

Received 11 November 2022, accepted 6 December 2022, date of publication 14 December 2022,
date of current version 22 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3229260

RESEARCH ARTICLE

Biometric-Based Security System for Smart Riding Clubs

ISLEM JARRAYA¹, (Member, IEEE), FATMA BEN SAID¹,
TAREK M. HAMDANI^{1,2}, (Senior Member, IEEE), BILEL NEJI³,
TAHA BEYROUTHY³, AND ADEL M. ALIMI^{1,4}, (Senior Member, IEEE)

¹REGIM-Laboratory—REsearch Groups in Intelligent Machines, National Engineering School of Sfax (ENIS), University of Sfax, Sfax 3038, Tunisia

²Higher Institute of Computer Science Mahdia (ISIMa), University of Monastir, Monastir 5147, Tunisia

³College of Engineering and Technology, American University of the Middle East, Egaila 54200, Kuwait

⁴Department of Electrical and Electronic Engineering Science, Faculty of Engineering and the Built Environment, University of Johannesburg, Johannesburg 2006, South Africa

Corresponding author: Islem Jarraya (islem.jarraya@enis.tn)

This work was supported by the Tunisian Ministry of Higher Education and Scientific Research under Grant LR11ES48.

ABSTRACT Horses, workers or riders need safety in a farm or a riding club. On account of the great value of the horse, the breeder needs to protect it from theft and disease. In this context, it is important to detect and recognize the identity of each worker or rider and horse for security reasons. In fact, this paper proposes a Smart Riding Club Biometric System (SRCBS) consisting in automatically detecting and recognizing horses as well as humans. The proposed system is based on the facial biometrics for a horse as well as the human gait biometrics due to their simplicity and intuitiveness in an uncontrolled environment. This work suggests a Siamese network based on DenseNet features for human gait recognition, a Sparse Neural Network (SNN) based on sparse features for horse face detection and a horse face recognition method based on Gabor features, LDA and SVM. Because of the unavailability of horse databases, this paper presents a new benchmark for horse detection and recognition in order to evaluate our proposed system. The proposed systems achieved an average accuracy of gait recognition equal to 95% in the 0° view, 100% in the 90° view and 98.90% in the 180° view on Casia-B dataset, an average precision equal to 90% for horse face detection and a recognition rate equal to 99.89% for horse face identification.

INDEX TERMS Horse face detection, human gait recognition, neural network, Siamese network, smart riding club, THDD.

I. INTRODUCTION

Object detection and identity recognition have been widely employed in recent years for several security reasons. Both animals and humans need safety. Farm animal control is required for verification of the source, the identity and the production process as well as for the livestock health surveillance. In addition, the outbreak of diseases such as bovine spongiform encephalopathy (BSE), foot-and-mouth disease (FMD) and classical swine fever (CSF) and the importance of export markets for domestic producers require the implementation of animal detection and identification programs that allow farmers to trace cattle from birth. Hence, the regulatory

The associate editor coordinating the review of this manuscript and approving it for publication was Yongming Li¹.

provisions (Orientation Law 1999) [1] made the equine identification compulsory to ensure the origin of each horse and certify their identities for the purpose of fighting against theft and fraudulent horses as well as assuring healthy monitoring.

In this context, making sure that both animals and humans are secure seems extremely interesting in a riding club. Owing to its great value, it is very necessary to help not only the owner but also the horse so that it could be controlled without obstacles or difficulties in the barn, the box or the race track.

Well-known traditional methods of animal identification use plastic ear tags, tattoos, freeze branding, hot-iron branding or RFID electronic ear tags. However, these techniques may threaten the well-being of animals [17] and require effort to identify the animal. Because of these problems, another detection and identification way should be used for

effective control by means of a stressed robust biometric marker that is invariant, fraud-proof, fast, accurate, inexpensive and non-invasive to capture. Researchers have resorted to other identification methods using different physiological biometric modalities such as muzzle patterns [51], [66] and retina pattern [65]. Although these patterns are effective, they require an effort to get the identity of the animal. In fact, these methods necessitate direct contact between the animal and the camera to capture the muzzle or retina patterns. Thus, a biometric detection and recognition system needs to be created for the comfort of both animals and humans taking into account the subject in motion for their well-being and the lack of direct contact between the subject and the sensor.

On the other hand, gait acquisition does not require contact with the sensor just like the face biometrics and is less likely to be exposed to obscurity than other biometrics. In fact, gait identification is easier to use and more secure than the other biometrics. The subject can be identified in any view and at any location. In a riding club, the gait biometrics are very useful for human identification but not for horses. In the stable, the horses remain in their boxes without walking and only the face appears. Therefore, the facial biometrics are the most suitable biometric traits for horse identification. In fact, the most suitable biometric traits that can be useful for a smart riding club security are the facial biometrics for horses and the gait biometrics for humans due to their simplicity and intuitiveness in an uncontrolled environment.

Face recognition is one of the most promising modalities for horse identification [28], [29], [72] despite the scarcity of research in this field. According to the 5th edition of the FEI,¹ the description of the horse face is particularly different. In fact, the horse hair color is extremely varied and contains different texture patterns such as the blaze, the strip, the snip, the star and the lip marking on the head. The variation in direction and shape of these texture patterns is different from one horse to another. Indeed, the use of the face pattern is more effective for the horse identification.

This paper suggests a new a Smart Riding Club Biometric System (SRCBS) using helpful modalities, features, methods and techniques to achieve an efficient automatic system with an improved performance. The proposed SRCBS is very important for breeders to ensure the safety of the horse due to its great value. Additionally, SRCBS helps breeders verify the presence of workers and riders in the club. Proposing a useful human gait recognition system as well as a horse face detection and recognition system is the main interest in this work. Regretfully, the face detection task is still very difficult mainly because of the large intra-class variation, the illumination change, the variable pose, the complex background, the partial occlusion and the uncontrolled environment. Face textures and shapes are grossly diverse, which makes the animal face detection extremely difficult. This is probably the reason why the number of related work approaches for

animals is small. Besides, face and gait recognition tasks are not less difficult than the face detection task with the same problems. Despite these difficulties, recent research has achieved significant progress to resolve the interesting detection and recognition problems. The detection and recognition rates have reached nearly 90% of the human face using boosting-based and CNN-based (Convolutional Neural Network) approaches [26], [45]. The same thing was noticed for gait recognition. The fact that Convolutional Neural Networks (CNNs) are strong and fascinating classification tools is among the reasons why Deep Learning is immensely popular and widely used for computer vision tasks. Despite the rapid development of CNN, there are different challenges during its training. First, CNN requires a huge dataset for training [50]. However, the riding club does not have enough horses, workers and riders to construct a dataset for CNN training. In fact, the number is much smaller. Second, pooling layers eliminate a great deal of information and ignore the relationship between image parts according to Hinton [58]. Finally, CNNs represent a big number of parameters and layers, leading to much training time and high computational complexity.

According to the above-listed problems, the challenging questions at this step are:

- How to verify the human gait using a network of small number of parameters and a small database for training.
- How to create a neural network as a backbone detector with a small number of parameters and layers in order to minimize the training time and the computational complexity.
- How to determine the powerful features and effectively detect horse faces using a small database for training.
- How to recognize horses using the simplest method for training and testing.

The challenging issue consists in establishing a safer smart riding club and proposing a new application to detect and recognize the horse face and the human gait with an easy and fast way without the need for a huge data for training. The different contributions in this paper are as follows:

- Proposing a new SRCBS using appropriate modalities and efficient methods.
- Making a Siamese network based on DenseNet features for human gait recognition without the need for a huge dataset for training.
- Making a Sparse Neural Network for horse face detection (SNN) based on sparse features using the smallest number of parameters and without the need for a huge horse data for training.
- Making a system for horse face detection based on the proposed SNN.
- Employing a sparse feature selection method for face detection. The feature selection field has been well adapted resorting to many learning methods for pattern recognition applications. However, these are not devoted to object detection applications.

¹Federation Equestre Internationale (2000), 5th edition (Identification of Horses with the narrative and the diagram), url:<https://www.fei.org/>

- Using the proposed sparse feature selection method as a training optimizer for a sparse hidden layer of SNN instead of such traditional algorithms as ADAM and SGD.
- Making a new horse database called Tunisian Horse Detection Database (THDD), which could contribute to the research community of the animal biometrics. To the best of our knowledge, this is the only dataset of public face image that is available for research on horse detection.
- Proving that our proposed SNN could get better performance than the CNNs of the traditional detectors. In the experiments, the efficiency of SNN was demonstrated for face classification compared to the other Convolutional Neural Networks using the smallest number of parameters and without the need for a huge dataset.
- Showing through experimental studies that the proposed recognition system for horses got a better performance than the related works.
- Proving that our proposed network for gait recognition had a better performance than the related works.

The rest of this paper is organized as follows: Section II presents the related works on human gait recognition, animal face detection and animal face recognition. However, the proposed SRCBS and its sub-systems are described in Section III. Section IV is devoted to the presentation of THDD which extends and generalizes THoDBRL'2015, whereas Section V focuses on the experimental study. The conclusion of this paper which also presents some possible future work is eventually drawn in Section VI.

II. RELATED WORKS

A. HUMAN GAIT RECOGNITION

This section presents brief survey developments in gait recognition and uses deep learning methods on the Casia-B dataset.

Song et al. [15], Zhang et al. [77] and Merlin et al. [37] are among the works based on Siamese architectures for gait classification and matching:

Song et al. [15] proposed a network known as GaitNet which is composed of two convolutional neural networks. One corresponds to gait segmentation and the other corresponds to classification. The supervision of both Siamese loss and soft-max loss were used to learn the final gait features. GaitNet achieved an average accuracy of 89.9%.

Zhang et al. [77] introduced a Siamese network which takes the gait energy image (GEI) as an input. In the training step, companion loss was employed in the middle layer using Multi-Layer Side-Output (MLSO) as a reference. Zhang et al. performed their network on the Casia-B dataset with a recognition rate equal to 75.17%.

Merlin et al. [37] proposed the CCGI gait feature which keeps more temporal and spatial differences in the gait patterns. The used model based on a convolutional neural network (CNN) gave an empirical evaluation to recognize and classify the discriminative changes of the CCGI feature. This method

achieved a mean accuracy in the normal walking conditions equal to 98.87%.

Other works proposed approaches based on CNN networks for gait identification and recognition such as [10], [24], [34], [64], [74], and [79], recently Chao et al. [23], Likai et al. [67], Hung-Min et al. [27], Liao et al. [48], Tianrui et al. [62], Shopon et al. [40] and Muhammad et al. [36]:

Chao et al. [23] suggested a network named GaitSet to extract both spatial and temporal information. This network is based on 2D-ConvNet as a backbone, Set Pooling to aggregate gait information, Horizontal Pyramid Mapping to focus on features of varying sizes to gather local and global information, and Triplet loss for classification. Chao et al. [23] found a mean accuracy of the 11 views equal to 96.1% [23] on the Casia-B dataset.

Likai et al. [67] proposed the STC-Att model based on a CNN branch taking silhouettes as input and a GCN branch taking skeletons as input. They found a recognition rate equal to 97.7% for the normal state.

Hung-Min et al. [27] proposed a gait recognition framework called Temporal Attention and Keypoint-guided Embedding (GaitTAKE). The framework merges the global appearance feature with the local appearance feature based on temporal attention and a temporal aggregated human pose feature. Experimental results showed that their method achieved a rank-1 accuracy of 98.0% on the normal set of the CASIA-B gait dataset.

Liao et al. [48] introduced a model-based gait recognition method, PoseMapGait, which consists of two streams: heatmap Convolutional Neural Networks (gaitMap-CNN) and Pose Graph Convolutional Networks (gaitPose-GCN). This method was evaluated on CASIA-B dataset. PoseMapGait achieved a mean accuracy of 11 gallery views in the normal walking conditions equal to 75.7%.

Tianrui et al. [62] proposed the LagrangeGait framework which is composed of three branches. The first branch extracts the second-order motion feature while the second one, the branch of the GaitGL backbone, extracts the appearance feature. The third branch predicts the view of input silhouette sequence. The recognition accuracy of this method was equal to 96.9% for normal walking.

Shopon et al. [40] introduced the RGCNN architecture (Residual Connection-based Graph Convolutional Neural Network). In fact, RGCNN backbone based on Residual Connections resulted in transformed body joints. The RGCNN architecture attained a testing accuracy of 98.86% for four views on the normal walking set.

Muhammad et al. [36] proposed using the ResNet101 for feature extraction using transfer learning. Besides they introduced the kurtosis-controlled entropy (KcE) approach for feature selection followed by a feature fusion step based on correlation. The multi-class OaA-SVM was applied for classification. The prediction accuracy of normal walking was 96%.

The proposed systems and networks have already taken long CNN architectures as backbones. However,

Convolutional Neural Networks (CNNs) represent a big number of parameters and layers and require a huge dataset for training. This leads to much training time and high computational complexity.

