بسم الله الرحمن الرحيم

دانشگاه الزهرا (س)

دانشکده فنی و مهندسی

پایان‌نامه

جهت اخذ درجه کارشناسی ارشد

رشته مهندسی کامپیوتر - هوش مصنوعی

عنوان

تخمین حالت ۳بعدی بدن انسان

اساتید راهنما

دکتر حمیدرضا ربیعی

دکتر بهروز قلی‌زاده

دانشجو

مهدیه عباسی

بهمن ۱۳۹۱

دانشگاه الزهرا (س)

دانشکده فنی و مهندسی

پایان‌نامه

جهت اخذ درجه کارشناسی ارشد

رشته مهندسی کامپیوتر - هوش مصنوعی

عنوان

تخمین حالت ۳بعدی بدن انسان

اساتید راهنما

دکتر حمیدرضا ربیعی

دکتر بهروز قلی‌زاده

استاد مشاور

دکتر رضا عزمی

دانشجو

مهدیه عباسی

بهمن ۱۳۹۱

دانشگاه اازهرا
دانشکده فنی و مهندسی

پایان‌نامه جهت اخذ درجه کارشناسی ارشد

تخمین حالت ۳بعدی بدن انسان

نگارش: مهدیه عباسی

اساتید راهنما: دکتر حمیدرضا ربیعی امضاء:

دکتر بهروز قلی‌زاده امضاء:

استاد ممتحن داخلی: دکتر مهران شرقی امضاء:

استاد ممتحن خارجی: دکتر نصرالله مقدم امضاء:

کلیه‌ی دستاوردهای این تحقیق
متعلق به دانشگاه الزهرا(س) می‌باشد.

تشکر و قدردانی

اینک که با لطف و یاری خداوند متعال موفق به انجام این تحقیق شده‌ام بر خود لازم می‌دانم از زحمات اساتید ارجمند آقای دکتر ربیعی، دکتر عزمی و دکتر قلی‌زاده تشکر کنم. به خصوص آقای دکتر ربیعی، در تمامی مراحل کار، با دقت و ریزبینی آن را هدایت کردند و همواره امکانات مورد نیاز را به نحو احسن برای اجرای پروژه در اختیار من قرار دادند.

همچنین لازم می‌دانم مراتب قدردانی خود را از همکاران اعضای گروه هوش مصنوعی آزمایشگاه رسانه دیجیتال دانشگاه صنعتی شریف، به خصوص جناب آقای دکتر رهبان به پاس همکاری ها و همفکری هاشان در این پژوهش اعلام نمایم. از آقای ذوالقدر، مدیر داخلی آزمایشگاه که شرایط مناسبی برای پژوهش در این آزمایشگاه را برای من فراهم آورده‌اند، کمال تشکر را دارم.

در نهایت و بیش از همه، از تمامی اعضای خانواده‌ام به ویژه پدر و مادر مهربان و دلسوزم به خاطر حمایت‌ها و محبت‌های بی‌شائبه‌شان در طول انجام این تحقیق و نگارش این پژوهش خالصانه و بی‌اندازه متشکرم.

عباسی

زمستان ۱۳۹۱

چکیده

یک دسته از رویکردها برای تخمین حالت ۳ بعدی بدن انسان، رویکرد تمایزی است که با یادگیری مدلی، حالت ۳ بعدی را از ویژگی‌های تصویر ورودی تخمین می‌زند. به دلیل نبود اطلاعات عمق در تصویر، این رویکرد با چالش ابهام در عمق روبرو شده است. برای حل این چالش که یکی از اساسی‌ترین چالش‌های مساله تخمین حالت ۳ بعدی است، روش‌های نظارتی متعددی ارایه شده است. اما این روش‌ها به داده‌های آموزشی بسیار زیادی (بیش از هزاران) نیاز دارند. بنابراین در سال‌های اخیر به منظور کاهش این نیاز، روش‌های نیمه‌نظارتی برای حل چالش ابهام در عمق مورد توجه قرار گرفتند.

