ChaLearn Looking at People 2015: Apparent Age and Cultural Event Recognition datasets and results

Sergio Escalera
Universitat de Barcelona and CVC

Junior Fabian
Computer Vision Center

Pablo Pardo
Universitat de Barcelona

Xavier Baró
Universitat Oberta de Catalunya
Computer Vision Center

Jordi Gonzàlez
Universitat Autònoma de Barcelona
Computer Vision Center

Hugo J. Escalante
INAOE Mexico

Dusan Misevic
Center for Research and Interdisciplinarity

Ulrich Steiner
Max-Planck Odense Center

Isabelle Guyon
Clopinet, ChaLearn

Abstract

Following previous series on Looking at People (LAP) competitions [14, 13, 11, 12, 2], in 2015 ChaLearn ran two new competitions within the field of Looking at People: (1) age estimation, and (2) cultural event recognition, both in still images. We developed a crowd-sourcing application to collect and label data about the apparent age of people (as opposed to the real age). In terms of cultural event recognition, one hundred categories had to be recognized. These tasks involved scene understanding and human body analysis. This paper summarizes both challenges and data, as well as the results achieved by the participants of the competition. Details of the ChaLearn LAP competitions can be found at http://gesture.chalearn.org/.

1. Introduction

The automatic analysis of the human body, also named Looking at People, in still images and image sequences keeps making rapid progress with and the constant improvement of new published methods that rapidly advance the state-of-the-art.

In 2015, ChaLearn organized new competitions and workshops on age estimation and cultural event recognition from still images. The recognition of continuous, natural human signals and activities is very challenging due to the multimodal nature of the visual cues (e.g., movements of fingers and lips, facial expressions, body pose), as well as technical limitations such as spatial and temporal resolution. Facial expressions analysis and age estimation are hot topics in Looking at People that serve as additional cues to determine human behavior and mood indicators. Finally, images of cultural events constitute a very challenging recognition problem due to a high variability of garments, objects, human poses and scene context. Therefore, how to combine and exploit all this knowledge from pixels constitutes an interesting problem.

These challenges motivated our choice to organize a new ICCV workshop and a competition on these topics to sustain the effort of the computer vision community. These new competitions come as a natural evolution from our previous workshops at CVPR2011, CVPR2012, ICPR2012, ICMI2013, ECCV2014 and CVPR2015. We continued using our website http://gesture.chalearn.org for promotion and challenge entries in the quantitative competition were scored on-line using the Codalab Microsoft-Stanford University platforms (http://codalab.org/).

In the rest of this paper, we describe in more detail both age estimation and cultural event recognition challenges, their relevance in the context of the state of the art, and describe the results achieved by the participants in both challenges.

2. Age estimation challenge

Age estimation is a difficult task which requires the automatic detection and interpretation of facial features. We have designed an application using the Facebook API for the collaborative harversting and labeling by the community in a gamified fashion (http://sunai.uoc.edu:8005/).

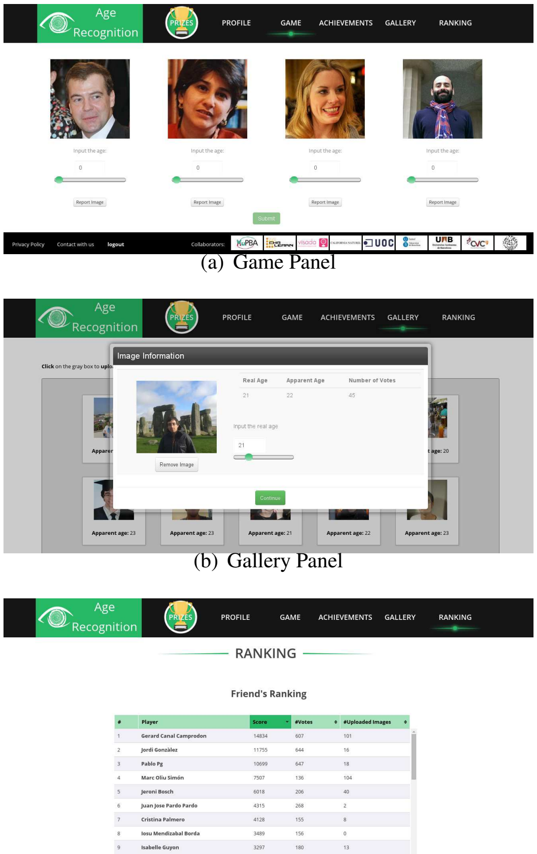
Age estimation has historically been one of the most challenging problems within the field of facial analysis [32, 17, 21]. It can be very useful for several applications, such as advanced video surveillance, demographic statistics collection, business intelligence and customer profiling, and search optimization in large databases. Different application scenarios can benefit from learning systems that predict the apparent age, such as medical diagnosis (premature aging due to environment, sickness, depression, stress, fatigue, etc.), effect of anti-aging treatment (hormone replacement therapy, topical treatments), or effect of cosmetics, haircuts, accessories and plastic surgery, just to mention a few. Some of the reasons age estimation is still a challenging problem are the uncontrollable nature of the aging process, the strong specificity to the personal traits of each individual, high variance of observations within the same age range, and the fact that it is very hard to gather complete and sufficient data to train accurate models.

