوریتم‌ها در مقایسه با روش‌های برخط کمتر است، اما سرعت بالای آن‌ها در هنگام اجرا توانایی مهم و مورد توجه است.

در ادامه ساختارهای مهم مورد استفاده در الگوریتم‌های ردیابی با رویکرد یادگیری عمیق تشریح شده‌اند.

۴-۱-شبکهٔ عصبی کامل متصل

شبکهٔ عصبی کامل متصل، شبکه‌ای است که تمامی سلول‌های آن به یکدیگر متصل بوده و وزن مخصوص به خود را دارند. این شبکه از لایه‌های مختلفی شامل لایهٔ ورودی، لایهٔ خروجی و لایه(های) مخفی میان این دو تشکیل شده‌است. تعداد لایه‌های مخفی قابل تغییر بوده و بسته به تعداد آن‌ها، پارامترهای آموزشی بیشتر شده و یادگیری شبکه را (شکل-۴) دشوارتر می‌کند. این شبکه توانمندی بالایی در یادگیری رابطه‌های غیرخطی و پیچیده در طول یک دورهٔ آموزش دارد. در پردازش تصاویر با ویژگی‌های عمیق، با توجه به تعداد بالای پیکسل‌های یک تصویر استفاده از این شبکه به‌صورت مستقیم مناسب نبوده و به‌طورمعمول از آن در دسته‌بندی یا رگرسیون انتهای شبکه استفاده می‌شود.



(شکل-۴): شماتیک شبکهٔ عصبی کامل متصل
(Figure-4): Fully connected neural network

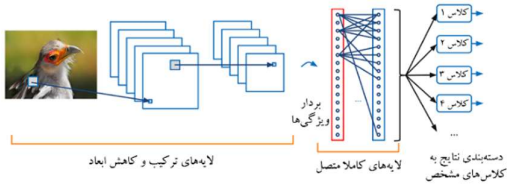
۴-۲-شبکهٔ عصبی پیچشی

شبکهٔ عصبی پیچشی یکی از قدرتمندترین انواع شبکه‌های عصبی مصنوعی، به‌خصوص در پیاده‌سازی‌های عملی، به‌منظور پردازش داده‌های تک‌بعدی نظیر سری‌های زمانی و همچنین داده‌های دوبعدی همچون تصاویر است. مبنای عملکردی این شبکه استفاده از عمل‌گر خطی ترکیب[3] است که به صورت ضرب ماتریسی اعمال می‌شود. با ترکیب یک ماتریس و همچنین هستهٔ[4] مشخص، ماتریس جدید با ابعاد کوچک‌تر از ماتریس ابتدایی ساخته می‌شود[۵۳].

به‌منظور استخراج ویژگی‌های تصاویر در ابتدا تصویر ورودی با ابعادی مشخص و در سه کانال رنگی قرمز، سبز و آبی به‌عنوان ماتریس اولیه وارد این شبکه شده و با استفاده از عمل‌گر ترکیب و با اعمال هسته‌های متفاوت ابعاد آن دستخوش تغییر می‌شود. این تغییرات به کاهش ابعاد تصویر و افزایش کانال‌های آن منجر شده و در انتها بردار ویژگی‌های عمیق تصویر استخراج می‌شود. در ادامه ویژگی‌های عمیق استخراج‌شده به‌وسیلهٔ لایه‌های کامل متصل (شکل-۵) دسته‌بندی می‌شوند. در سال‌های اخیر پیشرفت‌های گسترده‌ای بر روی شبکه‌های پیچشی و توانایی آن‌ها در پردازش، دسته‌بندی و استخراج ویژگی‌های تصاویر صورت گرفته و شبکه‌هایی با لایه‌های بسیار عمیق [۵۴] ارائه شده‌اند.

[1] Fine Tuning
[2] End-To-End
[3] Convolve
[4] Kernel

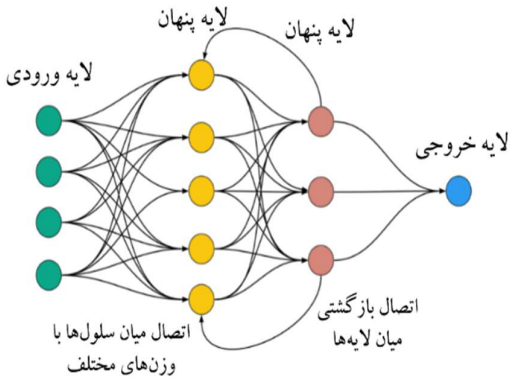
[DOI: 10.61186/jsdp.22.1.83]



(شکل-۵): شماتیک شبکهٔ عصبی پیچشی و دسته‌بندی نتایج
(Figure-5): Convolutional neural network and classification

۴-۳-شبکهٔ عصبی بازگشتی

شبکهٔ عصبی بازگشتی[1] یکی دیگر از انواع شبکه‌های عصبی است که به منظور پردازش داده‌های متوالی نظیر متن، صوت و فیلم مورد استفاده قرار می‌گیرد. تفاوت اصلی شبکهٔ بازگشتی با دیگر شبکه‌های عصبی با محاسبات روبه‌جلو[2]، استفاده از تاریخچهٔ زمانی مربوط به پردازش‌های قبلی است. این شبکه با حفظ اطلاعات به‌دست‌آمده از پردازش صورت‌گرفته بر روی یک مجموعهٔ متوالی و همچنین اشتراک‌گذاری آن‌ها با بخش‌های آتی به بررسی و تحلیل داده‌های متوالی می‌پردازد(شکل-۶) ؛ ازاین‌رو اقبال گسترده‌ای در استفاده از شبکه‌های بازگشتی در حوزه‌های مختلفی نظیر پردازش زبان طبیعی، ترجمهٔ متون، تبدیل متن به صوت و برعکس، دسته‌بندی صوت و متن و همچنین برآورد حالت[3] ایجاد شده‌است[۵۵]. با توسعهٔ روزافزون این شبکه، استفاده از آن در ردیابی مبتنی بر بینایی، که در آن پردازش داده‌های متوالی از نوع تصویر مطرح است، شدت گرفته‌است.



(شکل-۶): شماتیک شبکهٔ عصبی بازگشتی
(Figure-6): Recurrent neural network

به تناسب دادهٔ ورودی و دادهٔ خروجی، چهار نوع ساختار برای شبکه‌های بازگشتی تحت عنوان یک‌به‌یک، یک‌به‌چند، چندبه‌یک و چندبه‌چند درنظر گرفته می‌شود. در ادامه به تشریح این ساختارها پرداخته شده‌است.

[1] Recurrent Neural Network
[2] Feed Forward
[3] State Estimation

۴-۳-۱-ساختار یک‌به‌یک

در این ساختار ابعاد دادهٔ ورودی و خروجی برابر است و شبکه تنها یک ورودی و یک خروجی دارد؛ برای مثال یک واژه به‌عنوان ورودی شبکه قرار گرفته و شبکه وظیفه دارد تا یک واژه را به‌عنوان خروجی پیش‌بینی کند.

۴-۳-۲-ساختار یک‌به‌چند

در این ساختار شبکه یک ورودی دارد و به‌ازای آن چند خروجی دریافت می‌شود؛ برای مثال یک حرف به‌عنوان ورودی به شبکه داده شده و شبکه یک واژه را با پیش‌بینی حرف‌های بعدی آن خواهد ساخت؛ همچنین خروجی هر لایه به‌عنوان ورودی لایهٔ بعد مورد استفاده قرار می‌گیرد.

۴-۳-۳-ساختار چندبه‌یک

در این حالت شبکه چند ورودی دارد و به ازای آن تنها یک خروجی دریافت می‌شود. خروجی شبکه در لایهٔ آخر است؛ برای نمونه جمله‌ای شامل چندین واژه به‌عنوان ورودی به شبکه داده شده و یک واژه و یا یک عدد به‌عنوان خروجی شبکه درنظر گرفته می‌شود.

۴-۳-۴-ساختار چندبه‌چند

در این ساختار نیز همچون ساختار یک‌به‌یک ابعاد ورودی و خروجی برابر است، اما شبکه چند ورودی و چند خروجی دارد. در این حالت شبکه به‌ازای هر ورودی یک خروجی منحصر به فرد خواهد داشت؛ برای مثال در هر مرحله واژه‌ای به‌عنوان ورودی فرض شده و متناسب با آن واژه‌ای یا عددی به‌عنوان خروجی در نظر گرفته می‌شود. در مرحلهٔ بعد واژه‌ای دیگر وارد شده و خروجی متناسب با آن پیش‌بینی می‌شود. این در حالی است که برای این پیش‌بینی از اطلاعات مربوط به مرحلهٔ پیشین نیز استفاده می‌شود.

۴-۴-شبکهٔ عصبی بازگشتی LSTM

شبکه‌های بازگشتی توانایی حفظ وابستگی‌های زمانی میان داده‌های ورودی به شبکه را دارند. این در حالی است که حفظ ویژگی‌های مهم در داده‌های بسیار طولانی به‌سادگی امکان‌پذیر نیست؛ ازاین‌رو روش‌هایی به‌منظور بهبود این مشکل و افزایش توانایی شبکه‌های بازگشتی در حفظ وابستگی‌های طولانی‌تر ارائه شده‌است. در این میان یکی از موفق‌ترین روش‌ها، شبکهٔ بازگشتی با حافظهٔ کوتاه و بلندمدت[4] است[۵۶]. این شبکه علاوه‌بر پارامتر پنهان موجود در شبکه‌های بازگشتی پارامتر دیگری برای مدیریت تاریخچهٔ وابستگی‌های زمانی دارد. این پارامتر به

[4] Long Short-Term Memory (LSTM)

[DOI: 10.61186/jsdp.22.1.83]

شبکه اجازه می‌دهد تا میان حفظ یا حذف وابستگی‌ها تصمیم‌گیری کند؛ ازاین‌رو شبکه توانایی پردازش داده‌های طولانی‌تر را به نسبت شبکه‌های بازگشتی ساده خواهد داشت[۵۳].

۵- ویژگی‌های مکانی و زمانی در ردیابی

همان‌طور که پیش‌تر ذکر شد، اطلاعات مکانی و زمانی مبنای طراحی الگوریتم‌های ردیابی هستند. اطلاعات مکانی مربوط به محل حضور هدف در یک تصویر بوده و شامل مواردی همچون مختصهٔ مکانی و ویژگی‌های ظاهری مستخرج از پیکسل‌های تصویر است. اطلاعات زمانی با درنظرگرفتن تاریخچه‌ای از ویژگی‌های هدف ردیابی در میان تصاویر متوالی مانند ویژگی‌های حرکتی آن به‌دست می‌آید. این موضوع شاخصهٔ مهمی برای یک ردیابی موفق است؛ چرا که ردیابی در میان قاب‌های متوالی و مرتبط با یکدیگر صورت می‌گیرد.

در کنار ویژگی‌های ظاهری، توجه به شاخصه‌های حرکتی، رفتاری و ذاتی هدف ردیابی یکی از مهم‌ترین مواردی است که می‌توان در بهبود عملکرد الگوریتم‌های ردیابی درنظر گرفت. ویژگی‌های حرکتی را از دو جنبهٔ مکانی و زمانی می‌توان مورد استفاده قرار داد. ویژگی مکانی مربوط به موقعیت هدف در تصویر بوده و ویژگی زمانی در ایجاد ارتباط میان مکان هدف در قاب‌های متوالی است؛ برای مثال تغییرات مکان معرف سرعت حرکت و تغییرات سرعت معرف شتاب حرکت هدف است. واحد زمانی فاصله میان هر دو قاب درنظر گرفته می‌شود.

با گسترش روزافزون یادگیری عمیق و موفقیت آن در استخراج ویژگی‌های معنایی[1]، اطلاعات بیشتری از یک تصویر ساده به‌دست‌آمده و ازاین‌رو رشد چشم‌گیری در پردازش تصاویر حاصل شده‌است؛ ازاین‌رو الگوریتم‌های ردیابی نیز پیشرفت قابل توجهی داشته‌اند. شبکه‌های پیچشی توانایی شگرفی در استخراج و تفسیر اطلاعات عمیق مکانی از تصویر دارند، اما به تنهایی قادر به برقراری ارتباط زمانی میان قاب‌های متوالی نیستند؛ ازاین‌رو ضروری است تا با بهره‌گیری از ساختارهای دیگر مختصهٔ زمانی نیز درنظر گرفته شود.

بیشتر فعالیت‌های صورت‌گرفته در بهبود عملکرد الگوریتم‌های ردیابی مربوط به توسعه و توانمندسازی معماری‌ها در استخراج و تفسیر ویژگی‌های ظاهری و برقراری ارتباط بهتر میان این ویژگی‌ها در میان قاب‌های متوالی بوده‌است. در این میان ویژگی‌های حرکتی به‌عنوان عاملی مؤثر در فضای زمانی کمتر مورد توجه قرار گرفته‌اند.

این موضوع را می‌توان به چند دلیل دانست. به طور معمول حرکت‌های موجود در تصاویر متوالی منبع از پیش تعیین‌شده و مشخصی نداشته و حرکت دوربین، اجزای موجود در صحنه، هدف ردیابی و یا حرکت هم‌زمان آن‌ها می‌تواند موجب بروز چالش در خلال ردیابی شود؛ از سوی دیگر مدل‌سازی و پیش‌بینی حرکت صحنه یا دوربین امری دشوار است و این موضوع در حرکات پیچیده، ناگهانی و سریع تشدید می‌شود؛ اما از سوی دیگر، این امر به معنای موفقیت کامل ویژگی‌های ظاهری نیست و این ویژگی‌ها نیز با چالش‌های فراوانی مواجه‌اند. تغییرات ظاهری هدف، پس‌زمینه و محیط تأثیر بسزایی در ویژگی‌های ظاهری دارند و موجب کاهش دقت ردیابی می‌شوند؛ ازاین‌رو استفادهٔ ترکیبی از ویژگی‌های حرکتی و ظاهری یکی از نیازمندی‌های اساسی یک ردیابی موفق بوده و تأثیر به‌سزایی در عملکرد آن خواهد داشت. ویژگی‌های حرکتی با پیش‌بینی محتمل‌ترین مکان برای حضور هدف موجب کوچک‌تر و دقیق‌ترشدن فضای جست‌وجو در تصویر شده و در کنار ویژگی‌های ظاهری ردیابی مطلوب‌تری را موجب می‌شوند.

همچنین از جنبهٔ حجم محاسبات و سرعت پردازش، استخراج ویژگی‌های ظاهری امری پرهزینه است و استفاده از مدل‌های حرکتی چه برای پیش‌بینی محل دقیق هدف و چه برای جایابی محدودهٔ جست‌وجوی کوچک‌تر در قاب جدید کمک شایانی به کاهش حجم محاسبات و افزایش سرعت اجرا می‌کنند؛ ازاین‌رو استفاده از مدل‌های حرکتی محبوبیت بسیاری در پیاده‌سازی‌های عملی و بلادرنگ دارد. با توجه به جایگاه مهم این موضوع و از طرفی کم‌توجهی روش‌های نوین نسبت به آن، در این مقاله استفاده از ویژگی‌های زمانی به‌ویژه مدل‌های حرکتی با تمرکز بیشتری مورد بررسی قرار گرفته‌اند.

با پیشرفت چشم‌گیر در یادگیری عمیق و ارائهٔ معماری‌هایی با قابلیت یادگیری داده‌های متوالی، ساختارهای مطرح در این حوزه مورد استقبال روزافزون قرار گرفته‌اند. در این میان شبکه‌های بازگشتی توانایی بارزی در حفظ ویژگی‌های زمانی داشته و استفاده از آن‌ها موجب افزایش چشم‌گیر در عملکرد ردیابی شده‌است؛ ازاین‌رو در دسته‌بندی مقالات، استفاده از شبکه‌های بازگشتی به‌عنوان شاخصه‌ای مهم درنظر گرفته شده‌است. با توجه به گستردگی دسته‌بندی‌های مربوط به الگوریتم‌های ردیابی، تک‌هدفه یا چندهدفه‌بودن الگوریتم به‌عنوان مهم‌ترین دسته‌بندی درنظر گرفته شده‌است. در پایان توضیحات خلاصه‌ای از موارد بررسی شده در (جدول-۲) قرار داده شده‌است.

دسته‌بندی با استفاده از الگوریتم‌های یادگیری عمیق به طور معمول با استفاده از لایه‌های کامل متصل صورت می‌گیرد. روند کار به این صورت است که ویژگی‌های ظاهری پس از استخراج از تصویر به بردار تک‌بعدی ویژگی‌های ظاهری تبدیل شده و وارد لایهٔ نخست می‌شود.

