مجله ماشین‌بینایی و پردازش‌تصویر
سال پنجم، شماره دوم، زمستان ۱۳۹۷

# تخمین مکان و زاویه دید دوربین در دنباله محدود تصاویر با استفاده از یک روش مبتنی بر فیلتر کالمن توسعه یافته

محمدامین مهرعلیان[۱] و محسن سریانی[۲]

## چکیده

**استفاده از دنباله تصاویر برای تخمین مکان و زاویه دید دوربین در کاربردهایی چون واقعیت افزوده و ناوبری ربات بسیار مورد توجه قرار گرفته است. در این مقاله از میان رویکردهای موجود برای این منظور یک رویکرد ترکیبی پیشنهاد شده است. ایده اصلی این رویکرد،استفاده از فیلتر کالمن توسعه یافته برای تخمین مسیر حرکت یک دوربین با ۶ درجه آزادی است. تفاوت اصلی الگوریتم پیشنهادی با سایر روش‌های مبتنی بر فیلتر این است که نقاط سه‌بعدی محیط حرکت دوربین از بردار حالت فیلتر حذف شده‌اند و با استفاده از روش‌های مبتنی بر هندسه چنددیدی با قطعیت مقداردهی می‌شوند. با این ایده حجم محاسبات فیلتر کالمن که یکی از نقاط ضعف روش‌های مشابه به شمار می‌آید، کاهش می‌یابد. در نهایت روش ارائه شده با یکی از دقیق‌ترین روش‌های مبتنی بر PnPمقایسه شده است و نتایج آزمایش‌ها نشان می‌دهد که دقت تخمین جابجایی و چرخش بهبود می‌یابد.**

## کلیدواژه‌ها

**بینایی ماشین سه‌بعدی، تخمین موقعیت دوربین، هندسه چنددیدی، فیلتر کالمن توسعه‌یافته**

## ۱ مقدمه

تخمین حرکت دوربین با ۶ درجه آزادی با استفاده از دنباله تصاویر یک مساله پرسابقه و چالش برانگیز در بینایی ماشین است. در سال‌های اخیر روش‌های متعددی برای این منظور پیشنهاد شده است که بیشتر در حوزه واقعیت افزوده (AR)[۱] و ناوبری ربات به خصوص درمکان‌یابی و نقشه‌برداری همزمان (SLAM)[۲] مورد استفاده قرار گرفته است.

بیشتر کارهای انجام شده در این زمینه با استخراج نقاط کلیدی و ردگیری آن‌ها در طول دنباله تصاویر، به صورت همزمان محیط سه‌بعدی و حرکت دوربین را تخمین می‌زنند[۲]. هر چند راهکارهای متنوعی در این مورد وجود دارد اما می‌توان بیشتر کارهای انجام شده را در قالب دو دسته اصلی تقسیم‌بندی کرد:

أ) **روش‌هایی مبتنی بر فیلتر:**در این دسته از تحقیقات، مساله به صورت یک فرآیند احتمالاتی و در قالب دنباله‌ای از حالات تعریف می‌شود. متغیر حالت شامل نقاط سه‌بعدی صحنه و پارامترهای دوربین است که در هر گام از فرآیند بر اساس احتمالات پیشین، ابتدا تخمینی از آن محاسبه می‌شود و سپس بر مبنای مشاهدات به هنگام‌سازی می‌گردد. برای این منظور معمولا از فیلتر کالمن توسعه یافته (EKF[۱])، فیلتر ذرات(PF[۲]) یا فیلتر کالمن بی‌بو(UKF[۳]) استفاده می‌شود.

ب) **روش‌های مبتنی بر هندسه چند دیدی[۴]:** این روش‌ها بر اساس روابط هندسی موجود میان نقاط ویژگی دوبعدی در دنباله

این مقاله در اردیبهشت‌ماه ۱۳۹۶ دریافت، در مهرماه بازنگری و در آبان‌ماه پذیرفته شد.

[۱] دانشجوی دکتری هوش مصنوعی و مهندسی کامپیوتر، دانشگاه علم و صنعت ایران
رایانامه: mehralian@comp.iust.ac.ir
[۲] گروه مهندسی کامپیوتر، دانشگاه علم و صنعت ایران
رایانامه: soryani@iust.ac.ir

نویسنده مسئول: محسن سریانی

[1]Extended Kalman Filter
[2]Particle Filter
[3]Unscented Kalman Filter
[4]Multiple-view Geometry

تصاویر و تناظر آن‌ها با نقاط سه‌بعدی صحنه، مکان و راستای دوربین را محاسبه می‌کنند.

دسته اول با وجود آنکه سابقه طولانی‌تری نسبت به دسته دوم دارند ولی در کارهای انجام شده در سال‌های اخیر محبوبیت کمتری پیدا کرده‌اند و بیشتر کارهای جدید در این زمینه به سمت دسته دوم رفته‌اند[۱]. ویژگی روش‌های مبتنی بر فیلتر این است که یک چارچوب احتمالاتی برای مساله ارائه می‌کنند. در این روش‌ها پارامترهای دوربین به صورت قطعی مشخص نمی‌شوند بلکه هر کدام بر اساس یک توزیع احتمالاتی با میانگین و واریانس معین بازنمایی می‌شوند.

روش‌های دسته اول معمولا پیچیدگی محاسباتی بیشتری در مقایسه با روش‌های دسته دوم دارند[۳]. روش‌های مبتنی بر هندسه چند دیدی در یک فضای غیراحتمالاتی (قطعی)، محاسبات دقیقی را بر پایه روابط هندسی موجود بین نقاط دوبعدی در دنباله تصاویر و نقاطسه‌بعدی در صحنه ارائه می‌دهند اما اثر خطا در این روش‌ها بسیار زیاد است تا حدی که بدون استفاده از راهکارهای حذف خطا مانند RANSAC[۱] [۴] و بهینه‌سازی نتایج با استفاده از تنظیم دسته‌ای (BA[۲])[۵] قابل اتکا نخواهند بود. علاوه بر این بدون احتساب احتمالات در محاسبات، مشخص نخواهد شد که در هر مرحله از اجرا چه میزان عدم قطعیت در تخمین پارامترهای دوربین وجود دارد.

در روش‌های مبتنی بر فیلتر مشکل پیچیدگی محاسبات و نیاز به حافظه زیاد از آنجا ناشی می‌شود که پارامترهای مربوط به دوربین و نقاط سه‌بعدی موجود در صحنه به صورت همزمان تخمین زده می‌شوند و این موضوع ابعاد بردار حالت و میزان محاسبات را افزایش چشمگیری می‌دهد. به همین دلیل نقاط سه‌بعدی موجود در صحنه معمولا به تعداد مشخصی محدود خواهند شد[۳].

دراین مقاله یک رویکرد ترکیبی پیشنهاد شده است که در آن ابتدا با استفاده از روش‌های مبتنی بر هندسه چندیدی، یک نقشه سه‌بعدی قطعی از صحنه با دقت قابل قبولی ایجاد می‌شود و در گام‌های بعدی با استفاده از یک فیلتر کالمن توسعه یافته، پارامترهای دوربین به صورت احتمالاتی پیش‌بینی و بر اساس نقشه سه‌بعدی اولیه و موقعیت نقاط دوبعدی در تصویر به‌هنگام می‌شوند. در این الگوریتم با حذف نقاط سه‌بعدی صحنه از بردار حالت، حجم محاسبات تخمین حالت در فیلتر به شکل قابل توجهی کاهش می‌یابد. همچنین با این کار پیچیدگی معادلات تخمین حالت بسیار کمتر خواهد شد که در نتیجه مشکل کاهش دقت تخمین حالت در EKFبا معادلات غیرخطی با درجه بالارا مرتفع خواهد کرد. لازم به ذکر است که هرچند در الگوریتم پیشنهادی از فیلتر کالمن توسعه یافته استفاده شده است اما به لحاظ مراحل اجرای الگوریتم، این کار بسیار نزدیک به کارهای مبتنی بر هندسه چندیدی است و به همین دلیل مقایسه دقت الگوریتم با روش‌های چندیدی انجام شده است.

در ادامه ابتدا مروری بر کارهای مبتنی بر فیلتر خواهد شد. سپس چند مقاله در زمینه روش‌های مبتنی بر هندسه چندیدی مورد بررسی قرار می‌گیرد. در بخش سوم روش پیشنهادی برای تخمین موقعیت دوربین با استفاده از EKF معرفی خواهد شد و نهایتا در بخش ۴ آزمایش‌های انجام شده برای مقایسه عمکرد الگوریتم پیشنهادی با الگوریتم‌های مبتنی بر هندسه‌ی چندیدی ارائه می‌گردند.

