مقاله پژوهشی

مجله ماشین‌بینایی و پردازش تصویر

دوره ۱۲، شماره ۴، بهار ۱۴۰۵

تطبیق صحنه با ترکیب SIFT و ORB: بهبود کارایی و مقاوم‌سازی در پردازش تصویر

سجاد اردشیری[۱]، عاطفه سلیمی شهرکی[۲*] و مهدی عباسی[۳و۴]

چکیده

الگوریتم‌های تطبیق ویژگی زیربنای کلیدی بسیاری از کاربردهای بینایی ماشین به شمار می‌روند. با توجه به تنوع زیاد زمینه‌های کاربردی این الگوریتم‌ها و دامنه وسیع تغییرات بصری تصاویر، ارزیابی عملکرد الگوریتم‌ها یک جنبه نسبی داشته و انتخاب الگوریتم مناسب به تعادل بین دقت، سرعت، مقاومت و منابع محاسباتی بستگی دارد. در این تحقیق تلاش خواهد شد که ساختار الگوریتم‌های تطبیق کلاسیک و مبتنی بر یادگیری تحلیل شده و برخی از چالش‌های این الگوریتم‌ها بررسی شود. همچنین تلفیقی از این الگوریتم‌های SIFT و ORB در راستای بهینه‌سازی زمان پردازش در عین حفظ دقت تطبیق ارائه گردیده و SCO* نامگذاری می‌شود. به منظور تحلیل عملکرد الگوریتم‌های تطبیق، در یک سخت‌افزار دارای قابلیت پردازشی لازم به صورت بهینه شبیه‌سازی شده و عملکرد الگوریتم‌ها در رابطه با تصاویر چالش برانگیز مجموعه داده اکسفورد و با معیارهای استاندارد ارزیابی می‌گردد. نتایج ارزیابی نشان می‌دهد که زمان اجرای روش SCO به طور متوسط حدود ۰/۲۸ زمان اجرای SIFT و ۲/۶۷ برابر زمان اجرای ORBاست در حالیکه تعداد تطبیق‌های درست SCO به طور متوسط حدود ۰/۷۳ تطبیق‌های صحیح SIFT و حدود ۳/۳ برابر تطبیق‌های ORB می‌باشد.

کلید واژه ها

الگوریتم‌های تطبیق ویژگی کلاسیک و مبتنی بر یادگیری عمیق، ویژگی، تطبیق صحنه، نقطه کلیدی، توصیفگر، ماتریس انتقال تصویر

این مقاله در ۴ آذر ۱۴۰۴ دریافت و در ۶ خرداد ۱۴۰۵ پذیرفته شد

[۱] گروه مهندسی برق ،واحد اصفهان (خوراسگان)، دانشگاه آزاد اسلامی، اصفهان، ایران. رایانامه: slam.1329@chmail.ir

[۲*] گروه مهندسی برق، مرکز تحقیقات لیزر و بیوفوتونیک در فناوری‌های زیستی، واحد اصفهان (خوراسگان)، دانشگاه آزاد اسلامی، اصفهان، ایران.رایانامه: atefehsalimi@iau.ac.ir

[۳] گروه مهندسی کامپیوتر ، دانشکده مهندسی، دانشگاه بوعلی سینا، همدان، ایران. رایانامه: abbasi@basu.ac.ir

[۴] پژوهشکده علوم کامپیوتر، پژوهشگاه دانشهای بنیادی (IPM)، تهران، ایران. رایانامه: m.abasi@ipm.ir

۱- مقدمه

در عصر دیجیتال و توسعه فناوری‌های هوش مصنوعی، بینایی ماشین به عنوان یکی از مهم‌ترین و پرکاربردترین حوزه‌های مهندسی کامپیوتر، نقش محوری در درک و تحلیل تصاویر و ویدئوها توسط سیستم‌های هوشمند ایفا می‌کند. در این میان، استخراج ویژگی[1] و تطبیق صحنه[2] به عنوان دو پایه اساسی در اکثر الگوریتم‌های پردازش تصویر، به ویژه در کاربردهایی نظیر تشخیص اشیاء، ردیابی حرکت، مدل‌سازی سه‌بعدی، ناوبری خودکار و واقعیت افزوده، اهمیت حیاتی دارند.

---

* SIFT Combined with ORB

1 Feature Extraction

2 Scene Matching

شکل (۱) : دسته‌بندی الگوریتم‌های تطبیق

ویژگی‌ها به عنوان نقاط یا نواحی خاص در تصویر تعریف می‌شوند که دارای اطلاعات ساختاری، هندسی یا بافتی منحصربه‌فردی هستند و از تبدیلات هندسی (مانند چرخش، مقیاس و تغییر زاویه دید) و فوتومتریک (مانند تغییر نورپردازی و کنتراست) نسبتاً مستقل عمل می‌کنند. تطبیق صحنه نیز به فرآیند یافتن نقاط مشترک بین دو یا چند تصویر از یک صحنه گفته می‌شود که برای درک ساختار دنیای واقعی و تحلیل رابطه بین تصاویر ضروری است.

برخی از کاربردهای گسترده‌ی تطبیق صحنه شامل بازیابی تصویر[۱, ۲]، بازسازی سه‌بعدی[۳, ۴]، مکان‌یابی و نقشه‌برداری همزمان [۵] (SLAM[1]) ، تثبیت ویدئو[۶, ۷] و غیره. به عنوان مثال، در زمینه ربات‌های خدماتی، هدف این است که ربات‌ها بتوانند به طور خودکار حرکت کنند، از موانع دوری کنند، اشیا را تشخیص دهند و با محیط تعامل داشته باشند.

الگوریتم‌های کلاسیک تطبیق را می‌توان به دو دسته‌ی روش‌های مبتنی بر ناحیه[۸] و مبتنی بر ویژگی[۹] تقسیم کرد. روش‌های مبتنی بر ناحیه از بررسی مستقیم شباهت پیکسل‌های تصویر در یک پنجره لغزنده با اندازه از پیش تعیین شده برای تعیین تطابق بین تصاویر استفاده می‌کنند در حالیکه روش‌های مبتنی بر ویژگی انعطاف‌پذیری و پایداری بیشتری ارائه داده و به طور معمول در تطبیق تصویر استفاده می‌شوند. این رویکرد شامل تشخیص ساختارهای منحصر به فرد تصویر و سپس توصیف آن‌ها به صورت ویژگی با استفاده از مدل‌های ریاضی و در نهایت تبدیل مسئله تطبیق به یافتن نزدیکترین همسایه‌ها در فضای ویژگی با ابعاد بالا از توصیف‌گرها می‌باشد.[۱۰] الگوریتم‌های مبتنی بر ویژگی را بر اساس ویژگی‌های مورد علاقه آن‌ها می‌توان به ۳ دسته‌ی مبتنی بر گوشه، حباب و شدت روشنایی پیکسل تقسیم کرد.

امروزه روش‌های مبتنی بر یادگیری در دو حالت در حال توسعه هستند. یک رویکرد بکارگیری الگوریتم‌های مبتنی بر یادگیری عمیق در مراحل استخراج، توصیف و تطبیق ویژگی است و دیگری رویکرد بکارگیری سرتاسری[۲] است که به معنای استفاده از یک شبکه عصبی برای انجام یکپارچه مراحل استخراج، توصیف و تطبیق ویژگی از تصویر ورودی است. به طور کلی الگوریتم‌های تطبیق صحنه را می‌توان به صورت نشان داده شده در شکل (۱) دسته‌بندی کرد.

در سال‌های اخیر، الگوریتم‌های مختلفی برای استخراج ویژگی توسعه یافته‌اند که هر کدام با الهام از اصول ریاضی، آنالیز تصویر و یادگیری ماشین، به دنبال دستیابی به تعادل بهینه‌ای بین دقت، سرعت، مقاومت و مقیاس‌پذیری هستند. در این میان، الگوریتم‌هایی نظیرSIFT، SURF، ORB، KAZE/AKAZE و BRISK به دلیل عملکرد قابل توجه خود در شرایط متنوع، به عنوان معیارهای استاندارد در بسیاری از پژوهش‌ها و پیاده‌سازی‌های عملی مورد استفاده قرار گرفته‌اند. البته با توجه به توسعه الگوریتم‌های مبتنی بر یادگیری عمیق در سال‌های اخیر روش‌هایی مانند SuperPoint ، Key.Net، HardNet برای کاربردهای تطبیق توسعه یافته‌اند.

علیرغم ظهور روش‌های مبتنی بر هوش مصنوعی و توسعه چشمگیر این روش‌ها در سالیان اخیر، روش‌های سنتی تطبیق همچنان مورد توجه هستند. در واقع وجود مسائلی که در ادامه به برخی از آن‌ها

[1] Simultaneous Localization and Mapping (SLAM)

[2] End to end

اشاره می‌شود، باعث شده است که روش‌های سنتی تطبیق همچنان محبوبیت داشته باشند. مشکلاتی چون:

- نیاز به بکارگیری انواع سخت‌افزارهای نهفته در کاربردهای صنعتی که غالباً معماری آن‌ها GPU را در اختیار ندارند در حالیکه به روش‌های مبتنی بر یادگیری اغلب به دلیل پیچیدگی و حجم بالای محاسبات، در فرآیندهای بلادرنگ برای ایجاد امکان بکارگیری نیاز به سخت‌افزارهایی شامل GPU دارند.
- با توجه به بکارگیری معماری‌های مختلف در حوزه سخت افزارهای GPUهای مورد استفاده در انواع سیستم‌های نهفته عرضه شده، مشکلاتی برای پیاده‌سازی الگوریتم‌های مبتنی بر یادگیری وجود دارد و یا حداقل به سادگی نمی‌توان از قابلیت‌های آنها استفاده نمود. به بیان دیگر سازگار کردن کتابخانه‌های نرم‌افزاری عرضه شده برای الگوریتم‌ها با معماری GPUها پیچیدگی زیادی دارد.
- روش‌های مبتنی بر یادگیری هر چند عملکرد بهتری نسبت به توصیفگرهای کلاسیک دارند ولی عملکرد آن‌ها به شدت تحت تأثیر نمونه‌های آموزشی قرار می‌گیرد بطوریکه در سناریوهایی با بافت‌های ضعیف یا تصاویر ناشناخته، عملکرد مناسبی از خود نشان نداده و نسبت تطبیق نادرست بالایی در تطبیق از خود نشان می‌دهند. جهت تحقیقاتی در زمینه این الگوریتم‌ها نیز نشان دهنده توجه محققان به رفع این مشکلات است.

رویکرد دنبال شده در مقاله توسعه یا انتخاب الگوریتم تطبیق صحنه با قابلیت پیاده‌سازی در سخت‌افزارهای فاقد GPU و امکان استفاده در کاربردهای بلادرنگ است لذا مباحث بیان شده غالباً الگوریتم‌های کلاسیک تطبیق صحنه را شامل می‌شود.

انتخاب الگوریتم مناسب برای استخراج ویژگی و تطبیق صحنه با چالش‌های متعددی همراه است. از جمله برقراری تعادل بین دقت و سرعت و مقاوم بودن در برابر تغییرات محیطی مانند نویز، تغییرات خطی و غیرخطی روشنایی[1]، تغییرات مقیاس[2]، چرخش[3]، شلوغی صحنه[4]، انسداد[5] یا پوشیده شدن قسمتی از شی و تغییرات زاویه دید[6]. چنین مواردی در نهایت تنوع زیادی در نوع ویژگی‌های استخراجی ایجاد می‌کند که باید مد نظر محققان در انتخاب الگوریتم مناسب برای کاربرد خاص خود قرار گیرد.

هدف اصلی این مقاله ارائه یک تحلیل جامع از مراحل اجرای الگوریتم‌های شناخته شده استخراج ویژگی و تطبیق صحنه و سپس ارائه یک الگوریتم تطبیق پیشنهادی در راستای بهبود سرعت اجرا و پیچیدگی محاسباتی در عین حفظ مقاومت الگوریتم تطبیق در برابر تغییرات بصری تصویر است. با توجه به اینکه در بسیاری از سیستم‌های کاربردی محدودیت زمان پردازش و منابع سخت‌افزاری در دسترس از جمله چالش‌های مهم به شمار می‌رود، در راستای بهینه‌سازی محاسباتی و زمان پردازش، در مرحله تولید و تطبیق توصیفگرها، در این مقاله تلفیقی از دو الگوریتم شاخص پیشنهاد گردیده و عملکرد آن در مقایسه با الگوریتم‌های استاندارد بر روی یک پایگاه استاندارد مورد ارزیابی قرار گرفته است. همچنین در این مقاله تلاش شده است چالش‌های موجود در رابطه با الگوریتم‌های تطبیق تا حدودی شناسایی و پیشنهاداتی برای جهت‌های تحقیقاتی آینده در حوزه استخراج ویژگی ارائه شود.

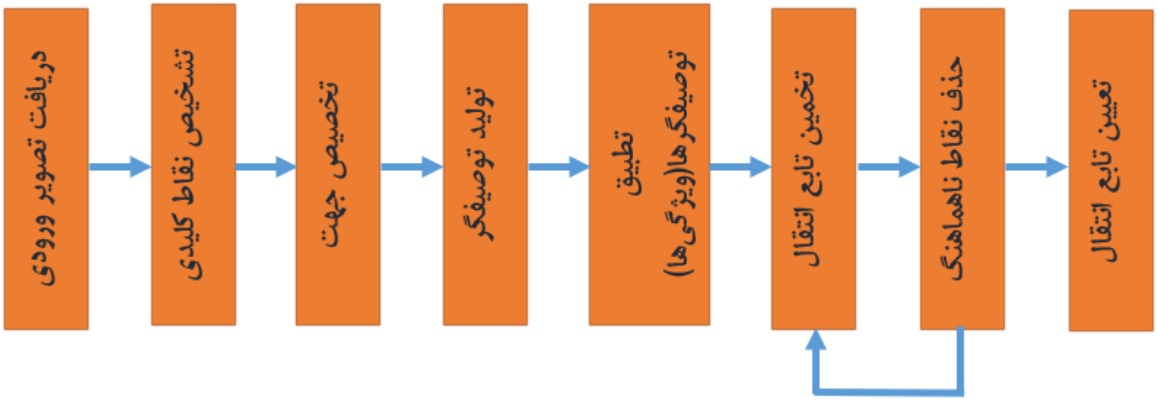
الگوریتم SIFT به دلیل مقاومت بسیار بالا در برابر تبدیلات هندسی و فوتومتریک، به عنوان یک معیار طلایی در استخراج ویژگی شناخته شده و پایه بسیاری از روش‌های بعدی قرار گرفته است. بررسی روند توسعه الگوریتم‌های تطبیق کلاسیک بعد از SIFT نشان می‌دهد که تمرکز اغلب این الگوریتم‌ها، بهینه‌سازی زمان اجرای الگوریتم SIFT با بکارگیری ابتکاراتی در مراحل آشکارسازی نقاط کلیدی و تولید توصیفگر می‌باشد. از طرفی ORB یکی از شاخص‌ترین الگوریتم‌های توسعه داده شده برای کاربردهای بلادرنگ با تمرکز با کاهش حجم محاسباتی در مرحله تطبیق توصیفگرها بوده است. لذا در این مقاله، با رویکرد توسعه الگوریتم تطبیق با ویژگی بهبود سرعت اجرا و کاهش پیچیدگی محاسباتی در عین حفظ مقاومت مناسب الگوریتم تطبیق در برابر تغییرات بصری تصویر است و قابلیت پیاده‌سازی در سیستم‌های نهفته با منابع سخت‌افزارهای محدود و کاربردهای بلادرنگ، یک تلفیق هوشمندانه از SIFT و ORB پیشنهاد تحت عنوان الگوریتم SCO پیشنهاد می‌شود.

