

# مطالعه‌ای روی طراحی رابط تعاملی انسان و رایانه، بر اساس فیله‌های واگرایی حرکتی برای تشخیص حرکت دست

علیرضا محمودی فرد<sup>۱</sup>، علی ملکی<sup>۲\*</sup>

<sup>۱</sup> دانشجوی کارشناسی ارشد ناپیوسته مدیریت صنعتی، دانشکده علوم انسانی، دانشگاه شاهد (و فارغ‌التحصیل کارشناسی ارشد مهندسی برق و مدرس دانشگاه‌ها)، تهران، ایران، alireza10.m10@gmail.com

<sup>۲</sup> دانشجوی کارشناسی ارشد ناپیوسته مهندسی مخابرات دانشگاه تربیت دبیر شهید رجایی، تهران، ایران، A.malekibme@gmail.com

## اطلاعات مقاله

ناریخچه مقاله:

تاریخ دریافت مقاله: ۱۴۰۲/۰۵/۱۰

تاریخ پذیرش مقاله: ۱۴۰۲/۰۶/۱۳

تاریخ انتشار مقاله: ۱۴۰۲/۰۶/۱۵

کلمات کلیدی:

HCI

تعامل انسان و کامپیوتر

TFIDF

استخراج الگوی فضایی

حرکت دست

فیله‌های واگرایی حرکتی

## چکیده

حرکات دست انسان برای تعامل انسان با رایانه (HCI)، شامل استفاده از حرکات یا وضعیت دست و انگشتان برای سیگنال دادن به یک فرمان یا کار به رایانه یا دستگاه تلفن همراه است، در حالی که پوزیشن‌های دو بعدی شامل سیگنال‌دهی دستورات با لمس یک سطح حسگر است؛ تبلت یا صفحه نمایش حرکات دو بعدی، ممکن است شامل حرکات مختلفی باشد، مانند تکان دادن برای تغییر پنجره یا باز کردن نوک انگشتان برای نشان دادن بزرگ شدن، اما محدود به حس کردن نوک انگشتان روی یک سطح است؛ اگرچه به‌طور کلی ردیابی حرکت مفصل دست دشوار است، رویکردهای مبتنی بر مثال، راه‌حلی قوی برای تشخیص ژست دست ارائه می‌دهند. احتمالاً، مجموعه‌ای غنی از حرکات دست پویا برای یک سیستم تشخیص معنادار مورد نیاز است. نحوه ساختن نمایش بصری برای الگوهای حرکتی، کلید تشخیص مقیاس‌پذیر است. با بررسی صورت گرفته، یک نمایش جدید بر اساس نقشه واگرایی میدان حرکت حرکتی پیشنهاد می‌شود که الگوهای حرکتی را به الگوهای فضایی تبدیل می‌کند؛ در مقالات مینا، با توجه به نقشه‌های واگرایی حرکتی، از آشکارسازهای ویژگی تصویر مدرن برای استخراج الگوهای فضایی برجسته، مانند حداکثر مناطق بیرونی پایدار استفاده شده است. یک توصیف‌گر محلی از هر منطقه استخراج می‌شود تا الگوی حرکت محلی را به تصویر بکشد؛ متعاقباً توصیف‌گرهای نمونه‌های اشاره با استفاده از درخت واژگان از پیش آموزش‌دیده نمایه‌سازی می‌شوند؛ سپس حرکات جدید به‌طور موثر با حرکات پایگاه داده با یک طرح TFIDF مطابقت داده می‌شوند؛ آزمایش‌های گسترده بر روی یک پایگاه داده بزرگ حرکات دست با ۱۰ دسته و ۱۰۵۰ نمونه ویدیویی، کارایی الگوهای حرکتی استخراج‌شده را برای تشخیص ژست تایید می‌کند؛ روش پیشنهادی مقاله مینا، به نرخ تشخیص کلی ۹۷٫۶۲٪ دست می‌یابد، در حالی که میانگین زمان تشخیص تنها ۳۴٫۵۳ میلی‌ثانیه است.

## ۱ - مقدمه

دست‌یابی به نتایج پیشرفته در تشخیص دوره‌ای کنش انسان استفاده کردند؛ با این حال، هیچ‌یک از این روش‌ها، راه‌حلی مقیاس‌پذیر برای تطبیق کارآمد زمانی که پایگاه داده نمونه بزرگ است، ارائه نمی‌دهد؛ در کار قبلی، برای بهبود سرعت زمان اجرا برای تشخیص ژست‌ها، بیشتر تلاش‌ها به افزایش سرعت پردازش تصویر و ویدئو با طراحی الگوریتم‌های کارآمد [۱۷] یا استفاده از واحدهای محاسباتی قدرتمندتر [۶] مانند GPU اختصاص یافته است؛ این قطعا مطلوب است [۳۵ و ۳۶ و ۳۸].

با این وجود، به موضوع ایجاد نمایش بصری کارآمد برای تطبیق حرکات دست مقیاس‌پذیر بر روی پایگاه‌های داده بزرگ نمونه به اندازه کافی پرداخته نشده است؛ یک نمایش بصری جدید از حرکات دست پویا، بر اساس میدان واگرایی حرکات جریان پیشنهاد می‌شود که الگوهای حرکتی حرکتی را به الگوهای تصویر فضایی تبدیل می‌کند؛ با توجه به یک توالی ژست، جریان نوری بین هر دو فریم متوالی استخراج می‌شود؛ میدان واگرایی آن مشتق شده است؛ سپس الگوهای فضایی برجسته، از میدان واگرایی با استفاده از آشکارساز ویژگی حداکثر پایدار مناطق بیرونی (MSER) [۱۳ و ۸] تشخیص داده می‌شود؛ سپس یک توصیف‌گر از هر ناحیه شناسایی شده برای مشخص کردن الگوهای حرکت محلی استخراج می‌شود؛ توصیف-گرهای استخراج شده از تمام دنباله‌های ژست‌های مثالی توسط یک درخت واژگان سلسله‌مراتبی از پیش آموزش‌دیده نمایه‌سازی می‌شوند؛ یک توالی پوزیشن جدید با تطبیق با پایگاه داده شناسایی می‌شود؛ یک طرح TF-IDF [۱۵]، برای پایگاه‌های داده بزرگ مقیاس‌پذیر است؛ خط لوله روش مذکور در شکل ۱ نشان داده شده است [۳۵ و ۳۶ و ۳۸]؛ با بهترین دانش ما، مقاله مبنا، اولین تلاش برای تبدیل الگوهای حرکت محلی به الگوهای تصویر فضایی با استفاده از میدان واگرایی است؛ این مورد ما را قادر می‌سازد تا از تکنیک‌های نمایه‌سازی تصویر پیشرفته برای تشخیص حرکات دستی مقیاس‌پذیر بدون استفاده از تشخیص دست، تقسیم‌بندی یا ردیابی استفاده کنیم؛ یک پایگاه داده قابل توجه از حرکات دست پویا با ۱۰ دسته و ۱۰۵۰ نمونه برای ارزیابی جمع‌آوری شده است که برای مطالعه بیشتر با جامعه پژوهشی به اشتراک گذاشته خواهد شد؛ نرخ تشخیص این روش در مجموعه داده، ارزیابی ۹۷٫۶۲٪ است، با میانگین زمان تشخیص ۳۴٫۵۳ میلی‌ثانیه؛ به عبارت دیگر، رویکرد پیشنهادی در مقاله مربوطه، نه تنها یک رویکرد جدید برای تجزیه و تحلیل الگوی حرکت است، بلکه یک چارچوب مقیاس‌پذیر برای تشخیص حرکات دست پویا با تعداد زیادی مثال را نیز ارائه می‌کند [۳۸].

