

تشخیص درب مبتنی بر بینایی ماشین در صحنه‌های بیرونی

* مهدی طالبی
* دانشجوی دکتری، دانشکده‌ی مهندسی کامپیوتر، دانشگاه اصفهان، اصفهان
** دانشیار، دانشکده‌ی مهندسی کامپیوتر، دانشگاه اصفهان، اصفهان
*** امیرحسین منجمی
*** دانشجوی دکتری، دانشکده‌ی مهندسی کامپیوتر، دانشگاه اصفهان، اصفهان
*** دانشیار، دانشکده‌ی مهندسی کامپیوتر، دانشگاه اصفهان، اصفهان
تاریخ دریافت: ۹۶/۰۲/۱۰ تاریخ پذیرش: ۹۶/۰۷/۰۶

چکیده

درب‌ها نشانه‌ی مهمی جهت ورود و خروج از ساختمان برای افراد نابینا و ربات‌ها می‌باشند. تشخیص درب در محیط‌های بیرونی به یکی از مسایل دشوار در بینایی کامپیوتر تبدیل شده است؛ زیرا معمولاً در درب‌های محیط‌های بیرونی، ویژگی‌های یک درب ساده مانند دستگیره، گوشه‌ها و فضای خالی بین درب و زمین آشکار نیستند. در این مقاله، روشی برای تشخیص درب در محیط‌های بیرونی ارائه می‌شود. پس از استخراج خطوط و حذف خطوط اضافی، ناحیه‌ی بین خطوط عمودی تشکیل می‌شود و ویژگی‌های هر ناحیه شامل ارتفاع، عرض، محل، رنگ، بافت و تعداد خطوط داخل ناحیه استخراج می‌گردند. سپس از دانش اضافی مانند وجود درب در پایین تصویر، ارتفاع و عرض معقول درب و اختلاف رنگ و بافت درب با ناحیه‌ی اطراف، برای تصمیم‌گیری وجود درب استفاده می‌شود. این روش بر روی مجموعه تصاویر eTRIMS و مجموعه تصاویر خودمان شامل درب‌های منازل، آپارتمان‌ها و فروشگاه‌ها امتحان شده است و نتایج ارائه‌شده، برتری روش پیشنهادی نسبت به روش‌های پیشین را نشان می‌دهد.

واژه‌های کلیدی: بینایی کامپیوتر، تشخیص درب، استخراج خطوط، رنگ، بافت

فرد نابینا، ربات و یا ماشین را تا نزدیک مقصد راهنمایی کنند تشخیص درب می‌تواند برای ورود به مکان مورد نظر بسیار مفید باشد. همچنین در مدل‌سازی سه‌بعدی شهری برای برنامه‌ی گوگل ارث^۲، تشخیص نمای ساختمان و اجزای آن شامل درب، جزء گام‌های ضروری است. در دهه‌ی اخیر کارهای زیادی روی تشخیص درب در محیط‌های درونی برای حل مسأله‌ی مکان‌یابی و نقشه‌برداری همزمان^۳ انجام شده است که می‌توان به سه دسته تقسیم کرد:

۱. تشخیص درب با حس‌گرهای تعیین فاصله‌ی لیزری [۲]:

۱. مقدمه

تشخیص درب در محیط‌های بیرونی به دلیل وجود انواع مختلف درب‌ها، شلوغی، سایه‌ها و اختفا یکی از موضوعات پیچیده می‌باشد. در میان هزاران اشیای مختلف که هنوز سیستمی به اندازه‌ی یک کودک دو ساله قادر به شناخت اشیا نیست [۱]، درب یکی از اشیای مهم است که در مسیریابی ربات‌ها و افراد نابینا می‌تواند نقش مهمی ایفا کند. همچنان که ابزار مجهز به جی‌پی‌اس^۱ می‌توانند

روش‌های تشخیص مبتنی بر گرامر [۱۳] می‌باشند. به هر حال در این مجموعه تصاویر نمای ساختمان، عناصر ساختمان مانند درب و دیوار و پنجره و بالکن با نظم خاصی قرار گرفته‌اند و از این محدودیت برای تشخیص استفاده می‌شود.

در بسیاری از مقاله‌ها، تشخیص درب به همراه اشیای دیگر به صورت تجزیه تصویر^۱ انجام می‌شود. منظور از تجزیه تصویر این است که برای هر پیکسل تصویر، برچسب معنایی مانند ساختمان، درب، زمین، آسمان و ... مشخص می‌شود. روش‌های تجزیه تصویر و یا قطعه‌بندی معنایی به دو صورت پارامتریک و غیرپارامتریک انجام می‌شوند. روش‌های پارامتریک مبتنی بر یادگیری هستند که پارامترهای مدل در فاز آموزش تخمین زده می‌شوند [۱۴-۱۷]. در این کارها، ابتدا تصاویر آموزشی به ناحیه‌هایی از پیکسل‌های پیوسته مشابه به نام ابرپیکسل^۱ تقسیم می‌گردند [۱۸] و پس از تعیین خصیصه‌های این ابرپیکسل‌ها، از روش‌های یادگیری ماشینی مانند ماشین‌های بردار پشتیبان^۲ استفاده می‌شوند تا بتوان تصاویر تست را برچسب‌گذاری کرد. از معایب این روش آن است که برای تصاویر جدید آموزشی، دوباره باید یادگیری و تخمین پارامترها انجام شود که بسیار وقت‌گیر است. به هر حال، برای مجموعه تصاویر کوچک با ده‌ها برچسب، روش‌های پارامتریک دقت تشخیص بالایی دارند. در روش غیر پارامتریک تجزیه تصویر، دانش تصاویر آموزشی برچسب‌گذاری شده به تصاویر تست منتقل می‌شود [۲۲-۱۹]. این روش برای مجموعه تصاویر زیاد با صدها برچسب، دقت تشخیص قابل مقایسه با روش پارامتریک را دارد در حالی که پیچیدگی‌های یادگیری را ندارد. به هر حال تجزیه تصویر بسیار پیچیده است و علاوه بر آن، چون چندین شیء با هم تشخیص داده می‌شوند دقت تشخیص درب پایین است در صورتی که درب بسیار مهم‌تر از اشیای دیگر مانند آسمان، بالکن یا پنجره می‌باشد.

