مجله‌ی ریاضی شریف
دوره‌ی سوم، شماره‌ی اول (بهار ۱۴۰۲)

احراز هویت خودکار بر اساس چهره

احمد رحیمی*

چکیده. امروزه در زندگی روزمره و بالاخص فعالیت‌های علمی، در سطوح مختلف، شاهد کاربردهای بسیار گسترده‌ی هوش مصنوعی هستیم. از ابزارهای مترجم متن گرفته تا جست‌وجو در تصاویر و ابزارهای تبدیل صوت به متن. یکی از این کاربردها که امروزه در بعضی لپ‌تاپ‌ها و گوشی‌های موبایل نیز مورد استفاده قرار می‌گیرد، قابلیت احراز هویت از طریق چهره است. این سیستم‌ها معمولاً به این شکل کار می‌کنند که در ابتدا تصویر یا فیلمی از چهره‌ی شما می‌گیرند و در دفعات بعدی که نیاز به احراز هویت باشد، با گرفتن عکس یا فیلمی جدید از چهره‌ی شما و تطبیق آن با تصویر یا فیلم اولیه، هویت شما را احراز می‌کنند. یکی از چالش‌های طراحی چنین سیستم‌هایی مطابقت‌دادن دو چهره با یکدیگر است. این مسأله که در حوزه‌ی بینایی کامپیوتر به طور مبسوط مورد مطالعه قرار گرفته، مسأله‌ی تأیید چهره نام دارد. در این نوشته ابتدا مدل Facenet را که از یادگیری متضاد استفاده می‌کند به عنوان راه‌حلی برای مسأله‌ی تأیید چهره معرفی می‌کنیم. سپس چالش زنده‌بودن را مطرح می‌کنیم و برای فائق‌آمدن بر آن، روشی برای تشخیص جهت صورت شخص در یک تصویر ارائه می‌دهیم. در نهایت اجزای مختلفی که ارائه شده را در کنار هم قرار داده و تصویری کلی از سیستم احراز هویت ارائه خواهیم داد.

۱. مقدمه

در قرن اخیر شاهد پیشرفت گسترده و پرشتابی در حوزه‌ی هوش مصنوعی بوده‌ایم؛ خصوصاً بعد از سال ۲۰۱۲ که ظهور شبکه‌های عصبی[1] باعث انقلابی در این حوزه شدند. یکی از حوزه‌هایی که شبکه‌های عصبی در آن منجر به پیشرفت شگرفی شده‌اند بینایی کامپیوتر است. بینایی کامپیوتر شاخه‌ای از هوش مصنوعی است که به کامپیوترها این قدرت را می‌دهد که اطلاعات معناداری از تصاویر، فیلم‌ها و سایر ورودی‌های بصری به دست آورند. مسائل متنوعی در بینایی کامپیوتر وجود دارد که از معروف‌ترین آن‌ها می‌توان به دسته‌بندی تصاویر[2]، تشخیص اشیا[3] و قطعه‌بندی تصاویر[4] اشاره کرد. مسأله‌ی ما که احراز هویت با استفاده از چهره می‌باشد نیز در حوزه‌ی بینایی کامپیوتر قرار می‌گیرد.

ما در این نوشته به دنبال طراحی سیستمی هستیم تا بتواند با دریافت یک ورودی بصری مانند تصویر یا فیلمی از چهره‌ی یک شخص و تطبیق آن با اطلاعاتی که پیشتر از آن شخص ذخیره کرده، مثلاً تصاویر و فیلم‌هایی که مطمئن است مربوط به این شخص هستند، هویت او را به صورت خودکار احراز کند. برای چنین سیستمی کاربردهای بسیاری می‌توان برشمرد؛ از استفاده در گوشی‌های هوشمند و لپ‌تاپ‌ها گرفته، تا انجام‌دادن کارهای اداری از راه دور. مورد اخیر خصوصاً در زمان همه‌گیری کرونا اهمیت پیدا می‌کند؛ چرا که افراد می‌توانند بدون نیاز به مراجعه‌ی حضوری و قرارگرفتن در معرض خطر ابتلا به بیماری کار اداری خود را انجام دهند.

مسأله‌ی مشابهی که در بینایی کامپیوتر وجود دارد، مسأله‌ی تأیید چهره[5] نام دارد. در این مسأله، دو تصویر متفاوت در حالت کلی از صورت انسان‌هایی داده شده و هدف این است که بفهمیم آیا این دو تصویر متعلق به شخص یکسانی هستند یا خیر. توجه کنید که این دو تصویر می‌توانند مربوط به زمان‌های متفاوتی باشند؛ پس‌زمینه‌های تصاویر می‌توانند با هم متفاوت باشند؛ جهت صورت در عکس‌ها می‌تواند یکسان نباشد و حتی فرد ممکن است در یکی از تصاویر با عینک و در دیگری بدون عینک باشد، یا در یکی از عکس‌ها ریش داشته باشد و در دیگری ریشش را زده باشد. در کنار این چالش‌ها، مقایسه‌ی دو تصویر از

[1] neural networks
[2] image classification
[3] object detection
[4] image segmentation
[5] face verification

صورت و تأیید چهره‌های آن دو، به خودی خود مسأله‌ی بسیار دشواری است و حل این سوال را سخت و چالش‌برانگیز خواهد کرد.