۲- مرور تاریخچه

توسعه تطبیق تصویر با استفاده از مجموعه‌ای از نقاط یا ویژگی‌های محلی می‌تواند به کار موراک[7] (۱۹۸۱) در تطبیق استریو با استفاده از یک آشکارساز گوشه برگردد.[۱۱] اشمید و مور[8] (۱۹۹۷) نشان دادند که تطبیق ویژگی‌های محلی پایدار را می‌توان به چالش‌های شناسایی تصویر متنوع گسترش داد که در آن یک ویژگی در یک پایگاه داده بزرگ از تصاویر مطابقت داده‌شد.[۱۲]

لاو[9] (۱۹۹۹) با معرفی الگوریتمSIFT ، انقلابی در این حوزه ایجاد کرد[۱۳]. این الگوریتم به دلیل مقاومت بالا در برابر تبدیلات هندسی و فوتومتریک، به سرعت به استاندارد طلایی تبدیل شد. در پی آن، بای و همکاران[10] (۲۰۰۶) الگوریتم SURF را معرفی کردند که با استفاده از تصاویر انتگرالی و تقریب هسین، سرعت بالاتری نسبت به SIFT ارائه داد.[۱۴] روش‌های ASIFT

[1] Illumination
[2] Scale
[3] Rotation
[4] Background clutter
[5] Occlusion
[6] Viewpoint change
[7] Moravec
[8] Schimd and Mohr
[9] Lowe
[10] Bay



شکل (۲): مراحل کلی استخراج و تطبیق ویژگی

و ASURF[1] که برای شناسایی شی حتی در حالت‌هایی که شی تغییرات افین[2] زیاد داشته‌باشد، ارائه شد[۱۴ـ۱۶].روبلی[3] و همکاران (۲۰۱۱) الگوریتم ORB را ارائه کردند که با ترکیب آشکارساز FAST و توصیفگرBRIEF ، گزینه‌ای سریع، کارآمد و بدون حق نشر ایجاد کردند.[۱۷] در همین سال لئوتنگر[4] و همکاران، الگوریتم BRISK را معرفی کردند که از نمونه‌برداری حلقوی برای دستیابی به مقیاس‌پذیری استفاده می‌کرد[۱۸]. همزمان، تحقیقاتی در جهت بهبود کیفیت استخراج ویژگی در محیط‌های پرنویز انجام شد. آلکانتاریلا[5] و همکاران (۲۰۱۲) با معرفی KAZE ، از فضای مقیاس غیرخطی استفاده کردند که عملکرد بهتری در تصاویر پرنویز داشت[۱۹]. نسخه بهبودیافته آن، AKAZE، با بهینه‌سازی محاسبات، سرعت بیشتری ارائه داد.

در سال‌های اخیر، تمرکز تحقیقات به سمت الگوریتم‌های مبتنی بر یادگیری عمیق مانند SuperPoint، Key.Net، HardNet معطوف شده است [۱۴] که عملکرد بهتری نسبت به روش‌های کلاسیک نشان می‌دهند. با این حال، الگوریتم‌های کلاسیک همچنان به دلیل شفافیت، قابلیت تفسیر، امکان پیاده سازی بر روی سخت افزارهای با پیچیدگی کمتر و عدم نیاز به داده‌های آموزشی، در بسیاری از پروژه‌ها محبوبیت دارند. در بین الگوریتم‌های کلاسیک الگوریتم‌های تطبیق کلاسیک SIFT به دلیل مقاومت بسیار بالا در برابر تغییرات تصویر همواره به عنوان یک الگوریتم شاخص مورد توجه بوده است، بطوریکه که انگیزه توسعه اغلب الگوریتم‌های پس از آن بر بهینه‌سازی زمان اجرا با حفظ دقت و مقاومت SIFT بوده است. از طرفی ORB و الگوریتم‌های توسعه یافته بر مبنای آن از بهترین الگوریتم‌های تطبیق از نظر زمان اجرا و قابلیت پیاده‌سازی در سیستم‌های نهفته بلادرنگ می‌باشد. بدیهی است که یک انتخاب هوشمندانه برای توسعه یک الگوریتم با قدرت مناسب و سرعت اجرای بهینه می‌تواند تلفیقی از این دو الگوریتم شاخص باشد.

۳- تحلیل ساختاری الگوریتم‌های تطبیق

بر اساس روند تطبیق ویژگی تشریح شده در [۲۰] و اغلب مقالات مرجع الگوریتم‌های تطبیق ویژگی می‌توان فرآیند استخراج ویژگی و تطبیق صحنه در الگوریتم‌های کلاسیک را طبق مراحل نشان داده شده در شکل (۲) طبقه‌بندی کرد.

در ادامه اصول و روابط پیاده‌سازی هر از مراحل شکل (۲) برای الگوریتم‌های مختلف تشریح خواهد شد.

۳ـ۱ ـ تشخیص (آشکارسازی) نقاط کلیدی (ویژگی)

در این مرحله، نقاطی با اطلاعات ساختاری منحصر به فرد (مانند گوشه‌ها یا نقاط با تغییر شدت بالا) شناسایی می‌شوند. تناظر بین توصیفگرهای تصویر و صحنه تنها در شرایطی می‌تواند برقرار شود که نقاط‌کلیدی با کیفیت مناسب در هر دو تصویر آشکار گردد. لذا این نقاط باید تکرار پذیر، متمایز، دارای تعداد کافی و دقت مناسب بوده و در کاربردهایی با محدودیت زمانی، دارای سرعت استخراج مناسب باشند.

در SIFT برای تشخیص کارآمد مکان‌های کلیدی پایدار نسبت مقیاس، از نقاط اکسترمم فضای مقیاس در تابع تفاضل گاوسی استفاده می‌شود، چراکه آشکارساز DoG تقریب مناسبی از لاپلاسین گوسی نرمالیزه شده با مقیاس ($\sigma^2\nabla^2G$) است[۱۳] که پایدارترین ویژگی‌های را در مقایسه با تعداد زیادی از دیگر توابع مانند گرادیان، هسین یا تابع گوشه هریس بدست می‌دهد. تابع تفاضل گوسی از کانوالو تصویر در تفریق دو تابع گوسی که مقیاس آن‌ها تنها در یک ضریب ثابت k متفاوت است، مطابق رابطه (۱ و ۲) بدست می‌آید.

$$D(x.y.\sigma) = (G(x.y.k\sigma) - G(x.y.\sigma)) * I(x.y) = L(x.y.k\sigma) - L(x.y.\sigma) \quad (۱)$$

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (۲)$$

[1] Affine SIFT and Affine SURF
[2] Affine
[3] Rublee
[4] Leutenegger
[5] Alcantarilla

به منظور پوشش کامل مقیاس، هرم مقیاس شکل (۳) در لایه‌های منظمی با تعداد سطوح مشخص S که دارای فاصله‌گذاری لگاریتمی از مقیاس پایه ($k= 2^{(1/S)}$)، تحت عنوان اکتاو تشکیل می‌شود. به منظور بهینه‌سازی هزینه محاسباتی تصاویر انتهایی هر اکتاو با نسبت ۲ زیرنمونه‌برداری شده و به عنوان تصویر ورودی اکتاوهای بالاتر بکار می‌رود.

هر نقطه نمونه از سطوح DoG که نسبت به همه‌ی ۲۶ نمونه در یک همسایگی ۳ × ۳ × ۳ در مقیاس فعلی و مقیاس‌های بالاتر و پایین‌تر خود بزرگتر یا کوچکتر باشد به عنوان نقطه‌ی کلیدی بالقوه لحاظ می‌شود. با اعمال دو آستانه بر مبنای نسبت معینی از مقادیر دترمینان و تریس ماتریس هسین $\frac{Tr(H)^2}{Det(H)} = \frac{(r+1)^2}{r}$ در هر نقطه کلیدی در تصویر اصلی، نقاط روی لبه (r=10) و نقاط ناپایدار (r=2) حذف می‌شوند.

SURF[1] به عنوان یک جایگزین سریع‌تر نسبت به SIFT با حفظ مقاومت قابل قبول ارائه گردید. در این الگوریتم به طور ویژه بهینه‌سازی‌هایی هم در مرحله آشکارسازی نقاط کلیدی (یا نقاط مورد علاقه) و مرحله تولید توصیفگر با استفاده از ویژگی تصاویر انتگرالی و فیلترهای جعبه‌ای به منظور تسریع در اجرای الگوریتم صورت گرفته است.

در SURF، آشکارساز نقاط کلیدی بر اساس ماتریس هسین بنا نهاده شده و تشخیص مکان و مقیاس نقاط کلیدی به معیار دترمینان هسین متکی است. نوآوری SURF تقریب فیلترهای گوسی پس از گسسته‌سازی و برش به فیلترهای جعبه‌ای ۳ سطحی گسسته با مقادیر تکراری است که به این ترتیب امکان استفاده از ویژگی تصویر انتگرالی در مراحل کانوالو فیلترهای مذکور در تصویر و تسریع چشمگیر در محاسبات حین تشکیل فضای مقیاس فراهم می‌شود. شکل (٤) نمونه‌ای از تقریب فیلتر گوسی برای محاسبه L_{yy} و L_{xy} مورد استفاده در محاسبه ماتریس هسین هر پیکسل را نشان می‌دهد.

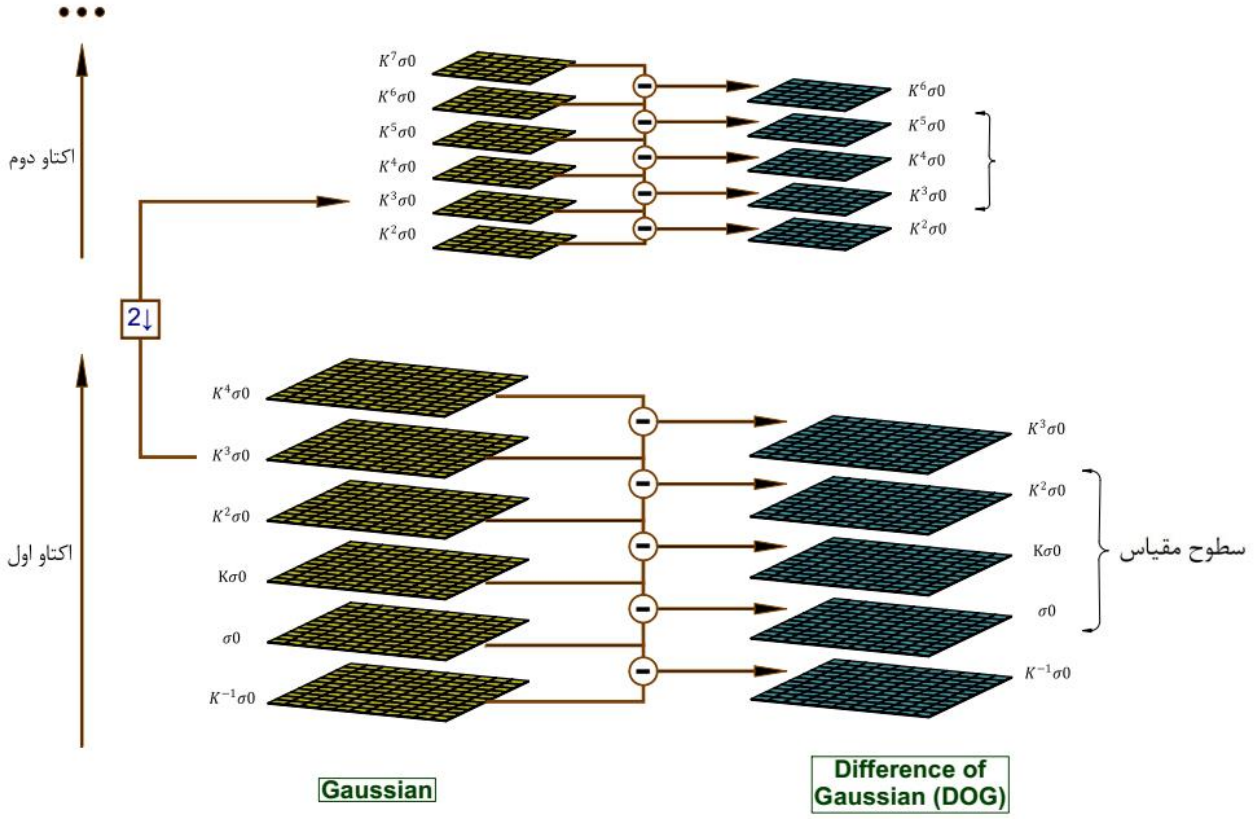
در SURF هرم مقیاس با اعمال مکرر فیلترهای جعبه‌ای مربوط به ماتریس هسین، با ابعاد معین روی تصویر اصلی ایجاد می‌شود. ابعاد فیلترها و گام افزایش ابعاد فیلترها متناسب با مقیاس هر اکتاو تعیین می‌شود. به طور مثال فیلتر ۹ × ۹ نشان داده شده در شکل (٤) مربوط به لایه مقیاس $S = 1.2$ است. پاسخ هسین دترمینان این ماتریس برای تمام پیکسل‌های تصاویر فضای مقیاس با استفاده از رابطه‌ی (۳) محاسبه می‌شود:

$$det(H_{approx}) = L_{xx}\,L_{yy} - 0.9L_{xy} \qquad (۳)$$

نقاط کلیدی بالقوه اکسترمم‌های پاسخ‌های هسین یک همسایگی ۳×۳×۳ خواهند بود. با قرار دادن دو آستانه‌ی محاسبه شده از روابط (٤) و (٥) نقاط ناپایدار حساس به نویز و نقاط روی لبه حذف و نقاط کلیدی نهایی انتخاب می‌شوند.

