GENDER RECOGNITION: METHODS, DATASETS AND RESULTS

Vito Santarcangelo[2], *Giovanni Maria Farinella*[1], *Sebastiano Battiato*[1]

[1]Department of Mathematics and Computer Science, University of Catania, Italy
[2]Centro Studi S.r.l., Buccino, Italy
Vito.Santarcangelo@centrostudi.biz, {gfarinella, battiato}@dmi.unict.it

ABSTRACT

Digital Out Of Home (DOOH) applications which exploit computer vision algorithms to automatically collect soft biometrics of people in front a smart screen are of great interest for industry. In the last years many gender recognition pipelines have been proposed in literature. Different benchmark datasets have been introduced and used for testing purpose. This paper gives an overview of the state-of-the-art in the context of gender recognition by highlighting features, classifiers and datasets which can be employed to reach the goal. Comparisons of the results obtained by different approaches are also presented.

Index Terms— Gender Recognition, Digital Out Of Home (DOOH), Digital Signage

1. INTRODUCTION AND MOTIVATION

Gender recognition is becoming a fundamental step in many human-interaction applications. In particular, gender recognition engines can be exploited in the context of Digital Out of Home (DOOH) to trace the revenue of an advertising campaign [8, 16, 17] and in all other applications based on customer behavior analysis. DOOH has received considerable interest from retailers and governmental institutions because of the benefits obtained in better manage and respond to the preferences of the users. As a result, Computer Vision related applications become more and more important to enrich digital signage experience [31]. Among others, applications such as face re-identification and age estimation have been discussed in [14] and [32]. In this work we focus on gender recognition as useful tool to improve the performances of active digital signage. Different gender recognition approaches nowadays available consider constrained datasets [4, 10, 15], so that in the real contexts can encounter some difficulties due to the variation of parameters such as illumination, pose, resolution, size and background. Approaches which consider unconstrained domain have been also presented in [1, 2, 3, 5]. In DOOH applications it is also useful to consider gender recognition on group of people. Thus the development of benchmark datasets of images with groups of people [9] have been also considered. An other important challenge for recognition is represented by the gender recognition of children [7] where the use of contextual informations can help [6].

In this study we compare recent approaches of gender recognition considering the related benchmark datasets, the features and the classification methods used, trying to highlight peculiarities and limitations of the different approaches. The aim is to present an up to date review of the current state of the art. Our work extend the analysis of [15] by including constrained and unconstrained face datasets, as well as a comparison of the performances obtained by different methods of the recent literature.

The paper is organized as follows: in Section 2 we describe the different datasets and the evaluation procedures employed in recent works. Section 3 presents the features descriptors and the classification methods in the state-of-the-art gender recognizer. Section 4 compares the different methods . Finally, in Section 5 we discuss open challenges about gender recognition in the context of DOOH.

2. DATASETS

To properly study the peculiarities of the different methods, in the various involved contexts, benchmark datasets have been introduced. Large dataset are usually required to properly test and measure the performances of an approach. Gender image datasets can be grouped in two main categories: constrained and unconstrained. Constrained datasets are mainly composed of faces usually used for biometrical application purpose. A constrained dataset is characterized by images with controlled poses of the acquired subjects and pre-defined scene conditions. Examples of constrained datasets are AR [20], Lab2 [21], FEI Database [22], FERET [23], PAL [24] as better detailed in the following.

- **AR dataset**
  It is characterized by over 3,000 color images of 116 people (63 men and 53 women) acquired at the Computer Vision Center of the U.A.B. under strictly controlled conditions with respect to the possible variabilities. The images are related to frontal view faces with different facial expressions, illumination conditions, and occlusions (e.g. sun glasses and scarf).

This dataset is public available at the following URI : http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html.

- **Lab2 dataset**
This dataset is characterized by visible light images and near-infrared images of 50 subjects (12 females and 38 males). For each subject there are 20 visible light face images and the same number of near-infrared face images. Variability is related to facial expression and pose. Moreover images were acquired under four different source point illumination conditions: frontal illumination, left illumination, right illumination, both left and right illumination. The dataset can be found at the following URI:
http://www.yongxu.org/databases.html.

