



روش جدید تناظریابی اتوماتیک در بینائی استریو با استفاده از شبکه های عصبی و همبستگی

شهریار برادران شکوهی

دانشگاه علم و صنعت، گروه مهندسی برق

bshokouhi@iust.ac.ir

محمد علیپور

صدا و سیمای جمهوری اسلامی ایران،

alipour@iribu.com

1- مقدمه :

بینائی استریو به معنی استخراج و تخمین عمق با استفاده از میزان عدم تطابق بین دو تصویر از یک صحنه می باشد. برای این کار سه مرحله کالیبراسیون، تعیین نقاط متناظر و حل مساله تطابق و نهایتاً بازیافت مختصات سه بعدی از روی آن مطرح می شود. در این مقاله سعی شده با روشهای نوین نسبت به حل دو مساله اخیر و نهایتاً استخراج کلید عمق اقدام شود. دو تکنیک عمده و خاصیت مهم در حل مساله تطابق نقش اساسی دارد:

• بر پایه محیط *area-based*

• بر پایه مختصات و شرایط *feature-based*

در مورد هر دو خاصیت به اندازه کافی صحبت شده است اما هر بار روشهایی برای اجرای موارد فوق مطرح می شود. در این میان شبکه های عصبی نیز در حل این مساله جایگاه و نقش خاصی دارند.

در [2,1] با استفاده از یک شبکه عصبی پس انتشار خطا یا پرسپترون شرایط کنترل کننده هندسی را برای استخراج عمق و پیدا کردن نقاط متناظر بکاربرده است. در [3,1] از شبکه عصبی برای خوشه بندی و کوانتیزه کردن جهت اجماع خواص و شرایط *feature-based* استفاده شده

چکیده: به منظور انتخاب روش مناسب شناسائی عمق، با تخمینی مناسب و در زمان کم الگوریتم هایی ارائه شده است یکی از این الگوریتم ها، بینائی استریو می باشد که به منظور استخراج عمق بکار میرود. در کل برای استخراج دقیق این داده ها روشهای متنوعی بصورت مکمل و در کنار هم استفاده می گردد.

در این مقاله با استفاده از شبکه عصبی، تصاویر را کلاس بندی و با *correlation* بیشینه تطابق نقاط متناظر سمت راست و چپ را یافته و در عین حال از یک شبکه عصبی دیگر برای یافتن نقشه عمق استفاده می کنیم.

واژه های کلیدی: نگاشت عمق¹، بینائی استریو چند دوربینه²، تشخیص الگو³، شبکه های عصبی⁴، اختلاف مکانی⁵.

¹ Depth Map

² Stereo – Multi view

³ Pattern Recognition

⁴ Neural Network

⁵ Disparity

جهت استخراج پارامترهای دوربین می باشد و همانطور که در رابطه 2 می بینید ماتریس M با آرایه های m شامل ترکیبی از متغیرهای فوق الذکر است که از کالیبراسیون حاصل می شود . روابط ارائه شده در 3 محاسبات پارامترهای داخلی و خارجی دوربین را نشان می دهد .



شکل 1- مکعب سه بعدی جهت کالیبراسیون

$$\begin{pmatrix} l_i x_i \\ l_i y_i \\ l_i \end{pmatrix} = M \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix} \quad (1)$$

می توان نوشت :

$$\begin{pmatrix} f_x r_{11} + o_x r_{31} & -f_x r_{12} + o_x r_{32} & -f_x r_{13} + o_x r_{33} & -f_x T_x + o_x T_z \\ f_y r_{21} + o_y r_{31} & -f_y r_{22} + o_y r_{32} & -f_y r_{23} + o_y r_{33} & -f_y T_y + o_y T_z \\ r_{31} & r_{32} & r_{33} & T_z \end{pmatrix} \quad (2)$$

که در آنها :

است . روش های دیگری برای حل مساله تطابق در راستای خواص گفته شده در مرجع [4] آمده است .

