دانشگاه خوارزمی تهران

دانشکده فنی و مهندسی

پایان نامه کارشناسی ارشد

گرایش هوش مصنوعی

عنوان:

بازشناسی اشیاء در تصاویر دیجیتال با استفاده از روشهای مبتنی بر متن

نگارش:

سهیلا شیخ بهائی

استاد راهنما:

دکتر جمشید شنبه زاده

استاد مشاور:

دکتر زینب قصابی

بهمن 1393

بسمه الله الرحمن الرحيم

به نام خدا

دانشگاه خوارزمی تهران

دانشکده فنی و مهندسی

رساله کارشناسی ارشد

عنوان: بازشناسی شئ با استفاده از روش های مبتنی بر متن

نگارش: سهیلا شیخ بهائی

کمیته ممتحنین:

استاد راهنما: دکتر جمشید شنبه زاده امضاء: .....

استاد مشاور: دکتر زینب قصابی امضاء: .....

استاد مدعو: امضاء: .....

تاریخ: .....

تقدیم به پدر و مادر عزیز و بزرگوارم

که زحمات بیشمار آنها را هیچگاه نمی توانم جبران نمایم.

و تقدیم به همسر گرانقدرم

که همواره مشوق و همراه و پشتیبان من بوده است.

چکیده:

در دنیای واقعی رابطه‌ای قوی بین محیط و اشیاء وجود دارد که استفاده از اطلاعات مربوط به روابط اشیاء و صحنه نقش مهمی در فرآیند بازیابی شئ ایفا می کنند. یعنی زمانی که در یک صحنه شئ خاصی جستجو می‌شود، یک بیننده در موقعیت‌هایی از تصویر تمرکز می‌کند که بیشترین احتمال پیشین را برای وجود شئ مورد نظر دارا هستند. بنابراین متن صحنه برای تصمیم گیری درمورد حرکات بعدی چشم بسیار تعیین کننده است و زمانیکه اطلاعات ظاهری شئ به دلیل کوچک بودن شئ، شلوغی صحنه، وجود مانع و یا هر نوع ابهام دیگری برای بازشناسی آن کافی نباشد، می توان از اطلاعات متن تصویر مانند نوع صحنه و اشیاء دیگر صحنه برای بازشناسی شئ کمک گرفت.

به منظور توسعه روشی دقیق برای بازشناسی شئ در تصاویر واقعی و با وجود اشیاء مختلف در این پایان نامه راه حل هایی ارائه شده است که بازیابی شئ را با استفاده از اطلاعات متنی یعنی اشیاء دیگر تصویر و صحنه انجام می‌دهد و هدف بهبود دقت بازیابی شئ است. در این روش اطلاعات متنی به صورت ماتریس همرخداد مدل می شوند و هر تصویر توسط برداری که نشان دهنده ی اشیاء دیگر تصویر و فرکانس آنها است نشان داده می‌شود از این بردارها برای آموزش کلاسبند استفاده می شود. به دلیل زیاد بودن تعداد داده های آموزشی و تست و زمانبر بودن استفاده از روش تطبیق و یا نمایه سازی از کلاسبند درخت تصمیم استفاده شده است. پس از آموزش کلاسبند، بردارهای تصاویر تست برای تعیین حضور شئ موردنظر بررسی می شوند و دقت بازیابی آنها تعیین می شود.
نکته مهم در روش هایی که از اطلاعات متنی استفاده می کنند این است که این مدل ها باید بر تصاویری تست شوند که شامل اشیاء مختلف و از صحنه های متنوع باشند تا بتوان از روابط اشیاء و صحنه کمک گرفت. مجموعه داده سان این ویژگی را فراهم می کند و دارای تصاویر بیش از 600 صحنه و بیش از 200 شئ است. تعداد اشیاء در هر تصویر بیش از ده کلاس شئ است درحالیکه در مجموعه داده های دیگر تعداد اشیاء یک تصویر حداکثر 5 کلاس شئ است. در این روش از مجموعه داده سان09 استفاده شده است و میانگین دقت بیش‌بینی حضور شئ با روش های مشابه مقایسه شده است و نتایج نشان می دهد که متوسط میانگین دقت برای تمام اشیاء در این روش بهبود یافته است.

واژه‌های کلیدی: بازشناسی شئ، اطلاعات متنی، بازیابی شئ، درک صحنه

با تقدیر وتشکر از استاد ارجمندم

جناب آقای دکتر جمشید شنبه زاده

جهت تشویق ها و راهنمایی های بی دریغ ایشان

با تقدیر و تشکر از خانم دکتر زینب قصابی

و خانم مهندس زهرا صادقی

که بسیار مرا راهنمایی و یاری نمودند.

فهرست مطالب

فهرست شکل‌ها

فهرست جدول ها

فهرست علائم و اختصارات

SVM	Support Vector Machine
HOG	Histogram of Gradient
BOW	Bag of Words
DPM	Discriminatively Trained Part Based Model
SIFT	Scale Invariant Feature Transform
DOG	Difference of Gaussians

فصل۱:

مقدمه و طرح تحقیق

۱–۱: مقدمه

امروزه علاقه و توجه به سیستم های بینایی بدون توقف ادامه دارد و دلیل آن پتانسیل سیستم های بینایی اتوماتیک در افزایش شدید ظرفیت تولید سازمان می باشد. عمدتا اساسی ترین بخش یک عامل برمبنای بینایی واحد شناسایی شئ آن است. تحقیقات در زمینه ی الگوریتم های بازشناسی شئ باعث پیشرفت‌هایی در خودکارسازی بسیاری از اعمال با استفاده از ایجاد سیستم های بازشناسی حروف، سیستم های بازرسی صنعتی خط تولید و همچنین سیستم های شناسایی خطا شده است. همچنین باعث پیشرفت های بزرگی در تصویربرداری پزشکی، دفاع و بیومتریک ها شده است[1].

پیشرفت‌های اخیر در بازشناسی شئ و درک صحنه باعث افزایش قدرت برنامه های کاربردی اصلاح تصویر نیمه خودکار شده است. از دیگر سیستم های کاربردی در این زمینه می توان به سیستم تکمیل صحنه اشاره کرد که در آن هدف حذف بخشی از صحنه و پرکردن آن با استفاده از بخش هایی از تصاویر دیگر است[2]. بازشناسی شئ در موارد مختلفی از جمله سیستم دستیار راننده، سیستم های کنترل خط تولید، روباتیک و پردازش تصاویر پزشکی نیز کاربرد دارد.

توانایی انسان برای بازشناسی هزاران دسته شئ در صحنه‌های شلوغ با وجود تغییرات در حالت و روشنایی و یا وجود مانع یکی از حیرت‌انگیزترین قابلیت‌های ادراک بصریست که هنوز توسط الگوریتم‌های بینایی کامپیوتر قابل دستیابی نیست. هدف بهبود فرآیند بازشناسی شئ با وجود کاربردهای مختلف و متنوع آن در زمینه‌های گوناگون محققین را واداشته است تا در این زمینه به تحقیقات گسترده‌ای بپردازند.

در بخش دوم این فصل سیستم بازشناسی شئ تعریف می شود و مراحل عملکرد آن معرفی می‌شود. در بخش سوم اطلاعات متنی را تعریف می کنیم و دلیل استفاده از آن و اهمیت استفاده از آن را توضیح می دهیم. همچنین انواع اطلاعات متنی و سطوح متنی را تعریف می کنیم و با ذکر مثال آن‌ها را توضیح می دهیم و شاخه ای که در این پایان نامه مدنظر است مشخص می شود.

در بخش چهارم این فصل درمورد جایگاه و اهمیت این موضوع توضیح داده می شود و اهمیت تحقیق و چالش‌ها مشخص می‌شود. بخش پنجم شامل هدف و سوال اصلی تحقیق می‌باشد. بخش ششم نیز روش تحقیق و گردآوری اطلاعات و حوزه مساله و جنبه نوآوری تحقیق را بیان می کند. در نهایت نیز در بخش هفتم این فصل ساختار رساله و اینکه هر فصل شامل چه مباحثی می باشد بیان شده است.

۱–۲:سیستم های بازشناسی شئ

در این بخش سیستم بازشناسی شئ تعریف می شود و چرخه کلی آن توضیح داده می شود و سپس بخش‌های مختلف آن معرفی و مراحل عملکرد آن ذکر می شود. در مرحله‌ی عملکرد یک سیستم بازشناسی شئ، دو مرحله کلی استخراج اطلاعات از تصویر و تعیین استراتژی بازشناسی شئ مبنی بر این اطلاعات بیان می شود. در فصل‌های آینده این مراحل با جزئیات بیشتری توضیح داده خواهند شد.

۱–۲–۱: تعریف سیستم های بازشناسی شئ

مساله بازشناسی شئ را می توان به عنوان یک مساله برچسب گذاری تعریف کرد. به این صورت که مدل هایی از اشیائی که باید بازشناسی شوند، درسیستم تعریف می شوند و سپس براساس این مدلها تصویر ورودی به سیستم، برچسب گذاری میشود. در این سیستم به دو سوال باید پاسخ داده شود:

- ✓ آیا شئ موردنظر در تصویر وجود دارد یا خیر؟
- ✓ در صورتی که وجود دارد موقعیت آن کجاست؟

درنتیجه ورودی تصویری است که شامل صفر یا تعداد بیشتری شئ موردنظر است. یک سری برچسب که متعلق به مجموعه ای از مدلهاست به سیستم شناسانده شده است، حال سیستم باید برچسب های درست را به نواحی درست تصویر تخصیص دهد. شکل 1-1 نمونه ای از یک تصویر و برچسب‌های اختصاص داده شده توسط سیستم بازشناسی شئ را برای دو شئ ماشین و شخص نشان می دهد.

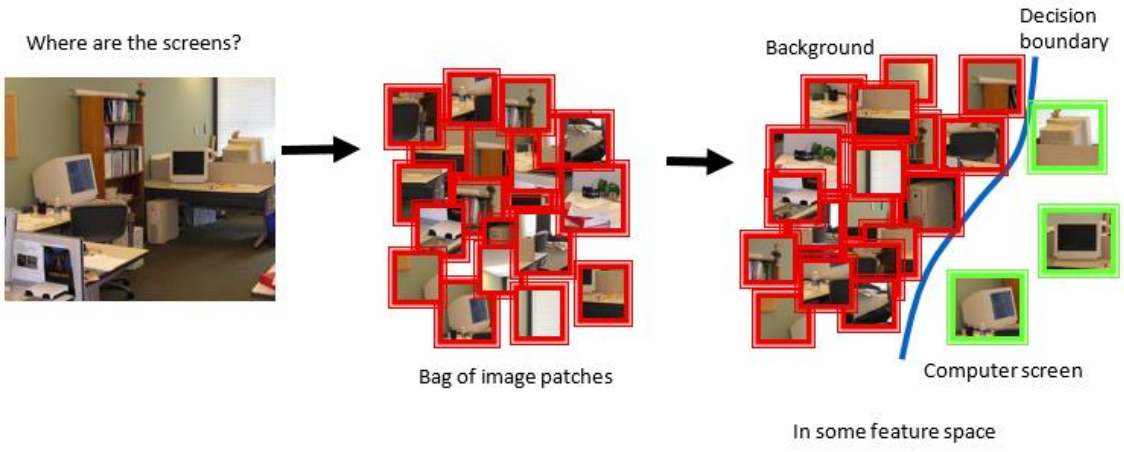
برای این کار ابتدا تصاویری که دارای برچسب هستند برای آموزش وارد سیستم می شوند. پس از مرحله آموزش تصاویری بدون برچسب وارد سیستم می شوند و پاسخ بدست آمده با پاسخ صحیح مقایسه می شود تا دقت سیستم بدست آید.

شکل 1-1 - نمونه ای از یک تصویر و برچسب های اختصاص داده شده توسط سیستم بازشناسی شئ برای دو شئ ماشین و شخص

۱-۲-۲: مراحل عملکرد سیستم بازشناسی شئ

چهارچوب کلی یک سیستم بازشناسی شئ شامل دو بخش است: استخراج اطلاعات از تصویر و دوم استفاده از این اطلاعات برای بازشناسی شئ براساس استراتژی مورداستفاده. درنتیجه در بازشناسی شئ دو مساله مهم وجود دارد: اول چگونگی بازنمایی یک طبقه شئ، و دوم چگونگی یادگیری سیستم و بازشناسی شئ موردنظر که عمدتا شامل ساخت یک کلاسبند با استفاده از داده های آموزشی و استفاده از این کلاسبند برای کلاسبندی داده های جدید است.

می توان یک مساله بازشناسی شئ را به عنوان یک مساله کلاسبندی درنظر گرفت. تصویر به پنجره-هایی که بایکدیگر همپوشانی دارند تقسیم می شود و سپس تصمیم گیری می شود که آیا پنجره موردنظر دارای شئ هدف هست یا خیر. شکل 2-1 این مراحل را نشان میدهد.



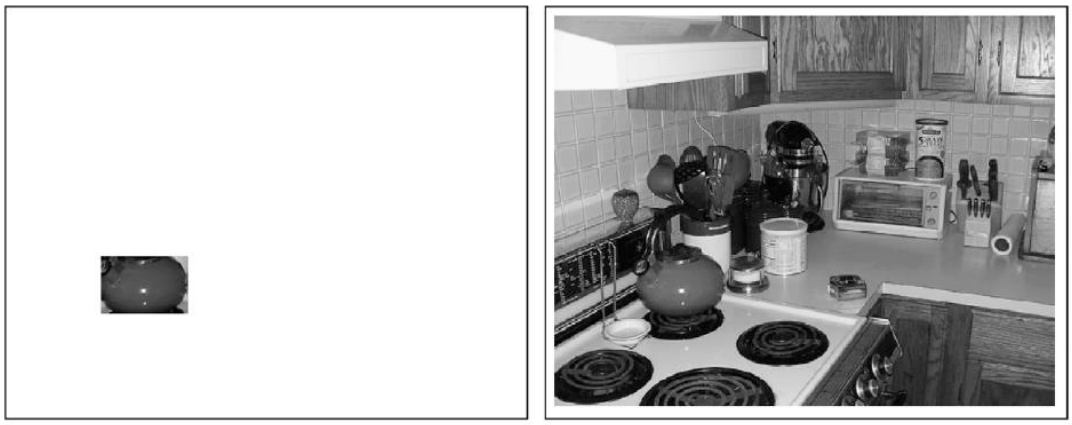
شکل 2-1- مراحل کلاسبندی در سیستم بازشناسی شئ مبتنی بر کلاسبند

روش‌های مختلفی جهت استخراج اطلاعات از تصویر برای فرآیند بازشناسی شئ وجود دارد. می‌توان این روش‌ها را به دو دسته تقسیم کرد: روش‌هایی که برمبنای اطلاعات شئ هستند و روش‌هایی که برمبنای اطلاعات صحنه و اشیاء دیگر تصویر هستند. اطلاعات مربوط به اشیاء دیگر و صحنه اطلاعات متنی[1] نامیده می‌شوند که در بخش بعد به تفصیل بررسی خواهند شد.

۱-۳: اطلاعات متنی

مساله‌ی اصلی در شناسایی شئ تطبیق نمایشی از شئ هدف با خصوصیات موجود تصویر و ردکردن خصوصیات زمینه است. در جستجوهای بصری معمولی متن صحنه مجموعه‌ای از مفهوم‌هایی رندم است که تنها فرآیند بازیابی شئ را سخت‌تر می‌کند، درحالی که در دنیای واقعی اشیاء دیگر در یک صحنه منبعی غنی از اطلاعات هستند که به جای مانع برای تشخیص و شناسایی اشیاء به این فرآیند کمک می‌کنند.

مدل های زیادی از اطلاعات متن اشیاء برای بهبود دقت بازشناسی استفاده می‌کنند. خصوصیات ظاهری مانند رنگ، لبه، بافت و علایم شکلی می توانند تا حدود مشخصی کلاس های اشیاء را تشخیص دهند. در صورت وجود شلوغی، نویز و تغییرات در حالت و روشنایی، ظاهر شئ دارای ابهام خواهد بود که برای رفع این ابهام می توان از پیوستگی اشیاء در یک صحنه استفاده کرد(شکل 3-1).



شکل 3-1 – مثالی از یک نمونه شئ که به تنهائی قابل شناسایی نیست اما با استفاده از اطلاعات متنی صحنه می‌توان شئ را بازشناسی کرد[3].

یکی از روش های متداول برای بازیابی و بازشناسی شئ استفاده از اطلاعات متنی است. استفاده از اطلاعات متنی به معنای استفاده از اطلاعات اشیاء دیگر تصویر و یا اطلاعات صحنه برای بازیابی شئی

[1] Contextual information

خاص است. معمولا در روش های بازشناسی و بازیابی شئ، اشیاء به تنهایی درنظر گرفته می شوند و صحنه ای که شئ در آن واقع است درنظر گرفته نمی شود. این یک محدودیت جدی است زیرا اطلاعات متنی نقش مهمی در عمل بازشناسی شئ در سیستم بینایی انسان ایفا می کنند. استفاده از اطلاعات متنی باعث بهبود عملکرد سیستم های بازشناسی شئ می شوند و همچنین راهنماهای معنایی مفیدی را برای درک صحنه فراهم می کنند.

خصوصیات متنی را می توان به سه دسته‌ی معنایی[۲]، مکانی[۳] و مقیاس[۴] دسته‌بندی کرد. مدل‌های بازشناسی شئ استفاده از اطلاعات متنی را از دو سطح سراسری و محلی تصویر درنظر گرفته اند. اطلاعات سراسری آماره‌های تصویر را در یک تصویر به عنوان یک کل درنظر می گیرد. اطلاعات محلی متنی اطلاعات متنی را برای ناحیه‌های همسایه شئ درنظر می گیرد.

۱-۳-۱: اطلاعات متنی معنایی

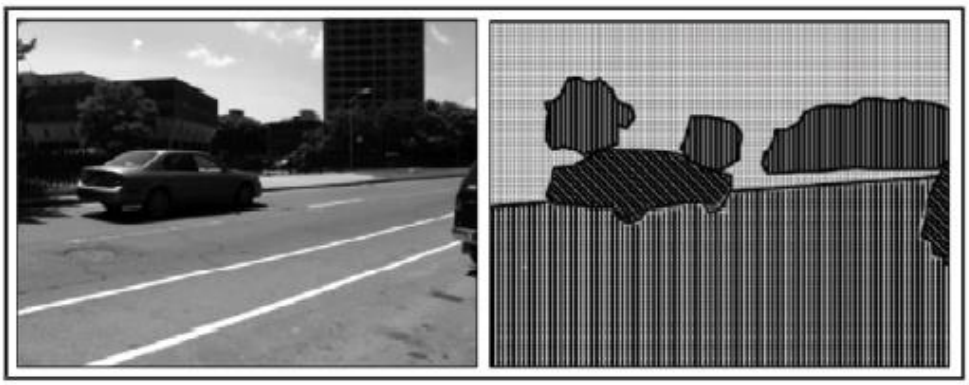
اطلاعات معنایی متن به احتمال وجود یک شئ در بعضی صحنه ها مربوط می شود. بنابراین می‌توان این اطلاعات را به صورت احتمال رخداد یک شئ با اشیاء دیگر و یا در یک صحنه خاص تعریف کرد.

مطالعات اخیر در روانشناسی و علوم شناختی[۵] نشان می دهد که اطلاعات معنائی متن به بازشناسی بصری در ادراک انسان کمک می کند. اطلاعات معنائی متن به صورت قوانین ازپیش تعیین شده تعریف می شوند. انتظار اولیه‌ی سیستم از جهان توسط فرضیه های مختلف (استراتژی های برمبنای قوانین) نمایش داده می شود. که وجود اشیاء دیگر در یک صحنه را پیش بینی می کند. این فرضیه ها توسط متخصصین بازشناسی انواع شئ تولید می شود.

در بعضی از روش های اخیر از روش های آماری برای عمومیت دادن و بهره برداری از اطلاعات معنائی متن استفاده شده است. همان طور که در شکل 4-1 نشان داده شده است، ولف[۶] و بیلسچی[7] از لایه های معنائی تصاویر آموزشی برای بدست آوردن اطلاعات معنائی متن استفاده کردند. لایه های معنائی وجود یک شئ مشخص را در یک تصویر نشان می دهند[3].

---

[2] Semantic context
[3] Spatial context
[4] Scale context
[5] Cognitive science
[6] Wolf
[7] Bileschi

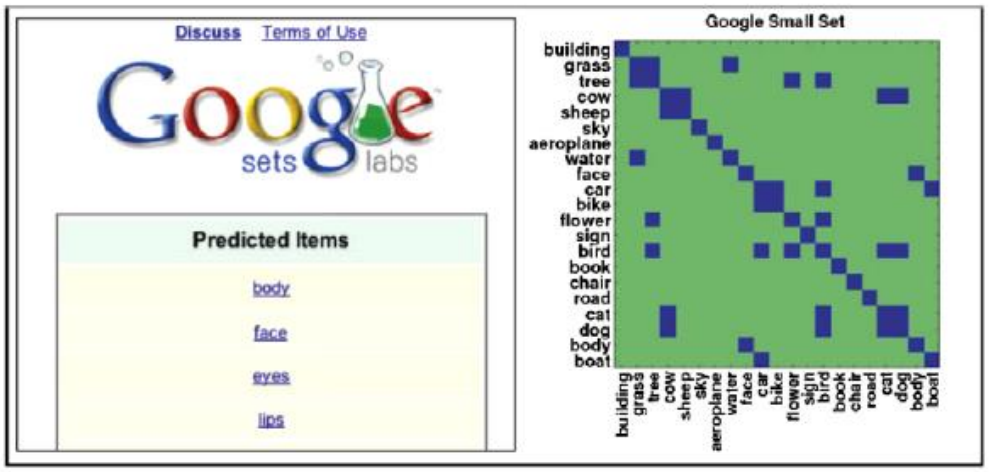


شکل 4-1- مثال تصاویر آموزشی و لایه های معنائی آن

هر تصویر چندین لایه معنائی را نشان می دهد. در یک لایه ی معنائی هر پیکسل با مقدار 1 یا صفر مقداردهی می شود که 1 نشان دهنده ی این است که این پیکسل متعلق به شئ هست و صفر نشان دهنده ی این است که پیکسل متعلق به شئ نیست . اطلاعات محتوائی متن به صورت لیستی از برچسب‌ها نمایش داده می شود که وقوع یک پیکسل در یک شئ مشخص را نشان می دهد.

اطلاعات متن را می توان از داده های آموزشی برچسب گذاری شده بدست آورد. روبینویچ و همکارانش[8] اطلاعات محتوائی متن را توسط پرسش از گوگل بدست آورد. گوگل لیستی از آیتم های به هم مرتبط را از تعداد کمی مثال بدست می آورد. این اطلاعات توسط ماتریس همرخداد[9] باینری نشان داده می شود . در صورتی که شئ 1 با شئ 2 رابطه داشته باشد، مولفه مربوطه در ماتریس 1 و درغیر این صورت صفر است. شکل 5-1 ماتریس مورد استفاده را نشان می دهد.

[8] Robinovich et al
[9] Co-occurrence



شکل 5-1 – برنامه تحت وب گوگل و ماتریس متن بدست آمده برای لیست آیتم های پیش بینی شده[3]

۱-۳-۲: اطلاعات متنی مکانی

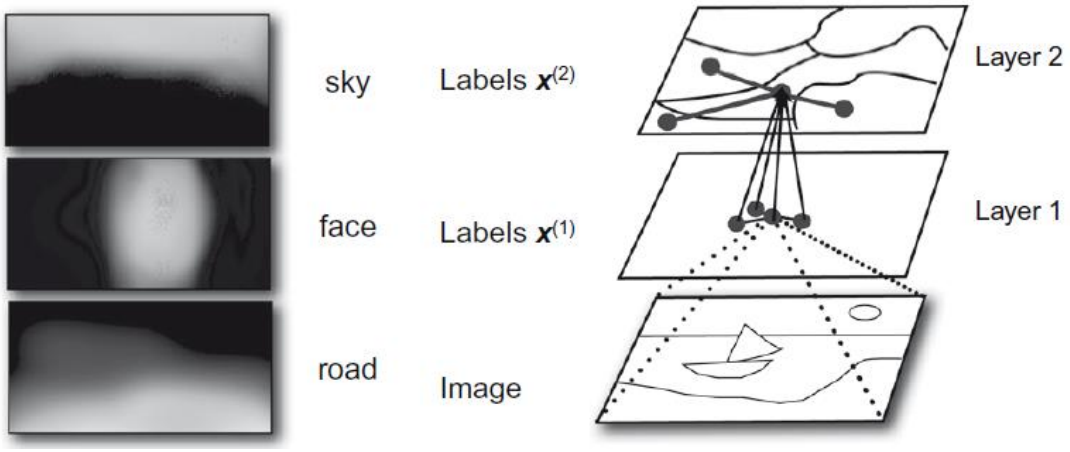
اطلاعات مکانی متن را می توان به صورت احتمال یافتن یک شئ در موقعیت های خاص با توجه به اشیاء دیگر در صحنه تعریف کرد. حضور اشیائی که تفسیر واحدی دارند بازشناسی اشیاء مبهم در صحنه را بهبود می دهد و روابط مکانی مناسب بین اشیاء نرخ خطا را در بازشناسی اشیاء کاهش می دهد[3].

یکی از کارهای اولیه ای که توسط فیشلر[10] انجام شد شمایی پایین به بالا را پیشنهاد می دهد که اشیاء مختلف را در صحنه شناسایی می کند. برای این کار ابتدا تصویر به نواحی بخش بندی می شود، هر ناحیه به عنوان یک شئ برچسب گذاری می شود و سپس برچسب های اشیاء با استفاده از اطلاعات مکانی متن تصحیح می شود. برای تصحیح می توان اشیاء را به بخش های کوچکتری تقسیم بندی کرد و محدوده ی امکان پذیر روابط مکانی این بخش ها را تعریف کرد. اطلاعات مکانی متن به شکل قوانین و ساختارهای شبیه گراف ذخیره شده اند.

