

۵-۱ مقدمه‌ای بر بینایی برجسته‌بین^۱: هندسه سه بعدی

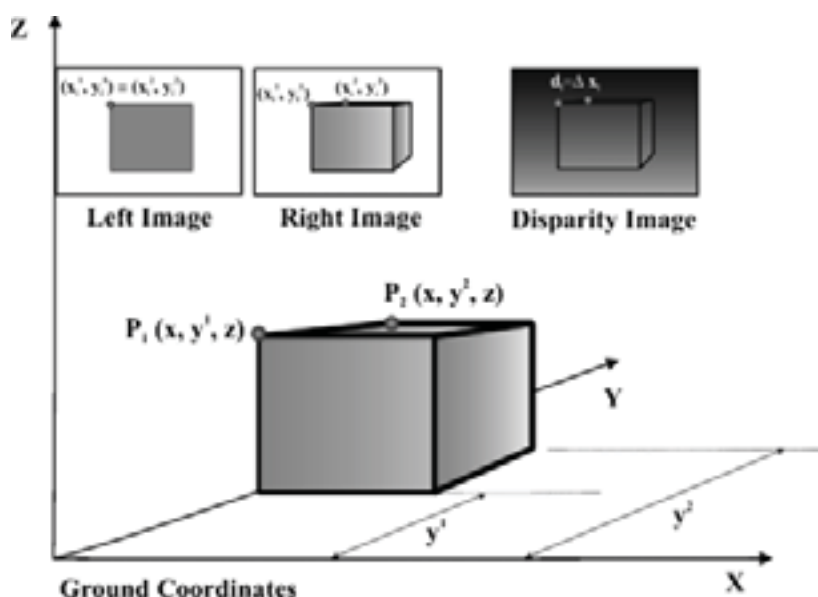
جهان قابل رؤیتی که ما در آن زندگی می‌کنیم دنیایی سه بعدی است که سه متغیر مستقل برای تعیین محل و مرزبایی اشیایی که با حواس خود می‌توانیم درک کنیم کفایت می‌کند. بلندا، پهنا و ژرفا اطلاعاتی کافی برای محصور کردن ناحیه‌ای از فضا که توسط یک شی بی‌جان یا جان‌دار اشغال شده است و همچنین برای موقعیت‌یابی دقیق آن، فراهم می‌آورند. بینایی شگفت‌انگیزترین حس انسان است. دو چشم ما، خردمندانه در ارتفاعی یکسان از زمین و فاصله‌ی افقی مشخصی از یکدیگر قرار دارند و به مغز اجازه مقایسه تصاویر ثبت شده توسط شبکه‌های دو چشم را می‌دهد تا فاصله اشیای پیرامون تا چشم‌ها را تخمین بزند. این اصل بنیادی برجسته‌بینی است که هم توسط انسان و هم ادراک رایانه‌ای به کار می‌رود. برای بررسی یک صحنه در ادراک رایانه‌ای به جای استفاده از مغز برای تطبیق تصاویر حاصل از چشم‌های چپ و راست و استنتاج خصوصیات سه بعدی (3D)، یک الگوریتم کامپیوتری یک جفت تصویر دیجیتالی را که هم‌زمان توسط یک دوربین استریو گرفته شده است به یکدیگر مرتبط می‌کند و اشیای یکسانی را در هر دو تصویر شناسایی کرده و تفاوت آنها را از لحاظ موقعیت پیکسل‌ها اندازه‌گیری می‌کند. این تفاوت عدم شباهت^۲ نامیده می‌شود و توأم با پارامترهای دیگر دوربین، رایانه‌ها را قادر می‌سازد که یک صحنه واقعی را در یک تصویر مجازی سه بعدی از نو خلق کنند. پیروی از برتری بینایی انسان آرزوی افرادی است که در زمینه روباتیک و هوش مصنوعی کار می‌کنند. اصول برجسته‌نمایی جدید نیستند، این اصول در قرن نوزدهم کشف شدند. افرادی که از خانه آبراهام لینکلن در اسپرینگفیلد (ایلینوی، ایالات متحده آمریکا) بازدید کرده‌اند یک استریوسکوپ برای مشاهده دوبینی^۳ تصاویر برجسته را به یاد خواهند آورد. این ابزار با ایجاد توهمی از عمق، یک تفریح بدیع در آن زمان محسوب می‌شد. پیش از این که فناوری روباتیک قادر به بهره‌گیری از اثر استریو^۴ باشد می‌بایست بیش از یکصد سال از ساخت این استریوسکوپ سپری می‌شد. اگرچه، ماشین بینایی تک‌چشمی در طی سه دهه گذشته به کرات مورد بررسی قرار گرفته است، اما تنها در طی ده سال گذشته حسگرهای استریو مورد محبوبیت قرار گرفته و ترویج یافتند. دلیل این تأخیر پیچیدگی در یافتن هم‌زمان یک شی در هر دو تصویر بود. در اواخر دهه ۱۹۹۰، پردازنده‌ها از سرعت کافی برخوردار بودند تا الگوریتم‌هایی که اجازه هم‌بستگی موفقیت‌آمیز هم‌زمان تصاویر استریو می‌دهند را اجرا کنند.