یک دسته از روش‌های نیمه نظارتی، روش‌های مبتنی بر گراف است که برای تخمین برچسب داده‌ها از گراف که تخمینی از منیفولد داده‌ها است، استفاده می‌کنند. یکی از مولفه‌های اصلی این روش‌ها، ساخت گراف همسایگی است به طوری که این گراف همسایگی از داده‌های بدون برچسب و برچسب‌دار ساخته می‌شود. بهبود کیفیت ساخت گراف تاثیر به سزایی در کارایی و دقت این روش‌ها دارد. اما به دلیل استفاده از فاصله‌ی اقلیدسی در ساخت گراف، گراف ساخته شده شامل یال‌های میان‌بر می‌باشد. این یال‌های میان‌بر، داده‌ها با برچسب‌های متفاوت را به یکدیگر متصل می‌نمایند.

در این پایان‌نامه روشی نیمه نظارتی مبتنی بر گراف جهت تخمین حالت ۳بعدی از دنباله‌ای از سیاه‌نماها ارایه شده است. اما به دلیل استفاده از فاصله‌ی اقلیدسی و چالش ابهام در عمق، گراف ساخته شده شامل یال‌های میان‌بر است. با بهره‌گیری از پنجره‌های زمانی لغزنده و مقایسه شباهت بین پنجره‌های زمانی، یال‌های میان‌بر در گراف همسایگی را کشف و حذف می‌نماییم. علاوه بر رفع چالش ابهام در عمق، به واسطه استفاده از رابطه‌ی بین داده‌های بدون برچسب و برچسب‌دار چالش نیاز به داده‌ی آموزشی زیاد نیز مرتفع می‌شود. روش پیشنهادی را بر روی فعالیت‌های گوناگونی آزمایش کردیم و نتایج حاصل از آن را با چند روش‌های نیمه‌نظارتی اخیر مقایسه نمودیم. کاهش میانگین مربع خطا (MSE) روش پیشنهادی نسبت به روش‌های اخیر و همچنین گراف اولیه نشان دهنده‌ی بهبود عملکرد روش پیشنهادی است.

کلمات کلیدی: *رویکرد تمایزی- تخمین حالت ۳ بعدی - فرض منیفولد- گراف همسایگی- ابهام در عمق- پنجره‌ی زمانی*

فهرست مطالب

۱ مقدمه ۱

۱.۱ بیان مساله ۱

۱.۱.۱ مشخصه ورودی ۲

۲.۱.۱ مشخصه خروجی ۴

۲.۱ اهمیت موضوع و کاربردها ۶

۳.۱ چالش‌ها ۶

۴.۱ رویکردهای حل مسأله ۹

۱.۴.۱ روش‌های بالا به پایین ۹

۲.۴.۱ روش‌های پایین به بالا ۱۱

۵.۱ معیار ارزیابی و مقایسه ۱۱

۶.۱ پایگاه داده ۱۳

۷.۱ راهکار پیشنهادی ۱۴

۸.۱ ساختار پایان‌نامه ۱۵

۲ پژوهش‌های پیشین ۱۶

۱.۲ یافتن فرد در تصویر ۱۶

۲.۲ استخراج ویژگی ۱۹

۱.۲.۲ زمینه شکل ۲۰

۲.۲.۲ توصیف کننده‌ی $SIFT$ ۲۲

۳.۲ روش‌های تخمین حالت ۳ بعدی بدن انسان ۲۳

۱.۳.۲ رویکرد مولد ۲۳

۲.۳.۲ رویکرد تمایزی ۲۷

۴.۲ نتیجه گیری ۴۳

۳ شرحی بر روش‌های نیمه نظارتی مبتنی بر منیفولد ۴۴

۱.۳ مقدمات ۴۵

۱.۱.۳ تعریف منیفولد ۴۵

۲.۱.۳ روش‌های ساخت منیفولد ۴۶

۲.۳ اهمیت روش‌های نیمه‌نظارتی مبتنی بر منیفولد ۴۷

۳.۳ اعمال فرض منیفولد ۴۸

۴.۳ نتیجه گیری ۵۰

۴ روش پیشنهادی و ارزیابی نتایج **۵۱**
۱.۴ فرضیات روش پیشنهادی ۵۲
۱.۱.۴ نشانه‌گذاری ۵۳
۲.۴ روش پیشنهادی ۵۳
۱.۲.۴ اصلاح گراف ۵۳
۲.۲.۴ بهینه‌سازی تابع هدف ۵۸
۳.۴ مجموعه داده ۶۰
۴.۴ مقایسه گراف‌های متفاوت و تحلیل خطای ناشی از آن‌ها ۶۱
۵.۴ مقایسه با کارهای پیشین ۶۳
۶.۴ نتیجه گیری ۶۵

