

تخمین چنددوربینی حالت سه‌بعدی انسان با برازش افکنش مدل اسکلت سه‌بعدی مفصل‌دار

در تصاویر سایه‌نما

قاسم خادمی و حسین ابراهیم‌نژاد

آزمایشگاه تحقیقاتی بینایی کامپیوتر، دانشکده مهندسی برق، دانشگاه صنعتی سهند

چکیده

تحلیل و ضبط خودکار حرکت انسان بر اساس تصاویر یا ویدیو به دلیل کاربردهای بسیار زیاد در پویانمایی، نظارت، بیومکانیک، تعامل بین انسان و رایانه، صنعت بازی و سرگرمی اهمیت بسیار زیادی در بینایی رایانه دارد. در این کاربردها، تخمین حالت سه‌بعدی انسان از قسمت‌های اساسی است و به همین دلیل دقت تخمین بر عملکرد این کاربردها تأثیر زیادی دارد. تخمین حالت سه‌بعدی انسان از مشاهدات تصویری با توجه به بازه وسیع تغییرات در ظاهر و مفصل‌بندی انسان، ابعاد بالا در فضای حالت انسان و پدیده خودانسدادی، یک مسئله بحث‌برانگیز است. در این مقاله، روشی جدید برای تخمین حالت سه‌بعدی انسان در رشته ویدیویی چنددوربینی معرفی می‌شود. در روش پیشنهادی، به جای جستجوی مستقیم فضای حالت ابعاد بالای انسان و استفاده از الگوریتم‌های استنتاج پیچیده، از یک روش جستجوی سلسله‌مراتبی، توابع هدف جداگانه برای قسمت‌های مختلف بدن و روش‌های بهینه‌سازی مستقیم استفاده شده است. مزایای روش پیشنهادی، ارزش‌دهی اولیه خودکار، برچسب‌زنی قسمت‌های مختلف کانتور بدن و استفاده از توابع هدف جداگانه برای قسمت‌های مختلف بدن است. نتایج آزمایش، نشان می‌دهد که روش پیشنهادی می‌تواند به‌طور مؤثری به‌عنوان یک سامانه بدون نشانه برای تخمین حالت سه‌بعدی انسان در رشته ویدیویی چنددوربینی به کار گرفته شود.

واژگان کلیدی: تخمین سه‌بعدی حالت انسان، ضبط حرکت انسان، مدل مفصل‌دار، بهینه‌سازی، تصاویر سایه‌نما و اسکلت سه‌بعدی.

۱- مقدمه

تحلیل و ضبط خودکار حرکت انسان به دلیل پیچیدگی مسئله و کاربردهای بسیار زیادش، یکی از زمینه‌های تحقیقاتی بسیار فعال است. این زمینه تحقیقاتی شامل تعدادی از مسائل دشوار مانند استنتاج حالت^۱ و حرکت یک جسم سه‌بعدی غیرصلب^۲ مفصل‌دار^۳ با خاصیت خودانسدادی^۴ از روی تصاویر است. پیچیدگی این مسائل، تحقیقات در این زمینه را از نقطه نظر آکادمیک بحث‌برانگیز می‌سازد. تکنیک‌های تخمین حالت سه‌بعدی انسان^۵ که از سامانه‌های مبتنی بر بینایی^۶ استفاده می‌کنند، کاربردهای

بسیار زیادی دارند. تخمین سه‌بعدی حالت انسان از قسمت‌های اساسی در این کاربردها و دقت تخمین به‌طور مستقیم بر عملکرد آنها اثرگذار است. برای نمونه، در صنعت بازی‌های رایانه‌ای و سرگرمی، حالت سه‌بعدی تخمین زده شده انسان برای ایجاد پویانمایی‌های واقعی از حرکات انسان به کار می‌رود. در علوم ورزشی، بازسازی دقیق حالت سه‌بعدی انسان به ورزشکاران کمک می‌کند تا به‌صورت دیداری حرکات خودشان را برای بهبود عملکرد فعالیت‌شان تحلیل نمایند. در زمینه فیزیوتراپی، برای شناسایی دلایل اساسی وضعیت‌های نامطلوب حرکتی بیمار که ممکن است به علت ضربه، فلج مغزی یا دیگر مشکلات عصبی و عضلانی باشد، از تحلیل راه رفتن انسان استفاده می‌شود. در زمینه کاربرد نظارت ویدیویی^۷ با استفاده از دوربین‌های ارزان

^۱Pose

^۲Non-rigid

^۳Articulated

^۴Self-occluding

^۵3D human pose estimation

^۶Vision-based

^۷Video surveillance

قیمت و ایجاد بهبود در قدرت محاسباتی، تحلیل دقیق حالت سه‌بعدی انسان از ویدیو، می‌تواند اپراتورهای نظارت را برای شناسایی رویدادهایی از قبیل دویدن، راه رفتن، دزدی از مغازه‌ها، بالا رفتن از دیوار، پرسه‌زدن و دیگر فعالیت‌های غیرطبیعی انسان کمک کند.

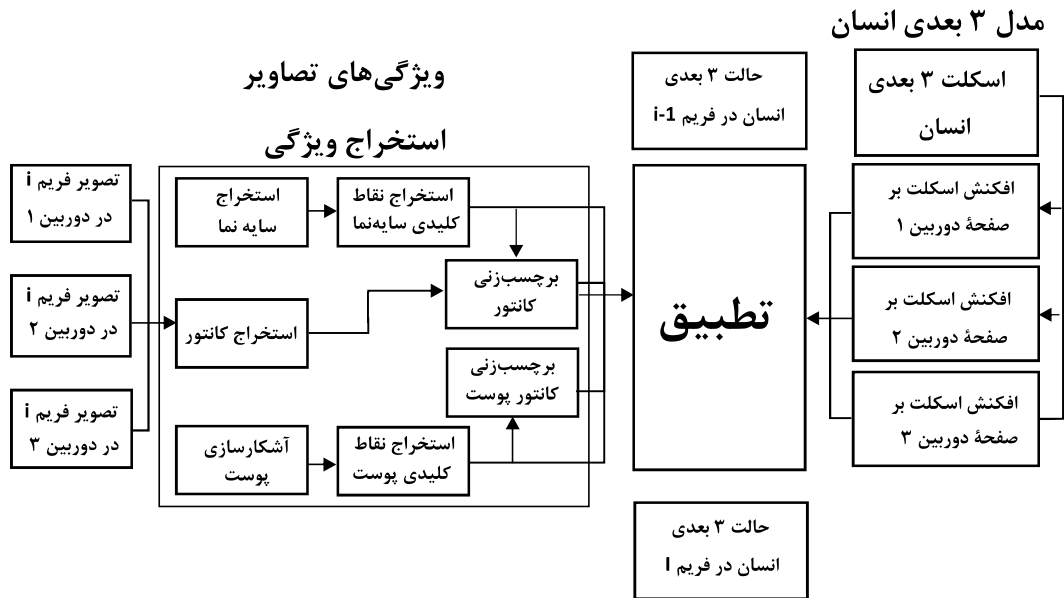
در این مقاله، یک روش خودکار برای تخمین و ردیابی حالت بدن انسان در فضای سه‌بعدی با استفاده از سه دوربین معرفی می‌شود. از یک مدل اسکلتی سینماتیک، برای مدل‌سازی بدن انسان استفاده می‌شود. این مدل به صورت پانزده نقطه در فضای سه‌بعدی که معرف مفاصل کلیدی بدن انسان و چهارده خط که معرف استخوان‌ها و یا قسمت‌های صلب بدن انسان هستند، در نظر گرفته می‌شود. تخمین حالت بدن انسان در هر فریم با استفاده از تطبیق اسکلت سه‌بعدی بر تصاویر سه دوربین از نماهای مختلف انجام می‌شود. پیکربندی اسکلت سه‌بعدی بهینه زمانی حاصل می‌شود که افکنش خطوط اسکلت سه‌بعدی در سه نما از دوربین‌های مختلف در وسط مرز عضوهای متناظر بدن انسان قرار گیرد. در واقع جستجوی زوایای اتصال مفاصل با این هدف انجام می‌شود که در تمام نماها تفاضل فاصله‌های افکنش خطوط اسکلت از مرز سایه‌نمای^۱ بدن کمینه شود. با استفاده از این ایده می‌توان درجه‌های آزادی هر یک از اعضای بدن را (زوایای اتصال مفاصل) در فضای سه‌بعدی تخمین زد. یکی از چالش‌های اصلی در تخمین و ردیابی حالت انسان مواجه شدن با یک فضای جستجوی ابعاد بالاست. از این رو بازیابی حالت کامل بدن انسان به‌طور مستقیم دشوار، اما محاسبه موقعیت هر یک از اعضای بدن به‌تنهایی راحت‌تر است. (شکل ۱) بلوک دیاگرام روش پیشنهادی را نشان می‌دهد. در روش پیشنهادی، حالت سه‌بعدی انسان از طریق تطبیق ویژگی‌های تصاویر سه دوربین مختلف و مدل اسکلتی سه‌بعدی استنتاج می‌شود. ابتدا ویژگی‌های تصاویر در هر نما استخراج و سپس ویژگی‌های استخراج شده از هر تصویر دوربین با افکنش مدل سه‌بعدی انسان بر آن صفحه تصویر از طریق یک تابع هدف پیشنهادی تطبیق داده می‌شوند. همان‌طور که در (شکل ۱) مشاهده می‌شود، ابتدا سایه‌نما، کانتور و نواحی پوستی بدن انسان از تصاویر سه دوربین مختلف در فریم‌آم استخراج می‌شوند. چند نقطه کلیدی از سایه‌نما به‌عنوان تخمین اولیه از محل ریشه بدن، مفاصل قوزک و زانوی پاها

استخراج می‌شوند. همچنین با استفاده از نواحی پوستی آشکارسازی شده چند نقطه کلیدی دیگر به‌عنوان یک تخمین اولیه مناسب از موقعیت سر و دست‌ها استخراج می‌شوند. با استفاده از نقاط کلیدی دوبعدی استخراج شده در صفحات تصویر می‌توان موقعیت‌های سه‌بعدی اولیه سر، گردن، دست‌ها و مفاصل را به‌دست آورد که در هم‌گرایی الگوریتم به پیکربندی حالت صحیح مؤثرند. در ادامه، با استفاده از روش پیشنهادی برجسب‌زنی کانتور، پیکسل‌های کانتور و نواحی پوست اعضای مختلف بدن از یکدیگر تفکیک و به‌عنوان ورودی تابع هدف پیشنهادی به‌منظور تخمین حالت سه‌بعدی انسان استفاده می‌شوند. همچنین استفاده از یک روش جستجوی سلسله‌مراتبی برای مهارکردن جستجو در این فضای ابعاد بالا پیشنهاد می‌شود؛ به‌طوری‌که در هر مرحله فقط پارامترهای مربوط به یک استخوان از بدن انسان با توجه به ترتیب سلسله‌مراتبی در درخت سینماتیک بدن انسان بهینه می‌شوند. با توجه به استفاده از روش‌های بهینه‌سازی مستقیم برای جستجوی درجه‌های آزادی هر عضو، مواجه‌شدن با کمینه‌های محلی یک مسئله، بحث‌برانگیز است. برای حل این مسئله در روش پیشنهادی، با استفاده از ویژگی‌های استخراج شده در هر فریم و اطلاعات فریم‌های قبلی، ارزش‌دهی اولیه مناسب برای پارامترها انجام شد و با به‌کارگیری روش‌های جستجوی محلی در اطراف مقادیر اولیه پارامترها، پارامترهای مدل بهینه می‌شوند. روش پیشنهادی نیازمند یک تابع هدف برای بهینه‌سازی هر عضو بدن است. این تابع هدف بر مبنای تفاضل فاصله از مرز کانتور هر عضو فرمول‌بندی می‌شود.

