



تشخیص حالت چهره با استفاده از توصیفگر محلی جهتدار نویسنده اول^۱ عبدالعلی دماوندی (نویسنده مسئول مقاله)، نویسنده دوم^۲ رضا روشنی

^۱موسسه آموزش عالی غیر انتفاعی لامعی گرگانی، گرگان، ایران Abdolalidamavandi54@gmail.com

^۲موسسه آموزش عالی غیردولتی انتفاعی لامعی گرگانی، گرگان، ایران

گروه مهندسی کامپیوتر، دانشگاه ملی مهارت، تهران، ایران r.roshany@gmail.com

چکیده

در سالهای اخیر تشخیص حالت چهره از مهمترین چالشهای پردازش تصویر محسوب شده و کاربردهای فراوانی پیدا کرده است. امروزه به دلیل وجود رابطه‌ی به اصطلاح عاطفی میان انسان و کامپیوتر در دنیای مجازی، استفاده از روشهای تشخیص حالت چهره اهمیت زیادی پیدا کرده است. در پژوهش حاضر روشی جدید برای تشخیص حالات چهره پیشنهاد شده است. به منظور اجتناب از محدودیتهای و حفظ سادگی و کارایی LBP سنتی، یک توصیفگر بافت مفهومی و محاسباتی ساده اما کارآمد را پیشنهاد میکنیم که با عنوان الگوی جهتدار سه گانه محلی (LDTP) نامیده میشود. مزیت اصلی توصیفگر پیشنهادی نسبت به آنهایی که در حال حاضر موجود است، این است که هر دو مفهوم اپراتورهای LTP و LDP را در یک طرح کدگذاری فشرده مشابه، که اطلاعات دقیقتر و تفکیکپذیرتری را فراهم میکند، ترکیب میکند. همچنین نشان داده شده است که استفاده از ترکیب ویژگیهای بهجای تنها یک ویژگی باعث تشخیص بهتر و مقاومتر حالت چهره میشود. بنابراین ویژگیهای LDTP و XCSLBP با هم ترکیب شده تا بردار ویژگی نهایی حاصل شود. سپس برای سرعت بیشتر، بردار ویژگی نهایی با استفاده از الگوریتم PCA، کاهش بعد داده میشود. حال این بردار ویژگی به SVM که قبلاً توسط دادههای آموزشی، آموزش داده شده است، داده میشود تا در نهایت یکی از ۷ حالت چهره مشخص شود. روش پیشنهادی از جنبههای گوناگون مورد بررسی قرار گرفت و با سایر روشهای موجود، مقایسه شده است. دقت روش پیشنهادی برای حالت چهره ۷ کلاسه حدود 96.22 درصد است که نسبت به سایر روشها از دقت بالاتری برخوردار است.

واژه‌های کلیدی: تشخیص حالت چهره، الگوی باینری محلی، الگوی جهتدار محلی، الگوی سه گانه محلی، ماشین بردار پشتیبان.

1. مقدمه

امروزه سیستمهای مبتنی بر تعامل کامپیوتر با انسان در بسیاری از کاربردهای زمان واقعی، برای رهگیری فوری و دقیق فعالیتهای انسانی از ویدئوها استفاده میشود. یکی از این زمینهها شناسایی و ردیابی تشخیص حالت و احساسات انسان از دنبالههای ویدئویی و با هدفهای مختلفی مانند تشخیص خستگی فیزیکی است [1]. جامعه بینایی کامپیوتری بیش از ده سال توجه به شناخت احساسات صورت را به خود جلب کرده است. کارهای بزرگی انجام شده و در حال انجام است تا با کمک بهبود تمام جنبههای تعامل بین کامپیوترها و انسانها، زندگی آسانی را برای افراد معلول (مثل کور، لال) و افراد سالمند فراهم کند. در حوزه HCI، بر استفاده عملی برای تشخیص خودکار یک حالت خاص صورت توسط یک لیست از پیش تعریف شده تاکید شده است. حالت چهره در روانشناسی بطور گسترده مورد مطالعه قرار گرفته است. نشان داده شده است که 7 درصد ارتباطات انسان توسط پیرازبان، 38 درصد توسط زبان و 55 درصد از طریق حالت چهره، صورت میپذیرد. بدین ترتیب، شناخت از طریق حالت چهره یک راه مهم در تعامل بین انسان و کامپیوتر است. مدل ریاضی حالت چهره با استفاده از استخراج ویژگیهای چهره، صورت میپذیرد. تحقیقات بسیاری برای تجزیه و تحلیل حالت چهره پیشنهاد شدهاند. این نتیجه دریافت شده است که منابع حالات چهره عبارتند از حالتی ذهنی (مانند احساسات پایدار، عقاید و تفکرات)، ارتباطات کلامی که از طریق تصویرگران، پاسخهای شنوندگان و تنظیم کنندهها متصل میشوند، ارتباطات غیر کلامی که نه تنها از طریق نشانها و شکلکهای اجتماعی بلکه از طریق احساسات ناپایدار مرتبط میشوند، فعالیتهای فیزیولوژیکی مانند یا ابزارهای دیگر مانند درد و خستگی [2].

تشخیص/آنالیز حالتی چهره یک حوزه تحقیقاتی بسیار فعال و باز در حوزه یادگیری ماشین است. که کاربردهای بسیاری در حوزههای مختلف موجود دارند. سیستمهای حالت چهره تمام خودکار و زمان واقعی در تفسیر حرکات غیر کلامی صورت استفاده شده در کاربردهای مختلفی مانند تحقیقات رفتاری، تعامل کامپیوتر و انسان، تماس تصویری و سیستمهای بینایی، سودمند هستند. اما امروزه چالش اصلی ایجاد سیستمی برای انجام تعاملات طبیعی بین انسان و یک ماشین است. بهعنوان مثال در HCI اگر کامپیوتر قادر به حس حالات کاربران باشد و بتواند با کمک حالات آنها اهدافشان را درک کند، آنگاه برای سیستم کمک به کاربران از طریق انتقال پاسخها و پیشنهادات مطابق با شرایط خاص در یک رفتار طبیعی، آسان خواهد بود که در واقع میتواند در کامپیوترهای تعاملی، سیستمهای سرگرم کننده برای کاربران، حسگرها، رباتها و غیره استفاده شود [3]. از آنجایی که کامپیوترها شروع به تبدیل شدن به بخشی از فضاهای زندگی و فضاهای کاری کردند و شروع به تعامل بیشتر و بیشتر با انسانها کردند، سیستمها باید در درک حالتی عاطفی انسانها دقیقتر باشند. داشتن یک سیستم تشخیص حالت چهره هوشمند، باعث میشود تا ایجاد یک رابط بصری خوب برای تعاملات بین انسان و کامپیوتر، آسانتر و مفید باشد. ارتباط بین انسانها موثر است زیرا آنها میتوانند پاسخهایی مطابق با حالتی دیگر اشخاص را دریافت کنند، بنابراین برای تعامل موثر با انسانها، کامپیوترها نیز باید به این توانایی دست یابند. رابطهای کامپیوتر و انسان و رباتها تنها کاربردهای سیستمهای تشخیص حالت چهره نیستند، بلکه کاربردهای دیگری در چندین حوزه مجزا از قبیل بازیهای ویدئویی، انیمیشن، روانپزشکی، نرمافزار آموزشی، موسیقی، علم پزشکی، دادرسی قانونی، مصاحبه جنایی و غیره دارند. همانطور که سیستمهای تشخیص حالت چهره مقاوم و زمان واقعی میشود، بسیاری استفادهها و کاربردهای خلاقانه دیگر هنوز دیده نشده است.

