

استریو ویژن: روشی برای فاصله یابی و غلبه بر مشکلات موجود در سیستم های فاصله یابی کنونی

بنیامین دوجی

دانشکده مهندسی برق و الکترونیک دانشگاه آزاد اسلامی واحد بجنورد
bendavaji@yahoo.com

چکیده: فاصله یابی یکی از مشکلات بزرگی است که برای حل آن راه های متعددی پیشنهاد شده است. هر یک از این راه حل ها مزایا و معایب خاصی دارند. استفاده از دید استریو روشی جدید مبتنی بر دو تصویر گرفته شده از یک صحنه می باشد که می تواند توسط آنها عمق یابی را انجام داد. البته برای تفهیم بیشتر ما یک از الگوریتم ساده و بازگشتی برای درک بهتر آن استفاده کرده ایم. پس از درک کلی مدل استریو به بهینه سازی آن می پردازیم تا از آن بتوان در کاربرد های نظامی ، صنعتی ، ناوبری رباتیک و مواردی از این قبیل استفاده کرد. البته لزوم استفاده از الگوریتم های کلریشن ما را مجبور به استفاده از نرم افزار Matlab در اثبات بازگشتی آن نموده است.

کلمات کلیدی: دید / استریو – ناوبری – کلریشن

۱- مقدمه

عملکرد دقیق یک سیستم بستگی به صحت اطلاعات ورودی و کیفیت طراحی خود سیستم دارد و تا زمانی که ورودی ها درست و به موقع به آن وارد نشوند هر مقدار هم که طراحی سیستم قوی و دقیق باشد از کارایی لازم برخوردار نخواهد بود. فاصله نیز به عنوان یک ورودی بسیار مهم در سیستم ها نقش بسزائی را داراست. برای مثال در یک سیستم نظامی داشتن فاصله دقیق از هدف و البته استخراج بلادرنگ (Real Time) این فاصله از اهمیت بسیار بالایی برخوردار است.

از این رو این مقاله به معرفی و بررسی مدل جدیدی برای استخراج فاصله از تصویر را ارائه می نماید. مدل استریو مدلی نو ظهور است که نظریه آن حدود یک دهه پیش مطرح شده و استفاده از آن نیز اندک زمانیست که شروع شده و در حال پیشرفت می باشد. این سیستم الهام گرفته شده است از سیستم بینایی و پدیده چشم غالب در انسان که در مورد آن بحث خواهد شد. در ابتدا به معماری سیستم ، نحوه عملکرد و اثبات الگوریتم آن خواهیم پرداخت و سپس برای بررسی روش های بهینه سازی و کاربرد های آن اقدام می نماییم. البته پیش نیاز پرداختن موارد به فوق تعریف فاصله سه بعدی و حل معادلات پرسپکتیو می باشد که پیشتر به آنها اشاره خواهیم داشت.

۲- روشهای متداول فاصله یابی (مزایا و معایب)

