چهارمین کنفرانس ماشین بینایی و پردازش تصویر ایران
۲۶-۲۵ بهمن ماه ۱۳۸۵

شناسایی حالت چهره با استفاده از پایگاه دادهٔ مکانی-زمانی ITMI و QIM

هادی صدوقی یزدی
گروه الکترونیک دانشکدهٔ فنی و مهندسی، دانشگاه تربیت معلم سبزوار
sadoghi@sttu.ac.ir

محمود امین طوسی
گروه ریاضی، دانشگاه تربیت معلم سبزوار
دانشکدهٔ مهندسی کامپیوتر، دانشگاه علم و صنعت ایران

محمود فتحی
دانشکدهٔ مهندسی کامپیوتر، دانشگاه علم و صنعت ایران
mahFathy@iust.ac.ir

چکیده: در این مقاله دو روش جدید پشته‌سازیِ مکانی-زمانی روی تصاویر ویدیویی ارائه شده و از آن برای شناسایی حالت چهره استفاده می‌شود. این روشها نوعی پایگاه دادهٔ مکانی-زمانی محسوب می‌شوند که اطلاعات زمانی و مکانیِ نقاط متحرکِ صحنه، روی یک کلیشه ذخیره می‌شود. روش اولِ پشته‌سازی، شاملِ زمان رخدادِ هر نقطهٔ متحرک است. در این روش، زمانهای رخدادِ هر حرکت ذخیره شده و نتیجهٔ نهایی به تعداد تصاویر دنباله، نرمالیزه می‌شود. از این پایگاهِ دادهٔ جدید که شامل اطلاعات مکان و زمان رخدادِ هر حرکت است ویژگیهایِ هندسی استخراج می‌شود. در روش دوم تکرار رخدادهایِ حرکت در هر ناحیه از تصویر بدست می‌آید. در این روش، تصاویر دریافتی به 30 ناحیه چندی سازی (کوانتیزه) می‌شوند و یک بردار ویژگی 30تایی نتیجهٔ مجموعهٔ فریمهای دریافتی است. با ویژگیهایِ استخراج شده از این دو پایگاهِ داده مبتنی بر پشته‌سازی فریمها و یک شبکهٔ عصبی روی مجموعهٔ 150 تایی از حالات چهرهٔ پایگاه دادهٔ کوهن-کاناد، نرخ شناسایی 80٪ بدست می‌آید. سهولت استخراج ویژگیهای پیشنهادی نسبت به ویژگیهای مورد استفاده در روشهای موجود، آنرا برای کاربردهای بلادرنگ مناسب می سازد.

واژه های کلیدی: پشته‌سازی مکانی-زمانی، شناسائی احساس، حالت چهره، تصاویر ویدیویی

1-مقدمه

در سیستم‌های پایگاهِ دادهٔ ویدیویی، یکی از مهمترین روشها برای توصیف تصاویرِ ویدیویی استفاده از روابط مکان و زمان بین اشیاء موجود در صحنه است. عموماً خاصیت مکانی یک شیء در تصویر شامل اطلاعاتی در مورد مکان یک شیء در یک فریم ویدیویی است. در مراجع [1]-[3] یک استراتژی متداول توصیف مکانیِ شیء، استفاده از موقعیت آن به شکل مختصات دو بعدی است. افزودن فاصلهٔ زمانیی که شیء در یک محدودهٔ مکانی واقع می‌شود و جهت نسبی اشیاء، اطلاعات بیشتری از مکان و زمانِ حضورِ شیء در اختیار قرار می‌دهد[4]. توصیف مکانی-زمانی در بازیابی حوادث و فشرده‌سازی اطلاعات ویدیویی کاربرد زیادی دارد[5]-[6]. در کاربردِ بازیابی حادثه، با استخراج ویژگیهایی از هر فریم و

تلفیق آن با یک مدل زمانی رویِ مجموعهٔ فریمها، یک مدل مکانی-زمانی از حادثه بدست می‌آید[7]-[8].

یکی از روشهای نمایش دانشِ مکانی و زمانی، روش پشته‌سازی فریمها است. تکنیک پشته‌سازی فریم‌ها[1] در[6] ارائه شده است. در این تکنیک چند فریم از یک عمل به‌نحوی با هم ترکیب می‌شوند که همواره نتیجهٔ ترکیب نوعی هموارسازی زمانی می‌باشد. ممکن است سطوح خاکستری یا ویژگی‌هایی از حوزهٔ تبدیلِ چند فریم، ترکیب شوند. از طرفی در مرجع فوق‌الذکر این قضیه مطرح می‌شود که ضرب داخلی دو بردار هموار شده، بطور متوسط از یک مقدار آستانهٔ معیّن بزرگتر است. بنابراین با هموارسازی مکانی با استفاده از فیلترهایی در حوزهٔ مکان وجمع چند فریم متوالی یک نوع هموارسازی مکانی-زمانی بدست می‌آید که مجموعه‌ای از آنها، حالات مختلف یک شیء را شامل می‌شود. روش پشته‌سازی فریم‌ها در لب‌خوانی برای تشخیص گفتار نیز استفاده می‌شود[9].