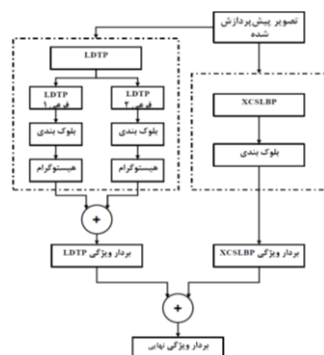
انگیزه در پشت هر حوزه تحقیقاتی، توانایی آن برای حل یک مسئله و کاربردهای آن است. هدف اصلی تمامی الگوریتمهای پردازش تصاویر و بینایی کامپیوتری این است که اطلاعات بصری مفیدی ایجاد کنند. از این رو در حوزه بینایی کامپیوتر، تشخیص حالت چهره با هدف مشابهی آغاز شده است. اهمیت و الزامات این حوزه تحقیقاتی بهدلیل کاربردهای آن در HCI (تعامل انسان و کامپیوتر) که در آن یک دید بصری از انسان، احساسات بینایی و لامسه و صدا در یک زمان مورد استفاده قرار میگیرند، رو به افزایش است. علاوه بر این روانشناسی اجتماعی بیان میکند که حالتی چهره در متناسب ساختن مکالمه پشتیبانی میشوند. حالتی چهره در صورت همراه بودن با اطلاعات بصری، 55٪ از نتیجه پیام گفتاری را ارائه میدهند. سهم کلمات گفتاری 7٪ است و تن صدا 38٪ سهم دارد [4] بنابراین حالت چهره مهمترین در میان تمامی حالتی های HCI است.

این امر باعث میشود که تحقیق در زمینه تشخیص حالت چهره ضروری باشد. علاوه بر این دیدگاه حوزه ربات، انیمیشن چهره و واقعیت مجازی نیز نیاز به آنالیز حالت چهره دارند. این سیستمها با هدف حذف خطای انسانی در کنترل خودرو طراحی شدهاند. شناخت احساسات پایه و شناخت واحدهای عملیاتی دو طبقهای است که تحقیقات براساس آنها انجام شده است. هدف از تشخیص احساس پایه، ردیابی حالتهای ذهنی است درحالیکه تشخیص واحدهای عملیاتی روی ارائه تغییر شکل عضلانی ماهیچههای صورت تمرکز دارد. ارتباطات اجتماعی، نیتها و حالتهای احساسی فرد بهعنوان حالتهای چهره در نظر گرفته میشوند. در حال حاضر محققان در حال کار بر روی تشخیص تصاویر چهرهای هستند که در آن تغییرات در حالات چالشهای بزرگی را برای این الگوریتمها به وجود آورده است. تشخیص احساسات کاربر برای درک یک کاربرد مهم سیستم تعامل کامپیوتر و انسان است که می تواند بهطور موثر با شناخت حالات چهره انجام شود. تشخیص حالت چهره در یک محیط بدون محدودیت برای ایجاد سیستم های HCI زمان واقعی بسیار مهم است.

شارما و همکاران [5] از روش ترکیبی بردارهای ویژگی CLBP و فیلترهای گابور برای تشخیص حالت چهره استفاده نمودند. در این مقاله احساسات با محاسبه بردار ویژگی تصویر چهره ورودی از طریق ترکیب CLBP و فیلترهای گابور و ارسال این بردارهای ویژگی به طبقهبندهای SVM آشنایی هار و Adaboost برای تشخیص حالت چهره ارائه شده است. سرعت اجرای الگوریتم Adaboost بالا است. این تحقیق دقت و زمان اجرای تشخیص حالت چهره را افزایش میدهد. به طور کلی تشخیص حالت چهره یکی از حوزههای تحقیقاتی داغ در ارتباط با هوش مصنوعی است و نقش مهمی در زمینه تعامل انسان و کامپیوتر ایفا میکند. تشخیص، بازشناسی و ردیابی چهره زمینههای تحقیقاتی مرتبط با تشخیص حالت چهره هستند و بسیار بهبود یافته و پیشرفته هستند که در نهایت نیازمند پیشرفتهای و دقت بیشتری در زمینه تشخیص حالت چهره را حاصل میشوند. به نظر میرسد که در آینده تشخیص حالت چهره در کاربردهای مختلفی استفاده شود که میتواند بهعنوان انگیزهای برای تحقیق در زمینههای مختلف از جمله درک تصویر، مطالعات روانشناختی، درجهبندی عصب صورت در پزشکی، انیمیشن صورت، رباتیک و بازیهای ویدئویی، برای ما باشد. تشخیص حالت چهره میتواند در طراحی دستگاههای تعاملی جدید بسیار مفید و مهم باشد که ممکن است منجر به توسعه ابعاد جدیدی برای ارتباط برقرار کردن با سیستم های کامپیوتری شود. بنابراین هدف از پژوهش حاضر تشخیص حالت چهره با استفاده از توصیفگر محلی جهتدار است.

