

روش جدید پشته‌سازی مکانی-زمانی حرکت در تصاویر ویدیویی و استفاده از آن در شناسایی حالت چهره

هادی صدوقی یزدی^۱

محمود امین طوسی^۲

محمود فتحی^۲

۱- گروه الکترونیک دانشکده فنی و مهندسی، دانشگاه تربیت معلم سبزوار

۲- دانشگاه علم و صنعت ایران، دانشکده مهندسی کامپیوتر

چکیده: در این مقاله دو روش جدید پشته‌سازی مکانی-زمانی روی تصاویر ویدیویی ارائه شده و از آن برای شناسایی حالت چهره استفاده می‌شود. این روشها نوعی پایگاه داده مکانی-زمانی محسوب می‌شوند که اطلاعات زمانی و مکانی نقاط متحرک صحنه، روی یک کلیشه ذخیره می‌شود. روش اول پشته‌سازی، شامل زمان رخداد هر نقطه متحرک است. در این روش، زمانهای رخداد هر حرکت ذخیره شده و نتیجه نهایی به طول سکانس، نرمالیزه می‌شود. از این پایگاه داده جدید که شامل اطلاعات مکان و زمان رخداد هر حرکت است ویژگیهای هندسی استخراج می‌شود. در روش دوم تکرار رخدادهای حرکت در هر ناحیه از تصویر بدست می‌آید. در این روش، تصاویر دریافتی به ۳۰ ناحیه کوانتیزه می‌شوند و یک بردار ویژگی ۳۰ تایی نتیجه مجموعه فریمهای دریافتی است. با ویژگیهای استخراج شده از این دو پایگاه داده مبتنی بر پشته‌سازی فریمها و یک شبکه عصبی روی مجموعه ۱۸۰ تایی از حالات چهره پایگاه داده کوهن-کاناد، نرخ شناسایی ۹۴/۹۶٪ بدست می‌آید.

کلمات کلیدی: پشته‌سازی مکانی-زمانی، پایگاه داده، حالت

چهره، تصاویر ویدیویی.

۱-مقدمه

در سیستم‌های پایگاه داده ویدیویی، یکی از مهمترین روشها برای توصیف تصاویر 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 [۱]-[۳] یک استراتژی متداول توصیف مکانی

شیء، استفاده از موقعیت آن به شکل مختصات دو بعدی است. افزودن فاصله زمانی که شیء در یک محدوده مکانی واقع می‌شود و جهت نسبی اشیاء، اطلاعات بیشتری از مکان و زمان حضور شیء در اختیار قرار می‌دهد [۴]. توصیف مکانی-زمانی در بازیابی حوادث و فشرده‌سازی اطلاعات ویدیویی کاربرد زیادی دارد [۵]- [۶]. در کاربرد بازیابی حادثه، با استخراج ویژگیهایی از هر فریم و تلفیق آن با یک مدل زمانی روی مجموعه فریمها، یک مدل مکانی-زمانی از حادثه بدست می‌آید [۷]-[۸].

یکی از روشهای نمایش دانش مکانی و زمانی، روش پشته‌سازی فریمها است. تکنیک پشته‌سازی فریمها^۱ در [۶] ارائه شده است. در این تکنیک چند فریم از یک عمل به‌نحوی با هم ترکیب می‌شوند که همواره نتیجه ترکیب نوعی هموارسازی زمانی می‌باشد. ممکن است سطوح خاکستری یا ویژگی‌هایی از حوزه تبدیل 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 آستانه معین بزرگتر است. بنابراین با هموارسازی مکانی با استفاده از فیلترهایی در حوزه مکان و جمع 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 [۹].