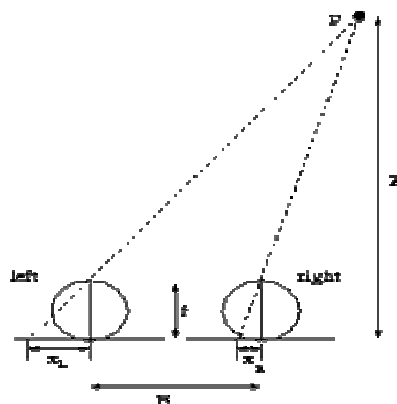
همان طور که از آن اطلاع داریم روشهای متعددی برای فاصله یابی وجود دارد که در ذیل به بررسی مزایا و معایب و محدودیت های هر یک و مقایسه آنها با یکدیگر خواهیم پرداخت. بکارگیری مادون قرمز یکی از متعارف ترین روش ها می باشد که کاربرد های آن به دلیل رنج محدود به کاربرد های خاصی محدود می شوند و تاثیرات محیطی نیز مزید بر علت گشته و بر محدودیت های کاربرد آن می افزایند. روش دیگری که دارای کاربرد های بسیار گسترده ای می باشد استفاده از سنسور های التراسونیک است که تفکیک زاویه ای کم آن یکی از مشکلات استفاده از آن می باشد. همان طور که میدانید در تصاویر سه بعدی فاصله هم به صورت سه بعدی تعریف شده و متفاوت از فاصله دو بعدی می باشد. برای مثال اگر فاصله دو بعدی در دو قسمت یک جسم نسبت به مبدا برابر باشند در فضای سه بعدی این فاصله ها برابر نخواهند بود. سیستم التراسونیک نیز برای تک تک اجزای تصویر نمی توان فاصله را استخراج نماید بلکه برای سطوح یک فاصله میانگین در نظر گرفته می شود و تفکیک زاویه ای کم استفاده از این روش را به کاربرد هایی که نیاز به دقت کمتری دارند، محدود می کند.

استفاده از لیزر نیز با توجه به پیشرفت های حاصله در این زمینه و نظر به واگرا بودن نور معمولی می تواند گزینه بسیار مناسبی باشد. البته دقت این روش بسیار بالا بوده ولی هزینه بالا و مصرف توان زیاد، استفاده از آنرا محدود به کاربرد هایی که احتیاج به دقت بسیار زیاد دارند، می نماید. در سیستم های نظامی از امواج رادار و امواج میلی متری فاصله یاب نیز استفاده می گردد که دارای دقت بسیار بالایی می باشند ولی تاثیرات محیطی آثار سوئی روی آن دارد که برای حذف این آثار و تقویت سیگنال به نویز احتیاج به سیستم های خاصی می باشد که بر هزینه های استفاده از آنها می افزاید. البته لازم به ذکر است که تمام سیستم های معرفی شده فوق اکتیو بوده و توان مصرف می نمایند.

با استریو ویژن یا دید دو گانه که به بررسی ساختار آن خواهیم پرداخت، به فاصله ای بادقت معقول خواهیم رسید که حجم بالای محاسباتی آن بلادرنگ ساختن آنرا با مشکل روبرو می سازد. اما با روش های بهینه سازی سخت افزاری و نرم افزاری می توان تا حدودی بر این مشکل غلبه نمود. در هر حال مناسب ترین سیستم برای هر کاربردی باید با لحاظ کردن تمام شرایط بکارگیری انتخاب شود.

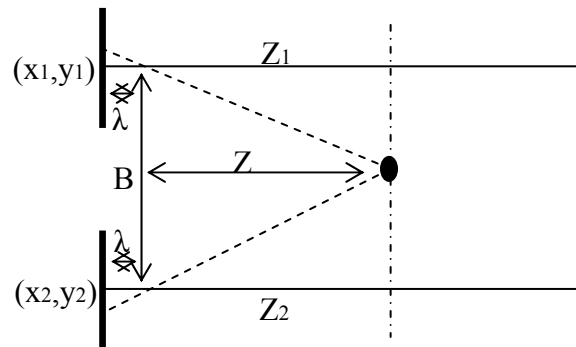
۳- حل معادلات پرسپکتیو و فاصله یابی به کمک آن

نگاشت صحنه سه بعدی بر روی صفحه تصویر تبدیل چند به یک است. یعنی نقطه تصویر به طور یکتا محل نقطه مورد نظر را معین نمی کند. با این حال اطلاعات عمق (Depth) از دست رفته را میتوان با روش های تصویر بر داری استریو به دست آورد.



شکل ۱-۳: محل دوربین ها و مختصات های تصویر.

