Human gender recognition with upper body gait kinematics

Author:
Tran, H; Pathirana, PN; Seneviratne, A

Publication details:
ACM International Conference Proceeding Series
v. 297
pp. 37 - 41
9781450353823 (ISBN)

Event details:
International Conference on Bioinformatics Research and Applications 2017
Barcelona, Spain
2017-12-08 - 2017-12-10

Publication Date:
2017-12-08

Publisher DOI:
https://doi.org/10.1145/3175587.3175596

Human Gender Recognition with Upper Body Gait Kinematics

Ha Tran
School of Engineering, Deakin University
Waurn Ponds, Victoria, AU
thuha@deakin.edu.au

Pubudu N. Pathirana
School of Engineering, Deakin University
Waurn Ponds, Victoria, AU
pubudu.pathirana@deakin.edu.au

Aruna Seneviratne
School of Electrical Engineering and Telecommunications
UNSW University, NSW, AU
a.seneviratne@unsw.edu.au

ABSTRACT

Human gender recognition has captured the attention of researchers particularly in computer vision and biometric arena. These investigations based on computer vision or image processing have highlighted applications in security systems, medical applications etc. This work is primarily aimed at investigating a possible characteristic difference in upper body movement kinematics between males and females and, in the affirmative, if that kinematic information alone is sufficient to distinguish each cohort from the other. We use a Microsoft Kinect© to capture the human upper body gait kinematics to uncover gender based kinematic variations.
Two groups of healthy volunteers (18 females, 16 males) were requested to walk along a linear pathway in front of the camera. Upper body movement kinematics were extracted from the male and female cohorts during walking. Principal component analysis (PCA) was employed to substantiate the differences between the two cohorts in terms of kinematic information. Finally, we use k-means clustering to classify and evaluate the performance of the classification system. Despite of the limitation of the dataset, e.g., the limited range of the Kinect© camera, the accuracy of the proposed approach reached up to 94%, indicating that upper body joint movements possess significant information content on human gender based features.

CCS CONCEPTS

• **Applied computing** → **Bioinformatics**; • **Computing methodologies** → *Object recognition*;

KEYWORDS

upper body kinematics; gender recognition; Kinect© sensor; skeleton; PCA; k-means clustering

1 INTRODUCTION

Recent developments in computer vision, image and video processing can be employed to extract meaningful information of human movements, which in turn can result in developing novel applications ranging from rehabilitation to security and surveillance systems. Recognition of human movement features has received considerable attention in the recent past and constitutes an integral part in numerous computer vision systems developments, security and surveillance systems. Motion based Human-Computer Interaction technology particularly in mobile devices is becoming a rapidly enhancing feature impacting modern daily life in numerous aspects.

Gender recognition research is typically based on features of human face, voice and gait. Indeed, the gender dependency on speech is an important and seemingly obvious characteristic, investigated in [5] where the first order statistics of the signal spectrum was used for feature extraction. The gender detection scheme subsequently used neural networks (NN) as a classifier. The system robustness addressing adverse audio quality has been highlighted in addition to the language independency in gender recognition. Besides using voice to recognize human gender, number of studies focused on combining the gender classification with automatic face detection and recognition [3, 12]. Here, firstly, the problem of face detection and feature extraction is addressed prior to engaging in the classification problem by evaluating number of classification algorithms resulting in gender recognition and subsequent performance evaluations.

Recently, gender dependency on gait has attracted the attention of researchers particularly in the bio-medical space. In addition to biometrics such as facial and vocal features walking style or gait patterns are considered to contain gender specific information [7]. In [4], a depth stream and skeleton stream from Kinect© sensor was utilized to capture human silhouettes as they consider gait as a sequence of repetitive walking patterns. Based on that, accumulation of the sequence of silhouette is used to obtain the Gait Energy Image (GEI) which can be considered as the average image of the sequence of silhouette in a gait cycle. Subsequently, the entire GEI is used as the feature vector for training and testing the SVM classifier. In [1], anthropometric and gait information captured from the subjects using a Kinect© sensor are used as features in the gender classification process. In order to obtain anthropometric information, several body segments of each frame were extracted and the Euclidean distance between the two tracked joints from skeleton stream was used.

