



# ارائه یک روش استخراج ویژگی از تصاویر چهره مبتنی بر اعمال تبدیل روی ویژگی‌های به دست آمده از شبکه‌های عصبی کانولوشن

مرتضی علی احمدی\* و روح‌الله دیانت

گروه مهندسی IT، دانشکده فنی و مهندسی، دانشگاه قم، قم، ایران

## چکیده

در این مقاله، یک روش استخراج ویژگی از داده ارائه شده است. ایده پیشنهادی، کلی بوده و قابل به‌کارگیری در استخراج ویژگی از هر نوع داده است. در این روش، بردار ویژگی آزمون، به ویژگی‌های موجود در همه دسته‌ها اضافه و سپس تبدیل مناسبی روی مجموعه ویژگی‌های هر دسته (با احتساب بردار آزمون اضافه‌شده)، اعمال می‌شود. نحوه اعمال تبدیل و مجموعه اقدامات بعد از آن، به نحوی صورت می‌گیرد که موجب می‌شود ویژگی‌های موجود در دسته‌ای که داده آزمون متعلق به آن است، دچار آسیب چندانی نشود و در مقابل، ویژگی‌های دسته‌هایی که داده آزمون متعلق به آنها نیست، دچار تخریب شوند. به‌طور شهودی می‌توان گفت، این امر، منجر به افزایش نرخ پذیرش به‌درستی (TP) در الگوریتم‌های دسته‌بندی یا شناسایی الگو می‌شود. به‌عنوان یک نمونه، ایده پیشنهادی، در مسأله شناسایی چهره با استفاده از شبکه‌های عصبی کانولوشن (CNN)، به‌عنوان یک پس‌پردازش و ویژگی‌های حاصل، به‌عنوان ویژگی‌های نهایی، برای عملیات شناسایی چهره به‌کار گرفته شد. نتایج پیاده‌سازی، نشان‌دهنده بهبود حدود ۳/۴٪ روی پایگاه داده LFW است.

واژگان کلیدی: استخراج ویژگی، شبکه‌های عصبی کانولوشن، تبدیل مویک، تبدیل فوریه

## Introducing a method for extracting features from facial images based on applying transformations to features obtained from convolutional neural networks

Morteza ali ahmadi\* & Rouhollah dianat

Department of IT Engineering, Faculty of Technical and Engineering,  
The University of Qom Qom, Iran

### Abstract

In pattern recognition, features are denoting some measurable characteristics of an observed phenomenon and feature extraction is the procedure of measuring these characteristics. A set of features can be expressed by a feature vector which is used as the input data of a system. An efficient feature extraction method can improve the performance of a machine learning system such as face recognition in the image field.

Most of the feature extraction methods in facial images are categorized as geometric feature extractor methods, linear transformation-based methods and neural network-based methods. Geometric features include some characteristics of the face such as the distance between the eyes, the height of the nose and the width of the mouth. In the second category, a linear transformation is applied to the original data and displaces them to a new space called feature space. In the third category, the last layer in the network, which is used for categorization, is removed, and the penultimate layer output is used as the extracted features. Convolutional Neural Networks (CNNs) are one of the most popular neural networks and are used in recognizing and verifying the face images, and also, extracting features.

\* Corresponding author

\*نویسندهٔ عهده‌دار مکاتبات

سال ۱۳۹۹ شماره ۳ پیاپی ۴۵

• تاریخ ارسال مقاله: ۱۳۹۷/۰۱/۲۱ • تاریخ پذیرش: ۱۳۹۸/۰۶/۱۱ • تاریخ انتشار: ۱۳۹۹/۰۹/۱۵ • نوع مطالعه: پژوهشی

فصلنامه



۱۴۱

The aim of this paper is to present a new feature extraction method. The idea behind the method can be applied to any feature extraction problem. In the proposed method, the test feature vector is accompanied with the training feature vectors in each class. Afterward, a proper transform is applied on feature vectors of each class (including the added test feature vector) and a specific part of the transformed data is considered. Selection of the transform type and the other processing, such as considering the specific part of the transformed data, is in such a way that the feature vectors in the actual class are encountered with less disturbing than the other ones. To meet this goal, two transformations, Fourier and Wavelet, have been used in the proposed method. In this regard, it is more appropriate to use transformations that concentrate the energy at low frequencies. The proposed idea, intuitively, can lead to improve the true positive (TP) rate. As a realization, we use the idea in CNN-based face recognition problems as a post-processing step and final features are used in identification. The experimental results show up to 3.4% improvement over LFW dataset.

**Keywords:** Feature extraction - Convolutional neural networks - Wavelet transform - Fourier transform.

بردارهای Fisher بوده است. این روش‌ها حتی بدون انجام ترازسازی و گام تعیین لندمارک‌ها، کارایی مناسبی دارند. در مقاله فعلی، رویکردی جهت استخراج ویژگی، ارائه شده است. ایده بیان‌شده، عمومی بوده که به‌طور بالقوه قابلیت به‌کارگیری را در هر مسأله دسته‌بندی و شناسایی الگو دارد. ایده کلی پیشنهادی آن است که داده آزمون به همه دسته‌های موجود در گالری (افراد)، اضافه شود. در ادامه، تبدیلی روی داده‌های هر دسته که اکنون شامل داده آزمون نیز است، اعمال می‌کنیم. نحوه اعمال تبدیل و فرایندهای انجام‌شده بعد از آن در نهایت منجر به آن می‌شود که اگر داده آزمون به دسته نامناسب یعنی به دسته‌ای که داده متعلق به آن نیست، اضافه شود، اثر مخرب ایفا کرده و بدین ترتیب، احتمال خطا (یعنی تشخیص نادرست دسته) را کمتر می‌کند.

در مقابل، افزودن داده آزمون به دسته مناسب، یعنی به دسته‌ای که داده در واقع به متعلق به آن است، اثر تخریبی کمتری دارد. بدین ترتیب، میزان خطا در شناسایی کاهش می‌یابد. شکل (۱) ایده رویکرد پیشنهادی را به شیوه ترسیمی نشان می‌دهد. برای سادگی بیان، ویژگی‌های دو بعدی، در نظر گرفته شده‌اند تا قابل ترسیم باشند. شکل (۱-الف)، وضعیت اولیه را نشان می‌دهد. اضافه کردن داده آزمون به همه دسته‌ها و اعمال تبدیل، منجر می‌شود که دسته D که داده آزمون به‌طور واقعی متعلق به آن است، به نمونه آزمون T نزدیک‌تر شود، زیرا با میزان تخریب کمتری مواجه شده است. در مقابل، بقیه دسته‌ها، از نمونه آزمون دورتر می‌شوند. این ویژگی در نهایت، منجر به افزایش دقت دسته‌بندی و تولید خطای کمتر، خواهد شد؛ لذا انتظار داریم استفاده از این روش، منجر به بهبود پارامتر TPIR<sup>4</sup> شود. در خصوص معیار FPIR<sup>5</sup>، تأثیر الگوریتم قابل پیش‌بینی

## ۱- مقدمه

شناسایی چهره<sup>۱</sup>، یکی از مهم‌ترین مباحث مطرح در حوزه تحلیل تصاویر چهره است. یک الگوریتم شناسایی چهره، یک تصویر چهره را به‌عنوان ورودی دریافت و تعلق یا عدم تعلق تصویر به افراد موجود را در یک گالری، تعیین می‌کند. مهم‌ترین قسمت یک الگوریتم شناسایی چهره، مرحله استخراج ویژگی است. ایده در نظر گرفته‌شده می‌بایست به‌صورتی باشد که منجر به تولید ویژگی‌های باشد که نسبت به موانعی چون تغییر وضعیت یا حالت چهره، سن، شرایط نوری مختلف و ...، مقاوم شود. الگوریتم‌های فراوانی برای استخراج ویژگی از تصویر چهره ارائه شده‌اند که در بخش ۲، مهم‌ترین آنها مرور خواهند شد. یک دسته از رویکردهایی که در سال‌های اخیر مورد توجه بوده و دقت‌های بالایی در استفاده از آنها در حوزه‌های مختلف شناسایی الگو، گزارش شده، روش‌های مبتنی بر شبکه‌های عصبی عمیق است. در حوزه شناسایی چهره، روش‌های مبتنی بر شبکه عصبی کانولوشن (CNN)<sup>۲</sup>، بسیار کارا بوده و با دقت‌های بالایی، عمل شناسایی را انجام می‌دهند. به‌عنوان نمونه می‌توان نتایج مسابقات ILSVRC<sup>۳</sup> را یاد کرد که در آن، دقت الگوریتم‌های دسته‌بندی مختلف روی مجموعه داده هزاردسته‌ای، آزمایش می‌شوند. در مسابقه ILSVRC برگزارشده در سال ۲۰۱۲ (ILSVRC 2012)، شبکه CNN پیشنهادشده توسط کریژفسکی و دیگران، به خطای دسته‌بندی Top-5 برابر ۱۵/۳٪ دست یافت که از الگوریتم رتبه دوم این مسابقه ۱۰/۹٪ بهتر بود [1]. گفتمنی است که الگوریتم رتبه دوم، سنتی بوده و مبتنی بر الگوریتم SIFT و

<sup>1</sup> Face recognition

<sup>2</sup> Convolutional neural network (CNN)

<sup>3</sup> ImageNet large scale visual recognition challenge (ILSVRC)

<sup>4</sup> True positive identification rate (TPIR)

<sup>5</sup> False positive identification rate (FPIR)

آن‌که در بخش پیاده‌سازی، الگوریتم پیشنهادی روی رویکرد مبتنی بر CNN اعمال شده، تعداد روش‌های بیشتری در این حوزه، مورد بررسی قرار گرفته است.