بنابراین مدتی بعد دوربین‌های دوچشمی فشرده در بازار حسگرها و الکترونیک با رشد فزاینده‌ای ظاهر شدند و کاربردهای ادراک برجسته برخط در بسیاری از زمینه‌ها، بخصوص روباتیک متحرک فراگیر شد. هرچند، دستیابی به این موفقیت فناوری محور، تأثیری بر خودروهای کشاورزی نداشت، مگر در مواردی نادر، اما در عین حال پتانسیل فوق‌العاده‌ای برای استفاده در خودروهای غیرجاده‌ای هوشمند داشت. ظرفیت نگاشت صحنه‌های کشاورزی در سه بعد گامی قابل توجه و رای تصویربرداری دوبعدی است که در حال حاضر در کاربردهای سنسجش از راه دور فراهم است. تشخیص آنی موانع، تعداد زیادی گزینه حفاظتی از خودروهای خودکار پیش رو قرار می‌دهد، که پس از تحلیل تصاویر استریو درختان و محصولات ردیفی می‌توان دستورات ناوبری را تعیین کرد.

یک صحنه سه بعدی را هرگز نمی‌توان به طور منحصر به فردی به وسیله یک تصویر دوبعدی نشان داد زیرا یک نقطه در صفحه تصویر را می‌توان کاملاً به وسیله مختصات تصویر آن (x^p, y^p, z^p) توصیف کرد، در حالیکه همان نقطه در فضا نیازمند اطلاعاتی از مختصات سه بعدی (x^p, y^p, z^p) است. ژرفا یا همان فاصله بین نقطه و صفحه تصویر اطلاعاتی است که در یک

- 1 - Stereoscopic Vision
- 2 - Disparity
- 3 - Cross-eyed
- 4 - Stereo Effect

تصویر ساده از یک صحنه سه بعدی از دست می‌رود. لذا یک تصویر ثانویه از همان صحنه گرفته می‌شود تا اثر استریو^۵ حاصل شود و با تطابق این دو تصویر بُعد سوم که در تصاویر دو بعدی از دست رفته است محاسبه شود. این واقعیت به این معنی نیست که بینایی ماشینی تک‌چشمی ارزشمند نیست، بلکه بالعکس به‌خصوص در ارتباط با دوربین‌های تک‌لنزی کاربردهای وسیعی دارد. البته، بینایی برجسته‌بینی احتمالات و روشی کاملاً جدید را برای درک اشیا می‌دهد. در واقع، هر دو فناوری ماشین‌بینایی مکمل هم بوده و اغلب زمانی که یکی با مشکل مواجه شود، دیگری بر آن برتری دارد، و بالعکس. طبیعت هر کاربرد فاکتور تصمیم‌گیری است که کدام یک انتخاب شود، البته در حالات کلی، روش تصویربرداری استریو راه حل کامل‌تری را ارائه می‌دهد زیرا مختصات سومی را به دو مختصات دیگر که در بینایی تک‌چشمی موجود است اضافه می‌کند. واضح است این پیچیدگی نیاز به محاسبات بسیار بیشتر پردازش تصویر منجر می‌شود که این مورد باید در طراحی ماشین ادراکی^۶ مورد توجه قرار گیرد.