To deal with the above-listed problems, this paper used the Siamese Network for human gait recognition. The proposed network effectively exploits the human gait features which is extracted from a pre-trained DenseNet network to obtain a robust and fast recognizer using a small number of operations and parameters without the need for a huge dataset.

B. ANIMAL FACE DETECTION

The number of works in this area is very limited due to the complication of the animal face detection and recognition tasks. The existing related works in this field are as follows:

Zhang et al. [73] proposed a set of Haar of Oriented Gradients (HOOG) to capture the texture and shape features on the animal head (such as cats, tigers, pandas, foxes and cheetahs). They used SVM for classification and decision calculation. Using the Cat Database, they found a precision equal to 95% and a recall equivalent to 99.8%.

Yamada et al. [9] proposed detecting dog and cat heads using edge-based features. They selected four directional features (Horizontal, Vertical, Upper Right and Upper Left) to detect the facial characteristics. They used a multi-layer classifier for features classification. Yamada et al. performed their method on a set of cat and dog images from the web. The recall rate was equal to 85% on the cat set and 90% on the dog set.

Mukai et al. [43] focused on the cat and dog face detection. They used the same Viola-Jones method and employed both the Haar and the HOG descriptors for feature extraction. Using 58 images from the Cat Database for the test, they found a recall equal to 96.6% and a precision equivalent to 75.7%. However, they achieved a recall equal to 98.3% and a precision equivalent to 90.8% using 60 images from the Stanford Dogs Dataset.

These traditional animal face detectors with handcrafted features, have been replaced in the recent works by deep convolutional neural networks with the ability to extract discriminating face features.

Vlachynska et al. [8] used the faster R-CNN proposed in [57] with ResNet-101 for dog face detection. They found an Average Precision equal to 98% on the Columbia Dogs Dataset.

Tureckova et al. [7] who used the YOLOv3 detector with DarkNet-53 for dog face detection noticed an Average Precision equivalent to 92% on the Columbia Dogs Dataset and the Oxford-IIIT Pet Dataset.

Xu et al. [12] used RetinaNet with ResNet-50 for cattle face detection. They found an average precision score of 99.8%.

Song et al. [59] optimized YOLOv3. They performed their proposed method on a sheep face dataset. The mAP was about of 97.2% of by clustering the anchor frames of the YOLOv3 compressed model based on DarkNet-53. The number of

parameters of the proposed model was reduced to 1/4 times the size of the original model.

The proposed detectors [7], [8] have already taken on known CNN architectures (ResNet and DarkNet) as backbones. In addition, other detectors for animal detection [63] based on CNN were propounded. However, Convolutional Neural Networks (CNNs) ignore the relationship between image parts, represent a big number of parameters and layers and require a huge dataset for training. This leads to much training time and high computational complexity.

To deal with the previously-mentioned problems, this paper introduces the Sparse Neural Network for animal face detection SNN. The proposed network effectively exploits the animal face characterization to obtain fast classification and detection using the smallest number of parameters and without the need for a huge dataset.

C. ANIMAL FACE RECOGNITION

In the last decades, the development of facial recognition systems has been achieved using manually-noted databases in order to locate the facial area in the image. Overall, facial recognition systems have not been automated by facial detection systems [5], [22], [28], [29], [41], [53], [72], [76]. However, these methods allow high recognition rates but their systems lack automatic face detection, which is why the animal face detection system is important to ensure safety and security.

Jarraya et al. [28] benefited from the horse face properties and proposed an approach for horse identity recognition using frontal facial features of 47 horses. They used the Gabor filters and the LDA for feature extraction. However, the Euclidian (Euc) distance and MahCosine (MC) distance were employed for classification. They validated our previous system using the THoDBRL'2015 and they achieved a recognition rate equal to 95.74%.

Jarraya et al. [29] suggested a multi-view horse face recognition using the THoDBRL'2015 that contains 47 horses. They used the Gabor filters for face characterization, the Stacked Auto-encoders (SA) to reduce the size of the feature vector and SVM for classification. Using 9 images for training, they obtained a recognition rate equal to 94.22% on the frontal images.

Ouarda et al. [72] propounded a new feature descriptor (RNGLBP) based on the Gabor and LBP features. They tested the proposed approach on the THoDBRL'2015 using the SVM classifier. They reached a recognition rate equal to 98.77%.

Shi et al. [76] propounded the Residual InterSpecies Equivariant Network (RiseNet) for deep cross-species feature learning. The features of the lower and the upper halves of faces were learned separately. They merged these features as additional information to improve the performance of the proposed RiseNet. Shi et al. performed the suggested method on the THoDBRL'2015. They found a recognition rate equal to 82.56%.

Face recognition was performed on other animals such as cattle by [5], [53] and dogs by [22]. Kumar et al. [53] used the SURF and the LBP for feature extraction and the Chi-Square for classification. The performance was about 92,5% with a database of 40 cattle.

Salama et al. [5] used the Bayesian optimization to automatically set the parameters for the DenseNet convolutional neural network. They performed their method on a database of 52 sheep. They achieved a recognition rate of 98%.

Mougeot et al. [22] proposed a deep learning approach based on a deep convolutional neural network for the face recognition and verification of dogs. They found a recognition rate equal to 88% on a database of 48 dogs.

Weng et al. [81] propounded a two-branch convolutional neural network for cattle face recognition. They performed their method on collected cattle face images. The performance was equal to 99.85% on cattle face images, 99.81% on cow face images, and 99.71% on beef cattle and cow mixed images.

Xu et al. [11] used RetinaFace-mobilenet with ArcFace Loss for cattle identification. The proposed CattleFaceNet outperformed on a dataset of 90 cows with identification accuracy of 91.3%.

Hitelman et al. [3] proposed to use ResNet-50V2 network with ArcFace loss function for sheep face identification. They performed their system on a database of 81 Assaf breed sheep. After transfer learning, the system achieved an average identification accuracy of 97%.

Xu et al. [20] enhanced the Siamese Neural Network for cow face recognition. Using a database of 63 cows, the system achieved an accuracy of 93%.

The previously-mentioned related works conducted their proposed methods on small datasets that contained on average between 40 and 50 subjects with no more than 10 images for each subject. This number is not sufficient for a deep neural network training. This explains the high recognition rate for works [28], [29], [53], [72] that used handcrafted features and the low rate for works [11], [20], [22], [76] which used deep neural network approaches. Hence, the proposed solution in this paper is to avoid the use of a deep neural network which represents a big number of parameters.

Owing to the importance of studying the horse, some works have been carried out recently and have given new computer solution for control and security objectives. In fact, according to North [55], the interaction between the horse and the computer is more substantial.

Hummel et al. [25] and Li et al. [80] proposed using the face pattern as it is rich in information about the life of horses such as pain, disease and feelings. The objective of Hummel et al. [25] was to recognize the pain in equines. They suggested employing the HOG features and SVM for pose estimation and the SIFT, LBP, HOG and VGG16 features as well as SVM for pain recognition. Using their own equine dataset, Hummel et al. found the F1 score to be equal to 0.89 for pose estimation and 0.53-0.87 for pain estimation.

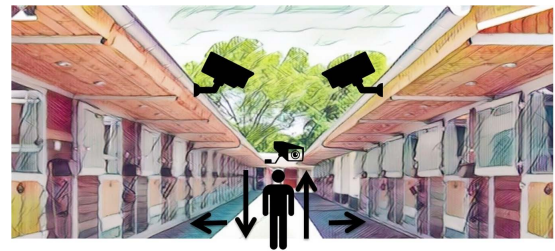


FIGURE 1. Smart riding club biometric system with three surveillance cameras.

The objective of Li et al. [80] was to detect EquiFACS units automatically from the horse face images. They suggested testing the DRML and AlexNet for horse facial AU recognition. They found an accuracy between 54.0% and 58.1% using DRML and an accuracy between 52.8% and 57.0% with AlexNet on their own dataset.

Bragança et al. [19] propounded improving the gait classification of horses using data generated by the Inertial Measurement Unit (IMU). They built a dataset of 120 horses which included 7.576 strides of 8 different gaits. Their gait classification model based on the LSTM network achieved 97% of accuracy.

III. THE PROPOSED SMART RIDING CLUB BIOMETRIC SYSTEM (SRCBS)

In this section, the proposed Smart Riding Club Biometric System (SRCBS) which involves three sub-systems for human (worker and rider) and horse recognition is introduced. The objective of the first sub-system (WRIR-GB system: Human Identity recognition based on Gait Biometrics) is to recognize the person from a specific distance using the gait modality. As for the second sub-system (HFD-SF system: Horse Face Detection using Sparse Features), it is used to detect the horse face. The objective of the third sub-system (HIR-FB system: Horse Identity Recognition based on Face Biometrics) is to develop a contactless solution for horse recognition using facial features. In fact, three Camera positions were proposed for human gait and horse face detection and recognition. The first one was placed at the end of the barn for gait capture of the worker and rider from a distance in front view (0° angle) and in rear view (180° angle). The other two cameras were put in the barn near the horses for the capture of their facial biometrics and the human gait in profile view (90° angle). Fig. 1 and Fig. 2 illustrates the proposed architecture of the new SRCBS system.

A. WRIR-GB SYSTEM: WORKER AND RIDER IDENTITY RECOGNITION BASED ON GAIT BIOMETRIC

The main advantage of the gait biometrics is that it offers a great potential for recognition at a low resolution and from a distance. CRF-RNN was used for body detection from image frame and background subtraction. The Siamese neural networks based on DenseNet features was introduced for gait recognition.

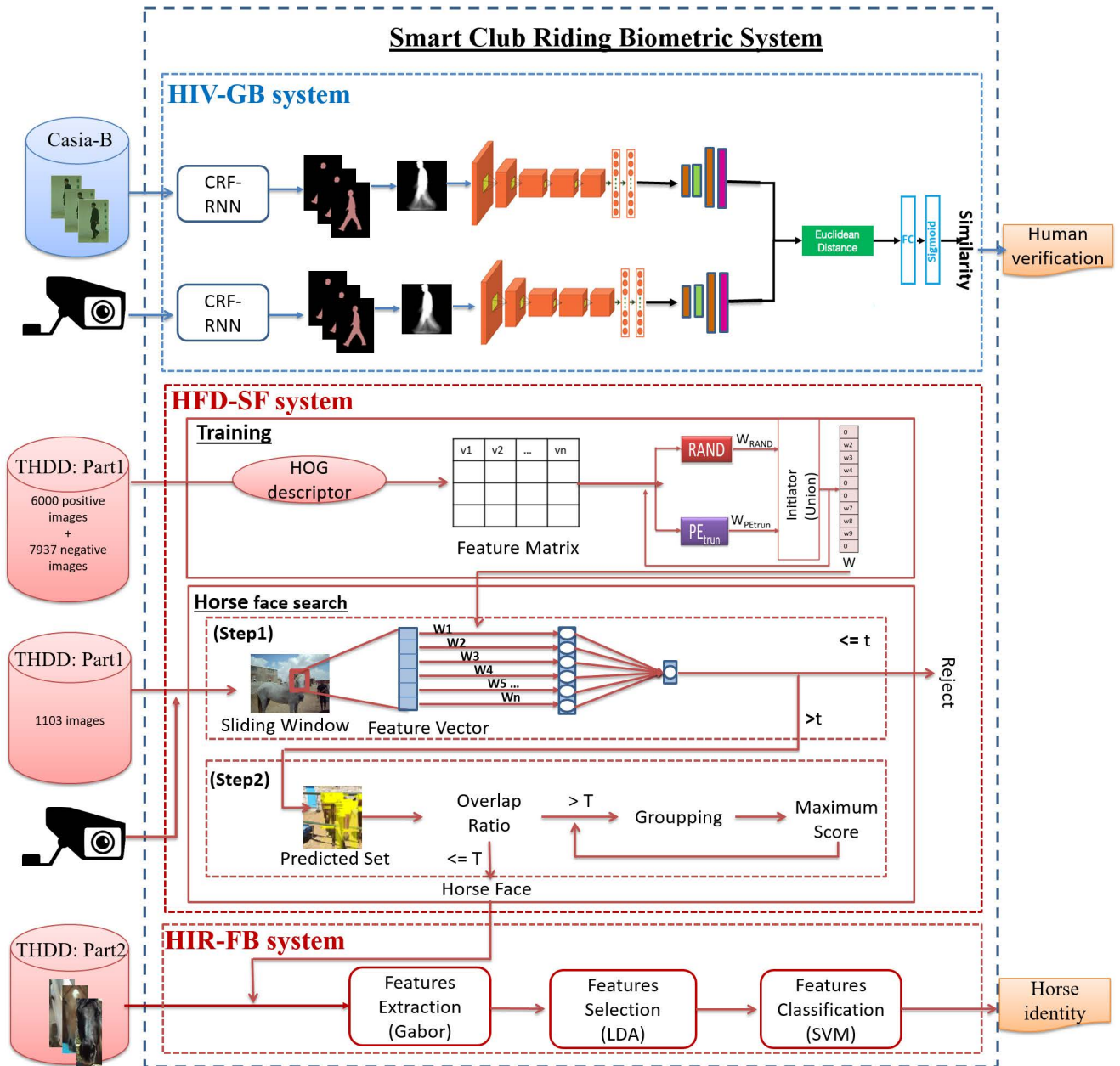


FIGURE 2. The proposed SRCBS: smart riding club biometric system.

