Arabian Horse Identification and Gender Determination System based on Feature Fusion and Gray Wolf Optimization

Ayat Taha[1]*   Ashraf Darwish[2]   Aboul Ella Hassanien[3]   Ahmed ElKholy[1]

[1] *Faculty of Science, Al-Azhar University, Cairo, Egypt*
[2] *Faculty of Science, Helwan University, Cairo, Egypt*
[3] *Faculty of Computers and Information, Cairo University, Cairo, Egypt*
* Corresponding author's Email: Ayat_taha@ymail.com

Abstract: In this paper, the identification system for the Arabian horse based on the biometric features is presented. This system used the Local Binary Pattern (LBP), Speed Up Robust Features (SURF) to extract the biometric features from the muzzle print images of the Arabian horse and to increase the robustness of the features against noise, illumination and rotation a fusion of SURF and LBP is presented to form LS features. The system also used the Support Vector Machine (SVM) classifier to perform the identification and gender determination process using the extracted features. To improve the accuracy of the proposed system Gray Wolf Optimization Algorithm (GWO) is used to optimize the SVM parameter. The experimental results represent that the proposed system achieved a promising accuracy of 96% using LS features without optimization. After Applying the GWO algorithm the accuracy of the system using LS features achieved the best accuracy 99.6% compared to SURF and LBP features.

Keywords: Animal biometric, SURF, LBP, SVM, Gray wolf optimization.

1. Introduction

All Arabian horses are a popular and eldest breed of horses around the world. They are famous for their endurance, speed, power, and beauty. Also, they involved in horse riding and many Western riding sports besides their beauty shows due to their intelligence and insomnia [1]. So Arabian horse identification is critical for many reasons first; keeping the purity of the breed, proof of ownership, and the identification of race. Second; tracking the spread of horse diseases by identifying and detecting infected horses. Third; reducing Arabian horse producers' losses by managing diseases. Fourth; reducing government costs by regulating, intervening and eradicating disease outbreaks [2].

The Arabian horse identification is performed using different methods that could be classified as traditional, electronic, and biometric methods [3]. Traditional methods may be permanent like freeze branding, hot-iron branding, hooves marker, and lip tattooing or temporary like ear tags, neckbands,

pastern bracelets, mark paints, and livestock chalk on his body. These traditional methods have may defect when used that causes painful, animal infections, duplication, fraud, difficult to read, easily remove and consume a lot of time during the identification process [4]. The electronic method which identifies the horse by inserting microchips into the neck of the horse that contains a unique alphanumeric code consisting of 15 digits also used radio frequency identification to scan the microchips and perform the identification process. The main disadvantage of this method is that the attached microchips can be misplaced, disconnected or damaged also the cost is height [5].

The biometric method depends on the features extracted from the horses, not on any external device like the human biometric identification system. The horse identification based on the biometric method can be performed using iris, retina, DNA, and muzzle that consider a unique identifier for animals. The identification based on the iris performed by the iris images obtained from the segmentation of the horse

eyes [6]. These images need to have height quality to can perform the segmentation process of the iris and extract the biometric features. Retina based identification needs a special device to get the retinal images but it achieves high accuracy in the identification system. DNA is used for identification and parentage testing in the Arabian horse but the cost and time consuming are very high. The muzzle print of the animal is widely used for cattle, dogs, and also acts identification [7].

The muzzle print is a biometric method of animal identification that is considered a unique identifier for each animal in the same species as human fingerprints. The inked muzzle print that uses ink to print the muzzle of the animal on paper is difficult to collect the data duo to the witness of the muzzle and the movement of the animals that lead to insufficient printed muzzle on the paper and consuming a lot of time. So the capture muzzle print images using a digital camera are more efficient as it produces a high-quality muzzle print image and reduces the consuming time [8].

In this paper, the Arabian horse identification system is introduced and also a gender determination. The proposed system performed on three stages that are feature extraction using SURF and LBP feature extraction techniques and also the fusion of them is performed to obtain LS features which is more robust. The second stage is the identification and gender determination performed by the SVM classifier, and finally Optimized SVM (OSVM) to enhance the accuracy of the proposed system.

The rest of this paper is organized as follow section 2 contains the related work of the animal identification using muzzle print or nose print. Section 3 represents the methods used for the Arabian horse identification system. The proposed identification system is described in detail in section 4. Experimental results and the conclusion are presented in sections 5 and 6, respectively.

2. Related work

Muzzle print is a robust biometric technique for animal identification that a chive high accuracy and less cost expensive compared to other techniques as proposed in many studies. Kumar and Singh [5] proposed a hybrid paradigm of feature extraction and classification for cattle identification using a muzzle point pattern. They applied a hybrid of eight different features extraction techniques to extract the texture features from the muzzle point images and used that texture features for recognition and classification using varying classifier methods that are K-nearest neighbor, Radial Basis Probability Network,

Decision Tree, Multilayer Perceptron, Fuzzy-K-NN, Naive Bayes, Gaussian Mixture and Probabilistic Neural Network. The proposed hybrid approach achieved the highest accuracy of 96.74% using the Fuzzy-K-NN classifier. The main limitation of the proposed paradigm is that the consuming time for the identification process due to using several classifiers in the system to enhance the accuracy rather than using the optimization techniques that give a better result.

Kumar et al. [9] proposed a framework for cattle recognition using the muzzle point image pattern. They extracted the features from the muzzle point images using a convolution neural network (CNN) and Deep Belief Neural Network (DBNN) and encode the muzzle features by stacked denoising autoencoder method. Finally, the recognition is performed by classifying the features by Deep Multilayer Neural Network (DMNN) and restricted Boltzmann machines. The framework of DBNN achieved the highest accuracy of 95.99% compared to the convolution neural network and stacked denoising autoencoder. The drawback of the proposed framework is that the feature extracted using CNN and DBNN is not discriminative features when compared to other texture features like LBP and SURF.

