تشخيص حالتهاي چهره از روي تصاوير متحرك به‌روش مبتنی بر ویژگی با استفاده از شبکه عصبي RBF و منطق فازی

ميرهادي سيدعربي دانشجوی دوره دکترای مخابرات، دانشکده مهندسی برق و کامپیوتر، دانشگاه تبریز
علي آقاگل‌زاده دانشیار دانشکده مهندسی برق و کامپیوتر، دانشگاه تبریز
سهراب خان‌محمدي استاد دانشکده مهندسی برق و کامپیوتر، دانشگاه تبریز

چکیده

در این مقاله روشی برای تشخیص حالتهای اصلی چهره براساس نقاط ویژگی صورت (FFP[1]) و ویژگیهای استخراج شده از تصاویر صورت از روبه‌رو ارایه شده‌است. تعداد ۲۱ نقاط ویژگی صورت در نظر گرفته شده و با استفاده از آنها ابتدا براي حذف اثر فاصلة دوربین نسبت به شخص، تمام اندازه‌ها نرمالیزه شده و اثر دوران سر نیز حذف گردیده است. این نقاط در اولین فریم مشخص شده و در فریمهاي بعدي با استفاده از روش بهبود یافته محاسبه شار نوری[2] براساس هم‌بستگی متقابل به‌طور اتوماتیک تعقیب می‌شوند. در روش اول براساس موقعیت این نقاط در فریمهای ابتدایي و انتهایي، یك بردار ویژگي براي هر یك از حالتهاي چهره تشکیل می‌گردد. با استفاده از بردار ویژگیها یك شبکة عصبي [3] RBF آموزش داده شده‌است. در روش دوم براساس حرکت نقاط ویژگی، هفت ویژگی استخراج شده و توسط شبکه عصبی RBF برای تشخیص شش حالت اصلی چهره مورد استفاده قرار می‌گیرد. در روش سوم با استفاده از یک پایگاه قواعد و براساس ویژگیهای استخراج شده، یک سیستم طبقه‌بندی براساس سیستم استنتاج فازی برای طبقه‌بندی شش حالت اصلی چهره مورد استفاده قرار می‌گیرد. نتایج به‌دست آمده نشان می‌دهد استفاده از تصاویر متحرك که از حالت نرمال شروع شده و به حالت احساسي مورد نظر ختم می‌گردند، قدرت تشخیص بهتري را ارائه می‌دهد. در مقایسه طبقه‌بندی کننده فازی با شبکه عصبی RBF، طبقه‌بندی کننده RBF دقت بهتری ارائه می‌دهد ولی به زمان پردازش بیشتری نیاز دارد و برای پیاده‌سازی زمان واقعی، طبقه‌بندی کننده فازی مناسب‌تر است.

کلمات کلیدی: تشخيص حالتهاي چهره، نقاط ویژگی صورت، شبکه عصبي RBF، سیستم استنتاج فازی.

Feature-Based Facial Expression Recognition from Image Sequences Using RBF Neural Networks and Fuzzy Inference System

M. H. Seyedarabi, A. Aghagolzadeh and S. Khanmohammadi
Faculty of Electrical and Computer Engineering, University of Tabriz

Abstract

In this Paper a facial expression recognition system based on facial characteristic points and facial features extracted from them in frontal image sequences is presented. In the first digitized frame, 21 key feature points were manually marked with a computer-mouse around facial landmarks. Selected facial feature points were automatically tracked using a cross-correlation based optical flow. At first a Radial Basis Function (RBF) neural network were used to classify four facial expressions using feature vector extracted from position of Facial Feature Points (FFP) in the first and the last frame. Then 7 features were extracted from FFP's movement and were used to classify six basic facial emotions using RBF neural networks and Fuzzy Inference System (FIS). The recognition results from proposed classifiers showed that comparing with static image classifier, using image sequences gives higher recognition rate. Comparing RBF neural networks and FIS classifier systems shows that RBF classifier acts better than FIS classifier and gives higher recognition rate but it takes more processing time and for the real time implementation FIS classifier is preferred .

Key words: Facial expressions recognition, Facial feature points, RBF neural networks, Fuzzy inference system.

1-Facial Feature Points
2- Optical Flow
3-Radial Basis Function

۱- مقدمه

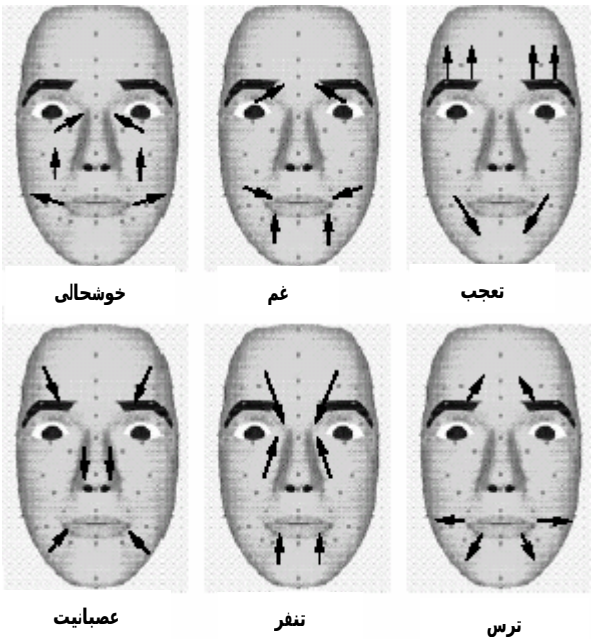
شش حالت اصلي چهره كه اغلب براي تشخيص در نظر گرفته مي‌شوند، عبارتند از: خوشحالي، ناراحتي، خشم، تعجب، ترس و تنفر. حالتهاي چهره با تغييراتي در نقاط مختلف صورت ايجاد مي‌گردد. هر كدام از اين قسمتها را واحد فعال (AU[1])مي‌نامند. نمي‌توان فقط با يك AU يك حالت چهره را توصيف نمود.

