بررسی روش‌های تشخیص چهره مبتنی بر الگوریتم‌های یادگیری عمیق

پژمان غلام‌نژاد*
دانشگاه علوم و فنون هوایی شهید ستاری، تهران، ایران
pezhman.gholamnezhad@gmail.com

احسان شریفی
دانشگاه علوم و فنون هوایی شهید ستاری، تهران، ایران
sharifi@ssau.ac.ir

تاریخ دریافت: ۱۴۰۰/۰۴/۲۵
تاریخ اصلاحات: ۱۴۰۱/۰۸/۰۹
تاریخ پذیرش: ۱۴۰۱/۰۹/۰۶

چکیده

امروزه با رشد فناوری اطلاعات، تشخیص چهره یک مسأله چالش برانگیز در زمینه تجزیه و تحلیل تصویر و بینایی رایانه است و به همین دلیل در چند سال گذشته به خاطر کاربردهای فراوان در حوزه‌های مختلف مورد توجه بسیاری قرار گرفته است. روش‌های زیادی برای پیاده‌سازی این فناوری مورد استفاده قرار می‌گیرد، اما روش کلی بر مبنای مقایسه مشخصه‌های خاصی از چهره افراد با یک پایگاه داده یا مجموعه اطلاعاتی از پیش ذخیره شده (که می‌تواند حاصل نمونه‌گیری از چهره افراد باشد) است. فناوری‌های مبتنی بر بیومتریک در سال‌های اخیر به‌عنوان امیدوارکننده‌ترین گزینه برای تشخیص هویت افراد، شناخته شده‌اند. به منظور پیاده‌سازی تشخیص چهره از روش‌های متفاوتی استفاده می‌شود. در این مقاله، مروری بر برخی از روش‌های شناخته‌شده پردازش تصویر انجام می‌شود و مزایا و معایب طرح‌های ذکرشده در آن بررسی شده است. همچنین فناوری‌های پیاده‌سازی سیستم‌های تشخیص چهره معرفی می‌گردد. سپس الگوریتم‌های تشخیص چهره براساس مشخصه‌های بیومتریک دسته‌بندی و معرفی می‌شوند. علاوه بر این، ضمن معرفی الگوریتم‌های مدل سلسله مراتبی و ایکس، مدل سلسله مراتبی باینری و ایکس و الگوریتم یُولُو، به بیانِ مفهوم ساختار تشخیصِ چهره‌ی عمیق، پرداخته شده است و برخی از جدیدترین الگوریتم‌های تولید شده برای این منظور ذکر شده است. در پایان، برخی از مهم‌ترین موارد کاربرد سیستم‌های تشخیص چهره مورد بررسی قرار گرفته است. هدف از این مقاله، معرفی و بیان الگوریتم‌های یادگیری عمیق در تشخیص چهره و بیان چالش‌های موجود می‌باشد.

واژگان کلیدی

تشخیص چهره عمیق؛ الگوریتمِ HMAX؛ الگوریتمِ B-HMAX؛ روش‌های تشخیصِ چهره؛ فناوریِ تشخیصِ چهره مبتنی بر بیومتریک.

۱- مقدمه

امروزه اکثر فعالیت‌ها در ادارات با کامپیوترها و با استفاده از نرم‌افزارها و برنامه‌های سازمان‌یافته‌ی داخلی انجام می‌شود. یکی از این مسائل مهم که شاخه‌ای از هوش مصنوعی است، پردازش تصویر می‌باشد. با پیشرفت روزافزون فناوری اطلاعات امکان ثبت، پردازش و انتقال تصاویر به صورت دیجیتال ممکن شده است. پردازش تصاویر امروزه بیشتر به موضوع پردازش تصویر دیجیتال گفته می‌شود که شاخه‌ای از دانش رایانه است که با پردازش سیگنال‌های دیجیتال که از طریق راه‌های مختلف بدست می‌آید، سر و کار دارند.

بیش از ۹۰ درصد اطلاعات پیرامون ما به وسیله‌ی مشاهده، صورت می‌پذیرد. تصاویر به صورت دیجیتال میسر شده است. پردازش تصاویر[1] امروزه بیشتر به موضوع پردازشِ تصاویر دیجیتال گفته می‌شود که شاخه‌ای از دانش رایانه است که با پردازشِ سیگنالِ دیجیتال که نماینده تصاویر برداشته شده با دوربین دیجیتال یا پویش‌شده توسط پویشگر هستند سر و کار دارد.

پردازش تصویر به صورت خلاصه بدان معنا است که با انجام فرایندهایی روی تصویر (چه عکس چه ویدئو)، خروجی مشخص و معینی از تصویر (باز هم به صورت تصویر) بگیرید [۱]. این فرایندها می‌توانند برای افزایش دقت تصویر[2] باشند یا می‌توانند برای جداکردن یک بخش مشخص از تصویر[3] باشند و ... . برنامه‌هایی مانند اینستاگرام، کم‌اسکنر[4] و ... از الگوریتم‌های مربوط به پردازش تصویر استفاده می‌کنند.

در حال حاضر در دنیا یکی از جذاب‌ترین بخش‌های پردازش تصویر، کاربرد تشخیص چهره[5] است.

در ادبیات تشخیص چهره، مفهومی مرتبط وجود دارد که از آن به‌عنوان مکان‌یابی چهره یاد می‌کنیم. هدف از مکان‌یابی چهره درست

1. Image Processing

2. Image Enhancement
3. Image Segmentation
4. Cam Scanner
5. Face Recognition

* نویسنده مسئول

همانند تشخیص چهره هست اما تفاوت اندکی موجود خواهد بود و آن این که در مکان‌یابی چهره تصویر موجود فقط شامل یک چهره در نظر گرفته می‌شود. یکی از روش‌های مرسوم در زمینه تشخیص اشیاء در نظر گرفتن قابی و چک روی تصویر اصلی و تشخیص این خواهد بود که آیا شیء مورد نظر در آن پنجره وجود دارد یا نه؟ پس اگر از این روش استفاده شود باید در جستجوی الگوریتمی بود تا توانایی تشخیص وجود یا عدم وجود چهره در یک قاب کوچک، متشکل از چند صد پیکسل داشته باشد. در این دیدگاه تشخیص چهره را می‌توان به صورت مسأله‌ی دسته‌بندی نیز در نظر گرفت. به این صورت که عامل هوش مصنوعی باید قاب‌های مختلف موجود در تصویر را در دو گروه چهره و غیرچهره در نظر گرفت. المان‌های مختلفی را می‌توان در ارزیابی یک سیستم تشخیص چهره مؤثر دانست مانند زمان یادگیری، زمان اجرا، تعداد مثال‌های مورد نیاز برای یادگیری و نسبت بین میزان تشخیص و خطای منفی. میزان تشخیص را می‌توان به نسبت تعداد چهره‌های درست تشخیص داده شده توسط عامل هوش مصنوعی به تعداد چهره‌های تشخیص داده شده توسط انسان تعریف کرد. در صورتی قابی توسط عامل تشخیص داده شده است به‌عنوان چهره در نظر گرفته می‌شود که قاب مورد نظر بیشتر از میزان خاصی از چهره‌ی فرد را پوشش دهد. از طرف دیگر خطای منفی زمانی رخ می‌دهد که عامل در تشخیص چهره ناموفق باشد که این خود ریشه در پایین‌بودن میزان تشخیص خواهد بود. در مقابل خطای منفی مفهوم دیگری به نام خطای مثبت وجود دارد که وقتی قابی به‌عنوان چهره از طرف عامل هوش مصنوعی معرفی می‌شود اما عامل انسانی تأیید نمی‌کند، رخ می‌دهد. فناوری تشخیص چهره وِری لوک[1] برای توسعه‌دهندگان و مجتمع‌سازان سیستم‌های بیومتریکی در نظر گرفته شده است [۲]. این فناوری بواسطه تشخیص چهره زنده، تشخیص چندین چهره بصورت همزمان و همچنین تطبیق سریع چهره در حالت‌های یک به یک و یک به چند، میزان کارایی، عملکرد و قابلیت اطمینان سیستم را تضمین می‌نماید. روش‌های تشخیص چهره که مبتنی بر ساختار چهره هستند، روش‌های بدون نظارتی می‌باشند که نسبت به تغییرات خطی که در تصویر رخ می‌دهد نتایج مناسبی را تولید می‌کنند. تحلیل مؤلفه‌های اساسی[2] یک تبدیل خطی است که ابزار قدرتمندی برای تجزیه و تحلیل داده‌هایی است که دارای تغییرات خطی می‌باشند [۳]؛ ولی برای تغییرات غیرخطی چهره ناشی از تغییرات حالت، روشنایی و ژست در تصویر چهره، مطلوب نمی‌باشد. فیلتر گابور یکی از روش‌های مبتنی بر ویژگی است که می‌تواند برای رفع نقطه ضعف تحلیل مؤلفه‌های اساسی[3] مورد استفاده قرار گیرد. روشی جدید برای تشخیص چهره با ترکیب روش‌های تحلیل مؤلفه‌های اساسی[4] و گابور ارایه شده است [۴]. بدین صورت که پس از اعمال فیلتر گابور بر روی هر چهره موجود در بانک اطلاعاتی، تعدادی تصویر حاصل از فیلتر گابور به‌دست می‌آید. میانگین تصاویر حاصل از فیلتر گابور به‌عنوان یک تصویر جدید در نظرگرفته می‌شود؛ سپس از مؤلفه‌های اصلی به‌دست آمده از اعمال تحلیل مؤلفه‌های اساسی[5] بر روی تصاویر میانگین برای تشخیص چهره استفاده می‌گردد. روش پیشنهادی بر روی پایگاه‌های یِل بی[6] و آ آرال[7] داده تصویر تحت شرایط مختلف بررسی گردید.

