Multiple Person Tracking based on Gait Identification using Kinect and OpenPose

Ryotaro Toma[1], Terumi Yaguchi[1], and Hiroaki Kikuchi[1]

School of Interdisciplinary Mathematical Sciences, Meiji University
4-21-1 Nakano, Tokyo 164-8525, Japan
kikn@meiji.ac.jp

Abstract. A gait provides the characteristics of a person's walking style and hence is classified as personal identifiable information. There have been several studies for personal identification using gait, including works using hardware such as depth sensors and studies using silhouette image sequences of gait. However, these methods were designed specialized for tracking a single walking person and the accuracy reduction when multiple people are simultaneously reflected in several angles of view is not clear yet. In addition, dependencies on hardware-based methods is not clarified yet. In this study, we focus on Kinect and OpenPose, the representative gait identification techniques with a function to detect multiple people simultaneously in real time. We investigate how many people can be identified for these devices and with the accuracy for tracking.

1 Introduction

Multiple human tracking refers to the task of simultaneously detecting and tracking multiple individuals in a given scene or video. The objective is to accurately locate and follow each person's movement throughout the sequence of frames or time. The goal of multiple human tracking is to provide a comprehensive understanding of the activities and interactions of multiple people in various applications, such as surveillance [1], crowd analysis [2], behavior understanding, human-computer interaction, and augmented reality.

The process of multiple human tracking typically involves several steps. First, individual humans need to be detected or localized in each frame, often using computer vision techniques such as object detection [3][4], face detection [5][7][34] or pose estimation [8][35]. Next, these detection or pose estimates are linked across frames to establish trajectories, ensuring consistent and accurate tracking over time. Various methods and algorithms are employed for data association and tracking. Muaaz et al.[31] proposed a person identification method using a smartphone-based accelerator. They used the acceleration information of an Android device in a person's front pocket as data. Preis et al. proposed a gait recognition method using Kinect [9]. They used a decision tree and a Naive Bayes classifier to recognize the gait. Han et al. [10] proposed the gait energy image (GEI). The advantages of GEI are the reduction of processing time, reduction of storage requirements, and robustness of obstacles. Backchy et al. [11] proposed a gait authentication method using Kohonen's self-organizing mapping (K-SOM). In this work, the authors used K-SOM to classify GEI and reported a 57%

recognition rate. Shiraga et al. proposed the GEINet [12] using a convolutional neural network to classify GEI images. The best EER obtained was 0.01.

In this study, we consider two approaches: the computer vision approach using OpenPose [20] and the depth sensor approach such as Kinect [14]. Computer vision approaches can be applied to a variety of applications and provide rich information including joint positions without any additional sensors or equipment, relying solely on visual data captured by cameras. However, computer vision methods primarily rely on 2D image data, which may lack depth information required for precise depth-related analysis. In contrast, depth sensors, like Kinect, provide depth data, enabling more accurate and detailed 3D tracking of human movements and positions. But, depth sensors often have limited tracking ranges, which may restrict their applicability to certain scenarios. The pros and cons are can vary depending on specific applications and implementations for tracking. Hence, it is not trivial to determine which is superior than others.

To evaluate the effectiveness of these two approaches, we employ the Dynamic Time Warping (DTW) [15] algorithm for individual identification. The DTW algorithm leverages the 3-dimensional coordinates obtained from either approach to recognize individuals. By calculating the DTW distance of time series data representing a complete walking cycle, it enables multiple human tracking, accommodating crowded scenes, and addressing environmental variations. Our objective is to utilize DTW for reliable and precise human tracking, facilitating a comprehensive understanding of behaviors and enabling in-depth analysis of human movements in urban environments.

In this study, we develop a testbed for human tracking, implementing two representative approaches: OpenPose and Kinect. We conduct small-scale experiments to assess the robustness and accuracy of estimation provided by these approaches in the context of multiple human tracking.

2 Preliminary

2.1 OpenPose

OpenPose [20] is an open-source computer vision library that enables real-time multi-person tracking from video and image data. OpenPose is capable of estimating the 25 positions of body joints, such as the shoulders, elbows, wrists, hips, knees, and ankles, for multiple individuals in a frame.

