

# افزایش سرعت آشکارسازی چهره در تصاویر ثابت با استفاده از هم‌جوشی داده‌های عمق و رنگ

علی سلمانی<sup>۱</sup> و مرتضی خادمی<sup>۲</sup>

## چکیده

پردازش چهره در اکثر کاربردهای بینایی ماشین موضوعی مهم به شمار می‌رود. این پردازش می‌تواند شامل مباحثی مثل آشکارسازی چهره، ردیابی چهره، شناخت حالات چهره و شناخت افراد شود. از میان این موارد، آشکارسازی چهره پایه‌ای‌ترین و کاربردی‌ترین شاخه پردازش چهره است. علت این موضوع، کاربردهای متفاوتی است که آشکارسازی چهره داراست. برای عملی کردن این کاربردها در ابتدا نیازمند یک الگوریتم سریع و دقیق برای آشکارسازی چهره می‌باشیم. روش‌های زیادی برای افزایش سرعت اجرای الگوریتم آشکارسازی چهره ارائه شده‌اند. اما معمولاً این روش‌ها دقت نهایی سیستم را کم می‌کنند. در سوی مقابل روش‌هایی که به دنبال افزایش دقت بوده‌اند، با تحمیل بار محاسباتی به سیستم، میزان سرعت را پایین آورده‌اند. در سال‌های اخیر با توجه به ارزان شدن و در دسترس عموم قرار گرفتن دوربین‌های دریافت عمق، امکان این که بتوان در یک دقت ثابت، سرعت الگوریتم را افزایش داد، فراهم شده است. در این تحقیق ما به دنبال ایجاد یک هم‌جوشی مناسب بین داده‌های عمق و رنگ برای غلبه بر مشکلات گذشته هستیم. بدین ترتیب که از ویژگی‌های داده‌های عمق به عنوان یک کاهنده فضای جستجو استفاده کرده تا بتوان سرعت مشخص‌سازی ناحیه چهره را در عین حفظ دقت، افزایش داد. نتایج شبیه‌سازی روش پیشنهادی نشان می‌دهد که با استفاده از این روش، سیستم آشکارسازی چهره با حفظ دقت، حدود ۲,۷۴ برابر سریع‌تر نسبت به الگوریتم ویولا جونز اجرا خواهد شد. این در حالی است که آخرین روش‌های همه‌جانبه موجود به حدود ۲,۵ برابر افزایش سرعت رسیده‌اند.

## کلید واژه‌ها

آشکارسازی چهره، هم‌جوشی داده‌های عمق و رنگ، کینکت

(قطعه‌بندی)، برای مثال لباس فرد و یقه لباس او را تشخیص می‌دهد، همچنین وجود چهره او را مشخص و مکان قرارگیری آن را تعیین می‌کند (آشکارسازی چهره<sup>۱</sup>)، می‌تواند او را شناسایی و هویت او را تعیین کند (شناخت چهره<sup>۲</sup>)، علاوه بر این می‌تواند حالت چهره او را تعیین کند (تشخیص حالات احساسی چهره<sup>۳</sup>) و در ادامه اگر این تصویر، اولین فریم از یک دنباله ویدیویی باشد، مکان چهره را در فریم‌های بعدی دنبال می‌کند (ردیابی چهره<sup>۴</sup>). اما سوال اساسی این است که یک الگوریتم بینایی ماشین غیرهوشمند

## ۱ مقدمه

پردازش چهره یکی از اساسی‌ترین موضوعات در حوزه الگوریتم‌های خودکار بینایی ماشین است. این پردازش در واقع بررسی همه‌جانبه یک چهره خواهد بود. یک تصویر دارای چهره، به خودی خود مجموعه‌ای از پیکسل‌هاست که با روشنایی‌های مختلف یک الگو را به وجود آورده‌اند. یک موجود هوشمند شبیه انسان، به راحتی ساختارهای آشنایی از آن استخراج می‌کند

این مقاله در شهریورماه ۱۳۹۴ دریافت، در دیماه بازنگری و در دیماه همان سال پذیرفته شد.

۱ گروه مهندسی برق، دانشکده مهندسی، دانشگاه فردوسی مشهد  
رایانامه: [ali.salmani@stu.um.ac.ir](mailto:ali.salmani@stu.um.ac.ir)

۲ گروه مهندسی برق، دانشکده مهندسی، دانشگاه فردوسی مشهد  
رایانامه: [khademi@um.ac.ir](mailto:khademi@um.ac.ir)

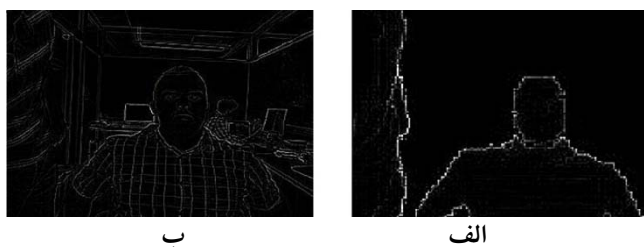
<sup>۱</sup> face detection

<sup>۲</sup> face recognition

<sup>۳</sup> facial expression recognition

<sup>۴</sup> face tracking

برای کنسول بازی xbox عرضه شد، اما به تدریج محققین بینایی ماشین به خاطر قابلیت‌ها و مزایای خوب آن سعی کردند تا از آن برای اخذ اطلاعات عمق استفاده کنند. در این تحقیق از تصاویر عمق و رنگ این دوربین به عنوان داده ورودی استفاده خواهد شد. ویژگی‌های تصویر عمق به طور خلاصه عبارتند از: الف: داده‌های عمق در برابر تغییرات روشنایی نور محیط، بدون تغییر هستند. این ویژگی امتیازی مهم برای آشکارسازی چهره‌ها در محیط‌هایی است که تغییرات روشنایی در آن‌جا زیاد است. ب: تصویر عمق به دلیل اینکه تنها از مقادیر شدت‌هایی تشکیل شده که فاصله فرد تا دوربین را مشخص می‌کند، می‌تواند تمایز بالایی را در درک صحنه ایجاد کند. مشخص است چنین ویژگی‌ای در تصویر رنگ نیست چرا که تصویر رنگ اولاً برای تمام نقاط ارزش یکسانی قائل است و ثانیاً از پیکسل‌هایی با رنگ‌ها و نورپردازی‌های متفاوت ایجاد شده است. ج: تصاویر عمق با سنجش فاصله اجسام تا دوربین می‌توانند به حذف اهداف غیرضروری در آشکارسازی کمک کرده تا مکان‌ها و اهداف مهم تصویر به راحتی استخراج شود. د: تصویر عمق ساختار همواری دارد. به عنوان شاهدی بر این ادعا می‌توان شکل ۱ را بررسی کرد. در شکل ۱-ب مشتق تصویر رنگ یک صحنه مشاهده می‌شود. به علت اینکه تصویر رنگ دارای شدت همواری نیست لبه‌های زیادی آشکار شده است. بالعکس در شکل ۱-الف، مشتق تصویر عمق همان صحنه نشان داده شده است. همانطور که ملاحظه می‌گردد به علت همواری بیشتر بین شدت‌ها در تصویر عمق، مشتق تصویر به سمت صفر میل کرده است.



شکل ۱: مشتق تصویر یک صحنه، (الف) تصویر عمق، (ب) تصویر رنگ

### ۳ مرور کارهای گذشته

روش‌ها و ایده‌های متفاوتی برای حل مساله آشکارسازی چهره ارائه شده است. در میان این روش‌ها، می‌توان الگوریتم آشکارسازی چهره ویولاجونز [۲] را به جرات یکی از دقیق‌ترین و سریع‌ترین الگوریتم‌های ارائه شده برای آشکارسازی چهره دانست. بلوک‌دیگرام الگوریتم ویولاجونز در مرحله آزمایش، در شکل ۲ نشان داده شده است. هدف از مرحله آزمایش این است که از طبقه‌بند آموزش دیده، برای جدا کردن تصاویر چهره‌ها از غیر چهره‌ها استفاده شود. اما باید توجه داشت که در یک تصویر آزمایشی ممکن است مکان قرارگیری چهره‌ها و هم‌چنین اندازه

چگونه قطعه‌بندی، آشکارسازی چهره، شناخت چهره، شناخت حالات احساسی چهره و ردیابی چهره را به خوبی یک انسان انجام دهد؟!.. از میان تمام جنبه‌های پردازش چهره که در بند قبل معرفی شدند، موضوع آشکارسازی چهره امری فوق‌العاده مهم است زیرا یک الگوریتم آشکارسازی چهره اولین گام برای ردیابی چهره، شناخت چهره و تشخیص حالات احساسی چهره است. به طور مشخص، یک آشکارساز چهره، یعنی سیستمی که قادر است در یک تصویر مشخص کند که آیا چهره انسانی در این تصویر وجود دارد یا خیر، و اگر وجود دارد محل آن را مشخص کند. در طراحی یک سیستم آشکارسازی چهره، سیستمی مطلوب است که زمان تشخیص چهره توسط آن کم و البته دقت مناسبی هم داشته باشد. برای رسیدن به هدف طراحی یک سیستم آشکارسازی سریع و دقیق، چالش‌هایی وجود دارد. برای مثال می‌توان به وجود عینک در برخی چهره‌ها، حالات احساسی متفاوت چهره، تغییرات روشنایی<sup>۱</sup>، تمام رخ یا نیم رخ بودن، رنگ پوست متفاوت افراد، انسداد<sup>۲</sup> و... اشاره کرد. تمام الگوریتم‌های آشکارسازی چهره به طور معمول در برابر برخی از این مشکلات سکوت کرده و راه‌حلی ارائه نمی‌دهند. هدف ما در این تحقیق استفاده از داده‌های عمق برای بهبود سرعت اجرای الگوریتم آشکارسازی چهره است. از این سو پس از استخراج برداری از تصویر عمق محیط، با توجه به مساله تکرار در طبیعت، ویژگی‌هایی از آن استخراج و از روی این ویژگی‌ها به دنبال استخراج محتوا و توصیفی از صحنه هستیم. بدیهی است که اگر بدانیم در یک تصویر چند نفر قرار گرفته‌اند و فاصله آن‌ها از دوربین چقدر است و... راحت‌تر می‌توان الگوریتمی هوشمندانه‌تر برای اجرای الگوریتم آشکارسازی چهره ساخت. در بخش دوم این مقاله، داده‌های عمق معرفی خواهند شد. در بخش سوم کارهای گذشته بررسی خواهند شد. در بخش چهارم، روش پیشنهادی این مقاله معرفی شده و نتایج شبیه‌سازی برای بررسی صحت عملکرد آن در بخش پنجم قرار می‌گیرد.