در این مقاله، مروری بر برخی از روش‌های شناخته‌شده پردازش تصویر انجام می‌شود و مزایا و معایب طرح‌های ذکرشده در آن بررسی شده است. همچنین فناوری‌های پیاده‌سازی سیستم‌های تشخیص چهره معرفی می‌گردد. سپس الگوریتم‌های تشخیص چهره براساس مشخصه‌های بیومتریک دسته‌بندی و معرفی می‌شوند. علاوه بر این، ضمن معرفی الگوریتم‌های مدل سلسله مراتبی و ایکس، مدل سلسله مراتبی باینری و ایکس و الگوریتم یُولُو، به بیانِ مفهوم ساختار تشخیصِ چهره‌ی عمیق، پرداخته شده است و برخی از جدیدترین الگوریتم‌های تولیدشده برای این منظور ذکر شده است.

۲- کاربردهای پردازش تصویر

زمینه‌های مختلف کاربرد پردازش تصویر عبارتند از صنعت، هواشناسی، شهرسازی، کشاورزی، علوم نظامی و امنیتی، نجوم و فضانوردی، پزشکی، فناوری‌های علمی، باستان‌شناسی، تبلیغات، سینما، اقتصاد، روان‌شناسی و زمین‌شناسی [۵].

شناسایی چهره توسط رایانه یکی از جذاب‌ترین زمینه‌های تحقیقات بیومتریک است که زمینه‌های علمی مختلفی از جمله بینایی ماشین، هوش محاسباتی، شناسایی الگو و روانشناسی را در بر می‌گیرد.

سیستم‌های شناسایی چهره کاربردهای متعددی دارند که می‌توان به‌طور خلاصه آن‌ها را به صورت زیر طبقه‌بندی کرد: تشخیص هویت با استفاده از کارت‌های اعتباری، گذرنامه و ...، استفاده از تصویر سابقه‌داران، برای تشخیص مجرمین. حفاظت از بانک‌ها و فروشگاه‌ها با آشکارسازی صورت در تصاویر، یافتن صورت در یک صحنه شلوغ.

۳- تشخیص چهره

تشخیص چهره یا به مجموعه علوم و فناوری‌هایی گفته می‌شود که هدف آن تشخیص هرچه دقیق‌تر و با سرعت بیشتر چهره انسان است. سیستم‌های تشخیص بصری[8] به سیستم‌هایی اطلاق می‌شود که قابلیت تشخیص یا تصدیق چهره افراد را از روی یک تصویر تصویر یا فیلم دارا می‌باشند. اگر بخواهیم به صورت علمی‌تر و یا فنی‌تر صحبت کنیم، سیستم تشخیص یک کاربرد بیومتریک مبتنی بر هوش مصنوعی است که

1. Veri Look
2. PCA
3. PCA
4. PCA
5. PCA
6. YaleB
7. ORL
8. Facial Recognition Systems

می‌تواند به صورت منحصر به فردی افراد را از طریق مقایسه الگوهای مبتنی بر بافت‌ها و هندسه صورت تشخیص دهد.

روش‌های زیادی برای پیاده‌سازی این فناوری مورد استفاده قرار می‌گیرد، اما روش کلی بدین شکل است که مشخصه‌های خاصی از چهره افراد با یک پایگاه داده یا مجموعه اطلاعاتی از پیش ذخیره شده (که می‌تواند حاصل نمونه‌گیری از چهره افراد باشد) مقایسه می‌شود.

درحالی‌که به صورت پایه این فناوری معمولاً در کامپیوترهای پردازشگر پیاده‌سازی می‌شود، اما با گسترش و توسعه الگوریتم‌ها و بهینه‌سازی‌های صورت گرفته بر روی آن‌ها، امروزه شاهد پیاده‌سازی و سفارشی‌سازی این الگوریتم‌های بهینه‌شده برای پلتفرم‌های دیگر مانند گوشی‌های هوشمند هستیم که کاربردهای بسیار جذابی را در آینده در این حوزه رقم خواهد زد.

فناوری‌های مبتنی بر بیومتریک در سال‌های اخیر به‌عنوان امیدوارکننده‌ترین گزینه برای تشخیص هویت افراد، به جای احراز هویت افراد بوسیله دسترسی آن‌ها به حوزه‌های فیزیکی و مجازی براساس کلمه عبور، پین کُد[1]، کارت‌های هوشمند، کارت‌های اعتباری، کلیدها و غیره شناخته شده‌اند؛ این روش‌ها، تعیین و یا تأیید هویت، ویژگی‌های روانی و یا رفتاری یک فرد را مورد بررسی قرار می‌دهند.

به خاطر آوردن کلمات عبور و پین کُدها کار سخت و دشواری می‌باشد و امکان سرقت یا حدس آن‌ها وجود دارد؛ کارت‌ها، توکن‌ها، کلیدها و مواردی از این دست احتمال دارد که گم شده و یا آن‌ها را دزدید یا جعل نمود؛ کارت‌های مغناطیسی می‌توانند خراب و غیرقابل خواندن شوند. اما، صفات بیولوژیکی یا زیستی یک فرد را نمی‌توان گم کرد، فراموش نمود، دزدید یا جعل کرد.

فناوری‌های مبتنی بر بیومتریک عبارتند از [۶]: شناسایی براساس ویژگی‌های روانی (نظیر چهره، اثرانگشت، شکل هندسی انگشت، شکل هندسی دست، رگ‌های دست، کف دست، عنبیه چشم، شبکیه چشم، گوش و صدا) و صفات رفتاری (نظیر شمار و اندازه گام، امضا و دینامیک ضربه کلید).

۳-۱- فناوری‌های پیاده‌سازی سیستم‌های تشخیص چهره

به منظور پیاده‌سازی تشخیص چهره از روش‌ها و الگوریتم‌های متفاوتی استفاده می‌شود. این الگوریتم‌ها را می‌توان به‌طور کلی در پنج گروه دسته‌بندی کرد. این ندی شامل: روش‌های سنتی، روش‌های ابعادی، روش دوربین حرارتی، روش مقایسه بافت پوست، روش‌های ترکیبی می‌باشند.

۳-۱-۱- روش‌های سنتی

روش‌هایی که در این بخش استفاده می‌شود، به دو دسته تقسیم‌بندی می‌شوند: تشخیص برمبنای مشخصات چهره، تشخیص بر مبنای مشخصات نرمالیزه شده.

تشخیص بر مبنای مشخصات چهره: الگوریتم‌هایی که در این گروه قرار می‌گیرند، مبنای عملکرد آن‌ها مقایسه مشخصاتی مانند موقعیت، اندازه و شکل اجزای صورت مانند چشم‌ها، بینی، دماغ و ... با مشخصات ثبت شده در پایگاه داده است. از جمله مزیت‌های این الگوریتم‌ها عبارت است از سرعت بالا و سادگی نسبی در پیاده‌سازی آن‌ها. همان‌طور که از نام این دسته نیز مشخص است، این الگوریتم‌ها جزء اولین روش‌های پیاده‌سازی سیستم‌های تشخیص چهره بوده است.

