Information Technology Innovations and Applied Communication

Vol. 3, No.4, Winter, 2026, Serial No. 8,pp.83-99

An overview of positioning methods in augmented reality with a focus on the ARCore platform

Sina Samadi Gharehveran[1*]

* PHD.Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran.

(**Received:** 29 September 2025**, Revised:** 21 November 2025, **Accepted:** 07 December 2025, **Published:** 1 January 2026)

Abstract

Augmented reality, as one of the emerging technologies in human-environment interaction, plays an important role in integrating virtual information with the real world, and to achieve this goal, accurate and reliable positioning is required. Vision-inertial (VIO) techniques that combine camera and motion sensor data have attracted widespread attention due to their high performance, especially in indoor environments without access to GPS. This paper reviews and systematically examines the positioning methods in the ARCore platform; ARCore is one of the most advanced and widely used development frameworks in the field of augmented reality. In the initial sections, basic concepts such as visual tracking, pose estimation, feature recognition, and environment mapping are introduced and reviewed. Then, the internal mechanisms of ARCore, especially the SLAM, depth sensing, and Kalman filter subsystems, are analyzed. Also, common VIO algorithms including MSCKF, OKVIS, VINS-Mono and ROVIO are introduced and compared with ARCore performance in terms of accuracy, computational complexity and adaptability to mobile data. The main innovation of the paper is to provide a comparative analysis between academic models and practical frameworks that helps to understand the advantages and limitations of each. The findings indicate that the future trend is towards lightweight, learning-based and robust models in unstable environmental conditions, which can significantly improve the quality and efficiency of augmented reality.

Keywords: Augmented Reality, Positioning, ARCore, Kalman Filter, Visual-Inertial Odometry

* E-mail: s.samadi@tabrizu.ac.ir

مجله علمی «نوآوری‌های فناوری اطلاعات و ارتباطات کاربردی»
سال سوم، شماره ۴، فصل زمستان، ۱۴۰۴؛ ص ۹۹-۸۳.

مروری - ترویجی

مروری بر روش‌های موقعیت‌یابی در واقعیت افزوده با تمرکز بر پلتفرم ARCore

سینا صمدی قره ورن[1*]

۱- دکتری، دانشکده مهندسی برق و رایانه، دانشگاه تبریز، تبریز، ایران.

(**دریافت:** ۰۷ مهر ۱٤۰٤؛ **بازنگری:** ۳۰ آبان ۱٤۰٤؛ **پذیرش:** ۱٦ آذر ۱٤۰٤؛ **انتشار:** ۱۱ دی ۱٤۰٤)

چکیده

واقعیت افزوده به‌عنوان یکی از فناوری‌های نوظهور در تعامل انسان با محیط، نقش مهمی در ادغام اطلاعات مجازی با دنیای واقعی ایفا می‌کند و برای تحقق این هدف نیازمند موقعیت‌یابی دقیق و قابل‌اعتماد است. فن‌های بینایی - اینرسی (VIO) که داده‌های دوربین و حسگرهای حرکتی را ترکیب می‌کنند، به‌ویژه در محیط‌های داخلی و بدون دسترسی به GPS، به دلیل کارایی بالا توجه گسترده‌ای جلب کرده‌اند. این مقاله به‌صورت مروری و نظام‌مند، روش‌های موقعیت‌یابی در پلتفرم ARCore را بررسی می‌کند؛ ARCore یکی از پیشرفته‌ترین و پرکاربردترین چارچوب‌های توسعه در حوزه واقعیت افزوده است. در بخش‌های ابتدایی، مفاهیم پایه‌ای مانند رهگیری بصری، برآورد وضعیت، تشخیص ویژگی‌ها و نگاشت محیط معرفی و مرور می‌شوند. سپس سازوکارهای داخلی ARCore، به‌ویژه زیرسامانه‌های SLAM، سنجش عمق و فیلتر کالمن، تحلیل شده‌اند. همچنین الگوریتم‌های متداول VIO شامل MSCKF، OKVIS، VINS-Mono و ROVIO معرفی شده و از منظر دقت، پیچیدگی محاسباتی و تطبیق‌پذیری با داده‌های موبایلی، با عملکرد ARCore مقایسه می‌شوند. نوآوری اصلی مقاله ارائه یک تحلیل مقایسه‌ای میان مدل‌های دانشگاهی و چارچوب‌های عملی است که به درک مزایا و محدودیت‌های هر یک کمک می‌کند. یافته‌ها نشان می‌دهند که گرایش آینده به سمت مدل‌های سبک‌وزن، یادگیری‌محور و مقاوم در شرایط ناپایدار محیطی است که می‌تواند کیفیت و کارایی واقعیت افزوده را به طور قابل‌توجهی ارتقا دهد.

کلیدواژه‌ها: واقعیت افزوده، موقعیت‌یابی، ARCore، فیلتر کالمن، ادیومتری بینایی - اینرسی

* رایانامه : E-mail: s.samadi@tabrizu.ac.ir

۱. مقدمه

واقعیت افزوده[1] فناوری نوینی است که با ترکیب دنیای واقعی و عناصر دیجیتال، تجربه‌ای تعاملی و همه‌جانبه برای کاربران فراهم می‌کند. در واقع واقعیت افزوده مرز میان محیط واقعی و مجازی را کم‌رنگ کرده و با افزوده‌سازی اطلاعات دیداری، شنیداری یا لمسی، به غنی‌سازی ادراک انسان از محیط می‌پردازد. با گسترش کاربردهای این فناوری در عرصه‌های گوناگون مانند آموزش، سرگرمی، سلامت، معماری و ناوبری، نیاز به ارائه محتوای مجازی دقیق، هم‌راستا و متناسب با محیط واقعی بیش‌ازپیش احساس می‌شود [۱].

در میان اجزای کلیدی یک تجربه موفق در واقعیت افزوده، موقعیت‌یابی دقیق دستگاه کاربر نقشی محوری ایفا می‌کند. بدون تخمین دقیق موقعیت و وضعیت[2] دستگاه، هیچ‌کدام از مؤلفه‌های دیگر از جمله پرسپکتیو صحیح، انسداد اشیای مجازی توسط موانع واقعی[3] و انطباق نوری مناسب قابل‌اجرا نخواهند بود [۲]. چالش اصلی آنجاست که در بسیاری از برنامه‌های واقعیت افزوده رایج، به‌ویژه در محیط‌های شلوغ یا دارای زیرساخت پیچیده، الگوریتم‌های موقعیت‌یابی دچار افت دقت می‌شوند که موجب اختلال در تجربه کاربری می‌گردد [۳].

باتوجه‌به اهمیت این موضوع، پژوهش‌های متعددی در سال‌های اخیر به توسعه الگوریتم‌های موقعیت‌یابی ترکیبی (بینایی - اینرسی) پرداخته‌اند. این الگوریتم‌ها با بهره‌گیری هم‌زمان از اطلاعات دوربین و حسگرهای اینرسی مانند شتاب‌سنج و ژیروسکوپ، تلاش می‌کنند تخمینی دقیق‌تر از وضعیت دستگاه ارائه دهند. الگوریتم‌هایی مانند MSCKF، OKVIS، ROVIO، VINS-Mono و سایر روش‌های مبتنی بر SLAM و ادیومتری بینایی - اینرسی از جمله دستاوردهای مهم در این حوزه محسوب می‌شوند. در این میان، پلتفرم ARCore توسعه‌یافته توسط گوگل، به دلیل دردسترس‌بودن و پشتیبانی گسترده در دستگاه‌های اندرویدی، به یکی از بسترهای محبوب برای پیاده‌سازی برنامه‌های AR تبدیل شده است. بااین‌حال، جزئیات فنی دقیق الگوریتم‌های مورداستفاده در ARCore منتشر نشده‌اند و بررسی تجربی عملکرد آن در مقایسه با سایر روش‌های علمی موجود، نیازمند تحلیل سامان‌مند و داده‌محور است. شکل (۱) نمایی از پلتفرم ARCore و نمایش قابلیت‌های واقعیت افزوده بر روی تلفن همراه که توسط شرکت گوگل توسعه پیدا کرده است را نشان می‌دهد.



شکل (۱). نمای کلی از عملکرد ARCore در تلفن همراه [۵]

در این مقاله، با مرور مبانی نظری، تحلیل مقایسه‌ای و بررسی نقاط ضعف و قوت هر روش، تلاش می‌شود تصویری جامع از وضعیت فعلی الگوریتم‌های موقعیت‌یابی در AR ارائه شود. هدف آن است که با شناسایی شکاف‌ها و محدودیت‌ها، زمینه‌ای برای انتخاب روش مناسب و بهینه‌سازی تجارب واقعیت افزوده در کاربردهای واقعی فراهم گردد. همچنین در این مقاله با مرور تاریخی و نظام‌مند الگوریتم‌های موقعیت‌یابی بینایی - اینرسی از نخستین روش‌های مبتنی بر SLAM تا نسل‌های جدید یادگیری‌محور، سیر تحول این حوزه نیز به‌صورت جامع تحلیل شده است.

باوجود پیشرفت‌های چشمگیر در الگوریتم‌های VIO، تاکنون مرور نظام‌مندی که به طور هم‌زمان معماری ARCore، عملکرد عددی الگوریتم‌های مطرح، و تحلیل مقایسه‌ای آن‌ها را پوشش دهد، در ادبیات‌پژوهشی ارائه نشده است. باتوجه‌به تنوع و پیچیدگی الگوریتم‌های موقعیت‌یابی بینایی - اینرسی، این مرور تحلیلی می‌تواند با ارائهٔ مقایسه‌ای ساختارمند میان روش‌های مطرح، به‌ویژه ARCore، به پژوهشگران تازه‌وارد در این حوزه کمک کند تا مسیر تحقیقاتی خود را هدفمندتر انتخاب کرده و توسعه‌دهندگان نیز بتوانند انتخاب دقیق‌تری در پیاده‌سازی سیستم‌های واقعیت افزوده بر بسترهای سخت‌افزاری محدود داشته باشند. یافته‌های این مرور می‌تواند به‌عنوان راهنمایی کاربردی برای توسعه‌دهندگان سامانه‌های واقعیت افزوده بر بسترهای موبایلی تلقی شود، به‌ویژه در انتخاب الگوریتم‌هایی با توازن مناسب میان دقت، هزینه پردازش و پایداری محیطی است.

نوآوری اصلی این مقاله در مرور جامع و تحلیلی روش‌های ناوبری بصری و ناوبری بصری - اینرسی و استخراج نقاط قوت و محدودیت‌های آن‌ها است. بر اساس این تحلیل، ما یک چارچوب بهبودیافته ترکیبی از روش‌های مقاوم، یادگیری مداوم و کاربردهای واقعیت افزوده و BIM ارائه کرده‌ایم که عملکرد پایدارتر و تعمیم‌پذیری بالاتری در محیط‌های پویا فراهم می‌کند. علاوه بر این، با بررسی راهکارهای موجود، پیشنهاداتی برای بهینه‌سازی وزن خطای فرایند و افزایش دقت ناوبری ارائه شده است که می‌تواند مسیر تحقیقات آینده را هدایت کند. این ترکیب

[1] Augmented Reality – AR
[2] Pose
[3] Occlusion

تحلیلی و کاربردی، چارچوبی نوآورانه برای طراحی سامانه‌های ناوبری دقیق و قابل‌اعتماد ارائه می‌کند.

هدف اصلی این پژوهش ارائه یک مرور جامع و نظام‌مند از روش‌های موقعیت‌یابی در واقعیت افزوده با تمرکز ویژه بر پلتفرم ARCore است. به طور مشخص، این مقاله در پی دستیابی به سه هدف کلیدی است: (۱) بررسی مفاهیم بنیادی و سازوکارهای ARCore شامل ادغام بینایی - اینرسی و زیرسامانه‌های مرتبط؛ (۲) مقایسه عملکرد ARCore با الگوریتم‌های مطرح دانشگاهی مانند MSCKF، OKVIS، VINS-Mono و ROVIO از منظر دقت، هزینه محاسباتی و سازگاری با داده‌های موبایلی؛ و (۳) شناسایی شکاف‌های پژوهشی و چالش‌های عملی موجود در مسیر دستیابی به راهکارهای سبک‌وزن و مقاوم برای موقعیت‌یابی در محیط‌های متغیر. تحقق این اهداف می‌تواند به پرکردن فاصله میان دستاوردهای دانشگاهی و چارچوب‌های کاربردی کمک کرده و راهنمای ارزشمندی برای پژوهشگران و توسعه‌دهندگان در حوزه واقعیت افزوده فراهم سازد.

