PROCEEDINGS

55. IWK

Internationales Wissenschaftliches Kolloquium
International Scientific Colloquium

13 - 17 September 2010

Crossing Borders within the ABC

Automation,

Biomedical Engineering and

Computer Science

Faculty of
Computer Science and Automation

www.tu-ilmenau.de

TECHNISCHE UNIVERSITÄT
ILMENAU

Home / Index:
http://www.db-thueringen.de/servlets/DocumentServlet?id=16739

Impressum
Published by

Publisher: Rector of the Ilmenau University of Technology
Univ.-Prof. Dr. rer. nat. habil. Dr. h. c. Prof. h. c. Peter Scharff

Editor: Marketing Department (Phone: +49 3677 69-2520)
Andrea Schneider (conferences@tu-ilmenau.de)

Faculty of Computer Science and Automation
(Phone: +49 3677 69-2860)
Univ.-Prof. Dr.-Ing. habil. Jens Haueisen

Editorial Deadline: 20. August 2010

Implementation: Ilmenau University of Technology
Felix Böckelmann
Philipp Schmidt

USB-Flash-Version.

Publishing House: Verlag ISLE, Betriebsstätte des ISLE e.V.
Werner-von-Siemens-Str. 16
98693 Ilmenau

Production: CDA Datenträger Albrechts GmbH, 98529 Suhl/Albrechts

Order trough: Marketing Department (+49 3677 69-2520)
Andrea Schneider (conferences@tu-ilmenau.de)

ISBN: 978-3-938843-53-6 (USB-Flash Version)

Online-Version:

Publisher: Universitätsbibliothek Ilmenau
ilmedia
Postfach 10 05 65
98684 Ilmenau

Home / Index:
http://www.db-thueringen.de/servlets/DocumentServlet?id=16739

SAISBECO - A SEMI-AUTOMATED AUDIOVISUAL SPECIES AND INDIVIDUAL IDENTIFICATION SYSTEM FOR BEHAVIORAL ECOLOGICAL RESEARCH AND CONSERVATION

*Alexander Loos**

Fraunhofer IDMT
Audio-Visual Systems
98693 Ilmenau, Germany

Hjalmar Kuehl†

Max Planck EVAN
Primatology
04103 Leipzig, Germany

ABSTRACT

The current biodiversity crisis and the accompanied catastrophical declining of species populations is startling. Therefore autonomous monitoring techniques become more and more important. However, the manual annotation of the accruing data is an enormous time consuming and tedious work. This paper brief y presents the current research project SAISBECO, which aims at the identif cation of species and individuals of great apes like gorillas and chimpanzees. A pre-study on gorilla face recognition shows the feasibility of human face recognition algorithms to identify also primate faces. Though, linear models are not intrinsically equipped to deal with variations in lighting, viewpoint and expression for instance. Therefore multilinear techniques based on the Higher-Order Singular Value Decomposition *HOSVD* promise to be more robust against suchlike variations.

1. INTRODUCTION

In this paper we brief y present a collaborative research project of Fraunhofer IDMT, Fraunhfoer IIS and the Max-Planck Institute for Evolutionary Anthropology and present f rst ideas of face recognition of great apes like gorillas and chimpanzees. The remainder of this paper is organized as follows: First, we give a overview of the motivation and aims of the project. In section 2 we are going to give an overview of existing approaches to identify individuals of other species including whales, elephants and bobcats. Section 3 covers a pre-study which was carried out at Fraunhofer IDMT regarding gorilla face recognition using several lighting normalization techniques and the well-known Fisherface analysis [1]. Then we are going to give further ideas of improvements of the applied face recognition algorithm. We conclude this paper by proposing future research in the f eld of individual recognition of animals in wild life scenarios.

*Thanks to Fraunhofer Insitute for funding.
†Thanks to Max Planck Insitute for funding.

1.1. Motivation

In the ongoing biodiversity crisis species extinction rates are very high and thousands of species populations are catastrophically declining [2, 3, 4]. An essential part of effective biodiversity and wildlife conservation management is population monitoring, to assess factors such as population size, and viability or to evaluate the success of implemented protection measures [5]. Therefore, the development of new monitoring approaches is a f eld of intense research [6]. Individual identif cation is a prerequisite for many approaches, such as capture-mark-recapture techniques.