## ۲ مروری بر کارهای انجام شده

### ۲-۱ روش‌های مبتنی بر فیلتر

ایده استفاده از فیلتر کالمن برای تخمین موقعیت دوربین سابقه طولانی دارد. یکی از برجسته‌ترین این موارد مربوط به دیوسن و همکارانش در[۶]است که از EKF برای این منظور استفاده کرده‌اند و می‌توان گفت چارچوب کلیالگوریتم پیشنهادی نیز بر اساس همین کار توسعه داده شده است. در این روش متغیرهای مربوط به دوربین (شامل مکان و زوایا به همراه سرعت‌های خطی و زاویه‌ای) و نقاط سه‌بعدی صحنه (نقشه سه‌بعدی) به صورت همزمان در قالب یک بردار حالت تخمین زده می‌شوند. بدین صورت که ابتدا ویژگی‌های تصویر اول استخراج شده و موقعیتسه‌بعدی آن‌ها روی خطی که از مرکز دوربین به آن نقاط متصل شده‌اند با یک عدم قطعیت بالا در عمق ۰٫۵ تا ۵ متر با توزیع یکنواخت تعیین می‌شوند. سپس در هر فریم جدید با تکرار مشاهده ویژگی‌ها، میزان عدم قطعیت کاسته خواهد شد.علاوه بر این چنانچه ویژگی‌های جدیدی در تصاویر ورودی مشاهده شوند، این نقاط نیز با عدم قطعیت بالا به نقاط سه‌بعدی اضافه خواهند شد.

در روش‌های مبتنی بر فیلتر، کارهای دیگری نیز انجام شده است که از فیلتر کالمن استفاده نکرده‌اند. تحقیقات نشان داده است در شرایطی که مدل پویایی[۳] و مدل مشاهدات[۴] در فیلتر، غیرخطی با درجه بالا باشند، EKF نتایج قابل قبولی ارائه نمی‌کند. در این صورت فیلترهای غیرخطی دیگری مانند فیلتر ذرات و فیلتر کالمن بی‌بو جایگزین‌هایی برای EKF خواهند بود. هرچند در این دو فیلتر نیازی به بسط مشتق معادلات موجود در روابط فیلتر کالمن نیست و پیچیدگی کمتری در روابط وجود دارد اما با این حال حجم محاسبات در این روش‌ها بیشتر از EKF است و این مساله در کاربردهای بلادرنگ بسیار با اهمیت است.

یکی از اولین کارهای انجام شده با استفاده از فیلتر ذرات مربوط به پاپیلی و همکارانش در [۷] است. ویژگی این فیلتر، دقت بالاتر در مدل‌های غیرخطی با درجه بالا و در شرایط وجود خطا با

[1]RANdom SAmple Consensus
[2]Bundle Adjustment
[3]Dynamic Model
[4]Observation Model

توزیع غیر گاوسی است. آزمایش‌ها نشان می‌دهد فیلتر ذرات علاوه بر افزایش مقاومت الگوریتم، توانایی تخمین موقعیت دوربین در محیط‌های پویا را نیز دارد.

در[۸] اِید و درومُند، مدعی هستند که ماهیت غیرخطی با درجه بالای مدل مشاهدات دقت EKF را کاهش می‌دهد و به همین دلیل استفاده از فیلتر ذرات را برای تخمین همزمان حرکت دوربین و محیط سه‌بعدی پیشنهاد کرده‌اند. با این حال به نظر می‌رسد ایده اصلی مقاله مربوط به استفاده از ترکیب فیلتر ذرات با گراف حالت و بهینه‌سازی روی گراف است. همچنین بکارگیری معکوس عمق نقاط سه‌بعدی صحنه، به جای استفاده مستقیم از مقدار عمق، ایده دیگر این مقاله در ساخت مدل نقشه سه‌بعدی است. بنابراین برتری گزارش شده در دقت تخمین نیز ناشی از این ایده بوده و می‌توان گفت استفاده از فیلتر ذرات تاثیر چندانی در این برتری نداشته است. علاوه بر این، مقایسه اعداد گزارش شده برای زمان اجرا نیز نشان می‌دهد این الگوریتم سرعت پایین‌تری نسبت به الگوریتم ارائه شده در [۶] دارد.

همانطور که اشاره شد یکی از مشکلات موجود در روش‌های مبتنی بر فیلتر، حجم زیاد محاسبات و حافظه مورد نیاز است. از آنجا که متغیرهای مربوط به دوربین و نقشه سه‌بعدی به صورت همزمان محاسبه می‌شوند بنابراین در مجموع، متغیر حالت شامل N=C+3M بُعد است که در آن C متغیرهای لازم برای حالت دوربین (معمولا ۱۳) و مابقی، متغیرهای لازم برایMنقطه سه‌بعدی از صحنه است. مرتبه زمانی اجرای یک فیلتر کالمن توسعه یافته با متغیر حالت N بعدی در حالت کلی برابر با $O(N^3)$ است. با این حال در [۶] نویسندگان مدعی شده‌اند که برای کاربرد SLAM نیازمند محاسباتی از مرتبه $O(N^2)$ هستیم. به همین دلیل نیز در برای اجرای بلادرنگ الگوریتم (۳۰ فریم در ثانیه) تعداد ویژگی‌ها به ۱۰۰ ویژگی محدود شده است.

در کار دیگری هولمس و همکارانش [۹]،از یک نسخه تغییریافته UKF تحت عنوان SRUKF[1] استفاده کرده‌اند. این فیلتر در حالت استاندارد با مرتبه زمانی $O(N^3)$ اجرا می‌شود. اما نویسندگان مقاله گفته‌اند که با اعمال تغییراتی در شیوه محاسبات به خصوص در تخمین نقشه سه‌بعدی، مرتبه اجرایی را به $O(N^2)$ کاهش داده‌اند. بنابراین می‌توان گفت در حالتی که نقاط سه‌بعدی نقشه افزایش پیدا کنند این روش سرعتی به مراتب بیشتر از EKF خواهد داشت. با این حال به گفته نویسندگان این برتری در مورد دقت تخمین موقعیت دوربین وجود ندارد.

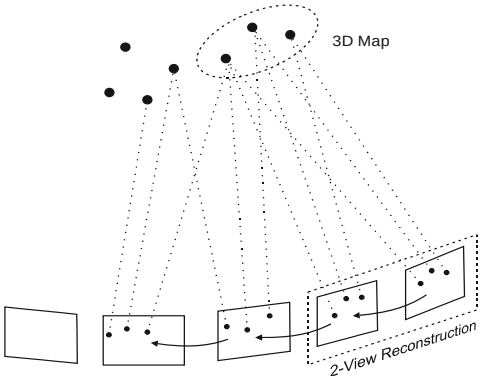
یکی از ویژگی‌های روش‌های مبتنی بر هندسه چندیدی این است که به وسیله الگوریتم RANSAC می‌توان نمونه‌های پرت[2] را شناسایی کرده و از فرآیند محاسبات حذف کرد. رویکرد مشابهی در روش‌های مبتنی بر فیلتر نیز وجود دارد. سیورا و همکارانش در [۱۰] و [۱۱] برای افزایش مقاومت الگوریتم مبتنی بر فیلتر کالمن توسعه یافته (EKF) در برابر نمونه‌های پرت، از یک روش شبه RANSAC استفاده کرده‌اند که 1-Point RANSAC نام گذاری شده است. در این روش علاوه بر افزایش مقاومت در برابر خطا، با حذف نمونه‌های پرت، حجم محاسبات نیز کاهش می‌یابد.

[1]Square Root Unscented Kalman Filter (SRUKF)
[2]Outlier

## ۲-۲ روش‌های مبتنی بر هندسه چند دیدی

روش‌های دیگری که برای تخمین مسیر حرکت دوربین مورد استفاده قرار می‌گیرد، بر اساس ارتباط هندسی میان نقاط متناظر در دو یا چند تصویر از یک صحنه شکل گرفته است، که به آن هندسه چندیدی گفته می‌شود [۱۲]. این دسته از روش‌ها بیشتر در حوزه استخراج ساختار از حرکت (SFM[3]) مطرح می‌شوند. این فرآیند در [۱۲] به ۳ مرحله تقسیم شده است: ۱-تشخیص ویژگی‌های متناظر در دنباله تصاویر ۲-محاسبه ساختار اولیه به عنوان نقطه شروع برای مرحله بعد ۳- تنظیم دسته‌ای ساختار (BA[4]).

