

MONITORING AGENT SYSTEM BASED ON HUMAN GAIT CLASSIFICATION

N. Balti

Tunis El Manar University, L3S, ENIT, Tunis, Tunisia
nidhal.balti@etudiant-enit.utm.tn

A. M. Aseere

King Khaled University, College of Computer Science, ABHA, KSA
amg@kku.edu.sa

M. Turki-Hadj Alouane

King Khaled University, College of Computer Science, ABHA, KSA
Tunis El Manar University, L3S, ENIT, Tunis, Tunisia
malouane@kku.edu.sa

Abstract - The need for monitoring and surveillance systems for safe access in private or public places is increasing nowadays and becoming very necessary. In the current research paper, we introduced a monitoring agent system that helps in recognition of suspicious human behavior. The paper focuses on the classification task of the agent rather than on its architecture that can be a conventional one. Based on RGB images provided by the camera of the agent, an extraction process extracts silhouette images from RGB ones. The classification agent used the images of the silhouettes to classify human behavior into normal or abnormal. Two classification algorithms are implemented: Convolutional Neural Network (CNN) and a Support Vector Machine. The CNN algorithm implicitly extracts the required features from the silhouette images to perform two steps of classification. However, before SVM perform classification step, appropriate features are extracted. These features correspond to interest points (corners) that describe the GAIT (walking manner). The results of performance, which are reported in this research showed that a very close to 93% classification rate, is achieved by CNN. The SVM classification rate is near to 75%. CNN obtained the classification model, has an average complexity as only two convolutional layers founded it

Keywords: Monitoring agent; Normal/abnormal human behavior; Silhouette images; Classification; CNN; SVM.

1. Introduction

Nowadays, with the rise in crime, people have a great need to feel safe in common, areas such as airports, train stations, shopping malls, and even in their neighborhoods. The timely identification of a suspicious person is an important factor in reducing and even preventing the occurrence of crime. Identification using biometrics has become an important research theme. Among biometrics, GAIT is attracting the interest of many researchers working in developing techniques for persons' identification. To identify people, they are focusing on people's body shapes and the way they walk, even when if the camera does not focus on their faces. Contrarily to face recognition techniques that require very close, high-resolution images of a person's face to work, GAIT based monitoring techniques can be efficient with medium or low camera resolution and a person at the distance of 50 m away from the camera. In most research works, involving GAIT, the focus is on persons' identification purposes [1] [2]. In [3] [4], the physical well-being of people is studied through the analysis of abnormal Gait. However, the use of GAIT in surveillance and control systems for the detection of abnormal behavior of persons is relatively recent [5] [6] [7].

The current research paper proposes a monitoring agent system whose objective is to recognize the behavior and type of the person based on the analysis of its GAIT. A person having abnormal gait may be categorized as suspicious and alarming actions may be taken. The proposed system can be a part of a global multimodal monitoring system.

Several agent systems are proposed for different monitoring and control tasks [8]-[12]. The efficiency of the proposed agent system relies principally on the technique used to distinguish between normal and abnormal GAIT. In fact, the GAIT classification techniques proposed in the literature are essentially performed using analytic models of motion [5][13][14] or machine learning-based models [15]-[17]. In this paper, the classification model built in the inference engine of the agent is computing using machine learning techniques. The conventional machine learning algorithms, such as Support Vector Machine, are still used and still show a high classification efficiency in several classification contexts. However, researchers are showing an increasing

interest in using deep learning algorithms such as Convolutional Neural Network designed to image and video processing.

The rest of the paper is divided into 5 sections. Section 2 talks about the proposed monitoring system based on the classification of the human GAIT. Section 3 describes the details of the system's functionality and architecture. In section 4, we present the proposed approaches for computing GAIT classification models. Indeed, we proposed a GAIT classification based on Support Vector Machine and CNN algorithms. Section 5 presents the results of the experiments conducted to evaluate the performance of two classification techniques.