مرجع [۲] با کمک شبکه‌های عصبی و با کمک‌گرفتن از روشی به نام یادگیری متضاد[1] تلاش کرده این مسأله را حل کند و به نتایج بسیار خوبی هم دست یافته است. در این مقاله یک شبکه‌ی عصبی در نظر گرفته شده که با ورودی گرفتن تصویری از صورت شخص یک بردار ویژگی ۵۱۲ بعدی خروجی می‌دهد. سپس این شبکه‌ی عصبی طوری آموزش داده می‌شود که بردارهای خروجی دو تصویر مربوط به یک شخص تا جای ممکن نزدیک به هم باشند و بردارهای خروجی دو تصویر از دو شخص متفاوت تا جای ممکن از هم دور باشند. به این ترتیب، اگر دو تصویر از صورت دو نفر داشته باشیم با دادن این دو تصویر به این شبکه‌ی عصبی و اندازه‌گیری فاصله‌ی بردارهای ویژگی خروجی آن از یک‌دیگر می‌توانیم در مورد یکی‌بودن شخص موجود در دو تصویر یادشده اظهارنظر کنیم. در بخش ۲ به جزئیات این روش خواهیم پرداخت.

نکته‌ی قابل توجه درباره‌ی روش Facenet این است که به مسأله به شکل کلی نگاه می‌کند؛ در حالی که در کاربردی که ما به دنبال آن هستیم می‌توانیم روی برخی شرایط تصویر محدودیت‌هایی قرار دهیم. برای مثال، یکی از چالش‌های بزرگ حل مسأله‌ی تأیید چهره در حالت کلی، همان‌طور که در بالا اشاره شد، تغییر زاویه‌ی صورت در تصاویر است؛ اما در احراز هویت با استفاده از چهره، می‌توانیم از کاربر بخواهیم از روبه‌رو از صورت خود عکس بگیرد و چهره‌اش حالت خاصی نداشته باشد (مثلاً خندان یا درهم نباشد). در نتیجه به مدلی نیاز داریم که بتواند در یک تصویر جهت صورت را تشخیص دهد تا اگر از روبه‌رو نبود از کاربر بخواهیم تصویر دیگری را که از روبه‌رو گرفته شده است برای ما ارسال کند.

چالش دیگری که در مسأله‌ی ما وجود دارد چالش زنده‌بودن تصویر یا فیلم ارسالی است؛ زیرا گاهی می‌خواهیم مطمئن شویم تصویر یا فیلمی که کاربر ارسال می‌کند در همان لحظه گرفته شده است. این مسأله چالشی حیاتی است چرا که در صورت حل‌نشدن این مشکل، شخصی ممکن است با داشتن یک تصویر یا فیلم از شخص دیگری بتواند به جای او احراز هویت شود و احتمالا خرابی‌هایی در حساب او ایجاد کند. با این‌حال، این چالش با داشتن مدل تشخیص جهت صورتی که در بالا ذکر شد به سادگی قابل حل است. کافی است دنباله‌ای از جهت‌های تصادفی تولید کنیم و به کاربر بدهیم و از او بخواهیم فیلمی برای ما ارسال کند که در آن صورت خود را طبق دنباله‌ی داده‌شده از جهات، حرکت دهد. سپس با استفاده از مدل تشخیص جهت صورت، صحت انجام این حرکات را بررسی کنیم.

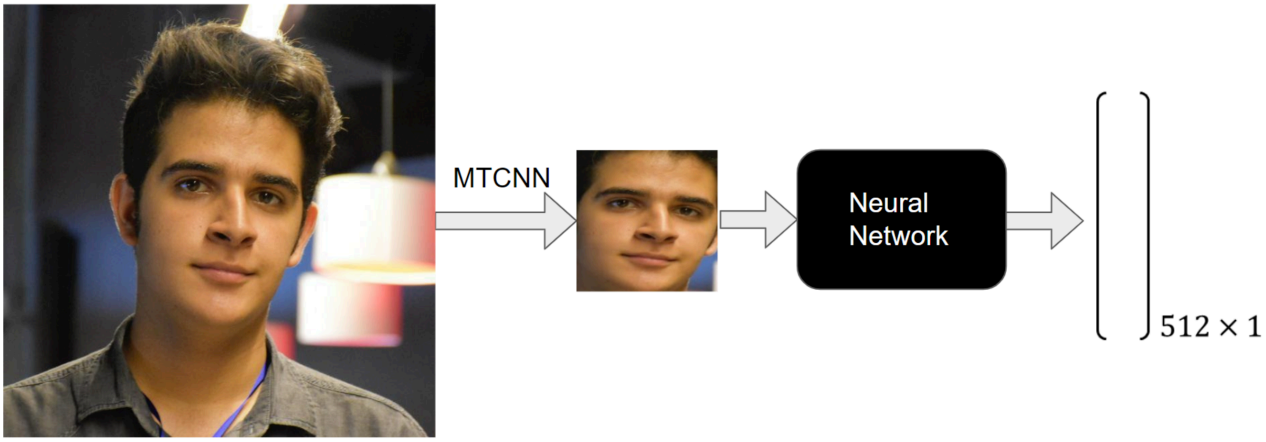
۲. روش Facenet

در این بخش روش Facenet را به اختصار توضیح می‌دهیم. هدف این است که خواننده در انتها اطلاعاتی کلی از نحوه‌ی کار کل سیستم به دست آورد. برای مطالعه‌ی بیشتر درباره‌ی این روش می‌توانید به [۲] مراجعه کنید.

