

ارائه یک روش تشخیص زبان علامت مبتنی بر رویکرد MLRF فازی با استفاده از اطلاعات عمق تصویر

ندا خانبانی^۱، کارشناس ارشد؛ امیرمسعود افتخاری مقدم^۲، دانشیار

۱- دانشکده مهندسی رایانه و فناوری اطلاعات - واحد قزوین - دانشگاه آزاد اسلامی - قزوین - ایران - neda.khanbani@yahoo.com

۲- دانشکده مهندسی رایانه و فناوری اطلاعات - واحد قزوین - دانشگاه آزاد اسلامی - قزوین - ایران - eftekhari@qiau.ac.ir

چکیده: تشخیص اشارات یک وظیفه مهم در حوزه‌هایی مانند تعامل انسان و کامپیوتر، تشخیص زبان اشاره و رباتیک می‌باشد. این مقاله یک الگوریتم تشخیص اشارات برای زبان اشاره آمریکایی مختص ناشنوایان ارائه می‌دهد. در این مقاله، یک روش بسیار دقیق برای تشخیص اشارات ایستا از یک تصویر عمق و درخت تصمیم فازی پیشنهاد شده است. در ابتدا، از تصاویر عمق برای استنتاج ویژگی‌های ثابت چرخش، تفسیر و مقیاس‌پذیری استفاده شده است. سپس، با استفاده از یک جنگل تصادفی چندلایه‌ای فازی، برای طبقه‌بندی بردارهای ویژگی آموزش داده، که برای تشخیص اشارات دست واگذار شده است. برای بهبود طبقه‌بندی، از انعطاف‌پذیری منطق فازی و مجموعه‌های فازی استفاده شده است. در حقیقت، این رویکرد، استحکام سیستم‌های طبقه‌بندی چندگانه، قدرت تصادفی بودن برای افزایش تنوع درختان، و انعطاف‌پذیری منطق فازی و مجموعه‌های فازی را برای مدیریت داده‌های ناقص ترکیب می‌کند.

واژه‌های کلیدی: جنگل تصادفی، درخت تصمیم فازی، مجموعه‌های فازی، جنگل تصادفی چندلایه‌ای فازی.

Presentation of a Sign Language Recognition Method Based on Fuzzy MLRF Approach Using Depth Image Information

N. Khanbani¹, MSc; A. M. Eftekhari Moghadam², Associate Professor

1- Department of Computer and IT Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran, Email: neda.khanbani@yahoo.com

2- Department of Computer and IT Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran, Email: eftekhari@qiau.ac.ir

Abstract: Gesture recognition is an important task in areas such as human computer interaction, sign language recognition and robotics. This paper presents a Gesture recognition algorithm for recognizing American Sign Language. In this article propose a highly precise method to recognize static gestures from a depth image and fuzzy decision tree. At first, article uses of depth images to derive rotation-, translation- and scale- invariant features. Then, by using of a fuzzy multi-layered random forest, trains to classify the feature vectors, which yields to the recognition of the hand signs. For improve classifying, uses flexibility of fuzzy logic and the fuzzy sets. This approach combines the robustness of multiple classifier systems, the power of the randomness to increase the diversity of the trees, and the flexibility of fuzzy logic and fuzzy sets for imperfect data management.

Keywords: Random forest, fuzzy decision tree, fuzzy sets, fuzzy multi-layered random forest.

تاریخ ارسال مقاله: ۱۳۹۵/۰۳/۰۳

تاریخ اصلاح مقاله: ۱۳۹۵/۰۵/۲۳، ۱۳۹۵/۰۶/۲۷ و ۱۳۹۵/۰۷/۱۸

تاریخ پذیرش مقاله: ۱۳۹۵/۰۷/۲۴

نام نویسنده مسئول: امیرمسعود افتخاری مقدم

نشانی نویسنده مسئول: ایران - قزوین - بلوار نخبگان - دانشگاه آزاد اسلامی واحد قزوین - دانشکده رایانه و فناوری اطلاعات.

۱- مقدمه

زبان اشاره شکل اولیه‌ای برای ارتباط شنوایی از دست‌رفته است. برای ارتباط یک شخص ساده با مردم ناشنوا، یک مترجم معمولاً زبان اشاره و زبان طبیعی را به یکدیگر ترجمه می‌کند. ولی از آنجایی که در دنیای مدرن امروز، ارتباطات شکل گسترده‌تری به خود گرفته است، می‌توان از تکنولوژی برای بهبود روابط ناشنوایان با سایر مردم و ابزارهای ارتباطی استفاده کرد. یکی از قسمت‌های اصلی بسیاری از زبان‌های اشاره و همچنین زبان اشاره آمریکایی^۱ اشارات دست است. همچنین از مشکلات مهم در بینایی کامپیوتر و یادگیری ماشین، زبان اشاره و تشخیص اشارات است. علاوه بر این، تحقیقات کمی در این حوزه انجام گرفته است. اخیراً، کارهای زیادی در اشارات ایستا و پویا و تشخیص ژست مشاهده شده است.

۱-۱- مروری بر کارهای پیشین

تشخیص اشاره با یک دوربین، مشکل چالش‌برانگیزی است که سال‌های طولانی حل نشده باقی مانده است. کارهای جدیدتر از انواع مختلفی از پیچیدگی تصویر برای تشکیل دادن بردارهای ویژگی مبنی بر یک تصویر RGB تک از دست استفاده کرده‌اند. در [۱]، نویسندگان از خانواده‌های موجی، محاسبه شده در تصاویر لبه، به عنوان ویژگی‌هایی برای آموزش یک شبکه عصبی در ۲۴ دسته مختلف از اشارات استفاده می‌کنند. ویژگی‌های شبه‌هارلت^۲ محاسبه شده در تصاویر سطح خاکستری و در شبیح‌ها در [۳] برای طبقه‌بندی ۱۰ شکل دست استفاده شده‌اند. تحلیل عنصر اصلی^۳ (PCA) به طور مستقیم در تصاویر برای استنتاج یک زیرفضای ژست‌های دست به کار رفته است، که سپس برای طبقه‌بندی ژست‌های دست ([۲] را ببینید) استفاده شده است.

در [۴]، یک تغییر از توصیف‌گرهای HOG برای بررسی اشارات ایستای زبان اشاره انگلیسی به کار گرفته شده است. ویژگی سیفت^۴ مبنی بر توصیف برای تشخیص اشارات زبان اشاره آمریکایی در [۲] استفاده شده است. تمام این متدها به سنگینی شرایط روشنایی، نمایش شیء و پیش‌زمینه بستگی دارند. بعلاوه، نرمال‌سازی ژست قبل از محاسبه ویژگی درخواست شده است، زیرا ویژگی‌های ذکر شده در بالا ثابت-چرخش، ثابت-موقعیت و ثابت-مقیاس نیستند. به‌علت نمایش برد حسگرها، چندین شکاف در حوزه تشخیص با استفاده از یک تک‌دوربین وجود دارد.

پیشرفت‌های شناخته‌شده‌تر در تشخیص ژست انسان با استفاده از حسگر مایکروسافت کینکت^۵ [۵] هستند. نویسندگان از یک نوع مخصوصی از ویژگی‌ها استفاده کرده‌اند، که به طور مستقیم در تصاویر عمق برای طبقه‌بندی قسمت‌های بدن انسان محاسبه شده است. همان ویژگی‌ها اخیراً برای مشکل تخمین ژست دست [۶] و طبقه‌بندی شکل [۴] به کار رفته‌اند.

کشف و قطع‌بندی دست در حالی از قابل دسترس بودن داده عمق یک وظیفه ساده‌تر هست، بنابراین یک تعدادی از متدها به‌تنهایی در

قطع‌بندی استنتاج‌شده از داده عمق، مثل متد مبنی بر حدفاصل، شرح داده‌شده در [۷]، و متد مبنی بر قانون، ارائه‌شده در [۸] ظاهر شده است. [۹] نیز الگوریتمی مبنی بر برش نرمالیزه گراف^۶ ارائه کرده است. در این مقاله، ضمن بیان هم‌ارزی رابطه ریاضی حاکم بر مسئله بدون مربی Ncut با معیار Fisher-Rao در طبقه‌بندی با مربی، از نگاهی نو به مسئله انتخاب بردار ویژه پرداخته شده است. همچنین، با پیشنهاد معیاری کارا از دیدگاه Fisher-Rao، گزینش و مرتب‌سازی بردارهای ویژه در مسئله هم‌ارز Ncut آن انجام شده است.

در بحث خوشه‌بندی نیز، الگوریتم‌های زیادی مطرح شده است. یکی از این روش‌ها در [۱۰] پیشنهاد شده است. در این مقاله، از خوشه‌بندی اسناد شبه‌بازخورد^۷ حاصل از نتایج اولیه، بر اساس شباهت مبتنی بر کلمه برای قرار دادن شبیه‌ترین اسناد کنار هم استفاده می‌شوند. تعدادی از خوشه‌ها طبق محتوایشان به عنوان خوشه‌های بازخورد انتخاب می‌شوند و از خوشه‌های بازخورد، اسناد رتبه‌بالا به عنوان اسناد بازخورد انتخاب می‌شوند. سپس، یک سند ترکیبی از روی اسناد انتخابی تشکیل می‌شوند و کلمات سند ترکیبی بر اساس تابع رتبه‌بندی^۸ مرتب می‌شوند. سپس، کلمات رتبه‌بالا برای بسط انتخاب می‌شوند.