OpenPose utilizes convolutional neural networks (CNNs) to analyze visual data and extract keypoint information. The library employs a two-step process: first, it generates a set of body part candidates through a body part detector, and then it associates these candidates to corresponding body parts and individuals through a series of refinement stages.

Fig. 1 shows the example of keypoint detection in OpenPose. It offers several advantages for person tracking: First, tracking multiple individuals simultaneously in real-time. It is a significant advantage for application that captures the movement of guest in shopping malls. Second, it detect human from images without any special equipment such as 3D depth sensors.

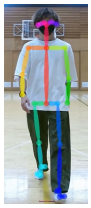


Fig. 1: Sample OpenPose execution

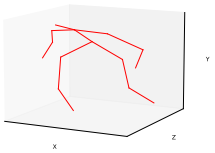


Fig. 2: Sample 3d-pose-baseline execution

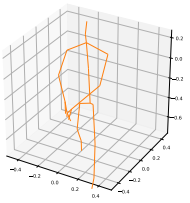


Fig. 3: Sample 3D skeleton data via Kinect

Despite its advantages, it has some limitations: (i) Limited to 2D Keypoint Estimation: OpenPose focuses on 2D keypoint estimation, meaning it provides positional information for body joints in the image plane. It lacks depth information and hence suffers low estimation accuracy. Its performance can be affected by image quality and variations in camera viewpoints. (ii) Resource Intensive: Real-time multi-person tracking with OpenPose can be computationally demanding. It requires substantial computational resources, including a powerful GPU, to achieve real-time performance.

To estimate 3D points from 2D keypoint estimated from OpenPose, Julieta et al. [29] proposed an effective model and developed open-source software, 3D-Pose-Baseline. It is based on an assumption that a 3D pose can be represented as a linear combination of a set of 3D basis points. Fig. 2 shows the sample of estimated 3D points of human. It demonstrates that some joints are accurately estimated.

2.2 Kinect

Kinect [14] is a motion-sensing input device developed by Microsoft for use with gaming. It was initially released as an accessory for the Xbox 360 gaming console in 2010. The Kinect sensor combines a depth camera, RGB camera, and multi-array microphone to provide a range of interactive and immersive experiences.

Kinect utilizes a structured light or time-of-flight technology to capture depth information of the surrounding environment. It measures the distance between the sensor and objects, enabling 3D depth perception. Kinect has built-in algorithms and software libraries for robust human body tracking and gesture recognition. It can detect and track the movements of multiple individuals within its field of view, allowing for natural and intuitive interaction in gaming, fitness, and person tracking. Fig. 3 shows the sample of 3D skeleton data captured via Kinect. We utilize Microsoft library Kinect for Windows v2 for retrieving the 3D points for this study.

Tracking more than six individuals becomes challenging due to the complexity of processing the depth data, distinguishing individual bodies, and maintaining accurate tracking in real-time. The hardware and computational resources of the Kinect sensor are optimized to handle a limited number of tracked bodies. Although the tracking limitation of six individuals is specific to the Kinect sensor, it apply to other depth-sensing devices or motion-tracking systems.

2.3 Related works

There are many approaches for multiple person tracking using various devices.

Multi-Camera Tracking Many works utilize multiple cameras to improve tracking accuracy. Each camera provides a different viewpoint, and sophisticated algorithms are used to use data from these cameras to track individuals as they move across different views. This is especially common in surveillance systems. Amosa et al. [16] categorized existing works based on on six crucial facets and summarized 30 state-of-the-art MCT algorithms on common datasets.

Depth Sensors Depth sensors have been widely used for multiple person tracking. They provide accurate depth information, which helps in distinguishing between individuals and handling occlusions more effectively. Preis et al. proposed a gait recognition method using Kinect [9]. They used a decision tree and a Naive Bayes classifier to recognize the gait. In their work, a success rate of 91.0% was achieved for nine subjects. Studies using depth sensors include [17, 18].

