

# شناسایی چهره در رشته‌های ویدئویی با استفاده از افکنش متعامد با حفظ ساختار محلی

سجاد شفیع پور یورداشahi<sup>۱</sup>، میرهادی سیدعربی<sup>۲</sup> و علی آقاگلزاده<sup>۳</sup>

<sup>۱</sup>دانشکده مهندسی برق و کامپیوتر، دانشگاه تبریز، تبریز، ایران

<sup>۲</sup>دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی نوشیروانی بابل، بابل، ایران

## چکیده

در این پژوهش، تلاش برای بهبود نرخ شناسایی و حل مشکلات ناشی از چرخش سر، تغییر شدت روشنایی و پوشش قسمتی از چهره در رشته‌های ویدئویی با استفاده از افکنش متعامد با حفظ ساختار محلی (OLPP<sup>۱</sup>) است. در این پژوهش ابتدا برای حذف تصویر زمینه، چهره در قاب‌های ویدئویی تشخیص داده می‌شود؛ سپس با ارائه یک روش مناسب، مجموعه تصاویر چهره که به صورت منیفوولد غیرخطی توزیع یافته است، خوشبندی شده و مراکز خوششها به عنوان نماینده هر خوشش تعیین می‌شوند. در این مقاله نشان داده می‌شود که با استفاده از افکنش OLPP قاب‌های کلیدی به یک فضای جدیدی منتقل می‌شوند که در آن قاب‌های متعلق به یک منیفوولد به هم نزدیک‌تر و از قاب‌های منیفوولدی دیگر دور می‌شوند. برای شناسایی، رشته ویدئویی آزمایشی به فضای جدید افکنش شده و فاصله بین منیفوولدها محاسبه می‌شود. مقایسه نتایج الگوریتم پیشنهادی با روش‌های دیگر نشان دهنده کارآیی بالای روش پیشنهادی است.

واژگان کلیدی: شناسایی چهره، افکنش متعامد با حفظ ساختار محلی، قاب کلیدی، زیرفضا، منیفوولد

## ۱- مقدمه

تغییراتی همچون روشنایی، حالات چهره، سن، چرخش سر و پوشش قسمتی از چهره، وجود ندارد. روش‌های شناسایی از جهت نوع اطلاعات ورودی به دو دستهٔ تصاویر ثابت و تصاویر ویدئویی تقسیم‌بندی می‌شوند. بیشتر الگوریتم‌های ارائه شده در زمینهٔ شناسایی چهره مربوط به تصاویر ثابت است (ژاو و همکاران، ۲۰۰۳)؛ اما به تازگی، با افزایش دوربین‌های امنیتی، شناسایی چهره بر اساس ویدئو بیشتر مورد توجه قرار گرفته است. روش‌های شناسایی چهره در رشته‌های ویدئو به دو دستهٔ تقسیم‌بندی می‌شوند. روش‌های مبتنی بر مجموعهٔ تصاویر که در آن ویدئو به صورت مجموعهٔ تصاویر در نظر گرفته می‌شود و از مزایای تعداد مشاهدات مختلف از چهره، استفاده و روش‌های مبتنی بر رشته که از خاصیت اطلاعات زمانی رشته‌های ویدیوئی برای شناسایی بهره‌برداری می‌کنند (بار و همکاران، ۲۰۱۲).

در طول چند سال گذشته پیشرفت‌های سریع در زمینه تعاملات انسان با رایانه و سامانه‌های هوشمند، موجب افزایش علاقهٔ پژوهش‌گران بینایی ماشین به مباحث شناسایی شخص شده است. ویژگی‌های فیزیکی و رفتاری افراد، همچون چهره، صدا، عنایی و اثر انگشت به طور گسترده در شناسایی شخص مورد استفاده قرار می‌گیرند. در میان این مشخصات، استفاده از چهره، با وجود این که دقیق بهاندازه روش‌های دیگر شناسایی همچون اثر انگشت ندارد؛ ولی به علت ماهیت غیر دخالت‌کننده‌ی آن و نیاز به همکاری کمتر افراد در اجرای پردازش شناسایی، بیشتر از روش‌های موجود دیگر مورد توجه قرار گرفته است. در طول دهه گذشته روش‌های مختلفی در زمینهٔ شناسایی چهره ارائه شده است؛ ولی با وجود پیشرفت‌های قابل توجه هنوز یک سامانه کامل که دارای دقت، سرعت بالا و مستقل از

<sup>۱</sup> Orthogonal Locality Preserving Projection



## ۲- بررسی منابع

یاماگوچی و همکارانش از روش زیر فضای متقابل (MSM<sup>۲</sup>) برای شناسایی چهره در رشته‌های ویدئویی استفاده کردند (یاماگوچی و همکاران، ۱۹۹۸). روش کار آن‌ها به این صورت بود که با استفاده از روش زیرفضای متقابل، قاب‌های ویدئو برای هر شخص به‌طور مجزا در نظر گرفته شده و فضای ویژه را برای آن شخص محاسبه کردند؛ سپس با محاسبه زوایایی<sup>۳</sup> بین زیرفضای مرجع و ورودی، معیار شباهت دو زیرفضا را به‌دست آوردند. نویسنده‌گان این مقاله نشان دادند که روش زیرفضای متقابل دارای عملکرد بهتر از روش مرسوم زیرفضای تصویر تنها است.

نیشی‌یاما و همکارانش روش زیرفضای متقابل را با استفاده از چند دوربین و انجام شناسایی سلسله‌مراتبی توسعه دادند (نیشی‌یاما و همکاران، ۲۰۰۷). آنها از سه لایه تطبیق استفاده کردند. نویسنده‌گان گزارش دادند که با استفاده از چند دوربین و تقسیم فرآیند تطبیق به زیرپردازها تأثیر تغییرات روشنایی و زاویه چهره در انجام شناسایی کاهش می‌یابد.

وطف و همکارش از یک کرنل برای نگاشت داده از فضای اصلی به فضای ویژگی غیرخطی استفاده کرده و با محاسبه زوایایی اصلی، شناسایی چهره را در رشته‌های ویدئویی انجام دادند (ولف و شاشیو، ۲۰۰۳). چنگ و همکارانش برای تعیین فواصل بین زیر فضاها از محاسبه زوایایی روی منیفولدی‌های گراسمن در فضای روشنایی استفاده کرده و ثابت کردند که پایگاه داده CMU PIE به علت مشخصات خطی اش در فضای روشنایی می‌تواند به‌طور کامل جدا شود (چنگ و همکاران، ۲۰۰۶). اگرچه شناسایی چهره در دنیای واقعی نمی‌تواند در چنین محیط کنترل شده‌ای عمل کند. کیم و همکارانش روش بهینه‌ای شبیه<sup>۴</sup> LDA را پیشنهاد دادند. بدین صورت که شناسایی چهره در رشته‌های ویدئویی با بیشینه کردن همبستگی‌های بنیادی مجموعه تصاویر درون‌طبقه‌ای و کمینه کردن همبستگی‌های بنیادی مجموعه تصاویر برونو طبقه‌ای انجام گرفت. رشته‌های آزمایشی با تابع جداساز یادگیری انتقال یافته و امتیاز شباهت، با همبستگی‌های بنیادی محاسبه شد. در این روش تعداد زیادی از داده‌های یادگیری لازم است (کیم و همکاران، ۲۰۰۷).

