Tech Science Press

Marker-Based and Marker-Less Motion Capturing Video Data: Person and Activity Identification Comparison Based on Machine Learning Approaches

Syeda Binish Zahra[1,2], Muhammad Adnan Khan[2,*], Sagheer Abbas[1], Khalid Masood Khan[2], Mohammed A. Al-Ghamdi[3] and Sultan H. Almotiri[3]

[1]School of Computer Science, National College for Business Administration and Economics, Lahore, 54000, Pakistan
[2]Department of Computer Science, Lahore Garrison University, Lahore, 54792, Pakistan
[3]Computer Science Department, Umm Al-Qura University, Makkah City, Saudi Arabia
*Corresponding Author: Muhammad Adnan Khan. Email: madnankhan@lgu.edu.pk
Received: 12 July 2020; Accepted: 28 August 2020

Abstract: Biomechanics is the study of physiological properties of data and the measurement of human behavior. In normal conditions, behavioural properties in stable form are created using various inputs of subconscious/conscious human activities such as speech style, body movements in walking patterns, writing style and voice tunes. One cannot perform any change in these inputs that make results reliable and increase the accuracy. The aim of our study is to perform a comparative analysis between the marker-based motion capturing system (MBMCS) and the marker-less motion capturing system (MLMCS) using the lower body joint angles of human gait patterns. In both the MLMCS and MBMCS, we collected trajectories of all the participants and performed joint angle computation to identify a person and recognize an activity (walk and running). Using five state of the art machine learning algorithms, we obtained 44.6% and 64.3% accuracy in person identification using MBMCS and MLMCS respectively with an ensemble algorithm (two angles as features). In the second set of experiments, we used six machine learning algorithms to obtain 65.9% accuracy with the k-nearest neighbor (KNN) algorithm (two angles as features) and 74.6% accuracy with an ensemble algorithm. Also, by increasing features (6 angles), we obtained higher accuracy of 99.3% in MBMCS for person recognition and 98.1% accuracy in MBMCS for activity recognition using the KNN algorithm. MBMCS is computationally expensive and if we re-design the model of OpenPose with more body joint points and employ more features, MLMCS (low-cost system) can be an effective approach for video data analysis in a person identification and activity recognition process.

Keywords: Marker-based motion capturing system; marker-less motion capturing system; support vector machine; K-nearest neighbor

1 Introduction

Motion capturing system is used for measuring and recording of body posture variation with time. Human motion constructs 3D representation using the subject's body parameters including subject's

orientation and position. In motion capturing, either we use marker-less or marker-based system. Marker-less system uses computer vision techniques whereas marker-based systems uses sensors or markers (active or passive), attached at a subject's body [1]. We used motion capturing systems to measure and record the human gait (lower body extremity) for analyzing each subject's movement to recognize him and various activities performed by him. Technically, gait analysis is the study of body movements in the forward direction, when one leg is in the air (swing phase) and the second leg (stance phase) supports the process of movement in a forward or backward manner to support the whole-body weight (the whole body weight is shifted with the action of swing and stance phases). Generally, gait is the cycle of swing and stance phases. Multiple applications are used for person gait analysis including animation industry, sports science, person verification and identification [2–5].

1.1 Marker-Based Motion Capturing System (MBMCS)

MBMCS extracts data during sensing and processing stages. In the sensing stage, joint orientation and position's data is collected using markers that are attached with the subject's body and high-speed cameras are used for recording the data. In the processing stage, a 3D representation of joint data is extracted using trajectories. Electro-magnetic, electromechanical and optical systems are based on a marker-based approach for motion analysis. In our research, data is captured using the VICON motion capturing system for the marker-based approach.

1.2 Marker-Less Motion Capturing System (MLMCS)

MLMCS, based on a series of video cameras (including vision-based software), extracts data of joint trajectories without attaching markers or sensors on the subject's body (e.g., visual surveillance). MLMCS perform its functionality in three stages. First stage is the subject's tracking and detection that employs predicted algorithms for tracking and estimating the subject's moving parameters. For tracking predictive methods, particle filtering, mean shift, condensation algorithm and Kalman filter are used. In the second stage, subject's feature extraction is performed using the whole body or body parts configuration to estimate a set of feature measurements. Feature estimated values are used in gait recognition (as measurements of joint angles), gesture-driven techniques (hand's configuration) and silhouette-based feature recognition and detection. The third phase is motion classification using the labeling process. This phase uses a variety of recognition methods for human motion (used in computer vision systems), such as the KNN method, SVM, and neural networks (NN). Gait analysis has become a more extensively used tool in research to access and evaluate human movements and characterizing locomotion disorders. In this research, the Tensorflow tool is used to characterize the human locomotion. Also, to avoid the expensive systems (motion capture system) and eliminate tough laboratory circumstances, data in our research is captured using the OpenPose motion capturing algorithm (MLMCS).

1.3 Contributions

A few major contributions in our research are compiled as follows:

1. We have used a person's lower joint angles for his identification and activity recognition. To the best of our knowledge, no one has ever used these angular gait parameters for a person's identification or activity recognition.

