MORF: Multi-Objective Random Forests for Face Characteristic Estimation

Dario Di Fina[1] Svebor Karaman[1,3] Andrew D. Bagdanov[2] Alberto Del Bimbo[1]
[1]MICC - University of Florence [2]CVC - Universitat Autonoma de Barcelona [3]DVMM Lab - Columbia University
{dario.difina, alberto.delbimbo}@unifi.it, svebor.karaman@columbia.edu, bagdanov@cvc.uab.es

Abstract

In this paper we describe a technique for joint estimation of head pose and multiple soft biometrics from faces (Age, Gender and Ethnicity). Our proposed Multi-Objective Random Forests (MORF) framework is a unified model for the joint estimation of multiple characteristics that automatically adapts the measure of information gain used for evaluating the quality of weak learners. Since facial characteristics are related in the feature space, estimating all of them jointly can be beneficial as trees can learn to condition the estimation of some characteristics on others. We reformulate the splitting criterion of random trees in our multi-objective formulation and evaluate it on publicly available face characteristic estimation imagery. These preliminary experiments show promising results.

1. Introduction

Facial characteristics like Gender, face pose, Age and Ethnicity are important to estimate in many computer vision applications. Estimating the Gender and Age can be used to adapt advertising displayed on nearby screens, and pose estimation can allow users to interact with devices by simply looking at them. Although head pose is not a biometric characteristic, it is related to the gaze of a person and therefore can be an important aspect of behaviour and social interaction understanding. Despite the attention received in recent years, estimation of multiple face characteristics, and especially multiple soft biometrics like Age, Gender and Ethnicity, remains a difficult problem and an active area of research in the computer vision community.

Most systems for estimating characteristics like Age, head pose [8] and Gender [11, 16] use their own sets of custom features and specific estimation techniques. In systems requiring simultaneous estimates of target characteristics in real-time, this can be wasteful since much work (like feature extraction) is duplicated. Moreover, it may be easier to estimate one characteristic after conditioning it on a subset of others. Instead of estimating characteristics individually, we believe estimators can and should share the same pool of features, and perhaps even the same estimators. In this way, estimation of multiple characteristics can be made more efficient and more robust.

We believe that random decision forests [2] can provide a unified framework for multi-objective estimation. In this article we show how they can be used to simultaneously estimate multiple characteristics using a single pool of features. We propose a new information gain formulation enabling the use of multiple (potentially heterogeneous) characteristics to train a random forest. We demonstrate the effectiveness of our approach for jointly estimating head pose, Gender, Age and Ethnicity from single face images.

In the next section we discuss the related literature on soft biometric and head pose estimation from face images. In section 3 we describe our approach to multi-objective estimation with random decision forests. We report on a series of experiments we performed to evaluate the potential of our proposed estimation framework in section 4, and we conclude in section 5 with a discussion of our contribution.

2. Related work

In this section we discuss some of the recent literature on face characteristic estimation from images.

Gender classification. Zheng et al. [16] proposed a support vector machine with automatic confidence (SVMAC) and the LGBP feature for gender classification. Guo et al. [5] demonstrated that Gender recognition based on LBP and HOG features can be helped by Age classification.

Age estimation. Gunay et al. [4] proposed an Age classification approach based on a nearest neighbour algorithm with LBP feature, obtaining about 80% accuracy on a subset of the FERET dataset [12]. In [13] the authors introduce the concept facial aging to improve the efficiency of Gender estimation and face recognition and proposing the "MORPH" dataset [14] as a benchmark for Gender and Age estimation.

Soft biometric estimation. The authors in [10] proposed a technique for face verification in uncontrolled environments, considering two different approaches: one based on attribute classifiers for Gender, Age and Ethnicity, and a second one to recognize facial landmarks.

One of the first works to jointly estimate multiple characteristics was presented in [15]. They use AdaBoost on LBPH features to jointly estimate Gender, Age and Ethnicity. With the same goal, Guo et al. in [6] investigated two different ways to estimate these same characteristics using Kernel Canonical Correlation Analysis (KCCA) and Partial Least Squares (PLS). They show that it is possible jointly estimate Age, Gender and Ethnicity, while significantly reducing feature dimensionality. We will use this work as our baseline.

