

استفاده از برآوردهای پویای دست مبتنی

بر مدل، برای تقلید عملکرد بازوی انسان توسط

ربات با داده‌های کینکت

مرویم زارع مهرجردی* و مهدی رضائیان

پردیس فنی مهندسی، گروه مهندسی کامپیوتر، دانشگاه یزد، یزد، ایران

چکیده

برای اجرای فرآیند ضبط حرکت، لازم است داده‌های مناسب، در طول زمان با دنبال کردن نقاط کلیدی از هدف مورد نظر استخراج شوند. با این داده‌ها و طی یک سری عملیات پس‌پردازشی کارهای زیادی از جمله ساخت مجدد آن حرکت در فضای سه‌بعدی می‌توان انجام داد. در این مقاله یک الگوی برآوردهای پویای دست مبتنی بر مدل با استفاده از روش ضبط حرکت بدون نشانه‌گذاری ارائه می‌شود. در این پژوهش حرکات بازوی عامل انسانی در قالب دنباله‌ای از تصاویر رنگی به همراه داده‌های عمق و اسکلت به دست آمده از کینکت (ابزاری برای ضبط حرکت بدون نشانه‌گذاری) با سرعت سی فریم در ثانیه به عنوان داده‌های ورودی استفاده شده‌اند. الگوی پیشنهادی، ویژگی‌های زمانی و مکانی از دنباله تصاویر ورودی استخراج می‌کند و روی تعیین موقعیت نوک انگشتان شست و اشاره و به دست آوردن زوایای مفاصل ربات، به منظور تقلید حرکت بازوی عامل انسانی در سه بعد در یک محیط کنترل نشده تمرکز دارد. در این پژوهش از بازوی ربات واقعی RoboTEK II ST240 استفاده شده و حرکات بازوی عامل انسانی به حرکات تعریف شده برای این بازوی ربات محدود شده است. بردار ویژگی جهت برآوردهای هر فریم، به مختصات x ، y و z عمقداری مفاصل و مختصات نوک انگشتان شست و اشاره نیازدارد. از داده‌های عمق و اسکلت برای تعیین زوایای مفاصل ربات استفاده می‌شود؛ ولی تعیین نوک انگشتان به طور مستقیم با داده‌های موجود امکان‌بزیر نیست؛ از این‌رو سه رویکرد برای شناسایی نوک انگشتان شست و اشاره با استفاده از داده‌های موجود ارائه می‌شود. در این رویکردها از مفاهیمی همچون آستانه‌گیری، لبه‌یابی، ساخت پوسته محدب، مدل کردن رنگ پوست و تفریق پس‌زمینه استفاده می‌شود. در پایان برای تقلید حرکت، با استفاده از بردارهای ویژگی بهزای هر فریم، حالت متناظر بر روی بازوی ربات اعمال می‌شود. برای ارزیابی تقلید حرکت، مسیرهای طی شده توسط قسمت نهایی دست عامل انسانی و قسمت مجری نهایی بازوی ربات با هم مقایسه شده‌اند. نمودارهایی که میزان تغییرات زوایای مفاصل را برای این دو مورد نشان می‌دهند، گویای مؤثربودن الگوی پیشنهادی در تقلید عملکرد بازوی انسانی است.

وازگان کلیدی: برآوردهای داده عمقداری، ضبط حرکت بدون نشانه‌گذاری، کینکت، مدل سه‌بعدی

Using of Model Based Hand Poses Estimation for Imitation of User's Arm Movements by Robot Arm

Maryam Zare Mehrjardi* & Mehdi Rezaeian

College of Engineering, Dept. of Electrical and Computer Engineering, University of Yazd, Yazd, Iran

Abstract

Pose estimation is a process to identify how a human body and/or individual limbs are configured in a given scene. Hand pose estimation is an important research topic which has a variety of applications in human-computer interaction (HCI) scenarios, such as gesture recognition, animation synthesis and robot control.

* Corresponding author

نویسنده عهده‌دار مکاتبات

However, capturing the hand motion is quite a challenging task due to its high flexibility. Many sensor-based and vision-based methods have been proposed to fulfill the task.

In sensor-based systems, specialized hardware is used for hand motion capture. Generally, vision-based hand pose estimation methods can be divided into two categories: appearance-based methods and model-based methods. In appearance-based approaches, various features are extracted from the input images to estimate the hand pose. Usually a lot of training samples are used to train a mapping function from the features to the hand poses in advance. Given the learned mapping function, the hand pose can be estimated efficiently. In model-based approaches the hand pose is estimated by aligning a projected 3D hand model to the extracted hand features in the inputs. Therefore, the desired information to be provided includes state at any time. These methods require a lot of calculations which are not possible in practice to implement them immediately. Hand pose estimation using (color/depth) images consist of three steps:

- 1) Hand detection and its separation
- 2) Feature extraction
- 3) Setting the parameters of the model using extracted feature and updating the model

To extract necessary features for pose estimation, depending on used model and usage of hand gesture analysis, features such as fingertips position, number of fingers, palm position and joint angles are extracted. In this paper a model-based markerless dynamic hand poses estimation scheme is presented. Motion Capture is the process of recording a live motion event and translating it into usable mathematical terms by tracking a number of key points in space over time and combining them to obtain a single 3D representation of the performance. The sequence of depth images, color images and skeleton data obtained from Kinect (a new tool for markerless motion capture) at 30 frames per second are as inputs of this scheme. The proposed scheme exploits both temporal and spatial features of the input sequences, and focuses on index and thumb fingertips localization and joint angles of the robot arm to mimic the user's arm movements in 3D space in an uncontrolled environment. The RoboTECH II ST240 is used as a real robot arm model. Depth and skeleton data are used to determine the angles of the robot joints. Three approaches to identify the tip of the thumb and index fingers are presented using existing data, each with its own limitations. In these approaches, concepts such as thresholding, edge detection, making convex hull, skin modeling and background subtraction are used. Finally, by comparing tracked trajectories of the user's wrist and robot end effector, the graphs show an error about 0.43 degree in average which is an appropriate performance in this research.

The key contribution of this work is hand pose estimation per every input frame and updating arm robot according to estimated pose. Thumb and index fingertips detection as part of feature vector resulted using presented approaches. User movements transmit to the corresponding Move instruction for robot. Necessary features for Move instruction are rotation values around joints in different directions and opening value of index and thumb fingers at each other.

Keywords: pose estimation, depth data, markerless motion capture, Kinect, 3d model

سامانه‌هایی حس‌گرهای مغناطیسی و الکترومکانیکی مثل دستکش‌ها و حس‌گرهای نوری به طور رایج برای اندازه‌گیری موقعیت‌های دست و مفصل‌بندی انگشت به کار گرفته می‌شوند؛ اگرچه این نوع سامانه‌ها اندازه‌گیری‌های دقیقی را فراهم می‌کنند و در عملکردهای زمان واقعی می‌توانند به کار گرفته شوند؛ اما چنین سامانه‌هایی برای استفاده گران و مایه زحمت هستند.

روش‌های مبتنی بر بینایی، جایگزین‌های ارزانی برای معادل‌های مبتنی بر حس‌گر هستند و برای تعاملات انسان و رایانه طبیعی‌تر عمل می‌کنند. این روش‌ها مشکلات خاص خود را دارند؛ از جمله اینکه تعداد ابعاد فضای پارامتری حرکت دست در حدود سی است و پیدا کردن یک حالت متناظر با یک تصویر ورودی به علت فضای جستجوی با ابعاد

۱- مقدمه

برآورد حالت^۱ به فرآیندی گفته می‌شود که طی آن مشخص می‌شود که اعضای بدن انسان نسبت به هم چه پیکربندی دارند [۱]. برآورد حالت دست موضوع پژوهشی مهمی است که کاربردهای متنوعی در حوزه تعامل انسان با رایانه مثل تشخیص ژست، ساخت انیمیشن و کنترل ربات دارد. هر چند به علت انعطاف بالای دست، ضبط حرکت دست وظیفه‌ای چالش برانگیز است؛ اما روش‌های مبتنی بر بینایی و مبتنی بر حس‌گر بسیاری برای حل این مشکلات ارائه شده‌اند. در سامانه‌های مبتنی بر حس‌گر، از سخت‌افزارهای خاصی برای ضبط حرکت استفاده می‌شود. در چنین

^۱ Pose estimation

دست و در ویژگی‌های رنگی از رنگ دست یا علائمی که بر روی دست گذاشته می‌شوند، برای این منظور استفاده می‌شود.

یکی از مهم‌ترین مزایای دوربین عمق نسبت به دوربین‌های معمولی، جداسازی بهتر دست است. در بسیاری از مقاله‌ها از آستانه‌گذاری جهت جداسازی دست استفاده شده است. در این روش فرض شده است که دست از سایر اعضای بدن جلوتر قرار می‌گیرد. این روش به صورت مستقیم در [3] تا [6] استفاده شده است.

برای استخراج ویژگی‌های لازم جهت برآورده حالت، بسته به مدل استفاده شده و نوع کاربردی که تحلیل ژست دست به دنبال دارد؛ ویژگی‌هایی از قبیل موقعیت نوک انگشتان، تعداد انگشتان، موقعیت کف دست^۴ و زاویه‌های بین مفاصل دست می‌توانند استخراج شوند.

طبق چهارچوب ارائه شده در [2] الگوی برآورده حالت دست مبتنی بر بینایی ماشین بدون نشانه‌گذاری، با استفاده از یک دنباله تصاویر عمق و رنگی ارائه شده است. اجزای اصلی چهارچوب ارائه شده، شامل مدل سه‌بعدی دست، مازول تجزیه دست، تشخیص دهنده نوک انگشت، تعیین کننده پارامترهای حرکت و حل کننده سینماتیک معکوس است. در این مقاله فضای پارامتری حرکت دست به وسیله بردار ویژگی $\phi l = \{\phi g, \phi l\}$ توصیف می‌شود؛ جایی که ϕg حرکت سراسری^۵ دست و ϕl حرکت محلی^۶ دست است. ϕg شامل انتقال و دوران سه‌بعدی دست و ϕl متاظر با حالت محلی دست با ۲۱ درجه آزادی است.

انتقال سراسری دست، به عنوان موقعیت مرکز کف دست ($Tg = (xg \ yg \ zg)$ در سامانه مختصات با قاعده دست راست که مبدأ مختصات در مرکز دوربین قرار دارد و دوران سراسری، به عنوان زوایای اویلر از دوران کف دست $\Theta g = (\Theta x \ \Theta y \ \Theta z)$ با قرارداد $x-z-y$ -تعريف می‌شود. برای یافتن مرکز کف دست، با فرض اینکه دست، نزدیک‌ترین شیء به دوربین است و با محدود کردن دوران دست با تعريف محدوده‌های خاصی برای هر یک از زوایای اویلر، ناحیه کف دست شناسایی می‌شود. با توجه به ریخت‌شناسی دست و اینکه کف دست موازی دوربین باشد ($0 \approx \Theta y \approx \Theta z \approx 0$) فرض می‌شود که بزرگ‌ترین حباب^۷ سراسری در ناحیه دست، کف دست است.