۲- کارهای مرتبط

در سال‌های اخیر کارهای عمده‌ای برای بازیابی حالت کامل بدن انسان از تصاویر و ویدیوهای چنددوربینی انجام شده است. در این بخش، روش‌هایی که مربوط به استنتاج سه‌بعدی چنددوربینی حالت انسان هستند به‌صورت کلی توصیف می‌شوند.

¹ Silhouette



(شکل ۱): نمودار جعبه‌ای روش پیشنهادی

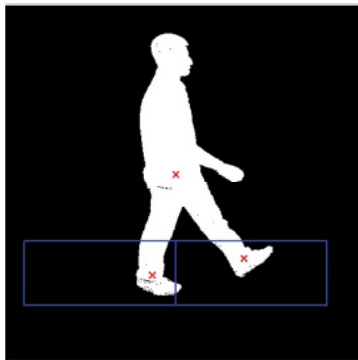
پوسته قابل مشاهده این است که بر جداکردن پس زمینه تقریباً کامل تکیه دارند. سایه‌نماهای نوفه‌ای حتی فقط از یک دوربین باعث ایجاد سوراخ‌هایی در حجم سه‌بعدی بازسازی شده می‌شود که به‌طور قابل ملاحظه‌ای بیان حجمی را نادرست می‌سازد. برای مقابله با این مشکل، روشی به نام شبکه اشغال احتمالاتی^۲ در (Franco and Boyer 2005) معرفی شده است که در آن پوسته قابل مشاهده معادل با در نظر گرفتن سطح همسان^۳ از چگالی در یک احتمال معین به دست می‌آید. هنگامی که حجم بازبایی شد، ردیابی شکل سه‌بعدی با استفاده از روش بهینه‌سازی تصادفی کاهشی متا^۴ (Kehl, Bray et al. 2005) یا الگوریتم نزدیک‌ترین نقطه تکراری^۵ (Mündermann, Corazza et al. 2006) انجام می‌شود.

روش‌های مبتنی بر افکنش مدل (Gavrila and Davis 1996; Deutscher, Blake et al. 2000; Sigal, Bhatia et al. 2004) از افکنش مدل بر تصاویر نماهای مختلف بدون در نظر گرفتن اجتماع مشاهدات سطح پایین تصویر در تمام نماها استفاده می‌کنند. همچنین، این روش‌ها برخلاف روش‌های مبتنی بر پوسته قابل مشاهده، نیازمند استفاده از داده‌های سایه‌نمای به‌طور تقریبی کامل نیستند. در این روش‌ها، نماهای مختلف با استفاده از یک تابع همانندی، مورد بررسی قرار می‌گیرند که در آنها اغلب، استقلال در تمام نماها فرض می‌شود (Deutscher, Blake et

al. 2003). اکثر روش‌هایی را که با چند دوربین سروکار دارند، می‌توان به دو دسته کلی تقسیم‌بندی کرد: یک دسته روش‌هایی که به‌صراحت از پوسته قابل مشاهده^۱ استفاده و دسته دیگر روش‌هایی که بدون بازسازی صریح بیان حجمی بدن انسان از افکنش یک مدل سه‌بعدی بر تصاویر استفاده می‌کنند. در هر دو مورد ذکر شده، معلوم‌بودن پارامترهای دوربین (هر دوی پارامترهای داخلی و خارجی) به منظور در نظر گرفتن ارتباط بین اطلاعات در دوربین‌های مختلف ضروری است.

روش‌های مبتنی بر پوسته قابل مشاهده به‌صراحت از ارتباط ویژگی‌های نماهای مختلف برای بازسازی هندسه محدود سه‌بعدی تقریبی جسم واقعی استفاده می‌کنند. زمانی که تعداد نماها افزایش پیدا می‌کند، می‌توان نشان داد که پوسته قابل مشاهده، به‌شکل واقعی جسم نزدیک می‌شود. اکثر روش‌های مبتنی بر پوسته قابل مشاهده (Cheung, Baker et al. 2003; Kehl, Bray et al. 2005) یک فرآیند تفریق پس‌زمینه خوب و سایه‌نماها به‌منظور تعریف مخروط سایه‌نما تکیه می‌کنند. مخروط‌های سایه‌نما با استفاده از مراکز دوربین‌ها و سایه‌نما بوجود می‌آیند. اشتراک مخروط‌های دوربین‌های مختلف، کران بالای فضای اشغال‌شده توسط جسم را تعریف می‌کند. همچنین روش‌هایی وجود دارند که با استفاده از رنگ‌آمیزی وکسل (Cheung, Baker et al. 2003) ثبات رنگ را نیز در تمام نماها در نظر می‌گیرند. مشکل اصلی روش‌های مبتنی بر

²Probabilistic occupancy grid³Isosurface⁴Stochastic Meta Descent (SMD)⁵Iterative Closest Point (ICP)¹Visual hull



(شکل ۲): نمایش نقاط کلیدی سایه‌نما

در روش آنها شامل چهارده نقطه کلیدی بدن انسان است که باتوجه به شباهت زیاد مدل اسکلت آنها با مدل اسکلت پیشنهادی در این مقاله، می‌توان نتایج الگوریتم پیشنهادی این مقاله را با روش آنها به صورت کمی مقایسه کرد.

۳- استخراج ویژگی

آشکارسازی قسمت‌های مختلف بدن، یکی از روش‌های رایج برای راه‌اندازی خودکار تخمین حالت انسان است (Sigal, Bhatia et al. 2004; Lee and Nevatia 2009) و برای کاربردهایی که نیاز به ارزش‌دهی اولیه خودکار دارند، مهم است. در روش پیشنهادی، با استفاده از استخراج ویژگی از سایه‌نما و آشکارسازی پوست، نامزدهایی برای موقعیت مفاصل بدن مانند سر، گردن، آرنج و زانوها تخمین زده می‌شود. بنابراین استخراج ویژگی‌هایی که تخمین اولیه مناسب از بعضی از مفاصل بدن ایجاد کنند ضروری است. همچنین با استفاده از روش پیشنهادی برچسب‌زنی کانتور، پیکسل‌های کانتور مربوط به قسمت‌های مختلف بدن انسان برچسب زده می‌شوند. نتایج به‌دست آمده از مرحله استخراج ویژگی به‌عنوان ورودی الگوریتم بهینه‌سازی پارامترهای حالت انسان مورد استفاده قرار می‌گیرند. در ادامه مراحل استخراج ویژگی‌های مختلف در بخش‌های جداگانه توضیح داده می‌شود.

۳-۱- استخراج سایه‌نما و نقاط کلیدی

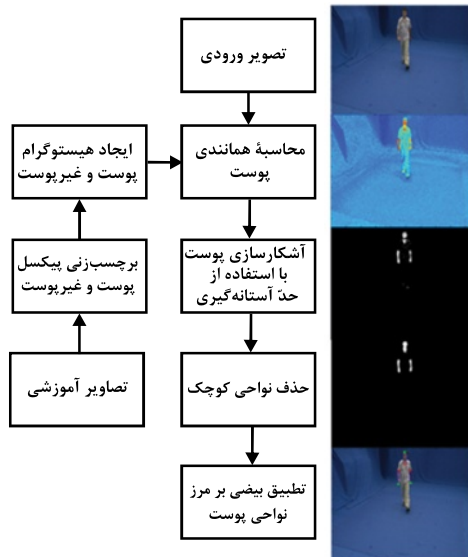
سایه‌نما یکی از ویژگی‌های رایج به‌کار گرفته شده در تخمین حالت انسان است. با استخراج سایه‌نما، بدن انسان در صفحه تصویر مکان‌یابی می‌شود که باعث کاهش فضای جستجوی الگوریتم برای پیکربندی صحیح بدن می‌شود. تفریق

ترکیب همانندی‌های مبتنی بر هر نما می‌تواند یک معیار کلی را برای تطبیق حالت استنتاج نمایند. در روش پیشنهادی نیز از افکنش مدل اسکلت سه‌بعدی انسان بر تصاویر سه‌نمای دوربین‌ها استفاده شده است. تابع هدف برای هر نمای دوربین به‌صورت مستقل تعریف شده است و تطبیق حالت سه‌بعدی بر مبنای ترکیب توابع هدف تعریف‌شده انجام شده است. ترکیب توابع هدف به‌صورت مجموع توابع هدف سه‌نمای دوربین‌ها در نظر گرفته شده است و تطبیق حالت سه‌بعدی انسان با ویژگی‌های تصاویر سه‌دوربین با استفاده از کمینه‌سازی تابع هدف ترکیب‌شده انجام می‌شود.

در (Hofmann and Gavrilu 2012) یک روش برای تخمین حالت قسمت بالای بدن انسان از روی تصاویر چند دوربین بر اساس ترکیب اطلاعات شکل و بافت ارائه شده است. در این روش، جهت‌بازیابی حالت شخص در یک فریم مشخص، از یک مرحله تولید فرض که در آن حالت‌های سه‌بعدی نامزد بر اساس مقایسه سلسله‌مراتبی شکل در هر دوربین ایجاد می‌شود، استفاده شده است. سپس در مرحله تأیید فرض، حالات سه‌بعدی نامزد به نماهای دوربین‌های دیگر باز افکنده شده و بر طبق یک اندازه‌شباهت، رتبه‌بندی می‌شوند. در ادامه، فریم‌هایی که تخمین حالت برای آنها با دقت خوبی انجام گرفته است، به تعداد مشخصی انتخاب شده و از آنها برای تولید یک مدل بافتی استفاده می‌گردد.

در (Zhu, Dariush et al. 2010) یک روش مبتنی بر مدل برای تخمین حالت سه‌بعدی انسان با استفاده از شناسایی مجموعه‌ای از نقاط کلیدی بدن انسان از روی تصاویر عمق به‌دست‌آمده از یک تصویربرداری زمان‌رفت و برگشت (TOF)^۱ معرفی شده است. سامانه پیشنهادی آنها در سه مرحله اصلی انجام می‌شود. در مرحله اول با استفاده از یک آشکارساز مبتنی بر چارچوب بی‌زین که از اطلاعات مکانی و زمانی در یک رشته ویدیویی استفاده می‌کند، نقاط کلیدی بدن انسان در فضای دوبعدی شناسایی و ردیابی می‌شوند. در مرحله دوم با استفاده از قیدهای حرکتی بدن انسان، نقاط دوبعدی شناسایی‌شده به مدل سه‌بعدی حرکتی انسان نگاشت می‌شوند. در مرحله سوم نیز نقاطی که توسط آشکارساز دوبعدی به علت انسداد شناسایی نشده‌اند، تخمین زده شده و ابهاماتی که در مرحله شناسایی نقاط کلیدی وجود دارد از بین می‌رود. مدل اسکلت سه‌بعدی

^۱ Time-of-flight (TOF)



(شکل ۳): آشکارسازی پوست

۳-۲- آشکارسازی پوست

یکی از فرض‌های رایج این است که بعضی از قسمت‌های بدن انسان توسط لباس پوشش داده نمی‌شود، که در این صورت می‌توان از رنگ پوست به‌عنوان یک ویژگی تشخیص برای این قسمت‌ها استفاده کرد. ویژگی‌های رنگ پوست سیگنال‌های مهمی دربارهٔ موقعیت سر، بازوها و بعضی از اوقات پاها می‌دهند. برای آشکارسازی پیکسل‌های پوست از بخشی از روش مرجع (Conaire, O'Connor et al. 2007) استفاده شده است. مراحل استخراج ویژگی‌های مبتنی بر پوست در (شکل ۳) نشان داده شده است. روش کار بدین‌صورت است که ابتدا مجموعه‌ای از تصاویر آموزشی در نظر گرفته می‌شود، سپس در این تصاویر آموزشی پیکسل‌های پوست و غیرپوست، به‌صورت دستی برچسب‌زنی می‌شوند. با استفاده از پیکسل‌های برچسب زده شده یک هیستوگرام RGB برای پیکسل‌های پوست و هیستوگرام دیگری برای پیکسل‌های غیرپوست ایجاد می‌شود که ابعاد این هیستوگرام‌ها $32 \times 32 \times 32$ هستند. عدد ۳۲ در هر محور بیان‌گر تعداد سطوح کوانتیزاسیون در رنگ اصلی متناظر است. همان‌طور که می‌دانیم هر تصویر رنگی دارای سه مؤلفه R، G و B برای هر پیکسل است. هر یک از مقادیر R، G و B دارای گستره ۰ تا ۲۵۵ هستند. یعنی سطح کوانتیزاسیون داریم. به‌منظور کاهش حجم محاسبات از ۳۲ سطح کوانتیزاسیون استفاده می‌شود که دقت خوبی برای تشخیص رنگ پوست از غیر پوست دارد.

