به نام خدا

دانشگاه صنعتی شریف

دانشکده مهندسی کامپیوتر

عنوان

گزارش پروژه نهایی درس

اعضای گروه

احمدرضا خناری

امیرحسین حاجی محمد رضایی

رامتین مسلمی

نام درس

درس مبانی بینایی سه بعدی کامپیوتری

نیم‌سال اول ۱۴۰۳-۱۴۰۲

نام استاد درس

دکتر شهره کسایی

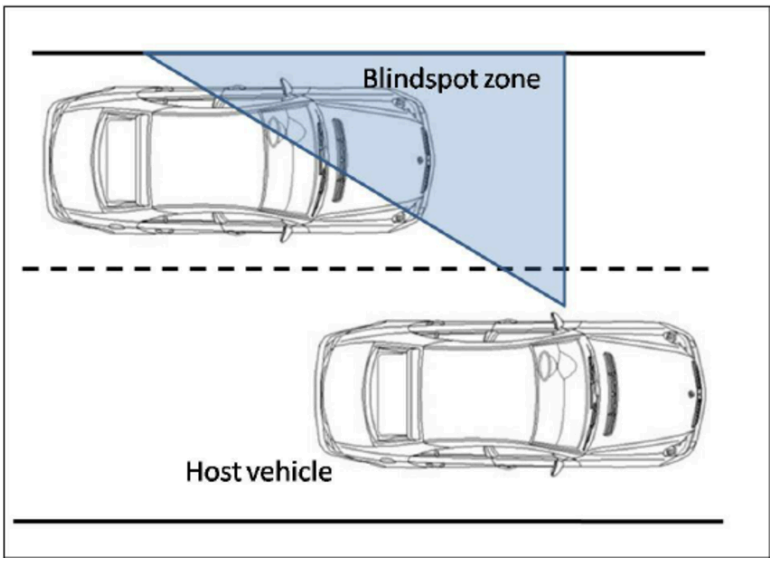
چکیده

ما در این پروژه قصد داریم یک نقشه سه‌بعدی از اطراف یک ماشین با تمرکز بر عمق آن ایجاد کنیم که برای کاربردهای مانند ماشین‌های خودران استفاده می‌شوند. در این کار، ابتدا از عکس‌های گرفته شده از سنسورهای ماشین در جهات مختلف، استفاده می‌کنیم و با ترکیب کردن آنها با یکدیگر و ایجاد یک تصویر پانوراما، در مرحله بعدی قصد ایجاد یک نقشه شامل عمق از این تصویر را داریم تا بدین ترتیب بتوان برای ادراک نزدیکی اشیا یا ماشین‌های دیگر در کاربرد ماشین‌های خودران از آن استفاده کرد.

مقدمه

ماشین‌های خودران یکی از مهم‌ترین تکنولوژی‌ها و مسائل مهم امروز هستند. همچنین چالش‌هایی که برای درک محیط اطراف با آن ها روبه‌رو هستیم، در حوزه بینایی کامپیوتر می‌گنجند. از طرفی، با توجه به حساسیت بسیار زیادی که در رابطه با این کاربرد وجود دارد، الگوریتم‌های مورد استفاده باید بسیار دقیق عمل کنند تا تعداد تصادفات و خطرات جانی که می‌توانند بوجود بیایند کم شده و ایمنی افزایش پیدا کند.

علاوه بر ماشین‌های خودران، رانندگان نیز می‌توانند از فناوری‌های بینایی کامپیوتر استفاده کنند. امروزه استفاده از سنسورها و دوربین‌ها در ماشین‌ها بسیار فراگیر شده است. یکی از نکات مهمی که به طور کلی در رانندگی وجود دارد، وجود داشتن نقطه کور در تصاویرِ در آینه (این مورد در ماشین‌های بزرگ حادتر است) یا دوربین‌های بغل ماشین است. در این حالت اگر یک موتور در این نقطه قرار بگیرد، در این حین، راننده یا کامپیوتر نمی‌تواند آن را تشخیص دهد و این خود می‌تواند باعث صدمات جانی و مالی شود.