- **FEI face database**
It is a Brazilian face database that contains 2800 images (14 images for each of the 200 individuals, 100 males and 100 females) which have been acquired with homogeneous background. The acquired images are characterized of profile rotation up to about 180 degrees. The dataset can be obtained at http://fei.edu.br/~cet/facedatabase.html.

- **FERET database**
FERET contains 14051 grayscale images of human faces with different views (frontal, left and right profiles). It represents one of the most known and used dataset for face recognition purposes. Information on how to obtain the dataset are available at the following website : http://www.itl.nist.gov/iad/humanid/feret/feret_master.html.

- **PAL dataset**
The PAL DB is characterized by 575 face images of adults ranging from 19 to 93 years of age (225 males and 350 female). The official website of this dataset is https://pal.utdallas.edu/facedb/.

Despite constrained dataset are frequently used in literature, in real application domain (such as DOOH) the images to be analyzed are taken in unconstrained settings. Proposed approaches could reach really good performances on constrained datasets but could not have the same accuracy on image acquired on real life. So, in the recent years unconstrained datasets have been introduced. Unconstrained datasets are built considering images acquired by real life, with different poses and scene conditions. These datasets are principally built collecting images from public repository (e.g., Flickr) with the use of web image crawlers. The unconstrained face datasets currently used in literature are LFW [13], Gallagher [25], Genki-4K [28], Image of Groups [26], KinFace [27]. The main characteristics of these datasets are reported in the following.

- **Labeled Faces in the Wild (LFW)**
It can be considered the most important unconstrained face dataset up today. It contains more than 13000 images of faces collected from the web (10256 male and 2977 female images). Each face has been labeled with the name of the person pictured. 1680 of the people pictured have two or more distinct photos in the data set. The official link to obtain the LFW dataset is http://vis-www.cs.umass.edu/lfw/.

- **Gallagher dataset**
It is a collection of 931 real life digital images of people. The dataset can be downloaded from the authors website: http://chenlab.ece.cornell.edu/people/Andy/GallagherDataset.html.

- **GENKI-4K dataset**
It contains over 3,000 color face images labeled as either smiling or non-smiling by human coders (1539 females and 1506 males extracted by Danisman et al. [1]). The images contains faces spanning a wide range of illumination conditions, geographical locations, personal identity, and ethnicity. GENKI-4K can be downloaded at the link http://mplab.ucsd.edu/wordpress/?page_id=398 .

- **Image of Groups**
This is a dataset of 5080 images containing groups of people. It contains 28231 faces labeled with age and gender. This is a useful dataset for the studies of groups of people in unconstrained settings. It can be downloaded at http://chenlab.ece.cornell.edu/people/Andy/ImagesOfGroups.html.

- **UB KinFace**
It comprises 600 images related to 400 people (most of them are real-world collections of public figures as celebrities and politicians from Internet). It can be organized into 200 groups (child, young parent and old parent images). The dataset is available online at : http://www1.ece.neu.edu/~yunfu/research/Kinface/Kinface.htm.

Table 1 summarizes the main characteristics of the datasets and, whenever it is available, the number of male/female subjects.

In Fig.1 the datasets are organized in a bipartite graph (constrained and unconstrained datasets) highlighting the correspondences with the related references. In the subsequent analysis it is possible to compare the performances of different algorithms which have been exploited the same datasets.

Table 1. Face Datasets for Gender Recognition

NAME	TYPE	Images	Male Faces	Female Faces
AR	Constrained	3016	1638	1378
Lab2	Constrained	2000	1520	480
FEI	Constrained	2800	1400	1400
FERET	Constrained	14051	nd	nd
PAL	Constrained	575	225	350
Gallagher	Uncostrained	931	nd	nd
LFW	Uncostrained.	13233	10256	2977
GENKI-4K	Uncostrained	3045	1506	1539
I.Groups	Uncostrained	5080	10303	9532
UB KinFace	Uncostrained	600	440	160

3. GENDER RECOGNITION APPROACHES

Gender recognition approaches share a pipeline composed by three main modules (see Fig.2): face detection, face representation and gender classification. In this Section we detail the components of a gender recognition method by considering the current state-of-the-art.

