On the Importance of Multi-dimensional Information in Gender Estimation from Face Images

Juan Bekios-Calfa[1], José M. Buenaposada[2], and Luis Baumela[3]

[1] Dept. de Ingeniería de Sistemas y Computación, Universidad Católica del Norte
Av. Angamos 0610, Antofagasta, Chile
juan.bekios@ucn.cl

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

[3] 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 demography from images is a problem that has recently been extensively studied because of its relevant applications. We review state-of-the-art approaches to gender classification and confirm that their performance drops significantly when classifying young or elderly faces. We hypothesize that this is caused by the existence of dependencies among the demographic variables that were not considered in traditional gender classifiers. In the paper we confirm experimentally the existence of such dependencies between age and gender variables. We also prove that the performance of gender classifiers can be improved by considering the dependencies with age in a multi-dimensional approach. The performance improvement is most prominent for young and elderly faces.

1 Introduction

By demographic classification we denote the problem of extracting personal attributes from face images [2,9], voice [12], clothing [16], names [2] or even gait [10]. This is a problem that has received very much attention recently because of its applications in human computer interaction, video indexing and video analytics for business intelligence [9]. The main demographic variables are gender, ethnicity and age. There is nevertheless a plethora of other interesting variables such as hairstyle, hair color, facial expression, wear glasses or not, have mustache or not, etc. We will concentrate on the gender attribute and its relation with age on near frontal face images.

Gender is perhaps the most widely studied facial demographic attribute in the Computer Vision field [14,3,11,4]. The state-of-the-art recognition rate in the Color FERET database [15] involving frontal faces with frontal illumination and 5 fold cross-validation is around 93% using either a Support Vector Machine

C. San Martin and S.-W. Kim (Eds.): CIARP 2011, LNCS 7042, pp. 264–271, 2011.

with Radial Basis function [14], pair-wise comparison of pixel values within a boosting framework [3] or linear discriminant techniques [4]. 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 [13], the performance drops to roughly 70% success rate [4]. 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 more ethnic groups and with a broader range of 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 suggest the existence of a dependency between gender and other demographic variable.

In this paper we will study the dependencies between gender and age given the facial appearance. Dependencies among demographic variables have also been previously considered in the literature. Gao and Ai [19] 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 [8], in experiments with the MORPH-II database, found that age estimation can have large errors if the influence of gender and ethnicity is no considered. Finally, Guo et al. [7] 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 estimating gender. 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.

2 Exploiting the Dependencies between Gender and Age

In this section we study the dependencies between the age and gender demographic variables given the facial apperance. First we will prove experimentally that those dependencies exist. Secondly we will exploit them to improve the accuracy of gender classification.

2.1 Are There Any Dependencies between Age and Gender?

If we assume that in any age range there is equal number of men and women and for any gender the distribution of population in ages is similar, then we are implicitly assuming 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.

To confirm whether A and G are independent variables we have trained a gender classifier as explained in Section 2.2 with the data in the GROUPS database (see 3 for a description) and tested it on PAL. The classifier has been trained with all men and women images in GROUPS but we compute the accuracy

Table 1. Accuracy in GROUPS/PAL experiment. First row shows results for a gender classifier trained with images from all age ranges. Second row displays results of four gender classifiers each one trained only with face images from a given age range. First four columns display results of four age ranges. 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

stratified into four age groups. In the first row of Table 1 we show the results of this experiment. Gender estimation for the range 20–36 are above the state-of-the-art in [4] whereas the results for the other age ranges are clearly below. This results indicates 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, in GROUPS and tested it on data from the same age range in PAL. This experiment provides information on the performance of a gender classifier that knows the age range of the subject. As we can see in the second row in Table 1, the classifier performance increases, most notably for elderly. 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 appearance of a face, X, depends on the gender, G, and age, A, of the subject, then age and gender are conditionally dependent, given the appearance of the face X, i.e. $P(A, G|X) \neq P(A|X)P(G|X)$.

