# تخمین چنددوربینی حالت سه‌بعدی انسان با برازش افکنش مدل اسکلت سه‌بعدی مفصل‌دار در تصاویرسایه‌نما

قاسم خادمی و حسین ابراهیم‌نژاد
آزمایشگاه تحقیقاتی بینایی کامپیوتر، دانشکدهٔ مهندسی برق، دانشگاه صنعتی سهند

## چکیده

**تحلیل و ضبط خودکار حرکت انسان بر اساس تصاویر یا ویدیو به‌دلیل 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 به‌همین دلیل دقّت تخمین بر عملکرد این کاربردها تأثیر زیادی دارد.تخمین حالت سه‌بعدی انسان از مشاهدات تصویری باتوجّه به بازهٔ وسیع تغییرات در ظاهر و مفصل‌بندی انسان، ابعاد بالا در فضای حالت انسان و پدیدهٔ خودانسدادی، یک مسئلهٔ بحث‌برانگیز است. در این مقاله، روشی جدید برای تخمین حالت سه‌بعدی انسان در رشته ویدئویی 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 اوّلیه 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 مؤثّری به‌عنوان یک سامانهٔ بدون نشانه برای تخمین حالت سه‌بعدی انسان در رشتهٔ ویدیویی چنددوربینی به‌کار 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 به‌دلیل پیچیدگی مسئله و کاربردهای بسیار زیادش، یکی از زمینه‌های تحقیقاتی بسیار فعّال است. این زمینهٔ تحقیقاتی شامل تعدادی از مسائل دشوار مانند استنتاج حالت[1] و حرکت یک جسم سه‌بعدی غیرصلب[2] مفصل‌دار[3] با خاصیت خودانسدادی[4] از روی تصاویر است. پیچیدگی این مسائل، تحقیقات در این زمینه را از نقطه نظر آکادمیک بحث‌برانگیز می‌سازد. تکنیک‌های تخمین حالت سه‌بعدی انسان[5] که از سامانه‌های مبتنی بر بینایی[6] استفاده می‌کنند، کاربردهای بسیار زیادی دارند. تخمین سه‌بعدی حالت انسان از قسمت‌های اساسی در این کاربردها و دقّت 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 به‌کار می‌رود. در علوم ورزشی، بازسازی دقیق حالت سه‌بعدی انسان به ورزش‌کاران کمک می‌کند تا به‌صورت دیداری حرکات خودشان را برای بهبود عملکرد فعّالیت‌شان تحلیل نمایند. در زمینهٔ فیزیوتراپی، برای شناسایی دلایل اساسی وضعیت‌های نامطلوب حرکتی بیمار که ممکن است به‌علّت ضربه، فلج مغزی یا دیگر مشکلات عصبی و عضلانی باشد، از تحلیل راه رفتن انسان استفاده می‌شود. در زمینهٔ کاربرد نظارت ویدیویی[7] با استفاده از دوربین‌های ارزان

[1] Pose
[2] Non-rigid
[3] Articulated
[4] Self-occluding
[5] 3D human pose estimation
[6] Vision-based
[7] Video survillance

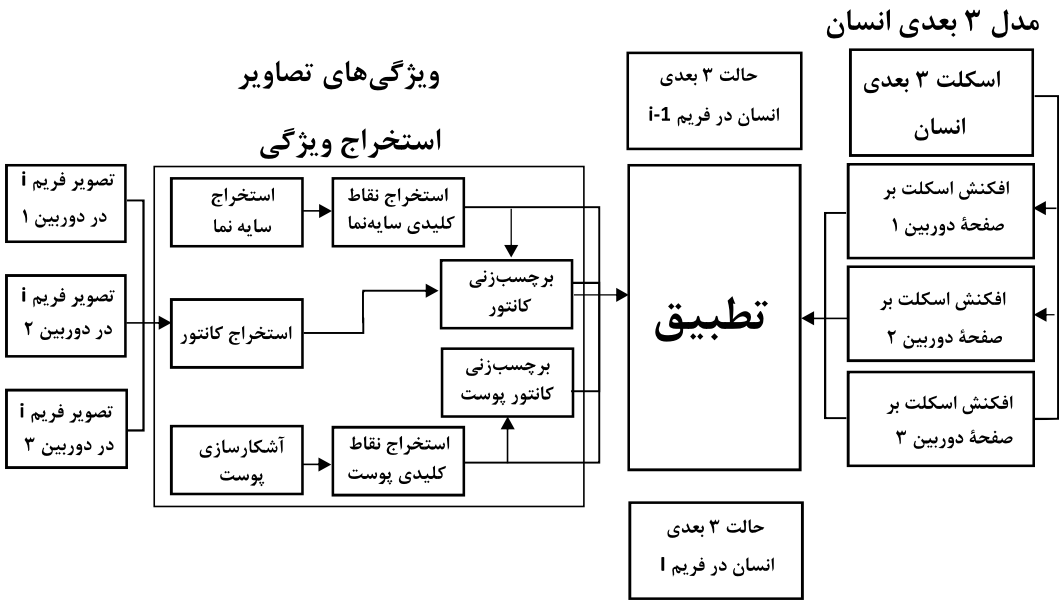
قیمت و ایجاد بهبود در قدرت محاسباتی، تحلیل دقیق حالت سه‌بعدی انسان از ویدیو، می‌تواند اپراتورهای نظارت را برای شناسایی رویدادهایی از قبیل دویدن، راه رفتن، دزدی از مغازه‌ها، بالا رفتن از دیوار، پرسه‌زدن و دیگر فعالیت‌های غیرطبیعی انسان کمک کند.

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 معرّف 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. درواقع جستجوی زوایای اتصال مفاصل با این هدف انجام می‌شود که در تمام نماها تفاضل فاصله‌های افکنش خطوط اسکلت از مرز سایه‌نمای[1] 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 این‌رو بازیابی حالت کامل بدن انسان به‌طور مستقیم دشوار، امّا محاسبهٔ 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 تصویر 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 فریم iأم استخراج می‌شوند. چند نقطهٔ کلیدی از سایه‌نما به‌عنوان تخمین اوّلیه از محل ریشهٔ بدن، مفاصل قوزک و زانوی پاها استخراج می‌شوند. همچنین با استفاده از نواحی پوستی آشکارسازی‌شده 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 اوّلیه سر، گردن، دست‌ها و مفاصل را به‌دست آورد که در هم‌گرایی الگوریتم به پیکربندی حالت صحیح مؤثّرند. در ادامه، با استفاده از روش پیشنهادی برچسب‌زنی 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 اطّلاعات فریم‌های قبلی، ارزش‌دهی اوّلیه 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 کلّی توصیف می‌شوند.

[1] Silhouette




**(شکل۱):نمودار جعبه‌ای روش پیشنهادی**

اکثر روش‌هایی را که با چند دوربین سروکار دارند، می‌توان به دو دستهٔ کلی تقسیم‌بندی کرد: یک دسته روش‌هایی که به‌صراحت از پوسته قابل مشاهده[1] استفاده و دستهٔ 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 درنظرگرفتن ارتباط بین اطّلاعات در دوربین‌های مختلف ضروری است.

