استفاده از برآورد حالت‌های پویای دست مبتنی بر مدل، برای تقلید عملکرد بازوی انسان توسط ربات با داده‌های کینکت

مریم 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 دنبال‌کردن 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 برآورد حالت‌های 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 به‌دست‌آمده از کینکت (ابزاری برای ضبط حرکت بدون نشانه‌گذاری) با سرعت سی فریم در ثانیه به‌عنوان داده‌های ورودی استفاده شده‌اند. الگوی پیشنهادی، ویژگی‌های زمانی و مکانی از دنبالهٔ تصاویر ورودی استخراج می‌کند و روی تعیین موقعیت نوک انگشتان شست و اشاره و به‌دست‌آوردن زوایای مفاصل ربات، به‌منظور تقلید حرکت بازوی عامل انسانی در سه بعد در یک محیط کنترل‌نشده تمرکز دارد. در این پژوهش از بازوی ربات واقعی RoboTEK II ST240 استفاده شده و حرکات بازوی عامل انسانی به حرکات تعریف‌شده برای این بازوی ربات محدود شده است. برداراویژگی جهت برآورد حالت به‌ازای هر فریم، به مختصات x، y و عمق برخی 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 به‌طور مستقیم با داده‌های موجود امکان‌پذیر 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، مدل‌کردن رنگ پوست و تفریق پس‌زمینه استفاده می‌شود. در پایان برای تقلید حرکت، با استفاده از بردارهای ویژگی به‌ازای هر فریم، حالت متناظر بر روی بازوی ربات اعمال می‌شود. برای ارزیابی تقلید حرکت، مسیرهای طی‌شده 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.

واژگان کلیدی: برآورد حالت، داده عمق، 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.

* نویسندهٔ عهده‌دار مکاتبات

* Corresponding author

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]

[DOI: 10.29252/jsdp.14.4.97]

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**

۱- مقدمه

برآورد حالت[1] به فرآیندی گفته می‌شود که طی آن مشخص می‌شود که اعضای بدن انسان نسبت به هم چه پیکربندی دارند [1]. 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 به‌علت انعطاف بالای دست، ضبط حرکت دست وظیفه ای 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 سامانه‌هایی حس‌گرهای مغناطیسی و الکترومکانیکی مثل دستکش‌ها و حس‌گرهای نوری به‌طور 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 مایهٔ 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 درحدود سی است وپیداکردن یک حالت متناظر با یک تصویر ورودی به‌علت فضای جستجوی با ابعاد

[1] Pose estimation

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]
[DOI: 10.29252/jsdp.14.4.97]

بالا به محاسبات زیادی نیاز دارد؛ به‌علاوه، با توجه به نوع مفصل‌بندی دست، تعداد هم‌پوشانی‌ها در تصاویر زیاد است. همچنین به‌علت وجود تغییرات روشنایی و پس‌زمینه درهم‌ریخته، محیط برای تعاملات انسان و رایانه غیرقابل کنترل می‌باشد. این موارد سختی‌های بیشتری را در شناسایی دست و استخراج ویژگی تحمیل می‌کنند [2].

روش‌های برآورد حالت دست مبتنی بر بینایی به دو دستهٔ روش‌های مبتنی بر ظاهر[1] و روش‌های مبتنی بر مدل[2] تقسیم می‌شوند. در روش‌های مبتنی بر ظاهر، تصویر ورودی یا ویژگی‌های استخراج‌شده از آن با پایگاه داده‌ای که شامل تصاویر دوبعدی و یا ویژگی‌های اشکال مختلف دست است، مقایسه می‌شود و نزدیک‌ترین حالت به تصویر ورودی به‌دست می‌آید. این روش از سرعت خوبی برخوردار، اما دقت آن به‌شدت وابسته به عمودبودن دست بر محور دوربین است. به‌طورمعمول تعداد زیادی نمونه‌های آموزشی برای آموزش تابع نگاشت از ویژگی‌ها به حالت‌های دست استفاده می‌شوند. با داشتن تابع نگاشت آموزش‌داده‌شده، حالت دست به‌طور مؤثر تخمین زده می‌شود [2].

در روش‌های مبتنی بر مدل، از یک مدل اولیه استفاده می‌شود که به‌طور پیوسته به‌وسیلهٔ ویژگی‌های استخراج‌شده از روی تصویر ورودی به‌روز می‌شود؛ ازاین‌رو اطلاعات مطلوبی شامل حالت در هر زمان فراهم می‌شود. این روش‌ها به محاسبات زیادی نیازمند هستند و درعمل پیاده‌سازی آنها به‌صورت آنی امکان‌پذیر نیست [2].

برآورد حالت دست با استفاده از تصاویر به‌دست‌آمده از دوربین‌ها (رنگی یا عمق) شامل سه مرحلهٔ زیر است:

۱. تشخیص محل دست و جداسازی آن
۲. استخراج ویژگی
۳. تنظیم پارامترهای مدل مورد استفاده با ویژگی‌های استخراج‌شده و در نتیجه به‌روزرسانی مدل

در مرحلهٔ تشخیص محل دست و جداسازی آن، ویژگی‌ها و علائمی برای یافتن ناحیه‌ای از تصویر که شامل دست است استخراج می‌شود. ویژگی‌های استفاده‌شده در تشخیص مکان دست در دوربین‌های رنگی را به دو دسته ویژگی‌های رنگی و شکلی می‌توان تقسیم‌بندی کرد. در ویژگی‌های شکلی از ریخت‌شناسی[3] دست و اعمال محدودیت‌هایی روی درجهٔ آزادی آن، جهت تشخیص مکان 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 به‌صورت مستقیم در [3] تا [6] استفاده شده است.

برای استخراج ویژگی‌های لازم جهت برآورد حالت، بسته به مدل استفاده‌شده و نوع کاربردی که تحلیل ژست دست به دنبال دارد؛ ویژگی‌هایی از قبیل موقعیت نوک انگشتان، تعداد انگشتان، موقعیت کف دست[4] و زاویه‌های بین مفاصل دست می‌توانند استخراج شوند.

طبق چهارچوب ارائه‌شده در [2] الگوی برآورد حالت دست مبتنی بر بینایی ماشین بدون نشانه‌گذاری، با استفاده از یک دنباله تصاویر عمق و رنگی ارائه شده است. اجزای اصلی چهارچوب ارائه‌شده، شامل مدل سه‌بعدی دست، ماژول تجزیه دست، تشخیص‌دهندهٔ نوک انگشت، تعیین‌کنندهٔ پارامترهای حرکت و حل‌کنندهٔ سینماتیک معکوس است. در این مقاله فضای پارامتری حرکت دست به‌وسیلهٔ بردار ویژگی $\phi=\{\phi g,\ \phi l\}$ توصیف می‌شود؛ جایی‌که ϕg حرکت سراسری[5] دست و ϕl حرکت محلی[6] دست است. ϕg شامل انتقال و دوران سه‌بعدی دست و ϕl متناظر با حالت محلی دست با ۲۱ درجه آزادی است.

انتقال سراسری دست، به‌عنوان موقعیت مرکز کف دست $Tg=(xg\ yg\ zg)$ در سامانهٔ مختصات با قاعده دست راست که مبدأ مختصات در مرکز دوربین قرار دارد و دوران سراسری، به‌عنوان زوایای اویلر از دوران کف دست $\Theta g=(\Theta x\ \Theta y\ \Theta z)$ با قرارداد $z\text{-}y\text{-}x$ تعریف می‌شود. برای یافتن مرکز کف دست، با فرض اینکه دست، نزدیک‌ترین شیء به دوربین است و با محدودکردن دوران دست با تعریف 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 ($\Theta x \approx \Theta y \approx 0$) فرض می‌شود که بزرگ‌ترین حباب[7] سراسری در ناحیه دست، کف دست است.

[1] Appearance-based methods
[2] Model-based methods
[3] Morphology
[4] Palm
[5] Global
[6] Local
[7] Blob

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]
[DOI: 10.29252/jsdp.14.4.97]

تجزیه دست به فرآیندی گفته می‌شود که طی آن به هر پیکسل از ناحیهٔ دست جداشده، یک برچسب از مجموعهٔ {کف دست، انگشت کوچک[1]، انگشت انگشتری[2]، انگشت میانی[3]، انگشت اشاره[4]، انگشت شست[5]} نسبت داده شود. بعد از شناسایی کف دست، وظیفه برچسب‌گذاری پیکسل‌های باقیمانده به‌عنوان یک مسئلهٔ دسته‌بندی مطرح که به‌کمک دسته‌بند بیزین[6] حل می‌شود. تشخیص‌دهندهٔ نوک انگشت سه‌بعدی از ایده بیشینه‌کردن فاصله ژئودزیک[7] الهام گرفته است. به‌طور ایده‌آل در هر قسمت برچسب‌گذاری‌شده، نقاطی که فاصله ژئودزیک تا مرکز کف دست را بیشینه می‌کنند، بایستی به‌عنوان نوک انگشت شناخته شوند [7]-[9].