نوآوری اصلی این پژوهش در ارائه یک مرور نظام‌مند، تاریخی و تحلیلی از الگوریتم‌های موقعیت‌یابی بینایی - اینرسی نهفته است؛ مرور سیر تکاملی روش‌ها از نخستین نسل‌های مبتنی بر فیلتر تا مدل‌های یادگیری‌محور جدید و برای نخستین‌بار، ارائه یک چارچوب مقایسه‌ای چندمعیاره میان الگوریتم‌های مطرح. در این چارچوب، الگوریتم‌ها بر اساس شاخص‌هایی نظیر دقت، پایداری نوری، هزینه محاسباتی، قابلیت اجرا در سخت‌افزارهای سبک و سازگاری در محیط‌های پویا ارزیابی شده‌اند. علاوه بر این، با ترکیب تحلیل تاریخی، کمی و کاربردی، مقاله شکاف میان رویکردهای کلاسیک و نسل جدید الگوریتم‌های یادگیری‌محور (مانند XR-VIO و HDVIO2.0) را روشن کرده و مسیر آینده پژوهش‌ها در این حوزه را ترسیم می‌کند.

ساختار این مقاله به شرح زیر است: در بخش دوم، مبانی نظری و پیشینه پژوهش مرور می‌شود. بخش سوم به معرفی پلتفرم ARCore و قابلیت‌های موقعیت‌یابی آن اختصاص دارد. در بخش چهارم، الگوریتم‌های بینایی - اینرسی مطرح و مقایسه آن‌ها با عملکرد ARCore ارائه می‌شود. بخش پنجم شامل بحث و نتیجه‌گیری است که ضمن جمع‌بندی یافته‌ها، مسیرهای آتی پژوهش نیز پیشنهاد می‌گردد.

۲. روش تحقیق

واقعیت افزوده یکی از شاخه‌های پیشرفته فناوری‌های ترکیبی است که هدف آن ادغام محتوای دیجیتال سه‌بعدی با دنیای واقعی است. در این مطالعه، فرض بر آن است که سامانه‌های موقعیت‌یابی در محیط‌های داخلی و بدون دسترسی به GPS عمل می‌کنند و داده‌های مورد استفاده از دوربین تک‌چشمی (Mono Camera) و حسگرهای اینرسی (IMU) به‌صورت هم‌زمان دریافت می‌شوند. تمامی مقایسه‌ها بر اساس نتایج گزارش‌شده در شرایط یکسان از نظر نوع داده، محیط آزمایش (داخلی یا بیرونی) و منبع داده (مجموعه‌داده‌های EuRoC، TUM-VI و Season۴) انجام شده است. همچنین فرض شده که نور محیط و بافت صحنه در محدوده‌ی متعارف برای برنامه‌های واقعیت افزوده قرار دارد تا ارزیابی منصفانه میان الگوریتم‌ها حاصل شود. برخلاف واقعیت مجازی که کاربر را به طور کامل از محیط واقعی جدا می‌کند، واقعیت افزوده عناصر مجازی را به‌گونه‌ای در میدان دید کاربر جای می‌دهد که با ادراک طبیعی از محیط فیزیکی تلفیق یابند و حس حضور در یک محیط واقعی - مجازی مشترک ایجاد شود [۴].

در سال ۱۹۹۵، مطالعات میلی گرام و همکارانش منجر به معرفی مفهومی به نام پیوستار واقعیت - مجازی[1] شد که جایگاه واقعیت افزوده را در طیف‌بین دنیای کاملاً واقعی و کاملاً مجازی تعریف می‌کند [۵]. در این پیوستار، یک سر آن محیط واقعی[2] قرار دارد که در آن همه عناصر قابل‌مشاهده واقعی هستند؛ و سر دیگر آن محیط مجازی[3] است که تمام محتوای آن ساختگی و دیجیتال است [۶]. واقعیت افزوده در میانه این پیوستار قرار دارد، جایی که اشیای مجازی به‌صورت زنده و هماهنگ با فضای واقعی قابل مشاهده‌اند. در مقابل آن، مفهوم واقعیت افزوده مجازی قرار دارد که در آن اشیای واقعی در محیط‌های مجازی تلفیق می‌شوند [۷].

بر اساس تعریف آزوما، برای ایجاد یک تجربه موفق واقعیت افزوده، چهارعنصر کلیدی باید محقق شوند:

- **موقعیت‌یابی دقیق دستگاه تلفن همراه:** برای هم‌راستاسازی صحیح عناصر مجازی با محیط واقعی، موقعیت و وضعیت دستگاه باید بادقت بالا تعیین شود. کوچک‌ترین خطا در این بخش می‌تواند کل تجربه را دچار اختلال کرده و موجب بی‌اعتمادی به محتوای افزوده گردد[۸].
- **نمایش از زاویه دید مناسب** :اشیای مجازی باید از همان زاویه‌ای که کاربر در آن لحظه مشاهده می‌کند رندر شوند تا انسجام فضایی حفظ شود [۹].
- **انسداد صحیح بین عناصر:** اگر یک شیء واقعی در جلوی شیء مجازی قرار دارد، سیستم باید آن را تشخیص دهد و قسمت‌های مربوط به شیء مجازی را که پشت آن پنهان می‌شود، حذف کند [۱۰].

[1] Reality-Virtuality Continuum
[2] Real Environment – RE
[3] Virtual Environment – VE

- **هماهنگی نوری و سایه‌ها**: نورپردازی و سایه‌های اشیای مجازی باید با محیط واقعی مطابقت داشته باشد تا توهم واقع‌گرایی حفظ شود [۱۱].

از میان این موارد، موقعیت‌یابی دقیق دستگاه به‌عنوان بنیادی‌ترین عامل در پیاده‌سازی موفق واقعیت افزوده تلقی می‌شود. چرا که در صورت ضعف در این بخش، حتی دقیق‌ترین رندر گرافیکی یا پیشرفته‌ترین مدل سه‌بعدی نیز نمی‌تواند تجربه‌ای واقع‌گرایانه فراهم کند. واقعیت افزوده در سال‌های اخیر، جایگاه مهمی در صنایع و خدمات مختلف یافته است. قابلیت ادغام دنیای واقعی با محتوای دیجیتال موجب شده تا واقعیت افزوده به‌عنوان یک ابزار کاربردی در آموزش، تجارت، معماری، سلامت، سرگرمی، گردشگری و بسیاری حوزه‌های دیگر موردتوجه قرار گیرد.

در این پژوهش، مقایسه الگوریتم‌های بینایی - اینرسی بر اساس نتایج گزارش‌شده در مقالات بنچ‌مارک معتبر و مجموعه‌داده‌های استاندارد انجام شده است. از جمله مهم‌ترین منابع داده‌ای می‌توان به EuRoC MAV Dataset، TUM-VI و Season Dataset[4] اشاره کرد که توسط اکثر پژوهش‌های روز دنیا برای ارزیابی دقت و پایداری الگوریتم‌های VIO استفاده می‌شوند.

ازآنجاکه هدف مقاله مرور نظام‌مند و تحلیلی بوده است، هیچ‌گونه شبیه‌سازی یا پیاده‌سازی مستقل انجام نشده و تمامی مقایسه‌ها از داده‌های واقعی و ارزیابی‌های منتشرشده در مقالات مرجع استخراج شده‌اند.

برای الگوریتم ARCore نیز، به دلیل بسته‌بودن کد منبع آن، از داده‌های مستندات رسمی گوگل و مقالات تحلیلی معتبر (نظیر RD-VIO و Kimera-VIO) برای تحلیل عملکرد استفاده شده است.

•تجارت الکترونیک و خرده‌فروشی

یکی از پرکاربردترین حوزه‌های واقعیت افزوده، خرده‌فروشی و تجربه خرید آنلاین است. برندهایی مانند IKEA با اپلیکیشن «IKEA Place»، به کاربران این امکان را می‌دهند که مدل‌های سه‌بعدی از مبلمان را به‌صورت مجازی در فضای واقعی منزل خود مشاهده کنند [۱۲]. این تجربه باعث تصمیم‌گیری بهتر قبل از خرید می‌شود و نرخ بازگشت کالا را کاهش می‌دهد. به طور مشابه، اپلیکیشن «Sephora Virtual Artist» به کاربران اجازه می‌دهد تا با دوربین گوشی خود، لوازم آرایشی را روی چهرهٔ خود امتحان کنند [۱۳].

•آموزش و یادگیری

در حوزهٔ آموزش، واقعیت افزوده موجب افزایش تعامل و یادگیری عمیق‌تر شده است. برنامه‌هایی مانند GoogleExpeditions به دانش‌آموزان امکان می‌دهند تا بدون ترک کلاس درس، به مکان‌هایی نظیر ایستگاه فضایی بین‌المللی یا شهرهای باستانی سفر کنند [۱۴]. همچنین، نرم‌افزار Anatomy 4D با نمایش مدل‌های سه‌بعدی از بدن انسان به‌صورت واقعیت افزوده، به دانشجویان پزشکی و زیست‌شناسی در درک بهتر ساختارهای پیچیده کمک می‌کند [۱۵].

•بازی و سرگرمی

صنعت بازی از پیش‌گامان استفاده از واقعیت افزوده بوده است. موفقیت جهانی Pokémon GO نمونهٔ بارزی از این کاربرد است که کاربران را وادار می‌کرد با استفاده از دوربین گوشی، پوکمون‌ها را در محیط واقعی پیدا کنند. بازی‌هایی مانند Harry Potter: Hogwarts Mystery نیز تجربه‌ای جادویی و تعاملی با محیط واقعی ایجاد کرده‌اند[۱۶].

•معماری و ساخت‌وساز

واقعیت افزوده در معماری ابزاری برای بصری‌سازی طرح‌ها پیش از اجراست. اپلیکیشن ARki یکی از نمونه‌های موفق در این زمینه است که به معماران امکان می‌دهد مدل‌های سه‌بعدی پروژه‌ها را در محل ساخت قرار دهند [۱۷]. این کار باعث می‌شود ذی‌نفعان مختلف، از کارفرما تا عموم مردم، پیش از اجرای پروژه نظر خود را اعلام کرده و تغییرات لازم اعمال شود. بااین‌حال، همان‌طور که در ادامه بحث خواهد شد، ARki نیز با چالش‌های موقعیت‌یابی دقیق، به‌ویژه در محیط‌های شلوغ و دارای زیرساخت‌های متراکم، مواجه است [۱۸].

•گردشگری و مسیریابی

در حوزهٔ گردشگری، اپلیکیشن‌هایی مانند SmartGuide اطلاعات زنده در مورد مکان‌های تاریخی یا دیدنی ارائه می‌دهند [۱۹]. کاربران می‌توانند با نشانه‌روی دوربین گوشی به یک بنای تاریخی، اطلاعات متنی و صوتی مربوط به آن را دریافت کنند. همچنین، Google Maps با ویژگی«Live View» ، امکان مسیریابی مبتنی بر واقعیت افزوده را فراهم کرده که مخصوصاً برای افراد پیاده در محیط‌های ناآشنا بسیار کاربردی است [۲۰].

۳. روش و رویکرد تحقیق

باوجود پیشرفت‌های قابل‌توجه در توسعه اپلیکیشن‌های واقعیت افزوده، هنوز چالش‌های اساسی در زمینهٔ موقعیت‌یابی دقیق و پایدار وجود دارد. یکی از ضعف‌های عمده در بسیاری از برنامه‌های فعلی، ناتوانی در تعیین دقیق موقعیت و وضعیت دستگاه کاربر[1] در محیط‌های واقعی پیچیده است. این مسئله به‌ویژه در کاربردهای معماری و زیرساخت‌های شهری که نیاز به

[1] Pose Estimation

قرارگیری دقیق یک مدل سه‌بعدی در مکان مشخص دارند، خود را بیشتر نشان می‌دهد [۲۱].

به‌عنوان‌مثال، اپلیکیشن ARki که برای نمایش مدل‌های معماری در محل واقعی طراحی شده، به دلیل مشکلات موقعیت‌یابی در محیط‌های شلوغ و پیچیده شهری، دچار چالش‌هایی در ارائه تصویر دقیق و منطبق با واقعیت می‌شود[۲۲].

