Alignment-free gender recognition in the wild

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. Gender is possibly the most common facial attribute automatically estimated from images. Achieving robust gender classification "in the wild," i.e. in images acquired in real settings, is still an open problem. Face pose variations are a major source of classification errors. They are solved using sophisticated face alignment algorithms that are costly computationally. They are also prone to getting stuck in local minima thus providing a poor pose invariance. In this paper we move the alignment problem to the learning stage. The result is an efficient pose-aware classifier with no on-line alignment. Our efficient procedure gets state of the art performance even with facial poses "in the wild." In our experiments using "The Images of Groups" database we prove that by simultaneously predicting gender and pose we get an increase of about 5% in the performance of a linear state-of-the-art gender classifier.

1 Introduction

Gender is perhaps the most widely studied facial demographic attribute in Computer Vision. Most of gender recognition algorithms have been evaluated in databases acquired in controlled conditions. The FERET Color Database [15] has been one of the most used benchmarks in gender recognition [14, 2, 13, 3]. In FERET the illumination is uniform and faces are frontal. Additionally, the range of ages is quite limited, since the number of senior subjects is low and there are no young subjects. Other databases have a broader demography. The Productive Aging Lab Face (PAL) database [3] (with adults and senior subjects) and MORPH II database [8] (with adults and seniors) are prominent examples. However, they were acquired in controlled conditions and algorithms trained on them would not generalise well to unconstrained real settings.

In recent years there is a trend to use face databases acquired "in the wild", i.e. in real settings. They emerged with different problems in mind: *Labelled Faces in the Wild* (LFW) [9] for face verification, *The Images of Groups Dataset*[4]

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

(GROUPS) [6] for age and gender recognition and finally the *Annotated Facial Landmarks in the Wild* (AFLW) [11] for face detection and pose estimation. The LFW database consists of 13,233 face images of 5,740 different subjects from the web and has been proposed as a benchmark for gender recognition[5]. However, LFW age distribution is quite limited, mostly consisting in middle-aged adults. It is also gender imbalanced ($\approx 10,000$ male and $\approx 3,000$ female faces) [5]. On the other hand, the GROUPS database consists of 28,231 faces labeled with age (seven discrete age ranges), gender and eyes position extracted from 5,800 images of groups of people. In this database gender is balanced. See Fig. 2 for some sample faces from this database. In this database Gallagger et *al.* [6] achieved 69.6% accuracy using 7 Linear Discriminant Analysis (LDA) projections (one per age range) and a K-Nearest Neighbour classifier ($K = 25$). Gallagher trained gender classifier using 23,218 images and tested it on 1,881. By using the Locally Binary Patterns (LBPs) as image descriptor and PCA projection followed by a Linear SVM on eye aligned faces, Dago-Casas et *al.* [5] achieved 86.34% 5 fold cross validation accuracy. Even when [6] and [5] employed aligned images using the eye positions, the classification is far from being perfect. The reason is that GROUPS is a real-world database: different illumination conditions, facial expressions, face poses and broad demographic distribution of faces. For this reason we will use GROUPS database in our experiments.

In face analysis problems (expressions recognition, face verification, gender recognition, age estimation, etc.) using real-world images the intra-class variability is usually larger than inter-class one. The changes in appearance produced by the orientation of the head, illumination, sun glasses or facial expressions can make the same face look quite different. By removing the intra-class variability the recognition performance can be improved significantly.

In this paper we are interested in dealing with face pose variations in gender recognition as a source for intra-class variability. Learning with pose variations can be done in three ways: 1) training with data separated into coherent groups (e.g. [1, 10]); 2) aligning the images to some canonical pose or configuration [5, 6]; and 3) using local features detectors invariant to rotations [16]. Options 1) and 2) are holistic methods (using the whole face image) whereas the third option is a feature-based approach. The problem with feature-based approaches is that they may not find enough features in reduced resolution surveillance footage. For this reason we will concentrate in holistic approaches.