در روش Facenet، تصویر ورودی ابتدا به الگوریتمی به نام MTCNN [۳] داده می‌شود تا مکان چهره در تصویر تشخیص داده شود. سپس تصویر ورودی را از آن قسمت برش[2] می‌دهیم و ابعاد آن را تغییر می‌دهیم تا قطعه عکسی مربعی به دست آوریم که تنها شامل صورت شخص و میزان پس‌زمینه‌ی آن تا جای ممکن کم است. سپس قطعه عکس حاصل را به شبکه‌ی عصبی می‌دهیم و بردار ویژگی ۵۱۲ بعدی عکس را از شبکه‌ی عصبی دریافت می‌کنیم (شکل ۱).

هدف این است که با انجام عمل فوق برای دو تصویر ورودی و مقایسه‌ی فاصله‌ی اقلیدسی بردارهای ویژگی نهایی آن‌ها، تشخیص دهیم دو تصویر از صورت یک شخص گرفته شده یا نه. برای آموزش شبکه‌ی عصبی در این روش، از یادگیری متضاد استفاده می‌شود. یادگیری متضاد به این شکل است که سه تصویر انتخاب می‌کنیم که دو تا از آن‌ها از یک شخص و سومی از شخص دیگری باشد. یکی از دو تصویری که از یک شخص گرفته شده را پایه[3] و دیگری را تصویر مثبت[4] می‌نامیم. تصویر سوم را نیز تصویر منفی[5] می‌نامیم. هدف این است که شبکه‌ی عصبی را طوری آموزش دهیم تا فاصله‌ی تصویر منفی از تصویر پایه بیشتر از فاصله تصویر مثبت از تصویر پایه شود (شکل ۲). به عبارت دیگر، اگر بردار ویژگی تصاویر پایه، مثبت و منفی را به ترتیب با v_a، v_p، v_n نشان دهیم، می‌خواهیم فاصله‌ی بردار ویژگی جفت تصویری که مربوط به یک شخص هستند، با یک

[1] contrastive learning
[2] crop
[3] anchor
[4] positive
[5] negative



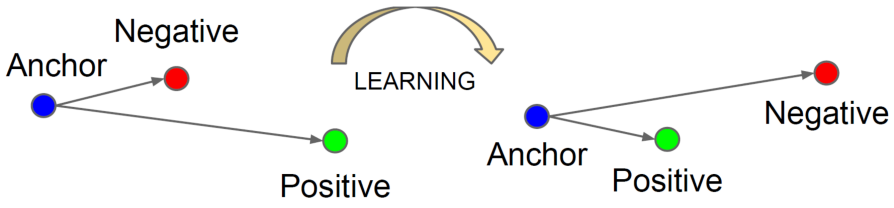
شکل ۱: روند روش Facenet که در آن ابتدا تصویر ورودی به مدل MTCNN داده می‌شود تا قطعه‌ای مربعی از صورت شخص در تصویر به دست آید. سپس آن را به شبکه‌ی عصبی می‌دهیم و بردار ویژگی را خروجی می‌گیریم.

حاشیه‌ی امن، کمتر از فاصله‌ی بردار ویژگی جفت تصویری که مربوط به دو شخص مختلف هستند شود:

$$\|v_a - v_p\|_۲^۲ + \alpha < \|v_a - v_n\|_۲^۲$$

که در آن α حاشیه‌ی امن بین جفت‌های مثبت و جفت‌های منفی است. در نتیجه شبکه‌ی عصبی را به گونه‌ای آموزش می‌دهیم که تابع هزینه‌ی زیر را کمینه کند:

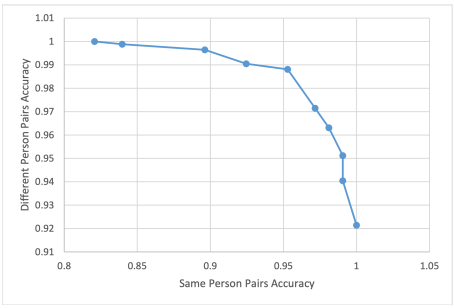
$$\max\left(\|v_a - v_p\|_۲^۲ - \|v_a - v_n\|_۲^۲ + \alpha, ۰\right)$$



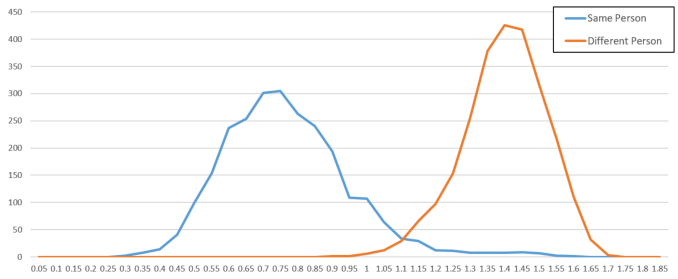
شکل ۲: یادگیری متضاد که در آن تلاش می‌شود بعد از یادگیری فاصله‌ی میان جفت پایه و مثبت کمتر از فاصله‌ی میان جفت پایه و منفی شود.