سحر فیکوحی، بنیان‌گذار ARki، در سخنرانی خود در اجلاس AWE اشاره می‌کند: وقتی با مدل‌های معماری بزرگ و پیچیده سروکار دارید، و در فضایی کار می‌کنید که خودروها حرکت می‌کنند و زیرساخت‌های مختلف در جریان هستند، ارائهٔ مدل دقیق در زمان واقعی بسیار دشوار می‌شود... تقریباً غیرممکن بود که رهگیری دوربین[1] خوبی داشته باشیم [۲۳].

در اغلب برنامه‌های واقعیت افزوده، کاربران می‌توانند اشیای مجازی را به‌صورت دستی در محیط قرار دهند و جابه‌جا کنند. اما در کاربردهایی مانند ARki یا سامانه‌های ناوبری، نیاز است که شیء مجازی فقط و فقط در محل واقعی و ازپیش‌تعیین‌شده خودنمایش داده شود. این موضوع وابسته به فن‌های دقیق موقعیت‌یابی است که هم در محیط‌های داخلی و هم خارجی کارایی داشته باشند [۲۴]. علاوه بر این، در اغلب تحقیقات علمی ارائه‌شده در این حوزه، الگوریتم‌های مکان‌یابی پیشرفته ارزیابی می‌شوند، ولی مقایسه مستقیمی بین آن‌ها و الگوریتم داخلی پلتفرم‌های مرسوم مثل ARCore صورت نگرفته است .این در حالی است که ARCore در حال حاضر در بسیاری از دستگاه‌های اندرویدی به‌صورت پیش‌فرض برای پیاده‌سازی AR استفاده می‌شود و بررسی توانایی آن در موقعیت‌یابی می‌تواند شکاف میان صنعت و پژوهش را کاهش دهد [۲۵]. چالش‌های اصلی در موقعیت‌یابی برنامه‌های واقعیت افزوده عبارت‌اند از:

- **موقعیت‌یابی نادقیق در محیط‌های شلوغ:** افت دقت در حضور اشیای متحرک، خودروها یا زیرساخت متراکم.
- **عدم پایداری در ردیابی دوربین:** قطع و وصل‌شدن رهگیری در زمان حرکت سریع یا تغییر نور.
- **عدم انطباق با مقیاس واقعی:** اختلاف بین اندازه و موقعیت شیء مجازی و محیط واقعی.
- **ناسازگاری نور و سایه‌ها :**موجب کاهش واقع‌گرایی در صحنه
- **فاصله میان روش‌های علمی و پلتفرم‌های تجاری:** نبود مقایسه دقیق با ابزارهایی مانند ARCore

۴. پلتفرم ARCore و قابلیت‌های آن در موقعیت‌یابی

ARCore، پلتفرم واقعیت افزودهٔ شرکت گوگل، به‌عنوان یکی از پراستفاده‌ترین ابزارهای توسعه واقعیت افزوده برای دستگاه‌های اندرویدی شناخته می‌شود. این پلتفرم مجموعه‌ای از ابزارها و APIها را فراهم می‌کند که امکان قرارگیری عناصر مجازی در دنیای واقعی را برای توسعه‌دهندگان فراهم می‌سازد. ARCore بر پایه سه قابلیت اصلی بنا شده است: ردیابی حرکت[2]، درک محیط[3] و تخمین نور[4] [۲۶].

•ردیابی حرکت

قابلیت ردیابی حرکت در ARCore با ترکیب داده‌های دوربین و دستگاه‌های اندازه‌گیری اینرسی IMU، تلاش می‌کند تا وضعیت دقیق دستگاه را در محیط تعیین کند. دستگاه‌های اندازه‌گیری اینرسی معمولاً شامل سه ژیروسکوپ نرخ متعامد و سه شتاب‌سنج متعامد هستند که به ترتیب، سرعت زاویه‌ای و شتاب خطی را اندازه‌گیری می‌کنند [۲۷]. با پردازش سیگنال‌های این دستگاه‌ها می‌توان موقعیت و جهت یک دستگاه را ردیابی کرد. در این فرایند، ARCore با شناسایی و پیگیری ویژگی‌های بصری مشخص در تصاویر دریافتی از دوربین، حرکت نسبی دستگاه را محاسبه می‌کند. هم‌زمان، اطلاعات به‌دست‌آمده از شتاب‌سنج و ژیروسکوپ نیز در این تخمین لحاظ می‌شوند[۲۸].

هرچند این ادغام داده‌ها نوعی هم‌جفت‌سازی tightly-coupled محسوب می‌شود، اما الگوریتم دقیق و جزئیات عملکرد داخلی ARCoreاز سوی گوگل به‌صورت عمومی منتشر نشده است. همچنین، اطلاعات مستقلی دربارهٔ دقت نهایی تخمین وضعیت در این پلتفرم نیز موجود نیست [۲۹].

•درک محیط

درک محیط در ARCore بر شناسایی سطوح افقی و عمودی در محیط اطراف تمرکز دارد. این کار از طریق تحلیل ویژگی‌های بصری تشخیص‌داده‌شده در تصاویر صورت می‌گیرد. اگر این ویژگی‌ها در یک صفحهٔ هندسی هم‌راستا باشند، آن ناحیه به‌عنوان یک سطح صاف شناسایی می‌شود. این قابلیت برای قراردادن اشیای مجازی روی میز، دیوار یا زمین بسیار حیاتی است. بااین‌حال، ARCore در شناسایی سطوح صاف و بدون بافت (مانند میز سفید ساده یا دیوار صاف) دچار ضعف است؛ چرا که برای تشخیص، نیاز به نقاط تمایز بصری[5] دارد [۳۰].

•تخمین نور

قابلیت تخمین نور در ARCore وضعیت نور محیط را بررسی کرده و شدت نور، دمای رنگ و جهت تابش را تخمین می‌زند. این ویژگی کمک می‌کند تا اشیای مجازی با نوری شبیه به نور واقعی

[1] Camera Tracking

[2] Motion Tracking
[3] Environmental Understanding
[4] Light Estimation
[5] Textured features

صحنه رندر شوند و در نتیجه، به‌صورت طبیعی‌تر با محیط تلفیق شوند [۳۱]. همچنین، ARCore می‌تواند سایه‌هایی متناسب با شرایط نوری برای اشیای مجازی ایجاد کند که به افزایش واقع‌گرایی کمک می‌کند [۳۲].

درحالی‌که ARCore یکی از پلتفرم‌های پیشرو در توسعه اپلیکیشن‌های واقعیت افزوده است، ارزیابی‌های مستقل درباره‌ی دقت الگوریتم‌های آن به‌ویژه در موقعیت‌یابی نسبتاً محدود هستند. همچنین، محدودیت‌های زیر برای آن طبق جدول (۱) گزارش شده‌اند:

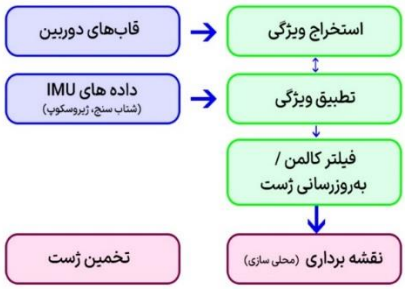
جدول (۱). مزیت و محدودیت ARCore در موقعیت‌یابی

جنبه	مزیت	محدودیت
ردیابی حرکت	استفاده از IMU و دوربین	جزئیات الگوریتم نامشخص؛ حساس به نور و حرکت سریع
درک محیط	شناسایی سطوح مختلف	ضعف در سطوح بدون بافت
نورپردازی	تطابق نوری مناسب با محیط	گاهی تأخیر در به‌روزرسانی نور واقعی محیط

یکی از مهم‌ترین چالش‌های فنی در توسعه سیستم‌های واقعیت افزوده، دستیابی به تخمین دقیق موقعیت و وضعیت دستگاه کاربر در زمان واقعی است. این تخمین باید در محیط‌های پیچیده، پویا و بدون زیرساخت اضافی انجام شود. ازآنجاکه اغلب دستگاه‌های همراه به دوربین و حسگرهای اینرسی مجهز هستند، بسیاری از الگوریتم‌های نوین موقعیت‌یابی از ترکیب داده‌های این دو منبع استفاده می‌کنند که به آن‌ها روش‌های بینایی - اینرسی[1] گفته می‌شود [۳۳].

در حوزهٔ موقعیت‌یابی بینایی - اینرسی، چند اصطلاح کلیدی به طور گسترده به کار می‌رود:

- **Fusion Techniques**: ادغام داده‌های دوربین و IMU می‌تواند به دو صورت، هر منبع به طور مستقل [2] تخمین زده می‌شود و سپس ادغام می‌گردد یا اندازه‌گیری‌ها در یک مدل مشترک ترکیب می‌شوند[3] انجام گیرد. روش دوم دقت بیشتری دارد.
- **Loop Closure**: قابلیت بازشناسی یک موقعیت قبلی و اصلاح مسیر کلی حرکت در صورت ورود دوباره به محیط‌های تکراری است. این ویژگی باعث کاهش خطای تجمعی می‌شود.
- **Reprojection Error**: اختلاف بین مکان پیش‌بینی‌شده یک نقطه در تصویر و مکان واقعی مشاهده‌شده آن است که به‌عنوان معیار دقت تخمین وضعیت استفاده می‌شود.
- **Bundle Adjustment**: فرایند بهینه‌سازی هم‌زمان وضعیت دوربین و مختصات سه‌بعدی نقاط برای کاهش خطای باز پراکنش است؛ اگرچه بسیار دقیق است؛ اما هزینه محاسباتی بالایی دارد.
- **Marginalization**: حذف تدریجی اطلاعات قدیمی (مانند فریم‌ها یا ویژگی‌ها) برای کاهش بار محاسباتی. برخی روش‌های جدید از *delayed marginalization* استفاده می‌کنند که در صورت اهمیت دوباره، داده‌ها را بازیابی می‌نماید.



شکل (۲). دیاگرام عملکرد داخلی ARCore برای موقعیت‌یابی بینایی - اینرسی

۵. معرفی الگوریتم‌های بینایی - اینرسی در واقعیت افزوده

۵-۱. Multi-State Constraint Kalman Filter (MSCKF)

یکی از نخستین الگوریتم‌های فیلتر محور در موقعیت‌یابی تک‌چشمی، MSCKF است که در سال ۲۰۰۷ معرفی شد [۳۴]. این روش با معرفی رویکرد multi-state constraint توانست بار محاسباتی را کاهش داده و امکان استفاده بلادرنگ را فراهم کند، هرچند فاقد مکانیزم loop-closure است و در مسیرهای طولانی خطای تجمعی بالایی دارد [۳۵]. این الگوریتم از فیلتر کالمن توسعه‌یافته برای ادغام داده‌های دوربین و IMU استفاده می‌کند و با روش Runge-Kutta مرتبه پنجم، حالت سیستم را در طول زمان پیش‌بینی می‌کند [۳۶]. نکته مهم در طراحی MSCKF، استفاده از وضعیت دوربین به‌جای مختصات سه‌بعدی نقاط است. در نتیجه، محاسبات کاهش‌یافته و سرعت بالایی حاصل می‌شود. این روش برای سیستم‌های فیلتر محور و بلادرنگ[4] مناسب است. این روش از نوع فیلتر محور[5] بوده و ساختاری tightly coupled دارد. بااین‌حال، فاقد قابلیت loop closure است که می‌تواند بر دقت کلی سیستم تأثیرگذار باشد. دقت این روش در سطح متوسط ارزیابی می‌شود، به‌طوری که در مسیرهای طولانی، خطای نسبی قابل‌توجهی مشاهده می‌شود؛ برای مثال، در یک مسیر آزمایشی به طول ۳.۲ کیلومتر در محیط بیرونی که داده‌ها از دوربین و IMU نصب‌شده روی خودرو جمع‌آوری شده‌اند، میزان خطای نهایی حدود ۱۰ متر گزارش شده است [۳۷]. روش

[1] Visual-Inertial Odometry
[2] Loosely coupled
[3] Tightly coupled
[4] Real-time
[5] Filter-based

MSCKF به‌عنوان نقطه آغاز بسیاری از فیلترهای بینایی - اینرسی مطرح است و با استراتژی نگهداری حالت‌های چندگانه توانست بار محاسباتی را کاهش دهد [۳۸].

