

Person-independent facial expression recognition based on prior knowledge from the new subject

A. Mohammadian¹, H. Aghaeinia², F. Towhidkhah^{3*}

¹Ph.D. Student, Bioelectric Department, Faculty of Biomedical Engineering, Amirkabir University of Technology, Tehran, Iran, A.mohammadian@aut.ac.ir

²Associate Professor, Department of Electrical engineering, Amirkabir University of Technology, Tehran, Iran, Aghaeini@aut.ac.ir

³Associate Professor, Bioelectric Department, Faculty of Biomedical Engineering, Amirkabir University of Technology, Tehran, Iran

Abstract

In this paper, a method is proposed based on the prior knowledge from a new subject to improve the performance of person-independent facial expression recognition. First, in order to obtain a basic system, a combination of geometric features and texture descriptor is compared with global features (i.e., mapped face images using the Kernel-PCA and raw data of face images). The results of comparison under noisy conditions were investigated and evaluated by person-dependent/independent cross-validation method. The obtained basic system was evaluated by leave-one-subject-out cross-validation. Since the same subjects are not introduced in both training and test phases, the basic recognition system is person-independent and its performance is substantially lower than that of person-dependent cross-validation case. To improve the performance of the basic system, a method is proposed in which virtual samples are generated based on the prior knowledge from the new subject and are used in learning process. The results show that the recognition rate increases up to 96% for the person-dependent basic system, kernel-PCA method is more sensitive than the others to interpersonal variability, and the recognition rate is significantly ($P<0.05$) improved up to 91.39% compared to that of person-independent case.

Key words: Facial expression recognition, person-independent system, prior knowledge, virtual samples.

*Corresponding author

Address: 424 Hafez Ave, Tehran, Iran, 15875-4413

Tel: +9821-64542363

Fax: +9821-66468186

E-mail: Towhidkhah@aut.ac.ir

بازشناسی جلوه‌های هیجانی چهره مستقل از فرد مبتنی بر دانش اولیه از شخص جدید

امین محمدیان^۱، حسن آقائی‌نیا^۲، فرزاد توحیدخواه^{۳*}

^۱دانشجوی دکتری، دانشکده مهندسی پزشکی، دانشگاه صنعتی امیرکبیر (پلی تکنیک)، تهران A.mohammadian@aut.ac.ir

^۲دانشیار، دانشکده مهندسی برق، دانشگاه صنعتی امیرکبیر (پلی تکنیک)، تهران، Aghaeini@aut.ac.ir

^۳دانشیار، دانشکده مهندسی پزشکی، دانشگاه صنعتی امیرکبیر (پلی تکنیک)، تهران.

چکیده

در این مقاله، روشی مبتنی بر دانش اولیه از شخص جدید با هدف افزایش قدرت تعمیم‌دهی سیستم بازشناسی جلوه‌های هیجانی چهره پیشنهاد شده است. به منظور بازشناسی مناسب، ترکیبی از ویژگی‌های هندسی و توصیفگرهای بافت چهره استفاده شد. این ویژگی‌ها با ویژگی‌های کل‌نگر (تحلیل مؤلفه‌های مستقل هسته-محور تصویر چهره و خود تصویر چهره) مقایسه شدند. برای تحلیل ویژگی‌های پیشنهادی، حساسیت نرخ بازشناسی آنها نسبت به تغییر نویز و تغییرات بین فردی بررسی شد. نتایج نشان داد با وابسته کردن سیستم به شخص بروز دهنده می‌توان نرخ بازشناسی را تا ۹۶٪ افزایش داد که این نتیجه مربوط به ویژگی‌های کل‌نگر است. بعلاوه روش کل‌نگر تحلیل مؤلفه‌های مستقل هسته-محور در مقایسه با دیگر ویژگی‌ها نسبت به تغییرات بین فردی حساسیت بیشتری داشته است. بر اساس دانش محدود از فرد جدید نمونه‌های مجازی تولید و برای تقویت یادگیری سیستم بازشناسی استفاده شد. نتیجه بازشناسی مستقل از فرد این روش در مقایسه با روش پایه به صورت معنی‌داری ($P < 0.05$) بهبود داشته و مقدار صحت تشخیص آن ۹۱٫۳۹٪ است.

کلیدواژگان: بازشناسی جلوه‌های هیجانی چهره، سیستم مستقل از فرد، دانش اولیه، نمونه مجازی.

*عهده‌دار مکاتبات

نشانی: تهران، خیابان حافظ، دانشگاه صنعتی امیرکبیر، دانشکده مهندسی پزشکی، گروه بیوالکترونیک، کدپستی: ۱۵۸۷۵-۴۴۱۳

تلفن: ۰۲۱-۶۴۵۴۲۳۶۳، دورنگار: ۰۲۱-۶۶۶۶۸۱۸۶، پیام نگار: Towhidkhan@aut.ac.ir

۱- مقدمه

همین روش‌ها در شرایط مستقل از فرد به ترتیب دارای صحت بازشناسی ۶۰.۲۳٪ و ۶۴.۷۷٪ بود که بیانگر تأثیرپذیری آهنگ بازشناسی نسبت به تغییرات بین فردی است [۱۰].

برای افزایش قابلیت اطمینان سیستم‌های بازشناسی در محیط‌های مختلف و مقاوم بودن آنها نسبت به افراد مختلف راهکارهایی پیشنهاد شده است. این روش‌ها را به طور کلی می‌توان به سه گروه افزایش تعداد نمونه در دادگان تعلیم، تکنیک‌های فرونت‌اند^۷، و تکنیک‌های بازشناسی یا تطبیق، دسته‌بندی کرد [۱۱].

مسئله دیگر بهره‌برداری از اطلاعات، هر چند محدود، از فرد جدید است. در بازشناسی گفتار از این ایده به منظور بهبود بازشناسی برای گوینده جدید استفاده شده است [۱۲]. به طوری که بر اساس اطلاعات محدود نمونه‌های مجازی تولید و در بازشناسی با تطبیق سیستم یادگیری استفاده می‌شوند [۱۲، ۱۳].

در مورد یافته‌های مغز انسان، محققان نشان داده‌اند که انسان‌ها ابتدا به تعیین موقعیت چهره و شناسایی چهره فرد پرداخته؛ سپس حالات هیجانی شخص را شناسایی می‌کنند [۱۴]. با الهام گرفتن از این مطلب به‌کارگیری ایده تعیین مشخصات فردی برای تشخیص بهتر نوع هیجان‌ات موضوع مناسبی برای بهبود بازشناسی آنها خواهد بود.