[1] Semantic Features

[DOI: 10.61186/jsdp.22.1.83]

(جدول-۲): دسته‌بندی الگوریتم‌های ردیابی بررسی‌شده به تفکیک ساختار اصلی الگوریتم و روش مورد استفاده

(Table-2): Classification of the studied tracking algorithms based on the architecture and tracking method

	روش ردیابی	ساختار اصلی الگوریتم	ردیاب
تک‌هدفه	ردیابی با شناسایی	الگوریتم شناسایی YOLO.v1 [۱۰] + شبکهٔ عصبی بازگشتی	[۵۷]
		الگوریتم شناسایی FR-CNN [۵۹] + شبکهٔ عصبی بازگشتی + مدل حرکتی	[۵۸]
		الگوریتم شناسایی نشان‌گر + مدل حرکتی	[۷۱]
	مدل مولد (تطبیق الگو)	شبکهٔ عصبی پیچشی + شبکهٔ عصبی بازگشتی + یادگیری تقویتی	[۱۴] [۶۰]
		شبکهٔ عصبی پیچشی + شبکهٔ عصبی بازگشتی	[۳۷] [۴۰] [۶۱] [۶۲] [۱۶] [۶۳] [۶۴] [۶۵]
		شبکهٔ عصبی پیچشی + مدل حرکتی	[۷۲]
		شبکهٔ عصبی پیچشی + شبکهٔ عصبی کامل متصل	[۴۴]
		شبکهٔ عصبی کامل پیچشی	[۱۷] [۴۳] [۴۷]
	مدل تمایزی (دسته‌بندی)	شبکهٔ عصبی پیچشی + شبکهٔ عصبی بازگشتی	[۴۱]
		شبکهٔ عصبی پیچشی + شبکهٔ عصبی RBF	[۶۶]
		شبکهٔ عصبی پیچشی + نقشه برجستگی + SVM	[۶۷]
		شبکهٔ عصبی کامل پیچشی	[۷۰]
		شبکهٔ عصبی کامل پیچشی + تولید ماسک	[۶۲] [۶۳] [۷۵]
	کلاسیک	عدم استفاده از ساختارهای آموزش‌محور عمیق	[۲] [۳۹] [۷۶] [۷۷]
چندهدفه	ردیابی با شناسایی	الگوریتم شناسایی + شبکهٔ عصبی بازگشتی +مدل حرکتی	[۵] [۵۵]
		الگوریتم شناسایی FR-CNN [۵۹] +مدل حرکتی	[۴۶] [۷۹] [۸۰] [۸۱]
	مدل مولد (تطبیق الگو)	شبکهٔ عصبی کامل پیچشی + مدل حرکتی	[۸۴]
	تمایزی (دسته‌بند)	شبکهٔ عصبی کامل پیچشی + مدل حرکتی	[۴۹] [۸۲] [۸۳]
	کلاسیک	مدل پیچیدهٔ تعاملی	[۸۵]

در ادامه با توجه به معماری مورد استفادهٔ لایه‌های میانی با تعداد مناسب ورودی به ترتیب پردازش و در پایان خروجی موردنظر با تعداد مشخص تولید می‌شود. در این روش بر حسب معمول از رگرسیون برای تولید خروجی استفاده می‌شود.

۵-۱-ردیاب‌های تک‌هدفه

همان طور که ذکر شد ویژگی‌های ظاهری عمیق با استفاده از شبکه‌های پیچشی از تصاویر استخراج می‌شوند، اما مهم‌تر از آن پردازش مناسب این ویژگی‌هاست که مرحله‌ای مهم و تأثیرگذار در عملکرد الگوریتم‌های ردیابی است. روش‌های مختلفی برای این امر وجود دارند که در یک دسته‌بندی کلی می‌توان آن‌ها را به روش‌های مبتنی بر یادگیری ماشین (کلاسیک) و یادگیری عمیق تقسیم کرد.

۵-۱-۱-استفاده از شبکهٔ بازگشتی در ردیاب

نیگ و همکاران [۵۷] با الهام از روش رگرسیون مورد استفاده در الگوریتم‌های شناسایی، ردیابی را با ساختاری مشابه ارائه کردند و در آن به پیش‌بینی مستقیم مکان هدف با استفاده از شبکهٔ بازگشتی پرداخته‌اند؛ از آنجا که الگوریتم‌های شناسایی هر تصویر را مستقل پردازش می‌کنند، استفاده از آن‌ها در ردیابی نیازمند افزودن شاخصه‌های زمانی است. این امر با افزودن شبکهٔ بازگشتی به الگوریتم شناسایی [۱۰] محقق شده‌است. هدف اصلی این مقاله افزایش میزان یادگیری شبکهٔ عصبی بازگشتی و تحلیل اطلاعات هم در دامنهٔ زمانی و هم مکانی بوده‌است. با استفاده از این روش دقت ردیابی افزایش یافته و این امر در چالش انسداد نمود بیشتری داشته‌است. استفاده از الگوریتم شناسایی در هر قاب و همچنین افزایش حجم پارامترهای شبکهٔ بازگشتی در خلال زمان موجب افزایش حجم محاسبات شده و سرعت اجرا را کاهش می‌دهد.

فنگ [۵۸] با استفاده از الگوریتم شناسایی قدرتمند [۵۹] هدف را ابتدا تشخیص داده و شیء شناسایی‌شده به فهرست شبکهٔ بازگشتی اضافه می‌شود. این شبکه با استفاده از اطلاعات مکانی و زمانی و تاریخچهٔ ردیابی شامل ویژگی‌های ظاهری و ویژگی‌های حرکتی استخراج‌شده از هدف در قاب‌های پیشین موجب پیش‌بینی دقیق‌تر می‌شود. در این ردیاب با کمک مدل حرکتی مکان جست‌وجو در قاب بعدی تعیین شده و چندین محدوده جدید در اطراف آن برای حضور هدف پیشنهاد داده می‌شود. با استفاده از ویژگی‌های ظاهری استخراج از این محدوده‌ها و مقایسهٔ آن با تاریخچهٔ ویژگی‌های ظاهری هدف، محدوده‌ها امتیازدهی و محتمل‌ترین آن‌ها انتخاب می‌شود. ردیاب پیشنهادی در حضور اهداف با شباهت ظاهری زیاد و انسداد دچار مشکل است و عملکرد مناسبی ندارد. این امر را می‌توان به‌دلیل سادگی مدل حرکتی استفاده‌شده در آن درنظر گرفت. در این

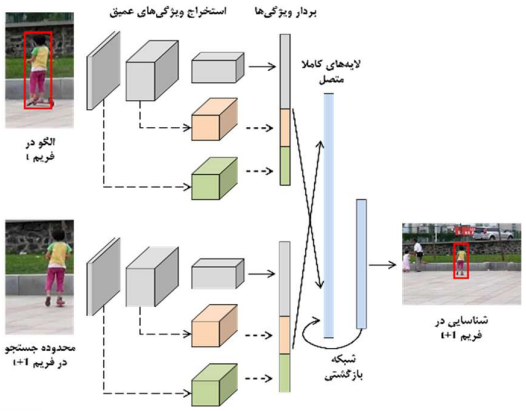
[DOI: 10.61186/jsdp.22.1.83]

مدل حرکتی سرعت میانگین هدف در ده قاب اخیر محاسبه و با استفاده از آن محل جدید هدف پیش‌بینی می‌شود.

ژانگ و همکارن [۱۴] شبکهٔ پیچشی و شبکهٔ بازگشتی را ترکیب کرده و آن را با روش یادگیری تقویتی[1] و به‌صورت برون‌خط آموزش داده‌اند. مبنای این مقاله افزودن وابستگی‌های زمانی به شبکهٔ پیچشی است؛ چرا که این شبکه در یافتن ویژگی‌های مکانی بسیار توانمند، اما فاقد اطلاعات زمانی میان ویژگی‌های استخراج‌شده از تصاویر متوالی است؛ برای مثال هدف ردیابی در قاب کنونی در همسایگی هدف شناسایی‌شده در قاب قبلی بوده و در مکانی نامرتبط و دور از آن نیست. در این ردیاب ویژگی‌های ظاهری هر قاب با استفاده از شبکهٔ پیچشی استخراج شده و به شبکهٔ بازگشتی وارد می‌شود. شبکهٔ بازگشتی با استفاده از تاریخچهٔ ذخیره‌شده در پارامترهای خود کادر مربوط به هدف را پیش‌بینی کرده و پارامترهای خود را به‌روزرسانی می‌کند. استفاده از این امر به بهبود ردیابی کمک می‌کند و تأثیر آن در افزایش موفقیت الگوریتم بیشتر از دقت آن بوده‌است. پس‌زمینه و صحنه‌های شلوغ و مبهم و همچنین اهداف نزدیک به هم در عملکرد الگوریتم مشکل ایجاد می‌کنند.

برای ردیابی ویژگی‌های ظاهری از تصویر استخراج‌شده، الگوهای ذخیره‌شده در مراحل پیشین از حافظه خوانده شده و با الگوی ابتدایی به‌دست‌آمده از قاب نخست ترکیب می‌شوند تا الگوی نهایی را تشکیل دهند. الگوی نهایی با ویژگی‌های تصویر داده‌شده ترکیب می‌شوند تا نقشهٔ ویژگی‌ها به‌دست‌آمده و محل حضور هدف تعیین شود. ساختار این مقاله پیچیدگی بسیاری دارد و از شبکهٔ بازگشتی با ورودی‌ها و پارامترهای متعدد استفاده شده‌است؛ هرچند به دقت مطلوبی نیز دست یافته‌است، اما سرعت اجرای الگوریتم بلادرنگ نیست. این ردیاب [۶۰] با ایجاد تغییراتی جزئی و افزودن ساختاری جدید برای نبود استفاده از الگوی اختلال‌برانگیز بهبود می‌یابد و کنترل بیشتری بر روی اهداف ذخیره شده دارد.

گوردون و همکاران [۳۷] به توسعهٔ الگوریتمی به‌منظور ردیابی اهداف عمومی پرداخته‌اند. در این مقاله به‌جای استفاده از یک الگوریتم به طور کامل ثابت[2] و بدون فرایند آموزش و یا آموزش برخط در هنگام ردیابی که موجب افزایش حجم پردازش و کاهش شدید سرعت اجرای الگوریتم می‌شود، از شبکهٔ بازگشتی استفاده شده و به وسیلهٔ آن اطلاعات مربوط به هدف در ۲پارامترهای شبکه ذخیره و در هر قاب به‌روزرسانی می‌شود. ورودی این ردیاب دو قاب متوالی پیشین و کنونی است و از شبکهٔ پیچشی برای استخراج ویژگی‌های ظاهری قاب‌ها، شبکهٔ بازگشتی برای به‌خاطرسپردن این ویژگی‌ها و لایه‌های به‌طور کامل متصل مبهنظور تعیین مکان هدف در قاب کنونی استفاده شده‌است شکل (۷). پرهیز از آموزش برخط و استفاده از پارامترهای شبکهٔ بازگشتی موجب دست‌یابی به سرعت اجرای بسیار بالا، اما با دقتی کمتر از روش آموزش‌محور شده‌است؛ همچنین به‌دلیل استفاده از شبکهٔ بازگشتی نتایج مطلوبی در مقابله با چالش انسداد به دست می‌آید. این مقاله از مدل حرکتی استفاده نکرده‌است و به‌دلیل آموزش بر روی داده‌های محدود، در ردیابی اهداف ناشناخته دقت کمتری دارد. از طرفی سرعت اجرای بالا و قابلیت آموزش مجدد آن را به یک الگوریتم مطلوب در پیاده‌سازی‌های عملی مبدل کرده‌است.



شکل (۷): ساختار الگوریتم ردیابی [۳۷] مبتنی بر یادگیری عمیق، استفاده از شبکهٔ بازگشتی در کنار شبکهٔ پیچشی به منظور استفاده از ویژگی‌های زمانی در خلال ردیابی

(Figure-7): The deep learning-based architecture of [37], using recurrent and convolutional networks to incorporate temporal features into the tracking

در پیاده‌سازی‌های عملی تشخیص پرنده در برابر اشیای هم‌شکل مانند برگ‌ها یا اشیا نازک یک چالش محسوب می‌شود. ترین و همکاران [۴۰] برای شناسایی پرندگان در اطراف توربین‌های بادی از ترکیب شبکهٔ پیچشی و شبکهٔ بازگشتی، به ترتیب برای استخراج ویژگی‌های ظاهری و حافظه‌ای برای ثبت تغییرات ظاهری هدف در طول زمان استفاده کرده‌اند. در این روش دوربین نظارتی ثابت است و هدف متحرک موجود در تصویر شناسایی می‌شود. این هدف برای چند قاب با الگوریتم ردیابی مجزایی دنبال شده و نتیجهٔ این ردیابی در ورودی شبکهٔ پیچشی و شبکهٔ بازگشتی استفاده می‌شود. شبکهٔ پیچشی ویژگی‌های مهم را یافته و شبکهٔ بازگشتی، پرنده‌بودن یا نبودن شیء شناسایی‌شده را با درنظرگرفتن تاریخچهٔ ظاهری و حرکتی آن تشخیص می‌دهد. نتایج این مقاله نشان داده‌است که استفاده از ترکیب این دو شبکه، از به‌کارگیری شبکهٔ پیچشی به‌تنهایی

[1] Reinforcement Learning
[2] Freeze

[DOI: 10.61186/jsdp.22.1.83]

مطلوب‌تر بوده و با بهره‌گیری از اطلاعات مکانی و زمانی عملکرد مناسب‌تری به‌دست خواهد آمد.

ارزیابی‌های یانگ و چان [۶۱] نشان می‌دهد که اعمال مستقیم ویژگی‌های استخراجی از شبکهٔ پیچشی به شبکهٔ بازگشتی باعث تضعیف اطلاعات مکانی می‌شود؛ بنابراین از شبکهٔ بازگشتی برای تولید پالایه استفاده شده‌است. به این منظور ابتدا تصویری از هدف ردیابی به‌عنوان الگو درنظر گرفته شده و پس از استخراج ویژگی‌های ظاهری آن به یک شبکهٔ بازگشتی داده می‌شود تا یک پالایهٔ ویژه برای آن تخمین بزند. در خلال ردیابی شبکهٔ پیچشی ویژگی‌های محدودهٔ جست‌وجو را استخراج کرده و با پالایهٔ تولیدشده توسط شبکهٔ بازگشتی در قاب‌های قبل هدف را در قاب کنونی شناسایی می‌کند. به منظور تطبیق ردیاب با تغییرات ظاهری هدف نتایج ردیابی در قاب‌های جدید به شبکهٔ بازگشتی پس‌خورد شده تا پالایه‌های تولیدشده با تغییرات سازگار شود. این روش تنها از ویژگی‌های ظاهری هدف استفاده می‌کند و به دقت مطلوبی دست یافته‌است، اما به‌دلیل استفاده از چند شبکهٔ پیچشی سرعت اجرای بالا مناسب پیاده‌سازی بلادرنگ ندارد.

آندروسکا و پوسنر [۶۲] با بهره‌گیری از شبکهٔ بازگشتی، الگوریتم ردیابی را برای مقابله با چالش انسداد اهداف و بهبود بینایی ربات ارائه داده‌اند. هدف این مقاله دست‌یابی به عملکردی مناسب در شناسایی و ردیابی اهداف با انسداد بسیار بالا بوده‌است. ارزیابی صورت‌گرفته در این مقاله در محیط‌های شبیه‌سازی‌شده و غیر واقعی صورت پذیرفته‌است. استفاده از شبکهٔ بازگشتی نتیجهٔ مطلوبی در حل چالش انسداد حاصل کرده‌است.

یانگ و چان [۱۶] شبکه‌ای با حافظهٔ پویا ارائه کرده‌اند تا هدف اولیه را به‌عنوان الگو درنظر گرفته و در ادامه این الگو را با تغییرات ظاهری هدف در حین ردیابی تطبیق دهد. در این مقاله شبکهٔ بازگشتی به‌عنوان کنترل‌کنندهٔ داده‌های موجود در حافظه مورد استفاده قرار گرفته‌است. ورودی این شبکه نقشه‌ای از ویژگی‌های ظاهری به‌دست‌آمده از شبکهٔ پیچشی بوده و خروجی آن سیگنال کنترلی برای خواندن و نوشتن بر حافظه است. از مکانیزم توجهی[1] نیز در ورودی شبکهٔ بازگشتی به منظور تمرکز بر الگو استفاده شده‌است. در هر قاب الگوی ردیابی به‌روزرسانی شده و ویژگی‌های ظاهری استخراج‌شده توسط شبکهٔ پیچشی برای به‌روزرسانی پارامترهای ردیاب در حافظه قرار می‌گیرد.

