

بازشناسی حالات ایستای دست با مدل مارکف مخفی بر اساس ویژگی شیب انحنای کانتور پیرامونی

نسبیه علیجانپور شلمانی^(۱) - حسین ابراهیم نژاد^(۲) - افشین ابراهیمی^(۲)

(۱) کارشناس ارشد - دانشکده مهندسی برق، دانشگاه صنعتی سهند تبریز

(۲) استادیار - دانشکده مهندسی برق، دانشگاه صنعتی سهند تبریز

تاریخ دریافت: بهار ۱۳۹۰ تاریخ پذیرش: بهار ۱۳۹۱

خلاصه: امروزه ارتباط انسان با رایانه از طریق صفحه کلید، ماوس و ... امکان پذیر می‌باشد. این گونه وسایل با محدودیت‌هایی از جمله سرعت عملکرد مواجه هستند. هدف نهایی دنیای فنآوری این است که تعامل بین انسان با کامپیوتر مانند تعامل بین انسانها با هم، طبیعی و آسان باشد. در این مقاله روشی برای بازشناسی حالات دست مبتنی بر ویژگی شیب انحنای منحنی Bspline دست ارائه می‌گردد. بدین ترتیب که ابتدا تصویر دست در فریم‌های مختلف استخراج شده و تعدادی نقطه روی کانتور پیرامونی دست به صورت متساوی الفاصله انتخاب می‌شود. سپس، منحنی Bspline مرتبه ۳ برای دسته‌های ۴ تایی از این نقاط محاسبه می‌گردد. در ادامه شیب انحنای منحنی‌های Bspline محاسبه و به عنوان بردار ویژگی برای طبقه بندی به مدل مخفی مارکف داده می‌شود. روش پیشنهادی با توجه به استفاده از ویژگی شیب انحنای نسبت به چرخش، انتقال و تغییر اندازه حساس نمی‌باشد. نتایج آزمایش روی ۱۵ مجموعه از دنباله‌های ویدیویی با تعداد فریم‌های متوسط ۱۵۰ فریم، نرخ بازشناسی 92/21% را به دست می‌دهد.

کلمات کلیدی: منحنی Bspline، مدل مخفی مارکف، استخراج ویژگی، بازشناسی دست.

۱- مقدمه

حرکات دست کار می‌شود و یک توالی از تصاویر حرکات دست در فریم‌های مختلف مورد بررسی قرار می‌گیرد (مثل حالت سلام یا خداحافظی به وسیله تکان دادن دست). در سالهای اخیر تحقیقات زیادی روی بازشناسی حرکات و حالات دست برای دستیابی به اهداف متفاوت انجام شده است. در برخی از روش‌ها نشانگرهایی روی دست نصب شده و در تصویر با مشاهده مکان نشانگر، مکان دست شناسایی شده است [۱]. در بسیاری از مقالات هم از نشانگرهای رنگی برای استخراج نوک انگشتان و مکان اتصالات یا برخی نقاط روی کف دست استفاده شده است. اما یکی از دشواریهای کار کردن با نشانگرها حساس بودن آنها به مانع‌ها است. ردیابی نشانگرها به دلیل وجود موانع ممکن است با مشکل مواجه شود. در مقاله [۲] مروری بر سیستم‌های بازشناسی دست که تا سال ۲۰۰۶ مطرح شده‌اند، انجام گرفته است. در حالت کلی، در سیستم‌های بازشناسی

انسانها از طریق بینایی و صدا با هم ارتباط برقرار می‌کنند. مزیت استفاده از این دو روش برای برقراری ارتباط انسان و رایانه، این است که ارتباط از طریق بینایی و صدا به فاصله کاربر از رایانه وابسته نمی‌باشد. در شرایط نویزی ارتباط بینایی نسبت به صوت ارجحیت دارد. می‌خواهیم سیستم‌های مبتنی بر بینایی ماشین را به نحوی گسترش دهیم که حالات و حرکات کاربر را بفهمد و تفسیر نماید. یکی از اهداف روش‌های مبتنی بر بینایی ماشین بازشناسی بر خط حالات و حرکات کاربر است. در حالت کلی حالات و حرکات دست به دو دسته حالات ایستا و پویا تقسیم می‌شوند. در بررسی حالت ایستای دست ساختار کلی دست مانند حالت دست با انگشتان کاملاً باز، حالت دست مشت کرده یا حالتی که فقط انگشت اشاره فرد فعال باشد، بررسی می‌شوند، یا به عبارتی حالت تصویر دست بازشناسی می‌شود. در بررسی حالت پویا، با

حالات دست، ویژگی‌ها را می‌توان به سه دسته ویژگی‌های سطح بالا، ویژگی‌های سطح پایین و ویژگی‌های سه بعدی تقسیم نمود. کانتور و لبه ویژگی‌های سطح پایینی هستند که در هر روش مبتنی بر مدل مورد استفاده قرار می‌گیرند. روش‌های مبتنی بر لبه در پس زمینه‌های پیچیده و به هم ریخته مؤثر نخواهند بود. در مقاله [۳] ترکیب جهت لبه و تطابق چامفر^۱ در مقابل ویژگی‌های متنی شکل مؤثرتر گزارش شده است. ترکیب ویژگی‌های چندگانه مثل لبه‌ها، جریان نوری و سایه در مرجع [۴] بررسی شده است. در [۵] و [۶]، روشی برای استخراج ویژگی از دست، تحت عنوان فضای مقیاس منحنی (CSS) مطرح شده است. روش CSS سه مرحله دارد. بخش‌بندی دست، استخراج ویژگی و بازشناسی. در [۷]، یک سیستم برای شناسایی حالات ایستای زبان اشاره تایوانی با استفاده از داده‌های سه بعدی و شبکه عصبی ارائه شده است. در این تحقیق، ۶ نشانگر روی کف دست و نوک انگشتان قرار داده شده است. فاصله تک‌تک این نشانگرها از هم ۱۵ ویژگی را به دست می‌دهد که این ویژگی‌ها برای بازشناسی حالت دست با شبکه عصبی BPNN استفاده شده است. در [۸] الگوریتمی برای شناسایی تعداد انگشتان دست برای کنترل ربات ارائه شده است. در این مقاله ابتدا با استفاده از یک شبکه عصبی ROC و با استفاده از اطلاعات رنگ پوست، دست را از پس زمینه جدا می‌سازند. سپس تصویر را باینری کرده و ویژگی‌هایی را با الگوریتم شمارش تعداد انگشتان، استخراج می‌کنند. "مفهوم شکل" یک روش مبتنی بر کانتور است [۹ و ۱۰]. ایده اصلی این روش بر این اساس است که هر نقطه در امتداد کانتور می‌تواند با یک هیستوگرام قطبی لگاریتمی از نقاط دیگر کانتور توصیف گردد. شباهت بین نقاط می‌تواند با این عنوان که شکل‌های با "مفهوم شکل" مشابه به هم شبیه‌ترند توجیه گردد. در [۱۱] ضرایب گشتاور HU برای توصیف شکل مورد استفاده قرار گرفته است. این گشتاور تابعی از گشتاور تصویر مرکزی است. این ویژگی در مقابل انتقال و تغییر اندازه و چرخش حساس نمی‌باشد. در [۱۲] یک رهیافت مبتنی بر خطی سازی محلی برای بازشناسی دست مطرح می‌شود. در این مقاله تصویر دست را به فضای YUV می‌برند. الگوریتم embedding local linear ارائه شده در این مقاله، نقاط را به بعد کمتر می‌برد در حالی که ارتباطات بین نقاط همجوار را حفظ می‌کند. برای تشخیص حالت از مینیمم سازی یک تابع هزینه بهره گرفته شده است. در [۱۳] از بازشناسی حرکات و حالات دست برای بازشناسی زبان اشاره استفاده شده است که در آن از تبدیل هاف برای استخراج ویژگی از دست استفاده می‌گردد. ویژگی‌های استخراج شده، به شبکه عصبی انتشار بازگشتی اعمال شده و با استفاده از این شبکه، حالات دست طبقه‌بندی و شناسایی می‌گردد. در مقاله [۱۴] از حالات دست برای کنترل و فرمان دادن به صندلی چرخدار استفاده شده است. با استفاده از ترکیب فیلتر ذره‌ای^۲ و شیفت میانگین روش مؤثری برای ردیابی دست مطرح شده است. در انتهای کار نیز روشی برای بازشناسی حرکت دست مطرح شده است. این روش MSEPF نام‌گذاری شده است. در [۱۵] از شبکه عصبی و مدل مخفی مارکوف (HMM) برای