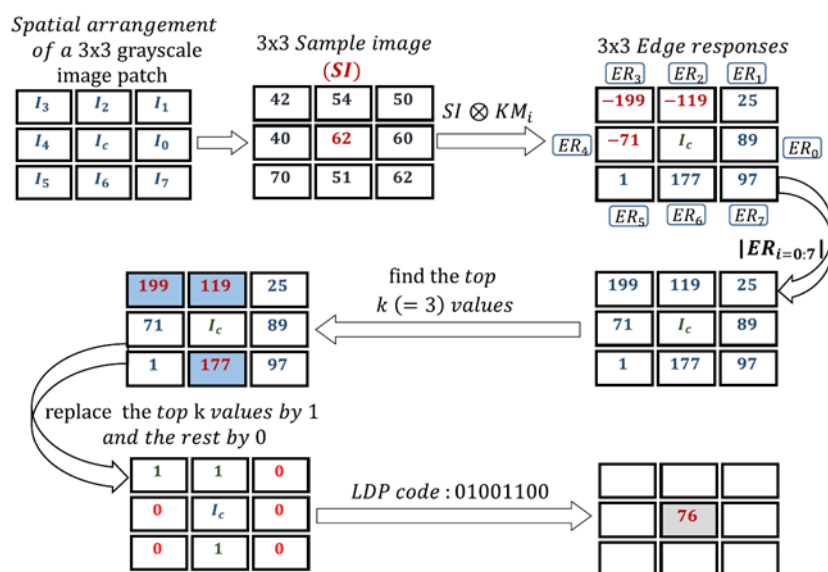
2. روش پژوهش

روش پیشنهادی پژوهش در سه گام اصلی (1) پیش پردازش، (2) استخراج ویژگی و (3) طبقه بندی انجام میشود. در بخش پیش پردازش پس از ورود تصاویر، ابتدا ناحیه صورت از پسزمینه جدا میشود و سپس تصویر صورت تغییر رزولوشن 32×32 ، 64×64 ، 96×96 و 128×128 داده میشود تا تمامی تصاویر دارای اندازه یکسان باشند. در بخش استخراج ویژگی ها، از ترکیب دو ویژگی LDTP استفاده کردیم. بردار ویژگی نهایی جهت طبقه بندی بهتر و زمان پردازشی کمتر کاهش بعد داده میشود و سپس وارد طبقهبندها میشود. بردار ویژگی از هر تصویر چهره توسط رویکرد ترکیبی، استخراج میشود که در شکل (1) آمده است. جهت مرحله طبقه بندی نیز از طبقهبندها SVM چندکلاسه استفاده شده است تا بتوانیم حالت چهره که هفت کلاسه است را طبقه بندی کنیم. خروجی این مرحله تشخیص یکی از هفت حالت عمومی چهره است.



شکل ۱. چارچوب روش ترکیبی استخراج ویژگی

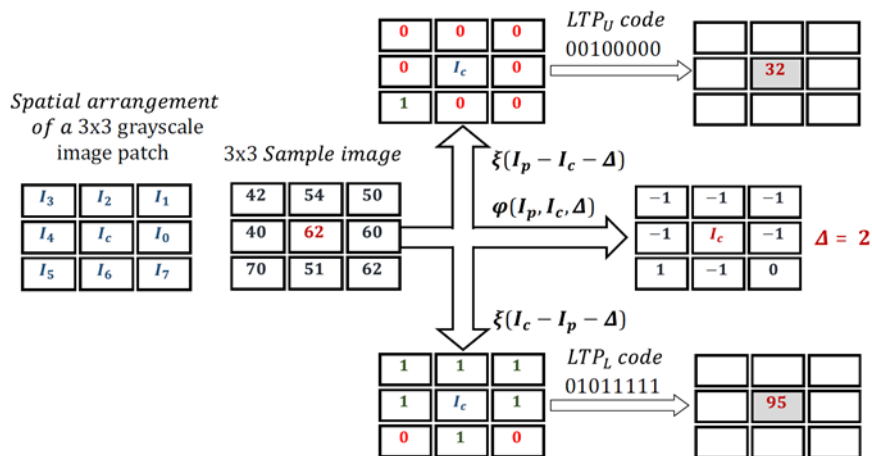
الگوی جهتدار محلی (LDP) یک الگوی سطح خاکستری است که ساختار مکانی بافت محلی تصویر را مشخص میکند. یک عملگر LDP مقادیر پاسخ لبه در تمامی هشت جهت در هر مکان پیکسل را محاسبه میکنند و یک کد از دامنه توان نسبی ایجاد میکنند. از آنجاییکه پاسخهای لبه نسبت به مقادیر شدت روشنایی، حساسیت بیشتری نسبت به نویز و تغییرات روشنایی دارند، ویژگی LDP نتیجه اولویتهای محلی شامل انواع مختلفی از منحنیها، گوشهها و اتصالات را پایدارتر توصیف میکند و اطلاعات بیشتری را حفظ میکند. با توجه به پیکسل مرکزی در تصویر، مقادیر پاسخ لبه هشت جهت $\{M_i\}, i=0,1,2,3,4,5,6,7$ توسط KM_i ماسک Kirsch در هشت جهت مختلف محاسبه میشود. این هشت دامنه پاسخ لبه برای تولید هشت بیت عدد باینری استفاده میشوند که میتواند الگوی لبه محلی یک پیکسل خاص را توصیف کند. با در نظر گرفتن یک همسایگی مربعی 3×3 و هشت ماسک، هشت پاسخ $[ER]_0, [ER]_1, [ER]_2, [ER]_3, [ER]_4, [ER]_5, [ER]_6, [ER]_7$ از کانولوشن الگوی تصویر 3×3 در هشت ماسک بدست میآید. مقادیر پاسخ در تمامی جهات اهمیت برابری ندارند. حضور گوشه یا لبه در برخی جهات خاص مقادیر پاسخ بالا را نشان میدهد. بنابراین، علاقه‌مند هستیم که مهمترین جهات را برای تولید LDP بدانیم. k مقدار بالاتر این هشت پاسخ لبه با 1 تنظیم میشوند و پاسخهای باقیمانده $(8-k)$ با 0 تنظیم میشوند. روند کلی در شکل (2) برای $k=3$ نشان داده شده است. کد LDP با رابطه (1) تعریف میشود که در آن k $[ER]_k$ امین پاسخ با بیشترین مقدار جهت است.



شکل ۲. روند محاسبه کد LDP برای $k=3$ با ماسک Kirsch 4.

الگوی سه گانه محلی (LTP) نوع بسط یافته LBP است که در آن با استفاده از حد آستانه Δ دو مقدار کانولوشن کدهای (LBP و 0) به سه مقدار کدهای سه گانه (0، -1 و 1) بسط یافته است. مقادیر سطح خاکستری در ناحیه‌های با عرض $\Delta \pm$ در اطراف سطح خاکستری I_c از پیکسل مرکزی به صفر مقداردهی میشود، آنهایی که زیر $I_c - \Delta$ هستن با -1 مقداردهی شده و آنهایی که بالای $I_c + \Delta$ هستند با +1 مقداردهی میشوند. تابع علامت $s(\cdot)$ در رابطه (2) با تابع 3 مقداردهی $\varphi(\cdot, \cdot)$ جایگزین میشود و کد LBP باینری با کد LTP سه گانه جایگزین میشود که در شکل (3) نشان داده شده است.

