

# تشخیص اجزای بدن انسان در تصاویر RGB-D با استفاده از ویژگی- های الگوی تغییرات عمق و تفاضل مکانی عمق

فرنوش عارفی<sup>۱</sup>، دانشجوی کارشناسی ارشد؛ علی نادیان<sup>۲</sup>، استادیار

۱- پژوهشکده‌ی فضای مجازی - دانشگاه شهید بهشتی - تهران - ایران - farnoosh.arefi@gmail.com

۲- پژوهشکده‌ی فضای مجازی - دانشگاه شهید بهشتی - تهران - ایران - a\_nadian@sbu.ac.ir

**چکیده:** تشخیص بخش‌های بدن انسان یکی از مهم‌ترین موضوعات پژوهش در دهه‌ی اخیر بوده است. این موضوع در حوزه‌هایی مانند تشخیص فعالیت، تشخیص حالت و سایر سامانه‌های مرتبط با فعالیت‌های حرکتی انسان، کاربرد گسترده‌ای دارد. هدف از سیستم تشخیص بخش‌های بدن انسان تعلق دادن هر پیکسل انسان به بخش‌های بدن می‌باشد. در تحقیقات اخیر نشان داده شده است، استفاده از نقشه‌ی عمق می‌تواند نتایج حاصل تشخیص بخش‌های بدن را بهبود بخشد. در این پژوهش ویژگی‌های جدیدی براساس تفاوت پیکسلی عمق ارائه شده است. ویژگی اول براساس تفاضل پیکسلی عمق بین پیکسل ورودی و همسایه‌های آن که براساس توزیع وزن دار حلقوی انتخاب شدند، تعیین شدند. ویژگی دوم تفاوت ضرایب چندجمله‌ای برازش شده از پیکسل ورودی در مقیاس‌های مختلف می‌باشد که موجب ایجاد ویژگی‌ای مستقل از مقیاس می‌شود. برای دسته‌بندی پیکسل‌ها از جنگل تصمیم تصادفی استفاده شده است. مقایسه نتایج روش پیشنهادی با روش‌های موجود نشان می‌دهد روش پیشنهادی توانسته است با دقت بیشتری اجزای مختلف بدن را تشخیص و تقسیم‌بندی کند.

**واژه‌های کلیدی:** تشخیص بخش‌های بدن، تصاویر RGB-D، ویژگی‌های مبتنی بر شکل، جنگل تصمیم تصادفی.

## Human body part detection in RGB-D image with pattern of depth difference and spatial depth difference features

F. Arefi<sup>1</sup>, M.Sc; A. Nadian<sup>2</sup>, Assistant Professor

1- Cyberspace Research Institute, Shahid Beheshti University, Tehran, Iran, Email: farnoosh.arefi@gmail.com

2- Cyberspace Research Institute, Shahid Beheshti University, Tehran, Iran, Email: a\_nadian@sbu.ac.ir

**Abstract:** Human body part detection has been an important research topic in the last decade. It is widely applicable in areas such as human activity recognition, pose detection and other applications related to human movements. The objective of a human body part detection system is to associate a body part to each human pixel. Recent studies show that applying depth maps significantly improves the results of body part detection. In this study, two new features based on pixel depth difference is proposed. First feature is based on pixel-wise depth difference between the input pixel and neighbor pixels selected using a weighted circular distribution. The second feature is the difference between coefficients of polynomials fitted to neighbors of the input pixel at difference scales, making the feature invariant scaling. Random decision forest was used for pixel classification. Comparison of results with the state of the art methods reveal that the proposed method is able to distinguish and differentiate the various components of the body more accurately.

**Keywords:** Human body part detection, RGB-D image, shape based feature, random decision forest.

تاریخ ارسال مقاله: ۱۳۹۶/۰۸/۰۵

تاریخ اصلاح مقاله: ۱۳۹۶/۱۰/۰۱

تاریخ پذیرش مقاله: ۱۳۹۶/۱۲/۱۳

نام نویسنده مسئول: علی نادیان

نشانی نویسنده مسئول: ایران - تهران - ولنجک - دانشگاه شهید بهشتی - پژوهشکده‌ی فضای مجازی

## ۱- مقدمه

تشخیص اجزای بدن نسبت به سایر روش‌های پیشین بالا رفته و اطمینان بیشتری در تشخیص اجزای بدن فراهم آورد.



شکل ۱: (الف) دستگاه کینکت، (ب) تصاویر رنگی و عمق، (ج) تقسیم‌بندی اجزای بدن

در ادامه در بخش ۲، به‌مرور کارهای پیشین پرداخته خواهد شد، سپس در بخش ۳ روش پیشنهادی شرح داده می‌شود. در بخش ۴ نحوه ی تولید بانک داده توضیح داده می‌شود، در بخش ۵ نتایج آزمایشگاهی را نشان داده خواهد شد و در بخش ۶، نتیجه‌گیری و کارهای آینده بیان می‌شود.

## ۲- کارهای پیشین

در پژوهش‌های گذشته، به‌صورت کلی محققان به دو صورت اقدام به تشخیص اجزای بدن به‌کمک تصاویر عمق کرده‌اند. اولین رویکرد، استفاده از ویژگی‌های مبتنی بر شکل و نقاط کلیدی و دومین رویکرد، استفاده از فناوری یادگیری عمیق در این حوزه می‌باشد. هدف کلی از روش‌های مبتنی بر توصیف شکل، ارائه‌ی هرچه بهتر اجزای بدن در حالت‌های مختلف می‌باشد. روش‌های این دسته عمدتاً تعدادی پیکسل آموزشی از ناحیه بدن [۳] یا لبه‌های تصویر بدن [۴] انتخاب می‌کنند و به اطلاعاتی که از این ویژگی‌ها به‌دست می‌آید، سعی بر توصیف مکانی نقاط آموزشی در همسایگی خود دارند. محبوب‌ترین روش تشخیص اجزا در رویکرد اول که در سال‌های اخیر استفاده‌ی زیادی از آن شده‌است، روش شاتون [۳] بوده‌است. این روش با آموزش یک مجموعه داده‌ی بسیار بزرگ مصنوعی تولید شده، توانست با دقت مناسبی اجزای مختلف بدن را در حرکات مختلف را شناسایی کند. شالوده‌ی این روش، براساس تولید داده‌های زیاد و آموزش ویژگی‌های وسیع توسط الگوریتم جنگل تصمیم تصادفی [۵] می‌باشد.

به‌صورت طبیعی، تهیه مجموعه داده‌ای از تصاویر عمق به‌همراه برچسب‌های آن کار بسیار سختی می‌باشد، زیرا افراد زیادی باید حرکات مختلفی را انجام دهند و همچنین برچسب‌زدن پیکسلی به هرکدام از بخش‌های این تصاویر عملاً کاری غیرممکن و طاقت‌فرسا هست. از این رو شاتون [۳] اقدام به تولید مجموعه داده مصنوعی کرد. برای آموزش اجزای مختلف این مجموعه داده، ابتدا از طریق ویژگی‌های مبتنی بر شکل اقدام به بازنمایی شکل هریک از اجزا کرد. سپس با استفاده از

سیستم تشخیص اجزای بدن انسان، بر اساس تصویر ورودی از انسان می‌تواند اجزای بدن انسان را افراز کند. به‌طور دقیق‌تر این سیستم می‌تواند پیکسل‌های مربوط به سر، دست‌ها، پاها و سایر اجزای اصلی بدن را تشخیص دهد (شکل ۱- ج). این سیستم استفاده‌ی زیادی در سامانه‌های تشخیص حالت و حرکت دارد. به‌طور مثال با استفاده از اجزای تشخیص داده شده در این سیستم، می‌توان اسکلتی سینماتیکی مثل روش [۱] طراحی کرد، که از طریق آن بتوان دامنه‌ی حرکتی از حرکات مختلف انسان را به‌صورت هوشمند اندازه‌گیری کرد. دامنه‌ی حرکتی تعریف شده برای انسان، بسیار بزرگ و حتی به شکلی می‌توان گفت نامحدود می‌باشد، از این رو در تحقیقات مختلف تلاش شده است در دامنه‌ی مشخصی از حرکات، اجزای مختلف بدن انسان تشخیص داده شود، و روش‌هایی معرفی شود که در صورت فراهم شدن داده‌های بیشتر، بتوانند در حرکات گوناگون، مکان اجزای بدن را تشخیص دهند. استفاده از تصاویر عمق نسبت به تصاویر رنگی، دامنه‌ی وسیع‌تری از ویژگی‌های مفید و متمایزکننده را برای تشخیص اجزای بدن فراهم می‌آورد، زیرا در این تصاویر فاصله‌ی هر بخش از بدن نسبت به بخش‌های دیگر مشخص است، بنابراین می‌توان از ویژگی‌های مبتنی بر فاصله و شکل برای تشخیص اجزای مختلف استفاده نمود. در تصاویر رنگی بخش‌های مختلف بدن، از نظر رنگ ویژگی متمایزکننده‌ی کمی دارند، بنابراین تشخیص اجزا در تصاویر رنگی کار پرچالش‌تری خواهد بود. از این رو، در این پژوهش، از تصاویر عمق برای تشخیص اجزای بدن انسان استفاده شده‌است. تصاویر عمق از دستگاه‌هایی به‌مانند کینکت<sup>۱</sup> یا دوربین زمان پرواز<sup>۲</sup> یا از سامانه‌ی بینایی استریو [۲] به‌دست می‌آیند. دستگاه کینکت، امروزه یکی از محبوب‌ترین دستگاه‌ها برای توسعه‌دهندگان در علوم بینایی ماشین در حوزه‌ی فعالیت انسان‌ها می‌باشد. این دستگاه، شامل دو دوربین رنگی و مادون قرمز می‌باشد که با دوربین مادون قرمز خود، می‌تواند فضای جلوی دوربین را به‌صورت سه‌بعدی ثبت کند و با دوربین رنگی، تصاویر دوبعدی رنگی را تولید کند. در شکل ۱ این دستگاه و خروجی تصویر رنگی و تصویر عمق مشاهده می‌شود.