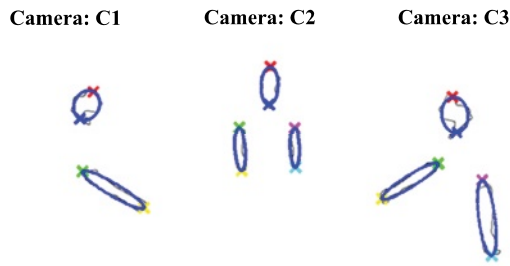
پس‌زمینه^۱ فرآیندی است که اغلب برای به‌دست آوردن سایه‌نما مورد استفاده قرار می‌گیرد. در مواردی که پس‌زمینه دارای یک رنگ ثابت است، مانند پایگاه داده iD3post (Gkalelis, Kim et al. 2009) که دارای پس‌زمینه ثابت آبی رنگ است، می‌توان با استفاده از یک حد آستانه‌گیری ساده، سایه‌نمای انسان از پس‌زمینه را به‌راحتی استخراج کرد.

در صورت استخراج یک تصویر سایه‌نمای کامل و بدون نوفه، از طریق پردازش این داده‌های باینری، می‌توان ویژگی‌هایی برای تخمین اولیه بعضی از مفاصل بدن استخراج کرد. البته، در مواردی که صحنه یا پس‌زمینه شلوغ باشد، اجسام متحرک دیگری در صحنه حضور داشته باشند، یا تغییرات نوری صحنه زیاد باشد و انسان حاضر در صحنه با اجسام متحرک دیگر دچار انسداد شود، استخراج یک سایه‌نمای کامل و بدون نوفه در صحنه دشوار خواهد بود. در این گونه موارد ممکن است حفره‌هایی در سایه‌نما ایجاد و یا اینکه به‌علت وجود مانع، گسستگی در سایه‌نما ایجاد شود که حتی با استفاده از عملگرهای مورفولوژی نیز سایه‌نمای تمیز و عاری از نوفه به‌دست نیاید. اولین ویژگی، مرکز ثقل^۲، سایه‌نمای باینری است. این ویژگی اطلاعاتی در مورد کلیت انسان که چگونه در زمان جابه‌جا می‌شود و همچنین یک تقریب مکانی از مفصل ران^۳ می‌دهد که با استفاده از آن می‌توان محل قسمت‌های دیگر بدن در تصویر باینری را نیز تخمین زد. تقریب مرکز ثقل باعث کاهش پیچیدگی محاسباتی در تخمین اولیه محل اسکلت بدن انسان می‌شود. کل سطح سایه‌نمای ایجاد شده از فرآیند تفریق پس‌زمینه برای محاسبه مرکز ثقل مورد استفاده قرار می‌گیرد. مقدار متوسط تمام موقعیت‌های پس‌زمینه در محور افقی و عمودی تصویر، معرف مکان مرکز ثقل سایه‌نما در تصویر است. مرحله بعدی پیدا کردن موقعیت پا در تصویر سایه‌نماست. مرکز ثقل یک پا، تقریبی مکانی از موقعیت قوزک پا در تصویر سایه‌نماست. موقعیت قوزک پا در روشی به‌طور کامل مشابه با محاسبه مرکز ثقل سایه‌نما به‌دست می‌آید، با این تفاوت که سطحی که برای محاسبه مرکز ثقل به‌کار می‌رود، باید محدود شود. نواحی در نظر گرفته شده به‌صورت مستطیل و نقاط کلیدی با علامت ضربدر در (شکل ۲) نشان داده شده‌اند.

^۱Background subtraction

^۲Centroid

^۳Hip



(شکل ۴): نمایش بیضی‌های پوست و نقاط نامزد محل سر، پایین گردن، آرنج‌ها و دست‌ها در ۳ نمای دوربین متفاوت

ضرب‌درهای قرمز و آبی نشان داده شده‌اند. مرحله آشکارسازی پوست به‌طور خودکار قادر به استخراج این اطلاعات است. از روی اطلاعات به‌دست آمده به‌عنوان مثال می‌توان تفسیر کرد که دست چپ در دوربین شماره ۱ توسط نیم‌تنه دچار انسداد شده است و دیده نمی‌شود، بنابراین در فرآیند بهینه‌سازی حالت دست چپ فقط از اطلاعات دوربین شماره ۲ و ۳ استفاده می‌شود.

۳-۳- برچسب‌زنی کانتور با استفاده از گراف

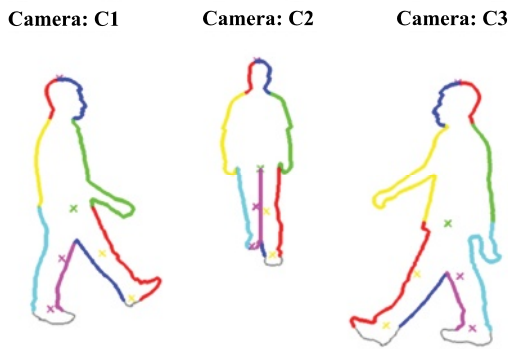
کانتور انسان، لبه کل یا طرح کلی انسان را ارائه می‌دهد؛ اما اینکه چگونه مرزهای قسمت‌های مختلف بدن انسان در یک کانتور از هم تفکیک شوند، یک مسئله بحث‌برانگیز است. در این قسمت روشی با استفاده از تئوری گراف، معرفی می‌شود که می‌تواند پیکسل‌های قسمت‌های مختلف کانتور انسان را از هم تفکیک کند. طرح کلی مسئله از اینجا آغاز می‌شود که چگونه می‌توان روی پیکسل‌های لبه یا کانتور یک شیء از یک مبدأ به یک مقصد حرکت کرد، علاوه‌بر این فرض می‌شود که مسیر حرکت، کوتاه‌ترین مسیر ممکن باشد. این مسئله به شکل‌های مختلف در زمینه‌های دیگر ظاهر شده است و راه‌حل‌های قوی با استفاده از تئوری گراف برای آنها پیشنهاد شده است. برای نمونه در مسیریابی بسته‌ها در اینترنت، طراحی شبکه تلفن، موتورهای جستجوی وب، برنامه‌ریزی خودکار مسیر سفر و موارد دیگر از تئوری گراف برای مدل‌سازی آنها استفاده شده است. در این بخش از یک مدل گراف برای مدل‌سازی پیکسل‌های کانتور انسان استفاده می‌شود و با استفاده از روش پیدا کردن کوتاه‌ترین مسیر در گراف مربوطه، مسئله برچسب‌زنی قسمت‌های مختلف کانتور انسان حل می‌شود. بنابراین نخست، تعاریفی در مورد تئوری گراف بیان می‌شود. در اصطلاحات علمی تئوری گراف،

برای هر پیکسل در موقعیت (i,j) با مقادیر رنگ‌های اصلی (r,g,b) میزان همانندی به پوست با استفاده از رابطه (۱) محاسبه می‌شود:

$$L^{(i,j)}(r,g,b) = \log \left(\frac{H^{(i,j)}(r,g,b)}{h^{(i,j)}(r,g,b)} \right) \quad (1)$$

که در این رابطه، H هیستوگرام پوست و h هیستوگرام غیر پوست و L نیز میزان همانندی به پوست برای پیکسل مورد نظر است. برای یک تصویر جدید، میزان شباهت به پوست برای پیکسل‌های تصویر محاسبه می‌شوند و با استفاده از یک حد‌آستانه‌گیری می‌توان تصمیم به پوست و غیر پوست بودن پیکسل مورد نظر گرفت. پس از حد‌آستانه‌گیری، نواحی کوچک که به‌اشتباه به‌عنوان پوست شناخته شده‌اند حذف و نواحی پوست باقیمانده با استفاده از عملگرهای شکل‌شناسی بهبود داده می‌شوند. یک بیضی بر مرز هر یک از نواحی که به‌عنوان پیکسل‌های پوست شناخته شده‌اند، تطبیق داده می‌شود. برای این کار از روش تطبیق حداقل مربعات مستقیم^۱ استفاده شده است، برای جزئیات کامل این روش می‌توان به مراجع (Pilu, Fitzgibbon et al. 1999; Fitzgibbon, Pilu et al. 1996) مراجعه کرد. بیضی‌های پوست ایجاد شده برای استنتاج موقعیت سر و سایر اعضای بدن به‌کار می‌روند. البته تفسیر بیضی‌های پوست به نوع لباس انسان وابستگی دارد. به‌عنوان مثال، اگر شخص لباس آستین کوتاه پوشیده باشد، بیضی پوست بیان‌گر ساعد دست بوده که نشان‌دهنده موقعیت‌های آرنج و دست است. در حالی که برای لباس آستین بلند، بیضی پوست فقط دست انسان را پوشش می‌دهد و فقط برای استنتاج موقعیت دست به‌کار می‌رود. بنابراین با استخراج بیضی‌های پوست مجموعه‌های مختلفی از تفاسیر با توجه به فرضیات نوع لباس پوشیدن فراهم می‌شود. برای هر یک از بیضی‌های پوست، دو نقطه انتهایی بیضی در امتداد محور اصلی استخراج می‌شود، این نقاط برای بیضی صورت معرف نامزدهای محل سر و پایین گردن هستند. همچنین برای بیضی‌های دست‌ها، اگر پوشش لباس شخص از نوع آستین کوتاه باشد، این نقاط به‌عنوان نامزدهای قابل قبول برای محل دست و آرنج هر یک از دست‌های چپ یا راست در نظر گرفته می‌شوند. در (شکل ۴) نقاط نامزد دست چپ با ضرب‌درهای فیروزه‌ای و سرخ‌آبی، نقاط نامزد دست راست با ضرب‌درهای سبز و زرد و نقاط نامزد سر و گردن با

^۱Direct least-square fitting



(شکل ۵): نتایج پیاده سازی روش برجسبزی کانتور بدن انسان در ۳ نمای دوربین متفاوت

پیکسل یا برجسبزه شده اند که با تعریف یک تابع هدف مناسب که در بخش های بعدی ذکر می شود، می توان نشان داد که اثر این پیکسل ها در محاسبه تابع هدف بدون تأثیرند. برجسبزی پیکسل های دست نیز به طور جداگانه بر اساس پیکسل های پوست انجام می شود.