⁴ Palm

⁵ Global

⁶ Local

⁷ Blob

بالا به محاسبات زیادی نیاز دارد؛ به علاوه، با توجه به نوع مفصل‌بندی دست، تعداد هم‌پوشانی‌ها در تصاویر زیاد است. همچنین به علت وجود تعییرات روش‌شناسی و پس‌زمینه در هم‌ریخته، محیط برای تعاملات انسان و رایانه غیرقابل کنترل می‌باشد. این موارد سختی‌های بیشتری را در شناسایی دست و استخراج ویژگی تحمیل می‌کنند [2]. روش‌های برآورده حالت دست مبتنی بر بینایی به دو دسته روش‌های مبتنی بر ظاهر^۱ و روش‌های مبتنی بر مدل^۲ تقسیم می‌شوند. در روش‌های مبتنی بر ظاهر، تصویر ورودی یا ویژگی‌های استخراج شده از آن با پایگاه داده‌ای که شامل تصاویر دو بعدی و یا ویژگی‌های اشکال مختلف دست است، مقایسه می‌شود و نزدیک‌ترین حالت به تصویر ورودی به دست می‌آید. این روش از سرعت خوبی برخوردار، اما دقت آن به شدت وابسته به عمودبودن دست بر محور دوربین است. به طور معمول تعداد زیادی نمونه‌های آموزشی برای آموزش تابع نگاشت از ویژگی‌ها به حالت‌های دست استفاده می‌شوند. با داشتن تابع نگاشت آموزش داده شده، حالت دست به طور مؤثر تخمین زده می‌شود [2].

در روش‌های مبتنی بر مدل، از یک مدل اولیه استفاده می‌شود که به طور پیوسته به وسیله ویژگی‌های استخراج شده از روی تصویر ورودی به روز می‌شود؛ از این روش اطلاعات مطلوبی شامل حالت در هر زمان فراهم می‌شود. این روش‌ها به محاسبات زیادی نیازمند هستند و در عمل پیاده‌سازی آنها به صورت آنی امکان‌پذیر نیست [2].

برآورده حالت دست با استفاده از تصاویر به دست آمده از

دوربین‌ها (رنگی یا عمق) شامل سه مرحله زیر است:

۱. تشخیص محل دست و جداسازی آن

۲. استخراج ویژگی

۳. تنظیم پارامترهای مدل مورد استفاده با ویژگی‌های

استخراج شده و در نتیجه به روزرسانی مدل در مرحله تشخیص محل محل دست و جداسازی آن، ویژگی‌ها و علائمی برای یافتن ناحیه‌ای از تصویر که شامل دست است استخراج می‌شود. ویژگی‌های استفاده شده در تشخیص مکان دست در دوربین‌های رنگی را به دو دسته ویژگی‌های رنگی و شکلی می‌توان تقسیم‌بندی کرد. در ویژگی‌های شکلی از ریخت‌شناسی^۳ دست و اعمال محدودیت‌هایی روی درجه آزادی آن، جهت تشخیص مکان

¹ Appearance-based methods

² Model-based methods

³ Morphology



تجزیه دست به فرآیندی گفته می‌شود که طی آن به هر پیکسل از ناحیه دست جداسده، یک برچسب از مجموعه {کف دست، انگشت کوچک^۱، انگشت انگشتتری^۲، انگشت میانی^۳، انگشت اشاره^۴، انگشت شست^۵} نسبت داده شود. بعد از شناسایی کف دست، وظیفه برچسب‌گذاری پیکسل‌های باقیمانده به عنوان یک مسئله دسته‌بندی مطرح که به کمک دسته‌بند بیزین^۶ حل می‌شود. تشخیص دهنده نوک انگشت سه‌بعدی از ایده بیشینه کردن فاصله ژئودزیک^۷ الهام گرفته است. به طور ایده‌آل در هر قسمت برچسب‌گذاری شده، نقاطی که فاصله ژئودزیک تا مرکز کف دست را بیشینه می‌کنند، بایستی به عنوان نوک انگشت شناخته شوند [۷]-[۹].

در [۱۰] روش برآورد حالت مبتنی بر مدل با استفاده از تصاویر رنگی و عمق به دست آمده از کینکت ارائه شده است. در این مقاله از مدل دست با ۲۶ درجه آزادی که شامل مش سه‌بعدی با رزولوشن متوسط، ۱۳۱۷ رأس و یک اسکلت متصل شده به مدل است، استفاده شده است. حالت دست بر مبنای بیست زاویه مابین مفاصل استخوان‌ها تعریف می‌شود. شش پارامتر دیگر، موقعیت و دوران دست را مشخص می‌کنند. الگوریتم ارائه شده برای برآورد حالت شامل چهار گام (۱) مقداردهی اولیه (۲) جداسازی ابر نقطه^۸ با استفاده از رشد ناحیه (۳) اصلاح اندازه و موقعیت دست و (۴) برآورد حالت نهایی دست است.

در [۱۱] ایده تشخیص انگشتان در گام‌های زیر خلاصه می‌شود: ۱) گرفتن تصویر عمق از کینکت (۲) دنبال کردن دست‌ها و تشخیص نقاط دست (۳) ناحیه‌بندی به وسیله عمق و (۴) تعیین موقعیت نوک انگشتان. در ابتدا با اعمال یک مقدار آستانه مناسب، قسمت دست جدا می‌شود. برای پیدا کردن موقعیت انگشتان ابتدا ناحیه کف دست به به وسیله اعمال یک فیلتر دایره‌ای بزرگ بر روی تصویر عمق به دست می‌آید، به گونه‌ای که همه انگشتان در تصویر حذف خواهند شد. در این مرحله ناحیه کف دست از تصویر اصلی دست کم می‌شود تا ماسک برای جداسه کردن انگشتان به دست

¹ Pinky

² Ring

³ Middle

⁴ Index

⁵ Thumb

⁶ Naïve Bayesian classification

⁷ geodesic extrema

⁸ Point cloud

آید. ماسک‌های انگشت با تصویر عمق اصلی ضرب شده تا نگاشت عمق از انگشتان به دست آید.
در [۱۲] و [۱۳] برای پیدا کردن نوک انگشتان ابتدا کانتور^۹ و پوسته محدب^{۱۰} اطراف دست پیدا و سپس محل تلاقی این دو ویژگی محاسبه می‌شود. از آنجایی که نوک انگشتان به همراه نقاط دیگری پیدا شده‌اند، نقاط اضافی در دو مرحله حذف می‌شوند؛ ابتدا نقاط بر اساس فاصله، خوشبندی شده و از میان نقاط نزدیک به هم تنها یک نقطه باقی می‌ماند؛ پس از آن با استفاده از روش تطبیق سه نقطه^{۱۱} نقاط اضافی باقیمانده از مرحله قبل حذف می‌شوند. در این روش برای هر نقطه باقیمانده، یک نقطه قبل و یک نقطه بعد از آن در فاصله‌های مشخص بر روی کانتور در نظر گرفته می‌شود؛ سپس فاصله خط گذرنده از این دو نقطه تا نقطه انتخابی محاسبه و اگر این فاصله از مقدار مشخصی کمتر باشد آن نقطه حذف می‌شود. مبنای این روش این است که انگشتان به دلیل اینکه ماهیت استوانه‌ای دارند، فاصله نقطه نوک انگشت از نقاط اطراف - نسبت به سایر نقاط که در سطوح هموارتری قرار دارند- بیشتر خواهد بود. در [۱۴] با استفاده از داده‌های عمق به دست آمده از کینکت، حالت‌هایی که توسط پلیس کنترل ترافیک انجام می‌شود، تشخیص داده شده است. در این مقاله بردار ویژگی که حالت کنترل ترافیک را توصیف می‌کند، شامل زوایای نسبی بین مفاصل اسکلت است و از خوشبند ماشین بردار پشتیبان^{۱۲} (SVM) برای تشخیص حالت استفاده شده است.

در [۱۵] و [۱۶] سامانه‌ای مبتنی بر استفاده از کینکت برای تعامل انسان با ربات طراحی شده است. در این سامانه، کاربران نیاز دارند تا از حالت‌های بدن‌شان استفاده کنند تا بتوانند با ربات تعامل داشته باشند. در این سامانه، کینکت حالت‌های مختلف بدن را تشخیص داده و یک واسطه تعاملی بین انسان و ربات تولید می‌شود. لازم به ذکر است که در این مقاله‌ها از حالت‌های ایستا استفاده شده است. روند کلی در این مقاله‌ها را در چهار گام (۱) جمع‌آوری حالت‌های بدن در فضای سه‌بعدی به وسیله کینکت (۲) پردازش تصاویر به دست آمده از کینکت و تعریف مازول‌های حالت^{۱۳} (۳) انتقال دستورها و داده‌های متناسب با مازول حالت

⁹ Contour

¹⁰ Convex hull

¹¹ Three point alignment

¹² Support vector machine

¹³ Gesture module

حول مفصل شانه^۵ به سمت داخل و خارج، حرکت دورانی حول مفصل شانه به سمت بالا و پایین و باز و بسته‌شدن انگشتان شست و اشاره از هم است)، توسط بازوی ربات، در یک محیط کنترل‌نشده تقليد شود. اين مسئله را برآورد حالت مبتنی بر مدل گويند. برای اين کار در ابتدا حرکت بازوی انساني در قالب دنباله‌اي از تصاویر يا فرييم‌ها به عنوان ورودي تهيه و سپس با استفاده از رو يكده‌اي استخراج ويژگي، بهازاي هر فرييم، بردار ويژگي جهت توصيف فضاي پaramتری حرکت بازو استخراج می‌شود. به منظور استخراج ويژگي‌ها از ترکيب داده‌های عمق، رنگي و اسكلت به دست آمده از كينكت و برای نمايش، از داده‌های رنگي استفاده شده است. برخی از ويژگي‌ها به طور مستقيم از اطلاعات عمق و اسكلت بهازاي هر فرييم استخراج می‌شوند؛ در حالی که ديگر ويژگي‌ها از جمله نوك انگشتان شست و اشاره با بهكارگيری مفاهيم بيناني ماشين، تحت سه رو يكدرد مطرح شده در ادامه که سهم اصلی کار اين پژوهش نيز است، به دست می‌آيند. در پایان بهازاي هر فرييم، بازوی ربات مناسب با بردار ويژگي استخراج شده، از طريق تنظيم پaramترهاي مربوط به مفاصل ربات، در حالت منتظر با حالت دست کاريبر قرار می‌گيرد. برآورد حالت توسط بازوی ربات بهازاي فرييم‌هاي متوالی يك دنباله ورودي، باعث تقليد حرکت می‌شود؛ به اين معناست که بهازاي هر فرييم ورودي، يك فرييم از عملکرد ربات به عنوان خروجي در دست خواهد بود. با کنار هم قراردادن فرييم‌هاي خروجي در مجموع، حرکت ربات به دست می‌آيد.

نكته قبل توجه اين است که به علت کندبوتن عملکرد بازوی ربات موجود برای قرارگرفتن در وضعیت مناسب، تقليد همزمان حرکت ممکن نیست. ربات برای قرارگيری در وضعیت مناسب بهازاي هر فرييم ورودي، به مدت زمانی نياز دارد که اين مدت زمان خيلي بيشتر از حرکت دست است؛ بنابراین اگرچه تقليد همزمان حرکت از اهداف اين پژوهش نیست؛ ولی در صورت استفاده از يك بازوی ربات سريعت، الگوريتم ارائه شده قابلیت اجرا به صورت برخط را نيز خواهد داشت.

در ادامه ساختار مقاله به ترتیب زير خواهد بود.
بخش ۲ مفاهيم اوليه، بخش ۳ روش پيشنهادي برای استخراج بردار ويژگي، بخش ۴ نتایج آزمایش‌ها و بخش ۵ نتیجه‌گيری را نشان می‌دهد.

⁵ Shoulder⁶ Elbow

تشخيص‌داده شده و ۴) اجرای عملکرد مناسب توسط ربات می‌توان خلاصه کرد.