3.1. FACE DETECTION

As first stage of a gender recognition engine, the faces of the people present into the acquired images have to be detected. To this purpose the well-known Viola and Jones object detection framework is usually used [18]. It is able of processing images in real time achieving high detection rates. The solution exploits the `Integral Imageërepresentation so that Haar-like features can be computed at any scale or location in constant time. A learning algorithm based on AdaBoost [19] is used to select the most discriminative features bases for classification purposes. Combining different classifiers in a cascade the background regions of the image are discarded while faces are detected.

3.2. FACE REPRESENTATION

When faces are detected a discriminative representation have to be computed. Usually a pre-processing step to remote geometric variabilities (e.g., by aligning faces with respect to the eyes positions) as well as photometric variabilities (e.g., light conditions) is perfomed [14]. The obtained images are hence processed to extract features to be used into the gender classification module. Different features methods have been used for gender recognition. Fig.3 shows the features which have used in the gender recognition systems. Almost all approaches for gender recognition are based on LBP, HOG, SIFT features. These features can be combined also with Color Histograms (CH) and Gabor features to improve the performances. It is useful to synthesize these approaches to better understand their application on gender recognition.

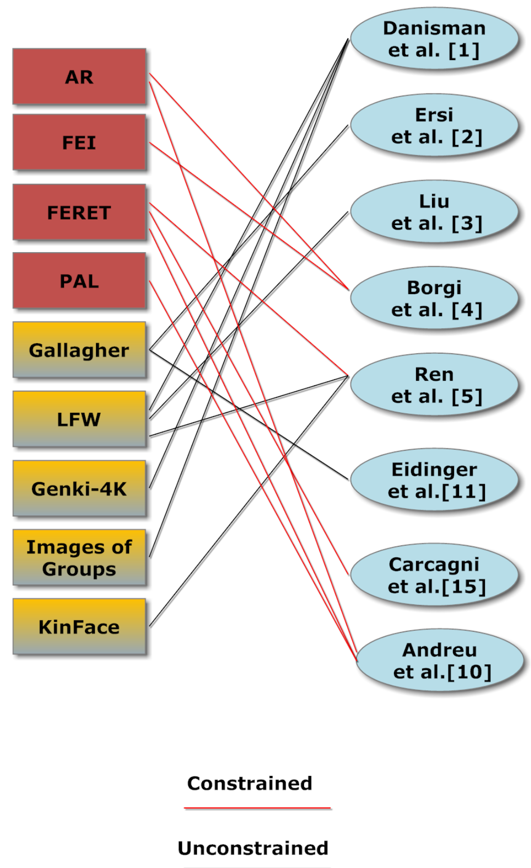

Fig. 1. FACE DATASETS AND METHODS

- **LBP descriptor**
  It considers spatial comparison of local neighborhood of a pixel and creates labels which are then aggregated in histograms. LBP descriptors (and the related variants) are robust to illumination and to the rotation variations. LBP are useful to capture textures (e.g., beard).

- **SIFT descriptor**
  The SIFT descriptor is invariant to scale, translation, rotation, partially invariant to illumination changes and robust to local geometric distortion.

- **HOG method**
  It considers the image divided into a block structure cell-based. The block is characterized by the histogram of oriented gradients as a function of the edges. Histograms are subjected to normalization of contrast.

LBP, SIFT and HOG approaches have been also combined with the CH and Gabor features for gender recognition [2, 5]. The histogram of the colors (CH) extracts the characteristics on the distribution of the ”colors” of the image, instead, Gabor filters are used to encode texture. Once the features are

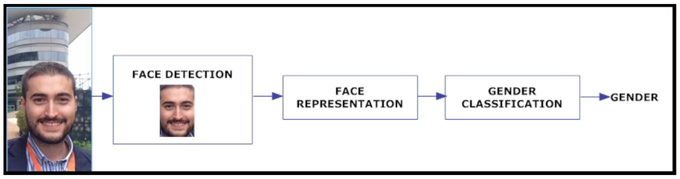

Fig. 2. GENERAL PIPELINE OF A GENDER RECOGNIZER

extracted and the image is represented, a classifier is applied.