1) PRE-PROCESSING (CRF-RNN FOR WORKER AND RIDER DETECTION)

Since CNNs have achieved a great success in natural image analysis and the CRF outperformed other existing solutions in structural learning, the Conditional Random Fields as Recurrent Neural Networks (CRF-RNN) [61] segmentation method has been employed for the human body detection. To describe the deep learning system for semantic image segmentation, it is necessary to understand how repeated iterations are organized as an RNN. One iteration of the algorithm could be formulated as a stack of CNN layers. The transformation done by one CRF-RNN iteration was denoted by f_{θ} using an image Img , pixel-wise unary potential values U and a marginal probability estimation H from the previous

iteration. The next estimation of marginal distributions after one iteration was given by $f_{\theta}(U, H, Img)$. Equations (1), (2) and (3) represent the behaviour of the network where T is the number of iterations while the θ vector represents the CRF parameters [61].

$$H_1(t) = \begin{cases} Softmax(U), & t = 0 \\ H_2(t - 1), & 0 < t \leq T \end{cases} \quad (1)$$

$$H_2(t) = f_{\theta}(U, H_1(t), Img), \quad 0 \leq t \leq T, \quad (2)$$

$$Y(t) = \begin{cases} 0, & 0 \leq t < T \\ H_2(t), & t = T \end{cases} \quad (3)$$

Using CRF-RNN, the output is a number of regions with different classes. The region reference map was obtained and

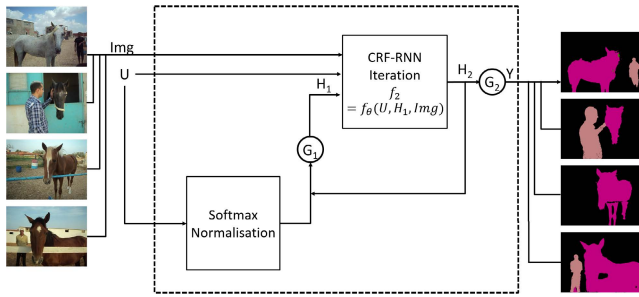


FIGURE 3. The CRF-RNN network.

only the human references have been selected. As shown in Fig. 3, the CRF-RNN of semantic image segmentation was benefited from in order to detect the human boundary area represented by the light pink color.

2) SIAMESE BASED ON DenseNet FEATURES FOR GAIT RECOGNITION

The proposed Siamese neural networks (Fig. 4) employ a unique structure to rank similarity between inputs. The used backbone is composed of a pre-trained DenseNet [21] network followed by a set of fully connected layers. The Euclidian distance, linear and sigmoid layers were placed on the network bottleneck. Once the first part of the backbone was pre-trained and frozen, the Siamese network could then benefit from powerful discriminant features for generalizing the predictive power of the network to recognize not only new data, but also entirely new classes. Using a feature vectors of the pre-trained DenseNet architecture, the system were able to achieve strong results using few parameters and less execution time.

B. HFD-SF SYSTEM: HORSE FACE DETECTION BASED SPARSE FEATURES

The difficult question, at this stage, is how to determine the powerful features and the most useful method to effectively detect horse faces using a small number of parameters and without the need for a huge set of data. Due to the large diversity of horse head textures, it would be a sophisticated task to develop a face detector. Although horses have distinctive ears, this characteristic cannot be focused on because the horse moves them frequently and changes their shape and position in a fascinating way. Therefore, our work concentrates on detecting the horse face without considering its ears. On the other hand, it has been observed that all horses have distinctive head forms, nose and profile eyes. The horse faces have a globally similar shape, but locally variant colors and textures. For this reason, finding out how to effectively use the shape features has been our focus in order to create a robust horse face detector. Based on this idea, a detection method that focuses on the most expressive oriented gradient features has been proposed. In fact, a Sparse Neural Network for horse face detection using sparse gradient features called (SNN) has been suggested.

1) GRADIENT FEATURES FOR HORSE FACE DESCRIPTION

Since gradient features have performed well in several high-level computer vision tasks such as object detection [56], testing their performance on horse face detection is proposed in the present work. This descriptor describes the apparent objects and shapes by estimating the direction of the edges or the intensity distribution [42]. The description was carried out by dividing the image into small adjacent regions, called cells, and by calculating the gradient the directions for each cell in the histogram. The magnitude (Mag) and direction in pixel (x, y) were calculated according to the following equations where I(x,y) is the brightness value of the image in (x,y):

$$d_y = I(x, y - 1) I(x, y + 1) \tag{4}$$

$$d_x = I(x - 1, y) I(x + 1, y) \tag{5}$$

$$Mag(x, y) = \sqrt{d_y^2 + d_x^2} \tag{6}$$

$$\theta(x, y) = \tan^{-1}\left(\frac{d_y}{d_x}\right) \tag{7}$$

For each cell, an oriented gradient feature vector was constructed by quantizing θ into K orientation bins weighted by the gradient magnitude [42]. The overlapped cells were grouped and normalized in order to form a wider spatial region (block). The concatenation of the block histograms formed the gradient descriptor [42]. There were significant challenges in adapting gradient features for horse face detection and selecting the effective ones of the complicated face textures.

2) SPARSE NEURAL NETWORK (SNN) ARCHITECTURE

To deal with the above-listed problems (in the Introduction), the smallest network with a limited number of parameters was suggested for horse face detection. In fact, the proposed Sparse Neural Network is composed of three layers; an input layer (gradient features), a sparse hidden layer and an output layer for feature classification as in the MLP network. The input layer represents the gradient feature vector. The sparse hidden layer had a size equal to the input layer. Each neuron in the input layer had a unique relationship with the opposite neuron in the hidden layer, which reduced the number of parameters. This layer was trained using a proposed sparse feature selection method that will be explained in the next section. In fact, this method produced a weight vector W through the input gradient features. W was used as the weight vector of the hidden layer. The weight vector contained a big number of zeros and therefore all the values corresponding to zero would be falling in the hidden layer giving rise only to the pertinent features that represented 10% of the input vector. The linear activation function was used for the hidden layer (8) where G is the gradient feature vector, n refers to the number of neurons in the input and in the hidden layer, w represents a weight value in the hidden layer and g stands for a gradient feature value in the input layer. The output layer contained only one neural because the classification is binary. This layer was trained by the

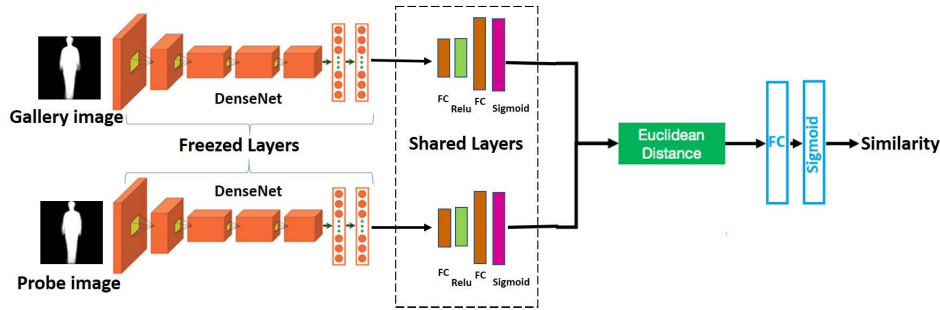


FIGURE 4. The proposed siamese network for human gait recognition.

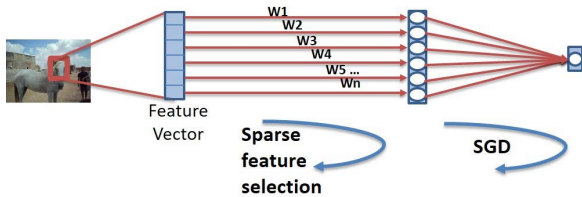


FIGURE 5. Sparse neural network (SNN) architecture.

Stochastic Gradient Descent (SGD) optimizer with momentum using a sigmoid transfer function. The SNN architecture is presented in Fig. 5. The code of this network is available at <https://github.com/Islem-Jarraya/Sparse-Neural-Network-SNN-for-horse-face-classification>

$$f(WG) = \begin{pmatrix} w_1 \\ w_2 \\ \dots \\ w_n \end{pmatrix} \begin{pmatrix} g_1 \\ g_2 \\ \dots \\ g_n \end{pmatrix} = \begin{pmatrix} wg_1 \\ wg_2 \\ \dots \\ wg_n \end{pmatrix} \quad (8)$$

3) SPARSE HIDDEN LAYER TRAINING BASED ON SPARSE FEATURE SELECTION METHOD

In other research works, the feature selection field has been well adapted resorting to many learning methods for pattern recognition applications. However, these are not devoted to object detection applications. The main challenge of feature selection methods is how to reach accurate results using a small number of active features. Despite the efficiency of these methods, they are not accurate enough to process real world data using a small number of features. In this section, a feature selection method that treats this limitation and integrates an automated negotiation process between the PE_{trun} and $RAND$ algorithms has been proposed for binary classification. The input pattern sequence was (x_t, y_t) where x is the input gradient features of d dimension, $t = 1, \dots, T$ refers to the number of iteration and y is the desired output. The sparse selection method required a classifier W_t which contained at most B non-zero elements (where $B > 0$ is a predefined constant). Thus, the classification of x_t depended only on B features and was made by the function: $sgn(W_t x_t)$. The classifier W_t would be updated in each trial t and the learner would classify the instance x_t . This scenario was repeated until $t = T$. It has been assumed that the learner

was provided with full inputs of every training instance. The $RAND$ algorithm which was described and used in [18] randomly selects and picks B features in a learning task. The PE_{trun} algorithm which was described and used in [18] is a perceptron modified by a simple truncation. Both of these algorithms should respect the following 6 steps:

- Step 1: Initializing the weight vector W by zero values
- Step 2: Applying the prediction function: $f_t = W' * x_t$
- Step 3: The feedback y_t ; the real class
- Step 4: Calculating the error: if $y_t * f_t \leq 0$, $err_{count} = err_{count} + 1$
- Step 5: Updating the weight vector: $W = W + y_t * x_t$
- Step 6: Selecting relevant features

Since our aim is to select only relevant features, unnecessary ones were reduced to zero. In the first step, all input features were considered to be irrelevant. Therefore, the weight vectors of the two algorithms were initialized by zeros. The role of $RAND$ and PE_{trun} algorithms was to keep zeros for irrelevant features and increase the weights of useful features. The first five steps were common between the two $RAND$ and PE_{trun} algorithms. The only difference was within the feature selection manner, which was processed in the final step. The $RAND$ and PE_{trun} algorithms participated both in the negotiation process and tried to select the best features. In fact, our key contribution was to incorporate the automated negotiation between the learning algorithms to improve the classification performance. The error rate was considered as the utility function of each negotiator. Negotiation, in this sense, involves the minimum number of mistakes to select the relevant features with the minimum execution time. The $RAND$ participant sends W_{RAND} while the PE_{trun} sends $W_{PE_{trun}}$ to the initiator. The initiator creates the union of W_{RAND} and $W_{PE_{trun}}$ into a W . The same scenario will be repeated and the participants will take the newly vector W in each iteration (Fig. 6). Algorithm 3 describes the different steps of the initiator upon receiving the W_{RAND} and $W_{PE_{trun}}$ proposals. Algorithm 2 and 1 represent the negotiation of the sparse feature selection method.

C. HIR-FB SYSTEM: HORSE IDENTITY RECOGNITION BASED ON FACIAL BIOMETRICS

According to [28], [29] it can be concluded that the Gabor descriptor is effective for horse face characterization. Gabor

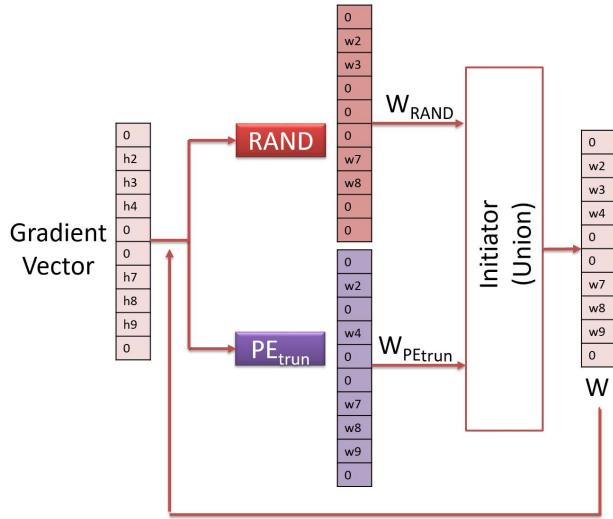


FIGURE 6. The proposed feature selection method.

descriptor has proven its sufficiency in many recent works for face recognition [2], [14], [38], [71]. Jarraya et al. [28] demonstrated that the LDA was better than the PCA for the selection of Gabor features. Multi-class SVM effectiveness for horse face feature classification was proven in [72]. Moreover, Multi-class SVM is effective even in recognizing human faces [70]. Thus, it would be interesting to use these techniques and propose an identification system using the Gabor filters for feature extraction, the LDA for feature selection and multi-class SVM for classification.