این حس‌گرها به ظاهر و رفتار شیء مثلاً رنگ و حرکت درب، حساس هستند.

۲. تشخیص درب با استفاده از اطلاعات بصری تصویر [۵، ۴، ۳]:

این مقاله‌ها خصیصه‌های درب شامل اختلاف رنگی درب و اطراف آن، عدم وجود بافت^۴ در قسمت پایین درب، خطوط عمودی و افقی درب، دستگیره‌ی درب، نقاط محو^۵ فاصله‌ی خالی پایین بین درب و زمین را در نظر می‌گیرند. سپس با استفاده از الگوریتم آدابوست^۶ همه‌ی خصیصه‌های درب ترکیب می‌شوند تا دسته‌بند قوی درب به دست آید و تصمیم‌گیری شود که درب پیدا شده است یا خیر.

۳. تشخیص درب با استفاده از ترکیب حس‌گر و دوربین [۶]: در این کار از حس‌گر مبتنی بر لیزر به همراه استخراج خصیصه‌های تصویر برای بالا بردن دقت تشخیص درب استفاده شده است.

به هر حال در نتایج این کارها گزارش شده است که ربات‌ها به صورت برخط با دقت بالای ۹۰ درصد می‌توانند درب‌ها را تشخیص دهند که البته این دقت برای درب‌های ساده‌ی داخل راهرو به دست آمده است.

با وجود توجه بسیار به تشخیص درب در محیط‌های درونی، دقت تشخیص درب در محیط‌های بیرونی به دلیل پیچیدگی آن، بسیار کم است. در بعضی از الگوریتم‌ها، از توزیع لبه‌ها و استخراج خصیصه‌های تصویر برای تشخیص درب استفاده می‌شود که مجموعه تصاویر آن‌ها بسیار خاص و محدود می‌باشند [۸، ۷]. در کارهای دیگر، تشخیص درب بر روی مجموعه تصاویر نمای ساختمان^۷ مانند ECP2011^۸ با ۱۰۴ تصویر [۹] و یا Graz2012 با ۵۰ تصویر [۱۰] انجام می‌گیرد. روش‌های موجود برای تجزیه نمای ساختمان به صورت روش‌های قطعه‌بندی^۹ معمولی [۱۲، ۱۱] و یا

^۴Texture

^۵Vanishing points

^۶Adaboost

^۷Facade

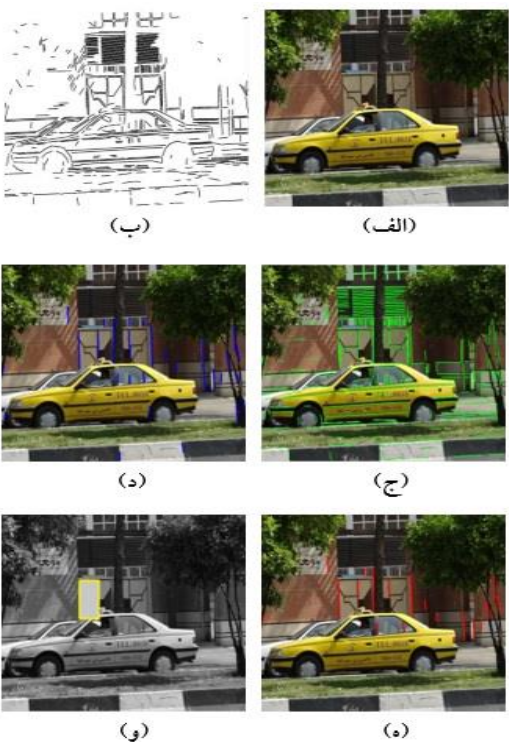
^۸Ecole Centrale Paris

^۹Segmentation

^۱Image parsing

^۲Superpixel

^۳Support vector machines (SVM)



شکل ۱: مراحل تشخیص درب در محیط بیرونی. (الف) تصویر ورودی، (ب) خطوط استخراج‌شده، (ج) تصویر و خطوط عمودی و افقی، (د) تصویر و خطوط عمودی، (ه) تصویر و خطوط عمودی کاندید ناحیه درب، (و) تصویر خروجی و قسمتی از درب تشخیص‌داده‌شده

۱-۲- استخراج خطوط

خطوط می‌توانند نقش مؤثری در شناسایی اشیای ساخته بشری مانند ساختمان و اجزای آن شامل درب و پنجره ایفا کنند. در ابتدا تصویر رنگی به خاکستری تبدیل می‌شود زیرا بر خلاف تصاویر رنگی که از چند کانال استفاده می‌کنند تصاویر خاکستری با یک کانال، باعث پردازش سریع‌تر می‌شوند و همچنین در بعضی مسائل مانند استخراج خطوط نیازی به رنگ تصویر نیست و شدت روشنایی تصویر کفایت می‌کند. در این‌جا از الگوریتم تشخیص قطعه خط (LSD) [۲۵] استفاده می‌شود که یک تصویر خاکستری را به عنوان ورودی می‌گیرد و در سه مرحله مطابق شکل ۳ خطوط تصویر را استخراج می‌کند (شکل ۱.ب).