3.3. GENDER CLASSIFICATION

The most used methods for gender classification are SVM (with RBF Kernel) and Adaboost combined with Linear SVM. A Support vector machine (SVM) [30] is a powerful classifier for two classes based problems: gender recognition problem (male vs female) is a good application scenario. The Adaptative boosting approach (Adaboost) [29] is an ensemble learning based method able to build a strong classifier from a combination of weak classifiers.
Fig.3 shows different features and classifiers which have been used in literature for gender classification.

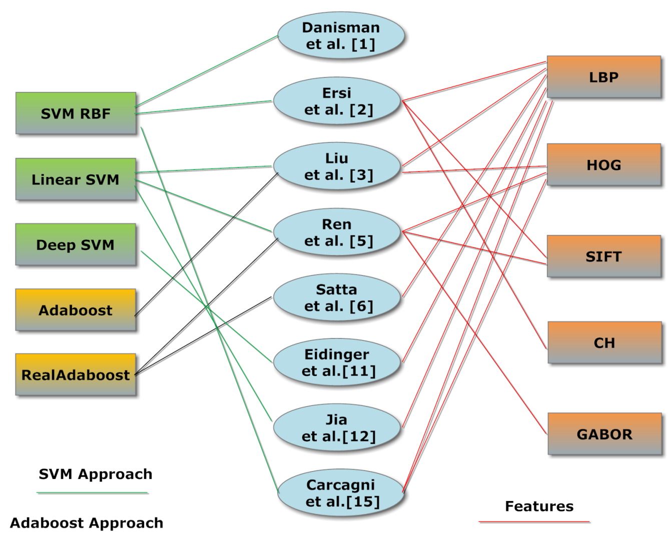

Fig. 3. FEATURES AND CLASSIFIERS

4. STATE OF THE ART RESULTS

In this section we review the results obtained by the current state-of-the-art approaches. For each approach we briefly describe the main ingredients used to build the gender recognizer, which dataset have been used in the experimental phase and well and the obtained results. The comparison among the different approaches is obtained by considering the results claimed by the authors in their papers. In this way approaches which used the same dataset in the experiments are straightforward compared. The comparison is reported in Table 2.

- **Danisman et al. [1]**
  This approach shows as the use of a pre-processing step can improve the performances of gender recognition on unconstrained datasets. The method proposed by the authors is composed by face detector, a face alignment step to remove geometric variability, and a histogram equalization in which the face images are normalized with respect to a specific probability density function obtained considering the average face of the training dataset to remove illumination variability. Classification is obtained exploiting the SVM classifier with RBF kernel. This method has been tested on unconstrained datasets. The accuracy obtained on the LFW, Genki-4k and Groups datasets are respectively 91.87%, 91.07%, 88.16%.

- **Ersi et al. [2]**
  The gender classification approach is based on the combination of LBP, SIFT and CH descriptors. Classification is obtained by exploiting a SVM with RBF kernel. The accuracy obtained on the Gallagher database (unconstrained images from the web) is 91.6%.

- **Liu et al. in [3]**
  The authors introduce a new feature called Self-Similarity of Gradients (GSS) which captures pairwise statistics of localized histogram of gradient distributions. For classification purpose both Adaboost and SVM are compared indipendendly or in cascade (boosting for feature selection and SVM for classification). The performances reached on LFW face dataset by considering only the GSS descriptor are 88.96%, whereas the combination of HOG31, LBP and GSS features achieved an accuracy of 95.76%.

- **Borgi et al. in [4]**
  This work proposes a new approach for gender classification called multi-regularized learning (MRL). It considers as first step a dimensional reduction of the faces feature space. Then the proposed multi-regularization feature learning approach is applied for classification purpose. The method obtained 92.83% of accuracy on the AR dataset, whereas an accuracy of 94% is reached on the FEI dataset.

