IDENTIFICATION OF GREAT APES USING FACE RECOGNITION

Alexander Loos, Martin Pfitzer* and Laura Aporius***

* Audio-Visual Systems Group, Fraunhofer IDMT, 98693 Ilmenau, Germany
** Department of Primatology, Max Planck Institute EVA, 04103 Leipzig, Germany
Email: {*loos, pfitmn*}*@idmt.fraunhofer.de, laura.aporius@eva.mpg.de*

ABSTRACT

In recent years, thousands of species populations declined catastrophically leaving many species on the brink of extinction. Several biological studies have shown that especially primates like chimpanzees and gorillas are threatened. An essential part of effective biodiversity conservation management is population monitoring using remote camera devices. However, due to the large amount of data, the manual analysis of video recordings is extremely time consuming and highly cost intensive. Consequently, there is a high demand for automatic analytical routine procedures using computer vision techniques to overcome this issue. In this paper we present a technique for the identification of great apes, in particular chimpanzees, using state-of-the-art algorithms for human face recognition in combination with several classification schemes. For benchmark purposes we provide a publicly available dataset of captive chimpanzees. In our experiments we applied several common techniques like the well known Eigenfaces, Fisherfaces, Laplacianfaces and Randomfaces approaches to identify individuals. We compare all of these methods in combination with the classification approaches Nearest Neighbor (NN), Support Vector Machine (SVM) and a new concept for face recognition, Sparse Representation Classification (SRC) based on Compressive Sensing (CS).

1. INTRODUCTION

The current biodiversity crisis and the accompanied catastrophical declining of species populations is startling and many thousands of species populations, especially great apes like chimpanzees or gorillas, are threatened [5, 3]. Therefore, autonomous monitoring techniques become more and more important. Especially individual identification is required for many questions in behavioral ecological research, ranging from wildlife epidemiology to interpopulation comparisons of social dynamics or the evolution of social behavior and cognition. In recent years the availability of digital recording devices has facilitated the collection of large amounts of data on species and individuals, for instance with remote camera traps or autonomous recording devices. Since the manual analysis of images and video recordings is not feasible for such a huge amount of data, there is a high demand for automated analytical routine procedures.

Recently, computer vision algorithms have been successfully applied to recognize animals of different species. Ardovini *et al.* [1] for instance proposed a system for semi-automatic recognition of elephants from photos based on shape comparison of the nicks characterizing the elephant's ears. Another automatic recognition approach to identify African Penguins on Robben Island was presented by [4]. The authors suggested to use a number of reference points on individually-specific coat patterns for identification. The proposed model employs boosted point-surround classifiers as local appearance descriptors. Most recently Lahiri *et al.* [8] described an algorithmic and experimental approach for the identification of zebras in the wild. Again, the authors use soft biometrics to recognize individuals.

All of these methods use characteristic patterns of fur and skin or other unique markings to distinguish between individuals. However, such a technique is hard to implement for great apes, especially if the resolution of the used images or videos is not sufficient to detect individual markings like wrinkles under the eyes. Starting from the assumption that humans and their closest relatives share similar properties of the face, we propose to use face recognition techniques for the recognition of primates. Over the past two decades there has been a rapidly increasing demand on technology for face recognition. One of the most successful and very well studied techniques for human face recognition are *appearance-based* methods. Here the face images of $h \times w$ pixels are usually represented as vectors of size $1 \times n$, where $n = h \cdot w$. In practice this n-dimensional space is too large for robust and fast face recognition. A well-known and common used attempt to overcome this issue is to use dimensionality reduction techniques. The most famous methods for this purpose are *Principle Component Analysis* (*Eigenfaces*) [9], *Linear Discriminant Analysis* (*Fisherfaces*) [2] and *Locality Preserving Projections* (*Laplacianfaces*) [7]. Recently, also randomly generated projection matrices, so called *Randomfaces*, in combination with a *Compressive Sensing* (*CS*) framework for classification, also known as *Sparse Representation Classification* (SRC), achieved excellent results even under difficult conditions like varying lighting or occlusion [10].