2. Presentation of the monitoring system

This paper is concerned with a real-time monitoring system based on the detection of abnormal human GAIT denoting a suspicious behavior. The block diagram of the proposed system is depicted in Fig. 1.

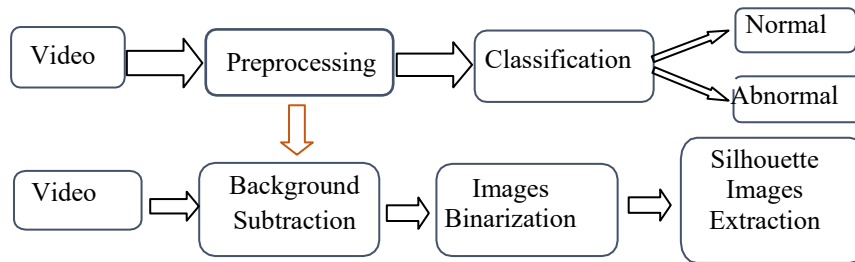


Fig. 1. Block diagram of the proposed Monitoring System

The video pre-processing step can be detailed as follows:

- Fragmentation of the video: the acquired video is sampled in frames. This means that we are processing a set of images.
- Background subtraction: this step, consists of subtracting the background from the current image in order to separate the object from the foreground. Therefore, the background image, which is also called a reference image, is subtracted from the current image to detect the object.
- Image binarization: in this step, the grayscale coded image is transformed into a binary-coded image; black and white. The binarization of the image corresponds to a thresholding operation of the gray level of each pixel. The threshold is fixed to 127.
- Silhouette extraction: this process is extracting the silhouette image of the filmed person which will finally be the input of the classifier.

The implementation of these classical and widely used steps in image and video processing is done using existing functions in the library of OPEN-CV software.

3. Monitoring Agent system

We are concerned with a real-time monitoring system. Our work is focused on the knowledge that the agent exploits to make a decision. In particular, the knowledge is based on an analysis of the perceived data through the machine learning techniques. For the design of the agent monitoring system, this work based on the agent design presented in [12], where three agents performed successive tasks to attain the aims of the system.

The system is supporting the monitoring process based on the analysis of human GAIT operates to achieve goals, organized in three layers, as shown in Fig. 2.

The Acquisition layer ensures guttering the video of the monitored environment by the digital camera of the agent. The agent in this layer must firstly sample the video in order to obtain a set of RGB images, and then apply an appropriate technique to subtract from each image its background and after that extracts the corresponding silhouette image. The extracted silhouette images are saved in the database.

The goals of the Machine Learning layer are analysis of the set of the extracted silhouette images by machine learning algorithm and the recognition of the human behavior type: normal or abnormal. The classification model categorizes the human behavior, computed during the classification algorithm learning process, based on the silhouette images or a collection of features derived from them. The classification model serves as a knowledge base for the agent.

The classification model can be updated, intermittently, using the updated Database. The distribution layer aims to send information about the human behavior detected at the end of the Machine Learning agent task, to the fusion system of a multimode monitoring system. The agent in this layer must distribute the information about human behavior according to an appropriate formatting protocol.

