Master in Artificial Intelligence (UPC-UB-URV)

Master of Science Thesis

Automatic Real and Apparent Age Estimation in Still Images

Pablo Pardo García

Supervisor:

Sergio Escalera Guerrero

Dept. Matemàtica Aplicada i Anàlisi, University of Barcelona (UB)

Co-supervisors:

Marc Oliu Simón

Dept. Matemàtica Aplicada i Anàlisi, University of Barcelona (UB)

Isabelle Guyon

President at ChaLearn and Clopinet, Berkeley, California

Barcelona, 20th April 2015

Acknowledgements

I would like to thank my supervisor Sergio Escalera for supporting and believing in me. This work would have been much harder without the incredible help of Marc Oliu and Xavi Baró who guided me though the deep learning and web developer world.

I also want to thank California Naturel, Visada and the rest of companies that sponsor the international challenge we are organizing, including Google and Microsoft Research among others. Also thank AgeGuess research group for their collaboration with our project.

And last but not least, this work would not have been the same without my friends, Jeroni, Denis, Lorenzo, Iosu, Anna, Hadi and all the master classmates. They are a source of energy and inspiration I needed. Hope this will hold in our new and soon coming adventures.

Abstract

Age Estimation tackles the problem of automatically predicting the age of people from visual data (still images, video data, depth maps, etc.). This is a very challenging problem because ageing is affected by many factors [88] such as bad habits, accidents, make up or facial hair.

This work focuses on three different age definitions, *real age* as the actual age (time elapsed since the person was born), *apparent age* as the perceived age from humans from the visual appearance and *estimated age* as the predicted age by a machine from the visual appearance.

In this work, a new age estimation face image database is presented containing for first time in the literature real and apparent age annotations. A study comparing the estimation of both type of ages was done by proposing two age estimation methods using the state of the art techniques, one based on Biologically Inspired Features (BIF) and the other based on Convolutional Neural Network (CNN).

The results of the two methods implemented in this master thesis show the superiority of the CNN over the BIF and also they show the difficulty of the age estimation problem with face images taken in an uncontrolled environment.

As a result of this work, an international challenge is being organized and the results will be presented in the ICCV conference 2015 (under revision).

Contents

List of Tables

List of Figures

1 Introduction

Computer Vision is a very important field in Artificial Intelligence that exists since the 1960s when digital image processing by computers became possible. Since the beginning of artificial vision, facial analysis has been a major interest in the research community (not just in Computer Vision but in other scientific areas such as biology [17], psychology [20], neuroscience [27] and sociology [59]) because its difficulty and its applications. Some of its applications include automatic detection of facial expressions [14], face detection [52], face recognition [94] [91] and automatic estimation of age [35], gender [3] and ethnicity [50].

Age estimation is a field within the facial analysis area in Computer Vision that tackles the problem of automatically predicting the age of people from visual data (still images, video data, depth maps, etc.). One of the main issues that the age estimation problem has is that there are many factors that influence human perception of age, some factors affect the aging of a person [88], such as smoking, drinking alcohol, doing sports, alimentation, etc. and others affect the face appealing such as scars, plastic surgery, make-up, facial hair, etc.

Some definitions should be established beforehand regarding the concept of human age:

- *Real age*: The actual age (time elapsed since the person was born).
- *Apparent age*: Perceived age from humans from the visual appearance.
- *Estimated age*: The predicted age by a machine from the visual appearance.

1.1 Goals

This work aims to study the differences (if any) of automatic age estimation from real age labels and apparent age labels. Given that does not exist any face image dataset with these two label annotations, a database with such requisites has been created. In order to do so, a web-based application was developed using the *Facebook API* to facilitate a collaborative and competitive collection of face images (`http://sunai.uoc.edu:8005/`).

As a consequence of this work, the first database in the literature containing real age and apparent age annotations for the face images was created and analysed. This database will allow researchers to tackle a different and new sub-problem of age estimation, *Apparent Age Estimation*. The most important methods in the state of the art are evaluated over our proposed database in this work.

In *Real Age Estimation* other external factors such as time evolution, habits and surgeries have to be taken into account. However *Apparent Age Estimation* is based purely in the perception field.

1.2 Age Challenge

Given the innovative aspect of the database, the HuPBA (Human Pose and Behaviour Analysis) research group[1] together with the non-profit organization ChaLearn[2] are going to prepare an international challenge competition later this year and present the results in a workshop organized by the team in the ICCV conference 2015 edition (under revision) within the ChaLearn Looking at People series [12] [23] [21] [22]. The challenge and the workshop will be sponsored by companies such as Google, Microsoft Research, Amazon and International Association for Pattern Recognition (IAPR) among others.

The challenge pretends to establish State of the Art techniques for Apparent Age Estimation and compare the methods and the results with the Real Age Estimation State of the Art.

Initial results regarding the database and the methods presented in this work have been published in the International Joint Conference on Neural Networks (IJCNN).

---

[1] Human Pose and Behaviour Analysis research group
`http://www.maia.ub.es/~sergio/soluciones2_008.htm`

[2] ChaLearn: Challenges in Machine Learning: `http://www.chalearn.org/`

2 State of the art

Age estimation has historically been one of the most challenging problems within the field of facial analysis [30][45]. Despite its multiple applications in many different areas there are relatively few publications compared to other topics in facial analysis. This difficulty is due to many factors:

- Depending on the application scenario, the age estimation problem can be taken as a multiclass classification problem or a regression problem.
- Large databases are difficult to collect, especially series of chronological image from the same individuals.
- The factors that affect the ageing process are uncontrollable and person specific [29][35][80].

The age estimation problem has generally two stages or blocks. The first one is the age representation in the images and the second one the learning algorithm. Several techniques have been published in order to deal with these two stages [30]. In the Table 2.1 there is a summary with some of the most important age estimation methods.

2.1 Age Representation

Age Representation, i.e. the extraction of the features that will represent the ageing patterns, is a very important step in the age estimation process. A good age representation will contain enough variation of the data to express the full complexity of the problem. There are many ways in the literature to represent ageing factors from an image. The most influential ones are described next.

2.1.1 Anthropomorphic Models

The first known work on age estimation from facial images was done by Y. Kwon and N. Lobo [61]. Their approach is based in cranio-facial development theory using geometrical ratios between different face regions to classify images into one of three age groups (babies, young adults and senior adults). They used frontal images in a very strict setup to be able to locate all face components. N. Ramanathan et al. [79, 81] used a similar approach in this case using 8 ratios rather than the 6 used by Y. Kwon et al.

The problem of this model is that it can only be applied to face images of young people in a growing age, since afterwards the facial geometry does not change as much. It is also a problem that the methods require frontal images, since it limits future applications.

2.1.2 Active Appearance Models

Active Appearance Models (AAM) is a statistical shape model proposed by T.F Cootes et al. [15]. This model contains the shape and grey-level appearance of the object of interest which can generalize to almost any valid example. This technique has been used to find the shape of faces by many researchers. A. Lanitis et al. [64] [65] [66] were the first to use the AAM model for age estimation by defining an ageing function $age = f(b)$, where b is a vector containing the parameters learned by the AAM.

A. Lanitis et al. [66] also tried different classifiers such as Quadratic Functions, Shortest Distance Classifier, Supervised Neural Network and Unsupervised Neural Network. Among all of them they reported that Quadratic Functions where the ones performing best.

Later on, K. Luu et al. [70, 71], use in [70] AAM with 68 facial landmarks to classify faces into young or adult classes and then use two specialised functions to finally determine the age. Luu's approach performs better than Lanitis' one [66]. In [71] they further improved the previous method by proposing Contourlet Appearance Models (CAM) which is more accurate and faster at calculating facial landmarks than AAM. This model has the ability of not only capturing global texture information like AAM but also local texture information using Nonsubsampled Contourlet Transform (NSCT) [16].

Chang et al. in [13] used AAM model with particular ranking formulation of support vectors, OHRank. The approach uses cost-sensitive aggregation to estimate ordinal hyperplanes (OH) and ranks them according to the relative order of ages.

This model captures shape and texture information and in general performs better than the *Anthropomorphic Models*. This method can deal with any range of ages rather than just with young ages like the previous model. However, as suggested by X. Geng et al. [34], the ageing function is empirically determined, so there is no evidence suggesting that the relation between face and age is described just by a quadratic function.

2.1.3 Ageing Pattern Subspace

X. Geng et al. [35, 34] were the first ones to explore AGing pattErn Subspace (AGES) model. They define an *ageing pattern* as a sequence of personal face images sorted in time order. Given a grey-scale face image I, where $I(x, y)$ determines the intensity of the pixel (x, y), then an ageing pattern can be represented as a three-dimensional matrix P, where $P(x, y, t)$ is the intensity of the pixel (x, y) in the face image at the time t. The images vector is filled with the available face images leaving empty the missing faces in the t axis. Now, the images in the age pattern vector can be precessed and transformed into meaningful feature vectors.

In order to extract the features X. Geng use AAM as used in [64] since they capture the shape and texture of the face images. By representing ageing patterns in this way, the concepts of identity and time are naturally integrated into the data without any pre-assumptions.

The principal drawback of the AGES method is that assumes that there are images of the same individual at different ages, which is not true in many age databases. The databases that fulfil these requirements are small such as FG-NET [62] with just 1000 images and very few representations of the same individual over time.

2.1.4 Age Manifold

Manifold learning methods are applied to find a sufficient embedding space and model the low-dimensional manifold data with a multiple linear regression function. Y. Fu et al. [28, 29] were the first to propose a manifold embedding approach for the age estimation problem.

The objective of this method is to find the low-dimensional representation in the embedded subspace capturing the intrinsic data distribution and geometric structure as well as its representation. G. Guo et al. [40] [41] shows that the Orthogonal Locality Preserving Projections (OLPP) [11] is a good an effective algorithm to connect the manifold learning with subspace learning. In a posterior work [38], G. Guo et al. introduces a new approach, using kernel partial least square (KPLS) regression which reduces feature dimensionality and learn the ageing function in a single step.

Then T. Wu et al. [95] proposed to model the facial shapes as points on a Grassmann manifold. Age estimation is then considered as regression and classification problems on this manifold. Then, they proposed a method for combining this shape-based approach with other texture-based algorithms.

The main drawback of the age manifold representation is the large number of training instances required to learn the embedded manifold with statistical sufficiency.

2.1.5 Appearance Models

Appearance Models focus on wrinkles, face texture and pattern analysis. From the beginning, researchers have tried to capture wrinkles and distinguish them from facial lines. Y. Kwon et al. [61] proposed a wrinkle detector based on snakelets [58] placed into key wrinkle areas of the face. Hayashi et al. [46] [47] [55] combined both shape and texture to estimate age and gender. In Hayashi's proposed approach, the skin is extracted based on a shape model and then a histogram equalization is applied to emphasize wrinkles.

Other researchers have used the texture descriptor Local Binary Patterns (LBP) [2] in the age estimation problem, such as [37] [100] obtaining good classification results with Nearest Neighbour and Support Vector Machines (SVM) classification algorithms. The Gabor [69] filter texture descriptor has also been used in the age estimation task [33], proving to be more discriminative than LBP.

G. Guo et al. [44] procposed to use Biologically Inspired Features (BIF) [84] for age estimation via faces. The BIF descriptor tries to mimic how the visual cortex works, with a hierarchy of increasingly sophisticated representations. The BIF original model proposed by Riesenhuber and Poggio [84] is based in a feed-forward model of the primate visual object recognition pathway, the "HMAX" model. The framework of the model contains alternative layers called simple (S) and complex (C) creating in each cycle a more elaborated representation. The S layers are created with a Gabor filtering on the input and the C layers generally operates a "MAX" operator over the previous S layer. G. Guo et al. [44] modifies tha BIF model by changing the operator in the complex layer (C1) from "MAX" to "STD".

Han et al. [45] used the BIF features in an hybrid classification framework improving the previous results with this descriptor. G. Guo et al. [39], in a recent paper (2014), also used the BIF features, and focus to investigate a proposed single-step framework for joint estimation of age, gender and ethnicity. Both the Canonical Correlation Analysis (CCA) [51] and Partial Least Square (PLS) based methods were explored under the joint estimation framework.

In [93], Weng et al. employs a similar ranking technique than Chang et al. in [13] called MFOR. LBP histogram features are combined with principal components of BIF, shape and textural features of AAM, and Principal Component Analysis (PCA) projection of the original image pixels. Fusion of texture and local appearance descriptors (LBP and HOG features) have independently also been used for age estimation by Huerta et al. in [54].