Tharwat et al. [10] proposed features and classifier fusion for cattle identification using the muzzle print images. The Gabor technique is used to extract the features from the muzzle print images at three different scales and then concatenate them into one feature vectors also linear discriminant analysis is used to reduce the feature vector dimension. The identification performed using support vector machine, K-nearest neighbor and minimum distance classifiers. The proposed biometric feature fusion achieved 99.5% identification accuracy. The drawback of the approach is that feature fusion depends on a different scale of the images that lead to a long feature vector dimension so the consuming time of the identification system increased especially for large datasets.

Chen et al. [11] proposed a novel biometric approach for cat recognition using their nose images. The nose images collected from 70 cats 10 images for each one. The Locality Constrained Sparse (LCS) is used to extract the cat nose features and the support vector machine is used for the recognition process which achieved 91.2% recognition accuracy. The main limitation of the proposed approach is that they do not provide the segmentation stage to remove the out layer of the capture nose images to can get the capture nose images without noise. It can affect the accuracy of the identifier proposed approach.

Tharwat et al. [12] also proposed a cattle identification approach using texture features. The texture features are extracted using LBP which is local invariant features. The identification performed using varying classifiers which are SVM, nearest neighbor, K-nearest neighbor, and Naive base. They also performed the approach at different illumination, occlusion, and rotation to prove their robustness. The results showed that SVM achieved 99.5% identification accuracy that is the best compared to other classifiers at all changes of rotations, illuminations, and occlusions. The limitation of this study is that no cross-validation method is used to verify the results of the experiments.

Samar et al. [13] proposed iris segmentation method for Arabian horse using Otsu and improved fruit fly optimization algorithm and also used Gabor filter and Discrete cosine transform (DCT) to extract the features from the iris patters and to enhance the features Dynamic Binary Particle Swarm Optimization (DBPSO) is performed to select the optimum features. They will apply this method later on iris dataset of the Arabian horse and no experimental results are reported for identification accuracy. Also, the iris pattern of animals, in general, is unstable at newborns and due to injury or infection. It is hard to obtain the dataset of iris images duo to horse movement.

3. Background

3.1 Local binary pattern

Local Binary Pattern (LBP) is an efficient image texture descriptor that encodes neighboring pixels based on the current pixel value [14]. It is robust against illumination variation and also has computational simplicity property for a texture descriptor [15]. The texture descriptor is computed using the binary pattern histogram over a region or an image. The operator compared each central pixel with the gray levels of its neighboring pixels. The value of the neighboring pixel is set to one if the gray level of it is greater than or equal to the gray level of the central pixel and the value is set to zero otherwise. These values are concatenated clockwise starting from top-left neighbor pixel to form a binary number that represents the central pixel [16]. The descriptor describes the result as a binary number (binary pattern) over the neighborhood which is LBP as indicated in Eq. (1).

$$LBP_{L,R} = \sum_{i=1}^{L-1} S(l_i - l_c) \, 2^i \qquad (1)$$

$$S(x) = \begin{cases} 1, & x \geq 0 \\ 0, & otherwise \end{cases} \qquad (2)$$

Where, $LBP_{L\,and\,R}$ determine the local binary pattern, l_c correspond to the central pixel of the neighborhood, l_i to every neighborhood pixel of l_c on a circle of radius R, L determine the number of pixels at a local region and $S(x)$ corresponding to the function of thresholding.

3.2 Speed up robust features

Speed Up Robust Features (SURF) is introduced by Bay et al. [17] which inspired by Scale Invariant Feature Transform (SIFT) descriptor algorithm but with much speed up and more robust against scale and transformation variation.
SURF detected the interesting point which is called the key point detector by using the determinant of the Hessian matrix for both scale and rotation [18]. For an image I at point x the Hessian matrix H represented as in Eq. (3).

$$H(x,\sigma) = \begin{pmatrix} L_{xx}(x,\sigma) & L_{yx}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{pmatrix} \qquad (3)$$

Where, H indicate the Hessian matrix, $L_{xx}(x,\sigma)$, $L_{xy}(x,\sigma)$, $L_{yx}(x,\sigma)$ and $L_{yy}(x,\sigma)$ are indicate the convolutions of the second-order partial derivatives of Gaussian of an image I at point x.

For more speed in computation process SURF approximated the second-order partial derivative of Gaussian by using box filters which can be easily calculated using integral images in parallel for different scales as in Eqs. (4) and (5).

$$Det(H) = D_{xx}D_{yy} - (w_f D_{xy})^2 \qquad (4)$$

Where $Det(H)$ indicate the determination of the hessian matrix, D_{xx}, D_{yy} and D_{xy} are indicate the approximation of Gaussian of an image I at point x. w_f Indicate the relative weight of the box filters that are determined by Eq. (5).

$$w_f = \frac{|L_{xy}| \ |D_{yy}|}{|L_{yy}| \ |D_{xy}|} \approx 0.9 \qquad (5)$$

For determinant of the Hessian, the relative weight w_f of the filter, responses are used to balance the expression. This is important for the conservation of energy between the Gaussian kernels and the approximate Gaussian kernels. We hold this element stable, as the outcomes of our studies did not have a noticeable effect.