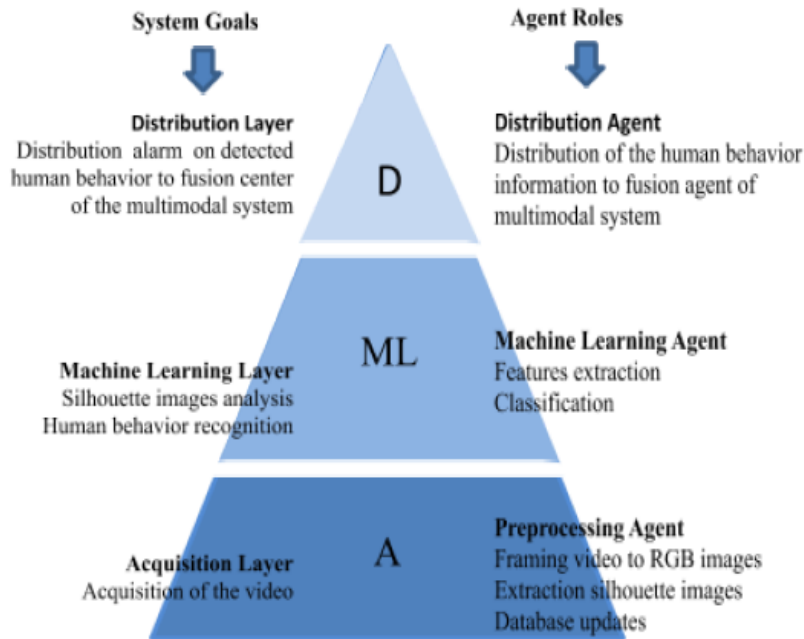


Fig. 2. Correspondence between System layers and agent tasks

Fig. 3 shows the architecture of the monitoring multi-agents system corresponding to the functional description. Agents work sequentially. When the pre-processing agent finishes the extraction of silhouette images, from the RGB ones, it saves them in the database and activates the Machine-learning agent. The latter extracts the features from the current silhouette images and use them as inputs of the classification task. Of course, this is in the case of SVM based classification. In the case of CNN based classification, the feature extraction task doesn't occur. As soon as the human behavior type (normal/abnormal) is recognized at the end of the classification process, the distribution agent is activated to distribute the information on the human behavior to the agent of the multimodal fusion center.

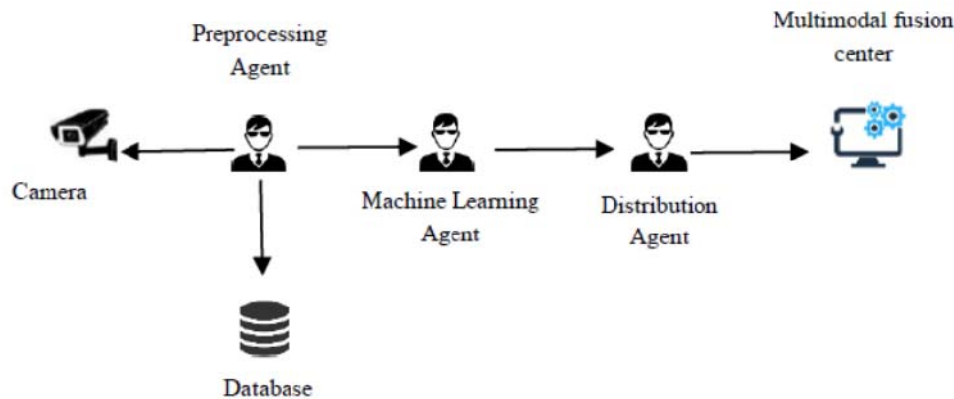


Fig. 3. Agent system architecture

4. Human behavior-based GAIT classification approaches

As announced in the introduction, we are focusing on the approach used to detect human behavior from silhouette images. Two phases are required to set the classification model, which is implemented in the Machine Learning agent. In this paper, we propose to experiment with two supervised classification algorithms: Convolutional Neural Network (CNN) and Support Vector Machine (SVM). The steps required by the classification process are slightly different from SVM to CNN. In fact, the CNN algorithm doesn't require the extraction of features since its inputs are Silhouette images. However, the SVM is a classical Machine Learning algorithm, so it performs classification basing on features that describe the person's GAIT in the silhouette image.

The learning process of CNN and SVM algorithms, for the construction of the respective classification models, is performed using a database of Silhouette images of persons walking in normal and abnormal ways. Hence, these images are labeled into two classes: normal and abnormal.

4.1 Supervised classification based on Convolutional Neural Network

According to fig. 4, the CNN algorithm is loaded with silhouette images. Implicitly, CNN's successive convolutional layers take out the relevant features.