یکی دیگر از روش‌های خوشه‌بندی، در مقاله [۱۱] ارائه شده، که یک الگوریتم خوشه‌بندی فازی با وزن‌دهی اتوماتیک محلی پیشنهاد می‌دهد. در این مقاله، تحلیل‌های ریاضی برای به‌دست آوردن توابع به‌روزرسانی مراکز خوشه‌ها و وزن‌های ویژگی‌ها مطرح شده است. همچنین تحلیل‌های ریاضی جهت اثبات همگرایی الگوریتم نیز ارائه شده است.

به‌علاوه، چون که داده عمق برای نمایش و تغییرات نمایش شی قوی است، تعدادی از متدها در محاسبه ویژگی‌های تصویر به طور مستقیم در تصاویر عمق ظاهر شده‌اند. این متدها هنوز برای نقطه دیدهای مختلف محاسبه نشده است.

بیشتر از یک رویکرد قراردادی برای تخمین اشاره دست به یک مدل سه‌بعدی، نگه‌داشته‌شده در حافظه و تطبیق داده‌شده در یک تصویر اعتماد می‌کند [۱۲]. رویکرد مقایسه مدل ارائه‌شده در برخی پیکربندی‌ها با تصویر اصلی با موفقیت هم در تشخیص اشاره و هم در تخمین ژست به کار می‌رود [۲]. این متدها به طور طبیعی برای موقعیت و چرخش دست متفاوت به شمار می‌روند، ولی اختلاف در چرخش، فضای پارامتر را توسعه می‌دهد و بنابراین پردازش تطبیق پایداری کمتر و زمان مصرفی بیشتری خواهد داشت.

بیشتر کارها بر روی تشخیص یک زیرمجموعه کوچک از اشارات تمرکز می‌کنند. همان‌طور که در حال حاضر ذکر شده، برای بیشتر آن‌ها تفاوت در نقطه دید (یا متناوباً، چرخش مختلف، موقعیت و مقیاس یک دست)، محیط و نمایش شی، زحمت قابل توجهی را نمایش می‌دهد.

در این کار، مشکلات ذکر شده در بالا توسط ویژگی‌های چرخش - موقعیت و مقیاس - ثابت عنوان می‌شود. به‌علاوه، یک جنگل تصادفی چندلایه‌ای فازی^۹ برای طبقه‌بندی به کار می‌گیریم. این به ما اجازه

می‌دهد که به اندازه قابل توجهی زمان آموزش و حافظه مصرفی را برای FMLRF کاهش دهیم.

۱-۲- مشکلات، چالش‌ها و شرح روش پیشنهادی

مشکل زبان علامت و تشخیص اشارات می‌تواند در دو دسته تشخیص اشارات ایستا و پویا تقسیم شود، همانند زبان علامت آمریکایی که شامل اشارات ایستا و پویا می‌باشد. قسمتی از داده، برای نیاز به تشخیص یک علامت ایستا از یک علامت پویا یا یک اشاره دست تصادفی وجود دارد. ابزارهای متفاوتی برای تشخیص اشارات وجود دارد؛ برخی از این ابزارها مبنی بر رویکردهای محدوده‌ای از مدل‌سازی احتمالاتی هستند. متدهای دیگر مبنی بر تکنیک‌های بینایی کامپیوتر و تشخیص الگو، کشف شیء، خوشه‌بندی، و طبقه‌بندی هستند. بیشتر متدهای دیگر مبنی بر تکنیک‌های پردازش تصویری همانند تحلیل و کشف شکل، بافت، رنگ، حرکت و طبقه‌بندی تصویر هستند.

در این کار، ما از الگوریتم طبقه‌بندی جنگل تصادفی به‌عنوان روش پایه استفاده می‌کنیم، اما برای بهبود کارایی و افزایش دقت جنگل تصادفی چندلایه‌ای فازی را پیشنهاد می‌دهیم. در ابتدا، دست را توسط آستانه‌گذاری مقادیر عمق کشف و قطع‌بندی می‌کنیم و یک نقطه انبوه برای پیش‌پردازش اضافی به‌دست می‌آوریم. در این کار از عمق تصویر استفاده می‌کنیم، بنابراین، در ابتدا این ویژگی‌ها را استخراج می‌کنیم.

در ادامه نیز از تکنیک طبقه‌بندی پیشنهادی FMLRF استفاده می‌کنیم. این تکنیک بر مبنای جنگل تصادفی است که یک تکنیک طبقه‌بندی و رگرسیون است، که به‌علت کارایی و سادگی‌اش محبوبیت زیادی دارد. برای بهبود طبقه‌بندی، از انعطاف‌پذیری منطق فازی و مجموعه‌های فازی استفاده کردیم. در الگوریتم پیشنهادی، به‌جای درخت تصمیم از درخت تصمیم فازی استفاده شده است. بنابراین هر درخت در جنگل یک درخت فازی خواهد بود. در حقیقت، این رویکرد، استحکام سیستم‌های طبقه‌بندی چندگانه، قدرت تصادفی بودن برای افزایش تنوع درختان، و انعطاف‌پذیری منطق فازی و مجموعه‌های فازی را برای مدیریت داده‌های ناقص را ترکیب می‌کند.

برای بهبود دقت طبقه‌بندی داده را از پیش خوشه‌بندی می‌کنیم. هدف مقاله، در این حالت تولید خوشه‌هایی است که شامل کلاس‌های بسیار کمتری از کل مجموعه آموزشی هستند. ما استفاده از جنگل تصادفی مبنی بر تکنیک خوشه‌بندی را انتخاب کردیم. در انتها نیز مرحله آموزش و آزمایش انجام می‌شود.

در حقیقت، مقاله یک رویکرد مبتنی بر استفاده از عمق تصویر، برای تشخیص علامت ایستا پیشنهاد می‌کند، که مبنی بر منطق فازی است. در الگوریتم پیشنهادی، مزیت‌های الگوریتم جنگل تصادفی چندلایه‌ای [۱] و درخت تصادفی فازی را ترکیب می‌کنیم.

در بحث تشخیص زبان اشاره مشکلات بسیاری وجود دارد. چالش‌های اساسی مشکل تشخیص اشاره ایستا عبارت‌اند از:

۱- تعداد علامت‌ها (برای مثال ۲۴ اشاره ایستا در زبان اشاره آمریکایی، یا ۲۶ اشاره ایستا در زبان علامت روسیه)

۲- شباهت بین برخی علامت‌ها

۳- تغییر در نمایش یک علامت به‌علت اختلاف در نقطه دید (تغییرات درون موضوع)

۴- تغییر در کارایی یک علامت توسط موضوعات مختلف (تغییرات بین موضوع)

که الگوریتم پیشنهادی در این مقاله بر روی رفع سه مورد اول چالش‌ها تمرکز می‌کند.

مقاله به این شکل سازمان‌دهی شده است: الگوریتم پیشنهادی (جنگل تصادفی چندلایه‌ای فازی) در بخش ۲ توضیح داده شده است. نتایج تجربی در بخش ۳ شرح داده شده است. در بخش ۴ نیز نتیجه‌گیری عنوان شده است.

۲- الگوریتم پیشنهادی

در این بخش الگوریتم طبقه‌بندی جنگل تصادفی فازی را شرح می‌دهیم که از الگوریتم طبقه‌بندی جنگل تصادفی به‌عنوان روش پایه استفاده کرده‌ایم، اما برای بهبود کارایی و افزایش دقت، جنگل تصادفی چندلایه‌ای فازی را پیشنهاد می‌دهیم. ما در الگوریتم خود، برای طبقه‌بندی از جنگل تصادفی چندلایه‌ای استفاده می‌کنیم، جنگل تصادفی از درخت‌های تصمیمی شکل گرفته است که ما در کار خود برای بالا بردن کیفیت این روش از درخت تصمیم فازی استفاده می‌کنیم. در ادامه، داده برای بهبود دقت طبقه‌بندی از پیش خوشه‌بندی می‌شود. خوشه‌بندی طیفی از بردارهای ویژگی ماتریس شباهت برای افزایش مجموعه داده استفاده می‌کند. در اغلب این روش‌ها مقدار ویژه اولویت بردارها را تعیین می‌کند. ولی این نحو انتخاب، تضمینی برای انتخاب بهترین بردارها نمی‌دهد. در آزمایش‌های انجام‌شده خوشه‌بندی طیفی برای تولید هر خوشه به‌علت عدم انتخاب بهترین بردارها، شکست می‌خورد. از آنجایی که هدف در این حالت تولید خوشه‌هایی است که شامل کلاس‌های بسیار کمتری از کل مجموعه آموزشی هستند. به همین دلیل، ما استفاده از جنگل تصادفی مبنی بر تکنیک خوشه‌بندی را انتخاب کردیم. در انتها نیز مراحل آموزش و آزمایش انجام می‌گیرند. شرح هر یک از این مراحل به‌تفصیل در زیر بیان شده است.