در پژوهش پیش‌رو از ویژگی‌های جدیدی به نام الگوی تغییرات عمق برای توصیف پیکسل‌های هر یک از اجزا استفاده شده‌است. الگوی تغییرات باعث می‌شود تغییرات سطوح در همسایگی پیکسل‌های مورد ارزیابی استخراج شود. این توصیف به‌دلیل مقاومت بیشتر نسبت به نویزها و تغییر حالات بدن، باعث دقت بیشتر در تشخیص اجزای بدن انسان می‌شود. همچنین در این پژوهش از ویژگی تفاضل مکانی پیکسل‌ها استفاده شده‌است. این ویژگی یک توصیف سطح پایین از پیکسل‌های مورد ارزیابی هست که به‌دلیل راحت بودن محاسبه می‌تواند در حذف کلاس‌های نامرتب کارایی بالایی داشته‌باشد. استفاده توأم از این دو ویژگی با استفاده از جنگل تصمیم تصادفی که خود یک الگوریتم یادگیری مقاوم به داده‌های پرت هست، باعث شده‌است، دقت در

[۱۱، ۱۲] از تمام این ویژگی‌ها استفاده نکرد، بلکه با تخمین ویژگی‌ها براساس میانگین جابه‌جایی و ارائه‌ی ضریبی وزن‌دار که باتوجه به پهنای ابعاد ویژگی‌ها تعریف می‌شد، سعی بر خلاصه‌سازی، ویژگی‌ها یا همان پیکسل‌های هم‌سایه‌ی طی شده در طول مسیر را داشت. این خلاصه‌سازی، سرعت را بسیار افزایش می‌داد. نتایج کمی ارائه شده در این مقاله نشان می‌دهد، این الگوریتم با همان ویژگی‌های مبتنی بر شکل که توسط شاتون پیشنهاد شده بود از الگوریتم جنگل تصمیم تصادفی کارایی بهتری دارد. سان [۱۳]، از ویژگی‌های مبتنی بر شکل استفاده کرد با این تفاوت که یک سری محدودیت‌های فیزیکی مربوط به هر شخص در انتخاب ویژگی دخیل می‌شد، به‌طور مثال در هنگام آموزش اگر عضوی از بدن، دارای اندازه بزرگ‌تری می‌بود، تعداد پیکسل‌های آموزشی بیشتری انتخاب می‌شد. همچنین، هر بخش توزیع مشخصی از پیکسل‌های هم‌سایه برای استخراج ویژگی خود داشت. روش‌هایی دیگری مثل [۱۷-۱۴]، با تغییر بسیار کمی در استخراج ویژگی ارائه شده توسط شاتون، سعی بر بهبود دقت حاصل شده داشتند، این تغییرات شامل تغییر توزیع پیکسل‌های هم‌سایه و یا اعمال وزن دهی به هر یک از پیکسل‌های هم‌سایه می‌شد.

روش‌هایی که براساس ویژگی مبتنی بر شکل ارائه شدند، معمولاً برای هر یک از اجزاء، تعدادی نقاط منتخب انتخاب می‌کنند، سپس برای هر پیکسل منتخب در هر یک از اجزاء، بردار ویژگی را طبق روش گفته شده می‌سازند. از لحاظ سرعت استخراج ویژگی این روش بسیار کارآمد است زیرا تنها عملیات صورت گرفته پیدا کردن هم‌سایه‌های پیکسل منتخب (که از پیش تعیین شده‌اند) و عملیات تفاضل مقدار عمق پیکسل‌های هم‌سایه از پیکسل‌های منتخب می‌باشد. یکی از مشکلات این دسته از ویژگی‌ها، طولانی بودن بردار ویژگی می‌باشد [۹] که موجب چالش‌های زیادی در هنگام آموزش سیستم می‌شود. از این روش‌هایی مانند [۱۸] از تحلیل مؤلفه‌ی اصلی<sup>۳</sup> [۱۹] برای کاهش بعد استفاده کرده‌اند. دسته‌ای دیگر از ویژگی‌های به کاررفته در تشخیص اجزای بدن، ویژگی‌های مبتنی بر توصیف نقاط کلیدی می‌باشد. در ویژگی مبتنی بر نقاط کلیدی، هدف پیدا کردن نقاطی است که به صورت مشخص بتوانند هر یک از اجزاء را توصیف کنند. در روش لیو [۲۰]، ابتدا با استفاده از فاصله‌ی ژئودزیک<sup>۴</sup>، نقاط اکسترمم شناسایی شدند. این نقاط شامل سر، دست‌ها و پاها می‌باشد. سپس برای توصیف هر یک از نقاط اکسترمم، فاصله‌ی آن‌ها با سایر نقاط اکسترمم و نقاطی دیگر از بدن محاسبه می‌شد. همین روند در سطح سوپراپیکسلی هم اجرا شد. ترکیب این دو سطح از ویژگی بردار ویژگی نهایی را در این روش مشخص می‌کرد. فاصله‌ی ژئودزیک<sup>۵</sup>، به کوتاه‌ترین مسیر بین دو رأس در گراف گفته می‌شود. از این فاصله در سیستم تشخیص حالت بدن استفاده‌ی زیادی شده است [۲۱، ۲۲]. در این روش‌ها، ابتدا تصویر عمق به گرایی با رأس‌های مشخص تبدیل می‌شود، سپس با استفاده از عملیات جستجو، نقاطی با بیشترین فاصله‌ی ژئودزیک ممکن از هم پیدا می‌شوند. نقاط پیدا شده، نقاط اصلی بخش‌های بدن می‌باشند که می

الگوریتم جنگل تصمیم تصادفی اقدام به آموزش این ویژگی‌ها نمود. این مجموعه داده شامل کاراکترهای مختلف با ابعاد مختلف هست که تحت داده‌های حرکتی مجموعه CMU [۶]، به حرکت درآورده می‌شوند. از تک‌تک حرکات یک تصویر عمق و یک تصویر رنگی رندر می‌شود. همان‌طور که در شکل ۲ ملاحظه می‌شود، تصویر رنگی یک پوشش برجستگی برای هر پیکسل عمق ایجاد می‌کند که می‌تواند هم در آموزش و هم برای ارزیابی به کار گرفته شود.



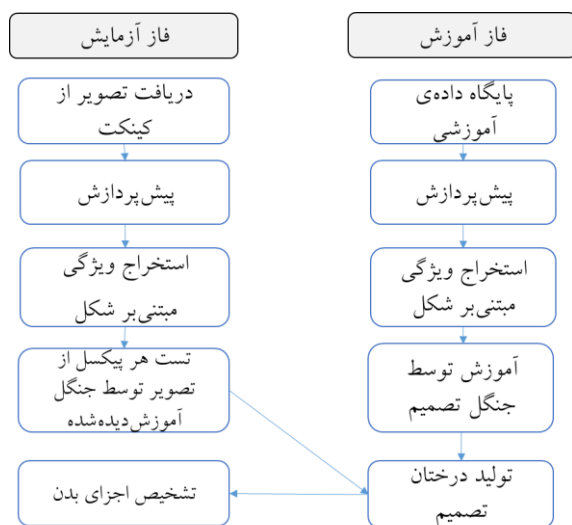
شکل ۲: تصاویر عمق و برجسب‌های بدن، (الف) برجسب‌های بدن، (ب) تصاویر عمق

ویژگی‌های مبتنی بر شکل، شکل هندسی اجزای مختلف بدن و نحوه‌ی قرارگیری هر یک از اجزاء را مشخص می‌کنند. در این روش برای استخراج ویژگی از یک پیکسل منتخب  $x$  در یک بخش مشخص، ابتدا یک سری پیکسل‌های هم‌سایه با یک توزیع خاص تعریف می‌شوند، سپس اختلاف عمقی پیکسل‌های هم‌سایه از مقدار عمق پیکسل  $x$  به عنوان بردار ویژگی برای آن پیکسل منتخب جهت آموزش و آزمایش استفاده می‌شود. در قسمت آموزش، کلاس هر بردار ویژگی برابر با برجستگی بخشی از بدن انسان هست که پیکسل ورودی در آن قرار گرفته است. در ادامه بردار ویژگی‌های به دست آمده در تصاویر آموزشی براساس جنگل تصمیم تصادفی دسته‌بندی می‌شود.