در حال حاضر تشخيص حالتهاي چهره كاربردهاي فراواني پيدا كرده است كه از جمله آنها مي‌توان ارتباط انسان با كامپيوتر را مثال زد. با استفاده از اين روش، يك ارتباط به اصطلاح قلبي ميان انسان و كامپيوتر قابل تصور است. روباتهايي كه داراي احساس هستند و با خوشحالي شما خوشحال و با ناراحتي شما ناراحت مي‌گردند، مي‌توانند مثالي از اين نمونه باشد [۴].

در طراحي انيميشن و فيلمهاي كارتوني، طراحي شخصيتهاي مختلف در بازيها و نيز در گرافيك كامپيوتري، تشخيص حالتهاي چهره مي‌تواند كاربرد مناسبي داشته باشد. اگر يك شخصيت كارتوني، حركات صورت خود را براي نمايش يك حالت، از چهرة يك انسان الهام بگيرد، نمايش حالت موردنظر بسيار واقعي‌تر بوده و طراحي نيز ساده‌تر مي‌گردد [۵]. تشخيص حالتهاي صورت از روي ويژگيهاي صورت كه مهم‌ترين آنها مربوط به چشم، دهان و ابروها هستند، انجام مي‌گيرد. مي‌توان با استفاده از قوانيني، حركات اين اعضا را كه توسط AUهاي مختلفي معين مي‌گردند، تقسيم‌بندي كرده و هر حالت مشخص چهره را با تعدادي AU توصيف‌نمود. ولي نحوة تغييرات در AUها در افراد مختلف، حتي براي نمايش يك حالت مشخص نيز متفاوت مي‌باشد.

شكل (۱) نحوه تغييرات AUهای نواحی مختلف صورت را برای نمايش شش حالت اصلی چهره نمايش می‌دهد.

در زمينة آناليز و سنتز حالتهاي چهره كارهاي زيادي انجام گرفته‌است. نظر به اهميت مرحلة آناليز، اغلب كارها و تحقيقات انجام گرفته روي مرحله آناليز و تشخيص حالتها متمركز است و بيشتر كارهاي انجام‌گرفته در مرحلة سنتز روي گرافيك كامپيوتري و انيميشن متمركز شده است.



شكل ۱- شش حالت احساسی اصلی چهره [۱۳]

در حالت كلي سه مرحله براي يك سيستم تشخيص اتوماتيك حالتهاي چهره بايد انجام‌گيرد:

- تشخيص صورت و ويژگيهای آن
- استخراج اطلاعات مربوط به حالتهاي چهره از تصوير ثابت يا دنباله‌اي از تصاوير
- طبقه‌بندي حالتهاي چهره

به دليل دقت كم تشخيص در تصاوير ثابت، اغلب كارهای انجام شده روی تصاوير متحرک متمركز است.

Cohn و همكارانش از يك مدل نقطه‌هاي مرزي صورت كه اطراف ويژگيهاي صورت متمركز شده‌اند استفاده نمودند [۶]. اين نقاط به‌صورت دستي توسط ماوس در اولين فريم مشخص مي‌گردند. در فريمهاي بعدي، يك روش سلسله مراتبي شار نوری براي تعقيب شار نوری در پنجره‌هاي ۱۳*۱۳ اطراف نقاط مورد نظر مورد استفاده قرار مي‌گيرد. جا به جايي هر يك از نقاط مرزي با كم كردن موقعيت نرماليزه شده آن در اولين فريم از موقعيت نرماليزة فعلي آن به‌دست مي‌آيد. تمام فريمهاي دنباله تصاوير ورودي به‌صورت دستي نرماليزه مي‌گردند. بردارهاي جابه‌جايي محاسبه شده بين اولين و آخرين فريم، اطلاعات لازم براي تشخيص حالت چهره را تشكيل مي‌دهند. در اين روش

1-Action Unit

صورت باید فاقد عینك و مو باشد و حرکات صلب[1] سر (حرکات سر به اطراف) نباید اتفاق بیفتد. اولین فریم نیز باید از حالت خنثي شروع گردد.

Kobayashi و Hara با مشخص‌کردن ۳۰ نقطة FFP به‌صورت دستي و استخراج چند ویژگي از روي این نقاط، با استفاده از شبکة عصبي روشي براي تشخیص حالتهاي ترس، تعجب، تنفر، خشم، خوشحالي و ناراحتي ارائه دادند که قدرت تشخیصي حدود ۷۸% با آموزش توسط ۵ چهره، ۸۳% با آموزش ۱۰ چهره و ۸۸% با آموزش توسط ۱۵ چهره به دست آوردند [۷و ۸].

Ushida و همکارانش با استفاده از CFS[2] و در نظر گرفتن ۲۷ نقطة FFP (۱۲ نقطه براي ابروها، ۷ نقطه براي چشم راست و ۸ نقطه براي دهان) روشي را براي تشخیص حالتهاي خشم، خوشحالي و ناراحتي ارائه نمودند. با آزمایش روي ۵۶ تصویر که نمایشدهندة یکي از سه حالت اشاره شده بود، سیستم پیشنهادي ایشان قدرت تشخیصي حدود ۷۸/۷% را ارائه نمود [۹].

Ebine و Nakamura با استفاده از منطق فازي روشي براي تشخیص چهار حالت خوش‌حالي، خشم، ناراحتي و تعجب ارائه دادند که با آزمایش روي ۵۹ تصویر در ۸۶/۷% موارد تشخیص صحیح ارائه مي‌داد. روش ارائه شده توسط ایشان در چند مرحله انجام مي‌گرفت [۱۰]:

- پیش‌پردازش شامل استخراج ناحیة مورد پردازش روي صورت.
- حذف حرکات ناشي از سر و نرمالیزه کردن سطح مورد پردازش روي صورت.
- اختصاص سطوحي مربوط به ناحیه‌هاي مختلف متحرك روي صورت و استخراج حرکات مربوط به آنها.
- تعیین قوانیني براي تشخیص هر حالت چهره از روي حرکات مربوط به ناحیه‌هاي صورت.

