www.czasopisma.pan.pl *Metrol. Meas. Syst.*, Vol. 25 (2018) No. 2, pp. 253–267. www.journals.pan.pl

METROLOGY AND MEASUREMENT SYSTEMS

Index 330930, ISSN 0860-8229

www.metrology.pg.gda.pl

GENDER DETECTION USING 3D ANTHROPOMETRIC MEASUREMENTS BY KINECT

Seda Camalan[1], Gokhan Sengul[2], Sanjay Misra[2,3], Rytis Maskeliūnas[4], Robertas Damaševičius[5]

1) *Atilim University, Department of Information Systems Engineering, Incek, Ankara 06836, Turkey* (*seda.camalan@atilim.edu.tr*)
2) *Atilim University, Department of Computer Engineering, Incek, Ankara 06836, Turkey* (*gokhan.sengul@atilim.edu.tr*)
3) *Covenant University, Department of Electrical and Information Engineering, Ota 1023, Nigeria* (*sanjay.misra@covenantuniversity.edu.ng*)
4) *Kaunas University of Technology, Department of Multimedia Engineering, Kaunas 51368, Lithuania* (*rytis.maskeliunas@ktu.lt*)
5) *Kaunas University of Technology, Department of Software Engineering, Kaunas 51368, Lithuania* (✉ *robertas.damasevicius@ktu.lt*, +370 609 43772)

Abstract

Automatic gender detection is a process of determining the gender of a human according to the characteristic properties that represent the masculine and feminine attributes of a subject. Automatic gender detection is used in many areas such as customer behaviour analysis, robust security system construction, resource management, human-computer interaction, video games, mobile applications, neuro-marketing *etc*., in which manual gender detection may be not feasible. In this study, we have developed a fully automatic system that uses the 3D anthropometric measurements of human subjects for gender detection. A Kinect 3D camera was used to recognize the human posture, and body metrics are used as features for classification. To classify the gender, KNN, SVM classifiers and Neural Network were used with the parameters. A unique dataset gathered from 29 female and 31 male (a total of 60 people) participants was used in the experiment and the Leave One Out method was used as the cross-validation approach. The maximum accuracy achieved is 96.77% for SVM with an MLP kernel function.

Keywords: gender detection, Kinect sensor, anthropometrics, measurement, gender issues.

1. Introduction

Automatic gender detection is a process of determining the gender of a human according to the characteristic properties that represent his/her masculinity or femininity. Gender detection is used in many areas, such as personalization and recommender systems [1], behaviour analysis [2], consumer research [3], digital forensics [4], security and biometrics [5], human computer interaction [6], mobile applications [7], *etc*.

Article history: received on Jan. 26, 2018; accepted on Mar. 18, 2018; available online on Jun. 30, 2018, DOI: 10.24425/119568.

Traditional approaches for gender recognition include face recognition [6] and handwriting analysis [4]. Some studies showed that the posture of human bodies can be used for gender classification [8–10]. The approaches for gender detection based on data derived from the human body can be classified as appearance-based and non-appearance-based [11]. The appearance-based approaches use static body features (such as face, fingernail, body shape [12] *etc*.), dynamic body features (gesture, motion, gait *etc*.), and apparel features (clothing, footwear *etc*.). The non-appearance-based approaches use biometric features (fingerprint, iris, ear, skin colour [13] *etc*.), bio-signals (DNA, EEG *etc*.), and social information (blog, email, handwriting *etc*.) for gender detection. The most commonly used features for gender detection are the features gathered from the human face [14, 15]. The face information is used not only for gender detection, but also for subject identification, age detection, emotion prediction *etc*. In order to use face images for gender detection, *high resolution* (HR) images acquired from a right angle that capture all parts of the face are needed. However, in daily life it is not practical to obtain HR face images because of the distance between the face and the camera, and the orientation angle between the camera and the subject being imaged. Finally, when some parts of the face are occluded, its image might be not usable for gender detection. In order to overcome this difficulty, the whole-body properties are used for gender detection. While considering the whole body as an input to gender detection, different measurements, such as: anthropometric [16], gait and motion properties [17] of a human, kinematical, dynamical, and motor control parameters of human skeleton and limbs [18] can be used. Low-cost range sensors such as a Kinect sensor (Microsoft, Redmond, WA, USA) are able to track the orientation and position of a human body and its limbs, and can be used to measure the anthropometric dimensions of a human [19]. Furthermore, the accuracy of measurements performed by a Kinect sensor was validated with metrological methods [20].

In this paper, an automatic gender detection system is proposed, which uses the anthropometric measurements of the human skeleton. The anthropometric measurements determine the size- and shape-related descriptors of a human body that are used in several domains of application, such as clothes, equipment and application materials, ergonomics and architecture. The measured objects include height, width and weight of a human body and its parts [21]. Since the body proportions are different for men and women, the coordinates of joints provided by the motion tracking technology enable to find individual measurements of lengths and proportions of humans [22]. Our approach is based on determining lengths between the joint points and heights of individuals. To verify this approach, the accuracy of gender detection is experimentally tested using 60 subjects in different age ranges. The 3D coordinates of joint points are obtained by a 3D camera. To predict the gender according to the features obtained by the measurements of lengths between joints and height of a subject, *K-Nearest Neighbour* (KNN), *Artificial Neural Network* (ANN), and *Support Vector Machine* (SVM) are used.

The paper is organized as follows. Section 2 describes the approaches used for gender detection with gait, anthropometric, gesture and body metric methods. Section 3 presents in detail the methods of collecting the data and obtaining the features employed in the study and explains the classification methods that are used. The results of the experiments are presented in Section 4. Finally, Section 5 concludes the paper.

2. Related work

The body-based gender detection studies can be categorized into four groups: 1) gait-based; 2) based on anthropometric measurements; 3) motion-based approaches; and 4) a combined approach using both gait and anthropometric measurements. In this section, we first summarize the gait-based and gesture-based approaches and then we focus on studying the use of anthropometric measurements, as well as combined gait and anthropometric measurements.

Body motions during walking were analysed to predict the gender in many studies with the image-based gait analysis, and the studies [23, 24] and [25] achieved accuracy rates of 83%, 94% and 96%, respectively. High accuracy rates are achieved even without using HR images. Subject identification and gender detection can be done with the help of human silhouettes while walking [24].