2.2 Multi-dimensional Classification of Gender and Age

We will simultaneously estimate gender and age using a multi-dimensional approach. The term *Multi-dimensional Classification* was introduced first by Gaag and Waal [5] to represent classification problems in which there are several class variables. This problem is a generalization of the multi-label problem [1]. In the multi-dimensional case each label is transformed into a dimension, which can have more than two values, in opposition to a label, which can only have two values.

Let be $D = \{d_1, \ldots, d_M\}$ the set of M dimensions of a given multi-dimensional problem and let be $V_i = \{c_1^i, \ldots, c_{N_i}^i\}$ the set of possible values for dimension d_i where $N_i = |V_i|$. Let be $D_\times = V_1 \times V_2 \times \ldots \times V_d$ the Cartesian product of all V_i dimension values sets. The output of a multi-dimensional classifier for an input instance, $\mathbf{x}$, is a vector $\mathbf{z} \in D_\times$. In the demography classification problem, one of the possible multi-dimensional formulations could be to have three dimensions, $D = \{age, gender, ethnicity\}$ and the corresponding values $V_{age} = \{young, adult, senior\}$, $V_{gender} = \{male, female\}$ and $V_{ethnicity} = \{Caucasian, African, mongoloid\}$.

One of the baseline approaches to multi-label classification is a problem transformation method: Label Powerset (LP) [17]. 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. The Label Powerset (LP) approach has an extension for the multi-dimensional case. The Dimension Powerset (DP) interprets every element in $D_\times$ as a single class in a multi-class classification problem. For each new instance the DP classifier outputs the estimated class that in fact represents a valid combination of dimensions values in the original multi-dimensional problem. The number of classes in the DP transformed problem is bounded by the minimum of N, number of training samples, and $|D_\times|$. Learning is difficult on the DP method with a low number of training samples in any combination of dimensions values (e.g. in demography classification the combination of Caucasian, male and young). In our case, we have 2 gender values and seven age categories, then we will have a 14 classes in the transformed (from a multi-dimensional one) multi-class classification problem. As the number of dimensions grows, it is very likely to find a combination of dimensions values with very low or no training data at all. To avoid such a problem we remove classes with few data.

In our approach we train the multi-class classifier for $D_\times$ in the PCA+LDA transformed subspace (LDA after Principal Components Analysis projection). In our experiments we use a K-Nearest Neighbor (KNN) classifier in the PCA+LDA transformed subspace. However, any multi-class classifier could be used within our framework. Depending on the amount of training data, the performance of the classifier built on PCA+LDA subspace decreases when retaining all PCA eigenvectors associated with non-zero eigenvalues. We select the dimension of the subspace resulting from the PCA step. We sort PCA eigenvectors in descending eigenvalue order. We then perform cross-validation and select the dimension with the best performance for the classification in the PCA+LDA subspace. This cross-validation driven feature selection (wrapper) approach is essential to correctly train a PCA+LDA procedure [4]. In the same cross-validation procedure we look for the dimension for the PCA initial step and the number of neighbors, k, that accounts for the best performance.

3 Experiments

In this section we evaluate the performance of the multi-dimensional framework estimating the combination of gender and age dimensions. We use the Images of Groups Dataset [6] (GROUPS database) and the Productive Aging Lab Face (PAL) database [13] for training and testing. We crop and re-size images to a base size of 25×25 pixels using OpenCV's[1] 2.1.0 face detector, which is based on [18]. Then we equalize the histogram to gain some independence from illumination changes. Finally, we also apply an oval mask to prevent the background from influencing our results.

[1] `http://opencv.willowgarage.com`

3.1 Face Databases

The Images of Groups Dataset[2] (in the following GROUPS database), consists of 28,231 faces labeled with age and gender extracted from 5,800 pictures of people groups. Most faces were automatically detected. The seven age categories used are: 0-2, 3-7, 8-12, 13-19, 20-36, 37-65, and 66+. In this database age labels are discrete. In our experiments we use those face detections from the GROUPS database that have at least 60×60 pixels (13,051 out of a total of 28,231 faces). See Fig. 1 for some examples of faces from this database. In Table 2 we give information on the number of images per age category.

