ACTA TECHNICA CORVINIENSIS – BULLETIN OF ENGINEERING
TOME XVI [2023] | FASCICULE 2 [APRIL – JUNE]

[1.]Adnan RAMAKIĆ, [2.]Dusanka BUNDALO, [3.]Zlatko BUNDALO

AN APPROACH TO GAIT RECOGNITION USING DEEP NEURAL NETWORK

[1.] Technical Faculty, University of Bihać, Bihać, BOSNIA & HERZEGOVINA
[2.] Faculty of Philosophy, University of Banja Luka, Banja Luka, BOSNIA & HERZEGOVINA
[3.] Faculty of Electrical Engineering, University of Banja Luka, Banja Luka, BOSNIA & HERZEGOVINA

Abstract: In this paper, an analysis was performed in the context of gait recognition. Gait recognition is an approach for people identification based on the gait characteristics of a person. Gait recognition methods essentially belong to the behavioral biometric methods. While walking, each person creates different patterns that may be used for the purpose of identification. For this reason, many gait recognition methods have been presented in recent years that use gait in different ways for the purpose of identification. Some of the presented methods were based on the silhouettes of a person, while others used different models based, for example, on different measurements of the human body. Following the above, the approach of gait recognition was analyzed, focusing on some important aspects of this type of identification. In addition, the use of a Deep Neural Network (DNN) for gait recognition was investigated.

Keywords: Biometric Methods, Deep Neural Network (DNN), Gait Energy Image (GEI), Gait Recognition, People Identification

INTRODUCTION

People identification is an important process in many aspects of human life. Most people are confronted with some kind of identification in the course of their lives, e.g., at airports, when crossing borders, with various security systems, e.g., to gain access to various facilities, etc. For this reason, different systems of identification have been introduced, using different methods for the purpose of identification. The mentioned methods are usually based on features related to and extracted from human body characteristics, i.e., physiological characteristics. In addition, behavioral characteristics of individuals are also used. The methods based on the mentioned characteristics (the so-called biometric characteristics) and the features extracted from them are called biometric methods. Accordingly, there are physiological and behavioral biometric methods.

Some examples of physiological biometric methods are fingerprint, methods based on eye features extracted from parts of the eye such as the iris and retina, methods using facial features, hand features, etc. Some examples of behavioral biometric methods are methods that use a person's voice, methods based on gait analysis, methods based on keystroke dynamics, signatures, etc. It is important to note that some methods are more reliable than others. For example, a method based on iris features is more reliable than a method based on keystroke dynamics. In general, the use of a particular biometric characteristic depends on the application. Each of the mentioned biometric characteristics has advantages and disadvantages, and it is not possible that one biometric characteristic satisfy all applications [9].

In this paper, gait recognition for people identification was analyzed. Gait recognition is a method based on the analysis of a person's gait. Gait is a behavioral biometric characteristic of a person, and various features and methods based on them have been presented in recent years. In the following chapters of this work, different aspects of gait recognition and some well-known gait recognition methods have been analyzed. Also, three experiments were conducted using Deep Neural Network (DNN) and results were presented.

GAIT RECOGNITION

Gait recognition is a method for people identification based on gait analysis. When walking, each person creates different patterns that may be used for identification. For this reason, two approaches to gait recognition are in use: the appearance-based approach and the model-based approach. An appearance-based approach usually uses the silhouettes of people or parts of them, i.e. different representations of the silhouettes. A model-based approach is usually based on a model that uses different measurements of the human body, e.g., the length of the legs or arms [18].

With the introduction of the Kinect device by Microsoft in 2010 [14], more attention began to be paid to gait recognition as a method for people identification. Microsoft Kinect is a device that can be used in conjunction with Microsoft's Xbox console for interaction without the need to use an intermediate device, such as a controller. It contains an RGB (Red, Green, Blue) camera, depth sensor, and microphone array. It makes it possible to obtain RGB and depth images, as well as additional features such as skeleton information. Later, Microsoft also introduced a version of Kinect for Windows with a corresponding software development kit (SDK). After that, Microsoft introduced the second generation of Kinect, but today the production of Kinect is not continued. It has been replaced by Azure Kinect. Microsoft Kinect for Xbox 360 console is shown in Figure 1.