Device-based Tracking Each smartphone continuously records accelerometer and gyroscope data, allowing us to capture motion patterns and changes in orientation. Muaaz et al.[31] introduced a multiple person tracking approach utilizing smartphone accelerometer data. Their method focuses on identifying individuals based on the accelerometer readings of an Android device placed in the front pocket of a person. During the registration phase, walking cycles are defined as templates, and multiple templates are enrolled. In the subsequent authentication phase, the system assesses the distances from all registered templates and considers the user as the correct individual if more than half of the templates fall within the predefined threshold.

3 Person Tracking based on Gait

Person tracking becomes feasible by utilizing gait data, which comprises a time-sequence of 3D points representing the primary joints of the human body. In this section, we describe the approach introduced by [18], which incorporates metrics quantifying the Dynamic Time Warping (DTW) [15]. This approach aims to recognize individuals by utilizing 3-dimensional coordinates obtained from motion capture sensors. It involves calculating the DTW distance of the time series data representing one complete cycle of walking. The method encompasses four key steps: cycle extraction, calculation of relative coordinates, computation of DTW distance, and person recognition.

3.1 Cycle Extraction

Let $a_\ell(t) = (x, y, z)$ be a time series of 3-dimensional absolute coordinates of joint ℓ in time t. The collection of these time series data, representing absolute coordinates at different points in time, is referred to as *skeleton data*.

From the skeleton data, we extract a single cycle of walking. In our specific context, each video stream observation typically contains approximately two complete walking cycles.

First, let $\Delta(t)$ be the distance between both feet in time t, defined using $a_{LF}(t)$ and $a_{RF}(t)$ as

$$\Delta(t) = \pm\|a_{RF}(t) - a_{LF}(t)\|. \tag{1}$$

If the right foot is in front, The sign of $\Delta(t)$ determines whether the right foot is positioned in front (positive) or not (negative).

Next, we apply Fourier transformation to the time series $\Delta(1), \ldots, \Delta(n)$ and employ a low pass filter to reduce noise and identify a single cycle. The resulting low-frequency components at a rate of $1/30$ are processed further. For cycle extraction, we define a *cycle* of walking as the period between peaks. It is important to note that the low-pass filter is solely used for cycle extraction purposes, while the DTW algorithm operates on the non-filtered data. In the cycle extraction phase, time t is a unit corresponding to the frame rate of the motion capture sensor. The frame rate is 30 fps. Suppose that we have one cycle as a series of features from the first peak ($t = 37$) to the second peak ($t = 70$). The data is normalized from t_1 to t_{35}.

3.2 DTW Distance

We compute the relative coordinates of joints while walking, with the choice of the coordinate origin being stable joints located at the center of the body (SpineMid). Given an absolute coordinate of center joints c at time t $a_c(t)$, the relative coordinate r is defined as $r_\ell(t) = a_\ell(t) - a_c(t)$.

We use a multi-dimensional Dynamic Time Warping [30], which is a technique used to measure the similarity between two temporal sequences. It is commonly employed in various fields, including time series analysis, speech recognition, gesture recognition, and pattern recognition. The goal of DTW is to find an optimal alignment or warping path between two sequences by stretching or compressing the time axes. This alignment aims to minimize the differences between corresponding elements of the sequences, allowing for comparison and similarity estimation even when the sequences have variations in length or speed.

Dynamic programming is used to find the optimal warping path through the cost matrix, providing distance between elements of two sequences. The algorithm iteratively computes the cumulative cost along different paths and identifies the path with the minimum total cost as follows.

Consider two sets of time series data denoted as $P = (p_1, p_2, \ldots, p_{n_P})$ and $Q = (q_1, q_2, \ldots, q_{n_Q})$. The distance between them, represented by $d(P, Q)$, is defined as $d(P, Q) = f(n_P, n_Q)$. The cost function $f(i, j)$ is calculated recursively as

$$f(i, j) = \|p_i - q_j\| + \min(f(i, j-1), f(i-1, j), f(i-1, j-1),) \tag{2}$$

with initial conditions; $f(0, 0) = 0$, and $f(i, 0) = f(0, j) = \infty$. When several features are aggregated, the distance is calculated as follows. Given two data sets (R_ℓ, R_m) and (R'_ℓ, R'_m), and data of joints ℓ and m, the integrated DTW distance $D((R_\ell, R_m), (R'_\ell, R'_m))$ is defined as an Euclidean distance of all DTW distances.