Yacoob و همکارانش با استفاده از شار نوری، حرکت AUهاي مختلف را در فریمهاي متوالي که نمایش‌دهندة یك حالت مشخص بود، به‌دست آوردند. این فریمها از حالت نرمال شروع و به یکي از حالتهاي اصلي چهره ختم مي‌شدند. با این روش براي هریك از حالتهاي چهره یك بردار ویژگي براساس حرکت AUها به‌دست آمده و براي آموزش یك شبکه عصبي RBF مورد استفاده قرار مي‌گرفت. سیستم فوق در تقسیمبندي فریمها به یکي از ۶ حالت اصلي چهره، قدرت تشخیصي حدود ۸۰% ارائه مي‌داد [۱۱، ۱۲ و۱۳].

Piat و Tsapatsoulis با مقایسة نسبي فواصل تعریف شده توسط FFPها در اولین فریم، که از حالت نرمال شروع مي‌شد و تعقیب آنها از یك فریم به فریم دیگر، یك روش تعقیب اتوماتیك براساس استنتاج فازي براي این ویژگیها ارائه نمودند. توابع عضویت استفاده شده براي هر یك از ویژگیها از نوع مثلثي بود که مابین ماکزیمم و مینیمم مقدار آن ویژگي محدود بوده و رأس آن در مقدار میانگین آن ویژگي متمرکز شده بود. ایشان با استفاده از ۸ ویژگي و توسط سیستم FIS قدرت تشخیصي بین۷۰% تا ۹۰% براي فریمهاي متوالي به‌دست آوردند [۱۴].

در این مقاله با استفاده از۲۱ نقطه مشخصة صورت مربوط به چشم، ابرو، دهان، بیني وگونه‌ها و تعقیب این نقاط در فریمهاي متوالي، ابتدا با تعیین اختلاف حرکت آنها در فریم اول و آخر، یك بردار ویژگي براي حالتهاي خوشحال، متعجب، ناراحت و عصباني محاسبه شده و با استفاده از شبکة عصبي RBF طبقه‌بندي شده است.در مرحله بعدی با استخراج چند ویژگی و طبقه‌بندی کننده RBF، سیستمی برای تشخیص شش حالت اصلی صورت ارائه شده است. در نهایت با استفاده از ویژگیهای استخراج شده و یک پایگاه قواعد از یک طبقه‌بندی کننده فازی با توابع عضویت گوسی برای طبقه‌بندی شش حالت اصلی چهره استفاده شده است. نسبت به کارهاي مشابهي که در این زمینه انجام شده است، تعداد نقاط ویژه صورت و نیز ویژگیهاي استفاده شده در این مقاله کمتر بوده و تعداد گره‌های ورودی شبکه عصبی RBF با قدرت تشخیص مناسب به حداقل رسیده است. با استفاده از چندین ویژگي جدید تعریف شده، سیستم طبقه‌بندی قدرت تمایز بالاتري ارائه مي‌دهد. در سیستم ارائه شده محدودیتهائی که در بعضی از کارهای انجام شده وجود دارد حذف شده است [۶]. از جمله این محدودیتها می‌توان به فاصله دوربین تا صورت، دوران سر به اطراف (حرکات صلب)، موی صورت و داشتن عینک اشاره نمود. در ادامه به بررسي خصوصیات این سیستم پرداخته و نتایج حاصل از آزمایش را ارائه مي‌دهیم.

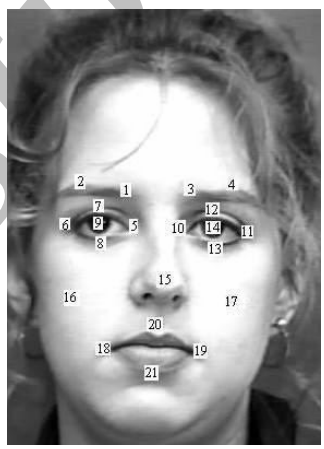
۲- نقاط ویژگی صورت (FFP)

در شکل (۲) تعداد ۲۱ نقطة FFP که در این تحقیق مورد استفاده قرار گرفته، نمایش داده شده است. ۱۰ نقطه مربوط به

1- Rigid motion
2-Conceptual Fuzzy Sets

چشم، ۴ نقطه مربوط به ابروها، يك نقطه مربوط به نوك بيني، ۴ نقطه مربوط به دهان و ۲ نقطه مربوط به گونه‌ها مي‌باشند. براي حذف اثر فاصلة دوربين نسبت به شخص، كميتي به‌نام BASE كه در واقع فاصلة بين انتهاي داخلي چشمها مي‌باشد، در نظر گرفته شده و تمام مقادير اندازهگيري شده نسبت به اين كميت نرماليزه مي‌شوند [۱و۲].

$$BASE=\sqrt{(x_5-x_{10})^2+(y_5-y_{10})^2} \qquad (۱)$$



شكل ۲- نقاط ويژگی صورت

براي حذف اثر دوران صورت، زاويه BASE نسبت به افق نيز در نظر گرفته شده‌است.

$$\theta = Arc\ \tan(\frac{y_5 - y_{10}}{x_5 - x_{10}}) \qquad (۲)$$

براي اين كه مختصات FFPها در تصاوير مختلف نسبت بـه يـك مرجع ثابتي اندازهگيري شوند، نقطه مرجـع(x_{org} , y_{org})كه تقريباً نوك بيني مي‌تواند فرض‌شود، به‌عنوان مبـدأ مقايـسة مختصات FFPها در تـصاوير مختلـف در نظـر گرفتـه مـي‌شـود. (x_0, y_0) مختـصات نقطـة وسـط BASE بـا روابـط زيـر مشخص ميگردند:

$$x_0 = \frac{x_5 + x_{10}}{2}$$
$$y_0 = \frac{y_5 + y_{10}}{2} \qquad (۳)$$

مختصات نقطة مبدأ (x_{org} , y_{org}) برابر خواهد بود با:

$$x_{org} = x_0 - BASE \times \sin\theta$$
$$y_{org} = y_0 - BASE \times \cos\theta \qquad (۴)$$

