

روش های اخیر و پایگاه های داده در تشخیص حرکات و اشارات (ژست) دست مبتنی بر

بینایی: یک بررسی

چکیده

تلاش های موفق در تحقیقات تشخیص حرکات و اشارات دست در طی دو دهه گذشته، راه را برای سیستم های تعامل طبیعی انسان-کامپیوتر هموار کرده است. چالش های حل نشده مانند شناسایی قابل اعتماد فاز ژست گرفتن، حساسیت به اندازه، شکل و تغییرات سرعت و مسائل ناشی از انسداد، تحقیقات در مورد تشخیص ژست دست را هنوز بسیار فعال نگهداشته است. ما یک بررسی از الگوریتم های تشخیص ژست دست مبتنی بر بینایی که در 16 سال گذشته گزارش ارائه می دهیم. روش های استفاده کننده از دوربین های RGB و RGB-D با مقایسه های کمی و کیفی الگوریتم ها بررسی می شوند. مقایسه کمی الگوریتم ها با استفاده از مجموعه ای از 13 معیار انتخاب شده از صفات مختلف الگوریتم و روش تجربی اتخاذ شده در ارزیابی الگوریتم انجام می شود. ما به لزوم در نظر گرفتن این معیارها همراه با دقت تشخیص الگوریتم برای پیش بینی موفقیت آن در کاربردهای دنیای واقعی اشاره می کنیم. این مقاله همچنین به بررسی 26 پایگاه داده ژست دست در دسترس عموم و ارائه لینک های وب برای دانلود آنها می پردازد.

کلید واژه ها: تشخیص ژست، تشخیص وضعیت، تخمین ژست دست، تشخیص زبان اشاره، پایگاه داده ژست دست،

پایگاه داده ژست، نظرسنجی

1. مقدمه

ارتباطات غیرشفاهی که شامل ارتباطات از طریق حرکات دست، حالات بدن و بیان های چهره می شود، حدود دو سوم کلیه ارتباطات بین انسان ها را تشکیل می دهد [1]. حرکات دست یکی از رایج ترین انواع زبان بدن هستند که برای ارتباط و تعامل استفاده می شوند. در حالی که بقیه بدن، نشان دهنده یک حالت عمومی تر عاطفی است، حرکات دست می توانند محتوای زبان خاص خود را در آن داشته باشند (2). با توجه به سرعت و ظرافت در تعامل، حرکات دست به طور گسترده در زبان های اشاره و سیستم های تعامل انسان-کامپیوتر استفاده می شود.

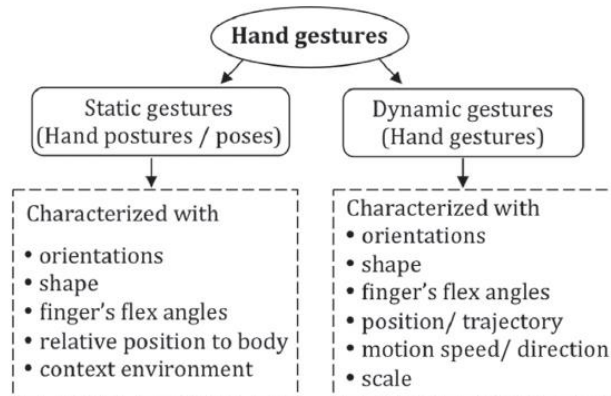
یک هدف مداوم در طراحی واسطه انسان-دستگاه، میسر ساختن تعامل موثر و درگیرانه است. به عنوان مثال، سیستم های تشخیص حرکت دست مبتنی بر بینایی (HGR) می توانند تعامل بدون تماس در محیط های استریل مانند اتاق های جراحی بیمارستان را میسر سازند یا به سادگی کنترل های درگیرانه برای برنامه های سرگرمی و بازی را فراهم کنند. با این حال HGR به اندازه تعامل مبتنی بر صفحه کلید و موس استاندارد، قوی نیست. مسائلی مانند حساسیت به اندازه و تغییرات سرعت، عملکرد ضعیف در برابر پس زمینه های پیچیده و شرایط مختلف نوردهی و تشخیص قابل اعتماد از فاز ژست، استفاده از حرکات دست را به عنوان یک روش معتبر در طراحی واسطه محدود کرده اند.

1.1 طبقه بندی ژست ها

روش های متعددی برای دسته بندی حرکات (ژست های) دست وجود دارد: (1) بر اساس ویژگی های قابل مشاهده و (2) بر مبنای تفسیر. در رده اول، حرکات براساس روابط زمانی، به دو نوع طبقه بندی می شوند: حرکات ایستا و پویا (شکل 1). حرکات استاتیک دست (مثلاً حالت های دست/ ژست های دست)، حرکاتی هستند که در آن، موقعیت دست در طول دوره ژست گیری تغییر نمی کند. حرکات استاتیک عمدتاً بر روی شکل و زوایای خمش انگشتان تکیه می کنند. در حرکات پویای دست، موقعیت دست به طور مداوم با توجه به زمان تغییر می کند. حرکات پویا عموماً سه مرحله حرکتی دارند: آماده سازی، ضربه، و انقباض [3]. پیام در یک ژست پویا عمدتاً در دنباله زمانی در مرحله

ضربه گنجانده می شود. حرکات پویا بر روی مسیرها و جهت های دست، به علاوه شکل و زاویه های خمش انگشت تکیه می کنند.

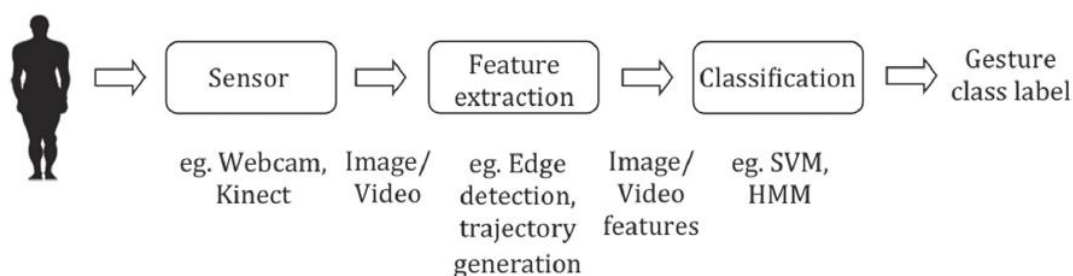
در رده دوم، حرکات بر اساس معنای تفسیری طبقه بندی می شوند. به عنوان مثال، نمادها، تصویرگرها، تنظیم کننده ها، نمایشگرها و آداپتورها [4، 5]، کلاس های معمول برای توصیف حرکات هستند. نشان ها (که همچنین به عنوان حرکات خود به خودی نامگذاری می شوند) حرکاتی هستند که می توانند برای کلمات گفتاری جایگزین شوند (به عنوان مثال، نشان دادن انگشت های شست رو به بالا به جای گفتن اینکه: بسیار خوب). تصویرگرها، ژست های مورد استفاده برای نشان دادن کلمات گفته شده هستند (مثلاً جهت دادن با اشاره). تنظیم کننده ها از تعامل و ارتباط بین سخنران و شنونده (به عنوان مثال، بالا بردن دست برای مدیریت نوبت گرفتن) پشتیبانی می کنند. نمایش های اثر، ابزارات چهره ای هستند که در هنگام ترکیب با حالت ها، شدت یک احساس را نشان می دهند (به عنوان مثال، برجسته کردن یک شی و حرکت بدن به پشت، منعکس کننده احساس ترس). تطبیق دهنده ها، حرکات مورد استفاده در برخی مواقع برای راحتی شخصی هستند اما به عادت تبدیل شده اند (برای مثال، تنظیم عینک ها در یک وضعیت تنگ شده).



شکل 1: طبقه بندی حرکات دست بر اساس ماهیت زمانی. حرکات استاتیک، مستقل از زمان هستند در حالیکه حرکات دینامیک، وابسته به زمان هستند.

1.2 تشخیص ژست دست

شکل 2، بلوک دیاگرام یک سیستم تشخیص ژست بدون تماس نوعی را نشان می دهد. سنسور، یک دوربین در سیستم های تشخیص ژست مبتنی بر بینایی است. Berman و همکاران [6]، سنسورهای مختلف مورد استفاده در سیستم های تشخیص حرکت را مورد بررسی قرار دادند و یک تحلیل جامع از ادغام سنسورها را در سیستم های تشخیص حرکت و تاثیر آنها بر عملکرد سیستم ارائه کردند. بر اساس استخراج ویژگی، سیستم های تشخیص ژست مبتنی بر بینایی به طور گسترده به دو دسته تقسیم می شوند، روش های مبتنی بر ظاهر و روش های مبتنی بر مدل سه بعدی دست (3D). روش های مبتنی بر ظاهر از ویژگی های تصویر آموزش برای مدلسازی ظاهر بصری استفاده می کنند و این پارامترها را با ویژگی های تصویر آزمون مقایسه می کنند. روش های سه بعدی مبتنی بر مدل با تخمین پارامترهای خطی و زاویه ای مدل، بر یک مدل سینماتیک سه بعدی تکیه می کنند.

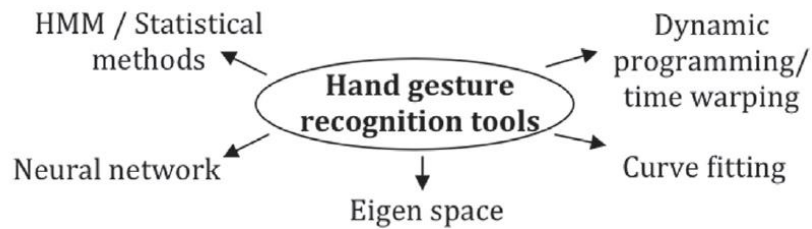


شکل 2. خط لوله شناخت ژست

1.3 بررسی و ارزیابی تکنیک های تشخیص ژست دست

مطالعه ما بر اساس تلاش های پیشین برای بررسی حوزه HGR استوار است. Mitra و همکاران [7]، یک پیمایش از روش های مختلف تشخیص ژست، شامل ژست های دست و بازو، ژست های سر و صورت و حرکات بدن را ارائه کردند. روش های HGR بررسی شده در این تحقیق محدود به مدل های مخفی مارکوف (HMM)، الگوریتم های فیلتر کردن ذرات و تراکم و شبکه های عصبی مصنوعی (ANN) بودند. روش های برآورد ژست مبتنی بر مدلسازی دست و مبتنی بر حرکت سه بعدی در [8] (با نادیده گرفتن طرح های طبقه بندی ژست) مورد بررسی قرار گرفته اند. یک تجزیه و تحلیل از زبان های نشانه، فرآیندهای دستوری در ژست های نشانه و مسائل مربوط به شناخت خودکار زبان های نشانه در [9] بحث شده اند. آخرین مقالات فوق (8)، توسعه ها تا ژانویه 2005 را پوشش دادند.

این بررسی نتیجه گیری کرد که روش های مورد مطالعه، آزمایشگاهی هستند و استفاده از آنها محدود به محیط های آزمایشگاهی می شود.



شکل 3. طبقه بندی تکنیک های بررسی شده شناخت ژست دست

با تمرکز بر تحولات در 16 سال گذشته، این مقاله به بررسی کارهای اخیر در HGR می پردازد. الگوریتم های استفاده کننده از دوربین های RGB معمولی (بخش 2) و همچنین دوربین های RGB-D نسل جدید (بخش 3) مورد بررسی قرار می گیرند که این بررسی را منحصر به فرد می سازند. روش های HGR، با توجه به روش استفاده شده برای طبقه بندی ژست طبقه بندی و تجزیه و تحلیل می شوند. ما یک مقایسه کمی از الگوریتم های HGR را بر اساس صفات مختلف الگوریتم و روش آزمایشگاهی دنبال شده در آزمایش الگوریتم انجام می دهیم. یک بازنگری در پایگاه های داده ژست دست (بخش 4) و بحث در مورد تحقیقات تشخیص ژست دست (بخش 5) نیز ارائه شده است. ما امیدواریم که این بررسی، با توجه به تلاش های تحقیقاتی روز افزون و گسترش بازار برای سیستم های تعاملی ژست به موقع باشد.

2. تشخیص مرسوم ژست دست: روش های مبتنی بر سنسور RGB

2.1 تشخیص ژست های پویای دست

تکنیک های مورد استفاده برای HGR پویا را می توان به صورت زیر دسته بندی نمود: الف) [10-23] HMM و سایر روش های آماری [24-31]، ب) [32-34] ANN و سایر روش های مبتنی بر یادگیری [35، 36]، ج) [37]

[38, d] روش های مبتنی بر فضای ویژه، تناسب منحنی [39]، و (e) برنامه نویسی پویا [40] / انحراف زمان پویا [41, 42] (شکل 3).

2.1.1 HMM و سایر روش های آماری

HMM پر استفاده ترین تکنیک HGR است. HMM یک مدل آماری است که در آن سیستم مدلسازی شده به عنوان یک فرآیند مارکوف با پارامترهای مجهول فرض می شود. HMM، نشان دهنده رفتار آماری توالی نماد قابل مشاهده با استفاده از شبکه ای از حالات پنهان با احتمالات انتقال و انتشار است. HMM را زمانی می توان برای تشخیص الگو استفاده کرد که پارامترهای پنهان با استفاده از داده های قابل مشاهده شناسایی شوند.

روشهای پویای تشخیص ژست مبتنی بر HMM عمدتاً از ویژگیهای زمانی و فضایی تصاویر ورودی استفاده می کنند. Chen و همکاران [14] از توصیفگر فوریه و تحلیل حرکت مبتنی بر جریان اپتیک به ترتیب به منظور مشخص کردن ویژگی های فضایی و زمانی استفاده نمودند. این الگوریتم با استفاده از ردیابی دست در زمان واقعی، شکل دست را از پس زمینه های پیچیده استخراج می کند. شناساگرهای مبتنی بر HMM، بهترین مدل ژست احتمال را برای یک الگوی معین شناسایی می کنند. تغییرات در ژست از یک الگوی مرجع، احتمال این ژست را با مدل کاهش می دهند.

Lee و Kim [10] یک مفهوم مدل آستانه مبتنی بر HMM را برای فیلتر کردن الگوها با احتمال کمتر معرفی کردند. جهت حرکت دست برای نشان دادن توالی های فضایی-زمانی ژست ها استفاده می شود. این روش به طور قابل اعتماد نقطه پایانی یک ژست را تشخیص می دهد و نقطه شروع را با ردیابی به عقب می یابد.

HMM مبتنی بر زنجیره های مارکف همگن است زیرا پویایی سیستم تنها توسط احتمالات گذار مستقل زمان تعیین می شود. Marcel و همکاران [15] یک فرمت HMM، یعنی مدل ورودی / خروجی مخفی مارکوف (IOHMM) را برای HGR پیشنهاد کردند. IOHMM مبتنی بر یک زنجیره غیرهمگن مارکوف است که احتمالات انتقال و انتشار آن به ورودی بستگی دارد. IOHMM یاد می گیرد تا توالی های ورودی، مشاهدات، توالی های

خروجی و کلاس های ژست را برای همه مشاهدات با استفاده از یادگیری اختیاری تحت نظارت یاد بگیرد. در مقایسه با HMM ها، IOHMM یک رویکرد تبعیض آمیز است زیرا به طور مستقیم احتمالهای خلفی را مدلسازی می کند. مطالعه در [15] محدود به 2 کلاس بود. Just و همکاران [13]، این مطالعه را برای شناختن حرکات تک و دوبل دست گسترش دادند و یک مقایسه از HMM و IOHMM ارائه کردند. آزمایشات انجام شده بر روی پایگاه های داده بزرگتر، اعم از 7 تا 16 کلاس ژست، نتیجه گرفتند که برای تعداد زیادی از کلاس ها، HMM عملکرد خوبی نسبت به IOHMM دارد.

ویژگی های موقعیت دست، زاویه و سرعت در [11] برای پیاده سازی HMM در HGR ترکیب می شوند. دست توسط تجزیه و تحلیل رنگ-پوست موقعیت یابی می شود و با اتصال مرکز حرکت ناحیه های دست ردیابی می شود. این مقاله، مطلوبیت سه ویژگی، موقعیت، زاویه و سرعت را مقایسه کردند و نتیجه گرفتند که ویژگی های زاویه ای، موثرترین هستند و دارای قدرت تشخیص بهتر هستند. ویژگی های مکان و سرعت، به ترتیب رتبه دوم و سوم را دارند. یک پیاده سازی HMM مشابه با استفاده از زوایای حرکت در امتداد مسیر مرکز دستی در [16] ارائه شده است.