در [10] روش برآورد حالت مبتنی بر مدل با استفاده از تصاویر رنگی و عمق به‌دست‌آمده از کینکت ارائه‌شده است. در این مقاله از مدل دست با ۲۶ درجهٔ آزادی که شامل مش سه‌بعدی با رزولوشن متوسط، ۱۳۱۷ رأس و یک اسکلت متصل‌شده به مدل است، استفاده شده است. حالت دست برمبنای بیست زاویه مابین مفاصل استخوان‌ها تعریف می‌شود. شش پارامتر دیگر، موقعیت و دوران دست را مشخص می‌کنند. الگوریتم ارائه‌شده برای برآورد حالت شامل چهار گام ۱) مقداردهی اولیه ۲) جداسازی ابرنقطه[8] با استفاده از رشد ناحیه ۳) اصلاح اندازه و موقعیت دست و ۴) برآورد حالت نهایی دست است.

در [11] ایدهٔ تشخیص انگشتان در گام‌های زیر خلاصه می‌شود: ۱) گرفتن تصویر عمق از کینکت ۲) دنبال‌کردن 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 پیداکردن موقعیت انگشتان ابتدا ناحیه کف دست به به‌وسیله اعمال یک فیلتر دایره‌ای بزرگ بر روی تصویر عمق به دست می آید، به‌گونه‌ای که همه انگشتان در تصویر حذف خواهند شد. در این مرحله ناحیه کف دست از تصویر اصلی دست کم می‌شود تا ماسک برای جداکردن انگشتان به دست آید. ماسک‌های انگشت با تصویر عمق اصلی ضرب شده تا نگاشت عمق از انگشتان به دست آید.

در [12] و [13] برای پیداکردن نوک انگشتان ابتدا کانتور[9] و پوسته محدب[10] اطراف دست پیدا و سپس محل تلاقی این دو ویژگی محاسبه می‌شود. از آنجایی که نوک انگشتان به‌همراه نقاط دیگری پیدا شده‌اند، نقاط اضافی در دو مرحله حذف می‌شوند؛ ابتدا نقاط بر اساس فاصله، خوشه‌بندی شده و از میان نقاط نزدیک به هم تنها یک نقطه باقی می‌ماند؛ پس از آن با استفاده از روش تطبیق سه نقطه[11] نقاط اضافی باقیمانده از مرحلهٔ 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 به‌دلیل 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 [14] با استفاده از داده‌های عمق به‌دست‌آمده 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 خوشه‌بند ماشین بردار پشتیبان[12] (SVM) برای تشخیص حالت استفاده شده است.

در [15] و [16] 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 واسطهٔ 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 به‌وسیلهٔ کینکت ۲) پردازش تصاویر به‌دست‌آمده از کینکت و تعریف ماژول‌های حالت[13] ۳) انتقال دستورها و داده‌های متناسب با ماژول حالت

[1] Pinky
[2] Ring
[3] Middle
[4] Index
[5] Thumb
[6] Naïve Bayesian classification
[7] geodesic extrema
[8] Point cloud

[9] Contour
[10] Convex hull
[11] Three point alignment
[12] Support vector machine
[13] Gesture module

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]
[DOI: 10.29252/jsdp.14.4.97]

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

در [17] سامانه تشخیص حالت پویای دست به‌صورت برخط[1] با استفاده از دوربین RGB-D ارائه شده است. قبل از دنبال‌کردن دست‌ها، با استفاده از تصاویر عمق و الگوریتم نگاشت فاصله چمفر[2]، دست‌ها تشخیص داده شده و موقعیتشان معلوم شده است. این سامانه‌هم از اطلاعات رنگی و هم اطلاعات عمق در فرآیند دنبال‌کردن دست استفاده کرده است. برای اینکه مسیر حرکت دست مشخص شود، لازم است که نقطه شروع و پایان مسیر حرکت تشخیص داده شود که برای حل این مسئله در حین عملیات دنبال‌کردن دست از دو حالت ثابت، یکی دست در حالت به‌طورکامل باز و دیگری دست در حالت مشت استفاده شده است. در پایان از مدل مخفی مارکوف با ساختار تسمه چپ به راست[3] به‌منظور طبقه‌بندی هر دنباله از مسیرها استفاده شده است.

در [18] مدلی از تقلید حرکت انسانی شامل حرکت‌های دو بازو مورد ارزیابی قرار می‌گیرد. این مدل، از شبکهٔ عصبی برای یادگیری تقلید استفاده می‌کند. مدل اشاره‌شده در شبیه‌سازی به‌وسیلهٔ یک آواتار با ۳۷ درجهٔ آزادی اعتبارسنجی می‌شود. حرکت‌های بازوی انسان به‌وسیلهٔ سامانه‌های دنبال‌کردن مبتنی بر نشانه‌گذاری[4] در قالب ویدئو ضبط می‌شوند و به‌عنوان ورودی به مدل داده می‌شوند.

در [19] برای کنترل حرکت بازوی یک ربات انسان‌نما به نام NAO، سامانه‌ای طراحی شده است که حالت‌های عامل انسانی شامل حرکت دو بازو را در راستاهای مختلف (بدون حرکت انگشتان) مورد تقلید قرار می‌دهد. اطلاعات مربوط به زاویهٔ بازوی عامل انسانی توسط کینکت به‌دست می‌آید. از کینکت به‌عنوان ماژول تعاملی جهت تبدیل اطلاعات زاویه به فرمت مناسب برای ربات استفاده می‌شود. این سامانه به‌صورت زمان واقعی کار می‌کند و برای حل مشکل هم‌پوشانی، ملزم به اعمال محدودیت‌هایی روی حرکت عامل انسانی است.

در این مقاله هدف این است تا با استفاده از روش ضبط حرکت بدون نشانه‌گذاری و مفاهیم بینایی ماشین، حرکت بازوی عامل انسانی (که شامل حرکت دورانی بازو حول مفصل شانه[5] به سمت داخل و خارج، حرکت دورانی حول مفصل شانه به سمت بالا و پایین، حرکت دورانی حول مفصل آرنج[6] به سمت بالا و پایین و باز و بسته‌شدن انگشتان شست و اشاره از هم است.)، توسط بازوی ربات، در یک محیط کنترل‌نشده تقلید شود. این مسئله را برآورد حالت مبتنی بر مدل گویند. برای این کار در ابتدا حرکت بازوی انسانی در قالب دنباله‌ای از تصاویر یا فریم‌ها به‌عنوان ورودی تهیه و سپس با استفاده از رویکردهای استخراج ویژگی، به‌ازای هر فریم، بردار ویژگی جهت توصیف فضای پارامتری حرکت بازو استخراج می‌شود. به‌منظور استخراج ویژگی‌ها از ترکیب داده‌های عمق، رنگی و اسکلت به‌دست‌آمده از کینکت و برای نمایش، از داده‌های رنگی استفاده شده است. برخی از ویژگی‌ها به‌طور مستقیم از اطلاعات عمق و اسکلت به‌ازای هر فریم استخراج می‌شوند؛ درحالی‌که دیگر ویژگی‌ها از جمله نوک انگشتان شست و اشاره با به‌کارگیری مفاهیم بینایی ماشین، تحت سه رویکرد مطرح‌شده در ادامه که سهم اصلی کار این پژوهش نیز است، به‌دست می‌آیند. در پایان به‌ازای هر فریم، بازوی ربات متناسب با بردار ویژگی استخراج‌شده، از طریق تنظیم پارامترهای مربوط به مفاصل ربات، در حالت متناظر با حالت دست کاربر قرار می‌گیرد. برآورد حالت توسط بازوی ربات به‌ازای فریم‌های متوالی یک دنباله ورودی، باعث تقلید حرکت می‌شود؛ به این معناست که به‌ازای هر فریم ورودی، یک فریم از عملکرد ربات به‌عنوان خروجی در دست خواهد بود. با کنار هم قراردادن فریم‌های خروجی در مجموع، حرکت ربات به‌دست می‌آید.

