تشخیص بر لحظه دست انسان در تصاویر ویدئویی

۹۱۱۰۷۲۳۰۵ـ محمد علی طیبی
دانشگاه آزاد واحد علوم و تحقیقات سیرجان
Email:mohamadali.tayebi@gmail.com

۹۱۱۰۵۷۹۱۵ـ محمد مهدی دهیده
دانشگاه آزاد واحد علوم و تحقیقات سیرجان
Email: mohamadmahdi90@yahoo.com

۹۱۱۰۵۸۱۲۱ـ سجاد باقری
دانشگاه آزاد واحد علوم و تحقیقات سیرجان
Email: sajad_bagheri@ymail.com

چکیده—با پیشرفت علم پردازش تصویر تشخیص خودکار بسیاری از اشیا و اعضا امکان‌پذیر شده است. تشخیص اعضای بدن انسان همواره به‌عنوان یکی از مسائل موردعلاقه محققین که با چالش‌های زیادی همراه بوده، قرار گرفته است. روش‌هایی که بتواند دست انسان را در میان ویدوئوهایی با حضور پس‌زمینه‌های شلوغ تشخیص دهد طراحی شده‌اند که هرکدام مزایا و معایب خود را بالطبع دارند. در این بین، روشی که بتواند پس از تشخیص دست انسان در یک فریم یا محدود زمانی توانایی دنبال کردن و ادامه حضور دست در تصویر را پیش‌بینی کند با حداقل سربار محاسباتی دارای چالش‌های فراوانی است. در این مستند سعی شده است روشی برای تشخیص دست انسان با پردازش برلحظه ارائه شود. در این مستند از روش‌های تشخیص پوست بدن انسان و از تشخیص‌دهنده sift استفاده شده است. نتایج این مستند قابلیت بالای این روش در تشخیص و دنبال کردن دست در تصویر را نمایش می‌دهد.

***کلمات کلیدی*—آمیختن ، تشخیص دست، پردازش برلحظه، سیفت،دنبال کردن شی.**

۱. مقدمه

به‌منظور پیش‌بینی و تشخیص دست انسان در تصویر با پس‌زمنیه سفید یا تک رنگ، می‌توانیم از روش‌های متعددی مانند تشخیص پوست دست انسان استفاده کنیم. در این روش به این دلیل که می‌دانیم پوست دست انسان دارای محدوده رنگ خاصی است با جداکردن این طیف رنگ می‌توان ادعا کرد که جسمی که دارای این طیف رنگ است قطعا پوست انسان است. پس از این مرحله به‌این دلیل که بدن انسان دارای بخش‌های متفاوتی است باید مرحله تشخیص دست و جداسازی دست از دیگر اعضای بدن را انجام دهیم. روش‌های متفاوتی به‌منظور تشخیص اشیا با اشکال خاص طراحی و ابداع گردیده‌اند که می‌توانند در این بخش مورد استفاده قرار گیرند. یکی از این روش‌ها تشخیص یک شی براساس شکل آن است. در این روش براساس شکل دست می‌توان دست اقدام به تشخیص آن در تصویر کرد. مشکل این روش این است که همیشه شکل دست ساختاری یکسان ندارد و حتی دست باز انسان با پنج انگشت نیز دارای اشکال متفاوتی می‌توان باشد. این مساله با بررسی بسته بودن دست نیز وارد فاز پیچیده‌تری از تشخیص و پردازش خواهد شد. همچنین تغییر اندازه دست نیز می‌تواند این روش را با چالش‌هایی مواجه کند. این تنها مشکل در تشخیص دست انسان و تمایز قائل شدن بین اعضای دیگر انسان نیست.

پس‌زمینه خلوت و تک رنگ آن چیزی نیست که محققین امروزه به‌دنبال آن باشند، بلکه تشخیص دست افراد در محیط‌های شلوغ و با حضور چند دست یک زمینه موفق تحقیقاتی و عملیاتی است. برای رفع این مشکل از تشخیص پوست انسان، روش sift و همچنین دنبال کردن رفتار دست در روش پیشنهادی خود استفاده کرده‌ایم که در بخش‌های ۲، ۳ و ۴ به آن‌ها پرداخته خواهد شد. در بخش۲ ، با بررسی روش‌ها به توضیح مختصری از آن‌ها خواهیم پرداخت. در بخش ۳ چگونگی ترکیب و استفاده از روش‌های گفته شده را بررسی خواهیم کرد و در بخش ۴ به بررسی چند تصویر و نتایج حاصل از استفاده از این روش خواهیم پرداخت.