در یک نگاه کلی، اغلب روش‌های معروف استخراج ویژگی را از چهره می‌توان در یکی از سه دسته زیر قرار داد:

- روش‌های مبتنی بر ویژگی‌های هندسی<sup>۱</sup>.
- روش‌های مبتنی بر تبدیلات خطی<sup>۲</sup>.
- روش‌های مبتنی بر شبکه‌های عصبی

## ۱-۲- روش‌های مبتنی بر استخراج ویژگی‌های

### هندسی

در اغلب روش‌های اولیه استخراج ویژگی از تصاویر چهره، ویژگی‌های به‌کار گرفته‌شده، از نوع ویژگی‌های هندسی بوده‌اند. منظور از ویژگی‌های هندسی، مواردی نظیر فاصله چشم‌ها از یکدیگر، ارتفاع بینی، عرض دهان... است. در جدول (۱) برای نمونه، سه روش از این دسته، مرور شده است.

(جدول-۱): چند روش هندسی استخراج ویژگی تصاویر چهره  
(Table-1): Some geometrical face feature extraction methods

روش	توضیحات
برونلی و پوگیو [2]	ابتدا تصویر چهره را با روش یادشده در [3] تراز کرده و سپس، ویژگی‌های هندسی مانند طول و عرض بینی، موضع دهان و شکل چانه به‌دست آوردند.
قیمیر و همکاران [4]	ابتدا، نقاط اصلی چهره با EFGM <sup>۴</sup> (در [5]) به‌دست آمده‌اند. از آنجایی که حالات چهره، با متراکم‌شدن چهره همراه است، برای ردیابی مجموعه مترامکی از نقاط چهره، روش KLT <sup>۵</sup> (6) را به‌کار بردند. نتیجه این ردیابی، منجر به ایجاد ویژگی‌های هندسی (نقطه، خط و مثلث) شد. متمایزترین ویژگی‌ها، با به‌کارگیری AdaBoost <sup>۷</sup> (7) به‌همراه دسته‌بندی ELM <sup>۸</sup> (8) انتخاب شدند. در انتها، ویژگی‌های هندسی به نام FER، از خطوط تقویت شده و مثلث‌ها به‌دست آمدند.
کوآردا و دیگران [9]	مجموعه‌ای از ویژگی‌های هندسی با استفاده از SVM، الگوریتم‌های ژنتیک و روشی به نام mRmR <sup>۹</sup> پیشنهاد شد. SVM به‌کار گرفته شده، مبتنی بر تابع هسته خطی، تابع پایه شعاعی <sup>۸</sup> و پرسپترون چندلایه‌ای می‌باشد.

<sup>1</sup> Geometric features

<sup>2</sup> Linear transformations

<sup>3</sup> Align

<sup>4</sup> Elastic bunch graph matching (EBGM)

<sup>5</sup> Kanade-Lucas-Tomaci (KLT)

<sup>6</sup> Extreme learning machine (ELM)

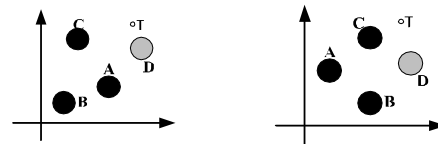
<sup>7</sup> Minimum redundancy maximum relevance (mRmR)

<sup>8</sup> Radial basis function

نیست؛ چون، تصویر آزمون، متعلق به هیچ یک از افراد نیست و اعمال الگوریتم، ویژگی‌های موجود در همه دسته‌ها را دچار تخریب می‌کند. البته در بخش پیاده‌سازی نشان دادیم که در این وضعیت هم، در مجموع، مقدار FPIR مناسبی به دست آمده است (تعریف و توضیح این دو پارامتر در بخش ۲-۴ بیان خواهد شد).

در این نوشتار، کاربرد ایده بیان‌شده در حوزه شناسایی چهره بیان شده و همچنین، فرض شده است که داده‌های ورودی به الگوریتم، ویژگی‌های به‌دست‌آمده به‌وسیله یک CNN باشد.

ساختار ارائه‌شده در ادامه به‌شرح زیر است: با توجه به آن‌که الگوریتم پیشنهادی، روی ویژگی‌های استخراجی مستخرج از تصاویر چهره از CNN، مورد ارزیابی قرار گرفته است، در بخش ۲، مرور کلی بر روش‌های استخراج ویژگی از تصاویر چهره و به‌خصوص، رویکردهای مبتنی بر CNN خواهیم داشت. در بخش بعد، روش پیشنهادی، ابتدا به‌صورت کلی بیان و سپس، برای استخراج ویژگی در حوزه تصاویر چهره، سفارشی شده است. بخش ۴ به بررسی و تحلیل نتایج حاصل از پیاده‌سازی رویکرد پیشنهادی در حوزه شناسایی چهره می‌پردازد. بخش پنجم نیز به نتیجه‌گیری اختصاص دارد.



(شکل-۱): بردار ویژگی مربوط به ۴ دسته فرضی A, B, C و D و همچنین بردار ویژگی مربوط به داده آزمون T قبل (الف) و بعد از اعمال روش پیشنهادی (ب) می‌توانند به‌صورت نشان‌شده در شکل باشند. این امر، منجر به افزایش دقت شناسایی می‌شود. T در واقع متعلق به دسته D است؛ لذا، بردار ویژگی مربوط به دسته D با رنگ دیگری نشان داده شده است.

(Figure-1): Feature vectors of 4 hypothetical class A, B, C and D accompanying test feature vector T (I) before and (II) after applying the proposed method can be as showed in figure. This leads to increased accuracy of identification. Since, T actually belongs to the class D, so the feature vector of the class D is shown in a different color.

## ۲- آشنایی با روش‌های استخراج ویژگی از تصاویر چهره

در این قسمت، روش‌های مختلف استخراج ویژگی از تصاویر چهره به‌صورت کلی مورد بررسی قرار می‌گیرند. با توجه به

## ۲-۲- روش‌های استخراج ویژگی مبتنی بر

### تبدیلات خطی

مبنای کلی این دسته، اعمال یک تبدیل خطی روی داده‌های اولیه بوده که در ادامه به فضای جدید که فضای ویژگی<sup>۱</sup> نام دارد، منتقل می‌شود. اگر داده اولیه، در ماتریس مشاهده<sup>۲</sup>  $X$  قرار داشته باشد، به ماتریس مشاهده جدید  $Y = AX$  در فضای ویژگی می‌رسیم که  $A$  ماتریس تبدیل<sup>۳</sup> نامیده می‌شود. تفاوت بین رویکردها، ناشی از تفاوت در نحوه محاسبه  $A$  است. ماتریس  $A$  شامل بردارهای پایه فضای جدید (فضای ویژگی) است. در جدول (۲) بعضی از مهم‌ترین روش‌های استخراج ویژگی مبتنی بر تبدیلات خطی در حوزه تصاویر چهره، آورده شده‌اند.

(جدول-۲): چند روش استخراج ویژگی مبتنی بر تبدیلات خطی

(Table-2): Some linear transform-based feature extraction methods

نام روش	توضیحات
EigenFace [10]	ماتریس $A$ طوری به دست آورده می‌شود که پراکندگی بین داده‌ها ماکزیمم شود. جهت کمی‌سازی معیار پراکندگی، از الگوریتم تحلیل مؤلفه اصلی (PCA) (PCA) استفاده شده است (در PCA، ماتریس تبدیل بر مبنای ماکزیمم‌سازی واریانس به دست می‌آید)
FisherFace [12]	در اینجا نیز، معیار محاسبه $A$ ماکزیمم ساختن پراکندگی است. با این تفاوت که به جای PCA، FLDA <sup>۵</sup> جایگزین شده است. در FLDA، پراکندگی به صورت نسبت پراکندگی بین دسته‌ای <sup>۶</sup> به پراکندگی درون دسته‌ای <sup>۷</sup> تعریف شده است.
نمایش تنک <sup>۸</sup> [13] و [14]	اگر داده اولیه در $X$ و تصویر آزمون، در بردار $y$ قرار داشته باشد، به دنبال تبدیلی مانند $y = Ax$ هستیم. $A$ برداری است که نقش هر عنصر $X$ (هر تصویر آموزشی) را در ساخت $y$ مشخص می‌کند. اگر تصویر آزمون، متعلق به فرد درون گالری باشد، ضرایب مربوط به آن فرد، سیار بزرگتر از ضرایب مربوط به تصاویر سایر افراد گالری می‌باشد.

## ۲-۳- روش‌های مبتنی بر شبکه‌های عصبی

بیش‌تر رویکردهای امروزی استخراج ویژگی از تصاویر چهره، از شبکه‌های عصبی و به‌خصوص، شبکه‌های عصبی عمیق، استفاده می‌کنند. در جدول (۳) چند شبکه عصبی عمیق

<sup>1</sup> Feature space

<sup>2</sup> Observation matrix

<sup>3</sup> Transformation matrix

<sup>4</sup> Principal component analysis (PCA)

<sup>5</sup> Fisher linear discriminant analysis (FLDA)

<sup>6</sup> Between-class scatter

<sup>7</sup> Within-class scatter

<sup>8</sup> Sparse representation

بیان‌شده در حوزه استخراج ویژگی از تصاویر چهره، آمده است.

یکی از کاربردهای مهم شبکه‌های عصبی، استخراج ویژگی از داده ورودی است. بدین منظور، لایه آخر موجود در شبکه را که برای دسته‌بندی به کار می‌رود، حذف کرده و خروجی لایه ماقبل آخر را به‌عنوان ویژگی‌های استخراج شده به کار می‌برند. در اینجا، شبکه عصبی نقش یک استخراج‌کننده ویژگی را دارد. در حوزه‌های شناسایی الگو مانند شناسایی چهره، به‌طور معمول شبکه عصبی به کار می‌رود.

