

تخمین ژست سر با استفاده از استخراج ناحیه‌ی بیضی‌گون در تصاویر سه‌بعدی چهره

و ویژگی‌های مبتنی بر هندسه سه‌بعدی

مسعود هنریار^۱، عاطفه اسدی^۲

^۱ مربی برق الکترونیک، دانشگاه آزاد اسلامی، واحد شهرری

^۲ کارشناسی ارشد برق الکترونیک، دانشگاه آزاد اسلامی، واحد شهرری، atefe.asadi@yahoo.com

تاریخ دریافت: ۹۳/۲/۲۱ تاریخ پذیرش: ۹۴/۵/۱

چکیده

تخمین ژست چهره یک فرایند بسیار مهم از شناسایی چهره است. مطالعه برای یافتن یک روش تمام خودکار و توانمند که قابل پاسخ برای ژست‌های گوناگون در تمام جهات باشد یک چالش واحد را برای تکنیک‌های موجود در سیستم‌های بینایی ماشین به وجود آورده است. همچنین در مقایسه با شناسایی و آشکارسازی چهره که کانون مقدماتی تحقیقات بینایی مربوط به چهره است، تخمین زاویه چهره به عنوان یک مرحله‌ی میان پردازشی بسیار مهم به حساب می‌آید. در این مقاله یک روش جدید ارائه شده است که در آن ابتدا ناحیه‌ی بیضی‌گون چهره از داده‌ی سه‌بعدی سر استخراج شده و سپس بردارهای ویژگی که به شدت به درک انسان از ژست سر تاثیر می‌گذارند، به عنوان سرنخ‌هایی کاملاً برجسته و نمایان از چهره در خصوص جهت‌گیری سر استخراج می‌شوند. در این مقاله از طبقه‌بندی SVM، روش شبکه عصبی RBF و روش K نزدیکترین همسایگی KNN برای آموزش بردارهای استخراج شده جهت قیاس و تمایز ویژگی‌ها و طبقه‌بندی داده‌ها استفاده شده است. این الگوریتم بر روی پایگاه‌های داده‌ی Frav3D و Gavab آزمایش شده و کسب نتیجه‌ی بیشینه میانگین تشخیص صحیح ۹۸/۴۸ درصد که با استفاده از روش KNN حاصل شده است، بیانگر دقت الگوریتم پیشنهادی می‌باشد.

کلیدواژه

تخمین ژست سر، تصاویر عمق، استخراج ویژگی، طبقه‌بندی.

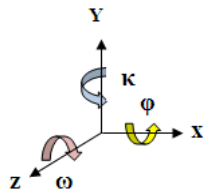
مقدمه

می‌باشد. اخیراً در تحقیقات بینایی ماشین رهیافت‌هایی با استفاده از داده سه‌بعدی برای غلبه بر کاستی‌های تصاویر دوبعدی توسعه و فزونی یافته است [۱-۵]. در واقع زمانی که از تصاویر دوبعدی برای تشخیص خودکار حالت چهره استفاده می‌شود، بخش مهمی از اطلاعات موجود در چهره مانند برآمدگی‌ها یا تورفتگی‌های سطح چهره که در اثر تغییر حالت به وجود می‌آید، از دست داده می‌شود. از طرفی، داده‌های تصاویر دوبعدی که با استفاده از نور بازگشتی از چهره جمع‌آوری می‌شوند، به نور محیط و جهت نور تابیده شده به چهره وابستگی زیادی دارند که به آن تغییر شدت روشنائی می‌گویند و همچنین از آنجاییکه چهره انسان حالت سه‌بعدی دارد، وقتی تحت زوایای مختلف تصویر برداری شود، به تصاویر متفاوتی منجر خواهد شد که به این مسئله تغییر زاویه‌ی چهره می‌گویند.

با پیشرفت تکنولوژی و ورود ربات‌ها از فعالیت‌های صنعتی به فعالیت‌های اجتماعی، تشخیص اتوماتیک حالت چهره به سبب کاربردش در تعاملات انسان و ربات به یکی از موضوعات تحقیقاتی مهم در پردازش تصویر مبدل شده است. انگیزه گرایش به این موضوع از اینجا ناشی می‌شود که نیاز به تخمین زاویه‌ی سر، بهنجارسازی و استخراج ویژگی از داده‌های کسب شده از چهره، پیش زمینه‌ای بسیار مهم در بسیاری از کاربردهای مرتبط با برهم کنش انسان و کامپیوتر است که مهمترین آنها شناسایی چهره می‌باشد.

تحقیقات روانشناسی نشان می‌دهد که سیستم بینایی انسان می‌تواند ویژگی‌های سه‌بعدی چهره را حتی وقتی به یک تصویر دوبعدی نگاه می‌کند استخراج کند، از اینرو تشخیص انسان از حالت چهره در یک تصویر دوبعدی بسیار بهتر از تشخیص ماشین

(یعنی حرکت عقب و جلو از گردن) زاویه‌ی $60/4^\circ$ تا $96/6^\circ$ را دربرمی‌گیرد، و در یک خمش جانبی از جلو (یعنی خمش راست به چپ از گردن) زاویه‌ی $40/9^\circ$ تا $36/3^\circ$ را دربرمی‌گیرد، و در یک دوران محور افقی (یعنی دوران راست به چپ از سر) از $79/8^\circ$ تا $75/3^\circ$ را در بر می‌گیرد [۹]. ترکیب دوران عضلانی و جهت‌گیری نسبی در فضا غالباً باعث ایجاد یک ابهام می‌شود، در واقع یک نمای مقطع جانبی از سر در مقایسه با وقتی که دوربین از نمای جلو تصویربرداری می‌کند و سر از پهلو چرخیده شده باشد، دقیقاً همان نمایی نیست که دوربین از آن ضلع تصویربرداری می‌کند. با وجود این مشکل، اغلب فرض می‌شود که سر شخص می‌تواند بعنوان یک شی صلب تجزیه شده در نظر گرفته شود. با این فرض سر شخص به ۳ درجه‌ی آزادی^۴ (DOF) در ژست محدود می‌شود که می‌توان بوسیله‌ی زاویه حول z (Roll)، زاویه حول y (Yaw) و زاویه حول محور x (Pitch) توصیف شود، در شکل ۱ دوران سر در سه جهت نشان داده شده است که ω ، κ و ϕ به ترتیب زوایای دوران حول محورهای z ، y و x می‌باشند.



شکل ۱. دوران چهره به ترتیب از راست حول محور z (roll)، y (yaw) و x (pitch).

مروری بر کارهای گذشته

الگوریتم‌های تخمین ژست سر در معیارهای متفاوتی طبقه‌بندی می‌شوند. یک معیار مهم، طبقه‌بندی کارکردی است که هر روش را با توجه به دامنه‌ی عملکردش دسته‌بندی می‌کند. در مقابل یک طبقه‌بندی کارکردی، سیستم‌ها با یک رهیافت مقدماتی که زمینه‌ساز پیاده‌سازی آنها است نیز دسته‌بندی می‌شود. مسئله‌ی مهم دیگر درجه‌ی خودکار بودن سیستم است. برخی از سیستمها ژست سر را به صورت تمام خودکار تخمین می‌زنند [۱۰]. اما در روش‌های دیگری که به روش‌های سنتی نیمه خودکار معروفند، پیش نیازهای چالش‌برانگیزی مانند مکان‌هایی از ویژگی‌های چهره باید از قبل شناخته شود [۲]. در این سیستم‌ها چهره‌ها را بوسیله استخراج علائم مشخصی یا ویژگی‌هایی از یک تصویر از صورت اشخاص نمادگذاری می‌کنند. این علائم اغلب بصورت دستی بر روی چهره قرار داده می‌شود. برای مثال، یک

اکثر روش‌های پیشنهادی برای شناسایی چهره، بر روی تصاویر تمام‌رخی کار می‌کنند که در محیط‌های آزمایشگاهی و با شرایط نورپردازی ثابت، تهیه شده‌اند. اما حقیقت این است که فرض تمام رخ بودن چهره یا ثابت بودن روشنایی، واقع‌گرایانه نمی‌باشد و این روش‌ها در شرایط غیر آزمایشگاهی نتایج جالب توجهی بدست نخواهند آورد. زیرا مردم بندرت احساسات خود را بدون حرکت ناخودآگاه سر و بدن بیان می‌دارند، از طرف دیگر، چرخش سر باعث تغییر شدت روشنایی خواهد شد که بخشی از چهره را غیر قابل رویت خواهد کرد. در تصاویر سه‌بعدی چهره حاوی اطلاعات عمق چهره می‌باشند، لذا نوع داده‌ها مستقل از شدت روشنایی و چرخش چهره می‌باشد. از اینرو استفاده از این نوع داده‌ها به عنوان جایگزینی برای داده‌های دوبعدی چهره به یک چالش جدید تبدیل شده است.