The outstanding results obtained by some of these works point out the suitability of BIF features for the age estimation via faces task.

2.1.6 Other Models

Depending on the available data the age estimation problem changes. Many researchers have tackled the age estimation problem with different types of data. A short overview is described below.

N. Ramanathan et al. [80] studied age progression of individual faces and proposed a method to perform face verification using a Bayesian age-difference classifier to improve the face verification algorithm.

A. Lanitis et al. [63] collected a database of head and mouse tracking movements and then attempted to perform age estimation with this data. This preliminary study shows the potential of using this type of data.

Y. Mikihara et al. [72] used a gait-based database to perform a viability study of age estimation using this type of data. The results show that in future research the combination of gait-based data and face-based age estimation could give very good results.

B. Xia et al. [96] were the first to attempt age estimation using 3D face images. The obtained results show that the depth dimension has a very discriminative power in this problem.

2.2 Age Estimation Learning Algorithm

Given an age representation, the next step is to determine the individual's age out of the ageing features. Age labels can be seen as a discrete set of classes or as a continuous label space, hence classification and regression learning methods can be used.

2.2.1 Classification Methods

The age estimation problem can be treated as a classification problem, where the solution space is discrete and the objective is to classify each face image into one of the age classes (a class could be an age range of several years or a single year).

A. Lanitis et al. [66] evaluated the performance of different classifiers such as quadratic function classifier, Artificial Neural Network (ANN) and k-Nearest Neighbours (kNN) classifier with their AAM model. Among the classification methods they tested, they claimed to perform better with the ANN, specifically the Multi Layer Perceptron (MLP), obtaining 4.78 Mean Absolute Error (MAE). The authors also proposed some extensions, for example, training age specific classifiers in a hierarchical fashion. With the extended methods the authors reduced the error to 4.38 MAE with the MLP and 3.82 MAE with the quadratic function (regression).

There have been previous proposals training neural networks, which are able to learn complex mappings and deal with outliers, for age estimation. In [66], Lanitis et al. used AAM encoded face parameters as an input for the supervised training of a neural network with a hidden layer. More recently, Geng et al. [36] tackled age estimation as a discrete classification problem using 70 classes, one for each age. The best algorithm proposed in this work (CPNN - Conditional Probability Neural Network) consists of a three-layered neural network, in which the input to the network includes both BIF features x and a numerical value for age y, and the output neuron is a single value of the conditional probability density function $p(y|x)$. An extensive comparison of these classification schemes for age estimation has been reported in Fernandez et al. [25]. In [98], Yang et al. used Convolutional Neural Network (CNN) for age estimation under surveillance scenarios and recently Dong Yi et al. [99] and Chenjing Yan et al. [97] have used CNN in the age estimation problem reporting very promising results in the MORPH dataset

(3.63 MAE).

K. Ueki [92] classified the images from the WIT_DB database into 11 age groups using Gaussian models in a low-dimensional 2DLDA+LDA feature space using the EM Algorithm. The accuracy rates they achieved were 46.3%, 67.8% and 78.1% for age groups that were in the 5-year, 10-year and 15-year range respectively.

SVM have been also used for age classification, Guo et al. [43] [40] trained an SVM for each pair of age classes and then using a binary tree search for testing, obtaining 5.55 MAE for females and 5.52 MAE for males in the YGA database and 7.16 MAE in the FG-NET database.

2.2.2 Regression Methods

The age of an individual is nothing else than the time passed from the individual's birth, and time is a continuous dimension. Hence, the age estimation problem can be formulated as a regression problem where the objective is to find a regression function that explains the ageing in terms of the feature space.

A. Lanitis et al. [66] evaluated three regression functions, linear, quadratic and cubic and claimed that the quadratic function was the one which better described the age from their feature space. Y. Fu et al. [28] [29] used linear, quadratic and cubic regression functions as learning algorithms for the manifold age representation. As Lenitis showed, Fu also reported superior performance on the quadratic regression function, pointing out that cubic functions lead to over-fitting while linear functions lead to under-fitting.

Guo et al. [43] [40] compared the performance of Support Vector Regressor (SVR) against the Local Adjusted Robust Regression (LARR) performance in the YGA and the FG-NET databases, concluding that LARR performs a more accurate estimation, achieving 5.25 MAE in YGA for female images, 5.30 in YGA for male images and 5.07 in FG-NET database. In a posterior work [44], the authors improve the SVR performance in the FG-NET to 4.77 MAE by using BIF features.

2.2.3 Hybrid Methods

Many authors have proposed mixture frameworks, using both classification and regression learning algorithms.

G. Guo et al. [42] proposed a probabilistic fusion approach. They use Bayes' rule to derive the predictor and then a sequential fusion strategy, so the output of the regressor is used as an intermediate decision which is then fed to the classifier to aid or affect the decision space of the classifier. Their fusion approach has better performance than other single step methods which they compare with.

SVM and SVR where used by Han et al. [45] in a hierarchical fashion. They proposed to use a binary decision tree with SVMs at each node to classify the images into different age ranges, which are coarsely assigned. Later, the age is fine grained by SVRs at the leaves. The SVRs are trained with 5-years-overlap between age ranges in order to reduce the misclassification error.

2.3 Applications

There are many real-world applications related to age estimation. Automatic age estimation is useful in situations where there is no need to specifically identify the individual, such as a government employee, but want to know his or her age.

2.3.1 Security Control and Surveillance Monitoring

In the last years security control and surveillance monitoring have gotten more relevant with the growth of internet content and the spread of technology that allows access to that content to under-age teenagers. Automatic age estimation systems can be used to prevent minors to buy alcohol in a grocery store, enter a bar or purchase tobacco from vending machines.

[1] Regularized Kernel Canonical Correlation Analysis
[2] Regularized Canonical Correlation Analysis

Publication	Year	Database (#subjects, #images)	Age Image Representation	Learning Method	MAE
A. Lanitis et al. [65]	2002	Private (60, 500)	AAM	Quadratic Aging Function	3.94 ± 3.8
A. Lanitis et al. [66]	2004	Private (40, 400)	AAM	Quadratic Aging Function	3.82 ± 5.58
X. Geng et al. [34]	2006	FG-NET	AGES	Regression	6.77
Y. Fu et al. [29]	2007	YGA	Manifold	Linear regression function	$5 \sim 6$
G. Guo et al. [44]	2009	FG-NET, YGA	BIF	SVR	FG-NET/YGA 4.77/F:3.91, M:3.47
K. Luu et al. [70]	2009	FG-NET	AAM	Hierarchical framework (SVM, SVR)	4.37
K. Luu et al. [71]	2011	FG-NET, PAL	Contourlet Appearance Model (CAM)	Hierarchical framework (SVM, SVR)	FG-NET/PAL 4.12 / 6.0
G. Guo et al. [38]	2011	MORPH-II	BIF	Kernel PLS Regression	4.18
H. Han et al. [45]	2013	FG-NET, MORPH-II, PCSO	BIF + Active Shape Models	Hierarchical framework (SVM, SVR)	FG-NET/PCSO 4.7 ± 24.8 / 7.2 ± 32.0
G. Guo et al. [39]	2014	MORPH-II	BIF	rKCCA[1] + Linear SVM	3.92
I. Huerta et al. [54]	2014	FG-NET, MORPH-II	BIF + HOG + LBP + SURF + Gradient	rCCA[2]	FG-NET/MORPH 4.17 / 4.25
Yi Dong et al. [99]	2014	MORPH-II	Local Aligned multi-scale patches	Multi-scale CNN	3.63

Table 2.1: Summary of some of the most important age estimation methods.

2.3.2 Biometrics

There are two types of biometric systems based on the number of traits used for recognition, unimodal biometric systems which consist on a single recognition trait and multimodal biometric systems, which combine evidences obtained from multiple sources [49] such as fingerprints, iris, face, etc. The multimodal system is more robust, more reliable and secure against spoofing attacks. However, the data acquisition is much more troublesome than the unimodal. In order to overcome these inconveniences, soft-biometrics [56], such as age, hight, weigh, gender, ethnicity and eye colour, are used in combination with classic biometric traits.

2.3.3 Age-based Indexing Face Databases

With the rise of interest for big data new and more efficient ways to retrieve data have to be developed. In large face image datasets, age can be used for indexing so that the queries to the dataset are simpler and faster. This is specially important in law enforcement where large image databases of suspects have to be filtered in order to find the most accurate suspects.

2.3.4 Human-Computer Interaction and e-Commerce

With the growth of e-commerce, companies want to offer a more personalized experience to their customers. Personalizing the offer or the product itself increases the user's satisfaction and the companies sells. Some examples of such policies are the following: Google [9] indexes the search results so the links that appear first appeal more to the user, Amazon [68] uses a recommender system to suggest products to the potential buyers according to their previous purchases, Netflix [60] held a competition in 2009 to create a film recommender system and gave a price of US $1,000,000. Age estimation systems could have an important role in the sector since age is a discriminative feature for different client profiles. Visada [57] is an example of the use of age estimation for product recommendation.

2.4 Age-based Datasets

There are many databases of faces in the literature, however, not so many capture the age of the individuals. This fact is due to the complexity of crawling such an information (if existent) from the usual fonts such as *Flickr* or *Facebook* and due to privacy issues. Moreover, the difficulty is even higher if the database contains chronological image series of individuals. Table 2.2 shows the most relevant databases used in the literature with number of samples, number of subjects, age range, type of age annotations and additional information if any. *FG-NET* [65] is one of the first and most consolidated age databases, it is used to compare with other age estimation methods.

After an initial interest in automatic age estimation from images dated back to the early 2000s [66] [65] [74], research in the field has experienced a renewed interest from 2006 on due to the availability of large databases like

Database	#Faces	#Subj.	Range	Type of age	Controlled Env.	Balanced age Distr.	Other annotation
FG-NET [65, 62]	1,002	82	0 - 69	Real Age	No	No	68 Facial Landmarks
GROUPS [32]	28,231	28,231	0 - 66+	Age group	No	No	-
PAL [74]	580	580	19 - 93	Age group	No	No	-
FRGC [77]	44,278	568	18 - 70	Real Age	Partially	No	-
MORPH2 [83]	55,134	13,618	16 - 77	Real Age	Yes	No	-
YGA [28]	8,000	1,600	0 - 93	Real Age	No	No	-
FERET[78]	14,126	1,199	-	Real Age	Partially	No	-
Iranian face [4]	3,600	616	2 - 85	Real Age	No	No	Kind of skin and cosmetic points [3]
PIE [89]	41,638	68	-	Real Age	Yes	No	-
WIT-BD [92]	26,222	5,500	3 - 85	Age group	No	No	-
Caucasian Face Database [10]	147	-	20 - 62	Real Age	Yes	No	Shape represented in 208 key points
LHI [1]	8,000	8,000	9 - 89	Real Age	Yes	Yes	-
HOIP [26]	306,600	300	15 - 64	Age Group	Yes	No	-
Ni's Web-Collected Database [76]	219,892	-	1 - 80	Real Age	No	No	-
OUI-Adience [19]	26.580	2.284	0 - 60+	Age Group	No	No	Gender

Table 2.2: Age-based Databases and its characteristics.

MORPH-Album 2 [83], which contain 55 times more age-annotated images than the *FG-NET* database.

[3] Surgical points, fracture or laceration on face.

3 Data Collection

As described in Section 2.4 there are many age-based databases of facial images. However, all existing datasets are based on real age estimation.

The idea of this work is to compare the performance between predicting real or apparent age labels. In order to do so, a web-based application has been developed using *Facebook's API* to collect a database with these annotations.

In order to increase the number of images in the database, an exhaustive research was done to find similar applications which were collecting the same, or very close, type data. The result of this search was AgeGuess[1] a web-based application developed by Dusan Misevic[2] and Ulrich Steiner[3], which collects nearly the same information as the collected for this work. The line of research of the AgeGuess team is focused on ageing and age perception from a biologically demographical point of view. The AgeGuess team agreed to partner with the HuPBA research group and the ChaLeran platform to joint effort in the data collection.

3.1 Web Application

The aim of the web-based application was to speed up the collection and labelling processes and reach more people with broader backgrounds to create an age database as diverse as possible. These processes were implemented in a gamified [4] fashion so the experience of the users with the application was satisfactory and engaging.

The application uses the API of Facebook to create a ranking with the user's Facebook friends and adding a factor of competitiveness to the game. It is also used to collect information about the labellers such as gender, age and nationality.