• تشخیص ژست دست مبتنی بر بینایی به دلیل کاربردهای گسترده آن در رابط‌های کامپیوتری انسانی، کنترل ربات، واقعیت ترکیبی/افزوده و غیره، یک موضوع تحقیقاتی فعال برای دهه‌ها بوده است. تشخیص ژست یک مرور کلی از روش‌های اخیر تشخیص ژست را می‌توان در [۱۴ و ۳۵] یافت؛ با توجه به اطلاعات حرکتی، حرکات دست پویا یک کانال ارتباطی غنی ارائه می‌دهد [۳۷ و ۳۸]. رویکردهای تشخیص حرکات دست پویا را می‌توان به دو حالت دسته‌بندی کرد: مبتنی بر مدل و مبتنی بر نمونه.

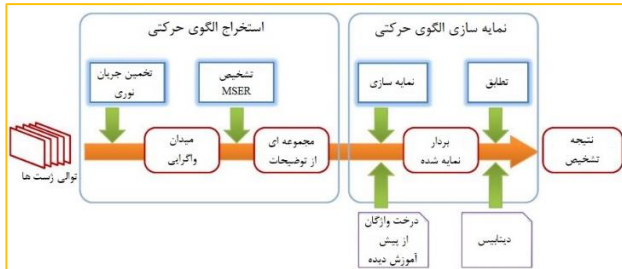
برای رویکرد مبتنی بر مدل، مدل‌های مارکوف پنهان (HMM) شاید بیشترین استفاده را داشته باشند [۲۲، ۱۹، ۵، ۲۱]. در [۲۲]، بردارهای ویژگی پویا با کوانتیزاسیون برداری به دنباله‌های نمادین تبدیل می‌شوند و متعاقبا توسط یک HMM گسسته مدل‌سازی می‌شوند؛ برخی از پیشرفت‌های اخیر نسبت به HMM سنتی شامل مدل شبکه معنایی (SNM) [۱۹]، HMM ناپارامتریک [۵] و میدان تصادفی شرطی پنهان [۲۱] است؛ این گونه‌ها یا تلاش‌های آموزشی را کاهش می‌دهند، یا دقت طبقه‌بندی را بهبود می‌بخشند. رویکردهای مبتنی بر شامل ماشین‌های حالت محدود (FSM) [۱۳]، شبکه‌های بیزی پویا (DBN) [۲۰] و شبکه‌های خودسازماندهی حفظ توپولوژی [۷] است؛ همه این رویکردها فرض می‌کنند که عقربه شناسایی و بیان شده است؛ حرکت ردیابی می‌شود، که یا با تقسیم‌بندی رنگ پوست، یا ردیابی دست مبتنی بر مدل سینماتیک [۱] به دست می‌آید؛ اگرچه آن‌ها نتایج امیدوارکننده‌ای ارائه کرده‌اند، استحکام این رویکردها، به موفقیت قبلی تشخیص دست و حرکت (که اغلب چالش‌برانگیز است) بستگی دارد؛ ردیابی علاوه بر این، آموزش این مدل‌ها قبل از اینکه بتوان آن‌ها را در تشخیص به کار برد، هم داده فشرده و هم از نظر محاسباتی دشوار است [۳۵ - ۳۸].

برای دور زدن مشکلات ردیابی و تشخیص دست، روش‌های دیگر سعی می‌کنند از نمایش بصری ثابت برای تطبیق و شناسایی استفاده کنند؛ در میان این توصیف‌گرها، ویژگی‌های مکانی-زمانی محلی، بیشترین بهره‌برداری را دارند [۹، ۱۸، ۴]. فریمن و راث [۹]، پیشنهاد می‌کنند که گرادینان‌های زمانی-مکانی ممکن است برای تشخیص ژست پویا مفید باشند؛ بیشتر توسط [۱۸] تایید می‌شود که در آن، آن‌ها از SVM برای طبقه‌بندی نقاط علاقه فضا-زمان برای تشخیص کنش استفاده می‌کنند [۴]؛ سایر ویژگی‌های بصری شامل مسیر حرکت [۲۳] و ویژگی‌های تراکم گاوسی اطراف نقاط علاقه منتخب [۱۱] است. اخیرا Chaudhry و همکاران [۲]، دنباله‌ای از هیستوگرام‌های جریان نوری جهت‌یافته را محاسبه کرده و از هسته‌های بینه کوشی در سیستم‌های دینامیکی غیرخطی برای

## ۲- الگوهای بصری حرکت ژست دست

• در این بخش مقاله، نمایش بصری پیشنهادی حرکات ژست دست ارائه شده است؛ در اصل، جریان نوری بین هر دو فریم متوالی تخمین زده می‌شود و میدان واگرایی حرکت استخراج می‌شود، که از آن مناطق MSER و توصیف‌گرهای حرکت محلی استخراج می‌شوند [۳۵ و ۳۶].

چنین تخمینی از حرکت، مبنای مناسبی برای تبعیض در آزمایش‌های مقاله مبنا بود، حتی اگر مقادیر جریان کاملاً دقیق نباشند؛ میدان واگرایی جریان بیشتر نویز جریان را در پس‌زمینه فیلتر کرده است و شکل واضحی از عقربه را ارائه می‌دهد، که تضمین می‌کند اکثر مناطق MSER روی دست قرار دارند؛ به ارائه استخراج توصیف‌گرهای حرکت محلی پرداخته می‌شود [۳۸ و ۳۶].



شکل ۱- خط لوله روش پیشنهادی [۳۸]

## ۳- میدان واگرایی جریان نوری

• در یک میدان برداری، واگرایی عملگری است که بزرگی منبع یا فرورفتگی میدان را اندازه‌گیری می‌کند؛ با توجه به یک بردار  $F = [F_1, F_2, \dots, F_n]^T$  که در فضای اقلیدسی  $n$  بعدی بوده، واگرایی  $F$  را می‌توان به صورت زیر محاسبه کرد [۳۸]:

(۱)

$$\text{div}F = \sum_{i=1}^n \frac{\partial F_i}{\partial x_i}$$

که در آن  $[x_1, x_2, \dots, x_n]^T$  مختصات دکارتی فضایی است که میدان برداری در آن تعریف شده است؛ بر این اساس، برای یک میدان بردار جریان نوری  $F(x, y) = [u(x, y), v(x, y)]^T$  بوده که در آن  $u(x, y)$  و  $v(x, y)$  اجزای عمودی جریان نوری در موقعیت  $(x, y)$  هستند و واگرایی  $F$  برابر است با [۳۸ و ۳۶]:

(۲)

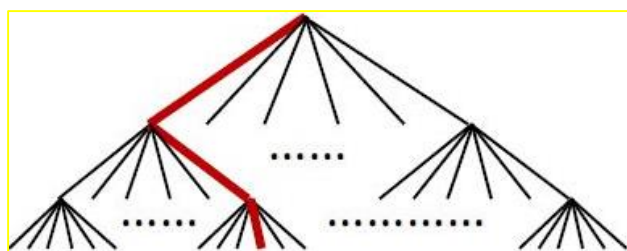
$$\text{div}F = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}$$