3.3 Human identification

Consider the set U representing all users. Let $R^{(u)}$ denote the time series data of k pieces of normalized relative coordinates for user u. Given a set of s data pieces ($R^u1, \ldots, R^us$), we define one of them as the template data $R^{(u)}$. Two users, u and v, are considered identical if the integrated DTW distance $D(R^{(u)}, R^{(v)})$ between their respective sets of time series data, $R^{(u)}$ and $R^{(v)}$, is less than a threshold value θ. The threshold θ^ℓ is determined using the Equal Error Rate (EER), which is an error rate at which the False Acceptance Rate (FAR) equals the False Rejection Rate (FRR).

4 Evaluation

4.1 Objectives

Our experiment aims to achieve the following objectives:

1. Evaluate the baseline accuracy of gait tracking using DTW distance for multiple humans.
2. Compare two tracking approaches: the computer vision approach utilizing OpenPose and the deep sensor approach using Kinect.

We have developed a testbed system to capture the 3D time-series data of walking humans. For the essential features, we the Processing V3 with the KinectPV2 library, which provides access to Kinect for Windows V2 [14].

4.2 Data

The experimental conditions are presented in Table 1, outlining the specifications. The data collected includes two types of scenarios: walking by a single individual and walking by multiple individuals. The observations took place in a gymnastic hall, where subjects walked without encountering any obstacles. During the experiment, the subjects were instructed to walk while being recorded from three different camera viewpoints: front camera, as well as cameras positioned obliquely at plus and minus 30 degrees.

We conducted a multiple person tracking evaluation by observing a varying number of walking subjects, denoted by m, ranging from 1 to 6. For each number of subjects, we explored different variations, including scenarios where all subjects walked in the same direction and cases where some subjects walked in a direction opposite to others. For each variation, we repeated the tracking process three times to ensure reliable results.

4.3 Methodology

We conducted an evaluation of two tracking approaches: one using computer vision with OpenPose and the other using a depth sensor with Kinect.

In the initial stage, we examined the quality of 3D skeleton data obtained from both methods. We observed that the accuracy of 3D estimation using OpenPose might be affected by the camera viewpoints due to the absence of depth information. On the other

Table 1: Experimental condition

item	value
date	July 16, 2022
venue	gymnastic hall, Meiji University
age	20's
population	7 (4 male, 3 female)

hand, the depth sensor method with Kinect had limitations in tracking the number of humans due to the complexity involved in processing depth data. Consequently, the 3D points estimated from these devices may contain erroneous data. To evaluate this, we analyzed the occurrence of detection failures by manually classifying three randomly sampled frames from three camera viewpoints into three categories: (a) *normal*, (b) *partially malfunctioning*, and (c) *completely malfunctioning* frames. Fig. 4 provides examples of these frames, including normal data with accurately estimated 3D points (Fig. 4a) , partially malfunctioning frames with failed detection of a specific joint (e.g., left knee) (Fig. 4b), and completely malfunctioning frames (Fig. 4c) where most of the points are incorrect.

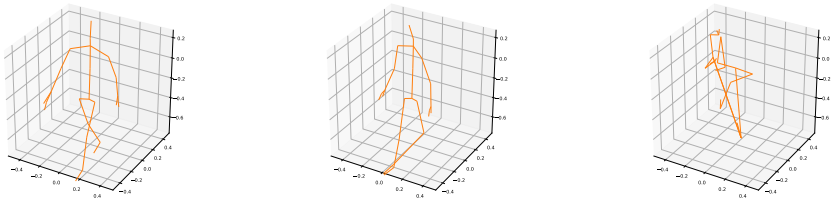
For the analysis involving a single walking subject, we investigated a total of n subjects captured from three viewpoints across three randomly chosen frames, resulting in a total of 63 frames ($7\times3\times3$). In the case of multiple subjects, we examined n subjects based on two randomly selected frames.