(جدول-۳): روش‌های استخراج ویژگی مبتنی بر شبکه‌های

### عصبی عمیق

(Table-3): Deep neural network-based feature extraction methods

نام روش	توضیحات
DeepID [15]	در این پژوهش، شبکه‌های عصبی کانولوشن (CNN) چندگانه‌ای پیشنهاد شده است که هر یک از شبکه‌ها وظیفه استخراج ویژگی‌هایی با نام DeepID از قسمت خاصی از چهره را داشتند و دسته‌بندی، براساس کنار هم قراردادن تمام DeepIDها صورت گرفته است. همین نویسندگان در سال‌های بعد شبکه‌های DeepID2 [16]، DeepID3 [17] و DeepID3 [18] را ارائه کرده‌اند که در آنها بهبودهایی چون استفاده از شبکه‌های عمیق‌تر، داده‌های آموزشی بیشتر و در نظر گرفتن پس‌پردازش‌هایی چون کاستن از اختلاف بین ویژگی‌های تصاویر یک فرد و افزودن اختلاف بین ویژگی‌های افراد مختلف، داده شده است.
DeepFace [19]	در DeepFace، نخست برای تراز کردن چهره از مدل‌سازی سه‌بعدی مبتنی بر تبدیل افاین <sup>۹</sup> استفاده کردند و ثانياً برای استخراج ویژگی و دسته‌بندی، CNN ۹ لایه‌ای به کار بردند که برخی لایه‌های آن، به جای لایه‌های کانولوشن استاندارد، لایه‌هایی با اتصال محلی <sup>۱۱</sup> بدون اشتراک وزن <sup>۱۲</sup> هستند.
VGGFace [20]	در این مقاله CNN بسیار عمیق (۳۷ لایه) جهت استخراج ویژگی‌های مقاوم از تصاویر چهره، ارائه شده است. سه معماری A، B و D براساس [21] در نظر گرفته می‌شود. معماری A شامل ۱۱ بلوک بوده که هر یک شامل یک عملگر خطی (لایه کانولوشن و یا لایه به‌طور کامل متصل <sup>۱۳</sup> ) و یک یا چند عملگر غیرخطی (لایه ReLU و لایه Max pooling) است. معماری B و D مشابه با A بوده با این تفاوت که به ترتیب دارای دو و پنج لایه کانولوشن اضافی هستند. آخرین لایه به‌طور کامل متصل در بلوک یازدهم، برای آموزش یک دسته‌بند به کار می‌رود.

<sup>9</sup> Convolutional neural network (CNN).

<sup>10</sup> Affine transformation

<sup>11</sup> Locally connected layers

<sup>12</sup> Weight sharing

<sup>13</sup> Fully connected

### ۳- توضیح روش پیشنهادی

در این قسمت، در ابتدا، ایده رویکرد پیشنهادی توضیح داده می‌شود. در ادامه، الگوریتم کلی کار بیان و توضیحاتی درخصوص بعضی از مراحل الگوریتم، بیان می‌شود. در انتهای این بخش، الگوریتم کلی بیان شده در کاربرد شناسایی چهره، سفارشی می‌شود.

#### ۳-۱- توضیح ایده روش پیشنهادی

در رویکردهای متداول دسته‌بندی، دو مجموعه مجزای "داده‌های آموزشی<sup>۱</sup>" و "داده‌های آزمون<sup>۲</sup>" در نظر گرفته می‌شود. با استفاده از داده‌های آموزشی، یک دسته‌بند ساخته و سپس، از دسته‌بند برای دسته‌بندی داده‌های آزمون استفاده می‌شود. در چنین ساختاری، طبیعتاً نباید هیچ یک از داده‌های آزمون در مجموعه داده‌های آموزشی وجود داشته باشد.

در این مقاله، به مسأله بیان شده در بالا، از جنبه دیگری نگریسته شده است. فرض کنید یک دسته‌بند با استفاده از تعدادی داده آموزشی، ساخته شده باشد و همچنین، یک داده به‌عنوان داده آزمون در اختیار داشته باشیم.

ایده‌ای که مبنای روش پیشنهادی است آن است که به‌طور شهودی می‌توان گفت اگر داده آزمون به دسته واقعی یعنی دسته‌ای که داده آزمون متعلق به آن است، اضافه شود، اتفاق خاصی نمی‌افتد؛ اما اگر این داده به دسته‌های دیگر تخصیص یابد، پارامترهای آماری آن دسته‌ها دچار آسیب و خدشه می‌شود. نکته دیگر آن است که در چنین ساختاری، چون اثر داده آزمون به همه دسته‌ها اضافه می‌شود، مشکل بیان شده در بالا، درخصوص عدم اشتراک داده‌های آزمون و آموزشی پیش نمی‌آید.

به‌منظور عملی‌تر کردن این ایده، مراحل کلی زیر پیشنهاد می‌شود:

الف- داده آزمون را به همه دسته‌ها اضافه کن و سپس، تبدیلی روی آن‌ها انجام بده و بخشی از آن را که در کاربرد فعلی مفید است، انتخاب کن. منظور از مفید بودن، آن است که اگر این بخش یادشده به داده‌های دسته‌ها اضافه شود، منجر به عدم تخریب اطلاعات دسته‌ای می‌شود که تصویر آزمون به آن متعلق است و در مقابل، موجب تخریب اطلاعات

<sup>1</sup> Training data

<sup>2</sup> Test data

سایر دسته‌ها می‌شود (جزئیات بیشتر در بخش توضیح الگوریتم پیشنهادی (۳-۲-۲)). بدین ترتیب یک مجموعه آموزشی جدید به‌وجود می‌آید.  
ب- با توجه به تغییر مجموعه آموزشی، دسته‌بند را دوباره با داده آموزشی جدید، آموزش بده.  
ت- اکنون داده آزمون را به دسته‌بند جدید بده و عمل دسته‌بندی را روی آن انجام بده.  
نکته: در صورتی که برای دسته‌بندی، از معیارهای فاصله یا انطباق استفاده شود، مراحل ب و ت به‌صورت زیر انجام می‌شود:  
ب- فاصله داده آزمون را با داده نماینده هر دسته به دست بیاور.  
ت- داده آزمون را به دسته‌ای منسوب کن که کمترین فاصله را با آن دسته داشته باشد.  
در بخش‌های آینده، نشان خواهیم داد که الگوریتم بیان شده توانسته این ایده را محقق سازد.

#### ۳-۲- بیان الگوریتم ارائه شده و توضیحاتی درخصوص بعضی مراحل آن

در این قسمت، متن الگوریتم پیشنهادی، بیان می‌شود و سپس، توضیحاتی درخصوص بعضی مراحل آن داده می‌شود.

##### ۳-۲-۱- بیان متن الگوریتم

ورودی: الگوریتم دارای دو دسته ورودی به شرح زیر است.  
• مجموعه داده‌های آموزشی که هر یک، متعلق به یکی از  $C$  دسته هستند. فرض می‌کنیم دسته نام، دارای  $n_i$  نمونه باشد. هر نمونه،  $d$  بعدی در نظر گرفته می‌شود. بدین ترتیب، به‌ازای هر دسته نام،  $n_i$  نمونه  $d$  بعدی خواهیم داشت.  
• داده آزمون که طبیعتاً آن هم  $d$  بعدی است و آن را با  $\Gamma$  نمایش می‌دهیم.  
الف- داده آزمون را به هر یک از دسته‌ها اضافه کن. اکنون دسته نام،  $n_i + 1$  نمونه  $d$  بعدی خواهد داشت. به بیان دیگر می‌توان گفت به‌ازای هر دسته نام، یک ماتریس از داده‌ها به ابعاد  $d \times (n_i + 1)$  خواهیم داشت.  
ب- اکنون به‌ازای هر دسته نام کارهای زیر را انجام بده

ب-1- روی هر بعد (هر ستون) ماتریس داده، به صورت مجزا، یک تبدیل مناسب، مانند فوریه، DCT<sup>1</sup> یا موجک اعمال کن. منظور از مناسب بودن، استفاده از تبدیلاتی است که تمرکز انرژی در فرکانس‌های پایین داشته باشند و هرچه یک تبدیل، تمرکز انرژی بیشتری در مؤلفه‌های پایین داشته باشد، در مجموع می‌تواند نتیجه بهتری را بدهد. در هر حال، ورودی تبدیل، ماتریس داده با ابعاد  $(n_i+1) \times d$  و خروجی تبدیل، یک ماتریس داده با ابعاد  $(n_i+1) \times d$  است<sup>2</sup>؛ در این رابطه،  $n_i$ ، تعداد نمونه‌های خروجی تبدیل است.

ب-2- در هر ستون، مقادیر ماتریس داده جدید (تبدیل یافته)، را به ترتیب نزولی مرتب کن.

ب-3- مقدار آخرین مؤلفه را در هر ستون در نظر بگیر. به این ترتیب، یک بردار  $d$  عنصری از این مقادیر تشکیل می‌شود.

ب-4- واضح است که بردار  $d$  بعدی به دست آمده در قسمت ب-3 در حوزه‌های متفاوت با حوزه داده‌های آموزشی (حوزه مکان) قرار دارد. از طرف دیگر، نحوه به دست آمدن داده‌های این بردار، به شکل انتخاب مقدار (کمینه) از تبدیلات اعمال شده روی ستون‌ها است، لذا نمی‌توان گفت که این داده‌ها، حتی در فضای تبدیل یافته (برای مثال فوریه) قرار دارند. به هر حال، واضح است که می‌بایست در این مرحله، تبدیلی اعمال شود که فضای بردار به دست آمده را به فضای مکانی<sup>3</sup> نزدیک گرداند. با توجه به آن که در فرایند محاسبه این بردار، سیگماهایی متشکل از ضرایب نمایی مختلف و همچنین عمل‌گر غیر خطی کمینه‌سازی مورد استفاده قرار گرفته است، در عمل، به دست آوردن چنین تبدیلی بسیار دشوار و حتی غیر ممکن به نظر می‌رسد. در پیاده‌سازی‌های انجام شده، در این مرحله، یک تبدیل معکوس مورد استفاده قرار گرفته است. به عنوان مثال، اگر تبدیل انجام گرفته در مرحله ب-1، تبدیل فوریه باشد، در اینجا تبدیل معکوس فوریه روی بردار  $d$  بعدی، انجام گرفته است.