هدف ما در این مقاله، تشخیص درب در محیط‌های بیرونی است که مراحل تشخیص آن در شکل ۱ نشان داده شده است. در این روش ابتدا خطوط تصویر با الگوریتم LSD [۲۳] استخراج می‌گردند (شکل ۱.ب). سپس خطوط مورب و بسیار کوتاه یا بلند حذف می‌شوند (شکل ۱.ج). پس از آن فقط خطوط عمودی در نظر گرفته می‌شوند (شکل ۱.د) و بین این خطوط عمودی آن خطوطی که پایین‌تر هستند و می‌توانند ناحیه‌های کاندید درب را تشکیل دهند انتخاب می‌گردند (شکل ۱.ه). آن‌گاه ناحیه‌هایی بین هر دو خط عمودی تشکیل می‌شوند و ویژگی‌های هر ناحیه استخراج می‌گردند: مختصات خط عمودی راست و چپ ناحیه، رنگ، بافت و تعداد خطوط عمودی و افقی داخل ناحیه. در آخر با استفاده از این ویژگی‌ها، تصمیم‌گیری می‌شود که کدام ناحیه می‌تواند درب باشد (شکل ۱.و). به هر حال در محیط‌های بیرونی معمولاً رنگ و بافت درب با ناحیه اطراف متفاوت است. همچنین مکان‌های درب در پایین تصویر و تعداد خطوط روی درب بیشتر از جاهای دیگر مانند درخت و ماشین است. مجموعه تصاویری از درب‌های خانه‌ها، آپارتمان‌ها و فروشگاه‌ها تهیه شده است که الگوریتم ما با دقت ۸۸/۴ درصد درب و یا بخشی از آن را مشخص می‌کند ضمن این‌که برای مجموعه تصاویر eTRIMS [۲۴] دقت تشخیص ۷۸/۳ درصد به دست آمده است.

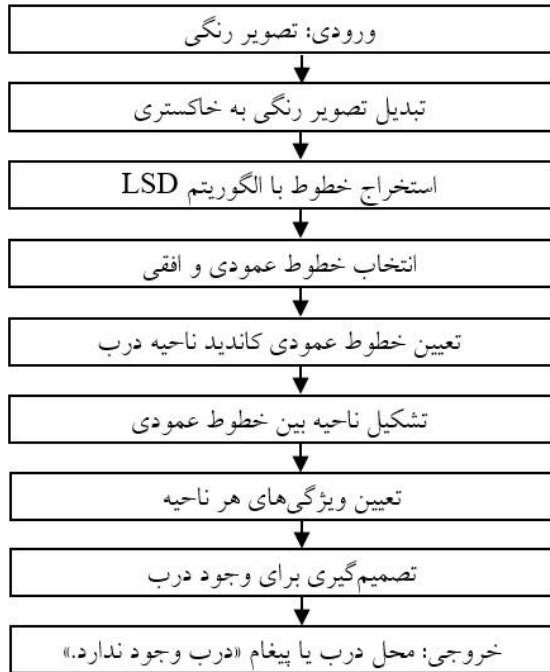
این مقاله به صورت زیر سازماندهی شده است: در بخش ۲ جزئیات روش و الگوریتم تشخیص درب توضیح داده می‌شود و سپس در بخش ۳ نتایج بررسی می‌گردند. در بخش ۴ هم به نتیجه‌گیری و کارهای آینده پرداخته می‌شود.

۲. روش پیشنهادی

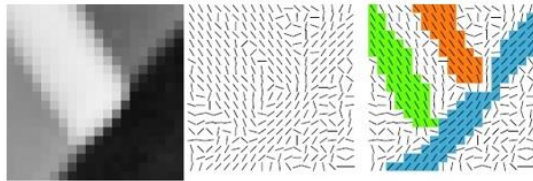
در این قسمت، مراحل تشخیص درب از یک تصویر محیط بیرونی شرح داده می‌شود که بلوک دیاگرام آن در شکل ۲ نشان داده شده است.

^۱Line Segment Detector (LSD)

^۲E-training for interpreting images of man-made scenes (eTRIMS)



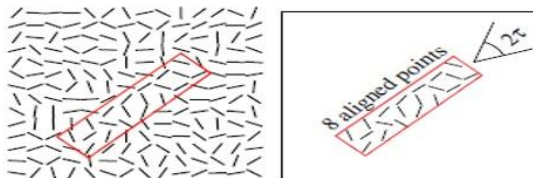
شکل ۲: بلوک دیاگرام روش پیشنهادی تشخیص درب



(الف) تصویر ورودی و تشکیل ناحیه‌های پشتیبان خط



(ب) تقریب مستطیلی ناحیه پشتیبان خط



(ج) بررسی نقاط هم‌راستا

شکل ۳: سه مرحله‌ی الگوریتم LSD [۲۵]

۱. تصویر را با گروه‌بندی پیکسل‌های دارای شیب یکسان، به ناحیه‌های پشتیبان خط تقسیم می‌کند (شکل ۳.الف).
 ۲. قطعه خط مستطیلی شکلی را می‌یابد که بهتر، ناحیه پشتیبان خط در مرحله‌ی ۱ را تقریب می‌زند (شکل ۳.ب).
 ۳. بر اساس تعداد پیکسل‌ها و تعداد نقاط هم‌راستا در ناحیه مستطیل شکل، بررسی می‌کند که آیا این ناحیه می‌تواند تشکیل یک خط دهد یا خیر (شکل ۳.ج).

فرض ما بر این است که تصاویر از دوربین در راستای افق به دست می‌آیند که معمولاً تصاویر هم در واقعیت و هم در مجموعه تصاویر به این صورت هستند. بنابراین فقط خطوط تصویر با ۱۵ درجه‌ی اختلاف با خط عمود و افق به عنوان خطوط عمودی و افقی در نظر گرفته شده‌اند. سپس به دلیل تعداد زیاد خطوط و برای پردازش سریع‌تر، خطوط بسیار کوتاه و بلند در مقایسه با میانگین طول خطوط حذف شده‌اند. همچنین از چند خط موازی بسیار نزدیک به هم فقط بلندترین آن‌ها نگه داشته شده‌اند (شکل ۱.ج).