یکی از کارهای اخیر که توسط شاتن و همکارانش[11] انجام شده است از اطلاعات بافت، چیدمان و اطلاعات مکانی استفاده می کند. اطلاعات مکانی متن توسط کار کومار[12] و هبرت[13] نیز معرفی شده است. روش آنها فرمولی سلسله مراتبی دولایه نمایش می دهد. همان طور که در شکل 6-1 نشان داده

[10] Fischler
[11] Shotton et al
[12] Kumar
[13] Hebert

شده است، لایه اول ارتباط متقابل ناحیه به ناحیه را مدل می کند و لایه دوم ارتباط متقابل شئ با شئ را مدل می کند[3].



شکل 6-1 – سمت چپ: کلاسبندهای مکانی متن برای سه کلاس مختلف آسمان، چهره و جاده. سمت راست: فرمول بندی دولایه ی سلسله مراتبی برای اطلاعات مکانی متن[3]

کارهای اخیر در بینایی کامپیوتر از داده های آموزشی همراه با تفسیر برای بدست آوردن اطلاعات مکانی متن استفاده می کنند. بعضی از روش ها نیز از قوانین از پیش تعریف شده استفاده کرده اند.

۱-۳-۳: اطلاعات متنی مقیاس

روش های متداول بازشناسی شئ نیاز به جستجوی فضای جستجوی بزرگ برای جستجوی مدل ها، موقعیت ها و اندازه های مختلف شئ داشتند. اطلاعات پیشین در مورد سایزهایی که اشیاء در صحنه‌ها ظاهر می شوند، فرآیند بازیابی شئ را بهبود می بخشد و نیاز به جستجوی اندازه های مختلف را کاهش می دهد. درنتیجه منابع محاسباتی را به مقیاس های محتمل تر اختصاص می دهد[3].

برای بدست آوردن اطلاعات مقیاس متن علاوه بر شناسایی حداقل یک شئ دیگر، نیاز به پردازش ارتباطات مکانی و عمق بین شئ موردنظر و این اشیاء دیگر است. سیستم کندر[14] که توسط استرت [15] و فیشلر[16] ارائه شد از اولین سیستم های بینایی کامپیوتر بود که اطلاعات مقیاس را به عنوان یک

[14] CONDOR
[15] Strat
[16] Fischler

ویژگی درنظر گرفت. اطلاعات مقیاس یک شئ از طریق موقعیت دوربین، جهت، فضای هندسی، ارتفاع زمین و نقشه محاسبه می شود.

تورالبا و همکارانش[17] چهارچوبی ساده برای مدل کردن رابطه بین متن و خصوصیات شئ معرفی کرد. اطلاعات متنی از طریق تصاویر آموزشی که در آن خصوصیات شئ برمبنای همبستگی های بین آماره های خصوصیات سطح پایین کل صحنه است، یاد گرفته می شوند. شکل 7-1 مثالی از یک تصویر آموزشی و تقسیر مربوط به آن را نشان می دهد.

شکل 7-1- مثالی از تصویر آموزشی و تفسیر استفاده شده برای آن[3]

اطلاعات مقیاس متن سخت ترین ارتباط قابل دستیابی است زیرا نیاز به اطلاعات جزئی درمورد شئ در صحنه دارد.

بیشتر مدل ها از یک یا دو نوع اطلاعات متنی استفاده می کنند. اطلاعات متنی مکانی و مقیاس بیشترین استفاده را داشته اند. به طور کلی اطلاعات متن معنائی به طور ضمنی در اطلاعات مکانی متن موجود است. همچنین در مورد مقیاس با توجه به اشیاء دیگر درنظر گرفته می شود. بنابراین استفاده از اطلاعات مکانی و مقیاس متن استفاده از تمام اطلاعات متنی را شامل می شود.

هرچند اطلاعات معنائی متن توسط انواع دیگر قابل دستیابی است اما تنها نوع اطلاعات متنی است که ارزشمندترین اطلاعات را به ما می دهد و همچنین دسترسی به آن بسیار ساده تر است و پردازش و محاسبات آن سریع تر است.

[17] Torralba et al

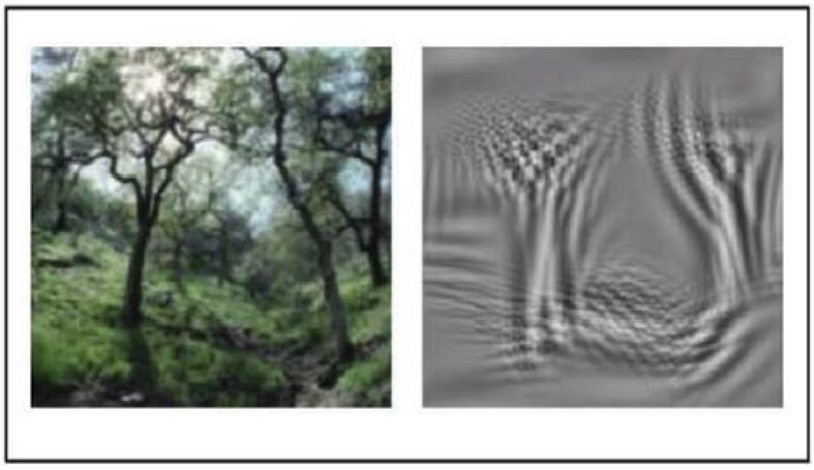
۱-۳-۴: سطوح متنی

مدل های بازشناسی شئ استفاده از اطلاعات متنی را از دو سطح سراسری و محلی تصویر درنظر گرفته‌اند. اطلاعات سراسری آماره های تصویر را در یک تصویر به عنوان یک کل درنظر می گیرد (مثلا یک آشپزخانه حضور یک فر را پیش بینی می کند). اطلاعات محلی متنی اطلاعات متنی را برای ناحیه های همسایه شئ درنظر می گیرد (مثلا یک چراغ خواب حضور یک ساعت شماطه ای را پیش بینی می کند)[3].

۱.۳.۴.۱اطلاعات متنی سراسری

ساختار یک صحنه را می توان با میانگین گرفتن ویژگی های سراسری تصویر بدست آورد، که خلاصه‌ای آماری از چیدمان مکانی فراهم می کند. بسیاری از چهارچوب های طبقه بندی شئ از این اطلاعات پیشین برای تعیین موقعیت استفاده کرده اند.

مورفی و همکارانش[18] از ویژگی های متنی با استفاده از "جیست[۱۹]"صحنه استفاده کرده است. جیست یک صحنه نمایشی کلی و سطح پایین از کل تصویر است. شکل 8-1 یک مثال از جیست یک جنگل را نشان می‌دهد. تورالبا و همکارانش نشان می دهند که فراهم کردن اطلاعات پیشین درمورد اینکه چه اشیائی در یک صحنه و در چه موقعیت و اندازه ای ظاهر خواهند شد، کافیست.



شکل 8-1 – یک تصویر آموزشی و جیست آن [3]

[18] Murphy et al
[19] Gist (Global Invariant Scale Transform)

استفاده از اطلاعات متنی با پردازش صحنه به عنوان یک کل و بدون بازیابی ابتدایی اشیاء دیگر به کاهش بازیابی های غلط می انجامد.

۱.۳.۴.۲ اطلاعات متنی محلی

اطلاعات متنی محلی از مناطق اطراف شئ شامل اشیاء دیگر، پیکسل ها و تکه ها بدست می آید. برای انجام این کار ابتدا باید شئ شناخته شود و قسمت هایی که شباهت لازم با شئ را نداشته باشند کنار گذاشته می شوند. زمانی که چهارچوب بازشناسی دارای تصاویری باشد که شئ در مرکز آن قرار داشته باشد، استفاده از اطلاعات متنی محلی باعث بهبود بازشناسی خواهد شد[3].

یکی از مزایای اطلاعات متنی سراسری نسبت به محلی این است که از نظر محاسباتی کم هزینه تر است. البته زمانیکه تعداد اشیائی که می خواهیم بازشناسی کنیم زیاد باشد اطلاعات متنی سراسری نمی تواند بین صحنه های مختلف تمایز قائل شود.

نمونه های متعددی از این ایده کلی تاکنون استفاده شده است. مدل های مبتنی بر بخش نوعی از اطلاعات متنی محلی را استفاده می کنند که بخش های مختلف باید با رابطه ی هندسی مناسبی نسبت به یکدیگر قرار بگیرند تا یک شئ را تشکیل دهند.

بسیاری از روش ها از جیست صحنه برای تصمیم گیری درمورد مکان محتمل برای اشیاء مختلف استفاده می کنند. برای مثال تورالبا و همکارانش سیستمی برای پیش بینی موقعیت عمودی اشیاء مانند افراد پیاده رو، ماشین ها و ساختمان ها براساس جیست صحنه آموزش داده اند. این مکان ها سپس توسط یابنده های شئ کلاسیک استفاده شده اند که عملکرد یابنده ها را بهبود می بخشند.

۱–۴: جایگاه و اهمیت موضوع

تکنولوژی های بازشناسی در صنعت غذا (مثلا برای خودکارسازی دسته بندی محصولات کشاورزی)، در صنعت خودرو و الکترونیک (برای اتوماتیک سازی مونتاژ و هدف های بازبینی صنعتی)، درصنعت داروسازی (برای دسته بندی قرص ها و کپسول ها)، در بهبود جستجوی آنلاین، در عکس برداری پیشرفته، در سیستم های بازشناسی بیومتریک برمبنای عنبیه یا اثرانگشت و همچنین شناسایی دستخط، در صنعت روبات، در وسایل نقلیه هوشمند و بسیاری کاربردهای دیگر استفاده می‌شوند.

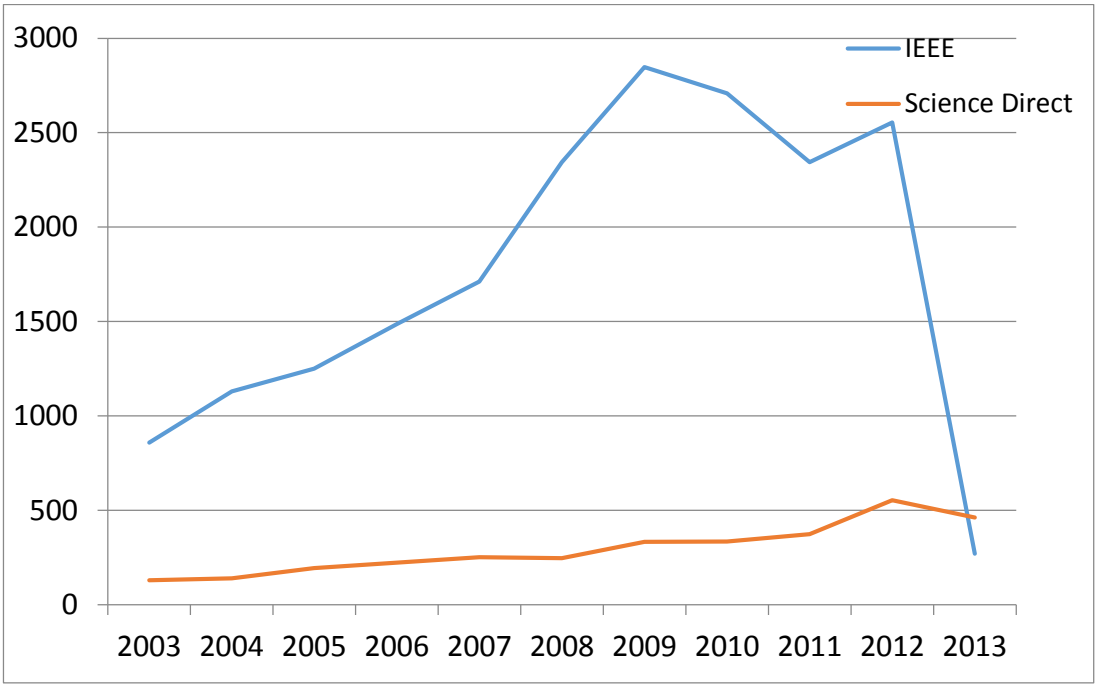
بسیاری از مدل های نمایش اشیاء به طور موثری توسط جامعه‌ی تصویربرداری پزشکی برای بخش بندی قدرتمند ساختارهای تشریحی مانند مغز و حفره های قلب استفاده می‌شوند.

روش های متداول بازشناسی شئ بر تصاویری آزموده می شوند که با تصاویر دنیای واقعی متفاوت بوده و معمولا دارای تعداد کمی شئ و در صحنه های خاصی می باشند. مجموعه داده ی سان[۲۰] دارای تصاویر از صحنه های متنوع می باشد و تعداد اشیاء موجود در این تصاویر نیز بسیار زیاد می باشند[4]. روش های بازشناسی متداول بر این مجموعه داده دقت پایینی را ارائه می دهند و دلیل آن پیچیدگی بیشتر فرآیند بازشناسی شئ به دلیل تنوع صحنه ها و شلوغی آنهاست. به همین دلیل استفاده از اطلاعات متنی جایگاه و اهمیت ویژه ای می یابد.

به دلیل اهمیت موضوع بازشناسی شئ مسابقات زیادی در سراسر دنیا در این زمینه برگزار می شود. یکی از این مسابقات رقابت پاسکال[۲۱] است که از سال 2005 هرساله برگزار شده است. یکی دیگر از این رقابت ها ایمیج نت[۲۲] است که هدف پیش بینی محتوای عکس ها برای انجام حاشیه نویسی خودکار و بازیابی تصاویر دیتاست ایمیج نت برای آموزش است. این رقابت نیز از سال 2010 هر ساله برگزار می‌شود و در بخش های مختلف عملکرد الگوریتم های شرکت داده شده سنجیده می شود و تعدادی از بهترین الگوریتم ها به عنوان برنده انتخاب می‌شوند.

روند تحقیقات در زمینه بازیابی و بازشناسی شئ در شکل 9-1 نشان داده شده است. همانطور که دیده می شود تعداد مقالات در دو پایگاه اطلاعاتی معتبر IEEE[23] و Science direct[24] از سال 2003 تا 2013 که بررسی شده است، روند روبه رشدی را داشته است. افزایش تعداد مقالات نشان دهنده ی اهمیت این موضوع و نیاز به تلاش و تحقیق در این زمینه است.

[20] SUN http://groups.csail.mit.edu/vision/SUN/
[21] PASCAL http://pascallin.ecs.soton.ac.uk/challenges/VOC/
[22] ImageNet http://www.image-net.org/
[23] http://ieee.org/
[24] http://sciencedirect.com/



شکل 9-1- نمودار تعداد مقالات در زمینه بازیابی شئ در IEEE و Science Direct

۱-۵:رویکرد پیشنهادی

در این پایان نامه هدف بهبود دقت فرآیند بازیابی شئ با استفاده از اطلاعات متنی معنایی است. سوال اصلی تحقیق این است که از چه اطلاعات معنایی استفاده شود و چگونه این اطلاعات مدل شوند که بتوان به کمک آنها بازیابی شئ انجام داد و بتوان بر مجموعه داده مناسب نتایج قابل قبول بدست آورد. مجموعه داده مناسب مجموعه داده ای است که دارای تصاویر از اشیاء مختلف و در صحنه های متنوع باشد و نتایج قابل قبول نتایجی هستند که نسبت به نتایج ارائه شده قبلی بر همان مجموعه داده برتری داشته باشند. یکی از راه های استفاده از اطلاعات معنایی استفاده از ماتریس همرخداد است که در این روش با اعمال تغییراتی از این ماتریس استفاده شده است. جهت کلاسبندی نیز از ماشین بردار پشتیبان[۲۵] استفاده شده است.

برای ارزیابی عملکرد این الگوریتم از مجموعه داده سان استفاده شده است. دقت بازیابی هر شئ به صورت نسبت پیش بینی های شئ انجام شده که درست بوده اند، تعریف می شود. برای هر شئ

[25] Support Vector Machine (SVM)

میانگین دقت پیش‌بینی‌هایی که در هر مرحله بدست آمده است به عنوان میانگین دقت[۲۶] پیش‌بینی حضور آن شئ درنظر گرفته می شود.

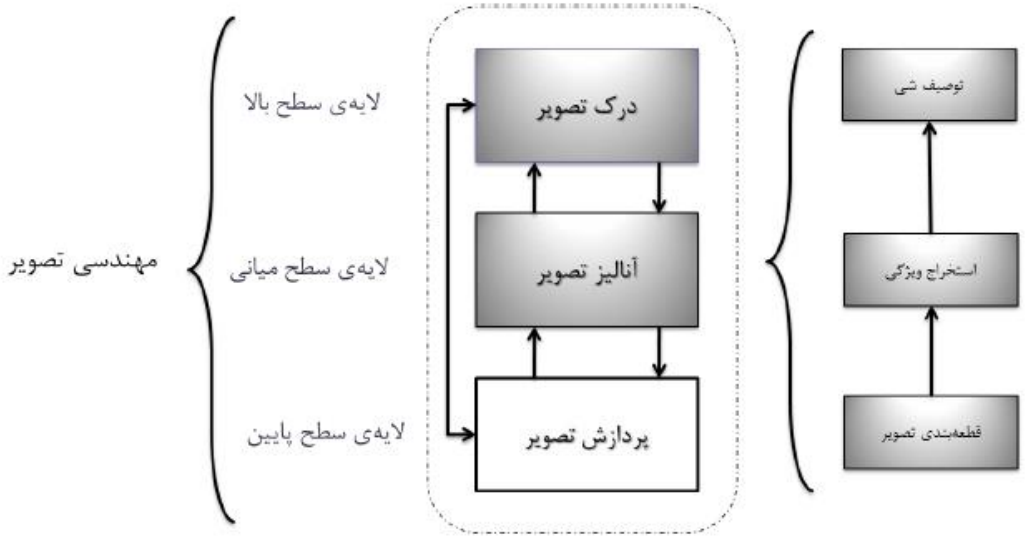
درنهایت نیز برای ارزیابی الگوریتم، متوسط میانگین دقت[۲۷] تمام اشیاء محاسبه و به عنوان معیار ارزیابی در نظر گرفته می شود. در دومقاله ای که با الگوریتم پیشنهادی مقایسه شده اند 107 شئ بررسی و میانگین دقت آنها محاسبه شده است. برای اینکه بتوان نتایج را مقایسه و الگوریتم را ارزیابی کرد در این الگوریتم نیز میانگین دقت برای همان 107 شئ محاسبه شده است.

۱-۶:حوزه مساله

یکی از مراحل مهم در انجام پایان نامه مشخص کردن حوزه مساله است تا از انجام فعالیت های غیرمرتبط با مساله جلوگیری شود، به همین دلیل در این بخش توضیح داده می شود که تحقیقات این پایان نامه در کدام حوزه ها قرار می گیرد.

فرآیند مهندسی تصویر در شکل 10-1 نشان داده شده است[5]. در این مراحل ابتدا در لایه سطح پایین تصویر پردازش می شود و نویز و مشکلات احتمالی تصویر برطرف می شود. این لایه در حوزه تحقیق قرار ندارد و تصویر ورودی تصویری آماده و پردازش شده است. اما دو لایه دیگر شامل آنالیز تصویر و درک تصویر در حوزه مساله قرار می گیرد و برای بازشناسی شئ آنالیز و درک تصویر انجام می‌شود.

[26] AP(Average precision)
[27] Mean AP



شکل 10-1- مهندسی تصویر شامل سه لایه اصلی پردازش تصویر، آنالیز تصویر و درک تصویر است که فرآیند بازشناسی شئ بر لایه آنالیز و درک تصویر تمرکز دارد[5].

اگر روش های بازشناسی شئ را به دو دسته مبتنی بر اطلاعات شئ و مبتنی بر اطلاعات متنی تقسیم کنیم، روش مبتنی بر اطلاعات متنی در حوزه این تحقیق قرار می گیرد و هدف یافتن اطلاعات متنی مناسب و مدل سازی آنها برای بازیابی شئ است تا بتوان دقت بازشناسی اشیاء را افزایش داد.

تصاویر موردنیاز این تحقیق از مجموعه داده سان09 فراهم شده است که شامل بیش از 600 صحنه و بیش از 200 شئ است. بیشتر تصاویر این مجموعه داده شامل بیش از ده کلاس شئ است، درحالی که بیشتر تصاویر در مجموعه داده های دیگر حداکثر در هر تصویر دارای پنج کلاس شئ هستند.

دو مقاله ای که تاکنون بهترین عملکرد را در حوزه بازشناسی شئ داشته اند برای مقایسه با روش پیشنهادی انتخاب شده اند و نتایج با هم مقایسه و در فصل های بعدی شرح داده شده اند. در این دو مقاله میانگین دقت بازیابی شئ برای 107 کلاس شئ محاسبه شده اند و نتیجه نهایی براساس میانگین دقت بازیابی 107 شئ مقایسه می شوند که هدف این تحقیق افزایش میانگین دقت بازیابی 107 کلاس شئ مشخص شده بر این مجموعه داده است. حاصل این تحقیق می تواند به صورت نرم افزار کاربردی در صنعت و خدمات استفاده شود و برای تصاویر متنوع از صحنه های مختلف و با تعداد اشیاء زیاد کاربرد دارد.

۱–۷: ساختار پایان نامه

در این پایان‌نامه پنج فصل تنظیم شده است. در فصل دو ابتدا به معرفی بخش های مختلف سیستم بازشناسی شئ می‌پردازیم و الگوریتم‌های مورداستفاده در هر بخش را معرفی می‌کنیم.

در فصل سوم روش‌های مطرح در این زمینه معرفی شده و نقاط قوت و ضعف آنها بررسی می‌شود، به ویژه بر روش‌های مبتنی بر اطلاعات متنی و مبتنی بر درخت تصمیم تمرکز شده است.

در فصل چهارم چالش‌های روش‌های قبلی مطرح می‌شوند و رویکرد کلی الگوریتم پیشنهادی پایان-نامه گفته می شود و سپس به طور دقیق بخش های مختلف آن توضیح داده می شوند. پس از آن مجموعه داده مناسب انتخاب و علت انتخاب آن ذکر می شود.

در فصل پنجم به بررسی عملکرد و ارزیابی الگوریتم پیشنهادی پرداخته می شود و نتایج دو روش با بهترین عملکرد ذکر می شوند و با الگوریتم پیشنهادی مقایسه می شوند و بررسی می شود که آیا این الگوریتم دقت میانگین بازشناسی اشیاء را افزایش داده است یا خیر.

درنهایت در فصل آخر نیز به نتیجه گیری و جمع بندی مطالب پرداخته می شود و کارهای آتی بیان می شوند.

فصل۲:

مروری بر کارهای انجام شده

۲–۱: مقدمه

در این فصل ابتدا ویژگی‌های مهم تصویر که به صورت کلی برای بازشناسی شئ از تصاویر دیجیتال استفاده می‌شوند، بررسی می شوند و مدل های مختلف استخراج این ویژگی ها توضیح داده می شوند. سپس در مورد بخش دوم یک سیستم بازشناسی شئ یعنی یادگیری سیستم و تعیین استراتژی بازشناسی شئ توضیح داده می‌شود و مدل‌های مختلف بررسی می‌شوند. پس‌ازآن تعدادی از الگوریتم‌های ارائه شده برای بازشناسی شئ را مرور می کنیم.

۲–۲: بررسی الگوریتم های استخراج ویژگی موجود در زمینه بازشناسی شئ

در این بخش به الگوریتم‌های موجود که در زمینه بازشناسی شئ به تعدد مورد استفاده قرار گرفته‌اند، می‌پردازیم. در ادامه ابتدا دسته بندی کلی ویژگی ها را بیان می کنیم و سپس روش‌های متداول بازشناسی شئ مبتنی بر این ویژگی ها را توضیح می دهیم.

۲–۲–۱: دسته بندی کلی ویژگی‌ها

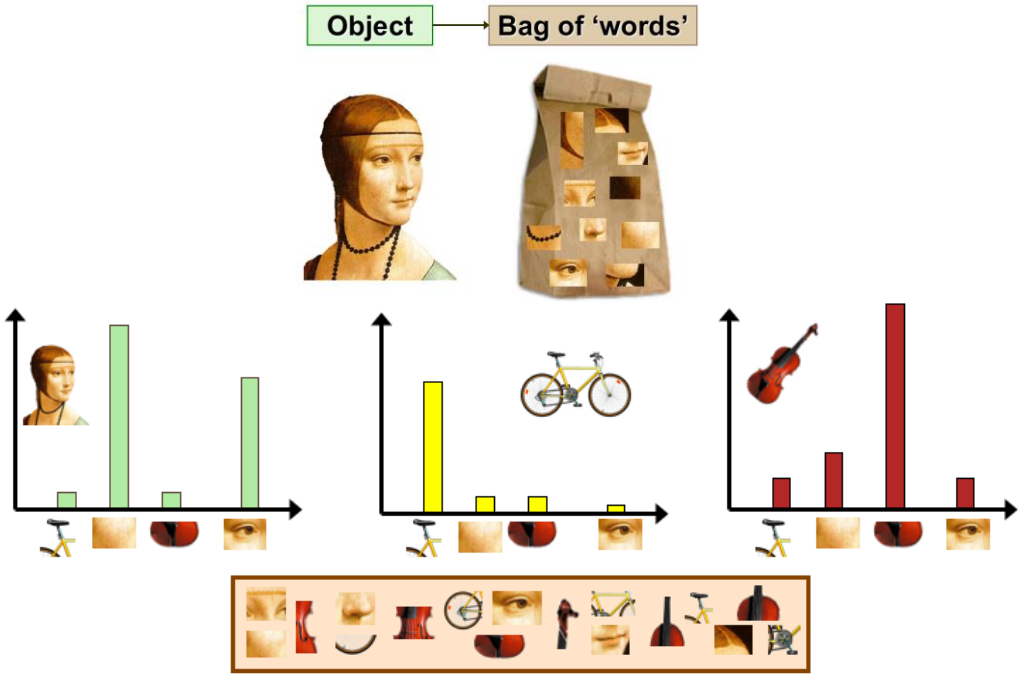
ساده ترین راه برای نمایش یک تصویر استفاده از تمام پیکسل های آن به عنوان ویژگی می باشد، اما این راه بسیار ابتدایی و دارای ابعاد زیاد می باشد. برای بدست آوردن اطلاعات مفید یا بردار ویژگی از تصاویر روش های مختلفی وجود دارد. از نظر ناحیه ای که برای بدست آوردن اطلاعات درنظر گرفته می شود، می توان ویژگی ها را به سه دسته تقسیم کرد:

- ویژگی های محلی: این ویژگی ها معمولا یک فضای کوچک قابل تمیز از یک ناحیه را بازنمایی می کنند. به عنوان مثال انحنا، قطعات مرزی و گوشه ها از ویژگی های محلی به شمار می آیند[6].
- ویژگی های سراسری: این ویژگی ها تصویر را به عنوان یک ناحیه درنظر می گیرند برخی از این ویژگی ها عبارتند از توصیفگرهای فوریه و گشتاور ها[6].
- ویژگی های رابطه ای: این ویژگی ها براساس موقعیت های نسبی موجودیت های مختلفی مانند نواحی، مرزهای بسته یا ویژگی های محلی هستند[6].