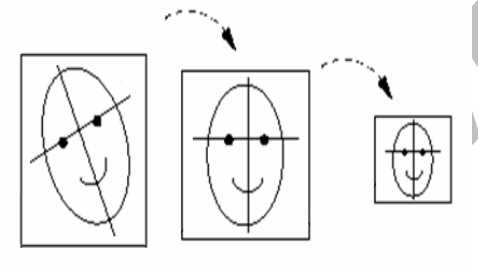
اگر مختصات يكي از FFP ها در تصوير اصلي (x_b, y_b) باشد مختصات جديد آن نسبت به مبدأ به‌دست آمده، به‌صورت زير محاسبه مي‌گردد:

$$x_a = x_b - x_{org}$$
$$y_a = y_b - y_{org} \qquad (۵)$$

اين مختصات به‌صورت زير نسبت به كميت BASE نرماليزه مي‌گردند:

$$x'_a = x_a / BASE$$
$$y'_a = y_a / BASE \qquad (۶)$$

شكل (۳) نحوه نرماليزه كردن فاصله‌ها وحذف اثر دوران سر را نشان می‌دهد:



شكل۳- حذف دوران سر و نرماليزه كردن فاصلهها

با استفاده از مختصات نرماليزه شده، اختلاف هر نقطه در اولين فريم و آخرين فريم محاسبه شده و به‌عنوان بردار ويژگي هر حالت در نظر گرفته مي‌شود.

نقاط ويژگی در اولين فريم به‌صورت دستی توسط ماوس مشخص شده و در فريمهای بعدی به‌طور اتوماتيک تعقيب می‌گردند.

۳- پايگاه داده‌های استفاده شده

در اين مقاله از پايگاه داده‌های Cohn-Kanade استفاده شده‌است [۱۵]. اين پايگاه داده‌ها شامل نمايش حالتهاي مختلف چهره توسط ۱۰۴ نفر مي‌باشد كه همگي از حالت نرمال شروع و به حالت احساسي مورد نظر ختم مي‌شوند. تصاوير داراي سرعت

۳۰ فریم در ثانیه بوده و تعداد متوسط فریمها براي نمایش یك حالت معین توسط افراد شرکت کننده در پایگاه داده‌ها ۱۰ فریم مي‌باشد.

به‌دلایلي پایگاه داده‌ها شامل نمایش تمام حالتها توسط همه افراد نمي‌باشد. به‌همین دلیل در این مقاله ۵۳ نفر که حداقل چهار حالت خوشحال، متعجب، ناراحت و عصباني را نمایش داده‌اند انتخاب شده است. مشخصات این پایگاه داده‌ها در جدول (۱) ارائه شده است.

جدول ۱- مشخصات پایگاه داده‌های Cohn-Kanade

سن	۱۸ الي ۵۰ سال
زن	%۶۹
مرد	%۳۱
اروپا-آمریکایي	%۸۱
افریقا-آمریکایي	%۱۳
سایر نژاد	%۶
وضوح (Grayscale)	۶۴۰×۴۹۰ پیکسل

نمایش حالتهاي احساسي اشاره شده توسط یکي از افراد این پایگاه داده‌ها در شکل (۴) نمایش داده شده است.

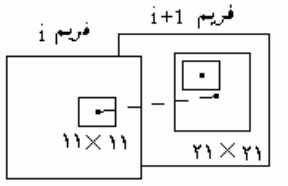
شکل ۴- حالتهاي متعجب، ناراحت، عصباني و خوشحال در پایگاه داده‌های Cohn-Kanade

۴- روش محاسبه شار نوری

شار نوری براي محاسبه حرکت قسمتهاي متحرك در تصویر استفاده مي‌گردد. روشهاي مختلفي براي محاسبه آن مي‌تواند استفاده گردد. در این مقاله براي محاسبه آن از روش همبستگی تقاطعی[1] استفاده شده‌است.

براي تخمین بهینه حرکت در روش شار نوری یك تخمین اولیه از سرعت حرکت و جهت آن مفید خواهد بود. در پایگاه داده‌های مورد استفاده، حداکثر سرعت حرکت یك پیکسل از یك فریم به فریم بعدي حدود pixel/frame ۱۰ به‌دست آمد. بر این اساس در اولین فریم اطراف هر یك از نقاط ویژگي انتخاب شده یك پنجره ۱۱×۱۱ در نظر گرفته شده و همبستگی تقاطعی آن با یك پنجره ۲۱×۲۱ در فریم بعدي و به مرکز همان نقطه محاسبه مي‌گردد. محلي که به ازاي آن بیشترین مقدار براي هم‌بستگی تقاطعی به‌دست مي‌آید به‌عنوان تخمین محل نقطه در فریم بعدي در نظر گرفته مي‌شود.

شکل (۵) نحوه انجام این روش را نشان مي‌دهد:



شکل ۵- محاسبه هم‌بستگی تقاطعی در فریمهاي متوالي

در این مقاله برای کاهش پیچیدگی و بهبود تخمین جهت حرکت، فرض شده است که نقاط ۲، ۴، ۷، ۸، ۱۲، ۱۳، ۲۰ و ۲۱ فقط دارای حرکت عمودی هستند (بعد از حذف حرکات صلب سر) و در نتیجه پنجره جستجو در فریم بعدی از ۲۱×۲۱ به ۵×۲۱ کاهش یافته است. همچنین برای نقطه ۱۵ به دلیل حساسیت آن (مبدا محور مختصات فرض شده است) و نقاط ۱۶ و ۱۷ به دلیل ابهامی که در شدت روشنائی اطراف آن وجود دارد پنجره اطرافشان از ۱۱×۱۱ به ۱۵×۱۵ افزایش می‌یابد. به غیر از این سه نقطه بقیه نقاط به نحوی انتخاب شده‌اند که با نقاط اطراف کنتراست بالائی داشته باشند و تخمین محل آنها در فریم بعدی سادهتر باشد.

استفاده از پنجره متغیر در مقایسه با روش‌هائی که از پنجره ثابت استفاده کرده‌اند [۶]، دقت و تخمین مناسبتری نشان داد و به غیر از مواردی که پلک زدن چشم اتفاق می‌افتاد، در بقیه موارد تخمین درستی از محل نقاط ویژگی در فریمهای متوالی ارائه کرد.

