دو فصلنامه علمی
فناوری اطلاعات و ارتباطات ایران

سال پانزدهم، شماره ۵۷ و۵۸ ، پاییز و زمستان ۱۴۰۲
صفحات:۹۴ الی ۱۱۰

Face Recognition and Liveness Detection Based on Speech Recognition for Electronic Authentication

Ahmad Dolatkhah[*], Behnam Dorostkar Yaghoti[**], Raheb Hashempour[***]

[*]Instructor, Department of Information and Communication Technology, Amin University of Law Enforcement Sciences, Tehran, Iran

[**]Assistant Professor, Department of Information and Communication Technology, Amin University of Law Enforcement Sciences, Tehran, Iran

[***]M.Sc., Computer Engineering, Amin University of Law Sciences, Tehran, Iran

Abstract

As technology develops, institutions and organizations provide many services electronically and intelligently over the Internet. The police, as an institution that provides services to people and other institutions, aims to make its services smarter. Various electronic and intelligent systems have been offered in this regard. Because these systems lack authentication, many services that can be provided online require a visit to +10 police stations. Budget and equipment limitations for face-to-face responses, limitations of the police force and their focus on essential issues, a lack of service offices in villages and a limited number of service offices in cities, and the growing demand for online services, especially in crisis situations like Corona disease, electronic authentication is becoming increasingly important. This article reviews electronic authentication and its necessity, liveness detection methods and face recognition which are two of the most important technologies in this area. In the following, we present an efficient method of face recognition using deep learning models for face matching, as well as an interactive liveness detection method based on Persian speech recognition. A final section of the paper presents the results of testing these models on relevant data from this field.

Keywords: Electronic Authentication, Face Recognition, Liveness Detection, Speech Recognition.

تطبیق چهره و تشخیص زنده بودن مبتنی بر بازشناسی گفتار برای احراز هویت غیرحضوری

احمد دولت خواه*×، بهنام درستکار یاقوتی**، راهب هاشم پور***

*مربی، گروه فناوری اطلاعات و ارتباطات، دانشگاه جامع علوم انتظامی امین، تهران، ایران

**استادیار، گروه فناوری اطلاعات و ارتباطات، دانشگاه جامع علوم انتظامی امین، تهران، ایران

***کارشناسی ارشد، مهندسی کامپیوتر، دانشگاه جامع علوم انتظامی امین

تاریخ دریافت: ۱۴۰۱/۰۹/۱۸ تاریخ پذیرش:۱۴۰۱/۱۲/۱۱

نوع مقاله: پژوهشی

چکیده

با گسترش فناوری بسیاری از خدمات نهادها و سازمان‌ها به صورت الکترونیکی و هوشمند، در بستر اینترنت ارائه می‌گردد. پلیس نیز به عنوان یک نهاد ارائه‌دهنده خدمات به مردم و سایر نهادها، به دنبال هوشمندسازی خدمات خود می‌باشد. در همین راستا نیز سامانه‌های الکترونیکی و هوشمند مختلفی را ارائه کرده است. به دلیل عدم احراز هویت کاربران در این سامانه‌ها، بسیاری از خدماتی که می‌توانند به صورت غیرحضوری ارائه گردد، نیاز به مراجعه به دفاتر پلیس+۱۰ را دارند. محدودیت بودجه و تجهیزات برای پاسخگویی حضوری، محدودیت نیروهای پلیس و تمرکز آن‌ها بر روی موضوعات مهم، محدودیت تعداد دفاتر خدماتی در شهرستان‌ها و عدم دسترسی روستاها به این دفاتر، رشد روزافزون خدمات برخط و افزایش تقاضای مردم برای آن، به ویژه در شرایطی مانند بحران بیماری کرونا، سبب شده است تا نیاز به احراز هویت غیرحضوری بسیار مورد توجه قرار بگیرد. در این مقاله، احراز هویت غیرحضوری و ضرورت استفاده از آن، روش‌های تشخیص زنده بودن و بازشناسی چهره که دو فناوری مهم در این حوزه است، مرور شده است. در ادامه یک روش کارآمد از مدل‌های یادگیری عمیق بازشناسی چهره برای تطبیق چهره و یک روش تشخیص زنده بودن تعاملی به وسیله‌ی بازشناسی گفتار فارسی ارائه شده است و در نهایت نتایج آزمایش این مدل‌ها بر روی داده‌های مربوط در این حوزه آورده شده است.

.

واژگان کلیدی: احراز هویت غیرحضوری، بازشناسی چهره، تشخیص زنده بودن، بازشناسی گفتار

× نویسنده مسئول: احمد دولتخواهict.ahmad93@gmail.com

۱. مقدمه

تغییرات و پیشرفت در دنیای فناوری و دیجیتال، سازمان‌ها را بر آن دا شته تا قبل از اینکه زیر امواج سهمگین این تحولات غرق شوند، خود را متناسب با تغییرات عصر حاضر وفق دهند. دولت‌ها در راستای حل مشکلات کشور، دست به‌سوی دنیای فناوری دراز کرده‌اند تا با بهره‌گیری از تحول دیجیتال چالش‌های کلان ک شور را حل و به بهبود شرایط زیست شهروندان، کسب‌وکارها و دولت کمک کنند. در پلیس هو شمند و ارائه خدمات به مردم در سال‌های اخیر با توجه به تغییرات فناوری و افزایش انتظارات مردم از سازمان ، لزوم تحول در پلیس احساس شده است. ایجاد سامانه‌های هوشمند و استفاده از تجهیزات نوین با بهره‌مندی از هوش مصنوعی در حوزه‌های مختلف از جمله گذرنامه، گواهینامه، خدمت سربازی گرفته تا خدمات الکترونیکی، دوربین‌های کنترل ترافیک و تکمیل باندهای اطلاعاتی و آگاهی از جمله اقداماتی است که می تواند ارائه خدمات به مردم را به صورت الکترونیکی و هو شمند انجام شود. در حال حاضر ارائه خدمات الکترونیک فرماندهی انتظامی به خوبی انجام می‌شود. اصلاح فرآیند هوشمندسازی جزء اولویت‌های فرماندهی انتظامی است و در حال حاضر برای دریافت گذرنامه دیگر نیازی به مراجعه به دفاتر «پلیس+۱۰» نیست البته به شرطی که احراز هویت برای پلیس در سایت انجام شده باشد و بعد از آن مردم می‌توانند گذرنامه خود را درب منازل تحویل بگیرند.

حرکت به سمت هو شمند سازی، اجتناب‌ناپذیر ا ست و به عبارتی دیگر باید میز پلیس را در منازل برده و مردم از طریق سامانه با فرماندهی انتظامی ارتباط داشته باشند و خدمات را با سرعت بیشتری به مردم ارائه دهند که این از اهداف پلیس هوشمند و زیر ساخت‌های آن در حال آماده سازی ا ست. سامانه‌های هو شمند فرماندهی انتظامی در بستر اپلیکیشن «پلیس همراه من» و سایت «پلیس راهور» و برخی سامانه‌های دیگر بارگذاری شده است. مردم با مراجعه به این سامانه‌ها می‌توانند به سهولت از خدمات راهنمایی و رانندگی فراجا بهره‌مند شوند.

احراز هو یت الکترونیکی (از راه دور)[1] (eKYC) به دلیل امنیت و قابل اعتماد بودن و همچنین سازگاری با استانداردهای جهانی، به عنوان یک راه‌حل منا سب برای پا سخگویی به نیاز بیان شده، ارائه شده است. به دلیل گسترش بیش از پیش تجهیزات دریافت تصویر در جهان و رشد فناوری دوربین‌های دیجیتال در ساختار گوشی های هوشمند امروزی و رایانه‌ها و همچنین افزایش استفاده از فناوری زیست‌سنجی چهره در بازار های تجاری و نیز وجود الگوریتم‌های قابل اطمینان در این حوزه، روش احراز هویت الکترونیکی از طریق تایید چهره به طور روزافزونی توسعه یافته ا ست. در را ستای ا ستفاده از فناوری باز شنا سی چهره برای احراز هویت، مسائلی مانند واقعی بودن ویدیوی ار سال شده به این معنی که مربوط به یک فرد زنده باشد، به وجود می‌آید. تشخیص زنده بودن به این معناست که تصویر و یا ویدیوی ارسال شده واقعی باشد و یک مصنوع تقلبی نباشد مانند نمایش یک عکس چهره‌ی چاپ شده و یا نمایش از روی صفحه‌ی تلفن همراه و یا ماسک. در همین جهت روش‌های مختلف باز شنا سی چهره و ت شخیص زنده بودن در ابتدا بررسی شده است. در بخش‌های سه و چهار از این پژوهش تلاش شده است تا به روزترین روش‌ها و نتایج آن‌ها در این دو حوزه بررسی گردد.

برای عمل‌یاتی کردن خدمات غیرحضوری، یکی از اصلی‌ترین چالش‌های پیش‌رو، مو ضوع امنیت و اعتبار سنجی هویت م شتریان[2] (KYC) است. این مساله در رویکرد سنتی، با مراجعه حضوری افراد به دفاتر و پیشخوان‌های فرماندهی انتظامی حل می‌شود، این در حالی است که برای ارائه خدمات مجازی باید از احراز هو یت الکترونیکی بهره گرفت. به طور خلاصه eKYC عبارتی است که برای توصیف دیجیتالی و الکترونیکی شدن فرآیندهای KYC ا ستفاده می شود. eKYC (م شتری خود را الکترونیکی ب شنا سید) فرآیندی از راه دور و بدون کاغذ است که هزینه‌ها و بوروکراسی سنتی مورد نیاز در فرآیندهای KYC را به حداقل می رساند. فرآیند eKYC دیجیتالی و از راه دور شدن فرآیند KYC سنتی است. شناسایی و تأیید هویت مشتری در زمان واقعی و بلافاصله اتفاق می افتد و به همین دلیل فرآیندهای eKYC رضایت مشتری را افزایش می‌دهند. در همین را ستا برای اطمینان از اینکه فرایندهای eKYC دارای استانداردهای ایمنی شناسایی هستند، راه‌حل‌های به کار رفته باید فرآیندهای شناسایی الکترونیکی را با سطح بالایی از ایمنی و قابلیت اطمینان و مطابق با قوانین تعیین شده پیاده سازی کنند. طیف گسترده‌ای از راه حل های eKYC مبتنی بر هوش مصنوعی و یادگیری ماشین ارائه شده است که سامانه‌های زیست‌سنجی و مفاهیم مربوط به آن از جمله‌ی مهم‌ترین این راه‌حل‌ها می‌باشند.

