

PKS: یک روش انتخاب فریم‌های کلیدی مبتنی بر فتوگرامتری برای سیستم‌های بینایی-اینرسی ساخته‌شده بر پایه ORB-SLAM3

آرش عظیمی فرد^۱

علی حسینی نوه احمدآبادیان^۲

تاریخ دریافت مقاله: ۱۴۰۰/۰۸/۲۷

تاریخ پذیرش مقاله: ۱۴۰۰/۱۰/۲۴

چکیده

به دلیل پیچیدگی‌های پردازش فریم برای تعیین موقعیت و تهیه نقشه در الگوریتم‌های ماشین بینایی و فتوگرامتری، روش‌های انتخاب فریم‌های کلیدی به منظور افزایش کارایی الگوریتم‌ها معرفی شدند که در عین حفظ دقت و استحکام الگوریتم، حجم پردازش‌ها را کاهش می‌دهند. یکی از معروف‌ترین الگوریتم‌های تعیین موقعیت و تهیه نقشه هم‌زمان مبتنی بر تصویر (ویژوال اسلم)، الگوریتم ORB-SLAM3^۳ است. انتخاب فریم کلیدی در این الگوریتم و سایر الگوریتم‌های این حوزه وابسته به حد آستانه‌های ابتکاری است. در این مقاله یک روش هندسی و بر پایه اصول طراحی شبکه تصویربرداری در فتوگرامتری به منظور انتخاب فریم‌های کلیدی در بهبود الگوریتم ORB-SLAM3 پیشنهاد شده است. در این روش، حد آستانه‌های ابتکاری با اصول فتوگرامتری جایگزین شده است که علاوه بر استحکام الگوریتم، کیفیت ابر نقطه حاصل از فریم‌های کلیدی را تضمین می‌کند. در روش پیشنهادی، ابتدا یک حد آستانه انطباقی در مورد مجاز بودن تعداد نقاطی که ناحیه مخروطی خط دید آن‌ها در یک مخروط چهار ناحیه‌ای تشکیل شده بر روی هر نقطه، تغییر کرده است، تصمیم می‌گیرد. سپس با تشکیل یک شبکه ۳ در ۳ در هر فریم و شمارش نقاط مؤثر در هر سلول این شبکه، معیار تعادل مرکز ثقل (ECOG)^۴ در مورد مناسب بودن توزیع نقاط داخل این فریم تصمیم می‌گیرد. از طرف دیگر سنسور اینرسی (IMU)^۵ در صورت مشاهده تغییرات شدید شتاب حرکت، مستقل از دوربین اقدام به اخذ فریم کلیدی می‌کند. به منظور ارزیابی روش پیشنهادی، آزمایش‌های وسیعی روی داده EuRoC^۶ در حالت تک‌دوربین و دو دوربین انجام شده است. ارزیابی‌های کیفی و کمی با مقایسه مسیر ردیابی شده هر الگوریتم با مسیر مرجع، مقایسه ابر نقطه تشکیل شده از فریم‌های کلیدی و مقایسه مقدار خطای مطلق مسیر حرکت (ATE)^۷ انجام شده است. همچنین زمان اجرای هر الگوریتم برای تمامی دنباله تصاویر داده EuRoC ارزیابی شده است. نتایج نشان می‌دهد، الگوریتم پیشنهادی در حالت دو دوربین ۱/۱۸٪ و در حالت تک‌دوربین ۴/۲۰٪ دقت تعیین موقعیت ORB-SLAM3 را بهبود داده و علاوه بر این ابر نقطه متراکمتری تولید کرده است.

واژه‌های کلیدی: اودومتری بینایی، اسلم بینایی، فتوگرامتری برد کوتاه، انتخاب فریم‌های کلیدی، قيود هندسی، حد آستانه انطباقی

۱- دانشجوی کارشناسی ارشد گروه فتوگرامتری و سنجش از دور، دانشکده مهندسی ژئودزی و ژئوماتیک، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران
(نویسنده مسئول) arashazimi 0032@gmail.com

۲- استادیار گروه فتوگرامتری و سنجش از دور، دانشکده مهندسی ژئودزی و ژئوماتیک، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران hosseininaveh@kntu.ac.ir

3- Oriented FAST and Rotated BRIEF-Simultaneous Localization And Mapping

4- Equilibrium of Center Of Gravity

5- Inertial Measurement Unit

6- European Robotic Challenge

7- Absolute Trajectory Error

۱- مقدمه

نقشه و ابر نقطه نیز استفاده شوند که انتخاب بهتر فریم‌های کلیدی می‌تواند منجر به تولید ابر نقطه متراکم‌تر و با کیفیت بالاتر شود.

همان‌طور که در بخش بعدی ملاحظه می‌شود، صرف‌نظر از روش‌های مبتنی بر شبکه عصبی عمیق که بستگی به دقت روش انتخاب فریم کلیدی در داده‌های آموزشی دارد و یا روش‌های انتخاب فریم کلیدی در کاربردهای پردازش ویدئو و SFM^۱ که در ابتدا به همه فریم‌ها نیاز دارند و نمی‌توانند به‌صورت آنی استفاده شوند، استراتژی‌های اتخاذشده در کاربردهای بینایی ماشین، مانند انتخاب فریم‌های کلیدی در فواصل مکانی یا زمانی ثابت، از نظر محاسباتی ساده هستند، اما انعطاف‌پذیری پایینی دارند. روش‌های دیگر مانند هم‌پوشانی تصویر، اختلاف پارالاکس متوسط، اختلاف آنترپی و غیره، عملکرد بهتری دارند، اما از نظر محاسباتی پیچیده هستند. علاوه بر این، همه روش‌های انتخاب فریم کلیدی ارائه‌شده در زمینه بینایی ماشین، تصمیم‌گیری برای اخذ فریم کلیدی را بر اساس آستانه‌های ابتکاری انجام می‌دهند و جنبه هندسی مسئله را نادیده می‌گیرند. الگوریتم ORB-SLAM3 بهترین عملکرد را در بین الگوریتم‌های ویژوال اسلم داشته است؛ اما روش انتخاب فریم کلیدی در این الگوریتم، همچون سایر روش‌ها، نیاز به مشخص کردن حد آستانه‌هایی برای معیارهای مختلف دارد که به‌صورت سعی و خطا محاسبه شده‌اند.

هدف در این مقاله، توسعه یک روش انتخاب فریم کلیدی بهبودیافته و بر پایه فتوگرامتری، براساس ORB-SLAM3، با نام Photogrammetric Keyframe Selection (PKS)، است که علاوه بر حفظ تعادل بین پیچیدگی و انعطاف‌پذیری محاسباتی، آستانه‌های ابتکاری را با اصول هندسی جایگزین می‌کند. با توجه به لحاظ کردن قیود هندسی برگرفته از اصول طراحی شبکه تصویربرداری در تولید ابر نقطه سه‌بعدی، روش پیشنهادی ابر نقطه‌ای کامل‌تر و با دقت بالاتری را تولید خواهد کرد.

مدت زیادی است که ناوبری خودکار ربات در کاربردهای مختلف از جمله کاربردهای نظامی، اکتشافات فضایی، امداد و نجات در بحران‌ها، نقشه‌برداری خودکار داخل ساختمان، اتومبیل‌های خود ران و غیره مورد توجه محققان قرار گرفته است. سنسورهای زیادی در این حوزه استفاده می‌شوند که از آن‌ها می‌توان به دوربین، IMU، اودومتر و لیزر اسکنر اشاره کرد که از بین آن‌ها، سنسور دوربین به دلیل داشتن مزایایی چون وزن کم، ارزان بودن، مصرف انرژی کم و اندازه کوچک‌تر نسبت به سایر سنسورها، امروزه بیشتر مورد توجه قرار گرفته است. تاکنون الگوریتم‌های زیادی در این حوزه توسعه داده شده‌اند که از آن‌ها می‌توان به VINS^۱ (Forster, Pizzoli, & Scaramuzza, 2014)، SVO^۲ (Engel, Koltun, & DSO^۳ (Qin, Li, & Shen, 2018)، mono^۴ (Leutenegger, Lynen, Bosse, Cremers, 2017)، OKVIS^۴ (Siegwart, & Furgale, 2015)، kimera (Rosinol, Abate, Bloesch, Burri, 2020)، ROVIO^۵ (Chang, & Carlone, 2020)، Omari, Hutter, & Siegwart, 2017) و در نهایت سرآمد آن‌ها یعنی ORB-SLAM3 (Campos, Elvira, Rodríguez, 2021) اشاره کرد که توانسته است تمامی الگوریتم‌های قبل از خود را از نظر دقت پشت سر بگذارد و پیشگام اصلی این حوزه به‌شمار می‌رود. چالش اصلی در الگوریتم‌های بینایی ماشین، حجم زیاد فریم‌ها برای پردازش می‌باشد که مانع اجرا شدن آنی این الگوریتم‌ها می‌شود. از این رو ایده پردازش فریم‌های کلیدی به‌جای کل فریم‌ها مطرح شد که به این ترتیب پیچیدگی محاسبات کاهش یافته و درعین حال دقت و استحکام الگوریتم حفظ می‌شود؛ بنابراین انتخاب مناسب فریم‌های کلیدی می‌تواند باعث بهبود دقت الگوریتم‌های بینایی ماشین شود. علاوه بر این، الگوریتم‌های SLAM^۶ می‌توانند در کاربردهای تولید

1- Fast Semi-Direct Monocular Visual Odometry

2- Visual Inertial Navigation Sistem mono

3- Direct Sparse Odometry

4- Open Keyframe-based Visual-Inertial SLAM

5- Robust Visual Inertial Odometry

6- Simultaneous Localization and Mapping

پارامترهای دیداری یا هندسی فریم‌ها، تصمیم به انتخاب فریم کلیدی می‌گیرد. این روش نتایج خوبی مخصوصاً از نظر سرعت داشته است؛ اما دقت این روش‌ها کاملاً وابسته به دقت روش انتخاب فریم‌های کلیدی در داده آموزشی می‌باشد. Rukiye savran و همکاران در (Savran, Kızıltepe, Gan, & Escobar, 2021) یک روش انتخاب فریم‌های کلیدی با ترکیب شبکه عصبی پیچشی^۱ و شبکه عصبی بازگشتی^۲، در کاربرد طبقه‌بندی ویدئو ارائه کردند. آن‌ها ابتدا ناحیه حاوی اطلاعات تصویر را شناسایی کرده و سپس با استفاده از شبکه عصبی ترکیبی و بر اساس شباهت بین این نواحی در فریم‌های متوالی، تصمیم به اخذ فریم کلیدی می‌گیرند. نتایج آن‌ها حاکی از بهبود دقت قابل‌توجهی در کاربرد بخش‌بندی ویدئو می‌باشد. همچنین changhai LV و همکاران در (Lv, Li, & Tian, 2021) یک روش انتخاب فریم‌های کلیدی در کاربرد بخش‌بندی ویدئوهای ورزشی، بر مبنای یادگیری عمیق ارائه کردند. آن‌ها با استفاده از شبکه FCN^۳ موقعیت نواحی فعال ورزشی تصویر را شناسایی کرده و سپس با استفاده از شبکه عصبی پیچشی CNN احتمال وقوع موقعیت هر فریم را تخمین زدند (موقعیت احتمالی هر فریم)؛ سپس با استفاده از اختلاف بین موقعیت احتمالی فریم‌های همسایه، فریم‌های کلیدی را انتخاب کردند. نتایج تحقیقات آن‌ها نشان می‌دهد که این روش می‌تواند براساس وضعیت بدن ورزشکاران، با دقت ۹۸٪ ویدئو را بخش‌بندی کند.

الگوریتم‌های یادگیری عمیق وابستگی شدیدی به روش انتخاب فریم‌های کلیدی در داده آموزشی دارند و نتایج نشان می‌دهد که اگر به‌تنهایی استفاده شوند، در بهترین حالت نتایجی مشابه با روش انتخاب فریم کلیدی در داده آموزشی ارائه می‌کنند؛ با این حال این الگوریتم‌ها سرعت پردازش بالایی دارند و در صورت ترکیب کردن آن‌ها با الگوریتم‌های سنتی، امکان دست یافتن به دقت‌های بالاتر،

در ادامه این مقاله، بعد از بیان مرور کارهای گذشته، روش پیشنهادی توضیح داده شده و سپس نحوه پیاده‌سازی و ارزیابی روش بیان شده است. در انتها هم نتایج و پیشنهادهایی برای کارهای آتی در بخش نتیجه‌گیری ارائه شده است.

۲- مروری بر الگوریتم‌های انتخاب فریم کلیدی

با توجه به اهمیت انتخاب فریم‌های کلیدی در الگوریتم‌های SLAM و Odometry، امروزه تحقیقات زیادی در این زمینه انجام گرفته است که در ادامه ارائه شده است.

۲-۱- انتخاب فریم‌های کلیدی با استفاده از شبکه

عصبی عمیق

در این دسته از روش‌ها از یک سری پارامترهای مناسب در انتخاب فریم کلیدی استفاده می‌شود. این پارامترها به دو دسته پارامترهای دیداری و پارامترهای هندسی تقسیم می‌شوند. از جمله پارامترهای دیداری، شار نوری صحنه و بافت صحنه هستند و پارامترهای عمق صحنه، موقعیت فریم‌ها، فاصله فریم‌ها از یکدیگر، تغییرات جهت‌گیری فریم‌ها و موقعیت و تعداد نقاط و لندمارک‌های داخل صحنه جز پارامترهای هندسی محسوب می‌شوند. این پارامترها در قسمت ورودی شبکه استفاده می‌شوند و پس از آموزش شبکه توسط دنباله تصاویری که فریم‌های کلیدی آن‌ها توسط سایر روش‌های سنتی از قبل انتخاب شده است، این شبکه در مورد کلیدی بودن یا نبودن یک فریم در یک دنباله تصاویر جدید تصمیم‌گیری می‌کند.

به‌عنوان کارهایی که در این زمینه انجام شده است می‌توان به کار Lu Sheng و همکاران (Sheng, Xu, Ouyang, & Wang, 2019) اشاره کرد. آن‌ها فرآیند انتخاب فریم کلیدی را به‌طور هم‌زمان با فرآیند اودومتري، در یک شبکه عمیق یکتا انجام می‌دهند. آن‌ها در ورودی شبکه عمیق خود هم پارامترهای هندسی و هم پارامترهای دیداری را مدنظر قرار می‌دهند و این شبکه علاوه بر تعیین موقعیت فریم‌ها، به‌طور هم‌زمان، با مشاهده تغییرات قابل توجه در یکی از

1- Convolution Neural Network

2- Recurrent Neural Network

3- Fully Convolutional Network

در عین حفظ سرعت پردازش آنی میسر می‌باشد.

