ChaLearn Looking at People 2015 new competitions: Age Estimation and Cultural Event Recognition

Sergio Escalera
University of Barcelona
Computer Vision Center, UAB
Email: sergio@maia.ub.es

Jordi Gonzàlez
Univ. Autònoma de Barcelona
Computer Vision Center, UAB
Email: poal@cvc.uab.es

Xavier Baró
Universitat Oberta de Catalunya
Computer Vision Center, UAB
Email: xbaro@uoc.edu

Pablo Pardo
University of Barcelona
Email: ppardoga7@gmail.com

Junior Fabian
Univ. Autònoma de Barcelona
Computer Vision Center, UAB
Email: jfabian@cvc.uab.es

Marc Oliu
University of Barcelona
Computer Vision Center, UAB
Email: moliusimon@gmail.com

Hugo Jair Escalante
INAOE
Email: hugo.jair@gmail.com

Ivan Huerta
University of Venezia
Email: huertacasado@iuav.it

Isabelle Guyon
Clopinet, Berkeley
Email: guyon@chalearn.org

***Abstract*—Following previous series on Looking at People (LAP) challenges [1], [2], [3], in 2015 ChaLearn runs two new competitions within the field of Looking at People: age and cultural event recognition in still images. We propose the first crowdsourcing application to collect and label data about apparent age of people instead of the real age. In terms of cultural event recognition, tens of categories have to be recognized. This involves scene understanding and human analysis. This paper summarizes both challenges and data, providing some initial baselines. The results of the first round of the competition were presented at ChaLearn LAP 2015 IJCNN special session on computer vision and robotics http://www.dtic.ua.es/∼jgarcia/IJCNN2015. Details of the ChaLearn LAP competitions can be found at http://gesture.chalearn.org/.**

I. Introduction

The automatic analysis of the human body in still images and image sequences, also known as Looking at People, keeps making rapid progress with the constant improvement of new published methods that push the state-of-the-art. Applications are countless, like Human Computer Interaction, Human Robot Interaction, communication, entertainment, security, commerce and sports, while having an important social impact in assistive technologies for the handicapped and the elderly.

In 2015, ChaLearn is organizing 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 expression, 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 context. Therefore, how to combine and exploit all this knowledge from pixels constitutes an interesting problem.

This motivates our choice to organize a new workshop and a competition on this topic to sustain the effort of the computer vision community. These new competitions come as a natural evolution from our previous workshops at CVPR 2011, CVPR 2012, ICPR 2012, ICMI 2013, and ECCV 2014. We will continue to use our website http://gesture.chalearn.org for promotion, while challenge entries in the quantitative competition will be scored on-line using the Codalab Microsoft-Stanford University platforms (http://codalab.org/), from which we have already organized international challenges related to Computer Vision and Machine Learning problems. The results of the first round of the competition will be presented at ChaLearn LAP 2015 IJCNN special session on computer vision and robotics http://www.dtic.ua.es/∼jgarcia/IJCNN2015.

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 in terms of application and data, and provide some initial baselines.

II. Age estimation challenge

Age estimation is a challenging 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/). We are currently collecting the data, which will be composed by thousands of images labeled by many users.

A. State of the art on age recognition

Age estimation has historically been one of the most challenging problems within the field of facial analysis [4], [5]. It can be very useful for several applications, such as advanced video surveillance ([5], [6]), demographic statistics collection,

business intelligence and customer profiling, and search optimization in large databases. Some of the reasons age estimation is yet a challenging problem are the uncontrollable nature of the aging process, the strong specificity to the personal traits of each individual [7], 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 [8].

One of the earliest works in age estimation was done by A. Lanitis et al. [9], [10], [11]. They propose two different aging estimation methods: weighted appearance specific method [10], [11]. The aging factor of a new individual is computed by the weighted sum of the aging functions of other individuals. They also use an age specific method [9], where the new individual is first classified into a cluster with similar aging factor patterns. Then, it is classified into different age ranges and an age-specific classifier is applied to estimate the final age.

There are some drawbacks to this "aging function" approach pointed out by X. Geng et al. [12]. The formula of the aging function is empirically determined, there is no evidence suggesting that the relation between face and age is described just by a quadratic function. Also, the new aging function for the unseen images is simply a linear combination of the already known aging functions. X. Geng et al. claimed to solve these problems in their new proposed method AGing pattErn Subspace (AGES) [12], where each face image is represented by a point in the aging pattern subspace.

