 ISBN 978-3-200-02328-4

TOWARDS AUTOMATED VISUAL IDENTIFICATION OF PRIMATES USING FACE RECOGNITION

Alexander Loos, Martin Pfitzer

Fraunhofer IDMT
Audio-Visual Systems
Ehrenbergstr. 31, 98693 Ilmenau, Germany

ABSTRACT

In this paper we propose and evaluate a face recognition approach for the individual identification of great apes. We extend our previous work to a more automatic approach using an unsupervised face alignment method known as congealing instead of a projective transform based on manually annotated facial feature points. Furthermore we present an improved version of the Randomfaces approach, called Hybridfaces, which complements the global recognition results with information obtained from local facial regions. We evaluate our approach on three publicly available primate databases of captive chimpanzees, free-living chimpanzees, and free-living western lowland gorillas. Our proposed framework shows promising results and the Hybridfaces approach clearly outperforms the previously used basic Randomfaces method on all three datasets.

Index Terms— Face Recognition,Congealing, Great Apes

1. INTRODUCTION

In recent years many species, especially great apes like chimpanzees or gorillas, are threatened and need to be protected. An essential part of biodiversity conservation management is non-invasive population monitoring using remote camera devices. Because of the huge amount of data, the manual annotation of the video files is highly cost and labor intensive. Especially individual identification of animals and evaluation of population sizes is a prerequisite for many biological questions like social network analysis, wildlife epidemiology or behavior analysis. Consequently, there is a high demand for automatic analytical routine procedures. The field of computer vision becomes more and more important for the individual recognition of animals. Several approaches for individual animal identification have been published.

An automatic recognition approach for individual identification of African Penguins was proposed and evaluated by [1]. The authors suggest to compactly capture penguin's individuality using the appearance context of a number of reference points on individually-specific coat patterns. Recently, an algorithm called StripeCodes was proposed by [2]. The authors claim that their algorithm efficiently extracts simple image features used for the comparison of zebra images to determine if the animal has been observed before.

The aforementioned approaches use characteristic coat patterns to distinguish between individuals. Such a technique cannot be used for species without obvious unique markings like primates for instance. Therefore, we proposed to use face recognition technologies to identify great apes in our previous work [3]. Assuming that primates and humans share similar face properties we used established appearance based face recognition approaches originally developed for human identification to distinguish between primate individuals.

In this paper we extend our previous work to a more automatic framework for face recognition for great apes, since we use an unsupervised alignment method called congealing. Furthermore we demonstrate that, if the face images are aligned, our Hybridfaces method outperforms the Randomfaces approach by [4]. The proposed method is evaluated on three publicly available datasets.

The rest of the paper is organized as follows: In the following section 2 we review or previous research from [3], followed by a brief presentation of the alignment technique we used in our experiments in 3. Section 4 covers the two face recognition approaches we compare in this paper. The first is the well-known Randomfaces approach by [4] followed by our improved Hybridfaces algorithm. In section 5 the results of the experiments we conducted are presented and discussed. Section 6 concludes the paper giving a summary of our work and future ideas of improvement.

2. PREVIOUS WORK

In general, the basic face recognition pipeline is as follows: After all possible faces of an image or a video file are detected, all faces should be aligned before applying the face recognition methods. This usually improves the accuracy of the identification process significantly, especially for appearance based methods. For the alignment step we previously used manually annotated facial feature points like eyes and mouth to apply a projective transform to every image in our datasets. Instead of automatic facial fiducial point detection methods, which are usually computationally very expensive, we suggest to use an unsupervised joint alignment method known as congealing [5]. One of the most established and well studied approaches for the following face recognition step are appearance based methods. Here the two dimensional gray level images of size $w \times h$ are reshaped to vectors of size $n = w \cdot h$. Since this high-dimensional feature space is too large to perform fast and robust face recognition, dimensionality reduction techniques like Principle Component Analysis (PCA) [6], Linear Discriminant Analysis (LDA) [7] or Locality Preserving Projections (LPP) [8] can be used to project the vectorized face images into a lower dimensional subspace. Recently, also a random projection has been successfully used for face recognition in combination with a Sparse Representation Classification (SRC) scheme [4]. This method is known as Randomfaces. Since facial images of primates carry a lot of individual-specific information in the local area around the eyes and nose, we expect improved recognition results by complementing the aforementioned method using additional information of these facial regions. We call this method Hybridfaces as it combines global and local feature extraction.