در [10] تاریخچهٔ حرکت در تصویر یا کلیشه‌ای با عنوان *MHI*[2] ذخیره می‌شود که در آن مکان و مدّت زمانِ رخداد حرکت درج می‌شود. در این تصویر، جهت حرکاتِ انجام شده، ثبت نمی‌شود. در تصویر دیگری اطلاعات موقعیت و چگونگی انجام عمل ذخیره می‌شود. این تصویر *MFH*[3] نامیده می‌شود. *MHI* طبق رابطهٔ زیر ایجاد می‌شود:

$$MHI(k,l)=\begin{cases} t\ , & if\ \left|m_x^{kl}(t)\right|+\left|m_y^{kl}(t)\right|\neq 0 \\ 0\ , & elsewhere \end{cases} \qquad (1)$$

که در رابطهٔ فوق t زمان وقوع عمل است و (k,l) موقعیت رخداد در تصویر است. $m_x^{kl}(t)$ و $m_y^{kl}(t)$ مولفه‌های بردار حرکت در لحظهٔ t و موقعیت (k,l) در راستاهای x و y است. بنابراین در تصویر *MHI* فقط زمان رخداد موقعیت‌های وقوع هر عمل ذخیره می‌شود. با توجه به این‌که عمل در طول زمان انجام شود، مقدار t زیاد شده و در تصویر نتیجه، آن نقاط روشنتر خواهد بود. تصویر جهتِ حرکتِ *MFH*، شامل موقعیت و چگونگی انجام عمل است یا مکان‌های رخداد عمل در تصویر و جهت حرکت آن عمل را نشان می‌دهد که طبق رابطهٔ زیر تعریف می‌شود:

$$MFH_d(k,l)=\begin{cases} m_d^{kl}(t) & if\ E\left[m_d^{kl}(t)\right]<T \\ M\left(m_d^{kl}(t)\right) & elsewhere \end{cases} \qquad (2)$$

که $E\left[m_d^{kl}(t)\right]=\left\|m_d^{kl}(t)-med\left(m_d^{kl}(t),...,m_d^{kl}(t-a)\right)\right\|$ و $M\left(m_d^{kl}(t)\right)=med\left(m_d^{kl}(t),...,m_d^{kl}(t-a)\right)$ که $m_d^{kl}(t)$ جزء افقی یا عمودی بردار حرکت یعنی $m_x^{kl}(t)$ یا $m_y^{kl}(t)$ است. a تعداد فریم‌های گذشته است که 3 تا 5 فریم در نظر گرفته شده است. رابطهٔ فوق نشان می‌دهد اگر بردار حرکت فریم فعلی نسبت به میانهٔ a فریم قبلی از حدی بیشتر (T)باشد نویز تلقی می‌شود و مقدار میانهٔ a فریم قبلی در تصویر ثبت می‌شود ولی اگر این بردار کمتر از میانه بود مقدار بردار حرکتِ فریم فعلی قرار می‌گیرد. این کار برای حذف بردارهای ناشی از نویز است. *MFH* و *MHI* مکمل هم هستند یعنی دارای اطلاعات مکانی و جهتی و زمانی می‌باشند. از این تصاویر که نوعی بانک مکانی زمانی‌اند، در شناسایی عمل استفاده شده‌اند[10].

در پایگاه مکانی-زمانیِ *MHI* حرکت مکرر در یک موقعیت در زمانهای مختلف، تفاوتی با حرکت در آخرین لحظه نمی‌کند، بعبارت ساده‌تر فقط زمان آخرین تغییر در هر موقعیت در کلیشهٔ *MHI* ذخیره می‌شود. این موضوع باعث می‌شود، ذخیره کردن الگوهای پیچیدهٔ ویدیویی دچار مشکل شود[10]. ما در مقالهٔ حاضر، به ارائهٔ یک *MHI* تصحیح شده می‌پردازیم بطوریکه تمامِ زمانهایِ رخداد، در هر موقعیت در کلیشهٔ پیشنهادی وجود دارد. از طرفی این مقاله با ارائهٔ یک کلیشهٔ مناسب، اطلاعات مکانیِ بدست آمده از مجموعه سکانسهای ویدیویی را ثبت می‌کند. استفاده از کوانتیزاسیون تصویر در ایجاد این کلیشه باعث کاهش حساسیت کلیشه نسبت به چرخش شیء در صحنه می گردد که این موضوع در شناسایی حالات چهره بسیار مناسب است.

بطور خلاصه نکات برجستهٔ این مقاله شامل

الف-ارائهٔ پایگاهِ دادهٔ مکانی-زمانی با توانایی ذخیرهٔ زمانهای رخداد یک عمل و موقعیت آنها در صحنه

[1] Stacking Frames
[2] Motion History Image
[3] Motion Flow History

ب-کاهش اثر چرخش و دوران در ذخیرهٔ یک عمل با استفاده از یک پایگاه مکانی مناسب

ج-آزمون دو پایگاه دادهٔ مکانی-زمانی در شناسایی حالات چهره روی پایگاه دادهٔ تهیه شده توسط کوهن-کانادا[11].

