



ISSN: 0067-2904

Gait Recognition Based on Deep Learning

Humam Khaled Jameel, Ban Nadeem Dhannoon

Department of Computer Science, College of Science, AL-Nahrain University, Baghdad, Iraq

Received: 28/10/2020

Accepted: 20/4/2021

Abstract

In current generation of technology, a robust security system is required based on biometric trait such as human gait, which is a smooth biometric feature to understand humans via their taking walks pattern. In this paper, a person is recognized based on his gait's style that is captured from a video motion previously recorded with a digital camera. The video package is handled via more than one phase after splitting it into a successive image (called frames), which are passes through a preprocessing step earlier than classification procedure operation. The pre-processing steps encompass converting each image into a gray image, cast off all undesirable components and ridding it from noise, discover difference between two successive images to discover the place motion occurs, converting the result to a binary image, and finally use morphological operation to close holes resulted from the previous steps. The last and most important stage in the system is the classification stage, which depends on deep neural network. The results obtained indicate a high quality of performance and an accuracy of 99.5%.

Keywords: Deep learning, CNN, Segmentation, classification, neural network, recognition.

التعرف على المشي على اساس التعلم العميق

همام خالد جميل، بان نديم ذنون
جامعة النهرين، كلية العلوم، قسم علوم الحاسوب

الخلاصه

في الجيل الحالي من التكنولوجيا، هناك حاجة إلى نظام أمان قوي يعتمد على السمات الحيوية (البايومترية) مثل مشية الإنسان، وهي مقياس حيوي سلس يساعد على فهم البشر من خلال أسلوب المشي. في هذه البحث، يتم التعرف على الشخص بناءً على أسلوب مشيته التي تم التقاطها من حركة فيديو تم تسجيلها مسبقاً باستخدام كاميرا رقمية. يتم التعامل مع حزمة الفيديو عبر أكثر من مرحلة بعد تقسيمها إلى صور متتالية (تسمى الإطارات)، والتي يتم تمريرها من خلال خطوات المعالجة المسبقة قبل عملية إجراء التصنيف. تشمل خطوات المعالجة المسبقة تحويل كل صورة إلى صورة رمادية، التخلص من جميع المكونات غير المرغوب فيها وتخليصها من الضوضاء، اكتشاف الفرق بين صورتين متتاليتين للحصول على المكان الذي تتراكم فيه الحركة، وتحويل النتيجة إلى صورة ثنائية، وأخيراً استخدم العملية المورفولوجية لإغلاق الثقوب الناتجة عن الخطوات السابقة. المرحلة الأخيرة والأكثر أهمية في النظام هي مرحلة التصنيف التي تعتمد على الشبكة العصبية العميقة. النتائج التي تم الحصول عليها تشير إلى جودة أداء عالية ودقة 99.5%.

1. Introduction

Researches in area of computerized person identification has been accelerated in recent years. Biometric systems provide a sturdy and dynamic performance as a verification system.

The verification overall performance of biometric machines will become more dependable in contrast to person verification [1].

Gait recognition is one of the most outstanding methods of biometric. Gait is the most promising trait of biometric which having various actual time application. This biometric trait used for verification of person based totally on fashion of human and strolling nature [2]. Biometrics which include iris, face, and fingerprint can be blended with gait recognition. Gait awareness has a blessing that verification and awareness is viable from a far distance. Gait recognition has blessing that human verification is viable from large distance degree [3]. This gait recognition is used in verification of character from video surveillance and improvement of smart video surveillance system. Sensible surveillances gadget is utilized for security and forensic primarily based applications. In 2016, Zhang, C et al. proposed and carried out a gait focus framework primarily based on Siamese neural network. The proposed gadget extracts the aspects for implementation of human verification system. The Siamese community is utilized for distance metric gaining knowledge to power the similarity metric to be small for pairs of gaits from the equal person, and giant for pairs from extraordinary persons [4]. In 2018, Wang, X. et al. proposed gait attention primarily based on novel Gabor wavelet. The proposed paper is carried out in three steps. First step, extract different orientation and scale information from Gabor wavelet. Second step, a two-dimensional principal component analysis method is employed to reduce feature space dimension by minimizing in-category distance and maximize between category distances. Third step, a multi-class support vector machine is adopted to recognize different gaits [5]. Another work in 2018 is proposed by Battistone F. et al. to dynamically learn graphs by using Long Short-Term Memory network. Experiments were made for popular gait recognition datasets, investigating the advantages of the proposed method with respect to state-of-the-art methods [6]. This research aims to design and build a robust recognition system for a collected data set, and provides person authentication by investigating the advantages of deep neural network.