2.1. Dataset

Due to the nature of the age estimation problem, there is a restricted number of publicly available databases providing a substantial number of face images labeled with accurate age information. Table 1 shows the summary of the existing databases with main reference, number of samples, number of subjects, age range, type of age and additional information. This field has experienced a renewed interest from 2006 on, since the availability of large databases like MORPH-Album 2 [34], which increased by 55× the amount of real age-annotated data with respect to traditional age databases. Therefore, this database has extensively been used in recent works by applying to it different descriptors and classification schemes. However, all existing datasets are based on real age estimation. In the present challenge, we propose the first dataset to recognize the apparent age of people based on the opinion of many subjects using a new crowd-sourcing data collection and labeling application and the data from AgeGuess platform[1].

We developed a web application in order to collect and label an age estimation dataset online by the community. The application uses the Facebook API to facilitate the access hence reach more people with a broader background. It also allows us to easily collect data from the participants, such as gender, nationality and age. We show some panels of the application in the Figure 1(a), 1(b) and 1(c).

The web application was developed in a gamified way, i.e. the users or players get points for uploading and labeling images, the closer the age guess was to the apparent age (average labeled age) the more points the player obtained. In order to increase the engagement of the players, we added



(a) Game Panel

Image Information

Real Age	Apparent Age	Number of Votes
21	22	45

(b) Gallery Panel

Friend's Ranking

#	Player	Score	#Votes	#Uploaded Images
1	Gerard Canal Camprodon	14834	607	101
2	Jordi Gonzàlez	11755	644	16
3	Pablo Pg	10699	647	18
4	Marc Oliu Simón	7507	136	104
5	Jeroni Bosch	6018	206	40
6	Juan Jose Pardo Pardo	4315	268	2
7	Cristina Palmero	4128	155	8
8	Iosu Mendizabal Borda	3489	156	0
9	Isabelle Guyon	3297	180	13

(c) Ranking Panel

Figure 1. Age Recognition Application. (a) User can see the images of the rest of participants and vote for the apparent age. (b) User can upload images and see their uploads and the opinion of the users regarding the apparent age of people in their photos. (c) User can see the points he/she achieves by uploading and voting photos and the ranking among his/her friends and all the participants of the application.

a global and friends leaderboard where the users could see their position in the ranking. We asked the users to upload images of a single person and we gave them tools to crop the image if necessary, we also asked them to provide the real age for images they uploaded themselves (or a good approximation), allowing more analysis and comparisons between real age and apparent age.

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 of data. The result of this search was AgeGuess, which collects nearly the same information as we do. The line of research of the AgeGuess team is focused on aging 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. The Table 2 shows the main characteristics

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

Table 1. Age-based Databases and its characteristics.

Database	#Faces	#Subj.	Range	Type of age	Controlled Enviroment	Balanced age Distrivution	Other annotation
FG-NET [25, 24]	1,002	82	0 - 69	Real Age	No	No	68 Facial Landmarks
GROUPS [19]	28,231	28,231	0 - 66+	Age group	No	No	-
PAL [27]	580	580	19 - 93	Age group	No	No	-
FRGC [31]	44,278	568	18 - 70	Real Age	Partially	No	-
MORPH2 [35]	55,134	13,618	16 - 77	Real Age	Yes	No	-
YGA [18]	8,000	1,600	0 - 93	Real Age	No	No	-
FERET[30]	14,126	1,199	-	Real Age	Partially	No	-
Iranian face [3]	3,600	616	2 - 85	Real Age	No	No	Kind of skin and cosmetic points
PIE [37]	41,638	68	-	Real Age	Yes	No	-
WIT-BD [42]	26,222	5,500	3 - 85	Age group	No	No	-
Caucasian Face Database [5]	147	-	20 - 62	Real Age	Yes	No	208 Shape Landmarks
LHI [1]	8,000	8,000	9 - 89	Real Age	Yes	Yes	-
HOIP [16]	306,600	300	15 - 64	Age Group	Yes	No	-
Ni's Web-Collected Database [28]	219,892	-	1 - 80	Real Age	No	No	-
OUI-Adience [9]	26.580	2.284	0 - 60+	Age Group	No	No	Gender

of the final database.

Table 2. ChaLearn-AgeGuess database characteristics.

Features		ChaLearn	AgeGuess	Total
Images		1506	3185	4691
Users	female	44	1828	1872
	male	110	1143	1253
Votes	female	1753	75136	76889
	male	14897	53117	68004

Some of the properties of the database and its associated challenge are: Thousands of faces labeled by many users; images with background; non-controlled environments; and non-labeled faces without landmarks, making the estimation problem even harder. This is one of the first datasets in the literature including estimated age labeled by many users to define the ground truth with the objective of estimating the age. The evaluation metric will be weighted by the mean and the variance of the labeling by the participants.