در بقیه فریم‌های داخل شات ساخته می‌شود. به این فریم کلیدی ساخته‌شده فریم با بیشترین وقوع در بازه زمانی (TMOF)^۱ گفته می‌شود. این روش انتخاب فریم کلیدی قدرت و کارایی بسیار بالایی دارد و فریم‌های کلیدی انتخابی نمایانگر خوبی از کل شات هستند.

۲-۲- انتخاب فریم‌های کلیدی در کاربردهای پردازش ویدئو

این دسته از روش‌های سنتی در کاربردهای غیر آنی و پس‌پردازشی مورد استفاده قرار می‌گیرند و در صورتی قابل استفاده هستند که از ابتدای امر تمامی فریم‌ها را در اختیار داشته باشند که خود به دو زیر دسته تقسیم می‌شوند.

۲-۳- استخراج عکس کلیدی در کاربردهای فتوگرامتری و بازسازی ساختار از حرکت (SFM)^۲ و تولید ابر نقطه متراکم

این دسته از روش‌ها نیز بیشتر در کاربردهای غیر آنی و پس‌پردازشی مانند مدل‌سازی سه‌بعدی اشیاء استفاده می‌شوند و معمولاً مسئله انتخاب تصاویر کلیدی با دید هندسی و فیزیکی مورد بررسی قرار می‌گیرد.

علی حسینی‌نوه و همکاران (Ahmadabadian, Robson, 2012; Hosseininaveh et al., 2013; Boehm, & Shortis, 2013) روش طراحی شبکه تصویر (IND)^۳ را با هدف استخراج زیرمجموعه‌ای بهینه از تصاویر (برای بازسازی سه‌بعدی) از بین مجموعه تصاویری که به صورت منظم و سیستماتیک از یک شیء گرفته شده‌اند ارائه کردند. در این روش، ابتدا با استفاده از تمامی تصاویر یک ابر نقطه پراکنده ایجاد می‌شود؛ سپس برای هر نقطه از نقاط موجود در نقشه یک مخروط که محور اصلی آن نرمال بر سطح است تشکیل می‌شود. این مخروط به چهار ناحیه تقسیم می‌شود که از مقطع مخروط به صورت چهار دایره با مرکز یکسان و شعاع‌های متفاوت دیده می‌شوند. با توجه به این نواحی و زاویه بین محور مخروط و بردار دید هر نقطه در هر تصویر، یک ماتریس دید تعریف و مقداردهی می‌شود. سطرهای ماتریس دید، چهار برابر نقاط موجود در ابر نقطه می‌باشد (به ازای هر ناحیه مخروط برای هر نقطه یک سطر) و هریک از ستون‌ها معرف موقعیت دوربین در لحظه اخذ هر یک از تصاویر است. با در نظر گرفتن دیده‌شدن هر نقطه در یک ناحیه مخروط در یک تصویر، عدد متناظر با آن ناحیه و تصویر

۲-۲-۱- روش‌های بر مبنای خوشه‌بندی

این دسته از روش‌ها تمامی فریم‌ها را براساس یک سری ویژگی‌ها، در خوشه‌های مختلف قرار می‌دهند و سپس از هر خوشه یک فریم (معمولاً فریمی که به مرکز خوشه نزدیک‌تر است) را به عنوان فریم کلیدی انتخاب می‌کنند.

اولین بار سال ۱۹۹۸ Zhuang و همکاران (Zhuang, Rui, 1998) روش‌های استخراج فریم کلیدی بر مبنای خوشه‌بندی را ارائه کردند. آن‌ها فریم‌ها را بر اساس ویژگی‌هایی مثل رنگ، بافت و شکل خوشه‌بندی کردند و از هر خوشه فریم نزدیک‌تر به مرکز خوشه را به عنوان فریم کلیدی انتخاب کردند. این روش در ویدئوهای معمولی مثل مسابقات ورزشی، اخبار و سریال‌های تلویزیونی نتیجه خوبی داشته است؛ اما در شرایطی که تغییرات نوری بین فریم‌ها بسیار زیاد باشد، این روش نتایج خوبی نداشته است.

۲-۲-۲- روش‌های بر مبنای شات

در این روش‌ها ابتدا با یک سری روش‌ها، ویدئو را به چندین شات تقسیم می‌کنند و سپس از هر شات، یک یا چند فریم را به عنوان فریم کلیدی انتخاب می‌کنند. این فریم می‌تواند فریم ابتدایی یا انتهایی و یا فریم وسط هر شات باشد.

Kin-Wai و همکاران (Sze, Lam, & Qiu, 2005) یک روش برای انتخاب فریم‌های کلیدی داخل هر شات، براساس اطلاعات آماری ارائه کردند. به این صورت که هر پیکسل از فریم کلیدی براساس احتمال وقوع پیکسل‌های متناظرش

1- Temporally Maximum Occurrence Frame

2- Structure From Motion

3- Image Network Designer

۲-۴- استخراج فریم‌های کلیدی در کاربردهای ماشین بینایی و الگوریتم‌های VO/VSLAM

برخلاف دسته‌های قبل، این دسته از روش‌ها بیشتر در کاربردهای آنی مورد استفاده قرار می‌گیرند که با توجه به ماهیت الگوریتم‌های ویژوال اسلم مستقیم و غیرمستقیم، به دودسته بر مبنای ظاهر^۲ و بر مبنای حرکت^۳ تقسیم می‌شوند:

۲-۴-۱- دسته بر مبنای ظاهر

این دسته از روش‌ها بیشتر در الگوریتم‌های اسلم و اودومتری مستقیم (که از اطلاعات درجه خاکستری تصویر استفاده می‌کنند) مورد استفاده قرار می‌گیرد و اطلاعاتی مانند شار نوری، درجه خاکستری پیکسل‌ها، بافت و شکل را به کار می‌گیرند.

الگوریتم اودومتری مستقیم پراکنده (DSO)^۴، به عنوان سرآمد الگوریتم‌های ویژوال اسلم مستقیم، در سال ۲۰۱۷، توسط Engle و همکاران (Engel et al., 2017) توسعه داده شد. این الگوریتم در ابتدا تعداد زیادی فریم را به عنوان فریم کلیدی انتخاب می‌کند و سپس به سرعت تعدادی از آن‌ها را کنار می‌گذارد. در کار آن‌ها سه معیار برای انتخاب فریم کلیدی در نظر گرفته شده است: ۱) اگر زاویه دید دوربین بیش از یک مقدار مشخص تغییر کند، فریم کلیدی انتخاب می‌شود که این معیار را با میانگین مربعات شار نوری اندازه‌گیری می‌کنند. ۲) اگر جابه‌جا شدن دوربین باعث شود که دوربین موقعیت خود را از دست بدهد باید فریم کلیدی انتخاب شود. برای اندازه‌گیری آن، میانگین مربعات شار نوری را بدون دوران در نظر می‌گیرند. ۳) اگر مدت زمان نوردهی دوربین بیش از مقدار مشخصی تغییر کند، باید فریم کلیدی جدید در نظر گرفته شود. برای محاسبه این فاکتور روشنایی نسبی بین دو فریم محاسبه می‌شود. در نهایت این سه معیار به صورت وزن‌دار با یکدیگر جمع شده و اگر این مقدار از یک حد آستانه تجاوز کند،

مدنظر، در ماتریس دید مقدار یک را خواهد گرفت. این محاسبه برای تمامی نقاط و تمامی تصاویر انجام می‌شود. دوربینی که بیشترین تعداد ناحیه از نقاط را پر کند ستونی از ماتریس است که مجموع اعداد آن بیشترین مقدار است. این تصویر به عنوان بهترین تصویر انتخاب شده و ستون مربوط به آن و تمامی سطریهایی که در این ستون مقدار یک داشته‌اند حذف می‌شوند. این فرآیند با موقعیت دوربین بعدی و نواحی باقی‌مانده از نقاط ادامه پیدا می‌کند و تصاویر بعدی شناسایی می‌شوند.

بعد از انتخاب سه تصویر، فرآیند تقاطع فضایی بین نقاط و تصاویر انتخاب شده انجام می‌شود و دقت تعیین موقعیت نقاط محاسبه می‌شود. این کار تا زمانی انجام می‌شود که هیچ ناحیه پر نشده‌ای از نقاط وجود نداشته باشد و دقت به دست آمده با آنچه توسط کاربر به عنوان دقت مورد انتظار در نظر گرفته شده بود رسیده باشد. به این ترتیب بهترین تصاویر انتخاب شده و بقیه کنار گذاشته می‌شوند. نتایج نشان می‌دهد که این روش نسبت به روش‌های دیگر، همچون CMVS^۱، نتایج بهتری در نایقینی مختصات ابر نقطه و همچنین در مدل تولید شده نهایی داشته است. روش گفته شده تنها بر روی مجموعه تصاویر اعمال شده است و هنوز در کاربردهای پردازش فریم و آنی مورد بررسی قرار نگرفته است.

علی حسینی‌نوه و همکاران به تازگی در (Hosseininaveh & Remondino, 2021) روش IND را به چهار شکل متفاوت برای تولید یک ابر نقطه متراکم از یک داده شبیه‌سازی و همچنین داده واقعی به کار گرفتند و نتایج تحقیقات آن‌ها نشان می‌دهد که روش به کار گرفته شده توانسته است بهترین ابر نقطه را از نظر دقت و تراکم در بین تمامی روش‌های موجود در این زمینه تولید کند. یکی از نتایج مهم در آن تحقیق، مشخص شدن بهترین زاویه برای نواحی مخروط چهار ناحیه‌ای بود که بعد از آزمایش‌های متعدد، مقدار ۱۰ درجه برای این زاویه محاسبه شده بود.

2- Appearance-based

3- Motion-based

4- Direct Sparse Odometry

1- Clustering Multi-Views Stereo

فریم کلیدی قرار می‌دهند و با انتخاب یک حد آستانه مناسب برای میزان هم‌پوشانی فریم کنونی و آخرین فریم کلیدی، تصمیم بر انتخاب یا عدم انتخاب فریم کلیدی جدید می‌گیرند. به‌طور مثال در (Leutenegger et al., 2015) OKVIS وقتی که تعداد نقاط متناظر در محدوده هم‌پوشانی فریم کنونی و فریم کلیدی قبلی کمتر از ۵۰٪-۶۰٪ کل نقاط ردیابی شده در آن فریم باشد، فریم کلیدی جدید انتخاب می‌شود. در (Tan, Liu, Dong, Zhang, & Bao, 2013) RD-SLAM برای هر فریم جدید سه شرط بررسی می‌شود: (۱) موقعیت دوربین با موفقیت تخمین زده شود، (۲) تعداد نقاط مشترک با فریم‌های کلیدی موجود از یک حد آستانه کمتر شود و (۳) تعداد نقاط جدید تولیدشده توسط فریم کنونی از یک حد آستانه بیشتر باشد. این دسته از روش‌ها، نسبت به دسته قبل، انعطاف‌پذیری و قدرت بیشتری داشته و در شرایط چالش‌برانگیز کارایی بهتری خواهند داشت.

۲-۴-۳- انتخاب فریم کلیدی براساس پارالاکس

در این دسته از روش‌ها، اختلاف پارالاکس فریم کنونی و آخرین فریم کلیدی مورد بررسی قرار می‌گیرد و با گذشتن این مقدار تغییرات از یک حد آستانه خاص، فریم کلیدی اخذ می‌شود. به‌طور مثال در (Qin et al., 2018) VINS-mono دو معیار برای انتخاب فریم کلیدی در نظر گرفته شده است: (۱) میانگین پارالاکس: اگر میانگین پارالاکس نقاط ردیابی شده از یک حد آستانه تجاوز کند، فریم کنونی به‌عنوان فریم کلیدی انتخاب می‌شود، (۲) کیفیت ردیابی: اگر تعداد نقاط ردیابی شده از یک حد آستانه کمتر باشد، فریم کنونی به‌عنوان فریم کلیدی انتخاب می‌شود.

۲-۴-۴- سایر روش‌ها

به‌غیر از روش‌های گفته‌شده در بالا، الگوریتم‌های دیگری وجود دارند که از هیچ‌یک از این روش‌ها و یا ترکیبی از آن‌ها در انتخاب فریم‌های کلیدی استفاده نمی‌کنند. در ادامه به چند مورد از آن‌ها اشاره شده است. به‌طور مثال Kerl و

یک فریم کلیدی جدید در نظر گرفته می‌شود. پس از انجام آزمایش‌ها بر روی مجموعه داده‌های مختلف، این روش انتخاب فریم‌های کلیدی، قدرت و نتایج بسیار مناسبی در شرایط نوری بد ارائه کرده است؛ اما تنظیم مناسب حد آستانه‌های این روش، چالش اصلی روش محسوب می‌شود.

۲-۴-۲- دسته بر مبنای حرکت

این دسته از روش‌ها در الگوریتم‌های اسلم و اودومتري غیرمستقیم (که از الگوریتم‌های استخراج گوشه استفاده می‌کنند) مورد استفاده قرار می‌گیرد. این روش‌ها به چند زیرمجموعه تقسیم می‌شوند:

۲-۴-۲-۱- انتخاب فریم کلیدی در فواصل مکانی یا

زمانی مشخص

در این حالت فریم‌های کلیدی بدون در نظر گرفتن هیچ معیار خاصی و صرفاً با گذشت زمان و یا فاصله خاصی قرار داده می‌شوند. به‌طور مثال در ردیابی و تهیه نقشه موازی (PTAM)^(۱) (Klein & Murray, 2007)، باید کیفیت ردیابی مناسب باشد تا فریم کلیدی در نظر گرفته شود و به این منظور، فاصله از فریم کلیدی قبلی نباید از یک حد آستانه تجاوز کند. SVO^(۲) (Forster et al., 2014) فقط از انتقال دوربین به‌عنوان معیاری برای انتخاب فریم کلیدی استفاده می‌کند و دوران را مدنظر قرار نمی‌دهد. این امر باعث می‌شود که فقط در حالت دید از بالا (مثل پهپادها) کارایی مناسبی داشته باشد و در حالت دید مستقیم عملکرد ضعیفی خواهد داشت. LSD-SLAM^(۳) (Engel, Schöps, & Cremers, 2014) فقط هنگامی یک فریم کلیدی انتخاب می‌کند که فاصله فریم کنونی از آخرین فریم کلیدی، از یک حد آستانه بیشتر شود.

