

بسم الله الرحمن الرحيم

شرکت مهندسی نرم افزاری هلو

# گزارش موقعیت قرار گیری دوربین ها

کاری از امیر علی نسیمی

## فهرست

۲	..... مقدمه
۲	..... روش‌های حل مسئله
۲	..... شیء مشترک
۲	..... مقدمه
۳	..... محدودیت‌ها
۷	..... مزایا
۷	..... محدوده خاص برای میز
۷	..... مقدمه
۸	..... محدودیت‌ها
۱۱	..... شبکه‌های عصبی
۱۱	..... مقدمه
۱۱	..... روش‌های پیشنهادی بررسی شده
۱۱	..... Image Correspondence or Feature Matching
۱۲	..... Stereo Vision
۱۳	..... Deep Learning-based Approaches
۱۳	..... Multi-View Object Recognition Using Generative Adversarial Networks (GANs)
۱۴	..... استفاده از سنسور برای شناسایی
۱۵	..... Few-Shot Recognition for Multi-View Object Recognition
۱۵	..... Active Learning for Multi-View Object Recognition
۱۷	..... مرجع

## مقدمه

مسئله Multiple-View Image از مسائل به روز در حوزه پردازش تصویر می باشد. تعریف مسئله این مورد بدین صورت می باشد که دو دوربین - از زوایای مختلف - از یک صحنه تصویر برداری می کنند. در این حالت شی خاصی در این صحنه وجود دارد که موقعیت مکانی آن از دید دوربین ها متفاوت است. هدف این است که مدل ریاضی تعبیه شود که بر اساس آن اشیاء مشترک در دو تصویر، متفاوت تلقی نشوند.

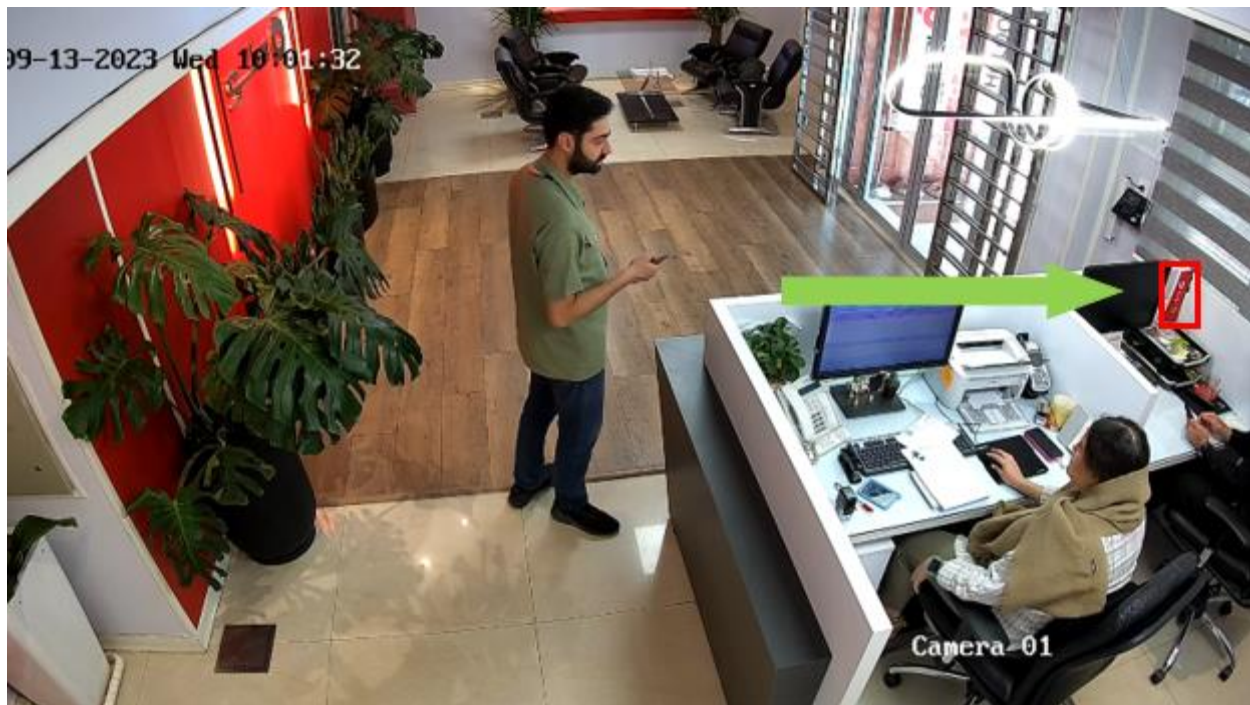
## روش های حل مسئله

با توجه به این مسئله، نوع نگاه های متفاوتی به این موضوع وجود دارد:

### شیء مشترک

#### مقدمه

در این نگاه، نقطه ای مشترک در صحنه انتخاب می شود. این نقطه در دو تصویر گرفته شده از دو دوربین نیز موجود است. یکی از تصاویر به عنوان تصویر مرجع انتخاب خواهد شد لذا بر این اساس مختصات نقطه مشترک در این تصویر نیز به عنوان نقطه مرجع انتخاب خواهد شد. میزان تغییرات این نقطه نسبت به تصویر بعدی نشان دهنده میزان تغییرات در زاویه ی دوربین خواهد بود. در این حالت نقطه مشترک باید بطور کامل در تصویر قابل رویت باشد. برای مثال در عکس زیر می توان مدلی را در نظر گرفت و آموزش داد که بر اساس آن از نماد هلو به عنوان نقطه مشترک استفاده شود.



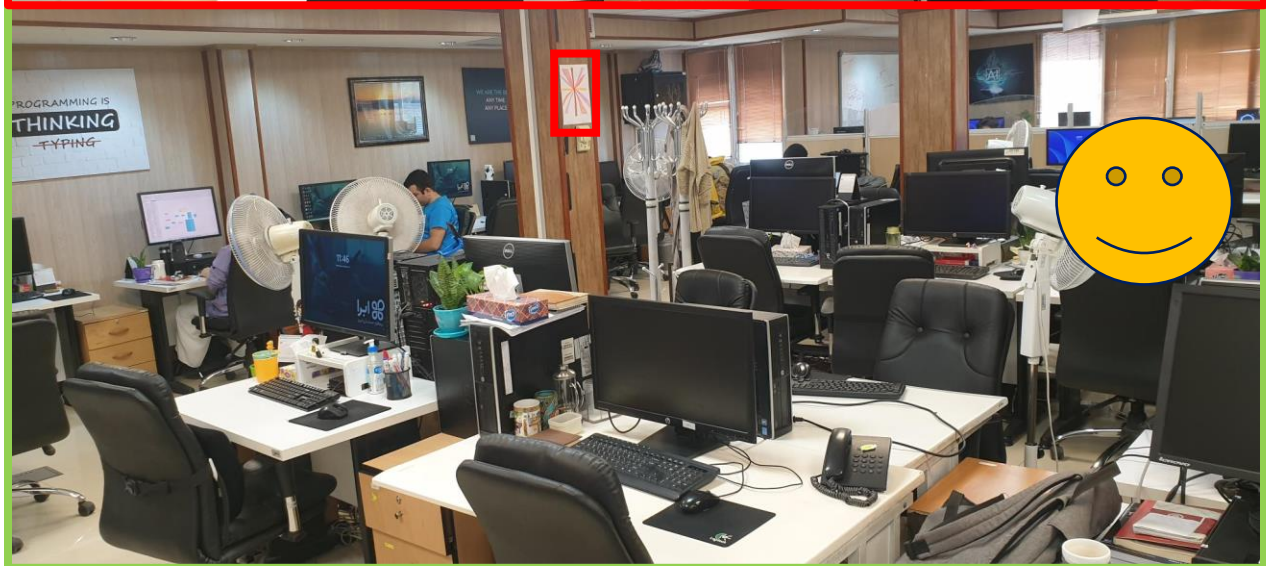
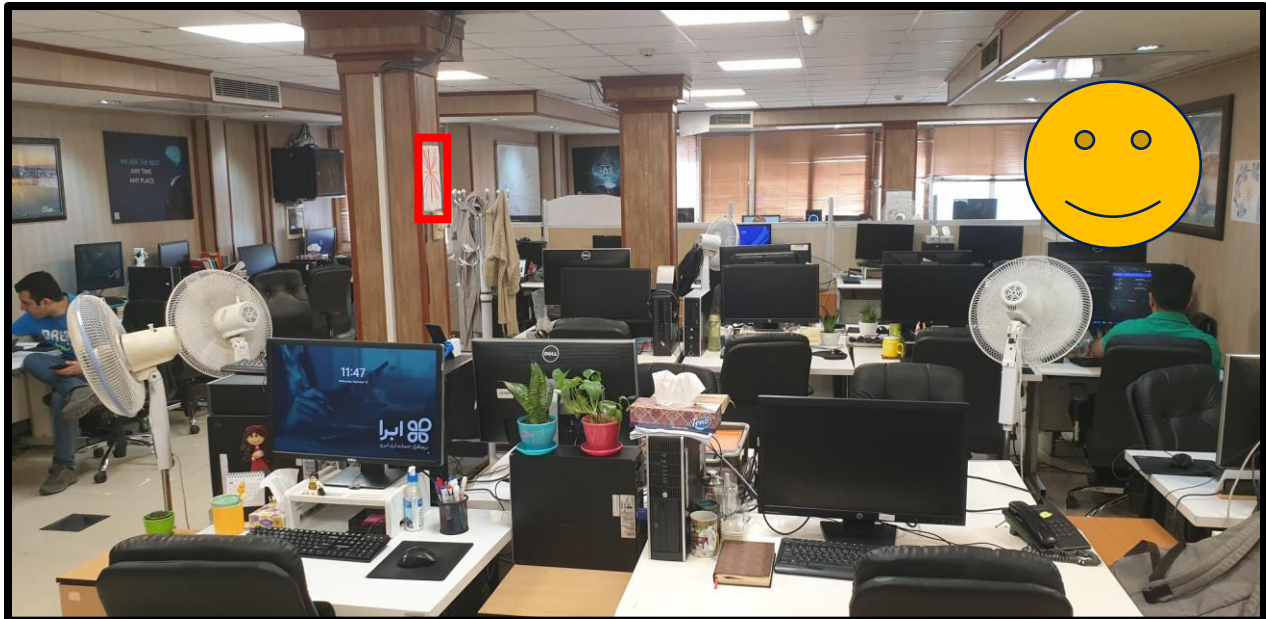
## محدودیت‌ها