Previous holistic face gender classification in real-world images (e.g. [5, 6]) aligned their images to a canonical pose. Face alignment requires the automated or manual detection of fiducial points [13, 6, 5] or a *congealling* previous step (aligning all the images in a set by reducing entropy [12]). These solutions are both costly in terms of computational resources and prone to errors, providing an incorrect alignment.

In this paper we transfer the alignment problem to the learning phase, removing the need for on-line alignment. Although there are some works about simultaneous alignment and classification in learning [1, 10] we will follow a dif-

[5] http://fipa.cs.kit.edu/downloads/LFW-gender-folds.dat

ferent path. We find clusters of face poses after face detection and use them in training. 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 classifier.

2 Face pose in frontal face detection

The GROUPS database provides the pixels coordinates of the eyes for every face. On the other hand, we use face detection output as the only aligment. The face detector in OpenCV 2.4 [6] is used in order to get the detector miss-aligment bias in the cropped faces. Detector parameters are set to miss the lowest number of faces (which increases the number of false positives). The false positives are removed by using the ground truth eye positions. The result from this process is 22,948 correctly detected faces (see Fig. 2): 11,932 female and 11,016 male. Our goal is to know the classes of miss-alignments produced by the face detector. This miss-alignments are possibly the largest source of intra-class variability.

Fig. 1. Some of the images from GROUPS database.

In order to get the clusters of different miss-alignments we use a procedure similar to the Poselets idea [4]. The ground truth eye coordinates are first changed to be relative to the top-left corner of the face detection window. We define a canonical image for face detections of 60×60 pixels and all ground truth relative eyes coordinates are transferred to this reference face size (see Fig.2). The feature vector we use for clustering is $(x_{le}, y_{le}, x_{re}, y_{re}^{\top})$ where le stands for canonical image left most eye coordinates and re stands for image right most eye coordinates. We use K-means clustering with an experimentally chosen value of $K = 6$. In Fig. 2 we show the clusters. In the left column we show the mean of eye position within each cluster. In the other columns we show some examples of images assigned to each cluster and the ground truth eye positions overlayed. The appearance changes between the face pose clusters are strong enough to distract any classification algorithm using holistic features.

The face pose classes have an easy interpretation:

- *Cluster 0, 1 and 2*, are frontal face detections where the degree of scene context included varies. The face detections include more or less background

[6] `http://opencv.willowgarage.com/wiki/`

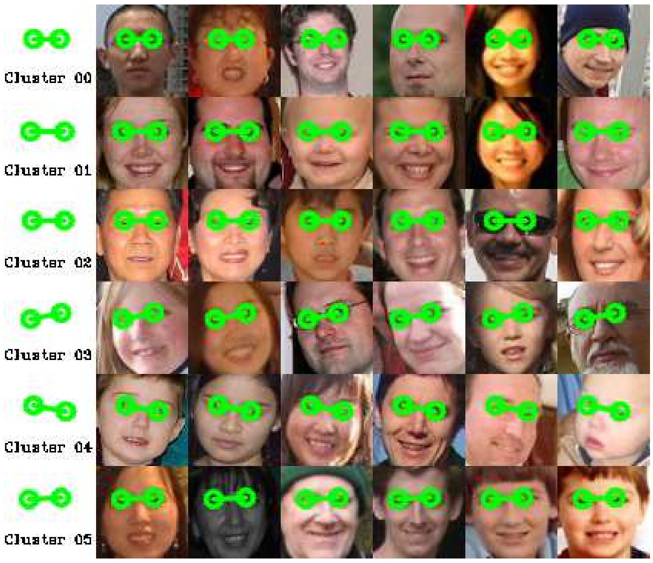

Fig. 2. Found clusters of eyes positions within the canonical face detection window.

depending on face distance to the camera, image resolution, background texture near the face or sliding window parameters (scale change and number of pixels displacement).

- *Cluster 3*, are faces with an in-plane rotation by a positive angle.
- *Cluster 4*, are faces with an in-plane rotation by a negative angle.
- *Cluster 5*, are faces with a subtle in-plane rotation by a positive angle.

The distribution of faces (see table 1) on each pose cluster and gender is balanced (a median of 7.34% of total data per subclass). Therefore, learning will not be biased by the amount of data per class.