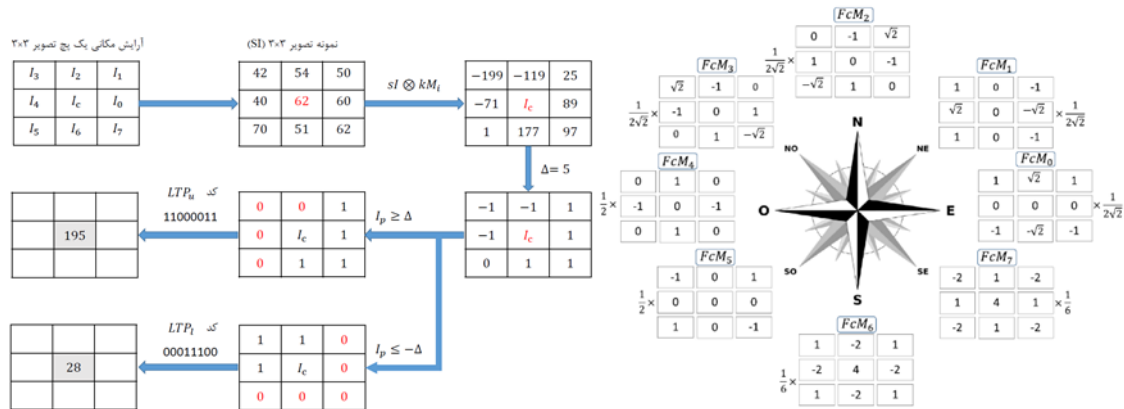
$$(2)\varphi(I_p, I_c, \Delta) = \begin{cases} +1 & I_p \geq I_c + \Delta \\ 0 & |I_p - I_c| < \Delta \\ -1 & I_p \leq I_c - \Delta \end{cases}$$



شکل 3. فرایند محاسبه LTP برای یک همسایگی مربعی 3×3 [4].

همچنین توصیف‌های سه سطحی LTP و ویژگی‌های جهتدار LDP ترکیب شده است تا یک توصیفگر ویژگی محلی جدید با عنوان الگوی جهتدار سه‌گانه محلی (LDTP) برای طبقه‌بندی بافت ایجاد شود. LDTP یک چارچوب است که شامل کدگذاری اطلاعات کنتراست و ویژگی‌های الگوی جهتدار مبتنی بر تغییرات مشتقی محلی در یک روش فشرده است. در نهایت، الگوی LDTP حاصل به دو قسمت متمایز الگوی جهتدار سه‌گانه محلی بالایی ($LDTP_u$) و الگوی جهتدار سه‌گانه محلی پایینی ($LDTP_l$) تقسیم میشود و بردار توصیف ویژگی نهایی با ترکیب خطی هر دو هیستوگرام $LDTP_u$ و $LDTP_l$ به دست می‌آید. برخلاف اپراتورهای LDP و LDN که فقط از یک نوع ماسک استفاده میکنند، توصیفگر پیشنهادی روی ماسکهای مختلف بررسی شده است. بنابراین اپراتور LDTP پیشنهادی بافت تصویر را با محاسبه پاسخهای لبه با استفاده از هشت جهت کدگذاری میکند.

در پژوهش حاضر سه ماسک مختلف کریش، فری چن و رابینسون مورد بررسی قرار گرفته است که در شکل 4 آمده است.



ب. نحوه محاسبه LDTP پیشنهادی
[با ماسک Kirsch].

الف. ماسک Frei-Chen برای محاسبه
7 جهت‌های LDTP

ج. اسکهای Robinson برای محاسبه جهت [LDTP 8].

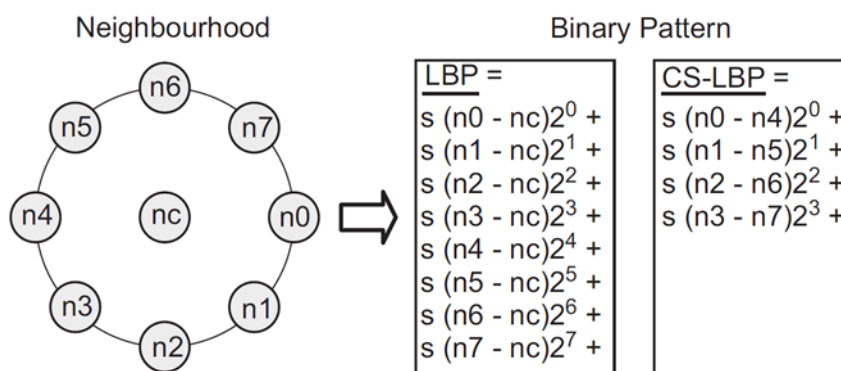
$$\begin{array}{cccc}
 \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} & \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{bmatrix} & \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} & \begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & -2 \end{bmatrix} \\
 M_0 & M_1 & M_2 & M_3 \\
 \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} & \begin{bmatrix} 0 & -1 & -2 \\ 1 & 0 & -1 \\ 2 & 1 & 0 \end{bmatrix} & \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} & \begin{bmatrix} -2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix} \\
 M_4 & M_5 & M_6 & M_7
 \end{array}$$

شکل 4. ماسک‌های استفاده شده در پژوهش

ایده‌ی اصلی [CSLBP 9] نیز مقایسه‌ی سطوح خاکستری جفت پیکسل‌های متقارن نسبت به مرکز بهجای مقایسه‌ی پیکسل مرکزی با پیکسل‌های همسایه در LBP است با فرض P عدد زوج از پیکسل‌های همسایه، عملگر CS-LBP به‌صورت رابطه (3) تعریف می‌شود.

که در آن g_i و $g_{i+(P/1)-2}$ (مقادیر سطح خاکستری جفت پیکسل‌های متقارن نسبت به مرکز هستند و S تابع آستانه‌ی تعریف شده به‌صورت رابطه (4) است.

که در آن T حد آستانه‌ی تعریف شده توسط کاربر است. از آنجاییکه سطوح خاکستری در بازه‌ی [0,1] نرمال می‌شوند، مؤلفین مقدار کوچکی را پیشنهاد می‌کنند. با استفاده از این روش طول بردار هیستوگرام حاصل از CS-LBP به $2^{(P/2)}$ (کاهش می‌یابد. مشاهده می‌شود که برای 8 همسایه، 256LBP الگوی باینری مختلف ایجاد می‌کند درحالیکه CS-LBP تنها 16 مقدار ایجاد می‌کند که به کاهش محاسبات منجر می‌شود.



شکل ۵. ویژگیهای LBP و CS-LBP برای ۸ همسایه [12].