این مقاله در 4 بخش ارائه می‌شود. بخش دوم این مقاله به ارائهٔ روشهای موجود در شناسایی حالات چهره اختصاص دارد. بخش سوم یک سیستم جدیدِ شناساییِ حالت چهره که از پایگاه دادهٔ مکانی-زمانی پیشنهادی سود می جوید معرفی شده نتایج بدست آمده ارائه می‌گردد و بخش نهایی شامل نتیجه‌گیری است.

2- کارهای انجام شده در شناسایی حالات چهره

شناسایی عواطف و احساسات کاربر به عنوان یکی از ابزار ارتباط غیر کلامی انسان و ماشین، تحقیقات زیادی را در دهه‌های اخیر به خود معطوف داشته است[26]. یکی از مهمترین تحقیقات انجام شده در مورد نحوه تغییر چهره توسط اکمن انجام پذیرفته است که منجر به تدوین ***(Facial Action Coding System) FACS*** شده است. در FACS هر واحد حرکت **(Action Unit)** به تغییری در چهره اطلاق می شود که اولا به تنهایی قابل انجام نباشد و ثانیاً قابل تقسیم نباشد. [12]-[14]. مثلا حرکت "باز کردن دهان همراه با بالا انداختن ابروها" گرچه یکباره انجام می‌گیرد به دو حرکت "بالا انداختن ابرو" و "باز کردن دهان تقسیم می شود که مستقل از هم می توانند انجام گیرند. از آنجا که این سیستم مبنای بسیاری از کارهای شناسائی اتوماتیک یا نیمه اتوماتیک احساس بوده است، شناسائی واحدهای حرکت خود به یکی از موضوعات مقالات تبدیل شده است[12]-[13]، [15]-[16].

در همین راستا کاناد، کوهن تیان پایگاه داده‌ای شامل 2105رشته تصویر از حالات مختلف 182 نفر که واحدهای حرکت آنها به صورت دستی استخراج شده است را آماده نموده اند[11] که در بسیاری از مطالعات انجام شده مورد استفاده قرار گرفته است.

در[17] با استفاده از سیستمهای خبره روی FACS نرخ شناسائی 90.57 برای 6 حالت بدست آمده است. در این مقاله تصاویر مورد استفاده از دو زاویه گرفته شده اند و 29 واحد حرکتی پوشش داده شده است. تصاویر ثابت بوده و پس از استخراج ویژگیها و واحدهای حرکتی با یک سیستم خبره و بر اساس شدت ظهور هر واحد، حالت چهره مشخص می شود.

مقدم [12] با سیستمی مبتنی بر روش کار پانتیک، نرخ شناسائی 80درصد را برای 6حالت اصلی و حالت معمولی و بر روی قسمتی از پایگاه داده کاناد-کوهن بدست آورده است. اولیور و همکارانش [18] با مدلسازی دهان، از مدل مخفی مارکوف برای شناسائی 5 حالت، روی داده‌های 2000 نمونه برخط 4 از 8 نفر استفاده نموده و دقت 95.95درصد را گزارش نموده اند.

کالدر با استفاده از آنالیز مؤلفه‌های اصلی برای شش حالت به 84درصد دقت دست یافته است[19]. در روشی مشابه اما بر اساس چهره‌های ویژه، جم زاد [24] به شناسایی 7 حالت با استفاده از یک طبقه‌بندِ ماشین بردار پشتیبان فازی می‌پردازد. این کار روی پایگاه تصاویر ***JAFFE*** [20] شامل 213 تصویر از ده نفر انجام شده که نرخ شناسایی 89/77 درصد بدست آمده است.

دابیشن با انجام اصلاحی روی PCA و ترکیب آن با درخت تصمیم به نرخ شناسائی 87.6 درصد برای 6 حالت رسیده اند[21]. چن متد کلاسه بندی جدیدی مبتنی بر LDA به نام ***CDA(Clustering based Discriminant Analysis)*** را پیشنهاد و به دقت 93 درصد برای شناسائی سه حالت معمولی خوشحال و عصبانی رسیده اند[22].

البته مکانیابی چهره در تصویر نیز احتیاج به الگوریتم‌های مناسبی دارد که یاکوب در [23] به آن پرداخته است و ژانگ [25] آنرا در نور مادون قرمز با یک ردیاب کالمن انجام می‌دهد. در بخش بعدی به شرح روش پیشنهادی می‌پردازیم.

[4] Real Time

3 -روش پیشنهادی

در این بخش به ارائهٔ یک پایگاه دادهٔ مکانی-زمانی جدید پرداخته، نحوه و نتایج استفاده از آنرا در یک سیستم شناسائی حالت چهره خواهیم دید.