۲. شیوه‌ی پیشنهادی

در تشخیص اشیا با شکل و ویژگی‌های خاص در تصاویر ویدوئی باید راه‌کاری به‌منظور استخراج و دسته‌بندی این ویژگی‌ها ارائه و طراحی کرد. به‌منظور تشخیص دست انسان از روش تشخیص پوست انسان در مرحله اول پردازش استفاده شده است و به این ترتیب اعضای انسان در مرحله اول از کل تصویر جدا شده و چالش اصلی فقط تشخیص شکل دست در تصویر موجود است. به‌منظور رفع مشکل تشخیص اعضای دیگر دست از روش‌های زیادی می‌توان استفاده کرد که روش پیشنهادی ما روش sift به‌علت سرعت و دقت بالای این روش است. با ترکیب این دو روش و استفاده از مکان فعلی دست و مکان بعدی آن قادر خواهیم بود به سادگی رفتار دست را تشخیص دهیم.

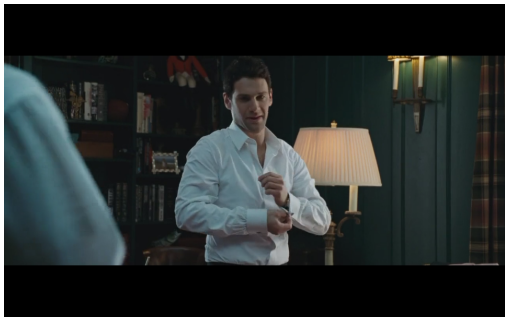
آ. تشخیص پوست انسان به‌کمک طیف رنگ

همان‌طور که پیش‌تر بحث شد، باید در مرحله اول تشخیص دست انسان، پوست انسان تشخیص داده شده و سپس از تصویر اصلی جدا شود. به این ترتیب بار پردازش و تشخیص دست کاهش چشم‌گیری خواهد داشت. در روش پیشنهادی از مدل رنگ YCbCr استفاده شده است و دلیل آن حساس بودن این مدل رنگ به شدت نور رنگ است. همچنین این ویژگی باعث کاهش خطای تشخیص دیگر اشیا به‌عنوان دست نیز می‌شود. در این روش سه مولفه Y شدت نور پیکسل، Cb تفاوت رنگ آبی و Cr تفاوت رنگ قرمز هر پیکسل به‌عنوان سه مولفه فضای رنگی معرفی شده‌اند. این سه مولفه براساس مقادیر RGB قابل محاسبه هستند. تفاوت رنگ آبی و قرمز هر پیکسل نیز براساس تفاوت سه مولفه قرمز، آبی و سبز مدل RGB قابل محاسبه است.

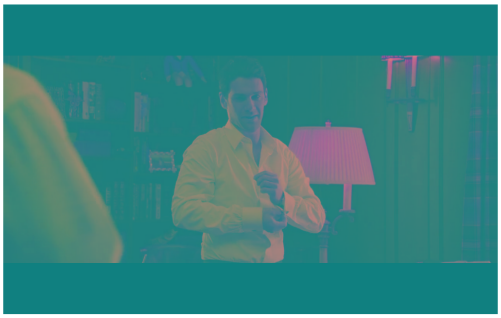
پس از محاسبه مقادیر مدل رنگ YCbCr، باید طیف رنگی را که پوست دست در آن قرار دارد را از تصویر جدا و باقی تصویر را حذف کنیم. طیف پیشنهادی برای تفاوت رنگ آبی ۱۲۷-۷۷ و برای تفاوت رنگ قرمز ۱۷۳-۱۳۳ است. در واقع مولفه اول که همان شدت نور است تاثیری در انتخاب و تشخیص پوست دست ندارد و این تنها تفاوت رنگ آبی و قرمز است که می‌تواند در تشخیص رنگ پوست دست کمک بکند. نتایج این بخش در تصاویر ۳گانه، یک تا سه قرار قابل مشاهده هستند.