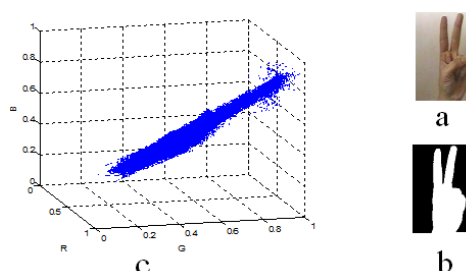
بازشناسی حالات و حرکات دست استفاده شده است. دو روش برای جداسازی مچ در این مقاله ارائه شده است. پس از جداسازی مچ، از تحلیل فوریه برای استخراج ویژگی از دست استفاده شده و از شبکه عصبی شعاعی برای بازشناسی استفاده می‌گردد. در تحقیق دیگری، ردیابی حالت دست از یک حالت از زبان اشاره به حالت دیگر با بهره‌گیری از ۳۰ مولفه اول ماتریس کواریانس و اعمال الگوریتم تکراری کمترین مربعات انجام شده است [۱۶]. در [۱۷] و [۱۸]، از بازشناسی حالات دست برای کنترل بازیهای رایانه‌ای استفاده شده است. در این مقاله دو مفهوم Convexity Hull و Convexity Defect برای استخراج ویژگی از دست استفاده می‌گردد. در [۱۹] بازشناسی دست با معیار فاصله Hausdorff (HD) و یک حافظه تصویری برای ذخیره‌سازی تصاویر، انجام می‌شود. در [۲۰] از تبدیل فوریه برای استخراج ویژگی استفاده شده است و بسیاری از مطالعات در سالهای اخیر بر اساس ویژگی‌هایی مانند اسکلت ساختاری تصاویر [۲۱] و [۲۲] و ویژگی‌های موجود در کانتور و منحنی پیرامونی شکل [۲۳] و [۲۴] و [۲۵] به بازشناسی تصویر می‌پردازند. در [۲۶-۲۹] روشهایی برای بازشناسی دست و اشکال مختلف توسط نویسندگان مقاله ارائه شده است. در [۲۶] پس از بخش بندی دست با استفاده از فضای رنگی و ردیابی آن با روش مرکز ثقل، از فاصله درونی به عنوان معیاری جهت بازشناسی حالت دست استفاده شده است. در [۲۷] از ویژگی‌های استخراج شده از دایره‌های قطع‌کننده جسم با مرکزیت واحد جهت شناسایی اشکال مختلف استفاده شده است. در [۲۸] یک روش ترکیبی با بهره‌گیری از ویژگی هندسی و ویژگی مفهوم شکلی برای بازشناسی حالت دست ارائه شده است. در [۲۹] نیز از ترکیب ویژگی‌های انگشتان فعال، مساحت تفاضلی و تبدیل هاف برای بازشناسی حالات دست استفاده شده است.

در این مقاله روشی برای بازشناسی حالات دست پیشنهاد می‌گردد که در آن از تغییرات انحنای کانتور پیرامونی دست استفاده می‌گردد. پس از استخراج تصویر دست در فریم‌های مختلف، منحنی پیرامونی دست را به قطعاتی تقسیم نموده و هر قطعه را با استفاده از منحنی BSpline مرتبه ۳ تقریب می‌زنیم. ویژگی تغییرات یا شیب انحنای برای این منحنی‌ها محاسبه شده و به عنوان بردار ویژگی برای طبقه‌بندی به HMM داده می‌شود.

ساختار این مقاله به این ترتیب است: در بخش (۲) جداسازی دست در فریم‌های ویدیویی مورد بررسی قرار می‌گیرد. در بخش (۳) به بیان نحوه استخراج ویژگی با استفاده از منحنی BSpline و محاسبه شیب انحنای پیرامونی می‌پردازیم. در بخش (۴) تطابق با استفاده از مدل‌های مخفی مارکوف را بررسی خواهیم نمود. در بخش (۵) نتایج حاصل از اعمال روش پیشنهادی روی دنباله‌های ویدیویی مختلف را نشان خواهیم داد و در بخش (۶) نیز جمع بندی از مقاله ارائه می‌شود.

۲- جداسازی دست

تصاویر استفاده شده در این مقاله شامل 2105 تصویر از حالات مختلف دست ۱۵ کاربر می باشد که در قالب فریم های ویدیویی و توسط نویسندگان مقاله ایجاد شده است. مزیت استفاده از این نوع داده، قابلیت الگوریتم پیشنهادی در بازشناسی حالت دست در دنباله های ویدیویی است. برای این منظور در صورتی که تنها شیء متحرک در تصویر، دست کاربر باشد، می توان با استفاده از تفاضل تصاویر، آن را بازشناسی نمود. به این ترتیب که تصویر ناحیه موجود در زاویه دید دوربین را هنگامی که دست در مقابل دوربین قرار نگرفته است، به عنوان پس زمینه در نظر بگیریم.



شکل (۱): a) دست در فضای RGB، b) پیکسل های موجود در ناحیه $1.2 < R/G < 1.6$ ، c) موقعیت R, G, B پیکسل ها

سپس این تصویر را از تصویر دوربین در حالی که دست در مقابل دوربین قرار گرفته، تفریق نمائیم. تصویر حاصل شامل دست کاربر و اشیای متحرک دیگر نسبت به پس زمینه خواهد بود. ایراد روش بالا این است که ممکن است غیر از دست کاربر اشیای متحرک دیگری هم در تصویر وجود داشته باشند که در پس زمینه موجود نیستند. علاوه بر این، دوربین باید در مکان ثابتی قرار داشته باشد تا پس زمینه در تمام مراحل تصویر برداری کاملاً ثابت و بدون تغییر باشد. ما در این مقاله از شیوه تفاضل تصویر و پس زمینه، به دلایل ایرادهایی که در بالا ذکر شد استفاده نکرده ایم. مقالات [۴] و [۳۰] و [۳۱] و [۳۲] هر یک به نحوی با تکیه بر اطلاعات رنگ پوست، پس زمینه را حذف می کنند یا دست را در کل تصویر با رنگ سفید و بقیه تصویر را با رنگ سیاه نشان می دهند. شیوه آشکارسازی دست در این مقاله مبتنی بر ویژگی رنگ پیکسل های پوست انسان می باشد. به این ترتیب که ما بیش از ۵۰ سکانس ویدیویی تهیه شده از حالات مختلف دست چپ و راست ۱۵ فرد متفاوت را مورد آزمایش قرار دادیم و به این نتیجه رسیدیم که پیکسل های هم رنگ با پوست در تمام تصاویر در ناحیه مشخصی قرار می گیرند. (شکل (۱)). با دقت روی این شکل و شکل های مشابه برای تصاویر مختلف، مشاهده می شود که نسبت R/G در پیکسل های رنگ پوست در محدوده های $1.2 < R/G < 1.6$ قرار گرفته است. محدود کردن R/G در این نواحی به طور قابل قبولی پوست را از دیگر اشیای موجود در تصویر متمایز می سازد. مشکلی که هنوز وجود دارد