This makes the Kinect device well suited for gait recognition tasks, as it is possible to implement methods based on both approaches to gait recognition, model-based and appearance-based. Using the skeleton feature provided by Kinect, many model-based gait recognition methods have been implemented, while with the availability of RGB and depth images, many appearance-based gait recognition methods have been implemented because the person silhouettes required for appearance-based approaches can be obtained from RGB and depth images. Figure 2 shows an example of the steps involved in implementing a gait recognition system and this can be realized as follows.

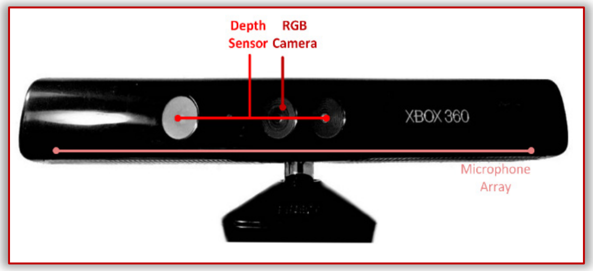

Figure 1. Microsoft Kinect for Xbox 360 Console

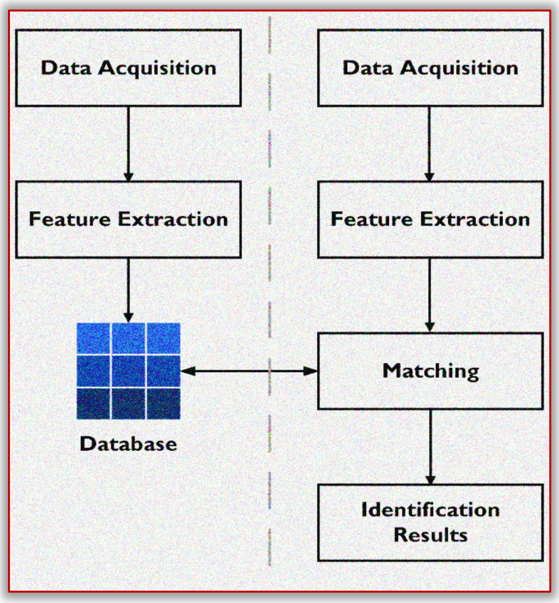

Figure 2. Gait Recognition System (Steps)

Some gait recognition system can be roughly divided into two parts. The first part is the creation of a database, and the second part is an identification part. This is illustrated in Figure 2, where the left side of the figure is related to the creation of a database, while the right part of the figure is an identification part. In the first part, features extracted from silhouette images, for example, must be stored for each person. In Figure 2 *Data Acquisition* refers to the acquisition of RGB or depth images, e.g., with Microsoft Kinect. Also, in Figure 2 may be added some additional steps between *Data Acquisition* and *Feature Extraction* related to data preparation, image processing, etc., but broadly the same as in Figure 2. *Feature Extraction* refers to the process of extracting features related to each person. This depends on the implementation and the defined type of features to be used in the gait recognition system. The database (*Database*) contains the features defined and extracted for each person.

The second part is the identification part (Figure 2, the right side of the figure), which consists of the data acquisition (*Data Acquisition*), the feature extraction (*Feature Extraction*), the matching of the features (*Matching*) and at the end the results of the matching, i.e. the identification results (*Identification Results*). An identification process works as follows. If there is a person to be identified, it is necessary to capture RGB or depth images with a device such as, e.g., Kinect. Depending on the implemented method, the features should be extracted, for example, from the silhouettes of a person. An example of the extracted silhouettes (unprocessed) from the RGB images for one person is shown in Figure 3. After the features are extracted, they should be matched with the features stored in the database. The best match between the extracted features and the features stored in the database results in the identified person.

Figure 3. The Extracted Silhouettes for One Person

In the previous text, an example scenario was described in which the system and the implemented method for gait recognition work with the silhouette of the person (the appearance-based approach). Similarly, can be implemented and some other system that uses some other elements for obtaining the features, for example, some measurements from the human body (the model-based approach). The steps shown in Figure 2 are essentially the similar for the most gait recognition systems.