۵-۲. OKVIS

الگوریتم OKVIS در سال ۲۰۱۵ معرفی شد و از روش بهینه‌سازی بر مبنای کلیدفریم[1] استفاده می‌کند [۳۹]. این چارچوب با بهره‌گیری از bundle adjustment محلی و انتخاب کلیدفریم‌های مناسب، دقت بالاتری نسبت به فیلترها نشان داده است، اما هزینه محاسباتی سنگین‌تری دارد [۴۰]. این روش، باز پراکنش[2] ویژگی‌های تصویری و داده‌های IMU را در یک تابع هزینه مشترک تلفیق می‌کند. کلیدفریم‌ها بر اساس تعداد تطابق ویژگی‌ها و پراکندگی آن‌ها در تصویر انتخاب می‌شوند. این الگوریتم به دلیل نداشتن loop closure، در مسیرهای طولانی دچار خطای تجمعی می‌شود. این روش از نوع بهینه‌سازی محور[3] بوده و همانند برخی روش‌های دیگر، دارای ساختاری tightly coupled است [۴۱]. بااین‌حال، همچنان فاقد قابلیت loop closure است. دقت آن نسبت به روش MSCKF بالاتر گزارش شده، اما در عوض، نیازمند منابع محاسباتی بیشتری است. بر اساس نتایج تجربی به‌دست‌آمده از آزمایش‌ها در محیط‌های داخلی و خارجی (با بهره‌گیری از GNSS و Vicon) سیستم OKVIS عملکرد دقیق‌تری نسبت به MSCKF از خود نشان داده است، ولی در مقابل، بار پردازشی بیشتری بر سیستم تحمیل می‌کند. OKVIS نشان داد که بهره‌گیری از بهینه‌سازی غیرخطی روی پنجره کلیدفریم‌ها می‌تواند دقت VIO را به طور محسوس افزایش دهد، به قیمت افزایش بار محاسباتی [۴۲].

۵-۳. ROVIO

الگوریتم ROVIO نیز در سال ۲۰۱۵ معرفی شد و از فیلتر کالمن توسعه‌یافته برای ادغام tightly coupled داده‌های تصویری و اینرسی استفاده می‌کند [۴۳]. نوآوری اصلی ROVIO استفاده از خطای شدت پیکسل‌ها به‌جای ویژگی‌های کلاسیک است که باعث مقاومت بیشتر در برابر نویز و تاری تصویری شده است[۴۴].

تفاوت اصلی ROVIO با OKVIS در روش عدم استفاده از ویژگی‌های تصویری کلاسیک (مثل SIFT یا ORB) است؛ بلکه از خطای شدت نور[4] در هر پیکسل به‌عنوان نوآوری در به‌روزرسانی فیلتر استفاده می‌کند. این روش از نوع فیلتر محور بوده و ساختاری tightly coupled دارد. همچنین، مانند برخی روش‌های مشابه، فاقد قابلیت loop closure است. از جمله ویژگی‌های بارز آن می‌توان به مقاومت در برابر نویز و تاری تصویر اشاره کرد که در شرایط نامساعد تصویری، مزیت محسوب می‌شود [۴۵]. در آزمایش‌هایی که با استفاده از سنجشگر VI-Sensor و سیستم حرکتی دقیق (MoCap) انجام شده‌اند، سیستم ROVIO عملکرد دقیق‌تری نسبت به OKVIS از خود نشان داده است؛ بااین‌حال، مقایسه عددی یا متریک دقیقی بین این دو روش ارائه نشده است. ROVIO معرفی‌کننده یک EKF مستقیم مبتنی بر خطای فوتو متریک بود که در محیط‌های با نویز تصویری عملکرد پایدارتری فراهم کرد [۴۶].

۵-۴. VIORB

الگوریتم Visual-Inertial ORB-SLAM (VIORB) در سال ۲۰۱۷ معرفی شد و توسعه‌ای بر نسخهٔ اولیه ORB-SLAM بود که با اضافه‌کردن داده‌های IMU ، برای محیط‌های بزرگ‌مقیاس بهینه شده است [۴۷]. ساختار این سیستم شامل سه بخش موازی است: رهگیری، نقشه‌برداری محلی[5] و بستن حلقه[6] [۴۸]. در بخش رهگیری، با دریافت هر فریم، وضعیت و سرعت دستگاه به همراه بایاس‌های IMU تخمین زده می‌شود. در بخش نقشه‌برداری، bundle adjustment محلی برای بهینه‌سازی نقشه انجام می‌شود. در بخش سوم، اگر تطابقی بین فریم جاری و فریم قبلی یافت شود، به‌صورت خودکار بستن حلقه و اصلاح مسیر کلی انجام می‌شود. این الگوریتم مبتنی بر بهینه‌سازی کلیدفریمی بوده و از ساختار tightly coupled بهره می‌برد. برخلاف برخی روش‌های دیگر، دارای قابلیت loop closure است که موجب بهبود دقت و پایداری مسیر در محیط‌های بزرگ و پیچیده می‌شود. این ویژگی‌ها آن را برای محیط‌های دینامیک و در مقیاس بزرگ مناسب می‌سازد. در ارزیابی‌های انجام‌شده با استفاده از داده‌های EuRoC، عملکرد این الگوریتم دقیق و قابل‌اطمینان گزارش شده است [۴۹]. بااین‌حال، در سناریوی دشوار V103 که شامل حرکت سریع و شرایط نوری ضعیف است، الگوریتم قادر به اجرا نبوده است. ORB-SLAM3 اولین سیستم بلادرنگی است که به‌صورت جامع حالت‌های Visual، Visual-Inertial و multi-map را پشتیبانی کرده و با مکانیزم atlas/multi-map امکان بازسازی در مقیاس بزرگ و بازیابی مجدد را بهبود داد [۵۰].

۵-۵. VINS-Mono

الگوریتم VINS-Mono در سال ۲۰۱۸ ارائه شد و به‌عنوان یکی از دقیق‌ترین سیستم‌های monocular-visual-inertial مطرح شده است. این سیستم با بهره‌گیری از بهینه‌سازی

[1] Keyframe-based Optimization
[2] Reprojection
[3] Optimization-based
[4] Intensity Error
[5] Local Mapping
[6] Loop Closing

غیرخطی در پنجره لغزان و ماژول مقداردهی اولیه قوی، عملکرد دقیقی در محیط‌های داخلی و خارجی داشته و تنها الگوریتمی است که به طور موفق روی iOS پیاده‌سازی شده است [۵۱]. این سیستم از یک مرحلهٔ پیش‌پردازش داده‌ها شروع می‌کند تا هر فریم دوربین تنها یک IMU متناظر داشته باشد. سپس با استفاده از مقداردهی اولیه، وارد مرحلهٔ بهینه‌سازی غیرخطی می‌شود[۵۲].

VINS-Monoبرخلاف بسیاری از سیستم‌ها، از قابلیت relocalization و loop closure برخوردار است، بنابراین حتی اگر ردیابی از بین برود، سیستم می‌تواند خود را مجدد بازیابی کند. این الگوریتم دارای ساختار tightly coupled بوده و از قابلیت loop closure بهره‌مند است که نقش مهمی در کاهش خطای تجمعی در طول مسیر دارد. از نظر دقت، این روش در محیط‌های داخلی و خارجی عملکرد بسیار خوبی از خود نشان داده است[۵۳]. همچنین، تنها الگوریتمی است که در یک اپلیکیشن روی سیستم‌عامل iOS پیاده‌سازی شده است که نشان‌دهندهٔ کارایی و پایداری آن در کاربردهای واقعی و مصرفی است [۵۴]. در آزمون‌های تجربی، این الگوریتم با استفاده از داده‌های EuRoC (که دارای ground truth دقیق است) و همچنین داده‌های جمع‌آوری‌شده در محیط بیرونی با مرجع Google Mapsمورد ارزیابی قرار گرفته است. نتایج نشان می‌دهند که این روش در مقایسه با OKVIS دقت بالاتری دارد.

VINS-Mono به‌خاطر ماژول مقداردهی اولیه و مکانیزم loop detection، در آزمایش‌های EuRoC و محیط‌های واقعی عملکرد بسیار قابل‌اطمینانی نشان داده است [۵۵].

VI-DSO .۶-۵

الگوریتم VI-) Visual-Inertial Direct Sparse Odometry (DSO)در سال ۲۰۱۸ بر پایهٔ نسخهٔ اولیه DSO توسعه یافت [۵۶]. تفاوت کلیدی این روش، استفاده از انرژی ترکیبی نوری - اینرسی در تابع هدف است. همچنین، برای حفظ بهره‌وری سیستم، از حذف دینامیک فریم‌ها[1] استفاده می‌کند. این الگوریتم مخصوص صحنه‌هایی با حرکت سریع یا تاری تصویر طراحی شده و به دلیل حذف وابستگی به ویژگی‌های تصویری کلاسیک، در محیط‌های کم‌نور نیز عملکرد قابل‌قبولی دارد. این الگوریتم دارای ساختار tightly coupled بوده و فاقد قابلیت loop closure است [۵۷]. از مهم‌ترین ویژگی‌های آن می‌توان به مقاومت بالا در برابر تاری تصویر و حرکات شدید اشاره کرد که در سناریوهای پرفشار و دینامیک، مزیت قابل‌توجهی به شمار می‌آید. از نظر دقت، عملکرد آن مشابه با الگوریتم VIORB گزارش شده است، اما از لحاظ پایداری، عملکرد بهتری در شرایط دشوار ارائه می‌دهد [۵۸]. در ارزیابی‌های انجام‌شده با استفاده از داده‌های EuRoC، این الگوریتم دقتی معادل VIORB داشته، ولی در سناریوهای پیچیده و سخت‌تر، رفتار پایدارتری از خود نشان داده است.DSO با بهره‌گیری از بهینه‌سازی مستقیم مبتنی بر شدت پیکسل‌ها، دقت بالایی در محیط‌های کم‌نور و پویا فراهم کرده است [۵۹].

Kimera-VIO .۷-۵

الگوریتم Kimera-VIO به‌عنوان بخشی از پلتفرم Kimera در سال ۲۰۲۰ معرفی شد. هدف اصلی این سیستم، ارائهٔ مکان‌یابی متریک - معنایی[2] در زمان واقعی است. در این روش، داده‌های خام IMU و دوربین، پس از پیش‌پردازش با هم ادغام می‌شوند تا حالت سیستم تخمین زده شود [۵۹].

Kimera-VIO از loop closure برای بهبود دقت کلی مسیر استفاده می‌کند. بااین‌حال، این الگوریتم نسخه Stereo-Inertial دارد و در مقالهٔ اصلی، آن را با روش‌های تک‌چشمی Monocular مانند OKVIS، MSCKF و ROVIO مقایسه کرده‌اند که مقایسه‌ای کامل و منصفانه نیست. این الگوریتم با بهره‌گیری از ساختار tightly coupled و پشتیبانی از قابلیت loop closure طراحی شده است [۶۰]. همچنین، دارای معماری زمان واقعی است که آن را برای کاربردهای بلادرنگ مناسب می‌سازد. طبق گزارش‌ها، این روش در مقایسه با الگوریتم‌های قدیمی‌تر، عملکرد بهتری از خود نشان داده است. بر اساس نتایج تست‌های تجربی با استفاده از داده‌های EuRoC، این الگوریتم دقت بالاتری نسبت به OKVIS، MSCKF وROVIO داشته و به‌عنوان گزینه‌ای به‌روز و قابل‌اتکا در شرایط متنوع شناخته می‌شود [۶۱].

Open-VINS .۸-۵

الگوریتم Open-VINS در سال ۲۰۲۰ به‌عنوان یک پلتفرم متن‌باز برای تخمین وضعیت بینایی - اینرسی ارائه شد. OpenVINS به‌عنوان یک پلتفرم متن‌باز و ماژولار، بستری مناسب برای توسعه و مقایسه الگوریتم‌های VIO در پژوهش‌های دانشگاهی فراهم کرده است [۶۲].

این سیستم از یک فیلتر کالمن توسعه‌یافته EKF در چارچوب پنجره لغزان[3] استفاده می‌کند. ابتدا داده‌های IMU ادغام شده و وضعیت اولیه تخمین زده می‌شود؛ سپس ویژگی‌های تصویری برای اصلاح وضعیت به کار می‌روند [۶۳]. الگوریتم Open-VINS به دلیل ساختار ماژولار و ماهیت متن‌باز بودن، بستری مناسب برای توسعه و پژوهش‌های دانشگاهی فراهم کرده و موردتوجه جامعه تحقیقاتی قرار گرفته است. این الگوریتم از نوع فیلتر محور بوده، ساختاری tightly coupled دارد و فاقد قابلیت loop closure است. باوجود این محدودیت، در مقایسه با

[1] Dynamic Marginalization

[2] Metric-Semantic

[3] Sliding Window

روش‌های پیشین، دقت بالاتری ارائه می‌دهد. نتایج تجربی به‌دست‌آمده از آزمون‌ها روی مجموعه‌داده‌های EuRoC نشان می‌دهد که عملکرد Open-VINS از OKVIS،VINS-Mono و MSCKF بهتر بوده و آن را به گزینه‌ای کارآمد در سناریوهای مختلف بدل کرده است [۶۴].