تولید تصاویر ویدئویی مجازی از موضوعاتی است که مورد توجه است و در مواردی همچون بینایی ماشین و حوزه‌های گرافیک نظیر تولید انیمیشن، واسط کاربر- ماشین و تولید نماهای انسانی^۸ استفاده می‌شود [۱۵]. تعدادی از روش‌ها به تولید قاب (فریم) اوج بروز جلوه هیجانی توجه کرده‌اند، به طوری که بر اساس عکس محدودی از فرد هدف و میانگین اطلاعات پویایی جلوه‌های هیجانی افراد منبع نمونه‌هایی تولید می‌کنند [۱۶]. در مطالعه دیگری ابتدا با ضرایب مدل ظهور فعال^۹، شکل و بافت چهره از یکدیگر

جلوه‌های هیجانی چهره افراد عموماً ناشی از حالات درونی آنها است و نقش مهمی در مطالعات پدیده‌های روانشناختی، توسعه ارتباطات غیرکلامی^۱ و رفتارهای اجتماعی دارند [۱].

دسته‌ای از روانشناسان بروز هیجان‌ات در گفتار و اطلاعات دیداری را عینی^۲ می‌دانند. سیستم برچسب‌زنی^۳ FACS نیز اینگونه عمل می‌کند و آنچه را که در ظاهر چهره مشهود است معیار قرار داده؛ از آن طریق هیجان‌ات را از یکدیگر تفکیک می‌کند [۲].

برخی تحقیقات گذشته در زمینه بازشناسی خودکار واحدهای کنشی و جلوه‌های هیجانی، بر اساس تحلیل تصاویر استاتیک یا فریم‌های جداگانه از تصاویر متوالی انجام شده است [۳-۵]. همچنین تحقیقات دیگری با استفاده از بافت‌های پویا در بازشناسی حالات چهره یا کد کردن پویایی واحدهای کنشی انجام شده است [۶]. در سال ۲۰۱۰ خادمی و همکارانش برای مدلسازی هر واحد کنشی از یک مدل مخفی مارکف استفاده و با شبکه عصبی تک‌تک واحدها و ترکیب آنها را بازشناسی کردند [۷]. در مطالعه دیگری با استفاده از شبکه‌های بیزین پویا^۴ و دانش FACS سیستمی برای مدلسازی رفتار پویا و تصادفی حالات خودبخودی چهره ارائه شده است [۸].

در پژوهش‌های انجام شده به هندسه چهره نیز توجه شده است؛ به طوری که با استفاده از اطلاعات نقاط مشخصی از چهره و با ایده تخمین وضعیت خنثای آن به بهبود بازشناسی پرداخته شده است [۹]. در سال ۲۰۰۳ کوهن^۵ و همکاران رابطه کارایی سیستم‌های تحلیل اطلاعات چهره را با مسئله وابسته به فرد بودن نشان دادند و چندین طبقه‌بند استاتیک و پویا را به‌کار گرفتند. نتیجه طبقه‌بندی وابسته به فرد با یک بیزین گوسی و کوشی به‌ترتیب دارای صحت متوسط ۷۹/۳۶٪ و ۸۰/۰۵٪ بر ۵ سوژه^۶ بود؛ اما

¹Non verbal Communication⁵Cohen⁹Active Appearance Model (AAM)²Objective⁴Subject³Facial Action Coding System⁷Front-end⁴Dynamic Bayesian network⁸Avatars

بخش ۳ بر اساس دانش اولیه از فرد جدید، سیستم پیشنهادی ارائه شده است. در ادامه در بخش ۴ به نتایج و در بخش ۵ به بحث و جمع بندی نتایج پرداخته شده است.

۲- بازشناسی جلوه‌های هیجانی چهره

یک سیستم بازشناسی جلوه‌های هیجانی دارای چند بخش اصلی است: انتخاب محدوده چهره، پیش پردازش، استخراج ویژگی و طبقه بندی. شکل (۱) روندنمای کلی روش‌های اجرا شده را برای بازشناسی جلوه‌های هیجانی چهره نشان می‌دهد.

یکی از عوامل ایجاد تنوع در بروز هیجانات، اختلاف اشخاص در بروز هیجانات است. هدف بسیاری از روش‌ها استخراج مشخصاتی از جلوه‌های هیجانی چهره است که در افراد مختلف اختلاف کمتری داشته باشند. برای کاربردی شدن روش‌ها می‌توان سیستم را مبتنی بر فرد نیز به کار گرفت؛ اما بمنظور کاربرپسند کردن و توسعه کاربردی سیستم‌ها، به ارائه روش‌های مستقل از فرد نیز نیاز است.

۲-۱- استخراج ویژگی

استخراج ویژگی‌های مناسب شروع مناسبی برای افزایش توان سیستم بازشناسی است. ویژگی‌های هندسی شامل شکل اجزای چهره و موقعیت آنها نسبت به یکدیگر است، اما در رویکرد مبتنی بر ظهور بر حرکت پوست و تغییرات بافت چون چین و چروک‌ها، شیارها و برآمدگی‌ها تأکید شده است. هر دو رویکرد مزایا و معایبی دارند. به طوری که روش هندسی از تغییرات بافت‌ها غافل بوده، و رویکرد دوم به تغییرات نور محیط و تغییرات بین فردی حساس است. در حوزه رویکردهای مبتنی بر ظهور، ویژگی‌هایی همچون ضرایب موجک گابور^{۱۱}، شار نوری و اندازه‌گیری مشخصات ظاهری مانند طول چین و چروک‌ها از جمله ویژگی‌هایی است که در مطالعات قبل پیشنهاد شده‌اند [۱۹،۲۰].

تفکیک شده؛ سپس نمونه مجازی تولید شده است [۱۷]. علاوه بر روش محاسباتی برای تولید نمونه ویدئویی مجازی پیشنهاد شده که در مقابل نویز مقاوم است و به تعداد نمونه آموزش کمی نیاز دارد. همچنین در این روش از طریق یک ساختار سلسله‌مراتبی تغییرات جابجایی تصویر منبع به فرد هدف منتقل می‌شود [۱۸].