در خصوص روش ذکر شده، محدودیت‌هایی وجود دارد که در ادامه به آن پرداخته خواهد شد:

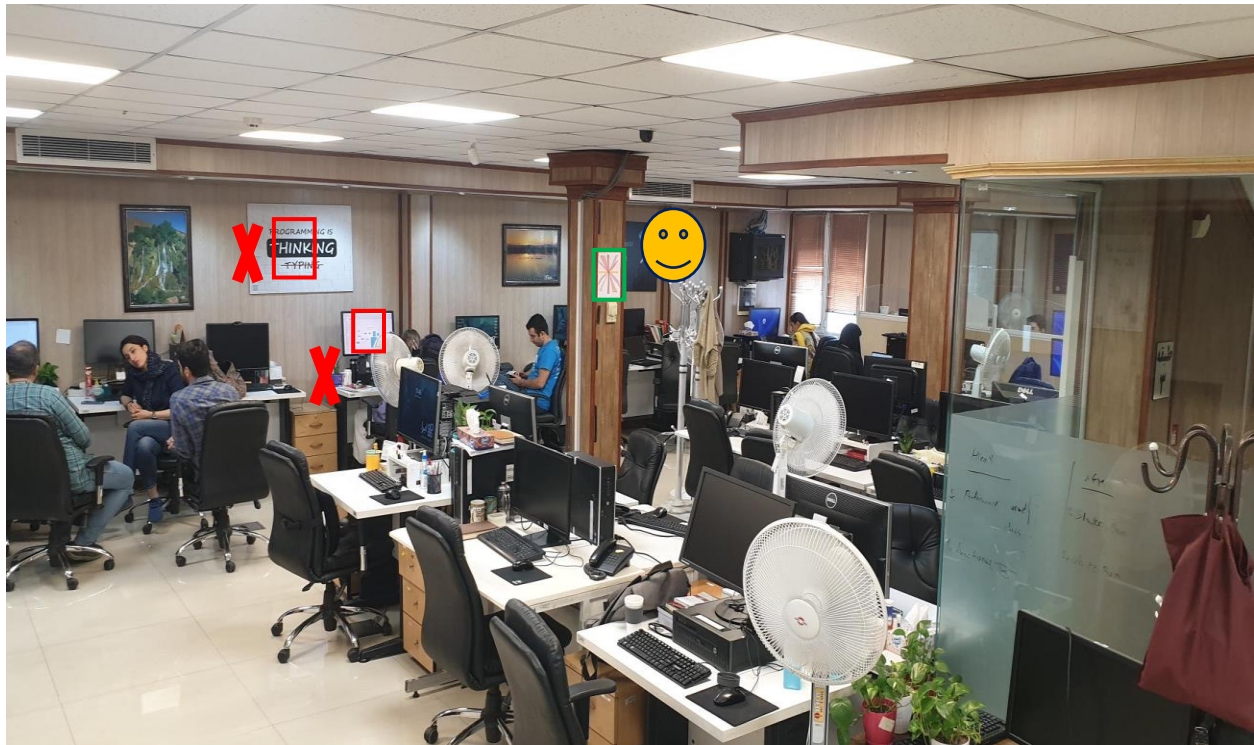
۱. در صورتیکه زاویه دوربین بسیار تغییر کند به نحوی که نقطه مشترک مورد نظر - بطور کامل - دیده نشود  
 آنگاه تخمین تغییرات مربوط به دوربین بسیار سخت و پیچیده خواهد شد. برای مثال در اشکال زیر این مسئله مورد بررسی قرار گرفته شده است:







۲. نماد انتخابی که به عنوان نقطه مشترک استفاده می شود ممکن است شبیه به تابلو بوده و سیستم اشتباهات تابلو را نیز به عنوان نماد در نظر گیرد. برای مثال در اشکال زیر این مسئله قابل رویت می باشد:



۳. موارد مربوط به شناسایی ممکن است در مسئله مشکلاتی ایجاد کند؛ برای مثال، مشکلات مربوط به نور، پیش پردازش و مواردی از این قبیل.

۴. خطای شهودی در این حالت وجود دارد. برای مثال موقعیت نقطه در فضا بصورت سه بعدی باید اندازه گیری شود تا جهت خطای دوربین اندازه گیری شود

۵. محدودیت مربوط به تعریف مرز برای میزهایی که نسبت به موقعیت مکانی دوربین، بسیار دور هستند

## مزایا

- مزیت استفاده از این روش سرعت آن می‌باشد. با آموزش مدلی برای پردازش تصویر، تنها مسئله پردازش استفاده از این مدل می‌باشد. پس از آن استفاده از روش‌های ریاضی جهت تخمین مختصات جدید، مورد نظر خواهد بود

## محدوده خاص برای میز

### مقدمه

در این نگاه نیاز به روشی جهت شناسایی میز وجود دارد. پس از اعمال شناسایی، برای هر میز یک محدوده در نظر خواهد گرفته شد. بر این اساس، زمانی که تغییرات بر دوربین اعمال شود - دوربین تکان شدید خورد یا .. - آنگاه بر اساس موقعیت قبلی میز و با در نظر گیری محدوده‌ای خاص برای هر میز هر چه قدر هم که میز تغییر داشته باشد بر اساس آن حد و محدوده، میز مربوطه همچنان قابل تشخیص خواهد بود. در تصویر زیر موقعیت میزها بررسی شده است.