Ramamoorthy و همکاران توسط ترکیب طرح مشخص سازی زمانی مبتنی بر HMM با یک سیستم تشخیص شکل استاتیک [12]، یک سیستم HGR را توسعه دادند. آنها از یک ردگیر خط دستی مبتنی بر فیلتر کالمن استفاده کردند که ویژگی های زمانی ژست را ارائه می کند. اشکال با استفاده از طبقه بند مبتنی بر متمایزکننده خط شناخته می شوند. این توصیفگرهای نمادین ژست ها برای آموزش HMM استفاده می شوند. این سیستم می تواند به طور قابل اطمینان، ژست های دینامیک را به رغم تغییرات حرکت و گسسته در ژست های دست تشخیص دهد. همچنین این الگوریتم توانایی تشخیص نقاط شروع و پایان توالی های ژست را دارد.

الگوریتم های تشخیص ژست پویا از یک طرح نقطه یابی رو به عقب استفاده می کنند که ابتدا نقطه پایان یک ژست را تشخیص می دهد و سپس به سمت نقطه شروع ردیابی می شود. Kim و همکاران [17] یک روش جایگزین، یک طرح نقطه گذاری مستقیم را پیشنهاد کردند که به طور همزمان بخش بندی و تشخیص ژست را اجرا می کند. نقاط

شروع و پایان ژست ها توسط عبورهای از صفر احتمال دیفرانسیل سیگنال شناسایی می شوند. مجموعه ای از ویژگی های مبتنی بر تشریح سه بعدی با استفاده از روش نگاشت ارتباطی استخراج می شوند که داده های شکل دو بعدی را به داده های مفصل سه بعدی متصل می کند. ژست ها با یک رای اکثریت با استفاده از HMM تجمعی طبقه بندی می شوند.

Davis و Shah [31] ژست ها را به چهار مرحله متمایز تقسیم نمودند که در یک نظم ثابت اتفاق می افتد و یک مدل ماشین حالت متناهی (FSM) را برای تشخیص توسعه دادند. نشانه زمانی حرکت دست استخراج می شود و ژست های دست با استفاده از FSM در [29] مدل سازی می شود. مفهوم انرژی حرکت برای تخمین حرکت غالب از یک توالی تصویر استفاده می شود. Hong و همکاران [30] از موقعیت های دوبعدی مراکز سر و دستان افراد برای توسعه FSM استفاده نمودند. مدل پویای شبکه بیزی در [24] برای تشخیص ژست های جدا شده و پیوسته دست ارائه شده است. ویژگی های استفاده شده عبارتند از کدهای جهت برای حرکت دست، رابطه موضعی بین دو دست و رابطه موقعیتی بین صورت و دست.

Chen و همکاران [25] از یک روش دو مرحله ای تجزیه و تحلیل آماری و نحوی به ترتیب برای تشخیص ژست های دستی استاتیک و پویا پیشنهاد کردند. سطح اول، تجزیه و تحلیل آماری، بر اساس یک الگوریتم یادگیری AdaBoost و ویژگی های Haar-مانند است. سطح دوم، تجزیه و تحلیل نحوی، مبتنی بر یک گرامر تصادفی عاری از زمینه (SCFG) است. ویژگی های Haar-مانند به طور موثر الگوی وضعیت دست را توصیف می کنند و الگوریتم AdaBoost, یک طبقه بند قوی را با ترکیب یک دنباله ای از طبقه بندهای ضعیف ایجاد می کند. موقعیت های شناسایی شده توسط سطح اول به دنباله ای از رشته های ترمینال با توجه به دستور زبان، در مرحله دوم تبدیل می شوند.

ANN 2.1.2 و سایر روشهای مبتنی بر یادگیری

Yang و همکاران [32، 33] از یک شبکه عصبی تاخیر زمانی (TDNN) برای یادگیری مسیرهای حرکت دو بعدی استفاده کردند. TDNN، یک شبکه چند لایه تغذیه-مستقیم است که از پنجره های جابجایی بین تمام لایه ها برای نشان دادن روابط زمانی بین وقایع استفاده می کند. طبقه بندی در TDNN، پویا است، زیرا این شبکه، تنها یک پنجره کوچک از الگوی حرکت ورودی را می بیند و پنجره بر روی داده های ورودی لغزش می کند در حالی که شبکه، یک سری تصمیمات موضعی می گیرد. این تصمیمات موضعی به صورت زمانی با یک تصمیم کلی در لایه خروجی یکپارچه می شوند.

الگوریتم های حرکت مبتنی بر ناحیه همانند [32] از روش های مبتنی بر شدت بهتر عمل می کنند. به عنوان مثال، اطلاعات حرکت در مناطقی با تغییر شدت کم در خطوط نواحی مرتبط گنجانده می شود. الگوریتم تقسیم بندی حرکت، تناظرها را برای چنین نواحی محاسبه می کند و بهترین تبدیل نسبت را پیدا می کند که تغییر شکل خط را تشخیص می دهد. پارامترهای تبدیل وابسته برای ناحیه در مقیاس های مختلف برای ایجاد یک میدان حرکت تک استفاده می شوند که سپس برای شناسایی نواحی متحرک بین دو فریم تقسیم بندی می شوند.

Chan و همکاران [34] ترکیبی از HMM و شبکه های عصبی بازگشت کننده (RNN) را پیشنهاد کردند که عملکردی بهتری را نسبت به HMM یا RNN به تنهایی ارائه دادند. ویژگی های مورد استفاده شکل بر اساس توصیفگرهای فوریه هستند که ورودی ها به شبکه تابع پایه شعاعی (RBF) برای یک طبقه بندی ژست اولیه هستند. بردار احتمال ژست از شبکه RBF همراه با اطلاعات حرکت، ورودی به دو طبقه بند مستقل، HMM و RNN است. خروجی ها از طبقه بندها به صورت خطی برای پیش بینی کلاس ژست ترکیب می شوند.

Shen و همکاران [35] روشی مبتنی بر نمونه را برای تشخیص ژست پیشنهاد کردند. ژست های دست با استفاده از میدان واگرایی حرکات جریان دست نشان داده می شوند. میدان های واگرایی جریان نوری بین فریم های متوالی تصویر استنتاج می شوند و مناطق برجسته از میدان واگرایی با استفاده از یک آشکارساز ویژگی نواحی خارجی پایدار حداکثر (MSER) شناسایی می شوند. توصیفگرها از هر منطقه تشخیص داده شده برای مشخص کردن الگوهای حرکت موضعی استخراج می شوند. دنباله های ژست پایگاه داده با توصیفگرهای آنها توسط یک واژگان سلسله

مراتبی از پیش آموزش دیده نشان داده می شوند. یک توالی ژست جدید با تطبیق آن با پایگاه داده تشخیص داده می شود.

2.1.3 روش مبتنی بر فضای-ویژه

Patwardhan و [37] Roy یک چارچوب مبتنی بر فضای ویژه را پیشنهاد دادند تا ژست های پویای دست شامل اطلاعات شکل و مسیر را مدلسازی نمایند. روش های مبتنی بر ویژگی شامل یک زمان جداگانه مصرف شده برای مرحله تشخیص ویژگی است که در این الگوریتم از آن اجتناب می شود. الگوریتم در برابر تغییر شکل های معمولی دست: چرخش، انتقال، مقیاس و برش، نامتغیر است.

2.1.4 تناسب منحنی

Shin و همکاران [39] یک روش هندسی با استفاده از منحنی های Beizer را برای تحلیل مسیر و طبقه بندی ژست های پویا پیشنهاد کردند. ژست های با استفاده از تناسب منحنی با مسیر حرکت سه بعدی دست شناخته می شوند. سرعت حرکت در الگوریتم گنجانده شده است تا امکان شناسایی دقیق مسیرهایی با تغییرات در سرعت فراهم شود.

2.1.5 برنامه نویسی پویا و انحراف زمان دینامیک

Kuremoto و همکاران. [40] یک رویکرد پویای تک-مرحله ای مبتنی بر برنامه نویسی پویا برای تشخیص حرکت پیشنهاد کردند. یک سیستم استخراج ویژگی بیولوژیکی مبتنی بر مدل شبکه چشم-1V پیشنهاد شده توسط Tohyama و Fukushima [43]، حرکت دست را برآورد می کند. ژست های دست به عنوان ترکیبی از قالب های ژست های ساده در نظر گرفته می شوند. جنبش ها برای ساخت یک مجموعه از 40 قالب ژست مورد استفاده قرار می گیرند.

| Work | Features used | Classification method | Application |
|------|----------------------------------------------------------------|-----------------------------|---------------------------------------------------------------|
| [10] | Direction of hand movement | HMM | Browsing commands in PowerPoint ^(R) presentation |
| [11] | Hand location, angle and velocity | HMM | HCI- recognizing alphanumeric characters and graphic elements |
| [14] | Fourier descriptors/optical flow | HMM | Taiwanese sign language |
| [12] | Hand shape and hand motion | HMM | Remote robot control |
| [17] | 3D articulation data | Accumulative HMM | Controlling lights and curtains in smart home |
| [13] | 3D trajectory, hand displacement, color and shape of hand blob | HMM and IOHMM | Interact-play, manipulation |
| [25] | Haar-like features | Statistical/syntactic anal. | Not specified |
| [24] | Directional features | DBN | Controlling media player |
| [39] | 3D motion trajectory | Curve fitting | 3D bioinformatics data visualization navigation |
| [37] | Hand shape/trajectory | Predictive eigen tracker | Audio player control |
| [32] | 2D motion field/trajectory | NN | American sign language |
| [34] | Fourier descriptors (shape of hand blob) | RBF, HMM and RNN | Manipulation of objects in windows user interface |
| [29] | Hand motion (motion energy) | FSM | HRI |
| [42] | 3D hand motion features | CDFD and Q-DFFM | Dutch sign language |

جدول 1: روش های تشخیص ژست دست: ویژگی های مورد استفاده، روش های طبقه بندی و برنامه های گزارش

شده

| Work | Accuracy | Class | Subject | Sample | UI | Spot | BG | Noise | Scale | Light | Extensibility | CV | Data |
|------|--------------|----------|----------|-----------|----|------|----|-------|-------|-------|---------------|----|------|
| [10] | 93.14 | 10 | 8 | 6.2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| [11] | 93.25 | 48 | 20 | 5 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| [14] | 93.6 | 20 | 20 | 3 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| [12] | 81.71 | 5 | 5 | 14 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| [17] | 95.42 | 8 | 1 | 60 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| [13] | 75 and 98 | 16 and 7 | 20 and 7 | 50 and 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| [25] | 87.21 | 4 | 1 | 25 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| [24] | 99.59 | 10 | 7 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| [39] | 97.9 | 10 | 4 | 2.38 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| [37] | 100 | 8 | 1 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| [32] | 96.21 | 40 | 1 | 7.6 | 0 | 1 | 1 | | 0 | 1 | 0 | 1 | 0 |
| [34] | 91.9 | 14 | 1 | 21.07 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| [29] | Not reported | 5 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| [42] | 92.3 | 120 | 75 | 15 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |

جدول 2: روش های تشخیص ژست دست: ویژگی های الگوریتم ها و روش تجربی اتخاذشده در آزمایش الگوریتم

(فهرست پایین، توضیح عناوین ستون را ارائه می دهد). ویژگی های ستون 6 به ترتیب باینری هستند، 1 نشان

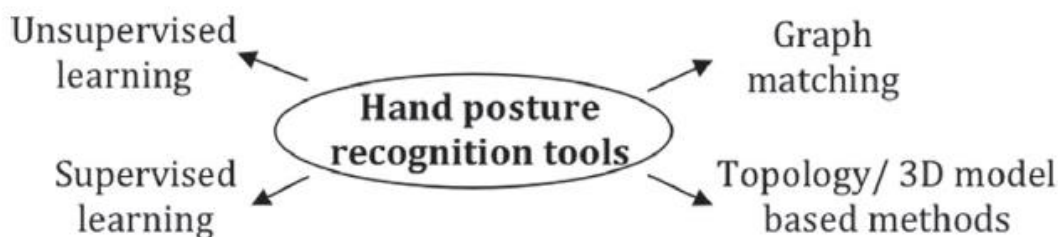
دهنده سازگاری کار با این ویژگی است، در حالی که 0 نشان دهنده عدم انطباق است.

انحراف پویای زمانی (DTW)، یک کاربرد برنامه نویسی پویا، به طور گسترده ای در شناخت ژست های جداگانه مورد استفاده قرار گرفته است. آندریا کورادینی [41]، با محاسبه یک تبدیل زمانی بین دو سیگنال که باید تطبیق یابند، یک رویکرد مبتنی بر الگو را با DTW برای هماهنگ سازی و نرمال سازی زمان پیشنهاد کرد. Lichtenauer و همکاران [42]، DTW آ (SDTW) را برای انحراف زمان و دو طبقه بند، یعنی آشکارسازهای ترکیبی ویژگی متمایز (CDFD) و طبقه بندی درجه دوم بر اساس نگاشت ویژگی های تشخیصی فیشر (Q-DFFM)، برای طبقه بندی پیشنهاد نمودند. طبقه بندها نشان می دهد که HMM و SDTW بهتر عمل می کنند.

یک خلاصه و مقایسه ویژگی های الگوریتم های تشخیص ژست دست در این بخش در جداول 1 و 2 ارائه شده است.

2.2 تشخیص حالات دست

روش های تشخیص حالت دست بررسی شده بدین صورت طبقه بندی می شوند: الف) روش های مبتنی بر یادگیری تحت نظارت [35, 44-60]، ب) روش های مبتنی بر یادگیری بدون نظارت [61]، ج) تطابق نمودار [62-67]، و د) روش های مبتنی بر مدل [68-723] D (شکل 4).



شکل 4: طبقه بندی تکنیک های تشخیص چهره دست بررسی شده.

2.2.1 یادگیری بدون نظارت

یک الگوریتم تعبیه خطی توزیع شده موضعی (DLLE) در [61] برای تشخیص وضعیت دست و ردیابی پویای ژست پیشنهاد شده است. تعبیه خطی موضعی [73] (LLE)، یک الگوریتم یادگیری بدون نظارت است که برای نگاشت داده ها با ابعاد بالا برای فضای کم-ابعاد تلاش می کند و در عین حال رابطه همسایگی را حفظ می کند. این مقاله، LLE را به DLLE تغییر داد تا خصوصیات ذاتی داده های ورودی را کشف کند، با توجه به این که برخی از قطعات اطلاعات مربوطه توزیع می شوند. DLLE، ساختار ذاتی داده هایی مانند رابطه همسایگی را استخراج می کند. فواصل بین نقاط داده های تصویرشده در فضای کم-ابعاد به شباهت تصاویر ورودی بستگی دارد. یک شبکه عصبی احتمالی (PNN) برای طبقه بندی وضعیت های مختلف بر اساس فواصل در فضای کم-ابعاد استفاده می شود. PNN دارای سرعت آموزش خوب و دقت طبقه بندی با زمان آموزش مجدد ناچیز است.

شکل 4: طبقه بندی تکنیک های تشخیص چهره دست بررسی شده است.

2.2.2 یادگیری نظارت شده

یادگیری نظارت شده در الگوریتم LLE در [52] برای تشخیص وضعیت ها در زبان نشانه گذاری چینی (CSL) معرفی می شود. LLE تحت نظارت (SLLE) استفاده از اطلاعات برچسب کلاس را در طول آموزش طبقه بند میسر می سازد. دست با استفاده از رنگ پوست شناسایی می شود و هندسه ذاتی دست برای تشخیص استفاده می شود. Zhao و همکاران [51]، یادگیری استنتاجی بازگشتی را بر اساس منطق ارزش-متغیر گسترش یافته برای تشخیص وضعیت دست پیشنهاد نمودند. در یادگیری استنتاجی، دانش توسط استنتاج قوانین از مجموعه ای از نمونه ها و یا مجموعه ای از بردارهای ویژگی به دست می آید. این مقاله، مفهوم قدیمی منطق ارزش-متغیر را در منطق ارزش-متغیر گسترش یافته (EVL) اصلاح نمود که یک نمایش قدرتمندتر را ارائه داد. یک الگوریتم اکتشافی یعنی RIEVL (استنتاج قانون با استفاده از منطق ارزش-متغیر گسترش یافته) برای یادگیری قوانین از نمونه ها و همچنین مجموعه قوانین پیشنهاد شده است. RIEVL قوانین جمع و جور تر را نسبت به سایر الگوریتم های استنتاجی تولید نمود. این قابلیت اجازه می دهد تا مجموعه ای بزرگ از ویژگی ها در هنگام آموزش برای وضعیت های دست به کار برده شود و استنتاج یک مجموعه قانون کاهش یافته با یک زیر مجموعه از ویژگی های آموزشی در هنگام تشخیص میسر می شود. این الگوریتم به طور خودکار مؤثرترین ویژگی ها را انتخاب می کند که آن را برای سیستم های تشخیص حرکت واقعی مناسب می سازد.