Gender/Pose	0	1	2	3	4	5
Female	1359 (5.92%)	2561 (11.16%)	2677 (11.66%)	1627 (7.08%)	1599 (6.96%)	2109 (9.19%)
Male	1987 (8.65%)	2565 (11.17%)	1713 (7.46%)	1660 (7.23%)	1526 (6.64%)	1565 (6.81%)

Table 1. Number of images per each Gender and pose clusters combinations (the percentage over the number of data are shown between parentheses).

3 Learning gender with pose subclasses

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. We can adapt the Label Powerset (LP) idea to our problem. The Powerset, $C_{\times} = gender \times pose$, of possible gender ($\{male, female\}$) and pose ($\{0, \ldots, 5\}$) values, is the Cartesian product of all gender and pose class values. In our case the cardinality of the Powerset is $|C_{\times}| = 12$.

By transforming the original gender classification problem into the Powerset subclasses we are dividing the female and male classes into subclasses. In order to get well separated and compact subclasses (with the same gender and pose on each) we perform Fisher Linear Discriminant Analysis (LDA) dimensionality reduction of the training data using the Powerset labels (12 classes giving a 11 dimensions feature space). Our approach for dimensionality reduction is related to Subclass Discriminant Analysis (SDA) [18] but we find the subclasses with face pose and gender labels.

After dimensionality reduction, and given that we have gender subclasses the problem is likely to be non-linearly separable. Therefore we use a K-Nearest-Neighbours classifier (K-NN) with binary labels (male or female). The K nearest neighbours votes are weighted by the inverse of the distance to the query point.

4 Experiments

In our experiments we compare the classifier described in section 3, termed *Pose Aware Classifier* (PaC) with a state-of-the-art linear gender classifier [3], termed *Linear Classifier* (LC). In the experiments we use a K-NN classifier and a 5 fold cross-validation scheme, using the same folds for all experiments. We sampled the folds in such a way that each of them had the same number of samples for each of the 2×6 classes. We cropped and re-sized the images to a base size of 25×25 pixels using OpenCV's[7] 2.4 face detector. After cropping we equalise the images to gain some independence from global brightness changes.

In the first experiment, we used all 22,948 face images from the GROUPS database. We obtained a 5 fold cross-validation accuracy of 73.68% with LC (see table 2). If we analyse the behaviour of LC stratified by poses (see table 3, LC row) the performance for the frontal images is 3% to 4% better than that for the other poses. Since 57.5% of the training data (see table 1) are in frontal position with almost no background (poses 1, 2 and 5) it is quite reasonable that the LC classifier trained with all images presents a bias towards frontal positions. With the PaC classifier (see table 2, PaC row) we get an accuracy of 78.05%. This result compared to that obtained with LC represents a significant improvement of 4.4% in the global gender recognition performance. The influence of pose in

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

learning is more evident if analyse the recognition results stratified by pose (see table 3, PaC row). We get an average improvement of about 6% in the poses with largest face rotation (poses 3 and 4).

These results show that the LaC is able to leverage on pose information to improve classification accuracy without the need of face alignment.

	Female	Male	Global
Pose-aware Classifier (PaC)	**79.28%**	**76.71%**	**78.05%** (K=385)
Linear Classifier (LC)	73.78%	73.57%	73.68% (K=377)

Results without children

Pose aware Classifier (PaC)	**81.69%**	**79.96%**	**80.88%** (K=391)
Linear Classifier (LC)	76.42%	76.30%	76.36% (K=387)

Table 2. 5-fold mean accuracy for gender recognition on the GROUPS database.

Guo et *al.* found that gender is more difficult to estimate on young faces than in adult ones [7]. Dago-Casas et *al.* [5] found exactly the same result with the GROUPS database. In this experiment we confirm that the difficulties estimating the gender for children have not biased our previous results. To this end we compare the performance of the classifiers with adult faces alone: 20,215 images (11.9% of data removed). The accuracy with LC classification is 76.36% (see table 2, LC row in the no children section). However the LaC classifier gets an improvement above 4%, with a global accuracy of 80.88%. Similarly, the pose stratified results in the "without children" section of table 3 also achieve the highest performance improvement (around 7%) in those poses with the largest face rotation. So, in this second experiment we have confirmed the previous result with no children in the database.