3. FACE ALIGNMENT

For the automatic face alignment we use a method called congealing developed by [5]. In this work the authors present a system that, given a collection of poorly aligned images of a specific class, automatically generates an "alignment machine" for that specific object class. The basic concept of congealing is as follows:

Given a set of possible feature values for a pixel location $\mathcal{X} = \{1, \ldots, M\}$ one can generate a distribution field over $\mathcal{X}$ for every pixel of an image stack. Congealing now proceeds by iteratively calculating the distribution field of the set of images, choosing the transformation out of a set of affine transformations that minimizes the entropy of the resulting distribution field, update the distribution field and iterate until convergence. Note that an equivalent formulation for minimizing the entropy of the distribution field is to maximize the likelihood of the images.
After congealing has been done for a set of training images, additional unseen images can then be aligned by maintaining the distribution fields from each iteration of congealing and then iteratively align test images according to the distribution field of the corresponding iteration.

For real world images the authors suggest to calculate SIFT descriptors around an 8×8 neighborhood for each pixel of the images and use those as features. Since every SIFT descriptor is a 32-dimensional vector in the original implementation, the corresponding feature space would be too large to efficiently estimate the distribution field. Therefore, instead of using the SIFT descriptors directly, the feature vectors are first clustered using kmeans, resulting in 12 cluster centroids which are finally used as feature values. Details of the underlying theory and the algorithm can be found in [5].

4. FACE RECOGNITION

4.1. Randomfaces

The Randomfaces approach was first published by [4]. Here, the n-dimensional image vectors $\{x_1, \ldots, x_N\}$ for N images are projected into a lower dimensional subspace of size m using a randomly generated projection matrix $W \in \mathbb{R}^{n \times m}$

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

where the m-dimensional vectors $\{y_1, \ldots, y_N\}$ represent the facial images in the feature space which can be used for classification. Usually the Randomfaces technique is used in combination with a Sparse Representation Classification (SRC) scheme. The core step within the SRC is the solution of a convex optimization problem via l_1-norm minimization:

$$\hat{p} = \arg\min_p \|p\|_1 \quad \text{subject to} \quad \tilde{y} = \tilde{A}p, \tag{2}$$

where $\tilde{A}$ is the normalized matrix of training samples transformed into the feature space and $\tilde{y}$ represents the normalized vectorized test image in the feature domain. The coefficients of the sparse vector p associated with the i-th class are 1 and the rest is 0. Finally, the test vector $\tilde{y}$ is assigned to the class that minimizes the residual $r_i(\tilde{y})$ between $\tilde{y}$ and $\tilde{A}\delta_i(\hat{p})$:

$$\min_i r_i(\tilde{y}) = \|\tilde{y} - \tilde{A}\delta_i(\hat{p})\|_2, \tag{3}$$

where δ_i is the characteristic function of class i. A detailed description of the Randomfaces approach and the SRC can be found in [4].

4.2. Hybridfaces

The aforementioned global Randomfaces method extracts features using the holistic face image. Consequently, every dimension of the feature vector contains information from every part of the face image. In contrast, local approaches extract feature vectors solely from certain local regions of the face image e.g. the region around the eyes. These regions carry a lot of individual characteristic information and offer a robust means for identifying individuals since they vary only slightly between different facial expressions. However, local approaches require the exact location of the regions of interest, which is a challenging task for real world applications. Furthermore, all individual-specific information that can be extracted from other parts of the facial image are not taken into account. With the Hybridfaces approach, we try to complement the global face recognition with local information gathered from the regions around the eyes and nose of individuals in an algorithm inspired by [9]. The detection of the local regions of interest is then performed as follows:

All face images of the training set are filtered with a set of 40 Gabor wavelets using 8 orientations and 5 different frequencies. A set of feature points is then generated for every Gabor filtered image by calculating the maximum absolute value within a sliding sensing window. These feature points have to be located in the upper two thirds of the face image to avoid detecting feature points around the mouth-region. We assume that this region is subject to a lot of variation between different poses and expressions and is therefore impractical for identification. All feature points that exceed a certain threshold t_1 are accumulated for all images in the training database. By using a second threshold t_2, only the most frequently appearing feature points are selected. The values for the thresholds t_1 and t_2 were found iteratively on a separate set not used for further training or testing. Figure 1(b) shows the selected feature points.

(a) (b) (c)

Fig. 1. (a) The original face image, (b) selected feature points derived from the complete training set, (c) location of the bounding boxes.

To define the local regions of interest, bounding boxes (patches) with a fixed size are initialized around a feature point and shifted to their final position using a mean-shift algorithm. This algorithm moves the bounding boxes iteratively towards regions with a high density of feature points. Figure 1(c) shows the three bounding boxes around the eyes, the nose, and the eyes and the nose in their final position. Features are extracted for the global image and all patches using Randomfaces resulting in $N_P + 1$ different feature vectors, where N_P are the number of patches. Each feature vector is classified separately using SRC. Then, a weighted classification scheme is applied to get the final result. Figure 2 illustrates the paradigm of combining local and global classification results.

The classification process is performed for every patch as well as the global face image and results in vectors that contain a ranking of the classes the tested individual is assumed to belong to. We

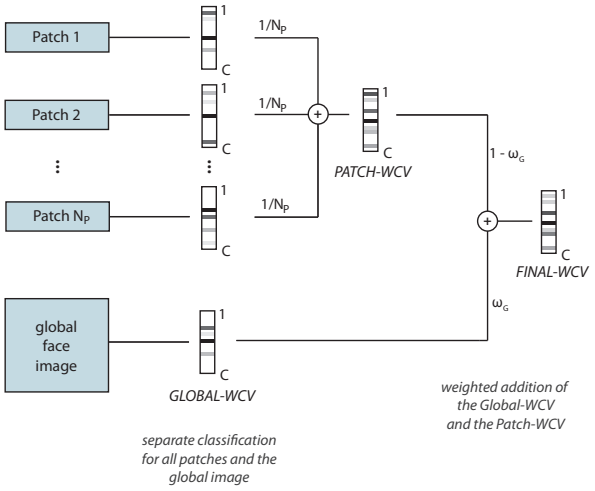


Fig. 2. Paradigm of the combination of local and global classification results for the Hybridfaces algorithm.

then assign weights to each class which are exponentially declining with increasing rank. With this information, *weighted class vectors* (WCVs) of size $1 \times C$ are generated for each classification result, where C is the number of classes. Every WCV contains the weights for each class. For the local patches, these weights are accumulated and normalized resulting in the Patch-WCV V_P. Then, the Global-WCV V_G and the Patch-WCV V_P are weighted and added up resulting in the Final-WCV V_F:

$$V_F = \omega_G V_G + (1 - \omega_G) V_P. \quad (4)$$

Finally, the entries of the Final-WCV are sorted descendingly and the according indices are the final classification results.