[1] AgeGuess: `http://www.ageguess.org/`

[2] Dusan Misevic is a researcher at Max-Planck Odense Center on the Biodemography of Aging, Biology department of the SDU

[3] Ulrich Steiner is a researcher at the Center for Research and Interdisciplinarity in Paris, of the INSERM Unit 1001

[4] Gamification is the use of game thinking and game mechanics in non-game contexts to engage users in solving problems and increase users' self contributions [18].

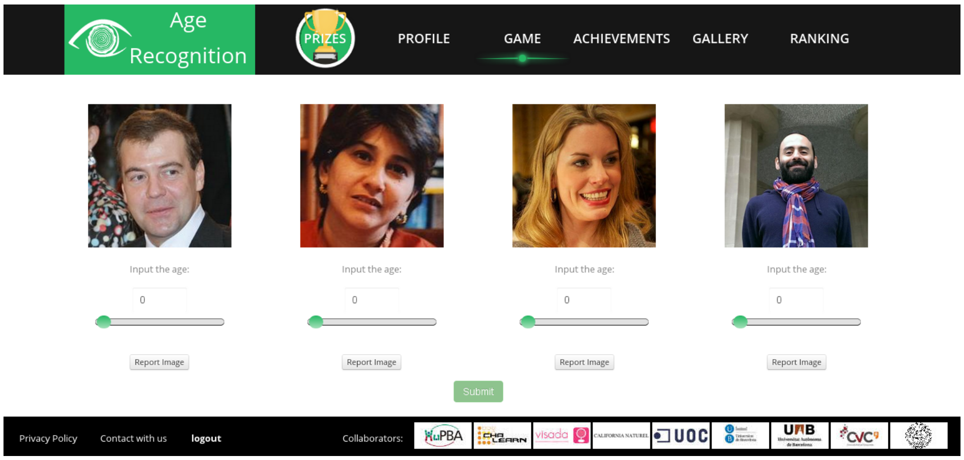


Figure 3.1: Age Recognition Application: Game Panel.

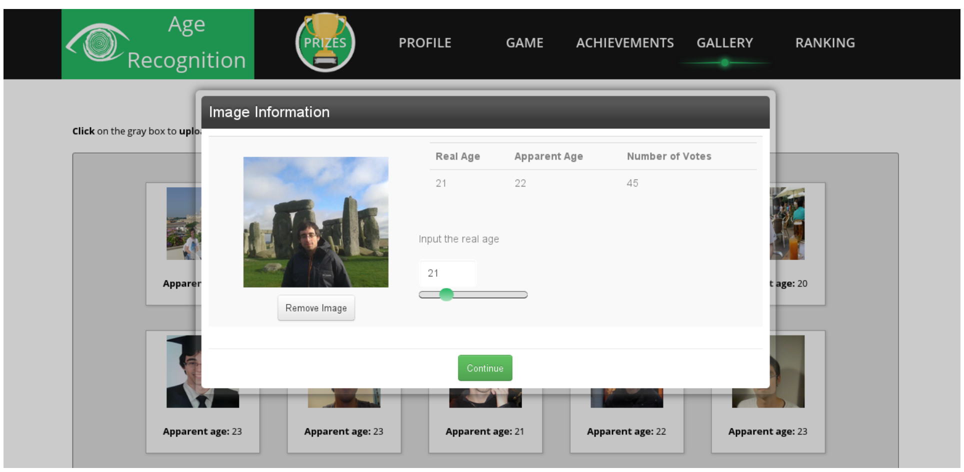


Figure 3.2: Age Recognition Application: Gallery Panel.

3.1.1 Gamification Strategy

The web-application is basically a platform to label and upload images, which is not funny or motivating. However, using gamification techniques the labeller engagement can be increased. The gamification strategies used were mainly four:

- The users or players get **points** for uploading and voting (labelling) images. The closer the vote is to the apparent age (average current

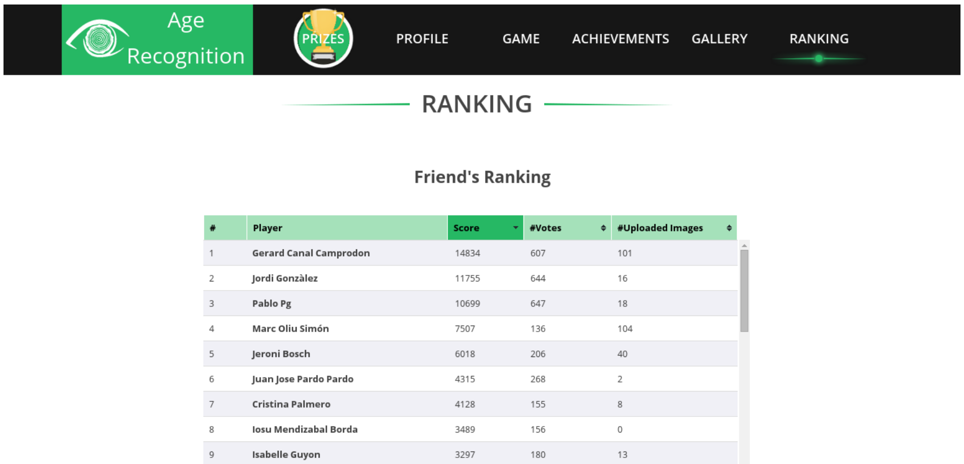


Figure 3.3: Age Recognition Application: Ranking Panel.

voted age) the more points the player gets. This strategy is pretended to persuade the users to wrongly label the images.

- Two **ranking** tables are shown to the users, one with the ranking positions of the users' Facebook friends and another showing the global classification. This strategy was created with the purpose of increase engagement between users, making them compete with each other.

- A system of **achievements** was implemented, including the four types of achievements shown in the Table 3.1.

Achievement	**Description**
Share it!	Invite your friends to play and you will get stars.
Precision	The better your guesses, the higher your rank.
Vote!	Vote on more images to get more stars.
Add Pictures!	Upload more images to get more stars.

Table 3.1: List of achievements with its description.

- Thanks to one of the Challenge sponsors, California Naturel [5], a bunch of cosmetic lots were offered as a **prize** for the ones who led the global ranking. The aim of the prize was to push participation further.

[5] `http://www.californianaturel.com/`

3.1.2 Application Structure

When users access the web for the first time they have to register using their Facebook account and accept the terms and conditions for the application usage (further explanation of the legal terms in the Appendix A). After registration the *How to play* panel will be displayed giving a short description of all the parts of the application.

As it is shown in Figures 3.1, 3.2 and 3.3, at the top of the site, there is a control menu that allows the users to move through the site with just one click. A description of the different panels is written below:

- **Profile**: In this section the users can keep track of their statistics (Number of uploaded images, number of votes, points obtained and global ranking position).
- **Game**: This is the main section (Figure 3.1), four images are shown at the same time and the users are asked to guess the age of all of them. They can report the images if any of them is considered offensive, bad quality, more than one person appears in the image, etc.
- **Achievements**: In this panel users can keep track of their achievements and shows the next goals.
- **Gallery**: In this section (Figure 3.2) players are able to see the images they have uploaded and see how many people have voted on them and get an estimation of their apparent age according to other users opinion. They are also able to upload new images in this panel, while uploading they are asked to crop (if necessary) and specify the real age of the person in the image.
- **Ranking**: The last panel (Figure 3.3) allows users to compare their scores with their Facebook friends in the *Friends Ranking* and to all the players in the *Global Ranking*.

3.1.3 Troubleshooting

When creating a public image database there are at least two things to keep in mind: image quality and image license.

To have good quality images the users are asked to upload images of a single person, preferable a portrait. In order to facilitate these requisites, the option of cropping the image is given in the uploading step. Also a minimum image size is enforced to avoid tiny images to be uploaded.

When users first register in the website they are asked to read and accept the terms and conditions which establish that the uploaded images rights must be hold by the user or the images have a license that allows the use of them and their redistribution free of charge. The users are also warned that the images will be used for research purposes and will be publicly available for this use.

3.2 HuPBA-AgeGuess Dataset

This section will describe and analyse the collected database called HuPBA-AgeGuess Dataset.

The Table 3.2 shows the main characteristics of the database. The web application gathered in 5 months more than 1500 images and almost 15000 votes, which means near 10 votes per image in average.

Features		HuPBA	AgeGuess	Total
Images		1506	33592	4865
Users	female	44	1828	1872
	male	110	1143	1253
Votes	female	1753	75136	76889
	male	14897	53117	68004

Table 3.2: HuPBA-AgeGuess Database Characteristics.

The ages distribution of the database (Figure 3.4) is not uniform as ideally should be, but instead contains more instances of people with ages between 20 and 30 than the rest of ages. There is also a very low incidence of individuals under 15 years old and over 70 years old. It can also be observed that both real and apparent age have a very similar distribution. This age distribution is justified by the fact that the people that most use social media are in between 18 and 49 years old [8].

Some of the properties of the database are listed below:

- Thousands of faces labelled by many users.
- Images with background.

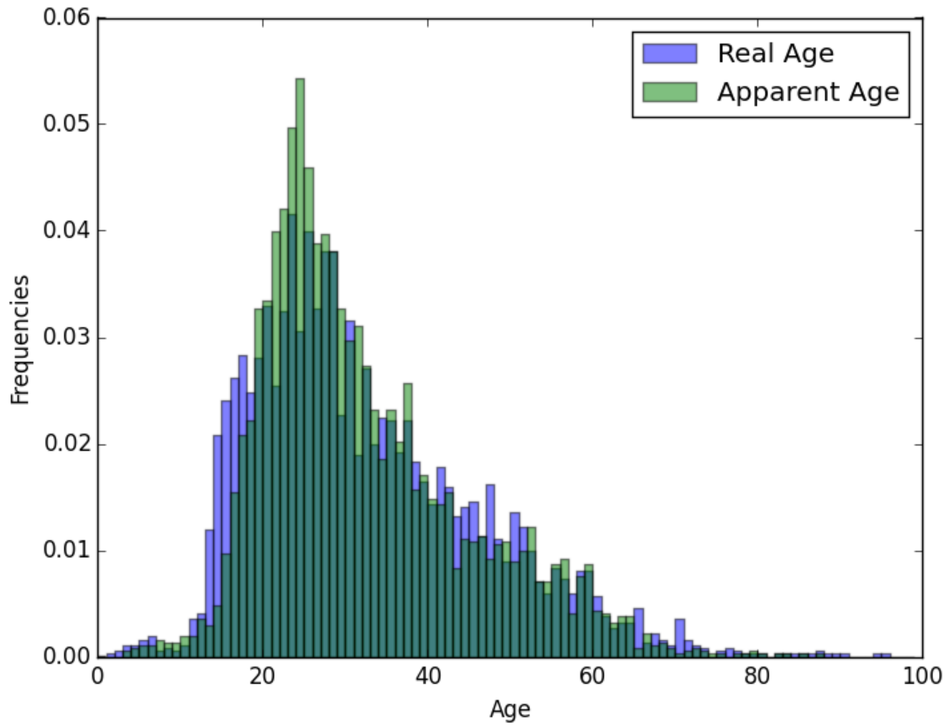


Figure 3.4: Labels Distribution.

- Non-controlled environments.
- Non-labelled faces neither landmarks, making the estimation problem even harder.
- The first datasets in the literature including estimated age labelled by many users to define the ground truth with the objective of estimating the age.
- The dataset also provides for each image the real age, although it is not used for recognition (just for analysis purposes).

3.2.1 Filtering Outliers

Outliers are a big problem in science because outliers show abnormality, fraudulent behaviour, human error, instrumental error, etc [48]. In this particular case that are many reasons to find outliers in the data since the users have many different backgrounds and the web-application was not developed by a professional. Some of the detected causes of outliers are listed below.

- To erroneously vote a different age than the one predicted.
- To deliberately vote a wrong or random ages.
- Server overload can slow down the loading process and make users vote wrongly.

This outliers are noise to the learning process so they can drop the performance of the learning algorithms. In order to disregard these outliers the votes are filtered with the following criterion,

$$f_V(v) = \begin{cases} 1, & \text{if } |v - \tilde{V_i}| \leq 10 \\ 0, & \text{otherwise} \end{cases}, \; v \in V \text{ and } i \in I \tag{3.1}$$

where V and I are the set of votes and images respectively and $V_i \subset V$ are the votes made to the image $i \in I$. Also note that $\tilde{V_i}$ is the median of the set of votes V_i. Basically this function filters out the votes which distance to the median is larger than 10 years.

The images are also filtered by the number of valid votes they have, setting the minimum accepted number of votes to 6. In this way the mean of all votes, i.e. the apparent age, is more reliable.