Similarly, individual identif cation is required for many questions in behavioral ecological research, ranging from wildlife epidemiology to between population comparisons of social dynamics or the evolution of social behavior and cognition. However, the problem is that for many species identif cation of individuals of wild populations is often a non-trivial and extremely time consuming task, which can take up to several years [7, 8]. In captive settings the individual identif cation is less of a problem; here the processing of large amounts of data is the tedious work. In recent years, the availability of digital recording devices has facilitated the collection of large amounts of species and individual data, for instance with remote camera traps or autonomous recording devices in the wild. However, the manual analysis of video and audio recordings is a very tedious work. Consequently, there is a high demand for automated analytical routine procedures of video and audio recordings for both accelerated data processing and improved data quality. This would not only help to collect and process large amounts of individual data, but would also release previously bounded resources for less repetitive and more innovative tasks.

The Fraunhofer Institutes contribute their long established experience in audio and video signal processing, face recognition and detection as well as machine learning to this project. The objective is an intensive exploitation of algorithms developed in previous research projects for the very new f eld of individual recognition of primates.

After a f rst meeting of the aff liated partners in the run-up to this project, existing technologies of both Fraunhofer Institutes have been successfully applied for the recognition and detection of primates within the scope of very f rst tests. However, a special challenge is the lack of comprehensive reference databases, which are essential for the development of algorithms for recognition and supervised learning. Even in the f eld of human face recognition, test data for unconstraint environments is still an urgent requirement after twenty years of research. Therefore the development of a comprehensive, reliable database is an important component of this project, which is supposed to be available to researches after this project has f nished. In the view of the originating conditions, semi-automatic approaches are specif cally taken into account, which can be optimized by interaction and user feedback. Thus researchers of the Max-Planck Institute are able to train the system even after the end of this project.

1.2. Aims and Objectives

The objective of this project is threefold:

1. to collect high quality visual and audio species and individual data
2. to develop procedures to automatically identify species, age and sex classes and individuals from these data, and
3. substantiate the value of the approaches taken with subsequent analysis of the processed data for conservation, behavioral and ecological research.

We will conduct this project with a particular emphasis on african great apes. Chimpanzees and gorillas are both threatened species and thus in the focus of conservation efforts, but they are also of particular interest in behavioral-ecological research in the human evolution context. In addition, f eld ape research is permanently confronted with the problem of limited sample sizes. This is because apes are highly elusive, and in order to collect necessary data researchers habituate them in a long lasting process to human observers. Therefore, any new approach providing larger sample sizes would be highly valuable. However, to demonstrate the wide applicability of the approaches to be taken, we will also include other apes (orangutans, gibbons, siamangs) and large mammal species into our study, such as smaller bodied primates and elephants. Figure 1 shows the schema of the project.

The specif c objectives of the proposed study are presented in chronological order:

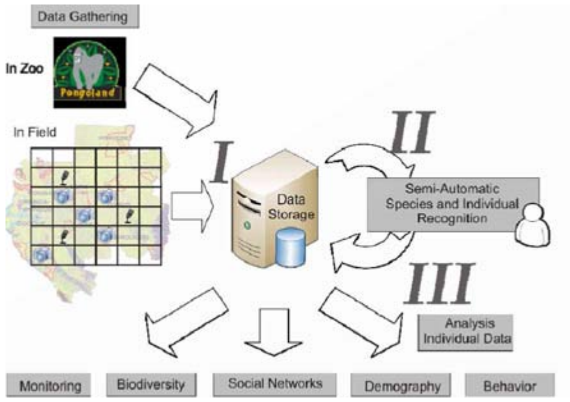

Fig.1: Schema of the project, consisting of three phases.

1. Collection of audiovisual data:
   The f rst purpose is to collect long-term video sequences and audio material with remote camera traps on great apes and the large mammal fauna in Loango National Park, Gabon and Leipzig zoo to create a database for the storage of collected video-, still image-, audio- and meta-data. An optimal placement and location of automonous recording devices is indispensible for a suff cient automated recognition. Therefore the evaluation of the camera type, positioning and light condition within an experimental f eld study will be conducted.
2. Semi-automatic identif cation:
   In this task we focus on the modif cation and evaluation of different existing approaches for the detection and recognition of ape faces as well as multiple object tracking algorithms and the development of visual features for species recognition. Besides that, algorithms for eff cient automatic extraction of target species vocalizatin from long-term recordings will be implemented and evaluated.
3. Analysis and synthesis of processed individual data:
   The f nal objective of this project ist the demonstration of the value of remotely collected and subsequently processed audiovisual species data for conservation monitoring, wildlife epidemiology and behavioral ecological research.