Although gait-based approach is an effective way to obtain walking style patterns, there are some drawbacks in the gait based classification problem: (i) using image as features usually demands a high computational cost, (ii) different thickness of clothes weakens the distinct differences between males and females when obtaining gait features. In order to overcome these problems and obtain pure

ICBRA 2017, December 8–10, 2017, Barcelona, Spain

ACM ISBN 978-1-4503-5382-3/17/12…$15.00
https://doi.org/10.1145/3175587.3175596

kinematic information, we employ a depth camera to capture relevant joint locations. Even though lower limbs are often used as the crucial biometric for gait analysis [10], it is difficult to collect this information in certain applications due to limited field of cameral view concerning the full body. Therefore, unlike in other gait based approaches, we are primary interested in ascertaining if there is any gender based information embedded in the upper body during the locomotor activity. Kinect© camera is used to capture upper body 3-dimensional joint depth information during gait cycle and subsequently a simplified skeleton is reconstructed. Females have a better upper body control strategy [9], allowing them to exhibit a more effective shoulder to head attenuation compared to males producing relatively smaller upper body accelerations. Though females have a better control strategy, their pelvises in general have higher acceleration than their male counterparts. Based on our observations in line with [9], joints positions located at the head, neck and shoulder were picked out to calculate the signal attributes of human accelerations when walking. Joints from hip were also taken into consideration in our experiment to establish which joints content the most significant information of human gender.

In the feature extraction stage, PCA is employed to obtain the respective component, which exhibits significant level of information of the total variance of each joint that was in turn considered as the feature vector for the succeeding phase. Feature vectors corresponding to all the relevant joints were combined and PCA is employed again to obtain the resulting principal component of the input data. At the last stage, k-means clustering technique is used to classify and evaluate the performance of the overall system.

The remainder of this paper is organized as follows. Section 3 provides a description of data collection involving a Kinect© sensor for kinematic information capture. The main contribution of our work is described in section 4, including feature extraction leading to the gender classification processes. Section 5 reports the experimental results with section 6 providing concluding remarks.

2 DATA ACQUISITION SYSTEM

Kinect© camera motion-sensing device substantially simplifies the data collection in term of capturing kinematic information. Indeed, the electro-optical information from the sensor is used to obtain the 3D joint positions and for our analysis we solely rely on this information. Our primary objective is to capture upper body kinematic information during the gait. Based on the output data from this device, we can readily extract specific kinematic information in terms of joint positions. Besides that, Kinect© camera has the advantage of capturing and tracking human motion without the use of markers, special clothing or and without any form of contact with the individual.

In the software functionality, there are three data streams for the Kinect© sensor including RGB stream, depth map stream and the skeleton stream. In our system, skeleton stream is utilized to take full advantage of reconstructing simplified skeleton and extracting 3D representations of the human joints as depicted in Fig. 1. Moreover, skeleton stream is not sensitive to lighting conditions or objective factors as thickness of clothes that typically results in distinct differences between males and females - particularly relevant to gait-based approaches.

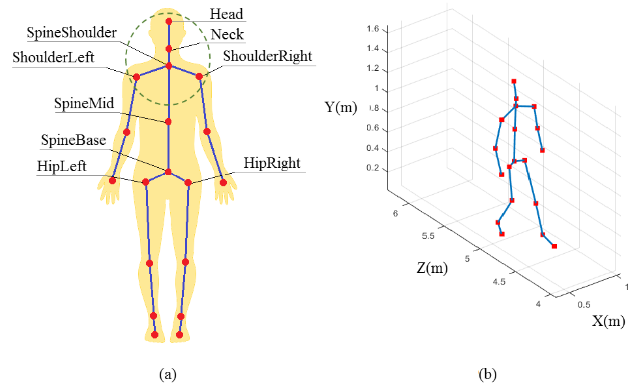

Figure 1: Data extracted from Kinect© sensor: (a) Skeleton joint location (b) Human motion reconstructed in 3D space.