روش استخراج ویژگی مبتنی بر شکل روش شاتون [۳] از روش لپتیت [۷] الهام گرفته است. محققان زیادی به صورت مشخص از این روش استفاده کرده‌اند. به‌طور مثال، لیانگ [۸]، با تغییر توزیع نقاط دوبعدی اولیه به توزیع مربعی و تغییر اندازه‌ی تطبیقی این نقاط با توجه به عمق، دقت تشخیص اجزاء بدن را بهبود داده است. یکی دیگر از پژوهش‌های موفق که از روش استخراج ویژگی مبتنی بر شکل الهام گرفته است، روش بویس [۹] است، بویس از روش شاتون برای تشخیص بخش‌های بدن بدون حذف پس‌زمینه استفاده کرده و سپس اجزای به دست آمده را با تصویر رنگی تطبیق داده و برجسب‌های اشتباه را حذف کرد. گیرشیک [۱۰]، با اقتباس از ویژگی ارائه شده توسط شاتون [۳]، به جای ذخیره‌سازی احتمال وجود هر کلاس در هر برگ، ویژگی‌های استخراج شده در هر گره را در برگ‌ها قرار می‌دهد. این روش، به جای استفاده از جنگل تصمیم تصادفی از جنگل تصمیم رگرسیون استفاده کرده است. جنگل تصمیم رگرسیون، همان مزیت‌های جنگل تصمیم تصادفی را دارد با این تفاوت که در این الگوریتم، خروجی هر برگ پیوسته خواهد بود. از این رگرشیک [۱۰] برخلاف

### ۳- روش پیشنهادی

مراحل کلی تشخیص بدن انسان به کمک روش پیشنهادی در شکل ۳ نشان داده شده است. چهارچوب کلی سیستم به دو بخش آموزش و آزمایش تقسیم بندی می شود. در بخش آموزش، ابتدا مجموعه تصاویر عمق و برچسب های آن ها، از پایگاه داده آموزشی دریافت می شوند. سپس عملیات پیش پردازش شامل حذف نویز و حذف پس زمینه صورت می گیرد. در مرحله ی بعد، استخراج ویژگی مبتنی بر شکل از نقشه عمق به دو صورت الگوی تغییرات و تفاضل مکانی انجام می گیرد. در ادامه، این ویژگی های مبتنی بر شکل براساس الگوریتم یادگیری جنگل تصمیم تصادفی آموزش می بینند. در فاز آزمایش، ابتدا فریم های عمق از پایگاه داده دریافت می شوند، سپس بر روی تصاویر عملیات پیش پردازش انجام می شود و استخراج ویژگی مبتنی بر شکل از تصاویر ورودی اصلاح شده انجام می شود. در مرحله ی بعد، هر یک از پیکسل های غیر پس زمینه، در هر یک از درخت های جنگل تصمیم آموزش دیده پیمایش می شوند و در نهایت برچسبی مناسب با اجزای بدن، توسط الگوریتم جنگل تصمیم به هر پیکسل تعلق می گیرد. در ادامه به بخش پیش پردازش تصویر پرداخته خواهد شد.



شکل ۳: چهارچوب کلی روش پیشنهادی

#### ۳-۱- پیش پردازش تصویر

در بخش پیش پردازش تصویر، هدف استخراج ناحیه بدن و کاهش نویز می باشد. برای حذف پس زمینه از آستانه گیری تصویر عمق استفاده می شود. ابتدا در تصویر عمق، یک آستانه ی مشخص تعریف می شود که مقادیر بالاتر از این آستانه به عنوان پیکسل های پس زمینه در نظر گرفته می شوند. با این کار در یک صحنه، اجزای نامرتبط به انسان جلوی دوربین حذف می شوند و تمرکز ویژگی های تعریف شده تنها بر روی تصویر عمق انسان خواهد بود.

یکی از اقدامات دیگر در پیش پردازش، حذف نویزهای ناشی از گیرنده مادون قرمز کینکت که ماهیتی مشابه با نویز نمک و فلفل دارد،

توان با استفاده از تخمین های از پیش تعریف شده براساس هندسه ی بدن، سایر اجزای بدن را نیز تشخیص داد.

دسته دوم از روش های تشخیص اجزای بدن روش های مبتنی بر یادگیری عمیق می باشد. جیو [۲۳] یک شبکه ی یادگیری عمیق کانولوشنی با اعمال محدودیت های همسایگی اجزای مختلف تعیین کرد. در این روش از یادگیری بدون نظارت و با نظارت به صورت توأم استفاده شده است. این روش به جای فراهم آوری یک مجموعه داده ی برچسب خورده شده، از ترکیبی از داده های بدون برچسب و با برچسب استفاده کرده است. سپس با اعمال یک تابع انرژی براساس میدان تصادفی مارکوف، برچسب های به دست آمده در مرحله قبل را اصلاح کرد. شفاعی [۲۴] اقدام به تولید مجموعه داده ی وسیع برای آموزش در شبکه ی کانولوشنی خود کرد. شفاعی [۲۴] داده های مناسب و وسیعی تولید کرد اما مجموعه داده ی تولید شده برخلاف مجموعه داده تشکیل شده توسط شاتون [۳] به شکل طبیعی نبود. شاتون به داده های مصنوعی خود، نویزهای دستگاه کینکت را اعمال می کرد تا داده ها هرچه بیشتر به تصاویر طبیعی نزدیک شوند. نتایج حاصله در این مقاله نشان از دقت بالا شبکه ی طراحی شده در تشخیص اجزای بدن دارد. روش های مبتنی بر یادگیری عمیق، برای داشتن کارایی بالا نیاز به داده های زیاد دارند، با توجه به کمبود داده در این حوزه در روش پیشنهادی از رویکرد ویژگی های مبتنی بر شکل استفاده شده است.

ویژگی های استفاده شده در روش های پیشین، اغلب برای غلبه بر نویزهای موجود در تصویر، تعداد ویژگی زیادی استخراج می کردند که باعث سخت شدن فرایند آموزش و آزمایش می شد.

به طور کلی، ویژگی های مبتنی بر شکلی که در کارهای گذشته ارائه شده اند چند ایراد عمده دارند که می توانند دقت تشخیص اجزای بدن را کاهش دهند، این موارد عبارت است از:

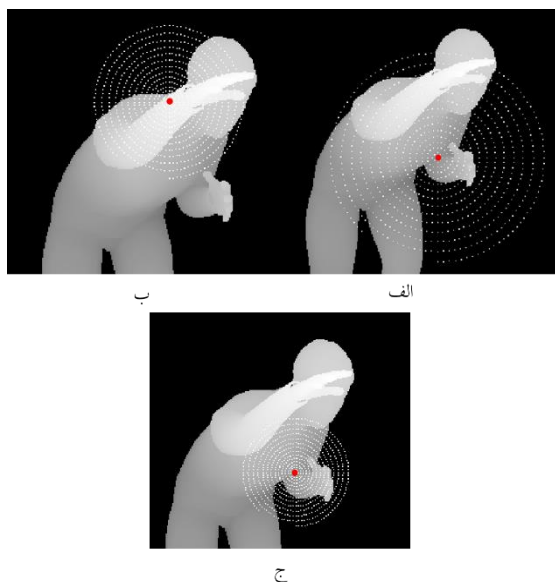
- تصادفی بودن ماهیت توزیع انتخابی برای پیدا کردن پیکسل های همسایه
- ارائه ی بردار ویژگی با ابعاد بسیار بالا
- ارائه الگویی حجیم و غیر کارآمد از نواحی همسایه به خصوص در تصاویر نویزی

در این پژوهش، دو ویژگی مبتنی بر تغییرات عمق برای افزایش دقت تشخیص اجزای بدن ارائه شده است. ویژگی تفاضل عمق پیکسل با پیکسل، که تفاضل عمق پیکسل ورودی با پیکسل های همسایه را محاسبه می کند. برای مشخص کردن مکان پیکسل های همسایه، یک توزیع وزن دار حلقوی ارائه شده است. ویژگی دوم، یعنی الگوی تغییرات عمق، از ضرایب چند جمله ای درجه سه تطبیق داده شده به یک دسته از پیکسل های مجاور پیکسل ورودی به دست می آید. این ویژگی به گونه ای طراحی شده است که نسبت به مقیاس و نویز تصویر عمق مقاوم می باشد. نتایج کمی نشان می دهد، استفاده از ویژگی الگوی تغییرات عمق تا حد زیادی توانسته است واریانس دقت در فریم های آزمایشی را کاهش و دقت تشخیص را افزایش دهد.

$b$  با توزیع مشخص  $R$  می‌باشد. اندیس‌های  $r$  و  $s$  نشان‌دهنده مکان پیکسل  $x$  در تصویر عمق می‌باشند.  $M$  ضریبی برای نرمال‌سازی و کنترل تأثیر مقدار جابه‌جایی برای تابع نگاشت می‌باشد. ضریب  $M$  توسط رابطه (۳) تعریف می‌شود.

$$M = \frac{f}{D(x)} \quad (3)$$

در این رابطه  $f$ ، ضریب ثابتی است که نشان‌دهنده تناسب فاصله‌ی پیکسلی با پیکسل  $x$  می‌باشد. اگر مقدار عمق پیکسل  $x$  بزرگ باشد، ضریب  $M$  کوچک خواهد شد، در نتیجه پیکسل‌های نگاشت نسبت به حالتی که مقدار عمق پیکسل کوچک است در مکان‌های نزدیک به پیکسل موردنظر قرار می‌گیرند. شکل ۵، تأثیر مقادیر مختلف  $f$  در نحوه‌ی انتخاب پیکسل‌های نگاشت با یک توزیع حلقوی را نشان می‌دهد.



شکل ۵: نحوه‌ی تأثیر پارامتر  $M$  در پیدا کردن پیکسل‌های همسایه با توزیع حلقوی، نقطه‌ی قرمز رنگ، پیکسل  $x$  و نقاط سفید پیکسل‌های نگاشت یا همسایه می‌باشند.