2. Dataset

A real-life video dataset was proposed and collected in this research for person identification and authentication based on his/her gait. Where nearly eighteen videos were captured, taken for eight different individuals, with more than one direction for each person, so that there are approximately four full gait cycles in each video. Video durations were different in terms of time. The number of frames that were used in the collected video dataset was 6100 frames.

3. The Proposed System

The design of any system is very important due to; it shows how the system is working and explains the exact steps and process that will be performed to get the desired need from it. The proposed recognition system can detect human and recognize him/her based on his/her captured image through a camera with a specific feature. SONY HXR camera model NX 200 with 4K capturing ability or FULL HD video accuracy were used to record videos and capture images. The used camera lens has three rings to control focus, aperture, and zoom as well as it has a property of image stabilizing. The collected images were saved into a dataset and used to build a classifier to make it easy in the recognizing process. Figure 1 shows phases of the proposed system that were used to recognize and detects person based on his/her walk and, the classifier building steps sequentially.

3.1 Preprocessing

Successive input images (frames) came from video require to be prepared and improved by diminishing any insignificant data to make it ready for following phase. At first, these images which are represented by three dimensional RGB bands are converted to gray scale, see Figure 2. This conversion is required due to the fact it reduces entered image graph facts from three channels to one. Two dimensional enter image graph named as gray scale image. Gray scale picture improves overall performance of verification and time for gadget execution.

Depth of gray scale image is saved in a structure of eight-bit integer which indicates the 256 special probabilities of gray scale.