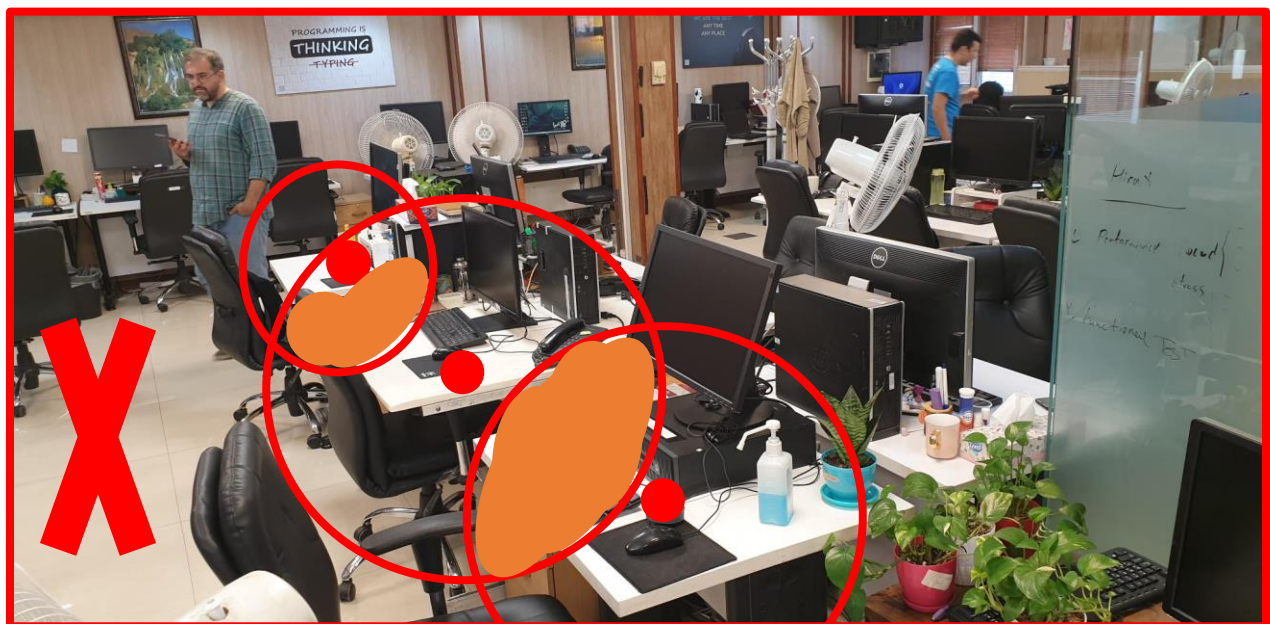
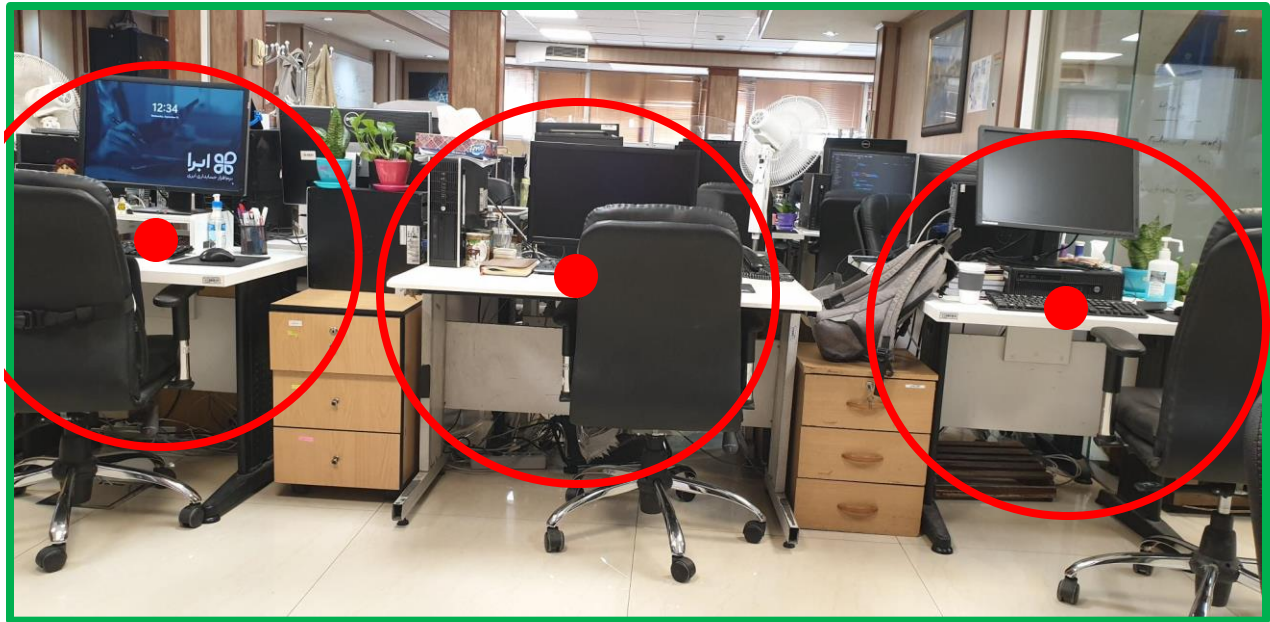