شکل ۲، نمونه‌ای از تبدیل یک میدان حرکت جریان به یک میدان واگرایی را نشان می‌دهد؛ شکل ۲ (a) اولین فریم یک جفت تصویر است و شکل ۲ (b) و (c) به ترتیب تجسم  $u$  و  $v$  هستند؛ میدان واگرایی مربوطه در شکل ۲ (d) نشان داده شده است؛ جریان نوری را با استفاده از الگوریتم Lucas-Kanade [۱۲] محاسبه می‌کنند و الگوریتم را روی GPU پیاده‌سازی می‌نمایند تا سرعت فریم پردازش را افزایش دهند؛ روش لوکاس-کانادا، همان‌طور که در اینجا اعمال می‌شود، به‌طور مستقیم تخمین می‌زند جریان در هر پیکسل چنین تخمینی بر کنتراست و بافت محلی متکی است و فقط برای حرکات کوچک معتبر است [۳۸]؛ از آنجایی که هیچ تخمین چند تفکیک-پذیری اعمال نمی‌شود، جریان محاسبه شده برای مناطق بزرگ با بافت کم، مانند فضای داخلی ناحیه دست در مثال، معتبر نیست؛ در نتیجه، جریان در این مناطق به‌سختی قابل برآورد است؛ با این حال،

## ۴- استخراج توصیف‌گر محلی

• هنگامی که میدان واگرایی، به‌دست آمد، شناسایی مناطق فوق‌العاده پایدار حداکثر ۱۳ (MSER) انجام می‌شود؛ ناحیه MSER منحصرًا توسط یک ویژگی فوق‌العاده تابع شدت در منطقه و در مرز بیرونی آن تعریف می‌شود و بنابراین دارای بسیاری از خواص مفید است، از جمله تغییرناپذیری نسبت به تغییر شدت تصویر، پایداری، و امکان تشخیص چند مقیاسی؛ MSER به‌طور گسترده در تطبیق تصویر و تشخیص اشیاء استفاده شده است و منجر به عملکرد تشخیص بهتر [۸] در چندین برنامه شده است؛ در چارچوب پیشنهادی، همان‌طور که در شکل ۲ (f) نشان داده شده است، هر ناحیه MSER با یک بیضی برازش داده می‌شود؛ از آنجایی که پس‌زمینه ثابت است، بیشتر نواحی MSER که در میدان واگرایی حرکتی شناسایی می‌شوند، در مرز عقربه یا درون دست قرار دارند [۳۸]؛ بنابراین ویژگی‌های استخراج شده از این مناطق با در هم‌ریختگی پس‌زمینه مخلوط نمی‌شوند؛ به‌عنوان یک مقایسه، شکل ۲ (e) نواحی MSER را نشان می‌دهد که مستقیماً از تصویر در شکل ۲ (a) مشتق شده‌اند، که در مناطق پس‌زمینه به همان اندازه در دست متحرک شناسایی می‌شوند؛ در هر ناحیه MSER شناسایی شده، یک هیستوگرام جهت‌گیری بردارهای جریان نوری محاسبه می‌شود؛ جهت‌گیری‌های جریان نوری را می‌توان از  $u(x, y)$  و  $v(x, y)$  محاسبه کرد و دارای محدوده  $[0, 2\pi]$  باشد؛ سپس همه جهت‌ها به‌صورت دو خطی کوانتیزه می‌شوند و در سطوح مجزا با قدر مطلق آن‌ها به‌عنوان وزن جمع می‌شوند. شکل ۲ (g) یک تصویر ساده به ما می‌دهد که در آن هیستوگرام دارای ۸ سطل است؛ در عمل، عدد bin را ۸۰

• در زمینه جستجوی تصویر و تشخیص شی، درخت واژگان ساختار سلسله‌مراتبی از مراکز خوشه‌ای و مجموعه‌ای از توصیف‌گرهای آموزشی است؛ این درخت برای اولین بار در [۳۸ و ۱۵] برای تشخیص اشیاء استفاده شد؛ شکل ۳ یک درخت واژگان ساده را با ۶ شاخه و ۳ سطح نشان می‌دهد؛ درخت واژگان را می‌توان با خوشه‌بندی سلسله‌مراتبی  $k$ -means ساخت؛ با توجه به مجموعه داده آموزشی، ابتدا یک فرآیند خوشه‌بندی  $k$ -means برای تعیین مراکز خوشه‌ای انجام می‌شود، جایی که  $k$  فاکتور شاخه درخت است، یعنی تعداد فرزندان هر گره. در شکل ۳،  $k = 6$ . این مراکز  $k$  گره‌های سطح اول درخت را نشان می‌دهند؛ سپس توصیف‌گرهای آموزشی با توجه به فاصله آن‌ها از مراکز خوشه‌ای به گروه‌های  $k$  منشعب می‌شوند؛ در هر گروه، خوشه‌بندی  $k$ -means بیشتر برای تعریف مراکز خوشه‌ای جدید انجام می‌شود، که سپس فرزندان مرکز اصلی هستند؛ همین فرآیند به صورت بازگشتی انجام می‌شود تا زمانی که درخت به حداکثر عمق از پیش تعریف شده دست یابد؛ برای یک درخت واژگان با  $k$  شاخه‌ها و سطوح  $l$ ، تعداد کل گره‌های برگ  $k^l$  خواهد بود [۳۵، ۳۶ و ۳۸].



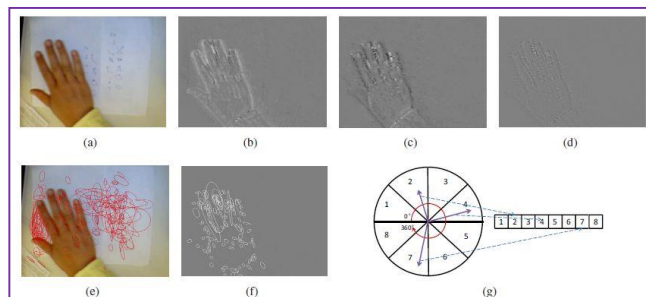
شکل ۳- درخت واژگانی با ۶ شاخه و ۳ سطح [۳۸]

هنگامی که درخت واژگان ساخته شد، توصیف‌گرهای یک تصویر را می‌توان با مقایسه با توصیف‌گرهای گره‌های  $k$  در هر سطح، کمی کرد و با نزدیک‌ترین آن‌ها مرتبط نمود؛ بنابراین هر توصیف‌گر مسیری از ریشه تا یک برگ درخت دارد؛ برای مثال، خط قرمز در شکل ۳، مسیری برای یک توصیف‌گر است؛ چنین مسیری را می‌توان با یک عدد صحیح در برگ کدگذاری کرد و برای نمایه‌سازی و مطابقت، همان‌طور که در بخش زیر توضیح داده شد، استفاده کرد؛ هر فرآیند کوانتیزاسیون برای یک توصیف‌گر فقط شامل مقایسه  $k * l$  است؛ هزینه محاسباتی در تعداد گره‌های برگ لگاریتمی است، که مزیت اصلی ارائه شده توسط ساختار سلسله‌مراتبی می‌باشد [۳۵، ۳۶ و ۳۸].

در خصوص نمایه‌سازی یک تصویر منفرد، پس از اینکه تمام توصیف‌گرهای یک تصویر پرس و جو از طریق درخت واژگان کوانتیزه شدند، می‌توان تصویر را با تصاویر پایگاه داده با مقایسه شباهت‌های مسیری‌های توصیف‌گر آن‌ها مطابقت داد؛ در نظر بگیرد که  $m_i$  و  $n_i$

قرار می‌دهند؛ هیستوگرام در نهایت نرمال می‌شود تا واحد  $L1$ -norm داشته باشد [۳۵ و ۳۶ و ۳۸].