In this paper, we demonstrate that both state-of-the-art algorithms and traditional methods for appearance based face recognition are not only capable of identifying humans but also great apes, especially chimpanzees (*Pan troglodytes*). Experimentation is conducted over a self prepared dataset of 24 chimpanzees, gathered in the zoo of Leipzig, Germany, to thoroughly compare each of the above mentioned approaches using three different classifiers: *Nearest Neighbor (NN)*, *Support Vector Machine (SVM)* and *Sparse Representation Classification (SRC)*.

The rest of the paper is organized as follows: In Section 2, we briefly explain the face recognition approaches and classifiers we used in our experiments. The methods for data collection and annotation are explained in Section 3. In Section 4, all performed evaluation experiments comparing the above-mentioned algorithms are described in detail to demonstrate the feasibility for the identification of primates using face recognition. Finally, we conclude our paper giving a summary of our work and future ideas of improvement.

2. BACKGROUND

2.1 Subspace Analysis

In appearance based face recognition techniques, the N high dimensional vectorized face images $\{x_1, \cdots, x_N\}$ of size n are usually projected into a lower dimensional subspace of size m using a unitary projection matrix $W \in \mathbb{R}^{n \times m}$.

$$y_k = W^T x_k \tag{1}$$

The resulting feature vectors $y_k \in \mathbb{R}^m$, with $k = 1, \cdots, N$, can then be used for classification.

2.1.1 Principal Component Analysis (Eigenfaces)

The famous Eigenfaces approach, introduced by Turk and Pentland in the early nineties [9], is one of the approaches for dimensionality reduction. This method uses Principal Component Analysis (PCA) to map the facial image vectors into a lower dimensional space. PCA aims to extract a subspace where the variance is maximized while the global structure of the image space is preserved. It's objective function is

$$w_{opt} = \arg\max_w (w^T S_{PCA} w). \tag{2}$$

The scatter matrix S_{PCA} is defined as

$$S_{PCA} = \frac{1}{N} \sum_{k=1}^{N} (x_k - \mu)(x_k - \mu)^T, \tag{3}$$

where $\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$ denotes the mean of all images. The output set of principle vectors $\{w_1, \cdots, w_m\}$ is an orthonormal set of vectors representing the eigenvectors of the sample covariance matrix associated with the $m \ll n$ largest eigenvalues.

2.1.2 Linear Discriminant Analysis (Fisherfaces)

While the Eigenfaces method tries to preserve the global structure of the image space, Linear Discriminant Analysis (LDA) aims to preserve the discriminating information searching for the directions that are efficient for distinction. The Fisherfaces method was first introduced in [2]. Again we consider a set of n-dimensional samples $\{x_1, \cdots, x_N\}$ associated to one of C classes $\{K_1, \cdots, K_C\}$, where N_i denotes the number of images in class K_i. The between-class scatter matrix S_b and the within-class scatter matrix S_w are defined as

$$S_b = \frac{1}{N} \sum_{i=1}^{C} N_i (\mu_i - \mu)(\mu_i - \mu)^T \tag{4}$$

and

$$S_w = \frac{1}{N} \sum_{i=1}^{C} \left[\sum_{j=1}^{N_i} (x_j^{(i)} - \mu_i)(x_j^{(i)} - \mu_i)^T \right], \tag{5}$$

respectively, where μ_i is the mean of all images in class K_i, μ is the total sample mean vector and $x_j^{(i)}$ is the j-th image of class i. LDA solves the Fisher criterion, *i.e.* the projection is chosen to maximize the ratio of the determinant of S_b of the projected samples to the determinant of S_w of the projected samples. Consequently, the objective function is as follows:

$$w_{opt} = \arg\max_w \left(\frac{w^T S_b w}{w^T S_w w} \right). \tag{6}$$

The optimal projection basis for LDA is the set of generalized eigenvectors of S_b and S_w associated with the m largest eigenvalues λ_i, $i = 1, \cdots, m$