2. Machine learning techniques are separated into unsupervised and supervised learning techniques [6,7]. We have used supervised machine learning algorithms such as ensemble, KNN, tree, linear discriminant (LD), logistic regression (LR), SVM and long short-term memory (LSTM) for a person's identification and activity recognition in a time series data and obtained higher accuracy rates (as mentioned in Tabs. 2–7).

3. In experiments, data gathering and plotting methods show state-of-the-art results using gold standard marker-based and marker-less approaches.

2 Comparison between Marker-Less and Marker-Based Approaches

To compare the performance of both systems, marker-less and marker-based approaches, we have computed gait parameters. In the previous research using gait parameters, Eman et al. [8] have presented a method to exact gait parameters using 2D marker-based approach. Using a "Gaussian Mixture Model" from an image, Labbe et al. [9] have computed gait parameter named "knee angle" (joint function) to evaluate the injury of knee. In our marker-based approach, we have used six angles and applied 16 markers at the lower body. For comparison, we have used only two angles (knee angle and hip angle) as in marker-less approach structure model, only 6 virtual markers are available at the lower body so we have computed only two angles using this approach. The beneficial point of a human body joints is that all points are transformed using "Rigid-Body", and all body joint points are moved using angular displacement [10].

Joint angles estimation is a crucial aspect using joint embedded frames in the marker-based 3D approach [11], while in the marker-less approach, only technical frames are used.

The other issue is about the different frame rate in both approaches. In the marker-based approach, we have to perform calibration to get accurate frames whereas in the marker-less approach, no such calibration is required. The marker-based approach is based on the full lab setup with proper connectivity of infrared cameras, high-speed video cameras, the marker-based software and the server that increases the cost of the whole project. Whereas, marker-less approach doesn't need such an expensive setup and it is a cost-effective approach. A static acquisition for tracking the subject is required by both systems. In both approaches, we get the trajectories of each subject. MATLAB is used for gathering the trajectories of marker-based approach and python is used in marker-less approach.

The surveillance system can provide a more natural and practical way to identify persons in videos. A person's re-identification cannot be performed using spatiotemporal information as surveillance videos lack algorithms or models which can compute spatiotemporal information [12]. For the computation of gait parameters, we have used the gold standard marker-based approach. Using this method, we have obtained all the trajectories at a higher frame rate and used these trajectories [13] for person identification. Wang et. al. [14] and Liu et al. [15], performed person identification using videos but they did not get or learn person features from a video sequence. Another downside of their research is that they have worked at low-level features but in our research, all the features of gait are used.

3 Methodology

There are a few loopholes in the gait recognition research using "Motion Capture" data. No research work covers or focuses on the discriminant power of features and previously, only two or three gait parameters are used for the analysis. In the previous research, video data is collected using ordinary cameras and all data was in two-dimensional (2D) format and only images-based analysis was performed.

In this research, three-dimensional (3D) video data is captured by the latest "Motion Capture System" with the highest frame rate and special cameras (high-speed DV cameras and infrared cameras) are used for data gathering. The main idea is to identify a person using his or her gait. To achieve gait identification, we have extracted parameters from a person's gait. These parameters are analyzed for an activity recognition and a person's identification.

3.1 Data Acquisition

3.1.1 MBMCS

In this research, the process of capturing the subject's movements in 3D is performed using high-speed cameras with the VICON System motion capturing technique. A reflected material is used in passive markers as they reflect infrared waves. During the data acquisition phase, infrared lights are produced by digital cameras (14 in numbers) at a frame rate of 250 frames per second and markers reflect this light.

3.1.2 Experimental Setup for MBMCS

All measurements are performed in biomechanics laboratory on a 3M walkway for capturing video frames as we have used 14 infrared cameras at a sampling rate of 250 Hz. Two high-speed video DVD cameras are also used and installed to record the front and right-side movement at 125 f/s. This walking path provides enough space to a subject for movements in a normal and natural way. 07 subjects were used in these experiments (all are male, M1 to M5). Fig. 1(a) shows the whole VICON motion analysis system. In the pre-processing stage, 16 markers are attached to a subject's lower body parts at both left and right side (hip, knee, ankle, thigh, tibia, heel and toe). Each marker position is labeled with the proper body part name that distinguishes left and right side of the body as shown in Fig. 1(b) and these marker positions have been identified in VICON Nexus. A lower body marker accurate placement protocol was used to enable the identification of lower body parts. The 16 reflective markers were attached at the lower-body using double sided tape. We got "motion trajectories" form a given sequence of video frames and get each position value for each angle in x, y, and z plane as shown in Fig. 1(c). The statistical formulation for dynamic analysis and mathematical calculation for angular analysis could be performed using joint point values while the single value is used in each angular joint kinematics that generates a unique pattern of lower limb motion in each "position" over time. Eq. (1) has the formulation for an angle and Eqs. (2) and (3) compute vector (u and v respectively) points for 3D data. We have computed 6 angles of lower extremity and extracted the following parameters.

1. hip-to-knee angle (ASI-TIB-KNE)
2. tibia-to-tight (TIB-KNE-THI)
3. knee-to ankle (AKN-TIB-KNE)
4. toe-to-tibia (TOE-ANK-TIB)
5. heel-to-tibia (HEE-ANK-TIB)
6. heel-to-toe (HEE-ANK-TOE).