یک مشکل رایج روشهای مبتنی بر آموزش، وابستگی آنها به داده های آموزشی است. به منظور افزایش عمومیت و استقلال کاربر، Licsar و Sziranyi [49]، [50] یک سیستم تشخیص حالت دست تطبیقی-کاربر را با آموزش آنلاین تعاملی پیشنهاد دادند. اگر دقت تشخیص کاهش یابد و سازگاری سریع با کاربران جدید تحقق یابد، سیستم برای حالات آشکار شده معیوب حفظ می شود. یک روش آموزش تحت نظارت برای کلاس های وضعیت تشخیص داده نشده اصلاح می شود و یک روش بدون نظارت به طور پیوسته اجرا می شود تا تغییرات جزئی در سبک های وضعیت اجرا شود.

یک راه حل برای مشکل پس زمینه پیچیده در تشخیص وضعیت دست در [44] ارائه شده است. این الگوریتم می تواند پس زمینه ها، از جمله پس زمینه های پیچیده رنگی-پوست را حل و فصل نماید. این سیستم از یک مدل

بیزی توجه بصری برای تولید یک نقشه برجستگی، آشکارسازی، شناسایی و تقسیم کردن ناحیه دست از پس زمینه های پیچیده استفاده می کند. توجه بصری مبتنی بر ویژگی با استفاده از ترکیبی از ویژگی های تصویر با سطح بالا (شکل، بافت) و ویژگی های سطح پایین (رنگ) انجام می شود. وضعیت های تقسیم بندی شده دست با استفاده از ویژگی های شکل و بافت با طبقه بند دستگاہ بردار پشتیبان (SVM) طبقه بندی می شوند.

Huang و همکاران [58] از یک الگوریتم برای تشخیص وضعیت دست را با توجه به شرایط حالت و روشنایی متغیر مطرح کردند. عدم تغییر له شرایط نوردهی با استفاده از یک روش تعویض مدل رنگ پوست به دست می آید. عدم حساسیت به تغییرات وضعیت دست با استفاده از یک روش اصلاح و برآورد زاویه حالت مبتنی بر فیلتر گابور به دست می آید. این وضعیت با استفاده از طبقه بند SVM طبقه بندی می شود.

2.2.3 الگوریتم های گراف

با شروع از اواخر دهه 1970، تکنیک های مبتنی بر گراف به عنوان یک ابزار قدرتمند برای نمایش و طبقه بندی الگو استفاده می شوند. پس از شور و شوق اولیه، الگوریتم های گراف عملاً برای مدت زمان طولانی استفاده نشدند. این به علت هزینه محاسبه بالای الگوریتم های گراف است که هنوز یک مشکل حل نشده باقی مانده است. با این حال، استفاده از گراف ها در بینایی کامپیوتری و تشخیص الگو به تازگی به دلیل توجه محاسباتی از الگوریتم های مبتنی بر گراف در حال حاضر با توجه به محاسبات قدرت کامپیوترهای نسل جدید، افزایش یافته است [74].

تطبیق گراف الاستیک (EGM)، که یک نوع تطبیق گراف است، یک معماری تشخیص الگوی الهام بخش عصبی است [75]. EGM دارای توانایی ذاتی برای حل و فصل اعوجاجات هندسی است، نیازی به یک تصویر ورودی کاملاً تقسیم بندی شده ندارد و می تواند به طور ظاهری تغییرات در ظاهر شیء را نشان دهد [63].

ناحیه های تصویر توسط رأس ها در یک نمایش گراف نشان داده می شوند. این رأس ها توسط لبه ها به یکدیگر مرتبط می شوند که بیانگر روابط ساختاری بین نواحی است. Triesch و همکاران [63-66]، تکنیک تطبیق گراف الاستیک (EGM) را برای توسعه یک سیستم در تشخیص وضعیت دست مستقل فرد در برابر پس زمینه های پیچیده مورد استفاده قرار دادند. وضعیت های دست توسط گراف های برچسب دار با یک توپولوژی دو بعدی پایه

نشان داده می شوند. متصل به گرہها، جت هایی قرار دارند، که یک توصیف تصویر موضعی (قابلیت تصویر) بر اساس فیلترهای گابور است. این رویکرد، تشخیص مستقل از کاربرد و نامتغیر در مقیاس را بدون تقسیم بندی صریح ناحیه دست ارائه داد. وضعیت های مختلف دست به صورت گراف های نسبت داده شده نشان داده می شوند و مقایسه هایی بین گراف های مدل (در پایگاه داده) و گراف داده (مربوط به تصویر واقعی) ساخته می شوند. گرہ ها با استفاده از یک تابع شباهت مقایسه می شوند و این الگو توسط محاسبه شباهت های گرہ متوسط تشخیص داده می شود.

گراف های دسته ای [76] برای مدلسازی تغییرپذیری در ظاهر شی استفاده می شود. تغییرپذیری طبیعی در ویژگی های نقاط متناظر در چندین تصویر (برای همان شیء یا یک کلاس از اشیاء) توسط برچسب گذاری هر گرہ با یک دسته از مقادیر ویژگی، که از نقاط مربوطه استخراج شده است گرفته می شود. این روش توسط Triesch و همکاران [63، 64] برای مدلسازی پس زمینه پیچیده در تصاویر حالت دست استفاده شده است. برای فرایند تطبیق، هر یک از مقدار ویژگی در دسته با اطلاعات موضعی تصویر در گراف داده مقایسه می شود و حداکثر شباهت ها به عنوان شباهت گراف دسته به دست می آید.

Li و Wachs [67]، یک الگوریتم سلسله مراتبی EGM برای تشخیص حرکت ژيروسکوپ را پیشنهاد دادند. بهبود عمده الگوریتم EGM، استفاده از سطوح سلسله مراتب های اختصاص داده شده به گرہ ها است. ویژگی های بصری با احتمال بیشتر (که روی تصویر هدف قرار می گیرند)، یک سطح سلسله مراتب بالاتر را در مقایسه با ویژگی های کمتر سازگار با مدل گراف دریافت می کنند.

2.2.4 توپولوژی / روش های مبتنی بر مدل سه بعدی

برازش مدل سه بعدی در برآورد حالت دست در [69] استفاده می شود. این روش، تمام زوایای اتصال بازسازی کننده دست را به عنوان یک مدل واکسل برآورد می کند. سپس برازش مدل بین مدل دست و مدل واکسل، در فضای سه بعدی انجام می شود. این روش تنها از اطلاعات هندسی مدل دست و مدل واکسل برای برازش مدل استفاده می کند و نیازی به هیچ اطلاعات اکتشافی یا پیشین ندارد. با این حال الگوریتم به پیاده سازی سریعتر برای برنامه ها در زمان واقعی نیاز دارد.

Xie و Yin [70] یک مدل بینش کامپیوتری از دست را به جای یک مدل سینماتیک معرفی کردند. این الگوریتم از پیچیدگی در برآورد پارامترهای زاویه ای و خطی مدل سینماتیک اجتناب می کند. آنها از ویژگی های توپولوژیک دست برای تشخیص وضعیت سه بعدی دست بهره گرفتند. نقطه لبه انگشتان به عنوان نقاط مورد نظر استخراج می شود. این دست از پس زمینه های پیچیده ای با استفاده از یک شبکه عصبی انرژی کولمب محدود (RCE) بر اساس تقسیم بندی رنگ تقسیم بندی می شود.

یک خلاصه و مقایسه ویژگی های الگوریتم های تشخیص دست در این بخش در جداول 3 و 4 ارائه شده است.

| Work | Features | Classification method | Application |
|------|------------------------------------------------------------|--------------------------------------|---------------------------------------------------|
| [61] | Geometric distance | DLLE/PNN | Manipulation of objects in windows user interface |
| [52] | Intrinsic geometry of hand | SLL | Chinese sign language |
| [51] | Multivalued features (centroid, compactness, area of hand) | RIL | Gesture commands |
| [50] | Discrete Fourier transform based distance metric | Nearest neighbour/maximum likelihood | Gesture commands |
| [44] | Shape, texture and color features | SVM | Recognition against complex backgrounds |
| [63] | Gabor jets | EGM | HRI |
| [69] | Joint angles | 3D model fitting | Not specified |
| [45] | Shape and texture features | Fuzzy-rough classifier | HRI |
| [62] | Shape features | EGM | Not specified |
| [58] | Gabor features | SVM | Recognition under varying illumination |
| [67] | Histogram of Oriented Gradient | EGM | Not specified |

جدول 3: روش های تشخیص وضعیت دست: ویژگی های استفاده شده، روش های طبقه بندی و برنامه های گزارش

شده

| Work | Accuracy | Class | Subject | Sample | UI | Spot | BG | Noise | Scale | Light | Extensibility | CV | Data |
|------|--------------|-------|---------|--------|----|------|----|-------|-------|-------|---------------|----|------|
| [61] | 93.2 | 14 | 1 | 20 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| [52] | 90.6 | 30 | 1 | 55 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| [51] | 94.4 | 20 | 1 | 45.4 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| [50] | 98.5 | 9 | 4 | 44.44 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| [44] | 94.36 | 10 | 40 | 5 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 |
| [63] | 85.8 | 12 | 19 | 1.48 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 |
| [69] | Not reported | 4 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| [45] | 98.75 | 10 | 19 | 18 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| [62] | 96.35 | 10 | 19 | 2.52 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| [58] | 96.1 | 11 | 10 | 6 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 |
| [67] | 99.85 | 10 | 24 | 2.7 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 |

جدول 4: روشهای تشخیص وضعیت دست: ویژگیهای الگوریتم و روش آزمایشی که در تست الگوریتم اتخاذ شده

است (فهرست در انتهای توصیف عناوین ستون) ارائه شده است. ویژگی های ستون 6 به ترتیب باینری هستند، 1

نشان دهنده سازگاری کار با این ویژگی است، در حالی که 0 نشان دهنده عدم انطباق است.

3. روندهای اخیر در تشخیص وضعیت دست: روش های مبتنی بر سنسور RGB-D

دوربین های عمق در بینش رایانه ای برای چندین سال استفاده شده اند. هرچند، قابلیت کاربرد دوربین های عمق به دلیل قیمت بالای آن و کیفیت پایین محدود شد. انتشار دوربین عمق-رنگ هزینه پایین Kinect (RGB-D) [77,78] توسط مایکروسافت، یک انقلاب را در تشخیص ژست با فراهم نمودن تصاویر عمق کیفیت-بالا با پرداختن به موضوعاتی مانند پس زمینه های پیچیده و تغییر روشنایی خلق شده است. این وسیله، یک نقشه سه بعدی از صحنه را با استفاده از یک ترکیب دوربین RGB و IR محاسبه می کند. اخیراً [79]. Hanetal یک بازنگری از میزان مفید بودن Kinect در پرداختن به مسائل بنیادی در بینش کامپیوتری فراهم نمود. سنسورهایی مانند *MicrosoftKinect(R)* و *ASUSXtionPROLIVE(R)* ردیابی قابل اطمینان از وضعیت های بدن انسان را سناریوهای بازی سازی فراهم می کنند. بر اساس ردیابی، این وسایل، ویژگی هایی مانند مختصات یک مدل اسکلتی را فراهم می کنند که برای تشخیص ژست استفاده می شوند.

داده های اسکلتی از این سنسورهای RGB-D به ویژگی های معنی دار تر و سطح بالایی تبدیل می شوند و الگوریتم ها باید برای طبقه بندی قوی حرکات طراحی شوند. تشخیص حرکات دست به ویژه به دلیل چیدمان پیچیده و نسبتاً کوچکتر منطقه دست چالش برانگیز است. علاوه بر این، یک الگوریتم تشخیص ژست قوی باید با توجه به اندازه و سرعت ژست، و جهت گیری ژستور، تغییری ایجاد کند. رافائل و همکاران [80] تأثیر اطلاعات عمق در فرایند تشخیص ژست را ارزیابی کردند و نتیجه گرفت که استفاده از شبیح عمق، دقت تشخیص را به طور قابل توجهی افزایش می دهد. Dominio و همکاران [81] یک الگوریتم را برای ترکیب چندین توصیف کننده عمق برای تشخیص حرکت ژيروسکوپ پیشنهاد کرد. دوربین های RGB-D اغلب برای تشخیص کامل ژست بدن [78، 82-85] استفاده می شوند، زیرا این دوربین ها ردیابی اسکلت را ارائه می دهند. این بخش از مقاله، الگوریتم های HGR مبتنی بر دوربین RGB-D را طبقه بندی کرده و نوشته های مرتبط را به دو دسته تقسیم می کند: الف) رویکردهای مبتنی بر Kinect و ب) سایر رویکردهای مبتنی بر سنسور RGB-D.

3.1 روشهای مبتنی بر Kinect

Zhang و همکاران [98] یک توصیفگر سطح جدید به نام هیستوگرام 3D Facets (H3DF) پیشنهاد کردند تا صریحاً اطلاعات سه بعدی شکل را از اطلاعات عمق سطح پایین تر رمزگذاری کنند. ویژگی های مبتنی بر Kinect برای تشخیص ژست دست پویا [89-93، 95-97، 99-111] و تشخیص حالت دست [88، 112-120] استفاده می شوند.

3.1.1 تشخیص حرکات دست پویا

Wu و همکاران [90] یک سیستم برای یادگیری حرکات فقط از یک مثال یادگیری برای هر کلاس پیشنهاد کردند، یعنی One-shot-learning. ویژگی ها بر اساس Extended-Motion-History-Image (Extended-MHI) (تصویر-تاریخ-حرکت-گسترش یافته) استخراج می شوند و حرکات با محاسبه حداکثر ضریب همبستگی طبقه بندی می شوند. تصاویر تاریخچه حرکت (MHI) [121] برای نشان دادن حرکات یک شی در یک ویدیو استفاده می شود. تمام فریم ها در یک توالی ویدیویی بر روی یک تصویر در محور زمانی تصویر می شوند تا اطلاعات زمانی دنباله حرکت را ضبط نمایند. به واسطه جبران کردن نواحی غیرمتحرک و اقدامات تکراری، Extended-MHI برای بهبود عملکرد MHI پیشنهاد شده است. الگوریتم تعبیه طیفی چند-دید (MSE) برای ترکیب داده های RGB و عمق به شیوه ای فیزیکی و معنی دار استفاده می شوند. الگوریتم MSE، رابطه درونی بین ویژگی های RGB و عمق را پیدا می کند و در سرعت تشخیص الگوریتم بهبود می بخشد.

[92] Lui، [99] یک الگوریتم تشخیص حرکت را بر اساس یک چارچوب رگرسیون غیر خطی روی منیفولدها پیشنهاد کرد. هندسه پایه و تناسب حداقل مربعات برای توسعه این الگوریتم استفاده می شود. رگرسیون حداقل مربعات به عنوان یک تابع ترکیبی با توجه به خصوصیات هندسی فرموله می شود. Gallo و همکاران [89] یک سیستم تشخیص ژست مبتنی بر Kinect را با کاربرد آن برای کاوش داده های تصویر پزشکی پیشنهاد کردند. حرکات مختلف برای توابعی مانند زوم، انیمیشن، ناحیه استخراج مورد نظر، چرخش و انتقال تصاویر پزشکی، توسط تجزیه و تحلیل توپولوژیکی منطقه دست تشخیص داده می شوند. ماتریس فاصله اقلیدسی و کوواریانس های یک

متریک لگاریتم-اقلیدسی به عنوان ویژگی ها [93] استفاده می شود. حرکات با استفاده از نزدیک ترین طبقه بند همسایه طبقه بندی می شوند.

یک روش جدید برای یادگیری یک لحظه ای برای تشخیص ژست از تصاویر عمق حرکت بر اساس تطبیق الگو در [100] ارائه شده است. این روش براساس محاسبه توصیفگرهای فضا-زمان از ویدیوی پرس و جو است که شباهت یک ژست در یک واژگان را اندازه گیری می کند. طبقه بند، بر اساس ضریب همبستگی انحراف استاندارد تبدیل فوریه تصویر و MHI است.