روش‌های مبتنی بر مجموعه تصاویر به ترتیب تصاویر چهره وابسته نیست در این روش‌ها تعداد و تنوع مشاهدات برای انجام شناسایی اهمیت دارد. این روش‌ها از جهت این که تلفیق، قبل و یا بعد از طبیق باشد با هم تفاوت دارند (راس، ۲۰۰۷؛ ساندرسون و پاریوال، ۲۰۰۴). قبل از طبیق، اطلاعات تصاویر می‌تواند در داده‌ها و یا سطوح ویژگی‌ها با هم تلفیق شوند. تکنیک‌های تفکیک‌پذیری بالا در این سطوح عمل کرده و قدرت تفکیک‌پذیری چهره را افزایش می‌دهند. به‌طور مشابه تکنیک‌های مدل‌سازی سه‌بعدی، برای ساختن مدل سه‌بعدی چهره از دیدهای مختلف چهره استفاده می‌کنند. همچنین در این روش‌ها مجموعه کامل مشاهدات را می‌توان به‌صورت یک منیفولد و یا یک توزیع آماری مدل‌سازی کرد. شناسایی با استفاده از این طرز نمایش نسبت به تغییرات زاویه چهره، روشنایی و حالات چهره می‌تواند مقاوم باشد. همچنین تلفیق می‌تواند در سطحی از امتیازات تطبیق‌ها و یا در زیرمجموعه‌ای از تصاویری که دارای شرایط ظاهری خاص و یا مشاهدات با کیفیت بالا هستند، انجام پذیرد.

برخلاف روش‌های مبتنی بر مجموعه تصاویر، روش‌های مبتنی بر رشته از اطلاعات زمانی رشته‌ها برای شناسایی استفاده می‌کنند. تکنیک‌های زمانی-مکانی<sup>۱</sup> هم از نظر ظاهر و هم از نظر حرکت برای دست‌یابی به شناسایی استفاده می‌کنند؛ در حالی که روش‌های زمانی فقط از حرکات ویژه صورت برای انجام شناسایی بهره‌برداری می‌کنند. در چهره، عمل کرد شناسایی را در شرایط انسداد چهره و یا تغییر شکل موقت در قسمتی از چهره بهبود بخشد.

در این پژوهش هر رشته ویدئویی به‌صورت یک منیفولد غیرخطی در نظر گرفته شده و شناسایی چهره از طریق محاسبه فاصله بین منیفولد‌ها انجام می‌شود. در الگوریتم پیشنهادی ابتدا منیفولد‌ها خوشه‌بندی شده و به تعدادی زیرفضای خطی تبدیل می‌شوند؛ سپس برای افزایش فاصله بین منیفولد‌ها با استفاده از یک افکنش مناسب زیرفضاهای خطی و قاب‌های کلیدی به فضای جدید منتقل می‌شوند و شناسایی با انتقال رشته ویدئویی آزمایشی به این فضای جدید و محاسبه فاصله بین زیرفضاهای و قاب‌های کلیدی انجام می‌پذیرد. در ادامه بعد از بررسی منابع، الگوریتم پیشنهادی توضیح داده و سپس نتایج پژوهش با روش‌های موجود مقایسه می‌شود.

<sup>2</sup> Mutual Subspace Method

<sup>3</sup> Principal angles

<sup>4</sup> Linear Discriminate Analysis



در فضای چهره صورت می‌پذیرد. در (بزرگ‌تبار و راد، ۲۰۱۱) بزرگ‌تبار و همکارش از برنامه‌ریزی ژنتیک (GP<sup>۴</sup>) برای شناسایی چهره استفاده کردند؛ بدین صورت که ابتدا توسط PCA ویژگی‌های چهره را استخراج، سپس از GP جهت دسته‌بندی تصاویر و شناسایی چهره استفاده کردند. در (شمسمی بابلی، ۲۰۱۳) شمسی بابلی و همکارانش عملیات شناسایی چهره را توسط یک زیرفضای بهینه که با استفاده از تجزیه و تحلیل المان اصلی چندخطی و تجزیه و تحلیل جداساز چند خطی به دست می‌آید، انجام دادند.

### ۳- الگوریتم پیشنهادی

هر رشته ویدئویی با منیفولد غیرخطی  $M_i$  نمایش داده می‌شود. بنابراین مسئله شناسایی به محاسبه فاصله بین منیفولدها تبدیل می‌شود. برای به دست آوردن این فاصله باستی منیفولدهای غیرخطی به تعدادی خوش که هر کدام به صورت یک زیرفضای خطی است، تبدیل شوند. بنابراین ابتدا یک الگوریتم برای خوشبندی مناسب منیفولدهای غیرخطی مطرح می‌شود. این الگوریتم علی‌رغم سادگی، بسیار کارا بوده و موجب خوشبندی هوشمندانه منیفولدها به تعدادی زیرفضای خطی می‌شود. فرض کنید  $T$  رشته آموزشی  $M = [M_1, M_2, \dots, M_T]$  را داریم که در آن  $M_i = [C_{i1}, C_{i2}, \dots, C_{im_i}]$   $C_{ij}$ ، زمین خوش (زیرفضای خطی) از  $\alpha$  امین منیفولد غیرخطی و  $m_i$  تعداد زیرفضاهای خطی محلی در  $\alpha$  امین منیفولد است. همچنین  $C_{ij} = [x_{ij1}, x_{ij2}, \dots, x_{ijN_{ij}}]$  است که در آن  $N_{ij}$  تعداد نمونه‌های تصویر در داخل زمین زیرفضای خطی از  $\alpha$  امین منیفولد و  $x_{ijk} \in R^d$  نشان دهنده  $k$  امین نمونه تصویر از  $\alpha$  امین زیرفضای خطی در  $\alpha$  امین منیفولد است که به صورت یک بردار ستونی با ابعاد  $d$  می‌باشد ( $1 \leq i \leq T, 1 \leq j \leq m_i, 1 \leq k \leq N_{ij}$ )

همان‌طوری که در شکل (۱-الف) نشان داده شده است، خوشها با زوایای چهره مختلف از یک رشته ویدئویی دارای فاصله بیشتر از خوشها رشته‌های مختلف با زوایای یکسان است.