SURF also described the interest keypoint by using the sum of the responses of the Haar wavelet filters which is called the SURF feature descriptor [19]. A square region is constructed around the keypoint making this point square center and divided into 4×4 square sub-regions of equal size. For each sub-region the responses of Haar wavelet filters is computed in both horizontal and vertical directions, also the absolute value of the wavelet responses is computed. The descriptor feature vector v of length 64 is formed by concatenating all 4×4 sub-regions as in Eq. (6).

$$v = \left(\sum d_x, \sum d_y, \sum |d_x|, \sum |d_x| \right) \qquad (6)$$

where v indicate the descriptor feature vector of SURF, d_x the response of Haar wavelet filters in the horizontal direction, d_y the response of Haar wavelet filters in the vertical direction and $|d_x|$, $|d_y|$ are the absolute value of the wavelet responses for horizontal and vertical direction respectively.

3.3 Gray wolf optimization algorithm

Gray Wolf Optimization Algorithm (GWO) is a meta-heuristic algorithm inspired by the social hierarchy and hunting mechanism of gray wolves in nature [20]. The leadership hierarchy of gray wolves consists of alpha, beta, delta, and omega. Alpha is considered the best solution, beta is the second, the delta is the third and the rest candidate solution described as omega. The wolves will move to encircle, hunting, and attacking to obtain the best solution in a search space [21]. The gray wolves encircle the prey using Eqs. (7) and (8).

$$\vec{D} = \left| \vec{c}.\vec{x}_p(t) - \vec{x}(t) \right| \qquad (7)$$

$$\vec{x}(t + 1) = \vec{x}_p(t) - \vec{A}.\vec{D} \qquad (8)$$

Where, $\vec{D}$ indicates the distance between the prey and the gray wolves, $\vec{x}_p$ is the position vector of the prey, $\vec{x}$ is the position vector of the gray wolf, t is the number of the current iteration, $\vec{c}$ and $\vec{A}$ are coefficient vectors calculated by Eqs. (9) and (10).

$$\vec{A} = 2\vec{a}.\vec{r}_1 - \vec{a} \qquad (9)$$

$$\vec{c} = 2.\vec{r}_2 \qquad (10)$$

Where, $\vec{a}$ indicates a linearly decreased vector from 2 to 0 during iteration, $\vec{r}_1$ and $\vec{r}_2$ are random vectors in [0, 1].

The responsibility of the hunting mechanism of the gray wolves lies on the alpha, also beta and delta are participating and omega follows them [22]. So we can simulate the hunting behavior of the gray wolves depend on the position of the three best search agents by Eqs. (11), (12), and (13)

$$\overrightarrow{D_\alpha} = \left| \vec{c}_1.\vec{x}_\alpha - \vec{x} \right|, \ \overrightarrow{D_\beta} = \left| \vec{c}_2.\vec{x}_\beta - \vec{x} \right|,$$
$$\overrightarrow{D_\delta} = \left| \vec{c}_3.\vec{x}_\delta - \vec{x} \right| \qquad (11)$$

$$\vec{x}_1 = \vec{x}_\alpha - \vec{A}_1.\left(\overrightarrow{D_\alpha} \right), \qquad \vec{x}_2 = \vec{x}_\beta - \vec{A}_2.\left(\overrightarrow{D_\beta} \right),$$
$$\vec{x}_3 = \vec{x}_\delta - \vec{A}_3.\left(\overrightarrow{D_\delta} \right) \qquad (12)$$

$$\vec{x}(t + 1) = \frac{\vec{x}_1 + \vec{x}_2 + \vec{x}_3}{3} \qquad (13)$$

Where, $\vec{x}(t + 1)$ indicates the updated position based on the position of the best search agent $\vec{x}_1$, $\vec{x}_2$, and $\vec{x}_3$.

3.4 Support vector machine

Support Vector Machine (SVM) is proposed by [23] which used to solve the classification problems. SVM based on statistical learning theory, its purpose is to construct an optimal hyperplane that separates the training data sets into two classes. Let $S = \{(x_i, y_i), i = 1,2, \dots n \}$ be the training data sets, where $x_i = x_{i1}, x_{i2}, \dots, x_{in}) \in R^n$ indicate the input vector and $y_i \in \{-1,1\}$ is the class label [24]. The optimal separating hyperplane determined by Eq. (14).

$$wx + b = 0 \qquad (14)$$

where w is the weight vector which determines hyperplane orientation, x is an input point lying on the hyperplane and b is the distance bias from the origin of the hyperplane [25]. The optimal hyperplane must have the maximum margin that separates between the two classes, so it can be found under the constrained indicated in Eqs. (15) and (16).

$$min \ \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{n} \xi_i \qquad (15)$$

$$s.t. \ \ y_i(w.x_i + b) \geq 1 - \xi_i,$$
$$\xi_i \geq 0, i = 1,2, \dots n \qquad (16)$$

Where, ξ_i indicate appositive slack variables, C is regularization constant that controls the tradeoff between the two competing criteria of maxi mizing margins and minimizing errors. Therefore
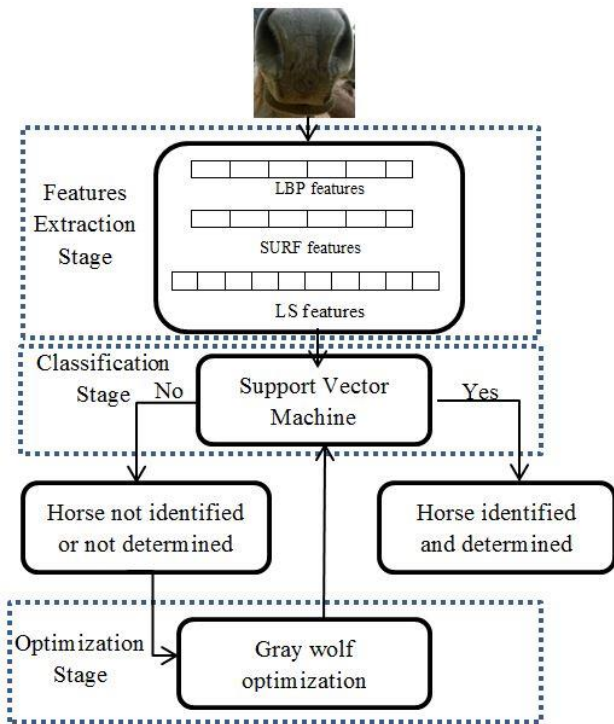
Figure. 1 Block diagram of Arabian horse identification and gender determination system

the SVM classification function can be obtained by Eq. (17).