For gender detection, Cao *et al.* [26], Guo *et al*. [27] and Collins *et al.* [28] used all body parts identified from 2D video images to extract the walking features of a person and to classify them according to the gender. Miyamoto and Aoki [29] used the time series variation, which was normalized with a linear interpolation of the joint positions taken from a Kinect v2, to determine the significant features for gender detection. In the study, not only three-dimensional coordinates of joint positions were used, but also projections of joints from a 3D space onto a 2D plane according to the exposure angle. The authors used SVM as the classifier to predict the gender, reaching accuracy of 99.12% with 12 subjects (six males and six females). Although gait-based approaches do not require HR images, one drawback of them is that they require the images of the full walking period of a person, which is not always possible in crowded urban environments.

To predict the human gender, gestures of the human body was used by Won et al. [30]. In the study, the non-verbal behaviours were used to predict the human gender. With different postures of twelve men and twelve women, their gestures were captured in front of a Kinect camera and used as the training data. The angles between shoulders and neck of human gestures were extracted as the movement features, and the lengths of body parts were extracted as the static length features for classification into ten different postures. After using a standard technique of reducing the number of features, top ten features were selected to classify the gender, reaching accuracy of 83%.

The gender detection approaches based on anthropometric measurements can be divided into 2 groups. The first group uses 2D anthropometric measurements, while the second group uses 3D measurements. The first introduced study of 2D anthropometric measurements was proposed by Adjeroh *et al.* [31] with an analysis of the correlation of anthropometric features from CAESAR database [32]. Also, by using anthropometric features a copula model was proposed using the same CAESAR database by Cao *et al*. [33] to the prediction of gender and weight. The proposed method using a combination of the copula model and SVM gives accuracies of 99.1% for the body-only feature, 87.4% for the head-only feature, and 99,4% for both head and body features.

To predict the gender from still images using the human metrology was also studied with ratios of anthropometric measurements for the LUPI (*Learning Using Privilege Information*) framework by Kakadiaris *et al.* [34], followed by a similar approach with Cao *et al*. [33]. In [33], the ratios of the anthropometric measurements are more accurate than those of the actual measurements, achieving accuracy of 98% for the LUPI framework. The study also shows the result of 3D pose estimation algorithm to obtain joint locations in real images from the PaSC and SARC3D datasets from 2D annotation. The best accuracy value for real images was 86%. However, the depth information was not obtained in the anthropometric measurements.

There are only a few studies using the 3D anthropometric measurements. One of them is that by Sandygulova *et al*. [32]. In the study, the predictions of gender and age of children were experimentally tested for an age range between 5 and 18 for 276 boys and 152 girls. Standard machine learning approaches were applied to features, which were obtained with Kinect by modelling 3D body metrics. The system performance was evaluated on volunteers by using an adaptive social robot in the wild. For gender detection, the shoulder height and hip-to-shoulder proportion were important features for classification. The gender detection reached accuracy of 73% in the real-time adaptive robotic system.

Another gender detection approach using the 3D anthropometric measurements was suggested by Andersson *et al*. [35]. In that study, the gender prediction and body mass index prediction were investigated using both anthropometric and gait information. Eighty features were used for classification: the Euclidean distance of tracked joints (20 attributes), and gait attributes, such as spatiotemporal (4 attributes) and kinematic (56 attributes). For training and testing, the images captured from 44 individuals were used for each set. For the classification part, KNN ($K = 3$), SVM and a *multilayer perceptron* (MLP) were used. The most accurate result of 95% was obtained by MLP using all attributes. Using the anthropometric measurements only, MLP achieved accuracy of 93%.

3. Materials and methods

3.1. Data collection

As no alternative datasets containing both skeleton joint position data and large enough gender information were found, we have collected our own dataset. For this study, 60 volunteers (29 – female, and 31 – male, all aged between 20 to 60) participated in the experiment. The test environment, shown in Fig. 1, was located in the laboratory in Atilim University, Turkey. To record the body metrics from volunteers, we used a KinectV1 camera.

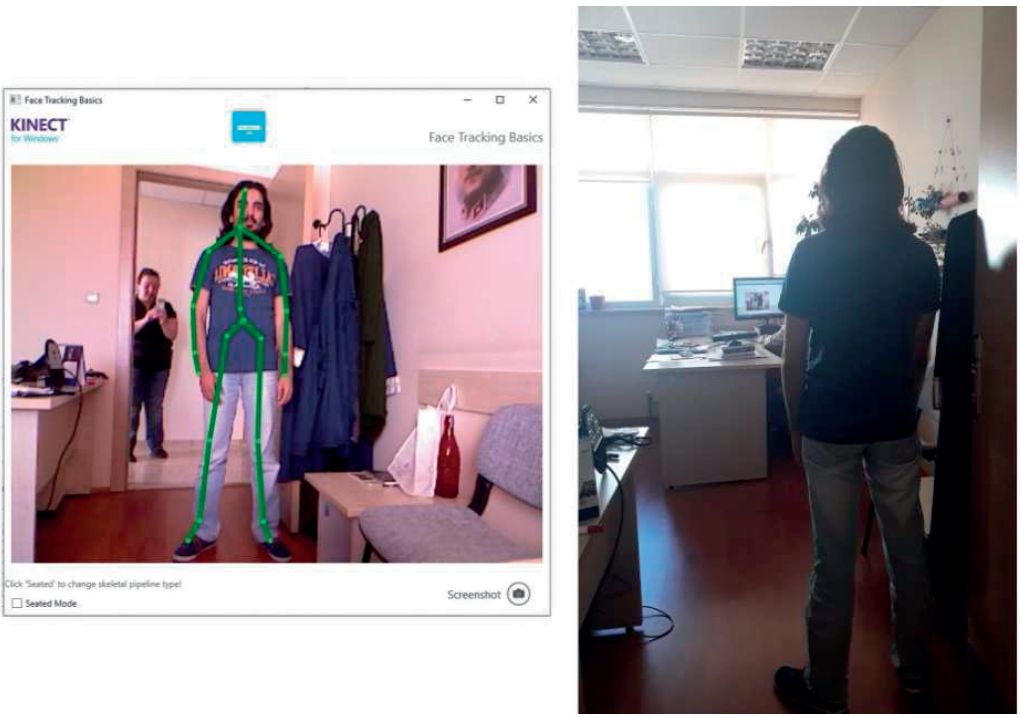

Fig. 1. The test Environment in Atilim University.