- **Ren et al. in [5]**
  This work considers a combination of the SIFT, HOG and Gabor filters as final descriptor for gender recognition. The classification is obtained through RealAdaboost with the use of a penalty term that considers the complexity of the feature combination. The combinations of the feature spaces with the penalty term reduces

Table 2. Performances

	Approaches							
	[1]	[2]	[3]	[4]	[5]	[11]	[12]	[15]
Groups	88.16							
Genki-4K	91.07							
LFW	91.87		95.80		98.00		96.86	
Gallagher		91.60				88.60		
AR				92.83				
FEI				94.00				
FERET					98.78			
KinFace					96.50			
Feret+Morph								89.70

the computational complexity. The approach has been tested on both constrained and unconstrained datasets obtaining the following results: 98.78% on FERET, 96.50% on KINFACE, 98.01% on LFW.

- **Eidinger et al. [11]**
  The authors presented a pipeline based on with four steps: detection, alignment, representation and classification. The detection is obtained by the use of Viola and Jones face detector, the alignment method is done considering the position of 68 specific facial features, the final classification is obtained by the use of a Dropout-SVM on LBP and FPLBP representation. The results of the method on Gallagher Dataset is of 88.6%.

- **Jia et al. [12]**
  This contribution address the important challeng of training the gender classifier by considering a big dataset represented in a high dimensional feature space (four million images and 60 thousand features). The proposed approach use an ensemble of linear classifiers, and achieve an accuracy of 96.86% on the most challenging public database, Labelled Faces in the Wild (LFW).

- **Carcagni et al. [15]**
  The paper presents a comparison among LBP, HOG and SWLD (Spatial Weber Local Descriptor) descriptors for gender recognition on constrained dataset. Specifically, the authors have used a fusion of FERET and MORPH datasets for testing purposes. The best accuracy is obtained by using HOG descriptors coupled with SVM and exploiting RBF Kernel (89,70%).

To complete the analysis, it is important to consider also the impact of the image resolution regarding the performances obtained by a gender classifier. Andreu et al. [10] performed tests on different dataset (FERET, PAL, AR) considering them at different resolution. The study shows that a size between 22x18 and 90x72 pixels is recommended for the problem of gender recognition. The authors pointed out that a size of 45x36 pixels provides enough information to infer the gender recognition from images.

The results reported above show that the best performances are obtained on constrained datasets (98.78%) [5]. Considering unconstrained datasets, the feature descriptor better performing is obtained by the combination of HOG, SIFT and Gabor descriptors [5] (98.01%). It is important to notice that some datasets (e.g., the one with images of groups) are more complex than others in gender recognition, due to the fact of the presence of groups of people, children, different ethnic groups and a large age gap.

5. CONCLUSION AND FUTURE WORKS

This paper discussed the problem of gender recognition. After a revision of the state-of-the-art approaches and the benchmark datasets available for testing purposes, a comparison of the performances has been done. Future works could be devoted on building a benchmark dataset by considering the needs of the DOOH application domain as well as considering multi-spectral Imaging (i.e., standard light, together with cross polarized flash and ultraviolet light) and 3D information to better determine the gender of a person from its face. Moreover, the combination of methods for age estimation and ethnicity identification could be a route to be explored to improve the results.

6. REFERENCES

[1] T. Danisman, I. M. Bilasco, C. Djeraba, Cross-database evaluation on normalized raw pixels for gender recognition under unconstrained settings, International Conference on Pattern Recognition (2014)

[2] F. Ersi, M. Pasandi, R. Laganiere, Age and gender recognition using informative features of various types, International Conference on Image Processing (2014)

[3] H. Liu, Y. Gao, C. Wang, Gender identification in unconstrained scenarios using self-similarity of gradients features, International Conference on Image Processing (2014)

[4] M.Borgi, M.ElArbi, C. BenAmar, D. Labate, Face,gender and race classification using multi-regularized features learning, International Conference on Image Processing (2014)

[5] H. Ren, Z. Li, Gender recognition using complety-aware local features, International Conference on Pattern Recognition (2014)

[6] R. Satta, J. Galbally, L. Beslay, Children gender recognition under unconstrained conditions based on contextual information, International Conference on Pattern Recognition (2014)

[7] Napa Sae-Bae, Xiaoxi Sun, Husrev T. Sencar, Nasir D. Memon, Towards automatic detection of child pornography, International Conference on Image Processing (2014)