Today, most gait recognition systems are implemented using machine learning and deep learning, where a model is created and trained using a deep neural network or some type of classifier. In recent years, many methods for gait recognition have been presented. The presented methods were based on either a model-based or an appearance-based approach. One of the most popular methods is the Gait Energy Image (GEI) [6]. GEI was defined as an image containing the silhouettes of a person over a gait cycle, normalized, aligned, and temporally averaged. Based on GEI, some other methods such as Backfilled Gait Energy Image (BGEI) [22] have also been developed. BGEI is similar to GEI, where silhouettes are filled from the foremost pixels to the back of the image. Also, other interesting presented methods are HGEI-i and HGEI-f [12] [20], where fusion of the information between GEI features and height of a person feature has been done. Gait Gaussian Image (GGI) [1] is a period-based gait recognition method intended for feature extraction from gait images over a gait cycle. An interesting approach presented in [8] divides the human body image into areas and then extracts features for each area. Gait Energy Volume (GEV) [21] extends the concept of GEI to 3D. Some other interesting methods for gait recognition can be found in [7] [3] [17] [19] [11].

The main advantage of gait recognition as a method of person identification is the possibility of person identification at a long distance. A stereo camera with a long range can be used for this purpose. Also, no interaction with a person to be identified is required. This means that a particular person can be identified without knowing that the identification process is in progress. It is important to note that gait is not as unique a human characteristic as, for example, fingerprint or iris, so the use of gait recognition methods often requires some additional features along with gait features. This makes implementation more complex and harder to implement.

EXPERIMENTAL SETUP

Three experiments were performed using a well-known gait recognition method called Gait Energy Image, or as it is acronym GEI. GEI was introduced by authors Han and Bhanu [6] and represents an image with multiple silhouettes of a person over a gait cycle that are normalized, aligned, and temporally averaged. Casia Dataset B [24] [23] [15] was used for experimental evaluation. Casia Dataset B is a gait database that contains 124 subjects with gait data obtained from 11 views and with three variations - viewing angle, clothing, and carrying condition changes [15]. Some examples of GEI images from Casia Dataset B [24] [23] [15] are shown in Figure 4. For all three experiments, 100 subjects were used.

(a) Normal gait (90 degrees viewing angle)

(b) Clothing changes (90 degrees viewing angle)

(c)Carrying changes (90 degrees viewing angle)

(d) Normal gait (180 degrees viewing angle)

Figure 4. Examples of GEI images from Casia Dataset B [24] [23] [15]

In the first experiment, GEI images with a viewing angle of 90 degrees were used for each of the 100 persons. In total, there were only six images in normal gait for each person. Thus, a total of 600 images. In the second experiment, GEI images with the viewing angle of 90 degrees were used for each of the 100 persons. In total, there were 10 images for each person (six images with the person in normal gait, two images with carrying condition changes, and two images with clothing changes). Thus, a total of 1.000 images. In the third experiment, GEI images with different viewing angle, carrying condition changes, and clothing changes were used for each of the 100 person. For each person, there were 110 GEI images (66 images with the person in normal gait, 22 images with carrying condition changes, and 22 images with clothing changes). A total of 11.000 images.

For the experiments, a deep neural network in Matlab was created. The neural network created consists of seven layers, the first layer being the feature input layer (*featureInputLayer*). Mentioned layers are *featureInputLayer*, *fullyConnectedLayer*, *batchNormalizationLayer*, *reluLayer*, *fullyConnectedLayer*, *softmaxLayer*, and *classificationLayer*.

Since the neural network works with features, it is necessary to obtain features from defined GEI images for each person. For this purpose, a bag of visual words (*bagOfFeatures* in Matlab, with parameters *VocabularySize* (500) and *PointSelection* as *Detector*) [4] [13] was used, where the visual vocabulary is created by default from Speeded-Up Robust Features (SURF) [2]. The features obtained were stored in a table.

Feature data were divided into a training and a test part in a ratio of 85% for training and 15% for testing. Other training options include 30 epochs, an initial learning rate of 0.01 (the first and second experiment) and 0,001 (the third experiment). And the Adaptive Moment Estimation Optimizer (Adam) [10] was used.

In addition to the Deep Neural Network, two classifiers were also used. The classifiers mentioned are Support Vector Machines (SVM) [5] and k-Nearest Neighbors (kNN) [16]. The same features, as in case of Deep Neural Network, were used for SVM and kNN classifiers, in the same ratio of 85% for training and 15% for testing.

RESULTS AND DISCUSSION

With certain experimental settings, the following results were obtained. In the first experiment, the accuracy was 90% in the case of the Deep Neural Network, using 600 images (six images for each person). For the two classifiers, the accuracy was 94.1% in the case of SVM and 84.9% in the case of kNN. Table 1 and Figure 5 show the results of the first experiment.