حتی در مرجع [1] ترکیبی از چند شبکه عصبی برای تخمین $disparity map$ استفاده شده است .

در این مقاله سعی می کنیم با استفاده از دو نوع شبکه عصبی و الگوریتم $correlation$ تخمینی اتوماتیک از $disparity map$ داشته باشیم .

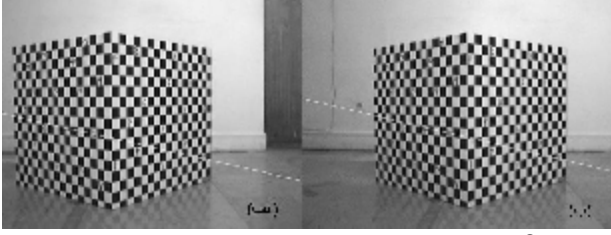
تازگی این روش در آموزش کاملاً اتوماتیک و هوشمند شبکه عصبی دوم که نقش عمومیت دادن یک شبکه عصبی را دارد ، برای تخمین $disparity map$ و بدنبال آن $depth map$ است .

در این مقاله سعی شده با آموزش صحیح شبکه عصبی دوم با استفاده از داده های شبکه عصبی اول و سایر راه کارها بطور اتوماتیک شبکه آموزش ببیند .

2- پردازشهای اولیه :

یکی از عملیات مهم و لازم جهت کاراستخراج پارامترهای دوربین از راه کالیبراسیون می باشد .

از مراجع [5,6,7] استفاده نموده تا با آزمایشی کالیبراسیون دوربین را بطور کامل انجام دهیم . یکی دیگر از کارهائی که در این پژوهش انجام شد مشخص کردن میزان غیر خطی بودن تابع تبدیل لنز است ، بایستی توجه شود که آموزش شبکه عصبی موقعی صحیح خواهد بود و به تبع آن به تخمین درستی از نگاشت عمق می رسیم که عملکرد دوربین ها خطی بوده یا لااقل در ناحیه خطی از آن عمل کنیم . بدین منظور ما دوربین ها را در حالت زوم بکار برده و کناره های تصویر را دور می ریزیم . (مرکز تصویر خطی تر است) . چنانچه در شکل 1 و رابطه 1 مشخص است هدف تعیین ماتریس M میباشد که با توجه به مجموعه روابط ارائه شده در 2 و 3 [6] پارامترهای لازم محاسبه می شود . در روابط ارائه شده O مبدا مختصات صفحه تصویر (IP) ، T ، ماتریس انتقال و r آرایه های ماتریس دوران و f فاصله کانونی دوربین می باشد . رابطه 1 رابطه کلی کالیبراسیون



شکل 2- خطوط اپی پلار متناظر در تصاویر راست و چپ

3- تناظر یابی :

میزان انطباق بین دو ویندو از این روابط حاصل می شود :

$$(5) \quad NCC = \frac{\sum_{(u,v) \in W} (I_l(u,v) \times I_r(x+u, y+v))}{\sqrt{\sum_{(u,v) \in W} I_l^2(u,v) \times \sum_{(u,v) \in W} I_r^2(x+u, y+v)}}$$

$$ZNCC = \frac{\sum_{(u,v) \in W} [(I_l(u,v) - \bar{I}_l) \times (I_r(x+u, y+v) - \bar{I}_r)]}{\sqrt{\sum_{(u,v) \in W} (I_l^2(u,v) - \bar{I}_l) \times \sum_{(u,v) \in W} (I_r^2(x+u, y+v) - \bar{I}_r)}}$$

اساس این محاسبات از روی همبستگی بین دو توزیع آماری حاصل میشود. [8,12]

کافی است روی خطوط اپی پلار در دو تصویر حرکت و میزان انطباق آنها را در آرایه ای ثبت و سپس ماگزیموم مقدار آن را پیدا نمود. اختلاف مکانی بین دو نقطه در دو تصویر سمت راست و چپ در هنگام رخداد متناسب با اطلاعات عمق آن نقطه خواهد بود.