۲-۱- استخراج ویژگی

اولین گام در این الگوریتم، استخراج ویژگی‌های لازم می‌باشد. در این کار از عمق تصویر استفاده می‌کنیم، بنابراین، در ابتدا این ویژگی‌ها را استخراج می‌کنیم. سپس توصیف‌گر ESF را که متشکل از مجموعه‌ای از هیستوگرام‌های متصل به هم می‌باشد، محاسبه می‌کنیم. جزئیات عملیات انجام‌شده در این گام در زیر به‌تفصیل بیان شده است.

پیش‌پردازش عمق تصویر: دست توسط آستانه‌گذاری مقادیر عمق کشف و قطع‌بندی می‌شود، و یک نقطه انبوه برای پیش‌پردازش اضافی به‌دست می‌آید [۱۳]. برای نرمال‌سازی عمق تصویر و انتقال آن

۲-۲- جنگل تصادفی چندلایه‌ای فازی (FMLRF)

جنگل تصادفی (RF): یک تکنیک طبقه‌بندی و رگرسیون است که به علت کارایی و سادگی‌اش محبوبیت زیادی دارد جنگل تصادفی از T درخت تصمیم t_i تشکیل شده است هر درخت تصمیم به صورت نرمال یک درخت دودویی است داده در میان هر درخت با شروع از ریشه درخت انتشار داده می‌شود. در هر گره داده با استفاده از یک تست دودویی در شکل $f(x) < \tau$ تقسیم‌بندی می‌شود که f و τ از داده آموزشی تأمین می‌شوند [۱۸]. اگر معادله برقرار باشد داده به شاخه چپ فرستاده می‌شود. در غیر این صورت - به شاخه راست ارسال می‌شود. همه درخت‌ها همان عمق D را دارند.

در هر گره برگ از یک درخت یک پیشگو ذخیره می‌شود. در این حالت این یک توزیع احتمالاتی از برچسب‌های کلاس یا خوشه است که این گره به مرحله آموزشی رسیده است. هر وقت که یک بردار داده x به گره k از درخت m برسد تخمین درخت m برای این بردار توسط پیشگوی ذخیره‌شده در برگ k شکل داده می‌شود. تخمین کل جنگل تصادفی، t برای x معمولاً توسط میانگین تخمین‌های درخت‌های مجزا شکل داده می‌شود:

$$t(x) = \frac{1}{T} \sum_i t_i(x) \quad (4)$$

که T تعداد درخت‌ها در جنگل تصادفی است.

در طی آموزش، همان‌طور که داده آموزشی به پایین درخت انتشار داده می‌شود. در هر گره چندین پیشنهاد برای f و T به صورت تصادفی نمونه‌گیری می‌شود و بهترین پارامتر مطابق با برخی شرایط (δ پیشنهاد مجزا برای هر گره در طی آموزش تولید می‌کنیم) انتخاب می‌شود. به‌علاوه در این کار ما از اطلاعات به‌دست‌آمده از هر شرایط استفاده می‌کنیم. همچون تابع f ، ما برای استفاده از طرح جداسازی دوبعدی انتخاب می‌کنیم، که دو مختصات به صورت تصادفی از بردار ویژگی نمونه‌گیری می‌شوند: $f(x) = x_{i1}w_1 + x_{i2}w_2$. بنابراین، هر گره از یک درخت تصمیم توسط پنج پارامتر در مجموع توصیف می‌شود. نتیجه به کار بردن جنگل تصادفی برای هر درخت در جنگل یک درخت تولیدشده خواهد بود [۱۹].

درخت تصمیم فازی: در الگوریتم پیشنهادی به جای درخت تصمیم از درخت تصمیم فازی استفاده می‌کنیم. بنابراین هر درخت در جنگل یک درخت فازی خواهد بود. الگوریتم درخت تصمیم فازی در شکل ۱ آمده است.

در این کار درخت‌های فازی بر روی الگوریتم پیشنهادی با استفاده از یکی از این دو استراتژی ممکن زیر اعمال شده است:

استراتژی ۱: اطلاعات را از برگ‌های مختلف به‌دست‌آمده در هر درخت برای گرفتن تصمیم درخت‌های انفرادی، ترکیب می‌کنیم و همان متد یا ترکیب دیگری را برای ساختن تصمیم جهانی FMLRF به کار می‌بریم. به‌منظور ترکیب اطلاعات برگ‌های به‌دست‌آمده در هر درخت، تابع $Faggre1_1$ استفاده شده و $Faggre1_2$ برای ترکیب خروجی‌های

به نقطه انبوه سه‌بعدی واقعی، اول عمق تصویر را به یک نقطه انبوه توسط انتقال پرسپکتیو معکوس انتقال می‌دهیم:

$$X_{3D} = \frac{x_{2D} - c_x}{f_x} I(x_{2D}, y_{2D}) \quad (1)$$

$$y_{3D} = \frac{y_{2D} - c_y}{f_y} I(x_{2D}, y_{2D}) \quad (2)$$

$$z_{3D} = I(x_{2D}, y_{2D}) \quad (3)$$

که (c_x, c_y, f_x, f_y) پارامترهای اصلی از یک حسگر عمق هستند. x_{2D} و y_{2D} نیز مختصات عمق تصویر هستند. در حالت حسگر kinect، m از مقادیر پیش‌فرض استفاده می‌کنیم.

توصیف گر ESF. توصیف گر ESF متشکل از یک مجموعه از هیستوگرام‌های متصل می‌باشد. سه تابع D_3 ، A_3 و D_2 که معرف سه هیستوگرام هستند، به شرح زیر توصیف می‌شوند.

✓ **تابع D_2 :** هیستوگرام اولیه توزیع فواصل بین دو نقطه تصادفی در نقطه انبوه را توصیف می‌کند.

✓ **تابع A_3 :** هیستوگرام دوم توزیع گوشه‌های بسته‌شده به‌وسیله دو خط ساخته‌شده توسط سه نقطه نمونه تصادفی را شرح می‌دهد.

✓ **تابع D_3 :** هیستوگرام سوم توزیع محیط‌های بسته‌شده توسط سه نقطه نمونه تصادفی را شرح می‌دهد.

علاوه بر این، برای هر خط اتصال دو نقطه معین شده است. اگر آن روی سطح انبوه پایین سطح، یا در میان سطح باشد. در هر حالت یک هیستوگرام جداگانه برای هر تابع ساخته شده است [۱۴]. برای تابع D_2 یک هیستوگرام شامل نسبت طول در حالت میان‌خط نیز محاسبه شده است. هر هیستوگرام طول ۶۴ دارد بنابراین کل توصیف‌گر، متشکل از ۱۰ هیستوگرام به طول ۶۴۰ است [۱۵]. مقاله [۱۶] نیز در کار خود از هیستوگرام وزن‌دارشده مدل هدف استفاده کرده و با اعمال مغایرت^{۱۰} یک نویز تصادفی در محل هدف، موقعیت ذرات نامزد در فریم بعد را پیشگویی می‌کند.

این توصیف‌گر ویژگی‌های زیر را دارد:

- بیان شده که کل شیء قابل مشاهده است، آن ثابت دورانی و انتقالی است، علاوه بر این، توصیف‌گر همچنین ثابت مقیاس است، چون که قبل از محاسبه هیستوگرام محدوده‌ی مقادیر ممکن از D_2 ، D_3 و A_3 نرمال‌سازی شده است.

- ثابت تحت تبدیل آینه‌ای

- محاسبات نرمال ناپایدار و زمان تحلیل درخواست نمی‌کند.

- گزارش شده است که تفکیک منصفانه‌ای انجام دهد.

دست در وابستگی به اشارات نمایش متفاوتی دارد، بنابراین ما به مسئله طبقه‌بندی علامت همچون مسئله طبقه‌بندی شیء رفتار می‌کنیم. یک توصیف‌گر ESF برای ناحیه دست بخش‌بندی شده محاسبه می‌کنیم [۱۷].

الگوریتم یادگیری درخت تصمیم فازی

۱. شروع با مجموعه نمونه ورودی، وزن نمونه‌ها (در گره ریشه) برابر با ۱ است.
۲. در هر گره N که هنوز بسط یافته نشده است، تعداد نمونه‌هایی از هر کلاس محاسبه می‌شود. نمونه‌ها در قسمتی یا در همه شاخه‌ها توزیع داده شده‌اند. مقدار توزیع شده از هر نمونه به یک شاخه همچون محصول وزن جاری و درجه‌ی عضویتش به گره به‌دست آمده است.
۳. محاسبه محتوای اطلاعات استاندارد.
۴. در هر گره مجموعه‌ای از صفات باقی‌مانده برای جدا کردن گره جستجو می‌شود. با هر معیار مجموعه صفات منتخب برای جدا کردن گره انتخاب می‌شود. محاسبه محتوای اطلاعات استاندارد به هر گره فرزند به‌دست‌آمده از هر صفت منتخب.
- انتخاب صفت منتخب از بیشینه آن اطلاعات به‌دست‌آمده.
۵. تقسیم N در زیر-گره‌هایی مطابق با خروجی‌های ممکن صفت انتخاب شده در گام قبلی.
۶. تکرار گام‌های ۲-۵ برای توقف معیار در همه گره‌ها راضی‌کننده است.