الف)  $f=600$  ,  $D(x)=148$  ، ب)  $f=600$  ,  $D(x)=239$  ، ج)  $f=300$  ,  $D(x)=148$

مهم‌ترین پارامتر در این روش از استخراج ویژگی، توزیع نقاط نگاشت می‌باشد که توسط تابع  $R$  محاسبه می‌شود. توزیع نگاشت، موقعیت پیکسل‌هایی را مشخص می‌کند که عمق آن‌ها باید از عمق پیکسل  $x$  کم شود. لازم به ذکر است که تابع توزیع تنها یک‌بار محاسبه می‌شود، به عبارتی، هنگام آموزش و آزمایش نمونه‌ها، مقادیر نقاط نگاشت یکسان می‌باشد. به صورت کلی هرچه توزیع تابع نگاشت، پیکسل‌های بیشتری در نزدیکی پیکسل منتخب  $x$  انتخاب کند، دقت بیشتر از حالتی هست که پیکسل‌ها از فاصله‌ی دورتری انتخاب شوند. با این وجود، در نظر گرفتن ارتباط پیکسل  $x$  با پیکسل‌های دورتر با شدت کمتری نسبت به پیکسل‌های همسایه، می‌تواند در مشخص کردن ناحیه‌ای از بدن که  $x$  به آن تعلق دارد مفید باشد.

می‌باشد. برای حذف این نویزها از فیلتر میانه‌ای استفاده شده است. شیوه این الگوریتم بدین صورت است که پنجره‌ی لغزانی با اندازه‌ی  $7 \times 7$  کل تصویر را پوشش می‌کند و سپس مقدار میانه هر پنجره را به جای پیکسل مرکز پنجره قرار می‌دهد. به کمک میانه‌گیری، پیکسل‌هایی با مقادیر بالا و یا پایین در تصویر حذف می‌شوند. تفاوت بین تصویر عمق خام و تصاویر عمق پس از مرحله پیش‌پردازش در شکل ۴ نشان داده شده است.



شکل ۴: الف) تصویر عمق خام، ب) تصویر بعد از پیش‌پردازش

یکی دیگر از عملیات پیش‌پردازش در این مرحله، انتقال محدوده انسان به مرکز تصویر برای هماهنگی بیشتر بین تصاویر آموزشی و آزمایشی می‌باشد.

### ۳-۲- استخراج ویژگی

پس از مرحله بهبود تصویر، ویژگی‌های توصیف‌کننده تصویر عمق استخراج می‌شود. ویژگی استخراج‌شده در روش پیشنهادی از جنس ویژگی‌های مبتنی بر شکل می‌باشد. این ویژگی، شکل هندسی هر یک از اجزای بدن را توصیف می‌کند. در مدل ارائه‌شده، ویژگی‌ها در دو سطح تفاضل مکانی عمق<sup>۵</sup> و الگوی تغییرات عمق<sup>۶</sup> استخراج می‌شوند. برخلاف کارهای پیشین که ویژگی‌ها تنها براساس اطلاعات موجود در سطح پیکسل‌ها به دست می‌آمدند، در روش پیشنهادی، علاوه بر در نظر گرفتن اطلاعات در سطح پیکسل، از اطلاعات ترکیبی پیکسل‌های مجاور نیز، استفاده شده است. ویژگی تفاضل مکانی عمق،  $SDD$  و ویژگی الگوی تغییرات عمق،  $PDD$  نامیده می‌شوند. برای هر پیکسل ورودی  $x$  یک بردار ویژگی تعریف می‌شود که به صورت  $FV = [SDD, PDD]$  نمایش داده می‌شود.

ویژگی تفاضل مکانی، نشان‌دهنده تفاضل عمق پیکسل ورودی با پیکسل‌های مجاور می‌باشد. بدین ترتیب، ویژگی  $SDD$  برای پیکسل  $x$  را می‌توان به صورت رابطه (۱) نوشت.

$$SDD(i)_x = D(Q_i(x)) - D(x) \quad (1)$$

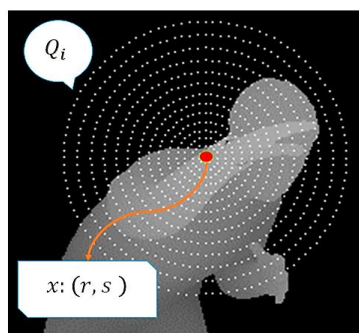
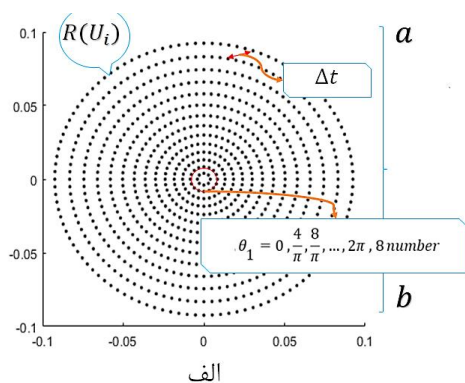
در این رابطه  $D$  تصویر عمق،  $i$  نشان‌دهنده  $i$  امین همسایه نقطه  $x$  و  $Q$  نشان‌دهنده مختصات نقاط همسایه  $x$  می‌باشد که به صورت رابطه (۲) تعریف می‌شود.

$$Q_i(x) = R(U_i) \times M + (r, s) \quad (2)$$

در رابطه (۲)،  $R$ ، تابع نگاشت می‌باشد که به ازای نقاط  $i$  در مجموعه  $U$  محاسبه می‌شود.  $U$  مجموعه‌ای از مختصات‌های دوبعدی، در بازه  $a$  تا



در روش پیشنهادی سعی بر این بود که در سطح‌های نزدیک به پیکسل  $x$  حلقه‌های بیشتری تولید شود، متغیر  $z$  وارد گام زاویه شده است. در رابطه بالا گام زاویه برابر  $4/\pi j$  در نظر گرفته شده است. متغیر افزایشی  $\Delta t$  شامل یک ضریب ثابت  $\tau$  می‌باشد که وظیفه کنترل بازه‌ی اعداد را به عهده دارد. این ضریب به صورت تجربی و با در نظر گرفتن فاصله تعریف شده‌ی  $a$  تا  $b$  به دست می‌آید. همچنین  $p$  متغیری است که به صورت پلکانی در هر سطح افزایش پیدا می‌کند و در  $\tau$  ضرب می‌شود. در شکل ۶-الف جزئیات پارامترها قابل مشاهده است. در شکل ۶-ب نقاط نگاشت حلقوی در فضای تصویر را مشاهده می‌فرمایید.



شکل ۶: الف) پارامترهای توزیع حلقوی، ب) پیکسل‌های نگاشت حلقوی. اختلاف عمق هر پیکسل نگاشت (سفید رنگ)، از پیکسل منتخب  $x$  (قرمز رنگ)، برابر با بردار ویژگی SDD این پیکسل منتخب خواهد بود. برچسب هر بردار ویژگی برای آموزش برابر محل قرارگیری پیکسل  $x$  می‌باشد، در این جا برچسب بردار ویژگی برای پیکسل  $x$ ، ساعد سمت چپ می‌باشد.

یکی از نقاط ضعف ویژگی‌های تفاضل مکانی پیکسلی، مقاوم نبودن در برابر نویز می‌باشد. برای مثال، با در نظر گرفتن شکل ۷ مشاهده می‌شود، تابع نگاشت نقطه‌ای از تصویر انتخاب کرده است که نشان گر محتوای اصلی ناحیه مورد نظر نیست، از این رو اگر به عنوان ویژگی انتخاب شوند، احتمال تصمیم‌گیری نادرست بیشتر می‌شود.

از این رو برای توصیف دقیق‌تر الگوی تغییرات عمق هر یک از اجزای بدن، از پیکسل  $x$  پاره‌خط‌هایی در جهت‌های مختلف رسم می‌شوند و سپس هر پاره‌خط در  $S$  مقیاس مختلف مورد بررسی قرار می‌گیرد. این

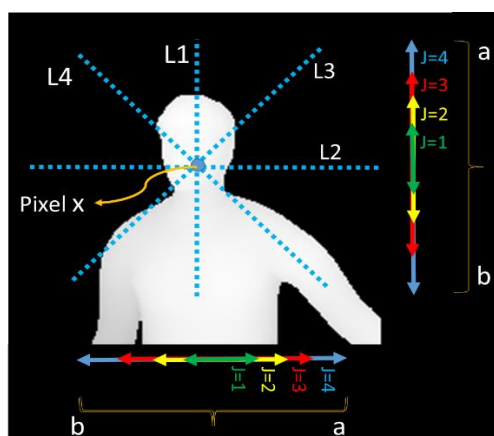
در کارهای گذشته، دو توزیع متفاوت برای مشخص کردن نقاط نگاشت معرفی شده است که در تحقیقات مختلفی مورد استفاده قرار گرفته‌اند. انتخاب تصادفی پیکسل‌ها در کل تصویر که توسط شاتون [۳] معرفی شد و تابع توزیع ثابت و تطبیقی که توسط لیانگ [۸] ارائه شده است. توزیع دوم دارای شکل مستطیلی و با دو نسخه ثابت و افزایشی است. در نسخه ثابت، نقاط نگاشت فاصله‌ی یکسانی از یکدیگر دارند، ولی در توزیع افزایشی، نقاط نگاشت بیشتری در فواصل نزدیک به پیکسل منتخب  $x$  نسبت به نواحی دورتر وجود دارند. روش شاتون [۳] و سایر روش‌هایی که از این روش الهام گرفته‌اند، به دلیل ماهیت توزیع تصادفی، این امکان وجود دارد پیکسل‌های نگاشتی تولید شوند که از پیکسل  $x$  فاصله‌ی بیشتری داشته باشند، که این باعث عدم دقت در این روش‌ها خواهد شد. علت این کاهش دقت این است که تصویر عمق انسان یک تصویر طبیعی است، در تصاویر طبیعی ناحیه‌های نزدیک به هم، خواص مکانی مشابه هم دارند و با حرکت‌ها و حالت‌های مختلف تغییر نمی‌کنند. از این رو ویژگی‌های نواحی نزدیک به پیکسل  $x$  در حالات مختلف بدن انسان، ثابت‌تر از نواحی می‌باشند که نسبت به این پیکسل دورتر می‌باشند. به طور مثال فرض شود پیکسل  $x$  در دست انسان واقع شده باشد، اگر تابع توزیع نگاشت تعریف شده به صورتی باشد که پیکسل‌هایی که در دست قرار دارند را انتخاب کند، از آنجا که این ویژگی‌ها با حرکت و حالات مختلف تغییر نمی‌کنند، اثربخشی بیشتری در دقت حاصل شده دارند. اما، اگر فرض شود تابع توزیع نگاشت ما پیکسل‌هایی از پا را انتخاب کند، چون آن نواحی وابستگی کمی به نقطه‌ی  $x$  دارند، باعث کاهش دقت می‌شوند.