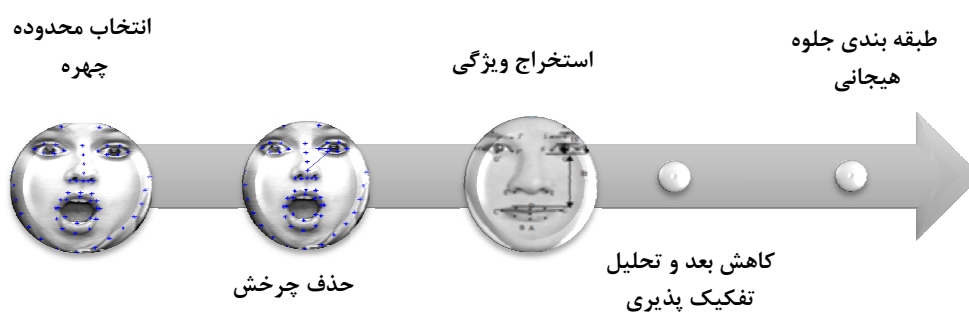
آنچه در این مقاله مورد توجه قرار گرفته است، اولاً تحلیل جایگاه تعمیم‌دهی سیستم بازشناسی نسبت به فرد جدید و ثانیاً ارائه روشی جهت افزایش تعمیم‌دهی به شخص جدید است. در همین راستا مسأله‌ای که در این مقاله به آن توجه شده، در نظر گرفتن اطلاعات مربوط به تولید هیجان در شخص جدید است، به طوری که با استفاده از حداقل اطلاعات از فرد جدید ویژگی‌های احتمالی بروز هیجانات تخمین زده شده؛ در تولید نمونه مجازی به کار گرفته می‌شود. سپس از این نمونه‌ها برای افزایش یادگیری سیستم مستقل از فرد استفاده می‌شود. در مقایسه با مطالعات انجام شده نوآوری‌های روش پیشنهادی به شرح زیر است:

- برای تولید نمونه مجازی مشابه بعضی الگوریتم‌ها روشی پیشنهاد شده است که علاوه بر به‌کارگیری مانیفولد^{۱۰} مخصوص هر جلوه، قادر به القای مناسب اطلاعات بروز فرد منبع به تصویر فرد هدف و تولید متنوعی از پروزهاست.
- از نمونه‌های مجازی بردارهای ویژگی برای افزایش قدرت تعمیم و بهبود سیستم بازشناسی استفاده شده است.
- این سیستم بر اساس اطلاعات کم از فرد جدید فعالیت می‌کند.

برای نشان دادن تأثیر عوامل موثر بر بازشناسی در بخش ۲ تحلیل‌هایی انجام شده است. بر اساس نتایج این بخش یک سیستم پایه بازشناسی پیشنهاد و پیاده‌سازی شده، و در

^{۱۰} Manifold

^{۱۱} Gabor



شکل (۱) - روندنمای کلی بازشناسی جلوه‌های هیجانی چهره

استفاده شده است [۲۳]. دسته سوم ویژگی‌ها اطلاعات مربوط به تصویر چهره 20×20 شده فرد است که به صورت سیگنالی با بعد 400 در آمده است.

۲-۲- طبقه‌بند

بمنظور در نظر گرفتن مشخصات استاتیک بروز هیجانات، دو طبقه‌بند سریع و توانمند بیزین و آدابوست^{۱۲} استفاده شدند. در طبقه‌بند بیزین خطی ابتدا با روش تفکیک‌پذیری خطی بردار ویژگی را به فضایی با بعد کمتر در راستای تفکیک‌پذیری بیشتر نگاشت کرده؛ سپس با طبقه‌بند بیزین و با فرض اشتراک ماتریس کواریانس برای طبقه‌ها یک طبقه‌بند خطی برای طبقه‌بندی بردارها به کار گرفته شد. طبقه‌بند آدابوست مبتنی بر الگوریتم تقویت‌کننده^{۱۳} با محوریت یادگیرنده ضعیف درخت تصمیم *CART* است. برای پیاده‌سازی این طبقه‌بند از جعبه ابزار برنامه *MATLAB* با عنوان *GML AdaBoost* دو طبقه و با ترکیب یکی و بقیه استفاده شد [۲۴].

در شکل (۲) جایگاه تحلیل‌های حساسیت نسبت به نویز و شخص بروز دهنده در قالب روندنمایی نشان داده شده است؛ به طوری که هدف مقاله در کنار معرفی ویژگی‌های تفکیک‌کننده هیجانات، تبیین اثرپذیری ویژگی‌ها نسبت به تغییرات است. بطور خاص در این پژوهش بر تغییرات ناشی از تغییر شخص بروز دهنده و تغییر محیط تصویربرداری- که با نویز نشان داده شده- توجه شده است.

بر طبق جدول پیوست هیجانات با واحدهای کنشی رابطه قاعده‌مند دارند، بنابراین بمنظور بازشناسی هیجانات می‌توان ابتدا فعالیت واحدهای کنشی را شناسایی و سپس هیجان را تشخیص داد و یا مستقیماً به تشخیص هیجانات پرداخت. برای استخراج ویژگی‌های مناسب در این قسمت از برجسب‌های مربوط به 68 نقطه مشخص در چهره که با روش مدل‌های ظهور فعال ارائه شده، استفاده می‌شود [۲۱]. در دسته اول حدود 119 مشخصه از هندسه چهره و اطلاعات بافت چهره استخراج شده است [۸] که قبلاً پیشنهاد شده‌اند [۲۲].

در دسته دوم- که مبتنی بر روش‌های کل‌نگر است- با روش یادگیری مانیفولد تحلیل مؤلفه‌های اساسی هسته محور ویژگی‌هایی از تصویر چهره استخراج شده است. برای استخراج اطلاعات از کاهش بعد غیر خطی سیگنال مربوط به تصویر با طول 20×20 - که با نمونه‌برداری از تصویر اصلی بدست آمده است- استفاده شد. روش کاهش بعد تحلیل مؤلفه‌های اساسی هسته محور بازنویسی جدیدی برای روش سنتی تحلیل مؤلفه‌های اساسی است. در این روش به جای بردارهای ویژه ماتریس کواریانس، بردارهای ویژه ماتریس هسته محاسبه می‌شود. این ماتریس از ضرب داخلی نقاط نگاشت شده توسط تابع هسته به دست می‌آید؛ برای پیاده‌سازی از جعبه ابزار کاهش بعد غیر خطی ماتن



شکل (۲)- روندنمایی تحلیل حساسیت روش های بازشناسی

فرضیه از نوع چندجمله‌ای تک متغیره درجه n باشد، بعد VC آن $n+1$ است. طبقات فرضیه‌ای که بزرگتر و پیچیده‌تر هستند بعد VC بیشتری دارند [۱۲].