$$S_w^{-1} S_b w_i = \lambda_i w_i. \tag{7}$$

The Fisherfaces method avoids the problem of the singularity of S_w by projecting the image set to a lower dimensional space using PCA before applying LDA. The set of eigenvectors of S_{PCA} corresponding to the $N-C$ largest eigenvalues is denoted as $W_{PCA} = [w_1, \cdots, w_{N-C}]$. Then the new between-class scatter matrix and within-class scatter matrix can be rewritten as

$$\tilde{S}_b = W_{PCA}^T S_b W_{PCA} \tag{8}$$

and

$$\tilde{S}_w = W_{PCA}^T S_w W_{PCA}. \tag{9}$$

Now $W_{LDA} = [w_1, \cdots, w_{C-1}]$ is the set of eigenvectors of $\tilde{S}_b$ and $\tilde{S}_w$ associated with the $C-1$ largest eigenvalues λ_i with $i = 1, \cdots, C-1$. The final projection is then simply given by the multiplication of W_{PCA} and W_{LDA}

$$W_{Final} = W_{PCA} W_{LDA}. \tag{10}$$

2.1.3 Locality Preserving Projections (Laplacianfaces)

The Localitiy Preserving Projections (LPP) approach assumes that the face images reside on a nonlinear submanifold hidden in the image space. Unlike the Eigenfaces or Fisherfaces method, which effectively only see the global euclidean structure, LPP finds an embedding that preserves local information and obtains a subspace that best detects the essential face manifold structure. To preserve the local structure of the face space, this manifold structure is modeled by a nearest-neighbor graph. We start by defining an adjacency graph G with m nodes. An edge is put between two nodes k and j if they are within an ε neighborhood, i.e. if $\|x_k - x_j\|^2 < \varepsilon$.

LPP will try to optimally preserve this graph in choosing projections. After constructing the graph, weights have to be assigned to the edges. Therefore a sparse symmetric matrix S of size $m \times m$ is created with $S_{k,j}$ having the weight of the edge joining vertices k and j, and 0 otherwise. The weights are calculated as follows:

$$S_{k,j} = \begin{cases} e^{\frac{\|x_k - x_j\|^2}{t}}, & \text{if } \|x_k - x_j\|^2 < \varepsilon \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

The constant values t and $\varepsilon > 0$ have to be chosen adaptively. Here, ε defines the radius of the local neighborhood. Therefore, the objective function of LPP is defined as

$$w_{opt} = \min \sum_{kj} (y_k - y_j)^2 S_{kj}. \tag{12}$$

Following some simple algebraic steps, it is possible to show that Eq. (12) finally results in a generalized eigenvalue problem:

$$XLX^T w = \lambda XDX^T w, \tag{13}$$

where D is a diagonal matrix whose entries are column sums of S and $L = D - S$ is the so called Laplacian matrix. The k-th column of matrix X is x_k.

The projection matrix W is constructed by concatenating the solution to the above equation, *i.e.* the column vectors

of $W_{LPP} = [w_1, \cdots, w_m]$ are ordered ascendingly according to their eigenvalues. Similar to the Fisherfaces approach, the image set is usually projected into the PCA subspace before applying LPP by deleting the smallest principle components. Thus, the final embedding is as follows:

$$W_{final} = W_{PCA} W_{LPP}. \tag{14}$$

Details about the algorithm and the underlying theory can be found in [7].

2.2 Classification

2.2.1 Classical Approaches

A Support Vector Machine (SVM) is a discriminative classifier, attempting to generate an optimal decision plane between feature vectors of the training classes. Oftentimes, classification with linear separation planes is not possible in the original feature space for real-world applications. Using a so called kernel trick, the feature vectors are transformed to a higher dimensional space in which they can be linearly separated. We used the RBF kernel in our experiments.