منطق انتخاب هیستوگرام جهت‌گیری جریان به‌عنوان توصیف‌گرهای محلی مقاله مبنای این است که سرعت حرکات به‌ویژه در بین کاربران مختلف به‌طور گسترده‌ای متفاوت است؛ از این‌رو یک الگوریتم تشخیص ژست خوب باید نسبت به سرعت انجام یک ژست حساس نباشد؛ این جهت‌گیری حرکت دست را به‌عنوان معیارهای مهم برای تشخیص نشان می‌دهد؛ آنچه در اینجا به‌دنبال آن هستیم، مجموعه‌ای از توصیف‌گرهای متمایز برای هر حرکت متمایز است [۳۸]؛ در آزمایش‌های انجامی، تایید شده است که چنین توصیف‌کننده‌هایی در حال حاضر بسیار متمایز هستند؛ صرف‌نظر از این واقعیت که یک الگوریتم ساده برای تخمین جریان نوری اتخاذ شده است؛ پس از استخراج توصیف‌گر محلی، هر فیلد واگرایی با مجموعه‌ای از توصیف‌گرهای محلی نشان داده می‌شود و یک حرکت دنباله‌ای از مجموعه‌های توصیف‌گر است؛ از تکنیک‌های تخمین حرکت مانند تقسیم‌بندی و ردیابی، صرف‌نظر می‌شود؛ تشخیص MSER را می‌توان با هزینه محاسباتی کم انجام داد؛ در مقاله مبنای یک پیاده‌سازی زمان خطی اخیر اعمال شده است [۱۶]؛ کل فرآیند استخراج ویژگی، بسیار کارآمد است و عملکرد با نرخ فریم بالا به‌راحتی حاصل می‌شود [۳۵ - ۳۸].



شکل ۲- استخراج الگوی حرکتی؛ (الف) فریم اول یک جفت تصویر (ب) مولفه  $u$  جریان نوری تخمینی [۳۸]، (ج) جزء  $v$  جریان تخمینی، (د) میدان واگرایی، (ه) تشخیص مستقیم MSER روی تصویر در (الف)، (و) تشخیص MSER در میدان واگرایی در (د)، (ز) محاسبه هیستوگرام جهت‌گیری جریان نوری با ۸ سطل از مناطق MSER [۳۸]

## ۶- نمایه‌سازی الگوی حرکتی

• در این بخش، بر روی یک طرح نمایه‌سازی مقیاس‌پذیر مبتنی بر نمایش بصری الگوهای حرکتی پیشنهادی در بخش ۲ برای تشخیص حرکات دست پویا با تعداد زیادی نمونه، تمرکز خواهد شد.

## ۷- درخت واژگان

TF-IDF (فرکانس معکوس - فرکانس سند) است؛ وقتی توصیف-گرهای تصویر پرس و جو همگی کوانتیزه می‌شوند، فقط فایل‌های معکوس گرهای مربوط به عناصر غیر صفر بردار پرس و جو جستجو می‌شوند [۳۸].

فاصله تصویر پرس و جو تا هر یک از تصاویر ثبت شده در فایل‌های وارونه را می‌توان به تدریج با استفاده از رابطه ۵ جمع کرد؛ با استفاده از فایل‌های معکوس برای فقط فواصل بین عناصر غیر صفر بردارها محاسبه می‌شود؛ این به ما امکان می‌دهد از فایل‌های معکوس استفاده کنیم تا از تطبیق مستقیم بردار پرس و جو با تمام بردارهای تصویر در پایگاه داده جلوگیری شود؛ یک فایل معکوس برای هر گر، فایلی است که تعداد تصاویری را که حداقل یک توصیف‌گر به آن گر کوانتیزه شده است و شناسه‌های این تصاویر به همراه تعداد توصیف-گرهای این تصاویر که به آن گر کوانتیزه می‌شوند، ثبت می‌کند؛ تحلیل متن را فرکانس اصطلاحی می‌نامند. اندازه‌گیری فاصله، شامل یک طرح وزن‌دهی TF-IDF (فرکانس معکوس فرکانس سند) است؛ وقتی توصیف‌گرهای تصویر پرس و جو همگی کوانتیزه می‌شوند، فقط فایل‌های معکوس گرهای مربوط به عناصر غیر صفر بردار پرس و جو جستجو می‌شوند؛ فاصله تصویر پرس و جو تا هر یک از تصاویر ثبت شده در فایل‌های وارونه را می‌توان به تدریج با استفاده از رابطه ۵ جمع کرد؛ با استفاده از فایل‌های معکوس برای تطبیق در طرح TF-IDF، هزینه محاسباتی تطبیق تصویر به‌طور قابل توجهی کاهش می‌یابد؛ این رویکرد حتی اگر میلیون‌ها گر برگ و تصاویر پایگاه داده وجود داشته باشد، جستجوی کارآمد را ممکن می‌سازد [۳۵ و ۳۶ و ۳۸].

#### ۸- نمایه‌سازی یک دنباله حرکت ژست

• تکنیک نمایه‌سازی تصویر معرفی شده در بخش سوم، فقط در جستجوی تک تصویر کار می‌کند، در حالی که در بخش دوم، یک حرکت دست به دنباله‌ای با طول متغیر از تصاویر واگرایی تبدیل می‌شود؛ در این بخش، نمایه‌سازی تصویر برای تطبیق توالی ژست‌ها گسترش می‌یابد؛ یک راه‌حل ساده، نمونه‌برداری یکنواخت از فریم‌ها از یک دنباله ژست است؛ با استفاده از روش در بخش سوم، هر قاب نمونه‌برداری شده را می‌توان برای تشکیل یک بردار نمایه کرد  $a^i = [a_1^i, a_2^i, \dots, a_M^i]^T$  که در آن  $i$  نشان‌دهنده  $i$ -امین قاب نمونه است و  $M$  تعداد گرهای برگ در درخت واژگان است. بردارهای تمام فریم‌های نمونه‌برداری شده را به هم متصل می‌کنند تا یک بردار جدید تشکیل دهند که نشان‌دهنده نتایج نمایه‌سازی کل دنباله ژست است؛ فاصله دو دنباله ژست بر این اساس را می‌توان به‌صورت رابطه ۶ محاسبه کرد [۳۵ و ۳۶ و ۳۸]:

به ترتیب تعداد توصیف‌گرهایی هستند که به گر  $i$ -امین در تصویر پرس و جو و در یک تصویر پایگاه داده کوانتیزه شده‌اند، فاصله بین تصویر پرس و جو و تصویر پایگاه داده می‌تواند به شکل رابطه ۳ تعریف شود [۳۷ و ۳۸]:

(۳)

$$a_i = n_i \omega_i, \quad b_i = m_i \omega_i$$

$$d(a, b) = \left\| \frac{a}{\|a\|} - \frac{b}{\|b\|} \right\|_p^p = \sum_i |a_i - b_i|^p$$