این است که در صورتی که قسمتهای دیگری از بدن کاربر (مثلاً صورت) در تصویر قرار داشته باشند نیز آشکارسازی خواهند شد. برای رفع این مشکل فرض می کنیم که دست کاربر در فاصله نزدیک تری از دوربین نسبت به صورت او قرار دارد. لذا بخش آشکار شده مربوط به دست نسبت به بخش صورت بزرگتر می باشد. در نتیجه بزرگترین بخش را متناظر با دست کاربر فرض نموده و به عنوان ورودی به مرحله بعد وارد می کنیم.

۳- استخراج ویژگی

روشی که ما در این مرحله برای استخراج ویژگی از شکل ها مطرح می نماییم بر مبنای منحنی های BSpline قرار دارد [۳۳] و [۳۴]. این منحنی ها که برای اولین بار در سال ۱۹۴۰ مطرح شدند، برای رسم منحنی با استفاده از تعدادی نقطه کنترلی مورد استفاده قرار می گیرند. روش پیشنهادی ما در این مقاله بر برازش منحنی های BSpline به کانتور پیرامونی دست استوار می باشد. استفاده از این ویژگی می تواند در هموارسازی لبه بسیار مؤثر بوده و حساسیت لبه را به نویز کاهش دهد. منحنی BSpline مورد استفاده از درجه سه می باشد و با عنوان Cubic BSpline شناخته می شود. برای محاسبه این منحنی ها باید ۴ نقطه کنترلی را در اختیار داشته باشیم. این نقاط که با $P_{i-1}, P_i, P_{i+1}, P_{i+2}$ نشان داده می شود، با استفاده از رابطه زیر در ترسیم منحنی تأثیر گذار خواهند بود.

$$P_i(t) = (t^3, t^2, t, 1)M \begin{pmatrix} P_{i-1} \\ P_i \\ P_{i+1} \\ P_{i+2} \end{pmatrix} \quad (1)$$

$$M = \begin{pmatrix} a_3 & b_3 & c_3 & d_3 \\ a_2 & b_2 & c_2 & d_2 \\ a_1 & b_1 & c_1 & d_1 \\ a_0 & b_0 & c_0 & d_0 \end{pmatrix}$$

که در آن $t \in [0,1]$ بوده و M یک ماتریس 4×4 با 16 مولفه مجهول می باشد. برای محاسبه این 16 مولفه مجهول ماتریس M باید 16 معادله مستقل از هم در اختیار داشته باشیم. لذا برای دست یابی به این هدف سه شرط زیر را خواهیم پذیرفت.

$$P_1(1) = P_2(0), \quad P_1'(1) = P_2'(0), \quad P_1''(1) = P_2''(0) \quad (2)$$

با استفاده از این سه شرط و با استفاده از روابط زیر ماتریس M را به دست می آوریم:

$$\begin{aligned} a_3 + a_2 + a_1 + a_0 &= 0, & b_3 + b_2 + b_1 + b_0 &= 0 \\ c_3 + c_2 + c_1 + c_0 &= 0, & d_3 + d_2 + d_1 + d_0 &= 0 \\ 3a_3 + 2a_2 + a_1 &= 0, & 3b_3 + 2b_2 + b_1 &= a_1 \\ 3c_3 + 2c_2 + c_1 &= b_1, & 3d_3 + 2d_2 + d_1 &= c_1 \\ 6a_3 + 2a_2 &= 0, & 6b_3 + 2b_2 &= 2a_2 \\ 6c_3 + 2c_2 &= 2b_2, & 6d_3 + 2d_2 &= 2c_2 \\ 0 &= d_0, & 0 &= d_1, & 0 &= 2d_2 \end{aligned} \quad (3)$$

معادله دیگری که می تواند برای استقلال داشتن روابط مورد استفاده قرار گیرد، به صورت زیر می باشد:

$$a(t) + b(t) + c(t) + d(t) \equiv 1 \quad (4)$$

$$\alpha = \arccos\left(\frac{y_f - y_{COG}}{\sqrt{(x_f - x_{COG})^2 + (y_f - y_{COG})^2}}\right) \quad (11)$$

همان طور که گفته شد با چرخش تصویر به اندازه این زاویه، اثر نامطلوب چرخش حذف شده و نقطه شروع منحنی B-Spline نیز در تمام تصاویر ثابت و مشخص می‌گردد.

پس از چرخش تصویر، n نقطه متساوی الفاصله روی کانتور شکل در نظر می‌گیریم. این نقاط را به طور متوالی به دسته‌های سه‌تایی تقسیم می‌نماییم. هر سه نقطه موجود در یک گروه و نقطه آغازین دسته مجاور را به عنوان ۴ نقطه کنترلی در نظر گرفته و منحنی B-Spline را برای آنها با استفاده از روابط (۱) - (۷) محاسبه می‌نماییم. اگر تعداد منحنی‌های محاسبه شده را با M نشان دهیم، این تعداد با رابطه زیر محاسبه می‌شود.

$$M = \frac{(n-1)}{3} \quad (12)$$

به عبارتی اگر تعداد نقاط نمونه انتخاب شده روی منحنی پیرامونی شکل برابر با $n=100$ باشد، تعداد منحنی‌های به دست آمده از طریق BSpline برابر با ۳۳ منحنی می‌باشد. هر منحنی محاسبه شده را در داخل یک مستطیل d به ابعاد (d_{in}, d_{hi}) محصور می‌نماییم. نقطه مرکزی این مستطیل $c = \frac{1}{2}(d_{in}, d_{hi})$ خواهد بود. لذا برای هر تصویر M نقطه مرکزی مربوط به منحنی‌های B-Spline آن تصویر را خواهیم داشت. در مرحله بعد که تطابق دو شکل مد نظر می‌باشد، تطابق بین این M نقطه از شکل‌ها را با هم مقایسه می‌نماییم. فاصله نرمالیزه نقطه مرکزی هر منحنی تا مرکز ثقل (COG) از رابطه زیر به دست خواهد آمد.

$$\text{NormDist} = \frac{\text{Dist}(c_i, \text{COG})}{\text{Dist}_{\max}} \quad (13)$$