5. EVALUATION

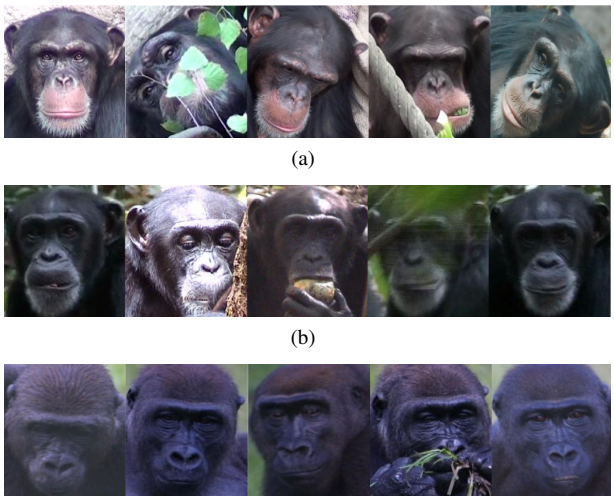
5.1. Databases

Due to the lack of publicly available benchmark databases for primates, we assembled 3 different sets of facial images of different great ape species and made them publicly available on our project website [1]. Table 1 gives an overview of the datasets we used in our experiments. Two of the datasets consist of different chimpanzee individuals, one for captured individuals from the zoo of Leipzig, Germany (ChimpZoo) and one for free-living primates from the Taï National Park, Africa (ChimpTaï). The third dataset contains images of free-living western lowland gorillas from the Odzala National Park, Republic of Congo, Africa (Gorilla). Examples of one single individual per dataset can be seen in Figure 3. All images were annotated by marking the region of the apes face and setting marker points for eyes and the mouth. We also assigned meta-information such as the name of the individual, species, gender, age and pose to every facial image. All of the annotated information was then saved into an XML-file associated with the corresponding image.

Dataset	Origin	Images	Individuals
ChimpZoo	Zoo Leipzig	1839	24
ChimpTaï	Tai NP	3193	71
Gorilla	Odzala NP	1387	230

Table 1. Overview of the datasets we used in our experiments.

[1] http://www.saisbeco.com/files/resources.html



(a) (b) (c)

Fig. 3. Three individuals (two chimps and one gorilla) with varying expression, lightings, and partial occlusion. Images were taken from the datasets (a) ChimpZoo (b) ChimpTaï (c) Gorilla.

5.2. Results

Since our datasets were gathered in uncontrolled environments as opposed to most available human face databases, all three applied datasets are very challenging as can bee seen in Figure 3. Different lighting situations, poses, and even partial occlusion hamper the recognition significantly. Therefore, we only focus on near-frontal faces with arbitrary vertical directions in our experiments. Additionally, we limit our investigations only to individuals with a minimal number of 5 frontal images in the database. The resulting subsets then contain 555 face images of 24 different chimpanzees for the ChimpZoo dataset and 81 images of 15 gorilla individuals for the Gorilla dataset. The number of images per individual for the ChimpZoo dataset varies between 11 and 28 and for the Gorilla dataset between 5 and 6. For the ChimpTaï dataset we could use 607 images of 34 different chimpanzees for our experiments resulting in 8 to 33 pictures per individual. We cropped the faces using the manual labeled regions for the primate faces and resized them to 100×80 pixels. Because we use the congealing algorithm in this work for alignment, we didn't apply any further transformation or manual preprocessing such as rotation. For Randomfaces, we found that a feature dimension of more than 540 doesn't improve the results significantly in a pre-study. The same holds true for the Hybridfaces approach which uses Randomfaces to generate the feature vectors. Here, 540 features were extracted for the global image and 270 features for each of the 3 patches. We found a weighting factor ω_G of 0.3 to perform best in our scenario (compare section 4.2). To validate our results, we used a 20-fold Monte Carlo cross-validation, *i.e.*, we randomly split our data into a training set and test set where we used 75% of the entire collection to train the classifier and the remaining 25% for testing. The rank-1 accuracies and the standard deviation over all 20 folds for Randomfaces (RF) and Hybridfaces

(HYB) with respect to the different alignment techniques are given in Tables 2 (a) - (c). To compare the results with our previous work, the accuracies if a projective transform with manually labeled facial markings is used to align the ape faces, are given as well.