روش ارائه شده در عمل حساسیت کمی نسبت به تغییرات شدت روشنائی از خود نشان داد ولی در تعقیب نقاطی که در اطراف چشم هستند، پلک زدن تاثیر نامطلوبی داشت.

1- Cross-Correlation

۵- سیستم تشخیص چهار حالت اصلی چهره توسط شبکهٔ عصبي RBF

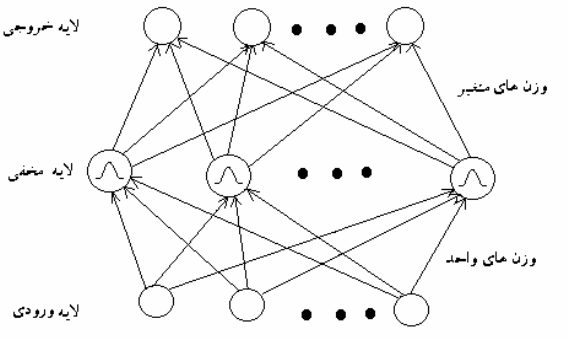
در این روش با استفاده از مختصات نرمالیزه شده، اختلاف هر نقطه در اولین فریم و آخرین فریم محاسبه شده و به‌عنوان بردار ویژگي هر حالت چهره در نظر گرفته مي‌شود. بردار ویژگیهاي محاسبه‌شده، ورودي شبکهٔ عصبي RBF را تشکیل می‌دهند. شکل (۶) ساختار روش استفاده شده را نمایش می‌دهد.

شکل۶- ساختار سیستم طبقهبندی با شبکه عصبی

شبکه عصبی شامل ۲۱ گره ورودي (هر ورودي براي يكي ازمؤلفه‌هاي بردار ویژگي)، يك لاية مخفي و چهار گره خروجي متناظر با يكي از حالتهاي خوشحال، متعجب، ناراحت و عصباني مي‌باشد.

در شکل (۷) شبکهٔ RBF سادهٔ Darken و Moody که شامل يك لاية ورودي، يك لاية مخفي و يك لاية خروجي است نشان داده شده است [۱۶].

لایه ورودي شامل واحدهایي است که به بردار ورودي متصل هستند. لایهٔ مخفي از واحدهایي تركب یافته که براساس تحریکات توابع Radial Basis عمل مي‌کنند. این واحدها میدان نامیده مي‌شوند. لایه خروجي شامل گره‌هایي است که به بردار خروجي متصل هستند.گره‌هاي ورودي به‌طور کامل با اتصالات با وزن واحد به فیلدهاي پذیرنده[1] در لایهٔ مخفي متصل هستند و فیلدهاي پذیرنده به‌طور کامل با اتصالات وزن‌دار به واحدهاي خروجي متصل هستند.



شکل۷- ساختار شبکه عصبي RBF

تابع Radial Basis استفاده شده براي تابع تحریك فیلدهاي پذیرنده مي‌تواند فرمهاي مختلفي داشته باشد. رایج‌ترین فرم براي تابع تحریك Radial Basis تابع گوسي N بعدي مي‌باشد. آموزش این شبکهٔ عصبي در واقع روي تغییر وزن اتصالات خروجي و تغییر مرکز وعرض توابع گوسي متمرکز است. تعداد گرههای ورودی برابر تعداد مولفههای بردار ویژگی ورودی و تعداد گرههای خروجی برابر تعداد کلاسهای خروجی است.

در این مقاله تعداد واحدهای لایه مخفی برابر تعداد بردارهای آموزش در نظر گرفته شده است. این کار سرعت مرحله آموزش شبکه را بسیار افزایش می‌دهد.

توابع تحریک استفاده شده براساس رابطه (۷) تعریف می‌شوند.

$$RBF(c_i) = \exp(-(||c_i - p|| b)^2) \qquad (۷)$$

c_i مرکز واحد RBF، p بردار ویژگی ورودی، b بایاس و $||\ .\ ||$ نشان دهنده نرم اقلیدسی است.

واحد RBF با مرکزی که فاصله زیادی با p داشته باشد خروجی نزدیک به صفر خواهد داشت و تاثیر ناچیزی روی نرونهای لایه خروجی خواهد گذاشت. از طرف دیگر، واحد RBF با مرکزی که فاصله کمی با p داشته باشد خروجی نزدیک به یک خواهد داشت. اگر خروجی واحدی یک باشد وزن خروجی آن این مقدار

1- Receptive fields

را به نرونهای خروجی انتقال خواهد داد. بایاس با رابطه زیر تعریف می‌گردد:

$$b = \frac{0.8326}{\sigma} \qquad (۸)$$

که σ عرض توابع گوسی است. فرض فوق باعث می‌شود تا اگر فاصله بردار ورودی و مرکز واحد RBF برابر $\pm\sigma$ باشد، خروجی تابع RBF برابر $\frac{1}{2}$ گردد. مراکز واحدهای RBF برابر بردارهای آموزش در نظر گرفته می‌شود و $\sigma = ۰/۹$ فرض می‌گردد. σ بایستی طوری انتخاب گردد که توابع مربوط به بردارهای نزدیک به هم هم پوشانی داشته باشند. از طرف دیگر نباید خیلی بزرگتر از محدوده بردارهای ورودی انتخاب گردد، چون در این صورت توابع گوسی بسیار پهن بوده و برای تمام ورودیها، خروجی خواهند داشت. انتخاب مناسب σ به شرایط مسئله بستگی دارد.

براي آموزش و آزمایش از ۵۳ چهرهٔ مختلف مربوط به پایگاه داده‌های Cohn-Kanade استفاده شده است. البته یک پایگاه داده‌ها نیز تهیه گردیده بود ولی به‌منظور مقایسه با کارهای انجام‌گرفته، بیشتر از این پایگاه داده‌ها که خصوصیات استاندارد دارد استفاده شده است.