ویژگی مورداستفاده برای بازشناسی شئ تاثیر بسیار زیادی بر عملکرد سیستم دارد. در ادامه چند مورد از روش های استخراج ویژگی متداول را توضیح می دهیم.

۲-۲-۲: بسته کلمات

در این روش بخش های مستقلی از یک تصویر به عنوان کلمه در نظر گرفته می شوند و فرکانس تکرار آنها در تصویر محاسبه می شود. در نتیجه نحوه نمایش یک تصویر با استفاده از هیستوگرام ویژگی های تخصیص یافته به هر بخش است. در شکل 1-2 چگونگی نمایش تصاویر با استفاده از این ویژگی ها نشان داده شده است.



شکل 1-2 - نمایش سه شئ به روش کیف کلمات

۲-۲-۳: مدل مبتنی بر اجزاء

در روش کیف کلمات تمام قطعات تصویر با احتمال یکسانی درنظر گرفته می شوند و همچنین اطلاعاتی در مورد محل این قطعات درنظر گرفته نمی شود. در روش مبتنی براجزاء یک شئ به عنوان مجموعه ای از اجزاء درنظر گرفته می شود و یک مدل مبتنی برظاهر است که موقعیت مکانی بین اجزاء را نیز درنظر می گیرد.

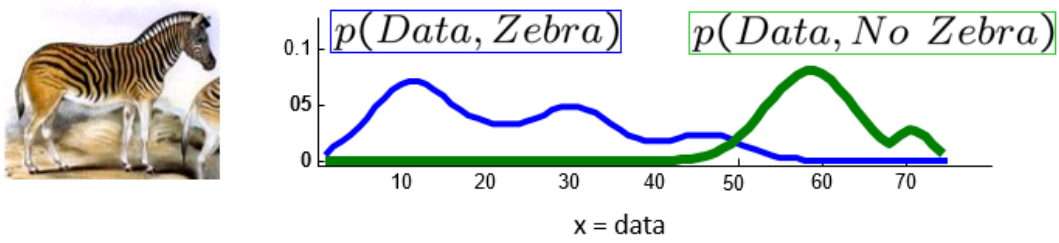
۲–۳: بررسی الگوریتم‌های یادگیری سیستم و بازشناسی شئ

مدلی که برای بدست آوردن برچسب هر ناحیه براساس ویژگی های بدست آمده استفاده می شود به سه دسته تقسیم می شود. مدل زایشی[۲۸]، مدل جداکننده[۲۹] و تابع کلاسبندی که درادامه هر یک توضیح داده می شوند.

۲–۳–۱: مدل زایشی

اگر ویژگی ها را به عنوان بردار x و برچسب ها را به عنوان بردار y در نظر بگیریم، در آن صورت در مدل زایشی هدف بدست آوردن احتمال توام x و y است. در شکل 2-2 برای مثال گورخر این احتمال نشان داده شده است. در این مدل احتمال توام داده های ورودی و گورخر بودن و گورخر نبودن شئ نشان داده شده است. در این مثال y به صورت باینری 0 و 1 درنظر گرفته شده است که در صورت 1 بودن شئ گورخر هست و درصورت 0 بودن شئ گورخر نیست.

این مدل به همراه قانون بیز می تواند برای کلاسبندی استفاده شود و یا اینکه با استفاده از این مدل می توان احتمال رخداد جفت های x و y را بدست آورد.

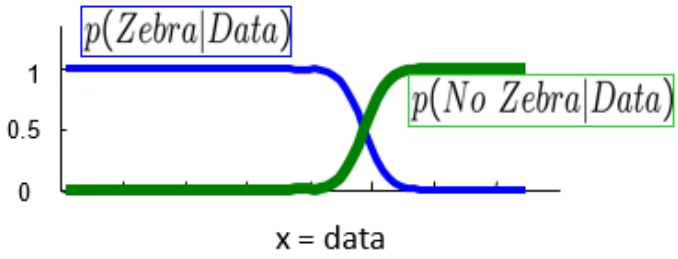


شکل 2-2 - یک نمونه مثال مدل زایشی. P(Data, Zebra) نشان دهنده‌ی احتمال گورخر بودن و داده‌های ورودی به طور همزمان است و P(Data, No Zebra) نشان دهنده‌ی احتمال گورخر نبودن و داده‌های ورودی است.

۲–۳–۲: مدل جداکننده

در این مدل احتمال شرطی y نسبت به x محاسبه می شود. یعنی با داشتن داده های x احتمال برچسب موردنظر محاسبه می شود. از این مدل می توان به عنوان کلاسبند استفاده کرد. شکل زیر مدل جداکننده را برای مثال گورخر نشان داده است.

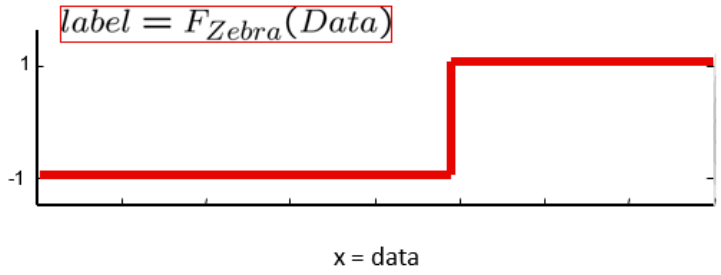
[28] Generative
[29] Discriminative



شکل 3-2- یک مثال از مدل جداکننده. p(Zebra|Data) نشان دهنده‌ی احتمال گورخر بودن به شرط داده‌های ورودی و p(No Zebra|Data) نشان دهنده‌ی احتمال گورخر نبودن به شرط داده‌های ورودی است.

۲-۳-۳: تابع کلاسبندی

دراین مدل برچسب، براساس تابعی از داده ها یا ویژگی ها محاسبه می شود. در شکل زیر از برچسب 1 و 1- برای نشان دادن اینکه آیا شئ موردنظر گورخر هست یا خیر استفاده شده است.



شکل 4-2 – یک مثال از تابع کلاسبندی. در این مثال توسط تابع F و ورودی Data تعیین می شود که آیا گورخر هست یا خیر.

روشهای مختلفی برای کلاسبندی وجود دارد که در این قسمت بعضی از متداول ترین روشها توضیح داده می شوند.

- روش نزدیک ترین همسایه

در این روش هر نمونه به نزدیک ترین کلاس نسبت داده می شود. هر کلاس توسط میانگین نمونه-های آن کلاس تعریف می شود و با استفاده از فاصله اقلیدسی نمونه جدید با تمام میانگین های

کلاس‌های مختلف در مورد کلاس نمونه جدید تصمیم گیری می شود. از مزایای این روش سادگی آن و از معایب این روش این است که باید نمونه های زیادی داشته باشیم تا دقت بالایی بدست آید.

- روش شبکه عصبی

شبکه عصبی روشی الهام گرفته از سیستم های یادگیر طبیعی است که در آنها یک مجموعه پیچیده از نرونهای به هم متصل در کار یادگیری دخیل هستند. این روش برای یادگیری توابع گوناگون نظیر توابع با مقادیر حقیقی، توابع با مقادیر گسسته و توابع با مقادیر برداری مناسب می باشد. اینگونه شبکه‌ها با موفقیت در مسائلی مانند شناسایی و تعبیر تصاویر، یادگیری روبات و شناسایی گفتار اعمال شده است. شبکه از تعداد دلخواهی سلول یا گره یا واحد یا نرون تشکیل می شود که مجموعه ورودی را به خروجی ربط می دهند.

این روش برای یادگیری توابع پیوسته و زمانی که زمان کافی برای یادگیری وجود داشته باشد مناسب است و نسبت به درخت تصمیم زمان بیشتری برای یادگیری نیاز دارد. وزن های یادگرفته شده توسط شبکه را به سختی می توان تعبیر کرد به همین دلیل زمانی استفاده می شود که نیازی به تعبیر تابع هدف نباشد. برای داده های آموزشی دارای نویز مانند داده های سنسورها نظیر دوربین و میکروفن نیز مناسب است.

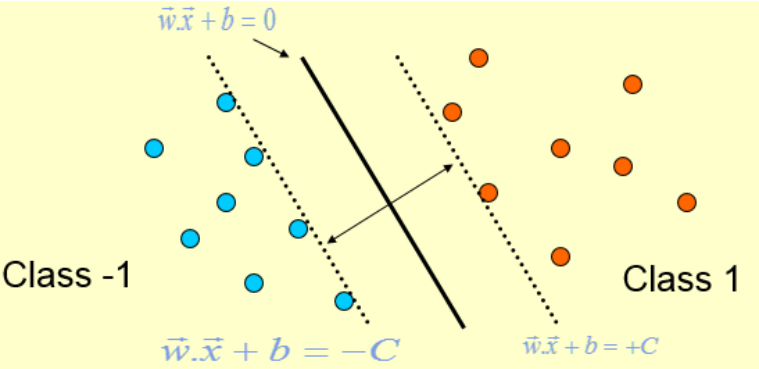
- روش ماشین بردار پشتیبان[۳۰]

روش ماشین بردار پشتیبان روشی در یادگیری ماشین برای دسته بندی است و شهرت آن به دلیل موفقیت در تشخیص حروف دست نویس است. در این روش در صورتی که دسته ها بصورت خطی جداپذیر باشند، ابرصفحه هائی با حداکثر حاشیه بدست می آید که دسته ها را جدا می کند. در مسائلی که داده ها بصورت خطی جداپذیر نباشند، داده ها به فضای با ابعاد بیشتر نگاشت پیدا می کنند تا بتوان در فضای جدید آنها را بصورت خطی جدا کرد [7].

به دلیل اینکه جداکننده ها با حداکثر حاشیه تنظیم می‌شوند این جداکننده ها قابلیت تعمیم بیشتری پیدا می کنند. طبق قضیه ای در تئوری یادگیری ماشین اگر مثالهای آموزشی بدرستی دسته

[30] Support Vector Machine (SVM)

بندی شده باشند، از بین جداسازهای خطی، آن جداسازی که حاشیه داده های آموزشی را حداکثر میکند خطای تعمیم را حداقل خواهد کرد.



شکل 5-2- جداسازی با حداکثر حاشیه در ماشین بردار پشتیبان

مزایای این روش عبارتند از: آموزش در این روش به سادگی انجام می شود، قابلیت تعمیم خوبی دارد، با نمونه های آموزشی کم نیز به خوبی کار می کند، برخلاف شبکه عصبی بهینه محلی نیست و بهترین پاسخ سراسری را پیدا می کند. از ضعف های این روش نیز نیاز به انتخاب تابع هسته خوب است درغیر این صورت پاسخ درستی بدست نخواهد آمد.

- روش بوستینگ[۳۱]

روش بوستینگ یک روش کلاسبندی براساس ترکیب تعدادی کلاسبند ضعیف برای ساختن یک کلاسبند قوی است. در هر الگوریتم کلاسبندی ضعیف سعی می شود یک ویژگی با کمترین خطا برای جدا کردن نمونه های مثبت و منفی به عنوان ویژگی جداکننده انتخاب شود. درنهایت نیز با ترکیب این کلاسبندها، کلاسبند نهایی بدست می آید.

- روش درخت تصمیم

ایده ی درخت تصمیم بسیار ساده و شهودی است و یک مثال نقطه شروع خوبی برای فرد ناآشنا است. این روش مانند بازی بیست سوالی است که براساس پاسخ هر سوال به هر نود حرکت انجام

[31] Boosting

می‌شود. دو فرآیند باید انجام شود: اول ساخت درخت و دوم کلاسبندی براساس آن . ابتدا به تعریف واژگان در این زمینه پرداخته می شود:

- مثالها: آنچه کلاسبندی می شود و برچسب گذاری می شود تا کلاسی که به آن تعلق دارد شناسایی شود.
- دیتاست: مجموعه ی مثالهاست.
- ویژگی ها: معیارهایی که از مثالها استخراج می شود مانند ویژگی مدوربودن که معیاری برای تشخیص این است که چقدر یک مثال مانند دایره است.
- سوالها: تعریف کننده ی مرزها در فضای ویژگی
- گره ها: درخت را می سازند و شامل مثالهایی هستند و اگر گره ی پایانی نباشند شامل سوالی هستند که پاسخ آن به عنوان گره دیگری نشان داده می شود.

برای ساخت درخت باید سوالهایی برای تقسیم دیتاست به صورت معناداری پیدا کرد. برای کلاسبندی یک مثال جدید درخت براساس پاسخ سوالها در هر گره پیمایش می شود، زمانیکه به گره برگ برسیم کلاسبندی براساس مثالهای یادگرفته شده در آن گره انجام شده است.

یک کلاسبندی خوب به درخت بستگی دارد. برای بازشناسی سریع و صحیح فرآیند آموزش پیچیده و زمانبر خواهد بود. برای اینکه چه سوالی در هر گره پرسیده شود محدودیتی وجود ندارد. مساله اساسی در درخت تصمیم این است که در هر گره راههای ممکن فراوانی برای تقسیم داده وجود دارد.

۲-۴: روش‌های متداول بازشناسی شئ

یکی از روش‌های مورد استفاده در سیستم‌های بازشناسی شئ که بسیار مورد توجه قرار گرفت توسط ویالا[۳۲] و جونز[۳۳] ارائه شد که در این روش تصویر توسط پنجره‌هایی با ابعاد مختلف بررسی می‌شود و در صورتی شئ شناسایی می‌شود که پنجره با ابعاد شئ روی آن قرار بگیرد. در واقع در این روش برای کشف حضور شئ از تطبیق الگو استفاده می‌شود. این روش برای محیط‌های خاص که روشنایی و حالت شئ کنترل شده باشد کارآمد است، اما زمانی که شئ دارای چرخش، تغییر اندازه، تغییر حالت یا

[32] Viola
[33] Jones

روشنایی باشد از لحاظ محاسباتی امکان پذیر نیست به خصوص زمانیکه بخشی از شئ نمایان باشد و زمانیکه پایگاه داده‌های بزرگ داشته باشیم[8, 9].

یک راه حل به جای جستجوی تمام موقعیت‌های تصویر و پیدا کردن تطابق، این است که ویژگی‌هایی از تصویر که نسبت به شکل تصویر تغییرناپذیرند استخراج شود و این ویژگی‌ها تطبیق داده شود. انواع خصوصیات کاندید شامل قطعات خط، گروه‌های لبه و ناحیه پیشنهاد و بررسی شدند. این ویژگی‌ها برای کلاس‌های خاصی از شئ به خوبی کار می‌کردند اما استحکام کافی برای ایجاد اساس بازشناسی قابل اعتماد را دارا نبودند.

ویژگی‌های سراسری و محلی هر کدام مزایا و معایب خود را دارا هستند. ویژگی‌های سراسری بازشناسی را سرعت می‌بخشند و ویژگی‌های محلی نسبت به تغییرات تصویر مانند شلوغی، تبدیلات و وجود مانع قدرت بیشتری دارند، اما تطبیق خصوصیات محلی بین اشیاء از زوایای دید مختلف زمان‌بر است. کارهای اخیر به توسعه‌ی مجموعه‌ی انبوه‌تری از ویژگی‌های تصویر پرداخته است. یکی از روش‌ها از بازیابی گوشه برای بازشناسی موقعیت‌های قابل تکرار تصویر استفاده می‌کند، به طور دقیق‌تر می‌توان گفت قله‌ی تغییرات محلی تصویر را بازشناسی می‌کند.

ژانگ[۳۴] و همکارانش از گوشه‌یاب هریس[۳۵] برای بازشناسی موقعیت‌های خصوصیات استفاده می‌کنند. در این روش علاوه بر تلاش برای یافتن ارتباط بین نواحی یک تصویر و تمام نواحی ممکن در تصویر دوم، تطبیق نواحی گوشه در هر تصویر با نواحی ممکن در همان تصویر نیز باعث صرفه‌جویی زیادی در زمان محاسبات انجام می‌شود.

گوشه‌یاب‌ها در روش‌های گفته شده یک مشکل بزرگ دارند، که تصویر را تنها در یک مقیاس بررسی می‌کنند و درصورت تغییر بزرگ در مقیاس این یابنده‌ها به نقاط مختلف تصویر واکنش نشان می‌دهند. همچنین به دلیل اینکه یابنده مقیاس شئ را نشان نمی‌دهد لازم است توصیفگر شئ ایجاد شود و تطبیق برای تعداد زیادی از مقیاس‌ها بررسی شود.

[34] Zhang
[35] Harris

بعضی روش‌ها نیز برای بازشناسی مبتنی بر ظاهر از تطبیق فضای پایه[۳۶] و هیستوگرام رنگ استفاده می‌کنند. این روش‌ها همه برای اشیاء جدا شده یا تصاویر بخش بندی شده با موفقیت عمل کرده‌اند اما به دلیل خصوصیات سراسری که در آنها استفاده می‌شود توسعه‌ی آنها به تصاویر شلوغ و تصاویری که بخشی از شئ پوشانده شده سخت است. اهبا[۳۷] و آیکوچی[۳۸] روش فضای پایه را با استفاده از تعداد زیادی پنجره‌های محلی کوچک با موفقیت برای تصاویر شلوغ استفاده کردند، اما این کار مانند روش تطبیق نمونه که باید برای هر پنجره کل تصویر پیمایش شود، نیاز به جستجوی سنگینی دارد.

در مقاله‌ی فلزنزواب و همکارانش [۳۹] سیستمی برای بازشناسی شئ براساس ترکیبی از مدل های جزئی چندمقیاسی منعطف[۴۰] شرح داده شده است. این سیستم قادر است کلاس های شئ با تنوع بالایی را نمایش دهد و بهترین نتایج را بر دیتابیس پاسکال بدست آورد. در این مقاله برای مدل سازی اشیاء از گرامرهای بصری استفاده شده است. مدل های مبنی برگرامر با نمایش اشیاء با استفاده از ساختارهای سلسله مراتبی متغیر، مدل های جزئی منعطف را تعمیم می دهند. هربخش در یک مدل مبنی برگرامر می تواند به طور مستقیم یا توسط اجزاء تعریف شود. مدل های مبتنی برگرامر تغییرات ساختاری را مدل می کنند[10].

به طور کلی بهبود عملکرد توسط غنی سازی مدل بسیار سخت است و مدل های ساده عموما بهتر از مدل های پیچیده عمل می کنند. دلیل عملکرد بهتر مدل های ساده این است که مدل های پیچیده در فاز آموزش با سختی های زیادی روبرو هستند. در بازیابی شئ، مدل های کیف ویژگی[۴۱] و قالب های سخت به سادگی با روش های جداکننده ماننده ماشین بردار پشتیبان آموزش داده می شوند. مدل های غنی تر برای آموزش سخت تر هستند زیرا معمولا از اطلاعات پنهان استفاده می کنند.

به عنوان مثال زمانیکه یک مدل مبنی بر جزء از تصاویری که دارای یک جعبه محصور اطراف اشیاء هستند آموزش داده می شود، موقعیت اجزاء برچسب ندارند و باید با آنها به عنوان متغیرهای پنهان در حین آموزش رفتار شود. اگر برچسب گذاری کامل تر باشد آموزش ضعیف تر خواهد بود. برچسب گذاری

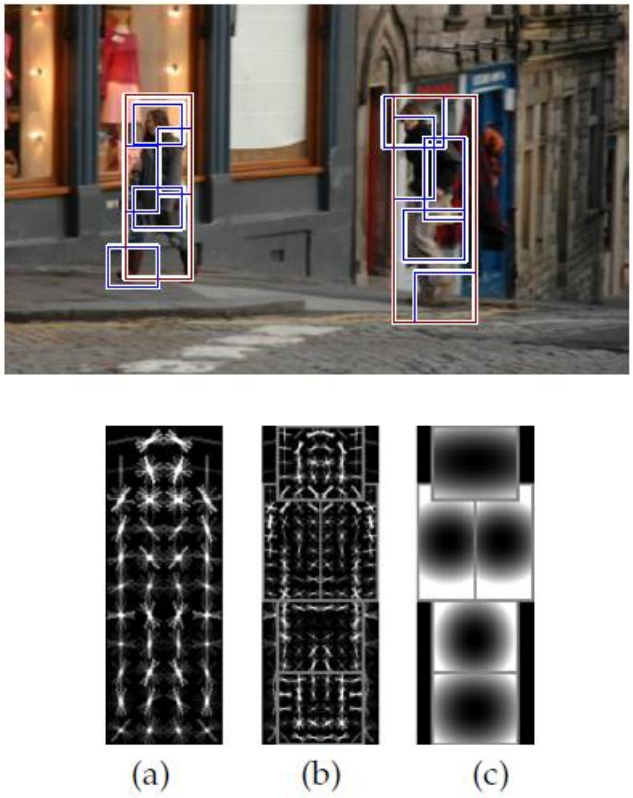
[36] Eigen space
[37] Ohba
[38] Ikeuchi
[39] Felzenszwalb, P. F., R. B. Girshick, et al.
[40] Multiscale deformable part models
[41] Bag of Features

جزئی اتوماتیک عملکرد بهتری دارد که به طور اتوماتیک اجزاء را پیدا می کند. برچسب گذاری با جزئیات بیشتر نیز بسیار پرهزینه و زمانبر است.

یابنده دالال- تریگز[42] که جایزه سال 2006 رقابت پاسکال را برنده شده است، از یک فیلتر بر ویژگی‌های هیستوگرام گرادیان ها استفاده می کند تا یک طبقه شئ را نمایش دهد. این یابنده از روش پنجره لغزان استفاده می کند که یک فیلتر به تمام موقعیت ها و به تمام اندازه ها در تصویر اعمال می‌شود. این یابنده یک کلاسبند است که به عنوان ورودی، تصویر و موقعیت و اندازه آن را می گیرد، و تصمیم می گیرد که آیا نمونه ای از شئ در موقعیت مورد نظر وجود دارد یا نه [11].

در این مقاله با استفاده از یک مدل با ساختار ستاره و مبتنی برجزء این ایده گسترش داده شده است. یک فیلتر ریشه و مجموعه ای از فیلترهای اجزاء تعریف می شوند. امتیاز فیلتر ریشه و اجزاء با ضرب نقطه ای فیلتر و زیرپنجره هرم ویژگی محاسبه شده از تصویر ورودی تعریف می شود. شکل زیر یک مدل ستاره برای طبقه شئ شخص را نشان می دهد.

[42] N. Dalal and B. Triggs



شکل 6-2- بازیابی های بدست آمده با یک مدل شخص با یک عنصر اصلی. مدل با یک فیلتر ریشه(a) چندین فیلتر اجزاء با رزولوشن بالاتر (b) و یک مدل مکانی برای موقعیت هر جزء نسبت به ریشه (c) تعریف شده است. فیلترها وزن های هیستوگرام گرادیان های ویژگی ها را مشخص می کنند[11].

برای آموزش مدل از اطلاعات برچسب گذاری جزئی استفاده شده که ماشین بردار پشتیبان پنهان[۴۳] نامیده شده است. بدلیل آموزش جداکننده مهم است که از مجموعه آموزشی بزرگ استفاده شود. در بازشناسی شئ مساله آموزش نامتوازن است زیرا زمینه بسیار بیشتر از اشیاء است.

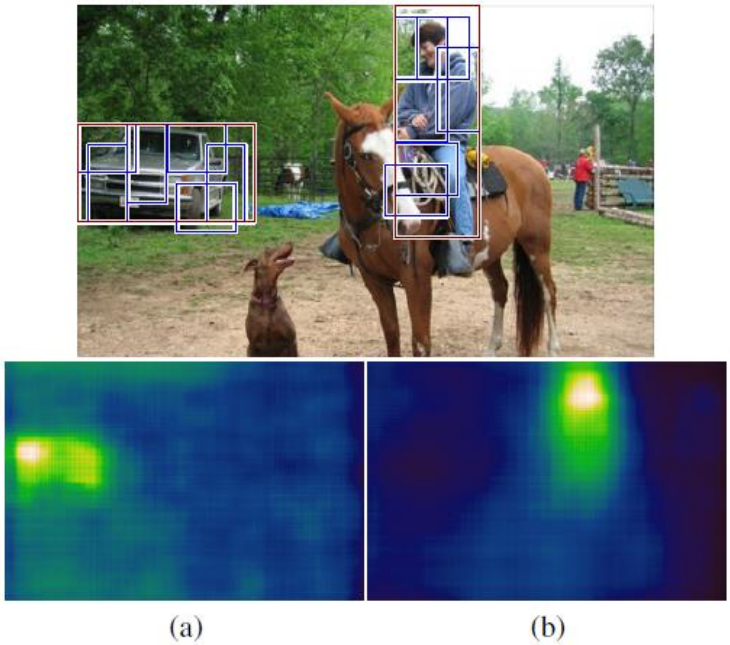
مدل های شئ توسط فیلترهایی که زیرپنجره های هرم ویژگی را امتیاز می دهند تعریف می شوند. مجموعه ویژگی شبیه ویژگی های هیستوگرام گرادیان استفاده شده است و با استفاده از تحلیل عناصر اصلی ابعاد آنها به طور قابل توجهی کاهش یافته است. همچنین بعضی مشکلات مرتبط با رقابت بازیابی شئ مجموعه داده پاسکال و مجموعه داده های مشابه برطرف شده است. یک روش ساده نیز برای جمع کردن خروجی یابنده های اشیاء مختلف اثبات شده است. ایده ی اصلی براین مبناست که وجود بعضی اشیاء دلیل برای وجود یا عدم وجود اشیاء دیگر در تصویر است. از این ایده به این صورت استفاده

[43] Latent SVM

می‌شود که در آموزش یک کلاسبند طبقه خاص هر بازیابی براساس امتیاز اصلی خود و بیشترین امتیاز از بقیه اشیاء دوباره امتیازدهی می شود.