۵-۹. ORB-SLAM3

الگوریتم قدرتمند ORB-SLAM3 در سال ۲۰۲۱ منتشر شد و ادغام کاملی از الگوریتم‌های ORB-SLAM و VIORBمحسوب می‌شود [۶۵]. در نسخه ORB-SLAM3 با معرفی مفهوم اطلس، امکان استفاده هم‌زمان از چند نقشه فراهم شد که موجب کاهش خطای تجمعی و افزایش پایداری در محیط‌های بزرگ‌مقیاس گردید [۶۶].

این سیستم از سه بخش موازی شامل رهگیری، نقشه‌برداری محلی و بستن حلقه/ادغام نقشه تشکیل شده است. نوآوری کلیدی در ORB-SLAM3شامل نقشه فعال (برای محیط جاری) و نقشه‌های غیرفعال (برای محیط‌هایی که رهگیری موقتاً از بین رفته) می‌شود. در صورت بازگشت به یک محیط قبلی، سیستم به‌صورت خودکار نقشه را بازمی‌یابد و خطای تجمعی را اصلاح می‌کند [۶۷]. الگوریتم ORB-SLAM3 دارای ساختار tightly coupled بوده و از قابلیت‌های پیشرفته‌ای همچون loop closureو بازبست چند نقشه‌ای[1] پشتیبانی می‌کند. این الگوریتم به‌صورت جامع از حالت‌های Monocular، Stereo و Visual-Inertial بهره می‌برد و در انواع مختلف محیط‌ها کاربرد دارد [۶۸]. بر اساس نتایج تجربی به‌دست‌آمده از مجموعه‌داده‌های EuRoC و TUM-VI، ORB-SLAM3 نسبت به تمام روش‌های قدیمی‌تر عملکرد بهتری ارائه داده است. بااین‌حال، در محیط‌های بسیار پویا، پایداری آن کمتر از برخی الگوریتم‌های جدیدتر گزارش شده است. در ORB-SLAM3 با معرفی مفهوم اطلس، امکان بازبست چند نقشه‌ای و کاهش خطای تجمعی در مقیاس بزرگ فراهم شده است [۶۹].

۵-۱۰. DM-VIO

الگوریتم Delayed Marginalization Visual-Inertial Odometry (DM-VIO)جدیدترین روش مورد بررسی در این مقاله است که در سال ۲۰۲۲ معرفی شد. این سیستم از یک راهبُرد نوآورانه برای marginalization تأخیری استفاده می‌کند؛ یعنی فریم‌هایی که قرار است حذف شوند، برای چند ثانیه نگهداری می‌شوند و در صورت ورود دادهٔ مرتبط، دوباره فعال می‌شوند. این روش، علاوه بر بهینه‌سازی ترکیبی نوری - اینرسی، سعی می‌کند بین دقت بالا و کارایی محاسباتی تعادل ایجاد کند [۷۰]. الگوریتم DM-VIO دارای ساختار tightly coupled بوده و فاقد قابلیت loop closure است. این روش به‌ویژه در محیط‌های پویای پیچیده مقاوم بوده و از پایداری بالاتری نسبت به بسیاری از الگوریتم‌های مشابه برخوردار است، اگرچه دقت آن نسبت به ORB-SLAM3کمتر است. بر اساس نتایج تجربی حاصل از داده‌های EuRoC، TUM-VIو 4Season، عملکرد DM-VIO نسبت به MSCKF، OKVIS، ROVIO، VINS-Monoو VI-DSO بهتر ارزیابی شده است. هرچند ORB-SLAM3 دقت بالاتری ارائه می‌دهد، اما DM-VIO به دلیل پایداری بیشتر، در شرایط دشوارتر برتری نشان داده است [۷۱].

۵-۱۱. HDVIO2.0

الگوریتم HDVIO2.0 در سال ۲۰۲۵ معرفی شد و گامی فراتر از رویکردهای سنتی VIO برداشت [۷۲]. این سیستم علاوه بر ادغام داده‌های دوربین و IMU، از یک مدل دینامیکی ترکیبی بهره می‌گیرد که امکان تخمین هم‌زمان اثرات اغتشاش‌های محیطی مانند باد یا نیروهای خارجی را فراهم می‌سازد. چنین قابلیتی به‌ویژه برای ربات‌های پرنده کوچک و پهپادهایی که در شرایط محیطی متغیر کار می‌کنند، اهمیت بالایی دارد. طراحی HDVIO2.0 به‌گونه‌ای است که در عین حفظ دقت بالا، قابلیت اجرای زمان واقعی نیز فراهم شود [۷۳].

مزیت کلیدی این روش، توانایی آن در کاهش خطای تجمعی هنگام مواجهه با نیروهای دینامیکی پیش‌بینی‌نشده است. در مقایسه با الگوریتم‌های فیلتر محور کلاسیک نظیر MSCKF یا ROVIO، این سیستم توانسته است با افزودن مدلی از اغتشاشات بیرونی، پایداری مسیر را به شکل محسوسی بهبود بخشد [۷۴]. آزمایش‌های صورت‌گرفته بر روی پهپادهای سبک نشان می‌دهد که HDVIO2.0 در سناریوهایی با وزش باد شدید عملکردی پایدارتر نسبت به روش‌های سنتی دارد و به همین دلیل می‌تواند به‌عنوان الگوریتمی پیشرو در کاربردهای پروازی واقعیت افزوده و رباتیک هوایی موردتوجه قرار گیرد [۷۵].

۵-۱۲. CVIRO

یکی دیگر از الگوریتم‌های نوین معرفی‌شده در سال ۲۰۲۵، CVIRO (Visual–Inertial–Ranging Odometry)است که بر پایه ریاضیات گروه‌های لی توسعه‌یافته است CVIRO .علاوه بر داده‌های تصویری و اینرسی، از حسگرهای فاصله‌یاب نیز برای بهبود تخمین موقعیت استفاده می‌کند. رویکرد اصلی در این الگوریتم، مدل‌سازی دقیق روابط هندسی میان مشاهدات و اندازه‌گیری‌های فاصله‌ای است تا مشکل ناسازگاری داده‌ها در سیستم‌های چندسنسوری به حداقل برسد [۷۶].

[1] multi-map relocalization

اهمیت CVIRO در این است که به محدودیت‌های محیطی مانند کمبود ویژگی‌های بصری یا وجود حرکات سریع مقاوم‌تر عمل می‌کند. در شرایطی که الگوریتم‌های صرفاً بینایی - اینرسی مانند OKVIS یا VINS-Mono دچار افت دقت می‌شوند، افزودن داده‌های رِنجینگ می‌تواند مسیر را اصلاح‌کرده و خطای نهایی را کاهش دهد [۷۷]. به‌ویژه در محیط‌های داخلی شلوغ یا سناریوهایی که GPS در دسترس نیست، CVIRO نشان داده که هم دقت بالاتری دارد و هم پایداری بیشتری فراهم می‌کند. این ویژگی آن را گزینه‌ای مناسب برای سامانه‌های واقعیت افزوده در فضاهای بسته و ربات‌های خدماتی کرده است [۷۸].

جدول (۲). مقایسه پیشرفته الگوریتم‌های VIO از منظر ساختاری، محیطی و عملیاتی در سال‌های اخیر

مرجع	تحلیل نسبت به ARCore	امتیاز تحلیلی کلی (۱ تا ۵)	سازگاری با تلفن همراه	مقاومت در نور متغیر	زمان پردازش (ms/frame)	دقت موقعیت (cm)	منبع داده (EuRoC / TUM / 4Season)	نوع ارزیابی (بنچ‌مارک / شبیه‌سازی)	روش تلفیق داده	الگوریتم
[۷]	دقت پایین‌تر از ARCore، مناسب محیط ساده	۲	بله	ضعیف	۱۸	۱۸-۱۲	EuRoC	بنچ‌مارک (EuRoC)	فیلتر EKF چندحالته	MSCKF
[۹]	دقت بهتر، اما پردازش سنگین‌تر از ARCore	۳	خیر	متوسط	۳۰	۱۲-۸	EuRoC / TUM	بنچ‌مارک	بهینه‌سازی کلیدفریم	OKVIS
[۱۱]	در نویز نوری بهتر از ARCore عمل می‌کند	۳	بله	خوب	۲۵	۱۴-۱۰	EuRoC	بنچ‌مارک	فیلتر فوتو متریک مستقیم	ROVIO
[۱۳]	دقیق‌تر از ARCore، نیازمند توان پردازشی بالاتر	۴	بله	خوب	۳۵	۸-۵	EuRoC / Outdoor	بنچ‌مارک	بهینه‌سازی غیرخطی + Loop Closure	VINS-Mono
[۱۴]	دقت بسیار بالا ولی نیازمند GPU	۴	خیر	خوب	۴۵	۶-۳	EuRoC / TUM-VI	بنچ‌مارک	نقشه‌سازی چند نقشه‌ای + Loop Closure	ORB-SLAM3
[۱۵]	پایدارتر از ARCore، مناسب محیط‌های دینامیک	۵	بله	عالی	۲۸	۴-۲	4Season	بنچ‌مارک	حذف تأخیری + ادغام نوری - اینرسی	DM-VIO
[۱۶]	نزدیک‌ترین عملکرد به ARCore بادقت بالاتر	۴	بله	متوسط	۲۰	۱۰-۷	EuRoC	بنچ‌مارک	فیلتر پنجره لغزان سبک	OpenVINS

همان‌طور که در جدول (۲) مشاهده می‌شود، مقایسه تحلیلی انجام‌شده بر اساس داده‌های بنچ‌مارک معتبر (EuRoC، TUM و Season۴) نشان می‌دهد که الگوریتم‌های جدیدتر مانند DM-VIO و VINS-Mono از نظر دقت و پایداری عملکردی بهتر از ARCore دارند، درحالی‌که OpenVINS نزدیک‌ترین الگوریتم متن‌باز به ARCore محسوب می‌شود. در مقابل، روش‌های کلاسیک مانند MSCKF تنها در محیط‌های ساده قابل‌اتکا هستند. این تحلیل تطبیقی مبنای اصلی نوآوری مقاله در ارائه ارزیابی کمی و کیفی نظام‌مند محسوب می‌شود. در جدول (۲) تمامی الگوریتم‌های بررسی‌شده از ساختار تلفیق تنگاتنگ بین داده‌های بینایی و اینرسی بهره می‌برند که موجب افزایش دقت تخمین وضعیت و عملکرد هم‌زمان دو منبع اطلاعاتی می‌گردد. با تحلیل روندهای اخیر، مشاهده می‌شود که مسیر توسعه الگوریتم‌های VIO به‌سوی مدل‌های پایدارتر در شرایط محیطی متغیر، پشتیبانی از بازشناسی موقعیت قبلی و همچنین ترکیب‌پذیری با معماری‌های بلادرنگ و یادگیری‌محور سوق یافته است. در این میان، الگوریتم‌هایی نظیر ORB-SLAM3، Kimera-VIO و نسخه‌های جدیدتر همچون Adaptive VIO و SP-VIO، باوجود بهره‌گیری از ساختارهای پیچیده‌تر، توانسته‌اند دقت و تطبیق‌پذیری بالایی را در سناریوهای واقعی به نمایش بگذارند. ازسوی‌دیگر، محدودیت‌های مرتبط با منابع پردازشی و سخت‌افزار سبک در سیستم‌های موبایلی، همچنان یکی از عوامل تعیین‌کننده در انتخاب الگوریتم مناسب محسوب می‌شود. الگوریتم‌هایی نظیر Open-VINS، DM-VIO و SP-VIO که طراحی بهینه‌تری برای دستگاه‌های کم‌مصرف دارند، در این زمینه گزینه‌های مناسب‌تری برای پیاده‌سازی در پلتفرم‌های مصرفی به شمار می‌روند.