۴- مدل سازی بدن انسان

در روش پیشنهادی این مقاله، از یک مدل اسکلتی سینماتیک برای مدل سازی بدن انسان استفاده می شود. این مدل به صورت پانزده نقطه در فضای سه بعدی که معرف مفاصل کلیدی بدن انسان و چهارده خط که معرف استخوان ها و یا قسمت های صلب بدن انسان هستند، در نظر گرفته می شود. (شکل ۶) مدل در نظر گرفته شده برای بدن انسان را نشان می دهد، مدل سه بعدی اسکلتی پیشنهادی به صورت مجموعه ای از پانزده نقطه در فضای سه بعدی در نظر گرفته شده است این نقاط در واقع معرف مفاصل کلیدی بدن انسان می باشند که در (شکل ۶) به صورت P1 تا P15 نشان داده شده اند. این مدل شامل چهارده خط در فضای سه بعدی می باشد. هر خط نماینده یکی از استخوان های بدن انسان است و هر کدام از آنها تعدادی درجه آزادی دارند که جزئیات آن ها در (جدول ۱) آمده است. در واقع، موقعیت هر جسم صلب در فضای سه بعدی و سامانه مختصات جهانی توسط شش پارامتر مشخص می شود که سه پارامتر X, Y, Z انتقال مرکز ثقل جسم نسبت به مبدأ مختصات را نشان می دهد و پارامترهای R_x, R_y, R_z نشان دهنده میزان چرخش جسم نسبت به محورهای مختصات x, y, z هستند که در این مقاله از بیان زوایای اوپلر به منظور نشان دادن چرخش یک جسم استفاده شده است. همچنین برای هر قسمت صلب از

مدل سازی به صورت مجموعه ای از رأس ها^۱ (V) و لبه ها^۲ (E) در نظر گرفته می شود که لبه ها معرف اتصالات بین رأس ها هستند. بنابراین یک گراف $G = (V, E)$ به صورت نشان داده می شود. یک مسیر، یک رشته از رأس ها $\langle v_0, v_1, \dots, v_k \rangle$ در گراف $G = (V, E)$ است؛ در صورتی که هر یک از لبه های (v_i, v_{i+1}) در مجموعه E قرار داشته باشند (یعنی در رشته مربوطه هر رأس به رأس بعدی متصل باشد). در مسئله کوتاه ترین مسیر، به هر لبه (u, v) یک وزن $w(u, v)$ داده می شود. وزن یک مسیر (یا طول مسیر) برابر مجموع وزن های لبه های مسیر می باشد:

$$w(p) = \sum_{i=0}^{k-1} w(v_i, v_{i+1}) \quad (2)$$

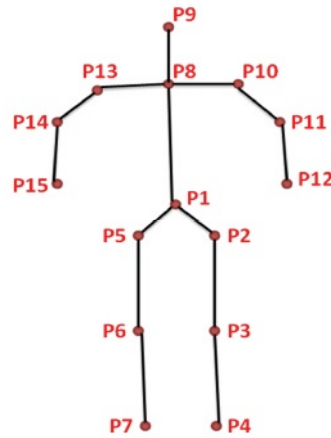
کوتاه ترین مسیر از رأس u به رأس v مسیری است که دارای کمینه ترین وزن در بین تمام مسیرهای ممکن باشد. برای یادگیری الگوریتم های تئوری گراف و نحوه پیاده سازی آنها می توان به مرجع (Siek, Lee et al. 2002) مراجعه کرد. برای مدل سازی تصویر باینری کانتور انسان با استفاده از یک گراف، ابتدا باید رأس ها، لبه ها و وزن بین رأس ها تعریف شوند. پیکسل هایی که در تصویر متعلق به کانتور انسان هستند (یعنی مقدار یک دارند) به عنوان رأس های گراف در نظر گرفته می شوند. اتصال بین رأس ها نیز با توجه به اتصال هشت گانه پیکسل متناظرشان در نظر گرفته می شود؛ بنابراین فقط رأس هایی که پیکسل متناظرشان در اتصال هشت گانه یکدیگر قرار دارند، به هم متصل هستند. اختصاص وزن بین رأس های متصل، به صورت فاصله اقلیدسی پیکسل متناظرشان تعریف می شود. بنابراین، با استفاده از مدل سازی تصویر کانتور انسان به وسیله یک گراف و مشخص کردن نقاط مبدأ و مقصد، می توان عملیات برجسبزی بین نقاط مبدأ و مقصد مشخص شده را به راحتی انجام داد. (شکل ۵) نتایج پیاده سازی روش پیشنهادی را در سه نمای متفاوت از دوربین ها نشان می دهد. در این شکل مرزهای قسمت های مختلف بدن مانند نیم تنه، سر و پاها با رنگ های مختلف نشان داده شده اند، پیکسل های برجسبزه شده در این مرحله به عنوان ورودی الگوریتم بهینه سازی حالت برای محاسبه تابع هدف در نظر گرفته می شوند. همان طور که در (شکل ۵) مشخص است پیکسل های کانتور دست به عنوان پیکسل نیم تنه و پا در بعضی موارد به عنوان

¹Vertex (V)

²Edges (E)

(جدول ۱): درجه‌های آزادی استخوان‌های مدل اسکلت

پیشنهادی	
درجه آزادی (DOF)	استخوان
$X, Y, Z, R_{X1}, R_{Y1}, R_{Z1}$	نیم‌تنه (P1-P8)
ندارد	مفصل ران چپ (P1-P2)
R_{X3}, R_{Y3}, R_{Z3}	استخوان ران چپ (P2-P3)
R_{X4}	ساق پای چپ (P3-P4)
ندارد	مفصل ران راست (P1-P5)
R_{X5}, R_{Y5}, R_{Z5}	استخوان ران راست (P5-P6)
R_{X7}	ساق پای راست (P6-P7)
R_{X9}, R_{Y9}, R_{Z9}	سر (P8-P9)
R_{Y10}, R_{Z10}	ترقوه چپ (P8-P10)
$R_{X11}, R_{Y11}, R_{Z11}$	بازوی چپ (P10-P11)
R_{X12}	ساعد چپ (P11-P12)
R_{Y13}, R_{Z13}	ترقوه راست (P8-P13)
$R_{X14}, R_{Y14}, R_{Z14}$	بازوی راست (P13-P14)
R_{X15}	ساعد راست (P14-P15)
29	تعداد کل (DOF)



(شکل ۶): مدل بدن انسان

مدل مفصل‌دار انسان یک اندیس در نظر گرفته شده است. به‌طور مثال، منظور از R_{X1} زاویه چرخش خط P1-P8 (نیم‌تنه) نسبت به محور X مختصات جهانی است.

۵- تطبیق مدل با ویژگی‌های تصاویر

در روش پیشنهادی، حالت سه‌بعدی انسان از طریق تطبیق ویژگی‌های تصاویر سه‌دوره‌بین مختلف و مدل اسکلتی سه‌بعدی استنتاج می‌شود. بنابراین، پس از استخراج ویژگی‌های مناسب، این ویژگی‌های از هر تصویر دوره‌بین با افکنش مدل سه‌بعدی انسان بر آن صفحه تصویر از طریق کمینه‌سازی تابع هدف پیشنهادی تطبیق داده می‌شوند.

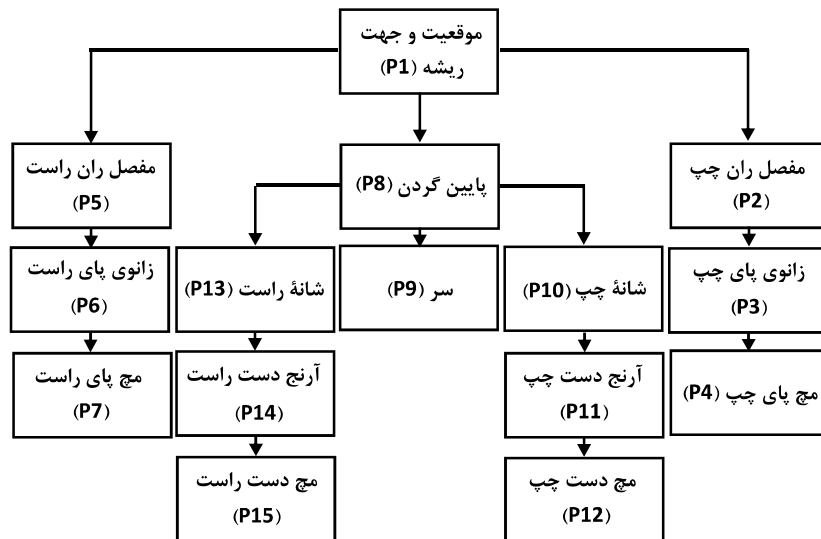
۵-۱- جستجوی سلسله‌مراتبی

جستجوی پیکربندی صحیح حالت سه‌بعدی مفصل‌دار انسان در یک فضای جستجوی ۲۹ بعدی از نظر محاسباتی پرهزینه و اغلب مهارنشده است. خوشبختانه، استفاده از خاصیت سلسله‌مراتبی در ساختار سینماتیک بدن انسان میسر می‌سازد که جستجو در یک توالی از مراحل انجام شود؛ طوری که در هر مرحله فقط یک زیرمجموعه از ۲۹ پارامتر بهینه شوند. این خاصیت سلسله‌مراتبی به‌صورت یک درخت سینماتیک است که در (شکل ۷) نشان داده شده است. به‌منظور پیدا کردن پانزده نقطه کلیدی بدن در فضای سه‌بعدی که در بخش قبل معرفی شدند، باتوجه به روند سلسله‌مراتبی (شکل ۷)، ابتدا نقطه P1 که معرف ریشه درخت سینماتیک بدن است، تخمین زده می‌شود. سپس با استفاده از نقطه P1، نقطه پایین گردن (نقطه P8) و مفصل ران چپ (نقطه P2 و P5) محاسبه می‌شوند. به همین

ترتیب، سایر نقاط باتوجه به ترتیب سلسله‌مراتبی در نظر گرفته شده محاسبه می‌شوند. جستجوی ۲۹ پارامتر مدل بدن انسان در شش مرحله جدا از هم انجام می‌شود (که معادل با جدا کردن فضای جستجوی ۲۹ بعدی به شش زیرفضای جستجو است). در مرحله اول نیم‌تنه، مرحله دوم سر و گردن، مرحله سوم و چهارم دست چپ و راست و در نهایت در مرحله پنجم و ششم پای چپ و راست جستجو می‌شوند. با توجه به روند سلسله‌مراتبی در نظر گرفته شده پارامترهای هر مرحله جداگانه تخمین زده می‌شوند که از این طریق فضای جستجوی ابعاد بالای حالت انسان مهار می‌شود.