۲-۴-۲-۲- انتخاب فریم کلیدی بر اساس هم‌پوشانی تصویر

این روش‌ها میزان هم‌پوشانی تصاویر را مبنای انتخاب

1- Parallel Tracking and Mapping

2- Fast Semi-direct monocular Visual Odometry

3- Large-Scale Direct Monocular SLAM

فرآیند بسته شدن حلقه^۲ را اجرا کرده و خطای بست را در کل فریم‌ها و نقاط نقشه سرشکن می‌کند. رشته اول در ORB-SLAM، چهار شرط با حد آستانه‌های ثابت زیر را برای اتخاذ فریم کلیدی در نظر می‌گیرد: (۱) بیشتر از ۲۰ فریم از آخرین موقعیت‌یابی مجدد^۳، یا فریم کلیدی گذشته باشد. (۲) هسته پردازشی مربوط به تهیه نقشه فعال نباشد. (۳) فریم کنونی حداقل ۵۰ نقطه ردیابی کند. (۴) فریم کنونی کمتر از ۹۰٪ نقاط موجود در آخرین فریم کلیدی را ردیابی کرده باشد (Campos et al., 2021). این حد آستانه‌ها، صرف‌نظر از داده و سخت‌افزار مورد استفاده، کاملاً ثابت بوده و قابلیت تنظیم‌شدن توسط کاربر را ندارند. این روش اتخاذ فریم کلیدی بسیار قدرتمند بوده و تقریباً ترکیبی از تمامی شروط مربوط به روش‌های قبل می‌باشد. علی‌رغم دقت بسیار خوبی که این الگوریتم نسبت به بقیه الگوریتم‌های این حوزه داشته است؛ اما همچنان از حد آستانه‌های ابتکاری با روش آزمون‌وخطا استفاده کرده است که انعطاف‌پذیری پایینی دارند و ممکن است با تغییر محیط و نحوه حرکت ربات، دقت تعیین موقعیت کاهش پیدا کند.

۳- روش پیشنهادی

روش پیشنهادشده در این مقاله بر پایه الگوریتم ORB-SLAM3 توسعه داده شده است و براساس اصول هندسی و فتوگرامتری، سعی می‌کند که تا حد امکان حد آستانه‌های ابتکاری و ابداعی انتخاب فریم کلیدی در الگوریتم ORB-SLAM3 را با حد آستانه‌های انطباقی جایگزین کند تا انعطاف‌پذیری الگوریتم افزایش پیدا کرده و قدرتمندی خود را در تمامی شرایط و محیط‌ها حفظ کند و همچنین بتواند یک ابر نقطه دقیق و متراکم از فریم‌های کلیدی استخراج‌شده، تولید نماید. علاوه بر این، از آنجا که روش پیشنهادشده، قابلیت استفاده در سیستم‌های باینای - اینرشیا را دارد، سنسور IMU را نیز در فرآیند انتخاب فریم کلیدی دخیل می‌کند. انتخاب فریم کلیدی توسط دو

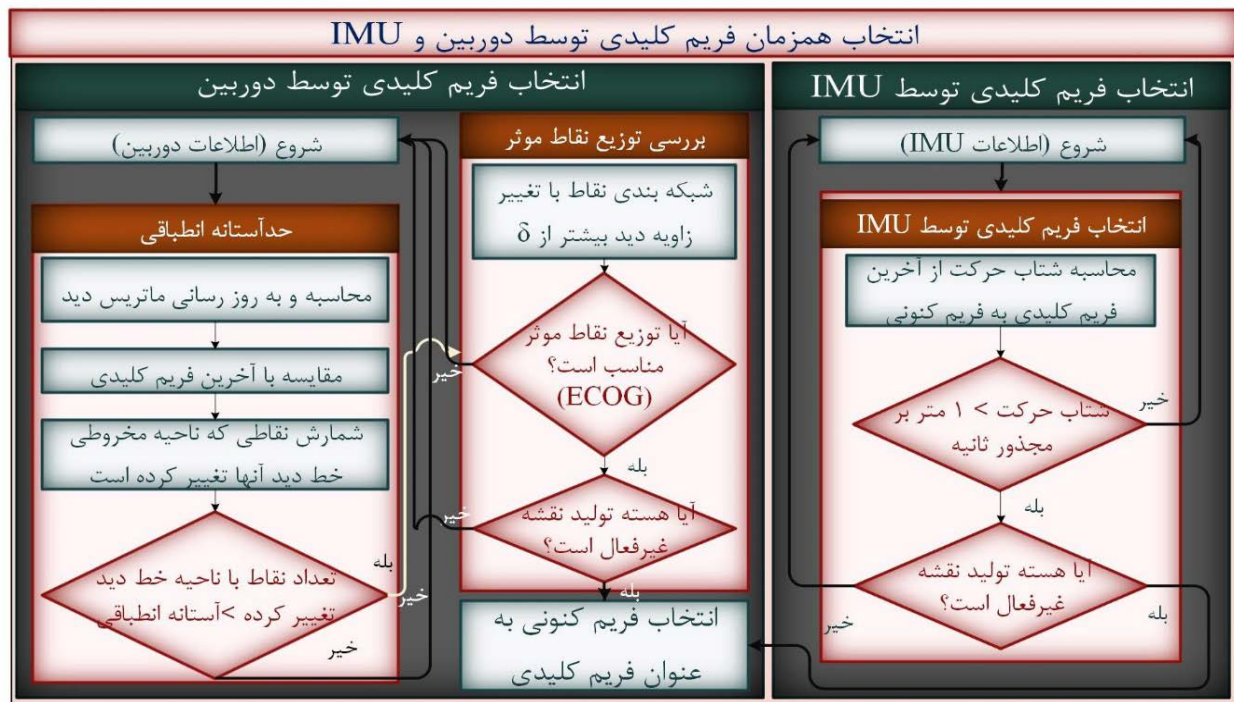
همکاران در (Kerl, Sturm, & Cremers, 2013) یک روش انتخاب فریم کلیدی بر اساس اختلاف آنروپی یک توزیع نرمال چند متغیره ارائه کردند که نتایج بسیار خوبی در محیط‌های بدون بافت داشته است؛ اما از نظر محاسباتی کارآمد نیست و آنی بودن فرآیند اودومتری را به خطر می‌اندازد. Xiaohu Lin و همکاران (Lin, Wang, Guo, & Zhang, 2019)، تصمیم‌گیری اتخاذ فریم کلیدی را بر اساس تغییرات المان‌های زوایای دوران دو فریم (که از تجزیه ماتریس ضروری به دست آمده است) نسبت به یکدیگر انجام می‌دهند. اگر تغییرات زوایا از یک حد آستانه بیشتر بود فریم‌های کلیدی با سرعت بالایی ایجاد می‌شوند. نتایج نشان می‌دهد که این روش با کاهش ۴۰٪ تا ۶۰٪ فریم‌های مازاد، سرعت الگوریتم را افزایش داده و درعین حال باعث کاهش دقت نمی‌شود. این روش برخلاف سایر روش‌ها، تا حد کمی توانسته است حد آستانه‌های ابتکاری را از فرآیند انتخاب فریم کلیدی حذف کند و یک روش تقریباً هندسی مبتنی بر زاویه دوران دوربین برای این مسئله ارائه دهد.

درنهایت، الگوریتم ORB-SLAM (Campos et al., 2021; Mur-Artal, Montiel, & Tardos, 2015; Mur-Artal & Tardós, 2017) مناسب‌ترین استراتژی انتخاب فریم کلیدی را اتخاذ نموده و سعی می‌کند که فریم‌های کلیدی در ابتدا با سرعت هرچه بیشتر وارد شوند و از طرف دیگر، برای جلوگیری از افزایش بی‌رویه حجم محاسبات، فریم‌های کلیدی اضافی را حذف می‌کند. این الگوریتم صفر تا صد فرآیند SLAM را به‌طور کامل در سه رشته^۱ موازی پیش می‌برد: (۱) رشته اول وظیفه ردیابی موقعیت هر فریم جدید ورودی را به روش بازسازی ساختار از حرکت بر عهده دارد. تصمیم‌گیری بر اخذ فریم کلیدی نیز توسط این رشته انجام می‌شود. (۲) رشته دوم با اخذ هر فریم کلیدی جدید، فرآیند تولید نقشه و محاسبه مختصات سه‌بعدی نقاط ردیابی شده در هر فریم کلیدی را پیش می‌برد. (۳) رشته سوم نیز به‌صورت هم‌زمان با مشاهده یک منطقه تکراری در مسیر حرکت،

2- Loop-Closing

3- Re-localization

1- Thread



نگاره ۱: روندنمای روش انتخاب فریم کلیدی پیشنهادشده (PKS)

نقطه آن‌ها از آخرین فریم کلیدی به فریم کنونی تغییر کرده است. در این قسمت یک حد آستانه انطباقی طراحی شده است که حد مجاز این تغییرات را کنترل می‌کند. شرط دوم نحوه توزیع و پخش نقاطی که ناحیه مخروطی خط واصل دوربین به نقطه آن‌ها بیشتر از ۳۰ درجه تغییر کرده است را، در داخل فریم بررسی می‌کند. معیار تعادل مرکز ثقل (ECOG)، توزیع و تقارن این نقاط را کنترل می‌کند. این معیار نیز مشابه با معیار قبل، کاملاً انطباقی طراحی شده و مطابق با تعداد نقاط متناظر و نقاط با ناحیه تغییر کرده داخل هر فریم، حد آستانه تقارن نقاط را تنظیم می‌کند. در ادامه این دو شرط به تفصیل بررسی شده‌اند.

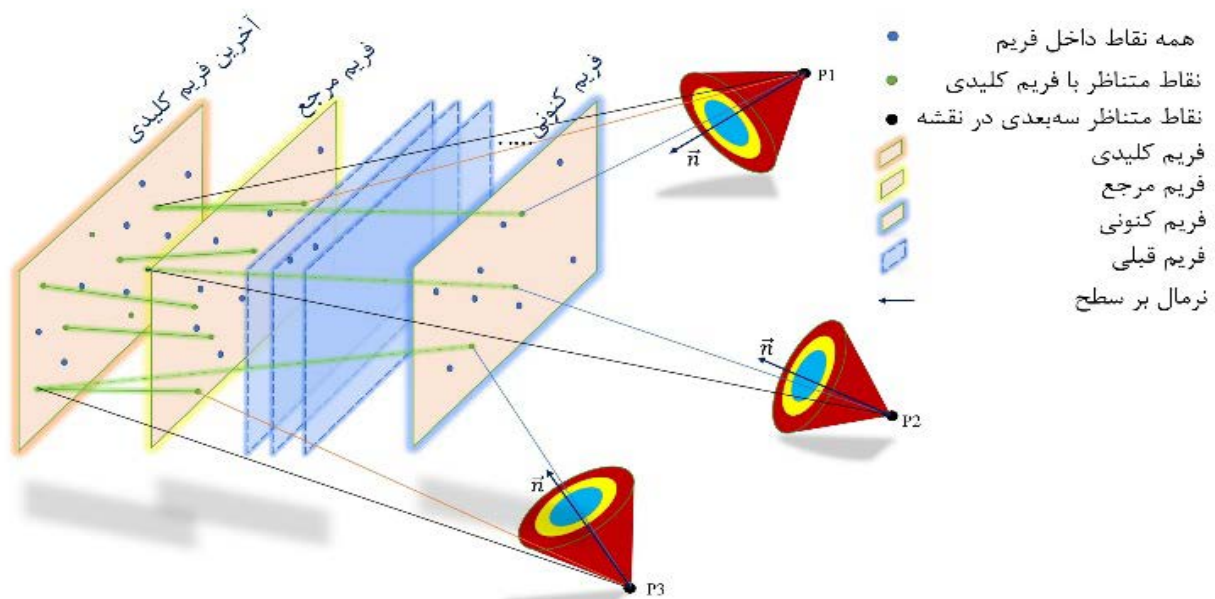
۳-۱-۱- محاسبه حد آستانه انطباقی

اساس فرآیند محاسبه حد آستانه انطباقی، بر مبنای شمارش تعداد نقاط در داخل هر فریم می‌باشد. ابتدا برای هر یک از نقاط متناظر بین فریم کنونی و آخرین فریم کلیدی، یک مخروط فضایی، مشابه با آنچه در روش IND گفته شده

سنسور دوربین و IMU به صورت کاملاً مستقل انجام شده و به صورت موازی با یکدیگر کار می‌کنند. سنسور دوربین با استفاده از روش‌ها و اصول هندسی و فتوگرامتری الهام گرفته از روش IND (Hosseiniaveh et al., 2012) تصمیم بر اخذ فریم کلیدی می‌گیرد؛ درحالی‌که به صورت هم‌زمان، سنسور IMU در صورت مشاهده تغییرات شدید در شتاب حرکت، اقدام به اخذ فریم کلیدی می‌کند. روندنمای روش انتخاب فریم کلیدی پیشنهادشده در نگاره ۱ ارائه شده است.

۳-۱- انتخاب فریم کلیدی توسط دوربین

همان‌طور که گفته شد سنسور دوربین و IMU به صورت هم‌زمان و مستقل از یکدیگر در مورد اخذ فریم کلیدی تصمیم می‌گیرند. در فرآیند انتخاب فریم‌های کلیدی توسط دوربین، دو شرط کاملاً هندسی بررسی می‌شوند و زمانی یک فریم کلیدی جدید اخذ می‌شود که این دو شرط به صورت هم‌زمان برقرار باشند. شرط اول، بررسی تعداد نقاطی است که ناحیه مخروطی خط واصل دوربین به



نگاره ۲: هندسه انتخاب فریم‌های کلیدی در روش PKS

و نزدیک‌ترین فریم به آخرین فریم کلیدی می‌باشد (فریم بلافاصله بعد از آخرین فریم کلیدی) به‌عنوان فریم رفرنس انتخاب شده است. رابطه پیشنهادشده برای حد آستانه انطباقی به‌صورت رابطه (۱) است:

$$Th_{adaptive} = Th_{initial}(1 + \theta - \beta - \alpha) \quad \text{رابطه (۱)}$$

جایی که $Th_{initial}$ یک مقدار اولیه برای حد آستانه انطباقی می‌باشد و α ، β ، θ ضرایب مورد نیاز برای انطباقی کردن حد آستانه اولیه می‌باشند که در ادامه نحوه شکل‌گیری این رابطه توضیح داده شده است.

