PCA-Based Animal Classification System

Emre Dandıl
Department of Computer Engineering
Bilecik Seyh Edebali University
Bilecik, Turkey
emre.dandil@bilecik.edu.tr

Rukiye Polattimur
Department of Computer Engineering
Bilecik Seyh Edebali University
Bilecik, Turkey
rukiyeozubek@gmail.com

***Abstract*— Missing, disappearing, swapping are fundamental problems encountered especially in pet animals that are very similar. Unfortunately, there are few methods that can overcome such problems. Traditionally, animals are known through their external appearances and patterns. Biometric based systems developed for the identification of animals are scarce and many of these systems are inadequate. In this study, a Principal Component Analysis (PCA) based system was developed for the recognition and classification of different species of animals. Thanks to the application software in the structure of the developed system, it is possible to identify the animals most resembling an animal in the image dataset. Experimental studies on cow, cat, dog, goat and rabbit animal species shows a success rate of 92% in the first nearest recognition and 83% in the second nearest recognition. It has been seen that the improving of this developed system can be used in the classification process of different kinds of animals.**

Keywords—animal classification, principal component analysis, PCA, software

I. INTRODUCTION

Despite of advances in software and technology, nowadays, detection and recognition of animals is still a difficult process and there is no single method that provides an effective solution. Traditionally, the problem in the detection and identification of animals is approached in the form of double pattern recognition [1]. It is expected that the image where the animal is located is divided into sections and each section separated will create a feature in itself. The training process is firstly carried out by a classification method with the features to be obtained from the image sections and then, when a new image is obtained, the recognition process is completed.

Animal recognition systems are generally examined in two sub-categories. In the first category, an animal image is compared with the other animal images in the data and after detection of similar ones, recognition process is provided according to a certain threshold value. In the second stage, an animal image is compared with the whole animal images in the data set, and then it is carried out confirmation of whether the animal image is in this dataset[2].

In the literature, it is seen that studies are carried out by using various methods on different animal groups in the context of animal classification and recognition. However, it is determined that the number of significant studies in this area is not much. Some examples of these studies are the study on different species of dogs [3], the classification of cats and dogs [4], and monitoring with face recognition of lions in the wild can be shown as examples [5]. In many of these studies, different techniques are used, and the differences in the data are prominent. Most of the studies using PCA [6]-[7] have few studies on the classification of animal images with PCA, but some studies [8] have been still found.

In this study, a biometric recognition system was developed to provide classification of PCA-based animals in a dataset generated from images collected from animals. The studies on created dataset are supported application software having designed a visual interface. In the proposed system, after any selected animal image has gone through an improvement phase with image pre-processing techniques and the class of the animal is identified by matching the relevant images in the dataset and a closest second class to that class is determined using the nearest calculated components of the PCA.

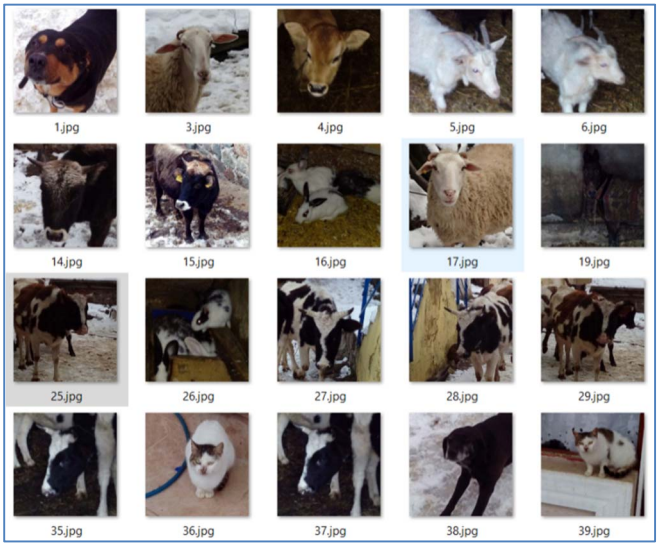
II. MATERIAL AND METHOD

A. Image Dataset

In this study, a dataset was created with images from certain animals. As a feature of the PCA algorithm used in the study, the size of all images is determined to be 702x702 because the whole images should have the same height and width. There are a total of 408 images, whose dimensions are corrected. The images of some animals in the dataset are shown in Figure 1. Images in the dataset to facilitate the processing of images are used in grayscale format that is lowest size and photographs were taken as '.pgm' format. There are 20 different classes (animal species) among the images in the dataset. At least two different images were recorded for each animal type. The fact that there are more images than the same animal is an advantage in the system of matching the images of the same animal.