۵ نتیجه گیری و کارهای آینده **۶۷**
مراجع ۶۸

آ PCA احتمالاتی **۷۴**

ب فرمت bvh **۷۶**

فهرست تصاویر

۱.۱ نمونه‌ای از ابهام در عمق [۳] ۴
۲.۱ مدل درختی حرکتی و مدل حجمی [۵] ۵
۳.۱ کاربرد تخمین حالت انسان ۷
۴.۱ چالش ابهام در عمق ۸
۵.۱ رویکردهای حل مساله تخمین حالت ۳ بعدی ۱۲

۱.۲ کشف مکان فرد در تصاویر با پس‌زمینه‌ی ثابت ۱۷
۲.۲ هیستوگرام قطبی -لگاریتمی [۲] ۲۱
۳.۲ بردارهای هیستوگرام قطبی-لگاریتمی ۲۱
۴.۲ ماتریس شباهت زمینه شکل و حالت واقعی ۳ بعدی ۲۲
۵.۲ استخراج نقاط مورد علاقه با روش هریس [۴۵]. ۲۳
۶.۲ مدل گرافیکی انتشار چگالی ۲۶
۷.۲ نحوه‌ی مقایسه‌ی تصویر بازسازی شده با تصویر اصلی ۲۷
۸.۲ یادگیری چند تابع برازش ۳۰
۹.۲ نحوه‌ی خوشه بندی داده‌ها ۳۱
۱۰.۲ منیفولد راه رفتن ۳۵
۱۱.۲ برازش تابع با فرایند نرمال ۳۷
۱۲.۲ نمایش گرافیکی روش PCA احتمالاتی و $GPLVM$ ۳۸

۱.۳ تخمین منیفولد داده‌ها با گراف همسایگی [۲۲] ۴۶
۲.۳ اهمیت روش نیمه‌نظارتی مبتنی بر منیفولد ۴۸

۱.۴ تفاوت فاصله‌ی اقلیدسی و منیفولدی ۵۴
۲.۴ منیفولد داده‌های راه رفتن ۵۴
۳.۴ ساخت گراف زمانی دنباله‌ی داده‌ها ۵۵
۴.۴ مقایسه‌ی پنجره‌های زمانی دو گره با یال متصل کننده ۵۵
۵.۴ گراف ۴ نزدیک ترین همسایه برای ۲۰۰ داده‌ی فعالیت راه رفتن دایره‌ای. همسایه‌ها با فاصله‌ی اقلیدسی در فضای ورودی مشخص شده‌اند. یال‌های قرمز رنگ، یال‌های میان‌بری است که توسط روش ارایه شده در بالا، پیدا شده‌اند. ۵۷
۶.۴ خطای زوایای بدن به ازای مقادیر متفاوت k ۶۲

۷.۴ مقایسه‌ی بصری زوایای واقعی با زوایای تخمین زده شده توسط روش پیشنهادی و گراف اولیه. ستون اول زوایای واقعی و ستون دوم نتیجه روش پیشنهادی و در ستون سوم نتایج گراف اولیه آورده شده است. ۶۴
۸.۴ محور عمودی میانگین خطا (MSE) و محور افقی ۱۸ زوایه از مفاصل بدن را نشان می‌دهد. در این نمودارها روش‌های پیشنهادی، $GC + RT$ ، گراف اولیه و TGP را با یکدیگر مقایسه نموده‌ایم. ۶۵