گن و همکاران [۶۳] ساختاری به طور کامل آموزش محور را برای ردیابی ارائه کرده‌اند به این صورت که برای آموزش ابتدا تا انتهای الگوریتم با هم در ارتباط بوده و برای انجام هدفی خاص آموزش می‌بینند. این نوع ساختار با یادگیری ویژگی‌های ظاهری و ویژگی‌های رفتاری و حرکتی هدف به صورت هم‌زمان، دقت مناسبی را در ردیابی به دست می‌آورد، اما به‌شدت به مجموعهٔ آموزشی مورد استفاده وابسته بوده و تنها قادر به ردیابی اهداف دیده‌شده و مشابه است. ساختار ارائه‌شده در این مقاله شبکهٔ پیچشی و شبکهٔ بازگشتی را ترکیب کرده، به‌طور کامل برون‌خط آموزش داده و در چند لایه از آن از سازوکار توجه استفاده کرده‌است. در ابتدا شبکهٔ پیچشی ویژگی‌های ظاهری هدف را استخراج کرده و به شبکهٔ بازگشتی ارسال می‌کند. این شبکه پیش‌بینی‌های مربوط به قاب‌های قبلی را به همراه ویژگی‌های ظاهری مربوط به قاب کنونی ترکیب کرده و محل حضور هدف و کادر مربوطه را برای قاب بعدی به‌صورت مستقیم پیش‌بینی می‌کند. ارزیابی الگوریتم بر روی شرایط محیط حقیقی و حرکات پیچیده‌تر و چالشی‌تر به کارهای آتی واگذار شده‌است.

الگوریتم ردیابی ارائه‌شده توسط ابراهیمی و همکاران [۶۴] به‌طور کامل آموزش‌محور بوده و از سه بخش اصلی تشکیل شده است. شبکهٔ بازگشتی توجهی برای تعیین محل جست‌وجو در هر قاب، شبکهٔ پیچشی برای استخراج ویژگی‌های ظاهری از تصویر و بخشی برای تعیین ویژگی‌های مهم به‌منظور یادگیری بهتر شبکه در حین فرایند آموزش مورد استفاده قرار می‌گیرد. ساختار ارائه‌شده این سه وظیفه را به‌خوبی انجام داده و می‌تواند بر روی یک مجموعهٔ آموزش‌دیده و توالی‌های مرتبط با آن و از پیش دیده‌شده را ردیابی کند. این چهارچوب برای ردیابی وابستگی بسیار زیادی به مجموعهٔ آموزشی دارد و این در حالی است که توانایی ردیابی اهداف غیر مرتبط را ندارد.

ونگ و همکاران [۶۵] برای حل چالش‌های مختلف ردیابی از شبکهٔ بازگشتی همراه با دو سازوکار توجه به‌منظور بهبود تفسیر ویژگی‌ها استفاده کرده‌اند. استخراج ویژگی‌های ظاهری توسط شبکهٔ پیچشی انجام می‌شود و اطلاعات مربوط به هر لایه از آن که شامل اطلاعات ابعادی و مکانی متفاوتی از هدف ردیابی است، در شبکهٔ بازگشتی مورد استفاده قرار می‌گیرد. این شبکه با ترکیب ویژگی‌های لایه‌های مختلف میزان شباهت میان نواحی نامزدشده در قاب کنونی و الگوی ردیابی در قاب قبلی را بررسی می‌کند. به‌منظور افزایش توانایی قدرت مقایسهٔ شبکهٔ بازگشتی و دنبال‌کردن ویژگی‌های اصلی مربوط به هدف ردیابی از سازوکار توجهی شامل دو بخش داخلی و خارجی استفاده شده‌است. بخش داخلی برای مقابله با چالش انسداد بوده و با مقایسهٔ اطلاعات مکانی هدف در قاب قبلی و ناحیه‌های انتخاب‌شده در قاب کنونی آن‌ها را امتیازدهی می‌کند. بخش

[1] Attentional Mechanism

[DOI: 10.61186/jsdp.22.1.83]

خارجی نیز برای افزایش کیفیت شناسایی و انتخاب بهتر محدودهٔ جست‌وجو با بهره‌گیری از اطلاعات زمانی در نظر گرفته شده‌است. نتایج نشان می‌دهد که ترکیب شبکهٔ بازگشتی و سازوکار توجه و همچنین آموزش هم‌زمان این دو باعث بهبود در عملکرد ردیاب در چالش‌های مختلف می‌شود، اما با افزایش پارامترها سرعت اجرا بسیار کاهش می‌یابد.

ردیابی بر مبنای شناسایی، یکی از روش‌های محبوب برای ردیابی است، اما مشکل اصلی آن خطا در دسته‌بندی اهداف با ظاهر مشابه است. فن و لینگ [۴۱] برای مقابله با این مشکل روشی را برای ردیابی عمومی ارائه کرده‌اند که در آن از اطلاعات زمانی و تغییرات ظاهری خود هدف در طول ردیابی برای تمایز آن از دیگر اهداف مشابه و نزدیک به آن استفاده شده‌است. برای این منظور از شبکهٔ بازگشتی برای مدل‌سازی ساختار هدف استفاده می‌شود و از ترکیب آن با شبکهٔ پیچشی برای بهبود مقاومت الگوریتم بهره‌گیری شده‌است. از آنجا که هر لایه از شبکهٔ پیچشی اطلاعات متفاوتی را ارائه می‌دهد برای هر لایه از آن یک شبکهٔ بازگشتی جداگانه درنظر گرفته شده‌است تا اطلاعات ظاهری هدف به‌خوبی استفاده شود. برای ردیابی در قاب کنونی چندین محل محتمل در اطراف محل حضور هدف در قاب پیشین انتخاب شده‌است و با استفاده از اطلاعات زمانی و امتیازدهی به وسیلهٔ شبکهٔ بازگشتی ردیابی صورت می‌گیرد. استفاده از ویژگی‌های لایه‌های مختلف شبکهٔ پیچشی و یک شبکهٔ بازگشتی جداگانه برای هریک از آن‌ها موجب افزایش دقت ردیابی شده است، اما سرعت اجرای الگوریتم به‌شدت کاهش یافته و هرگز توانایی اجرای بلادرنگ را نخواهد داشت.

۵-۱-۲-ردیاب‌های فاقد شبکهٔ بازگشتی

جین و همکاران [۶۶] الگوریتم ردیابی را به‌منظور بینایی بلادرنگ برای استفاده در پیاده‌سازی‌های عملی طراحی کرده‌اند. برای ردیابی طولانی با استفاده از شبکه‌های آموزش‌دیده و برای اهداف ازپیش تعیین‌نشده نیاز است که شبکه نسبت به تغییر ابعاد هدف و ویژگی‌های استخراج‌شده از آن مستقل بوده و با حرکت هدف و تغییر شکل آن همچنان پاسخ‌گو باشد؛ به این منظور از یادگیری عمیق در طراحی الگوریتم بهره‌گیری شده‌است و در آن ویژگی‌های ظاهری عمیق با استفاده از چندین لایه شبکهٔ پیچشی استخراج می‌شود. برای دسته‌بندی ویژگی‌ها و تعیین خروجی الگوریتم ردیابی از شبکهٔ عصبی RBF[1] استفاده شده‌است تا با دریافت تنها یک الگو از هدف، ردیابی صورت پذیرد. با کمک این شبکه یک نقشهٔ امتیازی[2] به‌دست می‌آید که با مقایسهٔ ویژگی‌های عمیق با الگوی ردیابی مکان جدید هدف تعیین می‌شود.

هنگ و همکاران [۶۷] یک شبکهٔ پیچشی را به‌صورت برون‌خط بر روی مجموعه‌ای بزرگ آموزش داده‌اند و ردیابی را با کمک این شبکهٔ پیچشی و همچنین نقشهٔ برجستگی‌ها[3] برای یادگیری هدفی خاص ارائه کرده‌اند. دسته‌بندی[4] ویژگی‌های به‌دست‌آمده از شبکهٔ پیچشی با استفاده از روش ماشین بردار پشتیبان[5] صورت می‌گیرد که از آن به‌منظور تمایز هدف از پس‌زمینه و آموزش برخط مدل ظاهری هدف استفاده شده‌است. این دسته‌بند در طول ردیابی آموزش می‌بیند تا با تغییرات ظاهری هدف تطبیق پیدا کند. خروجی این مرحله وارد یک شبکه برای تولید نقشهٔ امتیازی می‌شود. از آنجا که نقشهٔ برجستگی‌ها موقعیت مکانی هدف را با دقت بسیار مطلوبی تعیین می‌کند باعث افزایش دقت مکان‌یابی شده و کمک می‌کند تا مجموعه‌های پیکسلی موجود در تصویر توسط الگوریتم دسته‌بندی به شیوهٔ بهتری آموزش ببینند.

هلد و همکاران [۴۴] ردیابی قدرتمند و شاخص در حوزهٔ یادگیری عمیق را ارائه کرده‌اند. در این مقاله برای رسیدن به سرعت بالا در ردیابی از آموزش به‌طور کامل برون‌خط استفاده شده‌است. برای بهره‌گیری از ویژگی‌های زمانی و یافتن مکان جدید هدف در هر قاب دو برابر کادر محصورکنندهٔ آن در قاب پیشین و مکانی با همین خصوصیات، اما در قاب کنونی به‌عنوان ورودی به شبکه داده می‌شود. پس از استخراج ویژگی‌های ظاهری به‌وسیلهٔ شبکهٔ پیچشی این ویژگی‌ها در کنار یکدیگر ذخیره شده و در ادامه با استفاده از پنج لایهٔ به طور کامل متصل مکان هدف تعیین می‌شود. این الگوریتم با پرهیز از آموزش برخط در حین ردیابی به سرعت اجرای بسیار بالا دست یافته است که برای ردیاب‌های مطرح در حوزهٔ یادگیری عمیق موفقیتی چشمگیر به حساب می‌آید. این ردیاب در عین داشتن سرعت بالا در ردیابی از دقت قابل قبول، اما وابسته به مجموعهٔ آموزشی مورد استفاده برخوردار است.

برتینتو و همکاران [۱۷] روشی به‌طور کامل مبتنی بر شبکهٔ پیچشی را برای دنبال‌کردن هدفی دل‌خواه با استفاده از شبکهٔ سیامیس[6] [۶۸] ارائه کرده‌اند. این شبکه با بهره‌گیری از شبکهٔ عصبی پیچشی به بررسی میزان شباهت میان هدف تعیین‌شده و هدف در حال ردیابی می‌پردازد. برای حل مشکلات ناشی از کمبود دادهٔ آموزشی ابتدا یک شبکهٔ پیچشی به‌صورت برون‌خط آموزش داده می‌شود تا به‌صورت کلی شباهت‌های ظاهری میان اهداف مختلف را

[1] Radial Basis Fuction
[2] Confidence Map
[3] Saliency Map
[4] Classification
[5] Support Vector Machine (SVM)
[6] Siamese

[DOI: 10.61186/jsdp.22.1.83]

آموخته و سپس از آن در مرحلهٔ ردیابی استفاده شود. با استفاده از آموزش برون‌خط ابتدایی و عدم نیاز به آموزش برخط سرعت بسیار مناسبی به‌دست آمده‌است. در مرحلهٔ بعدی نیز شبکهٔ سیامیس آموزش داده می‌شود تا یک تصویر نمونه را به‌عنوان الگو در میان تصویر اصلی بزرگتر جست‌وجو کند. تمرکز این روش بر ویژگی‌های ظاهری است و تنها از مکان هدف قاب پیشین به‌عنوان شاخص زمانی استفاده می‌کند. روش ارائه‌شده در این مقاله به دقت مطلوب و سرعت بالایی در ردیابی دست‌یافته و مبنای توسعهٔ الگوریتم‌های ردیابی مبتنی بر تطبیق الگو شده‌است.

ما و همکاران [۴۳] ردیابی را با استفادهٔ ابتکاری از شبکهٔ پیچشی طراحی کرده‌اند که برای تشخیص اهداف مختلف آموزش دیده‌است. هر لایه از شبکهٔ پیچشی دارای اطلاعاتی منحصربه‌فرد است. خروجی لایه‌های آخر شبکهٔ پیچشی اطلاعات معنایی مهمی از هدف دارد که نسبت به تغییرات ظاهری آن مقاوم است؛ درحالی‌که مکان‌یابی به‌وسیلهٔ آن دقیق نخواهد بود. برخلاف این حالت لایه‌های ابتدایی شبکهٔ پیچشی موقعیت مکانی دقیقی را ارائه می‌دهند، اما نسبت به تغییرات ظاهری بسیار حساس‌اند. برای استفاده از تمامی این ویژگی‌ها بر روی هر لایه از شبکهٔ پیچشی یک پالایهٔ هم‌بستگی به‌منظور استفادهٔ دقیق‌تر از ویژگی‌های ظاهری هدف درنظر گرفته شده‌است. با نتیجهٔ به‌دست‌آمده از پالایه‌های چندگانه در تمامی لایه‌ها مکان هدف تعیین می‌شود. استفاده از لایه‌های مختلف موجب افزایش دقت الگوریتم می‌شود، اما از سوی دیگر بار پردازشی بالایی دارد و سرعت اجرا را کاهش می‌دهد. تنها شاخصهٔ زمانی مورد استفاده در این روش مکان هدف در قاب پیشین است.

ونگ و همکاران [۴۷] یک شبکهٔ پیچشی را به‌صورت برون‌خط آموزش داده‌اند و به‌دلیل کمبود دادهٔ آموزشی برچسب‌گذاری‌شده و دست‌یابی به دقت بالاتر از آموزش برخط نیز در خلال ردیابی استفاده می‌کنند. برای افزایش دقت در ابتدا شبکهٔ پیچشی برای شناسایی اهداف مهم، جداگانه آموزش داده شده‌است. این روش باعث تطبیق بهتر ردیاب با تغییرات ظاهری هدف در حین ردیابی شده‌است. به‌دلیل استفاده از آموزش برخط سرعت اجرای بلادرنگ به‌دست نیامده‌است.

یوآنو و همکاران [۶۹] یک الگوریتم ردیابی را برای مقابله با چالش تغییر ابعاد و چرخش هدف با ترکیب و یک‌پارچه‌سازی دو فرایند شناسایی و ردیابی پیشنهاد کرده‌اند. در بخش شناسایی، شیء موردنظر به‌صورت خودکار و با کمک نقشهٔ برجستگی‌ها و تفاوت آن با پس‌زمینه در هر قاب شناسایی و مکان‌یابی می‌شود. در بخش ردیابی نیز پالایهٔ کالمن به‌کار گرفته شده‌است تا با کمک پیش‌بینی‌های آن بر مبنای تاریخچهٔ زمانی حرکت هدف، تخمینی از موقعیت آن در قاب کنونی به‌دست آید و با تلفیق اطلاعات مکانی و زمانی عملکرد الگوریتم بهبود داده شود. ویژگی این ردیاب در مقایسه با دیگر روش‌ها نشان‌دهندهٔ موفقیت آن در ردیابی اهداف سریع، اما در پس‌زمینه‌های ساده است. به‌دلیل سادگی ساختار سرعت اجرا بسیار بالاست، اما در انسداد عملکرد مناسبی ندارد و هدف را در محیط‌ها و حرکت‌های پیچیده به راحتی از دست می‌دهد.

ژو و همکاران [۷۰] برای حل چالش‌هایی همچون تغییرات ظاهری شدید و انسداد، مدلی مولد با آموزش برون‌خط پیشنهاد کرده‌اند. در این مقاله با الهام از عمکرد الگوریتم‌های شناسایی تحت عنوان شبکهٔ پیشنهاددهندهٔ ناحیه[1] چندین بخش را به‌عنوان محل حضور هدف پیشنهاد داده و با دسته‌بندی یا رگرسیون محتمل‌ترین مکان برای حضور هدف شناسایی می‌شود. به‌دلیل استفاده از شناسایی‌های چندگانه برای مکان‌های مختلف در هر قاب حجم محاسباتی بسیار بالا و سرعت پردازش بسیار پایین است.

با وجود فعالیت‌های گستردهٔ انجام‌شده در زمینهٔ شناسایی و ردیابی موانع، اما همچنان انجام این کار خودکار و بلادرنگ یک چالش بزرگ محسوب می‌شود. بهاراتی و همکاران [۷۱] ردیابی با سرعت اجرای بالا با هدف تشخیص موانع و ردیابی هدفی مشخص جهت پیاده‌سازی بر روی یک پهپاد خودران طراحی کرده‌اند. برای این موضوع از ترکیب یک الگوریتم شناسایی با سرعت اجرای بالا با الگوریتم ردیابی سریع و مقاوم و برای شناسایی از روش نشان‌گر[2] و برای ردیابی از الگوریتم [۲] استفاده شده‌است. تمرکز این مقاله بر روی تقویت فرایند شناسایی اهداف است و در ردیابی به سرعت بالایی دست یافت، اما همچنان در چالش‌هایی نظیر انسداد ضعف وجود دارد؛ همچنین به‌منظور استفاده از ویژگی‌های زمانی تعیین محدودهٔ جست‌وجوی هدف از مدل حرکتی دینامیکی استفاده شده‌است.