مراکز خوشها به عنوان قاب‌های کلیدی هر خوشه انتخاب و از افکنش OLPP برای نزدیکی مراکز خوشها داخل یک منیفولد و دوری خوشها می‌نمی‌توانند. استفاده می‌شود؛ سپس با افکنش به دست آمده تمام قاب‌های

ی و همکارانش چهره هر شخص را، توسط منیفولد ظاهر با ابعاد پایین در فضای تصویر نشان دادند؛ سپس منیفولد ظاهر غیرخطی و پیچیده را با منیفولد زوایای چهره و ارتباطات بین آن‌ها نمایش دادند (لی و همکاران، ۲۰۰۳). به عبارت دیگر هر منیفولد زاویه چهره را، توسط یک صفحه شباهت تقریب زندند. برای ساختن این طرز نمایش، نمونه‌هایی از ویدئو، توسط الگوریتم K-means شد. هر خوش به عنوان یک صفحه، با استفاده از<sup>۱</sup> PCA محاسبه و نمایش داده شد. ارتباطات بین منیفولد زوایا، توسط احتمال انتقال بین تصاویر در هر منیفولد زاویه چهره، مدل سازی شده و توسط رشته‌های ویدئو، آموزش داده شد. وانگ و همکارانش با تبدیل منیفولدها به زیرفضاهای محاسبه فاصله بین آنها، شناسایی چهره را بر اساس مجموعه تصاویر انجام دادند (وانگ و همکاران، ۲۰۰۸). آنها در (وانگ و همکاران، ۲۰۱۱) نشان دادند که با استفاده از خاصیت بینایی انسان و فضای رنگی، نرخ شناسایی را می‌توان بهبود بخشید؛ سپس در (وانگ و همکاران، ۲۰۱۲) مجموعه تصاویر را با ماتریس کوواریانس مدل سازی کرده و یک روش جداسازی بهینه بر اساس این ماتریس پیشنهاد دادند. در (هو و همکاران، ۲۰۱۱) هو و همکارانش مجموعه تصاویر را توسط نمونه‌های هر مجموعه و مدل آفین هال<sup>۲</sup> نمایش داده و برای به دست آوردن فاصله بین مجموعه‌ها از یک روشی به نام<sup>۳</sup> SANP که براساس نزدیک‌ترین نقطه است، استفاده کردند. هو و همکارانش یک روش مقایسه‌ای بر اساس نمونه‌های اشخاص برای مقایسه بین مجموعه تصاویر ارائه کردند (مو و همکاران، ۲۰۱۳) که در آن برای هر منیفولد محلی از اطلاعات جداساز زیادی استفاده شده است. در (لین و اکسای، ۲۰۱۲) لین و همکارش از افکنش OLPP برای شناسایی چهره در تصاویر ثابت استفاده کردند. آنها به منظور مقاوم‌بودن شناسایی چهره به تغییرات روشنایی و حالات چهره، برای استخراج ویژگی‌های گابور از ویولت گابور با هشت جهت و پنج مقیاس متفاوت استفاده کردند.

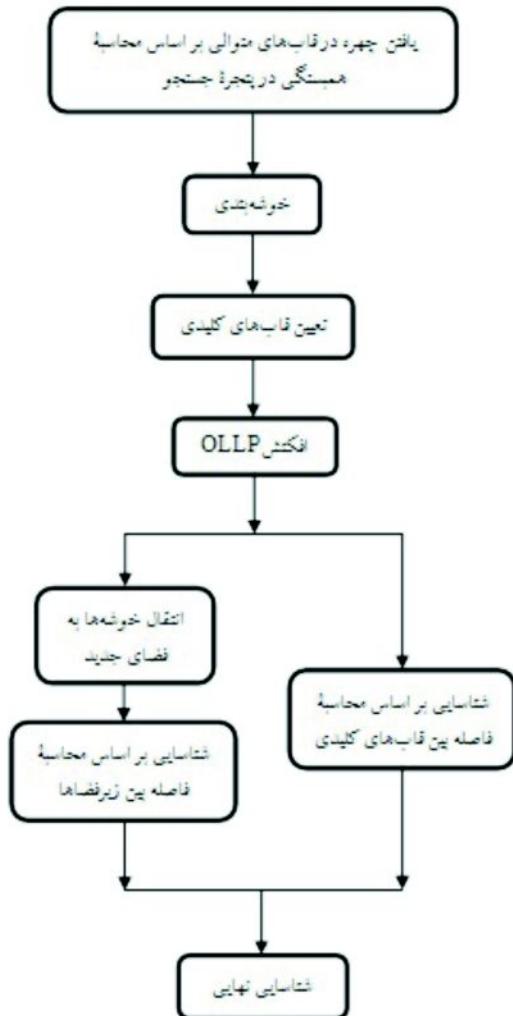
در ایران بیشتر پژوهش‌های انجام‌یافته در زمینه شناسایی چهره مربوط به تصاویر ثابت است. در (هرندی، ۲۰۰۹) هرندی الگوریتم پایه‌های محلی بهینه را ارائه کرده است. در این الگوریتم یک مجموعه نگاشت غیر ایزو متريک با هدف بيشينه کردن قدرت تفکيک‌پذيری طبقه‌های مجاور

<sup>1</sup> Principal Component Analysis

<sup>2</sup> Affine Hull

<sup>3</sup> Sparse Approximated Nearest Point

- ۲ خوشبندی رشته‌های ویدیویی
- ۳ افکنش OLLP
- ۴ شناسایی بر اساس قاب‌های کلیدی و زیرفضاهای

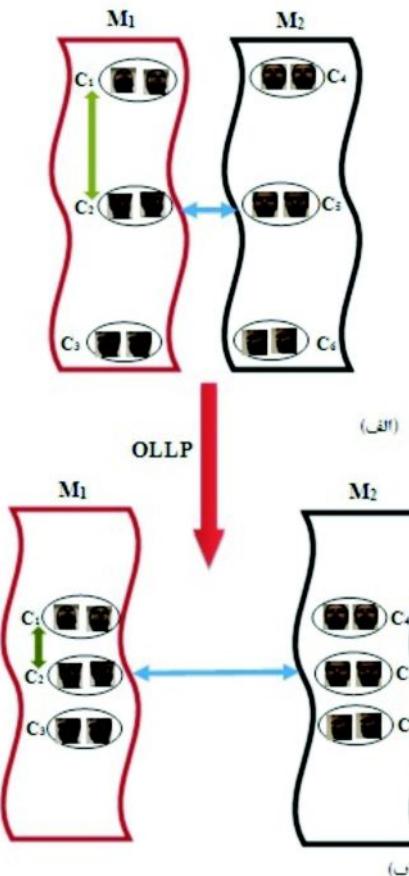


(شکل-۲) : نمودار جریانی الگوریتم پیشنهادی

### ۱-۳- تشخیص چهره در رشته‌های ویدئویی

در مباحث مربوط به شناسایی چهره بر اساس ویدئو تشخیص بلاذرنگ چهره اهمیت زیادی دارد. روش‌های تشخیص بلاذرنگ چهره، به دو گروه اصلی تقسیم‌بندی می‌شوند. روش Cascade AdaBoost که توسط ویولا و جونز پیشنهاد شده است (ویولا و جونز، ۲۰۰۱). گروه دوم از اطلاعات رنگ جهت تشخیص استفاده می‌کنند (یانگ و واپل، ۱۹۹۶).