$$f(x) = sign\left(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x_j) + b\right) \quad (17)$$

where α_i indicate the Lagrange multipliers. $K(x_i, x_j)$ is the SVM kernel function which can be the linear function as introduced in Eq. (18), or nonlinear kernel function like Eq. (19).

$$Linear\ Kernel: K(x_i, x_j) = x_i.x_j \quad (18)$$

$$polynomial: K(x_i, x_j) = (x_i.x_j + 1)^d \quad (19)$$

$$RBF: K(x_i, x_j) = exp(\frac{-\|x_i - x_j\|^2}{2\sigma^2}) \quad (20)$$

Where σ and d indicate a positive real number and the degree of the polynomial respective.

4. The proposed identification system

The proposed identification and gender determination system for Arabian horses using their muzzle print images are illustrated in detail in this section. The block diagram of the proposed identification system as shown in Fig. 1 is consists of three stages: feature extraction, classification and

finally the Gray Wolf optimization. First, the feature extraction stage includes three different feature extraction descriptors: LBP, SURF, and fusion of LBP and SURF feature descriptors to form LS features. LBP feature descriptors are obtained from the muzzle print images of the data set by dividing the image into cells and for each cell, the central pixel is compared to 8 neighboring pixels. The neighbors are selected using a circle pattern of radius equal to one to capture the details of this spatial scale. The neighboring pixel has one value if the gray level of it is greater than or equal to the gray level of the central pixel and has zero value otherwise. The value of the central pixel is obtained by concatenated all its neighbors clockwise starting from top-left. The histogram is computed for each cell and then concatenated together for all cells to form LBP features for the muzzle print image. The proposed recognition system is performed at a different number of cell sizes seeking the features that achieve the best accuracy. Second, the SURF feature descriptor is extracted from the muzzle print images as a form of the matrix for every image of length 64 dimensions and to get a feature vector that represents the muzzle print image a bag of features is used. The bag of features is performed using K-mean clustering and the histogram. K-mean divided the SURF features descriptor for all input muzzle print images into K cluster which represents the number of the images that have different sizes of the bag of features. The histogram then used to represent the feature vector for the input muzzle print images. The proposed system is performed at a different number of the bag of features seeking to improve identification accuracy. The LS feature is obtained from the fusion of LBP and SURF features. Assume $L = [x_{i1}, x_{i2}, x_{i3}, \ldots, x_{il}]$ is the feature vector obtained from the muzzle print images using LBP where $i = 1,2, \ldots, n$ and $S = [y_{i1}, y_{i2}, y_{i3}, \ldots, y_{is}]$ is the feature vector obtained from the muzzle print images using SURF where $i = 1,2, \ldots, n$ the fusion feature vector LS is constructed by concatenating L and S feature vectors to form $LS = [x_{i1}, x_{i2}, x_{i3}, \ldots, x_{il}, y_{i1}, y_{i2}, y_{i3}, \ldots, y_{is}]$ which is more robust against challenges such as rotation, illumination.

The classification using SVM is the second stage of the proposed system. The features obtained from the muzzle print images are partitioned into train features and test features. The trained features are used to get the learned models and the test features are used to evaluate the proposed system by assigning the features to each label. The Arabian horse is said to be identified and determined gender if the features extracted from the tested muzzle print image are

matched correctly to its label. Otherwise, if the tested features matched incorrectly the Arabian horse is said to be not identified or not determined gender. To increase the performance of the classification stage the SVM parameter is optimize using GWO.

Finally, the third stage in the proposed system is GWO optimization. GWO is used to optimize the parameter of SVM to improve the accuracy of the proposed system. The Arabian horse is identified if the accuracy satisfied the criteria. Otherwise, the GWO optimize the SVM parameter to reach to the highest accuracy that achieves the criteria.

5.   Experimental results and discussion

5.1 Dataset

The dataset used in the proposed system is consisting of 300 muzzle print images. These images are collected from EL Zahraa Farm- Ain Shams-Cairo-Egypt. The dataset captured from 15 female and 35 male Arabian horses to make a total 50 Arabian horse make a variety to the dataset and also determine the gender. Each Arabian horse has six muzzle print images so the dataset has 50 classes each class has six muzzle print images to make 300 samples of the Arabian horse muzzle print images. The images have different size starting from 600×700 pixels to 1400×1600 pixels for the female Arabian horse and 1000×1200 pixels to 1400×2100 pixels for female Arabian horse [26]. And also the images have different rotation, illumination and quality samples of the muzzle print images are illustrated in Fig. 2. Some preprocessing process is made to the muzzle print images such as notation, resize and cropping as shown in Fig. 3 it to 500×600 pixels.

Figure. 2 Sample of six muzzle print images of the Arabian horse and one face image
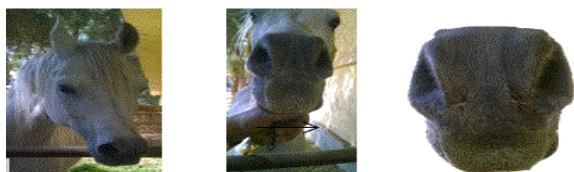

Figure. 3 the preprocessing steps to get the muzzle print images

5.2 Experimental results and discussion

The proposed system is performed using MATLAB software for the implementation process and experiments. The process and experiments were conducted using a PC with the following properties: processor has Core(TM) i5-3320M Intel(R) model with 2.6 GHz cash, and also 4.00 GB memory enabled. The operating system used is 64 bit, x64-based processor windows 8.1.