ژو و همکاران [۷۲] الگوریتمی را برای ردیابی اهداف در تصاویر ماهواره‌ای طراحی کرده‌اند. این ردیاب بر مبنای شبکهٔ سیامیس است و با دریافت الگو و محدودهٔ جست‌وجو به ردیابی می‌پردازد. در تصاویر هوایی تمایز میان اشیا دشوار است و ویژگی‌های ظاهری کارایی کمتری نسبت به تصاویر مایل خواهند داشت؛ ازاین‌رو لازم است تا از ویژگی‌های حرکتی نیز برای ردیابی دقیق‌تر استفاده شود. در تصاویر هوایی و ماهواره‌ای با توجه به زیادبودن ارتفاع، حرکت اهداف به‌طور معمول بدون جهش و ساده است؛ ازاین‌رو در این مقاله مدل حرکتی خطی برای افزایش دقت ردیابی ارائه شده‌است. برای

[1] Region Proposal Network
[2] Salient Object Detection

[DOI: 10.61186/jsdp.22.1.83]

استخراج ویژگی‌های ظاهری از الگو و محدودهٔ جست‌وجو از معماری [۷] اصلاح‌شده استفاده شده‌است. به‌دلیل کوچک‌بودن اهداف در تصاویر با افزایش ابعاد لایه‌های پایانی این معماری نقشهٔ ویژگی با ابعاد بالاتری به‌دست‌آمد که موجب غنی‌ترشدن ویژگی‌های ظاهری می‌شود. برای یافتن هدف در محدودهٔ جست‌وجو از محدوده‌های پیشنهادی هم‌اندازه با الگو استفاده می‌شود و الگو با آن ترکیب‌شده و امتیاز هریک به‌دست می‌آید؛ سپس با دسته‌بندی مکان هدف و با رگرسیون کادر مربوطه تعیین می‌شود.

یکی از رویکردهای جدید در ردیابی مبتنی بر بینایی تولید حاشیه اطراف هدف تحت عنوان ماسک به جای کادر محصورکننده به‌عنوان خروجی الگوریتم است [۷۳]، [۷۴]؛ در همین راستا [۷۵] الگوریتمی را برای ردیابی پایدار ارائه کرده است که خروجی آن به‌صورت ماسک است. ساختار این ردیاب از چهار بخش اصلی استخراج ویژگی‌های ظاهری توسط شبکهٔ پیچشی، شاخهٔ بخش‌بندی تصویر، شاخهٔ مکان‌یابی هدف و رمزگشای بخش‌بندی تشکیل شده‌است. برای هر قاب ابتدا ویژگی‌های ظاهری استخراج شده و با پردازش مجدد به دو دسته ویژگی‌های بخش‌بندی و ویژگی‌های دسته‌بندی تبدیل می‌شوند؛ ویژگی نخست برای تعیین دقیق ماسک مربوط به هدف و ویژگی دوم برای تعیین موقعیت مکانی هدف در تصویر مورد استفاده قرار می‌گیرند؛ در ادامه هر یک از این ویژگی‌ها وارد شاخهٔ مربوط به خود شده و پردازش می‌شوند؛ به‌منظور افزایش مقاومت الگوریتم در برابر چالش‌هایی مانند ازدست‌دادن هدف و انسداد، مدلی حافظه‌دار برای تطبیق با تغییرات هدف درنظر گرفته شده‌است؛ این مدل برای هر دو شاخه مکان‌یابی و بخش‌بندی به صورت جداگانه طراحی شده‌است، اما ورودی هر دو از نتایج شاخهٔ مکان‌یابی استفاده می‌کنند؛ دلیل این امر مقاومت بالای این شاخه نسبت به تشخیص مکان هدف است. در ادامه خروجی هر دو شاخه با یکدیگر ترکیب شده و از طریق یک شبکهٔ پیچشی چندلایه رمزگشایی و ماسک مربوط به هدف تولید می‌شود. از خروجی این بخش نیز یک پس‌خورد به شاخهٔ بخش‌بندی وارد می‌شود تا موجب افزایش دقت این شاخه شود. ارزیابی‌ها نشان‌دهندهٔ دقت مطلوب این ساختار است و در عین پیچیدگی سرعت مناسب سی قاب بر ثانیه را بر روی سخت‌افزارهای گرافیکی به‌دست آورده‌است.

۵-۱-۳-ردیاب‌های کلاسیک

توجه به ویژگی‌های ظاهری برای یک ردیاب امری ضروری است، اما در صورت شباهت زیاد میان اهداف موجود در تصویر کافی نخواهد بود. دیکل و همکاران [۷۶] با توجه به ویژگی‌های حرکتی یک ردیاب مبتنی بر دینامیک حرکت هدف و با حجم محاسبات بهینه برای حل چالش‌هایی همچون شباهت ظاهری میان اهداف، ازدست‌دادن هدف به خاطر خروج آن از تصویر، انسداد، تقاطع مسیر حرکت اهداف و حرکات دوربین ارائه داده‌اند؛ این روش با هر هدف دل‌خواه و با هر ابعادی کار کرده و نیازی به دانش قبلی نسبت به دینامیک حرکت ندارد؛ درحالی‌که دینامیک حرکت اهداف را پیش‌بینی می‌کند. ویژگی اصلی این مدل افزایش توانایی ردیاب در دنبال‌کردن اهداف با ظاهر مشابه است و دقت قابل قبولی برای اهداف با ظاهر ثابت و بدون تغییر دارد؛ عدم توجه به ویژگی‌های ظاهری و استفاده‌نکردن از شبکهٔ پیچشی برای استخراج ویژگی‌ها از ضعف‌های روش ارائه شده‌است.

فن و همکاران [۳۹] از شبکهٔ پیچشی برای تخمین مکان و ابعاد هدف در قاب کنونی با داشتن مکان و ابعاد قبلی آن استفاده کرده‌اند. این ساختار به‌طور کامل مبتنی بر شبکهٔ پیچشی بوده و هر دو اطلاعات زمانی و مکانی را مدیریت می‌کند. برای نیل به این هدف یک شبکهٔ پیچشی با دادهٔ آموزشی محدود به‌صورت برون‌خط به‌گونه‌ای آموزش داده شده‌است که هر دو ویژگی مکانی و زمانی را با درنظرگرفتن دو قاب متوالی یاد بگیرد. در این مقاله یک شبکهٔ پیچشی وابسته به تغییرات[1] طراحی شده‌است تا مشکل خطای ردیابی در هنگام شبیه‌بودن اشیای موجود در تصویر با هدف موردنظر در محیط‌های شلوغ کاهش یابد. در هنگام انسداد و در هنگام حضور اهداف شبیه به هم شبکهٔ پیچشی به‌دلیل عدم آموزش صحیح ناشی از کمبود دادهٔ آموزشی و همچنین عدم حضور ساختار مکمل عملکرد مناسبی برای حل این چالش نداشته‌است.

دنلجان و همکاران [۷۷] اشاره کردند که بیشتر روش‌های ردیابی در تصاویر با پس‌زمینهٔ پیچیده و مبهم و همچنین با تغییرات ابعادی زیاد در هدف ردیابی عملکرد مطلوبی نخواهند داشت. عملکرد الگوریتم‌های ردیابی تنها در صورت وجود تغییرات کم در شاخص‌های ابعادی قابل اطمینان‌اند. در این مقاله برای حل چالش تغییر ابعاد هدف و دست‌یابی به تخمینی مقاوم از آن روشی جامع مبتنی بر روش‌های موجود در الگوریتم‌های شناسایی ارائه شده‌است. رویکرد ارائه‌شده با آموزش پالایهٔ هم‌بستگی تمایزی[2] بر مبنای هرم توصیفی[3] در استخراج ویژگی‌های تصویر بوده‌است؛ همچنین پالایه‌های مختلف برای انتقال و تخمین ابعاد آموزش دیده و باعث بهبود در ردیابی شده‌اند.

۵-۲-ردیاب چندهدفه

۵-۲-۱-استفاده از شبکهٔ بازگشتی در ردیاب

صادقیان و همکاران [۵] ردیابی چندهدفه با آموزش برخط برای استفاده از وابستگی‌های زمانی با مدت طولانی پیشنهاد

[1] Shift Variant CNN
[2] Discriminative Correlation Filter
[3] Scale Pyramid Representation

[DOI: 10.61186/jsdp.22.1.83]

کرده‌اند. در این مقاله برای حل چالش‌هایی همچون انسداد، تمایز اهداف با ظاهر شبیه به هم، حفظ اطلاعات ظاهری به‌دست‌آمده، پیش‌بینی حرکت هدف و همچنین درنظرگرفتن تعاملات میان اهداف و محیط پیرامون آن از شبکهٔ بازگشتی استفاده شده‌است. شبکهٔ بازگشتی آموزش می‌بیند تا وابستگی‌های زمانی طولانی‌مدت را در میان نشانه‌های چندگانهٔ ظاهری، حرکتی و تعاملی حفظ کند؛ همچنین در این مقاله برای کاهش محدودهٔ جست‌وجو، افزایش سرعت پردازش و بهبود دقت الگوریتم از مدل حرکتی نیز استفاده شده‌است. در این الگوریتم به‌دلیل استفادهٔ مستقیم از ویژگی‌های ظاهری به‌دست‌آمده از شبکهٔ پیچشی در شبکهٔ بازگشتی و تحمیل بار محاسباتی بالا به الگوریتم سرعت اجرا بسیار پایین است.

یکی از چالش‌های اساسی در ردیابی چندهدفه، تغییر میان شناسهٔ[1] اهداف در حال ردیابی است و دلیل اصلی آن انسداد هدف به وسیلهٔ موانع موجود در صحنه است. روش‌های مختلفی برای مقابله با انسداد پیشنهاد می‌شود که یکی از آن‌ها افزودن یک مدل حرکتی خطی یا غیرخطی به الگوریتم ردیابی است. مدل خطی به‌دلیل ساده‌بودن در موارد چالش‌برانگیز عملکرد مناسبی ندارد. شبکهٔ بازگشتی یکی از مدل‌های غیرخطی با قابلیت آموزش بر روی داده‌های آموزشی مختلف است. بابایی و همکاران [۷۸] با بهره‌گیری از شبکهٔ بازگشتی با ساختاری جدید روشی برای مقابله با انسداد پیشنهاد کرده‌اند. در این مقاله معیاری برای تشخیص انسداد هدف درنظر گرفته شده و در صورت بروز آن به جای استفاده از الگوریتم شناسایی برای تشخیص و ردیابی اهداف، از مدل حرکتی مبتنی بر شبکهٔ بازگشتی استفاده می‌شود. این شبکه با دریافت سرعت تغییرات مرکز و ابعاد کادر محصورکنندهٔ هدف در طول ردیابی به پیش‌بینی کوتاه‌مدت حرکت هدف در قاب‌های آتی می‌پردازد. مدل مطرح‌شده در این مقاله در فصل مربوط به مدل‌های حرکتی مورد بررسی قرار خواهد گرفت.

میلان و همکاران [۵۵] الگوریتمی بر مبنای شبکهٔ بازگشتی برای ردیابی چندهدفه ارائه کرده‌اند. در این مقاله با الهام از روش بیزین یک شبکهٔ بازگشتی دارای ویژگی‌هایی از جمله پیش‌بینی حرکت اهداف، ایجاد ارتباط میان نتایج شناسایی، به‌روزرسانی محدودهٔ حضور اهداف و همچنین تشخیص ورودوخروج اهداف در تصویر در ساختار یک شبکهٔ یک‌پارچه و با اجرا به‌صورت پایان‌به‌پایان پیشنهاد شده‌است. از ویژگی‌های مهم مدل ارائه‌شده در این مقاله بی‌نیازی از دانش قبلی نسبت به شیوهٔ حرکت هدف است. در الگوریتم پیشنهادی از ویژگی‌های ظاهری استفاده نشده و تنها با استفاده از شبکهٔ بازگشتی به‌عنوان مدل حرکتی ردیابی صورت گرفته‌است؛ این امر هرچند موجب افزایش سرعت اجرای الگوریتم شد، اما در مقایسه با دیگر الگوریتم‌های چندهدفه نتایج بهتری در دقت ردیابی به‌دست نیامد. ویژگی این مقاله مدل حرکتی غیرخطی آن است که در بخش مربوط به مدل‌های حرکتی مورد بررسی قرار خواهد گرفت.

۵-۲-۲-ردیاب‌های فاقد شبکهٔ بازگشتی

بیلی و همکاران [۷۹] رویکردی عملی برای ردیابی چندهدفه به‌صورت بلادرنگ ارائه کرده‌اند. در این مقاله اشاره شده‌است که توانایی الگوریتم شناسایی عاملی بسیار تأثیرگذار است؛ به‌طوری که می‌توان بیست درصد از عملکرد ردیابی را وابسته به آن دانست. ردیاب پیشنهادشده برخلاف استفاده از روش‌های ساده نظیر پالایهٔ کالمن، به‌عنوان مدل حرکتی برای محاسبهٔ سرعت حرکت و تعیین مسیر حرکت هدف به نتیجهٔ مطلوب و قابل رقابتی دست یافته‌است. این مقاله از جمله ردیاب‌های مطرح‌شده در میان الگوریتم‌های ردیابی چندهدفه بوده و در پیاده‌سازی‌های عملی مورد استفاده قرار می‌گیرد.

محمودی و همکاران [۴۶] توسعهٔ ردیابی چندهدفه با حجم محاسبات کم، اما با دقت بالا را به‌عنوان هدف خود مطرح کرده‌اند. محاسبهٔ پیوستگی[2] یکی از مهم‌ترین مراحل در ردیابی چندهدفه است که برای آن روش‌های مختلفی همچون مدل‌سازی ظاهری، حرکتی، تعاملی و همچنین دریافت ویژگی‌های تصویر با شبکهٔ پیچشی وجود دارد. برای شناسایی و دریافت این ویژگی‌ها از شبکهٔ پیچشی [۵۹] به‌دلیل قدرت و مقاومت بالا استفاده شده‌است. در این مقاله با استفاده از اطلاعات مربوط به قاب‌های پیشین و با توجه به این امر که انسان‌هایی که گروهی و در یک دسته حرکت می‌کنند رفتار تعاملی مشابه با محیط و افراد اطراف خود دارند، از یک روش تعاملی جدید برای ردیابی دسته‌ای اهداف استفاده شده‌است. تمامی فرض‌های درنظر گرفته‌شده برای مدل‌های حرکتی و تعاملی بر این مبنا بوده‌است که حرکتی ناگهانی و تغییرات شدیدی در سرعت و شتاب اتفاق نخواهد افتاد. دقت ردیاب ارائه‌شده مطلوب است، اما سرعت ردیابی به‌دلیل حجم محاسبات بالا بلادرنگ نیست.

چو و همکاران [۴۹] چهارچوبی برخط برای ردیابی چندهدفه با استفادهٔ بهینه از روش‌های مختلف ردیابی بر مبنای شبکهٔ پیچشی ارائه کرده‌اند. برای کاهش اشتباهات ناشی از انسدادهای متناوب و ارتباطات میان اهداف از مکانیزم توجهی، هم در زمینهٔ مکانی و هم زمانی استفاده شده‌است؛ علاوه‌بر این برای افزودن اطلاعات حرکتی هدف به

[1] ID Switch

[2] Affinity

[DOI: 10.61186/jsdp.22.1.83]

الگوریتم از یک مدل حرکتی ساده نیز استفاده شده‌است. برای حل چالش انسداد از نقشهٔ امتیازی و آموزش برخط شبکهٔ پیچشی برای به‌روزرسانی مدل ظاهری استفاده می‌شود. به‌دلیل به‌روزرسانی برخط سرعت اجرای الگوریتم بسیار پایین است.

خان و همکاران [۸۰] ردیابی چندهدفه بر مبنای شناسایی ارائه کرده‌اند. از آنجا که هرچه الگوریتم شناسایی قوی‌تر باشد، نتیجهٔ ردیابی مطلوب‌تر خواهد بود به بهبود عملکرد الگوریتم شناسایی پرداخته شده‌است. بدین منظور از الگوریتم قدرتمند [۵۹] برای شناسایی با دقت و سرعت بالا استفاده می‌شود. در مرحلهٔ نخست در هر قاب هدف مورد نظر شناسایی شده و در مرحلهٔ دوم اهداف شناسایی‌شده در قاب کنونی و پیشین مطابقت داده می‌شوند. این امر برای دست‌یابی به مدل حرکتی اهداف است و دسته‌بندی بهتر اشیا و همچنین ارتباط میان قاب‌ها را افزایش می‌دهد.

ووک و همکاران [۸۱] روش مطرح در [۷۹] را به‌عنوان یک روش ساده و مؤثر و با قابلیت استفاده در پیاده‌سازی‌های عملی برای ردیابی چندهدفه توسعه داده‌اند. در این مقاله با انتقال بیشتر محاسبات به بخش برون‌خط سرعت ردیابی افزایش یافته و از طرفی با حفظ اطلاعات ظاهری هدف، ضمن وجود چالش‌هایی همچون انسداد و تغییر شناسهٔ اهداف، ردیابی در مدت زمان طولانی‌تری صورت می‌پذیرد. در [۷۹] به‌دلیل سادگی مدل پیشنهادی و پاسخ‌گوبودن آن تنها در شرایط بدون چالش همچون حرکات ساده و سرعت‌های کم، در حین ردیابی جابه‌جایی زیادی میان اهداف اتفاق افتاده و درنتیجه در انسدادها عملکرد مطلوبی ندارد. در [۸۱] با اضافه‌کردن ویژگی‌های ظاهری به این مدل بر این مشکل غلبه شده‌است. در این مقاله با درنظرگرفتن هشت متغیر حالت شامل مرکز کادر محصورکنندهٔ هدف، نسبت ابعاد کادر، ارتفاع و سرعت حرکت آن در دو جهت عمودی و افقی و همچنین با بهره‌گیری از پالایهٔ کالمن مدل حرکتی ایجاد شده‌است؛ درنهایت با استفاده از شبکهٔ پیچشی چالش‌ها بهتر مدیریت شده و تغییر شناسه میان اهداف تا ۴۵ درصد بهبود یافته‌است.