یک الگوریتم برای آشکارسازی و تشخیص حرکات دست به واسطه ترکیب DTW با تخمین های احتمالات در [102] پیشنهاد شده است. این الگوریتم دارای استحکام در برابر موقعیت و جهت گیری ژستور و سرعت ژست است. Cheng و همکاران [103، 104] الگوریتم های مبتنی بر DTW را برای تشخیص ژست سه بعدی پیشنهاد کردند. یک پنجره جستجوی پارامتریک در ماتریس هزینه روش DTW سنتی برای آشکارسازی شروع و پایان حرکات خاص از توالی های ژست مسیر نامتناهی معرفی می شود.

الگوریتم دیگر برای تشخیص ژست یادگیری یک لحظه ای از داده های RGB-D توسط Wan و همکاران ارائه شده است. [101] یک مایش ویژگی فضایی-زمانی جدید به نام تبدیل ویژگی نامتغیر در مقیاس حرکت ارتقایافته سه بعدی (EMoSIFT) استفاده می شود. مجموعه ویژگی های جدید در مقیاس و چرخش، نامتغیر است، زیرا داده های RGB-D را ترکیب می کند. یک روش کد گذاری نادر، یعنی تعقیب تطبیق متعامد شبیه سازی (SOMP) برای نشان دادن هر ویژگی بوسیله ترکیب خطی تعداد کمی از کد-کلمه ها مورد استفاده قرار می گیرد.

3.1.2 تشخیص وضعیت های دست

یک روش جدید ضبط حرکت دست بر اساس طرح پارتیشن دست 14-قسمتی در [117] برای جمع آوری مجموعه داده های حقیقی وضعیت در شرایط بدون محدودیت پیشنهاد شده است. Liang و همکاران [122]، یک طرح قوی تجزیه دست را برای استخراج یک توصیف سطح بالا از دست از تصاویر عمق پیشنهاد کردند. این روش برای تنظیمات دستی پیچیده، قوی است.

Ren و همکاران [88، 112] یک سیستم تشخیص وضعیت دست را پیشنهاد می کنند که دارای مقاومت در مقابل تغییرات در جهت گیری، مقیاس و بیان است. یک معیار فاصله به نام فاصله (FEMD) Mover Finger-Earth برای اندازه گیری عدم تشابه دست پیشنهاد شده است. این الگوریتم می تواند وضعیت های دست را به رغم تغییرات تشخیص دهد، زیرا فقط انگشتان (نه تمام شکل دست) را مطابقت می دهد. یک مقایسه از FEMD با الگوریتم زمینه شکل [123] در [113] ارائه شده است. الگوریتم مبتنی بر FEMD دارای دقت و سرعت محاسباتی بهتر است. علاوه بر این، Zhou و همکاران [113]، یک الگوریتم شناخت وضعیت دست مبتنی بر FEMD را برای بازی سودو کو ارائه دادند.

یک الگوریتم برای طبقه بندی شکل دست استاتیک و پویا با استفاده از جنگل های تصمیم گیری تصادفی در 91 پیشنهاد شده است. شکل دست با استفاده از داده ها از سنسورهای عمق طبقه بندی می شود. این سیستم مستقل از شرایط روشنایی عمل می کند و نیازی به مرحله ثبت دست ندارد. برچسب های کلاس به هر پیکسل در یک تصویر عمق اختصاص داده می شوند و برچسب کلاس نهایی توسط رای گیری تعیین می شود.

Kirac و همکاران [116] یک طرح برای استخراج اسکلت دست با استفاده از جنگل های رگرسیون تصادفی در زمان واقعی پیشنهاد کرد. این الگوریتم در خود-انسداد و وضوح کم دوربین عمق مقاوم است و می تواند موقعیت های مشترک را حتی اگر تمام پیکسل های مربوط به یک اتصال خارج از قاب دوربین باشند، برآورد نماید. مجموعه ای از ویژگی های توزیع شعاعی محور در [118] پیشنهاد شده است که می تواند به طور همزمان نوک انگشتان دست را موقعیت یابا نماید و وضعیت های دست را به طور کلی کدگذاری کند.

3.2 سایر روشهای مبتنی بر سنسور RGB-D

3.2.1 تشخیص حرکات پویای دست

Holte و همکاران [86] یک دوربین عمق-شدت (CSEM Swissranger SR-2) را برای توسعه یک الگوریتم تشخیص ژست نامتغیر دید استفاده کردند. بر خلاف روش مبتنی بر مسیر معمول، حرکات بر پایه حرکت های اولیه در داده های سه بعدی شناخته می شوند. موارد ابتدایی به صورت نامتغیر در دید با استفاده از زمینه شکل هارمونیک

نمایش داده می شوند. یک طبقه بند فاصله ویرایش احتمالی برای طبقه بندی استفاده می شود. الگوریتم دارای عدم تغییر در جهت گیری است، از یک نقطه نظر، بر روی داده ها آموزش داده می شود و بر روی داده ها از یک دیدگاه متفاوت آزمایش می شود.

الگوهای دوبعدی احتمالی ایجاد شده با استفاده از مسیر حرکت دستی در [94] برای شناخت حرکات پویا مورد استفاده قرار می گیرند. الگوی احتمالاتی، اعوجاجات مختلف مسیر با احتمالات مختلف را در نظر می گیرند. طولانی ترین طبقه بند متعاقب متعارف (LCS) برای محتمل ترین طبقه بند متعاقب عام طولانی (MPLCS) برای اندازه گیری شباهت بین الگوی احتمالاتی و نمونه ژست دست اصلاح می شود. [Erden et al 124] یک سیستم کنترل از راه دور مبتنی بر ژست دست را طراحی کرده است که ترکیبی از سنسورهای مادون قرمز با یک دوربین RGB است.

3.2.2 تشخیص وضعیت های دست

دوربین های زمان پرواز (ToF) و RGB در [87] ترکیب شده اند تا یک الگوریتم تشخیص دست بر اساس عمق و رنگ ایجاد کنند. موقعیت دست در سه بعد با وجود همپوشانی آن با قسمتهای بدن و دیگر دستها در پس زمینه، ردیابی می شود. ژست ها با استفاده از یک جستجوی نزدیکترین همسایه پس از کاهش ابعاد با استفاده از حداکثر سازی حاشیه همسایگی متوسط (ANMM) [125] شناسایی می شوند.

خلاصه و مقایسه ویژگی های ژست دست و الگوریتم های تشخیص وضعیت پیمایش شده در این بخش در جداول 5 و 6 ارائه شده است.

4. پایگاه های داده ژست دست

محققان دانشگاه کمبریج و تحقیقات میکروسافت، یک تحقیق [126] در مورد نحوه آموزش افراد برای توسعه بهترین مجموعه داده های ژست برای آموزش الگوریتم های یادگیری ماشین انجام داده اند. آنها از دو معیار، صحت و پوشش استفاده کردند تا بتوانند میزان خوب بودن مجموعه داده ها در تفسیر داده های جهان واقعی را از یک سیستم مستقر ارزیابی کنند. صحت اندازه گیری به شباهت حرکات فرد تا آنچه که برنامه نویس سیستم به انجام آنها

نیاز دارد اشاره می کند. این به درک فرد بستگی دارد. معیار پوشش به تکمیل بودن مجموعه داده ها در ارائه تغییرات طبیعی و ممکن الگوهای حرکت مرتبط اشاره می کند. پوشش توسط آزادی داده شده به فرد تعیین می شود. آنها، مناسب ترین روش نشانه شناسانه دستورالعمل ها و نظم آنها برای دستیابی به بهترین صحت و پوشش، برای مجموعه داده ها و سیستم تشخیص ژست یاد شده مورد بررسی قرار دادند. روشهای مورد بررسی شامل متن توصیفی، توالی تصویر ایستا و ویدیو بودند. ویدئو و پس از آن متن به عنوان بهترین مرتبه نشانه برای تسهیل درک و آزادی افراد انتخاب می شوند.

پایگاه داده های استاندارد ژست دست برای آزمایش قابل اطمینان و مقایسه الگوریتم های تشخیص ژست دست لازم هستند. دسترسی به پایگاه داده های ژست دست تا سال 2007 محدود بود و اخیراً افزایش یافته است (شکل 6). این بخش، یک بازنگری از مجموعه داده های دست ژست در دسترس عموم را ارائه می دهد. جدول 7، لیست پایگاه های داده وضعیت و ژست دست با لینک های وب برای دانلود آنها را فهرست نموده است. جدول 8 این مجموعه داده ها را با جزئیاتی مانند تعداد کلاس ها، افراد و نمونه های موجود در این مجموعه توضیح می دهد. آثار مورد استفاده از مجموعه داده ها نیز برای تسهیل مطالعه تطبیقی امکان پذیر گنجانده شده اند. در زمان انتشار این بررسی، در مجموع 25 مجموعه داده وجود دارد.

4.1 مجموعه داده های *Sebastien-Marcel* در مورد وضعیت و ژست دست

این مجموعه داده حاوی سه مجموعه داده وضعیت دست، پایگاه داده وضعیت استاتیک [64] Jochen Triesch، پایگاه داده موقعیت استاتیک دست [63] Jochen Triesch و پایگاه اطلاعات وضعیت استاتیک دست *Sebastien-Marcel* [133] و یک پایگاه داده دست ژست پویا، پایگاه داده دینامیک دست *Sebastien-Marcel* [15] می شود. مجموعه داده های وضعیت دست دارای پس زمینه های ساده و پیچیده است. ژست های پویا شامل سیگنال فرماندهی متغیر برای کلیک، توقف-بگیر-اوکی، چرخش، و شماره می شود.

4.2 مجموعه داده های ژست دست کمبریج

این مجموعه داده ها حاوی تصاویر وضعیت دست می شود. دارای دنباله هایی از تصاویر استاتیک مربوط به حرکات دست می شود که آن را برای آزمایش الگوریتم های تشخیص ژست پویا مناسب می سازد [137]. مجموعه داده ها شامل ژست های تعریف شده توسط 3 اشکال دست اولی (تخت، گسترش و V شکل) و 3 حرکت ابتدایی (سمت چپ، راست و قرارداد) می شود. وظیفه هدف برای این مجموعه داده ها، طبقه بندی اشکال و حرکات دست در یک زمان است. مجموعه داده ها دارای تغییرات درون گروهی نسبتاً زیادی در همترازی مکانی و زمانی حرکات دست است.

| Work | S/D | Sensor | Features | Classification method |
|------|-----|-----------------------|--------------------------------------------|----------------------------------------------|
| [86] | D | CSEM Swissranger SR-2 | Motion primitives | Probabilistic edit distance classifier |
| [87] | S | ToF and RGB camera | Haarlets | NeN |
| [88] | S | Kinect | Hand/finger shape | Template matching using FEMD |
| [89] | D | Kinect | Hand area | Classifier based on topology |
| [90] | D | Kinect | Extended-Motion-History-Image | Maximum correlation coefficient |
| [91] | S | Kinect | Depth pixel values | Randomized classification forests and voting |
| [92] | D | Kinect | Underlying geometry | Least squares fitting |
| [93] | D | Kinect | Euclidean and log-Euclidean distance | NeN |
| [94] | D | PrimeSense 3-D camera | Probabilistic 2D templates from trajectory | MPLCS classifier |
| [95] | D | Kinect | Spatial and motion features | Conditional density propagation |
| [96] | D | Kinect | Position, angle, and direction features | Probability; pairwise coupling |
| [97] | D | Kinect | Conditional distance | Dynamic time warping |

S—static, D—dynamic, ToF—time of flight, NeN—nearest neighbor, FEMD—finger earth mover's distance, and MPLCS—most probable longest common subsequence.

جدول 5 روش های مبتنی بر سنسور RGB-D؛ یک مقایسه

| Work | Accuracy | Class | Subject | Sample | UI | Spot | BG | Noise | Scale | Light | Extensibility | CV | Data |
|------|---------------|----------|---------------------|---------------------|----|------|----|-------|-------|-------|---------------|----|------|
| [86] | 92.9 | 4 | 10 | 3 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| [87] | 99.54 | 6 | 1 | 29.17 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| [88] | 93.9 | 10 | 10 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 |
| [89] | Not reported | 9 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| [90] | Not reported | 8–15 | Multiple (ChaLearn) | Multiple (ChaLearn) | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| [91] | 84.3 and 74.3 | 24 and 9 | 4 and 5 | 100 and 10 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 |
| [92] | 91.7 | 9 | 2 | 80 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 |
| [93] | 99.75 | 8 | 20 | 5 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| [94] | 98.7 | 10 | 8 | 5 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| [95] | 95.9 | 4 | 4 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| [96] | 97.26 | 10 | 6 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| [97] | 82 | 179 | 18 | Multiple (ChaLearn) | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 |

جدول 6 روش های مبتنی بر سنسور RGB-D؛ ویژگی های الگوریتم ها و روش شناسی آزمایشی اتخاذ شده در آزمایش الگوریتم (فهرست در پایین، توصیفی از عناوین ستون فراهم می کند). ویژگی ها در ستون 6 به سمت بیرون، دودویی هستند. 1 نشاندهنده انطباق کار با ویژگی است که در آن 0 نشان دهنده عدم انطباق می باشد.

4.3. مجموعه داده های ژست دست توسط Shen و همکاران

این پایگاه داده برای تست هر دوی الگوریتم های ژست دست و تشخیص حالت مفید است، زیرا حاوی هر دو الگوهای حرکت و اشکال دست خاص است [35]. دارای 10 کلاس از ژست های دست پویا (به عنوان مثال، حرکت به سمت راست، حرکت به سمت چپ، چرخش) می شود که با 7 وضعیت مختلف دست (به عنوان مثال انگشت شست، مشت، تمام انگشتان گسترش یافته) در مجموع 70 وضعیت دست متف در هر فرد انجام می شود.

4.4 پایگاه داده سیگنالهای حمل و نقل هوایی NATOPS

این پایگاه داده شامل 24 ژست های بدن و دست می شود که از سیگنال های حمل و نقل هوایی NATOPS (آموزش نیروی دریایی و استاندارد سازی عملیات) انتخاب شده است [132]. برای جمع آوری پایگاه داده، از یک دوربین استریو استفاده شد. این پایگاه شامل فیلم هایی با داده RGB و عمق می شود. همچنین شامل مجموعه هایی از ویژگی های بدن و دست استخراج شده در قالب های Matlab و CSV می شود.

4.5 مجموعه داده های ژست توسط یون و همکاران

این مجموعه داده شامل 48 ژست الفبایی کلاس (کاراکترهای الفبایی و عناصر گرافیکی) ضبط شده از 20 نفر، 10 بار در هر ژست می شود [11]. مجموعه داده حاوی توالی هایی از مختصات $x-y$ است که نشانگر حرکات ناخواسته است.

4.6 پایگاه داده بازی تعاملی Sebastien-Marcel

مجموعه داده شامل مسیرهای سه بعدی از ژست های دست تقسیم شده، از جمله مختصات سر و نیمتنه می شود [131، 13]. هر مسیر به عنوان یک فایل متنی در مجموعه داده ذخیره می شود. مجموعه داده ها دارای هر دو ژست های تک دست (مانند توقف، اشاره به سمت چپ، اشاره به سمت راست) و دو دست (مانند شنا، پرواز، کف زدن) می شود. مسیرهای ژست حاوی مختصات سه بعدی مرکز سر، دو دست و نیمتنه می شود.

4.7 مجموعه داده ژست Keck

مجموعه داده ژست شامل 14 ژست پویا می شود که زیر مجموعه هایی از سیگنال های نظامی (مانند چرخش به چپ، عقب رفتن و سرعت) می شوند [136]. مجموعه داده ها به دو مجموعه آموزشی و آزمایشی تقسیم می شوند.

مجموعه آموزشی با استفاده از یک دوربین ثابت با شخص مورد نظر در برابر یک پس زمینه ساده و استاتیک ضبط می شود. مجموعه آزمایش از یک دوربین متحرک، در حضور در هم ریختگی پس زمینه و دیگر اشیاء متحرک ضبط می شود.

4.8 پایگاه داده ژست حرکتی شش بعدی

پایگاه داده ژست حرکت 6 بعدی (6DMG)، داده های جامعی از ژست های حرکت، از جمله موقعیت، جهت، شتاب و سرعت زاویه ای را فراهم می کند. [130]. داده ها در فرم دودویی خام ذخیره می شوند و مجموعه داده ها با برنامه های ++C نمونه برای دسترسی و تجسم داده ها همراه می شوند.