مبنای عملکرد این گروه از الگوریتم‌ها مقایسه تنها برخی از مشخصات چهره با تصاویر موجود در پایگاه داده است. در واقع، پس از بررسی تصاویر موجود در پایگاه و نرمالیزه‌کردن مشخصات آن‌ها، فشرده‌سازی بر روی ویژگی‌های منتخب‌شده انجام شده و سپس این مشخصات با مشخصات موجود در تصویر چهره مقایسه می‌شود.

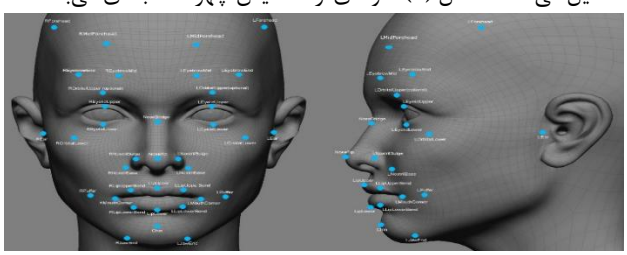
۳-۱-۲- روش‌های ابعادی (تشخیص چهره سه بعدی)

در این روش از داده‌های سنسور سه بعدی به منظور به‌دست آوردن اطلاعاتی در مورد شکل کلی چهره استفاده می‌شود. در مرحله بعد از این مشخصه‌ها به منظور تشخیص ویژگی‌های برجسته‌کننده سطح چهره مانند تشخیص مرزهای چشم‌ها، بینی و چانه و ... استفاده می‌شود.

یکی از ویژگی‌های روش تشخیص چهره سه بعدی این است که مانند دیگر روش‌ها نور محیط تأثیری بر آن نمی‌گذارد. از دیگر مزیت‌های آن قابلیت تشخیص چهره از زاویه‌های مختلف است.

نقاط داده به‌دست آمده از این روش می‌تواند کمک زیادی به بالارفتن دقت تشخیص چهره شود. استفاده از این روش همزمان با ساخته‌شدن سنسورهای سه بعدی پیشرفته که می‌توانند دقت بالاتری را در ثبت تصاویر سه بعدی ارائه نمایند به‌طور فزاینده‌ای افزایش پیدا کرده است.

روش دیگری که برای ثبت تصاویر سه بعدی استفاده می‌شود استفاده از سه دوربین مجزا است که در مقابل، کنار و در زاویه‌ای معین از چهره قرار می‌گیرند و به صورت سنکرون با یکدیگر تصویر چهره را به صورت همزمان تشکیل می‌دهند. شکل (۱) نمونه‌ای از تشخیص چهره سه بعدی می‌باشد.



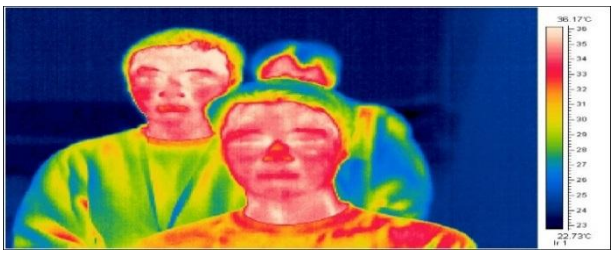
شکل ۱- تشخیص چهره سه بعدی

۳-۱-۳- روش دوربین حرارتی

یکی از روش‌های متفاوت برای به‌دست آوردن تصاویر چهره به منظور اعمال به سیستم تشخیص چهره، استفاده از دوربین‌های حرارتی است. دوربین‌های حرارتی فقط شکل و مشخصات کلی صورت را تشخیص

1. Postal Index Number (PIN) Code

می‌دهد و جزئیات صورت ثبت نمی‌گردد. شکل (۲) یک نمونه از تصویربرداری با دوربین حرارتی را نشان می‌دهد.



شکل ۲- تصویربرداری با دوربین حرارتی

برخلاف دوربین‌های معمولی، دوربین‌های حرارتی قابلیت ثبت تصاویر را در محیط‌هایی با نور بسیار پایین و در شب را دارا می‌باشند و ویژگی جالب دیگری آن‌ها نیز این است که نیازی به فلاش زدن ندارند که باعث می‌شود موقعیت دوربین پنهان بماند.

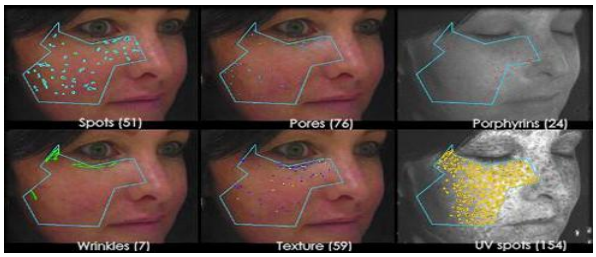
سنسورهای حرارتی مادون قرمز دارای قابلیت تصویربرداری از صحنه‌ها و اشیاء براساس بازتاب نور مادون قرمز یا تابش تابش مادون قرمز می‌باشد. تابش مادون قرمز، تابشی الکترومغناطیسی است که متناسب با گرمای تولیدشده / منعکس‌شده توسط یک جسم ساطع می‌شود و بنابراین از تصویربرداری مادون قرمز به‌عنوان تصویربرداری حرارتی یاد می‌شود. طول موج‌های مادون قرمز بیشتر از نور مرئی است، بنابراین مادون قرمز برای انسان نامرئی است. از طرف دیگر، دوربین‌های مادون قرمز نسبت به تغییرات دمای محیط بسیار حساس هستند و جزئیات کمتری نسبت به دوربین‌های نور مرئی ارائه می‌دهند، زیرا رنگی که در طیف مرئی گرفته می‌شود اطلاعات بسیار بیشتری را فراهم می‌کند و تفسیر آن آسان‌تر است. در تشخیص چهره با دوربین‌های حرارتی با روش‌های دمانگاری[1] سر و کار دارند. در مرجع [۷] مروری بر انواع روش‌های تشخیص چهره براساس دمانگاری صورت گرفته است.

یکی از مشکلات استفاده از تصاویر دوربین‌های حرارتی برای سیستم‌های تشخیص چهره عدم وجود پایگاه‌های داده مناسب برای این منظور است. اخیراً برخی از محققان در این زمینه فعالیت‌های خوبی انجام داده‌اند که منجر به ساختن یک پایگاه داده مناسب برای تحلیل تصاویر دوربین‌های حرارتی شده است.

۳-۱-۴- روش تحلیل بافت پوست

در این روش از جزئیات بصری پوست که در تصاویر استاندارد دوربین‌های دیجیتال یا تصاویر اسکن شده وجود دارد استفاده می‌کند. این روش که به نام تحلیل بافت پوست[2] معروف است جزئیاتی مانند خطوط، الگوها و نقاط موجود بر روی پوست صورت را از طریق تبدیلات ریاضیاتی به معادلات ریاضی تبدیل می‌کند [۸].

مراحل اجرای این روش به این شکل است که تصویری از بخشی از چهره[3] تهیه می‌شود، سپس این تصویر به اجزای کوچک‌تری تبدیل می‌شود و سپس توسط تبدیلات ریاضی این تکه تصویرها به فضای معادلات ریاضیاتی برده می‌شوند و از طریق این معادلات امکان تشخیص و تمیز بین خطوط، منافظ و بافت تصاویر وجود خواهد داشت. به‌عنوان مثال از جمله مزیت‌های این تکنیک می‌توان به تشخیص تصاویر دوقلوهای شبیه به هم اشاره کرد که در نرم‌افزارهای رایج تشخیص چهره امکان‌پذیر نمی‌باشد. شکل (۳) یک نمونه از تحلیل بافت پوست را نشان می‌دهد.



شکل ۳- تحلیل بافت پوست

نتایج تحقیقات نشان می‌دهند که استفاده از روش تحلیل بافت پوست می‌تواند باعث افزایش ۲۰ تا ۲۵ درصدی کارایی الگوریتم‌های تشخیص چهره شود.