نام و هان [۸۲] به منظور غلبه بر کمبود داده‌های آموزشی، ترکیبی از روش آموزش برون‌خط و برخط را پیشنهاد کرده‌اند. در این روش در مرحلهٔ برون‌خط شبکهٔ پیچشی با استفاده از داده‌های آموزشی موجود به صورت ابتدایی آموزش داده می‌شود تا اهداف عمومی موجود در مجموعهٔ آموزشی را شناسایی کند؛ سپس برای هر فیلم داده‌شده به شبکه شاخه‌ای مختص به آن تولید می‌شود تا در حین ردیابی بر روی قاب‌های ورودی به صورت برخط آموزش ببیند. به‌دلیل استفاده از آموزش برخط سرعت ردیابی بسیار پایین است، اما عملکرد مناسبی در ردیابی و حل چالش‌های آن به‌دست‌آمده‌است.

جانگ و همکاران [۸۳] به توسعه [۸۲] پرداخته و موفق شده‌اند با حفظ دقت ردیابی و با کاهش حجم محاسبات الگوریتم سرعت آن را ۲۵ برابر افزایش دهند؛ برای این منظور از شبکهٔ پیچشی با سرعت اجرای بالا و همچنین دقت بیشتر در استخراج ویژگی‌ها (به‌دلیل آموزش بر روی داده‌های متمایز و با تعداد بیشتر) استفاده شده که کمک شایانی به افزایش سرعت اجرای الگوریتم کرده‌است. [۸۲] به‌دلیل پردازش جداگانهٔ هر مکان محتمل برای حضور هدف همچنان حجم محاسبات بسیار بالا است که در نتیجهٔ آن سرعت ردیابی از حد بلادرنگی خارج می‌شود.

شوآی و همکاران [۸۴] ردیابی چندهدفه با توانایی بالا ارائه کرده‌اند که از ترکیب شبکهٔ سیامیس و مدل حرکتی برای ردیابی استفاده می‌کند. این الگوریتم از معماری [۱۲] برای استخراج ویژگی‌های حرکتی از ناحیه‌های با احتمال وجود هدف استفاده می‌کند. در ادامه با کمک شبکهٔ سیامیس برای تطبیق ویژگی‌های ظاهری و همچنین مدل حرکتی برای تطبیق ویژگی‌های مکانی به هر ناحیه، امتیازی اختصاص می‌دهد. اگر این امتیاز از آستانه‌ای بالاتر باشد هدف تأیید می‌شود و در خروجی الگوریتم قرار می‌گیرد. نتایج حاکی از آن است که افزودن مدل حرکتی تأثیر به‌سزایی در افزایش توانمندی ردیاب دارد؛ به طوری که بر روی معیارهای ارزیابی تا پنج درصد موجب بهبود عملکرد شده‌است.

۵-۲-۳-ردیاب‌های کلاسیک

پلگرینی و همکاران [۸۵] برای ردیابی چندهدفه در محیط‌های شلوغ روشی مبتنی بر مدلی پویا ارائه کرده‌اند. دلیل استفاده از این مدل دستیابی به محدودهٔ جست‌وجوی کوچکتر برای پیش‌بینی دقیق‌تر است. در این مقاله اشاره شده‌است که مدل‌های قدیمی تعاملات میان اهداف را درنظر نمی‌گیرند. چنین رویکردی جنبه‌های مهم رفتاری انسان‌ها را نادیده می‌گیرد. افراد در حرکت به سمت مقصدشان محیط پیرامون خود را درنظر می‌گیرند، برخورد با موانع را پیش‌بینی کرده و مسیر خود را در مراحل اولیه تنظیم می‌کنند. در این مقاله الگوی رفتار اجتماعی پویا با الهام از مدل‌هایی که برای شبیه‌سازی جمعیت تولید شده‌اند ارائه شده‌است. در طرح پیشنهادی یک انرژی پتانسیل برای عابر در مکانی که ایستاده فرض شده و یک نقطهٔ تخمینی برای نزدیکترین جهت‌گیری آن پیش‌بینی می‌شود. با استفاده از این نقطه به‌عنوان قوای محرکه برای تصمیم‌گیری استفاده شده و مسیری بهینه برای اهداف در

[DOI: 10.61186/jsdp.22.1.83]

نظر گرفته می‌شود. استفاده از این مدل موجب کاهش میزان پردازش شده و سرعت ردیابی بهبود یافته‌است. گفتنی است که این مدل برای تصاویر هوایی مناسب است.

۶- مدل‌های حرکتی در ردیابی

همان‌طورکه در بخش ۳-۱ ذکر شد بخش قابل توجهی از توسعه‌های ایجادشده در الگوریتم‌های ردیابی معطوف به مدل ظاهری است؛ ازاین‌رو مدل حرکتی یکی از بخش‌های مغفول‌مانده در ردیابی بوده و در مقایسه با مدل ظاهری بسیار کمتر به آن پرداخته شده‌است [۸۶] و این امر در ردیاب‌های تک‌هدفه بیشتر به چشم می‌خورد. مدل حرکتی نمایشی از نحوهٔ حرکت یک هدف با استفاده از تاریخچهٔ حرکتی آن در قاب‌های پیشین است که معرف ویژگی ذاتی و رفتاری آن هدف است. استفاده از این مدل در الگوریتم‌های ردیابی و تخمین محل حضور هدف در قاب آتی موجب کاهش ابعاد محدودهٔ جست‌وجو و درنتیجه کاهش حجم محاسبات می‌شود؛ همچنین با دقیق‌ترشدن محدودهٔ جست‌وجو در قاب آینده توانایی الگوریتم ردیابی در یافتن هدف نیز افزایش می‌یابد؛ ازاین‌رو با توجه به اهمیت مدل‌های حرکتی در حفظ ویژگی‌های زمانی و مکانی، مقالات شاخص موجود در این زمینه دقیق‌تر مورد بررسی قرار گرفته‌اند.

علاوه‌بر مدل حرکتی، دستهٔ دیگری از روش‌های ردیابی چندهدفه به مدل‌سازی رفتار اهداف و بررسی تعاملات میان آن‌ها می‌پردازند؛ برای مثال انسان‌هایی که در کنار هم مشغول پیاده‌روی‌اند به‌عنوان یک گروه درنظر گرفته شده و فرض می‌شود تا پایان در کنار یکدیگرند و سرعت حرکتی هم‌سانی خواهند داشت؛ همچنین رفتار انسان‌ها در عبور از کنار دیگر افراد موجود در صحنه و یا ممانعت از برخورد با موانع نیز یکی از فرضیات مطرح در این زمینه است. با استفاده از فرضیاتی از این دست مدل‌های مختلفی به منظور پیش‌بینی شیوهٔ رفتاری اهداف موجود در تصویر ارائه شده‌است[۴۶]. به‌دلیل پیچیدگی و حجم پردازش بالا کاربرد این روش‌ها محدود است و در این مقاله بررسی نشده‌اند.

در ردیابی چندهدفه به‌دلیل تعدد اهداف استفاده از مدل ظاهری به‌تنهایی بار محاسباتی سنگینی را ایجاد می‌کند. این امر هر چند موجب افزایش دقت الگوریتم می‌شود، اما از طرف دیگر استفادهٔ بلادرنگ از آن را غیرممکن می‌سازد؛ ازاین‌رو در الگوریتم‌های چندهدفه توجه بیشتری به مدل‌های حرکتی سادهٔ خطی و یا پیچیدهٔ غیرخطی شده‌است[۷۹]. علت این امر را می‌توان به ساختارهای این دسته از الگوریتم‌ها ارتباط داد. روش اصلی در ردیابی چندهدفه به این صورت است که یک الگوریتم شناسایی در هر قاب به تعیین اهداف موجود در صحنه می‌پردازد. اهداف شناسایی‌شده در دو دسته هدف جدید و هدف در حال ردیابی درنظر گرفته می‌شوند. در مرحلهٔ بعد با استفاده از معیارهایی همچون هم‌پوشانی محدوده‌ها[1] به‌منظور تعیین میزان تشابه کادر شناسایی‌شده و کادر پیش‌بینی‌شده به‌وسیلهٔ مدل حرکتی، نتایج شناسایی و اهداف موجود در صحنه با یکدیگر تطبیق داده می‌شوند؛ در این مرحله با استفاده از روش‌های مختلف مشخص می‌شود که هر یک از شناسایی‌های صورت‌گرفته چه ارتباطی با قاب‌های قبلی دارند، اهداف جدید مشخص و اهداف خارج‌شده از صحنه از فهرست حذف می‌شوند[۷۶].

با داشتن تخمینی از محل حضور هدف به‌کمک مدل‌های حرکتی و تطبیق نتایج شناسایی به اهداف در حال ردیابی حجم محاسبات کاهش یافته و درنتیجه سرعت اجرای الگوریتم بالا می‌رود؛ درنتیجه حضور یک مدل برای تخمین محل حضور هدف با کمک نتایج قاب‌های قبلی در الگوریتم‌های چندهدفه ضروری است[۸۱]، اما از طرف دیگر در الگوریتم‌های تک‌هدفه به‌دلیل حجم کم محاسبات ناشی از وجود تنها یک هدف برای ردیابی، تقویت و توسعهٔ استخراج ویژگی‌های ظاهری نحوهٔ استفاده از آن‌ها مورد توجه بیشتری بوده و ویژگی‌های حرکتی هدف درنظر گرفته نمی‌شوند[۳۷].

در کل می‌توان مدل‌های حرکتی را به دو دستهٔ خطی و غیرخطی تقسیم‌بندی کرد. مدل‌های خطی به‌دلیل سادگی، در حرکات پرچالش عملکرد مطلوبی نداشته و شرط جواب‌گویی آن‌ها وجود شرایطی نظیر حرکات آرام و یا با سرعت و شتاب محدود است. برای غلبه بر این موضوع مدل‌های غیرخطی با استفاده از شبکه‌های عصبی بازگشتی ارائه شده‌اند که توانایی‌های بیشتری در مقایسه با مدل‌های خطی دارند[۷۸]. مدل‌های خطی بیشتر شامل چند متغیر حالت همچون ابعاد کادر محصورکننده، نسبت ابعاد، مختصات مرکز کادر، سرعت و شتاب تغییر پارامترهای کادر با توجه به قاب قبل و یا میانگین چندین قاب و پارامترهایی مشابه که همگی مربوط به کادر محصورشده به دور هدف است هستند[۶۹]. با استفاده از این متغیرها و بهره‌گیری از الگوریتم‌هایی همچون پالایهٔ کالمن به پیش‌بینی محل حضور هدف در قاب‌های بعدی پرداخته می‌شود.

در مدل‌های غیرخطی با استفاده از شبکه‌های عصبی بازگشتی به پیش‌بینی محل حضور هدف پرداخته می‌شود.

[1] Intersection Over Union (IOU)

[DOI: 10.61186/jsdp.22.1.83]

در ورودی این شبکه‌ها از متغیرهای اشاره‌شده و همچنین در برخی موارد ویژگی‌های ظاهری استخراج‌شده به‌وسیلهٔ شبکه‌های پیچشی استفاده می‌شود[۵]. در بررسی‌های به‌عمل‌آمده مشخص شده‌است که اعمال مستقیم ویژگی‌های ظاهری به‌عنوان ورودی به شبکهٔ بازگشتی هزینهٔ محاسباتی بسیار بالایی دارد و سرعت اجرای الگوریتم به‌شدت کاهش می‌یابد[۵۵]؛ ازاین‌رو در بسیاری از موارد از اعمال مستقیم ویژگی‌های ظاهری خودداری شده و با استفاده از لایه‌های کامل متصل ابعاد بردار ویژگی‌ها کاهش یافته[1] و سپس به شبکه وارد می‌شوند؛ با این روش اگرچه از دقت ردیابی کاسته می‌شود، اما سرعت اجرای بالایی در پردازش به‌دست می‌آید[۷۸].

۶-۱-مدل‌های حرکتی خطی

در [۷۹] به‌منظور کاهش حجم محاسبات و افزایش سرعت اجرا از الگوریتم شناسایی تنها در قاب‌های خاصی استفاده شده و در قاب‌های میانی با استفاده از پالایهٔ کالمن با مدل سرعت ثابت محل حضور هدف تخمین زده می‌شود و ردیابی صورت می‌گیرد. مدل حرکتی ارائه‌شده شامل هفت شاخصه از کادر محصورکننده است: موقعیت مرکز کادر، نسبت ابعاد، مساحت و همچنین سرعت تغییرات مربوط به مرکز و ابعاد. در هر قابی که شناسایی انجام گیرد کادر شناسایی‌شده پارامترهای مدل را به‌روزرسانی می‌کند و در غیر این صورت مدل با پارامترهای قبلی و با سرعت خطی ثابت به کار خود ادامه می‌دهد. این مدل موجب افزایش سرعت پردازش الگوریتم شده و قابلیت اجرای بلادرنگ را برای آن ایجاد کرده‌است، اما از طرفی سادگی مدل و همچنین فرض سرعت خطی ثابت در آن دقت بالایی را در حالت‌های پیچیده به دست نمی‌دهد، به‌گونه‌ای که در صورت وجود حرکات ناگهانی مدل کارایی خود را از دست خواهد داد.

در [۸۱] به بهبود عملکرد [۷۹] در سرعت اجرا و دقت آن پرداخته شده‌است. این بهبود از یک‌سو با ایجاد تغییراتی در روند شناسایی و تخصیص اطلاعات مربوط به هر هدف و از سوی دیگر با پیشنهاد مدلی خطی با پارامترهایی بیشتر نسبت به [۷۹] صورت می‌گیرد. مدل خطی پیشنهادی در این مقاله شامل مختصات مرکز کادر، ارتفاع، نسبت ابعاد و همچنین سرعت تغییرات تمامی این پارامترها است. مدل پیشنهادی موجب بهبود عملکرد الگوریتم می‌شود، اما همچنان در حرکات چالشی و ناگهانی اهداف به‌ویژه در انسداد ضعیف است.

[۶۹] ردیابی را به‌منظور پیاده‌سازی عملی بر روی پهپادهای خودران ارائه کرده‌است. در این مقاله به‌منظور دست‌یابی به سرعت اجرای بلادرنگ از الگوریتم شناسایی نشان‌گر به‌دلیل سرعت اجرای بالا و در کنار آن از یک مدل حرکتی برای ردیابی بهره‌گیری شده‌است. این مدل حرکتی از شش متغیر حالت به‌عنوان ورودی استفاده می‌کند. این متغیرها شامل مختصات مرکز کادر، سرعت تغییرات آن‌ها و همچنین ابعاد کادر هستند. این مدل موجب کاهش ابعاد محدودهٔ جست‌وجو و افزایش دقت مکان‌یابی شده‌است. سرعت بالا در ردیابی و تشخیص موانع از قابلیت‌های این ردیاب است، اما به‌دلیل ضعف‌های موجود در الگوریتم شناسایی در چالش انسداد عملکرد ضعیفی دارد؛ همچنین مدل به کارگرفته‌شده توانایی مدیریت حرکات ناگهانی و سریع را ندارد و در صورت بروز این چالش‌ها هدف ردیابی از دست خواهد رفت.

[۴۹] به عدم توجه به مدل‌های حرکتی در ردیابی تک‌هدفه اشاره کرده و به بررسی مزایای استفاده از مدل‌های حرکتی پرداخته‌است. در این مقاله با ترکیب ردیاب‌های تک‌هدفه و افزودن مدل حرکتی به آن‌ها الگوریتم ردیابی چندهدفه ساخته شده‌است. در این مقاله از یک مدل حرکتی خطی ساده با سرعت ثابت استفاده شده‌است تا مرکز و ابعاد محدودهٔ جست‌وجو را در قاب بعد مشخص کند. به این صورت که اختلاف مراکز دو کادر متوالی محاسبه و به قاب آخر اضافه می‌شود. در این مدل نسبت ابعاد کادر محصورکننده ثابت فرض شده است؛ ازاین‌رو توانایی مقابله با حرکات ناگهانی و سریع را ندارد.

[۸۶] دو مدل حرکتی خطی برای بهبود عملکرد الگوریتم‌های ردیابی مبتنی بر بینایی ارائه کرده‌است. در این مقاله به این نکته اشاره شده که با پیشرفت چشم‌گیر در استخراج و استفاده از ویژگی‌های ظاهری عمیق از تصاویر، ویژگی‌های حرکتی در ردیابی تک‌هدفه بسیار کمتر از ویژگی‌های ظاهری مورد توجه قرار گرفته‌است؛ ازاین‌رو مدل‌های حرکتی مستقلی ارائه شده‌اند تا به ردیاب‌های مختلف اضافه شود.