Each volunteer was asked to stand in a free pose in front of the camera. Each volunteer was free to take any position or do any action for any period of time. There was no restrictions on time because the average values of all frames were taken for each target to extract features and generate a feature vector.

3.2. Origin of idea

To predict the human gender, differences between men and women have been investigated in many studies, some of which are mentioned above. In our problem, the body proportions and heights of humans are considered to predict the gender. The research proposed by Loomis [36] gives the ideal proportions in centimetres for both genders, as shown in Fig. 2.

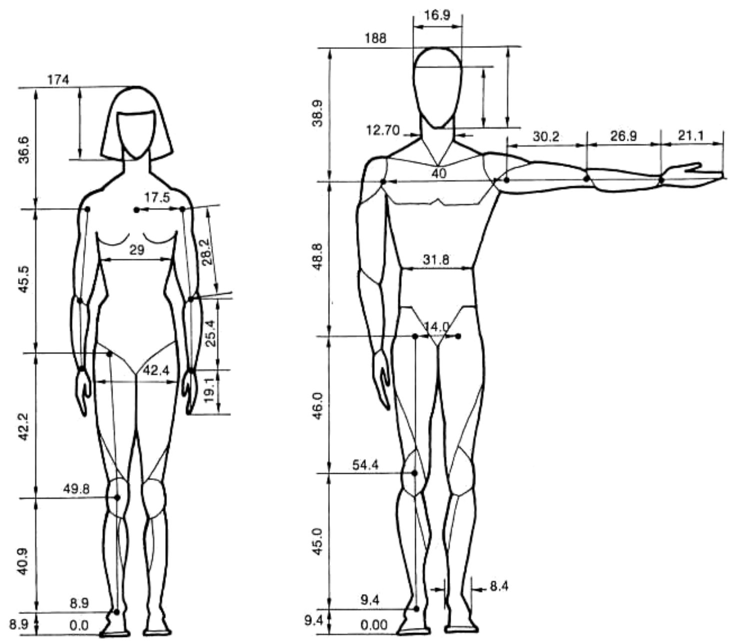

Fig. 2. The ideal gender proportions [36] in centimetres: female (left) and male (right).

The left of Fig. 2 corresponds to female and the right – to male and the proportions of each part of the body are given in centimetres. We can see that for men and women the proportions of some parts of the body are different. For instance, women' hips are larger than men' hips, and men' shoulders are larger than women' shoulders. In [22], these lengths are used with the height of humans to classify according to gender. In this study, similarly to the approach mentioned above, we assume that the distances between joint points might change depending on the gender of a human. Therefore, we have obtained the positions of the following joint points and performed the feature extraction from these joint points.

3.3. Obtained joint positions and features

In this study, we have used a Microsoft Kinect 3D camera, which is a motion capturing device. This camera not only acquires the standard RGB image of an imaged object but also the distance (depth information) between the object and the camera. The Kinect camera has five components: *infrared* (IR) Emitter, Colour Sensor, IR Depth Sensor, Tilt Motor and Microphone Array.

To acquire the depth image, IR Emitter emits IR light beams and IR Depth Sensor scans the reflected beams back, whereas the depth information is measured as the distance between the sensor and the target.

The Kinect skeletal model provides 20 joint points of a tracked person with the depth information [37] as follows: hip centre – 1, spine – 2, shoulder centre – 3, head – 4, left shoulder – 5, left elbow – 6, left wrist – 7, left hand – 8, right shoulder – 9, right elbow – 10, right wrist – 11, right hand – 12, left hip – 13, left knee – 14, left ankle – 15, left foot – 16, right hip – 17, right knee – 18, right ankle – 19, right foot –20 (see Fig. 3).

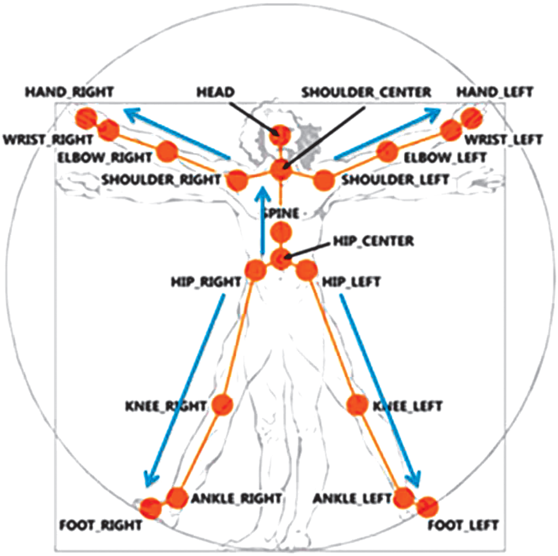

Fig. 3. Positions of skeleton joints obtained by Kinect v1 (adopted from [38]).

These 20 joint points are used to obtain the features that are used in the classification stage. The features are the Euclidian distance between any two joint points and the height of the target. If it is needed to indicate with a formula, the position of a joint is shown as p_i, defined as follows:

$$p_i = (x_i, y_i, z_i) \quad where \quad i \in \{1,2,3,\ldots,20\}. \tag{1}$$

The Euclidian distance between the joints is defined as:

$$dist P_{m,n} = \sqrt[2]{(x_m - x_n)^2 + (y_m - y_n)^2 + (z_m - z_n)^2}, \tag{2}$$

where m and n are numbers of two different joint points from 1 to 20, $m,n \in \{1,2,3,\ldots,20\}$. The mentioned joint points are listed with their numbers in this section.

The distance between each joint $dist P_{m,n}$ and any other joint is calculated using (2). There are 190 binary combinations of twenty joint points, therefore there are 190 features. Additionally, three more features (height by skeleton, height by estimation and height by wingspan) are calculated as given by (3) and (4) [22].

Metrol. Meas. Syst., Vol. 25 (2018), No. 2, pp. 253–267.