همچنین LBP جدید XCS-LBP به صورت رابطه (۵) تعریف میشود .

$$XCS-LBP(i, c) = \sum_{j=0}^{P-1} (s(g_j(i, c) - g_{j+1}(i, c)) \cdot 2^j) \quad (5)$$

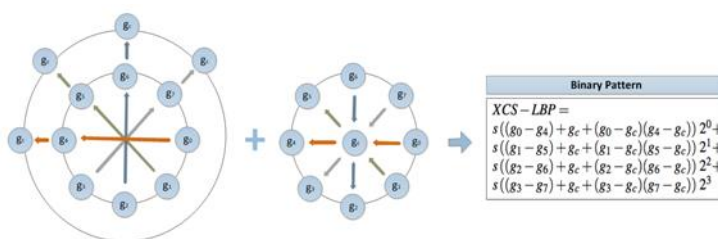
که در آن تابع آستانه‌ی s به صورت رابطه (۶) تعریف می شود

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

و همچنین $g_j(i, c)$ به صورت (۷) تعریف میشوند.

$$g_j(i, c) = \begin{cases} g_c & \text{if } j = 0 \\ g_{i+P/2} & \text{if } j = 1 \\ g_{i+P/4} & \text{if } j = 2 \\ \vdots \\ g_{i+P/2^{j-1}} & \text{if } j = P-1 \end{cases} \quad (7)$$

محاسبه XCS-LBP در شکل (۶) نشان داده شده است.



شکل ۶. توصیفگر XCS-LBP [10]

همچنین در این پژوهش ویژگیهای استخراج شده توسط PCA کاهش بعد داده شدند تا وارد مرحله طبقه بندی شوند. طبقه‌بند SVM شامل دو فاز آموزش و آزمون است. در فاز آموزش، طبقه‌بند توسط نمونه‌های موجود در پایگاه داده آموزش داده میشود و در فاز آزمون، حالت چهره تصویر ورودی توسط طبقه‌بند آموزش داده شده، به یکی از هفت حالت متداول طبقه‌بندی میشود. دو روش معمول برای تشخیص حالت چهره با استفاده از SVM وجود دارد: روش یک در برابر یک و روش یک در برابر همه. روش یک در برابر یک، طبقه‌بندی بین هر جفت از کلاسها است. روش یک در برابر همه، طبقه‌بندی بین هر کلاس و تمامی کلاسهای باقیمانده است. در آزمایشات ما روش یک در برابر همه برای طبقه‌بندی مورد استفاده قرار گرفته است. در روش یک در برابر همه، عدد SVM آموزش داده میشوند. هر SVM یک کلاس را از بقیه کلاسهای باقیمانده جدا میکند.

ایجاد یک طبقه بند n کلاسه با استفاده از روشهای تفکیک کننده دو کلاسه اغلب توسط پروسه زیر انجام میشود ایجاد n تابع تصمیم دو کلاسه $1d_k(x)$, $k=1, \dots, n$ ، که نمونه‌های کلاس k را از نمونه‌های آموزشی تمامی کلاسهای دیگر جدا میکند.

$$1d_k(x) = \begin{cases} 1 & \text{if } x \text{ belongs to class } k \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

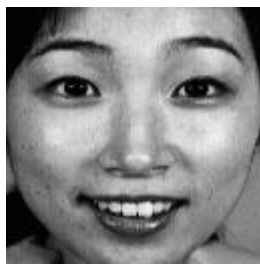
SVM با SVM 1 برچسبگذاری می شود. به نوبت میتوان SVM را برای هر حالت چهره بدست آورد و با SVM 1, SVMn, ... نامگذاری نمود.

n عدد SVM نمونه ها را به n کلاس تقسیم میکنند. زمانی که یک نمونه آزمون به نوبت به هر SVM وارد میشود، چندین مورد وجود دارد (1) اگر نمونه در یک زمان یکسان، توسط SVMi مثبت و توسط دیگر SVM ها منفی تشخیص داده شود، نمونه به عنوان کلاس i طبقه بندی میشود (2) اگر نمونه به طور همزمان توسط چند SVM منفی و توسط چند SVM دیگر مثبت تشخیص داده شود، آنگاه طبقه بندی دچار خطا میشود (3) اگر نمونه به طور همزمان توسط تمام SVM ها منفی در نظر گرفته شود، آنگاه تصمیم گرفته میشود که نمونه به هیچ کلاس حالت چهره های متعلق نیست.

نرخ تشخیص در محیط OS Windows 8.1 با پردازنده Intel Core i7 @ 2.3 GHz و 8 GB RAM محاسبه شده است. همچنین برای تمامی محاسبات مجموعه آموزشی با استفاده از روش k-fold به 10 بخش تقسیم شده و هر بار یک بخش به عنوان داده آزمون و بقیه به عنوان داده آموزش به طبقه بندی داده شده اند، این فرایند 10 بار تکرار شده و نرخ تشخیص نهایی از میانگین 100 نتیجه حاصل بدست آمده است. همچنین در این پایان نامه از پایگاه داده JAFFE استفاده شده است که پایگاه داده های شامل 213 تصویر از زنان ژاپنی است، تصاویر برای 10 فرد مختلف در 7 حالت اصلی چهره موجود است. تصاویر در اندازه 256*256 پیکسل و به فرمت سیاه-سفید ارائه شده اند. واحدهای حرکتی صورت در این پایگاه داده علامتگذاری نشده و تنها حالات چهره نامگذاری شده اند.

3. یافته ها

برای استخراج ویژگی در این پژوهش از ترکیب دو ویژگی XCSLBP و LDTP استفاده شده است. نتیجه حاصل از روش های استخراج ویژگی LDTP با ماسک های مختلف برای تصویر با رزولوشن 128*128 پیکسل در شکل (7) نشان داده شده است. همچنین تصویر خروجی حاصل از استخراج ویژگی XCSLBP با شعاع های مختلف در شکل (8) نشان داده شده است.



(الف)



(ت)



(پ)



(ب)



(چ)

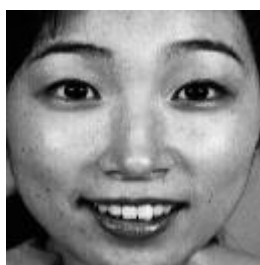


(ج)



(ث)

با $u = 1$ LDTP]] با ماسکهای مختلف. (الف) تصویر اصلی، (ب) LDTP]] شکل 7. نتیجه حاصل از روشهای استخراج ویژگی ، (ث) Frei-Chen با ماسک $u = 1$ LDTP]] ، (ت) Kirsch با ماسک $u = 1$ LDTP]] ، (پ) Kirsch ماسک با ماسک $u = 1$ LDTP]] و (چ) Robinson با ماسک $u = 1$ LDTP]] ، (ج) Frei-Chen با ماسک $u = 1$ LDTP]] Robinson.



(الف)



(ب)



(پ)



(ت)

شکل ۸. تصویر خروجی حاصل از استخراج ویژگی XCSLBP با شعاعهای مختلف. (الف) تصویر اصلی (ب) شعاع ۱، (پ) شعاع ۲ و (ت) شعاع ۳.