همان طور که شکل فوق نشان میدهد فاصله دو مرکز عدسی، محور عدسی (Base line) نامیده میشود و هدف پیدا کردن مختصات (X, Y, Z) نقطه W با داشتن نقاط تصویر (x_1, y_1) و (x_2, y_2) است. در ابتدا برای اثبات و تفهیم بهتر فرض میشود که دوربین ها کاملاً هم راستا هستند و تنها در محل مبدا با هم تفاوت دارند.



شکل ۲: مختصات سیستم استریو و مولفه های کالیبراسیون.

طبق شکل فوق مبدا مختصات را مرکز دوربین اول می گیریم و مختصات نقاط دیگر را بر آن منطبق می نماییم، از تبدیلات پرسپکتیو نتیجه می شود که W روی خط L با مختصات $X_1 = (x_1 / \lambda)(\lambda - z_1)$ قرار دارد. زیر نویس های X, Z دلالت بر این دارد که دوربین اول بر مبدا مختصات قرار دارد و W و دوربین دوم به آن منطبق می باشند و اگر با حفظ آرایش نسبی دوربین دوم را مبدا مختصات بگیریم و نقاط دیگر را بر آن منطبق کنیم، W روی خطی با مختصات جزئی زیر قرار خواهد داشت:

$$X_2 = (x_2 / \lambda)(\lambda - z_2)$$

در هر حال به دلیل فاصله بین دوربین ها و نیز به دلیل اینکه مختصه Z مربوط به W برای هر دو سامانه مختصات دوربینی

یکسان است، نتیجه می شود که:

$$X_2 = x_1 + B ; z_2 = z_1 = Z$$

که B طول خط پایه می باشد.

با جایگذاری معادلات فوق در معادلات خطوط قبلی داریم:

$$X_1 - B = (x_2 / \lambda)(\lambda - Z)$$

با تفریق معادله های بالا و حل حاصل برای Z داریم:

$$Z = \lambda - [\lambda B / (x_2 - x_1)]$$

که نشان می دهد که اگر بتوان تفاضل بین دو مختصه تصویر مربوط به x_1, x_2 را تعیین کرد و طول خط پایه و فاصله کانونی

معلوم باشد، محاسبه مختصه Z مربوط به نقطه W کار ساده ای است و X, Y با استفاده از معادلات زیر از یکی از مختصات (x_1, y_1) و

یا (x_2, y_2) بدست می آید:

$$X = (x_0 / \lambda)(\lambda - Z) ; Y = (y_0 / \lambda)(\lambda - Z)$$

مشکل ترین کار استفاده از معادله $Z = \lambda - [\lambda B / (x_2 - x_1)]$ برای بدست آوردن Z، پیدا کردن دو نقطه متناظر در دو تصویر مختلف یک صحنه است. چون این نقاط معمولاً در نزدیکی هم هستند، یک روش متداول انتخاب یک نقطه از یک ناحیه کوچک و در یکی از نما های تصویر و سپس تلاش برای پیدا کردن بهترین ناحیه تطبیق کننده (Matching Region) در نمای دیگر با استفاده از فنون همبستگی می باشد. اما وقتی صحنه، ویژگی های مجزایی نظیر گوشه های مهم دارد معمولاً روش تطبیق ویژگی راه حل سریع تری برای تعیین تناظر می باشد. البته در این روش ها فاصله از تمام نقاط تصویر (پیکسل ها) استخراج می شود که مورد نیاز نمی باشد. معمولاً پس از انتخاب هدف مورد نظر فقط استخراج همبستگی برای آن شکل صورت می گیرد. مثلاً در تصاویر نظامی پس از شناخت هدف به وسیله الگوریتم های تشخیص الگو، با استفاده از روش تطبیق ویژگی ها و کلاسه بندی و بانک اطلاعاتی خاص، عمق فقط برای الگو یا همان

هدف مورد نظر استخراج می شود البته یکی از مشکلات اساسی استفاده از مدل استریو بلادرنگ کردن آن است . به علت حجم پردازش بسیار بالا نمی توان با تراشه های معمولی تمام محاسبات را بلادرنگ انجام داد، پس لزوم استفاده از تراشه های سرعت بالا همچون DSB ها نمود پیدا می کند.