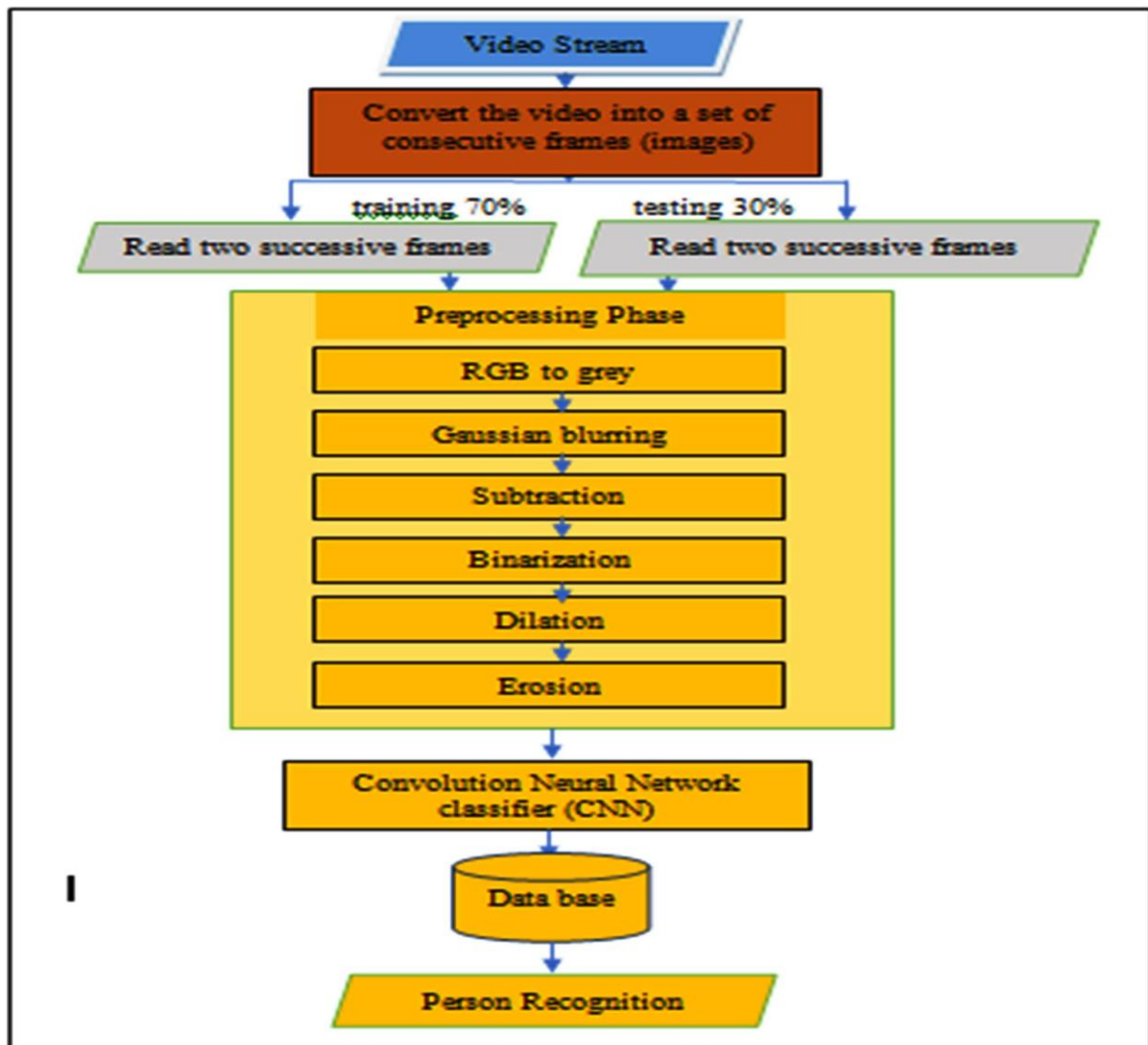


Figure 1- Flowchart of the proposed system



Figure 2- Resulted grayscale image.

A. Blurring the image based on Gaussian filter

Filtering process involves getting rid of as much noise as possible while losing minimum information. Gaussian filtering is used to blur images, removes noise, and unwanted details. Gaussian smoothing is very effective due to it is a linear operation as well as, weights give higher significance to pixels near edge. In this paper, Gaussian blurring filter is applied into grey image with kernel value (5*5) as illustrated in algorithm (1).

Choosing a suitable kernel size is very important because whenever size of kernel mask becomes bigger, more noise will be removed but its side effect can be seen in removing some details in the image like object boundaries and details that they are very important in detection and recognition of objects in the image. The resulted blurred image after applying the Gaussian mask on the inputted greyscale image resulted from the previous operation is shown in Figure 3.

Algorithm (1): Gaussian blurring filter
Input: grayscale image, kernel matrix with size 5*5 Output: B image //smoothed greyscale image Begin B = Greyscale // copy the gray scale image to B N=0 For k= 0 to 4 For L= 0 to 4 N= N+ F(k, L) End for L End for k For i=2 to Grayscale.hight-3 For j=2 to Grayscale.width-3 Image (i,j)=0 For k= -2 to 2 For L= -2 to 2 If k<>0 && L<>0 then // skip the center point Image(i,j) = Image (i,j) + image(i-k, j-L) * F(k+2,L+2) End if End for L End for k B(i,j)= Greyscale (i,j) / N End For j End For i Return smoothed image B End



Figure 3-Blurred image using Gaussian function

B. Image subtraction process

Image subtraction is a mathematical operation that carried out on image by placing a digital value of a pixels in the image subtracted from the digital value of the same pixel in another image. The advantage of subtraction system is to realize the alternate between two images. In this paper, the subtraction is used to seize the place the motion has been brought via subtracting two successive blurred images. Resulted image includes distinction between them which indicates the motion of the pixels. Algorithm (2) describes the subtraction procedure steps, the movement pixels by subtracting the two blurred consecutive images shows in Figure 4.