در [17] سامانه تشخيص حالت پويای دست به صورت برخط^۱ با استفاده از دوربین RGB-D ارائه شده است. قبل از دنبال‌کردن دست‌ها، با استفاده از تصاویر عمق و الگوريتم نگاشت فاصله چمفر^۲، دست‌ها تشخيص داده شده و موقعیت‌شان معلوم شده است. اين سامانه هم از اطلاعات رنگی و هم اطلاعات عمق در فرآيند دنبال‌کردن دست استفاده کرده است. برای اينکه مسیر حرکت دست مشخص شود، لازم است که نقطه شروع و پایان مسیر حرکت تشخيص داده شود که برای حل اين مسئله در حين عمليات دنبال‌کردن دست از دو حالت ثابت، يكى دست در حالت به طور کامل باز و ديگري دست در حالت مشت استفاده شده است. در پایان از مدل مخفی مارکوف با ساختار تسمه چپ به راست^۳ به منظور طبقه‌بندی هر دنباله از مسیرها استفاده شده است.

در [18] مدلی از تقليد حرکت انسانی شامل حرکت‌های دو بازو مورد ارزیابی قرار می‌گيرد. اين مدل، از شبکه عصبی برای يادگيری تقليد استفاده می‌کند. مدل اشاره‌شده در شبیه‌سازی به وسیله يك آواتار با ۳۷ درجه آزادی اعتبارسنجی می‌شود. حرکت‌های بازوی انسان به وسیله سامانه‌های دنبال‌کردن مبتنی بر نشانه‌گذاري^۴ در قالب ويدئو ضبط می‌شوند و به عنوان ورودی به مدل داده می‌شوند.

در [19] برای کنترل حرکت بازوی يك ربات انسان‌نما به نام NAO، سامانه‌اي طراحی شده است که حالت‌های عامل انسانی شامل حرکت دو بازو را در راستاهای مختلف (بدون حرکت انگشتان) مورد تقليد قرار می‌دهد. اطلاعات مربوط به زاویه بازوی عامل انسانی توسط کينكت به دست می‌آيد. از کينكت به عنوان مازول تعاملی جهت تبديل اطلاعات زاویه به فرمت مناسب برای ربات استفاده می‌شود. اين سامانه به صورت زمان واقعی کار می‌کند و برای حل مشکل هم‌پوشانی، ملزم به اعمال محدودیت‌هایي روی حرکت عامل انسانی است.

در اين مقاله هدف اين است تا با استفاده از روش ضبط حرکت بدون نشانه‌گذاري و مفاهيم بيناني ماشين، حرکت بازوی عامل انسانی (که شامل حرکت دورانی بازو

¹ Online² Chamfer distance matching algorithm³ Left-Right banded structure⁴ marker-based tracking systems

۲- مفاهیم اولیه

۱- تشخیص اسکلت بدن با استفاده از داده عمق

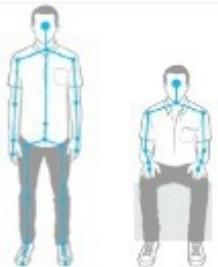
شاتون و همکارانش برای پیش‌بینی سریع و دقیق موقعیت مفاصل بدن، روشی مبتنی بر تک تصویر عمق و بدون دخالت هیچ اطلاعات زمانی را نداشت. در این روش از ایده شناسایی شیء^۱ و طراحی یک نمایش میانی از بخش‌های بدن^۲، برای نگاشت مسئله برآورده حالت‌های سخت، به مسئله ساده‌تر طبقه‌بندی بهازای هر پیکسل استفاده شده است [20].

استفاده از پایگاه داده آموزشی بسیار بزرگ و متنوع، به خوشبند اجراه می‌دهد تا نسبت به شکل بدن، نوع لباس پوشیدن و انواع حالت‌ها، غیرقابل تغییر باشد. از آنجایی که خوشبند از اطلاعات زمانی استفاده نمی‌کند، تنها از حالت‌های ایستا و نه حرکت^۳ برای آموزش استفاده می‌شود. تغییرات در حالت یک فریم از حرکت، با فریم بعدی حرکت بسیار کوچک و قابل صرف‌نظر کردن است. بنابراین بسیاری از حالت‌های مشابه اضافی در دنباله حرکت، با خوشبندی دورترین همسایگی^۴ دور ریخته می‌شوند.

برای ساخت نمایش میانی از بخش‌های بدن، سهم کلیدی کار این بوده که کل بدن توسط ۳۱ بخش مطابق شکل (۱) برچسب‌گذاری شده است: اگر بدن انسان به دو قسمت چپ و راست تقسیم شود، برای هر کدام از قسمت‌ها، پانزده برچسب با نام‌های، قسمت بالایی سر، قسمت پایینی سر، شانه، قسمت بالایی بازو، قسمت پایینی بازو، آرنج، مج، دست، قسمت بالایی کمر، قسمت پایینی کمر، ران پا، زانو، ساق پا، قوزک پا، مج پا وجود دارد. ۱۳۰مین برچسب متعلق به گردن است. بعضی از بخش‌ها به‌طور مستقیم موقعیت مفصل خاصی را مشخص می‌کنند؛ در حالی که دیگر بخش‌ها شکاف‌ها را پر می‌کنند یا در ترکیب با هم برای شناسایی مفاصل دیگری می‌توانند استفاده شوند. با تعریف این نمایش میانی، مسئله را با الگوریتم‌های خوشبندی می‌توان حل کرد. جفت تصاویر عمق و تصاویر برچسب‌گذاری شده متناظر با آن برای آموزش خوشبند استفاده می‌شوند. با استفاده از



(شکل-۱): ۳۱ برچسب برای حالت‌های مختلف [20]
(Figure-1): 31 labels for different poses [20]



(شکل-۲): مفاصل شناسایی شده در حالت ایستاده و ایستاده [20]
(Figure-2): recognized joints in standing and sedentary poses [20]

۲- درجه‌بندی عملیاتی ربات

در این پژوهش از بازوی ربات RoboTEK II ST240 استفاده شده است که با سه درجه آزادی دورانی از گروه ربات‌های آرنجی است. بازوی ربات مورد استفاده دارای واسط گرافیکی (شکل ۳) است. در واسط گرافیکی، مفصل‌های موردن استفاده و محدوده کاری هر مفصل نشان داده شده است. محدوده کاری هر مفصل با اعداد ۱ تا ۱۲۷ نشان داده می‌شود. میزان دوران بهازای هر یک واحد در این محدوده نه تنها برای همه مفاصل یکسان نیست، بلکه معادل ۱ تا ۱۲۷ درجه هم نیست. بنابراین با درنظرگرفتن یک حالت پایه برای هر مفصل، میزان دوران بهازای هر یک واحد برای

^۵ Randomized decision forests classifiers

^۶ Mass

^۱ Object recognition

^۲ Intermediate body parts representation

^۳ Motion

^۴ Furthest neighbor clustering



(شکل-۵): حالت پایه در نظر گرفته شده برای مفصل آرنج معادل صفر درجه در عمل

(Figure-5): basic state for elbow joint equivalent to zero degree in action

جدول (۱) درجه‌بندی واقعی مفاصل ربات را برای چهار زاویه پایه، شانه، آرنج و مجری نهایی، برای بعضی از اعداد در محدوده ۱ تا ۱۲۷ نشان می‌دهد.

(جدول-۱): درجه‌بندی عملیاتی ربات بر حسب درجه
(Table-1): robot calibration in degree

زاویه محدوده کاری	زاویه پایه	زاویه شانه	زاویه آرنج	زاویه مجری نهایی
1	-42.2433	-33.1163	-30.5783	18.6360
10	-42.2433	-32.8639	-30.5783	18.6792
30	-42.2433	-20.6063	-27.8330	20.5033
40	-35.7324	-14.0581	-24.0501	22.7577
50	-25.6143	-6.9820	-18.4935	26.3843
60	-15.6035	0	-12.3195	30.5143
70	-7.0276	7.0962	-5.5725	34.5187
78	0	13.1374	0	37.3032
80	2.0498	14.5336	1.4143	37.9189
90	10.4688	22.2658	8.6109	40.7529
100	18.7528	29.8554	16.1278	43.0342
110	27.5585	37.3623	23.5571	43.4195
120	35.9247	44.0439	30.7408	43.4985
127	41.2824	46.4810	35.9064	43.5322

آن مفصل به طور تجربی به دست می‌آید. درجه‌بندی عملیاتی ربات به صورت دستی انجام و بخشی از خطاهای موجود در قسمت نتیجه نیز از همین جا ناشی می‌شود. برای مثال، شکل (۴) محدوده کاری مفصل آرنج در راستای بالا-پایین را با تنظیم کردن متغیر Elbow در واسطه گرافیکی، یک بار به‌ازای عدد ۱ و بار دیگر به‌ازای عدد ۱۲۷ نشان می‌دهد. شکل (۵) حالت پایه در نظر گرفته شده را برای مفصل آرنج به منظور درجه‌بندی عملیاتی نشان می‌دهد. واسطه گرافیکی مورد استفاده قادر به دریافت دستورها به عنوان ورودی است؛ و حالت متناظر با آن دستور را بر روی ربات می‌زنیم. اعمال می‌کند. دستور ورودی به ربات نیازمند چهار مقدار ورودی به ترتیب، میزان دوران بازو حول مفصل شانه به سمت داخل و خارج Base در واسطه گرافیکی، میزان دوران حول مفصل شانه به سمت بالا و پایین Shoulder و میزان بازشدن قسمت مجری نهایی Elbow است. برای تمام مفاصل میزان دوران، نسبت به حالت پایه به کار گرفته شده برای آن مفصل (چون حالت پایه در عمل معادل صفر درجه است). در نظر گرفته می‌شود. در ادامه برای فهم بهتر و اشتباہ‌نگرفتن زاویه شانه و آرنج واسطه گرافیکی با مفاصل شانه و آرنج، به جای زاویه‌های Base، Elbow، Shoulder، Jaw و بازویه‌های زاویه شانه، زاویه آرنج و زاویه مجری نهایی استفاده می‌شود.



(شکل-۳): واسطه گرافیکی بازوی ربات

(Figure-3): graphical interface for robot arm



الف) وضعیت ربات با تنظیم عدد ۱ در محدوده کاری مفصل آرنج
ب) state of robot with setting 127 in elbow workspace

(شکل-۴): محدوده کاری مفصل آرنج در عمل

(Figure-4): the workspace of elbow joint in action

¹ End Effector

۳- روش پیشنهادی

۱-۳- جمع‌آوری پایگاه داده

در این پژوهش با استفاده از دوربین کینکت، حرکات بازوی عامل انسانی در قالب دنباله تصاویر رنگی، عمق و داده‌های اسکلت به طور همزمان با سرعت سی فریم بر ثانیه ضبط شدند. شکل (۶) نمونه‌ای از هر یک از داده‌های ورودی را نشان می‌دهد.



الف) تصویر رنگی
a) color image



ب) تصویر عمق
b) depth image



پ) اطلاعات اسکلت
c) skeleton data

(شکل-۶): نمونه‌ای از داده‌های ورودی
(Figure-6): sample of input data

۲-۳- استخراج بردار ویژگی از روی داده‌ها

بردار ویژگی جهت برآورد حالت، شامل مختصات دو بعدی (متغیر x و y) و در برخی موارد سه بعدی (متغیر x ، y و عمق) برای نقاط مشخص شده در شکل (۷) است. با استفاده از داده‌های اسکلت به دست آمده از دوربین کینکت، مختصات x و y مفصل شانه، مفصل آرنج، مفصل مج بر حسب پیکسل به ازای هر فریم به دست می‌آید. مقدار فاصله مفاصل مج و شانه از دوربین کینکت، با استفاده از تصویر عمق بر حسب میلی‌متر به دست می‌آید. موقعیت نوک انگشتان اشاره و شست به طور مستقیم از روی داده‌های موجود در دسترس نیست. بنابراین به دست آوردن مختصات نوک انگشتان به عنوان بخشی از بردار ویژگی و سهم اصلی کار این پژوهش، با به کار گیری رویکردهایی که در ادامه بیان می‌شوند، یکی از اهداف این پژوهش است.