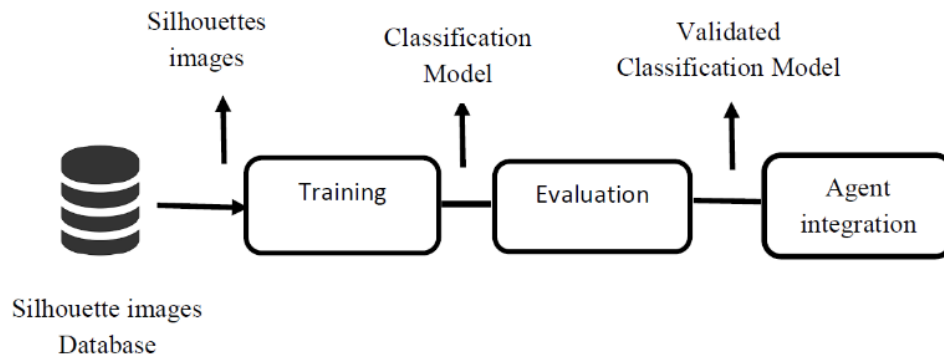


Fig. 4. Process for building the classification based CNN model

Convolutional Neuron Networks (ConvNets or CNN), which are part of the Deep Learning family of algorithms, are dedicated to image and video processing, in particular classification and prediction [18].

The block diagram in Fig. 5 shows the general architecture of CNN. The whole network can be divided into a feature learning section and a classification section. The convolution layers act as an extractor of the features representing the content of an image.

- Convolutional layer

This layer learns the characteristics of the image using a set of linear filters applied to the blocks of the image. The convolution operation requires two inputs: an image block and a filter of the same size. At the level of each convolution layer, a set of filters is used, their number depends on the application and of course on the content of the images. The deeper the convolution layer, the more sophisticated descriptors can be extracted. However, the first layers extract basic features like corners, edges, etc.

- Pooling layer

This layer makes a down-sampling operation along the spatial dimensions (Width x Height) of the input volume for the next Convolutional layer. The procedure is done by either taking the full value (Max Pooling) or the average of values. Max pooling was chosen over the others because of its better performance.

- Classification

The last stage is a classifier. We need to convert the output layer of the convolutional part (the pooled feature map) into a one-dimensional feature vector, this operation is called "flattening". The flattening layer is the input of the fully connected layer. This is used to combine the features with more attributes that predict the output (classes) more accurately. Softmax converts the output to probability values for each class.

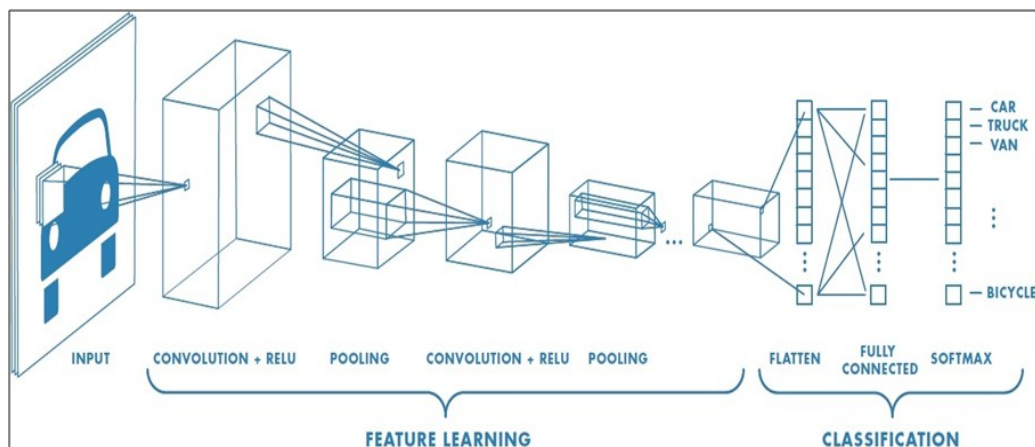


Fig. 5. Block diagram of CNN mechanism [19]

CNN's GAIT classification involves setting a bunch of parameters, such as the number of neurons on each convolutional layer, the number of layers, filter sizes for each layer, batch size, backpropagation algorithm's learning rate. The latter is gradient descent based algorithm that ensures the updating of the filters' weights at each epoch. Since the classification is based on silhouette images which are binary ones, we expect that the number of convolutional layers required to achieve satisfactory classification performance, will not be very high.