مدل نخست بر این فرض استوار است که حرکت هدف به‌طور کامل تصادفی بوده و از مدل مشخصی تبعیت نمی‌کند. در این مدل با بررسی آماری یک دادهٔ آموزشی و محاسبهٔ چگالی احتمال سرعت حرکت هدف پارامترهای توزیع آماری تعیین شده و در ردیابی مورد استفاده قرار می‌گیرند. به این صورت که محل هدف در قاب بعدی با افزودن مقدار تصادفی سرعت حرکت به مکان کنونی آن به‌دست خواهد آمد؛ هر چند این مدل موجب تقویت عملکرد ردیابی حتی به میزان کمی شده‌است، اما پارامترهای آن تحت تأثیر مجموعهٔ آموزشی است و در تمامی حالات و تحت چالش‌های مختلف ردیابی قابل اطمینان نخواهد بود.

[1] Down Sampling

[DOI: 10.61186/jsdp.22.1.83]

در مدل دوم مختصات مکان هدف در تصویر به‌عنوان دو سری زمانی درنظر گرفته شده‌است و الگوی حرکت هدف با استفاده از بردار خودتنظیم پیش‌بینی می‌شود. در این مدل معادلات مربوطه به فرم ضرب ماتریس ضرایب ثابت در مختصات مکانی هدف در قاب کنونی نوشته شده‌است و حاصل آن، مکان هدف در قاب بعدی خواهد بود. پارامترهای ماتریس ضرایب در سی قاب ابتدایی هر فیلم تعیین شده‌اند و در ادامه بدون تغییر مورد استفاده قرار می‌گیرند؛ این بدین معناست که در ابتدای هر فیلم مکان حقیقی هدف باید مشخص باشد و در آموزش مدل مورد استفاده قرار بگیرد که ضعف بزرگی برای این مدل حرکتی است. بدیهی است که مدل دوم به نتایج بهتری نسبت به مدل نخست دست یافته‌است.

[۷۲] یک مدل حرکتی خطی برای بهبود ردیابی در تصاویر هوایی و ماهواره‌ای ارائه کرده‌است؛ این مدل از نتایج شبکهٔ سیامیس تا قاب مشخص و قابل اعتمادی به‌عنوان ورودی خود استفاده می‌کند. ویژگی حرکتی مورد استفاده در این مدل سرعت حرکت هدف است و با محاسبهٔ اختلاف مکان هدف میان دو قاب متوالی محاسبه می‌شود. مقدار سرعت میانگین در دو جهت افقی و عمودی برای تعداد قاب مشخص به‌دست می‌آید. با داشتن مکان هدف در قاب فعلی و افزودن سرعت میانگین، مکان آن در قاب بعدی قابل پیش‌بینی خواهد بود. با توجه به این پیش‌بینی اگر نتیجهٔ شبکهٔ سیامیس در محدودهٔ مطمئنی از آن باشد مورد تأیید قرار می‌گیرد و به‌عنوان خروجی ردیاب درنظر گرفته می‌شود؛ در غیر این صورت خروجی مدل حرکتی به‌عنوان نتیجهٔ نهایی مورد استفاده قرار می‌گیرد. طبق ارزیابی‌ها با استفاده از ۱۴۰ قاب به‌عنوان تعداد قاب مشخص برای تعیین سرعت میانگین بهترین نتیجه حاصل شده‌است و با استفاده از مدل حرکتی نتایج تا ده درصد در دقت و مقاومت افزایش داشته‌اند.

۶-۲-مدل‌های حرکتی غیرخطی

با توجه به بررسی‌های صورت‌گرفته مدل‌های غیرخطی ارائه‌شده در مقالات بیشتر مبتنی بر روش‌های یادگیری عمیق هستند. در میان مقالات بررسی‌شده مدل‌های غیرخطی پیچیده‌ای نیز موجود است که از روش‌هایی غیر از یادگیری عمیق در توسعهٔ آن‌ها استفاده شده‌است [۷۶، ۸۷] که به‌دلیل پیچیدگی بالا و عدم کارایی قابل توجه مورد بررسی قرار نگرفته‌اند.

[۵] از شبکهٔ LSTM برای ذخیرهٔ تاریخچه‌ای از ویژگی‌های ظاهری و حرکتی در ردیابی چندهدفه استفاده کرده‌است. در این مدل به‌دلیل استفادهٔ مستقیم از ویژگی‌های ظاهری به‌دست‌آمده از شبکهٔ پیچشی به‌عنوان ورودی شبکهٔ بازگشتی، حجم پردازش بسیار بالا و سرعت اجرا بسیار پایین است. در بخش مدل حرکتی ابتدا سرعت حرکت هدف از طریق محاسبهٔ اختلاف موقعیت آن در دو قاب متوالی به‌دست می‌آید. این مقدار به‌همراه سرعت‌های مربوط به قاب‌های قبلی به‌عنوان ورودی شبکهٔ بازگشتی مورد استفاده قرار می‌گیرند؛ درنهایت خروجی این شبکه به‌همراه ویژگی‌های ظاهری در کنار هم قرار گرفته و به یک شبکهٔ بازگشتی دیگر وارد می‌شوند. وظیفهٔ این شبکه تعیین کادر محصورکنندهٔ هدف در قاب کنونی است. عدم استفاده از ورودی‌های مختلف، مستقل‌نبودن مدل حرکتی از دیگر بخش‌های الگوریتم و همچنین سرعت اجرای بسیار پایین از ضعف‌های مدل ارائه‌شده در این مقاله هستند.

[۵۵] به استفادهٔ هم‌زمان از شبکهٔ بازگشتی و LSTM در ردیابی چندهدفه پرداخته‌است. در معماری این ردیاب از ویژگی‌های ظاهری به‌صورت مستقیم استفاده نشده‌است؛ بلکه در کنار آن از الگوریتم شناسایی برای تعیین محل حضور اهداف در هر قاب بهره‌گیری می‌شود. شبکهٔ بازگشتی به‌عنوان مدل حرکتی است و با دریافت مختصهٔ مکانی به ردیابی می‌پردازد. برای تخصیص نتایج شناسایی‌شده به اهداف در حال ردیابی و همچنین تعیین خروج از صحنه و یا ورود اهداف جدید از LSTM استفاده شده‌است. ورودی‌های درنظر گرفته‌شده برای شبکهٔ بازگشتی تنها مرکز و ابعاد کادر محصورکنندهٔ مربوط به هر هدف است و دیگر ویژگی‌ها نظیر سرعت یا شتاب تغییرات به کارهای آینده واگذار شده‌است. در این مقاله به تولید داده‌های آموزشی ساختگی با کمک داده‌های آماری مربوط به مجموعه‌های آموزشی موجود پرداخته شده‌است. روش ارائه‌شده در این مقاله به‌دلیل استفاده نکردن از ویژگی‌های ظاهری، هرچند به دقت الگوریتم‌های مطرح در حوزهٔ ردیابی چندهدفه نرسیده‌است، اما ایدهٔ استفاده از شبکهٔ بازگشتی به‌عنوان مدل حرکتی غیرخطی و از سوی دیگر سرعت بالا در اجرای الگوریتم از ویژگی‌های این روش هستند. این مقاله نشان داده است که استفاده از شبکه‌های بازگشتی در پیش‌بینی حرکت اهداف و ردیابی با استفاده از آن ممکن بوده و دقت قابل توجهی را به دست خواهد داد.

در [۷۸] به‌منظور مقابله با چالش انسداد و کاهش اثر آن در ردیابی چندهدفه از شبکهٔ عصبی بازگشتی استفاده شده‌است. هدف این مقاله استفاده از این شبکه به‌صورت موقتی است و تنها در مواردی مورد استفاده قرار می‌گیرد که هدف ردیابی دچار انسداد شود. در این مدل با استفاده از اطلاعات حرکت هدف در طول ردیابی شامل سرعت تغییرات مرکز و ابعاد کادر به‌عنوان ورودی شبکه به پیش‌بینی مکان و ابعاد کادر در قاب آینده پرداخته می‌شود؛ به این صورت که با محاسبهٔ اختلاف مرکز و ابعاد

[DOI: 10.61186/jsdp.22.1.83]

کادر در تمامی قاب‌های متوالی، به‌عنوان سرعت حرکت هدف و اعمال آن به شبکهٔ مدل حرکتی آماده به کار باقی می‌ماند. در حین ردیابی به محض ازدست‌دادن هدف به‌دلیل بروز انسداد، ردیابی با استفاده از شناسایی متوقف شده و ادامهٔ کار به مدل حرکتی مبتنی بر شبکهٔ بازگشتی سپرده و پس از خروج هدف از انسداد ادامهٔ ردیابی به حالت قبل باز می‌شود. در این مقاله اشاره شده‌است که استفاده از تغییرات پارامترهای کادر در میان قاب‌های متوالی بهتر از استفاده از خود پارامترها است. این امر هم در زمان آموزش و هم پیش‌بینی‌های بهتر به وسیلهٔ شبکه اثر خود را نشان می‌دهد. مدل حرکتی ارائه‌شده در این مقاله موفق به بهبود عملکرد الگوریتم‌های چندهدفه در حل چالش انسداد شده‌است.

[۸۴] به بررسی دقیق مدل‌های حرکتی در ردیابی چندهدفه پرداخته است. در این مقاله دو مدل حرکتی ضمنی و صریح ارائه شده و نتایج افزودن آن به ردیاب‌های مختلف بررسی شده‌است. مدل ضمنی از یک شبکهٔ کامل متصل استفاده می‌کند تا مدل حرکتی و موقعیت هدف را در قاب پیش‌رو تخمین بزند. این مدل مبتنی بر شبکه‌ای عمیق و به‌طور کامل آموزش‌پذیر است و باید تا پیش از استفاده آموزش ببیند و سپس بدون نیاز به تنظیم مجدد مورد استفاده قرار گیرد. ورودی این مدل ویژگی‌های ظاهری قاب کنونی و پیشین بوده و خروجی آن نقشهٔ امتیازی و موقعیت مکانی هدف است. بخشی از نتایج در (جدول-۳) قابل مشاهده است.

(جدول-۳): نتایج [۸۴] میزان افزایش دقت و موفقیت الگوریتم‌های ردیابی چندهدفه در صورت افزودن مدل حرکتی به آن‌ها

(Table-3): The results of [84], the increase in precision and success of multi-target tracking algorithms by incorporating motion model to their architectures

مجموعه ارزیابی	معیار ارزیابی	[۹۸]	[۹۹]	[۱۰۰]
MOTChallenge 2016 [۱۰۱]	دقت	۰.۴	۰.۱	۰.۱
	موفقیت	۰.۴	۱.۱	۰.۷

مدل صریح مبتنی بر عمل‌گر هم‌بستگی است و با استفاده از آن نقشهٔ امتیازی و میزان جابه‌جایی هدف میان دو قاب متوالی به‌دست می‌آید. این مدل به جای استفاده از ویژگی‌های عمیق استخراج‌شده از تصویر مستقیم بر روی کانال‌های هر قاب اعمال می‌شود و یک تابع برای مقایسه میان قاب‌های متوالی تولید می‌کند؛ درنتیجه نسبت به مدل ضمنی نظارت دقیق‌تری به‌دست می‌آید که پیش‌بینی‌های اشتباه را کاهش می‌دهد. نتایج حاصله نیز تأییدکنندهٔ این امر است و با استفاده از مدل دوم نتایج دقیق‌تری (تا ۲ درصد) حاصل شده که گفتنی است هرچند این مدل به دقت بهتری دست یافته‌است، اما به‌دلیل محاسبات پیچیده‌تر سرعت اجرای کمتری نسبت به مدل نخست دارد؛ به طوری که سرعت اجرا تا ۱۲ درصد کاهش می‌یابد.

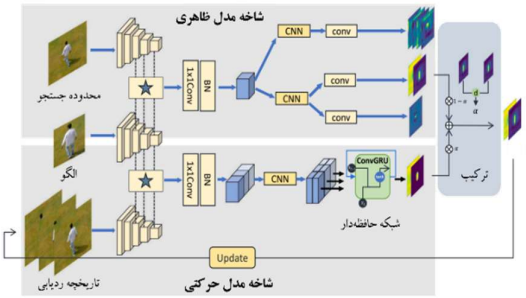
[۸۸] یک مدل حرکتی غیرخطی مبتنی بر شبکهٔ عمیق بازگشتی ارائه کرده‌است؛ این مدل کامل مستقل، از ساختار الگوریتم ردیابی است و قابلیت اضافه‌شدن به هر نوع ردیاب را داراست. ورودی این مدل نتایج الگوریتم ردیابی بوده و برای غنی‌ترشدن ویژگی‌های حرکتی علاوه‌بر موقعیت مکانی هدف، سرعت و شتاب جابه‌جایی آن نیز به‌عنوان ورودی مدل درنظر گرفته شده‌است. نتایج حاصل از افزودن این مدل به الگوریتم‌های ردیابی تک‌هدفه نشان از تأثیر قابل توجه ویژگی‌های حرکتی در عملکرد ردیاب دارد. بخشی از نتایج حاصله در (جدول-۴) قابل مشاهده است.

(جدول-۴): نتایج [۸۸] میزان افزایش دقت و موفقیت الگوریتم‌های ردیابی تک‌هدفه در صورت افزودن مدل حرکتی به آن‌ها

(Table-4): The results of [88], the increase in precision and success of single-target tracking algorithms by incorporating motion model to their architectures

مجموعه ارزیابی	معیار ارزیابی	[۴۴]	[۱۷]	[۱۰۲]
[۲۰] OTB50	دقت	۹	۳.۹	۵.۴
	موفقیت	۶.۲	۱.۹	۴.۱
[۳۶] OTB100	دقت	۸	۳.۶	۱.۹
	موفقیت	۱.۷	۱.۶	۶.۸

ردیاب‌های مبتنی بر شبکه‌های سیامیس نتایج چشم‌گیری را در زمینهٔ ردیابی کسب کرده‌اند، اما ساختارهای مربوطه تنها به ویژگی‌های ظاهری بسنده کرده و از ویژگی‌های زمانی بهره‌گیری نمی‌کنند. برای غلبه بر این ضعف، [۸۹] مدل حرکتی هدف را به این نوع ردیاب مطابق (شکل-۸) اضافه کرده‌است.



(شکل-۸): ساختار الگوریتم ردیابی [۸۹]

(Figure-8): The architecture of [89]

در این ساختار ویژگی‌های زمانی و تاریخچهٔ حرکتی هدف مورد استفاده قرار می‌گیرند. ورودی مدل حرکتی محدودهٔ جست‌وجو در سه قاب پیشین است و با اعمال هم‌بستگی عمقی[1] میان آن‌ها و الگوی ردیابی، ویژگی‌های

[1] Depth-Wise Cross Correlation

[DOI: 10.61186/jsdp.22.1.83]

مربوطه استخراج می‌شوند. در ادامه به‌منظور کاهش حجم پردازش ابعاد این ویژگی‌ها کاهش می‌یابد و وارد شبکهٔ حافظه‌دار شده و موقعیت هدف پیش‌بینی می‌شود. شبکهٔ حافظه‌دار با درنظرگرفتن پیش‌بینی‌های پیشین، خروجی مدل حرکتی را غنی‌تر می‌کند. موازی با مدل حرکتی از مدل ظاهری نیز استفاده شده و موقعیت هدف نیز با کمک آن مشخص می‌شود. برای ترکیب نتایج ظاهری و حرکتی فرض شده‌است که هدف ردیابی جابه‌جایی شدید و ناگهانی نخواهد داشت؛ ازاین‌رو هر یک از دو بخش ظاهری و حرکتی که نتایج دور از انتظاری داشته باشند امتیاز کمتری دریافت خواهند کرد. نتایج نشان‌دهندهٔ بهبود چشم‌گیر در ردیابی هستند. بخشی از ارزیابی‌های صورت‌گرفته در (جدول-۵) قرار گرفته است. در این جدول الگوریتم‌های ردیابی مختلف با ساختار مبتنی بر شبکهٔ سیامیس مقایسه شده‌اند که در میان آن‌ها [۸۹] شامل مدل حرکتی است.