4.9 داده های ژست ChaLearn

این مجموعه داده به عنوان بخشی از چالش شناخت ژست ایجاد شده است. چالش ژست [82] ChaLearn، 90-92، 100، 127]. داده های ژست ChaLearn 2011 شامل مجموع 62000 نمونه می باشد. مجموعه داده ها از 20 نفر تشکیل شده است که به 100 دسته تقسیم می شوند. داده ها با دوربین Kinect ضبط می شوند و شامل فیلم های RGB و عمق ژست های پویا می باشند. این مجموعه داده همچنین 8000 نمونه از داده های ترجمه شده، مقیاس پذیر و محصور شده را شامل می شود.

در مقایسه با مجموعه داده های دیگر، ژست ها در داده های ژست ChaLearn در دامنه های کاربرد گسترده مفید هستند. این شامل 9 دسته از ژست های مربوط به دامنه های کاربردی مختلف می شود. رده بندی ها عبارتند از: الف) نشان ها (به عنوان مثال Mudras هندی)، ب) تصویرگرها (به عنوان مثال ژست ایتالیایی)، ج) تنظیم کننده ها (اشاره ها با سر و دست به همراه گفتار)، د) پانتومیم ها (ژست های صورت گرفته برای تقلید اقدامات)، e) نشانه ها (از زبان های نشانه برای کورها)، f) سیگنال ها (مانند سیگنال های مارشالینگ برای هدایت ماشین یا وسیله نقلیه)؛ ج) ژست های زبان بدن (مانند سر زدن، عبور دست ها)؛ ه) اقدامات (مثلا نوشیدن یا نوشتن)؛ هر مجموعه داده شامل تعدادی از اقدامات ارائه شده به طور جداگانه یک بار برای هدف آموزش می شود. ترکیبی از یک یا چند عمل در یک توالی ویدیویی برای آزمایش وجود دارد.

| No. | Name, year | Source |
|-----|-----------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | ChaLearn gesture ^a data ^b , 2011 | http://gesture.chalearn.org/data |
| 2 | MSRC-12 Kinect gesture ^a dataset ^b , 2012 | http://research.microsoft.com/en-us/um/cambridge/projects/msrc12/ |
| 3 | ChaLearn multi-modal gesture data ^b , 2013 | http://sunai.uoc.edu/chalearn/ |
| 4 | NUS hand posture dataset-II, 2012 | http://www.ece.nus.edu.sg/stfpage/elepv/NUS-HandSet/ |
| 5 | 6D motion gesture database ^a , 2011 | http://www.ece.gatech.edu/6DMG/6DMG.html |
| 6 | Sebastien Marcel interact play database, 2004 | http://www.idiap.ch/resource/interactplay/ |
| 7 | NATOPS aircraft handling signals database ^a , 2011 | http://groups.csail.mit.edu/mug/natops/ |
| 8 | Sebastien Marcel hand posture and gesture datasets, 2001 | http://www.idiap.ch/resource/gestures/ |
| 9 | Gesture dataset by Shen et al., 2012 | http://users.eecs.northwestern.edu/~xsh835/GestureDataset.zip |
| 10 | Gesture dataset by Yoon et al., 2001 | Available on e-mail request to yoonhs@etri.re.kr |
| 11 | ChAirGest multi-modal dataset ^b , 2013 | https://project.eia-fr.ch/chairgest/Pages/Download.aspx |
| 12 | Sheffield Kinect Gesture Dataset ^b , 2013 | http://lshao.staff.shef.ac.uk/data/SheffieldKinectGesture.htm |
| 13 | Keck gesture dataset, 2009 | http://www.umiacs.umd.edu/~zhuolin/Keckgesturedataset.html |
| 14 | NUS hand posture dataset-I, 2010 | http://www.ece.nus.edu.sg/stfpage/elepv/NUS-HandSet/ |
| 15 | Cambridge hand gesture data set, 2007 | http://www.iis.ee.ic.ac.uk/~tkkim/ges_db.htm |
| 16 | Posture dataset by Ren et al. ^a , 2011 | http://eeeweaba.ntu.edu.sg/computervision/people/home/renzhou/HandGesture.htm |
| 17 | ColorTip dataset ^b , 2013 | https://imatge.upc.edu/web/res/colortip |
| 18 | NYU Hand Pose Dataset ^b , 2014 | http://cims.nyu.edu/~tompson/NYU_Hand_Pose_Dataset.htm |
| 19 | General-HANDS data-set ^b , 2014 | http://wildhog.ics.uci.edu:9090 |
| 20 | VPU Hand Gesture dataset (HGds), 2008 | http://www-vpu.eps.uam.es/DS/HGds/ |
| 21 | Dataset by Kawulok et al., 2014 | http://sun.aei.polsl.pl/~mkawulok/gestures/ |
| 22 | ASL Finger Spelling Dataset ^b , 2011 | http://personal.lee.surrey.ac.uk/Personal/N.Pugeault/index.php?section=FingerSpellingDataset |

جدول 7: پایگاه های داده ژست در دسترس عمومی و منابع آنها. برای توصیف، جدول 8 را ببینید.

| No. | Description | S/D | Works |
|-----|------------------------------------------------------------------------------------------------------------------------------------|---------|--------------------|
| 1 | ChaLearn Gesture Challenge, 62,000 samples | D | [82,90-92,100,127] |
| 2 | 12 classes, 30 subjects, 6244 samples | D | [126] |
| 3 | 20 classes, 27 subjects, 13,858 samples | D | [128] |
| 4 | 10 classes, 40 subjects, 2750 samples, complex background | S | [44,129] |
| 5 | 20 classes, 28 subjects, 5600 samples | D | [130] |
| 6 | 16 classes, 22 subjects, 50 samples/ subject | D | [13,131] |
| 7 | 24 classes, 20 subjects, 9600 samples | S and D | [132] |
| 8 | Three hand posture datasets, with 10 (gray scale), 12 (color), and 6 (gray scale) classes. One hand gesture dataset with 4 classes | S and D | [63-65,133] |
| 9 | 10 classes, 15 subjects, 1050 samples | S and D | [35] |
| 10 | 48 classes, 20 subjects, 9600 samples | D | [11] |
| 11 | 10 classes, 10 subjects, 1200 samples recorded with Kinect and inertial motion units | D | [134] |
| 12 | 10 classes, 6 subjects, 2160 samples recorded with Kinect and RGB cameras | D | [135] |
| 13 | 14 classes, 3 subjects, 126 training and 168 testing samples | D | [136] |
| 14 | 10 classes, 1 subject, 240 samples, color as well as gray scale | S | [45] |
| 15 | 9 classes, 2 subjects, 900 image sequences, with different illumination conditions | S and D | [137] |
| 16 | 10 classes, 10 subjects, 1000 samples, color as well as depth maps, cluttered background | S | [88] |
| 17 | 7 subjects, 9 classes, 7 training sequences of between 600 and 2000 depth frames | S | [118] |
| 18 | 2 users, data from 3 Kinects (frontal and 2 sides), 72,757 and 8252 frames in training and test sets | S | [138] |
| 19 | 22 sequences, different view-points, scales, poses, and occlusions | S | - |
| 20 | 12 classes, 11 subjects, 1 video per gesture (252 frames) | S | [139] |
| 21 | 32 classes, 18 subjects, gestures from Polish Sign Language and American Sign Language (ASL) | S | [140] |
| 22 | 24 classes, 9 subjects, 65,000 samples | S | [141] |

S--static, and D--dynamic.

جدول 8: شرح پایگاه های دست ژست در دسترس عموم (همانگونه که در جدول 7 آمده است)

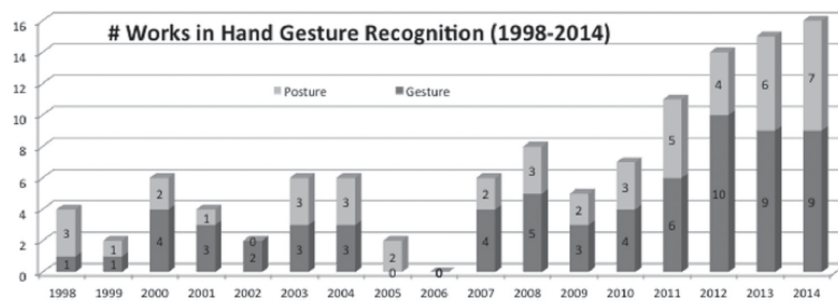
4.10 داده های ژست چندنمایی ChaLearn

در مقایسه با داده های ژست ChaLearn، تست با استفاده از داده های ژست چندنمایی [128] ChaLearn چالش برانگیزتر است. داده های ژست چندنمایی ChaLearn شامل ضبط توالی های پیوسته، حضور ژست های آشفته، تعداد نسبتاً زیادی از رده ها، توالی های ژست طولانی و ژست های مختلف توسط انواع کاربران می شود.

چندین روش در مجموعه داده ها شامل صدا، RGB، نقشه های عمق، ماسک های کاربر و مدل اسکلت کاربر ارائه می شوند.

4.11 مجموعه داده های چندنمایی ChAirGest

این داده ها با استفاده از یک دوربین Kinect و 4 واحد حرکتی اینرسی متصل به بازوی راست و گردن افراد به دست آمده است. ژست ها از 3 حالت مختلف استراحت آغاز می شوند و در دو حالت روشنایی ضبط می شوند [134].



شکل 5. نمودار نشاندهنده تلاش های تحقیقاتی در حال رشد در ژست دست و تشخیص وضعیت

4.12 مجموعه داده های ژست Kinect Sheffield (SKIG)

مجموعه داده SKIG [135] دارای 10 رده از ژست های دست می شوند که از 6 فرد با استفاده از دوربین های RGB و Kinect ضبط شده است. مجموعه داده ها با 3 پس زمینه مختلف (تخته چوبی، کاغذ سفید و کاغذ با کاراکترها) و 2 شرایط روشنایی (نور و تاریکی) ضبط می شود.

4.13 مجموعه داده های ژست MSRC-12 kinect

Microsoft Research Cambridge-12 (MSRC-12) یک مجموعه داده ژست پویا 12 کلاس ضبط شده با استفاده از داده های اسکلتی از Kinect است [126]. مجموعه داده ها شامل توالی هایی از حرکات انسانی می شود که با استفاده از مکان های بخش-بدن (20 مفاصل اسکلتی) نشان داده می شوند. مجموعه داده شامل 594 توالی و 719,359 فریم می شود.

4.14 مجموعه داده های I-وضعیت دست NUS

وضعیت ها در این مجموعه داده ها با موقعیت و اندازه مختلف دست در فریم تصویر ضبط می شوند. هر دو نسخه های رنگی و خاکستری از مجموعه داده ها در دسترس هستند. وضعیت های دست در مجموعه داده از نظر ظاهر دارای تنوع بین کلاسی کمتر می باشند، که کار شناسایی را به چالش می کشد [45].

4.15 مجموعه داده های II-وضعیت دست NUS

این مجموعه داده وضعیت دست پس زمینه پیچیده [44] سه زیر مجموعه دارد. A، B و C. زیرمجموعه A دارای تصاویر با پس زمینه های پیچیده طبیعی است و زیر مجموعه B دارای تصاویری با نویزهایی مانند بدن / چهره ژست گیرنده یا گروهی از افراد دیگر در پس زمینه است. زیرمجموعه C تنها تصاویر پس زمینه را شامل می شود (برای تصاویر منفی برای تشخیص موقعیت دست استفاده می شود). موقعیت ها دارای اشکال و اندازه های مختلف دست هستند و از افراد با قومیت های مختلف جمع آوری می شوند. زیر مجموعه A دارای 2000 تصویر است، B دارای 750 عکس است و C دارای 2000 تصویر است.

4.16 داده های ذخیره شده بوسیله Ren و همکاران

این یک مجموعه داده کینکت است که دارای 10 کلاس است. این شامل تصاویر رنگی و نقشه های عمق می شود. مجموعه داده ها تحت پس زمینه های در هم و برهم جمع آوری شده است.

4.17 مجموعه داده ColorTip

حاشیه نویسی ژست دست در این مجموعه داده [118] بین 9 کلاس ژست انجام می شود و مجموعه داده دارای تغییرات درونی کلاس است. حاشیه نویسی انگشت در مجموعه داده ها با استفاده از دستکش های رنگی انجام می شود، که کاهش تشخیص و محل زدن انگشتان دست می شود.

4.18 مجموعه داده های حالت دست NYU

مجموعه داده های ژست دست [138] [NYU] به ترتیب 72757 و 8252 عدد فریم در مجموعه های آموزشی و آزمون دارد. داده ها با استفاده از 3 سنسور Kinect، ضبط می شوند که یک دید جلو و 2 دید جانبی را فراهم می کند. مجموعه آموزش از کاربر 1 ضبط می شود و مجموعه آزمون از 2 کاربر ضبط می شود.

4.19 مجموعه داده های HANDS (دستان)-عمومی

این مجموعه داده شامل انواع 22 توالی می شود که دیدگاه ها، مقیاس ها، مواضع، انعکاس ها و فن آوری های مختلف دوربین را نشان می دهد. مجموعه داده ها برای ارزیابی الگوریتم های تشخیص دست و تخمینی ژست مفید است.

4.20 مجموعه داده های ژست دست VPU

این مجموعه داده ژست دست [139] شامل 12 طبقه داده از 11 فرد می شود. همچنین شامل داده های مصنوعی تولید شده می شود و در ارزیابی الگوریتم های تشخیص وضعیت دست مفید است.

4.21 مجموعه داده ها توسط Kawulok و همکاران.

این مجموعه داده [140] شامل ژست ها از زبان نشانه لهستانی و ASL می باشد و به صورت سه سری به دست آمده در شرایط مختلف به دست می آید. این کلاس 32 کلاس ژست از 18 فرد مختلف دارد.

4.22 مجموعه داده های تلفظ انگشت ASL

مجموعه داده های تلفظ انگشت ASL [141] شامل 24 ژست دست، حروف انگلیسی از a تا y می شود به جز j. این شامل یک مجموعه آسان، مجموعه داده های A، ضبط شده از 5 فرد بدون تغییر نور، و یک مجموعه سخت، مجموعه داده های B، ضبط شده از 9 فرد با تغییرات نور می شود.

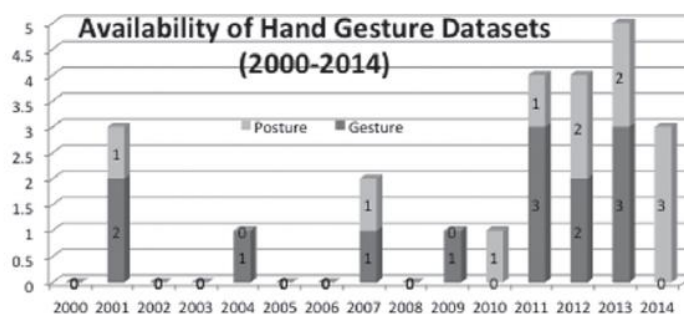
4.23 دیگر پایگاه های داده مربوط به دست

چندین پایگاه داده قابل دسترس در دسترس عمومی مربوط به تحقیقات تشخیص ژست دست عبارتند از: (a) MSRA Hand Tracking Database [142] ([http:// research. microsoft.com/en-us/um/people/yichenw/handtracking/index.html](http://research.microsoft.com/en-us/um/people/yichenw/handtracking/index.html)) (ب) پایگاه داده زبان لاتین دیجیتال آمریکایی

[143] (<http://www.bu.edu/asllrp/cslgr>)؛ و (c) پایگاه داده های دستی Bosphorus (<http://bosphorus.ee.boun.edu.tr/hand/Home.aspx>)

5. بحث

نمودار 5 نشان می دهد که رشد سریع در تحقیقات تشخیص ژست دست و در شکل 6، نشان دهنده رشد در انتشار پایگاه های ژست دست است. به رغم این پیشرفت ها، چالش های حل نشده در تشخیص ژست وجود دارد. این بخش به طور خلاصه برخی از مسائل حل نشده در این زمینه را مرور می کند، مقایسه ای از رویکردهای مختلف را ارائه می دهد و چندین مسیر تحقیق در آینده را مورد بحث قرار می دهد.