4.2 Supervised classification based Support Vector Machine

Fig. 6 represents the block diagram of the overall process for building the classification model based SVM algorithm.

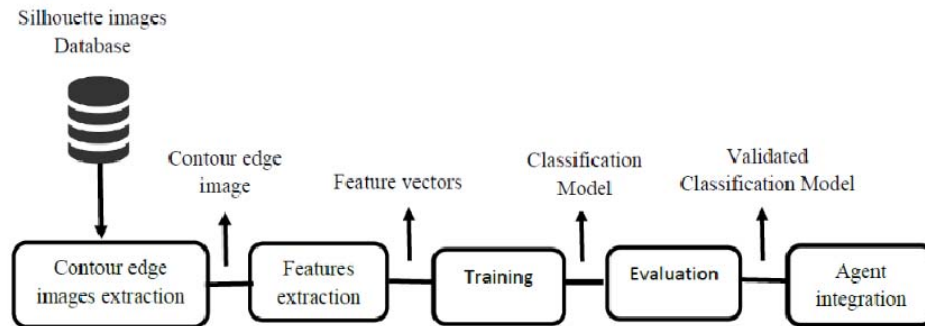


Fig. 6. Process for building the classification based SVM model

The features, input of the SVM algorithm, are extracted from the contour edge images. A conventional transformation of a silhouette image, provide a contour edge image. From the contour image, we extract the most relevant corners, and then we compute the Euclidean distances between the corners and the center of the contour edge image (reference point). These distances are used for classification.

Corners are the vital features in the image, and in common, they are characterized as interest points. They are invariant to translation, rotation, and illumination. Fig. 7 shows an example of a silhouette image (a), its corresponding edge corner image (b) and the computed interest points (c).

For the detection of the corners, the Shi-Tomasi corner detector algorithm is used [20]. It is very similar to the famous Harris corner detector algorithm. It differs in computing of the score value which is compared to a threshold to decide if a pixel is a corner. This slight change helps the Shi-Tomasi corner detector algorithm to outperform the Harris corner detector one.

Through setting the number of the corner, the minimum distance between successive corners and the threshold, the Shi-Tomasi corner detector extracts the most relevant corners characterizing the GAIT.

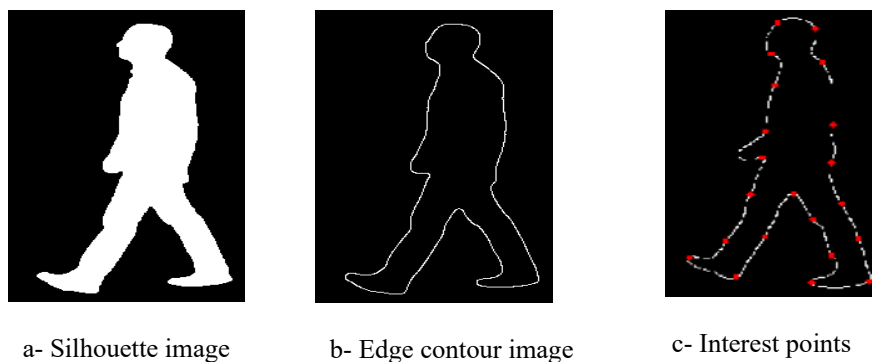


Fig. 7. Example of a silhouette image, Edge contour image and interest points.

