

تشخیص چهره سریع و قوی با کمک بافت محلی

چکیده

در نظارت تصویری، تشخیص چهره می‌تواند نشانه مهمی برای مقداردهی اولیه الگوریتم ردیابی باشد. کارهای اخیر در زمینه‌ی psychophysics اشاره به اهمیت بافت محلی یک چهره برای تشخیص قوی، مانند خطوط سر دارد. این مقاله یک آشکارساز ارائه می‌دهد که فعلا نه از ایده بافت محلی بهره می‌گیرد. هدف، به دست آوردن استحکامی فراتر از توانایی‌های آشکارساز سنتی است که جهت بررسی و نظارت جالب است. عملکرد آشکارساز پیشنهادی از نظر دقت و سرعت در مجموعه داده‌هایی از 2000 PETS و 2003 PETS ارزیابی شده و با روش شی گرای می مقایسه شده است. توجه خاص بیشتر بر نقش وضوح تصویر در دسترس تکیه دارد.

1. مقدمه و کارهای مرتبط

تشخیص سریع و قوی هدف از یک تصویر واحد، قابلیت مطلوبی برای سیستم‌های ردیابی و نظارت است. ابتدا، اجازه می‌دهد تا ارتباطات موجود در ردیابی اهداف بررسی شود و سپس فرضیه‌های ناخواسته به منظور ردیابی کارآمدتر نادیده گرفته می‌شود. ثانیاً اجازه می‌دهد تا با ردیابی شکست اهداف از دست رفته، بازگردانده و ترمیم شوند. ثالثاً، می‌تواند به طور موثر ردیابی را در مواردی که در آن هدف حرکت بسیار کم و یا جنبش بسیار ناپیوسته‌ای دارد (به عنوان مثال حرکات از دست رفته در فریم‌های ویدئویی) تکمیل کند. تمام این مثال‌ها نیاز دارند که تشخیص سریع و قوی باشد. اغلب، تشخیص مردم در محوطه جالب توجه است. نشانه مهم برای تشخیص افراد با استفاده از بینایی کامپیوتر، چهره انسان [1، 2]، سر [3، 4]، تمام بدن از جمله پاها [5] و همچنین پوست انسان [6] است. که در تشخیص چهره در

تصاویر ثابت احتمالا محبوب‌ترین است. الگوریتم‌های اخیر از ویژگی‌های با محاسبه سریع و ساختار آبخاری برای رسیدن به کارآیی زمان واقعی در سطح بالایی از دقت [7] استفاده می‌کنند. یک آشکارساز از این نوع برای محلی‌سازی چشم در دنباله ویدیو FGNET (ورکشاپ 2003PETS [8]) با موفقیت به کار گرفته شده است.



شکل 1

با این حال، جای تعجب دارد که ببینید چگونه آشکارساز چهره به راحتی می‌تواند در شرایطی که انسان هیچ مشکلی با تشخیص چهره قابل اعتماد ندارد فریب خورده است. چنین مواردی به صورت نظام‌مند در آزمایش‌های روانشناسی توسط Sinha و Torralba [9, 10] مورد مطالعه قرار گرفته است. یکی از یافته‌های رایج این است که سیستم بینایی انسان می‌تواند چهره واقعی را از الگوهای چهره با وضوح بسیار کم تشخیص دهد. از سوی دیگر سیستم محاسباتی نه تنها به جزئیات بیشتری از صورت برای تشخیص چهره در صحنه‌های واقعی نیاز دارد، بلکه هشدارهای کاذب که به درستی توسط انسان رد می‌شود باید قابل تشخیص باشد.

آزمایش Torralba نشان می‌دهد که سطح جزئیات همانند مناطق محلی اطراف صورت، استفاده انسان از بافت محلی را کاهش می‌دهد. که در تضاد با این فرض روش شی محور است که، تنها ویژگی‌هایی از تصویر که مربوط به تشخیص یک شی در یک مکان هستند از ویژگی‌های بالقوه متعلق به شی و غیرمرتبط به پس‌زمینه هستند.

تصویر شکل 1 چند نمونه از چهره‌ها را در بافت محلی و جهانی خود نشان می‌دهد. این مقاله یک آشکارساز را به کار می‌گیرد که به طور فعال از بافت محلی به عنوان یک پیش‌بینی برای تشخیص چهره محاسباتی بهره می‌گیرد. هدف

به دست آوردن استحکامی است که فراتر از توانایی‌های سنتی بین آشکارسازهای شی محور صورت به‌ویژه مربوط به بصری سازی است.