۳-۱-۵- روش‌های ترکیبی

هر یک از روش‌های قبلی، دارای مزایا و معایبی است. به همین دلیل شرکت‌هایی که بر روی پیاده‌سازی الگوریتم‌ها و ساخت محصولات تشخیص چهره فعالیت دارند سعی می‌کنند به منظور بهینه‌سازی الگوریتم‌های خود از ترکیبی از روش‌های معرفی شده استفاده نمایند. به‌عنوان مثال بسیاری از شرکت‌ها از ترکیبی از روش‌های تشخیص سه بعدی و تحلیل بافت چهره استفاده می‌کنند تا بتوانند قابلیت اطمینان سیستم خود را افزایش دهند.

از جمله ویژگی‌هایی که در اثر ترکیب روش‌های تشخیص چهره به‌وجود می‌آید عبارت است از عدم حساسیت سیستم به حالات مختلف صورت (لبخند، اخم، چمشک‌زدن و ...) و همچنین تغییر حالات ریش و سبیل و عینک است.

۳-۲- چالش‌ها

چالش‌های مرتبط با تشخیص چهره را می‌توان به عوامل زیر نسبت داد [۹و۱۰]:

- حضور و یا عدم حضور مؤلفه‌های ساختاری: ویژگی‌های چهره نظیر ریش، سبیل و عینک می‌توانند وجود داشته و یا نداشته باشد و تغییرپذیری زیادی میان این مؤلفه‌ها از جمله شکل، رنگ و اندازه وجود دارد.

1. Thermogramy
2. Skin Texture Analysis
3. Patch

- فرم صورت و قیافه: تصاویر چهره به خاطر فرم و حالت چهره دوربین تغییر می‌کنند.
- حالت چهره و هیجانات: ظاهر چهره مستقیماً با حالت چهره و هیجانات شخص، تشخیص داده می‌شود.
- انسداد یا احتباس: چهره‌ها ممکن است تا حد زیادی با اشیاء دیگر محو شوند و به‌طور کامل رؤیت نشوند. به‌طور مثال، در یک تصویر با گروهی از افراد، برخی از چهره‌ها ممکن است در حد جزئی چهره‌های دیگر را مسدود نمایند و تصویر برخی از افراد به‌طور کامل دیده نشود.
- جهت و موقعیت تصویر: تصاویر چهره برای چرخش‌های متفاوت، حول محور اپتیکی دوربین، مستقیماً تغییر می‌کنند.
- شرایط تصویربرداری: زمانی‌که تصویر تشکیل می‌شود، عواملی نظیر نور و ویژگی‌های دوربین ظاهر چهره را مشخص می‌کنند.
- فرسایش تصویر بدلیل قدمت آن: تصاویر گرفته شده بعد از شکاف یک یا دو سال ممکن است با تصاویر موجود در پایگاه داده مطابقت نداشته باشند.

۴- الگوریتم‌های تشخیص چهره

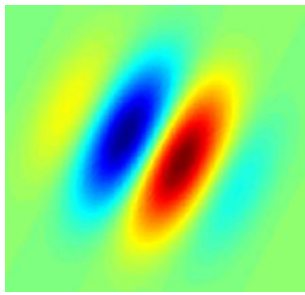
بیومتریک‌ها مشخصه‌های منحصر به فرد و قابل اندازه‌گیری برای شناسایی هویت افراد هستند. مشخصه‌های بیومتریک به دو دسته اصلی تقسیم می‌شوند: مشخصه های فیزیولوژی: برای اندازه‌گیری مشخصات فیزیکی خارجی. مشخصه‌های رفتاری: برای اندازه‌گیری و شناسایی رفتارهای یادگیری شده.

چهره نقش اساسی در شناسایی افراد و نمایش احساسات آن‌ها در سطح جامعه دارد. بازشناسی چهره نسبت به دیگر روش‌های بیومتریک مانند تشخیص اثر انگشت و عنبیه، مزایایی دارد. طبیعی و غیرقابل بروزبودن این نوع بازشناسی، مهم‌ترین مزیت بازشناسی چهره است که می‌تواند در هر فاصله‌ای گرفته و پوشش داده شود.

۴-۱- الگوریتم HMAX[1]

ساختار HMAX شامل یک توالی متناوب از سلول‌های ساده و لایه‌های سلول‌های پیچیده است. لایه‌های سلول‌های ساده (S1 و S2) برای استخراج ویژگی‌ها از تصویر ورودی هستند و لایه‌های سلول‌های پیچیده (C1 و C2)، باعث عدم تغییر مکان و مقیاس آن ویژگی‌ها می‌شوند. لایه اول، S1، از یک بانک فیلترهای گابور[2] تشکیل شده است که هر یک به یک جهت و مقیاس خاص حساس هستند [۱۱].

در پردازش تصویر، یک فیلتر گابور که به نام دنیس گابور نامگذاری شده است، یک فیلتر خطی است که برای تحلیل بافت استفاده می‌شود، به این معنی که تحلیل می‌کند که آیا محتوای فرکانسِ خاص در تصویر، در جهتِ خاص، در یک منطقه محلی در اطراف نقطه یا منطقه تجزیه و تحلیل وجود دارد [۱۲]. بسیاری از دانشمندان دیدگاه معاصر ادعا می‌کنند که فرکانس و جهت‌گیری نمایش‌های فیلترهای گابور شبیه به سیستم بصری انسان می‌باشند، هرچند هیچ شواهد تجربی و هیچ منطقی عملی برای حمایت از این ایده وجود ندارد. آن‌ها به ویژه برای نمایش و تبعیض بافت مناسب هستند. همچنین در حوزه فضایی، یک فیلتر گابور دوبعدی، یک تابع هسته گاووسی است که توسط یک موج مسطح سینوسی مدولاسیون شده است. شکل (۴)، یک نمونه از فیلتر گابور دوبعدی را نشان می‌دهد.



شکل ۴- یک مثال از فیلتر گابور دوبعدی

تابع گابور دوبعدی و تبدیل فوریه آن بصورت زیر است:

$$g(x,y)=exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)+2\pi jWx\right]$$

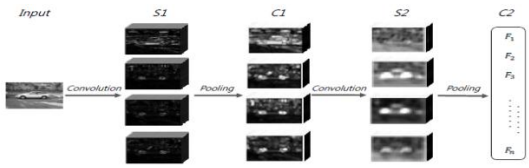
$$G(u,v)=exp\left\{-\frac{1}{2}\left[\frac{(u-W)^2}{\sigma_u^2}+\frac{v^2}{\sigma_v^2}\right]\right\}$$

$$\sigma_u=\frac{1}{2\pi\sigma_x},\sigma_v=\frac{1}{2\pi\sigma_y}$$

به منظور فراهم‌شدن سطح اول عدم تغییر در مکان‌ها و مقیاس‌ها، هر واحد از C1، از همسایگی واحدهای S1، دو تصویر در جهت یکسان و دو مقیاس متوالی، ترسیم می‌کند و حداکثر مقدار را استخراج می‌نماید.

قبل از استفاده از مدل، لایه سوم نیاز به یک مرحله آموزش دارد تا مؤثر واقع شود. در حین مرحله آموزش، چندین تکه در اندازه‌های مختلف از تصاویر از C1، در موقعیت‌ها و مقیاس‌های تصادفی بریده می‌شوند. هر تکه در چهار موقعیت در همان موقعیت و مقیاس بریده می‌شود و تمام مقادیر عناصر برش خورده به منظور تشکیل یک بردار به صورت سریالی در می‌آیند. بنابراین، برش تکه‌ی $S \times S$ منجر به تولید بردار ۴ مؤلفه‌ای می‌شود. در طول تغذیه رو به جلو ، تکه‌ها از نتیجه C1، در تمام مکان‌ها و مقیاس‌ها برش داده می‌شوند و تمام واحدهای S2 با اندازه‌های مربوطه را تغذیه می‌کنند.

سرانجام یک روش برای انتخاب بهترین نتایج از S2، به‌کار گرفته می‌شود. شکل (۵) مراحل الگوریتم HMAX را نشان می‌دهد.