که در آن  $\omega_i$  وزن گر  $i$  امین درخت واژگان است که Lp-norm را نشان می‌دهد؛ در آزمایشات مقاله، از L1-norm استفاده شده است؛ وزن  $\omega_i$  را می‌توان بر اساس آنتروپی رابطه ۴ تعریف کرد [۳۸]:

$$\omega_i = \log \frac{N}{N_i} \quad (۴)$$

که در آن  $N$  تعداد تصاویر در پایگاه داده و  $N_i$  تعداد تصاویری است که دارای توصیف‌گرهای کوانتیزه شده به گر  $i$ -امین هستند که در تحلیل متن، به آن فرکانس سند معکوس می‌گویند؛ در عمل مشاهده می‌شود که گرهای برگ، حاوی بیشترین اطلاعات هستند و گاهی اوقات فقط گرهای برگ در تطبیق تصویر برای راحتی استفاده می‌شوند؛ معمولاً هزاران گر برگ در یک درخت واژگان وجود دارد که تنها صدها یا ده‌ها توصیف‌کننده در یک تصویر وجود دارد؛ در نتیجه،  $a$  و  $b$  هر دو بردارهای پراکنده هستند؛ بعد از اینکه  $a$  و  $b$  را عادی کردیم تا قدر واحد داشته باشیم، فاصله  $a$  و  $b$  در معادله ۳ را می‌توان به‌صورت رابطه ۵ بازنویسی کرد [۳۵ - ۳۸]:

(۵)

$$d(a, b) = 2 - \sum_{a_i \neq 0, b_i \neq 0} (|a_i - b_i|^p + |a_i|^p + |b_i|^p)$$

فقط فواصل بین عناصر غیر صفر بردارها محاسبه می‌شود؛ این به ما امکان می‌دهد از فایل‌های معکوس استفاده کنیم تا از تطبیق مستقیم بردار پرس و جو با تمام بردارهای تصویر در پایگاه داده جلوگیری کنیم؛ یک فایل معکوس برای هر گر، فایلی است که تعداد تصاویری را که حداقل یک توصیف‌گر به آن گر کوانتیزه شده است و شناسه‌های این تصاویر به همراه تعداد توصیف‌گرهای این تصاویر که به آن گر کوانتیزه می‌شوند، ثبت می‌کند؛ تحلیل متن را فرکانس اصطلاحی می‌نامند؛ اندازه‌گیری فاصله، شامل یک طرح وزن‌دهی

(۶)

$$d(a, b) = \sum_{i=1}^n d(a^i, b^i) = \sum_{i=1}^n \sum_j |a_j^i - b_j^i|^p$$

که در آن  $n$  تعداد کل فریم‌های نمونه‌برداری شده است؛ این پسوند، اگرچه ساده است، اما به دو دلیل زیر، در روش مذکور بسیار مهم است:

(۱) ژست‌ها را به بردارهایی با طول یکسان عادی می‌کند، که فاکتور مدت ژست را حذف می‌کند و تشخیص ژست‌ها را با تغییرات قابل توجهی در سرعت ممکن می‌سازد.

(۲) جهت جریان نوری به‌طور مداوم در برخی از حرکات (مانند کشیدن یک دایره) تغییر می‌کند؛ نتایج تجربی تایید کرده‌اند که با نمونه‌برداری بالاتر از نرخ بحرانی، تغییرات دینامیکی الگوهای حرکتی را می‌توان با موفقیت در بردارهای پیوسته حفظ کرد؛ پس از مراحل نمایه‌سازی، حرکات  $k$  نامزد برتر برگردانده می‌شوند؛ سپس موقعیت مناطق MSER برای تخمین مرکز هندسی دست برای هر قاب نمونه‌برداری شده استفاده می‌شود؛ مراکز موجود در پرس و جو و مراکز موجود در کاندیداها با هم مقایسه می‌شوند تا به هر داوطلب امتیازی اختصاص دهند؛ این یک مرحله طبیعی پس از تایید در جستجوی تصویر است؛ نمرات جمع‌آوری شده برای هر دسته اشاره را می‌توان از این امتیازات برتر به دست آورد و سپس حرکت پرس و جو به دسته‌ای که بالاترین امتیاز را دارد، اختصاص می‌یابد [۳۵] و [۳۶].

## ۹- آزمایش و پایگاه داده

• ژست‌های موجود در پایگاه داده، حرکات دو بعدی دست در یک پس‌زمینه دلخواه، با یک دوربین ثابت مستقیماً بالای دست هستند؛ سناریوهایی را شبیه‌سازی می‌کند که در آن کاربران در مقابل دوربینی که روی میز نشسته است، حرکات دست انجام می‌دهند که می‌تواند در تعاملات انسان با دستگاه‌های تلفن همراه اعمال شود؛ در مجموع ۱۰ حرکت دست پویا وجود دارد، از جمله: حرکت به راست، حرکت به چپ، چرخش به بالا، چرخش به پایین، حرکت به سمت پایین به راست، حرکت به سمت راست به پایین، دایره در جهت عقربه‌های ساعت، دایره خلاف جهت عقربه‌های ساعت، "Z" و "صلیب"، همان‌طور که نشان داده شده است. در شکل ۴ (ب). در فرآیند جمع‌آوری، از هر فرد خواسته می‌شود که این ده حرکت را با هفت وضعیت که در شکل ۴ (الف) نشان داده شده است، انجام دهند؛ شست، انگشت اشاره، دست - انگشتان دراز، "خوب" (حلقه انگشت شست و سبابه)، مشت، اشاره انگشت با چرخش ۹۰ درجه و دست با

انگشتان کشیده با چرخش ۹۰ درجه؛ هر موضوع ۷۰ نمونه اشاره به پایگاه داده ما کمک می‌کند [۳۵ - ۳۸]؛ ۱۰۵۰ ژست نمونه جمع‌آوری شده‌اند که توسط ۱۵ نفر انجام شده است؛ شکل ۴ (ج) چند نمونه توالی را ارائه می‌دهد؛ همان‌طور که قابل رویت است، پس‌زمینه و همچنین رنگ پوست دست‌ها بسیار متنوع است و سکانس‌های گرفته شده حاوی تارهای حرکتی شدید هستند؛ هر دو شرایط در کاربردهای واقعی مشترک هستند؛ این را یک پایگاه داده معرف در نظر می‌گیرند که نه تنها برای تشخیص حرکات دست پویا بلکه برای تخمین حالت استاتیک دست نیز مفید است. از آنجایی که روش مقاله مذکور به تشخیص حرکات دست پویا می‌پردازد، بر روی تشخیص ۱۰ حرکت پویا در آزمایشات خود تمرکز کرده‌اند؛ بنابراین نمونه‌هایی با عملکرد یکسان در حالت‌های مختلف دست به‌عنوان یک دسته در نظر گرفته می‌شوند و بر این اساس، هر دسته دارای ۱۰۵ نمونه است [۳۶ و ۳۸].

## ۱۰- تعیین پارامترها

• دو پارامتر اصلی در چارچوب مقاله وجود دارد که باید تعیین شوند: اندازه درخت واژگان (تعداد گره‌های برگ) و تعداد فریم‌های نمونه‌گیری شده در یک دنباله حرکتی برای نمایه‌سازی، همان‌طور که در بخش قبل توضیح داده شد؛ در مجموع ۲۲۵۷۸۵۱ توصیف‌گر جمع‌آوری شده‌اند و فاکتورها و سطوح مختلف شاخه را برای ساختن درخت واژگان انتخاب نموده‌اند؛ عملکرد تشخیص با اندازه‌های مختلف درخت در شکل ۵ نشان داده شده است؛ برخلاف مشاهدات [۱۵] مبنی بر اینکه یک درخت واژگان بزرگتر عملکرد را بهبود می‌بخشد، در آزمایشات مقاله مربوطه، نرخ تشخیص در حال حاضر با تنها ۵۱۲ گره برگ بسیار بالا است؛ عملکرد زمانی که اندازه کوچکتر از ۲۰۰۰ باشد ثابت می‌ماند و زمانی که اندازه بزرگتر است، کمی کاهش می‌یابد؛ با این حال، نرخ تشخیص هنوز بالای ۹۶٪ است که اندازه آن بیش از ۱۰۰۰۰ باشد؛ درخت واژگان با ۹ شاخه و ۳ سطح (۷۲۹ گره برگ) بالاترین میزان تشخیص را دارد که ۹۷٫۶۲٪ است؛ این درخت در آزمایشاتی استفاده شده است.