The Nearest Neighbor (NN) classifier matches a given test sample to a specific class based on the smallest distance in the feature space between the test sample and all training samples. We used the Euclidean distance for the NN classifier in this paper.

2.2.2 Sparse Representation Classification

Another classification approach is the so called Sparse Representation Classification (SRC), which is based on Compressive Sensing (CS), a technique for signal measurement and reconstruction. Recently, SRC has been successfully applied to face recognition and promising results were obtained even under difficult lighting conditions and partial occlusion [10]. It is assumed, that all training samples of a single class lie on one mutual subspace. Given a sufficiently large number of training samples a testvector y of a class C_i can be represented as a linear combination of training samples of this class. Taking the training samples of all classes into account, y can be expressed as

$$y = Ap_0, \tag{15}$$

where A is a matrix containing all vectorized facial training images as column vectors and p_0 is a vector holding the coefficients for the linear combination of training vectors that represent y. Since this vector is naturally sparse if the number of classes is large enough, an unknown sample y can be classified by finding the sparsest representation p_0 solving the above equation.

Since this equation is usually underdetermined, the sparsest solution can be found via a convex optimization problem using l_1-norm minimization:

$$\hat{x} = \arg\min_x \|x\|_1 \quad \text{subject to} \quad y = Ax, \tag{16}$$

Ideally, the nonzero entries in the sparse coefficient vector x will all be associated with the columns of A which represent a single class C_i. However, in real-world examples, noise and modeling error may lead to small nonzero entries associated with different object classes.

Thus, the minimal residual $r_i(y)$ between y and $A\delta_i(\hat{x})$ is chosen to be most likely to indicate the class the test image y belongs to:

$$\min_i r_i(y) = \|y - A\delta_i(\hat{x})\|_2, \tag{17}$$

where δ_i is the characteristic function of class C_i.

Usually, the high dimensional face images are first projected into a lower dimensional subspace using a sensing matrix which underlies the so called Restricted Isometry Property (RIP) [6]. It can be shown that even a randomly generated projection matrix can be used for that purpose. Such a matrix can simply be generated by sampling zero-mean independent identically distributed gaussian entries.

A detailed description of the Compressive Sensing based face recognition algorithm and Sparse Representation Classification in general can be found in [10].

3. VIDEO DATA COLLECTION AND ANNOTATION

The study subjects were 24 chimpanzees (*Pan troglodytes*) separated into two groups, all from the zoo of Leipzig, Germany. Video material from each individual was collected between June 2010 and December 2010. We placed a High Definition camera (Sony Handycam, 3.1 MegaPixel, 25x optical zoom) with a tripod on one of five observation platforms from which we had a barrier-free view down into the enclosures of the study groups. In general, individuals were recorded for one to five minutes depending on their activity level (the more active the focal animal, the longer the recording time). The aim was to capture diverse poses and different expressions under varying lighting conditions for each individual. We then used an annotation tool to mark the region of the faces and the position of eyes, mouth and earlobes to normalize the facial images before applying the face recognition algorithms explained above. In addition to the position of the head and facial feature points, we also added metainformation to the facial images, such as lighting, species, gender, pose, expression, age and identity. The whole chimpanzee database contains 1839 images of 24 different individuals. The entire annotated dataset we used in our experiments is publicly available at http://www.saisbeco.com/files/resources.html.

4. EXPERIMENTS AND RESULTS

The dataset we gathered in the zoo is very challenging for face recognition because it contains images with huge variation in lighting, expression and pose. For an accurate identification of the individuals, it is necessary to either control the conditions in which the images are taken or to apply a face recognition technique that is robust to those kinds of variation.

Since the habitat of the primates should be kept in its original state and the primates cannot be in direct contact with humans, the latter method can hardly be realized in a real-life scenario. Since the face recognition methods we apply in our experiments are very sensitive to extreme changes in pose and occlusion, we only use frontal face images with different vertical directions and reasonable occlusion. Examples of 3 different individuals with different expressions and lighting conditions as well as partially occluded primate faces can be seen in Figure 1. For our experiments we used 517 facial images of 24 individuals.