در [۱۰] تاریخچه حرکت در تصویر یا کلیشه‌ای با عنوان MHI^۲ ذخیره می‌شود که در آن مکان و مدت زمان رخداد حرکت درج می‌شود. در این تصویر، جهت حرکات انجام شده، ثبت نمی‌شود. در تصویر دیگری اطلاعات موقعیت و چگونگی انجام

^۱ Stacking Frames

^۲ Motion History Image

عمل ذخیره می‌شود. این تصویر MFH¹ نامیده می‌شود. MHI طبق رابطه زیر ایجاد می‌شود:

$$MHI(k, l) = \begin{cases} \tau, & \text{if } |m_x^{kl}(\tau)| + |m_y^{kl}(\tau)| \neq 0 \\ 0, & \text{elsewhere} \end{cases} \quad (1)$$

که در رابطه فوق τ زمان وقوع عمل است و (k, l) موقعیت رخداد در تصویر است. $m_x^{kl}(\tau)$ و $m_y^{kl}(\tau)$ مولفه‌های بردار حرکت در لحظه τ 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 این که عمل در طول زمان انجام شود، مقدار τ زیاد شده و در تصویر نتیجه، آن نقاط روشنتر خواهد بود. تصویر جهت حرکت MFH، شامل موقعیت و چگونگی انجام عمل است یا مکان‌های رخداد عمل در تصویر و جهت حرکت آن عمل را نشان می‌دهد که طبق رابطه زیر تعریف می‌شود:

$$MFH_d(k, l) = \begin{cases} m_d^{kl}(\tau) & \text{if } E[m_d^{kl}(\tau)] < T \\ M(m_d^{kl}(\tau)) & \text{elsewhere} \end{cases} \quad (2)$$

که $E[m_d^{kl}(\tau)] = \|m_d^{kl}(\tau) - med(m_d^{kl}(\tau), \dots, m_d^{kl}(\tau - \alpha))\|$ و $M(m_d^{kl}(\tau)) = med(m_d^{kl}(\tau), \dots, m_d^{kl}(\tau - \alpha))$ که $m_d^{kl}(\tau)$ جزء افقی یا عمودی بردار حرکت یعنی $m_x^{kl}(\tau)$ یا $m_y^{kl}(\tau)$ است. α تعداد فریم‌های گذشته است که ۳ تا ۵ فریم در نظر گرفته شده است. رابطه فوق نشان می‌دهد اگر بردار حرکت فریم فعلی نسبت به میانه α فریم قبلی از حدی بیشتر (T) باشد نوین تلقی می‌شود و مقدار میانه α فریم قبلی در تصویر ثبت می‌شود ولی اگر این بردار کمتر از میانه بود مقدار بردار حرکت فریم فعلی قرار می‌گیرد. این کار برای حذف بردارهای ناشی از نوین 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 [۱۰].

در پایگاه مکانی-زمانی MHI حرکت مکرر در یک موقعیت در زمانهای مختلف، تفاوتی با حرکت در آخرین لحظه نمی‌کند، عبارت ساده‌تر فقط زمان آخرین تغییر در هر موقعیت در کلیشه MHI ذخیره می‌شود. این موضوع باعث می‌شود، ذخیره کردن الگوهای پیچیده و ویدیویی دچار مشکل شود [۱۰]. ما در مقاله حاضر، به ارائه یک MHI تصحیح شده می‌پردازیم بطوریکه تمام زمانهای رخداد، در هر موقعیت در کلیشه پیشنهادی وجود دارد. از طرفی این مقاله با ارائه یک کلیشه مناسب، اطلاعات مکانی 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 می‌گردد که این موضوع در شناسایی حالات چهره بسیار مناسب است.

بطور خلاصه نکات برجسته این مقاله شامل

- الف- ارائه پایگاه داده مکانی-زمانی با توانایی ذخیره زمانهای رخداد یک عمل و موقعیت آنها در صحنه
- ب- کاهش اثر چرخش و دوران در ذخیره یک عمل با استفاده از یک پایگاه مکانی مناسب
- ج- آزمون دو پایگاه داده مکانی-زمانی در شناسایی حالات چهره روی پایگاه داده تهیه شده توسط کوهن-کاناد [۱۱].