روش‌های مبتنی بر پوستهٔ قابل مشاهده به‌صراحت از ارتباط ویژگی‌های نماهای مختلف برای بازسازی هندسه محدودهٔ سه‌بعدی تقریبی جسم واقعی استفاده می‌کنند. زمانی‌که تعداد نماها افزایش پیدا می‌کند، می‌توان نشان داد که پوسته قابل مشاهده، به‌شکل واقعی جسم نزدیک می‌شود. اکثر روش‌های مبتنی بر پوستهٔ قابل مشاهده (*Cheung, Baker et al. 2003; Kehl, Bray et al. 2005*) بر یک فرآیند تفریق پس‌زمینهٔ 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 پوستهٔ قابل مشاهده این است که بر جداکردن پس زمینه تقریبا کامل تکیه دارند. سایه‌نماهای نوفه‌ای حتّی فقط از یک دوربین باعث ایجاد سوراخ‌هایی در حجم سه‌بعدی بازسازی‌شده می‌شود که به‌طور قابل ملاحظه‌ای بیان حجمی را نادرست می‌سازد. برای مقابله با این مشکل، روشی به نام شبکهٔ اشغال احتمالاتی[2] در (*Franco and Boyer 2005*) معرفی شده است که در آن پوستهٔ قابل مشاهده معادل با درنظرگرفتن سطح همسان[3] از چگالی در یک احتمال معین به‌دست می‌آید. هنگامی‌که حجم بازیابی شد، ردیابی شکل سه‌بعدی با استفاده از روش بهینه‌سازی تصادفی کاهشی متا[4] (*Kehl, Bray et al. 2005*) یا الگوریتم نزدیک‌ترین نقطه تکراری[5] (*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 درنظرگرفتن اجتماع مشاهدات سطح پایین تصویر در تمام نماها استفاده می‌کنند. همچنین، این روش‌ها برخلاف روش‌های مبتنی بر پوستهٔ 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*

[1] Visual hull

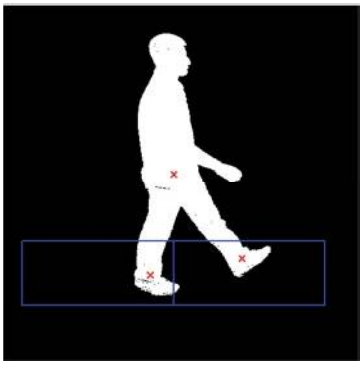
[2] Probabilistic occupancy grid
[3] Isosurface
[4] Stochastic Meta Descent (SMD)
[5] Iterative Closest Point (ICP)

*al. 2000; Sigal, Bhatia et al. 2004*) و با استفاده از ترکیب همانندی‌های مبتنی بر هر نما می‌توانند یک معیار کلی را برای تطبیق حالت استنتاج نمایند. در روش پیشنهادی نیز از افکنش مدل اسکلت سه‌بعدی انسان بر تصاویر سه نمای 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 سه نمای دوربین‌ها در نظر گرفته شده است و تطبیق حالت سه‌بعدی انسان با ویژگی‌های تصاویر سه دوربین با استفاده از کمینه‌سازی تابع هدف ترکیب‌شده انجام می‌شود.

در (*Hofmann and Gavrila 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، جهت بازیابی حالت شخص در یک فریم مشخص، از یک مرحله تولید فرض که در آن حالت‌های سه‌بعدی نامزد بر اساس مقایسهٔ سلسله‌مراتبی شکل در هر دوربین ایجاد می‌شود، استفاده شده است. سپس در مرحلهٔ تأیید فرض، حالات سه‌بعدی نامزد به نماهای دوربین‌های دیگر باز افکنده شده و بر طبق یک اندازهٔ شباهت، رتبه‌بندی می‌شوند. در ادامه، فریم‌هایی که تخمین حالت برای آنها با دقّت خوبی انجام گرفته است، به تعداد مشخّصی 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 تصویربردار زمان رفت و برگشت (*TOF*)[1] معرفی شده است. سامانهٔ پیشنهادی آنها در سه مرحلهٔ اصلی انجام می‌شود. در مرحلهٔ اوّل با استفاده از یک آشکارساز مبتنی بر چارچوب بیزین که از اطلاعات مکانی و زمانی در یک رشتهٔ ویدیویی استفاده می‌کند، نقاط کلیدی بدن انسان در فضای دوبعدی شناسایی و ردیابی می‌شوند. در مرحلهٔ دوم با استفاده از قیدهای حرکتی بدن انسان، نقاط دوبعدی شناسایی‌شده به مدل سه‌بعدی حرکتی انسان نگاشت می‌شوند. در مرحلهٔ سوم نیز نقاطی که توسط آشکارساز دوبعدی به علّت انسداد شناسایی نشده‌اند، تخمین زده شده و ابهاماتی که در مرحلهٔ شناسایی نقاط کلیدی وجود دارد از بین می‌رود. مدل اسکلت سه‌بعدی

[1] Time-of-flight (TOF)



**(شکل۲): نمایش نقاط کلیدی سایه‌نما**

در روش آنها شامل چهارده نقطهٔ کلیدی بدن انسان است که باتوجّه به شباهت زیاد مدل اسکلت آنها با مدل اسکلت پیشنهادی در این مقاله، می‌توان نتایج الگوریتم پیشنهادی این مقاله را با روش آنها به‌صورت کمّی مقایسه کرد.

## ۳- استخراج ویژگی

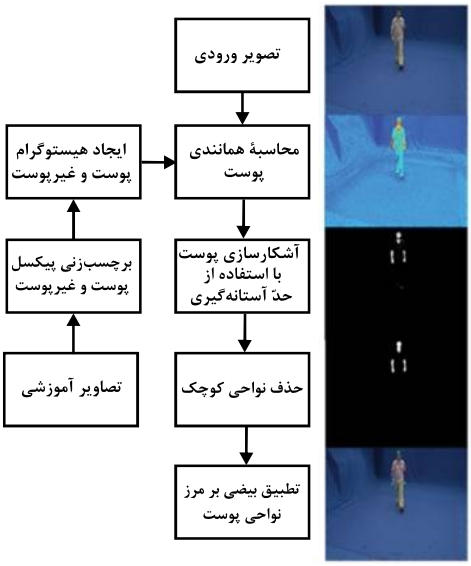
آشکارسازی قسمت‌های مختلف بدن، یکی از روش‌های رایج برای راه‌اندازی خودکار تخمین حالت انسان است (*Sigal, Bhatia et al. 2004; Lee and Nevatia 2009*) و برای کاربردهایی که نیاز به ارزش‌دهی اوّلیه خودکار دارند، مهم است. در روش پیشنهادی، با استفاده از استخراج ویژگی از سایه‌نما و آشکارسازی پوست، نامزدهایی برای موقعیت مفاصل بدن مانند سر، گردن، آرنج و زانوها تخمین زده می‌شود. بنابراین استخراج ویژگی‌هایی که تخمین اوّلیه 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 مرحلهٔ 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 صفحهٔ تصویر مکان‌یابی می‌شود که باعث کاهش فضای جستجوی الگوریتم برای پیکربندی صحیح بدن می‌شود. تفریق

پس‌زمینه[1] 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 (*GkaleIis, Kim et al. 2009*) که دارای پس‌زمینهٔ ثابت آبی رنگ است، می‌توان با استفاده از یک حدّ آستانه‌گیری ساده، سایه‌نمای انسان از پس‌زمینه را به‌راحتی استخراج کرد.

درصورت 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 به‌علّت وجود مانع، گسستگی در سایه‌نما ایجاد شود که حتّی با استفاده از عملگرهای مورفولوژی نیز سایه‌نمای تمیز و عاری از نوفه به‌دست نیاید. اوّلین ویژگی، مرکزثقل[2]، سایه‌نمای باینری است. این ویژگی اطّلاعاتی در مورد کلیت انسان که چگونه در زمان جابه‌جا می‌شود و همچنین یک تقریب مکانی از مفصل ران[3] 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 اوّلیه محل اسکلت بدن انسان می‌شود. کل سطح سایه‌نمای ایجاد شده از فرآیند تفریق پس‌زمینه برای محاسبهٔ مرکزثقل مورد استفاده قرار می‌گیرد. مقدار متوسط تمام موقعیت‌های پس‌زمینه در محور افقی و عمودی تصویر، معرّف مکان مرکز ثقل سایه‌نما در تصویر است. مرحلهٔ بعدی پیداکردن موقعیت پا در تصویر سایه‌نماست. مرکزثقل یک پا، تقریبی مکانی از موقعیت قوزک پا در تصویر سایه‌نماست. موقعیت قوزک پا در روشی به‌طور کامل مشابه با محاسبهٔ مرکزثقل سایه‌نما به‌دست می‌آید، با این تفاوت که سطحی که برای محاسبهٔ مرکزثقل به‌کار می‌رود، باید محدود شود. . نواحی در نظر گرفته شده به‌صورت مستطیل و نقاط کلیدی با علامت ضرب‌در در (شکل ۲) نشان داده شده‌اند.