در مقاله فلزنزواب و همکارانش[44] یک روش عمومی برای ساخت کلاسبندهای آبشاری از مدل های منعطف مبتنی برجزء مانند ساختارهای تصویری ارائه شده است. تمرکز اصلی در این مقاله بر مدلهای با ساختار ستاره است و نشان داده می شود که چطور یک الگوریتم ساده براساس هرس فرضیه های جزئی بازیابی شئ را سرعت می بخشند. در این الگوریتم فرضیه های جزئی با ترتیبی از آستانه ها هرس می شوند. درنهایت یک الگوریتم آبشاری بازیابی برای یک کلاس عمومی مدل‌ها با یک ظاهر گرامری تعریف می شود[12].

این الگوریتم یک روش بازیابی 20 بار سریعتر از بازیابی های استاندارد ارائه می دهد که براساس برنامه نویسی پویا و تبدیلات فاصله تعمیم داده شده و بدون کاهش عملکرد است. شکل زیر کاری را که توسط این الگوریتم در موقعیت های مختلف یک تصویر با استفاده از دو مدل مختلف انجام شده است، نشان داده است.



شکل 7-2- قسمت (a) از مدل ماشین و قسمت (b) ازمدل شخص استفاده شده است[12].

[44] P. Felzenszwalb, R. Girshick, D. McAllester

۲–۵: خلاصه و نتیجه گیری

در این فصل پس از دسته بندی کلی وِیژگی‌هایی که از تصویر استخراج می‌شوند، الگورتیم‌های متداول استخراج ویژگی در بازشناسی شئ معرفی شدند. سپس در مورد استراتژی‌های بازشناسی شئ و انواع روش یادگیری سیستم توضیح داده شد. پس از آن تاریخچه ای از روش‌های مختلف بازشناسی شئ مطرح و معایب و مزایای هر کدام بررسی شدند.

فصل۳:

پیش زمینه تئوری

۳–۱: مقدمه

در دنیای واقعی، اشیاء با اشیاء دیگر و محیط‌های مخصوص تغییر می‌کنند و منبعی غنی از پیوستگی‌های متنی فراهم می‌کنند که می‌توانند توسط سیستم‌های بصری استفاده شوند. مشاهدات زیادی در ادراک بصری[۴۵]، بینایی کامپیوتر و عصب شناسی[۴۶] نشان داده است که اطلاعات متن بر کارآمدی جستجو و شناسایی اشیاء تاثیرگذار است. اشیائی که در یک زمینه سازگار یا آشنا باشند با دقت بیشتری شناسایی می‌شوند و سریع‌تر از اشیائی که در یک صحنه‌ی ناسازگار هستند، پردازش می‌شوند[13]. مدل های زیادی برای طبقه بندی شئ از ارتباطات معنایی بیدرمن[۴۷] بهره برده اند. اطلاعات متنی عبارتست از هر اطلاعاتی که به طور مستقیم توسط ظاهر شئ تولید نشود. این اطلاعات را می توان از تگ ها، حاشیه نویسی ها، وجود و موقعیت اشیاء دیگر بدست آورد[3].

در این فصل یک مدل مبتنی بر متن که بهترین عملکرد را در بازشناسی شئ و بر تصاویر با تعداد اشیاء زیاد داشته است را بررسی می‌کنیم و جزئیات روش را شرح می‌دهیم. همچنین مروری بر روش-هایی که از درخت تصمیم استفاده کرده اند خواهیم داشت. درنهایت نیز الگوریتم کارت که برای ساخت درخت تصمیم استفاده شده است را معرفی خواهیم کرد.

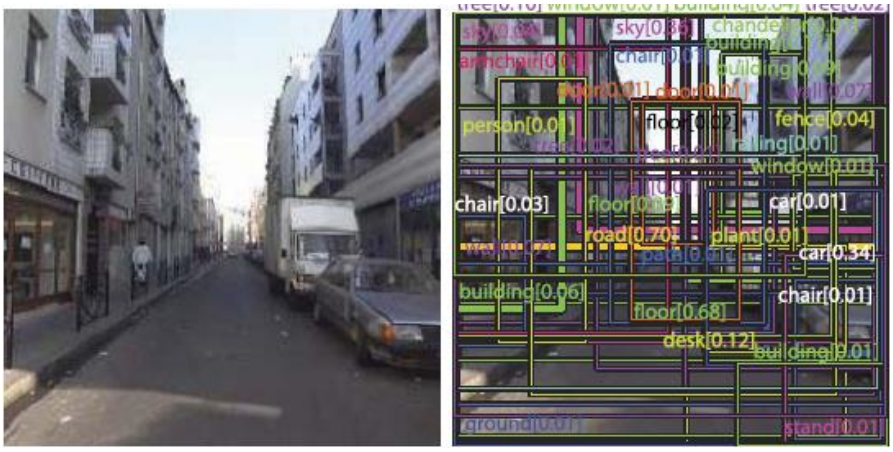
۳–۲: مدل مبتنی بر اطلاعات متنی سلسله مراتبی

با استفاده از تشخیص‌دهنده‌هایی که برای کشف یک شئ استفاده می‌شوند تنها می‌توان یک دسته از انواع شئ را استخراج کرد. برای کشف تعداد شئ بیشتر نیاز به اجرای یابنده‌های مجزا برای هر دسته شئ است. به دلیل اینکه هرکدام از این یابنده‌ها به طور مجزا کار می‌کنند، نتیجه ممکن است غلط باشد. برای بهبود عملکرد شناسایی شئ، می‌توان علاوه بر خصوصیات محلی اطلاعات متنی مانند خصوصیات سراسری یک تصویر (مانند اینکه صحنه خیابان است) و روابط بین انواع اشیاء (مانند اینکه خیابان و ماشین معمولا باهم دریک صحنه قرار دارند) را نیز استفاده کرد. همانطور که در شکل21 نشان داده شده است، با داشتن دسته‌های زیاد شئ تشخیص‌های اشتباه زیادی اتفاق می‌افتد. 6 تشخیص با اطمینان بیشتر در شکل22 نشان داده شده است که ترکیبی از اشیاء درونی و بیرونی است، درحالی

[45] Visual cognition
[46] Cognitive neuroscience
[47] Biederman

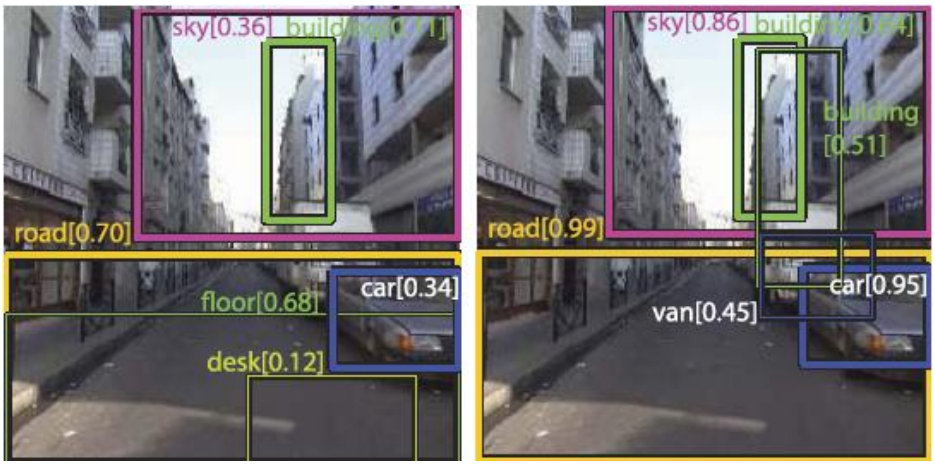
که اگر مدل متنی را نیز درنظر بگیریم احتمال کمتری برای اشیاء درونی مانند کف اطاق و میز وجود خواهد داشت. (شکل22)

اشیاء داخلی و اشیاء خارجی معمولا دریک صحنه اتفاق نمی‌افتند، در مقاله چی[۴۸] و تورالبا از یک مدل درخت که تمام اشیاء خارجی در یک زیردرخت و تمام اشیاء داخلی در یک زیردرخت دیگر هستند استفاده کرده است، که دو درخت با وزن منفی به یکدیگر متصل هستند. به طور مشابه اشیاء مرتبط با آشپزخانه مانند یخچال، ظرفشویی و غیره همه در یک زیردرخت هستند و وزن‌های مثبت دارند. با استفاده از وابستگی‌های ساختار درختی بین اشیاء می‌توان این مدل را برای بیش از صد نوع شئ استفاده کرد. با ترکیب این مدل از روابط اشیاء با خروجی‌های شناساگرهای محلی و خصوصیات سراسری تصویر می‌توان تمام نمونه‌های مختلف انواع اشیاء را شناسایی و تعیین موقعیت کرد[4]. در ادامه به شرح این روش پرداخته می شود.



شکل 1-3- تصویر ورودی و نتیجه اعمال 107 یابنده کلاس

[48] Choi



شکل 2-3 – سمت چپ: خروجی شش یابنده با اطمینان بالاتر، سمت راست: خروجی شش یابنده با استفاده از مدل متنی استفاده شده

۳-۲-۱: .مدل سازی اولیه

هر شئ را می‌توان با یک متغیر باینری که نشان‌دهنده‌ی حضور یا عدم حضور یک شئ در تصویر است نشان داد و برای نمایش محل شئ از متغیر گوسی استفاده می‌شود.

در یک درخت باینری اگر هر گره b_i نمایشگر این باشد که شئ i در تصویر وجود دارد یا خیر، احتمال توأم متغیرهای باینری طبق ساختار درخت به صورت زیر محاسبه می‌شوند:

$$p(b) = p(broot)\prod p\left(bi\middle|bpa(i)\right) \qquad (3-1)$$

که $Pa(i)$ والد گره‌ی i است. b نشان‌دهنده‌ی تمام متغیرهای مربوطه است $b \equiv \{b_i\}$. والد و فرزند ممکن است رابطه‌ی مثبت (مثل رابطه‌ی کف اطاق با دیوار) و یا منفی (مثل رابطه‌ی کف اطاق و آسمان) داشته باشند.

اشیاء در موقعیت‌های مخصوصی نسبت به یکدیگر قرار می‌گیرند. برای مثال، صفحه‌ی کامپیوتر معمولا بالای یک کیبورد و موس قرار می‌گیرد. برای ذخیره‌ی این اطلاعات مکانی می‌توان متغیرهای موقعیت را به مدل درخت اضافه کرد.

در این مدل به جای استفاده از شئ جداشده از تصویر از یک قاب محاط استفاده شده است که کوچکترین قاب حصارکننده برای تمام نقاط بخش جداشده است. از این قاب برای نمایش محل یک نمونه شئ استفاده می‌شود. فرض می کنیم l_x و l_y به ترتیب مختصات افقی و عمودی مرکز قاب محاط

و l_w و l_h به ترتیب پهنا و ارتفاع قاب هستند. همچنین فرض می‌شود ارتفاع تصویر به یک نرمال شده است و $l_x=0$، $l_y=0$ مرکز تصویر است. فاصله‌ی موردانتظار بین مراکز اشیاء به سایز آنها بستگی دارد، اگر یک کیبورد و یک موس کوچک باشند فاصله‌ی بین مراکز آنها نیز باید کوچک باشد. مدل صورت فلکی اطلاعات موقعیت را به فضای تغییرناپذیر مقیاسی تبدیل می‌کند. هویم[۴۹] و همکارانش تغییرات مقیاس را به اطلاعات 3بعدی واضح مرتبط می‌سازند[14]. طبق این روش برای تعیین موقعیت‌های شئ تبدیلات مختصات زیر در مختصات سه‌بعدی انجام می‌شود.

$$Lx = \frac{lx}{lh} Hi, \qquad Ly = \frac{ly}{lh} Hi, \qquad Lz = \frac{1}{lh} Hi, \qquad (3-2)$$

L_z فاصله‌ی بین ناظر و شئ است و H_i ارتفاع فیزیکی شئ i است. ارتفاع هر نوع شئ به صورت دستی براساس سایز اشیاء واقعی کد می شود. (مثلا شخص=1.7 متر، ماشین =1.5 متر)

موقعیت‌های ارتباطی افقی اشیاء به دلیل دیدهای مختلف از یک تصویر با تصویر دیگر به طور قابل توجهی متفاوت است و نشان داده شده است که موقعیت‌های افقی عموما اطلاعات متنی ضعیفی دارند. بنابراین L_x درنظر گرفته نشده است و تنها L_y و L_z را برای ذخیره موقعیت عمودی و وابستگی‌های مقیاسی درنظر گرفته شده است. فرض شده است که L_yS و L_zS مستقل هستند، مثلا موقعیت عمودی یک شئ مستقل از فاصله‌ی آن از سطح تصویر است. L_yS را مانند گوسین توام و L_zS با استفاده از توزیع‌های log نرمال مدل مدل شده است، زیرا همیشه دارای مقادیر مثبت هستند و اطراف مقادیر کوچک توزیع شده‌اند. یک متغیر موقعیت برای شئ i را به صورت زیر بازتعریف می‌شود.

$$Li = (Ly, \log Lz) \qquad (3\text{-}3)$$

برای سادگی روابط مکانی بین انواع شئ مدل شده است، نه بین تک تک نمونه‌ها بنابراین اگر چندین نمونه از شئ i در یک تصویر وجود داشته باشد، L_i موقعیت میانه‌ی تمام نمونه‌ها را نشان می‌دهد.

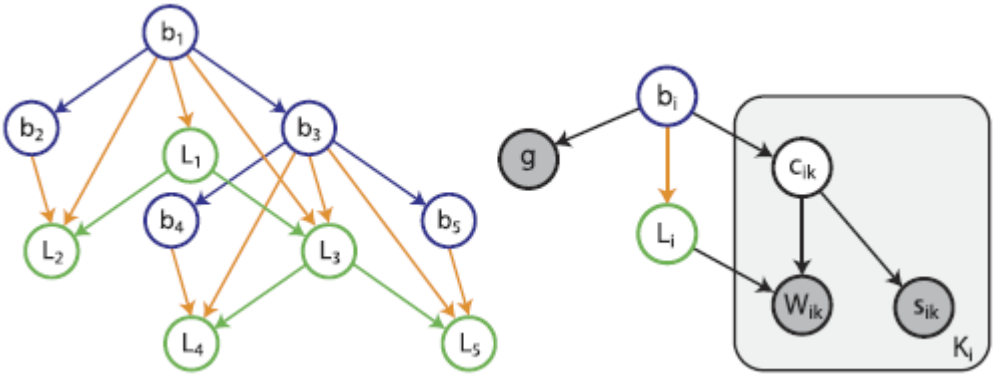
با فرض شرط بر متغیر وجود b، ساختار وابستگی L_iS ساختار درختی مشابهی با درخت باینری دارد:

[49] Hoiem

$$p(L|b) = p(L_{root}|b_{root}) \prod p\left(L_i \middle| L_{pa(i)}, b_i, b_{pa(i)}\right) \qquad (3-4)$$

که پتانسیل هر لبه توزیع موقعیت فرزند به شرط موقعیت والد و وجود یا نبودن فرزند و والد را کد می‌کند. شکل4 مدل گرافیکی مربوط به متغیرهای وجودی b_is و متغیرهای موقعیت L_is را نشان می‌دهد. با ترکیب 1 و 3 توزیع توام تمام متغیرهای باینری و گوسی به صورت زیر نشان داده می‌شود:

$$\begin{aligned} p(b, L) &= p(b)p(L|b) \\ &= p(b_{root})p(L_{root}|b_{root}) \\ &\times \prod p\left(b_i \middle| b_{pa(i)}\right) p\left(L_i \middle| L_{pa(i)}, b_i, b_{pa(i)}\right). \end{aligned} \qquad (3-5)$$



شکل 3-3 – چپ: مدل اولیه که ارتباط متغیرهای وجود شئ و متغیرهای موقیت شئ را نشان می دهد. راست: مدل اندازه گیری برای شئ i.

۳-۲-۲: مدل اندازه‌گیری

یکی کردن خصوصیات سراسری تصویر

توصیف‌گر جیست نمایش با ابعاد کم یک تصویر است، که بافت درشت و لایه‌ی مکانی یک صحنه را تسخیر می‌کند. برای استفاده از خصوصیات سراسری تصویر جیست به عنوان معیاری برای حضور هر متغیر b_i معرفی شده است. این کار به مدل امکان بدست آوردن نوع صحنه که برای پیش بینی اشیاء داخلی و خارجی مفید است را می‌دهد.

یکی کردن خروجی‌های یابنده‌ی محلی

برای بازیابی و تعیین موقعیت نمونه‌های شئ در یک تصویر ابتدا یابنده‌های بازیابی تک شئی اعمال شده است و مجموعه‌ای از پنجره‌های کاندید برای هرنوع شئ بدست آمده است. i نشان دهنده‌ی نوع شئ و k برای فهرست کردن پنجره‌های کاندید که توسط یابنده‌های پایه تولید می‌شوند استفاده می‌شود. هر خروجی یابنده یک امتیاز s_{ik} و یک جعبه محصورکننده فراهم می‌کند که تبدیل 2 برآن اعمال می‌شود تا موقعیت متغیر بدست آید $W_{ik}=(L_y, \log L_z)$ به هر پنجره یک متغیر باینری c_{ik} نسبت داده شده است تا نشان دهد آیا بازیابی درستی است $(c_{ik}=1)$ یا خیر $(c_{ik}=0)$. شکل 4 مدل اندازه گیری برای شئ i برای یکی کردن جیست و یابنده پایه در مدل پیشین را نشان می‌دهد که در آن از نشانه‌گذاری برای نمایش K_i پنجره مختلف کاندید استفاده شده است.

اگر یک پنجره کاندید به درستی شئ i را بازیابی کرده باشد $(c_{ik}=1)$ سپس موقعیت آن W_{ik} یک بردار گوسی با میانگین L_i موقعیت شئ i را نشان می‌دهد، و درغیراین صورت $(c_{ik}=0)$ W_{ik} مستقل از L_i است و یک توزیع یکنواخت دارد.

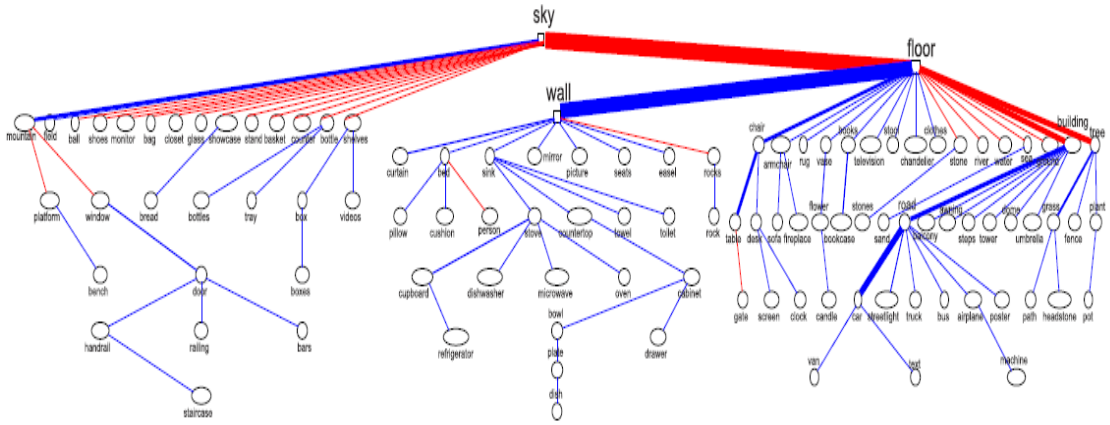
۳-۲-۳: یادگیری ساختار وابستگی شئ

برای یادگیری ساختار وابستگی بین اشیاء از مجموعه‌ای از تصاویر برچسب گذاری شده استفاده شده است. یادگیری یک مدل درخت توسط الگوریتم ساده و کارآمد چو-لو[50] امکان پذیر است[15]. این الگوریتم ابتدا اطلاعات متقابل تجربی مقادیر نمونه تمام جفت متغیرها را محاسبه می‌کند. سپس ماکسیموم وزن درخت پوشا که وزن یالها برابر با اطلاعات متقابل بین متغیرهایی که توسط یال به یکدیگر متصل شده‌اند، است را پیدا می‌کند. ساختار درخت با استفاده از نمونه‌های b_{is} در یک مجموعه از تصاویر برچسب‌گذاری شده یادگرفته می‌شود. حتی باوجود بیش از 100 شئ و هزاران تصویر آموزشی، یک مدل درخت در چند ثانیه در متلب یادگرفته می‌شود.

شکل زیر ساختار درخت یادگرفته شده از دیتاست سان 09 را نشان می‌دهد. به‌دلیل اینکه خروجی الگوریتم چو-لو یک درخت غیرجهت دار است، sky را به عنوان ریشه درخت انتخاب کرده است تا یک ساختار درختی جهت دار بدست آید. نکته موردتوجه این است که از اطلاعات مربوط به ساختار سلسله مراتبی اصلی بین انواع اشیاء در حین فرآیند یادگیری استفاده نمی‌شود. الگوریتم چو-لو به‌سادگی وابستگی‌های دوبدوی قوی را انتخاب می‌کند. ساختار درخت یادگرفته شده اشیاء را در یک سلسله‌مراتب

[50] Chow-liu

طبیعی سازماندهی می‌کند. برای مثال یک زیردرخت با ریشه‌ی ساختمان اشیاء زیادی دارد که در صحنه خیابان قرار دارند، و یک زیردرخت با ریشه‌ی sink شامل اشیایی است که به طور متداول در آشپزخانه وجود دارند. بنابراین بسیاری از گره‌های غیربرگ به همانند ابرشئ[۵۱] با اندازه‌ی درشت‌تر عمل می‌کنند. به عبارت دیگر ساختار درخت یادگرفته شده سلسله‌مراتب اصلی بین اشیاء و صحنه‌ها را ذخیره می‌کند. درنتیجه بازشناسی شئ و عملکرد فهم صحنه بهتر اتفاق می‌افتد.



شکل 4-3– ساختار وابستگی اشیاء که از دیتاست SUN 09 یادگرفته شده است. یال های قرمز نشان دهنده ی رابطه ی منفی بین طبقه ها است. کلفتی هریال قدرت آن رابطه را نشان می دهد[16]

۳-۲-۴: یادگیری پارامترهای مدل

برای یادگیری پارامترهای مدل پیشین از برچسب های صحیح تصاویر آموزشی استفاده می شود. برای محاسبه $p(b_i|b_{pa(i)})$ همرخداد جفت شئ های والد و فرزند به سادگی شمارش می شود. برای هر جفت والد و فرزند از سه توزیع گوسی مختلف برای $p(L_i|L_{pa(i)}, b_i, b_{pa(i)})$ استفاده می شود. زمانیکه هردو شئ حضور داشته باشند ($b_i=1$, $b_{pa(i)}=1$) در این صورت موقعیت شئ فرزند L_i به موقعیت والد $L_{pa(i)}$ بستگی دارد. زمانیکه شئ وجود داشته باشد ولی والد آن وجود نداشته باشد موقعیت فرزند مستقل از والد است. زمانیکه شئ وجود نداشته باشد، فرض می کنیم L_i مستقل از تمام موقعیت های اشیاء دیگر است و میانگین آن میانگین موقعیت شئ i در تمام تصاویر است.

در مدل اندازه گیری، $p(g|b_{(i)})$ را می توان توسط توصیفگرهای جیست که از هر تصویر آموزشی محاسبه می شوند، آموزش داد. برای یادگیری بقیه ی پارامترها در مدل اندازه گیری از یابنده های

[51] Meta-object

محلی برای هر نوع شئ در تصاویر آموزشی استفاده می شود. امتیازات یابنده ی محلی مرتب می شوند و s_{ik}، kامین امتیاز بالا برای طبقه i است و $p(c_{ik}|s_{ik})$ با استفاده از رگرسیون منطقی[52] یادگرفته می شود که از طریق آن می توان $p(s_{ik}|c_{ik})=p(c_{ik}|s_{ik})p(s_{ik})/p(c_{ik})$ را محاسبه کرد.

۳-۲-۵: استفاده از مدل

با داشتن جیست (g) و موقعیت های کاندید پنجره ها (W) و امتیازات آنها (S)، حضور اشیاء (b)، بازیابی صحیح (c) و موقعیت های موردانتظار تمام اشیاء (L) با حل مساله بهینه سازی زیر بدست می‌آیند:

$$b, c, L = \arg max_{b,c,L}\, p(b, c, L|g, W, s). \qquad (3 - 6)$$

در مرحله ی اول اطلاعات موقعیت W درنظر گرفته نمی شود و تخمین MAP متغیرهای b و c براساس جیست تصویر و امتیازات پنجره های کاندید انجام می شود:

$$b, c = \arg max_{b,c}\, p(b, c|s, g). \qquad (3 - 7)$$

سپس براساس این تخمین ها موقعیت های موردانتظار اشیاء به صورت زیر با استفاده از درخت گوسی به دست می آیند:

$$L = \arg max_{L}\, p(\, L|b, c, W). \qquad (3 - 8)$$

سپس براساس تخمین های موقعیت حضور و متغیرهای یابنده دوباره تخمین زده می شوند:

$$b, c = \arg max_{b,c}\, p(b, c|s, g, L, W) = \arg max_{b,c}\, p(b, c|s, g)p(L, W|b, c), \qquad (3 - 9)$$

در این گام جفت اشیاء یا پنجره ها در موقعیت های محتمل در تصویر برای حضور بررسی می شوند.