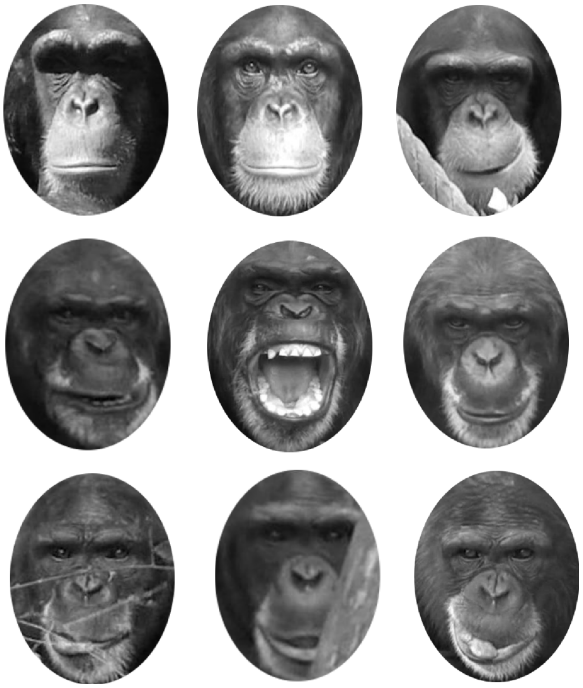

Figure 1: Three different chimpanzee individuals (rows) with different expressions, lightings and partial occlusion (e.g. third row). ©Laura Aporius - MPI EVA (2010); WKPRC (Zoo Leipzig)

The number of images per individual varies between 13 and 38. Using the data from the annotation process, we eliminated all images with low quality. We cropped and rotated the images into an upright position and applied a projective transformation to ensure comparability of the primate faces. All images were scaled to a size of 100×80 pixel, converted to gray scale, vectorized and normalized to unity. We applied a 10-fold Monte Carlo cross-validation in our experiments to validate our results, *i.e.* we randomly split our data into training and test data using 75% of all the data per individual for training and the rest for testing. We repeated this procedure for all of the 10 iterations and averaged the results over all folds.

For PCA we used energy thresholding to remove the last few eigenvectors and took only the first p eigenvectors to improve the performance such that

$$\min_p \frac{\sum_{k=1}^{p} \lambda_k}{\sum_{k=1}^{n} \lambda_k} \geq \tau, \tag{18}$$

where λ_k ($\lambda_k \geq \lambda_{k+1}$) is the eigenvalue of the k-th eigenvector, n is the total number of eigenvalues and τ is a threshold. We found $\tau = 0.85$ to perform best in a pre-experiment resulting in 63 Eigenfaces. Note that for Fisherfaces, the maximal number of features is limited to $C-1$, where C is the number of classes. We chose to have the maximal number of 23 features for Fisherfaces in our test scenario. In the Laplacianfaces approach, we achieved the optimal results using 160 Laplacianfaces. For Randomfaces, we used a feature dimension of 540 as suggested by [10] for high recognition results. The performance statistics are reported as cumulative match scores. The horizontal axis of the cumulative accuracy graph is the rank and the vertical axis is the cumulated accuracy in percent. Figure 2 provides the cumulative accuracy function for all three classifiers with Eigenfaces (PCA), Fisherfaces (LDA), Laplacianfaces (LPP) and Randomfaces (RF). The mean rank-1 accuracy and its standard deviation over all 10 folds for every approach and every classifier are shown in Table 1. With a 540 dimensional feature vector the CS-framework in conjunction with RF achieved the best results with a rank-1 accuracy of 74.89% and therefore outperformed all other approaches.