**(شکل۳): آشکارسازی پوست**

## ۳-۲- آشکارسازی پوست

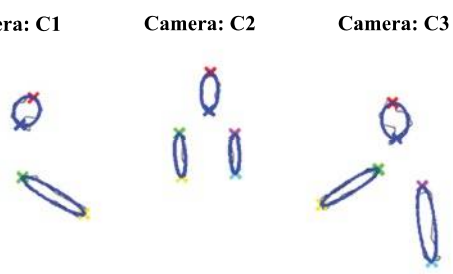
یکی از فرض‌های رایج این است که بعضی از قسمت‌های بدن انسان توسط لباس پوشش داده نمی‌شود،که در این صورت می‌توان از رنگ پوست به‌عنوان یک ویژگی تشخیص برای این قسمت‌ها استفاده کرد. ویژگی‌های رنگ پوست سیگنال‌های مهمی دربارهٔ موقعیت سر، بازوها و بعضی از اوقات پاها می‌دهند. برای آشکارسازی پیکسل‌های پوست از بخشی 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 درنظر 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 ۳۲×۳۲×۳۲ هستند. عدد ۳۲ در هر محور بیان‌گر تعداد سطوح کوانتیزاسیون در رنگ اصلی متناظر است. همان‌طور که می‌دانیم هر تصویر رنگی دارای سه مؤلّفهٔ R، G و B برای هر پیکسل است. هر یک از مقادیر R، G و B دارای گسترهٔ ۰ تا ۲۵۵ هستند. یعنی ۲۵۶ سطح کوانتیزاسیون داریم. به‌منظور کاهش حجم محاسبات از ۳۲ سطح کوانتیزاسیون استفاده می‌شود که دقّت خوبی برای تشخیص رنگ پوست از غیر پوست دارد.

[1] Background subtraction
[2] Centroid
[3] Hip

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

$$L^{(i,j)}(r,g,b) = \log\left(\frac{H^{(i,j)}(r,g,b)}{h^{(i,j)}(r,g,b)}\right) \qquad (۱)$$

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 حدّ آستانه‌گیری، 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[1] استفاده شده است، برای جزئیات کامل این روش می‌توان به مراجع (*Pilu, Fitzgibbon et al. 1996; Fitzgibbon, Pilu et al. 1999*) 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 نشان‌دهندهٔ موقعیت‌های آرنج و دست است. درحالی‌که 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 نقطهٔ 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




**(شکل۴): 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. مرحلهٔ آشکارسازی پوست به‌طور خودکار قادر به استخراج این اطّلاعات است. از روی اطّلاعات به‌دست آمده به‌عنوان مثال می‌توان تفسیر کرد که دست چپ در دوربین شمارهٔ ۱ توسط نیم‌تنه دچار انسداد شده است و دیده نمی‌شود، بنابراین در فرآیند بهینه‌سازی حالت دست چپ فقط از اطّلاعات دوربین شمارهٔ ۲ و ۳ استفاده می‌شود.

## ۳-۳- برچسب‌زنی کانتور با استفاده از گراف

کانتور انسان، لبهٔ کل یا طرح کلّی انسان را ارائه می‌دهد؛ امّا اینکه چگونه مرزهای قسمت‌های مختلف بدن انسان در یک کانتور از هم تفکیک شوند، یک مسئلهٔ 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 کلّی مسئله از اینجا آغاز می‌شود که چه‌طور می‌توان روی پیکسل‌های لبه یا کانتور یک شئ 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 شبکهٔ 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، مسئلهٔ برچسب‌زنی قسمت‌های مختلف کانتور انسان حل می‌شود. بنابراین نخست، تعاریفی در مورد تئوری گراف بیان می‌شود. در اصطلاحات علمی تئوری گراف،

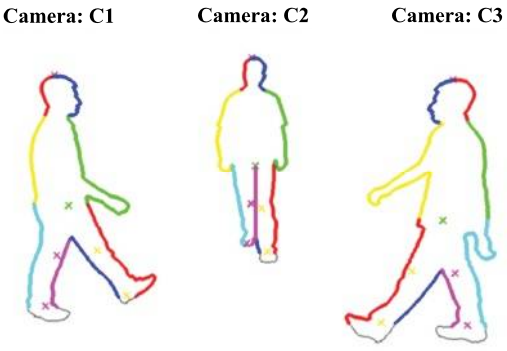
[1] Direct least-square fitting

مدل‌سازی به‌صورت مجموعه‌ای از رأس‌ها[1] (V) و لبه‌ها[2] (E) در نظر گرفته می‌شود که لبه‌ها معرّف اتصالات بین رأس‌ها هستند. بنابراین یک گراف $G$ به‌صورت $G=(V,E)$ نشان داده می‌شود. یک مسیر، یک رشته از رأس‌ها $\langle v_0,v_1,...,v_k \rangle$ در گراف $G=(V,E)$ است؛ درصورتی‌که هر یک از لبه‌های $(v_i,v_{i+1})$ در مجموعه $E$ قرار داشته باشند (یعنی در رشته مربوطه هر رأس به رأس بعدی متصل باشد). در مسئله کوتاه‌ترین مسیر، به هر لبه $(u,v)$ یک وزن $w(u,v)$ داده می‌شود. وزن یک مسیر (یا طول مسیر) برابر مجموع وزن‌های لبه‌های مسیر می‌باشد:

$$w(p)=\sum_{i=0}^{k-1} w(v_i,v_{i+1}) \qquad (۲)$$

کوتاه‌ترین مسیر از رأس $u$ به رأس $v$ مسیری است که دارای کمینه‌ترین وزن در بین تمام مسیرهای ممکن باشد. برای یادگیری الگوریتم‌های تئوری گراف و نحوهٔ پیاده‌سازی آنها می‌توان به مرجع (*Siek, Lee et al. 2002*) مراجعه کرد. برای مدل‌سازی تصویر باینری کانتور انسان با استفاده از یک گراف، ابتدا باید رأس‌ها، لبه‌ها و وزن بین رأس‌ها تعریف شوند. پیکسل‌هایی که در تصویر متعلّق به کانتور انسان هستند (یعنی مقدار یک دارند) به‌عنوان رأس‌های گراف در نظر گرفته می‌شوند. اتصال بین رأس‌ها نیز با توجه به اتصال هشت‌گانه پیکسل متناظرشان در نظر گرفته می‌شود؛ بنابراین فقط رأس‌هایی که پیکسل متناظرشان دراتصال هشت‌گانهٔ یکدیگر قرار دارند، به هم متّصل هستند. اختصاص وزن بین رأس‌های متصل، به‌صورت فاصلهٔ اقلیدسی پیکسل متناظرشان تعریف می‌شود. بنابراین، با استفاده از مدل‌سازی تصویر کانتور انسان به‌وسیلهٔ یک گراف و مشخّص‌کردن نقاط مبدأ و مقصد، می‌توان عملیات برچسب‌زنی بین نقاط مبدأ و مقصد مشخّص‌شده را به‌راحتی انجام داد. (شکل ۵) نتایج پیاده‌سازی روش پیشنهادی را در سه نمای متفاوت از دوربین‌ها نشان می‌دهد. در این شکل مرزهای قسمت‌های مختلف بدن مانند نیم‌تنه، سر و پاها با رنگ‌های مختلف نشان داده شده‌اند، پیکسل‌های برچسب‌زده شده در این مرحله به‌عنوان ورودی الگوریتم بهینه‌سازی حالت برای محاسبهٔ تابع هدف در نظر گرفته می‌شوند. همان‌طور که در (شکل ۵) مشخّص است پیکسل‌های کانتور دست به‌عنوان پیکسل نیم‌تنه و یا در بعضی موارد به‌عنوان