۲-۲- تعیین خطوط عمودی کاندید ناحیه درب

خطوط عمودی برای پردازش و تعیین ناحیه درب در نظر گرفته شده‌اند (شکل ۱.د) و در مراحل بعد، از خطوط افقی برای تصمیم‌گیری وجود درب استفاده شده است. سپس با استفاده از الگوریتم خوشه‌بندی k -means^۵، این خطوط عمودی بر اساس مختصات پایین‌ترین نقطه به چند ناحیه از خطوط عمودی موازی که در یک ردیف باشند تقسیم شده‌اند. چون در مجموعه تصاویر، درب در پایین تصویر وجود دارد خطوط عمودی پایین‌ترین ناحیه برای پردازش در نظر گرفته شده‌اند. البته اگر در موارد خاص تعداد خطوط ناحیه پایین و میانگین طولی این خطوط، کمتر از یک‌چهارم ناحیه بالاتر باشد ناحیه بالاتری برای وجود درب در نظر گرفته شده است. به عنوان مثال در شکل ۱.د) خطوط عمودی ردیف پایین روی بلوار، درخت و تاکسی حذف و خطوط ردیف بالاتر در شکل ۱.ه) برای تشکیل ناحیه در نظر گرفته شده است.

^۵Clustering

۲-۳- تشکیل ناحیه بین خطوط عمودی

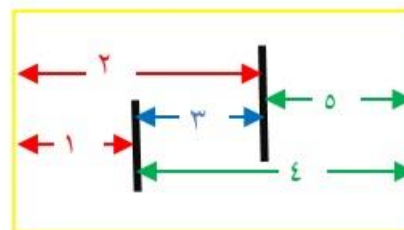
چون در صحنه‌های بیرونی انواع درب‌ها با اندازه‌های مختلف وجود دارند و شاید تصویر از نمای دور گرفته شده باشد و یا قسمتی از درب مثلاً توسط ماشین مخفی شده باشد هر ناحیه کوچک یا بزرگی می‌تواند یک درب باشد. بنابراین هر خط عمود با هر یک از خطوط دیگر و انتهای راست و چپ تصویر تشکیل یک ناحیه می‌دهد. برای مثال، مطابق رابطه‌ی (۱) جهت تعیین تعداد ناحیه‌ها، برای دو خط عمود در شکل ۴ تعداد ۵ ناحیه ایجاد می‌گردد.

$$(1) \quad n + \frac{n \times (n-1)}{2} + n = \text{تعداد ناحیه‌ها}$$

که n تعداد خطوط عمودی است.

۲-۴- تعیین ویژگی‌های هر ناحیه

در این بخش، ویژگی‌های هر ناحیه مطابق جدول ۱ استخراج می‌گردند.



شکل ۴: پنج ناحیه ایجاد شده برای دو خط عمودی

جدول ۱: ویژگی‌های هر ناحیه

مختصات نقطه‌ی ابتدایی و انتهایی خط عمودی راست و چپ
رنگ: میانگین مقادیر HSV: $C(1:3)$ هیستوگرام hue در ۵ قسمت: $C(4:8)$ هیستوگرام saturation در ۳ قسمت: $C(9:11)$
بافت: (با استفاده از بانک فیلتر لیانگ و مالیک ^{۱۹} (LM)) میانگین پاسخ فیلتر برای هر یک از ۱۵ فیلتر: $T(1:15)$ هیستوگرام پاسخ بیشینه در ۱۵ قسمت: $T(16:30)$
تعداد خطوط عمودی و افقی هر ناحیه

۲-۴-۱- مختصات خطوط عمودی راست و چپ ناحیه

برای هر ناحیه، مختصات نقطه‌ی ابتدایی و انتهایی هر یک از دو خط عمودی راست و چپ در نظر گرفته می‌شود.

۲-۴-۲- رنگ

معمولاً رنگ درب با رنگ اطراف درب مثل دیوار متفاوت می‌باشد. در این‌جا از فضای رنگی HSV^۶ استفاده می‌شود زیرا بر خلاف فضای رنگی RGB^{۱۷} سطح خاکستری تصویر و اطلاعات رنگی را جدا می‌کند و به علت توانایی آن در قبال تغییرات نور، در کاربردهای بینایی کامپیوتر بسیار مفید است. بنابراین برای هر ناحیه، میانگین مقادیر HSV، هیستوگرام hue در پنج قسمت^{۱۸} و هیستوگرام saturation در سه قسمت مشخص می‌گردد.

۲-۴-۳- بافت

بافت درب هم مانند رنگ معمولاً با بافت اطراف درب متفاوت است. در این‌جا از قسمتی از بانک فیلتر طراحی شده به وسیله‌ی لیانگ و مالیک^{۱۹} (LM) استفاده می‌شود [۲۶]. برای کاهش زمان پردازش ضمن افزایش دقت تشخیص درب، فقط تعدادی از فیلترها با پارامترهای زیر انتخاب می‌گردند: ساختار ۱۹×۱۹ پیکسلی با مقیاس $\sqrt{2}$ و ۶ جهت. در کل ۱۵ فیلتر شامل فیلترهای با ۶ لبه، ۶ میله، ۱۲ گاسین و ۲ لاپلاسیان فیلتر گاسین^۱ وجود دارد. سپس برای هر ناحیه، میانگین پاسخ فیلتر برای هر یک از ۱۵ فیلتر و هیستوگرام پاسخ‌های بیشینه در ۱۵ قسمت تعیین می‌شوند.