همانطور که مشاهده شد تعداد نمونه مناسب با بعد VC مرتبط است. این موضوع ترغیب می‌کند که از طبقه فرضیات با بعد کم استفاده شود، در حالی که استفاده از اینگونه فرضیات کافی نیست، زیرا حتی اگر در این طبقات بهترین تابع پیدا شود خطای تعمیم افزایش خواهد یافت. بطور مثال می‌توان به حالتی اشاره کرد که لازم است تابعی با L نمونه و یک شبکه با n نرون میانی تخمین زده شود. همانطور که در رابطه (۱) نشان داده شده است دو نوع خطا خواهیم داشت یکی تقریب و دیگری تخمین که اولی مربوط به محدودیت تعداد پارامتر مدل است و دیگری به کمبود داده ارتباط دارد. هرچه توان بازنمایی داده افزایش می‌یابد توان تقریب آن کاهش می‌یابد [۱۲].

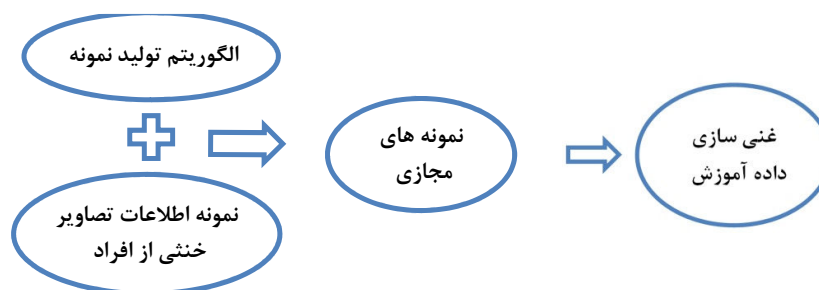
$$\|f_0 - \hat{f}_{n,l}\| \leq e_{app}(n) + e_{est}\left(\frac{VC(\mathcal{H}_n)}{L}\right) \quad (1)$$

اگر تعداد پارامتر مدل افزایش یابد خطای تقریب کاهش می‌یابد، در حالی که خطای تخمین آن با فرض ثابت بودن داده افزایش می‌یابد. بنابراین باید مصالحه‌ای بین پارامترهای مدل انجام شود. مشخص است اگر طبقه فرضیاتی مورد نظر باشند که محدودیتی ندارند، نیازمند داده بسیار زیاد برای

۳- روش پیشنهادی مبنی بر به‌کارگیری دانش اولیه

بر اساس تحلیل‌های بخش قبل چالش مستقل کردن سیستم بازشناسی از فرد، به‌عنوان چالش مهمتر مطرح شده است [۲۲، ۱۰]. به همین منظور روشی برای بهبود بازشناسی در حالت مستقل از فرد پیشنهاد شده است. از آنجایی که سبک فرد جدید فضای جدیدی نسبت به فضای یادگیری از فضای افراد آموزش ایجاد می‌کند، مسأله تعمیم مطرح می‌شود. بر اساس مرور پیشینه تحقیقات یک راه‌حل برای بهبود تعمیم سیستم بازشناسی، بهره‌گیری از اطلاعات هر چند محدود از فرد جدید است [۱۲]. می‌توان چنین تحلیل کرد که استفاده از اطلاعات اولیه نقشی تنظیم‌کننده^{۱۴} دارد. این موضوع بصورت تحلیلی اثبات شده است و مواردی از آن در ادامه آورده شده است [۱۲].

اگر یادگیری به‌صورت تخمین تابع در نظر گرفته شود و اگر تعداد نمونه کافی برای یادگیری وجود داشته باشد، امید است خروجی تابع به هدف مورد نظر نزدیک شود. در نتیجه امکان تعمیم بر نمونه جدیدی که قبلاً در آموزش نبوده است، فراهم می‌شود. در یادگیری مبتنی بر مثال، تعیین تعداد نمونه لازم برای داشتن تعمیمی مناسب، فرضیه‌ای کلیدی است. مسأله پیچیدگی نمونه ابتدا توسط نظریه وپنیک و چرووننکیس^{۱۵} مشخص شد؛ که به توصیف قاعده‌ای کلی می‌پردازد. به طوری که همه ماشین‌های یادگیری مبتنی بر احتمال باید از آن پیروی کنند. اگر یادگیری یک فرضیه h ($h \in \mathcal{H}$) بر اساس مجموعه مثال در نظر گرفته شود، خطای تعمیم‌دهی آن با تعداد مثال‌ها یعنی $\sqrt{VC(\mathcal{H})/L}$ وابسته است که در آن $VC(\mathcal{H})$ بعد VC طبقه \mathcal{H} و L تعداد نمونه تعلیم است. بطور مثال اگر طبقه



شکل (۳) - روندنمای نحوه تولید نمونه مجازی و غنی سازی داده آموزش

CK است. در این دادگان در مجموع تصاویر ۱۱۸ سوژه موجود است که ۳۲۶ داده ترتیبی است. برای استخراج مشخصات هندسی چهره افراد از وضعیت نقاط مشخصه چهره که با روش مدل ظهور فعال مشخص شده، استفاده شد. این نقاط شامل ۶۸ نقطه مشخصه از چهره بود. برای تحلیل از قاب‌هایی که برچسب ۷ جلوه هیجانی را داشتند استفاده و سه فریم آخر هر توالی برداشته شد. متوسطی از تصاویر ۷ جلوه هیجانی در دادگان در شکل (۴) نشان داده شده است [۲۱].