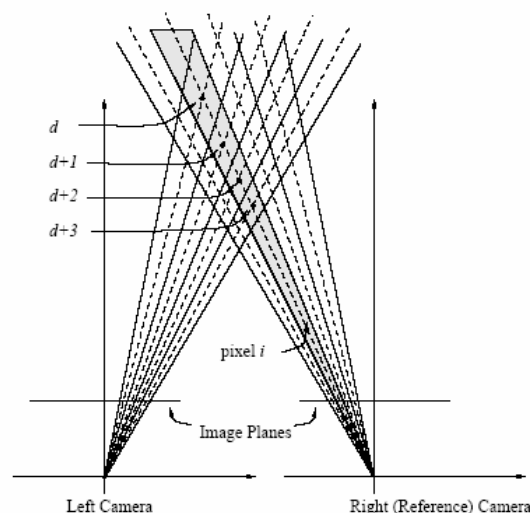
۴- بینایی انسان و چشم غالب

استریو ویژن اولین بار از نحوه بینایی انسان الهام گرفته شد . در سیستم بینایی انسان پدیده جالبی بنام چشم قالب و جود دارد طبق این پدیده یکی از چشم های انسان دقیقا مانند دوربین (مبدا مختصات) مرجع به عنوان چشم مرجع عمل می نماید و برای یافتن این چشم می توان آزمایش ساده ای به شرح زیر ترتیب داد:

ورقی را لوله نموده از درون آن به تصویری با فاصله تقریبا متوسط از خود نگاه می کنیم. در این حال با دست دیگر جلوی یکی از چشمانمان را سد می کنیم. این کار را برای چشم دیگر نیز به روشی مشابه انجام داده خواهید دید که نتایج مختلف از این دو آزمایش پدید می آید. وقتی یکی از چشمان را بگیریید هیچ تفاوتی در تصویری که از داخل کاغذ لوله شده دیده می شود رخ نمی دهد اما با گرفتن چشم مرجع تصویر داخل کاغذ شیفیت پیدا می کند که به این چشم ، چشم مبنا در سیستم بینایی انسان گفته می شود. در سیستم دید استریو نیز یک دوربین به نام دوربین مبنا انتخاب می شود و نقطه مبدا مختصات روی آن قرار می گیرد و فاصله (x_1-x_2) که توسط الگوریتم های شباهت استخراج می شود بر طبق این نقطه بدست می آید. به این فاصله شیفیت افقی یا دیسپیریتی (Disparity) گفته می شود. چون در یک سیستم استریو ایدال با شرایط کالیبراسیونی در نظر گرفته شده (منظور فاصله کانونی - نوع دوربین ها - مختصه های پیم و تیلت در آنها می باشد). دو تصویر فقط در یک شیفیت افقی با هم اختلاف خواهند داشت که با داشتن مقدار این شیفیت افقی می توان فاصله را از تصاویر بکمک فرمول زیر بدست آورد:

$$Z = \lambda B / d$$

در آن d همان شیفیت افقی در دو تصویر می باشد.