۵-۲- تابع هدف

در روش پیشنهادی از یک روش بهینه‌سازی مستقیم برای یافتن پارامترهای حالت بدن انسان استفاده شده است. بنابراین، فرمول‌بندی یک تابع هدف خوب برای فرآیند بهینه‌سازی پارامترهای بدن مورد نیاز است. همان‌طور که ذکر شد، در روش پیشنهادی، از یک روش جستجوی سلسله‌مراتبی برای مهار کردن جستجوی فضای ابعاد بالای حالت بدن انسان استفاده شده است. بنابراین تابع هدف برای



(شکل ۷): سلسله‌مراتب در مدل سینماتیک بدن انسان

استخوان A_m در تصویر دوربین شماره ۱ را با $Seg_{1,i}$ نشان می‌دهیم که به صورت یک خط دویبعی در صفحه تصویر در نظر گرفته می‌شود. تعداد N نقطه روی این خط با فاصله یکسان در نظر گرفته می‌شود؛ این نقاط با $S_{1,i,j}, j = 1, \dots, N$ نشان داده می‌شوند. پیکربندی صحیح اسکلت سه‌بعدی، زمانی به دست می‌آید که افکشن هر یک از خطوط اسکلت در وسط مرز عضو متناظرش در سه نمای دوربین‌ها قرار گیرد. فرض کنید مرز عضو مربوط به $Seg_{1,i}$ را با $Bound_{1,i}$ نشان داده شود، بدیهی است که مرز یک جسم شامل دو قسمت است؛ این دو قسمت با $Bound_{1,1,i}$ و $Bound_{1,2,i}$ نشان داده می‌شود. نقاط متناظر با هر یک از این مرزها با $B_{1,1,i,k}, k = 1, \dots, M_1$ و $B_{1,2,i,k}, k = 1, \dots, M_2$ علامت‌گذاری می‌شود که M_1 و M_2 تعداد نقاط هر یک از مرزهای $Bound_{1,1,i}$ و $Bound_{1,2,i}$ می‌باشند. حال تابع $F_{1,i}(X_{i,t}, I_{1,t})$ به صورت زیر تعریف می‌شود:

$$F_{1,i}(X_{i,t}, I_{1,t}) = \left| \begin{array}{l} \sum_{j=1}^N \text{Min}_k (\|S_{1,i,j} - B_{1,1,i,k}\|) \\ - \sum_{j=1}^N \text{Min}_k (\|S_{1,i,j} - B_{1,2,i,k}\|) \end{array} \right| \quad (4)$$

محاسبه $F_{2,i}(X_{i,t}, I_{2,t})$ و $F_{3,i}(X_{i,t}, I_{3,t})$ نیز با توجه به تصاویر متناظرشان مشابه با رابطه (۴) انجام می‌شود. پیکربندی صحیح عضو A_m در زمان t با کمینه‌سازی تابع هدف $F_i(X_{i,t}, I_{i,t})$ (مجموع سه تابع هدف محاسبه شده در سه نمای مختلف) به صورت زیر به دست می‌آید:

هر یک از اعضای بدن جداگانه تعریف می‌شوند. پارامترهای بدن انسان در زمان t به صورت X_t تعریف می‌شود که X_t شامل ۲۹ پارامتر حالت بدن انسان است. باتوجه به روند جستجوی سلسله‌مراتبی، پارامترهای حالت هر یک از اعضای بدن، جداگانه بهینه می‌شوند. بنابراین، پارامترهای مدل انسان به صورت $X_t = \{X_{i,t}\}, i = 1, \dots, 14$ در نظر گرفته می‌شود که $X_{i,t}$ بیان‌گر پارامترهای عضو A_m بدن انسان در زمان t است. (تعداد درجه‌های آزادی هر عضو مطابق (جدول ۱) است). تصاویر مشاهده شده در زمان t را با I_t نشان می‌دهیم. در اینجا هدف، تعریف یک تابع هدف $F_i(X_{i,t}, I_{i,t})$ برای عضو A_m بدن انسان است. تابع هدف $F_i(X_{i,t}, I_{i,t})$ به صورت زیر تعریف می‌شود:

$$F_i(X_{i,t}, I_{i,t}) = F_{1,i}(X_{i,t}, I_{1,t}) + F_{2,i}(X_{i,t}, I_{2,t}) + F_{3,i}(X_{i,t}, I_{3,t}) \quad (3)$$

در رابطه فوق $F_{1,i}(X_{i,t}, I_{1,t})$ ، $F_{2,i}(X_{i,t}, I_{2,t})$ و $F_{3,i}(X_{i,t}, I_{3,t})$ به ترتیب معرف تابع هدف عضو A_m در صفحه تصویر دوربین شماره ۱، دوربین شماره ۲ و دوربین شماره ۳ می‌باشند و $I_{1,t}$ ، $I_{2,t}$ و $I_{3,t}$ به ترتیب بیان‌گر تصویر دوربین‌های شماره ۱، ۲ و ۳ در زمان t هستند. حال با تعریف $F_{1,i}(X_{i,t}, I_{1,t})$ می‌توان این تعریف را به بقیه حالات تعمیم داد.

همان‌طور که در بخش‌های قبل ذکر شد، مدل بدن انسان به صورت مجموعه‌ای از نقاط (مفاصل) و خطوط (استخوان‌های صلب بدن) در نظر گرفته می‌شود. افکشن

$$X_{i,t}^* = \arg \min_{X_{i,t}} (F_i(X_{i,t}, I_t)) \quad (5)$$

در رابطه (5) معرف پارامترهای بهینه عضو t در زمان t است. باتوجه به روند سلسله‌مراتبی تعریف‌شده در بخش‌های قبل، پارامترهای هر یک از عضوهای بدن به‌طور جداگانه با کمینه‌سازی تابع هدف متناظرشان به‌دست می‌آیند. بنابراین، پارامترهای بهینه حالت کامل بدن انسان X_i^* در زمان t با استفاده از کنار هم قرار دادن پارامترهای بهینه حالت هر یک از عضوهای بدن به‌صورت رابطه (6) به‌دست می‌آید:

$$X_t^* = \{X_{i,t}^*\}, i = 1, \dots, 14 \quad (6)$$

5-3- الگوریتم کمینه‌سازی تابع هدف

پس از تعریف تابع هدف برای هر یک از اعضای بدن، کمینه‌سازی توابع هدف معرفی‌شده، به‌منظور به‌دست آوردن پارامترهای بهینه حالت هر عضو با استفاده از الگوریتم *Nelder-Mead* (Lagarias, Reeds et al. 1998) انجام می‌شود. بنابراین در این بخش توضیحات مختصری در مورد این الگوریتم داده می‌شود.

الگوریتم بهینه‌سازی *Nelder-Mead* یک تکنیک بهینه‌سازی غیرخطی برای کمینه‌سازی یک تابع مقدار حقیقی $f(x), x \in R^n$ است. این روش از یک مفهوم به نام سیمپلکس¹ استفاده می‌کند. سیمپلکس یک چندوجهی با $n+1$ رأس در فضای n بعدی است. مثال‌هایی از سیمپلکس‌ها شامل یک بخش خطی روی یک خط، یک مثلث روی یک صفحه، یک چهاروجهی در فضای سه‌بعدی و غیره است. الگوریتم بهینه‌سازی *Nelder-Mead* یک الگوریتم تکراری است که در یک فضای n بعدی، ابتدا $n+1$ رأس یک سیمپلکس را با استفاده از یک مقدار اولیه x_0 ایجاد می‌کند. هدف از هر تکرار الگوریتم، جایگزین کردن بدترین نقطه با یک نقطه بهتر است. این فرآیند تکرار تا زمانی که تمام $n+1$ نقطه روی یک موقعیت با یک تلورانس مشخص هم‌گرا شوند، ادامه پیدا می‌کند.

در هر تکرار الگوریتم، مقدار تابع در هر $n+1$ رأس سیمپلکس ارزیابی می‌شود و این رأس‌ها به‌صورت $f(x_1) \leq f(x_2) \leq \dots \leq f(x_{n+1})$ مرتب می‌شوند. بهترین رأس x_1 ، بهترین رأس دوم x_2 و به همین ترتیب

¹Simplex

بدترین رأس x_{n+1} است. در هر تکرار از الگوریتم، بدترین رأس (x_{n+1}) در سیمپلکس کنار گذاشته می‌شود و یک نقطه دیگر جایگزین آن می‌شود و با اینکه تمام رأس‌های سیمپلکس بر طبق مقدار بهترین نقطه (x_1) جایگزین می‌شوند. چهار پارامتر عددی برای تعریف روش *Nelder-Mead* مورد نیاز است. این پارامترها، ضرایب انعکاس² (ρ)، توسعه³ (χ)، ادغام⁴ (γ) و انقباض⁵ (σ) هستند. این پارامترها در الگوریتم استاندارد *Nelder-Mead* به‌صورت رابطه (7) در نظر گرفته می‌شوند:

$$\rho = 1, \chi = 2, \gamma = 0.5, \sigma = 0.5 \quad (7)$$

مرکز ثقل تمام نقاط به جز نقطه x_{n+1} با استفاده از رابطه $\bar{x} = \sum_{i=1}^n x_i / n$ محاسبه می‌شود. تعدادی رأس دیگر با

استفاده از \bar{x} و پارامترهای الگوریتم محاسبه می‌شوند که محاسبه مقدار این رأس‌ها در روابط زیر نشان داده می‌شود:

$$x_r = (1 + \rho)\bar{x} - \rho x_{n+1} \quad (8)$$

$$x_e = \bar{x} + \chi(x_r - \bar{x}) \quad (9)$$

$$x_c = x_{n+1} + \gamma(\bar{x} - x_{n+1}) \quad (10)$$

$$x_i = x_1 + \sigma(x_i - x_1), i \in \{2, \dots, n+1\} \quad (11)$$

(شکل 8) نمودار جعبه‌ای الگوریتم جستجوی *Nelder-Mead* را نشان می‌دهد. رأس‌های سیمپلکس در تکرارهای الگوریتم مطابق نمودار جعبه‌ای (شکل 8) تغییر می‌نمایند تا زمانی که معیار توقف برآورده شود. شرط توقف الگوریتم بدین‌صورت است که قطر سیمپلکسی که در هر تکرار تولید می‌شود، کم‌تر از یک تلورانس مشخص شود.

6- نتایج پیاده‌سازی

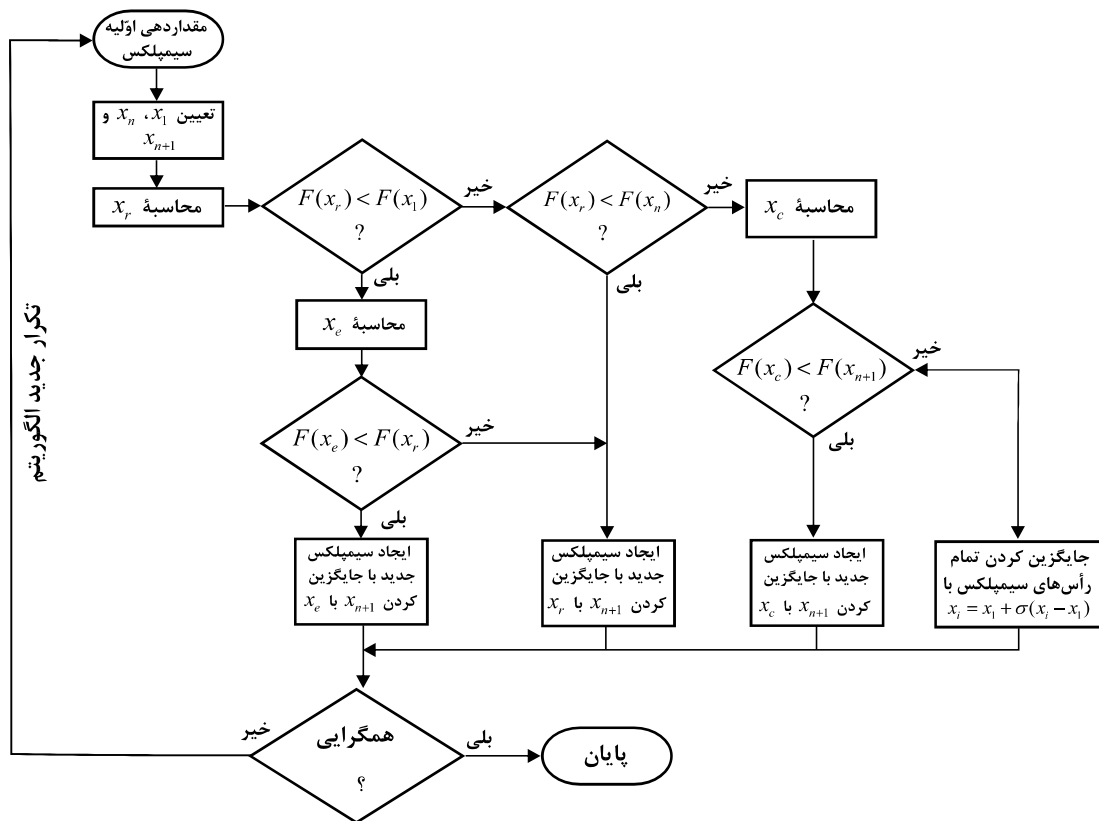
در روش پیشنهادی، ورودی به‌صورت رشته ویدیویی از سه نمای مختلف در نظر گرفته شده است و خروجی با استفاده از 29 پارامتر حالت سه‌بعدی بدن انسان را بیان می‌کند. تمام مراحل روش پیشنهادی مانند استخراج سایه‌نما، آشکارسازی پوست و برجسب‌زنی کانتور به‌صورت خودکار انجام می‌شوند. در مرحله آشکارسازی پوست، برای ایجاد توابع هیستوگرام پوست H و هیستوگرام غیر پوست h ، فریم‌های مختلف از تصاویر دوربین‌های مختلف پایگاه داده *i3DPost*