[1] Vertex (V)
[2] Edges (E)

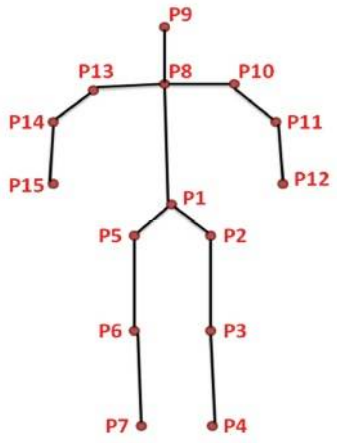



**(شکل۵): نتایج پیاده‌سازی روش برچسب‌زنی کانتور بدن انسان در ۳ نمای دوربین متفاوت**

پیکسل پا برچسب‌زده شده‌اند که با تعریف یک تابع هدف مناسب که در بخش‌های بعدی ذکر می‌شود، می‌توان نشان داد که اثر این پیکسل‌ها در محاسبه تابع هدف بدون تأثیرند. برچسب‌زنی پیکسل‌های دست نیز به‌طور جداگانه بر اساس پیکسل‌های پوست انجام می‌شود.

## ۴- مدل‌سازی بدن انسان

در روش پیشنهادی این مقاله، 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. (شکل ۶) مدل در نظر گرفته‌شده برای بدن انسان را نشان می‌دهد، مدل سه‌بعدی اسکلتی پیشنهادی به‌صورت مجموعه‌ای از پانزده نقطه در فضای سه‌بعدی در نظر گرفته شده است این نقاط درواقع معرف مفاصل کلیدی بدن انسان می‌باشند که در (شکل ۶) به‌صورت P1 تا P15 نشان داده شده‌اند. این مدل شامل چهارده خط در فضای سه‌بعدی می‌باشد. هر خط نماینده یکی از استخوان‌های بدن انسان است و هر کدام از آنها تعدادی درجهٔ آزادی دارند که جزئیات آن‌ها در (جدول ۱) آمده است. درواقع، موقعیت هر جسم صلب در فضای سه‌بعدی و سامانهٔ مختصات جهانی توسط شش پارامتر مشخص می‌شود که سه پارامتر X,Y,Z انتقال مرکز ثقل جسم نسبت به مبدأ مختصات را نشان می‌دهد و پارامترهای Rx,Ry,Rz نشان‌دهندهٔ میزان چرخش جسم نسبت به محورهای مختصات x,y,z هستند که در این مقاله از بیان زوایای اویلر به‌منظور نشان‌دادن چرخش یک جسم استفاده شده است. همچنین برای هر قسمت صلب از




**(شکل۶): مدل بدن انسان**

مدل مفصل‌دار انسان یک اندیس در نظر گرفته شده است. به‌طور مثال، منظور از $R_{X1}$ زاویهٔ 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 سه دوربین مختلف و مدل اسکلتی سه‌بعدی استنتاج می‌شود. بنابراین، پس از استخراج ویژگی‌های مناسب، این ویژگی‌های از هر تصویر دوربین 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 ۲۹بعدی از نظر محاسباتی پرهزینه و اغلب مهارنشدنی 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 نقطهٔ کلیدی بدن در فضای سه‌بعدی که در بخش قبل معرفی شدند، باتوجّه به روند سلسله‌مراتبی (شکل ۷)، ابتدا نقطهٔ P1 که معرّف ریشهٔ درخت سینماتیک بدن است، تخمین زده می‌شود. سپس با استفاده از نقطهٔ P1، نقطهٔ پایین گردن (نقطهٔ P8) و مفصل ران چپ (نقطهٔ P2 و P5) محاسبه می‌شوند. به همین

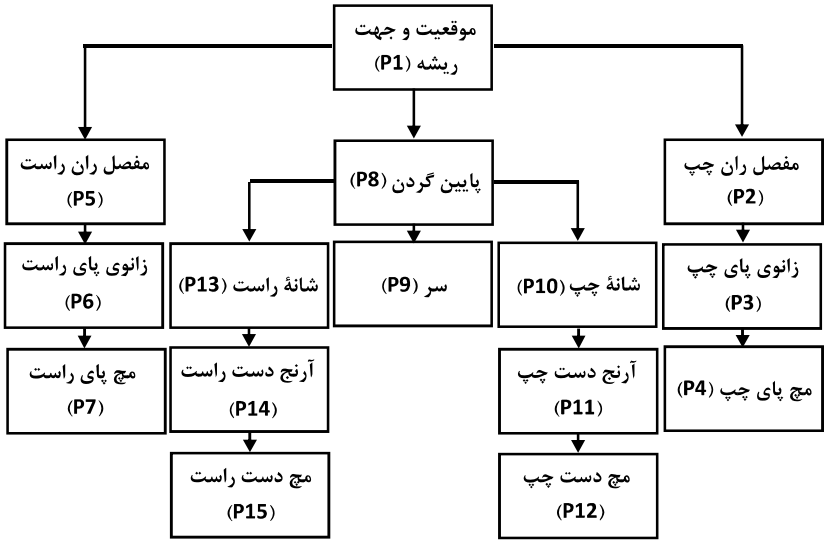
**(جدول ۱): درجه‌های آزادی استخوان‌های مدل اسکلت پیشنهادی**

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

ترتیب، سایر نقاط باتوجّه به ترتیب سلسله‌مراتبی در نظر گرفته شده محاسبه می‌شوند. جستجوی ۲۹ پارامتر مدل بدن انسان در شش مرحلهٔ جدا از هم انجام می‌شود (که معادل با جدا کردن فضای جستجوی ۲۹بعدی به شش زیرفضای جستجو است). در مرحلهٔ اوّل نیم‌تنه، مرحلهٔ دوم سر و گردن، مرحلهٔ سوم و چهارم دست چپ و راست و درنهایت در مرحلهٔ 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 مهارکردن جستجوی فضای ابعاد بالای حالت بدن انسان استفاده شده است. بنابراین تابع هدف برای




**(شکل۷): سلسله‌مراتب در مدل سینماتیک بدن انسان**

هر یک از اعضای بدن جداگانه تعریف می‌شوند. پارامترهای بدن انسان در زمان $t$ به‌صورت $X_t$ تعریف می‌شود که $X_t$ شامل ۲۹ پارامتر حالت بدن انسان است. باتوجّه به روند جستجوی سلسله‌مراتبی، پارامترهای حالت هر یک از اعضای بدن، جداگانه بهینه می‌شوند. بنابراین، پارامترهای مدل انسان به‌صورت $X_t=\{X_{i,t}\}, i=1,...,14$ در نظر گرفته می‌شود که $X_{i,t}$ بیان‌گر پارامترهای عضو iاُم 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 $F_i\left(X_{i,t},I_t\right)$ برای عضو iاُم بدن انسان است. تابع هدف $F_i\left(X_{i,t},I_t\right)$ به‌صورت زیر تعریف می‌شود:

$$F_i\left(X_{i,t},I_t\right)=F_{1,i}\left(X_{i,t},I_{1,t}\right)+F_{2,i}\left(X_{i,t},I_{2,t}\right)+F_{3,i}\left(X_{i,t},I_{3,t}\right) \quad (۳)$$

در رابطهٔ فوق $F_{1,i}\left(X_{i,t},I_{1,t}\right)$، $F_{2,i}\left(X_{i,t},I_{2,t}\right)$ و $F_{3,i}\left(X_{i,t},I_{3,t}\right)$ به‌ترتیب معرّف تابع هدف عضو iاُم در صفحهٔ تصویر دوربین شمارهٔ ۱، دوربین شمارهٔ ۲ و دوربین شمارهٔ ۳ می‌باشند و $I_{1,t}$، $I_{2,t}$ و $I_{3,t}$ به‌ترتیب بیان‌گر تصویر دوربین‌های شماره ۱ ، ۲ و ۳ در زمان $t$ هستند. حال با تعریف $F_{1,i}\left(X_{i,t},I_{1,t}\right)$ می‌توان این تعریف را به بقیّهٔ حالات تعمیم داد.