شکل 6. نمودار نشاندهنده رشد در مجموعه داده های ژست و وضعیت دست در دسترس عموم

5.1 شناسایی تصویرگرها

تشخیص تصویرگرها چالش برانگیز است، زیرا معنی این ژست ها بستگی به زمینه دارد. مرجع زمینه علاوه بر شناسایی یک ژست تصویرگر باید تشخیص داده شود. در میان تصویرگرهای مختلف، ژست اشاره در برنامه هایی مثل فرماندهی ربات بسیار مفید است. درک ژست اشاره در سه بعد شامل تشخیص ژست، پیدا کردن موقعیت دست و شناسایی جهت اشاره شده می شود. برآورد دقیق جهت اشاره، تشخیص ژست نشانه را به چالش می کشد. ابتکاری مانند جهت که به نظر می رسد فرد به آن نگاه می کند، در تشخیص یک حرکت اشاره مفید است. به عنوان مثال یک خط متصل کننده مرکز چشمان با نوک انگشت اشاره می تواند برآورد جهت اشاره شده را فراهم کند که به نوبه خود می تواند برای شناسایی نقطه هدف مورد استفاده قرار گیرد [144، 145].

برآورد جهت اشاره با استفاده از خط دست-سر زمانی که دست اشاره کننده به سمت بیرون گسترش یابد و روی سطح یک نیمکره خیالی محور به مرکز شانه قرار گیرد [146، 147]، موثر است. با این حال، این روش در مورد ژست های اشاره ای جمع و جور که در آن فرد تنها ساعد را حرکت می دهد، موثر نیست. چنین ژست های اشاره گر کوچک را می توان با مدل سازی ویژگی های سینماتیک ساعد و انگشت اشاره تشخیص داد [146]. جهت گیری سر را می توان به عنوان یک ویژگی برای بهبود عملکرد تشخیصگر ژست اشاره کننده استفاده کرد [148]. در مقایسه، برآورد جهت با استفاده از خط سر-دست بهتر عمل می کند که بر اساس جهت گیری ساعد، در مورد یک ژست اشاره طبیعی است. [148]. Raheja و همکاران [149]، یک الگوریتم برای آشکارسازی مکان اشاره ژست دست پیشنهاد کرد که براساس مکان های سر، شانه ها و آرنج ها است. روش پیشنهاد شده توسط Pateraki و همکاران. [150، 151]، ژست چهره و جهت گیری سر را با جهت دست ترکیب نمود.

5.2 مقایسه روش ها، ویژگی ها و روش های طبقه بندی

5.2.1 رویکردهای مبتنی بر ظاهر و مدل

رویکردهای مبتنی بر ظاهر، عملکرد بهتر در زمان واقعی را در مقایسه با رویکردهای مبتنی بر مدل سه بعدی دست ارائه می کنند، زیرا فرآیند استخراج ویژگی تصویر، سریعتر است. مدل های مبتنی بر ظاهر منجر به الگوریتم های کارآمدتر از نظر محاسباتی می شوند که در شرایط محدود به خوبی کار می کنند، اما فاقد تعمیم مطلوب برای تعامل کامپیوتری انسان هستند. روش های مبتنی بر ظاهر عمدتاً از داده های شکل دو بعدی دست استفاده می کنند که وابسته به زاویه دید است. استفاده از چنین روش هایی توسط چشم اندازهای دیدن محدود شده است. یک کلاس گسترده از ژست های دست را می توان با رویکرد مبتنی بر مدل دست سه بعدی تحت پوشش قرار داد، زیرا این مدل ها، راهی برای مدلسازی مشروح ژست دست ارائه می دهند. با این حال، مدل های سه بعدی به پایگاه داده تصویر بزرگ برای پوشش تمام اشکال مشخصه و تغییرات آن تحت دیدگاه های مختلف نیاز دارند. تطبیق تصویر آزمون با تمام مدل های موجود در پایگاه داده، وقت گیر و از نظر محاسباتی گران است که استفاده از مدل های سه بعدی را برای برنامه ها در زمان واقعی محدود می کند.

5.2.2 ویژگی ها

انتخاب پذیری و عدم تغییر، دو کیفیت مورد نظر برای هر فرآیند شناسایی الگوی مبتنی بر تصویر هستند. رویکردهای مبتنی بر الگو، انتخاب پذیری مطلوب را برای الگوهای شکل بدون انحراف فراهم می کنند. رویکردهای مبتنی بر هیستوگرام دارای ویژگی عدم تغییر هستند. با این وجود رویکردهای هیستوگرام، اطلاعات تصویر یکپارچه را در نظر می گیرند، که این کار را برای وظایف تشخیص شکل مانند تشخیص ژست بدنه نامناسب می سازد. الگوهای شکل-بافت استخراج شده با استفاده از رویکردهای الهام گرفته زیست شناسی [152]، ویژگی هایی را فراهم می کنند که دارای قابلیت انتخاب و عدم تغییر هستند و در تشخیص وضعیت دست مفید هستند [44].

ویژگی های جهت گیری و زاویه ای ژست ها در مقایسه با ویژگی های موقعیتی، تغییری را ارائه می دهند. از سوی دیگر، ویژگی های موقعیتی ساده هستند و می توانند با دقت بیشتری استخراج شوند. ویژگی های مبتنی بر بافت، توانایی ضبط خواص مکانی بهتر را در مقایسه با موارد ضبط شده توسط ویژگی هایی مانند رنگ دارند.

سنسورهای RGB-D، صرف نظر از پس زمینه های پیچیده و تغییرات در مقیاس، نورپردازی و نقاط مشاهده، امکان استخراج ویژگی های غیر مجاز را فراهم می کنند. داده های دقیق عمق و اطلاعات موقعیت از این حسگر ها، استخراج مدل های دست را سرعت می بخشند که موجب افزایش مطلوبیت رویکردهای مبتنی بر مدل می شوند.

5.2.3 روشهای طبقه بندی

روش های مبتنی بر HMM، موثر هستند و برای HGR به طور گسترده ای مورد استفاده قرار می گیرند. با این وجود رویکردهای مبتنی بر HMM نیاز به تعداد زیادی از نمونه های آموزشی دارند و دارای عیب روش آموزش دقیق هستند. هزینه های محاسباتی الگوریتم های مبتنی بر HMM با واژگان ژست افزایش می یابد. علاوه بر این، عملکرد الگوریتم های مبتنی بر HMM، زمانی کاهش می یابد که تنوعات بین شرایط آموزش و آزمایش می شود. پیدا کردن مجموعه های مطلوب پارامتر و پراکندگی مسیر برای تقسیم بندی زمانی، دیگر تنگناها در استفاده از HMM هستند.

طراحی یک TDNN جذاب است، زیرا ساختار جمع و جور، وزن ها را کاهش می دهد و امکان شناسایی ویژگی های عمومی را فراهم می آورد. سلسله مراتب تاخیرات در TDNN، این آشکارسازهای ویژگی را با افزایش دامنه آنها در هر لایه بهینه سازی می کند. یکپارچه سازی زمانی ویژگی ها در لایه خروجی، باعث عدم تغییر جابجایی شبکه می شود (غیر حساس به موقعیت دقیق دست). تعداد کل وزن ها در شبکه نسبتاً کوچک است از آنجا که فقط یک پنجره کوچک از الگوی ورودی به TDNN در هر لحظه تغذیه می شود. این به کاهش زمان آموزش کمک می کند.

الگوریتم های مبتنی بر گراف دارای عیب پیچیدگی محاسباتی بالا هستند که منجر به عدم پذیرش آن برای برنامه های در زمان واقعی می شود. با این حال هر گره در گراف می تواند با مجموعه ای از ویژگی های گره مدل سازی شود، که در حل مسائل با توجه به زمینه های پیچیده [63] و اندازه ها و تغییرات شکل مفید است.

5.3 چالش ها و دستورالعمل های تحقیقاتی آینده

شناسایی فاز ژست گرفتن، یک چالش عمده در HGR است. حضور حرکات دست غیر-ژست غیرقابل پیش بینی و مبهم، این کار را چالش برانگیز می کند. توانایی مردود کردن کلاس های ناشناخته یکی از الزامات مهم برای تشخیص دهنده خودکار ژست است. مفهوم مدل آستانه معرفی شده توسط Lee و Kim [10] برای این منظور مفید است. الگوریتم تقسیم بندی و تشخیص ژست پیشنهاد شده توسط Kim و همکاران. [17] از یک برآورد احتمالی پیوسته ژست و غیر ژست برای پیدا کردن نقاط شروع / پایان استفاده کرد. Kang [153] و همکاران یک طرح نقطه یابی ژست مبتنی تشخیص را برای فیلتر کردن حرکات ناخواسته پیشنهاد کردند. اخیراً Yin و همکاران [154] از یک HMM تلفیقی، برای تکمیل نقطه گذاری ژست در جریان داده های مداوم استفاده می کند که نتایج دلگرم کننده ای را حاصل می کند.

حرکات گذار بین ژست های مجاور یکی دیگر از مسائل مربوط به شناخت خودکار ژست های پیوسته، به ویژه در برنامه های کاربردی مانند تشخیص نشانه زبان است. Yang و همکاران [155] با استفاده از رویکرد مبتنی بر برنامه نویسی پویا، مساله مربوط به افزودن صدا در یک کلمه را با استفاده از رویکرد مبتنی بر برنامه نویسی مورد بررسی قرار دادند. Li و همکاران [156]، سه روش مبتنی بر یک مدل ژست را برای موقعیت یابی نقطه پایانی پیشنهاد

دادند. روش های مورد بررسی یک جستجوی چندگانه، انحراف زمان پویا و برنامه ریزی پویا هستند. در مقایسه، روش مبتنی بر برنامه نویسی پویا از دو مورد دیگر بهتر عمل نمود. یک رویکرد برنامه نویسی پویای مبتنی بر سطح-ساختمان تودرتو توسط Sarkar و همکاران، [157] برای پرداختن به عدم قطعیت های مرزهای علامت در جملات پیشنهاد شده است.

تطبیق یک توالی تصویر برای یک مدل، یک مسئله اصلی در HGR است. Yang و همکاران [158]، یک الگوریتم بهینه سازی را برای تطبیق گروه های ابتدایی تصویر با مدل های آماری (HMM) و غیر آماری (نمونه بردار) پیشنهاد نمودند. الگوریتم نیازی به تقسیم کامل صحنه و ردیابی ویژگی ها در فریم ها نداشت.

روند اخیر یادگیری یک لحظه ای [90، 97، 100، 101] در تشخیص ژست امیدوار کننده است. یادگیری یک لحظه ای شامل یادگیری یک ژست با مشاهده یک نمونه از آن ژست، مشابه یادگیری در انسان می شود. این روند، فرصتی را برای به دست آوردن چالش استخراج ویژگی های متمایزکننده و همچنین طراحی کلاس های رقابتی با استفاده از تنها یک مثال آموزشی برای هر کلاس ایجاد کرده است. همچنین یادگیری یک لحظه ای، مقایسه بحرانی الگوریتم های تشخیص حرکت را تسهیل می کند.

ژست های دست استفاده شده در سیستم های تشخیص ژست موجود، محدود به یک واژه نامه دقت انتخاب شده از ژست های نمادین (نشان ها و تصویرگرها) می شوند که عمدتاً برای صدور دستورات استفاده می شود. تشخیص ژست ها از تنظیم کننده ها، نمایشگرها و تطبیق دهنده ها (بخش 1) برای تعامل طبیعی بین انسان ها و ماشین ها ضروری است. الگوریتم ها با قابلیت های انحرافی بهتر، دارای توانایی تشخیص تعدادی گسترده از کلاس ها بدون آموزش گسترده، برای ساخت ماشین هایی با توانایی برای درک اهداف و الگوهای رفتاری بهتر می شود.

تاثیر تعاملات تجسمی از طریق حرکات بر افزایش پردازش بصری و توجه کمتر مورد بررسی قرار گرفته است. به عنوان مثال بررسی میزان مفید بودن ضبط توجه خاص توسط یک دست در حال موج برای توسعه مکانیسم توجه به یک روبات تعاملی مفید خواهد بود. یکی دیگر از رویکردهای آینده پژوهشی، کشف نواحی مغز پیشوا برای توسعه مدل های محاسباتی برای تقلید از فرآیند تشخیص الگو ژست است.

