

Robust gender recognition by exploiting facial attributes dependencies

Juan Bekios-Calfa^a, José M. Buenaposada^b, Luis Baumela^b

^a*Dept. de Ingeniería de Sistemas y Computación, Universidad Católica del Norte
Av. Angamos 0610, Antofagasta, Chile
juan.bekios@ucn.cl*

^b*Dept. de Ciencias de la Computación, Universidad Rey Juan Carlos
Calle Tulipán s/n, 28933, Móstoles, Spain
josemiguel.buenaposada@urjc.es*

^c*Dept. de Inteligencia Artificial, Universidad Politécnica de Madrid
Campus Montegancedo s/n, 28660 Boadilla del Monte, Spain
lbaumela@fi.upm.es*

Abstract

Estimating human face gender from images is a problem that has been extensively studied because of its relevant applications. Recent works report significant drops in performance for state-of-the-art gender classifiers when evaluated “in the wild,” i.e. with uncontrolled demography and environmental conditions. We hypothesize that this is caused by the existence of dependencies among facial demographic attributes that have not been considered when building the classifier. In the paper we study the dependencies among gender, age and pose facial attributes. By considering the relation between gender and pose attributes we also avoid the use of computationally expensive and fragile face alignment procedures. In the experiments we confirm the existence of dependencies among gender, age and pose facial attributes and prove that we can improve the performance and robustness of gender classifiers by exploiting these dependencies.

Keywords: gender estimation, facial attribute dependencies

1. Introduction

Visual attributes are properties observable in images that have a semantic meaning, for example *has clothes, furry, male, young*. Attribute-based representations have recently received much attention because they have been

successfully used for image retrieval (Yu et al., 2012), for recognizing objects (Duan et al., 2012; Wang et al., 2009), for describing unknown objects (Farhadi et al., 2009), and even for learning new unseen object models from descriptions (Farhadi et al., 2009; Lampert et al., 2009). Facial attributes have a key role in human-computer interaction applications, image and video retrieval and surveillance. They have also been successfully used for facial verification (Kumar et al., 2009). There is a plethora of interesting facial attributes such as hairstyle, hair color, facial expression, etc. However, the main facial demographic attributes are gender, ethnicity and age. In this paper we will consider the gender attribute and its relation with age and pose.

Gender is perhaps the most widely studied facial demographic attribute in the Computer Vision field (Moghaddam and Yang, 2002; Baluja and Rowley, 2007; Mäkinen and Raisamo, 2008; Bekios-Calfa et al., 2011). The state-of-the-art recognition rate for the Color FERET database (Phillips et al., 2000) involving frontal faces with frontal illumination and 5 fold cross-validation is around 93% using either a Support Vector Machine with Radial Basis function (Moghaddam and Yang, 2002), pair-wise comparison of pixel values within a boosting framework (Baluja and Rowley, 2007) or linear discriminant techniques (Bekios-Calfa et al., 2011). This performance drops significantly if classifiers are trained and tested on different databases. For example, if we train our classifier with the FERET database and test it with images from PAL (Minear and Park, 2004), the performance drops to roughly 70% success rate (Bekios-Calfa et al., 2011). This is mainly due to the different demographic distributions in both databases. FERET is a database with mostly Caucasian adult subjects, whereas PAL includes people from a broader range of ethnic groups and ages. In general, when a gender classifier is trained with a data set with limited demography and tested with a data set with more general samples the classification rate drops significantly. This suggests the existence of a dependency between gender and other demographic variables.