آ.۱ نمایش گرافیکی روش PCA احتمالاتی ۷۵

فهرست جداول

۱.۴ مقایسه روش پیشنهادی با گراف اولیه و با دو روش نیمه نظارتی ۶۴

ب.۱ فرمت bvh. ۷۷

فصل ۱

مقدمه

۱.۱ بیان مساله

با توجه به ارزان شدن دوربین‌های دیجیتال و افزایش قدرت محاسباتی رایانه‌ها، سامانه‌های هوشمندی که قابلیت کشف انسان در تصویر، دنبال کردن، تخمین حالت ۳بعدی و تشخیص فعالیت‌های در حال انجام او را داشته باشند، در زمینه‌های پزشکی و امنیتی و همچنین در تولید بازی‌های رایانه‌ای کاربرد فراوانی پیدا کرده است. هسته‌های اصلی این سامانه‌های هوشمند، بایستی توانایی کشف و دنبال کردن انسان و تخمین حالت در فضای ۳ بعدی و تشخیص فعالیت را دارا می‌باشند. در حال حاضر برای تامین امنیت در مراکز حساس، با قرار دادن چندین دوربین در نقاط مختلف رفتار غیر نرمال مشتریان تشخیص داده می‌شود. برای تحقق امنیت به اپراتورهایی برای نظارت بر رفتار مشتریان نیاز است که علاوه‌بر غیر خودکار بودن به دلیل خطاهای انسانی، غیر مطمئن نیز هستند. پس ایجاد سامانه‌ای که بتواند به طور خودکار و دقیق با هزینه راه اندازی کم، سوژه را دنبال و حالت ۳بعدی و فعالیت‌های آن را تشخیص دهد، موردنیاز است.

با نصب تعداد زیادی نشانگرهای لیزری بر روی بدن و نصب چندین دوربین در دیدهای متفاوت در یک محیط آزمایشگاهیِ کنترل شده، حالت ۳ بعدی بدن می‌تواند دقیقاً تخمین زده شود. اما این روش به دلیل نصب چندین دوربین و نشانگرهای لیزری متصل به بدن و ایجاد محیطی همانند محیط آزمایشگاه (محیط‌های کنترل شده) کاری هزینه‌بر است و برای محیط‌های واقعی کاربردی نیست[۲۴]. به همین دلیل ایده‌ی استفاده از این نشانگرهای لیزری راه‌حل بهینه‌ای به شمار نمی‌رود. راه حل دیگری که ارائه شده است، استفاده از چندین دوربین در زوایای دید متفاوت می‌باشد. اما این راه‌حل نیز به دلیل استفاده از چند دوربین

مورد استقبال واقع نشد. از همین رو محققان روش‌هایی را ارائه می‌دهند تا تنها با یک دوربین، حالت ۳ بعدی بدن را به طور نسبتاً دقیق تخمین بزنند. اما به دلیل چالش‌های موجود (بخش ۳.۱) نتایج مطلوب و دقیقی که برای همه‌ی محیط‌ها کارا و کاربردی باشد، بدست نیامده است.

در این پایان نامه به مسأله‌ی تخمین حالت ۳ بعدی بدن انسان از یک دید (با استفاده از یک دوربین)، که یکی از وظایف این سامانه‌های هوشمند است، می‌پردازیم. مسأله‌ی تخمین حالت ۳ بعدی بدن انسان از یک دید، به کل بدن انسان به عنوان یک جسم مفصل‌دار نگاه می‌کند و با استفاده از تصویر یا دنباله‌ای از تصاویر حاوی اطلاعات ۲ بعدی، زوایای مفاصل بدن در فضای ۳ بعدی را تخمین می‌زند. در ادامه‌ی این بخش ابتدا مشخصه‌ی ورودی و سپس مشخصه‌ی خروجی در این مسأله را بیان می‌کنیم.

۱.۱.۱ مشخصه ورودی

همان گونه که بیان کردیم قصد داریم با استفاده از تک دوربین، حالت ۳بعدی بدن انسان را بازسازی کنیم. پس واضح است که ورودی مساله تنها یک تصویر یا دنباله‌ای از تصاویر است. این تصاویر حاوی اطلاعات اضافی و بی‌ربط به مساله از قبیل بافت لباس، تغییرات نور، رنگ پوست و پس‌زمینه و ... است که نباید آن‌ها را در نظر گرفت. علاوه‌براین اطلاعات بی‌ربط، هر تصویر مقداری نویزی نیز هست. بنابراین مناسب است که بدون درگیر شدن با این اطلاعات اضافی، ورودی ساده‌ای که حاوی اطلاعات ارزشمندی برای حل این مسأله است، در نظر بگیریم. به همین منظور از روش‌های معمول در پردازش تصویر [۱] استفاده می‌شود. این روش‌ها اطلاعات ساده اما مرتبط با مساله را از عکس استخراج می‌کنند.

علاوه‌بر این دقت رویکرد تمایزی (به ۲.۴.۱ رجوع شود) به توصیفی که از این ورودی ارائه شده است، بستگی دارد. این توصیف بایستی توصیف سطح بالایی باشد به طوری که حالت هر عضو به خوبی در آن توصیف شده باشد و با تغییرات در پس زمینه، شرایط نور، پوشش فرد و رنگ پوست او تغییر نکند.