3.2.2 Further Analysis

The collected data contains much more information than just the real and apparent age. As explained in Section 3.1 the database contains information about the users or labellers. An analysis of human age perception from a sociological and biological point of view can be done. This work does not aim to attempt such a study, however a simple and superficial study has been done.

One of the first things to be analysed is the relation between the real age and the apparent age. Figure 3.5 shows the apparent age of all the images sorted by real age. The apparent age is fitted into a curve for a better comparison. The variance of the data is also plotted as a blue area. The two curves (real and apparent) show how younger people (0 - 25 years old) is overestimated while older people (50 - 100) is under estimated. The Figure 3.5 also shows that the ages where the apparent age has more variance are in

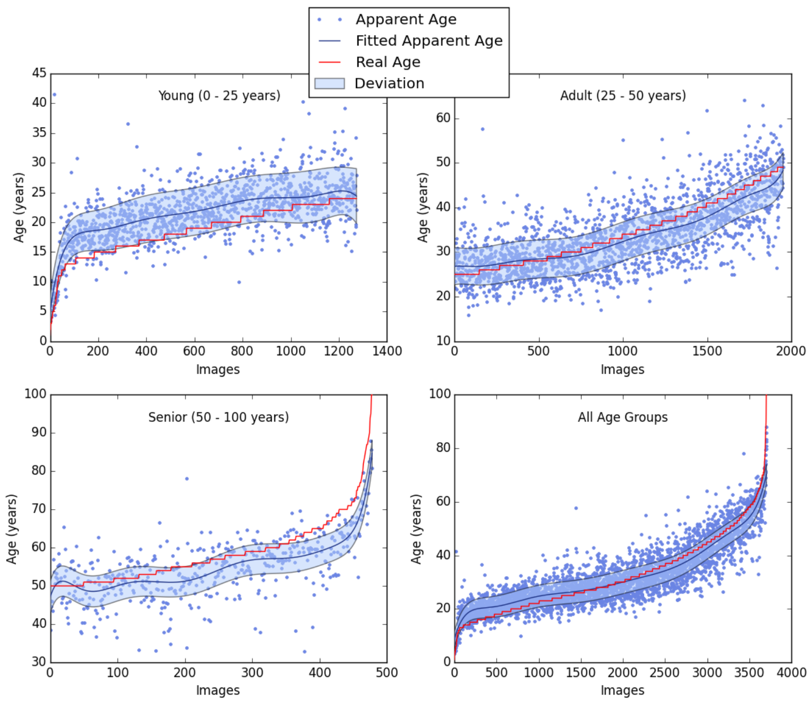


Figure 3.5: Apparent age compared to Real age in different age groups.

between 18 and 30 years old. The same data is illustrated in Figure 3.6, but distinguishing the male and female votes for calculating the apparent age. It is easily observed that there are no significant differences between male and female apparent ages.

Figure 3.7 culminates the study by showing all the votes between real and apparent age axis. The intensity of the colour shows the distribution of the votes in the space. The red dashed line shows the subspace where both real and apparent age match. This line also confirms that young people are overestimated and old people are underestimated.

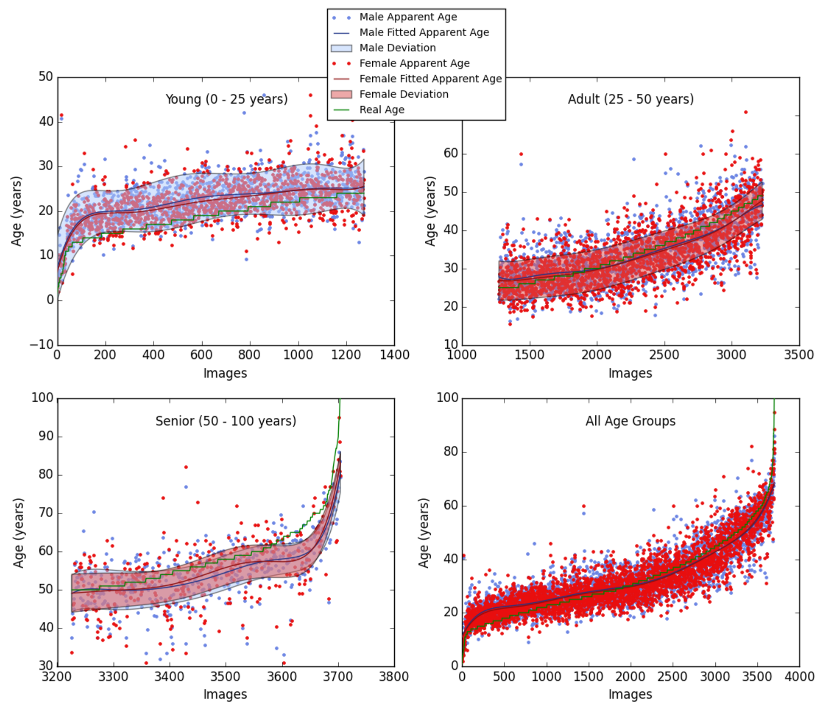


Figure 3.6: Comparison between Male and Female age perception.

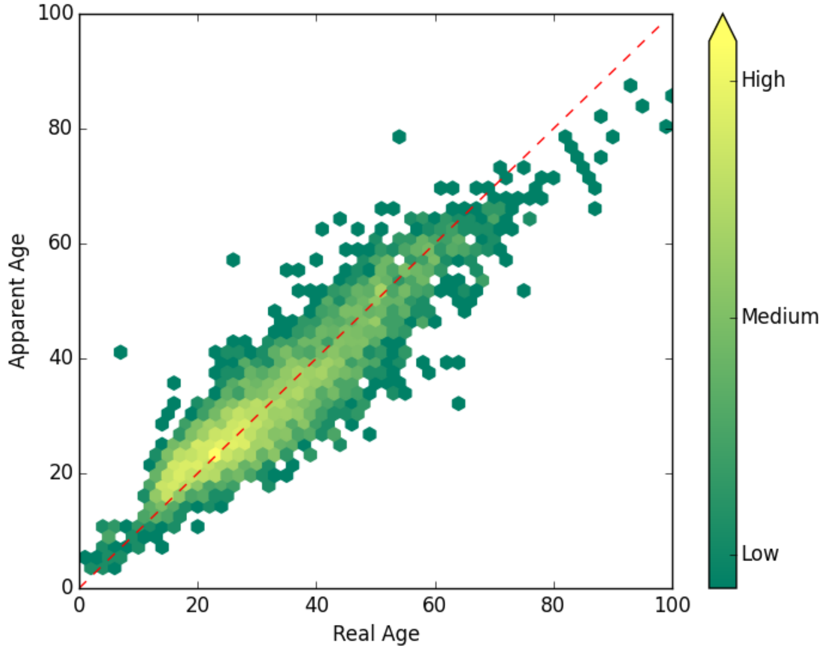


Figure 3.7: Real vs. Apparent age.

4 Method

In this work tow methods combining state of the art techniques have been proposed. A full description is delineated in this Chapter.

4.1 Preprocessing

Given that the background information is not needed in any of the two proposed methods, it is required to first detect the faces from the images and second align all of them (Figure 4.1).

- **Face detection** was performed using the Viola Jones algorithm from the open source library OpenCV [7]. An heuristic method was used to detect faces. Given that each images of the database contains a single face the detection algorithm should choose among all the detected faces. A Haar Cascade classifier was trained with nearly 5000 face images from different mugshot databases and more than 9500 negative samples. This classifier plus the standard frontal and lateral OpenCV cascade classifiers were used to detect faces in the images. Then an eye cascade classifier was used among the detected faces to drop out the ones where no eyes were detected. Finally, the biggest detected face is selected. After detecting the face, the image is cropped and resized into a 200×200.

- **Face alignment** or shape regression consists in, beginning with an initial shape S_0, S is refined by estimating the shape incremental ΔS. The general equation of the increment ΔS^t at time t is,

  $$\Delta S^t = W^t \Phi^t(I, S^{t-1}) \ , \tag{4.1}$$

  where I is the image, S^{t-1} is the estimated shape at time $t-1$, Φ^t is the feature mapping and W^t is the linear regression matrix. Shaoqing Ren et al. [82] contribution consist of calculating Φ^t and W^t in a consecutive step learning local feature mapping function to calculate local binare features for the shape landmarks and estimate W^t using linear regression. Their approach is fast and effective.

  In this work 68 facial landmarks were extracted for every detected face with the algorithm proposed by Shaoqing Ren et al. [82]. The algorithm was trained with almost 5000 face labelled images from FG-NET, AFW, HELEN, IBUG and LFPW face databases. After extracting all the landmarks all the faces are aligned.

All the images where transformed into **grey scale** because of the variation of the data in terms of colour (some images are in grey scale other in RGB).

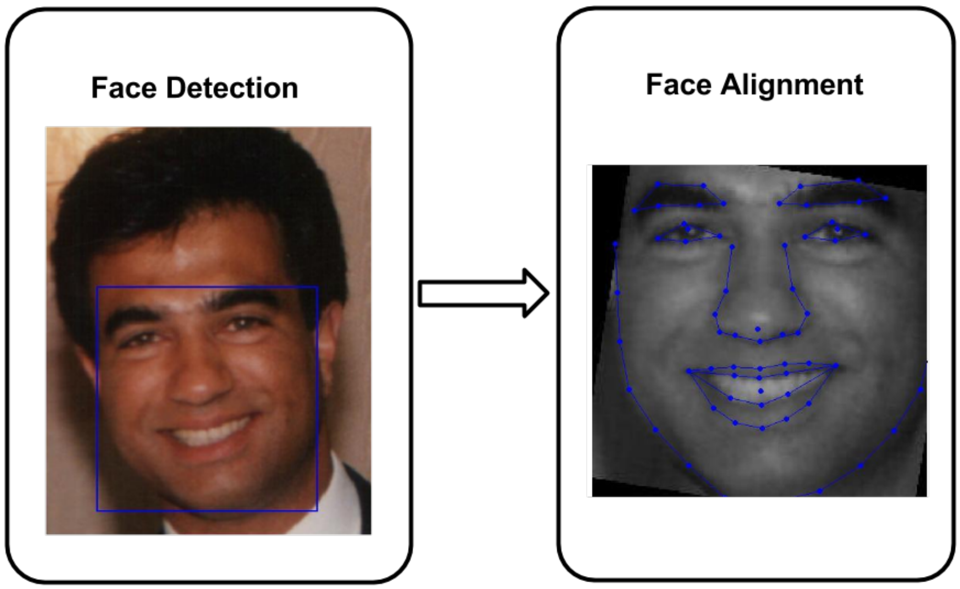


Figure 4.1: Face Detection and Face Alignment.

4.2 Biologically Inspired Method

As mentioned in Section 2.1.5 Biologically Inspired Features (BIF) have been proven to work specially good in age estimation [44][45]. In this work a method is proposed using BIF as age representation and a hierarchical framework with SVM and SVR as a learning method.

4.2.1 Background

Inspired by how the visual cortex works, Maximilian Riesenhuber and Tomaso Poggio [84] proposed the BIF model for object recognition called "HMAX" based on the hierarchical model of the visual nervous system proposed by D. H. Hubel and T. N. Wiesel [53].

The "HMAX" model is composed by layers that will contain increasingly sophisticated representations. There are two type of layers called simple (S_1) and complex (C_1). Layer S_1 consists a battery of Gabor filters with different orientation and scales and layer C_1 is a pooling layer, in this model is a "MAX" pooling layer. There are many variations and extensions of this model made by many authors, Serre et al. [87][86] introducing two more layers S_2 and C_2 or Mayers and Wolf [73] uses a spacial S_2 layer called S_2 facial features (S2FF) for face recognition.

C_1 **layer**			S_1 **layer**		
Scale Band S	Pool grid $(q \times q)$	Overlap Δn	Filter size n	Gabor σ	Gabor λ
Band 1	6×6	3	5×5 7×7	2.0 2.8	2.5 3.5
Band 2	8×8	4	9×9 11×11	3.6 4.5	4.6 5.6
Band 3	10×10	5	13×13 15×15	5.4 6.3	6.8 7.9
Band 4	12×12	6	17×17 19×19	7.3 8.2	9.1 10.3
Band 5	14×14	7	21×21 23×23	9.2 10.2	11.5 12.7
Band 6	16×16	8	25×25 27×27	11.3 12.3	14.1 15.4
Band 7	18×18	9	29×29 31×31	13.4 14.6	16.8 18.2
Band 8	20×20	10	33×33 35×35	15.8 17.0	19.7 21.2

Table 4.1: S_1 and C_1 parameters.