In the next stage, we applied the DTW human identification algorithm using 3D time-series data as described in Section 3. By varying the number of subjects m from 1 to 6, we tested the accuracy of identification using the 3D point data obtained from Kinect and OpenPose. The accuracy of identification was measured as the fraction of correctly identified subjects. Additionally, we evaluated the *top-k accuracy*, which is a common performance metric used in classification tasks. The top-k accuracy measures the proportion of correct identifications where the correct person label is among the top-k predicted labels. In other words, if the true class label is among the k highest-ranked persons based on the DTW distance to the given data, the tracking is considered correct. We calculated the top-k accuracy for values of k ranging from 1 to 5.

4.4 Results

Quality of detection Table 2 shows the successful tracking rate using Kinect. It is evident that the rate of successful tracking decreases significantly when the camera viewpoint is not frontal. When the camera is positioned obliquely, 33% of the frames exhibit partial malfunctions, while 19% of the frames are not utilized at all.

Table 3 shows the successful tracking rate in relation to the number of walking individuals simultaneously. The rate of success decreases as the number of walking persons m increases. Specifically, at $m = 4$, the success rate is 0.36, which is half of the rate observed at $m = 3$ (0.64). It should be noted that the maximum number of individuals that Kinect can track is specified as 6. The findings reveal that the quality of



(a) Normal

(b) partially malfunctioning frames

(c) completely malfunctioning frames

Fig. 4: Detection failures

Table 2: successful tracking rate with regard to orientation (Kinect)

orientation	success	malfunctioning	
		partially	totally
front	1.0 (21 / 21)	0.0 (0 / 21)	0.0 (0 / 21)
side	0.48 (20 / 42)	0.33 (14 / 42)	0.19 (8 / 42)
total	0.65 (41 / 63)	0.22 (14 / 63)	0.13 (8 / 63)

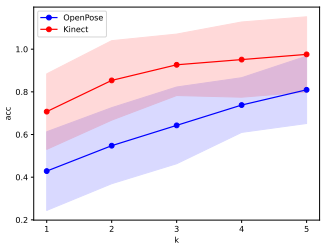
3D point detection diminishes even before reaching the specified tracking limitation of 6 individuals.

As a result, the accuracy of 3D point estimations is negatively affected when the camera viewpoint is not frontal or when there are multiple individuals walking in different directions. The accuracy of person tracking, therefore, depends on these factors, including the orientation of walking and the number of individuals being tracked.

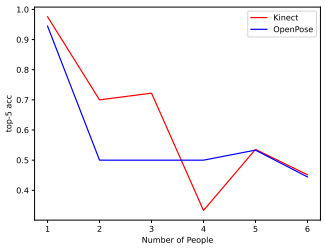
Multiple Person Tracking Fig. 5a shows the tracking accuracy based on DTW, for different top-k values ranging from 1 to 5. Confidence intervals are included for both tracking approaches, Kinect and OpenPose. It can be observed that as the top-k value

Table 3: successful tracking rate with regard to population

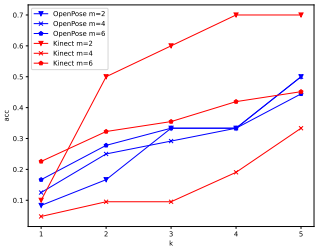
population m	success	failure	
		partially	totally
1	0.65 (41 / 63)	0.22 (14 / 63)	0.13 (8 / 63)
2	0.67 (8 / 12)	0.17 (2 / 12)	0.17 (2 / 12)
3	0.64 (9 / 14)	0.29 (4 / 14)	0.071 (1 / 14)
4	0.36 (5 / 14)	0.43 (6 / 14)	0.21 (3 / 14)
5	0.29 (4 / 14)	0.50 (7 / 14)	0.21 (3 / 14)
6	0.43 (6 / 14)	0.29 (4 / 14)	0.29 (4 / 14)



(a) single ($m = 1$)



(b) population m (top-5)



(c) top k

Fig. 5: Person tracking accuracy

increases, the accuracy for both approaches improve. Overall, the accuracy of Kinect is superior to that of OpenPose, with a difference ranging from 0.3 to 0.2.