Our data was recorded under indoor laboratory conditions. Two groups of healthy volunteers including 18 females and 16 males were requested to walk 3 meters along a straight line path directly in front of the camera.

3 ANALYSIS

With calibrated measurements from the Kinect© sensor here we look at feature extraction of normalized data streams for gender classification.

3.1 Normalization

Let the position information pertaining to each posture is given by,

$$P = (X_1, Y_1, Z_1, \cdots, X_n, Y_n, Z_n)^T, \tag{1}$$

where n is number of joints of the skeleton, (X_i, Y_i, Z_i) is the coordinates of joint i in the X, Y, Z Cartesian co-ordinate systems.

In the dataset, the subjects have different heights, ranging from 1.6 meters to 1.8 meters for males, and 1.56 meters to 1.6 meters for females. In this study, as we are primarily interested in the kinematics, and as the difference in heights can influence the classification, we have normalized each of the silhouette. In fact, the height differences can cause upper body movement accelerations, adversely affecting the objective of the underlying study (i.e. average male height is greater than that of the female). Therefore, prior to feature extraction, each subject is normalized by its corresponding size scale factor that is defined by finding a least-square solution to the equation (2) [11],

$$P_{j,0} = s_j \frac{1}{n} \sum_{k=1}^{n} P_{k,0}. \tag{2}$$

Here, $p_(j, 0)$ is the static full body posture of subject j. Similarly, $p_(k, 0)$ is the static full body posture of subject k, and s_j is the size scale factor of subject j and n is the number of subjects in our dataset.

This size normalization process avoids the misclassification of each gender for the case where a male's height is close to a female's height. Furthermore, it also reduces the subject height influencing the classification process while the emphasis of the peripheral acceleration characteristics of the upper body is subjugated.

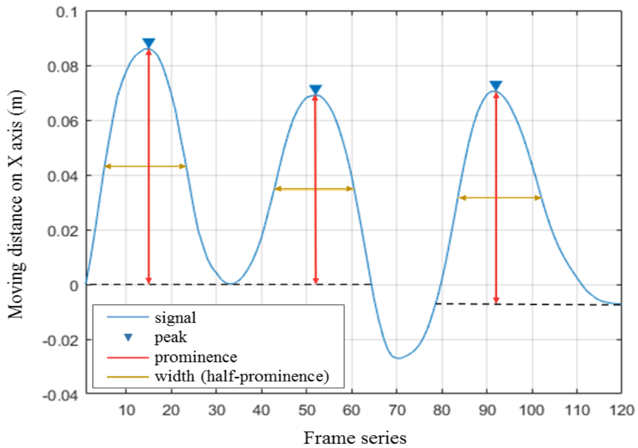

Figure 2: Sine waveform signal of motion of joint on X axis.

3.2 PCA based orthogonalization

According to [9], the females have a better control strategy, allowing them to exhibit a more effective shoulder to head attenuation compared to males resulting in smaller upper body accelerations than those of the males. In other words, there exists a difference in gender via the control of the upper body accelerations during walking. Therefore, in this study, five joints at head, neck, shoulder and spine (indicated by a green dash circle in Fig. 1), were used to extract specific characteristics for the gender classification.

After examining the movement of the upper body of the subject while walking on XY, YZ, XZ planes, we observed that the XY plane contains more information relevant to the motion. For each joint, we obtain the sine waveform signal of its motion on the X axis based on frame series. In the signal, each peak contains the information of one walking step (half gait cycle) of the subject while walking. During the gait cycle, the upper body moves to the left or right on each step, so the prominence of each peak gives us the information about the length of the motion. Besides that, the width of the peak represents the time that the subject spent on one walking step. Therefore, prominence and width of peaks were extracted and considered as specific characteristics of the current joint (Fig. 2).