نکته قابل توجه این است که به‌علت کندبودن عملکرد بازوی ربات موجود برای قرارگرفتن در وضعیت مناسب، تقلید همزمان حرکت ممکن نیست. ربات برای قرارگیری در وضعیت مناسب به‌ازای هر فریم ورودی، به مدت زمانی نیاز دارد که این مدت زمان خیلی بیشتر از حرکت دست است؛ بنابراین اگرچه تقلید همزمان حرکت از اهداف این پژوهش نیست؛ ولی در صورت استفاده از یک بازوی ربات سریعتر، الگوریتم ارائه شده قابلیت اجرا به‌صورت برخط را نیز خواهد داشت.

در ادامه ساختار مقاله به‌ترتیب زیر خواهد بود. بخش۲ مفاهیم اولیه، بخش۳ روش پیشنهادی برای استخراج بردار ویژگی، بخش۴ نتایج آزمایش‌ها و بخش۵ نتیجه‌گیری را نشان می‌دهد.

[1] Online
[2] Chamfer distance matching algorithm
[3] Left-Right banded structure
[4] marker-based tracking systems
[5] Shoulder
[6] Elbow

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]
[DOI: 10.29252/jsdp.14.4.97]

۲- مفاهیم اولیه

۲-۱- تشخیص اسکلت بدن با استفاده از دادهٔ عمق

شاتون و همکارانش برای پیش‌بینی سریع و دقیق موقعیت مفاصل بدن، روشی مبتنی بر تک‌تصویر عمق و بدون دخالت هیچ اطلاعات زمانی ارائه دادند. در این روش از ایده شناسایی شیء[1] و طراحی یک نمایش میانی از بخش‌های بدن[2]، برای نگاشت مسئله برآورد حالت‌های سخت، به مسئلهٔ ساده‌تر طبقه‌بندی به‌ازای هر پیکسل استفاده شده است [20].

استفاده از پایگاه دادهٔ آموزشی بسیار بزرگ و متنوع، به خوشه‌بند اجازه می‌دهد تا نسبت به شکل بدن، نوع لباس پوشیدن و انواع حالت‌ها، غیرقابل تغییر باشد. از آنجایی‌که خوشه‌بند از اطلاعات زمانی استفاده نمی‌کند، تنها از حالت‌های ایستا و نه حرکت[3] برای آموزش استفاده می‌شود. تغییرات در حالت یک فریم از حرکت، با فریم بعدی حرکت بسیار کوچک و قابل صرف‌نظرکردن است. بنابراین بسیاری از حالت‌های مشابه اضافی در دنبالهٔ حرکت، با خوشه‌بندی دورترین همسایگی[4] دور ریخته می‌شوند.

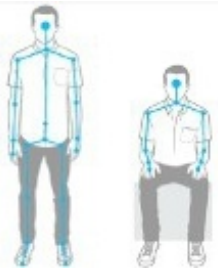
برای ساخت نمایش میانی از بخش‌های بدن، سهم کلیدی کار این بوده که کل بدن توسط ۳۱ بخش مطابق شکل (۱) برچسب‌گذاری شده است. ۳۱ برچسب مشخص‌شده برای بدن بدین قرار است: اگر بدن انسان به دو قسمت چپ و راست تقسیم شود، برای هر کدام از قسمت‌ها، پانزده برچسب با نام‌های، قسمت بالایی سر، قسمت پایینی سر، شانه، قسمت بالایی بازو، قسمت پایینی بازو، آرنج، مچ، دست، قسمت بالایی کمر، قسمت پایینی کمر، ران پا، زانو، ساق پا، قوزک پا، مچ پا وجود دارد. ۳۱امین برچسب متعلق به گردن است. بعضی از بخش‌ها به‌طور مستقیم موقعیت مفصل خاصی را مشخص می‌کنند؛ درحالی‌که 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 خوشه‌بندی می‌توان حل کرد. جفت‌تصاویر عمق و تصاویر برچسب‌گذاری‌شده متناظر با آن برای آموزش خوشه‌بند استفاده می‌شوند. با استفاده از خوشه‌بند جنگل تصادفی[5] و استخراج ویژگی از روی تصویر عمق، به‌ازای هر پیکسل از تصویر عمق، برداری شامل ۳۱ مقدار به‌دست می‌آید. این بردار نشان‌دهندهٔ این مطلب است که هر پیکسل با چه احتمالی متعلق به هر یک از ۳۱ بخش برچسب‌گذاری‌شدهٔ بدن می‌تواند باشد. درنهایت، محاسبه مرکزهای توده[6] احتمالی برای هر بخش شناسایی‌شده، یک راهنما برای به‌دست‌آوردن موقعیت مفاصل است.

با استفاده از این الگوریتم برای حالت ایستاده بیست مفصل و برای حالت نشسته ده مفصل را می‌توان شناسایی کرد (شکل ۲).



(شکل-۱): ۳۱ برچسب برای حالت‌های مختلف [20]

(Figure-1): 31 labels for different poses [20]



(شکل-۲): مفاصل شناسایی‌شده در حالت نشسته و ایستاده [20]

(Figure-2): recognized joints in standing and sedentary poses [20]

۲-۲- درجه‌بندی عملیاتی ربات

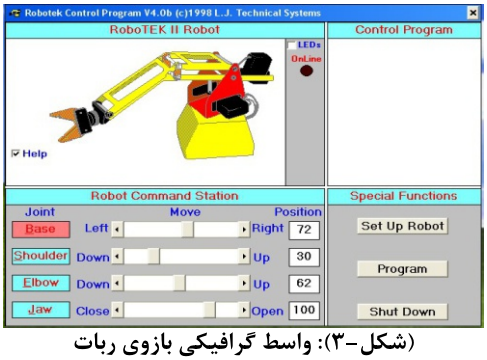
در این پژوهش از بازوی ربات RoboTEK II ST240 استفاده شده است که با سه درجه آزادی دورانی از گروه ربات‌های آرنجی است. بازوی ربات مورد استفاده دارای واسط گرافیکی (شکل۳) است. در واسط گرافیکی، مفصل‌های مورد استفاده و محدودهٔ کاری هر مفصل نشان داده شده است. محدودهٔ کاری هر مفصل با اعداد ۱ تا ۱۲۷ نشان داده می‌شود. میزان دوران به‌ازای هر یک واحد در این محدوده نه‌تنها 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

[1] Object recognition
[2] Intermediate body parts representation
[3] Motion
[4] Furthest neighbor clustering
[5] Randomized decision forests classifiers
[6] Mass

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]

[DOI: 10.29252/jsdp.14.4.97]

آن مفصل به‌طور تجربی به‌دست می‌آید. درجه‌بندی عملیاتی ربات به‌صورت دستی انجام و بخشی از خطاهای موجود در قسمت نتیجه نیز از همین جا ناشی می‌شود. برای مثال، شکل (۴) محدودهٔ کاری مفصل آرنج در راستای بالا-پایین را با تنظیم‌کردن متغیر Elbow در واسط گرافیکی، یک بار به‌ازای عدد ۱ و بار دیگر به‌ازای عدد ۱۲۷ نشان می‌دهد. شکل (۵) حالت پایهٔ درنظرگرفته‌شده را برای مفصل آرنج به‌منظور درجه‌بندی عملیاتی نشان می‌دهد. واسط گرافیکی مورد استفاده قادر به دریافت دستورها به‌عنوان ورودی است؛ و حالت متناظر با آن دستور را بر روی بازوی ربات واقعی اعمال می‌کند. دستور ورودی به ربات نیازمند چهار مقدار ورودی به‌ترتیب، میزان دوران بازو حول مفصل شانه به سمت داخل و خارج (Base در واسط گرافیکی)، میزان دوران حول مفصل شانه به سمت بالا و پایین (Shoulder)، میزان دوران حول مفصل آرنج به سمت بالا و پایین (Elbow) و میزان بازشدن قسمت مجری نهایی[1] (Jaw) است. برای تمام مفاصل میزان دوران، نسبت به حالت پایهٔ به‌کارگرفته‌شده برای آن مفصل (چون حالت پایه درعمل معادل صفر درجه است.) در نظر گرفته می‌شود. در ادامه برای فهم بهتر و اشتباه‌نگرفتن زاویهٔ شانه و آرنج واسط گرافیکی با مفاصل شانه و آرنج، به جای زاویه‌های Base، Shoulder، Elbow و Jaw واسط گرافیکی، به‌ترتیب زاویهٔ پایه، زاویهٔ شانه، زاویهٔ آرنج و زاویهٔ مجری‌نهایی استفاده می‌شود.