IV. THDD DATABASE

To the best of our knowledge, there is no public horse face image pattern database that can be used for the evaluation of detection algorithms. Therefore, the THDD was prepared for the horse face detection and recognition system. THDD is an extension of the THoDBRL’2015 which was used for

Algorithm 1 : Negotiation Based on PE_trun Method

```

Input
B: The number of selected features
Initialization
W_RAND
W_1 = 0
W_PEtrun = 0
1: for t=1, 2, ..., T do
2:   Receive x_t
3:   Make prediction sgn(x_t^T W_t)
4:   Receive y_t
5:   if y_t x_t^T W_t ≤ 0 then
6:     W_{t+1} = W_t + y_t x_t
7:     W_PEtrun = Truncate(W_{t+1}, B)
8:     W_{t+1} = initiator(W_RAND, W_PEtrun)
9:   Send W_PEtrun
    
```

Algorithm 2 : Negotiation Based on RAND Method

```

Input
B: The number of selected features
Initialization
W_PEtrun
W_1 = 0
W_RAND = 0
1: for t=1, 2, ..., T do
2:   Receive x_t
3:   Make prediction sgn(x_t^T W_t)
4:   Receive y_t
5:   if y_t x_t^T W_t ≤ 0 then
6:     W_{t+1} = W_t + y_t x_t
7:     V = Zeros(size(W_{t+1}, 1), 1)
8:     perm_t = randperm(size(W_{t+1}, 1))
9:     C_t = perm_t(1 : B)
10:    V(C_t) = 1
11:    W_RAND = W_{t+1} V
12:    W_{t+1} = initiator(W_RAND, W_PEtrun)
13:   Send W_RAND
    
```

Algorithm 3 : W = initiator(W_RAND, W_PEtrun)

```

Input
W_RAND: The output of the RAND algorithm
W_PEtrun: The output of the PE_trun algorithm
Initialization W = 0
NumFeatures: initial number of features
1: for i=1, 2, ..., NumFeatures do
2:   if W_RAND(i) ≈ 0 then
3:     W(i) = W_RAND(i)
4:   else
5:     if W_PEtrun(i) ≈ 0 then
6:       W(i) = W_PEtrun(i)
    
```

the facial recognition system of the horse. In fact, THoDBRL’2015 became the first part (THDD-part1) of THDD. The second part (THDD-part2) was used for the facial detection system.

The THDD-part1 (First part of THDD) is a multi-view horse face database. The digital images of this dataset were taken at a distance of about 1 meter from the horses when they were in the barn. The capture was achieved by the video camera using a digital still camera of 10.1 Mega Pixels and at a resolution of 640 * 480 pixels. The horse face data were captured from 3 views: frontal view, right view profile and left view profile of the horse. This dataset contains 470 frontal face images for 47 Barbaro, Arabian and hybrid horses (Table 1), 470 left profile images and 470 right profile images. In fact, following the same database construction process of the most related works [4], [6], [52], [53], 10 frontal face images, 10 left profile images and 10 right profile images were taken for each horse.



FIGURE 7. Horse images from the training set of THDD (part1 and part2).



FIGURE 8. Horse images from the test set of THDD-part2.

TABLE 1. Some details about the used database.

Database	Subjects	Views	Breeds	Cite
THDD-part1	47	Frontal Right Left	Barbaro Arabian Hybrid (Barbaro-Arabian)	37
THDD-part2	60	Frontal	Barbaro Arabian Hybrid (Barbaro-Arabian)	-

In this dataset, there are three sets: The first one contains the captured videos, the second set includes selected images and the third one has the cropped images.

The THDD-part2 (Second part of THDD) includes a set of frontal horse images. The digital images were taken at different distances ranging from 1 to 2 meter from horses. The capture was performed by a video camera. Two digital cameras were used. The first is of 10.1 Mega Pixels and at a resolution of 640 * 480 pixels, whereas the second is of 15 Mega Pixels and at a resolution of 12080 * 720 pixels. The collected data set consists of 1103 horse images and 6000 cropped face images for 60 Barbaro, Arabian and hybrid horses (Table 1). In fact, there were two sets: The first one which was for horse face classification includes 6000 positive images and 7937 negative images while the second set which was for horse face detection includes 1103 images. The first set was used as a training set and the second one was employed as a test set in this paper. Most of these animal images had near frontal view. Fig. 7 shows some sample images from the training set of the database. Fig. 8 shows some sample images from the test set of the database.

A. DATA COLLECTION

1) HORSES IN MOTION

The animal is not sane and it is impossible to fix its head and keep its stability. It may change its place and position at any time. These changes of place spoil the distance of 1



FIGURE 9. An illustration of luminance variation in three horses howeverrom THDD.



FIGURE 10. An illustration of background variation in some horses from THDD.

meter (for THDD-part1) and 1-2 meters (for THDD-part2) between the camera and the animal. In order to fix the head of the animal as much as possible to obtain adequate images, the capture of horse videos was done when they were in barns. The movements of the horse were reduced but did not disappear. Thus, the distance between the horse face and the camera slightly changed on the two databases.

2) NATURAL CONDITIONS

Videos were taken from four equestrian centers in Sfax (a city in the east coast of Tunisia) at daylight. The camera was hand-held and positioned in front of the photographer eyes at a distance of about 10 cm from his face. In order to guarantee a maximum fixation of the animal head and to obtain adequate images, the videos were taken when the horses were in barns in natural conditions and without any pressure.

There was a change of the lighting in the horse face according to the position of the animal head and the sun. In addition, there were shadows such as the shadow of the walls or the leaves (Fig. 9). The background also varied in each of the captured videos (Fig. 10).

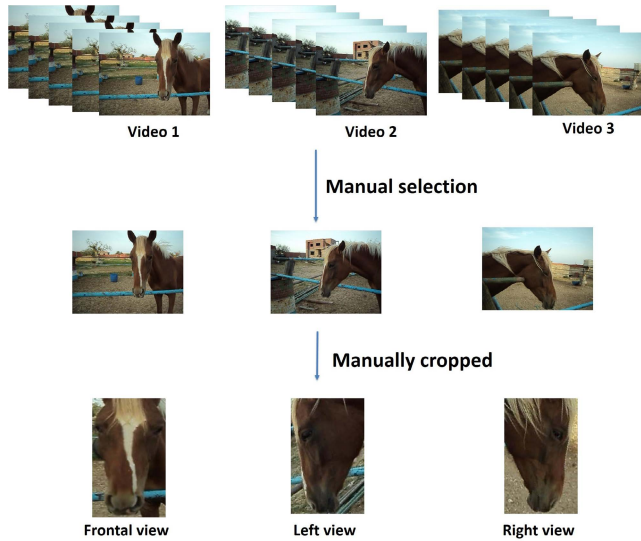


FIGURE 11. THDD-part1 collection procedure.



FIGURE 12. THDD-part2 collection procedure.

B. DATA PROCESSING PROCEDURE

As the horse cannot be fixed or kept stable, a capture of a video for each horse of about 50 seconds was opted for in order to obtain adequate poses. Based on human observation, the best photo was selected with different poses, background and luminance. Selected images were neither 100% front nor 100% profile. The view could be inclined to the right or to the left. Consequently, the facial images of our database had almost the same size and resolution. The difference was not very large.

In THDD-part1, each face area was manually cropped of the selected image for three views (Fig. 11).

In THDD-part2, each face area was manually detected with four ROIs (Xmin, Ymin, Width, Height) of the selected image (Fig. 12).

V. EXPERIMENTAL STUDY OF SRBCS

A. IMPLEMENTATION DETAILS

1) DenseNet

200 epochs were used to train the DenseNet121² network with an input size of 224×224 .

2) GRADIENT FEATURES

According to the experiment, nine unsigned orientations, a cell size equal to 8×8 and a grouped block including 2×2 cells were used.

²<https://pytorch.org/vision/main/models/generated/torchvision.models.densenet121.html>

3) SPARSE HIDDEN LAYER OF SNN

A small fraction of features equal to 10% of the feature dimensions was selected as proposed in [18]. This fraction is enough to find favorable prediction results [18]. Indeed, 90% of the weight vector W of the hidden layer were equal to zero.

4) INTERSECTION OVER UNION (IoU)

Following the Pascal challenge [35], A detected bounding box was considered as a true positive detection only when the Intersection over Union (IoU) ratio was equal to or larger than 50%. For a more accurate evaluation, the metrics proposed by COCO³ challenge were used. 10 different IoU thresholds were considered from 0.5 to 0.95 in steps of 0.05. The average precision was calculated over 10 IoU.

5) BLOCK GENERATION FOR DETECTION PROCESS

A very popular searching strategy has been proposed in [47] to detect face instances in the image. A sequential scan of all possible regions in the image was done by a sliding window. The highly accurate real-time human frontal face detector presented by [46] and [47] used the sliding window strategy. This technique is still used for its sufficiency in many recent detection works [33], [49], [78]. In this paper, this detection strategy was chosen owing to its interesting performance.

The digital images of the database were taken at distances ranging from 1 to 2 meter from the horses. Hence, the camera was considerably not too far from the animal and, consequently, the horse face region was a bit wide in the image. Thus, the horse face size in the image varied between 80×155 to 360×640 . Taking this characteristic into account, the first stage consisted in scanning the entire image with a resized window (from 80×155 to 360×640) with 8 pixels in width, 20 pixels in height and 10 pixels stride. To control the presence of the horse face in each window, the output value of the SNN was monitored. Each window classified as face, would be kept to collect all predicted windows. The predicted windows were filtered applying the Non-Maximum Suppression algorithm (NMS).

6) GABOR DESCRIPTOR

The same parameters used in [28] and [29] were employed.

7) MULTI-CLASS SVM

The linear kernel of the multi-class SVM was used.

B. EVALUATION METRICS

The Receiver Operating Characteristic (ROC), the Cumulative Match Characteristic (CMC) and the precision-recall curves and the classification, recognition and verification rates were recorded using different metrics such as accuracy, precision, average precision, recall, sensitivity, specificity, negative Predictive value and F1 Score. The accuracy (ACC)

³<https://competitions.codalab.org/competitions/12061>

is calculated by dividing the total number of two correct predictions ($T_p + T_n$) by the total images number of a dataset (N):

$$ACC = \frac{T_p + T_n}{N} \quad (9)$$

The recognition rate (RR) is the same as the accuracy (ACC) with the T_n equal to zero and the T_p is the number of recognized images.

Verification Rate (VR) is calculated at a False Acceptance Rate (FAR). False Acceptance Rate is a unit used to determine the accuracy level of a biometric security system. The FAR is calculated by dividing the number of false acceptances by the number of identification attempts (FA is the number of False Acceptances and TA is the Total number of Attempts):

$$FAR = \frac{FA}{TA} \quad (10)$$

The precision (P or PPV) calculates the percentage of the detector predictions which are correct. Precision is determined when dividing the number of true positives (T_p) by the number of true positives plus the number of false positives (F_p):

$$P = \frac{T_p}{T_p + F_p} \quad (11)$$

Recall or sensitivity (R , TPR) measures how good the detector finds all the positives. Recall is calculated by dividing the number of true positives (T_p) by the number of true positives plus the number of false negatives (F_n):

$$R = \frac{T_p}{T_p + F_n} \quad (12)$$

Specificity is also called true negative rate (TNR or SPC). It is calculated by dividing the number of correct negative predictions by the total number of negatives:

$$TNR = \frac{T_n}{T_n + F_p} \quad (13)$$

The negative predictive values (NPV) refer to the proportions of negative results in statistical tests which are true negative results. A strong result can be interpreted using this statistic. The NPV is defined as follows:

$$NPV = \frac{T_n}{T_n + F_n} \quad (14)$$

F1-score ($F1$) can be useful, but it is less frequently used than the other basic measures. F1-score is a harmonic mean of precision and recall:

$$F1 = \frac{2 * P * R}{P + R} \quad (15)$$

The Intersection over Union IoU ratio is computed as a ratio between the intersection and the union of the predicted bounding box and the ground-truth bounding boxes:

$$IoU = \frac{Area\ of\ Intersection}{Area\ of\ Union} \quad (16)$$

Following the Pascal VOC challenge [35], every true positive detection has an IoU ratio equal or larger than 50%.