In order to reduce noise present in raw skeleton data provided by Kinect© sensor, we applied the Moving Average filter [2] to the entire dataset.

With the purpose of combining our concerned joints to utilize their correlation information, firstly we applied PCA to the pair features (prominence and peak width) of these joints to extract their significant values. We then considered these values as a feature dataset to distinguish between males and females. The PCA manly projects the current feature space to a new coordinate space to maximize the variance of the dataset, reduce the original data dimension without missing information of the data [6]. We defined the observations-variables matrix X for each joint as follows,

$$X = \begin{pmatrix} \overline{pr}_1 & \overline{w}_1 \\ \vdots & \vdots \\ \overline{pr}_n & \overline{w}_n \end{pmatrix}. \tag{3}$$

Table 1: Percentage of Total Variance Denoted by The Principal Components of Each Joint

	Head	Shoulder_ Right	Shoulder_ Left	Spine_ Shoulder	Neck
PC1	90.96	91.72	88.36	87.12	87.96
PC2	9.04	8.28	11.64	12.88	12.04

Here, $\overline{pr}_1$, $\overline{w}_1$ are mean values of prominence and width of all peaks in the sine waveform signal of joint i respectively and n is the number of subjects.

In order to estimate the transformation, the eigenvectors and eigenvalues of the covariance matrix of the matrix X are calculated as follows,

$$C = X^T X, \tag{4}$$

$$Cv = \lambda v. \tag{5}$$

Here C is the covariance of X, v is eigenvector and is eigenvalue of C. For each joint, the prominences and width variables are in different units resulting in different variances. Therefore, we calculate the inverse variances of each variable vector in X as a weight matrix to perform PCA. For a random variable vector A made up of n scalar observations, the variance is defined as,

$$V = \frac{1}{n-1} \sum_{i=1}^{n} |A_i - \mu|^2, \tag{6}$$

where μ is the mean of A,

$$\mu = \frac{1}{n} \sum_{i=1}^{n} A_i. \tag{7}$$

Then, weight matrix is calculated by inverting variance values in A as,

$$W = \begin{pmatrix} {V_1}^{-1} & {V_2}^{-1} \end{pmatrix}. \tag{8}$$

The eigenvalues in PCA tell us how much variance can be expected by its associated eigenvector, the highest eigenvalue indicates the highest variance in the data which was observed in the direction of its eigenvector. In order to estimate the percentage of the total variance explained by each principal component, we apply the above weight matrix to the eigenvalues and then calculate the contribution E as follows,

$$E = \frac{W * \lambda}{\sum (W * \lambda)} \times 100, \tag{9}$$

Table 1 shows the contribution of each principal component on the variation of each joint. According to this experiment, we can see that the first component, which contains 87% to 91% of the total variance, denoting more information of the joint in a new coordinate system. Therefore, it is realizable if we discard the second component and keep the first one as the feature vector of each joint for the next process.

Subsequently, we combine feature vectors of all joints from the upper body and employ PCA to find the new coordinate frame resulting in a better separation of the input data. The process of this step is similar to the previous step without the use of a weight matrix. Observations-variables matrix in this step is defined in (10), where the observation in fact is the subject in our dataset, while the variables of each observation is the first principal component of all 5 joints in the previous step.

$$Y = \begin{pmatrix} PC_{1,1} & \dots & PC_{1,5} \\ \vdots & \ddots & \vdots \\ PC_{n,1} & \dots & PC_{n,5} \end{pmatrix}, \tag{10}$$

Here, n is the number of subjects in our dataset, $PC_{i,j}$ is the feature value which is extracted from the previous step of the i^{th} joint of subject j. In the new coordinate system, the first two components account for 96% of the total variance of the data.