شکل ۱: الگوریتم یادگیری درخت تصمیم فازی

الگوریتم طبقه‌بندی با استفاده از استراتژی ۱

- شروع
- DecisionOfTrees (ورودی: e و FMLRF؛ خروجی: T_F)
 DecisionOfForest (ورودی: e و FMLRF؛ خروجی: T_F)
- پایان
- DecisionOfTrees (ورودی: e و FMLRF؛ خروجی: T_F)
- شروع
- ۱- نمونه e را از بین هر درخت برای به‌دست آوردن ماتریس L_F اجرا می‌کنیم.
- ۲- برای هر درخت t {برای هر کلاس i } $T_F_{t,i} = \{ Faggore1_1(t, i, L_F) \}$.
- پایان
- DecisionOfForest (ورودی: T_F ؛ خروجی: c)
- شروع
- ۱- برای هر کلاس i $D_F_i = Faggore1_2(i, T_F)$.
- ۲- کلاس c را برای نمونه e به الگوریتم پایه الحاق می‌کنیم،
 همچون $c = \arg \max_{i=1, \dots, l} \{ D_F_i \}$.
- پایان

شکل ۲: الگوریتم طبقه‌بندی با استفاده از استراتژی ۱

الگوریتم طبقه‌بندی با استفاده از استراتژی ۲

- شروع
- ۱- نمونه e را از بین هر درخت برای به‌دست آوردن ماتریس L_F اجرا می‌کنیم.
- ۲- برای هر کلاس i $D_i = Faggre2(i, L_F)$.
- ۳- کلاس c را برای نمونه e به الگوریتم پایه الحاق می‌کنیم،
 همچون $c = \arg \max_{i=1, \dots, l} \{ D_F_i \}$.
- پایان

شکل ۳: الگوریتم طبقه‌بندی با استفاده از استراتژی ۲

به‌دست‌آمده با $Faggre1_1$ به کار می‌رود. در شکل ۲ شرح این الگوریتم آمده است.

استراتژی ۲: اطلاعات را از تمام برگ‌های به‌دست‌آمده از همه درخت‌ها برای ساخت تصمیم جهانی FMLRF ترکیب می‌کنیم. ما از تابع $Faggre2$ برای ترکیب اطلاعات ساخته‌شده توسط تمام برگ‌ها استفاده می‌کنیم. در شکل ۳ شرح این الگوریتم آمده است. در ادامه، توابع این دو استراتژی را تعریف می‌کنیم که در گام ۲ الگوریتم‌های تعریف‌شده در شکل ۲ و شکل ۳ برای این‌اند که هر برگ به‌دست‌آمده به یک رأی ساده در کلاس اکثریت الحاق شود، به کار برده می‌شوند [۲۰].

تابع $Faggre1_1$ همچون زیر تعریف می‌شود:

$$Faggre1_1(t, i, L_F) = \begin{cases} 1 & \text{if } i = \arg \max_{j=1, \dots, l} \{ \sum_{n=1}^{N_t} L_F_{t,n,j} \} \\ 0 & \text{در غیر این صورت} \end{cases} \quad (5)$$

که در آن L_F یک ماتریس است که هر عنصر ماتریس یک بردار با اندازه I می‌باشد که شامل پشتیبان هر کلاس فراهم‌شده توسط هر برگ فعال در هر درخت t است [۲۰].

تابع $Faggre1_2$ شامل وزن تعریف‌شده توسط تابع عضویت فازی $\mu_{pond}(x)$ (در ادامه تعریف شده است) برای هر درخت است:

$$Faggre1_2(t, T_F) = \sum_{t=1}^T \mu_{pond} \left(\frac{errors_{(OOB_t)}}{size_{(OOB_t)}} \right) \cdot T_{F,t,i} \quad (6)$$

T_F یک ماتریس است که شامل اطمینان الحاق شده توسط هر درخت t برای هر کلاس i است. عناصر ماتریس از پشتیبان هر کلاس در برگ‌ها زمانی که برای برخی متدهای ترکیبی به کار می‌روند، به‌دست آمده است [۲۰].

تابع عضویت فازی $\mu_{pond}(x)$ به‌صورت زیر تعریف می‌شود [۲۰]:

$$\mu_{pond}(x) = \begin{cases} 1 & 0 \leq x \leq (pmin + marg) \\ \frac{(pmax + marg) - x}{(pmax - pmin)} & (pmin + marg) \leq x \leq (pmax + marg) \\ 0 & (pmax + marg) \leq x \end{cases} \quad (7)$$

- که $pmax$ حداکثر نرخ خطا در درخت‌های گروه FMLRF
 $(pmax = \max_{t=1, \dots, T} \left\{ \frac{errors_{(OOB_t)}}{size_{(OOB_t)}} \right\})$. نرخ خطاها در یک درخت t همچون $\frac{errors_{(OOB_t)}}{size_{(OOB_t)}}$ به دست می‌آید که $errors_{(OOB_t)}$ تعداد خطاهای طبقه‌بندی درخت t (با استفاده از مجموعه داده (OOB_t) به‌عنوان مجموعه آزمایشی) است و کاردینال مجموعه داده می‌باشد.

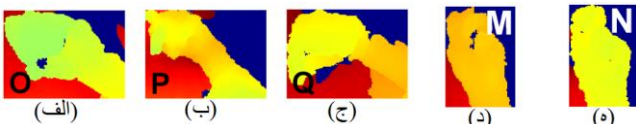
- $pmin$ حداقل نرخ خطا در درخت‌های گروه FMLRF می‌باشد.
 $marg = \frac{pmax - pmin}{4}$

حاضر شامل اطلاعات کافی برای پیدا کردن علامت‌های مشابه می‌باشد در عین حال شامل نویزهای بسیار کمتری است [۲۱]. L به عنوان پارامتر سطح تراکم هیستوگرام تعریف می‌شود، $L=0$ هیستوگرام کامل 64-bin تعریف می‌شود، $L=1$ اندازه هیستوگرام را به 32-bin کاهش می‌دهد، $L=2$ ، 16-bin تولید می‌کند و غیره. بنابراین الگوریتم خوشه‌بندی سراسری گام‌های زیر را دارد:

- ۱- محاسبه ویژگی‌ها در یک سطح تراکم داده شده L
- ۲- ساخت دو خوشه مصنوعی یکی شامل همه داده‌های آموزشی، و دیگری شامل داده نمونه تصادفی
- ۳- آموزش یک جنگل برای تفاوت قائل شدن بین دو کلاس و محاسبه p برای این جنگل
- ۴- به کار بردن الگوریتم خوشه‌بندی سراسری برای یافتن خوشه‌ها در داده

دیده می‌شود که خوشه‌بندی در واقع نتایج معناداری را تولید می‌کند. علامت‌های مشابه با یکدیگر در یک خوشه گروه‌بندی می‌شوند و برای هر خوشه تعداد داخل کلاس‌ها کاهش داده می‌شود. برای مثال برای داده‌های واقعی علامت‌های 'o', 'p', 'q' با هم در یک خوشه گروه‌بندی می‌شوند و علامت 'n', 'm' در خوشه دیگری گروه‌بندی می‌شوند. در شکل ۴ نشان داده شده است که تصاویر عمق این سه علامت کاملاً مشابه هستند بنابراین به عنوان توصیف‌گر استفاده می‌شوند [۲۲].

خوشه‌بندی داده مصنوعی به طور کامل کلاس‌ها را جدا می‌کند خوشه‌ها را تولید می‌کند فقط شامل یک کلاس در حالی که داده واقعی شامل نویز بیشتری است و بنابراین خوشه‌ها شامل بخش‌های کوچکی از کلاس‌های متفاوت می‌باشند.



شکل ۴: تصاویر عمق از علامت‌ها، شکل گرفتن دو خوشه مختلف در خوشه‌بندی نمایش داده شده در شکل ۷، تصاویر الف، ب، ج از یک خوشه هستند د و ه از خوشه دیگری می‌باشند.

۳- نتایج تجربی

این الگوریتم در موقعیت‌های مختلف بر روی داده مصنوعی پایگاه داده قابل دسترس عمومی شامل اشاره‌های ASL و نیز داده عمق جمع شده از دوربین عمق اینتل [۴] ارزیابی شده است. مقاله نتایج متد را با متدهای شناخته شده و در شرایط خطای طبقه‌بندی (تعداد نمونه‌های طبقه‌بندی شده نادرست به اندازه مجموعه آزمایشی تقسیم شده است) زمان آموزش و نیز حافظه مصرفی مقایسه می‌کند.

مقاله از یک پیاده‌سازی جنگل تصادفی در محیط برنامه‌نویسی مطلب^{۱۱} و نیز به صورت غیرموازی استفاده کرده است. زمان‌های محاسباتی برای یک CPU دو هسته‌ای به کار گرفته شده در محاسبات فراهم شده‌اند.

D_F در الگوریتم‌های نمایش داده شده در شکل‌های ۲ و ۳، یک بردار است که به اطمینان الحاق شده توسط الگوریتم پایه به هر کلاس i اشاره می‌کند. عناصر ماتریس از پشتیبان برای هر کلاس در برگ‌ها، زمانی که برای برخی متدهای ترکیبی به کار می‌روند، به دست آمده است [۲۰].