شکل ۵- الگوریتم HMAX [۱۱]

1. Hierarchical Model And X (HMAX)
2. Gabor

از نظر بازشناسی الگو، مدل HMAX استخراج ویژگی از تصویر است. ویژگی‌هایی که این مدل از تصاویر استخراج می‌کند، نسبت به تغییر موقعیت و مقیاس اشیاء در تصاویر، مقاوم هستند. در مرحله یادگیری مدل HMAX، تعداد زیادی از قسمت‌های تصاویر در اندازه‌های گوناگون که "تکه"، نامیده می‌شود، در موقعیت‌های تصادفی از مجموعه‌های تصاویر آموزشی استخراج می‌شود. مدل HMAX با وجود دقت بالا به دلیل پیچیدگی محاسبات و مقایسه پیکسل‌ها دارای سرعت بالا نیست.

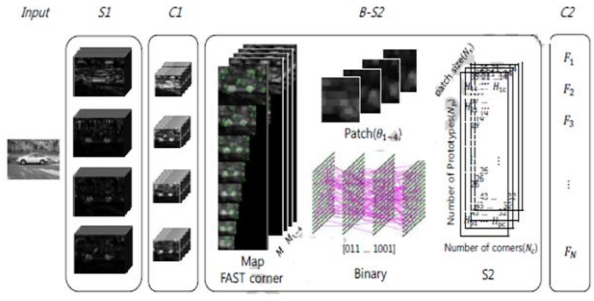
روش HMAX به دلیل آن که در برابر تغییرات اندازه‌ی چهره در تصویر و دوری و نزدیکی تصویر، مقاوم است، در زمینه شناسایی چهره کاربرد فراوان دارد. با توجه به گوناگونی اندازه و ویژگی‌های تصاویر چهره، طول ویژگی‌های استخراج‌شده از تصاویر نیز متغیر است.

۴-۲- الگوریتم B-HMAX[1]

تکه‌های استخراج‌شده در الگوریتم HMAX از اجزای اصلی متمایزکننده می‌باشند و توانایی زیادی در دقتِ تشخیص دارند. در الگوریتم HMAX در لایه‌ی C1، تکه‌ها بصورت تصادفی در فرایند آموزش، استخراج می‌شوند. برخی از این تکه‌ها که بصورت تصادفی انتخاب می‌شوند، فاقد اطلاعات هستند که دقت تشخیص را محدود می‌نمایند.

در این الگوریتم، لایه‌ی C1 و S1، همانند الگوریتم HMAX هستند و لایه‌ی B-S2 ارائه شده است. در این لایه M_i، شامل تمام اطلاعات مقایس‌ها می‌باشد که نگاشت نامیده می‌شود.

همچنین واحدهای S2، با فاصله اقلیدسی بین تکه‌های ذخیره‌شده و تمام نقاط موجود در لایه C1، محاسبه می‌شوند که این امر هزینه‌ی محاسباتی بالایی را دارد. برای رفع این دو مشکل، الگوریتم B-HMAX، ارائه شد که در آن تکه‌های انتخاب‌شده از تمایز بهتری برخوردار هستند و هزینه‌های محاسباتی کمتری دارند [۱۲]. شکل (۶)، ساختار کلی الگوریتم B-HMAX را نشان می‌دهد.



شکل ۶- ساختار کلی الگوریتم B-HMAX [۱۲]

الگوریتم فَست[2]، برای تشخیص گوشه‌ها، به‌عنوان نقاط مورد علاقه استفاده می‌شود که از روش‌های یادگیری ماشین استفاده می‌کند و بسیار سریع‌تر از سایر الگوریتم‌های ردیاب گوشه می‌باشد و قابلیت تکرار عالی دارد و همچنین دارای سرعت بالایی می‌باشد. این الگوریتم براساس تست بخش شتاب[3]، ساخته شده است. در این روش، یک نقطه‌ی کاندیدِ P را که دارای شدت I_p است را به‌عنوان یک گوشه، کلاس‌بندی می‌کند اگر n پیکسلِ مجاور در دایره‌ی برسنهام[4]، روشن‌تر از $I_p + t$ یا تاریک‌تر از $I_p - t$ باشد که t یک حد آستانه‌ی از پیش تعریف شده است. سپس به هر گوشه یک امتیاز داده می‌شود که به‌عنوان بزرگ‌ترین حدِ آستانه برای P تعریف می‌شود.

نقاط کلیدی اطرافِ اشیاء را می‌توان با تعدادِ کمی از نقاطِ گوشه‌ی قوی تشخیص داد. سپس یک رشته بیت از طریقِ مقایسه‌ی یک زوج شدتِ پیکسل‌ها، بصورت تصادفی از یک تکه با ۴ نگاشتِ جهت، انتخاب می‌شوند.

۴-۳- تشخیص چهره عمیق

از سال ۲۰۱۴، یادگیری عمیق[5]، چندین لایه پردازشی را برای یادگیریِ داده‌ها با چندین سطح استخراج ویژگی اعمال می‌کند [۱۳].

در تشخیص چهره عمیق، سه ماژول برای تشخیص چهره نیاز است و بصورت زیر بیان می‌شود:

$$M\left[F\left(P_i\left(I_i\right)\right),F\left[P_j\left(I_j\right)\right]\right]$$

که I_i و I_j، دو تصویر صورت هستند، P مخفف پردازش صورت برای کنترل تغییرات درون شخصی قبل از آموزش و آزمایش است، مانند حالت‌ها، نورها، عبارات و انسدادها. F بیانگر استخراج ویژگی است که اطلاعات هویت را رمزگذاری می‌کند.

در ابتدا از یک کشف‌کننده‌ی چهره برای متمرکزشدن بر روی چهره‌ها در تصاویر یا فیلم‌ها استفاده می‌شود. در مرحله‌ی دوم چهره‌ها با مختصات متعارف عادی، مطابقت داده می‌شوند و در مرحله سوم تشخیص چهره انجام می‌شود که شامل پردازش چهره، استخراج ویژگی‌های عمیق و تطبیق با چهره است.

استخراج‌کننده‌ی ویژگی، توسط تابع زیان[6]، یادگیری را انجام می‌دهند. تابع زیان می‌تواند بر مبنای فاصله اقلیدسی باشد و یا مبتنی بر زاویه/ کسینوس باشد. این روش باعث می‌شود که ویژگی‌های آموخته‌شده با فاصله‌های بزرگ‌تر از زاویه / کسینوس قابل تفکیک باشند. همچنین تابع زیان می‌تواند تابع بیشینه‌ی هموار[7] (تابع نمایی نرمال‌سازی شده) باشد.

M به معنای الگوریتمی همسان‌سازی چهره است که برای محاسبه مقدار شباهت ویژگی‌ها، برای تعیین هویتِ خاص چهره‌ها استفاده می‌شود.

1. Binary-Hierarchical Model And X (B-HMAX)

2. Fast
3. Accelerated Segment Test (AST)
4. Bresenham
5. Deep Learning
6. Loss Function
7. Softmax Function

متفاوت از طبقه‌بندی اشیاء، هویت‌های آزمایش معمولاً از داده‌های آموزش در تشخیص چهره جدا نیستند، که باعث می‌شود از طبقه‌بندی آموخته شده برای تشخیص چهره‌های آزمایش استفاده نشود. بنابراین، الگوریتم تطابق چهره یکی از قسمت‌های اساسی تشخیص چهره است.

در پردازش چهره (مرحله‌ی اول)، شرایط مختلف، مانند حالت‌ها، نورها، عبارات و انسدادها، بر عملکرد تشخیص چهره عمیق تأثیر می‌گذارد. بنابراین، پردازش صورت معرفی و به‌کار گرفته شده است. روش‌های پردازش صورت به دو گروه زیر دسته‌بندی می‌شوند:

افزایش یک به چند[1] و عادی‌سازی چند به یک[2].