در مقاله سعی شده است توالی‌ها در تعداد متفاوتی از فریم‌های نمونه‌برداری شده، رمزگذاری شود و نتایج در شکل ۶ نشان داده شود؛ همان‌طور که توسط خط قرمز در شکل ۶ نشان داده شده است، عملکرد تشخیص در نرخ‌های نمونه‌گیری فریم‌های مختلف کاملاً قوی است؛ حتی اگر فقط از ۳ فریم نمونه‌برداری شود، نرخ تشخیص در حال حاضر نزدیک به ۹۶٪ رسیده است؛ هنگامی که نرخ نمونه‌برداری فریم بزرگ‌تر از ۷ باشد، عملکرد را نمی‌توان بیشتر بهبود بخشید؛ احتمالاً به این دلیل است که برای حرکات متمایز در پایگاه

روش آن‌ها در سناریوهای ما برای مقایسه منصفانه کار کند، هنگامی- که دنباله‌ای از هیستوگرام‌ها استخراج می‌شود، فریم‌ها نیز برای عادی‌سازی طول دنباله، همان‌طور که در بخش ۳.۳ پیشنهاد شده است، دوباره نمونه می‌شوند و فاصله  $x^2$  را برای هیستوگرام‌های پیوسته محاسبه می‌کنند [۳۵ و ۳۸]:

(۷)

$$d(a,b) = \frac{1}{2} \sum_i \frac{|a_i - b_i|^2}{a_i + b_i}$$

سپس نتیجه تشخیص توسط طبقه‌بندی‌کننده  $k$ -Nearest Neighbor تعیین می‌شود؛ بهترین نرخ تشخیص این روش ۹۴,۳۸٪ است که قدرت تمایز هیستوگرام‌های جریان نوری جهت‌دار را تایید می‌کند؛ با این حال، روش مذکور مقاله مبنا، قوی‌تر از هیستوگرام جهانی است؛ شکل ۷، نتایج تشخیص سه روش را با تعداد متفاوت نمونه‌های برتر برای طبقه‌بندی نشان می‌دهد؛ نرخ تشخیص روش مذکور، همیشه بالاتر از هیستوگرام جهانی است؛ علاوه بر این، روش هیستوگرام سراسری نسبت به نرخ نمونه‌گیری حساس‌تر است؛ همان‌طور که در شکل ۶، نشان داده شده است، عملکرد طبقه‌بندی هیستوگرام‌های سراسری، زمانی که نرخ نمونه‌برداری کمتر از ۷ باشد، به شدت کاهش می‌یابد؛ حرکات دست در همین حال، با بهره‌مندی از طرح نمایه-سازی، میانگین زمان تشخیص روش مذکور ۳۴,۵۳ میلی‌ثانیه است، که تنها حدود ۲۰٪ از زمان تشخیص تطابق هیستوگرام جهانی است، همان‌طور که در شکل ۸ (b) نشان داده شده است [۳۷ و ۳۸]؛ زمان تشخیص تطابق هیستوگرام از نظر تعداد حرکات در پایگاه داده خطی است؛ با توجه به پایگاه داده‌ای با بیش از ۱۰ هزار حرکت، شناسایی بلادرنگ توسط تطبیق نزدیک‌ترین همسایه در حال حاضر امکان‌پذیر نیست، در حالی- که روش مربوطه به راحتی برای پایگاه‌های داده بزرگ مقیاس-پذیر است؛ نمایه‌سازی مستقیم تصویر نیز نسبت به روش مقاله، زمان تشخیص بیشتری را می‌طلبد، که احتمالاً به این دلیل است که مناطق MSER بیشتری مستقیماً از تصاویر شناسایی می‌شوند تا از میدان‌های واگرایی جریان نوری آن‌ها. شکل ۹، ماتریس سردرگمی نتیجه تشخیص روش مقاله مربوطه را نشان می‌دهد؛ طبقه‌بندی اشتباه برخی از نمونه‌ها به دلیل شباهت برخی حرکات و همچنین ابهام زمانی است که کاربران این حرکات را انجام می‌دهند؛ به‌عنوان مثال، در شکل ۱۰، کاربر ژست Rotate Up را انجام می‌دهد؛ با این حال، چرخش دست

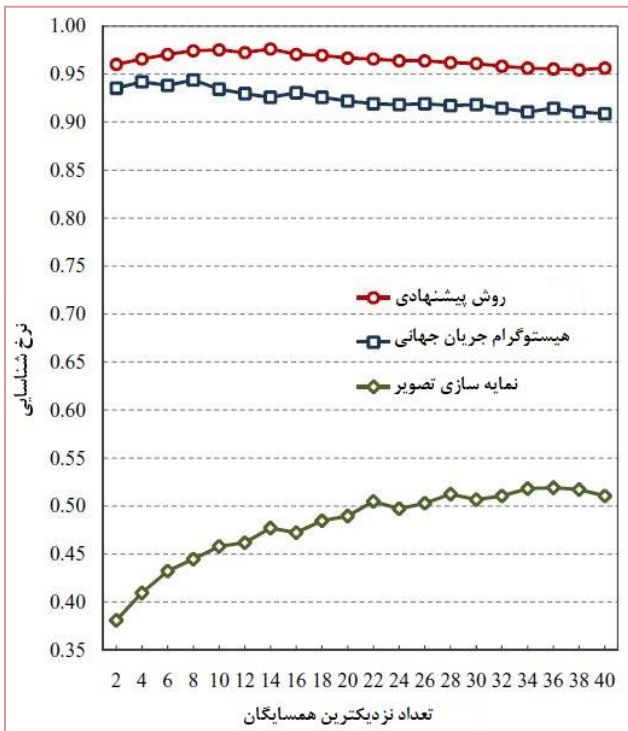
داده مربوطه، اطلاعات پویا از الگوهای حرکت قبلاً به‌طور کامل در ۷ فریم گرفته شده است [۳۶-۳۸].