For comparison with the MLMCS we have also compute two angles, knee angle and hip angle.

$$\theta = arccos\left(\frac{u.v}{|u||v|}\right) \tag{1}$$

$$u = \begin{bmatrix} kne_x \\ kne_y \\ kne_z \end{bmatrix} - \begin{bmatrix} thi_x \\ thi_y \\ thi_z \end{bmatrix} \tag{2}$$

$$v = \begin{bmatrix} thi_x \\ thi_y \\ thi_z \end{bmatrix} - \begin{bmatrix} asi_x \\ asi_y \\ asi\_z \end{bmatrix} \tag{3}$$
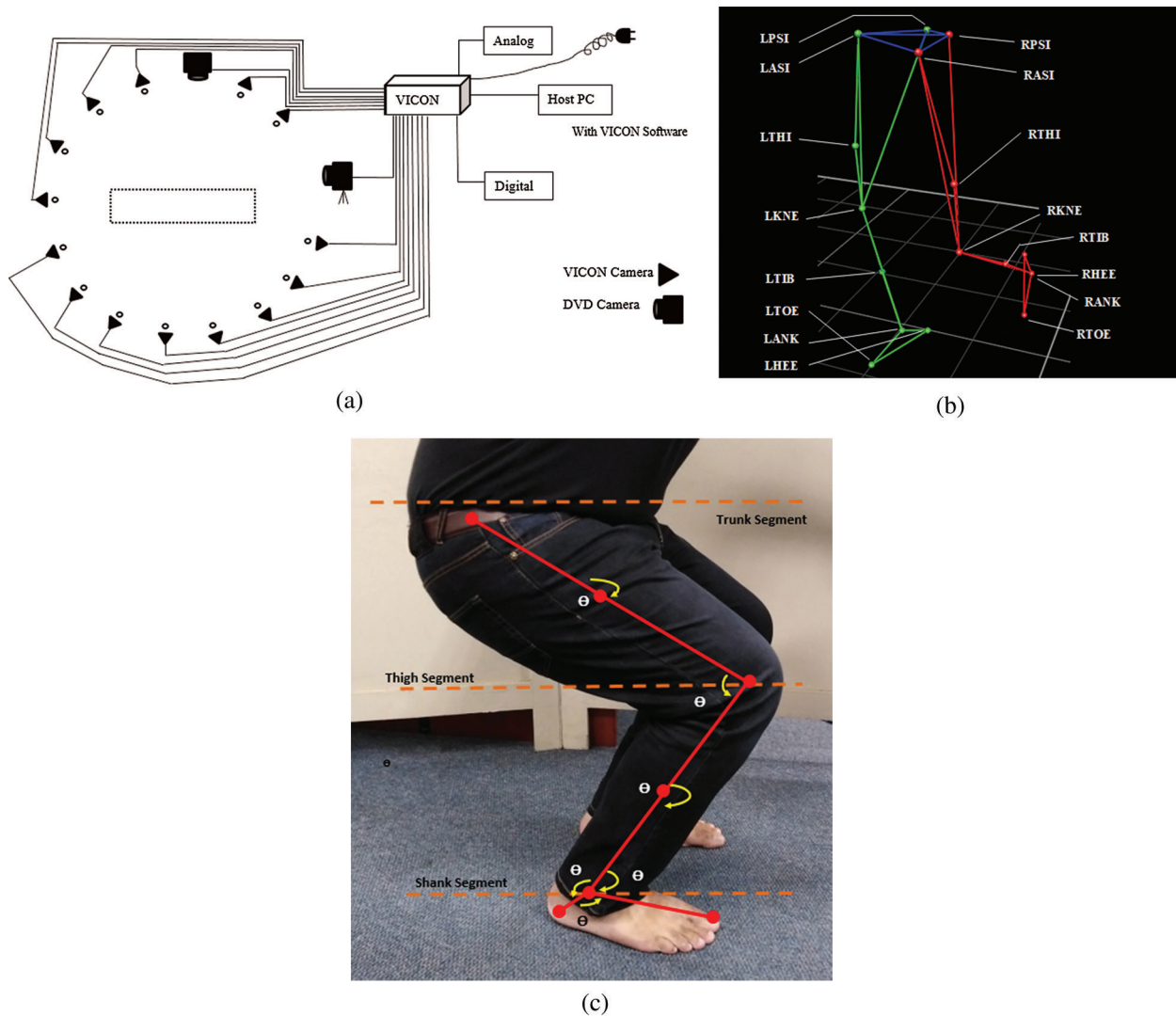
(a)

(b)

(c)

Figure 1: (a) VICON, motion capturing system, (b) 16 reflecting marker names with marker positions, (c) Angular kinematics along 6 angles and their positions

3.1.3 MLMCS

In this research, we have employed OpenPose for capturing a subject's movements. OpenPose is the workable application to detect and provide face and body points such as arms, legs, hips, ankle body joint points in 2D [16,17]. The drawback of OpenPose is that if it fails to detect the body then no source of recovery exists in the live streaming. The second problem is cost as in real-time approach and for multiple person body point detection, the cost of this application is high. The process of OpenPose starts with the detection algorithm that detects body key-points [18]. Then it joins the key-point values and builds a network for each point using interpolation and approximation [17].