خوشه‌بندی داده: در سال‌های اخیر کاربردهای زیادی در حوزه ساخت جنگل تصادفی چندلایه‌ای ارائه شده است. آزمایش‌های نشان می‌دهد که جنگل تصادفی در تشخیص بین تعداد کوچک‌تری از کلاس‌ها بهتر اجرا شده است، بنابراین داده برای بهبود دقت طبقه‌بندی از پیش خوشه‌بندی شده است هدف اصلی در این حالت تولید خوشه‌هایی است که شامل کلاس‌های بسیار کمتری از کل مجموعه آموزشی هستند. مقاله استفاده از جنگل تصادفی مبنی بر تکنیک خوشه‌بندی را انتخاب کرده است چون که در آزمایش‌های انجام شده خوشه‌بندی بینایی برای تولید هر خوشه شکست می‌خورد.

خوشه‌بندی خودش به معنی پیدا کردن ساختارهایی در داده می‌باشد که معادل تشخیص بین داده، داده شده و داده تصادفی با همان رنج مقدار است. بنابراین یک مجموعه داده آموزشی مصنوعی شامل کلیه داده‌هایی که در یک کلاس خوشه‌بندی شده‌اند، می‌سازیم، و به صورت تصادفی داده را در کلاس دیگر تولید می‌کنیم. این داده توسط مقادیر نمونه‌گیری تصادفی از هر توزیع حاشیه‌ای از مختصات بردار داده آموزشی تولید شده است. یکی از ویژگی‌های جنگل تصادفی این است که در برخی حالت‌هایی که نمونه‌ها مشابه هستند به بخش پایین همان گره در طی آموزش تمایل پیدا می‌کند. این می‌تواند توسط ماتریس مجاورتی به نام p از اندازه $N * N$ توصیف شود که N اندازه مجموعه آموزشی است:

$$P_{ij} = P_{ji} = \frac{| \{m | x_i, x_j \in L_k^m\} |}{T} \quad (8)$$

در اینجا L_k^m مجموعه‌ای از همه نمونه‌های x است که انتهای گره k برای درخت تصمیم فازی m است. ماتریس مجاورت p می‌تواند همچون یک اندازه‌گیری شباهت ملاحظه شود و بنابراین می‌تواند همچون یک ورودی برای یک الگوریتم خوشه‌بندی استفاده شود. ما ماتریس مجاورت را به یک معیار فاصله $\sqrt{1 - P_{ij}}$ انتقال می‌دهیم. می‌توان دید که این متریک غیر اقلیدسی است و بنابراین یک الگوریتم خوشه‌بندی مناسب نیاز است.

استفاده از خوشه‌بندی سلسله‌مراتبی [۲۰] را برای خوشه‌بندی داده مبنی بر ماتریس‌های شباهت تولید شده انتخاب می‌کنیم. برای خوشه‌بندی سلسله‌مراتبی از پیوند کامل برای تراکم داده در یک درخت خوشه استفاده می‌کنیم. همچنین از bin‌های کمتری در توصیف‌گر ESF برای خوشه‌بندی استفاده می‌کنیم زیرا هم سرعت آموزش را افزایش می‌دهد و هم نویز را در توصیف‌گرهای محاسبه شده کاهش می‌دهد. مطابق با آزمایش‌های انجام شده بیشتر هیستوگرام‌های Coavse در حال

بعد از این که داده خوشه‌بندی شد، ما اولین سطح جنگل تصادفی را در بردارهای ویژگی متراکم آموزش می‌دهیم. این جنگل یک برچسب خوشه به هر بردار به‌دست‌آمده الحاق می‌کند و سپس برای هر یک از خوشه‌های مشتق‌شده، یک جنگل تصادفی جداگانه در بردارهای ویژگی کامل برای تشخیص بین علامت‌های مشابه آموزش می‌دهیم. عمق انباشته از FMLRF را به‌عنوان مجموع عمق‌های جنگل تصادفی در اولین و دومین لایه تعریف می‌کنیم [۲۴].

همچنین تعداد درخت‌ها را برای هر جنگل در اولین و دومین لایه جایگذاری می‌کنیم. درخت‌ها با استفاده از منطق فازی ساخته می‌شوند. یک گام مهم نرمال‌سازی دوباره بردارهای ویژگی برای هر جنگل در سطح دوم است. نرمال‌سازی دوباره ضروری است، زیرا بردارهای مشابه باهم خوشه‌بندی می‌شوند و بنابراین مغایرت در میان هر خوشه نسبت به مغایرت کل نمونه خیلی کم‌تر است. مغایرت توسط کلاس‌هایی مستقل می‌شود که در این خوشه‌ها شامل شده‌اند و گذشتن از نرمال‌سازی دوباره می‌تواند کارایی طبقه‌بندی را کاهش دهد.

آزمایش: در طی مرحله آزمایش هر نمونه در میان جنگل اولین مرحله از هر کلاس که نمونه متعلق به آن است گذرانده می‌شود. جنگل مرحله اول برچسب خوشه یک نمونه را مشخص می‌کند، سپس نمونه برای جنگل متعلق به سطح دوم برای تعیین برچسب کلاسش گذرانده می‌شود.

۳-۱- داده مصنوعی

در یک محیط طبیعی، موقعیت دست یک شخص اجراکننده اشارات مختلف نسبت به یک دوربین می‌تواند به اندازه بسیار زیادی تغییر کند. بنابراین با استفاده از ویژگی‌های قوی برای برخی تغییرات کارایی افزایش پیدا می‌کند.

برای نشان دادن قدرت ویژگی‌های ESF از پایگاه‌داده تولیدشده توسط [۴] استفاده می‌کنیم. [۴] داده مصنوعی را با استفاده از مدل دست از [۲۵] برای ساختن تصاویر عمق از ۲۶ حروف ایستای ASL تولید کرده است. برای مجموعه آموزشی [۴] ۵۰۰ تصویر برای هر حرف تولید کرده است قابل ذکر است که در تمام حالات، دست، موازی با دوربین است. برای مجموعه آزمایشی [۴] چرخش‌هایی پیرامون سه محور اصلی برای ارزیابی خطای طبقه‌بندی در مقابل زاویه چرخش مدل دست به کار برده است.

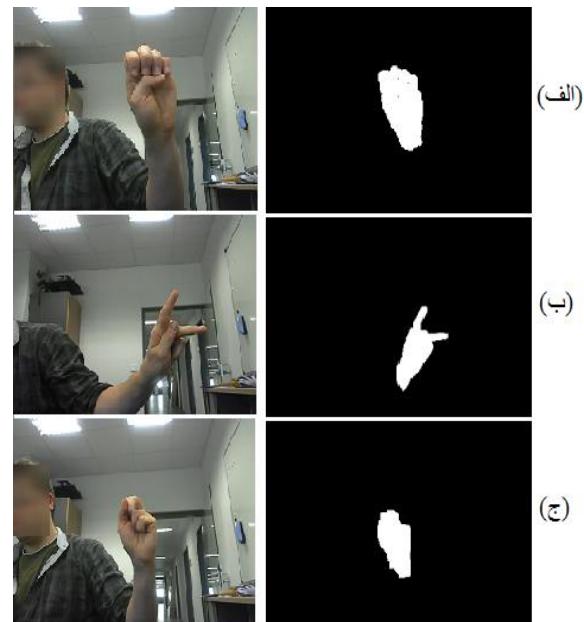
[۴] چرخش پیرامون محور X و Y و پیرامون محور Z دوربین را جدا کرده است. همان‌طور که در شکل ۷ نشان داده شده است حرف هیچ تأثیری در کل کارایی ندارد چون که شکل توده نقطه خودش به‌دقت به همان اندازه گویا است. چرخش قوی‌تر پیرامون محور X و Y باعث کاهش قوی‌تر کارایی می‌شود به‌دلیل اینکه نمایش توده نقطه‌ای را تغییر می‌دهد (شکل ۶ را ببینید). این نتیجه نشان می‌دهد که برای توانایی تشخیص اشارات قوی از یک نقطه دید مختلف در عموم داده‌های آموزشی مشابه هنوز مورد نیاز است. به‌رحال چرخش یک دست در

شکل ۵ نمونه‌ای از تصاویر پایگاه‌داده [۴] را نمایش می‌دهد. در این پایگاه‌داده تمامی حروف الفبای آمریکایی دسته‌بندی شده‌اند و در هر دسته تعداد زیادی از تصاویر گرفته‌شده از حروف الفبا وجود دارد. هر یک از این تصاویر از زوایای گوناگون گرفته شده‌اند.