Fig. 1. Some of the images from GROUPS database

The PAL Database 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, we use 579 images in our PAL experiments. The right profile and some facial expressions are also available for some subjects. There are 219 male and 357 female subjects divided into four groups depending on their age. In this database, the actual continuous face age is available for each image. In our experiments we only use frontal images and only one image per subject. See some sample images in Fig. 2. Again, the number of images per age category in this database is shown in Table 2.

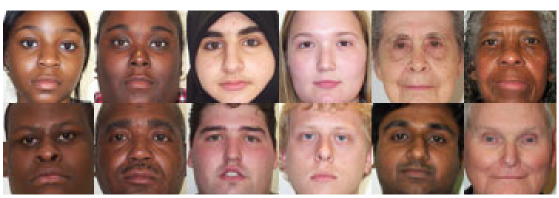

Fig. 2. Some cropped images after face detection, from the PAL database

3.2 Gender Estimation Results

To evaluate the performance of a gender estimation algorithm we are interested in the algorithm's generalization capabilities. To this end we train our algorithm using one database and test it on a different one. We train our algorithms

[2] `http://chenlab.ece.cornell.edu/people/Andy/ImagesOfGroups.html`

Table 2. Number of image per age category

Database/Age Range	0-2	3-7	8-12	13-19	20-36	37-65	66+	Total
GROUPS	460	807	368	777	6972	3109	558	13051
PAL	-	-	-	32	221	123	203	579

Table 3. Multi-dimensional gender accuracy in GROUPS/PAL experiment

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

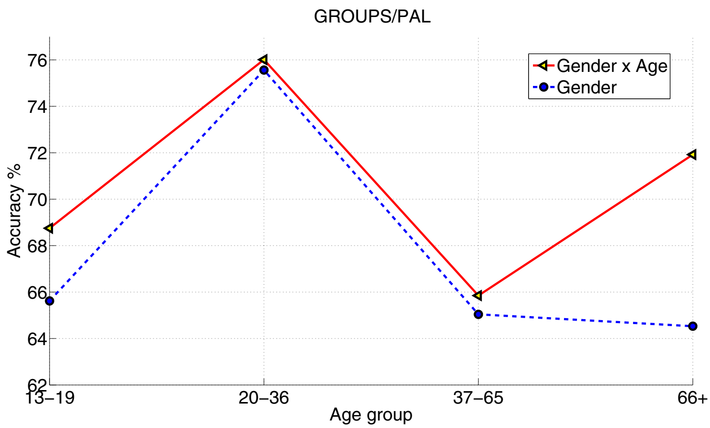

Fig. 3. GROUPS/PAL experiment comparison between uni-dimensional and multi-dimensional approaches

with GROUPS and test them with PAL (GROUPS/PAL experiment). We use GROUPS for training because it is large (≈13000 data) and it has a broad demography with realistic acquisition conditions. On the other hand, we use PAL for testing because it is a difficult database for gender recognition mostly due to the broad demography [4].

As already said in Section 2 we use a PCA+LDA dimensionality reduction procedure and a KNN classifier. We use a five-fold cross-validation scheme to estimate the classifier parameters (number of retained eigenvectors and number of neighbors in KNN). We have performed two experiments, one using only the *Gender* dimension (see Table 1) and another using *Gender*×*Age* dimension powerset (see Table 3). Note that in PAL there are no faces in the 0 to 12 years age ranges. The global accuracy (computed by weighted mean of the per age category accuracy using the proportion of data on each category) of the multi-dimensional approach (*Gender*×*Age*) is better than the uni-dimensional one (only Gender) by 3%. Most interestingly, the multi-dimensional approach outperforms the uni-dimensional one in Table 1 for all age categories (see also Fig. 3). Moreover, the most difficult categories for the uni-dimensional classifier, 13-19 and 66+, are those for which the improvement is highest, 3.13% and 7.39% respectively. This shows that the multi-dimensional procedure is able to exploit the dimension combination and improve the gender estimation accuracy.