5. Experimental results

In this section, we present the performance analysis of the classification techniques proposed in the paper. The INIT GAIT database [21], is used for training and the evaluation of the classification algorithms. The database was originally registered for the study of GAIT medical abnormalities. However, the simulated abnormalities of the GAITS can also indicate suspicious behaviors, which motivate us to use this open-access database for our work. It is important to notice that the existing open-access GAIT databases are all oriented to GAIT identification and recognition.

5.1 CNN classification results

Classically, the performance of CNN can be measured by the Loss function which indicates the mean square error between the actual output and the algorithm, and the accuracy that corresponds to the percentage of the right classified images. Both assessment criteria are highly correlated. So, in this research, we are focusing on the accuracy rate.

Many simulations are performed to set the parameters of the Convolutional Neural Network architecture that achieves a satisfactory classification rate. The most suitable CNN architecture is defined by the following parameters in the light of different results obtained: two convolution layers, 32 filters for size 3x3, the learning rate set to 10^{-6} , a sigmoid as an activation function, a batch size of 32 silhouette images and a dropout of 25 percent. Indeed, the dropout is necessary to avoid the overfitting aspect.

We notice that the increase in the number of convolution layers to 3 or 4 layers, doesn't enhance the accuracy of the model. Such results are expected as the silhouette images don't contain complex information that requires a deep network.

The above architecture of CNN is tested on 2924 silhouette images of the database; 2339 images (80%) are used for the training phase and 585 images (20%) for the validation phase. Ninety percent classification accuracy is obtained, as shown in Fig. 8, where we map the precision of the training and testing phases against the number of epochs.

For a prediction step, we used the obtained CNN classification model to classify 64 new silhouette images unused in training and validation tests. Normal behavior is predicted with an accuracy of 80%, however, an abnormal one is predicted with an accuracy of 79%.

In order to enhance the detection performance of the CNN classification model, we trained the CNN model on a larger database of silhouette images (75000). The obtained variations of the accuracy versus the number of epochs are reported in Fig. 9. A classification rate of almost 98% is reached.

For the detection of the GAIT type of new silhouette images, we obtain a classification rate of 97% for normal GAIT and 89% for an abnormal one. As expected, increasing the size of the database enhances the performance of the classification CNN model that must be introduced in the inference engine of the Learning agent.

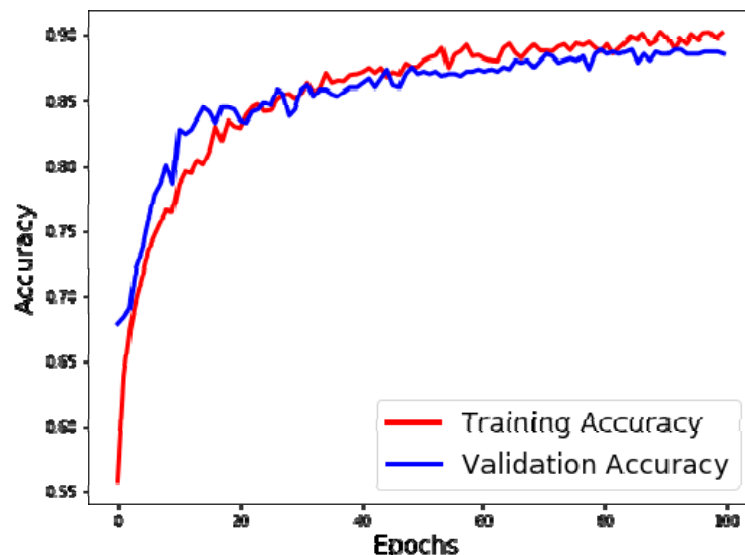


Fig. 8. Performance of CNN (database of 2924 silhouette images)