آموزش: جنگل تصادفی چندلایه‌ای یک جنگل تصادفی است که

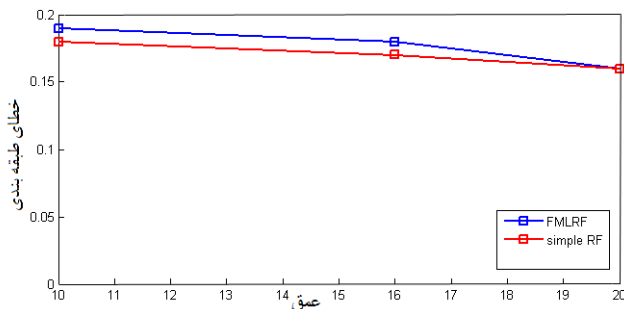
شامل دو لایه می‌باشد برای آموزش FMLRF اول خوشه‌بندی داده اجرا می‌شود. در آغاز تعداد خوشه‌ها $k=20$ جایگذاری می‌شود و خوشه‌ها از داده واقعی که نویزی است و شامل بخش‌های کوچکی از داده بسیاری کلاس‌ها هستند ساخته می‌شوند. هدف ساختن خوشه‌هایی شامل تعداد کمی از کلاس‌ها می‌باشد بنابراین خوشه‌ها توسط موارد زیر هرس می‌شوند:

- ادغام خوشه‌های کوچک با خوشه‌های بزرگ‌تر؛ کوچک‌ترین اندازه خوشه را به ۵۰٪ از اندازه یک کلاس در داده آموزشی جایگذاری می‌کنیم؛ اگر یک خوشه کوچک‌تر از این اندازه هست، برای هر عنصر از این خوشه یک خوشه شامل بیشترین عناصر همان کلاس پیدا می‌کنیم و این عنصر را با خوشه جدید ادغام می‌کنیم، به‌علت این آستانه تعداد خوشه‌ها در حقیقت به‌صورت اتوماتیک برای هر مجموعه‌داده معین شده است [۲۳]. می‌توانیم مشاهده کنیم که در حالت مسئله طبقه‌بندی علامت‌های ASL، ۱۲-۱۵ خوشه بزرگ شکل داده شده است.



شکل ۵: در هر ردیف شکل‌های سمت چپ، تصاویر اصلی را نشان می‌دهد و تصاویر سمت راست، تصاویر به‌دست‌آمده از اجرای الگوریتم قطعه‌بندی است. در ردیف (الف) زبان اشاره حرف (e)، ردیف (ب) زبان اشاره حرف (k) و ردیف (ج) زبان اشاره حرف (s) می‌باشد.

- در خوشه‌های بزرگ کلاس‌هایی که تعداد نماینده آن‌ها کم است را پاک می‌کنیم.



شکل ۹: خطای طبقه بندی نسبت به عمق جنگل‌ها

۳-۲- داده واقعی

برای این که دقت و کارایی الگوریتم خود را آزمایش کنیم علاوه بر انجام آزمایش‌های بر روی داده مصنوعی بر روی داده واقعی نیز آزمایش‌هایی را انجام می‌دهیم.

۳-۲-۱- پایگاه داده عمومی

پایگاه داده عمومی [۲۱]، مجموعه داده‌ای شامل ۲۴ اشاره ایستا از زبان اشاره آمریکایی (ASL) است، که توسط پنج شخص اجرا شده است. تغییرپذیری بالایی در اجرای علامت‌ها توسط انسان و در موقعیت وابسته به دوربین وجود دارد اگرچه تغییرپذیری در ژست برای یک شخص نسبتاً کم است.

برای ارزیابی متد پیشنهادی، از دو سناریو به شرح زیر استفاده می‌کنیم:

✓ سناریو ۰-۰-۱ جنگل‌ها را به چهار نفر آموزش داده و سپس کارایی آن‌ها را ارزیابی می‌کنیم.

✓ سناریو h-h در این حالت نیمی از داده را برای آزمایش و نیمی را برای آموزش به کار می‌بریم.

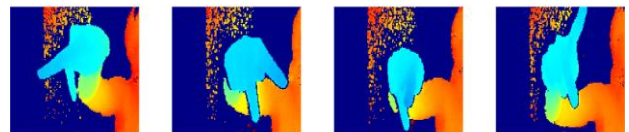
در جدول ۱ نتایج آزمایش ۰-۰-۱ گزارش شده است. می‌توان دید خطای آموزشی در این آزمایش‌ها تقریباً ده درصد کاهش داده شده است. به علاوه در آزمایش h-h توانایی کاهش خطا به پانزده درصد را نیز داریم. علاوه بر این در جدول ۱ کارایی الگوریتم پیشنهادی ما با نتایج ۲ مقاله دیگر [۲۱] و [۴] مقایسه شده است. مقاله [۲۱] الگوریتم GF+RF را معرفی کرده است و همچنین مقاله [۴] نیز الگوریتم MLRF را پیشنهاد داده است که در جدول ۱ بررسی کردیم. همان‌طور که از نتایج مشهود است روش پیشنهادی مقاله در مقایسه با دو روش دیگر زمان کم‌تر، پیچیدگی الگوریتم کم‌تر و همچنین حافظه مصرفی کم‌تری را دارا می‌باشد.

برای تصدیق استفاده از FMLRF، کارایی آن را در برابر کارایی یک جنگل تصادفی استاندارد مقایسه می‌کنیم. عمق جنگل تصادفی را تغییر داده و خطای طبقه بندی RF ساده را با FMLRF مقایسه کرده و زمان آموزشی را نیز اندازه‌گیری می‌کنیم (شکل ۱۰ و شکل ۱۱). به دلیل اینکه اندازه کل پایگاه داده در حدود ۶۵۰۰۰ تصویر است FMLRF در زمان

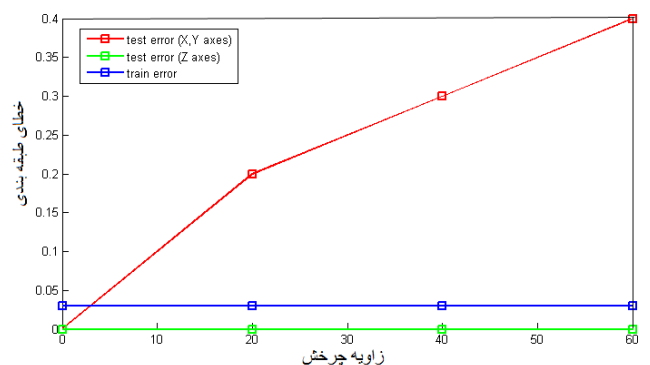
محدودیت‌های ۱۰ درجه کارایی طبقه بندی کاهش معنی داری پیدا نکرده است [۲۶-۲۸].

در این کار برای تصدیق استفاده از RF چندلایه‌ای فازی زمان آموزش یک RF قراردادی را در مقابل FMLRF با همان عمق انباشته ارزیابی کرده‌ایم. می‌توان دید آموزش RF چندلایه‌ای فازی در حدود ۵ مرتبه سریع‌تر از یک RF ساده است (شکل ۸).

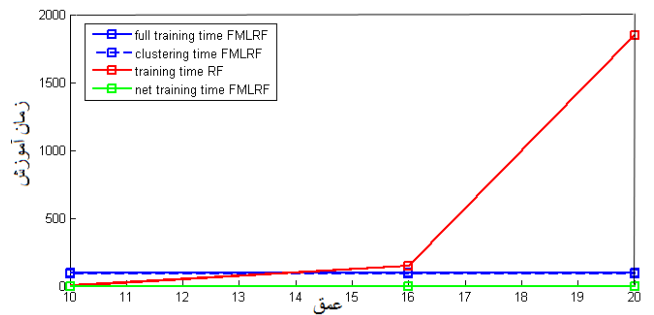
یادداشت اینکه خطای طبقه بندی برای هر دو متد همان (شکل ۹) شروع از عمق ۲۰ D= است. خطای طبقه بندی برای FMLRF در عمق کمتر کمی بالاتر است. آزمایش‌های انجام شده نشان می‌دهد که داشتن عمق کمتر از ۱۰ اولین سطح جنگل را برای طبقه بندی داده در خوشه صحیح ضعیف می‌سازد.



شکل ۶: تفاوت در نمایش علامت 'P' نسبت به دیدهای گوناگون (پایگاه داده [۷])



شکل ۷: خطای آموزش و آزمایش نسبت به محدودیت‌های زاویه چرخش نمونه آزمایشی جنگل استفاده شده تعداد T=۱۰ درخت با عمق D=۲۰ دارد.



شکل ۸: زمان‌های مورد نیاز برای یک RF ساده و برای RF چندلایه‌ای فازی با T=۱۰ نسبت به عمق جنگل

نمایش داده شده است سرعت بالا و محدوده نزدیک دوربین را در حوزه تشخیص اشارات مورد استفاده می‌سازد [۲۹، ۳۰].

مجموعه داده شامل سه شخص در حال اجرای ۲۴ علامت از زبان اشاره ASL می‌باشد. برای هر حرف مقاله [۴، ۳۱] در حدود ۲۵۰ داده فریم جمع‌آوری کرده است. هدف از این مجموعه داده محاسبه تغییر طبیعی در ژست دست نسبت به دوربین است و بنابراین شرکا برای چرخش دست نسبت به دوربین توسط ۳۰ درجه در همه مسیرها برای ساختن تغییرپذیری بالاتر پرسیده می‌شود (برای یک مثال شکل ۸ را ببینید).

چون که داده وضوح^{۱۴} کمتری (۳۲۰x۲۴۰) در مقابل ۶۴۰x۴۸۰ پیکسل برای Kinect دارد انتظار برخی کاهش‌های کارایی را در تشخیص داریم. در حقیقت نتایج (جدول ۲) برخی کاهش‌های کارایی را نشان می‌دهد که می‌توان توسط رزولوشن پایین دوربین و مغایرت بالا در ژست دست توجیه کرد.