The precision-recall curve introduces the relation between the precision and the recall calculated for different detection thresholds. Consequently, the area under the precision recall curve presents the average precision (AP) of the detector.

A Receiver Operating Characteristic curve (ROC curve) illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. A ROC curve plots the relationship between the true-positive rate (TPR) of detection or T_p rate and the false positive rate (FPR) of error or F_p rate at various threshold settings. The ROC curve is a graph with:

$$\text{The } x \text{ axis showing } FPR = \frac{F_p}{F_p + T_n} \quad (17)$$

$$Acc_k = R \quad (18)$$

Cumulative Matching Characteristics (CMC) curves are a popular assessment measures for identification methods. Assuming there is only one instance for each identity, for each query, the classification algorithm will classify all samples based on their distances to the query from small to large. The CMC top-k precision is:

$$Acc_k = \begin{cases} 1 & \text{Query} \in \text{top}_k \\ 0 & \text{otherwise} \end{cases} \quad (19)$$

The CMC curve is calculated by averaging the shifted step functions over all queries.

C. EXPERIMENTAL RESULTS

1) WRIR-GB SYSTEM

The proposed Siamese network consists of two shared streams with a 224×224 input size of GEI image. SGD was selected as an optimizer to train the DenseNet network while ADAM was used for the bottleneck layers. The models were trained with NVIDIA GeForce GPU and a 18 GB memory.

CASIA-B dataset [13] is a popular public gait dataset. It was established at the Institute of Automation of Chinese Academy of Sciences (CASIA) in 2005. It contains 93 males and 31 females. The total number of subjects is 124. For each subject there are 10 states, including 6 states of normal walking (NM), 2 states of walking with a bag (BG) and 2 states of walks with a coat (CL). In addition, for each state, every subject has 11 video sequences in 11 angles of view $0^\circ, 18^\circ, \dots, 180^\circ$.

In order to evaluate the proposed system, our experimental framework was applied on the same partitions used in related works. The first 74 subjects were placed in the training set and the other 50 subjects were placed in the test set. In the test set, the gallery set consisted of the first 4 normal walking sequences, and the probe set consisted of the last 2 normal walking sequences.

Table 2 shows comparative results with different research studies. It is noted that our approach was a competitor to the other approaches and achieved a high performances. The

TABLE 2. A comparison between related work results on CASIA-B dataset.

Probe Angle NM #5-6	0°	90°	180°
GaitSet [24]	90.8%	91.7%	85.8%
3DLocal [79]	96.0%	94.2%	95.2%
CSTL [74]	97.8%	95.2%	96.7%
GLFE+GLConv [10]	96.0%	95.4%	94.0%
GaitGraph [64]	85.3%	86.5%	81.9%
GaitSet [23]	91.1%	94.5%	88.0%
STC-Att [67]	97.0%	97.3%	96.0%
GaitTAKE [27]	96.7%	96.3%	96.4%
LagrangeGait [62]	95.2%	94.6%	91.5%
RCC-PFL [75]	87.4%	82.6%	80.3%
Siamese (ours)	95%	100%	98.90%

proposed WRIR-GB has proven its performance in the experimental part for the three views needed in the equestrian club. In fact, the results were very encouraging with a recognition rate equal to 95.0% in the 0° view, 100% in the 90° view and 98.90% in the 180° view. The proposed architecture obtained not only good results but also presented a small number of parameters equal to 1 million in training. This small number of parameters allowed for quick execution which ensured its use in online recognition in a smart riding club.

2) HFD-SF SYSTEM

In the literature, the detection procedure usually includes three steps: block generation (multi-scale sliding windows or region proposals), face classification (in the backbone of the detector) and post-processing (non-maximum suppression and bounding box regression). In fact, the performance of face detectors is mainly influenced by the face classification network also known as the backbone. Duan et al. [32] discovered that the detector and the classifier of the general object detection have comparable performances using the same backbone. This explains that the designed backbone for the classification dataset is applied easily to the general object detection which gets an excellent mAP (mean Average Precision) score. Therefore, it is necessary to evaluate the proposed SNN and the HFD-SF system for horse face classification and detection on the THDD-part2.

a: CLASSIFICATION EVALUATION

For a good evaluation, the proposed SNN was compared with other CNNs such as MobileNetV2 [39] and GoogLeNet [16] which have represented the smallest number of parameters as well as ResNet-50 [60] which has been widely used for human face detection [13], [68]. In fact, MobileNetV2 contains 3.4 million parameters, GoogLeNet contains 5 million parameters and ResNet-50 contains 25 million parameters. CNNs need a huge data for training to perform at its best. During the training of ResNet-50, GoogLeNet and MobileNetV2, three types of data augmentation were used. Reflection, X and Y axis translation, and random scaling were applied to the images of the THDD-part2 training set. The training was done with a mini-batch size equal to 10. A transfer learning of the three pre-trained CNNs (on ImageNet) was made for

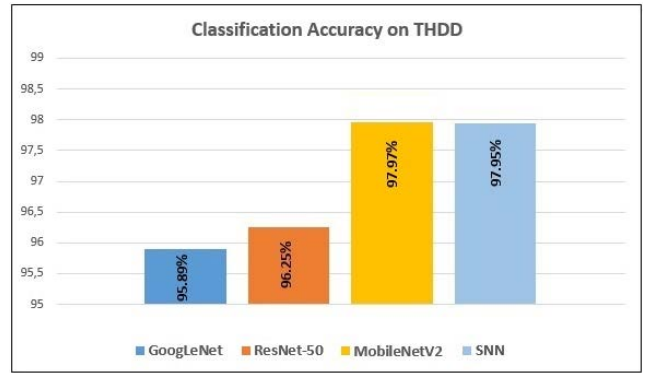


FIGURE 13. Classification results on THDD-part2 using SNN, GoogLeNet, MobileNetV2 and ResNet-50.

classification on the training set of the THDD-part2. The last Fully Connected (FC) layer in these CNNs was replaced with another FC layer having two outputs (face/non-face). The classification was made on the test set of the THDD-part2 which contains 1124 horse faces and 10000 negative images randomly selected from the background.

The comparison includes the accuracy of classification and the number of parameters of each network. Fig. 13 shows the accuracy of the classifications. It is very obvious that the accuracy of SNN was very large compared to GoogLeNet and ResNet-50. The accuracy of MobileNetV2 and SNN is almost the same (Table 3). SNN presents competitive results with a small number of parameters equal to 76.928.

b: DETECTION EVALUATION

Since human and horse faces share similar structures (two eyes, nose and mouth), starting with the existing human frontal face detection approaches has been proposed. Unfortunately, applying these approaches directly on horses met some obvious difficulties. In fact, the horse faces have large appearance variations and intricate textures compared to the human face. Due to the small amount of data in our database, CNN human face detection methods could not be used as they require a large number of images for effective training. Because there are no related works to compare the results with, our approach was contrasted with the Viola-Jones detector that used for cat face detection by the authors in [43]. Our detector was found to achieve a higher performance. The Average Precision of the two detectors (HFD-SF and Viola-Jones) was 90% and 70% respectively. The detection results were compared (shown in Table 4) using two kinds of features, gradient features and intensity features (using LBP features). The average precision and recall of the two descriptors (gradient, intensity) using the proposed detection system were (90%, 90.39%) and (71.40%, 83.27%) respectively. Fig. 16 shows precision-recall curves on the THDD-part2. This figure represents the performance of the proposed detector HFD-SF on the THDD-part2 and reports that the proposed method had the biggest critical region. The Haar (Fig. 14) and weighted LBP features (Fig. 16) gave the

TABLE 3. Classification rates of different networks on THDD-part2.

CNN	ACC	PPV	TPR	SPC	NPV	F1
ResNet-50	96.25%	95.61%	65.93%	99.66%	96.30%	78.04%
GoogLeNet	95.89%	97.99%	60.59%	99.86%	95.75%	74.88%
MobileNetV2	97.97%	90.35%	89.15%	98.96%	98.78%	89.88%
SNN	97.95%	89.02%	90.93%	98.74%	98.98%	89.96%

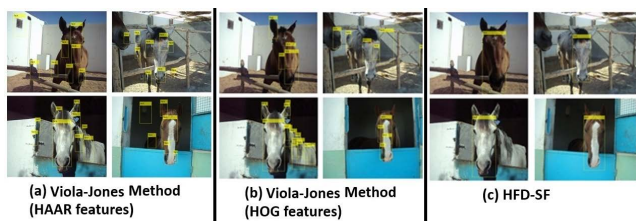


FIGURE 14. Comparison between the detection results of Viola-Jones method [46] and the proposed HFD-SF.



FIGURE 15. Detection results on the THDD-part2.

TABLE 4. A comparison between the results of the gradient features and the intensity features.

Features	AP	Recall	Precision	F1
Viola-Jones(HOG)	70%	-	-	-
HFD-SF(Intensity)	71.40%	83.27%	-	-
HFD-SF(Gradient)	90%	90.39%	63.50%	84.87%

poorest performance because of the large texture variations and shapes of the horse head. With the help of the proposed SNN, the performance has been improved (Fig. 15). Using the evaluation metrics of the COCO challenge, Fig. 17 displays precision-recall curves calculated at 10 different IoU thresholds. Using the proposed SNN, the average precision varied between 84% and 90% for the first four IoU thresholds. To reduce the search time for the horse faces in the image, the CRF-RNN was applied for regions selection to select the horse bodies. The face search was restricted to the selected area only. The experiment showed that the search with Sliding Window (SW) gave better results (Fig. 16).

The proposed SNN proved its performance on the last experiments for classification and detection. Fig 13 presents the classification accuracy on the THDD-part2. It is very obvious that the SNN gave competitive results compared to the other networks. In fact, the SNN overcame the accuracy of the other CNNs by about 1.25% on the THDD-part2 (Table 3).

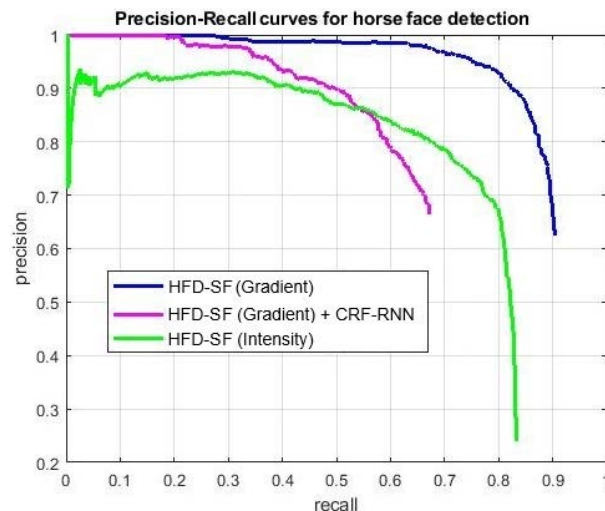


FIGURE 16. Precision-Recall curves on THDD-part2.

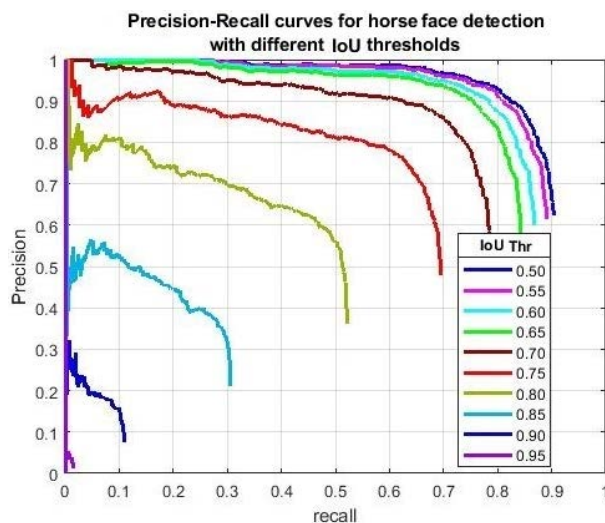


FIGURE 17. Precision-Recall curves calculated at various IoU thresholds according to the COCO challenge.

The same thing was noticed for the detection process. HFD-SF based on SNN gave encouraging results with an Average Precision equal to 90%.

The use of a sparse feature selection method as a learning algorithm enhanced the information transmitted to the output layer. In fact, the proposed sparse hidden layer and training methodology contributed to a proper distinction between true and false detection (face/non-face). SNN extracted the relevant features using the sparse hidden layer and then



FIGURE 18. Samples of frontal face images from the training set of four horses; (a) first horse, (b) second horse, (c) third horse, (d) fourth horse.

classified the candidate bloc using the output layers. Unlike the other networks, SNN kept as much information as possible by minimizing the number of operations and parameters. Consequently, the sparse hidden layer positively influenced decisions and brought detection closer to reality.