References

- [1] K. Hogan, R. Stubbs, *Can't get Through 8 Barriers to Communication*, Pelican Publishing Company, Gretna, LA, 2003.
- [2] D.K. Spencer, M.M. Sarah, R. Sabrina, *Gesture gives a hand to language and learning: perspectives from cognitive neuroscience, developmental psychology and education*, *Lang. Linguist. Compass* 2 (4) (2008) 569–588.
- [3] A. Kendon, *Current issues in the study of gesture*, in: *The Biological Foundation of Gestures: Motor and Semiotic Aspects*, Psychology Press, 1986, pp. 23–47.
- [4] L.L. Barker, L.A. Malandro, A.B. Deborah, *Nonverbal Communication*, 2nd ed., Addison-Wesley, MA, 1989.
- [5] A. Kendon, *Gesture and speech: how they interact*, in: John M. Wiemann, Randall P. Harrison (Eds.), *Nonverbal Interaction*, Sage Publications, Beverly Hills, 1983.
- [6] S. Berman, H. Stern, *Sensors for gesture recognition systems*, *IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev.* 42 (3) (2012) 277–290.
- [7] S. Mitra, T. Acharya, *Gesture recognition: a survey*, *IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev.* 37 (3) (2007) 311–324.
- [8] A. Erol, G. Bebis, M. Nicolescu, R.D. Boyle, X. Twombly, *Vision-based hand pose estimation: a review*, *Comput. Vis. Image Underst.* 108 (2007) 52–73.
- [9] S.C.W. Ong, S. Ranganath, *Automatic sign language analysis: a survey and the future beyond lexical meaning*, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (6) (2005) 873–891.
- [10] K.H. Lee, J.H. Kim, *An HMM based threshold model approach for gesture recognition*, *IEEE Trans. Pattern Anal. Mach. Intell.* 21 (10) (1999) 961–973.
- [11] H.S. Yoon, J. Soh, Y.J. Bae, H.S. Yang, *Hand gesture recognition using combined features of location, angle, and velocity*, *Pattern Recognit.* 34 (2001) 1491–1501.
- [12] A. Ramamoorthy, N. Vaswani, S. Chaudhury, S. Banerjee, *Recognition of dynamic hand gestures*, *Pattern Recognit.* 36 (2003) 2069–2081.
- [13] A. Just, S. Marcel, *A comparative study of two state-of-the-art sequence processing techniques for hand gesture recognition*, *Comput. Vis. Image Underst.* 113 (4) (2009) 532–543.
- [14] F.S. Chen, C.M. Fu, C.L. Huang, *Hand gesture recognition using a real-time tracking method and hidden Markov models*, *Image Vis. Comput.* 21 (2003) 745–758.
- [15] S. Marcel, O. Bernier, J.E. Viallet, D. Collobert, *Hand gesture recognition using input/output hidden Markov models*, in: *Proceedings of IEEE Automatic Face and Gesture Recognition, FG, 2000*, pp. 456–461.
- [16] N. Liu, B.C. Lovell, P.J. Kootsookos, *Evaluation of HMM training algorithms for letter hand gesture recognition*, in: *Proceedings of the 3rd IEEE International Symposium on Signal Processing and Information Technology, Darmstadt, Germany, 2003*, pp. 648–651.
- [17] D. Kim, J. Song, D. Kim, *Simultaneous gesture segmentation and recognition based on forward spotting accumulative HMMS*, *Pattern Recognit.* 40 (11) (2007) 3012–3026.
- [18] W.H.A. Wang, C.L. Tung, *Dynamic hand gesture recognition using hierarchical dynamic Bayesian networks through low-level image processing*, in: *Proceedings of the 7th International Conference on Machine Learning and Cybernetics, Kunming, China, 2008*, pp. 3247–3253.
- [19] C.L. Huang, M.S. Wu, S.H. Jeng, *Gesture recognition using the multi-PDM method and hidden Markov model*, *Image Vis. Comput.* 18 (11) (2000) 865–879.
- [20] J. Beh, D.K. Han, R. Durasiwami, H. Ko, *Hidden Markov model on a unit hypersphere space for gesture trajectory recognition*, *Pattern Recognit. Lett.* 36 (2014a) 144–153.
- [21] J. Beh, D. Han, H. Ko, *Rule-based trajectory segmentation for modeling hand motion trajectory*, *Pattern Recognit.* 47 (4) (2014b) 1586–1601.
- [22] J.H. Lee, T. Delbruck, et al., *Real-time gesture interface based on event-driven processing from stereo silicon retinas*, *IEEE Trans. Neural Netw. Learning Syst.* 25 (12) (2014) 2250–2263.
- [23] S. Theodorakis, V. Pitsikalis, P. Maragos, *Dynamic-static unsupervised sequentiality, statistical subunits and lexicon for sign language recognition*, *Image Vis. Comput.* 32 (8) (2014) 533–549.
- [24] H.I. Suk, B.K. Sin, S.W. Lee, *Hand gesture recognition based on dynamic Bayesian network framework*, *Pattern Recognit.* 43 (9) (2010) 3059–3072.
- [25] Q. Chen, N.D. Georganas, E.M. Petriu, *Hand gesture recognition using Haar-like features and a stochastic context-free grammar*, *IEEE Trans. Instrum. Meas.* 57 (8) (2008) 1562–1571.
- [26] G. Caridakis, K. Karpouzis, A. Drosopoulos, S. Kollias, *SOMM: self-organizing Markov map for gesture recognition*, *Pattern Recognit. Lett.* 31 (1) (2010) 52–59.
- [27] M. Abid, E. Petriu, E. Amjadian, *Dynamic sign language recognition for smart home interactive application using stochastic linear formal grammar*, *IEEE Trans. Instrum. Meas.* 64 (3) (2014) 596–605.
- [28] W.W. Kong, S. Ranganath, *Towards subject independent continuous sign language recognition: a segment and merge approach*, *Pattern Recognit.* 47 (3) (2014) 1294–1308.
- [29] M. Yeasin, S. Chaudhuri, *Visual understanding of dynamic hand gestures*, *Pattern Recognit.* 33 (11) (2000) 1805–1817.
- [30] P. Hong, M. Turk, T.S. Huang, *Gesture modeling and recognition using finite state machines*, in: *Proceedings of IEEE Automatic Face and Gesture Recognition, FG, 2000*, pp. 410–415.
- [31] J. Davis, M. Shah, *Recognizing hand gestures*, in: *Proceedings of the European Conference on Computer Vision, 1994*, pp. 331–340.
- [32] M.H. Yang, N. Ahuja, M. Tabb, *Extraction of 2D motion trajectories and its application to hand gesture recognition*, *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (8) (2002) 1061–1074.
- [33] M.H. Yang, N. Ahuja, *Extraction and classification of visual motion patterns for hand gesture recognition*, in: *Proceedings of IEEE Computer Vision and Pattern Recognition, CVPR, Santa Barbara, CA, USA, 1998*, pp. 892–897.
- [34] C.W. Ng, S. Ranganath, *Real-time gesture recognition system and application*, *Image Vis. Comput.* 20 (2002) 993–1007.
- [35] X.H. Shen, G. Hua, L. Williams, Y. Wu, *Dynamic hand gesture recognition: an exemplar-based approach from motion divergence fields*, *Image Vis. Comput.* 30 (3) (2012) 227–235.
- [36] J. Cheng, C. Xie, W. Bian, D.C. Tao, *Feature fusion for 3d hand gesture recognition by learning a shared hidden space*, *Pattern Recognit. Lett.* 33 (4) (2012) 476–484.
- [37] K.S. Patwardhan, S.D. Roy, *Hand gesture modelling and recognition involving changing shapes and trajectories, using a predictive eigentracker*, *Pattern Recognit. Lett.* 28 (2007) 329–334.
- [38] K. Daniel, M. John, M. Charles, *A person independent system for recognition of hand postures used in sign language*, *Pattern Recognit. Lett.* 31 (2010) 1359–1368.
- [39] M.C. Shin, L.V. Tsap, D.B. Goldgof, *Gesture recognition using Bezier curves for visualization navigation from registered 3-d data*, *Pattern Recognit.* 37 (5) (2004) 1011–1024.
- [40] T. Kuremoto, Y. Kinoshita, L. Feng, S. Watanabe, K. Kobayashi, M. Obayashi, *A gesture recognition system with retina-v1 model and one-pass dynamic programming*, *Neurocomputing* 116 (2012) 291–300.
- [41] A. Corradini, *Dynamic time warping for off-line recognition of a small gesture vocabulary*, in: *Proceedings of IEEE International Workshop on Computer Vision, ICCVW, 2001*, pp. 82–89.
- [42] J.F. Lichtenauer, E.A. Hendriks, M.J.T. Reinders, *Sign language recognition by combining statistical DTW and independent classification*, *IEEE Trans. Pattern Anal. Mach. Intell.* 30 (11) (2008) 2040–2046.
- [43] K. Tohyama, K. Fukushima, *Neural network model for extracting optic flow*, *Neural Netw.* 18 (5–6) (2005) 549–556.
- [44] P.K. Pisharady, P. Vadakkepat, A.P. Loh, *Attention based detection and recognition of hand postures against complex backgrounds*, *Int. J. Comput. Vis.* 101 (3) (2013) 403–419.
- [45] P.K. Pisharady, P. Vadakkepat, A.P. Loh, *Hand posture and face recognition using a fuzzy-rough approach*, *Int. J. Humanoid Robot.* 7 (3) (2010) 331–356.
- [46] P.K. Pisharady, P. Vadakkepat, A.P. Loh, *Fuzzy-rough discriminative feature selection and classification algorithm, with application to microarray and image datasets*, *Appl. Soft Comput.* 11 (4) (2011) 3429–3440.
- [47] P.K. Pisharady, Q.S.H. Stephanie, P. Vadakkepat, A.P. Loh, *Hand posture recognition using neuro-biologically inspired features*, in: *Proceedings of International Conference on Computational Intelligence, Robotics and Autonomous Systems, CIRAS, Bangalore, 2010*.
- [48] J. Alon, V. Athitsos, Q. Yuan, S. Sclaroff, *A unified framework for gesture recognition and spatiotemporal gesture segmentation*, *IEEE Trans. Pattern Anal. Mach. Intell.* 31 (9) (2009) 1685–1699.
- [49] A. Licsar, T. Sziranyi, *Dynamic training of hand gesture recognition system*, in: J. Kittler, M. Petrou, M. Nixon (Eds.), *Proceedings of International Conference on Pattern Recognition, ICPR, Cambridge, England, 2004*, pp. 971–974.
- [50] A. Licsar, T. Sziranyi, *User-adaptive hand gesture recognition system with interactive training*, *Image Vis. Comput.* 23 (2005) 1102–1114.
- [51] M. Zhao, F.K.H. Quek, X. Wu, *RIEVL: recursive induction learning in hand gesture recognition*, *IEEE Trans. Pattern Anal. Mach. Intell.* 20 (11) (1998) 1174–1185.
- [52] X. Teng, B. Wu, W. Yu, C. Liu, *A hand gesture recognition system based on local linear embedding*, *J. Vis. Lang. Comput.* 16 (2005) 442–454.
- [53] S. Marcel, O. Bernier, J.E. Viallet, D. Collobert, *Hand gesture recognition using input/output hidden Markov models*, in: *Proceedings of IEEE Automatic Face and Gesture Recognition, FG, 2000*, pp. 456–461.
- [54] N. Liu, B.C. Lovell, P.J. Kootsookos, *Evaluation of HMM training algorithms for letter hand gesture recognition*, in: *Proceedings of the 3rd IEEE International Symposium on Signal Processing and Information Technology, Darmstadt, Germany, 2003*, pp. 648–651.
- [55] D. Kim, J. Song, D. Kim, *Simultaneous gesture segmentation and recognition based on forward spotting accumulative HMMS*, *Pattern Recognit.* 40 (11) (2007) 3012–3026.
- [56] W.H.A. Wang, C.L. Tung, *Dynamic hand gesture recognition using hierarchical dynamic Bayesian networks through low-level image processing*, in: *Proceedings of the 7th International Conference on Machine Learning and Cybernetics, Kunming, China, 2008*, pp. 3247–3253.
- [57] C.L. Huang, M.S. Wu, S.H. Jeng, *Gesture recognition using the multi-PDM method and hidden Markov model*, *Image Vis. Comput.* 18 (11) (2000) 865–879.
- [58] J. Beh, D.K. Han, R. Durasiwami, H. Ko, *Hidden Markov model on a unit hypersphere space for gesture trajectory recognition*, *Pattern Recognit. Lett.* 36 (2014a) 144–153.
- [59] J. Beh, D. Han, H. Ko, *Rule-based trajectory segmentation for modeling hand motion trajectory*, *Pattern Recognit.* 47 (4) (2014b) 1586–1601.
- [60] J.H. Lee, T. Delbruck, et al., *Real-time gesture interface based on event-driven processing from stereo silicon retinas*, *IEEE Trans. Neural Netw. Learning Syst.* 25 (12) (2014) 2250–2263.
- [61] S. Theodorakis, V. Pitsikalis, P. Maragos, *Dynamic-static unsupervised sequentiality, statistical subunits and lexicon for sign language recognition*, *Image Vis. Comput.* 32 (8) (2014) 533–549.
- [62] H.I. Suk, B.K. Sin, S.W. Lee, *Hand gesture recognition based on dynamic Bayesian network framework*, *Pattern Recognit.* 43 (9) (2010) 3059–3072.
- [63] Q. Chen, N.D. Georganas, E.M. Petriu, *Hand gesture recognition using Haar-like features and a stochastic context-free grammar*, *IEEE Trans. Instrum. Meas.* 57 (8) (2008) 1562–1571.
- [64] K.S. Patwardhan, S.D. Roy, *Hand gesture modelling and recognition involving changing shapes and trajectories, using a predictive eigentracker*, *Pattern Recognit. Lett.* 28 (2007) 329–334.
- [65] K. Daniel, M. John, M. Charles, *A person independent system for recognition of hand postures used in sign language*, *Pattern Recognit. Lett.* 31 (2010) 1359–1368.
- [66] M.C. Shin, L.V. Tsap, D.B. Goldgof, *Gesture recognition using Bezier curves for visualization navigation from registered 3-d data*, *Pattern Recognit.* 37 (5) (2004) 1011–1024.
- [67] T. Kuremoto, Y. Kinoshita, L. Feng, S. Watanabe, K. Kobayashi, M. Obayashi, *A gesture recognition system with retina-v1 model and one-pass dynamic programming*, *Neurocomputing* 116 (2012) 291–300.
- [68] A. Corradini, *Dynamic time warping for off-line recognition of a small gesture vocabulary*, in: *Proceedings of IEEE International Workshop on Computer Vision, ICCVW, 2001*, pp. 82–89.
- [69] J.F. Lichtenauer, E.A. Hendriks, M.J.T. Reinders, *Sign language recognition by combining statistical DTW and independent classification*, *IEEE Trans. Pattern Anal. Mach. Intell.* 30 (11) (2008) 2040–2046.
- [70] K. Tohyama, K. Fukushima, *Neural network model for extracting optic flow*, *Neural Netw.* 18 (5–6) (2005) 549–556.
- [71] P.K. Pisharady, P. Vadakkepat, A.P. Loh, *Attention based detection and recognition of hand postures against complex backgrounds*, *Int. J. Comput. Vis.* 101 (3) (2013) 403–419.
- [72] P.K. Pisharady, P. Vadakkepat, A.P. Loh, *Hand posture and face recognition using a fuzzy-rough approach*, *Int. J. Humanoid Robot.* 7 (3) (2010) 331–356.
- [73] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, A. Blake, *Real-time human pose recognition in parts from single depth images*, in: *Proceedings of IEEE Computer Vision and Pattern Recognition, CVPR, Colorado Springs, 2011*.
- [74] J. Han, L. Shao, D. Xu, J. Shotton, *Enhanced computer vision with microsoft Kinect sensor: a review*, *IEEE Trans. Cybern.* 43 (5) (2013) 1318–1334.
- [75] R. Muñoz-Salinas, R. Medina-Carnicer, F.J. Madrid-Cuevas, A. Carrama-Puyato, *Depth silhouettes for gesture recognition*, *Pattern Recognit. Lett.* 29 (3) (2008) 319–329.
- [76] F. Dominio, M. Donadeo, P. Zanuttigh, *Combining multiple depth-based descriptors for hand gesture recognition*, *Pattern Recognit. Lett.* 50 (2014) 101–111.
- [77] M.R. Malfredy, I. Inwogu, V. Govindaraju, *A temporal Bayesian model for classifying, detecting and localizing activities in video sequences*, in: *Proceedings of IEEE Computer Vision and Pattern Recognition Workshops, CVPRW, 2012*, pp. 43–48.
- [78] J. Sung, C. Ponce, B. Selman, A. Saxena, *Human activity detection from RGBD images*, in: *Proceedings of AAAI workshop on Pattern, Activity and Intent Recognition, PAIR, 2011*.
- [79] J. Sung, C. Ponce, B. Selman, A. Saxena, *Unstructured human activity detection*