4 Conclusions

In the paper we have reviewed the state-of-the-art in gender recognition from near frontal face images. We have confirmed previous results reporting the existence of dependencies between age and gender. Our approach explores the combination of various demographic variables and proves the interest of exploiting variable combination to improve classifier performance. In future research lines it would be interesting to use both better visual descriptors, as used in [7] within the multi-dimensional approach.

Acknowledgment. The authors gratefully acknowledge funding from the Spanish *Ministerio de Ciencia e Innovación* under contracts `TIN2010-19654` and the *Consolider Ingenio* program contract `CSD2007-00018`.

References

1. de Carvalho, A., Freitas, A.A.: A tutorial on multi-label classification techniques. Foundations of Computational Intelligence. Function Approximation and Classification, vol. 5. Springer, Heidelberg (2009)
2. Gallagher, A.C., Chen, T.: Estimating age, gender and identity using first name priors. In: Proc. of CVPR (2008)
3. Baluja, S., Rowley, H.A.: Boosting sex identification performance. International Journal of Computer Vision 71(1) (January 2007)
4. Bekios-Calfa, J., Buenaposada, J.M., Baumela, L.: Revisiting linear discriminant techniques in gender recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(4), 858–864 (2011)
5. van der Gaag, L.C., de Waal, P.R.: Multi-dimensional bayesian network classifiers. In: European Conference on Probabilistic Graphical Models, pp. 107–114 (2006)
6. Gallagher, A.C., Chen, T.: Understanding images of groups of people. In: Proc. of CVPR, pp. 256–263 (2009)
7. Guo, G., Dyer, C.R., Fu, Y., Huang, T.S.: Is gender recognition affected by age? In: Proc. of IEEE International Workshop on Human-Computer Interaction (HCI 2009), pp. 2032–2039 (2009)
8. Guo, G., Mu, G.: 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 2010), pp. 79–86 (2010)
9. Kumar, N., Berg, A.C., Belhumeur, P.N., Nayar, S.K.: Attribute and Simile Classifiers for Face Verification. In: Proc. of ICCV (October 2009)
10. Li, X., Maybank, S., Yan, S., Tao, D., Xu, D.: Gait components and their application to gender recognition. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 38(2), 145–155 (2008)

11. Mäkinen, E., Raisamo, R.: Evaluation of gender classification methods with automatically detected and aligned faces. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(3), 541–547 (2008)
12. Metze, F., Ajmera, J., Englert, R., Bub, U., Burkhardt, F., Stegmann, J., Muller, C., Huber, R., Andrassy, B., Bauer, J., Littel, B.: Comparison of four approaches to age and gender recognition for telephone applications. In: Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2007, vol. 4, pp. 1089–1092 (2007)
13. Minear, M., Park, D.C.: A lifespan database of adult facial stimuli. Behavior Research Methods, Instruments and Computers 36, 630–633 (2004)
14. Moghaddam, B., Yang, M.H.: Learning gender with support faces. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(5), 707–711 (2002)
15. Phillips, P., Moon, H., Rauss, P., Rizvi, S.: The feret evaluation methodology for face recognition algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(10), 1090–1104 (2000)
16. Sivic, J., Zitnick, C.L., Szeliski, R.: Finding people in repeated shots of the same scene. In: Proc. BMVC (2006)
17. Tsoumakas, G., Katakis, I., Vlahavas, I.: Random k-labelsets for multi-label classification. IEEE Transactions on Konwledge and Data Engineering (2010)
18. Viola, P., Jones, M.J.: Robust real-time face detection. International Journal of Computer Vision 57(2), 137–154 (2004)
19. Gao, W., Ai, H.: Face gender classification on consumer images in a multiethnic environment. In: Tistarelli, M., Nixon, M.S. (eds.) ICB 2009. LNCS, vol. 5558, pp. 169–178. Springer, Heidelberg (2009)