که در آن $i = 1, 2, \dots, M$ می‌باشد. این فاصله را به عنوان یک ویژگی برای هر کدام از نقاط مرکزی منحنی‌ها در نظر خواهیم گرفت. می‌توان ضرایب چندجمله‌ای را که از رابطه (۷) به دست آمد به عنوان ویژگی‌هایی برای هر قطعه منحنی در نظر گرفت. همان طور که گفته شد با استفاده از چرخش اولیه نقاط شروع منحنی‌ها را بر هم منطبق نمودیم، اما معادله منحنی به زاویه محور میانی منحنی بسیار حساس می‌باشد. لذا اگر اندکی در چرخش اولیه دچار اشتباه شویم معادلات منحنی‌ها و ضرایب چند جمله‌ای‌ها بسیار تغییر خواهد کرد. لذا به ویژگی محکم‌تری در مقابل چرخش و یا انتقال برای بازشناسی شکل نیاز خواهیم داشت. فرض کنیم که منحنی را بچرخانیم و یا آن را انتقال داده و یا اندازه آن را به یک اندازه کوچک یا بزرگ نماییم. بر اثر چرخش ضرایب چند جمله‌ای تغییر خواهد نمود و تغییر اندازه و انتقال ضریب ثابت چند جمله‌ای را تغییر خواهد داد. در این میان پارامتری از منحنی که همواره ثابت می‌ماند شیب انحنا می‌باشد. این ویژگی غیر حساس به انتقال و چرخش و تغییر اندازه بوده و می‌تواند ویژگی مناسبی برای بازشناسی باشد. اگر هر قطعه منحنی را با $P(t)$ نشان

از این معادله، روابط زیر را می‌توان استخراج نمود:

$$a_3 + b_3 + c_3 + d_3 = 0, \quad a_2 + b_2 + c_2 + d_2 = 0 \quad (5)$$

$$a_1 + b_1 + c_1 + d_1 = 0, \quad a_0 + b_0 + c_0 + d_0 = 0$$

با بهره‌گیری از روابط بالا معادله منحنی درجه ۳ مربوط به ۴ نقطه کنترلی مذکور از رابطه زیر به دست خواهد آمد:

$$P_i(t) = \frac{1}{6}(-t^3 + 3t^2 - 3t + 1)P_{i-1} + \frac{1}{6}(3t^3 - 6t^2 + 4) \quad (6)$$

$$P_i + \frac{1}{6}(-3t^3 + 3t^2 + 3t + 1)P_{i+1} + \frac{t^3}{6}P_{i+2}$$

که می‌توان آنرا به صورت زیر بازنویسی نمود:

$$P_i(t) = \left(-\frac{P_{i-1}}{6} + \frac{P_i}{2} - \frac{P_{i+1}}{2} + \frac{P_{i+2}}{6}\right)t^3 + \left(\frac{P_{i-1}}{6} - P_i + \frac{P_{i+1}}{2}\right)t^2 + \left(-\frac{P_{i-1}}{2} + \frac{P_{i+1}}{2}\right)t + \left(\frac{P_{i-1}}{6} + \frac{2P_i}{3} + \frac{P_{i+1}}{6}\right) \quad (7)$$

نقاط شروع و پایان هر منحنی نیز از روابط زیر محاسبه می‌شوند.

$$K_i = \frac{1}{6}(P_{i-1} + 4P_i + P_{i+1}) \quad (8)$$

$$K_{i+1} = \frac{1}{6}(P_i + 4P_{i+1} + P_{i+2})$$

۳-۱- استفاده از ویژگی B-Spline برای بازشناسی شکل

بازشناسی بر اساس منحنی‌های B-Spline به نقطه شروع حساس خواهد بود. لذا برای استفاده از این ویژگی در بازشناسی شکل‌ها در مرحله اول باید نقطه شروع مناسب برای اعمال ویژگی مشخص شود. برای این کار، ابتدا مرکز ثقل تصویر داده شده را محاسبه نموده و نقطه‌ای از کانتور شکل که بیشترین فاصله را از این مرکز دارد مشخص می‌نماییم. این نقطه را به مرکز ثقل متصل نموده و زاویه خط واصل را با محور عمودی محاسبه می‌نماییم و تصویر داده شده را به اندازه این زاویه می‌چرخانیم تا بیشترین شعاع تصویر در راستای عمودی قرار گیرد. این فرایند اثر نامطلوب چرخش را در بازشناسی از بین خواهد برد. مرکز ثقل منحنی را می‌توان با استفاده از مفهوم گشتاور محاسبه نمود. اگر فرض نماییم که گشتاور مرتبه $(p+q)$ ام تصویر با استفاده از رابطه زیر به دست آید:

$$m_{p,q} = \sum_x \sum_y (x^p y^q f(x, y)) \quad (9)$$

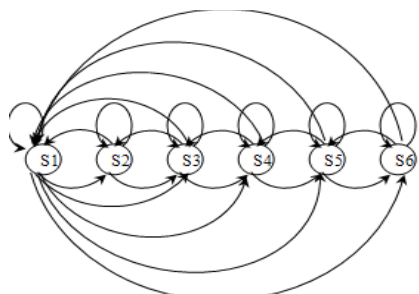
که در آن $x = 1, 2, \dots, m$ و $y = 1, 2, \dots, n$ تعداد ستونهای تصویر و n تعداد سطرهای آن می‌باشد. مرکز ثقل (x_c, y_c) که در واقع مرکز هندسی دست می‌باشد از رابطه (۱۰) محاسبه می‌گردد.

$$x_{COG} = \frac{m_{1,0}}{m_{0,0}}, \quad y_{COG} = \frac{m_{0,1}}{m_{0,0}} \quad (10)$$

اگر دورترین نقطه کانتور از این مرکز را با (x_f, y_f) نشان دهیم، زاویه خط واصل این نقطه با راستای عمودی از رابطه (۱۱) محاسبه می‌شود:



شکل (۲): حالات ایستای بررسی شده در این مقاله [۵]
Fig. (2): Static gestures that are studied in this paper [5]



شکل (۳): ارتباط حالات مختلف در مدل های مخفی مارکوف
Fig. (3): State relations in hidden markov models

در زمانهای $t=1,2,\dots$ حالات سیستم تغییر خواهد نمود. معمولاً حالت واقعی را با q_t نمایش می‌دهند. توصیف کامل سیستم در فرآیند مارکوف به مشخصات حالت فعلی و تمام حالت‌های پیشین نیاز دارد. اما در فرآیند مارکوف درجه اول این توصیف احتمالی تنها به حالت کنونی و حالت قبلی بستگی خواهد داشت. به عبارتی رابطه زیر را خواهیم داشت.

$$p(q_t = s_j | q_{t-1} = s_i, q_{t-2} = s_k, \dots) = p(q_t = s_j | q_{t-1} = s_i) \quad (17)$$

اگر خروجی مجموعه‌ای از حالات یا مشاهدات در زمان باشد، فرآیند را آشکار می‌نامند. اما اگر مشاهدات خود تابعی تصادفی از حالات باشند، مدل نهایی که مدل مخفی مارکوف نامیده می‌شود، یک فرآیند تصادفی پنهان متناظر همراه با یک فرآیند تصادفی پایه‌ای است که قابل مشاهده نمی‌باشد. اما توسط مجموعه دیگری از فرآیندهای تصادفی که ترتیب مشاهدات را تولید می‌کنند، قابل مشاهده است. مدل مورد استفاده در این مقاله، مدل مخفی مارکوف خواهد بود که پارامترهای زیر را برای آن در نظر می‌گیریم. N تعداد حالات در مدل می‌باشد (در این مقاله $N=6$) و M تعداد سمبل‌های مشاهده در هر حالت است. این سمبل‌ها متناظر با خروجی یک سیستم مدل شده می‌باشند. M ، در این مقاله همان بردار ویژگی‌هایی است که در مرحله استخراج ویژگی به دست آمد. مشاهدات را به صورت زیر نشان خواهیم داد. $V = \{v_1, v_2, \dots, v_M\}$. ماتریس توزیع احتمال تغییر حالت شامل وزن‌هایی است که برای خطوط ارتباطی حالات مختلف در شکل (۴) در نظر گرفته می‌شود. آن را با $A = \{a_{ij}\}$ نشان خواهیم داد و مولفه‌های آن با رابطه زیر محاسبه می‌شوند.