3.3 Classification

In order to classify the input data and evaluate the significance of the features extracted from the previous step, we employ the unsupervised learning method of k-means clustering which automatically cluster the set of subjects into 2 groups: male and female. Here, all subjects are partitioned into k ($k = 2$) distinct clusters based on the Squared Euclidean distance between their feature values to the centroid value of a cluster. Indeed, the set of observations $(x_1, x_2, \dots, x_n)$, where n is the number of subjects with each observation is a 2-dimensional feature vectorare partitioned into two sets $G = \{G1, G2\}$ so as to minimize the within-cluster sum of squares [8],

$$\arg\min \sum_{i=1}^{2} \sum_{x \in G_i} x - \mu_i{}^2 = \arg\min \sum_{i=1}^{2} |G_i| \, Var\ G_i. \tag{11}$$

Here μ_i is the mean of points in G_i. This is equivalent to minimizing the pairwise squared deviations of points in the same cluster.

4 EXPERIMENTAL RESULTS AND DISCUSSION

In our experiment, data is collected from two groups of healthy volunteers including 18 females and 16 males under an indoor laboratory conditions.

In order to identify the joints containing the most significant information that intrinsically consists gender related information, prominence and peak width of nine joints from the upper body named in Fig. 1 are taken into consideration. Fig. 3 is the distribution of the input data using the first principal component values when we apply PCA for prominence and peak width of the joint under consideration. In our experiment, joints from shoulder to head give the better variance information of the dataset where the distance of mean values are greater and the standard deviations are smaller compared to joints in the hip area. Our experiment also points out that the joints in the head and right shoulder higher contribution when we combining with the other joints in gender based classification.

From the upper body, three sets of joints were extracted and were subjected to the same process to evaluate the correlation between the movements of joints and the difference in human gender. We define three different cases as follows,

- *Case 1*: Five joints including Head, Neck, ShoulderLeft, ShoulderRight and SpineShoulder (green dash circle in Fig. 1).
- *Case 2*: Six joints including SpineMid joint in addition to the joints in *Case 1*.
- *Case 3*: Nine joints including *Case 2* and HipLeft, HipRight and SpineBase joints.

Table 2: Group Distribution and Classification Accuracy for Different Combinations of Joints

k-means clustering					
	Accuracy	μ_1	σ_1	μ_2	σ_2
Case 1	32/34	0.59	1.25	-0.86	1.24
Case 2	32/34	0.64	1.35	-0.94	1.34
Case 3	31/34	1.65	1.73	-0.95	1.83

In *Case 1*, we considered joints from shoulder to head, which gave better variance information as information as per the preceding description. In *Case 2* we included SpineMid joint to *Case 1* joints. We considered SpineMid in *Case 2* instead of *Case 1* as we have noticed that joints from shoulder to head have similar motion directions and attitudes and hence resulting in a better correlation when prominence and peak width from their waveform signal were used. According to [9], even though females have a better control strategy, their pelvis have higher acceleration compared to men. Therefore, we defined *Case 3* to combine the joints from shoulder to head and also joints from hip to investigate if the gait based distinguishing characteristic between female and male are emphasized.

According to the output from k-means clustering, our data can be clustered into two groups automatically, with accuracy reaching up to 94.12% (*Case 1*). In Fig. 4, female subjects are located in light red area and male subjects are in light blue area. There are two female subjects misclassified into the male group. Moreover, the accuracy, mean μ and standard deviation (SD) σ of two clusters are used to evaluate the results of the above three cases. In generic terms, larger distance between the two mean values of each cluster and smaller SD value in each cluster implies the cluster separation is greater.