در نتیجه با آموزش شبکه‌ی عصبی به شیوه‌ی فوق، به ازای دو تصویر ورودی، از آن‌ها طبق روند شکل ۱ دو بردار ویژگی به دست می آوریم. با داشتن یک مقدار آستانه و مقایسه‌ی فاصله‌ی دو بردار ویژگی با این مقدار آستانه، می‌توان راجع به یکی‌بودن افراد موجود در تصویر اظهارنظر کرد؛ به این شکل که اگر فاصله‌ی دو بردار ویژگی کمتر از این مقدار آستانه باشد، این دو تصویر از صورت یک شخص گرفته شده‌اند و در غیر این صورت اشخاص موجود در دو تصویر متفاوت هستند.
برای بررسی بهتر نتیجه‌ی این روش، ۵۰۰۰ جفت تصویر در نظر گرفته‌ایم که ۲۵۰۰ تا از آن‌ها از یک شخص گرفته شده و ۲۵۰۰ تای دیگر تصاویر مربوط به اشخاص متفاوتی هستند. سپس این عکس‌ها را به الگوریتم Facenet داده و دو بردار ویژگی ۵۱۲ بعدی خروجی گرفته‌ایم. شکل ۳(آ) هیستوگرام فاصله‌ی این بردارها است.همان‌طور که در هیستوگرام می‌توان دید، جفت تصاویر با افراد مختلف به خوبی با استفاده از فاصله‌ی بردارهای ویژگی از جفت تصاویر مربوط به یک فرد جدا شده‌اند. در واقع، اگر مقدار آستانه را برابر با ۱/۱ قرار دهیم و طبق روشی که در پاراگراف قبل ذکر شد عمل کنیم، وضعیت ۹۹ درصد از جفت تصاویر را می‌توانیم به درستی تشخیص دهیم که دقت بالایی به شمار می‌رود. هم‌چنین در شکل ۳(ب) دقت برای جفت تصویرهای مربوط به اشخاص مختلف و مربوط به جفت تصویرهای مربوط به یک شخص را به ازای مقادیر مختلف آستانه نشان داده‌ایم.

به این ترتیب هر کاربری با توجه به نیاز و میزان حساسیتش روی اشتباه‌کردن در هر یک از این دو دسته می‌تواند مقدار آستانه‌ی مناسبی برای خود انتخاب کند. به بیان دیگر این روش در میزان حساسیت مدل نیز انعطاف خواهد داشت.



(ب)



(آ)

شکل ۳: نتایج مدل Facenet. (آ) هیستوگرام فواصل میان بردارهای خروجی از روش Facenet برای جفت عکس‌هایی از یک‌نفر (رنگ آبی) و جفت عکس‌هایی از افراد مختلف (رنگ نارنجی). (ب) دقت در جفت تصویرهای مربوط به اشخاص مختلف و در جفت تصویرهایی از یک شخص به ازای مقادیر آستانه‌ی مختلف.

۳. مدل تشخیص جهت صورت

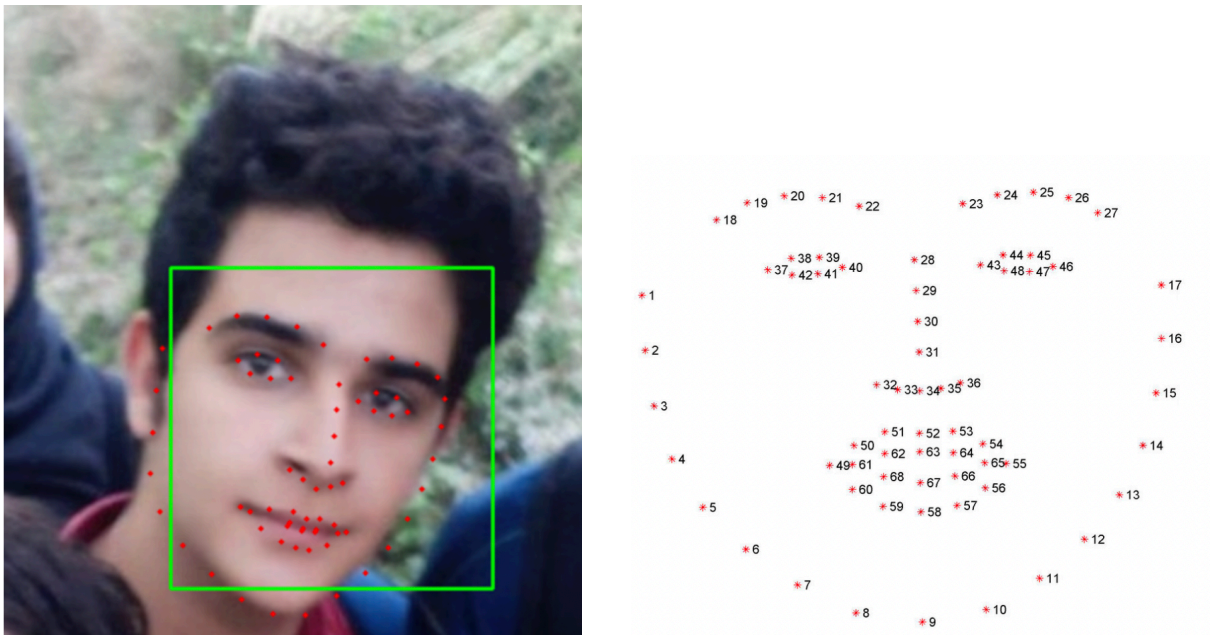
در این بخش مدل تشخیص جهت صورت را توضیح می‌دهیم. این مدل با ورودی‌گرفتن یک تصویر از چهره‌ی یک فرد، جهت صورت او در عکس را خروجی می‌دهد. هدف این است که یکی از پنج حالت روبه‌رو، بالا، پایین، چپ و راست را خروجی دهد. برای سادگی در این بخش فقط روش تشخیص از روبه‌رو بودن را توضیح می‌دهیم. برای چهار حالت دیگر نیز به طور مشابه می‌توان عمل کرد. برای تشخیص جهت صورت، از نقاط خاص صورت[1] استفاده می‌کنیم. نخست استخراج نقاط خاص صورت با استفاده از کتابخانه‌ی Dlib [۱] را معرفی می‌کنیم. سپس دو روش متفاوت را ارائه می‌کنیم که برای تشخیص از روبه‌رو بودن تصویر، از نقاط خاص صورت استفاده می‌کنند. در نهایت راهی ارائه می‌دهیم که دو روش ذکرشده را با هم ترکیب کرده و یک مدل نهایی قدرتمند به ما می‌دهد.