## محدودیت‌ها

در خصوص روش ذکر شده، محدودیت‌هایی وجود دارد که در ادامه به آن پرداخته خواهد شد:

۱. مسئله میزان نزدیک بودن میزها: در این حالت در صورتیکه میزها در شرایطی بسیار به هم نزدیک باشند آنگاه سبب ایجاد خطا خواهد شد. بطور دقیق‌تر، تعریف یک محدوده برای هر میز بطور دقیق بسیار سخت و در شرایطی ناممکن خواهد شد. در این حالت همپوشانی بر کلیه مرزهای بین میزها اتفاق خواهد افتاد







۲. محدودیت مربوط به تعریف مرز برای میزهایی که نسبت به موقعیت مکانی دوربین، بسیار دور هستند



## شبکه‌های عصبی

### مقدمه

در این گونه روش‌ها با توجه به ویژگی‌های مربوط به تصاویر و ارتباط اشیا با یکدیگر، تخمین زده می‌شود که کدام شی موجود در دو تصویر – که از زوایای دیگر از دوربین گرفته شده است – در واقع یک شی است.

### روش‌های پیشنهادی بررسی شده

در این بخش به بررسی روش‌های مطالعه شده، پرداخته خواهد شد:

#### *Image Correspondence or Feature Matching*

موضوع [1]: پیدا کردن نقاط یا مناطق متناظر بین دو تصویر از دو نقطه دید مختلف. گام‌های زیر برای ادامه کار مورد نظر است:

۱. Feature Extraction: ویژگی‌های خاص یا نقاط کلیدی متمایز در هر دو تصویر شناسایی شود. این ویژگی‌ها نقاط محلی و متمایزی هستند که برای شناسایی نقاط متناظر در تصاویر مورد استفاده قرار می‌گیرند. الگوریتم‌های معروف تشخیص ویژگی ممکن است شامل SIFT، ORB و یا Harris corners باشند.

۲. Feature Description: بعد از شناسایی نقاط کلیدی، توصیف‌ها<sup>۱</sup> برای هر یک از آنها محاسبه شود. این توصیف‌ها نماینده‌های کوچکی از ناحیه محلی در تصویر در اطراف هر کدام از نقاط کلیدی هستند. این توصیف‌ها به منظور مقاومت در برابر تغییرات در دیدگاه و شرایط نوری طراحی شده و به نوعی پیش پردازش هستند.

---

<sup>1</sup> descriptors



۳. Feature Matching: تطبیق توصیف‌ها بین دو تصویر به منظور یافتن ویژگی‌های متناظر انجام شود. می‌توان از تکنیک‌هایی مانند نزدیک‌ترین همسایه<sup>۲</sup> استفاده کرد که توصیف‌های نقاط کلیدی در یک تصویر را با توصیف‌های نقاط کلیدی در تصویر دیگر مقایسه کرده و بهترین تطبیق‌ها را انتخاب می‌کند. روش‌های پیشرفته‌تر مانند RANSAC (Random Sample Consensus) می‌توانند برای حذف تطبیق‌های نادرست و تخمین تبدیل‌های هندسی استفاده شوند.

مثال: تصور کنید که دو تصویر از یک شی از دو زاویه مختلف دارید. تطبیق ویژگی‌ها می‌تواند نقاط کلیدی در هر دو تصویر را شناسایی کرده و تطبیق‌های متناظر را برقرار کند. این تطبیق‌ها برای برنامه‌های مختلفی مانند تخمین موقعیت سه‌بعدی شی یا ایجاد نقشه اختلاف عمق برای تخمین عمق مفید هستند.