2. STATE OF THE ART

A small number of articles provide f rst approaches to the autimatic identif cation of species or individuals. For the purpose of this workshop we want to focus on approaches using video or image analysis.

Araabi et. al. for instance presented a syntactic and semantic string representation scheme as well as a string matching method as part of a computer-assisted system to identify dolphins from photographs of their dorsal f n [9].

A similar approach was used by Ardovini et. al. to develop a semiautomatic elephant photo identif cation system based on the shape comparison of the nicks characterizing the elephant's ears [10]. The method can deal with very cluttered and noisy images as the ones commonly used by zoologists for wild elephant photo identif cation. Diff cult segmentation problems were solved using rough position information input by the user. The system achieved a top-10-recognition rate of 85%, meaning that an individual could be found within a set of 10 individuals suggested by the system. However, that an individual was ranked on the f rst place was just 46%.

Ranguelova et. al. presented a photo-id system for humpack whales [11]. Individuals of this species can be uniquely identif ed by the light and dark pigmentation patches on their tails. The authors developed an interface that assisted the user in segmenting the animal's tail from the sea and f tting an aff ne invariant coordinate grid to it. A numerical feature vector capturing the patch-distribution with respect to the grid was then automatically extracted and used to match the individual against a database of similarly processed images. A recognition rate of 70% was obtained.

A study of Heilbrun et. al. evaluated whether bobcats (Lynx rufus) could be photographed effectively using automatically triggered cameras and whether adult bobcats could be individually identif ed by their natural markings [12]. They obtained 76 photographs of bobcats, of which 88% contributed to individual identif cation. The ability to identify bobcats was signif cantly inf uenced by the body position in which they were photographed. The most common positions enabled the identif cation of unique pelt and facial markings.

All these methods use characteristic pattern of fur and skin or individual nicks in ears or the dorsal f n to distinguish between individuals. Such a technique is not feasible for the individual recognition of great apes because primates like chimpanzees or gorillas don't have such special markings on their fur. Furthermore the system proposed by Ardovini et. al. [10] is only suited for Serengeti elephants due to their ear nicks. Other elephants usually don't have such an easy to f nd feature for individual identif cation. A system that individually recognizes apes on the basis of audio-visual features or even by facial recognition algorithms is, to the best of our knowledge, not available yet.

3. PILOT STUDY - GORILLA FACE RECOGNITION

A pre-study on the individual recognition of gorilla faces based on a combination of different lightning-normalization techniques and the well known Fisherface analysis [1] showed, that an identif cation of primates is in principle possible.

Figure 2 shows a selection of the analyzed gorilla faces.

Fig.2: Faces of 5 different gorillas.

Nine individuals and a maximal number of nine images per class mostly with frontal faces were considered. For this database, six sets of experiments were performed with 4 up to 9 pictures per class for training and one for testing. This was repeated such that each observation was used once as the validation data and the mean accuracy has been calculated. A Support Vector Machine (SVM) was used as classif er. Two lightning normalization techniques, the Contrast Limited Adaptive Histogram Equalization (CLAHE) and the normalization based on Gross' and Bravjovic's work [13], were compared. Figure 3 shows the results.

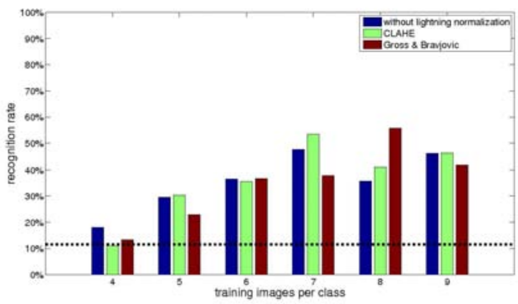

Fig.3: Results of the face recognition pre-study.