در این مقاله یک روش جدید برای تخمین ژست سه‌بعدی چهره ارائه شده است که در آن ابتدا داده‌های اضافی که منجر به ایجاد خطا در معرفی چهره به سیستم بینایی ماشین می‌شود حذف می‌شود و ناحیه‌ی چهره از داده سه‌بعدی سر که شامل موها و گردن است استخراج می‌شود و سپس از نقاط اکسترمم منحنی سطح چهره برای تعیین بردارهای ویژگی استفاده می‌شود. ویژگی تعیین شده توسط روش‌های ماشین‌های بردار پشتیبان^۱ (SVM) [۶]، روش شبکه‌های عصبی مبتنی بر توابع پایه‌ای شعاعی^۲ (RBF) [۷] و روش K نزدیکترین همسایگی^۳ KNN [۸] آموزش داده شده و نتایج با هم مقایسه شده است.

ساختار این مقاله به شرح زیر تنظیم شده است. به دنبال این بخش در بخش بعدی تخمین زاویه سر در بینایی ماشین تعریف شده است. در ادامه به مروری بر کارهای گذشته پرداخته شده است و سپس پایگاه داده معرفی شده است. الگوریتم پیشنهادی به ترتیب در زیر بخش‌های پیش پردازش تصویر، انتخاب نقاط ویژگی و طبقه‌بندی شرح داده شده است. در نهایت نتایج شبیه‌سازی و نتیجه‌گیری آورده شده است.

تعریف تخمین زاویه‌ی سر در بینایی ماشین

در محتوای بینایی ماشین، تخمین زاویه‌ی سر غالباً بعنوان توانایی تفسیر زاویه‌ی سر شخص در مقایسه با زاویه از یک دوربین بیان می‌شود. با نگاه دقیق‌تر به این مسئله می‌توان گفت که تخمین زاویه‌ی سر توانایی تفسیر زاویه‌ی سر در مقایسه با یک سیستم مختصاتی عمومی می‌باشد، اما این تفاوت دقیق در تعریف مسئله نیاز به اطلاعات پارامترهای ذاتی دوربین برای درک انحراف زاویه‌ای دورنمای سه‌بعدی دارد. رنج زاویه‌ی حرکت سر برای یک شخص مرد بزرگسال بالغ، در یک خمش سهمی و انبساط کششی

1 Support Vectors Machine
2 Radial Basis Function
3 k-Nearest Neighbors

4 Degree of Freedom

آشکارسازی چشم‌ها مبتنی بر اطلاعات تراکم لبه است که بر اساس این مشاهده است که بیشترین تراکم لبه پیرامون نواحی چشم با نواحی دیگر در چهره مقایسه شده است. روش دوم یک روش ترکیبی است که اطلاعات رنگ، لبه و روشنایی را برای آشکارسازی چشم‌ها ترکیب می‌کند. هنگامیکه گوشه‌های چشم‌ها و دهان آشکارسازی شد، سپس مراکز آنها با استفاده از معادلاتی که بر اساس نرخ طول دو چشم است محاسبه می‌شود. ژست‌ها در این روش تا حدی قابل شناسایی هستند که در چرخش سر چشم‌ها از تصویر خارج نشده باشند.

الگوریتم دوم روش پیشنهادی آقای Tobias Bar و همکارانش می‌باشد [۵]. در این الگوریتم ژست سر با تخمین خط نگاه چشم در حالت سه بعدی با استفاده از الگوریتم ICP^۵ تخمین زده می‌شود. تکنیک ICP دارای عیوبی است که باعث کاهش راندمان آن در هنجارسازی داده‌ی سه‌بعدی می‌شود. در واقع معیار چرخش ICP زمانی خوب جواب می‌دهد که نقاط به یک تطبیق بسیار خوب نزدیک بشوند. اگر دو مجموعه داده، به صورت تقریبی در یک تنظیم اولیه نباشند این تکنیک در معرض گیر افتادن در یک کمینه‌ی محلی قرار می‌گیرد.

در روش ارائه شده در این مقاله ابتدا با اعمال پیش پردازش در چند مرحله ناحیه‌ی چهره از سر استخراج می‌شود و داده‌های اضافی مانند مو و گردن تا حد زیادی حذف می‌شوند. سپس با انتخاب ویژگی مناسب و آموزش طبقه‌بندی در روش‌های ذکر شده، ژست سر تخمین زده می‌شود. الگوریتم پیشنهادی با دو الگوریتم مراجع [۵] و [۱۶] در بخش نتایج شبیه‌سازی مورد مقایسه قرار گرفته است.

پایگاه داده

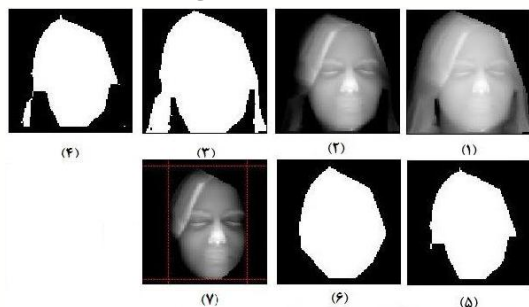
پایگاه‌های داده‌ای که در این مقاله از آنها استفاده شده است پایگاه داده‌ی Frav3D [۱۹] و پایگاه داده‌ی Gavab [۲۰] می‌باشند. پایگاه داده‌ی Frav3D شامل ۱۰۶ فرد متفاوت است که به ازای هر فرد ۱۶ تصویر عمق و بافت متفاوت وجود دارد. در این الگوریتم تنها از داده عمق استفاده شده است. در واقع ۷ ژست سه‌بعدی متفاوت برای هر فرد وجود دارد که شامل تصاویر ژست روبرو با حالت عادی (۸ تصویر)، تصاویر دارای چرخش حول محور y (نگاه به چپ و راست هرکدام دو ژست)، تصاویر دارای چرخش حول محور z (دو تصویر چرخش به سمت راست و چپ)، دو تصویر دارای چرخش حول محور x (دو تصویر نگاه به بالا و پایین) می‌باشد. پایگاه داده‌ی Gavab شامل تصاویر سه بعدی چهره از ۵۴۹ نفر می‌باشد. تصاویر از ۶۱ نفر (۴۵ مرد و ۱۶ زن) گرفته شده‌اند که به ازای هر فرد ۹ تصویر قرار داده شده است. ۹ تصویری که به ازای هر فرد قرار داده شده است شامل دو تصویر

الگوریتم ممکن است مکان مشخص شده، اندازه و یا شکل چشم‌ها، بینی، استخوان گونه و فک را تحلیل کند. این ویژگی‌ها سپس برای جستجوی علائم از تصاویر دیگر از طریق تطبیق ویژگی استفاده می‌شوند.

بطور کلی رهیافت‌های تخمین ژست سر به ۸ دسته ذیل طبقه‌بندی می‌شود:

- **روش‌های الگوی ظاهر:** در این دسته از روش‌ها یک تصویر جدید از یک سر با یک مجموعه از نمونه‌هایی که هر یک با یک ژست مجزا برچسب‌گذاری می‌شود برای یافتن بیشترین شباهت به آن مقایسه می‌شود [۱۱].
 - **روش‌های آرایه آشکارساز:** در این دسته از روش‌ها یک سلسله از آشکارسازهای سر که هر یک با یک ژست ویژه وفق داده شده است و به آن اختصاص یافته با یک آشکارساز با بیشترین پشتیبانی تعلیم داده می‌شود [۱۲].
 - **روش‌های رگرسیون غیرخطی:** در این دسته از روش‌ها از ابزارهای رگرسیون غیرخطی برای گسترش یک نگاشت تابعی از تصویر یا داده ویژگی به یک مقدار اندازه‌گیری ژست سر استفاده می‌شود [۱۳].
 - **روش‌های تعیبه‌ی مانیفولد:** در این روش‌ها مانیفولدهایی با ابعاد پایین که تغییرات پیوسته ژست سر را مدلسازی می‌کنند محاسبه می‌شود. تصاویر جدید می‌توانند در این مانیفولدها جاگذاری شوند و سپس برای تطبیق نمونه‌های جاگذاری شده یا رگرسیون استفاده شوند [۱۴] و [۱۵].
 - **مدل‌های انعطاف پذیر:** مدل‌های انعطاف‌پذیر یک مدل غیرصلب را به ساختار چهره‌وار هر شخص در سطح تصویر برازش می‌کند. ژست سر از مقایسه‌های سطح ویژگی یا از نمونه‌ای از پارامترهای مدل، تخمین زده می‌شود [۳].
 - **روش‌های هندسی:** روش‌های هندسی از مکان‌های ویژگی‌هایی مانند چشم، دهان و نوک بینی برای تعیین ژست از پیکربندی منسوب به آنها استفاده می‌کند [۴] و [۱۶].
 - **روش‌های ردیابی:** روش‌های ردیابی تغییر کلی ژست سر را از روی تغییر مکان مشاهده شده بین فریم‌های ویدیو بازیابی می‌کنند [۱۷].
 - **روش‌های ترکیبی:** در آن دسته از روش‌ها یک یا چند تا از روش‌های ذکر شده برای غلبه بر محدودیت‌های ذاتی در هر یک از رهیافت‌ها با هم ترکیب می‌شوند [۱۸].
- الگوریتم پیشنهادی با دو الگوریتم جدید مورد مقایسه قرار گرفته است. الگوریتم اول یک روش هندسی است که توسط آقای M. Shafi و همکارانش ارائه شده است [۱۶]. در این روش بر برجسته‌ترین ویژگی‌های چهره مانند چشم‌ها و دهان تکیه شده است. در این بخش دو روش برای مکانیابی چشم‌ها و یک روش برای آشکارسازی دهان شرح داده شده است. یکی از روش‌های

از ناحیه‌ی سر و حذف داده‌های اضافی دارای اهمیت زیادی است. در مرحله‌ی پیش پردازش برای بالا بردن قابلیت اطمینان و توانمند بودن الگوریتم، داده‌های اضافی در چندین مرحله از چهره حذف می‌شوند. مراحل پیش پردازش تصویر سه بعدی در شکل ۳ نشان داده شده است. در این شکل، مراحل استخراج چهره از تصویر سر یک زن با موهای بلند دیده می‌شود.

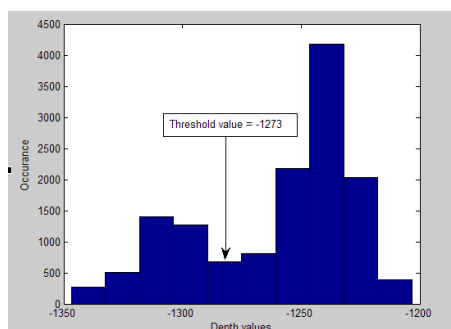


شکل ۳. مراحل پیش پردازش تصویر سه بعدی و حذف داده‌های اضافی (موها و گردن).

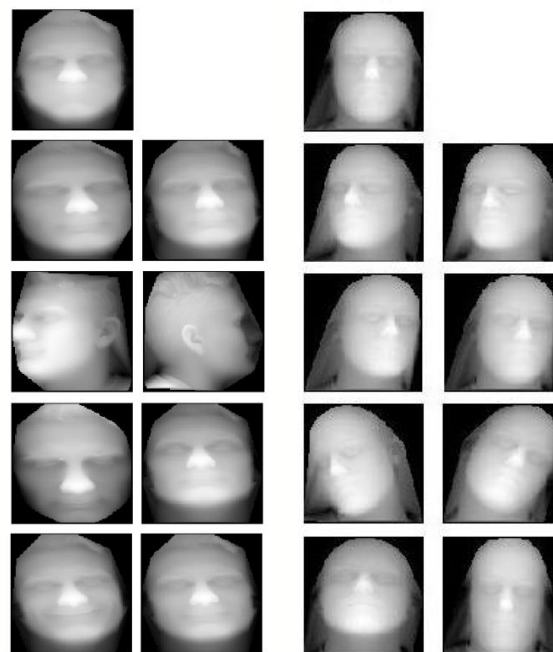
شکل ۱-۳ تصویر اصلی می‌باشد. برای حذف داده‌های اضافی از چهره در اولین مرحله، الگوریتم آستانه‌گذاری OTSU [۲۱] به کار برده شده است. OTSU یک معیار را پیشنهاد کرده که در آن با کمینه کردن انحراف درون کلاسی که بعنوان یک مجموع وزن دار شده از انحراف دو کلاس تعریف شده، یک حد آستانه T را جستجو می‌کند. با توجه به رابطه‌ی زیر:

$$\sigma_w^2(k) = c_1(k) \cdot \sigma_1^2(k) + c_2(k) \cdot \sigma_2^2(k) \quad (1)$$

$c_1(k)$ و $c_2(k)$ به ترتیب مجموع احتمالات هیستوگرام کلاس اول و دوم با مقادیر خاکستری $[0, \dots, k]$ و $[k+1, \dots, I]$ می‌باشند، σ_1^2 و σ_2^2 به ترتیب واریانس‌های دو کلاس C_1 و C_2 هستند و در نهایت σ_w^2 واریانس درون کلاسی می‌باشد که باید کمینه شود. در واقع OTSU نشان داده است که کمینه کردن انحراف درون کلاسی همان بیشینه کردن انحراف میان کلاسی است. در شکل ۴ هیستوگرام عمق تصویر چهره‌ی شکل ۱-۳، مقدار آستانه‌ی منتج و نیز هیستوگرام تصویر بعد از اعمال حد آستانه نشان داده شده است.



روبرو با حالت طبیعی، دو تصویر با چرخش حول محور x ها با حالت عادی (نگاه به بالا و پایین)، دو تصویر با چرخش حول محور y ها با حالت عادی (نگاه به چپ و راست)، سه تصویر روبرو با حالات چهره متفاوت شامل خنده، لبخند و یک حالت به انتخاب فرد می‌باشد. نمونه‌هایی از ژست‌های مختلف دو فرد در پایگاه داده‌های Frav3D و Gavab در شکل ۲ آورده شده است.



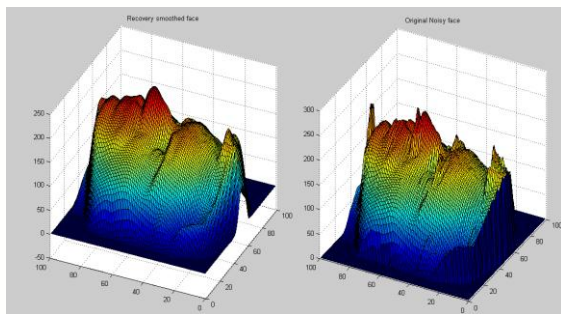
شکل ۲. نمونه‌های کامل از ژست‌های مختلف یک فرد در پایگاه داده‌ی Frav3D (دو ستون سمت راست)، پایگاه داده‌ی Gavab (دو ستون سمت چپ).

شرح الگوریتم پیشنهادی

روش پیشنهادی در الگوریتم ارائه شده دارای سه مرحله‌ی پیش پردازش، استخراج ویژگی و آموزش بردارهای ویژگی با استفاده از روش‌های طبقه‌بندی SVM، شبکه عصبی RBF و روش KNN می‌باشد. در مرحله‌ی پیش پردازش تصویر صورت از ناحیه‌ی سر در چند مرحله استخراج شده و داده‌های نویزی نیز از آن حذف می‌شوند. در مرحله‌ی استخراج ویژگی نقاط اکستریم و نوک بینی از تصویر صورت مکانیابی شده و مجموع فاصله‌ی اقلیدسی نقاط اکستریم در نواحی مختلف نسبت به موقعیت نوک بینی به عنوان بردار ویژگی آموزش داده می‌شود. در ادامه جزئیات مراحل پیشرفت الگوریتم شرح داده می‌شود.

پیش پردازش تصویر

مرحله‌ی پیش پردازش در این الگوریتم به دلیل توانمند بودن در حذف داده‌های نویزی، موجب جامعیت بخشیدن به پاسخ‌گویی آن در برابر چهره‌های مختلف در ژست‌های متفاوت شده است. از آنجاییکه در تصویر سه بعدی فقط اطلاعات عمق موجود است، داده‌های اضافی مانند موها و گردن با محدوده‌ی اصلی چهره بافت یکسانی دارد. از اینرو بکارگیری روش مناسب برای استخراج چهره



شکل ۵. هموارسازی تصویر با استفاده از روش پیشنهادی مرجع [۲۲] تصویر نویزی (سمت راست)، تصویر بدون نویز (سمت چپ).