بخش 2 ایده بافت محلی و تحلیل تفاوت در آشکارسازهای نتیجه شده را هنگامی که با بافت محلی یا بدون بافت محلی آموزش می‌بینند تعریف می‌کند. آشکارساز پیشرفته آبخاری به‌عنوان یک آشکارساز زمینه استفاده می‌شود. تشخیص قابلیت‌های آشکارساز بافت محلی در مقایسه با روش شی محور در بخش 3 بیان شده است. مجموعه آزمون به کار براساس داده‌ی ویدئو کنفرانس FGNET (2003PETS) و محوطه پارکینگ PETS2000 است. بخش 4 خلاصه نتیجه‌گیری از این کار است.

2. بافت محلی در مقابل آشکارساز شیء محور

بنا به مقاله [11] مشتق رسمی برای بافت محلی به شرح زیر است: ابتدا، احتمال برای حضور شی O می‌تواند با استفاده از قانون بیز تجزیه شود مانند:

$$P(O | \mathbf{v}) \simeq P(O | \mathbf{v}_L) = \frac{P(\mathbf{v}_L | O)}{P(\mathbf{v}_L)} P(O) \quad (1)$$

که در آن اندازه تصویر \mathbf{v} یک اندازه محلی است، که $\mathbf{v} = \mathbf{v}_L$ است. هدف روش شی محور با $P(\mathbf{v}_L | O)$ نشان داده شده است و $P(O)$ جسم قبلی است. باین حال، به‌منظور جذب وابستگی بین یک شی و بافت آن بردار اندازه‌گیری می‌توان توسعه یابد که شامل ویژگی‌های خارج شی هدف

$$\mathbf{v} = \{\mathbf{v}_L, \mathbf{v}_C\} \quad (2)$$

است که در آن \mathbf{v}_C اندازه‌هایی از متن جسم هستند. در حال حاضر به کار بردن قانون بیز مانند عبارتی است که در آن همه احتمالات شرطی هستند

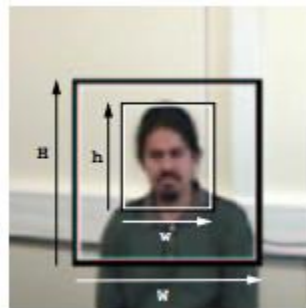
$$P(O | \mathbf{v}) = \frac{P(O, \mathbf{v})}{P(\mathbf{v})} = \frac{P(\mathbf{v}_L | O, \mathbf{v}_C)}{P(\mathbf{v}_L | \mathbf{v}_C)} P(O, \mathbf{v}_C) \quad (3)$$

برای پیاده‌سازی روش بافت محلی، آشکارساز با نمونه‌هایی که حاوی تمام سر یک فرد، گردن و بخشی از قسمت فوقانی بدن اوست تعلیم می‌بیند. به‌طور مستقیم این انتخاب حاوی نشانه‌های مهمی برای حضور چهره‌ها است. داده‌های تعلیم داده شده‌ی نتیجه شده کاملاً متفاوت از روش شی محور هستند که در آن‌ها تنها چهره در نظر گرفته می‌شود. شکل 2 نمونه‌های تعلیم داده شده از هر دو پارادایم را برای مقایسه نشان می‌دهد.



شکل 2

در طول تشخیص محل صورت واقعی که با فرض یک موقعیت ثابت در داخل پنجره تشخیص استنباط می‌شود. اندازه و محل صورت به‌طور مستقیم از محاسبه عرض و ارتفاع در بافت محلی شناسایی شده است. که در شکل 3 نشان داده شده است.

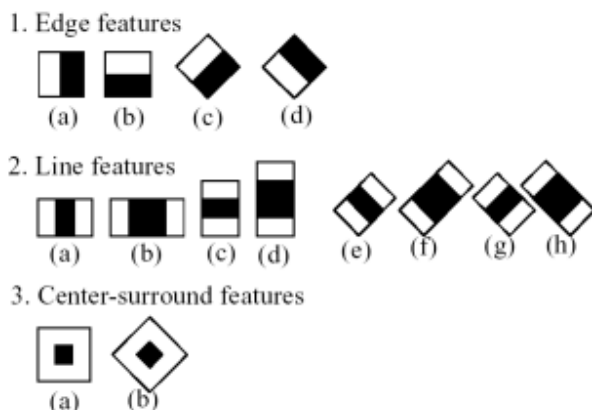


شکل 3

چارچوب آشکارساز به کار رفته، یک نسخه اصلاح شده از آشکارساز ویولا-جونز و در دسترس از طریق کتابخانه بصری کامپیوتر است [12]. اطلاعات بیشتر در مورد الگوریتم ارائه شده، رویکرد یادگیری و پارامترها در یک بخش جداگانه توضیح داده شده است (بخش 2.1).