<sup>1</sup> Discrete cosine transform (DCT)

<sup>2</sup> به عنوان مثال اگر ماتریس داده با ابعاد  $5 \times 1000$  باشد، با اعمال 1000 تبدیل فوریه سه نقطه‌ای که هر کدام روی یک ستون ماتریس اعمال می‌شود، به ماتریس داده خروجی  $3 \times 1000$  می‌رسیم.

<sup>3</sup> Spatial

نتایج پیاده‌سازی‌های انجام شده، رضایت‌بخش بوده است. به نظر می‌رسد در اینجا، گرفتن تبدیل معکوس، هر چند راه‌کار ایده‌آل و اصلی نیست، اما تا اندازه‌ای فضای این داده‌ها را به فضای مکانی، نزدیک می‌کند. برای مثال، در تبدیل فوریه چون محاسبه معکوس بر مبنای توان‌های نمایی با علامت مثبت، انجام می‌پذیرد، بعضی از توان‌های نمایی با علامت منفی (که در تبدیل فوریه به کار برده می‌شوند) را خنثی می‌کند.

ب-5- اکنون، داده‌ای را که در قسمت الف به دسته اضافه شده است، حذف کن و داده به دست آمده در قسمت ب-4 جایگزین آن کن. اکنون، دسته همچنان دارای  $(n_i+1)$  نمونه  $d$  عنصری خواهد بود.

ب-6- میانگین نمونه‌های دسته را محاسبه کن و آن را به عنوان نماینده این دسته در نظر بگیر. این میانگین را با  $r_i$  نمایش بده.

مراحل الگوریتم پیشنهادی در شکل (۲) به صورت شبه کد نیز نشان داده شده است.

(شکل-۲): شبه کد مربوط به الگوریتم پیشنهادی

(Figure-2): Pseudo-code of the proposed algorithm

Algorithm 1:

Input:

$tr$ : Training data ( $tr_j$ :  $j$ th training data for class  $i$  ( $d$ -dimensional vector))

$td$ : Test data,

$C$ : Number of classes

$n_i$ : Number of samples in Class  $i$

$Trans$ : A suitable transform (Fourier, Wavelet)

$InvTrans$ : Inverse transform

a. for  $i=1, 2, \dots, C$ :

$$a.1: O_{(n_i \times d)}^1 \leftarrow [tr_{i1}; tr_{i2}; \dots; tr_{in_i}]$$

$$a.2: O_{(n_i+1) \times d}^1 \leftarrow [tr_{i1}; tr_{i2}; \dots; tr_{in_i}; td]$$

$$a.3: T_i(:, j) \leftarrow Trans\{O_i^1(:, j)\} \text{ (for } j=1, 2, \dots, d)$$

$$a.4: S_i(:, j) \leftarrow Sort(T_i(:, j), 'descend')$$

$$a.5: D_i \leftarrow [S_i(1, end) \ S_i(2, end) \ \dots \ S_i(d, end)]$$

$$a.6: I_i = InvTrans\{D_i\}$$

$$a.7: O_{i \times d}'' \leftarrow DeleteRow(O_{(n_i+1) \times d}^1, td)$$

$$a.8: O_{(n_i+1) \times d}'' \leftarrow AppendRow(O_{i \times d}'', I_i)$$

end

Output:  $O_{(n_i+1) \times d}''$

در بخش نتایج پیاده‌سازی، کارایی این الگوریتم در کاربرد شناسایی چهره نشان داده است. در آن بخش، با انجام

آزمایش‌های متعدد نشان داده‌ایم که به‌کارگیری ویژگی‌های به‌دست‌آمده از رویکرد پیشنهادی، منجر به دستیابی به کارایی بهتر شده است.

### ۲-۲-۳- توضیحاتی در خصوص مراحل الگوریتم

با توجه به وضوح الگوریتم، توضیح خاصی در خصوص مراحل اجرای آن، لازم به نظر نمی‌رسد؛ تنها به چند نکته، اشاره می‌شود:

- در مرحله ب-۱- تبدیلی نظیر فوریه، DCT یا موجک روی داده، اعمال شده است. نوع تبدیل و اندازه خروجی آن، می‌تواند موارد متعددی باشد. در بخش پیاده‌سازی، تبدیلات فوریه و موجک<sup>۱</sup> مورد ارزیابی قرار گرفته است.
- ممکن است، این سؤال مطرح شود که می‌توان هیچ تبدیلی اعمال نکرد و تنها خود داده آزمون را به هر دسته اضافه کرد و سپس عمل محاسبه نماینده و فاصله را انجام داد؛ اما آزمایش‌های انجام‌شده نشان داد که این کار، نتایج مناسبی به‌دنبال نخواهد داشت و لذا، از پرداختن به این موضوع، صرف نظر شد.
- همان‌گونه که در بخش توضیح ایده روش پیشنهادی (۳-۱) بیان شد، یکی از مراحل مهم ایده پیشنهادی، اعمال تبدیل و در نظر گرفتن بخش مفید تبدیل است. در این قسمت، منظور از مفیدبودن را توضیح می‌دهیم. فرض کنید داده‌های مربوط به یک دسته، به‌طور کامل مانند یکدیگر و بر هم منطبق باشند. همچنین، فرض کنید داده آزمون نیز، در واقع، مربوط به همین دسته و باز، مساوی داده‌های دسته باشد. در این صورت، اگر در هر ستون، تبدیل فوریه گرفته شود فقط مؤلفه نخست تبدیل، غیر صفر بوده و سایر مؤلفه‌ها برابر صفر هستند. در نتیجه، بردار  $d$  عنصری تشکیل‌شده در مرحله ب-۳ الگوریتم، بردار تمام صفر است؛ اما اگر، داده آزمون، جزو آن دسته نباشد، مؤلفه آخر (کوچک‌ترین مؤلفه)، بردار غیر صفر می‌شود و در واقع، اثر مخرب دارد. گفتنی است که در حالت تعلق داده آزمون به دسته صحیح (تمام صفر بودن بردار)، میانگین تحت الشعاع قرار می‌گیرد که به‌علت اضافه‌شدن یک نمونه به دسته و در نتیجه، تقسیم بر  $n+1$  به جای تقسیم بر  $n$  است؛ اما تأثیر این امر در دسته واقعی به‌مراتب، کمتر از دسته‌های دیگر است؛

<sup>۱</sup> Wavelet

زیرا در آنها با اعداد غیر صفر مواجه هستیم. اگر از فرض ایده‌آل تمام صفر بودن بردار هم، کوتاه بیاییم، با توجه به اینکه داده‌های دسته با هم نزدیکی و شباهت زیادی دارند، اضافه‌کردن یک داده غیر متعلق به دسته، اثر تخریبی خود را بر جای می‌گذارد.

### ۳-۳- استفاده از الگوریتم ارائه‌شده در شناسایی چهره

در الگوریتم یادشده در قسمت قبل، هیچ‌گونه قید و شرطی در خصوص نحوه استخراج ویژگی و یا فرایند دسته‌بندی، گذاشته نشده است. لذا، ایده، کلی بوده و روی هر روش دسته‌بندی یا شناسایی الگو، قابل اعمال است. در این قسمت، نسخه سفارشی‌شده این الگوریتم را در کاربرد شناسایی چهره، بیان می‌کنیم.

همان‌گونه که در بخش ۲ بیان شد، رویکردهای مبتنی بر CNN، در حوزه‌های مربوط به شناسایی الگو، دقت‌های بسیار مطلوب دارند. همچنین، در همان بخش بیان شد که در حوزه شناسایی چهره، به‌طور معمول از این شبکه‌ها به‌عنوان استخراج‌کننده ویژگی استفاده، سپس، ویژگی‌های به‌دست‌آمده، برای شناسایی به یک دسته‌بند داده می‌شود. در نتیجه، ورودی دسته‌بند، ویژگی‌های به‌دست‌آمده از بلاک CNN است.

الگوریتم یادشده در بخش ۲-۳ با توجه به نکات زیر می‌تواند در نقش یک الگوریتم شناسایی چهره ظاهر شود:

- در مسأله شناسایی چهره، تعدادی تصویر از هر فرد، در اختیار داریم. این مجموعه تصاویر، در اصطلاح گالری نامیده می‌شود.
- در یک الگوریتم شناسایی چهره، افراد موجود در گالری، نقش دسته‌ها را ایفا می‌کنند. اگر تعداد افراد  $n$  نفر باشند، شناسایی چهره، یک مسأله دسته‌بندی  $n$  دسته‌ای است.
- فرض کنید از فرد مشخص در گالری،  $k$  تصویر در اختیار باشد. هر یک از این تصاویر به‌طور مجزا به بلاک CNN داده شده و بردار ویژگی مربوط به آن، به‌دست می‌آید. بدین ترتیب اگر از فرد  $i$ ام،  $n_i$  تصویر در اختیار باشد، به‌ازای این فرد،  $n_i$  بردار ویژگی به‌دست می‌آید. بردارهای ویژگی متعلق به تصاویر افراد داخل گالری، مجموعه داده آموزشی تشکیل می‌دهند.

• یک الگوریتم شناسایی چهره می‌بایست در مواجهه با یک تصویر مجهول، تعیین کند که این تصویر متعلق به چه فردی در گالری است. در رویکردهای مبتنی بر CNN، تصویر مجهول نیز به بلاک CNN داده شده و بردار ویژگی مربوط به آن محاسبه می‌شود. این بردار ویژگی، همان داده آزمون الگوریتم ارائه‌شده ما، است.