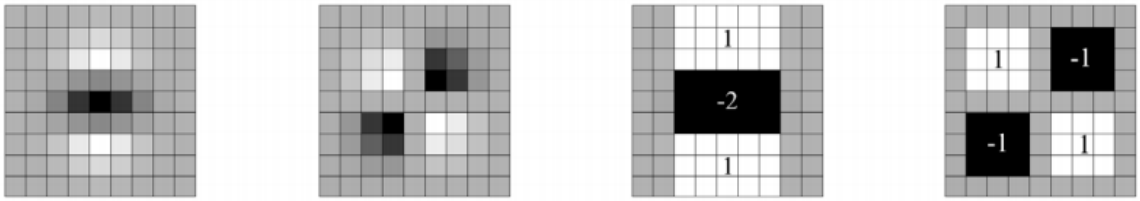
$$R_1 = trace(H) = L_{xx} + L_{yy} \qquad (۴)$$

$$R_2 = det(H) - \frac{1}{3}\,trace(H)^2 \qquad (۵)$$

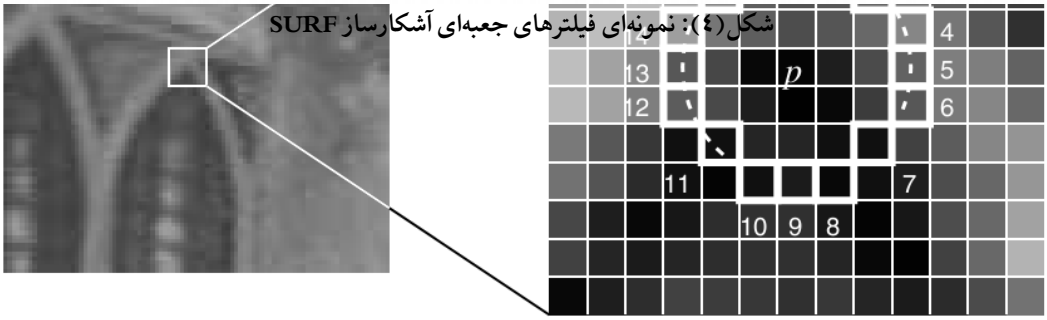


شکل (۳) : روش ایجاد فضای مقیاس و تفاضل‌های گوسی تصویر

[1] Speeded-Up Robust Features



شکل(٤): نمونه‌ای فیلترهای جعبه‌ای آشکارساز SURF



شکل (۵): آشکارسازی گوشه به روش FAST

ORB[1] با ترکیب نسخه‌های بهبود یافته آشکارساز گوشه FAST و توصیفگر باینری BRIEF از نظر تخصیص جهت و تطبیق، برای استفاده در سیستم‌های تعبیه‌شده و کاربردهای مبتنی بر پردازش تصویر بلادرنگ (مانند SLAM، تشخیص اشیا و ناوبری ربات) توسعه یافته است.

FAST[2] با بررسی یک دنباله از ۱۶ پیکسل روی محیط دایره‌ای با شعاع ۳ پیکسل حول یک پیکسل مرکزی (تحت بررسی) تعیین می‌کند که آیا آن نقطه یک گوشه است یا خیر. شرط گوشه‌بودن این است که در مرحله اول حداقل n (حداقل n برابر ۹ و مقدار مطلوب بیش از ۱۲ پیکسل است) پیکسل متوالی از ۱۶ پیکسل بالا همگی حداقل t مقدار خاکستری روشن‌تر یا تیره‌تر از مقدار مرکزی باشند. اگر در همسایگی پیکسل مورد بررسی(۸ پیکسل مجاور) پیکسلی با n یا t بیشتر وجود داشته باشد پیکسل مورد بررسی گوشه نیست. نمونه‌ای از ناحیه تعریف شده در FAST در شکل (۵) نشان داده شده است.

FAST ویژگی‌های چند مقیاسی تولید نمی‌کند لذا در ORB از هرم مقیاس تصویر استفاده شده و ویژگی‌های FAST (فیلتر شده توسط هریس) را در هر سطح هرم تولید می‌شوند ولی برای ایجاد هرم مقیاس شامل نسخه‌های متفاوتی از تصویر اصلی، به جای فیلتر گاوسی و محاسبات پیچیده DoG به سادگی ابعاد تصویر را با فاکتورهای مختلف (معمولاً ۲) کاهش می‌دهد.

روش‌های مبتنی بر یادگیری عمیق در مرحله استخراج ویژگی از شبکه‌های عصبی کانولوشنال برای تولید یک نقشه ویژگی استفاده می‌کنند. مفهوم تشخیص ویژگی در این الگوریتم ها نگاشت تصویر به یک نقشه امتیاز بدست آمده از طریق آموزش یادگیری ماشین است. میزان امتیاز هر پیکسل را می‌توان به عنوان احتمال اینکه یک ویژگی برجسته باشد تفسیر کرد. در این روش‌ها بر خلاف روش‌های سنتی نقاط ویژگی لزوماً نقاطی با "تغییرات محلی سریع" نبوده بلکه نقاط تکرارپذیر پیش‌بینی می‌شوند لذا امکان تشخیص نقاط ویژگی حتی در نواحی صاف تصویر نیز وجود دارد. SuperPoint یک شبکه عصبی که ویژگی‌های بهینه‌ای را یاد می‌گیرد و عملکرد بهتری نسبت به SIFT دارد[۲۱].

۳ـ۲ـ تخصیص جهت[3]

با تخصیص یک جهت غالب به هر نقطه کلیدی بر اساس مشخصه‌های محلی آن و استفاده از این جهت در مرحله تولید توصیفگر، قابلیت چرخش‌پذیری٤ در الگوریتم ایجاد می‌شود.

به منظور تعیین جهت غالب در SIFT، مقدار و زاویه گرادیان پیکسل‌ها در یک همسایگی مربعی به ابعاد ١٦×١٦ پیکسل اطراف نقطه کلیدی در تصویر گوسی با نزدیک‌ترین سیگما به مقیاس نقطه کلیدی مطابق روابط (٦ و٧) محاسبه می‌شوند.

(٦)

$$m(x,y)=\sqrt{\begin{matrix}\big(L(x+1,y)-L(x-1,y)\big)^2+\\(L(x,y+1)-L(x,y-1))^2\end{matrix}}$$

(٧)

$$\theta(x,y)=tan^{-1}\left(\frac{L(x+1,y)-L(x-1,y)}{L(x,y+1)-L(x,y-1)}\right)$$

مقادیر گرادیان‌ها بر اساس یک تابع گوسی با انحراف معیار ۱/۵ برابر مقیاس نقطه کلیدی و مرکز نقطه کلیدی وزن‌دهی می‌شوند و یک هیستوگرام جهت‌دار گرادیان‌ها[5] با ۳٦ دسته[6]ی ۱۰ درجه‌ای تشکیل می‌شود. یک آستانه ۸۰٪ مقدار ماکزیمم مقدار هیستوگرام اعمال شده و در نهایت با درونیابی مکعبی حول ماکزیمم‌های

[1] Oriented FAST and Rotated BRIEF
[2] Feature from Accelerate Segment Test (FAST)
[3] Orientation Assignment
[4] Rotation Invariance

[5] Histogram of Oriented Gradients (HOG)
[6] Bin

هیستوگرام و یک زاویه به عنوان جهت غالب به نقطه کلیدی تخصیص داده می‌شود. در این مرحله به تعداد ماکزیمم‌های درونیابی شده هیستوگرام نقاط کلیدی با زاویه‌های غالب و موقعیت مشابه تولید می‌شود و چنانچه در یک نقطه کلیدی بیش از ۳ جهت غالب وجود داشت، نقطه ناپایدار بوده و حذف می‌شود.

در SURF فیلترهای موجک هار نشان داده شده در شکل (٦) در یک همسایگی دایره‌ای در اطراف نقطه کلیدی با شعاع $6S$ در جهت‌های x و y روی تصویر اصلی اعمال شده و با یک تابع گوسی به مرکز نقطه کلیدی و انحراف معیار σ=2.5S وزن‌دهی می‌شوند.

در مرحله کانوالو فیلتر موجک روی تصویر نیز با توجه به ویژگی تکرار مؤلفه‌های فیلتر موجک هار از خاصیت تصویر انتگرالی برای کاهش هزینه محاسباتی استفاده می‌شود. ناحیه همسایگی دایره‌ای به ٦ قطاع ٦٠ درجه‌ای تقسیم و در هر قطاع مجموع پاسخ‌های افقی و عمودی محاسبه شده و سپس اندازه و زاویه هر قطاع با روابط (۸ و ۹) بدست می‌آید:

$$m(Haar_{\forall}) = \sqrt{(d_x)^2 + (d_y)^2} \qquad (۸)$$

$$\theta(Haar_{\forall}) = tan^{-1}(\frac{d_x}{d_y}) \qquad (۹)$$

θ متناظر با قطاعی که دارای ماکزیمم اندازه است، مشخص‌کننده جهت‌گیری غالب نقطه کلیدی می‌باشد.

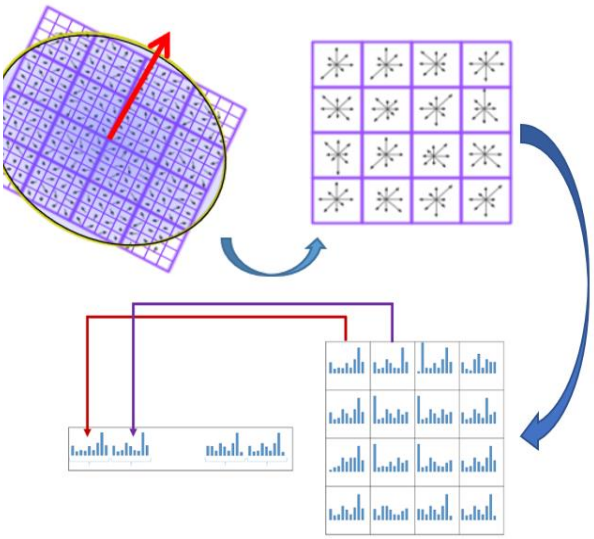
ORB از الگوریتم FAST به دلیل ویژگی‌های محاسباتی آن برای آشکارسازی نقاط کلیدی استفاده می‌کند؛ ولی FAST بر خلاف آشکارسازهایی مانند SIFT و SURF فاقد عملگر جهت‌گیری است. لذا در ORB یک جزء جهت‌گیری سریع و دقیق به FAST افزوده شده و تحت عنوان oFAST[1] نامگذاری می‌شود. به این منظور فرض می‌شود که مرکز ثقل شدت گوشه از مرکز آن جدا شده است. لذا می‌توان یک بردار از مرکز گوشه، O، به مرکز ثقل شدت آن، ($\overrightarrow{OC}$) ایجاد کرد. سپس یک ناحیه به شعاع r حول نقطه کلیدی تعریف شده و جهت‌گیری گوشه بر اساس محاسبات ممان‌های تصویر در راستای عمودی و افقی (m01,m10) در این ناحیه به صورت رابطه (۱۰) تعیین می‌شود:

$$\theta=atan2(\ m01\ \ ,m10) \qquad (۱۰)$$

شکل(۶): فیلترهای استفاده شده برای محاسبه پاسخ موجک هار در راستای افقی و عمودی

۳ـ۳ـ تولید توصیفگر

توصیفگر یک بردار عددی حول نقطه کلیدی است که اطلاعات محلی (گرادیان، شدت، بافت) را کدگذاری می‌کند. هدف در طراحی توصیفگرها، تمرکز بر اطلاعات محلی به‌گونه‌ایست که توصیفگر نسبت به برخی متغیرها مثل نقطه‌ی دید و روشنایی ثابت بماند و در عین حال به اندازه‌ی کافی اطلاعات ذخیره کند تا از توصیفگرهای بقیه‌ی نقاط‌کلیدی متمایز باشد. علاوه بر این باید کوتاه و کارآمد باشد چرا که ابعاد توصیفگر در هزینه محاسبات مرحله تطبیق ویژگی تأثیرگذار است. شکل (۷) ساختار توصیفگر را نشان می دهد.



شکل(۷): ساختار توصیفگر SIFT

به منظور تولید توصیفگر در SIFT، یک ناحیه ١٦×١٦ پیکسل حول نقطه کلیدی در تصویر اصلی در نظر گرفته می‌شود. این همسایگی به ١٦ زیر ناحیه‌ی مربعی ٤×٤ پیکسل تقسیم شده و در هر زیر ناحیه اندازه و زاویه‌ی گرادیان‌ها برای تمام پیکسل‌ها محاسبه می‌شود. زوایای گرادیان‌ها نسبت به زاویه غالب نقطه‌ی کلیدی تراز شده و اندازه گرادیان‌ها بر اساس یک تابع گوسی با انحراف معیار معادل ۰/۸ برابر مقیاس نقطه کلیدی وزن‌دهی می‌شوند. سپس بر اساس مقدار و زوایای گردایان‌ها در هر زیرناحیه یک HoG با ۸ دسته‌ی ٤٥ درجه‌ای تشکیل می‌شود. در نهایت مطابق شکل (۷)، با چینش مقادیر هیستوگرام‌های ١٦ ناحیه پشت سر هم بردار ۱۲۸ بُعدی پس از نرمالیزه کردن و اعمال یک آستانه ۰/۲ به منظور حذف اثرات خطی فوتومتریک (مانند تغییر نورپردازی و کنتراست) و غیرخطی فوتومتریک (اشباع دوربین و انعکاس‌های شدید نور)، توصیفگر SIFT با مؤلفه‌های نوع اعشاری ٤ بایتی با دقت معمولاً ۱۰ رقم، ایجاد می‌شود.

توصیفگرهای ویژگی سنتی برای استخراج ویژگی بر تبدیلات ریاضی صریح برای تجزیه و تحلیل تفاوت‌های مقیاس خاکستری بین تصاویر تکیه دارند در حالیکه رویکردهای مبتنی بر یادگیری عمیق به ویژگی‌های دست‌ساز[2] صریح متکی نبوده بلکه از دانش الگوریتم‌های یادگیری عمیق بهره می‌برند.