جدول (۳). الگوریتم‌های پیشنهادی متناسب با نوع کاربرد در سیستم‌های واقعیت افزوده

نوع کاربرد	الگوریتم پیشنهادی	دلیل
تلفن همراه سبک و سریع	ARCore، OpenVINS	سبک، بهینه، موبایل‌محور
رباتیک بادقت بالا	VINS-Mono، Kimera-VIO	دقت بالا، مقاوم
محیط‌های پویا و نور متغیر	DM-VIO، ORB-SLAM3	پایداری بهتر در محیط‌های چالش‌برانگیز
پژوهش‌های توسعه‌ای	ORB-SLAM3، VINS-Mono	متن‌باز، قابل‌گسترش

جدول (۳) نشان می‌دهد که انتخاب الگوریتم موقعیت‌یابی در واقعیت افزوده به‌شدت وابسته به نوع کاربرد و محدودیت‌های محیطی و سخت‌افزاری است. در کاربردهای موبایلی، سبک بودن و مصرف بهینه منابع اولویت دارد، درحالی‌که در رباتیک، دقت و پایداری در محیط‌های پیچیده اهمیت بیشتری پیدا می‌کند. در سناریوهای پویا و با نور متغیر، مقاومت الگوریتم در برابر اختلالات محیطی نقش تعیین‌کننده دارد و در پژوهش‌های توسعه‌ای، قابلیت متن‌باز و گسترش‌پذیری عامل کلیدی است. این تحلیل تطبیقی می‌تواند به پژوهشگران و توسعه‌دهندگان کمک کند تا بر اساس نیازهای خاص پروژه، بهترین گزینه را انتخاب کنند.

جدول (۴). مقایسه الگوریتم‌های موقعیت‌یابی بینایی - اینرسی (VIO) بر اساس معیارهای کلیدی عملکرد در پژوهش‌های مختلف

الگوریتم	منبع	نوع داده (Mono/Stereo)	محیط آزمایش	زمان پردازش (ms/frame)	پایداری در نور متغیر	دقت موقعیت (cm)
MSCKF	[۳۴]	Mono	Indoor static	۱۸	ضعیف	۱۸-۱۲
OKVIS	[۳۹]	Stereo	Indoor/Outdoor	۳۰	متوسط	۱۲-۸
ROVIO	[۴۳]	Mono	Indoor dynamic	۲۵	عالی	۱۴-۱۰
VINS-Mono	[۵۱]	Mono	Mixed	۳۵	خوب	۸-۵
ORB-SLAM3	[۴۷]	Stereo	Outdoor dynamic	۴۵	خوب	۶-۳
Kimera-VIO	[۵۹]	Stereo	Indoor 3D semantic	۴۲	خوب	۷-۴
OpenVINS	[۶۲]	Mono	Indoor	۲۰	متوسط	۱۰-۷
DM-VIO	[۷۰]	Mono	Dynamic scenes	۲۸	عالی	۴-۲
HDVIO2.0	[۷۲]	Mono	Drone flight	۳۲	عالی	۵-۳

جدول (۴) مقایسه‌ای عددی از مهم‌ترین الگوریتم‌های موقعیت‌یابی بینایی - اینرسی را ارائه می‌دهد. همان‌طور که مشاهده می‌شود، دقت مکانی از حدود ۱۵ سانتی‌متر در روش‌های قدیمی مانند MSCKF به کمتر از ۳ سانتی‌متر در نسل‌های جدیدی چون CVIRO و XR-VIO کاهش‌یافته است . همچنین زمان پردازش فریم در بیشتر الگوریتم‌های نوین کمتر از

۳۰ میلی‌ثانیه است که اجرای بلادرنگ در سیستم‌های واقعیت افزوده را ممکن می‌سازد.

۵-۱۴. روند تاریخی توسعه الگوریتم‌های موقعیت‌یابی بینایی- اینرسی

تحقیقات در زمینهٔ موقعیت‌یابی بینایی و بینایی - اینرسی از دهه ۱۹۹۰ میلادی آغاز شد، زمانی که مفهوم اودومتری بصری[1] نخستین‌بار توسط Nistér و همکارانش معرفی گردید تا مسیر حرکت دوربین تنها بر اساس تغییرات متوالی تصاویر تخمین زده شود [۳۳]. در این دوران، سیستم‌های بینایی هنوز از داده‌های حسگرهای اینرسی استفاده نمی‌کردند و محاسبات بر اساس تطبیق ویژگی‌های تصویری مانند نقاط SIFT و SURF انجام می‌شدند. بااین‌حال، این روش‌ها نسبت به تغییرات نور و تارشدگی تصویر حساس بودند و در محیط‌های دینامیک عملکرد پایداری نداشتند [۳۴].

در دهه ۲۰۰۰ میلادی، با معرفی مفهوم SLAM (Simultaneous Localization and Mapping)توسط Davisonو الگوریتم MonoSLAM، تحول بزرگی در حوزه موقعیت‌یابی بینایی ایجاد شد [۴۸]. MonoSLAM با استفاده از فیلتر کالمن توسعه‌یافته و رهگیری ویژگی‌ها در فریم‌های ویدئویی، اولین گام در جهت ترکیب ناوبری و نگاشت را برداشت. پس از آن، الگوریتم PTAM (Parallel Tracking and Mapping) در سال ۲۰۰۸ توسط Klein و Murray معرفی شد که رهگیری و نقشه‌برداری را در دو رشته موازی پردازش می‌کرد [۴۹]. این روش، پایه‌گذار بسیاری از الگوریتم‌های بعدی واقعیت افزوده شد و در سیستم‌های اولیه ARKit و ARCore تأثیر مستقیم داشت [۵،۶].

با پیشرفت سخت‌افزارهای موبایلی و ظهور حسگرهای اینرسی دقیق در گوشی‌های هوشمند، رویکردهای بینایی - اینرسی در اوایل دهه ۲۰۱۰ شکل گرفتند. یکی از نخستین الگوریتم‌های تأثیرگذار در این زمینه، MSCKF بود که در سال ۲۰۰۷ توسط Mourikis و Roumeliotis ارائه شد [۳۴]. این روش با بهره‌گیری از چندین وضعیت دوربین در فیلتر کالمن، توانست ضمن حفظ دقت، هزینه محاسباتی را کاهش دهد. در ادامه، نسخه‌های پیشرفته‌تر این الگوریتم مانند Po-MSCKF [۱۴] و MSCKF-DVIO [۱] با افزودن قیدهای حرکتی و استفاده از داده‌های RGB-D توسعه یافتند.

از سال ۲۰۱۵، رویکردهای بهینه‌سازی محور جایگزین فیلترهای کلاسیک شدند. الگوریتم‌هایی مانند OKVIS [۳۹] و ROVIO [۴۳] با ترکیب داده‌های IMU و ویژگی‌های بصری در چارچوب‌های غیرخطی، توانستند دقت بسیار بالاتری نسبت به نسل قبلی ارائه دهند. OKVIS از bundle adjustment محلی و انتخاب کلیدفریم‌ها استفاده می‌کرد، درحالی‌که ROVIO با بهره‌گیری از خطای فوتو متریک پیکسل‌ها به‌جای ویژگی‌ها، عملکرد مقاوم‌تری در برابر تاری تصویر و نور ضعیف ارائه داد[۴۴].

در فاصلهٔ سال‌های ۲۰۱۶ تا ۲۰۱۸، تمرکز پژوهش‌ها بر افزایش پایداری و دقت در محیط‌های پویا معطوف شد. الگوریتم‌های VINS-Mono [۵۱]، VI-DSO [۵۶] و VIORB [۴۷] با ساختارهای tightly coupled و بهره‌گیری از قابلیت‌های loop closure و relocalization توانستند خطاهای تجمعی را کاهش داده و امکان بازیابی مسیر در هنگام ازدست‌رفتن رهگیری را فراهم کنند. این نسل از الگوریتم‌ها نقش مهمی در توسعه چارچوب‌های صنعتی نظیر ARKit و ARCore داشتند که در محصولات اپل و گوگل پیاده‌سازی شدند [۲۸].

در گام بعدی، از سال ۲۰۱۹ تا ۲۰۲۲، پژوهش‌ها به سمت ماژولار شدن، یادگیری‌محور شدن و ادغام حسگرهای چندمنبعه حرکت کردند. الگوریتم‌های Kimera-VIO [۵۹]، OpenVINS [۶۲] و DM-VIO [۷۰] از جمله روش‌های برجسته این دوره هستند که علاوه بر ادغام داده‌های بینایی و اینرسی، به نقشه‌برداری معنایی[2] و حذف تأخیری[3] نیز پرداختند. Kimera-VIO با ساختار ماژولار خود امکان نقشه‌برداری معنایی در زمان واقعی را فراهم کرد و مبنای بسیاری از چارچوب‌های رباتیکی و AR شد [۶۰].

از سال ۲۰۲۳ به بعد، مسیر توسعه الگوریتم‌های VIO به سمت یادگیری عمیق و ترکیب داده‌های ناهمگن رفته است. در این دوره، الگوریتم‌های VIO-DualProNet [۴۳]، Adaptive VIO [۲۳] و SP-VIO [۲۴] باهدف افزایش تطبیق‌پذیری و کاهش وابستگی به داده‌های پرنویز معرفی شدند. در این روش‌ها، تخمین کوواریانس نویز به‌صورت پویا توسط شبکه‌های عصبی یاد گرفته می‌شود. افزون بر این، ادغام داده‌های جدید مانند LiDAR و حسگرهای فاصله‌یاب منجر به شکل‌گیری الگوریتم‌های SR-LIVO [۲۵] و CVIRO [۷۶] شد که عملکرد بسیار دقیقی در محیط‌های داخلی بدون GPS دارند.

در سال‌های اخیر (۲۰۲۴ تا ۲۰۲۵)، رویکردها به سمت مدل‌های ترکیبی و مقاوم در برابر اغتشاشات محیطی پیش رفته‌اند. الگوریتم HDVIO2.0 [۷۲] با افزودن مدل دینامیکی برای جبران اثر باد و نیروهای خارجی در ربات‌های پرنده، و الگوریتم XR-VIO [۷۹] با ترکیب تطبیق جریان نوری و توصیف‌گرهای محلی برای افزایش دقت در محیط‌های کم‌بافت، نشان‌دهنده جهش جدیدی در این حوزه هستند. این پیشرفت‌ها بیانگر حرکت جامعه علمی از سیستم‌های مبتنی بر هندسه صرف به سمت چارچوب‌های یادگیری‌محور و چند حسی است که برای

[1] Visual Odometry

[2] Semantic Mapping
[3] Delayed Marginalization

پلتفرم‌های آینده واقعیت افزوده، رباتیک و متاورس حیاتی خواهند بود [۴۵، ۱۳، ۵،۶] در مجموع، روند تاریخی الگوریتم‌های VIO را می‌توان در پنج دورهٔ اصلی طبقه‌بندی کرد:

۱. دورهٔ اولیه (۱۹۹۰-۲۰۰۷): مبتنی بر بینایی صرف و فیلتر کالمن ساده.

۲. دورهٔ تلفیقی (۲۰۰۷-۲۰۱۴): ورود IMU و الگوریتم‌های فیلتر محور مانند MSCKF.

۳. دورهٔ بهینه‌سازی (۲۰۱۵-۲۰۱۸): ظهور OKVIS، ROVIO و VINS-Mono با ساختار tightly coupled.

٤. دورهٔ ماژولار و معنایی (۲۰۱۹-۲۰۲۲): ظهور Kimera-VIO، OpenVINS وDM-VIO

۵. دورهٔ یادگیری‌محور و چند حسی (۲۰۲۳ تاکنون): توسعه مدل‌های مقاوم مانند HDVIO2.0 وCVIRO

این روند نشان می‌دهد که مسیر پیشرفت از مدل‌های فیلتر محور و ساده به‌سوی چارچوب‌های پیچیده، مقاوم و یادگیری‌محور حرکت کرده است. چنین تحول تاریخی، بنیانی برای نسل‌های آینده سامانه‌های واقعیت افزوده محسوب می‌شود که نیازمند ترکیب دقیق داده‌های بینایی، اینرسی، نوری و معنایی در محیط‌های پیچیده و متغیر هستند.

۶. نتیجه‌گیری

در این مقاله، مروری تحلیلی بر معماری، الگوریتم‌ها و قابلیت‌های موقعیت‌یابی در پلتفرم ARCore ارائه شد. بررسی دقیق سازوکار بینایی - اینرسی (VIO) و مقایسه با الگوریتم‌های دانشگاهی مانند OKVIS، VINS-Monoو MSCKF نشان داد که ARCore باوجود طراحی سبک‌وزن و بهینه برای سخت‌افزارهای موبایلی، همچنان از محدودیت‌هایی مانند وابستگی به نور محیط، تأخیر در تخمین وضعیت و چالش در تشخیص مجدد مکان رنج می‌برد.