## Stereo Vision

موضوع [2]: تخمین عمق و بازسازی سه بعدی از جفت تصاویر. گام‌های زیر مورد نظر است:

۱. Stereo Images: از جفت تصاویری که از نقاط دید متفاوتی گرفته شده استفاده شود.

۲. Disparity Estimation: هدف اصلی در دید سه بعدی، یافتن اختلافات عمق<sup>۳</sup> است که نماینده شیب‌های افقی پیکسل‌ها<sup>۴</sup> برای مطابقت نقاط یا ویژگی‌ها بین دو تصویر می‌باشد. روش‌هایی مانند مطابقت بلوک<sup>۵</sup>، برش‌های گراف<sup>۶</sup> یا روش‌های مبتنی بر یادگیری عمیق معمولاً برای تخمین اختلاف عمق استفاده می‌شوند.

مثال: این مدل می‌تواند در صنایع رباتیک و ساخت نقشه‌های سه بعدی از محیط مورد استفاده قرار گیرد. با محاسبه اختلافات عمق، می‌توانید فاصله اشیاء از دوربین استریو را تعیین کرده و ساختار سه بعدی محیط را بازسازی کنید.

<sup>2</sup> KNN

<sup>3</sup> disparities

<sup>4</sup> horizontal pixel shifts

<sup>5</sup> block matching

<sup>6</sup> graph cuts

## Deep Learning-based Approaches

موضوع [3]: استفاده از شبکه‌های عصبی عمیق برای تطبیق تصاویر:

۱. شبکه‌های سیامی<sup>۷</sup>: از شبکه‌های سیامی به عنوان نوعی از معماری شبکه‌های عصبی عمیق برای یادگیری معیارهای تشابه استفاده می‌شود. این شبکه‌ها بر روی جفت‌های تصاویر چند نما<sup>۸</sup> آموزش داده می‌شوند که شبکه یاد می‌گیرد تا یک معیار تشابه بین دو تصویر ورودی ارائه دهد که به او امکان می‌دهد تا تشخیص دهد که تصاویر چقدر مشابه یا متفاوت هستند. این رویکرد برای شناسایی مناطق متناظر در تصاویر چند نما استفاده می‌شود.

۲. شبکه‌های سه‌تایی<sup>۹</sup>: شبکه‌های سه‌تایی یک رویکرد یادگیری دیگر هستند که از سه تصویر تشکیل شده‌اند: تصویر مرجع، تصویر مثبت (مشابه تصویر مرجع) و تصویر منفی (متفاوت از تصویر مرجع). این شبکه‌ها یاد می‌گیرند تا تصویر مرجع و تصویر مثبت را در فضای ویژگی نزدیک به یکدیگر قرار دهند و تصویر منفی را دور کنند. به این ترتیب، آنها زیرفضایی را یاد می‌گیرند که برای شناسایی اشیاء یا مناطق متناظر مورد استفاده قرار می‌گیرد.

## Multi-View Object Recognition Using Generative Adversarial Networks (GANs)

موضوع [4]: بهره‌گیری از شبکه‌های مقابله‌ای تولیدی (GANs) برای تشخیص اشیاء در تصاویر چند نما:

۱. GAN-Based Data Augmentation: از GAN برای تولید داده‌های تکمیلی یا افزایش داده استفاده می‌کنیم. با آموزش GANs بر روی تصاویر موجود، می‌توانیم نماهای مصنوعی از اشیاء با زوایای و شرایط نوری مختلف ایجاد کنیم تا مجموعه داده آموزشی خود را گسترش دهیم.

<sup>۷</sup> siami

<sup>۸</sup> multi-view

<sup>۹</sup> triplet

۲. Discriminative Networks: شبکه‌های مربوطه را آموزش می‌دهیم تا اشیاء را بر اساس نماهای واقعی و مصنوعی تشخیص دهند. شبکه تمیزدهنده به نوعی یک امتیاز به تصاویر اختصاص می‌دهد و به مدل تشخیص دهنده امکان می‌دهد که بهترین تطبیق‌ها را برای تصاویر با زوایای مختلف انجام دهد.

مثال: با استفاده از GANs، می‌توانیم تصاویر مصنوعی از اشیاء با زوایای مختلف ایجاد کرده و سپس از شبکه‌های تمیزدهنده برای تشخیص اشیاء در تصاویر واقعی با شرایط نوری مختلف استفاده کنیم. این رویکرد می‌تواند به افزایش دقت تشخیص اشیاء کمک کند.