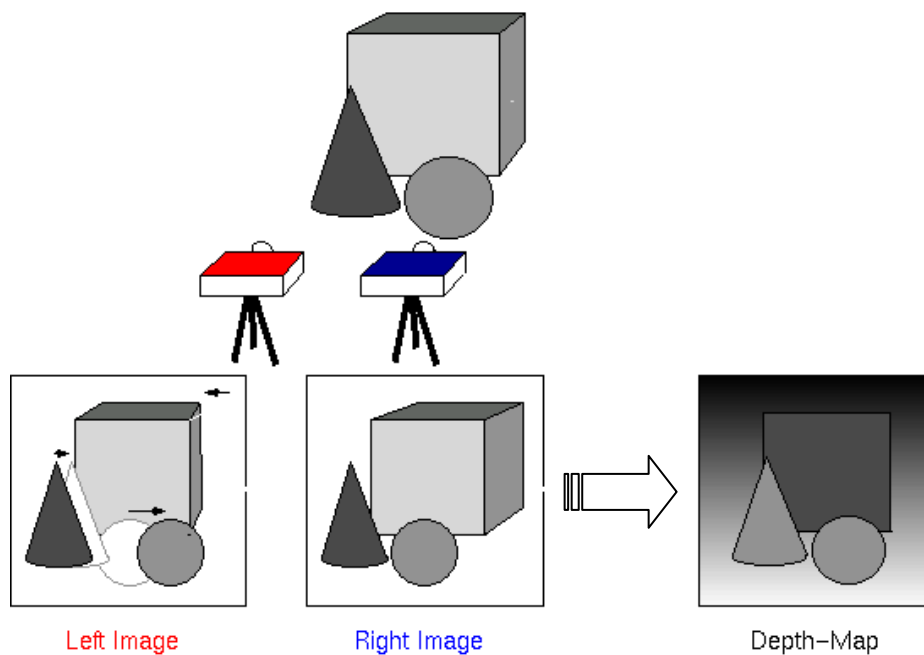
شکل ۴-۱: نحوه استخراج پیکسل ها در سیستم استریو.

۵- استخراج دیسپیریتی از تصاویر استریو

پس از اعمال شرایط کالیبراسیونی مورد نظر به دوربین های استریو، تصویر استریو مشخص شده و به سیستم ورودی وارد می شود. در ابتدا برای بهبود نتیجه یک سری فیلتر ها برای تضعیف نویز صورت می گیرد و در صورت لزوم از الگوریتم های تشخیص الگو

برای تشخیص هدف استفاده می شود و پس از تشخیص هدف ناحیه مورد نظر از تصویر برای محاسبات ارجاع می یابد، که به کمک این روش حجم محاسباتی بسیار کم و سرعت استخراج نتیجه بالا می رود. پس از این کار نوبت به استخراج تناظر می رسد که این کار را می توان توسط الگوریتم ها و ماسک های مختلف انجام داد که در موارد مختلف نتایج متفاوتی را می دهد و برای هر کاربرد روش خاصی که بهترین جواب را می دهد بهترین گزینه خواهد بود. ما در این مقاله به روش همبستگی برای استخراج بیشترین تناظر می پردازیم.