با توجه به ملاحظات بالا، اکنون می‌توان از این الگوریتم، به‌عنوان یک پردازش‌گر ویژگی‌های به‌دست‌آمده به‌وسیلهٔ CNN، جهت شناسایی چهره استفاده کرد.

#### ۴- نتایج پیاده‌سازی

در این بخش، کارایی رویکرد پیشنهادی برای استخراج ویژگی را در حوزه شناسایی چهره<sup>۱</sup> مورد بررسی و تجزیه و تحلیل قرار می‌دهیم. با انجام آزمایش‌ها متعدد نشان خواهیم داد، ویژگی‌های به‌دست‌آمده توسط الگوریتم یادشده، منجر به دستیابی به نتایج بهتر در حوزه شناسایی چهره شده‌اند. یک الگوریتم شناسایی چهره، دارای دو ورودی است:

- ✓ یک گالری، شامل تصاویر تعدادی از افراد مشخص.
- ✓ یک تصویر آزمون، با نام فرضی  $\Gamma$ .

هدف الگوریتم آن است که مشخص کند نخست آیا این تصویر متعلق به فردی داخل گالری است یا خیر. در صورت تعلق، الگوریتم باید فرد را نیز مشخص کند. مراحل کلی یک الگوریتم شناسایی چهره به‌صورت زیر است:

- ✓ به‌ازای هر تصویر موجود در گالری، با استفاده از یک روش استخراج ویژگی، یک بردار ویژگی به‌دست بیاور.
- ✓ میانگین بردار ویژگی‌های مربوط به تصاویر هر فرد، را محاسبه کن. فرض کنید به‌ازای فرد شماره  $i$  میانگین، برابر  $m_i$  باشد.
- ✓ فاصله داده آزمون با هر کدام از  $m_i$  ها را به‌دست بیاور.

$$D_i = \text{dist}(\Gamma, m_i) \quad (1)$$

✓ اگر  $D_i$  بیشتر از یک مقدار آستانه  $\tau$  باشد، تصویر آزمون را متعلق به هیچ فردی ندان. در غیر این صورت، تصویر را به دسته‌ای (فردی) که کمترین  $D_i$  را موجب شده منسوب کن.

$$k = \arg, \min D_i \quad (2)$$

<sup>1</sup> Face recognition

خروجی:  $k$  (شماره دسته‌ای (فردی) که منجر به ایجاد فاصله مینیمم شده است).

در ادامه این بخش، نخست، توصیف شرایط آزمایش (شامل مجموعه داده، سخت افزار و نرم افزار به‌کار رفته) انجام می‌شود. در ادامه، معیارهای ارزیابی کارایی یک الگوریتم شناسایی چهره بیان شده و در انتها، با بیان نتایج آزمایش‌های متعدد، نشان می‌دهیم به‌کاربردن رویکرد پیشنهادی به عنوان استخراج ویژگی، منجر به بهبود مناسبی در کارایی شناسایی چهره خواهد شد.

#### ۴-۱- توصیف مجموعه داده، سخت‌افزار و

##### نرم‌افزار

در جدول (۴)، ۴ مجموعه دادهٔ مورد استفاده، توضیح داده شده‌اند. همچنین، در شکل‌های (۳ و ۴) تصاویر مربوط به بعضی از آنها، نشان داده شده است.

در جدول (۵) نیز، مشخصات سخت‌افزار و همچنین ابزارهای نرم‌افزاری مورد استفاده در پیاده‌سازی‌ها، آورده شده است.

(جدول-۴): مجموعه داده‌های مورد استفاده  
(Table-4): Datasets used in the paper

نام	توضیحات
AT&T	شامل تصاویر چهل نفر و از هر نفر ده تصویر است. تصاویر در وضعیت‌های نوری و حالت چهره مختلف گرفته شده‌اند. ابعاد تصاویر، $92 \times 112$ می‌باشد [22].
MUCT	شامل ۳۷۵۵ تصویر از ۲۷۶ نفر با ابعاد $480 \times 640$ است که در وضعیت‌های مختلف چهره و شرایط نوری مختلف، گرفته شده‌اند [23].
LFW	تصاویر از وب گرفته شده‌اند و لذا یک مجموعه داده کنترل‌نشده و چالشی است. ابعاد تصاویر، $250 \times 250$ است. برای ۱۶۸۰ نفر دو و یا بیشتر تصویر وجود دارد. تعداد کل این تصاویر ۹۱۶۴ است. همچنین ۴۰۶۹ نفر، فقط دارای یک تصویر هستند. برای آنکه محاسبه TPIR ساده‌تر انجام شود، فقط تصاویر مربوط به ۱۶۸۰ نفر در پیاده‌سازی مورد استفاده شده است [24].
100LFW	شامل پانصد تصویر از صد نفر (هر نفر پنج تصویر) است. تصاویر، مربوط به افراد مجموعه LFW است. در مواردی که تعداد تصاویر فرد در LFW کمتر از پنج بود، با جستجو در وب، تصاویر مربوط به آن فرد، تا رسیدن به پنج تصویر، تکمیل شده است.

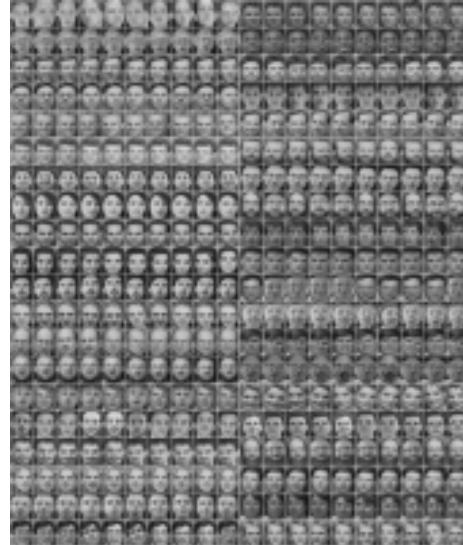
(جدول-۵): سخت‌افزار و نرم‌افزار مورد استفاده

(Table-5): Hardware and software used in the paper

نرم افزار	سخت افزار
MATLAB 2017- MatConvNet ([25])	CPU: Corci7 (7700k)- RAM: 16GB- Hard: (SSD-512GB)-Cache: 8MB- GPU: 1070 (8 GB)

(شکل-۳): مجموعه تصاویر پایگاه داده AT&T ([26])

(Figure-3) Collection of AT&T database images



(شکل-۴): بعضی تصاویر مجموعه داده LFW (جمع‌آوری شده

توسط نویسندگان)

(Figure-4): Some of the LFW database images (collected by authors)



## ۴-۲- معیارهای ارزیابی کارایی

برای سنجش کارایی یک الگوریتم شناسایی چهره، دو معیار "نرخ تشخیص پذیرش به‌اشتباه" (FPIR<sup>۱</sup>) و "نرخ تشخیص پذیرش به‌درستی" (TPIR<sup>۲</sup>) " در نظر گرفته می‌شوند [27]. معیار FPTR به‌صورت زیر تعریف می‌شود:

$$FPTR(\tau) = \frac{F(\tau)}{T} \quad (۳)$$

<sup>1</sup> False positive identification rate

<sup>2</sup> True positive identification rate

در این رابطه،  $T$ ، تعداد کل تصاویر غیر متعلق به گالری است که برای آزمایش به الگوریتم داده شده و  $F$ ، تعدادی از این تصاویر است که الگوریتم، آنها را به‌اشتباه، متعلق به گالری دانسته است. پارامتر  $\tau$ ، مقدار آستانه قرار داده‌شده روی معیار فاصله است که طبیعتاً وابسته به آستانه انتخابی  $\tau$  است؛ همچنین، پارامتر TPTR نیز از رابطه زیر به‌دست می‌آید:

$$TPTR(R, \tau) = \frac{P(R, \tau)}{S} \quad (۴)$$

در این رابطه  $S$ ، تعداد تصاویر متعلق به گالری است که برای آزمایش به الگوریتم داده شده است.  $P$  نیز تعدادی از این تصاویر است که در آنها، الگوریتم به درستی فرد مورد نظر را مشخص کرده است.

$R$ ، پارامتر رتبه<sup>۳</sup> است. به‌طورمعمول یک الگوریتم شناسایی چهره به‌جای یک نتیجه،  $R$  نامزد به‌ترتیب دقت تعلق، خروجی می‌دهد. هر چه  $R$  بیشتر باشد، مقدار TPTR بیشتر خواهد بود. مشابه FPIR،  $\tau$  مقدار آستانه است. طبیعتاً، زمانی که رتبه برابر یک باشد، فقط یک فرد به‌عنوان نتیجه شناسایی، اعلام می‌شود.

یک الگوریتم شناسایی چهره، زمانی الگوریتم مناسبی است که اگر تصویر آزمون متعلق به هیچ فرد داخل گالری نباشد آن را اعلام کند و اگر تصویر، متعلق به فردی در گالری باشد، به‌درستی آن فرد را تشخیص دهد. پارامتر FPIR، دقت الگوریتم در تشخیص تعلق یا عدم تعلق تصاویر به گالری را می‌سنجد و TPTR بیان‌گر میزان کارایی الگوریتم در تشخیص فرد صحیح است.

پارامتر 1-TPTR در اصطلاح "نرخ تشخیص رد به‌اشتباه" (FNIR<sup>۴</sup>) نامیده می‌شود.  $FNIR(R, \tau)$  بیان‌گر آن است که از بین تصاویر آزمون که در واقع متعلق به فردی در داخل گالری هستند، الگوریتم با چه نسبتی فرد مورد نظر را به‌درستی تشخیص نداده است.

واضح است که هر دو پارامتر TPTR و FPIR در بازه  $[0, 1]$  هستند. البته، برای این‌که یک الگوریتم شناسایی چهره عملکرد مناسبی داشته باشد، می‌بایستی FPIR مقدار کم (نزدیک به صفر) و TPTR، مقدار بزرگ (نزدیک به یک) داشته باشد.