4- کلاسه بندی کردن تصاویر از روی

مشخصات مهم تصویر :

یکی از روشهای مهم حل مساله انطباق روش *feature-based* است. [14, 13, 1]

اینک با استفاده از شبکه های عصبی مشخصاتی را هم زمان برای تصویر تعیین و بر مبنای آن تصاویر را کلاسه بندی می کنیم. همچنان که در روابط 6 آورده شده است خواص مهمی چون لبه ها، گرادیان و جهت تغییرات آن مورد نظر ما در این مقاله بوده است :

$$T_z = s m_{34}$$

$$r_{3i} = s m_{3i}$$

$$o_x = q_1^T q_3$$

$$o_y = q_2^T q_3$$

$$f_x = \sqrt{q_1^T q_1 - o_x^2}$$

$$f_y = \sqrt{q_2^T q_2 - o_y^2}$$

$$r_{1i} = s(o_x m_{3i} - m_{1i}) / f_x$$

$$r_{2i} = s(o_y m_{3i} - m_{2i}) / f_y$$

$$T_x = s(o_x m_{34} - m_{14}) / f_x$$

$$T_y = s(o_y m_{34} - m_{24}) / f_y$$

$$f_x = f / \text{pixel width}$$

$$f_y = f / \text{pixel height}$$

$$(3)$$

با بدست آوردن پارامترهای لازم از کالبراسیون و مشخص شدن پارامترهای محدود کننده اپی پلار به اعمال عملیات *correlation* بر روی خط اپی پلار با عرض ویندوی 9×9 می پردازیم.

با توجه به روابط مثلثاتی می توان روابط 4 را بدست آورد که در آن اندیس k شماره دوربین k ام می باشد که در این تحقیق برای بینائی دو دوربینه $k=2$ می باشد. اندیس im برای مختصات تصویر بکار رفته است.

$$X_{c1} = -Z_{c1} \cdot (x_{im1} - o_{x1}) / f_{x1}$$

$$Y_{c1} = -Z_{c1} \cdot (y_{im1} - o_{y1}) / f_{y1}$$

$$\begin{bmatrix} X_{ck} \\ Y_{ck} \\ Z_{ck} \end{bmatrix} = R_k \cdot R_1^{-1} \cdot \begin{bmatrix} X_{c1} - T_{1x} \\ Y_{c1} - T_{1y} \\ Z_{c1} - T_{1z} \end{bmatrix} + \begin{bmatrix} T_{kx} \\ T_{ky} \\ T_{kz} \end{bmatrix}$$

$$x_{imk} = -f_{xk} \cdot X_{ck} / Z_{ck} + o_{xk}$$

$$y_{imk} = -f_{yk} \cdot Y_{ck} / Z_{ck} + o_{yk}$$

$$(4)$$

در شکل 2 خطوط متناظر اپی پلار حاصل از محاسبات فوق برای دو تصویر راست و چپ آمده است :

کنیم. در اینصورت برای هر مختصاتی در یک تصویر متناظری در تصویر دیگر پیدا می شود. میزان اختلاف مکانی¹ های بدست آمده را در ماتریسی جمع کرده و دو ماتریس شامل اختلاف مکانی ها را برای دو تصاویر اصلی و تصاویر کلاسه بندی شده با هم مقایسه می کنیم. دو مقدار اختلاف مکانی که باهم حد اکثر 3 نقطه اختلاف دارند (تعیین آستانه) به عنوان یک اختلاف مکانی معتبر و در غیر اینصورت اختلاف مکانی غیر معتبر می نامیم، (طراحی تابع تصمیم گیری). اگر d_c میزان اختلاف مکانی نقاط کلاسه بندی شده و d_m مقدار اختلاف مکانی نقاط تصاویر اصلی باشد برای هر نقطه (x, y) می توان نوشت:

$$\begin{aligned} & \text{if}(\text{abs}(d_m - d_c) \leq 3) \\ & \{ \\ & \quad \text{point is valid} \\ & \quad \text{save in file} \\ & \} \end{aligned} \quad (7)$$

که محور اصلی تابع تصمیم گیری ما را تشکیل می دهد. نقاط تایید شده در تابع مذکور نقاط آموزش شبکه دوم برای تخمین نقشه عمق کل تصویر خواهد بود.