بجای بکارگیری زمانِ حرکت در رابطهٔ (1) برای ثبت زمانِ آخرین حرکت، زمانِ رخدادِ تمامِ حرکات، ذخیره می‌شود. ماتریسِ بدست آمده را ماتریسِ تجمع زمانِ حرکت می‌نامیم(*ITMI*[5]).

$$ITMI_T(k,l) = \begin{cases} \dfrac{\left(\sum_i i + ITMI_i(k,l)\right)}{N}, & if\ |d(k,l)| > Thre \\ 0, & elsewhere \end{cases} \quad (3)$$

که i شمارهٔ فریم است و (k,l) موقعیت رخداد در تصویر است. $d(k,l)$ تفاضل فریم i از فریم اولیه است و N تعداد کل فریمهای یک حالت از هر فرد است. *Thre* آستانهٔ بکار رفته برای آشکارسازی حرکت است که در شناسایی حالت چهره عدد 30 انتخاب می‌شود. در محاسبهٔ *ITMI* یک هموارسازی میانگین استفاده شده است که باعث کاهش نویز می‌شود که در آن مقدار اولیهٔ $ITMI_0(k,l) = 0$ است. تصویر *ITMI* نرمالیزه است و طول سکانسهای یک حالت، اثری بر آن ندارد. همچنین هر تغییری در هر لحظه در محاسبهٔ *ITMI* موثر است و برخلافِ محاسبهٔ *MHI* در رابطهٔ (1)، اثرات حرکاتِ قبلی از بین نمی‌روند. در این روش تمام زمانهای رخدادِ هر حرکت جمع می‌شوند. وزنِ هر حرکت، زمان یا شمارهٔ فریم آن است و نتیجهٔ نهایی به طول سکانس، نرمالیزه می‌شود. این تصویر که نوعی پایگاهِ داده است شامل اطلاعات مکان و زمان رخدادِ هر حرکت است. همچنین با توجه به قضیهٔ ارائه شده در پیوست، در تصویر *ITMI* سطح *DC* حذف می‌شود و اثرات نویز کاهش می‌یابد. حال آنکه *MHI* [10]، شامل سطح *DC* مخربی است و اطلاعات زیادی حذف می‌گردد.

با افزودن تعداد رخدادهای هر حرکت به این پایگاه می‌توان اطلاعات بیشتری برای ایجاد یک پایگاهِ مناسب، جمع‌آوری کرد. همچنین برای کاهش میزان محاسبات و کاهش اثر حرکات ناخواسته در ایجاد پایگاه مکانی-زمانی، از کوانتیزاسیون تصویر استفاده می‌کنیم و یک ماتریسِ چندی سازی (کوانتیزه) شدهٔ تکرارِ حرکت (*QIM*[6])ارائه می‌کنیم. برای هر پیکسل که دارای $|d(k,l)| > Thre$ باشد *QIM* یک واحد افزایش می‌یابد.

$$QIM_t(m,n) = QIM_{t-1}(m,n) + 1 \quad (4)$$

(k,l) پیکسلی است که دارای حرکت است و در یکی از m در n ناحیه(m تعداد سطرها است) قرار می‌گیرد. m و n تعداد نواحیی است که تصویر به این نواحی تقسیم می‌شود. در این کاربرد m و n بترتیب 6و 5 می‌باشند.

در ادامهٔ این بخش پس از بیان چارچوب کلی سیستم پیشنهادی برای شناسائی احساس از روی حالت چهره به جزئیات نحوه استخراج ویژگیها از *ITMI* و *QIM* می‌پردازیم.

واحدهای حرکتی ذکر شده در *FACS* – که در اغلب تحقیقات شناسائی احساس مورد استفاده قرار گرفته است – به واقع مبتنی بر حرکات ماهیچه های صورت در حین بروز یک احساس هستند. در سیستم پیشنهادی به جای استفاده از واحدهای حرکتی، از *ITMI* و *QIM* بسادگی برای مدل کردن تغییرات ماهیچه های صورت بهره گرفته شده است که استخراج آنها به مراتب از استخراج *FACS* ساده تر بوده و در عین حال کارائی خوبی نیز از خود نشان می دهند.

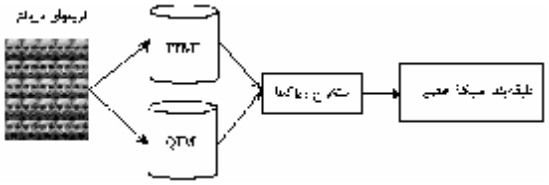
اولین مرحله در سیستم جداسازی صورت از تصاویر ورودی می باشد. در برخی ار کارهای انجام شده این مرحله به صورت خودکار درآمده است. اما ما در اینجا آنرا بصورت دستی انجام داده ایم. به این ترتیب که ناحیه صورت در اولین تصویر از هر دنباله از تصاویر به صورت دستی مشخص شده و با توجه به حرکت کم سر در سایر تصاویر هر دنباله، از این مختصات در سایر فریم ها نیز استفاده شده است.یک نمونه از تصاویر بکار رفته و ناحیه صورت استخراج شده در شکل 1 دیده می شود:

[5] Integrated Time Motion Image

[6] Quantized Iterance of Motion

شکل 1: استخراج صورت

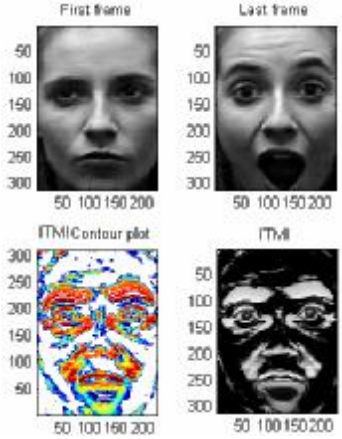
در مرحله دوم ویژگیهای ذکر شده استخراج شده و پس از تنظیم وزنهای یک شبکه *MLP* در مرحله آموزش، از این ویژگیها برای شناسائی حالات چهره بهره گرفته شده است. چار چوب کلی سیستم در شکل 2نشان داده شده است.