برای ارائه خدمات غیرحضوری و از راه دور به افراد جامعه، لازم است امنیت ارائه خدمات به ویژه احراز هو یت افراد با اطمینان مطلوب تامین شود که برای این کار از ویژگی‌های زیست‌سنجی چهره آن‌ها به عنوان معیار شناسایی استفاده می‌شود. بدین صورت که فرد متقاضی با بیان یکی از شناسه‌های هویتی خود مانند کد ملی، شماره گواهینامه یا گذرنامه و همچنین ارائه تصویری از خود به صورت برخط درخوا ست تایید هویت[3] می‌کند. علاوه بر تایید هویت

[1] electronic Know Your Customer (eKYC)

[2] Know Your Customer (KYC)

[3] Verification

مبتنی بر چهره، موضوع مهم دیگر این طرح، تشخیص زنده بودن است که در آن زنده بودن ویدئوی دریافتی بررسی می‌شود، در شکل ۱ این فرایند مشاهده می‌شود.

شکل ۱. نمایی از فرایند احراز هویت غیرحضوری

در ادامه این مقاله، پس از بیان ضرورت احراز هویت غیرحضوری و کاربرد فناوری‌های مرتبط در حوزه انتظامی، به بررسی این حوزه و علوم مربوط به احراز هویت از راه دور از جمله مفاهیم زیست‌سنجی به طور خلاصه و رویکردهای بازشناسی چهره و تشخیص زنده بودن پرداخته می‌شود. پس از آن، روش‌های ارائه شده در این پژوهش برای تطبیق چهره و تشخیص زنده بودن مبتنی بر تشخیص گفتار ارائه می‌گردد.

۲. ضرورت به‌کارگیری احراز هویت غیرحضوری در فرماندهی انتظامی

در برنامه توسعه به ازای هزار نفر، پنج پلیس باید باشد و این تعداد در بعضی از کشورها هشت نفر است ولی در حال حاضر در ایران حدود نصف این تعداد پلیس مشغول خدمت‌رسانی هستند که موجب فشار بر روی فرماندهی انتظامی می‌شود. موضوع هوشمندسازی، بخشی از خلأ فرماندهی انتظامی را می‌پوشاند و هرچه این هوشمند سازی در جهت درست و کامل پیش رود علاوه بر احتیاج به نیروی انسانی کمتر به راحتی کاربر و همچنین کاهش هزینه های سازمان کمک می‌کند که قدم اول در این راستا، پیاده‌سازی احراز هویت برخط (آنلاین) می‌باشد.

قابل به ذکر است که فرماندهی انتظامی جمهوری اسلامی ایران یک نهاد عظیم با رسته‌های مختلف است. شمار نیروی انسانی آن در سراسر ایران به بیش از چند صد هزار نفر می‌رسد. از این رو، ارائه خدمات به نیروهای این نهاد به دلیل گستردگی و پراکندگی، خود یکی دیگر از چالش‌های موجود است. کنترل این افراد و نظارت بر آن‌ها به خصوص به دلایل امنیتی نیازمند احراز هویت می‌باشد که به دلایل بیان شده، این نظارت تنها در صورت الکترونیکی بودن (از راه دور) دارای توجیه اقتصادی، امنیتی و مدیریتی می باشد. هم‌چنین در بعضی از شهرستان‌ها به دلیل کمبود دفاتر پلیس+۱۰ و یا ایام خاص سال مانند ایام اربعین، مراجعه به این دفاتر افزایش پیدا می‌کند. اما در صورت کاهش لزوم مراجعه به این دفاتر به دلیل انتقال خدمات آن‌ها به سامانه‌های هوشمند از طریق احراز هویت الکترونیکی، می‌توان آسایش کاربران به خصوص برای شهرستان‌ها و مناطق روستایی را تامین نمود.

باید به این نکته اشاره کرد که با تحول دیجیتالی و هوشمندسازی، همواره نیازهای جدیدی به وجود می‌آید که راه‌حل‌های مناسب آن‌ها نیز در همین حوزه ارائه می‌شود. وجود بسترهایی مانند احراز هویت الکترونیکی (از راه دور) می‌تواند انعطاف کافی برای اجرای چنین راه‌حل‌هایی را در زمینه امنیتی و خدمات پلیس ایجاد کرده و توانایی کنترل یکپارچه بر روی کاربران و مدیریت اطلاعات آن‌ها را در اختیار نهادها قرار دهد. با توجه به موارد بیان شده، ضرورت ارائه خدمات غیرحضوری با رشد روزافزون خدمات برخط و افزایش تقاضای مردم برای آن، به ویژه در شرایطی مانند بحران بیماری کرونا، موضوعی بدیهی است که همه سازمان‌ها و نهادهای ارائه‌دهنده خدمات را به سمت بهره‌گیری از آن سوق داده است و مورد تاکید نهادهای بالادستی کشور شامل قانون‌گذاران و سیاست گذاران است. احراز هویت، پیش‌نیاز ارائه هرگونه خدمات غیرحضوری توسط پلیس است و لازم است افراد قبل از دریافت خدمات (به ویژه خدمات مهم و حساس)، احراز هویت شوند. بنابراین، همه سرویس‌های ارائه شده به مردم، قبل از دریافت توسط افراد، با فراخوانی سرویس احراز هویت، فرد گیرنده خدمات را شناسایی می‌کنند و این کار باید به صورت غیرحضوری و از راه دور باشد.

تقاضای روزافزون ارائه خدمات الکترونیکی و غیرحضوری از سمت مردم و لزوم بهبود تجربه مشتری[1] به دلیل سادگی و سرعت کار، تاکید نهادهای قانون‌گذار بر ارائه خدمات الکترونیکی و غیرحضوری به مردم به ویژه با تشدید موضوع در شرایط بحران کرونا، نیاز به کاهش مراجعات حضوری افراد (از نظر سلامتی، ترافیک و غیره)، ضرورت افزایش امنیت و اشراف اطلاعاتی پلیس با تکمیل پایگاه های داده افراد به ویژه در تکمیل اطلاعات زیست‌سنجی، تهیه زیرساخت‌های استفاده از آن‌ها، یکپارچه‌سازی و پیگیری (شفافیت)، لزوم کاهش خطاهای انسانی و سوء استفاده افراد از اطلاعات و اسناد از دلایل مهم ضرورت اجرای احراز هویت غیرحضوری توسط فرماندهی انتظامی می‌باشد.

[1] User Experience

۳. بازشناسی چهره و رویکردهای آن

شناسایی افراد توسط انسان ها با مشاهده چهره فرد مقابل قابل انجام است و در زندگی روزمره به این امر به عنوان تشخیص هویت یاد می شود. انسان این قدرت را دارد که چهره فردی را که قبلا مشاهده کرده را بعد از گذشت زمان نیز به یادآورده و تشخیص به درستی صورت پذیرد. در بین زمینه‌های زیست سنجی نیز، احراز هویت به کمک چهره بسیار مورد توجه قرار گرفته است. مخصوصا در سه دهه‌ی اخیر، موضوع بازشناسی چهره از یک موضوع تحقیقاتی علمی عبور کرده و پا به عرصه‌ی تکنولوژی و محصولات تجاری گذاشته است و کاربردهای این تکنولوژی از تشخیص هویت افراد در مرز های بین المللی و جستجو به دنبال مجرمان تا نشانه‌گذاری[1] صورت‌ها در شبکه‌های اجتماعی گسترده شده است[1]. اولین تلاش‌ها برای دسته‌بندی چهره در مقاله‌ای در سال ۱۸۸۸ میلادی مورد بررسی قرار گرفت [2]. روش پیشنهادی نویسنده در این مقاله بدین صورت است که خطوط نیم‌رخ چهره به صورت برداری ذخیره شود و با محاسبه‌ی میانگین این بردارها و محاسبه‌ی فاصله‌ی هر بردار تا بردار میانگین، دسته‌بندی خطوط انجام شود[3]. تحقیقات انجام شده نشان‌دهنده‌ی این موضوع است که سه عامل تغییراتی که به واسطه‌ی سن، تغییرات نور و تغییرات زاویه‌ی تصویربرداری ایجاد می شوند، مهم‌ترین مشکلاتی است که سیستم‌های بازشناسی چهره با آن مواجه هستند[4]. عملیات مقایسه در فرآیند بازشناسی چهره با استفاده از یک عکس خاکستری مانند هر سیستم زیست‌سنجی دیگر، مراحل مشابهی را طی می‌کند. به این صورت که ابتدا سیستم یک عکس حاوی چهره دریافت می‌کند، مکان چهره‌ی انسان را در عکس تشخیص می‌دهد، قسمت چهره از عکس بریده شده، نرمال می شود و ویژگی‌های آن استخراج می‌شود و بدین ترتیب الگوی تصویر صورت تشکیل می‌شود. در هنگام تشخیص هویت، این الگوی دریافت شده با الگوهای موجود در پایگاه داده مقایسه می‌شود. بدین ترتیب دو بخش اصلی این الگوریتم؛ ۱) مکان‌یابی چهره و نرمال‌سازی و ۲) تشخیص هویت چهره خواهد بود. الگوریتم‌هایی که هر دو بخش را در بر می‌گیرند، الگوریتم‌های تشخیص چهره‌ی تمام اتوماتیک و الگوریتم‌هایی که تنها بخش دوم را شامل می‌شوند الگوریتم‌های نیمه خودکار نامیده می‌شوند[5]. در ادامه‌ی این بخش، جزئیات بیشتری در مورد تشخیص هویت چهره ارائه می‌شود.

در یک سیستم تشخیص هویت به کمک چهره، پس از مکان‌یابی چهره در تصویر و پیش‌پردازش آن، وارد مرحله‌ی بعدی یعنی استخراج ویژگی از چهره و تشکیل الگوی چهره می‌شود. الگوریتم‌های بازشناسی چهره را می‌توان در یک دسته‌بندی کلی به دو بخش تقسیم‌بندی کرد: ۱) روش‌های کلی ۲) روش‌های بر پایه‌ی اجزای چهره[5]. در روش های کلی، ویژگی های کل چهره در یک بردار ذخیره می‌شود. اما در روش‌های بر پایه‌ی اجزای چهره، هر یک از اجزاء به صورت جداگانه مکان‌یابی شده و ترکیب آن اجزاء با یکدیگر در تشخیص هویت چهره به کار می‌روند. مزیت روش بر پایه‌ی اجزای چهره در مقایسه با روش‌های کلی این است که برای تغییر زاویه‌های جزئی در چهره، تغییراتی که در هر یک از اجزاء به تنهایی ایجاد می‌شود، به نسبت تغییرات کلی چهره بسیار کمتر است و بدین ترتیب سیستم نسبت به چرخش و تغییر حالت مقاومت بیشتر نشان خواهد داد.[6].