در واقع، از ویژگی‌های این آشکارساز تفاوت‌های وزنی در انتگرال نواحی مستطیل شکل است. شکل 4 مجموعه‌ای از انواع ویژگی‌های موجود (گرفته شده از [13]) را به تصویر می‌کشد که در آن مستطیل‌های سیاه و سفید مربوط به وزن مثبت و منفی هستند. ویژگی‌ها شامل چهار نوع ویژگی مختلف لبه، هشت ویژگی خطی و دو ویژگی مرکزی است.

الگوریتم یادگیری (بنا به یادگیری درخت تصمیم) ویژگی‌ها را به‌طور خودکار با توجه به تمام ویژگی‌های تفکیک شده‌ی ممکن، اندازه و مکان انتخاب می‌کند. مقایسه ویژگی‌های انتخاب شده در روش شی محور و روش بافت محلی جالب است.



شکل 4



شکل 5

شکل 5 یک بصری‌سازی از ویژگی‌های مستطیل انتخاب شده است (ردیف بالا: مورد شی محور، دو ردیف پایین: مورد بافت محلی). دو مورد از تعلق‌های مختلف (افراد) در مورد بافت محلی به‌منظور ارائه نشان داده شده است. ویژگی‌های نمایش داده شده در اینجا در مرحله اول از آشکارساز آبخاری ارزیابی می‌شود بنابراین به‌عنوان مهم‌ترین

ویژگی‌ها در نظر گرفته می‌شود. در آنجا 9 ویژگی در مورد شی محور و 14 ویژگی در مورد بافت محلی وجود دارد (الگوریتم یادگیری تعدادی از ویژگی‌های مورد نیاز در هر مرحله را به‌طور خودکار تعیین می‌کند).

در موارد شی محور سه ویژگی اول، ضبط نواحی داخلی چهره، به ویژه اطراف چشم و اطراف دهان و بینی با استفاده از امکانات خط افقی و عمودی است. ویژگی‌های اضافی عمدتاً ویژگی‌های لبه برای گرفتن خطوط سر و چانه است (5F, 8F و 9F در شکل).

بنابراین، برای بافت محلی اولین ساختار در کل مسیر گسترش می‌یابد، به‌عنوان مثال آن واقعا از بافت محلی استفاده می‌کند. ویژگی‌های زیر خطوط سر و بدن (ویژگی‌های 5-F2F) را ضبط می‌کند. ویژگی‌های دیگر شانه سمت چپ و راست (ویژگی‌های 9F و 10F در مثال بالا، ویژگی 14F در مثال پایین) را ضبط می‌کند. از این رو کاملاً متفاوت از آشکارسازهای سنتی چهره است، که به تنهایی بر روی قطعات صورت تکیه داشتند.

2.1. روش آموزش و جزئیات پیاده‌سازی

چارچوب آشکارساز استفاده شده در این ایده بر اساس افزایش آشکارساز آبخاری طبقه‌بندی شده است (نگاه کنید به [7]) اما قابلیت تنظیم اصلی را گسترش می‌دهد و انواع مختلف تقویت یادگیری را ارائه می‌دهد [13]. این بخش خلاصه‌ای از ضروری‌ترین جزئیات پیاده‌سازی در مورد ویژگی‌های الگوریتم یادگیری و پارامترهای آموزشی است. انواع ویژگی‌های مشخص شده در شکل 4 یادآور موجک‌ها و ویژگی‌های اولیه از مسیر بینایی انسان مانند مرکز فراگیر و جهت پاسخ است. مزیت اصلی آنها این است که می‌توان آنها را باهم و در زمان ثابتی در هر مقیاسی محاسبه کرد. هر ویژگی با جمع پیکسل در مستطیل‌های کوچکتر قابل محاسبه است.

$$\text{feature}_I = \sum_{i \in I = \{1, \dots, N\}} \omega_i * \text{RecSum}(r_i) \quad (4)$$