شکل (۴) - متوسط تصویر ۷ حالت هیجانی چهره در دادگان CK+ برای همه افراد

۴-۲- تحلیل حساسیت نسبت به استقلال از شخص در حالت ارزیابی ۱۰ دسته ۱۰ بار و ویژگی‌های دسته اول صحت بازشناسی ۸۱٪ برای طبقه‌بند بی‌زین با کاهش بعد تفکیک‌پذیر خطی و صحت بازشناسی ۸۷٪ برای آداپوست بدست آمد. با وجود بیشتر بودن میانگین نرخ بازشناسی

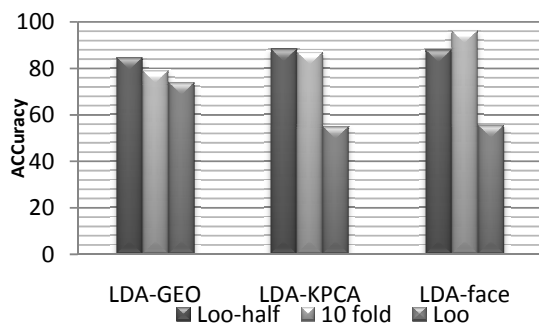
داشتن خطای تخمین کم هستند و اگر طبقه فرضیه محدود انتخاب شود خطای تقریب برای داشتن تعمیم بهتر افزایش می‌یابد. برای حل این مسأله اگر دانش اولیه‌ای از مسأله موجود باشد قابل حل است. اگر اطلاعات کمی در مورد تابع هدف وجود داشته باشد معادل با داشتن توابع هدف کوچکتر است و مسأله‌ای با خطای تعمیم کم وجود خواهد داشت. یک راه برای به‌کارگیری دانش اولیه استفاده از نمونه‌های مجازی است.

در این پژوهش تولید نمونه مجازی بر اساس الگوریتمی مبتنی بر روش تجزیه مقادیر ویژه و با تفاوت‌هایی پیاده‌سازی شد [۱۳]. از نمونه‌های مجازی تولید شده برای افزایش توان یادگیری سیستم استفاده شد. روندنمایی کلی روش در شکل (۳) نشان داده شده است. در این روش در کنار مراحل قبلی که برای بازشناسی جلوه‌های هیجانی ارائه شده است، قسمت غنی‌سازی داده آموزش اضافه شده است. همانطور که شکل (۳) نشان می‌دهد روش پیشنهادی با استفاده از اطلاعاتی از تصویر ختای فرد جدید و افراد آموزش نمونه‌هایی مجازی برای غنی‌سازی داده آموزش ایجاد کرده که در نقش تنظیم‌کننده باعث افزایش تعمیم‌دهی سیستم بازشناسی می‌شوند.

۴- نتایج

۴-۱- دادگان Cohn-Kanade-2010

در این مقاله از دادگان CK+ استفاده شد که توسعه یافته



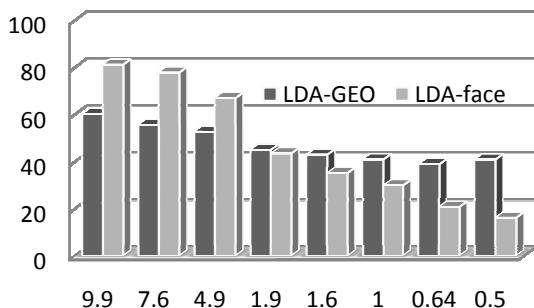
شکل (۵)- مقایسه نرخ بازشناسی سه روش در ارزیابی با روش Loo و ۱۰ دسته ۱۰ بار و Loo-half

۴-۳- تحلیل حساسیت نسبت به نویز

برای بررسی حساسیت روش‌ها نسبت به نویز اطلاعات تصاویر با نویز گوسی با میانگین صفر و واریانس متفاوت جمع زده و اثر صحت بازشناسی در حالات ۱۰ دسته ۱۰ بار با یکدیگر مقایسه شد. جهت محاسبه سیگنال به نویز از فرمول (۲) استفاده شد [۲۵].

$$SNR = \frac{\sqrt{\sum_{xy}(f(x,y) - avg(f))^2}}{\sqrt{\sum_{xy}(n(x,y) - avg(f))^2}} \quad (2)$$

مقدار سیگنال به نویز از حدود ۰.۵ تا ۱۰ تغییر داده شده است. فرض بر این بود که تنها داده آزمایش به نویز آغشته است. صحت روش‌ها دارای کاهش نرخ متفاوتی به صورت نمودار شکل (۶) است. در این قسمت تنها اثر نویز بر ویژگی‌های بافت بررسی شد.



شکل (۶)- مقایسه نرخ بازشناسی دو روش در سیگنال به نویزهای مختلف

روش آدابوست توزیع صحت بر هیجان‌نا مناسب بود؛ به همین دلیل در ادامه روال، از طبقه‌بند بی‌زین و کاهش بعد تفکیک‌پذیر خطی استفاده شد. برای گزارش حالت مستقل از فرد روش ارزیابی LOO^{13} استفاده شد. در این روش در هر مرحله داده هر فرد جدا شده و بقیه داده به عنوان آموزش استفاده شد. نرخ بازشناسی همه روش‌ها در این مرحله کاهش پیدا کرد و بهترین نتیجه 76.92% بود که با روش کاهش بعد تفکیک‌پذیر خطی و طبقه‌بند بی‌زین بر ویژگی‌های هندسه-بافت بدست آمد.

همچنین در حالت LOO نیمی از داده فرد برای آموزش استفاده شد که روش ارزیابی $LOO-half$ نام گذاری شده است. در این حالت- که وابسته بودن سیستم بازشناسی به شخص را نشان می‌دهد- روش‌ها بهترین نتیجه را داشته‌اند که بیانگر جایگاه اصلی اطلاعات شخصی در بروز تنوعات است. در کنار روش LOO از روش ارزیابی $10-fold$ که به صورت تصادفی داده افراد را برای آموزش و آزمون به‌کار می‌گیرد- نیز استفاده شده است.

همانطور که در شکل (۵) نشان داده شده است، در هر سه روش نرخ بازشناسی در حالت LOO کاهش پیدا کرده است. در حالت استفاده از ویژگی‌های هندسه-بافت و کاهش بعد تفکیک‌پذیر خطی و طبقه‌بند بی‌زین نرخ بازشناسی از 84.74% در حالت $LOO-half$ به 74% در حالت LOO کاهش و در روش کاهش بعد تفکیک‌پذیر خطی و تحلیل مؤلفه‌های اساسی هسته محور کل تصویر و طبقه‌بند بی‌زین نیز نرخ بازشناسی از 88.35% به 55% در حالت LOO کاهش پیدا کرده است. بهترین نتیجه مربوط به روش $LDA-face$ توسط روش $LOO-half$ است.