Owing to the photos taken close to the pets in the used databases, the faces of the animals are not very small and the system easily detect and recognize them. Indeed, the proposed system cannot detect very small faces when the animal present in the photo is very far away. This was due to the poor resolution of the facial area as well as the lack of important details and information. However, the performance of the sparse feature selection method decreases as the amount of useful information is reduced. The more information there is, the fairer the SNN does the classification.

3) HIR-FB SYSTEM

Among the THDD-part1 images, only the frontal face images were interesting and seven face images were chosen per horse (329 images in total) for the training step of the facial recognition system. To better verify the performance and the efficiency of the HFD-SF system in SRCBS and achieve an end-to-end recognition system, an investigation on horse identity recognition has been integrated based on our detection algorithm results. Using the proposed HFD-SF system, faces were detected on 3 images from the test set of THDD-part2 for each horse (47 horses of the THDD-part1). These images were used for evaluation on the HIR-FB system. In fact, three face images per horse (141 images in total) were used as test images.

In the identification of applications, the camera is fixed and the background was static. It is easy to take a background model and identify objects of interest by detecting changes in the background. In our system, the case was different. In fact, the capture was done in natural conditions with varied background. Because of these reasons and the variety of luminance in our database, the facial area of the horse was cropped with Matlab from the original images of the database and resized (the same image size 160 * 380) as shown in Fig. 18.

Table 5 shows that the proposed recognition system gave the highest performance in two processes: Verification and Identification with interesting results compared to the results of the related works. Fig. 19 of ROC curves and Fig. 20 of CMC curves disclose that the HIR-FB had the biggest critical

TABLE 5. Evaluation of the performance of HIR-FB system on THDD-part1.

Approach	Rank-1 RR	VR at FAR=0.001	VR at FAR=0.01	VR at FAR=0.1
Gabor + PCA + Euc [28]	82.98%	77.19%	92.10%	99.08%
Gabor + PCA + MC [28]	80.85%	77.19%	89.14%	99.09%
Gabor + LDA + Euc [28]	95.74%	95.19%	99.08%	99.08%
Gabor + LDA + MC [28]	95.74%	95.19%	99.08%	99.08%
Gabor + SA [29]	94.22%	-	-	-
RNGLBP + SVM [72]	98.77%	99.08%	100%	100%
HIR-FB	99.89%	100%	100%	100%

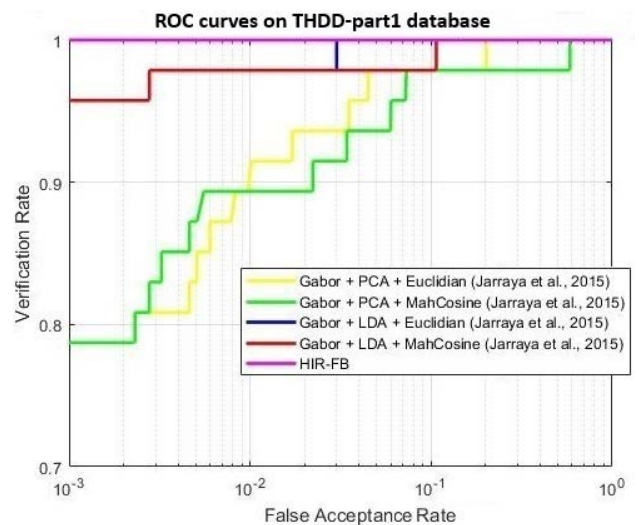


FIGURE 19. ROC curves of HIR-FB system and the approaches of [28] on THDD-part1.

region compared to the others. HIR-FB recorded an 8.5% and a 6.62% improvement in the identification and verification rates respectively. It is worth mentioning that there had already been a suitable horse face recognition system in [28], [29], and [76]. However, their previously published results were not as prominent as the present results.

For more evaluation, the proposed approaches were compared to some famous and standard systems of face recognition including VGG-Face [44], SphereFace [69], ArcFace [30] and RiseNet [76].

VGG-Face [44] is a convolutional neural network composed of 37 layers including input, convolutional, relu, max pooling and softmax layers.

SphererFace [69] is based on a convolutional neural network of 64 layers including convolutional, fully connected and softmax layers with an angular softmax loss (A-Softmax) that enables the CNN to learn angularly discriminative features. ArcFace (Additive Angular Margin Loss) was proposed by [30] to get highly discriminative features for human face

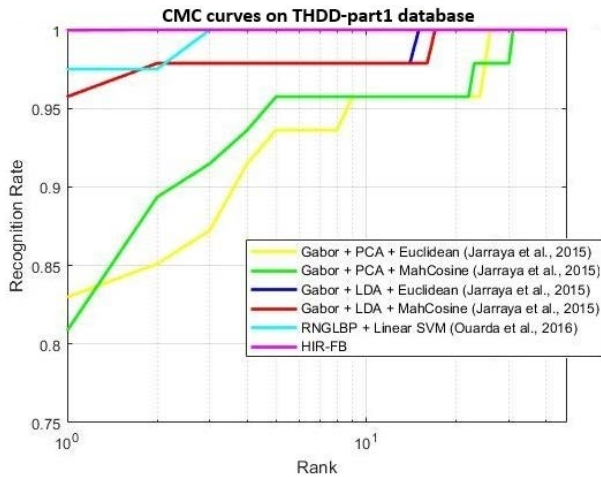


FIGURE 20. CMC curves of HIR-FB system and the approaches of [28], [72] on THDD-part1.

TABLE 6. Comparison between the horse face recognition rates of the proposed approaches and the Related works.

Ref	Methods	Accuracy
[44]	VGG Face	72.62%
[69]	SphererFace	74.98%
[30]	ArcFace	75.32%
[76]	Fine-tuning of Arcface	81.79%
[76]	RiseNet	82.56%
HIR-FB	Gabor+LDA+SVM	99.89%

recognition. ArcFace is incorporated within CNN architectures such as ResNet-50 and ResNet-100.

RiseNet proposed by [76] is an animal face recognition framework.

The above-listed systems are based on pre-trained networks on the ImageNet [31]. A transfer learning on the THDD-part1 obtained the results shown in Table 6. This table proves the superiority of the proposed approach based on Gabor features, LDA and SVM. The proposed HIR-FB proved its performance on the experimental part for the classification and verification processes. In fact, the results were very encouraging with a recognition rate equal to 99.89%. The proposed system improved the recognition rate by 22.44% in comparison with the standard systems for face recognition [30], [44], [69], [76]. According to these results, facial pattern can be considered as a good biometric marker for horse identification.

The multi-class SVM was used to separate the data in n classes by optimal hyperplanes. It uses different kernels for feature transformation in a new helpful representation to facilitate the optimization of the margin by reducing feature complexity. The choice of the best and simplest kernel was proposed in this work to easily separate the feature classes. The use of the multi-class SVM based on a linear kernel is justified by the linearity of data using the Linear Discriminant Analysis (LDA) which represents the features in a new space based on the projection of eigenvectors. In addition, the

Gabor features and LDA proved their efficiency for horse face recognition.

VI. CONCLUSION

In this paper, a new biometric system, SRCBS, was introduced for smart riding club security. This system is based on three main sub-systems to detect and recognize both humans and horses.

In order to obtain an efficient SRCBS system, with supplementary services for humans, our paper has proposed a new method for human gait recognition based on Siamese network. Using the gait modality, the CRF-RNN for human body detection and the proposed Siamese network, the recognition rates were equal to 95% in the 0° view, 100% in the 90° view and 98.90% in the 180° view on Casia-B database. These results were better than the recognition rates of the related works.

The proposed HFD-SF system for horse face detection is based on the proposed network SNN. The SNN gave competitive classification results compared to the other networks. In fact, the SNN overcame the accuracy of the other CNNs by about 1.25%. This detection system achieved encouraging results. The experiments on the THDD-part2 showed that our system is efficient by reaching a useful detection rate equal to 90%. The use of a sparse feature selection method as a learning algorithm enhanced the information transmitted to the output layer. In fact, the proposed sparse hidden layer and training methodology contributed to proper distinction between true and false detections (face/non-face). Unlike the other networks, SNN kept as many information as possible using a smaller number of parameters equal to 76928.

The proposed HIR-FB for horse face recognition proved their performance by a recognition rate equal to 99.89% on the THDD-part1. In fact, HIR-FB enhanced the recognition rate by 22.44% in comparison with the standard systems for horse face recognition. Thus, the facial pattern can be considered as a good biometric marker for horse identification according the obtained results.

The THDD was prepared for the horse face detection and recognition system. To the best of our knowledge, there is no public horse face image pattern database that can be used for the evaluation of detection algorithms. Therefore, the THDD could contribute to the research community of the animal biometrics. In fact, this is the only dataset of public face image that is available for research on horse face detection and recognition.

Our future perspectives can be summarized as follows:

- Developing the SRCBS system for horse and human detection and identification at real time.
- Expanding the proposed HFD-SF system in two directions. First, it is important to improve its performance and effectiveness by designing more discriminant features. Second, extending the HFD-SF system to other animals is among our plans for the future.

ACKNOWLEDGMENT

The authors would like to acknowledge that this research was supported by four Riding Clubs in Sfax, Tunisia: Equestrian Clubs of Road Mahdia, Road Tunis in Sakit Ezzit, Road Ain Km 17, and Road Saltnia Km 17.