(a)

978-1-5386-4184-2/18/$31.00 ©2018 IEEE



(b)

Fig. 1. (a),(b) Images of some animals in the image dataset

B. *Principal Component Analysis (PCA)*

Principal Component Analysis, a wide range of uses in the field, such as, face recognition, image compression, pattern recognition, based on statistical methods is a method [9]. While PCA simultaneously turn into small-sized and unrelated vectors from vector-based large-size data and interrelated vectors, it also provides the transformation of the original data. After this conversion, the original data is represented in a different dimension as the main component. The features obtained by representation in different size are called the principal components of their initial features. Variance value of the first principal component is the highest and variance values of other essential components are descendingly listed [19]. Figure 2 show the flow diagram of the PCA.

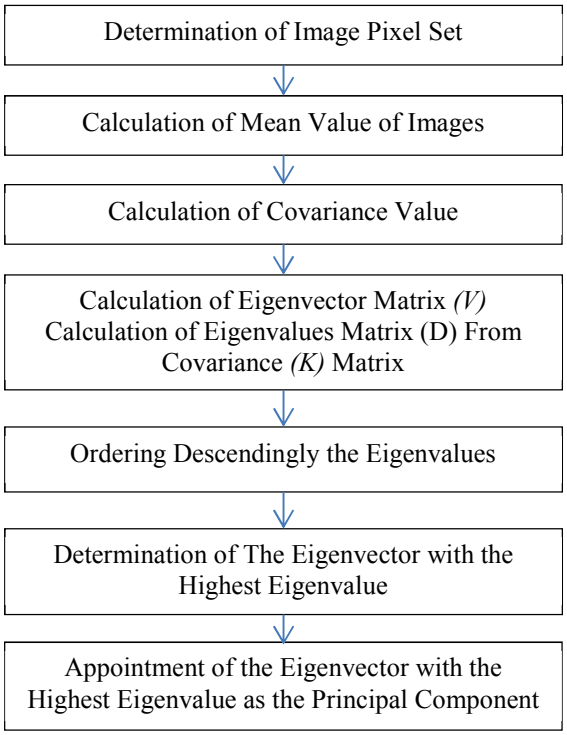


Fig. 2. The flowchart diagram of PCA

Principal component analysis (PCA) is a linear reduction technique that requires the data to be tracked toward the highest variability direction. The most important goal in the principal component approach is to find the strongest pattern. The first major component in TBA shows the basic vector for the highest variability direction. The second main component forms the base vector in a direction perpendicular to the primary principal component and other components continue in the same way. The percentage of total variability of the data during the calculation is set equally for selecting the number of main components. Then, the calculation of the basic components, the calculation of the covariance matrix of the data, the eigenvalue separation, the order of the eigenvectors according to the descending order of the eigenvalues, and finally the basic components are determined, respectively. The detailed flow procedure of the PCA can be given in the following order.

Step 1: Calculation of the Covariance matrix from the data (Equation 1)

$$C = (X - \bar{x})(X - \bar{x})^T \quad (1)$$

Equation 1 X represents the data matrix, and $\bar{x}$ represents the mean vector of X.

Step 2: Calculation of the Eigenvectors (V) matrix and the diagonal matrix of Eigenvalues (D) (Equation 2)

$$V^{-1}\, C\, V = D \quad (2)$$

Step 3: Eigenvectors in V are sorted in descending order of eigenvalues in D and the data is reflected in these eigenvector directions from the inner product between the data matrix and the ordered eigenvector matrix.

$$\text{Predicted data}=[V^T (X - \bar{x})^T]^T \quad (3)$$

III. ANIMAL CLASSIFICATION APPLICATION

In this study, PCA based an application software was developed in MATLAB in order to classify animal images. All applications in the study were performed in the application created with MATLAB software. All experimental studies were performed with a computer equipped with 2.8 Ghz i7 processor and 16 GB memory.

The main form screen of the developed application software is shown in Figure 3. As can be seen from this software form, an interface has been created in which selected image of an animal can be loaded and then the images closest to this image can be displayed.