۴-۴- نتایج روش مبتنی بر دانش اولیه

بر اساس تحلیل‌های بخش قبل چالش مستقل کردن سیستم بازشناسی از فرد، به‌عنوان چالش مهم‌تر مطرح شده است [۱۰، ۲۲]. به همین منظور روشی برای بهبود بازشناسی در حالت مستقل از فرد پیشنهاد شد. با توجه به حساسیت کمتر اطلاعات هندسی به تغییرات بین فردی، در سیستم پایه و روش پیشنهادی این ویژگی‌ها مورد توجه قرار گرفت. پارامترهای مدل ظهور فعال استخراجی از چهره افراد یعنی موقعیت افقی و عمودی نقاط مشخصه چهره به‌عنوان بردار ویژگی انتخاب شدند. برای هر قاب در زمان t بردار ویژگی $F_f(t)$ شامل ۶۸ موقعیت نقاط مشخصه است که برای مختصات x و y یک بردار ۱۳۶ بعدی مشابه بردار (۳) می‌شود:

$$f_1(p_{i,y,t}) = p_{i,y,t} f_2(p_{i,x,t}) = p_{i,x,t} \quad \sqrt{i=1:68} \quad (3)$$

$$F_f(t) = [f_1(t) f_2(t)]$$

در اینجا $p_{i,y,t}$ موقعیت y نقطه i در زمان t و $p_{i,x,t}$ موقعیت x نقطه i در زمان t است. این ویژگی‌ها با تحلیل مؤلفه‌های اساسی کاهش بعد داده شد تا اطلاعات اضافی از آن کسر شود.

بمنظور نشان دادن تأثیر روش پیشنهادی برای بهبود بازشناسی مستقل از فرد، سیستم پایه‌ای برای بازشناسی پیاده‌سازی شد. در سیستم پایه ویژگی‌های استخراجی با روش تحلیل مؤلفه‌های اساسی و تحلیل تفکیک‌پذیر خطی کاهش بعد پیدا کردند و با طبقه‌بند بیزین طبقه‌بندی شدند.

نتایج بازشناسی به ازای ۷ جلوه هیجانی با روش ارزیابی LOO برای روش پایه و راهکار پیشنهادی برای مواجهه با تنوعات ناشی از فرد جدید در جدول (۱) آورده شده است. همانطور که در جدول (۱) مشاهده می‌شود، چهار حالت برای هر روش در نظر گرفته شده است. در حالت ۱ و حالت ۳ داده آموزش شامل موارد مجازی تولید شده برای

افراد آموزش نبوده و تنها از نمونه مجازی تولیدی برای فرد جدید بمنظور غنی‌سازی استفاده شده است. در حالی که در حالت‌های ۲ و ۴ اطلاعات مجازی تولید شده برای افراد داده آموزش و فرد جدید، هر دو برای غنی‌سازی داده آموزش و در نتیجه آموزش سیستم استفاده شده است. در حالت‌های ۱ و ۲ در داده آموزش، همه قاب‌های یک ترتیب داده استفاده شده‌اند، در حالی که در حالت‌های ۳ و ۴ تنها ۳ قاب نزدیک به نقطه اوج بروز برای یادگیری استفاده شدند. نتیجه روش پیشنهادی حالت ۳ در مقایسه با روش پایه در حالت ۱ به‌صورت معنی‌داری ($p < 0.05$) بهبود داشته است.

جدول (۱) - نتیجه نرخ بازشناسی ۷ جلوه هیجانی سیستم پایه و روش پیشنهادی (%).

روش/حالت	مد ۱	مد ۲	مد ۳	مد ۴
روش پایه	۸۸.۴۷	۸۲.۴۵	۹۰.۱۷	۸۵.۴۶
روش پیشنهادی	۸۸.۰۴	۸۲.۹۱	۹۱.۳۹	۸۶.۴۲

برای تولید نمونه مجازی به ازای بردار تصویر خنثای هر فرد از ۷ جلوه هیجانی ۵ نفر دیگر استفاده شد که در مجموع ۳۵ ترتیب داده مجازی برای هر تصویر خنثی ساخته شد. در بین روش‌های اجرا شده میانگین بهترین نتیجه مربوط به روش پیشنهادی در حالت ۳ بود که مطابق جدول (۱) دارای صحت ۹۱.۳۹٪ است.

علاوه بر میانگین نرخ بازشناسی ۷ جلوه هیجانی، ماتریس اختلاط بین جلوه‌های هیجانی برای بهترین روش به شرح جدول (۲) آورده شده است. در این جدول بدترین نرخ مربوط به هیجان تحقیر و بهترین نتیجه مربوط به هیجان تنفر است. اختلاط تحقیر با عصبانیت و ناراحتی، بالاترین اختلاط است.

جدول (۲) - ماتریس اختلاط روش پیشنهادی در مد ۳

	عصبانیت	تحقیر	تفر	ترس	خوشحالی	ناراحتی	تعجب
عصبانیت	۸۹	۱	۳	۰	۰	۷	۰
تحقیر	۱۷	۶۷	۰	۰	۰	۱۷	۰
تفر	۱	۱	۹۸	۰	۰	۰	۰
ترس	۰	۰	۱	۸۱	۱۰	۸	۰
خوشحالی	۰	۳	۰	۰	۹۷	۰	۰
ناراحتی	۶	۷	۰	۰	۰	۸۷	۰
تعجب	۰	۳	۰	۲	۰	۰	۹۵

جدول (۳) - مقایسه نتایج بازشناسی جلوه‌های هیجانی با نتایج دیگران

مراجع	تعداد	طبقه	پویایی	ارزیابی	نرخ
[۲۶]	۹۴	۷	خیر	۱۰ گروه	۹۱/۵۱
[۲۷]	۹۶	۷	خیر	۱۰ گروه	۹۱/۴۰
[۲۱]	۱۱۸	۷	خیر	LOO	۸۳/۳۲
پیشنهادی	۱۱۸	۷	-	LOO	۹۱/۳۹

در این بخش بهترین نتیجه بدست آمده از روش پیشنهادی، با نتایج تعدادی از روش‌های قبلی به‌کار گرفته شده بر روی دادگان CK، در جدول (۳) مقایسه شده است. این نکته قابل ذکر است که نتایج همه مطالعات مستقیماً قابل مقایسه نیستند؛ زیرا دارای اختلاف در انتخاب تعداد سوژه، تعداد هیجان پایه و یا نحوه ارزیابی هستند.

نتایج روش‌های قبل که با تعداد سوژه کمتر و با معیار ارزیابی ۱۰ دسته ۱۰ گروه انجام گرفته است، با وجود تنوعات بین فردی کمتر در دادگان نتیجه‌ای نزدیک به نتیجه روش پیشنهادی دارند [۲۶،۲۷].