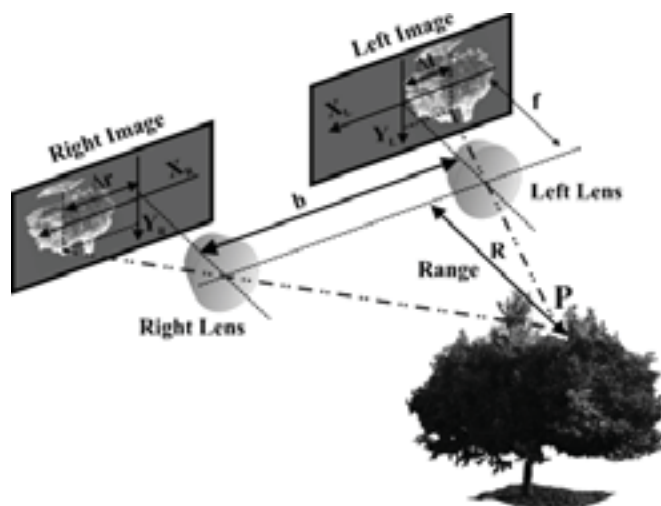
شکل ۵-۱. نمایش مفهومی تولید یک تصویر سه بعدی

یک تصویر سه بعدی با تخمین موقعیت‌های نقاط در صفحات موازی با صفحه تصویر (X و Z مختصات سه بعدی از مختصات تصویر x_i و y_i) ساخته می‌شود و سپس عمق (مختصات y) از تصاویر اختلافی^۷ محاسبه می‌شود. بیابید مسئله بازسازی یک صحنه سه بعدی از تصاویر دوبعدی در طراحی بصری تصویر ۵-۱ را ترسیم کنیم. در این صحنه سه بعدی، P_1 و P_2 دو نقطه مجزا بر شیء مورد نظر هستند، اما در دامنه تصویر دوبعدی مرجع (تصویر سمت چپ) در یک مکان نمایش داده می‌شوند. این دو نقطه تنها از لحاظ موقعیت عمق (مختصات y) تفاوت دارند. اگر تصویر دومی (تصویر سمت راست) باشد که شرایط برجسته‌بینی را برآورده سازد، مشخص می‌شود که نقاط P_1 و P_2 موقعیت‌های مختلفی در تصویر سمت راست دارند که عمق آن‌ها را می‌توان به وسیله تصاویر اختلافی محاسبه کرد. به طوری که اختلاف در هر دو نقطه مقادیر متفاوتی دارند. مختصات کارتزین سه بعدی این دو نقطه توسط $P_1(x, y', z)$ و $P_2(x, y'', z)$ مشخص می‌شود. در ادامه این بخش روابط ریاضیاتی لازم برای انجام این انتقالات و ساخت تصاویر سه بعدی با دوربین‌های برجسته‌بین بیان می‌شود.

هدف محاسبات استریو ساخت یک تصویر سه بعدی از یک جفت تصویر استریو است که هم‌زمان توسط یک دوربین استریو دریافت شده‌اند. این تصاویر استریو دو تصویر دیجیتالی متداول (متشکل از پیکسل) هستند که برای حصول اثر استریو تحت شرایط خاصی گرفته‌اند. این ساختار ویژه، در ساده‌ترین حالت برای نیل به هندسه استریویی^۸، هر دو سنجنده تصویر

- 5 - Stereo effect
- 6 - Perception engine
- 7 - Disparity
- 8 - Stereo geometry

(آرایه‌های الکترونیکی حساس) را وادار می‌کند که لنزها با طول کانونی یکسان^۹ f و در مکان‌هایی که تصویر بردارها در یک فاصله مشخص از صفحه X_1Y_1 قرار گرفته و تصاویر سمت و سمت راست هم‌سطح^{۱۰} گردد. بدین معنی که بازه یا عمق هر نقطه سنجیده شده برای هر دو تصویر جفت استریو دقیقاً یکسان باشد. به‌علاوه، محورهای افقی (X_1) هر دو تصویر باید در یک خط مستقیم واقع شوند که به‌عنوان محدودیت اپیپلار^{۱۱} شناخته می‌شود. شکل ۲-۵ شماتیک یک صفحه استریو گرفته شده به‌وسیله یک دوربین استریو دو چشمی را نشان می‌دهد.