Head pose estimation. Head pose, while not a soft biometric, is a fundamental characteristic in video surveillance as it can provide an estimate of interest for profiling-at-a-distance applications [9], or evidence of the direction in which a person is moving [1]. In [1] the authors propose an unsupervised method to estimate gaze direction with random decision trees over Histogram of Oriented Gradients (HOG) and color features. In [7] the authors use partial least squares (PLS) and a 3-level pyramid of HOGs to estimate head pose in the presence of misalignment.

Random forests for face characteristic estimation. The random forest model has been used for head pose estimation [2, 8]. Existing approaches based on random forests focus on a single characteristic estimation, or consider multiple characteristics independently. Random forests have been shown to be a powerful tool for head pose estimation [8], and we believe they can be naturally extended to simultaneously estimation multiple, heterogeneous properties (including soft biometrics) from face imagery.

3. Multi-objective Random Forests (MORF)

In this section we describe our multi-objective estimation approach using random forests.

3.1. Feature representation for faces

We use the HOG feature descriptor to describe faces. We first resize all face images to a canonical dimension of 41×54 pixels. Using the standard cell configuration of 8×8 pixels, the gradient orientations are extracted in each cell and quantized into 8 orientations. Block normalization is then applied to 4×4 blocks of cells, and normalized outputs are concatenated to the final HOG vector of 1024 bins.

3.2. Random forests for supervised estimation

In this section we describe the basic, single-objective estimation model upon which MORF are based and how we extend it to multi-objective estimation model using random forests.

Training single-objective random trees. We consider supervised estimation problems. Assume we have a set of N face images available, each represented as a d-dimensional vector and labeled with a single characteristic c:

$$S_0 = \{(\mathbf{x}_i, y_i) \mid i = 1, \ldots, N\}, \tag{1}$$

where $\mathbf{x}_i = (x_1, x_2, \ldots x_d)$ and $y_i \in \mathcal{Y}^c$, the set of labels of characteristic c. A random tree is built by recursively splitting this initial set $\mathcal{S}_0$ of labeled examples in such a way that the mutual information between the set of examples at each node and the characteristic c being estimated is maximized.

We define a splitting function that compares two dimensions in the face descriptor:

$$h(\mathbf{x}; \theta_1, \theta_2) = \begin{cases} 1 & \text{if } x_{\theta_1} > x_{\theta_2} \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Given a parametrization $\boldsymbol{\theta} = (\theta_1, \theta_2)$ of the splitting function, we define the left and right child sets of $\mathcal{S}_i$ as:

$$\begin{aligned} \mathcal{S}_i^L(\boldsymbol{\theta}) &= \{\mathbf{x} \in \mathcal{S}_i \mid h(\mathbf{x}; \theta_1, \theta_2) = 1\} \\ \mathcal{S}_i^R(\boldsymbol{\theta}) &= \{\mathbf{x} \in \mathcal{S}_i \mid h(\mathbf{x}; \theta_1, \theta_2) = 0\}. \end{aligned} \tag{3}$$

By construction, the splitting function $h(\mathbf{x})$ guarantees that $\mathcal{S}_i^L(\boldsymbol{\theta})$ and $\mathcal{S}_i^R(\boldsymbol{\theta})$ partition $\mathcal{S}_i$. The quality of splits is measured by the *information gain* in the resulting subsets with respect to the characteristic being estimated:

$$I_i(\boldsymbol{\theta}) = H(\mathcal{S}_i) - \sum_{j \in \{L,R\}} \frac{|\mathcal{S}_i^j(\boldsymbol{\theta})|}{|\mathcal{S}_i(\boldsymbol{\theta})|} H(\mathcal{S}_i^j(\boldsymbol{\theta})), \tag{4}$$

where $H(\mathcal{S})$ is the characteristic entropy in set $\mathcal{S}$: $H(\mathcal{S}) = -\sum_{y \in \mathcal{Y}^c} p(y|\mathcal{S}) \log p(y|\mathcal{S})$ for the discrete estimate of $p(y|\mathcal{S})$ from set $\mathcal{S}$: $p(y|\mathcal{S}) = \frac{|\{\mathbf{x}_i \in \mathcal{S} \mid y_i = y\}|}{|\mathcal{S}|}$. Using information gain to evaluate splits produces trees in which the entropy of the class distributions associated with the nodes decreases when descending in the tree, and thus prediction confidence increases.