عمده روش‌های قبل بر توسعه ویژگی‌های استخراجی از تصاویر جلوه هیجانی متمرکز بوده‌اند و پیش‌بینی می‌شود همانطور که در این کار نشان داده شد با اضافه کردن فرایند غنی‌سازی دادگان آموزش از طریق داده مجازی قدرت تعمیم آنها نیز افزایش یابد و در نتیجه نرخ بازشناسی افزایش پیدا کند. به دلیل اضافه کردن این قسمت به فرایند بازشناسی حجم محاسبات افزایش داشته و در نتیجه روش پیشنهادی در مقایسه با کارهای قبل کندتر عمل می‌کند [۲۶،۲۷].

نتیجه‌ای که به عنوان پایه نتایج بر روی دادگان CK+ گزارش شده است، دارای صحت طبقه‌بندی بر ۴۹۰ تصویر از هیجان‌ات مختلف است که با روش ارزیابی LOO دارای میانگین ۸۳٪ است. نتایج این روش در جدول (۴) با بهترین نتیجه مستقل از فرد ارائه شده در این کار با میانگین ۹۱/۳۹٪ مقایسه شده است [۲۱].

همچنین نتایج نشان می‌دهند روش‌های کل‌نگر تأثیرپذیری بیشتری ۱/۰ نسبت به اطلاعات فرد بروز دهنده دارند و کاهش نرخ بازشناسی آنها بیشتر است. روش‌های کل‌نگر تلاش می‌کنند تا تمام توان تشخیص الگوهای مؤثر در بازشناسی جلوه‌های هیجانی را از طریق محاسبات حل کنند، در حالی که دانش مربوط به تمایز جلوه‌ها توانمندتر قادر است این اطلاعات را دست‌یافتنی کند. در نتیجه با توانمندسازی روش‌های محاسباتی توسط قواعد، توان بازشناسی افزایش پیدا خواهد کرد.

جدول (۴) - مقایسه نرخ بازشناسی به ازای هر جلوه هیجانی روش پیشنهادی در مد ۳ و نتیجه پایه دادگان CK+ (%)

	عصبانیت	تحقیر	تفر	ترس	خوشحالی	ناراحتی	تعجب	ختی
[۲۱]	۷۵	-	۹۴.۷۴	۶۵.۲۲	۱۰۰	۶۸	۷۷.۰۹	۱۰۰
پیشنهادی	۸۹	۶۷	۹۸	۸۱	۹۷	۸۷	۹۵	-

۴- نتیجه گیری

همانطور که نشان داده شد مستقل کردن سیستم بازشناسی از فرد بروز دهنده از عوامل ایجاد تنوعات و کاهش نرخ بازشناسی است. همچنین در بررسی حساسیت به نویز روش های کل نگر با فرض حفظ اطلاعات هندسی چهره نسبت به نویز حساس تر است و نرخ بازشناسی آنها به شدت کاهش می یابد. بنابراین با وجود اینکه بهترین نتیجه مربوط به روش کل نگر بود، این روش در بخش حساسیت نسبت به نویز و استقلال سیستم رتبه پائین تری داشت.

بر اساس تحلیل های انجام شده ویژگی های هندسی چهره با دانش کمی از اطلاعات فرد جدید به کار گرفته شد تا چالش کاهش نرخ ناشی از تنوع بروز فرد جدید برطرف شود. این کار از طریق تولید نمونه های مجازی و به کارگیری آنها در تقویت سیستم یادگیری صورت گرفت. نتیجه روش پیشنهادی در حالت های مختلف بررسی شد و روش پیشنهادی در حالت ۳ قادر به بازشناسی با دقت ۹۱/۳۹٪ شد که بصورت معنی داری در مقایسه روش پایه در نظر گرفته شده، بهبود داشت. این نتیجه با مطالعات قبل نیز قابل مقایسه است.

در این کار خطای تشخیص نقاط مشخص چهره در نتیجه وجود نویز در نظر گرفته نشده است و بهتر است در قسمت بررسی نرخ بازشناسی و حساسیت آن به نویز این موضوع نیز لحاظ شود. در این مقاله در کنار ویژگی های هندسی به ویژگی های ساده ای از چین و چروک های چهره توجه شد. بعلاوه لازم است ویژگی های مناسب تری از بافت همچون توصیفگر الگوهای باینری محلی را به کمک ویژگی های هندسی آورد [۲۸]. رویکرد آینده افزایش تعمیم دهی از طریق تولید نمونه مجازی، تولید توالی های مجازی با کیفیت، با تنوع زیاد و مبتنی بر اطلاعات ذاتی فرد جدید است تا تغییرات احتمالی ناشی از بروز یک جلوه در

در ادامه تحلیل های انجام شده یک سیستم مستقل از فرد برای بازشناسی جلوه های هیجانی چهره به کار گرفته شد و روشی برای بهبود بازشناسی آن پیشنهاد شد. روش پیشنهادی مبتنی بر دانش اولیه ای از فرد (تصویر خنثی فرد) است و استفاده از روش های تولید توالی مجازی نمونه هایی از جلوه های مختلف را بر اساس تصویر خنثای فرد تولید می کند. این نمونه ها بمنظور تقویت سیستم یادگیری و افزایش تعمیم سیستم برای فرد جدید به کار گرفته شدند. همانطور که پیشتر بدان پرداخته شد، به کارگیری دانش اولیه نقشی تنظیم کننده داشته؛ تعمیم سیستم بازشناسی را افزایش داده است. همانطور که نتایج جدول (۱) نشان می دهد افزایش قدرت تعمیم دهی به صورت معنی داری ($P < 0.05$) نرخ بازشناسی سیستم را افزایش داده است.

روش پیشنهاد شده برای تولید نمونه مجازی در این کار دو فرق با روش قبل داشت. یکی اینکه تابع انتقال در زیر فضا بصورت پیوسته از فریم خنثی تا اوج انتقال داده شد تا اطلاعات بروز هیجان هر فرد منبع در فضای اطلاعات تصویر خنثای فرد جدید القا شود و سپس تصاویر مربوط به اوج از ترتیب داده انتخاب شد. فرق دوم اینکه به جای به کارگیری الگوریتم تولید نمونه مجازی بر داده خام تصاویر این الگوریتم بر فضای بردار ویژگی های استخراجی از تصاویر اعمال شد که کارایی بیشتری ایجاد کرده است [۱۳، ۱۶].