$$a_{ij} = P(q_{t+1} = s_j | q_t = s_i), \quad 1 \leq i, j \leq N \quad (18)$$

$$a_{ij} > 0 \quad \forall i, j$$

ماتریس توزیع احتمال سمبل‌های مشاهده در حالت j به شکل $B = \{b_j(k)\}$ تعریف می‌شوند که در آن:

دهیم. از آن جایی که این منحنی از ۴ نقطه کنترلی تشکیل شده است، رابطه (۱) را می‌توان به صورت زیر بازنویسی نمود.

$$[x_i(t), y_i(t)] = (t^3, t^2, t, 1) \begin{pmatrix} a_3 & b_3 & c_3 & d_3 \\ a_2 & b_2 & c_2 & d_2 \\ a_1 & b_1 & c_1 & d_1 \\ a_0 & b_0 & c_0 & d_0 \end{pmatrix} \quad (14)$$

$$\begin{bmatrix} x_{i-1}, y_{i-1} \\ x_i, y_i \\ x_{i+1}, y_{i+1} \\ x_{i+2}, y_{i+2} \end{bmatrix}, \quad t \in [0, 1]$$

لذا منحنی $P(t)$ به صورت مجموعه نقاطی با مختصات x, y نشان داده می‌شود. ویژگی شیب انحنا یا (κ) برای این منحنی از رابطه زیر محاسبه می‌گردد:

$$\kappa = \frac{dT}{ds} = \frac{|x'y'' - y'x''|}{(x'^2 + y'^2)^{\frac{3}{2}}} \quad (15)$$

که در این رابطه خواهیم داشت:

$$x' = \frac{dx}{dt}, \quad y' = \frac{dy}{dt} \quad (16)$$

این ویژگی را برای هر کدام از قطعات منحنی B-Spline به دست آورده و برای بازشناسی مورد استفاده قرار می‌دهیم. به این ترتیب به ازای هر منحنی ۲ ویژگی و در کل M منحنی خواهیم داشت. لذا $2M$ ویژگی غیر حساس به انتقال، چرخش و تغییر اندازه برای هر تصویر محاسبه می‌شود. بردار ویژگی‌های مورد استفاده یک ماتریس $M \times 2$ خواهد بود. به عنوان مثال اگر ۱۰۰ نقطه روی کانتور شکل انتخاب نماییم، آنها را با ۳۳ مرکز قطعه منحنی‌ها جایگزین خواهیم نمود و برای این ۳۳ نقطه ۶۶ ویژگی به دست خواهیم آورد. از این ویژگی‌ها برای بازشناسی شکل بهره خواهیم گرفت. این ویژگی‌ها به عنوان ورودی به مرحله بعد وارد می‌شوند.

۴- تطابق

در مراجع روش‌های متفاوتی برای طبقه‌بندی و کلاس‌بندی ویژگی‌ها مورد استفاده قرار گرفته است [۳۴-۳۶]. یکی از روش‌هایی که برای کلاس‌بندی ویژگی‌ها مورد استفاده قرار می‌گیرد، مدل مخفی مارکوف می‌باشد [۳۷-۳۸]. در تحقیق حاضر از این سیستم برای بازشناسی استفاده شده است. سیستم به گونه‌ای طراحی شده است که در هر زمان می‌توان آن را به صورت یکی از N حالت مشخص S_1, \dots, S_N بیان کرد. در این مقاله N برابر با ۶ در نظر گرفته می‌شود. این ۶ حالت در شکل (۲) نشان داده شده‌اند. (این تصاویر از مرجع [۵] انتخاب شده‌اند)

حالات مختلف و گذار مربوطه از حالت S_1 به حالات دیگر را می‌توان در شکل (۳) مشاهده نمود. از ترسیم گذار مربوط به حالات دیگر به دلیل دشواری مشاهده تصویر خودداری نموده‌ایم. اما این نکته را در نظر خواهیم گرفت که تمامی حالات به همین ترتیب با هم ارتباط خواهند داشت.

نهایت پارامترهای مدل با روابط زیر به دست خواهد آمد که از آنها در ساخت مدل استفاده می‌نماییم.

$$\pi_i = \frac{\pi_i \frac{\partial p}{\partial \pi_i}}{\sum_{k=1}^N \pi_k \frac{\partial p}{\partial \pi_k}} \quad (27)$$

$$a_{ij} = \frac{a_{ij} \frac{\partial p}{\partial a_{ij}}}{\sum_{k=1}^N a_{ik} \frac{\partial p}{\partial a_{ik}}} \quad (28)$$

$$b_j(k) = \frac{b_j(k) \frac{\partial p}{\partial b_j(k)}}{\sum_{l=1}^M b_j(l) \frac{\partial p}{\partial b_j(l)}} \quad (29)$$

به این ترتیب مدل مخفی مارکوف که در این مقاله از آن استفاده نموده‌ایم ساخته می‌شود. از این مدل برای بازشناسی حالت نهایی استفاده می‌نماییم. نتایج بازشناسی در بخش بعد آورده شده است.

۵- نتایج

الگوریتم مطرح شده در این مقاله را روی ۲۱۰۵ تصویر که در قالب فریم‌های ویدیویی از حرکات دست ۱۵ کاربر متفاوت جمع آوری شده است، اعمال نمودیم. این دنباله‌ها با استفاده از یک دوربین سامسونگ با سرعت تصویر برداری 24 fps تهیه شده‌اند. هر توالی شامل تقریباً 150 فریم می‌باشد و وضوح تصویر 240×320 پیکسل می‌باشد. برنامه‌ها با نرم‌افزار MATLAB نوشته شده‌اند و با نوت بوک با مشخصات زیر پیاده‌سازی شده‌اند: processor 1.7GHz, Intel(R) Celeron(R) M و 512 GRAM جدول (۱) و نمودارهای (۱) و (۲)، نتایج حاصل از بازشناسی برای سه سکانس ویدیویی نمونه که به طور تصادفی از میان سکانس‌های موجود، انتخاب شده‌اند را نشان می‌دهند.