(شکل-۳): واسط گرافیکی بازوی ربات

(Figure-3): graphical interface for robot arm



الف) وضعیت ربات با تنظیم عدد ۱ در محدوده کاری مفصل آرنج

a) state of robot with setting 1 in elbow workspace



ب) وضعیت ربات با تنظیم عدد ۱۲۷در محدوده کاری مفصل آرنج

b) 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



(شکل-۵): حالت پایهٔ درنظرگرفته‌شده 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 زاویهٔ 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

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]

[DOI: 10.29252/jsdp.14.4.97]

۳- روش پیشنهادی

۳-۱- جمع‌آوری پایگاه داده

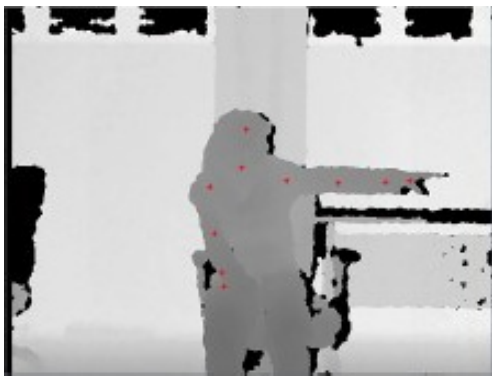
در این پژوهش با استفاده از دوربین کینکت، حرکات بازوی عامل انسانی در قالب دنباله تصاویر رنگی، عمق و داده‌های اسکلت به‌طور هم‌زمان با سرعت سی فریم بر ثانیه ضبط شدند. شکل (۶) نمونه‌ای از هر یک از داده‌های ورودی را نشان می‌دهد.



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

۳-۲- استخراج بردار ویژگی از روی داده‌ها

بردار ویژگی جهت برآورد حالت، شامل مختصات دوبعدی (مختصات x و y) و در برخی موارد سه‌بعدی (مختصات x ،y و عمق) برای نقاط مشخص‌شده در شکل (۷) است. با استفاده از داده‌های اسکلت به‌دست‌آمده از دوربین کینکت، مختصات x و y مفصل شانه، مفصل آرنج، مفصل مچ بر حسب پیکسل به‌ازای هر فریم به‌دست می‌آید. مقدار فاصلهٔ مفاصل مچ و شانه از دوربین کینکت، با استفاده از تصویر عمق بر حسب میلی‌متر به‌دست می‌آید. موقعیت نوک انگشتان اشاره و شست به‌طور مستقیم از روی داده‌های موجود در دسترس نیست. بنابراین به‌دست‌آوردن مختصات نوک انگشتان به‌عنوان بخشی از بردار ویژگی و سهم اصلی کار این پژوهش، با به‌کارگیری رویکردهایی که در ادامه بیان می‌شوند، یکی از اهداف این پژوهش است.



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

۳-۳- شناسایی نوک انگشتان اشاره و شست

به‌دست‌آوردن موقعیت نوک انگشتان اشاره و شست برای تعیین میزان بازشدن مجری نهایی ربات لازم و ضروری است. بعد از پیداکردن نوک انگشتان بایستی نقطه‌ای برای قرارگرفتن در مبدأ مختصات مشخص شود تا مجموع زوایایی که نوک انگشتان با محور مختصات می‌سازند، میزان بازشدن انگشتان از یکدیگر را نشان دهد (شکل۸). لازم به ذکر است که در این مورد موقعیت کف دست حاصل از کینکت برای استفاده به‌عنوان مبدأ مختصات مناسب نیست. ازاین‌رو با پیداکردن کمینهٔ محلی در مرز دست بین انگشتان شست و اشاره، مبدأ مناسب به‌دست می‌آید.

همان‌طورکه درقبل اشاره شد، مختصات سه نقطه از مجموعهٔ نقاط شکل (۷) (مفصل شانه، مفصل آرنج، مفصل

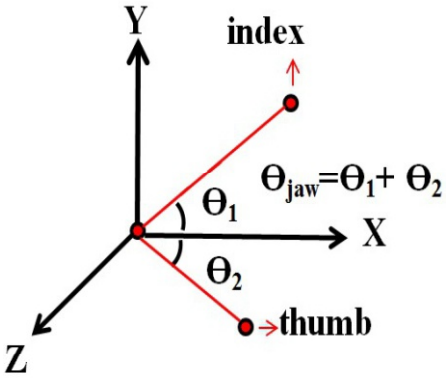
[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]

[DOI: 10.29252/jsdp.14.4.97]

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]

[DOI: 10.29252/jsdp.14.4.97]

مچ) به‌طور مستقیم از روی داده‌های ورودی به‌دست می‌آیند. برای پیداکردن مختصات سه نقطهٔ باقیمانده (نوک انگشتان و نقطه بین دو انگشت)، در ادامه سه رویکرد ارائه می‌شود.



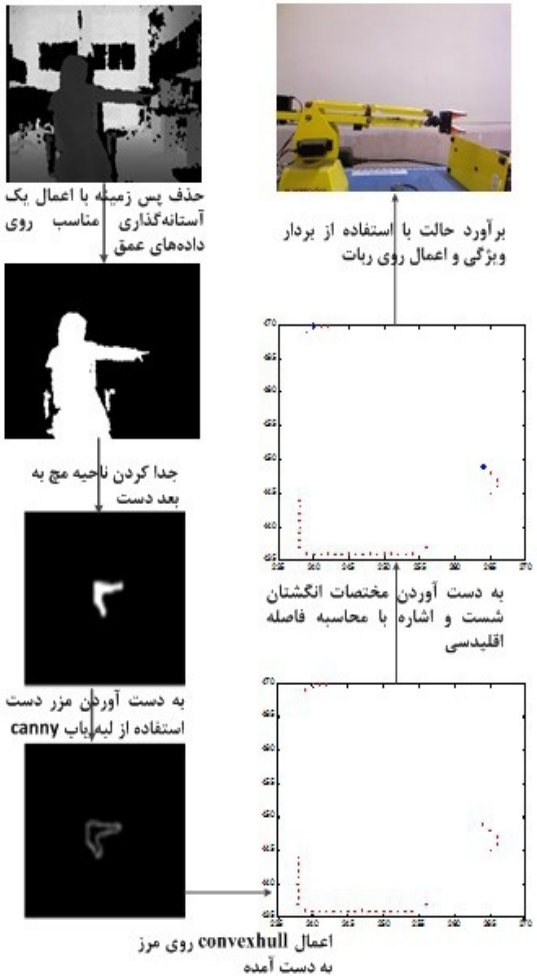
(شکل-۸): محاسبه مقدار بازشدن انگشتان از هم

(Figure-8): calculating value of the opening fingertips at each other

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

در این روش با الهام از مقاله‌های [3] تا [6]، برای به‌دست‌آوردن 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 مچ دست در داده‌های اسکلت، از قسمت مچ دست به جلو نگه داشته شده و مابقی حذف می‌شوند. در ادامه نقاط حاصل از تلاقی کانتور و پوسته محدب اعمال‌شده بر روی ناحیه باقیمانده به‌دست 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 فاصلهٔ اقلیدسی، نقطهٔ انگشت اشاره را نشان می‌دهد. انگشت شست نیز در ادامه با استفاده از میزان فاصله‌های اقلیدسی به‌دست‌آمده و اعمال این محدودیت که انگشت شست همیشه پایین‌تر از انگشت اشاره قرار می‌گیرد به‌دست می‌آید. اعمال این محدودیت به‌دلیل این است که قسمت مجری نهایی ربات مورد استفاده، قابلیت چرخش ندارد و همواره ثابت است. ازاین‌رو برای کاربر نیز فرض کردیم که در طول حرکت، انگشت اشاره همواره بالاتر از انگشت شست قرار دارد. شکل (۹) مراحل ذکرشده و به‌روزرسانی ربات توسط بردار ویژگی به‌دست‌آمده را نشان می‌دهد.

در این روش به‌دلیل ازدست‌رفتن برخی اطلاعات عمق، تعیین موقعیت نوک انگشت شست در برخی از فریم‌ها امکان‌پذیر نیست (شکل ۱۰). این مشکل را با فرض اینکه انگشتان شست و اشاره، هماهنگ با هم و به یک اندازه در دو جهت مخالف حول نقطه مبدأ دوران داده می‌شوند، می‌توان برطرف کرد و به تعیین موقعیت نوک انگشت اشاره بسنده کرد. دو برابر زاویهٔ به‌دست‌آمده برای انگشت اشاره (Θ_1 در شکل۸)، میزان بازشدن انگشتان را از هم نشان می‌دهد.