Although the Color FERET database has often been used as a benchmark for gender estimation, it can hardly predict the performance of a gender classifier in a real setting, since it was acquired in laboratory conditions. In recent years there is a trend to use face databases acquired “in the wild,” i.e. in real settings. *The Images of Groups Dataset*¹ (GROUPS) (Gallagher

¹<http://chenlab.ece.cornell.edu/people/Andy/ImagesOfGroups.html>

and Chen, 2009) is a large database acquired from internet images of groups of people and labeled with age and gender data. Training with 23,218 eye-aligned images and testing with 1,881 images from this database Gallagher *et al.* (Gallagher and Chen, 2009) achieved 69.6% accuracy using 7 Linear Discriminant Analysis (LDA) projections (one per age range) and a K-Nearest Neighbor classifier (K-NN). Even when (Gallagher and Chen, 2009) employed aligned images using the eye positions, the classification is far from being perfect. The reason is that images in GROUPS represent real-world situations: different illumination conditions, facial expressions, face poses and broad demographic distribution of faces. When estimating facial attributes using real-world images the intra-class variability is usually larger than inter-class one. The changes in appearance produced by the the head pose, illumination or facial expressions make the facial appearance change dramatically. By removing the intra-class variability the recognition performance can be improved significantly.

In this paper we hypothesize that face pose variations and age are sources of intra-class variability for gender estimation. We will exploit the appearance-based dependencies among gender, age and pose to remove the intra-class variability and improve the recognition performance.

Previous gender classification in real-world settings align the face image to a canonical pose (Mäkinen and Raisamo, 2008; Gallagher and Chen, 2009). Face alignment requires the automated or manual detection of fiducial points (Mäkinen and Raisamo, 2008; Gallagher and Chen, 2009) or a *congealling* previous step (aligning all the images in a set by reducing entropy (Learned-Miller, 2006)). In their work, (Mäkinen and Raisamo, 2008) report that, although manual face alignment does increase gender classification rates, the performance improvements achieved by automatic methods are not significant. This is caused by the lack of reliability of automated approaches. So, alignment-based solutions are both costly in terms of computational resources and fragile because they are not robust and often reach an incorrect alignment.

In this paper, by considering the pose attribute in the classification, we transfer the alignment problem to the learning phase, removing the need for alignment. This will increase both the efficiency and robustness of the classification. Although there are previous works that simultaneously learn to align and classify (Babenko *et al.*, 2008; Kim *et al.*, 2010), in our work we follow a different path. We find clusters of face poses after face detection and use them in training. Then we train a classifier that simultaneously predicts

pose and gender. We test our procedure in the GROUPS database achieving an increase of about 5% in the performance of a pose-aware classifier compared to a standard state-of-the-art results..

To address the influence of age in gender estimation, we also study the relation between gender and age attributes. Dependencies among demographic variables have also been previously considered in the literature. Wei and Ai (Wei et al., 2009) showed experimentally that by exploiting the relation between ethnicity and gender a boost of 4-5% in gender classification accuracy can be obtained for mongoloid and African faces. Guo and Mu (Guo and Mu, 2010), in experiments with the MORPH-II database, found that age estimation can have large errors if the influence of gender and ethnicity is not taken into account. Finally, Guo et al. (Guo et al., 2009) considered the dependencies between age and gender. They found that gender recognition accuracy was 10% higher in adult faces than in young and senior faces and studied the influence of different image features (LBP, HOG and BIF). In this paper we also consider the influence of age in the estimation of gender, but from a completely different perspective. We will study whether the accuracy in gender can be improved by jointly estimating age and gender.

The organization of the rest of the paper is as follows. In section 2 we introduce a classification procedure that combines several facial attributes. In section 3 we analyze the dependencies of gender with age and pose. Finally in sections 4 and 5 we validate experimentally the hypothesis in the paper and draw conclusions.

2. The Facial Attributes Powerset for classification

One of the baseline approaches to multi-label classification is a problem transformation method: Label Powerset (LP) (Tsoumakas et al., 2011). The LP approach explores all possible label combinations. LP interprets every possible subset of labels (a combination of labels) appearing on the multi-label training set as a single class in a multi-class problem.

We can adapt the LP idea to our problem, denoting it Attributes Powerset (AP). Let $T = \{a_1, \dots, a_N\}$ be the set of N facial attributes of a given problem and let $V_i = \{c_1, \dots, c_{M_i}\}$ be the set of possible values for attribute a_i where $M_i = |V_i|$. Let $C_\times = V_1 \times V_2 \times \dots \times V_d$ be the Cartesian product of all V_i attributes values sets. The output of an AP classifier for an input instance, \mathbf{x} , is a vector $\mathbf{z} \in C_\times$. For example, in the facial demographic attribute classification problem, one of the possible formulations could be

to have two attributes, $T = \{age, gender\}$ and the corresponding values $V_{age} = \{young, adult, senior\}$, $V_{gender} = \{male, female\}$ In this case the AP is given by: $C_{\times} = \{(young, male), (young, female), \dots, (senior, female)\}$, where powerset cardinality is $|C_{\times}| = 6$.