(شکل-۷): نقاط تشکیل دهنده بردار ویژگی
(Figure-7): points in feature vector

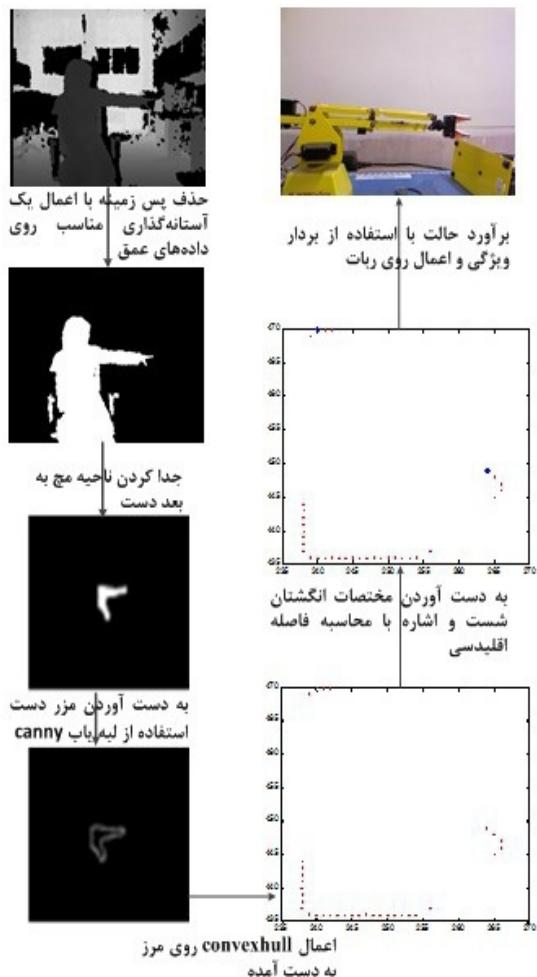
۳-۳- شناسایی نوک انگشتان اشاره و شست

به دست آوردن موقعیت نوک انگشتان اشاره و شست برای تعیین میزان بازشدن مجری نهایی ربات لازم و ضروری است. بعد از پیدا کردن نوک انگشتان با استیتی نقطه‌ای برای قرار گرفتن در مبدأ مختصات مشخص شود تا مجموع زوایایی که نوک انگشتان با محور مختصات می‌سازند، میزان بازشدن انگشتان از یکدیگر را نشان دهد (شکل ۸). لازم به ذکر است که در این مورد موقعیت کف دست حاصل از کینکت برای استفاده به عنوان مبدأ مختصات مناسب نیست. از این‌رو با پیدا کردن کمینه محلی در مرز دست بین انگشتان شست و اشاره، مبدأ مناسب به دست می‌آید.

همان‌طور که در قبل اشاره شد، مختصات سه نقطه از مجموعه نقاط شکل (۷) (مفصل شانه، مفصل آرنج، مفصل

فصل سی

انگشتان شست و اشاره، هماهنگ با هم و به یک اندازه در دو جهت مخالف حول نقطه مبدأ دوران داده می‌شوند، می‌توان بطرف کرد و به تعیین موقعیت نوک انگشت اشاره بسنده کرد. دو برابر زاویه به دست آمده برای انگشت اشاره (Θ_1 در شکل ۸)، میزان بازشدن انگشتان را از هم نشان می‌دهد.

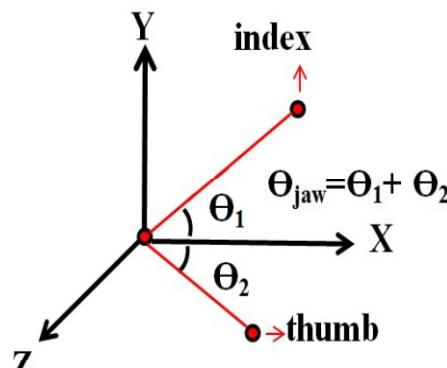


(شکل-۹): به دست آوردن موقعیت نوک انگشتان و به روزرسانی ربات
(Figure-9): fingertips identification and update the robot



(شکل-۱۰): از دست رفتن اطلاعات عمق برای انگشت شست
(Figure-10): losing some of depth information

مج) به طور مستقیم از روی داده‌های ورودی به دست می‌آیند. برای پیدا کردن مختصات سه نقطه باقیمانده (نوک انگشتان و نقطه بین دو انگشت)، در ادامه سه رویکرد ارائه می‌شود.



شکل-۸): محاسبه مقدار بازشدن انگشتان از هم
(Figure-8): calculating value of the opening fingertips at each other

۱-۳-۳-۳- استفاده از داده عمق و اسکلت برای شناسایی انگشتان

در این روش با الهام از مقاله‌های [3] تا [6]، برای به دست آوردن مختصات انگشتان با انتخاب یک مقدار آستانه مناسب روی تصویر عمق، قسمت پس‌زمینه حذف می‌شود (کاربر در فاصله بین یک تا دو متری کینکت قرار دارد). از بخش‌های باقیمانده با داشتن مختصات مج دست در داده‌های اسکلت، از قسمت مج دست به جلو نگه داشته شده و مابقی حذف می‌شوند. در ادامه نقاط حاصل از تلاقی کانتور و پوسته محدب اعمال شده بر روی ناحیه باقیمانده به دست می‌آیند. انگشت شست و اشاره در بین این نقاط هستند. برای همه نقاط میزان فاصله اقلیدسی تا نقطه کف دست که از داده‌های اسکلت است، محاسبه می‌شود. نقطه‌ای با بیشینه فاصله اقلیدسی، نقطه انگشت اشاره را نشان می‌دهد. انگشت شست نیز در ادامه با استفاده از میزان فاصله‌های اقلیدسی به دست آمده و اعمال این محدودیت که انگشت شست همیشه پایین‌تر از انگشت اشاره قرار می‌گیرد به دست می‌آید. اعمال این محدودیت به دلیل این است که قسمت مجری نهایی ربات مورد استفاده، قابلیت چرخش ندارد و همواره ثابت است. از این‌رو برای کاربر نیز فرض کردیم که در طول حرکت، انگشت اشاره همواره بالاتر از انگشت شست قرار دارد. شکل (۹) مراحل ذکر شده و به روزرسانی ربات توسط بردار ویژگی به دست آمده را نشان می‌دهد.

در این روش به دلیل از دست رفتن برخی اطلاعات عمق، تعیین موقعیت نوک انگشت شست در برخی از فریم‌ها امکان‌پذیر نیست (شکل ۱۰). این مشکل را با فرض اینکه

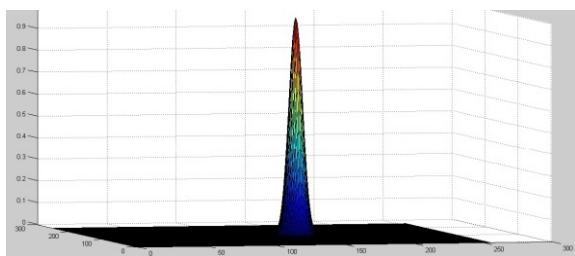
خوبی را برای جداسازی پیکسل پوست از غیر پوست به دست نمی آوری.



(شکل-۱۱): استفاده از فضای رنگی RGB برای جداسازی پیکسل پوست و غیر پوست

(Figure-11): separating skin and non-skin pixels using RGB color space

میزان روشنایی ممکن است در روی دست یک فرد متغیر باشد و معیار مناسبی در جاداکردن نواحی پوست از غیر پوست نیست. بعضی از فضاهای رنگی، افروزنگی موجود در کانال‌های رنگی RGB را کاهاش می‌دهند و رنگ را با اجزای مستقل آماری نمایش می‌دهند. در این فضای رنگی، شدت روشنایی و اجزای اصلی رنگ به صورت واضح از هم مجزا می‌شوند. این فضاهای انتخاب مطلوبی برای تشخیص پوست هستند [24] تا [28]. فضای رنگی YCbCr، نمونه‌ای از این فضای رنگی است. در فضای YCbCr، Y میزان روشنایی تصویر را مشخص می‌کند و Cb و Cr به ترتیب تفاوت میان میزان روشنایی با رنگ آبی (Y-blue) و رنگ قرمز (Y-red) را نشان می‌دهند. روشنایی از نمایش رنگی در این فضا می‌تواند حذف شود. Cb و Cr به عنوان رنگ‌های خالص در غیاب روشنایی شناخته می‌شوند؛ بنابراین در این رویکرد برای شناسایی انگشتان، با جمع‌آوری تعداد زیادی پیکسل پوست از افراد مختلف در شرایط نورپردازی متفاوت، رنگ پوست با استفاده از یکتابع گوسی در فضای رنگی YCbCr می‌تواند مدار شود (شکل ۱۲).



(شکل-۱۲): مدل کردن رنگ پوست در فضای YCbCr با استفاده از تک تابع گوسی

(Figure-12): fitting skin color into a Gaussian distribution in YCbCr color space

۲-۳-۳- مدل کردن رنگ پوست برای شناسایی انجشتن

در این روش برای شناسایی انگشتان از اطلاعات رنگ پوست استفاده می‌شود. گام‌های مقدماتی برای شناسایی پوست در یک تصویر بهوسیله اطلاعات رنگی به قرار زیر است [21]:
 ۱) نمایش پیکسل‌های پوست در فضای رنگی مناسب
 ۲) مدل کردن پیکسل‌های پوست و غیر پوست بهوسیله توزیع آماری مناسب^۳ طبقه‌بندی کردن توزیع‌های مدل شده.

انتخاب فضای رنگی به عنوان گام ابتدایی در طبقه‌بندی رنگ پوست در نظر گرفته می‌شود. فضای رنگی همچنین تعیین می‌کند که چقدر به طور مؤثر توزیع رنگ پوست را می‌توان مدل کرد. فضای رنگی RGB، فضای رنگی پیش فرض برای اغلب فرمتهای تصاویر موجود است. به طور معمول فضاهای رنگی دیگر از تبدیل خطی یا غیر خطی از فضای RGB به دست می‌آیند. تبدیل فضای رنگی، برای کاهش همپوشانی بین پیکسل‌های پوست و غیر پوست فرض شده است و از این‌رو به فرآیند دسته‌بندی پیکسل‌های پوست کمک می‌کند و دقت بالایی را در شرایط نورانی متغیر فراهم می‌کند. به طور معمول اجزای اصلی رنگ پوست برای افراد مختلف با توجه به میزان شدت روشنایی با هم متفاوت هستند [22]؛ از این‌رو یک عملکرد رایج، حذف شدت روشنایی، برای وظیفه دسته‌بندی پوست است.