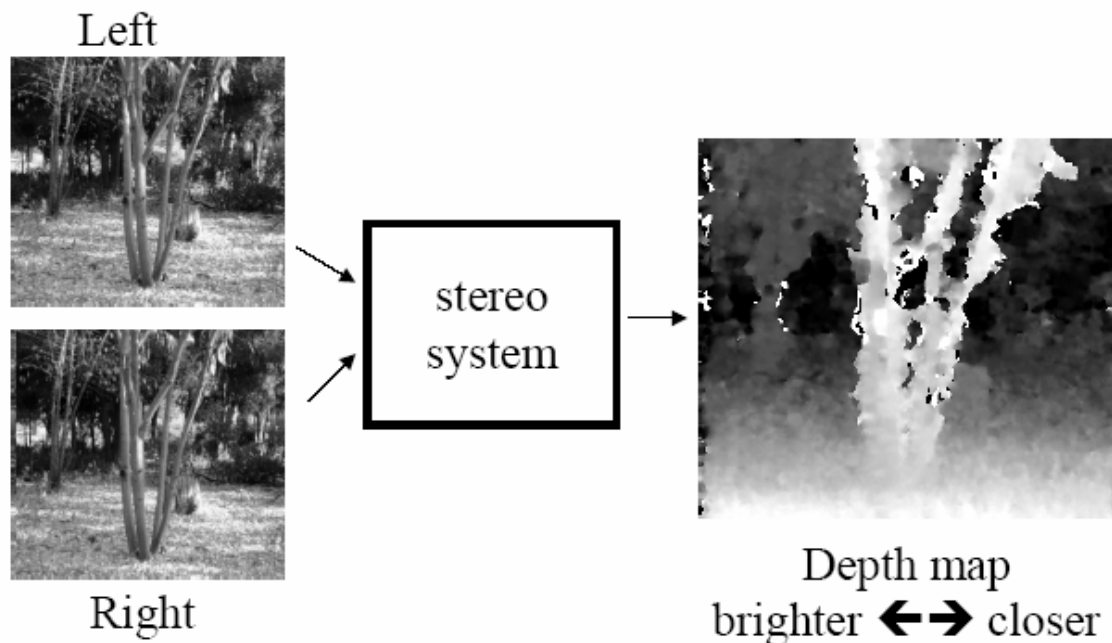
کلریشن یا همبستگی یک نقطه از تصویر اول را با تصویر دوم استخراج می کنیم و با مقایسه این مقدار با مقدار اتو کرولیشن یا خود همبستگی تصویر اول با خودش مقدار شیفت افقی نقطه مورد نظر از تصویر اول را در تصویر دوم پیدا می کنیم. مهمترین مسئله مورد نظر نحوه مقایسه و الگوریتم آن می باشد که باز هم با توجه به شرایط محیطی و دقت مورد نیاز در مقایسه با سرعت استخراج داده متفاوت بوده و بهترین گزینه طبق نتایج حاصله از مقایسه کاربرد پیدا می نماید. در این مقاله از روشی به نام SAD استفاده می کنیم که در ذیل به شرح آن پرداخته شده است:



شکل ۱-۵: شیفت افقی در تصاویر استریو.

مجموع مقادیر اختلافات مقدار همبستگی هر نقطه از تصویر اولی با تصویر دوم و خود همبستگی تصویر اول با خودش را که باید دارای مقداری مثبت بوده و در غیر اینصورت باید در داخل قدر مطلق قرار داده و مقدار حاصله دیسپیریتی برای آن نقطه را به دست می دهد. همانطور که پیش تر گفتیم این الگوریتم مقایسه گر تنها یکی از شیوه های استخراج مقدار دیسپیریتی (شیفت افقی) برای دو تصویر است و با توجه به دقت و سرعت آنالیز داده مورد نیاز از روشی مناسب برای این کار استفاده می نمایند. چون الگوریتم برای هر یک از نقاط تصویر محاسبات لازم را انجام می دهد نتیجه حاصل ماتریسی به نام ماتریس دیسپیریتی است که مقادیر هر یک از آرایه های آن برابر با مقدار دیسپیریتی آرایه متناظر با آن در تصویر اول می باشد، که در این ماتریس هرچه جسم نزدیک تر باشد پررنگ تر (سفید تر) و هرچه جسم دور تر باشد تیره تر دیده می شود. لازم به ذکر است که مقادیر این ماتریس می توانند دارای مقادیر ۰ تا ۲۵۵ باشند که دور ترین نقطه دارای مقدار ۰ و نزدیک ترین نقطه دارای مقدار ۲۵۵ به عنوان دیسپیریتی می باشد. با داشتن مقدار دیسپیریتی برای هر نقطه و دانستن شرایط و مقادیر کالیبراسیونی می توان از فرمول زیر به استخراج فاصله پرداخت:

$$Z = \lambda B / d$$



شکل ۲-۵: یک نمونه دیسپریتی استخراج شده.