در ردیف افقی این نمودارها تعداد نقاط انتخاب شده روی مرزهای شکل (که آن را با n نشان می‌دهیم) را از ۲۰ نقطه تا ۲۰۰ نقطه افزایش می‌دهیم. نتایج بازشناسی درست به ازای nهای مختلف در نمودار (۱) و زمان لازم برای پردازش هر فریم و تشخیص حالت دست در هر فریم در نمودار (۲) نشان داده شده است. همان طور که در نمودار (۱) می‌توان مشاهده نمود، به ازای n < 60 میزان بازشناسی برای سکانس‌های متفاوت تغییرات شدید داشته و در محدوده وسیعی نوسان می‌کند. به ازای n > 80 اگرچه بازشناسی به روند تقریباً ثابتی می‌رسد اما میزان آن کاهش می‌یابد و علاوه بر این همان طور که در نمودار (۲) نشان داده شده است زمان پردازش فریم‌ها افزایش یافته و به عبارتی هرچه n افزایش یابد، سیستم بازشناسی کندتر خواهد بود. لذا انتخاب n در محدوده 60 < n < 80 می‌تواند انتخاب مناسبی برای بازشناسی باشد. در این مقاله n = 70 در نظر گرفته شده است.

$$b_j(k) = (v_k \text{ in time}(t) \mid q_t = s_j) \quad (19)$$

$$1 \leq j \leq N, \quad 1 \leq k \leq M$$

و نهایتاً ماتریس توزیع حالت ابتدایی به شکل $\pi = \{\pi_i\}$ بوده و با رابطه زیر به دست خواهد آمد:

$$\pi_i = P(q_1 = s_i), \quad 1 \leq i \leq N \quad (20)$$

مشخصات کامل یک HMM نیاز به مشخص شدن دو پارامتر مدل M, N, سمبل‌های مشاهده و مقادیر احتمال A, B, π دارد. در حالت کلی از شکل نمایش $\lambda = (A, B, \pi)$ استفاده می‌نمایند. فرایند بازشناسی به سه مسأله اصلی تقسیم می‌شود. که عبارتند از:

(۱) رشته مشاهدات $O = o_1 o_2 \dots o_T$ داده شده‌اند. مدل $\lambda = (A, B, \pi)$ نیز داده شده است. هدف یافتن $P(O \mid \lambda)$ به نحوی است که از بین مدل‌های مختلف مدلی که بهترین سازگاری با رشته مشاهدات را داشته باشد، انتخاب نماییم. الگوریتمی که از آن برای حل این مسأله استفاده نموده‌ایم، پیشرو-پسرو نام دارد. در این الگوریتم دو متغیر تعریف می‌شود. متغیر پیشرو یا α و متغیر پسرو یا β که متغیر پیشرو با استفاده از رابطه زیر به دست می‌آید:

$$\alpha_t(i) = P(o_1 o_2 \dots o_t, q_t = s_i \mid \lambda) \quad (21)$$

این متغیر را می‌توان با استفاده از راه حل استقرایی سه مرحله‌ای زیر محاسبه نمود:

$$\alpha_1(i) = \pi_i b_i(o_1), \quad 1 \leq i \leq N \quad (22)$$

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(o_{t+1}) \quad (23)$$

$$P(O \mid \lambda) = \sum_{i=1}^N \alpha_T(i) \quad (24)$$

یک متغیر پسرو نیز در بخش پسرو الگوریتم مذکور مطرح می‌گردد. این متغیر پسرو از رابطه (۲۵) به دست می‌آید که می‌توان آن را با رابطه (۲۶) بازنویسی نمود.

$$\beta_t(i) = P(o_{t+1} o_{t+2} \dots o_T \mid q_t = s_i, \lambda) \quad (25)$$

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(o_{t+1}) \beta_{t+1}(j) \quad (26)$$

$$t = Y - 1, T - 2, \dots, 1$$

با استفاده از این الگوریتم و مدل‌های مخفی مارکوف، ماتریس ویژگی به دست آمده در مرحله قبل را به یکی از مدل‌های موجود نسبت می‌دهیم. مسأله دومی که در مدل‌های مخفی مارکوف از آن تشکیل شده است این است که:

(۲) رشته مشاهدات و مدل داده شده‌اند. هدف یافتن ترتیب حالات $Q = q_1 q_2 \dots q_T$ است که بهترین توصیف‌کننده مشاهدات باشد. برای این بخش از الگوریتم ویتربی استفاده می‌نماییم. این بخش در ضمیمه توضیح داده شده است.

(۳) در بخش سوم و نهایی مدل‌های مخفی مارکوف باید پارامترهای مدل را باید به نحوی تعیین نماییم که $P(O \mid \lambda)$ ماکزیمم گردد. برای حل این مسأله نیز یک الگوریتم تکراری مورد استفاده قرار می‌گیرد. در

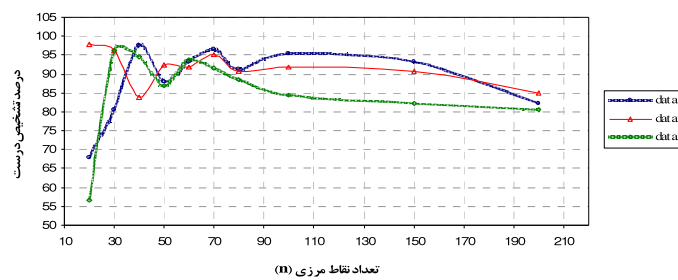
Table (1): Number of correct gesture recognition and consumed time for each frame of three sample video sequences

جدول (۱): تعداد تشخیص حالات درست و زمان موردنیاز برای تشخیص حالت هر فریم برای سه سکانس ویدیویی نمونه

تعداد نقاط مرزی	نام داده	تعداد درست	زمان پردازش هر فریم (ثانیه)	درصد تشخیص درست
N=20	Data1 (174 frames)	118	0.59	67.8
	Data2 (185 frames)	181	0.4988	97.84
	Data3 (129 frames)	73	0.58	56.6
N=30	Data1 (174 frames)	140	0.54	80.46
	Data2 (185 frames)	178	0.6	96.21
	Data3 (129 frames)	124	0.583	96.12
N=40	Data1 (174 frames)	170	0.59	97.7
	Data2 (185 frames)	155	0.683	83.78
	Data3 (129 frames)	122	0.61	94.57
N=50	Data1 (174 frames)	153	0.657	87.9
	Data2 (185 frames)	171	0.598	92.4
	Data3 (129 frames)	112	0.578	86.82
N=60	Data1 (174 frames)	162	0.64	93.1
	Data2 (185 frames)	170	0.651	91.89
	Data3 (129 frames)	121	0.583	93.8
N=70	Data1 (174 frames)	168	0.641	96.55
	Data2 (185 frames)	176	0.655	95.13
	Data3 (129 frames)	118	0.623	91.47
N=80	Data1 (174 frames)	159	0.666	91.38
	Data2 (185 frames)	168	0.657	90.81
	Data3 (129 frames)	114	0.61	88.37
N=100	Data1 (174 frames)	166	0.66	95.4
	Data2 (185 frames)	170	0.64	91.89
	Data3 (129 frames)	109	0.612	84.5
N=150	Data1 (174 frames)	162	0.6788	93.1
	Data2 (185 frames)	168	0.651	90.81
	Data3 (129 frames)	106	0.63	82.17
N=200	Data1 (174 frames)	143	0.7159	82.18
	Data2 (185 frames)	157	0.75	84.86
	Data3 (129 frames)	104	0.75	80.62

نمودارهای (۱) و (۲) جدول (۱) را تشریح می کنند.