تصویر ۱. Blindspot zone یا نقطه کور [1]

در همین راستا، ما در این کار قصد داریم که با استفاده از اطلاعات تصاویر دریافت شده توسط دوربین‌های عقب، راست و چپ ماشین و با استفاده از روش‌های ترکیب تصاویر (image stitching)، یک تصویر پانوراما از اطراف ماشین بدست بیاوریم تا در این صورت بتوانیم مشکل ایجاد نقطه کور را حل کنیم. در ادامه، برای اینکه بتوان به سیستم خودران کمک کرد تا درک بهتری از دوری یا نزدیکی ماشین‌ها و اجسام اطراف خود داشته باشد، از تصویر ایجاد شده در بخش قبل استفاده می‌کنیم تا یک نقشه عمق بدست بیاید که در نتیجه، این تصویر برای جلوگیری از تصادف کردن در ماشین‌های خودران بسیار کمک‌کننده است.

به‌طور خلاصه، کاری که ما در اینجا قصد پیاده کردن و بررسی آن را داریم:

۱- ایجاد یک تصویر پانوراما با استفاده از تصاویر بدست‌آمده از دوربین‌های ماشین که در جهت‌های عقب، راست و چپ قرار گرفته‌اند.

۲- ایجاد یک نقشه عمق با استفاده از تصویر بدست‌آمده در بخش قبل.

با بکارگیری روش‌های گفته شده می‌توان سامانه‌ای ایجاد کرد که نیاز به آینه‌ها در ماشین‌ها را از بین می‌برد و آن‌ها را با تعدادی دوربین و سنسور جایگزین می‌کند. بسیاری از صاحب‌نظران بر این باورند که ماشین‌های خودران تا اواسط دهه‌ی ۲۰۳۰ میلادی فراگیر نخواهند شد. ایجاد چنین سامانه‌ای و پیاده‌سازی آن در یک مقیاس صنعتی، فرآیند جمع‌آوری حجم گسترده‌ای از داده‌های مربوط به رانندگی را تسهیل می‌بخشد و از طرفی نسبت به ایجاد یک سامانه‌ی خودران ساده‌تر ،پیش نیاز آن به شمار می‌آید.

مرور کارهای پیشین

در این بخش به کارهای مرتبط با ترکیب تصاویر و ایجاد نقشه عمق که در ماشین‌های خودران استفاده شده اند، اشاره خواهیم کرد.

ترکیب تصاویر (image stitching)

ایده اصلی و کلی برای ترکیب تصاویر و ایجاد یک تصویر پانوراما از چندین عکس این است که در ابتدا با تشخیص نقاط کلیدی و مهم در هر دو عکس و انطباق آنها با یکدیگر بتوان ماتریس homography بین دو عکس را بدست آورد و در نهایت با استفاده از این ماتریس و انجام یک projective transformation بر یکی از تصاویر بتوان آنها را با هم منطبق کرد [2]. در این راستا، برای ایجاد یک تصویر که دارای یک میدان دید طبیعی و مانند وقتی که انسان به آن منظره نگاه می‌کند است، می‌توان از باز نگاشت تصویر پانوراما‌ی ایجاد شده بر روی یک سطح استوانه‌ای و یا کروی استفاده کرد و همچنین با روش‌های بهینه‌سازی، کیفیت آن را بهتر کرد [3].

در راستای کاربرد ماشین‌های خودران نیز، ترکیب تصاویر مشابه انجام شده است. بیشتر این کارها بر روی عکس‌هایی انجام شده است که توسط دوبین‌های fisheye تهیه شده و در واقع در این عکس‌ها اثرات عدسی دوربین نیز قابل مشاهده است. از جمله این کارها می توان به [4] اشاره کرد که از روش ترکیب تصاویر و کمینه کردن مجموع مربعات برای یافتن بهتر ماتریس homography بهره می‌برد.