The number of images per individual was increased from 4 up to 9 images, which is illustrated on the x-axis in f gure 3. The y-axis shows the results of the Fisherface technique and the different normalization methods. The blue bars are the results in case no lightning-normalization is used, the green bars show the results of the CLAHE- technique and the red bars sign the results of the method developed by Gross and Bravjovic. The dashed line shows the result that a random classif er would produce. A signif cant decrease of the recognition rate for all normalization techniques could be seen if more than seven images per class are considered. This is because the eighth and ninth image in every class are semi-frontal faces. The best results were obtained by the lightning normalization technique based of Gross and Bravjovic with a detection rate of 55.69%, which is way above the results of the random classif er (11.1%).

The pre-study proves the possibility to identify individuals in an ape- population with the combination of lightning normalization techniques and face recognition methods, like the Fisherface analysis. However, the results also show that the success of face recognition is limited because of the sensitivity to extreme changes in pose and expression, which is supposed to appear quite often in a wild-life scenario.

Modern techniques like methods based on tensor analysis, promise to be more invariant against changes in pose, expression and lightning but were not yet tested on primates. We will give a short introduction to face recognition using tensor analysis and higher-order singular value decomposition in the following section.

4. TENSOR BASED FACE RECOGNITION

Facial image recognition across different lightings, poses and expression is after more than 20 years of research still a very challenging task. Linear methods such as Eigenfaces [14] or Fisherfaces [1] usually are insuff - cient if there are variations in lighting, expression etc. and are therefore only suitable when the identity of the person is the only factor to be considered.

Recently multilinear algorithms [15, 16] have been proposed to accommodate variations across multiple factors in a natural way resulting in higher accuracy then linear models and lower computational costs then non-linear methods.

4.1. Tensor Model of Face Database

Vasilescu and Terzopoulos [15] presented a multilinear modeling technique that arranges faces of different persons under different lightings, viewpoints and expressions in a tensor structure. A tensor is therefore nothing else than a multidimensional array. For simplicity, let us assume that the face database only contains images of different individuals with variations in lighting and viewpoint. Then, the tensor representation of the face database is given by

$$\mathcal{T}(i_p, i_l, i_v) = I_{P_{i_p}, L_{i_l}, V_{i_v}} \quad (1)$$

where $I_{P_{i_p}, L_{i_l}, V_{i_v}}$ is the vecorized image of the i_p-th person with i_l-th lighting and i_v-th viewpoint or pose. Here $\mathcal{T} \in \mathbb{R}^{N_p \times N_l \times N_v \times N_x}$ is a 4-D tensor representing the unif ed images of the whole dataset, where N_p is the number persons or individuals, N_l and N_v are the number of lightings and viewpoints respectively. N_x is the number of pixels of the image faces.

4.2. Higher Oder Singular Value Decomposition

The Multilinear Singular Value Decomposition or Higher-Order Singular Value Decomposition *HOSVD*, presented by De Lathauwer et. al. in 2000 [17], is the multi dimensional counterpart of the well known matrix SVD. Computing the SVD of a data matrix $D \in \mathbb{R}^{I_1 \times I_2}$ results in the decomposition of the matrix D and orthogonalizes the two associated vector spaces U and V as,

$$D = U\Sigma V^T = \Sigma \times_1 U \times_2 V \quad (2)$$

where Σ is a pseudo-diagonal matrix containing the eigenvalues in descending order, U is a unitary matrix containing the eigenvectors of $D \cdot D^T$ as columns, V is also a orthonormal matrix given by the eigenvectors of $D^T \cdot D$ and $\times_k$ denotes the *mode-k* product. That means that in the same way that U is associated with the column space of D , V is associated with the row space of D. This theorem can easily be generalized to the n-th order case.

Similarly the HOSVD yields n orthogonal subspaces for an n-th order tensor, where each subspace corresponds to one particular mode of variation. If $\mathcal{D} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_n}$ the HOSVD of $\mathcal{D}$ can be calculated as follows:

$$\mathcal{D} = \mathcal{S} \times_1 U^1 \times_2 U^2 \ldots \times_k U^k \ldots \times_n U^n \quad (3)$$

where the orthonormal matrix U^k contains the ordered principal components for the k-th mode of $\mathcal{D}$. $\mathcal{S}$ is called the *core tensor* and is for higher order cases not necessarily dioganal. Though, the core tensor $\mathcal{S}$ fulf lls the property of *all-orthogonality*, meaning that two arbitrary subtensors of $\mathcal{S}$ along one arbitrary dimension are orthogonal.