شکل ۲-۵. پارامترهای مهم درگیر در برجسته‌بینی

تصویر شکل ۲-۵ شامل همه پارامترهای کلیدی درگیر در محاسبات استریو است. فاصله بین مراکز کانونی لنزهای سمت چپ و راست به‌عنوان خط مبنای دوربین استریو تعریف و در تصویر با b نشان داده شده است. طول کانونی هر دو لنز دوربین (که برای هر دو باید یکسان باشد) با f نمایش داده شده و نهایتاً بازه یا فاصله سنجیده شده تا شیء توسط R مشخص می‌شود. عمل ضروری در مشاهده استریویی محاسبه نگاشت‌های اختلافی در برگزیده مثلث‌بندی است. به‌منظور ایجاد نمایشی سه‌بعدی از درخت، تصویر ۲-۵، می‌بایست دو تصویر (چپ و راست) هم‌زمان گرفته شوند. در نتیجه، هر نقطه بر روی درخت، فرضاً نقطه P ، در محل‌های مختلف در تصاویر استریو، (x_L, y_L) برای تصویر چپ و (x_R, y_R) برای تصویر سمت راست نگاشت خواهند شد. محدودیت اپیپلار برجسته‌بینی به مختصات X اجازه مقایسه می‌دهد. اگر فواصل افقی تصویر نقطه P بر روی صفحات تصویر توسط Δr برای تصویر سمت راست و Δl برای تصویر سمت چپ نمایش داده شوند، مقدار تفاوت آن‌ها d با اختلاف آن‌ها به‌دست می‌آید که در معادله ۱-۵ بیان می‌گردد:

$$d = \Delta r - \Delta l \quad (5.1)$$

هر دو فاصله $(\Delta l$ و $\Delta r)$ بر حسب پیکسل اندازه‌گیری می‌شوند و علائم آن‌ها به محل قرارگیری شیء مورد نظر و سیستم مختصاتی انتخاب شده برای تصاویر سمت چپ و راست بستگی دارد. اختلاف پارامتر بنیادی مورد استفاده برای محاسبه بازه نقطه P به روش مثلث‌بندی است. شکل ۳-۵، پارامترهای هندسی اصلی در محاسبه بازه نقطه P است که Δr و Δl با توجه به مراکز هندسی تصاویر چپ و راست اندازه‌گیری می‌شوند. یادآوری می‌شود که با توجه به سیستم مختصاتی انتخاب شده در تصاویر ۲-۵ و ۳-۵، Δl منفی و Δr مثبت است. محور X بر خلاف جهت تعریف شده بود، علائم Δr و Δl باید برعکس می‌گشت. اما بزرگی اختلاف باید همان مقدار به‌دست آمده، یعنی برابر با افزوده هر دو کمیت باقی بماند.

- 9 - Identical
- 10 - Coplanar
- 11 - Epipolar

خط مبنا (b)، نشان داده شده در شکل ۳-۵ به فاصله AC موازی با صفحات تصویر را می‌توان با معادله ۳-۵ محاسبه کرد که تشابه مثلث‌ها معادله ۲-۵ را شکل می‌دهد. به طوری که در تصویر ۳-۵ برقرار است.

$$\begin{cases} \frac{R}{AB} = \frac{f}{|\Delta l|} \rightarrow \overline{AB} = \frac{|\Delta l|.R}{f} \\ \frac{R}{BC} = \frac{f}{|\Delta r|} \rightarrow \overline{BC} = \frac{|\Delta r|.R}{f} \end{cases} \quad (5.2)$$

$$b = \overline{AC} = \overline{AB} + \overline{BC} = \frac{|\Delta l|.R}{f} + \frac{|\Delta r|.R}{f} = (|\Delta l| + |\Delta r|) \cdot \frac{R}{f} = (\Delta r - \Delta l) \cdot \frac{R}{f} = \frac{d.R}{f} \quad (5.3)$$