روش های بازشناسی چهره را می‌توان از جنبه های مختلف طبقه‌بندی کرد که در ادامه رایج‌ترین آن‌ها ارائه می‌گردد. رویکردهای دوبعدی را می‌توان به چهار دسته مختلف تقسیم کرد: ۱) روش‌های جامع (کلی)، ۲) روش‌های محلی (هندسی)، ۳) روش های مبتنی بر یادگیری کم عمق و ۴) روش‌های مبتنی بر یادگیری عمیق [7]. طبق این پژوهش‌ها، الگوریتم‌های بازشناسی چهره به صورت زیر به چهار دسته قابل توصیف می باشند[5]:

یادگیری کلی: در این روش‌ها که بیشتر در دهه‌ی ۱۹۹۰ و اوایل دهه‌ی ۲۰۰۰ میلادی مورد توجه قرار گرفتند، تلاش بر این بود که به کمک یک پراکندگی فرضی، یک بازنمایی با تعداد ابعاد محدود برای هر چهره ارائه شود. اولین و بارزترین نمونه‌ی روش، «مقادیر ویژه‌ی چهره» است[8]. این روش‌ها تحت شرایط محیطی مختلف معمولاً با مشکل مواجه می‌شوند. این روش که در ابتدای دهه‌ی ۱۹۹۱ میلادی ارائه شد، یکی از زمینه‌های رشد زمینه‌ی بازشناسی چهره به شمار می‌رود [9]. الگوریتم‌های بر پایه‌ی تطابق گراف‌ها[2]، مدل مخفی مارکف[3]، تطابق ویژگی هندسی[4]، تطابق نمونه‌ها[5]، نقشه‌ی خطوط لبه[6] و همچنین SVM نیز از دیگر روش هایی هستند که در مسئله‌ی تشخیص هویت به کمک چهره به کار رفته‌اند.

روش های مبتنی بر ویژگی های محلی: در دهه‌ی ۲۰۰۰ میلادی، روش‌هایی بر پایه‌ی ویژگی‌های محلی (مانند نتایج فیلترهای گابور) ارائه شد. این روش‌ها تا حدودی نسبت به شرایط محیطی مختلف

[1] Tagging
[2] Graph matching
[3] Hidden Markov model
[4] Geometrical feature matching
[5] Template matching
[6] Line edge map (LEM)

مقاومت نشان می‌دادند اما فشردگی کافی را نداشتند و همچنین قابلیت ایجاد متمایز در آن ها کافی نبود. یک روش بارز در این بخش، بر پایه‌ی فیلترهای گابور ارائه شد [10].

یادگیری کم‌عمق: در اوایل دهه‌ی ۲۰۱۰ میلادی روش هایی ارائه شدند که در آن‌ها توصیف‌گرهای محلی بر پایه‌ی یادگیری معرفی شدند. در واقع در این روش ها با توجه به پایگاه داده، فیلترهایی آموزش داده می شوند که بیشترین ایجاد تمایز را ایجاد می‌کنند. اما هنوز این روش ها مقاومت کافی در برابر تبدیل های غیر خطی و پیچیده‌ی چهره را ندا شتند. پژوهش نمونه روش ارائه شده در این زمینه است.

یادگیری عمیق: در سال ۲۰۱۴ میلادی با ارائه‌ی الگوریتم DeepFace توسط تیم تحقیقاتی شرکت Facebook سری دیگری از روش‌های بازشناسی چهره بر پایه‌ی یادگیری عمیق کلید خورد [11]. در این روش ها بر خلاف روش های یادگیری کم‌عمق، تعداد لایه‌های زیادی به صورت متوالی به منظور استخراج ویژگی و تبدیل آن ها در نظر گرفته شده و بدین ترتیب در سطوح ویژگی‌های مختلفی با سطوح پیچیدگی مختلف شناسایی می‌شوند و این ویژگی‌ها نسبت به حالت چهره و شرایط محیطی نیز مقاوم هستند. لازم به ذکر است DeepFace برای اولین بار دقت الگوریتم های باز شنا سی چهره را به دقت باز شنا سی چهره تو سط انسان (حدود ۹۷ در صد) ر سانید. پس از ارائه‌ی DeepFace الگوریتم‌های دیگری نیز بر پایه‌ی یادگیری عمیق باز شنا سی چهره کردند از جمله‌ی این روش‌ها می‌توان به FaceID ، VGGFace، VGGFace2 و FaceNet اشاره کرد [12] تا [15].

در ادامه پیشرفته‌ترین روش‌های بازشناسی چهره مبتنی بر یادگیری عمیق به همراه توضیحات آن آورده شده است.

DeepFace: از یک شبکه عصبی عمیق نه لایه با بیش از ۱۲۰ میلیون پارامتر برای بازشناسی چهره استفاده می‌کند و از خطای Softmax برای آموزش شبکه ا ستفاده شده ا ست و مجموعه داده های آموزش، یک مجموعه داده خصوصی با چهار میلیون تصویر چهره با بیش از ۴۰۰۰ هویت است [16]. این سامانه همچنین روش پیش‌پردازشی مؤثری را که از یک مدل سه‌بعدی برای تراز کردن چهره‌ها در موقعیت ا ستاندارد چهره ا ستفاده می شود، پیاده سازی می‌کند. به طور خلاصه، موفقیت DeepFace به سه عامل اصلی مربوط می شود: (۱) مرحله پیش‌پردازش دقیق، (۲) معماری شبکه و (۳) داده های آموزش در مق یاس بزرگ. علاوه بر سا ما نه‌ی پی شنهادی، DeepFace همچنین یک سی ستم تأیید چهره انتها به انتها[1] را با استفاده از یک شبکه Siamese ارائه می‌دهد. پس از آموزش، شبکه شامل یک لایه طبقه‌بندی است که برای تولید ویژگی برای دو تصویر به طور همزمان، تکرار می‌شود. بردارهای ویژگی تولید شده برای تصمیم‌گیری اینکه آیا دو تصویر از یک شخص هستند با یکدیگر مقایسه می‌شوند.

VGGFace: با الهام از VGGNet که نشان داد پیچیدگی‌های عمیق تر می‌توانند در تشخیص تصویر در مقیاس بزرگ مؤثرتر باشند، طراحی شده است، VGGFace همان مفهوم را برای بازشناسی چهره به کار می‌برد. نویسندگان از نسخه ا صلاح شده معماری ارائه شده در VGGNet است فاده کرده‌ا ند و روی مجمو عه داده VGGFace آموزش داده‌اند. نویسندگان دو تابع محاسبه خطا، softmax triplet را ارز یابی کردند و نتیجه گرفتند که از خطای سه‌گانه قطعاً عملکرد کلی بهتری را ارائه می‌دهد. با این وجود ، گزارش شده که آموزش شبکه برای طبقه‌بندی با خطای softmax، آموزش را به میزان قابل توجهی آسان و سریع‌تر می‌کند. بعدها از سامانه‌ی VGGFace برای یادگیری انتقالی با تطبیق الگو استفاده شد. در این پیاده‌سازی، ویژگی‌های CNN عمیق حاصل از VGGNet از پیش آموزش‌دیده با SVMهای خطی آموزش دیده، در زمان آزمون ترکیب می شود. گزارش شده ا ست که SVM های خطی one-vs-rest، قدرت تفکیک فضای ویژگی را افزایش می دهند.

OpenFace: یک سامانه باز شنا سی چهره تحت مجوز Apache 2.0 است. این سامانه با هدف از بین بردن فاصله بین سامانه‌های بازشناسی چهره در دسترس عموم و پیشرفته‌ترین سامانه‌های خ صو صی با کیفیت بالا، تو سعه یافته ا ست. این سامانه مبتنی بر مفاهیمی ا ست که در GoogleNet و FaceNet معرفی شده ا ست. OpenFace از نسخه اصلاح شده شبکه nn4 از GoogleNet استفاده می‌کند که در FaceNet نیز مورد استفاده قرار گرفته است. DNN با استفاده از خطای سه‌گانه آموزش داده می‌شود. بردارهای ویژگی خروجی به دست آمده از این مدل آموزش‌دیده دارای ۱۲۸ بعد هستند. طبقه‌بندی چهره با ا ستفاده از SVM خطی انجام می شود. با توجه به تصاویر چهره دارای برچسب از داده‌های آموزش، این سامانه برای هر چهره بردارهای مشخصه تولید می‌کند. سپس، بردارهای ویژگی به SVM داده می‌شوند که مدلی را بر اساس بردارهای ویژگی چهره ایجاد می‌کند. هنگامی که یک بردار ویژگی های چهره از یک تصویر چهره نا شناخته ارائه می شود، مدل SVM چهره ناشناخته را طبقه‌بندی می‌کند.

ArcFace: مجمو عه‌ای از DNN ها (ResNet-100، ResNet-50 و ResNet-34) را همراه با خطای ArcFace پیاده‌سازی می‌کند [17]و [18]. این سیستم یک بردار ویژگی ۵۱۲ بعدی را برای تصاویر چهره تولید می‌کند. DNNها روی نسخه اصلاح شده مجمو عه داده Ms Celeb آموزش د ید ند. در یک سری ن تایج

[1] end-to-end

آزمایشی، نویسندگان نشان می‌دهند که این پیاده‌سازی از اکثر نتایج پیشرفته گزارش شده بهتر است.