تخمین عمق

روش‌هایی که برای تخمین عمق و فاصله با توجه به تصاویر وجود دارد، بیشتر مبتنی بر روش‌های یادگیری عمیق و بر عکس‌های تک چشمی عمل می‌کنند. به‌عنوان اولین مورد، میتوان به مدل monodepth2 اشاره کرد [5]. در این روش با استفاده از روش self-supervised learning و تغییرات در تابع loss و ماسک‌های برای عمق، نتایج دقیق‌تری نسبت به مدل mondepth برای عکس‌های در محیط ماشین‌های خودران ایجاد می‌شود [6]. همچنین مشابه این کار میتوان از عکس‌های دوربین‌های fisheye با داده اضافی از نقشه LIDAR محیط استفاده کرد تا تخمین عمق را برای این عکس‌ها بتوان انجام داد [8, 7]. همچنین، به‌عنوان مورد آخر می‌توان به مدل panodepth اشاره کرد [9]. این مدل یک تصویر پانوراما ۳۶۰ درجه از محیط خود دریافت می‌کند و با استفاده از مدل‌های یادگیری عمیق و بازتولید عکس از منظره‌های مختلف و مقایسه این عکس‌ها با یکدیگر، عمق اشیا را تخمین می‌زند.

روش‌های استفاده شده

ترکیب تصاویر (image stitching)

برای ترکیب تصاویر می‌توانیم ماتریس هموگرافی را بدست آورده و فرآیند stitching را به کمک این ماتریس انجام دهیم. ولی ما در این پروژه با چندین چالش مواجه می‌شویم:

۱. ما ابعاد و زوایای دوربین‌ها را نداریم و از dataset های آماده استفاده می‌کنیم. به همین علت تخمین ماتریس هموگرافی را باید خودمان انجام دهیم. از آنجایی که در این روش ما باید در شرایط غیرایده‌آل تخمین زدن را انجام دهیم امکان دارد کمی خطا داشته باشیم. در این صورت به کمک SIFT نقاط متناظر را پیدا کرده و ماتریس هموگرافی را به این روش تخمین می‌زنیم.

۲. به طور کلی در دوربین‌های معمولی FoV یا همان Field of View حدود ۳۰ الی ۴۰ درجه می‌باشد. پس برای اینکه بتوانیم فرآیند stitching را انجام دهیم نیاز داریم که یک نمای ۳۶۰ درجه از اطراف ماشین داشته باشیم که میان دوربین‌های استفاده شده مقداری overlap نیز وجود داشته باشد. بسیاری از dataset های موجود برای این کار مناسب نمی‌باشند.

برای حل چالش دوم چاره‌ای جز گزینش بهترین dataset های موجود را نداریم و از همین روی مقداری اختیاراتمان محدود می‌شود. در حالت ایده‌آل می‌بایست خودمان طراحی را انجام داده و دوربین‌های متعدد را به گونه‌ای نصب کنیم تا نقاط متناظر مشترک در آن‌ها وجود داشته باشد. سپس با رانندگی در سطح شهر یک dataset جدید مطابق با نیازهای خود بسازیم. بدیهیست که این کار زمان و هزینه‌ی نسبتا قابل توجهی می‌برد و ما با استفاده بهینه از dataset های موجود تلاش می‌کنیم تا بهترین نتیجه را بگیریم.

برای حل چالش اول می‌توانیم به صورت iterative فرآیند تخمین ماتریس هموگرافی را انجام دهیم. با توجه به اینکه تصاویر موجود مربوط به رانندگی در سطح شهر هستند، و البته به چالش دوم که پیش‌تر به آن پرداختیم، امکان دارد نقاط متناظر خوبی پیدا نکنیم. به همین علت می‌توانیم از فریم‌های متوالی استفاده کنیم تا ماتریس هموگرافی را بهتر و دقیق‌تر تخمین بزنیم. برای انجام این کار ابتدا برای فریم اول ماتریس هموگرافی را بدست می‌آوریم و با دسترسی به فریم‌های بعدی این ماتریس را به صورت جمع وزندار مقدار فعلی و ماتریس جدید بروزرسانی می‌کنیم:

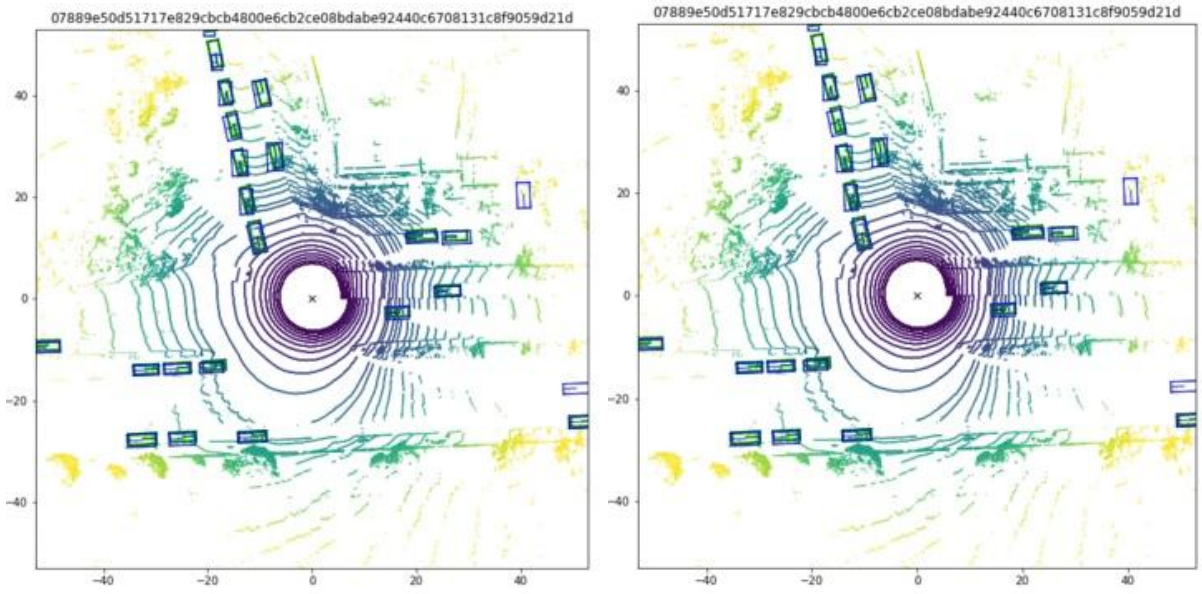
$$H = \alpha H' + (1 - \alpha)H, \qquad \alpha = \frac{1}{cnt}$$

که در این رابطه مقدار آلفا به مرور زمان کاهش می‌یابد تا به مقدار ثابتی همگرا شویم (cnt شمارنده‌ی فریم‌هاست). اگر بخواهیم در صورت drift نیز دوربین‌های ما به درستی کار کنند، می‌توانیم یک پنجره یا window تعریف کنیم و بروزرسانی را محدود به این بازه انجام دهیم.

در نظر داشته باشید که امکان دارد در یک فریم نتوانیم تعداد مناسبی نقطه‌ی مشترک استفاده کنیم، در این صورت ماتریس بدست آمده بسیار خطا خواهد داشت و نتیجه‌ی مطلوبی از آن حاصل نخواهد شد. در این صورت ما اختلاف ماتریس‌هایمان (ماتریس مربوط به این فریم و ماتریس جمع موزون) را حساب کرده و نرم آن را محاسبه می‌کنیم. در صورتی که این مقدار بزرگ باشد (با توجه به نتایج عملی ما این مقدار اگر بیشتر از ۱۰۰ باشد می‌توان انتظار داشت که این خطا رخ داده است) می‌توان حدس زد که در این فریم ماتریس هموگرافی بدست آمده مناسب نبوده، پس آن را کنار گذاشته و با استفاده از ماتریس‌های پیشین (جمع موزون ماتریس‌های هموگرافی فریم‌های پیش) فرآیند stitching این فریم را انجام می‌دهیم و در جمع موزون هم این ماتریس را در نظر نمی‌گیریم.