Table 1. The Obtained Results for the First Experiment

THE FIRST EXPERIMENT	
The Method Used	Accuracy
Deep Neural Network	90%
SVM Classifier	94.1%
kNN Classifier	84.9%

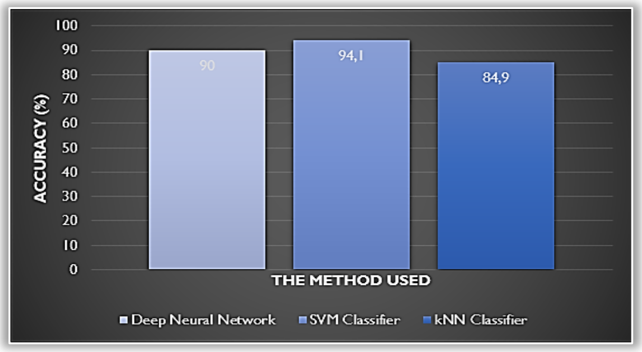

Figure 5. The Obtained Results for the Methods Used (The First Experiment)

In the case of the Deep Neural Network and the defined settings for the second experiment, the accuracy was 84.7%. In the second experiment, there were 1.000 images, for each of 100 subjects 10 GEI images. For the classifiers used, the accuracy was about 74.8% in the case of SVM and 67.4% in the case of kNN. The obtained results, in terms of accuracy, for the second experiment are shown in Table 2 and Figure 6.

Table 2. The Obtained Results for the Second Experiment

THE SECOND EXPERIMENT	
The Method Used	Accuracy
Deep Neural Network	84.7%
SVM Classifier	74.8%
kNN Classifier	67.4%

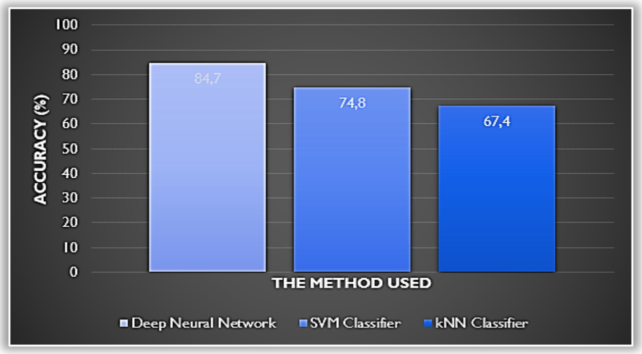

Figure 6. The Obtained Results for the Methods Used (The Second Experiment)

In the third experiment, slightly different settings were used and the learning rate was 0,001 instead of 0.01 as in the first and second experiment. The reason for changing the settings was better overall results. Also, in the third experiment, there were 11.000 images, 110 GEI images for each of the 100 subjects. In the case of the Deep Neural Network, the results in terms of accuracy were 54.7%. For the classifiers used, the accuracy was 54.2% for SVM and 48.8% for kNN. The obtained results, in terms of accuracy, for the third experiment are shown in Table 3 and Figure 7.

Table 3. The Obtained Results for the Third Experiment

THE THIRD EXPERIMENT	
The Method Used	Accuracy
Deep Neural Network	54,7%
SVM Classifier	54,2%
kNN Classifier	48,8%

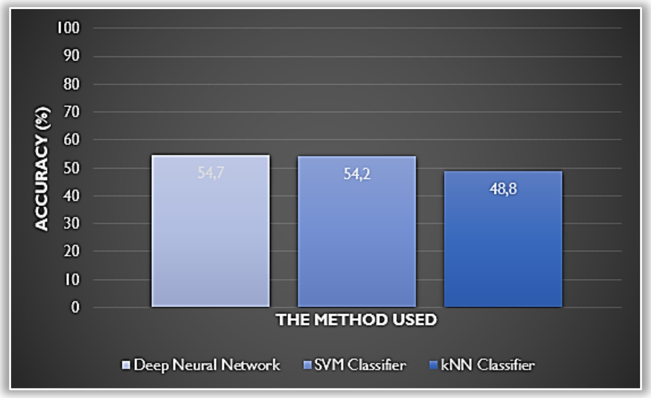

Figure 7. The Obtained Results for the Methods Used (The Third Experiment)

The comparison of the obtained results for the three performed experiments is shown in Figure 8.