در افزایش یک به چند، وصله‌ها یا تصاویر زیادی از تنوع ژست را از یک تصویر واحد تولید می‌کنند تا شبکه‌های عمیق را قادر به یادگیری نمایش‌های ثابت کند. در عادی‌سازی چند به یک، نمای متعارف تصاویر چهره را از یک یا چند تصویر از نمای غیر پیشانی بازیابی می‌کنند. جدول (۱)، رویکردهای مختلف روش‌های پردازش صورت را نشان می‌دهد:

جدول ۱- رویکردهای مختلف روش‌های پردازش صورت

رویکرد	روش پردازش صورت	شماره مرجع
مدل ۳- بعدی	یک به چند	[۱۴]،[۱۵]،[۱۶]،[۱۷]،[۱۸]،[۱۹]،[۲۰]،[۲۱]
افزایش داده	یک به چند	[۲۲]،[۲۳]،[۲۵]،[۲۶]،[۲۷]،[۲۸]
مدل ۲- بعدی عمیق	یک به چند	[۲۹]،[۳۰]،[۳۱]
مدل رمزگذار خودکار[3]	چند به یک	[۳۲]،[۳۳]،[۳۴]،[۳۵]،[۳۶]
شبکه عصبی کانولوشن	چند به یک	[۳۷]،[۳۸]
شبکه زایای دشمن‌گونه دشمن‌گونه[4]	چند به یک	[۳۹]،[۴۰]،[۴۱]،[۴۲]

در استخراج ویژگی‌های عمیق (مرحله‌ی دوم)، معماری شبکه‌ها به دو گروه زیر دسته‌بندی می‌شود: شبکه‌های مازه‌ای(ستون فقراتی)[5] و شبکه‌های مجتمع[6].

شبکه‌های مازه‌ای، اشاره به شبکه‌های استخراج ویژگی دارد که در آن از معماری دیپ لَب[7] [۴۲] استفاده می‌شود. این استخراج‌کننده ویژگی برای رمزگذاری ورودی شبکه در نمایشِ ویژگی قطعی استفاده می‌شود. چارچوبِ دیپ لَب، ویژگی‌های "پیچیده" اطرافِ این استخراج‌کننده ویژگی‌ها است. با این کار می‌توان استخراج‌کننده ویژگی را رد و بدل کرد و مدلی را انتخاب کرد که از نظر دقت، کارایی و ... متناسب با وظیفه موجود باشد.

در دیپ لَب، اصطلاح مازه‌ای (ستونِ فقرات) ممکن است به مدل‌هایی مانند رِس نِت[8] [۴۳و۴۴] و ... اشاره داشته باشد.

دیپ لَب، تقسیم‌بندی تصویر[9] با شبکه‌های عصبی پیچشی[10]، پیچشیِ آتُرس[11] و میدانِ تصادفیِ شرطیِ کاملاً متصل[12] است.

شبکه عصبیِ پیچشیِ آتُرس، یک گزینه جایگزین برای لایه نمونه‌برداری است و زمینه پذیرش را افزایش می‌دهد درحالی‌که بُعدِ مکانیِ نگاشت‌های ویژگی را حفظ می‌کند.

میدانِ تصادفیِ شرطی، یک مدل گرافیکی برای طبقه‌بندی است که دارای دو مدل جریمه می‌باشد، یکی برای طبقه‌بندی گره (شدتِ پیکسل[13]: یک حدسِ خوب برای یک لَبه که بین دو کلاسِ شدتِ پیکسل است که با وزن‌دهی لبه‌ی اشیاء انجام می‌شود) و دیگری برای لبه‌ها، جایی‌که اختلاف گره‌های همسایه جریمه می‌شود (مجاورتِ پیکسل: برای پیکسل‌هایی که به هم نزدیک هستند، احتمال تعلق آن‌ها به یک کلاس زیاد است. بدون این، لبه‌های داخل یک شیء را در نمونه‌هایی از کلاس دیگر تفسیر می‌شوند). برای تقسیم‌بندی تصویر، معمولاً هر پیکسل، به‌عنوان یک گره در گراف در نظر گرفته می‌شود و پیکسل‌های مجاور آن‌ها، همسایگان آن‌ها هستند (۴ یا ۸ همسایه در یک تصویر ۲ بعدی)، وزنِ لبه‌ها سعی در اجرایِ پیکسل‌هایِ مجاوری را دارند که دارای برچسب‌های مشابه هستند. گرافِ حاصل از آن بسیار پراکنده است و محاسبه میدانِ تصادفیِ شرطی، سریع است. هنگامی‌که میدانِ تصادفیِ شرطی کاملاً متصل است، هر گره‌ای در مجاورت یکدیگر قرار دارند، که این امر محاسبه را بسیار پرهزینه می‌کند. می‌توان بهینه‌سازی را در گراف‌های تصویری با وزن لبه گاوسی به صورت کارآمد انجام داد. در این حالت، نه تنها همسایگی هر پیکسل را برای بدست آوردن کلاسِ آن، بلکه هر پیکسل دیگر در تصویر، در نظر گرفته می‌شود.

شبکه‌های مجتمع، شبکه با توجه به تعداد کلاس‌هایی که باید تشخیص دهد به چندین شبکه‌ی فرعی تقسیم می‌شود. به‌عنوان مثال شبکه‌های چند منظوره[14] و شبکه‌های چند ورودی[15] در تشخیص چهره استفاده می‌شوند. جمع‌آوری نتایج شبکه‌های مونتاژشده باعث افزایش عملکرد در مقایسه با یک شبکه فردی می‌شود [۴۵].

مطابقت چهره با ویژگی‌های عمیق (مرحله‌ی سوم)، را می‌توان به دو گروه تأییدِ چهره[16] و شناسایی چهره[17] دسته‌بندی نمود. در هر دو روش، در ابتدا، مجموعه‌ای از موضوعاتِ شناخته شده در سیستم (گالری) ثبت می‌شوند و در حین آزمایش، موضوع جدیدی (کاوشگر) ارائه می‌شود. بعد از این که شبکه‌های عمیق با نظارت بر یک تابع از دست‌دادن مناسب،

1. One-to-many Augmentation
2. Many-to one Normalization
3. Auto-encoder Model
4. Generative Adversarial Network (GAN)
5. Backbone
6. Ensemble
7. DeepLab
8. ResNet
9. Image Segmentation
10. Convolutional Neural Networks
11. Atrous Convolutional
12. Fully Connected CRF (Conditional Random Field)
13. Pixel Intensity
14. Multi-task Networks
15. Multi-input Networks
16. Face Verification
17. Face Identification

روی داده‌های عظیم آموزش می‌بینند، هر یک از تصاویرِ آزمون از طریق شبکه‌ها منتقل می‌شود تا یک نمایش ویژگی عمیق بدست آورد. با استفاده از فاصله کسینوس یا فاصله L2، تأیید چهره شباهت یک به یک بین گالری و کاوشگر را محاسبه می‌کند تا تعیین کند که آیا این دو تصویر از یک موضوع هستند، درحالی‌که شناسایی چهره، شباهت یک به چند را برای تعیینِ هویتِ خاصِ یک کاوش محاسبه می‌کند. علاوه بر این، روش‌های دیگری نیز برای پردازش ویژگی‌های عمیق ارائه می‌شود، به‌طوری‌که مطابقتِ چهره به‌طور کارآمد و دقیق انجام می‌شود، مانند یادگیری متریک، طبقه‌بندی مبتنی بر نمایشِ پراکنده.

۴-۴- تشخیص اشیاء با استفاده از یُولُو[1] (یک نگاه)

تشخیص اشیاء یکی از مشکلات غالب و چالش برانگیز در بینایی کامپیوتری است. در دهه‌ی اخیر، با تکامل سریع یادگیری عمیق، محققان به‌طور گسترده‌ای روش‌های مختلف را آزمایش نموده و در بهبود عملکرد تشخیص اشیاء و وظایف مرتبط، مانند طبقه‌بندی اشیاء، محلی‌سازی و تقسیم‌بندی با استفاده از مدل‌های عمیق مشارکت داشته‌اند.

به‌طور کلی، آشکارسازهای اشیاء به دو دسته طبقه‌بندی می‌شوند: آشکارسازهای شی دو مرحله‌ای و تک مرحله‌ای[2]. آشکارسازهای دو مرحله‌ای مرحله‌ای عمدتاً بر راهبرد پیشنهادهای منطقه انتخابی از طریق معماری پیچیده تمرکز می‌نمایند، درحالی‌که، آشکارسازهای تک مرحله‌ای بر روی تمام پیشنهادات منطقه فضایی برای تشخیصِ احتمالی اشیاء، از طریق معماری نسبتاً ساده‌تر در یک عکس[3] تمرکز می‌نمایند. عملکرد هر آشکارسازِ آشکارسازِ شی، از طریق دقتِ تشخیص و زمان استنتاج ارزیابی می‌شود. به‌طور کلی، دقتِ تشخیصِ آشکارسازهای دو مرحله‌ای از آشکارسازهای شیء تک مرحله‌ای بهتر است. با این حال، زمان استنتاج آشکارسازهای تک مرحله‌ای در مقایسه با همتایان خود بهتر است. علاوه بر این، با ظهور یُولُو [۴۶] (در یک نگاه) و جانشین‌های معماری آن، دقت تشخیص به‌طور قابل توجهی بهبود می‌یابد و گاهی اوقات بهتر از آشکارسازهای دو مرحله‌ای است. یُولُوها در کاربردهای مختلف عمدتاً به دلیل استنتاج سریع‌تر آن‌ها به جای در نظر گرفتن دقت تشخیص، مورد استفاده قرار می‌گیرند. به‌عنوان مثال، دقت تشخیص برای یُولُو و شبکه عصبی بازگشتی سریع[4]، به ترتیب ۶۳/۴ و ۷۰ است، با این حال، زمان استنتاج در مورد یُولُو حدود ۳۰۰ برابر سریع‌تر است [۴۷].