مجموعه دادگان LFW یک دادگان محک عمومی برای بازشناسی چهره است که به آن تطبیق جفت نیز می‌گویند. در جدول ۱ عملکرد برخی از الگوریتم‌های معروف بر روی این مجموعه داده را مشاهده می‌کنید که در وبسایت آن آورده شده است.[1]

نتایج مرور جدیدترین روش‌ها و پیشرفت‌های اخیر به ما نشان می دهد که افزایش چشمگیری در تحقیقات این حوزه طی پنج سال گذشته رخ داده است، به ویژه با ظهور رویکرد یادگیری عمیق که از محبوب‌ترین روش‌های بینایی ماشین به حساب می‌آید. علاوه بر این، پایگاه داده‌های متعدد چهره (دولتی و خصوصی) برای اهداف تحقیقاتی و تجاری در دسترس هستند و ویژگی‌های اصلی آن‌ها و پروتکل‌های ارزیابی ارائه شده است. تمرکز بر روی چهره‌های برچسب زده شده در پایگاه داده LFW از نظر روش، معماری، معیارها، دقت و پروتکل‌ها لازم است تا محققان بتوانند نتایج خود را با این پایگاه داده مرجع مقایسه کنند. می‌توان بیان کرد که تشخیص چهره دو بعدی هنوز به تحقیق و توسعه فنی برای دستیابی به تجزیه و تحلیل تصاویر نیاز دارد. از طرفی با توجه به پیشرفت تجهیزات نوین سخت افزاری، بازشناسی چهره سه‌بعدی نیز مورد توجه قرار گرفته است. توسعه اخیر حسگرهای سه‌بعدی رویکرد جدیدی را برای بازشناسی چهره نشان می‌دهد که می‌تواند بر محدودیت‌های اصلی فناوری‌های دوبعدی غلبه کند، به عنوان مثال، تغییرات ظاهری، عامل پیری، حالت، تغییرات در شدت نور و به طور کلی در حالات چهره، داده‌های از دست رفته، لوازم آرایشی و انسداد. اطلاعات هندسی ارائه شده توسط داده‌های چهره سه‌بعدی می‌تواند دقت تشخیص چهره را در شرایط نامساعد اکتسابی بهبود بخشد. با این حال، فقدان پایگاه داده بازشناسی چهره سه‌بعدی مانع بهره‌برداری از روش‌های مبتنی بر یادگیری عمیق می شود. همچنین، تفسیر حالت چهره سه‌بعدی، شناسایی تغییرات در سن و یادگیری انتقالی سه چالش دیگر این روش است که هنوز در آغاز کار خود هستند و نیاز به تحقیقات بیشتری دارد. به طور طبیعی، این پیشرفت‌های جدید در بازشناسی چهره باید چهار هدف را برآورده کند: سریع بودن (پاسخ فوری از دید کاربر)، دقت نزدیک به ۱۰۰، امنیت مطلوب، تجهیزات مینیاتوری و قابل حمل.

۴. تشخیص زنده بودن و رویکردهای آن

یک سامانه زیست‌سنجی می‌تواند تحت حملات مختلفی قرار بگیرد [۲۹]. با ایمن‌سازی نقاط خاصی از سامانه تشخیص، از جمله کانال‌های ارتباطی، تجهیزات و زیرساخت‌های درگیر، می‌توان از حملات غیرمستقیم جلوگیری کرد. روش‌های مورد نیاز برای بهبود این ماژول‌ها بیشتر مربوط به امنیت سایبری است تا زیست‌سنجی، بنابراین در این بحث پوشش داده نمی‌شوند. از طرف دیگر، حملات نمایش فقط یک آسیب‌پذیری زیست‌سنجی است که با سایر راه‌حل‌های امنیتی فناوری اطلاعات مشترک نیست و نیاز به اقدامات متقابل خاصی دارد.

حمله نمایش (PA) عبارتست از ارائه یک مصنوع ساخت انسان به یک سنسور دریافت داده در سامانه زیست‌سنجی. یک سامانه تشخیص زیست‌سنجی از زیرسامانه‌های مختلفی مانند ضبط مشخصه‌های زیست‌سنجی، پردازش سیگنال و استخراج ویژگی، مقایسه، تصمیم‌گیری و زیرسامانه ثبت در پایگاه داده تشکیل شده است. در صورت اضافه کردن زیرسامانه‌ی تشخیص حمله‌ی نمایش به این سامانه، این زیرسامانه می‌تواند در محل‌های مختلفی واقع شود: ۱) پس از زیرسامانه ضبط داده، ۲) در زیرسامانه ضبط داده، ۳) پس از زیرسامانه پردازش سیگنال و ۴) پس از زیرسامانه مقایسه یا تصمیم‌گیری. شکل ۲ تصویری از چارچوب کلی از سامانه زیست‌سنجی با تشخیص حمله نمایش را نشان می‌دهد [۲۹].

در راستای تعیین روش مناسب برای اجزای مختلف یک سامانه تشخیص زنده بودن مورد بررسی قرار می گیرد.

روش‌ها و گونه‌شناسی‌های موجود و رایج در مرجع [30]. به صورت زیر ارائه شده است: ۱) روش‌هایی که از حسگرهای رایج و موجود برای دریافت سیگنال‌های لازم جهت استخراج ویژگی‌هایی برای تشخیص زنده بودن استفاده می‌کنند. ۲) استفاده از سخت‌افزار اختصاصی برای استخراج شواهد کافی از زنده بودن است که این روش همیشه امکان استقرار ندارد. ۳) یک روش مبتنی بر تولید چالش-پاسخ که در آن از کاربر خواسته می‌شود تا با سامانه تعامل داشته باشد و بر اساس چالش تولید شده به سامانه پاسخ دهد. ۴) استفاده از الگوریتم‌های تشخیص ابتکاری که در سامانه پیاده‌سازی شده است [30].

از نگاه دیگری می‌توان این روش‌ها را به سه دسته طبقه‌بندی کرد: ۱) روش‌های مبتنی بر بافت ۲) روش‌های مبتنی بر حرکت ۳) روش های یادگیری عمیق.

در تحقیقی دیگر روش‌های موجود در برابر حملات نمایش دوبعدی را به پنج دسته طبقه‌بندی شده است: مبتنی بر بافت، مبتنی بر کیفیت تصویر (در این حوزه با توجه به ویژگی‌های کیفی تصویر، تحلیل انجام می پذیرد و در حوزه احراز هویت برخط، کیفیت تصویر یک چالش بوده و نیاز است با روش هایی کیفیت به میزان

[1] http://viswww.cs.umass.edu/lfw/

قابل قبول ارتقاء یابد.)، رویکردهای پویا، ویژگی های آموزش داده شـــده و روش‌های ترکیبی. طرح‌های مبتنی بر بافت عمدتاً تفاوت الگوهای ریز بافتی چهره‌ها و مصـــنوعات واقعی را با کمک توصـــیف گرهای مختلف کشف می‌کنند[31].

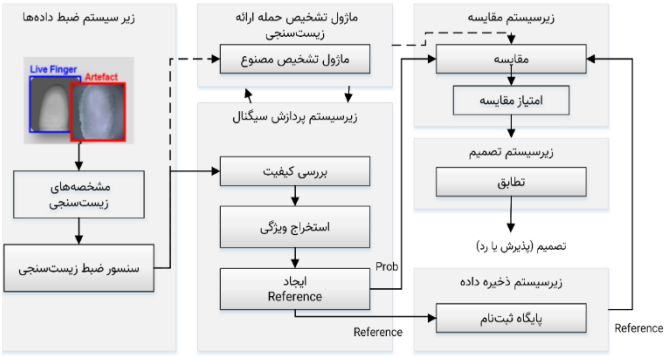
رویکردهای پویا، از اطلاعات زمانی برای کشـــف الگوهای حرکتی در فریم‌هـــای ویـــدیویی بهره‌برداری می‌کننـــد. یـــادگیری عمیق برای ا ستخراج ویژگی‌های سازگار ا ستفاده و رویکرد دیگر، تو سعه روش های مبتنی بر ترکیب است که با ترکیب ویژگی‌های مختلف از نقاط قوت هر زمینه بهره‌مند می‌شوند.

این پژوهش تطبیق چهره و تشخیص زنده بودن بر اساس تشخیص گفتار مورد بررسی قرار گرفته و در حال صنعتی می‌توان نرم‌افزار مرتبط را به صورت بومی تولید نمود. با توجه به پژوهش‌های پیشین، ما پیشنهاد می‌کنیم روش‌های موجود در این حوزه به سه نوع اصلی تقسیم شود (شکل۳): (۱) روش‌های تعاملی، (۲) روش‌های غیرتعاملی (منفعل) و (۳) ترکیب روش‌ها. روش‌های غیر تعاملی تنها ویدیو و یا تصویر کاربر ضبط شده و مورد تحلیل و پردازش قرار می‌گیرد. در روش تعاملی کاربر بایستی در مقابل دوربین یک وظیفه معین را انجام دهد. روش‌های تعاملی به این دلیل که در آن‌ها روش تشخیص زنده بودن به کاربر گفته می‌شود از امنیت کمتری برخوردار هستند اما معمولا از دقت بیشتری نسبت به روش‌های غیرتعاملی برخوردار هستند ازین رو ترکیب این دو روش می‌تواند دقت قابل توجهی را در اختیار ما قرار دهد. در شکل ۳ یک طبقه‌بندی از روش‌های تشخیص زنده بودن را مشاهده می‌کنید.

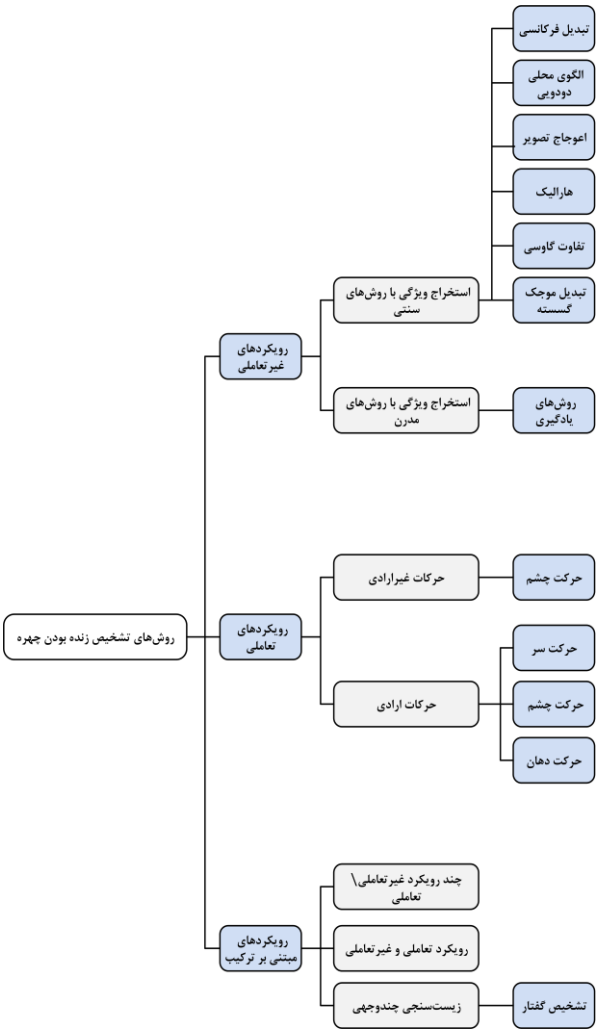
در جدول ۲ اطلاعات الگوریتم‌های مختلف، به طور خلاصـــه آورده شده است.