جدول ۱. تاثیر تغییر رزولوشن با اندازه بلوکهای مختلف بر ابعاد بردار ویژگی

XCSLBP+LD TP	روشها					تعداد بلوک	اندازه بلوک	اندازه تصویر
	XCSLBP	LDTP	LDP	LTP	LBP			
33792	1024	32768	16384	32768	16384	64	4×4	32×32
19008	576	18432	9216	18432	9216	36	6×6	
135168	4096	131072	65536	131072	65536	256	4×4	64×64
63888	1936	61952	30976	61952	30976	121	6×6	
33792	1024	32768	16384	32768	16384	64	8×8	
135168	4096	131072	65536	131072	65536	256	6×6	96×96
76032	2304	73728	36864	73728	36864	144	8×8	
52800	1600	51200	25600	51200	25600	100	10×10	
33792	1024	32768	16384	32768	16384	64	12×12	
135168	4096	131072	65536	131072	65536	256	8×8	128×128

89232	2704	86528	43264	86528	43264	169	10×10
63888	1936	61952	30976	61952	30976	121	12×12
33792	1024	32768	16384	32768	16384	64	16×16

در جدول ۱ تاثیر تغییر رزولوشن با اندازه بلوکهای مختلف بر ابعاد بردار ویژگی نشان داده شده است. تعداد میلههای هیستوگرام هر بلوک LBP و LDP برابر ۲۵۶ میله، تعداد میلههای هیستوگرام هر بلوک LTP و LDTP برابر ۵۱۲ میله و تعداد میلههای هیستوگرام هر بلوک XCSLBP برابر ۱۶ میله است. بنابراین برای یک تصویر با رزولوشن و اندازه بلوک یکسان، ابعاد بردار ویژگی LTP و LDTP دو برابر LBP و LDP است و همچنین ابعاد بردار ویژگی LBP و LDTP برابر XCSLBP است. بنابراین استفاده از روش XCSLBP به خودی خود باعث کاهش زمان اجرا میشود و ترکیب آن با دیگر روشها علاوه بر بهبود دقت تشخیص، تاثیر زیادی در زمان اجرا نمیگذارد. در جدول ۲ دقت تشخیص محاسبه شده در رزولوشنها با اندازه بلوکهای مختلف برای سه ماسک استفاده شده در پژوهش آمده است. در رزولوشنهای مختلف، دقت تشخیص متفاوت است، بیشترین میزان دقت تشخیص در رزولوشن ۹۶×۹۶ اتفاق افتاده است که گزینه مناسبی برای انتخاب است زیرا مصالحه بین دقت تشخیص و ابعاد بردار ویژگی (زمان اجرا) در آن رعایت شده است. میانگین دقت تشخیص برای ماسک Kirsh در تمامی رزولوشنها و اندازه بلوکها برابر ۹۱.۵۷ درصد است که نسبت به دیگر ماسکها بیشتر است. همچنین بیشترین دقت تشخیص در ۳ روش با میانگین ۹۲.۹۰ درصد در اندازه تصویر ۹۶×۹۶ با اندازه بلوک ۸×۸ بالاترین دقت است.

جدول ۲. تاثیر تغییر ماسک LDTP بر دقت تشخیص

با ماسکهای مختلف LDTP					
اندازه تصویر	اندازه بلوک	Kirsh-LDTP	Frei-Chen-LDTP	Robinson-LDTP	میانگین
32×32	4×4	89.97	91.44	90.24	90.55
	6×6	83.61	90.59	85.63	86.61
64×64	4×4	92.20	90.04	92.39	91.54
	6×6	93.34	92.11	93.07	92.84
	8×8	90.05	92.41	89.41	90.62
96×96	6×6	93.02	88.79	92.16	91.32
	8×8	94.28	91.25	93.17	92.90
	10×10	91.55	89.07	90.66	90.43
	12×12	90.69	89.44	90.03	90.05
128×128	8×8	92.94	86.59	92.76	90.76

90.97	93.04	86.88	92.98	10×10	
91.68	93.47	88.37	93.19	12×12	
89.24	91.10	84.02	92.60	16×16	
90.73	91.32	89.31	91.57	-	میانگین

در جدول 3 تاثیر تغییر شعاع XCSLBP با شعاع 1 (همسایگی 3×3)، شعاع 2 (همسایگی 5×5) و شعاع 3 (همسایگی 7×7) بر دقت تشخیص حالت چهره مورد بررسی قرار گرفته است. مشاهده میشود که بیشترین دقت تشخیص در XCSLBP با شعاع 3 با میانگین 90.84 درصد اتفاق میافتد و همچنین در اینجا نیز بیشترین دقت تشخیص در اندازه تصویر 96×96 با اندازه بلوک 8×8 اتفاق میافتد. همچنین با افزایش شعاع دقت تشخیص بهتر شده است که این امر به دلیل در نظر گرفتن روابط بین پیکسلها در همسایگیهای بزرگتر است. یافته ها نشان می دهد که همواره زیاد شدن اندازه تصویر به دقت تشخیص بهتر نمیانجامد بهطوریکه میانگین دقت تشخیص در اندازه تصویر 96×96 از بقیه رزولوشنها بالاتر است. همچنین با توجه به اندازه تصویر دقت تشخیص در رزولوشنهای مختلف متفاوت است. بنابراین انتخاب اندازه بلوک به اندازه تصویر مرتبط است. بهطوریکه بهترین دقت تشخیص در اندازه تصویر 32×32 در اندازه بلوک 4×4، بهترین دقت تشخیص در اندازه تصویر 64×64 در اندازه بلوک 6×6، بهترین دقت تشخیص در اندازه تصویر 96×96 در اندازه بلوک 8×8 و بهترین دقت تشخیص در اندازه تصویر 128×128 در اندازه بلوک 10×10 است.

جدول 3. تاثیر تغییر شعاع XCSLBP بر دقت تشخیص

اندازه تصویر	اندازه بلوک	با شعاعهای مختلف XCSLBP			میانگین
		r=3	r=2	r=1	
32×32	4×4	86.63	88.35	87.46	87.48
	6×6	83.83	82.22	79.82	81.96
64×64	4×4	91.73	92.11	90.12	91.32
	6×6	91.19	93.27	91.73	92.06
	8×8	90.71	90.52	89.96	90.40
96×96	6×6	92.85	91.13	91.86	91.95
	8×8	94.11	92.44	92.07	92.87
	10×10	91.83	91.43	91.06	91.44
	12×12	91.60	89.81	87.34	89.58
128×128	8×8	91.23	91.73	91.50	91.49

92.34	92.58	93.14	91.29	10×10	
91.82	91.98	92.44	91.05	12×12	
89.13	90.63	89.04	87.73	16×16	
90.30	90.84	90.59	89.46	-	میانگین

از آنجایی که بهترین دقتها در اندازه تصویر 96×96 و در اندازه بلوک 8×8 بدست آمده است، بنابراین روی این پارامترها آزمایشات بعدی انجام شده است. در این قسمت تاثیر کاهش بعد PCA با کاهش بعدهای 50، 100، 150 و 200 روی انواع مختلفی از روشها بررسی شده است (جدول 4). همانطور که مشاهده میشود دقت تشخیص با کاهش بعد PCA نه تنها دقت تشخیص را کاهش نمیدهد بلکه در اکثر موارد باعث افزایش دقت تشخیص میشود. کاهش بعد PCA باعث کاهش زمان اجرا بهخصوص برای پایگاه دادههای بزرگ میشود. همچنین با افزایش ابعاد از 50 تا 200 دقت افزایش مییابد که بیشترین دقتها در کاهش بعد با ابعاد 200 بدست آمده است. بیشترین دقت برای LDTP با ماسک Kirsch با دقت 94.46 در کاهش بعد 200 حاصل شده است.