استخراج ساختار از حرکت معمولا بر اساس خط‌مشی‌های متفاوتی انجام می‌گیرد که رویکرد افزایشی[5] بیشتر از سایر رویکردها مورد توجه قرار گرفته است. در این رویکرد تخمین ساختار و موقعیت دوربین از دنباله تصاویر معمولا با یک گام تشخیص و انطباق (یا ردگیری) ویژگی برای دو فریم متوالی آغاز می‌شود. بر اساس تناظر صورت گرفته میان دو فریم و با کمک روش مثلث‌بندی[6]، یک مدل سه‌بعدی اولیه از صحنه ایجاد می‌شود. البته باید توجه داشت که فاصله میان دو دوربین در این دو فریم به اندازه کافی بزرگ باشد تا نتایج محاسبات اپیپلار[7] قابل اتکا باشد. سپس به صورت افزایشی، موقعیت دوربین در فریم‌های جدید بر اساس تناظر نقاط سه‌بعدی صحنه و ویژگی‌های دوبعدی ردگیری شده در تصاویر جدید، محاسبه می‌شود. با دریافت تصاویر جدید، بخش‌های جدیدی از صحنه نیز قابل مشاهده خواهد بود. لذا توسعه مدل سه‌بعدی صحنه با استفاده از نقاط جدید و الگوریتم مثلث‌بندی ضروری است. شکل ۱ نمایی از مراحل گفته شده برای تخمین مسیر حرکت دوربین با استفاده از رویکرد افزایشی را نشان می‌دهد.




**شکل ۱ تخمین مسیر حرکت دوربین با استفاده از رویکرد افزایشی در هندسه چندیدی**

[3]Structure from Motion (SFM)
[4]Bundle Adjustment (BA)
[5]Incremental
[6]Triangulation
[7]Epipolar

در محاسبات هریک از مراحل گفته شده همواره خطایی وجود دارد که در صورت افزایش آن‌ها الگوریتم دچار انحراف زیادی در نتایج خواهد شد. تنظیم دسته‌ای یک بهینه‌سازی غیرخطی از موقعیت دوربین و مدل سه‌بعدی صحنه است که در آن خطای بازافکنش[1] کمینه می‌شود. به بیانی دیگر در کمینه‌سازی خطای بازافکنش تلاش می‌شود تا خطای ناشی از نگاشت نقاط سه‌بعدی صحنه به نقاط دوبعدی روی تصاویر با استفاده از پارامترهای دوربین به نحوی کمینه شود که کمترین فاصله را با ویژگی‌های ردگیری شده داشته باشد. روش‌هایی که برای تنظیم دسته‌ای بکار گرفته می‌شوند برای همگرا شدن معمولا نیازمند یک پاسخ اولیه با دقت قابل قبول برای شروع هستند. از طرفی مشکل اصلی این روش‌ها حجم بالای محاسبات و کندی در اجرا است. بنابراین بهتر است که تعداد اجرای آنرا تا حد امکان کاهش داد. در واقع تنظیم دسته‌ای به علت کندی در اجرا، برای بهبود بلادرنگ تخمین موقعیت دوربین استفاده نمی‌شود بلکه خطای تجمعی نقشه سه‌بعدی که در فریم‌های بعدی به عنوان معیار استفاده می‌گردد را کاهش می‌دهد.به طوری که در برخی از کارهای انجام شده در این زمینه مانند [۱۳] تنظیم دسته‌ای در یک پردازش مجزا و برای تعداد محدودی از فریم‌ها (فریم‌های کلیدی[2]) اعمال می‌شود.

همانطور که گفته شد موقعیت دوربین در فریم‌های جدید براساس نقاط سه‌بعدی صحنه و ویژگی‌های دوبعدی ردگیری شده در تصاویر جدید، محاسبه می‌شود. برای این کار روش‌های متفاوتی وجود دارد که یکی از این روش‌ها استفاده از تناظرهایسه‌بعدی به دوبعدی برای تخمین موقعیت دوربین است که معمولا بر اساس الگوریتم‌های مبتنی بر PnP[3] عمل می‌کنند[۲].
می‌کنند[۲].

ایده استفاده از هندسه چندیدی برای تخمین مسیر حرکت دوربین با استفاده از دنباله تصاویر، اولین بار توسط نیستر و همکارانش در [۱۴] و [۱۵] مطرح شد. وی این ایده را به منظور ناوبری یک ربات زمینی بکار گرفت و از عنوان مسافت‌سنج بصری (VO)[4] برای آن استفاده کرد.در [۱۴] از الگوریتم گفته شده برای تخمین مکان و راستای دوربین و همچنین نقشه سه‌بعدی استفاده شده است و برای این منظور ابتدا ویژگی‌ها در تعداد مشخصی از تصاویر ردگیری می‌شوند. سپس با استفاده از الگوریتم«۵نقطه»[5][۱۶] و به کمک RANSAC موقعیت نسبی دوربین در ۳ فریم از این دنباله محاسبه می‌شود. در ادامه با استفاده از مثلث‌بندی در فریم ابتدایی و انتهایی، موقعیتسه‌بعدی نقاط صحنه محاسبه می‌شود. از الگوریتم P3P[۱۷] برای تخمین موقعیت دوربین در فریم‌های جدید استفاده شده است و بعد از دریافت تعداد مشخصی از فریم‌ها مجددا نقاط سه‌بعدی جدیدی به کمک مثلث‌بندی به نقشه سه‌بعدی اضافه می‌شوند. با انتشار مقاله نیستر و همکارنش، الگوریتم ۵-نقطه که ایده اصلی کار آن‌ها بود در مقالات بسیاری مورد توجه قرار گرفت که از جمله آن‌ها می‌توان به [۱۸] و[۱۹] اشاره کرد که کارهای برجسته‌تری هستند.

کار انجام شده در [۲۰] نیز تقریبا از همین الگو پیروی می‌کند، علاوه بر این، ایده دیگری که در این مقاله مورد استفاده قرار گفته است مربوط به فریم‌های کلیدی است. با توجه به اینکه در فریم‌های جدید طبیعتا بخشی از ویژگی‌ها از صحنه خارج شده و تعداد ویژگی‌های ردگیری شده کاهش می‌یابند، نیازمند یک مرحله تشخیص ویژگی مجدد هستیم. بنابراین با کاهش یافتنویژگی‌ها به تعداد مشخصی (۴۰۰ ویژگی) یک فریم کلیدی تعریف خواهد شد که در آن نقاط جدید تشخیص داده می‌شوند و به نقشه سه‌بعدی اضافه خواهند شد. همچنین یک بهینه‌سازی محلی روی موقعیت دوربین‌ها و نقاط متناظر دوبعدی به سه‌بعدی انجام خواهد گرفت که تنظیم دسته‌ای محلی نامگذاری شده است.

## ۳ روش پیشنهادی

بررسی کارهای انجام شده در هریک از دو دسته روش نشان می‌دهد هر کدام دارای ضعف‌ها و نقاط قوتی هستند. در روش‌های مبتنی بر فیلتر استفاده از رویکرد احتمالاتی برای نقشه سه‌بعدی، علاوه بر افزایش حجم محاسبات باعث کاهش دقت آن نیز می‌شود. در مقابل استفاده از یک مدل احتمالاتی بر اساس موقعیت‌های قبلی دوربین برای تخمین پارامترهای آن در گام‌های بعدی بسیار منطقی به نظر می‌رسد.

نقطه قوت روش‌های مبتنی بر هندسه چند دیدی این است که با استفاده از روابط هندسی میان ویژگی‌ها می‌توانند یک تخمین دقیق و قطعی از نقشه سه‌بعدی ارائه کنند. در عین حال مشکل اصلی استفاده از الگوریتم‌های مبتنی بر PnP در دنباله تصاویر این است که هر مرحله از اجرا، موقعیت دوربین به صورت مستقل محاسبه می‌شود و از موقعیت‌های پیشین برای تخمین وضعیت جدید دوربین هیچ استفاده‌ای نمی‌شود. در واقع استفاده از اطلاعات توالی حرکت دوربین‌ها تا زمانی که با استفاده از تنظیم دسته‌ای (BA) یک بهینه‌سازی محلی یا سراسری بر روی توالی آن‌ها انجام گیرد، به تاخیر می‌افتد.

در روش پیشنهادی تلاش شده است الگوریتم مربوط به تخمین پارامترهای دوربین در روش‌های مبتنی بر هندسه چندیدی با یک فیلتر کالمن توسعه یافته جایگزین شود. در این صورت همان مراحل لازم برای اجرای الگوریتم که در بخش ۲-۲ تشریح گردید مورد استفاده قرار می‌گیرد با این تفاوت که به جای استفاده از الگوریتم‌های مبتنی بر PnP از فیلتر کالمن برای تخمین موقعیت دوربین استفاده می‌شود.

---
[1] Reprojection Error
[2] Key-frames
[3] Perspective-n-Point
[4] Visual Odometry (VO)
[5] five-point Algorithm

بر این اساس روش پیشنهادی شامل مراحل زیر خواهد بود:

أ) تخمین مدل سه‌بعدی صحنه به صورت قطعی و با استفاده از هندسه اپیپلار در دو فریم متوالی در ابتدای حرکت دوربین.