جدول ۱. نتایج رویکردهای مختلف بازشناسی چهره بر روی مجموعه داده LFW

الگوریتم	مرجع	بازیابی چهره(٪)	الگوریتم	مرجع	بـــازیـــابـــی چهره(٪)
Deep Face	[16]	97.35	FaceNet	[22]	99.63
DeepFR	[19]	98.95	DeepID2+	[19]	99.47
Center Face	[20]	99.2	8 Baidu	[20]	99.13
SphereFace	[21]	99.42	VGGFace	[12]	99.13
Face++	[22]	99.50	FR+FCN	[19]	96.45
DeepID	[23]	97.45	GaussianFace	[21]	98.52
DeepID2	[24]	99.15	DeepID3	[11]	99.53
YouTu Lab, Tencent	[25]	99.80	PingAn AI Lab	[16]	99.80
Fisher vector faces	[26]	93.03	CMD+SLBP	[27]	92.58
Simile classifiers	[27]	84.72	DFD	[28]	84.02
LBP LDA	[28]	87.33	LBP multishot	[24]	85.17



شکل ۲. چارچوب کلی سامانه زیست‌سنجی با تشخیص حمله نمایش [29]



شکل ۳. گونه‌شناسی روش‌های تشخیص زنده بودن با استفاده از چهره

جدول ۲: خلاصه‌ی اطلاعات روش‌های بیان شده برای تشخیص زنده بودن.

نویسنده و سال	روش	پایگاه داده	نوع حمله	نتایج
Matta و همکاران [32] ۲۰۱۱	مبتنی بر الگوهای دودویی محلی (LBP)، طبقه‌بندی با ماشین بردار پشتیبان SVM	NUAA	تصویر چاپ شده و نمایش از روی صفحه نمایشگر	سطح زیر نمودار (AUC) برابر با ۰,۹۹ و نرخ خطای تساوی (EER) معادل با ۲,۹
Agarwal و همکاران[33] ۲۰۱۶	استفاده از ویژگی‌های هارالیک بافت چهره، جداسازی کانال‌های رنگی تصویر، استفاده از تبدیل موجک گسسته، بلوک‌بندی تصویر، طبقه‌بندی با SVM	3DMAD, CASIAFASD, MSU-MFSD	ماسک و بازپخش ویدیو	بدون تقسیم به بلوک‌ها، HTER برابر با ۴,۱٪ مبتنی بر بلوک‌ها، HTER برابر با ۳٪، استفاده از کانال قرمز با استخراج ویژگی بلوک‌ها، HTER برابر با صفر
Yang و همکاران[34] ۲۰۱۳	استخراج ویژگی با هیستوگرام شیب‌های جهت‌دار (HoG) و تفاوت گاوسی(DoG)، استفاده از LBP، ناحیه‌بندی صورت، طبقه‌بندی با SVM	NUAA, CASIA, PRINT-ATTACK	تصویر چاپ شده و نمایش از روی صفحه نمایشگر	بهترین نتیجه حاصل از ترکیب روش‌ها و تقسیم صورت به ناحیه‌های مختلف می‌باشد. دقت‌ها بر روی داده‌های NUAA: ۰,۹۷۷
He و همکاران [35] ۲۰۱۹	ترکیب مدل‌های مختلف فضاهای رنگی، ویژگی‌های LBP و CM، طبقه‌بندی با SVM	Replay-Attack, CASIA-FASD	بازپخش ویدیو	بهبود عملکرد تشخیص با استخراج ویژگی های LBP در فضای رنگی YCbCr و Luv و ویژگی‌های لحظه‌ای رنگ در فضای رنگی HSV
Mahore و همکاران [36] ۲۰۱۸	روش ترکیبی مبتنی بر ویژگی‌های بافت، ترکیب تبدیل موجک گسسته و الگوی دودویی محلی در فضای رنگی YCbCr، بلوک‌بندی تصویر	3D Mask Attack	ماسک سه‌بعدی	بدون استفاده از بلوک‌ها، دقت ۹۷,۱٪ و HTER ۲,۵٪ بلوک‌بندی با اندازه بلوک ۱۶*۱۶ با دقت ۹۹,۹۶٪ و HTER ۰,۰۱٪
Uzun وهمکاران [37] ۲۰۱۸	ترکیب زیست‌سنجی صدا و گفتار، استفاده از چالش و پاسخ (کپچا)، تبدیل گفتار به متن	CASIA Face Anti-Spoofing	تصویر چاپ شده، ماسک و بازپخش ویدیو	۹۷٪ از نمونه های اصلی به درستی شناسایی شده در سامانه تشخیص گفتار دقت درستی پاسخ به کپچا توسط شرکت‌کنندگان ۸۹,۲ درصد
Zhang و همکاران [38] ۲۰۲۰	استخراج ویژگی با شبکه عصبی، استفاده از ویژگی معنایی و هندسی تصویر برای طبقه‌بندی	Celeba spoof	تصویر چاپ شده، ماسک دوبعدی و بازپخش ویدیو	عملکرد بهبود سامانه با ویژگی‌های معنایی به ویژه نوع جعل، تاثیر اطلاعات هندسی در تشخیص در بازپخش ویدیو

از نتایج ارزیابی ارائه شده، می‌توان دریافت که تشخیص زنده بودن چهره هنوز یک مشکل بسیار چالش برانگیز است. به طور خاص، عملکرد روش های فعلی هنوز پایین‌تر از الزامات اکثر برنامه های کاربردی در دنیای واقعی (به ویژه از نظر قابلیت تعمیم) هستند.

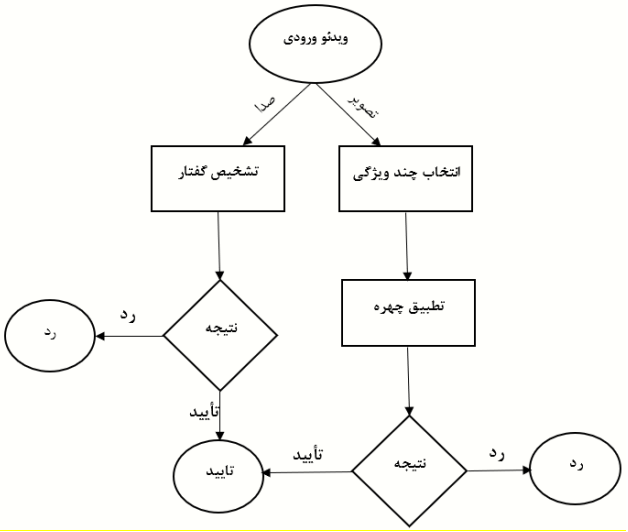
با این حال، همه ویژگی‌هایی که به صورت دستی استخراج می‌شوند توانایی تعمیم محدودی را دارند، زیرا به اندازه کافی قدرتمند نیستند که بتوانند همه تغییرات احتمالی را در شرایط مختلفِ ضبط چهره به دست آورند. ویژگی‌های آموزش دیده که توسط شبکه‌های عصبی عمیق استخراج می‌شوند در مقایسه با حجم محدود داده‌های آموزشی، دارای ابعاد بسیار بالایی هستند و از بیش‌برازش و در نتیجه تعمیم‌پذیری ضعیف رنج می‌برند. بنابراین، ویژگی های یادگیری که قادر به تمایز بین یک چهره واقعی و هر نوع PA باشند، احتمالاً با توجه به شرایط، بسیار متفاوت هستند. همان‌طور که قبلاً گفته شد، ویژگی‌های یادگیری که بتوانند به حد کافی بین چهره‌های واقعی و PAهای مختلف، تمایز قائل شوند هنوز یک چالش بزرگ است. البته، این نوع مسائل (مربوط به قابلیت‌های تعمیم مدل‌های داده محور) در زمینه بینایی ماشین، بسیار فراتر از تشخیص زنده بودن چهره است. برای مقابله با تمام حملاتی که قبلاً دیده شده است؛ روند مناسب، ترکیب چندین روش است. با این حال، با توجه به چالش‌های ذکر شده در ایجاد مجموعه داده و همچنین پیشرفت های تکنولوژیکی که کاربران کلاهبردار می‌توانند به منظور توسعه حملات پیچیده به آن دسترسی پیدا کنند، روش‌های تشخیص زنده بودن ممکن است مجبور به شناسایی حملاتی شوند که در مجموعه داده آموزشی آن گنجانده نشده است.

روش‌های تشخیص زنده بودن می‌توانند در سناریوهایی که کاربر فقط به دوربین‌های RGB دستگاه‌های عمومی دسترسی دارد، به کار گرفته شود. روش‌های مبتنی بر بافت که بیشترین استفاده را در تشخیص زنده بودن دارند و به ویژه روش‌های مبتنی بر بافت پویا، قادر به تشخیص تقریباً همه انواع حملات هستند. علاوه بر این، روش های مبتنی بر ویژگی های بافتی که با استفاده از یادگیری عمیق آموزش دیده‌اند، در مقایسه با روش‌های مبتنی بر ویژگی‌های بافت دستی، به طور چشمگیری عملکردهای تشخیص زنده بودن چهره را بهبود بخشیده است.

همان‌طور که پیش‌تر نیز به آن اشاره گردید ترکیب روش‌ها یکی از مواردی است که دقت یک سامانه تشخیص زنده بودن را به طور چشمگیری افزایش می‌دهد یکی از این روش‌ها، ترکیب چند روش زیست سنجی می‌باشد. از این رو در این پژوهش سعی بر آن شده که تشخیص گفتار که یکی از فناوری‌های به روز به خصوص در زبان فارسی می‌باشد به عنوان یک روش تشخیص زنده بودن مورد مطالعه قرار گرفته و یک روش مناسب برای آن ارائه گردد. با توجه به الگوریتمی که در این پژوهش بر روی آن کار شده است، تشخیص گفتار می‌تواند به عنوان یک رویکرد تعاملی نیز طبقه‌بندی شود.

۵. روش پیشنهادی و کارهای انجام شده

در احراز هویت غیرحضوری دو سرویس اصلی تطبیق چهره و تشخیص زنده بودن به کار گرفته می‌شود که پیاده‌سازی آن‌ها به عنوان هسته مرکزی پردازش هوش مصنوعی شناخته می‌شود. این پردازش مرکزی با بهره‌گیری از مدل‌های یادگیری عمیق مبتنی بر پردازش تصویر و گفتار انجام می‌شود که ادامه برای دو موضوع تطبیق چهره و تشخیص زنده بودن تشریح می‌شود. معماری روش پیشنهادی در شکل ۴ ارائه شده است. تشخیص گفتار و همچنین تطبیق چهره دو عامل مهم برای احراز هویت فرد محسوب می‌شود که در صورت تایید هردو بخش، هویت فرد تایید می‌گردد. از آنجاییکه در تشخیص گفتار این امکان وجود دارد که فرد دیگری جملات را تکرار کند لذا این روش به صورت تنها، دارای امنیت بالا نیست و از طرفی صدای مرجع برای همه افراد وجود ندارد و نیاز است با ترکیب روش های دیگر زنده بودن غیرفعال، احراز هویت انجام می پذیرد.