با وزن‌های $\omega_i \in R$ ، مستطیل I_i و شماره آن‌ها یعنی N ، تنها ترکیب وزنی مجموع پیکسل‌های دو مستطیل در نظر گرفته می‌شود، که $N=2$ است. وزن‌ها علاماتی مخالف دارند (با نقاط سیاه و سفید در شکل نشان داده شده‌اند)، و برای جبران تفاوت در منطقه استفاده می‌شود. محاسبات کافی با استفاده از خلاصه جدول‌ها قابل دستیابی است. ویژگی‌های

چرخشی و ویژگی‌های فراگیر مرکزی مجموعه‌ای از ویژگی‌های اصلی از ویولا-جونز توسط Lienhart و همکارانش [13] با استفاده از خلاصه جداول چرخش است. مجموعه اصلی تنها شامل ویژگی‌های (1a)، (1b)، (2a)، (2c) به‌عنوان یکی از ویژگی‌های مورب است که توسط ویژگی‌های چرخشی رده‌بندی می‌شود. مجموعه‌ای از ویژگی افزوده برای گسترش بیان و تطبیق‌پذیری ویژگی‌های اصلی منجر به تشخیص دقیق‌تر نشان داده شده است. توجه داشته باشید که نمایش ویژگی‌ها نیازی به محاسبه هرم تصویر برای جستجو در مقیاس‌های مختلف ندارد.

الگوریتم یادگیری آبخاری مشابه درخت تصمیم یادگیری است. اساساً، طبقه‌بند آبخاری می‌تواند به‌عنوان یک درخت تصمیم‌گیری بیان شود. برای هر مرحله در آبخار زیر طبقه‌بند جداگانه‌ای برای تشخیص تمام اهداف اشیاء آموزش داده شده است درحالی که بخش خاصی از الگوهای غیر شیء صرف‌نظر می‌شوند. میزان شناسایی و در نتیجه میزان مثبت کاذب آبخاری توسط رابطه زیر قابل محاسبه است

$$F = \prod_{i=1}^K f_i \quad D = \prod_{i=1}^K d_i \quad (5)$$

برای مثال، اگر یک آشکارساز 20 مرحله‌ای به گونه‌ای آموزش دیده باشد که در هر مرحله 50 درصد از الگوهای غیرشیء حذف شوند (هدف قرار دادن میزان مثبت کاذب) درحالی که تنها 0.1٪ از الگوهای شیء (نرخ تشخیص هدف) به اشتباه از بین رفته است سپس انتظار می‌رود نرخ تشخیص کلی $0.99920 \approx 0.98$ با میزان مثبت کاذب 0.520 $\approx 0.9 * 6-10$ باشد. در نهایت، تعداد مورد نظری از مراحل، هدف نرخ مثبت کاذب و نرخ تشخیص هدف در هر مرحله اجازه می‌دهد تا دقت و سرعت طبقه‌بندی مورد نظر مبادله شود. همچنین توضیح می‌دهد که تعداد مختلفی از ویژگی‌ها در مرحله اول برای آشکارساز شیء محور و برای آشکارساز بافت محلی در شکل 5 نشان داده شده است.

مراحل جداگانه با استفاده از تقویت که مجموعه‌ای از "یادگیرنده‌های ضعیف" نامیده می‌شود ترکیب می‌شوند ("طبقه‌بندی قوی"). در این مورد یک یادگیرنده ضعیف معادل یکی از ویژگی‌های خاص و (به‌طور خودکار آموزش دیده) آستانه باینری در مقدار آن است. هر دور از افزایش یادگیرنده ضعیفی را (به‌عنوان مثال از ویژگی‌ها و آستانه) انتخاب می‌کند که مجموعه آموزش را به بهترین شکل ممکن طبقه‌بندی می‌کند. اولین افزایش، فرض می‌کند که وزن داده‌های آموزشی یکنواخت است درحالی که مراحل پی‌درپی وزن بالاتری به موارد آموزشی که به اشتباه طبقه‌بندی شده‌اند

نگاشت می‌کند. این مسئله موجب می‌شود بر روی الگوریتم در موارد "سخت" بیشتر تمرکز شود. تفاوت افزایش در انواع گسسته، AdaBoost واقعی و آرام به‌طور عمده توسط چارچوب OpenCV در نحوه تعیین وزن‌های جدید متفاوت از داده‌های آموزشی ارائه شده است. برای جزئیات بیشتر توصیه می‌شود به [14] مراجعه کنید. به‌صورت تجربی در [12] تجربی نشان داده شده است که AdaBoost آرام در موارد گسسته بهتر عمل می‌کند و AdaBoost واقعی برای کارهای تشخیص چهره دقت و سرعت بیشتری دارد. بنابراین AdaBoost آرام (GAB) انتخاب نهایی شده است.