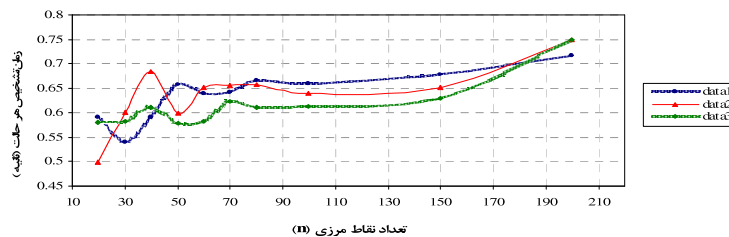
درصد تشخیص درست برای n های مختلف



نمودار (۱): درصد تشخیص درست حالت دست به ازای تعداد نقاط مرزی متفاوت برای سه نمونه از سکانس های ویدیویی

Chart (1): Correct recognition rate of hand gesture versus contour points number for three video sequences

زمان لازم برای تشخیص هر حالت به ازای n های مختلف (ثانیه)



نمودار (۲): زمان لازم برای تشخیص حالت دست در هر فریم (ثانیه) به ازای تعداد نقاط مرزی متفاوت برای سه نمونه از سکانس های ویدیویی

Chart (2): Required time (in second) for recognition of hand gesture in each frame versus contour points number for three sample video sequences

۶- جمع بندی

در این مقاله یک روش مبتنی بر HMM برای بازشناسی برخی حالات ایستای دست ارائه نمودیم. روش پیشنهادی بر ویژگی شیب انحنا منحنی BSpline تصویر دست متکی بوده است. در این مقاله، ابتدا منحنی پیرامونی دست را به 70 قطعه تقسیم نموده و هر قطعه منحنی را با یک منحنی Bspline مرتبه ۳ تقریب زدیم. برای غیر حساس بودن به چرخش و تغییر اندازه و انتقال به جای استفاده مستقیم از ضرایب منحنی، شیب انحنا برای هر منحنی محاسبه شده است. این ویژگی را برای تمام منحنی‌های تشکیل دهنده دست محاسبه نموده و بردار ویژگی حاصل را برای کلاس بندی به مدل‌های مخفی مارکوف داده‌ایم. HMM مورد استفاده از ۶ زنجیره مارکوف تشکیل شده است و ۶ حالت ایستای دست را تشخیص می‌دهد. نتایج حاصل از بازشناسی نشان داد که روش پیشنهادی در 92.21% از موارد بازشناسی درستی خواهد داشت.

ضمایم

الگوریتم ویتربی

الگوریتم Viterbi برای یافتن بهترین تک رشته حالت $Q = \{q_1, q_2, \dots, q_T\}$ برای رشته مشاهدات داده شده $O = \{o_1, o_2, \dots, o_T\}$ مقدار $\delta_t(i)$ را به صورت زیر تعریف می‌نماید.

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P(q_1 q_2 \dots q_t = i, o_1 o_2 \dots o_t | \lambda) \quad (I)$$

یعنی $\delta_t(i)$ در زمان t در مسیری را دارا می‌باشد که t مشاهده اول به s_j ختم شود. با استقراء خواهیم داشت:

$$\delta_{t+1}(i) = \max_{1 \leq l \leq N} [\delta_t(l) \cdot a_{lj}] \cdot b_i(o_{t+1}) \quad (II)$$

الگوریتم زیر رابطه (II) را برای هر t و j ماکزیمم می‌کند، که به الگوریتم معروف است.

>> Start

$$\delta_1(i) = \pi_i b_i(o_1), \quad 1 \leq i \leq N \quad (III)$$

$$\psi_1(i) = 0$$

>> return

$$\delta_t(j) = \max_{1 \leq l \leq N} [\delta_{t-1}(l) \cdot a_{lj}] \cdot b_j(o_t) \quad (IV)$$

$$\psi_t(j) = \arg \max_{1 \leq l \leq N} [\delta_{t-1}(l) \cdot a_{lj}] \quad (IV)$$

$$2 \leq t \leq T, \quad 1 \leq i \leq N$$

>> end

$$p^* = \max_{1 \leq i \leq N} [\delta_T(i)] \quad (V)$$

$$q^* = \arg \max_{1 \leq i \leq N} [\delta_T(i)] \quad (V)$$

>> end return

$$q^* = \psi_{t+1}(q_{t+1}^*), \quad t = T-1, T-2, \dots, 1 \quad (VI)$$

>>end

پی‌نوشت:

- 1- Chamfer
- 2- Particle filter

جدول (۲) نتایج حاصل از اعمال روش پیشنهادی در این مقاله را بر روی ۲۱۰۵ که بیانگر ۶ حالت از حالات ایستای دست کاربرهای مختلف می‌باشند نشان می‌دهد. همانطور که در جدول (۲) نشان داده شده است روش پیشنهادی بهترین بازشناسی را در مورد حالت (۵) شکل (۲) و بدترین بازشناسی را در مورد حالت (۱) شکل (۲) به دست آورده است.

Table (2): Final recognition results for each gesture of six studied gestures in this paper

جدول (۲): نتایج نهایی بازشناسی برای هر کدام از ۶ حالت ایستای بررسی شده در این مقاله







حالت ایستا	تعداد کل فریم	تعداد درست	درصد درست
 1	213	173	%81.22
 2	336	288	%85.71
 3	343	328	%95.63
 4	364	339	%93.13
 5	417	401	%96.16
 6	432	412	%95.37

Table (3): Final recognition of static hand gestures

جدول (۳): بازشناسی نهایی حالات ایستای دست

تعداد کل فریم‌ها	تعداد تشخیص‌های درست	درصد تشخیص درست	متوسط زمان پردازش یک فریم
2105	1941	%92.21	0.71 ثانیه

Table (4): Comparison of proposed method with method in [29]

جدول (۴): مقایسه روش پیشنهادی با روش [۲۹]

روش مرجع [۲۹]	روش پیشنهادی
%89.3	%92.21

نتایج نهایی بازشناسی که از اعمال روش پیشنهادی مقاله روی 2105 تصویر که در قالب بیش از ۱۵ سکانس ویدیویی تهیه شده‌اند، در جدول (۳) نشان داده شده است. همان طور که در جدول نیز نشان داده شده است، الگوریتم پیشنهادی ما در 92.21% از موارد بازشناسی صحیحی را انجام خواهد داد. همچنین جدول (۴) نرخ بازشناسی کل روش پیشنهادی را با روش ارائه شده در [۲۹] که ترکیبی از سه ویژگی تعداد انگشتان فعال و نیمه‌فعال، مساحت محصور شده و تبدیل هاف را برای بازشناسی مورد استفاده قرار می‌دهد، مقایسه می‌کند.