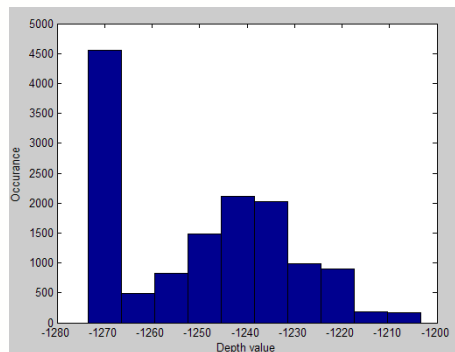
انتخاب نقاط ویژگی

انتخاب ویژگی‌های مناسب جهت رسیدن به نتایج مناسب هدف بعدی می‌باشد. در واقع استخراج ویژگی ایجاد مجموعه‌ای از توصیفات معنادار از یک تصویر است که برای پیدا کردن یک مجموعه از ویژگی‌های چهره با ابعاد کمتر در مرحله‌ی آموزش مورد استفاده قرار می‌گیرد. در این الگوریتم برای مدل کردن تصویر با یک بردار ویژگی، ابتدا نقاط برجسته‌ای مانند نقاط اکسترم که شامل نقاط عطف سه‌بعدی یعنی مینترم‌ها و ماکسترم‌ها هستند آشکارسازی می‌شوند. از آنجاییکه که قالب کلی چهره برای همه افراد دارای ساختار یکسانی می‌باشد، تعیین این نقاط دارای وجه تمایز قابل قبولی در تغییرات ژست‌های مختلف هستند. نتایج استخراج نقاط اکسترم در ژست‌های در شکل ۶ نشان داده شده است.

یکی از نقاط مهم و کلیدی در چهره که غالباً به عنوان مبنا در نظر گرفته می‌شود موقیت نوک بینی است. الگوریتم‌های مختلفی برای مکانیابی نوک بینی معرفی شده است. در مرجع [۱۵] با فرض اینکه بینی دارای بیشترین داده عمق در میان تمام نقاط می‌باشد، با استفاده از یک پنجره ۳ در ۳، مجموع مقدار عمق پیکسل‌های منطبق محاسبه می‌شود و مختصات پیکسل مرکزی از پنجره با بیشترین مقدار، بعنوان نوک بینی انتخاب می‌شود. آشکارسازی بینی در این روش در ژست‌هایی که سر حول محور x چرخش دارد با خطا همراه است. در این حالت در اکثر موارد در چرخش سر به سمت بالا و پایین به ترتیب چانه و پیشانی به عنوان بیشترین ارتفاع انتخاب می‌شوند. روشی که در این الگوریتم برای یافتن نوک بینی استفاده شده است، انتخاب نزدیکترین نقطه اکسترم به تمام نقاط اکسترم در یک ژست است. در واقع اگر برای هر ژست، مجموعه‌ای از نقاط اکسترم به صورت زیر تعریف شود:

$$P = \{P_1, P_2, \dots, P_m\} \quad (2)$$

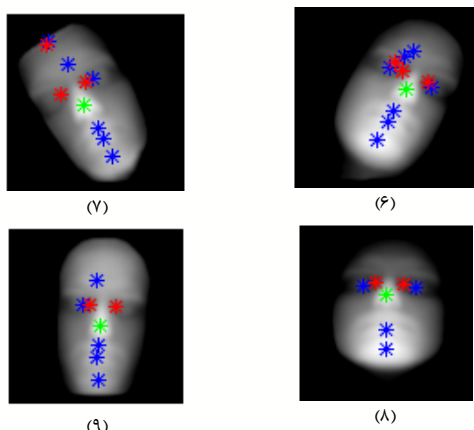
که P_i ها نقاط اکسترم با مختصات (x_i, y_i) می‌باشند و m تعداد نهایی نقاط برای ژست‌های مختلف متفاوت است. اگر d_{ij} فاصله اقلیدسی بین دو نقطه P_i و P_j به صورت زیر تعریف شود:



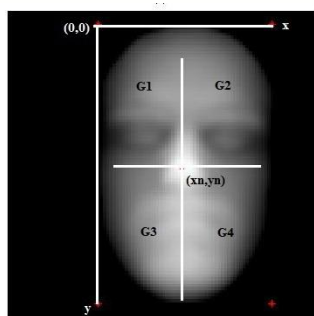
شکل ۴. هیستوگرام عمق تصویر و مقدار حد آستانه (تصویر بالا) و هیستوگرام تصویر بعد از اعمال آستانه‌گذاری OTSU (تصویر پایین).

نتایج اعمال آستانه‌گذاری بر روی تصویر اصلی در شکل ۲-۳ نشان داده شده است. در ادامه برای کاهش موهای بلند، تصویر آستانه‌گذاری شده به تصویر باینری تبدیل می‌شود (شکل ۳-۳). در این حالت با استفاده از عملیات شکل شناسی شاخه‌های اضافی از ناحیه بیضی‌گون چهره تا حد زیادی هرس می‌شوند (شکل ۳-۴) و سپس برای اطمینان از حذف کامل آنها، بزرگترین مؤلفه متصل از شکل به عنوان ناحیه‌ی چهره استخراج می‌شود (شکل ۳-۵) و همچنین برای جلوگیری از انتخاب داده‌های ناهموار مرزی به عنوان نقاط اکسترم در مرحله‌ی بعد، با اعمال عملیات مورفولوژی تولید سطح محدب، محدوده‌ی مرزی چهره از ناحیه‌ی سر با مرزهای هموارتری استخراج می‌شود (شکل ۳-۶). همانطور که در شکل ۳-۷ مشاهده می‌شود ناحیه‌ی صورت نسبت به داده‌ی اولیه در شکل ۳-۱ دارای اطلاعات دقیقتری از صورت می‌باشد.

مرحله‌ی بعدی در پیش پردازش تصویر سه بعدی، حذف داده‌های نویزی از تصویری است که توسط پویسگر سه‌بعدی Minotla در پایگاه داده‌ی Frav3D گرفته شده است. این نوع نویز در اثر خطای عملکرد سنسورهای دوربین تصویری وجود آمده است. در این نویز پیکسل‌های نویزی در یک همسایگی محدود، اغلب دارای بیشترین یا کمترین سطح خاکستری می‌باشند. اعمال فیلترهای خطی باعث مات شدن تصویر می‌گردند و در برخی از فیلترهای غیرخطی که تمام پیکسل‌های تصویر با یک مقدار میانه در یک همسایگی خاص جابجا می‌شود، جزئیات تصویر از بین برده می‌رود و وضوح تصویر را کاهش می‌یابد. روشی که برای حذف این نوع نویز در این الگوریتم استفاده شده روش هموارسازی مبتنی بر مرجع [۲۲] می‌باشد. این روش مبتنی بر روش کمینه کردن حداقل مربعات داده در میانگین‌های تبدیل فوریه گسسته‌ی تصویر انجام می‌گیرد. شکل زیر تصویر چهره را قبل و بعد از اعمال الگوریتم هموارسازی مرجع [۲۲] نشان می‌دهد.



شکل ۶. نقاط اکسترمم، مینترمها (نقاط قرمز رنگ)، ماکسترمها (نقاط آبی رنگ)، نوک بینی (نقاط سبز رنگ).



شکل ۷. ناحیه بندی تصویر نسبت به نوک بینی.

با توجه به موقعیت نقاط اکسترمم در ژستهای مختلف در شکل ۶ مشاهده می‌شود که در ژستهای روبرو نقاط اکسترمم حول موقعیت نوک بینی دارای پراکندگی و فاصله عمقی نسبتاً یکسانی می‌باشند (شکل‌های ۱-۳، ۸-۳ و ۹-۳)، در حالیکه در تصاویر شکل‌های ۲-۶ تا ۷-۶ این تفاوت در طرفین به شدت احساس می‌شود. در تصاویر شکل‌های ۲-۶ تا ۵-۶ که دوران حول محور y می‌باشد پراکندگی نقاط اکسترمم در نواحی G_4 و G_2 برای تصاویر شکل‌های ۲-۶ و ۳-۶ و همچنین در نواحی G_3 و G_1 برای تصاویر شکل‌های ۴-۶ و ۵-۶ بیشتر است. در دوران تصویر حول محور z پراکندگی نقاط در نواحی G_4 و G_1 برای تصویر شکل ۶-۶ (چرخش به سمت راست) و در نواحی G_3 و G_2 برای تصویر شکل ۶-۷ (چرخش به سمت چپ) بیشتر است. از اینرو برای استخراج ویژگی از مجموع فواصل اقلیدسی بین نقاط اکسترمم و محور $y=y_n$ در هر ناحیه استفاده می‌شود. همچنین برای جلوگیری از احتمال وجود نتایج مشابه در تصاویر ۲-۶ تا ۷-۶ بهتر است این پراکندگی برای هر ناحیه به صورت جداگانه محاسبه شود. همانطور که در روابط ۶ تا ۹ نشان داده شده است ویژگی‌های f_1, f_2, f_3 و f_4 به صورت مجموع فواصل اقلیدسی نقاط اکسترمم در هر ناحیه نسبت به کل فواصل تمام نقاط اکسترمم نسبت به خط $y=y_n$ محاسبه شده است.