شکل ۴. معماری روش پیشنهادی برای احراز هویت

۵-۱ تطبیق چهره در روش پیشنهادی

برای انجام تطبیق چهره در دو عکس، اولین مرحله مکان‌یابی چهره[1] می‌باشد. مکان یابی چهره یک فناوری رایانه‌ای مبتنی بر هوش

[1] Face Detection

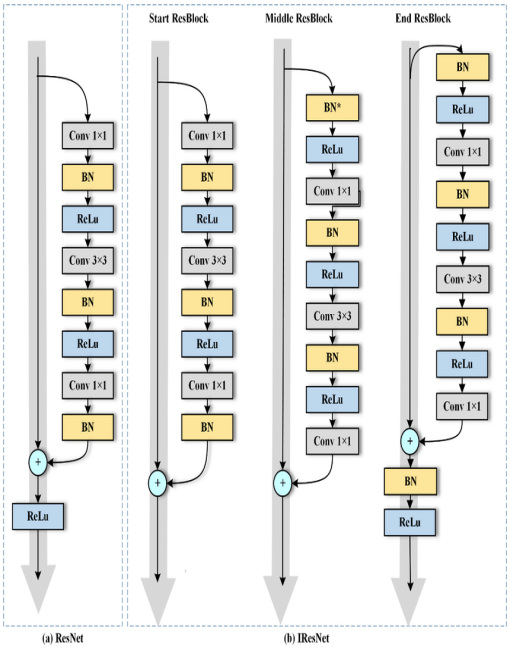
مصنوعی است که برای یافتن و شناسایی چهره انسان در تصاویر دیجیتال استفاده می‌شود. فناوری مکان‌یابی چهره را می‌توان در زمینه‌های مختلف، از جمله امنیت و زیست‌سنجی به کار برد تا نظارت و ردیابی افراد را در زمان واقعی ارائه دهد. مکان‌یابی چهره از تکنیک‌های پایه‌ای بینایی ماشین تا شبکه‌های عصبی مصنوعی پیچیده، برای پیدا کردن مکان چهره بهره گرفته است و اکنون نقش مهمی را به عنوان اولین گام در بسیاری از برنامه‌های کلیدی ایفا می‌کند، از جمله ردیابی چهره، تجزیه و تحلیل چهره و بازشناسی چهره. مکان‌یابی چهره تأثیر قابل توجهی بر نحوه انجام عملیات در این برنامه‌ها را دارد.

این بخش به عنوان یکی از بخش‌های پیش‌نیاز در هسته هوش مصنوعی محسوب می‌شود. به منظور بررسی هر دو مولفه‌ی سرعت و دقت مدل‌های مختلف مکان‌یابی چهره مورد بررسی قرار گرفته است. تصاویر وارد شده به سیستم را می‌توان به دو دسته تصاویر کنترل شده و تصاویر کنترل نشده دسته‌بندی نمود. تصاویر کنترل شده تصاویری هستند که زاویه، اندازه و کیفیت چهره در آن‌ها دارای استانداردهای کافی باشد. با توجه به دو شاخص دقت و سرعت مدل های مختلفی ارائه گردید. یکی از مدل ها برای کاربردهایی که در آن تصویر ورودی کنترل‌شده باشد قابل قبول بوده و به لطف پیچیدگی پرداز شی پایین آن، سرعت بسیار بالایی دارد. این مدل با بهره‌گیری از ویژگی های HOG تصویر، مکان چهره‌ها در تصویر را محاسبه می‌کند.

در یکی از بررسی‌های دیگر سعی بر آن شد تا این مدل علاوه بر داشتن سرعت پردازش نسبتا بالا، دقت مناسبی را نیز ارائه دهد و در اکثر موارد گزینه مناسبی برای استفاده در کاربردهای احراز هویت از راه دور است. این مدل بر پایه شبکه عصبی و از لایه‌های پیچشی سبک برای استخراج ویژگی استفاده شده است و در نهایت یک مدل مورد بررسی قرار گرفت که دقت مکان‌یابی چهره این مدل بسیار بالا بوده ولی نسبت به دو مدل قبلی پیچیدگی محاسباتی بیشتری داشته و سرعت کمتری دارد. در این مدل استخراج ویژگی با استفاده از مدل پایه ResNet-50 صورت می‌گیرد و در مواردی که تصویر یا ویدئو ورودی کنترل‌شده نباشد، می‌توان آن را مورد استفاده قرار داد. بنا بر نوع تصاویر ورودی، دقت مورد نیاز و توان سخت‌افزاری بستری که سامانه در آن اجرا می‌شود، می‌توان ماژول‌های مختلفی را برای استفاده در سرویس‌ها فعال‌سازی کرد.

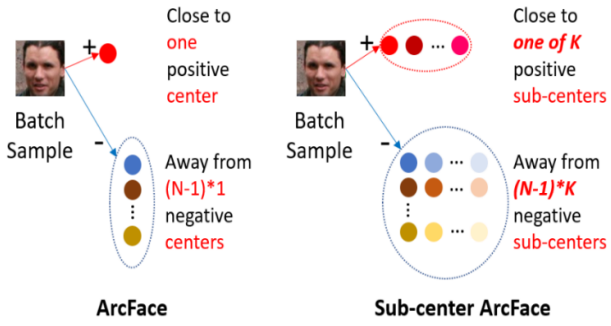
در این مقاله بر روی روش‌های مختلف تطبیق چهره کار شده است. همان‌طور که در بخش‌های قبلی نیز به آن اشاره گردید، شبکه‌های عصبی توانسته‌اند نتایج قابل توجهی را در این حوزه ارائه دهند در نتیجه سعی بر آن شد که بر روی مدل‌های مجزایی برای تطبیق چهره کار شود که در هر کدام از معماری‌های متفاوتی برای شبکه استفاده کرده‌اند. در نتیجه‌ی بررسی‌ها، معماری‌های مبتنی بر شبکه‌های عمیق ResNet توانسته است نتایج خوبی را به خود اختصاص دهد.

مدل تطبیق چهره این سامانه بر پایه شبکه باقی‌مانده بهبودیافته (improved residual network یا iResNet) پیاده‌سازی شده است [39]. معماری iResNet به گونه‌ای طراحی شده که اجازه افزایش عمق شبکه عصبی نسبت به ResNet را می‌دهد؛ در این سامانه از مدلی با عمق ۱۰۰ لایه برای استخراج ویژگی از چهره استفاده شده است. تفاوت بلوک‌های iResNet در مقابل ResNet در شکل ۴ ملاحظه می‌شود.

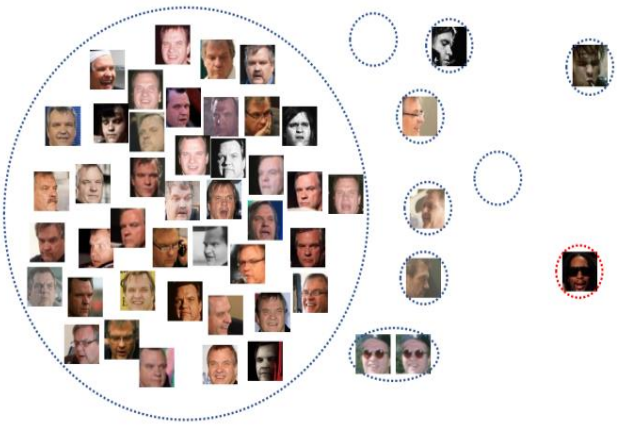


شکل ۴. مقایسه بلوک iResNet در مقابل ResNet [39]

نحوه آموزش شبکه با توجه به مقاله Sub-Center ArcFace انجام شده است [17]. خطای حین آموزش در این مقاله خود بر اساس مقاله ArcFace [18] طراحی شده است. در حین آموزش با تابع خطای ArcFace، شبکه تلاش می‌کند که فاصله بین ویژگی‌های به دست آمده از یک چهره و مرکز دسته خود را نسبت به مرکز دسته‌های دیگر (هویت‌های دیگر) را کاهش دهد. اما در روش Sub-Center-ArcFace با درنظر گرفتن چندین مرکز دسته برای هر هویت آن را بهبود داده است. شکل ۵ تفاوت‌های این دو روش را به نمایش می‌گذارد و شکل ۶ نمونه‌هایی از مراکز دسته که در روش Sub-Center ArcFace انتخاب می‌شود را نشان می‌دهد.



شکل ۵. مقایسه ArcFace در مقابل Sub-Center-ArcFace [17]



شکل ۶. مراکز دسته نمونه الگوریتم Sub-Center-ArcFace [17]

آموزش اولیه این شبکه با استفاده از دادگان Ms-Celeb-1M با ۱۰۰ هزار هویت و ۱۰ میلیون تصویر انجام شده است و برای همخوانی بیشتر با چهره‌های ایرانی با دادگان ایرانی نیز بازآموزی شده است.

۵-۲ تشخیص زنده بودن مبتنی بر تشخیص گفتار در روش پیشنهادی

همان‌طور که در پیشینه پژوهش بیان شد، روش های غیرتعاملی در تشخیص زنده بودن از امنیت بیشتری برخوردار هستند، به این دلیل اینکه سناریوی تشخیص زنده بودن به کاربر گفته نمی‌شود. با این وجود قرار گرفتن آن‌ها در کنار روش‌های تعاملی، دقت آن‌ها را تا مقدار قابل توجهی افزایش می‌دهد. یکی از روش‌های تشخیص زنده بودن تعاملی که در بین این روش‌ها از دقت بسیار بالایی برخوردار است، روش مبتنی بر تشخیص گفتار می‌باشد.

تشخیص گفتار فرایندی است که قادر است گفتار انسان را به متن تبدیل کند. فناوری تشخیص گفتار به رایانه‌ای که توانایی دریافت صدا را دارد برای مثال به یک میکروفن مجهز است این قابلیت را می‌دهد که صحبت کاربر را متوجه شود، این فناوری در تبدیل گفتار به متن یا به عنوان جایگزین برای ارتباط با رایانه کاربرد دارد. برقراری ارتباط گفتاری با رایانه‌ها به جای استفاده از صفحه کلید و ماوس یکی از زمینه‌های تحقیقاتی مهم چند دهه‌ی اخیر است.