The proposed method in this work used the BIF model described by Guo et al. [44] that uses "STD" pool operator instead of "MAX" operator (Figure 4.2).

S_1 **layer** is formed by S_1 units which take as input a grey image of size 200×200. These units are usually modelled by Gabor filters, given (x, y)

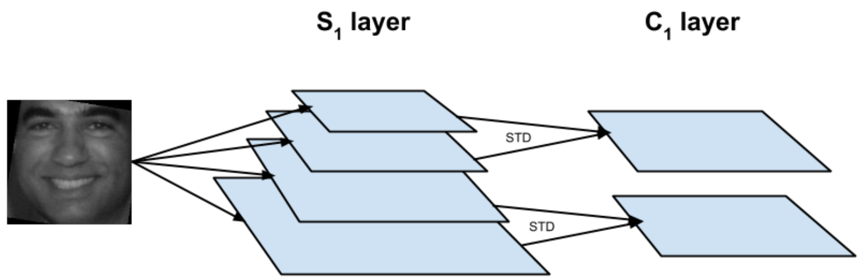


Figure 4.2: Biologically Inspired Features (BIF).

pixel coordinates of the input image,

$$G(x,y) = exp\left(-\frac{X^2+\gamma^2Y^2}{2\sigma^2}\right)\cdot cos(\frac{2\pi}{\lambda}X) \ , \tag{4.2}$$

where $X = x\cos\theta + y\sin\theta$ and $Y = -x\sin\theta + y\cos\theta$ are the rotations of the filter with angle $\theta \in [0,\pi]$. The aspect ration γ is set to 0.3, the width σ, the wavelength λ and the filter sizes n are adjusted as in Table 4.1. These parameters are empirically determined based on reactions of the visual cortex to real stimuli [87].

C_1 **layer** is formed by C_1 units which pool over the S_1 units with the same orientation and from the same scale band (see Table 4.1). Each scale band S contains two adjacent filter sizes, for instance, scale band 1 contains filter with sizes 5×5 and 7×7. The scale band also determines the sizes of the neighbourhood over which the C_1 units pool ($q\times q$). The pooling operator used in this method is the “STD” operator proposed by Guo et al. [44],

$$std_{j,j+1} = \sqrt{\frac{1}{q\times q}\sum_{i=1}^{q\times q}(F_i-\bar{F})^2} \ , \tag{4.3}$$

where F_i is the maximum value of two consecutive S_1 units output in the same scale band at pixel index i,

$$F_i = \max(x_i^j, x_i^{j+1}) \ , \tag{4.4}$$

where x_i^j and x_i^{j+1} are the filtered values with scales j and $j+1$ at position i. $\bar{F}$ is the mean of the filtered values in the neighbourhood $q \times q$.

The C_1 units are concatenated into a single feature vector for each input image.

4.2.2 System overview

This method combines a hierarchical framework and a hybrid face model, mixing BIF features and face shape landmarks to optimize the performance. The pipeline consists of three steps (see Figure 4.3).

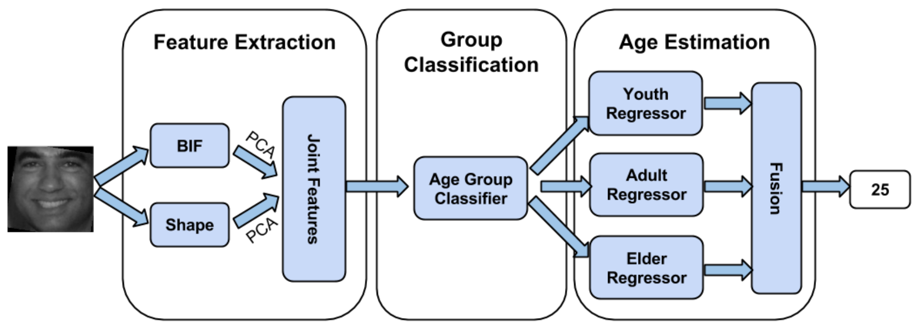


Figure 4.3: Biologically Inspired Method Pipeline.

Feature extraction: After the face alignment is done, BIF features are extracted from each face image as described above. Because of the high dimensionality of the data ($\sim$ 3000 BIF features and 138 shape landmark coordinates) feature reduction is needed. PCA is used to reduce dimensionality of both shape and BIF. As shows Figure 4.3 BIF and shape features are concatenated into a single feature vector.

Age Group Classification: In this stage the images are classified into three age groups with age ranges (0-18), (19-45) and (45-100). The classifier used is a linear SVM as suggested in [41].

Age Estimation: Three SVR with Radial Basis Function (RBF) kernel were trained to perform age regression in a specific age range. The data used to fit the regressors were images within 5-years-overlap between age ranges to reduce the misclassification error as proposed in [45].

In order to improve the system performance a last fusion step is done by averaging the two nearest age regressors, i.e. if the age group classifier predicts that a given face belongs to the adult age group (19-45)] and the adult-specific regression function predicts that the exact age is 43, then in the fusion step would average the adult regressor with the elder regressor.

4.3 Deep Learning Method

Deep Learning methods have not been fully explored in the age estimation problem. However, the rise of Deep Learning methods these last few years makes very likely that more works will base their approaches on this technique. In this work a method based on CNN is analysed.

4.3.1 Background

CNN were first proposed by Kunihiko Fukushima [31] in 1980 and then further improved by Y. LeCun et al. [67] in 1998. This particular type of Neural Network has been widely used in Computer Vision because of the way its ability to capture spacial relations. Some of the most successful applications were seen in the ImageNet Large Scale Visual Recognition Challenge [85] in 2014 where the winner [90] achieved object detection to 0.439329, and reduced classification error to 0.06656 in the ImageNet database which contains millions of images and thousands of classes. Also early this year S. Farfade et al. [24] proposed a method of multi-view face detection using CNN achieving a hight accuracy being able to detect faces from an up-side down view or partially occluded.

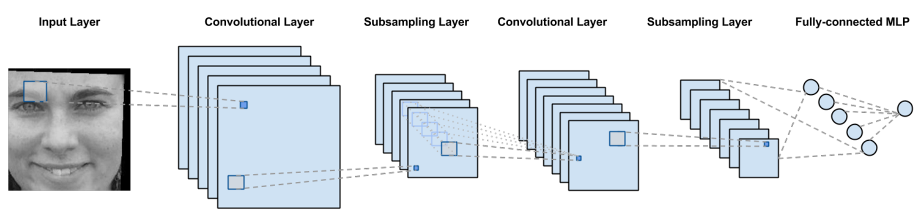


Figure 4.4: Example of Convolutional Neural Network.

CNN consists of a number of convolutional and subsampling layers optionally followed by fully connected layers (see Figure 4.4). The CNN input is an image of size $m \times m \times r$. The convolutional layer contains k kernels (filters) of size $n \times n \times q$ where $n \leq m$ and $q \leq r$. The filters will produce k feature maps of size $m - n - 1$. Each map is then subsampled typically with max or mean pulling over $p \times p$ contiguous region where p is small (typically between 2 and 5).

After the convolutional layers there may be any number of fully connected layers.

4.3.2 System overview

The CNN proposed in this work for age estimation is inspired on [99] [97]. The network was implemented using the Python library Theano [5][6] to optimize the training using the GPU. The proposed network topology is composed by three convolutional layers each of these followed by subsampling pooling layers and three fully connected layers at the end.

More in detail Table 4.2 shows the network topology layer by layer. The pooling layers use all the "MAX" operator and the fully-connected layers have ReLu [75] activation function. The proposed CNN is designed to do regression returning a single value for each input instance.

Layer	**Input size**	**Output size**	**Filter size**	**Pooling size**
Conv1	200×200	190×190	$10 * (11 \times 11)$	-
Pool1	190×190	95×95	-	$(2, 2)$
Conv2	95×95	89×89	$20 * (7 \times 7)$	-
Pool2	89×89	44×44	-	$(2, 2)$
Conv3	44×44	40×40	$40 * (5 \times 5)$	-
Pool3	40×40	20×20	-	$(2, 2)$
Full1	$16,000$	500	-	-
Full2	500	200	-	-
Full3	200	1	-	-

Table 4.2: Proposed CNN topology.

5 Results

This Chapter describes the experiments done in this work and the results achieved.

5.1 Datasets

Two datasets were used in the experiments, our collected dataset HuPBA-AgeGuess and the classic benchmark database FG-NET.

FG-NET consists of 1002 frontal face images of 82 different individuals. The image quality varies a lot in the dataset since there are images in grey scale and RGB. The face position is frontal and under similar illumination conditions. The dataset also contain 68 facial landmarks for each face image.

Figure 5.1: FG-NET image samples.

HuPBA-AgeGuess dataset (5.2) used in this project is a subset of the 4865 images filtered out by a minimum number of votes per image. This subset contains 3398 face images. The images are captured in the wild so the faces position vary up to $\pm 90^{o}$ and the illumination is also different in every picture.

Given the characteristics of the used descriptors in this work, both databases could be incremented by computing the mirror image, doubling the number of faces of each dataset (i.e. 2004 faces for the FG-NET and 6796 face images for the HuPBA-AgeGuess dataset).

Figure 5.2: HuPBA-AgeGuess image samples.

5.2 Evaluation Metrics

The two most commonly used evaluation metrics in age estimation are Mean Absolute Error (MAE) and Cumulative Score (CS).

The MAE is described as,

$$MAE = \frac{1}{n_s} \sum_{i=0}^{n_s - 1} |e_i|, \tag{5.1}$$

where n_s is the number of samples and e_i is the error of the ith instance, i.e. $e_i = |\hat{y}_i - y_i|$ where y_i is the real label and $\hat{y}_i$ is the predicted label. This metric tells the average number of years that the prediction is wrong.

CS is defined as the percentage of test images such that the absolute error is not higher than a threshold, t (in years). i.e.,

$$CS(t) = (1 - \frac{1}{n_s} \sum_{i=0}^{n_s - 1} h(|\hat{y}_i - y_i| - t)) \cdot 100 \tag{5.2}$$

$$h(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{otherwise,} \end{cases} \tag{5.3}$$

where y_i is the age label of the ith test image and $\hat{y}_i$ is the age prediction of the ith test image.

5.3 Experimental Settings

This section describes the experimental setup and the parameters used in the two proposed methods.

5.3.1 Biologically Inspired Method

As described in Section 4.2 a SVM classifier and three SVR regressors were trained for this method.

In order to find the best parameters a grid search was performed to find the best parameters. The parameters that formed the search space were the ones required by the SVM and SVRs and its kernels, which are the following:

- **Penalty term (C)**: This is the Support Vector parameter that deals with the cost of a misclassification over all the classification task.

  The C parameters tried in the search space were between 0.1 and 2. The best parameters in the Age Group Classification were between 1 and 2 and the best parameters for the SVRs were between 0.1 and 1.

- **Influence term (γ)**: It is the RBF kernel parameter. It determines how far the influence of a single training example reaches.

  The γ parameters used were between 0.001 and 1, being between 0.001 and 0.01 the ones with better performance.

In order to train and validate the parameters a nested 10-fold cross validation technique was used.

To evaluate the performance of the age group classifier it is needed to take into account that the labels are a disjoint segmentation of the spectrum of possible ages. Therefore, the in the performance evaluation the problem is treated as a regression problem rather than a classification, qualifying the predictions based on how far they are from the real label. The error measure chose is the MAE, normalizing the labels by the number of classes first.

5.3.2 Deep Learning Method

The FG-NET dataset has only 1002, clearly not enough to train the described deep CNN in Section 4.3. Therefore, this method has only been tested with the HuPBA-AgeGuess database.

Depending on the dataset the experimental setting were different. In order to validate and test the method a 10-fold cross validation was performed splitting the data into 80% training set, 10% validation set and 10% testing set. The data was split into batches each of them containing 213 instances. In this way each fold will contain an exact number of batches, 3. The error function used during the training was the mean square error. The network also needs the following parameters to be set,

- **Learning Rate (η)**: The learning rate is used in the backpropagation stage to modify the net weights.

  The learning rate used in the experiments was set to 0.002.

- **Epoch**: In training a neural network, the term epoch is used to describe a complete pass through all of the training patterns. The error is reduced over the epoch since the weights are updated within each epoch.

  The number of epochs used in the training of the proposed CNN were 400.

5.4 Analysis of the Experiments

In this section the results of all the experiments is described and analysed in detail. The experiments were run in the two datasets, FG-NET and HuPBA-AgeGuess, amd with the two different age labels (just in HuPBA-AgeGuess).