ب. استخراج ویژگی به کمک SIFT

تبدیل ویژگی مستقل از مقیاس یا همان روش SIFT یکی از دقیق‌ترین و سریع‌ترین روش‌ها در استخراج ویژگی از تصاویر مختلف است. یکی از کاربردهای گسترده SIFt در شناسایی تصویر و روش تطبیق نمونه (باتصویر) است. روش استخراج ویژگی نه تنها به انتقال، پیمایش و چرخش وابسته نیست بلکه تغییرات نورپردازی و مکان جغرافیایی نیز تاثیری در تشخیص و استخراج ویژگی‌های همسان ندارند و از این‌رو توانایی تشخیص شی را در هر زاویه و اندازه‌ای با نمونه موردنظر دارد. در روش SIFT، ابتدا تصویر به‌صورت فیلتر کردن آبشاری به‌کمک فیلتر گاوسی با سیگما ۱ تا ۲ تبدیل می‌شود به یک‌سری تصویر جدید تبدیل



شکل ۱: تصویر اصلی در فرمت RGB



شکل ۲: تصویر در فرمتYCbCr

می‌شود.سپس این هر تصویر جدید از همسایه خود از دید سیگما کم شده و تفاضل گاوسی را به‌وجود می‌آورد. پس از این مرحله کمینه‌ها و بیشینه‌ها به‌عنوان نقاط کلیدی تفاضل گاوسی انتخاب می‌شوند. تفاضل گاوسی براساس عبارت ۱ و با تعریف دو معیار سیگما همسایه قابل محاسبه است و همچنین L مورد بحث در عبارت ۱ از کانولوشن تصویر با فیلتر گاوسی است که در عبارت ۲ به آن اشاره شده است.

$$D(x, y, sigma) = L(x, y, k_i\sigma) - L(x, y, k_j\sigma) \qquad (۱)$$

$$L(x, y, k\sigma) = G(x, y, k\sigma) * I(x, y)) \qquad (۲)$$

نقاط به‌دست آمده توسط روش مطرح شده بسیار زیاد هستند و مشکل اینجاست که بسیاری از این نقاط قابلیت استفاده شدن به‌منظور نقاط کلیدی را ندارند. به‌منظور حذف نقاط کم‌ـبی کاربرد باید نقاطی که دارای کنتراست پایین هستند و در مقابل اغتشاش مقاوم نیستند حذف شوند. همچنین نقاطی که به‌شکل ضعیفی در طول یک لبه توزیع و محاسبه شده‌اند، نقاط مناسبی نیستند و باید حذف شوند. در ابتدا باید به‌منظور تشخیص نقاط مفید و نقاط بی‌فایده، به درون‌یابی نقاط محاسبه شده پرداخت و این‌کار براساس محاسبه همسایه هر نقطه کلیدی صورت می‌گیرد. این ویژگی محاسبه برحسب همسایه باعث مقاوم بودن این روش در مقابل تغییر اندازه، تغییر زاویه و تغییر نور می‌شود. درون‌یابی نقاط کاندید براساس بسط تیلور درجه دوم بر روی تفاضل گاوسی‌های به‌دست آمده و پیکسل‌ها و به‌کمک عبارت۳ صورت می‌گیرد. دقت داریم که DوX به‌ترتیب مشتق و پیکسل مربوط به فیلتر گاوسی با سیگما خاص هستند.

شکل ۳: بخش‌های جدا شده تصویر که به‌عنوان پوست شناسایی شده‌اند

$$D(x) = D + \frac{\partial D^T}{\partial x} + \frac{۱}{۲} * x^T * \frac{\partial D}{\partial x^۲} * x \qquad (۳)$$

پس از محاسبه درون‌یابی، مقدار اکسترمم(x^-) را براساس صفر قرار دادن مشتق تابع تیلور درجه دوم و محاسبه آن به‌دست خواهد آمد. سپس به‌منظور حذف نقاط با کنتراست پایین، با محاسبه شده برای مشتق درجه دوم بسط تیلور با مقدار (x^-) نقاط کاندیدی که مقداری کمتر از ۰٫۰۳ دارند حذف و باقی نقاط باقی خواهند ماند. حتی پس از این هرس کردن نیز نقاط دیگری باقی می‌مانند که در طول لبه‌ها پاسخ‌های قوی دارند و حتی در صورت حذف این نقاط نیز در مرحله قبل همچنان مقداری از نقاط باقی‌مانده نقاط مغشوشی هستند که پاسخ‌های آن‌ها نسبت به لبه‌ها مانع از حذف آن‌ها شده است. به‌منظور حذف نقاط کاندید با این ویژگی از روش تشخیص گوشه به‌کمک ماتریس هسین به‌جای مشتق دوم مرتبه دوم استفاده می‌شود.