۶- اثبات الگوریتم به شیوه باز گشتی با کمک نرم افزار

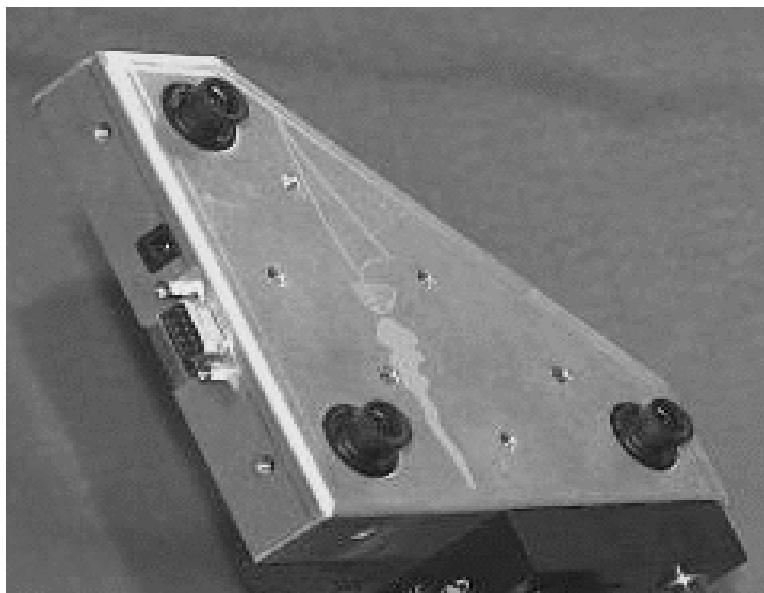
با چند فرض کلی و محدود کننده به بررسی شیوه عملکرد این سیستم می پردازیم که شرایط مفروض و محدودیت های کالیبراسیونی به شرح ذیل می باشند:

نوع دوربین ها باید کاملاً یکسان فرض شود و هر پیکسل در تصویر باید دارای مولفه های مستقل از پیکسل های دیگر باشد و نیز فرض می شود که تمام نقاطی که روی یک ستون از تصویر قرار دارند در فاصله یکسانی نسبت به دوربین ها قرار دارند که این فرض یک فرض غیر ممکن می باشد ولی برای اثبات و تفهیم بهتر این فرض لحاظ شده و در پایان اصلاح می گردد تا مطابق با شرایط واقعی گردد. در این حین با داشتن مقایر فاصله کانونی و فاصله پایه دوربین ها تنها مجهول باقی مانده مقدار دیسپریتی می باشد که به شیوه زیر به استخراج آن می پردازیم:

یک تصویر را در نظر گرفته و آنرا به صورت تصاویر سطح خاکستری در می آوریم و دو سطح برابر از این تصویر را که فقط در راستای افقی با هم تفاوت دارند یعنی شیفته یافته هم هستند را در نظر می گیریم و مقدار شیفته نیز که همان دیسپریتی است را می توان تنظیم نمود. سپس الگوریتم همبستگی را برای آن به شرح زیر در نرم افزار مطلب پیاده سازی می کنیم و خواهیم دید که نتیجه حاصله با دیسپریتی مفروض برابر خواهد شد.

```
F=imread('pic.jpg');
D1=F(300:400,300:400);
D2=F(300:400,310:410);
figure; imshow(D1)
figure; imshow(D2)
GD1=double(rgb2gray(D1));
GD2= double(rgb2gray(D2));
temp1=xcorr2(GD1,GD1);
for i=1:100
    R=xcorr(GD1,GD2(:,i));
    out(i)=sum(abs(R-temp1));
end
[S1 S2]=min(out);
```

طبق الگوریتم فوق تصاویر به صورت آرایه double تعریف شده اند تا بتوانند تمام مقادیر غیر صحیح را نیز بپذیرند و S1 مقدار دیسپرستی . S2 شماره سطر را می دهد که دقیقا منطبق با مفروضات مسئله می باشند. همانطور که دیدید این الگوریتم برای تمام ستون های تصویر محاسبات را انجام داد و اگر الگوریتم دیگری تمام کار های فوق را برای سطر ها نیز انجام دهد آن هنگام با ترکیب نتایج حاصله از دو بررسی دیگر نیازی به فرض یکسان بودن فاصله روی سطر یا ستون خاصی نمی باشد. و در واقع با ترکیب این دو روش روشی نو به نام الگوریتم سه گوش یاسه گانه که از ترکیب دو سیستم استریو استفاده می نماید بهره جسته ایم.