جدول 1

Parameter	Loc.Context	Obj-centered
Positive examples	960	5000
Negative examples	1232	3000
Stages	21	25
Min hit rate	0.995000	N/A
Max false alarm rate	0.500000	N/A
Width	20	24
Height	20	24

در نهایت، مهم‌ترین پارامترهای آموزشی مانند نوع و تعداد نمونه آموزش، نرخ هدف و اندازه پنجره تشخیص به‌صورت خلاصه بیان شده است. حدوداً 1000 نمونه بافت محلی از جهان گسترده وب و از مجموعه عکس‌های خصوصی جمع آوری شده است. به‌منظور محدود کردن میزان تنوع و در نتیجه افزایش تفکیک، تنها دیدگاه روبه‌جلو برای آموزش و موارد تقریباً تراز وسط استفاده شده است. هر مثال مثبت به 20×20 پیکسل کوچک شده است.

برای جمع‌آوری نمونه‌های منفی یک زیرمجموعه از پایگاه داده‌های تصویر WuArchive استفاده شده است که شامل تمام افراد نیست. این تصاویر بارها و بارها توسط الگوریتم یادگیری برای جستجوی الگوهای شی بررسی شده است. "مرز موارد" اجازه می‌دهد تا کران تصمیم مجدداً تعریف شود (این روش گاهی اوقات به عنوان "خود راه انداز" نامیده می‌شود). آموزش آشکارساز بافت محلی حدوداً 48 ساعت در یک ماشین پنتیوم GHZ1 زمان می‌برد. جدول 1 پارامترهای آشکارساز بافت محلی را با آشکارساز جسم محور که همراه با کتابخانه OpenCV است مقایسه می‌کند که در آزمایشات بخش بعدی مورد استفاده قرار گرفته است.

3. ارزیابی عملکرد

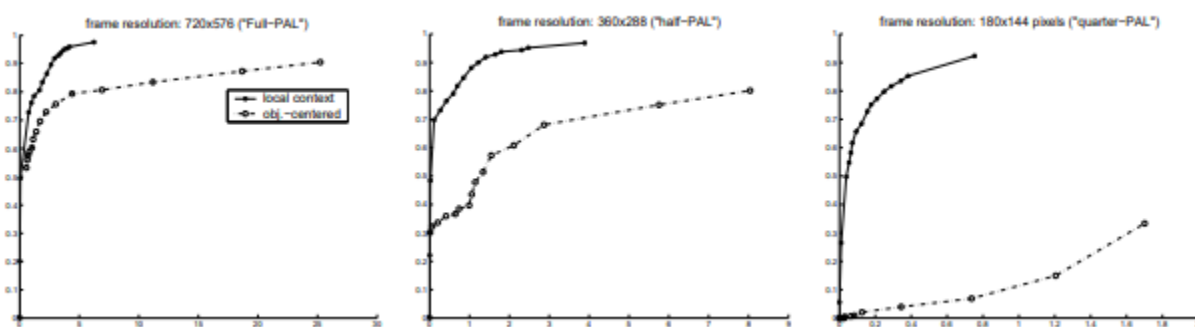
برای درک ارتباط آزمایشات متفاوت بافت محلی که در دنباله ویدئو PETS انجام شده است، داده‌های کنفرانس تصویری FGNET از PETS 2003 و محوطه پارکینگ از PETS 2000 جمع‌آوری شد. هر دو مجموعه داده مجزا از داده‌های مورد استفاده برای آموزش هستند.

3.1. فضای داخل

در داده ویدئو کنفرانس FGNET، هر 100 فریم متوالی از دنباله A، B، D (دوربین‌های 1 و 2) در آزمایشات زیر مورد استفاده قرار گرفته است. این نتایج در کل حاوی 502 قاب محتوی 1160 چهره، که 160 تا از آن‌ها مشخصات پروفایل آن‌هاست (حدود 14 درصد).



شکل 6



شکل 7

هر فریم دارای رزولوشن 576×720 پیکسل و چهره‌ها حدود 64×48 پیکسل بزرگ هستند. دنباله، یک اتاق کنفرانس در سراسر هر دو طرف میز را با مردمی که گاهی اوقات از اتاق ورود و خروج می‌کنند نشان می‌دهد.