[1] Oriented FAST

[2] handcrafted

در SURF یک ناحیه مربعی به ابعاد $20S$ اطراف نقطه کلیدی تعریف شده و بر اساس جهت غالب نقطه کلیدی تراز می‌شود. ناحیه همسایگی به ۴ × ۴ زیرناحیه مساوی تقسیم و فیلترهای موجک افقی و عمودی متناسب با تراز نواحی همسایگی با جهت غالب، به ابعاد $2S$ به هر زیرناحیه اعمال می‌شود. در این مرحله نیز از ویژگی فیلتر موجک و تصویر انتگرالی برای تسریع محاسبات استفاده می‌شود. به منظورکاهش تأثیر تغییر شکل هندسی و خطای مکانیابی مقادیر پاسخ‌های عمودی و افقی با استفاده از فیلتر گوسی با مرکز نقطه کلیدی و انحراف معیار σ=3.3S وزن‌دهی می‌شوند. مجموع پاسخ‌های افقی و عمودی به همراه مجموع قدرمطلق پاسخ‌ها در هر زیرناحیه مولفه‌های بردار توصیفگر را به صورت نشان داده شده در رابطه‌ی (۱۱) تشکیل می‌دهند.

$$V = (\sum d_x , \sum d_y , \sum |d_x| , \sum |d_y|) \quad (۱۱)$$

به این ترتیب بردار توصیفگر SURF برای هر نقطه کلیدی در مجموع ۴ × ۱۶ یعنی ۶۴ مؤلفه خواهد داشت.

ORB از الگوریتم BRIEF[1] که یک توصیفگر باینری سریع و کارآمد است، استفاده می‌کند. این توصیفگر به جای محاسبه توصیف‌گرهای پیچیده مبتنی بر گرادیان (مانند SIFT یا SURF) از توزیع الگوهای باینری حاصل از مقایسه‌های محلی پیکسل‌ها استفاده می‌کند. ابتدا یک ناحیه مربعی ۳۱ × ۳۱ پیکسلی (تحت عنوان patch) حول نقطه کلیدی در تصویر با یک فیلتر میانگین‌گیر هموار می‌شود که در این مرحله از ویژگی تصویر انتگرالی برای تسریع محاسبات استفاده می‌شود. مجموعه‌ای ۲۵۶ تایی از جفت پیکسل‌های (p_i , q_i) در یک زیرپنجره‌ی ۵ × ۵ که به صورت تصادفی محدوده patch را جارو می‌کند، انتخاب شده و در آزمون باینری BRIEF شرکت می‌کنند. البته مکانیابی زیرپنجره‌ی انتخاب جفت پیکسل‌ها دارای یک توزیع گوسی حول مرکز patch است تا توزیع انتخاب‌ها بر اساس فاصله تا نقطه کلیدی وزن‌دهی شود. برای هر جفت پیکسل تصادفی انتخاب شده آزمون باینری به صورت رابطه‌ی (۱۲) برقرار و با مقایسه شدت آن‌ها، یک بیت تولید می‌شود:

$$f_i = \begin{cases} 1 & if\ I(p_i) < I(q_i) \\ 0 & otherwise \end{cases} \quad (۱۲)$$

لذا برای مجموعه ۲۵۶ جفت پیکسل تصادفی یک رشته باینری $f = [f_1 , f_2 , \dots , f_n]$ به طول ۲۵۶ بیت ایجاد خواهد شد.

BIERF نسبت به چرخش تصویر حساس است و به منظور ایجاد قابلیت چرخش‌پذیری BIERF بر اساس جهت نقاط کلیدی توجیه شده و الگوریتمی موسوم به BRIEF توجیه شده[2] شکل می‌گیرد که در آن یک نسخه توجیه شده (S_θ) از ماتریس مکانی جفت پیکسل‌های تصادفی (S) بر اساس ماتریس دوران مربوط به جهت نقطه کلیدی (R_θ) ایجاد می‌شود. یکی از الزامات اساسی BRIEF واریانس زیاد بیت‌های توصیفگر است که تمایزپذیری توصیفگر را تأمین می‌کند. ویژگی مطلوب دیگر ناهمبستگی آزمون‌ها (مؤلفه‌های توصیفگر) است تا هر آزمون نقش مؤثری در توصیف ویژگی داشته باشند لذا دو معیار همبستگی و واریانس در کیفیت توصیفگر تعیین کننده هستند. تحلیل انجام شده توسط روبلی و همکاران [۱۷] نشان می‌دهد که هم در توصیفگرهای BRIEF و هم BRIEF توجیه شده همبستگی مؤلفه‌ها زیاد است. از طرفی BRIEF توجیه شده واریانس بسیار کمتری نسبت به BRIEF دارد که می‌تواند ناشی از جهت‌گیری تصادفی نقاط کلیدی در BRIEF باشد؛ لذا توجیه‌کردن BRIEF هر چند توصیفگر را چرخش‌پذیر کرده است ولی تمایز آن را تضعیف می‌کند. به منظور بهبود این دو معیار یک روش یادگیری برای انتخاب زیرمجموعه مناسبی از آزمون‌های باینری ایجاد می‌شود. استراتژی به این ترتیب است که در یک مجموعه آموزشی بزرگ متشکل از حدود ۳۰۰ هزار نقطه کلیدی از تصاویر مجموعه PASCAL 2006 با روشی موسوم به Greedy Search، آزمون باینری برای ۲۰۵۵۹۰ جفت نقطه اجرا می‌شود. سپس با استفاده از PCA، ۲۵۶ ویژگی جدید شناسایی می‌شود که واریانس بالایی دارند و در یک مجموعه آموزشی بزرگ غیرهمبسته هستند. پس از استخراج الگوی مکانی این ۲۵۶ جفت نقاط بدست آمده، این الگو برای تولید توصیفگر باینری تمام نقاط کلیدی استفاده می‌شوند. به این معنا که در یک ناحیه مربعی ۳۱ × ۳۱ پیکسلی حول نقطه کلیدی در تصویر، طبق یک الگوی مکانی معین که بر اساس جهت نقطه کلیدی توجیه شده است، جفت پیکسل‌ها انتخاب شده و نتایج آزمون باینری آن‌ها توصیفگر ORB را تشکیل می‌دهد. نتیجه بکارگیری این روش در انتخاب نقاط که الگوریتم rBRIEF[3] نامیده می‌شود، بهبود قابل توجهی در واریانس و همبستگی نسبت به BRIEF توجیه‌شده بوجود می‌آورد لذا توصیفگر ایجاد شده دارای مقاومت و تمایزپذیری مناسبی است.

BRISK[4] یک توصیفگر باینری مقیاس‌پذیر ارائه می‌دهد که ترکیبی از سرعت ORB و مقاومت SURF است. این الگوریتم از یک الگوی نمونه‌برداری حلقوی برای تشخیص نقاط کلیدی و تولید توصیفگر استفاده می‌کند. این الگو شامل حلقه‌های مقیاس[5] برای پوشش نواحی مختلف در مقیاس‌های متفاوت برای تشخیص نقاط کلیدی و حلقه‌های مکانی[6] برای نمونه‌برداری از نقاط در فاصله‌های مختلف از مرکز برای تولید توصیفگر می‌باشد. در واقع توصیفگر BRISK از دو نوع مقایسه شامل مقایسه‌های مقیاس[7] و مقایسه‌های مکانی[8] تشکیل شده است و هر مقایسه یک بیت ایجاد می‌کند.

[1] Binary Robust Independent Elementary Features (BRIEF)
[2] Steered BRIEF
[3] Rotation-Aware BRIEF
[4] Binary Robust Invariant Scalable Key-points (BRISK)
[5] Scale Rings
[6] Spatial Rings
[7] Scale-space comparisons
[8] Spatial comparisons

الگوریتم KAZE[1] اولین روشی است که به‌جای فضای مقیاس خطی (مانند گاوسی در SIFT وSURF)، از فضای مقیاس غیرخطی بر پایه معادلات انتشار غیرخطی[2] برای استخراج ویژگی‌های کلیدی استفاده کرد. این نوآوری باعث شد KAZE در تصاویر پرنویز و با کنتراست پایین، عملکرد بهتری نسبت به روش‌های مرسوم داشته باشد. KAZE از یک آشکارساز و توصیفگر مشابه SURF به نام M-SURF[3] برای تشخیص نقاط کلیدی استفاده می‌کند که بر اساس محاسبه ماتریس هسین در فضای مقیاس غیرخطی عمل می‌کند و نقاط کلیدی را در فضای مقیاس غیرخطی توصیف کرده و یک توصیفگر حقیقی (معمولاً ۶۴ یا ۱۲۸ بعدی) تشکیل می‌دهد. با وجود دقت بالای KAZE ، سرعت پایین آن (ناشی از حل عددی کامل معادلات PDE) مانع از بکارگیری این الگوریتم در کاربردهای بلادرنگ می‌شد. برای رفع این مشکل، نسخه AKAZE[4] معرفی شد[۲۲] که به جای M-SURFاز نسخه بهینه‌شده‌ای از BRIEF به نام M-LDB[5]، استفاده می‌کند. M-LDB یک توصیفگر باینری است و در مرحله تطبیق فاصله همینگ را برای مقایسه بکار می‌برد. این تغییر سرعت تطبیق را به‌طور چشمگیری افزایش می‌دهد.

توصیفگرهای ویژگی سنتی برای استخراج ویژگی بر تبدیلات ریاضی صریح برای تجزیه و تحلیل تفاوت‌های مقیاس خاکستری بین تصاویر تکیه دارند در حالیکه رویکردهای مبتنی بر یادگیری عمیق به ویژگی‌های دست‌ساز صریح متکی نبوده بلکه از دانش الگوریتم‌های یادگیری عمیق بهره می‌برند. HardNet یک شبکه برای یادگیری تولید توصیفگر است که با تولید توصیفگرهای بهینه و دارای ساختار معنایی مناسب و تمایز لازم در صحنه‌های با بافت ضعیف، دقت تطبیق را بهبود می‌بخشد.[۲۳] هر چند توصیفگرهای متکی بر یادگیری عملکرد بهتری نسبت به توصیفگرهای سنتی دارند ولی اگر با مجموعه داده‌های محدودی آموزش داده می‌شوند، عملکرد آن‌ها به شدت تحت تأثیر نمونه‌های آموزشی قرار می‌گیرد بطوریکه در سناریوهایی با ساختارهای تار یا ضعیف عملکرد ضعیفی از خود نشان می‌دهند.

۳ـ۴ـ تطبیق ویژگی

در این مرحله، توصیفگرهای استخراج شده از دو تصویر با یکدیگر مقایسه می‌شوند. برای توصیفگرهای حقیقی، از فاصله اقلیدسی و برای توصیفگرهای باینری، از فاصله همینگ استفاده می‌شود. جفت‌هایی که فاصله آن‌ها کمتر از یک آستانه باشد، به عنوان جفت تطبیق‌یافته یا تطبیق‌های درست در نظر گرفته می‌شوند. در SIFT فاصله اقلیدسی هر توصیفگرهای تصویر الگو از تک‌تک توصیفگرهای تصویر صحنه محاسبه و در یک ماتریس تحت عنوان ماتریس فاصله ذخیره می‌شود. برای هر توصیفگر تصویر الگو کمترین فاصله نسبت به توصیفگرهای تصویر صحنه به عنوان نزدیکترین همسایه و کمترین فاصله بعدی به عنوان دومین همسایه نزدیک تعیین و نسبت آن‌ها به عنوان " نسبت فاصله‌ی نزدیک‌ترین همسایه[6]" محاسبه می شود. چنانچه NNDR کمتر از یک آستانه باشد یعنی نقطه کلیدی مربوط به توصیفگر مورد بررسی از تصویر الگو دارای تطبیق صحیحی در تصویر صحنه است و تطبیق متناظر آن نقطه‌ای از صحنه است که توصیفگر آن نزدیکترین همسایه را محقق کرده است؛ به این ترتیب این دو نقطه به عنوان یک جفت تطبیق صحیح ذخیره می‌شوند. اگر NNDR از یک آستانه بیشتر باشد، به معنی آن است که این دو همسایه نزدیک‌تر از مقدار مجاز بوده و تطبیق دارای ابهام بوده و تناظر رد می‌شود.

در اکثر الگوریتم‌ها مرحله تطبیق ویژگی به صورت دستی کدنویسی می‌شود لذا قسمت عمده هزینه زمانی و پیچیدگی اجرای الگوریتم‌ها مربوط به این مرحله است. از این نظر هر چه ابعاد توصیفگر کوچک‌تر بوده و حجم حافظه کمتری را اشغال کند پیچیدگی محاسباتی الگوریتم کاهش چشمگیری می‌یابد لذا اغلب بهینه‌سازی‌های صورت گرفته در الگوریتم‌ها معطوف به ساده‌سازی توصیفگرها می‌باشد. به طور مثال در ORB و AKAZE که از توصیفگرهای باینری استفاده می‌کنند، در مرحله تطبیق فاصله همینگ را برای مقایسه بکار می‌برند که این تغییر سرعت تطبیق را به‌طور چشمگیری افزایش می‌دهد.

۳ـ۵ـ تعیین تابع انتقال

با استفاده از جفت تطبیق‌های بدست آمده در مرحله‌ی قبل، تابع انتقال بین نقاط تصویر الگو و تصویر صحنه بگونه‌ای تعیین می‌شود که تحت این تابع انتقال بتوان هر نقطه از تصویر الگو را به نقطه متناظر در تصویر صحنه نگاشت کرد و بنابراین پس از تعیین این تابع انتقال، محل هر شیء در تصویر صحنه به‌سادگی بدست می‌آید. در [۱۳] این انتقال افین فرض شده است؛ البته در کاربردهای عملی ممکن است تابع انتقال از نوع پیچیده‌تری مانند انتقال پرسپکتیو باشد که در این شرایط می‌تواند پیش‌پردازش‌هایی روی تصویر الگو قبل ورود به مراحل جستجوی ویژگی اعمال شده و تابع انتقال به افین نزدیک شود.

در تابع انتقال افین رابطه‌ی بین نقاط متناظر دو تصویر به صورت $\begin{bmatrix}u\\v\end{bmatrix}=\begin{bmatrix}m_1 & m_2\\m_3 & m_4\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}+\begin{bmatrix}t_x\\t_y\end{bmatrix}$ مدل می‌شود که در این رابطه $(x,y)^{\mathrm{t}}$ مختصات نقطه‌ای در تصویر الگو، $(u,v)^{\mathrm{t}}$ مختصات نقطه‌ی نظیر آن در تصویر صحنه، $(t_x,t_y)^{\mathrm{t}}$ مدل جابجایی در دو راستای x و y و m_{i} پارامترهای چرخش و مقیاس می‌باشند؛ بنابراین ۳ تطبیق (که نقاط آن در امتداد یک خط نباشند) برای تعیین ۶ پارامتر تبدیل افین کفایت می‌کند.