در گام نهایی احتمال حاشیه ای هر متغیر وجود و احتمال حاشیه ای هرمتغیر یابنده برای تعیین موقعیت شئ به صورت زیر محاسبه می شود:

[52] Logistic regression

$$p(c_{ik} = 1|s, g, L, W) \qquad (3-10)$$

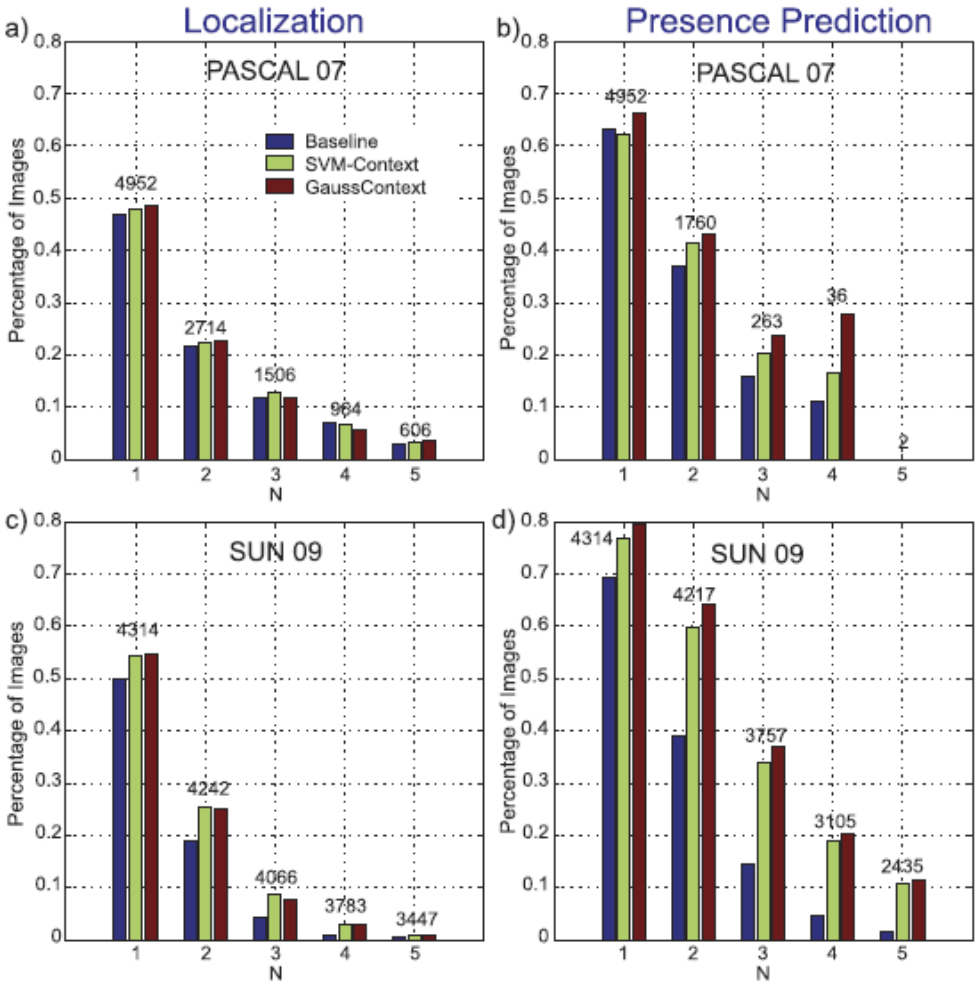
$$p(b_i = 1|s, g, L, W) \qquad (3-11)$$

۳-۲-۶: نتایج

برای آموزش این مدل از 4367 تصویر و برای تست از 4317 تصویر از مجموعه داده سان09 استفاده شده است. برای استفاده از نمونه‌های آموزشی کافی برای یابنده‌های پایه 26000 تصویر اضافی دیگر نیز تفسیر شده و برای آموزش یابنده‌های پایه استفاده شده اند. این تصاویر شامل یک شئ هستند و برای یادگیری مدل متنی استفاده نشده اند.

این مجموع تصاویر از مجموعه داده سان دارای بیش از 200 شئ است اما یابنده‌های پایه برای بعضی از اشیاء حتی با تفسیرهای بیشتر نیز کیفیت پایینی ارائه می‌دهند. به دلیل اینکه مدل متنی از خروجی‌های یابنده‌های پایه استفاده می‌کند و احتمال بازیابی صحیح را برای هرپنجره بدست می‌آورد، اگر کاندیدی برای بعضی از نمونه اشیاء توسط یابنده پایه وجود نداشته باشد نمی‌تواند بازیابی انجام دهد. بنابراین کلاس‌های شئ که یابنده پایه نتوانسته حداقل 4 کاندید صحیح در کل مجموعه داده برای آن تولید کند حذف شده اند و از 107 کلاس شئ دیگر استفاده شده است.

در شکل زیر عملکرد تعیین موقعیت و پیش‌بینی حضور شئ این مدل بر مجموعه داده سان و پاسکال نشان داده شده است و با دو الگوریتم یابنده پایه[10] و ماشین بردار پشیبان متنی[10] مقایسه شده است. برای این کار N تا از بازیابی‌های با اطمینان بیشتر در هر تصویر بررسی می‌شوند که آیا درست هستند یا خیر. اعداد روی هر ستون نشان دهنده‌ی تعداد تصاویر که شامل حداقل N نمونه شئ برای تعیین موقعیت و N طبقه شئ برای پیش‌بینی حضور هستند، می‌باشد. همانطور که دیده می‌شود با بزرگتر شدن N تعداد تصاویر در پاسکال بسیار کم می‌شود، زیرا بیشتر تصاویر در این مجموعه داده تنها شامل یک یا دو طبقه شئ هستند.

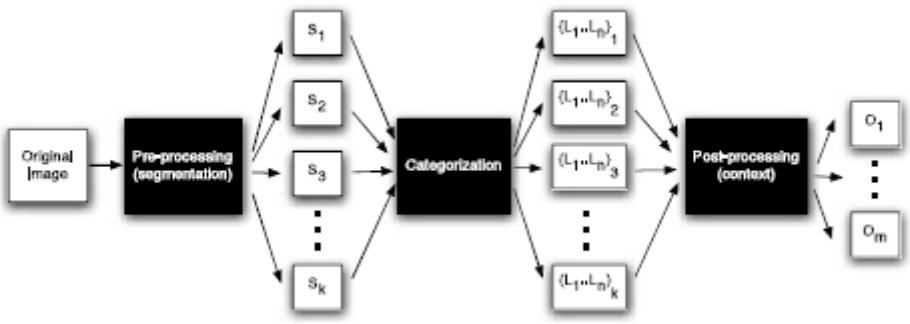


شکل 5-3- عملکرد تعیین موقعیت و پیش‌بینی حضور بر مجموعه داده سان 9 و پاسکال7 a و c درصد تصاویری که N تا از بیشترین بازیابی‌های با اطمینان بیشتر صحیح بوده اند. اعداد روی هر ستون تعداد تصاویری که حداقل شامل N نمونه شئ بوده اند، را نشان می‌دهد. B و d درصد تصاویری که N تا از پیش‌بینی حضورها با بیشترین احتمال درست بوده‌اند. اعداد نوشته شده بر هر ستون نشان دهنده‌ی تعداد تصاویری که حداقل شامل N طبقه شئ مختلف بوده‌اند را نشان می‌دهد.

پیش‌بینی اینکه چه اشیائی در تصویر وجود دارند در درک محتوای صحنه موثر است و می‌تواند برای پرسش در مورد تصاویری که شامل چند شئ مشخص هستند استفاده شود. نتایج بیشتر این مدل در فصل بعد نشان داده می‌شوند و نتایج الگوریتم پیشنهادی نیز با بخش عملکرد پیش بینی حضور مقایسه خواهد شد.

۳–۳: مدل مبنی بر اطلاعات متنی معنایی

در مقاله [17] گامی برای استفاده از اطلاعات متنی معنایی به عنوان پس پردازش در هر مدل بازشناسی شئ ارائه شده است. در این روش توسط یک چهارچوب میدان تصادفی شرطی[۵۳]، سازگاری برچسب اشیاء طبق ارتباطات متنی بیشینه می شود. مدل کیف ویژگی[۵۴] توسط استفاده از تعاملات متنی بین اشیاء، توسعه داده شده است. به عنوان گام پیش پردازش ابتدا بخش بندی انجام شده است که این کار باعث دسته بند مکانی در مدل بازشناسی می شود. فلوچارت این روش در شکل زیر نشان داده شده است.



شکل 3-6- طبقه بندی شئ با استفاده از اطلاعات متنی معنایی. $S_1 \dots S_k$ مجموعه k بخش یک تصویر که توسط چندین بخش بندی بدست آمده است را نشان می دهد. $L_1 \dots L_n$ لیست n برچسب هر بخش را نشان می دهد. $O_1 \dots O_m$ نیز مجموعه m طبقه شئ در تصویر اصلی است.

برای استفاده از اطلاعات متنی معنایی در طبقه بندی شئ از چهارچوب میدان تصادفی شرطی استفاده شده است که در ای روش دو تفاوت عمده دارد. اول اینکه بین برچسب بخش ها یک گراف کاملا متصل استفاده شده است و دوم اینکه به جای یکی کردن مدل متنی با مدل طبقه بندی، میدان تصادفی شرطی بر مسائل ساده تر بر تعداد کمتری بخش آموزش داده شده است.

با داشتن تصویر I و بخش های $S_1 \dots S_k$ هدف یافتن برچسب های $C_1 \dots C_k$ است به گونه ای که این برچسب ها متناسب با محتوای بخش ها باشند و نسبت به یکدیگر تناسب متنی داشته باشند. فرض شده است که برچسب ها از یک مجموعه محدود C هستند.

[53] Conditional Random Field (CRF)
[54] Bag-of-features (BoF)

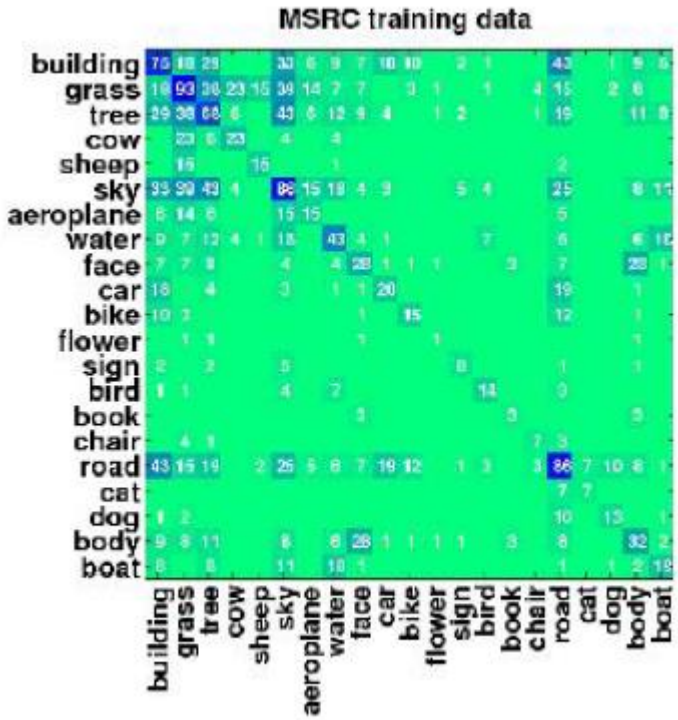
این روابط به عنوان توزیع احتمالی به صورت زیر مدل می شوند:

$$p(C_1 \dots C_k|S_1 \dots S_k) = \frac{B(C_1 \dots C_k)\prod A(i)}{Z(\phi, S_1 \dots S_k)} \qquad (3-12)$$

$$\mathrm{A(i)} = \mathrm{p}(C_i|S_i);\ \ \mathrm{B}(C_1 \dots C_k) = \exp\Big(\sum_{i,j=1}^{k} \phi(C_i, C_j)\Big) \qquad (3-13)$$

$Z(0)$ تابع بخش بندی است. بخش حاشیه ای $\mathrm{p}(C|\mathrm{S})$ توسط سیستم بازشناسی فراهم می شود.

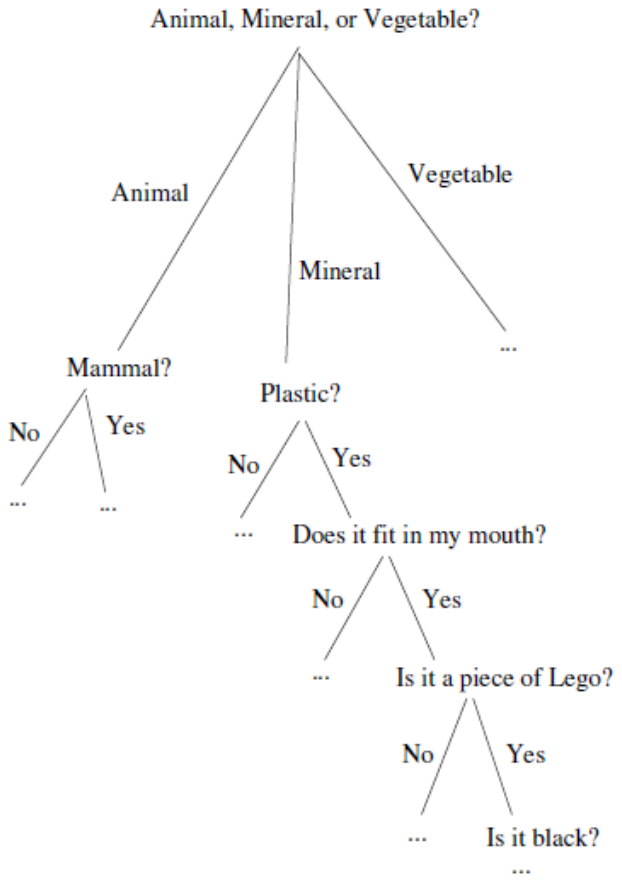
برای ترکیب اطلاعات متنی معنایی با بازشناسی شئ و در چهارچوب میدان تصادفی شرطی، ماتریس‌هایی ساخته می شوند. این ماتریس ها متقارن و غیرمنفی هستند که شامل فرکانس همرخداد برچسب های اشیاء در مجموعه آموزشی مجموعه داده هستند. به عنوان مثال ماتریس زیر از مجموعه داده MSRC بدست آمده است.



شکل 7-3- ماتریس متنی. فرکانس همرخدادی اشیاء برای مجموعه داده MSRC

۳–۴: استفاده از درخت تصمیم در بازشناسی شئ

روشی که در مقاله بلک وال و آستین[55] استفاده شده است، با این فرض است که شئ از تصویر جدا شده است و فقط باید تشخیص داده شود که چه شئی است. بخش کوچکی از یک نمونه درخت تصمیم در شکل زیر نشان داده شده است. هر شئ در دیتابیس دارای مقادیری برای ویژگی های مختلف است، توسط بخشی از این دیتابیس درخت تصمیم ساخته می شود و توسط بقیه داده ها تست می شود. روش متداول، تعریف مجموعه مناسبی از ویژگی ها برای دیتابیس است. هدف در این مقاله توسعه ی ویژگی های جدید نبوده است و ویژگی های نسبتا ساده و شناخته شده استفاده شده اند[18].



شکل 8-3- بخش کوچکی از یک نمونه درخت تصمیم[18]

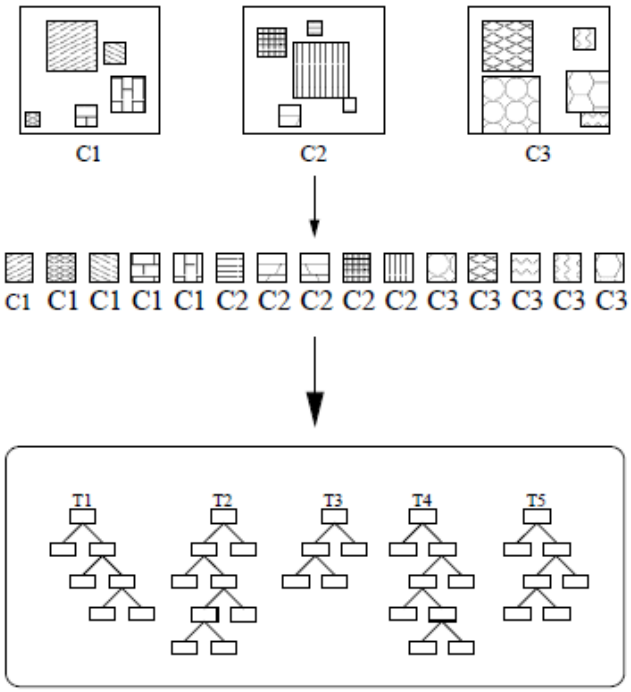
در مقاله ولیکینگ وروفر[56] فرآیند بازشناسی روبات که برمبنای کلاسبندی درخت تصمیم است، ارائه می شود. با انجام پیش پردازش هایی مانند بخش بندی برمبنای رنگ، بازیابی مرز و تولید خط ویژگی

[55] Blackwell, P & Austin, D
[56] D. Wilking, T. Röfer

های مختلف ارائه می شوند. (به عنوان مثال مساحت، محیط و زاویه همسایگی) برای ساخت درخت ویژگی با بیشترین آنتروپی محاسبه می شود. با هرس درخت می‌توان از فرایادگیری آن جلوگیری کرد[19].

در مقاله کی، وانگ و همکارانشان[۵۷] روشی برای بازشناسی شئ ارائه شده است که یادگیری آدابوست[۵۸] را با استفاده از روش درخت تصمیم و یک معماری سلسله مراتبی جدید با ویژگی های محلی هیستوگرام ترکیب می کند[20].

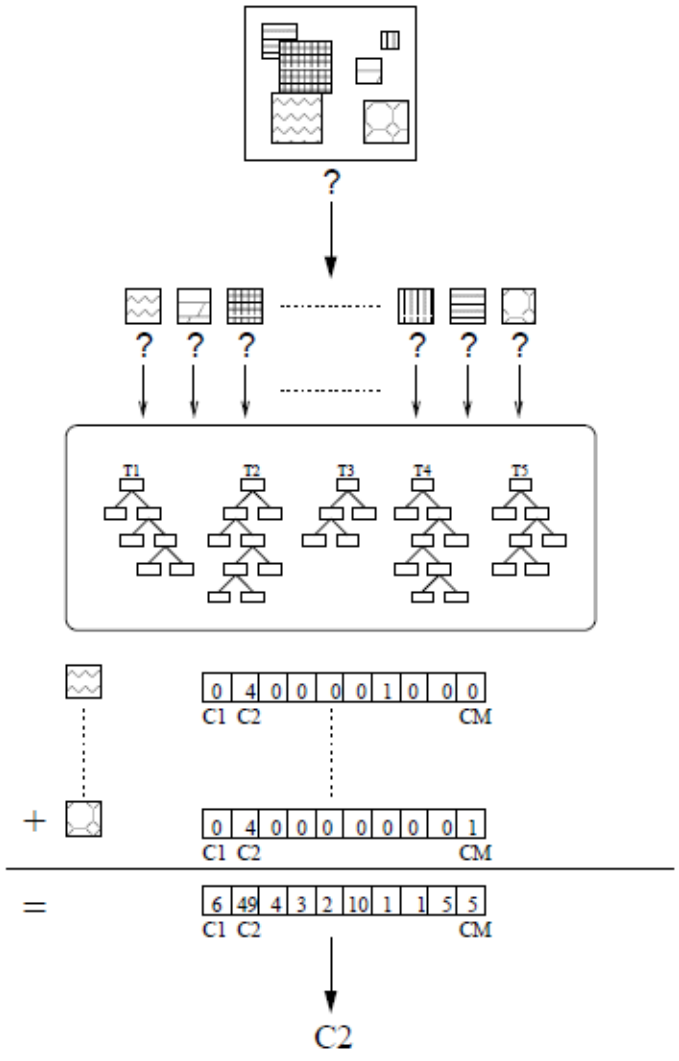
مقاله مری و همکارانش[۵۹] براساس استخراج رندم زیرپنجره هایی به صورت تکه های مربعی و کلاسبندی آنها توسط درخت تصمیم است. زیرپنجره ها از موقعیت هایی به صورت رندم و با سایز رندم انتخاب می شوند و پس از تغییر سایز درخت تصمیم براساس آنها ساخته می شود. این مراحل در شکل زیر نشان داده شده است[21].



شکل9-3 - فرآیند یادگیری: استخراج رندم زیرپنجره ها از تصاویر آموزشی، تغییر سایز آنها و سپس ساخت درختان تصمیم[21]

[57] Qi, wang et al..
[58] Adaboost
[59] R Marée, P Geurts, J Piater, L Wehenkel.

سپس برای کلاسبندی یک تصویر تست، هر درخت تصمیم احتمال شرطی برای هر زیرپنجره تخمین می زند. تمام تخمین ها متوسط گرفته می شوند و سپس کلاس متعلق به بیشترین احتمال به تصویر تخصیص می یابد. این مراحل در شکل زیر نشان داده شده اند. در اینجا 5 درخت آموزش داده شده است. آموزش یک درخت نیز حالت خاصی از این مساله خواهد بود.



شکل 10-3- فرآیند بازشناسی: زیرپنجره هایی که به صورت رندم استخراج شده اند بین درخت ها منتشر می شوند و پس از جمع آوری رای ها کلاس با رای اکثریت به تصویر تخصیص می یابد[21].

در مقاله ی ماتاس [60] حافظه ی بصری به عنوان درخت تصمیم باینری سازماندهی می شود که زمان تصمیم میانگین را کمینه کند. برگ های این درخت تصمیم بعضی از نواحی محلی را نمایش

[60] Obdrzalek, S. and J. Matas

می‌دهند و هر گره غیره برگ مربوط به یک کلاسبند ضعیف است. در فاز بازشناسی، یک معیار ثابت مشخص می کند که یک بخش تصویر در کدام زیردرخت جستجو شود[22].

هر تکنیکی که ورودی بصری جاری را یکی یکی با مدل های ذخیره شده مقایسه کند خطی است. تکینیک هایی که مساله ی آنها تطبیق دو تصویر است، زمان پاسخ قابل قبولی ندارند. جستجو و اندیس گذاری موضوعاتی هستند که مطالعات بسیاری در رابطه با آنها انجام شده است و دو روش درهم سازی و جستجوی درختی فراگیرتر هستند.

۳-۵: معرفی الگوریتم کارت

این الگوریتم سرانجام درخت کلاسبندی است که انقلابی در زمینه‌ی آنالیز پیشرفته ایجاد کرده است و یکی از مهم‌ترین ابزارهای داده‌کاوی مدرن است[23]. یک درخت کارت یک درخت تصمیم باینری است که با تقسیم هر گره به دو گره فرزند به طور مکرر ایجاد می‌شود. گره ریشه شامل تمام نمونه‌های آموزشی است. متغیرهای مختلف در جدول زیر معرفی شده‌اند[24].

جدول 1-3- تعریف متغیرهای استفاده شده در تعریف الگوریتم ساخت درخت تصمیم

نام	تعریف
Y	متغیر هدف که می‌تواند مقداری مطلق وصفی یا اسمی و یا مقدار پیوسته داشته باشد. در صورتی که شامل j کلاس باشد یکی از مقادیر $C=\{1,\ldots,j\}$ را خواهد داشت.
X_m , m=1,..,M	مجموعه‌ی متغیرهای مورداستفاده برای پیشگویی که می‌تواند مقادیر مطلق وصفی یا اسمی و یا مقادیر پیوسته داشته باشد.
h= $\{X_n, Y_n\}$ n=1,..,N	کل نمونه‌های یادگیری
h(t)	نمونه‌های یادگیری که در گره t قرار می‌گیرند.
W_n	وزن مورد n

f_n	وزن فرکانس مربوط به مورد n
$\pi(j)$, j=1,…,J	احتمال پیشین y=j ،j=1,…,J
$p(j,t)$	احتمال اینکه یک مورد هم در کلاس j باشد و هم از گره tعبور کند.
$\mathbf{p(t)}$	احتمال عبور از گره t
$p(j\|t)$	احتمال اینکه یک مورد از کلاس j باشد به شرطی که بدانیم از گره t عبور می کند.
$C(i\|j)$	هزینه کلاسبندی اشتباه کلاس j به عنوان کلاس i درنتیجه C(j | j) = 0

برای ساخت درخت باید درمورد اینکه چگونه در هر گره داده‌ها تقسیم شود، تصمیم‌گیری شود. این تقسیم‌بندی از بین تمام تقسیم‌بندی‌ها باید به گونه‌ای انجام شود که گره‌های فرزند به خالص‌ترین حالت برسند. در هر گره تقسیم‌بندی به مقدار یک متغیر وابسته است. در صورتی که X یک مقدار پیوسته با K مقدار مختلف داشته باشد، K-1 تقسیم بندی مختلف بر X می‌تواند وجود داشته باشد. یک درخت با گره‌ی ریشه شروع می‌شود و با تکرار گام‌های زیر ایجاد می‌شود.

روش الگوریتم کارت

گام 1. یافتن بهترین تقسیم‌بندی هر متغیر

برای هر متغیر X، تمام تقسیم‌بندی‌های ممکن امتحان می‌شود.

گام 2. پیدا کردن بهترین تقسیم‌بندی گره

انتخاب تقسیم‌بندی که معیار مشخص شده را بیشینه می‌کند، دربین بهترین تقسیم‌بندی‌های یافت شده در گام 1

گام 3. انجام تقسیم‌بندی یافت شده در گام دوم درصورتی‌که شرط پایان برقرار نیست

۳-۵-۱: معیار تقسیم‌بندی و مقیاس خالص بودن

در گره‌ی t بهترین تقسیم‌بندی s به گونه‌ای انتخاب می‌شود که معیار تقسیم $\Delta i(s,t)$ را بیشینه کند. در گره t مقادیر $p(j,t)$، $p(t)$ و $p(j|t)$ به صورت زیر بدست می‌آیند:

$$p(j,t) = \frac{\pi(j) N_{w,j}(t)}{N_{w,j}}, \qquad (3\text{-}14)$$

$$p(t) = \sum_j p(j,t), \qquad (3\text{-}15)$$

$$p(j|t) = \frac{p(j,t)}{p(t)} = \frac{p(j,t)}{\sum_j p(j,t)} . \quad (3-16)$$

که $N_{w,j}$ و $N_{w,j}(t)$ به صورت زیر تعریف می‌شوند:

$$N_{w,j} = \sum_{n \epsilon h} (w_n f_n I(y_n = j)) \qquad (3-17)$$

$$N_{w,j}(t) = \sum_{n \epsilon h(t)} (w_n f_n I(y_n = j)) \qquad (3-18)$$

تابع $I(a=b)$ تابع شاخصی است که در صورتی که a=b باشد مقدار 1 و درغیر این صورت دارای مقدار صفر است.