۱.۳. نقاط خاص صورت. با استفاده از کتابخانه‌ی Dlib [۱] می‌توان نقاط خاص صورت موجود در یک تصویر را به دست آورد. در واقع می‌توان یک تصویر شامل صورت یک انسان را به این کتابخانه داد و مختصات ۶۸ نقطه‌ی خاص از صورت در تصویر را خروجی گرفت. این ۶۸ نقطه در یک شمایل صورت در شکل ۴(آ) نشان داده شده‌اند. هم‌چنین یک تصویر واقعی شامل صورت انسان به این کتابخانه ورودی داده شده و در مختصات خروجی آن نقاط قرمز کشیده شده که نتیجه را می‌توانید در شکل ۴(ب) مشاهده کنید.

۲.۳. روش اول. در اولین روش، به نحوی به دنبال سنجش میزان تقارن در صورت هستیم. هرچه صورت موجود در عکس متقارن‌تر باشد، زاویه‌ی صورت شخص به روبه‌رو نزدیک‌تر است. به این منظور، ابتدا از نقاط روی بینی، وسط لب و وسط چانه بهترین خط ممکن را عبور می‌دهیم. سپس هر نقطه از سمت چپ این خط را نسبت به این خط قرینه کرده و فاصله‌ی قرینه‌شده‌ی آن نقطه با نقطه‌ی متناظرش در سمت راست خط را محاسبه می‌کنیم. مجموع تمام این فواصل، به ما معیاری از میزان متقارن‌بودن و در نتیجه از روبه‌رو بودن صورت می‌دهد. در واقع هرچه این حاصل‌جمع عدد کمتری باشد، صورت متقارن‌تر است. برای فهم بهتر این روش می‌توانید به شکل ۵ مراجعه کنید.

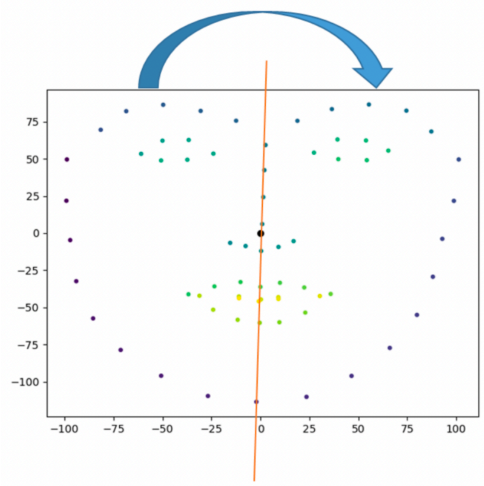
۳.۳. روش دوم. در این روش هدف تطبیق نقاط خاص به‌دست‌آمده از تصویر هدف با نقاط خاص به‌دست‌آمده از یک تصویر ذخیره‌شده است که می‌دانیم از روبه‌رو است. این تصویر را تصویر نمونه می‌نامیم. ابتدا یک تابع پردازش اولیه‌ی[2] ساده روی نقاط خاص صورت معرفی می‌کنیم که شامل یک انتقال، یک دوران و یک تجانس است. ابتدا نقاط خاص را به گونه‌ای