From the outcomes depicted in Table 2, for *Case 1* and *Case 2*, distance between two mean points of two cluster are 1.45, 1.58, respectively, so in *Case 2*, this distance is relatively significant. However, the SD values are greater than the value for *Case 1*. For *Case 3*,two clusters are further apart. However, the accuracy rate in *Case 3* is lower compared to the other two cases. Therefore, among three cases, *Case 1* gives the best result with higher accuracy with a better distribution in each cluster. For *Case 3*, the combination of joints from *Case 1* and joints from the hip does not contributed to increase the accuracy; in fact it decreases the correlation values.

5 CONCLUSION

According to the experimental results, with gender classification accuracy exceeding 94%, we can deduce that the upper body accelerations are vital features for human gender representation. Indeed the upper body kinematics contains gender specific characteristics during gait. Combining joints at head, neck, and shoulder provides enhanced information content with regards to the distinct differences between males and females engaged in the locomotor activity. Among these, joints at the head and right shoulder contain the most distinct information between males and females. Moreover, even

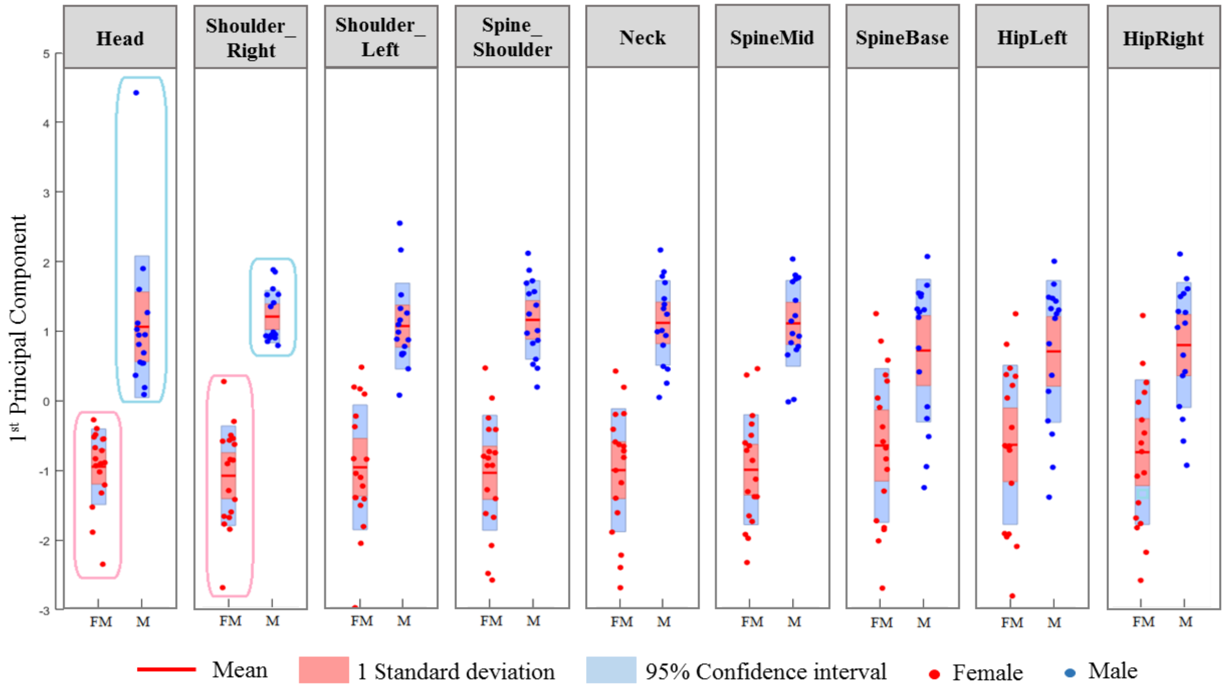

Figure 3: Mean and Standard deviation of the first principal component of upper body joints for male and female subjects.