ب) مقدار دهی اولیه پارامترهای فیلتر کالمن توسعه یافته بر اساس پارامترهای تخمین زده شده برای دوربین در محاسبات اپیپلار.

ج) ردگیری ویژگی‌ها و تخمین موقعیت دوربین در فریم‌های بعدی با استفاده از فیلتر کالمن توسعه یافته.

در این مقاله فرض بر این استکه دوربین مورد استفاده از قبل واسنجی[1] شده است و از پارامترهای داخلی آن مطلع باشیم. باید توجه داشت که در این روش نیز مانند دیگر روش‌های مبتنی بر دوربین تک‌دید[2]، مسیر حرکت دوربین به صورت نسبی و با یک مقیاس نامعلوم تخمین زده می‌شود و معمولا برای تعیین مقیاس از اطلاعات دیگری مانند ابعاد اشیاء موجود در صحنه استفاده می‌شود. به عنوان نمونه در [۶] شروع حرکت دوربین از مقابل یک صفحه با ابعاد مشخص انجام می‌گیرد.

در ادامه فیلتر کالمن و چگونگی تطبیق روابط آن به منظور بکارگیری در تخمین پارامترهای حرکتی دوربین در دنباله تصاویر تشریح خواهد شد.

## ۳-۱ تخمین موقعیت دوربین با استفاده از فیلتر کالمن

فیلتر کالمن الگوریتمی است که در آن، بردار حالت یک فرآیند پویا به کمک مجموعه‌ای از مشاهدات تخمین زده می‌شود. این الگوریتم در دو گام پیش‌بینی و به‌هنگام‌سازی اجرا می‌شود. در گام پیش‌بینی با استفاده از تخمین‌های حالات در بازه‌های زمانی پیشین، تخمینی برای حالت فعلی بدست می‌آید. این تخمین پیش‌بینی شده همان دانش پیشین است زیرا تنها به تخمین‌های قبلی وابسته است و هیچ مشاهده‌ای در حالت فعلی فرآیند را دربر نمی‌گیرد. در گام به‌هنگام‌سازی، تخمین پیشین با مشاهدات فعلی ترکیب می‌شود تا تخمینی دقیق‌تری از حالت فعلی فرآیند ارائه کند.

از آنجا که در بیشتر مسائل دنیای واقعی پویایی‌هاییک فرآیند به صورت غیرخطی تغییر می‌کنند، از نسخه غیرخطی فیلتر کالمن یعنی فیلتر کالمن توسعه یافته(EKF) استفاده می‌شود. بنابراین برای تعریف روابط گام پیش‌بینی بر اساس مدل غیرخطی پویایی فرآیند، خواهیم داشت:

$$s_{k|k-1} = f(s_{k-1|k-1}, \mathbf{u}_k) \qquad (۱)$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^{\mathrm{T}} + \mathbf{G_k Q_k G_k^T} \qquad (۲)$$

رابطه (۱) معادله تخمین حالت پیش‌بینی شده براساس ورودی متغیر حالت پیشین ($s_{k-1|k-1}$) و ورودی مستقل ($\mathbf{u}_k$) است. رابطه (۲) معادله کوواریانس تخمین حالت پیش‌بینی شده است که در آن $\mathbf{F}_k$ مشتق مدل پویایی فرآیند نسبت به متغیر حالت و $\mathbf{G_k^T}$مشتق مدل پویایی نسبت به ورودی $\mathbf{u_k}$است.

در گام به‌هنگام‌سازی می‌بایست بر اساس مشاهدات و حالت تخمینی، یک به‌هنگام‌سازی در پارامترهای حالت انجام دهیم. در این صورت برای روابط **گام به‌هنگام‌سازی** خواهیم داشت:

$$s_{k|k} = s_{k|k-1} + \mathbf{K}_{k|k-1}\left(z_k - \mathbf{h}(s_{k|k-1})\right) \qquad (۳)$$

$$\mathbf{P}_{k|k} = \left(I - \mathbf{K}\mathbf{H}_{k|k-1}\right)\mathbf{P}_{k|k-1} \qquad (۴)$$

$$\mathbf{K}_{k|k-1} = \mathbf{P}_{k|k-1}\mathbf{H}_{k|k-1}^T\left(\mathbf{H}_{k|k-1}\mathbf{P}_{k|k-1}\mathbf{H}_{k|k-1}^T + \mathbf{R}_k\right)^{-1} \qquad (۵)$$

$\mathbf{h}(s_{k|k-1})$ تابع مدل مشاهدات بر اساس حالت تخمینی در گام پیش‌بینی است و $\mathbf{H}_{k|k-1}$ نیز مشتق آن نسبت به متغیر حالت می‌باشد. $z_k$ متغیر مشاهدات فرآیند است و هر چه میزان اختلاف آن با خروجی مدل مشاهدات کمتر باشد، تخمین دقیق‌تری در پیش‌بینی قبل انجام شده است. پارامتر $\mathbf{K}_{k|k-1}$ که به آن نرخ بهره می‌گویند مشخص کننده میزان تاثیرپذیری بردار حالت، از اختلاف میان مشاهدات و خروجی مدل مشاهدات است. به صورت مشابه در روابط (۴) و (۵)مقدار کواریانس بردار حالت و نرخ بهره به‌هنگام می‌شوند. ماتریس $\mathbf{R}_k$ نیز کواریانس نویز مشاهدات در به‌هنگام‌سازی نرخ بهره است.

## ۳-۲ مدل حرکت دوربین

در حالتی که دوربین حرکتی با ۶ درجه آزادی داشته باشد از تبدیل اقلیدسی برای نمایش پارامترهای آن در فضای سه‌بعدیاستفاده می‌شودکه در آن t شامل جابجایی در راستای $x$، $y$ و $z$ است و $R$ نیز ماتریس دوران است که یک ماتریس متعامد بوده و می‌دانیم $R^T = R^{-1}$ و $|R| = 1$.

در اینجا، مشابه کار انجام شده در [۶] برای نشان دادن زاویه در فضای سه‌بعدی بجای ماتریس دوران از بازنمایی چهارگان استفاده شده است. استفاده از چهارگان دارای مزایای بسیاری است که برخی از آن‌ها عبارتند از: ۱-برای نمایش چرخش به جای ۹ پارامتر از ۴ پارامتر استفاده می‌کند ۲-محدودیت‌های ماتریس چرخش ($R^T = R^{-1}$ و $|R| = 1$) در آن به سادگی حفظ می‌شود. ۳- ترکیب چند چرخش در بازنمایی چهارگان با محاسبات کمتری قابل اعمال است. ۴- معکوس یک چرخش در بازنمایی چهارگان به راحتی و با منفی کردن ۳ مؤلفه از آن امکان پذیر است.

برای محاسبه حرکت دوربین از دنباله تصاویر، متغیر حالتی شامل جابجایی و چرخش دوربین نسبت به مبدا مختصات جهانی تعریف می‌شود که فیلتر کالمن در هر مرحله تخمینی از آن را محاسبه می‌کند. این متغیر حالت به صورت یک بردار ۱۳ بعدی شامل جابجایی و چرخش در سه بُعد و سرعت‌های خطی و زاویه‌ای تعریف می‌شود:

$$s_{13\times1} = [d_{3\times1} q_{4\times1} v_{3\times1} \omega_{3\times1}]^T \qquad (۶)$$

[1] Calibration
[2] Monocular

که در آن $d$ بردار جابجایی، $q$ بردار چرخش در بازنمایی چهارگان و $v$ و $\omega$ به ترتیب سرعت‌های خطی و زاویه‌ای هستند.در مدل حرکت دوربین فرض می‌شود سرعت جابجایی و چرخش دوربین ثابت باشد و شتاب نیز به صورت یک ورودی مستقل(بردار $\mathbf{u}_k$ در رابطه (۱)) بعنوان نویز گاوسی با میانگین صفر، $n = [\mathrm{a}\ \alpha]^T$ مدل می‌شود. بنابراین مدل حرکت دوربین به صورت زیر تعریف می‌شود:

(۷)

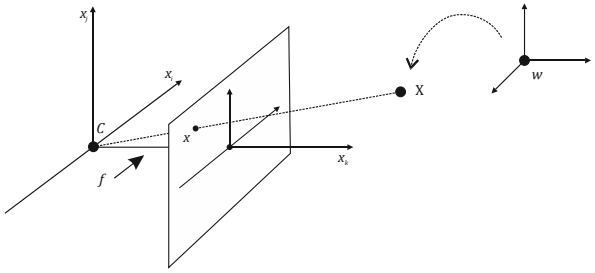
$$f(s,n)_{13\times1} = \begin{pmatrix} d_{k+1} \\ q_{k+1} \\ v_{k+1} \\ \omega_{k+1} \end{pmatrix} = \begin{pmatrix} d_k + (v_k + V_k)\Delta t \\ q_k * q\{(\omega + \Omega)\Delta t\} \\ v_k + V \\ \omega_k + \Omega \end{pmatrix}$$

که در آن $V = \mathrm{a}\Delta t$ ، $\Omega = \alpha\,\Delta t$ و $q\{(\omega + \Omega)\Delta t\}$ معادل چهارگان برای زاویه $(\omega + \Omega)\Delta t$ است. باید توجه داشت که علامت $*$ به معنای ضرب چهارگان است.