Acc. [%] (Std. [%])	None	Congealing	Projective Transform [3]
(a) ChimpZoo			
RF	**43.11 (3.46)**	64.63 (3.36)	77.70 (3.67)
HYB	42.43 (3.18)	**70.00 (3.49)**	**82.97 (2.62)**
(b) ChimpTaï			
RF	**52.01 (3.32)**	59.91 (4.30)	63.60 (3.53)
HYB	49.70 (2.73)	**61.59 (3.98)**	**64.54 (2.50)**
(c) Gorilla			
RF	**28.67 (6.34)**	69.67 (7.64)	69.83 (6.53)
HYB	27.00 (5.81)	**72.83 (5.75)**	**73.83 (6.42)**

Table 2. Rank-1 accuracy and standard deviation for Randomfaces (RF) and Hybridfaces (HYB) in combination with different alignment methods for the datasets: (a) ChimpZoo (b) ChimpTaï (c) Gorilla.

For all datasets one can see that our Hybridfaces approach consistently outperforms the standard Randomfaces approach if the face images are sufficiently aligned using either congealing or a projective transformation with manually annotated facial landmark points. Not only the rank-1 accuracy improves but also the standard deviation across the folds decreases using our Hybridfaces method. One can also see the dramatic improvement of the results if an alignment step is applied before the recognition. Especially for the Gorilla dataset the results improve significantly when congealing or a projective transform is applied. The results increase from 27.00% accuracy with no alignment to 72.83% and 73.83% for congealing and a projective transform, respectively. This manner holds true for the other datasets as well, even though the improvement is not as high as for the gorilla dataset.

6. CONCLUSION AND FUTURE WORK

In this paper we extended our previous work from [3] to a more automatic approach using a face alignment method called congealing instead of applying a projective transform using manually annotated facial feature points. We also presented an improvement of the Randomfaces approach [4] applying this approach not only to the whole facial image but additionally to regions of the apes face which, in our assumption, are very discriminative between individuals. We evaluated our approach in combination with two alignment methods on three publicly available datasets of chimpanzee and gorilla individuals. The presented Hybridfaces approach outperformed the standard Randomfaces approach on all three datasets if the facial images are sufficiently aligned. Even though the identification results when congealing is used as alignment technique are not as high as when a projective transform is applied, our experiments indicate that congealing is a powerful unsupervised alignment algorithm. Therefore, it can be used as a pre-processing step for automatic identification great apes in the wild as well as in captive environments. In future works we want to implement and evaluate a full automatic face recognition framework for primates including face detection, alignment and face recognition.

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 Ivorian authorities for long term support, especially the Ivorian Ministére de l'Environnement, des Eaux et Forêts and the Ministére de l'Enseignement Supérieur et de la Recherche Scientifique, the directorship of the Taï National Park, the OIPR and the CSRS in Abidjan. Financial support is gratefully acknowledged from the Swiss Science Foundation. We would like to thank especially Dr. Tobias Deschner for collecting videos and pictures over the last years and for providing invaluable assistance during the data collection. We thank all the numerous field assistants and students for their work on the Taï Chimpanzee Project. 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. We also thank Laura Aporius for providing videos and pictures in 2010. We thank Dr. Damien Caillaud for providing gorilla pictures and data from November 2003 until December 2004. His work was supported by the Station Biologique de Paimpont, the ECOFAC program, the National Geographic Society. We thank all the field assistants from Mbomo for their work on the project. This work was supported by the Max Planck Society. We also thank Laura Aporius for the annotation of data.

7. REFERENCES

[1] T. Burghardt and N. Campell, "Individual Animal Identification using Visual Biometrics on Deformable Coat-Patterns," in *5th International Conference on Computer Vision Systems (ICVS)*, 2007.

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

[3] A. Loos, M. Pfitzer, and L. Aporius, "Identification of great apes using face recognition," in *19th European Signal Processing Conference (EUSIPCO)*, 2011.

[4] 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.

[5] B. G. Huang and V. Jain E. Learned-Miller, "Unsupervised joint alignment of complex images," in *International Conference on Computer Vision (ICCV)*, 2007.

[6] 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.

[7] 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.

[8] 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.

[9] Yu Su, Shiguang Shan, Xilin Chen, and Wen Gao, "Hierarchical Ensemble of Global and Local Classifiers for Face Recognition," *IEEE Transactions on Image Processing*, vol. 18, no. 8, pp. 1885–1896, 2009.