The proposed Arabian system is designed using three scenarios. The first scenario is the comparison between SURF, LBP as a feature extractor method and the effect of the fusion of SURF and LBP to form LS feature. The second scenario is conducted to choose the kernel of SVM that gives the highest accuracy and stable performance of the proposed system. Three kernels are used linear, RBF and polynomial. The third scenario of the proposed system is the optimization scenario. GWO is used to optimize the SVM parameter to enhance the identification system.

The accuracy is calculated using the true positive, true negative, false positive and false negative as illustrated in Eq. (21).

$$Accuracy = (TP + TN)/(TP + FP + TN + FN) \quad (21)$$

Where,   TP is the true positive or the correctly identified horses, TN indicates the true negative or correctly not identified horses, FP represents the false positive or incorrectly identified horses and FN reprent the false negative or incorrectly not identified horses.

In the first scenario SURF, LBP and LS are used as three different features vector extracted from the muzzle print images. The SURF feature extractor method as illustrated in Fig. 4 is performed using the bag of the feature at the different clusters. The number of clusters indicates the number of bags of SURF features. The experimental process is performed at 10, 50, 100,300 and 500 clusters or bag of SURF features.

The accuracy of the Arabian horse identification system at 500 SURF bag of features is 94% and reduced to 93% at 300 bags of SURF feature. The other bag of SURF features reduced to 92%, 87% and 60% for 100, 50 and 10 bags of SURF features respectively. It's illustrated that the proposed system achieves the best accuracy for the SURF feature extractor feature at 500-mean clusters.
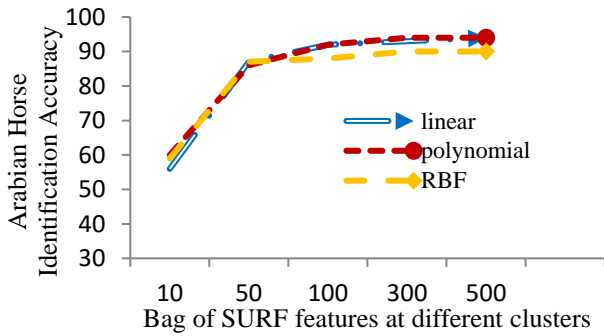
Figure. 4 Accuracy identification of the Arabian horse by using SURF features and SVM classification at linear, polynomial and RBF kernels.
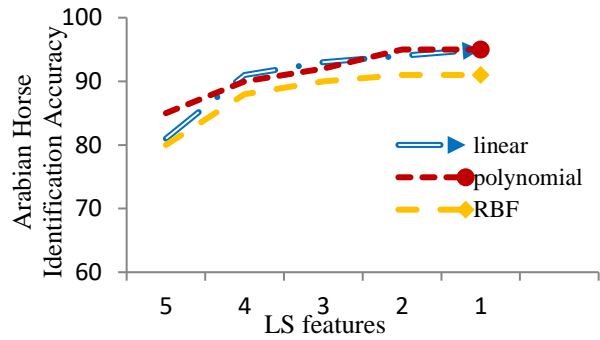

Figure. 6 Accuracy identification of the Arabian horse by using LS features and SVM classification at linear, polynomial and RBF kernels
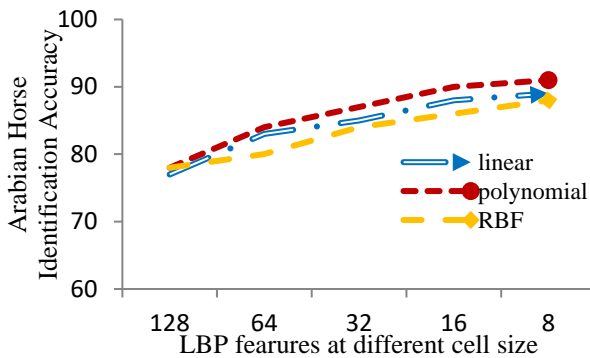

Figure. 5 Accuracy identification of the Arabian horse by using LBP features and SVM classification at linear, polynomial and RBF kernels
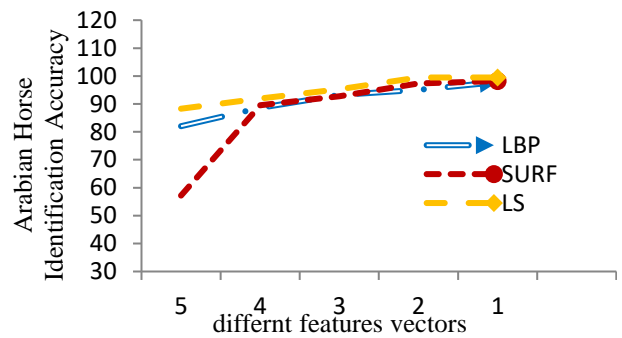

Figure. 7 Accuracy identification of the Arabian horse by using LS, LBP and SURF features and SVM classification at linear kernels

LBP extracted the features from the muzzle print image at varying cell sizes for tuning the LBP parameters and monitor the number of features extracted from the dataset at each cell size that achieve the highest accuracy. As illustrated in Fig. 5 the LBP features at cell size 8 achieve the highest Arabian horse identification accuracy which is 91% and reduced by only 1% at cell size 16, however at cells size 32 and 64 reduced by 4% and 7% respectively. At cell size 128 achieves unacceptable accuracy. The time consuming for extract LBP features is varying from 16 seconds to 10 seconds at the five cell size so there is no noticeable difference in the time consuming regardless of the identification accuracy.