برای محاسبه کواریانس تخمین حالت از رابطه (۲)، می‌بایست مشتقات مدل حرکت دوربین نسبت به متغیر حالت و متغیر نویز محاسبه شود که در این صورت نیازمند محاسبه ژاکوبین $\frac{\partial f}{\partial s}$ و $\frac{\partial f}{\partial n}$ هستیم اما با توجه به حجم روابط در این مقاله به آن اشاره نمی‌شود.

## ۳-۳ مدل مشاهدات

ویژگی‌های دوبعدی موجود در دنباله تصاویر، نقش مشاهدات را در فیلتر بر عهده خواهند داشت. به همین دلیل برای فهم مدل مشاهدات بر اساس پارامترهای دوربین و نقاط سه‌بعدی صحنه، چگونگی افکنش محیط سه‌بعدی به تصویر دوبعدی تشریح می‌شود. افکنش نقاط از فضای سه‌بعدی در مختصات جهانی به تصویر دوبعدی در دو گام انجام می‌شود. همانطور که در شکل ۲ نیز نشان داده شده است، ابتدا نقاط سه‌بعدی از مختصات مرجع جهانی به مختصات دوربین منتقل شده و سپس با استفاده از مدل دوربین‌های روزنه‌ای[1] به صفحه تصویر نگاشت می‌گردند.




**شکل ۲افکنش نقشه سه‌بعدی از مختصات جهانی به صفحه تصویر**

برای انتقال نقطه سه‌بعدی $X = [X_i, X_j, X_k]$به مختصات دوربین خواهیم داشت:

$$X^c = R(X^w - C) \qquad (۸)$$

که در آن C موقعیت مرکز دوربین در مختصات جهانی و $R$ مقدار دوران آن است. این معادله با فرض استفاده از بازنمایی چهارگان برای چرخش به صورت زیر بازنویسی می‌شود:

$$[0, X^c] = q[0, X^w - C]q^* \qquad (۹)$$

که در آن $q$ بازنمایی چهارگان از ماتریس چرخش $R$ است. همانطور که در رابطه بالا دیده می‌شود برای استفاده از بازنمایی چهارگان می‌بایست ابعاد نقاط سه‌بعدیاز ۳ به ۴ افزایش یابند که به همین جهت هر نقطه سه‌بعدی با اضافه شدن یک بُعد با مقدار صفر در ابتدای آن به ۴ بعد تبدیل می‌شود (در واقع یک چهارگان با بخش موهومی صفر و بخش حقیقی $[X_i, X_j, X_k]$ ایجاد می‌شود.)

برای نگاشت نقطه سه‌بعدی$X^c$در مختصات دوربین به نقطه $x = [x_i, x_j]$ در تصویر نیز از رابطه زیر استفاده می‌شود:

$$x = \left[f.\frac{X_i}{X_k} + c_i, f.\frac{X_j}{X_k} + c_j\right] \qquad (۱۰)$$

که در آن $f$ فاصله کانونی دوربین و $(c_i, c_j)$ مرکز صفحه تصویر است. در اینجا فرض می‌شود دوربین مورد استفاده از قبل واسنجی شده و پارامترهای داخلی آن معلوم است.

برای n ویژگی موجود در تصاویر، بردار مشاهدات به صورت زیر خواهد بود:

$$z_{2n\times1} = \left[x_i^1, x_j^1, x_i^2, \dots, x_i^n, x_j^n\right]^t \qquad (۱۱)$$

طبیعتا مدلی که مشاهدات دوبعدی را از یک صحنه سه‌بعدی ایجاد خواهد کرد، رابطه افکنش سه‌بعدی به دوبعدی دوربین است. در این مدل، نقشه سه‌بعدی بر اساس متغیرهای دوربین (متغیر حالت) به یک تصویر دوبعدی نگاشت خواهد شد که در صورت دقیق بودن تخمین‌ها، باید با مشاهدات (ویژگی‌های دوبعدی ردگیری شده) منطبق باشد. لذا برای معادله $h(s)$ در رابطه (۳)در ۲ گام خواهیم داشت:

$$\hat{h}_{n\times4} = \begin{pmatrix} q(0, X^1 - d)q^* \\ q(0, X^2 - d)q^* \\ \vdots \\ q(0, X^n - d)q^* \end{pmatrix} \qquad (۱۲)$$

که در آن $d$ همان فاصله مرکز دوربین از مبدا مختصات جهانی است. در گام دوم نیز خواهیم داشت:

$$h_{n\times2} = \begin{pmatrix} f\frac{\hat{h}_i^1}{\hat{h}_k^1} + c_i & f\frac{\hat{h}_j^1}{\hat{h}_k^1} + c_j \\ \vdots & \\ f\frac{\hat{h}_i^n}{\hat{h}_k^n} + c_i & f\frac{\hat{h}_j^n}{\hat{h}_k^n} + c_j \end{pmatrix} \qquad (۱۳)$$

برای به هنگام‌سازی کواریانس تخمین حالت و نرخ بهره با استفاده از روابط (۴) و (۵) می‌بایست مشتق این معادلات نیز محاسبه شوند که به علت پیچیدگی‌های موجود در آن‌ها، از بازنویسی روابط خودداری می‌شود.

[1] Pinhole Camera

## ۴ آزمایش‌ها

به منظور ارزیابی روش پیشنهادی لازم است مقایسه‌ای میان این روش و یکی از دو دسته روش‌های مبتنی بر فیلتر یا مبتنی بر هندسه چنددیدی انجام گیرد. هرچند در الگوریتم پیشنهادی از فیلتر کالمن توسعه یافته استفاده شده است اما به لحاظ مراحل اجرای الگوریتم، این کار بسیار نزدیک به کارهای مبتنی بر هندسه چنددیدی است. در هر دو مورد ابتدا می‌بایست یک بازسازی سه‌بعدی از صحنه انجام گیرد و در تصاویر بعدی با تناظر نقاط سه‌بعدی و ویژگی‌های دوبعدی تصویر، موقعیت جدید دوربین تخمین زده شود. در حالی که در روش‌های مبتنی بر فیلتر نقشه سه‌بعدی اولیه قطعی وجود ندارد و با گذشت زمان به صورت احتمالاتی ایجاد می‌شود. بنابراین در آزمایش‌های انجام شده از روش‌های مبتنی بر هندسه چنددیدی برای تخمین موقعیت دوربین استفاده شده است که معمولا بر اساس الگوریتم‌های مبتنی بر PnP عمل می‌کنند.

برای مقایسه الگوریتم‌های مبتنی بر PnP و روش پیشنهاد شده در این مقاله، از EPnP[1] [۲۱] که یکی از سریع‌ترین و دقیق‌ترین روش‌های مبتنی بر PnP است و پیاده‌سازی آن نیز در دسترس عموم قرار گرفته، استفاده شده. در ادامه دو آزمایش مجزا، یکی برای مقایسه میزان دقت هریک از دو الگوریتم و دیگری برای ارزیابی مقاومت آن‌ها در برابر نویز ارایه می‌شود.