سیاه نما

می‌خواهیم اطلاعاتی از تصویر ورودی را بدون در نظر گرفتن نویزهایش به گونه‌ای استخراج کنیم که مرتبط‌ترین توصیف را با مساله تخمین ۳ بعدی داشته باشد و شکل کلی حالت بدن را بدون در نظر داشتن ظاهر و پوشش فرد، استخراج کند. به همین منظور سیاه‌نما یکی از مشخصه‌های مناسبِ ورودی محسوب می‌شود.

[۱] image processing

سیاه‌نما، تصویری دودویی از فرد است که با تفریق کل تصویر از پس زمینه‌ی ثابت به دست می‌آید (۱.۱). در مقاله [۲] دلایلی برای انتخاب سیاه نما برای این مساله ذکر شده است:

۱. استخراج سیاه‌نماها از تصاویر با پس زمینه ثابت با فرض فقدان سایه با اطمینان قابل قبولی انجام می‌پذیرد.

۲. سیاه‌نماها بدون درگیر شدن با پوشش و بافت و رنگ لباس فرد و پس زمینه تصویر می‌تواند حالت ۳ بعدی را تا حد معقولی نمایش دهد.

اگرچه سیاه‌نماها، ورودی مناسبی برای حل مساله تخمین حالت ۳ بعدی است اما همیشه استخراج دقیق سیاه‌نما از تصویر به دلیل سایه‌ها و تار شدن قسمتی از بدن به‌علت حرکات سریع امکان‌پذیر نیست و سیاه‌نمای استخراج شده از دقت لازم برخوردار نیست. از دیگر مشکلات سیاه نماها از دست رفتن اطلاعات عمق است. به دلیل از دست رفتن اطلاعات عمق، ممکن است سیاه‌نماهای دو تصویر در دو حالت متفاوت کاملاً مشابه باشد (شکل ۱.۱).

لبه‌ها

لبه‌های تصویر فرد نیز اطلاعات مفید دیگری است که علاوه بر اینکه استخراج آنها هزینه‌بر نیست، با افزودن آنها تا حدودی مشکل ابهام در عمق حل خواهد شد. همانطور که در شکل ۱.۱ مشخص است در صورتی که لبه‌ها در این سیاه‌نما مشخص شده بود، ابهامی مربوط به این که پای راست یا پای چپ جلو قرار دارد، رفع می‌شد. اما از مشکلات استخراج لبه‌ها این است که تصاویر افراد با لباس‌هایی با لبه‌های زیاد، منجر به استخراج لبه‌های نامربوط زیادی می‌شود.

بافت و رنگ

ویژگی رنگ و بافت برای مشخص کردن قسمت‌های گوناگون بدن مثل صورت و دست‌ها و پاها و بازوان مناسب است. این ویژگی با تغییر در مقیاس و جابجا شدن قسمت‌های بدن، با ثابت ماندن شرایط نورپردازی تغییر نمی‌کند. مقاله [۲۳] با پیدا کردن رنگ و بافت پوست در عکس، سر و دست و گاهی پاها را مشخص می‌کند. اما این روش هنگامی‌که برخی از اشیا هم‌رنگ با پوست است، دچار خطا می‌شود؛ به خصوص که برخی قسمت‌های اجزای بدن با اجسام همرنگ با پوست پوشانده شده باشد.

شکل ۱.۱: نمونه‌ای از ابهام در عمق [۳]

حرکت

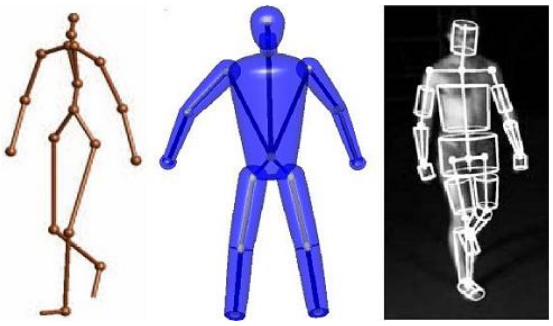
با فرض ثابت ماندن روشنایی پیکسل‌های قسمت‌های مختلف بدن و ثابت ماندن پیش‌زمنیه با تفاضل‌گیری از دو فریم متوالی، حرکت به‌دست آورده می‌شود. بسیاری از روش‌ها از حرکت برای تخمین دقیق‌تر حالت بدن استفاده کرده‌اند. حرکت مشخصه‌ای مناسب برای کاهش مشکل ابهام در عمق است. حرکت می‌تواند به صورت مستقیم در قالب مدل‌های مخفی مارکوف استفاده شود یا علاوه‌بر سایر مشخصه‌ها مانند سیاه‌نما برای افزایش اطلاعات مشخصه‌ی ورودی از آن بهره‌برداری شود [۵].