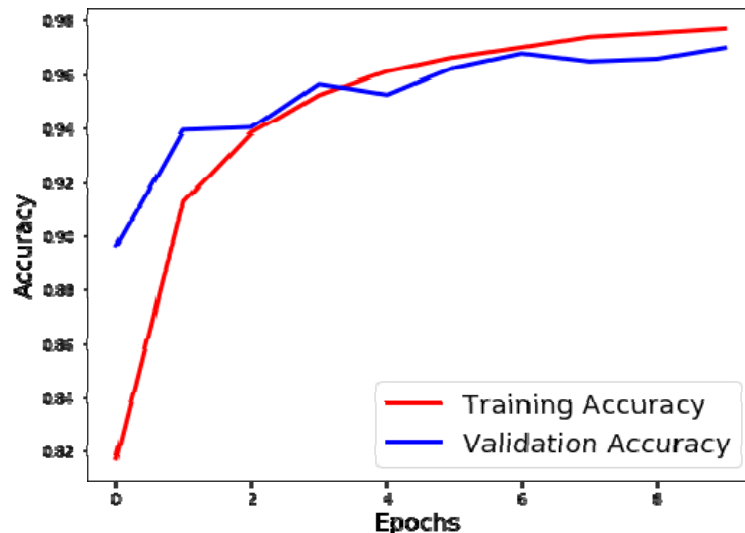


Fig. 9. Performance of the CNN model on a large database (75000 images).

5.2 SVM classification results

Many tests were conducted to extract the relevant corners to form the silhouette images. The number and the quality of extracted corners are fixed through the parameters of the Shi-Tomasi corner detector algorithm. Many sets of values are tested in order to extract the same number of corners for all images of the database, such thing is important to construct vectors of features of the same length for all the images. Indeed, the length of the vector of features, which corresponds to the number of corners extracted from each silhouette image, influences the classification accuracy of the SVM algorithm. In fact, with less than 20 corners per image, the obtained classification rate is insufficient.

After several experiments, we obtain a classification rate near to 74% with a database of around 2400 images. The parameters of the detector are fixed in such a way that we extract the most relevant 40 corners for each silhouette image. However, as expected with the SVM algorithm, increasing the database doesn't really improve the classification rate obtained by the SVM classification model. Even increasing the number of corners from 40 to 50 doesn't improve the classification rate.

6. Conclusion and future works

In this paper, a monitoring agent system based on the detection of human abnormal GAIT is presented. The classification model plays an important role in the proposed conventional architecture of the monitoring agent. Two approaches for building the classification model are proposed. Thus, the Convolutional Neural Network and the Support Vector Machine algorithms are used for the supervised classification of Human GAIT into normal and abnormal. The classification rate obtained with CNN is an average of 93%, however, with SVM it is around 74%. Future works will focus on improving the performance of the classification model that must be built in the inference engine of the learning agent. Two lines of work are envisaged. Firstly, the construction of a richer database and particularly better adapted to the monitoring context. Indeed, the existing database can be enriched by new images corresponding to new scenarios of abnormal behaviors. Secondly, the use of an algorithm that takes into account a set of silhouette images corresponding to a GAIT cycle. In the case of SVM, this corresponds to a feature vector of very high dimensionality, which generates a great complexity of calculation may be unmanageable by existing versions of the algorithm. CNN requires a high batch size which doesn't guarantee an interesting classification rate.

References

- [1] Amit Kale, Aravind Sundaresan, A. N. Rajagopalan, Naresh P. Cuntoor, Amit K. Roy-Chowdhury, Volker Krüger, and Rama Chellappa: Identification of Humans Using Gait. IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 13, NO. 9, 1163-1173, 2004.
- [2] Nikolaos V. Boulgouris, Dimitrios Hatzinakos, Konstantinos N. Plataniotis: Gait Recognition: A challenging signal processing technology for biometric identification. IEEE signal processing magazine 22 (6), 78-90, 2005.
- [3] C. Bauckhage, J. K. Tsotsos, and F. E. Bunn, Automatic detection of abnormal GAIT, Image and Vision Computing, vol. 27, no. 1, pp. 108-115, 2009.
- [4] M. Nieto-Hidalgo, F. J. Ferrndez-Pastor, R. J. Valdivieso-Sarabia, J. Mora-Pascual, and J. M. Garca-Chamizo: A vision based proposal for classification of normal and abnormal GAIT using rgb camera. Journal of Biomedical Informatics, vol. 63, pp. 82 - 89, 2016.
- [5] Naveen Rohila et. al.: Abnormal Gait Recognition. International Journal on Computer Science and Engineering, Vol. 02, No. 05, 1544-1551, 2010.
- [6] Ji-il Park, Min Kyu Yu, Jong-woo Lee, Sam-Hyeon Yoo: Analysis of the Gait Characteristics of Soldier between the Normal and Loaded Gait. International Journal of Mechanical and Mechatronics Engineering Vol:10, No:7, 2016.
- [7] <https://apnews.com/bf75dd1c26c947b7826d270a16e2658a>