The dataset contains the real age for each image, although this is not used for recognition but only for data analysis purposes. In the same way for all the labelers (the users of the platforms which make estimates of the age of the person in the photo). We have their nationality, age, and gender, which will allow analyzing demographic and other interesting studies among the correlation of labelers.

In relation to the properties of existing datasets shown in Table 1, ours include labels of the real age of the individuals and the apparent age given by the collected votes, both age distributions are shown in Figure 2. The images of our database have been taken under very different conditions, which makes it more challenging for recognition purposes. Different application scenarios can benefit from learning systems that predict the apparent age, such as medical diagnosis (premature aging due to environment, sickness, depression, stress, fatigue, etc.), effect of anti-aging treatment (hormone replacement therapy, topical treatments), or effect of cosmetics, haircuts, accessories and plastic surgery, just to mention a few.

2.2. Apparent age challenge results

More than 100 participants registered into the competition. Finally, at the test step, the teams that submitted their predictions are shown in Table 3. Each prediction is evaluated as $\epsilon = 1 - e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, where x is the prediction, μ and σ are the mean and standard deviation of the human labels. The summary of the methods is shown in Table 4. The summary of the first top ranked methods is shown next.

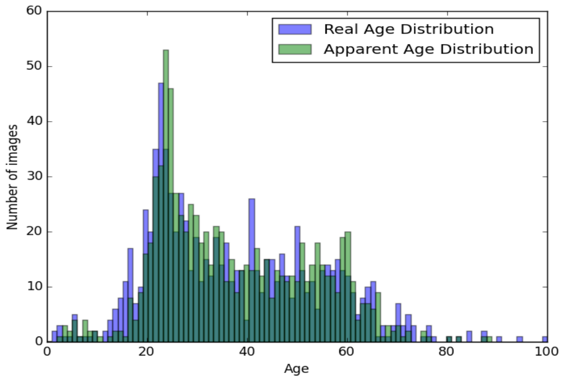


Figure 2. Real and Apparent age distributions in our database.

Table 3. Results of the Age Estimation challenge.

Age Estimation			
Position	**Team**	**Development**	**Test**
1	**CVL_ETHZ**	**0.295116**	**0.264975**
2	**ICT-VIPL**	**0.292297**	**0.270685**
3	AgeSeer	0.327321	0.287266
3	**WVU_CVL**	**0.316289**	**0.294835**
4	SEU-NJU	0.380615	0.305763
5	UMD	-	0.373352
6	Enjuto	0.370656	0.37439
7	Sungbin Choi	-	0.420554
8	Lab219A	0.477079	0.499181
9	Bogazici	0.483337	0.524055
10	Notts CVLab	-	0.594248

Table 4. Table of Age Methods.

Team	**Proposed Method**
CVL_ETHZ	Face detection using [26]. 20 CNN models [38] on cropped faces. External data: ImageNet, IMDB and Wikipedia. System prediction: Expected value of 101 softmax-normalized output neurons.
ICT-VIPL	Face detection using Boosting+Neural Networks and Face landmark detection using CFAN [46]. Model used: GoogleNet. External data: CASIA-WebFace database [45], Cross-Age Celebrity Dataset [6], MORPH. System prediction based on three cascade CNN: face classification, real age, apparent age.
AgeSeer	Face and face landmark detection using a comercial software. Model used: CNN VGG. External data: Celebface+, Morph, FGNet, Adience. Prediction of age codes, fusion of regressors, such as lasso, global and local quadratic regressor, and random forest.
WVU_CVL	Face and landmark detection using Face++, MS Project Oxford. Model used: GoogleNet. External data: WebFace and 240k age data from (CACD, Adience, MORPH, FGnet, own collected data). Prediction based on Multiple deep features, 10 age groups, RF, SVR, and fusion.
SEU-NJU	Face detection based on [26] and face landmark detection based on [39]. Model used: VGG16 + CNN. External data: MORPH, own collected data. Prediction using fusion of different network setups, softmax loss and KL-divergence for training, including aligned and non-aligned faces for training with different color spaces and filters on the input data.
UMD	Face detection using [33] and face landmark detection using [22]. Model: CNN [7]. External data: Adience [10], MORPH. Prediction using CNN regression model with the challenge gaussian loss function, classifying in three age groups, and then regressing the age.
Enjuto	Face detection using [26] [23] and fusion by overlap. Face landmark detection using [22]. Model used: 6 CNN. External data: CNN regression model with the challenge gaussian loss function, classifying in three age groups, and then regressing the age. Prediction using fusion of some CNN global face learning and others local face image parts learning.
Sungbin Choi	Model used: GoogleNet. External data: The image of Groups Dataset, Cross-age Celebrity Dataset (CACD), Adience [10], ImageNet, Faces and Labeled Faces in the Wild. Final average for classification.
Lab219A	Face detection using Viola & Jones and face landmark detection using STASM. Model: CNN. External data: Adience [10].
Bogazici	Face detection using Viola & Jones, face landmark detection using SDM. Model: Kernel ELM regression. Prediction using Fusion of HOG GIST LBP SIFT, PCA reduction.
Notts CVLab	Face detection using [47], face landmark detection using [41]. Model: Cascaded Ridge Regression. Prediction using LGBP features [36].