LS features represent the fusion of LBP and SURF features extracted from the muzzle print images. The fusion is performed at vary number of features from LBP and SURF to determine the best fusion for Arabian horse identification system represented from 1 to 5 feature vector as illustrated in Fig. 6. The results represented that at feature vector 1 the Arabian horse identification accuracy reached 96% and decreased to 95% at feature vector 2. Also at feature vector 3 and 4 decreased to 93% and 90%.

At the last feature vector 5, the accuracy reached 87% that the least accuracy obtained. The LS feature vector reached the best accuracy of 96% for the Arabian horse identification system compared to LBP and SURF features that reached 91% and 94% respectively without any optimization as illustrated in Table 1.

In the second scenario, three kernels of SVM are used to evaluate the performance of the system. The results demonstrate that when using the SURF features the linear and polynomial kernels obtained the same accuracy 94% however reduced by 4% at RBF kernel as shown in Fig. 4. Also, polynomial kernel achieves the best accuracy when using the LBP features but linear and RBF kernels reduced by 2% and 3% respectively as shown in Fig. 5. At LS feature the linear kernel achieves the best accuracy 96% and reduced by 1% at the polynomial kernel and by 4% at RBF kernel as shown in Fig. 6. After the SVM optimization, the polynomial kernel achieves the best accuracy for the proposed system also linear kernel achieves near results for LBP, SURF and LS as illustrated in Table 1 and Figs. 7 and 8. However, the RBF achieves the least accuracy that reduced by a proximately 2% for LBP, SURF and LS as shown in table 1 and Fig. 9.

152

Table 1. A comparison between the accuracy of the Arabian horse identification system in terms of SVM and Optimized SVM by GWO at linear, polynomial and RBF kernels

Feature extracted method	SVM (%)			Optimized SVM (%)		
	Lin-ear	Polynomial	RBF	Lin-ear	Polynomial	RBF
LBP	89	91	88	97.3	97.8	95.3
SURF	94	94	90	98.2	98.6	96.2
LS	96	95	92	99.47	99.6	97.5

Table 2. A comparison between the accuracy of gender determination of the Arabian horse in terms of SVM and Optimized SVM by GWO at linear, polynomial and RBF

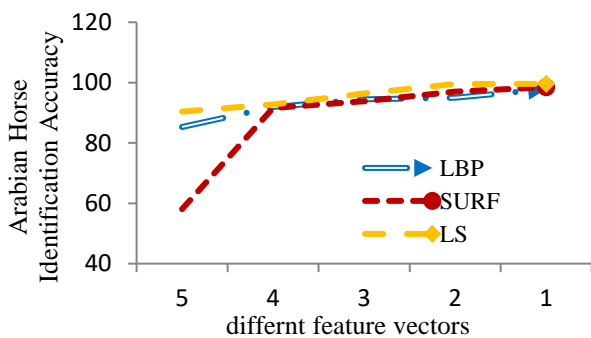| Feature extracted method	SVM (%)			Optimized SVM(%)		
	Linear	Polynomial	RBF	Linear	Polynomial	RBF
LBP	99.2	99.2	95.8	100	100	100
SURF	98.8	99.2	97.4	100	100	100
LS	100	100	98.2	_	_	100

Figure. 8 Accuracy identification of the Arabian horse by using LS, LBP and SURF features and SVM classification at polynomial kernels
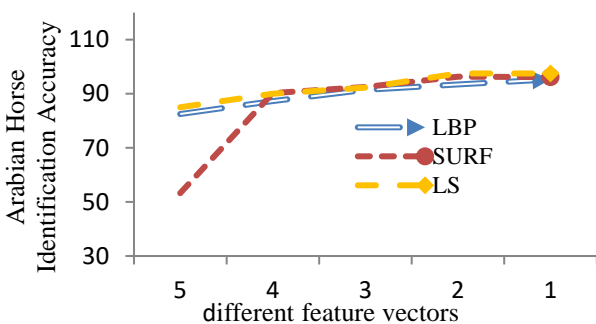

Figure. 9 Accuracy identification of the Arabian horse by using LS, LBP and SURF features and SVM classification at RBF kernels

The third scenario of the proposed system is the optimization of the SVM classifier using the GWO algorithm at linear, polynomial and RBF kernels and compare between LBP, SURF and LS feature vectors

to determine the best feature vectors for Arabian horse identification and gender determination system. At LBP features the optimized SVM increased the accuracy of the proposed system at linear, polynomial and RBF kernels which increased by approximately 6% compared to non-optimized SVM as shown in table 1. The accuracy of the system at linear and polynomial kernels is higher than the RBF kernel by 2%. Also at the SURF features the accuracy of the optimized SVM at linear and polynomial kernels increased by 4% and increased by 6% at RBF kernel compared to non-optimized SVM but still the best accuracy achieved at polynomial kernel. At the LS features the accuracy of the proposed system after the optimization increased by 3% at the linear kernel, by 4% at the polynomial kernel and by 5% at RBF kernel. The optimized SVM at polynomial and linear kernels using LS features achieve higher accuracy compared to RBF kernel.

The comparison of LBP, SURF and LS features after GWO of SVM represented in Figs. 7, 8, and 9 at linear, polynomial and RBF respectively. The comparison demonstrates that the Arabian horse identification and gender determination system achieve the best accuracy using the LS features at all SVM kernels compared to LBP and SURF.

The accuracy of the gender determination of the Arabian horse has proposed at Table 2 the result demonstrated that the accuracy with LS features with linear and polynomial kernels of SVM achieve the best accuracy and not need to optimization stage.