Height By Skeleton: the total length from foot to head plus half of the head height, because the head position coordinates give the centre of the head].

$$\text{Height by Skeleton} = \mathrm{L_{H\text{-}SC}} + \mathrm{L_{SC\text{-}S}} + \mathrm{L_{S\text{-}HC}} + \mathrm{L_{HC\text{-}K}} + \mathrm{L_{K\text{-}A}} + \mathrm{L_{A\text{-}F}} \tag{3}$$

where

$\mathrm{L_{H\text{-}SC}} = dist\,P_{\text{head-shoulderCentre}}$, $\mathrm{L_{SC\text{-}S}} = dist\,P_{\text{shoulderCentre-spine}}$,
$\mathrm{L_{H\text{-}SC}} = dist\,P_{\text{spine-hipCentre}}$, $\mathrm{L_{HC\text{-}K}} = dist\,P_{\text{hipCentre-kneeLeft/kneeRight}}$,
$\mathrm{L_{K\text{-}A}} = dist\,P_{\text{kneeLeft/kneeRight-ankleLeft/ankleRight}}$, $\mathrm{L_{A\text{-}F}} = dist\,P_{\text{ankleLeft/ankleRight-footLeft/footRight}}$.

Height By Estimation: This height calculation method depends on the idea that the whole body is not captured by a camera and some of upper or lower parts of the body are not seen. The proposed method [36], which recommends that the distance between shoulders and knees is equal to 52% of the overall height according to the ideal body proportions, is applied to the height estimation.

Height By Wingspan: This height calculation method depends on the idea that the subject is considered to be in a sitting position. According to the ideal body proportions (Fig. 2), the length of wingspan is equal to the height of the target.

$$\begin{gathered}\text{Height by Wingspan} = \mathrm{L_{HL\text{-}WL}} + \mathrm{L_{WL\text{-}EL}} + \mathrm{L_{EL\text{-}SL}} + \\ + \mathrm{L_{SL\text{-}SC}} + \mathrm{L_{SC\text{-}SR}} + \mathrm{L_{SR\text{-}ER}} + \mathrm{L_{ER\text{-}WR}} + \mathrm{L_{WR\text{-}HR}}\end{gathered} \tag{4}$$

where

$\mathrm{L_{HL\text{-}WL}} = dist\,P_{\text{handLeft-wristLeft}}$, $\mathrm{L_{WL\text{-}EL}} = dist\,P_{\text{wristLeft-elbowLeft}}$,
$\mathrm{L_{EL\text{-}SL}} = dist\,P_{\text{elbowLeft-shoulderLeft}}$, $\mathrm{L_{SL\text{-}SC}} = dist\,P_{\text{shoulderLeft-shoulderCentre}}$,
$\mathrm{L_{SC\text{-}SR}} = dist\,P_{\text{shoulderCentre-shoulderRight}}$, $\mathrm{L_{SR\text{-}ER}} = dist\,P_{\text{shoulderRight-elbowRight}}$,
$\mathrm{L_{ER\text{-}WR}} = dist\,P_{\text{elbowRight-wristRight}}$, $\mathrm{L_{WR\text{-}HR}} = dist\,P_{\text{wristRight-handRight}}$.

Together with these three features, a total of 193 features are obtained for each frame. In order to avoid any unbalanced feature number depending on a frame and some missing joints that may be obtained by emitting the reflected IR light beams from the Kinect camera, the average values of each feature are calculated for all frames to prepare the feature vector for the classification stage.

While collecting the data, the actual heights of the participants were also recorded to verify our height calculations and the ability of capturing data by the Kinect camera. When the actual and experimental results are compared, the results are close to each other.

3.4. Feature selection methods

In the classification stage, all features were used both individually and jointly for each classification method. However, the accuracy rate may be lower when all features are used for classification. Testing each feature with different combination sizes for 193 features would be a very time-consuming process, so that not all combinations of 193 features were tested. On the other hand, the features should be combined based on a criterion to increase the classification accuracy. The criterion was identified according to the accuracy rate of an individual feature performance. Starting from feature 1 to 193, each accuracy rate of classification was checked. If the accuracy rate was greater than 70%, the individually classified feature was selected to combine with other selected features. Otherwise, the feature was not added to the set of selected features. After the set of selected features was generated, an exhaustive search [39] was done to combine the selected features that satisfy the criterion. Therefore, all possible combinations of selected features were tested for KNN and SVM classifiers. For the neural network, the feature selection was not applied because of its structure of weighted neurons.

3.5. Classification methods

To predict the gender, we have used KNN, SVM and *Artificial Neural Network* (ANN) in this study. These approaches are summarized below.

The *K-Nearest Neighbour* (KNN) method classifies the test data by comparing them with the K-most similar data (neighbours) in the training set; here K is the number of neighbours considered for classification. The similarity of data is measured by a distance function which is most commonly chosen as the Euclidean Distance:

$$d_{Euclidean}(x,y) = \sqrt{\sum_i (x_i - y_i)^2}, \tag{5}$$

where $x_i = x_1, x_2, \ldots, x_m$ and $y_i = y_1, y_2, \ldots, y_m$ denote feature vectors.

KNN was used for gender recognition in the previous studies [40, 41].

SVM separates the data points by maximizing the distance of each class from a hyper-plane, which intersects the feature space [42]. The classifier is constructed by finding an optimal hyper-plane. Kernel functions are used to find an optimal hyper-plane by mapping the input space onto a multi-dimensional feature space to separate the data points linearly. In this work, we have used linear, quadratic, polynomial, *radial basis function* (RBF) and *multilayer perceptron function* (MLP) kernels. For the description and formal definition of the methods refer to [43]. SVM was also used for gender recognition in the previous studies [44, 45].

ANN is widely used in pattern recognition and also gender detection [46, 47] because of the good performance in its applications. The ANN structure consists of layers composed of individual neurons. Each neuron is associated with its weight which is learned during the training stage to reduce the network error. In an ANN, there are three layers: an input layer, a hidden layer, and an output layer. The input layer is the first layer where each sample of the network is used as an input. The hidden layer is the second layer where the error reduction is performed in the network. The last layer is the output layer where the desired outputs are determined with the same number of neurons to represent each output separately.