First place (CVL_ETHZ). The proposed Deep EXpectation (DEX) method first detects the face in the test image and then extracts the CNN predictions from an ensemble of 20 networks on the cropped face. Each network (VGG-16 architecture) was pre-trained on Imagenet and then fine-tuned on face images from IMDB and Wikipedia. The resulting network was then again fine-tuned on the actual dataset from the challenge. The networks were trained for classification with 101 output neurons, each neuron corresponding to an integer age (0-100). The final prediction is the expected value of the softmax-normalized output of the last layer, averaged over the 20 networks.

Second place (ICT-VIPL). The proposed approach is an end-to-end learning framework based on general to specific deep transfer learning. The main steps are: 1) Pre-train 22 layer large-scale deep convolutional neural network for multiclass face classification using the CASIA-WebFace database; 2) Fine-tune 22 layer large-scale deep convolutional neural network for age estimation on large outside age dataset. In this work, two kind of losses are involved. The authors adopted a Euclidean loss for single dimension age encoding and a cross-entropy loss of label distribution learning based age encoding; 3) Fine-tune 22 layer large-scale deep convolutional neural network on the final apparent age training set; 4) And finally, Ensemble Learning, where the final age estimation output is the fusion of 10 deep neural network. The basic NN architecture is based on GoogleLet and the loss layer is depended on the task. For multi-class face classification, they adopted the softmax loss, while Euclidean loss and cross-entropy loss were used for age estimation task.

Third place (WVU_CVL). The method is based on multiple GoogleNet deep networks, which are trained on 240.000 public facial images with real age label. Then the data provided from this challenge is processed using data augmentation. The deep network is fine-tuned using augmented competition data and the feature vectors from deep networks are extracted. Age grouping is further applied so that each image is classified into one of ten age groups. Within each age group, Random Forest and SVR are used to train the age estimator. Score level fusion is applied to fuse all the predictions to get the final prediction result.

3. Cultural Event Recognition

Inspired by the Action Classification challenge of PASCAL VOC 2011-12 successfully organized by Everingham et al. [15], we planned to run a competition in which 99 categories corresponding to different world-wide cultural events and 1 non-class would be considered. In all the image categories, garments, human poses, objects, illumination, and context do constitute the possible cues to be exploited for recognizing the events, while preserving the inherent inter- and intra-class variability of this type of images. Thousands of images were downloaded and manually labeled, corresponding to cultural events like Carnivals (Brasil, Italy, USA), La Tomatina (Spain), Holi Festival (India) and Inti Raymi (Peru), among others. Figure 3 depicts in shades of blue the amount of cultural events per country.

Following the success of the 2014 cultural event recognition challenge that we organized, in this paper we present the results of our second edition of the challenge organized in 2015. We introduce an extended database based on cultural events [2] and the second cultural event recognition challenge. In this section, we discuss some of the works most closely related to it.

Action Classification Challenge [15] This challenge belongs to the PASCAL - VOC challenge which is a benchmark in visual object category recognition and detection. In particular, the Action Classification challenge was introduced in 2010 with 10 categories. This challenge consisted in predicting the action(s) being performed by a person in a still image. In 2012 there were two variations of this competition, depending on how the person (whose actions are to be classified) was identified in a test image: (i) by a tight bounding box around the person; (ii) by only a single point located somewhere on the body.

Table 5. Comparison between our dataset and others present in the state of the art.

Dataset	#Images	#Categories	Year
Action Classification Dataset [15]	5,023	10	2010
Social Event Dataset [29]	160,000	149	2012
Event Identification Dataset [4]	594,000	24,900	2010
Cultural Event Dataset I [2]	11,776	50	2015
Cultural Event Dataset II	28,705	100	2015

Social Event Detection [29] This work is composed of three challenges and a common test dataset of images with their metadata (timestamps,tags, geotags for a small subset of them). The first challenge consists of finding technical events that took place in Germany in the test collection. In the second challenge, the task consists of finding all soccer events taking place in Hamburg (Germany) and Madrid (Spain) in the test collection. The third challenge aims at finding demonstration and protest events of the *Indignados* movement occurring in public places in Madrid.

Event Identification in Social Media [4] In this work the authors introduce the problem of event identification in social media. They presented an incremental clustering algorithm that classifies social media documents into a growing set of events.