۲-۴-۴- تعداد خطوط عمودی و افقی هر ناحیه

تعداد خطوط عمودی و افقی در داخل هر ناحیه شمارش شده است زیرا در اکثر اوقات تعداد خطوط عمودی و افقی در ناحیه درب بیشتر از ناحیه‌های شامل انسان، درخت و یا ماشین است. همچنین اگر تعداد این خطوط برای ناحیه‌ای

^۶Hue Saturation Value (HSV)

^{۱۷}Red Green Blue (RGB)

^{۱۸}Bin

^{۱۹}Leung-Malik(LM)

^۱Bar

^۲Laplacian of Gaussian (LoG)

نزدیک به صفر باشد این ناحیه نمی‌تواند کاندید درب باشد هر چند اختلاف رنگ و بافت این ناحیه با ناحیه‌های همسایه راست و چپ زیاد باشد.

۵-۲- تصمیم‌گیری برای وجود درب

در این بخش، با استفاده از ناحیه‌ها و ویژگی‌های آن‌ها، ناحیه درب مشخص می‌گردد. در الگوریتم ۱، ابتدا عرض و ارتفاع هر ناحیه و توسط تابعی مجموع اختلاف رنگ و بافت هر ناحیه از ناحیه‌های مجاور محاسبه می‌شود (بخش ۲-۵-۱). همچنین در بخش ۲-۵-۲ درصد اشغال ناحیه توسط حداکثر رنگ و بافت محاسبه می‌گردد. سپس ناحیه‌های غیر معقول حذف می‌شوند و در نهایت، ناحیه‌ای که درصد یکنواختی و اختلاف رنگ و بافت آن بیشتر است به عنوان درب در نظر گرفته می‌شود.

تابع ۱: محاسبه‌ی مجموع اختلاف رنگ و بافت ورودی: ویژگی‌های رنگ و بافت (جدول ۱) خروجی: مجموع اختلاف رنگ و بافت هر ناحیه از ناحیه‌های راست و چپ شروع
گام ۱: اختلاف رنگ و بافت ناحیه از ناحیه راست را محاسبه کن:
(CR, TR): ویژگی‌های رنگی و بافتی ناحیه راست: (جدول ۱)
 $diffR(1:3) = abs(C(1:3)-CR(1:3));$
 $diffR(4) = JSD(C(4:8), CR(4:8));$
 $diffR(5) = JSD(C(9:11), CR(9:11));$
 $diffR(6:20) = abs(T(1:15)-TR(1:15));$
 $diffR(21) = JSD(T(16:30), TR(16:30));$
گام ۲: اختلاف رنگ و بافت ناحیه از ناحیه چپ را محاسبه کن:
(CL, TL): ویژگی‌های رنگی و بافتی ناحیه چپ: (جدول ۱)
 $diffL(1:3) = abs(C(1:3)-CL(1:3));$
 $diffL(4) = JSD(C(4:8), CL(4:8));$
 $diffL(5) = JSD(C(9:11), CL(9:11));$
 $diffL(6:20) = abs(T(1:15)-TL(1:15));$
 $diffL(21) = JSD(T(16:30), TL(16:30));$
گام ۳: مجموع این دو اختلاف را برگردان.
 $return sum(diffR(1:21)) + sum(diffL(1:21));$
پایان

تابع ۱: محاسبه‌ی مجموع اختلاف رنگ و بافت

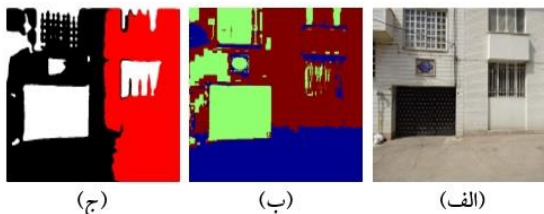
۵-۲-۱- مجموع اختلاف رنگ و بافت از ناحیه‌های مجاور

در تابع ۱، مجموع اختلاف رنگ و بافت هر ناحیه از ناحیه‌های راست و چپ محاسبه می‌گردد. از ویژگی‌های رنگ و بافت در جدول ۱ استفاده می‌شود که ۱۱ ویژگی رنگی هر ناحیه در C و ۳۰ ویژگی بافتی در T قرار می‌گیرند. برای اندازه‌گیری شباهت بین دو توزیع احتمال و یا دو هیستوگرام P و Q از روش واگرایی جنسن-شانون (JSD) ^{۲۲} استفاده می‌شود که بر اساس روش کالک-

الگوریتم ۱: الگوریتم تشخیص درب ورودی: تصویر و ویژگی‌های هر ناحیه خروجی: محل درب یا پیغام «درب وجود ندارد». شروع
گام ۱: برای هر ناحیه مقادیر زیر را محاسبه کن:
۱-۱- عرض و ارتفاع.
۲-۱- مجموع اختلاف رنگ و بافت از ناحیه‌های راست و چپ (تابع ۱: بخش ۲-۵-۱)
۳-۱- درصد یکنواختی رنگ و بافت ناحیه (بخش ۲-۵-۲) گام ۲:
۱-۲- مقادیر آستانه را با توجه به تصویر و ویژگی‌های ناحیه‌ها مشخص کن.
۲-۲- ناحیه‌ها را حذف کن اگر (مجموع اختلاف رنگ و بافت کوچک‌تر از مقدار آستانه ۱ است) یا (درصد یکنواختی رنگ و بافت ناحیه کوچک‌تر از مقدار آستانه ۲ است) یا (عرض و ارتفاع ناحیه کوچک‌تر از مقدار آستانه ۳ است) یا (تعداد خطوط عمودی و افقی کوچک‌تر از مقدار آستانه ۴ است) گام ۳:
اگر (ناحیه‌ای وجود ندارد) پیغام «درب وجود ندارد» را چاپ کن.
در غیر این صورت ناحیه‌ای را که درصد یکنواختی و مجموع اختلاف رنگ و بافت آن، بیشینه است به عنوان درب در نظر بگیر.
پایان

^{۲۲}Jensen-Shannon Divergence (JSD)

- ۳.۲. نمونه‌ها به نزدیک‌ترین مراکز تخصیص می‌یابند.
- ۳.۳. مراکز k خوشه دوباره محاسبه می‌گردند.
- ۳.۴. اگر مراکز خوشه‌ها ثابت ماند k خوشه مشخص شده است و کار تمام است. در غیر این صورت به مرحله ۳.۲. برگرد و مراحل را دوباره تکرار کن.