تخمین عمق

روش‌هایی که برای تخمین عمق می‌توانیم استفاده کنیم به طور کلی به دو دسته‌ی stereo vision و deep learning تقسیم‌بندی می‌شوند. با توجه به اینکه بسیاری از dataset ها فاقد stereo هستند ما از روش‌های ژرف استفاده می‌کنیم. با استفاده از مدل موجود که توسط **nuScenes** در اختیارمان قرار دارد می‌توانیم عمق اشیا را در جهات مختلف بیابیم. در نظر داشته باشید برای پیاده‌سازی این بخش ما باید به زوایای دوربین‌ها نیز دسترسی می‌داشتیم که این مقادیر نیز توسط nuScenes در اختیارمان قرار گرفته بود. از طرفی برای تشخیص اشیا به کمک عمق آن‌ها نیز از مدل‌های ژرف همین dataset استفاده می‌کنیم. به این صورت می‌توان تصاویر زیر را بدست آورد:

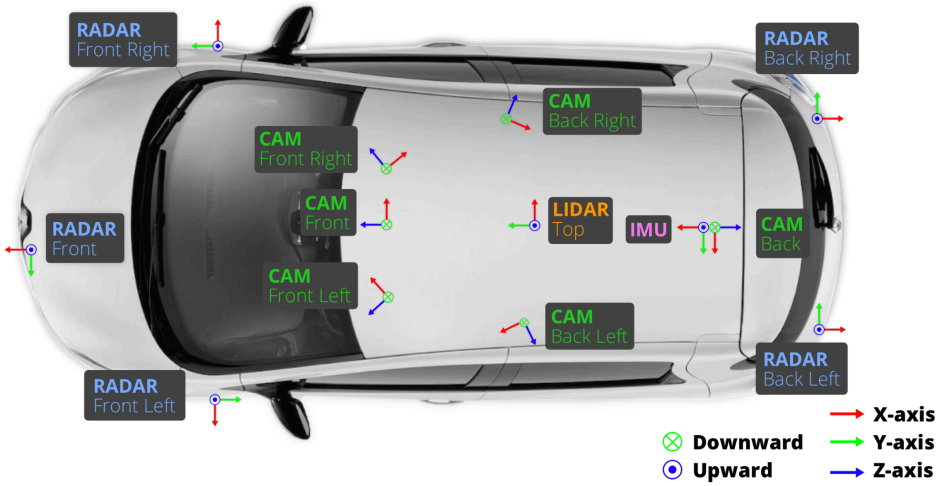


در این تصاویر x نشان دهنده‌ی موقعیت ماشین است. مستطیل‌های آبی رنگ ماشین‌های تشخیص داده شده می‌باشند.

مجموعه داده‌ها

مجموعه داده nuScenes

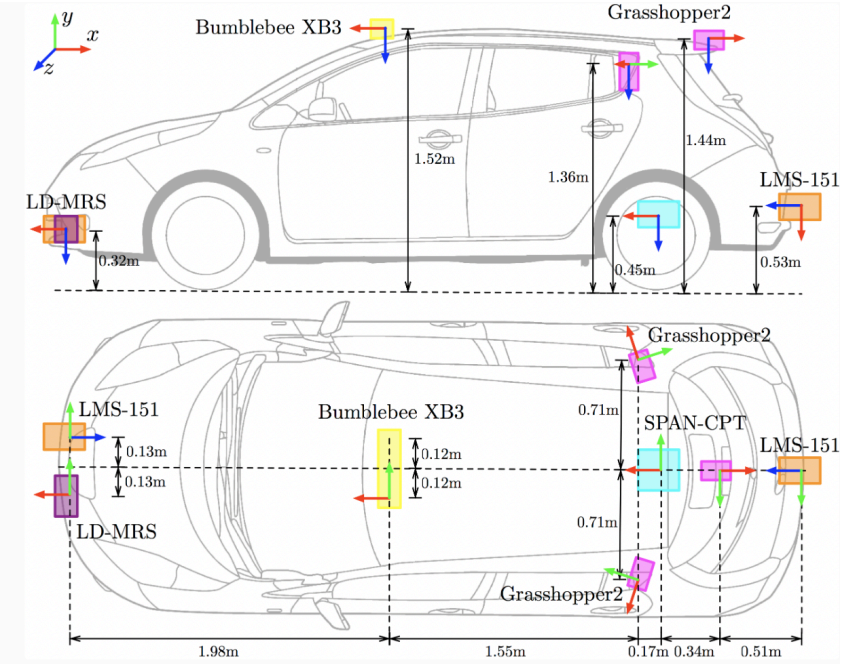
مجموعه داده nuScenes، یک مجموعه داده عمومی در مقیاس بزرگ برای رانندگی خودکار است که توسط تیم Motional توسعه یافته است و شامل تقریباً 1.4 میلیون تصویر دوربین، 390 هزار پیمایش LIDAR، 1.4 میلیون حرکت رادار و 1.4 میلیون جعبه محدودکننده شی در 40 هزار فریم کلیدی است [10]. برای این منظور 1000 صحنه رانندگی در بوستون و سنگاپور جمع آوری شده است؛ دو شهری که به دلیل ترافیک متراکم و موقعیت های رانندگی بسیار چالش برانگیز شناخته شده اند. در این مجموعه داده، صحنه های 20 ثانیه ای به صورت دستی انتخاب شده‌اند تا مجموعه ای متنوع از مانورهای رانندگی، موقعیت های ترافیکی و رفتارهای غیر منتظره را نشان دهند. پیچیدگی nuScenes، توسعه روش هایی را تشویق می کند که رانندگی ایمن را در مناطق شهری با ده ها شی در هر صحنه امکان پذیر می کند [10].