(شکل-۹): به‌دست‌آوردن موقعیت نوک انگشتان و به‌روزرسانی ربات

(Figure-9): fingertips identification and update the robot



(شکل-۱۰): ازدست‌رفتن اطلاعات عمق برای انگشت شست

(Figure-10): losing some of depth information

۲-۳-۳- مدل‌کردن رنگ پوست برای شناسایی انگشتان

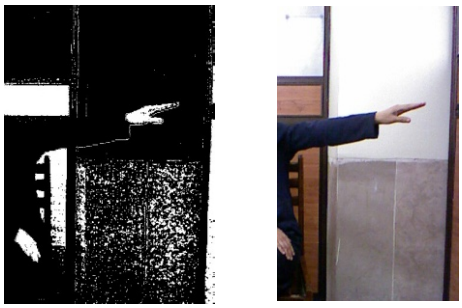
در این روش برای شناسایی انگشتان از اطلاعات رنگ پوست استفاده می‌شود. گام‌های مقدماتی برای شناسایی پوست در یک تصویر به‌وسیلهٔ اطلاعات رنگی به قرار زیر است [21]: ۱) نمایش پیکسل‌های پوست در فضای رنگی مناسب ۲) مدل‌کردن پیکسل‌های پوست و غیر پوست به‌وسیلهٔ توزیع آماری مناسب ۳) طبقه‌بندی‌کردن توزیع‌های مدل‌شده.

انتخاب فضای رنگی به‌عنوان گام ابتدایی در طبقه‌بندی رنگ پوست در نظر گرفته می‌شود. فضای رنگی همچنین تعیین می‌کند که چقدر به‌طور مؤثر توزیع رنگ پوست را می‌توان مدل کرد. فضای رنگی RGB، فضای رنگی پیش فرض برای اغلب فرمت‌های تصاویر موجود است. به‌طورمعمول فضاهای رنگی دیگر از تبدیل خطی یا غیر خطی از فضای RGB به‌دست می‌آیند. تبدیل فضای رنگی، برای کاهش هم‌پوشانی بین پیکسل‌های پوست و غیر پوست فرض شده است و ازاین‌رو به فرآیند دسته‌بندی پیکسل‌های پوست کمک می‌کند و دقت بالایی را در شرایط نورانی متغیر فراهم می‌کند. به‌طورمعمول اجزای اصلی رنگ پوست برای افراد مختلف با توجه به میزان شدت روشنایی با هم متفاوت هستند [22]؛ ازاین‌رو یک عملکرد رایج، حذف شدت روشنایی برای وظیفهٔ 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 مشخص‌کردن رنگ پوست مناسب نیست. شکل (۱۱) تصویر رنگی RGB و تصویر دودویی متناظر آن، که با اعمال مقادیر آستانه اشاره‌شده برای هر جز سه‌گانه در فرمول (۱) [23]، پیکسل‌های پوست را از غیر پوست جدا می‌کند، نشان می‌دهد.

$$R>95 \text{ and } G>40 \text{ and } B>20 \text{ and } \max\{R,G,B\}-\min\{R,G,B\}>15 \text{ and } |R-G|>15 \text{ and } R>G \text{ and } R>B \quad (۱)$$

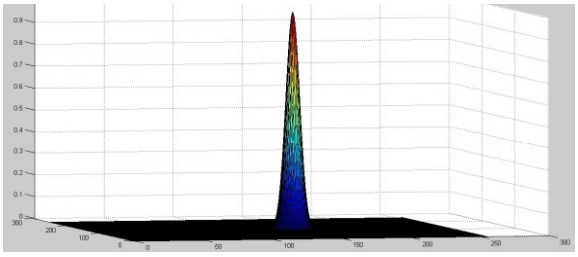
هر پیکسلی که در رابطه (۱) صدق کند، به‌عنوان پوست در نظر گرفته می‌شود و در غیر این صورت آن پیکسل متعلق به طبقه پوست نیست؛ اما به‌دلیل این‌که در یک محیط کنترل‌نشده، در پس‌زمینه، رنگ‌های مشابه رنگ پوست وجود دارند و تصاویر رنگی به‌دست آمده از کینکت، کیفیت بالایی ندارند؛ استفاده از فضای رنگی RGB، نتیجهٔ خوبی را برای جداسازی پیکسل پوست از غیر پوست به‌دست نمی‌آورد.



(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 جداکردن نواحی پوست از غیر پوست نیست. بعضی از فضاهای رنگی، افزونگی موجود در کانال‌های رنگی RGB را کاهش می‌دهند و رنگ را با اجزای مستقل آماری نمایش می‌دهند. در این فضای رنگی، شدت روشنایی و اجزای اصلی رنگ به‌صورت واضح از هم مجزا می‌شوند. این فضاها 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 به‌ترتیب 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 به‌عنوان 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 یک تابع گوسی در فضای رنگی YCbCr می‌تواند مدل شود (شکل۱۲).



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

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

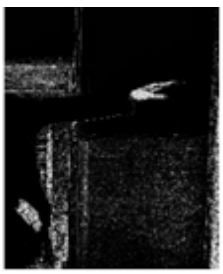
[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]

[DOI: 10.29252/jsdp.14.4.97]

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[1] 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) به‌ترتیب تصویر رنگی، تصویر سطح خاکستری نمایش‌دهندهٔ مقدار احتمال‌ها و تصویر دودویی جداکنندهٔ رنگ پوست و غیر پوست را نشان می‌دهند.



الف) تصویر رنگی
a) color image



ب) تصویر سطح خاکستری نشان‌دهندهٔ میزان احتمال‌ها
b) gray scale image showing the likelihood of pixels



پ) تصویر باینری جداکنندهٔ رنگ پوست و غیرپوست
c) binary image separating skin and non-skin color

(شکل -۱۳): استفاده از مدل‌کردن رنگ پوست برای شناسایی پوست

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

[1] Likelihood

از آنجایی‌که پس‌زمینه شامل رنگ‌هایی، مشابه رنگ پوست و جمع‌آوری پیکسل پوست مورد نیاز برای مدل‌کردن رنگ پوست با مشکلاتی همراه است؛ تعیین مقدار آستانه مناسب برای جداکردن پیکسل پوست از غیر پوست دشوار است. برای بهبود فرآیند شناسایی پیکسل‌های پوست، با استفاده از داده‌های عمق کینکت، عمق متناظر نقاط را در قسمت (پ) از شکل (۱۳) می‌توان به دست آورد و با آستانه‌گذاری مناسب، پیکسل‌های غیر پوست را حذف کرد. همان‌طورکه در قسمت الف از شکل (۱۳) دیده می‌شود دست‌های کاربر نسبت به دیوار پشت سر، به کینکت نزدیک‌تر هستند. شکل (۱۴) مؤثربودن به‌کارگیری اطلاعات عمق برای دور ریختن پیکسل‌های غیر پوست را نشان می‌دهد. نوک انگشتان با استفاده از تلاقی کانتور و پوسته محدب اعمال‌شده بر روی مرز دست، همانند رویکرد قبلی حاصل می‌شوند.

برای برخی فریم‌ها که قسمت مچ دست اشتباه تشخیص داده شده است (پس‌زمینه هم به‌عنوان بخشی از مچ دست تشخیص داده شده است.) و اطلاعات عمق نیز در آن نواحی در دست نیست که جهت جداسازی پس‌زمینه مورد استفاده قرار گیرد؛ به‌کارگیری این روش، مختصات نوک انگشتان را به‌درستی تشخیص نمی‌دهد.



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

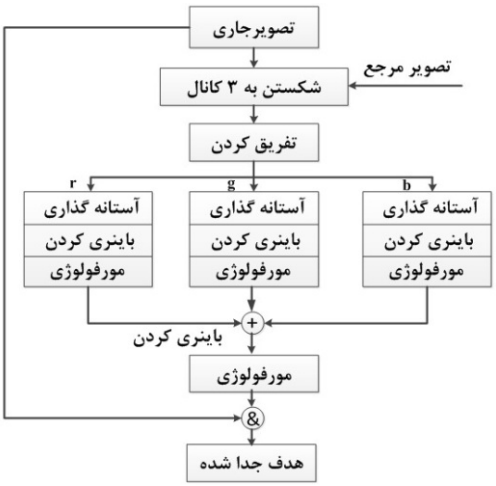
(Figure-14): using of depth data for removing non-skin pixels

۳-۳-۳- استفاده از روش تفریق برای شناسایی انگشتان

ساده‌ترین روش برای به‌دست‌آوردن قسمت هدف، تفریق‌کردن 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.