این مقاله در ۴ بخش ارائه می‌شود. بخش دوم این مقاله به ارائه روشهای موجود در شناسایی حالات چهره اختصاص دارد. بخش سوم به ارائه پایگاه داده مکانی-زمانی پیشنهادی می‌پردازد و از آن در یک سیستم جدید شناسایی حالت چهره، استفاده می‌شود و نتایج بدست آمده ارائه می‌گردد و بخش نهایی شامل نتیجه‌گیری است.

۲- 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 [۲۶]. یکی از مهمترین تحقیقات انجام شده در مورد نحوه تغییر چهره توسط اکمن انجام پذیرفته است که منجر به تدوین FACS (Facial Action Coding System) شده است. در FACS هر واحد حرکت (Action Unit) به تغییری در چهره اطلاق می‌شود که اولاً به تنهایی قابل انجام نباشد و ثانیاً قابل تقسیم نباشد. [۱۲]-[۱۴]. مثلاً حرکت "باز کردن دهان همراه با بالا انداختن ابروها" گرچه یکباره انجام می‌گیرد به دو حرکت "بالا انداختن ابرو" و "باز کردن دهان تقسیم می‌شود که مستقل از هم می‌توانند انجام گیرند. از آنجا که این سیستم مبنای بسیاری از کارهای شناسایی اتوماتیک یا نیمه اتوماتیک احساس بوده است، شناسایی واحدهای حرکت خود به یکی از موضوعات مقالات تبدیل شده است [۱۲]-[۱۳]، [۱۵]-[۱۶].

در همین راستا کاناد، کوهن تیان پایگاه داده‌ای شامل ۲۱۰۵ رشته تصویر از حالات مختلف ۱۸۲ نفر که واحدهای حرکت آنها به صورت دستی استخراج شده است را آماده نموده‌اند [۱۱] که در بسیاری از مطالعات انجام شده مورد استفاده قرار گرفته است.

در [۱۷] با استفاده از سیستمهای خبره روی FACS نرخ شناسایی ۹۰،۵۷٪ برای ۶ حالت بدست آمده است. در این مقاله تصاویر مورد استفاده از دو زاویه گرفته شده‌اند و ۲۹ واحد حرکتی پوشش

¹ Motion Flow History

ماتریس بدست آمده را ماتریس تجمع زمان حرکت می نامیم (ITMI²).

(۳)

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

که i شماره 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. $Thre$ آستانه بکار رفته برای آشکارسازی حرکت است که در شناسایی حالت چهره عدد ۳۰ انتخاب می شود. در محاسبه ITMI یک هموارسازی میانگین استفاده شده است که باعث کاهش نویز می شود که در آن مقدار اولیه $ITMI_0(k, l) = 0$ است. تصویر ITMI نرمالیزه است و طول سکانسهای یک حالت، اثری بر آن ندارد. همچنین هر تغییری در هر لحظه در محاسبه ITMI موثر است و برخلاف محاسبه MHI در رابطه (۱)، اثرات حرکات قبلی از بین نمی روند. در این روش تمام زمانهای رخداد هر حرکت جمع می شوند. وزن هر حرکت، زمان یا شماره فریم آن است و نتیجه نهایی به طول سکانس، نرمالیزه می شود. این تصویر که نوعی پایگاه داده است شامل اطلاعات مکان و زمان رخداد هر حرکت است. همچنین با توجه به قضیه ارائه شده در پیوست، در تصویر ITMI سطح DC حذف می شود و اثرات نویز کاهش می یابد. حال آنکه MHI [۱۰]، شامل سطح DC مخربی است و اطلاعات زیادی حذف می گردد.

با افزودن تعداد رخدادها هر حرکت به این پایگاه می توان اطلاعات بیشتری برای ایجاد یک پایگاه مناسب، جمع آوری کرد. همچنین برای کاهش میزان محاسبات و کاهش اثر حرکات ناخواسته در ایجاد پایگاه مکانی-زمانی، از کوانتیزاسیون تصویر استفاده می کنیم و یک ماتریس کوانتیزه شده تکرار حرکت (QIM³) ارائه می کنیم. برای هر پیکسل که دارای $|d(k, l)| > Thre$ باشد QIM یک واحد افزایش می یابد.