² Reflection

³ Expansion

⁴ Contraction

⁵ Shrinkage

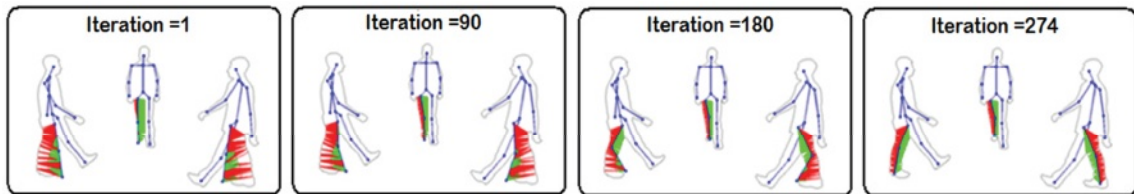


(شکل ۸): نمودار جعبه‌ای الگوریتم جستجوی Nelder-mead

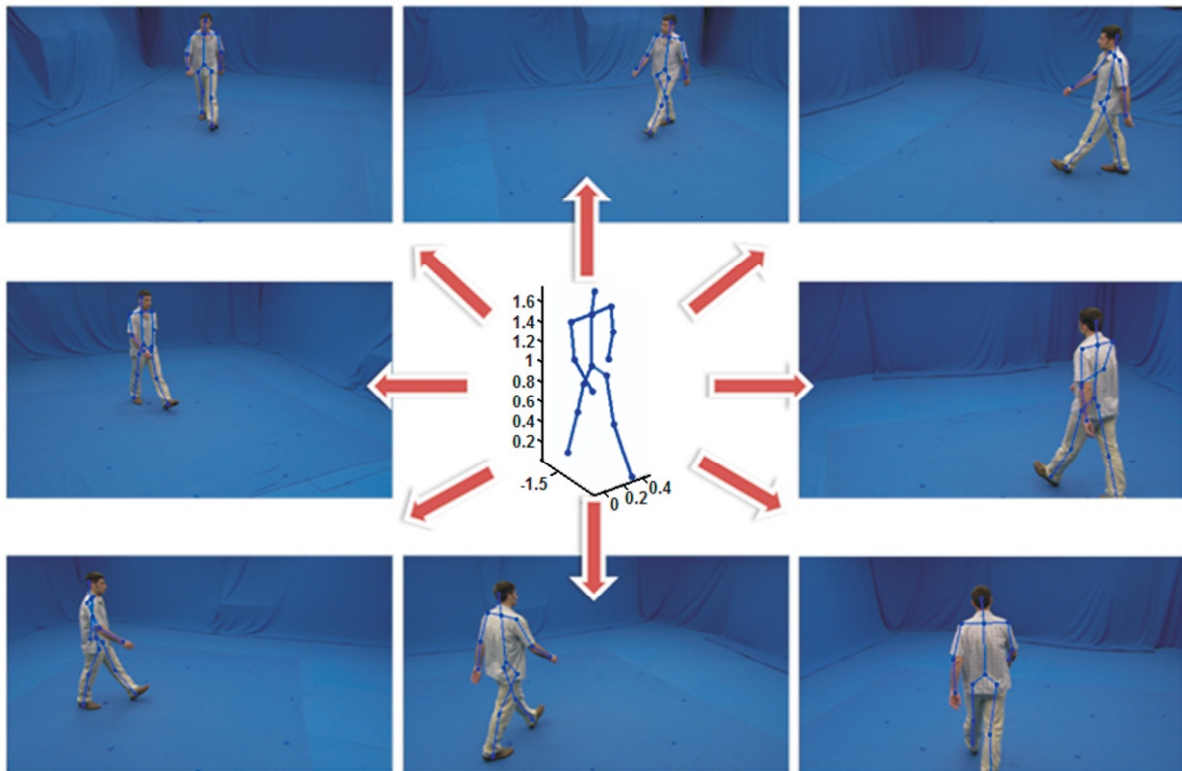
تطبیق داده می‌شوند. پس از استخراج ویژگی از تصاویر سه دوربین مختلف، تطبیق بین افکشن مدل اسکلت سه‌بعدی با ویژگی‌های تصویر انجام می‌شود. این تطبیق، از طریق کمینه‌سازی توابع هدف حالت اعضای بدن انجام می‌شود. بهینه‌سازی پارامترهای حالت بدن انسان در شش مرحله با توجه به ترتیب سلسله‌مراتبی درخت سینماتیک مدل بدن انسان انجام می‌شود. پیکربندی صحیح اسکلت سه‌بعدی، زمانی به دست می‌آید که افکشن هر یک از خطوط اسکلت در وسط مرز عضو متناظرش در سه نمای دوربین‌ها قرار گیرد. این کار از طریق تعریف تابع هدف پیشنهادی و کمینه‌سازی آن انجام می‌شود. برای نمونه، فرآیند تخمین پارامترهای حالت پای راست در تکرارهای مختلف الگوریتم بهینه‌سازی برای رسیدن به حالت بهینه در (شکل ۹) نشان داده شده است. همان‌طور که در این شکل مشاهده می‌شود در تکرار ۲۷۴ حالت بهینه پای راست حاصل شده است. در (شکل ۹) تعداد یکصد نقطه بر روی افکشن خط بیان‌گر پای راست در سه نما از دوربین‌ها در نظر گرفته شده است. خطوط قرمز و سبز، به ترتیب بیان‌گر کمینه‌ترین فاصله هر یک از این یکصد نقطه از مرز یک و دوی پای راست می‌باشند. با توجه به تابع هدف تعریف‌شده، هستند.

مورد استفاده قرار گرفته و در مرحله آموزش، نواحی پوست و غیر پوست به صورت دستی انتخاب می‌شود. بدین ترتیب میزان حضور مؤلفه‌های رنگی (r, g, b) در تصاویر پوست به صورت یک هیستوگرام چهاربندی استخراج می‌شود که در آن مقدار H در هر نقطه دلخواه (r, g, b) از حجم بیان‌گر احتمال پوست بودن آن رنگ است. همچنین مقدار h در هر نقطه دلخواه (r, g, b) از حجم بیانگر احتمال غیر پوست بودن آن رنگ است.

باتوجه به اینکه فرض شده شخص در حالت ایستاده است، ابتدا قد شخص تخمین زده می‌شود و طول‌های اولیه استخوان‌های اسکلت به مقیاس قد شخص تغییر داده می‌شوند. حالت شخص در هر فریم به صورت خودکار تخمین زده می‌شود. روش پیشنهادی بر روی رشته ویدیویی چنددوربینی راه رفتن یک شخص از پایگاه داده $i3DPost$ (*Gkalelis, Kim et al. 2009*) آزمایش شده است. در روش پیشنهادی فقط از سه دوربین استفاده شده است. این دوربین‌ها در پایگاه داده فوق با شماره‌های ۳ و ۵ و ۷ علامت گذاری شده‌اند. همان‌طور که در (شکل ۱) مشاهده می‌شود، ویژگی‌های تصویر از سه نمای مختلف استخراج و با افکشن مدل اسکلت سه‌بعدی پیشنهادی در سه نمای مختلف



(شکل ۹): نمایش فرآیند تخمین پارامترهای حالت پای راست در تکرارهای مختلف الگوریتم بهینه‌سازی



(شکل ۱۰): نمایش حالت سه‌بعدی بازیابی شده و افکنش آن بر ۸ صفحه دوربین متفاوت

در هر فریم به‌دست می‌آید که حالت سه‌بعدی بازیابی شده در فریم ۵۲ از رشته ویدیویی راه‌رفتن از پایگاه داده i3DPost در (شکل ۱۰) نشان داده شده است. با توجه به نامعلوم بودن موقعیت سه‌بعدی واقعی مفاصل بدن شخص، برای ارزیابی کیفی حالت سه‌بعدی بازیابی شده، افکنش پیکربندی اسکلت بازیابی شده بر روی هشت نما از دوربین‌های مختلف در (شکل ۱۰) نشان داده شده است؛ در صورتی که در روش پیشنهادی فقط از سه نما استفاده شده است. نتایج آزمایش، توانمندی روش پیشنهادی در تخمین حالت را بدون داشتن هیچ‌گونه اطلاعاتی در مورد حالت شخص نشان می‌دهد. با فرض اینکه جهش بزرگ در حرکت انسان از یک فریم به فریم بعدی اتفاق نمی‌افتد، پارامترهای تخمین زده شده در فریم‌های قبلی به‌عنوان شرایط اولیه خوبی برای فریم‌های بعدی مورد استفاده قرار می‌گیرند که این در افزایش سرعت هم‌گرایی الگوریتم تأثیرگذار است.