شکل ۴- پایگاه داده جمع آوری شده، با ۱۰ دسته از حرکات و ۱۰۵۰ نمونه در مجموع [۳۸]

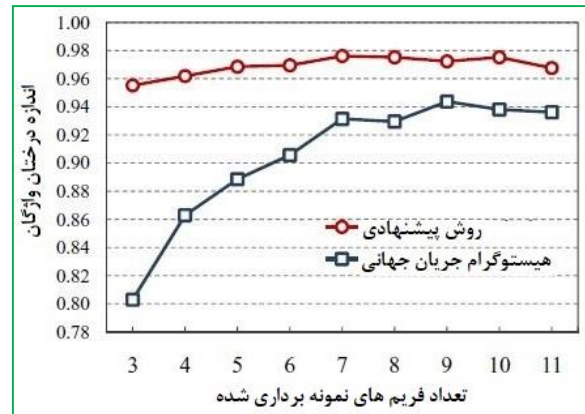
#### ۱۱- مقایسه‌ها

• روش مذکور، با دو روش دیگر مقایسه شده است؛ در ارزیابی برای هر سه روش، از اعتبارسنجی متقابل ترک یک خارج استفاده می‌شود؛ روش اول، یک روش پایه است که در آن تشخیص MSER مستقیماً روی دنباله‌های تصویر بدون تخمین جریان نوری انجام می‌شود؛ سپس توصیف‌گرهای تصویر HOG مانند با استفاده از روش در بخش ۳ نمونه‌برداری و نمایه‌سازی می‌شوند؛ در مقایسه با روش فعلی که الگوهای حرکتی را از میدان واگرایی نمایه می‌کند و مطابقت می‌دهد، این روش پایه، مستقیماً با الگوهای ظاهری دست مطابقت دارد؛ به نظر می‌رسد تغییرات ظاهری نیاز به یک مجموعه آموزشی بسیار بزرگ دارد و تطبیق تا حد زیادی تحت‌تأثیر ویژگی‌های استخراج شده از پس‌زمینه است؛ عملکرد این روش در مجموعه داده فعلی ضعیف است، به‌سختی بالای ۵۰٪، همان‌طور که در شکل ۷ و شکل ۸ (a) نشان داده شده است؛ روش دوم، روشی است که از Chaudhry و همکاران اقتباس شده است [۲، ۳۵ و ۳۶]؛ چودری و همکاران، در مقاله اصلی خود از یک هیستوگرام برای کل میدان جریان نوری جهت‌یافته استفاده کرده و سپس دینامیک هیستوگرام‌ها را برای تشخیص عملکرد دوره‌ای استخراج کردند؛ با این حال، از آنجایی‌که فقط دینامیک (تغییرات) در هیستوگرام‌ها استفاده می‌شود، نمی‌توان انتظار داشت که این رویکرد برخی از حرکات را در پایگاه داده متمایز کند (مانند حرکات ثابت چپ یا راست، که در آن هیستوگرام‌ها لزوماً تغییر نمی‌کنند)؛ برای اینکه

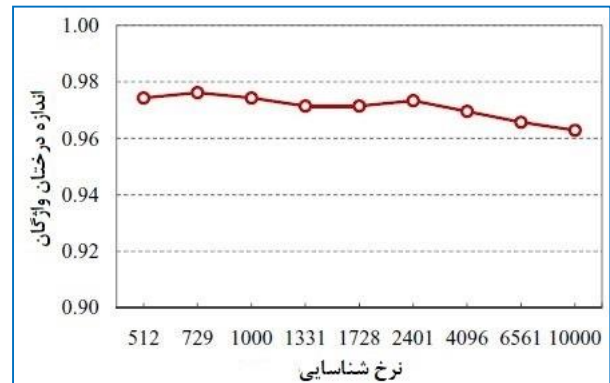


شکل ۷- اجرا با نامزدهای برتر  $k$  [۳۸]

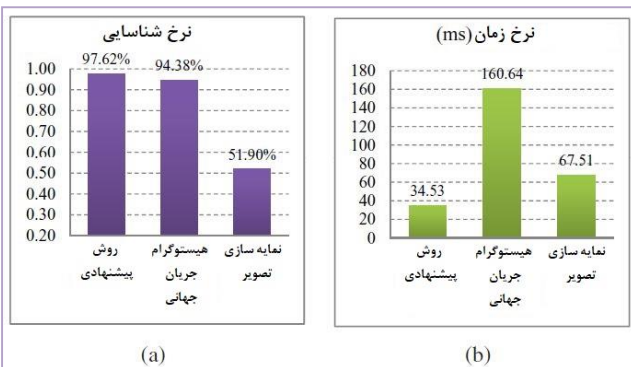
کافی نیست و روش مربوطه، ژست را به اشتباه به عنوان Move Right طبقه‌بندی می‌کند [۳۵، ۳۶ و ۳۸].



شکل ۵- نتایج تشخیص با اندازه درخت واژگان مختلف [۳۸]



شکل ۶- نتایج شناسایی با نرخ نمونه‌گیری متفاوت



شکل ۸- مقایسه سه روش در تشخیص عملکرد و زمان؛ روش مقاله مینا در کوتاه‌ترین زمان به بهترین عملکرد می‌رسد [۳۸]

	1	2	3	4	5	6	7	8	9	10
1	103	0	2	0	0	0	0	0	0	0
2	0	105	0	0	0	0	0	0	0	0
3	6	0	96	0	0	1	0	0	0	2
4	0	2	0	99	2	2	0	0	0	0
5	0	0	0	0	105	0	0	0	0	0
6	0	0	0	0	0	104	1	0	0	0
7	0	0	0	0	0	0	103	0	1	1
8	0	0	0	0	1	0	0	104	0	0
9	0	0	0	0	1	0	1	0	103	0
10	0	0	1	0	0	0	1	0	0	103

شکل ۹- ماتریس سردرگمی نتیجه تشخیص مقاله؛ هر دسته شامل ۱۰۵ نمونه است؛ نرخ شناخت کلی ۹۷،۶۲٪ است.

## ۱۲- برنامه‌های کاربردی

• در مقاله مینا، یک سیستم نمونه ساخته شد و روش خود برای تشخیص حرکات دست زنده به کار برده شد؛ شکل ۱۱ تصویری از این سیستم را نشان می‌دهد؛ تشخیص اشاره، یعنی تشخیص خودکار شروع و پایان ژست‌ها، توسط یک آستانه حرکت جهانی در سیستم پیاده‌سازی می‌شود؛ این سیستم می‌تواند حرکات دست زنده را در زمان واقعی (۳۰ فریم در ثانیه) برای دنباله‌های ویدیویی  $320 \times 240$  با کیفیت تشخیص بالا تشخیص دهد [۳۶ و ۳۵].

## ۱۳- نتیجه‌گیری و بحث‌های آینده



چارچوب پیشنهادی مقاله، تا یک پایگاه داده بزرگ مقیاس‌پذیر است و به‌طور همزمان به دقت تشخیص بالایی دست می‌یابد؛ محققان بر این باورند که رویکرد پیشنهادی برای وظایف عمومی‌تر تشخیص فعالیت نیز قابل استفاده است و شایستگی مطالعه بیشتر را دارد. در مقالات آتی تلاش می‌شود که پژوهش‌های بیشتری در این زمینه انجام شود و تا حد امکان نیز نتایج، بر مبنای تحلیل‌های برگرفته از نرم‌افزارها باشد.

۱۲- B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In IJCAI, pages 674-679, 1981.

۱۳- J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In BMVC, 2002.

۱۴- S. Mitra and T. Acharya. Gesture recognition: A survey. IEEE Trans. on Systems, Man and Cybernetics - Part C, 37(3): 311-324, 2007.

۱۵- D. Nistér and H. Stewénius. Scalable recognition with a vocabulary tree. In CVPR, pages 2006, 2161-2168.

۱۶- D. Nistér and H. Stewénius. Linear time maximally stable extremal regions. In ECCV, pages 2008, 183-196.

۱۷- R. Lockton and A. W. Fitzgibbon. Real-time gesture recognition using deterministic boosting. In BMVC, 2002.

۱۸- C. Schödl, I. Laptev, and B. Caputo. Recognizing human actions: A local svm approach. In ICPR, 2004.

۱۹- S. Rajko, G. Qian, T. Ingalls, and J. James. Real-time gesture recognition with minimal training requirements and on-line learning. In CVPR, 2007.