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]

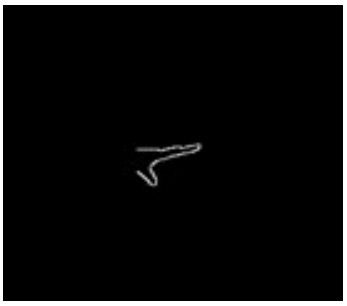
[DOI: 10.29252/jsdp.14.4.97]



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

اختلاف ۳ کانال ($r\ g\ b$) هر تصویر با کانال‌های متناظر در تصویر مرجع محاسبه می‌شود. نقاطی که میزان اختلاف بیشتری دارند، به‌عنوان هدف، در نظرگرفته می‌شوند. در این مرحله با انتخاب آستانهٔ مناسب برای هر کانال ، هدف را به‌درستی 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 جداشده 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 جداکردن هدف از پس‌زمینه، با داشتن مختصات مچ دست در داده‌های اسکلت، ناحیه مچ دست به جلو مطابق شکل (۱۶) جدا می‌شود. در ادامه با اعمال لبه‌یاب و پیداکردن کمینه و بیشینهٔ محلی در مرز به‌دست‌آمده به‌ترتیب مبدأ مختصات موردنظر و نوک انگشتان به‌دست می‌آیند.



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

۴-۳- تبدیل ویژگی‌های استخراج‌شده به زوایای مناسب ربات

همان‌طورکه درقبل بیان شد، به‌ازای هر فریم، یک بردار ویژگی استخراج می‌شود. برخی از ویژگی‌ها، شامل مختصات x و y نقطهٔ خاصی از روی تصویر هستند و درواقع موقعیت پیکسل خاصی را نشان می‌دهند و اطلاعات پیکسلی هستند؛ درحالی‌که 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 تنظیم‌شده 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 اشاره‌شده عمل می‌کند و در مبدأ مختصات قرار می‌گیرد. مطابق شکل (۱۷) زاویه‌ای که مفصل مچ 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 مچ و شانه، از روی داده‌های عمق به‌دست‌آمده از کینکت به دست می‌آیند. لازم به ذکر است از آنجایی‌که بازوی ربات موجود قابلیت دوران نسبت به مفصل آرنج به سمت داخل و خارج را ندارد، استفاده از مفصل مچ و آرنج در این مورد یکسان است.

با تغییر زاویهٔ 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.

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]
[DOI: 10.29252/jsdp.14.4.97]

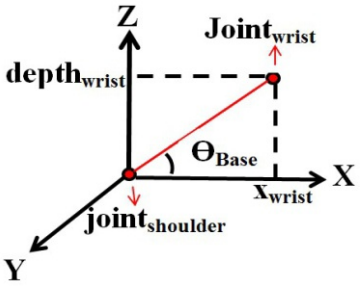
[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]

[DOI: 10.29252/jsdp.14.4.97]

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

زاویه مچری نهایی نیز مطابق شکل (۸) بعد از به‌دست‌آمدن نوک انگشتان و نقطه میانی دو انگشت، با استفاده از فرمول (۵) به‌دست می‌آید.

مقدار زاویه‌های به‌دست‌آمده بر حسب درجه هستند؛ بنابراین بعد از به‌دست‌آوردن 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 واسط گرافیکی می‌توان انتخاب کرد.



(شکل-۱۷): به‌دست‌آوردن Θ_{Base} در محور مختصات X-Z

(Figure-17): calculating Base angle in x-z coordinate plane

$$\Theta_{Base} = \text{Arctan}\left(\frac{depth_{shoulder} - depth_{wrist}}{x_{wrist} - x_{shoulder}}\right) \quad (۲)$$

$$\Theta_{Shoulder} = \text{Arctan}\left(\frac{y_{elbow} - y_{shoulder}}{x_{elbow} - x_{shoulder}}\right) \quad (۳)$$

$$\Theta_{Elbow} = \text{Arctan}\left(\frac{y_{wrist} - y_{elbow}}{x_{wrist} - x_{elbow}}\right) - \Theta_{Shoulder} \quad (۴)$$

$$\Theta_{Jaw} = \text{Arctan}\left(\frac{y_{index} - y_{palm}}{x_{index} - x_{palm}}\right) + \text{Arctan}\left(\frac{y_{thumb} - y_{palm}}{x_{thumb} - x_{palm}}\right) \quad (۵)$$

با توجه به یکسان‌نبودن کمینه و بیشینهٔ بازشدن مچری نهایی ربات و انگشتان شست و اشاره دست از هم، برای تناظریابی بین حالت مچری نهایی ربات و قسمت مچ دست، زاویه‌های به‌دست‌آمده 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 نرمالیزه‌شده مشخص می‌شوند.

۴- نتایج آزمایش‌ها

به‌منظور مقایسهٔ سه رویکرد مطرح‌شده برای شناسایی موقعیت نوک انگشتان شست و اشاره که از مقالات مختلف الهام گرفته شده‌اند، برای سی فریم تصادفی که لزوماً فریم‌های متوالی یک دنباله نیستند، جدول (۲) به‌دست آمده است.

(جدول-۲): نتیجه مقایسه رویکردهای شناسایی نوک انگشتان برای سی فریم

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

	روش اول [۳]	روش دوم [۲۲]	روش سوم روش پیشنهادی
تعداد فریم‌های درست تشخیص‌داده‌شده	۲۱	۱۸	۳۰

ذکر این نکته نیز ضروری است که هدف ما از تقلید یک حرکت، تقلید تمام فریم‌های متوالی بین حرکت است؛ یعنی حرکت از یک حالت اولیه شروع شده و به‌تدریج به سمت حالت نهایی پیش می‌رود. حال اگر در حین حرکت تدریجی، برای یک فریم، یک حالت به‌شدت متفاوت با فریم‌های قبلی اتفاق افتد، توالی حرکت از بین می‌رود. به‌همین 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 متوالی مورد تقلید قرار گرفته است.

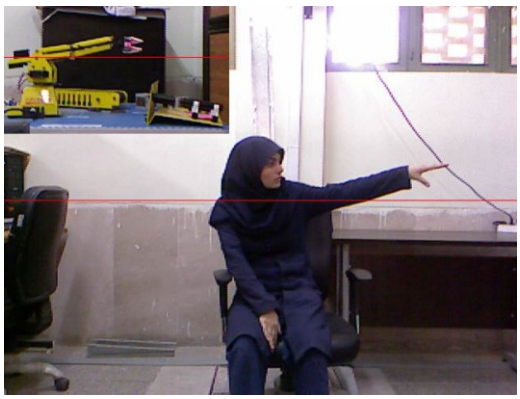
دنبالهٔ شکل (۱۸) حرکت دورانی را نسبت به مفصل آرنج، ابتدا به سمت بالا و سپس به سمت پایین نشان می‌دهد. رسم خطوط قرمزرنگ برای فهم بهتر خواننده است تا تغییرات حرکت در طول فریم‌ها را بهتر درک کند.

شکل (۱۹) حرکت دورانی بازو به سمت داخل و سپس به سمت خارج، با مبنا قراردادن خطوط قرمزرنگ را نشان می‌دهد. شکل (۲۰) حرکتی را نشان می‌دهد که باز و بسته‌شدن انگشتان شست و اشاره از هم با استفاده از قسمت مچری 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 جابه‌جایی محسوسی ندارد. ازاین‌رو استخراج ویژگی‌ها از روی فریم ورودی با استفاده از فرمول‌های ارائه‌شده ممکن است.