(جدول-۵): مقایسهٔ [۸۹] شامل مدل حرکتی با ردیاب‌های مختلف بدون مدل حرکتی بر روی مجموعهٔ OTB100 [۳۶]

(Table-5): Comparison [89] including motion model with trackers without motion model on OTB100 [36]

ردیاب	دقت	موفقیت
[۹۰]	۸۶.۰	۶۴.۱
[۹۱]	۸۹.۹	۶۷.۱
[۹۲]	۸۹.۳	۶۹.۶
[۹۳]	۹۱.۵	۶۹.۶
[۹۴]	۹۱.۰	۶۹.۶
[۹۵]	۸۹.۸	۶۹.۷
[۹۶]	۹۱.۶	۷۰.۱
[۹۷]	۹۱.۵	۷۰.۹
[۸۹]	۹۲.۵	۷۰.۷

۷-جمع‌بندی و نتیجه‌گیری

در این مقاله روش‌های ردیابی مبتنی بر بینایی مورد بررسی قرار گرفته‌است. گستردگی معماری، روش و کاربرد ردیاب‌ها تقسیم‌بندی آن‌ها را دشوار می‌کند، اما در کل می‌توان ردیاب‌ها را در دو دستهٔ تک‌هدفه و چندهدفه درنظر گرفت. الگوریتم‌های ردیابی از دو ویژگی ظاهری و حرکتی استفاده می‌کنند. بررسی‌های انجام‌شده در این مقاله نشان می‌دهد که الگوریتم‌های چندهدفه سهم بیشتری را برای مدل حرکتی درنظر گرفته‌اند و به‌دلیل تعدد اهداف موجود در صحنه تمرکز کمتری بر روی ویژگی‌های ظاهری دارند، اما در مقابل الگوریتم‌های تک‌هدفه بیشتر از مدل ظاهری استفاده می‌کنند؛ این در حالی است که طبق ارزیابی‌ها مدل‌های حرکتی موجب بهبود عملکرد ردیابی شده و در قیاس با ویژگی‌های ظاهری حجم پردازش کمتری را به خود اختصاص می‌دهند.

بررسی‌ها نشان می‌دهد که با وجود گستردگی روش‌های ارائه‌شده و پیشرفت چشمگیر در بینایی ماشین، در استفاده از الگوریتم‌های ردیابی با قابلیت اجرای بلادرنگ، عملکرد قابل اطمینان و بدون نقصی را نمی‌توان انتظار داشت. این امر در صورت بروز یکی از چالش‌های مطرح در این حوزه تشدید می‌شود؛ چالش‌هایی نظیر حرکات ناگهانی و سریع توسط هدف، انسداد توسط موانع یا دیگر اهداف موجود در صحنه، تغییرات شدید در ظاهر و ابعاد هدف و همچنین ورود و خروج از صحنه که موجب واماندگی الگوریتم‌های ردیابی می‌شوند؛ اگرچه روش‌های با آموزش برخط در خلال ردیابی از دقت مطلوبی برخوردارند، اما به‌دلیل حجم پردازش بالا استفادهٔ بلادرنگ از آن‌ها در پیاده‌سازی‌های عملی ممکن نیست؛ به عبارت دیگر می‌توان گفت که الگوریتم‌های ردیابی با قابلیت اجرای بلادرنگ تنها در صورت برقراری شرایط ایدئال و نبود چالش‌های جدی، عملکردی قابل اطمینان خواهند داشت.

یکی از مشکلات مطرح‌شده در مقالات ایجاد تعاملی میان دقت عملکرد الگوریتم و سرعت اجرای آن است. الگوریتم‌های شناسایی، که به‌منظور تشخیص اهداف موجود در یک تصویر مستقل شناخته می‌شوند، دقت قابل قبولی را از خود نشان داده‌اند، اما امکان استفاده از آن‌ها در هر قاب برای یک ردیابی بلادرنگ ممکن نیست؛ چرا که از یک‌سو به‌دلیل حجم پردازش بالای این الگوریتم‌ها، سرعت اجرای ردیاب را تحت‌الش قرار داده و از سوی دیگر تنها قادر به شناسایی کلاس‌های خاصی هستند؛ این درحالی است که مقصود از یک ردیابی عمومی، دنبال‌کردن یک شئ در دنباله‌ای از تصاویر صرف‌نظر از نوع آن است؛ همچنین باید این نکته را درنظر داشت که هدف ردیابی درنظرگرفتن وابستگی‌های زمانی و مکانی در میان قاب‌های متوالی است؛ ازاین‌رو به‌دلیل احتمال بروز خطا در عملکرد الگوریتم‌های شناسایی، عدم تشخیص هدف در پاره‌ای از قاب‌ها و یا از سوی دیگر وجود چندین هدف هم‌نوع و با ظاهر همسان، استفاده از الگوریتم‌های شناسایی برای ردیابی ساده نخواهد بود؛ ازاین‌رو در بیشتر موارد از الگوریتم شناسایی در ردیاب‌هایی استفاده می‌شود که تنها دنبال‌کردن هدف خاصی موردنظر باشد. نمونه‌ای از این ردیاب‌ها الگوریتم‌های چندهدفه به‌منظور ردیابی عابران پیاده است که همگی از نوع انسان هستند.

[DOI: 10.61186/jsdp.22.1.83]

با توسعهٔ شبکه‌های عصبی بازگشتی و توانایی شگرف آن‌ها در پردازش داده‌های متوالی همچون متن، صوت و فیلم استفاده از آن‌ها در الگوریتم‌های ردیابی رو به افزایش است. بهره‌گیری از این شبکه‌ها به‌دلیل داشتن حافظهٔ کوتاه‌مدت و بلندمدت در حفظ ویژگی‌های مهم در حین ردیابی کمک شایانی به بهبود عملکرد الگوریتم‌های ردیابی کرده‌است. روش‌های مختلفی از تلفیق شبکه‌های عصبی پیچشی و بازگشتی ارائه شده‌اند که ایراد اصلی در بیشتر آن‌ها سرعت پردازش پایین است. بررسی‌ها نشان داده‌است که اعمال مستقیم ورودی‌ها با ابعاد زیاد نظیر ویژگی‌های استخراج‌شده از تصویر موجب کاهش شدید سرعت پردازش شبکه‌های بازگشتی می‌شود؛ ازاین‌رو در برخی مقالات با رویکرد توسعهٔ ردیاب بلادرنگ، با کاهش ابعاد ورودی شبکه‌های بازگشتی، در ازای کاهش دقت ردیابی، موفق به طراحی الگوریتم‌های بلادرنگ و با سرعت اجرای بسیار بالا شده‌اند.

به‌دلیل طاقت‌فرسابودن برچسب‌گذاری داده‌های متوالی همچون فیلم، مجموعه‌های آموزشی محدودی در زمینهٔ ردیابی در دسترس است و همین امر به یکی از محدودیت‌های موجود در این زمینه تبدیل شده‌است. به منظور غلبه بر این محدودیت راه‌کارهای مختلفی نظیر تولید داده‌های آموزشی ساختگی و یا تقویت و افزایش مجموعه‌های آموزشی موجود ارائه شده‌است. در روشی دیگر گروهی از الگوریتم‌ها به آموزش برخط بر روی قاب‌های فیلم در حین ردیابی پرداخته‌اند؛ در این میان بعضی نیز به آموزش ابتدایی الگوریتم به صورت برون‌خط پرداخته و در حین ردیابی با آموزش برخط و یا تنظیم مجدد شبکه بر روی قاب‌ها موجب افزایش دقت ردیابی می‌شوند. با استفاده از این روش‌ها نتایج قابل قبولی در ردیابی حاصل شده‌است، اما از آنجا که آموزش شبکه‌های عصبی حجم محاسباتی بالایی را می‌طلبد ردیابی بلادرنگ با استفاده از این الگوریتم‌ها ممکن نیست.

گروهی دیگر از الگوریتم‌ها به افزایش داده‌های آموزشی موجود پرداخته و آموزش الگوریتم را تنها به صورت برون‌خط انجام می‌دهند. برای این منظور علاوه‌بر فیلم، از داده‌های آموزشی به صورت عکس نیز استفاده و در آن با ایجاد حرکتی ساختگی از طریق جابه‌جاکردن پیکسل‌های عکس، دو قاب متوالی از یک فیلم را برای شبکه تداعی کرده و از این ترفند در آموزش شبکه استفاده می‌کنند. این شبکه‌ها هرچند دقت ردیابی روش‌های با آموزش برخط را ندارند، اما سرعت اجرای بالا در عین داشتن دقتی قابل قبول استفاده از آن‌ها را در پیاده‌سازی‌های عملی ممکن می‌سازد. ضعف اصلی مشاهده‌شده در استفادهٔ عملی از این الگوریتم‌ها، تمایل آن‌ها به جابه‌جایی میان اهداف موجود در مجموعهٔ آموزشی است؛ برای نمونه اگر در کنار هدف مورد نظر شی شناخته‌شده‌تری، به‌دلیل حضور بیشتر در دادهٔ آموزشی قرار داشته باشد، الگوریتم ردیابی بر روی آن جابه‌جا می‌شود و به ردیابی هدف شناخته‌شده‌تر می‌پردازد. به‌منظور مقابله با این مشکل از ترکیب شبکه‌های پیچشی و بازگشتی استفاده شده و نتایج چشمگیری در بهبود این ضعف حاصل شده‌است.

یکی از بخش‌های مغفول‌مانده در توسعهٔ الگوریتم‌های تک‌هدفه استفاده از مدل حرکتی است؛ این در حالی است که مطالعات صورت‌گرفته نشان از موفقیت این مدل‌ها در بهبود عملکرد ردیابی به‌ویژه ردیاب‌های چندهدفه دارد. پیش از مطرح‌شدن شبکه‌های پیچشی و موفقیت چشم‌گیر آن‌ها در استخراج ویژگی‌های عمیق از تصویر، مدل‌های حرکتی بیشتر مورد استفاده قرار می‌گرفتند، اما در روش‌های جدید به ویژه در ردیاب‌های تک‌هدفه از ویژگی‌های ظاهری استفادهٔ بیشتری می‌شود. در الگوریتم‌های تک‌هدفه وجود تنها یک هدف در تصویر و حجم محاسباتی کمتر در مقایسه با الگوریتم‌های چندهدفه موجب استفادهٔ آزادانه‌تر از ویژگی‌های ظاهری است، اما این امر در ردیابی چندهدفه به‌دلیل تعدد اهداف ممکن نیست و مدل‌های حرکتی در این الگوریتم‌ها استفادهٔ بیشتری دارند؛ ازاین‌رو در این مقاله به بررسی دقیق‌تر مدل‌های حرکتی و تأثیر آن‌ها در عملکرد ردیابی پرداخته شده‌است. نتایج نشان می‌دهند که مدل‌های حرکتی در عین سادگی و حجم پردازش کم، تأثیر عمیقی در بهبود عملکرد ردیابی دارند و استفاده از آن‌ها در این حوزه رو به افزایش است.

۸-مراجع 8-References

[1] M. Biglari, A. Soleimani, and H. Hassanpour, "Using Discriminative Parts for Vehicle Make and Model Recognition," Signal Data Process., vol. 15, no. 1, 2018, doi: 10.29252/jsdp.15.1.41.

[2] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2012, pp. 702–715. doi: 10.1007/978-3-642-33765-9_50.

[3] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.

[4] P. Li, D. Wang, L. Wang, and H. Lu, "Deep visual tracking: Review and experimental comparison," Pattern Recognit, vol. 76, pp.

[DOI: 10.61186/jsdp.22.1.83]

323–338, 2018, doi: 10.1016/j.patcog.2017.11.007.
[5] A. Sadeghian, A. Alahi, and S. Savarese, "Tracking the Untrackable: Learning to Track Multiple Cues with Long-Term Dependencies," Proceedings of the IEEE International Conference on Computer Vision, vol. 2017-Octob, pp. 300–311, 2017, doi: 10.1109/ICCV.2017.41.
[6] W. Liu et al., "SSD: Single shot multibox detector," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2016, pp. 21–37. doi: 10.1007/978-3-319-46448-0_2.
[7] J. S. He, Kaiming, Xiangyu Zhang, Shaoqing Ren, "Deep residual learning for image recognition," In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778, 2016.
[8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, pp. 1–14, 2015.
[9] Krizhevsky Alex, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Adv Neural Inf Process Syst, pp. 145–151, 2012, doi: 10.1145/3383972.3383975.
[10] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), New York, NY, USA: ACM, Sep. 2016, pp. 779–788.
[11] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.
[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," Adv Neural Inf Process Syst, vol. 28, pp. 91–99, 2015.
[13] R. He, K., Gkioxari, G., Dollár, P., & Girshick, "Mask r-cnn," In Proceedings of the IEEE international conference on computer vision, pp. 2961–2969, 2017.
[14] D. Zhang, H. Maei, X. Wang, and Y.-F. Wang, "Deep reinforcement learning for visual object tracking in videos," arXiv preprint arXiv:1701.08936, 2017.
[15] R. Spilger et al., "A Recurrent Neural Network for Particle Tracking in Microscopy Images Using Future Information, Track Hypotheses, and Multiple Detections," IEEE Transactions on Image Processing, vol. 29, pp. 3681–3694, 2020, doi: 10.1109/TIP.2020.2964515.
[16] T. Yang and A. B. Chan, "Learning dynamic memory networks for object tracking," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 11213 LNCS, pp. 153–169, 2018, doi: 10.1007/978-3-030-01240-3_10.
[17] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. S. Torr, "Fully-convolutional siamese networks for object tracking," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9914 LNCS, pp. 850–865, 2016, doi: 10.1007/978-3-319-48881-3_56.
[18] G. Plastiras, C. Kyrkou, and T. Theocharides, "You Only Look Once: Unified, Real-Time Object Detection," ArXiv, 2019.
[19] K. Remya and C. V Vipin Krishnan, "Survey of Generative and Discriminative Appearance Models in Visual Object Tracking," International Journal of Advance Research, Ideas and Innovations in Technology, vol. 4, no. 1, pp. 343–346, 2018.
[20] Y. Wu, J. Lim, and M. H. Yang, "Online object tracking: A benchmark," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2411–2418, 2013, doi: 10.1109/CVPR.2013.312.
[21] Y. Wu, J. Lim, and M. H. Yang, "Object tracking benchmark," IEEE Trans Pattern Anal Mach Intell, vol. 37, no. 9, pp. 1834–1848, 2015, doi: 10.1109/TPAMI.2014.2388226.
[22] H. J. C. Kristan, Matej, Aleš Leonardis, Jiří Matas, Michael Felsberg, Roman Pflugfelder, Joni-Kristian Kämäräinen, "The Tenth Visual Object Tracking VOT2022 Challenge Results," In Computer Vision–ECCV 2022 Workshops, pp. 431–460, 2023.
[23] K. H. Huang, Lianghua, Xin Zhao, "Got-10k: A large high diversity benchmark for generic object tracking in the wild," IEEE Trans Pattern Anal Mach Intell, pp. 1562–1577, 2019.
[24] H. L. Fan, Heng, Liting Lin, Fan Yang, Peng Chu, Ge Deng, Sijia Yu, Hexin Bai, Yong Xu, Chunyuan Liao, "Lasot: A high-quality benchmark for large-scale single object tracking," In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 5374–5383, 2019.
[25] B. G. Muller, Matthias, Adel Bibi, Silvio Giancola, Salman Alsubaihi, "Trackingnet: A large-scale dataset and benchmark for object tracking in the wild," In Proceedings of the European conference on computer vision (ECCV), pp. 300–317, 2018.
[26] S. L. Kiani Galoogahi, Hamed, Ashton Fagg, Chen Huang, Deva Ramanan, "Need for speed: A benchmark for higher frame rate object tracking," In Proceedings of the IEEE International Conference on Computer Vision, pp. 1125–1134, 2017.
[27] B. G. Mueller, Matthias, Neil Smith, "A benchmark and simulator for uav tracking," In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14, Springer International Publishing, pp. 445–461, 2016.
[28] M. S. et al Fan, Heng, Longyin Wen, Dawei Du, Pengfei Zhu, Qinghua Hu, Haibin Ling, "Visdrone-sot2020: The vision meets drone