خوشبندی به فضای جدید منتقل می‌شوند. علت استفاده از مراکز خوشبندی به جای تمام قاب‌های خوشبندی کاهش پیچیدگی سامانه است. شکل (۱- ب) خوشبندی منیفولدها را در فضای جدید نشان می‌دهد.



(شکل-۱) : (الف) فاصله بین خوشبندی یک منیفولد با زوایای چهره مختلف از فاصله بین خوشبندی منیفولدهای مختلف با زوایای چهره یکسان بیشتر است. (ب) با یک افکنش مناسب می‌توان خوشبندی یک منیفولد را به هم نزدیک تر و خوشبندی منیفولدهای مختلف را از هم دور کرد.

بعد از افکنش قاب‌های کلیدی و زیرفضاهای تمامی منیفولدها، برای بهبود دقیق شناسایی از ترکیب شناسایی بر اساس قاب‌های کلیدی و شناسایی بر اساس زیرفضاهای خطی استفاده می‌شود. شکل (۲) نمودار جریانی کلی الگوریتم پیشنهادی را نشان می‌دهد. در ادامه درخصوص مراحل الگوریتم پیشنهادی که شامل موارد زیر است، توضیح داده می‌شود.

- ۱- تشخیص چهره در رشته‌های ویدئویی

$$A^{(k-1)} = [o_1, o_2, \dots, o_{k-1}] \quad (1)$$

<sup>1</sup> Locality Preserving Projection

سال ۱۳۹۵ شماره ۲ پیاپی ۲۸

- خطی بودن مدل‌ها، دقت بالایی دارد.
- به علت موجود بودن اطلاعات بردارهای حرکت این روش بسیار ساده است.
- با توجه به تغییرات بردارهای حرکت جهت چرخش سر را می‌توان پیش‌بینی کرده و در هنگام محاسبه فاصله بین زیرفضاهای تعداد مقایسه‌ها را به میزان قابل توجهی کاهش داد.

### ۳-۳- نگاشت حفظ ساختار محلی متعامد

افکنش OLPP بر اساس روش<sup>1</sup> LPP است که در آن تلاش بر افکنش داده ورودی به فضای جدید با حفظ ساختار هندسی داده‌ها است؛ با این تفاوت که در این روش پایه‌های افکنش متعامد هستند. در (کای و همکاران، ۲۰۰۶) اثبات شده است که OLPP دارای قدرت جداسازی بالاتری نسبت به LPP است. برای مجموعه قاب‌های کلیدی در منیفولد‌های غیرخطی  $\{x_{11}, x_{12}, \dots, x_{Tm_1}, x_{T1}, x_{T2}, \dots, x_{Tm_T}\}$  که در آن  $\{x_{11}, x_{12}, \dots, x_{lm_l}\}$  فیژه‌های کلیدی  $l$  امین منیفولد است.

الگوریتم تولید OLPP (کای و همکاران، ۲۰۰۶)

به صورت زیر است:

- ۱- ابتدا مجموعه تصاویر ورودی به زیرفضای PCA انتقال داده می‌شوند که در آن از اجزای مطابق با مقادیر ویرثه  $W_{PCA}$  صفر، صرف نظر می‌شود. این ماتریس انتقال، با نمایش داده می‌شود. با انتقال داده‌ها به این زیرفضا، داده‌ها ناهمبسته می‌شوند. مهم‌ترین مزیت انتقال داده‌ها به این زیرفضا کاهش حجم آن‌ها و درنتیجه ذخیره‌سازی اطلاعات با حجم کمتر امکان‌پذیر می‌شود.
- ۲- گراف  $G$  مطابق با تعداد تصاویر چهره ورودی در نظر گرفته می‌شود که در آن ماتریس شباهت برای دو قاب کلیدی تنها زمانی که هر دو قاب متعلق به یک منیفولد یکسان باشند، برابر یک است.
- ۳- ماتریس قطری  $D$  و ماتریس لاپلاسین  $L$  به صورت  $L = D - S$  و  $D_{ii} = \sum_j S_{ji}$  به دست آوردن بردار پایه‌های متعامد  $A^{(k-1)}$  و  $B^{(k-1)}$  است.  $W_{OLPP} = \{o_1, o_2, \dots, o_k\}$  به صورت زیر تعریف می‌کنیم (کای و همکاران، ۲۰۰۶):

برای تعیین ناحیه دقیق چهره، محل چشم‌ها در نخستین قاب به دست آمده و ناحیه چهره بر اساس آن مشخص می‌شود. برای تعیین محل چشم‌ها روش‌های مختلفی ارائه شده است. مهم‌ترین این روش‌ها شامل روش تطبیق الگو، روش فضای ویرثه و استفاده از تبدیل هاف است. برای تعیین محل چهره در قاب‌های متوالی دیگر از این واقعیت که در یک ویدئو، قاب‌های متوالی دارای همبستگی بالا هستند استفاده می‌شود. مکان چهره در دومین قاب، با محاسبه همبستگی محل چهره در نخستین قاب با قاب بعدی در حوالی ناحیه مشخص شده به عنوان چهره و تعیین محل بیشترین همبستگی، به دست می‌آید. این روش به صورت تکراری برای کل قاب‌ها انجام می‌گیرد. تعیین اندازه پنجره جستجو نقش اساسی در میزان محاسبات دارد.

### ۲-۳- خوشبندی قاب‌ها

جهت ساختن مدل خطی محلی از منیفولد غیرخطی، چندین روش پیشنهاد شده است. بیشترین آن‌ها از روش خوشبندی بر اساس تکرار (به طور مثال k-means) برای نسبت دادن نمونه‌ها به خوشبندی مختلف استفاده می‌کنند. این روش‌ها دو محدودیت اساسی دارند. نخست این که در ابتدای اجرای الگوریتم بایستی تعداد خوشبندی به صورت دستی مشخص شوند. دوم این که خاصیت خطی بودن مدل‌های محلی، قطعی نیست. در این مقاله یک الگوریتم کارا برای خوشبندی ویدئوها پیشنهاد می‌شود؛ بدین صورت که برای تبدیل منیفولد غیرخطی به مدل‌های خطی از اطلاعات موجود در یافتن چهره در قاب‌های متوالی در رشته ویدئویی استفاده می‌شود. به عبارت دیگر با توجه به محل چهره یافته شده در داخل پنجره جستجو، بردارهای حرکت چهره به دست می‌آیند که با استفاده از تحلیل این بردارهای حرکت در قاب‌های متوالی می‌توان حرکت سر را پیش‌بینی کرده و خوشبندی قاب‌ها را انجام داد. به عنوان مثال با حرکت سر به سمت راست بردارهای حرکت نیز به سمت راست حرکت می‌کنند که قاب‌های مربوطه را می‌توان به عنوان یک خوشبندی قاب‌ها را انجام داد. زیر است:

- تعیین تعداد خوشبندی و مدل‌های خطی به طور کامل هوشمندانه است و با اجرای این روش با توجه به حرکات سر، تعداد خوشبندی متفاوت برای هر رشته ویدئویی به دست می‌آید.

$$P_1^T P_2 = Q_{12} \Lambda Q_{21}^T \quad (4) \quad (\text{وانگ و همکاران}, ۲۰۰۸)$$

$$\Lambda = \text{diag}(\sigma_1, \dots, \sigma_r)$$

که در آن  $Q_{12}$  و  $Q_{21}$  ماتریس‌های متعامد هستند.  
مقادیر منفرد  $\sigma_1, \dots, \sigma_r$  کسینوس زوایای اصلی است؛ به عبارت دیگر همبستگی اصلی به صورت زیر به دست می‌آید:

$$\cos(\theta_k) = \sigma_k \quad k = 1, 2, \dots, r \quad (5)$$

$V = P_2 Q_{21} = [v_1, \dots, v_{d_2}]$  بردارهای  
اصلی می‌باشند.

از این زوایا برای محاسبه فاصله بین زیرفضاهای  
شیوه‌های گوناگون استفاده شده است که برای نمونه  
روش‌های زیر را می‌توان نام برد:

(6) (ادلمن و همکاران, ۱۹۹۹)

$$D_P(C_1, C_2) = \left( \sum_i \sin^2 \theta_i \right)^{\frac{1}{2}} \quad (7) \quad (\text{ولف و شاشیو, ۲۰۰۳})$$

$$D_{BC}(C_1, C_2) = \left( 1 - \prod_i \cos^2 \theta_i \right)^{\frac{1}{2}} \quad (8) \quad (\text{چکیوز, ۲۰۰۳})$$

با توجه به این که در بین این زوایای اصلی، زاویه نخست از اهمیت بالایی دارد، است لذا در این مقاله طبق بررسی به عمل آمده، از کسینوس زاویه نخست به عنوان معیار ارزیابی فاصله بین زیرفضاهای استفاده شده است. این معیار با نماد  $d_V(C_{ij}, C_{kl})$  نشان داده می‌شود.

در چندین روش دیگر ارائه شده، نمونه‌های منیفولد جهت اندازه‌گیری شباهت مدل‌ها استفاده شده است. این روش‌ها به عنوان روش اندازه‌گیری بر اساس نمونه شناخته شده می‌باشند. به عبارت دیگر در این روش فاصله بین دو منیفولد غیر خطی توسط کوچکترین فاصله بین قاب‌های کلیدی محاسبه می‌شود. این معیار با نماد  $d_E(C_{ij}, C_{kl})$  نشان داده می‌شود. بهتر است برای به دست آوردن اندازه‌گیری کامل فاصله هر دو معیار را در نظر بگیریم. به عبارت دیگر ترکیب دو معیار مذکور را می‌توان به صورت زیر نوشت:

$$B^{(k-1)} = [A^{(k-1)}]^T (XDX^T)^{-1} A^{(k-1)} \quad (2)$$

که در آن پایه‌های متعامد  $\{o_1, o_2, \dots, o_k\}$  را می‌توان به صورت زیر به دست آورد:

$$o_1 \text{ بردار ویژه مطابق با کوچکترین مقدار ویژه } (XDX^T)^{-1} XDX^T \\ - o_k \text{ بردار ویژه مطابق با کوچکترین مقدار ویژه } \{I - (XDX^T)^{-1} A^{(k-1)} [B^{(k-1)}]^{-1} [A^{(k-1)}]^T\} (XDX^T)^{-1} \\ . XDX^T \text{ محاسبه می‌شود.}$$

- ۴- ماتریس انتقال افق‌گش OLPP توسعه  $W = W_{PCA} W_{OLPP}$  محاسبه می‌شود.

### ۴-۳ شناسایی بر اساس قاب‌های کلیدی و زیرفضاهای

بعد از به دست آوردن ماتریس افق‌گش OLLP تمامی قاب‌های موجود در رشته‌های ویدئویی به فضای جدید انتقال می‌یابند. برای بالا بردن دقیق شناسایی از ترکیب زوایایی بر اساس قاب‌های کلیدی و شناسایی بر اساس زوایایی استفاده می‌شود. در روش‌هایی که برای شناسایی از زوایایی اصلی استفاده می‌کنند، به طور اساسی مدهای مشترک تغییرات بین دو زیرفضا منعکس می‌شوند و اطلاعات موجود در هر زیرفضا نادیده گرفته می‌شود. این روش به عنوان اندازه‌گیری بر اساس تغییرات معروف است. استفاده از مفهوم زوایایی اصلی جهت محاسبه فاصله بین زیرفضاهای برای نخستین بار توسط هاتلینگ مطرح شده است (Hatteling, ۱۹۳۶).

زوایایی اصلی  $0 \leq \theta_1 \leq \theta_2 \leq \dots \leq \theta_r \leq (\pi/2)$  بین دو زیرفضای خطی  $C_1$  و  $C_2$  به صورت زیر تعریف می‌شود:

$$\cos(\theta_k) = \max_{u_k \in C_1} \max_{v_k \in C_2} u_k^T v_k \quad (3) \quad (\text{وانگ و همکاران}, ۲۰۰۸)$$

$$u_k^T u_k = v_k^T v_k = 1, u_k^T u_i = v_k^T v_i = 0 \\ i = 1, 2, \dots, k-1$$

که در آن  $(C_1, C_2)$  است  $r = \min(\dim(C_1), \dim(C_2))$  و  $u_k, v_k$  امین جفت از بردارهای اصلی است. کسینوس‌های زوایایی اصلی، همبستگی‌های اصلی نامیده می‌شوند. الگوریتم محاسبه زوایای اصلی، بر اساس SVD<sup>۱</sup> ارائه شده است. اگر  $P_1 \in R^{D \times d_1}$  و  $P_2 \in R^{D \times d_2}$  به ترتیب نشان‌دهنده پایه‌های متعامد دو زیرفضای  $C_1$  و  $C_2$  باشد و  $d_1$  و  $d_2$  ابعاد زیرفضاهای باشند، SVD<sup>۲</sup> به صورت زیر محاسبه می‌شود:

فصل نهم



<sup>۱</sup> Singular Value Decomposition

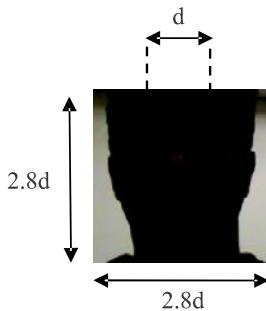
سال ۱۳۹۵ شماره ۲ پیاپی ۲۸

بعد از تشخیص چهره و چشم، ناحیه چهره توسط مربعی با ابعاد  $2.8d \times 2.8d$  مشخص می‌شود که در آن  $d$  فاصله بین دو چشم است. شکل (۴) این ناحیه مستطیلی را در پایگاه داده Honda/UCSD نشان می‌دهد.