مجموع تفاضل اندازه خطوط قرمز و سبز در هر نما محاسبه می‌شوند. این محاسبه، در رابطه (۴) برای دوربین شماره ۱ و بخش ۱۱ مدل اسکلت بدن انسان نشان داده شده است. در این رابطه مقدار N برابر ۱۰۰ در نظر گرفته شده است. در نهایت با استفاده از رابطه (۵) پارامترهای حالت بهینه پای راست به‌دست می‌آیند. پس از کمینه‌سازی تابع هدف پای راست، همان‌طور که در (شکل ۹) مشاهده می‌شود، افکنش خط بیان‌گر پای راست در سه نمای متفاوت در وسط مرز پای راست در هر نما قرار گرفته شده است؛ به طوری که تفاضل اندازه خطوط قرمز و سبز در هر نما کمینه شده است. از این طریق پارامترهای بهینه پای راست در فضای سه‌بعدی به‌دست می‌آید. به‌طور مشابه، پس از بهینه‌سازی پارامترهای حالت سایر بخش‌های مدل اسکلت سه‌بعدی، افکنش خط بیان‌گر هر بخش اسکلتی در هر نما در وسط مرز عضو متناظرش قرار می‌گیرد. پس از بهینه‌سازی پارامترهای بدن، حالت سه‌بعدی صحیح شخص

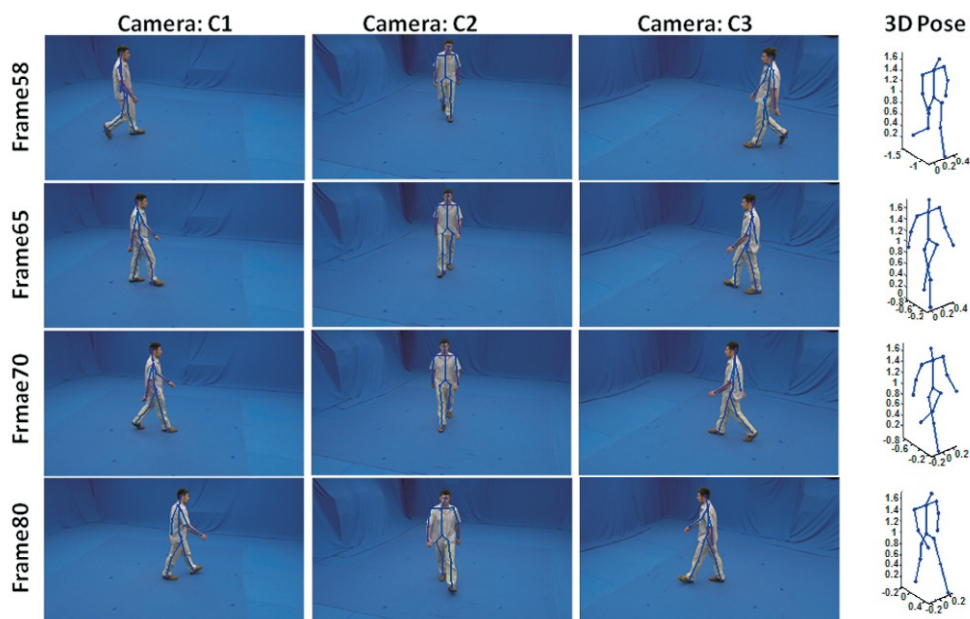
پیشنهادی بر حسب میلی‌متر در طول پنجاه فریم مطابق با رابطه (۱۲) در (شکل ۱۲) نشان داده شده است. مقدار متوسط خطا در طول پنجاه فریم در این شکل با خط قرمز نشان داده شده است. همان‌طور که در این شکل دیده می‌شود، مقدار متوسط خطا برای رشته ویدیویی آزمایش شده حدود ۷.۵ سانتی‌متر (۷۵ میلی‌متر) است. نتایج عددی مربوط به اختلاف مکان پانزده نقطه مفروض در دو اسکلت واقعی و اسکلت تخمینی برای روش پیشنهادی و روش (Zhu, Dariush et al. 2010) در (جدول ۲) گزارش شده است. مدل اسکلت سه‌بعدی مورد استفاده در (Zhu, Dariush et al. 2010) شامل چهارده نقطه کلیدی بدن انسان است که باتوجه به شباهت زیاد مدل اسکلت آنها با مدل اسکلت مورد استفاده در این مقاله، می‌توان نتایج الگوریتم پیشنهادی را با روش آنها به‌صورت کمی مقایسه کرد. در این جدول میانگین و انحراف معیار خطای هر یک از ابعاد نقاط مفروض در مدل اسکلت سه‌بعدی و همچنین میانگین فاصله اقلیدسی نقاط تخمین‌زده شده با نقاط واقعی برای کل فریم‌های حرکتی در ستون‌های جداگانه برای دو روش مذکور نشان داده شده است. مقایسه میانگین خطای روش پیشنهادی با روش (Zhu, Dariush et al. 2010) در هر یک از ابعاد X, Y, Z و همچنین میانگین خطای اقلیدسی (ردیف آخر جدول ۲) حاکی از دقت بیش‌تر روش پیشنهادی است. همان‌طور که در (جدول ۲) نیز مشاهده می‌شود از مزایای روش پیشنهادی ما دقت بسیار خوب در تخمین نقاط مربوط به دست‌هاست در صورتی که روش (Zhu, Dariush et al. 2010) دقت پایینی در تخمین این نقاط دارد. از دلایل دقت تخمین خوب نقاط کلیدی دست‌ها، استفاده از آشکارساز پوست و عدم انسداد دست چپ و راست در حرکت راه رفتن است. بیشترین خطای تخمین در روش پیشنهادی ما مربوط به تخمین محل سر است. البته باید توجه داشت که تابع هدف معرفی شده به‌صورت تفاضل فاصله افکنش هر یک از استخوان‌های بدن از مرز تصاویر متناظرشان در هر نمای دوربین تعریف شده است. بنابراین طبیعی است که در قسمت‌هایی از بدن، که تقارن کامل وجود ندارد (مانند سر)، دقت تخمین کم‌تر باشد. البته می‌توان برای اعضای نامتقارن بدن، (مانند سر و نیم‌تنه) تابع هدف را به‌صورت تفاضل وزن‌دار از مرزهای عضو مورد نظر در نماهای مختلف در نظر گرفت. یکی دیگر از مواردی که از روی اعداد به‌دست آمده از (جدول ۲) قابل استنتاج است پایین بودن دقت تخمین در نقاط کلیدی پاها نسبت به سایر

به‌طور اصولی با استفاده از یک سامانه ضبط حرکت مبتنی بر نشانه^۱ می‌توان محل واقعی مفاصل کلیدی بدن در فضای سه‌بعدی را به‌دست آورد. باتوجه به عدم دسترسی به پایگاه داده‌ای که شامل داده‌های ضبط حرکت مبتنی بر نشانه نیز باشد، به‌منظور ارزیابی کمی الگوریتم پیشنهادی در این مقاله از یک روش دستی برای به‌دست آوردن محل واقعی مفاصل کلیدی بدن انسان استفاده می‌شود. بدین‌صورت که ابتدا به‌صورت دستی محل مفصل‌های کلیدی بدن در تصاویر چند نمای مختلف برچسب‌زنی می‌شوند. سپس با استفاده از پارامترهای دوربین در هر نما، این نقاط به یک خط در فضای سه‌بعدی نگاشت می‌شوند. محل تلاقی خطوط نگاشت‌شده از هر نما در فضای سه‌بعدی، معرف محل سه‌بعدی، هر مفصل کلیدی خواهد بود؛ بنابراین با استفاده از روش ذکر شده می‌توان محل واقعی پانزده نقطه مفروض در مدل اسکلت پیشنهادی برای هر فریم از یک رشته ویدیویی را به‌دست آورد. فرض کنید محل واقعی پانزده نقطه مفروض در مدل اسکلت پیشنهادی به‌صورت $X_t = \{P_1, P_2, \dots, P_{15}\}$ و محل تخمین‌زده شده این نقاط توسط الگوریتم پیشنهادی به‌صورت $\hat{X}_t = \{\hat{P}_1, \hat{P}_2, \dots, \hat{P}_{15}\}$ بیان شوند که در این بیان‌ها $P_k, \hat{P}_k \in \mathbb{R}^3$ موقعیت سه‌بعدی نقطه k ام در سامانه مختصات جهانی را نشان می‌دهند. با مقایسه این داده‌ها از طریق رابطه (۱۲) می‌توان خطای مکان‌یابی الگوریتم پیشنهادی برای هر فریم را محاسبه کرد.

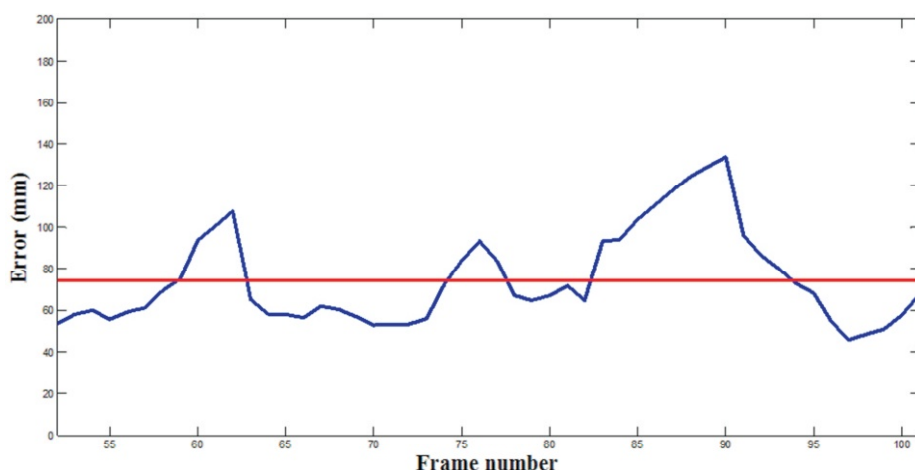
$$Error(X_t, \hat{X}_t) = \frac{1}{15} \sum_{k=1}^{15} \|P_k - \hat{P}_k\| \quad (12)$$

روش پیشنهادی بر روی یک رشته ویدیویی چنددوربینی، راه رفتن یک شخص از پایگاه داده *i3DPost* آزمایش شده است. در روش پیشنهادی فقط از سه دوربین استفاده شده است. این دوربین‌ها در پایگاه‌داده فوق با شماره‌های ۳ و ۵ و ۷ علامت‌گذاری شده‌اند. به‌دلیل رؤیت کامل بدن انسان در تصویر دوربین‌های در نظر گرفته شده، این رشته ویدیویی از فریم ۵۲ تا ۱۰۱ در نظر گرفته شده است. نتایج پیاده‌سازی به‌صورت کیفی و کمی در (شکل‌های ۱۱ و ۱۲) نشان داده شده است. در (شکل ۱۱)، حالت سه‌بعدی بازیابی‌شده شخص در چندین فریم نمایش داده شده است که از روی این شکل می‌توان به‌صورت کیفی و بصری، عملکرد الگوریتم پیشنهادی را مورد ارزیابی قرار داد. همچنین به‌منظور ارزیابی کمی روش پیشنهادی، نمودار متوسط خطای تخمین مکان نقاط مفروض در مدل اسکلت

¹ Marker-based motion capture system



شکل ۱۱: نتایج تصویری تخمین حالت انسان در رشته ویدیویی راه رفتن از پایگاه داده i3DPost



شکل ۱۲: نتایج کمی تخمین حالت انسان در رشته ویدیویی راه رفتن از پایگاه داده i3DPost با روش پیشنهادی به صورت خطای میانگین در هر فریم

نقاط مدل بدن انسان می باشد. دلیل این مسئله، مواجه شدن با انسداد پای چپ و راست در حرکت راه رفتن می باشد که در این حالت مرزهای کانتور پای چپ و راست بر روی یکدیگر قرار می گیرند و تخمین حالت دقیق پای چپ و راست ناممکن می شود. البته با افزایش تعداد دوربین ها، می توان دقت تخمین در این نقاط را نیز بهبود داد. در مجموع از مقایسه روش پیشنهادی با روش (Zhu, Dariush et al. 2010) می توان نتیجه گرفت که بهره گیری از سامانه تخمین حالت چنددوربینی و استفاده از اطلاعات مرزی در سایه نماها در مقایسه با یک سامانه تخمین مبتنی بر تصاویر عمق دقت بیشتری برخوردار است. همچنین، به منظور ارزیابی دقت عملکرد روش پیشنهادی، مقدار میانگین خطای مکانی حالت انسان با روش مقاله (Hofmann and Gavril