153

Table 3. A comparison between our proposed Arabian horse identification method and some related methods in terms of, accuracy, feature extraction, classification, and biometric methods

Authors	Feature extract methods	classification method	Biometric method	Results (%)
Kumar and Singh [5]	hybrid feature	Fuzzy-K-NN	Cattle muzzle	96.74
Kumar et al. [9]	DBN	DMNN	Cattle muzzle	95.99
Tharwat et al. [10]	Gabor	SVM	Cattle muzzle	99.5
Chen et al. [11]	LCS	SVM	Cat nose	91.2
Tharwat et al. [12]	LBP	SVM	Cattle muzzle	99.5
Samar et al. [13]	DCT DBPSO	Not reported	Horse iris	Not reported
Our Proposed system	LBP SURF LS	SVM OSVM	Horse muzzle	99.6

However, LBP and SURF need optimization to enhance the accuracy of the gender determination of the Arabian horse.

To further show the effectiveness of the proposed system comparison is performed with the most related work as represented in Table.3. Since there is no public muzzle print images dataset for the Arabian horse and there is one study of the Arabian horse identification using its iris without any reported results [13], the comparison performed with the identification of cattle and cat based on their muzzle print or nose print images. As represented in Table 3 the proposed Arabian horse identification system achieved the best results inaccuracy. That is due to two reasons: the use of the features fusion that is robust against rotation, illumination, and variation and the optimization of GW to SVM classifier (feature fusion and GWO discussed in detail in Section 4).

6.   Conclusions and future work

In this study, anew system for Arabian horse identification and gender determination using muzzle print images was presented. This system used LBP, SURF, and LS to extract the muzzle print features that are robust against noise, rotation, and illumination. SVM classifier is used to perform the identification process at linear, polynomial and RBF kernels. It also optimized using the GWO algorithm to increase the accuracy of the system. The cell size parameter of LBP and the number of the bag of features of SURF are tuned to determine the feature vector that achieving the best accuracy. Also, the LS feature fusion is performed at vary number of LBP and SURF features vector. The experimental results demonstrate that at cell size 8 of LBP, at 500 bags of features for SURF and the feature vector 1 of LS the system achieves the best performance over the features. The result demonstrates that the best feature extraction method was LS achieved 96%, 95%, and 92% accuracy at linear, polynomial and RBF kernel

respectively which indicate the linear kernel is the suitable kernel for the proposed system. Moreover, the accuracy after SVM optimization using the GWO algorithm was increased to 99.47%, 99.6% and 97.5% at linear, polynomial and RBF kernels respectively which reflect the superiority of the polynomial kernel. Also, the proposed Arabian horse identification system achieved the best accuracy compared to other related work as illustrated in Table 3. In future work, the dataset will extend to evaluate the proposed system against a larger dataset and will perform different algorithms such as deep learning algorithms.

Conflicts of Interest

Data collection for this study was undertaken by Ayat Taha from Al-zahraa farm after their permission. All research results presented in this manuscript belong to the authors alone, and not any institution to which they are or were affiliated. The Authors declares that there is no conflict of interests.

Author Contributions

The research is performed by the authors and the individual contributions is provided as follows: conceptualization, Ashraf Darwish and Aboul Ella; methodology, Ayat Taha; software, Ayat Taha; validation, Ayat Taha, Ashraf Darwish, and Aboul Ella formal analysis, Ahmed ElKholy ; investigation, Ahmed ElKholy; resources, Ayat Taha; data curation, Ayat Taha; writing—original draft preparation, Ayat Taha; writing—review and editing, Ayat Taha; visualization, Ayat Taha; supervision, Ashraf Darwish , Aboul Ella and Ahmed ElKholy.

Notation

$LBP_{L,R}$ = LBP feature vector
l_c = central pixel of a region
l_i = neighborhood pixel of l_c
R = circle radius

L	=	number of pixels at a local region
$S(x)$	=	1 or 0 according to Eq.2
H	=	Hessian matrix
L_{xx}, L_{yx}, L_{yy}	=	second-order partial derivatives of Gaussian of an image I at point x
$Det(H)$	=	Hessian matrix approximation
D_{xx}, D_{yy} D_{xy}	=	approximation of Gaussian of an image I at point x
w_f	=	weight ≈ 0.9 as given by Eq.5
v	=	SURF feature vector
d_x	=	the response of Haar wavelet filters in the horizontal direction
d_y	=	the response of Haar wavelet filters in the vertical direction
$|d_x|$	=	absolute value of d_x
$|d_y|$	=	absolute value of d_y
$\vec{D}$	=	distance between prey and GW
$\vec{x}_p$	=	the position vector of the prey
$\vec{x}$	=	the position vector of the gray wolf
t	=	the number of the current iteration
$\vec{c}$ and $\vec{A}$	=	coefficient vectors are given by Eq.9
$\vec{a}$	=	2,0
$\vec{r}_1, \vec{r}_2$	=	[0, 1]
$\vec{x}(t+1)$	=	The new position in the search space
w	=	the weight vector is given by Eq. 15 and Eq. 16
b	=	distance bias from the origin of the plane
ξ_i	=	appositive slack variables
C	=	regularization constant
α_i	=	Lagrange multipliers
$K(x_i, x_j)$	=	SVM kernel function

References

[1] M. Bowling, D. Pendell, D. Morris, Y. Yoon, K. Katoh, K. Belk, and G. Smith, "Review: identification and traceability of cattle in selected countries outside of North America", *The Professional Animal Scientist*, Vol. 24, No. 4, pp. 287–294, 2008.

[2] C. Shanahan, B. Kernan, G. Ayalew, K. McDonnell, F. Butler, and S. Ward, "A framework for beef traceability from farm to slaughter using global standards: an Irish perspective", *Computers and Electronics in Agriculture,* Vol. 66, No. 1, pp. 62–69, 2009.