In Fig. 4, the Neural Network is presented. There are three layers: Input Layer, Hidden Layer, and Output Layer. In the input layer, F_1, F_2, ..., F_{193} denote all features and $w_{1,1}$, w_{193,n_s} denote

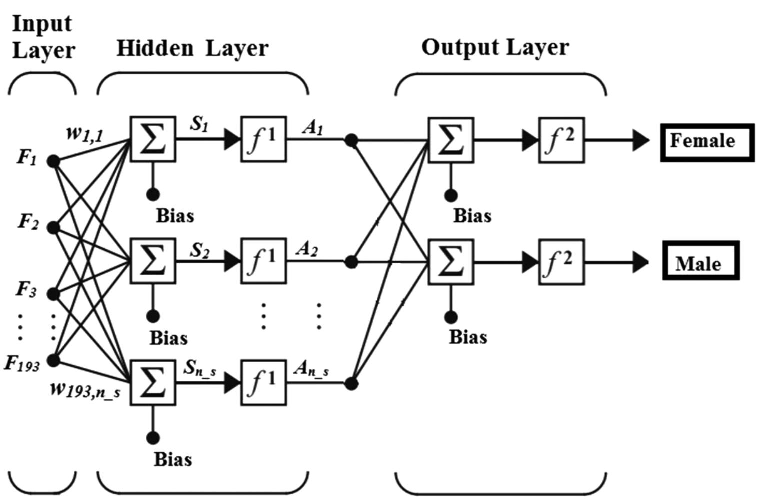

Fig. 4. ANN Structure of the Proposed System.

weights of the related features, where w_{193,n_s} is a weight of input feature 193 in a neuron size n_s (where $n\_s = 10, 15, 20, 25, 30, 35, 40, 45$) of the hidden layer. In the hidden layer f^1 is an activation function, $S_1, S_2, \ldots, S_{n\_s}$ are weighted sums added to bias. $A_1, A_2, \ldots, A_{n\_s}$ are activation function results.

$$\begin{aligned}
&S_1 = w_{1,1}F_1 + w_{1,2}F_2 + \ldots + w_{1,193}F_{193} + Bias; \quad A_1 = f(S_1)\\
&S_2 = w_{2,1}F_1 + w_{2,2}F_2 + \ldots + w_{2,193}F_{193} + Bias; \quad A_2 = f(S_2)\\
&\ldots\\
&S_{n_s} = w_{n_s,1}F_1 + w_{n_s,2}F_2 + \ldots + w_{n_s,193}F_{193} + Bias; \quad A_{n_s} = f(S_{n_s})
\end{aligned} \tag{6}$$

All procedures of the work are presented in the following algorithm.

Algorithm of the study:

- For each volunteer, repeat the following process:
  - Capture all joint points' coordinates from Kinect: $p_i = (x_i, y_i, z_i)$ *where* $i \in \{1, 2, 3, \ldots, 20\}$
  - For a joint point $m, m = 1, 2, \ldots, 20$, calculate its distance to another joint point n, $n = 1, 2, \ldots, 20$, *where* $n \neq m$: Using (2)
  - Calculate ***Height By Skeleton (HbS)***, ***Height By Estimation (HbE)*** and ***Height By Wingspan (HbW)***
  - All calculated distances and heights are concatenated to create the feature vector for a person x where $x = 1, \ldots, 60$, $P_{xFeatures}$: $P_{xFeatures} = [distP_{1,1}, distP_{1,2}, \ldots, distP_{20,19}, HbS, HbE, HbW]$
- Each feature is considered individually and for each classification method find its accuracy $Acc\_P_{ALL\_Features[x]}$, where $x = 1, 2, \ldots, 193$
- If the accuracy $Acc\_P_{ALL\_Features[x]} \geq 70\%$, define the feature x as the selected one from 193 features
- From the selected features, different numbers of combinations of features are classified for each classification
- The results are recorded and the most accurate feature combination is reported for each classification method

4. Results and discussion

To examine the proposed approach, the data collected from volunteers, and the information obtained from the Kinect camera were processed to extract features used to classify the gender. For classification, we used SVM with different kernel functions, KNN with different K values and ANN with three training functions (Scaled Conjugate Gradient, Bayesian regularization, Levenberg-Marquardt optimization). They were all used by the *Leave One Out* (LOO) method to predict the gender of subjects.

Figure 5 shows the classification accuracy when the KNN classifier is used. The black dashed line shows the accuracy when all the features are used, and the red bash-dot line shows the accuracy when the feature set obtained by the exhaustive search giving the best accuracy is used.

The best accuracy achieved is 85.48% with $K = 1$ and $K = 2$ values, while using the following feature combinations: Neck – Right Knee, Neck – Left Foot, Left Shoulder – Right Hip, Right Shoulder – Right Ankle. As the k values increase, the accuracy tends to decrease – as is seen in Fig. 5.

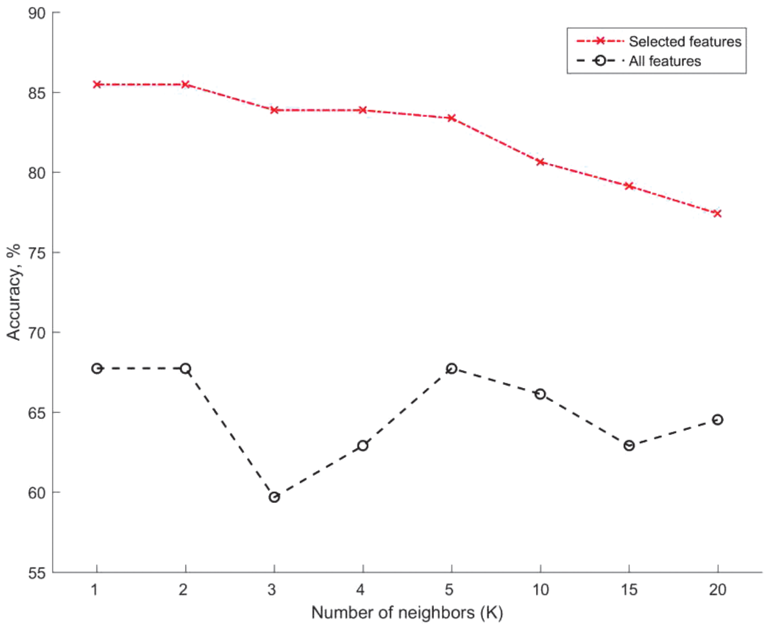

Fig. 5. KNN classifier accuracy results for gender detection vs different K values.