تابع توزیع پیشنهادی در این پژوهش مکان پیکسل‌های نگاشت را بر روی حلقه‌هایی در نظر می‌گیرد که شعاع آن‌ها به صورت افزایشی نسبت به پیکسل  $x$  تغییر می‌کند (شکل ۶). رابطه (۴) توزیع پیشنهادی  $R$  را نشان می‌دهد.

$$R(U_i) = \{U_i \rightarrow (\Delta t \cos(\theta_j), \Delta t \cos(\theta_j)), \quad \theta_j = [0; \frac{4}{\pi j}; 2\pi), i \in U\}, \Delta t = \Delta t + \tau \times p \quad (4)$$

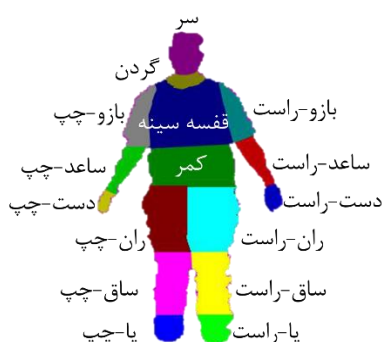
دلیل استفاده از توزیع حلقوی، پوشش زیاد و متراکم نقاط در این حالت می‌باشد. همچنین در این توزیع، به علت این که نقاط مختلف در هر سطح با فاصله‌ی شعاعی یکسانی از مرکز قرار می‌گیرند الگوی تغییرات عمق در فاصله‌ی یکسان در سطوح مختلف بهتر مشخص می‌شود. در رابطه (۴)،  $z$  تعداد سطوح دایره‌ای شکل را نشان می‌دهد. ضریب  $\Delta t$  متغیر افزایش دهنده‌ای می‌باشد که فاصله‌ی بین سطوح را مشخص می‌کند، در این رابطه  $\Delta t$  ابتدا دارای مقدار صفر است و با افزایش سطح مقدار آن به صورت تصاعدی افزایش پیدا می‌کند. این بدان معنی است که فاصله‌ی سطح‌های حلقوی نزدیک به پیکسل  $x$  فاصله‌ی کمتری نسبت به سطح‌های دورتر دارند. همچنین در این رابطه  $\theta_j$  آرایه‌ای از زوایای تشکیل دهنده حلقه است. هر چه مقدار گام طی شده از ۰ تا  $360^\circ$  درجه کوچک‌تر باشد، تعداد نقاط نگاشت بیشتر خواهد بود. از آنجایی که

برچسب هر بردار ویژگی، برابر موقعیت قرارگیری آن پیکسل در هر بخش بدن می‌باشد. هر بخش از بدن دارای شماره‌ی مشخص می‌باشد که به صورت قراردادی تعیین شده‌است.



شکل ۸: نمایش خطوط چهارگانه و نمایش سطوح مختلف در راستای افقی و عمودی (سطوح در راستای ارب هم به صورت سطوح در راستای افقی و عمودی مقیاس بندی شده‌است)

در این پژوهش برچسب گذاری اجزای بدن را ساس شکل ۹ انجام شده‌است. در روش پیشنهادی از ۴۸ ضریب رگرسیون چندجمله‌ای و ۷۹۲ ویژگی تفاضل مکانی با توزیع حلقوی استفاده شده‌است.

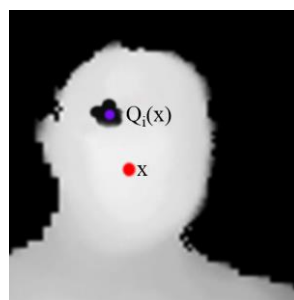


شکل ۹: نام و شماره‌ی اجزای تشخیص داده شده در این پژوهش

### ۳-۳- آموزش و آزمایش توسط جنگل تصمیم تصادفی

جنگل تصمیم یک روش متداول طبقه بندی در یادگیری ماشین می‌باشد که قادر است داده‌های بزرگ را با دقت زیاد طبقه بندی کند. این روش یکی از روش‌های یادگیری گروهی<sup>۷</sup>، و همچنین به شکلی دیگر، از روش‌های پیش‌بینی براساس نزدیک‌ترین همسایه می‌باشد. ایده‌ی اصلی این الگوریتم، این است که گروهی از یادگیرنده‌های ضعیف با ترکیب مناسب می‌توانند یک یادگیرنده قوی بسازند [۵]. درختان تصمیم‌گیری اغلب یا واریانس بالا یا بایاس بالا دارند، جنگل تصمیم تلاش می‌کند، متوسط طبیعی میان این دو را پیدا کند. با توجه به شرایط عملکردی جنگل تصمیم در داده‌های بزرگ و متنوع، برای آموزش ویژگی‌های استخراج شده از جنگل تصمیم استفاده می‌شود.

مقیاس بندی بدین منظور انجام می‌شود تا ویژگی پیشنهادی در برابر تغییر حالات مقاوم باشد.



شکل ۱۰: جای گیری نادرست پیکسل نگاشت در تصاویر نویزی. نقطه ی قرمز، پیکسل x و نقطه‌ی بنفش نقطه‌ی پیکسل نگاشت می‌باشد. پیکسل نگاشت در مکانی قرار گرفته‌است که نشان گر محتوی اصلی همسایه‌های پیکسل x نیست.

به کمک پیکسل‌های منطبق بر هر تکه از پاره خط، الگوی تغییرات عمق به کمک یک چندجمله‌ای درجه  $n$  تخمین زده می‌شود. این خطوط و مقیاس آن‌ها در شکل ۸ نمایش داده شده‌اند. برای تعریف بردار ویژگی  $PDD$  رابطه‌ی (۵) به صورت زیر تعریف می‌شود:

$$\vec{P}_{ij} = X\vec{N}_{ij} + \vec{e}, \quad \vec{N}_{ij} = (X^T X)^{-1} X^T \vec{P}_{ij} \quad (5)$$

$$PDD(x) = [\vec{N}_{ij} - \vec{N}_{ij-1}], i = 1 \dots L, j = S \dots 1$$

در رابطه (۵) مقادیر عمق پیکسل‌های خط  $i$  ام در مقیاس  $j$  را نشان می‌دهد.  $X$  ماتریس ضرایب پایه‌های چندجمله‌ای،  $\vec{e}$  خطای تصادفی و  $\vec{N}_{ij}$ ، ضرایب حاصل از برازش بر روی  $P_{ij}$  را نشان می‌دهد. بردار ویژگی  $PDD$  برابر با اختلاف دوه‌دوی ضرایب چندجمله‌ای در سطوح مختلف  $j$  ام هر خط  $i$  می‌باشد. بدین صورت الگوی تغییرات عمق برای  $x$  با انتخاب  $L$  پاره خط، یک بردار ویژگی با  $4 * L * (S-1)$  امان خواهد بود. در این مقاله  $L = 4$  و  $n = 3$  انتخاب شده‌اند. با توجه به تقسیم بندی هر خط به  $S = 4$  مقیاس و اعمال تفاضل دوه‌دو برای هر یک از ضرایب، ۱۲ ضریب به دست می‌آید. در مجموع با تعریف ۴ خط در راستای مختلف از پیکسل  $x$  جهت آموزش و آزمایش ۴۸ مقدار برای توصیف تغییرات الگوی عمق تولید می‌شود.

لازم به ذکر است، هر پاره خط به فضای مشخص شده  $a$  و  $b$  که قبلاً برای مشخص کردن نقاط  $U$  ذکر شده بود، محدود شده‌است و همچنین پیکسل‌های منطبق بر هر خط با نرخ نمونه برداری مشخص (به ازای هر دو پیکسل، یک پیکسل) استخراج می‌شوند.