In order to get well separated and compact subclasses we perform Fisher Linear Discriminant Analysis (LDA) dimensionality reduction of the training data. Our approach for dimensionality reduction is related to Subclass Discriminant Analysis (SDA) (Zhu and Martinez, 2006). However, instead of finding subclasses with a clustering procedure, we find the subclasses with the facial attributes labels. If the number of training data is low or the number of classes arising from the powerset is high, the results from LDA can be seriously compromised (Bekios-Calfa et al., 2011). In this case we use Principal Component Analysis (PCA) with cross-validation to reduce the dimensionality of the data prior to performing LDA (Bekios-Calfa et al., 2011).

3. Robust gender recognition

Age and face pose variations are sources of intra-class variability that reduce the performance of gender classifiers. In this section we will analyze the relation of gender with age and pose attributes.

3.1. On the dependence between age and gender attributes

Age and gender variables are statistically independent. If we consider that in any age range there is equal number of men and women, and for any gender the distribution of people in ages is similar, then we can conclude that age and gender demographic variables are statistically independent. That is, $P(A, G) = P(A)P(G)$, were A and G denote respectively age and gender variables and P the probability of an event.

However, from our intuition we know that there is a conditional dependence between gender and age given the facial appearance, since, from a face image, the gender of small children is almost indistinguishable, whereas it can be easily established for adult faces. To confirm this dependence we have trained a state-of-the-art gender classifier (Bekios-Calfa et al., 2011) with the GROUPS database, and tested it with PAL (see Section 4 for a description of these databases) . The classifier was trained with the images in GROUPS for all age ranges. In the first row of Table 1 we show the performances stratified into four age groups. Gender estimation results for the age

Table 1: Gender and age dependence. We separate the images in 4 age ranges. Each row shows the result of a different training process. In the first row the gender classifier is trained using all age ranges. In the second row the gender classifier is trained independently in each age range. The first four columns display the results stratified by age range. The last column shows average results for all age ranges.

Experiment/Age category	13-19	20-36	37-65	66+	Global
Gender	65.62	75.56	65.04	64.53	68.73
Gender Age	65.62	76.47	73.98	74.87	74.78

range 20–36 are clearly above those in the Global column, that represent the state-of-the-art. The results for the other age ranges are below this value. These results confirm that the performance of the classifier depends on the age range. Moreover, to confirm the existence of a dependency between age and gender we have trained four gender classifiers, one per age range, with GROUPS and tested them with data from the same age range in PAL. This experiment provides information about the performance of a gender classifier that knows the appearance and the age range of the subject. As we can see in the second row in Table 1, the difference in performance increases with the age range. This experiment clearly shows the existence of a dependency between age, A , and gender, G , given the facial appearance, represented by the classifier discriminating variables X .

In summary, if the performance of an appearance-based gender classifier changes when we provide it with information about age, then $P(G|X) \neq P(G|X, A)$. This implies that the gender, G , and age, A , demographic attributes are conditionally dependent given the appearance of the face, X .

3.2. On the dependence between gender and pose attributes

In this section we are interested in dealing with face alignment variations within a face detection window. Face detectors are usually trained to be rather insensitive to face pose (see Fig. 2 for a sample of detected faces with different poses). This feature of detection algorithms makes facial attribute estimation more difficult. To make gender classifiers robust to face pose, images are aligned to some canonical pose or configuration (Mäkinen and Raisamo, 2008; Gallagher and Chen, 2009). This procedure is computationally very intensive, prone to errors, and, as reported by Mäkinen and Raisamo (Mäkinen and Raisamo, 2008), it does not increase significantly the performance of gender classifiers. Our proposal in this section is to transfer

the alignment problem to the learning phase, avoiding the need for on-line alignment. To this end, we will find clusters of face poses after face detection and use them to train a classifier that recognizes both gender and face pose.