در صورتی که Y پیوسته باشد، معیار تقسیم زیر مورد استفاده قرار می‌گیرد. معیار ناخالصی نیز همانطور که در معادله 7 نشان داده شده است، با مربع کمینه‌ی انحراف معیار[۶۱] تعیین می‌شود.

$$\Delta i(s,t) = i(t) - p_L i(t_L) - p_R i(t_R) \qquad (3\text{-}19)$$

$$i(t) = \frac{\sum_{n \epsilon h(t)} (w_n f_n (y_n - y(t))^2)}{\sum_{n \epsilon h(t)} (w_n f_n)}, \quad (3-20)$$

[61] Least Squares Deviation (LSD)

$$p_L = \frac{N_w(t_L)}{N_w(\mathrm{t})}, p_R = \frac{N_w(t_R)}{N_w(\mathrm{t})},$$

$$N_w(\mathrm{t}) = \sum_{n \epsilon h(t)} (w_n\, f_n\,), \quad (3-21)$$

$$y(t) = \frac{\sum_{n \epsilon h(t)} (w_n\, f_n y_n)}{N_w(\mathrm{t})} \quad (3-22)$$

۳-۵-۲: قوانین پایان

کنترل فرآیند رشد درخت تصمیم توسط قوانین پایان کنترل می‌شود. قوانین پایان عبارتند از:

- اگر یک گره خالص شود، یعنی تمام موارد در یک گره مقدار یکسانی از تابع هدف را داشته باشند.
- اگر تمام موارد یک گره مقادیر یکسانی برای تمام متغیرهای ورودی داشته باشند.
- اگر کاربر عمق ماکسیمم را برای درخت مشخص کرده باشد و به آن عمق رسیده باشد.
- اگر کاربر مینیمم سایز یک گره را مشخص کرده باشد و سایز یک گره کمتر از آن باشد.
- اگر تقسیم‌بندی یک گره باعث ایجاد فرزندی می‌شود که سایز آن کمتر از مقدار مینیمم مشخص شده توسط کاربر باشد.
- اگر بهترین تقسیم بندی در گره t، باعث بهبودی شود که این بهبود از حداقل بهبود تعیین شده توسط کاربر کمتر باشد.

۳-۶: خلاصه و نتیجه گیری

بازشناسی شئ یکی از مسائل پیچیده در بینائی کامپیوتر است. بخشی از این پیچیدگی به دلیل ابهامات در ویژگی های سطح پایین است که به دلیل شلوغی زمینه، موانع و غیره ایجاد می شود. مطالعات اخیر نشان داده است که مدل کردن روابط متنی می تواند به برطرف کردن این پیچیدگی ها کمک کند و عملکرد بازشناسی شئ را بهبود بخشد[25]. در این فصل چند مورد الگوریتم‌های بازشناسی شئ بررسی شدند و جزئیات و مراحل آنها شرح داده شدند. در فصل بعد الگوریتم پیشنهادی بر مبنای این الگوریتم‌ها و با بهبود عملکرد ارائه می‌شود.

فصل ۴:

بررسی و تحلیل الگوریتم پیشنهادی

۴-۱: مقدمه

در این فصل به بررسی الگوریتم پیشنهادی پایان نامه پرداخته می شود. در این الگوریتم تلاش می‌شود ویژگی های متنی استخراج شود و توسط ساخت کلاسبند با توجه به ارتباطات متنی دوبه‌دوی اشیاء و با توجه به تعداد اشیاء در هر تصویر، دقت بازیابی شئ افزایش یابد.

با توجه به مطالب مطرح شده در فصل قبل چند اشکال اساسی در روش های بازیابی شئ و روش هایی که مبتنی بر متن هستند وجود دارد. در ادامه این فصل ابتدا به بررسی چالش هایی که در الگوریتم های پیشین که به تفصیل در فصل دوم بیان شد، می پردازیم و دلایل لزوم الگوریتم پیشنهادی را بررسی می کنیم. سپس روشی ارائه می‌دهیم تا از مزایای این روشها سود ببرد به طوری که اشکالات آنها را نیز دارا نباشد. سپس الگوریتم به صورت مرحله به مرحله تحلیل و بررسی می شود.

۴-۲: چالشهای موجود در روش های قبلی

در این بخش به چالش‌های الگوریتم‌های بازیابی شئ می پردازیم و اینکه چگونه روش‌های متنی می‌توانند این چالش‌ها را برطرف کنند و همچنین به طور ویژه چالش‌های مربوط به الگوریتم‌های بازیابی شئ مبتنی بر متن را بررسی می‌کنیم. سپس بررسی می‌کنیم که الگوریتم پیشنهادی چگونه سعی در برطرف کردن این چالش‌ها دارد.

۴-۲-۱: چالش‌های روش‌های بازیابی شئ غیرمتنی

در این قسمت چالش هایی که معمولا الگوریتم های بازیابی شئ با آن ها روبه رو هستند، مطرح می‌شود و راه حل برخورد با آنها بررسی می شود. برخی از این چالشها عبارتند از:

- مبهم بودن و ناآشنا بودن شئ
- کوچک بودن شئ موردنظر
- کافی نبودن ویژگی‌های محلی شئ
- رزولوشن پایین
- وجود مانع در مقابل شئ

شکل زیر چند نمونه از تصاویری را نشان می‌دهد که دارای این دشواری‌ها هستند. در این شکل تصویر اول بیانگر چالش اول است و همان‌طور که در تصویر مشاهده می‌شود به دلیل ناآشنا بودن شئ نمی‌توان توسط روش‌های متداول این شئ را بازیابی کرد.

تصویر دوم نشان دهنده چالش دوم می‌باشد و همان‌طور که مشاهده می‌شود کوچک بودن شئ باعث شده تا شئ موردنظر به اشتباه بازیابی شود و نمی‌توان توسط روش‌های متداول این مشکل را برطرف نمود. تصویر سوم نیز چالش سوم را دربرمی‌گیرد که ویژگی‌های محلی برای شناسایی کافی نبوده اند و بازیابی به درستی صورت نگرفته است.

شکل 1-4 - برخی دشواری‌هایی که معمولا در بازیابی شئ با آنها روبرو هستیم. تصویر اول بیانگر این مساله است که بعضی اشیاء که دارای ظاهری ناآشنا و مبهم هستند که بازیابی آنها توسط روش‌های متداول امکان پذیر نیست. تصویر دوم این واقعیت را بیان می‌کند که بعضی اشیاء کوچک هستند و به

همین دلیل نمی توان آنها را به درستی بازیابی کرد. تصویر سوم نیز ناکافی بودن ویژگی‌های محلی در بعضی تصاویر را نشان می‌دهد که باعث بازیابی خطا می‌شود.

همانطور که دیده می‌شود این چالش‌ها را نمی‌توان توسط اطلاعات شئ موردنظر برطرف نمود و نیاز به استفاده از اطلاعات دیگر که همان اطلاعات متنی هستند، روشن می‌شود. در روش‌های متنی توسط اطلاعات مربوط به اشیاء دیگر تصویر می‌توان شئ موردنظر را بازیابی نمود.

۴-۲-۲: چالش‌های روش‌های بازیابی شئ مبتنی براطلاعات متنی

در روش‌های مبتنی بر اطلاعات متنی نیز چالش‌هایی وجود دارد که در این پایان نامه تلاش شده است تا بعضی از این چالش‌ها برطرف شود و دقت بازیابی شئ افزایش یابد. در ادامه به این چالش‌ها پرداخته می‌شود.

مهم‌ترین نکته در روش‌های مبتنی برمتن این است که این روش‌ها باید بر تصاویری که دارای اشیاء مختلف و از صحنه‌های مختلف هستند تست شوند تا روابط اشیاء و صحنه بتوانند به بهبود فرآیند بازشناسی شئ کمک کنند. روش‌های متنی بسیاری ارائه شده‌اند اما به دلیل اینکه بر تصاویر با تعداد شئ کم تست شده‌اند، مزیت روش‌های متنی خود را نشان نداده است.

مقاله مبتنی براطلاعات سلسله مراتبی [4] که در فصل قبل به تفصیل بررسی شد و جزء بهترین الگوریتم‌هایی است که تاکنون ارائه شده است، و برای تست این الگوریتم از مجموعه داده سان با اشیاء مختلف و صحنه‌های متنوع استفاده شده است. همان طور که در فصل قبل توضیح داده شد، در این مدل ارتباط بین اشیاء را به صورت سلسله مراتبی ذخیره می‌کند. به عنوان مثال بیشتر اشیائی که معمولا در یک آشپزخانه وجود دارند، در زیردرخت گره‌ی sink قرار می‌گیرند و تمام وسایل نقلیه نیز در زیردرخت شئ جاده قرار می‌گیرند. مساله‌ای که در این مدل وجود دارد این است که ممکن است در تصاویری وسایل آشپزخانه وجود داشته باشند اما شئ sink وجود نداشته باشد، درواقع در این مدل برای بازیابی هر شئ تنها از اطلاعات مربوط به والدهای آن تا ریشه استفاده می‌شود و اشیاء دیگری که ممکن است آنها نیز موثر باشند درنظر گرفته نمی‌شود که این چالش در مدل پیشنهادی بررسی شده است.

موارد دیگری که می‌توان در در این مدل به آنها اشاره کرد این است که استفاده از روابط فیزیکی بین اشیاء مانند اینکه یک ماشین همیشه بالای یک جاده قرار دارد و موارد مشابه می‌تواند به طور قابل

توجهی پیچیدگی محاسباتی را کاهش دهد و فرآیند بازیابی شئ را بهبود بخشد که در این مقاله به آنها پرداخته نشده است. این مساله با توجه به اینکه مربوط به مدل مکانی مقاله مرجع است و در حوزه پایان نامه قرار ندارد، در این رساله بررسی نشده است و در حوزه کارهای آینده قرار می‌گیرد.

۴-۳: رویکرد کلی الگوریتم پیشنهادی

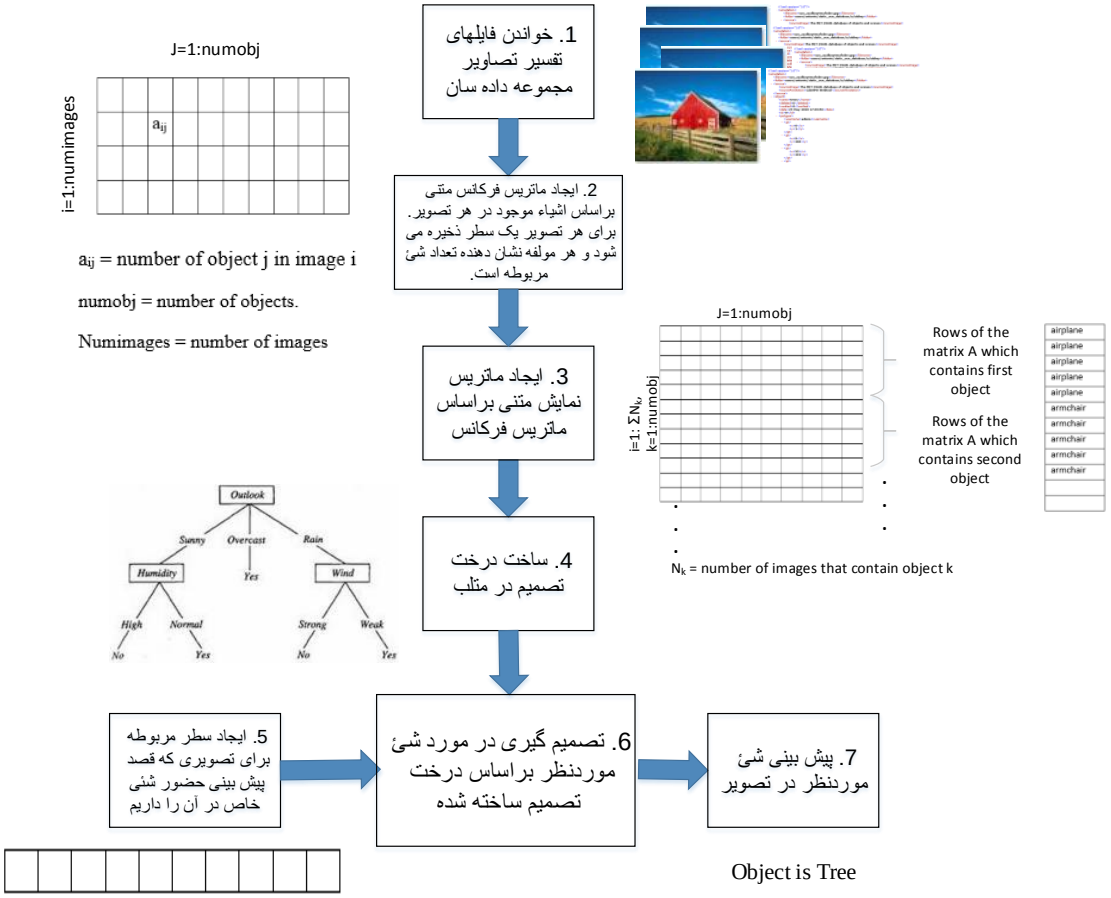
در این بخش به بیان رویکرد کلی الگوریتم پیشنهادی می پردازیم. همانطور که تاکنون توضیح داده شده است، اطلاعات متنی معنایی یک تصویر از ارزشمندترین اطلاعات متنی است و برای بازشناسی شئ مفید است. این اطلاعات شامل صحنه و اشیاء دیگر موجود در تصویر است. هدف این پایان نامه استفاده از این اطلاعات برای بازشناسی شئ است.

اطلاعات متنی موردنظر اشیاء دیگر و تعداد آنها هستند که برای ساخت کلاسبند و آموزش آن از فایلهای تفسیر تصاویر آموزشی مجموعه داده سان استفاده می‌شود. برای این کار فایلهای تفسیر موجود در مجموعه داده سان را به صورت ساختار ذخیره می‌کنیم و سپس به استخراج ویژگی پرداخته می‌شود و اطلاعات موردنیاز این فایلها استخراج و ذخیره می‌شوند.

به دلیل وجود اشیاء مختلف و صحنه های متنوع در مجموعه داده‌ی سان[16] و عدم درنظر گرفتن تمام اشیاء تصویر و تعداد آنها در مدل بازشناسی شئ مبتنی براطلاعات سلسله مراتبی [4]، عملکرد این مدل بر مجموعه داده سان چندان قابل قبول نیست. در این بخش مدلی مبتنی بر اطلاعات متنی ارائه می‌شود که برای تصاویر با اشیاء زیاد موثر باشد. اطلاعات متنی موردنظر اشیاء دیگر تصویر و فرکانس آنها می باشد. این الگوریتم شامل سه مرحله اصلی می‌باشد: اول ساخت ماتریس فرکانس متنی براساس تمام تصاویر آموزشی، دوم ساخت درخت تصمیم براساس ماتریس بدست آمده در مرحله قبل و سوم استفاده از سطر متنی هر تصویر تست برای تعیین برچسب شئ موردنظر در داده‌های تست. در تصویر زیر فلوچارت این الگوریتم نشان داده شده است.

در قسمت (1) تصاویر مختلف مجموعه داده به همراه فایل های تفسیر که توسط مجموعه داده‌ی سان فراهم شده است، نشان داده شده است. در قسمت (2) که مربوط به مرحله اول الگوریتم پیشنهادی می‌باشد، ماتریس بدست آمده از فایل های تفسیر نشان داده شده است که هر سطر از این ماتریس بردار ویژگی یک تصویر است و شامل اطلاعات متن تصویر می باشد که جزئیات آن در ادامه توضیح داده خواهد شد. برای استفاده از این ماتریس برای آموزش درخت تصمیم لازم است برای هر شئ نمونه‌های

آموزشی به طور مستقل بدست آید. به همین دلیل در قسمت (3) ماتریس نمایش متنی براساس ماتریس قبلی بدست می‌آید. چگونگی ساخت این ماتریس در بخش بعد توضیح داده خواهد شد. قسمت (4) شامل مرحله ساخت کلاسبند که یک درخت تصمیم است می‌باشد. تا این مرحله درخت تصمیم ساخته شده است. حال برای شناسایی یک شئ خاص در یک تصویر جدید با استفاده از اطلاعات متنی به این صورت عمل می‌شود: (5) بدست آوردن بردار ویژگی متنی مربوطه از تصویر موردنظر (6) کلاسبندی و تعیین کلاس شئ موردنظر با توجه به درخت تصمیم ساخته شده در مرحله قبل. در بخش بعد این مراحل با جزئیات بیشتری تشریح خواهند شد.



شکل2-4- فلوچارت الگوریتم پیشنهادی

فایل های تفسیر در واقع شامل اطلاعات مختلفی از جمله اشیاء تصویر، محل آنها و مواردی از این دست است. در بخش بعد جزئیات هر کدام از مراحل با شرح بیشتری توضیح داده خواهد شد.

۴-۴: بررسی دقیق الگوریتم پیشنهادی پایان نامه

برای این روش از بخشی از فایل های تفسیر مجموعه داده سان که شامل بیش از 8000 تصویر از بیش از 600 صحنه مختلف است، استفاده شده است. در این تصاویر بیش از 200 شئ مختلف وجود دارد که می توان آنها را توسط روش ارائه شده شناسایی کرد. هدف اثبات تاثیر کانتکست و اطلاعات متن تصویر در بازشناسی شئ است. در این روش با دانستن حضور اشیاء دیگر تصویر و فرکانس آنها می‌توان شئ موردنظر را شناسایی کرد.

3.1. ساخت ماتریس فرکانس متنی

ابتدا با استفاده از تصاویر مجموعه داده سان ماتریسی ساخته می شود که به تعداد تصاویر سطر دارد. ستون‌های این ماتریس نیز به تعداد شئ‌های موردنظر است که در این مقاله 107 شئ مشابه با مقاله مبتنی براطلاعات متنی سلسله مراتبی انتخاب شده است تا نتایج قابل مقایسه باشند. اطلاعات لازم برای ساخت این ماتریس، بعضی از اشیاء دیگر تصویر است. هر مولفه a_{ij} در این ماتریس نشان دهنده‌ی حضور شئ j در تصویر i است. برای دخالت دادن رابطه بین اشیاء از وقوع مشترک آنها استفاده شده است. درنتیجه تمام اشیاء موجود در تصویر به غیر از شئی که قرار است شناسایی شود در این ماتریس مقدار خواهند گرفت. شئی که قرار است شناسایی شود نیز دارای مقدار نخواهد بود و به عنوان برچسب برای سطر موردنظر ذخیره می شود. در تصویر زیر بخشی از این ماتریس برای چهار تصویر نشان داده شده است. در تصویر اول شئی که باید بازیابی شود شئ grass است و شئ person و floor و table در این تصویر وجود ندارد و شئ flower نیز یک نمونه در این تصویر وجود دارد. به همین ترتیب تصویر دوم دارای هیچ یک از چهارشئ grass، flower، floor و table نیست و هدف پیش‌بینی حضور شئ person است. برای هر تصویر سطری با تعداد 107 مولفه که نشان دهنده‌ی مولفه مربوطه است ذخیره می‌شود.

	grass	person	flower	floor	table
	NaN	0	1	0	0
	0	NaN	0	0	0
	0	0	4	NaN	0
	NaN	0	5	0	1

شکل3-4 - یک مثال برای چگونگی ذخیره ماتریس فرکانس متنی

تعداد اشیاء توسط متغیر numobj و تعداد تصاویر توسط متغیر numimages ذخیره می‌شوند. حال اشیائی که در هر سطر می خواهیم شناسایی کنیم را نیز به صورت جدا در برداری به شکل زیر ذخیره می کنیم. از این ماتریس و بردار زیر که براساس همه تصاویر آموزشی و اشیاء ساخته می شود، برای ساخت درخت تصمیم استفاده می کنیم.

grass
person
floor
grass

شکل4-4- برداری که شامل اشیائی هست که می خواهیم شناسایی شود و برای آموزش به درخت تصمیم داده می شود.

به این ترتیب در صورتی که این ماتریس و بردار را برای تصاویر بیشتری از مجموعه داده‌ی سان بسازیم، ماتریسی با اندازه (تعداد اشیاء)*(تعداد تصاویر) خواهیم داشت. حال از این ماتریس برای ساخت ماتریس دوم استفاده می‌کنیم. برای هر شئ تمام سطرهای مربوط به تصاویر شامل آن شئ به ماتریس جدید اضافه می‌شود و در سطرهای موردنظر برای آموزش هر شئ، مولفه‌ی مربوط به آن مقدار NaN خواهد گرفت. به این ترتیب هر تصویر به تعداد اشیاء موجود در آن در این ماتریس تکرار خواهد شد، اما هر بار برای آموزش بازشناسی یکی از اشیاء موجود در تصویر در این ماتریس قرار می‌گیرد. در نهایت ماتریس بدست آمده را به عنوان نمونه آموزشی و برچسب اشیاء را به عنوان برچسب این نمونه های آموزشی برای ساخت درخت تصمیم استفاده می‌کنیم.

3.2. ساخت درخت تصمیم

استفاده از درخت تصمیم دو مزیت عمده برای بازشناسی شئ به همراه دارد، اول اینکه برای دیتاست-های بزرگ نیز قابل استفاده است و دوم اینکه ویژگی های موثر را انتخاب می کند و برای کلاسبندی یک نمونه تست نیازی به محاسبه‌ی تمام ویژگی ها نیست.

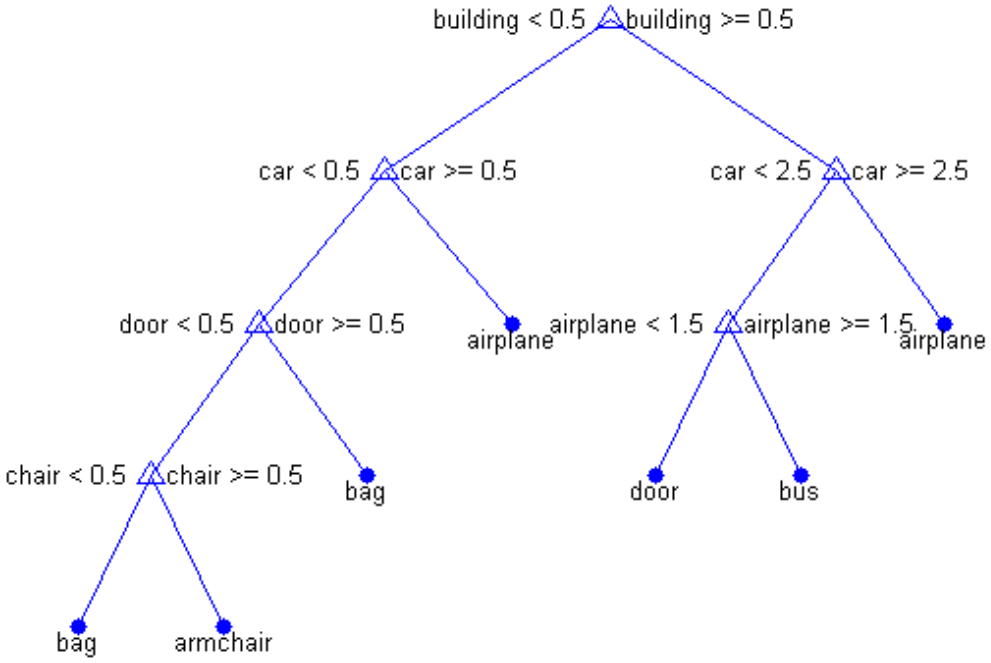
کلاسبند درخت تصمیم یکی از روش‌های تصمیم‌گیری چند مرحله ایست. ایده‌ی اولیه در هر روش چندمرحله‌ای، شکافتن یک تصمیم پیچیده به چندین تصمیم ساده‌تر است. یک گراف $G = (V, E)$ شامل مجموعه‌ای از گره‌های غیرتهی V و مجموعه‌ای از یال‌های E است. گره‌ای که فرزند نداشته باشد، برگ نامیده می‌شود.

در یادگیری درخت تصمیم هدف ایجاد مدلی است که مقدار تابع هدف را براساس متغیرهای مختلف ورودی بدست می‌آورد. هر برگ نمایش یک مقدار از متغیر هدف است که براساس مقادیری از متغیرهای ورودی که توسط مسیر از ریشه تا برگ وجود دارند، بدست می‌آید[26]. یادیگری درخت تصمیم یکی از موفق‌ترین تکنیک‌های یادگیری کلاسبندی باناظر است. یک درخت تصمیم یا درخت کلاسبندی، درختی است که هر گره‌ی داخلی در آن توسط ویژگی ورودی برچسب گذاری شده است. یک درخت با تقسیم مجموعه ورودی به زیربخش‌ها براساس یک مقدار ویژگی یادگرفته می‌شود.

الگوریتمی که در درخت تصمیم استفاده شده است الگوریتم کارت[۶۲] [24] است که در فصل قبل جزئیات مربوط به این الگوریتم توضیح داده شده است. این الگوریتم می‌تواند روابط داده‌های مهم را سریعا مشخص کند. طی سالها این الگوریتم به عنوان سریع‌ترین و تطبیق‌پذیرترین الگوریتم مدل‌سازی پیشگو برای آنالیز بوده است و به عنوان پایه برای بسیاری از روش‌های داده‌کاوی مدرن براساس بوستینگ و غیره بوده است[23].

در الگوریتم پیشنهادی سوالهای هر گره مربوط به اطلاعات متنی یا اشیاء دیگر تصویر است که براساس آنها برچسب گره برگ یعنی شئی که می‌خواهیم بازیابی کنیم بدست می‌آید. به عنوان مثال اگر این درخت برای تعداد شئ کمتر ساخته شود تصویری به صورت شکل 5-4 خواهد داشت:

[62] CART



شکل5-4- درخت تصمیم حاصل از الگوریتم پیشنهادی با درنظر گرفتن 40 شئ

تصمیم گیری براساس اعداد مشخص شده در هر گره صورت می گیرد به این ترتیب در هر گره تعداد شئ مربوطه با رابطه مربوط به گره مقایسه می‌شود و سپس هر کدام از شاخه ها که رابطه مربوط به آن برقرار باشد تا رسیدن به برگ پیمایش می شود. درنهایت براساس برچسب گره برگ شئی که حضور آن در تصویر مشخص نبود، یافت می‌شود. درصورتی‌که این درخت را برای کل اشیاء بسازیم، برای هر شئ قاعده ای بدست می‌آید که توسط آن می‌توان حضور یا عدم حضور آن شئ را دریافت.

۵-۴: خلاصه و نتیجه گیری

در این فصل ابتدا به بررسی چالش‌های موجود در روش‌های بازیابی شئ و سپس چالش‌های موجود در روش‌های بازیابی شئ مبتنی براطلاعات متنی پرداخته شد. پس از آن رویکرد کلی الگوریتم پیشنهادی ارائه و جزئیات هر یک از بخش‌های آن توضیح داده شد. در فصل بعد به بررسی نتایج اعمال این الگوریتم بر مجموعه داده سان09 پرداخته می شود و نتایج آن با الگوریتم‌هایی که بهترین نتایج را داشته‌اند مقایسه می‌شود.