<sup>3</sup> Rank

<sup>4</sup> False negative identification rate (FNIR)

در پیاده‌سازی‌های انجام‌شده، میانگین وزن‌دار TPIR و FPIR محاسبه شده است. این میانگین وزن‌دار بر حسب تعداد تصاویر متعلق به افراد داخل گالری و تصاویر غیر مرتبط در نظر گرفته شده و تحت عنوان نمره نهایی (FS)<sup>1</sup> استفاده شده است. این پارامتر، برآیند کارایی کلی الگوریتم را نشان می‌دهد.

به‌منظور محاسبه یک مقدار مناسب برای آستانه  $\tau$ ، از معیار نرخ خطای معادل (EER)<sup>2</sup> استفاده شده است. در محاسبه EER، ابتدا مقداری از  $\tau$  را پیدا می‌کنیم که به‌ازای آن  $FNR(R, \tau) = FPIR(\tau)$  یک مقدار ثابت در نظر گرفته می‌شود. اگر این مقدار آستانه را  $\hat{\tau}$  بنامیم  $EER = FNR(R, \hat{\tau}) = FPIR(\hat{\tau})$  عدد  $\hat{\tau}$  می‌تواند به‌عنوان یک مقدار مناسب برای آستانه، محسوب شود.

### ۳-۴- نتایج پیاده‌سازی، بررسی و تحلیل

در این بخش به بیان، بررسی و تحلیل نتایج الگوریتم ارائه شده در بخش ۳-۳ می‌پردازیم. نکات زیر درخصوص این پیاده‌سازی قابل ذکر است:

- CNN مورد استفاده برای استخراج ویژگی‌های اولیه در این قسمت، همان شبکه VGGFace ارائه شده به‌وسیلهٔ برخی است. این شبکه با ۲۶۲۲ نفر و از هر نفر هزار تصویر، آموزش داده شده است (درجمع ۲۶۲۲۰۰۰ تصویر). جزئیات مربوط به لایه‌ها و نحوه ساخت شبکه در [20] ارائه شده است.
- در پیاده‌سازی انجام‌شده، خروجی لایه ۳۶ به‌عنوان بردار ویژگی، در نظر گرفته است. با توجه به اینکه، تعداد عناصر این بردار ویژگی، ۴۰۹۶ عدد است، در نهایت، بردار ویژگی که در این پروژه برای هر تصویر مورد استفاده قرار می‌گیرد، ۴۰۹۶ مقداری خواهد بود.
- ارزیابی کارایی با استفاده از هر دو معیار TPIR و FPIR انجام گرفته است.
- هر چهار مجموعه داده AT&T، MUCT، LFW و 100LFW مورد ارزیابی قرار گرفته‌اند. شبکه VGG-

Face فرض می‌کند، تصاویر ورودی رنگی و با اندازه  $۲۲۴ \times ۲۲۴$  باشند و لذا، روی هر چهار داده، یک تغییر اندازه<sup>۴</sup>، انجام شده است؛ به‌علاوه، داده AT&T به‌صورت تصاویر خاکستری هستند. با سه تکرار مؤلفه‌های خاکستری، یک تصویر، که هر پیکسل دارای سه مقدار است، ساخته و این داده جدید، به شبکه اعمال شد.

در ادامه، نتایج یک سری از آزمایش‌ها و پیاده‌سازی‌ها بیان شده و مورد تحلیل و بررسی قرار می‌گیرند. با توجه به مطالب یادشده، برای انجام پیاده‌سازی، حالات زیر در نظر گرفته شده است:

- به‌منظور مقایسه بردار ویژگی آزمون و بردار ویژگی‌های مربوط به هر فرد در داده آموزشی، سه معیار فاصله کسینوسی، همبستگی و فاصله اقلیدسی در نظر گرفته است.
  - سه حالت برای تبدیل به‌کار گرفته شده است: بدون استفاده از تبدیل (ویژگی‌های به‌دست‌آمده از CNN)، با استفاده از تبدیل فوریه و تبدیل موجک
  - سه معیار برای ارزیابی در نظر گرفته شده است: TPIR، FPIR و ERR با مقادیر آستانه مختلف.
  - چهار مجموعه‌داده برای ارزیابی به‌کار برده شده‌اند: AT&T، MUCT، LFW و 100LFW.
- بدین ترتیب تعداد حالات بسیار زیادی برای پیاده‌سازی وجود دارند که با توجه محدودیت‌های موجود، تنها بخشی از حالات بیان‌شده، تحت قالب آزمایش‌های زیر، مورد ارزیابی قرار گرفته‌اند:
- با توجه به آن‌که، هدف، بررسی تأثیر پس‌پردازش ارائه‌شده است، مقایسه روی دو حالت استفاده از پس‌پردازش و عدم استفاده از آن، انجام گرفته و نتایج حاصله در جداول، آورده شده است.
- در جدول (۶) تأثیر ویژگی‌های به‌دست‌آمده از اعمال تبدیلات مختلف، در افزایش کارایی الگوریتم شناسایی چهره مبتنی بر شبکه VGGFace نشان داده شده است. به‌کارگیری تبدیلات فوریه و موجک، منجر به حفظ کارایی و یا افزایش آن در بعضی موارد شده است. جدول (۷) مانند جدول (۶) است؛ با این تفاوت که به جای فاصله کسینوسی، فاصله همبستگی در نظر گرفته است. نتایج مناسب روش پیشنهادی روی انواع مختلف معیارهای سنجش شباهت، نشان‌دهنده آن است که بهبود کارایی، ناشی از ایده ارائه‌شده بوده و ارتباطی با نوع معیار سنجش شباهت مورد استفاده ندارد.

<sup>4</sup> Resizing

<sup>1</sup> Final score (FS)

<sup>2</sup> Equal error rate (ERR)

<sup>۳</sup> درعمل، برای داده‌های واقعی، ممکن است هیچ‌گاه این دو معیار، دقیقاً با هم مساوی نشوند؛ لذا روال به این صورت است که ابتدا به‌ازای مقادیر مختلف  $\tau$ ، FPIR و FNIR را محاسبه می‌کنند. سپس،  $\hat{\tau}$  را برابر با مقداری می‌گیرند که کم‌ترین تفاضل را بین FPIR و FNIR موجب شده باشد.

ایده بیان‌شده منجر به بهبود وضعیت  $TPIR$  و یا به بیان معادل  $FNIR$  شده است. در مقابل،  $FPIR$  رویکرد ارائه‌شده در بعضی موارد، وضعیت نامناسب‌تری نسبت به حالت بدون تبدیل دارد. بدین ترتیب به‌کارگیری الگوریتم پیشنهادی زمانی مناسب است که بدانیم تصاویر آزمون، متعلق به گالری هستند و الگوریتم فقط می‌خواهد فردی را که تصاویر متعلق به او است، در گالری بیابد.

در شکل‌های (۵) و (۶)، نمودار  $FPIR$  و  $FNIR$  به‌ازای ۲۰ مقدار آستانه مختلف از صفر تا یک با گام ۰/۰۵ و با فرض استفاده از فاصله‌های همبستگی و اقلیدسی، رسم شده است. داده 100LFW مورد استفاده قرار گرفته و با توجه به نمودارهای رسم‌شده، پارامتر EER (محل تقاطع نمودارهای  $FPIR$  و  $FNIR$ )، به‌دست آمده و در جدول‌های (۸ و ۹) آورده شده است. جدول (۱۰) مشابه جدول‌های (۶ و ۷) است؛ با این تفاوت که فاصله اقلیدسی به‌جای فاصله‌های کسینوسی و همبستگی جایگزین شده است. با توجه به آن که از نظر تئوری، فاصله اقلیدسی می‌تواند هر مقداری در بازه  $(0, +\infty)$  باشد، امکان در نظر گرفتن آستانه، فراهم نیست. در این حالت، تصویر چهره آزمون، به فردی در داخل گالری منسوب می‌شود که کمترین فاصله را موجب شده باشد. به‌علت عدم در نظر گرفتن مقدار آستانه، پارامتر  $FPIR$  قابل محاسبه نبوده و لذا در جدول، صرفاً معیار  $TPIR$  آورده شده است. نتایج حاصل از اعمال فاصله اقلیدسی نیز نشان‌دهنده کارایی بهتر رویکرد ارائه‌شده است.

نکته قابل توجه، آن است که نتایج پیاده‌سازی الگوریتم پیشنهادی، هم روی دادگان کنترل‌شده (آزمایشگاهی) مانند AT&T و MUCT و هم روی مجموعه داده‌های واقعی (کنترل‌نشده و جمع‌آوری‌شده از وب) یعنی LFW و 100LFW نتایج مناسبی داشته است و این امر خود، مؤیدی بر کارآوردن ایده ارائه‌شده است؛ به‌خصوص آن که نتایج الگوریتم، روی مجموعه‌داده چالشی‌ای چون LFW، بهبود چشم‌گیری داشته است؛ در ضمن گفتنی است که نتیجه رویکرد پیشنهادی روی داده‌های آزمایشگاهی که بدون اعمال روش نیز دقت بالایی دارند، به‌طور کامل مناسب بوده و کارایی را پایین نیاورده است. در داده AT&T دقت شناسایی بدون اعمال تبدیل، ۱۰۰٪ بوده که بعد از اعمال تبدیل، این دقت هیچ کاهشی نداشته است. برای داده MUCT حتی افزایش بسیار جزئی در کارایی داشته‌ایم (در جدول (۱۰)).