### ۲ داده‌های عمق

بدیهی است که هرچه اطلاعاتمان از صحنه افزایش یابد به سمت آشکارسازی دقیق‌تر و سریع‌تر پیش خواهیم رفت. یکی از راه‌های رسیدن به افزایش اطلاعات، استفاده از داده‌های عمق است. در سال‌های اخیر با توجه به ارزان شدن و در دسترس عموم قرار گرفتن دوربین‌های دریافت اطلاعات عمق صحنه، امکان استفاده از ویژگی‌های ممتاز اطلاعات عمق، برای کاربردهای متفاوت فراهم شده است. یکی از معروفترین دوربین‌های دریافت اطلاعات عمق صحنه، دوربین کینکت<sup>۳</sup> است. این دوربین توسط شرکت مایکروسافت، در ابتدا برای کاربردهای سرگرمی و به طور خاص

<sup>۱</sup> illumination

<sup>۲</sup> occlusion

<sup>۳</sup> Kinect

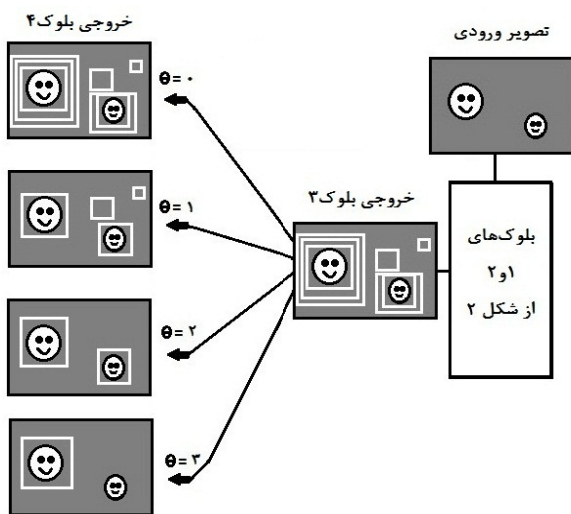
<sup>۴</sup> discrimination

<sup>۵</sup> edge

هر زیرتصویر تولیدی توسط بلوک ۱، وارد طبقه‌بندهای آموزش دیده (بلوک ۲) می‌شود و برای چهره یا غیرچهره بودن هر زیرتصویر، طبقه‌بندها تصمیم‌گیری می‌کنند. بدیهی است تعیین چهره بودن یا چهره نبودن، براساس آموزشی است که طبقه‌بند دیده است. در الگوریتم ویولاجونز برای اینکه ناحیه داخل یک پنجره بتواند به عنوان چهره یک فرد انتخاب شود، باید از تعدادی طبقه‌بند (و نه یک طبقه‌بند) که به صورت متوالی پشت سر هم قرار گرفته‌اند، عبور کند. اگر در هر مرحله مشخص شود که ویژگی استخراج شده مربوط به یک چهره نیست، زیرتصویر قرار گرفته زیرپنجره به عنوان تصویر غیرچهره تلقی می‌شود. در مقابل برای اینکه ناحیه به عنوان چهره یک فرد آشکار شود، باید از تمام طبقه‌بندها با موفقیت عبور کند. ایجاد چنین ساختاری به افزایش سرعت و دقت الگوریتم پیشنهادی ویولاجونز کمک می‌کند.

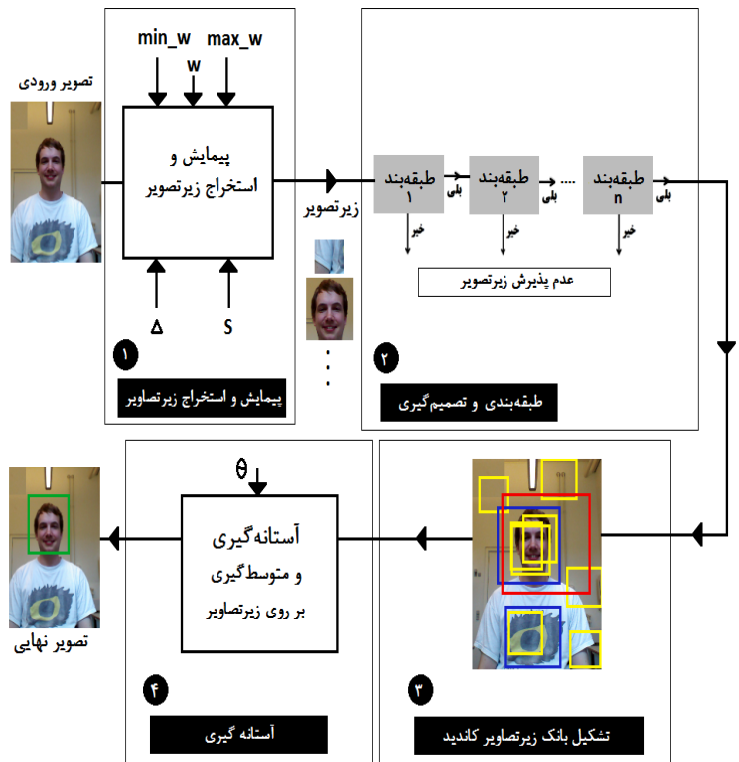
در ادامه، یک بانک اطلاعاتی از آن دسته از زیرتصاویری که با موفقیت از طبقه‌بندها عبور کرده‌اند، تشکیل می‌شود (بلوک ۳). این زیرتصاویر در واقع مکان‌های کاندید چهره بوده که در ابعاد مختلف بدست آمده‌اند. این مکان‌ها، به ورودی بخش آستانه‌گیری (بلوک ۴) اعمال می‌شود. در این بلوک باید با استفاده از عددی با نام آستانه  $\theta$ ، مکان قرارگیری چهره‌ی نهایی را به طور صحیح از بین مکان‌های کاندید انتخاب کرد. وقتی مقدار این پارامتر برابر صفر انتخاب شود، تمام مکان‌های کاندید در تصویر نهایی وجود خواهند داشت (به عبارتی می‌توان بلوک ۴ را حذف کرد).

برای درک بهتر چگونگی استفاده از این پارامتر در الگوریتم آشکارسازی چهره ویولاجونز، باید به شکل ۳ توجه کرد. در این شکل تصویر خروجی بلوک ۳، مکان‌های کاندید آشکارشده پس از اتمام مراحل پیمایش و جست‌وجو و تصمیم‌گیری را نشان می‌دهد.



شکل ۳: نحوه‌ی عملکرد  $\theta$  بر روی آشکارسازی

آن‌ها متفاوت باشد. برای این که بتوان در چنین تصویری تمام چهره‌ها را به طور صحیح آشکار کرد باید این تصویر را به شکلی مناسب برای ورودی طبقه‌بند، آماده کرد. این کار در بلوک پیمایش و استخراج زیرتصویر<sup>۱</sup> (بلوک ۱) انجام می‌گیرد. در این بلوک با استفاده از یک پنجره با ابعاد  $w * w$  (اندازه  $w$  بین دو مقدار  $min\_w$  و  $max\_w$  محدود می‌شود) ناحیه‌های مختلف تصویر ورودی جدا می‌گردد. این ناحیه‌ها را زیرتصویر می‌نامیم. برای جدا کردن زیرتصاویر در تمام تصویر، ابتدا پنجره با اندازه  $min\_w$  در نقطه  $(0,0)$  تصویر ورودی قرار می‌گیرد و زیرتصویر قرار گرفته تحت پنجره را جدا می‌کند. سپس به اندازه  $\Delta$  پیکسل، به جلو حرکت کرده و زیرتصویر دیگری را جدا می‌کند. بنابراین میزان حرکت پنجره جست‌وجو از چپ به راست و از بالا به پایین بر حسب پیکسل و به میزان  $\Delta$  خواهد بود. بعد از این که پنجره یک بار کامل تمام تصویر ورودی را پیمود، اندازه پنجره با ضریبی با نام  $S$  بزرگتر شده و پنجره جدید بار دیگر روی تصویر ورودی به دنبال چهره‌ها می‌گردد. این عمل بزرگ‌شدن تا جایی ادامه می‌یابد که ابعاد پنجره جست‌وجو برابر با ابعاد  $max\_w$  شود. به طور معمول در الگوریتم ویولاجونز و در حالت پیش‌فرض، اندازه  $min\_w$  برابر با ۲۴ (ابعاد تصاویر آموزشی) است. اندازه پیش‌فرض  $max\_w$  نیز برابر با اندازه تصویر ورودی برای آشکارسازی است. مقدار پیش‌فرض  $\Delta$  برابر یک پیکسل است. اندازه پیش‌فرض  $S$  نیز برابر ۱٫۱ است.

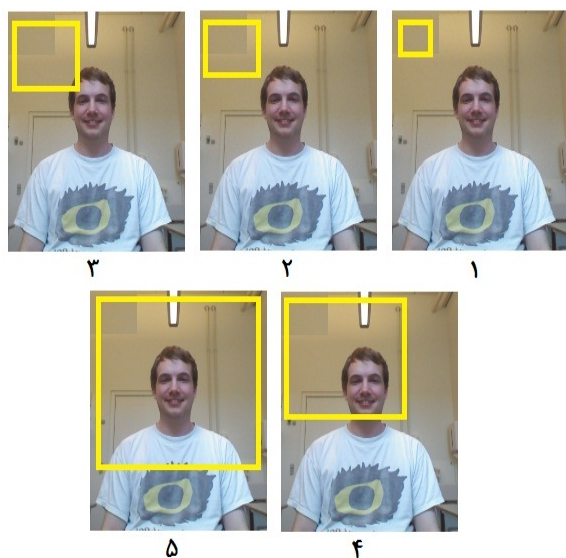


شکل ۲: طرح کلی الگوریتم ویولاجونز

<sup>۱</sup> Sub-image

**میزان حرکت پنجره ( $\Delta$ ):** این عامل نقش مهمی در سرعت اجرای الگوریتم آشکارسازی چهره خواهد داشت. با افزایش اندازه این عامل سرعت افزایش یافته ولی با توجه به نوع تصویر ورودی ممکن است دقت نهایی سیستم کاهش یابد. اگر از میزان پیچیدگی اجسام و اهداف موجود در تصویر اطلاعاتی داشته باشیم، می‌توان در نواحی‌ای از تصویر که دارای پیچیدگی کمتری است اندازه این عامل را افزایش داد. البته افزایش این عامل نباید باعث کاهش دقت سیستم شود.

**ابعاد پنجره جست‌وجو ( $W$ ):** هرچه محدودیت بیشتری بر روی این پنجره قایل شویم، یعنی اندازه‌های غیرمعمول را برای صورت فرد حذف کنیم، زمان آشکارسازی کاهش خواهد یافت. فرض کنید در شکل ۴، برای آشکارسازی چهره فرد، نیاز به ۵ مرحله پیمایش و جست‌وجو است. در مرحله اول  $w=24$  است. در مرحله بعد اندازه  $w$ ،  $s$  برابر می‌شود. این روند تا جایی ادامه می‌یابد که ابعاد پنجره جست‌وجو برابر با عرض تصویر ورودی شود (مرحله ۵). در همین فرایند جست‌وجو، اگر از اندازه چهره مورد جست‌وجو در تصویر اطلاعاتی داشته باشیم آنگاه می‌توان برخی از اندازه‌های دور از انتظار را حذف کرد. مثلاً در شکل ۴ می‌توان با اطلاعاتی که از اندازه چهره فرد داریم، پیمایش و جست‌وجو را برای مراحل ۱ و ۴ و ۵ انجام ندهیم. با این عمل می‌توان از جست‌وجوهای غیرضروری جلوگیری کرده، زمان آشکارسازی چهره را کاهش داد.



شکل ۴: امکان حذف بعضی از ابعاد پنجره نامتناسب با اندازه چهره

البته علاوه بر ۴ عامل بالا، عوامل دیگری نظیر ابعاد و بافت تصویر ورودی نیز بر روی سرعت آشکارسازی موثر است. بدیهی است که این ۲ عامل تحت کنترل الگوریتم آشکارسازی نیست و در این تحقیق برای ورودی الگوریتم پیشنهادی و الگوریتم ویولاجونز، ثابت در نظر گرفته می‌شوند.

تصاویر ستون سمت چپ (خروجی بلوک ۴) نیز مکان‌هایی را نشان می‌دهد که پس از آستانه‌گیری، به ازای چند مقدار  $\theta$ ، به عنوان چهره‌های نهایی آشکار شده‌اند. فرض می‌کنیم که تصویری بعد از عبور از بلوک‌های ۱ و ۲ از شکل ۲، دارای ۷ ناحیه کاندید به عنوان چهره است (خروجی بلوک ۳). در سطر اول، مقدار  $\theta$  برابر صفر انتخاب شده است. بدین ترتیب تمام مکان‌های کاندید (۷ مکان) در تصویر نهایی به عنوان چهره مشخص می‌شوند. در سطر دوم، مقدار آستانه برابر با یک است. این بدین معنی است که برای انتخاب ناحیه‌ای به عنوان چهره نهایی، باید حداقل یک بار آن ناحیه، کاندید شده باشد، و اگر بیشتر از یک بار آشکار شده باشد، می‌بایست با متوسط‌گیری بر روی چند مربع، یک مربع نهایی به عنوان ناحیه هدف انتخاب کرد. به همین ترتیب، در حالت  $\theta=2$  باید هر ناحیه برای اینکه به عنوان هدف برگزیده شود، حداقل دو بار کاندید شده باشد، که در شکل ۳ فقط برای دو ناحیه چهره چنین اتفاقی افتاده است. این همان حالت ایده‌آل است. اما در حالت  $\theta=3$ ، چون ناحیه‌ی چهره کوچکتر تنها دوبار کاندید شده است، پس از شرط عبور نکرده و آشکار نمی‌شود و دقت سیستم کاهش می‌یابد. لازم به ذکر است که مقدار پیش فرض  $\theta$ ، در الگوریتم ویولاجونز برابر ۴ است.