ترکیب ویژگی‌ها

ممکن است ترکیب مشخصه‌های ذکر شده، کارایی بهتری نسبت به استفاده از تنها یک مشخصه داشته باشد؛ به دلیل اینکه که اگر یک ویژگی در سناریویی خاص منجر به شکست شد می‌توان از اطلاعات ویژگی دیگر استفاده کرد. اغلب، ترکیب مشخصه‌ها در سیاه‌نماها کارایی را افزایش می‌دهد. یک راهکار ساده برای ترکیب ویژگی‌ها، اختصاص وزن‌هایی به هر ویژگی است. این وزن‌ها معمولاً به طور مکاشفه‌ای تنظیم می‌شود.

۲.۱.۱ مشخصه خروجی

برای تخمین حالت ۳ بعدی انسان بایستی مدلی برای توصیف حالت ۳ بعدی داشته باشیم. این مدل خروجی شامل پارامترهایی است که تعداد آنها با توجه به رویکرد اتخاذ شده و درجه‌ی نمایش جزییات از بدن فرد، قابل تغییر هستند. هر رویکرد بایستی مدل خروجی مناسبی به منظور مقایسه‌ی بصری و مقایسه‌ی عددی



شکل ۲.۱: مدل درختی حرکتی و مدل حجمی [۵]

داشته باشد.

معروف‌ترین مدل برای نمایش خروجی مساله، مدل درخت اسکلت حرکتی[۲] است که شامل اتصال اجزا گوناگون اسکلت بدن است (شکل ۲.۱). در کاربردهای رایج و به طور معمول این مدل با ۱۸ اتصال و هر اتصال با ۳ درجه‌ی آزادی درنظر گرفته می‌شود که هر درجه آزادی یک زاویه را مشخص می‌کند. بنابراین درخت اسکلت حرکتی در مجموع ۵۴ پارامتر دارد که بایستی توسط الگوریتم‌ها تخمین زده شود. البته با توجه به کاربرد و دقت مورد نیاز، تعداد پارامترها (زوایا) از ۳۰ تا ۶۰ نیز قابل تغییر است. اگرچه با افزایش درجه آزادی و تعداد اتصالات به واقعیت نزدیک‌تر می‌شویم اما بعد مساله و در نتیجه پیچیدگی افزایش یافته است. بنابراین بایستی مصالحه‌ای بین درجه‌ای از جزییات و پیچیدگی داشته باشیم.

رویکرد مولد بدون در نظر گرفتن توصیف سطح پایینی از تصویر، پارامترهای مدل خروجی را تخمین می‌زند و پس از آن با این پارامترهای بدست آمده، تصویر ۲ بعدی را دوباره تولید می‌کند. برای تولید تصویری شبیه به تصویر ورودی بایستی علاوه بر تخمین دقیقی از زوایای مفصلی مدل خروجی، شکل هر عضو از اندام فرد را در نظر گرفت. به همین منظور علاوه بر مدل کردن حالت ۳ بعدی بدن با زوایای مفاصل، از یک مدل حجمی نیز استفاده می‌شود. به عبارت دیگر این مدل علاوه بر تخمین زوایای مفاصل بدن در فضای ۳ بعدی، حجم هر عضو بدن را نیز تخمین می‌زند. برای نمایش حجم هر عضو از بدن فرد می‌توان از حجم‌های هندسی گوناگون استفاده کرد. در مقالات گوناگون از حجم‌های گوناگونی مثل استوانه [۱۴] و مخروط بیضوی [۴۴] برای مدل کردن شکل هر عضو بدن استفاده شده است. اما به علت متغیر بودن اندازه‌ی اعضای بدن هر فرد، این مدل‌ها نیز نمی‌توانند دقیقاً شکل عضو هر فرد را نشان دهند و بعلاوه برای تخمین اندازه‌ی هر عضو بدن، پیچیدگی مساله افزایش یافته است.