Later N. Ramanathan et al. [13], [14] approached the age estimation problem in two different scenarios, estimating the age difference between two face images of the same individual based on a Bayesian age-difference classifier [13] and estimating the age of young faces using the facial growth geometry [14]. The problem of the last approach 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.

Y. Fu et al. were the first to approach the problem through manifold analysis methods [15], [16]. Each face image is assigned to a low-dimensional representation via manifold embedding. Following this approach G. Guo et al. [17] proposed a new method based on a study of different dimensionality reduction and manifold embedding techniques and added a robust regression step to the previous framework. In a posterior work [18], G. Guo et al. introduced a new approach, using kernel partial least square (KPLS) regression to reduce feature dimensionality and learn the aging function in a single step.

G. Guo et al. also proposed different approaches to the age estimation problem such as [19], where they define a probabilistic fusion approach, or [20] where they introduce the Biologically Inspired Features (BIF) for the age estimation problem and propose some changes adding a novel operator. In a recent paper [21], Guo et al. used the BIF features, and focused on investigating a novel single-step framework for joint estimation of age, gender and ethnicity. Both the CCA (Canonical Correlation Analysis) and PLS (Partial Least Square) based methods were explored under the joint estimation framework.

Under the same idea as Y. Fu et al. [15], K. Luu et al. [22], [23] reduced dimensionality by using facial landmarks and Active Shape Models (ASM) [22] and an improved version, Contourlet Appearance Model (CAM) [23], where they prove the efficiency of using facial landmarks. Then T. Wu et al. [24] proposed to use facial landmarks and project them into a Grassmann manifold to model the age patterns.

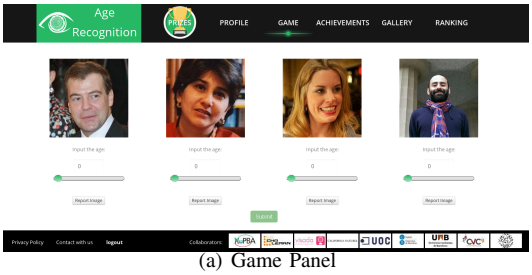
With regards to the learning algorithm, Support Vector Machines (SVM) have commonly been used for age classification and regression, as in [20]. A binary decision tree with SVMs at each node is proposed in [6] by Han et al. Age ranges are coarsely assigned, and later are more precisely estimated by Support Vector Regressors (SVR) at the leaves. Chang et al. in [25] used a 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. In this paper Active Appearance Model (AAM) is used. In [7], Weng et al. employs a similar ranking technique, where Local Binary Patterns (LBP) histogram features are combined with principal components of BIF, shape and textural features of AAM, and 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 [26].

There have been previous proposals training neural networks, which are able to learn complex mappings and deal with outliers, for age estimation. In [9], 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. in [8] tackle 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. [27]. In [28], Yang et al. used Convolutional neural networks for age estimation under surveillance scenarios.

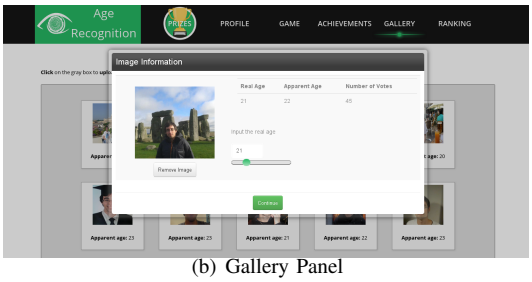
Other different variations of the problem has been addressed, A. Lanitis et al. [29] performed a first approach to age estimation using Head and Mouse tracking movements, Y. Makihara et al. [30] used a gait-based database to estimate the age, B. Xia et al. [31] proposed an age estimation method based on 3D face images.

B. Dataset

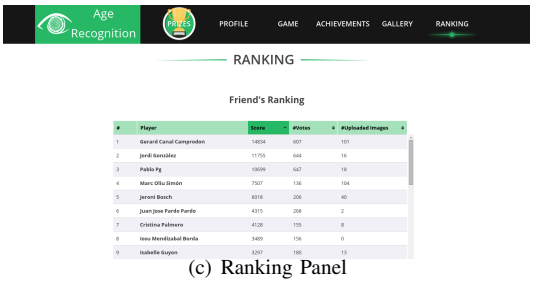
Due to the nature of the age estimation problem, there is a restricted number of publicly available databases providing a substantial number of face images labeled with accurate age information. Table I shows the summary of the existing databases with main reference, number of samples, number of subjects, age range, type of age and additional information. After an initial interest on automatic age estimation from images dated back to the early 2000s [9], [11], [32], research in the field has experienced a renewed interest from 2006 on, since the availability of large databases like MORPH-Album 2 [33], which increased by 55$\times$ the amount of real age-annotated data with respect to traditional age databases.