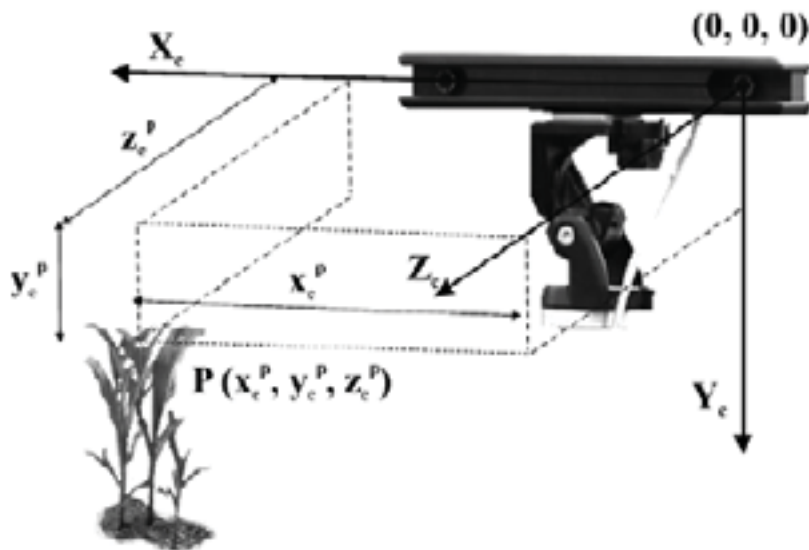
معادله ۳-۵، بیانی را برای بازه به‌عنوان تابعی از خط مبنا، طول کانونی و اختلاف فراهم می‌آورد. البته توجه داشته باشید که یک فاکتور تبدیل بین پیکسل‌ها و فاصله‌ها باید درگیر گردد تا اطمینان حاصل شود که این ارتباط با توجه به واحدها همگن است. این فاکتور اندازه پیکسل‌هایی است که آرایه سنسور - یعنی رزولوشن سنسور - را تشکیل می‌دهند که به‌وسیله ω (mm/pixel) در عملیات پایانی برای بازه نشان داده می‌شود (معادله ۵-۴). استفاده از این معادله ساده است. مثلاً یک دوربین استریو با یک خط مبنا ۹۰ mm و پیکسل‌های مربعی با ابعاد ۰/۰۰۷۵ میلی‌متر یک بازه ۹ m را برای اختلاف ۱۰ پیکسل می‌دهد.

$$R [mm] = \frac{b [mm] \cdot f [mm]}{d [pixel] \cdot \omega [mm \cdot pixel^{-1}]} \quad (5.4)$$

اگر بازه متناظر با محور Z در یک سیستم مختصات دوربین معمولی مانند آنچه که در شکل ۵-۴ نشان داده شده است، باشد، معادله ۵-۴ مختصات z^P نقطه مطلوب P را مستقیماً فراهم می‌کند، اما دو مختصات دیگر X و Y در دنیای واقعی نیز باید محاسبه شوند. فرایند مورد استفاده برای تعیین آنها معرف مثلث‌بندی انجام شده برای کاهش بازه است. اما در عوض مقدار بازه R و مختصات تصویر (x_i^P, y_i^P) ثبت شده در تصویر مرجع (چپ یا راست مطابق با دوربین مخصوص نصب شده بر روی ماشین) استفاده شدند. عبارت پایانی که یک پیکسل را از فضای تصویر به یک نقطه در فضای سه بعدی تبدیل می‌کند توسط معادله ۵-۵ به‌دست می‌آید:

$$\begin{cases} x^P = \frac{x_i^P \cdot \omega \cdot R}{f} \\ y^P = \frac{y_i^P \cdot \omega \cdot R}{f} \end{cases} \quad (5.5)$$

$$z^P = R.$$



شکل ۵-۴. تعریف مختصات دوربین برای سنسورهای استریو دوچشمی

پس از این که سامانه درک^{۱۲} توسط انتخاب پارامترهای بهینه دوربین مانند خط مبنا، لنز، و سنسور تصویر بردار انجام گرفت، تمایل بر این است که این پارامترها ثابت باقی بمانند و بنابراین معادله ۵-۵ را می توان با معرفی ثابت تبدیل مختصات T که به وسیله معادله ۶-۵ تعیین می شود، ساده سازی کرد. معادله مورد استفاده برای تبدیل مختصات تصویر (x_i^P, y_i^P) در فضاییکسل به مختصات سه بعدی در واحدهای طول (x^P, y^P, z^P) به شکل ماتریس در معادله ۷-۵ نشان داده می شود.

$$T = \frac{\omega}{f} [\text{pixel}^{-1}] \quad (5.6)$$

$$\begin{bmatrix} x^P \\ y^P \\ z^P \end{bmatrix} = R_{\text{meters}} \cdot \begin{bmatrix} T & 0 & 0 \\ 0 & T & 0 \\ 0 & 0 & 1 \end{bmatrix}_{\text{pixel}^{-1}} \cdot \begin{bmatrix} x_i^P \\ y_i^P \\ 1 \end{bmatrix}_{\text{pixel}} \quad (5.7)$$