۲۰- H. Suk, B. Sin, and S. Lee. Recognizing hand gestures using dynamic bayesian network. In FG, 2008.

۲۱- S. Wang, A. Quattoni, L. P. Morency, D. Demirdjian, and T. Darrell. Hidden conditional random fields for gesture recognition. In CVPR, 2006.

۲۲- J. Yamato, J. Ohya, and K. Ishii. Recognizing human action in time-sequential images using hidden markov model. In CVPR, 1992.

۲۳- M. H. Yang, N. Ahuja, and M. Tabb. Extraction of 2d motion trajectories and its application to hand gesture recognition. IEEE Trans. on PAMI, 24(10): 1061-1074, 2002.

۲۴- Alpern, M., & Minardo, K. 2003. Developing a car gesture interface for use as a secondary task. In CHI 2003 Extended Abstracts on Human Factors in

در مقاله مبنای، یک روش جدید برای تشخیص حرکات دست پویا ارائه شد که در آن یک دنباله ژست دست به دنباله‌ای از زمینه‌های واگرایی حرکتی تبدیل می‌شود؛ استخراج میدان واگرایی جریان نوری، الگوهای حرکتی را به الگوهای تصویر فضایی متمایز تبدیل می‌کند، که سپس با استفاده از آشکارساز MSER استخراج می‌شوند و توسط یک درخت واژگان آموزش‌دیده نمایه‌سازی می‌شوند؛ یک طرح TF-IDF برای تطبیق یک حرکت پرس و جو با توالی‌های پایگاه داده نمایه شده و برای تولید نتیجه تشخیص نهایی استفاده می‌شود؛

## ۱۴ - مراجع

۱- L. Bretzner, I. Laptev, and T. Lindeberg. Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering. In FG, pages 423-428, 2002.

۲- R. Chaudhry, A. Ravichandran, G. Hager, and R. Vidal. Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions. In CVPR, pages 2009, 1932-1939.

۳- J. Davis and M. Shah. Recognizing hand gestures. In ECCV, 1994.

۴- P. Dollár, V. Rabaud, G. Cottrell, and S. Belongie. Behavior recognition via sparse spatio-temporal features. In ICCV VSPETS, 2005.

۵- A. Elgammal, V. Shet, Y. Yacoob, and L. Davis. Learning dynamics for exemplar-based gesture recognition. In CVPR, 2002.

۶- C. L. H. F. S. Chen, C. M. Fu. Hand gesture recognition using a real-time tracking method and hidden markov models. Image and Vision Computing, 21(7): 45-58, 2003.

۷- F. Flórez, J. M. García, J. García, and A. Hernández. Hand gesture recognition following the dynamics of a topology-preserving network. In FG, 2002.

۸- P. Forssen and D. Lowe. Shape descriptors for maximally stable extremal regions. In ICCV, 2007.

۹- W. T. Freeman and M. Roth. Orientation histograms for hand gesture recognition. In FG, 1995.

۱۰- P. Hong, M. Turk, and T. S. Huang. Gesture modeling and recognition using finite state machines. In FG, 2000.

۱۱- T. Kirishima, K. Sato, and K. Chihara. Real-time gesture recognition by learning and selective control of visual interest points. IEEE Trans. on PAMI, 27(3): 351-364, 2005.

Gesture Set for Human-Computer Interaction, Human Factors and Ergonomics Society, Vol. 57, No. 4, June 2015, pp. 607- 621.

37- Hossein Mousavi Hondori, Maryam Khademi, Lucy Dodakian, Alison McKenzie, Cristina V. Lopes, Steven C. Cramer (2015), Choice of Human-Computer Interaction Mode in Stroke Rehabilitation, Neurorehabilitation and Neural Repair, pp 1-8.

38- E.A. Onibere, S. Morgan, E.M. Busang, D. Mpoeleng (2001), Human computer interface design issues for a multi-cultural and multi lingual English speaking country Botswana, Interacting with Computers 13 (2001), pp 497-512.

39- Xiaohui Shen, Gang Hua, Lance Williams, Ying Wu (2011), Motion Divergence Fields for Dynamic Hand Gesture Recognition, 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG).

Computing Systems (pp. ۹۳۲-۹۳۳ New York, NY: .ACM

Bao, S., Howard, N., Spielholz, P., Silverstein, B-۲۰ Polissar, N. Interrater reliability of posture & .observations. Human Factors, ۵۱' ۲۹۲-۳۰۹' ۲۰۰۹

Baudel, T., & Beaudouin-Lafon, M. Charade: -۲۶ Remote control of objects using freehand gestures. .Communications of the ACM, ۳۶' ۲۸-۳۵' ۱۹۹۳

Bowman, D., & Wingrave, C. (۲۰۰۱) / Design and -۲۷ evaluation of menus systems for immersive virtual environments. In Proceedings of IEEE Virtual Reality .pp ۱۴۹-۱۵۶ Piscataway, NJ: IEEE, ۲۰۰۱

de la Barré, R., Chojecki, P., Leiner, U., -۲۸ Mühlbach, L., & Ruschin, D. Touchless interaction: Novel chances and challenges. In M. Kurosu (Ed.), Human-computer interaction: Novel interaction methods and techniques pp. ۱۶۱-۱۶۹ Berlin, .Germany: Springer., , ۲۰۰۹

Farhadi-Niaki, F., Etemad, S. A., & Arya, A. -۲۹ Design and usability analysis of gesture-based control for common desktop tasks. In M. Kurosu Ed. Human-computer interaction: Interaction modalities and techniques pp. ۲۱۵-۲۲۴ Berlin, Germany: Springer, .۲۰۱۳

Feuerstein, M., Carosella, A. M., Burrell, L. M., -۳۰ Marshall, L., & Decaro, J. Occupational upper extremity symptoms in sign language interpreters: Prevalence and correlates of pain, function, and work disability. Journal of Occupational Rehabilitation, .۱۹۹۷, ۷۱۸۷-۲۰۵

Fikkert, W., Vet, P. van der, Veer, G. van der, & -۳۱ Nijholt, A. Gestures for large display control. In S. Kopp & I. Wachsmuth Eds. Lecture notes in computer science: Vol . ۵۹۳۴ / Gesture in embodied communication and human-computer interaction (pp. .Berlin, Germany: Springer, ۲۴۵-۲۵۶' ۲۰۱۰

Guiard, Y. Asymmetric division of labor in human -۳۲ skilled bimanual action: The kinematic chain as a model. Journal of Motor Behavior, , ۱۹' ۴۸۶-۵۱۷' .۱۹۸۷

Guiard, Y., & Ferrand, T. Asymmetry in bimanual -۳۳ skills. In D. Elliott & E. A. Roy Eds, Manual asymmetries in motor performance pp-, ۱۷۵-۱۹۵' ., Boca Raton, FL: CRC Press ۱۹۹۶

Hignett, S., & McAtamney, L. Rapid entire body -۳۴ assessment (REBA). Applied Ergonomics, ۳۱' ۲۰۱- .۲۰۰۰, ۲۰۵

Horsky, J., Kaufman, D. R., Oppenheim, M. I. & -۳۵ Patel, V. L. A framework for analyzing the cognitive complexity of computer-assisted clinical ordering. .Journal of Biomedical Informatics, ۳۶' ۴-۲۲' ۲۰۰۳

-۳۶ Anna Pereira, Juan P. Wachs, Kunwoo Park and David Rempel, (2015), A User-Developed 3-D Hand