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]

[DOI: 10.29252/jsdp.14.4.97]



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

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



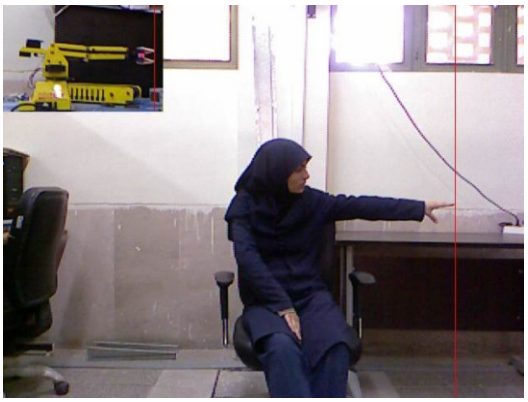










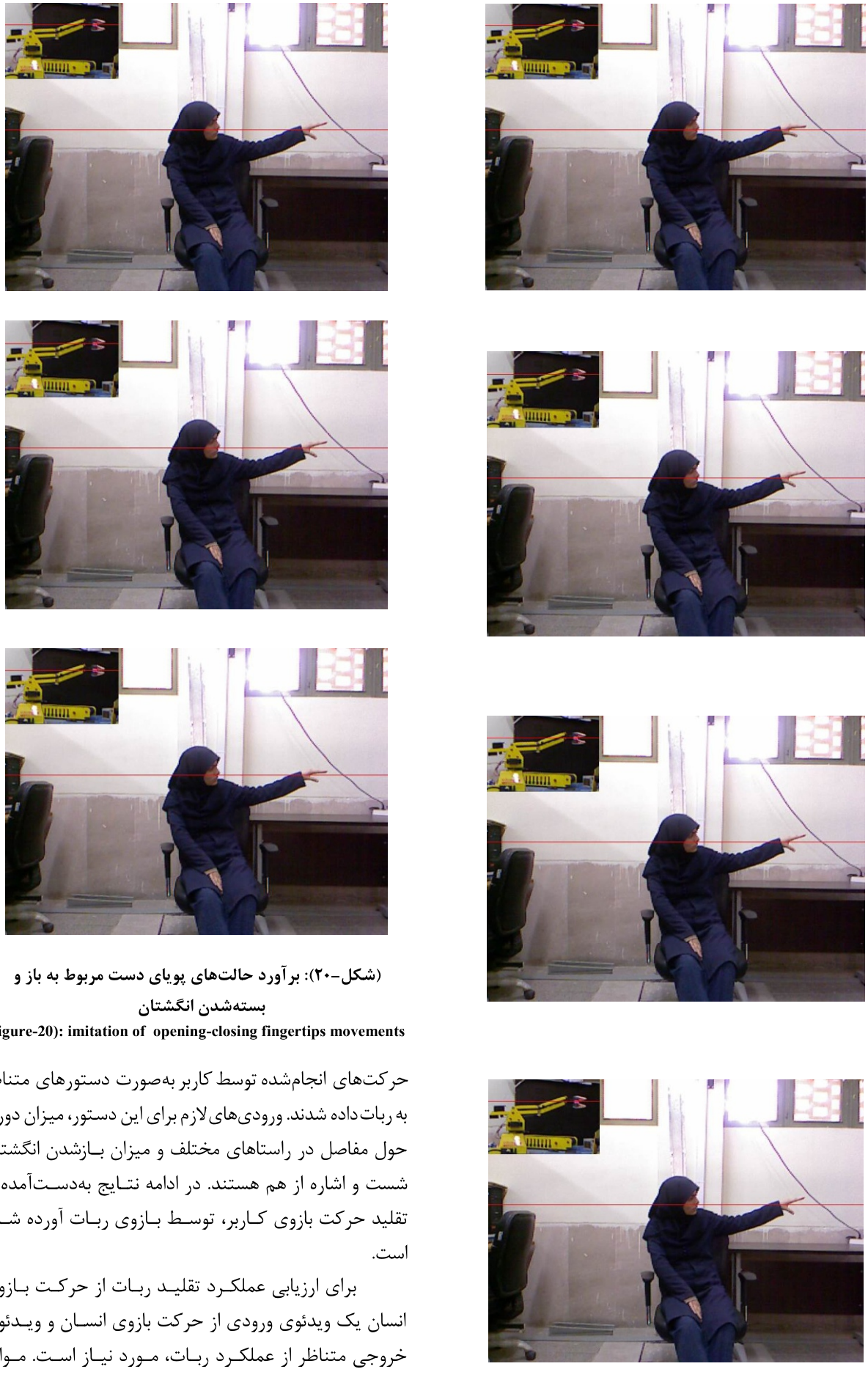


(شکل-۱۹): برآورد حالت‌های پویای دست مربوط به حرکت دورانی به سمت داخل و خارج

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

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]

[DOI: 10.29252/jsdp.14.4.97]



(شکل-۲۰): برآورد حالت‌های پویای دست مربوط به باز و بسته‌شدن انگشتان

(Figure-20): imitation of opening-closing fingertips movements

حرکت‌های انجام‌شده توسط کاربر به‌صورت 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 به‌دست‌آمده 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

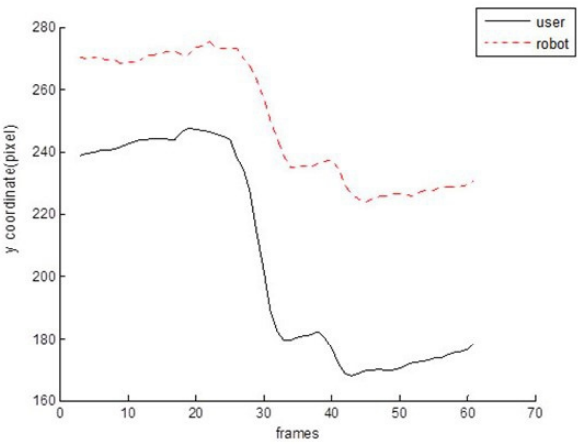
[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]

[DOI: 10.29252/jsdp.14.4.97]

اشاره‌شده در ادامه برای هر دو ویدئو محاسبه و مقایسه می‌شوند.

مسیر طی‌شده (مختصات y 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 طی‌شده توسط مجری نهایی ربات (رنگ قرمز) که با استفاده از الگوریتم دنبال‌کردن[1] کانادە-لوکاس-توماسی[2] (KLT) به‌دست آمده‌اند، در شکل (۲۱) مقایسه شده است. ویدئوی مورد استفاده جهت برآورد حالت شامل ۶۳ فریم است. به‌علت متفاوت‌بودن طول بازوی ربات و عامل انسانی، دو نمودار موجود در شکل (۲۱) روی هم قرار نگرفته‌اند. شکل‌های۲۲، 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.

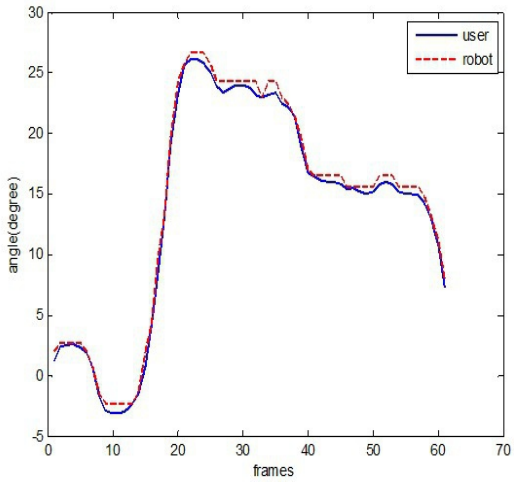
شکل (۲۵) زاویهٔ بین مجری نهایی ربات را با زاویهٔ بین انگشتان دست در فضای نرمالیزه‌شده در طول فریم‌ها مقایسه می‌کند. در جدول (۳) میانگین و واریانس اختلاف زاویه‌های برآوردشده برای زاویه‌های پایه، شانه، آرنج و مجری نهایی بازوی ربات با زوایای مفاصل بازوی کاربر نشان داده شده است. همان‌طورکه درقبل اشاره شد، بازوی ربات مورد استفاده فاقد درجه‌بندی دقیق و درجه‌بندی به‌صورت دستی انجام گرفته است. بخشی از این اختلاف‌ها از همین مسئله ناشی شده است. اعداد به‌دست‌آمده برای زاویهٔ مجری نهایی در فضای نرمالیزه‌شده هستند.