5.4.1 Face Detection and Alignment

This preprocessing step was just applied to the HuPBA-AgeGuess dataset since the FG-NET already contains the 68 facial landmarks manually placed. The database HuPBA-AgeGuess was preprocessed as described in Section 4.1. The face detection was 92.52% accurate and face alignment was 84.54% accurate, hence the total accuracy of the preprocessing was 78.26%. Some examples of good and bad face detections and alignments are shown in Figure 5.3. The detection and alignment were tricky since the images are not taken in a controlled environment, the multiple illumination changes and the multiple head orientations made the preprocessing very difficult.

(a) Well detected and aligned faces.

(b) Samples of bad face detection.

(c) Samples of bad face alignment.

Figure 5.3: Samples of preprocessed faces.

5.4.2 Biologically Inspired Method

The MAE achieved by this method in both databases is shown in the Table 5.1. As it was expected the algorithm performed better with the FG-NET dataset since the landmarks were manually placed. Between both experiments with the HuPBA-AgeGuess dataset this method performed best in the apparent age labels.

Database	MAE
FG-NET	7.99
HuPBA-AgeGuess real age	10.73
HuPBA-AgeGuess apparent age	9.34

Table 5.1: MAE achieved with BIF.

Figure 5.4 shows the CS obtained with this method. It shows that the performance is better over apparent age than real age.

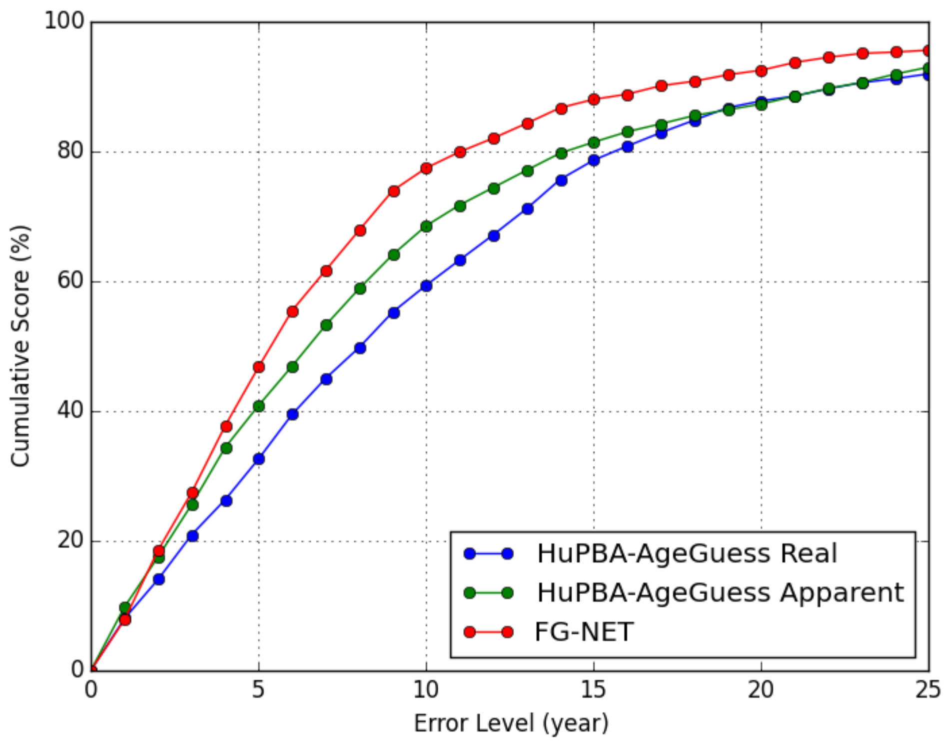
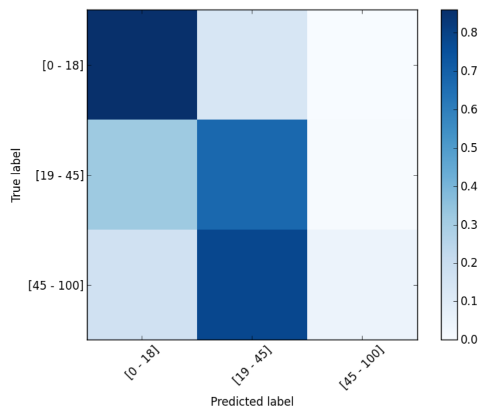
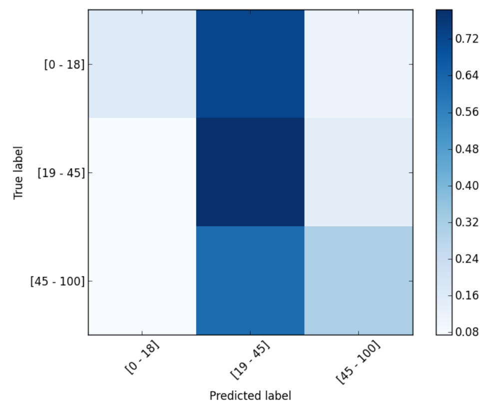


Figure 5.4: CS achieved with the Biologically Inspired Method.

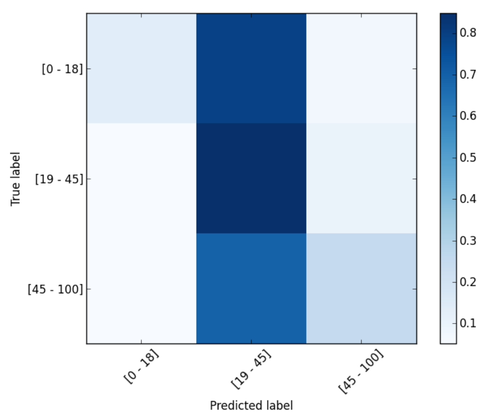
The group classification achieves 0.895 accuracy with FG-NET and 0.87 with HuPBA-AgeGuess dataset. Figure 5.5 shows the confusion matrix obtained in each one of the three datasets. In the FG-NET database there is confusion between the adult and the elder faces, and in the HuPBA-AgeGuess young and elder are confused with adults. This bad result is product of different factors, the high number of face instances with age between 18 and 45 years old bias the classifier towards the adult class.



(a) FG-NET dataset.



(b) HuPBA-AgeGuess with Real Age labels.



(c) HuPBA-AgeGuess with Apparent Age labels.

Figure 5.5: Confusion Matrices of the Age Group Classification.

The poor age group classification causes an overall bad performance of the age estimation algorithm with the HuPBA-AgeGuess dataset.

Study the CS and MAE by group of ages (maybe images in Appendix)

5.4.3 Deep Learning Method

The Table 5.2 shows the achieved MAE with this method with the HuPBA-AgeGuess database and both label annotations. The MAE of the method predicting real age is 40.77% higher than predicting apparent age.

Database	MAE
FG-NET	-
HuPBA-AgeGuess real age	10.29
HuPBA-AgeGuess apparent age	8.71

Table 5.2: MAE achieved with CNN.

Figure 5.6 shows the CS obtained with this method. It shows that the performance is better over apparent age than real age.

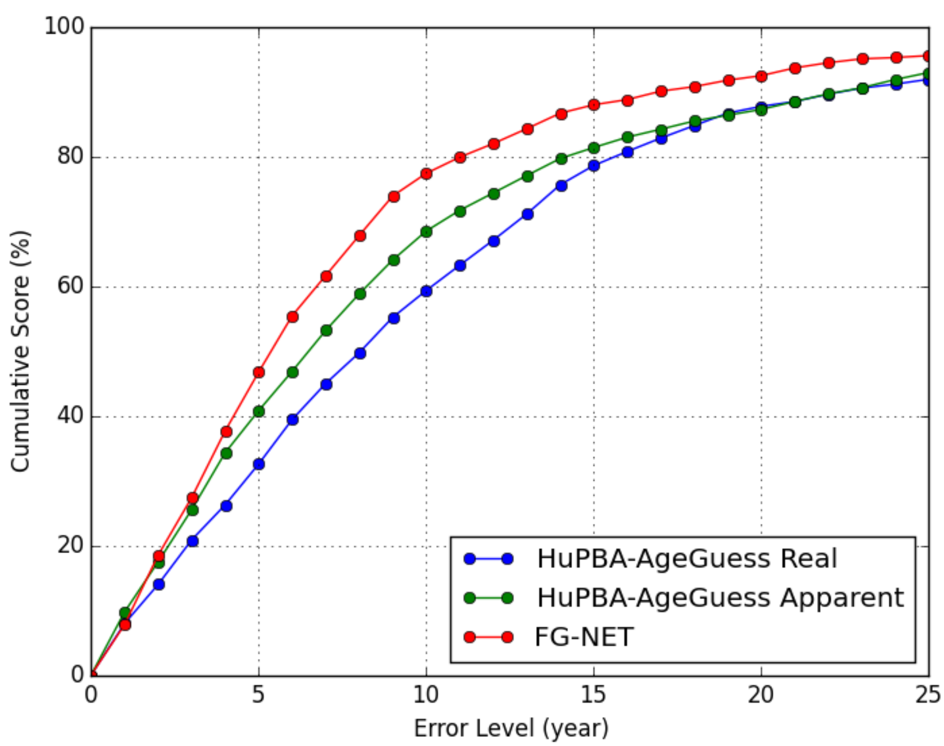


Figure 5.6: CS achieved with the CNN Method.

Show weight masks (maybe in the Apendix)

5.4.4 Comparison

From the two tables of results (Tables 5.1 and 5.2) can be seen that the CNN method achieves better results with the HuPBA-AgeGuess, reducing the MAE 0.65 years with apparent age and 0.44 years with real age annotations.

It also can be observed that the performance with apparent age is in all the cases higher than with real age.

Despite the fact that the FG-NET database could not be tested with the CNN method, the lowest MAE achieved in the experiments was using the

In both CNN and BIF methods the FG-NET dataset achieved better results. This is because the images in that dataset are frontal faces and the 68 landmarks were manually placed, while the HuPBA-AgeGuess database contains non frontal face images and the landmarks are automatically regressed.

6 Conclusions and Future work

6.1 Conclusions

todo

6.2 Future Work

As a future work, the three main parts of this project -proposed database, proposed methods and challenge organization- can be further explored.

The web-application could be improved by collecting further information about the database face images such as gender, ethnic, whether the individual is wearing make-up or no, etc. It should also be maintained in order to keep the users playing, uploading more images, guessing the already existent ones and engaging other users to play.

There are several aspect of the proposed methods that could be exhaustively explored.

- *Input Quality*: Determine until which extent the quality of the input images affects the performance of the state of the art method for age estimation.
- *Landmark Regression*: This step in the preprocessing is crucial since the features will be extracted locally in the aligned face. Therefore it is very important to improve the landmark regression to being able to localise the landmarks correctly despite the occlusions and facial expressions.
- *Face Frontalization*: The feature extraction could be improved by frontalizing the faces.
- *Local Multi-scale Patch BIF*: Calculate the BIF features from localized multi-scale patches using the landmark position could give less noisy and more reliable information about the analysed face.
- *Multi-modal Age Extimation*: Analyse the utility of depth and thermal images in a multi-modal fashion to determine how they complement discriminative information for age estimation.

After the organized challenge, the methods proposed by the contestants should be further analysed and see how they tackle the problems described in this work.

References

[1] LHI image database. URL http://www.lotushill.org/LHIFrameEn.html. Available at http://www.lotushill.org/LHIFrameEn.html, 2010.

[2] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(12):2037–2041, December 2006. ISSN 0162-8828. doi: 10.1109/TPAMI.2006.244. URL http://dx.doi.org/10.1109/TPAMI.2006.244.

[3] Luís A Alexandre. Gender recognition: A multiscale decision fusion approach. *Pattern Recognition Letters*, 31(11):1422–1427, 2010.

[4] A. Bastanfard, M.A. Nik, and M.M. Dehshibi. Iranian face database with age, pose and expression. In *Machine Vision, 2007. ICMV 2007. International Conference on*, pages 50–55, Dec 2007. doi: 10.1109/ICMV.2007.4469272.

[5] Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, James Bergstra, Ian J. Goodfellow, Arnaud Bergeron, Nicolas Bouchard, and Yoshua Bengio. Theano: new features and speed improvements. Deep Learning and Unsupervised Feature Learning NIPS 2012 Workshop, 2012.

[6] James Bergstra, Olivier Breuleux, Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, Guillaume Desjardins, Joseph Turian, David Warde-Farley, and Yoshua Bengio. Theano: a CPU and GPU math expression compiler. In *Proceedings of the Python for Scientific Computing Conference (SciPy)*, June 2010. Oral Presentation.