[1] Key-point Alignment with Zoom and Enhancement
[2] Non-Linear Diffusion PDEs
[3] Modified SURF
[4] Accelerated-KAZE
[5] Modified Local Difference Binary
[6] Nearest Neighbor Distance Ratio (NNDR)

۳ـ۶ـ حذف نقاط ناسازگار[1]

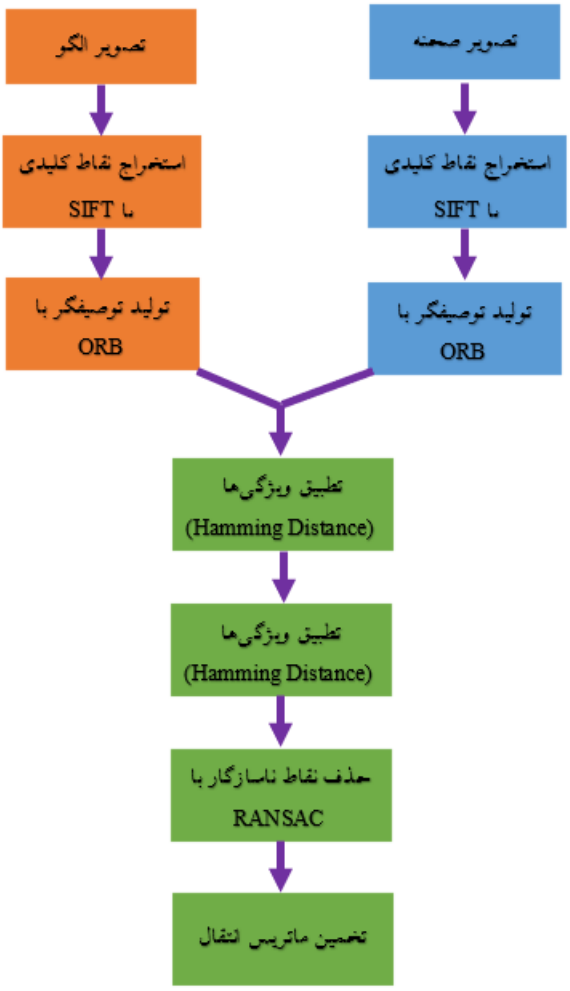
به دلیل وجود تطبیق‌های نادرست ناشی از پس‌زمینه و نویز ، تابع تبدیل برازش شده در مرحله قبل دارای خطا بوده و لازم است تأثیر نقاط ناسازگار در این تخمین حذف شود. به این منظور از روش‌هایی دارای حلقه تکرار مانند موارد زیر استفاده می‌شود. نمونه‌هایی از روش‌های متداول برای حذف نقاط ناسازگار، RANSAC[2] [۲۴] (روش معمول برای برازش مدل هندسی مانند همگرایی یا انتقال) و MAGSAC++ [۲۵] که نسخه بهبودیافته RANSAC می‌باشند.

یکی از روش‌های پیاده‌سازی RANSAC استفاده از رویکرد غیرخطی یا تطبیقی[3] است. در این روش برای هر تابع فرضیه‌ی افین یک p_n به‌صورت [۲۶] $p_n = \frac{N_{in}}{N_{Total}}$ تعریف می‌شود که در این رابطه N_{in} تعداد نمونه‌هایی است که فرضیه‌ی nام را تایید می‌کنند و N_{Total} تعداد کل تطبیق‌ها است. پس از محاسبه‌ی p_n برای هر فرضیه، N_n با قرار دادن p_n در معادله‌ی $N = \frac{\log(1-P)}{log(1-p^k)}$ محاسبه می‌شود. تا زمانی که تعداد فرضیه‌های (n) ساخته شده کوچکتر از مینیمم N_n ها باشد فرضیه‌ساختن ادامه پیدا می‌کند. یک روش بهینه‌تر از نظر هزینه محاسباتی استفاده از تمام جفت تطبیق‌ها در تخمین اولیه تابع انتقال است. پس از محاسبه تابع انتقال، تمام تطبیق‌ها بر این تابع فرضیه اعمال شده و تطبیق‌های ناسازگار کنار گذاشته شده و فرآیند تکرار می‌شود. معادله تبدیل افین پس از جداسازی مقادیر معلوم و پارامترهای مجهول و به ازای تمام جفت تطبیق‌ها به صورت بیان شده در رابطه (۱۳) قابل بازنویسی است:

$$\begin{bmatrix} x_1 & y_1 & 0 & 0 & 1 & 0 \\ 0 & 0 & x_1 & y_1 & 0 & 1 \\ x_2 & y_2 & 0 & 0 & 1 & 0 \\ 0 & 0 & x_2 & y_2 & 0 & 1 \\ & \cdots & & & \cdots & \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ t_x \\ t_y \end{bmatrix} = \begin{bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \\ \vdots \end{bmatrix} \qquad (۱۳)$$

این سیستم خطی را می‌توان به شکل $AX = b$ نوشت و پارامترهای تابع انتقال افین با معادله $X = A^{-1}\, b$ محاسبه می‌شوند. با توجه به تعداد زیاد نقاط متناظر استخراج شده در یک تصویر معمول بدست آوردن A^{-1} پیچیدگی محاسباتی بالایی دارد. به این منظور می‌توان از ویژگی‌های شبه‌وارون[4] بهره برد و پارامترهای انتقال افین را از رابطه $X = (A^T . A)^{-1} A^T . b$ تعیین کرد که از نظر هزینه محاسباتی بسیار بهینه‌تر است. تابع انتقال بدست آمده را بر تمام تطبیق‌ها اعمال کرده و فاصله نقاط نگاشت شده‌ی تحت تابع انتقال $(\hat{u}, \hat{v})^t$ با نقطه متناظر ناشی از تطبیق توصیفگرها با بکارگیری رابطه‌ی $\varepsilon = \sqrt{(u-\hat{u})^2 + (v-\hat{v})^2}$ مقایسه می‌شوند. چنانچه خطای بدست آمده کمتر از ۵ پیکسل باشد، زوج تطبیق مورد بررسی برای این تابع انتقال سازگار[5] بوده و در غیر این صورت ناسازگار[6] است و از فرآیند تخمین تابع انتقال کنار گذاشته می‌شود. در مرحله بعد مجدداً تابع انتقال با بردار X , b متشکل از نقاط سازگار مرحله قبل قبل محاسبه شده و نقاط سازگار و ناسازگار برای تابع انتقال جدید تعیین می‌شود. این فرآیند تا جایی تکرار می‌شود که موارد ناسازگار به کمتر از ۵٪ تطبیق‌های باقیمانده برسد؛ تابع بدست آمده در این مرحله به عنوان تابع انتقال بین تصویر الگو و صحنه لحاظ می‌شود.

LightGlue نمونه‌ای از الگوریتم تطبیق هوشمند است که سریع‌تر و دقیق‌تر از RANSAC در مرحله حذف نقاط ناسازگار و تخمین تابع انتقال دقیق عمل می‌کند.[۲۷] در واقع تطبیق‌دهنده‌های مبتنی بر شبکه عصبی هنوز در مرحله توسعه هستند و این روش‌ها به یک نقطه تمرکز تحقیقات در زمینه تطابق ویژگی‌ها تبدیل شده است. با این حال، این روش‌ها از مشکلات زیر رنج می‌برند: (۱) آنها اغلب به مجموعه اولیه تطابق‌ها یا نقاط کلیدی اولیه متکی هستند، که کشف تطابق‌های اضافی با در نظر گرفتن تصاویر تطبیق یافته اصلی را دشوار می‌کند. (۲) برخی از روش‌های مبتنی بر توجه به دقت بالایی دست می‌یابند، اما در بکارگیری این الگوریتم‌ها چالش پیچیدگی محاسباتی به ویژه هنگام برخورد با تعداد زیادی نقطه



شکل(۸): ساختار الگوریتم SCO

[1] Outlier Rejection
[2] Random Sample Consensus
[3] Non-linear RANSAC or Adaptive RANSAC
[4] Pseudo inverse
[5] Inlier
[6] Outlier

کلیدی، که منجر به افزایش قابل توجهی در پیچیدگی و حجم منابع پردازشی، حافظه و توان و احیاناً لزوم استفاده از سخت‌افزار GPU مطرح است.

رویکرد سرتاسری مبتنی بر شبکه عصبی، تطبیق ویژگی‌ها شامل استفاده از یک شبکه عصبی برای انجام هر دو مرحله استخراج ویژگی و تطبیق ویژگی است. این رویکرد با تولید مستقیم رابطه تطبیق صحیح از تصاویر ورودی، با روش‌های سنتی دو مرحله‌ای تفاوت دارد و در واقع یک شبکه عصبی پیچیده آموزش داده می‌شود تا مستقیماً از تصویر ورودی، نتیجه‌ی نهایی را پیش‌بینی کند. مزیت رویکرد سرتاسری این است که شبکه یاد می‌گیرد که به طور خودکار ویژگی‌های مهم را استخراج کند، آن‌ها را با یکدیگر مقایسه کند و تطابق‌ها را پیدا کند و این یکپارچگی منجر به تطبیق‌های بهتر می‌شود ولی این شبکه‌ها پیچیدگی بیشتری هم از نظر پیاده‌سازی و هم منابع محاسباتی مورد نیاز دارند و از طرفی وابستگی آن‌ها به داده‌های آموزشی بیشتر است. SuperGlue نمونه‌ای از روش سرتاسری است [۲۷].

۴- تحلیل عملکرد الگوریتم‌های تطبیق

الگوریتم SIFT به دلیل مقاومت بسیار بالا در برابر تبدیلات هندسی و فوتومتریک، به عنوان یک معیار طلایی در استخراج ویژگی شناخته شده و پایه بسیاری از روش‌های بعدی قرار گرفته است. بررسی روند توسعه الگوریتم‌های تطبیق کلاسیک بعد از SIFT نشان می‌دهد که تمرکز اغلب این روش‌ها، بهینه‌سازی زمان اجرای الگوریتم SIFT با بکارگیری ابتکاراتی در مراحل آشکارسازی نقاط کلیدی و تولید توصیفگر می‌باشد. از طرفی ORB یکی از شاخص‌ترین الگوریتم‌های توسعه داده شده برای کاربردهای بلادرنگ با تمرکز با کاهش حجم محاسباتی در مرحله تطبیق توصیفگرها است. لذا در این مقاله، با رویکرد توسعه الگوریتم تطبیق با ویژگی بهبود سرعت اجرا و کاهش پیچیدگی محاسباتی در عین حفظ مقاومت مناسب الگوریتم تطبیق در برابر تغییرات بصری تصویر است و قابلیت پیاده‌سازی در سیستم‌های نهفته با منابع سخت‌افزارهای محدود و کاربردهای بلادرنگ، یک تلفیق هوشمندانه از SIFT و ORB پیشنهاد تحت عنوان الگوریتم SCO پیشنهاد می‌شود.

در مرحله‌ی استخراج نقاط کلیدی، SIFT نقاطی با دقت تطبیق بالا و مقاوم نسبت به چرخش، مقیاس، نویز و تغییرات شدت نور را استخراج می‌کند در حالیکه ORB در این زمینه کیفیت پایین‌تری دارد. در مراحل بعد SIFT توصیفگرهای ۱۲۸ بُعدی از نوع عدد اعشاری شناور تولید می‌کند که علیرغم دقت بالا، پیچیدگی محاسباتی زیادی هم در مرحله تولید توصیفگرها (روش مبتنی بر آمار گرادیان) و هم در انطباق ویژگی‌ها در دو تصویر دارد؛ در حالی که ORB با بکارگیری الگوهای حاصل از یادگیری بدست آمده از پایگاه داده‌های متنوع تحت عنوان ORB-pattern، توصیفگرهای باینری ارائه می‌دهد که با وجود دقت بالا، سرعت اجرای مناسب و پیچیدگی محاسباتی بسیار پایین در تولید توصیفگرها داشته و تطبیق آن‌ها دارد. به این ترتیب تلفیقی از SIFT در مرحله تشخیص نقاط کلیدی و ORB در بخش تولید توصیفگرها که ساختار آن در شکل (۸) نشان داده شده است، می‌تواند ایده بسیار جذابی برای بهینه‌سازی از نظر سرعت اجرا و پیچیدگی محاسباتی باشد.

در این بخش، عملکرد الگوریتم‌های شاخص تطبیق صحنه شامل SIFT، SURF، ORB، BRISK، KAZE، AKAZE و الگوریتم پیشنهادی SCO به روش شبیه‌سازی تحلیل می‌شود. مقایسه‌ها از جنبه‌های سرعت اجرا، تعداد ویژگی‌های استخراج‌شده و تعداد تطبیق‌های درست خواهد بود. شبیه‌سازی در یک محیط نرم‌افزاری با استفاده از سخت‌افزار با مشخصات مندرج در جدول (۱) انجام شد.

جدول(۱): سخت افزار و نرم افزار مورد استفاده در شبیه‌سازی

مولفه ها	مقادیر مولفه ها
سیستم عامل	Ubuntu 24.04.2 LTS
مشخصات CPU	Intel® Core™ i5-8400 CPU 2.80GHZ
RAM	16G Byte
نسخه Python	3.12.4
نسخه OpenCV	4.12.0

در فرآیند ارزیابی عملکرد الگوریتم‌ها دو نمونه تصویر یکی با بافت شلوغ و ویژگی‌های قوی و دیگری با ساختار یکنواخت و بافت دارای ویژگی‌های ضعیف استفاده شده است که در مراحلی اغتشاشاتی مانند افزودن نویز، تغییر کنتراست و روشنایی[1] روی آن‌ها اعمال می‌شود. علاوه بر این عملکرد الگوریتم‌ها پایگاه تصاویر استاندارد اکسفورد[۲۸] که دارای ۶ دسته از تصاویر با اعمال تبدیلاتی چون چرخش، تغییر مقیاس، بلورشدگی، تغییرات روشنایی و اثرات فشرده‌سازی است نیز مورد استفاده قرار گرفته است. معیارهای ارزیابی عملکرد تعداد تطبیق‌های درست (تطبیق‌های سازگاربرای تابع انتقال بدست آمده) و فراخوانی[2] و عدم دقت[3] است. فراخوانی مطابق رابطه (۱۴) در واقع میزان توانایی الگوریتم در تعیین تطبیق‌های درست از کل نواحی متناظر[4] در دو تصویر را بیان می‌کند.

$$recall = \frac{\#\ correct\ matches}{\#\ correspondences} \qquad (۱۴)$$

در رابطه (۱۴)، صورت کسر تعداد تطبیق‌های درست و مخرج کسر تعداد ناحیه یا نقاط متناظر دو تصویر است. نواحی متناظر در واقع مجموع تطبیق درست[5] و عدم تطبیق نادرست[6] (نقاط متناظری که

[1] Brightness
[2] Recall
[3] 1-Precision
[4] Corresponding regions
[5] True Positive
[6] False Negative

الگوریتم قادر به تشخیص آن‌ها نشده است) می‌باشد. برای پایگاه تصاویر استاندارد استفاده شده، تعداد نقاط متناظر که همان مقدار مخرج کسر است به صورت استاندارد محاسبه و برابر ۸۸۷ تناظر تعیین شده است.

$$1 - percision = \frac{\#\ false\ matches}{\#false\ matches\ +\ \#\ correct\ matches} \quad (۱۵)$$

معیار عدم دقت مطابق با رابطه (۱۵) به صورت نسبت تطبیق‌های نادرست به کل تطبیق‌های خروجی هر الگوریتم تعریف می‌شود. در واقع این رابطه معیاری از عدم دقت یعنی میزان تطبیق‌های اشتباه به کل تطبیق‌ها را بیان می‌کند. هر چه فراخوانی بیشتر باشد به معنای قابلیت بالاتر الگوریتم در تشخیص تطبیق‌های صحیح بوده و هر چه عدم دقت پایین‌تر باشد به معنی کمتر بودن تشخیص‌های نادرست الگوریتم است.