$$d_{ij} = \|P_i - P_j\| \quad (3)$$

و S_i مجمع فواصل اقلیدسی برای هر نقطه اکسترمم از مابقی نقاط اکسترمم در هر ژست بصورت زیر تعریف شود:

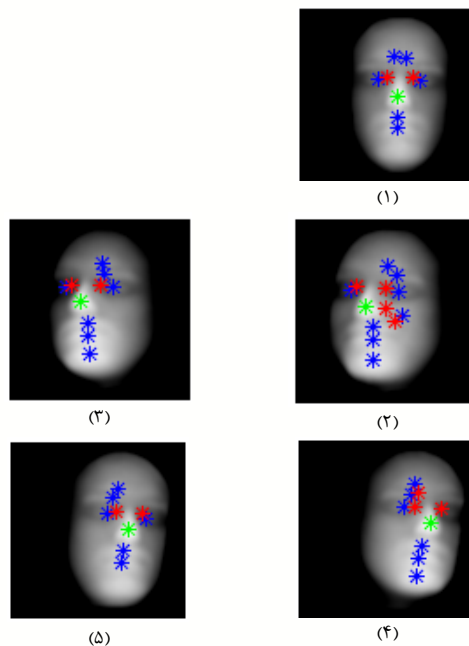
$$S_i = \sum_{j=1, j \neq i}^m d_{ij} \quad (4)$$

آنگاه می‌توان بردار مجموع فاصله برای تمام نقاط را به صورت زیر تعریف کرد:

$$S = [S_1, \dots, S_m] \quad (5)$$

شاخص کمترین مقدار آرایه در بردار S به عنوان شاخص نقطه مورد نظر برای موقعیت نوک بینی با مختصات (x_n, y_n) انتخاب می‌شود. در شکل ۶ نقاط اکسترمم شامل مینترمها، ماکسترمها و همچنین موقعیت نوک بینی برای هر ژست نشان‌گذاری شده است.

با توجه به شکل ۶ تغییرات در پراکندگی نقاط اکسترمم در نواحی مختلف تصویر با تغییرات ژست قابل مشاهده است. از اینرو قطعه‌بندی تصویر جهت دسته‌بندی نقاط اکسترمم به عنوان یک مرحله پیشنهاد شده است. در این حالت تصویر به ۴ ناحیه G_1, G_2, G_3 و G_4 همانطور که در شکل ۷ نشان داده شده است تقسیم می‌شود، سپس هر ژست با ویژگی‌های مجموع فاصله‌های اقلیدسی نقاط اکسترمم نسبت به محورهای گذرا از موقعیت نوک بینی مدلسازی می‌شوند.



در این ژست محاسبه‌ی گرادیان عمق در خط متقاطع بین صفحه‌ی $x=x_n$ (صفحه موازی محور y ها و گذرنده از نوک بینی) و سطح رویه چهره می‌باشد. قطعاً در تصاویر ژست چرخش به سمت بالا تعداد نقاط با گرادیان مثبت نسبت به تعداد نقاط با گرادیان منفی در نقاط بیشتر است و به همین ترتیب برای در تصاویر ژست چرخش به سمت پایین تعداد نقاط با گرادیان منفی بیشتر می‌باشد و در تصاویر روبرو نسبت این مقادیر با خطای کم با هم برابر هستند. اگر F مجموعه نقاط خط مذکور در نظر گرفته شود (رابطه‌ی ۱۰) و مجموعه G مجموعه گرادیان عمق نقاط مجموعه F در رابطه‌ی ۱۱ نظر گرفته شود، می‌توان در رابطه‌ی ۱۲ مقدار ویژگی f_5 را نسبت تعداد مقادیر منفی در مجموعه گرادیان G نسبت به کل نقاط خط در نظر گرفت.

(۱۰)

$$F = \{ (x_n, y_n, z_i) : 0 < i < \text{all points in } y \text{ axis} \}$$

(۱۱)

$$G = \nabla F = \left\{ g_1, g_2, \dots, g_{100} \right\}$$

(۱۲)

$$f_5 = \frac{\text{The number of negative points in } G}{\text{All Points on } y \text{ axis}}$$

و در نهایت بردار ویژگی برای هر تصویر به صورت زیر تعریف می‌شود:

$$\text{Feature_vector} = [f_1, f_2, f_3, f_4, f_5]$$

(۱۳)

طبقه بندی

فرایند طبقه‌بندی در واقع نوعی یادگیری با ناظر می‌باشد که در طی دو مرحله‌ی آموزش و آزمایش انجام می‌گردد. در مرحله آموزش مجموعه‌ای از داده‌ها که در آن هر داده شامل مقادیر ویژگی و یک خصوصیت بنام خصوصیت کلاس می‌باشد، برای ایجاد یک مدل بکار می‌روند که این داده‌ها توصیف کننده مفهوم و خصوصیات مجموعه داده‌هایی است که این مدل از روی آنها ایجاد شده است. مرحله‌ی آزمایش فرایند طبقه‌بندی است که اعمال مدل داده ایجاد شده بر روی داده‌هایی است که شامل تمام خصوصیات داده‌هایی است که برای ایجاد مدل داده بکار گرفته شده‌اند. هدف از عمل طبقه‌بندی تخمین مقدار کلاس می‌باشد. در این مقاله از روش‌های ماشین‌های بردار پشتیبان (SVM)، روش شبکه عصبی RBF و روش K نزدیکترین همسایگی KNN استفاده شده است. روش SVM به دلیل وجود حاشیه امن در تفکیک کلاس‌ها دارای توانمندی زیادی می‌باشد. SVM یک روش دسته‌بندی دوتایی^۶ می‌باشد. برای حل مسائل چند کلاسه باید از ترکیب دسته‌بندی دوتایی استفاده کرد. از اینرو می‌توان از روش

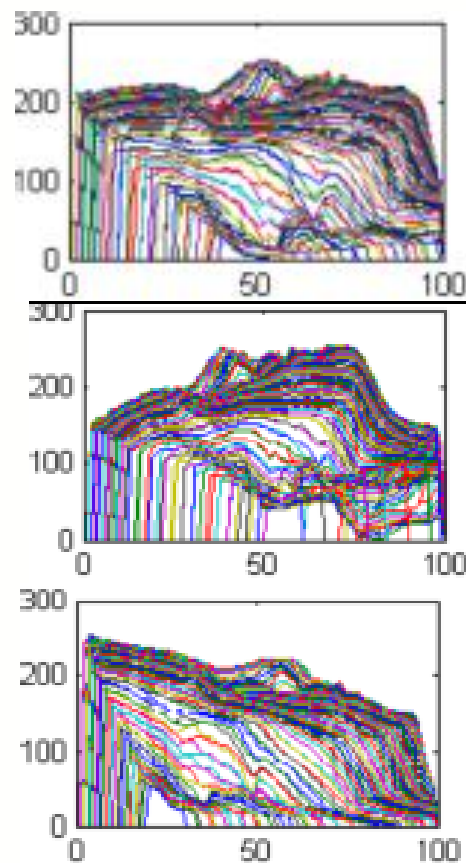
$$f_1 = \frac{\sum_{i \in \{G_1 \cap P\}} \|x_i - x_{nose}\|}{\sum_{i \in P} \|x_i - x_{nose}\|} \quad (۶)$$

$$f_2 = \frac{\sum_{i \in \{G_2 \cap P\}} \|x_i - x_{nose}\|}{\sum_{i \in P} \|x_i - x_{nose}\|} \quad (۷)$$

$$f_3 = \frac{\sum_{i \in \{G_3 \cap P\}} \|x_i - x_{nose}\|}{\sum_{i \in P} \|x_i - x_{nose}\|} \quad (۸)$$

$$f_4 = \frac{\sum_{i \in \{G_4 \cap P\}} \|x_i - x_{nose}\|}{\sum_{i \in P} \|x_i - x_{nose}\|} \quad (۹)$$

در شکل ۸ نمای $y-z$ تصاویر ۱-۶، ۸-۶ و ۹-۶ مورد بررسی قرار گرفته شده‌اند.