جدول ۱: مقایسه کارایی روش GF+RF و روش MLRF با الگوریتم

| پیشنهادی (در تمام جنگل‌ها، T=۱۰) | | | | | |
|----------------------------------|-------|-------|----------|------------------|------------|
| روش | h-h | l-o-o | زمان (s) | پیچیدگی الگوریتم | حافظه (Mb) |
| GF+RF | ٪۳۱ | ٪۵۱ | ۳۲۷۰ | $O(n^3)$ | ۴۲۱ |
| RF(D=۲۰) | ٪۱۵ | ٪۴۹/۱ | ۴۴۸۵ | $O(n^3)$ | ۵۶۰ |
| FMLRF(D=۲۰) | ٪۱۹/۳ | ٪۴۸/۳ | ۳۶۱ | $O(n \log n)$ | ۵/۰۹ |
| FMLRF(D=۳۰) | ٪۱۱/۳ | ٪۴۰ | ۹۸۹ | $O(n \log n)$ | ۱۷۹ |
| MLRF(D=۲۰) | ٪۲۳/۴ | ٪۵۰/۱ | ۵۲۲ | $O(n^2)$ | ۶/۰۵ |
| MLRF(D=۳۰) | ٪۱۳ | ٪۴۳ | ۱۱۳۲/۳ | $O(n^2)$ | ۱۹۲/۵ |

جدول ۲: ارزیابی کارایی FMLRF برای یک شخص در داده [۲۱] و در داده به دست آمده از دوربین اشارات اینتل [۴].

| پایگاه داده [۲۱] | داده دوربین اینتل [۴] | |
|------------------|-----------------------|---------------|
| ٪۵/۵ | ٪۱۹ | خطای D=۲۰ |
| ۲۱۰ | ۶۷/۰۵ | زمان (s) D=۲۰ |
| ٪۲/۲۳ | ٪۱۴/۶ | خطای D=۳۰ |
| ۷۴۶ | ۴۵۱ | زمان (s) D=۳۰ |

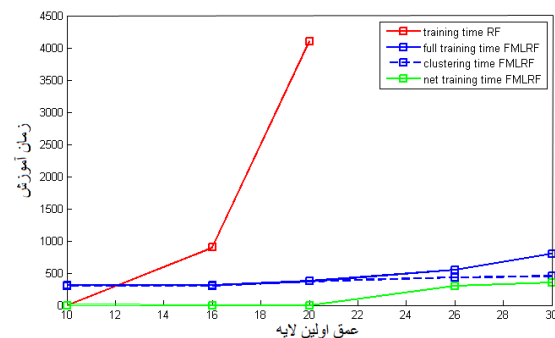
۳-۳- تأثیر داده‌های نویزی در مجموعه داده‌ها

مجموعه داده‌های همراه با نویز به روش زیر به دست می‌آیند: به احتمال ۱۰٪ مقدار مجموعه داده را توسط جایگذاری آن با یک مقدار انتخاب شده یکنواخت در حالت تصادفی از مقادیر ممکن دیگر تغییر می‌دهیم. این فقط برای مجموعه داده آموزشی می‌باشد افزایش درصد خطای میانگین طبقه‌بندی که در جدول ۳ نشان داده شده است به شکل:

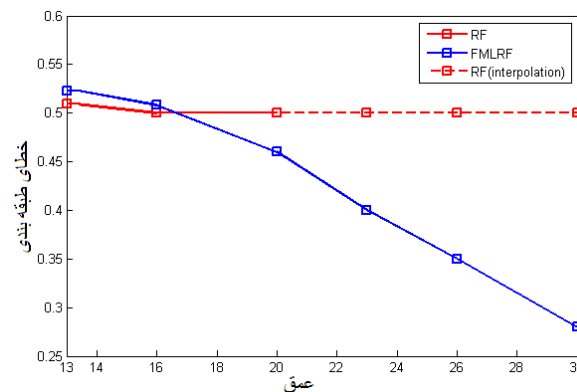
$$\% \text{increase error} = 100 \cdot \frac{CE(\text{noise}) - CE(\text{original})}{CE(\text{original})} \quad (9)$$

آموزشی اکثراً به پیروزی دست پیدا می‌کند که در حدود ۱۰ برابر سریع‌تر از RF ساده برای عمق D=۲۰ است.

در [۱۷] کارایی طبقه‌بندی شکل برای پایگاه داده [۲۱] ارائه شده است در آزمایش l-o-o و h-h دقت و کارایی این متد و متد ارائه شده در این مقاله را مقایسه و نتیجه را در جدول ۱ ثبت کردیم. به هر حال همان‌طور که در شکل ۱۰ مشاهده می‌کنید زمان آموزشی گزارش شده در حدود ۴۰۰۰ ثانیه بر درخت در یک PC ۲ هسته‌ای است. در متد پیشنهادی، زمان آموزشی برای یک نسخه مطلب الگوریتم FMLRF کم‌تر از یک‌هزارم ثانیه در یک هسته‌ای است. در آزمایش‌های یک‌نفره متد پیشنهادی دقت ۹۷٫۶٪ را نشان می‌دهد (جدول ۲ را ببینید). بنابراین در آزمایش‌هایی که هدف یک نفر است، متد ارائه شده کارایی مشابهی را دارد؛ در حالی که زمان‌های آموزش اجازه calibrate دوباره سیستم را برای اضافه شدن یک شخص جدید، حتی در سرعت‌های خیلی زیاد می‌دهد.



شکل ۱۰: زمان آموزش نسبت به عمق جنگل‌ها



شکل ۱۱: خطای طبقه‌بندی نسبت به عمق جنگل‌ها.

۳-۲-۲- ارزیابی در پایگاه داده جدید

برای نشان دادن قابلیت اجرای متد تشخیص یا حسگرهای دیگر از یک پایگاه داده که با استفاده از دوربین اشاره اینتل^{۱۳} جمع‌آوری شده است استفاده می‌کنیم. اطلاعات وسیله که در [۴] شرح داده شده است؛ عبارت است از: یک دوربین ارزان قیمت time-of-flight با محدوده ۱۰ cm تا ۱ حد اکثر نرخ فریم ۵۰ fps و وضوح ۳۲۰x۲۴۰ پیکسل (برای داده عمق)

مراجع

- [1] W. S. Lin, Y. Wu, W. C. Hung and C. Y. Tang, "Real time hand gesture recognition using sift", *Advances in Intelligent Systems and Applications*, vol. 2, Springer, Smart Innovation, Systems and Technologies, vol. 21, pp. 235-246, 2013.
- [2] Z. Ma, E. Wu, "Real-time and robust hand tracking with a single depth camera", *The Visual Computer*, Springer, vol. 30, Issue 10, pp. 1133-1144, October 2014.
- [3] S. Mutlu, T. Hu and O. Lanz, "Learning the Scene Illumination for Color-Based People Tracking in Dynamic Environment", *Image Analysis and Processing - ICIAP 2013*, Springer, vol. 8157, pp. 683-692, 2013.
- [4] A. Kuznetsova, L. Leal-Taixe, and B. Rosenhahn, "Real-time sign language recognition using a consumer depth camera", *Computer Vision Workshops (ICCVW), IEEE International Conference on*, pp. 83-90, 2013.
- [5] A. State, F. Coleca, E. Barth and T. Martinetz, "Estimating 3D Hand Pose from a Cluttered Image", *Advances in Self-Organizing Maps*, Springer, vol. 198, pp. 115-124, 2013.
- [6] W. S. Lin, Y. L. Wu, W. C. Hung and C. Y. Tang, "Hand Gesture Recognition using SIFT", *Advances in Intelligent Systems and Applications*, Springer, vol. 21, pp. 235-246, 2013.
- [7] M. Nachamai, "Alphabet recognition of american sign language: a hand gesture recognition approach using sift algorithm", *International Journal of Artificial Intelligence & Applications*, vol. 4 Issue 1, p105, January 2013.
- [8] A. Kuznetsova and B. Rosenhahn. "Hand pose estimation from a single rgb-d image", *In Proceedings of the 9th International Symposium on Visual Computing, ISVC'13*, Springer, Verlag, 2013.
- [9] معصومه محسنی، مهدی ازوجی و رضا قادری، «قطعه‌بندی تصویر مبتنی بر برش نرمالیزه گراف از دیدگاه میزان اطلاعات جداکننده»، *مجله مهندسی برق دانشگاه تبریز*، دوره ۴۶، شماره ۱، صفحه ۳۱۰-۳۰۳، بهار ۱۳۹۵.
- [۱۰] رضا خدایی، محمدعلی بالافر و سیدناصر رضوی، «اثربخشی بسط پرس‌وجو مبتنی بر خوشه‌بندی اسناد شبه‌بازخورد با الگوریتم K_NN»، *مجله مهندسی برق دانشگاه تبریز*، دوره ۴۶، شماره ۱، صفحه ۱۵۱-۱۴۳، بهار ۱۳۹۵.
- [۱۱] سمیرا رفیعی و پرهام مرادی، «بهبود عملکرد الگوریتم خوشه‌بندی فازی سی‌مینز با وزن‌دهی اتوماتیک و محلی ویژگی‌ها»، *مجله مهندسی برق دانشگاه تبریز*، دوره ۴۶، شماره ۲، صفحه ۷۵-۸۶، تابستان ۱۳۹۵.
- [12] J. Wang, J.Y. Peng, X. Y. Feng, L. Q. Li and D. J. Li, "An Improved Camshift-Based Particle Filter Algorithm for Face Tracking", *Intelligent Science and Intelligent Data Engineering*, Springer, vol. 7202, pp. 278-285, 2012.
- [13] M. Schroder, C. Elbrechter, J. Maycock, and R. Haschke, et al., "Real-Time Hand Tracking with a Color Glove for the Actuation of Anthropomorphic Robot Hands", *Humanoid Robots (Humanoids)*, 2012 12th IEEE-RAS International Conference on, pp. 262 - 269, Nov. 29 2012-Dec. 1 2012.
- [14] Z. H. Khan, I. Y. H. Gu, A. G. Backhouse, "A Robust Particle Filter-Based Method for Tracking Single Visual Object Through Complex Scenes Using Dynamical Object Shape and Appearance Similarity", *Journal of Signal*

محاسبه می‌شود که CE(noise) خطای طبقه‌بندی برای مجموعه داده با نویز است و CE(original) خطای طبقه‌بندی برای داده اصلی است. نتایج در جدول ۳ نمایان است با استفاده از آزمایش Wilcoxon مقایسه نتایج [۸]، [۳۱] و FMLRF اختلافات قابل توجهی در ۲/۳/۹۷ به دست می‌آوریم. با این نتایج FMLRF زمانی که نویز را در صفت کلاس با یک افزایش خطای کوچک‌تر از مقدار [۸] تولید می‌کنیم رفتار خوبی دارد.