۴ـ۱ـ مقایسه عملکرد الگوریتم‌های تطبیق روی تصاویر با بافت شلوغ

شکل (۸) نتایج بصری اجرای الگوریتم های تطبیق را روی جفت تصاویر بافت شلوغ نشان می‌دهد. طبق جدول (۲) برای تصاویر دارای ویژگی‌های بصری بارز، زمان اجرای الگوریتم ORB نسبت به سایرین مینیمم است و الگوریتم AKAZE در جایگاه دوم و روش SCO در رتبه بعد قرار دارد. از لحاظ تعداد تطبیق‌های صحیح نیز برای روش SCO در محدوده قابل قبول قرار دارد و در مقایسه با SIFT ضعیف‌تر و نسبت به ORB بهتر است.

جدول(۲): نتایج عملکرد الگوریتم‌ها روی تصاویر با بافت شلوغ

الگوریتم	زمان اجرا بر حسب ثانیه	تعداد نقاط کلیدی	تعداد نقاط تطبیق درست
SIFT	2.1759	9910	2521
ORB	0.7657	500	141
BRISK	3.1207	34071	4026
KAZE	5.7663	15717	3781
AKAZE	0.9812	12138	1889
SURF	1.4418	2138	1392
SCO	1.0320	9910	1382

جدول(۳): نتایج عملکرد الگوریتم‌ها روی تصاویر با بافت شلوغ بلوری و نویزی

الگوریتم	Blur_1.5	Blur_2.0	Noise_5	Noise_10
SIFT	27	11	750	160
ORB	0	0	114	18
BRISK	0	0	2255	311
KAZE	0	0	3720	390
AKAZE	26	1	2937	461
SURF	6	1	244	102
SCO	4	1	220	3

شکل(۸): خروجی تطبیق الگوریتم‌ها روی جفت تصاویر بافت شلوغ، الف) SIFT (ب) ORB (ج) SCO

جدول(۴): نتایج عملکرد الگوریتم‌ها روی تصاویر با بافت شلوغ دارای تغییرات کنتراست و روشنایی

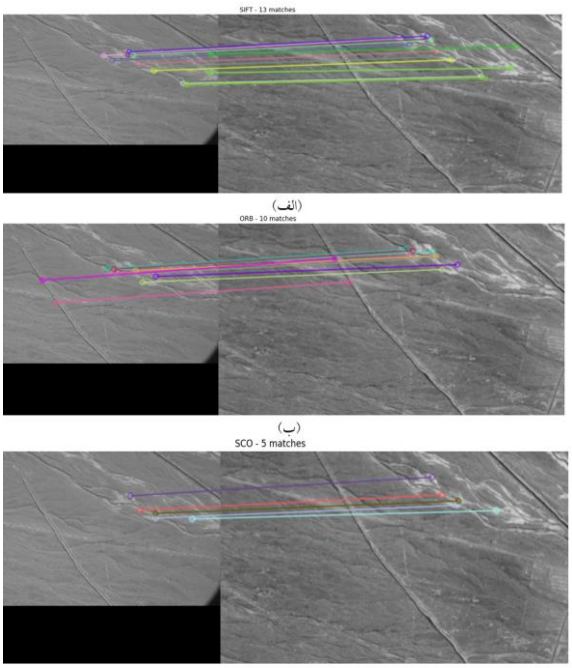
الگوریتم	Contrast_0.50	Contrast_2.0	Brightness_-50	Brightness_50
SIFT	6989	6257	7511	7585
ORB	309	273	371	340
BRISK	22555	19420	18842	25456
KAZE	11676	10854	9971	12589
AKAZE	9736	9050	8022	10364
SURF	3447	3089	3589	3738
SCO	3437	3124	3677	3764

به صورت کمی می‌توان گفت زمان اجرای روش SCO حدود ۰/۴۷ زمان اجرای SIFT و ۱/۳۵ برابر زمان اجرای ORB است در حالیکه تعداد تطبیق‌های درست SCO حدود ۰/۵۵ تطبیق‌های صحیح SIFT و ۹/۸ برابر تطبیق‌های ORB می‌باشد.

جداول (۳) و (۴) نشان می‌دهند که الگوریتم‌های نسبت به بلور شدگی حساس هستند در حالیکه نسبت به نویز و تغییرات روشنایی و کنتراست مقاوم‌اند.

۴ـ۲ـ مقایسه عملکرد الگوریتم‌های تطبیق روی تصاویر با بافت ضعیف

شکل (۹) نتایج بصری اجرای الگوریتم‌های تطبیق را روی جفت تصاویر بافت ضعیف نشان می‌دهد.



شکل(۹): خروجی تطبیق الگوریتم‌ها روی جفت تصاویر بافت ضعیف،(الف) SIFT (ب) ORB (ج) SCO

در جدول(۵) نتایج عملکرد الگوریتم‌های تطبیق مختلف روی تصاویر بافت ضعیف بیان شده است.

جدول(۵): نتایج عملکرد الگوریتم‌ها برای تصویر با بافت ضعیف

الگوریتم	زمان اجرا (ثانیه)	تعداد نقاط کلیدی	تعداد نقاط تطبیق درست
SIFT	0.0409	29	13
ORB	0.00806	188	10
BRISK	0.0053	75	5
KAZE	0.0744	0	0
AKAZE	0.0090	0	0
SURF	0.0227	36	9
SCO	0.0203	29	5

همان‌طور که در جدول (۵) مشاهده می‌شود، در تصاویر با بافت ضعیف الگوریتم‌های SIFT و ORB عملکرد نسبتاً مناسبی دارند، اما الگوریتم پیشنهادی SCO در این نوع تصاویر ضعیف‌تر عمل می‌کند .به این دلیل که در این الگوریتم از آشکارساز SIFT استفاده شده است، تعداد نقاط کلیدی تشخیص داده شده در SCO مشابه SIFTبوده و کمتر از ORB است.

در مرحله تولید توصیفگر و تطبیق آن‌ها SCO با بکارگیری تعداد نقاط کلیدی کمتر نسبت به ORB، موفقیت کمتری را با بکارگیری rBRIEF نسبت به ORB داشته و نقاط متناظر کمتری را تأمین کرده است. به این ترتیب می‌توان گفت زمان اجرای روش SCO به طور تقریبی حدود ۰/۵ برابر زمان اجرای SIFT و ۲/۵۲ برابر زمان اجرای ORB است، در حالی که تعداد تطبیق‌های درست روش SCOحدود ۰/۳۸ برابر تطبیق‌های صحیح SIFT و ۰,۵ برابر تطبیق‌های ORB می‌باشد.

بررسی عملکرد الگوریتم‌های تطبیق در تصاویر با بافت ضعیف به منظور ارزیابی آن‌ها در پیاده‌سازی روی سخت‌افزارهای دارای سیستم نهفته با کاربرد ناوبری تصویری است. با توجه نتایج بدست آمده، راهکار پیشنهادی بررسی لحظه‌ای تعداد تطبیق‌های درست و مقایسه آن با یک سطح آستانه بهینه و فعال کردن موقت الگوریتم SIFT بجای SCO حین عبور از مناطق با تصاویر با بافت ضعیف می‌باشد.

۴ـ۳ـ مقایسه عملکرد الگوریتم‌های تطبیق روی پایگاه داده استاندارد

در جداول (۶) تا (۹) نتایج عملکرد الگوریتم‌های تطبیق مختلف روی تصاویر پایگاه داده استاندارد بیان شده است.

نتایج ارزیابی نشان می‌دهد که زمان اجرای روش SCO به طور متوسط حدود ۰/۲۸ زمان اجرای SIFT و ۲/۶۷ برابر زمان اجرای ORBاست در حالیکه تعداد تطبیق‌های درست SCO به طور متوسط حدود ۰/۷۳ تطبیق‌های صحیح SIFT و حدود ۳/۳ برابر تطبیق‌های ORB می‌باشد. معیار فراخوانی برای SIFT به طور متوسط ۲۱/۶٪، برای ORB به طور متوسط ۱۱/۰۵٪ و برای SCO به طور متوسط ۱۶/۳۹٪ است. معیار عدم دقت برای SIFT به طور متوسط ۷/۷۸٪، برای ORB به طور متوسط ۱۵/۳۸٪ و برای SCO به طور متوسط ۱۱/۳۵٪ می‌باشد. به طور خلاصه می‌توان گفت روش پیشنهادی SCO از لحاظ زمان پردازش مابین SIFT و ORB است و در کلیه موارد به جز Zoom از لحاظ عملکردی بین SIFT و ORB قرار دارد.

بر اساس نتایج جداول مقایسه‌ای شماره ۶ تا ۹، هر یک از سه الگوریتمSIFT ، ORB و روش ترکیبی پیشنهادی (SCO) دارای کارایی متمایزی هستند. الگوریتم SIFT بالاترین دقت تطبیق را در اغلب شرایط به‌دست می‌دهد (میانگین نرخ فراخوانی حدود ۲۱/۶٪ و کمترین نرخ تطابق اشتباه حدود ۷/۸٪). در مقابل، الگوریتم ORBاگرچه بسیار سریع‌ترین زمان اجرا را دارد، اما نرخ فراخوانی کمتری (میانگین حدود ۱۱٪) و خطای بیشتری (حدود ۱۵٪) نشان می‌دهد. روش ترکیبی (SCO) SIFT-ORB توانسته است از نظر سرعت و دقت عملکردی مابین دو الگوریتم مذکور ارائه دهد؛ به‌طوری‌که به طور متوسط زمان اجرای آن حدود ۰/۲۸ برابر SIFT (یعنی تقریباً یک‌سوم) و ۲/۶۷ برابر ORB بوده و از نظر تعداد تطابق صحیح نیز حدود ۰/۷۳ برابر SIFT و ۳/۳ برابر ORB به‌دست آورده است. این نتایج کلی تأیید می‌کند که روش SCO از نظر سرعت پردازش بسیار کارآمدتر از SIFT بوده و در عین حال نسبت به ORB تطابق‌های صحیح به‌مراتب بیشتری پیدا می‌کند. نرخ فراخوانی متوسط SCO حدود ۱۶/۴٪ (میان SIFT و ORB) و نرخ تطابق اشتباه آن حدود ۱۱/۳٪ گزارش شده است که بین مقادیر دو روش دیگر قرار می‌گیرد. به طور خلاصه، الگوریتم پیشنهادی ترکیبی در اکثر سناریوها موفق شده تعادلی میان دقت و سرعت برقرار کند و عملکردی بین SIFT و ORB داشته باشد.

جدول (۶): مقایسه عملکرد الگوریتم‌ها روي تصاویر بلورشده

الگوریتم	زمان اجرا (ثانیه)	تعداد نقاط کلیدی	تعداد تطبیق درست	تعداد تطبیق نادرست	Recall (%)	1-Precision (%)
SIFT	0.0849	1402	201	16	22.6	7.4
ORB	0.00806	500	143	24	16.1	14.4
SCO	0.0203	1402	174	21	19.6	10.8
SCO/SIFT	0.24	1	0.86	-	-	-
SCO/ORB	2.52	2.8	1.22	-	-	-

جدول (۷): مقایسه عملکرد الگوریتم‌ها روي تصاویر با تغییر روشنایی

الگوریتم	زمان اجرا (ثانیه)	تعداد نقاط کلیدی	تعداد تطبیق درست	تعداد تطبیق نادرست	Recall (%)	1-Precision (%)
SIFT	0.0637	985	168	15	18.9	8.2
ORB	0.00555	680	114	25	12.8	16.2
SCO	0.0203	985	139	19	15.67	12.0
SCO/SIFT	0.32	1	0.83	-	-	-
SCO/ORB	3.66	1.45	1.22	-	-	-

جدول (۸): مقایسه عملکرد الگوریتم‌ها روي تصاویر با اثرات فشرده‌سازی

الگوریتم	زمان اجرا (ثانیه)	تعداد نقاط کلیدی	تعداد تطبیق درست	تعداد تطبیق نادرست	Recall (%)	1-Precision (%)
SIFT	0.0811	1624	232	21	26.2	8.3
ORB	0.0097	680	22	4	2.5	15.4
SCO	0.0203	1624	226	27	25.4	10.6
SCO/SIFT	0.25	1	0.97	-	-	-
SCO/ORB	2.09	2.39	10.3	-	-	-

جدول (۹): مقایسه عملکرد الگوریتم‌ها روي تصاویر اثرات تغییر مقیاس

الگوریتم	زمان اجرا (ثانیه)	تعداد نقاط کلیدی	تعداد تطبیق درست	تعداد تطبیق نادرست	Recall (%)	1-Precision (%)
SIFT	0.0637	985	168	13	18.9	7.2
ORB	0.0084	584	114	21	12.8	15.5
SCO	0.0203	985	44	6	4.9	12.0
SCO/SIFT	0.32	1	0.26	-	-	-
SCO/ORB	2.42	1.69	0.39	-	-	-

تحلیل جزئی‌تر نتایج نشان می‌دهد که در سناریوی تصاویر تاری (جدول ۶)، الگوریتم SIFT با تشخیص ۲۰۱ تطابق صحیح، بالاترین فراخوانی (~۲۲/۶٪) را کسب کرده است. در این حالت ORBتنها ۱۴۳ تطابق صحیح (فراخوانی ~۱۶/۱٪) پیدا کرده و روش ترکیبی SCO با ۱۷۴ تطابق صحیح (فراخوانی ~۱۹/۶٪) عملکردی مابین آن دو دارد. از نظر صحت تطابق‌ها نیز SIFT دقت بالاتری نشان داده و درصد تطابق‌های اشتباه آن فقط ~۷/۴٪ بوده است، در حالی که این میزان برای ORB به ~۱۴/۴٪ رسیده و SCOخطایی حدود ~۱۰/۸٪ داشته است. نکته قابل توجه دیگر در تصاویر تار، اختلاف قابل ملاحظه در زمان پردازش است : SIFTبه دلیل استخراج ویژگی‌های پرتعداد و توصیف‌گرهای پرهزینه، کندترین روش بوده (حدود ۰/۰۸۵ ثانیه برای هر جفت تصویر)؛ ORB با پردازش بسیار چابک (~۰/۰۰۸ ثانیه) سریع‌ترین الگوریتم محسوب می‌شود، و SCO هرچند نسبت به ORBکندتر است (حدود ۲/۵ برابر زمانORB)، اما زمان اجرای آن تقریباً یک‌چهارم زمان SIFT بوده است. به بیان دیگر، روش ترکیبی با کاهش چشمگیر هزینه محاسباتیSIFT ، هنوز بخش عمده‌ای از تطابق‌های درست آن را حفظ کرده است. الگوریتم SIFT اگرچه در تصاویر تار بهترین فراخوانی را دارد، اما سایر روش‌ها (خصوصاًORB) در این شرایط افت محسوسی در تعداد تطابق‌های صحیح نشان می‌دهند که بیانگر حساسیت بیشتر آن‌ها به محوشدگی تصویر است.