4.3. Overview of Existing Approaches

A tensor based decomposition approach was f rst developed by [15] to provide a algorithm that performs better than linear models and on the same time requires less computational costs than non-linear models. However, this method used only the person-mode decomposition effectively, whilst decomposition in other modes was neglected. Therefore, in [16] the authors criticized that the previous approach was not suitable for face recognition under unknown lightings and viewpoints as the projection basis do not contain information on lighting and viewpoint variation. However, this approach was ref ned and improved in [16] by developing a unif ed tensor framework for face recognition that unites a number of algorithms. All tensor based approaches are based on the HOSVD and they differ only in the way that the projection basis is calculated by multiplying and reshaping the tensor-product of the resulting core tensor and the specif c eigen-mode matrices.

Beginning with the datatensor $\mathcal{T} \in \mathbb{R}^{N_p \times N_l \times N_v \times N_x}$ from equation 1, the HOSVD yields four orthogonal subspaces and is given by

$$\mathcal{T} = \mathcal{S} \times_1 U^P \times_2 U^L \times_3 U^V \times_4 U^X \quad (4)$$

where the columns of U^P, U^L, U^V and U^X def ne the person, lighting, viewpoint and the pixel subspace respectively. The columns in U^X represent the traditional eigenfaces and the columns of U^P, U^L and U^V represent the *eigen-person*, *eigen-lighting* and *eigen-viewpoint* respectively.

In the following we only present the *MPCA-PS* algorithm as it promises the best results for unkown lightings and viewpoints, which is expected to occur quite often in a wild-life scenario. MPCA-PS utilising *person-specif c eigen-modes* as the basis of projection by calculating the mode-1 product of the core-tensor $\mathcal{S}$ and the matrix of the person-subspace:

$$\mathcal{A} = \mathcal{S} \times_1 U^P \quad (5)$$

In a next step the tensor $\mathcal{A}$ is unfolded in the person-mode, resulting in,

$$A_{person} = \begin{bmatrix} I^e_{P_1 L_1 V_1} & \cdots & I^e_{P_1 L_{N'_l} V_{N'_v}} \\ I^e_{P_2 L_1 V_1} & \cdots & I^e_{P_2 L_{N'_l} V_{N'_v}} \\ \vdots & \vdots & \vdots \\ I^e_{P_{N_p} L_1 V_1} & \cdots & I^e_{P_{N_p} L_{N'_l} V_{N'_v}} \end{bmatrix} \quad (6)$$

where $I^e_{P_{i_p} L_{i_l} V_{i_v}}$ is the *eigen-image* of the i_p-th person at i_l-th *eigen-lighting* and i_v *eigen-viewpoint*. Furthermore N_p, $N'_l (\leq N_l)$ and $N'_v (\leq N_v)$ is the number of dominant eigenvectors of the person, lighting- and viewpoint subspaces respectively.

As testing criteria, the minimal reconstruction error is used. The algorithm for testing is def ned as follows, where I_T ist the vectorized test image:

1. Calculate: $I'_T = I_T \times U^X$
2. For every training image $i = 1, \ldots, N_p$ calculate
   - $c_i = I'_T \times A_i^+$
   - $I_T^{\prime recon} = c_i \times A_i$
   - $error = \|I'_T - I_T^{\prime recon}\|$
3. The training image with the minimal reconstruction error $\min(error)_{i=1}^k$ is assigned to the test image.

Here A_i for $i = 1, \ldots, N_p$ are the person-specif c bases and can be calculated as

$$A_i = \begin{bmatrix} I^e_{P_i L_1 V_1} \\ I^e_{P_i L_2 V_1} \\ \vdots \\ I^e_{P_i L_{N'_l} V_{N'_v}} \end{bmatrix} \quad (7)$$

and A_i^+ is the Moore-Penrose-Pseudoinverse of A_i.