۵- کاربردهای سیستم‌های تشخیص چهره

برخی از مهم‌ترین موارد کاربرد سیستم‌های تشخیص چهره به شرح زیر می‌باشد [۴۸]: شبکه‌های اجتماعی، احراز هویت، سرویس‌های امنیتی، رباتیک، قفل گوشی‌های هوشمند، کشف کلاه‌برداری، خرید آنلاین، مراکز درمانی و تبلیغات.

شبکه‌های اجتماعی از قابلیت‌های فناوری تشخیص چهره به منظور جذب مخاطبان بیشتر بهره می‌گیرند. یکی از نرم‌افزارهای این حوزه لوک‌سِری[5] است که به شما امکان تغییر و ویرایش تصویر و ویدئوی خود به صورت بلادرنگ می‌دهد.

دیپ‌فِیس[6]، سیستم تشخیص چهره یادگیری عمیق است که توسط یک یک گروه تحقیقاتی در فیسبوک ایجاد شده است و می‌تواند چهره انسانی را در تصاویر دیجیتالی شناسایی کند. این شرکت از یک شبکه عصبی ۹ لایه با بیش از ۱۲۰ میلیون وزن اتصال استفاده کرده و مدل آن‌ها در چهار میلیون تصویر آپلود شده توسط کاربران فیس‌بوک آموزش دیده است.

لنزهای انیمیشنی اسنب چت[7] یکی دیگر از نمونه‌های بکارگیری تشخیص چهره در شبکه‌های اجتماعی است که به کاربران امکان اضافه‌کردن انواع فیلترها را می‌دهد. با استفاده از این فیلترها کاربران می‌توانند چهره‌ی خود را تغییر دهند، خود را پیرتر کرده و یا اشیائی را به چهره‌ی خود اضافه کنند.

در حال حاضر نرم‌افزار احراز هویت در حال ظهور در سرویس‌های احراز هویت است. بسیاری از شرکت‌ها در حال فعالیت در این زمینه هستند تا خدمات احراز هویت خود را به بانک‌ها، شرکت‌های تجارت الکترونیک و دیگر سازمان‌ها ارائه دهند. سرویس احراز هویتِ شرکت هوش داده مهتاب به صورت آنلاین در دسترس است.

یکی از کاربردهای اولیه‌ی فناوری تشخیص چهره در مسائل امنیتی است. در فرودگاه‌ها، گمرک و ادارات و سازمان‌هایی که نیاز به کنترل دسترسی و تردد افراد به بخش‌های خاصی را دارند، نرم‌افزار تشخیص چهره می‌تواند مؤثر واقع شوند. هم‌چنین می‌توان برای حضور و غیاب کارمندان نیز از این سیستم استفاده کرد. پلیس می‌تواند از نرم‌افزار تشخیص چهره در سیستم‌های خود به منظور شناسایی مجرمان و نظارت بهره گیرد.

ربات‌ها برای ارتباط برقرارکردن با دنیای واقعی نیازمند تشخیص اشیاء و انسان‌های افراد خود هستند. قابلیت‌های تشخیص چهره یکی از نیازمندی‌های اساسی در صنعت رباتیک است. ربات‌ها نیازمند تشخیص و شناسایی چهره‌ی انسان‌ها به صورت بلادرنگ هستند.

توسعه‌دهندگان برنامه‌های موبایل از تشخیص چهره برای تمیزدادن چهره ی انسان از عکس انسان استفاده می‌کنند. گاهی این فناوری قادر به تمایز قائل‌شدن بین عکس و چهره‌ی واقعی انسان نیست. برای این مسأله محققان از روش‌های یادگیری عمیق استفاده می‌کنند. کشف کلاهبرداری

1. YOLO (You Only Look Once)
2. Two-stage and single-stage Object Detectors
3. Shot
4. Fast-RCNN
5. Looksey
6. Deep Face
7. Snap Chat

با استفاده از تشخیص چهره در مراکز بهداشتی درمانی، بانک‌ها، صنایع نظامی نیز کاربرد دارد.

سازندگان گوشی‌های هوشمند مانند اپل و گوگل نرم‌افزار تشخیص چهره را در گوشی‌های خود پیاده‌سازی کرده‌اند و مشتریان می‌توانند تراکنش‌های مالی را به راحتی با گوشی‌های خود انجام دهند.

علاوه بر این که تشخیص چهره می‌تواند به‌عنوان یک سیستم امنیتی در بیمارستان‌ها مورد استفاده قرار بگیرد، شامل کاربردهای دیگری هم می‌شود. برای مثال می‌توان احساسات فرد بیمار را با استفاده از این فناوری تشخیص داد و در طی دوره‌ای که بیمار در آن مرکز درمانی مستقر است می‌تواند حالات او بررسی کرد. علاوه بر این امکان ردیابی بیمار هم با بکارگیری این فناوری میسر می‌شود.

تشخیص چهره نقش مهمی را در جمع‌آوری داده‌های شخصی افراد بازی می‌کند. فرض کنید مشتری بین ۱۸ تا ۲۵ ساله توسط دوربین‌ها تشخیص داده شده است، صفحه نمایش با دریافت این اطلاعات محصولات مرتبط با سن این فرد را به او نمایش می‌دهند.

۶- نتیجه‌گیری

در این مقاله، به بررسی الگوریتم‌های نوین تشخیص چهره پرداخته شد و ضمن بیان مفاهیم پایه‌ی پردازش تصویر، تشخیص چهره و فناوری‌های آن بیان شدند. تشخیص چهره مبتنی بر بیومتریک‌ها می‌باشد که مشخصه‌های منحصر به فرد و قابل اندازه‌گیری برای شناسایی هویت افراد هستند. به منظور پیاده‌سازی تشخیص چهره از روش‌ها و الگوریتم‌های متفاوتی استفاده می‌شود. این الگوریتم‌ها را می‌توان به‌طور کلی در پنج گروه دسته‌بندی کرد. این دسته‌بندی شامل: روش‌های سنتی، روش‌های ابعادی، روش دوربین حرارتی، روش مقایسه بافت پوست، روش‌های ترکیبی می‌باشند.

سپس چالش‌های فعلی تشخیص چهره بیان گردید و سه گروه از الگوریتم‌های تشخیص چهره معرفی و توضیح داده شدند. در ابتدا الگوریتمِ HMAX معرفی و توضیح داده شد. سپس ساختارِ الگوریتمِ B-HMAX بیان شد. همچنین به بیان ساختارِ تشخیصِ چهره‌ی عمیق پرداخته شد و سه بخش که برای تشخیص چهره‌ی عمیق لازم است، توضیح و معرفی گردید:

در ابتدا از یک کشف‌کننده‌ی چهره برای متمرکزشدن بر روی چهره‌ها در تصاویر یا فیلم‌ها استفاده می‌شود. در مرحله‌ی دوم چهره‌ها با مختصات متعارف عادی، مطابقت داده می‌شوند و در مرحله سوم تشخیص چهره انجام می‌شود که شامل پردازش چهره، استخراج ویژگی‌های عمیق و تطبیق با چهره است. همچنین، آشکارسازهای شیء تک مرحله‌ای مبتنی بر یُولُو و پیشرفت‌های معماری آن‌ها که مبتنی بر شبکه‌های عصبی بازگشتی از پیش آموزش دیده شده می‌باشند را به همراه و عملکرد آن‌ها معرفی شد. در پایان کاربردهای سیستم‌های تشخیص چهره بیان شد.