2012) که یک روش تخمین حالت مبتنی بر چند دوربین می باشد، مورد مقایسه قرار گرفته است. لازم به توضیح است که پایگاه داده مورد استفاده در این روش HumanEva است که دارای داده های آموزشی، آزمایشی و همچنین برنامه هایی برای محاسبه برآورد خطاست. متأسفانه دسترسی به این پایگاه داده برای دانشگاه های کشور ایران محدود شده است و در عمل امکان پیاده سازی روش پیشنهادی با این پایگاه داده و مقایسه روشها در شرایط یکسان میسر نشد؛ با وجود این، بررسی سکانس های ویدیویی مختلف و مقایسه مقدار میانگین خطای مکانی حالت انسان در شرایط فیلم برداری چنددوربینی که در هر دو پایگاه داده HumanEva¹ و i3DPost² لحاظ شده است، می تواند معیار به نسبه خوبی

¹ <http://vision.cs.brown.edu/humaneva/>

² http://kahlan.eps.surrey.ac.uk/i3dpost_action/

(جدول ۲): میانگین و انحراف معیار خطای مکانی اجزای بدن انسان در مدل اصلی و تخمینی در کل فریم‌های حرکتی (میلی متر)

روش پیشنهادی							روش (Zhu, Dariush et al. 2010)							نقاط	
														مفصلی	
$(\mu, \sigma)_{\Delta X}$	$(\mu, \sigma)_{\Delta Y}$	$(\mu, \sigma)_{\Delta Z}$	μ_d	$(\mu, \sigma)_{\Delta X}$	$(\mu, \sigma)_{\Delta Y}$	$(\mu, \sigma)_{\Delta Z}$	μ_d	$(\mu, \sigma)_{\Delta X}$	$(\mu, \sigma)_{\Delta Y}$	$(\mu, \sigma)_{\Delta Z}$	μ_d	$(\mu, \sigma)_{\Delta X}$	$(\mu, \sigma)_{\Delta Y}$	$(\mu, \sigma)_{\Delta Z}$	μ_d
18	20	27	94	17	23	36.6	30	40	71	25	37	18	85.5	P1	
22	27	11	79	8	18	25.9	10	79	146	39	6	23	146.5	P2	
37	52	38	52	41	43	67.0	33	52	64	106	62	32	95.0	P3	
41	53	39	67	45	46	72.3	17	52	56	67	161	34	171.3	P4	
25	39	20	71	15	18	35.4	19	72	133	34	22	23	136.1	P5	
51	63	46	71	52	61	86.1	52	57	43	121	51	38	84.6	P6	
39	49	40	64	41	42	69.3	60	65	29	98	155	56	168.7	P7	
11	45	9	14	23	40	27.0	-	-	-	-	-	-	-	P8	
21	53	130	118	12	19	132.2	7	31	22	25	41	26	47.1	P9	
9	24	14	33	25	31	30.0	74	39	2	35	10	17	74.7	P10	
5	8	12	15	9	15	15.8	58	34	6	47	58	23	82.2	P11	
7	6	7	15	5	15	11.1	108	28	20	60	141	51	178.7	P12	
14	24	12	18	26	35	31.9	86	20	1	47	11	25	86.7	P13	
9	11	7	13	7	24	13.4	46	40	13	36	85	18	97.5	P14	
8	21	9	28	3	6	12.4	135	36	21	61	143	43	197.8	P15	
21.1	33	28.1	50.1	21.9	29.1	44.4	52.5	46.1	44.8	57.2	70.2	30.5	118.0	متوسط	

(جدول ۳): میانگین خطای مکانی تخمین حالت انسان برای چند رشته ویدیویی

روش پیشنهادی	(Hofmann and Gavrilu 2012)						روش مورد استفاده برای تخمین حالت
i3DPost	HumanEva-I						پایگاه داده مورد استفاده
Walking	Walking-I		ThrowCatch-I		Gestures-I		نام سکانس ویدیویی مورد استفاده
	Subject 2	Subject 1	Subject 2	Subject 1	Subject 2	Subject 1	
7.5	7.4	4.7	8.9	5.8	6.5	6.3	خطای میانگین در کل سکانس (سانتی متر)

قسمت‌های مختلف بدن و از روش‌های بهینه‌سازی مستقیم برای تخمین حالت انسان استفاده شده است. سامانه پیشنهادی قادر به تشخیص انسدادهای و تخمین به‌نسبه دقیق اعضای بدن مانند نیم‌تنه، سر و دست‌هاست. از مزایای روش معرفی‌شده این است که باتوجه به معرفی روش برچسب‌زنی کانتور و تابع هدف معرفی‌شده، حتی با استفاده از ساده‌ترین روش‌های بهینه‌سازی، الگوریتم قادر به هم‌گرایی به حالت صحیح است. از دیگر مزایای روش پیشنهادی ارزش‌دهی اولیه خودکار آن است؛ به طوری که روش معرفی‌شده قادر به تخمین خودکار حالت انسان در فریم اول یک رشته ویدیویی چنددوربینی است. درحالی‌که سامانه معرفی‌شده، قادر به تخمین حالت انسان به‌طور مؤثر است، ولی با این حال لازم به ذکر است که در مواردی روش پیشنهادی دچار شکست می‌شود. یکی از مواردی که روش پیشنهادی در آن دچار مشکل می‌شود مواجه شدن با انسدادهای پای چپ و راست است. در این حالت مرزهای کانتور پای چپ و راست بر روی یکدیگر قرار می‌گیرند و تخمین حالت دقیق پای چپ و راست ناممکن می‌شود. البته باید توجه داشت که در این حالت می‌توان با افزایش تعداد دوربین‌ها و یا استفاده از ویژگی‌های لبه تصویر این نقصان را برطرف کرد. یکی دیگر از مواردی که روش پیشنهادی را دچار شکست می‌کند این

برای ارزیابی روش پیشنهادی باشد. نتایج این مقایسه در (جدول ۳) آورده شده و همان‌طور که ملاحظه می‌شود خطای روش پیشنهادی قابل رقابت با خطای روش (Hofmann and Gavrilu 2012) است.

آزمایش روش پیشنهادی بر روی یک پردازنده ۲.۲ گیگا هرتزی اینتل با حافظه رم یک گیگا بایت انجام شده است. زمان کلی پردازش برای هر فریم حدود ۱۲۹ ثانیه است، این زمان محاسباتی، سامانه معرفی‌شده را برای کاربردهای بلادرنگ^۱ نامناسب می‌سازد. البته باید توجه کرد که پیاده‌سازی بر روی وضوح تصویری بالا ۱۹۲۰×۱۰۸۰ انجام شده است و با بهینه‌سازی برنامه‌های نوشته شده می‌توان زمان محاسباتی نیز را کاهش داد.

۷- نتیجه‌گیری

در این مقاله، روشی جدید برای تخمین حالت سه‌بعدی انسان در رشته ویدیویی چنددوربینی معرفی شد. در روش پیشنهادی، به جای جستجوی مستقیم فضای حالت ابعاد بالای انسان و استفاده از الگوریتم‌های استنتاج پیچیده، از یک روش جستجوی سلسله‌مراتبی، توابع هدف جداگانه برای

¹Real-time

Lee, M. W. and R. Nevatia (2009). "Human pose tracking in monocular sequence using multilevel structured models." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 31(1): 27-38.

Mündermann, L., S. Corazza, et al. (2006). Measuring human movement for biomechanical applications using markerless motion capture. *Proceedings of SPIE, Citeseer*.

Pilu, M., A. W. Fitzgibbon, et al. (1996). Ellipse-specific direct least-square fitting. *IEEE International Conference on Image Processing, IEEE*.

Siek, J., L. Q. Lee, et al. (2002). *The Boost Graph Library: User Guide and Reference Manual*, Addison-Wesley.

Sigal, L., S. Bhatia, et al. (2004). Tracking loose-limbed people. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), IEEE*.

Zhu, Y., B. Dariush, et al. (2010). "Kinematic self retargeting: A framework for human pose estimation." *Computer Vision and Image Understanding* 114(12): 1362-1375.



قاسم خادمی در سال ۱۳۸۵ مدرک کارشناسی خود را در رشته مهندسی برق، مخابرات از دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران) اخذ و سپس با ادامه تحصیل در دانشگاه

صنعتی سهند، مدرک کارشناسی ارشد خود را در گرایش مخابرات سیستم در سال ۱۳۹۰ دریافت کرد. زمینه‌های تحقیقاتی مورد علاقه ایشان بینایی کامپیوتر، پردازش تصویر و شناسایی الگو است.

نشانی رایانامه ایشان عبارتست از:

g_khademi@sut.ac.ir



حسین ابراهیم‌نژاد مدرک کارشناسی و کارشناسی ارشد خود را به ترتیب در سال‌های ۱۳۷۲ و ۱۳۷۵ در رشته مهندسی برق- الکترونیک و برق مخابرات از دانشگاه تبریز و دانشگاه

صنعتی خواجه نصیرالدین طوسی اخذ کرد. همچنین مدرک دکتری خود را در گرایش مخابرات سیستم در سال ۱۳۸۶ از دانشگاه تربیت مدرس دریافت کرد. زمینه‌های تحقیقاتی مورد علاقه ایشان بینایی کامپیوتر، پردازش مدل سه‌بعدی، پردازش تصویر، شناسایی الگو و محاسبات نرم بوده و در حال حاضر عضو هیأت علمی دانشگاه صنعتی سهند می‌باشد. نشانی رایانامه ایشان عبارتست از:

ebrahimezhad@sut.ac.ir

است که در صحنه‌های واقعی خارج از محیط آزمایشگاه، به دلیل وجود پس‌زمینه شلوغ، ممکن است سایه‌نما و کانتور استخراج‌شده آغشته به نوفه باشند که در این حالت ایجاد گسستگی در کانتور بدن انسان، روش برچسب‌زنی کانتور را دچار مشکل خواهد کرد. البته باید توجه داشت که در روش پیشنهادی، فقط از ویژگی‌های سایه‌نما، کانتور و پوست استفاده شده است که می‌توان با استفاده از اطلاعات لبه و بافت، تخمین حالت بدن انسان را به صورت قوی‌تری انجام داد.

۸- منابع

Cheung, K., S. Baker, et al. (2003). Shape-from-silhouette of articulated objects and its use for human body kinematics estimation and motion capture. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), IEEE*.

Conaire, C. O., N. E. O'Connor, et al. (2007). Detector adaptation by maximising agreement between independent data sources. *IEEE International Workshop on Object Tracking and Classification Beyond the Visible Spectrum, IEEE*.

Deutscher, J., A. Blake, et al. (2000). Articulated body motion capture by annealed particle filtering. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), IEEE*.

Fitzgibbon, A., M. Pilu, et al. (1999). "Direct least square fitting of ellipses." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21(5): 476-480.

Franco, J. S. and E. Boyer (2005). Fusion of multiview silhouette cues using a space occupancy grid. *IEEE International Conference on Computer Vision (ICCV), IEEE*.

Gavrila, D. M. and L. S. Davis (1996). 3-D model-based tracking of humans in action: a multi-view approach. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), IEEE*.

Gkalelis, N., H. Kim, et al. (2009). The i3dpost multi-view and 3d human action/interaction database. *Conference for Visual Media Production., IEEE*.

Hofmann, M. and D. Gavrila (2012). "Multi-view 3D human pose estimation in complex environment." *International journal of computer vision* 96(1): 103-124.

Kehl, R., M. Bray, et al. (2005). Full body tracking from multiple views using stochastic sampling. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), IEEE*.

Lagarias, J. C., J. A. Reeds, et al. (1998). "Convergence Properties of the Nelder--Mead Simplex Method in Low Dimensions." *SIAM Journal on Optimization* 9(1): 112-147.