In Table 1, the classification results of SVM classifier are presented. The best accuracy is 96.77% while using the MLP kernel function with seven features (Head – Right Foot, Left Shoulder – Right Hip, Right Shoulder – Left Foot, Spine – Right Foot, Waist – Left Foot, Left Knee – Right Knee, and Right Knee – Left Ankle). The features are presented in Fig. 6. The distances between the cross joint points of feature combinations, which depend upon the widths of shoulder and hip (identifiers for gender), give the best accuracy for gender detection.

Table 1. Accuracies of gender detection in the SVM classification with feature combinations giving the best performance.

Kernel Function	All Features	Best Accuracy	Feature Combinations Giving the Best Performance
Linear	63.33%	83.87%	Left Shoulder – Right Hip, Right Shoulder – Left Hip, Right Shoulder – Left Foot, Spine – Left Hip, Spine – Left Foot, Spine – Right Foot, Waist – Left Foot, Left Hip – Left Foot, Right Hip – Left Foot, Left Knee – Right Knee, Right Knee – Left Ankle, Right Knee – Right Foot
Quadratic	65.00%	93.55%	Head – Right Knee, Neck – Right Shoulder, Neck – Left Knee, Left Shoulder – Left Foot, Left Knee – Right Knee, Right Knee – Left Ankle
Polynomial	76.66%	93.55%	Neck – Right Shoulder, Left Shoulder – Right Hip, Spine – Left Hip, Left Knee – Right Knee
RBF	5.00%	93.55%	Head – Right Knee, Neck – Right Shoulder, Neck – Left Knee, Left Shoulder – Left Foot, Left Knee – Right Knee, Right Knee – Left Ankle
MLP	63.33%	**96.77%**	Head – Right Foot, Left Shoulder – Right Hip, Right Shoulder – Left Foot, Spine – Right Foot, Waist – Left Foot, Left Knee – Right Knee, Right Knee – Left Ankle

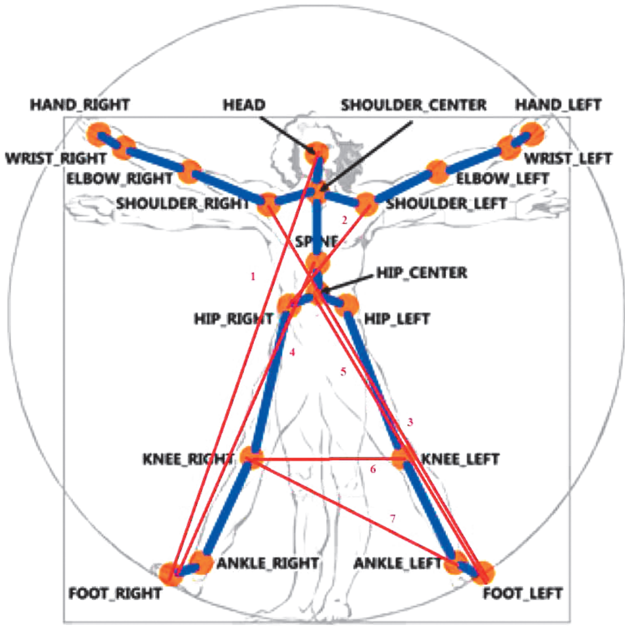

Fig. 6. A combination of features giving the best accuracy with the MLP kernel function of SVM.

As it was mentioned before, the best accuracy rates of the classification methods are obtained with different feature combinations as inputs. Not only the number of selected features, but also the sets of the selected features may be differentiated based on the classification methods. In Table 1, the SVM classifier with the Linear Kernel function gives an 83.87% accuracy rate for twelve features used as inputs. Moreover, the SVM classifier with the MLP Kernel function gives the highest accuracy rate of 96.77% when seven features, which did not exactly match the features used in the Linear Kernel function of SVM, were used as inputs of the classifier. Other methods employed also different numbers of selected features to obtain the best accuracy. But all these combinations of selected features were obtained by testing every possibility. This process was time-consuming, thus using some methods like Fisher Discriminant Ratio would be good to select features that give the best accuracy. In the future work, we intend to use Fisher Discriminant Ratio or similar feature selection methods.

In Fig. 7, the accuracy results obtained with an ANN are presented according to the number of neurons in the hidden layer, their training functions, and momentum. The training functions are defined to update the weight and bias values according to the Scaled Conjugate Gradient method, Bayesian regularization and Levenberg-Marquardt optimization. The best accuracy result (76.66%) is obtained with the Levenberg-Marquardt optimization training function, 35 neurons in the hidden layer and the momentum value of 0.9.

From the results of the experiment, we can see that using the distances between each pair of joint points gives a better accuracy than using the lengths of two neighbouring joint points. The feature combination, which gives the best accuracy results, uses the distances between joint points, which are not directly connected to each other.

We can compare our results with those of Andersson *et al.* [34] and Sandygulova *et al.* [22]. The differences between each study and our study are as follows. The study [22] was performed

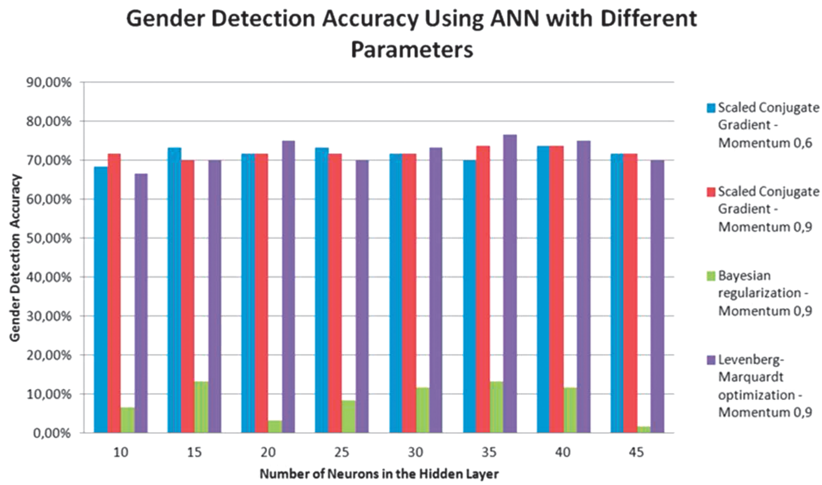

Fig. 7. Gender detection accuracy using an ANN with different parameters.

in a group of individuals between 5–18 of age and the accuracy rate was 73%. In our study, the age range was between 20 and 60 years and the accuracy was 96.77 %. The study [34] achieved 95% accuracy using the gait and anthropometric measurements, while we managed to achieve 96.77% accuracy using only the anthropometric measurements. Furthermore, even when the camera cannot see the entire body, we can predict the gender with the help of the ratio of different joint points.