فصل ۵:

پیاده سازی، بررسی کارایی و مقایسه با روش های دیگر

۵–۱:مقدمه

در این فصل ابتدا در مورد مجموعه داده سان بیشتر توضیح داده می شود و امکاناتی که این مجموعه داده فراهم کرده است، معرفی می شود . پس از آن دلیل استفاده از این مجموعه داده بیان می شود و تصاویر آن با تصاویر مجموعه داده های دیگر مقایسه می شود. سپس عملکرد روش پیشنهادی پایان نامه ارزیابی می شود و نتایج آن با روش‌های دیگر مقایسه می‌شود.

۵–۲: مجموعه داده

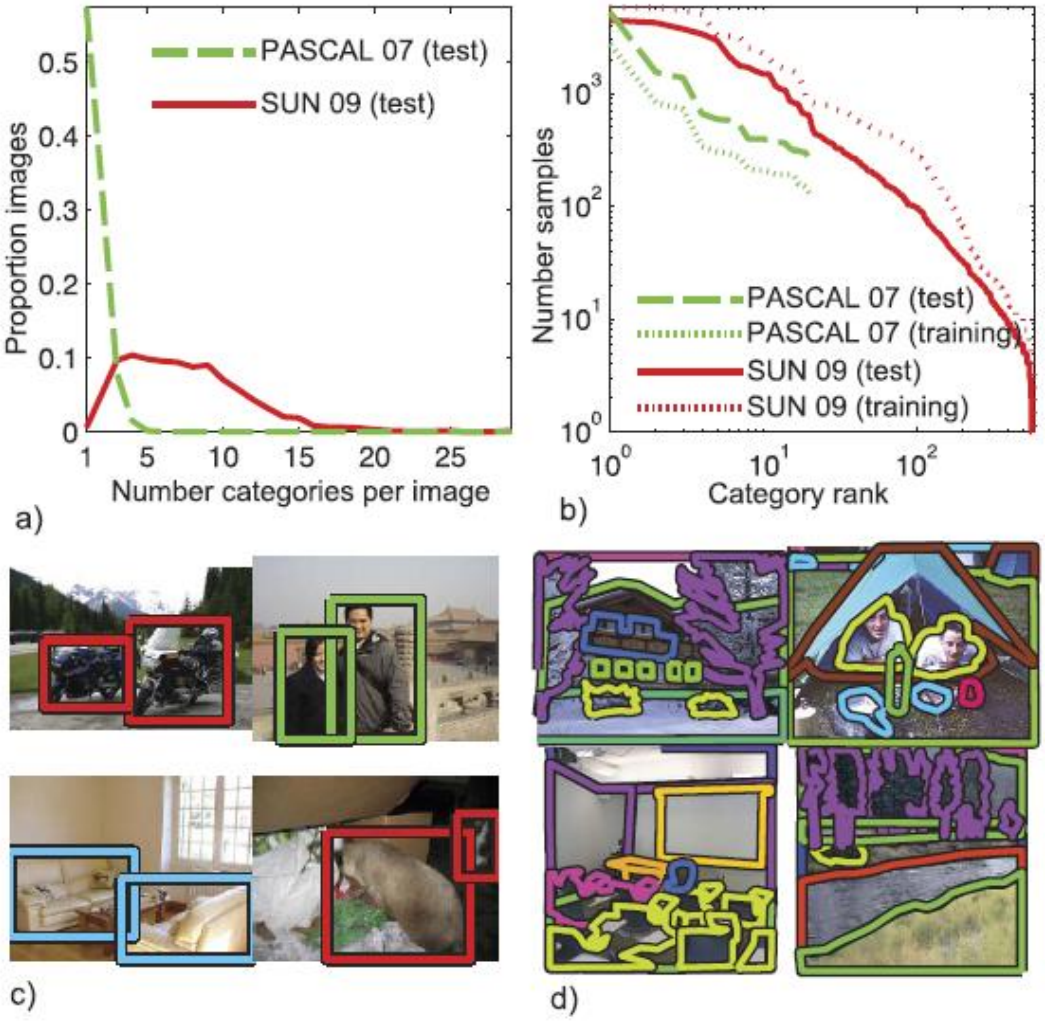
برای انجام تحقیقات شناسایی شئ، شامل یادگیری مدل‌های بصری شئ و انواع صحنه‌ها و تشخیص و تعیین محل نمونه‌ها در مدل‌های مختلف در تصاویر و ارزیابی عملکرد الگوریتم‌های شناسایی نیاز به مجموعه‌داده‌های مناسب است[27]. درسال‌های اخیر دیتاست‌های علامت‌گذاری شده‌ی زیادی که شامل کلاس‌های زیادی از اشیاء هستند، به وجود آمده‌اند و استفاده از روش‌های کمّی برای ارزیابی، سیستم‌های شناسایی را سرعت بخشیده‌اند. در این بخش تعدادی از دیتاست‌های متداول برای تست‌کردن کارآیی کلاس‌بندی شئ و صحنه و الگوریتم‌های تعیین موقعیت شئ بررسی شده اند.

مجموعه داده سان[۶۳] یک مجموعه داده وسیع برای درک صحنه می باشد. هدف این پایگاه داده فراهم کردن مجموعه وسیعی از تصاویر یادداشت نویسی شده از صحنه های مختلف، موقعیت ها و اشیاء با تنوع بالا برای پژوهشگران در حوزه بینایی کامپیوتر، ادراک انسانی، علوم شناختی، علوم اعصاب، یادگیری ماشین، داده کاوی، گرافیک کامپیوتر و روباتیک است.

سان 09 نیز که زیرمجموعه ای از این پایگاه داده است شامل بیش از 200 نوع شئ در رنج وسیعی از انواع صحنه‌هاست، شامل 12000 تصویر علامت گذاری شده و 152000 نمونه شئ علامت‌گذاری شده است که برای استفاده از اطلاعات متنی مناسب است. این تصاویر از منابع مختلف از جمله موتور جستجوی گوگل، فلیکر[۶۴]، آلتاویستا[۶۵] و لیبل می[۶۶] بدست آمده‌اند. در این تصاویر هر تصویر با یک شئ یا تصاویر با زمینه های سفید حذف شده‌اند تا تنها تصاویر مربوط به صحنه‌ها نگه‌داشته شوند. فرآیند تفسیر با یک تفسیرگر طی یک سال با استفاده از ابزارهای لیبل‌می[۶۷] انجام شده است. تصاویر

[63] SUN (Scene UNderstanding)
[64] Flickr
[65] Altavista
[66] Labelme
[67] http://labelme.csail.mit.edu/Release3.0/

برچسب‌گذاری شده به دقت برای استحکام و برچسب‌های مترادف بررسی می‌شوند و یکی می‌شوند[4]. تفسیرهای حاصل کیفیت بیشتری نسبت به لیبل می خواهند داشت. شکل زیر آمارهایی از این دیتاست را نشان می‌دهد و آنها را با دیتاست پاسکال 7 مقایسه می‌کند. همانطور که دیده می شود تعداد تصاویر با کلاس‌های شئ بیشتر از 2 شئ در مجموعه داده سان بسیار بیشتر است و در مجموعه داده‌ی پاسکال در یک تصویر حداکثر 5 شئ وجود دارد.



شکل 1-5- مقایسه‌ی دیتاست پاسکال7 و سان 9 a) هیستوگرام تعداد طبقه‌های شئ درهر تصویر b) توزیع نمونه‌های آموزشی و تست نمونه تصویرها در هر طبقه از شئ c)نمونه‌هایی از تصاویر پاسکال d) نمونه‌هایی از تصاویر سان[4]

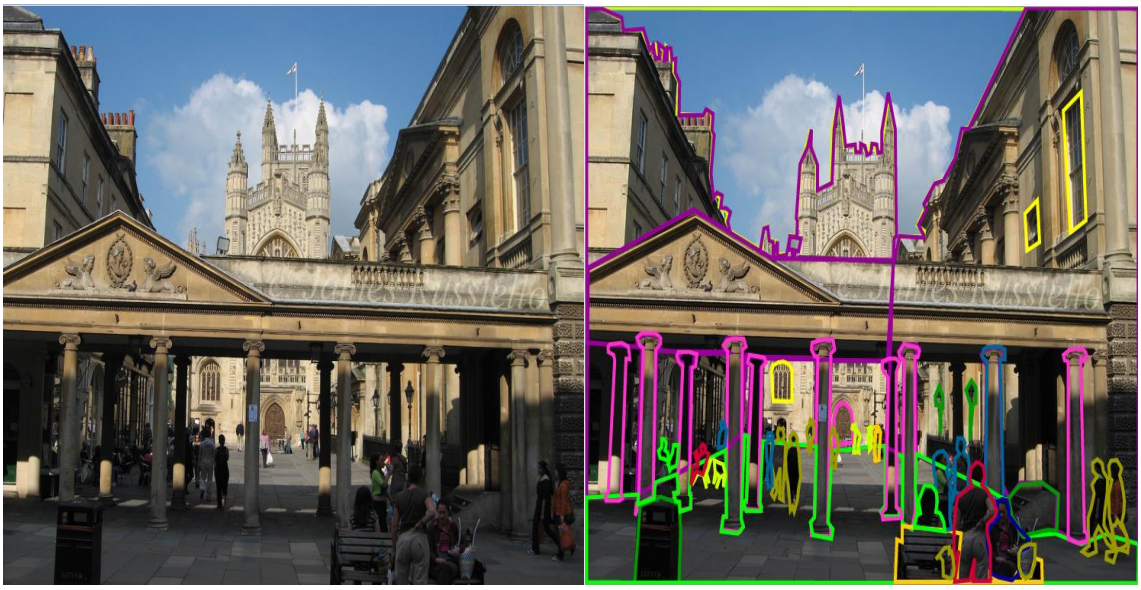
مزیت استفاده از مدل‌های متنی هنوز چندان خود را نشان نداده‌اند و دلیل آن این است که بیشتر روش‌های قبلی بر دیتاست‌هایی تست شده‌اند که تنها دسته‌های محدودی از اشیاء را شامل می‌شوند و بیشتر تصاویر شامل یک یا دو دسته شئ بوده‌اند[4].

در این گزارش از مجموعه داده سان استفاده شده است، زیرا در این مجموعه داده تصاویر براساس صحنه دسته بندی شده اند و تصاویر متعددی از صحنه های مختلف جمع آوری شده اند. تعداد زیاد اشیاء و وجود اشیاء مختلف در یک تصویر نیز از دلایل دیگر انتخاب این مجموعه داده بوده است زیرا در الگوریتم ارائه شده روابط بین شئی که قرار است شناسایی شود با اشیاء دیگر بسیار اهمیت دارد.

آزمایشات و ارزیابی عملکرد بر تصاویر مجموعه داده سان با انواع صحنه ها انجام شده است تا کارایی الگوریتم پیشنهادی را بر گروه های مختلف تصاویر نشان دهد. تصاویر مجموعه داده سان از صحنه های مختلفی هستند و در هر تصویر اشیاء زیادی وجود دارد. بنابراین این مجموعه داده برای روش هایی که مبتنی بر روابط اشیاء و صحنه هستند مناسب تر است. این مجموعه تصاویر در مقاله های [4, 16, 25] برای هدف بازشناسی شئ استفاده شده است و نتایج آزمایشات براین تصاویر بدست آمده اند.

شکل زیر یکی از تصاویر مجموعه داده سان با نام صحنه کلیسا را نشان می دهد. این تصویر دارای 21 نمونه شئ است که در شکل سمت راست هر کدام از این اشیاء نشان داده شده اند. لیست اشیاء موجود در این تصویر را می توان در شکل بعدی مشاهده نمود.

فایل های تفسیر تمام تصاویر مجموعه داده سان شامل اشیاء و صحنه و موقعیت های آنها، در این مجموعه داده موجود است. از این فایل های تفسیر می توان برای آموزش و تست الگوریتم های بازیابی شئ استفاده کرد. در این پایان نامه هدف پیش بینی حضور هر یک از این اشیاء با دانستن بقیه اشیاء و تعداد آنها است.



شکل2-5- نمونه ای از تصویر مجموعه داده سان از صحنه کلیسا. شکل بالا تصویر و شکل پایین تصویر که تمام اشیاء آن مشخص شده اند را نشان می دهد.

لیست اشیاء موجود در تصویر
Sky
Building
Column
Door
Window
Litter bin crop
Sidewalk
Steps
Streetlight
Spotlight
Bench occluded
Person occluded
Person sitting occluded
Person sitting crop
Person crop
Person walking
Person standing
Chair
Person sitting
Person
Bag

شکل3-5- لیست اشیاء موجود در تصویر قبل

در ادامه برای روشن شدن دلیل استفاده از مجموعه داده سان تعدادی از تصاویر مجموعه داده‌ی پاسکال که در مقاله های بسیاری برای بررسی عملکرد الگوریتم ها استفاده شده اند، بررسی می شوند.

شکل بعد سه نمونه از تصاویر مجموعه داده پاسکال را نشان می دهد. تصویر اول یک تصویر از باند فرود در یک فرودگاه است. این تصویر شامل دو شئ هواپیما و شخص است. تصویر دوم که تصویری از یک منظره است شامل یک شئ قایق می باشد و اشیاء دیگر در این مجموعه داده درنظر گرفته نشده‌اند. تصویر سوم تصویری از دو مانیتور است. این تصویر نیز تنها شامل یک شئ مانیتور است. همانطور که دیده می شود تصاویر مجموعه داده پاسکال دارای تعداد کمی شئ است و بنابراین برای ارزیابی الگوریتم-هایی که برمبنای روابط اشیاء هستند مناسب نیست. برای این الگوریتم ها بهتر است از تصاویری که دارای تعداد شئ زیاد است استفاده شود و تمام اشیاء در تصویر درنظر گرفته شوند.

شکل4-5- سه نمونه از تصاویر مجموعه داده پاسکال. تصویر 1 شامل دو شئ هواپیما و شخص است. شکل 2 شامل یک شئ قایق و شکل 3 شامل یک شئ مانیتور است.

همان طور که در فصل قبل توضیح داده شد، یکی از ابزارهای بسیار مناسب تفسیر که بصورت آنلاین فراهم شده است لیبل می است که امکان ایجاد فایل های تفسیر تصاویر مختلف را به صورت آنلاین فراهم می کند و فایل های تفسیر مجموعه داده سان توسط این ابزار فراهم شده اند. همچنین یکی از جعبه ابزارهای بسیار مهم و کاربردی نیز توسط همین پروژه ارائه شده است که امکانات بسیار گسترده و مهمی را برای کار با مجموعه داده خود و مجموعه داده سان و همچنین پاسکال فراهم می‌کند. این جعبه ابزار دارای بخش های مختلف و تابع های متنوع برای اعمال عملیات موردنظر بر روی فایلهای تصویر و تفسیر این مجموعه داده ها می باشد. برای استفاده از این توابع می توان این جعبه ابزار را نصب و سپس از توابع مربوطه استفاده کرد.

در این پایان نامه ابتدا تمام فایل های تفسیر تصاویر مجموعه داده سان که در آدرسی که در جعبه ابزار ذکر شده است، ذخیره شده اند، دانلود و ذخیره شدند. سپس با استفاده از جعبه ابزار لیبل می این فایل ها به صورت ساختاری در برنامه متلب ذخیره می شوند. چند مورد از توابع این جعبه ابزار که توسط این پایان نامه استفاده شده اند در ادامه بیان می شوند:

- ✓ تابعی با نام مجموعه داده سان که پس از اجرا تمام فایل های تفسیر موجود در آدرس ذخیره شده را به صورت ساختاری ذخیره می کند.

```
SUNdatabase();
```

- ✓ تابعی که نام و تعداد تمام اشیاء موجود در کلیه تصاویر مجموعه داده یا هر زیر مجموعه ای از آن را که در فایل ساختاری ذخیره شده اند را بر می گرداند.

```
LMobjectnames(D);
```

- ✓ تابعی که با گرفتن نام شئ موردنظر تعداد تکرار آن را در کل مجموعه یا در زیرمجموعه ای از آن برمی گرداند.

```
LMcountobject(D, objectname, method);
```

- ✓ تابعی که هر زیرمجموعه از مجموعه داده با شرایط مشخص شده را برمی گرداند.

```
LMquery(D, fieldName, content, method)
```

در ادامه به ارزیابی روش پیشنهادی و مقایسه آن با روش های مشابه پرداخته می شود.

۵-۳: ارزیابی روش پیشنهادی پایان نامه

در این بخش به ارزیابی نتایج این الگوریتم و مقایسه با الگوریتمی که بهترین عملکرد را بر روی مجموعه داده سان داشته است می پردازیم. این الگوریتم در فصل قبل توضیح داده شده است و در این بخش نتایج آنها بررسی می شود. در ادامه ابتدا به بیان شرایط ارزیابی و ذکر معیارهای مورد استفاده پرداخته می شود و پس از آن نتایج ارزیابی را بیان کرده و در نهایت به بررسی نتایج ارزیابی می پردازیم.

۵-۳-۱: شرایط ارزیابی

در این ارزیابی الگوریتم پیشنهادی را با الگوریتم روش مبتنی بر اطلاعات متنی سلسله مراتبی که در بخش قبل به تفصیل بررسی شدند و روش مدل‌های جداکننده ، مقایسه می کنیم. دلیل مقایسه با روش مدل‌های جدا کننده این است که این مدل در سال 2011 بهترین عملکرد را بر مجموعه داده پاسکال داشته است و یک روش غیرمتنی است. برای نشان دادن بهبود حاصل از روش متنی و بر روی مجموعه داده سان، نتیجه ای مدل برای مقایسه ذکر شده است.

تصاویر از مجموعه داده سان09 هستند که شامل 4367 تصویر آموزشی و 4317 تصویر تست می‌باشد. این تصاویر شامل بیش از 200 شئ هستند که مواردی همچون جاده، آسمان، ساختمان، ماشین، تخت خواب و حتی اشیاء دگردیس پذیری چون رودخانه، پرده و حوله ازجمله از آنها هستند. این تصاویر از بیش از 600 صحنه مختلف انتخاب شده اند و هدف بازیابی 107 کلاس شئ مشخص می باشد.

۵-۳-۲: معیارهای ارزیابی

در مسائل مختلف از معیارهای ارزیابی متفاوتی استفاده می شود. در این پایان نامه نیز تلاش شده است تا الگوریتم پیشنهادی با معیارهای استانداردی موردارزیابی قرار گیرد. این معیارها به صورت زیر می باشند:

مثبت درست[۶۸]: تعداد مواردی که جزء دسته موردبررسی هستند و در این دسته قرار گرفته اند.

مثبت غلط[۶۹]: تعداد مواردی که جزء دسته موردبررسی نیستند ولی در این دسته قرار گرفته اند.

منفی غلط[۷۰]: تعداد مواردی که جزء دسته موردبررسی هستند ولی در این دسته قرار نگرفته اند.

منفی درست[۷۱] : تعداد مواردی که جزء دسته موردنظر نیستند و در آن هم قرار نگرفته اند.

دقت[۷۲]: نسبت مواردی که به درستی در دسته موردنظر قرار گرفته اند به تمام مواردی که در دسته موردنظر قرار گرفته اند .

دقت متوسط: متوسط دقت بازیابی تمام اشیاء

حساسیت[۷۳]: نسبت مواردی که در دسته موردنظر قرار گرفته اند به تمام مواردی که به دسته موردنظر تعلق دارند

۵-۳-۳: نتایج ارزیابی

این پایان نامه بر مجموعه داده سان ارزیابی شده است. الگوریتم هایی که از روابط اشیاء و صحنه برای بازیابی شئ استفاده می‌کنند، بر مجموعه داده سان ارزیابی می شوند. بنابراین باید نشان داده شود دقت بازیابی شئ در این مجموعه داده توسط الگوریتم پیشنهادی بهبود می یابد.

در این پایان نامه دو دسته آزمایش انجام می شود تا عملکرد الگوریتم پیشنهادی به صورت کمی و کیفی ارزیابی شود. ارزیابی کمی براساس معیار دقت بازیابی شئ انجام می شود که در الگوریتم های

---

[68] True Positive
[69] False Positive
[70] False Negative
[71] True Negative
[72] Precision
[73] Recall or Sensitivity

بازشناسی شئ محاسبه می شود. تمام آزمایشات توسط متلب 2012 و در ویندوز 8 بر کامپیوتر با پردازنده 2.40 گیگاهرتز و حافظه موقت 4 گیگابایت انجام شده است. در ادامه نتایج مقایسه الگوریتم ها با جزئیات شرح داده خواهند شد.

۵.۳.۳.۱ ارزیابی کمی

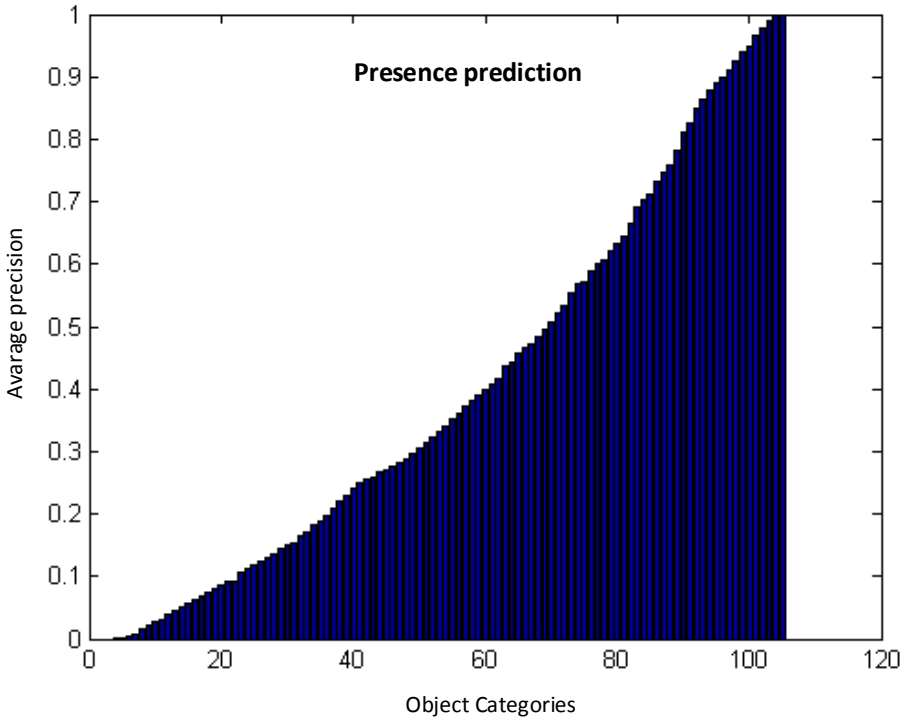
برای ارزیابی نتایج بازیابی شئ در تصاویر مجموعه داده سان، ارزیابی کمی الگوریتم های مختلف ضروری است. مقایسات توسط معیار میانگین دقت بازیابی که توسط معادله زیر تعریف شده است انجام می‌شود. دقت بازیابی هر شئ به صورت نسبت پیش بینی های شئ انجام شده که درست بوده اند، تعریف می‌شود.

$$\text{دقت بازیابی هر شئ} = \frac{\text{پیش بینی های درست آن شئ}}{\text{کل پیش بینی های آن شئ}}$$

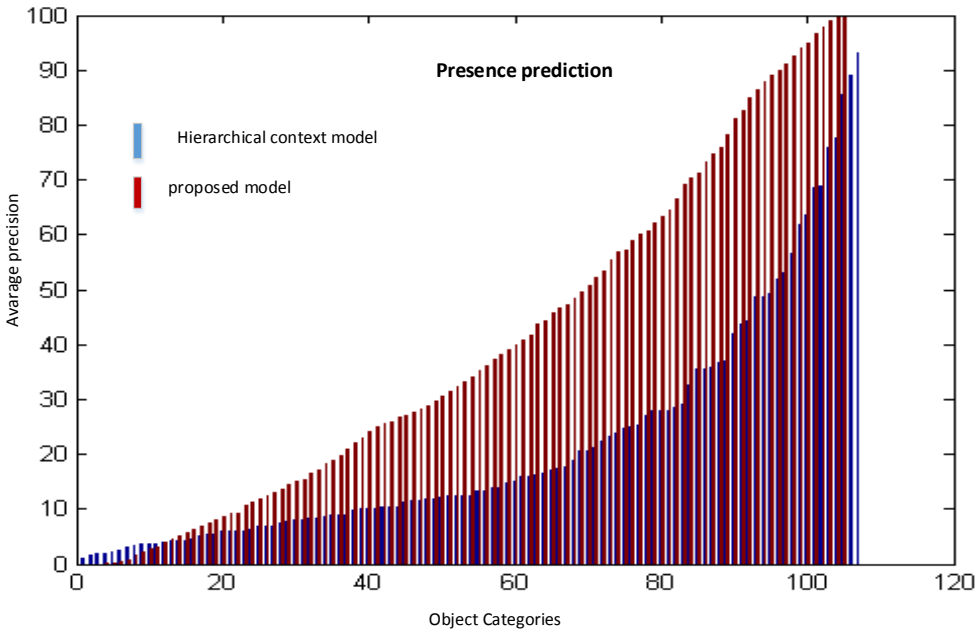
برای هر شئ میانگین دقت بازیابی هایی که در هر مرحله بدست آمده است به عنوان میانگین دقت[۷۴] بازیابی آن شئ درنظر گرفته می شود. سپس برای ارزیابی الگوریتم، متوسط میانگین دقت[۷۵] تمام اشیاء محاسبه و مقایسه می شود. در دومقاله ای که با الگوریتم پیشنهادی مقایسه شده اند 107 شئ بررسی و میانگین دقت آنها محاسبه شده است. برای اینکه بتوان نتایج را مقایسه و الگوریتم را ارزیابی کرد در این الگوریتم نیز میانگین دقت برای همان 107 شئ محاسبه شده است.