در مرحله بعد به‌منظور انتخاب نقاط کلیدی، هر نقطه کاندید براساس جهت گرادیان به یک یا چند جهت تخصیص پیدا می‌کند. هدف این‌کار یافتن جهت چرخش نقاط نسبت به تصویر است و درواقع این ویژگی است که نقاط کلیدی نهایی را مقاوم و نابسته نسبت به چرخش شی در تصویر می‌کند. در ادامه اندازه گرادیان m(x،y) و جهت $\theta(x, y)$ را طبق عبارات ۴ و ۵ محاسبه می‌کنیم. هیستوگرام جهت چرخش با ۳۶ بین که هر بین برای ۱۰ درجه است را رسم می‌کنیم. پیک‌های این هیستوگرام، جهت‌های برتر را نشان می‌دهند.

$$m = \sqrt{((L(x+۱, y) - L(x-۱, y))^۲ + (L(x, y+۱) - L(x, y-۱))^۲} \qquad (۴)$$

$$\theta = atan۲(L(x, y+۱) - L(x, y-۱), L(x+۱, y) - L(x-۱, y)) \qquad (۵)$$

در این مرحله که مرحله نهایی است باید هر نقطه کلیدی انتخاب شده را در مقابل چرخش، نور، تغییر اندازه و... مقاوم کرد. برای این‌کار در مجموع ۱۲۸ ویژگی(گاهی نیز ۶۴ ویژگی) برای هر نقطه کلیدی ساخته می‌شود و در یک بردار توصیف‌گر آن نقطه قرار داده می‌شود. برای مثال به‌منظور ایجاد مقاومت در برابر چرخش برای هر نقطه کلیدی از هیستوگرام چرخش که بر روی همسایگی ۴در ۴ و با ۸ بین تعریف می‌شود و ۱۶ ویژگی چرخش را می‌سازد استفاده می‌شود. شکل۴ نمایش نحوه الگوریتم را به‌خوبی نشان می‌دهد. همچنین در شکل‌های ۵و۶ ویژگی‌های استخراج شده از تصویر نمونه و یک تصویر برای تطبیق قابل مشاهده است. دقت داشته باشیم همان‌طور که مشخص است استفاده تنها از الگوریتم sift قطعا باعث تطابق تصویر و نمونه نخواهد شد و باید به‌کمک تطبیق روش‌های بررسی شده و روش پیشنهادی به یک راه‌حل

(آ) انتخاب نقاط ابتدایی

(ب) حذف نقاط با کنتراست پایین

(ج) حذف نقاط مغشوش و پاسخ‌های ضعیف در کنار لبه‌ها

شکل ۴: نحوه کارکرد الگوریتم sift

مناسب رسید.