شکل ۸. نمای $y-z$ از تصویر نمای روبرو و تصاویر دوران داده شده حول محور x .

آنچه در شکل ۸ به وضوح مشهود است، تفاوت عمق بین دو نیمه تصویر حول صفحه $x=x_n$ می‌باشد. در واقع با چرخش سر به سمت بالا، نقاط پایین‌تر از موقعیت نوک بینی دارای عمق کمتری نسبت به نقاط بالایی می‌شوند و برعکس. در واقع در چرخش چهره به سمت بالا میزان عمق در هر نقطه نسبت به نقطه‌ی مجاور خود در حال افزایش می‌باشد و بالعکس در چرخش چهره به سمت پایین میزان این مقدار در حال کاهش می‌باشد. از اینرو بهترین ویژگی

چرخش به سمت بالا و به همین ترتیب تعداد ۶۰ بردار ویژگی به کلاس شش برای آموزش تصاویر با ژست چرخش حول محور x ها با چرخش به سمت پایین آموزش داده می‌شود. در نهایت تعداد $۷۳۶=۱۶ \times ۴۶$ تصویر از مابقی افراد به عنوان داده آزمایشی به طبقه‌بندها داده می‌شود.

پایگاه داده‌ی Gavab شامل ۵ ژست متفاوت می‌باشد و در نتیجه برای آن ۵ کلاس متفاوت در نظر گرفته می‌شود. نسبت داده‌های آموزشی به آزمایشی به نسبت تعداد افراد ۴۰ به ۲۱ در نظر گرفته شده است. از اینرو در این پایگاه داده $۳۶۰=۹ \times ۴۰$ نمونه آزمایشی و $۱۸۹=۹ \times ۲۱$ نمونه آزمایشی می‌باشد. از آنجاییکه در این مقاله تنها ژست داده‌ها مورد بررسی قرار گرفته است، در نتیجه تمام داده‌های روبرو که تعداد آنها $۲۰۰=۵ \times ۴۰$ است به عنوان کلاس صفر در نظر گرفته می‌شوند. همچنین تعداد $۴۰=۱ \times ۴۰$ نمونه برای هر یک از ژستهای تصاویر با چرخش حول محور y با نگاه به سمت راست و چپ و ژستهای تصاویر با چرخش حول محور x با نگاه به سمت بالا و پایین به ترتیب با برچسب کلاس یک، دو، سه و چهار در نظر گرفته می‌شود.

در این آزمایش نمونه‌های مربوط به هر نفر فقط در یکی از گروه‌های آموزشی و آزمایشی قرار گرفته است. همچنین انتخاب قرارگیری هر فرد در هر یک از مجموعه آموزشی و آزمایشی به صورت تصادفی و بدون ملاحظه‌ی خاصی صورت گرفته است. در آزمایش صورت گرفته، از جعبه ابزار `stprtool` برای پیاده‌سازی SVM چند کلاسه استفاده شده است. برای به دست آوردن جواب برای یک نمونه ورودی، نمونه مذکور به هر یک از دسته‌بندی کننده‌ها داده می‌شود. دسته‌بندی کننده‌ای که بزرگترین مقدار خروجی را تولید کند، کلاس نمونه ورودی را معین خواهد کرد. هسته‌های مورد استفاده در این تابع هسته‌ی چند جمله‌ای و هسته‌ی گوسین (RBF) می‌باشند. همچنین برای درجه‌ی چند جمله‌ای پارامتر P ، مقادیر ۱ تا ۴ مورد آزمایش قرار گرفته است و بهترین نتایج از درجات ۲ و ۳ حاصل آمده که در جدول ۱ نشان داده شده است. با توجه به نتایج نشان داده شده در جدول ۱، میانگین تشخیص صحیح در پایگاه داده‌ی Frav3D در روش SVM با هسته‌ی چندجمله‌ای با درجه‌ی ۲ مقدار $۹۴/۹۷$ درصد و با درجه‌ی ۳ مقدار $۹۶/۷۶$ درصد می‌باشد و همچنین میانگین تشخیص صحیح در پایگاه داده‌ی Gavab در این روش با درجه‌ی ۲ مقدار $۷۱/۲۴$ درصد و با درجه‌ی ۳ مقدار $۷۳/۱۸$ درصد می‌باشد.

یکی در مقابل همه^۷ و یا روش یکی در مقابل دیگری^۸ استفاده نمود. در این مقاله از روش یکی در مقابل همه استفاده شده است. RBF یک شبکه‌ی عصبی مصنوعی است که از توابع پایه‌ای شعاعی به عنوان توابع فعالیت استفاده می‌کند، خروجی این شبکه یک ترکیب خطی از توابع پایه‌ی شعاعی برای پارامترهای ورودی و نرونها است. شبکه‌ی RBF معمولاً شامل سه لایه است: لایه‌ی ورودی که یک بافر است، لایه دوم یا لایه‌ی مخفی که نگاشتی غیرخطی از فضای ورودی به فضایی با ابعاد بزرگ‌تر است و امکان جداسازی خطی الگوها را میسر می‌سازد و لایه خروجی که در اهداف دسته بندی از تابع سیگموئید استفاده می‌شود. همچنین یکی از بهترین طبقه‌بندها، طبقه‌بند KNN یا K نزدیکترین همسایگی است. این طبقه‌بند، نمونه آزمایشی را متعلق به کلاسی می‌داند که بیشترین آرا را در بین K نزدیکترین همسایگان آن داشته باشد. برای بدست آوردن نزدیکترین همسایگان یک نمونه، معمولاً از فاصله اقلیدسی استفاده می‌شود. طبقه‌بند KNN به دلیل قابلیت درک بالا و عدم نیاز به ایجاد فرضیه روی داده‌ها، روشی ساده و پرکاربرد محسوب می‌شود.

نتایج شبیه سازی

به منظور بررسی عملکرد الگوریتم پیشنهادی در این مقاله، در این قسمت نتایج شبیه سازی مدل ارائه شده با استفاده از نرم‌افزار Matlab پیاده‌سازی شده است.

به منظور تخمین ژست چهره ابتدا تصاویر ۱۰۶ نفر در پایگاه داده‌ی Frav3D به دو قسمت برای آموزش و آزمایش به نسبت تعداد افراد، ۶۰ به ۴۶ تقسیم‌بندی می‌شوند. در این روش تعداد ۷ دسته‌بندی کننده باینری به تعداد ژستهای مجزا ساخته می‌شود که هر کدام از آنها یکی از اعداد ۰ تا ۶ را از بقیه اعداد جدا می‌کند. تعداد ژستهای روبرو برای هر فرد ۸ تصویر در پایگاه داده می‌باشد که به طور مجموع $۴۸۰=۸ \times ۶۰$ بردار ویژگی به ازای هر فرد به کلاس صفر برای آموزش اختصاص داده می‌شود و تعداد $۱۲۰=۲ \times ۶۰$ بردار ویژگی به کلاس یک برای آموزش تصاویر با ژست چرخش حول محور y ها با نگاه به سمت راست و به همین ترتیب ۱۲۰ بردار ویژگی به کلاس دو برای آموزش تصاویر با ژست چرخش حول محور y ها با نگاه به سمت چپ آموزش داده می‌شود. تعداد $۶۰=۱ \times ۶۰$ بردار ویژگی به کلاس سه برای آموزش تصاویر با ژست چرخش حول محور x ها با چرخش به سمت راست و به همین ترتیب ۶۰ بردار ویژگی به کلاس چهار برای آموزش تصاویر با ژست چرخش حول محور z ها با چرخش به سمت چپ آموزش داده می‌شود. و همچنین تعداد $۶۰=۱ \times ۶۰$ بردار ویژگی به کلاس پنج برای آموزش تصاویر با ژست چرخش حول محور x ها با

جدول ۱. نتایج حاصل از تخمین ژست در روش SVM با هسته‌ی چندجمله‌ای. ۷۷/۶۶ درصد می‌باشد و همچنین میانگین تشخیص صحیح در پایگاه داده‌ی Frav3D در روش KNN میزان ۹۸/۴۸ درصد و در پایگاه داده‌ی Gavab میزان ۸۰/۴ درصد می‌باشد.

جدول ۳. نتایج حاصل از تخمین ژست با استفاده از شبکه عصبی RBF و روش KNN.