اصول انتخاب فریم کلیدی این روش براساس تعداد نقاطی که ناحیه خط دید آن‌ها تغییر کرده، می‌باشد؛ چراکه افزایش تعداد نقاطی که ناحیه خط دید آن‌ها تغییر کرده است، به معنی تغییر وضعیت معنادار فریم کنونی و افزایش نیاز به فریم کلیدی جدید است. بنابراین تعیین حد آستانه این تغییرات، مهم‌ترین امر در اخذ فریم کلیدی می‌باشد. این حد آستانه، ابتدا با فرض یکسان بودن وضعیت فریم کنونی با فریم مرجع و با توجه به تعداد نقاط متناظر و نقاط ناحیه خط دید تغییر کرده، یک حد آستانه اولیه برای تعداد

است، تشکیل می‌شود. این مخروط به چهار ناحیه با زوایای تفکیک ۱۰ درجه تقسیم می‌شود و ناحیه خطوط واصل دوربین به این نقاط مشخص می‌شود. سپس این نواحی، با نواحی متناظر در آخرین فریم کلیدی مقایسه شده و نقاطی که ناحیه خط واصل دوربین به نقطه آن‌ها از آخرین فریم کلیدی به فریم کنونی تغییر کرده است، شمارش می‌شوند. سرانجام حد آستانه انطباقی که در ادامه روابط آن ارائه شده است، در مورد مجاز بودن تعداد این تغییرات تصمیم‌گیری می‌کند. زاویه ۱۰ درجه نواحی مخروطی براساس بیشینه زاویه انحرافی که اکثر الگوریتم‌های شناسایی نقاط کلیدی می‌توانند نقاط متناظر را در تصاویر پیدا کنند، مشخص شده است. این زاویه در روش IND (Hosseinineh et al., 2012) با آزمایش‌های متعدد به‌دست آمده است.

در نگاره ۲، هندسه مسئله انتخاب فریم‌های کلیدی در روش پیشنهادشده، ارائه شده است. در تعریف حد آستانه انطباقی، پارامترهای تعداد کل نقاط داخل فریم، تعداد نقاطی که با نقاط آخرین فریم کلیدی مطابقت داده شده‌اند و تعداد نقاطی که ناحیه مخروطی آن‌ها از آخرین فریم کلیدی تغییر کرده است، استفاده شده‌اند. همچنین یک فریم که شبیه‌ترین

یک نسبت تناسب به دست آورد که در رابطه (۳) مشاهده می شود:

$$\frac{D_r}{M_r} = \frac{D_c}{M_c} \Rightarrow D_c = \frac{D_r M_c}{M_r} \quad \text{رابطه (۳)}$$

در رابطه (۳) D_c تعداد نقاطی از فریم کنونی را نشان می دهد که ناحیه مخروطی خط واصل دوربین به آن ها از آخرین فریم کلیدی به فریم کنونی تغییر کرده است. همچنین می توان رابطه دوم در صورت کسر رابطه ۲ را به طور مشابه به دست آورد.

اما همان طور که گفته شد مقدار اولیه با فرض مشابه بودن فریم کنونی و فریم رفرنس به دست آمده است؛ در حالی که در واقعیت به دلیل دور شدن این فریم از فریم رفرنس، این فرض برقرار نیست و مقدار اولیه، حد آستانه ساده ای برای تعداد مجاز نقاطی که ناحیه آن ها تغییر خواهد کرد، می باشد؛ بنابراین باید اصلاحاتی به این مقدار اولیه وارد شود که در ادامه مورد بحث قرار گرفته است. اولین ضریبی که برای دشوارتر کردن این مقدار اولیه استفاده می شود، به صورت رابطه (۴) تعریف می شود:

$$\theta = 1 - \frac{M_c}{M_r} \quad \text{رابطه (۴)}$$

ضریب θ به نسبت کاهش تعداد نقاط متناظر از فریم رفرنس به فریم کنونی، حد آستانه اولیه را دشوار می کند. کاهش تعداد نقاط متناظر بین فریم کنونی و آخرین فریم کلیدی، نشان دهنده افزایش تغییرات وضعیت فریم کنونی نسبت به آخرین فریم کلیدی می باشد؛ بنابراین به همین نسبت می توان حد آستانه تعداد نقاط با ناحیه تغییر کرده در فریم کنونی را افزایش داد؛ بنابراین، این ضریب به صورت رابطه (۵) و به عنوان ضریبی از مقدار حد آستانه اولیه به خود آن اضافه می شود:

$$th_{adaptive} = th_{initial}(1 + \theta) \quad \text{رابطه (۵)}$$

حد آستانه انطباقی که تاکنون به دست آمده است، همچنان شرایط ایده آلی را شبیه سازی می کند که به دلیل تأثیر عوامل دیگر، این شرایط ایده آل به ندرت رخ می دهد؛ چراکه در این

تغییرات شبیه سازی می کند و سپس این حد آستانه اولیه را با ضرایبی که وابسته به وضعیت هندسی فریم ها هستند به واقعیت مسئله نزدیک می کند.

مقدار اولیه حد آستانه بر اساس این اصل انتخاب شده است که چه تعداد نقطه با ناحیه تغییر کرده در فریم کنونی نیاز است تا فریم کنونی مشابه با فریم رفرنس باشد؟ منظور از نقاط با ناحیه تغییر کرده، نقاطی است که ناحیه خط واصل دوربین به نقطه آن ها از آخرین فریم کلیدی به فریم کنونی تغییر کرده است. این حد آستانه انطباقی اولیه از رابطه (۲) به دست می آید:

$$Th_{initial} = \frac{\frac{M_c D_r}{M_r} + \frac{A_c D_r}{A_r}}{2} \quad \text{رابطه (۲)}$$

جایی که M_r معرف تعداد نقاط فریم رفرنس که متناظر آن ها در آخرین فریم کلیدی وجود دارد، M_c معرف تعداد نقاط فریم کنونی که متناظر آن ها در آخرین فریم کلیدی وجود دارد، D_r تعداد نقاطی از فریم رفرنس را نشان می دهد که ناحیه مخروطی خط واصل دوربین به آن ها از آخرین فریم کلیدی به فریم کنونی تغییر کرده است. همچنین A_c و A_r به ترتیب، تعداد کل نقاط موجود در فریم کنونی و فریم رفرنس را نشان می دهند. این مقدار اولیه، از میانگین گیری از دو مقدار اولیه به دست آمده است: یکی مقدار اولیه به دست آمده از کل نقاط و دیگری مقدار اولیه به دست آمده از نقاط متناظر بین فریم کنونی و فریم کلیدی آخر. همان طور که از رابطه مقدار اولیه مشخص است، این مقدار اولیه با این فرض به دست آمده است که فریم کنونی کاملاً مشابه با فریم رفرنس باشد؛ چراکه در صورت این رابطه می بینیم که هر یک از رابطه ها از یک نسبت تناسب، به این صورت به دست آمده اند که اگر تعداد نقاط متناظر با آخرین فریم کلیدی در فریم رفرنس برابر M_r و تعداد نقاطی که ناحیه آن ها از فریم کلیدی آخر به فریم رفرنس تغییر کرده است، برابر D_r باشد، می توان تعداد مجاز نقاطی که ناحیه آن ها در فریم کنونی تغییر خواهد کرد را با داشتن تعداد نقاط متناظر با آخرین فریم کلیدی آن از

احتمال اخذ یک فریم کلیدی جدید را افزایش دهد.

$$\alpha = \frac{D_c}{M_c} - 0.5 \quad \text{رابطه (۸)}$$

ضریب α متناسب با تغییر ناحیه بیش از نصف نقاط متناظر داخل هر فریم، حد آستانه را ساده می‌کند. وقتی که ناحیه مخروطی بیش از نصف نقاط متناظر داخل هر فریم تغییر کند، فریم به سمت ناپایدار شدن می‌رود و نیاز به اخذ فریم کلیدی جدید افزایش پیدا می‌کند؛ بنابراین رابطه حد آستانه انطباقی به‌صورتی تبدیل می‌شود که رابطه (۹) حاصل آن است.

$$th_{adaptive} = th_{initial}(1 + \theta - \beta - \alpha) \quad \text{رابطه (۹)}$$

این حد آستانه قابلیت انطباق در تمامی شرایط را داراست و به‌خوبی می‌تواند مطابق با نحوه حرکت ربات و جابه‌جایی‌های دوربین و با توجه به تعداد نقاط متناظر و نقاط با ناحیه تغییر کرده داخل هر فریم، نسبت به اخذ فریم کلیدی تصمیم‌گیری کند؛ اما این حد آستانه هیچ کنترلی بر شرایط و پایداری فریم ندارد و نمی‌تواند تضمین کند که اگر این فریم به‌عنوان فریم کلیدی انتخاب شود، فریم مستحکمی خواهد بود یا خیر؛ به همین دلیل معیار دیگری تعریف شده است که با بررسی توزیع نقاط داخل هر فریم، تعادل و توازن و در نتیجه استحکام فریم را بررسی می‌کند و فریم‌هایی که از فیلتر حد آستانه انطباقی عبور می‌کنند، برای بررسی پایداری وارد معیار بعد می‌شوند که در قسمت بعد شرح داده شده است.

۳-۱-۲- محاسبه معیار تعادل مرکز ثقل (ECOG)^۱

حد آستانه انطباقی که در قسمت قبل شرح داده شد، فقط قادر است که تغییرات و جابه‌جایی‌های فریم کنونی نسبت به آخرین فریم کلیدی را مدل‌سازی کند و از این‌رو با دور شدن فریم و افزایش میزان تغییرات، احتمال صادر شدن پاسخ مثبت به‌منظور اخذ فریم کلیدی توسط این حد آستانه افزایش می‌یابد؛ درحالی‌که هیچ کنترلی بر شرایط

رابطه فرض بر این است که افزایش تعداد نقاطی که ناحیه آن‌ها تغییر کرده است دقیقاً متناسب با نسبت کم شدن تعداد نقاط متناظر می‌باشد؛ درحالی‌که در عمل این‌گونه نیست و حتی ممکن است که کم شدن نقاط متناظر به‌دلیل ردیابی نشدن آن‌ها در فریم جدید و بد بودن شرایط نوری صحنه باشد و یا حتی افزایش تعداد نقاطی که ناحیه آن‌ها تغییر کرده است به‌دلیل جابه‌جایی بیش‌ازحد فریم باشد نه تغییر وضعیت بیش‌ازحد آن؛ به هر حال از آنجایی که این ضریب در یک شرایط کاملاً ایده‌آل تصور شده است، یک ضریب سخت برای حد آستانه اولیه می‌باشد که در اکثر مواقع اتفاق نمی‌افتد و تعداد نقاطی که ناحیه آن‌ها تغییر کرده است به این حد آستانه نمی‌رسد و هیچ فریم کلیدی اخذ نمی‌شود؛ به همین دلیل ضریب β به‌صورت رابطه (۶) محاسبه شده است تا به میزان افزایش نسبت نقاط با ناحیه تغییر کرده به نقاط متناظر از فریم رفرنس به فریم کنونی، حد آستانه را ساده‌تر کند تا بر شرایط واقعی منطبق شود:

$$\beta = \frac{D_c}{M_c} - \frac{D_r}{M_r} \quad \text{رابطه (۶)}$$

افزایش این ضریب به معنی کاهش پایداری فریم و دور شدن و تغییر وضعیت شدید آن نسبت به فریم رفرنس می‌باشد که این شرایط علاوه بر اینکه نیاز به اخذ فریم کلیدی جدید را افزایش می‌دهد، نشان‌دهنده بهبود هندسه مثلث تقاطع نیز می‌باشد؛ بنابراین به‌منظور اخذ سریع‌تر فریم کلیدی جدید نیاز است تا به نسبت افزایش این ضریب، حد آستانه ساده شود و در نهایت رابطه حد آستانه انطباقی به‌صورت رابطه (۷) خواهد بود:

$$th_{adaptive} = th_{initial}(1 + \theta - \beta) \quad \text{رابطه (۷)}$$

رابطه (۷) تا حد زیادی شرایط واقعی را مدل‌سازی کرده است و انطباق خوبی با شرایط متفاوت دارد؛ اما نیاز است ضریب دیگری برای آن در نظر گرفته شود تا در مواردی که تغییرات وضعیت فریم بحرانی می‌شود و ممکن است از قواعد قبلی پیروی نکند، حد آستانه را ساده‌تر کرده تا

ماتریس و فاصله بین آن‌ها به صورت رابطه (۱۰) تعریف می‌شود:

$$ECOG = \frac{d + max_2}{2\sqrt{2} max_1} \quad \text{رابطه (۱۰)}$$

جایی که max_1 بزرگ‌ترین عدد و max_2 دومین عدد بزرگ داخل ماتریس و d فاصله بین آن‌ها داخل ماتریس می‌باشد از طریق رابطه (۱۱) محاسبه می‌شود:

$$d = \sqrt{(r_{max_2} - r_{max_1})^2 + (c_{max_2} - c_{max_1})^2} \quad \text{رابطه (۱۱)}$$

جایی که r و c مؤلفه‌های سطر و ستون ماتریس هستند. مقدار $2\sqrt{2}$ که در مخرج قرار گرفته است، بیشترین فاصله ممکن بین دو درایه از یک ماتریس ۳ در ۳ می‌باشد. از رابطه (۱۰) مشخص است که با کاهش فاصله بین دو عدد بزرگ این ماتریس، معیار ECOG هم کاهش یافته و این کاهش فاصله به معنی تجمع اکثریت نقاط در یک سمت ماتریس و در نتیجه بد بودن توزیع نقاط و استحکام فریم می‌باشد. برای کنترل این مقدار، از مرکز ثقل ماتریس به عنوان حد آستانه استفاده شده است و تا زمانی که معیار ECOG از اندازه بردار مرکز ثقل ماتریس کمتر باشد، توزیع نقاط داخل فریم مناسب است و این فریم می‌تواند فریم کلیدی باشد. اندازه بردار مرکز ثقل از رابطه (۱۲) به دست می‌آید:

$$th = \sqrt{X_c^2 + Y_c^2} \quad \text{رابطه (۱۲)}$$

در رابطه فوق X_c و Y_c مؤلفه‌های بردار مرکز ثقل می‌باشند. استحکام فریم‌هایی که از حد آستانه انطباقی عبور می‌کنند توسط معیار ECOG بررسی می‌شوند و در صورت برآورده شدن این معیار، فریم به عنوان فریم کلیدی انتخاب می‌شود. همان‌طور که گفته شد معیار حد آستانه انطباقی و معیار ECOG در تضاد با یکدیگر هستند و فقط هنگامی یک فریم کلیدی جدید اخذ می‌شود که این دو معیار به صورت هم‌زمان آن فریم را تأیید کنند؛ بنابراین تا جایی که هم تعادل و توزیع نقاط داخل فریم مناسب باشد و هم تغییرات فریم