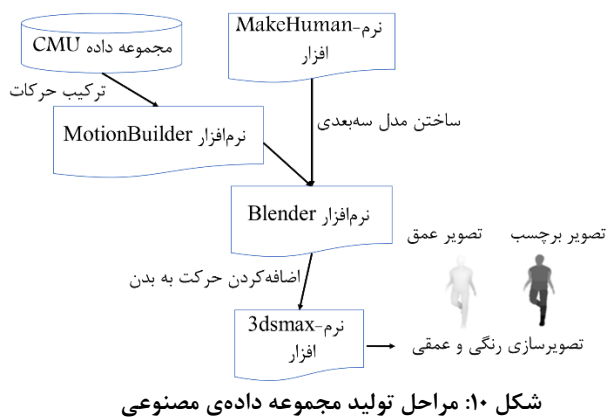
بردار ویژگی نهایی برای پیکسل منتخب  $x$  ترکیبی از ضرایب رگرسیون چندجمله‌ای و اختلاف عمق پیکسل همسایه از پیکسل  $x$  تحت توزیع نگاشت حلقوی می‌باشد.

برای آموزش اجزای مختلف، به ازای هر بخش از بدن، در هر فریم آموزشی تعدادی پیکسل انتخاب می‌شوند و سپس بردار ویژگی برای آن‌ها استخراج می‌شود.

شدند، ساخته شد. در شکل ۲ این دو تصویر را پیش‌تر مشاهده کردید.

در شکل ۱۰ فرآیند تولید فریم‌های آموزشی را مشاهده می‌شود. ابتدا، در نرم‌افزار شبیه ساز Makehuman<sup>TM</sup> [۲۵]، یک بدن سه‌بعدی که دارای مفصل مرتبط با بانک داده CMU [۶] می‌باشد ساخته می‌شود. سپس در نرم‌افزار Blender<sup>TM</sup> [۲۶]، حرکت بانک داده CMU [۶] به بدن مصنوعی تولید شده در مرحله قبل اضافه می‌شود. خروجی این مرحله یک انسان سه‌بعدی می‌باشد که دارای حرکت مشخصی است. از آن‌جا که تعداد حرکات زیاد می‌باشد، حرکات مختلف باهم ترکیب شدند، فرآیند ترکیب‌سازی در نرم‌افزار شبیه سازی Motionbuilder<sup>TM</sup> [۲۷] انجام شد. برای تصویربرداری از حرکات تولید شده باید مشخصات دستگاه کینکت یا هر دستگاهی که نقشه‌ی عمق را تصویربرداری می‌کند در محیط شبیه سازی به صورت مجازی ساخته می‌شود. این کار تحت زبان برنامه‌نویسی Maxscript در نرم‌افزار 3dsmax<sup>TM</sup> [۲۸] شبیه سازی شد. لازم به توضیح است به بدن ساخته شده در محیط شبیه سازی 3dsmax<sup>TM</sup> [۲۸] یک پوشش برجسیبی نیز اختصاص داده شد.

این پوشش برجسیبی شامل رنگ‌هایی است که به هریک از اجزای بدن اختصاص یافته‌اند. بنابراین هر فریم تصویربرداری شده از هر دوربین شامل یک تصویر عمق و یک تصویر شامل برجسب می‌باشد. تصاویر تولید شده به روش ذکر شده در مجموعه‌ای به نام Farpose [۲۹] قرار داده شده و در دسترس عمومی قرار گرفته است.



## ۵- نتایج آزمایشگاهی

در این بخش کارایی روش پیشنهادی برای تشخیص اجزای بدن بررسی می‌شود. در قسمت اول آزمایش، نحوه تغییر دقت حاصل شده با تغییر پارامترهای مربوط به روش پیشنهادی، طبق معیار صحت کل سنجیده خواهد شد. صحت کل<sup>۹</sup> برابر با میانگین پیکسل‌های به درستی تشخیص داده نسبت به تعداد پیکسل‌های بدن انسان در هر فریم  $I$  خواهد بود (رابطه ۸).

کلیدی آزمایش‌های پیش‌رو در مجموعه داده Farpose [۲۹] ارزیابی شدند. این مجموعه داده شامل ۳۴۵ فریم آموزشی و ۳۴۷ فریم آزمایش

همانند آنچه در [۷] توضیح داده شد، از جنگل تصمیم برای آموزش ویژگی‌های مبتنی بر شکل استفاده می‌شود. در مرحله‌ی آموزش به‌ازای هر بخش از بدن، ۲۵ پیکسل به‌صورت تصادفی جهت آموزش در جنگل تصمیم استفاده شد و طبق روش پیشنهادی، بردار ویژگی با دو سطح از ویژگی‌های PDD و SDD از تمام پیکسل‌های منتخب استخراج شد. خروجی فاز آموزش، شامل درخت‌هایی است که در هر گره آن ویژگی و آستانه‌ی جهت تقسیم گره قرار دارد و در هر برگ آن احتمال پسین هر کلاس قرار گرفته است. در فاز آزمایش هر پیکسل  $x$ ، از تمامی درختان پیمایش شده و مطابق با آستانه‌های هر گره از درخت به سمت برگ‌ها هدایت می‌شود. این فرآیند تا زمانی که به برگ‌های هر درخت برسیم ادامه پیدا می‌کند. به‌ازای هر درخت موجود  $t$  در جنگل تصمیم، احتمال پسین  $P_t(label|x)$  محاسبه می‌شود. لازم به توضیح است به‌ازای هر  $x$   $P_t(label|x)$  به تعداد برجسب‌های موجود دارای احتمالی مشخص است که مجموع این احتمالات برابر یک می‌باشد. برجسب نهایی هر پیکسل طبق رابطه‌ی ۶ و ۷ محاسبه می‌شود. در رابطه‌ی ۶،  $T$  تعداد درختان جنگل می‌باشد، احتمال پسین برجسب پیکسل  $x$  برابر با میانگین مجموع احتمال پسین برجسب پیکسل  $x$  در هر درخت می‌باشد. همچنین برجسب نهایی پیکسل  $x$ ، از رابطه ۷ محاسبه می‌شود، در این رابطه، برجسبی که بیشترین احتمال را داشته باشد، به‌عنوان برجسب نهایی پیکسل  $x$  انتخاب می‌شود.

$$P(label|x) = \frac{1}{T} \sum_{t=1}^T P_t(label|x) \quad (6)$$

$$label^* = \arg \max_{label} (P(label|x)) \quad (7)$$

## ۴- تولید بانک داده

بانک داده حرکت استفاده شده در دستگاه‌های تشخیص اجزای بدن، به‌صورت مرسوم بانک داده‌ی CMU [۶] می‌باشد، این بانک داده شامل حرکات واقعی از انسان می‌باشد که توسط سیستم ضبط حرکات گردآوری شده‌اند. در این بانک داده، حرکات مختلفی از انسان وجود دارد که توسط افراد مختلفی اجرا شده است. این حرکات شامل دویدن، نشستن، حرکات آکروباتیک، مشت‌زدن و سایر حرکات متداول روزمره و ورزشی می‌باشد. این بانک داده تنها شامل مفصل سه‌بعدی از بدن انسان می‌باشد، بنابراین، هیچ داده‌ی تصویری از این مجموعه داده موجود نمی‌باشد. محدود پژوهش‌هایی مثل شاتون [۳] و شفاعی [۲۴]، بانک داده‌ی تصویری از بانک داده CMU [۶] تولید کرده‌اند که در دسترس عموم قرار ندارند. بنابراین در این پژوهش، یک بانک داده جدید از بانک حرکتی CMU [۶] تولید شد. برای تولید بانک داده از هریک از حرکات، یک انسان مجازی یا به اصطلاح مصنوعی ساخته شد که براساس حرکات بانک داده CMU [۶] حرکت می‌کند. با تصویرسازی<sup>۸</sup> رنگی و عمقی در هر یک از فریم‌های هر حرکت، فریم‌های آموزش و آزمایشی که شامل دو تصویر عمق و برجسب‌های بدن می



بعدهای بیشتری از آن ویژگی در هر گره موردبررسی قرار می‌گیرند، بنابراین می‌توان از این تعداد آزمون برای تقسیم هر ویژگی استفاده نمود. در نمودار شکل ۱۱-ج، دقت روش پیشنهادی با افزایش عمق حداکثری درخت با شیب کمی افزایش پیدامی‌کند. صرف‌نظر از افزایش دقت حاصل‌شده، به‌دلیل زمان‌بر بودن آموزش این نوع درخت، تعداد عمق متوسط یعنی عمق ۲۲ کارتر به نظر می‌رسد.

در آزمایش دیگر، تأثیر پیش‌پردازش بر روی دقت حاصل‌شده سنجیده شده‌است. طبق این آزمایش کارایی روش پیشنهادی با اضافه شدن بخش پیش‌پردازش، صحت کل را در مجموعه داده Farpose [۲۹]، ۵.۲۳ درصد افزایش داده‌است.

در قسمت دوم آزمایش، کارایی روش پیشنهادی با تفکیک ویژگی‌ها و سپس در مقایسه با روش‌های دیگر بررسی می‌شوند. تنظیمات این سری از آزمایش‌ها به این صورت است که در هر فریم آموزشی به‌ازای هر بخش از بدن ۲۵ پیکسل برای استخراج ویژگی و آموزش به‌صورت تصادفی انتخاب شدند، با توجه به اینکه هدف تفکیک و شناسایی ۱۶ بخش بدن هست، در هر فریم آموزشی ۴۰۰ پیکسل برای آموزش انتخاب شده‌است. برای آموزش جنگل تصمیم از ۵ درخت با عمق ۲۲ و تعداد ۲۳ آزمون برای هر ویژگی استفاده شده‌است.