(a) Game Panel



(b) Gallery Panel



(c) Ranking Panel

Fig. 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.

Therefore, this database has deeply been employed in recent works by applying over it different descriptors and classification schemes. However, all existing datasets are based on real age estimation. In our proposed challenge, we propose the first dataset to recognize the apparent age of people based on the opinion on many subjects using a new crowsourcing data collection and labeling application.

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 obtains. In order to increase the engagement of the players we add a global and friends leaderboard where the users can see their position in the ranking. We ask the users to upload images of a single person and we give them tools to crop the image if necessary, we also ask them to give the real age (or as close as possible)

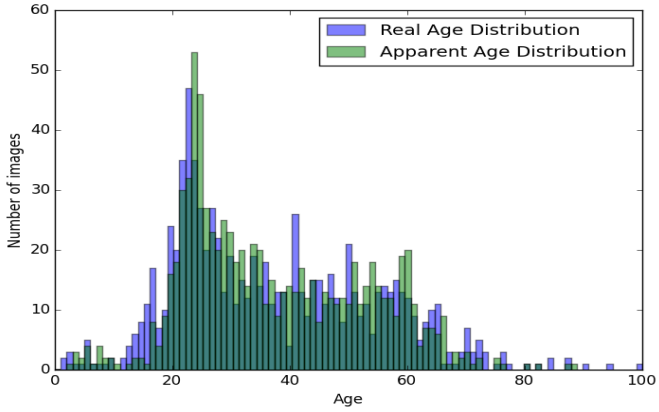


Fig. 2. Real and Apparent age distributions in our database.

of the person in uploaded image, allowing more analysis and comparisons with real age and apparent age.

Few weeks after release the application we have already collected near 1000 images and near 10000 votes. These numbers will continue growing in order to generate the future competition. Some of the properties of the database which is being collected with the web application are listed below:

- Thousands of faces labeled by many users.
- Images with background.
- Non-controlled environments.
- Non-labeled faces neither landmarks, making the estimation problem even harder.
- 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 pondered by the mean and the variance of the labeling by the participants.
- The dataset also provides for each image the real age although not used for recognition (just for analysis purposes). In the same way for all the labelers 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 I, 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 the Figure 2. The images of our database has 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.

C. Baseline

We propose a simple baseline for the age estimation problem based on Biological Inspired Features and a hierarchical age estimation. The steps proposed are described below:

TABLE I. AGE-BASED DATABASES

Database	#Faces	#Subj.	Range	Type of age	Controlled Env.	Balanced age Distr.	Other annotation
FG-NET [11], [34]	1,002	82	0 - 69	Real Age	No	No	68 Landmarks
GROUPS [35]	28,231	28,231	0 - 66+	Age group	No	No	-
PAL [32]	580	580	19 - 93	Age group	No	No	-
FRGC [36]	44,278	568	18 - 70	Real Age	Partially	No	-
MORPH2 [33]	55,134	13,618	16 - 77	Real Age	Yes	No	-
YGA [16]	8,000	1,600	0 - 93	Real Age	No	No	-
FERET[37]	14,126	1,199	-	Real Age	Partially	No	-
Iranian face [38]	3,600	616	2 - 85	Real Age	No	No	Kind of skin and cosmetic points[1]
PIE [39]	41,638	68	-	Real Age	Yes	No	-
WIT-BD [40]	26,222	5,500	3 - 85	Age group	No	No	-
Caucasian Face Database [41]	147	-	20 - 62	Real Age	Yes	No	Shape represented in 208 key points
LHI [42]	8,000	8,000	9 - 89	Real Age	Yes	Yes	-
HOIP [43]	306,600	300	15 - 64	Age Group	Yes	No	-
Ni's Web-Collected Database [44]	219,892	-	1 - 80	Real Age	No	No	-
OUI-Adience [45]	26.580	2.284	0 - 60+	Age Group	No	No	Gender

• **Face Preprocessing**: Because the image capture conditions are not restricted in color an illumination, we decided to convert the images to grey-scale. A Viola-Jones face detector was used to crop the faces from the images. The faces were resized to 200×200 pixels resolution.