[3] C. Roberts, "Radio frequency identification (RFID)", *Computers and Security,* Vol. 25, No. 1, pp.18–26, 2006.

[4] A. Awad, "From classical methods to animal biometrics: A review on cattle identification and tracking", *Computers and Electronics in Agriculture,* Vol. 123, pp. 423–435, 2016

[5] S. Kumar and S. Singh, "Automatic identification of cattle using muzzle point pattern: hybrid feature extraction and classification paradigm", *Multimedia Tools and Application*, Vol. 76, pp. 26551–26580, 2017.

[6] P. Ahrendt, T. Gregersen, and H. Karstoft, "Development of a real-time computer vision system for tracking loose-housed pigs", *Computers and Electronics in Agriculture*, Vol. 76, No. 2, pp. 169-174, 2011.

[7] W. Petersen, "The identification of the bovine using nose-prints", *Journal of Dairy Science,* Vol. 5, No. 3, pp. 249-258, 1922.

[8] H. Minagawa, T. Fujimura, M. Ichiyanagi, and K. Tanaka, "Identification of beef cattle by analyzing images of their muzzle patterns lifted on paper", In: *Proc. of International Conf. for Information Technology in Agriculture*, Beijing, China, pp. 596–600, 2002.

[9] S. Kumar, A. Pandey, K.Satwik, S. Kumar, S.Singh, A. Singh, and A. Mohan, "Deep Learning Framework for Recognition of Cattle using Muzzle Point Image Pattern", *Measurement*, Vol. 116, pp. 1-17, 2017.

[10] A. Tharwat, T. Gaber, and A. Hassanien, "Two biometric approaches for cattle identification based on features and classifiers fusion", *International Journal of Image Mining*, Vol. 1, No. 4, pp. 342–365, 2015.

[11] Y. Chen, S. Hidayati, W. Cheng, M. Hu, and K. Hua, "Locality Constrained Sparse Representation for Cat Recognition", *In: Tian Q., Sebe N., Qi GJ., Huet B., Hong R., Liu X. (eds) MultiMedia Modeling. MMM 2016. Lecture Notes in Computer Science, Vol. 9517. Springer, Cham*, 2016.

[12] A. Tharwat, T. Gaber, A. Hassanien, H. Hassanien, and M. Tolba, "Cattle Identification Using Muzzle Print Images Based on Texture Features Approach", In: *Proc. of International Conf. on Bio-Inspired Computing and Applications*, *Advances in Intelligent Systems and Computing*, Vol. 303, 2014.

[13] S. Zekrallah, M. Soliman and A. Hassanien, "Feature selection using dynamic binary particle swarm optimization for Arabian horse identification system", In: *Proc. of International Conf. on Computer Engineering and Systems*, Cairo, pp. 669-675, 2017.

[14] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face Recognition with Local Binary Patterns", *In: Pajdla T., Matas J. (eds) Computer Vision - ECCV 2004. ECCV 2004. Lecture Notes in*

Computer Science, vol 3021. Springer, Berlin, Heidelberg, 2004.

[15] M. Hassaballah, H. Alshazly, and A. Ali, "Ear recognition using local binary patterns: A comparative experimental study", *Expert Systems with Applications*, Vol. 118, pp. 182-200, 2018.

[16] L. Liu, P. Fieguth, G. Zhao, M. Pietikainen, and D. Hu, "Extended Local Binary Patterns for Face Recognition", *Information Sciences*, Vol. 358-359, pp. 56-72, 2016.

[17] B. Herbert, E. Andreas, T. Tuytelaars, and L. Gool, "Speeded-Up Robust Features (SURF)", *Computer Vision and Image Understanding*, Vol. 110, No. 3, pp. 346-359, 2008.

[18] Y. Bakhshi, S.Kaur, and P.Verma, "Face Recognition using SIFT, SURF and PCA for Invariant Faces", *International Journal of Engineering Trends and Technology*, Vol. 34, No. 1, pp. 39-42, 2016.

[19] B. Anand and P. Shah, "Face Recognition using SURF Features and SVM Classifier", *International Journal of Electronics Engineering Research,* Vol. 8, No. 1, pp. 1-8, 2016.

[20] S. Mirjali, S. Mirjalili, and A. Lewis, "Grey Wolf Optimizer", *Advances in Engineering Software*, Vol. 69, pp. 46–61, 2014.

[21] S. Kamel, R. YaghoubZadeh, and M. Kheirabad, "Improving the performance of support-vector machine by selecting the best features by Gray Wolf algorithm to increase the accuracy of diagnosis of breast cancer", *Journal of Big Data*, Vol. 6, No. 90, 2019.

[22] Y. Wei, N. Liu, H. Chen, M. Wang, Q. Li, X. Cui, and H. Ye, "An Improved Grey Wolf Optimization Strategy Enhanced SVM and Its Application in Predicting the Second Major", *Mathematical Problems in Engineering,* Vol. 2017, 2017.

[23] C. Cortes and V. Vapnik, "Support-vector networks", *Machine Learning*, Vol. 20, pp. 273-297, 1995

[24] J. Wei, Z. Jian-qi, and Z. Xiang, "Face recognition method based on support vector machine and particle swarm optimization", *Expert Systems with Applications*, Vol. 38, pp. 4390–4393, 2011.

[25] H. Tsai1 and Y. Chang, "Facial expression recognition using a combination of multiple facial features and support vector machine", *Soft Computing*, Vol. 22, pp. 4389–4405, 2018.

[26] A. Taha, A. Darwish, and A. Hassanien, "Arabian Horse Identification Benchmark Dataset", *arXiv: 1706.04870v1*, 2017.