شکل ۱-۶: یک نمونه استریو مثلثی (سه گانه).

۷- نتیجه گیری

همانطور که نشان دادیم مدل استریو یک مدل جدید برای فاصله یابی است که دارای دقت معقول و قابل تنظیم است که ماتریسی از فاصله در اختیار ما می گذارد که نشان دهنده تفکیک زاویه ای بالای آن است که می توان از این مزیت در مدلسازی های محیط برای اهداف ناوبری در رباتیک و اهداف نظامی از آن استفاده نمود. البته در سینمای مجازی نیز داشتن فاصله بصورت زنده و سه بعدی از اهمیت خاصی برخوردار است و استفاده از سیستم های استریو در کمپانی های فیلم سازی نیز رواج یافته و در حال رشد می باشد. البته روشی بنام پچلت (patchlet) نیز وجود دارد که محاسبات را بر روی المانهای سطح انجام می دهد و برتری شگرفی در سرعت و دقت نسبت به محاسبات روی هر پیکسل یا سطر و ستون خاص دارد. و با توجه به نو ظهور بودن و وارد نشدن این سیستم به کشور تا کنون جا دارد که به مطالعات و تحقیقات برای دست یابی به آن اقدام نماییم تا بتوان از مزایای این روش مفید و نسبتا ارزان در کاربرد های مختلف استفاده نماییم.

۸- قدر دانی و تشکر

افراد زیادی در به ثمر رساندن این مقاله به من کمک کردند که از همه آنها تشکر می نمایم و لازم می دانم از مهندس شهپری برای تشکیل یک گروه پردازش تصویر در دانشگاه قدر دانی نمایم و از زحمات تمام دوستانم در این گروه به ویژه تیم استریو ویژن تشکر می نمایم، و از دوست عزیزم آقای احمد بهرمند که در ویرایش و آماده سازی این مقاله مرا یاری نمودند تشکر و قدر دانی می نمایم.

۹- مراجع

[۱] گونزالس, پردازش تصویر رقمی- مترجم- دکتر خادمی, دانشگاه فردوسی مشهد, پاییز ۸۳ .

[۲] دوجی, ب و جوهری نیا, س و رحیمی, آ – استفاده از سیستم استرو ویژن در ناوبری و نقشه برداری ربات – سومین سمینار سراسری باشگاه دانش پژوهان جوان, ۱۳۸۳.

[3] D. Murray and J. Little. Surface segmentation from stereo scenes. In 3DPVT 2004.

[4] D. Murray. Patchlets: a method of interpreting correlation stereo 3D data. PhD thesis, University of British Columbia, Vancouver, Canada, 2004.

[5] D. Murray and J. J. Little. Using real-time stereo vision for mobile robot navigation. *Autonomous Robots*, 8(2):161–171, 2000.

[6] C. Martin and S. Thrun. Real-time acquisition of compact volumetric maps with mobile robots. In *Proceedings of the IEEE*

[7] International Conference on Robotics and Automation (ICRA 2002), Washington, DC, 2002. M. Shimizu and M. Okutomi.

Precise sub-pixel estimation on area-based matching. In Proceedings of the 9th IEEE International Conference on Computer Vision (ICCV 2001), volume I, pages 90–97, Vancouver, BC, Canada, 2001.

[8] O D. Faugeras. Three-Dimensional Computer Vision: A Geometric Viewpoint. MIT Press, 1993.

[9] S. Zucker and R. Hummel. A three-dimensional edge operator. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 3(3):324 { 331, 1981.