شکل 2: سیستم پیشنهادی مبتنی بر پایگاهِ دادهٔ مکانی-زمانی در شناسایی حالت چهره

3-1- استخراج ویژگی از *ITMI*

در ابتدا ویژگیهای استخراج شده از *ITMI* توضیح داده می‌شود. نمونه‌ای از حالتِ تعجب (اولین و آخرین فریم) و تصویر *ITMI* در شکل زیر نشان داده شده است.



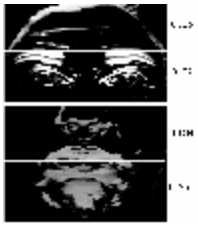
شکل 3: حالت تعجب و تصویر *ITMI*. تصاویر بالایی اولین و آخرین فریم از حالت تعجب است و دو تصویر پایینی تصویر اصلی *ITMI* و کانتور آن است.

همانطور که در تصویر کانتور شکل 3 دیده می‌شود، کلیشه *ITMI*ای استخراج شده به خوبی تغییرات ماهیچه ها در فرآیند بروز احساس تعجب را نشان می دهد. پرتحرک‌ترین نقاط در تعجب این شخص، ابروها و ماهیچه‌های اطراف بینی است. با توجه به خصوصیات تصویر *ITMI* می‌توان ویژگیهایی از آن استخراج کرد که در تشخیص 6 حالت چهره مناسب باشد. پنج ویژگیِ هندسی از *ITMI* استخراج می‌شود که عبارتند از:

ویژگی 1: برای داشتن یک نگرش کلی به تصویر *ITMI* مجموع انرژی فوقانی (مقادیر موجود در ماتریس *ITMI*)تصویر را به نیمهٔ پایینی آن بدست می‌آوریم. همانطور که در نمونهٔ نشان داده شده در شکل 4 دیده می‌شود، حالتِ خوشحال دارای *ITMI* نامتقارن و حالتِ تعجب دارای *ITMI* متقارن می‌باشد. بنظر می‌رسد با همین یک ویژگی بتوان این دو حالت را تشخیص داد ولی بدلیل متنوع بودن حالات و تنوع بروز حالات در افراد مختلف این ویژگی برای تشخیص کافی نیست. در ادامه، چهار ویژگی هندسی دیگر استخراج می‌شوند.

شکل 4: تفاوت بین نیمهٔ فوقانی و تحتانی در تصویر *ITMI* برای دو حالت تعجب(تصویر بالایی) و خوشحال(تصویر پایینی).

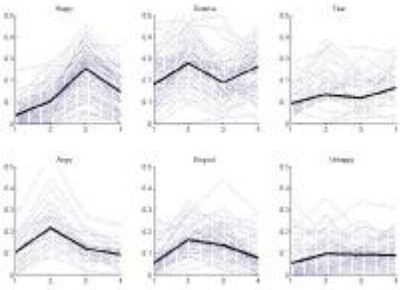
ویژگیهای 2 الی 5: این ویژگیها نوعی بررسی هر واحد حرکت (*Action Unit*) است. تصویر *ITMI* به 4 ناحیهٔ افقی مساوی شبیه شکل زیر تقسیم می‌شود و متوسط سطوح آن به عنوان یک ویژگی استخراج می‌شود. (شکل 5) می توان گفت که به نحوی این چهار ویژگی به ترتیب نمایانگر میزان تغییرات چهره در قسمت پیشانی، چشم و ابرو، بینی و دهان و چانه می باشند.



شکل 5: مقادیر ویژگیهای 2 تا 5 و نواحی مربوط به هر ویژگی مستخرج از *ITMI*

با استفاده از ویژگیهای بدست آمده می‌توان انتظار داشت یک طبقه‌بند بتواند حالات مختلف را شناسایی کند، این موضوع بدلیل آنست که ویژگیهای ارائه شده در تشخیص این حالات می‌تواند بخوبی این الگوها را متمایز سازد. درشکل 6 ویژگی های شمارهٔ 2 تا 5 برای 6 حالت و 300 نمونه ، این ویژگیها تصویر شده اند که تفاوت قابل ملاحظه‌ای در مقدار متوسط آن – که در شکل با خطوط توپر نشان داده شده است– برای هر حالت دیده می‌شود (به جز حالات 3و6).