Trees are built by randomly sampling the parameter space $\boldsymbol{\theta} = (\theta_1, \theta_2)$ at each node $\mathcal{S}_i$ in the tree. At each internal node, T random parameters $\boldsymbol{\theta}_t$ for $t = \{1, \ldots, T\}$ are generated, and the split resulting in the highest information gain is chosen:

$$\boldsymbol{\theta}_i^* = \arg\max_t I_i(\boldsymbol{\theta}_t). \tag{5}$$

The number of tests T controls the randomicity of the resulting tree. Clearly, if $T = d(d-1)$ the best overall split at each node will be selected and there will be no randomicity in the resulting tree. Child nodes are added until a maximum depth D is reached in the tree, or until a minimum number of training elements remains in the set $\mathcal{S}_i$ at node i.

Estimation with forests of random trees. A random tree $\mathcal{T}$ is thus defined by the split parameters at each internal node:

$$\mathcal{T} = \{\boldsymbol{\theta}_i\}_{i=1}^{i=|\mathcal{T}|}. \tag{6}$$

To ensure diversity and avoid overfitting possible with a single tree, a forest of trees is defined as $\mathcal{F} = \{\mathcal{T}_1, \ldots, \mathcal{T}_F\}$. The trees in $\mathcal{F}$ are trained independently and their outputs averaged to provide some sort of regularization.

Given an unlabeled test sample $\mathbf{x}$, we use the hierarchy of tests defined by each $\mathcal{T}$ to determine to which leaf node it arrives. Denoting the leaf node that $\mathbf{x}$ arrives to in tree $\mathcal{T}$ as $l(\mathbf{x}; \mathcal{T})$, we can estimate the unknown label y of $\mathbf{x}$ using the training elements $\mathbf{x}_i$ that arrive to the same leaf node:

$$\text{label}(\mathbf{x}; \mathcal{F}) = \arg\max_{y} \frac{1}{|\mathcal{F}|} \sum_{\mathcal{T} \in \mathcal{F}} p(y | L(\mathbf{x}; \mathcal{T})), \tag{7}$$

where $L(\mathbf{x}; \mathcal{T})$ is the set of training examples in the same leaf as $\mathbf{x}$ in tree $\mathcal{T}$:

$$L(\mathbf{x}; \mathcal{T}) = \{(\mathbf{x}_i, y_i) \in \mathcal{S}_0 \mid l(\mathbf{x}_i; \mathcal{T}) = l(\mathbf{x}; \mathcal{T})\}. \tag{8}$$

Multi-objective estimation with random forests. We now assume that each training sample $\mathbf{x}_i$ is labeled with $C > 1$ characteristics we wish to estimate:

$$\mathcal{S}_0 = \{(\mathbf{x}_i, \{y_i^j\}_{j=1}^C) \mid i = 1 \ldots N\} \tag{9}$$

If constructed properly, the tree model should learn how to condition the estimation of one characteristic on the estimation of another, thus simplifying the problem. The fundamental difference between single-objective and multi-objective estimation is in how the information gain driving the splitting process is defined. In particular, there is no guarantee that the information gain in one characteristic is comparable in scale with the information gain in another.