• **Facial Shape Extraction**: In order to achieve face alignment of the images and extract the shape vector we used the method proposed by Shaoqing Ren et al. [46], the regressor is trained with labeled images from five different datasets (AFW [47], FG-NET [34], HELEN [48], IBUG [49], LFPW [50]). The shape vector consists of 68 landmarks for each face image.

• **Feature Extraction**: We calculated the BIF features from the face images using the same Gabor filters and pooling method as G. Guo et al. [20]. We then performed a PCA to the extracted BIF and the 68 landmarks coordinates separately in order to reduce dimensionality. We then merged both feature sets.

• **Group Age Classification**: A SVM classifier with RBF kernel was trained to classify the instances into three disjoint age groups youth (0-25), adult (26-50) and middle-aged/pensioners (51-100).

• **Age Estimation**: Three specialized SVR (with RBF kernel) were trained, one for each age group. The data used to train them was face images within each age group up to 5 years outside the age range to mitigate error due to age miss-classification [51]. The age estimation is improved by fusing the prediction of different age estimators with adjacent age ranges when the predicted age is close to the borders of its age range.

The classifier and the regressors were trained and tested with a 10-Fold Cross Validation. In Figure 3 we show the performance of the baseline method for the current set of images of the dataset, using both the real age and the apparent age labels. Interestingly, one can see that the performance is significantly better with the apparent age than with the real age, which implies that current recognition approaches are better to approximate human opinion rather than real age. The Mean Average Error is 13.3 ± 0.3 using real age and 12.3 ± 0.2 using apparent age, the large error is due to the complexity of the database.

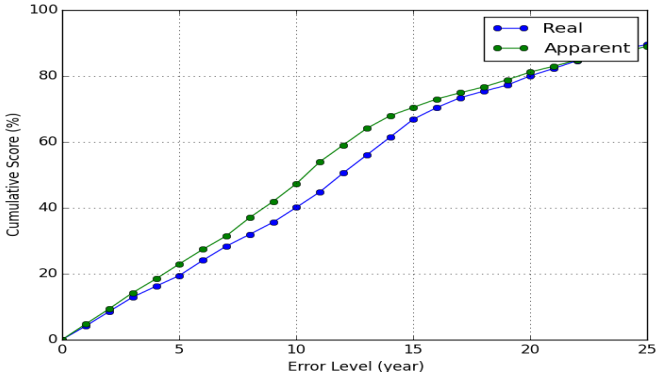


Fig. 3. Cumulative Score on our database with Real age and Apparent age estimations. Error level axis indicates the absolute number of deviation years between the predicted year and the ground truth to indicate a sample as correctly or incorrectly classified. Cumulative error shows (1 - recognition_accuracy) for each particular error level.

TABLE II. COMPARISON BETWEEN OUR DATASET AND OTHERS PRESENT IN THE STATE OF THE ART.

Dataset	#Images	#Categories	Year
Action Classification Dataset [52]	5,023	10	2010
Social Event Dataset [53]	160,000	149	2012
Event Identification Dataset [54]	594,000	24,900	2010
Cultural Event Dataset	11,776	50	2015

III. CULTURAL EVENT RECOGNITION

Inspired by the Action Classification challenge of PASCAL VOC 2011-12 successfully organized by Everingham et al. [52], we planned to run a competition in which 50 categories corresponding to different world-wide cultural events 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 Carnival (Brasil, Italy, USA), Oktoberfest (Germany), San Fermin (Spain), Holi Festival (India) and Gion Matsuri (Japan), among others. Figure 4 depicts in shades of green the amount of cultural events selected by country.

A. State of the art on cultural event recognition

In this work, we introduce the first database based on cultural events and the first cultural event recognition challenge. In this section, we discuss some of the works most closely related to it.

Action Classification Challenge [52] 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 on 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.

Social Event Detection [53] 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 in the test collection.

Event Identification in Social Media [54] 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 II 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 [53] and [54] 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.