Fig. 5b illustrates the distribution of top-5 accuracy while varying the number of individuals m from 1 to 6. Similar to the results obtained for 3D point quality, the accuracy of person tracking substantially decreases with an increasing number of individuals. When comparing the accuracies of OpenPose and Kinect, the reduction in accuracy with Kinect is more strongly influenced by the number of individuals. The accuracy of Kinect appears to be more unstable with respect to m. Therefore, we concluded that OpenPose demonstrates greater robustness in handling different numbers of individuals, although the overall accuracy is relatively lower.

Fig. 5c shows the accuracy plot for the number of individuals, specifically for $m = 2$, $m = 4$, and $m = 6$. The differences in accuracy are evident when multiple people are simultaneously tracked using Kinect. In particular, the accuracy for $m = 4$ is consistently the lowest across all top-k values. This finding provides evidence of the robustness of OpenPose in comparison to Kinect.

4.5 Discussion

Kinect's sensing error Through our observations, we noticed a significant decline in tracking accuracy when using Kinect as the number of individuals increased. In Kinect, the estimation of joint locations relies on depth information captured by the sensor.

However, when obstructed viewpoints are encountered, the accuracy of depth measurement can be compromised, resulting in missing joints. This inherent limitation leads to difficulties in tracking multiple individuals using Kinect.

Based on the findings from our experiments, we assert that the maximum number of individuals that can be reliably tracked using Kinect is 3, which is less than the specified limitation according to Kinect's specifications.

Robustness of multiple person tracking Our findings indicate that the computer vision approach, specifically OpenPose, exhibits greater robustness when considering the number of individuals being tracked. The experiment demonstrates that the accuracy of OpenPose surpasses that of Kinect, particularly when dealing with a larger number of individuals. This observation aligns with the fact that the computer vision approach possesses a higher tracking capacity, as specified by its capabilities.

However, it is important to acknowledge the limitations of our experiment. These include the limited number of subjects involved, variations in environmental conditions such as brightness, and the impact of obstacles on sensing accuracy. Additionally, it should be noted that the performance of a specific device, Kinect, cannot be generalized to other depth sensors, as different sensors may have varying characteristics and performance.

Privacy Concerns Privacy regulations like the GDPR [32] and the CCPA [33] strictly forbid the collection of personal information without explicit individual consent. Gait information is categorized as a form of personal data. Consequently, employing multiple person tracking methods based on gait information raises concerns regarding privacy regulation compliance. To harness the insights derived from tracking individuals, it becomes imperative to adopt privacy-enhancing technologies, including techniques such as data anonymization and obfuscation.

An illustrative approach, "VideoDP" proposed by Wang et al. [19], offers a potential method for identifying individuals within video data by incorporating noise into statistical data, ensuring the application of differential privacy principles. However, there is a pressing need for specialized privacy-enhanced technologies tailored to the unique characteristics of gait information.

5 Conclusions

In this study, we have examined the performance of multiple human tracking using two approaches: OpenPose and Kinect. By employing the DTW distance metric, we have demonstrated the feasibility of tracking multiple humans based on time-series 3D point data. Our experimental results indicate that the depth sensor, Kinect, is capable of accurately tracking multiple individuals. However, its accuracy diminishes when there are more than three individuals walking simultaneously or when their walking orientations differ. Consequently, we conclude that person tracking is influenced by factors such as the orientation of walking and the number of individuals being tracked.

As a future area of investigation, we plan to explore algorithms for large-scale human tracking. Additionally, we are interested in addressing privacy concerns arising from this type of tracking.

Acknowledgment. Part of this work was supported by JSPS KAKENHI Grant Number 23K11110 and JST, CREST Grant Number JPMJCR21M1, Japan.