While our results are still lower than those achieved using face images (*e.g.* 95%, 97% and 98% in [48, 14, 49], respectively), an advantage of our approach is that we do not need high resolution face images of a human, and the gender can still be predicted if the face images cannot be acquired).

5. Conclusions

In this paper, a new approach to gender detection is proposed by using both anthropometric body metrics and posture of individuals. Our main contribution is that the anthropometric measurements of a person are automatically obtained with the developed system and using a 3D camera. The second contribution is the collection of a unique dataset of skeleton joint positional data with the subject gender information. The dataset is available on request.

For gender detection, we have used three classifiers (KNN, SVM, and Neural Network) with different parameters. On the basis of a dataset of records collected from 60 subjects, aged from 20 to 60, the best performance for gender detection is 96.77% achieved with the MLP kernel of SVM classifier, while using eight selected features (Head – Right Foot, Left Shoulder – Right Hip, Right Shoulder – Left Foot, Spine – Right Foot, Waist – Left Foot, Left Knee – Right Knee, and Right Knee – Left Ankle).

The best accuracy is obtained for a combination of features, which does not directly depend on the widths of shoulder and hip but is indirectly related to the widths. This approach, considering both anthropometric and postural information, is different from the previous gender detection approaches using the gait or bone-length analysis.

In the future work, we plan to use both anthropometric measurements and face images for gender detection in order to increase its accuracy. Besides, we also plan to investigate whether the anthropometric measurements can be used also for age detection.

References

[1] You, Q., Bhatia, S., Sun, T., Luo, J. (2014). The Eyes of the Beholder: Gender Prediction Using Images Posted in Online Social Networks. *2014 IEEE International Conference on Data Mining Workshop*, Shenzhen, 1026–1030.

[2] Zhang, J., Du, K., Cheng, R., Wei, Z., Qin, C., You, H., Hu, S. (2016). Reliable Gender Prediction Based on Users' Video Viewing Behavior. *2016 IEEE 16th International Conference on Data Mining (ICDM)*, Barcelona, 649–658.

[3] Duong, D., Tan, H., Pham, S. (2016). Customer gender prediction based on E-commerce data. *2016 Eighth International Conference on Knowledge and Systems Engineering (KSE)*, Hanoi, 91–95.

[4] Topaloglu, M., Ekmekci, S. (2017). Gender detection and identifying one's handwriting with handwriting analysis. *Expert Syst. Appl.*, 79, C, 236–243.

[5] Dantcheva, A., Elia P., Ross, A. (2016). What Else Does Your Biometric Data Reveal? A Survey on Soft Biometrics. *IEEE Transactions on Information Forensics and Security*, 11(3), 441–467.

[6] Abouelenien, M., Pérez-Rosas, V., Mihalcea, R., Burzo, M. (2017). Multimodal gender detection. *19th ACM International Conference on Multimodal Interaction (ICMI 2017).* ACM, 302–311.

[7] Seneviratne, S., Seneviratne, A., Mohapatra, P., Mahanti A. (2015). Your Installed Apps Reveal Your Gender and More! *SIGMOBILE Mob. Comput. Commun. Rev.*, 18(3), 55–61.

[8] Cao, L., Dikmen, M., Fu, Y., Huang, T.S. (2008). Gender recognition from body. *16th ACM International Conference on Multimedia (MM '08)*, 725–728.

[9] Guo, G.-D., Mu, G., Fu, Y. (2009). Gender from body: a biologically-inspired approach with manifold learning. *9th Asian Conference on Computer Vision*, 236–245.

[10] Wu, Q., Guo, G. (2014). Gender Recognition from Unconstrained and Articulated Human Body. *The Scientific World Journal*, Article ID 513240.

[11] Lin, F., Wu, Y., Zhuang, Y., Long, X., Xu, W. (2016). Human gender classification: a review. *Int. J. Biom.*, 8, 275–300.

[12] Jalal, A., Kamal, S., Kim, D. (2015). Shape and Motion Features Approach for Activity Tracking and Recognition from Kinect Video Camera. *2015 IEEE 29th International Conference on Advanced Information Networking and Applications Workshops*, 445–450.

[13] Farooq, A., Jalal, A., Kamal, S. (2015). Dense RGB-D Map-Based Human Tracking and Activity Recognition using Skin Joints Features and Self-Organizing Map. *KSII Transactions on Internet and Information Systems*, 9(3), 1856–1869.

[14] Han, H., Otto, C., Liu, X., Jain, A.K. (2015). Demographic estimation from face images: Human vs. machine performance. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37, 1148–1161.

[15] Andreu, Y., García-Sevilla, P., Mollineda, R.A. (2014). Face gender classification: A statistical study when neutral and distorted faces are combined for training and testing purposes. *Image Vis. Comput.*, 32, 27–36.

[16] Farina, G. L., Spataro, F., De Lorenzo, A., Lukaski, H.A. (2016). Smartphone Application for Personal Assessments of Body Composition and Phenotyping. *Sensors*, 16(12), 2163.

[17] Riaz, Q., Vögele, A., Krüger, B., Weber, A. (2015). One Small Step for a Man: Estimation of Gender, Age and Height from Recordings of One Step by a Single Inertial Sensor. *Sensors*, 15(12), 31999–32019.

[18] Scano, A., Chiavenna, A., Malosio, M., Tosatti, L.M. (2017). Kinect V2 Performance Assessment in Daily-Life Gestures: Cohort Study on Healthy Subjects for a Reference Database for Automated Instrumental Evaluations on Neurological Patients. *Applied Bionics and Biomechanics*.

[19] Buffa, R., Mereu, E., Lussu, P., Succa, V., Pisanu, T., Buffa, F., Marini, E.A. (2015). New, Effective and Low-Cost Three-Dimensional Approach for the Estimation of Upper-Limb Volume. *Sensors*, 15, 12342–12357.