We define a new normalized measure of information gain for multi-objective random forests, it weights the information of each characteristic c by the ratio between the local entropy in $\mathcal{S}_i$ with respect to the root entropy $H_c(\mathcal{S}_0)$. This *locally weighted information gain* I_{lw} is defined as:

$$I_{\text{lw}}(\mathcal{S}_i, \boldsymbol{\theta}) = \max_{c} \frac{H_c(\mathcal{S}_i)}{H_c(\mathcal{S}_0)} I^c(\mathcal{S}_i, \boldsymbol{\theta}). \tag{10}$$

The main idea behind the definition of I_{lw} is to update weights during the training process in order to scale each characteristic information gain based on how much entropy remains at the current depth. During training, our approach selects the split function parametrization $\boldsymbol{\theta}$ which maximizes the *locally weighted information gain* I_{lw} according to one characteristic. Note that the characteristic c is selected automatically at each node of each tree. A detailed experimental analysis of our learning procedure is conducted in section 4.2.

4. Experimental results

We report on experiments comparing our approach to a state-of-the-art baseline [10] (based on KCCA) for the joint estimation of multiple soft biometrics and face characteristics. We evaluate performance on a dataset we create using a subset of images from the FERET [12] and CAS-PEAL [3] datasets.

	FERET [12]	**CAS-PEAL-R1** [3]	**MIX Dataset**
# Persons	994	1040	210
Pan Angles	$\{-90, -75, -67.5, -45, -22.5, -15, 0, +15, +22.5, +45, +67.5, +90\}$	$\{-45, -30, -15, 0, +15, +30, +45\}$	$\{-67, -45, -22, -15, 0, +15, +22, +45, +67\}$
Tilt Angles	0	$\{-45, 0, +45\}$	0
Age	19 years of birth	Y, M, O	Y, M, O
Ethnicity	9 ethnic groups	Asian only	5 ethnic groups
# Images per subjects	At least 5 by subject	~ 21	At least 5 by subject

Table 1: Characteristics of the FERET, CAS-PEAL and our MIX datasets.

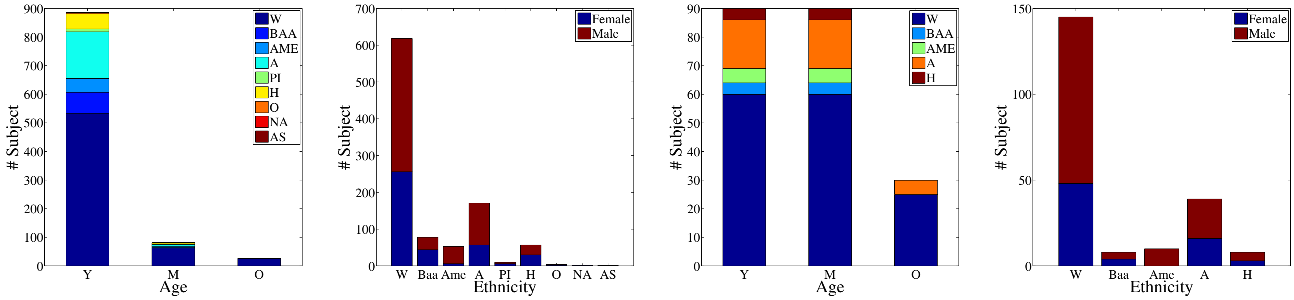
4.1. Datasets and experimental protocols

We evaluate our approach on a subset of the FERET [12] and CAS-PEAL-R1 [3] datasets. The main characteristics of these datasets are summarized in table 1. We first detail each of these datasets, and then describe how we merge the two of them to create a combined dataset with a richer and more balanced set of soft biometrics to estimate.

The FERET Dataset. The FERET dataset [12] is a benchmark for face-recognition algorithms. It is composed of images of 994 subjects (591 males and 403 females). It is composed of 11,338 images, annotated with 9 different ethnic groups, 12 different Pan angles, and year of birth.

The CAS-PEAL-R1 Face Dataset. The complete CAS-PEAL dataset [3] is one of the largest datasets (99,594 images of 1,040 subjects) for evaluating Gender recognition, head pose estimation, and face recognition methods. The dataset is composed only of Chinese persons (595 males and 445 females) imaged in different poses and with varying expressions, accessories, and lighting. The publicly available version, called CAS-PEAL-R1, contains 30,863 images of 1,040 subjects, with about twenty images of each person. Each image is acquired considering a combination of the Tilt and Pan angles reported in Table 1 and the Age is quantized into three classes Young (Y) from 10 to 44 years, Middle-Age (M) from 45 to 59 years and Old (O) above 60 years.