References

1. D. Kim and S. Park, "A Study on Face Masking Scheme in Video Surveillance System", Tenth International Conference on Ubiquitous and Future Networks (ICUFN 2018), pp. 871-873, 2018.
2. CreÅču, AM., Monti, F., Marrone, S. et al., "Interaction data are identifiable even across long periods of time", Nat Commun 13, 313, 2022 (https://doi.org/10.1038/s41467-021-27714-6).
3. Viola, P. and Jones, M.: Rapid object detection using a boosted cascade of simple features, Proc. Computer vision and Pattern Recognition 2001 (CVPR 2001), pp.I-511-I-518, 2001.
4. Shakhnarovich, G., Viola, P. and Moghaddam,B.: A unified learning framework for real time face detection and classification, Proc. Automatic Face and Gesture Recognition 2002 (FG2002), 2002.
5. Viola, P. and Jones, M.: Robust Real-Time Face Detection, International Journal of Computer Vision(IJCV), Vol.57, No.2, pp. 134-157, 2004.
6. Karen Simonyan and Andrew Zissserman(2014), "Very Deep Convolutional Networks for Large-Scale Image Recognition", pp.1409-1556, ICLR, 2014.
7. Jie Wang and Zihao Li, "Research on Face Recognition Based on CNN", IOP Conference Series: Earth and Environmental Science, Volume 170, Issue 3, 2018.
8. Kazemi, V., Sullivan, J., "One Millisecond Face Alignment with an Ensemble of Regression Trees", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1867-1874, 2014.
9. Preis, J., Kessel, M., Werner, M., and Linnhoff-Popien, C. "Gait recognition with Kinect", Proceedings of the First Workshop on Kinect in Pervasive Computing, 2012.
10. Han, J. and Bhanu, B., "Individual recognition using gait energy image", *IEEE Trans., Pattern Anal. Mach. Intell.*, 28(2), pp. 316âĂŞ322, 2006.
11. S. C. Bakchy, M. R. Islam and A. Sayeed, "Human identification on the basis of gait analysis using Kohonen self-organizing mapping technique," 2nd International Conference on Electrical, Computer and Telecommunication Engineering (ICECTE), pp. 1-4, 2016.
12. Shiraga, K., Makihara, Y., Muramatsu, D., Echigo, T., and Yagi, Y. "Geinet: View-invariant gait recognition using a convolutional neural network", 2016 International Conference on Biometrics (ICB), pp. 1âĂŞ8, 2016.
13. B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "Openface: A general- purpose face recognition library with mobile applications", Technical report, CMU School of Computer Science, CMU-CS-16-118, 2016.
14. MicroSoft, "Kinect v2 library for Processing", (https://github.com/ThomasLengeling/KinectPV2, 2016).
15. Berndt, D. J. and Clifford, J., "Using dynamic time warping to find patterns in time series", The Third International Conference on Knowledge Discovery and Data Mining, pp. 359-370, 1994.