روش KNN		شبکه عصبی RBF		شماره ژست
Gavab	Frav3D	Gavab	Frav3D	
۹۹/۰۴	۹۹/۱۸	۹۸/۵۰	۹۸/۱۰	۰
۵۳/۰۶	۹۷/۸۳	۴۸/۹۸	۹۷/۸۳	۱
۵۵/۱۰	۹۸/۹۱	۵۱/۰۲	۹۷/۸۳	۲
۹۶/۸۶	۹۷/۸۳	۹۳/۸۸	۹۵/۶۵	۳
۹۷/۹۶	۹۸/۹۱	۹۵/۹۲	۹۷/۸۳	۴
-	۹۷/۸۳	-	۹۵/۶۵	۵
-	۹۸/۹۱	-	۹۷/۸۳	۶

با مقایسه‌ی نتایج میانگین تشخیص صحیح در جدول‌های ۱ تا ۳ مشاهده شد که بهترین نتیجه‌ی حاصل شده در روش KNN می‌باشد و روش‌های SVM با هسته‌ی RBF با سیگما برابر با ۴ و روش شبکه عصبی RBF در مراتب دوم و سوم قرار دارند. به منظور نمایش کارایی الگوریتم پیشنهادی، این روش با دو روش جدید مراجع [۱۶] و [۵] مورد مقایسه قرار گرفته است. نتایج پیاده‌سازی الگوریتم‌های مراجع مذکور بر روی پایگاه داده‌ی Frav3D در جدول ۴ آورده شده است.

جدول ۴. نتایج حاصل از تخمین ژست با استفاده از دو روش پیشنهادی مراجع بر روی پایگاه داده‌ی Frav3D.

نتایج میزان تشخیص (درصد)		شماره ژست
روش پیشنهادی مرجع M.Shafi [۱۶]	تکنیک مرجع ICP [۵]	
۸۸/۲۸	۸۳/۰۷	۰
۸۰/۲۱	۷۸/۱۳	۱
۸۱/۲۵	۷۹/۱۷	۲
۶۰/۴۲	۸۵/۴۲	۳
۶۲/۵۰	۸۷/۵۰	۴
۲/۶۴	۳۲/۰۸	۵
۱۲/۵۵	۴۳/۱۶	۶

با توجه به نتایج کسب شده در جدول ۴ مشاهده می‌شود که در این الگوریتم‌ها ژست‌های دوران داده شده حول محور x (چرخش سر به سمت پایین و بالا) تا حد زیادی جزء ژست‌های روبرو به حساب می‌آید، که در الگوریتم پیشنهادی به دلیل ارائه‌ی روش مستقل در این نوع ژست، این خطا به صورت قابل ملاحظه‌ای کاهش یافته است.

جدول ۲. نتایج حاصل از تخمین ژست در روش SVM با هسته‌ی چندجمله‌ای.

شماره ژست	نتایج میزان تشخیص (درصد) پایگاه داده Gavab		نتایج میزان تشخیص (درصد) پایگاه داده Frav3D	
	P=3	P=2	P=3	P=2
۰	۹۶/۵۰	۹۵/۰۰	۹۷/۱۴	۹۶/۰۹
۱	۴۴/۹۰	۴۲/۸۶	۹۶/۸۸	۹۵/۸۳
۲	۴۲/۸۶	۴۰/۸۲	۹۷/۹۲	۹۵/۸۳
۳	۹۱/۸۴	۸۹/۸۰	۹۵/۸۳	۹۱/۶۷
۴	۸۹/۸۰	۸۷/۷۶	۹۵/۸۳	۹۵/۸۳
۵	-	-	۹۷/۹۲	۹۵/۸۳
۶	-	-	۹۵/۸۳	۹۳/۷۵

در آزمایش دوم از روش SVM با هسته‌ی RBF استفاده شده است. نتایج آزمایش به ازای مقادیر مختلف پارامتر سیگما (σ) در هسته‌ی RBF در جدول ۲ آورده شده است. میانگین تشخیص صحیح در پایگاه داده‌ی Frav3D در روش SVM با هسته‌ی RBF با قرار دادن سیگما برابر با ۳ میزان ۹۵ درصد، در سیگما برابر با ۴ میزان ۹۷/۶۲ درصد و در سیگما برابر با ۵ میزان ۹۷/۴۷ درصد می‌باشد. همچنین میانگین تشخیص صحیح در پایگاه داده‌ی Gavab با قرار دادن سیگما برابر با ۳ میزان ۷۳/۳۸ درصد، در سیگما برابر با ۴ میزان ۷۵/۹۲ درصد و در سیگما برابر با ۵ میزان ۷۴/۵۰ درصد می‌باشد.

جدول ۲. نتایج حاصل از تخمین ژست در روش SVM با هسته‌ی RBF.

شماره ژست	نتایج میزان تشخیص (درصد) GavaB			نتایج میزان تشخیص (درصد) Frav3D		
	$\sigma=5$	$\sigma=4$	$\sigma=3$	$\sigma=5$	$\sigma=4$	$\sigma=3$
۰	۹۷/۰۰	۹۸/۰۰	۹۷/۵۰	۹۷/۹۲	۹۸/۹۶	۹۷/۶۶
۱	۴۶/۹۴	۴۸/۹۸	۴۴/۹۰	۹۶/۸۸	۹۷/۹۲	۹۵/۸۳
۲	۴۴/۹۰	۴۶/۹۴	۴۲/۸۶	۹۷/۹۲	۹۸/۹۶	۹۶/۸۸
۳	۹۱/۸۴	۹۱/۸۴	۸۹/۸۰	۹۵/۸۳	۹۵/۸۳	۹۳/۷۵
۴	۹۱/۸۴	۹۳/۸۸	۹۱/۸۴	۹۷/۹۲	۹۷/۹۲	۹۵/۸۳
۵	-	-	-	۹۷/۹۲	۹۷/۹۲	۹۷/۹۲
۶	-	-	-	۹۷/۹۲	۹۵/۸۳	۹۳/۷۵

برای پیاده‌سازی شبکه عصبی RBF از تابع newrb استفاده شده است که به تعداد نورونهای بیشتری در هر مرحله پرداخته است تا جایکه خطا از مقدار مورد نظر کمتر شود. در این پیاده‌سازی پارامتر spread که تعیین کننده‌ی محدوده‌ی انتخاب تابع گوسین است ۱۰ و خطای مورد نظر ۰/۱ در نظر گرفته شده است و در نهایت پارامتر K در پیاده‌سازی تابع KNN مقدار ۹ در نظر گرفته شده است. در جدول ۳ نتایج شبیه‌سازی شده در دو روش شبکه عصبی RBF و KNN نشان داده شده است.

میانگین تشخیص صحیح در پایگاه داده‌ی Frav3D در روش شبکه عصبی RBF میزان ۹۷/۲۴ درصد و در پایگاه داده‌ی Gavab میزان

کمترین زمان را نسبت به روش RBF به میزان 0.9507 ثانیه و نسبت به روش SVM، با هسته‌ی RBF به میزان $13/70$ ثانیه صرف کرده است.

همچنین غالباً، خطای پیدا کردن بینی که ناشی از موهای اطراف صورت و تغییرات ارتفاع در چرخش‌های سر بخصوص در چرخش سر حول محور x (چرخش به سمت پایین و بالا) به وجود می‌آید، در روش پیشنهادی به صورت قابل ملاحظه‌ای کاهش یافته است. همچنین استخراج ناحیه‌ی چهره از قالب سر به روش ذکر شده نوآوری به حساب می‌آید و استفاده از روش مرجع [۲۲] جهت هموارسازی سطح سه‌بعدی چهره برای اولین بار صورت گرفته است.

در بررسی پوشیدگی بخشی از چهره^۹ و در برابر تغییر حالات چهره^{۱۰} می‌توان ادعا کرد از آنجاییکه در این روش از اطلاعات بافت و رنگ چهره استفاده نشده است و از داده سه بعدی ویژگی استخراج شده است، در صورتی که انسداد دارای عمق متقارن باشد (مانند عینک و پوشاندن دهان)، با چرخش چهره نیز نقاط اکسترم به یک میزان تغییر می‌کنند و انسداد متقارن تغییر زیادی در نتیجه ایجاد نخواهد کرد. همچنین در این روش به دلیل استخراج ویژگی مناسب، از تغییرات ظاهری که متناظر با تغییرات ژست نیستند چشم‌پوشی شده است. معمولاً در تغییر حالات چهره مانند خنده، اخم و بستن یک چشم، در قالب کلی و سه‌بعدی چهره تغییری به وجود نمی‌آید و بافت موضعی چهره تحت تاثیر قرار می‌گیرد که استفاده از روش هموارسازی مرجع [۲۲] این بافت را تا حدی هموار می‌کند.

