Pattern Recognition Letters 36 (2014) 228–234

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Robust gender recognition by exploiting facial attributes dependencies

Juan Bekios-Calfa [a,c], José M. Buenaposada [b,c,*], Luis Baumela [b,c]

[a] Dept. de Ingeniería de Sistemas y Computación, Universidad Católica del Norte, Av. Angamos 0610, Antofagasta, Chile
[b] Dept. de Ciencias de la Computación, Universidad Rey Juan Carlos, Calle Tulipán s/n, 28933 Móstoles, Spain
[c] Dept. de Inteligencia Artificial, Universidad Politécnica de Madrid, Campus Montegancedo s/n, 28660 Boadilla del Monte, Spain

ARTICLE INFO

Article history:
Available online 13 May 2013

Communicated by Luis Gomez Deniz

Keywords:
Gender estimation
Facial attribute dependencies
Linear Discriminant Analysis

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.

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 suggest 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 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

* Corresponding author at: Dept. de Ciencias de la Computación, Universidad Rey Juan Carlos, Calle Tulipán s/n, 28933 Móstoles, Spain. Tel.: +34 914888129.

E-mail addresses: juan.bekios@ucn.cl (J. Bekios-Calfa), josemiguel.buenaposada@urjc.es (J.M. Buenaposada), lbaumela@fi.upm.es (L. Baumela).

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

0167-8655/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2013.04.028

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 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 (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 (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. (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 $C_\times = V_1 \times V_2 \times \cdots \times V_d$ 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), ..., (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

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