Algorithm (2): image subtraction
<p>Input: Two greyscale images with equal size</p> <p>Output: SubImage</p> <p>Begin</p> <p>For i=0 to Greyscale.width-1</p> <p> For j=0 to Greyscale. high-1</p> <p> Abs (SubImage(i,j))= Greyscale (i,j)-B(i,j)</p> <p> End for</p> <p>End for</p> <p>Return SubImage</p> <p>End</p>



Figure 4-The resulted image from the subtraction

C. Image binarization process

The image binarization technique is used for separating image pixels into two groups based totally on pixel value. The distinguish manner is performed primarily based on white and black pixel color. The white pixels are utilized as foreground or object pixels, while black pixels are used as background pixels. Last pixel transformation operation will be thresholding method. The global threshold is used in binarization operation which minimizes interclass variance between black and white pixels. Steps of binarization operation are described in algorithm (3) and Figure 5 shows the resulted image after applying binarization process.

Algorithm (3): Image binarization**Input** :SubImage**Output**:BinaryImage**Begin**

thr= 128

For i=0 to SubImage..high-1

For j=0 to SubImage.width-1

If SubImage (i,j)> = thr

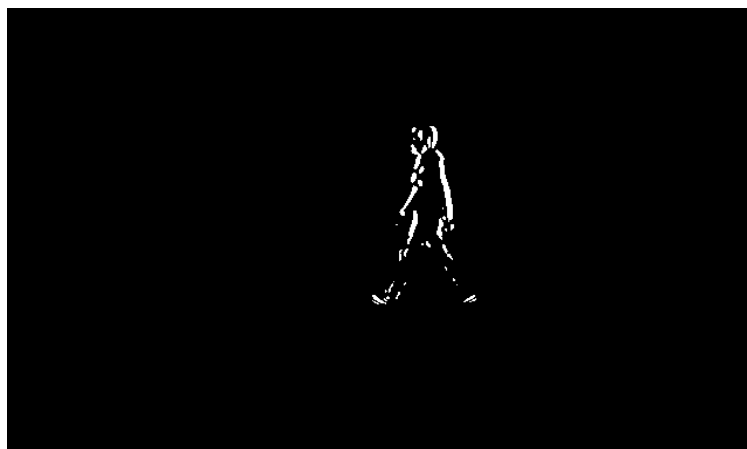
BinaryImage (i,j)=1

Else BinaryImage (i,j)=0

End if

End for j

End for i

End**Figure 5**-Resulted image from applying binary process**D. Image dilation process**

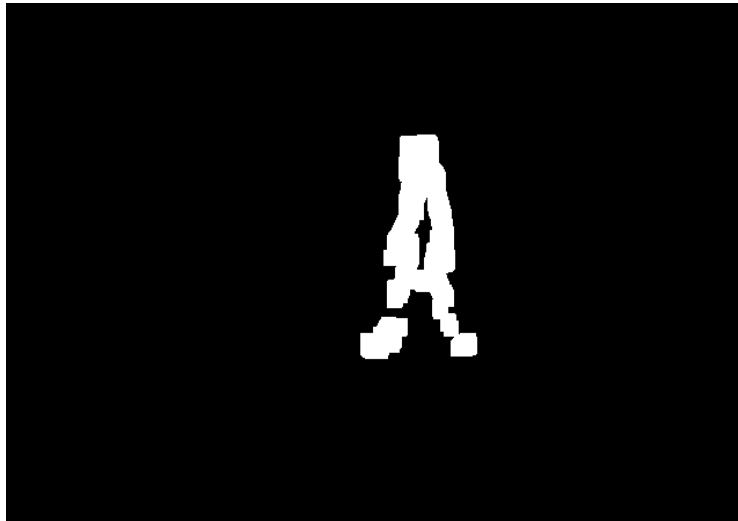
The dilation process is a fundamental step of preprocessing and used to form totally function extraction. Based on this operation, structure and morphology based characteristics of images will be extracted. This operation relying upon relative ordering, not on their manual. It adds a layer of pixels to both the inner and outer boundaries of region values. Accordingly, it closes holes and gaps between different regions which become smaller. The implementation steps are described in algorithm (4). Figure 6 indicates resulted image after making use of two instances dilation procedures.