### استفاده از سنسور برای شناسایی

موضوع [5]: ترکیب داده‌ها از چندین حسگر برای شناسایی قوی‌تر اشیاء.

Sensor Array: یک آرایه از سنسورها مانند دوربین‌ها، LiDAR و حسگرهای عمق را مستقر می‌کنیم تا داده‌های چندین تصویر از زوایای مختلف به دست آید.

Sensor Data Fusion: الگوریتم‌های جهت تلفیق اطلاعات این دستگاه‌ها را تولید می‌کنیم. تکنیک‌هایی مانند ادغام حسگری<sup>۱۰</sup>، ادغام بیزین<sup>۱۱</sup> یا ادغام مبتنی بر یادگیری عمیق می‌توانند برای این موضوع استفاده شوند.

Recognition: الگوریتم‌های شناسایی اشیاء را روی داده ادغامی اعمال می‌کنیم و از اطلاعات غنی‌تر و متنوع‌تری که از سنسورهای متعدد به دست آمده استفاده می‌کنیم.

<sup>10</sup> sensor fusion

<sup>11</sup> Bayesian fusion

## *Few-Shot Recognition for Multi-View Object Recognition*

موضوع [6]: امکان تشخیص اشیاء با تعداد محدودی نمونه از دیده‌شدگی‌های مختلف.

۱. Few-Shot Dataset: یک مجموعه داده تشخیص اشیاء با تعداد محدودی نمونه از هر شیء از زوایای مختلف ایجاد می‌کنیم.

۲. Meta-Learning: از تکنیک‌های یادگیری متا مانند MAML (یادگیری متا بدون وابستگی به مدل) یا Reptile استفاده می‌کنیم تا مدل‌ها را آموزش دهیم که به سرعت به اشیاء و دیدگاه‌های جدید با تعداد محدودی نمونه پیش از آموزش بتوانند سازگار شوند.

۳. Recognition

## *Active Learning for Multi-View Object Recognition*

موضوع [7]: استفاده از استراتژی‌های یادگیری فعال برای انتخاب مشاهدات<sup>۱۲</sup> مفهومی برای تشخیص اشیاء.

Initial Model: با یک مدل تشخیص اولیه شروع خواهیم کرد بطوریکه این مدل قبلاً بر روی یک مجموعه داده محدود آموزش داده شده است.

View Selection: تکنیک‌های یادگیری فعال را پیاده‌سازی می‌کنیم تا مشاهدات بیشتری که اطلاعات مفهومی را برای برچسب‌گذاری<sup>۱۳</sup> دارند انتخاب کنید. این مشاهدات ممکن است بر اساس عدم اطمینان<sup>۱۴</sup>، تنوع<sup>۱۵</sup> یا معیار<sup>۱۶</sup>های دیگر انتخاب شوند.

<sup>12</sup> views

<sup>13</sup> labeling

<sup>14</sup> uncertainty

<sup>15</sup> diversity

<sup>16</sup> criteria



**Model Updating:** به طور مداوم مدل تشخیص را با استفاده از داده‌هایی که به تازگی برچسب گذاری شده است به‌روزرسانی می‌کنیم. مدل به مرور زمان به طور متناسب تر در تشخیص اشیاء از زوایا و مشاهدات مختلف تقویت خواهد شد.

- [1] Lowe, D. G. (2004). *Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision*, 60(2), 91-110.
- [2] Scharstein, D., & Szeliski, R. (2002). *A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision*, 47(1-3), 7-42.
- [3] Chopra, S., Hadsell, R., & LeCun, Y. (2005). *Learning a similarity metric discriminatively, with application to face verification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Vol. 1, pp. 539-546).*
- [4] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). *Generative adversarial nets. In Advances in neural information processing systems (NeurIPS).*
- [5] Kaaniche, W., Eude, F., Ouezdou, F. B., & Atri, M. (2020). *Multi-sensor data fusion: A review of the state-of-the-art. Information Fusion*, 57, 109-123.
- [6] Finn, C., Abbeel, P., & Levine, S. (2017). *Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In Proceedings of the International Conference on Machine Learning (ICML).*
- [7] Settles, B. (2012). *Active Learning. Synthesis Lectures on Artificial Intelligence and Machine Learning*, 6(1), 1-114.