پایداری و استحکام فریم وجود ندارد و این درحالی است که با افزایش تغییرات فریم کنونی نسبت به آخرین فریم کلیدی، احتمال کاهش استحکام فریم بالا رفته و این فریم نمی‌تواند به خوبی با آخرین فریم کلیدی تراز شود؛ بنابراین حد مجاز تغییرات فریم کنونی جایی است که استحکام فریم از دست نرود و به این منظور معیار ECOG به صورت کاملاً انطباقی و در تضاد با حد آستانه انطباقی تعریف شده است تا با بررسی توزیع نقاط متناظر در فریم کنونی، پایداری فریم و میزان تغییرات آن را کنترل کند؛ علاوه بر این برای تشکیل یک ابر نقطه متراکم و دقیق نیازمند انتخاب فریم‌های کلیدی هستیم که نقاط مؤثر آن‌ها توزیع مناسبی داشته باشند. نقاط مؤثر نقاطی هستند که زاویه دید آن‌ها از آخرین فریم کلیدی به فریم کنونی بیشتر از δ تغییر کرده باشد. این زاویه با بررسی هم‌زمان مقدار خطای مطلق مسیر حرکت و تراکم ابر نقطه تشکیل شده و به صورت تجربی (در بخش پیاده‌سازی ۴-۲) برابر ۳۰ درجه انتخاب شده است. با توجه به اصول تشکیل مثلث فتوگرامتری محکم، مقدار ۳۰ درجه برای زاویه δ مقداری معقول می‌باشد.

برای پیاده‌سازی این معیار ابتدا یک گرید ۳ در ۳ در فریم تشکیل می‌شود و داخل هر سلول از این گرید تعداد نقاطی که ناحیه مخروطی آن‌ها از آخرین فریم کلیدی تاکنون بیشتر از ۳۰ درجه تغییر کرده است، شمارش می‌شوند. این فرآیند یک ماتریس ۳ در ۳ برای هر فریم ایجاد می‌کند. ایده‌آل‌ترین حالت برای این ماتریس هنگامی است که تمامی درایه‌های آن یکسان باشند و در این حالت بهترین توزیع برای نقاط وجود دارد؛ اما این حالت ایده‌آل دور از واقعیت بوده و به ندرت رخ می‌دهد؛ بنابراین تعریف حالت ایده‌آل به این صورت تغییر می‌کند که دو عدد بزرگ داخل این ماتریس بیشترین فاصله از یکدیگر را داشته باشند. معیار ECOG به نحوی تعریف می‌شود که در این حالت ایده‌آل مقدار یک و در بقیه حالت‌ها مقداری بین صفر و یک داشته باشد. به این ترتیب این معیار به صورت ترکیبی از اعداد داخل

EuRoC انجام شده است. ارزیابی کیفی با مقایسه شهودی مسیر حرکت پیموده‌شده توسط ربات با مسیر حرکت مرجع و همچنین مقایسه ابر نقطه تشکیل‌شده از فریم‌های کلیدی انتخاب‌شده توسط الگوریتم‌ها حاصل شده است. به‌منظور ارزیابی کمی دقت تعیین موقعیت، از معیار خطای مطلق مسیر حرکت (ATE) و به‌منظور ارزیابی کمی ابر نقاط، اختلاف نقاط هر ابر نقطه با ابر نقطه مرجع محاسبه شده و نمودار فراوانی^۱ خطاها مقایسه شده است. در آخر، زمان پردازش هر الگوریتم در تمامی دنباله تصاویر داده EuRoC نیز مقایسه شده است. روش پیشنهادی توسط یک لپ‌تاپ Core i7 - 4510U (4 cores @ 2 GHz) و با 8GB RAM و در سیستم عامل لینوکس ubuntu انجام شده است. این روش در محیط برنامه‌نویسی ++C به‌عنوان یک تابع به توابع موجود در الگوریتم ORB-SLAM3 اضافه شده است. کد توسعه داده‌شده در آدرس <https://github.com/arashazimi0032/PKS> قرار داده شده است.

۴-۱- داده مورد استفاده

داده EuRoC^۲ یکی از مشهورترین داده‌های مورد استفاده در بررسی و ارزیابی دقت الگوریتم‌های ماشین بینایی در ناوبری خودکار می‌باشد. این مجموعه داده در دو دسته تهیه شده است که هر دو دسته شامل تصاویر استریو و اندازه‌گیری‌های IMU می‌باشند. اولین دسته که machine hall نامگذاری شده، در یک محیط صنعتی جمع‌آوری شده است و دسته دوم که Vicon Room نام‌گذاری شده، داخل یک اتاق جمع‌آوری شده است و شامل موقعیت واقعی اندازه‌گیری شده توسط یک سیستم ردیابی لیزر، با دقت میلی‌متر می‌باشند. همچنین داده vicon شامل ابر نقطه متراکم تولیدشده توسط سیستم لیزر اسکن می‌باشد که به‌عنوان ابر نقطه مرجع برای این داده قرار داده شده است. درمجموع، یازده دنباله داده در این مخزن داده وجود دارد که شامل

کنونی نسبت به آخرین فریم کلیدی به‌قدری محسوس باشد که بتوان تشکیل مثلث فتوگرامتری محکم را تضمین کرد، یک فریم کلیدی جدید می‌تواند اخذ شود.

۳-۲- انتخاب فریم کلیدی توسط سنسور IMU

در سیستم‌های بینایی-اینرشیا لطف وجود سنسور IMU می‌توان حرکات سریع و ناگهانی را که دوربین توانایی مدیریت آن‌ها را ندارد، کنترل کرده و استحکام الگوریتم را در این شرایط حفظ کرد. در روش پیشنهادشده در این مقاله، از سنسور IMU برای اخذ فریم کلیدی در حرکات ناگهانی که موجب تغییرات شدید در موقعیت و وضعیت فریم و صحنه پیشرو می‌شود، استفاده شده است. به این صورت هنگامی که شتاب حرکت تغییر شدیدی داشته باشد، سنسور IMU به‌صورت مستقل و جدا از دوربین اقدام به اخذ فریم کلیدی می‌کند. از آنجایی که این سنسورها کاملاً مستقل از یکدیگر هستند و انتخاب فریم کلیدی توسط IMU ارتباطی به وضعیت فریم‌ها ندارد، با یک حد آستانه ساده که به‌صورت تجربی برابر ۱ متر بر مجذور ثانیه به‌دست آمده است، می‌توان نسبت به اخذ فریم کلیدی اقدام کرد. با توجه به اینکه در حرکات با شتاب حرکت بیشتر از ۱ متر بر مجذور ثانیه، موقعیت دوربین در کسری از ثانیه تغییر چشمگیری می‌کند، اخذ هرچه سریع‌تر فریم کلیدی در این شرایط، احتمال تناظریابی بهتر بین فریم کنونی و آخرین فریم کلیدی را افزایش می‌دهد. لازم به‌ذکر است که منظور از حرکات ناگهانی، حرکات و جابه‌جایی‌های سریع سکو است و شامل لرزش‌ها و نوسان‌های واردشده به سکو نمی‌شود. چراکه اولاً لرزش‌ها توسط IMU‌های رایج قابل تشخیص نیستند و ثانیاً جابه‌جایی‌هایی که موجب ایجاد وضعیت هندسی جدید برای فریم می‌شود را ایجاد نمی‌کنند.

۴- پیاده‌سازی و ارزیابی

به‌منظور ارزیابی کمی و کیفی روش پیشنهادشده در این مقاله، آزمایش‌های گسترده‌ای بر روی مجموعه داده

1- histogram

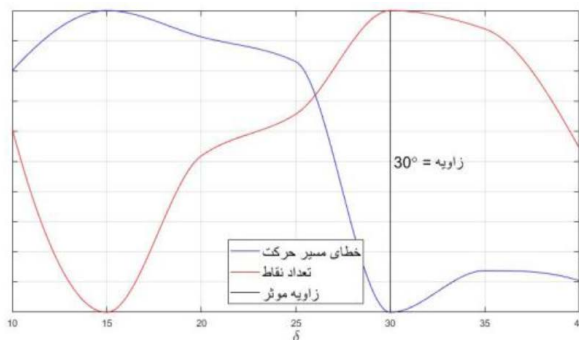
2- European Robotic Challenge <https://projects.asl.ethz.ch/datasets/doku.php?id=kmauvisualinertialdatasets>



نگاره ۳: نمونه‌ای از تصاویر داده‌های EuRoC. (الف) داده Vicon. (ب) داده Machine Hall

در این مقاله با تغییر مقدار این زاویه از ۱۰ درجه تا ۴۰ درجه و مقایسه خطای مطلق مسیر حرکت و همچنین تراکم ابر نقطه تولیدشده، بر روی یکی از دنباله تصاویر موجود (V101)، مقدار بهینه به نحوی انتخاب شده است که متراکم‌ترین ابر نقطه و درعین حال کمترین خطای مسیر حرکت حاصل شود. به عبارت دیگر رابطه (۱۳) باید بیشینه شود.

رابطه (۱۳) $\text{argmax}_{\delta \in \mathbb{R}} (\text{density}(\delta) - \text{ATE}(\delta))$ نتایج اجرای الگوریتم با مقادیر مختلف زاویه مؤثر در نگاره ۴ ارائه شده است. به منظور کنار هم نمایش دادن دو پارامتر خطای مسیر حرکت و تعداد نقاط ابر نقطه، این دو پارامتر در بازه صفر تا یک نرمال شده‌اند. همان‌طور که مشخص است بهترین مقدار زاویه مؤثر برابر ۳۰ درجه می‌باشد.



نگاره ۴: مقادیر خطای مسیر حرکت و تراکم ابر نقطه برای مقادیر متفاوت δ

پروازهای ربات در شرایط و چالش‌های متفاوت، شامل پرواز آرام و نرم (داده‌های آسان) و پروازهای پرافت‌وخیز و با حرکات سریع و در شرایط نوری بد (داده‌های متوسط و دشوار) می‌باشد. در نگاره ۳ نمونه‌ای از این داده‌ها مشاهده می‌شود.

به منظور بررسی دقت تعیین موقعیت روش PKS، این روش برای تمامی یازده داده موجود در مخزن داده EuRoC در دو حالت دو دوربین و تک دوربین اجرا شده است و نتایج آن‌ها به دو صورت کیفی و کمی، با نتایج اجرای الگوریتم ORB-SLAM3 مقایسه شده است. به منظور جلوگیری از به دست آمدن نتایج تصادفی، هر یک از الگوریتم‌ها در هر یک از دو حالت تک دوربین و دو دوربین، برای هر دنباله داده، ۱۰ بار اجرا شده است و میانگین نتایج آن‌ها قرار داده شده است.

۴-۲- محاسبه مقدار بهینه زاویه مؤثر نقاط (δ)

زاویه مؤثر نقاط، زاویه‌ای است که اگر نقاطی که زاویه خط دید آن‌ها بیشتر از این زاویه تغییر کرده است توزیع مناسبی داخل فریم داشته باشند، می‌توان گفت آن فریم از وضعیت هندسی خوبی برخوردار است و می‌تواند به عنوان فریم کلیدی انتخاب شود. انتخاب مناسب این زاویه باعث بهبود دقت تعیین موقعیت ربات و همچنین بهبود کیفیت ابر نقطه تولیدشده از فریم‌های کلیدی می‌شود.

ردیابی شده توسط الگوریتم، به سیستم مختصات زمینی انتقال داده می‌شود. مسیرهای خروجی حاصل از اجرای هر دو الگوریتم در دو حالت تک دوربین و دو دوربین در مقایسه با مسیرهای واقعیت زمینی برای بعضی از دنباله تصاویر موجود در مخزن داده EuRoC در نگاره‌های ۵ و ۶ ارائه شده است. لازم به ذکر است که داده‌های V103, MH04, MH05, V203، داده‌های سطح دشوار در مجموعه داده EuRoC هستند.

واضح است که الگوریتم PKS نسبت به ORB-SLAM3، در هر دو حالت تک دوربین و دو دوربین انحرافات کمتری از مسیر واقعیت زمینی داشته است و این مسئله در پیچ‌های مسیر بیشتر مشهود است.

در نگاره‌های ۷ و ۸ مسیر فریم‌های کلیدی در مقایسه با مسیر واقعیت زمینی ارائه شده است. همان‌طور که مشخص است، فریم‌های کلیدی نیز انطباق بیشتر مسیر حرکت الگوریتم PKS را نسبت به الگوریتم ORB-SLAM3 تأیید می‌کند. نکته حائز اهمیت دیگر در مورد مسیر فریم‌های کلیدی این است که در بعضی از دنباله تصاویر (اکثراً داده Vicon) و مخصوصاً در حالت دو دوربین، در محل‌هایی که پیچیدگی مسیر زیاد باشد (به‌طور مثال در پیچ‌های تند) الگوریتم ORB-SLAM3 فریم‌های کلیدی بیشتری گرفته و با مسیر واقعیت زمینی منطبق‌تر به نظر می‌رسد؛ اما با توجه به نتایج ATE که در قسمت بعد ارائه شده است، فریم‌های کلیدی اخذ شده در شرایطی با تغییرات سریع زاویه دید، به دلیل هندسه ضعیف فریم‌ها در این مواقع و توزیع نامناسب نقاط کلیدی در فریم، فریم‌های مناسبی نیستند و استحکام شبکه فریم‌های کلیدی را کاهش می‌دهند؛ اما الگوریتم PKS به دلیل بررسی استحکام فریم‌ها توسط معیار ECOG، در شرایطی با تغییرات سریع مسیر حرکت، فریم‌های کلیدی زیادی را اخذ نمی‌کند و به این ترتیب استحکام شبکه فریم‌های کلیدی را حفظ می‌کند. این مسئله در داده V203 و در حالت دو دوربین به وضوح مشخص است.