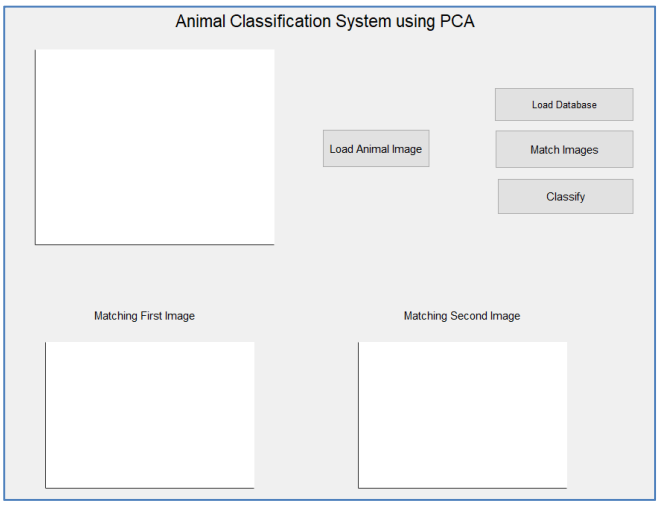
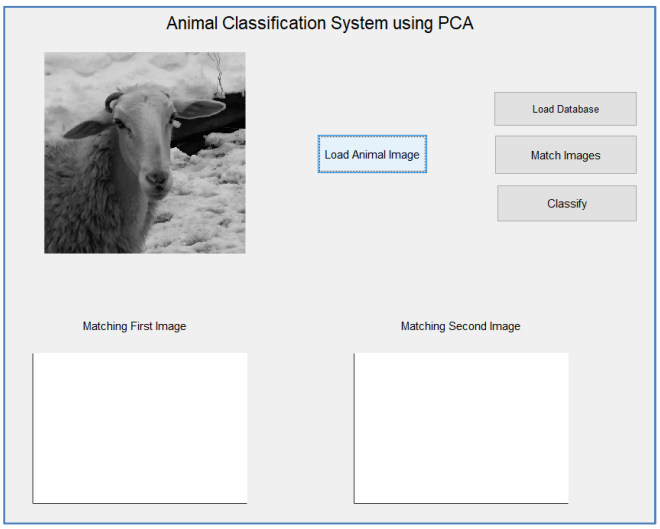
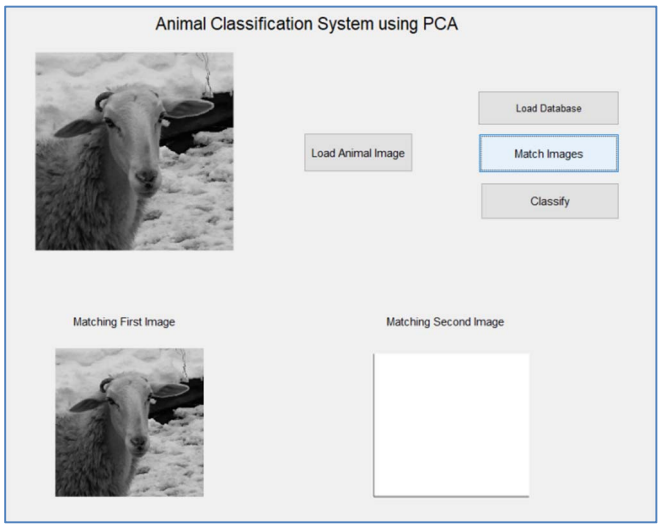


Fig. 3. Main form screen of the developed application software

The first operation after executing the application is the selection of any animal image from the database. After the animal image is uploaded to the software, a number of operations are carried out; image improvement / the elimination of roughness on the image by using pre-processing techniques, removal of small debris, elimination of errors during shooting, respectively. After the image pre-processing phase is completed, the Scan of Matching Images operation is started and the image closest to the loaded image is shown on the software as shown in Figure 4.



(a)



(b)

Fig. 4. (a) Loading of the animal image, (b) the results of the TBA based classification process and showing of the matching image

Figure 5 shows the matching results for both components in a different animal species. Both results show successful results.