۷- مراجع

1- S. V. Khedaskar, M. A. Rokade, B. R. Patil, and P. Tatwadarshi, "A survey of image processing and identification techniques," VIVA-Tech International Journal for Research and Innovation, vol. 1, no. 1, pp. 1-10, 2018.
2- P. Pradhan, "Development Of Face Recognition System Using Verilook Software Development Kit," 2008.
3- T. Kurita, "Principal component analysis (PCA)," Computer Vision: A Reference Guide, pp. 1-4, 2019.
4- K. Sudhakar and P. Nithyanandam, "An accurate facial component detection using gabor filter," Bulletin of Electrical Engineering and Informatics, vol. 6, no. 3, pp. 287-294, 2017.
5- P. R. G. Kurka and A. A. D. Salazar, "Applications of image processing in robotics and instrumentation," Mechanical Systems and Signal Processing, vol. 124, pp. 142-169, 2019.
6- J. Hernandez-Ortega, J. Galbally, J. Fiérrez, and L. Beslay, "Biometric quality: Review and application to face recognition with faceqnet," arXiv preprint arXiv:2006.03298, 2020.
7- M. Krišto and M. Ivasic-Kos, "An overview of thermal face recognition methods," in 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2018, pp. 1098-1103: IEEE.
8- M. M. Oghaz, M. A. Maarof, M. F. Rohani, A. Zainal, and S. Z. M. Shaid, "An optimized skin texture model using gray-level co-occurrence atrix," Neural Computing and Applications, vol. 31, no. 6, pp. 1835-1853, 2019.
9- Z. Cheng, X. Zhu, and S. Gong, "Face re- identification challenge: Are face recognition models good enough?," Pattern Recognition, vol. 107, p. 107422, 2020.
10- Z. Cheng, X. Zhu, and S. Gong, "Surveillance face recognition challenge," arXiv preprint arXiv:1804.09691, 2018.
11- C. Liu and F. Sun, "HMAX model: A survey," in 2015 International Joint Conference on Neural Networks (IJCNN), 2015, pp. 1-7: IEEE.
12- M. Nazarkevych, I. Klyujnyk, I. Maslanych, B. Havrysh, and H. Nazarkevych, "Image filtration using the Ateb-Gabor filter in the biometric security systems," in 2018 XIV-th International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH), 2018, pp. 276-279: IEEE.
13- I. Masi, Y. Wu, T. Hassner, and P. Natarajan, "Deep face recognition: A survey," in 2018 31st SIBGRAPI conference on graphics, patterns and images (SIBGRAPI), 2018, pp. 471-478: IEEE.
14- I. Masi, A. T. Trần, T. Hassner, J. T. Leksut, and G. Medioni, "Do we really need to collect millions of faces for effective face recognition?," in European conference on computer vision, 2016, pp. 579-596: Springer.
15- I. Masi, T. Hassner, A. T. Tran, and G. Medioni, "Rapid synthesis of massive face sets for improved face recognition," in 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), 2017, pp. 604-611: IEEE.
16- E. Richardson, M. Sela, and R. Kimmel, "3D face reconstruction by learning from synthetic data," in 2016 fourth international conference on 3D vision (3DV), 2016, pp. 460-469: IEEE.
17- E. Richardson, M. Sela, R. Or-El, and R. Kimmel, "Learning detailed face reconstruction from a single image," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1259-1268.
18- P. Dou, S. K. Shah, and I. A. Kakadiaris, "End-to-end 3D face reconstruction with deep neural networks," in proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 5908-5917.
19- Y. Guo, J. Cai, B. Jiang, and J. Zheng, "Cnn-based real-time dense face reconstruction with inverse-rendered photo-realistic

face images," IEEE transactions on pattern analysis and machine intelligence, vol. 41, no. 6, pp. 1294-1307, 2018.
20- A. Tuan Tran, T. Hassner, I. Masi, and G. Medioni, "Regressing robust and discriminative 3D morphable models with a very deep neural network," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 5163-5172.
21- A. Tewari et al., "Mofa: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction," in Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 1274-1283.
22- J. Liu, Y. Deng, T. Bai, Z. Wei, and C. Huang, "Targeting ultimate accuracy: Face recognition via deep embedding," arXiv preprint arXiv:1506.07310, 2015.
23- E. Zhou, Z. Cao, and Q. Yin, "Naive-deep face recognition: Touching the limit of LFW benchmark or not?," arXiv preprint arXiv:1501.04690, 2015.
24- C. Ding and D. Tao, "Robust face recognition via multimodal deep face representation," IEEE Transactions on Multimedia, vol. 17, no. 11, pp. 2049-2058, 2015.
25- Y. Sun, X. Wang, and X. Tang, "Sparsifying neural network connections for face recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4856-4864.
26- D. Wang, C. Otto, and A. K. Jain, "Face search at scale," IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 6, pp. 1122-1136, 2016.
27- Y. Sun, X. Wang, and X. Tang, "Deeply learned face representations are sparse, selective, and robust," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 2892-2900.
28- Y. Sun, D. Liang, X. Wang, and X. Tang, "Deepid3: Face recognition with very deep neural networks," arXiv preprint arXiv:1502.00873, 2015.
29- Z. Zhu, P. Luo, X. Wang, and X. Tang, "Multi-view perceptron: a deep model for learning face identity and view representations," Advances in Neural Information Processing Systems 27 (NIPS 2014), 2014.
30- J. Zhao et al., "Dual-Agent GANs for Photorealistic and Identity Preserving Profile Face Synthesis," in NIPS, 2017, vol. 2, p. 3.
31- A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning from simulated and unsupervised images through adversarial training," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2107-2116.
32- M. Kan, S. Shan, H. Chang, and X. Chen, "Stacked progressive auto-encoders (spae) for face recognition across poses," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 1883-1890.
33- Y. Zhang, M. Shao, E. K. Wong, and Y. Fu, "Random faces guided sparse many-to-one encoder for pose-invariant face recognition," in Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 2416-2423.
34- J. Yang, S. Reed, M.-H. Yang, and H. Lee, "Weakly-supervised disentangling with recurrent transformations for 3d view synthesis," arXiv preprint arXiv:1601.00706, 2016.
35- M. Xi, L. Chen, D. Polajnar, and W. Tong, "Local binary pattern network: A deep learning approach for face recognition," in 2016 IEEE international conference on Image processing (ICIP), 2016, pp. 3224-3228: IEEE.
36- Z. Zhu, P. Luo, X. Wang, and X. Tang, "Recover canonical-view faces in the wild with deep neural networks," arXiv preprint arXiv:1404.3543, 2014.
37- L. Hu, M. Kan, S. Shan, X. Song, and X. Chen, "LDF-Net: Learning a displacement field network for face recognition across pose," in 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), 2017, pp. 9-16: IEEE.
38- E. Zhou, Z. Cao, and J. Sun, "Gridface: Face rectification via learning local homography transformations," in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 3-19.
39- R. Huang, S. Zhang, T. Li, and R. He, "Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis," in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2439-2448.
40- L. Tran, X. Yin, and X. Liu, "Disentangled representation learning gan for pose-invariant face recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1415-1424.
41- J. Deng, S. Cheng, N. Xue, Y. Zhou, and S. Zafeiriou, "Uv-gan: Adversarial facial uv map completion for pose-invariant face recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7093-7102.
42- L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," IEEE transactions on pattern analysis and machine intelligence, vol. 40, no. 4, pp. 834-848, 2017.
43- X. Zhang, Z. Fang, Y. Wen, Z. Li, and Y. Qiao, "Range loss for deep face recognition with long-tailed training data," in Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 5409-5418.
44- W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 212-220.
45- G. Hu et al., "When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition," in Proceedings of the IEEE international conference on computer vision workshops, 2015, pp. 142-150.
46- T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: challenges, architectural successors, datasets and applications," Multimedia Tools and Applications, pp. 1-33, 2022.
47- B. Rekha and D. Athiya Marium, "Literature survey on object detection using yolo," International Research Journal of Engineering and Technology (IRJET), 2020.
48- Q. Huang and L. Cui, "Design and Application of Face Recognition Algorithm Based on Improved Backpropagation Neural Network," Revue d'Intelligence Artificielle, vol. 33, no. 1, pp. 25-32, 2019.