در این پژوهش سعی بر آن شده است تا مدلی ارائه گردد که وظیفه تشخیص زنده بودن تعاملی با استفاده از گفتار را ارائه دهد. لازم به ذکر است که از آنجاییکه امکان دارد گفتار توسط شخص دیگری بیان شود، برای احراز هویت علاوه بر پردازش گفتار از روش‌های احراز هویت غیرفعال استفاده می‌شود.

برای استفاده از بازشناسی گفتار در تشخیص زنده بودن تصویر، تعدادی جمله اختصاصی طراحی شده است که دارای ویژگی‌های خاصی باشد مانند کوتاه، آسان و پر تکرار، موجود در واژگان سامانه بازشناسی گفتار و ساختار دستوری جملات بر اساس ساختار نحوی استاندارد فارسی باشد. این ویژگی‌ها از این جهت مد نظر بوده است که علاوه بر عمومی بودن جملات و استاندارد بودن آن‌ها، سامانه دارای دقت بالایی باشد. بهره‌گیری کاربر از یک مدل تشخیص گفتار در چند مرحله انجام می‌شود، یک جمله فارسی به کاربر نمایش داده می‌شود. این جمله می‌تواند از بین جملات ساده و روان در موضوعات مختلف استخراج شده باشد. از کاربر خواسته می‌شود که در یک ویدئو جمله‌ی مشخص‌شده را بخواند. در مرحله‌ی بعد ویدئوی دریافتی با استفاده از ماژول محاسبه امتیاز نتیجه نهایی اعلام می‌شود که این امتیاز که با پردازش صوت کاربر به دست آمده است. از معیار کمینه فاصله ویرایشی برای محاسبه شباهت دو جمله استفاده می شود. بر روی میزان شباهت به دست آمده با در نظر گرفتن میزان حساسیت سرویس (که هنگام فراخوانی دومین درخواست دریافت شده)، آستانه‌های مناسب اعمال می‌شود و نتیجه تشخیص زنده بودن مشخص خواهد شد. مدل تشخیص گفتار فارسی مورد استفاده در سامانه مبتنی بر یادگیری عمیق نویسا[1] است که آموزش یافته با حدود ۱۰۰۰ ساعت گفتار فارسی است که گفتار مورد استفاده از گوینده‌های فارسی زبان در سراسر کشور با تنوع لهجه، جنسیت، سن، تحصیلات و با جملات محاوره/رسمی تهیه شده است.

۶. نتایج و ارزیابی

در این بخش، دادگان‌ها و معیارهای ارزیابی مدل‌ها معرفی و در انتها نتایج ارزیابی آن‌ها بر روی مجموعه داده‌های مختلف ارائه می‌گردد. دادگانی که برای ارزیابی الگوریتم های تطابق چهره استفاده شده است عبارتند از LFW پاک‌سازی شده، مجموعه‌داده LFW استاندارد، مجموعه داده LFW با ۱ میلیون Non-match و

[1] PersianSpeech.com

مجموعه داده LFW با ۱۰ میلیون Non-match که در جدول ۳ توضیحات مربوط به آن‌ها را مشاهده می‌کنید.

برای تشخیص چهره از شبکه عصبی CNN و resnet برای استخراج ویژگی استفاده شده است. برای تشخیص گفتار از ویژگی تبدیل فرکانسی مل[1] (MFCC) استفاده شده است.

برای بررسی مدل تطبیق چهره از معیارهای نرخ خطای برابر[2](EER)، خطای ناحیه زیر منحنی[3] در نمودار منحنی مشخصه‌ی عملکرد پذیرنده[4](ROC) که نشان‌دهنده‌ی عملکرد یک مدل طبقه‌بندی در تمام آستانه‌های طبقه‌بندی است و دقت[5] طبقه‌بندی (ACC) استفاده می شود. نتایج ارزیابی الگوریتم تطابق چهره ارائه شده را با معیارهای EUC ، EER و ACC را در جدول ۴ مشاهده می‌کنید.

از آنجا که نتایج روش های مختلف دیگر بر روی دادگان LFW در جدول ۱ ارائه شده است، مقایسه نتایج روش پیشنهادی بیانکر کارایی بالای این روش در مقایسه با روش‌های مشابه است. برای ارزیابی تشخیص زنده بودن تعاملی با روش تشخیص گفتار از دو دادگان تهیه شده برای زبان فارسی استفاده شد که خلاصه مشخصات آن‌ها در جدول ۵ آورده شده است.

علاوه بر دادگان‌های بزرگ بیان شده، یک ارزیابی روی تعدادی از گوینده‌ها (۱۰ نفر) که جملات طراحی شده برای تشخیص زنده بودن در این پژوهش را خوانده‌اند صورت گرفته است که نتایج مدل تشخیص گفتار روی این سه دادگان در جدول ۶ آورده شده است که در آن معیار ارزیابی نرخ خطای کلمه[6] (WER) است که بیانگر درصد کلمات تشخیص داده شده به صورت نادرست در جملات تست است. همان‌طور که مشخص است، کارایی سرویس روی جملات مورد استفاده برای تشخیص زنده بودن (به دلیل سادگی جملات و عدم وجود کلمات خارج از واژگان) بالاست. از آنجا که روش مرجع مرتبطی برای مقایسه نتایج این بخش از کار در زبان فارسی وجود ندارد، امکان مقایسه مستقیم با پژوهش‌های پیشین وجود ندارد و طرح پیشنهادی در صنعت به صورت بومی تولید و مورد استفاده قرار گرفته است.

[1] Mel Frequency Cepstral Coefficients
[2] Equal Error Rate (EER)
[3] Area Under the Curve
[4] Receiver Operating Characteristic (ROC)
[5] Accuracy
[6] Word error rate (WER)

جدول ۳. اطلاعات دادگان آزمایش برای مدل تطبیق چهره.

مجموعه داده	توضیحات
مجموعه‌داده LFW پاک‌سازی شده	این مجموعه‌داده دارای ۱۲۰۰ تصویر است. به‌منظور ساخت این مجموعه‌داده اقدام به پالایش مجموعه‌داده LFW استاندارد شد. بدین صورت که تصاویر با بیش از یک چهره از داخل مجموعه‌داده حذف شدند. سپس اقدام به ساخت دسته‌های match و nonmatch شد. تعداد pairهای تولید شده برای هر کدام از دسته‌های match و nonmatch برابر با ۶۰۰ جفت است.
مجموعه‌داده LFW استاندارد	این مجموعه‌داده دارای ۱۳۲۳۳ تصویر متعلق به ۵۷۴۹ شخصیت می‌باشد.۳۰۰۰ جفت تصویر به‌عنوان match و ۳۰۰۰ جفت تصویر به‌عنوان non-match در نظر گرفته شده‌اند.
مجموعه داده LFW با ۱ میلیون Non-match	این مجموعه‌داده همان مجموعه‌داده LFW استاندارد است با این تفاوت که ۱۲۱۱۲۸۵ جفت تصویر به‌عنوان non-match و ۲۴۲۲۵۷ به‌عنوان match در نظرگرفته شده است.
مجموعه‌داده LFW با ۱۰ میلیون Non-match	این مجموعه‌داده همان مجموعه‌داده LFW استاندارد است با این تفاوت که ۱۰ میلیون جفت تصویر به‌عنوان non-match و ۲۴۲۲۵۷ به‌عنوان match در نظر گرفته شده است.

جدول ۴. نتایج مدل تطبیق چهره بر روی دادگان آزمایش.

مجموعه داده	ACC	EUC	EER
مجموعه‌داده LFW پاک‌سازی شده	۰,۹۹۵	۰,۰۰۲۲۶۲	۰,۰۰۳۲۸۹
مجموعه‌داده LFW استاندارد	۰,۹۹۸	۰,۰۰۰۹۱۴۸	۰,۰۰۳
مجموعه داده LFW با ۱ میلیون Non-match	۰,۹۹۹۸	۰,۰۰۰۱۱۹	۰,۰۰۰۳۷۹۱
مجموعه‌داده LFW با ۱۰ میلیون Non-match	۰,۹۹۹۹	۰,۰۰۰۱۱۸۹	۰,۰۰۰۳۷۹۲

جدول ۵. اطلاعات دادگان آزمایش مدل تشخیص گفتار.

مجموعه داده	توضیحات
DeepMine	شامل ۱۰۲۱ جمله فارسی است که توسط گوینده‌های زن و مرد (با نسبت حدود ۴۰٪ زن و ۶۰٪ مرد) با خواندن جملات مختلف جمع‌آوری شده است. در جمع آوری داده‌ها از یک اپ موبایل استفاده شده و جملات در شرایط واقعی و در محیط‌های متنوع ضبط شده است که در آنها نویز و جملات محاوره هم وجود دارد.
میکروفونی	شامل ۱۷۸۰ جمله رسمی فارسی ضبط شده با صدای زن و مرد (با نسبت حدود ۳۰٪ زن و ۷۰٪ مرد) با استفاده از میکروفون‌های متصل به رایانه و در محیط‌های واقعی است.
جملات نمونه خوانده شده	شامل جملات طراحی شده برای تشخی‌ص زنده بودن از ۱۰ نفر(۵ نفر زن و ۵ نفر مرد) هرکدام ۱۰۰ جمله خوانده شده و در مجموع ۱۰۰۰ جمله خواند شده است.

جدول ۶: ارزیابی سرویس تشخیص گفتار با معیار نرخ خطای کلمه.

مجموعه داده	WER
DeepMine	۰,۰۹۱
میکروفونی	۰,۰۴۸
جملات نمونه خوانده شده	۰,۰۱۹

۷. خلاصه و جمع‌بندی

احراز هویت الکترونیکی به عنوان یک راه‌حل مناسب برای پاسخگویی به عدم احراز هویت کاربران در سامانه‌های خدماتی موجود در بسیاری از سازمان‌ها ارائه شده است. افزایش تقاضای مردم، به ویژه در شرایطی مانند بحران بیماری کرونا، سبب شده است تا نیاز به احراز هویت غیرحضوری بسیار مورد توجه قرار بگیرد. احراز هویت الکترونیکی باعث افزایش سرعت در یافت خدمات، کاهش مراجعات حضوری به دفاتر و دستیابی به مزایای ناشی از آن (ترافیک، زمان، سلامتی و ...)، فراهم کردن دسترسی شبانه‌روزی و حتی در روزهای تعطیل به خدمات انتظامی، امکان دریافت خدمات به صورت ساده و آسان، صرفه جویی در زمان افراد با حذف مراجعه حضوری و منتظر ماندن در دفاتر، صرفه جویی در هزینه با توجه به کاهش تردد و کمک به سلامتی و جلوگیری از شیوع در مواردی مانند بحران کرونا می‌شود.