معیار ارزیابی در این سری از آزمایش براساس، بازخوانی<sup>۱</sup>، دقت<sup>۱۱</sup>، همسازي<sup>۱۲</sup>، صحت کل و واریانس می‌باشد. معیار بازخوانی، نسبت پیکسل‌های درست تشخیص داده‌شده در هر اندام ( $\alpha_j$ ) را به کل پیکسل‌های آن اندام ( $\eta_j$ ) مشخص می‌کند (رابطه ۸). معیار دقت، نسبت پیکسل‌های درست تشخیص داده‌شده در هر اندام را نسبت به تعداد کل پیکسل‌های تشخیص داده‌شده در آن اندام ( $\chi_j$ ) محاسبه می‌کند (رابطه ۱۰). همسازي، معیاری توأمان از معیارهای بازخوانی و دقت می‌باشد (رابطه ۱۱). واریانس صحت کل به‌دست آمده نیز در رابطه ۱۲ مشخص می‌باشد.

$$\text{Recall} = \frac{1}{\text{Frame}} \sum_i \frac{1}{\text{Part}} \sum_j \left\{ \frac{\alpha_j}{\eta_j} \right\} \quad (9)$$

$$\text{precision} = \frac{1}{\text{Frame}} \sum_i \frac{1}{\text{Part}} \sum_j \left\{ \frac{\alpha_j}{\chi_j} \right\} \quad (10)$$

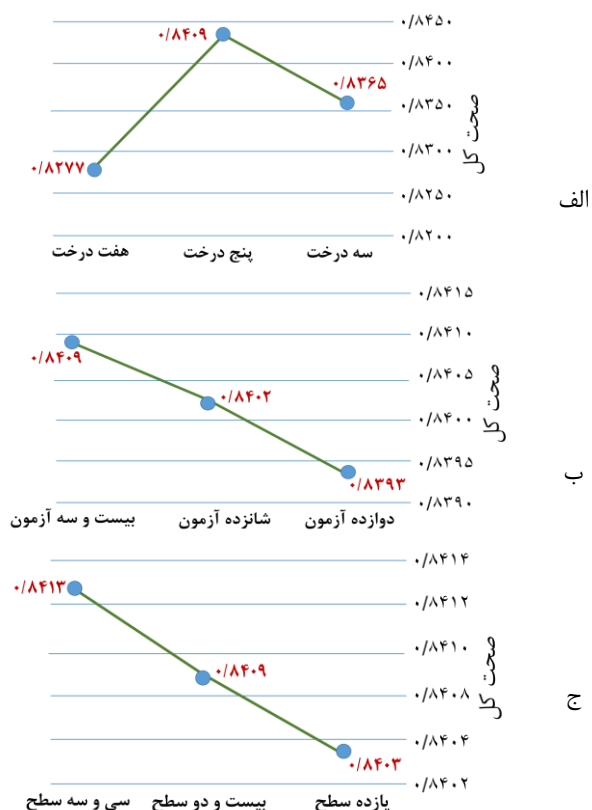
$$\text{F-measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (11)$$

$$\text{Variance(Var)} = \frac{1}{\text{Frame}} \sum_i \left( \left\{ \frac{\text{Correct pixel detected in } i}{\text{Total human pixel in } i} \right\} - \text{TA} \right)^2 \quad (12)$$

در آزمایش اول، ابتدا، دقت هر یک از ویژگی‌های PDD، ویژگی SDD و ترکیب این دو ویژگی را در مجموعه داده Farpose [۲۹] نشان داده خواهد شد. نتایج این آزمایش در جدول ۱ نشان داده شده‌است. همان‌طور که در جدول ۱ مشاهده می‌کنید، اضافه کردن ویژگی‌های PDD علاوه بر بهبود یک درصدی در صحت کل، توانسته است واریانس

با ابعاد ۴۸۰×۶۴۰ می‌باشد. در نمودار شکل ۱۱، تغییر صحت کل حاصل‌شده با تغییر درختان آموزشی قابل مشاهده می‌باشد.

$$\text{Total-Accuracy(TA)} = \frac{1}{\text{Frame}} \sum_i \left\{ \frac{\text{Correct pixel detected}}{\text{Total human pixel in } i} \right\} \quad (8)$$

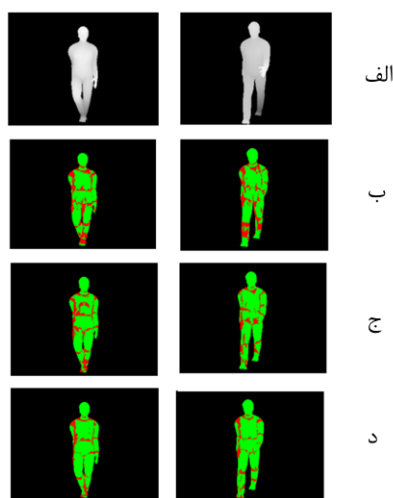


شکل ۱۱: تغییر دقت با (الف) تغییر درختان، (ب) تغییر آزمون‌ها، (ج) تغییر حداکثر سطوح عمق درخت

با نگاه به شکل ۱۱-الف به این نکته می‌توان اشاره کرد، تعداد درختان، در افزایش و کاهش دقت مؤثر بوده‌است. همان‌طور که مشاهده می‌شود، وجود ۵ درخت، در ساختن جنگل تصمیم کارایی بیشتری نسبت به ۳ و ۷ درخت فراهم می‌آورد. به‌صورت کلی اگر تعداد درختان کم باشد، چون تصمیم‌گیری براساس مجموعه‌ای از نمونه‌های زیاد در هر درخت صورت می‌گیرد، قابلیت انعطاف تصمیم‌گیری مناسب کاهش می‌یابد، درحقیقت رسیدن به یک نقطه‌ی بهینه باوجود داده‌های زیاد در هر درخت کاهش می‌یابد و هر چه تعداد درختان زیاد شود، تصمیم‌گیری با تعداد نمونه‌ی کم در هر درخت صورت می‌گیرد، با توجه به احتمال تعدد الگوهای مختلف در درختان مختلف، پیدا کردن یک الگوی مناسب کاهش می‌یابد. بهترین روش برای انتخاب تعداد درخت مناسب، توجه توأمان به تعداد ویژگی و تعداد نمونه‌ها می‌باشد، که در این پژوهش این تعداد، برابر ۵ در نظر گرفته شده‌است.

در نمودار شکل ۱۱-ب، تعداد ۲۲ آزمون برای تقسیم ویژگی نسبت به ۱۲ آزمون و ۱۶ آزمون با دقت بیشتری، توانسته اجزای بدن را شناسایی کند، علت این امر این است که با تقسیم بیشتر هر ویژگی

به کارایی استفاده از توزیع حلقوی پی برد. این توزیع در مقایسه با توزیع تصادفی ارائه شده توسط شاتون [۳] و توزیع شبه مستطیلی توسط لیانگ [۸] به ترتیب حدود ۷ و ۱ درصد بهبود داشته است. از نظر زمان برچسب زنی در هر پیکسل روش پیشنهادی به طور میانگین ۳۲ میلی ثانیه، روش لیانگ ۲۳ میلی ثانیه و روش شاتون ۱۸ میلی ثانیه داشته است. باین که در این پژوهش از ابتدا معیار کارایی تشخیص اجزای بدن مهم بوده است و نه برخط بودن سیستم، ولی باین حال، از نظر زمان سرعت محاسبه این روش بسیار نزدیک به سایر روش ها می باشد. در شکل ۱۲، برخی از فریم های آزمایش شده را در مقایسه با روش های دیگر مشاهده می کنید. در شکل ۱۳، دقت هر یک از بخش های بدن را مشاهده می کنید. بیشترین دقت به دست آمده مربوط به بخش های سر و گردن و کمترین دقت به دست آمده در قسمت سینه می باشد.



شکل ۱۲: مقایسه روش های مختلف در دقت تفکیک اجزای بدن در دو نمونه فریم آزمایشی، نواحی سبز، نواحی درست تشخیص داده شده، نواحی قرمز، نواحی اشتباه الف) فریم های ورودی، ب) خروجی روش شاتون [۳]، ج) خروجی روش لیانگ [۸]، د) خروجی روش پیشنهادی



شکل ۱۳: دقت تشخیص هر یک از اجزا در مجموعه داده Farpose [۲۹]

## ۶- نتیجه گیری

در این پژوهش، با ارائه ی ویژگی های جدید بر مبنای توصیف شکل اجزای بدن، دقت تشخیص اجزای مختلف در تصاویر عمق بهبود داده شد. ویژگی های استفاده شده موجب افزایش دقت و کاهش واریانس نسبت به روش های پیشین گردید. روش ارائه شده به علت کشف الگوی سطح بالا

صحت کل را حدود ۴ درصد بهبود دهد. همچنین در سایر معیار کارایی، این افزایش مشخص می باشد.

## جدول ۱: تفکیک دقت به دست آمده با اعمال هر یک از ویژگی

های SDD, PDD در مجموعه داده Farpose [۲۹]

| ویژگی SDD | ویژگی PDD | ویژگی SDD+PDD |          |
|-----------|-----------|---------------|----------|
| ۰/۸۳۴۹    | ۰/۷۴۴۴    | ۰/۸۳۱۹        | بازخوانی |
| ۰/۷۴۶۴    | ۰/۶۴۶۳    | ۰/۷۳۷۹        | دقت      |
| ۰/۷۸۸۲    | ۰/۶۹۱۹    | ۰/۷۸۳۰        | همسازي   |
| ۰/۸۴۰۹    | ۰/۷۲۴۶    | ۰/۸۳۱۰        | صحت کل   |
| ۰/۳۰۰۸    | ۰/۵۷۹۰    | ۰/۳۴۰۶        | واریانس  |

دومین آزمایش مقایسه روش پیشنهادی با روش های شاتون [۳] و لیانگ [۸] می باشد. از روش های مورد مقایسه همانند روش پیشنهادی به تعداد ۸۴۰ ویژگی انتخاب شده است.