شکل سمت چپ در شکل 7 نشان می‌دهد که عملکرد تشخیص چهره بر روی مجموعه داده FGNET منحنی ROC است. درصد چهره‌های بازیابی شده در محور عمودی داده شده است و تعداد مثبت کاذب در هر فریم در محور افقی نشان داده شده است. نقطه‌ها در منحنی بوسیله نمونه‌برداری متفاوت آستانه از مرحله نهایی در آشکارساز آبخاری قابل تشخیص است:

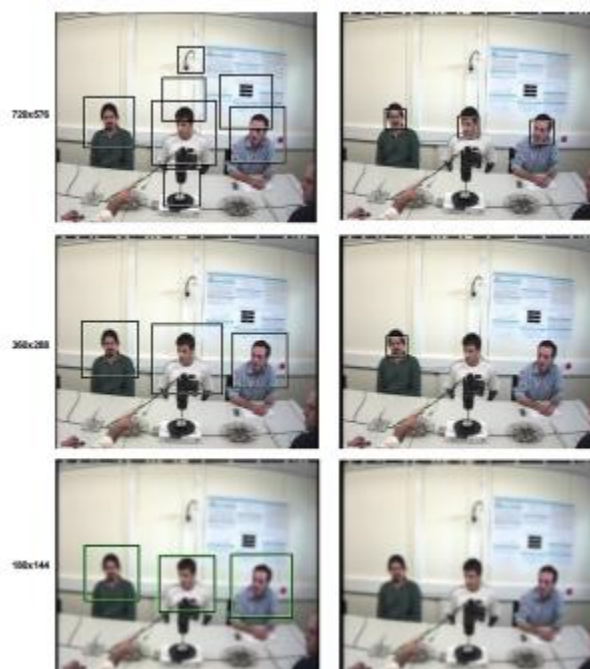
$$H = \text{sign} \left[\sum_{i=1}^K h_i + b \right] \quad (6)$$

که در آن H خروجی طبقه‌بندی نهایی است و h_i نشان‌دهنده مراحل فردی است. آبخاری پی در پی به منظور تشخیص عملکرد اضافی حاصل می‌شود. عملکرد هر دو آشکارساز در شی محور و بافت محلی نشان داده شده است. برای نسخه شی محور، آشکارساز صورت توسط Lienhart و همکارانش مورد استفاده قرار گرفته است. آشکارساز برای عملکرد عالی در مقایسه با موارد دیگر نشان داده شده است [12]. همانطور که می‌توان در شکل سمت چپ شکل 7 دید، 5 اخطار کاذب آشکارساز شی محور در 80 درصد از چهره‌ها و منحنی مسطح قابل مشاهده است. بنابراین، آشکارساز بافت محلی 95٪ از چهره‌ها را با همان سطح از اخطار کاذب تشخیص می‌دهد.

این را می‌توان با بازدید پروفایل چهره‌های موجود در داده توضیح داد که توسط آشکارساز جسم محور که در اینجا مورد استفاده قرار گرفته قابل شناسایی نیست. باین حال، بافت محلی، توسط تغییرات مطرح چهره تحت تاثیر قرار نمی‌گیرد و به این ترتیب می‌تواند نمایش از روبه‌رو و کنار را تشخیص دهد. به‌عنوان مثال قاب حاوی نمایش جانبی که در شکل 6 نشان داده شده است نیز خروجی هر دو آشکارساز را نشان می‌دهد. در روش شی محور یک نفر باید به‌صورت جداگانه نمونه مشخصات را جمع‌آوری کند و آشکارساز تخصصی را به‌منظور دستیابی به یک سطح قابل مقایسه‌ای از عملکرد آموزش می‌دهد. باین حال، به‌طور کلی در [15، 16] مشخص شده است که یافتن مشخصات به دلیل حفظ تفکیک زیراالگوها (مانند تقارن چشم به چشم در دید روبه جلو) بیشتر مستعد خطا است. آشکارساز بافت محلی با موفقیت بر این مشکلات غلبه می‌کند.

آزمایش‌های اضافی جهت بررسی تاثیر وضوح تصویر در دسترس انجام شد. برای این منظور هر قاب، از رزولوشن اصلی (720 × 576 که تقریباً مربوط به وضوح تصویر PAL است) به 360 × 288 ("نیمه PAL") و به 180 × 144

پیکسل (" PAL چهارم ") کاسته می‌شود. براین اساس، رزولوشن در دسترس چهره از 64×48 به 32×24 و 12 \times 16 پیکسل کاهش می‌یابد. شکل 8 یک قاب را نشان می‌دهد که در آن هر سطر مربوط به رزولوشن‌های مختلف است (پس از آن تمام فریم‌ها برای بصری‌سازی اهداف تغییر اندازه داده شده‌اند).