از فرایندهای اصلی احراز هویت الکترونیکی تایید هویت فرد می‌باشد که این کار می‌تواند به کمک تطبیق چهره انجام گردد. به دلایل دسترسی آسان، عدم تماس فیزیکی، موجود در تمام بانک‌های اطلاعاتی چهره می‌تواند انتخاب مناسبی باشد. یکی از چالش‌های تطبیق چهره در دنیای واقعی این است که فردی که مقابل دوربین قرار گرفته است، شخص واقعی است به این معنا که، نمایشی از تصویر چاپ شده فرد و یا نمایش تصویر وی از روی صفحه نمایشگر و یا ماسک چهره‌ی او نباشد که توسط فردی سودجو به سامانه ارائه شده است. از این رو مفهومی به نام تشخیص زنده بودن مطرح می‌گردد که وظیفه حل این چالش را بر عهده دارد. رویکردهای مختلفی برای تشخیص زنده بودن تصویر چهره فرد وجود دارد که یکی از طبقه‌بندی‌های معروف در این حوزه بدین شرح است: رویکرد غیرتعاملی، رویکرد تعاملی و ترکیب رویکردها. همان‌طور که در بررسی روش‌های تشخیص زنده بودن به آن اشاره گردید روش های غیرتعاملی به این دلایل که از نظر امنیتی در سامانه‌های زیست‌سنجی چهره، قوی‌تر هستند، فرایند روان‌تر و آسان‌تری دارد، سریع‌تر است و نرخ از قلم انداختن را به میزان قابل توجهی کاهش می‌دهد، دارای اهمیت بیشتری در صنعت می‌باشند. اما موضوع قابل بحث در مورد این سیستم‌ها دقت تشخیص آن‌ها می‌باشد. به دلیل وجود دوربین‌های مختلف با کیفیت‌های متفاوت و وجود ابزارهای حمله متفاوت تنوع داده‌ی مورد پردازش باید بسیار بالا باشد اما تا به امروز مجموعه داده‌ای ارائه نشده است که بتواند این محدودیت را تا حد قابل توجهی بپوشاند، از این رو روش‌های تعاملی در کنار روش های غیرتعاملی قرار می‌گیرند. در این میان روش‌های ترکیب زیست سنجی مختلف دقت سیستم را تا میزان قابل توجهی افزایش می‌دهد. قرار گرفتن یک مدل بازشناسی چهره با معماری شبکه پیچیده با دقت ۰,۹۹۵ در کنار یک سیستم تشخیص گفتار با قابلیت تولید جملات اتفاقی (که امکان پیش‌بینی جملات توسط کاربر را پایین می‌آورد) که نرخ خطای کلمه در جملات استاندارد ۰,۰۱۹ می‌باشد، می‌تواند یک سامانه احراز هویت الکترونیکی کارآمد در دنیای واقعی را ارائه دهد.

مراجع

[1] Givens, G. H., Beveridge, J. R., Phillips, P. J., Draper, B., Lui, Y. M., and Bolme, D., "Introduction to face recognition and evaluation of algorithm performance," *Comput. Stat. Data Anal.*, vol. 67, pp. 236–247, 2013.

[2] FRANCIS GALTON, "Personal Identification and Description 2," *Nature*, vol. 38, pp. 173–177, 1888.

[3] Hazim Barnouti, N., Sameer Mahmood Al-Dabbagh, S., and Esam Matti, W., "Face Recognition: A Literature Review," *Int. J. Appl. Inf. Syst.*, vol. 11, no. 4, pp. 21–31, 2016.

[4] Ding, X. and Fang, C., "Discussions on some problems in face recognition," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3338, pp. 47–56, 2004.

[5] Ahmad Tolba,Ali El-Baz, A. A. E.-H., "Face Recognition: A Literature Review," *Int. J. Appl. Inf. Syst.*, vol. 11, no. 4, pp. 21–31, 2016.

[6] Heisele, B., Ho, P., and Poggio, T., "Face recognition with support vector machines: Global versus component-based approach," in *Proceedings of the IEEE International Conference on Computer Vision*, 2001, vol. 2, pp. 688–694.

[7] Adjabi, I., Ouahabi, A., Benzaoui, A., and Taleb-Ahmed, A., "Past, present, and future of face recognition: A review," *Electron.*, vol. 9, no. 8, pp. 1–53, 2020.

[8] matthew a.turk, A. p. pentlan., "Face

I., and Zafeiriou, S., "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 5962–5979, 2022.

[19] Bowyer, K. W., Chang, K., and Flynn, P., "A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition," *Comput. Vis. Image Underst.*, vol. 101, no. 1, pp. 1–15, 2006.

[20] Li, X., Jia, T., and Zhang, H., "Expression-insensitive 3D face recognition using sparse representation," *2009 IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2009*, pp. 2575–2582, 2009.

[21] Drira, H., Ben Amor, B., Srivastava, A., Daoudi, M., and Slama, R., "3D Face recognition under expressions, occlusions, and pose variations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2270–2283, 2013.

[22] Gupta, S., Markey, M. K., and Bovik, A. C., "Anthropometric 3D face recognition," *Int. J. Comput. Vis.*, vol. 90, no. 3, pp. 331–349, 2010.

[23] Koudelka, M. L., Koch, M. W., and Russ, T. D., "A prescreener for 3D face recognition using radial symmetry and the Hausdorff fraction," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2005-September, pp. 1–8, 2005.

[24] Cover, T. M. and Hart, P. E., "Nearest Neighbor Pattern Classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[25] Tang, H., Yin, B., Sun, Y., and Hu, Y., "3D face recognition using local binary patterns," *Signal Processing*, vol. 93, no. 8, pp. 2190–2198, 2013.

[26] Lei, Y., Bennamoun, M., and El-Sallam, A. A., "An efficient 3D face recognition approach based on the fusion of novel local low-level features," *Pattern Recognit.*, vol. 46, no. 1, pp. 24–37, 2013.

[27] Berretti, S., Del Bimbo, A., and Pala, P., "3D face recognition using isogeodesic stripes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2162–2177, 2010.

[28] Chang, K. I., Bowyer, K. W., and Flynn, P. J., "Multiple nose region matching for 3D face recognition under varying facial recognition using eigenfaces," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1991.

[9] Sharif, M., Naz, F., Yasmin, M., Shahid, M. A., and Rehman, A., "Face recognition: A survey," *J. Eng. Sci. Technol. Rev.*, vol. 10, no. 2, pp. 166–177, 2017.

[10] Liu, C. and Wechsler, H., "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 467–476, 2002.

[11] Taigman, Y., Yang, M., Ranzato, M., and Wolf, L., "DeepFace: Closing the gap to human-level performance in face verification," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1701–1708, 2014.

[12] Sun, Y., Chen, Y., Wang, X., and Tang, X., "Deep learning face representation by joint identification-verification," *Adv. Neural Inf. Process. Syst.*, vol. 3, no. January, pp. 1988–1996, 2014.

[13] Parkhi, O. M., Vedaldi, A., and Zisserman, A., "Deep Face Recognition - Abstract only," *Procedings Br. Mach. Vis. Conf. 2015*, no. Section 3, pp. 41.1-41.12, 2015.

[14] Cao, Q., Shen, L., Xie, W., Parkhi, O. M., and Zisserman, A., "VGGFace2: A dataset for recognising faces across pose and age," *Proc. - 13th IEEE Int. Conf. Autom. Face Gesture Recognition, FG 2018*, no. May, pp. 67–74, 2018.

[15] Schroff, F., Kalenichenko, D., and Philbin, J., "FaceNet: A unified embedding for face recognition and clustering," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June-2015, pp. 815–823, 2015.

[16] Samir, C. *et al.*, "An Intrinsic Framework for Analysis of Facial Surfaces To cite this version□: HAL Id□: hal-00665862 An Intrinsic Framework for Analysis of Facial Surfaces," pp. 80–95, 2012.

[17] Deng, J., Guo, J., Liu, T., Gong, M., and Zafeiriou, S., "Sub-center ArcFace: Boosting Face Recognition by Large-Scale Noisy Web Faces," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12356 LNCS, pp. 741–757, 2020.

[18] Deng, J., Guo, J., Yang, J., Xue, N., Kotsia,

expression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1695–1700, 2006.

[29] Newton, E., "Overview of the ISO / IEC 30107 Project Authentication Use Case Comparison," pp. 1–13.

[30] Hernandez-Ortega, J., Fierrez, J., Morales, A., and Galbally, J., "Introduction to face presentation attack detection," *Adv. Comput. Vis. Pattern Recognit.*, no. April, pp. 187–206, 2019.

[31] Tseng, T. C., Shih, T. F., and Fuh, C. S., "Anti-spoofing of live face authentication on smartphone," *J. Inf. Sci. Eng.*, vol. 37, no. 3, pp. 605–616, 2021.

[32] Määttä, J., Hadid, A., and Pietikäinen, M., "Face spoofing detection from single images using texture and local shape analysis," *IET Biometrics*, vol. 1, no. 1, pp. 3–10, 2012.

[33] Agarwal, A., Singh, R., and Vatsa, M., "Face anti-spoofing using Haralick features," *2016 IEEE 8th Int. Conf. Biometrics Theory, Appl. Syst. BTAS 2016*, no. September, 2016.

[34] Yang, J., Lei, Z., Liao, S., and Li, S. Z., "Face liveness detection with component dependent descriptor," *Proc. - 2013 Int. Conf. Biometrics, ICB 2013*, 2013.

[35] He, J. and Luo, J., "Face Spoofing Detection Based on Combining Different Color Space Models," *2019 IEEE 4th Int. Conf. Image, Vis. Comput. ICIVC 2019*, pp. 523–528, 2019.

[36] Mahore, A. and Tripathi, M., "Detection of 3D Mask in 2D face recognition system using DWT and LBP," *2018 IEEE 3rd Int. Conf. Commun. Inf. Syst. ICCIS 2018*, pp. 18–22, 2019.

[37] Uzun, E., Chung, S. P. H., Essa, I., and Lee, W., "rtCaptcha: A Real-Time CAPTCHA Based Liveness Detection System," pp. 1–15, 2018.

[38] Zhang, Y. *et al.*, "CelebA-Spoof: Large-Scale Face Anti-spoofing Dataset with Rich Annotations," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12357 LNCS, pp. 70–85, 2020.

[39] Duta, I. C., Liu, L., Zhu, F., and Shao, L., "Improved residual networks for image and video recognition," *Proc. - Int. Conf. Pattern Recognit.*, pp. 9415–9422, 2020.