همانگونه که از سطر دوم شکل 4 پیداست در حین بروز احساس خوشحالی بیشترین تغییرات چهره مربوط به قسمت سوم چهره و پس از آن مربوط به چهارمین قسمت می باشد. (متناظر با حالت میانگین احساس خوشحالی –*happy*– درشکل 6) و همانگونه که می دانیم در حین بروز این احساس لب و دهان بیشترین تغییرات را دارند. نگاهی به سطر اول شکل 43 مشخص می سازد که بیشترین تغییرات چهره در حین بروز احساس تعجب مربوط به قسمتهای دوم و چهارم می باشد. این نکته را می توان مترادف با تغییرات در ناحیه چشم و ابرو و لب و فک پایین دانست (متناظر با حالت میانگین احساس تعجب –*surprise*– درشکل 6).



شکل 6: نمایش و مقایسهٔ ویژگیهای 2 تا 5 برای 6 حالت چهره روی 300 نمونه – خطوط توپر متوسط هر گروه را نشان می دهند.

استفاده از فقط این 4 ویژگی و بکار گرفتن الگوریتم کلاسبندی نزدیکترین همسایه برای دو حالت اول و دوم نرخ شناسائی 92٪، برای حالات اول تا سوم 77 درصد و برای تمام شش حالت نرخ شناسائی 58 درصد را در حالت متوسط نشان داد.

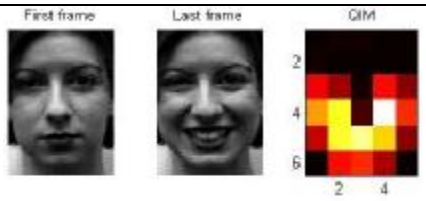
همچنین نتایج پیاده سازی سیستم پیشنهادی نشان داد که تنها استفاده از این ویژگی برای تفکیک دو حالت خوشحال و متعجب، می تواند نرخ شناسائی 100٪ را به همراه داشته باشد. اما افزایش تعداد حالات نرخ شناسائی را به شدت کاهش می دهد.

3-2- ویژگیهای مستخرج از *QIM*

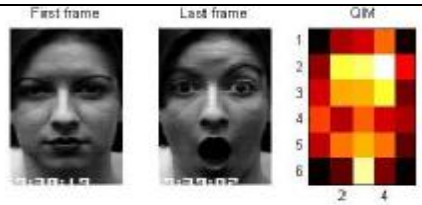
همانطور که گفته شد *QIM* یک ماتریس 6 در 5 است که در شکل 7 نشان داده شده است که هر عنصر آن نشان دهندهٔ میزان تکان در یکی از 30 ناحیه می‌باشد. تکانهای زیاد در بعضی از نواحی باعث روشنتر شدن نواحی در ستون آخر شکل 7 می‌شود که بر افزایش شمارندهٔ ناحیه دلالت می‌کند.

بنابراین ویژگیهای 6 تا 35 مربوط به این 30 ناحیه می‌باشد. مراجعه به شکل 7 نشان می دهد که *QIM* تقریب بسیار خوبی از میزان تغییرات ماهیچه ها در هر ناحیه است. به عنوان مثال آخرین تصویر در سطر اول شکل 7 نشان می دهد که در حین بروز احساس خوشحالی بیشترین تغییرات مربوط به گونه ها و لبها می باشد.یا در سطر دوم شکل 7 روشن ترین خانه های *QIM* متناظر با تغییرات ماهیچه های پیشانی و حرکت فک پایین در حالت تعجب می باشد.

حالت: خوشحال (*Happy*)



حالت: متعجب (*Surprise*)



شکل 7: نمایش ویژگیهای *QIM* برای دو حالت خوشحال و متعجب

یکی از نتایج استفاده از این کوانتیزاسیون کاهش حساسیت کلیشه نسبت به چرخش سر می باشد که این موضوع در شناسایی حالات چهره حائز اهمیت است. برخی از محققین به

منظور اجتناب از اثرات منفی چرخش سر سعی در حذف آن داشته اند[16]

3-2 نتایج پیاده سازی

سیستم پیشنهادی روی پایگاه تصاویر کوهن-کاناد روی 6 حالت چهره و 300 نمونه ساخته و آزمایش می‌شود. این سیستم شامل پایگاه داده‌های مکانی-زمانیِ *ITMI* و *QIM* است. ویژگیهای مناسب شامل یک بردار 35 تایی از آنها استخراج می‌شود و شبکهٔ عصبی 4 لایه آموزش می‌بیند (روی 150 نمونه). نتایج بدست آمده روی 150 نمونه آزمون نرخ شناسایی 80٪ را نشان می‌دهد. بدلیل کوانتیزه بودن *QIM* و ویژگیهای ناحیه‌ای مستخرج از *ITMI* حرکات محدود افقی و عمودی تاثیری بر بدست آوردن *ITMI* و *QIM* ندارد.