شکل ۵: قطعه‌بندی تصویر بر اساس رنگ و بافت. (الف) تصویر، (ب) قطعه‌بندی به سه رنگ، (ج) قطعه‌بندی به سه بافت

۳. نتایج

این روش بر روی دو مجموعه تصاویر اجرا شده است. (۱) مجموعه تصاویر eTRIMS [۲۴] با ۶۰ تصویر از نماهای مختلف ساختمان، (۲) مجموعه تصاویر خود ما با ۳۰۰ تصویر با رزولوشن 640×480 پیکسل. از کد برنامه‌نویسی استخراج خطوط LSD به زبان C [۲۵] استفاده گردیده و الگوریتم به زبان C و MATLAB R2014a بر روی لپ‌تاپ با حافظه ۸ گیگابایت و پردازنده intel core i7 2.11GH پیاده‌سازی شده است. دقت تشخیص درب در مجموعه تصاویر eTRIMS، ۷۸/۳ درصد و در مجموعه تصاویر خود ما ۸۸/۴ درصد است. همچنین وابسته به تعداد خطوط استخراجی، زمان اجرا متفاوت می‌باشد که میانگین آن ۵/۶ ثانیه به دست آمده است.

در شکل ۶ خروجی‌های تشخیص درب نشان داده شده است. شکل (۶.الف، ب) مربوط به مجموعه تصاویر eTRIMS و بقیه از مجموعه تصاویر خودمان هستند. در شکل (۶.ج) با این‌که فقط ناحیه بالای درب آشکار است اما الگوریتم، تشخیص درستی دارد. همچنین این الگوریتم در تصاویر با چند درب در شکل (۶.د، ه) فقط یک درب آن را تشخیص می‌دهد و در بعضی از تصاویر مانند شکل (۶.ه) که به دلیل دوری تصویر یا هم‌رنگی درب با محیط اطراف،

لیبل^۲(KL) است و به صورت رابطه‌ی (۲) محاسبه می‌گردد:

$$(2) \quad JSD(P, Q) = \sqrt{\frac{1}{2}KL(P, M) + \frac{1}{2}KL(Q, M)}$$

$$M = \frac{1}{2}(P + Q), \quad KL(P, Q) = \sum P \log \frac{P}{Q}$$

۲-۵-۲- درصد اشغال ناحیه با حداکثر رنگ و بافت

ابتدا تصویر با استفاده از فضای رنگی $L^*a^*b^*$ و الگوریتم خوشه‌بندی k -means به سه رنگ قطعه‌بندی می‌شود (شکل ۵.ب). عدد ۳ که با آزمایش انتخاب شده است باعث می‌شود که رنگ‌های مشابه در یک دسته قرار گیرند. چون در واقعیت هم اگر رنگ درب مختلف باشد معمولاً ترکیبی از رنگ‌های مشابه می‌باشد پس مثلاً اگر رنگ درب ترکیبی از رنگ‌های سفید، خاکستری، شیری و قهوه‌ای روشن باشد بهتر است کل درب را به عنوان یک قطعه در نظر بگیرد. سپس مشخص شده است رنگ و بافتی که حداکثر فضای ناحیه را پوشش داده‌اند چند درصد کل ناحیه را شامل می‌شوند. برای نمونه، اگر ناحیه‌ای با ۸۰ درصد رنگ اول، ۱۵ درصد رنگ دوم و ۵ درصد رنگ سوم پوشیده شده باشد ۸۰ درصد در نظر گرفته شده است که این یعنی ۸۰ درصد ناحیه توسط یک رنگ اشغال شده است. از آن‌جا که در اکثر تصاویر، درب‌ها دارای رنگ یکسان هستند پس می‌توان گفت اگر این پارامتر مقدار کوچکی باشد یعنی در این ناحیه، پراکندگی رنگی زیاد است و نمی‌توان به عنوان درب در نظر گرفت مانند یک ناحیه بزرگ که شامل درخت، دیوار و ماشین باشد.

همچنین تصویر به سه ناحیه بافتی مانند شکل (۵.ج) قطعه‌بندی می‌شود. سیستم قطعه‌بندی بافتی در سه مرحله‌ی زیر انجام می‌گیرد [۲۷]:

۱. تجزیه تصویر ورودی با استفاده از بانک فیلتر گابور^۳
 ۲. استخراج ویژگی
 ۳. خوشه‌بندی k -means
- ۳.۱. مراکز k خوشه به صورت تصادفی انتخاب می‌شوند.

^۲Kullback-Leibler (KL)

^۳Gabor

^۲Resolution

خطوط بلند به درستی استخراج نمی‌شوند فقط قسمتی از درب مشخص می‌شود.

به هر حال واضح است که چون الگوریتم ما فقط از خطوط، رنگ و بافت استفاده می‌کند در بعضی تصاویر مانند شکل ۷ دچار اشتباه می‌شود. در شکل (الف.۷) از مجموعه تصاویر eTRIMS با این که یک درب در سمت چپ تصویر وجود دارد اما به دلیل هم‌رنگی بسیار زیاد آن با ساختمان، خطوط آن استخراج نشده است و برنامه به اشتباه پیغام «درب وجود ندارد.» را نشان می‌دهد که البته تشخیص درب در این تصویر با چشم خود ما هم کمی مشکل به نظر می‌آید. همچنین در بعضی تصاویر دور و یا شلوغ با وجود ماشین، انسان و درخت که خطوط به درستی استخراج نمی‌شوند نیز الگوریتم دچار اشتباه می‌شود مانند شکل (ب.۷) که قسمت بزرگی از درب به خاطر وجود ماشین مشخص نیست.