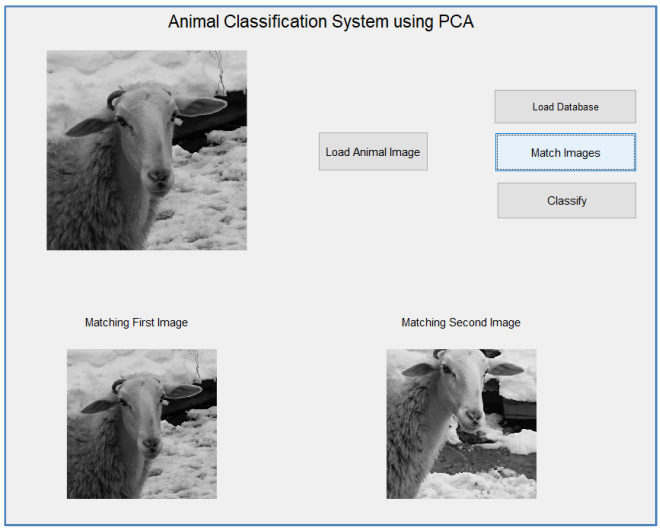


Fig. 5. Results of PCA based classification

Figure 6 shows the final stage of the animal classification process. After determining the matching first and second images on the PCA based software, it is decided what the type of animal is. As seen on Figure 6, even though the image of animal was taken from different angles, the software has identified both matching images as successful in the same class as the loaded image.

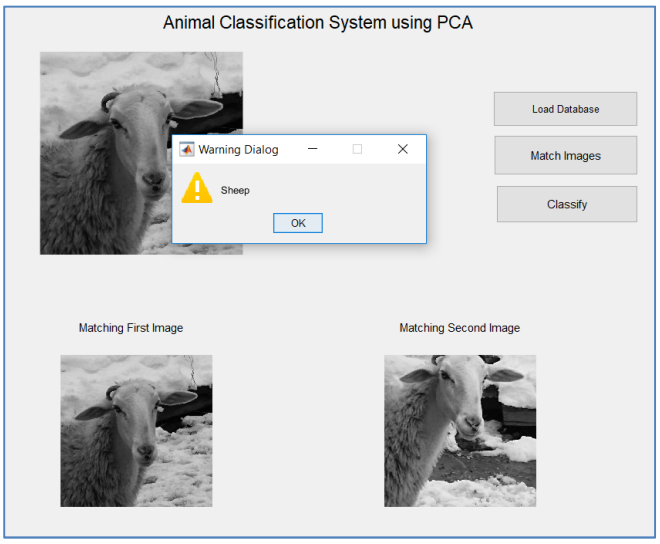


Fig. 6. Completion of animal classification process

Figure 7 shows the classification and recognition process of a different species of animal. It appears that the images matching the uploaded image belong to the same class and are classified correctly.

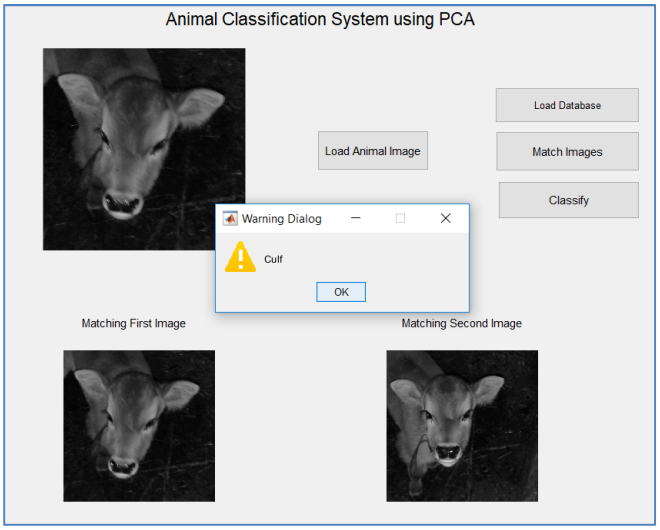


Fig. 7. Completion of different types of animal classification process

Figure 8 also shows the classification and recognition process of a different species of animal species. However, in this test, the first image matching the loaded image was determined to belong to the same class, while the second matching image was determined by the software that was not of the same class (animal).