Table 5 shows a comparison between our cultural event dataset and the others present in the state of the art. Action Classification dataset is the most closely related, but the amount of images and categories is smaller than ours. Although the number of the images and categories in the datasets [29] and [4] are larger than our dataset, these dataset are not related to cultural events but to events in general. Some examples of the events considered in these dataset are soccer events (*football games that took place in Rome in January 2010*), protest events (*Indignados movement occurring in public places in Madrid*), etc.

3.1. Dataset

The Cultural Event Recognition challenge aims to investigate the performance of recognition methods based on several cues like garments, human poses, objects, background, etc. To this end, the cultural event dataset contains significant variability in terms of clothes, actions, illumination, localization and context.

The Cultural Event Recognition dataset consists of images collected from two images search engines (Google Images and Bing Images). To build the dataset, we chose 99 important cultural events in the world and we created several queries with the names of these events. In order to in-

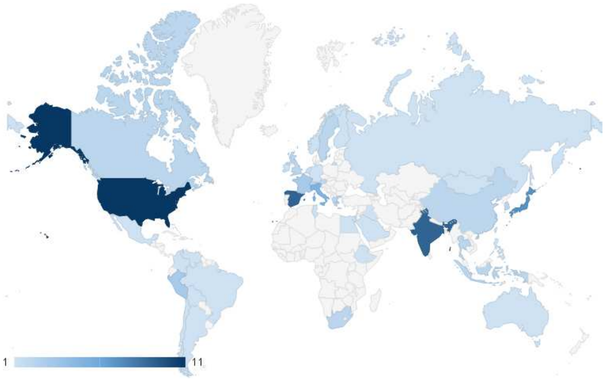


Figure 3. Cultural events by country, dark blue represents greater number of events.

crease the number of retrieved images, we combined the names of the events with some additional keywords (festival, parade, event, etc.). Then, we removed duplicated URLs and downloaded the raw images. To ensure that the downloaded images belonged to each cultural event, a process was applied to manually filter each of the images. Next, all exact duplicate and near duplicate images were removed from the downloaded image set using the method described in [8]. While we attempted to remove all duplicates from the database, there may exist some remaining duplicates that were not found. We believe the number of these is small enough to not significantly affect the research outcomes. In order to have a more challenging competition we aggregated a new class called non-class, this class represents a distractor class and is composed by 2,000 images containing no information related to cultural events. To build this non-class, we randomly downloaded images from ImageNet and then we filtered each of the images. In the end, our dataset is composed of 28,705 images.

The database can be downloaded at the following web address: https://www.codalab.org/competitions/4081. Some of the properties of the database and its associated challenge are: First database on world-wide cultural events; more than 28,000 images of 100 different events; and high intra- and inter-class variability. For this type of images, different cues can be exploited like garments, human poses, crowds analysis, objects and background scene. The evaluation metric will be the recognition accuracy.

Table 6 lists the 100 categories, country they belong and the number of images considered for this challenge.

There is no similar database in the literature. For example, the ImageNet competition does not include the cultural event taxonomy as considered in this specific track. In comparison with the Action Classification challenge of PASCAL VOC 2011-12, the database we constructed and used include more than twice the number of images, and over 10 times more categories.

Table 7. Table of Cultural Methods.

Team	Proposed Method
VIPL-ICT-CAS	Model used: VGGNet, GoogLeNet. Prediction using Logistic Regression and LDA for final classification.
FV	Model used: VGG16, VGG19, Place-CNN. Prediction using 5 CNNs, fusion of 5 feature vectors and final logistic regression classification [20].
MMLAB	Models used: GoogLeNet, VGGNet. Prediction using Object and scenes activations [44], CNN features and Fisher Vectors [43], final SVM classification.
NU&C	Model: CaffeNet based on ImageNet and Places205. Combination of Object CNN stream and Scene CNN stream for prediction.
CVL_ETHZ	Model: VGG-16 based on ImageNet and Places205. Prediction using Pooled and LDA-projected CNN features with Iterative Nearest Neighbors-based classifier [40].
SSTK	Model: CNN. Prediction using a combination of 13 pretrained CNN models on places-205 and ImageNet.
MIPAL_SNU	Model: GoogLeNet. 4 trained networks: region, image, person, face, and CNN features combined and RF prediction.
ESB	Models used: VGG16, GoogleNet. Prediction using RF classification.
Sungbin Choi	Model: GoogleNet based on MIRFLICKR-1M and ImageNet. Final average for classification.
UPC-STP	Model: AlexNet. Final SVM classification

3.2. Cultural event top ranked methods

More than 50 participants were registered to join the competition. Finally, at the test step of the competition, the teams that submitted their predictions (based on average precision performance) are shown in Table 8 and the their methods are described in Table 7. The first top ranked methods are detailed next.