Algorithm (4): Binary image dilation

```

Input: BinaryImage
Output: DilatedImage
Begin
Set a window (3x3)
For i=1 to BinaryImage.hight-2
  For j=1 to BinaryImage.width-2
    DilatedImage(i,j) =0
    For K= -1 to 1
      For L = -1 to 1
        If BinaryImage(i+K, j+L) ==1 then
          DilatedImage(i,j) =1
          L=2
          K=2
        End If
      End for L
    End for k
  End For j
End For i
End

```

**Figure 6-** Outcome of dilated image**E. Erosion process**

It is a non-linear operation related to shape or morphology of features in an image. Erosion procedure will shrink person's body in the image which is considered the object in that image by removing its boundaries as well as, holes and gaps between different regions which become larger, and small details are eliminated. The erosion process steps are shown in the algorithm (5). Figure 7 shows resulted images from the erosion procedure.

Algorithm (5): Binary image Erosion

```

Input: DilatedImage
Output: ErodedImage
Begin
Set a window (3x3)
For i=1 to DilatedImage.hight-2
  For j=1 to DilatedImage.width-2
    ErodedImage(i, j) =0
    For K= -1 to 1
      For L = -1 to 1

```



```

        If DilatedImage(i+K, j+L) ==0 then
            ErodedImage(i,j) =0
            L=2
            K=2
        End If
    End for L
End for k
End For j
End For i
End

```

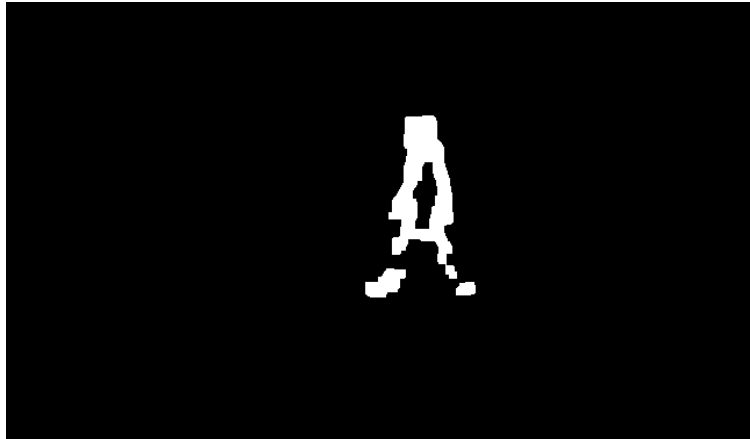


Figure 7- Outcome of eroded image

G. Detecting human body process

This process is very important to speedup classification phase. Instead of entering whole image to the classifier, a detected one will be entered, so size of input neurons will be reduced. Algorithm (6) shows body detection process, while Figure 8 shows resulted image of the body detection process.



Figure 8- Resulted image of body detection process.