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

(k, l) پیکسلی است که دارای حرکت است و در یکی از m در n ناحیه (m تعداد سطرها است) قرار می گیرد. m و n تعداد نواحی است که تصویر به این نواحی تقسیم می شود. در این کاربرد m و n بترتیب ۶ و ۵ می باشند. در ادامه این بخش به ویژگیهای استخراجی

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 [۱۲] با سیستمی مبتنی بر روش کار پاتیک، نرخ شناسایی ۸۰ درصد را برای ۶ حالت اصلی و حالت معمولی و بر روی قسمتی از پایگاه داده کاناد-کوهن بدست آورده است. اولیور و همکارانش [۱۸] با مدلسازی دهان، از مدل مخفی مارکوف برای شناسایی ۵ حالت، روی داده های برخط^۱ استفاده نموده و دقت ۹۵.۹۵ درصد را گزارش نموده اند.

کالدر با استفاده از آنالیز مؤلفه های اصلی برای شش حالت به ۸۴ درصد دقت دست یافته است [۱۹]. در روشی مشابه اما بر اساس چهره های ویژه، جم زاد [۲۴] به شناسایی ۷ حالت با استفاده از یک طبقه بند ماشین بردار پشتیبان فازی می پردازد. این کار روی پایگاه تصاویر JAFFE [۲۰] شامل ۲۱۳ تصویر از ده نفر انجام شده که نرخ شناسایی ۸۹/۷۷ درصد بدست آمده است.

دایشن با انجام اصلاحی روی PCA و ترکیب آن با درخت تصمیم به نرخ شناسایی ۸۷،۶ درصد برای ۶ حالت رسیده اند [۲۱]. چن متد کلاسه بندی جدیدی مبتنی بر LDA به نام CDA(Clustering based Discriminant Analysis) را پیشنهاد و به دقت ۹۳ درصد برای شناسایی سه حالت معمولی خوشحال و عصبانی رسیده اند [۲۲].

البته مکانیابی چهره در تصویر نیز احتیاج به الگوریتم های مناسبی دارد که یاکوب در [۲۳] به آن پرداخته است و ژانگ [۲۵] آنرا در نور مادون قرمز با یک ردیاب کالمن انجام می دهد ولی در این مقاله ما عمل مکانیابی را بطور دستی انجام می دهیم. پس از مکانیابی روی پایگاه داده کوهن-کاناد، از یک طبقه بند شبکه عصبی پرسپترون ۴ لایه با یادگیری بر اساس حداقل مربعات پس انتشار خطا برای شناسایی استفاده می شود البته نکته قابل ملاحظه سیستم پیشنهادی در ویژگیهای استخراجی از یک پایگاه داده مکانی-زمانی جدید ارائه شده و اعمال شده روی تصاویر چهره می باشد. در بخش بعدی به شرح روش پیشنهادی می پردازیم.

۳- روش پیشنهادی

در این بخش به ارائه یک پایگاه داده مکانی-زمانی جدید می پردازیم. بجای بکارگیری زمان حرکت در رابطه (۱) برای ثبت زمان آخرین حرکت، زمان رخداد تمام حرکات، ذخیره می شود.

² Integrated Time Motion Image

³ Quantized Iterance of Motion

¹ Real Time

از ITMI و QIM می‌پردازیم و سپس سیستم پیشنهادی را معرفی می‌کنیم.

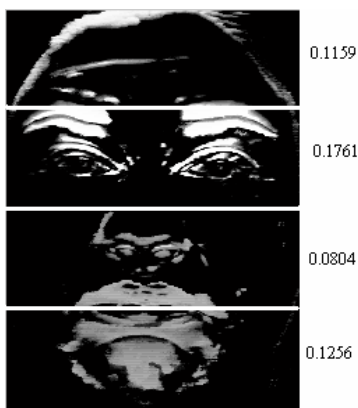
۳-۱- استخراج ویژگی از ITMI

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



شکل ۲: تفاوت بین نیمه فوقانی و تحتانی در تصویر ITMI برای دو حالت تعجب (تصویر بالایی) و خنده (تصویر پایینی).