[1] facial landmarks
[2] pre processing function



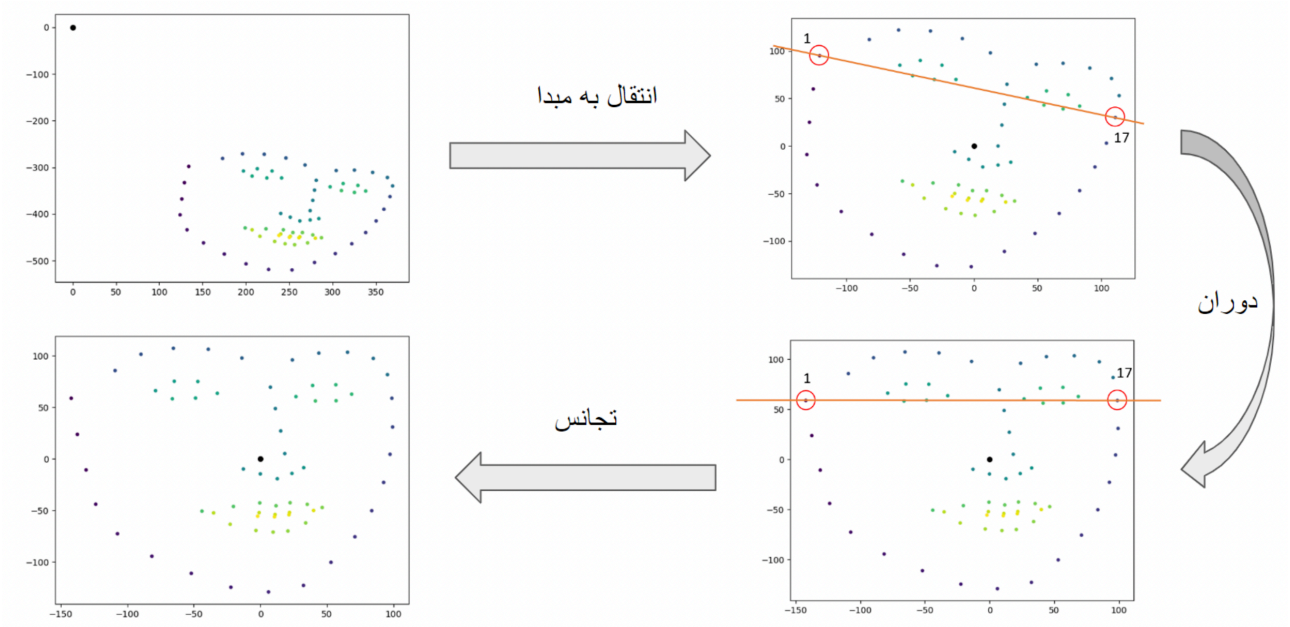
(آ) (ب)

شکل ۴: ۶۸ نقطه‌ی خاص صورت (آ) در یک شمایل و (ب) در یک صورت واقعی.



شکل ۵: ابتدا بهترین خط ممکن از نقاط روی بینی، وسط لب و وسط چانه را به دست آورده، نقاط سمت چپ خط را نسبت به خط قرینه کرده و فاصله‌ی نقاط حاصل از نقاط متناظرشان در سمت راست خط را محاسبه می‌کنیم و با هم جمع می‌زنیم.

انتقال می‌دهیم که میانگین آن‌ها در مبدأ قرار گیرد. سپس نقاط را به گونه‌ای دوران می‌دهیم که خط واصل نقاط ۱ و ۱۷ (به شکل ۴(آ) مراجعه کنید.) خطی افقی شود. به این ترتیب، صورت موجود در تصویر صاف می‌شود. در نهایت روی نقاط خاص به دست آمده یک تجانس را به گونه‌ای اعمال می‌کنیم که یک مربع به ضلع ۲۵۰ به آن‌ها محیط شود. مراحل این تابع پردازش اولیه را در شکل ۶ می‌توانید مشاهده کنید. حال نقاط خاص تصویر نمونه (که از روبه‌رو می‌باشد) و تصویر هدف (تصویر ورودی) را استخراج کرده و به تابع پردازش اولیه‌ای می‌دهیم که در بالا تعریف کردیم. فرض کنید نقاط خاص تصویر هدف بعد از اعمال تابع پردازش اولیه در ماتریس M و نقاط خاص تصویر نمونه نیز بعد از اعمال تابع پردازش اولیه در ماتریس N آمده باشد که



شکل ۶: مراحل سه‌گانه‌ی تابع پردازش اولیه.

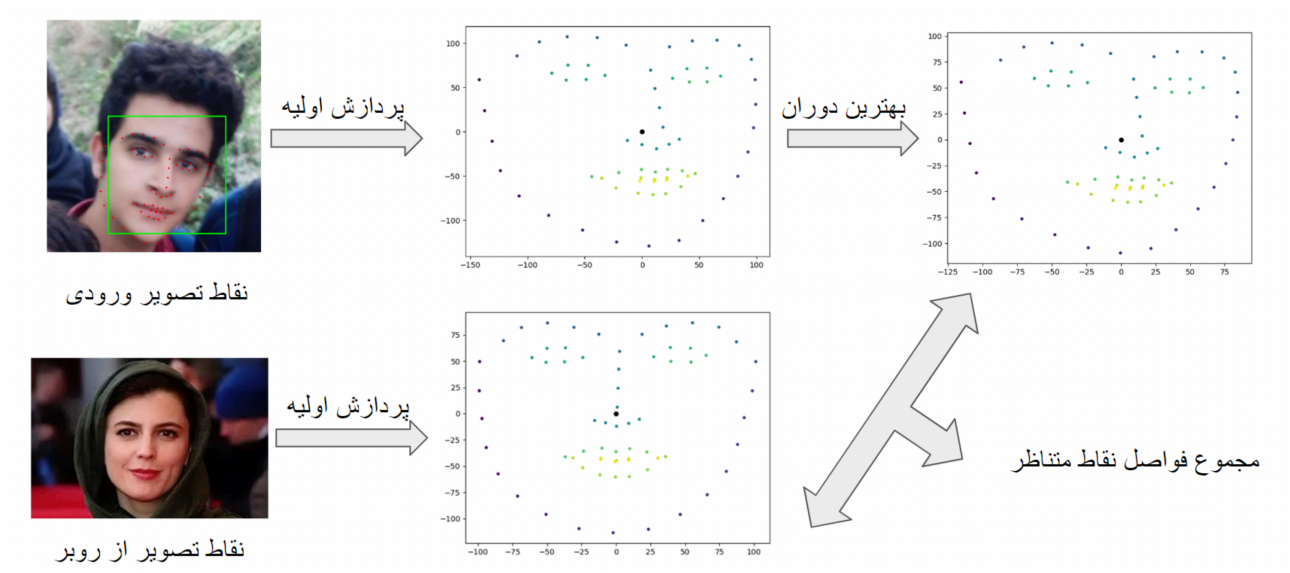
M, N به شکل زیر هستند:

$$M = \begin{bmatrix} x_۱ & x_۲ & \dots & x_{۶۸} \\ y_۱ & y_۲ & \dots & y_{۶۸} \end{bmatrix}, \quad N = \begin{bmatrix} x_۱^t & x_۲^t & \dots & x_{۶۸}^t \\ y_۱^t & y_۲^t & \dots & y_{۶۸}^t \end{bmatrix}$$

می‌خواهیم بهترین دورانی را پیدا کنیم که با اعمال آن روی نقاط M، آن‌ها را تا جای ممکن به نقاط N شبیه کند. اگر قرار دهیم $H = MN^T$ و $H = USV^T$ تجزیه‌ی طیفی H باشد، می‌توان دید بهترین دوران مذکور از رابطه‌ی $R = VU^T$ به دست می‌آید. این دوران را بر نقاط M اعمال می‌کنیم، سپس فاصله‌ی نقاط به‌دست‌آمده را از نقاط متناظرشان در N محاسبه می‌کنیم. حاصل‌جمع تمامی این فواصل، خروجی این روش است. تصویری کلی از این روش را می‌توانید در شکل ۷ ببینید.

در حقیقت این روش شباهت نقاط تصویر ورودی را با نقاط تصویری که می‌دانیم از روبه‌رو است می‌سنجد. هرچه خروجی این روش عددی بزرگ‌تر باشد یعنی این دو دسته از نقاط با هم متفاوت‌تر هستند و هرچه عدد کوچکتری باشد یعنی این دو دسته از نقاط به هم شبیه‌تر بوده و در نتیجه تصویر ورودی از روبه‌رو است.

۴.۳. ترکیب روش‌های اول و دوم و به‌دست‌آوردن مدل نهایی. هر یک از دو روشی که در بالا توضیح داده شدند ممکن است ضعف‌هایی داشته و در شرایطی کارکرد مطلوبی نداشته باشند. به همین دلیل در این بخش می‌خواهیم دو روش فوق را به گونه‌ای ترکیب کنیم تا مدلی نهایی به دست آوریم که از هر کدام از دو روش فوق قوی‌تر باشد. درواقع می‌خواهیم مقادیر مجهول a, b, c را به گونه‌ای بیابیم که اگر x خروجی روش اول و y خروجی روش دوم برای یک تصویر باشد، با محاسبه‌ی $ax + by$ و مقایسه‌ی آن با c بفهمیم آن تصویر از روبه‌رو گرفته شده یا خیر. به این منظور، ابتدا ۴۵ تصویر مختلف را انتخاب کرده و آن‌ها را بر حسب از روبه‌رو بودن مرتب کرده‌ایم. ۲۵ تصویر اول به عنوان تصویر از روبه‌رو و باقی تصاویر غیر روبه‌رو در نظر گرفته شده‌اند. سپس به آن‌ها وزن‌هایی مطابق شکل ۸(آ) داده شده که وزن هر تصویر اهمیت پیش‌بینی درست آن برای مدل ما را نشان می‌دهد. حال روش‌های اول و دوم را روی این ۴۵ تصویر اعمال کرده و عدد خروجی آن‌ها را ثبت می‌کنیم. در نتیجه می‌توان به هر تصویر به عنوان یک نقطه در فضای دوبعدی نگاه کرد که مؤلفه‌ی اول آن خروجی روش اول و مؤلفه‌ی دوم آن خروجی روش دوم می‌باشد. هم‌چنین اگر نقاط مربوط به ۲۵ تصویر اول را سبز به نشانه‌ی روبه‌رو بودن و باقی نقاط را قرمز به نشانه‌ی غیر روبه‌رو بودن در نظر بگیریم، پیدا کردن a, b, c که به آن اشاره شد معادل پیدا کردن بهترین خط $ax + by = c$ است که نقاط سبز را از قرمز جدا کند (شکل ۸(ب)). حل چنین سوالی یک مسأله در یادگیری ماشین کلاسیک می‌باشد که SVM



شکل ۷: تصویری کلی از روش دوم.

وزن‌دار با هسته‌ی خطی[1] نام دارد و به تفصیل مطالعه شده است. با حل این سوال، مقادیر $a = ۰/۰۰۲, b = ۰/۰۹۱, c = ۷/۳۳۶$ به دست می‌آید. در نتیجه مدل نهایی به این شکل حاصل می‌شود که نقاط خاص تصویر ورودی را به دست آورده و آن‌ها را به روش‌های اول و دوم می‌دهیم. خروجی روش اول را در a و خروجی روش دوم را در b ضرب کرده و حاصل را با هم جمع می‌کنیم. در صورتی که مقدار حاصل از c کم‌تر بود می‌گوییم تصویر حاصل از روبه‌رو بوده و در غیر این صورت از روبه‌رو بودن آن را رد می‌کنیم.