First place (VIPL-ICT-CAS). The method is based on a combination of visual features extracted from deep convolutional neural networks. Specifically, authors investigate two off-the-shelf architectures, VGGNet and GoogLeNet, and adapt them to the task by performing event-specific fine-tuning on both global and local images. For global scheme, authors take the whole image as input; while for local scheme, authors first generate a batch of region proposals in each image and take these local regions as inputs. In recognition stage, it is employed two kinds of linear classifiers, Logistic Regression (LR) and Linear Discriminant Analysis (LDA) on image features from the different deep models and decision scores are fused.

Second place (FV). The approach is named DSP (Deep Spatial Pyramid). By considering utilizing the spatial infor-

Table 6. List of the 100 cultural event categories.

	Cultural Event	**Country**	**#Images**
1	Annual Buffalo Roundup	USA	230
2	Ati-atihan	Philippines	221
3	Ballon Fiesta	USA	270
4	Basel Fasnacht	Switzerland	225
5	Boston Marathon	USA	236
6	Bud Billiken	USA	261
7	Buenos Aires Tango Festival	Argentina	230
8	Carnaval de Dunkerque	France	253
9	Carnival of Venice	Italy	281
10	Carnivale Rio	Brazil	282
11	Castellers	Spain	322
12	Chinese New Year	China	289
13	Correfocs	Spain	251
14	Desert Festival of Jaisalmer	India	211
15	Desfile de Silleteros	Colombia	226
16	Da de los Muertos	Mexico	240
17	Diada de Sant Jordi	Spain	221
18	Diwali Festival of Lights	India	216
19	Falles	Spain	361
20	Festa del Renaixement	Spain	263
21	Festival de la Marinera	Peru	260
22	Inti Raymi	Peru	288
23	Fiesta de la Candelaria	Peru	243
24	Gion matsuri	Japan	258
25	Harbin Ice and Snow Festival	China	276
26	Heiva	Tahiti	236
27	Helsinki Samba Carnaval	Finland	229
28	Holi Festival	India	255
29	Infiorata di Genzano	Italy	239
30	La Tomatina	Spain	248
31	Lewes Bonfire	England	223
32	Macys Thanksgiving	USA	235
33	Maslenitsa	Russia	226
34	Midsommar	Sweden	259
35	Notting hill carnival	England	244
36	Obon Festival	Japan	253
37	Oktoberfest	Germany	329
38	Onbashira Festival	Japan	228
39	Pingxi Lantern Festival	Taiwan	225
40	Pushkar Camel Festival	India	243
41	Quebec Winter Carnival	Canada	244
42	Queens Day	Netherlands	246
43	Rath Yatra	India	264
44	SandFest	USA	235
45	San Fermin	Spain	306
46	Songkran Water Festival	Thailand	245
47	St Patrick's Day	Ireland	248
48	The battle of the Oranges	Italy	207
49	Timkat	Ethiopia	297
50	Viking Festival	Norway	241
51	July 4th	USA	301
52	AfrikaBurn	South Africa	422
53	Aomori nebuta	Japan	401
54	Apokries	Greece	290
55	Asakusa Samba Carnival	Japan	463
56	Australia day	Australia	241
57	Bastille day	France	232
58	Beltane Fire	Scotland	361
59	Boryeong Mud	South Korea	300
60	Carnaval de Oruro	Bolivia	248
61	Carnevale Di Viareggio	Italy	644
62	Cascamorras	Spain	223
63	Cheongdo Bullfighting Festival	South Korea	212
64	Crop over	Barbados	262
65	Eid al-Adha	Egypt	250
66	Eid al-Fitr Iraq	Iraq	282
67	Epiphany	Greece	270
68	Festa Della Sensa	Italy	204
69	Frozen Dead Guy Days	USA	230
70	Galugan	Indonesia	356
71	Grindelwald Snow Festival	Switzerland	142
72	Hajj	Saudi Arabia	308
73	Halloween Festival of the Dead	USA	275
74	Highland Games	Scotland	515
75	Junkanoo	Bahamas	290
76	Kaapse Klopse	South Africa	229
77	Keene Pumpkin Festival	USA	275
78	Krampusnacht Festival	Austria	142
79	Los Diablos danzantes	Venezuela	306
80	Magh Mela	India	200
81	Mardi Gras	USA	333
82	Monkey Buffet Festival	Thailand	188
83	Naadam Festival	Mongolia	360
84	Passover	Israel	273
85	Pflasterspektakel	Austria	218
86	Phi Ta Khon	Thailand	252
87	Sahara Festival	Tunisia	234
88	Sapporo Snow Festival	Japan	227
89	Spice Mas Carnival	Grenada	224
90	Sweden Medieval Week	Sweden	264
91	Tamborrada	Spain	451
92	Tapati rapa Nui	Chile	244
93	Thaipusam	India	318
94	Thrissur Pooram	India	336
95	Tokushima Awa Odori Festival	Japan	354
96	Tour de France	France	278
97	Up Helly Aa Fire Festival	Scotland	224
98	Vancouver Symphony of Fire	Canada	214
99	Waisak day	Indonesia	220
100	Non-class	-	2000