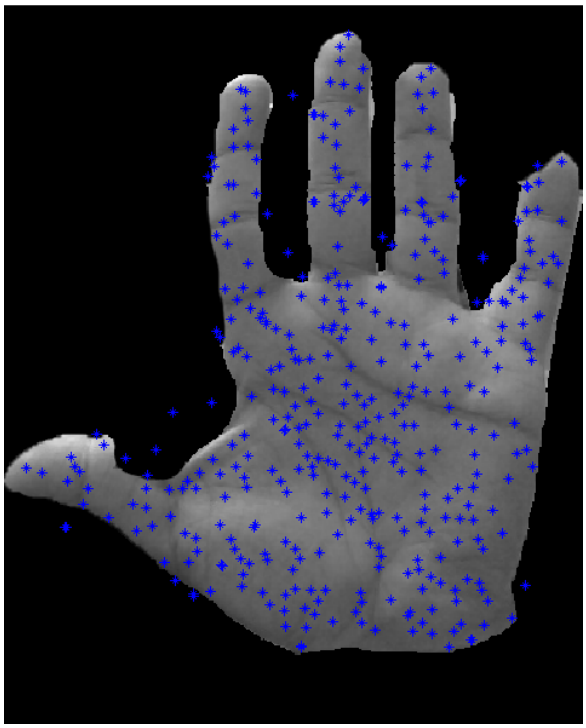
۳. نتایج پیاده‌سازی

همان‌طور که در بخش‌های پیشین بررسی و مشاهده شد. از روش تشخیص پوست به‌منظور تشخیص پوست دست انسان استفاده می‌شود. چون این روش به‌تنهایی قادر به تشخیص دست انسان و تمایز آن از باقی اعضا مانند صورت نیست، روش SIFT به‌منظور تطبیق ویژگی‌های دست(از جمله مچ، انگشتان و دست باز) با تصاویر ویدئویی به‌صورت برلحظه پیشنهاد شده‌اند. همان‌طور که در شکل‌های ۵و۶ مشاهده کردید، استفاده تنها از Sift نیز مانند استفاده تنها از روش تشخیص رنگ پوست جواب مناسبی نخواهد داد. بنابراین در این مرحله پیشنهاد ترکیب این دو روش را می‌دهیم. در واقع انتظار داریم با ترکیب این دو روش پس از جداسازی بدن انسان از تصویر با پس‌زمینه شلوغ بتوان به‌کمک SIFT نمونه دست را با دست‌های موجود در تصویر تطبیق داد.

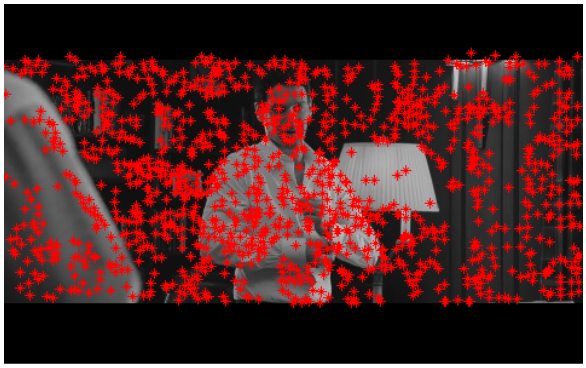
نتایج حاصل از روش پیشنهادی بر روی شکل۱ را می‌توان در شکل۷ مشاهده کرد. همان‌طور که در شکل۷ می‌توان مشاهده کرد و نقاط قرمز روی تصویر نشان‌دهنده نقاط تطبیقی بین نمونه دست شکل ۸ و تصویر شکل ۱ هستند. در واقع شکل۷ به‌خوبی نشان می‌دهند که باتوجه به نقاط قرمز این دو روش بررسی شده به‌تنهایی قادر به تشخیص دست انسان نیستند. در ادامه روشی به‌منظور حل این مشکل و بررسی و تشخیص دست انسان را در تصاویر ویدویی بررسی و تحلیل خواهیم کرد.

آ. روش پیشنهادی

پس از بررسی روش ترکیبی تشخیص پوست انسان و SIFT و مشاهده معایب آن، در این بخش سعی داریم روشی را ارائه دهیم که توسط آن بتوانیم با دقت بیشتری دست موجود در تصاویر ویدوئی را بررسی کنیم. به‌این منظور برای غلبه بر مشکلات به‌وجود آمده از مراحل قبل و باتوجه به وجود چندین فریم در هر ثانیه ضبط شده از هر دوربین، می‌دانیم که تغییرات تصویر در هر ثانیه و در چند ثانیه را به‌سادگی می‌توان بررسی و مشخص کرد. از این ایده به‌منظور تشخیص دست انسان سعی داریم



شکل ۵: نتیجه روش SIFT بر روی تصویر نمونه یک دست



شکل ۶: نتیجه SIFT بر روی یک تصویر نمونه

شکل ۷: تاثیر روش بررسی شده بر روی تصویر با پس زمینه شلوغ

(آ) تصویر نمونه انگشتان

(ب) تصویر نمونه دست باز

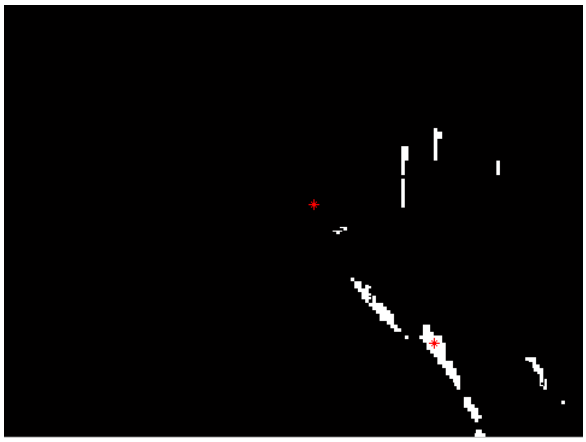
(ج) تصویر نمونه مچ دست

شکل ۸: نمونه‌های استفاده شده در این مستند به‌منظور تشخیص دست انسان در تصاویر ویدئویی