(شکل-۳) : چند نمونه از چهره‌های افراد در پایگاه داده‌های Honda/UCSD (الف) CMU MoBo (ب)

همان‌طور که در قبل توضیح داده شد در روش خوشبندی ارائه شده، از اطلاعات به دست آمده از موقعیت چهره در داخل پنجره جستجو استفاده می‌شود. پنجره جستجوی مورد استفاده در این شبیه‌سازی، دارای ابعاد  $10 \times 10$  پیکسل است. تغییرات در جهت افقی و عمودی نسبت به قاب قبلی به ترتیب با  $dx$  و  $dy$  نشان داده می‌شود. به عبارت بهتر  $dx$  و  $dy$  ابعاد بردارهای حرکت هستند.



(شکل-۴) : ناحیه چهره بر اساس فاصله بین دو چشم در پایگاه داده Honda/UCSD

شکل (۵) تغییرات  $dx$  و  $dy$  در فریم‌های متواالی برای دو رشته از ویدئوهای پایگاه داده Honda/UCSD را نشان می‌دهد. همان‌طوری که در شکل مشخص است، بردارهای حرکت مناسب با حرکت سر در جهت‌های مختلف حرکت می‌کنند با توجه به حرکت این بردارها، می‌توان خوشبندی مناسب را انجام داده و با وجود آوردن امکان مقایسه خوشبندی‌های معادل در دو رشته ویدئویی، دقیق

$$d(C_{ij}, C_{kl}) = (1 - \lambda).d_E(C_{ij}, C_{kl}) + \lambda.d_V(C_{ij}, C_{kl}) \quad (9)$$

در این روش از قاب‌های کلیدی هر خوش بعنوان نماینده هر خوش برای شناسایی استفاده می‌شود. برای تعیین قاب‌های کلیدی، بعد از به دست آوردن خوش‌ها، قابی را که در داخل هر خوش بیشترین تغییرات بردارهای حرکت در جهت افقی و عمودی را دارد، به عنوان قاب کلیدی آن خوش انتخاب می‌شود. این فریم‌ها به همراه نخستین قاب هر رشته به عنوان قاب‌های کلیدی آن رشته ویدئویی در نظر گرفته می‌شوند. با توجه به این که در الگوریتم پیشنهادی از نخستین قاب برای مشخص کردن موقعیت چشم‌ها و سپس مشخص کردن ناحیه چهره استفاده می‌شود، لذا این الگوریتم از لحاظ مکانی به انسداد نواحی چشم در نخستین قاب حساس است. از طرف دیگر با توجه به این که قاب‌های انسدادیافته در یک خوش قرار می‌گیرند، لذا روش ارائه شده زیاد به انسداد از نظر تعداد قاب حساس نیست.

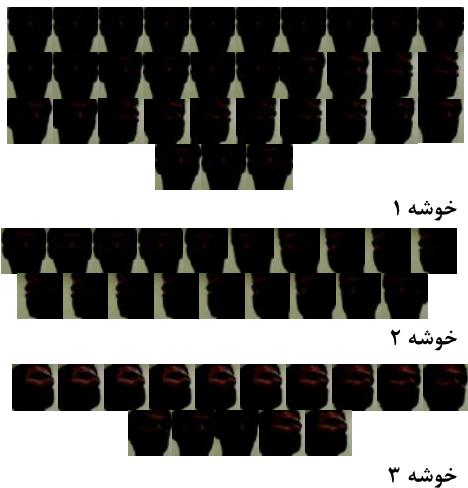
#### ۴- نتایج الگوریتم پیشنهادی

در این مقاله برای پیاده‌سازی الگوریتم پیشنهادی از دو پایگاه داده Honda/UCSD (الی و همکاران، ۲۰۰۳) و CMU MoBo (گراس و شی، ۲۰۰۱) که در پژوهش‌های شناسایی چهره در رشته‌های ویدئویی معروف هستند، استفاده شده است.

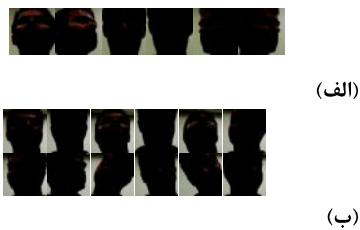
در پایگاه داده Honda/UCSD هر رشته ویدئویی در شرایط بیرونی با سرعت پانزده قاب در ثانیه ضبط شده که طول هر کدام از رشته‌ها حداقل پانزده ثانیه است. تفکیک پذیری هر رشته ویدئویی  $480 \times 480$  است. چهره هر شخص حداقل در دو رشته ویدئویی ضبط شده است. با توجه به این که تغییرات زاویه سر بزرگ‌ترین چالش در شناسایی چهره است، همه رشته‌ها شامل چرخش‌های سر دو بعدی و سه بعدی هستند. در هر ویدئو شخص سرش را مطابق میل خود در جهت و سرعت موردنظر می‌چرخاند.

پایگاه داده CMU MoBo شامل ۹۶ رشته از ۲۴ شخص متفاوت است که در حال قدمزدن هستند. از هر شخص چهار رشته ویدئویی توسط چند دوربین ضبط شده است.

شکل (۳) چند نمونه از چهره‌های افراد در پایگاه داده‌های CMU MoBo و Honda/UCSD را نشان می‌دهد.



(شکل-۶) : سه نمونه از خوشه‌های بهدستآمده با استفاده از تحلیل بردارهای حرکت



(الف)

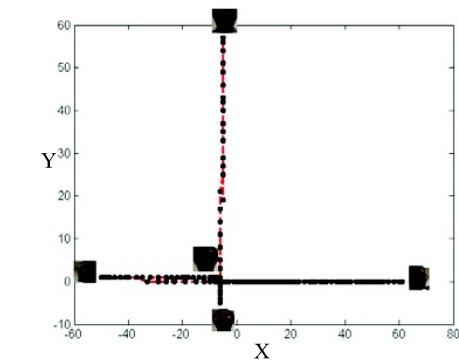
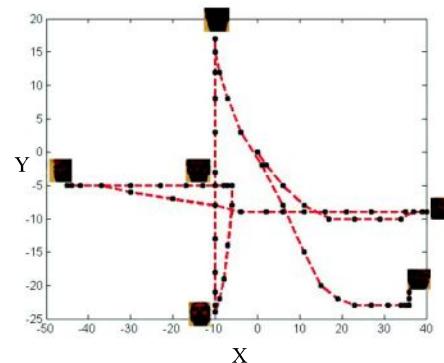


(ب)

(شکل-۷) : تصاویر بالا فریم‌های کلیدی برای دو رشته از ویدئوهای پایگاه داده Honda/UCSD را نشان می‌دهد.