(شکل-۲۱): مسیر طی‌شده توسط مجری نهایی و قسمت نهایی دست کاربر در راستای بالا به پایین و برعکس

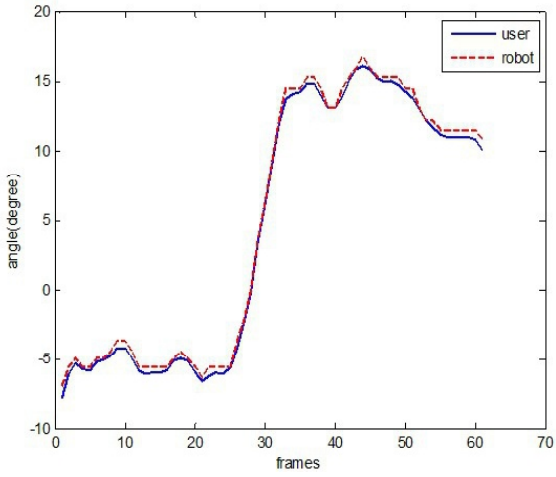
(Figure-21): tracked trajectory by user's wrist and robot end effector in up-down and reverse

[1] Tracking algorithm
[2] Kanade-Lucas-Tomasi



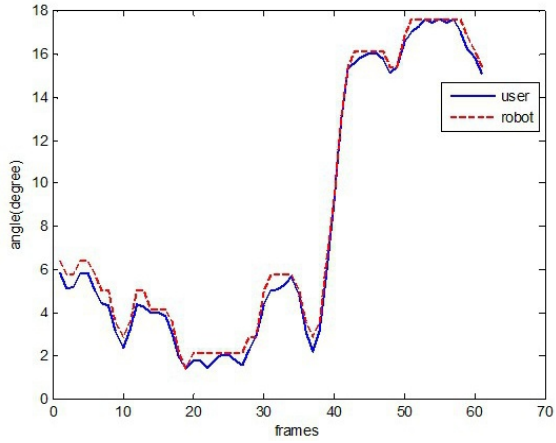
(شکل-۲۲): مقایسه تغییرات زاویه پایه برای کاربر و ربات

(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

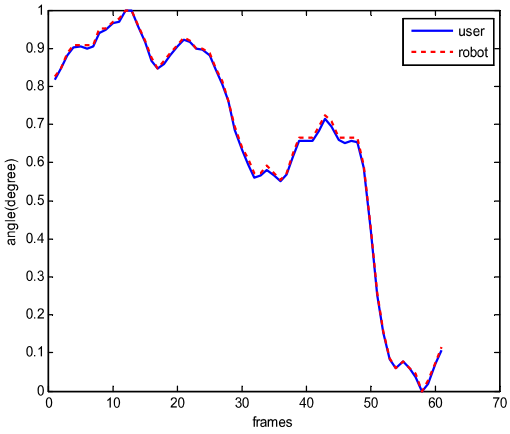


(شکل-۲۴): مقایسه تغییرات زاویه آرنج برای کاربر و ربات

(Figure-24): changes of elbow angle for user and robot

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]

[DOI: 10.29252/jsdp.14.4.97]



(شکل-۲۵): مقایسه تغییرات زاویه مجری نهایی برای کاربر و ربات

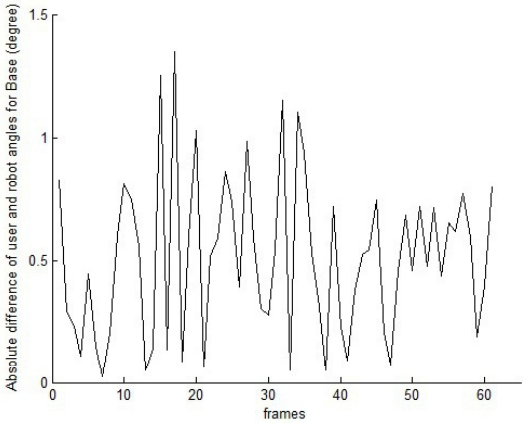
(Figure-25): changes of end effector angle for user and robot

نتایج عددی در شکل‌های (۲۶) تا (۲۹) نشان می دهند که سامانهٔ پیشنهادی، نتیجهٔ قابل قبولی را برای تقلید حرکت در بازهٔ تغییرات نشان‌داده‌شده برای پارامترها در جدول (۱) ارائه می‌دهد.

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

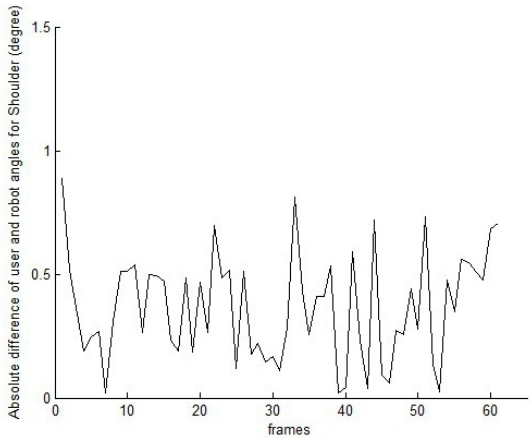
(Table-3): average and variance difference user and robot angles

	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



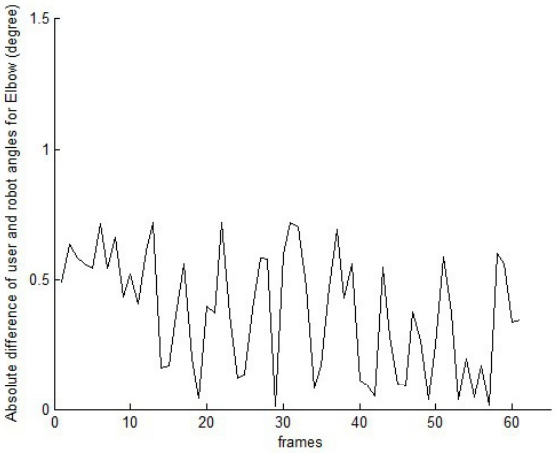
(شکل-۲۶): اختلاف زاویه پایه برای کاربر و ربات در طول فریم‌ها

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



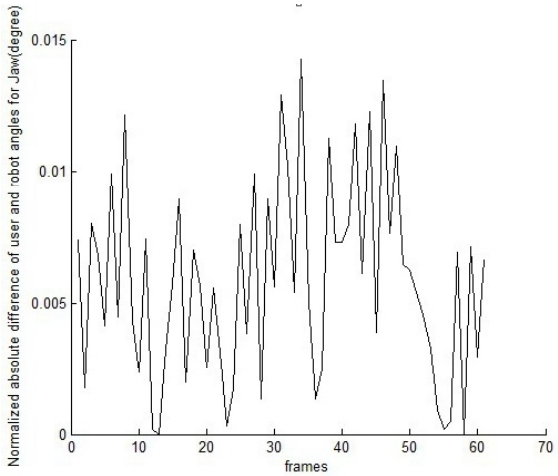
(شکل-۲۷): اختلاف زاویه شانه برای کاربر و ربات در طول فریم‌ها

(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

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]

[DOI: 10.29252/jsdp.14.4.97]

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 دنبالهٔ تصاویر عمق، رنگی و داده‌های اسکلت به‌دست‌آمده از کینکت به‌عنوان ورودی استفاده شد. مختصات سه‌بعدی مفاصل شانه، آرنج و مچ از روی تصاویر عمق و داده‌های اسکلت به‌دست‌آمده از کینکت به‌ازای هر فریم به‌دست آمدند. با استفاده از روش‌های پردازش تصویر مثل آستانه‌گیری، لبه‌یابی، ساخت پوسته محدب، مدل‌کردن رنگ پوست در فضای YCbCr و تفریق پس‌زمینه، نوک انگشتان اشاره و شست که به‌طور 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 محاسبه‌شده به‌روزرسانی 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، به‌طور خلاصه شاخص‌های مورد توجه در این مطالعه به‌شرح 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 دهانهٔ ربات، مختصات نوک انگشتان شست و اشاره مورد نیاز است. این اطلاعات به‌طور مستقیم با استفاده از کینکت به‌دست نمی‌آیند؛ ازاین‌رو 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.
- به‌منظور ارزیابی عملیاتی چنین سامانه‌ای، تقلید عملکرد بازوی کاربر بر روی یک بازوی ربات واقعی RoboTEK II ST240 صورت گرفته است.

۶- مراجع
6- References

[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

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]

[DOI: 10.29252/jsdp.14.4.97]

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 رشتهٔ مهندسی کامپیوتر (گرایش نرم‌افزار) و کارشناسی ارشد را در همان رشته (گرایش هوش مصنوعی و رباتیک) در سال‌های ۱۳۹۴-۱۳۹۲ از دانشگاه یزد گرفت. زمینهٔ مورد علاقهٔ ایشان شامل پردازش تصویر، رمزنگاری تصویروشناسایی‌الگواست.

نشانی رایانامهٔ ایشان عبارت است از:

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



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

نشانی رایانامهٔ 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 Syst-ems, 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. Makrogiannis, 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

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-06]

[DOI: 10.29252/jsdp.14.4.97]