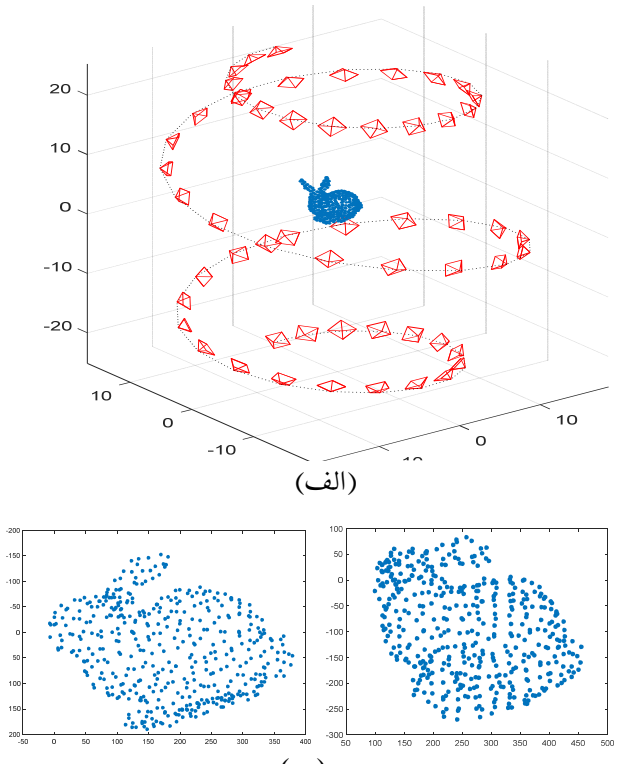
برای اندازه‌گیری دقت هریک از روش‌ها، خطای جابجایی و چرخش به صورت جداگانه گزارش شده است. در جابجایی از نسبت اختلاف فاصله اقلیدسی تا درستی مرجع بر کل جابجایی و برای خطای چرخش از نُرم اختلاف ۳ زاویه roll، pitch، yaw با درستی مرجع استفاده شده است.

### ۴-۱ داده‌های مصنوعی

از آنجا که فرآیند تخمین پارامترهای دوربین در یک دنباله متوالی از تصاویر، شامل مراحل دیگری چون تشخیص و ردگیری ویژگی‌ها است، در اینآزمایش‌ها برای بررسی دقت الگوریتم پیشنهادی، از داده‌های مصنوعی تولید شده توسط نویسندگان استفاده شده است. استفاده ازداده‌های مصنوعی تولیدشده این امکان را فراهم می‌کند تا بدون اثرپذیری از خطای سایر مراحل، الگوریتم به صورت مستقل ارزیابی شود. در واقع با این روش برای جلوگیری از انتشار خطا فرض شده است که یک بازسازی اولیه از دو تصویر اول انجام گرفته است تا تنها اثر خطای تخمین مکان و راستای دوربین مورد بررسی قرار گیرد.

علاوه بر این، دنباله حرکت دوربین با مسیر نسبتا پیچیده و به نحوی طرح ریزی شده است که نقاط ویژگی همواره در تصاویر حاضر باشند و با گذشت زمان تعداد ویژگی‌ها کاهش پیدا نکند. زیرا این مساله نیز در دقت الگوریتم‌ها مؤثر خواهد بود.

شکل ۳(الف) داده‌های مصنوعی ایجاد شده برای تصویربرداری توسط دوربین در ۵۵ موقعیت مختلف از یک شیسه‌بعدی را نشان می‌دهد. برای این منظور از یک دوربین مجازی با وضوح ۶۴۰×۴۸۰ پیکسل و فاصله کانونی ۲۴ میلیمتر استفاده شده است. سپس دوربین در یک مسیر مشخص حرکت کرده و تصویربرداری از شیسه‌بعدی را در ۵۵ موقعیت انجام می‌دهد. حرکت دوربین به نحوی است که در هر گام، تمامی ۶ درجه آزادی میان دو موقعیت متوالی تغییر کرده باشد.همچنین فاصله دوربین از نقاط ۳بعدی ابتدا به صورت کاهشی و در حال نزدیک شدن به آن‌ها است و سپس روند افزایشی پیدا می‌کند و از آن‌ها دور می‌شود. شیسه‌بعدی مورد تصویربرداری نیز از اسکن سه‌بعدی یک مجسمه خرگوش با ۴۵۳ نقطه ویژگی حاصل شده است. براساس پارامترهای داخلی و خارجی دوربین یک افکنش سه‌بعدی به دوبعدی انجام می‌شود و نقاط دوبعدی به عنوان ویژگی‌های تصویر انتخاب می‌شوند. شکل ۳ (ب) دو نمونه از تصاویر گرفته شده از دوربین را نشان می‌دهد.




**شکل ۳(الف) نمایی از داده‌های مصنوعی استفاده شده برای انجام آزمایش‌ها (ب) افکنش نقاط سه‌بعدی به تصویر در دو موقعیت مختلف از دوربین.**

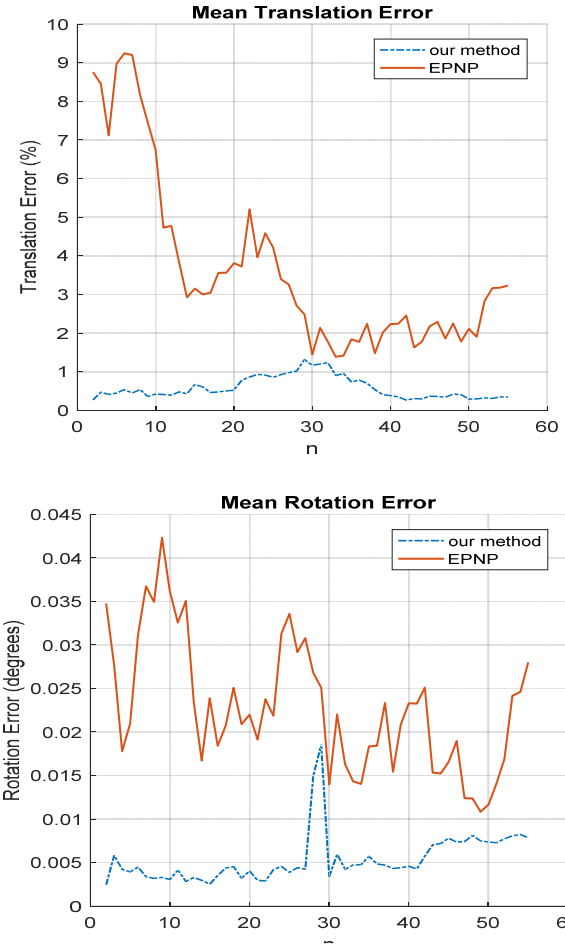
با توجه به اینکه مسیر حرکت دوربین در روش‌های مبتنی بر دوربین تک‌دید به صورت نسبی تخمین زده می‌شود، برای مقایسه این مسیر با مسیر درستی مرجع[2]، اطلاعات مرحله اول الگوریتم یعنی تخمین مدل سه‌بعدی و موقعیت دوربین در دو فریم اول از مرجع درستی استخراج شده است که باعث رفع ابهام در مقیاس می‌گردد.

[1] Efficient PnP

[2] Ground Truth

### ۴-۱-۱- مقایسه خطای دو روش

مهمترین معیار ارزیابی روش پیشنهادی در مقابلروش‌های مبتنی بر PnP، مقایسه خطای دو روش در تخمین مکان و زاویه دید دوربین است. در این آزمایش، هر دو الگوریتم در شرایط وجود نویز یکسان در مکان نقاط دوبعدی، مورد بررسی قرار گرفته‌اند و برای این منظور از نویز گاوسی استاندارد با میانگین۳ پیکسل استفاده شده است.




**شکل ۴ خطای تخمین موقعیت دوربین در روش پیشنهادی و EPnP به ازای نویز گاوسی با میانگین ۳ پیکسل در مکان نقاط دوبعدی**

شکل ۴ خطای جابجایی و چرخش را برای هر یک از ۵۵ موقعیت دوربین در طول زمان نشان می‌دهد. این خطا از میانگین ۱۰ مرتبه اجرا بدست آمده است. همانطور که در نتایج دیده می‌شود خطای روش EPnP در همه موارد بیش از خطای روش پیشنهادی است.

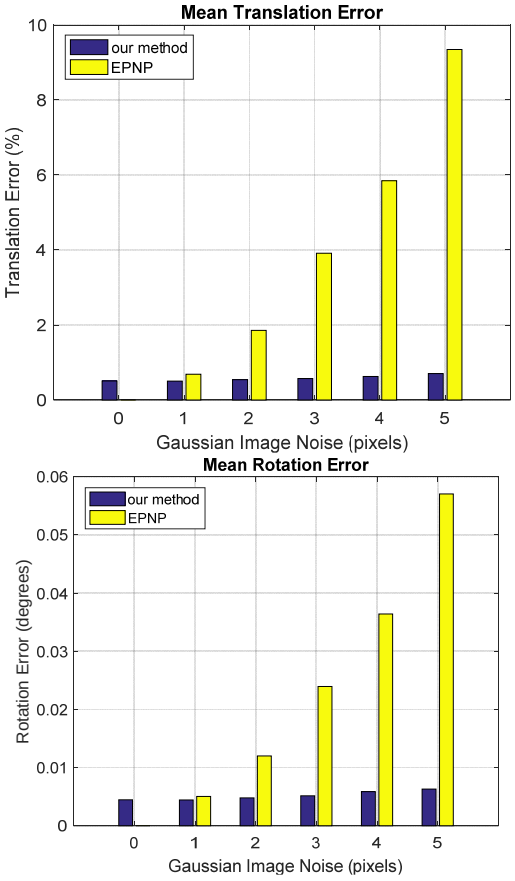
همانطور که در نمودار خطای جابجایی دیده می‌شود، روش EPnP در ابتدا و انتهای دنباله حرکت دوربین خطای بیشتری داشته است. از آنجا که در ابتدا و انتهای مسیر مورد آزمایش، فاصله دوربین از نقاط ۳بعدی به حداکثر مقدار خود می‌رسد، نتایج نشان داده است که الگوریتم EPnPدر مقایسه با الگوریتم پیشنهادی، نسبت به فاصله از نقاط سه‌بعدی حساسیت بیشتری دارد. البته در مورد خطای مربوط زوایا به نظر می‌رسد یکنواختی بیشتری در نتایج وجود دارد.