پیاده سازی سیستم روی یک کامپیوتر شخصی با پردازنده *Intel Celeron 2.2 GHz*، زمانهای متوسط 2.7 ثانیه و 0.003 ثانیه را به ترتیب برای محاسبه ویژگیها برای یک دنباله با 22 تصویر – صرفنظر از زمان خواندن فایل- و شناسائی حالت برای هر دنباله پس از استخراج ویژگیها را در پی داشته است. استفاده از تعداد افراد و حالات نسبتاً زیاد (78 نفر و 300 نمونه) در مقایسه با برخی از روشها – که از 8 یا ده نفر استفاده نموده اند، قابلیت اعتماد سیستم را افزایش داده است. در برخی از مقالات حالت معمولی هم به عنوان یک حالت در نظر گرفته شده است که با توجه به ویژگیهای پیشنهادی، اضافه کردن این حالت به داده های مورد استفاده، مطمئناً نرخ شناسائی را افزایش خواهد داد.

4- نتیجه‌گیری

روش جدید پشته‌سازیِ مکانی-زمانی روی تصاویر ویدیویی و آزمون آن در شناسایی حالات چهره موضوع این مقاله است. نوعی پایگاه دادهٔ مکانی-زمانی که اطلاعات زمانی و مکانی نقاط متحرکِ صحنه، را در خود دارد، ارائه شد. روش اولِ پشته‌سازی شامل زمان رخدادِ هر نقطهٔ متحرک است. این پایگاههای داده در اصل کلیشه‌هایی است که از مجموعهٔ فریمها بدست می‌آید و نوعی روش پشته‌سازی فریمهاست. با ویژگیهایِ استخراج شده از این دو پایگاهِ داده و استفاده از یک شبکهٔ عصبی روی یک مجموعهٔ آموزشی 150 تائی و مجموعه آزمون 150تائی حالات چهره از پایگاه دادهٔ کوهن-کاناد، نرخ شناسایی 80٪ بدست آمد. استخراج ویژگیهای مورد استفاده در شیوه پیشنهادی برخلاف بسیاری از ویژگیهای معمولِ روشهای موجود به سادگی قابل محاسبه بوده و سیستم را برای کاربردهای بلادرنگ مناسب می سازد.

تشکر و قدردانی

در خاتمه مولفین وظیفه خود می دانند که از جناب ***Nicki Ridgeway*** که پایگاه داده کوهن-کاناد را در اختیار مولفین قرار دادند تشکر و قدردانی به عمل آورند.

پیوست: پشته‌سازیِ با وزن زمانیِ تفاضل بین فریمی

در این پیوست نشان می‌دهیم که پشته‌سازیِ روی تفاضل بین فریم‌ها با وزنِ زمان رخدادِ آنها باعث ایجاد یک سیگنالِ جدید، بدون سطح *DC* می‌شود، همچنین باعث افزایش سیگنال به نویز سیگنال منتجّه می‌شود. به این منظور آنرا در حوزهٔ پیوسته و روی توابع مشتق پذیر بررسی می‌کنیم و سپس می‌توان آنرا به حوزهٔ اعداد گسسته تعمیم داد و از آن در پشته‌سازی فریمها با وزنِ زمان رخداد، اعمال کرد.

قضیه: اگر $f(t)$ یک تابع پیوسته و مشتق‌پذیر باشد آنگاه $G(t)$، برابر $f(t)$ مدوله شده بدون سطح DC است و سیگنال به نویز بیشتری است.

$$G(t)=\int_{t=0}^{t=t} t\ f(t)df(t) \qquad \text{(الف-1)}$$

اثبات:

$$G(t)=\int_{t=0}^{t=t} t\ f(t)dt \qquad \text{(الف-2)}$$
$$=tf(t)-\int_{t=0}^{t=t} f(t)dt$$

عبارت فوق را بصورت زیر می‌توان نوشت.

$$G(t)=tf(t)-t\times\frac{1}{t}\int_{t=0}^{t=t} f(t)dt \qquad \text{(الف-3)}$$
$$=t(f(t)-f_{DC}(t))$$

که در عبارت فوق $f_{DC}(t)$ مولفهٔ DC تابع $f(t)$ است. بنابراین در $G(t)$ ابتدا سطح DC $f(t)$ حذف، سپس در t

مدوله می‌شود. همانطور که در حوزهٔ فوریه دیده‌ایم مدوله شدن در حوزهٔ زمان معادل مشتق‌گیری در حوزهٔ فرکانس می‌شود یعنی

$$t\,\hat{f}(t) \xrightarrow{\text{Fourier Domain}} \frac{d}{df}\hat{f}(f) \qquad \text{(الف–4)}$$

که $\hat{f}(t)$ همان $f(t)$ است که سطح DC آن حذف شده است و داریم:

$$G(f) = \frac{d}{df}\hat{f}(f) \qquad \text{(الف–5)}$$

حال اگر در طیف سیگنال $\hat{f}(t)$ نویزی با تابع توزیع یکنواخت باشد با توجه به رابطهٔ فوق، در $G(t)$ نویز کاهش می‌یابد. بعبارت ساده‌تر مشتق چنین نویزی دارای طیفی با دامنهٔ صفر است که بمعنی حذف کامل نویز از سیگنال در حالت ایده‌ال است.

مراجع

[1] J.Z. Li, M.T. Ozsu, D. Szafron, "Modeling of moving objects in a video database", Proceedings of IEEE Int. Conf. on Multimedia Computing and Systems, Ottawa, Canada, pp. 336–343, June 1997.