در سناریوی تغییر شدید روشنایی و کنتراست (جدول ۷)، روندی مشابه مشاهده می‌شود. الگوریتم SIFT همچنان بالاترین کارایی را ارائه داده و توانسته است ~۱۶۸ تطابق صحیح را شناسایی کند (فراخوانی ~۱۸/۹٪) در حالی که ORB حدود ۱۱۴ تطابق صحیح داشته است (فراخوانی ~۱۲/۸٪). روش SCO با پیدا کردن ~۱۳۹ تطابق صحیح (فراخوانی ~۱۵/۷٪) در حد فاصل دو روش دیگر عمل کرده است. از منظر دقت تطبیق نیز، SIFT کمترین خطا (~۸/۲٪ تطابق اشتباه) را داشته و ORB بیشترین (~۱۶/۲٪)؛ SCO نیز نرخ خطای میانی (~۱۲٪) را نشان داده است. در این سناریو نیز ORB بسیار سریع‌تر از رقبا بوده (حدود ۰/۰۰۶ ثانیه) و SIFT کندترین (۰/۰۶۴ ثانیه)، در حالی که زمان SCO(~۰/۰۲ ثانیه) تقریباً میانه‌ی آن‌ها است. جالب این‌که تغییرات شدید نور و کنتراست تأثیر مخربی بر دقت هیچ‌یک از روش‌ها نداشته و همگی همچنان درصد قابل توجهی از تطابق‌های صحیح را حفظ کرده‌اند. این موضوع بیانگر مقاومت نسبی الگوریتم‌هایSIFT ، ORB و ترکیبی در برابر دگرگونی‌های روشنایی است.

اما در شرایط وجود اعوجاج‌های نویزی ناشی از فشرده‌سازی (جدول ۸)، تفاوت عملکرد الگوریتم‌ها بارزتر می‌شود. الگوریتم SIFTدر برابر نویز فشرده‌سازی بسیار پایدار بوده و با تشخیص ۲۳۲ نقطه متناظر صحیح (فراخوانی ~۲۶/۲٪) تقریباً هیچ افتی نسبت به حالت عادی نشان نداده است. روش ترکیبی SCO نیز عملکردی نزدیک به SIFT داشته و ۲۲۶ تطابق صحیح یافته است (فراخوانی ~۲۵/۴٪) .در مقابل، ORBبه‌شدت تحت تأثیر نویز فشرده‌سازی قرار گرفته و تنها ۲۲ تطابق صحیح را برقرار کرده است (فراخوانی ~۲/۵٪). این کاهش چشمگیر نشان می‌دهد که توصیف‌گر دودویی ORB (BRIEF) در برابر مصنوعات فشرده‌سازی JPEG مقاومت چندانی ندارد، در حالی‌که توصیف‌گر SIFT(مبتنی بر گرادیان) و نیز ترکیب SCO بسیار مقاوم‌تر عمل کرده‌اند. دقت تطبیق (درصد صحت) نیز برای SIFT (~۹۱/۷٪ صحیح، خطای ~۸/۳٪) و SCO (~۸۹٪ صحیح، خطای ~۱۰/۶٪) بالا باقی مانده، اما ORB با خطای ~۱۵/۴٪ تطابق‌های نادرست، کم‌دقت‌ترین نتایج را ارائه کرده است. به بیان دیگر، در حضور نویز شدید، روش ORB بخش عمده‌ای از تطابق‌های ممکن را از دست می‌دهد در حالی که SIFT و الگوریتم ترکیبی تقریبا تمامی تطابق‌های صحیح را حفظ می‌کنند. سرعت ORB کماکان بالاترین بوده (۰/۰۱ ثانیه) و هرچند SIFT حدود ۰/۰۸ ثانیه زمان برده، روش SCO با ~۰/۰۲ ثانیه زمان اجرا نشان می‌دهد که می‌توان بدون افت محسوس دقت، زمان پردازش را به‌طور چشمگیری نسبت به SIFT کاهش داد.

سخت‌ترین سناریو برای روش ترکیبیSCO ، آزمون تغییر مقیاس زیاد (Zoom) است (جدول ۹). نتایج نشان می‌دهد که در تصاویر با تغییر اندازه شدید، الگوریتم SIFT مطابق انتظار بهترین عملکرد را داشته و ۱۶۸ تطابق صحیح (فراخوانی ~۱۸/۹٪) یافته است. الگوریتم ORB در این حالت ۱۱۴ تطابق صحیح برقرار کرده (فراخوانی ~۱۲/۸٪) که کمتر از SIFT است اما بالاتر از روش ترکیبی .SCO روش SCO تنها ۴۴ نقطه متناظر صحیح میان تصاویر مقیاس متفاوت تشخیص داده است (فراخوانی ~۴/۹٪) که حتی از ORB نیز پایین‌تر بوده و بیانگر ضعف این روش در مواجهه با تغییرات مقیاس بزرگ است. علت این افت، آن است که اگرچه آشکارساز قوی SIFT در SCO به‌کار رفته و نقاط کلیدی پایدار را شناسایی می‌کند، اما توصیف‌گر ORB (rBRIEF) علی‌رغم چرخش‌ناپذیری، در برابر تغییرات مقیاس به اندازه توصیف‌گر برداری SIFT مقاوم نیست. در نتیجه در سناریویZoom ، بسیاری از نقاط کلیدی شناسایی‌شده توسط SIFT با توصیف‌گر ORBقابل تطبیق دقیق نبوده و تطابق‌های صحیح کاهش می‌یابد. این نکته در جمع‌بندی نتایج هم تأکید شده که تنها در سناریوی بزرگ‌نمایی تصویر عملکرد SCO پایین‌تر از ORB قرار گرفته است، در حالی‌که در سایر موارد این روش مابین دو الگوریتم عمل می‌کند. از نظر دقت تطبیق نیز SIFT حتی در این شرایط کمترین خطا (~۷/۲٪ تطابق اشتباه) را داشته و ORB بیشترین (~۱۵/۵٪)، در حالی که SCO با خطای ~۱۲٪ میان آن‌ها است. بنابراین ترکیب SIFT و ORB در بحث مقیاس‌پذیری کامل تصویر دچار ضعف بوده که می‌تواند یکی از زمینه‌های بهبود در کارهای آتی باشد.

با کنار هم گذاشتن یافته‌های فوق، می‌توان جمع‌بندی کرد که هر یک از این الگوریتم‌ها دارای نقاط قوت و ضعف مشخصی هستند . SIFT به عنوان یک روش کلاسیک و دقیق، تقریباً در همه سناریوها بیشترین تعداد تطابق صحیح را استخراج می‌کند و نسبت به اغلب اعوجاج‌های هندسی و تغییرات شدت روشنایی بسیار پایدار است (به ویژه در برابر نویز و فشرده‌سازی مقاومت عالی دارد). نقطه قوت اصلی SIFT دقت بالا و قابلیت اعتماد آن در تطبیق ویژگی‌های پیچیده است. با این حال، نقطه ضعف بارز SIFT کند بودن آن و نیاز به منابع محاسباتی بیشتر است که استفاده از آن را در کاربردهای بلادرنگ محدود می‌سازد. در مقابل، ORB مزیت عمده‌اش سرعت بسیار بالا و سبُک بودن الگوریتم است؛ این روش حتی روی سخت‌افزارهای نهفته با منابع محدود قابل اجراست و برای کاربردهای بلادرنگ مناسب ارزیابی می‌شود .ORB همچنین در صحنه‌های بافت‌کم می‌تواند نسبت به SIFT نقاط کلیدی بیشتری بیابد (زیرا آشکارساز FAST در سطوح مختلف تصویر فعال است) و در محیط‌های با نور متفاوت یا نویز ملایم نیز تطابق‌های نسبتاً قابل قبولی ارائه می‌کند .ضعف‌های ORB اما در دقت تطبیق و حساسیت آن به برخی اعوجاج‌ها نمایان می‌شود؛ چنان‌که دیدیم در تصاویر بسیار پرجزئیات، ORB به دلیل محدودیت تعداد ویژگی (۵۰۰ نقطه) تطابق‌های بسیار کمتری از SIFT برقرار کرد، و نیز در شرایط نویز شدید (مانند فشرده‌سازی) بخش بزرگی از تطابق‌های ممکن را از دست می‌دهد. به طور کلی ORB نسبت به روش‌های دقیق‌تر، نرخ فراخوانی پایین‌تر و خطای بالاتری دارد که آن را بیشتر برای کاربردهایی مناسب می‌سازد که سرعت و منابع محدود اولویت بالاتری نسبت به حداکثر دقت دارند.

روش پیشنهادی SCO که ترکیبی از SIFT و ORB است، تلاش کرده تا مزایای دو رویکرد را همزمان به‌دست‌آورد. نتایج نشان داد که این روش تا حد زیادی تعادل مؤثری بین دقت و سرعت ایجاد کرده است SCO :در اکثر سناریوها توانست تعداد تطابق‌های صحیح را به‌مراتب بیشتر از ORB (گاهی چندین برابر) فراهم کند، در حالی که زمان اجرای آن تنها کسری از زمان SIFT بود. به عنوان نمونه، در همه مجموعه تصاویر استاندارد (به‌جز حالت تغییر مقیاس زیاد)، SCO موفق شد فراخوانی خود را میان SIFT و ORB نگه دارد و همزمان از نظر سرعت پردازش نیز فاصله‌ی زیادی با ORB نداشت. این نقطه قوت عمده‌ی روش ترکیبی است که با کاهش ۷۲٪ از زمان SIFT تنها حدود ۲۷٪ کمتر از تطابق‌های صحیح آن را پیدا می‌کند، و در مقایسه با ORB با وجود صرف زمانی بیشتر، تطابق‌های درست بسیار بیشتری (به طور متوسط بیش از سه برابر) ارائه می‌دهد. البته همان‌گونه که اشاره شد، نقطه ضعف SCO بروز می‌کند آنجا که یکی از اجزای تشکیل‌دهنده‌اش محدودیت دارد: وابستگی به آشکارساز SIFT باعث می‌شود در تصاویر با بافت ضعیف، تعداد ویژگی‌های کمتری نسبت به ORB داشته باشد و در نتیجه تطابق کمتری بیابد. همچنین استفاده از توصیف‌گر ORB موجب می‌شود در تغییر مقیاس شدید، تطابق‌های استخراج‌شده کاهش یابد. با این حال در سایر شرایط متداول، روش SCO توانسته است به خوبی از ترکیب دو رویکرد سود برده و کارایی قابل قبولی ارائه کند؛ به طوری که می‌توان آن را راه‌حلی میانی دانست که دقت بالاتر از ORB و سرعت بیشتر از SIFT را به شکل همزمان فراهم می‌آورد.

برای درک بهتر تفاوت عملکرد این الگوریتم‌ها، شکل(۸) خروجی تطبیق آن‌ها بر روی یک جفت تصویر با بافت غنی (صحنه شلوغ) را به‌صورت کیفی نشان می‌دهد. همان‌طور که در این شکل و داده‌های جدول مربوطه مشاهده می‌شود، الگوریتم SIFT (و نیز روش‌های دقیقی مانند KAZE) در صحنه‌های پربافت قادر به یافتن هزاران تطابق صحیح است، در حالی که ORB به دلیل محدودسازی تعداد ویژگی‌ها و ساده‌تر بودن توصیف‌گر، تنها کسری کوچکی از این تطابق‌ها را پیدا می‌کند. روش ترکیبی SCO نیز هرچند به اندازه SIFT تطابق پیدا نکرده، اما تعداد نقاط متناظر شناسایی‌شده توسط آن به‌مراتب بیشتر از ORB بوده است و در حد قابل قبولی قرار دارد. در نقطه‌ی مقابل، شکل (۹)عملکرد این الگوریتم‌ها را بر روی تصاویر با بافت بسیار ضعیف نمایش می‌دهد. در یک صحنه کم‌بافت، الگوریتم‌های SIFT و ORB هنوز قادر به یافتن تعدادی تطابق صحیح (هرچند محدود) هستند، اما الگوریتم پیشنهادی SCO به‌مراتب تطابق‌های کمتری برقرار کرده است. این مشاهده همسو با نتایج کمی جدول (۵) است که نشان داد در تصاویر بافت‌ضعیف، به دلیل استفاده‌ی SCO از آشکارساز SIFT(و در نتیجه شناسایی تعداد ویژگی مشابه SIFT که کمتر از ORB است)، تطابق‌های موفق آن کمتر از دو روش دیگر خواهد بود. بنابراین دو شکل(۸) و (۹) مصداق بصری روشنی از نقاط قوت و ضعف نسبی الگوریتم‌ها ارائه می‌کنند SIFT :در محیط‌های پربافت بسیار قدرتمند ظاهر می‌شود، ORB در بافت‌های ضعیف نیز همچنان چند تطابق را می‌یابد، و SCO بسته به غنای ساختاری صحنه ممکن است عملکردی نزدیک به یکی از آن دو داشته باشد (در بافت قوی شبیه SIFT و در بافت ضعیف متمایل به ORB).