رنگ پوست را در فضاهای رنگی گوناگونی می‌توان مدل کرد. اجزای سه‌گانه (r,g,b) در فضای رنگی RGB، تنها رنگ را نشان نمی‌دهند؛ بلکه همچنین روشنایی را نشان می‌دهند. بنابراین نمایش رایج RGB در تصاویر رنگی برای مشخص کردن رنگ پوست مناسب نیست. شکل (۱۱) تصویر رنگی RGB و تصویر دودویی متناظر آن، که با اعمال مقادیر آستانه اشاره شده برای هر جز سه‌گانه در فرمول (۱) [۲۳]، پیکسل‌های پوست را از غیر پوست جدا می‌کند، نشان می‌دهد.

$$R > 95 \text{ and } G > 40 \text{ and } B > 20 \text{ and } \max\{R, G, B\} - \min\{R, G, B\} > 15 \text{ and } |R - G| > 15 \text{ and } R > G \text{ and } R > B \quad (1)$$

هر پیکسلی که در رابطه (1) صدق کند، به عنوان پوست در نظر گرفته می‌شود و در غیر این صورت آن پیکسل متعلق به طبقه پوست نیست؛ اما به دلیل این که در یک محیط کنترل نشده، در پس زمینه، رنگ‌های مشابه رنگ پوست وجود دارند و تصاویر رنگی به دست آمده از کینکت، کیفیت بالایی ندارند؛ استفاده از فضای رنگی RGB، نتیجه

از آنجایی که پس زمینه شامل رنگ هایی، مشابه رنگ پوست و جمع آوری پیکسل پوست مورد نیاز برای مدل کردن رنگ پوست با مشکلاتی همراه است؛ تعیین مقدار آستانه مناسب برای جدا کردن پیکسل پوست از غیر پوست دشوار است. برای بهبود فرآیند شناسایی پیکسل های پوست، با استفاده از داده های عمق کینکت، عمق متناظر نقاط را در قسمت (پ) از شکل (۱۳) می توان به دست آورد و با آستانه گذاری مناسب، پیکسل های غیر پوست را حذف کرد. همان طور که در قسمت الف از شکل (۱۳) دیده می شود دست های کاربر نسبت به دیوار پشت سر، به کینکت نزدیک تر هستند. شکل (۱۴) مؤثربودن به کارگیری اطلاعات عمق برای دور ریختن پیکسل های غیر پوست را نشان می دهد. نوک انگشتان با استفاده از تلاقي کانتور و پوسته محدب اعمال شده بر روی مرز دست، همانند روی کرد قبلی حاصل می شوند.

برای برخی فریم‌ها که قسمت مج دست اشتباه تشخیص داده شده است (پس زمینه هم به عنوان بخشی از مج دست تشخیص داده شده است). و اطلاعات عمق نیز در آن نواحی در دست نیست که جهت جداسازی پس زمینه مورد استفاده قرار گیرد؛ به کارگیری این روش، مختصات نوک انگشتان را به درستی تشخیص نمی‌دهد.



(شکل-۱۴): استفاده از داده عمق برای حذف پیکسل های غیر پوست

(Figure-14): using of depth data for removing non-skin pixels

ساده‌ترین روش برای به دست آوردن قسمت هدف، تفریق کردن تصویر، از یک تصویر مرجع است. در این رویکرد فرض شده است که دوربین ثابت است و تنها جسم متحرک در صحنه، هدف مورد نظر است. در این روش از ایده مقاله [29] مطابق شکل (۱۵) استفاده شده و لازم است تا یک تصویر مرجع از پس‌زمینه بدون حضور هدف موجود باشد.

با داشتن این مدل گوسی، برای هر تصویر رنگی، کلیه پیکسل‌ها از فضای رنگی RGB به فضای YCbCr انتقال داده می‌شوند و به‌ازای هر پیکسل، یک جفت مقدار (cb,cr) به‌دست می‌آید. مقدار درستنمایی^۱ پوست برای هر پیکسل با استفاده از معادله‌ای مناسب به‌دست می‌آید. این مدل رنگ پوست، تصویر رنگی را به تصویر سطوح خاکستری می‌تواند منتقل کند؛ به‌طوری‌که مقادیر خاکستری در هر پیکسل مقدار احتمال متعلق‌بودن آن پیکسل به طبقهٔ پوست را نشان می‌دهد. با آستانه‌گذاری مناسب، تصاویر سطوح خاکستری به تصاویر دودویی که نواحی پوست و غیر پوست را نشان می‌دهند، می‌توانند منتقل شوند. قسمت‌های الف، ب و پ در شکل (۱۳) به ترتیب تصویر رنگی، تصویر سطوح خاکستری نمایش‌دهندهٔ مقدار احتمال‌ها و تصویر دودویی جاکنندهٔ رنگ پوست و غیر پوست را نشان می‌دهند.



الف) تصویر رنگی
a) color image



b) تصویر سطح خاکستری نشان دهنده میزان احتمال‌ها
b) gray scale image showing the likelihood of pixels



c) binary image separating skin and non-skin color

(شکل ۱۳): استفاده از مدل کردن رنگ پوست برای شناسایی پوست

(Figure-13): using of skin color modeling for skin detection

1 Likelihood

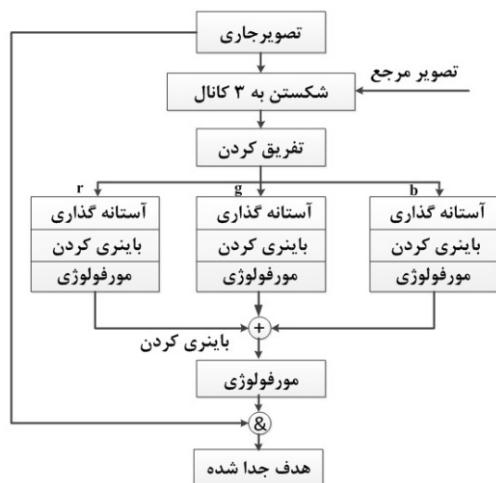
۳-۴- تبدیل ویژگی‌های استخراج شده به زوایای مناسب ربات

همان‌طور که در قبل بیان شد، به‌ازای هر فریم، یک بردار ویژگی استخراج می‌شود. برخی از ویژگی‌ها، شامل مختصات x و y نقطه خاصی از روی تصویر هستند و درواقع موقعیت پیکسل خاصی را نشان می‌دهند و اطلاعات پیکسلی هستند؛ درحالی که برخی از ویژگی‌ها میزان عمق یا فاصله از دوربین آن نقاط خاص بر حسب میلی‌متر را شامل می‌شوند. این اطلاعات از یک نوع نیستند و برای به‌دست‌آوردن زاویه مفاصل ربات، اطلاعات باید از یک نوع باشند. این مسئله با دردست‌داشتن بعضی از اطلاعات دوربین کینکت مثل فاصله کانونی و مقیاس عمق با استفاده از توابع کینکت رفع می‌شود.

بازوی ربات موجود برای قرارگرفتن در یک وضعیت خاص، نیازمند تعیین چهار زاویه است. با تغییر زاویه پایه در واسط گرافیکی، بازوی ربات حول محور خاصی که ثابت است، بسته به عدد تنظیم شده در واسط گرافیکی، به سمت داخل یا خارج حرکت داده می‌شود. درواقع با تغییر زاویه پایه، حرکت بازو، در صفحه مختصات $x-z$ و حول محور ثابت خواهد بود.

برای تعیین زاویه پایه از روی فریم ورودی، مفصل شانه در نقش محور ثابت اشاره شده عمل می‌کند و در مبدأ مختصات قرار می‌گیرد. مطابق شکل (۱۷) زاویه‌ای که مفصل مج با محور مختصات در صفحه $x-z$ می‌سازد و با فرمول (۲) به‌دست می‌آید، میزان دوران به سمت داخل یا خارج را نشان می‌دهد. مختصات z برای دو مفصل مج و شانه، از روی داده‌های عمق به‌دست‌آمده از کینکت به دست می‌آیند. لازم به ذکر است از آنجایی که بازوی ربات موجود قابلیت دوران نسبت به مفصل آرنج به سمت داخل و خارج را ندارد، استفاده از مفصل مج و آرنج در این مورد یکسان است.

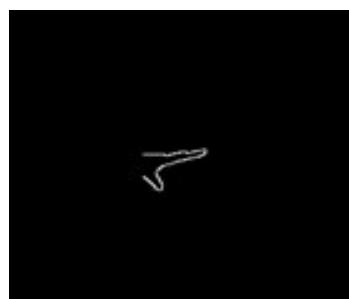
با تغییر زاویه شانه در واسط گرافیکی، بازوی ربات حول محور ثابت دیگری که عمود بر محور مربوط به زاویه پایه است، در راستای بالا-پایین دوران داده می‌شود. درواقع حرکت در صفحه مختصات $x-y$ و حول محور ثابت است. برای تعیین زاویه شانه از روی فریم ورودی، مفصل شانه در نقش محور ثابت اشاره شده عمل می‌کند و در مبدأ مختصات قرار می‌گیرد. زاویه‌ای که مفصل آرنج با محور مختصات در صفحه $x-y$ با فرمول (۳) می‌سازد؛ میزان دوران به سمت بالا یا پایین را نشان می‌دهد.



(شکل-۱۵): جداسازی هدف از پس‌زمینه [29]
(Figure-15): separating of target from background

اختلاف ۳ کanal (r g b) هر تصویر با کانال‌های متناظر در تصویر مرجع محاسبه می‌شود. نقاطی که میزان اختلاف بیشتری دارند، به عنوان هدف، در نظر گرفته می‌شوند. در این مرحله با انتخاب آستانه مناسب برای هر کانال، هدف را به درستی می‌توان تشخیص داد. در این روش ابتدا با استفاده از اطلاعات عمق، هدف را جدا کرده و با استفاده از توابع کینکت ناحیه متناظر هدف جداسده در تصویر رنگی به‌دست می‌آید. با توجه به ناحیه هدف در تصویر رنگی برای هر کدام از کانال‌ها حد بالا و پایین برای آستانه‌گذاری جهت جداسازی هدف در هر کانال مشخص می‌شود.

بعد از جدا کردن هدف از پس‌زمینه، با داشتن مختصات مج دست در داده‌های اسکلت، ناحیه مج دست به جلو مطابق شکل (۱۶) جدا می‌شود. در ادامه با اعمال لبه‌یاب و پیدا کردن کمینه و بیشینه محلی در مرز به‌دست‌آمده به ترتیب مبدأ مختصات موردنظر و نوک انگشتان به‌دست می‌آیند.



(شکل-۱۶): ناحیه مج دست جداسده بعد از اعمال لبه‌یاب
(Figure-16): the wrist after applying edge detector

۴- نتایج آزمایش‌ها

به منظور مقایسه سه رویکرد مطرح شده برای شناسایی موقعیت نوک انگشتان شست و اشاره که از مقالات مختلف الهام گرفته شده‌اند، برای سی فریم تصادفی که لزوماً فریم‌های متواالی یک دنباله نیستند، جدول (۲) به دست آمده است.

(جدول-۲): نتیجه مقایسه رویکردهای شناسایی نوک

انگشتان برای سی فریم

(Table-2): comparing different fingertips detection approaches for 30 frames

روش سوم پیشنهادی	روش دوم [۲۲]	روش اول [۳]	تعداد فریم‌های در تشخیص داده شده
۳۰	۱۸	۲۱	تعداد فریم‌های در تشخیص داده شده

ذکر این نکته نیز ضروری است که هدف ما از تقلید یک حرکت، تقلید تمام فریم‌های متواالی بین حرکت است؛ یعنی حرکت از یک حالت اولیه شروع شده و به تدریج به سمت حالت نهایی پیش می‌رود. حال اگر در حین حرکت تدریجی، برای یک فریم، یک حالت بهشدت متفاوت با فریم‌های قبلی اتفاق افتد، متواالی حرکت از بین می‌رود. به همین دلیل در این پژوهش با توجه به جدول (۲) روش پیشنهادی برای شناسایی نوک انگشتان، عملکرد بهتری دارد. که دقت خروجی نهایی را می‌تواند بهبود دهد.