[2] E. Oomoto, K. Tanaka, OVID: "design and implementations of a video-object database system", IEEE Trans. on Knowledge and Data Engineering, vol.5, no.4, pp.629–643, 1993.

[3] D. Papadias, Y. Theodoridis, "Spatial Relations, Minimum Bounding Rectangles and Spatial Data Structures", Int. Journal of Geographical Information Science, vol.11, pp.111–138, 1997.

[4] M. Koprulu, N. K. Cicekli, A. Yazici, "Spatio-Temporal Querying in Video Databases", Information Science, vol. 160, pp.131-152, 2004.

[5] F. M. Idris, S. Panchanathan, "Spatio-Temporal Indexing of Vector Quantized Video Sequences", IEEE Trans. on Circuit and Systems for Video Technology, vol. 7, no. 5, pp.728-740, Oct. 1997.

[6] M. Osadchy, D. Keren, "A Rejection-Based Method for Event Detection in Video," IEEE Trans. on Circuits and Systems for Video Technology, vol.14, no.4, pp.534-541, Apr.2004.

[7] D. Hogg, "Model-based vision: A program to see a walking person," Image Vis. Comput., vol. 1, no. 1, pp. 5–20, 1983.

[8] C. Bregler, "Learning and recognizing human dynamics in video sequences," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 568–574, Puerto Rico, 1997.

[9] N. Li, S. Dettmer, and M. Shah, "Visually Recognizing Speech Using Eigensequences," in Motion-Based Recognition, pp. 345–371, 1997.

[10] R. V. Babua, K. R. Ramakrishnanb, "Recognition of Human Actions Using Motion History Information Extracted from the Compressed Video," Image and Vision Computing, vol.22, pp.597–607, 2004.

[11] T. Kanade, J.F. Cohn, and Y. Tian, "Comprehensive Database for Facial Expression Analysis", Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, March 2000.

[12] م. منصوری زاده، ن. مقدم چرکری، ا. کبیر، "سیستم خبره شناسائی احساس از روی تصویر ویدیویی چهره" ، نهمین کنفرانس سالانه انجمن کامپیوتر ایران، دانشگاه صنعتی شریف - صفحات 353-361، 1382.

[13] J. F. Cohn, T. Kanade, "Automated Facial Image Analysis for Measurement of Emotion Expression, To appear in J. A. Coan & J. B. Allen (Eds.), The handbook of emotion elicitation and assessment. Oxford University Press Series, Oxford.

[14] FACS - Facial Action Coding System, http://www-2.cs.cmu.edu/~face/facs.htm

[15] M. Pantic, L.J. M. Rothkrantz, "Facial Action Recognition for Facial Expression Analysis From Static Face Image",IEEE Trans. on Sys, Man, and Cyber. —Part B: Cybernetics, Vol.34, No.3, 2004.

[16] J. Lien, T. Kanade, J. F. Cohn, C.C. Li, "Detection, tracking, and classification of action units in facial expression", Robotics and Autonomous Systems, vol.31, pp.131–146, 2000.

[17] M. Pantic, L.J.M. Rothkrantz, "Expert system for Automatic Analysis of Facial Expressions," Image and Vision Computing, vol.18, pp. 881–905, 2000.

[18] N. Oliver, A.Pentland, F. Bérard, "LAFTER: a real-time face and lips tracker with facial expression recognition", Pattern Recognition, vol.33, pp.1369-1382, 2000.

[19] A. J. Calder, A. M. Burton, P. Miller, A. W. Young, S. Akamatsu,"A Principal Component Analysis of Facial Expressions", Vision Research, vol.41, pp.1179–1208, 2001.

[20] JAFFE Japanese Analysis Female Facial Expression Database.

[21] S.Dubuission, F. Davoine, M. Masson, "A Solution for facial expression representation and recognition", Signal Proccessing: Image Communication, vol.17, pp.657-673, 2002.

[22] X. Chen, T. Huang, "Facial expression recognition: A clustering-based approach", Pattern Recognition Letters 24, 1295–1302, 2003.

[23] Y. Yacoob, L. S. Davis, "Computing Spatio-Temporal Representations of Human Faces," IEEE Conf. Computer Vision and Pattern Recognition, 1994.

[24] س. خلیفی، م. جمزاد، " آشکارسازی هیجانات چهره با استفاده از چهره‌های ویژه و ماشین بردار پشتیبان فازی،" سومین کنفرانس ماشین بینایی و پردازش تصویر ایران، جلد 2، ص. 453-460، تهران، اسفند 1383.

[25] Y. Zhang, Q. Ji, "Active and Dynamic Information Fusion for Facial Expression Understanding from

Image Sequences," IEEE Trans. On Pattern Analysis and Machine Intelligence, vo.27, no.5, pp. 699-714, May 2005.

[26] M. Pantic, L. J. M. Rothkrantz, "Automatic Analysis of Facial Expressions: The State of the Art," IEEE Trans. On Pattern Analysis and Machine Intelligence, vol.22, no.12, pp.1424-1445, Dec. 2000.