A dataset for multiple soft biometric estimation. Neither CAS-PEAL-R1 nor FERET alone are satisfactory for evaluation of multiple soft biometric estimation. For example there are only 4 "Old" subjects and 10 "Middle-Age" subjects in the CAS-PEAL-R1 dataset, and it also composed only of Asian subjects. In contrast, the FERET dataset contains very few "Old" subjects and 81 "Middle-Age" ones, and 9 different Ethnicity classes. As seen in Fig. 1(a), there are mostly "White/Young" persons, and few "Middle-Age" and "Old" ones. Additionally, as shown



(a) Ethnicity/Age (FERET) (b) Gender/Ethnicity (FERET) (c) Ethnicity/Age (MIX) (d) Gender/Ethnicity (MIX)

Figure 1: Characteristic distribution in the FERET and MIX datasets. See text for definition of MIX dataset.

in 1(b) Gender is not uniformly distributed within Ethnicity and some ethnicities have very few subjects.

We define a dataset that mixes images from the FERET and CAS-PEAL-R1 (which we call the "MIX" dataset in what follows). When joining the two datasets, we exclude images having a Pan angle of $\{-90, -75, -30, +30, +90\}$, because -90, -75 and $+90$ are present only in FERET and $\{-30, +30\}$ are present only in CAS-PEAL. We remove all the images with a nonzero Tilt angle, since FERET contains no images with nonzero tilt angles, and set to -67 and 67 the Pan angles -67.5 and 67.5, respectively. The images from 4 ethnic groups that have very few subjects were removed. We hence maintain the following 5 ethnicities: White (W), Black-or-African-American (BAA), Asian-Middle-Eastern (AME), Asian (A) and Hispanic (H). Note that the authors of [6] addressed the imbalance in ethnicities by maintaining only the White and Asian classes. Finally, we project the Age labels from the FERET dataset into the set of Age classes of CAS-PEAL: Young, Middle and Old, as these three classes are the only Age information provided in CAS-PEAL.

In our dataset the majority number of subjects are "White" or "Asian", and a significant part is composed by "Young" people as shown in Fig. 1(a). To avoid bias in classification results, we consider ten randomly drawn subsets of these subjects. Our random subset divisions consider the number of subjects for each Age class so that each subset contains 90 "Young" subjects, 90 "Middle-Age" subjects and 30 "Old" subjects, where the 90 "Young" are selected randomly from the mixed dataset, so that we have the same number of subject for each Ethnicity class both for "Young" and "Middle-Age" people. Each split contains from 1,858 to 1,995 images, randomly divided in half of the subjects for each Ethnicity class to form the training and test sets.

Experimental protocol. Given the random nature of our approach, and the ten different subsets, all results reported are averages over multiple runs of the algorithm (three trials for each split of the dataset). All trees used in these experiments have maximum depth of 9, which was found to work well in preliminary tests. Our random forest is composed of 200 trees. At each node of a tree, $T = 500$ parametrizations of the split function are randomly generated and the best one (according to Eq. 10) is selected.

We evaluate Pan and soft biometrics estimation in terms of recognition accuracy, and in terms of mean average error (MAE). While recognition accuracy and mean average error are commonly used for performance evaluation, we feel that – especially for soft biometrics in unbalanced datasets – precision and recall on individual biometrics is a better metric for evaluating performance. We thus also give precision/recall plots for all estimated soft biometrics.