به‌منظور نشان‌دادن کارایی الگوریتم در مواجهه با روش‌های غیر مبتنی بر شبکه‌های عصبی، در جدول (۱۱) نتایج حاصل از روش‌های استخراج ویژگی LDA<sup>۱</sup> و KFA<sup>۲</sup> نیز آورده شده است (برای آشنایی با LDA و KFA به ترتیب به [28] و [29] مراجعه شود). مجموعه‌داده AT&T در این آزمایش در نظر گرفته شده است. نتایج جدول نیز حاکی از کارایی مناسب الگوریتم ارائه‌شده است. گفتنی است با توجه به دقت بسیار کم روش‌های سنتی روی داده‌های کنترل‌شده، این آزمایش روی سایر مجموعه داده‌ها، انجام نشده است.

در انتهای این بخش، در خصوص سربرار محاسباتی ایده ارائه‌شده، توضیحاتی ارائه می‌شود. قسمت محاسبه تبدیلات در الگوریتم، از نظر زمان اجرا وضعیت مناسبی دارد. به‌عنوان مثال، محاسبه تبدیل فوریه یا معکوس آن، با الگوریتم مشهور پروانه<sup>۳</sup> است که با مرتبه زمانی  $O(dN_i \log N_i)$  قابل انجام است. در رابطه یادشده،  $N_i$  تعداد نقاط خروجی تبدیل و  $d$ ، تعداد ویژگی‌ها هستند. بقیه بخش‌های الگوریتم نیز شامل مواردی مانند مرتب‌سازی و پیمایش نقاط بوده و مرتبه‌های زمانی مناسبی دارند. از نظر سربرار حافظه‌ای نیز با توجه به آن که داده آزمون به هر دسته اضافه شده است، سربرار به‌نسبه کوچکی تحمیل می‌شود که این سربرار متناسب با تعداد دسته‌ها است. با توجه به موارد بیان‌شده، ایده پیشنهادی، دارای سربرار معقول و مناسبی است.

## ۵- نتیجه‌گیری

در این مقاله، روشی برای استخراج ویژگی‌های تصاویر چهره به‌منظور استفاده از آنها در شناسایی چهره ارائه شد. ایده پیشنهادی، کلی بوده و قابل به‌کارگیری در هر روش استخراج ویژگی و هر الگوریتم دسته‌بندی یا شناسایی الگو است.

در الگوریتم ارائه‌شده، داده آزمون به همه دسته‌های موجود در گالری، اضافه شده و تبدیلی روی داده‌های هر دسته که اکنون شامل داده آزمون نیز است، اعمال می‌کنیم. فرآیند تبدیل و سایر پردازش‌های انجام‌شده بعد از آن، باعث می‌شود که اگر داده آزمون به دسته نامناسب یعنی به دسته‌ای که داده متعلق به آن نیست، اضافه شود، اثر مخرب ایفا کرده و بدین ترتیب، احتمال خطا (یعنی تشخیص

<sup>1</sup> Linear Discriminant Analysis (LDA)

<sup>2</sup> Kernel Fisher Analysis (KFA)

<sup>3</sup> Butterfly algorithm

نادرست دسته) را کمتر می‌کند؛ همچنین، افزودن داده آزمون به دسته مناسب، یعنی به دسته‌ای که داده در واقع متعلق به آن است، اثر تخریبی کمتری دارد. در اعمال الگوریتم پیشنهادی، دو نوع تبدیل معروف (فوری و موجک) مورد استفاده قرار گرفته‌اند. به منظور نشان دادن توانایی روش بیان شده، کارایی آن در بهبود ویژگی‌های استخراج شده از

شبکه VGGFace که یکی از قوی‌ترین CNN‌های ارائه شده در حوزه پردازش تصاویر چهره است، نشان داده شده است؛ به علاوه، کارایی مناسب رویکرد ارائه شده، روی داده‌های چالشی (مانند LFW) و روی داده‌های آزمایشگاهی (مانند AT&T) اثبات شده، به طوری که بر روی LFW بهبودی در حدود ۳/۴٪ داشته است.

(جدول ۶-): تأثیر ویژگی‌های به دست آمده از اعمال تبدیلات مختلف در کارایی الگوریتم شناسایی چهره مبتنی بر شبکه VGGFace.

(R=1 و فاصله کسینوسی با  $\tau = 0.3$  در نظر گرفته شده است. در تبدیل موجک، DB11 مورد استفاده قرار گرفته است).

(Table -6): Effect of the features obtained from applying different transformations on the efficiency of the face recognition algorithm based on the VGGFace network (R=1 and the cosine distance is considered by  $\tau = 0.3$ . In the wavelet transform, the DB11 has been used.)

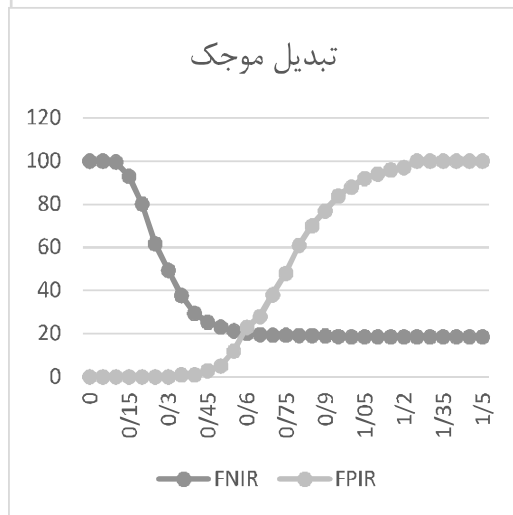
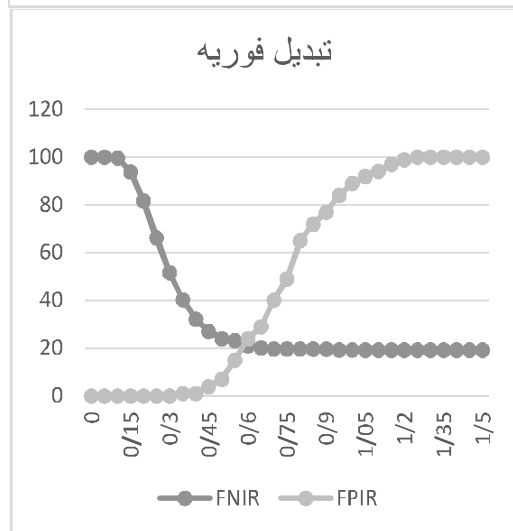
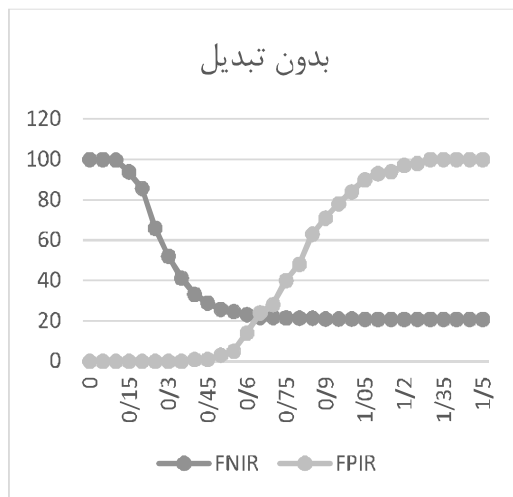
Final score	FPIR	TPIR	بدون تبدیل
100	0	100	AT&T
99.74	2	99.97	MUCT
86.52	8	86.22	LFW
90.5	0	81	100LFW
Final score	FPIR	TPIR	تبدیل فوری
100	0	100	AT&T
99.74	2	99.97	MUCT
86.57	8	86.27	LFW
90.5	0	81	100LFW
Final score	FPIR	TPIR	تبدیل موجک
100	0	100	AT&T
99.74	2	99.97	MUCT
90.33	18	89.655	LFW
91.1	0	82.2	100LFW

(جدول ۷-): تأثیر ویژگی‌های به دست آمده از اعمال تبدیلات مختلف در کارایی الگوریتم شناسایی چهره مبتنی بر شبکه VGGFace.

(R=1 و فاصله همبستگی با  $\tau = 0.5$  در نظر گرفته شده است. در تبدیل موجک، DB11 مورد استفاده قرار گرفته است).

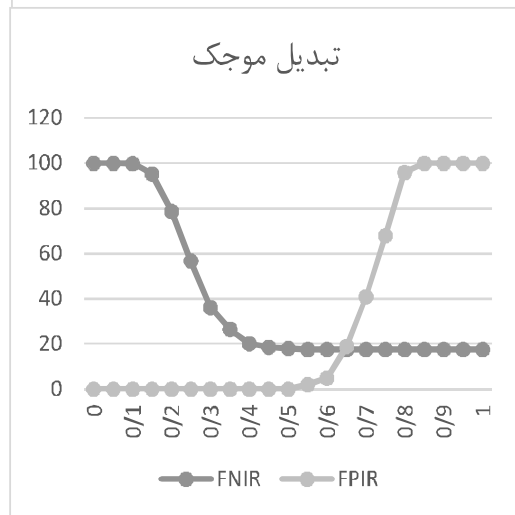
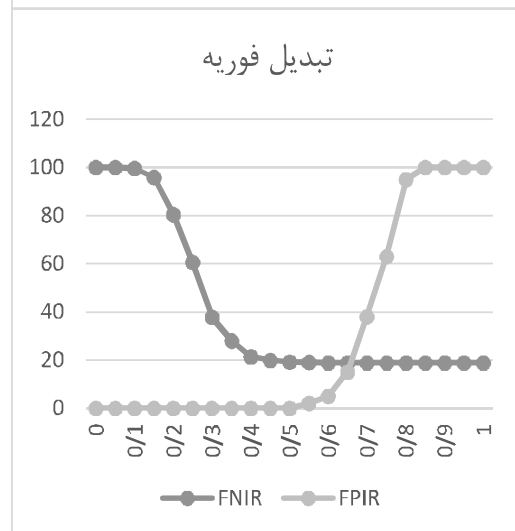
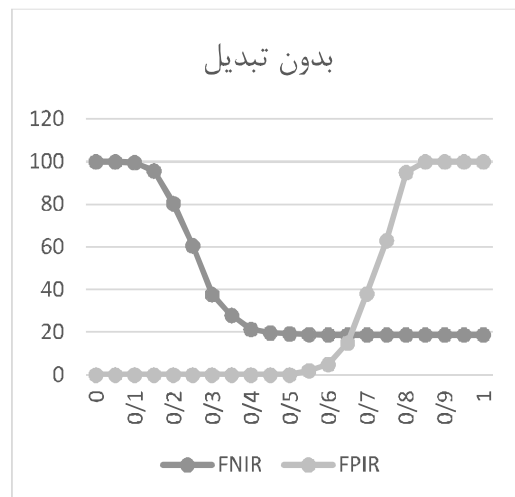
(Table -7): Effect of the features obtained from applying different transformations on the efficiency of the face recognition algorithm based on the VGGFace network (R=1 and the correlation distance is considered by  $\tau = 0.5$ . In the wavelet transform, the DB11 has been used.)