<p>Algorithm (6): Detecting a body</p> <p>Input: ErodedImage</p> <p>Output: L1, L2, K1, K2 // left upper and right bottom of the boundary box pixels applied on original image</p> <p>Begin</p> <p>Flag=F</p> <p>j=0</p> <p>For i=0 to ErodedImage.hight-1 // detect the left upper box</p> <p> For j=0 to erodedImage.width-1</p>

```

    If ErodedImage(i, j) ==1
        L1=i
        K1=j
        j= erodedImage.width+5
    End if
    End For j

        // detect the right bottom box
    For j= erodedImage.width-1 to 0 step -1
        If ErodedImage(i, j) ==1
            K2=j
            j= -2
        End If
        Else if j == 0
            L2=i
        End if
    End For j
    End For i
End

```

4 Convolutional Neural Networks Classifier

Convolutional neural network is one of main categories to do images recognition. It takes an input image, process it and classify it under certain classes. A free open-source deep learning library, Keras written in Python, was used to implement CNN model. Each preprocessed image will pass through a series of convolution layers with filters (Kernels), Pooling, fully connected configuration layers (FC) to recognize an object with probabilistic values between multi classes.

In our CNN model, initialize a random value for all parameters needed in the CNN. Summary in Table-1.

Table 1-The parameters initialize in current CNN	
Parameter	Value
Activation function	Leaky ReLU, Linear, Softmax
Kernel Size of filters	3
Gradient Descent Optimizer	Adam
Learning rate	0.00001
Stride	1

1. Nine convolutional layers depending on problem complexity with kernel size =3. Eight layers have Leaky ReLU activation function. First layer contains 16 filters. Second layer contains 32 filters. Third layer contains 64 filters. Fourth layer contains 128 filters. Fifth layer contains 265 filters. Sixth layer contains 512. Seventh and eight layers contains 1024 filters, and the last one contains 40 filters.

2. Six max-pooling layers to extract global patterns, each layer contains pool size = 2*2, and strides =1.

3. Fully connected layers: a flattening layer with Softmax activation function to predict whether a source image input was any of Eight classes of human body in database of the system.

As shown in Figure 9, the input image (resulted from previous preprocessing operations) is divided into a matrix $13 * 13$, and each part of the 13 is passed to the CNN for this work, where the input part is divided again into $13 * 13$ continuously until there will be a matrix of $13 * 13$ pixels (cannot be divided). By passing all parts of the image to the CNN layers and until it reaches the output layer, which determines whether the model can recognize the human body and is it within the system's database

The human body classification by convolution neural network is explains in Algorithm (7), CNN trained for this work on our dataset with 6100 record, each record has own attributes.