4.2. Analysis of multi-objective learning

In Figs. 2(a), 2(b) and 2(c) we show, respectively, the evolution of the entropy, the information gain, and the ranking (in terms of information gain) of each estimated characteristic as a function of the depth in the random trees comprising a MORF. The curves were computed from the average of all of the trees of all trials using our information gain combination approach (Eq. 10) with 200 trees. At the first levels of the trees, the Pan characteristic has the highest entropy (see Fig. 2(a)) and indeed as we see in Fig. 2(c) Pan is the highest ranked splitting characteristic is and also results in the highest information gain (Fig. 2(b)). The early levels of each tree thus specialize in discriminating Pan angles and consequently implicitly condition the estimation of the other characteristics on Pan angle. Going deeper into the trees, the entropy of each soft biometric characteristic decreases gradually, and Age, Gender and Ethnicity begin to have more importance as splitting characteristics. Trees learn to specialize first on Pan, then Gender and Age, and finally on Ethnicity.

4.3. Comparison with the state-of-the-art

In this section we compare the performance of our proposed MORF approach for Pan and soft biometric esti-

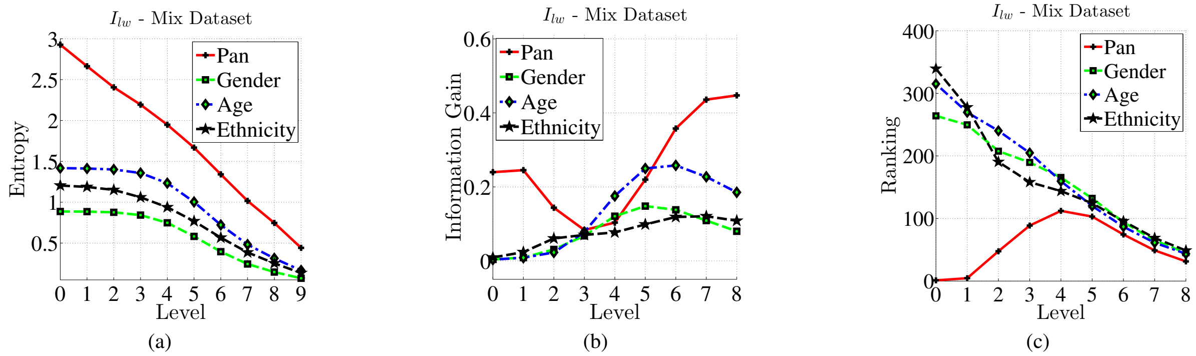


Figure 2: MORF performance on the MIX dataset. (a) Mean entropy per level for each characteristic. (b) Mean information gain of selected split per level. (c) Mean rank (in terms of information gain) of selected split per level.

	Pan		Age		Gender		Ethnicity	
Approach	Accuracy	MAE	Accuracy	MAE	Accuracy	MAE	Accuracy	MAE
MORF	**63.56%**	**10.65**	**58.89%**	**0.43**	79.86%	1.41	**78.42%**	**0.56**
CCA	31.27%	19.29	47.91%	0.58	76.28%	1.66	43.71%	1.01
KCCA	40.30%	14.59	57.12%	**0.43**	**83.63%**	**1.15**	40.04%	0.89

Table 2: Comparison between our approach and the state of the art on Pan, Age, Gender and Ethnicity estimation.

mation with those of the Canonical Correlation Analysis (CCA) and Kernel CCA (KCCA) approaches reported in [6]. These approaches have the same objective as ours: to reliably and simultaneously estimate multiple face characteristics. They work by learning a subspace in which correlation between image features and desired characteristics is maximized, and then fitting a least squares prediction model from projected image features and characteristics. We evaluate baselines with standard linear CCA, and KCCA with a radial basis kernel.

We give accuracy and mean absolute error (MAE) performance of our methods and both baselines in Table 2. We give Mean Average Precision (MAP) for all methods and for each label separately in Table 3. We plot in Fig. 3 the precision/recall curves of our approach compared with CCA and KCCA for: age (Figs. 3(a) and 3(b)), gender (Figs. 3(c) and 3(d)), and ethnicity estimation (Figs. 3(e) and 3(f)).

Pan estimation. We compare the Pan estimation performance of our approach with the state-of-the-art in the first two columns of Table 2 in terms of accuracy and MAE. The MAE measures how close predictions are to the expected outcomes. Our approach outperforms the state-of-the-art baselines by a significant margin. The Pan characteristic has 9 different labels ($-67, -45, -22, -15, 0, +15$, $+22, +45, +67$), this high number of labels may be an issue for the CCA and KCCA baselines.