۴-۳- مقایسه شهودی مسیر حرکت ربات در دو الگوریتم با مسیر واقعیت زمینی

مقایسه شهودی مسیر حرکت ربات یکی از مرسوم‌ترین روش‌های کیفی بررسی وضعیت کلی تعیین موقعیت الگوریتم می‌باشد. با نمایش دادن مسیر حرکت ربات در کنار مسیر واقعیت زمینی، می‌توان تخمین کلی از کیفیت و کارایی الگوریتم به دست آورد. به‌طور مثال، می‌توان میزان انحراف الگوریتم‌ها را از مسیر واقعیت زمینی با یکدیگر مقایسه کرد و به صورت کیفی تعیین کرد که کدام الگوریتم کارایی بهتری داشته است.

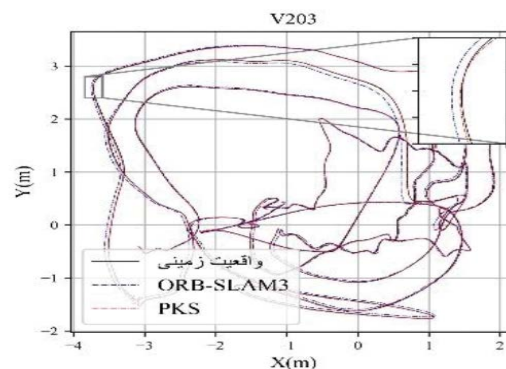
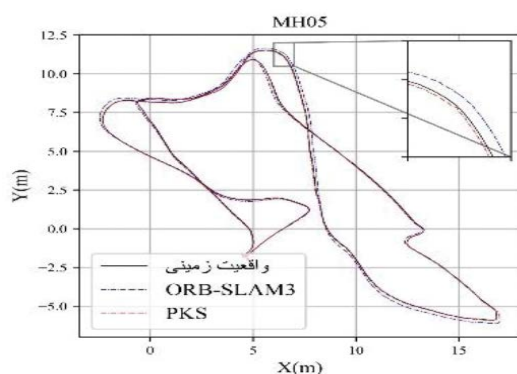
پس از اجرای الگوریتم PKS و همچنین الگوریتم ORB-SLAM3 (در هر دو حالت تک دوربین و دو دوربین) بر روی تمامی دنباله تصاویر موجود در مخزن داده EuRoC، موقعیت‌های خروجی از هر الگوریتم استخراج شد و با نمایش دادن این موقعیت‌ها در یک فضای دوبعدی در کنار موقعیت‌های موجود از داده‌های واقعیت زمینی، بررسی کیفی مسیر حرکت طی شده توسط هر الگوریتم انجام شد.

لازم به ذکر است، از آنجایی که موقعیت‌های به دست آمده از الگوریتم در یک سیستم مختصات دلخواه هستند و قابل مقایسه با مسیر حرکت واقعیت زمینی نیستند، ابتدا لازم است تا موقعیت‌های خروجی هر الگوریتم توسط یک تابع تبدیل مشابهت^۱ (Zhang & Scaramuzza, 2018) به سیستم مختصات زمینی منتقل شوند و سپس با موقعیت‌های واقعی مقایسه شوند.

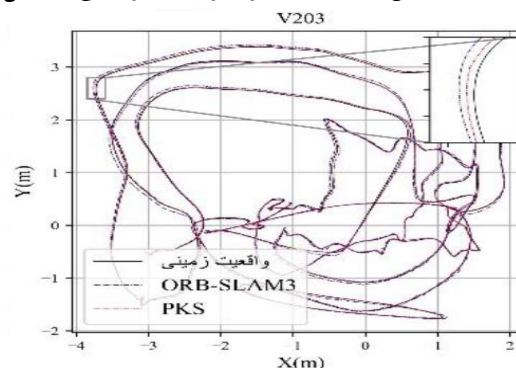
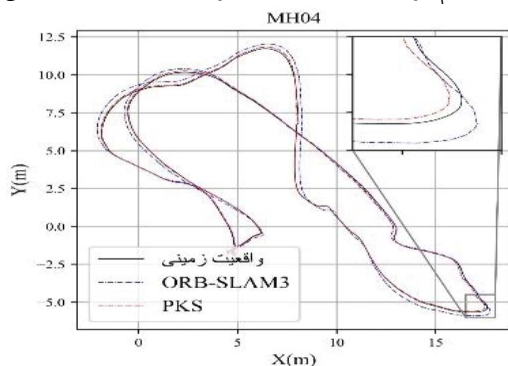
برای انتقال مسیر ردیابی شده توسط الگوریتم از سیستم مختصات محلی به سیستم مختصات زمینی، ابتدا مختصات تعدادی فریم از این مسیر انتخاب شده و با استفاده از زمان دریافت هر فریم، مختصات فریم متناظر آن‌ها در داده واقعیت زمینی استخراج می‌شود. سپس با در نظر گرفتن این نقاط به عنوان نقطه کنترل و با استفاده از یک تبدیل کانفورمال سه‌بعدی^۲، موقعیت مسیر

1- Similarity Transform

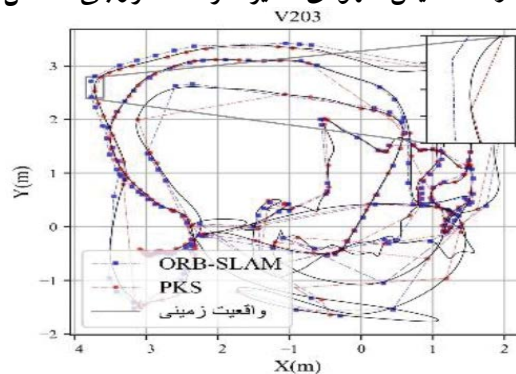
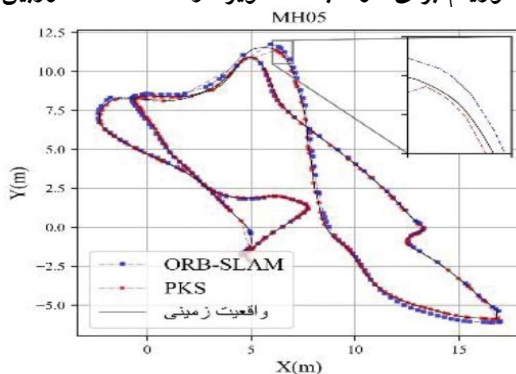
2- 3D conformal



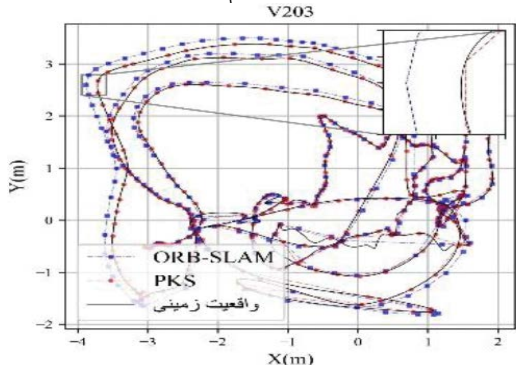
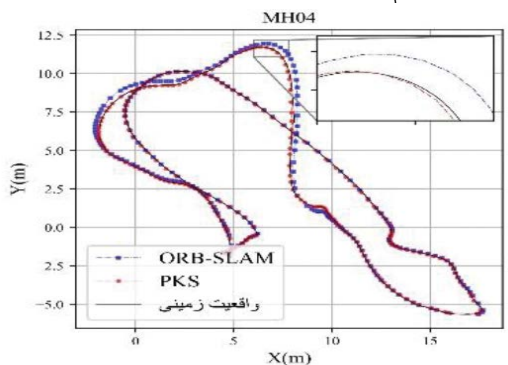
نگاره ۵: نمایش شهودی مسیر حرکت خروجی حاصل از هر دو الگوریتم برای دو دنباله تصویر در حالت دو دوربین



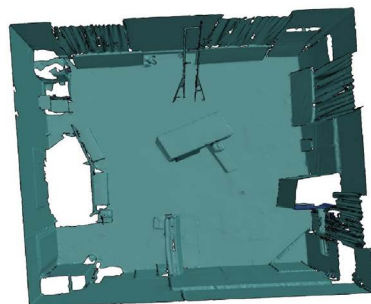
نگاره ۶: نمایش شهودی مسیر حرکت خروجی حاصل از هر دو الگوریتم برای دو دنباله تصویر در حالت تک دوربین



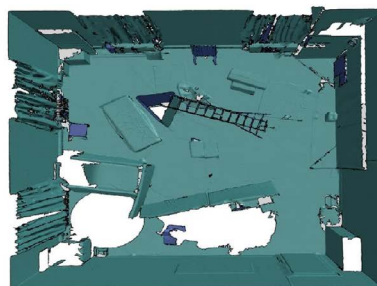
نگاره ۷: نمایش شهودی مسیر فریم‌های کلیدی خروجی حاصل از هر دو الگوریتم برای دو دنباله تصویر در حالت دو دوربین



نگاره ۸: نمایش شهودی مسیر فریم‌های کلیدی خروجی حاصل از هر دو الگوریتم برای دو دنباله تصویر در حالت تک دوربین



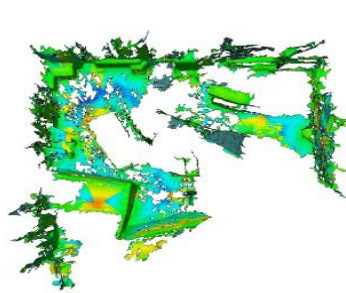
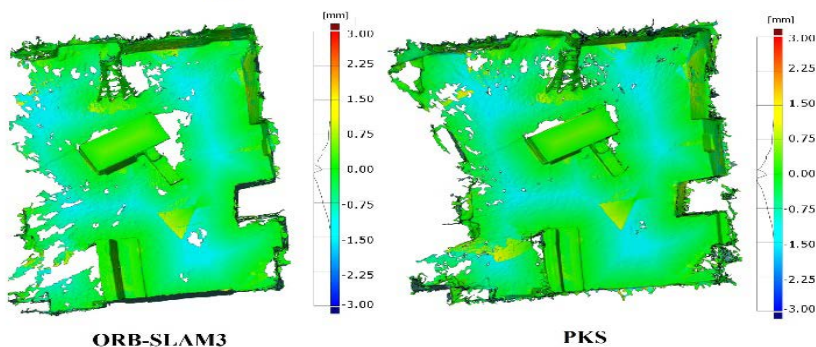
داده مرجع - V101



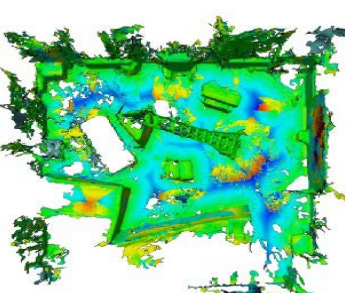
داده مرجع - V202 - V203

نگاره ۹: ابر نقطه مرجع داده V101 و V202

نگاره ۱۰: ابر نقطه حاصل از فریم‌های کلیدی هر دو الگوریتم در داده V101



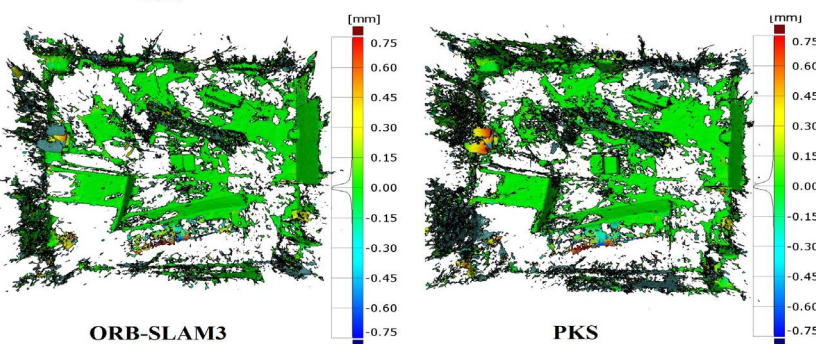
ORB-SLAM3



PKS

نگاره ۱۱: ابر نقطه حاصل از فریم‌های کلیدی هر دو الگوریتم در داده V202

نگاره ۱۲: ابر نقطه حاصل از فریم‌های کلیدی هر دو الگوریتم در داده V203



ORB-SLAM3

PKS

۴-۴- مقایسه کیفی ابر نقاط

ابرنقطه متراکم در نرم‌افزار metashape شد. مدل‌های نهایی تولیدشده از لحاظ کیفی و کمی با ابر نقطه مرجع مقایسه شده‌اند و نتایج آن‌ها در نگاره‌های ۹ تا ۱۲ ارائه شده است. در نگاره‌های ۹ و ۱۰ و ۱۱ و ۱۲، کیفیت و تراکم بهتر ابر نقطه الگوریتم PKS و همچنین استخراج کامل‌تر اجسام

برای ارزیابی کیفیت ابر نقاط الگوریتم‌ها، فریم‌های کلیدی انتخابی آن‌ها برای سه دنباله تصویر (یک ساده، یک متوسط و یک دشوار) از بین کل فریم‌ها جدا شده و برای تشکیل مدل سه‌بعدی ابر نقطه وارد فرآیند SFM و تولید

PKS، از معیار خطای مطلق مسیر حرکت استفاده می‌شود که مطابق با روش گفته‌شده در (Zhang & Scaramuzza, 2018) محاسبه می‌شود.

معیار ATE برای الگوریتم PKS و الگوریتم ORB-SLAM3 در دو حالت تک دوربین و دو دوربین، برای تمامی دنباله تصاویر محاسبه شده و نتایج آن در ادامه ارائه شده است. این نتایج که حاصل از میانگین‌گیری از ۱۰ بار اجرای هر الگوریتم در هر مجموعه داده می‌باشد نشان می‌دهد که در هر دو حالت تک دوربین و دو دوربین، الگوریتم PKS نتایج بهتری داشته است. همچنین از آنجایی که مقدار انحراف معیار الگوریتم PKS نیز کمتر می‌باشد، می‌توان گفت که استحکام و پایداری این روش بیشتر است؛ چراکه انحراف معیار بالاتر در نتایج ATE نشان‌دهنده نتایج پراکنده و بی‌ثبات الگوریتم بوده که این مسئله در روش PKS به‌وضوح بهبود پیدا کرده است (نگاره‌های ۱۳ تا ۱۶).