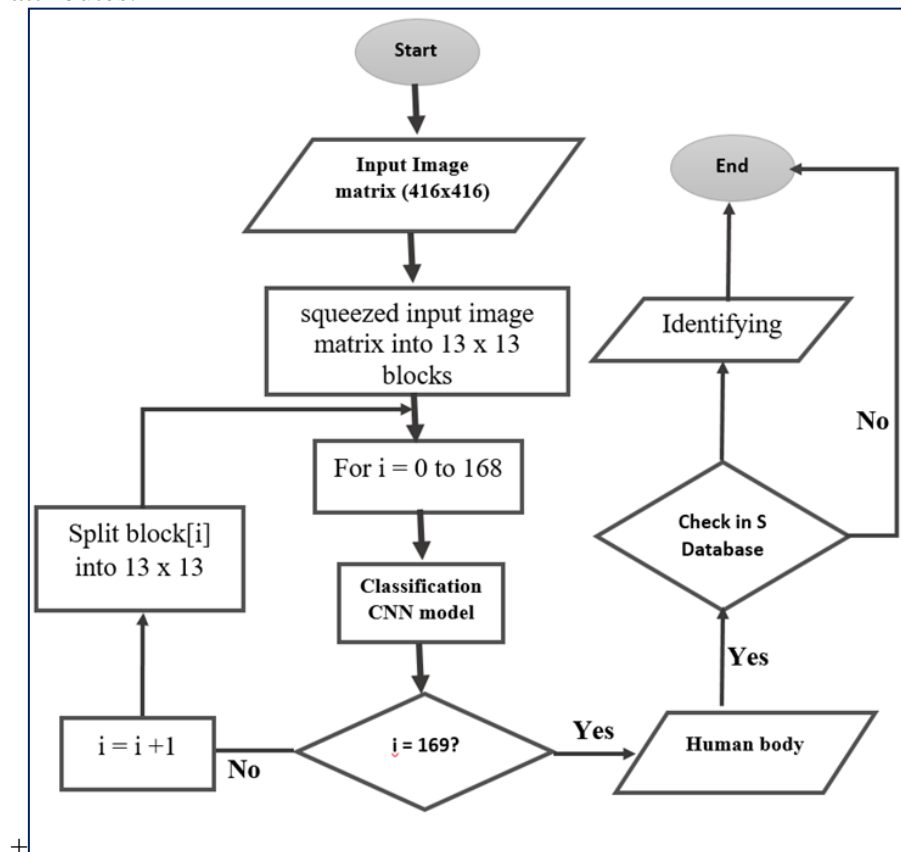


Figure 9-Classification and Identification by CNN model

Algorithm (7) : CNN model for Human body Classification
Input: Image with 416 x 416 from dataset
Output: One of eight classes
<ol style="list-style-type: none"> 1. Initialize a random value for all parameters learning-rate = 0.00001 Max_No_epoch = 1000 (batch's size = 64 for each epoch) Stride size = 1, Pool size = 2x2 2. For input image do <ol style="list-style-type: none"> A. squeezed input image matrix into 13 x 13 blocks B. Parsing and Reformat image into numeric data representation. 3. For each block of 13 x 13 do <ol style="list-style-type: none"> A. Split Block matrix into 13 x 13 B. Split data into training / test sets data

4.	Call CNN: Apply network layers
A.	Conv 1, with 16 filter, AF = Leaky ReLU Max_pool, with pool_size=2 x 2
B.	Conv 2, with 32 filter, AF = Leaky ReLU Max_pool, with pool_size=2 x 2
C.	Conv 3, with 64 filter, AF = Leaky ReLU Max_pool, with pool_size= 2 x2
D.	Conv 4, with 128 filter, AF = Leaky ReLU Max_pool, with pool_size=2 x 2
E.	Conv 5, with 256 filter, AF = Leaky ReLU Max_pool, with pool_size= 2 x2
F.	Conv 6, with 512 filter, AF = Leaky ReLU Max_pool, with pool_size=2 x 2
G.	Conv 7, with 1024 filter, AF = Leaky ReLU
H.	Conv 8, with 1024 filter, AF = Leaky ReLU
I.	Conv 9, with 40 filter, AF = Linear
J.	Flatten output of conv 9
K.	Fully connected layer with Dense layer, AF = softmax
L.	Classify the outcome for a give dataset
M.	update weight
5.	Return to Step 3
6.	CNN performance measures by confusion matrix

3.3 Results of Proposed System

As the video starts playing, detection is made automatically by analyzing video into a number of frames, each frame is checked and a person detection procedure is accomplish depending on his/her gait. After selecting a video, the video stream frames will be read and resized to 416*416 before storing it in a dataset to separate them into two groups: training and testing. The classifier is implemented using conventional neural network classification algorithm. The first step after attaining images is preprocessing steps which include multi important steps in making these images suitable when making the classification procedure works probably. The preprocessing steps include: converting image into grey, blurring image, apply subtraction algebra operation, binaries image, and finally apply dilation and erosion on image. A CNN is used as a classifier with nine convolved layers as shown in table (2), forward the result to a fully connected layer which is responsible to recognize the person depending on his/her gait.