mation with fully convolutional activations, it is formed a natural deep spatial pyramid by partitioning an image into sub-regions and computing local features inside each sub-region. In practice, it is just needed to spatially partition the cells of activations in the last convolutional layer, and then pool deep descriptors in each region separately using Fisher Vectors. In order to capture variations of the activations caused by variations of objects in an image, a multiple scale pyramid approach with different rescaled versions of the original input image is applied. Images of all different scales are feed into a pre-trained CNN model. In each scale, the corresponding rescaled image is encoded. Then, vectors are merged into a single vector by average pooling. Data augmentation is also performed on the 99 cultural events classes and not for the non-cultural class. In particular VGG16, VGG19 and Place-CNN models were employed. Meanwhile, for VGG16 and VGG19, also fine-tune is performed using the training and validation images and crops. Finally, one test image is then represented by five equally instances, and at the test stage, the prediction scores of these five instances, obtained by a logistic regression and softmax prediction, are averaged to get the final score.

Third place (MMLAB). The approach uses a deep architecture to perform event recognition by extracting visual information from the perspectives of object and scene. Specifically, the proposed OS-CNN is composed of object net and scene net. Based on OS-CNN, it is presented an effective image representation, by extracting the activations of fully connected layers and convolutional layers. Average pooling is applied to aggregate the activations of fully connected layers. Then, Fisher vector encodes those convolutional layers, and an SVM is used for classification.

4. Conclusion

We reviewed the apparent age estimation and cultural event recognition challenges. We presented the first state of the art data set for apparent age estimation rather than real age. We also proposed a large dataset of cultural events composed by a hundred of categories and tens of thousands of samples. We ran a challenge for each of these two data sets. Results show that most of the participants based their solutions on deep learning architectures.

Acknowledgements

We thank to our sponsors: Microsoft Research, University of Barcelona, Amazon, INAOE, Google, NVIDIA corporation, Facebook, and Disney Research. This research has been partially supported by projects TIN2012-39051 and TIN2013-43478-P. We wish to thank Visada and California Naturel for providing the initial idea of the project, contributing advice and prizes for data collection, in particularly John Bernard Duler and Praveen Srinivasan.

Table 8. Results of the Cultural Event Recognition challenge.

Cultural Event Recognition

Position	Team	Development	Test
1	**VIPL-ICT-CAS**	**0.783**	**0.854**
2	**FV**	**0.770**	**0.851**
3	**MMLAB**	**0.717**	**0.847**
4	NU&C	0.387	0.824
5	CVL_ETHZ	0.662	0.798
6	SSTK	0.740	0.770
7	MIPAL_SNU	0.801	0.763
8	ESB	0.729	0.758
9	Sungbin Choi	-	0.624
10	UPC-STP	0.503	0.588

References

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

[2] X. Baro, J. Gonzàlez, J. Fabian, M. A. Bautista, M. Oliu, H. J. Escalante, I. Guyon, and S. Escalera. Chalearn looking at people 2015 challenges: action spotting and cultural event recognition. In *ChaLearn LAP Workshop, CVPR*, 2015.

[3] A. Bastanfard, M. Nik, and M. Dehshibi. Iranian face database with age, pose and expression. In *Int. Conf. Machine Vision, 2007*, pages 50–55, Dec 2007.

[4] H. Becker, M. Naaman, and L. Gravano. Learning similarity metrics for event identification in socialmedia. In *Proceedings WSDM*, 2010.

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

[6] B.-C. Chen, C.-S. Chen, and W. H. Hsu. Cross-age reference coding for age-invariant face recognition and retrieval. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2014.

[7] J.-C. Chen, V. M. Patel, and R. Chellappa. Unconstrained face verification using deep cnn features. *CoRR*, abs/1508.01722, 2015.

[8] O. Chum, J. Philbin, M. Isard, and A. Zisserman. Scalable near identical image and shot detection. In *ACM International Conference on Image and Video Retrieval*, 2007.

[9] 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.

[10] 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.

[11] S. Escalera, X. Baro, J. Gonzàlez, M. Bautista, M. Madadi, M. Reyes, V. Ponce, H. Escalante, J. Shotton, and I. Guyon. Chalearn looking at people challenge 2014: Dataset and results. *ChaLearn LAP Workshop, ECCV*, 2014.

[12] S. Escalera, J. Gonzàlez, X. Baro, P. Pardo, J. Fabian, M. Oliu, H. J. Escalante, I. Huerta, and I. Guyon. Chalearn looking at people 2015 new competitions: Age estimation and cultural event recognition. In *IJCNN*, 2015.

[13] S. Escalera, J. Gonzàlez, X. Baro, M. Reyes, I. Guyon, V. Athitsos, H. Escalante, A. Argyros, C. Sminchisescu, R. Bowden, and S. Sclarof. Chalearn multi-modal gesture recognition 2013: grand challenge and workshop summary. *ICMI*, pages 365–368, 2013.