(د)



(ه)

شکل ۶: تشخیص درب در محیط‌های بیرونی. تصاویر (الف)، (ب) از مجموعه تصاویر eTRIMS و تصاویر



(الف)

درب وجود ندارد.



(ب)

شکل ۷: دو تصویری که الگوریتم ما در تشخیص درب اشتباه کرده است.



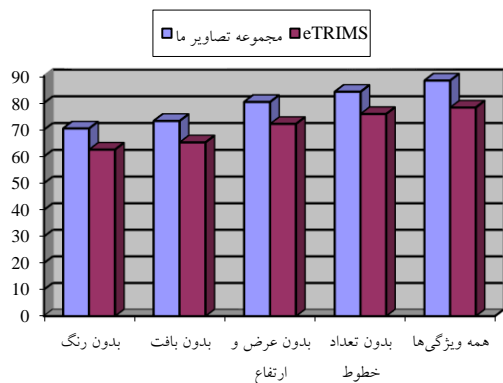
(الف)



(ب)



(ج)



شکل ۸: دقت تشخیص درب برای مجموعه ویژگی‌های مختلف

۴. نتیجه‌گیری و کارهای آینده

در این مقاله، با استفاده از خطوط، رنگ، بافت و تشکیل ناحیه بین خطوط عمودی، روشی برای تشخیص درب در محیط‌های بیرونی ارائه شده است. چون در تصاویر، مشخصات ظاهری درب مانند دستگیره و یا گوشه‌ها آشکار نبوده است از محدودیت‌های زمینه‌ای یعنی دانش محیط‌های ساخته‌ی بشری بهره برده شده است. به عنوان مثال در مجموعه تصاویر، درب‌ها در پایین تصویر وجود دارند یا در ناحیه‌های درب، تجمع خطوط بیشتر از ناحیه‌های دیگر شامل درخت، ماشین و یا انسان است. الگوریتم پیشنهادی بر روی مجموعه تصاویر خودمان و eTRIMS آزمایش شده است و با این‌که مجموعه تصاویر ما شامل انواع گوناگون درب‌ها است ولی به علت استخراج بسیار دقیق خطوط با الگوریتم LSD و استفاده از دانش موجود، تشخیص درب قابل قبولی داشته است.

در آینده، می‌خواهیم با استفاده از قطعه‌بندی تصویر به زمین، آسمان و سطوح عمودی [۲۸] و تشکیل مجموعه تصاویر کامل‌تری از انواع ساختمان‌ها با استفاده از یادگیری عمیق^{۲۹} و شبکه‌ی عصبی کانولوشنال^{۳۰} [۲۹-۳۱] به تشخیص دقیق ساختمان و درب پردازیم.

چون یک مجموعه تصاویر استاندارد، مخصوص درب‌های ورودی در محیط‌های پیچیده‌ی بیرونی وجود ندارد برای مقایسه‌ی عادلانه، روش‌های پیشین روی مجموعه تصاویر خودمان پیاده‌سازی و اجرا شده است. نتایج در جدول ۲ نشان داده می‌شود. در الگوریتم پیشنهادی ما، دقت تشخیص درب ۸۸/۴ درصد به دست آمده است که نسبت به روش لیو و همکاران^{۲۶} [۷] ۵/۱ درصد و در مقایسه با روش کانگ و همکاران^{۲۷} [۸] ۸/۸ درصد برتری داشته است. در الگوریتم ۱ برای تشخیص درب، از چهار ویژگی رنگ، بافت، عرض و ارتفاع، و تعداد خطوط عمودی و افقی استفاده شده است. در شکل ۸، ابتدا اثرهای همه‌ی ویژگی‌ها بدون در نظر گرفتن یک ویژگی و در انتها همه ویژگی‌ها برای تشخیص درب در مجموعه تصاویر خودمان و eTRIMS بررسی شده است. شکل ۸ نشان می‌دهد که در تشخیص درب، ویژگی‌های رنگ و بافت بسیار مؤثر هستند اما ویژگی‌های عرض و ارتفاع و تعداد خطوط تأثیر کمتری دارند. همچنین رنگ، نقش مؤثرتری نسبت به بافت برای تشخیص درب بازی می‌کند.

جدول ۲: مقایسه‌ی روش پیشنهادی با سایر روش‌ها برای تشخیص درب در مجموعه تصاویر خودمان

روش‌ها	دقت تشخیص درب
لیو و همکاران ^{۲۶} [۷]	۸۳/۳
کانگ و همکاران ^{۲۷} [۸]	۷۹/۶
الگوریتم پیشنهادی	۸۸/۴

^۲Context

^۲Deep learning

^۳Convolutional neural network (CNN)

^{۲۶}Liu et al.

^{۲۷}Kang et al.

Conf. on Computer Vision and Pattern Recognition (CVPR), USA, pp. 1640-1647, 2012.

11. M. Mathias, A. Martinovic, and L. Van Gool, "ATLAS: A Three-Layered Approach to Facade Parsing," *Int. Journal of Computer Vision (IJCV)*, Vol. 118, No. 1, pp. 22-48, 2016.

12. A. Cohen, A.G. Schwing, and M. Pollefeys, "Efficient Structured Parsing of Facades Using Dynamic Programming," *Proc. Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, USA, pp. 3206-3213, 2014.

13. R. Gadde, R. Marlet, and N. Paragios, "Learning Grammars for Architecture-Specific Façade Parsing," *Int. Journal of Computer Vision (IJCV)*, Vol. 117, No. 3, pp. 290-316, 2016.

14. C. Zhou and C. Liu, "Semantic Image Segmentation using Low-Level Features and Contextual Cues," *Computers and Electrical Engineering*, Vol. 40, pp. 844-857, 2014.

15. S. Gould and X. He, "Scene Understanding by Labeling Pixels," *Communications of the ACM*, Vol. 57, No. 11, pp. 68-77, 2014.