در نمودار 6-5 میانگین دقت 107 شئ برای الگوریتم پیشنهادی نشان داده شده است. در شکل 7-5 نیز میانگین دقت الگوریتم پیشنهادی و میانگین دقت الگوریتم مبتنی براطلاعات متنی سلسله مراتبی مقایسه شده اند. ستون های قرمز نشان دهنده دقت الگوریتم پیشنهادی برای هر شئ و ستون های آبی نشان دهنده دقت الگوریتم مبتنی بر اطلاعات متنی سلسله مراتبی است.

[74] AP(Average precision)
[75] Mean AP



شکل 6-5- دقت 107 شئ بازیابی شده در مدل پیشنهادی. اشیاء براساس دقت بازیابی مرتب شده‌اند.



شکل7-5- دقت بازیابی 107 شئ. ستون های قرمز مربوط به الگوریتم پیشنهادی و ستون های آبی مربوط به اگلوریتم مبتنی براطلاعات متنی سلسله مراتبی است.

در جدول زیر دقت بازیابی تعدادی از این اشیاء به عنوان نمونه در الگوریتم پیشنهادی و الگوریتم مبتنی بر اطلاعات متنی سلسله مراتبی مقایسه شده است.

جدول 1-5- دقت نمونه هایی از 107 شئ بازیابی شده توسط الگوریتم پیشنهادی

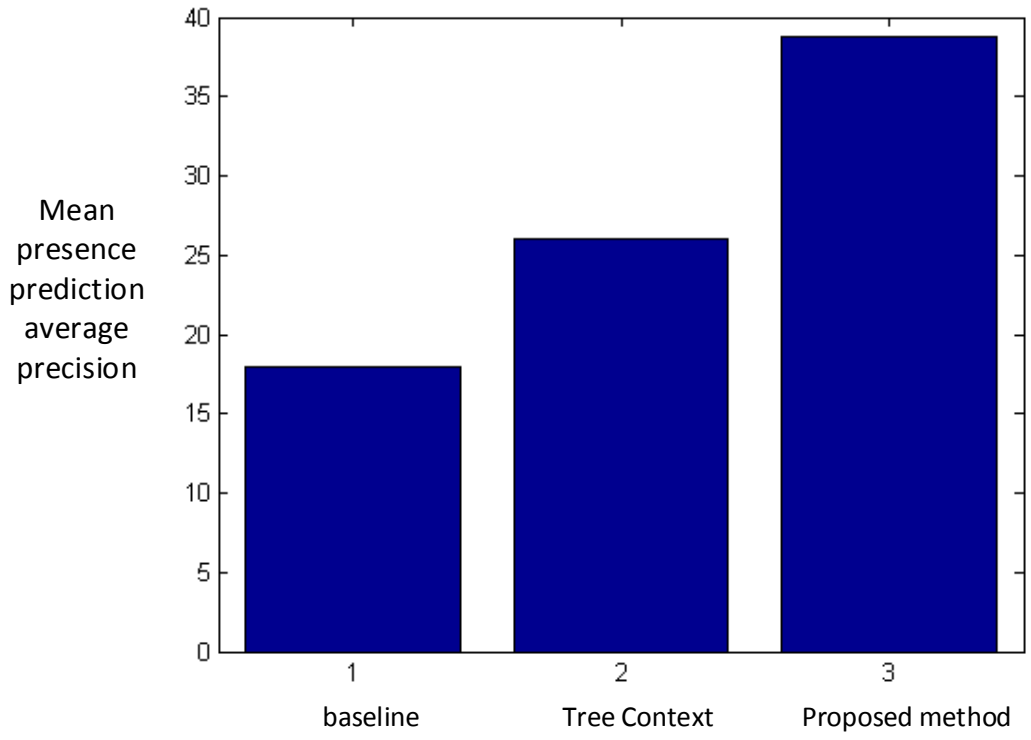
شئ بازیابی شده	دقت الگوریتم بازیابی شئ مبتنی بر اطلاعات متنی سلسله مراتبی[16]	دقت روش پیشنهادی
'airplane'	**44.39**	18.62
'armchair'	20.56	**33.91**
'awning'	10.17	**28.11**
'bag'	6.00	**30.36**
'balcony'	8.24	**40.41**
'ball'	3.61	**6.48**
'bars'	8.00	**40.71**

8.86	**8.98**	'basket'
86.98	51.84	'bed'
29.36	3.72	'bench'
39.69	27.10	'bookcase'
47.94	23.31	'books'
29.49	13.17	'bottle'
1.15	**2.47**	'bottles'
13.33	**15.83**	'bowl'
47.34	14.98	'box'
87.17	75.88	'building'
12.22	4.33	'bus'

با توجه به اینکه تعداد اشیاء موجود در تصاویر مجموعه داده سان بسیار بیشتر از مجموعه داده های دیگر است، نویز این مجموعه داده نیز از بقیه بیشتر است و نتایج بهترین الگوریتم‌ها نیز دقت پایینی را نشان می دهد. در جدول زیر دقت متوسط بازیابی سه الگوریتم [4, 10, 16]و الگوریتم پیشنهادی برای 107 دسته شئ نشان داده شده است[4].

جدول 2-5- دقت بازیابی دو الگوریتم با بهترین عملکرد بر مجموعه داده سان

	مدل های جدا کننده [10]	الگوریتم بازیابی شئ مبتنی بر اطلاعات متنی سلسله مراتبی[16]	الگوریتم پیشنهادی
دقت متوسط	17.93	26.08	38.81



شکل 8-5- مقایسه میانگین دقت الگوریتم پیشنهادی و الگوریتم های قبلی. الگوریتم 1 مدلهای جداکننده[28]، الگوریتم 2 الگوریتم مبتنی بر اطلاعات متنی سلسله مراتبی[16] و الگوریتم 3 الگوریتم متنی با کلاسبندی مبتنی بر مجموعه[25] است.

۵.۳.۳.۲ ارزیابی کیفی

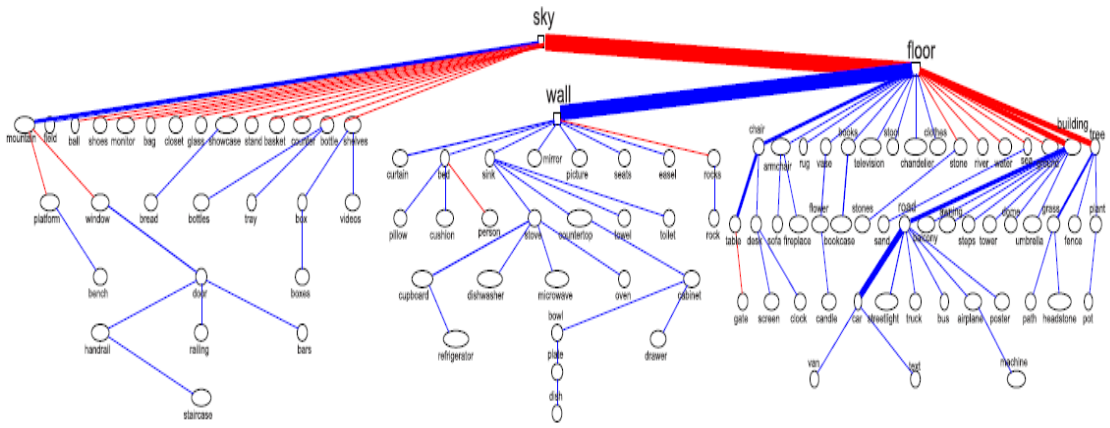
در این بخش ابتدا نشان داده می‌شود که در الگوریتم مبتنی براطلاعات سلسله مراتبی به دلیل درنظر نگرفتن تمام اشیاء موجود در آن و تعداد آنها، به دلیل تعداد زیاد شئ در تصاویر و تنوع آنها عملکرد مطلوبی ندارد. سپس نشان داده می‌شود که الگوریتم پیشنهادی این مساله را برطرف نموده است و با در نظر گرفتن تمام اشیاء دیگر موجود در آن و تعداد آنها توانسته است وجود شئ موردنظر را پیش‌بینی کند.

مقاله [4] که در فصل قبل به تفصیل بررسی شد و جزء بهترین الگوریتم‌هایی است که تاکنون ارائه شده است، ارتباط بین اشیاء را به صورت سلسله مراتبی ذخیره می‌کند. به عنوان مثال اشیائی که معمولا در یک آشپزخانه وجود دارند، در زیردرخت گره‌ی sink قرار می‌گیرند و تمام وسایل نقلیه نیز در زیردرخت شئ جاده قرار می‌گیرند. مساله‌ای که ممکن است در این مدل به وجود آید این است که

ممکن است در تصاویری وسایل آشپزخانه وجود داشته باشند اما شئ sink وجود نداشته باشد، درواقع در این مدل برای بازیابی هر شئ تنها از اطلاعات مربوط به والدهای آن تا ریشه در درخت متنی ساخته شده که در شکل 5-9 نشان داده شده است، استفاده می‌شود و اشیاء دیگری که ممکن است آنها نیز موثر باشند درنظر گرفته نمی‌شود که این چالش در مدل پیشنهادی بررسی شده است.

در مدل مبتنی بر اطلاعات متنی سلسله مراتبی هر شئ را می‌توان با یک متغیر باینری که نشان‌دهنده‌ی حضور یا عدم حضور یک شئ در تصویر است نشان داد. در یک درخت باینری هر گره b_i نمایشگر این است که شئ i در تصویر وجود دارد یا خیر و احتمال توأم متغیرهای باینری طبق ساختار درخت به صورت معادله (3-1) محاسبه می‌شوند.

برای بدست آوردن احتمال بالا همان طور که در فصل قبل توضیح داده شد از ساختار درختی زیر استفاده شده است. برای نشان دادن چالش مطرح شده در مورد این مدل با یک مثال توضیح داده خواهد شد.



شکل9-5– ساختار وابستگی اشیاء که از دیتاست SUN 09 یادگرفته شده است. یال های قرمز نشان دهنده ی رابطه ی منفی بین طبقه ها است. کلفتی هریال قدرت آن رابطه را نشان می دهد[16]

برای محاسبه احتمال حضور هر شئ در تصویر از احتمال حضور اشیاء والد آن تا ریشه استفاده می‌شود. شئی که بیشترین تعداد گره تا ریشه را دارد شئ dish است که احتمال حضور آن به صورت زیر محاسبه می‌شود:

$$P(b_{dish}) = P(b_{sky})P(b_{floor}|b_{sky})P(b_{wall}|b_{floor})P(b_{sink}|b_{wall})P(b_{countertop}|b_{sink})$$

$$P(b_{cabinet}|b_{countertop})P(b_{bowl}|b_{cabinet})P(b_{plate}|b_{bowl})P(b_{dish}|b_{plate})$$

در این حالت برای پیش‌بینی حضور شئ dish از احتمالات شرطی حضور 8 شئ دیگر کمک گرفته شده است. در مورد اشیاء دیگر تعداد این اشیاء کمتر است و برای پیش‌بینی احتمال حضور بیشتر اشیاء از احتمال شرطی حضور دو یا سه شئ دیگر استفاده شده است. این درحالی است که درصورت استفاده از اطلاعات اشیاء بیشتر می‌توان دقت بازیابی را افزایش داد. در مدل پیشنهادی برای پیش‌بینی حضور هر شئ از تمام اشیاء دیگر استفاده شده است و درنتیجه دقت به طور قابل ملاحظه ای افزایش یافته است.

به عنوان مثال در تصویر زیر که شامل 10 شئ نشان داده شده می‌باشد در مدل متنی مبتنی بر اطلاعات سلسله مراتبی اشیاء با بیشترین احتمال (احتمال 1) grass، headstone و box تخمین زده شده است در حالیکه در این تصویر هیچ یک از این اشیاء وجود ندارد. احتمال وجود هواپیما در این تصویر 0.05 پیش بینی شده است با توجه به اینکه این احتمال نزدیک به صفر است، نتیجه میگیریم که این الگوریتم پیش بینی کرده است شئ هواپیما در این تصویر وجود ندارد. اما توسط الگوریتم پیشنهادی وجود هواپیما تخمین زده شده است.

لیست اشیاء موجود در تصویر
Sky
airplane
Car
unmatched
building
fence
finger
person
window
road

شکل10-5- تصویری از مجموعه داده سان که به عنوان تصویر تست استفاده شده است و دارای 10 کلاس شئ مختلف است. احتمال وجود هواپیما در این تصویر در الگوریتم مبتنی بر اطلاعات سلسله مراتبی 0.05 است اما در الگوریتم پیشنهادی وجود آن پیش بینی شده است.

به عنوان مثالی دیگر می‌توان تصویر بعدی را درنظر گرفت که در الگوریتم مبتنی بر اطلاعات سلسله مراتبی احتمال وجود هواپیما 0.35 تخمین زده شده است درحالیکه وجود هواپیما در الگوریتم پیشنهادی به درستی تخمین زده شده است. همچنین در تصویر 11 احتمال وجود توپ 0.005 که عددی نزدیک به صفر است تخمین زده شده است ولی الگوریتم پیشنهادی وجود این شئ را پیش بینی کرده است.

لیست اشیاء موجود در تصویر
Sky
Control tower
airplane
building
field
finger
airport cart
road

شکل 11-5- – تصویری از مجموعه داده سان که به عنوان تصویر تست استفاده شده است و دارای 8 کلاس شئ مختلف است. احتمال وجود هواپیما در این تصویر در الگوریتم مبتنی بر اطلاعات سلسله مراتبی 0.35 است اما در الگوریتم پیشنهادی وجود آن پیش بینی شده است.

لیست اشیاء موجود در تصویر
Ceiling
Staircase
Handrail
Door
Furniture
Mirror
Armchair
Side table
Desk lamp
Painting
Fireplace
Table
Swivel chair
Billiard table
Cue
Ball
Door
Painting
Ceiling lamp
Floor lamp
Plant
Wall
Floor

شکل 12-5 – تصویری از مجموعه داده سان که به عنوان تصویر تست استفاده شده است و دارای 23 کلاس شئ مختلف است. احتمال وجود توپ در این تصویر در الگوریتم مبتنی بر اطلاعات سلسله مراتبی 0.005 است اما در الگوریتم پیشنهادی وجود آن پیش بینی شده است.

۵.۳.۳.۳ ماتریس تقابل

برای بررسی عملکرد مدل پیشنهادی، ماتریس تقابل که ماتریسی با تعداد سطر و ستون مساوی با تعداد اشیاء یعنی 107*107 است، محاسبه می شود. اشیاء ستون ها اشیاء خروجی کلاسبند و اشیاء سطرها اشیاء واقعی هستند. به ازای هر پیش بینی حضور شئ i به عنوان شئ j مولفه مربوطه یعنی C_{ij} یکی اضافه می شود و تمام مولفه های ماتریس محاسبه می شود. با توجه به تعریف این ماتریس می‌توان دریافت که هر چه عناصر قطری ماتریس تداخل مقادیر بیشتر و عناصر غیرقطری مقادیر کمتری داشته باشد، پیش بینی ها به واقعیت نزدیکتر بوده اند و تعداد پیش بینی های درست بیشتر بوده است.

ماتریس تداخل برای 10 شئ از این 107 شئ که دارای بیشترین دقت بازیابی بوده‌اند در شکل بعد نشان داده شده است. همانطور که دیده می‌شود بیشترین مقادیر مربوط به عناصر قطری هستند و بیشتر مقادیر غیرقطری دارای مقادیر صفر هستند، در نتیجه پیش بینی ها تاحدود خوبی به درستی انجام شده است. میزان درستی[۷۶] این ماتریس برای تمام اشیاء که با جمع عناصر قطری ماتریس بر جمع تمام عناصر ماتریس بدست می آید برابرست با 66.53 و نسبت کلاسبندی های غلط[۷۷] آن برابر با 33.47 است.

	sky	window	floor	road	chair	wall	person	sea	building	bed
sky	2065.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
window	0.00	972.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
floor	0.00	0.00	1547.00	0.00	2.00	0.00	1.00	0.00	0.00	0.00
road	0.00	0.00	5.00	633.00	0.00	0.00	0.00	0.00	0.00	0.00
chair	0.00	1.00	1.00	0.00	443.00	0.00	2.00	0.00	1.00	0.00
wall	0.00	0.00	2.00	0.00	1.00	1841.00	5.00	1.00	0.00	0.00
person	0.00	0.00	8.00	2.00	2.00	2.00	757.00	0.00	0.00	0.00
sea	0.00	0.00	0.00	0.00	3.00	0.00	0.00	156.00	0.00	0.00
building	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	184.00	0.00
bed	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	205.00

شکل 13-5- ماتریس تداخل برای 10 شئ با بیشترین دقت بازیابی از مجموعه داده سان

[76] Accuracy
[77] Error

۵-۴: جمع بندی

در این فصل ابتدا به معرفی مفصل‌تر مجموعه داده استفاده شده در پایان نامه و جعبه ابزار کاربردی لیبل می پرداخته شد و دلیل استفاده از این مجموعه داده توضیح داده شد. سپس نتایج بررسی شده بر این مجموعه داده برای اشیاء شناسایی شده ذکر شدند و همچنین این روش با روشی با بهترین عملکرد بر روی این دیتابیس مقایسه شد. در دو بخش ارزیابی کمی و کیفی به ارزیابی این الگوریتم با روشی که بهترین عملکرد را بر مجموعه داده سان داشته است، به صورت کمی و کیفی پرداخته شده است.

فصل ۶:

نتیجه‌گیری و کارهای آتی

در این پایان نامه ابتدا به معرفی یک سیستم بازشناسی شئ و سپس به معرفی روشهای مختلف بازشناسی شئ پرداخته شد. بعضی از این روشها از ویژگی های محلی و بعضی از ویژگی های سراسری تصویر استفاده می کنند. هر کدام از این روشها مزایا و معایب خود را دارند. بعضی روش ها نیز از ترکیب این دو روش استفاده می کنند.

روشهای دیگری که در این پایان نامه به آنها پرداخته شد، روش های مبتنی بر متن بودند که در این روش ها از اطلاعات صحنه و اشیاء دیگر برای شناسائی شئ موردنظر استفاده می شود. با توجه به اینکه به دلیل کوچک بودن بعضی اشیاء، وجود مانع و یا مبهم بودن یک شئ خاص در یک تصویر به دلیل شرایط مختلف نمی توان بازشناسی شئ را انجام داد، روش های مبتنی برمتن می توانند به شناسایی شئ در این موارد خاص کمک کنند. در نتیجه از نوع صحنه و اشیاء دیگر صحنه برای شناسایی شئی خاص استفاده می شود.

در قسمت ارزیابی نشان داده شده که نتایج الگوریتم پیشنهادی امیدبخش است و می توان از این روش در شرایط مختلف استفاده کرد. یکی از کارهایی که در ادامه این کار قابل انجام است، طبقه بندی اشیاء توسط روش ارائه شده است. به این صورت که توسط یک سری از اشیاء درخت تصمیم ساخته شود و سپس توسط اشیائی جدید درخت تصمیم پیمایش شود حال می توان اشیائی را که در یک گروه متنی قرار می گیرند بدست آورد و به این ترتیب می توان اشیاء را طبقه بندی کرد.

بدست آوردن اشیاء مختلف که عموما با یکدیگر و در صحنه های یکسان ظاهر می شوند می تواند در بازشناسی شئ با استفاده از تعریف قوانین وابستگی[۷۸] بسیار موثر باشد.

در فرآیند بازشناسی شئ در مغز اطلاعات بسیار زیادی از محیط پیرامون و روابط اشیاء دیگر حتی اشیاء ظاهرا نامرتبط دریافت می شود و با سرعت بالایی پردازش می شود و درنتیجه در زمان کوتاهی فرآیند بازشناسی شئ انجام می شود. در مدل پیشنهادی تلاش شده است تا نمایشی ارائه شود که اطلاعات کلاس های مختلف اشیاء بتوانند بر نمایش شئ تاثیر گذار باشند همانطور که در مغز اطلاعات اشیاء مختلف و ظاهرا نامرتبط دریافت و پردازش می شوند.

[78] Association rules

در مدل پیشنهادی پیش بینی در مورد موقعیت اشیاء انجام نمی شود و این بخش به عنوان کارهایی آینده درنظر گرفته شده است. همچنین در این مدل از روابط اشیاء برای ایجاد نمایش یک کلاس شئ استفاده شده است و در ماتریس مربوطه هر تصویر به تعداد اشیاء موجود در آن تکرار شده است، بنابراین ممکن است به نظر برسد که این نمایش دارای افزونگی می باشد، در کارهای آینده تلاش می شود با بهبود ماتریس نمایش ویژگی این افزونگی را برطرف کرد.

فهرست منابع

1. Andreopoulos, A. and J.K. Tsotsos, *50 Years of object recognition: Directions forward*. Computer Vision and Image Understanding, 2013. **117**(8): p. 827-891.
2. Grauman, K. and B. Leibe, *Visual Object Recognition.* Synthesis Lectures on Artificial Intelligence and Machine Learning, 2011. **5**(2): p. 1.
3. Galleguillos, C. and S. Belongie, *Context based object categorization: A critical survey.* Computer Vision and Image Understanding, 2010. **114**(6): p. 712-722.
4. Myung Jin, C., A. Torralba, and A.S. Willsky, *A Tree-Based Context Model for Object Recognition.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2012. **34**(2): p. 240-252.
5. R. C. Gonzalez, R.E.W., and S. L. Eddins, *Digital image processing using MATLAB*. 2004.
6. Jain, R., R. Kasturi, and B.G. Schunck, *Machine vision*. 1995: McGraw-Hill, Inc. 549.
7. Cristianini, N. and J. Shawe-Taylor, *Support Vector Machines*. 2000: Cambridge: Cambridge University Press.
8. Viola, P. and M. Jones. *Rapid object detection using a boosted cascade of simple features*. in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. 2001.
9. Viola, P. and M. Jones, *Robust Real-time Object Detection.* International Journal of Computer Vision - to appear, 2002.
10. Felzenszwalb, P.F., et al., *Object Detection with Discriminatively Trained Part-Based Models.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2010. **32**(9): p. 1627-1645.
11. Dalal, N. and B. Triggs. *Histograms of oriented gradients for human detection*. in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. 2005.
12. Felzenszwalb, P.F., R.B. Girshick, and D. McAllester. *Cascade object detection with deformable part models*. in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. 2010.
13. Oliva, A. and A. Torralba, *The role of context in object recognition.* Trends in cognitive sciences, 2007. **11**(12): p. 520-527.
14. Hoiem, D., A. Efros, and M. Hebert. *Putting Objects in Perspective*. in *CVPR*. 2006.
15. Chow, C. and C. Liu, *Approximating discrete probability distributions with dependence trees.* Information Theory, IEEE Transactions on, 1968. **14**(3): p. 462-467.
16. Myung Jin, C., et al. *Exploiting hierarchical context on a large database of object categories*. in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. 2010.
17. Rabinovich, A., et al. *Objects in context*. in *Computer vision, 2007. ICCV 2007. IEEE 11th international conference on*. 2007. IEEE.

18. Blackwell, P.A., D, *Appearance Based Object Recognition with a Large Dataset using Decision Trees*, in *Australasian Conference on Robotics and Automation* 2004.
19. Wilking, D. and T. Röfer, *Realtime Object Recognition Using Decision Tree Learning*, in *RoboCup 2004: Robot Soccer World Cup VIII*, D. Nardi, et al., Editors. 2005, Springer Berlin Heidelberg. p. 556-563.
20. Zhiquan, Q., et al. *Robust Object Detection Based on Decision Trees and a New Cascade Architecture*. in *Computational Intelligence for Modelling Control & Automation, 2008 International Conference on*. 2008.
21. R Marée, P.G., J Piater, L Wehenkel. , *Decision Trees and Random Subwindows for Object Recognition.* , in *ICML workshop on Machine Learning Techniques for Processing Multimedia Content* (2005).
22. Obdrzalek, S. and J. Matas. *Sub-linear Indexing for Large Scale Object Recognition*. in *BMVC 2005*. 2005.
23. *CART Classification And Regression Trees*. Available from: http://www.salford-systems.com/products/cart#cart.
24. Breiman, L., et al., *Classification and Regression Trees (Wadsworth Statistics/Probability)*. 1984: Chapman and Hall/CRC.
25. Cinbis, R. and S. Sclaroff, *Contextual Object Detection Using Set-Based Classification*, in *Computer Vision – ECCV 2012*, A. Fitzgibbon, et al., Editors. 2012, Springer Berlin Heidelberg. p. 43-57.
26. Safavian, S.R. and D. Landgrebe, *A survey of decision tree classifier methodology*. Systems, Man and Cybernetics, IEEE Transactions on, 1991. **21**(3): p. 660-674.
27. Ponce, J., et al., *Dataset Issues in Object Recognition*. 2006. p. 29-48.
28. Desai, C., D. Ramanan, and C.C. Fowlkes, *Discriminative Models for Multi-Class Object Layout.* Int. J. Comput. Vision, 2011. **95**(1): p. 1-12.

Abstract

In real world there is strong relation between objects and environment, so information of object relations and scene plays an important role in object detection. When a special object is searched in a scene, spectator will focus on locations with highest prior probability for object existence. Therefore scene context is very important for decision making about eye movement. Contextual influences on object recognition become evident if the local features are insufficient because the object is small, occluded or camouflaged.

In order to develop a precise method for object detection in real images with various objects, some solutions are proposed in this thesis. Object detection in this method is based on contextual information or scene and other objects of the image in order to increase detection precision. Contextual information are modeled as co-occurrence matrix and each image is shown by a vector displaying other objects of the image and their frequencies. These vectors are used to train the classifier. Because of the large number of training and test data and because indexing and matching methods are time-consuming, decision tree classifier is used. After training the classifier, test images vectors are surveyed to decide presence of each object and detection precision is computed.

An important point in contextual methods is to test them on images with high number of various objects and from different scenes, to use object relations for improvement in object detection. SUN dataset images are from more than 600 scenes and contain more than 200 object categories. The number of objects in each image is more than 10 classes, while in other datasets the maximum number of objects is 5 classes. In this method we use SUN 09 dataset and object presence prediction precision is compared with similar methods and is improved.

Index term: Object Recognition, Contextual Information, Object Detection, Scene Understanding

Kharazmi University

Faculty of engineering

M. Sc. Thesis

Artificial Intelligence

Context-based object recognition in digital images

By:

Soheila Sheikhbahaei

Superviser:

Dr. Jamshid Shanbehzadeh

Advisor:

Dr. Zeinab Ghasabi

2015, winter