در آزمایش هر یك از حالتها، شبکهٔ عصبي با تصاویر مربوط به ۴۰ نفر آموزش دیده و توسط ۱۳ نفر باقي مانده مورد آزمایش قرار گرفته و این کار پنج بار به‌صورت دوره‌اي براي آزمایش تمام افراد انجام مي‌گیرد و چهره‌هاي مورد آزمایش در هر مرحله در آموزش شبکه دخالت داده نمي‌شوند. اجرای برنامه با نرمافزار Matlab 6.5 و کامپیوتر Pentium IV 2.4 و دارای 512 MB RAM به‌طور متوسط برای ۱۰ فریم به ازای هر حالت نمایش داده شده، حدود ۲۰ ثانیه طول کشید. نتایج حاصل از آزمایش در جدول (۲) نمایش داده شده است.

جدول ۲- نتایج تشخیص حاصل از شبکه عصبي RBF

تشخیص(%)	خوش‌حال	متعجب	ناراحت	عصباني
خوشحال	۹۱	۶	۳	۰
متعجب	۳	۹۳	۰	۴
ناراحت	۴	۶	۸۶	۴
عصباني	۱۰	۶	۳	۸۱

ایـن روش همـان‌طـور کـه از جـدول مـشخص اسـت، قـدرت تشخیص متوسطي برابر با ۸۷/۷۵% ارائه مي‌دهد کـه در مقایـسه بـا کارهـاي انجـام گرفتـه [۱۱، ۱۲و۱۳] نتیجـهٔ رضـایتبخـشي مي‌باشد.

۶- طبقه‌بندی شش حالت اصلی چهره با شبکه عصبی RBF

در روش قبل با استفاده از بردار ویژگی به‌دست آمـده از روی حرکت نقاط ویژگی صورت، سیستمی برای تشخص چهار حالـت اصلی چهره ارائه شد. این روش به‌دلیل عدم استفاده از ویژگیهای استخراج شده از روی نقاط ویژگی، در تشخیص حالتهای ترس و تنفر با مشکل مواجه بود و با افزودن ویژگیهای این دو حالـت در مرحله آموزش، سیستم طبقه‌بندی دقت مناسبی ارائـه نمـی‌داد. در مرحله دوم با استفاده از هفت ویژگی اسـتخراج شـده از روی حرکت نقاط ویژگی و با طبقه‌بنـدی آنهـا توسـط شـبکه عـصبی RBF دقت مناسبی برای تشخیص شش حالت اصلی چهره ارائـه گردید [۳].

۶-۱- ویژگیهای استخراج شده

براساس تغییرات صورت که برای هر یک از حالتهای احساسی ششگانه در شکل (۱) نمایش داده شده است، با استفاده از حرکت ۲۱ نقطه مشخصه صورت هفت ویژگی استخرج گردید. این ویژگیها نشانگر تغییرات ویژگیهای اصلی چهره، یعنی چشم‌ها، ابروها و دهان هستند.

عرض چشم‌ها:

$$we = \frac{(x_{11} - x_{10}) + (x_5 - x_6)}{2} \qquad (۹)$$

ارتفاع ابروها (۱):

$$he1 = \frac{(y_{15} - y_4) + (y_{15} - y_2)}{2} \qquad (۱۰)$$

ارتفاع ابروها (۲):

$$he2 = \frac{(y_{15} - y_3) + (y_{15} - y_1)}{2} \qquad (۱۱)$$

عرض دهان:

$$wm = x_{19} - x_{18} \qquad (۱۲)$$

میزان باز بودن دهان:

$$om = y_{21} - y_{20} \quad (۱۳)$$

فاصله نوک بینی تا گوشه‌های لب:

$$nl = \frac{(y_{18} - y_{15}) + (y_{19} - y_{15})}{2} \quad (۱۴)$$

فاصله مردمک تا گونه:

$$ec = \frac{(y_{16} - y_9) + (y_{17} - y_{14})}{2} \quad (۱۵)$$

(x_i, y_i) مختصات نقاط ویژگی صورت است که مبدا مختصات نوک بینی فرض شده است. این ویژگیها با استفاده از موقعیت نرمالیزه شده نقاط ویژگی در اولین و آخرین فریم محاسبه شده و نسبت مقادیر در فریم آخر به فریم اول به عنوان مولفههای بردار ویژگی در نظر گرفته می‌شوند.

۶-۲- سیستم طبقه‌بندی با شبکه عصبی RBF

در این مرحله از سیستم طبقه‌بندی شبکه عصبی RBF که دارای هفت گره ورودی متناظر با هفت ویژگی استخراج شده و شش گره خروجی متناظر با شش حالت اصلی چهره بود استفاده شد. در این روش هم تعداد ورودیهای شبکه نسبت به کار قبلی کاهش یافته و هم تعداد حالتهای طبقه‌بندی شده افزایش یافته‌است. نتایج حاصل از تشخیص توسط سیستم فوق در جدول (۳) آورده شده است:

در مقایسه با کارهائی‌که از ویژگیهای مشابه و طبقهبندی با شبکههای عصبی استفاده نموده‌اند [۷و۸]، نتایج به دست آمده رضایت بخش می‌باشند. اجرای برنامه با همان شرایطی که در بخش ۵ ذکر شد، حدود ۳۰ ثانیه طول کشید.

۷- سیستم طبقه‌بندی با منطق فازی

سیستم استنتاج فازی (FIS) فرایند فرموله کردن نگاشت ورودی معین به خروجی با استفاده از منطق فازی است. شکل (۸) سیستم طبقه‌بندی با استفاده از FIS را نمایش می‌دهد. مهمترین قسمت تعیین پایگاه قواعد مناسب می‌باشد.

در این مرحله با استفاده از پایگاه قواعد تعریف شده در جدول (۴) یک سیستم استنتاج فازی برای طبقه‌بندی شش حالت اصلی چهره استفاده شد. ویژگیهای استخراج شده همان ویژگیهای استفاده شده در مرحله قبل می‌باشند [۳].