We will use face detection output as the only alignment procedure. Our goal then is to model and learn the miss-alignments produced by the face detector. To this end we exploit the fact that GROUPS provides the eyes co-ordinates for all faces in the database. We then set the parameters of our face detector to a very low number of false negatives, which indeed increases the number of false positives. False detections are subsequently removed using the ground truth eye positions. The result from this process is a set of 22,948 correctly detected faces (see Fig. 2): 11,932 female and 11,016 male.

We learn the classes of face miss-alignment by clustering eye poses. To this end, we define a canonical image with size 60×60 pixels. The ground truth eye coordinates are expressed relative to the top-left corner of the face detection window and expressed relative to the canonical image (see Fig.1). For clustering we use the feature vector $(x_{le}, y_{le}, x_{re}, y_{re})$ where le stands for canonical image left eye coordinates and re stands for canonical image right eye coordinates. We find the clusters using a K -means algorithm. We chose a value of $K=6$ because it provides a well differentiated set of clusters and the distribution of samples (see Table 2) is not too unbalanced, so that learning would not be biased by the amount of data per class. In Fig. 1 we show the resulting eye clusters.

Table 2: Number of samples per each Gender and Pose cluster combination.

Gender/Pose	0	1	2	3	4	5
Female	1359	2561	2677	1627	1599	2109
Male	1987	2565	1713	1660	1526	1565

If we assume that for any eye pose, L , there is equal number of men and women and for any gender, G , the distribution of eye poses is similar, then we are implicitly assuming that eye poses and gender are statistically independent. That is, $P(L, G) = P(L)P(G)$.

However, to assess whether L and G are conditionally dependent given the facial appearance, X , we have tested a state-of-the-art gender classifier (Bekios-Calfa et al., 2011) using 5-fold cross-validation with the 22,948 images from the GROUPS database used to find the clusters. In the first row of Table 3 we show the results of this classifier stratified by eye pose. Various

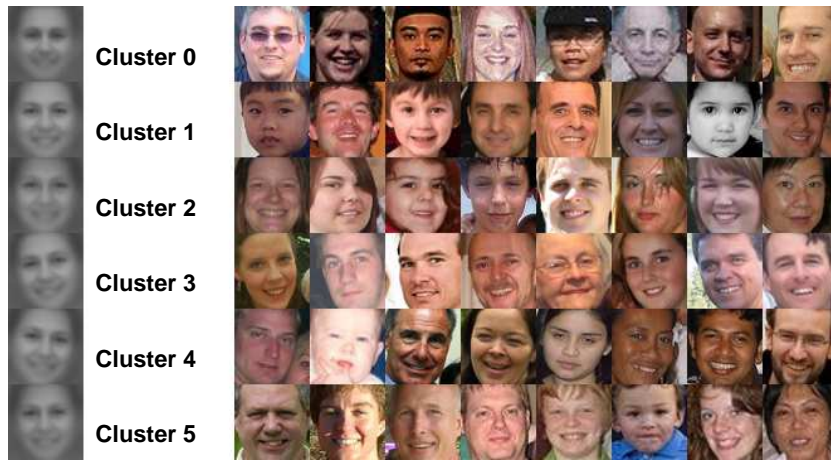


Figure 1: Clusters of eyes positions within the canonical face detection window. In the left most column we show the mean of each cluster. In the other columns we show the nearest images to the cluster mean on each of the clusters.

poses provide different classification results. To confirm the existence of a dependency between pose and gender we have now tested, using the same procedure, seven gender classifiers, one per pose cluster. This experiment provides information about the performance of a gender classifier that knows the appearance and the pose of the subject. As we can see in the second row in Table 3, the poses with the largest appearance change have the best improvement in performance. This experiment clearly shows the existence of a dependency between pose, L , and gender, G , given the facial appearance, represented by the classifier discriminating variables X .