[8] R. Ravnik, F. Solina, Interactive and Audience Adaptive Digital Signage Using Real-Time Computer Vision, International Journal of Advanced Robotic Systems (2013)

[9] Yi-I Chiu, Ru-Yun Hsu, Chun-Rong Huang, Spatial face context with gender information for group photo similarity assessment, International Conference on Pattern Recognition (2014)

[10] Y. Andreu, J. Lopez-Centelles, R. Mollineda and P. Garca-Sevilla, Analysis of the effect of image resolution on automatic face gender classification, International Conference on Pattern Recognition (2014)

[11] E. Eidinger, R. Enbar, T. Hassner, Age and gender estimation of unfiltered faces, IEEE Transactions on Information Forensics and Security (2014)

[12] S. Jia, N. Cristianini, Learning to classify gender from four million images, Pattern Recognition Letters, VOL 58 NO.1 35-41 (2015)

[13] G. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled faces in the wild: a database for studying face recognition in unconstrained environments, TECHNICAL REPORT (2008)

[14] G. M. Farinella, G. Farioli, S. Battiato, S. Leonardi, G. Gallo, Face Re-Identification for Digital Signage Applications, VAAM (2014)

[15] P. Carcagni, M. Del Coco, P. Mazzeo, A. Testa, C. Distante, Features descriptors for demographic estimation: a comparative study, VAAM (2014)

[16] B. Batagelj, R. Ravnik, F. Solina, Computer vision and digital signage, International Conference on Multimodal Interfaces (2008)

[17] J. Muller, J. Exeler, M. Buzeck, A. Kruger, ReflectiveSigns: digital signs thatadapt to audience attention, LNCS, Vol. 5538, pp. 1724. Springer, Heidelberg (2009)

[18] P. Viola, M. Jones, Robust Real-time Object Detection,International Journal of Computer Vision, Vol. 57 No.2, 137-154 (2004)

[19] R. E. Shapire , A brief introduction to boosting, International Joint Conference on Artificial intelligence (1999)

[20] A. M. Martinez, R. Benavente, The AR Face Database. CVC Technical Report 24 (1998)

[21] J. Wen, Y. Xu, J. Tang, Y. Zhan, Z. Lai, Joint video frame set division and low-rank decomposition for background subtraction, IEEE Transactions on Circuits and Systems for Video Technology (2014)

[22] C. E. Thomaz, G. A. Giraldi, A new ranking method for Principal Components Analysis and its application to face image analysis, Image and Vision Computing, VOL. 28, NO. 6, pp. 902-913 (2010)

[23] P. J. Phillips, H. Moon, P. J. Rauss, S. Rizvi, The FERET evaluation methodology for face recognition algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 10 (2000)

[24] M. Minear, D. C. Park, A lifespan database of adult facial stimuli. Behavior Research Methods, Instruments Computers. 36, 630-633 (2004)

[25] A. Gallagher, T. Chen , Clothing Cosegmentation for Recognizing People, IEEE Conference on Computer Vision and Pattern Recognition (2008)

[26] A. Gallagher, T. Chen , Understanding Groups of Images of People, IEEE Conference on Computer Vision and Pattern Recognition (2009)

[27] S. Ming, X. Siyu, Y. Fu, Genealogical Face Recognition based on UB KinFace Database, IEEE CVPR Workshop on Biometrics (2011)

[28] The MPLab GENKI Database, GENKI-4K Subset, http://mplab.ucsd.edu (2011)

[29] Y. Freund, R. Schapire, A Short Introduction to Boosting, Journal of Japanese Society for Artificial Intelligence (1999)

[30] B. Boser, I. Guyon, V. Vapnik, A Training Algorithm for Optimal Margin Classifiers, Proceedings of the fifth annual workshop on Computational learning theory (1992)

[31] C. Distante, S. Battiato, A. Cavallaro, Video Analytics for Audience Measurement,VAAM 2014, 123 LNCS 8811, Springer (2014)

[32] A. Torrisi, G.M. Farinella, G. Puglisi, S. Battiato, Selecting discriminative CLBP patterns for age estimation, Workshop on visual analytics for audience measurement in retail and digital signage (2015)