در جدول ۲، کارایی تشخیص اجزای بدن را در مقایسه با روش های شاتون و لیانگ نشان داده شده است.

## جدول ۲: مقایسه ی دقت به دست آمده، در تشخیص بخش های بدن در

مجموعه داده Farpose [۲۹]

| روش پیشنهادی | لیانگ [۸] | شاتون [۳] |          |
|--------------|-----------|-----------|----------|
| ۰/۸۳۴۹       | ۰/۸۱۹۶    | ۰/۷۷۲۱    | بازخوانی |
| ۰/۷۴۶۴       | ۰/۷۲۰۰    | ۰/۶۷۳۷    | دقت      |
| ۰/۷۸۸۲       | ۰/۷۶۶۶    | ۰/۷۱۹۶    | همسازي   |
| ۰/۸۴۰۹       | ۰/۸۲۱۸    | ۰/۷۶۶۸    | صحت کل   |
| ۰/۳۰۰۸       | ۰/۳۵۰۶    | ۰/۳۴۷۲    | واریانس  |

روش پیشنهادی ارائه شده نسبت به روش شاتون [۳] حدود ۸ درصد و نسبت به روش لیانگ [۸] ۲ درصد در دقت کل و در همسازي نیز ۵ و ۲ درصد بهبود داشته است. نکته ی قابل تأمل کاهش حدود ۵ درصدی واریانس دقت در روش پیشنهادی نسبت به سایر روش ها می باشد، این معیار به این معنی است که کمترین دقت به دست آمده در روش پیشنهادی ما نسبت به میانگین دقت به دست آمده، فاصله ی کمتری نسبت به روش های دیگر دارد. به عبارت دیگر، در حرکات پیچیده یا حرکاتی که در فریم آزمایشی بودند ولی در فریم آموزشی نبودند روش های مورد مقایسه دقت کمتری نسبت به روش پیشنهادی این پژوهش داشته است.

علت این امر این است، که ویژگی الگوی تغییرات یک ویژگی سطح بالا از نواحی اجزای بدن تعریف می کند، یعنی این ویژگی مقاومت بیشتری در هنگام وجود انسداد توسط نواحی دیگر و یا وجود نویزهای مختلف نسبت به سایر ویژگی های استفاده شده در پژوهش های مورد-مقایسه از خود نشان می دهد، بنابراین در صورت عدم وجود یک حرکت در فریم آموزشی، روش پیشنهادی می تواند، با دقت بهتری، اجزای بدن در این گونه فریم ها را تقسیم بندی کند. همچنین در جدول ۱ و ۲ می توان

- [12] J. Gall and V. Lempitsky, "Class-specific Hough forests for object detection", Decision forests for computer vision and medical image analysis. pp. 143-157, 2009.
- [13] S. Min, P. Kohli and J. Shotton. "Conditional regression forests for human pose estimation", Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, pp. 3394-3401, 2012.
- [14] C. Ju Yong, S. Woo Nam. "Fast Random-Forest-Based Human Pose Estimation Using a Multi-scale and Cascade Approach", ETRI Journal, vol. 35, pp. 949-959, 2013.
- [15] T. Daniel, H. Liang and J. Yuan, "First-Person Palm Pose Tracking and Gesture Recognition in Augmented Reality", International Joint Conference on Computer Vision, Imaging and Computer Graphics, pp. 3-15, 2015.
- [16] Y. Jung, H. Lee, S. Heo and Y. Dong Yun, "Random tree walk toward instantaneous 3D human pose estimation", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2467-2474, 2015.
- [17] W. Xiaolin, P. Zhang and J. Chai, "Accurate realtime full-body motion capture using a single depth camera", ACM Transactions on Graphics, vol. 31 188, 2012.
- [18] J. Ahmad, Y. Kim and D. Kim. "Ridge body parts features for human pose estimation and recognition from RGB-D video data", Computing, Communication and Networking Technologies (ICCCNT), 2014 International Conference on. IEEE, pp. 1-6, 2014.
- [19] E. Brian, G. Dunn. "Principal components analysis", Applied Multivariate Data Analysis, Second Edition, pp. 48-73, 1993.
- [20] L. Yazhou, P. Lasang, M. Siegel and Q. Sun. "Geodesic invariant feature: A local descriptor in depth", IEEE Transactions on Image Processing, vol. 24, pp. 236-248, 2015.
- [21] S. Schwarz, L. Arthur, A. Mkhitarian, D. Mateus and N. Navab, "Human skeleton tracking from depth data using geodesic distances and optical flow", Image and Vision Computing, vol. 30, pp. 217-226, 2012.
- [22] C. Plagemann, V. Ganapathi, D. Koller and S. Thrun, "Real-time identification and localization of body parts from depth images", Robotics and Automation (ICRA), 2010 IEEE International Conference on. IEEE, pp. 3108-3113, 2010.
- [23] J. M., C. Wolf, G. Taylor and A. Baskurt, "Human body part estimation from depth images via spatially-constrained deep learning", Pattern Recognition Letters, vol. 50 pp. 122-129, 2014.
- [24] A. Shafaei, J. James, "Real-Time Human Motion Capture with Multiple Depth Cameras", Computer and Robot Vision (CRV), 2016 13th Conference on. IEEE, pp. 24-31, 2016.
- [25] <http://www.makehuman.org/>, December 10, 2017.
- [26] <https://www.blender.org/>, December 10, 2017.
- [27] <https://www.autodesk.com/products/motionbuilder/overview>, December 10, 2017.
- [28] <https://www.autodesk.com/products/3ds-max/overview>, December 10, 2017.
- [29] <https://github.com/noronet/sbupose>, December 10, 2017.

از اجزای مختلف توانسته نسبت به سایر روش‌های مورد مقایسه دقت بهتری را در هنگام تغییر حالت بدن در فریم‌های آزمایشی رقم بزنند. همچنین در این پژوهش یک مجموعه داده‌ی عمومی از تصاویر عمق به همراه برجسب‌های زمینه تهیه گردید که می‌تواند منبع مناسبی برای محققان آینده باشد. از کارهای آینده می‌توان به ارائه روشی برای اصلاح برجسب‌های اشتباه بعد از افراز کردن بدن انسان اشاره کرد.

## مراجع

- [۱] محمد امین نعمت‌اللهی، سید علی اکبر صفوی، محمد علی حاج‌عباسی، محمد رحیم همتیان، «حل سینماتیک وارون روبات‌های فرافزونه‌ای با استفاده از شبکه عصبی موجکی»، مجله مهندسی برق دانشگاه تبریز، دوره ۴۰، شماره ۱، صفحه‌ی ۵۷-۶۸، ۱۳۸۹.
- [۲] رسول قربانی، حمید ابریشمی مقدم، «استفاده از بینای استریو به منظور ارائه یک سامانه واقعیت افزوده مبتنی بر بینایی»، مجله مهندسی برق دانشگاه تبریز، دوره ۴۵، شماره ۱، صفحه‌ی ۱-۱۱، ۱۳۹۲.
- [3] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman and A. Blake, "Real-time human pose recognition in parts from a single depth image", Communications of the ACM, vol. 56 pp. 116-124, 2011.
- [4] L. He, G. Wang, Q. Liao and J. Xue, "Depth-images-based pose estimation using regression forests and graphical models" Neurocomputing, vol. 164, pp. 210-219, 2015.
- [5] L. Breiman. "Random forests", Machine learning, vol. 45, pp. 5-32, 2001.
- [6] <http://mocap.cs.cmu.edu/>, December 10, 2017
- [7] L. Vincent, P. Laguerre and P. Fua. "Randomized trees for real-time keypoint recognition", Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 2, pp. 775-781, 2005.
- [8] L. Hui, J. Yuan, and D. Thalmann. "Parsing the hand in depth images", IEEE Transactions on Multimedia vol. 16, pp. 1241-1253, 2014.
- [9] K. Buys, C. Cagniat, A. Baksheev, T. De Laet, J. De Schutter and C. Pantofaru, "An adaptable system for RGB-D based human body detection and pose estimation", Journal of visual communication and image representation, vol. 25. pp. 39-52, 2014.
- [10] R. Girshick, J. Shotton, P. Kohli, A. Criminisi and A. Fitzgibbon, "Efficient regression of general-activity human poses from depth images", Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, pp. 415-422, 2011.
- [11] B. Leibe, A. Leonardis and B. Schiele. "Robust object detection with interleaved categorization and segmentation", International journal of computer vision, vol. 77, pp. 259-289, 2008.

## زیر نویس‌ها

- <sup>7</sup> Ensemble learning
- <sup>8</sup> Render
- <sup>9</sup> Total-accuracy (TA)
- <sup>10</sup> Recall
- <sup>11</sup> Precision
- <sup>12</sup> F-measure

- <sup>1</sup> Kinect
- <sup>2</sup> Time of flight camera
- <sup>3</sup> Principal Component Analysis
- <sup>4</sup> Geodesic
- <sup>5</sup> Spatial depth difference
- <sup>6</sup> Pattern of depth difference