همان‌طور که در بخش‌های قبل ذکر شد، مدل بدن انسان به‌صورت مجموعه‌ای از نقاط (مفاصل) و خطوط (استخوان‌های صلب بدن) در نظر گرفته می‌شود. افکنش استخوان iاُم در تصویر دوربین شمارهٔ ۱ را با $Seg_{1,i}$ نشان می‌دهیم که به‌صورت یک خط دوبعدی در صفحهٔ تصویر درنظرگرفته می‌شود. تعداد $N$ نقطه روی این خط با فاصلهٔ یکسان درنظرگرفته می‌شود؛ این نقاط با $S_{1,i,j}, j=1,...,N$ نشان داده می‌شوند. پیکربندی صحیح اسکلت سه‌بعدی، زمانی به‌دست می‌آید که افکنش 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 $B_{1,1,i,k}, k=1,...,M_1$ و $B_{1,2,i,k}, k=1,...,M_2$ 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 $F_{1,i}\left(X_{i,t},I_{1,t}\right)$ به‌صورت زیر تعریف می‌شود:

$$F_{1,i}\left(X_{i,t},I_{1,t}\right)=\left|\sum_{j=1}^{N}\underset{k}{Min}\left(\left\|S_{1,i,j}-B_{1,1,i,k}\right\|\right)-\sum_{j=1}^{N}\underset{k}{Min}\left(\left\|S_{1,i,j}-B_{1,2,i,k}\right\|\right)\right| \quad (۴)$$

محاسبهٔ $F_{2,i}\left(X_{i,t},I_t\right)$ و $F_{3,i}\left(X_{i,t},I_t\right)$ نیز باتوجّه به تصاویر متناظرشان مشابه با رابطهٔ (۴) انجام می‌شود. پیکربندی صحیح عضو iاُم در زمان $t$ با کمینه‌سازی تابع هدف $F_i\left(X_{i,t},I_t\right)$ (مجموع سه تابع هدف محاسبه شده در سه نمای مختلف) به‌صورت زیر به‌دست می‌آید:

$$X_{i,t}^{*} = \arg\min_{X_{i,t}} \left( F_i \left( X_{i,t}, I_t \right) \right) \quad (۵)$$

در رابطهٔ (۵) $X_{i,t}^{*}$ معرف پارامترهای بهینهٔ عضو iاُم در زمان $t$ است. باتوجّه 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 $X_t^{*}$ در زمان $t$ با استفاده از کنار هم قرار دادن پارامترهای بهینه حالت هر یک از عضوهای بدن به‌صورت رابطهٔ (۶) به‌دست می‌آید:

$$X_t^{*} = \left\{ X_{i,t}^{*} \right\}, i = 1,...,14 \quad (۶)$$

## ۵-۳- الگوریتم کمینه‌سازی تابع هدف

پس از تعریف تابع هدف برای هر یک از اعضای بدن، کمینه‌سازی توابع هدف معرّفی‌شده، به‌منظور به‌دست آوردن پارامترهای بهینهٔ 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 سیمپلکس[1] 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 اوّلیهٔ $x_0$ ایجاد می‌کند. هدف از هر تکرار الگوریتم، جایگزین کردن بدترین نقطهٔ با یک نقطه بهتر است. این فرآیند تکرار تا زمانی‌که تمام $n+1$ نقطه روی یک موقعیت با یک تلورانس مشخّص هم‌گرا شوند، ادامه پیدا می‌کند.

در هر تکرار الگوریتم، مقدار تابع در هر $n+1$ رأس سیمپلکس ارزیابی می‌شود و این رأس‌ها به‌صورت $f(x_1) \le f(x_2) \le ... \le f(x_{n+1})$ مرتب می‌شوند. بهترین رأس $x_1$، بهترین رأس دوم $x_2$ و به همین ترتیب بدترین رأس $x_{n+1}$ است. در هر تکرار از الگوریتم، بدترین رأس ($x_{n+1}$) در سیمپلکس کنار گذاشته می‌شود و یک نقطهٔ دیگر جایگزین آن می‌شود و یا 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 انعکاس[2] ($\rho$)، توسعه[3] ($\chi$)، ادغام[4] ($\gamma$) و انقباض[5] ($\sigma$) هستند. این پارامترها در الگوریتم استاندارد *Nelder-Mead* به‌صورت رابطهٔ (۷) درنظر گرفته می‌شوند:

$$\rho = 1,\ \chi = 2,\ \gamma = 0.5,\ \sigma = 0.5 \quad (۷)$$

مرکزثقل تمام نقاط به جز نقطهٔ $x_{n+1}$ با استفاده از رابطهٔ $\bar{x} = \sum_{i=1}^{n} x_i / n$ محاسبه می‌شود. تعدادی رأس دیگر با استفاده از $\bar{x}$ و پارامترهای الگوریتم محاسبه می‌شوند که محاسبهٔ مقدار این رأس‌ها در روابط زیر نشان داده می‌شود:

$$x_r = (1+\rho)\bar{x} - \rho x_{n+1} \quad (۸)$$

$$x_e = \bar{x} + \chi(x_r - \bar{x}) \quad (۹)$$

$$x_c = x_{n+1} + \gamma(\bar{x} - x_{n+1}) \quad (۱۰)$$

$$x_i = x_1 + \sigma(x_i - x_1),\ i \in \{2,...,n+1\} \quad (۱۱)$$

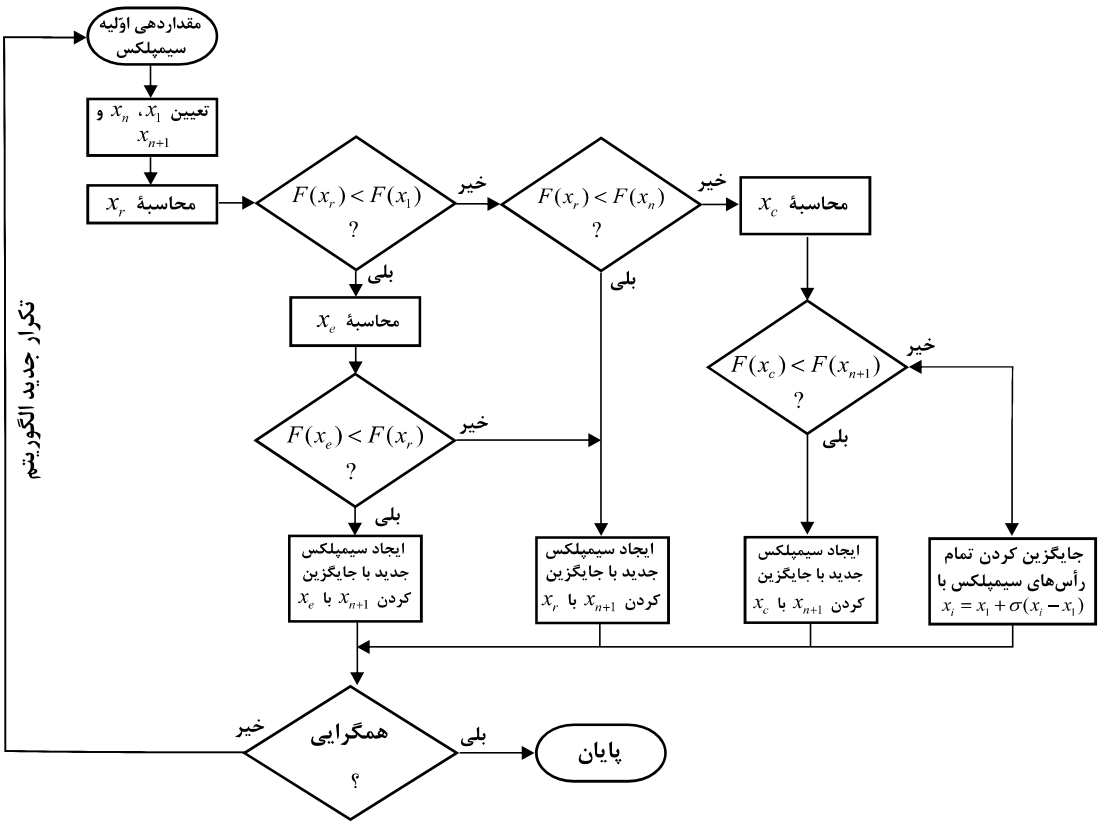
(شکل ۸) نمودار جعبه‌ای الگوریتم جستجوی *Nelder-Mead* را نشان می‌دهد. رأس‌های سیمپلکس در تکرارهای الگوریتم مطابق نمودار جعبه‌ای (شکل ۸) تغییر می‌نمایند تا زمانی که معیار توقف برآورده شود. شرط توقف الگوریتم بدین‌صورت است که قطر سیمپلکسی که در هر تکرار تولید می‌شود، کم‌تر از یک تلورانس مشخّص شود.

## ۶- نتایج پیاده‌سازی

در روش پیشنهادی، ورودی به‌صورت رشتهٔ ویدیویی از سه نمای مختلف در نظر گرفته شده است و خروجی با استفاده از ۲۹ پارامتر حالت سه‌بعدی بدن انسان را بیان می‌کند. تمام مراحل روش پیشنهادی مانند استخراج سایه‌نما، آشکارسازی پوست و برچسب‌زنی کانتور به‌صورت خودکار انجام می‌شوند. در مرحلهٔ آشکارسازی پوست، برای ایجاد توابع هیستوگرام پوست $H$ و هیستوگرام غیر پوست $h$، فریم‌های مختلف از تصاویر دوربین‌های مختلف پایگاه داده *i3DPost*

[1] Simplex

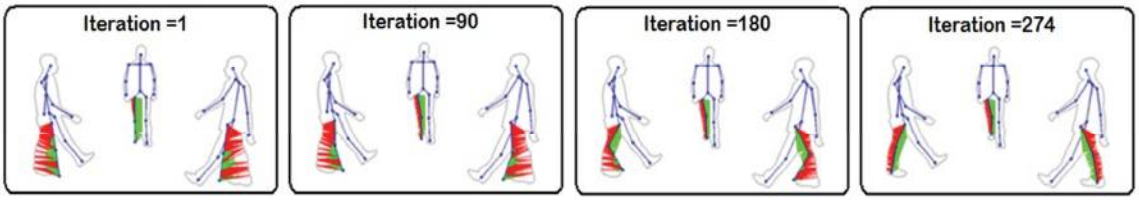
[2] Reflection
[3] Expansion
[4] Contraction
[5] Shrinkage



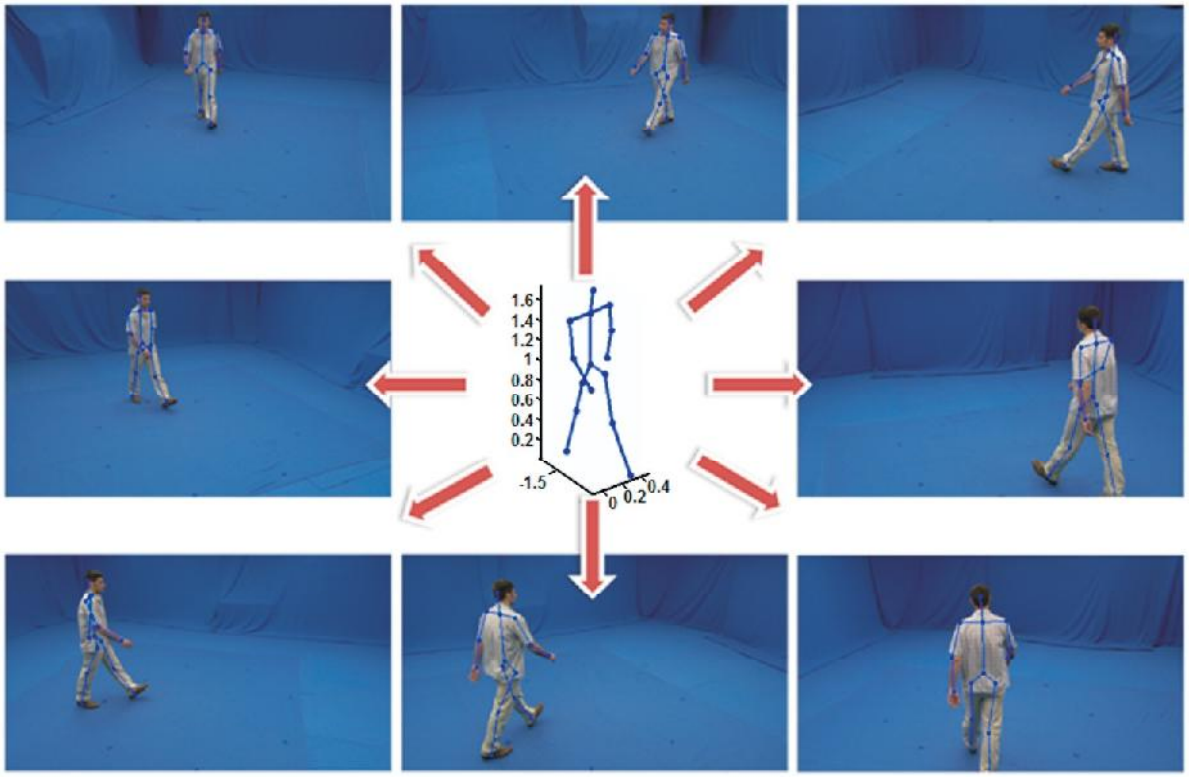

**(شکل۸): نمودار جعبه‌ای الگوریتم جستجوی Nelder-mead**

مورد استفاده قرار گرفته و در مرحلهٔ آموزش، نواحی پوست و غیر پوست به‌صورت دستی انتخاب می‌شود. بدین ترتیب میزان حضور مؤلّفه‌های رنگی (r,g,b) در تصاویر پوست به‌صورت یک هیستوگرام چهاربُعدی استخراج می‌شود که در آن مقدار H در هر نقطهٔ 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.