6- استخراج نقاط برای آموزش شبکه عصبی دوم برای تخمین نگاشت عمق:

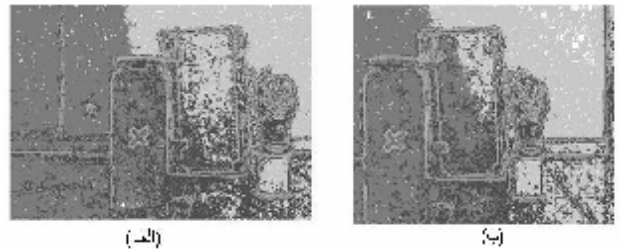
نقاط معتبر حاصل از پروسه بند 5 را بصورت اتفاقی و به میزان حدود 15٪ کل نقاط تصویر در نظر می گیریم و مختصات آنها (x, y) و شدت روشنایی تصاویر اصلی متناظر $f(x, y)$ و شدت روشنایی تصاویر کلاسه بندی شده متناظر آنها $f_c(x, y)$ را به عنوان ورودی به یک شبکه عصبی پس انتشار با سه لایه میانی به ترتیب با 7، 15 و 3 سلول دارای تابع خطی جهت آموزش اعمال می کنیم. شبکه به سرعت همگرا می شود و آموزش می بیند.

$$\begin{aligned} & x \\ & y \\ & f(x, y) \\ & \nabla f(x, y) = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2} \\ & q = tg^{-1}\left(\frac{\partial y}{\partial x}\right) \end{aligned} \quad (6)$$

با اینکار توانسته ایم تا تمام مشخصات ویژه یک تصویر را یکجا جمع کنیم، شکل 4 نشان دهنده این کلاسه بندی به 7 کلاس برای دو تصویر شکل 3 می باشد. حال با استفاده از متد دیگری خاصیت *area_based* را برای حل مساله تطابق در نظر می گیریم و آنهم اجرای الگوریتم *correlation* در امتداد خطوط اپی پلار متناظر هم می باشد.



شکل 3- دو تصویر سمت راست و چپ



شکل 4- تصاویر کلاسه بندی شده (7 کلاس) تصاویر راست و چپ

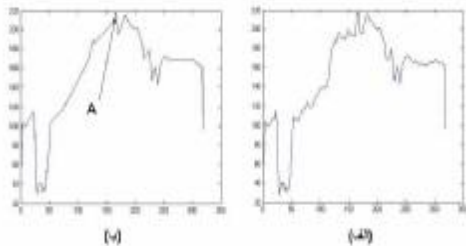
5- محاسبه میزان انطباق تصاویر اصلی و کلاسه بندی شده و استخراج نقاط اختلاف مکانی معتبر با تعریف یک تابع تصمیم گیری:

حال محدودیت خط اپی پلار و *correlation* را روی دو تصویر اصلی و تصاویر کلاسه بندی شده اعمال می