### ۳-۱ بررسی تاثیر عوامل مختلف بر روی سرعت و دقت الگوریتم ویولاجونز

با توجه به توضیحاتی که در مورد الگوریتم ویولاجونز داده شد، ۴ عامل به شرح زیر در دقت و سرعت الگوریتم ویولاجونز دخیل‌اند. هر ایده‌ای که به دنبال افزایش سرعت یا دقت الگوریتم ویولاجونز باشد باید به نحوی بر روی بعضی از این عوامل تاثیرگذار باشد.

**آستانه تشخیص ( $\theta$ ):** این عامل، ادغام‌کننده مکان‌های کاندید چهره در اندازه‌های مختلف تصویر است. با توجه به توضیحات داده شده، مقدار  $\theta$ ، بیش از آنکه بر روی سرعت اجرای الگوریتم موثر باشد، بر روی میزان دقت سیستم موثر است. به عبارت دیگر، اگر اندازه  $\theta$ ، بیشتر از اندازه مناسب انتخاب شود ممکن است ناحیه چهره نیز آشکار نشود و بالعکس، اگر کمتر انتخاب گردد ممکن است علاوه بر ناحیه چهره، نواحی دیگری نیز به اشتباه به عنوان چهره انتخاب شوند.

**مقیاس ( $S$ ):** همانطور که گفته شد، پارامتر  $s$  میزان بزرگ شدن پنجره جست‌وجو را در هر بار فرایند جست‌وجو نشان می‌دهد. این پارامتر یک عدد بدون واحد بوده و همواره بزرگتر از یک است. بدیهی است هرچه مقدار این پارامتر افزایش یابد، سرعت اجرای الگوریتم افزایش خواهد یافت. البته وابسته به نوع تصویر، دقت نیز کاهش خواهد داشت.

### ۳-۲ مروری بر روش‌های بهبود الگوریتم ویولاجونز با استفاده از اطلاعات عمق

یکی از اولین کارها برای افزایش سرعت الگوریتم ویولاجونز با ترکیب داده‌های عمق و رنگ، در [۳] گزارش شده است. در [۳] با توجه به اطلاعاتی که از ارتفاع قرارگیری دوربین و نوع هدف در اختیار است، محدودیتی بر روی ارتفاع و اندازه اهداف موجود در صحنه قرار می‌گیرد. در نتیجه الگوریتم ویولاجونز فقط بر روی مکان‌هایی عمل می‌کند که در بازه محدود شده قرار گیرند. با استفاده از روش ارائه شده در [۳] می‌توان از محاسبات غیرضروری اجتناب کرده و ۸۵ درصد زمان آشکارسازی را کاهش داد. یکی از مهمترین مشکلات موجود در [۳] عدم استفاده از یک دوربین عمق مناسب و دقیق است. استفاده از دوربین کینکت می‌تواند باعث کاهش بیشتر زمان آشکارسازی شود.

از دیگر کارها برای افزایش سرعت الگوریتم ویولاجونز با به کارگیری داده‌های عمق و رنگ، می‌توان به [۴] اشاره کرد. در [۴] با استفاده از دو تصویر مختلف از یک صحنه، نقشه عمق صحنه به دست می‌آید. برای افزایش سرعت و کاهش بار محاسبات از نقشه بدست آمده در فواصل معینی (۱۰ پیکسل) نمونه برداری شده تا تصویر جدیدی تولید شود. با توجه به تصویر عمق تولید شده، محدودیتی بر روی اندازه چهره‌ها قرار می‌گیرد. [۴] توانسته است حدود ۲,۸ برابر سرعت آشکارسازی را افزایش دهد. اما الگوریتم ارائه شده در [۴] نیازمند پردازش ابتدایی برای ساخت تصویر عمق است. می‌توان این مشکل را با دوربین کینکت حل کرد.

یکی از اولین کارها برای بهبود سرعت الگوریتم ویولاجونز با استفاده از تصاویر عمق و رنگ کینکت، در [۵] گزارش شده است. در [۵] با استفاده از اطلاعات عمق صحنه، تخمینی از اندازه چهره به دست آمده است. با استفاده از این تخمین، مکان‌های احتمالی حضور چهره افراد در تصویر محاسبه شده است. با داشتن اطلاعات این مکان‌ها، مکان‌های غیرضروری برای جستجو حذف شده‌اند. بدین ترتیب با یک پردازش مناسب بر روی داده‌های عمق، فضای جستجو در الگوریتم اصلی کاهش یافته که این امر در نهایت به کاهش بار محاسبات می‌انجامد. در بهترین شرایط، روش ارائه شده در [۵] توانسته است، ۶ برابر بار محاسبات را کاهش دهد. گرچه [۵] توانسته زمان آشکارسازی چهره را کاهش دهد ولی این موضوع باعث کاهش دقت سیستم نیز شده است. این کاهش دقت مواقعی رخ می‌دهد که چهره فرد به دوربین نزدیک یا خیلی از آن دور باشد. هم‌چنین [۵] به دنبال بهینه کردن سایر پارامترهای الگوریتم ویولاجونز، نیز نبوده است.

تحقیق [۶] نیز از همان ایده‌های کلی ارائه شده در [۵] استفاده کرده است. روش طراحی شده در [۶] از دو قسمت کلی تشکیل شده است. در گام اول، نواحی موجود در تصویر عمق

بخش‌بندی<sup>۱</sup> می‌شوند. عمل بخش‌بندی ناحیه‌ها براساس مکان قرارگیری ناحیه‌ها و میانگین شدت عمق هر ناحیه چهره می‌گیرد. با انجام این عمل، ناحیه‌های غیر ضروری برای جستجو چهره حذف می‌شوند. در گام دوم، از تصویر بخش‌بندی شده درگام اول، برای آشکارسازی چهره استفاده شده است. نتایج شبیه‌سازی در [۶] نشان می‌دهد که با استفاده از این روش، می‌توان ۴,۷ برابر، سرعت آشکارسازی چهره را افزایش داد. اما این افزایش فقط برای تصاویری با بافت ساده است. اگر بافت تصویر پیچیده شود، آنگاه دقت آشکارسازی کم خواهد شد.

در [۸] با استفاده از هم‌چوشی داده‌های عمق و رنگ دوربین کینکت، دقت الگوریتم ویولاجونز افزایش یافته است. در [۸] ابتدا با استفاده از تکنیک آشکارسازی پوست<sup>۲</sup>، نواحی ممکن برای حضور صورت مشخص می‌شوند. سپس با استفاده از اطلاعات عمق، اندازه‌ی احتمالی چهره‌های موجود در تصویر مشخص می‌شوند. سپس بافت‌هایی که ممکن نیست چهره‌ای در آن‌جا باشد، نظیر بالای سقف، روی درخت و... حذف می‌شوند. با استفاده از این ایده‌ها، [۸] توانسته است به دقت خوبی در الگوریتم ویولاجونز برسد.

از دیگر کارهای انجام شده برای افزایش سرعت، می‌توان به کار meyer و همکاران [۹] اشاره کرد. در [۹] هدف ایجاد یک محدودیت هوشمند در فضای جستجو الگوریتم ویولاجونز است. در این تحقیق دو فرض اساسی در نظر گرفته شده است. نخست اینکه با داشتن نقشه عمق می‌توان اندازه تقریبی سر فرد را بدست آورد. بدین ترتیب نیازی نیست تا پنجره جستجو تمام تصویر را در مقیاس‌ها و مکان‌های مختلف کاوش کند. فرض دوم وجود ناپیوستگی در چهره فرد و پس زمینه است. بدین ترتیب اگر زیر تصویری دارای لبه نباشد، از حوزه جستجو خارج خواهد شد. نکته مهم در این مقاله یافتن بیشترین و کمترین اندازه زیر تصویر است که باید طبق چند فرض بدست آیند. در [۹] بعد از انجام محاسبات بر پایه فرض‌های بالا، نقشه‌ای بدست می‌آید که مکان‌های احتمالی حضور چهره را نشان می‌دهد. از مهمترین مشکلات این تحقیق شکست در عملیات آشکارسازی چهره در فاصله‌هایی است که فرد به دوربین نزدیک شده است. هم‌چنین برای چهره‌هایی که تمام رخ در برابر دوربین قرار نگرفته‌اند نیز، عملیات آشکارسازی به درستی انجام نمی‌شود. روش ارائه شده در [۹] توانسته است حدود ۲,۵۴ برابر سرعت آشکارسازی را تحت دقت تقریباً ثابت، افزایش دهد. درصد تشخیص‌های نادرست در این روش حدود ۱۸,۳۷ درصد بوده است، این در حالی است که میزان درصد خطا برای الگوریتم ویولاجونز برابر ۱۷,۷ درصد می‌باشد. این مقاله مرجع مقایسه ما در این تحقیق خواهد بود. علت انتخاب [۹] این است که تحقیق مذکور اولاً جدید بوده، ثانیاً

<sup>۱</sup> segmentation

<sup>۲</sup> skin detection

شد. بدیهی است بزرگتر شدن (مثلاً ۶۴\*۶۴) نیز اگرچه باعث افزایش دقت در یافتن نواحی و ساختارهای منحصر به فرد گردیده، ولی به همان نسبت سرعت کاهش خواهد یافت. به همین علت، ایجاد یک مصالحه در انتخاب اندازه و زمان پردازش، ضروری است. به طور خلاصه با کوچک کردن ابعاد تصویر عمق، بدون اینکه خطری متوجه ساختارهای موجود در تصویر شود، حجم محاسبات بر روی تصویر کاهش خواهد یافت.

محدود به کاربردی خاص نبوده و در چنین شرایطی بهترین نتایج را بدست آورده است و ثالثاً تصاویر عمق مورد استفاده مربوط به دوربین کینکت هستند.

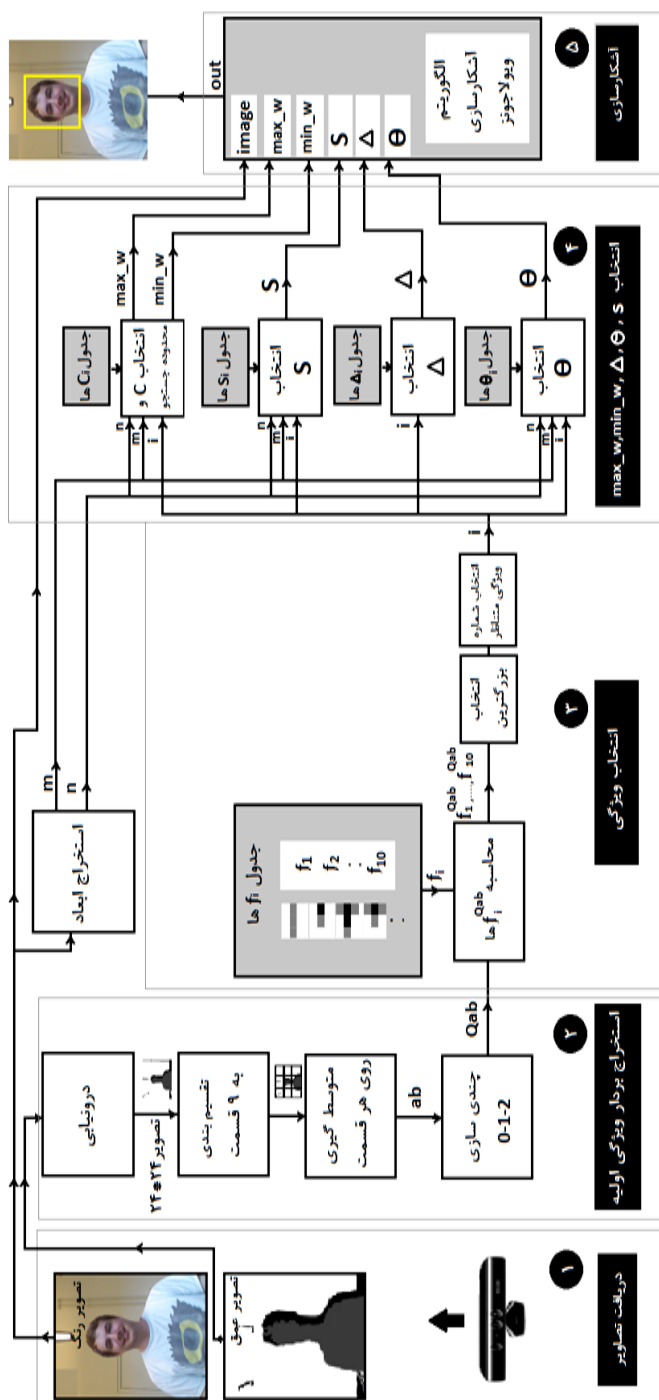
## ۴ روش پیشنهادی

در بخش ۳-۱ تاثیر عوامل مختلف بر روی سرعت و دقت الگوریتم آشکارسازی چهره بررسی شد. به طور خلاصه،  $\theta$  نقش مهمی بر روی دقت نهایی سیستم داشت. پارامترهای  $S$  و  $\Delta$ ، در میزان سرعت اجرای الگوریتم موثر بودند. کاهش ابعاد و پیچیدگی تصویر ورودی نیز به افزایش سرعت آشکارسازی کمک می‌کرد (هرچند این عوامل وابسته به تصویر بوده و در کنترل سیستم نیست). و در نهایت اعمال محدودیت بر روی پنجره جست‌وجو نیز منجر به افزایش سرعت می‌شد. هدف این تحقیق انتخاب مناسب  $S$ ،  $\theta$ ،  $\Delta$ ،  $\min\_w$  و  $\max\_w$  از روی ویژگی‌های تصویر ورودی است. بلوک‌دیگرام سیستم پیشنهادی در شکل ۵ نشان داده شده است. خلاصه این شکل را می‌توان در چند گام زیر بیان کرد:

**گام اول:** دریافت تصاویر عمق و رنگ صحنه به طور همزمان  
**گام دوم:** استخراج بردار ویژگی اولیه Qab از تصویر عمق صحنه  
**گام سوم:** انتخاب ویژگی  $f_i$  بر اساس بیشترین همبستگی بین بردار Qab و ویژگی‌های ذخیره‌شده در سیستم (نحوه‌ی استخراج و شماره‌گذاری این ویژگی‌ها در ادامه توضیح داده خواهد شد).  
**گام چهارم:** انتخاب عوامل موثر بر روی سرعت آشکارسازی (S,  $\theta$ ,  $\Delta$ ,  $\min\_w$  و  $\max\_w$ ) با استفاده از طول و عرض تصویر رنگ و ویژگی انتخابی  $f_i$   
**گام پنجم:** اعمال الگوریتم ویولا-جونز و آشکارسازی چهره از روی تصویر رنگ با عوامل انتخابی در گام چهارم

### ۴-۱ استخراج بردار ویژگی اولیه

طبق شکل ۵ در ابتدا، توسط دوربین کینکت تصاویر عمق و رنگ به طور همزمان از صحنه استخراج می‌شوند. ابعاد تصویر رنگ و عمق دریافتی از دوربین کینکت نیز قابل تغییر است. ورودی بخش ۲ از سیستم شکل ۵، تصویر عمق صحنه و خروجی آن برداری با نام Qab با ۹ مولفه می‌باشد. این بردار نماینده تصویر عمق خواهد بود. برای استخراج این بردار باید پردازش‌هایی بر روی تصویر عمق انجام گیرد. بدین منظور در ابتدا با یک عملیات درون‌یابی، تصویر عمق را به اندازه ۲۴\*۲۴ می‌رسانیم (این ابعاد بر اساس سعی و خطا بدست آمده است). ایده استفاده از درون‌یابی مشخص است. با استفاده از کوچک کردن ابعاد تصویر می‌توان پردازش‌های قدرتمندی با هزینه زمانی کمتر انجام داد، در حالیکه لطمه‌ای به ساختارهای منحصر به فرد در تصویر عمق، وارد نشود. ابعاد ۲۴\*۲۴ می‌تواند کوچکتر شود، اما کوچک کردن آن باعث از بین رفتن اطلاعات مهم تصویر خواهد



شکل ۵: بلوک دیگرام روش پیشنهادی

الگوریتم‌های مختلفی برای کوچک کردن ابعاد تصویر و درون‌یابی آن نظیر نزدیکترین همسایگی<sup>۱</sup>، دوخطی<sup>۱</sup> و دو مکعبی<sup>۲</sup> وجود دارد.

<sup>۱</sup> nearest

مشخص می‌کند. بر این اساس بهترین انتخاب برای هر یک از درایه‌های این ماتریس، متوسط فاصله قرارگیری افراد و اجسام تا دورین است. پس میانگین شدت پیکسل‌های موجود برای هر قسمت محاسبه شده و در ماتریس  $ab$  ذخیره می‌شود.

درایه‌های ماتریس بدست آمده در این مرحله باید به یکی از سه سطح ۰ یا ۱ یا ۲، چندی‌سازی<sup>۱</sup> شوند تا ماتریس Qab (Quantization of ab) بدست آید. تعداد سطوح گرچه می‌تواند بیشتر یا کمتر شود اما انتخاب سه سطح با توجه به ساختار تصویر عمق، باعث تمایز خوب در عین حال بار محاسباتی مناسب، خواهد شد. عمل چندی‌سازی بدین‌گونه است که اگر متوسط فاصله قرارگیری اهداف از  $L1$  کمتر باشد، سطح ۰، اگر متوسط فاصله قرارگیری اهداف بین  $L1$  و  $L2$  باشد، سطح ۱، و اگر متوسط فاصله قرارگیری اهداف از  $L2$  بیشتر باشد، سطح ۲ انتخاب می‌شود. برای یافتن مرز سطوح چندی‌سازی  $(L2, L1)$ ، دو رویکرد استفاده از عملیات سعی و خطا و روش وفقی از روی خود تصویر، وجود دارد. یکی از روش‌های استخراج سطوح به صورت وفقی استفاده از روش اوتسو (Nubuyuki Otsu) است. در این تحقیق این مرزها به صورت وفقی با روش اوتسو از روی تصویر عمق استخراج می‌شوند. استفاده از روش اوتسو با توجه به اینکه، اولاً بر اساس سنجش آماری شدت پیکسل‌ها به صورت وفقی و از روی خود تصویر عمل کرده و ثانیاً زمان زیادی صرف نمی‌کند، انتخاب مناسبی برای این قسمت الگوریتم است.

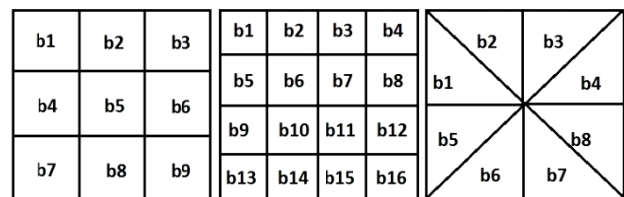
با چندی‌سازی ماتریس Qab و تبدیل آن به بردار Qab، برداری با ۹ مولفه خواهیم داشت که نشانگر ساختارهای مهم موجود در تصویر عمق است. در واقع این بردار نماینده چگونگی قرارگیری اهداف و اجسام مقابل دورین است. پس از استخراج بردار Qab از تصویر ورودی، باید بر اساس آن به سوالاتی درباره محتوای تصویر پاسخ داد، سوالاتی مانند اینکه چند چهره در تصویر وجود دارد؟ فرد ایستاده است یا نشسته؟ فاصله چهره‌ها تا دورین نزدیک است یا دور؟. بلوک‌های بعدی سیستم پیشنهادی (شکل ۵) به این سوالات پاسخ خواهند داد.

#### ۴-۲ انتخاب ویژگی

برای یافتن مقدار مناسبی برای پارامترهای الگوریتم ویولاجونز و به منظور تسریع در روند آشکارسازی، باید از تصویر عمق به عنوان اطلاعات اضافی استفاده کنیم. اما سوال مهم این است که چگونه از این اطلاعات اضافی بهره بگیریم؟ برای جواب به این سوال ابتدا باید دید که چه ویژگی‌های منحصر به فردی در تصاویر عمق مربوط به اشخاص دیده می‌شود؟! ویژگی‌های منحصر به فرد تصویر عمق در بخش ۲ توضیح داده شد. در آنجا بیان شد که تصویر عمق برای پیکسل‌های مختلف، ارزش‌های متفاوتی قائل است، ساختار همواری دارد، مکان‌ها و اهداف مهم تصویر به

در این تحقیق از روش درونیابی دوخطی استفاده شده است. علت این انتخاب رعایت دقت در عین کاهش زمان است. روش نزدیکترین همسایگی دارای زمان پردازش کم ولی دقت پایین و روش دومکعبی بالعکس دارای زمان پردازش زیاد ولی دقت بالاست.

تصویر عمق کوچک شده، به منظور بررسی دقیق‌تر و اصولی‌تر، باید به ۹ قسمت مساوی با ابعاد  $8*8$  تقسیم شده و برای هر قسمت نام‌گذاری انجام شود. برای تقسیم‌بندی<sup>۲</sup> تصویر می‌توان هر ساختاری را برگزید، اما این ساختار باید سریع ساخته شده و ویژگی‌های مورد نظر به سرعت از داخل آن بدست آید. برخی از ساختارهایی که برای تقسیم‌بندی تصویر می‌توانند به کار گرفته شوند در شکل ۶ نشان داده شده است. برای هر قسمت از ساختارهای شکل ۶ نام‌گذاری نیز انجام شده است  $(b1, b2, \dots)$ . نوع الف در شکل ۶ ساختار محاسباتی پیچیده داشته و نوع ب ساختار اطلاعاتی هم‌پوشانی دارد. اما نوع ج در همین شکل، علاوه بر تقارن مناسب (تعداد قسمت‌های فرد)، ساختار محاسباتی ساده‌ای هم دارد. به طور معمول و در تصاویر طبیعی، با ایجاد ساختاری نظیر نوع ج، می‌توان حداکثر ویژگی‌ها را در عین سادگی محاسبات، از تصویر استخراج کرد. بنابراین نوع ج، ساختار منتخب ما در این تحقیق است.



شکل ۶: سه نوع تقسیم‌بندی تصویر برای ایجاد ساختاری منظم در روند استخراج ویژگی‌ها

بعد از تقسیم‌بندی تصویر به ۹ قسمت، نوبت محاسبه مقادیر عددی  $(b1, b2, \dots, b9)$  این ۹ قسمت است. این مقادیر عددی در قالب ماتریس  $ab$  (average of block) ذخیره می‌شوند. مقادیر درایه‌های ماتریس  $ab$  باید به شکلی مناسبی انتخاب شوند تا حداکثر اطلاعات مهم صحنه را در بر داشته باشند. برای یافتن این مقادیر ابتدا باید نگاهی به ساختار تصویر عمق داشت. هر تصویر عمق، یک تصویر معمولی است که شدت هر پیکسل برای مختصات  $(x, y)$ ، در واقع نماینده فاصله قرارگیری آن مختصات تا لنز دورین است. محدوده قابل قبول بر اساس دورین کینکت بین ۴۰۰ میلی متر و ۳۰۰۰ میلی‌متر (۳ متر) است. تصویر عمق برای نواحی‌ای که خیلی به دورین نزدیک باشند مقدار صفر و برای مکان‌ها و اجسام خیلی دور تا دورین مقداری نظیر ۴۰۹۶ را

<sup>۱</sup> bilinear

<sup>۲</sup> bicubic

<sup>۳</sup> blocking

<sup>۴</sup> quantization

کرد. این ویژگی‌ها در ادامه به ترتیب قرار گرفته‌اند. برای هر ویژگی بردار ویژگی به همراه شکل کلی ساختار آن ویژگی قرار گرفته است. رنگ سیاه معادل سطح ۲، رنگ خاکستری معادل سطح یک و رنگ سفید معادل سطح صفر است. در جدول ۱ (جدول  $f_i$  ها) خلاصه‌ای از بردارهای ویژگی، فهرست شده‌اند. در شکل ۸ نیز ساختار این ده ویژگی نشان داده شده است.