4. Experiments

Our aim in this section is to show that by exploiting the dependencies of gender with age and eye pose attributes, we can improve the gender estimation accuracy in real-world problems. To this end we evaluate the performance of gender estimation within the AP method in different estimation problems involving age and eye position attributes. In all the experiments, we crop and re-size images to a base shape of 25×25 pixels using OpenCV's² Haar-like-wavelets face detector, which is based on (Viola and Jones, 2004)

²<http://opencv.willowgarage.com>

Table 3: Gender and pose dependence. 5-fold cross-validation with GROUPS for six different face detection miss-alignments. Each row shows the results of a different test. In the first row the gender classifier is evaluated using images from all six miss-alignments. In the second row the gender classifier is evaluated independently for each miss-alignment group. The first six columns display the results of the six miss-alignments classes. The last column shows the average accuracy across all miss-alignments. Below each performance result we show the cross-validation standard deviations preceded by the \pm symbol.

Experiment/Pose	0	1	2	3	4	5	Global
Gender	73.40	78.22	76.65	74.20	73.05	78.34	76.02
	± 0.77	± 0.79	± 1.89	± 1.41	± 1.62	± 0.68	± 0.38
Gender Pose	76.77	78.46	76.26	78.94	77.66	77.40	77.58
	± 1.41	± 1.31	± 1.51	± 1.27	± 0.73	± 0.89	± 1.21

work.

4.1. Face databases

Given that there is actually no single public database with all facial attributes we need, we have used various databases for different problems:

- **Gender and Age estimation in controlled conditions.** For this problem we have used the PAL database (Minear and Park, 2004). It consists of frontal pictures of 576 individuals. There is only one frontal face image per subject, although 3 individuals have two pictures in the database. Therefore, in our PAL experiments, we use 576 images, 219 male and 357 female subjects. The age is also available for each image. In our experiments we only use one frontal image per subject. See some sample face detections in Fig. 2.
- **Gender and Age estimation in the wild.** For this problem we use the GROUPS database (Gallagher and Chen, 2009)³ It consists of 28,231 faces, labeled with gender and age category (0-2, 3-7, 8-12, 13-19, 20-36, 37-65, and 66+), extracted from 5,800 pictures of people groups. Most faces were automatically detected. See Fig. 2 for some examples of face detections from this database. GROUPS is a real-world database involving different illumination conditions, facial expressions, face poses and a broad demographic distribution of faces.

³<http://chenlab.ece.cornell.edu/people/Andy/ImagesOfGroups.html>

- **Gender classification in the wild.** For this task we can also use the “Labeled Faces in the Wild” (LFW) database (Huang et al., 2007), that has been proposed as a benchmark for gender recognition⁴. It consists of 13,233 face images of 5,749 subjects (4263 male, 1486 female) obtained from the web (see Fig. 2). The age distribution is quite limited mostly consisting of middle-aged adults. Images are also gender imbalanced, 10256 male and 2977 female samples. We use the LFW dataset to make cross-database experiments, training with the GROUPS and testing with LFW, since both databases have challenging real-world imaging conditions.



Figure 2: Example images. First and second rows, PAL Database. Third and fourth rows GROUPS database. Last two rows LFW database.

- **Gender and 3D pose classification.** For this problem we use the CMU Multi-PIE database (Gross et al., 2010). It is a database of facial images acquired in different illuminations and 3D poses. It contains 337 subjects, imaged under 15 view points and 19 illumination conditions in

⁴<http://fipa.cs.kit.edu/downloads/LFW-gender-folds.dat>

up to four recording sessions. This is the perfect database for testing the influence of 3D pose on gender recognition. In our experiments we use all the illuminations from the first recording session for each subject. We do not use the facial expressions images. See Fig. 3 for a sample of the images.



Figure 3: Sample faces after detection from the Multi-PIE database.

4.2. Gender and age estimation

We have performed two experiments, one using the *Gender* attribute alone (see Table 1) and another using the *Gender* \times *Age* AP.