وجود قله در نمودار مربوط به خطای جابجایی در روش پیشنهادی ناشی از فرض سرعت ثابت در فیلتر کالمن است زیرا در قسمت میانه دنباله، سرعت حرکت دوربین بیشترین تغییرات را دارد و این موضوع اثرش را در نتایج نشان داده است.

### ۴-۱-۲- تاثیرپذیری از میزان نویز

آزمایش دیگری که تحلیل آن بسیار با اهمیت خواهد بود، مربوط به تاثیرپذیری دو روش از میزان نویز در مکان نقاط دوبعدی است. این نویز می‌تواند حاصل عوامل متعددی از جمله خطاهای بسیار کوچک ناشی از رقمی‌سازی تصویر تا خطای ناشی از ردگیری نقاط در دنباله تصاویر باشد.

در این آزمایش به ازای مقادیر متفاوتی از نویز، میانگین خطای جابجایی و چرخش برای همه موقعیت‌های دوربین محاسبه شده است. علاوه بر این برای هر مقدار از نویز، آزمایش ۱۰ مرتبه تکرار شده است و میانگین آن‌ها گزارش شده است.




**شکل ۵ بررسی تاثیر پذیری روش پیشنهادی و EPnP از میزان نویز در مکان نقاط دوبعدی**

همانطور که در نتایج شکل ۵مشخص است روش EPnP نسبت به نویز بسیار حساس است در صورتی که روش پیشنهادی کمتر تحت تاثیر نویز قرار گرفته است. علاوه بر این باید توجه داشت که در شرایطی که مکان نقاط دوبعدی با دقت بالا مشخص شده‌اند (نویز گاوسی با میانگین صفر) خطای روش EPnPنزدیک به صفر است و این موضوع ناشی از قطعیت بالا در مکان نقاط دوبعدی است که باعث دقت در نتایج محاسبات عددی می‌شود.

## ۴-۲ داده‌های واقعی

این آزمایش برای ارزیابی عملکرد الگوریتم پیشنهادی برای داده‌های واقعی انجام گرفته است. برای این منظور از داده‌های گردآوری شده در [۲۲] استفاده شده است که چند نمونه از تصاویر آن در شکل ۶ آمده است.

**شکل ۶ نمونه از تصاویر مورد استفاده**

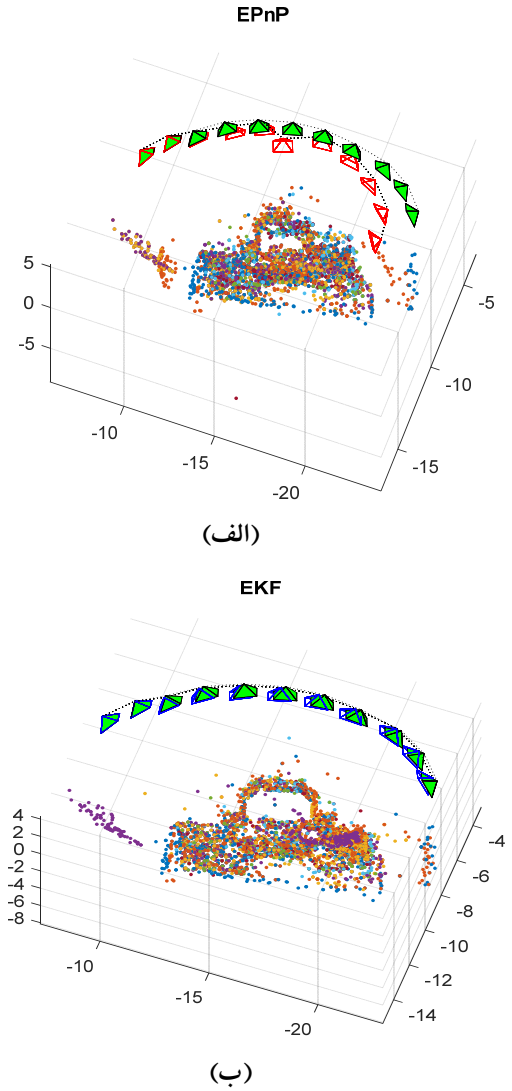
مراحل انجام آزمایش بدین صورت است که ابتدا ویژگی‌ها[1] و توصیفگرهای[2] SIFT[3][۲۳] در هر تصویر استخراج شده و سپس تناظریابی میان آن‌ها انجام می‌شود. تناظریابی در دو گام انجام می‌شود: ۱- استفاده از روش پیشنهادی در [۲۳] که بر اساس نسبت «اولین نزدیک‌ترین همسایه» به «دومین نزدیک‌ترین همسایه» است. در این روش دو ویژگی هنگامی با یکدیگر متناظر خواهند بود که علاوه بر نزدیکی مقادیر توصیفگرهایشان، با دومین نزدیکترین همسایه فاصله معناداری داشته باشد. به همین دلیل از نسبت «اولین نزدیک‌ترین همسایه» به «دومین نزدیک‌ترین همسایه» استفاده می‌شود. ۲- استفاده از روش RANSAC برای حذف نقاط پرت.

به منظور مقایسه نتایج آزمایش با درستی مرجع نیاز به محاسبه مسیر حرکت دوربین با مقیاس صحیح است. به این دلیل موقعیت دوربین در دو فریم اول از درستی مرجع استخراج شده است که باعث رفع ابهام در مقیاس می‌گردد. سپس نقاط سه‌بعدی با استفاده از الگوریتم مثلث‌بندی ایجاد می‌شوند.

در هر فریم جدید ویژگی‌هایی که در دو فریم قبل دیده شده‌اند و معادل سه‌بعدی آن‌ها نیز در مدل صحنه وجود دارد برای تخمین موقعیت دوربین استفاده می‌شود. در ادامه ویژگی‌هایی از فریم جدید که در مدل صحنه وجود ندارند با موقعیت جدید دوربین و به کمک مثلث‌بندی به صحنه سه‌بعدی اضافه می‌شوند. طبیعتا با گذشت زمان و افزایش خطا در تخمین موقعیت دوربین مدل صحنه سه‌بعدی نیز دچار خطا خواهد شد. این روال برای هر فریم جدید تکرار می‌شود.

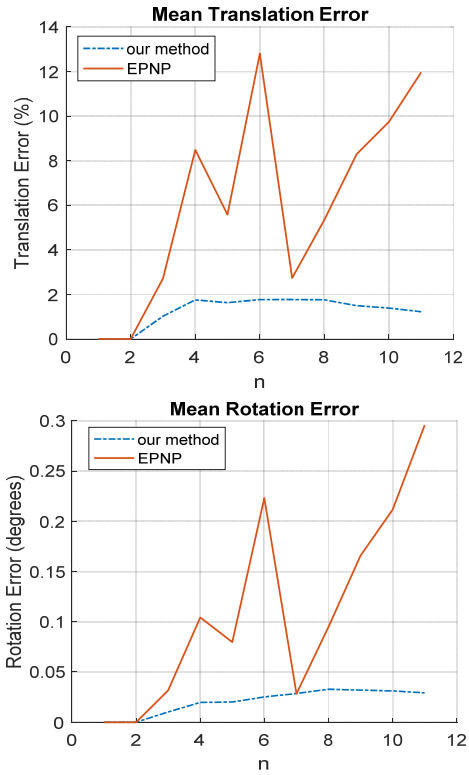
شکل ۷ نتایج حاصل از تخمین مسیر حرکت دوربین و مدل سه‌بعدی صحنه را نشان می‌دهد. در این شکل موقعیت دوربین در درستی مرجع با استفاده از یک هرم توپر (سبز رنگ) و نتیجه اجرای الگوریتم با یک هرم تو خالی مشخص شده است. با توجه به اینکه با افزایش خطا در تخمین پارامترهای دوربین بعضی از نقاط سه‌بعدی دچار انحراف زیادی می‌شوند، برای نمایش بهتر خروجی، برخی از این نقاط حذف شده‌اند تا حرکت دوربین بهتر دیده شود.