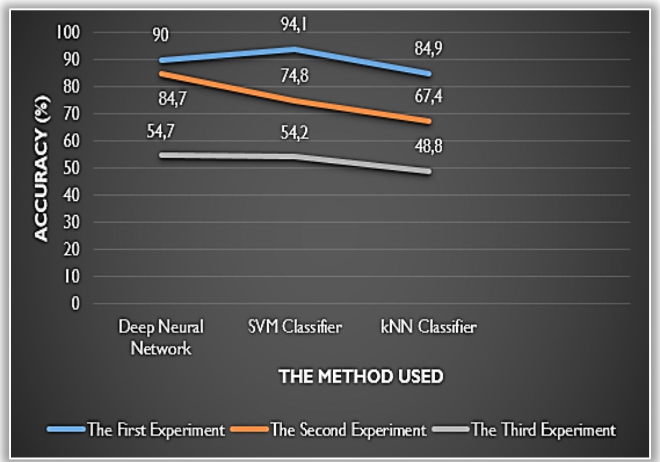

Figure 8. Comparison of the Results for the Performed Experiments

In the first experiment, the Deep Neural Network achieved the second highest result (accuracy) of 90% for defined experimental settings and the learning rate of 0,01. The SVM classifier achieved the highest result of 94.1%. Somewhat lower was the result for kNN with 84.9%. In the second experiment, the Deep Neural Network had the highest accuracy of 84.7%. This was achieved with defined experimental settings where the learning rate was 0.01. Compared to the Deep Neural Network, SVM and kNN classifiers had lower accuracy of 74.8% and 67.4%, respectively. In the third experiment, the Deep Neural Network achieved the highest results of 54.7% with learning rate of 0,001. Also, the two classifiers used achieved 54.2% in the case of SVM and 48.8% in the case of kNN.

In two out of three experiments, the Deep Neural Network achieved the best results compared to the classifiers used. In the experiment with a smaller number of images (the first experiment), the SVM classifier achieved better results. With increasing number of images for each subject and overall, the Deep Neural Network achieved the best results. It is important to note that the duration of the training process for the Deep Neural Network was several minutes, using standard laptop, in the case of the third experiment, where a large number of images were used.

The GEI images were also used as input instead of features. Instead of the feature input layer (*featureInputLayer*), image input layer (*imageInputLayer*) with customized options was used. More specifically, a Convolutional Neural Network (CNN) was created and used. The settings for the experiments with CNN were similar to those in the case of using extracted features (for the three experiments described above), with 85% of the images used for training and 15% for testing. In this case (for used GEI images), the results for defined experiments were significantly lower in terms of accuracy and the duration of the training process was much longer (several hours in the case of the third experiment).

It is important to note that besides SVM and kNN classifiers, other classifiers were also used (such as Tree) and the results obtained were lower compared to SVM and kNN classifiers and also compared to Depp Neural Network.

CONCLUSION

In this paper, it was analyzed gait recognition for people identification. Gait recognition is one of the behavioral biometric methods that analyze the features of gait and use them for purpose of identification, re-identification etc. During a gait cycle, each person forms certain patterns that can be utilized in purpose of identification. In recent years, different methods have been presented that rely on a model-based or appearance-based approach, meaning that are based on some model or on some person's silhouettes representation.

In this paper, a deep neural network was developed for use with a well-known gait recognition method called Gait Energy Image or GEI. Also, the Casia Dataset B for experimental evaluation was used. Three experiments were defined using 100 subjects from the Casia Dataset B in different ways. An accuracy of about 90% was achieved in certain settings. Also, two classifiers, kNN and SVM, were used for comparison with created Deep Neural Network, which achieved slightly lower results in the two out of three experiments.

References

[1] Arora, P. and Srivastava, S. (2015). Gait Recognition Using Gait Gaussian Image. In: 2nd International Conference on Signal Processing and Integrated Networks (SPIN), 791-794. IEEE.
[2] Bay, H., Ess, A., Tuytelaars, T. and Van Gool, L. (2008). Speeded-Up Robust Features (SURF). Computer Vision and Image Understanding, 110(3), 346-359. Elsevier.
[3] Chattopadhyay, P., Roy, A., Sural, S. and Mukhopadhyay, J. (2014). Pose Depth Volume Extraction from RGB-D Streams for Frontal Gait Recognition. Journal of Visual Communication and Image Representation, 25(1), 53-63. Elsevier.
[4] Csurka, G., Dance, C., Fan, L., Willamowski, J. and Bray, C. (2004). Visual Categorization with Bags of Keypoints. In: Workshop on Statistical Learning in Computer Vision (ECCV), 1-2.