Assuming that the subject age is known before estimating gender (as done in section 3.1) is not realistic, since age is usually unknown. For this reason we use the AP (see section 2) to simultaneously estimate gender and age. There are two gender values and seven age categories (0-2, 3-7, 8-12, 13-19, 20-36, 37-65, 66+) labeled in the GROUPS database. In consequence, there will be 14 classes in the AP classification problem. We use five-fold cross-validation to train the 14 classes K-NN classifier in the LDA-transformed space. We compute the global accuracy considering the proportion of data in each category. The performance of the *Gender* \times *Age* AP improves in 3% that of the state-of-the art Gender classifier (see Table 4 and Table 1). Concerning the results in Table 4, note that PAL has no faces in the 0 to 12 years range. Most interestingly, the AP approach outperforms the single attribute one for all age categories. The most difficult categories for the single attribute Gender classifier, 13-19 and 66+, are those for which the improvement is highest, 3.13% and 7.39% respectively. This shows that the AP procedure is able to exploit the attributes dependencies and improve the gender estimation accuracy.

Table 4: Gender×Age AP accuracy for the GROUPS/PAL experiment. The first four columns display the results stratified by age range. The last column shows average results for all age ranges.

Experiment/Age category	13-19	20-36	37-65	66+	Global
Gender×Age	68.75%	76.01%	65.85%	71.92%	72.01%

4.3. Unaligned gender classification

As we have discussed in section 3.1, there is a dependency between gender and pose, given the facial appearance. Here we will exploit those dependencies using the AP to improve gender estimation. Whenever we mention the GROUPS database in this section we actually refer to the subset of that database used for obtaining the pose clusters. We have performed two groups of experiments both for Gender alone and using the AP of *Gender×Pose*:

- **Intra-database experiments:** In this case we evaluate the classifiers with the GROUPS database using five-fold cross validation.
- **Cross-database experiments:** Here we train the classifiers with GROUPS and test gender recognition with the LFW database.

In section 3.2 we analyzed the results of part of our intra-database experiment (see Table 3). We have also trained the same classifier on the AP of gender and eye poses. In Table 5 we show the results of this experiment as *Gender×Pose*. The *Gender* row in Table 3 shows the results of a gender classifier trained on miss-aligned images obtained from raw face detection. We can see in Table 5 that considering face pose, i.e. using the AP, we can get an global improvement of about 2%. With almost frontal images (i.e. 0, 1, 2, miss-alignment poses) the improvement is modest, between 1% and 2%. However, we get the best improvement, between 3% and 4%, in the clusters corresponding to the faces with largest in-plane rotation. So, we are actually improving gender recognition for the toughest cases.

The performance of intra-database experiments tends to be optimistic because images in one database usually have similar demography and acquisition conditions (Bekios-Calfa et al., 2011). The cross-database experiments inform us of the generalisation capabilities of our classifier. For these tests we chose the LFW database because it is also a real-world database, however with limited age ranges (only adults). In the column *GROUPS/LFW*

Table 5: Gender×Pose AP accuracy for 5-fold cross-validation with GROUPS. The first six columns display the results of the six pose clusters. The last column shows the average gender accuracy. Standard deviations are shown below each data preceded by the symbol \pm .

Experiment/Pose	0	1	2	3	4	5	Global
Gender×Pose	75.70	79.92	77.10	77.73	77.02	78.90	77.89
	± 1.37	± 1.24	± 1.84	± 1.19	± 1.16	± 1.36	± 0.33

of Table 6, we show the accuracy obtained in this test. Here again the *Gender×Pose* classifier provides better results than the *Gender* alone one. This confirms that the previous results generalize well to other databases.

Table 6: Gender and Pose cross-database experiments, training with GROUPS and testing with LFW. The rows show the results of a Gender and a Gender×Pose AP classifiers. The third column shows the results when training without the children age ranges (0-2, 3-7 and 8-12).

Experiment/Data Bases	GROUPS/LFW	GROUPS/LFW (w/o children)
Gender	77.95%	78.33%
Gender×Pose	79.11%	79.53%

It is well known that gender recognition in children images is very difficult because of the little appearance differences between genders. Therefore, children faces can be seen as noise for facial gender recognition algorithms. In the last experiment we test our classifier without child faces. We have repeated the intra- and cross-database experiments training with GROUPS without the age ranges 0-2, 3-7 and 8-12 years old. Table 7 shows an improvement in gender recognition in more than 2% compared with the experiment using the complete database. On the other hand, the align-aware gender classifier improves the results of the one using gender alone by more than 2%. However, in the cross-database experiment (see third column in Table 6), and because LFW has no children, the results are equivalent to those with children faces.