شکل 8

ستون سمت چپ تشخیص را بر اساس بافت محلی نشان می‌دهد، ستون سمت راست روش شی محور است. در رزولوشن نیمه PAL (ردیف وسط) آشکارساز شی محور کمتر تغییرات چهره را احتمال می‌دهد، این مورد ناشی از تغییرات ژست فرد است. به‌عنوان نتیجه نمی‌تواند چهره را در رزولوشن‌های پایین‌تر تشخیص دهد. از سوی دیگر آشکارساز بافت محلی تنها بر روی جزئیات صورت تکیه می‌کند و هنوز هم می‌تواند تمام چهره‌ها را با موفقیت تشخیص دهد. وضعیت تشدید برای رویکرد شی محور مانند رزولوشن بیشتر کاهش یافته است. در رزولوشن چهارم PAL هیچ تشخیصی را بر نمی‌گرداند در این مثال رویکرد بافت محلی دوباره تمام چهره را تشخیص می‌دهد.

گزارشی چند از این آزمایش توسط نمودارها در وسط و سمت راست شکل 7 منحنی ROC هر دو آشکارساز را زمانی که به مجموعه داده پایینی اعمال می‌شود نشان می‌دهد. در رزولوشن نیمه PAL (شکل وسط) آشکارساز شی محور بازده حدود 70 درصدی از چهره را با 5 اخطار کاذب بیان می‌کند که در مقایسه با وضوح کامل 15٪ است. که به‌طور

مستقیم با کاهش جزئیات صورت تحت تاثیر قرار می‌گیرد. بنابراین، عملکرد آشکارساز بافت محلی با توجه به تعداد یکسان در 95٪ موارد پایدار باقی می‌ماند. این اثر در رزولوشن چهارم PAL قوی‌تر است. ROC مربوطه در سمت راست همان شکل نشان داده شده است. در مورد PAL چهارم رویکرد شی محور تنها 10 درصد در 1 هشدار اشتباه در هر فریم را تشخیص می‌دهد درحالی‌که آشکارساز بافت محلی موفق به تشخیص بیش از 90٪ از چهره است. به‌طور کلی آشکارساز بافت محلی بهبودی با نرخ تشخیص 15٪، 25٪ و 80٪ در سطح هشدار کاذب را فراهم می‌کند.

امکان کار کردن در رزولوشن پایین سرعت بالای قابل توجهی در جست‌وجوی صورت ارائه می‌کند. برای تصویر، مواردی را در نظر بگیرید که می‌خواهیم 80 درصد از چهره با کمتر از 5 مثبت کاذب در هر فریم را به‌دست آوریم. استفاده از آشکارساز چهره تنها در بالاترین رزولوشن ممکن است (PAL کامل)، که در آن پردازش تک فریم 1.2 ثانیه طول می‌کشد. بنابراین، در حال حاضر آشکارساز بافت محلی به این دقت در روش رزولوشن چهارم PAL در 0.08 ثانیه در هر فریم رسیده است. این مربوط به افزایش 15 برابری سرعت است.

در اینجا باید تاکید کرد که آشکارساز بافت محلی مرتبا بهینه‌سازی نمی‌شود (به‌عنوان مثال، توسط تست پارامترهای متفاوت آموزش) و در نتیجه نتایج گزارش می‌تواند احتمالا بیشتر بهبود یابد.

3.2. فضای باز

جهت بررسی مناسب بودن آنها برای بصری‌سازی، هر دو آشکارساز در محوطه پارکینگ PETS2000 به کار گرفته شد. توالی ویدیو محوطه یک پارکینگ را با کمک تنها دوربین ثابت از بالا نشان می‌دهد. این ویدئو نشان می‌دهد که اتومبیل‌ها و افراد به محوطه وارد می‌شوند و از محوطه خارج می‌شوند. هر 10 فریم مورد استفاده قرار می‌گیرد که در نتیجه در 146 قاب محتوی 210 چهره است. هر فریم دارای رزولوشن 576×756 پیکسل است. چهره‌ها به کوچکی 12×14 پیکسل هستند.