3.1.4 Experimental Setup for MLMCS

Open pose performs the key-point detection based on state-of-the-art initial body key-point network architecture. In our research, we have used the part affinity field (PAF) network architecture. The PAF part to part federation is performed iteratively using confidence map detection. Vectors are defined in PAF

as 2D orientation and this vector point is obtained from one body key-point to another key-point as mentioned in Fig. 2. Convolutional network N analyzes the initial image represented as I and generates a feature set FS (in our case angles). In the next step, FS is used as input to the network stage I, which performs the prediction (containing some values) on each set of PAFs (known as Level Value LV). At each level of PAF, previous level value LV^{i-1} is concatenated to FS and the current level LV^I is refined. These values (Eqs. (4)–(6)) are sent into network N, which performs prediction PV Eq. (7) (predicted key-point value) on key-point values. Finally, full-body detection of points for each subject is performed using the "Bipartite Graph Matching" [19].
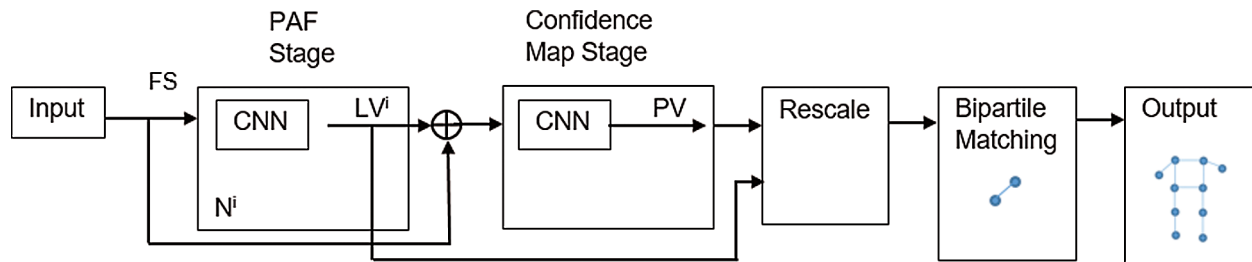

Figure 2: Part affinity field network architecture in OpenPose

$$LV\ (1)\ =\ N\ (1)\ FS \tag{4}$$

$$LV\ (i)\ =\ N\ I\ (FS,\ L\ (i-1)),\ \text{for all}\ 2\ \leq\ i\ \leq\ T \tag{5}$$

$$LV\ =\ LV^{\ T} \tag{6}$$

$$PV\ =\ N\ (FS,\ LV) \tag{7}$$

Different libraries like Alpha-Pose [20] and Mask R-CNN are available for "Pose Estimation" in 2D. By implementing these libraries, it works in their own "Frame Reader" (video streaming or images) and generates direct results in visual forms and key-point values in 2D that results in the output file (in our case it generates .csv files). OpenPose libraries can run on different platforms (embedded systems, MAC, Windows and Ubuntu) whereas in our experiments, we have used Windows and the python platform. Selected video data is served as an input and the output video labeled with body key-points is obtained that contains a .csv file with x and y coordinate values of each frame.

3.2 Participants

To capture gait data, 07 healthy participants (all men, aged 21–26 years, height 5 to 6 feet) were used in our experiments. The summary of participants is shown in Tab. 1. The experiments were performed in closed and control conditions at the biomechanical laboratory in LUMS. Height and body mass were properly measured using a stadiometer and weight scale respectively. Functional assessments are performed using biomechanical gait analysis, (static, clockwise and anticlockwise walking and running) with 14-cameras based VICON systems.

3.3 Tools Used

- VICON® motion capturing system is used for data gathering with the marker-based scheme.
- OpenPose (with TensorFlow) is used for data gathering with the marker-less scheme.
- MATLAB will be used to simulate this work.

Table 1: Participant summary

|	Age	Height	Weight
Subject 1	19	5'11"	50 kg
Subject 2	21	5'11"	73 kg
Subject 3	23	5'8"	70 kg
Subject 4	19	6'0"	56 kg
Subject 5	22	5'7"	70 kg
Subject 6	23	5'11"	82 kg
Subject 7	24	5'8"	78 kg