جدول ۳: مقایسه الگوریتم پیشنهادی [۸] و الگوریتم پیشنهادی ما با

| داده نویزی | | مجموعه داده |
|--------------------|------------------|-----------------------|
| خطای افزایشی FMLRF | خطای افزایشی [۸] | |
| ۳۷/۴ | ۵۴/۸۴ | Wisconsin breast C. |
| ۷۱/۵ | ۱۶/۹۸ | BUPA liver disorders |
| ۳۷/۰ | ۲۳/۶- | Contraceptive method |
| ۳۱/۲- | ۰.۸/۲- | Statlog heart disease |
| ۹۸/۳ | ۷۴/۵ | Pima Indian diabetes |
| ۲۹/۷۹ | ۵۶/۱۱۵ | Image segmentation |
| ۰.۲/۰ | ۰.۷/۳ | Attitude smoking |
| ۲/۱۵ | ۰.۸/۲۳ | Thyroid disease |
| ۰.۲/۱ | ۷۶/۰- | Vehicle |

۴- نتیجه‌گیری

زبان اشاره برای ارتباط با ناشنویان تهیه شده و در هر زبانی حروف الفبا و کلمات آن زبان با اشاره طراحی شده است. در این کار زبان اشاره آمریکایی ایستا را محوریت قرار دادیم. برای انجام این کار الگوریتم پیشنهادی تحت عنوان الگوریتم جنگل تصادفی چندلایه‌ای فازی را ارائه کردیم. در ابتدا از ویژگی عمق تصویر استفاده کردیم و عمل کشف و قطعه‌بندی را انجام دادیم. پس از شناسایی دست از توصیف‌گر ESF استفاده کردیم. برای طبقه‌بندی الگوریتم پیشنهادی خود را به کار بردیم. در نهایت نیز گام‌های خوشه‌بندی و آموزش و آزمایش را انجام دادیم. الگوریتم پیشنهادی تشخیص زبان اشاره بنام جنگل تصادفی چندلایه‌ای فازی خود را در موقعیت‌های مختلف بر روی داده مصنوعی پایگاه داده قابل دسترس عمومی شامل اشاره‌های ASL و نیز داده عمق جمع شده از دوربین عمق اینتل^{۱۵} [۴] ارزیابی کردیم و نتایج متد خود را با متدهای شناخته شده و در شرایط خطای طبقه‌بندی (تعداد نمونه‌های طبقه‌بندی شده نادرست به اندازه مجموعه آزمایشی تقسیم شده است) زمان آموزش و نیز حافظه مصرفی مقایسه کردیم و در نتایج به دست آمده، مشاهده کردیم که الگوریتم پیشنهادی ارائه شده در مقایسه با جنگل تصادفی چند برابر سریع‌تر است. مشکل شباهت اشارات زبان اشاره ایستا با استفاده از کشف و قطعه‌بندی بر اساس عمق تصویر و نیز الگوریتم پیشنهادی ما در حوزه طبقه‌بندی حل می‌شود زمان آموزشی و حافظه مصرفی این الگوریتم نیز در شرایطی که سرعت آن بالا است، بسیار مناسب است.

- [23] W. Wohlkinger and M. Vincze. "Ensemble of shape functions for 3d object classification", In *ROBIO*, pp. 2987-2992, IEEE, 2011.
- [24] C. Liu, P. Liu, J. Liu, J. Huang, and X. Tang, "2D Articulated Pose Tracking Using Particle Filter with Partitioned Sampling and Model Constraints", *Journal of Intelligent and Robotic Systems*, Springer, vol. 58, Issue 2, pp. 109-124, May 2010.
- [25] P. P. Bonissone, J. M. Cadenas, M. D. Carmen Garrido and R. Andrés Díaz-Valladares, "A Fuzzy Random Forest: Fundamental for Design and Construction", *Foundations of reasoning under Uncertainty*, Springer, Studies in Fuzziness and Soft Computing, vol. 249, pp. 23-42, 2010.
- [26] L. Oikonomidis, N. Kyriazis and A. Argyros, "Efficient model-based 3D tracking of hand articulations using Kinect. In Jesse Hoey", *Proceedings of the British Machine Vision Conference*, pp. 101.1-101.11, BMVA Press, September 2011.
- [27] N. Petersen and D. Stricker, "Real-time 3-D Hand Posture Estimation based on 2-D Appearance Retrieval Using Monocular Camera", *Advances in Artificial Intelligence*, Springer, vol. 5803, pp. 106-113, 2009.
- [28] C. C. Wang and K. C. Wang, "Hand Posture Recognition Using Adaboost with SIFT for Human Robot Interaction", *Recent Progress in Robotics: Viable Robotic Service to Human*, Springer, vol. 370, pp. 317-329, 2008.
- [29] D. Freitas, "Rao-Blackwellised Particle Filtering for Fault Diagnosis", *Aerospace Conference Proceedings, IEEE*, vol. 4, 2008.
- [30] S. Wildermann and J. Teich, "3D Person Tracking with a Color-Based Particle Filter", *Robot Vision*, Springer, vol. 4931, pp. 327-340.
- [31] L. Shan, D. Metaxas, D. Samaras and J. Oliensis, "Using Multiple Cues for Hand Tracking and Model Refinement", *Computer Vision and Pattern Recognition, Proceedings. 2003 IEEE Computer Society Conference on*, vol. 2, pp. 443-450.
- Processing Systems*, Springer, vol. 65, Issue 1, pp. 63-79, October 2011.
- [15] A. Ortiz, J. Antich and G. Oliver, "A particle filter-based approach for tracking undersea narrow telecommunication cables", *Machine Vision and Applications*, Springer, vol. 22, Issue 2, pp. 283-302, March 2011.
- [۱۶] عقیل عبیری و محمدرضا محزون، «ردیابی اهداف متحرک هوایی با استفاده از تخمین چگالی کرنل بر اساس الگوریتم فیلتر ذره»، *مجله مهندسی برق دانشگاه تبریز*، دوره ۴۵، شماره ۳، صفحه ۹۷-۱۰۷، پاییز ۱۳۹۴.
- [17] V. Rudakova, S. K. Saha and F. A. Cheikh, "Multiple Collaborative Cameras for Multi-Target Tracking Using Color-Based Particle Filter and Contour Information", *Digital Information and Communication Technology and Its Applications*, Springer, vol. 166, 2011, pp. 315-332.
- [18] X. Dai and S. Payandeh, "Multi-camera Relay Tracker Utilizing Color-Based Particle Filtering", Springer, vol. 6753, pp. 395-405, 2011.
- [19] D. Mohr, G. Zachmann and S. Free, "Area-Based Articulated Object Tracking", *Advances in Visual Computing*, Springer, vol. 6938, pp. 112-123, 2011.
- [20] P. Bonissone, J. M. Cadenas, M. Carmen Garrido and R. A. Díaz-Valladares, "A fuzzy random forest", *International Journal of Approximate Reasoning*, Elsevier, vol. 51, Issue 7, pp. 729-747, September 2010.
- [21] N. Pugeault and R. Bowden. Spelling it out: "Real-time asl fingerspelling recognition", In *ICCV Workshops*, pp. 1114-1119, 2011.
- [22] Z. Ren, J. Yuan, and Z. Zhang. "Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera", In *K. S. Candan, S. Panchanathan, B. Prabhakaran, H. Sundaram, W. chi Feng, and N. Sebe, editors, ACM Multimedia*, pp. 1093-1096, ACM, 2011.

زیرنویس‌ها

⁹ FMLRF

¹⁰ Variance

¹¹ Intel Creative Depth Camera

¹² MATLAB

¹³ Intel Creative Gesture Camera

¹⁴ Resolution

¹⁵ Intel Creative Depth Camera

¹ American Sign Language

² Haarlet-like

³ Principle Component Analysis

⁴ SIFT

⁵ Microsoft Kinect

⁶ Ncut

⁷ CPRF

⁸ TF-IDF