[DOI: 10.61186/jsdp.22.1.83]

single object tracking challenge results," In Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part IV 16, pp. 728-749. Springer International Publishing, 2020.
[29] J. Z. et al Chen, Guanlin, Wenguan Wang, Zhijian He, Lujia Wang, Yixuan Yuan, Dingwen Zhang, "VisDrone-MOT2021: The vision meets drone multiple object tracking challenge results," In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2839-2846, 2021.
[30] M. S. et al Du, Dawei, Longyin Wen, Pengfei Zhu, Heng Fan, Qinghua Hu, Haibin Ling, "Visdrone-cc2020: The vision meets drone crowd counting challenge results," In Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part IV 16, pp. 675-691. Springer International Publishing, 2020.
[31] Q. T. Du, Dawei, Yuankai Qi, Hongyang Yu, Yifan Yang, Kaiwen Duan, Guorong Li, Weigang Zhang, Qingming Huang, "The unmanned aerial vehicle benchmark: Object detection and tracking," In Proceedings of the European conference on computer vision (ECCV), pp. 370–386, 2018.
[32] L. L.-T. Patrick Dendorfer, Hamid Rezatofighi, Anton Milan, Javen Shi, Daniel Cremers, Ian Reid, Stefan Roth, Konrad Schindler, "MOT20: A benchmark for multi object tracking in crowded scenes," arXiv:2003.09003, 2020.
[33] D. R. Achal Dave, Tarasha Khurana, Pavel Tokmakov, Cordelia Schmid, "Tao: A large-scale benchmark for tracking any object," In European Conference on Computer Vision, 2020.
[34] N. S. Lin, Weiyao, Huabin Liu, Shizhan Liu, Yuxi Li, Rui Qian, Tao Wang, Ning Xu, Hongkai Xiong, Guo-Jun Qi, "Human in events: A large-scale benchmark for human-centric video analysis in complex events," arXiv preprint arXiv:2005.04490, 2020.
[35] M. Kristan et al., "The Eighth Visual Object Tracking VOT2020 Challenge ResultsKristan, M., Leonardis, A., Matas, J., Felsberg, M., Pflugfelder, R., Kämäräinen, J.-K., Danelljan, M., Zajc, L. Č., Lukežič, A., Drbohlav, O., He, L., Zhang, Y., Yan, S., Yang, J., Fernández, G.," pp. 547–601, 2020.
[36] Y. Wu, J. Lim, and M. H. Yang, "Object tracking benchmark," IEEE Trans Pattern Anal Mach Intell, vol. 37, no. 9, pp. 1834–1848, 2015, doi: 10.1109/TPAMI.2014.2388226.
[37] D. Gordon, A. Farhadi, and D. Fox, "Re3 : Real-Time Recurrent Regression Networks for Object Tracking," IEEE Robot Autom Lett, vol. 3, pp. 788–795, 2018.
[38] G. Ciaparrone, F. Luque Sánchez, S. Tabik, L. Troiano, R. Tagliaferri, and F. Herrera, "Deep learning in video multi-object tracking: A survey," Neurocomputing, vol. 381, pp. 61–88, 2020, doi: 10.1016/j.neucom.2019.11.023.
[39] J. Fan, W. Xu, Y. Wu, and Y. Gong, "Human tracking using convolutional neural networks," IEEE Trans Neural Netw, vol. 21, no. 10, pp. 1610–1623, 2010, doi: 10.1109/TNN.2010.2066286.
[40] T. T. Trinh, R. Yoshihashi, R. Kawakami, M. Iida, and T. Naemura, "Bird detection near wind turbines from high-resolution video using lstm networks," World Wind Energy Conference, 2016.
[41] H. Fan and H. Ling, "SANet: Structure-Aware Network for Visual Tracking," IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, vol. 2017-July, pp. 2217–2224, 2017, doi: 10.1109/CVPRW.2017.275.
[42] F. Bi et al., "Review on Video Object Tracking Based on Deep Learning," Journal of New Media, vol. 1, no. 2, pp. 63–74, 2019, doi: 10.32604/jnm.2019.06253.
[43] X. Yang, C. Ma, J.-B. Huang, and M.-H. Yang, "Hierarchical Convolutional Features for Visual Tracking," Proceedings of the IEEE international conference on computer vision, pp. 3074–3082, 2015, doi: 10.1109/ICCV.2015.352.
[44] D. Held, S. Thrun, and S. Savarese, "Learning to track at 100 FPS with deep regression networks," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9905 LNCS, pp. 749–765, 2016, doi: 10.1007/978-3-319-46448-0_45.
[45] G. E. H. Krizhevsky, Alex, Ilya Sutskever, "Imagenet classification with deep convolutional neural networks," Adv Neural Inf Process Syst, pp. 1–1432, 2012, doi: 10.1201/9781420010749.
[46] N. Mahmoudi, "Multi-target tracking using CNN-based features : CNNMTT," Multimedia Tools and Applications 78.6 (2019): 7077-7096., 2019.
[47] N. Wang, S. Li, A. Gupta, and D.-Y. Yeung, "Transferring Rich Feature Hierarchies for Robust Visual Tracking," arXiv preprint arXiv:1501.04587, 2015.
[48] L. Wang, W. Ouyang, X. Wang, and H. Lu, "STCT: Sequentially training convolutional networks for visual tracking," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2016-Decem, pp. 1373–1381, 2016, doi: 10.1109/CVPR.2016.153.
[49] Q. Chu, W. Ouyang, H. Li, X. Wang, B. Liu, and N. Yu, "Online Multi-object Tracking Using CNN-Based Single Object Tracker with Spatial-Temporal Attention Mechanism," Proceedings of the IEEE International Conference on Computer Vision, vol. 2017-Octob, pp. 4846–4855, 2017, doi: 10.1109/ICCV.2017.518.
[50] A. W. M. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, "Visual tracking: An experimental survey," IEEE Trans Pattern Anal Mach Intell, vol. 36, no. 7, pp. 1442–1468, 2014, doi: 10.1109/TPAMI.2013.230.
[51] O. Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge," Int J Comput Vis, vol. 115, no. 3, pp. 211–252, 2015, doi: 10.1007/s11263-015-0816-y.

[DOI: 10.61186/jsdp.22.1.83]

Multimedia, vol. 21, no. 4, pp. 930–942, 2019, doi: 10.1109/TMM.2018.2869277.
[66] J. Jin, J. Bates, C. Farabet, and E. Culurciello, “Tracking with Deep Neural Networks,” 2013 47th Annual Conference on Information Sciences and Systems (CISS). IEEE, no. 1, 2013.
[67] S. Hong, T. You, S. Kwak, and B. Han, “Online tracking by learning discriminative saliency map with convolutional neural network,” 32nd International Conference on Machine Learning, ICML 2015, vol. 1, pp. 597–606, 2015.
[68] G. Koch, “Siamese Neural Networks for One-shot Image Recognition,” 2011.
[69] Y. Wu, Y. Sui, and G. Wang, “Vision-Based Real-Time Aerial Object Localization and Tracking for UAV Sensing System,” IEEE Access, vol. 5, pp. 23969–23978, 2017, doi: 10.1109/ACCESS.2017.2764419.
[70] G. Zhu, F. Porikli, and H. Li, “Robust Visual Tracking with Deep Convolutional Neural Network based Object Proposals on PETS,” Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2016.
[71] S. P. Bharati, Y. Wu, Y. Sui, C. Padgett, and G. Wang, “Real-Time Obstacle Detection and Tracking for Sense-and-Avoid Mechanism in UAVs,” IEEE Transactions on Intelligent Vehicles, vol. 3, no. 2, pp. 185–197, 2018, doi: 10.1109/tiv.2018.2804166.
[72] K. Zhu et al., “Single object tracking in satellite videos: Deep siamese network incorporating an interframe difference centroid inertia motion model,” Remote Sens (Basel), vol. 13, no. 7, 2021, doi: 10.3390/rs13071298.
[73] X. Y. Yan, Bin, Xinyu Zhang, Dong Wang, Huchuan Lu, “Alpha-refine: Boosting tracking performance by precise bounding box estimation,” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5289–5298, 2021.
[74] B. L. Voigtlaender, Paul, Jonathon Luiten, Philip HS Torr, “Siam r-cnn: Visual tracking by re-detection,” In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 6578–6588, 2020.
[75] M. Paul, M. Danelljan, C. Mayer, and L. Van Gool, “Robust Visual Tracking by Segmentation,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 13682 LNCS, pp. 571–588, 2022, doi: 10.1007/978-3-031-20047-2_33.
[76] C. Dicle, O. I. Camps, and M. Sznaier, “The way they move: Tracking multiple targets with similar appearance,” Proceedings of the IEEE International Conference on Computer Vision, pp. 2304–2311, 2013, doi: 10.1109/ICCV.2013.286.
[77] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, “Accurate scale estimation for robust visual tracking,” BMVC 2014 -
[52] M. Kristan, “The sixth Visual Object Tracking VOT2018 challenge results,” Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018.
[53] I. Goodfellow, Y. Bengio, and A. Courville, “Deep Learning,” p. 800, 2017.
[54] C. Szegedy et al., “Going deeper with convolutions,” Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 07-12-June, pp. 1–9, 2015, doi: 10.1109/CVPR.2015.7298594.
[55] A. Milan, S. H. Rezatofighi, A. Dick, I. Reid, and K. Schindler, “Online multi-target tracking using recurrent neural networks,” in 31st AAAI Conference on Artificial Intelligence, AAAI 2017, 2017, pp. 4225–4232.
[56] S. Hochreiter and J. Urgen Schmidhuber, “Long Short term Memory,” Neural Comput, vol. 9, no. 8, p. 17351780, 1997.
[57] G. Ning et al., “Spatially supervised recurrent convolutional neural networks for visual object tracking (ROLO),” Proceedings - IEEE International Symposium on Circuits and Systems, no. 1, pp. 1–4, 2017, doi: 10.1109/ISCAS.2017.8050867.
[58] K. Fang, “Track-RNN: Joint Detection and Tracking Using Recurrent Neural Networks,” 29th Conference on Neural Information Processing Systems (NIPS 2016), no. Nips, 2016.
[59] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” IEEE Trans Pattern Anal Mach Intell, vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.
[60] T. Yang and A. B. Chan, “Visual Tracking via Dynamic Memory Networks,” IEEE Trans Pattern Anal Mach Intell, vol. 14, no. 8, pp. 1–1, 2019, doi: 10.1109/tpami.2019.2929034.
[61] T. Yang and A. B. Chan, “Recurrent Filter Learning for Visual Tracking,” Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017, vol. 2018-Janua, pp. 2010–2019, 2017, doi: 10.1109/ICCVW.2017.235.
[62] P. Ondrúška and I. Posner, “Deep tracking: Seeing beyond seeing using recurrent neural networks,” 30th AAAI Conference on Artificial Intelligence, AAAI 2016, pp. 3361–3367, 2016.
[63] Q. Gan, Q. Guo, Z. Zhang, and K. Cho, “First Step toward Model-Free, Anonymous Object Tracking with Recurrent Neural Networks,” arXiv preprint arXiv:1511.06425, pp. 1–13, 2015.
[64] S. E. Kahou, V. Michalski, R. Memisevic, C. Pal, and P. Vincent, “RATM: Recurrent Attentive Tracking Model,” in IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 1613–1622. doi: 10.1109/CVPRW.2017.206.
[65] Q. Wang, C. Yuan, J. Wang, and W. Zeng, “Learning attentional recurrent neural network for visual tracking,” IEEE Trans

[DOI: 10.61186/jsdp.22.1.83]

Proceedings of the British Machine Vision Conference 2014, doi: 10.5244/c.28.65.
[78] M. Babaee, Z. Li, and G. Rigoll, "Occlusion Handling in Tracking Multiple People Using RNN," Proceedings - International Conference on Image Processing, ICIP, pp. 2715–2719, 2018, doi: 10.1109/ICIP.2018.8451140.
[79] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," Proceedings - International Conference on Image Processing, ICIP, vol. 2016-Augus, pp. 3464–3468, 2016, doi: 10.1109/ICIP.2016.7533003.
[80] G. Khan, Z. Tariq, and M. U. G. Khan, "Multi-Person Tracking Based on Faster R-CNN and Deep Appearance Features," Visual Object Tracking in the Deep Neural Networks Era. IntechOpen, vol. i, no. tourism, p. 13, 2019, doi: http://dx.doi.org/10.5772/57353.
[81] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," Proceedings - International Conference on Image Processing, ICIP, vol. 2017-Septe, pp. 3645–3649, 2018, doi: 10.1109/ICIP.2017.8296962.
[82] H. Nam and B. Han, "Learning Multi-domain Convolutional Neural Networks for Visual Tracking," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2016-Decem, pp. 4293–4302, 2016, doi: 10.1109/CVPR.2016.465.
[83] I. Jung, J. Son, M. Baek, and B. Han, "Real-Time MDNet," Proceedings of the European Conference on Computer Vision (ECCV), 2018.
[84] J. T. Shuai, Bing, Andrew Berneshawi, Xinyu Li, Davide Modolo, "Siammot: Siamese multi-object tracking," In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 12372–12382, 2021.
[85] S. Pellegrini, A. Ess, K. Schindler, and L. Van Gool, "You'll never walk alone: Modeling social behavior for multi-target tracking," Proceedings of the IEEE International Conference on Computer Vision, pp. 261–268, 2009, doi: 10.1109/ICCV.2009.5459260.
[86] J. Qiu, L. Wang, Y. H. Hu, and Y. Wang, "Two motion models for improving video object tracking performance," Computer Vision and Image Understanding, vol. 195, no. March, p. 102951, 2020, doi: 10.1016/j.cviu.2020.102951.
[87] B. Yang and R. Nevatia, "Multi-target tracking by online learning a CRF model of appearance and motion patterns," Int J Comput Vis, vol. 107, no. 2, pp. 203–217, 2014, doi: 10.1007/s11263-013-0666-4.
[88] M. Shahbazi, M. H. Bayat, and B. Tarvirdizadeh, "A motion model based on recurrent neural networks for visual object tracking," Image Vis Comput, vol. 126, p. 104533, 2022, doi: 10.1016/j.imavis.2022.104533.
[89] Z. Kang, T. Xu, X. F. Zhu, and X. J. Wu, "Learning Motion-Perceive Siamese network for robust visual object tracking," Pattern Recognit Lett, vol. 173, pp. 23–29, Sep. 2023, doi: 10.1016/j.patrec.2023.07.011.
[90] H. Zhang, J. Zhang, G. Nie, J. Hu, and W. J. (Chris) Zhang, "Residual memory inference network for regression tracking with weighted gradient harmonized loss," Inf Sci (N Y), vol. 597, pp. 105–124, Jun. 2022, doi: 10.1016/j.ins.2022.03.047.
[91] Z. Zhang, H. Peng, J. Fu, B. Li, and W. Hu, "Ocean: Object-aware Anchor-free Tracking," Jun. 2020.
[92] J. Wang, C. Lai, W. Zhang, Y. Wang, and C. Meng, "Transformer tracking with multi-scale dual-attention," Complex & Intelligent Systems, vol. 9, no. 5, pp. 5793–5806, Oct. 2023, doi: 10.1007/s40747-023-01043-1.
[93] B. Li, W. Wu, Q. Wang, F. Zhang, J. Xing, and J. Yan, "SiamRPN++: Evolution of Siamese Visual Tracking with Very Deep Networks," Dec. 2018.
[94] Z. Chen, B. Zhong, G. Li, S. Zhang, and R. Ji, "Siamese Box Adaptive Network for Visual Tracking," Mar. 2020.
[95] M. Danelljan, L. Van Gool, and R. Timofte, "Probabilistic Regression for Visual Tracking," Mar. 2020.
[96] D. Guo, J. Wang, Y. Cui, Z. Wang, and S. Chen, "SiamCAR: Siamese Fully Convolutional Classification and Regression for Visual Tracking," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Jun. 2020, pp. 6268–6276. doi: 10.1109/CVPR42600.2020.00630.
[97] F. Xie, C. Wang, G. Wang, Y. Cao, W. Yang, and W. Zeng, "Correlation-Aware Deep Tracking," Mar. 2022.
[98] R. U. Geiger, Andreas, Martin Lauer, Christian Wojek, Christoph Stiller, "3d traffic scene understanding from movable platforms," IEEE transactions on pattern analysis and machine intelligence 36, no. 5, pp. 1012–1025, 2013.
[99] J. M. Rehg. Kim, Chanho, Fuxin Li, Arridhana Ciptadi, "Multiple hypothesis tracking revisited," In Proceedings of the IEEE international conference on computer vision, pp. 4696–4704, 2015.
[100] A. C. Sanchez-Matilla, Ricardo, Fabio Poiesi, "Online multi-target tracking with strong and weak detections," In Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part II 14, pp. 84-99. Springer International Publishing, 2016.
[101] K. Schindler. Milan, Anton, Laura Leal-Taixé, Ian Reid, Stefan Roth, "MOT16: A benchmark for multi-object tracking," arXiv preprint arXiv:1603.00831, 2016.
[102] G. Bhat, M. Danelljan, L. Van Gool, and R. Timofte, "Learning discriminative model prediction for tracking," Proceedings of the IEEE International Conference on Computer Vision, vol. 2019-Octob, pp. 6181–6190, 2019, doi: 10.1109/ICCV.2019.00628.

[DOI: 10.61186/jsdp.22.1.83]

محمدحسین بیات دانشجوی دکترای مکاترونیک در دانشگاه تهران است. زمینهٔ پژوهشی وی بینایی ماشین و کنترل بینایی‌پایه است.
نشانی رایانامهٔ ایشان عبارت‌است از:
mhbayat@ut.ac.ir

بهرام تارویردی‌زاده دانشیار گروه مکاترونیک در دانشکدگان علوم و فناوری‌های نوین در دانشگاه تهران است. زمینهٔ پژوهشی مورد علاقهٔ ایشان سامانه‌های بلادرنگ، ربات‌های توان‌بخشی و پردازش سیگنال است.
نشانی رایانامهٔ ایشان عبارت‌است از:
bahram@ut.ac.ir

محمد شهبازی استادیار گروه ساخت و تولید در دانشکدهٔ مکانیک دانشگاه علم و صنعت است. زمینهٔ پژوهشی مورد علاقهٔ ایشان بینایی ماشین، اتوماسیون، سامانه‌های هیبرید و ربات‌های پادار است.
نشانی رایانامهٔ ایشان عبارت‌است از:
shahbazi@iust.ac.ir

[DOI: 10.61186/jsdp.22.1.83]