تصویر ۱. موقعیت دوربین‌های مجموعه داده nuScenes برای جمع‌آوری داده‌ها[10]

مجموعه داده oxford robotcar

در طول دوره May 2014 تا December 2015، با استفاده از پلت فرم Oxford RobotCar، یک نیسان LEAF، مسیری را با شروع از مرکز آکسفورد شروع کرده است. این منجر به بیش از 1000 کیلومتر رانندگی ضبط شده به همراه تقریباً 20 میلیون تصویر جمع آوری شده از 6 دوربین نصب شده روی خودرو، همراه با GPS، LIDAR و INS شده است. داده ها در تمام شرایط آب و هوایی از جمله باران شدید، شب، نور مستقیم خورشید و برف جمع آوری شده است و همچنین عملیات‌های راه‌سازی و ساختمانی در طول یک سال به طور قابل توجهی بخش های مسیر را از ابتدا تا انتهای جمع آوری داده ها تغییر داد. با پیمودن مکرر یک مسیر در طول یک سال، امکان بررسی مکان‌یابی و نقشه‌برداری بلندمدت برای وسایل نقلیه خودران در محیط‌های شهری واقعی و پویا را امکان‌پذیر می‌کنیم. تصویر ۲ موقعیت و جهت هر سنسور را در خودرو نشان می دهد. کالیبراسیون های بیرونی دقیق برای هر سنسور در ابزار توسعه گنجانده شده است [9].



تصویر۲. مکان دوربین‌های نصب‌شده بر ماشین برای ایجاد مجموعه داده [9]

شبیه‌ساز CARLA

CARLA برای پشتیبانی از توسعه، آموزش و اعتبارسنجی سیستم‌های رانندگی خودران توسعه یافته است [11]. علاوه بر کدها و پروتکل‌های منبع باز، CARLA دارای طرح‌های دیجیتال (طرح‌بندی شهری، ساختمان‌ها، وسایل نقلیه) است. پلتفرم شبیه‌سازی از مشخصات انعطاف‌پذیر مجموعه‌های حسگر، شرایط محیطی، کنترل کامل همه بازیگران استاتیک و پویا، تولید نقشه‌ها و موارد دیگر پشتیبانی می‌کند. با استفاده از api ای که این مجموعه داده در اختیار ما قرار داده است، میتوان مکان و جهت سنسورهای روی ماشین را در شبیه‌ساز تنظیم کرد و خروجی این دوربین‌ها را به عنوان عکس در مجموعه داده جمع‌آوری کرد. ما از CARLA برای ایجاد مجموعه دادگان تصاویر که بتوان دوربین‌ها و میدان دید آنها را در موقعیتی قرار داد که خروجی ترکیب تصاویر در این مجموعه داده نسبت به مجموعه داده‌هایی که در دنیای واقعی جمع‌آوری شده‌اند طبیعی‌تر باشد که در نتیجه در تولید نقشه عمق از این تصاویر، تصاویر طبیعی‌تر و با دقت بالاتری را بتوان ایجاد کرد. برای این منظور، ما از یک سورس کد موجود در github برای دانلود و تنظیم این مجموعه داده استفاده کرده‌ایم [12].