به محض آن که پارامترهای دوربین اصلی برای استفاده خارج از جاده ای انتخاب شدند (بخش ۵-۴)، مهم است که برآوردی از کمترین تغییر در بازه ای که سیستم می تواند تشخیص دهد به دست آید. به بیان دیگر، زمانی که اختلاف یک واحد تغییر می کند، چه تغییری در بازه روی می دهد؟ عدم تطابقها به طور معمول برحسب پیکسل اندازه گیری می شوند، اما دوربین های با عملکرد بالا عموماً در سطوح کوچک تر از پیکسل کار می کنند تا تطابق^{۱۳} را افزایش دهند. تنوع در بازه با توجه به عدم تطابق را می توان با دیفرانسیل گرفتن از معادله ۴-۵ به دست آورد (معادله ۸-۵):

$$\frac{\partial R}{\partial d} = -\frac{b \cdot f}{\omega} \cdot \frac{1}{d^2} = -\frac{b \cdot f}{\omega} \cdot \frac{1}{\frac{b^2 \cdot f^2}{\omega^2 \cdot R^2}} = -\frac{\omega}{b \cdot f} \cdot R^2 \quad (5.8)$$

با جایگزینی دیفرانسیل های معادله ۸-۵ با افزایش های جزئی و چشم پوشی از علامت منفی، معادله ای جبری که اجازه دهد تا دقت برد^{۱۴} به عنوان تابعی از اختلاف برای یک سیستم استریو برآورد شود در معادله ۹-۵ نشان داده می شود:

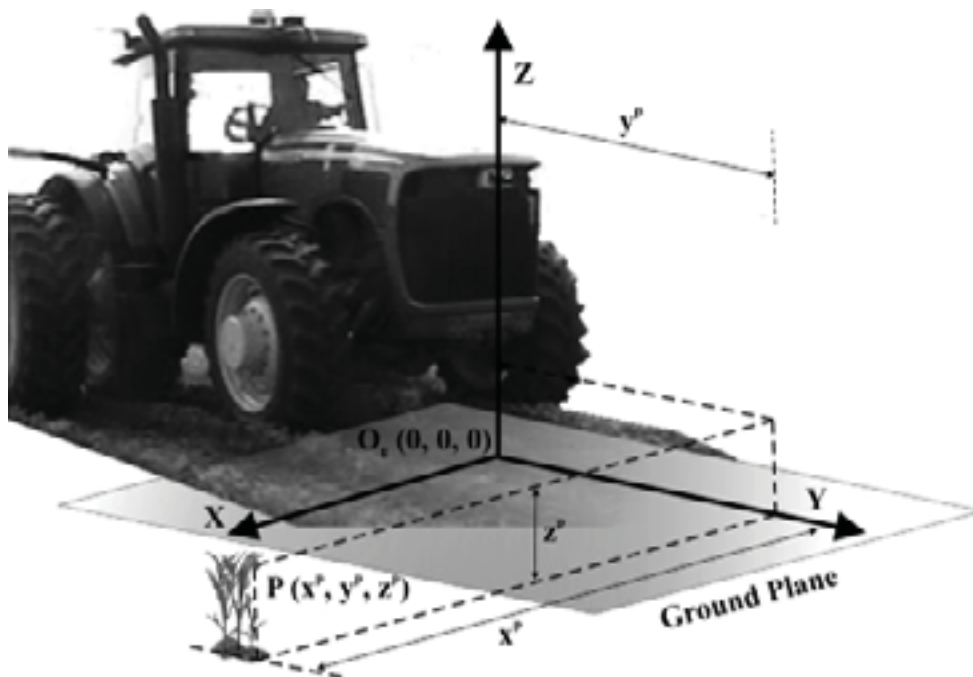
$$\Delta R = \frac{\omega}{b \cdot f} \cdot R^2 \cdot \Delta d \rightarrow \Delta R = \frac{T}{b} \cdot R^2 \cdot \Delta d \quad (5.9)$$

توضیح دقت برد و معادله ۹-۵ داده شده دلالت بر آن دارد که خطوط مبنای وسیع، طول های کانونی بلند، و آرایه های سنسور پیکسل های ریز برای دست یافتن به ΔR کوچک، که منجر به نتایج دقیق تر برای یک عدم قطعیت مشخص Δd می شوند، مطلوب هستند. اگرچه، دقت برد با مربع بازه کاهش می یابد، بدین معنی که با دور شدن شی مورد نظر از دوربین، کاهش دقت برد سریعاً افزایش می یابد.