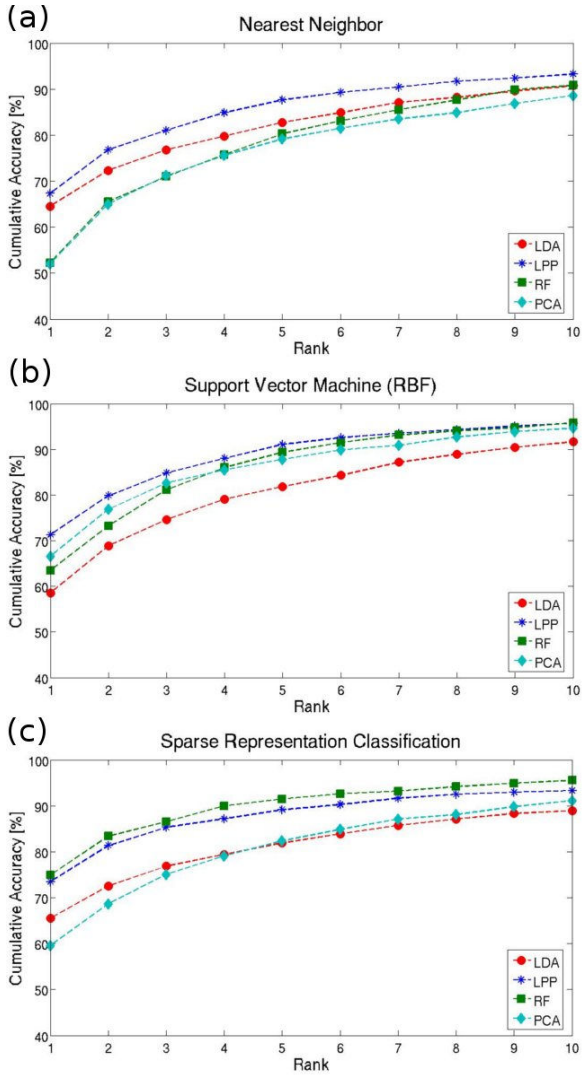

Figure 2: Cumulative Accuracy for Eigenfaces (PCA), Fisherfaces (LDA), Laplacianfaces (LPP) and Randomfaces (RF) with three different classifiers: (a) Nearest Neighbor (NN), (b) Support Vector Machine (SVM) and (c) Sparse Representation Classification (SRC).

The best results achieved by NN and SVM are 67.30% and 71.28% respectively, using the Laplacianfaces algorithm which finds an embedding that preserves local information. It is worth noting that, surprisingly, Eigenfaces and Randomfaces outperformed Fisherfaces using an SVM classifier.

This might be because for SVM the higher feature dimensionality of RF and PCA is more appropriate than the lower dimension of LDA.

Acc. [%] (Std. [%])	NN	SVM	SRC
PCA	51.99 (4.54)	66.80 (2.64)	59.57 (2.46)
LDA	64.54 (2.80)	58.58 (4.44)	65.53 (5.71)
LPP	**67.30 (2.10)**	**71.28 (2.25)**	73.48 (2.88)
RF	52.27 (3.24)	63.48 (2.98)	**74.89 (2.30)**

Table 1: Rank-1 accuracy and standard deviation for Eigenfaces (PCA), Fisherfaces (LDA), Laplacianfaces (LPP) and Randomfaces (RF) with three different classifiers: Nearest Neighbor (NN), Support Vector Machine (SVM) and Sparse Representation Classification (SRC).

For NN, Fisherfaces achieved expectably better results than Eigenfaces and Randomfaces, where RF and PCA have almost the same recognition results. Again, Randomfaces performed slightly better than Eigenfaces because of the higher feature dimension. Note that for LDA, LPP and RF the Sparse Representation Classification achieved better results than the classification by Nearest Neighbor and even Support Vector Machines.