نمونه‌ای از تصاویر ایجادشده توسط شبیه‌ساز [12]

پیاده‌سازی و نتایج

برای مشاهده‌ی نتایج می‌توانید به نوتبوک‌های ما مراجعه کنید. در این بخش می‌توانید تعدادی از فریم‌های حاصل از stitching را مشاهده کنید. این تصاویر مربوط به سمت چپ ماشین می‌باشند:

چالش‌ها در پروژه

یکی از چالش‌های مهمی که در جمع‌آوری مجموعه داده‌ها وجود داشت، حجم بسیار زیاد آنها برای استفاده بود که تعدادی از آنها برای دانلود و ذخیره به فضای بیشتر از ۱۵ گیگ نیاز داشتند و با توجه به اینکه فضای ذخیره‌سازی به این میزان را نداشتیم، در بعضی از موارد مجبور به استفاده از بخشی جزیی از مجموعه داده‌ها شدیم تا بتوانیم بر آنها پیاده‌سازی را انجام دهیم. چالش دیگری که با آن مواجه بودیم میدان دید مجموعه داده‌هایی بود که در دنیای واقعی جمع‌آوری شده بودند. به دلیل میدان دید محدودی که این دوربین‌ها داشتند، باعث می‌شد که در روند ترکیب کردن تصاویر تعداد نقاط کلیدی مشترک بین تصاویر کاهش چشمگیری داشته باشد و در نتیجه تخمین خوبی را از ماتریس homography نتوانیم ایجاد کنیم.

آینده‌نگری و استراتژی

شرکت های خودروسازی می توانند با تمرکز بر خودرو های خودران و حتی تغییر و حذف آینه بغل و جایگزین کردن آن ها با دوربین ها و با استفاده از روش ذکر شده در بالا، می‌توانند به همراه بالا بردن امنیت و دقت و کمک به راننده، به جمع آوری مجموعه داده نیز بپردازند و به این مورد ارزشمند دست پیدا کنند.

منابع و مراجع

[1] Blindspot zone. link:
https://www.researchgate.net/figure/The-blindspot-zone-description-We-define-the-blindspot-of-a-driver-as-the-zone-he-can_fig1_221355854

[2] Brown, Matthew, and David G. Lowe. "Automatic panoramic image stitching using invariant features." *International journal of computer vision* 74 (2007): 59-73.

[3] Lin, Chung-Ching, et al. "Adaptive as-natural-as-possible image stitching." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.

[4] Ho, Tuan, et al. "360-degree video stitching for dual-fisheye lens cameras based on rigid moving least squares." *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017.

[5] Godard, Clément, et al. "Digging into self-supervised monocular depth estimation." *Proceedings of the IEEE/CVF international conference on computer vision*. 2019.

[6] Johnston, Adrian, and Gustavo Carneiro. "Self-supervised monocular trained depth estimation using self-attention and discrete disparity volume." *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*. 2020.

[7] Kumar, Varun Ravi, et al. "Monocular fisheye camera depth estimation using sparse lidar supervision." *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018.

[8] Kumar, Varun Ravi, et al. "Fisheyedistancenet: Self-supervised scale-aware distance estimation using monocular fisheye camera for autonomous driving." *2020 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2020.

[9] Li, Yuyan, et al. "Panodepth: A two-stage approach for monocular omnidirectional depth estimation." *2021 International Conference on 3D Vision (3DV)*. IEEE, 2021.

[10] Maddern, Will, et al. "1 year, 1000 km: The oxford robotcar dataset." *The International Journal of Robotics Research* 36.1 (2017): 3-15.

[10] Caesar, Holger, et al. "nuscenes: A multimodal dataset for autonomous driving." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.

[11] Dosovitskiy, Alexey, et al. "CARLA: An open urban driving simulator." *Conference on robot learning*. PMLR, 2017.

[12] Nett, Ryan, "CARLASim". Github: https://github.com/rnett/CARLASim, 2020.