[20] Skalski, A., Machura, B. (2015). Metrological Analysis Of Microsoft Kinect In The Context Of Object Localization. *Metrol. Meas. Syst.*, 22(4), 469–478.

[21] Fryar, C.D., Gu, Q., Ogden, C.L., Flegal, K.M. (2016). Anthropometric reference data for children and adults: United States, 2011–2014. *National Center for Health Statistics. Vital Health Stat*, 3(39), 7–9.

[22] Sandygulova, A., Dragone, M., O'Hare, G.M.P. (2014). Real-time adaptive child-robot interaction: Age and gender determination of children based on 3d body metrics. *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, RO-MAN, 826–831.

[23] Yoo, J.H., Hwang, D., Nixon, M.S. (2005). Gender classification in human gait using support vector machine. *Proc. of the 7th international conference on Advanced Concepts for Intelligent Vision Systems ACIVS*, Springer, 5, 138–145.

[24] Lee, L., Grimson, W.E.L. (2002). Gait analysis for recognition and classification. *Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, 155–162.

[25] Makihara, Y., Mannami, H., Yagi, Y. (2010). Gait analysis of gender and age using a large-scale multi-view gait database. *Asian Conference on Computer Vision*, Springer, 440–451.

[26] Cao, L., Dikmen, M., Fu, Y., Huang, T.S. (2008). Gender recognition from body. *Proc. of the 16th ACM international conference on Multimedia*, ACM, 725–728.

[27] Guo, G., Mu, G., Fu, Y. (2009). Gender from body: A biologically-inspired approach with manifold learning. *Asian Conference on Computer Vision*, Springer, 236–245.

[28] Collins, M., Zhang, J., Miller, P., Wang, H. (2009). Full body image feature representations for gender profiling. *2009 IEEE 12th International Conference on, Computer Vision Workshops (ICCV Workshops), IEEE*, 1235–1242.

[29] Miyamoto, R., Aoki, R. (2015). Gender prediction by gait analysis based on time series variation on joint position. *J. Syst. Cybern. Informatics*, 13, 75–82.

[30] Won, A.S., Yu, L., Janssen, J.H., Bailenson, J.N. (2012). Tracking gesture to detect gender. *Proc. of the International Society for Presence Research Annual Conference*, 24–26.

[31] Adjeroh, D., Cao, D., Piccirilli, M., Ross, A. (2010). Predictability and correlation in human metrology. *2010 IEEE international workshop on Information forensics and security (WIFS)*, 1–6.

[32] Robinette, K.M., Blackwell, S., Daanen, H., Boehmer, M., Fleming, S. (2002). *Civilian American and European Surface Anthropometry Resource (CAESAR), Final Report. 1. Summary*. Sytronics Inc. Dayton Oh.

[33] Cao, D., Chen, C., Adjeroh, D., Ross, A. (2012). Predicting gender and weight from human metrology using a copula model. *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 162–169.

[34] Kakadiaris, I.A., Sarafianos, N., Nikou, C. (2016). Show me your body: Gender classification from still images. *2016 IEEE International Conference on Image Processing (ICIP)*, 3156–3160.

[35] Andersson, V.O., Amaral, L.S., Tonini, A.R., Araujo, R.M. (2015). Gender and Body Mass Index Classification Using a Microsoft Kinect Sensor. *FLAIRS Conference*, 103–106.

[36] Loomis, A. (1943). *Figure drawing for all it's worth*. Viking Pr.

[37] Zhang, Z. (2012). Microsoft Kinect Sensor and Its Effect. *IEEE MultiMedia*, 19(2), 4–10.

[38] Skeleton Position and Tracking State. *https://msdn.microsoft.com/en-us/library/jj131025.aspx*. (Jun. 2017)

[39] Tsai, C.Y., Huang, C.H., Tsao, A.H. (2016). Graphics processing unit-accelerated multi-resolution exhaustive search algorithm for real-time keypoint descriptor matching in high-dimensional spaces. *IET Comput. Vis.*, 10, 212–219.

[40] Bui, L., Tran, D., Huang, X., Chetty, G. (2012). Face Gender Classification Based on Active Appearance Model And Fuzzy k-Nearest Neighbors. *The 2012 International Conference on Image Processing, Computer Vision, and Pattern Recognition*, CSREA Press, Las Vegas, USA, 17, 617–621,

[41] Çamalan, S., Sengül G. (2016). Gender prediction by using Local Binary Pattern and K Nearest Neighbor and Discriminant Analysis classifications. *2016 24th Signal Processing and Communication Application Conference (SIU)*, 2161–2164.

[42] Cortes, C., Vapnik, V. (1995). Support-vector networks. *Mach. Learn.*, 20, 273–297.

[43] Signoretto, M., Suykens, J.A.K. (2015). Kernel Methods. *Kacprzyk J., Pedrycz W. (eds) Springer Handbook of Computational Intelligence*. Springer, Berlin, Heidelberg.

[44] Munsell, B.C., Temlyakov, A., Qu, C., Wang, S. (2012). Person identification using full-body motion and anthropometric biometrics from kinect videos. *European Conference on Computer Vision*, Springer, 91–100.

[45] Gianaria, E., Grangetto, M., Lucenteforte, M., Balossino, N. (2014). Human classification using gait features. *International Workshop on Biometric Authentication*, Springer, 16–27.

[46] Jaswante, A., Khan, A.U., Gour, B. (2014). Back Propagation Neural Network Based Gender Classification Technique Based on Facial Features. *Int. J. Comput. Sci. Netw. Secur.*, 14, 91.

[47] Rudra, S., Mitra, S., Das, S., Roy, A., Guha, S., Seal, D.B., Mukherjee, S., Chatterjee, S. (2016). Gender classification system from offline survey data using neural networks. *Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), IEEE Annual*, 1–5.

[48] Hu, M., Zheng, Y., Ren, F., Jiang, H. (2014). Age estimation and gender classification of facial images based on Local Directional Pattern. *Cloud Computing and Intelligence Systems (CCIS), IEEE 3rd International Conference*, 103–107.

[49] Basha, A.F., Jahangeer, G.S.B. (2012). Face gender image classification using various wavelet transform and support vector machine with various kernels. *Int. J. Comput. Sci.*, 9, 150–157.