در شکل‌های (۱۸) تا (۲۰) حرکت بازوی عامل انسانی توسط بازوی ربات برای سه دنباله از فریم‌های متواالی مورد تقلید قرار گرفته است.

دنباله شکل (۱۸) حرکت دورانی را نسبت به مفصل آرنج، ابتدا به سمت بالا و سپس به سمت پایین نشان می‌دهد. رسم خطوط قرمزرنگ برای فهم بهتر خواننده است تا تغییرات حرکت در طول فریم‌ها را بهتر درک کند.

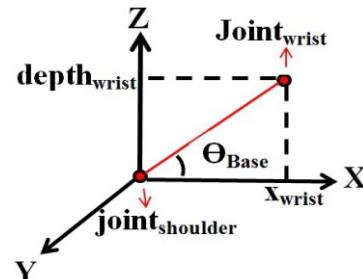
شکل (۱۹) حرکت دورانی بازو به سمت داخل و سپس به سمت خارج، با مبدأ قراردادن خطوط قرمزرنگ را نشان می‌دهد. شکل (۲۰) حرکتی را نشان می‌دهد که باز و بسته شدن انگشتان شست و اشاره از هم با استفاده از قسمت مجری نهایی ربات مورد تقلید قرار گرفته است.

در این موارد فرض شده که مفصل شانه در طول حرکت بازوی انسان در ویدئوی ورودی ثابت است و جایه‌جایی محسوسی ندارد. از این‌رو استخراج ویژگی‌ها از روی فریم ورودی با استفاده از فرمول‌های ارائه شده ممکن است.

برای تعیین زاویه آرنج از روی فریم ورودی، به اطلاعات دو مفصل آرنج و مج نیاز است. با قراردادن مفصل آرنج در مبدأ مختصات، زاویه‌ای که مفصل مج با محور مختصات در صفحه $y-x$ می‌سازد، به دست می‌آید. زاویه به دست آمده مجموع زاویه آرنج و شانه است. از این‌رو در فرمول (۴) زاویه به دست آمده از Θ_{shoulder} کم شده است تا زاویه آرنج به دست آید.

زاویه مجری نهایی نیز مطابق شکل (۸) بعد از به دست آمدن نوک انگشتان و نقطه میانی دو انگشت، با استفاده از فرمول (۵) به دست می‌آید.

مقدار زاویه‌های به دست آمده بر حسب درجه هستند؛ بنابراین بعد از به دست آوردن هر چهار زاویه، با داشتن فضای کاری هر مفصل مطابق جدول (۱)، نزدیک‌ترین عدد متناظر در محدوده ۱ تا ۱۲۷ را برای آن زاویه در واسطه گرافیکی می‌توان انتخاب کرد.



(شکل-۱۷): به دست آوردن Θ_{Base} در محور مختصات X-Z
(Figure-17): calculating Base angle in x-z coordinate plane

$$\Theta_{\text{Base}} = \text{Arctan}\left(\frac{\text{depth}_{\text{shoulder}} - \text{depth}_{\text{wrist}}}{x_{\text{wrist}} - x_{\text{shoulder}}}\right) \quad (2)$$

$$\Theta_{\text{Shoulder}} = \text{Arctan}\left(\frac{y_{\text{elbow}} - y_{\text{shoulder}}}{x_{\text{elbow}} - x_{\text{shoulder}}}\right) \quad (3)$$

$$\Theta_{\text{Elbow}} = \text{Arctan}\left(\frac{y_{\text{wrist}} - y_{\text{elbow}}}{x_{\text{wrist}} - x_{\text{elbow}}}\right) - \Theta_{\text{Shoulder}} \quad (4)$$

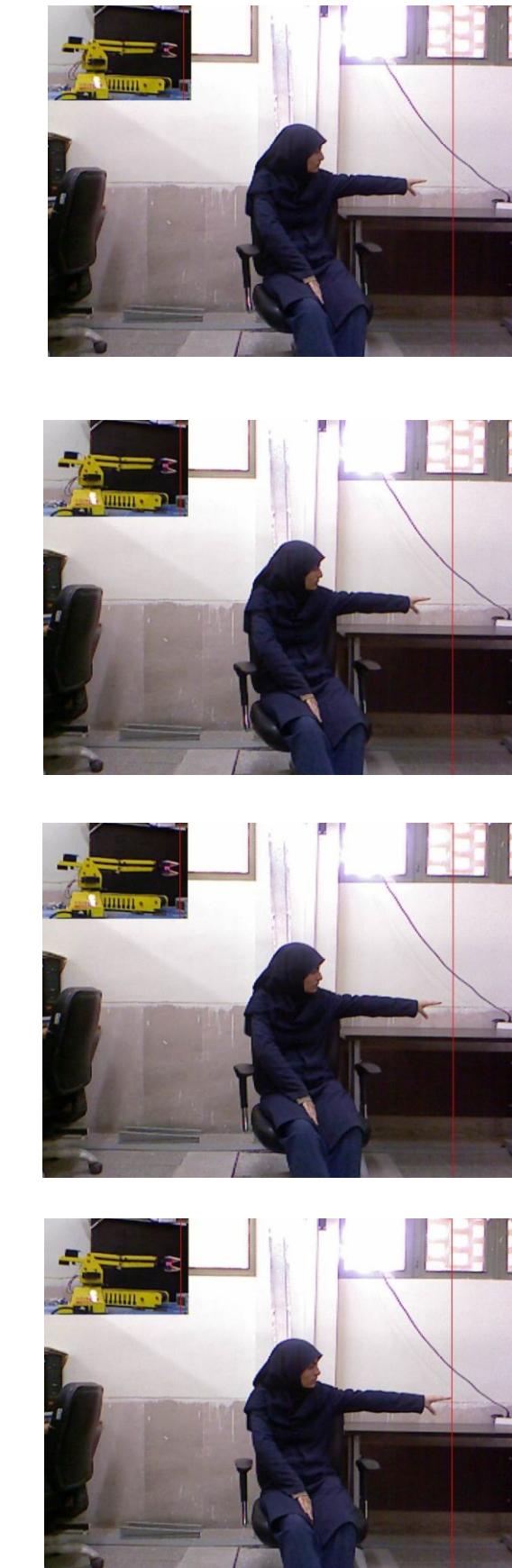
$$\Theta_{\text{Jaw}} = \text{Arctan}\left(\frac{y_{\text{index}} - y_{\text{palm}}}{x_{\text{index}} - x_{\text{palm}}}\right) + \text{Arctan}\left(\frac{y_{\text{thumb}} - y_{\text{palm}}}{x_{\text{thumb}} - x_{\text{palm}}}\right) \quad (5)$$

با توجه به یکسان‌بودن کمینه و بیشینه بازشدن مجری نهایی ربات و انگشتان شست و اشاره دست از هم، برای تناظریابی بین حالت مجری نهایی ربات و قسمت مج دست، زاویه‌های به دست آمده برای هر دوی آنها به محدوده صفر تا یک انتقال داده می‌شوند. با این کار عمل نرم‌السازی انجام می‌شود و حالت‌های نزدیک به هم در فضای نرمال‌بازه شده مشخص می‌شوند.