REFERENCES

- [1] (1999). *LOI N 99-574 DU 9 Juillet 1999 d'Orientation Agricole, JORF n°0158*. [Online]. Available: <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000395813/>
- [2] A. Boughida, M. N. Kouahla, and Y. Lafifi, "A novel approach for facial expression recognition based on Gabor filters and genetic algorithm," *Evolving Syst.*, vol. 2021, pp. 331–345, Jul. 2021, doi: [10.1007/s12530-021-09393-2](https://doi.org/10.1007/s12530-021-09393-2).
- [3] A. Hitelman, Y. Edan, A. Godo, R. Berenstein, J. Lepar, and I. Halachmi, "Biometric identification of sheep via a machine-vision system," *Comput. Electron. Agricult.*, vol. 194, Mar. 2022, Art. no. 106713, doi: [10.1016/j.compag.2022.106713](https://doi.org/10.1016/j.compag.2022.106713).
- [4] A. I. Awad, H. M. Zawbaa, H. A. Mahmoud, E. H. H. A. Nabi, R. H. Fayed, and A. E. Hassanien, "IA robust cattle identification scheme using muzzle print images," in *Proc. Federated Conf. Comput. Sci. Inf. Syst.*, 2013, pp. 529–534.
- [5] A. Salama, A. E. Hassanien, and A. Fahmy, "Sheep identification using a hybrid deep learning and Bayesian optimization approach," *IEEE Access*, vol. 7, pp. 31681–31687, 2019, doi: [10.1109/ACCESS.2019.2902724](https://doi.org/10.1109/ACCESS.2019.2902724).
- [6] A. Tharwat, T. Gaber, and A. E. Hassanien, "Cattle identification based on muzzle images using Gabor features and SVM classifier," in *Proc. Int. Conf. Adv. Mach. Learn. Technol. Appl.*, vol. 488, 2014, pp. 236–247, doi: [10.1007/978-3-319-13461-1_23](https://doi.org/10.1007/978-3-319-13461-1_23).
- [7] A. Tureckova, T. Holik, and Z. K. Oplatkova, "Dog face detection using YOLO network," *Mendel Soft Comput. J.*, vol. 26, pp. 2571–3701, Jan. 2020, doi: [10.13164/mendel.2020.2.017](https://doi.org/10.13164/mendel.2020.2.017).
- [8] A. Vlachyńska, Z. K. Oplatkova, and T. Turecek, "Dogface detection and localization of dogface's landmarks," in *Proc. Comput. Sci. Line Conf.*, 2018, pp. 465–476, doi: [10.1007/978-3-319-91189-2_46](https://doi.org/10.1007/978-3-319-91189-2_46).
- [9] A. Yamada, K. Kojima, J. Kiyama, M. Okamoto, and H. Murata, "Directional edge-based dog and cat face detection method for digital camera," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2011, pp. 87–88, doi: [10.1109/ICCE.2011.5722916](https://doi.org/10.1109/ICCE.2011.5722916).
- [10] B. Lin, S. Zhang, and X. Yu, "Gait recognition via effective global-local feature representation and local temporal aggregation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2021, pp. 14648–14656.
- [11] B. Xu, W. Wang, L. Guo, G. Chen, Y. Li, Z. Cao, and S. Wu, "CattleFaceNet: A cattle face identification approach based on RetinaFace and ArcFace loss," *Comput. Electron. Agricult.*, vol. 193, Feb. 2022, Art. no. 106675, doi: [10.1016/j.compag.2021.106675](https://doi.org/10.1016/j.compag.2021.106675).
- [12] B. Xu, W. Wang, L. Guo, G. Chen, Y. Wang, W. Zhang, and Y. Li, "Evaluation of deep learning for automatic multi-view face detection in cattle," *Agriculture*, vol. 11, no. 11, p. 1062, 2021.
- [13] B. Zhang, J. Li, Y. Wang, Y. Tai, C. Wang, J. Li, F. Huang, Y. Xia, W. Pei, and R. Ji, "ASFD: Automatic and scalable face detector," 2020, *arXiv:2003.11228*.
- [14] C. Li, Y. Huang, W. Huang, and F. Qin, "Learning features from covariance matrix of Gabor wavelet for face recognition under adverse conditions," *Pattern Recognit.*, vol. 119, Nov. 2021, Art. no. 108085, doi: [10.1016/j.patcog.2021.108085](https://doi.org/10.1016/j.patcog.2021.108085).
- [15] C. Song, Y. Huang, Y. Huang, N. Jia, and L. Wang, "GaitNet: An end-to-end network for gait based human identification," *Pattern Recognit.*, vol. 96, Dec. 2019, Art. no. 106988, doi: [10.1016/j.patcog.2019.106988](https://doi.org/10.1016/j.patcog.2019.106988).
- [16] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2016, pp. 2818–2826, doi: [10.1109/CVPR.2016.308](https://doi.org/10.1109/CVPR.2016.308).
- [17] D. S. Edwards, A. M. Johnston, and D. U. Pfeiffer, "A comparison of commonly used ear tags on the ear damage of sheep," *Animal Welfare*, vol. 10, no. 2, pp. 141–151, 2001.
- [18] F. BenSaid and A. M. Alimi, "Online feature selection system for big data classification based on multi-objective automated negotiation," *Pattern Recognit.*, vol. 110, pp. 107–629, Feb. 2021, doi: [10.1016/j.patcog.2020.107629](https://doi.org/10.1016/j.patcog.2020.107629).
- [19] F. M. S. Bragança, S. Broomé, M. Rhodin, S. Björnsdóttir, V. Gunnarsson, J. P. Voskamp, E. Persson-Sjodin, W. Back, G. Lindgren, M. Novoa-Bravo, A. I. Gmel, C. Roepstorff, B. J. van der Zwaag, P. R. Van Weeren, and E. Hernlund, "Improving gait classification in horses by using inertial measurement unit (IMU) generated data and machine learning," *Sci. Rep.*, vol. 10, no. 1, pp. 1–9, Oct. 2020, doi: [10.1038/s41598-020-73215-9](https://doi.org/10.1038/s41598-020-73215-9).
- [20] F. Xu, J. Gao, and X. Pan, "Cow face recognition for a small sample based on Siamese DB capsule network," *IEEE Access*, vol. 10, pp. 63189–63198, 2022, doi: [10.1109/ACCESS.2022.3182806](https://doi.org/10.1109/ACCESS.2022.3182806).
- [21] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2261–2269, doi: [10.1109/CVPR.2017.243](https://doi.org/10.1109/CVPR.2017.243).
- [22] G. Mougeot, D. Li, and S. Jia, "A deep learning approach for dog face verification and recognition," in *Proc. Pacific Rim Int. Conf. Artif. Intell.*, 2019, pp. 418–430, doi: [10.1007/978-3-030-29894-4_34](https://doi.org/10.1007/978-3-030-29894-4_34).
- [23] H. Chao, K. Wang, Y. He, J. Zhang, and J. Feng, "GaitSet: Cross-view gait recognition through utilizing gait as a deep set," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 3467–3478, Jul. 2022, doi: [10.1109/TPAMI.2021.3057879](https://doi.org/10.1109/TPAMI.2021.3057879).
- [24] H. Chao, Y. He, J. Zhang, and J. Feng, "GaitSet: Regarding gait as a set for cross-view gait recognition," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 8126–8133, doi: [10.1609/aaai.v33i01.33018126](https://doi.org/10.1609/aaai.v33i01.33018126).
- [25] H. I. Hummel, F. Pessanha, A. A. Salah, T. J. P. A. M. van Loon, and R. C. Veltkamp, "Automatic pain detection on horse and donkey faces," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, Nov. 2020, pp. 793–800, doi: [10.1109/FG47880.2020.00114](https://doi.org/10.1109/FG47880.2020.00114).
- [26] H. Jiang and E. Learned-Miller, "Face detection with the faster R-CNN," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, May 2017, pp. 650–657.
- [27] H.-M. Hsu, Y. Wang, C.-Y. Yang, J.-N. Hwang, H. L. U. Thuc, and K.-J. Kim, "GAITTAKE: Gait recognition by temporal attention and keypoint-guided embedding," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2022, pp. 2546–2550, doi: [10.1109/ICIP46576.2022.9897409](https://doi.org/10.1109/ICIP46576.2022.9897409).
- [28] I. Jarraya, W. Ouarda, and A. M. Alimi, "A preliminary investigation on horses recognition using facial texture features," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2015, pp. 2803–2808, doi: [10.1109/SMC.2015.489](https://doi.org/10.1109/SMC.2015.489).
- [29] I. Jarraya, W. Ouarda, and A. M. Alimi, "Deep neural network features for horses identity recognition using multiview horses' face pattern," in *Proc. 9th Int. Conf. Mach. Vis. (ICMV)*, vol. 10341, Mar. 2017, pp. 52–56, doi: [10.1117/12.2269064](https://doi.org/10.1117/12.2269064).
- [30] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 4685–4694, doi: [10.1109/CVPR.2019.00482](https://doi.org/10.1109/CVPR.2019.00482).
- [31] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255, doi: [10.1109/CVPR.2009.5206848](https://doi.org/10.1109/CVPR.2009.5206848).
- [32] J. Duan, S. Liao, S. Zhou, and S. Z. Li, "Face classification: A specialized benchmark study," in *Proc. Chin. Conf. Biometric Recognit.*, 2016, pp. 22–29, doi: [10.1007/978-3-319-46654-5_3](https://doi.org/10.1007/978-3-319-46654-5_3).
- [33] L. Kulanuwat, C. Chantrapornchai, M. Maleewong, P. Wongchaisuwat, S. Wimala, K. Sarinnapakorn, and S. Boonya-Aroonue, "Anomaly detection using a sliding window technique and data imputation with machine learning for hydrological time series," *Water*, vol. 13, no. 13, p. 1862, 2021, doi: [10.3390/w13131862](https://doi.org/10.3390/w13131862).
- [34] M. E. Eshaimir, I. Jarraya, W. Ouarda, and A. M. Alimi, "Human gait identity recognition system based on gait Pal and Pal entropy (GPPE) and distances features fusion," in *Proc. Sudan Conf. Comput. Sci. Inf. Technol. (SCCSIT)*, Nov. 2017, pp. 1–5, doi: [10.1109/SCCSIT.2017.8293061](https://doi.org/10.1109/SCCSIT.2017.8293061).
- [35] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Jun. 2010, doi: [10.1007/s11263-009-0275-4](https://doi.org/10.1007/s11263-009-0275-4).
- [36] M. I. Sharif, M. A. Khan, A. Alqahtani, M. Nazir, S. Alsubai, A. Binbusayyis, and R. Damaevius, "Deep learning and kurtosis-controlled, entropy-based framework for human gait recognition using video sequences," *Electronics*, vol. 11, no. 3, p. 334, Jan. 2022, doi: [10.3390/electronics11030334](https://doi.org/10.3390/electronics11030334).