ویژگی اول ( $f_1$ ): این اصلی‌ترین ساختار موجود بوده و مربوط به زمانی است که فردی در مقابل دوربین نشسته و فاصله متوسطی تا دوربین دارد. بردار ویژگی در این حالت عبارتست از:

$$f_1 = [000010121]$$

ویژگی دوم ( $f_2$ ): این ساختار مربوط به زمانی است که فردی در مقابل دوربین نشسته و فاصله آن تا دوربین کمتر از ۸۰۰ میلی‌متر است. بردار ویژگی در این حالت عبارتست از:

$$f_2 = [010020121]$$

ویژگی سوم ( $f_3$ ): این ساختار مربوط به زمانی است که فردی در مقابل دوربین نشسته و فاصله آن تا دوربین بیشتر از ۲۰۰۰ میلی‌متر است. بردار ویژگی در این حالت عبارتست از:

$$f_3 = [000010020]$$

ویژگی چهارم ( $f_4$ ): این ویژگی مربوط به حالتی است که فردی در برابر دوربین در فاصله نزدیک یا متوسط ایستاده است. بردار ویژگی در این حالت عبارتست از:

$$f_4 = [010010010]$$

ویژگی پنجم ( $f_5$ ): این ویژگی مربوط به حالتی است که فردی در برابر دوربین در فاصله دور ایستاده است. بردار ویژگی در این حالت عبارتست از:

$$f_5 = [000010010]$$

ویژگی ششم ( $f_6$ ): دو نفر در فاصله دور در برابر دوربین قرار گرفته‌اند. بردار ویژگی در این حالت عبارتست از:

$$f_6 = [000212222]$$

ویژگی هفتم ( $f_7$ ): این ویژگی هیچ توصیفی از صحنه ارایه نمی‌دهد، در واقع هیچ هدفی در برابر دوربین قرار ندارد. بردار ویژگی در این حالت عبارتست از:

$$f_7 = [222222222]$$

(این ویژگی یک جسم افقی در تصویر را  $f_8$  ویژگی هشتم) مشخص می‌کند. بردار ویژگی در این حالت عبارتست از:

$$f_8 = [000111000]$$

ویژگی نهم ( $f_9$ ): این ویژگی مشخص می‌کند که احتمال حضور هدف در گوشه‌های تصویر بیشتر از سایر مکان‌هاست. بردار ویژگی در این حالت عبارتست از:

$$f_9 = [000101101]$$

ویژگی دهم ( $f_{10}$ ): ویژگی دهم را ویژگی شطرنجی می‌نامیم. در این حالت ساختار آشنایی در تصویر وجود ندارد. این حالت همان حالت پیش‌فرض است. بردار ویژگی در این حالت عبارتست از:

$$f_{10} = [101010101]$$

راحتی از آن استخراج می‌شود، و در برابر تغییرات روشنایی مقاوم است.

با توجه به همواری تصویری عمق و ارزش متفاوت پیکسل‌های این تصویر برای مکان‌های متفاوت، با بررسی تصاویر مختلف عمق صحنه از یک جنبه یکسان (مانند تصویر یک فرد نشسته) می‌توان ویژگی‌هایی آشنا از آن‌ها استخراج کرد. شکل ۷ تصاویر ۳ فرد مختلف در ابعاد نابرابر را نشان می‌دهد که همگی جلوی دوربین نشسته‌اند. مقایسه بردار ویژگی اولیه (Qab) این ۳ تصویر به راحتی نشان می‌دهد که آنها ساختار تقریباً یکسانی دارند. این ساختار را می‌توان با تصویر د از همین شکل نمایش داد (ساختار د را بر اساس تصاویر الف، ب و ج به صورت ذهنی ترسیم می‌کنیم که در واقع نمایشی واحد برای تصاویر الف تا ج است). اگر برای چندین تصویر نشسته دیگر نیز بررسی فوق را انجام دهیم بازهم به ساختار تقریباً یکسانی با تصویر د شکل ۷ خواهیم رسید. بدین ترتیب می‌توانیم به این ساختار نام یک ویژگی آشنا را بدهیم. از این پس این ویژگی در هر تصویر عمق مشخص می‌کند که فردی در برابر دوربین نشسته است.



شکل ۷: تصاویر عمق سر افراد در ابعاد نابرابر و ساختارهای یکسان (ساختار کلی شکل های الف، ب و ج شبیه ساختار د است.)

اگر برای تصاویر در حالت‌های دیگر نیز چنین تجزیه و تحلیلی انجام شود می‌توان ساختارهای آشنای دیگری را بدست آورده و بدین وسیله جدول ویژگی‌های ذخیره شده در سیستم ( $f_i$ ) را توسط آنها تشکیل داد. در این تحقیق ۱۰ بردار ویژگی  $f_i$  پس از بررسی تصاویر مختلف بدست آمده است. این ویژگی‌ها را می‌توان در دو دسته تشخیص و توصیف صحنه دسته‌بندی کرد. در دسته اول که از  $f_1$  شروع و تا  $f_6$  ادامه دارند، ویژگی صحنه تشخیص داده می‌شود. برای مثال ویژگی  $f_4$  بیان می‌کند که در صحنه فردی در فاصله متوسط تا دوربین ایستاده است. تشخیص اینکه صحنه دربردارنده چه ویژگی‌هایی است باعث می‌شود تا بتوان تصمیم بهتری در مورد پارامترهای الگوریتم آشکارسازی چهره گرفت. دسته دوم که از  $f_7$  تا  $f_{10}$  ادامه دارد، ویژگی صحنه را توصیف می‌کنند. مثلاً ویژگی  $f_7$  مشخص می‌کند که در صحنه هیچ چیز قابل ذکری وجود ندارد. بدیهی است درک چنین توصیف‌هایی از صحنه برای یک الگوریتم آشکارسازی غیرهوشمند کمک شایانی به الگوریتم آشکارسازی چهره خواهند



$i$ ، برای انتخاب پارامترهای آشکارسازی  $(S, \theta, \Delta, \min_w$  و  $\max_w$ ) مورد نیاز خواهد بود. در روابط (۱) و (۲)، امید ریاضی از نسبت مجموع مولفه‌ها به تعداد مولفه‌ها بدست می‌آید.

$$f_i^{Qab} = \frac{E\{(f_i - \mu_{f_i})(Qab - \mu_{Qab})\}}{\sqrt{E\{f_i - \mu_{f_i}\}^2 * E\{Qab - \mu_{Qab}\}^2}} \quad i = 0, \dots, 10 \quad (۱)$$

$$= \frac{\sum(f_i - \mu_{f_i})(Qab - \mu_{Qab})}{\sqrt{\sum(f_i - \mu_{f_i})^2 * \sum(Qab - \mu_{Qab})^2}} \quad i = 0, \dots, 10 \quad (۲)$$

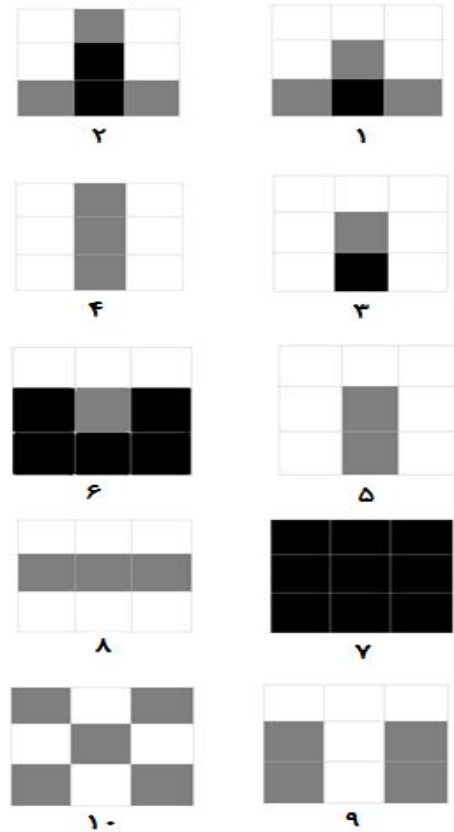
### ۳-۴ انتخاب پارامترها

بخش دیگر در سیستم پیشنهادی (شکل ۵)، انتخاب مناسب پارامترهای آشکارسازی از روی داده‌های عمق است. در اولین گام طبق نیاز الگوریتم ویولاجونز، به دنبال ایجاد محدودیتی برای ابعاد پنجره جست‌وجو هستیم. در این صورت لازم است حداقل  $(\min_w)$  و حداکثر  $(\max_w)$  ابعاد پنجره جست‌وجو را تعیین کنیم. این بازه باید بر اساس شماره ویژگی  $i$  و ابعاد تصویر رنگ صحنه  $(m, n)$  بدست آید. روابط (۳) و (۴) مقادیر حداکثر و حداقل انتخابی را برای ابعاد پنجره نشان می‌دهد. بعد از یافتن  $\max_w$  به طور تقریبی  $\frac{3}{4}$  اندازه آن را برای کران پایین بازه انتخاب می‌کنیم.

$$\max_w = \frac{m + n}{2c} \quad (۳)$$

$$\min_w = \frac{3}{4} \max_w \quad (۴)$$

انتخاب ثابت  $C$  از روی جدول  $C_i$  ها و نحوه تعیین عناصر این جدول بر اساس یک بررسی ساده بر روی ویژگی‌های  $f_i$  صورت می‌گیرد. برای مثال  $C_1$  را می‌توانیم با بررسی تصویر فردی که در برابر دوربین در فاصله متوسط نشسته است (شکل ۹- الف)، بدست آوریم. بر اساس شکل ۹- الف مشخص است که ابعاد پنجره‌ای که بر روی صورت محاط می‌شود حدودا برابر با  $\frac{1}{5.5}$  ابعاد طول و عرض تصویر است. پس می‌توانیم  $C_1 = 5.5$  در نظر بگیریم. به همین ترتیب در شکل ۹- ب،  $C_2 = 2$  در نظر گرفته می‌شود. دیگر ثابت‌ها نیز بر همین اساس بدست می‌آیند. این مقادیر در جدول ۲ فهرست شده‌اند. برای ویژگی‌های هفتم و دهم با توجه به اینکه نمی‌توانیم حضور هدفی را در صحنه تشخیص دهیم، ثابتی در نظر گرفته نشده، در نتیجه محدودیتی بر روی ابعاد پنجره اعمال نمی‌شود. هم‌چنین با توجه به اینکه



شکل ۸: ده ساختار آشنا

جدول ۱: جدول  $f_i$  ها

$i$	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
$f_i$	[000010121]	[010020121]	[000010020]	[010010010]	[000010010]	[000212222]	[222222222]	[000111000]	[000101101]	[101010101]

بر طبق بلوک دیاگرام سیستم پیشنهادی (شکل ۵)، بردار  $Qab$  حاصله باید با ۱۰ بردار ویژگی ذخیره شده  $f_i$  در سیستم (جدول ۱) مقایسه و نزدیکترین  $f_i$  به این بردار مشخص شود. برای یافتن شبیه‌ترین ویژگی  $f_i$  به بردار ویژگی  $Qab$  از ضریب همبستگی بین دو بردار استفاده می‌کنیم. رابطه (۱) ضریب همبستگی بین این دو بردار را نشان می‌دهد [۱۰]. در این رابطه میانگین عناصر بردار  $f_i$  و  $\mu_{Qab}$  مقدار میانگین عناصر بردار  $Qab$  می‌باشد. می‌توان رابطه (۱) را با ساده‌سازی به شکل رابطه (۲) نوشت. (یادآوری می‌شود که بردارهای  $f_i$  و  $Qab$  بردارهایی با ۹ مولفه می‌باشند). مقادیر ضریب همبستگی  $(f_i^{Qab})$  بین هر یک از ده بردار  $f_i$  و بردار  $Qab$  محاسبه شده و  $f_i$  متناظر با بزرگترین  $f_i^{Qab}$ ، به عنوان ویژگی سیستم اعلام می‌شود. بدیهی است پس از تعیین نزدیکترین  $f_i$  به بردار  $Qab$ ، تنها شماره (اندیس) ویژگی

ویژگی‌های هشتم و نهم، تنها توصیفی از صحنه ارایه می‌دهند، پس بر اساس همین توصیف صحنه، مقداری تخمینی برای ثابت‌ها در نظر گرفته می‌شود.