5. CONCLUSION AND FUTURE WORK

In this paper, we proved that state of the art algorithms for appearance based human face recognition are not only capable to identify humans but also primates like chimpanzees. To perform our experiment we first annotated a new dataset consisting of 24 chimpanzee individuals, collected in a zoo, which we published as a public benchmark for the given classification task. Because this primate face database was gathered in an uncontrolled environment, it shows a huge variety of different viewpoints, lightings, expressions and even occlusion and is therefore a very challenging dataset to thoroughly compare different face recognition techniques for individual identification. Besides the position of the head and facial landmarks, such as eyes and mouth, we additionally annotated metadata like occlusion, lightning, pose and image quality as well as other useful information like gender, age and identity. Afterwards, we evaluated different state-of-the-art algorithms for human face recognition, including Eigenfaces, Fisherfaces, Laplacianfaces and a novel technique called Randomfaces, in combination with different classifiers for the identification of frontal primate faces. Despite the fact that the dataset we used in our experiments was gathered in a real life scenario, most of the applied face recognition algorithms achieved good results. The best results were obtained by the Sparse Representation Classification (SRC) using a randomly generated projection matrix with a rank-1 recognition rate of 74.89% and therefore outperformed all other appearance based face recognition approaches and classifiers. In future works we want to extend our work by taking other primate species like gorillas and gibbons into account. Therefore we need to build several high-quality publicly available datasets of different great ape species gathered in the zoo and in the field as a public benchmark for primate identification. We also expect more accurate results for face recognition on great apes using global and local descriptors in combination with a hierarchical classification paradigm or multiple kernel learning. Additionally we will extend our work to face recognition in video recordings including the detection of primate faces and facial feature detection.

6. ACKNOWLEDGMENTS

This work was funded by the German Federal Ministry of Education and Research (BMBF) under the "pact for research and innovation".
We thank the Zoo Leipzig and the Wolfgang Köhler Primate Research Center (WKPRC), especially Josep Call and all the numerous research assistants, zoo-keepers and Josefine Kalbitz for support and collaboration. Financial support is gratefully acknowledged from the Max Planck Society. We also thank Dipl. Biol. Laura Aporius for providing videos and pictures in 2010 and for the annotation of data.

REFERENCES

[1] A. Ardovini, L. Cinque, and E. Sangineto. Identiying elephant photos by multi-curve matching. In *Journal of the Pattern Recognition Society, Vol. 41, 1867 - 1877*, 2007.

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. In *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, pp. 711 - 720*, 1997.

[3] S. Blake, S. Strindberg, P. Boudjan, C. Makombo, I. Bila-Isia, O. Ilambu, F. Grossmann, L. Bene-Bene, B. de Semboli B, V. M. D. S'hwa, R. Bayogo, L. Williamson, M. Fay, J. Hart, and F. Maisels. Forest Elephant Crisis in the Congo Basin. In *PLoS Biol, Vol. 5*, 2007.

[4] T. Burghardt. *Visual Animal Biometrics - Automatic Detection and Individual Identification by Coat Patterns*. PhD thesis, University of Bristol, Faculty of Engineering, Department of Computer Science, 2008.

[5] G. Campbell, H. Kuehl, P. N. Kouamé, and C. Boesch. Alarming decline of West African chimpanzees in Côte d'Ivoire. In *Current Biology, Vol. 18, No. 19, R904 - 905*, 2008.

[6] E. J. Candes and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies? In *IEEE Transaction on Information Theory, Vol 52 (12), pp. 5406 - 5425*, 2006.

[7] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang. Face Recognition Using Laplacianfaces. In *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 3, pp. 328 - 340*, 2005.

[8] M. Lahiri, R. Warungu, D. I. Rubenstein, T. Y. Berger-Wolf, and C. Tantipathananandh. Biometric animal databases from field photographs: Identification of individual zebra in the wild. In *ACM International Conference on Multimedia Retrieval (ICMR)*, 2011.

[9] M. A. Turk and A. P. Pentland. Face recognition using Eigenfaces. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR, pp. 3 - 6*, 1991.

[10] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust Face Recognition via Sparse Respresentation. In *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 31, No. 2, pp. 210 - 226*, 2009.