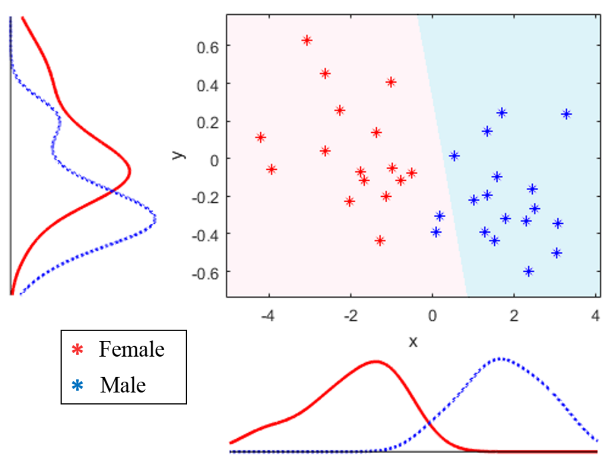

Figure 4: Male and female recognition using k-means clustering.

though the pelvis acceleration is also a signification characteristic to distinguish between female and male, the prominence and peak width feature of the sine waveform signal considered in our research are not the appropriate features for the joints at hip. By using features from joints in the skeleton stream, the proposed approach also overcomes the drawbacks in cloth thickness dependencies and computational demand requirements, which is applicable in number of practical applications.

ACKNOWLEDGMENTS

This work was supported by CSIRO's Data61.

REFERENCES

[1] Virginia Andersson, Livia Amaral, Aline Tonini, and Ricardo Araujo. 2015. Gender and Body Mass Index Classification Using a Microsoft Kinect Sensor. (2015). https://www.aaai.org/ocs/index.php/FLAIRS/FLAIRS15/paper/view/10409

[2] Mehran Azimi. 2017. Skeletal Joint Smoothing White Paper. (2017). Retrieved September 27, 2017 from https://msdn.microsoft.com/en-us/library/jj131429.aspxl

[3] B. Cerit, S. A. BÃűlÃijk, and M. F. Demirci. 2016. Analysis of the effect of image resolution on automatic face gender and age classification. In *2016 24th Signal Processing and Communication Application Conference (SIU)*. 853–856.

[4] M Eltaher, Y Yang, and J Lee. 2015. Gait-based gender classification using kinect sensor. 122 (01 2015).

[5] H. Harb and Liming Chen. 2003. Gender identification using a general audio classifier. In *Multimedia and Expo, 2003. ICME '03. Proceedings. 2003 International Conference on*, Vol. 2. II–733–6 vol.2. https://doi.org/10.1109/ICME.2003.1221721

[6] Ian Jolliffe. 2002. *Principal component analysis)*. Wiley Online Library.

[7] F. M. Kazemi, W. Banzhaf, and M. Gong. 2016. Human recognition through walking styles by multiwavelet transform. In *2016 Eighth International Conference on Information and Knowledge Technology (IKT)*. 47–53. https://doi.org/10.1109/IKT.2016.7777786

[8] J. MacQueen. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. University of California Press, Berkeley, Calif., 281–297. http://projecteuclid.org:443/euclid.bsmsp/1200512992

[9] C. Mazzà, M. Iosa, P. Picerno, F. Masala, and A. Cappozzo. 2009. *Gender differences in the control of the upper body accelerations during level walking*. Springer Berlin Heidelberg, Berlin, Heidelberg, 43–46.

[10] Mark S. Nixon, Tieniu N. Tan, and Rama Chellappa. 2005. *Human Identification Based on Gait (The Kluwer International Series on Biometrics)*. Springer-Verlag New York, Inc. Secaucus, NJ, USA 2005.

[11] Nikolaus F. Troje. 2002. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision* 2, 5 (2002), 2. https://doi.org/10.1167/2.5.2 arXiv:/data/journals/jov/933498/jov-2-5-2.pdf

[12] X. Wang, A. Mohd Ali, and P. Angelov. 2017. Gender and Age Classification of Human Faces for Automatic Detection of Anomalous Human Behaviour. In *2017 3rd IEEE International Conference on Cybernetics (CYBCONF)*. 1–6.