4 Experimental Results

We have used two motion capturing systems, one is marker-based and other is marker-less. For MBMCS, we have used the VICON system and gathered all the lower body movement data. For MLMCS, we have used video data of the subject's movements. Each subject performs two activities (1) walk and (2) running. For MBMCS, we used MATLAB to get all the movement trajectories in 3D and for MLMCS, we used python to get all the movement trajectories in 2D. We computed two angles (1) knee angle and (2) hip angle and used these angle calculations in-person identification and activity recognition. For person identification, we used five machine learning algorithms named ensemble algorithm, KNN, Decision tree, SVM, and LD analysis. For activity recognition, we have employed six machine learning algorithms named ensembles, KNN, Decision tree, SVM, LD analysis and LR. For person identification, we have achieved 44.6% accuracy in MBMCS with ensemble algorithm as mentioned in Fig. 6 and Tab. 5. For person identification, we have achieved 64.3% accuracy in MLMCS with an ensemble algorithm as mentioned in Fig. 7 and Tab. 6. For activity recognition, 65.9% accuracy level is achieved in MBMCS with the KNN algorithm as mentioned in Fig. 3 and Tab. 2. For activity recognition, 74.6% accuracy level is achieved in MLMCS with an ensemble algorithm as mentioned in Fig. 4 and Tab. 3. Our main result finding shows that MBMCS is costly (as we have mentioned earlier that it needs a lab setup and costly equipment) and performance in terms of accuracy is low. The MLMCS approach generates high accuracy when we compare it with MBMCS, and if we improve the model of OpenPose using more body joint points (by adding more body joint points we can actually increase features) then we can achieve more accuracy using MLMCS which endorses that it is an effective approach in terms of cost and even no lab setup is required.

We have computed six angles to increase the accuracy of our system. It is evident that after increasing the features (6 angles instead of 2 angles) we get higher accuracy 99.3% with KNN algorithm for person identification as shown in Fig. 8 and Tab. 7. The confusion matrix is presented in Fig. 9 and 98.1% accuracy with the KNN algorithm for activity recognition as shown in Fig. 5 and Tab. 4 whereas their confusion matrix is shown in Fig. 10. We got these 6 angles in MBMCS as we increase the massive markers on the human body, and this cannot be performed in MLMCS.

Table 2: Marker based 2 angle: Activity

	Accuracy	Precision	Recall	F1
Ensembles	64.0%	62.7%	**76.3%**	**73.5%**
KNN	**65.9%**	**65.7%**	72.2%	68.7%
Linear Discriminant	61.8%	62.8%	64.9%	63.9%
Tree	64.8%	63.7%	74.9%	68.9%
SVM	65.8%	65.4%	72.5%	68.8%
Logistic Regression	61.8%	62.8%	65.1%	63.9%
Mean	64	63	70.98	67.95
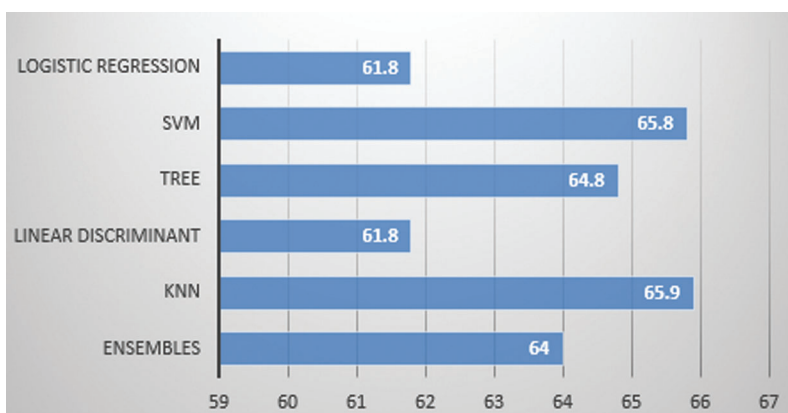
Figure 3: Marker based 2 angle: Activity

Table 3: Marker less 2 angle: Activity

	Accuracy	Precision	Recall	F1
Ensembles	**74.6%**	**74.3%**	76.4%	75.4%
KNN	74.5%	73.1%	**79.2%**	**75.9%**
Linear Discriminant	55.0%	54.1%	76.7%	63.5%
Tree	67.3%	69.6%	63.5%	66.4%
SVM	67.3%	67.7%	68.4%	68.1%
Logistic Regression	54.9%	54.1%	76.6%	63.4%
Mean	60.6	65.53	73.47	68.78
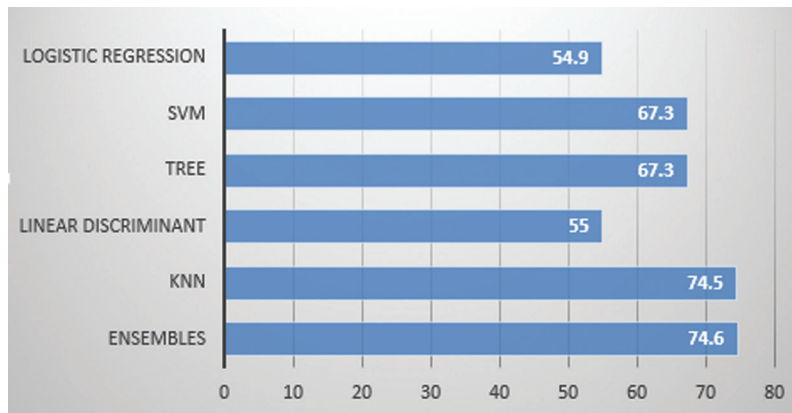
Figure 4: Marker less 2 angle: Activity