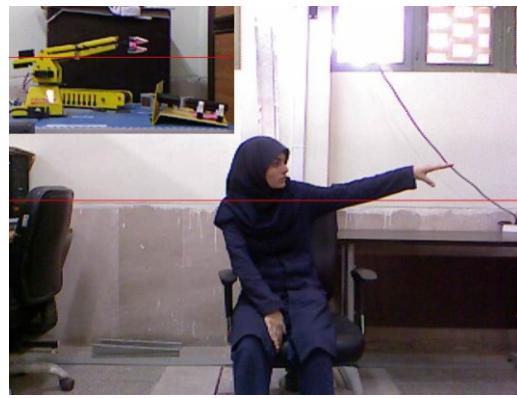
فصلنامی

پردازش داده
ش. علامه حاج
۱۳۹۶ سال
۴ شماره پیاپی
۳۴



(شکل-۱۸): برآوردهای پویای دست مربوط به حرکت دورانی نسبت به آرنج

(Figure-18): imitation of elbow movements for some frames



(شکل-۱۸): برآوردهای پویای دست مربوط به حرکت دورانی نسبت به آرنج

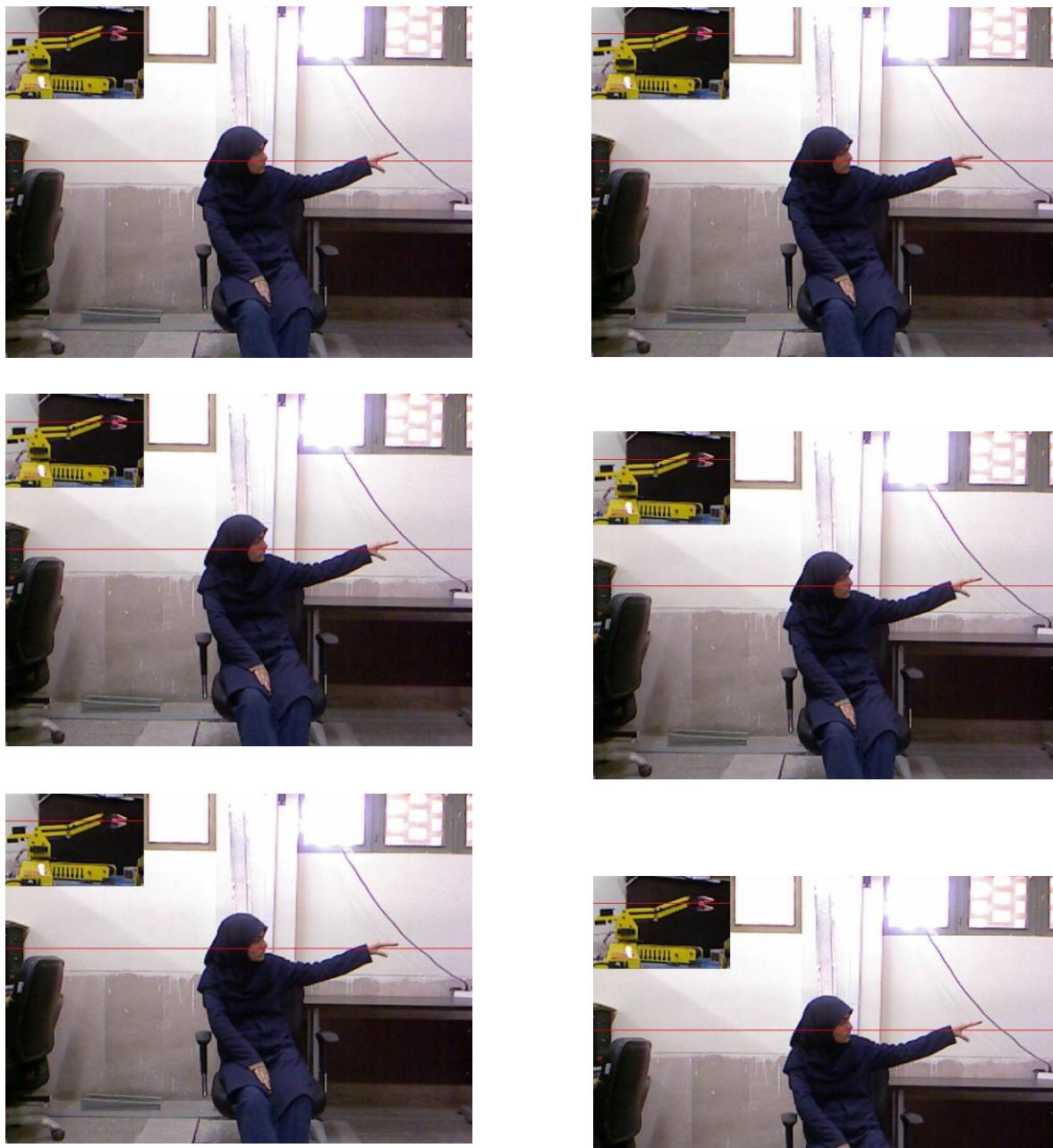
حرکت دورانی نسبت به آرنج

(Figure-18): imitation of elbow movements for some frames



(شکل-۱۹): برآوردهای پویای دست مربوط به حرکت دورانی به سمت داخل و خارج

(Figure-19): imitation of inside-outside hand movements

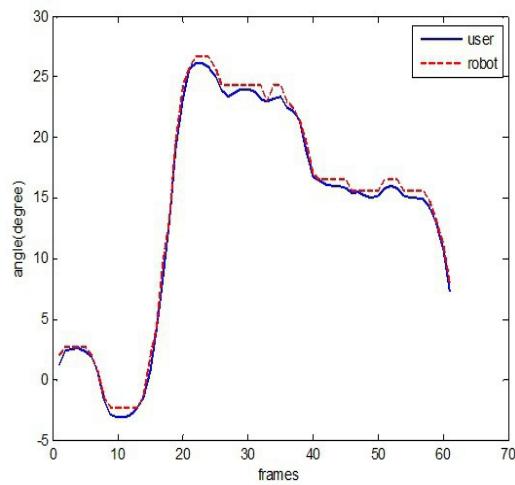


(شکل-۲۰): برآوردهای پویای دست مربوط به باز و بسته شدن انگشتان
 (Figure-20): imitation of opening-closing fingertips movements

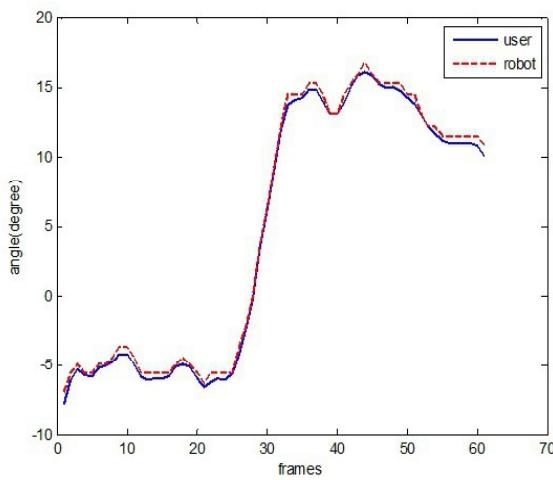
حرکت‌های انجام شده توسط کاربر به صورت دستورهای متناظر به ربات داده شدند. ورودی‌های لازم برای این دستور، میزان دوران حول مفاصل در راستاهای مختلف و میزان بازشدن انگشتان شست و اشاره از هم هستند. در ادامه نتایج به دست آمده از تقلید حرکت بازوی کاربر، توسط بازوی ربات آورده شده است.

برای ارزیابی عملکرد تقلید ربات از حرکت بازوی انسان یک ویدئوی ورودی از حرکت بازوی انسان و ویدئوی خروجی متناظر از عملکرد ربات، مورد نیاز است. موارد

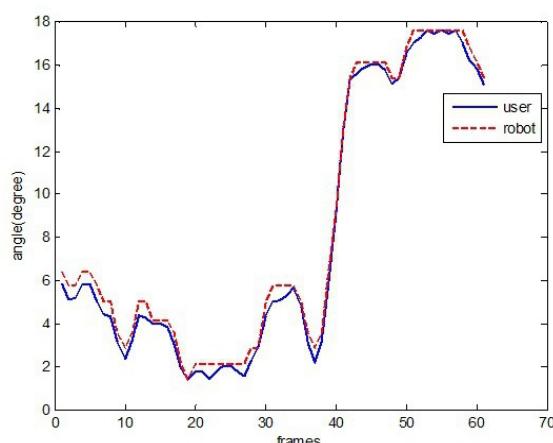




(شکل-۲۲): مقایسه تغییرات زاویه پایه برای کاربر و ربات
(Figure-22): changes of base angle for user and robot



(شکل-۲۳): مقایسه تغییرات زاویه شانه برای کاربر و ربات
(Figure-23): changes of shoulder angle for user and robot

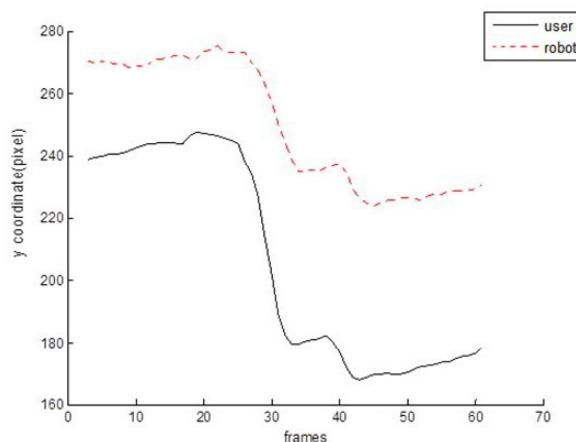


(شکل-۲۴): مقایسه تغییرات زاویه آرنج برای کاربر و ربات
(Figure-24): changes of elbow angle for user and robot

اشارة شده در ادامه برای هر دو ویدئو محاسبه و مقایسه می‌شوند.

مسیر طی شده (مختصات z بر حسب پیکسل در راستای بالا به پایین که از تغییرات زوایای شانه و آرنج حاصل می‌شوند). توسط قسمت نهایی دست کاربر (رنگ سیاه) با مسیر متناظر طی شده توسط مجری نهایی ربات (رنگ قرمز) که با استفاده از الگوریتم دنبال کردن^۱ کاناده-لوکاس-توomasی^۲ (KLT) به دست آمداند، در شکل (۲۱) مقایسه شده است. ویدئوی مورد استفاده جهت برآورد حالت شامل ۶۳ فریم است. به علت متفاوت بودن طول بازوی ربات و عامل انسانی، دو نمودار موجود در شکل (۲۱) روی هم قرار نگرفته‌اند. شکل‌های ۲۲، ۲۳ و ۲۴ تغییرات زاویه‌های پایه، شانه و آرنج بر حسب درجه را برای کاربر در طول ویدئوی ورودی و ربات در طول ویدئوی خروجی نشان می‌دهند.

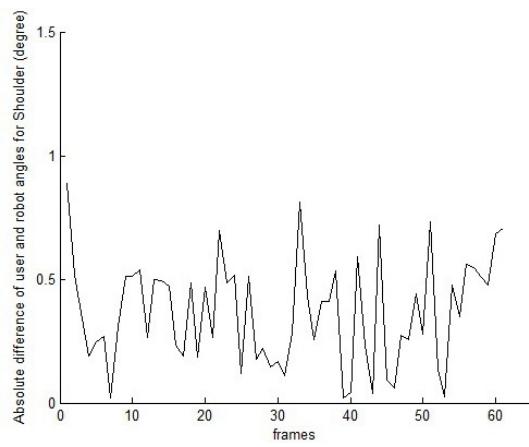
شکل (۲۵) زاویه بین مجری نهایی ربات را با زاویه بین انگشتان دست در فضای نرمالیزه شده در طول فریم‌ها مقایسه می‌کند. در جدول (۳) میانگین و واریانس اختلاف زاویه‌های برآورده شده برای زاویه‌های پایه، شانه، آرنج و مجری نهایی بازوی ربات با زوایای مفاصل بازوی کاربر نشان داده شده است. همان‌طور که در قبل اشاره شد، بازوی ربات مورد استفاده قادر درجه‌بندی دقیق و درجه‌بندی به صورت دستی انجام گرفته است. بخشی از این اختلاف‌ها از همین مسئله ناشی شده است. اعداد به دست آمده برای زاویه مجری نهایی در فضای نرمالیزه شده هستند.



(شکل-۲۱): مسیر طی شده توسط مجری نهایی و قسمت نهایی دست کاربر در راستای بالا به پایین و بر عکس
(Figure-21): tracked trajectory by user's wrist and robot end effector in up-down and reverse

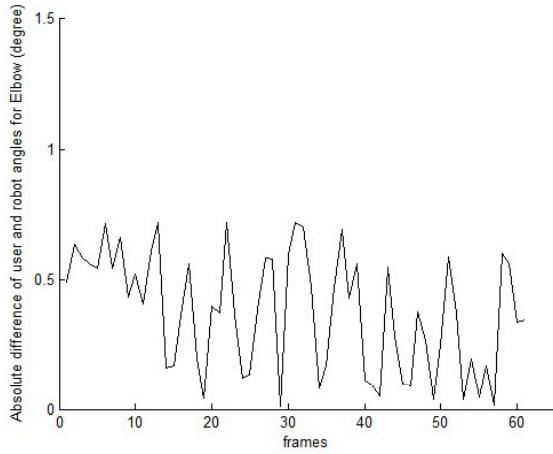
¹ Tracking algorithm

² Kanade-Lucas-Tomasi



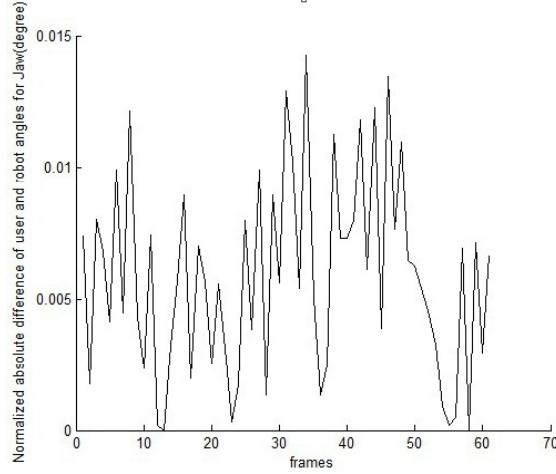
(شکل-۲۷): اختلاف زاویه شانه برای کاربر و ربات در طول فریمها

(Figure-27): difference of shoulder angle for robot and user during frames



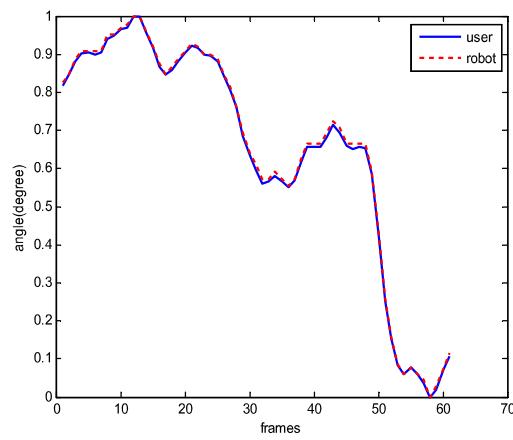
(شکل-۲۸): اختلاف زاویه آرنج برای کاربر و ربات در طول فریمها

(Figure-28): difference of elbow angle for robot and user during frames



(شکل-۲۹): اختلاف زاویه نرمالیزه شده مجری نهایی برای کاربر و ربات در طول فریمها

(Figure-29): difference of normalized end effector angle for robot and user during frames



(شکل-۲۵): مقایسه تغییرات زاویه مجری نهایی

برای کاربر و ربات

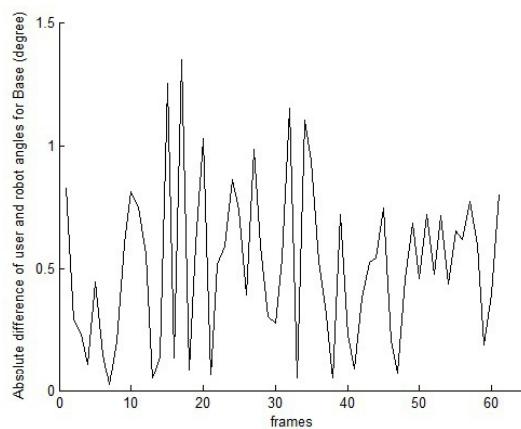
(Figure-25): changes of end effector angle for user and robot

نتایج عددی در شکل های (۲۶) تا (۲۹) نشان می دهند که سامانه پیشنهادی، نتیجه قابل قبولی را برای تقلید حرکت در بازه تغییرات نشان داده شده برای پارامترها در جدول (۱) ارائه می دهد.