[14] S. Escalera, J. Gonzàlez, X. Baró, M. Reyes, O. Lopés, I. Guyon, V. Athitsos, and H. J. Escalante. Multi-modal gesture recognition challenge 2013: Dataset and results. In *ChaLearn Multi-Modal Gesture Recognition, ICMI Workshop*, 2013.

[15] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *IJCV*, 88(2):303–338, June 2010.

[16] S. J. Foundation. Human and Object Interaction Processing (HOIP) Face Database. Available at `http://www.hoip.jp/`, 2014.

[17] Y. Fu, G. Guo, and T. Huang. Age synthesis and estimation via faces: A survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(11):1955–1976, Nov 2010.

[18] Y. Fu and T. Huang. Human age estimation with regression on discriminative aging manifold. *Multimedia, IEEE Transactions on*, 10(4):578–584, June 2008.

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

[20] B.-B. Gao, X.-S. Wei, J. Wu, and W. Lin. Deep spatial pyramid: The devil is once again in the details. *arXiv:1504.05277*, 2015.

[21] H. Han, C. Otto, and A. K. Jain. Age estimation from face images: Human vs. machine performance. In *ICB'13*, pages 1–8, 2013.

[22] V. Kazemi and J. Sullivan. One millisecond face alignment with an ensemble of regression trees. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[23] D. E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research (JMLR)*, 10:1755–1758, 2009.

[24] A. Lanitis. FG-NET Aging Data Base, November 2002.

[25] A. Lanitis, C. Taylor, and T. Cootes. Toward automatic simulation of aging effects on face images. volume 24, pages 442–455, 2002.

[26] M. Mathias, R. Benenson, M. Pedersoli, and L. V. Gool. Face detection without bells and whistles. In *European Conference on Computer Vision (ECCV)*, 2014.

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

[28] B. Ni, Z. Song, and S. 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.

[29] S. Papadopoulos, E. Schinas, V. Mezaris, R. Troncy, and I. Kompatsiaris. Social event detection at mediaeval 2012: Challenges, dataset and evaluation. In *Proc. MediaEval 2012 Workshop*, 2012.

[30] P. Phillips, H. Wechsler, J. Huang, and P. J. Rauss. The {FERET} database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16(5):295 – 306, 1998.

[31] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the Face Recognition Grand Challenge. In *CVPR*, pages 947–954. IEEE, 2005.

[32] N. Ramanathan, R. Chellappa, and S. Biswas. Computational methods for modeling facial aging: A survey. *Journal of Visual Languages and Computing*, 20(3):131 – 144, 2009.

[33] R. Ranjan, V. M. Patel, and R. Chellappa. A deep pyramid deformable part model for face detection. *CoRR*, abs/1508.04389, 2015.

[34] K. Ricanek and T. Tesafaye. MORPH: a longitudinal image database of normal adult age-progression. In *Int. Conf. FG*, pages 341–345, 2006.

[35] K. Ricanek and T. Tesafaye. Morph: a longitudinal image database of normal adult age-progression. In *Int. Conf. FG*, pages 341–345, April 2006.

[36] T. Senechal, V. Rapp, H. Salam, R. Seguier, K. Bailly, and L. Prevost. Combining aam coefficients with lgbp histograms in the multi-kernel svm framework to detect facial action units. In *Int. Conf. FG*, pages 860–865, March 2011.

[37] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression (pie) database. In *Int. Conf. FG*, pages 46–51, May 2002.

[38] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556, 2014.

[39] Y. Sun, X. Wang, and X. Tang. Deep convolutional network cascade for facial point detection. In *CVPR*, pages 3476–3483. IEEE Computer Society, 2013.

[40] R. Timofte and L. V. Gool. Iterative nearest neighbors for classification and dimensionality reduction. In *CVPR*, CVPR '12, pages 2456–2463, Washington, DC, USA, 2012. IEEE Computer Society.

[41] G. Tzimiropoulos. Project-out cascaded regression with an application to face alignment. June 2015.

[42] K. Ueki, T. Hayashida, and T. Kobayashi. Subspace-based age-group classification using facial images under various lighting conditions. In *Int. Conf. FG*, pages 43–48, 2006.

[43] L. Wang, Y. Qiao, and X. Tang. Action recognition with trajectory-pooled deep-convolutional descriptors. June 2015.

[44] L. Wang, Z. Wang, W. Du, and Y. Qiao. Object-scene convolutional neural networks for event recognition in images. In *CVPR Workshop*, 2015.

[45] D. Yi, Z. Lei, S. Liao, and S. Z. Li. Learning face representation from scratch. *arXiv preprint*, arXiv:1411.7923, 2014.

[46] J. Zhang, S. Shan, M. Kan, and X. Chen. Coarse-to-fine auto-encoder networks (CFAN) for real-time face alignment. In *ECCV*, pages 1–16, 2014.

[47] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *CVPR*, pages 2879–2886, Washington, DC, USA, 2012. IEEE Computer Society.