Table 4: Marker based 6 angle: Activity

	Accuracy	Precision	Recall	F1
Ensembles	95.4%	**97.7%**	93.7%	95.7%
KNN	**98.1%**	97.6%	**98.6%**	**98.2%**
Linear Discriminant	72.7%	68.8%	86.8%	76.8%
Tree	76.6%	70.6%	94.1%	80.7%
SVM	95.3%	94%	97.2%	95.6%
Logistic Regression	71.0%	68.2%	82.7%	74.8%
Mean	84.85	82.82	92.18	86.97
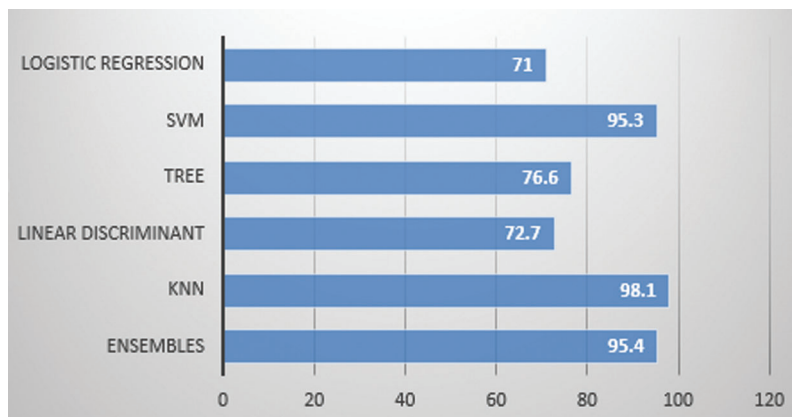
Figure 5: Marker based 6 angle: Activity

Table 5: Marker based 2 angle: Person

	Accuracy	Precision	Recall	F1
Ensembles	**44.6%**	43.9%	**43.7%**	**43.8%**
KNN	43.1%	42.4%	41.5%	41.9%
Linear Discriminant	23.9%	16.1%	21.4%	18.1%
Tree	34.9%	**44.4%**	32.2%	37.3%
SVM	39.0%	40.8%	36.6%	38.6%
Mean	37.1	37.52	35.08	35.94
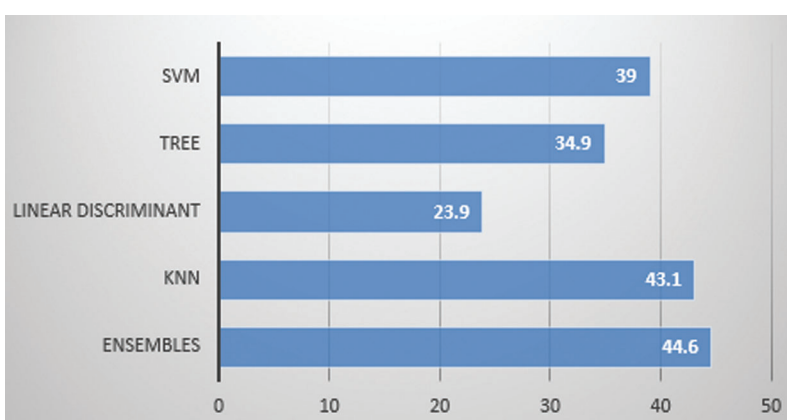
Figure 6: Marker based 2 angle: Person

Table 6: Marker less 2 angle: Person

	Accuracy	Precision	Recall	F1
Ensembles	**64.3%**	**63.4%**	**63.2%**	**63.3%**
KNN	61.1%	60.1%	59.5%	59.8%
Linear Discriminant	23.3%	15.9%	30.2%	20.9%
Tree	39.9%	36.9%	37.7%	37.3%
SVM	48.5%	47.0%	46.3%	46.6%
Mean	47.42	44.66	47.38	45.58
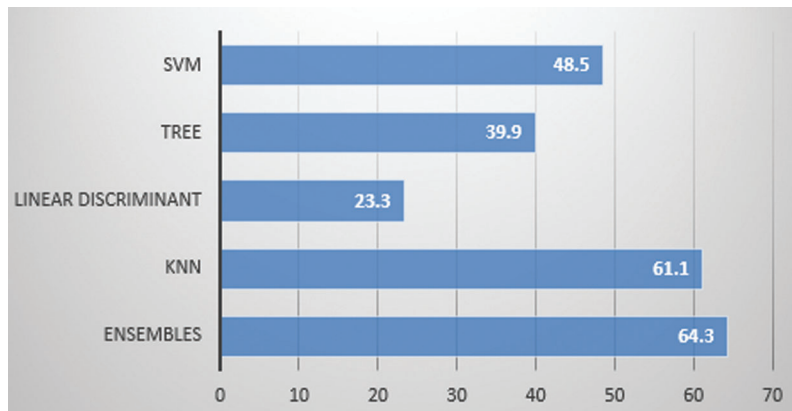
Figure 7: Marker less 2 angle: Person