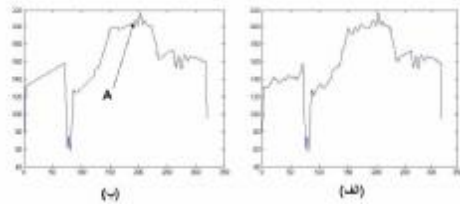
¹ disparity

$$\begin{bmatrix} 38.9312 \\ -100.3587 \\ 149.3000 \end{bmatrix} \quad (9)$$

رابطه 8 مقادیر عملی $R_k R_l^{-1}$ بازای $K=2$ و رابطه 9 مقادیر عملی ماتریس انتقال T می باشد که در بند 2 توضیح داده شد. چون دو دوربین در جایگاه های خاص مکانیکی ثابت شده اند، دیگر مرحله استخراج این ماتریسها برای آزمایشات بعدی ضروری نخواهد بود. در جدول 1 با توجه به نقاط مهم روی منحنی حاصل از شدت روشنایی پیکسل های در طول خط اپی پلار (شکل 7 و 8)، میزان صحت نقاط متناظر بدست آمده در دو تصویر بصورت آماری و ارزش انتظاری در حالت اولیه حاصل از تنها الگوریتم *Correlation*، در مقایسه با مقدار متوسط صحت تطابق بعد از تخمین آورده شده است زمانهای پردازش بسیار قابل توجه اند.

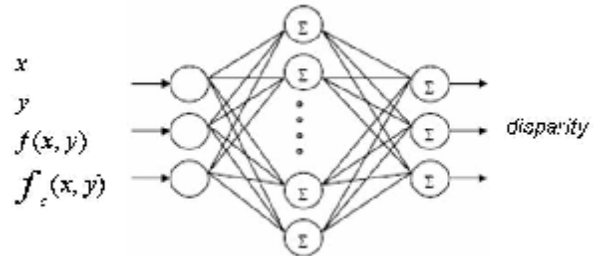


شکل 7 - نمودار تغییرات شدت روشنایی پیکسل ها در تصویر سمت راست (الف)؛ با جزئیات (ب)؛ با تعداد نقطه کم ولی شامل اطلاعات اصلی



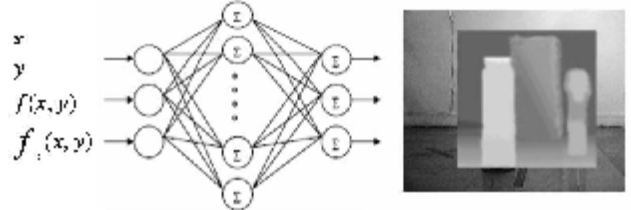
شکل 8 - نمودار تغییرات شدت روشنایی پیکسل ها در تصویر سمت چپ (الف)؛ با جزئیات (ب)؛ با تعداد نقطه کم ولی شامل اطلاعات اصلی

تخمین بعد از اعمال شبکه عصبی (پردازش موازی)	تنها الگوریتم <i>Correlation</i>	
بیش از 90٪	حداکثر 60٪	متوسط میزان صحت تطابق
2 دقیقه	8 دقیقه برای	زمان پردازش (پیچیدگی زمانی)



شکل 5- اساس آموزش و عملکرد شبکه عصبی دوم

مشاهده می شود که شبکه دوم بطور اتوماتیک آموزش دید و این نقطه برجسته این پروژه میباشد. حال چنانچه در شکل 6 نمایش داده شده است به شبکه آموزش دیده بصورت اتوماتیک فوق سایر نقاط را اعمال و میزان عمق را استخراج می کنیم.



شکل 6- نتیجه نهائی عملکرد شبکه عصبی دوم و تخمین نگاشت عمق

7- نتایج عملی :

این آزمایشات با دوربین *creative* و تصاویر گرفته شده در ابعاد 640×480 پیکسل می باشد. برای اجرای برنامه از یک کامپیوتر *dual CPU Supermicro PIV* استفاده شده و کد برنامه *MATLAB* می باشد. توجه شود که بعلت غیر خطی بودن دوربین و زمان بر بودن محاسبات (تعداد پیکسل زیاد) قسمت وسط تصویر که دوربین در آن ناحیه خطی تر عمل میکند مورد استفاده قرار گرفته است. درضمن این نتایج در آزمایشگاه صدا و سیمای جمهوری اسلامی ایران بدست آمده است.