4.4. Head out-of-plane rotations and gender estimation

In section 3.2 we analyzed the dependencies between 2D pose (in-plane misalignment) and gender. In this section we will show that gender and 3D

Table 7: Gender 5-fold cross-validation accuracy with GROUPS w/o children. First row shows the results of a gender classifier trained with images from all 2D poses. Second row displays the results for the Gender×Pose AP. The first six columns display the results for the six miss-alignments classes. Last column shows the average gender accuracy. Below each performance result we show the standard deviation preceded by symbol \pm .

Experiment/Pose	0	1	2	3	4	5	Global
Gender	74.43	82.91	78.47	76.77	75.80	80.19	78.49
	± 2.54	± 0.93	± 2.02	± 1.68	± 1.22	± 0.63	± 0.61
Gender×Pose	77.71	84.15	78.80	81.56	80.06	81.16	80.53
	± 1.23	± 1.08	± 2.27	± 0.84	± 1.75	± 1.33	± 0.63

face pose are also dependent and we can exploit that dependency to improve gender estimation.

In our experiments we have used face images from 11 points-of-view in the Multi-PIE database. We discarded the two extreme profile and two top head views. We used a frontal and a profile detector from OpenCV to achieve multi-view face detection (see Fig. 3). For this test we used a Leave One Person Out (LOPO) validation strategy training with the images of all subjects except one and testing with the images of the subject left out. To speed-up the validation process, instead of K-NN we used an euclidean distance classifier. In Table 4.4 we show the classification accuracy for Gender alone and for the Gender×3D Pose AP. In this case, the pose-aware classifier improves in 4% the gender recognition. This result, together with that related to 2D eye pose in section 3.2, shows clearly that it is possible to leverage on facial attribute dependencies to improve gender estimation.

Table 8: Gender and 3D Pose experiment. First row shows results for a gender alone classifier. The second row displays results the Gender×3D Pose AP.

Experiment/Data Base	Multi-PIE
Gender	84.31%
Gender×3D Pose	88.04%

5. Conclusions

In this paper we have studied the problem of gender recognition from a multi-attribute perspective. Gender recognition in laboratory conditions (e.g. the Color FERET database) is a well known problem for which state-of-the-art algorithms provide performances well above 90%. However, when these algorithms are tested in a real setting, they suffer a significant drop in performance. We have found that this is caused by the existence of dependencies among facial attributes that have not been considered when building the classifier. In the paper we exploit these dependencies to improve gender recognition.

We have confirmed previous results reporting the existence of dependencies between age and gender, and introduced new ones related to face pose. Our approach explores the combination of various demographic variables. We prove the interest of exploiting variable combination to improve classifier performance. Moreover, the combination of gender and pose attributes also avoids the use of computationally expensive and fragile face alignment procedures. In the experiments we confirm the existence of conditional dependencies among gender, age and pose facial attributes and prove that we can improve the performance of gender classifiers by exploiting these relations.

Our aim in this paper is to prove that by considering gender dependencies with other facial attributes we can improve the gender estimation accuracy and efficiency. This is more important when using images from real settings, such as those in the GROUPS database. Gallagher *et al.* (Gallagher and Chen, 2009) achieved 69.6% accuracy in GROUPS using an appearance-based classifier. In our experiments we achieved a 80.53% success rate for the same database using also an appearance-based multi-attribute linear classifier based on (Bekios-Calfa *et al.*, 2011).

In our classifiers we have used simple appearance-based features, since our goal in this paper was not obtaining the best gender classifier, but showing how to improve gender recognition. In future research we will analyze the best features for multi-view and multi-attribute facial demography estimation.