References

- [1] C. Manresa, J. Varona, R. Mas, F.J. Perales, "Real-time hand tracking and gesture recognition for human-computer interaction", *Elec. Lett. on Comp. Vis. and Image Anal.*, pp.1-7, 2000.
- [2] A. Erol, G. Bebis, M. Nicolescu, R.D. Boyle, X. Twombly, "Vision-based hand pose estimation: A review", *Comp. Vis. and Image Under.*, Vol.108, No.1-2, pp.52-73, Oct. 2007.
- [3] A. Thayananthan, B. Stenger, P.H.S. Torr, R. Cipolla, "Shape context and chamfer matching", *Conf. Comp. Vis. and Patt. Recog.*, IEEE Comp. Soci., Vol.1, pp.127-136, 2003.
- [4] S. Lu, D. Metaxas, D. Samaras, J. Oliensis, "Using multiple cues for hand tracking and model refinement", *Conf. on Comp. Vis. and Patt. Recog.*, IEEE Comput. Soci., Vol.2, pp.443-450, 2003.
- [5] C.C. Chang, C.Yi. Liu, W.K. Tai, "Feature alignment approach for hand posture recognition based on curvature scale space", *Neurocomputing*, Elsevier, Vol.71, No.10-12, pp.1-7, Jun. 2008.
- [6] C.C. Chang, "Adaptive multiple sets of CSS features for hand posture recognition", *Neurocomputing*, Elsevier, Vol.69, No.16-18, pp.2017-2025, Oct. 2006.
- [7] Y.H. Lee, C.Y. Tsai, "Taiwan sign language (TSL) recognition based on 3D data and neural networks", *Expert Systems with Applications*, Elsevier, Vol.36, No.2, pp.1-6, Mar. 2009.
- [8] X. Yin, M. Xie, "Finger identification and hand posture recognition for human-robot interaction", *Image and Vis. Comp.*, Elsevier, Vol.25, No.8, pp.1291-1300, Aug. 2007.
- [9] Y. Wu, J. Lin, T.S. Huang, "Analyzing and capturing articulated hand motion in image sequences", *IEEE Trans. on Patt. Anal. and Mach. Intelli.*, IEEE Computer Society, Vol.27, No.12, pp.1910-1922, Dec. 2005.
- [10] H. Guan, J.S. Chang, L. Chen, R.S. Feris, M. Turk, "Multi-view appearance-based 3d hand pose estimation", *Comp. Vis. and Patt. Recog. Workshop*, pp.154-154, Jun. 2006.
- [11] R. Rosales, V. Athitsos, L. Sigal, S. Sclaroff, "3d hand pose reconstruction using specialized mapping", *Inter. Conf. on Comp. Vis.*, Vol.1, pp.378-385, Jul. 2001.
- [12] X. Yin, M. Xie, "Finger identification and hand posture recognition for human-robot interaction", *Image and Vision Computing*, Elsevier, Vol.25, No.8, pp.1291-1300, Aug. 2007.
- [13] Q. Munib, M. Habeeb, B. Takruri, H.A. Al-Malik, "American sign language (ASL) recognition based on Hough transform and neural networks", *Expert Systems with Applications*, Elsevier, Vol.32, No.1, pp.24-37, Jan. 2007.
- [14] C. Shan, T. Tan, Y. Wei, "Real-time hand tracking using a mean shift embedded particle filter", *Patt. Recog.*, Elsevier, Vol.40, No.7, pp.1958 - 1970, July 2007.
- [15] C.W. Ng, S. Ranganath, "Real-time gesture recognition system and application", *Image and Vis. Comp.*, Elsevier, Vol.20, No.13-14, pp.993-1007, Dec. 2002.
- [16] R. Bowden, M. Sarhadi, "A non-linear model of shape and motion for tracking finger spelt American sign language", *Image and Vis. Comp.*, Elsevier, Vol.20, No.9-10, pp.597-607, Aug. 2002.
- [17] E. Ong, R. Bowden, "A boosted classifier tree for hand shape detection", *Inter. Conf. on Auto. Face and Gesture Recog.*, pp.889- 894, May 2004.
- [18] E.S. Nielsen, L.A. Canalís, M.H. Tejera, "Hand gesture recognition for human-machine interaction", *Jou. of WSCG*, Vol.12, No.1-3, Plzen, Czech Republic, pp.1-8, Feb. 2003.
- [19] C. Manresa, J. Varona, R. Mas, F.J. Perales, "Hand tracking and gesture recognition for human-computer interaction", *Elec. Lett. on Comp. Vision and Image Anal.*, Vol.5, No.3, pp.96-104, 2005.
- [20] T.H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein, "Introduction to algorithms", second edition, MIT Press, 2001.
- [21] X. Bai, L. Latecki, "Path similarity skeleton graph matching", *IEEE Trans. Patt. Anal. and Mach. Intell.*, Vol.30, No.7, pp.1282-1292, July 2008.
- [22] M. Kolsch, M. Turk, "Robust hand detection", *Inter. Conf. on Auto. Face and Gesture Recog.*, pp.614-619, May 2004.
- [23] J. Shotton, A. Blake, R. Cipolla, "Multiscale categorical object recognition using contour fragments", *IEEE Trans. Patt. Anal. and Mach. Intell.*, Vol.30, No.7, July 2008.
- [24] S.M. Lee, A.L. Abbott, N.A. Clark, P.A. Araman, "Spline curve matching with sparse knot sets", *Asian Conf. on Comp. Vision*, pp.246-251, 2004.
- [25] S. Belongie, J. Malik, J. Puzicha, "Shape matching and object recognition using shape context", *IEEE Trans. patt. Anal. and Mach. Intell.*, Vol.24, No.4, pp.509-522, Apr. 2002.
- [26] N. Alijanpour, H. Ebrahimnezhad, A. Ebrahimi, "Inner distance based hand gesture recognition for devices control", *Inter. Conf. on Innov. in Info. Tech.*, IEEE, United Arab Emirates, pp.742-746, Dec. 2008.
- [27] N. Alijanpour, H. Ebrahimnezhad, "Shape recognition using combined features of crossing circles", *Iranian Conf. on Elect. Engi. (ICEE2009)*, Tehran, pp.614-619, May 2009.
- [28] N. Alijanpour, H. Ebrahimnezhad, A. Ebrahimi, "Hand gesture recognition by combining the geometric and shape context features to control the wheelchair", *15th Iranian Conf. on Biom. Engi. (ICBME 2008)*, Mashhad, Iran, pp.316-321, Feb. 2009 (in Persian).
- [29] N. Alijanpour, H. Ebrahimnezhad, A. Ebrahimi, "Hand gesture recognition by combining the features of active fingers, differential area and Hough transform", *Proc. of 17th Iranian Conf. on Elect. Engi. (ICEE2009)*, pp.59-64, May 2009 (in Persian).

- [30] K. Imagawa, H. Matsuo, R. Taniguchi, D. Arita, S. Lu, S. Igi, "Recognition of local features for camera-based sign language recognition system", Inter. Conf. on Patt. Recog., Barcelona, Spain, pp.849-853, 2000.
- [31] R. Rosales, V. Athitsos, L. Sigal, S. Sclaroff, "3d hand pose reconstruction using specialized mappings", IEEE Inter. Conf. on Comp. Vision, Vancouver, BC, Canada, Vol.1, pp.378 - 385, 2001.
- [32] H. Fillbrandt, S. Akyol, K. Kraiss, "Extraction of 3d hand shape and posture from image sequences for sign language recognition", IEEE Inter. Workshop on Analy. and Model. of Faces and Gestures, pp.181-186, Oct. 2003.
- [33] R.C. Gonzalez, R.E. Woods, "Digital image processing", Second Edition, Prentice Hall, 2002.
- [34] D. Salomon, "Curves and surfaces for computer graphics", Springer Science+Business Media, the United States of America, 2006.
- [35] A.R. Webb, Q.L. Malvern, UK, Statis. Patt. Recog., First edition, Butterworth Heinemann, 2002.
- [36] R. Bowden, D. Windridge, T. Kadir, A. Zisserman, M. Brady, "A linguistic feature vector for the visual interpretation of sign language", Europ. Conf. on Comp. Vision, ECCV04, Springer-Verlag, pp.391-401, 2004.
- [37] L.F. Costa, R.M. Cesar, "Shape analysis and classification", CRC Press LLC, 2001
- [38] M.S. Nixon, A.S. Aguado, "Feature extraction and image processing", Second Edition, Elsevier Ltd, 2008.