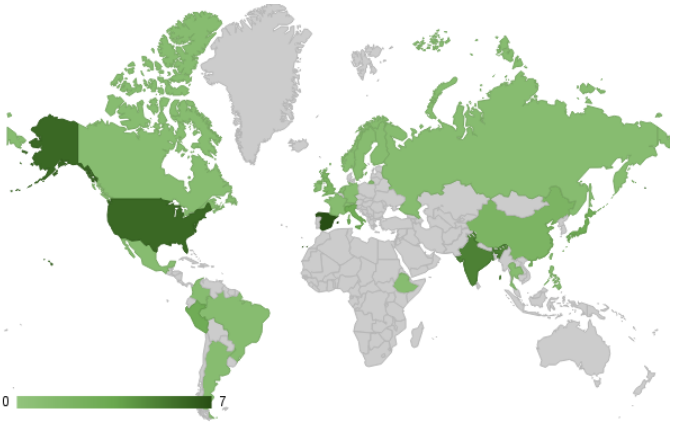


Fig. 4. Cultural events by country, dark green represents greater number of events.

B. 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 50 important cultural events in the world and we created several queries with the names of these events. In order to increase 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 [55]. 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 so that they will not significantly impact research. After all this preprocessing, our dataset is composed of 11,776 images.

The database can be viewed and downloaded at the following web address: https://www.codalab.org/competitions/2611. Some additional details and main contributions of the cultural event database are described below:

- First database on cultural events from all around the globe.

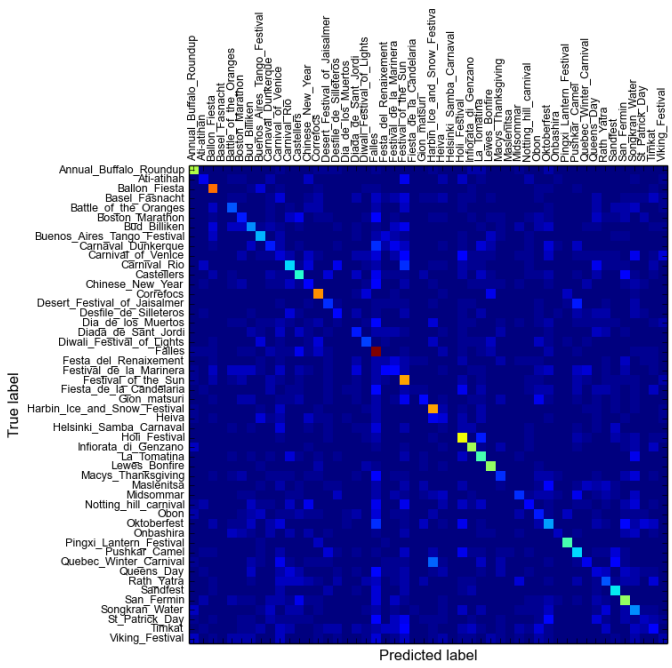


Fig. 5. Confusion Matrix obtained in our experiments for the cultural event dataset.

• More than 11,000 images representing 50 different categories.

• 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 III lists the 50 selected cultural events, country they belong, number of images considered for this challenge and the average precision (AP) obtained in our experiments.

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. Considering the Action Classification challenge of PASCAL VOC 2011-12, the number of images will be similar, around 11,000, but the number of categories will be here increased more than 5 times.

C. Baseline

To build a baseline in the cultural event database we have evaluated an approach based on bag-of-words. The first step of our approach consists of extracting low-level features. In this sense, we used a sparse-sampling approach based on the Speeded Up Robust Features (SURF) algorithm. Thereafter, we compute a mid-level image representation using a visual dictionary. We select the most representative points of interest according to each particular cultural event by means of clustering. The final set of selected points of interest represents a projection space onto which the points of interest found in any image are projected creating its representative feature vector. Given a set of "words" from the visual dictionary, we find the feature vector representing each image of the collection analyzing and assigning each of its PoIs to the closest visual word in the dictionary. Finally, we use this representation, in a third step, to train classifiers which provide decision scores to each test image. The main steps of our approach are formalized bellow:

Low-level Feature Extraction: For each image in the cultural event database, the points of interest are extracted using the SURF descriptor. Here, the objective is to find

TABLE III. LIST OF THE 50 CULTURAL EVENTS.