در نهایت، شکل (۱۰) به یک چالش مهم در حوزه روش‌های نوین مبتنی بر یادگیری عمیق اشاره می‌کند. این شکل خروجی دو الگوریتم یادگیری‌عمیق تطبیق را بر روی دو تصویر کاملاً غیرمرتبط نشان می‌دهد. همان‌گونه که ملاحظه می‌شود، هر دوی این الگوریتم‌ها در کمال تعجب تعداد قابل توجهی نقاط متناظر بین دو تصویر بی‌ارتباط پیدا کرده‌اند. این نتیجه به معنای آن است که روش‌های یادگیری عمیق در صورت عدم تنظیم مناسب، حتی بین تصاویری که هیچ وجه اشتراکی ندارند نیز تطابق برقرار می‌کنند که بیانگر نرخ بالای تطابق‌های کاذب در خروجی آن‌ها است. علاوه بر این، این دسته الگوریتم‌ها وابستگی شدیدی به داده‌های آموزشی خود دارند و دیده شده است که روی تصاویر با ساختار متفاوت یا بافت‌های تار و ضعیف، عملکرد مناسبی نداشته و درصد خطای آن‌ها افزایش می‌یابد. در مقابل، الگوریتم‌های تطبیق کلاسیک مانند SIFT و ORB به دلیل ماهیت مبنایی و عدم اتکا به آموزش داده، در چنین

شرایطی معمولاً بسیار محتاطانه‌تر عمل کرده و چنانچه دو تصویر هیچ شباهتی نداشته باشند، یا اصلاً تطابقی گزارش نمی‌کنند یا تعداد بسیار اندکی نقطه مشابه می‌یابند.

بدین‌ترتیب می‌توان گفت روش‌های کلاسیک از منظر قابلیت اعتماد ذاتی همچنان برتری‌هایی در سناریوهای خارج از توزیع داده‌ها دارند. البته باید توجه داشت که ظهور الگوریتم‌های یادگیری عمیق در سال‌های اخیر پیشرفت چشمگیری در افزایش دقت تطبیق ویژگی‌ها به ارمغان آورده است؛ با این وجود به دلایلی همچون شفافیت و قابلیت تفسیر، عدم نیاز به داده‌ی آموزشی و سازگاری با پلتفرم‌های سخت‌افزاری محدود، روش‌های کلاسیک (نظیر SIFTوORB) همچنان از ارکان اساسی بینایی ماشین بوده و معیار استانداردی برای ارزیابی روش‌های جدید باقی خواهند ماند.

۴ـ۴ـ بررسی چالش عملکرد الگوریتم‌های تطبیق مبتنی بر یادگیری عمیق روی تصاویر غیرمرتبط

شکل (۱۰) نتایج بصری اجرای دو نمونه از الگوریتم‌های تطبیق مبتنی بر یادگیری عمیق روی جفت تصاویر غیرمرتبط را نشان می‌دهد.

ملاحظه می‌شود که هر دو الگوریتم برای تصاویر غیرمرتبط تعداد زیادی نقاط متناظر تعیین کرده‌اند که به معنای تطبیق دو تصویر غیرمرتبط توسط الگوریتم‌ها است. علاوه بر این چالش اساسی دیگر این روش‌ها (در مواردی که از وزن‌های آماده استفاده شود) وابستگی شدید آن‌ها به مجموعه‌ی آموزشی است به طوری که روی تصاویر با ساختار و مشخصات جدیدی یا تصاویر با بافت تار یا ضعیف عملکرد نامطلوبی از خود نشان می‌دهند و نسبت False Positive ها در پاسخ آن‌ها زیاد می‌شود.

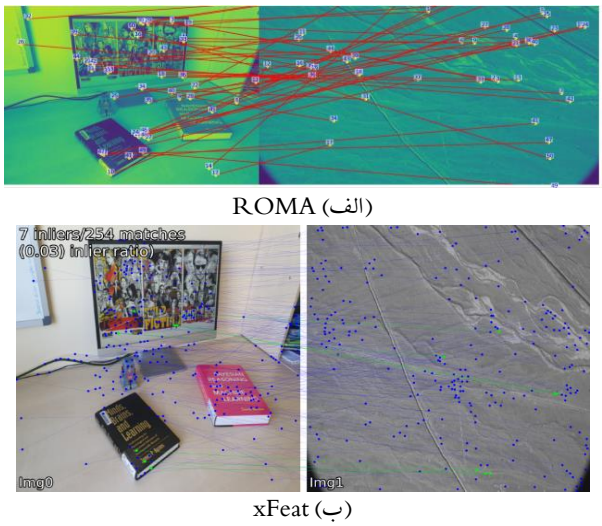
۵- نتیجه‌گیری

الگوریتم‌های استخراج ویژگی، ستون فقرات بسیاری از سیستم‌های بینایی ماشین هستند. هیچ الگوریتمی کلی و فراگیری وجود ندارد که از همه جوانب و در تمام شرایط بهترین عملکرد را داشته باشد. انتخاب الگوریتم مناسب به تعادل بین دقت، سرعت، مقاومت و منابع محاسباتی بستگی دارد.

علیرغم ظهور و پیشرفت چشمگیر الگوریتم‌های مبتنی بر یادگیری عمیق در سال‌های اخیر، الگوریتم‌های کلاسیک به دلیل شفافیت، قابلیت تفسیر و عدم نیاز به داده آموزشی و همچنین سازگاری با سخت‌افزارهای نهفته دارای منابع پردازشی و توان محدود همچنان جایگاه مهمی در پژوهش و صنعت دارند و به عنوان معیارهای استاندارد برای ارزیابی روش‌های جدید باقی خواهند ماند. جهت‌های تحقیقاتی نشان می‌دهد که آینده‌ی این حوزه به سمت ترکیب هوشمند الگوریتم‌های کلاسیک و یادگیری عمیق پیش خواهد رفت تا به نوعی ترکیبی از قابلیت تفسیرپذیری روش‌های کلاسیک با دقت بالای یادگیری عمیق ایجاد نماید.

۶- جهت‌گیری‌های آینده

با توجه به یافته‌ها و چالش‌های مطرح‌شده در این مقاله، چند مسیر مهم برای تحقیقات آینده به شرح زیر پیشنهاد می‌شود: اول، توسعه الگوریتم‌های ترکیبی که در آن‌ها آشکارسازهای کلاسیک مانند FASTو DoG با توصیف‌گرهای مبتنی بر شبکه‌های عصبی (به عنوان مثال HardNet برای بهبود ORB) ترکیب شوند. دوم، طراحی مدل‌های یادگیری خودنظارتی که بدون نیاز به داده‌های برچسب‌دار و تنها با استفاده از تبدیلات هندسی و فوتومتریک، ویژگی‌های پایا و مقاوم استخراج کنند. سوم، بهینه‌سازی الگوریتم‌های تطبیق ویژگی برای پیاده‌سازی کارآمد روی سخت‌افزارهای تعبیه‌شده با محدودیت‌های توان، حافظه و زمان. در نهایت، یکی از زمینه‌های کاربردی مهم، بهبود تطبیق صحنه در شرایط خارج از توزیع است؛ شرایطی مانند سطوح بازتابنده، اجسام شفاف و تغییرات شدید روشنایی که همچنان چالش‌برانگیزترین سناریوها برای الگوریتم‌های تشخیص و توصیف ویژگی به شمار می‌آیند.



ROMA (الف)

xFeat (ب)

شکل(۱۰): خروجی تطبیق الگوریتم‌ها روی جفت تصاویر غیر مرتبط

مراجع

[1] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, "NetVLAD: CNN architecture for weakly supervised place recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 5297–5307.

[2] H. Noh, A. Araujo, J. Sim, T. Weyand, and B. Han, "Large-scale image retrieval with attentive deep local features," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 3456–3465.

[3] S. Agarwal *et al.*, "Building rome in a day," *Communications of the ACM,* vol. 54, no. 10, pp. 105–112, 2011.

[4] S. Zhu *et al.*, "Very large-scale global sfm by distributed motion averaging," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4568–4577.

[5] L. Cui and C. Ma, "SDF-SLAM: Semantic depth filter SLAM for dynamic environments," *IEEE Access,* vol. 8, pp. 95301–95311, 2020.

[6] Y. Wang, Q. Huang, C. Jiang, J. Liu, M. Shang, and Z. Miao, "Video stabilization: A comprehensive survey," *Neurocomputing,* vol. 516, pp. 205–230, 2023.

[7] Y. Wang, Q. Huang, J. Liu, C. Jiang, and M. Shang, "Adaptive video stabilization based on feature point detection and full-reference stability assessment," *Multimedia Tools and Applications,* vol. 83, no. 11, pp. 32497–32524, 2024.

[8] Z. Zhang, Y. Zhang, Z. Shi, and S. Huang, "Image matching based on region template correlation," *Infrared Laser Eng,* vol. 33, no. 6, pp. 615–618, 2004.

9] Z. Diao, G. Peng, and L. Fu, "Fast high-precision reliable image matching algorithm based on shape," *J. Comput. Appl,* vol. 30, no. 2, p. 441, 2010.

[10] Q. Huang, X. Guo, Y. Wang, H. Sun, and L. Yang, "A survey of feature matching methods," *IET Image Processing,* vol. 18, no. 6, pp. 1385–1410, 2024.

[11] H. P. Moravec, "Rover Visual Obstacle Avoidance," in *IJCAI*, 1981, vol. 81, pp. 785–790.

[12] C. Schimd and R. Mohr, "Local gray value invariants for image retrieval," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* pp. 530–534, 1997.

[13] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the seventh IEEE international conference on computer vision*, 1999, vol. 2: Ieee, pp. 1150–1157.

[14] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *European conference on computer vision*, 2006: Springer, pp. 404–417.

[15] J.-M. Morel and G. Yu, "ASIFT: A new framework for fully affine invariant image comparison," *SIAM journal on imaging sciences,* vol. 2, no. 2, pp. 438–469, 2009.

[16] C. Lin, J. Liu, and L. Cao, "Image matching by affine speed-up robust features," in *MIPPR 2011: Pattern Recognition and Computer Vision*, 2011, vol. 8004: SPIE, pp. 120–124.

[17] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *2011 International conference on computer vision*, 2011: Ieee, pp. 2564–2571.

[18] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *2011 International conference on computer vision*, 2011: Ieee, pp. 2548–2555.

[19] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, "KAZE features," in *European conference on computer vision*, 2012: Springer, pp. 214–227.

[20] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision,* vol. 60, no. 2, pp. 91–110, 2004.

[21] D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superpoint: Self-supervised interest point detection and description," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 224–236.

[22] P. F. Alcantarilla and T. Solutions, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," *IEEE Trans. Patt. Anal. Mach. Intell,* vol. 34, no. 7, pp. 1281–1298, 2011.

[23] M. Pultar, "Improving the hardnet descriptor," *arXiv preprint arXiv:2007.09699,* 2020.

[24] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM,* vol. 24, no. 6, pp. 381–395, 1981.

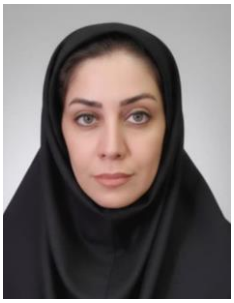
[25] D. Barath, J. Noskova, M. Ivashechkin, and J. Matas, "MAGSAC++, a fast, reliable and accurate robust estimator," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 1304–1312.

[26] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[27] P. Lindenberger, P.-E. Sarlin, and M. Pollefeys, "Lightglue: Local feature matching at light speed," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 17627–17638.

[28] U. o. O. Visual Geometry Group. "Affine Image Transformations." http://www.robots.ox.ac.uk/vgg/affine.html (accessed October 10, 2025).

سجاد اردشیری دانش‌آموخته کارشناسی ارشد مهندسی برق از دانشگاه آزاد اسلامی واحد خوراسگان در سال ۲۰۲۵، دارای تخصص طراحی سیستم‌های الکتروآپتیکی و مکاترونیکی و پیاده‌سازی کنترلرهای خطی مدل‌مبنا روی میکروکنترلرهای DSP شرکت Texas Instrument است. ایشان از سال ۲۰۱۳ به عنوان کارشناس الکترونیک مشغول به کار بوده و تجربه مناسبی در زمینه الگوریتم‌های پردازش و تطبیق تصاویر و توسعه‌ی سامانه‌های مکاترونیکی نظیر سرومکانیزم‌های الکترمکانیکی و مکانیزم‌های پایدارسازی مبتنی بر Rate-Gyro دارند.



عاطفه سلیمی شهرکی دانش‌آموخته دکترای مهندسی برق از دانشگاه صنعتی اصفهان در سال ۲۰۱۳، از همان سال تاکنون به‌عنوان عضو هیات علمی در گروه مهندسی برق، دانشکده مهندسی، دانشگاه آزاد اسلامی واحد اصفهان (خوراسگان) مشغول به فعالیت است. در سال ۲۰۱۷ به‌عنوان پژوهشگر پسادکتری در دانشگاه یورک، تورنتو، کانادا حضور داشته است. حوزه‌های پژوهشی او شامل طراحی دیجیتال ASIC/FPGA، معماری کامپیوتر، حسگرهای زیست‌پزشکی و شتاب‌دهنده‌های سخت‌افزاری برای توالی‌یابی DNA و پردازش سیگنال‌های زیستی می‌باشد.

مهدی عباسی مدرک کارشناسی و کارشناسی‌ارشد خود را در رشته‌ی مهندسی کامپیوتر به‌ترتیب در سال‌های ۲۰۰۰ و ۲۰۰۵ از دانشگاه صنعتی شریف دریافت کرده است. وی در سال ۲۰۱۲ نیز موفق به اخذ مدرک دکتری مهندسی کامپیوتر از دانشگاه اصفهان شد. از سال ۲۰۱۲ تا ۲۰۲۳، او عضو هیئت علمی دانشگاه بوعلی سینا بوده و به‌عنوان دانشیار در گروه مهندسی کامپیوتر فعالیت داشته است. از سال ۲۰۲۴ نیز به‌عنوان پژوهشگر مقیم با پژوهشکده علوم کامپیوتر، مؤسسهٔ پژوهش در علوم بنیادی (IPM) تهران و محقق دانشگاه اومئو سوئد همکاری دارد. زمینه های تحقیقاتی او معماری سیستم‌های کامپیوتری، طراحی بهینه سیستم‌های دیجیتال، هوش مصنوعی توزیع شده، هوش لبه و معماری شتاب‌دهنده‌های سخت افزاری است.