16. J. Xiao, T. Fang, P. Zhao, M. Lhuillier, and L. Quan, "Image-based Street-side City Modeling," *ACM Transactions on Graphics*, Vol. 28, No. 5, 2009.

17. B. Shuai, Z. Zuo, G. Wang, and B. Wang, "Scene Parsing with Integration of Parametric and Non-parametric Models," *IEEE Trans. on Image Processing*, Vol. 25, No. 5, pp. 2379-2391, 2016.

18. P.F. Felzenszwalb and D.P. Huttenlocher, "Efficient Graph-Based Image Segmentation," *Int. Journal of Computer Vision (IJCV)*, Vol. 59, No. 2, pp. 167-181, 2004.

19. J. Tighe and S. Lazebnik, "Superparsing: Scalable Nonparametric Image Parsing with Superpixels," *Int. Journal of Computer Vision (IJCV)*, Vol. 101, No. 2, pp. 329-349, 2013.

20. C. Liu, J. Yuen, and A. Torralba, "Nonparametric Scene Parsing via Label Transfer," *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 33, No. 12, pp. 2368-2382, 2011.

21. P. Razzaghi and S. Samavi, "A New Fast Approach to Nonparametric Scene Parsing," *Pattern Recognition Letters*, Vol. 42, pp. 56-64, 2014.

۵. منابع

1. R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer, London, 2011.

2. D. Anguelov, D. Koller, E. Parker, and S. Thrun, "Detecting and Modeling Doors with Mobile Robots," *Proc. Int. Conf. on Robotics and Automation, USA*, pp. 3777-3784, 2004.

3. Z. Chen, Y. Li, and S.T. Birchfield, "Visual Detection of Lintel-Occcluded Doors by Integrating Multiple Cues Using Data-Driven Markov Chain Monte Carlo Process," *Robotics and Autonomous Systems*, Vol. 59, No. 11, pp. 966-976, 2011.

4. J. Hensler, M. Blaich, and O. Bittel, "Real-Time Door Detection Based on Adaboost Learning Algorithm," *Proc. Int. Conf. on Research and Education in Robotics, France*, pp. 61-73, 2009.

5. A.C. Murillo, J. Kosecka, J.J. Guerrero, and C. Sagues, "Visual Door Detection Integrating Appearance and Shape Cues," *Robotics and Autonomous Systems*, Vol. 56, No. 6, pp. 512-521, 2008.

6. R. Sekkal, F. Pateau, M. Babel, B. Brun, and I. Leplumey, "Simple Monocular Door Detection and Tracking," *Proc. Int. Conf. on Image Processing, Australia*, pp. 3929-3933, 2013.

7. J. Liu, T. Korah, V. Hedau, V. Parameswaran, R. Grzeszczuk, and Y. Liu, "Entrance Detection from Street-View Images," *Proc. Int. Conf. on Computer Vision and Pattern Recognition Workshop (CVPR)*, USA, 2014.

8. S.J. Kang, H.H. Trinh, D.N. Kim, and K.H. Jo, "Entrance Detection of Buildings Using Multiple Cues," *Proc. Int. Conf. on Intelligent Information and Database Systems, Vietnam*, pp. 251-260, 2010.

9. O. Teboul, I. Kokkinos, L. Simon, P. Koutsourakis, and N. Paragios, "ShapeGrammar Parsing via Reinforcement Learning," *Proc. Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, USA, pp. 2273-2280, 2011.

10. H. Riemenschneider, U. Krispel, W. Thaller, M. Donoser, S. Havemann, D. Fellner, and H. Bischof, "Irregular Lattices for Complex Shape Grammar Façade Parsing," *Proc. Int.*

No. 12, pp. 1167-1186, 1991. Code available:

28. A.K. Jain and F. Farrokhnia, "Unsupervised Texture Segmentation using Gabor Filters," *Pattern Recognition*, Vol. 24, No. 12, pp. 1167-1186, 1991. Code available: <http://note.sonots.com/SciSoftware.html>

29. D. Hoiem, A.A. Efros, and M. Hebert, "Recovering Surface Layout from an Image," *Int. Journal of Computer Vision (IJCV)*, Vol. 75, No. 1, pp. 151-172, 2007.

29. S. Bu, P. Han, Z. Liu, and J. Han, "Scene Parsing using Inference Embedded Deep Networks," *Pattern Recognition*, Vol. 59, pp. 188-198, 2016.

30. Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep Learning for Visual Understanding: A Review," *Neurocomputing*, Vol. 187, pp. 27-48, 2016.

31. E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 39, No. 4, pp. 640-651, 2017.

22. M. Najafi, S. Taghavi Namin, M. Salzmann, and L. Petersson, "Sample and Filter: Nonparametric Scene Parsing via Efficient Filtering," *Proc. Int. Conf. on Computer Vision and Pattern Recognition (CVPR), USA*, pp. 607-615, 2016.

23. R.G. Von Gioi, J. Jakubowicz, J.M. Morel, and G. Randall, "LSD: A Fast Line Segment Detector with a False Detection Control," *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 32, No. 4, pp. 722-732, 2010.

24. F. Korc and W. Forstner, "eTRIMS Image Database for Interpreting Images of Man-Made Scenes," *Technical Report, University of Bonn*, 2009.

25. R.G. Von Gioi, J. Jakubowicz, J.M. Morel, and G. Randall, "LSD: a Line Segment Detector," *Image Processing On Line*, Vol. 2, pp. 35-55, 2012.

26. T. Leung and J. Malik, "Representing and Recognizing the Visual Appearance of Materials using Three-dimensional Textons," *Int. Journal of Computer Vision (IJCV)*, Vol. 43, No. 1, pp. 29-44, 2001.

27. A.K. Jain and F. Farrokhnia, "Unsupervised Texture Segmentation using Gabor Filters," *Pattern Recognition*, Vol. 24,