7-1 نتیجه کالیبراسیون برای کل دورانهای موقعیت دوربین راست نسبت به دوربین سمت چپ (رابطه 8) و میزان انتقال آن به میلی متر (رابطه 9) در روابط زیر آورده شده است :

$$\begin{bmatrix} -0.8404 & 0.0021 & 0.5419 \\ 0.0011 & 1.0000 & -0.0022 \\ 0.5419 & 0.0012 & 0.8404 \end{bmatrix} \quad (8)$$

8- مراجع :

- [1]Jung-Hua:National Taiwan Ocean University and Chih-Ping :Aetex Biometric Group, "On Disparity Matching in Stereo Vision via a Neural Network Framework",1999.
- [2]Ruichek, Y: Univ. of Technol. Of Belfort , France, "Multilevel- and neural network Stereo matching",2005.
- [3]Jia Li , Robert M.Gray :Stanford University : "Image Compression and Classification with Vector Quantization",1999.
- [4]Daniel Scharstein:Middlebury college , "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms",IEEE,2002.
- [5]Horst A.Beyer," Advances in characterization and of calibration of digital imaging system ", Institute Geodesy and photogrametry,1992.
- [6]"Camera Calibration", computer vision ,lecture9:<http://hompagees.inf.ed.ac.uk/rbf/CVonline.2005>.
- [7] "Real-Time Camera Calibration for Virtual Studio" : <http://www.idealibrary.com> , 2000, Academic Press .
- [8]Alan V.Oppenheim , "Discrete-Time Signal Processing" , Prentice-Hall International 1999.
- [9]Louis J. Galbiati, Jr. "Machine Vision and Digital Image Processing Fundamentals ",State University of New York,1990.
- [10] Andre Redert, Jan Biemond, " Correspondence Estimation in Image Pairs ",IEEE Signal Processing Magazine,2000.
- [11]M.Hariyama ,T.Takeuchi and M.Kameyama, " VLSI Processor for Reliable Stereo Matching Based on Adaptive Window-Size Selection ",Proceedings of the 2001 IEEE International Conference on Robotics & Automation pp. 1168-1173,May 2001.
- [12]Jung-Hua,dep,Hsien-Chu:National Taiwan Ocean University ,Ru-Feng:Hsien-Chu:UMAX Data System Inc, "A Progressive Constraint Search Approach for Disparity Matching in Stereo Vision",1999.
- [13]Devedra P.Garg and Manish Kumar:Dept. of Mechanical Engineering and Materials Science ,Duke University , " Object Classification via Stereo Vision Sensing in a Flexible Manufacturing Work Cell",IEEE,2003.
- [14]Ruichek, Y , University of Belfort ,France, "Multilevel and neural-network-based stereo-matching method for real-time obstacle detection using linear cameras",IEEE,2005.
- [15]Li Hong and George Chen , "Segment-based Stereo Matching Using Graph Cuts",IEEE,2004.

آموزش و نتایج نهائی کسری از ثانیه	پردازش تصویر 200x180	الگوریتم ها
-----------------------------------	-------------------------	-------------

جدول 1- جدول مقایسه

7-2) نتیجه کلاسه بندی تصویر ها :

شکل 9 نمایش خروجی عملیات کلاسه بندی برای یک تصویر می باشد .



الف) تصویر اصلی ب) تصویر کلاسه بندی شده (9 شکل)

شکل 9- تصویر اصلی (الف) و تصویر کلاسه بندی شده (ب)

7-3) تخمین نگاشت عمق برای تصاویر. (جهت برآورد صحیح سعی شده تا تصاویر با زوایائی گرفته شوند تا کمترین هم پوشانی را داشته باشند در ضمن نور پردازی طوری صورت گرفته تا سایه کمتر ایجاد شود و توجه کنید که در صحنه تغییرات شدید عمق را نداریم .) شکل 10 تخمینی از عمق برای صحنه فوق را نشان می دهد .



شکل 10- تخمین نگاشت عمق

(بعلت ورود اطلاعات feature-base به شبکه عصبی دوم شکاف رنگ دیوار با کد رنگ دیگری به اشتباه برآورد شده است)