۵.۳. جمع‌بندی. در این بخش قطعه‌های مختلف سیستم که در قسمت قبل توضیح داده شدند را در کنار هم قرار می‌دهیم تا یک تصویر کلی از سیستم احراز هویت داشته باشیم.

سیستم کلی احراز هویت بر اساس چهره را می‌توان با الگوریتم زیر توضیح داد:

(۱) یک دنباله‌ی تصادفی از جهت‌های بالا، پایین، چپ، راست و روبه‌رو بساز که شامل حداقل یک جهت روبه‌رو باشد.

(۲) از کاربر بخواه از خود فیلمی بفرستد که در آن با صورتش دنباله‌ی فوق از جهت‌ها را انجام دهد.

(۳) با استفاده از مدل تشخیص جهت صورت، جهت صورت او را در طول فیلم محاسبه کن. در صورتی که با دنباله‌ی تولیدشده در مرحله‌ی ۱ مطابق نبود، اعلام خطا کن و به ۱ برو. در غیر این صورت به مرحله‌ی بعد برو.

(۴) فریمی از فیلم ارسالی کاربر که خروجی مدل تشخیص روبه‌رو بودن برای آن کم‌ترین مقدار ممکن است را جدا کن. این مقدار باید حتماً از مقدار c کم‌تر باشد و در نتیجه این عکس به عنوان عکسی که از روبه‌رو گرفته شده است شناسایی شود؛ چرا که در دنباله‌ی تصادفی حتماً یک جهت روبه‌رو قرار داشته و چون فیلم ارسالی با آن دنباله تطابق داشته حتماً فریمی با مقدار کمتر از c پیدا می‌شود و در نتیجه فریم با کمترین مقدار ممکن نیز از c کمتر است.

(۵) این فریم را با تصویری که از او داریم به روش Facenet بده. اگر طبق این روش، تصاویر با هم تطابق داشتند و مربوط به یک شخص بودند، او را احراز هویت کن. در غیر این صورت خطا نشان بده.

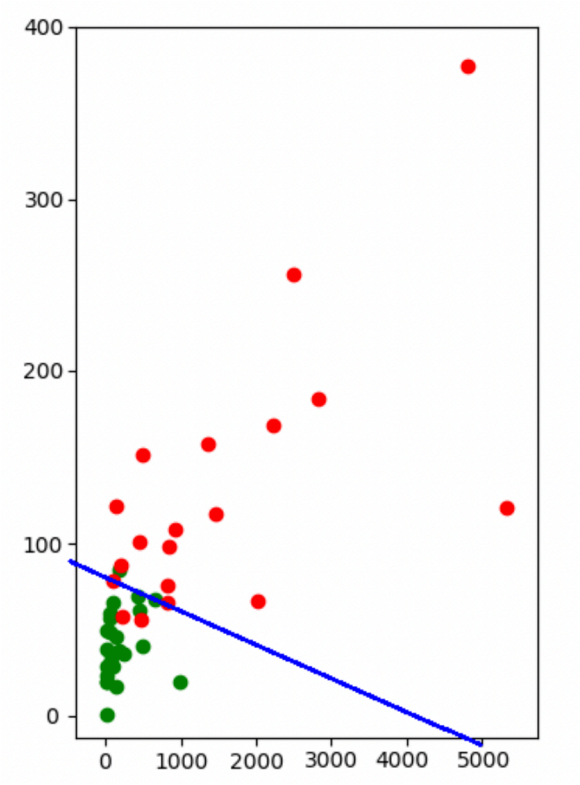
مراجع

[1] King, D. E. (2009). Dlib-ml: A machine learning toolkit. The Journal of Machine Learning Research, 10, 1755-1758.

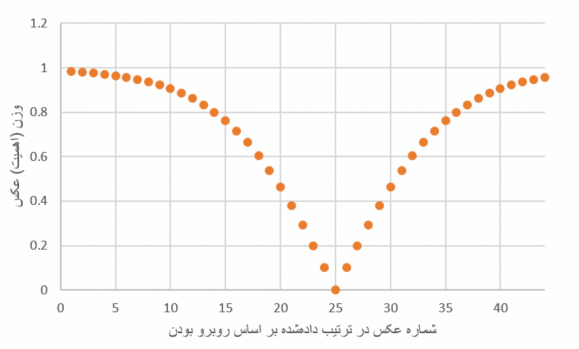
[2] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 815-823).

[3] Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. IEEE signal processing letters, 23(10), 1499-1503.

[1] weighted SVM with linear kernel



(ب)



(آ)

شکل ۸: ادغام روش‌های اول و دوم تشخیص روبه‌روبودن یک تصویر. (آ) وزن‌دهی به تصاویر مرتب‌شده بر اساس میزان روبه‌روبودنشان را نشان می‌دهد. (ب) به هر یک از ۴۵ عکس به چشم یک نقطه در فضا نگاه شده که مؤلفه‌ی اول آن خروجی روش اول برای آن عکس و مؤلفه‌ی دوم خروجی روش دوم است. هم‌چنین یک نقطه سبز است در صورتی که عکس متناظر با آن از روبه‌رو باشد و در غیر این صورت قرمز است. خط آبی بهترین خطی است که نقاط قرمز و سبز را از هم جدا می‌کند و با استفاده از روش SVM به دست آمده است.

* دانشجوی دکتری علوم‌کامپیوتر، اکول پلی‌تکنیک فدرال لوزان (EPFL)
رایانامه: ahmadrahimiuni@gmail.com