Cultural Event	Country	#Images	AP
1. Annual Buffalo Roundup	USA	334	0.699
2. Ati-atihan	Philippines	357	0.173
3. Ballon Fiesta	USA	382	0.669
4. Basel Fasnacht	Switzerland	310	0.040
5. Boston Marathon	USA	271	0.086
6. Bud Billiken	USA	335	0.158
7. Buenos Aires Tango Festival	Argentina	261	0.161
8. Carnival of Dunkerque	France	389	0.090
9. Carnival of Venice	Italy	455	0.045
10. Carnival of Rio	Brazil	419	0.269
11. Castellers	Spain	536	0.253
12. Chinese New Year	China	296	0.110
13. Correfocs	Catalonia	551	0.704
14. Desert Festival of Jaisalmer	India	298	0.118
15. Desfile de Silleteros	Colombia	286	0.082
16. Día de los Muertos	Mexico	298	0.051
17. Diada de Sant Jordi	Catalonia	299	0.098
18. Diwali Festival of Lights	India	361	0.254
19. Falles	Spain	649	0.417
20. Festa del Renaixement	Catalonia	299	0.026
21. Festival de la Marinera	Peru	478	0.099
22. Festival of the Sun	Peru	514	0.441
23. Fiesta de la Candelaria	Peru	300	0.044
24. Gion matsuri	Japan	282	0.159
25. Harbin Ice and Snow Festival	China	415	0.605
26. Heiva	Tahiti	286	0.106
27. Helsinki Samba Carnival	Finland	257	0.036
28. Holi Festival	India	553	0.479
29. Infiorata di Genzano	Italy	354	0.580
30. La Tomatina	Spain	349	0.401
31. Lewes Bonfire	England	267	0.724
32. Macys Thanksgiving	USA	335	0.167
33. Maslenitsa	Russia	271	0.024
34. Midsommar	Sweden	323	0.126
35. Notting hill carnival	England	383	0.088
36. Obon Festival	Japan	304	0.219
37. Oktoberfest	Germany	509	0.158
38. Onbashira Festival	Japan	247	0.035
39. Pingxi Lantern Festival	Taiwan	253	0.770
40. Pushkar Camel Festival	India	433	0.272
41. Quebec Winter Carnival	Canada	329	0.141
42. Queens Day	Netherlands	316	0.074
43. Rath Yatra	India	369	0.202
44. SandFest	USA	237	0.357
45. San Fermin	Spain	418	0.343
46. Songkran Water Festival	Thailand	398	0.252
47. St Patrick's Day	Ireland	320	0.096
48. The Battle of the Oranges	Italy	276	0.292
49. Timkat	Ethiopia	425	0.107
50. Viking Festival	Norway	262	0.048
		mAP	**0.239**

(a) Quebec Winter Carnival

(b) Harbin Ice and Snow Festival

Fig. 6. Example of mislabeled cultural events.

scale-invariant interest points such that we have a robust representation.

Compute Intermediate-level Features: Low-level features are not enough to fully represent images of cultural events. In this sense, we use the concept of visual vocabularies to increase the descriptor generalization. We use the standard Bag-of-Words framework, in which the K-means technique is applied over the points of interest to obtain k visual words and create a visual dictionary representing the categories of interest. In this step, we assess three vocabulary sizes: 100, 500 and 1,000 words. Then, we select the best-performing dictionary to evaluate our experiments.

Classification: Finally, we use a standard supervised learning technique (SVM, guassian kernel) to define the score for each test image.

We assess our experiments by using 50% of the dataset as the training set, 20% as the validation set and the remaining 30% to test our experiments, as it this defined within the competition for the participants. Table III shows the average precision (AP) obtained in our experiments for each cultural event. The mean average precision (mAP) is also showed in Table III, this measure represents our baseline for the competition. In our obtained results, we can note that the best accurate category is *Pingxi Lantern Festival* (AP $= 0.77$). This is because the bag of words approach is very successful in classifying images with well-defined objects, and in this case most of the images of this event contains explicit objects (lanterns).

Additionally, in Figure 5 we show the confusion matrix to evaluate the quality of our baseline. By analyzing this matrix we can observe the following:

- *Falles* is the cultural event with more correct predictions.
- *Quebec Winter Carnival* is mislabeled with *Harbin Ice and Snow Festival*. The reason is that these events have many images related to ice sculptures. Figure 6 depicts examples of images of these events.
- Most of mislabeled have been between festivals and carnivals. This is because in most of the carnivals, people wear costumes as well as in festivals.

The obtained baseline and the confusion matrix show that our dataset is a challenging problem. The main reason is that the cultural events have a high variability in terms of clothes, actions, illumination, localization and context. This allows the participants of this track the possibility to fully exploit and combine different image features.