- [37] M. L. George, T. Govindarajan, K. A. Rajasekaran, and S. R. Bandi, "A robust similarity based deep Siamese convolutional neural network for gait recognition across views," *Wiley Online Library*, vol. 36, no. 3, pp. 1290–1319, Aug. 2020, doi: [10.1111/coin.12361](https://doi.org/10.1111/coin.12361).
- [38] M. M. Daisy and P. Kannan, "Investigation of rotated local Gabor features in face recognition using fusion techniques," *J. Ambient Intell. Humanized Comput.*, vol. 12, p. 5895–5908, Jun. 2021, doi: [10.1007/s12652-020-02134-4](https://doi.org/10.1007/s12652-020-02134-4).
- [39] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "Identification of two methylated fragments of an SDC2 CpG island using a sliding window technique for early detection of colorectal cancer," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520, doi: [10.1109/CVPR.2018.00474](https://doi.org/10.1109/CVPR.2018.00474).
- [40] M. Shopon, G.-S. J. Hsu, and M. L. Gavrilova, "Multiview gait recognition on unconstrained path using graph convolutional neural network," *IEEE Access*, vol. 10, pp. 54572–54588, 2022, doi: [10.1109/ACCESS.2022.3176873](https://doi.org/10.1109/ACCESS.2022.3176873).
- [41] M. Zaied, C. B. Amar, and A. M. Alimi, "Beta wavelet networks for face recognition," *J. Decis. Syst.*, vol. 14, pp. 109–122, Apr. 2012, doi: [10.3166/jds.14.109-122](https://doi.org/10.3166/jds.14.109-122).
- [42] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, vol. 1, no. 1, pp. 886–893, doi: [10.1109/CVPR.2005.177](https://doi.org/10.1109/CVPR.2005.177).
- [43] N. Mukai, Y. Zhang, and Y. Chang, "Pet face detection," in *Proc. Nicograph Int. (NicoInt)*, Jun. 2018, pp. 52–57, doi: [10.1109/NICOINT.2018.00018](https://doi.org/10.1109/NICOINT.2018.00018).
- [44] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. Brit. Mach. Vis. Conf.*, 2015, pp. 471–478.
- [45] P. Taunk and G. Jayasri, "Face detection using Viola Jones with Haar cascade," *Test Eng. Manage.*, vol. 83, Jan. 2020.
- [46] P. Viola and M. Jones, "Robust real-time face detection," in *Proc. 8th IEEE Int. Conf. Comput. Vis.*, Oct. 2001, p. 747, doi: [10.1109/ICCV.2001.937709](https://doi.org/10.1109/ICCV.2001.937709).
- [47] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004, doi: [10.1023/B:VISL.0000013087.49260.fb](https://doi.org/10.1023/B:VISL.0000013087.49260.fb).
- [48] R. Liao, Z. Li, S. S. Bhattacharyya, and G. York, "PoseMapGait: A model-based gait recognition method with pose estimation maps and graph convolutional networks," *Neurocomputing*, vol. 501, pp. 514–528, Aug. 2022, doi: [10.1016/j.neucom.2022.06.048](https://doi.org/10.1016/j.neucom.2022.06.048).
- [49] R. Li, B. Qu, K. Wan, C. Lu, T. Li, F. Zhou, and J. Lin, "Identification of two methylated fragments of an SDC2 CpG island using a sliding window technique for early detection of colorectal cancer," *FEBS Open Bio*, vol. 11, no. 7, pp. 1941–1952, Jun. 2021, doi: [10.1002/2211-5463.13180](https://doi.org/10.1002/2211-5463.13180).
- [50] S. Joshi, D. K. Verma, G. Saxena, and A. Paraye, "Issues in training a convolutional neural network model for image classification," in *Proc. Int. Conf. Adv. Comput. Data Sci.*, vol. 1046, Jul. 2019, pp. 282–293.
- [51] S. Kumar, A. Pandey, K. S. R. Satwik, S. S. K. K. Singh, A. K. Singh, and A. Mohan, "Deep learning framework for recognition of cattle using muzzle point image pattern," *Measurement*, vol. 116, pp. 1–17, Feb. 2017.
- [52] S. Kumar and S. Singh, "Biometric recognition for pet animal," *J. Softw. Eng. Appl.*, vol. 7, no. 5, pp. 470–482, May 2014, doi: [10.4236/jsea.2014.75044](https://doi.org/10.4236/jsea.2014.75044).
- [53] S. Kumar, S. Tiwari, and S. K. Singh, "Face recognition for cattle," in *Proc. Int. Conf. Image Inf. Process. (ICIIP)*, 2015, pp. 65–72.
- [54] S. Manoj, S. Rakshith, and V. Kanchana, "Identification of cattle breed using the convolutional neural network," in *Proc. Int. Conf. Signal Process. Commun.*, 2021, pp. 503–507, doi: [10.1109/ICSPC51351.2021.9451706](https://doi.org/10.1109/ICSPC51351.2021.9451706).
- [55] S. North, "Do Androids dream of electric steeds? The allure of horse-computer interaction," *Interactions*, vol. 23, pp. 50–53, Jan. 2016.
- [56] S. Rathor, S. Kumari, R. Singh, and P. Gupta, "Two layers machine learning architecture for animal classification using HOG and LBP," in *Proc. Int. Conf. Commun. Artif. Intell.*, in Lecture Notes in Networks and Systems, vol. 192, V. Goyal, M. Gupta, A. Trivedi, and M. L. Kolhe, Eds., May 2021, pp. 445–453, doi: [10.1007/978-981-33-6546-9_42](https://doi.org/10.1007/978-981-33-6546-9_42).
- [57] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: [10.1109/TPAMI.2021.3057879](https://doi.org/10.1109/TPAMI.2021.3057879).
- [58] S. Sabour, N. Frosst, and G. Hinton, "Dynamic routing between capsules," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 3859–3869.
- [59] S. Song, T. Liu, H. Wang, B. Hasi, C. Yuan, F. Gao, and H. Shi, "Using pruning-based YOLOv3 deep learning algorithm for accurate detection of sheep face," *Animals*, vol. 12, no. 11, p. 1465, Jun. 2022, doi: [10.3390/ani12111465](https://doi.org/10.3390/ani12111465).
- [60] S. Zagoruyko and N. Komodakis, "Wide residual networks," 2016, *arXiv:1605.07146*.
- [61] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. S. Torr, "Conditional random fields as recurrent neural networks," in *Proc. ICCV*, Dec. 2015, pp. 1529–1537, doi: [10.1109/ICCV.2015.179](https://doi.org/10.1109/ICCV.2015.179).
- [62] T. Chai, A. Li, S. Zhang, Z. Li, and Y. Wang, "Lagrange motion analysis and view embeddings for improved gait recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 20217–20226, doi: [10.1109/CVPR52688.2022.019161](https://doi.org/10.1109/CVPR52688.2022.019161).
- [63] T. Kishore and A. Jha, "Deep CNN based automatic detection and identification of Bengal tigers," in *Proc. Int. Conf. Comput. Intell. Commun. Bus. Anal.*, 2021, pp. 503–507.
- [64] T. Teepe, A. Khan, J. Gilg, F. Herzog, S. Hörmann, and G. Rigoll, "Gait-graph: Graph convolutional network for skeleton-based gait recognition," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2021, pp. 2314–2318, doi: [10.1109/ICIP42928.2021.9506717](https://doi.org/10.1109/ICIP42928.2021.9506717).
- [65] U. G. Barron, G. Corkery, B. Barry, F. Butler, K. McDonnell, and S. Ward, "Assessment of retinal recognition technology as a biometric method for sheep identification," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 156–166, Mar. 2008, doi: [10.1016/j.compamp.2007.07.010](https://doi.org/10.1016/j.compamp.2007.07.010).
- [66] W. Kusakunniran, A. Wiratsudakul, U. Chuachan, S. Kanchanapreechakorn, and T. Imaromkul, "Automatic cattle identification based on fusion of texture features extracted from muzzle images," in *Proc. IEEE Int. Conf. Ind. Technol. (ICIT)*, Feb. 2018, pp. 1484–1489, doi: [10.1109/ICIT.2018.8352400](https://doi.org/10.1109/ICIT.2018.8352400).
- [67] W. Likai, H. Ruizhe, C. Jinyan, and F. Wei, "Combining the silhouette and skeleton data for gait recognition," 2022, *arXiv:2202.10645*.
- [68] W. Liu, I. Hasan, and S. Liao, "Center and scale prediction: Anchor-free approach for pedestrian and face detection," 2019, *arXiv:1904.02948*.
- [69] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "SphereFace: Deep hypersphere embedding for face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 11, Jun. 2017, pp. 6738–6746, doi: [10.1109/CVPR.2017.713](https://doi.org/10.1109/CVPR.2017.713).
- [70] W. Ouarda, H. Trichili, A. M. Alimi, and B. Solaiman, "Face recognition based on geometric features using support vector machines," in *Proc. 6th Int. Conf. Soft Comput. Pattern Recognit. (SoCPar)*, Aug. 2014, pp. 89–95, doi: [10.1109/SOCPAR.2014.7007987](https://doi.org/10.1109/SOCPAR.2014.7007987).
- [71] W. Ouarda, H. Trichili, A. M. Alimi, and B. Solaiman, "MLP neural network for face recognition based on Gabor features and dimensionality reduction techniques," in *Proc. Int. Conf. Multimedia Comput. Syst. (ICMCS)*, Apr. 2014, pp. 127–134, doi: [10.1109/ICMCS.2014.6911265](https://doi.org/10.1109/ICMCS.2014.6911265).
- [72] W. Ouarda, H. Trichili, M. A. Alimi, and B. Solaiman, "Towards a novel biometric system for smart riding club," *J. Inf. Assurance Secur.*, vol. 11, no. 4, pp. 201–213, Jan. 2016.
- [73] W. Zhang, J. Sun, and X. Tang, "From tiger to panda: Animal head detection," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1696–1708, Jun. 2011, doi: [10.1109/TIP.2010.2099126](https://doi.org/10.1109/TIP.2010.2099126).
- [74] X. Huang, D. Zhu, H. Wang, X. Wang, B. Yang, B. He, W. Liu, and B. Feng, "Context-sensitive temporal feature learning for gait recognition," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2021, pp. 12909–12918.
- [75] X. Ren, S. Hou, C. Cao, X. Liu, and Y. Huang, "Progressive feature learning for realistic cloth-changing gait recognition," 2022, *arXiv:2207.11720*.
- [76] X. Shi, C. Yang, X. Xia, and X. Chai, "Deep cross-species feature learning for animal face recognition via residual interspecies equivariant network," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 667–682, doi: [10.1109/TPAMI.2021.3057879](https://doi.org/10.1109/TPAMI.2021.3057879).
- [77] Y. Zhang, J. Qin, L. Lv, and Z. Wang, "Based on Siamese network with self-attention model for gait recognition," in *Proc. IEEE Int. Conf. Mechatronics Autom. (ICMA)*, Oct. 2020, pp. 1118–1122, doi: [10.1109/ICMA49215.2020.9233689](https://doi.org/10.1109/ICMA49215.2020.9233689).
- [78] Z. Hang, Y. Wang, and S. Huang, "Accurate heavy hitter detection with sliding window approach," in *Proc. IEEE 6th Int. Conf. Comput. Commun. Syst. (ICCCS)*, Apr. 2021, pp. 512–516, doi: [10.1109/ICCCS52626.2021.9449202](https://doi.org/10.1109/ICCCS52626.2021.9449202).
- [79] Z. Huang, D. Xue, X. Shen, X. Tian, H. Li, J. Huang, and X.-S. Hua, "3D local convolutional neural networks for gait recognition," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 14900–14909, doi: [10.1109/ICCV48922.2021.01465](https://doi.org/10.1109/ICCV48922.2021.01465).

- [80] Z. Li, S. Broomé, P. H. Andersen, and H. Kjellström, "Automated detection of equine facial action units," 2021, *arXiv:2102.08983*.
- [81] Z. Weng, F. Meng, S. Liu, Y. Zhang, Z. Zheng, and C. Gong, "Cattle face recognition based on a two-branch convolutional neural network," *Comput. Electron. Agricult.*, vol. 196, May 2022, Art. no. 106871, doi: 10.1016/j.compag.2022.106871.



ISLEM JARRAYA (Member, IEEE) was born in Sfax, Tunisia, in 1986. She received the degree in computer science and multimedia and the master's degree in computer science and multimedia from the Multimedia, Information Systems and Advanced Computing Laboratory (MIRACL-Laboratory), Higher Institute of Computer Science and Multimedia of Sfax, Tunisia, in 2009 and 2014, respectively. She is currently pursuing the Ph.D. degree in computing system engineering with the Research Group on Intelligent Machines (ReGIM-Laboratory), ENIS. She focuses her research on applying intelligent methods (deep neural networks, feature extraction, and recognition algorithms) to deep machine learning, computer vision, face image processing, intelligent pattern recognition, and analysis of large scale complex systems.



FATMA BEN SAID was born in Sfax, Tunisia, in 1981. She received the degree in applied informatics in management from the Faculty of Economic and Management (FSEG), University of Sfax, Tunisia, in June 2004, the master's degree in computer science, dedicated computer system new technologies (NTSID) from the National School of Engineers (ENIS), University of Sfax, Tunisia, in July 2006, and the Ph.D. degree from the Computing System Engineering, Research Group on Intelligent Machines, ENIS, University of Sfax, in 2017. She is currently pursuing the Ph.D. degree with the REGIM-Laboratory. Her research interests include artificial intelligence, machine learning, feature selection, pattern recognition, and classification.



TAREK M. HAMDANI (Senior Member, IEEE) was born in Tunis, Tunisia, in 1979. He received the M.Sc., Ph.D., and H.D.R. degrees in computer science engineering from the National Engineering School of Sfax, Tunisia, in 2003, 2011, and 2019, respectively. He is currently an Associate Professor in computer science at the University of Monastir, Monastir. He focuses his research on applying intelligent methods (neural networks, fuzzy logic, and evolutionary algorithms) to machine learning, computer vision, natural language processing, intelligent pattern recognition, and analysis of large scale complex systems. He is a Reviewer of the *Pattern Recognition Letters*, *Neurocomputing*, *Neural Processing Letter*, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, and the *Soft Computing* journal. He is a Senior Member of CIS and SMC Societies.



BILEL NEJI was born in Tunisia, in May 1983. He received the joint B.Eng. degree in electrical engineering from the National Engineering School of Sfax, Tunisia, in cooperation with Valenciennes University, France, in 2007, the joint M.S. degree in new technologies in computer systems from the National Engineering School of Sfax, in cooperation with Lille University, France, in 2008, the joint Ph.D. degree in electrical engineering focused on scientific satellite's subsystems from the National Engineering School of Sfax, in cooperation with Wurzberg University, Germany, in 2014, and the Ph.D. degree in electrical engineering

focused on MEMS and micro/nano sensors design and fabrication from the State University of New York at Buffalo, USA, in 2015. He co-founded BAK USA Technologies Corporation, New York, USA, in 2014, and worked as the Director of Engineering, until 2018. He joined the American University of the Middle East, Kuwait, AUM, in September 2018, as an Assistant Professor. He is currently an Associate Professor at AUM, where he works as the Research Office Coordinator of the College of Engineering and Technology. He has been conducting research in different areas, including embedded systems, MEMS & sensors design, artificial intelligence, and renewable energy. He received the Fulbright Science & Technology Award, in 2011, from the U.S. Department of States, USA, and several other prestigious awards worldwide.



TAHA BEYROUTHY received the Ph.D. degree in micro and nano electronics from the Grenoble Institute of Technology, in 2009, and the degree in engineering education from IMT-Atlantique (Télécom-Bretagne). He is currently an Associate Professor in electrical engineering, he has authored/coauthored more than 100 peer-reviewed publications in micro and nano electronics, robotics, artificial intelligence, and applied physics. He joined the American University of the Middle East in Kuwait (AUM), in November 2013, as an Assistant Professor and was promoted to an Associate Professor, in 2017. He has been the Dean of Engineering and Technology at AUM, since September 2017. He has been instrumental in AUM growth of higher education through his broad experience in academic leadership and commitment to both a studentcentered education and a technologically empowered teaching and learning environment.



ADEL M. ALIMI (Senior Member, IEEE) received the Graduate degree in electrical engineering, in 1990, the Ph.D. degree in electrical and computer engineering from the Ecole Polytechnique of Montreal, Canada, in 1995, and the H.D.R. degree in electrical and computer engineering, in 1995. He has been a Full Professor in electrical engineering at the National Engineering School of Sfax (ENIS), University of Sfax, since 2006. He is currently the Founder and the Director of Intelligent Machines with the Research REGIM Laboratory. He was the Director of the Tunisia Erasmus+ Office (2018–2020). He was the Director of the National Engineering School of Sfax, University of Sfax, Tunisia (2005–2011). His research interests include applications of intelligent methods (neural networks, fuzzy logic, and evolutionary algorithms) to pattern recognition, robotic systems, vision systems, and industrial processes. He focuses his research on intelligent pattern recognition, learning, analysis, and intelligent control of large scale complex systems. He is an Associate Editor and a member of the Editorial Board of many international scientific journals (e.g., IEEE TRANSACTIONS ON FUZZY SYSTEMS, *Neurocomputing*, *Neural Processing Letters*, *International Journal of Image and Graphics*, *Neural Computing and Applications*, *International Journal of Robotics and Automation*, and *International Journal of Systems Science*). He was a Guest Editor of several special issues of international journals (e.g., *Fuzzy Sets and Systems*, *Soft Computing*, *Journal of Decision Systems*, *Integrated Computer Aided Engineering*, and *Systems Analysis Modelling and Simulations*). He is also the Founder and the Chair of many IEEE Chapter in Tunisia Section, he is the IEEE Sfax Subsection Chair (2011), IEEE ENIS Student Branch Counselor (2011), IEEE Systems, Man, and Cybernetics Society Tunisia Chapter Chair (2011), and IEEE Computer Society Tunisia Chapter Chair (2011), he is also an Expert Evaluator of the European Agency for Research. He was the General Chairperson of the International Conference on Machine Intelligence ACIDCA-ICMI'2005 & 2000.

...