شکل ۹: چگونگی انتخاب ثابت‌های  $C_1$  تا  $C_6$

محدوده معقولی قرار گیرد. (S به طور کلی از بین چهار مقدار ۱،۱، ۱،۲، ۱،۵ و ۱،۹، انتخاب می‌شود). سپس الگوریتم آشکارسازی با این عدد، در عین ثابت بودن سایر پارامترها، اجرا شده و مقدار نهایی آن پس از بررسی روی تصاویر مختلف براساس کسب بهترین نتیجه و به روش سعی و خطا بدست می‌آید. بدیهی است که این مقادیر، مقادیر بهینه نیستند، اما به طور متوسط زمان آشکارسازی را روی تصاویر مختلف، کاهش خواهند داد. اما انتخاب مقدار نهایی S علاوه بر شماره ویژگی  $I$  به ابعاد تصویر رنگ ( $m * n$ ) نیز وابسته است. بدین ترتیب اگر حاصل  $m * n$  تصویر از مقدار خاصی (۱۶۰۰۰۰) بیشتر باشد (تصاویر با ابعاد بالا)، مقدار S از سطر اول جدول انتخاب شده و اگر از مقدار خاصی (۲۲۵۰۰) کمتر باشد (تصاویر با ابعاد پایین) مقدار S از سطر سوم انتخاب می‌شود. در غیر این صورت مقدار S از سطر دوم انتخاب خواهد شد.

جدول ۳: جدول  $S_i$  ها

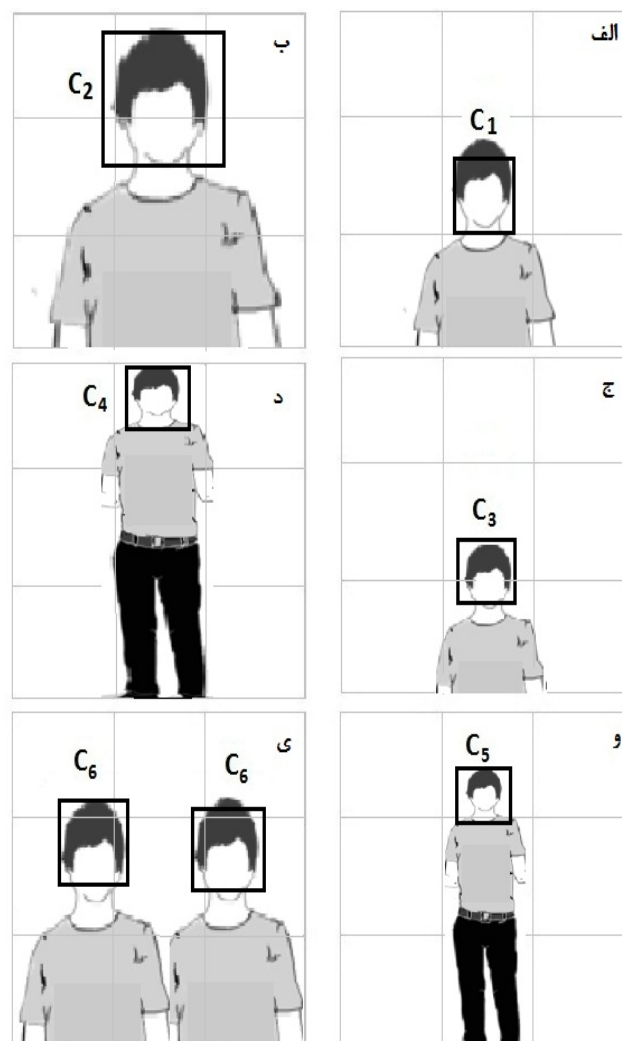
S	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
[بالا]	۱،۹	۱،۹	۱،۹	۱،۹	۱،۵	۱،۵	۱،۹	۱،۹	۱،۹	۱،۹
[متوسط]	۱،۲	۱،۲	۱،۱	۱،۵	۱،۲	۱،۲	۱،۵	۱،۵	۱،۵	۱،۵
[پایین]	۱،۲	۱،۱	۱،۱	۱،۲	۱،۱	۱،۱	۱،۵	۱،۲	۱،۲	۱،۲

به موازات یافتن مقدار S، باید مقدار مناسبی برای  $\theta$  و  $\Delta$ ، با همان شیوه انتخاب S، طوری که موجب کاهش دقت نگردد، انتخاب کرد. اگر اندازه  $\theta$ ، بیشتر از اندازه مناسب انتخاب شود ممکن است ناحیه چهره نیز آشکار نشود و بالعکس، اگر کمتر انتخاب گردد ممکن است علاوه بر ناحیه چهره، نواحی دیگری نیز به اشتباه به عنوان چهره انتخاب شوند. جدول ۴ مقدار مناسب این پارامترها را بر اساس شماره ویژگی  $I$  نشان می‌دهد.

جدول ۴: جدول  $\theta_i$  ها و جدول  $\Delta_i$  ها

$\theta$ و $\Delta$	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
[بالا]	۴	۳	۴	۳	۴	۳	۴	۳	۳	۲
[متوسط]	۳	۲	۳	۳	۳	۳	۲	۲	۲	۲
[پایین]	۲	۲	۲	۲	۲	۲	۲	۲	۲	۱
	۱	۱	۱	۱	۱	۱	۲	۱	۱	۱

به هنگام استفاده از روش پیشنهادی، ممکن است سوالی مهم پیش آید، و آن اینکه اگر شرایط تصویر طوری باشد که ساختاری را به وجود آورد که ناخودآگاه شبیه یکی از ویژگی‌های انتخابی ما شود آنگاه چگونه می‌توان از صحت عملکرد روش پیشنهادی بردار مطمئن بود؟... برای مثال می‌دانیم طبق روش پیشنهادی بردار ویژگی فردی که در مقابل دوربین نشسته و فاصله آن تا دوربین

شکل ۹: چگونگی انتخاب ثابت‌های  $C_1$  تا  $C_6$ جدول ۲: جدول  $C_i$  ها

$C_i$	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
	۵،۵	۲	۶	۶،۵	۷	۵،۵	-	۸	۸	-

بعد از انتخاب بازه‌ای برای ابعاد پنجره جست‌وجو، به دنبال انتخاب عامل S هستیم. عوامل مختلفی بر روی اندازه این پارامتر موثرند. برای مثال می‌توان به ابعاد تصویر و ویژگی‌های منحصر به فرد در تصویر عمق اشاره کرد (این ویژگی‌ها در بخش ۲ معرفی شدند). در این تحقیق پارامتر S را بر اساس همین دو عامل انتخاب کرده‌ایم. بدیهی است که مقدار انتخابی S باید در عین افزایش سرعت، دقت را کاهش ندهد. جدول ۳ مقادیر S را نشان می‌دهد. برای انتخاب مقادیر جدول ۳، در ابتدا بر اساس شماره ویژگی  $I$ ، عددی را انتخاب می‌کنیم. این عدد طوری انتخاب می‌شود که با توجه به ابعاد تصویر و مقدار پیش فرض  $S=1.1$ ، در

دو نفر در فواصل مختلف در برابر دوربین قرار می‌گیرند، و در بیشتر کاربردهای بینایی ماشین نیز چنین حالتی وجود دارد. به طور خلاصه، گرچه نمی‌توان منکر این شد که دسته‌بندی‌های بیشماری از افراد و ژست‌ها وجود دارد، اما می‌توان ادعا کرد که تنها تعدادی از این دسته‌ها (ویژگی‌های  $f_1$  تا  $f_6$ ) سهم عظیمی (در واقع اگر بتوانیم تمام تصاویر تولیدی چهره در سراسر دنیا را گردآوری کنیم، می‌توانیم بدون آزمایش کمی، و به صورت ذهنی ادعا کنیم که بیشتر این تصاویر را می‌توان در گروه‌های آشنا و قابل حدسی قرار داد.) از تعداد کل را در خود جای خواهند داد. چه سیستمی وجود یکی از این دسته‌ها را تشخیص بدهد؛ چه ندهد، به جرات می‌توان گفت ابهام سیستم در مورد ویژگی تصویر ورودی تا حدود ۶۰٪ (۶ ویژگی از مجموع ۱۰ ویژگی) برطرف خواهد شد. بدیهی است این رفع ابهام به انتخاب هوشمندانه پارامترها کمک شایانی خواهد کرد، زیرا لازم نیست برای هر تصویر ورودی روندی تکراری و غیرهوشمندانه طی شود.

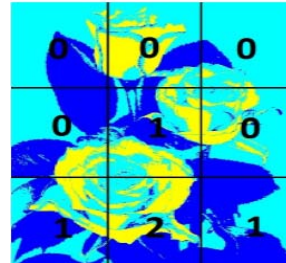
شکل ۱۱: تصاویری با تعداد افراد بیش از دو نفر با ژست‌های متفاوت



## ۵ نتایج شبیه‌سازی

بر طبق شکل ۵، در اولین قسمت از شبیه‌سازی‌ها، بردار ویژگی Qab را برای هفت تصویر نمونه مختلف در شکل ۱۲ بدست می‌آوریم. تصاویری که برای این قسمت انتخاب شده‌اند، دارای حالات و ویژگی‌های متفاوتی هستند. **تصویر نمونه اول:** در این تصویر فرد بدون چرخش سر و حالت خاصی به دوربین نگاه می‌کند. چنین حالتی ایده‌آل برای یک سیستم آشکارسازی است. لازم به ذکر است که ساختارهای بدست آمده از تصویر عمق این تصاویر در ستون سمت چپ شکل ۱۲ نشان داده شده است. **تصویر نمونه دوم:** در این تصویر دو چهره در برابر دوربین قرار گرفته است. **تصویر نمونه سوم:** این نیز تصویری مشکل برای آشکارسازی است. در این حالت فردی ایستاده در مقابل دوربین و چهره او نسبت به دوربین دور است. ساختار بدست آمده از تصویر عمق این صحنه در شکل سوم ۱۲ چپ نشان داده شده است. ماتریس و بردار Qab استخراجی از این تصویر نیز ارائه گردیده است. **تصویر نمونه چهارم:** در این تصویر، حالت متفاوتی نسبت به حالت عادی یک چهره وجود دارد. ساختار بدست آمده از تصویر عمق این صحنه در شکل چهارم چپ نشان داده شده است. ماتریس و بردار Qab استخراجی از این تصویر نیز ارائه گردیده

متوسط است به شکل  $f_1 = [0 0 0 0 1 0 1 2 1]$  است. حال اگر ساختار تصویر عمق شکل ۱۰ را در نظر بگیریم آنگاه بردار Qab استخراجی از آن دقیقاً شبیه ویژگی فردی می‌شود که در برابر دوربین در فاصله متوسط نشسته است. در چنین حالتی تکلیف سیستم چه خواهد شد؟!



ب



الف

شکل ۱۰: تصویر یک گل، (الف) تصویر رنگ، (ب) تصویر عمق

پاسخ به این سوال بسیار ساده است. اگر تصویری ساختاری شبیه به یکی از ویژگی‌های مشخص داشت، مثلاً شبیه یک فرد نشسته در فاصله نزدیک مقابل دوربین بود، ولی چهره‌ای در تصویر مشخص نبود، به طور طبیعی الگوریتم چهره‌ای را آشکار نخواهد کرد، چون طبق توضیحات بخش ۳، طبقه‌بند تصویر چهره‌ای را تشخیص نداده و بدین ترتیب هر انتخابی برای S و  $\theta$  و سایر پارامترها، تاثیری در این موضوع (تشخیص اشتباه ویژگی در تصاویر بدون چهره) نخواهد داشت. برای نمونه نتیجه شبیه‌سازی الگوریتم آشکارسازی چهره ویولاجونز بر روی تصویر گل در شکل ۱۰، هیچ نتیجه‌ای در پی نداشت.