[5] Gunn, S. R. (1998). Support Vector Machines for Classification and Regression. ISIS Technical Report, 14(1), 5-16.
[6] Han, J. and Bhanu, B. (2005). Individual Recognition Using Gait Energy Image. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(2), 316-322. IEEE.
[7] Hofmann, M., Bachmann, S. and Rigoll, G. (2012). 2.5D Gait Biometrics Using the Depth Gradient Histogram Energy Image. In: 5th International Conference on Biometrics: Theory, Applications and Systems (BTAS), 399-403. IEEE.
[8] Iwashita, Y., Uchino, K. and Kurazume, R. (2013). Gait-based Person Identification Robust to Changes in Appearance. Sensors, 13(6), 7884-7901. MDPI.
[9] Jain, A. K., Ross, A. and Prabhakar, S. (2004). An Introduction to Biometric Recognition. IEEE Transactions on Circuits and Systems for Video Technology, 14(1), 4-20. IEEE.
[10] Kingma, D. P. and Ba, J. (2014). Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980.
[11] Kumar, M. N. and Babu, R. V. (2012). Human Gait Recognition Using Depth Camera: A Covariance Based Approach. In: Proceedings of the 8th Indian Conference on Computer Vision, Graphics and Image Processing, 1-6.
[12] Lenac, K., Sušanj, D., Ramakić, A. and Pinčić, D. (2019). Extending Appearance Based Gait Recognition with Depth Data. Applied Sciences, 9(24), 5529. MDPI.
[13] Nister, D. and Stewenius, H. (2006). Scalable Recognition with a Vocabulary Tree. In: Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2161-2168. IEEE.
[14] Official Web Page of Microsoft - Kinect Device. Link: https://learn.microsoft.com/en-us/windows/apps/design/devices/kinect-for-windows [Accessed 05/1/2023]
[15] Official Web Page of the Institute of Automation, Chinese Academy of Sciences. Link: http://www.cbsr.ia.ac.cn/english/Gait\%20Databases.asp [Accessed 05/1/2023]
[16] Peterson, L. E. (2009). K-Nearest Neighbor. Scholarpedia, 4(2), 1883.
[17] Preis, J., Kessel, M., Werner, M. and Linnhoff-Popien, C. (2012). Gait Recognition with Kinect. In: 1st International Workshop on Kinect in Pervasive Computing, 1-4. New Castle, UK.
[18] Ramakić, A. and Bundalo, Z. (2023). Gait Recognition as an Approach for People Identification. In: International Symposium on Innovative and Interdisciplinary Applications of Advanced Technologies, 717-726. Springer.
[19] Ramakić, A., Bundalo, Z. and Bundalo, D. (2020). A Method for Human Gait Recognition from Video Streams Using Silhouette, Height and Step Length. Journal of Circuits, Systems and Computers, 29(7), 2050101. World Scientific.
[20] Ramakić, A., Sušanj, D., Lenac, K. and Budalo, Z. (2020). Depth-based Real-time Gait Recognition. Journal of Circuits, Systems and Computers, 29(16), 2050266. World Scientific.
[21] Sivapalan, S., Chen, D., Denman, S., Sridharan, S. and Fookes, C. (2011). Gait Energy Volumes and Frontal Gait Recognition Using Depth Images. In: International Joint Conference on Biometrics (IJCB), 1-6. IEEE.
[22] Sivapalan, S., Chen, D., Denman, S., Sridharan, S. and Fookes, C. (2012). The Backfilled GEI-A Cross-capture Modality Gait Feature for Frontal and Side-view Gait Recognition. In: International Conference on Digital Image Computing Techniques and Applications (DICTA), 1-8. IEEE.
[23] Yu, S., Tan, D. and Tan, T. (2006). A Framework for Evaluating the Effect of View Angle, Clothing and Carying Condition on Gait Recognition. In: 18th International Conference on Pattern Recognition (ICPR), 441-444. IEEE.
[24] Zheng, S., Zhang, J., Huang, K., He, R. and Tan, T. (2011). Robust View Transformation Model for Gait Recognition. In: 18th International Conference on Image Processing, 2073-2076. IEEE.

ISSN: 2067-3809