همان‌طور که انتظار می‌رفت و در قسمت قبل هم دیدیم، مقادیر کمی دقت الگوریتم PKS بهتر از الگوریتم ORB-SLAM3 بوده و بهبود دقت روش PKS به‌وضوح دیده می‌شود. نتایج ATE تجمعی نشان می‌دهد که علی‌رغم کاهش دقت ناشی از روش‌های تصادفی حل ماتریس ضروری در مسئله SLAM، در بعضی از اجراهای الگوریتم، در مجموع و با افزایش تعداد اجراها، دقت الگوریتم PKS بهتر از دقت الگوریتم ORB-SLAM3 می‌باشد.

در ادامه، نتایج میانگین‌گیری شده مقادیر ATE به همراه مقادیر انحراف معیار آن‌ها در قالب نمودارهای میله‌ای نمایش داده شده است (نگاره ۱۷).

از نمودارهای نگاره ۱۷ می‌توان دریافت که در مجموع، الگوریتم PKS در حالت دو دوربین در ۱۰ دنباله تصویر و در حالت تک دوربین در ۸ دنباله تصویر از ۱۱ دنباله موجود نتایج بهتری داشته است. همچنین انحراف معیار این الگوریتم در حالت دو دوربین برای ۶ دنباله تصویر و در حالت تک دوربین برای ۷ دنباله تصویر از ۱۱ دنباله موجود بهتر از الگوریتم ORB-SLAM3 بوده است.

و دیوارهای اتاق، مخصوصاً در داده V202 کاملاً مشهود می‌باشد. همان‌طور که قبلاً هم گفته شد، الگوریتم‌های اسلم و اودومتري تمرکز چندانی به تولید ابرنقطه از فریم‌های کلیدی ندارند و در نتیجه فریم‌هایی را به‌عنوان فریم کلیدی انتخاب می‌کنند که صرفاً نمایانگر خوبی از کل مسیر حرکت باشند و توجهی به همپوشانی فریم‌های کلیدی برای تولید ابر نقطه متراکم ندارند؛ اما در الگوریتم PKS با انتخاب مناسب زاویه مؤثر نقاط و همچنین زاویه ناحیه‌های مخروطی، امکان کنترل بهتر همپوشانی فریم‌ها به‌منظور تولید ابر نقطه متراکم وجود دارد.

۴-۵- مقایسه کمی ابر نقاط

در مورد مقدار کمی خطاها و نمودارهای فراوانی خطا که در کنار هر ابر نقطه ارائه شده است، اختلاف بین الگوریتم‌ها چندان زیاد نیست و نمودارهای فراوانی، نتایج تقریباً یکسانی را نشان می‌دهند؛ اما نکته قابل توجه این است که این مقادیر یکسان خطا در حالی ایجاد شده است که تعداد نقاط ابر نقطه الگوریتم PKS بیشتر از الگوریتم ORB-SLAM3 می‌باشد که خود این مسئله باعث افزایش تعداد و همچنین انحراف معیار خطاها می‌شود؛ اما با این حال می‌بینیم که حتی در داده V202 مقدار انحراف معیار خطاهای ابرنقطه الگوریتم PKS (پهنای نمودار فراوانی) کمتر از ORB-SLAM3 می‌باشد و این موضوع نشان‌دهنده قدرت بالاتر الگوریتم PKS در تولید ابر نقطه می‌باشد.

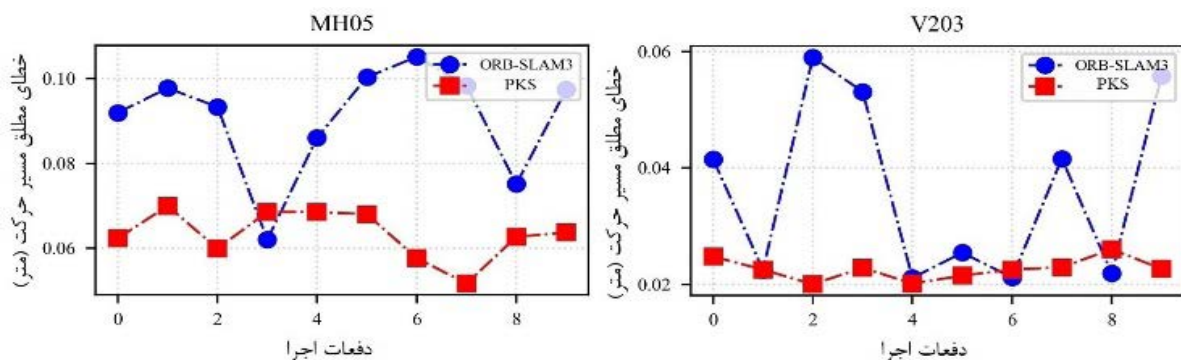
جدول ۱: تعداد نقاط ابر نقطه تشکیل‌شده توسط هر الگوریتم

	V101	V202	V203
PKS	۱۲۲۹۶۳۳	۱۳۹۰۱۸۱	۱۱۲۸۴۵۲
ORB-SLAM3	۹۵۹۵۳۵	۶۰۴۵۹۴	۹۹۲۲۸۳

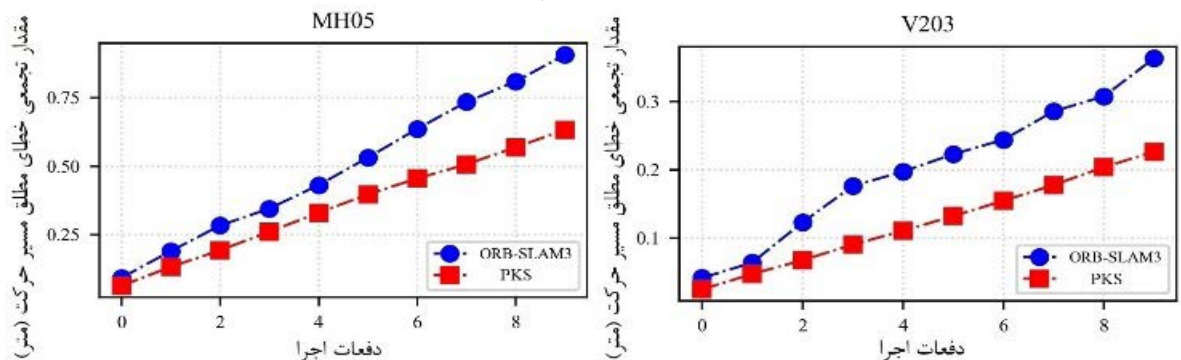
۴-۶- مقایسه کمی دقت تعیین موقعیت الگوریتم‌ها توسط

معیار ATE

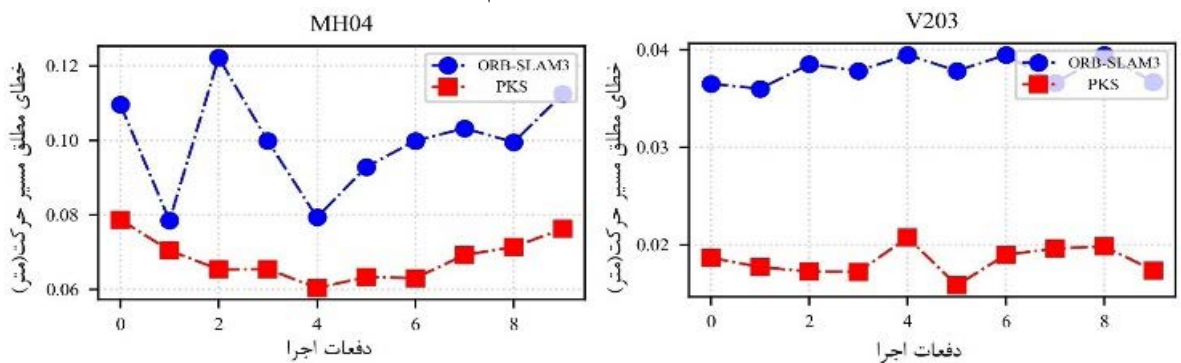
به‌منظور ارزیابی کمی دقت تعیین موقعیت الگوریتم



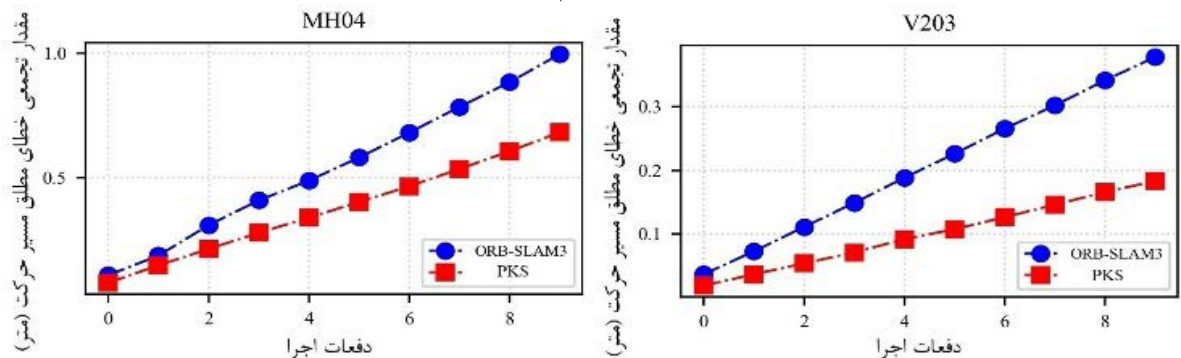
نگاره ۱۳: مقادیر ATE حاصل از ۱۰ بار اجرای هر الگوریتم بر روی چند دنباله تصویر در حالت دو دوربین



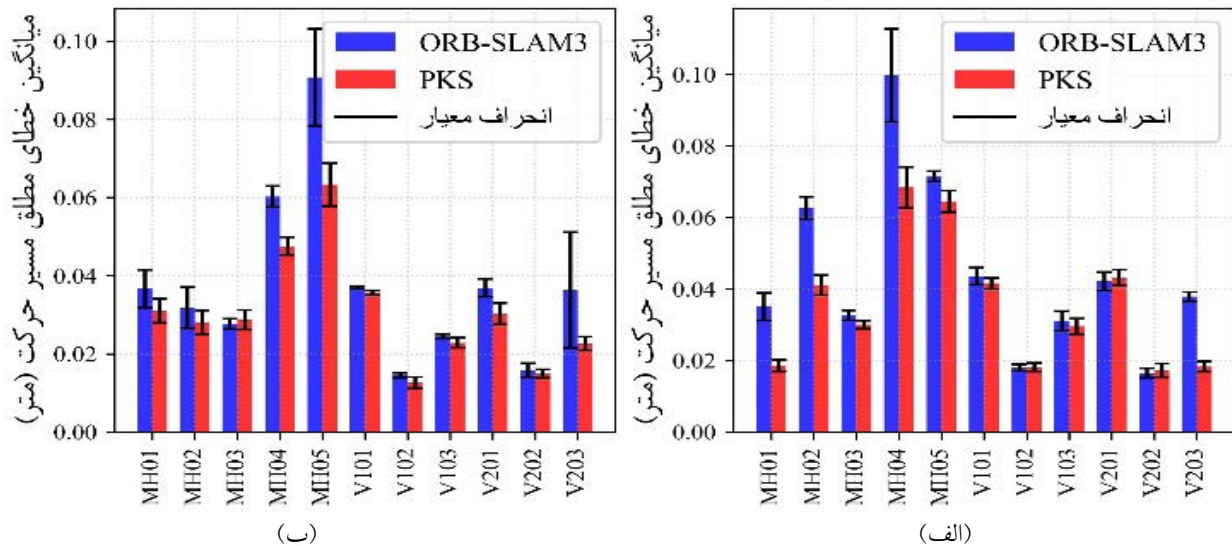
نگاره ۱۴: نمایش تجمعی مقادیر ATE برای ۱۰ بار اجرای هر الگوریتم بر روی چند دنباله تصویر در حالت دو دوربین



نگاره ۱۵: مقادیر ATE حاصل از ۱۰ بار اجرای هر الگوریتم بر روی تمامی دنباله تصاویر در حالت تک دوربین



نگاره ۱۶: نمایش تجمعی مقادیر ATE برای ۱۰ بار اجرای هر الگوریتم بر روی تمامی دنباله تصاویر در حالت تک دوربین



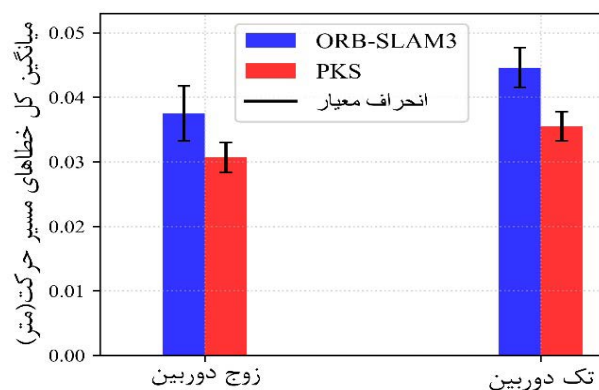
نگاره ۱۷: میانگین مقادیر ATE و انحراف معیار حاصل از هر الگوریتم برای هر دنباله تصویر. (الف) حالت تک دوربین. (ب) حالت دو دوربین.

نکته حائز اهمیت آخر که از نمودارها و جدول ۲ نیز مشخص است، این است که اختلاف دقت کلی الگوریتم PKS در دو حالت تک دوربین و دو دوربین، کمتر از این اختلاف برای الگوریتم ORB-SLAM3 می باشد و این موضوع تأییدی بر استحکام بالای این الگوریتم است. از آنجایی که الگوریتم PKS از اصول هندسی و فتوگرامتری برای انتخاب فریم های کلیدی استفاده می کند، وابستگی آن به سیستم مورد استفاده (دو دوربین و یا تک دوربین) و همچنین شرایط محیط اطراف آن، کمتر از الگوریتم ORB-SLAM3 که از شروط و حد آستانه های ابتکاری برای انتخاب فریم کلیدی استفاده می کند، می باشد.