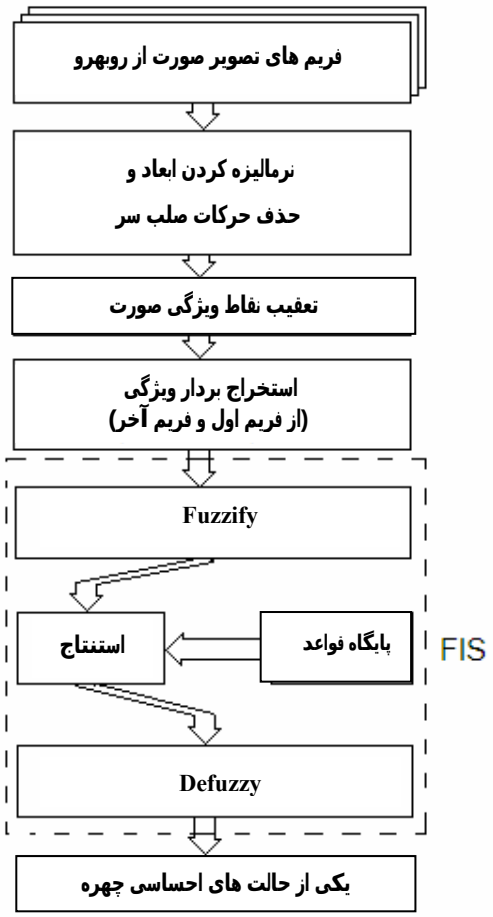
جدول۳- نتایج طبقهبندی شش حالت اصلی چهره با شبکه RBF

تشخیص (%)	خوشحال	متعجب	ناراحت	عصباني	تنفر	ترس
خوشحال	۱۰۰	۰	۰	۰	۰	۰
متعجب	۱/۴	۹۶/۲	۰	۰	۰	۲/۴
ناراحت	۰/۹	۰	۸۷/۱	۶	۶	۰
عصباني	۰	۰	۹	۸۹/۶	۱/۴	۰
تنفر	۰/۹	۰	۳/۴	۰	۹۵/۷	۰
ترس	۰	۱۲	۶/۸	۰	۰	۸۱/۲

جدول۴- پایگاه قواعد برای سیستم استنتاج فازی

VS: very small, **S**: small, **M**: medium
L: large, **VL**: very large

WE	HE1	HE2	WM	OM	NL	EC	
M	M	M	VL	S	S	S	خوشحال
VL	VL	L	M	VL	VL	L	متعجب
S	M	L	M	VS	M	M	ناراحت
M	M	S	S	S	M	S	عصبانی
M	VS	S	M	S	S	S	تنفر
L	L	L	S	M	M	M	ترس

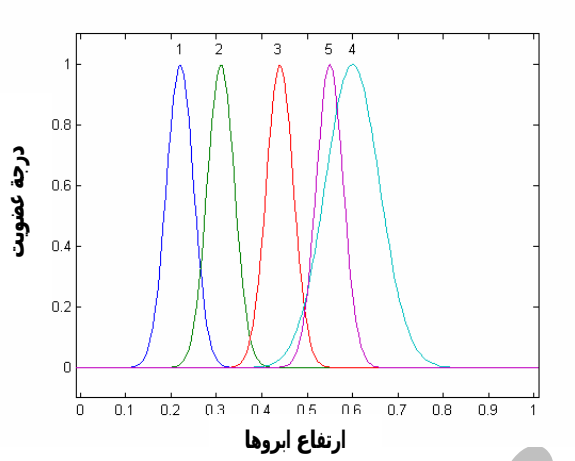


شکل ۸- سیستم طبقه‌بندی فازی

با توجه به شکل (۱) و تغییراتی که ویژگیهای چهره برای نمایش یک حالت احساسی پیدا می‌کنند، پایگاه قواعد به صورت تجربی و براساس ویژگیهای استخراج شده در بخش **۶-۱** تشکیل می‌گردد.

سیستم استنتاج براساس AND قواعد برای تشخیص هر حالت عمل می‌کند. هر یک از ویژگیهای استخراج شده، به پنج قسمت تقسیم می‌شوند: خیلی کوچک، کوچک، متوسط، بزرگ و خیلی بزرگ. این مقادیر در مرحله آموزش براساس میانگین دادههای آموزش برای هر ویژگی تعیین می‌گردند.

توابع عضویت انتخاب شده برا ی ویژگیها، از نوع گوسی هستند. مرکز گوسی روی هر یک از این پنج گروه متمرکز شده و عرض توابع گوسی براساس واریانس مقادیر در دادههای مرحله آموزش تعیین می‌شود. شکل (۹) توابع استفاده شده برای یکی از ویژگیها ارتفاع ابروها را نمایش می‌دهد.



شکل ۹- توابع عضویت گوسی برای یکی از ویژگیها

برای خروجی نیز از شش تابع عضویت گوسی برای هر یک از شش حالت احساسی استفاده شده است. مرکز این توابع روی اعداد ۱ تا ۶ که هر عدد متناظر با یکی از حالتهای احساسی شش گانه است، متمرکز شده است.

نتایج حاصل از تشخیص توسط این سیستم در جدول (۴) آورده شده است:

جدول۴- نتایج طبقه‌بندی شش حالت اصلی چهره با سیستم استنتاج فازی

تشخیص(%)	خوشحال	متعجب	ناراحت	عصباني	تنفر	ترس
خوشحال	۹۳	۷	۰	۰	۰	۰
متعجب	۲/۸	۹۶/۲	۰	۰	۰	۱
ناراحت	۰	۰	۸۶/۳	۷/۵	۳/۴	۲/۸
عصباني	۰	۰	۱۱	۸۶/۲	۲/۸	۰
تنفر	۲/۸	۰	۲/۸	۳/۴	۹۱	۰
ترس	۰	۱۱/۷	۶/۸	۰	۰	۸۱/۵

در مقایسه با روش‌هائی که از طبقه‌بندی فازی برای حالتهای چهره استفاده کرده‌اند [۹، ۱۰ و۱۴]، نتایج به دست آمده رضایتبخش به نظر می‌رسد. اجرای برنامه با همان شرایطی که در بخش ۵ ذکر شد، حدود ۲۴ ثانیه طول کشید.