[۲] kinematic tree

۲.۱ اهمیت موضوع و کاربردها

همانطور که در بخش قبل ذکر شد، تخمین حالت ۳ بعدی انسان با استفاده از یک دوربین که یکی از هسته‌های اصلی یک سامانه‌ی هوشمند است، در بسیاری از زمینه‌ها از جمله تولید و ساخت بازی‌های رایانه‌ای و انیمیشن و تعامل انسان-رایانه [۳] کاربرد دارد. در این بخش قصد داریم برخی از کاربردها (شکل ۳.۱) را شرح دهیم.

- **تحلیل خودکار ویدیو** تحلیل ویدئوها هم در زمینه‌های امنیتی و هم در زمینه‌های پزشکی کاربرد دارند. تحلیل تعداد عظیمی از ویدئوها به طور سنتی کاری پرهزینه است و معمولاً به دلیل خطاهای انسانی، از دقت بالایی برخوردار نیست. در حالیکه با خودکار شدن تحلیل، هم هزینه‌ها کاهش و هم دقت افزایش می‌یابد. به طور مثال رفتارهای نامعمول و مشکوکی از قبیل نفوذ به مناطق امنیتی، بالا رفتن از دیوارهای حصار غیره به طور خودکار قابل کشف خواهند بود. مثال دیگر در زمینه‌ی پزشکی است؛ پزشک معالج با به کار بردن این سامانه‌ی هوشمند می‌تواند رفتار بیمار را از راه دور تحت کنترل داشته باشد و در صورت بروز رفتارهای نامعمول ناشی از سکته یا گرفتگی عضلات و حملات عصبی، او را به سرعت خبردار سازد و برای نجات بیمار اقدامات لازم صورت پذیرد.

- **تحلیل حرکت** تحلیل حرکت در زمینه‌ی فیزیوتراپی و ورزش به طور گسترده‌ای مورد نیاز است. به طور مثال تحلیل حرکات ورزشکاران که ترکیبی از چندین فعالیت با سرعت و شتاب متفاوت است، می‌تواند مربیان را برای ارزیابی و یا آموزش حرکات پیچیده‌تر یاری دهد.

- **سرگرمی** تولیدکنندگان انیمیشن و بازی‌های رایانه‌ای برای داشتن تولیداتی با کیفیت بالا، نیاز به شبیه‌سازی حرکات واقعی انسان دارند. برای تولید حرکات نزدیک‌تر به واقعیت در بازی‌های یارانه‌ای و تولید انیمیشن، دوربین‌های دیجیتال هوشمند که بتواند حالت ۳ بعدی انسان را به دقت تخمین بزند، به تولیدکنندگان سرگرمی‌ها نیز کمک می‌کند.

۳.۱ چالش‌ها

همان‌طور که در بالا گفته شد، محققان بینایی ماشین تلاش می‌کنند در محیط‌های واقعی با استفاده از یک تصویر یا دنباله‌ای از تصاویر، حالت شی موردنظرشان را در فضای ۳ بعدی بازسازی کنند. ساختن جسم ۳

[۳] Human-Computer Interaction

شکل ۳.۱: در حال حاضر با استفاده از نشانگرها و چندین دوربین در محیط‌های آزمایشگاهی، تخمین حالت ۳ بعدی انسان انجام می‌گیرد. شکل سمت راست و وسط: کاربرد آن را در ساخت و تولید بازی‌های رایانه‌ای و انیمیشن نشان می‌دهد. در شکل سمت چپ: برای بررسی کارایی و آموزش ورزشکاران از این سامانه استفاده شده است [۲۴].

بعدی در شرایط بدون قید و شرط (محیط‌های واقعی) با به کارگرفتن تنها یک دوربین محققان را با چالش‌هایی روبرو کرده است. در ادامه این بخش این چالش‌ها بیان می‌شود.