(جدول-۳): میانگین و واریانس اختلاف زوایای مفاصل بازوی ربات و کاربر

(Table-3): average and variance difference user and robot angles

	زاویه پایه	زاویه شانه	زاویه آرنج	زاویه مجری نهایی
میانگین	0.5616	0.3601	0.3812	0.0058
واریانس	0.1304	0.0468	0.0470	1.4062e-05



(شکل-۲۶): اختلاف زاویه پایه برای کاربر و ربات در طول فریمها

(Figure-26): difference of base angle for robot and user during frames

6- References**۶- مراجع**

- [1] T. B. Moeslund and E. Granum, “A survey of computer vision-based human motion capture”, Computer vision and image understanding, vol. 81, no. 3, pp. 231–268, 2001.
- [2] H. Liang, J. Yuan, D. Thalmann, and Z. Zhang, “Model-based hand pose estimation via spatial-temporal hand parsing and 3D fingertip localization”, vol. 29, no. 6–8, pp. 837–848, 2013.
- [3] K. Fujimura and X. Liu, “Sign recognition using depth image streams”, 7th International Conference on Automatic Face and Gesture Recognition, pp. 381–386, 2006.
- [4] J. MacCormick and A. Blake, “A probabilistic exclusion principle for tracking multiple objects”, International Journal of Computer Vision, vol. 39, no. 1, pp. 57–71, 2000.
- [5] M. Moghaddam, M. Nahvi, and R. H. Pak, “Static Persian Sign Language Recognition Using Kernel-Based Feature Extraction”, 7th Iranian Machine Vision and Image Processing (MVIP), pp. 1–5, 2011.
- [6] P. Breuer, C. Eckes, and S. Müller, “Hand gesture recognition with a novel IR time-of-flight range camera—a pilot study”, in Computer Vision/Computer Graphics Collaboration Techniques, Springer, pp. 247–260, 2007.
- [7] C. Plagemann, V. Ganapathi, D. Koller, and S. Thrun, “Real-time identification and localization of body parts from depth images”, IEEE International Conference on Robotics and Automation (ICRA), pp. 3108–3113, 2010.
- [8] A. Baak, M. Müller, G. Bharaj, H.-P. Seidel, and C. Theobalt, “A data-driven approach for real-time full body pose reconstruction from a depth camera”, Consumer Depth Cameras for Computer Vision. Springer London, pp. 71–98, 2013.
- [9] L. A. Schwarz, A. Mkhitaryan, D. Mateus, and N. Navab, “Estimating human 3d pose from time-of-flight images based on geodesic distances and optical flow”, IEEE International Conference on Automatic Face & Gesture Recognition and Workshops, pp. 700–706, 2011.
- [10] A. Kuznetsova and B. Rosenhahn, “Hand pose estimation from a single rgb-d image”, Advances in Visual Computing. Springer Berlin Heidelberg, pp. 592–602, 2013.
- [11] J. L. Raheja, A. Chaudhary, and K. Singal, “Tracking of fingertips and centers of palm using kinect”, 3th international conference on

۵- نتیجه‌گیری

در این پژوهش رویکردهایی برای استخراج ویژگی‌های لازم جهت برآورد حالت‌های پویای بازوی یک عامل انسانی در یک محیط کنترل‌نشده ارائه شد. از دنباله تصاویر عمق، رنگی و داده‌های اسکلت به دست آمده از کینکت به عنوان ورودی استفاده شد. مختصات سه‌بعدی مفاصل شانه، آرنج و مچ از روی تصاویر عمق و داده‌های اسکلت به دست آمده از کینکت به‌ازای هر فریم به دست آمدند. با استفاده از روش‌های پردازش تصویر مثل آستانه‌گیری، لبه‌یابی، ساخت پوسته محدب، مدل‌کردن رنگ پوست در فضای YCbCr و تغیریق پس‌زمینه، نوک انگشتان اشاره و شست که به طور مستقیم از روی داده‌های ورودی نمی‌توانستند به دست آیند، شناسایی شدند. از بازوی ربات واقعی استفاده و این بازو به‌ازای هر فریم با استفاده از بردار ویژگی محاسبه شده به روزرسانی شد. به این ترتیب با استفاده از روش برآورد حالت مبتنی بر مدل، حرکت بازوی یک عامل انسانی توسط بازوی ربات مورد تقلید قرار گرفت.

در قیاس با پژوهش‌های پیشین، به طور خلاصه شاخص‌های مورد توجه در این مطالعه به شرح زیر هستند:

- آزمایش‌ها در یک محیط کنترل‌نشده انجام شده است و هیچ‌گونه محدودیتی برای شرایط نورپردازی و یا رنگ‌های موجود در پس‌زمینه در نظر گرفته نشده است.
- برای انجام این پژوهش از داده‌های تصویر و داده‌های عمق دوربین کینکت که ارزان و قابل دسترس می‌باشد، استفاده شده است.
- برای شبیه‌سازی عملکرد باز و بسته‌شدن دهانه ربات، مختصات نوک انگشتان شست و اشاره مورد نیاز است. این اطلاعات به طور مستقیم با استفاده از کینکت به دست نمی‌آیند؛ از این‌رو برای دست‌یابی به مختصات نوک انگشتان، رویکردهایی با استفاده از مبانی بینایی ماشین ارائه شده است.
- برای شناسایی نوک انگشتان، حرکت نوک انگشتان در طول فریم‌ها شناسایی شدند.
- به منظور ارزیابی عملیاتی چنین سامانه‌ای، تقلید عملکرد بازوی کاربر بر روی یک بازوی ربات واقعی RoboTEK II ST240 صورت گرفته است.

- Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 696–706, 2002.
- [25] D. Chai and A. Bouzerdoum, “*A Bayesian approach to skin color classification in YCbCr color space*”, TENCON Proceedings, vol. 2, pp. 421–424, 2000.
- [26] D. Chai and K. N. Ngan, “Locating facial region of a head-and-shoulders color *image*”, Third IEEE International Conference on Automatic Face and Gesture Recognition, pp. 124–129, 1998.
- [27] K.-W. Wong, K.-M. Lam, and W.-C. Siu, “*A robust scheme for live detection of human faces in color images*”, Signal Processing: Image Communication, vol. 18, no. 2, pp. 103–114, 2003.
- [28] J. J. De Dios and N. Garcia, “*Face detection based on a new color space YCgCr*”, International Conference on Image Processing, vol. 3, pp. 906–909, 2003.
- [29] H.-S. Yeo, B.-G. Lee, and H. Lim, “*Hand tracking and gesture recognition system for human-computer interaction using low-cost hardware*”, Multimedia Tools and Applications, vol. 74, no. 8, pp. 2687–2715, 2013.



مریم زارع مهرجردی مدرک کارشناسی خود را در رشته مهندسی کامپیوتر (گرایش نرم افزار) و کارشناسی ارشد را در همان رشته (گرایش هوش مصنوعی و رباتیک) در سال های ۱۳۹۲-۱۳۹۴ از دانشگاه یزد گرفت. زمینه مورد علاقه ایشان شامل پردازش تصویر، رمزنگاری تصویر و شناسایی الگو است.

نشانی رایانامه ایشان عبارت است از:

zaremaryam@stu.yazd.ac.ir
maryamzare1388@yahoo.com



مهندی رضائیان دارای دکترای فتوگرامتری از دانشگاه ETH سوییس بوده و مدرک کارشناسی ارشد خود را در رشته مهندسی پزشکی (بیوالکترونیک) از دانشگاه تهران و کارشناسی خود را در رشته مهندسی برق (الکترونیک) از دانشگاه صنعتی خواجه نصیرالدین طوسی اخذ کرده است. وی هم‌اکنون عضو هیئت علمی گروه مهندسی کامپیوتر دانشگاه یزد است. زمینه پژوهشی مورد علاقه ایشان شامل بینایی ماشین، پردازش تصویر و تشخیص الگو است.

نشانی رایانامه ایشان عبارت است از:

mrezaeian@yazd.ac.ir

Computational intelligence, modelling and simulation, pp. 248–252, 2011.

- [12] Y. Li, “*Hand gesture recognition using Kinect*”, IEEE 3rd International Conference on Software Engineering and Service Science, pp. 196–199, 2012.
- [13] Z. Li and R. Jarvis, “*Real time hand gesture recognition using a range camera*”, Australasian Conference on Robotics and Automation, pp. 21–27, 2009.
- [14] Q. K. Le, C. H. Pham, and T. H. Le, “*Road Traffic Control Gesture Recognition using Depth Images*”, journal of IEEK Transactions on Smart Processing and Computing, vol. 1, no. 1, pp. 1–7, 2012.
- [15] L. Cheng, Q. Sun, H. Su, Y. Cong, and S. Zhao, “*Design and implementation of human-robot interactive demonstration system based on Kinect*”, 24th Chinese Control and Decision Conference, pp. 971–975, 2012.
- [16] K. Qian, J. Niu, and H. Yang, “*Developing a Gesture Based Remote Human-Robot Interaction System Using Kinect*”, International Journal of Smart Home, vol. 7, no. 4, 2013.
- [17] D. Xu, X. Wu, Y.-L. Chen, and Y. Xu, “*Online Dynamic Gesture Recognition for Human Robot Interaction*”, Journal of Intelligent & Robotic Systems, pp. 1–14, 2014.
- [18] A. Billard and M.J. Matarić, “*Learning human arm movements by imitation:: Evaluation of a biologically inspired connectionist architecture*”, Robotics and Autonomous Systems, vol. 37, no. 2 , pp.145-160, 2001.
- [19] S. Filiatrault and A.M. Cretu. “*Human arm motion imitation by a humanoid robot*”, IEEE International Symposium on Robotic and Sensors Environments, pp.31-36, 2014.
- [20] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore, “*Real-time human pose recognition in parts from single depth images*”, Communications of the ACM, vol. 56, no. 1, pp. 116–124, 2013.
- [21] P. Kakumanu, S. Makrigiannis, and N. Bourbakis, “*A survey of skin-color modeling and detection methods*”, Pattern Recognition, vol. 40, no. 3, pp. 1106–1122, 2007.
- [22] J. Yang, W. Lu, and A. Waibel, “*Skin-color modeling and adaptation*,” Springer, 1997.
- [23] V. Vezhnevets, V. Sazonov, and A. Andreeva, “*A survey on pixel-based skin color detection techniques*”, Proc. Graphicon, vol. 3, pp. 85–92, 2003.
- [24] R.-L. Hsu, M. Abdel-Mottaleb, and A. K. Jain, “*Face detection in color images*”, IEEE