هدف از انتقال به دست آمده از بکارگیری معادله ۷-۵ محاسبه همه مختصات است که در بردارنده یک صحنه واقعی در یک سیستم کارتیزین سه بعدی مختصاتی است. اگر داده های خروجی دوربین های استریو فضای باز به این فرمت باشند، مهندسی که پیرامون رباتیک کشاورزی فعالیت می کنند باید تلاش خود را متوجه توسعه دادن خودروهای با عملکرد بالا کنند و نه به دنبال بهبود کارهای داخلی سنسورهای خاص باشند که معمولاً شرکتهای تجاری به نحو احسن انجام می دهند. مختصات تصویر و سیستم های مختصاتی تصویر از فرعیات هستند. بالعکس، سیستم مختصاتی توده های نقطه ای سه بعدی^{۱۵} که توسط دوربین استریو ایجاد می شوند و مورد استفاده متخصصین رباتیک و طراحان خودرو هستند امری حیاتی است. مختصات (x^P, y^P, z^P) حاصل از معادله ۷-۵ برای نقطه P به عنوان مختصات دوربین شناخته می شوند و باید مناسب تر برچسب-زنی (x_c^P, y_c^P, z_c^P) شوند. مختصات دوربین نشان داده شده در شکل ۴-۵ برای یک دوربین که لنز سمت چپ آن مرجع است (از پشت دیده می شود)، به طوری که با مبدأ $O_c(0,0,0)$ نشان داده می شود. در این سیستم مختصاتی، صفحه $X_c Y_c$ منطبق با

- 12 - Perception
- 13 - Precision
- 14 - Range Resolution
- 15 - 3D point clouds

صفحه تصویر بردارها^{۱۶} است. بازه^{۱۷} Z_c عمود بر صفحه $X_c Y_c$ است، محور Y_c بعد قائم بر تصویر مرجع که به سمت پایین، روند افزایشی دارد را دنبال کرده و محور X_c منطبق با بعد افقی تصویر مرجع است (از راست به چپ مقادیر مثبت افزایش می یابند). مبدأ مختصات دوربین در مرکز اپتیکی یکی از لنزها است، که طراح دوربین به دلخواه درباره آن تصمیم می گیرد. مختصات دوربین نشان داده شده در شکل ۵-۴ موارد استاندارد هستند که در آن‌ها نقاط توده‌های سه بعدی بیان می شوند، هر چند کاملاً به موقعیت و جهت دوربین وابسته هستند. حرکت دادن سنسور از بالای کابین ماشین برداشت، که سمت پایین را هدف قرار داده است، به دماغه برداشت منجر به تغییری کلی در سیستم مرجع می شود: به سادگی یک تغییر افقی ساده نیست بلکه چرخش را هم شامل می شود، زیرا دوربین ممکن است در راستای مستقیم نگاه کند به جای این که نگاهش روبه جلو باشد. به منظور پیشگیری از این وابستگی، لازم است که یک سیستم جدید مختصاتی تعریف شود. این مختصات باید مستقل از موقعیت و جهت سنسور باشد. تجهیزات خاص مورد استفاده سرانجام خصوصیات مشخصه این سیستم را تعریف می کند، اما بعضی از خصوصیات عمومی را می توان در اینجا بیان کرد. از آنجا که یکی از خصوصیات اصلی روباتیک کشاورزی این است که خودروهای هوشمند باید زمینی^{۱۸} باشند، چنین مختصات زمینی را می توان به صورتی که در ادامه می آید تعریف کرد. محور Z ارتفاع شی را نشان می دهد، محور Y فاصله بین دوربین و شی را تأمین می کند و محور X موقعیت‌های افقی را با توجه به مرکز مختصات ثبت می کند. مبدأ مختصات زمینی بر روی زمین قرار دارد ($Z=0$) و موقعیت دقیق آن در ارتباط با خودرو به وسیله هر دستگاه تعریف می شود. شش مورد مطالعاتی در پایان این فصل مثال‌هایی را از سیستم‌های مختصات زمینی و تبدیلات آن‌ها از مختصات دوربین را فراهم می آورند. شکل ۵-۵ یک سیستم مختصات زمینی عمومی و یک شی که به آن مرجع دهی شده است را نشان می دهد.



شکل ۵-۵. مختصات زمینی برای یک خودروی غیرجاده‌ای هوشمند.

- 16 - Imagers
- 17 - Range
- 18 - Grounded