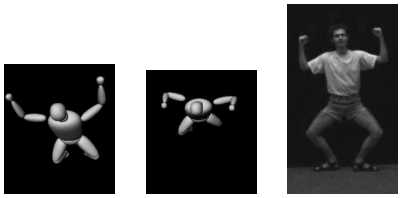
- **ابهام در عمق**

تولید عکس ۲بعدی از فضای ۳بعدی توسط دوربین‌ها به راحتی امکان پذیر است. ایجاد این فضای ۲بعدی تنها با تبدیل افکنش[۴] پرسپکتیو از فضای ۳ بعدی انجام می‌گیرد. تبدیل از فضای ۳ بعدی به فضای ۲ بعدی، تابعی یک به یک است که هر نقطه از فضای ۳ بعدی را تنها به‌یک نقطه در فضای ۲بعدی نگاشت می‌کند. اما نگاشت از فضای دو بعدی به فضای ۳ بعدی، تبدیلی یک به چند است و این امکان وجود دارد که به ازای نقطه‌ای در فضای ۲ بعدی، چندین نقطه در فضای ۳ بعدی موجود باشد. به طور مثال در شکل ۴.۱ برای دست‌های فرد در تصویر دو حالت متفاوت در فضای ۳ بعدی وجود دارد. . کمبود اطلاعات عمق، مسئله تخمین ۳ بعدی اشیاء را به مسئله‌ای بد تعریف [۵] تبدیل کرده است.

برای رفع چنین مشکلی می‌توان از دانش پیشین یا از یک مدل زمانی پویا که حالت قبلی را نیز در نظر دارد، استفاده کرد. در مدل‌های زمانی برای تخمین حالت ۳ بعدی بدن در زمان t، از حالت در زمان قبلی،t-1، کمک گرفته می‌شود. در کارهای دیگری [۳۴، ۴] برای رفع چنین مشکلی ایده‌ی به کار بردن چند تابع برازش برای تخمین حالت انسان ارائه شده است. در فصل ۲ این روش‌ها شرح داده شده است.

[۴] projection

[۵] ill-posed



(ج) (ب) (آ)

شکل ۴.۱: نمونه‌ای از ابهام در عمق. تصویر فرد در (آ) شبیه به تصاویر با دو حالت متفاوت (ب) و (ج) است [۲۳].

- **بعد بالای داده‌های خروجی**

بدن انسان دارای تعداد زیادی مفصل است که با اتصال این مفاصل کل بدن ساخته می‌شود. می‌توان برای ساخت اسکلت خروجی در فضای ۳ بعدی زاویه‌ی این مفاصل را در فضای ۳ بعدی تخمین زد. حداقل ۳۰ زاویه‌ی مفصلی در فضای ۳ بعدی باید تخمین زده شود تا بتوان به تقریب کلی از اسکلت ۳ بعدی دست یافت. در برخی از کاربردها که نیاز به جزییات بیشتری است، تعداد زاویای مفصلی به ۶۰ هم افزایش می‌یابد. جستجو در چنین فضایی برای یافتن جواب بهینه سراسری بسیار هزینه‌بر است. هم‌چنین برای مدل کردن این فضا به داده‌ی آموزشی زیادی نیاز است از طرفی مدل‌های مورد استفاده برای توصیف این فضا، پیچیدگی بالایی دارند.

- **انسداد**[۶]

در شرایط داشتن تنها یک دوربین تنظیم شده در یک جهت، ممکن است در تصویر قسمتی از بدن توسط محیط یا خود فرد پوشانده شود. این امر در بسیاری از فعالیت‌های روزمره‌ی افراد مثل راه رفتن و دویدن و غیره رخ می‌دهد. برای تخمین قسمت پوشیده شده، اطلاعاتی در دسترس نیست و نمی‌توان زوایای مفصلی این قسمت را به درستی تخمین زد. برای حل این مشکل از این واقعیت استفاده شده است که اجزای بدن به‌یکدیگر مرتبط هستند. در صورتی که‌یک قسمت از بدن را در تصویر نداشته باشیم می‌توان از رابطه‌ای که این جز با دیگر اجزا دارد، آن را تخمین زد. پس برای حل این مشکل مناسب است کواریانس بین اجزا را درنظر بگیریم.

بعلاوه انسان‌ها در پوشش‌های متفاوت ومتنوعی ظاهر می‌شوند؛ پوششی‌هایی که باعث پنهان ماندن قسمتهای از بدن می‌شود و مساله تخمین ۳ بعدی را دچار مشکل می‌سازند. همچنین پوشش‌هایی با رنگ و بافت‌هایی شبیه به رنگ و بافت پوست بدن انسان یا پس‌زمینه می‌تواند باعث خطاهایی در یافتن انسان در عکس شود.

[۶] Occlusion