IV. Conclusion

We summarized the two new Looking at People competitions ChaLearn is organizing for 2015, which include age estimation and cultural event recognition. We proposed the first crowd-sourcing application to collect and label data about the apparent age of people instead of the real age. In terms of cultural event recognition, tens of categories have to be recognized. This involves scene understanding and human analysis. We provided initial baseline for the two challenge competitions. The results of the first round of the competition will be presented at ChaLearn LAP 2015 IJCNN special session on computer vision and robotics http://www.dtic.ua.es/∼jgarcia/IJCNN2015. Details of the ChaLearn LAP competitions can be found at http://gesture.chalearn.org/.

Acknowledgment

The authors would like to thank John Bernard Duler, Praveen Srinivasan, Xavier Pérez-Sala, and Josep Gonfaus for their support during the design of both age and cultural event competitions. Also we would like to thank to the sponsors of these competitions: Microsoft Research, University of Barcelona, Amazon, INAOE, VISADA, and California Naturel. This research has been partially supported by research projects TIN2012-39051 and TIN2013-43478-P. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Tesla K40 GPU used for creating the baseline of the Cultural Event Recognition track.

References

[1] 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 Grand Challenge and Workshop, 15th ACM International Conference on Multimodal Interaction*, 2013.

[2] S. Escalera, J. Gonzalez, 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," *15th ACM International Conference on Multimodal Interaction*, pp. 365–368, 2013.

[3] S. Escalera, X. Baro, J. Gonzalez, 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 Looking at People, European Conference on Computer Vision*, 2014.

[4] N. Ramanathan, R. Chellappa, and S. Biswas, "Computational methods for modeling facial aging: A survey," *Journal of Visual Languages and Computing*, vol. 20, no. 3, pp. 131 – 144, 2009.

[5] Y. Fu, G. Guo, and T. Huang, "Age synthesis and estimation via faces: A survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 11, pp. 1955–1976, Nov 2010.

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

[7] R. Weng, J. Lu, G. Yang, and Y.-P. Tan, "Multi-feature ordinal ranking for facial age estimation," in *AFGR*. IEEE, 2013.

[8] X. Geng, C. Yin, and Z.-H. Zhou, "Facial age estimation by learning from label distributions," in *TPAMI*, vol. 35. IEEE, 2013, pp. 2401–2412.

[9] A. Lanitis, C. Draganova, and C. Christodoulou, "Comparing different classifiers for automatic age estimation," *Trans. Sys. Man Cyber. Part B*, vol. 34, no. 1, pp. 621–628, Feb. 2004.

[10] A. Lanitis, C. Taylor, and T. Cootes, "Modeling the process of ageing in face images," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 1, 1999, pp. 131–136 vol.1.

[11] A. Lanitis, C. Taylor, and T. Cootes, "Toward automatic simulation of aging effects on face images," vol. 24, no. 4, 2002, pp. 442–455.

[12] X. Geng, Z.-H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *TPAMI*, vol. 29, no. 12, pp. 2234–2240, 2007.

[13] N. Ramanathan and R. Chellappa, "Face verification across age progression," *Image Processing, IEEE Transactions on*, vol. 15, no. 11, pp. 3349–3361, Nov 2006.

[14] N. Ramanathan and R. Chellappa, "Modeling age progression in young faces," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 1, 2006, pp. 387–394.

[15] Y. Fu, Y. Xu, and T. Huang, "Estimating human age by manifold analysis of face pictures and regression on aging features," in *Multimedia and Expo, 2007 IEEE International Conference on*, July 2007, pp. 1383–1386.

[16] Y. Fu and T. Huang, "Human age estimation with regression on discriminative aging manifold," *Multimedia, IEEE Transactions on*, vol. 10, no. 4, pp. 578–584, June 2008.

[17] G. Guo, Y. Fu, C. Dyer, and T. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *Image Processing, IEEE Transactions on*, vol. 17, no. 7, pp. 1178–1188, July 2008.

[18] G. Guo and G. 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*, June 2011, pp. 657–664.

[19] G. Guo, Y. Fu, C. Dyer, and T. Huang, "A probabilistic fusion approach to human age prediction," in *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on*, June 2008, pp. 1–6.

[20] G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features." in *CVPR*. IEEE, pp. 112–119.

[21] G. Guo and G. Mu, "A framework for joint estimation of age, gender and ethnicity on a large database," *Image and Vision Computing*, vol. 32, no. 10, pp. 761 – 770, 2014.

[22] K. Luu, K. Ricanek, T. D. Bui, and C. 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*, ser. BTAS'09. Piscataway, NJ, USA: IEEE Press, 2009, pp. 314–318.