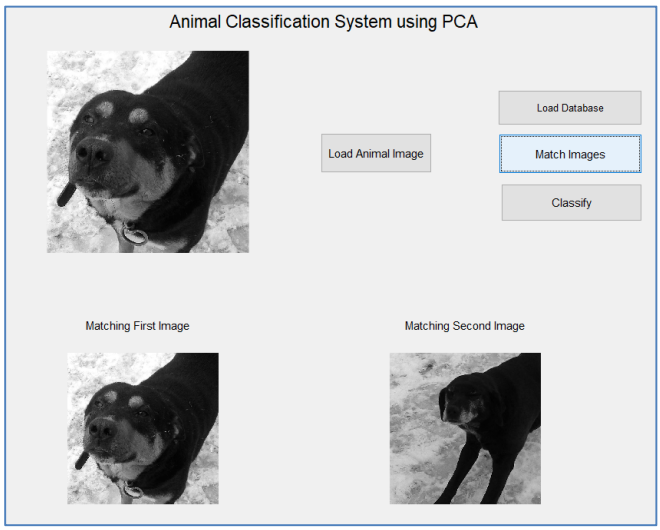


Fig. 8. Different types of animal classification process and belonging status to different classes of matching images

In Table 1, percent of the correct classification of the first and second images which matched the results of the test procedures for animal species such as cows, cats, dogs, goats and rabbits are given. As can be seen this table, the success ratio in the first matching image was 92%, while the performance ratio in the second matching image was calculated as 83%. The reason for the low ratio in the second matching image can be explained by the fact that the first one of the components calculated in the PCA algorithm consists of the data that is closest to the image.

TABLE I. CLASSIFICATION RESULTS

	Accuracy	
Animal types	1.Matching Image	2.Matching Image
Cow	96%	86%
Dog	95%	84%
Cat	89%	78%
Goat	92%	83%
Rabbit	90%	82%
Average	92%	**83**%

IV. DISCUSSION AND CONCLUSION

In this study, PCA based an application software has been developed to classify different species of animal images on a specially generated dataset. During the tests, the success ratio in the first matching image was 92%, while the performance ratio in the second matching image was calculated as 83%. The reason for the low ratio in the second matching image can be explained by the fact that the first one of the components calculated in the TBA algorithm consists of the data that is closest to the image. The high rate of success achieved in this study will provide a basis for the identification of animal species in future. In later studies, the PCA's performance rate can be compared with distinct classification algorithms. In addition, the increase in the number of images used in the dataset is expected to increase performance rate.

ACKNOWLEDGEMENTS

This Project was supported by Bilecik Seyh Edebali University BAPK with Project Number: 2017-01.BŞEU.03-09.

REFERENCES

[1] Z. Xie, , A. Singh, J. Uang, K. S. Narayan and P. Abbeel, "Multimodal blending for high-accuracy instance recognition", In: 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. Tokyo: IEEE, 2013, pp. 2214–2221.

[2] T. Trnovszky, P. Kamencay, R.Orjesek, M. Benco, and P. Sykora, "Animal recognition system based on convolutional neural network", Advances in Electrical and Electronic Engineering, 15(3), pp. 517., 2017.

[3] S. Kumar, and S.Kumar Singh, "Biometric Recognition For Pet Animal", Journal of Software Engineeringand Applications, 7, 470-482, 2014.

[4] B. Tuyen Ly, J. Hammer, "Classification of Cats and Dogs", Math 521 Spring, Csulb Dept. of Mathematics, 2012.

[5] T. Burghardt, J. Calic, and B. T. Thomas, "Tracking Animals in Wildlife Videos Using Face Detection." In EWIMT, October, 2004.

[6] A. A.Oommen, C. S. Singh, and M. Manikandan, "Design of face recognition system using principal component analysis." Int. J. Res. Eng. Technol, 3(1), pp. 6-10, 2014.

[7] S. Satonkar Suhas, B. Kurhe Ajay, and B. Prakash Khanale, "Face recognition using principal component analysis and linear discriminant analysis on holistic approach in facial images database.", Int Organ Sci Res, 2(12), pp. 15-23., 2012.

[8] S. P. Bahurupi, and D. S. Chaudhari, "Principal component analysis for face recognition.", International Journal of Engineering and Advanced Technology (IJEAT), 2012.

[9] I. Yazar, H. S. Yavuz, M. A. Çay, "Temel Bileşen Analizi Yönteminin ve Bazı Klasik ve Robust Uyarlamalarının Yüz Tanıma Uygulamaları", Eskişehir Osmangazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, XXII(1), 2009.

[10] H. Durucasu, "Asal Bileşen Analizi ve Bir Uygulama Denemesi", Yüksek Lisans Tezi, Anadolu Üniversitesi Fen Bilimleri Enstitüsü, 89, 1991.