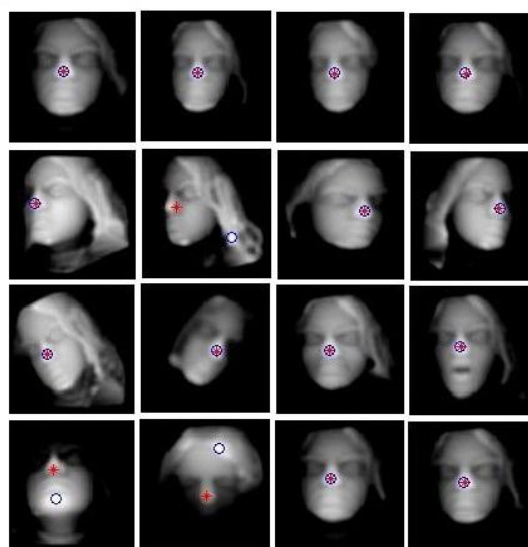
استخراج ویژگی در روش پیشنهادی با طبقه‌بند KNN به دلیل داشتن زمان اجرای بسیار کم، قابلیت پیاده‌سازی بر روی سیستم‌هایی را که به صورت برخط کار می‌کنند را دارا می‌باشد.

اگرچه تا رسیدن به یک سیستم مستقل از زاویه تصویر ورودی فاصله زیادی وجود دارد ولی استفاده از روش‌های فویتر می‌تواند به بهبود نرخ تخمین ژست در جهت بازشناسی چهره کمک کند که در راستای این هدف، برای داشتن یک تصویر کامل‌تر از چهره‌ی افراد می‌توان ایده استفاده از فرض تقارن در چهره را برای حالاتی که به دلیل پنهان‌سازی، بخشی از اطلاعات چهره برای بازسازی موجود نباشد را در تحقیقات آتی پیشنهاد کرد.

مرجع‌ها

- [1] Paysan, P., Knothe, R., Amberg, B., Romdhani, S., "A 3D face model for pose and illumination invariant face recognition," Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 296-301, 2009.
- [2] Xiaoguang Lu, Member, IEEE, and Anil K. Jain, "Deformation Modeling for Robust 3D Face

در نهایت نمونه‌ای از نتایج تشخیص صحیح نوک بینی در الگوریتم پیشنهادی و الگوریتم ارائه شده در مرجع [۱۵] در شکل ۹ نمایش داده شده است. مکان تقریبی نوک بینی در روش پیشنهادی با ستاره قرمز رنگ و در روش مرجع [۱۵] با دایره آبی مشخص شده است. همانطور که در شکل دیده می‌شود در مرجع [۱۵] به دلیل تعیین بیشترین ارتفاع به عنوان نوک بینی، با گردش بیش از حد چهره به سمت بالا و پایین (چرخش حول محور x) بیشترین ارتفاع صورت در ناحیه‌ی چانه و پیشانی قرار می‌گیرد و همچنین با گردش بیش از حد چهره به سمت راست و چپ (چرخش حول محور y) بیشترین ارتفاع در طرفین صورت قرار می‌گیرد و سبب ایجاد خطا مکانیابی نوک بینی شده است.



شکل ۹. نتایج تشخیص صحیح نوک بینی در روش پیشنهادی با ستاره قرمز رنگ و روش مرجع [۱۵] با دایره آبی رنگ.

نتیجه‌گیری

در این مقاله با استفاده از اطلاعات سه‌بعدی رویه چهره، ژست تصاویر غیر روبرو تخمین زده شد. در روش پیشنهادی شده ناحیه‌ی صورت ابتدا از داده سه‌بعدی سر استخراج شده و سپس مختصات نوک بینی در چهره مکانیابی شده و ویژگی‌های مبتنی بر موقعیت نوک بینی از ناحیه‌ی صورت استخراج شده و در روش‌های طبقه‌بندی SVM، شبکه‌های عصبی RBF و روش K نزدیکترین همسایگی KNN آموزش داده شده است. با مقایسه‌ی نتایج روش پیشنهادی با روش‌های جدید [۱۶] و [۵] بر روی پایگاه داده‌ی Frav3D و کسب نتیجه‌ی بیشینه $99/18$ درصد برای ژست روبرو و همچنین حصول میانگین $98/48$ درصد برای کل ژست‌ها در این پایگاه داده و به همین ترتیب در پایگاه داده‌ی Gavab کسب نتیجه‌ی بیشینه $99/04$ درصد برای ژست روبرو همچنین حصول میانگین $80/4$ درصد برای کل ژست‌ها در روش KNN می‌توان مدعی دقت و توانمندی الگوریتم پیشنهادی شد و همچنین از لحاظ سرعت بار محاسباتی روش KNN به میزان 0.2615 ثانیه

9 Occlusion
10 Expression

- [17] G. Zhao, L. Chen, J. Song, and G. Chen, "Large Head Movement Tracking Using SIFT-Based Registration," Proc. ACM Int'l Conf. Multimedia, pp. 807-810, 2007.
- [18] J. Wu and M. Trivedi, "A Two-Stage Head Pose Estimation Framework and Evaluation," Pattern Recognition, vol. 41, no. 3, pp. 1138-1158, 2008.
- [19] <http://archive.today/B1WeX>
- [20] <http://www.gavab.etsii.urjc.es>
- [21] N. Otsu, "A threshold selection method from gray-level histogram," IEEE Transactions on systems Man and Cybernetics, Vol.8, No.1, pp.62-67, 1979.
- [22] Garcia D, "Robust smoothing of gridded data in one and higher dimensions with missing values," Computational Statistics & Data Analysis, pp. 1167-1178, 2010.
- Matching," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol.2, pp.1377-1383, 2006.
- [3] J. Xiao, S. Baker, I. Matthews, and T. Kanade, "Real-Time Combined 2D+ 3D Active Appearance Models," Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol.2, pp. 535-542, 2004.
- [4] J.-G. Wang and E. Sung, "EM Enhancement of 3D Head Pose Estimated by Point at Infinity," Image and Vision Computing, vol. 25, no. 12, pp. 1864-1874, 2007.
- [5] Tobias Bar, Jan Felix Reuter, Manuel Yguel, J.Marius Zollner, "Driver Head Pose and Gaze Estimation based on Multi-Template ICP 3D Point Cloud Alignment." IEEE Conference on Intelligent Transportation Systems, pp.1797-1802, 2012.
- [6] Corinna cortes, Vladimir Vapnik "Support vector networks," Machine Learning, vol. 20, No.3, pp. 273-297, 1995.
- [7] Martin D. Buhmann, "Radial Basis Functions: Theory and Implementations," Cambridge University Press, 2003.
- [8] Alpaydin E. "Voting over Multiple Condensed Nearest Neighbors," Artificial Intelligence Ressearch 11, pp. 115-132, 1997.
- [9] V.F. Ferrario, C. Sforza, G. Serrao, G. Grassi, and E. Mossi, "Active Range of Motion of the Head and Cervical Spine: A Three-Dimensional Investigation in Healthy Young Adults," J. Orthopaedic Research, vol. 20, No.1, pp. 122-129, 2002.
- [10] Woo Won Kim, Biometrics Dept., Yonsei Univ, "Automatic head pose estimation from a single camera using projective geometr," Information, Communications and Signal Processing (ICICS) 8th International Conference on, pp. 1-5, Dec. 2011.
- [11] J. Ng and S. Gong, "Composite Support Vector Machines for Detection of Faces Across Views and Pose Estimation," Image and Vision Computing, vol. 20, No. 5, pp. 359-368, 2002.
- [12] M. Jones and P. Viola, "Fast Multi-View Face Detection," Technical Report 096, Mitsubishi Electric Research Laboratories, 2003.
- [13] R. Osadchy, M. Miller, and Y. LeCun, "Synergistic Face Detection and Pose Estimation with Energy-Based Models," J. Machine Learning Research, vol. 8, pp.1197-1215, 2007.
- [14] V. Balasubramanian, J. Ye, and S. Panchanathan, "Biased Manifold Embedding: A Framework for Person-Independent Head Pose Estimation," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1-7, 2007.
- [15] Amr Rashad and Alaa Hamdy, "3D face recognition using 2DPCA," IJCSNS International Journal of Computer Science and Network Security, vol. 9, No.12, pp. 149-155 Dec. 2009.
- [16] M. Shafi, P.W.H. Chung, "Face pose estimation from eyes and mouth," International Journal of Advanced Mechatronic Systems (IJAMECHS), Vol. 2, No. 1/2, 2010.