	0	1	2	3	4	5
Pose-aware Classifier (PaC)	**75.88%**	**79.65%**	**77.26%**	**77.88**	**77.37%**	**79.45%**
Linear Classifier (LC)	71.60%	75.84%	74.60%	71.76%	70.72%	75.69%

Results without children

Pose-aware Classifier (PaC)	**78.14%**	**84.03%**	**78.91%**	**81.43%**	**80.28%**	**81.56%**
Linear Classifier (LC)	72.84%	81.62%	75.57%	74.51%	73.33%	77.79%

Table 3. Per pose 5-fold mean accuracies for gender recognition over GROUPS database.

5 Conclusions

Face pose is a source of strong changes in appearance variability that hamper image analysis algorithms. It has traditionally been ignored, by assuming faces in frontal position, e.g. [3], or solved using sophisticated face alignment algorithms [13]. However, face alignment procedures are costly computationally and prone to getting stuck in local minima such that they do not increase the classification rates [13]. In this paper we have addressed the problem of estimating the gender of face images captured "in the wild." To this end we have established a set of five standard poses obtained from clusters of automatically detected faces. By introducing a classifier that simultaneously predicts gender and face pose we have moved the alignment problem to the learning stage. We have tested our approach with "The Images of Groups" database. In our experiments we have proved that by simultaneously predicting gender and pose we get an increase of about 5% in the performance of a state-of-the-art gender classifier.

Acknowledgement

Juan Bekios-Calfa was founded by a spanish *UPM-Santander* studentship. The authors gratefully acknowledge funding from the Spanish *Ministerio de Ciencia e Innovación* under *Consolider Ingenio* program contract `CSD2007-00018`.

References

1. Babenko, B., Dollár, P., Tu, Z., Belongie, S.: Simultaneous Learning and Alignment: Multi-Instance and Multi-Pose Learning. In: Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition. Marseille, France (2008)
2. Baluja, S., Rowley, H.A.: Boosting sex identification performance. International Journal of Computer Vision 71(1) (January 2007)
3. 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 (April 2011)
4. Bourdev, L., Malik, J.: Poselets: Body part detectors trained using 3d human pose annotations. In: Proc. of International Conference on Computer Vision. pp. 1365 –1372 (oct 2009)
5. Dago-Casas, P., Gonzalez-Jimenez, D., Yu, L.L., Alba-Castro, J.: Single- and cross-database benchmarks for gender classification under unconstrained settings. In: Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on. pp. 2152 –2159 (nov 2011)
6. Gallagher, A.C., Chen, T.: Understanding images of groups of people. In: Proc. of Int. Conference on Computer Vision and Pattern Recognition. 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'09). 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'10). pp. 79–86 (2010)

9. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Tech. Rep. 07-49, University of Massachusetts, Amherst (October 2007)
10. Kim, T., Stenger, B., Woodley, T., Cipolla, R.: Online multiple classifier boosting for object tracking. In: Proc. IEEE Workshops on Computer Vision and Pattern Recognition (2010)
11. Koestinger, M., Wohlhart, P., Roth, P.M., Bischof, H.: Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization. In: First IEEE International Workshop on Benchmarking Facial Image Analysis Technologies (2011)
12. Learned-Miller, E.: Data driven image models through continuous joint alignment. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(2), 236 –250 (feb 2006)
13. 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 (March 2008)
14. Moghaddam, B., Yang, M.H.: Learning gender with support faces. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(5), 707–711 (May 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 (October 2000)
16. Toews, M., Arbel, T.: Detection, localization, and sex classification of faces from arbitrary viewpoints and under occlusion. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(9), 1567 –1581 (sept 2009)
17. Tsoumakas, G., Katakis, I., Vlahavas, I.: Random k-labelsets for multi-label classification. IEEE Transactions on Konwledge and Data Engineering (2010)
18. Zhu, M., Martinez, A.: Subclass discriminant analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(8), 1274 –1286 (aug 2006)