Age estimation. We obtain similar or slightly higher performance with respect to the CCA and KCCA baselines. Accuracy and MAE are given in Table 2. We obtain an accuracy of 58.89%, which is 1.7% higher than the KCCA baseline. In the first three rows of Table 3 we report MAP performance on Age estimation. We plot in Fig. 3 the precision/recall curves of our approach compared with CCA and KCCA baseline. The trend is similar for our approach and the two baselines, but our curves tend to stay above the CCA curves (except for the Old class) and a bit above the KCCA curves (except for the Young class for lower recall values).

Gender estimation. Our MORF approach gives lower performance that the KCCA baselines both in terms of accuracy and MAE as reported in Table 2. However MORF outperforms the CCA baseline on these two metrics. From Table 3 where per class MAP are given and from the Precision/Recall curves in Fig. 3(c), we can see our method works better on female class with respect to the CCA, while in Fig. 3(d) we can see that the performance on the Female class is lower than KCCA. For all methods the performance on the female class is lower, this could be related to the fact that only one third of the subjects in our dataset are females.

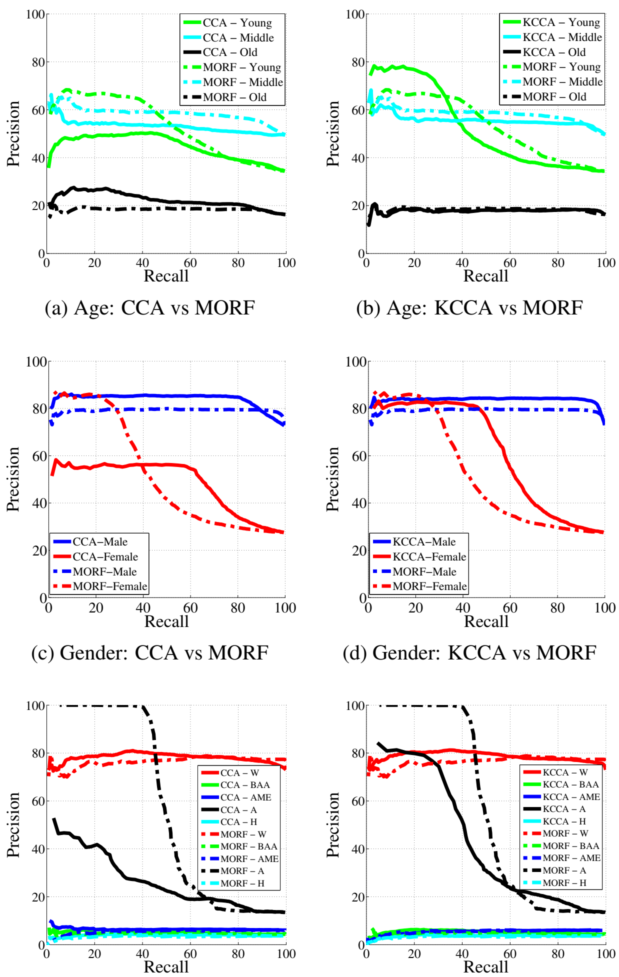
Ethnicity estimation. We significantly outperforms the baselines on Ethnicity estimation. MORF obtains an accuracy of 78.42% while the CCA and KCCA baseline are around 40%, see Table 2. In terms of MAP, results in Table 3, the performances of our approach are comparable with the baselines for most of labels except for the Asian class where our approach performs much better. This is confirmed by the precision/recall plot in Figs. 3(e) and 3(f), where the Asian curve of our approach is much higher.