۸- نتیجه‌گیري

در مرحله اول روشي براي تشخيص حالتهاي چهره شامل حالتهاي خوشحال، ناراحت، متعجب و عصباني ارائه گرديد. با استفاده از شبكة عصبي RBF، سيستم با بردار ويژگي استخراج شده براي هر حالت آموزش ديده و قدرت تشخيص متوسطي حدود ۸۷/۷۵%، براي چهره‌هايي كه مورد آموزش قرار نگرفته‌اند ارائه‌داد.

استفاده از پنجره متغییر در تعقیب حرکت نقاط ویژگی، علاوه بر کاهش پیچیدگی باعث افزایش دقت تخمین محل نقاط در فریمهای بعدی گردید. تغییرات شدت روشنائی تأثیر کمی روی دقت تخمین داشت ولی پلک زدن چشم تأثیر نامطلوبی روی آن داشت.

در مرحله دوم روشي براي تشخيص شش حالت اصلی چهره با استفاده از شبكة عصبي RBF ارائه گرديد. سيستم به‌جای استفاده مستقیم از بردار حرکت مربوط به نقاط ویژگی صورت، با هفت ويژگي استخراج شده از روی این نقاط آموزش دیده و قدرت تشخيصي حدود ۹۱/۶%، براي چهره‌هايي كه مورد آموزش قرار نگرفته‌اند ارائه داد. در مقایسه با مرحله قبلی هم تعداد گره‌های ورودی کمتر شده است و هم تعداد حالتهای تشخیص داده شده (گره‌های خروجی) افزایش داده شده است.

در مرحله سوم سيستمي براساس FIS پيادهسازي شده و براساس توابع عضويت گوسی و استفاده از قوانيني براي هر يك از حالتهاي چهره، روشي ارائه شد كه در ۸۹/۱% موارد تشخيص صحيح ارائه مي‌داد.

سپاس گزاریها

در اینجا لازم است از مرکز تحقیقات مخابرات ایران به دلیل حمایت مالی از این تحقیق و نیز آقایان Jeffrey F. Cohn و Takeo Kanade از دانشگاه Pittsburgh که مجوز استفاده از پایگاه داده‌های مورد استفاده در این مقاله را داده‌اند، تشکر و قدرداني گردد.

مراجع

[۱] ميرهادي سيدعربي، علي آقاگل‌زاده و سهراب خان‌محمدي، "تشخيص حالتهاي چهره براساس نقاط مشخصه صورت در تصاوير ثابت با استفاده از شبكه عصبي RBF و منطق فازي،" دومين كنفرانس بينايي ماشين و پردازش تصوير ايران(MVIP 2003)، بهمن ۱۳۸۱.

[۲] ميرهادي سيدعربي، علي آقاگل‌زاده و سهراب خان‌محمدي "تشخيص حالتهاي چهره از روي تصاوير متحرك با استفاده از شار نوری و شبكه عصبي RBF،" دوازدهمين كنفرانس مهندسی برق ايران (ICEE2004)، دانشگاه فردوسی مشهد، ارديبهشت ۱۳۸۳.

[3] H. Seyedarabi, A. Aghagolzadeh and S .Khanmohammadi ," Recognition of Six Basic Facial Expressions by Feature-Points Tracking using RBF Neural Network and Fuzzy Inference System", IEEE ICME 2004, Jun. 2004,Taipei, Taiwan.

[4] H. Kobayashi and F. Hara," Recognition of Six Basic Facial Expressions and Their Strength by Neural Network ," IEEE International Workshop on Robot and Human Communication, ROMAN' 92, pp. 381-386 , 1992.

[5] T. Goto, S. Kshirsagar and N. Magnenat, "Automatic Face Cloning and Animation," IEEE signal processing Magazine, pp. 17-25, May 2001

[6] T. Kanade, J. Cohn, and Y. Tian "Recognizing Action Units for Facial Expression Analysis," IEEE Transactions on Pattern Analysis and. Machine Intelligence, vol. 23, no.1, pp. 97-115, 2001.

[7] H. Kobayashi and F. Hara, "Real - Time Recognition of Six Basic Facial Expression," IEEE Int. Workshop on Robot and Human Communication, pp. 179- 185, 1995

[8] H. Kobayashi," Study on Face Robot Platform as a KANSEI Medium," Journal of Robotics and Mechatronics, vol. 13, no. 5, pp. 497-504, 2001.

[9] H. Ushida, T. Yamaguchi and T. Takagi, "Fuzzy-Associative-Memory-Based Knowledge Construction with an Application to a Human-Machine Interface", IEEE

Transaction on Industrial Electronics, vol. 46, no. 4, pp. 851-960, 1999.

[10] H. Ebine, Y. Shiga, M. Ikeda and O. Nakamura," Automatic Detection of Reference Face and The Recognition of Transition of Facial Expressions," Transactions of The Institute of Electrical Engineers of Japan vol. 121, no. 10, 2001.

[11] M. Rosenblum , Y. Yacoob and L. Davis ,"Human Expression Recognition from Motion using a Radial Basis Function Network Architecture," IEEE Trans. on Neural Networks, vol. 7, No. 5, Sep. 1996.

[12] Y. Yacoob and L. Davis," Computing Spatio-Temporal Representation of Human Faces," Proc. of Computer Vision and Pattern Recognition Conf., pp. 70-75, 1994.

[13] M. j. Black and Y. Yacoob, "Recognizing Facial Expressions in Image Sequences Using Local Parameterized Model of Image Motion," Int'l J. Computer Vision, vol. 25, no. 1, pp. 23-48, 1997.

[14] F. Piat and N. Tsapatsoulis, "Exploring the Time Course of Facial Expressions with a Fuzzy System," IEEE ICME 2000, PP. 615-618 , NY, U.S.A, 2000.

[15] T. Kanade, J. Cohn, and Y. Tian., Comprehensive database for facial expression analysis, 2000.

[16] J. Moody and C. Darken, "Learning with Localized receptive fields" in proc, 1988 Connectionist Models Summer School. San Matco. CA: Morgan-Kaufmann 1988.