This approach of tensor-based face recognition gives excellent results on human face databases like Weizmann face database, Extended YaleB [18] and PEAL [19] even under unknown lightings and viewpoints and outperforms linear models like Eigenfaces. However, this algorithm was never tested for recognizing ape faces. Nevertheless, we are conf dent that tensor based face recognition will peform well for the identif cation of primates.

5. CONCLUSION AND FUTURE WORK

In this paper we gave a brief overwiew on our current interdisciplinary joint project of Fraunhofer and Max-Planck Institutes on the identif cation of species and individuals in wild life scenarios. We presented a pre-study by the Fraunhofer Institute of Digital Media Technology dealing with gorilla face recognition. Therefore we exploited the well known Fisherface analysis in combination with several lighting normalization techniques. Additionally we presented a novel face recognition technique developed by [16], exploiting the interaction of subspaces, resulting from multilinear analysis. This approach outperforms traditional face recognition methods based on linear algebra and is even capable of handling unseen variations in images.

In future works we want to create a public high-quality dataset of great apes like gorillas and chimpanzees. Furthermore, we are going implement and compare several face recognition algorithms, including multilinear methods as well as non-linear algorithms, to achieve high recognition rates.

6. REFERENCES

[1] Belhumeur et. al., "Eigenfaces vs. f sherfaces: recognition using class specif c linear projection," in *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, pp. 711 - 720*, 1997.

[2] Campbell et. al., "Alarming decline of west african chimpanzees in cte d?ivoire," in *Current Biology, Vol. 18, No. 19, R904 - 905*, 2008.

[3] Blake et. al., "Forest elephant crisis in the congo basin," in *PLoS Biol 5(4): e111*, 2007.

[4] Walsh et. al., "Catastrophic ape decline in western equatorial africa," in *Nature 422:611 - 614*, 2003.

[5] Nichols and Williams, "Monitoring for conservation," in *Trends in Ecology and Evolution 21, pp. 668 - 673*, 2006.

[6] Buckland et. al., "Introduction to distance sampling," in *Oxford University Press, Oxford*, 2001.

[7] Bertolani and Boesch, "Habituation of wild chimpanzees (pan troglodytes) of the south group at tai forest, cote d'ivoire: Empirical measure of progress," in *Folia Primatol (Basel) 79:162 - 171*, 2007.

[8] Boesch and Boesch-Achermann, "The chimpanzees of the tai forest," in *Oxford: Oxford University Press*, 2000.

[9] Araabi et. al., "A string matching computer-assisted system for dolphin photoidentif cation," in *Annals of Biomedical Engineering, Vol. 28, 1269 - 1279*, 2000.

[10] Ardovini et. al., "Identiying elephant photos by multi-curve matching," in *Journal of the Pattern Recognition Society, Vol. 41, 1867 - 1877*, 2007.

[11] Ranguelova et. al., "Towards computer-assisted photo-identif cation of humpback whales," in *IEEE International Conference on Image Processing, pp. 1727 - 1730*, 2004.

[12] Heilbrun et. al., "Estimating bobcat abundance using automatically triggered cameras," in *Wildlife Society Bulletin, Vol 34, 69 - 73*, 2006.

[13] Gross and Bravjovic, "An image preprocessing algorithm for illumination invariant face recognition," in *4th International Conference on Audio- and Video-Based Biometric Person Authentif cation, pp. 10 - 18*, 2003.

[14] Turk and Pentland, "Face recognition using eigenfaces," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR, pp. 3 - 6*, 1991.

[15] Vasilescu and Terzopoulos, "Multilinear analysis of image ensembles: Tensorfaces," in *European Conference on Computer Vision ECCV, pp. 447 - 460*, 2002.

[16] Rana et. al., "A unif ed tensor framework for face recognition," in *Journal on Pattern Recognition, Vol. 42, pp. 2850 - 2862*, 2009.

[17] De Lathauwer et. al., "A multilinear singular value decomposition," in *SIAM Journal on Matrix Analysis and Applications, Vol. 21, pp. 1253 - 1278*, 2000.

[18] Lee et. al., "Acquiring linear subspaces for face recognition under variable lighting," in *IEEE PAMI, Vol 27(5), pp. 684 - 698*, 2005.

[19] Gao et. al., "The cas-peal large-scale chinese face database and baseline evaluations," in *IEEE SMC Part A, Vol 38(1), pp. 149 - 161*, 2008.