Table 7: Marker based 6 angle: Person

	Accuracy	Precision	Recall	F1
Ensembles	97.9%	85.4%	97.6%	91.2%
KNN	**99.3%**	**99.2%**	**99.2%**	**99.2%**
Linear Discriminant	79.6%	79.5%	78.5%	79.0%
Tree	85.4%	84.5%	84.5%	84.5%
SVM	98.9%	98.8%	98.8%	98.8%
Mean	92.22	89.48	91.72	90.54
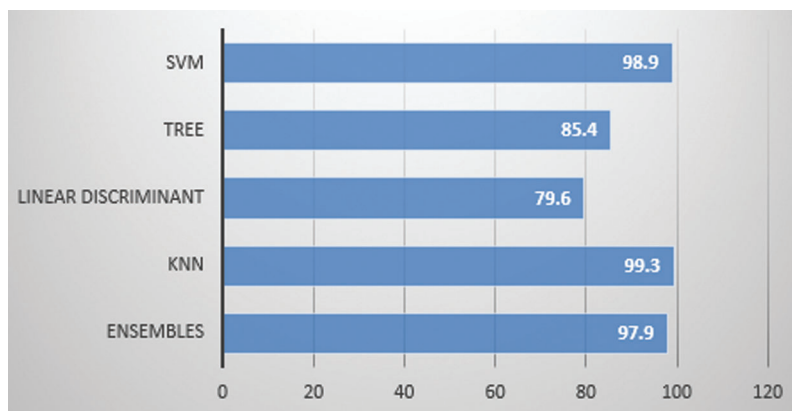
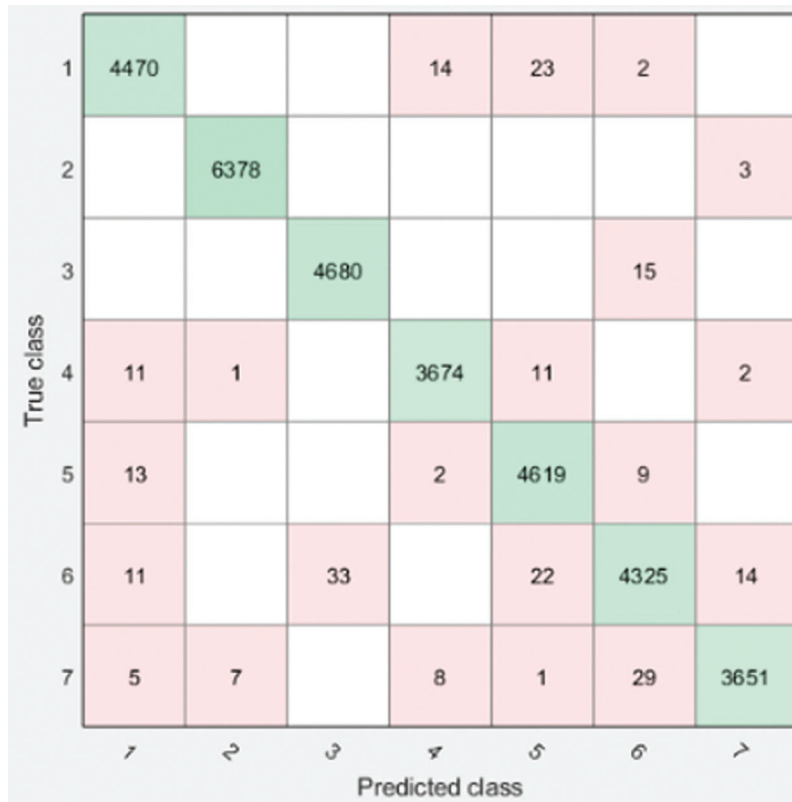**Figure 8:** Marker based 6 angle: Person

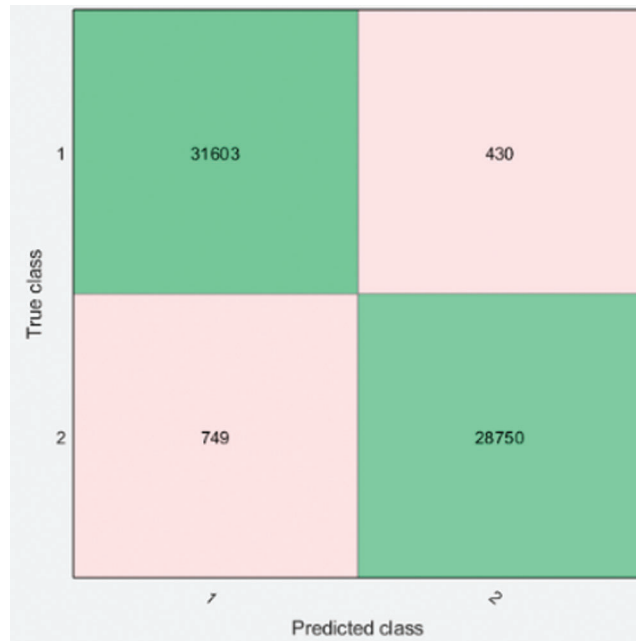**Figure 9:** Confusion matrix marker based 6 angle person

Figure 10: Confusion matrix marker based 6 angle activity

5 Conclusions

Our research has presented a comparison between MLMCS and MBMCS to identify gait patterns. We have computed trajectories for all participants and performed angle computation to identify a person and recognize the activity (walk and running). For person identification, we have employed five algorithms while for activity recognition, we have used six machine learning algorithms. For person identification, we have achieved **44.6%** accuracy in MBMCS and **64.3%** accuracy in MLMCS with an ensemble algorithm. For activity recognition, we **65.9%** accuracy level is achieved in MBMCS using the KNN algorithm and **74.6%** accuracy in MLMCS using the ensemble algorithm. We have also computed six angles to increase the accuracy of our system. It is evident that increasing the number of features (6 angles instead of 2 angles), we can achieve higher accuracy **99.3%** for person identification and **98.1%** for activity recognition using only the KNN algorithm in both approaches. MBMCS is computationally expensive and employs costly lab setup. If we re-design the OpenPose model with more detailed body joint points, thus capturing t more features, then this low-cost system will be an effective approach for analyzing video data in a person's identification and activity recognition problem.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] S. Mohammed, A. Samé, L. Oukhellou, K. Kong, W. Huo *et al.,* "Recognition of gait cycle phases using wearable sensors," *Robotics & Autonomous Systems*, vol. 75, pp. 50–59, 2016.