جدول ۴. تاثیر کاهش بعد بر دقت تشخیص

PCA اندازه کاهش بعد					روش
200	150	100	50	بدون کاهش بعد	
93.26	89.47	90.94	88.03	93.36	LBP
93.97	91.86	89.18	87.72	93.64	LTP
94.10	93.31	92.48	91.38	93.88	LDP
94.61	94.40	94.08	91.31	94.28	Kirsh-LDTP
85.23	87.67	86.58	84.04	91.25	Frei-Chen-LDTP
93.65	93.59	93.43	90.59	93.17	Robinson-LDTP
93.19	91.09	87.94	86.78	92.07	XCSLBP-r1
94.11	92.26	87.12	85.91	92.44	XCSLBP-r2
94.53	92.31	90.41	86.80	94.11	XCSLBP-r3

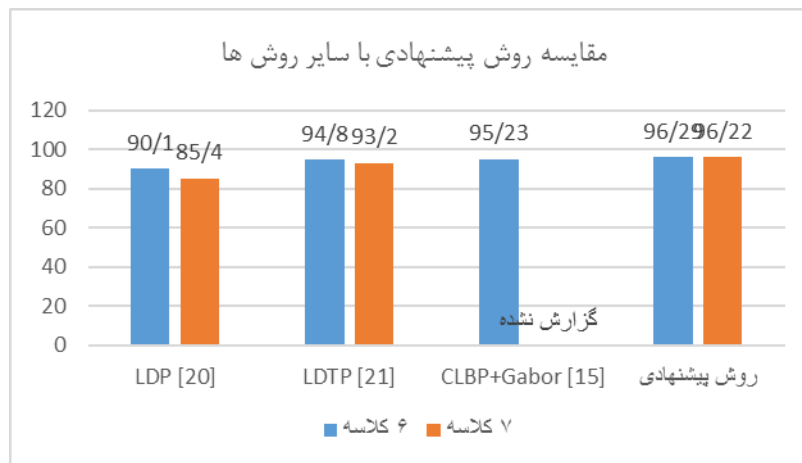
همچنین انواع مختلفی از ترکیبها مورد بررسی قرار گرفته است. برای این منظور اندازه تصویر 96×96 ، اندازه بلوک 8×8 در نظر گرفته شده و کاهش ابعاد 50 تا 200 مورد بحث و بررسی قرار گرفته است (جدول 5). در تمامی روشها کاهش ابعاد باعث افزایش دقت تشخیص شده است که این امر به دلیل انتخاب ویژگیهای بهینه است. بیشترین دقت مربوط به ترکیب دو روش LDTP با ماسک Kirsch و XCSLBP با شعاع 3 بدست آمده است که برابر 96.22 درصد است. همچنین روش پیشنهادی ترکیب LDTP و XCSLBP با سایر روشهای موجود در ادبیات مورد مقایسه قرار گرفته است. همانطور که در جدول 6 مشاهده میشود روش پیشنهادی نسبت به روش LDTP برای 7 کلاسه حدود 11 درصد افزایش داشته و همچنین نسبت به روش LDTP مطرح شده در مرجع [11] در 7 کلاسه حدود 3 درصد افزایش داشته است.

جدول ۵. دقت تشخیص ترکیب روشهای مختلف

PCA اندازه کاهش بعد					روش
200	150	100	50	بدون کاهش بعد	
94.70	94.77	93.65	92.90	93.52	LBP+LDP
94.54	94.46	93.38	93.34	94.48	LTP+LDP
95.68	95.28	95.25	92.29	93.68	Kirsh- LDTP+LBP
96	95.29	95.12	91.53	94.38	Kirsh-LDTP+ XCSLBP-r1
96.05	95.35	95.28	92.62	94.54	Kirsh-LDTP+ XCSLBP-r2
96.22	95.99	95.51	92.46	94.58	Kirsh-LDTP+ XCSLBP-r3
94.16	93.69	93.68	91.10	93.63	Robinson- LDTP+LBP
94.90	94.23	94.47	91.18	93.80	Robinson- LDTP+ XCSLBP-r1
95.06	94.20	94.08	92.02	93.85	Robinson- LDTP+ XCSLBP-r2
95.25	94.60	94.30	92.51	94.16	Robinson- LDTP+ XCSLBP-r3

جدول ۶. مقایسه روش پیشنهادی با سایر روشها

روش	6 کلاسه	7 کلاسه
LDP [11]	90.10	85.40
LDTP [12]	94.80	93.20
CLBP+Gabor [7]	95.23	-
روش پیشنهادی	96.29	96.22



شکل ۹. نمودار مقایسه دقت تشخیص روش پیشنهادی با سایر روشها

۴. بحث و نتیجه گیری

در پژوهش حاضر روشی جدید برای تشخیص حالت چهره مبتنی بر استفاده از خصوصیات سه ویژگی LDP، LTP و XCSLBP مطرح شده است. روش پیشنهادی در ۳ گام انجام شده است. ابتدا پیشپردازش روی صورت انجام گرفته و ناحیه صورت از باقی تصویر جدا میشود، سپس تصویر به اندازه رزولوشنهای مختلف تغییر اندازه داده میشود. در این مقاله اندازههای مختلف از 32×32 تا 128×128 مورد بررسی قرار گرفت که مشخص گردید در رزولوشنهای پایین دقت تشخیص کم است و با افزایش رزولوشن دقت تشخیص بهبود مییابد اما این افزایش از حدی که بیشتر میشود دوباره دقت تشخیص کم میشود. بنابراین اکثرا بهترین دقت تشخیص در اندازه تصویر 96×96 پیکسل حاصل شد که اندازه تصویر مناسبی است زیرا در آن مصالحه بین دقت تشخیص و زمان اجرا در نظر گرفته میشود. در گام استخراج ویژگی ابتدا از خصوصیات دو ویژگی LDP و LTP استفاده شده و روش جدید LDTP برای استخراج ویژگی ساخته میشود. برای این منظور از سه ماسک مختلف استفاده شد که بهترین نتیجه برای ماسک معمول Kirsch حاصل شد. همچنین اندازه بلوکهای مختلف با توجه به رزولوشنهای تصویر در نظر گرفته شدند که مشخص گردید با توجه به رزولوشن اندازه بلوک باید انتخاب گردد. هرچه اندازه تصویر کوچکتر باشد، اندازه بلوک نیز باید کوچکتر باشد. همچنین از استخراج ویژگی XCSLBP استفاده شد که تاثیر اندازه شعاع همسایگی برای