Table 2-CNN Layers Detailed Information

Type	Filter	Size / Stride	Output
Convolution	16	3x3/1	416x416
Max pooling		2x2/2	208x208
Convolution	32	3x3	208x208
Max pooling		2x2/2	104x104
Convolution	64	3x3	104x104
Max pooling		2x2/2	52x52
Convolution	128	3x3	52x52
Max pooling		2x2/2	26x26
Convolution	256	3x3	26x26
Max pooling		2x2/2	13x13
Convolution	512	3x3/1	13x13
Max pooling		2x2/1	-
Convolution	1024	3x3/1	-
convolution	1024	3x3/1	-
convolution	40	1x1/1	-

The accuracy of the proposed model is 99.5%, obtained from 20% of the captured videos' images (frames), that is 1220 image. Each image is resized to 416*416, and then the same pre-processing steps is implemented before entering to 9 convolved and max-polling layers. Finally, two fully connected layers were used with soft max layer for recognition process.

4. Conclusions

The individuals' gait style is attained for recognition purpose to get know person and provide his authenticity to retrieve his personal information from system such as name, age, his enrollment in the system, etc. The proposed system provides an acceptable recognition rate with a minimum loss amount. Max-pooling is better shrinking layer because is retain most informative image information. As well as learning rate is small, but it gave best results in the proposed system. In testing phase, required time for recognition is fast, since recognition process requires feed forward process only, so it can be considered powerful reliable, and fast in concept of speed and can provide security for the protected systems.

References

- [1] Gafurov, Davrondzhon, "A Survey of Biometric Gait Recognition: Approaches, Security and Challenges," Annual Norwegian Computer Science Conference, Norway, pp. 19-21, 2007.
- [2] Idrus, Syed Zulkarnain Syed, Estelle Cherrier, Christophe Rosenberger, and Jean-Jacques Schwartzmann, "A Review on Authentication Methods," 2013.
- [3] Kale, Amir, NareshCuntoor, B. Yegnanarayana, A. N. Rajagopalan, and Rama Chellappa, "Gait Analysis For Human Identification," International Conference on Audio-and Video-Based Biometric Person Authentication, Springer, Berlin, Heidelberg, pp. 706-714, 2003.
- [4] Zhang, Cheng, Wu Liu, Huadong Ma, and Huiyuan Fu., "Siamese Neural Network Based Gait Recognition for Human Identification," 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, pp. 2832-2836, 2016.
- [5] Wang, Xiuhui, Jun Wang, and Ke Yan, "Gait Recognition Based on Gabor Wavelets and (2D) 2 PCA," *Multimedia Tools and Applications*, vol. 77, no. 10, pp. 12545-12561, 2018.
- [6] Battistone, Francesco & Petrosino, Alfredo, "TGLSTM: A Time Based Graph Deep Learning Approach to Gait Recognition," *Pattern Recognition Letters*. 10.1016/j.patrec.2018.05.004.
- [7] Xue, Zhaojun, Dong Ming, Wei Song, Baikun Wan, and Shijiu Jin, "Infrared Gait Recognition Based on Wavelet Transform and Support Vector Machine," *Pattern Recognition*, vol. 43, no. 8, pp. 2904-2910, 2010.
- [8] Xu, Dong, Yi Huang, ZinanZeng, and XinxingXu, "Human Gait Recognition Using Patch Distribution Feature and Locality-Constrained Group Sparse Representation," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 316-326, 2011.
- [9] Zeng, Wei, Cong Wang, and Yuanqing Li, "Model-Based Human Gait Recognition Via Deterministic Learning," *Cognitive Computation*, vol. 6, no. 2, pp. 218-229, 2014.
- [10] Salih, A. Mahdi, Dhannoon, B. Nadeem, "Color Model Based Convolutional Neural Network for Image Spam Classification," *Al-Nahrain Journal of Science ANJS*, vol.23, no. 4, pp. 44-48, 2020.
- [11] Hussien, Z. Khalaf, Dhannoon, B. N., "Anomaly Detection Approach Based on Deep Neural Network and Dropout," *Baghdad Science Journal*, vol. 17, no. (2 Special Issue) NICST, pp. 701-709, 2020.