16. Temitope Ibrahim Amosa, Patrick Sebastian, Lila Iznita Izhar, Oladimeji Ibrahim, Lukman Shehu Ayinla, Abdulrahman Abdullah Bahashwan, Abubakar Bala, Yau Alhaji Samaila, "Multi-camera multi-object tracking: A review of current trends and future advances," Neurocomputing, Volume 552, 2023.
17. Mori, T. and Kikuchi, H. "Person tracking based on gait features from depth sensors", The 21st International Conference on Network-Based Information Systems (NBiS-2018), 22, pp. 743-751, 2018.
18. Mori, T., and Kikuchi, H., "Robust person identification based on DTW distance of multiple-joint gait pattern", In P. Mori, S. Furnell, and O. Camp (Eds.), ICISSP 2019 - Proceedings of the 5th International Conference on Information Systems Security and Privacy, pp. 221-229, 2019.
19. Han Wang, Shangyu Xie, Yuan Hong, "VideoDP: A Flexible Platform for Video Analytics with Differential Privacy", Proc. Priv. Enhancing Technol. 2020(4): 277-296, 2020.
20. Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields", IEEE Trans. Pattern Anal. Mach. Intelligence, 43, 1, 172-186, 2021.
21. A. Espitia-Contreras, P. Sanchez-Caiman and A. Uribe-Quevedo, "Development of a Kinect-based anthropometric measurement application", 2014 IEEE Virtual Reality (VR), 2014, pp. 71-72, DOI: 10.1109/VR.2014.6802056.
22. Marijke M. Booij, Martijn S. van Noorden, Irene M. van Vliet, Nathaly Rius Ottenheim, Nic J.A. van der Wee, Albert M. Van Hemert, Erik J. Giltay, "Dynamic time warp analysis of individual symptom trajectories in depressed patients treated with electroconvulsive therapy", Journal of Affective Disorders, Volume 293, 2021, pp. 435-443, 2021. https://doi.org/10.1016/j.jad.2021.06.068.
23. Jingren Tang, Hong Cheng, Yang Zhao, Hongliang Guo, "Structured dynamic time warping for continuous hand trajectory gesture recognition", Pattern Recognition, Volume 80,2018, pp. 21-31, 2018. https://doi.org/10.1016/j.patcog.2018.02.011.
24. R. MartÄśnez-Felez, R. Alberto-Mollineda and J. Salvador-Sanchez. "Gender Classification from Pose-Based GEIs", Proceedings of the Computer Vision and Graphics (ICCVG), pp.501-508, 2012.
25. P. Barra, C. Bisogni, M. Nappi, D. Freire-ObregÃşn and M. CastrillÃşn-Santana, "Gender classification on 2D human skeleton", 2019 3rd International Conference on Bioengineering for Smart Technologies (BioSMART), pp. 1-4, 2019, doi: 10.1109/BIOSMART.2019.8734198.
26. Bewes J, Low A, Morphett A, Pate FD, Henneberg M., "Artificial intelligence for sex determination of skeletal remains: Application of a deep learning artificial neural network to human skulls", J Forensic Leg Med., 62, pp. 40-43, 2019.
27. Hukkelas, H., Mester, R., Lindseth, F., "DeepPrivacy: A Generative Adversarial Network for Face Anonymization", In: , et al. Advances in Visual Computing. ISVC 2019. Lecture Notes in Computer Science(), vol 11844. Springer, Cham., 2019.
28. Jun Liu, Amir Shahroudy, Mauricio Perez, Gang Wang, Ling-Yu Duan, Alex C. Kot, "NTU RGB+D 120: A Large-Scale Benchmark for 3D Human Activity Understanding", IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2019.
29. Julieta, Rayat, Javier, James, "A simple yet effective baseline for 3d human pose estimation", ICCV, pp. 2640-2649, 2017.
30. Ten Holt, G. A., Reinders, M. J. T., and Hendriks, E. A., "Multi-dimensional dynamic time warping for gesture recognition", Thirteenth annual conference of the Advanced School for Computing and Imaging, 2007.
31. Muaaz, M. and Mayrhofer, R.: "Smartphone-Based Gait Recognition: From Authentication to Imitation," IEEE Trans. Mobile Computing, Vol.16, No.11, pp.3209-3221, 2017.

32. European Parliament, Council of the European Union, "General Data Protection Regulation", 2016 (accessed from https://eur-lex.europa.eu/eli/reg/2016/679/oj).
33. "California Consumer Privacy Act of 2018", Civil Code, Division 3, Part 4, Title 1.81.5, 2020. (accessed from https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?lawCode=CIV&division=3.&title=1.81.5.&part=4.&chapter=&article=)
34. R. Mishra, "Persuasive Boundary Point Based Face Detection Using Normalized Edge Detection in Regular Expression Face Morphing," 2023 International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), pp. 1-4, Ballar, India, 2023.
35. S. W. Wong, Y. C. Chiu and C. Y. Tsai, "A Real-time Affordance-based Object Pose Estimation Approach for Robotic Grasp Pose Estimation," 2023 International Conference on System Science and Engineering (ICSSE), Ho Chi Minh, Vietnam, pp. 614-619, 2023.