۴-۷- مقایسه زمان پردازش الگوریتم ها

به منظور ارزیابی زمان پردازش الگوریتم ها، هر الگوریتم برای تمامی یازده دنباله تصویر EuRoC اجرا شده و زمان اجرای هر الگوریتم در هر دنباله تصویر ثبت شده است. همان طور که انتظار می رفت، با توجه به پیچیدگی های محاسباتی اخذ فریم کلیدی در الگوریتم PKS، زمان اجرای این الگوریتم بیشتر از ORB-SLAM3 می باشد؛ چراکه

بهتر بودن انحراف معیار الگوریتم PKS نشان دهنده استحکام و ثبات بیشتر آن بوده که در شرایط و صحنه های مختلف نتایج تقریباً یکسانی را نشان می دهد و کمتر تحت تأثیر شرایط صحنه قرار می گیرد. سرانجام، نمودار میله ای نتایج میانگین گیری شده نشان می دهد که مقدار میانگین ATE الگوریتم PKS در حالت دو دوربین ۱۸/۱٪ و در حالت تک دوربین ۲۰/۴٪ بهبود یافته است. علاوه بر آن، انحراف معیار الگوریتم PKS در حالت دو دوربین ۴۶/۱٪ و در حالت تک دوربین ۲۷/۱٪ بهبود داشته است.



نگاره ۱۸: میانگین مقادیر ATE و انحراف معیار تمامی دنباله تصاویر برای هر دو الگوریتم در حالت تک دوربین و دو دوربین

جدول ۲: زمان اجرای هر الگوریتم در هر دنباله تصویر از داده EuRoC

نام داده	PKS (sec)	ORB-SLAM3 (sec)	نام داده	PKS	ORB-SLAM3 (sec)
MH01	۳۷۹/۰۹	۳۵۰/۰۲	V102	۱۷۳/۰۲	۱۵۸/۱۰
MH02	۳۰۶/۰۳	۲۸۰/۰۳	V103	۲۰۸/۰۶	۱۹۰/۰۸
MH03	۲۷۷/۰۱	۲۵۲/۰۹	V201	۲۱۹/۰۷	۱۹۲/۰۱
MH04	۱۹۷/۰۴	۱۹۰/۰۰	V202	۲۴۰/۰۵	۲۲۰/۰۳
MH05	۲۳۰/۰۶	۲۱۲/۰۹	V203	۱۹۰/۰۶	۱۷۹/۰۴
V101	۳۰۱/۰۱	۲۶۶/۰۵			

آستانه‌های تطبیقی، ناگزیر در بعضی از بخش‌های روش PKS، حد آستانه‌های تجربی استفاده شده‌اند که در ادامه، مورد بحث قرار می‌گیرد.

۵-۱- زاویه ناحیه‌های مخروطی

همان‌طور که گفته شد، زاویه ناحیه‌های مخروطی، بر اساس قواعد مثلث محکم فتوگرامتری و اصول هندسی گفته شده در (Hosseiniaveh et al., 2012)، ۱۰ درجه انتخاب شده است که این ۱۰ درجه، یک حد آستانه قابل تغییر است و با تغییر آن می‌توان نتایج متفاوتی را به دست آورد. با کمتر شدن مقدار این زاویه، حرکات و جابه‌جایی‌های کم دوربین باعث تغییر ناحیه -خط دید نقطه توسط دوربین- می‌شود و فریم‌های کلیدی سریع‌تر انتخاب می‌شوند و در نتیجه ابر نقطه بهتری تشکیل می‌شود. این موضوع باعث افزایش حجم محاسبات می‌شود؛ اما همچنان دقت تعیین موقعیت و کیفیت ابر نقطه خروجی مناسب می‌باشد. از طرفی، افزایش این زاویه باعث انتخاب دیرتر فریم‌های کلیدی شده و به این ترتیب استحکام شبکه فریم‌های کلیدی کاهش پیدا می‌کند و همچنین به دلیل پوشش کم فریم‌های کلیدی، ابر نقطه تشکیل شده از آن‌ها کیفیت مناسبی را نخواهد داشت.

۵-۲- حد آستانه شتاب حرکت برای سنسور IMU

از آنجایی که سنسور IMU تصمیم اخذ فریم کلیدی را به‌طور مستقل از سیستم بینایی و وضعیت فریم‌ها می‌گیرد، اصول هندسی در آن مشارکتی ندارند و بنابراین لازم است

الگوریتم ORB-SLAM3 از حد آستانه‌های ابداعی و ثابت در اخذ فریم کلیدی استفاده می‌کند، درحالی که الگوریتم PKS از قیود و اصول فتوگرامتری در اخذ فریم کلیدی بهره می‌برد و در نتیجه به ازای هر فریم، مقدار تقریباً ثابتی به زمان پردازش‌های الگوریتم افزوده می‌شود؛ بنابراین الگوریتم PKS در عین بهبود دقت تعیین موقعیت و کیفیت ابر نقطه، محاسبات پیچیده‌تری را به الگوریتم تحمیل می‌کند؛ اما این محاسبات اضافی، تأثیر قابل لمسی بر اجرای آنی الگوریتم نمی‌گذارد و همچنان این الگوریتم به‌صورت آنی اجرا می‌شود.

۵- نتیجه‌گیری و پیشنهادات

با توجه به نیاز به یک روش کلی و عمومی^۱ برای انتخاب فریم‌های کلیدی در الگوریتم‌های اودومتری و اسلم بینایی که قابلیت تعمیم برای تمامی شرایط و صحنه‌ها را داشته باشد، در این مقاله یک روش هندسی و بر پایه اصول فتوگرامتری برای انتخاب فریم‌های کلیدی معرفی شد که این روش با توسعه و تحول روش انتخاب فریم‌های کلیدی در الگوریتم ORB-SLAM3 ایجاد شده است. این روش، انعطاف و تعمیم‌پذیری انتخاب فریم‌های کلیدی را با اعمال قیود هندسی و حذف حد آستانه‌های ابتکاری که انعطاف الگوریتم را در شرایط و محیط‌های مختلف تضمین نمی‌کنند، افزایش می‌دهد. باوجود تلاش برای حذف حد آستانه‌های ابتکاری و جایگزینی روش‌های هندسی و حد

ابر نقطه حاصل از دو الگوریتم، از لحاظ کیفیت و تراکم مورد بررسی قرار گرفت و خطای هر ابرنقطه از ابرنقطه مرجع، در قالب نمودار فراوانی نمایش داده شد. به طور کلی نتایج این آزمایش‌ها نشان از بهبود دقت ۲۰ تا ۴۰ درصدی مسیر حرکت الگوریتم پیشنهادی می‌دهد.

در مطالعات آینده، می‌توان روش انتخاب فریم کلیدی پیشنهاد شده را برای دوربین‌های چشم ماهی^۱ و همچنین سیستم‌های فقط-بینایی تعمیم داد و شروط هندسی دیگری (همچون شرط تعداد نقاط دور و نزدیک و یا لحاظ کردن زاویه رأس مثلث تشکیل شده بین نقطه در فریم کنونی، نقطه متناظر در نقشه و نقطه متناظر در آخرین فریم کلیدی در محاسبات انتخاب فریم کلیدی) به مسئله اضافه کرد. همچنین با توجه به کاهش سرعت پردازش و افزایش دقت در الگوریتم PKS، این الگوریتم می‌تواند به عنوان الگوریتم پایه، در تولید داده آموزشی برای الگوریتم‌های یادگیری عمیق استفاده شود و به این ترتیب، سرعت اجرای این الگوریتم نیز بهبود پیدا کند.

منابع و مأخذ

- 1- Ahmadabadian, A. H., Robson, S., Boehm, J., & Shortis, M. (2013). Image selection in photogrammetric multi-view stereo methods for metric and complete 3D reconstruction. Paper presented at the Videometrics, Range Imaging, and Applications XII; and Automated Visual Inspection.
- 2- Bloesch, M., Burri, M., Omari, S., Hutter, M., & Siegwart, R. (2017). Iterated extended Kalman filter based visual-inertial odometry using direct photometric feedback. The International Journal of Robotics Research, 36(10), 1053-1072.
- 3- Campos, C., Elvira, R., Rodríguez, J. J. G., Montiel, J. M., & Tardós, J. D. (2021). ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multimap SLAM. IEEE Transactions on Robotics.
- 4- Engel, J., Koltun, V., & Cremers, D. (2017). Direct sparse odometry. IEEE transactions on pattern analysis

1- Fish-eye

که حد آستانه تغییرات شتاب حرکت، با دقت انتخاب شود. از این رو، حد آستانه تغییرات شتاب حرکت باید براساس نوع ربات مورد استفاده انتخاب شود. مقدار این حد آستانه در مقاله حاضر برابر ۱ (متر بر مجذور ثانیه) انتخاب شده است که این مقدار برای یک ربات پرنده، به صورت تجربی و با بررسی مقادیر شتاب حرکت در تمامی دنباله تصاویر موجود در مخزن داده EuRoC به دست آمده است. با توجه به حرکات نرم تر و آهسته تر ربات‌های زمینی، حد آستانه انتخاب شده ممکن است سخت گیرانه باشد و بهتر است که براساس نوع ربات مورد استفاده بهینه شود.

این مقاله یک روش هندسی انتخاب فریم کلیدی برای سیستم‌های بینایی - اینرشال معرفی کرده است که بر پایه ORB-SLAM3 استوار شده است. به منظور بررسی و ارزیابی الگوریتم پیشنهاد شده، آزمایش‌های گسترده‌ای با مجموعه داده EuRoC، در حالت تک دوربین و دو دوربین انجام شد. نتایج نشان می‌دهد که روش PKS، یک روش کاملاً هندسی انتخاب فریم‌های کلیدی می‌باشد که بدون نیاز به حد آستانه‌های ابتکاری، عملکرد پایدار و مستحکمی را در شرایط مختلف خواهد داشت و علاوه بر آن، برخلاف سایر الگوریتم‌های اسلم و اودومتري که تمرکز چندانی بر روی تولید ابر نقطه و تشکیل مدل سه بعدی ندارند، الگوریتم PKS به دلیل به کارگیری قیود هندسی و فتوگرامتری، از طریق تنظیمات متفاوت زاویه مؤثر و زاویه ناحیه‌های مخروطی، قادر به کنترل کیفیت و تراکم ابر نقطه تشکیل شده نیز خواهد بود و بنابراین ابر نقطه بسیار متراکم و با کیفیتی از فریم‌های کلیدی خروجی آن تولید خواهد شد. به منظور ارزیابی کیفی دقت تعیین موقعیت، مسیر ردیابی شده توسط الگوریتم ORB-SLAM3 و الگوریتم PKS، به صورت شهودی، با مسیر واقعیت زمینی مقایسه شد. از جنبه کمی، مقدار خطای مطلق مسیر حرکت (ATE) دو الگوریتم با یکدیگر مقایسه و نتایج آن برای تمامی دنباله تصاویر موجود در مخزن داده EuRoC، در قالب جدول و نمودارهای مختلف نشان داده شد. همچنین،

1255-1262.

16- Qin, T., Li, P., & Shen, S. (2018). Vins-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics*, 34(4), 1004-1020.

17- Rosinol, A., Abate, M., Chang, Y., & Carlone, L. (2020). Kimera: an open-source library for real-time metric-semantic localization and mapping. Paper presented at the 2020 IEEE International Conference on Robotics and Automation (ICRA).

18- Savran Kızıltepe, R., Gan, J. Q., & Escobar, J. J. (2021). A novel keyframe extraction method for video classification using deep neural networks. *Neural Computing and Applications*, 1-12.

19- Sheng, L., Xu, D., Ouyang, W., & Wang, X. (2019). Unsupervised collaborative learning of keyframe detection and visual odometry towards monocular deep slam. Paper presented at the Proceedings of the IEEE International Conference on Computer Vision.

20- Sze, K.-W., Lam, K.-M., & Qiu, G. (2005). A new key frame representation for video segment retrieval. *IEEE transactions on circuits and systems for video technology*, 15(9), 1148-1155.

21- Tan, W., Liu, H., Dong, Z., Zhang, G., & Bao, H. (2013). Robust monocular SLAM in dynamic environments. Paper presented at the 2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR).

22- Zhang, Z., & Scaramuzza, D. (2018). A tutorial on quantitative trajectory evaluation for visual (-inertial) odometry. Paper presented at the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).

23- Zhuang, Y., Rui, Y., Huang, T. S., & Mehrotra, S. (1998). Adaptive key frame extraction using unsupervised clustering. Paper presented at the Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No. 98CB36269).

and machine intelligence, 40(3), 611-625.

5- Engel, J., Schöps, T., & Cremers, D. (2014). LSD-SLAM: Large-scale direct monocular SLAM. Paper presented at the European conference on computer vision.

6- Forster, C., Pizzoli, M., & Scaramuzza, D. (2014). SVO: Fast semi-direct monocular visual odometry. Paper presented at the 2014 IEEE international conference on robotics and automation (ICRA).

7- Hosseininaveh, A., & Remondino, F. (2021). An Imaging Network Design for UGV-Based 3D Reconstruction of Buildings. *Remote Sensing*, 13(10), 1923.

8- Hosseininaveh, A., Serpico, M., Robson, S., Hess, M., Boehm, J., Pridden, I., & Amati, G. (2012). Automatic image selection in photogrammetric multi-view stereo methods.

9- Kerl, C., Sturm, J., & Cremers, D. (2013). Dense visual SLAM for RGB-D cameras. Paper presented at the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems.

10- Klein, G., & Murray, D. (2007). Parallel tracking and mapping for small AR workspaces. Paper presented at the 2007 6th IEEE and ACM international symposium on mixed and augmented reality.

11- Leutenegger, S., Lynen, S., Bosse, M., Siegwart, R., & Furgale, P. (2015). Keyframe-based visual-inertial odometry using nonlinear optimization. *The International Journal of Robotics Research*, 34(3), 314-334.

12- Lin, X., Wang, F., Guo, L., & Zhang, W. (2019). An automatic key-frame selection method for monocular visual odometry of ground vehicle. *IEEE Access*, 7, 70742-70754.

13- Lv, C., Li, J., & Tian, J. (2021). Key Frame Extraction for Sports Training Based on Improved Deep Learning. *Scientific Programming*, 2021.

14- Mur-Artal, R., Montiel, J. M. M., & Tardos, J. D. (2015). ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics*, 31(5), 1147-1163.

15- Mur-Artal, R., & Tardós, J. D. (2017). Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE Transactions on Robotics*, 33(5),