شکل ۹: تصویر حاصل از حذف زمینه براساس تفاضل دو فریم

استفاده کنیم. در واقع در روش پیشنهادی معتقدیم که دست انسان دارای تغییرات در یک ویدئو است که هرچند این تغییرات کوچک هستند ولی قابل بررسی هستند. این تغییرات به ما این قابلییت را می‌دهند که با از بین بردن بخش‌های ثابت تصویر به‌کمک روش تفاضل پس‌زمینه داده‌های مفیدی نظیر حرکت انسان و بدن انسان را از تصویر بیرون بکشیم. به‌این ترتیب قادر خواهیم بود با تصویر ساده‌تری که فقط تغییرات ساده بدن انسان را ذخیره کرده‌اند سروکار داشته باشیم. روش پیشنهادی در مرحله اول، تغییرات تصویر نسبت به صحنه قبل را محاسبه کرده و سپس دست به تشخیص پوست انسان می‌زند. بدیهی است در این شرایط بار محاسباتی کمتری متوجه سیستم برای پردازش و تشخیص دست در تصویری است که تفاضل آن از تصویر قبلی محاسبه شده است. شکل ۹ تغییرات ثبت شده بین دو فریم یک تصویر را نشان می‌دهد. همان‌طور که مشاهده می‌کنید در شکل۹، تطبیق شکل دست باقی‌مانده در تصویر به‌دلیل ثابت بودن محیط بسیار ساده‌تر و راحت‌تر است.

شکل ۱۰، نتیجه حاصل از تطبیق تصویر و نمونه را نمایش می‌دهد.



شکل ۱۰: تصویر حاصل از حذف زمینه براساس تفاضل دو فریم

شکل ۱۱: .(تصویر سمت راست تصویر اصلی و تصویر سمت چپ پیش‌بینی برنامه است)تشخیص دست، فلش آبی نشانگر نقطه قرمز شناسایی شده توسط برنامه است

نقاط قرمز نشانگر تطابق تصویر و نمونه هستند. نکته‌ای که لازم است ذکر شود این‌است که، تعداد نقاط یافت شده توسط روش SIFT برای تطبیق بیشتر از تعداد مشخص شده در تصویر هستند ولی با قرار دادن حد آستانه که براساس معیار فاصله اقلیدسی است، نقاط تطبیقی ضعیف را از تصویر حذف کرده‌ایم تا مشکل تشخیص اشتباه رخ ندهد.

ب. بررسی و نمایش نتایج

به‌منظور بررسی برلحظه ویدئویی دریافتی و یا کار بر روی تصاویر گرفته شده از دوربین برخط، تغییرات اندکی در روند اجرای الگوریتم پیشنهادی قرار دادیم. در واقع پردازش هر فریم با وجود حتی ۱۰ فریم بر ثانیه، کار پردازش و تشخیص دست انسان را بسیار کند خواهد کرد. در نتیجه به ازای هر ثانیه دو یا سه فریم را بررسی می‌کنیم. آزمایشات نشانگر آن‌است که حتی بررسی این تعداد فریم در مقایسه با حالت قبل نمایش‌دهنده تغییرات بدن انسان هستند و منجر به تشخیص دست خواهند شد. نکته لازم به ذکر دیگر آن‌است که، به این دلیل که پس از گذشت حداقل دو بررسی (دو فریم و نزدیک به یک ثانیه) اقدام به تشخیص دست می‌کنیم، شاید به‌نظر برسد که عملکرد روش پیشنهاد شده کند است. البته با دقت به این‌که به حداقل ۲ فریم نیازمندیم که کمتر از یک ثانیه خواهد بود، متوجه خواهیم شد، تاخیر ایجاد شده قطعا قابل چشم‌پوشی است. همچنین برنامه توانایی این‌را دارد که اگر در ۳ یا ۴ فریم متوالی یکی در میان دست را تشخیص داد به‌دلیل اشکالات موجود در تشخیص رای به حضور دست بدهد در صورتی که حداقل به تعداد نصف به‌علاوه یک‌بار دست را تشخیص داده باشد. در ادامه نتایج حاصل از کار را می‌توان مشاهده کرد.