5. Discussion

In this paper we described a technique for simultaneously estimating multiple facial characteristics. We proposed a new normalized measure of multi-objective information gain that is used with our Multi-Objective Random

Characteristic	Label	MORF	CCA	KCCA
Age	Y	**54.66%**	46.15%	53.75%
	M	**59.20%**	53.74%	56.53%
	O	19.64%	**23.85%**	19.09%
Gender	F	53.75%	50.09%	**63.07%**
	M	80.69%	84.76%	**85.08%**
Ethnicity	W	78.65%	79.55%	**79.86%**
	BAA	5.64%	**7.62%**	6.55%
	AME	6.33%	**7.62%**	6.55%
	A	**57.73%**	27.85%	45.63%
	H	4.08%	**4.94%**	4.35%

Table 3: Comparison with the state-of-the-art in terms of MAP on Age, Gender and Ethnicity estimation.



(a) Age: CCA vs MORF

(b) Age: KCCA vs MORF

(c) Gender: CCA vs MORF

(d) Gender: KCCA vs MORF

(e) Ethnicity: CCA vs MORF

(f) Ethnicity: KCCA vs MORF

Figure 3: Precision/Recall plots of Age, Gender and Ethnicity estimation compared with CCA and KCCA.

Forests (MORF) framework for estimation of multiple characteristics from a single feature representation. On average, MORF outperforms subspace methods like KCCA and CCA for simultaneous estimation of multiple biometrics.

We feel that simultaneous estimation of characteristics is an interesting direction for future research. Ongoing work is focused on better features for soft biometric estimation, and on addressing problems where many characteristics (such as attributes) can be estimated for each sample. Methods to overcome the imbalance of the training sets are also something we plan to study.

Acknowledgments

A. D. Bagdanov acknowledges the support of a Ramón y Cajal Fellowship (RYC-2012-11776).

References

[1] B. Benfold and I. Reid. Unsupervised learning of a scene-specific coarse gaze estimator. In *Proc. of ICCV*, 2011.

[2] A. Criminisi, J. Shotton, and E. Konukoglu. Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Foundations and Trends® in Computer Graphics and Vision*, 7(2):81–227, 2012.

[3] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao. The cas-peal large-scale chinese face database and baseline evaluations. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans.*, 38(1):149–161, 2008.

[4] A. Gunay and V. Nabiyev. Automatic age classification with lbp. In *Proc. of ISCIS*, 2008.

[5] G. Guo, C. R. Dyer, Y. Fu, and T. S. Huang. Is gender recognition affected by age? In *Proc. of ICCVW*, 2009.

[6] G. Guo and G. Mu. A framework for joint estimation of age, gender and ethnicity on a large database. *Image and Vision Computing*, 32(10):761 – 770, 2014.

[7] M. A. Haj, J. Gonzalez, and L. Davis. On partial least squares in head pose estimation: How to simultaneously deal with misalignment. In *Proc. of CVPR*, 2012.

[8] C. Huang, X. Ding, and C. Fang. Head pose estimation based on random forests for multiclass classification. In *Proc. of ICPR*, 2010.

[9] S. Karaman, A. D. Bagdanov, G. D'Amico, L. Landucci, A. Ferracani, D. Pezzatini, and A. Del Bimbo. Passive profiling and natural interaction metaphors for personalized multimedia museum experiences. In *Proc. of ICIAP*. 2013.

[10] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and simile classifiers for face verification. In *Proc. of ICCV*, 2009.

[11] H. Lian and B. Lu. Multi-view gender classification using local binary patterns and support vector machines. *Advances in Neural Networks-ISNN*, 3972:202–209, 2006.

[12] P. Phillips, H. Moon, S. Rizvi, and P. Rauss. The feret evaluation methodology for face-recognition algorithms. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.

[13] N. Ramanathan, R. Chellapa, and S. Biswas. Age progression in human faces: A survey. *Journal of Visual Languages and Computing*, 15:3349 – 3361, 2009.

[14] K. Ricanek and T. Tesafaye. Morph: a longitudinal image database of normal adult age-progression. In *Proc. of FGR*, 2006.

[15] Z. Yang and H. Ai. Demographic classification with local binary patterns. In *Advances in Biometrics*, volume 4642, pages 464–473. Springer Berlin Heidelberg, 2007.

[16] J. Zheng and B. Lu. A support vector machine classifier with automatic confidence and its application to gender classification. *Neurocomputing*, 74(11):1926–1935, 2011.