[7] G. Bradski. *Dr. Dobb's Journal of Software Tools.*

[8] Joanna Brenner and Aaron Smith. 72% of online adults are social networking site users. *Washington, DC: Pew Internet & American Life Project*, 2013.

[9] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.*, 30(1-7):107–117, April 1998. ISSN 0169-7552. doi: 10.1016/S0169-7552(98)00110-X. URL http://dx.doi.org/10.1016/S0169-7552(98)00110-X.

[10] D Michael Burt and David I Perrett. Perception of age in adult caucasian male faces: Computer graphic manipulation of shape and colour information. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 259(1355):137–143, 1995.

[11] Deng Cai, Xiaofei He, Jiawei Han, and Hong-Jiang Zhang. Orthogonal laplacianfaces for face recognition. *IEEE Transactions on Image Processing*, 15(11):3608–3614, 2006.

[12] ChaLearn. ChaLearn Looking at People. URL `http://gesture.chalearn.org/`. Available at `http://gesture.chalearn.org/`, 2015.

[13] Kuang-Yu Chang, Chu-Song Chen, and Yi-Ping Hung. Ordinal hyperplanes ranker with cost sensitivities for age estimation. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 585–592, June 2011. doi: 10.1109/CVPR.2011.5995437.

[14] Ira Cohen, Nicu Sebe, Ashutosh Garg, Lawrence S Chen, and Thomas S Huang. Facial expression recognition from video sequences: temporal and static modeling. *Computer Vision and image understanding*, 91(1):160–187, 2003.

[15] Timothy F. Cootes, Gareth J. Edwards, and Christopher J. Taylor. Active appearance models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(6):681–685, June 2001. ISSN 0162-8828. doi: 10.1109/34.927467. URL `http://dx.doi.org/10.1109/34.927467`.

[16] A.L. da Cunha, Jianping Zhou, and M.N. Do. The nonsubsampled contourlet transform: Theory, design, and applications. *Image Processing, IEEE Transactions on*, 15(10):3089–3101, Oct 2006. ISSN 1057-7149. doi: 10.1109/TIP.2006.877507.

[17] Charles Darwin. *The expression of the emotions in man and animals / by Charles Darwin.* New York ;D. Appleton and Co.,, 1916. URL `http://www.biodiversitylibrary.org/item/24064`. http://www.biodiversitylibrary.org/bibliography/4820.

[18] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. From game design elements to gamefulness: Defining "gamification". In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, MindTrek '11, pages 9–15, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0816-8. doi: 10.1145/2181037.2181040. URL `http://doi.acm.org/10.1145/2181037.2181040`.

[19] E. Eidinger, R. Enbar, and T. Hassner. Age and gender estimation of unfiltered faces. *Information Forensics and Security, IEEE Transactions on*, 9(12):2170–2179, Dec 2014. ISSN 1556-6013. doi: 10.1109/TIFS.2014.2359646.

[20] P. Ekman, W.V. Friesen, and J.C. Hager. *Facial Action Coding System (FACS): Manual*. A Human Face, Salt Lake City (USA), 2002.

[21] Sergio Escalera, Jordi Gonzàlez, Xavier Baró, Miguel Reyes, Oscar Lopes, Isabelle Guyon, Vassilis Athitsos, and Hugo Escalante. Multimodal gesture recognition challenge 2013: Dataset and results. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*, ICMI '13, pages 445–452, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2129-7. doi: 10.1145/2522848.2532595. URL `http://doi.acm.org/10.1145/2522848.2532595`.

[22] Sergio Escalera, Jordi Gonzàlez, Xavier Baró, Miguel Reyes, Isabelle Guyon, Vassilis Athitsos, Hugo Jair Escalante, Leonid Sigal, Antonis Argyros, Cristian Sminchisescu, Richard Bowden, and Stan Sclaroff. Chalearn multi-modal gesture recognition 2013: grand challenge and workshop summary. In Julien Epps, Fang Chen, Sharon Oviatt, Kenji Mase, Andrew Sears, Kristiina Jokinen, and Björn Schuller, editors, *ICMI*, pages 365–368. ACM, 2013. ISBN 978-1-4503-2129-7. URL `http://dblp.uni-trier.de/db/conf/icmi/icmi2013.html#EscaleraGBRGAESASBS13`.

[23] Sergio Escalera, Xavier Baro, Jordi Gonzalez, Miguel Angel Bautista, Meysam Madadi, Miguel Reyes, Victor Ponce, Hugo J. Escalante, Jaime Shotton, and Isabelle Guyon. Chalearn looking at people challenge 2014: Dataset and results. In *ChaLearn Looking at People Workshop*, 2014.

[24] Sachin Sudhakar Farfade, Mohammad Saberian, and Li-Jia Li. Multiview face detection using deep convolutional neural networks. *CoRR*, abs/1502.02766, 2015. URL `http://arxiv.org/abs/1502.02766`.

[25] Carles Fernández, Ivan Huerta, and Andrea Prati. A comparative evaluation of regression learning algorithms for facial age estimation. In *FFER in conjunction with ICPR, in press*. IEEE, 2014.

[26] Softopia Japan Foundation. Human and Object Interaction Processing (HOIP) Face Database. URL `http://www.hoip.jp/`. Available at `http://www.hoip.jp/`, 2014.

[27] Winrich A Freiwald, Doris Y Tsao, and Margaret S Livingstone. A face feature space in the macaque temporal lobe. *Nature neuroscience*, 12(9):1187–1196, 2009.

[28] Yun Fu and T.S. Huang. Human age estimation with regression on discriminative aging manifold. *Multimedia, IEEE Transactions on*, 10 (4):578–584, June 2008. ISSN 1520-9210. doi: 10.1109/TMM.2008.921847.

[29] Yun Fu, Ye Xu, and T.S. Huang. Estimating human age by manifold analysis of face pictures and regression on aging features. In *Multimedia and Expo, 2007 IEEE International Conference on*, pages 1383–1386, July 2007. doi: 10.1109/ICME.2007.4284917.

[30] Yun Fu, Guodong Guo, and T.S. Huang. Age synthesis and estimation via faces: A survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(11):1955–1976, Nov 2010. ISSN 0162-8828. doi: 10.1109/TPAMI.2010.36.

[31] Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36:193–202, 1980.

[32] A. Gallagher and T. Chen. Understanding images of groups of people. In *Proc. CVPR*, 2009.

[33] Feng Gao and Haizhou Ai. Face age classification on consumer images with gabor feature and fuzzy lda method. In *Proceedings of the Third International Conference on Advances in Biometrics*, ICB '09, pages 132–141, Berlin, Heidelberg, 2009. Springer-Verlag. ISBN 978-3-642-01792-6. doi: 10.1007/978-3-642-01793-3_14. URL `http://dx.doi.org/10.1007/978-3-642-01793-3_14`.

[34] Xin Geng, Zhi-Hua Zhou, Yu Zhang, Gang Li, and Honghua Dai. Learning from facial aging patterns for automatic age estimation. In *Proceedings of the 14th Annual ACM International Conference on Multimedia*, MULTIMEDIA '06, pages 307–316, New York, NY, USA, 2006. ACM. ISBN 1-59593-447-2. doi: 10.1145/1180639.1180711. URL `http://doi.acm.org/10.1145/1180639.1180711`.

[35] Xin Geng, Zhi-Hua Zhou, and K. Smith-Miles. Automatic age estimation based on facial aging patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(12):2234–2240, Dec 2007. ISSN 0162-8828. doi: 10.1109/TPAMI.2007.70733.

[36] Xin Geng, Chao Yin, and Zhi-Hua Zhou. Facial age estimation by learning from label distributions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(10):2401–2412, Oct 2013. ISSN 0162-8828. doi: 10.1109/TPAMI.2013.51.

[37] A. Gunay and V.V. Nabiyev. Automatic age classification with lbp. In *Computer and Information Sciences, 2008. ISCIS '08. 23rd International Symposium on*, pages 1–4, Oct 2008. doi: 10.1109/ISCIS.2008.4717926.

[38] Guodong Guo and Guowang Mu. Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 657–664, June 2011. doi: 10.1109/CVPR.2011.5995404.

[39] Guodong Guo and Guowang Mu. A framework for joint estimation of age, gender and ethnicity on a large database. *Image and Vision Computing*, 32(10):761 – 770, 2014. ISSN 0262-8856. doi: http://dx.doi.org/10.1016/j.imavis.2014.04.011. URL `http://www.sciencedirect.com/science/article/pii/S0262885614000791`. Best of Automatic Face and Gesture Recognition 2013.

[40] Guodong Guo, Yun Fu, C. R. Dyer, and T. S. Huang. Image-based human age estimation by manifold learning and locally adjusted robust regression. *Trans. Img. Proc.*, 17(7):1178–1188, July 2008. ISSN 1057-7149. doi: 10.1109/TIP.2008.924280. URL `http://dx.doi.org/10.1109/TIP.2008.924280`.

[41] Guodong Guo, Yun Fu, C.R. Dyer, and T.S. Huang. Image-based human age estimation by manifold learning and locally adjusted robust regression. *Image Processing, IEEE Transactions on*, 17(7):1178–1188, July 2008. ISSN 1057-7149. doi: 10.1109/TIP.2008.924280.

[42] Guodong Guo, Yun Fu, C.R. Dyer, and T.S. Huang. A probabilistic fusion approach to human age prediction. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on*, pages 1–6, June 2008. doi: 10.1109/CVPRW.2008.4563041.

[43] Guodong Guo, Yun Fu, T.S. Huang, and C.R. Dyer. Locally adjusted robust regression for human age estimation. In *Applications of Com-*

puter Vision, 2008. WACV 2008. IEEE Workshop on, pages 1–6, Jan 2008. doi: 10.1109/WACV.2008.4544009.

[44] Guodong Guo, Guowang Mu, Yun Fu, and Thomas S. Huang. Human age estimation using bio-inspired features. In *CVPR*, pages 112–119. IEEE, 2009. ISBN 978-1-4244-3992-8. URL `http://dblp.uni-trier.de/db/conf/cvpr/cvpr2009.html#GuoMFH09`.

[45] Hu Han, Charles Otto, and Anil K. Jain. Age estimation from face images: Human vs. machine performance. In *ICB'13*, pages 1–8, 2013.

[46] J. Hayashi, M. Yasumoto, H. Ito, and H. Koshimizu. Method for estimating and modeling age and gender using facial image processing. In *Virtual Systems and Multimedia, 2001. Proceedings. Seventh International Conference on*, pages 439–448, 2001. doi: 10.1109/VSMM.2001.969698.

[47] J. Hayashi, M. Yasumoto, H. Ito, Y. Niwa, and H. Koshimizu. Age and gender estimation from facial image processing. In *SICE 2002. Proceedings of the 41st SICE Annual Conference*, volume 1, pages 13–18 vol.1, Aug 2002. doi: 10.1109/SICE.2002.1195171.

[48] Victoria Hodge and Jim Austin. A survey of outlier detection methodologies. *Artif. Intell. Rev.*, 22(2):85–126, October 2004. ISSN 0269-2821. doi: 10.1023/B:AIRE.0000045502.10941.a9. URL `http://dx.doi.org/10.1023/B:AIRE.0000045502.10941.a9`.

[49] Lin Hong, Anil K. Jain, and Sharath Pankanti. Can multibiometrics improve performance. Technical Report MSU-CSE-99-39, Department of Computer Science, Michigan State University, East Lansing, Michigan, December 1999.

[50] Satoshi Hosoi, Erina Takikawa, and Masato Kawade. Ethnicity estimation with facial images. In *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pages 195–200. IEEE, 2004.

[51] Harold Hotelling. Relations between two sets of variates. *Biometrika*, pages 321–377, 1936.

[52] Rein-Lien Hsu, Mohamed Abdel-Mottaleb, and Anil K Jain. Face detection in color images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(5):696–706, 2002.

[53] D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology (London)*, 160:106–154, 1962.

[54] Ivan Huerta, Carles Fernández, and Andrea Prati. Facial age estimation through the fusion of texture and local appearance descriptor. In *Soft Biometrics in conjunction with ECCV, in press*. IEEE, 2014.

[55] Jun ichiro Hayashi, Mamoru Yasumoto, Hideaki Ito, and Hiroyasu Koshimizu. Age and gender estimation based on wrinkle texture and color of facial images. In *ICPR (1)*, pages 405–408, 2002. URL `http://dblp.uni-trier.de/db/conf/icpr/icpr2002-1.html#HayashiYIK02`.