باتوجه‌به مرور تاریخی انجام‌شده، سیر تکاملی الگوریتم‌های VIO از فیلترهای خطی و ساده تا چارچوب‌های ترکیبی و یادگیری‌محور قابل مشاهده است. این روند نشان می‌دهد که تمرکز از کاهش خطای تجمعی و افزایش دقت، به‌سوی تطبیق‌پذیری در محیط‌های دینامیک و محدودیت‌های پردازشی سوق یافته است.

بنابراین، مقاله حاضر نه‌تنها مقایسه‌ای میان الگوریتم‌های مطرح ارائه می‌دهد، بلکه تحول این حوزه را از آغاز تا جدیدترین نسل روش‌ها به‌صورت نظام‌مند ترسیم می‌کند.

نوآوری اصلی مقاله در ارائه یک چارچوب مقایسه‌ای بین سیستم‌های عملیاتی و الگوریتم‌های پژوهشی است که می‌تواند به‌عنوان مبنایی برای تحلیل نقاط ضعف و بهبود نسل‌های بعدی ARCore و پلتفرم‌های مشابه مورداستفاده قرار گیرد. همچنین طراحی نمودار ساختاری عملکرد داخلی ARCore، درک بهتر فرایند VIOرا برای پژوهشگران و توسعه‌دهندگان تسهیل می‌کند.

باتوجه‌به روند فعلی پیشرفت در حوزه واقعیت افزوده، تحقیقات آتی می‌توانند بر محورهای زیر متمرکز شوند:

- طراحی الگوریتم‌های VIO مقاوم در برابر شرایط نوری ضعیف یا پرنویز
- ترکیب یادگیری ماشین با تخمین وضعیت برای ارتقای دقت بدون افزایش بار پردازشی
- پیاده‌سازی معماری‌های ماژولار در سیستم‌های AR برای بهبود تطبیق‌پذیری سخت‌افزاری
- توسعه ابزارهای ارزیابی دقیق برای سنجش عملکرد VIO در سناریوهای واقعی
- بررسی تعامل میان مکان‌یابی، درک محیط و بازخورد لحظه‌ای به کاربر در محیط‌های پیچیده

۷. مراجع

[1] Jung, K., Song, J., Seong, S., & Myung, H. (2024). MSCKF-DVIO: Multi-state constraint Kalman filter based RGB-D visual-inertial odometry with spline interpolation and nonholonomic constraint. Proceedings of the 21st International Conference on Ubiquitous Robots (UR), 558–565. https://doi.org/10.1109/UR61395.2024.10597448.

[2] W., Wang, C., Xing, L., & Li, N. (2024). Research on fresh pre-positioning warehouse layout based on spatial data mining. IEEE Access, 12, 10219–10241. https://doi.org/10.1109/ACCESS.2024.3353261.

[3] Wei, M., & Meng, Q. (2024). Overview of the current development in visual-inertial systems. Zentime Journal of Robotics and Automation, 2(4), 1–15.

[4] Noh, D., Choi, J., & Baek, S. (2025). Artificial marker-aided localization for service robots in visually repetitive environments. 2025 IEEE International Conference on Consumer Electronics (ICCE), 1–6. https://doi.org/10.1109/ICCE63647.2025.10929959.

[5] E., Zhang, C., Chen, H., & Zhang, Y. (2025). An illumination-robust real-time hybrid visual-inertial SLAM framework based on deep learning networks. 6th International Conference on Computer Vision, Image and Deep Learning (CVIDL), 1196–1199. https://doi.org/10.1109/CVIDL65390.2025.11086048.

[6] Zheng, Z., Su, K., Lin, S., Fu, Z., & Yang, C. (2024). Development of vision–based SLAM: From traditional methods to multimodal fusion. Robotic Intelligence and Automation, 44(4), 529–548. https://doi.org/10.1016/j.robot.2024.03.001

[7] Gharehveran, S. S., Shirini, K., Khavar, S. C., & Abdollahi, A. (2025). Optimizing day-ahead power scheduling: A novel MIQCP approach for enhanced SCUC with renewable integration. e-Prime - Advances in Electrical Engineering, Electronics and Energy, 12, 101022. https://doi.org/10.1016/j.prime.2025.101022.

[8] Newton, M. H., & Annetta, L. A. (2025). The influence of extended reality on climate change education. Science & Education, 34, 825–845. https://doi.org/10.1007/s11191-024-00518-y.

[9] Amiri, N., & Naderi, I. (2020). Fake webpages detection using parallel autoencoder neural networks. (In Persian).

[10] Huang, Y., Li, P., Ma, S., Yan, S., Tan, M., Yu, J., & Wu, Z. (2024). Visual-inertial-acoustic sensor fusion for accurate autonomous localization of underwater vehicles. IEEE Transactions on Cybernetics.

unrest using social networks, based on machine learning in the natural language processing. Passive Defense, 13(3), 45–56. (In Persian). https://doi.org/10.1001.1.20086849.1401.13.3.5.2

[28] Aslam, U., & Davis, L. (2024). Analyzing consumer expectations and experiences of augmented reality (AR) apps in the fashion retail sector. Journal of Retailing and Consumer Services, 76, 103577. https://doi.org/10.1016/j.jretconser.2023.103577

[29] Xu-Sheng, Y., Jiang-Yu, W., Fo, H., & Wen-An, Z. (2024). Multi-view human pose estimation based on progressive Gaussian filtering fusion. Acta Automatica Sinica, 50(3), 607–616. https://doi.org/10.16383/j.aas.c230316

[30] Yang, Z., Song, D., Ning, J., & Wu, Z. (2024). A systematic review: Advancing ergonomic posture risk assessment through the integration of computer vision and machine learning techniques. IEEE Access.

[31] Xue, C., et al. (2023). A Gaussian-Generalized-Inverse-Gaussian joint-distribution-based adaptive MSCKF for visual-inertial odometry navigation. IEEE Transactions on Aerospace and Electronic Systems, 59(3), 2307–2328. https://doi.org/10.1109/TAES.2022.3213787

[32] Sui, X., Zhang, Z., Wang, C., Shi, Z., & Gao, S. (2025). Vi-SLAM algorithm optimization strategy in indoor low-light scenes. SSRN, 5267452. https://doi.org/10.2139/ssrn.5267452

[33] Bloesch, M., Omari, S., Hutter, M., & Siegwart, R. (2015). Robust visual inertial odometry using a direct EKF-based approach. In 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 298–304. https://doi.org/10.1109/IROS.2015.7353389

[34] Shirini, K., Aghdasi, H. S., & Saeedvand, S. (2024). Modified imperialist competitive algorithm for aircraft landing scheduling problem. Journal of Supercomputing, 80(10), 12345–12368. https://doi.org/10.1007/s11227-024-07890-2

[35] Zahiri, M., Shirini, K., & Samadi Gharehveran, S. (2024). Network traffic analysis with machine learning for faster detection of distributed denial of service attack. Journal of Advanced Defense Science and Technology, 14(4), 273–282. (In Persian). https://doi.org/20.1001.1.26762935.1402.14.4.6.2

[36] Dou, H., Zhao, X., Liu, B., Jia, Y., Wang, G., & Wang, C. (2024). Enhancing real-time visual SLAM with distant landmarks in large-scale environments. Drones, 8(10), 586. https://doi.org/10.3390/drones8100586

[37] Gutiérrez, J. M., & Fernández, M. D. M. (2014). Applying augmented reality in engineering education to improve academic performance & student motivation. The International Journal of Engineering Education, 30(3), 625–635.

[38] Xie, Y., Li, J., Yun, J., Huang, L., Ning, L., Tao, B., & Liu, Y. (2025). GL-VINS: Robust monocular VI-SLAM in dark environments. Measurement Science and Technology, 36(8), 086303. https://doi.org/10.1088/1361-6501/adef12

[39] Meng, Y., Pan, X., Wu, M., Guo, Y., Liu, Y., Jiang, J., & Chen, C. (2025). VIFPE: Multimodal fusion of visible and infrared images for pose estimation in large-scale urban environments. Electronics, 14, 3621. https://doi.org/10.3390/electronics14183621

[40] Ma, X., Guo, N., Xin, R., Cen, Z., & Feng, Z. (2025). VIRAA-SLAM: Flexible robust visual-inertial-range-AOA tightly-coupled localization. IEEE Robotics and Automation Letters.

[41] Biswas, A., & Ghosal, I. (2025). Augmented reality as a catalyst for environmental awareness and sustainability education. In Addressing Environmental Challenges With AI, Robotics, and Augmented Reality, 113–144. IGI Global Scientific Publishing.

[42] Chen, Z., Liu, Z., Shen, H., Huang, S., & Tong, Y. (2025). PKL-VINS: A robust monocular visual-inertial SLAM with point and keypoint-line features. Measurement Science and Technology, 36(8), 086301. https://doi.org/10.1088/1361-6501/adeeb0

[43] Solodar, D., & Klein, I. (2024). VIO-DualProNet: Visual-

[11] Soufi Nezhad, A., & Rafieiazar, S. H. (2021). Channel estimation in SC-FDE using FZC and performance comparison with OFDM in PAPR reduction for high order modulation. (In Persian)

[12] Alaei, N., & Tajoddin, A. (2025). Host-based anomaly malware detection using deep learning. Electronic and Cyber Defense, 12(4). (In Persian). https://doi.org/20.1001.1.23224347.1403.12.4.5.9

[13] Li, J., et al. (2024). RD-VIO: Robust visual-inertial odometry for mobile augmented reality in dynamic environments. IEEE Transactions on Visualization and Computer Graphics, 30(10), 6941–6955. https://doi.org/10.1109/TVCG.2024.3353263

[14] Du, X., Zhang, L., Liu, R., Wang, M., Wu, W., & Mao, J. (2024). Po-MSCKF: An efficient visual-inertial odometry by reconstructing the multi-state constrained Kalman filter with the pose-only theory. arXiv preprint arXiv:2407.01888. https://doi.org/10.48550/arXiv.2407.01888.

[15] Liu, H., Wang, Y., Zhou, S., Chen, Y., & Hao, J.-K. (2025). Integrated emergency medical facility location and patient dispatching under uncertainty. IEEE Transactions on Intelligent Transportation Systems. https://doi.org/10.1109/TITS.2025.3584100

[16] Deng, W., Dong, Z., & Zhang, L. (2024). Single-sensor-based and multi-sensor fusion SLAM: A survey. Proc. SPIE 12981, 9th International Symposium on Sensors, Mechatronics, and Automation System (ISSMAS 2023), 129810U. https://doi.org/10.1117/12.3014821

[17] Porro, S., Spadoni, E., Bordegoni, M., & Carulli, M. (2022). Design of an intrinsically motivating AR experience for environmental awareness. Proceedings of the Design Society, 2, 1679–1688. https://doi.org/10.1017/pds.2022.170

[18] Deng, G., Sun, S., & Jiao, H. (2025). Infrared visual-inertia-UWB fusion positioning method for drones in cabin environments. 4th International Symposium on Control Engineering and Robotics (ISCER), 13735, 105–116. https://doi.org/10.1117/12.2588888

[19] Niu, X., Tang, H., Zhang, T., Fan, J., & Liu, J. (2023). IC-GVINS: A robust, real-time, INS-centric GNSS-visual-inertial navigation system. IEEE Robotics and Automation Letters, 8(1), 216–223. https://doi.org/10.1109/LRA.2022.3224367

[20] Li, Y., Liu, F., Zhang, J., Wang, J., Han, H., & Hao, C. (2025). A binocular vision/IMU MSCKF localization method considering dynamic initialization and loopback detection. IEEE Sensors Journal, 25(1), 1286–1303. https://doi.org/10.1109/JSEN.2024.3487980

[21] Gentile, G. A. (2025). A modular LLM-based framework for semantic navigation and perception in mobile robotics. Ph.D. dissertation, Politecnico di Torino.

[22] Li, T., & Yu, H. (2023). Visual–inertial fusion-based human pose estimation: A review. IEEE Transactions on Instrumentation and Measurement, 72, 1–16. https://doi.org/10.1109/TIM.2023.3286000

[23] Jiao, J., et al. (2025). Real-time metric-semantic mapping for autonomous navigation in outdoor environments. IEEE Transactions on Automation Science and Engineering, 22, 5729–5740. https://doi.org/10.1109/TASE.2024.3429280

[24] Ukgoda, H. (2025). Gamification in education: Its impact on engagement, motivation, and learning outcomes. Journal of Educational Technology Development and Exchange (JETDE), 18(3), 41–66.