- [53] M. Hasanuzzamana, T. Zhanga, V. Ampornaramveth, H. Gotoda, Y. Shirai, H. Ueno, Adaptive visual gesture recognition for human-robot interaction using a knowledge-based software platform, *Robot. Auton. Syst.* 55 (8) (2007) 643–657.
- [54] Y. Wu, T.S. Huang, View-independent recognition of hand postures, in: Proceedings of IEEE Computer Vision and Pattern Recognition, CVPR, vol. 2, 2000, pp. 88–94.
- [55] O. Eng-Jou, R. Bowden, A boosted classifier tree for hand shape detection, in: Proceedings of IEEE International Conference on Face and Gesture Recognition, FG, 2004, pp. 889–894.
- [56] E. Stergiopoulou, N. Papamarkos, Hand gesture recognition using a neural network shape fitting technique, *Eng. Appl. Artif. Intell.* 22 (8) (2009) 1141–1158.
- [57] P. Premaratne, S. Ajaz, M. Premaratne, Hand gesture tracking and recognition system using Lucas-Kanade algorithms for control of consumer electronics, *Neurocomputing* 116 (2013) 242–249.
- [58] D.Y. Huang, W.C. Hu, S.H. Chang, Gabor filter-based hand-pose angle estimation for hand gesture recognition under varying illumination, *Expert Syst. Appl.* 38 (5) (2011) 6031–6042.
- [59] P.K. Pisharady, P. Vadakkepat, A.P. Loh, Computational Intelligence in Multi-Feature Visual Pattern Recognition – Hand Posture and Face Recognition Using Biologically Inspired Approaches, Springer, Singapore, 2014.
- [60] P.K. Pisharady, Computational Intelligence Techniques in Visual Pattern Recognition (Ph.D. thesis), National University of Singapore, August, 2011.
- [61] S.S. Ge, Y. Yang, T.H. Lee, Hand gesture recognition and tracking based on distributed locally linear embedding, *Image Vis. Comput.* 26 (2008) 1607–1620.
- [62] P.K. Pisharady, P. Vadakkepat, A.P. Loh, Graph matching based hand posture recognition using neuro-biologically inspired features, in: Proceedings of International Conference on Control, Automation, Robotics and Vision, ICARCV, Singapore, 2010.
- [63] J. Triesch, C. Malsburg, A system for person-independent hand posture recognition against complex backgrounds, *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (12) (2001) 1449–1453.
- [64] J. Triesch, C. Malsburg, Robust classification of hand postures against complex backgrounds, in: Proceedings of IEEE International Conference on Face and Gesture Recognition, FG, Killington, VT, USA, 1996, pp. 170–175.
- [65] J. Triesch, C. Malsburg, A gesture interface for human-robot-interaction, in: Proceedings of IEEE International Conference on Face and Gesture Recognition, FG, Nara, Japan, 1998, pp. 546–551.
- [66] J. Triesch, C. Eckes, Object recognition with multiple feature types, in: Proceedings of the 8th International Conference on Artificial Neural Networks, ICANN'98, Skovde, Sweden, 1998.
- [67] Y.-T. Li, J.P. Wachs, HEGM: a hierarchical elastic graph matching for hand gesture recognition, *Pattern Recognit.* 47 (1) (2014) 80–88.
- [68] V. Athitsos, S. Sclaroff, Estimating 3D hand pose from a cluttered image, in: Proceedings of IEEE Computer Vision and Pattern Recognition, CVPR, vol. 2, 2003, pp. 432–439.
- [69] E. Ueda, Y. Matsumoto, M. Imai, T. Ogasawara, A hand-pose estimation for vision-based human interfaces, *IEEE Trans. Ind. Electron.* 50 (4) (2003) 676–684.
- [70] X. Yin, M. Xie, Estimation of the fundamental matrix from uncalibrated stereo hand images for 3d hand gesture recognition, *Pattern Recognit.* 36 (2003) 567–584.
- [71] J. Lee, T. Kunii, Model-based analysis of hand posture, *IEEE Comput. Graph. Appl.* 15 (5) (1995) 77–86.
- [72] A. El-Sawah, N.D. Georganas, E. Petriu, A prototype for 3-D hand tracking and posture estimation, *IEEE Trans. Instrum. Meas.* 57 (8) (2008) 1627–1636.
- [73] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, *Science* 290 (5500) (2000) 2323–2326.
- [74] D. Conte, P. Foggia, C. Sansone, M. Vento, Thirty years of graph matching in pattern recognition, *Int. J. Pattern Recognit. Artif. Intell.* 18 (3) (2004) 265–298.
- [75] M. Lades, J.C. Vorbruggen, J. Buhmann, J. Lange, C. Malsburg, R.P. Wurtz, W. Konen, Distortion invariant object recognition in the dynamic link architecture, *IEEE Trans. Comput.* 42 (3) (1993) 300–311.
- [76] L. Wiskott, J.M. Fellous, N. Kruger, C. Malsburg, Face recognition by elastic bunch graph matching, *IEEE Trans. Pattern Anal. Mach. Intell.* 19 (7) (1997) 775–779.
- [77] Z. Zhang, Microsoft kinect sensor and its effect, *IEEE MultiMed.* 19 (2) (2012) 04–10.
- [106] M.G. Jacob, J.P. Wachs, Context-based hand gesture recognition for the operating room, *Pattern Recognit. Lett.* 36 (2014) 196–203.
- [107] O. Mendels, H. Stern, S. Berman, User identification for home entertainment based on free-air hand motion signatures, *IEEE Trans. Syst. Man Cybern.: Syst.* 44 (11) (2014) 1461–1473.
- [108] S.-Z. Li, B. Yu, W. Wu, S.-Z. Su, R.-R. Ji, Feature learning based on SAE-PCA network for human gesture recognition in RGBD images, *Neurocomputing* 151 (2015) 565–573.
- [109] Y. Ming, Hand fine-motion recognition based on 3d mesh mosift feature descriptor, *Neurocomputing* 151 (2015) 574–582.
- [110] R. Schramm, C.R. Jung, E.R. Miranda, Dynamic time warping for music conducting gestures evaluation, *IEEE Trans. Multimed.* 17 (2) (2015) 243–255.
- [111] F.A. Kondori, S. Yousefi, J.-P. Kouma, L. Liu, H. Li, Direct hand pose estimation for immersive gestural interaction, *Pattern Recognit. Lett.* (2015) in press, doi:10.1016/j.patrec.2015.03.013.
- from RGBD images, in: Proceedings of IEEE International Conference on Robotics and Automation, ICRA, 2012.
- [85] P.K. Pisharady, M. Saerbeck, Kinect based body posture detection and recognition system, in: Proceedings of International Conference on Graphic and Image Processing (ICGIP), 2012.
- [86] M.B. Holte, T. Moeslund, P. Fihl, Fusion of range and intensity information for view invariant gesture recognition, in: Proceedings of IEEE Computer Vision and Pattern Recognition Workshops, CVPRW, 2008, pp. 1–7.
- [87] M. Van den Bergh, D. Carton, R. De Nijs, N. Mitsou, C. Landsiedel, K. Kuehnlenz, D. Wollherr, L. Van Gool, M. Buss, Real-time 3d hand gesture interaction with a robot for understanding directions from humans, in: Proceedings of IEEE International Symposium on Robot and Human Interactive Communication, IEEE RO-MAN, 2011.
- [88] R. Zhou, Y. Junsong, Z. Zhengyou, Robust hand gesture recognition based on finger-earth movers distance with a commodity depth camera, in: Proceedings of ACM Multimedia, 2011.
- [89] L. Gallo, A.P. Placitelli, M. Ciampi, Controller-free exploration of medical image data: experiencing the Kinect, in: Proceedings of International Symposium on Computer-Based Medical Systems, CBMS, 2011.
- [90] W. Di, Z. Fan, S. Ling, One shot learning gesture recognition from RGBD images, in: Proceedings of IEEE Computer Vision and Pattern Recognition Workshops, CVPRW, 2012.
- [91] C. Keskin, F. Kirac, Y. Kara, L. Akarun, Randomized decision forests for static and dynamic hand shape classification, in: Proceedings of IEEE Computer Vision and Pattern Recognition Workshops, CVPRW, 2012, pp. 31–46.
- [92] Y.M. Lui, A least squares regression framework on manifolds and its application to gesture recognition, in: Proceedings of IEEE Computer Vision and Pattern Recognition Workshops, CVPRW, 2012, pp. 13–18.
- [93] K. Lai, J. Konrad, P. Ishwar, A gesture-driven computer interface using Kinect, in: Proceedings of IEEE Southwest Symposium on Image Analysis and Interpretation, SSIAI, 2012, pp. 185–188.
- [94] D. Frolova, H. Stern, S. Berman, Most probable longest common subsequence for recognition of gesture character input, *IEEE Trans. Cybern.* 43 (3) (2013) 871–880.
- [95] H. Jiang, B.S. Duerstock, J.P. Wachs, A machine vision-based gestural interface for people with upper extremity physical impairments, *IEEE Trans. Syst. Man Cybernet. Syst.* 44 (5) (2014) 630–641.
- [96] P.K. Pisharady, M. Saerbeck, A robust gesture detection and recognition algorithm for domestic robot interactions, in: Proceedings of International Conference on Control, Automation, Robotics and Vision, ICARCV, Singapore, 2014.
- [97] R. Krishnan, S. Sarkar, Conditional distance based matching for one-shot gesture recognition, *Pattern Recognit.* 48 (4) (2015) 1302–1314.
- [98] C. Zhang, T. Yingli, Histogram of 3d facets: a depth descriptor for human action and hand gesture recognition, *Comput. Vis. Image Underst.* (2015) in press, doi:10.1016/j.cviu.2015.05.010.
- [99] Y.M. Lui, Human gesture recognition on product manifolds, *J. Mach. Learn. Res.* 13 (2012) 3297–3321.
- [100] U. Mahbub, H. Imtiaz, T. Roy, M. Rahman, M. Ahad, A template matching approach of one-shot-learning gesture recognition, *Pattern Recognit. Lett.* 34 (2013) 1780–1788.
- [101] J. Wan, Q. Ruan, W. Li, S. Deng, One-shot learning gesture recognition from RGB-D data using bag of features, *J. Mach. Learn. Res.* 14 (2013) 2549–2582.
- [102] P.K. Pisharady, M. Saerbeck, Robust gesture detection and recognition using dynamic time warping and multi-class probability estimates, in: Proceedings of IEEE Symposium on Computational Intelligence for Multimedia, Signal and Vision Processing, CIMSIVP, 2013.
- [103] H. Cheng, Z. Dai, Z. Liu, Image-to-class dynamic time warping for 3d hand gesture recognition, in: Proceedings of IEEE International Conference on Multimedia and Expo, ICME, 2013.
- [104] H. Cheng, J. Luo, X. Chen, A windowed dynamic time warping approach for 3d continuous hand gesture recognition, in: Proceedings of IEEE International Conference on Multimedia and Expo, ICME, 2014.
- [105] E. Ohn-Bar, M. Trivedi, Hand gesture recognition in real time for automotive interfaces: a multimodal vision-based approach and evaluations, *IEEE Trans. Intell. Transp. Syst.* 15 (6) (2014) 2368–2377.
- [134] S. Ruffieux, D. Lalanne, E. Mugellini, Chairgest: a challenge for multimodal mid-air gesture recognition for close HCI, in: Proceedings of 15th ACM on International Conference on Multimodal Interaction, ICMI, 2013.
- [135] L. Liu, L. Shao, Learning discriminative representations from RGB-D video data, in: Proceedings of International Joint Conference on Artificial Intelligence, IJCAI, 2013.
- [136] J. Zhuolin, L.S. Davis, Recognizing actions by shape-motion prototype trees, in: Proceedings of IEEE International Conference on Computer Vision, ICCV, 2009, pp. 444–451.
- [137] T.-K. Kim, S.-F. Wong, R. Cipolla, Tensor canonical correlation analysis for action classification, in: Proceedings of IEEE Computer Vision and Pattern Recognition, CVPR, 2007, pp. 1–8.
- [138] J. Tompson, M. Stein, Y. Lecun, K. Perlin, Real-time continuous pose recovery of human hands using convolutional networks, *ACM Trans. Graph.* 33 (2014) 169:01–169:10.

- [112] Z. Ren, J. Yuan, J. Meng, Z. Zhang, Robust part-based hand gesture recognition using Kinect sensor, *IEEE Trans. Multimed.* 15 (5) (2013) 1110–1120.
- [113] R. Zhou, M. Jingjing, Y. Junsong, Depth camera based hand gesture recognition and its applications in human computer interaction, in: *Proceedings of International Conference on Information, Communications and Signal Processing, ICICS, 2011*.
- [114] Y. Li, Hand gesture recognition using kinect, in: *Proceedings of the 3rd International Conference on Software Engineering and Service Science, ICSESS, 2012*.
- [115] P. Doliotis, V. Athitsos, D. Kosmopoulos, S. Perantonis, Hand shape and 3d pose estimation using depth data from a single cluttered frame, *Adv. Vis. Comput.* 7431 (2012) 148–158.
- [116] F. Kirac, Y.E. Kara, L. Akarun, Hierarchically constrained 3d hand pose estimation using regression forests from single frame depth data, *Pattern Recognit. Lett.* 50 (2014) 91–100.
- [117] Y. Yao, Y. Fu, Contour model based hand-gesture recognition using Kinect sensor, *IEEE Trans. Circuits Syst. Video Technol.* 24 (11) (2014) 1935–1944.
- [118] X. Suau, M. Alcoverro, A. Lopez-Mendez, J. Ruiz-Hidalgo, J.R. Casas, Real-time fingertip localization conditioned on hand gesture classification, *Image Vis. Comput.* 32 (8) (2014) 522–532.
- [119] F. Kirac, Y.E. Kara, L. Akarun, Hierarchically constrained 3d hand pose estimation using regression forests from single frame depth data, *Pattern Recognit. Lett.* 50 (2014) 91–100.
- [120] C. Wang, Z. Liu, S.-C. Chan, Superpixel-based hand gesture recognition with Kinect depth camera, *IEEE Trans. Multimed.* 17 (1) (2015) 29–39.
- [121] A. Bobick, J. Davis, The recognition of human movement using temporal templates, *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (3) (2001).
- [122] H. Liang, J. Yuan, D. Thalmann, Parsing the hand in depth images, *IEEE Trans. Multimed.* 16 (5) (2014) 1241–1253.
- [123] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (3) (2002) 509–522.
- [124] F. Erden, A.E. Cetin, Hand gesture based remote control system using infrared sensors and a camera, *IEEE Trans. Consum. Electron.* 60 (4) (2014) 675–680.
- [125] F. Wang, C. Zhang, Feature extraction by maximizing the average neighborhood margin, in: *Proceedings of IEEE Computer Vision and Pattern Recognition, CVPR, 2007*, pp. 1–8.
- [126] F. Simon, M.M. Helena, K. Pushmeet, N. Sebastian, Instructing people for training gestural interactive systems, in: *Proceedings of International Conference on Human Factors in Computing Systems, CHI, ACM, 2012*, pp. 1737–1746.
- [127] I. Guyon, V. Athitsos, P. Jangyodsuk, B. Hamner, H. Escalante, Chalearn gesture challenge: design and first results, in: *Proceedings of IEEE Computer Vision and Pattern Recognition Workshops, CVPRW, 2012*, pp. 1–6.
- [128] S. Escalera, J. González, X. Baró, M. Reyes, O. Lopes, I. Guyon, V. Athitsos, H.J. Escalante, Multi-modal gesture recognition challenge 2013: dataset and results, in: *Proceedings of the 15th ACM International Conference on Multimodal Interaction, ICMI, Sydney, Australia, 2013*.
- [129] Y. Chuang, L. Chen, G. Chen, Saliency-guided improvement for hand posture detection and recognition, *Neurocomputing* 133 (2014) 404–415.
- [130] M. Chen, G. AlRegib, B.H. Juang, 6dmg: a new 6d motion gesture database, in: *Proceedings of IEEE Computer Vision and Pattern Recognition Workshops, CVPRW, 2011*.
- [131] A. Just, O. Bernier, S. Marcel, HMM and IOHMM for the recognition of mono- and bi-manual 3d hand gestures, in: *Proceedings of British Machine Vision Conference, BMVC, 2004*.
- [132] S. Yale, D. David, D. Randall, Tracking body and hands for gesture recognition: NATOPS aircraft handling signals database, in: *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, FG, Santa Barbara, CA, 2011*, pp. 500–506.
- [133] S. Marcel, Hand posture recognition in a body-face centered space, in: *Proceedings of the Conference on Human Factors in Computer Systems, CHI, 1999*.
- [139] E. Kollorz, J. Penne, J. Hornegger, A. Barke, Gesture recognition with a time-of-flight camera, *Int. J. Intell. Syst. Technol. Appl.* 5 (2007) 334–343.
- [140] M. Kawulok, J. Kawulok, J. Nalepa, Spatial-based skin detection using discriminative skin-presence features, *Pattern Recognit. Lett.* 41 (2014) 3–13.
- [141] N. Pugeault, R. Bowden, Spelling it out: real-time ASL fingerspelling recognition, in: *Proceedings of International Conference on Computer Vision, ICCV 2012, 2011*.
- [142] Q. Chen, S. Xiao, W. Yichen, T. Xiaoou, S. Jian, Realtime and robust hand tracking from depth, in: *Proceedings of IEEE Computer Vision and Pattern Recognition, CVPR, 2014*.
- [143] C. Neidle, A. Thangali, S. Sclaroff, Challenges in development of the american sign language lexicon video dataset corpus, in: *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, LREC, 2012*.
- [144] R. Kehl, L. Van Gool, Real-time pointing gesture recognition for an immersive environment, in: *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, FG, Seoul, Korea, 2004*, pp. 577–582.
- [145] N. Jovic, B. Brumitt, B. Meyers, S. Harris, T. Huang, Detection and estimation of pointing gestures in dense disparity maps, in: *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, FG, Grenoble, France, 2000*, pp. 1000–1007.
- [146] C.B. Park, S.W. Lee, Real-time 3d pointing gesture recognition for mobile robots with cascade HMM and particle filter, *Image Vis. Comput.* 29 (1) (2011) 51–63.
- [147] C.B. Park, M.C. Roh, S.W. Lee, Real-time 3D pointing gesture recognition in mobile space, in: *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, FG, 2008*.
- [148] K. Nickel, R. Stiefelhagen, Visual recognition of pointing gestures for human-robot interaction, *Image Vis. Comput.* 25 (12) (2007) 1875–1884.
- [149] J.L. Raheja, A. Chaudhary, S. Maheshwari, Hand gesture pointing location detection, *Optik* 125 (3) (2014) 993–996.
- [150] M. Pateraki, H. Baltzakis, P. Trahanias, Visual estimation of pointed targets for robot guidance via fusion of face pose and hand orientation, *Comput. Vis. Image Underst.* 120 (2014) 1–13.
- [151] M. Pateraki, H. Baltzakis, P. Trahanias, Visual estimation of pointed targets for robot guidance via fusion of face pose and hand orientation, *Proceedings of IEEE International Conference on Computer Vision Workshops, ICCVW 2011, 2011*.
- [152] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, T. Poggio, Robust object recognition with cortex-like mechanisms, *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (3) (2007) 411–426.
- [153] H. Kang, W.L. Chang, K.C. Jung, Recognition-based gesture spotting in video games, *Pattern Recognit. Lett.* 25 (15) (2004) 1701–1714.
- [154] Y. Yin, R. Davis, Gesture spotting and recognition using saliency detection and concatenated hidden Markov models, in: *Proceedings of 15th ACM International conference on multimodal interaction, ICMI, 2013*, pp. 489–494.
- [155] R.D. Yang, S. Sarkar, B. Loeding, Handling movement epenthesis and hand segmentation ambiguities in continuous sign language recognition using nested dynamic programming, *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (3) (2010) 462–477.
- [156] L. Hong, G. Michael, Model-based segmentation and recognition of dynamic gestures in continuous video streams, *Pattern Recognit.* 44 (8) (2011).
- [157] S. Sarkar, B. Loeding, R. Yang, S. Nayak, A. Parashar, Segmentation-robust representations, matching, and modeling for sign language, in: *Proceedings of IEEE Computer Vision and Pattern Recognition Workshops, CVPRW, 2011*, pp. 13–19.
- [158] R.D. Yang, S. Sarkar, Coupled grouping and matching for sign and gesture recognition, *Comput. Vis. Image Underst.* 113 (6) (2009) 663–681.
- [159] A. Betancourt, P. Morerio, E. Barakova, L. Marcenaro, M. Rauterberg, C. Regazzoni, A dynamic approach and a new dataset for hand-detection in first person vision, in *International Conference on Computer Analysis of Images and Patterns (2015)*.