انتخاب ۸ همسایه پیکسل مرکزی مورد مقایسه قرار گرفت. مشخص شد با افزایش شعاع، دقت تشخیص افزایش میابد زیرا ابعاد گستردهتری از همسایگان پیکسل و ارتباط بین پیکسلها را در نظر میگیرد. که بهترین دقت تشخیص در شعاع همسایگی ۳ بدست آمد. در نهایت دو ویژگی XCSLBP و LDTP با هم ترکیب شدند. در گام طبقه‌بندی از طبقه‌بند SVM یک در برابر همه استفاده شده است که دلیل این انتخاب هم افزایش دقت تشخیص و هم پیچیدگی محاسباتی کمتر آن نسبت به روش یک در برابر یک است. که این امر در بسیاری از روشهای موجود در ادبیات آمده است .

جهت پژوهش های آتی میتوان LDTP را همانند XCSLBP برای شعاعهای بزرگتر بسط داد تا ساختارهای محلی وسیع آشکار شوند و نتیجه آن بر دقت تشخیص را بررسی نمود. در گام طبقه‌بندی نیز میتوان طبقه‌بندهای مختلف مانند knn و AdaBoost را استفاده نمود و نتیجه آنها بر دقت تشخیص و طبقه‌بندی را با SVM مقایسه کرد.

۵. منابع و مراجع

- [1] H. Salih and L. Kulkarni, "Study of video based facial expression and emotions recognition methods," International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), pp. 692-696, 2017.
- [2] N. U. Khan, "A comparative analysis of facial expression recognition techniques," 3rd IEEE International Advance Computing Conference (IACC), pp. 1262-1268, 2013.
- [3] R. Jameel, A. Singhal and A. Bansal, "A comprehensive study on Facial Expressions Recognition Techniques," 6th International Conference - Cloud System and Big Data Engineering (Confluence), pp. 478-483, 2016.
- [4] A. Rathi and B. N. Shah, "Facial Expression Recognition A Survey", International Research Journal of Engineering and Technology (IRJET), vol. 3, pp. 540-545, 2016.
- [5] R. L. Kristensen, Z. Tan, Z. Ma and J. Guo, "Binary pattern flavored feature extractors for Facial Expression Recognition: An overview," 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 1131-1137, 2015.
- [6] P. A. Gavade and V. S. Bhat, "A Survey on Facial Expression Recognition using DIP Techniques", International Conference on Signal, Image Processing Communication and Automation – ICSIPCA, pp.51-58, 2017.
- [7] S. A. Khan, A. Hussain and M. Usman, "Facial expression recognition on real world face images using intelligent techniques: A survey", vol. 127, pp. 6195-6203, 2016.
- [8] D. Tyagi, A. Verma and S. Sharma, "An improved method for facial expression recognition using hybrid approach of CLBP and Gabor filter," International Conference on Computing, Communication and Automation (ICCCA), pp. 1019-1024, 2017.
- [9] V. Krishna Gudipati, O. Ray Barman, M. Gaffoor, Harshagandha and A. Abuzneid, "Efficient facial expression recognition using adaboost and haar cascade classifiers," Annual Connecticut Conference on Industrial Electronics, Technology & Automation (CT-IETA), pp. 1-4, 2016.
- [10] I. El khadiri, A. Chahi, Y. El merabet, Y. Ruichek and R. Touahni, "Local directional ternary pattern: A New texture descriptor for texture classification", Computer Vision and Image Understanding, vol. 169, pp. 14-27, 2018.
- [11] T. Jabid, H. Kacir and O. Chae, "Robust Facial Expression Recognition Based on Local Directional Pattern", ETRI Journal, vol. 32, pp. 784-794, 2010.
- M. Heikkilä, M. Pietikäinen and C. Schmid, "Description of interest regions with local binary patterns", Pattern Recognition, vol. 42, pp. 425-436, 2009.

[12] C. Silva, T. Bouwmans and C. Frelicot, "An eXtended Center-Symmetric Local Binary Pattern for Background Modeling and Subtraction in Videos", 10th International Conference on Computer Vision Theory and Applications, 2015.

1. Francisco: Jossey Bass.
2. Boxall, P., Purcell, J. and Wright, P. (2007). The oxford handbook of human resource management. Oxford university press.
3. Boxall, P., & Purcell, J. (2003). Strategy and Human Resource Management. New York: Palgrave Macmillan.
4. Gratton, L., Hailey, V. H., and Truss, C. (2000). Strategic Human Resource Management. New York: Oxford University Press.
5. Guest, D. E. (1987) Human resource management and industrial relations, *Journal of Management Studies*, 14(5), pp 503–521.
6. Herzberg, F. (1968). One more time: how do you motivate employees? *Harvard Business Review*, January–February, pp 109–120.

Facial expression detection using directional local descriptor

¹ Abdul Ali Damavandi (responsible author of the article), ² Reza Roshni (second author of the article)

¹ lamei Gorgani Institue of Higher Education Gorgan, Iran
Abdolalidamavandi54@gmail.com

² lamei Gorgani Institue of Higher Education Gorgan, Iran
Department of Mechanical Engineering, National University of, Skill (NUS), Tehran, Iran
r.roshany@gmail.com

Abstrac

In recent years, facial expression recognition is considered one of the most important challenges in image processing and has found many applications. Nowadays, due to the so-called emotional relationship between humans and computers in the virtual world, the use of facial recognition methods has become very important. In this research, a new method for recognizing facial expressions is proposed. In order to avoid the limitations and maintain the simplicity and efficiency of traditional LBP, we propose a simple but efficient conceptual and computational texture descriptor, which is called Local Triple Directional Pattern (LDTP). The main advantage of the proposed descriptor over the existing ones is that it combines both concepts of LTP and LDP operators in a similar compact coding scheme, which provides more accurate and separable information. It has also been shown that using a combination of features instead of just one feature makes facial expression better and more robust. Therefore, the features of LDTP and XCSLBP are combined to obtain the final feature vector. Then, for more speed, the final feature vector is dimensionally reduced using the PCA algorithm. Now, this feature vector is given to the SVM, which has been previously trained by the training data, to finally identify one of the 7 facial expressions. The proposed method was examined from various aspects and compared with other existing methods. The accuracy of the proposed method for facial expressions of 7 classes is approx

Keywords:

Facial expression recognition, local binary pattern, local directional pattern, local ternary pattern, support vector machine.