- تطبیق نزدیک‌ترین همسایگی در (الف) Eigenface (بلهیمر و همکاران، ۱۹۹۷) و Fisherface
- تطبیق نزدیک‌ترین همسایگی در خوشبندی K-means<sup>۱</sup> و LLE<sup>۲</sup> (هدید و پیتیکاین، ۲۰۰۴).
- روش افکنش Schur-OLPP (لی و زیو، ۲۰۰۷). (در این روش لی و همکارش الگوریتم تاحدودی ساده‌تری برای بهدست آوردن بردارهای افکنش متعدد ارائه کردند).
- روش افکنش OLPP و ویولت گابور (Gabor-OLPP) (لین و اکسای، ۲۰۱۲).
- روش زیرفضای متقابل که یک نمونه از روش شناسایی بر اساس مقایسه زیرفضاهای است (یاماگوچی و همکاران، ۱۹۹۸).
- روش محاسبه فاصله بین منیفولدها (MMD) (وانگ و همکاران، ۲۰۰۸).

شناسایی را بهبود بخشدید. شکل (۶) سه نمونه از خوشه بهدست آمده را با روش مذکور نشان می‌دهد.



(شکل-۵) : تغییرات  $dx$  و  $dy$  در دو رشته از ویدئوهای پایگاه داده Honda/UCSD

شکل (۷) قاب‌های کلیدی را برای دو رشته از ویدئوهای پایگاه داده Honda/UCSD نشان می‌دهد. با توجه به این که ابعاد و روش‌نایی‌های چهره‌های بهدست آمده از رشته‌های ویدئویی متفاوت هستند، بنابراین چهره‌ها از نظر ابعاد هنجارسازی شده و از متعادل‌کننده هیستوگرام برای حذف اثرات تغییرات روش‌نایی استفاده می‌شود. مقدار ضریب ترکیبی دو روش شناسایی (۶) در رابطه (۹) به  $0/5$  تنظیم می‌شود.

الگوریتم پیشنهادی علاوه بر روش خوشبندی ارائه شده، با استفاده از خوشبندی K-means<sup>۱</sup> (K=۵) شبیه‌سازی شده و با شش روش زیر مقایسه می‌شود. چهار روش نخست از جمله روش‌های شناسایی براساس نمونه و دو روش بعدی از جمله روش‌های شناسایی بر اساس تغییرات است.

<sup>1</sup> locally linear embedding  
<sup>2</sup> Manifold to Manifold Distance



شده است. این الگوریتم از بردارهای حرکت به دست آمده از تشخیص چهره در داخل پنجره جستجو استفاده می‌کند. مرکز هر زیرفضای خطی (خوش) به عنوان نماینده آن زیرفضاهای انتخاب و با استفاده از افکنش OLPP این مراکز، به منظور نزدیکی قاب‌های متعلق به یک منیفولد و دوری قاب‌های متعلق به منیفولددهای مختلف، به فضای جدیدی منتقل می‌شوند. بعد از افکنش قاب‌های کلیدی و زیرفضاهای تمامی منیفولدها، برای بهبود دقیق شناسایی از ترکیب شناسایی بر اساس قاب‌های کلیدی و شناسایی بر اساس زیرفضاهای خطی استفاده می‌شود. درنهایت مقایسه نتایج به دست آمده با روش‌های موجود دیگر، نشان دهنده کارآیی بالای الگوریتم پیشنهادی است.

## ۶- مراجع

- Barr, J. R., Bowyer, K. W., Flynn, P. J., & Biswas, S. (2012). Face recognition from video: A review. *Journal of pattern Recognition and Artificial Intelligence*, 26(5).
- Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *PAMI*, 19(7), 711–720.
- Bozorgtabar, B., Rad, G. A. R. (2011). A Genetic Programming-PCA Hybrid Face Recognition Algorithm. *Journal of Signal and Information Processing*, (Vol. 2), 170-174.
- Cai, D., He, X., Han, J., & Zhang, H. J. (2006). Orthogonal Laplacianfaces for Face Recognition. *IEEE Trans. Image Proces* 15(11): 3608-3614.
- Cheng, J. M., Beveridge, J., Draper, B., Kirby, M., Kley, H., & Peterson, C. (2006). Illumination face spaces are idiosyncratic. *The International Conference on IPCVPR*.
- Chikuse, Y. (2003). Statistics on special manifolds, lecture notes in statistics. (Vol. 174), New York, Springer.
- Edelman, A., & Arias, T. A. Smith, S. T. (1999). The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl* (Vol. 20), 303-353.
- Fukui, K., & Yamaguchi, O. (2003). Face recognition using multi-viewpoint patterns for robot vision. *Int. Symp. Of Robotics Res.*, 192-201.
- Hadid, A., & Pietikäinen, M. (2004). From Still Image to Video-Based Face Recognition: An Experimental Analysis. *FG*, 813–818.

برای محاسبه نرخ متوسط شناسایی، برای هر شخص، از پایگاه داده مورد استفاده یک ویدئو جهت آموزش و بقیه رشته‌ها برای آزمایش مورد استفاده قرار می‌گیرند. همچنین آزمایش‌ها برای پنج ترکیب مختلف رشته‌های آموزشی و آزمایشی اجرا می‌شود. جدول (۱) مقایسه روش‌های مختلف را نشان می‌دهد. در این جدول برای ذکر مقادیر روش‌های MSM، Fisherface، LLE+K-means، Eigenface و MM از مرجع (وانگ و همکاران، ۲۰۰۸) استفاده شده است.

(جدول -۱) : نرخ متوسط شناسایی روش‌های مختلف شناسایی

در پایگاه داده‌های Honda/UCSD

روش‌های مورد استفاده	نرخ متوسط شناسایی در پایگاه داده Honda/UCSD	نرخ متوسط شناسایی در پایگاه داده CMU MoBo
Eigenface	۷۴/۲	۸۱/۰
Fisherface	۷۹/۲	۸۸/۳
LLE+K-means	۹۱/۸	۸۹/۸
MSM	۸۸/۲	۸۵/۱
MMD	۹۶/۹	۹۳/۶
Schur-OLPP	۹۴/۰	۹۰/۰
Gabor-OLPP	۹۵/۲	۹۱/۹
روش پیشنهادی با استفاده از K-means و OLPP	۹۳/۲	۹۱/۱
روش پیشنهادی با استفاده از الگوریتم خوشبندی ارائه شده و OLPP	۹۹/۰	۹۶/۹

با مقایسه روش پیشنهادی با استفاده از الگوریتم k-means خوشبندی ارائه شده و الگوریتم خوشبندی مشخص می‌شود که الگوریتم خوشبندی ارائه شده که از اطلاعات تشخیص چهره در قاب‌های متوالی رشتۀ ویدئویی استفاده می‌کند، نقش مهمی در بهبود نرخ شناسایی دارد.