مقادیر خطای جابجایی و چرخش برای هر دو روش در شکل ۸ آمده است. در مسائل واقعی معمولا موقعیت دوربین‌ها و نقاط سه‌بعدی به عنوان ورودی برای الگوریتم تنظیم دسته‌ای (BA) مورد استفاده قرار می‌گیرد بنابراین هر چه به مقادیر واقعی نزدیک‌تر باشد، نتایج تنظیم دسته‌ای نیز بهتر خواهد بود.همانطور که در نتایج دیده می‌شود روش پیشنهادی اثرپذیری کمتری از خطا داشته است و انتشار خطا نیز با سرعت کمتری در آن رخ داده است.




**شکل ۷ نتایج آزمایش بر اساس دادگان واقعی. مسیر حرکت و مدل سه‌بعدی صحنه با استفاده از (الف) روش EPnP (ب) روش پیشنهادی.**

[1] Features (keypoints)
[2] Descriptors
[3] Scale Invariant Feature Transform (SIFT)




**شکل ۸مقایسه خطای جابجایی و چرخش در هریک از دو روش نسبت به درستی مرجع**

## ۵ جمع‌بندی

در این مقاله یک رویکرد جدید به منظور تخمین حرکت دوربین با ۶ درجه آزادی ارائه گردید. برای این منظور با استفاده از روش‌های مبتنی بر هندسه چندديدی ابتدا یک بازسازی اولیه از صحنه ایجاد می‌شود. سپس با ردگیری نقاط ویژگی در دنباله تصاویر و با اطلاع از موقعیت سه‌بعدی ویژگی‌ها در محیط، تخمینی از مکان و زاویه دید دوربین محاسبه می‌شود. این تخمین بر خلاف روش‌های مبتنی بر هندسه چند دیدی که از روابط ریاضی قطعی استفاده می‌کنند، مبتنی بر یک چارچوب احتمالاتی است که بر پایه فیلتر کالمن توسعه یافته ایجاد شده است. استفاده از احتمالات در محاسبات حرکت دوربین سبب می‌شود در شرایط واقعی و حضور نویز در محاسبات، انحراف کمتری در نتایج ایجاد شود. علاوه بر این در هر گام عدم قطعیت در پارامترها به صورت عددی قابل نمایش است.

آزمایش‌های انجام شده نشان می‌دهد که الگوریتم ترکیبی پیشنهادی در مقایسه با روش‌هایی که منحصرا مبتنی بر هندسه چندديدی هستند مقاومت بهتری نسبت به نویز دارد. البته در آینده لازم است دامنه این کار از دنباله‌های محدود به فضای بزرگتری گسترش یابد تا ارزیابی بهتری از عملکرد الگوریتم‌ها انجام گیرد.

## مراجع

[1] E. Marchand, H. Uchiyama and F. Spindler, "Pose Estimation for Augmented Reality: A Hands-On Survey," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 12, pp. 2633-2651, Dec 2016.

[2] D. Scaramuzza and F. Fraundorfer, "Visual Odometry [Tutorial]," *IEEE Robotics \& Automation Magazine*, vol. 18, no. 4, pp. 80-92, dec 2011.

[3] H. Strasdat, J. Montiel and A. J. Davison, "Visual SLAM: Why filter?," *Image and Vision Computing*, vol. 30, no. 2, pp. 65-77, 2012.

[4] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381-395, #jun# 1981.

[5] B. a. M. P. F. a. H. R. I. a. F. A. W. Triggs, "Bundle Adjustment --- A Modern Synthesis," in *Vision Algorithms: Theory and Practice: International Workshop on Vision Algorithms Corfu, Greece, September 21--22, 1999 Proceedings*, B. a. Z. A. a. S. R. Triggs, Ed., Berlin, Heidelberg, Springer Berlin Heidelberg, 2000, pp. 298-372.

[6] A. Davison, I. Reid, N. Molton and O. Stasse, "MonoSLAM: Real-Time Single Camera SLAM," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1052-1067, June 2007.

[7] M. L. Pupilli and A. D. Calway, "Real-Time Camera Tracking Using a Particle Filter," in *Proceedings of the British Machine Vision Conference*, 2005.

[8] E. Eade and T. Drummond, "Monocular SLAM as a Graph of Coalesced Observations," in *2007 IEEE 11th International Conference on Computer Vision*, 2007.

[9] S. A. Holmes, G. Klein and D. W. Murray, "An O(N2) Square Root Unscented Kalman Filter for Visual Simultaneous Localization and Mapping," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 7, pp. 1251-1263, July 2009.

[10] J. Civera, O. G. Grasa, A. J. Davison and J. Montiel, "1-point RANSAC for EKF-based structure from motion," in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, 2009.

[11] J. Civera, O. G. Grasa, A. J. Davison and J. M. M. Montiel, "1-Pointt RANSAC for Extended Kalman Filtering: Application to Real-time Structure from Motion and Visual Odometry," *J. Field Robot.*, vol. 27, no. 5, pp. 609-631, #sep# 2010.

[12] R. Hartley, Multiple view geometry in computer vision, 2nd ed., Cambridge, UK ; New York: Cambridge University Press, 2003.

[13] G. Klein and D. Murray, "Parallel Tracking and Mapping for Small AR Workspaces," in *6th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, 2007.

[14] D. Nister, O. Naroditsky and J. Bergen, "Visual odometry," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, 2004.

[15] D. Nistér, O. Naroditsky and J. Bergen, "Visual odometry for ground vehicle applications," *Journal of Field Robotics*, vol. 23, no. 1, pp. 3-20, 2006.

[16] D. Nister, "An efficient solution to the five-point relative pose problem," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 6, pp. 756-770, June 2004.

[17] B. M. a. L. C.-N. a. O. K. a. N. M. Haralick, "Review and analysis of solutions of the three point perspective pose estimation problem," *International Journal of Computer Vision*, vol. 13, no. 3, pp. 331-356, 1994.

[18] M. Lhuillier, "Automatic scene structure and camera motion using a catadioptric system," *Computer Vision and Image Understanding*, vol. 109, no. 2, pp. 186-203, 2008.

[19] J. P. Tardif, Y. Pavlidis and K. Daniilidis, "Monocular visual odometry in urban environments using an omnidirectional camera," in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2008.

[20] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser and P. Sayd, "Real Time Localization and 3D Reconstruction," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 2006.

[21] V. a. M.-N. F. a. F. P. Lepetit, "EPnP: An Accurate O(n) Solution to the PnP Problem," *International Journal of Computer Vision*, vol. 81, no. 2, pp. 155-166, 2008.

[22] C. Strecha, W. von Hansen, L. V. Gool, P. Fua and U. Thoennessen, "On benchmarking camera calibration and multi-view stereo for high resolution imagery," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[23] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.

**محمدامین مهرعلیان** در سال ۱۳۸۸ مدرک کارشناسی خود را در رشته مهندسی کامپیوتر- نرم‌افزار از دانشگاه اراک اخذ کرد. سپس در همان سال دوره کارشناسی ارشد را در گرایش هوش مصنوعی در دانشگاه صنعتی امیرکبیر شروع کرد و در سال ۱۳۹۰ فارغ التحصیل شد. وی از سال ۱۳۹۲ تا کنون در دوره دکتری هوش‌مصنوعی و رباتیک در دانشگاه علم و صنعت ایران مشغول به تحصیل است. زمینه های پژوهشی مورد علاقه وی شامل پردازش تصویر، بینایی‌ماشین و رباتیک است.

**محسن سریانی** در سال ۱۳۵۹ از دانشگاه علم و صنعت ایران مدرک کارشناسی خود را در رشته مهندسی برق – الکترونیک اخذ کرد و در سال ۱۳۶۶ از دانشکده مهندسی برق و کامپیوتر دانشگاه هریوت-وات در شهر ادینبورو در اسکاتلند در رشته مهندسی الکترونیک – تکنیک‌های دیجیتال در مقطع کارشناسی ارشد فارغ‌التحصیل شد. وی سپس در دوره دکتری در همان دانشگاه در گرایش پردازش تصویر ادامه تحصیل داد و در سال ۱۳۶۹ پس از فراغت از تحصیل، در دانشکده فنی دانشگاه مازندران به عنوان عضو هیات علمی مشغول به کار گردید. در سال ۱۳۸۱ به گروه هوش مصنوعی و رباتیک دانشکده مهندسی کامپیوتر دانشگاه علم و صنعت ایران آمد و در حال حاضر در این گروه با مرتبه دانشیاری مشغول به کار است. زمینه‌های پژوهشی مورد علاقه وی شامل پردازش و تحلیل تصویر، بینایی‌ماشین و پردازش تصاویر پزشکی است.