Acknowledgment

Juan Bekios-Calfa was funded by a spanish *UPM-Santander* studentship. The authors gratefully acknowledge funding from the Spanish *Ministerio de*

References

- Babenko, B., Dollár, P., Tu, Z., Belongie, S., 2008. Simultaneous learning and alignment: multi-Instance and multi-Pose Learning, in: Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, Marseille, France.
- Baluja, S., Rowley, H.A., 2007. Boosting sex identification performance. *International Journal of Computer Vision* 71, 111–119.
- Bekios-Calfa, J., Buenaposada, J.M., Baumela, L., 2011. Revisiting linear discriminant techniques in gender recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, 858–864.
- Duan, K., Parikh, D., Crandall, D., Grauman, K., 2012. Discovering localized attributes for fine-grained recognition, in: Proc. of Int. Conference on Computer Vision and Pattern Recognition, pp. 3474 – 3481.
- Farhadi, A., Endres, I., Hoiem, D., Forsyth, D., 2009. Describing objects by their attributes, in: Proc. of Int. Conference on Computer Vision and Pattern Recognition, pp. 1778–1785.
- Gallagher, A.C., Chen, T., 2009. Understanding images of groups of people, in: Proc. of Int. Conference on Computer Vision and Pattern Recognition, pp. 256–263.
- Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S., 2010. Multi-pie. *Image and Vision Computing* 28, 807 – 813.
- Guo, G., Dyer, C.R., Fu, Y., Huang, T.S., 2009. Is gender recognition affected by age?, in: Proc. of IEEE International Workshop on Human-Computer Interaction (HCI'09), pp. 2032–2039.
- Guo, G., Mu, G., 2010. A study of large-scale ethnicity estimation with gender and age variations, in: IEEE Int. Workshop on Analysis and Modeling of Faces and Gestures (AMFG'10), pp. 79–86.
- Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E., 2007. Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical Report 07-49. University of Massachusetts, Amherst.

- Kim, T., Stenger, B., Woodley, T., Cipolla, R., 2010. Online multiple classifier boosting for object tracking, in: Proc. IEEE Workshops on Computer Vision and Pattern Recognition.
- Kumar, N., Berg, A.C., Belhumeur, P.N., Nayar, S.K., 2009. Attribute and simile classifiers for face verification, in: Proc. of International Conference on Computer Vision, pp. 365–372.
- Lampert, C.H., Nickisch, H., Harmeling, S., 2009. Learning to detect unseen object classes by betweenclass attribute transfer, in: Proc. of Int. Conference on Computer Vision and Pattern Recognition, pp. 951–958.
- Learned-Miller, E., 2006. Data driven image models through continuous joint alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28, 236–250.
- Mäkinen, E., Raisamo, R., 2008. Evaluation of gender classification methods with automatically detected and aligned faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 541–547.
- Minear, M., Park, D.C., 2004. A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments and Computers* 36, 630–633.
- Moghaddam, B., Yang, M.H., 2002. Learning gender with support faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 707–711.
- Phillips, P., Moon, H., Rauss, P., Rizvi, S., 2000. The feret evaluation methodology for face recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 1090–1104.
- Tsoumakas, G., Katakis, I., Vlahavas, I., 2011. Random k-labelsets for multilabel classification. *IEEE Transactions on Knowledge and Data Engineering* 23, 1079–1089.
- Viola, P., Jones, M.J., 2004. Robust real-time face detection. *International Journal of Computer Vision* 57, 137–154.
- Wang, J., Markert, K., Everingham, M., 2009. Learning models for object recognition from natural language descriptions, in: Proc. British Machine Vision Conference.

- Wei, G., , Ai, H., 2009. Face gender classification on consumer images in a multiethnic environment, in: Proc. of third International Conferences, ICB 2009, Alghero, Italy, Springer-Verlag. pp. 169–178.
- Yu, F., Ji, R., Tsai, M.H., Ye, G., Chang, S.F., 2012. Weak attributes for large-scale image retrieval, in: Proc. of Int. Conference on Computer Vision and Pattern Recognition, pp. 2949–2956.
- Zhu, M., Martinez, A., 2006. Subclass discriminant analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28, 1274 –1286.