Final score	FPIR	TPIR	بدون تبدیل
100	0	100	AT&T
99.62	3	99.97	MUCT
87.28	17	87.52	LFW
90.4	0	80.8	100LFW
Final score	FPIR	TPIR	تبدیل فوری
100	0	100	AT&T
99.62	3	99.97	MUCT
87.28	17	87.52	LFW
90.4	0	80.8	100LFW
Final score	FPIR	TPIR	تبدیل موجک
100	0	100	AT&T
99.51	4	99.97	MUCT
89.06	29	90.05	LFW
91	0	82	100LFW



(شکل-۶): FNIR و FPIR به‌ازای بیست مقدار آستانه مختلف از صفر تا ۱/۵ با گام ۰/۰۵ با فرض استفاده از فاصله اقلیدسی ( $R=1$  و داده 100LFW استفاده شده است).

(Figure-6): FPIR and FNIR for 20 different threshold values from zero to 1.5 with step 0.05 assuming the use of Euclidian distance ( $R=1$  and 100LFW is used.)



(شکل-۵): FNIR و FPIR به‌ازای بیست مقدار آستانه مختلف از صفر تا یک با گام ۰/۰۵ با فرض استفاده از فاصله همبستگی ( $R=1$  و داده 100LFW استفاده شده است).

(Figure-5): FPIR and FNIR for 20 different threshold values from zero to one with step 0.05 assuming the use of correlation distance ( $R=1$  and 100LFW is used.)

(Table-11): Results of the proposed algorithm on features obtained from non-neural networks-based methods with different transformations. AT&T dataset has been considered. (R=1, Euclidian distance and DB11 wavelet transform have been used)

روش	بدون تبدیل	تبدیل فوریه	تبدیل موجک
LDA	93.5	93.5	93.75
KFA	96.25	97	97

## 6- References

## ۶- مراجع

- [1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097-1105, 2012.
- [2] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," *IEEE transactions on pattern analysis and machine intelligence*, vol. 15, no. 10, pp. 1042-1052, 1993.
- [3] R. J. Baron, "Mechanism of human facial recognition," *International Journal of Man Machine Studies*, vol. 15, pp. 137-178, 1981.
- [4] D. Ghimirc, J. Lcc, Z.-N. Li, and S. Jeong, "Recognition of facial expressions based on salient geometric features and support vector machines," *Multimedia Tools and Applications*, vol. 76, no. 6, pp. 7921-7946, 2017
- [5] L. Wiskott, N. Krüger, N. Kuiger, and C. Von Der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 19, no. 7, pp. 775-779, 1997.
- [6] J. yves Bouguet, "Pyramidal implementation of the lucas kanade feature tracker," *Intel Corporation, Microprocessor Research Labs*, 2000.
- [7] Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *European conference on computational learning theory*, 1995: Springer, pp. 23-37.
- [8] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 2, pp. 513-529, 2012.
- [9] W. Ouarda, H. Trichili, A. M. Alimi, and B. Solaiman, "Face recognition based on geometric features using Support Vector Machines," in *Soft Computing and Pattern Recognition (SoCPaR), 2014 6th International Conference of*, 2014: IEEE, pp. 89-95.
- [10] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Computer Vision and Pattern Recognition, 1991. Proceedings*

(جدول-۸): مقایسه کارایی الگوریتم ارائه شده با اعمال تبدیلات مختلف بر حسب پارامتر EER. روی داده 100LFW (R=1) ، فاصله همبستگی و تبدیل موجک DB11 مورد استفاده قرار گرفته اند.

(Table-8): Comparison of the efficiency of the proposed algorithm with different transformations according to the EER parameter over 100LFW data (R=1, correlation distance and DB11 wavelet transform have been used).

بدون تبدیل	تبدیل فوریه	تبدیل موجک	100LFW
16.9	16.9	18.3	EER

(جدول-۹): مقایسه کارایی الگوریتم ارائه شده با اعمال تبدیلات مختلف بر حسب پارامتر EER. روی داده 100LFW (R=1) ، فاصله اقلیدسی و تبدیل موجک DB11 مورد استفاده قرار گرفته اند.

(Table-9): Comparison of the efficiency of the proposed algorithm with different transformations according to the EER parameter over 100LFW data (R=1, Euclidian distance and DB11 wavelet transform have been used)

بدون تبدیل	تبدیل فوریه	تبدیل موجک	100LFW
22.9	22.5	21.6	EER

(جدول-۱۰): تأثیر ویژگی های به دست آمده از اعمال تبدیلات مختلف در کارایی الگوریتم شناسایی چهره مبتنی بر شبکه VGGFace . (R=1) . فاصله اقلیدسی و تبدیل موجک DB11 مورد استفاده قرار گرفته اند.

(Table-10): Effect of the features obtained from applying different transformations on the efficiency of the face recognition algorithm based on the VGGFace network (R=1, Euclidian distance and DB11 wavelet transform have been used)

تبدیل فوریه	TPIR
AT&T	100
MUCT	99.95
LFW	91.02
100LFW	80.8
تبدیل موجک	TPIR
AT&T	100
MUCT	99.95
LFW	91.63
100LFW	81.4
بدون تبدیل	TPIR
AT&T	100
MUCT	99.92
LFW	87.19
100LFW	79.2

(جدول-۱۱): نتایج الگوریتم ارائه شده با فرض استفاده از ویژگی های به دست آمده از روش های غیر مبتنی بر شبکه های عصبی با اعمال تبدیلات مختلف. داده در نظر گرفته شده، AT&T می باشد (R=1) . فاصله اقلیدسی و تبدیل موجک DB11 مورد استفاده قرار گرفته اند

- [22] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proceedings of 1994 IEEE workshop on applications of computer vision*, 1994: IEEE, pp. 138-142.
- [23] S. Milborrow, J. Morkel, and F. Nicolls, "The MUCT landmarked face database," *Pattern Recognition Association of South Africa*, vol. 201, no. 0, 2010.
- [24] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," in *Workshop on faces in Real-Life Images: detection, alignment, and recognition*, 2008.
- [25] A. Vedaldi and K. Lenc, "Matconvnet: Convolutional neural networks for matlab," in *Proceedings of the 23rd ACM international conference on Multimedia*, 2015: ACM, pp. 689-692.
- [26] GTDLBench. *The Database of Faces (AT&T)* [Online]. Available: [https://github.com/GTDLBench/datasets/att\\_face\\_dataset/](https://github.com/GTDLBench/datasets/att_face_dataset/).
- [27] P. Grother and M. Ngan, "Performance of face identification algorithms," *NIST Inter-agency Internal Report 8009*, 2014.
- [28] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of eugenics*, vol. 7, no. 2, pp. 179-188, 1936.
- [29] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.-R. Mullers, "Fisher discriminant analysis with kernels," in *Neural networks for signal processing IX: Proceedings of the 1999 IEEE signal processing society workshop (cat. no. 98th8468)*, 1999: IEEE, pp. 41-48.
- [11] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of educational psychology*, vol. 24, no. 6, p. 417, 1933.
- [12] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 19, no. 7, pp. 711-720, 1997.
- [13] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 2, pp. 210-227, 2009.
- [14] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma, "Toward a practical face recognition system: Robust alignment and illumination by sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 2, pp. 372-386, 2012.
- [15] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1891-1898.
- [16] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *Advances in neural information processing systems*, 2014, pp. 1988-1996.
- [17] 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.
- [18] Y. Sun, D. Liang, X. Wang, and X. Tang, "Deepid3: Face recognition with very deep neural networks," *arXiv preprint arXiv:1502.00873*, 2015.
- [19] Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1701-1708.
- [20] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," in *BMVC*, 2015, vol. 1, no. 3, p. 6.
- [21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.



مرتضی علی احمدی تحصیلات

کارشناسی خود را در رشته علوم رایانه در دانشگاه قم، در سال ۱۳۹۱ و همچنین تحصیلات کارشناسی ارشد خود را در رشته مهندسی فناوری اطلاعات، در همان

دانشگاه، در سال ۱۳۹۳ به پایان رسانده و همچنین در حال حاضر دانشجوی دکترای رشته مهندسی فناوری اطلاعات است. زمینه پژوهشی و مورد علاقه وی، پردازش صدا و تصویر و داده‌کاوی است.

نشانی رایانامه ایشان عبارت است از:

ma.ahmadi@qom.ac.ir



**روح‌الله دیانت** تحصیلات کارشناسی ارشد

و دکترای خود را به ترتیب در سال‌های

۱۳۸۲ و ۱۳۸۹ در دانشگاه صنعتی شریف

به اتمام رسانده و هم‌اکنون استادیار

دانشگاه قم (گروه مهندسی کامپیوتر و فناوری اطلاعات)

است. زمینه اصلی فعالیت وی، پردازش چندرسانه‌ای (صدا،

تصویر و ویدئو) است.

نشانی رایانامه ایشان عبارت است از:

**rdianat@qom.ac.ir**