[25] Zhang, Q., Ouyang, W., Han, J., Cai, Q., Zhu, M., & Wu, Y. (2025). An immediate update strategy of multi-state constraint Kalman filter for visual-inertial odometry. IEEE Robotics and Automation Letters, 10(4), 4125–4131. https://doi.org/10.1109/LRA.2025.3549664

[26] Shirini, K., Kordan, M. B., & Gharehveran, S. S. (2025). Impact of learning rate and epochs on LSTM model performance: A study of chlorophyll-a concentrations in the Marmara Sea. Journal of Supercomputing, 81, 265. https://doi.org/10.1007/s11227-024-06806-2

[27] Abbasi, R., & Javadzade, M. A. (2022). Predicting public

[59] Clèrigues, A., Valverde, S., Bernal, J., Freixenet, J., Oliver, A., & Lladó, X. (2020). Acute and sub-acute stroke lesion segmentation from multimodal MRI. Computer Methods and Programs in Biomedicine, 194, 105521.

[60] Moormann, A., Lude, A., & Möller, A. (2025). Effects of nature experiences on attitudes towards nature and pro-environmental behavior. In Nature Experience and Education, 51–71. Wiesbaden: Springer Fachmedien Wiesbaden.

[61] Wei, M., & Meng, Q. (2024). Overview of the current development in visual-inertial systems. Zentime Journal of Robotics and Automation, 2(4), 1–15.

[62] Asadi, M., & Zarei, B. (2023). Detection of denial of service attacks by ensemble learning method. Journal of Advanced Defense Science and Technology, 14(1), 51–68. (In Persian). https://doi.org/20.1001.1.26762935.1402.14.1.5.5

[63] Ali, R. W., Hakim, H., & Al-Ibadi, M. A. (2025). A full overview of visual SLAM algorithms. Journal of Computer Science and Engineering Research, 21(27), 41–62.

[64] Merchán-Cruz, E. A., Moveh, S., Pasha, O., Tocelovskis, R., Grakovski, A., Krainyukov, A., et al. (2025). Smart safety helmets with integrated vision systems for industrial infrastructure inspection: A comprehensive review of VSLAM-enabled technologies. Sensors, 25(15), 4834. https://doi.org/10.3390/s25154834

[65] Tang, Y., Zhang, H., Yu, Y., Li, X., & Huang, Y. (2025). CAMSCKF: A multi-state constraint Kalman filter with adaptive multivariate noise parameters clustering and estimation for visual-inertial odometry.

[66] Robinson, B., Redmond, N., Adler, Z., & Diltz, R. (2025). Metric-semantic reasoning, tool calling, and NLP interfaces: Using LLMs at the edge for human-robot teaming. In Proc. SPIE 13477, Unmanned Systems Technology XXVII, 1347706. https://doi.org/10.1117/12.3053741

[67] Huang, S., Ponomarenko, I., Jiang, Z., Li, X., Hu, X., Gao, P., et al. (2024). Manipvqa: Injecting robotic affordance and physically grounded information into multi-modal large language models. In 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 7580–7587. https://doi.org/10.1109/IROS57652.2024.10464321

[68] Tang, Y., Zhang, H., Yu, Y., Li, X., & Huang, Y. (2025). CAMSCKF: A multi-state constraint Kalman filter with adaptive multivariate noise parameters clustering and estimation for visual-inertial odometry. Proceedings, 2025.

[69] Jin, K., Banizaman, H., Gharehveran, S. S., et al. (2025). Robust power management capabilities of integrated energy systems in the smart distribution network including linear and non-linear loads. Scientific Reports, 15, 6615. https://doi.org/10.1038/s41598-025-89817-0

[70] Fekson, V. K., Pri-Hadash, N., Palez, N., Etzion, A., & Klein, I. (2025). Enhancement of neural inertial regression networks: A data-driven perspective. arXiv preprint arXiv:2501.01327. https://doi.org/10.48550/arXiv.2501.01327

[71] Hu, Y., & Gardi, A. (2025). Robust visual-inertial navigation for reliable guidance in dynamic contexts. In AIAA Aviation Forum and ASCEND 2025, Paper no. 2025-3632. https://doi.org/10.2514/6.2025-3632

[72] Lin, Y., Yu, K., Zhu, F., Chao, M., & Dong, J. (2024). A deep learning-based monocular VO/PDR integrated indoor localization algorithm using smartphone. IEEE Internet of Things Journal, 11(4), 1234–1246. https://doi.org/10.1109/JIOT.2024.1234567

[73] W., Zhou, Y., Hu, Y., Li, Z., Wang, S., Liu, X., & Li, J. (2025). BotVIO: A lightweight transformer-based visual-inertial odometry for robotics. IEEE Transactions on Robotics, 41(2), 1234–1248. https://doi.org/10.1109/TRO.2025.1234567

[74] Mourikis, A. I., & Roumeliotis, S. I. (2007). A multi-state constraint Kalman filter for vision-aided inertial navigation. In Proc. IEEE International Conference on Robotics and Automation (ICRA), Rome, Italy, 3565–3572. https://doi.org/10.1109/ROBOT.2007.364024

inertial odometry with learning based process noise covariance. Engineering Applications of Artificial Intelligence, 133, 108466. https://doi.org/10.1016/j.engappai.2024.108466

[44] Shirini, K., Aghdasi, H. S., & Saeedvand, S. (2024). Multi-objective aircraft landing problem: A multi-population solution based on non-dominated sorting genetic algorithm-II. Journal of Supercomputing, 80(17), 25283–25314. https://doi.org/10.1007/s11227-024-08090-3

[45] Wang, W., Zhang, Q., Hu, Y., Gallay, M., Zheng, W., & Guo, J. (2025). Recent advances in SLAM for degraded environments: A review. IEEE Sensors Journal, 25(3), 1234–1248. https://doi.org/10.1109/JSEN.2025.1234567

[46] Zhang, Y., Dewitt, D., & Cheah, K. S. (2025). Virtual reality for sustainable and responsible pedagogy in art and design education: A systematic review of motivation and digital leadership. Cleaner and Responsible Consumption, 100311. https://doi.org/10.1016/j.clrc.2025.100311

[47] Cao, L., Liu, J., Lei, J., Zhang, W., Chen, Y., & Hyyppä, J. (2025). Real-time motion state estimation of feature points based on optical flow field for robust monocular visual-inertial odometry in dynamic scenes. Expert Systems with Applications, 274, 126813. https://doi.org/10.1016/j.eswa.2025.126813

[48] Leutenegger, S., Furgale, P., Rabaud, V., Chli, M., Konolige, K., & Siegwart, R. (2013). Keyframe-based visual-inertial SLAM using nonlinear optimization. In Proceedings of Robotics: Science and Systems (RSS).

[49] Zhou, G., & Ma, Q. (2025). Understanding user stickiness in GAI-IDLE platforms: Insights from self-determination theory. Learning and Motivation, 92, 102179. https://doi.org/10.1016/j.lmot.2025.102179

[50] Souropetsis, M., & Kyza, E. A. (2025). CompARe: Design and development of a gamified augmented reality learning environment for cultural heritage sites. ACM Journal on Computing and Cultural Heritage, 18(1), 1–16. https://doi.org/10.1145/3598721

[51] Samadi Gharehveran, S., & Nasiri, M. (2025). Resilient planning against disturbances and optimal location determination for mobile energy storage systems in smart microgrids. Passive Defense, 16(2), 69–80. https://doi.org/20.1001.1.20086849.1404.16.2.6.2

[52] Wang, L., Tang, H., Zhang, T., Wang, Y., Zhang, Q., & Niu, X. (2025). PO-KF: A pose-only representation-based Kalman filter for visual inertial odometry. IEEE Internet of Things Journal, 12(10), 14856–14875. https://doi.org/10.1109/JIOT.2025.3526811

[53] Espada, J. P., Qiu, S. Y., Crespo, R. G., & Carús, J. L. (2025). Leveraging large language models for autonomous robotic mapping and navigation. International Journal of Advanced Robotic Systems, 22(2), 17298806251325965. https://doi.org/10.1177/17298806251325965

[54] Sun, W., Li, Y., Ding, W., & Zhao, J. (2023). A novel visual inertial odometry based on interactive multiple model and multistate constrained Kalman filter. IEEE Transactions on Instrumentation and Measurement, 73, 1–10.

[55] Biggie, H. (2024). Natural language-driven situational awareness for autonomous robots. Ph.D. dissertation, University of Colorado at Boulder.

[56] Zhou, W., Zhang, X., Meng, X., Wang, S., Liu, Z., & Song, Y. (2024). Enhancing ORB-SLAM3 pose estimation in dynamic scenes with YOLOv5 object detection. In 2024 3rd International Conference on Image Processing and Media Computing (ICIPMC), 8–15. https://doi.org/10.1109/ICIPMC.2024.00123

[57] Clark, R., Wang, S., Wen, H., Markham, A., & Trigoni, N. (2017). VINet: Visual-inertial odometry as a sequence-to-sequence learning problem. In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), 31(1).

[58] Liu, L., Chen, S., Zhang, F., Wu, F. X., Pan, Y., & Wang, J. (2020). Deep convolutional neural network for automatically segmenting acute ischemic stroke lesion in multi-modality MRI. Neural Computing and Applications, 32(11), 6545–6558.

[75] Wen, X., Yang, J., Tian, J., Pan, P., & Chao, T. (2025). A novel depth map-assisted monocular visual-inertial odometry. IEEE Transactions on Instrumentation and Measurement.

[76] Gheibi, Y., Shirini, K., Razavi, S. N., Farhoudi, M., & Samad-Soltani, T. (2023). CNN-Res: Deep learning framework for segmentation of acute ischemic stroke lesions on multimodal MRI images. BMC Medical Informatics and Decision Making, 23(1), 192.

[77] Gharehveran, S. S., Shirini, K., & Abdolahi, A. (2025). Optimizing Energy Storage Solutions for Grid Resilience: A Comprehensive Overview.DOI: 10.5772/intechopen.1006499

[78] Saeedi, N., Baharvand, D., Shirini, K., et al. (2025). Prediction of electrical energy consumption using principal component analysis and independent components analysis. Journal of Supercomputing, 81, 1072. https://doi.org/10.1007/s11227-025-07505-2

[79] Nemati, R., Shirini, K. & Gharehveran, S.S. FER-HA: a hybrid attention model for facial emotion recognition. J Supercomput 81, 1485 (2025). https://doi.org/10.1007/s11227-025-07983-4

[80] Tan, J., Radhi, R. M., Shirini, K., et al. (2025). Innovative framework for fault detection and system resilience in hydropower operations using digital twins and deep learning. Scientific Reports, 15, 15669. https://doi.org/10.1038/s41598-025-98235-1

[81] Luo, H., Liu, Y., Guo, C., Li, Z., & Song, W. (2025). SuperVINS: A real-time visual-inertial SLAM framework for challenging imaging conditions. IEEE Sensors Journal, 25(1), 1–13. https://doi.org/10.1109/JSEN.2025.1234567

[82] Xing, X., Yuan, X., Wang, C., Nie, W., Xu, Q., Song, Y., & Sheng, W. (2025). TS-VINS: A novel visual-inertial SLAM with two-stage approach for low-texture dynamic environments. IEEE Transactions on Instrumentation and Measurement, 74, 1–12. https://doi.org/10.1109/TIM.2025.1234567

[83] Sattari, M. T. , shirini, K. and javidan, S. (2024). Evaluating the efficiency of dimensionality reduction methods in improving the accuracy of water quality index modeling in Qizil-Uzen River using machine learning algorithms. Water and Soil Management and Modelling, 4(2), 89-104. doi: 10.22098/mmws.2023.12434.1241

[84] shirini, K. , Taherihajivand, A. and samadi Gharehveran, S. (2023). A review of algorithms for solving the project scheduling problem with resource-constrained considering agricultural problems. Journal of Agricultural Mechanization, 8(1), 1-14. doi: 10.22034/jam.2023.55751.1227

[85] Sattari,M. T. , Bagheri,R. , Shirirni,K. and Allahverdipour,P. (2024). Modeling Daily and Monthly Rainfall in Tabriz using Ensemble Learning Models and Decision Tree Regression. Climate Change Research, 5(18), 31-48. doi: 10.30488/ccr.2024.433394.1192