## ۴-۴ تعمیم روش پیشنهادی به حالت‌های پیچیده

یکی از موضوعات مهم برای بررسی صحت روش پیشنهادی، تعمیم آن به حالت‌های پیچیده است. در یک تصویر نمونه، ممکن است سه نفر، چهار نفر یا بیشتر در برابر دوربین قرار گرفته باشند. یا افرادی که در برابر دوربین قرار گرفته‌اند، ژست‌های متفاوت و پیچیده و منحصر به فردی داشته باشند. شکل ۱۱ دو نمونه از این تصاویر را نشان می‌دهد. در این حالات روش پیشنهادی چه پاسخی خواهد داشت...؟

برای پاسخ به این پرسش باید به دو موضوع توجه داشت. اول اینکه، فرض کنیم برای تصویری مشخص شود که این تصویر یکی از ویژگی‌های  $f_1$  تا  $f_6$  را ندارد، یعنی تعداد نفرات مقابل دوربین یک یا دو نفر نیست. گرچه این تشخیص منفی است، اما نتیجه مثبتی در پی خواهد داشت. به عنوان مثال، هنگامی که سیستم تشخیص دهد که تنها یک نفر در برابر دوربین قرار نگرفته است، در این حالت می‌توان پارامترها را طوری تنظیم کرد که برای این حالت مناسب نباشد، این موضوع کمک بزرگی به افزایش سرعت آشکارسازی خواهد کرد. البته علاوه بر این، باید به موضوعی دیگر نیز توجه داشت. در بیشتر تصاویر طبیعی، یک یا

برای ۷ تصویر نمونه، بردار  $Qab$  بدست آمد. برای بررسی مقادیر همبستگی بین بردار ویژگی اولیه  $Qab$  و بردارهای ویژگی ذخیره شده در سیستم، مقادیر همبستگی برای تصاویر فوق‌الذکر در جدول ۵ فهرست شده‌اند. با توجه به جدول واضح است که برای مثال در تصویر سوم بیشترین مقدار همبستگی بین بردار  $Qab$  با ویژگی‌های  $f_i$  در  $f_5$  رخ داده است. ویژگی  $f_5$  نشان می‌دهد که فردی در فاصله دور تا دوربین ایستاده است. این ادعا با بررسی تصویر نمونه سوم اثبات خواهد شد. این حالت برای تصویر نمونه چهارم در  $f_1$  رخ داده است یعنی فردی در مقابل دوربین در فاصله متوسط نشسته است. (در جدول ۵ خانه‌هایی که با رنگ مشکی مشخص شده‌اند، مقدار بیشینه و خانه‌های توسی، نزدیکترین مقدار را به بیشینه مشخص می‌کنند).

برای بررسی و تفسیر نتایج جدول ۵، مقادیر همبستگی برای تصویر نمونه سوم، در نمودار شکل ۱۳ نمایش داده شده است. در این شکل میزان نزدیکی اعداد به یکدیگر به صورت گرافیکی نشان داده شده است که این روش نمایش می‌تواند برخی از نکات را بهتر روشن کند. با توجه به این نمودار، کمترین میزان همبستگی (از لحاظ اندازه) با ویژگی  $f_{10}$  رخ داده است. این بدان معنی است که حتماً در این تصویر ساختار آشنا و منظمی وجود دارد. البته به طور معمول چنین اتفاقی رخ خواهد داد چون در بیشتر تصاویر طبیعی که در آن چهره فرد یا افرادی باشد، نواحی و ساختارها در آن تصاویر کمترین شباهت را با ویژگی  $f_{10}$  ایجاد خواهند کرد. علاوه بر این، ضریب همبستگی بین بردار  $Qab=[1\ 1\ 1\ 1\ 2\ 1\ 1\ 2\ 1]$  و  $f_5$  و  $f_3$  تصویر نمونه سوم بیشترین مقادیر را با ویژگی‌های  $f_5$  و  $f_3$  داراست. گرچه در این حالت، نزدیکترین مقدار ضریب همبستگی (بیشترین مقدار ضریب همبستگی) ویژگی به بردار  $Qab$ ، ویژگی  $f_5$  است اما با توجه به نزدیکی مقدار  $f_3$  به  $f_5$  ممکن است در برخی تصاویر شبیه به تصویر نمونه سوم، مقدار ضریب همبستگی بین  $Qab$  و  $f_3$  کمی از مقدار ضریب همبستگی بین  $Qab$  و  $f_5$  بیشتر شده، در نتیجه سیستم به اشتباه ویژگی تصویر را  $f_3$  اعلام کند. گرچه بروز چنین اشتباهی کمی دور از ذهن به نظر می‌رسد، اما باید توجه داشت حتی اگر چنین اشتباهی نیز رخ دهد، یعنی به جای اینکه ویژگی سیستم ایستاده بودن فردی در فاصله دور اعلام شود، نشسته بودن فرد در فاصله دور اعلام شود، مقادیر پارامترهای انتخابی و نتیجه آشکارسازی تغییر زیادی نخواهند کرد، چون ویژگی اصلی هر دو حالت، یعنی دور بودن فرد تا دوربین ثابت مانده است. بررسی نتایج ضریب همبستگی برای تصویر نمونه پنجم نیز نشان می‌دهد که نزدیکترین ویژگی به ویژگی انتخابی سیستم ( $f_1$ )،  $f_6$  است، با توجه به اینکه در این تصویر فرد دست خود را بلند کرده است، این نتیجه معقول به نظر می‌رسد چون ویژگی  $f_6$ ، حالتی را بیان می‌کند که دو نفر در برابر دوربین قرار گرفته‌اند. هم چنین برای تصویر نمونه هفتم، روش پیشنهادی ویژگی  $f_6$  را به عنوان نتیجه اعلام

است. تصویر نمونه پنجم: در این تصویر فردی دست خود را بلند کرده است.



شکل ۱۲: هفت تصویر نمونه برای استخراج و انتخاب ویژگی

در چنین حالتی ممکن است سیستم به اشتباه، حضور دو نفر را در برابر دوربین اعلام کند. ساختار بدست آمده از تصویر عمق این صحنه در شکل پنجم چپ نشان داده شده است. ماتریس و بردار  $Qab$  استخراجی از این تصویر نیز ارائه گردیده است. تصویر نمونه ششم: در این تصویر، چهره فرد فاصله نزدیکی با دوربین دارد. ساختار بدست آمده از تصویر عمق این صحنه در شکل ششم چپ نشان داده شده است. ماتریس و بردار  $Qab$  استخراجی از این تصویر نیز ارائه گردیده است. تصویر نمونه هفتم: در این تصویر دو نفر در برابر دوربین قرار گرفته‌اند. ساختار بدست آمده از تصویر عمق این صحنه در شکل هفتم چپ نشان داده شده است. ماتریس و بردار  $Qab$  استخراجی از این تصویر نیز ارائه گردیده است.

روی یک پردازنده اینتل ۲،۲ گیگاهرتز با حافظه ۲ مگابایت انجام شده است.

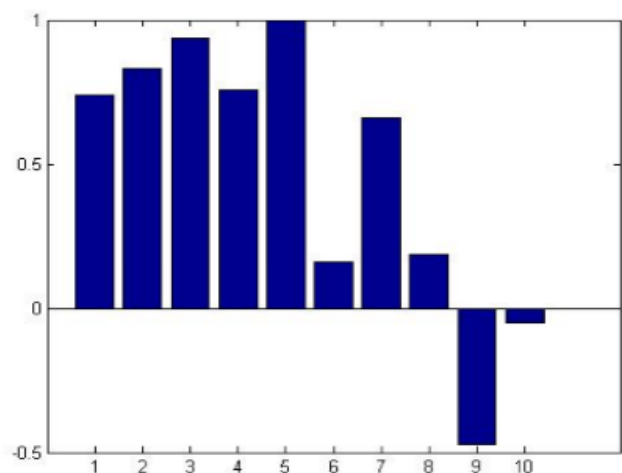
برای ارزیابی هر سیستم پیشنهادی، از نسبت سرعت آن به سرعت سیستم ویولاجونز استفاده می‌گردد. این عدد معیاری برای مقایسه روش پیشنهادی با سایر روش‌هاست. سیستم ویولاجونز دارای  $S=1.1$  و  $\theta=4$  و  $\Delta=1$  و  $W$  غیر محدود است. اما بدیهی است که استفاده از این مقادیر پیش‌فرض باعث جست‌وجویی غیرهوشمندانه شده و در نهایت سیستم کند و غیرمنعطفی برای آشکارسازی به وجود خواهد آمد. در خروجی سیستم ویولاجونز، چهره‌ی تعدادی از تصاویر به طور صحیح آشکار شده و طبیعتاً برخی نیز آشکار نمی‌شوند. در حالت کلی متوسط زمان آشکارسازی برای سیستم پیشنهادی نسبت به سیستم ویولاجونز برای ۲۷۰ تصویر حدود ۲،۷۴ برابر بهتر شده است. دقت سیستم پیشنهادی نیز در حد ثابتی باقی مانده است (دقت سیستم پیشنهادی برای ۲۷۰ تصویر، حدود ۱،۰۵ برابر بیشتر شده است). لازم به ذکر است که دقت آشکارسازی از نسبت عددی تعداد مکان‌های صحیح آشکار شده به تعداد کل مکان‌های آشکار شده بدست آمده است. در حالت کلی سیستم پیشنهادی حدود ۲،۷۴ برابر نسبت به سیستم ویولاجونز سریع‌تر شده است. این میزان به طور متوسط حدود ۱،۱ برابر از نتایج کار [۹] بهتر شده است. جدول ۶ نتایج سرعت و دقت آشکارسازی را برای ۲۰ تصویر مختلف از مجموع ۲۷۰ تصویر نشان می‌دهد. در حالت کلی برای مقایسه میزان سرعت دو روش آشکارسازی، دو نگاه متفاوت از زمان آشکارسازی وجود دارد. اولین حالت مقایسه سرعت دو سیستم آشکارسازی است. مقایسه دو سیستم آشکارسازی پیشنهادی با سیستم ویولاجونز، ۲،۷۴ برابر سرعت بیشتر سیستم پیشنهادی را نتیجه می‌دهد. بدیهی است منظور از سیستم، تمام بخش‌های نشان داده شده در شکل ۵ است. یکی دیگر از معیارهای میزان افزایش سرعت، مقایسه سرعت دو الگوریتم آشکارسازی است. منظور از الگوریتم، بخش ۵ نشان داده شده در شکل ۵ است. سرعت یک الگوریتم در واقع مدت زمان خالص برای پیمایش، جست‌وجو، تصمیم‌گیری و آستانه‌گیری برای یافتن محل چهره‌هاست. این حالت در واقع صرفاً نماینده زمان لازم برای اجرای الگوریتم آشکارسازی، بدون توجه به عوامل دیگری نظیر تاخیر در خواندن تصاویر ورودی رنگ و عمق، برنامه نویسی غیربهبوده برای اجرای الگوریتم‌های پیش‌پردازش روی داده‌ها و تاخیر و اتلاف زمان در نمایش تصویر نهایی است. در این تحقیق، با استفاده از روش پیشنهادی، به فراخور تصویر ورودی پارامترهای آشکارسازی انتخاب شدند. این در حالی است که در الگوریتم ویولاجونز، بدون توجه به تصویر ورودی و ویژگی‌های آن، همواره پارامترهای یکسانی در نظر گرفته می‌شود. به عنوان مثال اندازه پارامتر  $S$  برای دو تصویر با ابعاد بالا و پایین یکسان

کرده است، یعنی دو نفر در برابر دوربین قرار گرفته‌اند. در این حالت نزدیکترین ویژگی در  $f_1$  رخ داده است، یعنی تنها یک نفر در فاصله متوسط تا دوربین قرار دارد، با توجه به تصویر این نتیجه نیز معقول به نظر می‌رسد زیرا فرد سمت چپ تاثیر زیادی بر روی صحنه ندارد.