باتوجّه به اینکه فرض شده شخص در حالت ایستاده است، ابتدا قد شخص تخمین زده می‌شود و طول‌های اوّلیه استخوان‌های اسکلت به مقیاس قد شخص تغییر داده می‌شوند. حالت شخص در هر فریم به‌صورت خودکار تخمین زده می‌شود. روش پیشنهادی بر روی رشتهٔ ویدیویی چنددوربینی راه رفتن یک شخص از پایگاه‌داده i3DPost (*GkaleIis, Kim et al. 2009*) آزمایش شده است. در روش پیشنهادی فقط از سه دوربین استفاده شده است. این دوربین‌ها در پایگاه دادهٔ 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 افکنش مدل اسکلت سه‌بعدی پیشنهادی در سه نمای مختلف تطبیق داده می‌شوند. پس از استخراج ویژگی از تصاویر سه دوربین مختلف، تطبیق بین افکنش 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 باتوجّه به ترتیب سلسله‌مراتبی درخت سینماتیک مدل بدن انسان انجام می‌شود. پیکربندی صحیح اسکلت سه‌بعدی، زمانی به‌دست می‌آید که افکنش 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 بهینهٔ پای راست حاصل شده است. در (شکل ۹) تعداد یکصد نقطه بر روی افکنش خط بیان‌گر پای راست در سه نما از دوربین‌ها در نظر گرفته شده است. خطوط قرمز و سبز، به‌ترتیب بیان‌گر کمینه‌ترین فاصله هر یک از این یکصد نقطه از مرز یک و دوی پای راست می‌باشند. با توجه به تابع هدف تعریف‌شده، هستند.




(شکل۹): نمایش فرآیند تخمین پارامترهای حالت پای راست در تکرارهای مختلف الگوریتم بهینه‌سازی




(شکل۱۰): نمایش حالت سه‌بعدی بازیابی‌شده و افکنش آن بر ۸ صفحه دوربین متفاوت

مجموع تفاضل اندازهٔ خطوط قرمز و سبز در هر نما محاسبه می‌شوند. این محاسبه، در رابطهٔ (۴) برای دوربین شمارهٔ ۱و بخش أم 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 رابطهٔ (۵) پارامترهای حالت بهینهٔ 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؛ به‌طوری که تفاضل اندازهٔ خطوط قرمز و سبز در هر نما کمینه شده است. از این طریق پارامترهای بهینهٔ حالت 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 در هر فریم به‌دست می‌آید که حالت سه‌بعدی بازیابی‌شده درفریم ۵۲ از رشتهٔ ویدیویی راه‌رفتن از پایگاه‌داده i3DPost در (شکل ۱۰) نشان داده شده است.

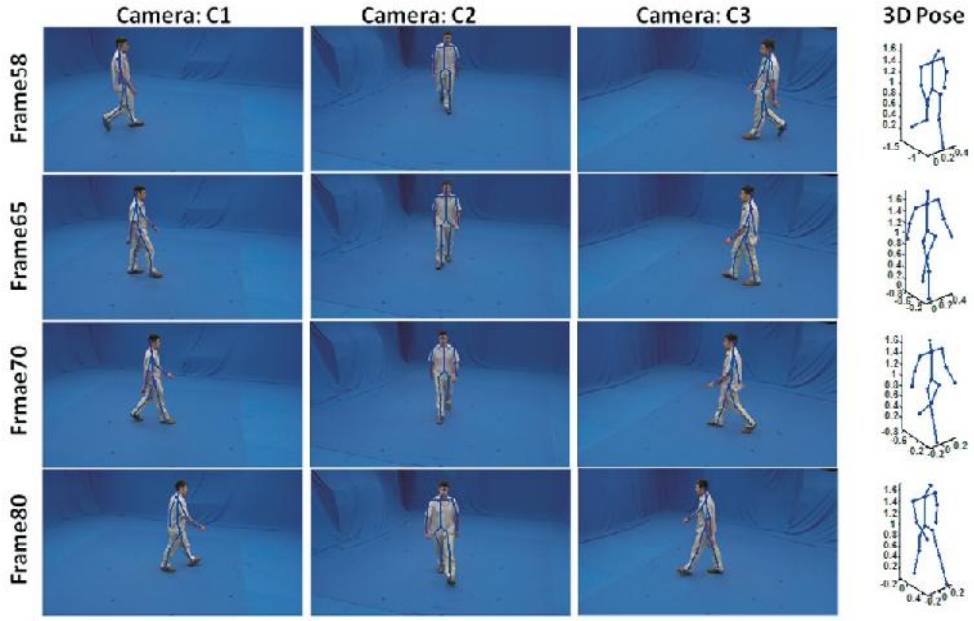
باتوجّه به نامعلوم بودن موقعیت سه‌بعدی واقعی مفاصل بدن شخص، برای ارزیابی کیفی حالت سه‌بعدی بازیابی‌شده، افکنش پیکربندی اسکلت بازیابی‌شده بر روی هشت نما از دوربین‌های مختلف در (شکل ۱۰) نشان داده شده است؛ درصورتی‌که در روش پیشنهادی فقط از سه نما استفاده شده است. نتایج آزمایش، توانمندی روش پیشنهادی در تخمین 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 تخمین‌زده شده در فریم‌های قبلی به‌عنوان شرایط اوّلیه خوبی برای فریم‌های بعدی مورد استفاده قرار می‌گیرند که این در افزایش سرعت هم‌گرایی الگوریتم تأثیرگذار است.

به‌طور اصولی با استفاده از یک سامانهٔ ضبط حرکت مبتنی بر نشانه[1] می‌توان محل واقعی مفاصل کلیدی بدن در فضای سه‌بعدی را به‌دست آورد. باتوجّه 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 نقطهٔ مفروض در مدل اسکلت پیشنهادی برای هر فریم از یک رشتهٔ ویدیویی را به‌دست آورد. فرض کنید محل واقعی پانزده نقطهٔ مفروض در مدل اسکلت پیشنهادی به‌صورت $X_t = \{P_1, P_2, ..., P_{15}\}$ و محل تخمین‌زده شدهٔ این نقاط توسط الگوریتم پیشنهادی به‌صورت $\hat{X}_t = \{\hat{P}_1, \hat{P}_2, ..., \hat{P}_{15}\}$ بیان شوند که در این بیان‌ها $P_k, \hat{P}_k \in \mathbb{R}^3$ موقعیت سه‌بعدی نقطهٔ $k$اُم در سامانهٔ مختصات جهانی را نشان می‌دهند. با مقایسهٔ این داده‌ها از طریق رابطهٔ (۱۲) می‌توان خطای مکان‌یابی الگوریتم پیشنهادی برای هر فریم را محاسبه کرد.

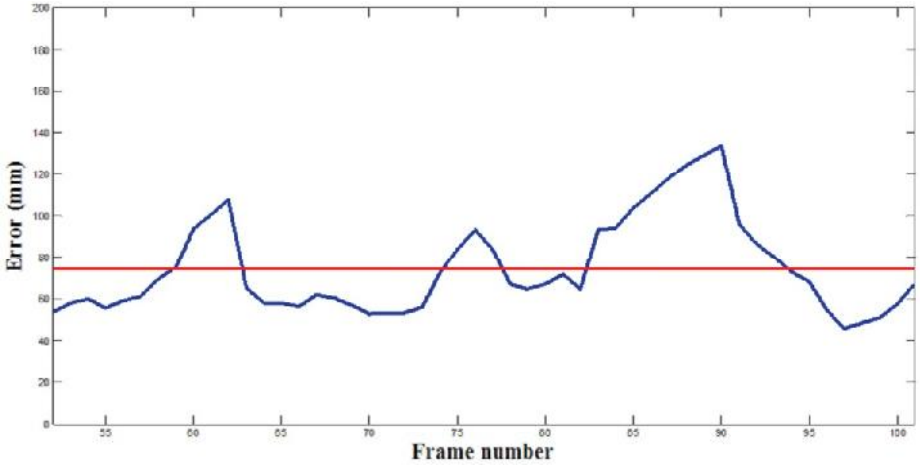
$$Error\left(X_t, \hat{X}_t\right) = \frac{1}{15}\sum_{k=1}^{15}\left\|P_k - \hat{P}_k\right\| \qquad (۱۲)$$