ویژگیهای ۲ الی ۵: این ویژگیها نوعی بررسی هر واحد حرکت (Action Unit) است. تصویر ITMI به ۴ ناحیه افقی مساوی شبیه شکل زیر تقسیم می‌شود و متوسط سطوح آن به عنوان یک ویژگی استخراج می‌شود.

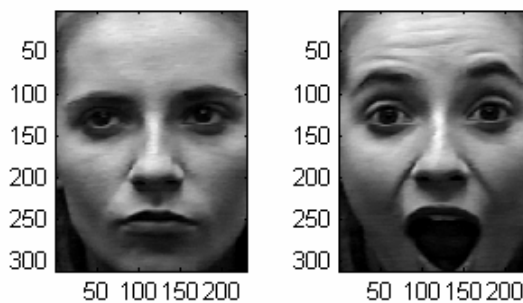


شکل ۳: مقادیر ویژگیهای ۲ تا ۵ و نواحی مربوط به هر ویژگی مستخرج از ITMI

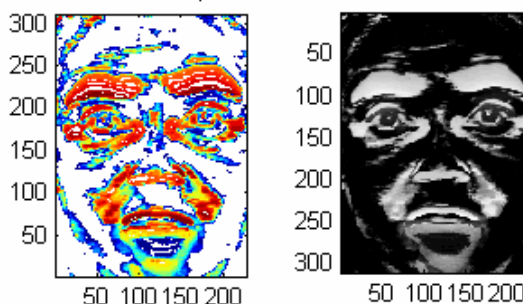
۳-۲- ویژگیهای مستخرج از QIM

همانطور که گفته شد QIM یک ماتریس ۶ در ۵ است که در شکل ۴ نشان داده شده است که هر عنصر آن نشان دهنده میزان تکان در یکی از ۳۰ ناحیه می‌باشد. تکانهای زیاد در بعضی از نواحی باعث روشتتر شدن نواحی در شکل ۴ می‌شود که بر افزایش شمارنده ناحیه دلالت می‌کند.

First frame Last frame



ITMI Contour plot



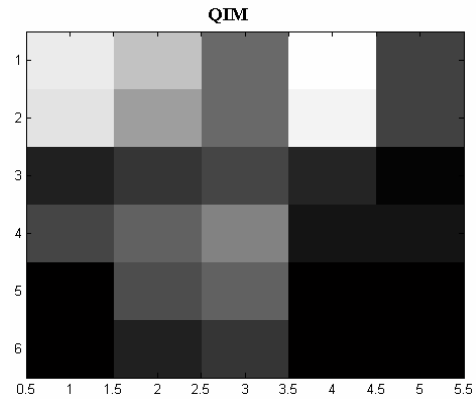
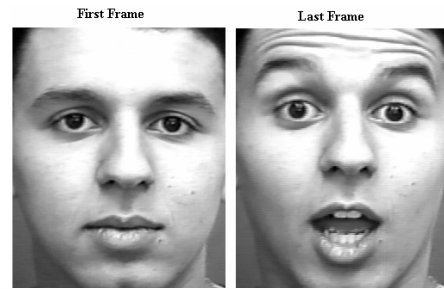
شکل ۱: حالت تعجب و تصویر ITMI، تصاویر بالایی اولین و آخرین فریم از حالت تعجب است و دو تصویر پایینی تصویر اصلی ITMI و کانتور آن است.

همانطور که در تصویر کانتور ITMI دیده می‌شود 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 ۶ حالت چهره مناسب باشد. پنج ویژگی هندسی از ITMI استخراج می‌شود که عبارتند از:

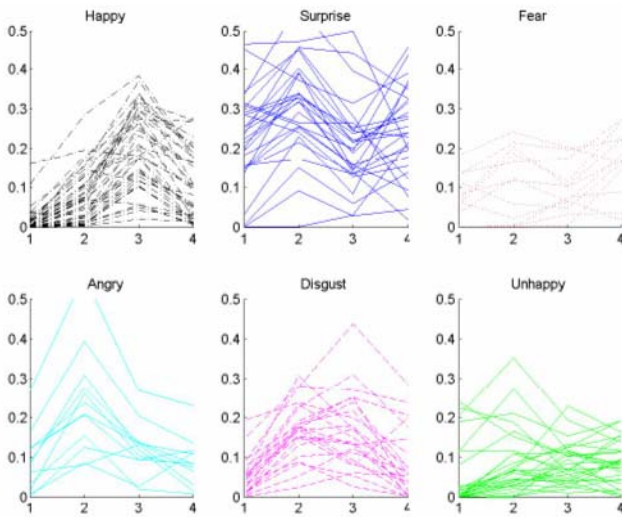
ویژگی ۱: برای داشتن یک نگرش کلی به تصویر ITMI مجموع انرژی فوقانی (مقادیر موجود در ماتریس ITMI) تصویر را به نیمه پایینی آن بدست می‌آوریم. همانطور که در نمونه نشان داده شده در شکل ۲ دیده می‌شود، حالت خنده دارای نامتقارن و حالت تعجب دارای ITMI متقارن می‌باشد. بنظر می‌رسد با همین یک ویژگی بتوان این دو حالت 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.

۴- نتیجه گیری

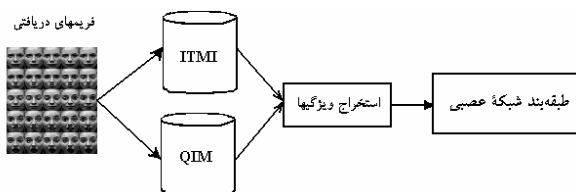
روش جدید پشته‌سازی مکانی-زمانی روی تصاویر ویدیویی و آزمون آن در شناسایی حالات چهره موضوع این مقاله است. نوعی پایگاه داده مکانی-زمانی که اطلاعات زمانی و مکانی نقاط متحرک صحنه، را در خود دارد، ارائه شد. روش اول پشته‌سازی شامل زمان رخداد هر نقطه متحرک است. این پایگاه‌های داده در اصل کلیشه‌هایی است که از مجموعه فریم‌ها بدست می‌آید و نوعی روش پشته‌سازی فریم‌هاست. با ویژگی‌های استخراج شده از این دو پایگاه داده و استفاده از یک شبکه عصبی روی مجموعه حالات چهره ۱۸۰ تایی از پایگاه داده کوهن-کانادا، نرخ شناسایی ۹۴/۹۶٪ بدست آمد.



شکل ۴: نمایش ویژگی‌های QIM برای حالت تعجب چهره



شکل ۵: نمایش و مقایسه ویژگی‌های ۲ تا ۵ برای ۶ حالت چهره روی ۱۵ نفر



شکل ۶: سیستم پیشنهادی مبتنی بر پایگاه داده مکانی-زمانی در شناسایی حالت چهره

پیوست: پشته‌سازی با وزن زمانی تفاضل بین فریمی

در این پیوست نشان می‌دهیم که پشته‌سازی روی تفاضل بین فریم‌ها با وزن زمان رخداد آنها باعث ایجاد یک سیگنال جدید، بدون سطح DC می‌شود، همچنین باعث افزایش سیگنال به نویز سیگنال منتجه می‌شود. به این منظور آنها در حوزه پیوسته و روی

بنابراین ویژگی‌های ۶ تا ۳۵ مربوط به این ۳۰ ناحیه می‌باشد. حال از این ویژگی‌ها برای شناسایی ۶ حالت چهره استفاده می‌کنیم.

۳-۳- سیستم پیشنهادی

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. بعنوان مثال ویژگی شماره ۲ تا ۵ که مستخرج از ITMI است در شکل ۵ نشان داده شده است. در این شکل برای ۶ حالت و ۱۵ نفر، این ویژگی تصویر شده است که تفاوت قابل ملاحظه‌ای در مقدار هر ویژگی برای هر حالت دیده می‌شود.