[56] Anil K. Jain, Sarat C. Dass, and Karthik Nandakumar. Soft biometric traits for personal recognition systems. In David Zhang and Anil K. Jain, editors, *ICBA*, volume 3072 of *Lecture Notes in Computer Science*, pages 731–738. Springer, 2004. ISBN 3-540-22146-8. URL `http://dblp.uni-trier.de/db/conf/icba/icba2004.html#JainDN04`.

[57] Praveen Srinivasan JB Duler. Visada, Personalized Beauty Discovery from a Selfie. `http://visada.me/`, 2015. [Online; accessed 08-March-2015].

[58] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *INTERNATIONAL JOURNAL OF COMPUTER VISION*, 1(4):321–331, 1988.

[59] T.D. Kemper. *A Social Interactional Theory of Emotions*. A Wiley-interscience publication. Wiley, 1978. ISBN 9780471014058. URL `http://books.google.es/books?id=3499AAAAMAAJ`.

[60] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, August 2009. ISSN 0018-9162. doi: 10.1109/MC.2009.263. URL `http://dx.doi.org/10.1109/MC.2009.263`.

[61] Young H. Kwon and Niels da Vitoria Lobo. Age classification from facial images. *Comput. Vis. Image Underst.*, 74(1):1–21, April 1999. ISSN 1077-3142. doi: 10.1006/cviu.1997.0549. URL `http://dx.doi.org/10.1006/cviu.1997.0549`.

[62] A. Lanitis. FG-NET Aging Data Base, November 2007. URL `http://sting.cycollege.ac.cy/~alanitis/fgnetaging/index.htm`. Avail-

able at `http://www.eecs.qmul.ac.uk/~yf300/FG_NET_data/index.html`.

[63] A. Lanitis. Age estimation based on head movements: A feasibility study. In *Communications, Control and Signal Processing (ISCCSP), 2010 4th International Symposium on*, pages 1–6, March 2010. doi: 10.1109/ISCCSP.2010.5463396.

[64] A. Lanitis, C.J. Taylor, and T.F. Cootes. Modeling the process of ageing in face images. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 1, pages 131–136 vol.1, 1999. doi: 10.1109/ICCV.1999.791208.

[65] A. Lanitis, C.J. Taylor, and T.F. Cootes. Toward automatic simulation of aging effects on face images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(4):442–455, Apr 2002. ISSN 0162-8828. doi: 10.1109/34.993553.

[66] A. Lanitis, C. Draganova, and C. Christodoulou. Comparing different classifiers for automatic age estimation. *Trans. Sys. Man Cyber. Part B*, 34(1):621–628, February 2004. ISSN 1083-4419. doi: 10.1109/TSMCB.2003.817091. URL `http://dx.doi.org/10.1109/TSMCB.2003.817091`.

[67] Yann Lecun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. In *Proceedings of the IEEE*, pages 2278–2324, 1998.

[68] Greg Linden, Brent Smith, and Jeremy York. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1):76–80, January 2003. ISSN 1089-7801. doi: 10.1109/MIC.2003.1167344. URL `http://dx.doi.org/10.1109/MIC.2003.1167344`.

[69] Chengjun Liu and H. Wechsler. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *Trans. Img. Proc.*, 11(4):467–476, April 2002. ISSN 1057-7149. doi: 10.1109/TIP.2002.999679. URL `http://dx.doi.org/10.1109/TIP.2002.999679`.

[70] Khoa Luu, Karl Ricanek, Tien D. Bui, and Ching Y. Suen. Age estimation using active appearance models and support vector machine regression. In *Proceedings of the 3rd IEEE International Conference on*

Biometrics: Theory, Applications and Systems, BTAS'09, pages 314–318, Piscataway, NJ, USA, 2009. IEEE Press. ISBN 978-1-4244-5019-0. URL http://dl.acm.org/citation.cfm?id=1736406.1736456.

[71] Khoa Luu, Keshav Seshadri, Marios Savvides, Tien D. Bui, and Ching Y. Suen. Contourlet appearance model for facial age estimation. In Anil K. Jain, Arun Ross, Salil Prabhakar, and Jaihie Kim, editors, *IJCB*, pages 1–8. IEEE, 2011. ISBN 978-1-4673-0396-5. URL http://dblp.uni-trier.de/db/conf/icb/ijcb2011.html#LuuSSBS11.

[72] Y. Makihara, M. Okumura, H. Iwama, and Y. Yagi. Gait-based age estimation using a whole-generation gait database. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–6, Oct 2011. doi: 10.1109/IJCB.2011.6117531.

[73] Ethan Meyers and Lior Wolf. Using biologically inspired features for face processing. *Int. J. Comput. Vision*, 76(1):93–104, January 2008. ISSN 0920-5691. doi: 10.1007/s11263-007-0058-8. URL http://dx.doi.org/10.1007/s11263-007-0058-8.

[74] Meredith Minear and Denise C Park. A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, 36(4):630–633, 2004.

[75] Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. In Johannes Fürnkranz and Thorsten Joachims, editors, *ICML*, pages 807–814. Omnipress, 2010. ISBN 978-1-60558-907-7. URL http://dblp.uni-trier.de/db/conf/icml/icml2010.html#NairH10.

[76] Bingbing Ni, Zheng Song, and Shuicheng Yan. Web image mining towards universal age estimator. In *Proceedings of the 17th ACM International Conference on Multimedia*, MM '09, pages 85–94, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-608-3. doi: 10.1145/1631272.1631287. URL http://doi.acm.org/10.1145/1631272.1631287.

[77] P. Jonathon Phillips, Patrick J. Flynn, Todd Scruggs, Kevin W. Bowyer, Jin Chang, Kevin Hoffman, Joe Marques, Jaesik Min, and William Worek. Overview of the Face Recognition Grand Challenge. In *CVPR*, pages 947–954. IEEE, 2005.

[78] P.Jonathon Phillips, Harry Wechsler, Jeffery Huang, and Patrick J. Rauss. The {FERET} database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16(5):295 – 306,

1998. ISSN 0262-8856. doi: http://dx.doi.org/10.1016/S0262-8856(97)00070-X. URL `http://www.sciencedirect.com/science/article/pii/S026288569700070X`.

[79] N. Ramanathan and R. Chellappa. Modeling age progression in young faces. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 387–394, June 2006. doi: 10.1109/CVPR.2006.187.

[80] N. Ramanathan and R. Chellappa. Face verification across age progression. *Image Processing, IEEE Transactions on*, 15(11):3349–3361, Nov 2006. ISSN 1057-7149. doi: 10.1109/TIP.2006.881993.

[81] N. Ramanathan, R. Chellappa, and S. Biswas. Computational methods for modeling : A survey. *Journal of Visual Languages and Computing*, 20(3):131 – 144, 2009. ISSN 1045-926X. doi: http://dx.doi.org/10.1016/j.jvlc.2009.01.011. URL `http://www.sciencedirect.com/science/article/pii/S1045926X09000032`.

[82] Shaoqing Ren, Xudong Cao, Yichen Wei, and Jian Sun. Face alignment at 3000 fps via regressing local binary features. In *CVPR 2014*, 2014.

[83] K. Ricanek and T. Tesafaye. Morph: a longitudinal image database of normal adult age-progression. In *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, pages 341–345, April 2006. doi: 10.1109/FGR.2006.78.

[84] Maximilian Riesenhuber and Tomaso Poggio. Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2:1019–1025, 1999.

[85] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael S. Bernstein, Alexander C. Berg, and Li Fei-Fei. Imagenet large scale visual recognition challenge. *CoRR*, abs/1409.0575, 2014. URL `http://arxiv.org/abs/1409.0575`.

[86] T. Serre, L. Wolf, and T. Poggio. Object recognition with features inspired by visual cortex. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 994–1000 vol. 2, June 2005. doi: 10.1109/CVPR.2005.254.

[87] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio. Robust object recognition with cortex-like mechanisms. *Pattern Analysis and*

Machine Intelligence, IEEE Transactions on, 29(3):411–426, March 2007. ISSN 0162-8828. doi: 10.1109/TPAMI.2007.56.

[88] Roy J Shephard et al. *Aging, physical activity, and health.* Human Kinetics Publishers, 1997.

[89] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression (pie) database. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 46–51, May 2002. doi: 10.1109/AFGR.2002.1004130.

[90] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *CoRR*, abs/1409.4842, 2014. URL `http://arxiv.org/abs/1409.4842`.

[91] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 1701–1708. IEEE, 2014.

[92] Kazuya Ueki, Teruhide Hayashida, and Tetsunori Kobayashi. Subspace-based age-group classification using facial images under various lighting conditions. In *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, FGR '06, pages 43–48, Washington, DC, USA, 2006. IEEE Computer Society. ISBN 0-7695-2503-2. doi: 10.1109/FGR.2006.102. URL `http://dx.doi.org/10.1109/FGR.2006.102`.

[93] Renliang Weng, Jiwen Lu, Gao Yang, and Yap-Peng Tan. Multi-feature ordinal ranking for facial age estimation. In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*, pages 1–6, April 2013. doi: 10.1109/FG.2013.6553772.

[94] John Wright, Allen Y Yang, Arvind Ganesh, Shankar S Sastry, and Yi Ma. Robust face recognition via sparse representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(2):210–227, 2009.

[95] Tao Wu, Pavan K. Turaga, and Rama Chellappa. Age estimation and face verification across aging using landmarks. *IEEE Transactions on Information Forensics and Security*, 7(6):1780–1788, 2012. URL `http://dblp.uni-trier.de/db/journals/tifs/tifs7.html#WuTC12`.

[96] Baiqiang Xia, Boulbaba Ben Amor, Mohamed Daoudi, and Hassen Drira. Can 3D Shape of the Face Reveal your Age? In *International Conference on Computer Vision Theory and Applications*, Lisbonne, Portugal, January 2014. URL `https://hal.inria.fr/hal-00904007`.

[97] Chenjing Yan, Congyan Lang, Tao Wang, Xuetao Du, and Chen Zhang. Age estimation based on convolutional neural network. In WeiTsang Ooi, CeesG.M. Snoek, HungKhoon Tan, Chin-Kuan Ho, Benoit Huet, and Chong-Wah Ngo, editors, *Advances in Multimedia Information Processing – PCM 2014*, volume 8879 of *Lecture Notes in Computer Science*, pages 211–220. Springer International Publishing, 2014. ISBN 978-3-319-13167-2. doi: 10.1007/978-3-319-13168-9_22. URL `http://dx.doi.org/10.1007/978-3-319-13168-9_22`.

[98] Ming Yang, Shenghuo Zhu, Fengjun Lv, and Kai Yu. Correspondence driven adaptation for human profile recognition. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 505–512, June 2011. doi: 10.1109/CVPR.2011.5995481.

[99] Dong Yi, Zhen Lei, and Stan Z Li. Age estimation by multi-scale convolutional network. In *Computer Vision – ACCV 2014*. Springer International Publishing, 2014.

[100] J. Ylioinas, A. Hadid, and M. Pietikainen. Age classification in unconstrained conditions using lbp variants. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 1257–1260, Nov 2012.

Glossary

AAM Active Appearance Models. 4, 5, 7, 8, 49

AGES AGing pattErn Subspace. 5, 49

ANN Artificial Neural Network. 8, 49

BIF Biologically Inspired Features. 6–9, 49

CNN Convolutional Neural Network. 49

kNN k-Nearest Neighbours. 8, 49

LBP Local Binary Patterns. 6, 7, 49

MAE Mean Absolute Error. 8, 9, 49

MLP Multi Layer Perceptron. 8, 49

PCA Principal Component Analysis. 7, 49

SVM Support Vector Machines. 6, 9, 10, 49

SVR Support Vector Regressor. 9, 10, 49

Appendices

A Terms & Conditions

The Terms and Conditions of the web-application are an adaptation to the ones of ChaLearn challenges[1]. The general terms and conditions added to the ones described in ChaLearn are displayed below.

> *The purpose of this Facebook application is to gather images of anonymous people labeled with an approximation of the estimated age. ChaLearn intends to use the data collected to organize a challenge in computer vision in which the participants will need to write programs that estimate a person's age from a photo portrait.*
>
> *By using this application you agree to:*
>
> - *Upload only images of photo portraits that you own or are authorized to use.*
> - *Grant ChaLearn the right to use these images in a public scientific competition and to conduct research in computer vision.*
> - *Not to upload images that are offending or illegal.*

[1] ChaLearn privacy policy: `http://www.chalearn.org/privacy.html`