## ۵- نتیجه‌گیری

در این مقاله یک الگوریتم کارا برای بهبود نرخ شناسایی و حل مشکلات ناشی از چرخش سر، تغییر شدت روش‌نایابی و پوشش قسمتی از چهره در رشته‌های ویدئویی با استفاده از افکنش OLPP ارائه شده است. ابتدا یک روش ساده برای استخراج زیرفضاهای خطی از هر رشتۀ ویدئویی توضیح داده

- Proc. Conf. Computer Vision and Pattern Recognition, 511- 518.
- Wang, R., Shan, S., Chen, X., & Gao, W. (2008). Manifold-Manifold Distance with Application to Face Recognition based on Image Set. computer vision and pattern recognition.
- Wang, S. J., Yang, J., Zhang, N., & Zhou, C. G. (2011). Tensor Discriminant Color Space for Face Recognition. IEEE Transactions on Image Processing, 20(9), 2490-2501.
- Wang, R., Guo, H., Davis, L. S., & Dai, Q. (2012). Covariance Discriminative Learning: A Natural and Efficient Approach to Image Set Classification. IEEE Conference on Computer Vision and Pattern.
- Wolf, L., & Shashua, A. (2003). Learning over sets using kernel principal angles. Journal of Machine Learning Research, (Vol. 4), 913–931.
- Wolf, L., & Shashua, A. (2003). Learning over sets using kernel principal angles. J. Mach. Learn. Res. (Vol. 4), 913-931.
- Yamaguchi, O., Fukui, K., & Maeda, K. (1998). Face Recognition Using Temporal Image Sequence. FG, 318–323.
- Yang, J., & Waibel, A. A real-time face tracker. (1996). In Proceedings of the Third IEEE Workshop on Applications of Computer Vision, Sarasota, FL, 142-147.
- Zhao, W., Chellappa, R., Phillips, P., & Rosenfeld, A. (2003). Face recognition: a literature survey. ACM Comput Surv, 35(4), 399–458.
- Gross, R., & Shi, J. (2001). The CMU Motion of Body (MoBo) database. Technical Report CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University.
- Harandi, M. T. (2009). Face recognition through optimal local basis learning. PhD thesis, University of Tehran.
- Hotelling, H. (1936). Relations between two sets of variates. Biometrika, (Vol. 28), 321–372.
- Hu, Y., Mian, A. S., & Owens, R. (2011). Sparse Approximated Nearest Points for Image Set Classification. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 121–128.
- Hu, Y., Mian, A. S., & Owens, R. (2012). Face Recognition Using Sparse Approximated Nearest Points between Image Sets. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(10), 1992–2004.
- Kim, T. K., Kittler, J., & Cipolla, R. (2007). Discriminative learning and recognition of image set classes using canonical correlations. PAMI, 29(6), 1–14.
- Lee, K. C., Ho, J., Yang, M. H., & Kriegman, D. (2003). Video-based face recognition using probabilistic appearance manifolds. Computer Vision and Pattern Recognition, IEEE Computer Society Conference, 313-320.
- Li, R., & Zhu, L. (2007). Face Recognition Based on an Alternative Formulation of Orthogonal LPP. IEEE International Conference on Control and Automation, Guangzhou, CHINA.
- Lin, G., & Xie, M. (2012). A Face Recognition Algorithm Using Gabor Wavelet and Orthogonal Locality Projection. IEEE ICCP2012 Proceeding.
- Mu, Y., Ding, W., & Tao, D. (2013). Local discriminative distance metrics ensemble learning. Pattern Recognition, 46(8), 2337–2349.
- Nishiyama, M., Yuasa, M., Shibata, T., Wakasugi, T., Kawa-hara, T., & Yamaguchi, O. (2007). Recognizing faces of moving people by hierarchical image-set matching. CVPR, Min-neapolis, MN.
- Ross, A. (2007). An Introduction to Multibiometrics. Proc. 15th European Signal Processing Conference (EUSIPCO).
- Sanderson, J.C., & Paliwal, K. K. (2004). On the Use of Speech and Face Information for Identity Verification, IDIAP.
- Shams-Baboli, A. A., Kaffashpour-Yazdi, M., Shams-Baboli, A., Araghi, S. (2013). Face Recognition with the Mixture of MDA and MPCA. Journal of Basic and Applied Scientific Research, 3(3), 1110-1117.
- Viola, P., & Jones, M. (2001). Rapid Object Detection Using a Boosted Cascade of Simple Features.



**سجاد شفیع پور یوردشاھی دوره کارشناسی خود را در دانشگاه ارومیه در رشته مهندسی برق-الکترونیک در سال ۱۳۷۲ پشت سر گذاشت و مدرک کارشناسی ارشد خود را در رشته برق- مخابرات در سال ۱۳۷۵ از دانشگاه فردوسی مشهد دریافت کرد. ایشان از سال ۱۳۸۸ دانشجوی دوره دکترای رشته برق-مخابرات دانشگاه تبریز هستند. زمینه‌های مورد علاقه ایشان پردازش تصویر، شناسایی الگو و ویدئو کدینگ است. نشانی رایانامه ایشان عبارت است از:**

**shafeipour@gmail.com**



**میرهادی سیدعربی تحصیلات خود را در مقطع کارشناسی مهندسی برق- الکترونیک در دانشگاه تبریز در سال ۱۳۷۲ به اتمام رساند. مقاطع کارشناسی ارشد و دکترای مهندسی برق-مخابرات را**

فصل نهم

سال ۱۳۹۵ شماره ۲ پیاپی ۲۸



به ترتیب در دانشگاه‌های خواجه نصیرالدین طوسی و تبریز در سال‌های ۱۳۷۵ و ۱۳۸۵ تکمیل کرد. ایشان از سال ۱۳۸۵ تاکنون عضو هیئت علمی دانشگاه تبریز بوده و با رتبه دانشیاری در دانشکده مهندسی برق و کامپیوتر مشغول به کار هستند. زمینه‌های پژوهشی مورد علاقه ایشان پردازش تصویر، بینایی ماشین، پویانمایی و مدل‌سازی سه‌بعدی چهره و ارتباط متقابل انسان با رایانه است.

نشانی رایانمۀ ایشان عبارت است از:

[seyedarabi@tabrizu.ac.ir](mailto:seyedarabi@tabrizu.ac.ir)



علی آق‌آگلزاده تحصیلات خود را در مقطع کارشناسی مهندسی برق در دانشگاه تبریز در سال ۱۳۶۴ به اتمام رساند. مقاطع کارشناسی ارشد و دکترای مهندسی برق را به ترتیب در دانشگاه‌های Illinois و Purdue در سال‌های ۱۳۶۷ و ۱۳۷۰ تکمیل کرد. ایشان در حال حاضر با رتبه استاد تمام در دانشگاه صنعتی نوشیروانی بابل مشغول به کار هستند. زمینه‌های پژوهشی مورد علاقه ایشان پردازش سیگنال و تصویر، بینایی ماشین، ویدئو کدینگ و سامانه‌های مخابرات دیجیتال است.

نشانی رایانمۀ ایشان عبارت است از:

[aghagol@nit.ac.ir](mailto:aghagol@nit.ac.ir)