سیستم پیشنهادی در شکل ۷ نشان داده شده است. این سیستم روی پایگاه تصاویر کوهن-کانادا روی ۶ حالت چهره و ۱۸۰ نفر ساخته و آزمایش می‌شود. این سیستم شامل پایگاه داده‌های مکانی-زمانی ITMI و QIM است. ویژگی‌های مناسب شامل یک بردار ۳۵ تایی از آنها استخراج می‌شود و شبکه عصبی ۴ لایه آموزش می‌بیند (روی ۹۰ نفر). نتایج بدست آمده روی ۹۰ نفر آزمون نرخ شناسایی ۹۴،۹۶٪ نشان می‌دهد. بدلیل کوانتیزه بودن QIM و ویژگی‌های ناحیه‌ای مستخرج از ITMI حرکات محدود افقی و عمودی تأثیری بر بدست آوردن ITMI و QIM ندارد.

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. Boston, MA: Kluwer, 1997, pp. 345–371.
- [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 (FG'00), March 2000, Grenoble, France.

[۱۲] م. منصورى زاده، ن. مقدم چركرى، ا. كبرى، "سيستم خبره شناسائى احساس از روى تصوير وىديوئى چهره" - نهمين كنفرانس

توابع مشتق پذير بررسى مى كنيم و سپس مى توان آنرا به حوزه اعداد گسسته تعميم داد و از آن در پشته سازى فرمها با وزن زمان رخداد، اعمال كرد.

قضيه: اگر $f(t)$ يك تابع پيوسته و مشتق پذير باشد آنگاه $G(t)$ (رابطه الف-۱)، برابر $f(t)$ مدوله شده بدون سطح DC است و سيگنال به نويز بيشترى است.

$$G(t) = \int_{\tau=0}^{\tau=t} \tau f(\tau) d\tau \quad (\text{الف-۱})$$

اثبات:

$$G(t) = \int_{\tau=0}^{\tau=t} \tau f(\tau) d\tau \quad (\text{الف-۲})$$

$$= tf(t) - \int_{\tau=0}^{\tau=t} f(\tau) d\tau$$

عبارت فوق را بصورت زير مى توان نوشت.

$$G(t) = tf(t) - t \times \frac{1}{t} \int_{\tau=0}^{\tau=t} f(\tau) d\tau \quad (\text{الف-۳})$$

$$= t(f(t) - f_{DC}(t))$$

که در عبارت فوق $f_{DC}(t)$ مولفه DC تابع $f(t)$ است. بنابراین در $G(t)$ ابتدا سطح DC حذف، سپس در t مدوله می شود. همانطور که در حوزه فوریه دیده ایم مدوله شدن در حوزه زمان معادل مشتق گیری در حوزه فرکانس می شود یعنی

$$t \hat{f}(t) \xrightarrow{\text{Fourier Domain}} \frac{d}{df} \hat{f}(f) \quad (\text{الف-۴})$$

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

$$G(f) = \frac{d}{df} \hat{f}(f) \quad (\text{الف-۵})$$

حال اگر در طیف سیگنال $\hat{f}(t)$ نویزی با تابع توزیع یکنواخت باشد با توجه به رابطه فوق، در $G(t)$ نویز کاهش می یابد. عبارت ساده تر مشتق چنین نویزی دارای طیفی با دامنه صفر است که بمعنی حذف کامل نویز از سیگنال در حالت ایده آل است.

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.

سالانه انجمن کامپیوتر ایران - دانشگاه صنعتی شریف - ۱۳۸۲ -
صفحات ۳۵۳-۳۶۱

[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 in Affective Science. New York: 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, June 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. Dubuisson, F. Davoine, M. Masson, "A Solution for facial expression representation and recognition", Signal Processing: Image Communication, vol.17, pp.657-673, 2002.

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

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

[۲۴] س. خلیفی، م. جمزاد، "آشکارسازی هیجان‌ها با استفاده از چهره‌های ویژه و ماشین بردار پشتیبان فازی"، سومین کنفرانس ماشین بینایی و پردازش تصویر ایران، جلد ۲، ص. ۴۵۳-۴۶۰، تهران، اسفند ۱۳۸۳.

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