جدول ۵: مقادیر پارامتر همبستگی ( $f_i^{Qab}$ ) برای ۷ تصویر نمونه

تصویر	ت ۱	ت ۲	ت ۳	ت ۴	ت ۵	ت ۶	ت ۷
$f_1^{Qab}$	۰،۷۴	۰،۵۴	۰،۷۳	۰،۹۱	۰،۶۹	۰،۸۴	۰،۵۸
$f_2^{Qab}$	۰،۵۵	۰،۳۵	۰،۸۳	۰،۸۴	۰،۵۵	۱	۰،۲۵
$f_3^{Qab}$	۰،۸۸	۰،۱۷	۰،۹۳	۰،۸۱	۰،۴۹	۰،۷۷	۰،۱۲
$f_4^{Qab}$	۰،۵۰	-۰،۱	۰،۷۵	۰،۴۵	۰،۲۷	۰،۸۰	-۰،۱
$f_5^{Qab}$	۰،۶۶	۰،۱۸	۱	۰،۷۳	۰،۳۶	۰،۸۳	۰،۰۵
$f_6^{Qab}$	۰،۳۰	۰،۸۵	۰،۱۶	۰،۵۱	۰،۵۹	۰،۲۲	۰،۷۵
$f_7^{Qab}$	*۰،۱	*۰،۱	*۰،۶	*۰،۲	۰،۰۶	*۰،۳	*۰،۳
$f_8^{Qab}$	-۰،۲	۰،۲۵	۰،۱۸	-۰،۲	-۰،۱	-۰،۱	-۰،۱
$f_9^{Qab}$	-۰،۳	۰،۶۷	-۰،۴	۰،۰۷	۰،۲۱	-۰،۳	۰،۵۵
$f_{10}^{Qab}$	-۰،۳	۰،۰۳	۰،۰۵	۰،۰۷	۰،۲۱	۰،۰۳	-۰،۱

شکل ۱۳: نمودار اندازه مقادیر همبستگی برای تصویر نمونه سوم



پس از یافتن بردار ویژگی و پارامترهای متناظر برای چند تصویر نمونه، در ادامه به دنبال بررسی میزان افزایش سرعت روش پیشنهادی می‌باشیم. لازم به ذکر است که در این تحقیق هدف ما آموزش سیستم نیست. بنابراین برای شبیه‌سازی روش پیشنهادی، از یک طبقه‌بند آموزش دیده بر اساس آداپوست [۱] در نرم افزار متلب استفاده می‌شود. برای شبیه‌سازی و بررسی صحت سیستم پیشنهادی، حدود ۱۵۰ تصویر از مجموعه داده دانشگاه آلبورگ [۷] و حدود ۱۲۰ تصویر از اینترنت انتخاب شد. در این تصاویر رنگ پوست متفاوت افراد، حالات احساسی متفاوت چهره، ابعاد متفاوت، نزدیکی و دوری چهره به دوربین، به چشم می‌خورد. لازم به ذکر است که شبیه‌سازی روش پیشنهادی و روش ویولاجونز، بر

عمق و ویژگی‌های آن، می‌تواند در سایر کاربردهای بینایی ماشین نیز به کار رود.

## ۶ جمع‌بندی و پیشنهادات

هدف ما در این تحقیق استفاده از داده‌های عمق برای بهبود سرعت اجرای الگوریتم آشکارسازی چهره بود. از این سو پس از استخراج برداری از تصویر عمق محیط، ویژگی‌هایی از آن استخراج و از روی این ویژگی‌ها به دنبال استخراج محتوا و توصیفی از صحنه بودیم. بدیهی است که اگر بدانیم در یک تصویر چند نفر قرار گرفته‌اند و فاصله آن‌ها از دوربین چقدر است و... راحت‌تر می‌توان الگوریتمی هوشمندانه‌تر برای اجرای الگوریتم آشکارسازی چهره ساخت. روش پیشنهادی این تحقیق هرگز ادعا نمی‌کند که می‌تواند ویژگی هر تصویر ورودی را به طور دقیق شناسایی کند، بدیهی است چنین ادعایی کاملاً غیرمعمولانه است زیرا هر تصویری می‌تواند ویژگی پیچیده‌ای و منحصر به فردی داشته باشد، اما روش پیشنهادی این تحقیق می‌تواند تا حدودی صحنه را توصیف یا ویژگی صحنه را تشخیص دهد. در این تحقیق ویژگی‌هایی از تصاویر استخراج و در غالب ۱۰ بردار  $f_i$  ارایه شد. ممکن است انتخاب این ویژگی‌ها به شکل سطحی به نظر برسند، اما بر اساس ایده ارایه شده، این ویژگی‌های سطحی در عمل پرکاربردترین و آشناترین ویژگی‌های تصاویر دارای چهره هستند، پس می‌توان آن‌ها را مهم‌ترین ویژگی‌های موجود در تصاویر دانست. به عنوان یک روش جدیدتر می‌توان از یک سیستم آموزشی نظیر شبکه عصبی برای استخراج و دسته‌بندی این بردارها استفاده کرد. هم‌چنین در طراحی بلوک‌ها می‌توان از بلوک‌هایی با اندازه نامتقارن استفاده کرد. برای انتخاب قاعده‌ی طراحی بلوک‌های نامتقارن می‌توان از نحوه‌ی توجه طبیعی انسان به محیط اطرافش الهام گرفت که چرا انسان با نگاه به یک منظره برای نقاط مختلف ارزش‌های متفاوتی قائل می‌شود و اصلاً به بعضی از نقاط توجه نمی‌کند؟! یا به عبارتی چرا انسان به سرعت هدف اصلی و مهم صحنه را با اشتباه بسیار کم تشخیص می‌دهد حتی اگر این هدف در ابعادی کمتر از چند ده پیکسل باشد؟! بسیار بدیهی است که اگر بتوان براساس چنین توانایی جالبی از انسان سیستمی طراحی کرد که قابلیت تشخیص نقاط مهم صحنه را داشته باشد آنگاه به راحتی می‌توان به سرعت بیشتری در آشکارسازی دست یافت چرا که نیازی نیست به نواحی کم اهمیت توجهی کنیم. مثلاً استفاده از این ایده که انسان‌ها معمولاً در روند آشکارسازی چهره به صورت طبیعی به مرکز تصویر توجه می‌کنند و... می‌تواند باعث بهبود در روند استخراج ویژگی از تصویر شود.

## مراجع

[1] Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to

است. چنین انتخابی، غیرهوشمندانه بوده و باعث اتلاف زمان آشکارسازی می‌شود. در حالت مقایسه دو الگوریتم آشکارسازی، با توجه به اینکه عوامل تأخیر در نظر گرفته نشده، میزان بهبود سرعت اجرای الگوریتم برای ۲۷۰ تصویر، به حدود ۳,۵۵ برابر می‌رسد.

جدول ۶: مقایسه نتایج آشکارسازی بین روش پیشنهادی و روش ویولاجونز

نتایج دو سیستم	سرعت سیستم ویولاجونز (ثانیه)	سرعت سیستم پیشنهادی (ثانیه)	میزان دقت روش پیشنهادی (درصد)	میزان دقت روش ویولاجونز (درصد)	شماره تصویر
۱	۳,۳۶	۲,۰۰	۱۰۰	۱۰۰	۱۰۰
۲	۰,۹۳	۱,۲۰	۱۰۰	۱۰۰	۱۰۰
۳	۳,۵۹	۱,۷۱	۱۰۰	۱۰۰	۱۰۰
۴	۳,۶۲	۱,۹۶	۱۰۰	۱۰۰	۱۰۰
۵	۳,۴۶	۱,۹۳	۱۰۰	۱۰۰	۱۰۰
۶	۳,۵۹	۱,۶۹	۵۰	۱۰۰	۱۰۰
۷	۳,۷۱	۱,۷۲	۱۰۰	۱۰۰	۱۰۰
۸	۳,۸۹	۱,۸۸	۱۰۰	۱۰۰	۱۰۰
۹	۳,۶۵	۱,۶۶	۱۰۰	۱۰۰	۱۰۰
۱۰	۳,۸۱	۱,۸۲	۰	۰	۰
۱۱	۳,۴۵	۱,۶۵	۱۰۰	۰	۱۰۰
۱۲	۰,۸۹	۰,۹۱	۱۰۰	۱۰۰	۱۰۰
۱۳	۰,۹۸	۱,۰۲	۰	۱۰۰	۰
۱۴	۵۲/۳	۶۰/۱	۰	۳۳	۰
۱۵	۳,۳۵	۱,۶۷	۱۰۰	۱۰۰	۱۰۰
۱۶	۰,۸۸	۰,۹۳	۰	۰	۰
۱۷	۳,۶۵	۱,۷۵	۱۰۰	۱۰۰	۱۰۰
۱۸	۳,۷۱	۱,۶۳	۱۰۰	۱۰۰	۱۰۰
۱۹	۳,۶۱	۱,۸۰	۱۰۰	۱۰۰	۱۰۰
۲۰	۳,۷۸	۱,۹۲	۰	۰	۰

گرچه عنوان این تحقیق، افزایش سرعت آشکارسازی چهره است، اما باید توجه داشت، این هدف تنها کاربردی از روش پیشنهادی است، استفاده از ایده تکرار در طبیعت و بررسی نظم و ساختارهای موجود در پدیده‌های طبیعی با توجه به ذات تصاویر



**علی سلمانی** مدرک کارشناسی خود را در سال ۱۳۹۱ در رشته مهندسی برق-الکترونیک از دانشگاه تربیت معلم سبزوار دریافت کرد و در سال ۱۳۹۴ تحصیلات مقطع کارشناسی ارشد خود را در رشته مهندسی برق - مخابرات گرایش سیستم در دانشگاه فردوسی مشهد به پایان رساند. زمینه‌های تحقیقاتی مورد علاقه نامبرده مخابرات تصویری، پردازش تصویر و ویدیو، بینایی ماشین، هوش مصنوعی و شناسایی الگو است.



**مرتضی خادمی** مدرک کارشناسی خود را در گرایش الکترونیک از دانشگاه صنعتی اصفهان اخذ نمود (سال ۱۳۶۴). نامبرده تحصیلات خود را در همان دانشگاه ادامه داده و موفق شد کارشناسی ارشد را در گرایش مخابرات به پایان برساند (۱۳۶۶). نامبرده بلافاصله پس از فراغت از تحصیل در دانشگاه فردوسی مشهد به

عنوان عضو هیأت علمی و با درجه مربی استخدام گردیده و به مدت چهار سال به امر تدریس اشتغال ورزید. در سال ۱۳۷۰ با بورس دانشگاه برای ادامه تحصیل به کشور استرالیا عزیمت نموده و در سال ۱۳۷۴ با اخذ مدرک Ph.D. به میهن اسلامی مراجعت نمود. هم‌اکنون نامبرده به عنوان استاد گروه برق دانشگاه فردوسی مشهد مشغول به کار می‌باشد. زمینه‌های تحقیقاتی مورد علاقه نامبرده، مخابرات تصویری، فشرده سازی تصویر و صوت و پردازش سیگنالهای پزشکی می‌باشد. وی تاکنون بیش از صد مقاله در کنفرانسها و مجلات داخل و خارج کشور در زمینه‌های فوق به چاپ رسانیده است.

- boosting". Journal of Computer and System Sciences, 55(1):119{139, 1997.
- [2] P.Viola, M.Jones, "Robust real-time face detection," International Journal of Computer Vision, vol. 57, no. 2, pp. 137–154, 2004.
- [3] M.Dixon, F.Heckel, R.Pless, and W.D.Smart, "Faster and more accurate face detection on mobile robots using geometric constraints," IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1041–1046, 2007.
- [4] H.Wu, K.Suzuki, T. Wada, and Q. Chen, "Accelerating face detection by using depth information," in Advances in Image and Video Technology vol. 5414 of Lecture Notes in Computer Science, pp. 657–667. Springer Berlin Heidelberg, 2009.
- [5] W. Burgin, C.Pantofaru, and W.D.Smart. "Using depth information to improve face detection". in proceedings of the 6th acm/IEEE international conference on human-robot Interaction, lausanne, switzerland, 2011.
- [6] S.Bodiroza, "Using image depth information for fast face detection". toward autonomous robotic systems, vol.6856,pp 424–425, 2011.
- [7] R.-I.Hg, P.Jasek, C.Rofidal, K.Nasrollahi, T.B. Moeslund, and G.Tranchet, "An RGB-D Database Using Microsoft's Kinect for Windows for Face Detection", Laboratory of Visual Analysis of People, Aalborg University, Aalborg, Denmark, 2012.
- [8] L.Nanni, A.Lumini, F.Dominio, P.Zanuttigh, "Effective and precise face detection based on color and depth data" Journal of Applied Computing and Informatics, 2014.
- [9] G.P.Meyer, S.Alfano, M.N.Do, "Accelerating face detection with rgb-d images". IEEE International conference on image processing, IEEE icip in Paris, 2014.
- [10] A.papoulis, S.pillai, "Probability, Random Variables and stochastic Processes" McGraw-Hill, Fourth Edition, 2002.