- [8] Sara Rodríguez, Juan F. De Paz, Javier Bajo, Dante I. Tapia, Belén Pérez: Stereo-MAS: Multi-Agent System for Image Stereo Processing. In: Bio-Inspired Systems: Computational and Ambient Intelligence. 10th International Work-Conference on Artificial Neural Networks, IWANN 2009, Salamanca, Spain, June 10-12, Springer Berlin Heidelberg (2009).
- [9] Rodríguez S., Bajo J.: Multi-agent System for Management and Monitoring of Routes Surveillance. In: Corchado E., Abraham A., Pedrycz W. (eds) Hybrid Artificial Intelligence Systems. HAIS 2008. Lecture Notes in Computer Science, vol 5271. Springer, Berlin, Heidelberg.
- [10] Rao Muzamal Liaqat, Ali Athar, Nazar Abbas Saqib: Intelligent Agent Based System for Monitoring and Control of Hospital Management System. In: 2nd International Conference on Information Science and Security (ICISS) 2015, 14-16 Dec. 2015, Seoul, South Korea.
- [11] Nicksson Ckayo Arrais de Freitas and Marcelino Pereira dos Santos Silva: Multiagent System for Image Mining. In: Multi-agent Systems, intechopen, pp 123-150, 2017.
- [12] Ioannis N. Athanasiadis and Pericles A. Mitkas: An Agent-Based Intelligent Environmental Monitoring System. In: Management of Environmental Quality, An International Journal vol.15, no.3, pp.238-249,2004. Emerald Publishers.
- [13] Faezeh Tafazzoli, Reza Safabakhsh: Model-based human gait recognition using leg and arm movements. Engineering Applications of Artificial Intelligence, 23(2010)1237–1246.
- [14] S.L. Dockstader; M.J. Berg; A. Murat Tekalp: Performance analysis of a kinematic human motion model. In Proceedings. IEEE International Conference on Multimedia and Expo, Lausanne, Switzerland, 26-29 Aug. 2002.
- [15] M. R. Talib, A. Sha Que, M. K. Hanif, and M. U. Sarwar: GAIT identification using neural network, International Journal of Advanced Computer Science and Applications, vol.8, no.9, 2017.
- [16] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi: Multi-view large population GAIT dataset and its performance evaluation for cross-view GAIT recognition. IPSJ Transactions on Computer Vision and Applications, vol. 10, no. 1, p. 4, Feb 2018.
- [17] S. Ramachandran, L. Harika, and C. Giridharan: An intelligent system to detect human suspicious activity using deep neural networks. Journal of Intelligent Fuzzy Systems, pp. 1-12, (02) 2019.
- [18] Charu C. Aggarwal: Neural Networks and Deep Learning. Springer International Publishing AG, part of Springer Nature 2018.
- [19] <https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148>
- [20] J. Shi and C. Tomasi: Good features to track. In Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94), pages 593–600, Seattle, Washington, June 1994.
- [21] Javier Ortells, María Trinidad Herrero-Ezquerro, Ramón A. Mollineda: Vision-based gait impairment analysis for aided diagnosis. Medical & Biological Engineering & Computing, September 2018, Volume 56, Issue 9, pp 1553–1564.