شکل 9 یک قاب نمونه از دنباله با تشخیص آشکارساز بافت محلی نشان می‌دهد. این وضعیت بسیار مشکل‌تر از تنظیمات ویدئو کنفرانس است: پس‌زمینه پیچیده است و به ناچار منجر به هشدارهای نادرستی می‌شود. علاوه بر این،

چشم‌انداز، از دوربین که به سمت پایین است و مردمی که بعضی مواقع قسمت فوقانی بدن خود را به اطراف برمی‌گردانند دشوار است.



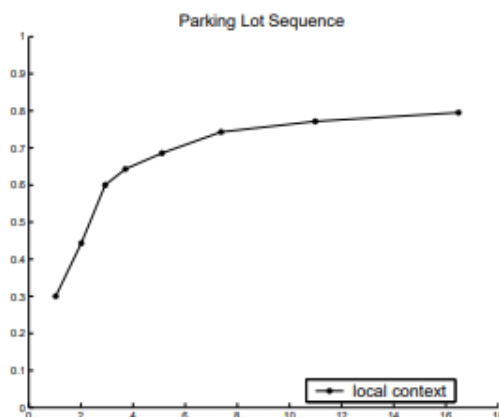
شکل 9

حداکثر منطقه تحت پوشش توسط چهره‌ها تنها 0.1٪ در هر فریم در مقایسه با 2٪ در فضاهای داخلی است. این به این معنی است که حداقل 1000000 ناحیه کاندید برای به‌دست آوردن نرخ هشدار غلط قابل قبولی باید در هر فریم به درستی رد شود. توجه داشته باشید که پاهای مردم در این فضا گاهی اوقات بسته است، به‌عنوان مثال زمانی که آنها در بین خودروهای پارکینگ شده راه می‌روند. این مسئله موجب دشواری سناریو تشخیص عابر پیاده در روش [5] می‌شود. با این حال، قسمت فوقانی بدن آنها به‌طور دائم قابل مشاهده است که موجب می‌شود تشخیص بافت محلی به سهولت انجام گیرد.

منحنی ROC برای آشکارساز بافت محلی در شکل 10 نشان داده شده است. روش شی محور به‌طور کامل در این فضا شکست خورده است: این امر نیاز به تشخیص هر گونه چهره با وضوح دوبرابری تصویر در دسترس دارد. از این رو تنها ROC آشکارساز بافت محلی در اینجا قابل مشاهده است. همانطور که دیده می‌شود، حدود 75٪ از چهره واقعی با 7 هشدار کاذب در هر فریم را بازیابی می‌کند. در حال حاضر نشان می‌دهد که آشکارساز بافت محلی فراتر از توانایی‌های آشکارساز شی محور است. علاوه بر این، در یک نرم‌افزار بصری می‌تواند تعداد بیشتری هشدار کاذب را با کمک حذف پس‌زمینه ([8]) کاهش دهد.

4. خلاصه و نتیجه گیری

در این مقاله عملکرد آشکارساز بافت به‌عنوان وسیله‌ای جهت پیدا کردن چهره انسان مورد بررسی قرار گرفت. آشکارساز با استفاده از یک چارچوب که بخشی از کتابخانه باز بصری کامپیوتر و آزادانه در دسترس است پیاده‌سازی شده است.



شکل 10

این رویکرد در دو دنباله‌ی متفاوت PETS مورد بررسی قرار گرفت: یک فضای داخلی (ویدئو کنفرانس) و یک فضای خارجی (محوطه پارکینگ). به کمک تحلیل ROC مشاهده شد که بافت محلی بهتر از آشکارساز صورت عمل می‌کند. که عمدتاً به دلیل استحکام بیشتر آن در تغییرات مطرح شده، تنوع در ظاهر فردی و همچنین وضوح پایین تصویر است. عملیات مقاوم در وضوح کم نه‌تنها فرایند جستجو را تسریع می‌کند، بلکه از توجه خاصی برای نظارت در جاهایی که اغلب عکس نزدیکی از افراد در دسترس نیست برخوردار است. تجزیه و تحلیل در فضای باز نشان می‌دهد که آشکارساز بافت محلی فراتر از قابلیت‌های تشخیص چهره شی‌محور و استنتاج مستقیم مکان از چهره انسان به کوچکی 9×12 پیکسل است.

تشکر و قدردانی

نویسنده دوم توسط Beleyma و Unelco از طریق دانشگاه Fundacion Canaria، و توسط پروژه‌های تحقیقاتی 042/ 2000PI از Gobierno de Canarias و 16/ 2002UNI از دانشگاه de Las Palmas de Gran Canaria حمایت می‌شد.