[2] A. Nandy, S. Bhowmick, P. Chakraborty and G. C. Nandi, "Gait biometrics: An approach to speed invariant human gait analysis for person identification," *Advances in Intelligent Systems and Computing*, vol. 236, pp. 729–737, 2014.

[3] M. Jarolav, "Normative human gait databases," *Statistical Research Letters*, vol. 2, no. 3, pp. 69–74, 2013.

[4] P. Chandra, G. Kanika, M. Anshul, K. Rajesh, L. Vijay *et al.,* "Passive marker based optical system for gait kinematics for lower extremity," in *Proc. Int. Conf. on Advanced Computing Technologies and Applications (ICACTA)*, pp. 176–185, 2015.

[5] A. Frizera, A. Elias, A. J. Del-Ama, R. Ceres, T. F. Bastos *et al.,* "Characterization of spatio-temporal parameters of human gait assisted by a robotic walker," in *4th IEEE Int. Conf. on Biomedical Robotics and Biomechatronics (BioRob)*, pp. 1087–1091, 2012.

[6] O. Simeone, "A very brief introduction to machine learning with applications to communication systems," *IEEE Transitions on Cognitive Communications and Networking*, vol. 4, no. 4, pp. 648–664, 2018.

[7] C. Turkay, R. Laramee and A. Holzinger, "On the challenges and opportunities in visualization for machine learning and knowledge extraction: A research agenda," in *Proc. Int. Cross-Domain Conf. for Machine Learning Knowledge Extraction*, Cham, Switzerland: Springer, pp. 191–198, 2017.

[8] E. F. Mashagba, F. F. Mashagba and M. O. Nassar, "Simple and efficient marker-based approach in human gait analysis using Gaussian mixture model," *Austrian Journal of Basic and Applied Sciences*, vol. 8, no. 1, pp. 137–147, 2014.

[9] D. R. Labbe, A. Fuentes, J. A. de Guise , R. Aissaoui and N. Hagemeister, "Three-dimensional biomechanical assessment of knee ligament ruptures," in *Biomechanics and Biomaterials in Orthopedics*, P. D. London, Ed., Springer, pp. 509–527, 2016.

[10] A. Phinyomark, S. T. Osisa and R. Ferber, "Analysis of big data in running biomechanics: Application of multivariate analysis and machine learning methods," in *Proc. of the 39th Canadian Medical and Biological Engineering Conf.*, Calgary, Canada, pp. 1–4, 2016.

[11] E. Ceseracciu, Z. Sawacha, S. Fantozzi, M. Cortesi, G. Gatta *et al.,* "Markerless analysis of front crawl swimming," *Journal of Biomechanics*, vol. 44, no. 12, pp. 2236–2242, 2011.

[12] B. Prosser, W. S. Zheng, S. Gong, T. Xiang, Q. Mary *et al.,* "Person re-identification by support vector ranking," in *British Machine Vision Conf.*, pp. 21.1–21.11, 2010.

[13] T. Wang, S. Gong, X. Zhu and S. Wang, "Person re-identification by video ranking," in *13th European Conf. on Computer Vision*, Zurich, Switzerland, pp. 688–703, 2014.

[14] K. Liu, B. Ma, W. Zhang and R. Huang, "A spatio-temporal appearance representation for video-based person re-identification," in *Int. Conf. on Computer Vision*, pp. 3810–3818, 2015.

[15] R. Zhao, W. Ouyang and X. Wang, "Person re-identification by salience matching," in *Proc. of IEEE Int. Conf. on Computer Vision*, pp. 2528–2535, 2013.

[16] L. Pishchulin, E. Insafutdinov, S. Tang, B. Andres and M. Andriluka, "Deepcut: Joint subset partition and labeling for multi person pose estimation," in *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 4929–4937, 2016.

[17] S. Johnson and M. Everingham, "Clustered pose and nonlinear appearance models for human pose estimation," *British Machine Vision Conf.*, Leeds, UK, pp. 1–5, 2010.

[18] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

[19] E. Levinkov, J. Uhrig, S. Tang, M. Omran, E. Insafutdinov *et al.,* "Joint graph decomposition & node labeling: Problem, algorithms, applications," *Computer Vision Foundation*, pp. 6012–6020, 2017.

[20] A. Newell, Z. Huang and J. Deng, "Associative embedding: End-to-end learning for joint detection and grouping," in *Advances in Neural Information Processing Systems*, pp. 2277–2287, 2017.