غنی سازی داده آموزش در حوزه شناسایی چهره مبتنی بر تک تصویر از فرد قبلاً پیشنهاد شده است [۱۳، ۱۶]؛ اما به کارگیری این ایده در حوزه بازشناسی جلوه های هیجانی از تصاویر چهره- تا آنجا که اطلاع وجود دارد- تنها پژوهش ارائه شده است.

فرد جدید را پیش‌بینی کرده؛ در یادگیری استفاده شود. قابل
 ذکر است مقاوم کردن روش‌ها نسبت به نویز و تنوع ناشی
 از بروز جلوه‌ها در افراد مختلف به توسعه روش‌های

پیوست : رابطه بین ۶ هیجان پایه و واحدهای کنشی و ویژگی‌های گذرا [۸]										
نوع هیجان	علائم اصلی قابل مشاهده					علائم فرعی قابل مشاهده				
	شماره واحد کنشی					شماره واحد کنشی				
خوشحالی	۶	۱۲				۲۵	۲۶	۱۶		
ناراحتی	۱	۱۵	۱۷			۴	۷	۲۵	۲۶	
تنفر	۹	۱۰				۱۷	۲۵	۲۶		
تعجب	۵	۲۶	۲۷	۲+۱						
عصبانیت	۲	۴	۷	۲۳	۲۴	۱۷	۲۵	۲۶	۱۶	
ترس	۲۰	۵+۱	۷+۵			۴	۵	۷	۲۵	۲۶

۵- مراجع

- [6] Pantic M., Patras I., Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences; Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 2006; 36(2): 433-449.
- [7] Calder A.J., Young A.W., Understanding the recognition of facial identity and facial expression; Nature Reviews Neuroscience, 2005; 6(8): 641-651.
- [8] Zhang Y., Ji Q., Active and Dynamic Information Fusion for Facial Expression Understanding from Image Sequences; Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2005; 27(5): 699-714.
- [9] Ulukaya S., Affect recognition from facial expressions for human computer interaction; MScİstanbul, 2011.
- [10] Cohen I., Sebe N., Garg A., et al., Facial expression recognition from video sequences: temporal and static modeling; Computer Vision and Image Understanding, 2003; 91: 160-187.
- [1] Lien J., Automatic Recognition of Facial Expressions Using Hidden Markov Models and Estimation of Expression Intensity; Electrical Engineering, Washington University, 1998.
- [2] Gallegos D.R., Tranel D., Positive facial affect facilitates the identification of famous faces; Brain and Language, 2005; 93(3): 338-348.
- [3] Bartlett M.S., Littlewort G., Frank M., et al., Fully Automatic Facial Action Recognition in Spontaneous Behavior; in Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on, 2006: 223-230.
- [4] Littlewort G., Bartlett M.S., Fasel I., et al., Dynamics of facial expression extracted automatically from video; Image and Vision Computing, 2006; 24: 615-625.
- [5] Kotsia I., Pitas I., Facial expression recognition in image sequences using geometric deformation features and support vector machines; IEEE Trans. Image Processing, 2007; 16(1): 172-187.

- [20] Freitag C., Schwarzer G., Influence of emotional facial expressions on 3–5-year-olds' face recognition; *Cognitive Development*, 2011; 26(3): 230-247.
- [21] Lucey P., Cohn J.F., Kanade T., et al., The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression; in *Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE Computer Society Conference on, 2010: 94-101.
- [22] Mohammadian A., Aghaeinia H., Towhidkhah F., Geometric and texture based facial expression recognition," in *BioMedical Engineering*, 19th Iranian Conference on, Tehran, 2012.
- [23] Van der Maaten L., Postma E., Van Den Herik H., Dimensionality reduction: A comparative review," *Journal of Machine Learning Research*, 2009; 10: 1-41.
- [24] Friedman T.H.J., Tibshirani R., Additive logistic regression: A statistical view of boosting; *The Annals of Statistics*, 200; 38(2): 337–374.
- [25] Chen G.Y., Bui T.D., Krzyżak A., Invariant pattern recognition using radon, dual-tree complex wavelet and Fourier transforms; *Pattern Recognition*, 2009; 42(9): 2013-2019.
- [26] Gu W., Xiang C., Venkatesh Y.V., et al., Facial expression recognition using radial encoding of local Gabor features and classifier synthesis; *Pattern Recognition*, 2012; 45(1): 80-91.
- [27] Shan C., Gong S., McOwan P.W., Facial expression recognition based on Local Binary Patterns: A comprehensive study; *Image and Vision Computing*, 2009; 27(6): 803-816.
- [28] Jafarzadeh M., Mohammadian A., Using Dimension Reduction of Local Binary Patterns To improve Facial Expression Recognition; in *Electrical Engineering*, 20th Iranian Conference on, Tehran, 2012.
- [11] Xuan T.W., Rapid speaker adaptation by variable reference model subspace; *Université de Rennes* 2008.
- [12] Niyogi P., Girosi F., Poggio T., Incorporating prior information in machine learning by creating virtual examples; *Proceedings of the IEEE*, 1998; 86(11): 2196-2209.
- [13] Mohammadzade H., Hatzinakos D., Projection into Expression Subspaces for Face Recognition from Single Sample per Person; *Affective Computing*, *IEEE Transactions on*, 2013; 4(1): 69-82.
- [14] Gu W., C. X. Ā, Venkatesh Y.V., et al., Facial expression recognition using radial encoding of local Gabor features and classifier synthesis; *Pattern Recognition*, 2012; 45(1): 80-91.
- [15] Fan W., Bouguila N., Novel approaches for synthesizing video textures; *Expert Systems with Applications*, 2012; 39(1): 828-839.
- [16] Mohammadzade H., Hatzinakos D., An expression transformation for improving the recognition of expression-variant faces from one sample image per person; in *Biometrics: Theory Applications and Systems (BTAS)*, 2010 Fourth IEEE International Conference on, 2010: 1-6.
- [17] Spangler S.M., Schwarzer G., Korell M., et al., The relationships between processing facial identity, emotional expression, facial speech, and gaze direction during development; *Journal of Experimental Child Psychology*, 2010; 105(1-2): 1-19.
- [18] Yang Y., Zheng N., Liu Y., et al., Expression transfer for facial sketch animation; *Signal Processing*, 2011; 91(11): 2465-2477.
- [19] Tian Y.I., Kanade T., Cohn J.F., Recognizing action units for facial expression analysis; *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on*, 2001; 23(2): 97-115.