[23] K. Luu, K. Seshadri, M. Savvides, T. D. Bui, and C. Y. Suen, "Contourlet appearance model for facial age estimation." in *IJCB*, A. K. Jain, A. Ross, S. Prabhakar, and J. Kim, Eds. IEEE, pp. 1–8.

[24] T. Wu, P. K. Turaga, and R. Chellappa, "Age estimation and face verification across aging using landmarks." *IEEE Transactions on Information Forensics and Security*, no. 6, pp. 1780–1788.

[25] K.-Y. Chang, C.-S. Chen, and Y.-P. Hung, "Ordinal hyperplanes ranker with cost sensitivities for age estimation," in *CVPR*. IEEE, 2011, pp. 585–592.

[26] I. Huerta, C. Fernández, and A. Prati, "Facial age estimation through the fusion of texture and local appearance descriptor," in *Soft Biometrics in conjunction with ECCV, in press*. IEEE, 2014.

[27] C. Fernández, I. Huerta, and A. Prati, "A comparative evaluation of regression learning algorithms for facial age estimation," in *FFER in conjunction with ICPR, in press*. IEEE, 2014.

[28] M. Yang, S. Zhu, F. Lv, and K. Yu, "Correspondence driven adaptation for human profile recognition," in *CVPR*. IEEE, 2011, pp. 505–512.

[29] A. Lanitis, "Age estimation based on head movements: A feasibility study," in *Communications, Control and Signal Processing (ISCCSP), 2010 4th International Symposium on*, March 2010, pp. 1–6.

[30] 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*, Oct 2011, pp. 1–6.

[31] B. Xia, B. Ben Amor, M. Daoudi, and H. Drira, "Can 3D Shape of the Face Reveal your Age?" in *International Conference on Computer Vision Theory and Applications*, Lisbonne, Portugal, Jan. 2014.

[32] M. Minear and D. C. Park, "A lifespan database of adult facial stimuli," *Behavior Research Methods, Instruments, & Computers*, vol. 36, no. 4, pp. 630–633, 2004.

[33] K. Ricanek and T. Tesafaye, "MORPH: a longitudinal image database of normal adult age-progression," in *Automatic Face and Gesture Recognition*, 2006, pp. 341–345.

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

[35] A. Gallagher and T. Chen, "Understanding images of groups of people," in *Proc. CVPR*, 2009.

[36] 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*. IEEE, 2005, pp. 947–954.

[37] P. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The {FERET} database and evaluation procedure for face-recognition algorithms," *Image and Vision Computing*, vol. 16, no. 5, pp. 295 – 306, 1998.

[38] A. Bastanfard, M. Nik, and M. Dehshibi, "Iranian face database with age, pose and expression," in *Machine Vision, 2007. ICMV 2007. International Conference on*, Dec 2007, pp. 50–55.

[39] 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*, May 2002, pp. 46–51.

[40] K. Ueki, T. Hayashida, and T. 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*, ser. FGR '06. Washington, DC, USA: IEEE Computer Society, 2006, pp. 43–48.

[41] D. M. Burt and D. 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*, vol. 259, no. 1355, pp. 137–143, 1995.

[42] "LHI image database," available at http://www.lotushill.org/LHIFrameEn.html, 2010.

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

[44] B. Ni, Z. Song, and S. Yan, "Web image mining towards universal age estimator," in *Proceedings of the 17th ACM International Conference on Multimedia*, ser. MM '09. New York, NY, USA: ACM, 2009, pp. 85–94.

[45] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," *Information Forensics and Security, IEEE Transactions on*, vol. 9, no. 12, pp. 2170–2179, Dec 2014.

[46] S. Ren, X. Cao, Y. Wei, and J. Sun, "Face alignment at 3000 fps via regressing local binary features," June 2014.

[47] D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, ser. CVPR '12. Washington, DC, USA: IEEE Computer Society, 2012, pp. 2879–2886.

[48] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang, "Interactive facial feature localization," in *Proceedings of the 12th European Conference on Computer Vision - Volume Part III*, ser. ECCV'12. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 679–692.

[49] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 faces in-the-wild challenge: The first facial landmark localization challenge," in *The IEEE International Conference on Computer Vision (ICCV) Workshops*, December 2013.

[50] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar, "Localizing parts of faces using a consensus of exemplars." in *CVPR*. IEEE, 2011, pp. 545–552.

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

[52] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, Jun. 2010.

[53] 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.

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

[55] 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.