روش پیشنهادی بر روی یک رشتهٔ 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 درنظرگرفته‌شده، این رشتهٔ ویدیویی از فریم ۵۲ تا ۱۰۱ درنظرگرفته‌شده است. نتایج پیاده‌سازی به‌صورت کیفی و کمّی در (شکل‌های۱۱ و ۱۲) نشان داده شده است. در (شکل ۱۱)، حالت سه‌بعدی بازیابی‌شدهٔ 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 پیشنهادی بر حسب میلی‌متر در طول پنجاه فریم مطابق با رابطهٔ (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 رشتهٔ ویدیویی آزمایش‌شده حدود ۷.۵ سانتی‌متر (۷۵ میلیمتر) است. نتایج عددی مربوط به اختلاف مکان پانزده نقطهٔ 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 نقطهٔ کلیدی بدن انسان است که باتوجّه 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 فاصلهٔ 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 دقّت بیش‌تر روش پیشنهادی است. همان‌طور که در (جدول ۲) نیز مشاهده می‌شود از مزایای روش پیشنهادی ما دقّت بسیار خوب در تخمین نقاط مربوط به دست‌هاست درصورتی‌که روش (*Zhu, Dariush et al. 2010*) دقّت پایینی در تخمین این نقاط دارد. از دلایل دقّت 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 توجّه داشت که تابع هدف معرفی‌شده به‌صورت تفاضل فاصلهٔ 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)، دقّت 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 دقّت تخمین در نقاط کلیدی پاها نسبت به سایر

[1] Marker-based motion capture system




(شکل۱۱): نتایج تصویری تخمین حالت انسان در رشتهٔ ویدیویی راه رفتن از پایگاه‌داده i3DPost




(شکل۱۲): نتایج کمی تخمین حالت انسان در رشتهٔ ویدیویی راه رفتن از پایگاه‌داده i3DPost با روش پیشنهادی به‌صورت خطای میانگین در هر فریم

نقاط مدل بدن انسان می باشد. دلیل این مسئله، مواجه‌شدن با انسداد پای چپ و راست در حرکت راه‌رفتن می باشد که 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 با افزایش تعداد دوربین‌ها، می‌توان دقّت تخمین در این نقاط را نیز بهبود داد. در مجموع از مقایسهٔ روش پیشنهادی با روش ( *Zhu, Dariush et al. 2010*) می‌توان نتیجه گرفت که بهره‌گیری از سامانهٔ تخمین حالت چنددوربینی و استفاده از اطّلاعات مرزی در سایه‌نماها در مقایسه با یک سامانهٔ تخمین مبتنی بر تصاویر عمق دقّت بیشتری برخوردار است. همچنین، به‌منظور ارزیابی دقت عملکرد روش پیشنهادی، مقدار میانگین خطای مکانی حالت انسان با روش مقاله ( *Hofmann and Gavrila 2012*) که یک روش تخمین حالت مبتنی بر چند دوربین می‌باشد، مورد مقایسه قرار گرفته است. لازم به توضیح است که پایگاه داده مورد استفاده در این روش HumanEva است که دارای داده‌های آموزشی، آزمایشی و همچنین برنامه‌هایی برای محاسبهٔ برآورد خطاست. متأسفانه دسترسی به این پایگاه داده برای دانشگاه‌های کشور ایران محدود شده است و درعمل امکان پیاده‌سازی روش پیشنهادی با این پایگاه داده و مقایسه روشها در شرایط یکسان میسر نشد؛ با وجود این، بررسی سکانس‌های ویدیویی مختلف و مقایسه مقدار میانگین خطای مکانی حالت انسان در شرایط فیلم‌برداری چنددوربینی که در هر دو پایگاه داده HumanEva [1] و i3DPost [2] لحاظ شده است، می‌تواند معیار به‌نسبه خوبی

[1] http://vision.cs.brown.edu/humaneva/

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

(جدول۲): میانگین و انحراف معیار خطای مکانی اجزای بدن انسان در مدل اصلی و تخمینی در کل فریم‌های حرکتی (میلی متر)

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

(جدول۳): میانگین خطای مکانی تخمین حالت انسان برای چند رشتهٔ ویدیویی

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

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 Gavrila 2012)* است.

آزمایش روش پیشنهادی بر روی یک پردازنده ۲.۲ گیگا هرتزی اینتل با حافظهٔ رم یک گیگا بایت انجام شده است. زمان کلی پردازش برای هر فریم حدود ۱۲۹ ثانیه است، این زمان محاسباتی، سامانهٔ معرفی‌شده را برای کاربردهای بلادرنگ[1] نامناسب می‌سازد. البته باید توجّه کرد که پیاده‌سازی بر روی وضوح تصویری بالا ۱۹۲۰×۱۰۸۰ انجام شده است و با بهینه‌سازی برنامه‌های نوشته شده می توان زمان محاسباتی نیز را کاهش داد.

## ۷- نتیجه‌گیری

در این مقاله، روشی جدید برای تخمین حالت سه‌بعدی انسان در رشتهٔ 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 قسمت‌های مختلف بدن و از روش‌های بهینه‌سازی مستقیم برای تخمین حالت انسان استفاده شده است. سامانهٔ پیشنهادی قادر به تشخیص انسداد و تخمین به‌نسبه دقیق اعضای بدن مانند نیم‌تنه، سر و دست‌هاست. از مزایای روش معرفی‌شده این است که باتوجّه به معرّفی روش برچسب‌زنی کانتور و تابع هدف معرفی‌شده، حتّی با استفاده از ساده‌ترین روش‌های بهینه‌سازی، الگوریتم قادر به هم‌گرایی به حالت صحیح است. از دیگر مزایای روش پیشنهادی ارزش‌دهی اوّلیهٔ خودکار آن است؛ به‌طوری‌که روش معرفی‌شده قادر به تخمین خودکار حالت انسان در فریم اوّل یک رشتهٔ ویدیویی چنددوربینی است. درحالی‌که سامانهٔ معرفی‌شده، قادر به تخمین حالت انسان به‌طور مؤثّر 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 انسداد 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 توجّه داشت که در این حالت می‌توان با افزایش تعداد دوربین‌ها و یا استفاده از ویژگی‌های لبهٔ تصویر این نقصان را برطرف کرد. یکی دیگر از مواردی که روش پیشنهادی را دچار شکست می‌کند این

[1] Real-time

است که در صحنه‌های واقعی خارج از محیط آزمایشگاه، به‌دلیل وجود پس‌زمینهٔ 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 به‌صورت 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.

GkaleIis, 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.

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 رشتهٔ مهندسی برق، مخابرات از دانشگاه صنعتی امیرکبیر (پلی‌تکنیک تهران) اخذ و سپس با ادامهٔ تحصیل در دانشگاه صنعتی سهند، مدرک کارشناسی ارشد خود را در گرایش مخابرات سیستم در سال ۱۳۹۰ دریافت کرد. زمینه‌های تحقیقاتی مورد علاقهٔ ایشان بینایی کامپیوتر، پردازش تصویر و شناسایی الگو است.

نشانی رایانامهٔ ایشان عبارتست از:

**g_khademi@sut.ac.ir**



**حسین ابراهیم‌نژاد** مدرک کارشناسی و کارشناسی ارشد خود را به‌ترتیب در سال‌های ۱۳۷۲ و ۱۳۷۵ در رشتهٔ 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 علاقهٔ 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 رایانامهٔ ایشان عبارتست از:

**ebrahimnezhad@sut.ac.ir**