Demographic Classification: Do Gender and Ethnicity Affect Each Other?

Giovanna Farinella
EURECOM,
Sophia Antipolis, France
Email: giovanna.farinella@eurecom.fr

Jean-Luc Dugelay
EURECOM,
Sophia Antipolis, France
Email: jean-luc.dugelay@eurecom.fr

***Abstract*—Gender and ethnicity classification are challenging topics in the field of face analysis. Some features, like skin color, are relevant only for ethnicity but not for gender; some others, like face geometry, are important for both. The impact of ethnicity in gender perception, as the effect of gender on ethnicity disambiguation, is not clear. This paper provides a study to check if gender and ethnicity affect each other during the classification. Three different well-established algorithms have been implemented to provide significant experiments. These algorithms are used for both gender and ethnicity classification. Ethnicity-specific gender classifiers are trained and tested using faces from a specific ethnicity; the accuracies achieved are compared with the ones obtained using generic gender classifiers, trained and tested with faces from different ethnic groups. With a similar procedure we compare gender-specific ethnicity classifiers, trained and tested selecting faces with a specific gender, with generic ethnicity classifiers, trained and tested with both male and female faces. The study shows that specific and generic classifiers perform equally. That means, for the features selected, gender and ethnicity do not affect each other.**

I. Introduction

The demographic classification, such as gender, ethnicity or age estimation, has been attracted a lot of interest in the last ten years, finding applications in many fields such as forensic art, electronic customer relationship management, surveillance monitoring ([1]), biometrics ([2]) and video indexing ([3]).

Many efforts have been spent to study sexual and race discriminative characteristics in human faces, not only in the field of computer vision, but even from an anthropologic point of view. Some discriminative characteristics, like skin color, allow us to distinguish different ethnic groups, but do not help in the gender disambiguation. Some characteristics, like the presence of beard, help in the gender disambiguation but are irrelevant for ethnicity. Some other features, like face geometry, are important for both gender and ethnicity perception. The impact of ethnicity on the gender recognition, as the impact of gender on ethnicity classification is hard to express. Our aim is to check if, using benchmark algorithms for demographic classification, gender and ethnicity affect each other.

To study the influence of ethnicity in gender classification we compare ethnicity-specific gender classifiers with generic gender classifiers. An ethnicity-specific gender classifier is trained and tested using faces from a specific ethnicity. A generic gender classifier is trained and tested using faces from different ethnic groups.

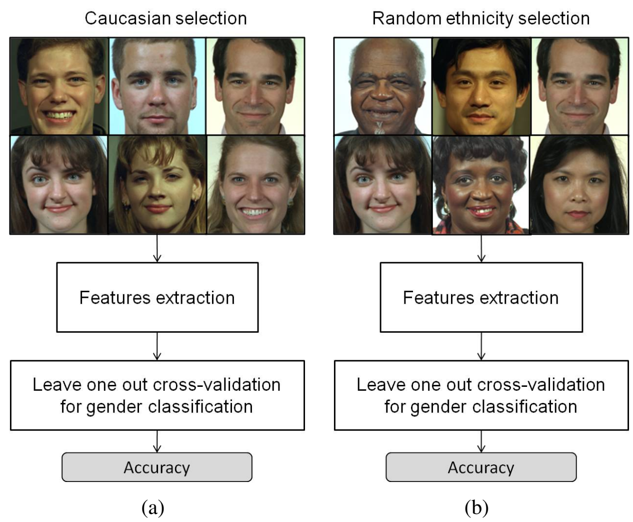

Fig. 1. (a): Caucasian-specific gender classification. (b): Generic gender classification.

According to a similar protocol, to study the influence of gender in ethnicity classification we compare gender-specific ethnicity classifiers with generic ethnicity classifiers. Figure 1 shows the general idea.

If the classification performances, obtained using specific and generic classifiers, have significant differences, we can deduce that, gender and ethnicity affect each other during the classification; in that case the correct way to classify sex and race should take into account the interdependences. Otherwise, if the performances obtained are close together, we conclude that, for the considered algorithms, the ethnicity of a subject does not matter while evaluating the gender and/or, vice versa, that the gender of the subject does not matter while evaluating the ethnicity; that means that these two traits could be evaluated independently.

The study is drafted testing three different methods widely used in the state-of-art for both gender and ethnicity classification. Support Vector Machines are used for the classification with Leave One Out protocol.

This paper is organized as follows: Section II reports related works; in Section III some considerations on human perception and computer vision are presented; in Section IV the classification methods used in the experiments are described. Section V illustrates the experimental set up; results are shown in Section VI. Finally, in Section VII conclusions are presented.

II. Related Work

Many efforts have been spent in the last years for gender classification, a little bit less for ethnicity. Table I and Table II summarize the state of the art for gender and ethnicity classification respectively. Each table compares some existing classification techniques evaluated on several benchmark databases. Some of these works test a set of features on a set of classifiers; in this case we choose to report the maximum classification rate obtained for each classifier. When more than one database is tested separately, we decided to report the classification rate just for one of them. In this case, the name of the database is indicated near the classification rate. When one database is used for the training and another one is used for the test, the following notation is used: (Train/Test). If the work consider both 2D and 3D classification problems, only results for the 2D are reported.

Golomb et al. with SEXNET are the pioneers of gender classification [4]. After, different methods have been tried. Most of them use simple features like pixel intensity ([4],[5]), LBP([6]) or haar-like([7]) as input for learning machines such as SVM([8],[5]), Neural Network([9],[10]) and Adaboost([11],[12]).

To the best of our knowledge, less efforts have been spent for ethnicity classification. In [12] LLBPH is used with an Adaboost classifier. In [1] both pixel intensity values and Biologically Inspired Model (BIM), features are used for soft biometry computation. In [13] the Linear Discriminant Analysis (LDA) based scheme is presented for the two-class (Asian vs. non-Asian) ethnicity classification task.

The investigation of interdependencies among different classes is a quite unexplored field. In [14] it has been shown that a gender classifier trained on a dataset with limited demography does not work well if the test data contain more general samples. This result suggests the existence of a correlation between different demographic variables such as gender, ethnicity and age. Ethnic factor in gender classification have been taken under consideration by Gao and Ai in [15]; they built an ethnicity-specific gender classifier based on Adaboost; the specific classifiers outperform the generic ones only in some cases making the results hard to interpret.

Our aim is not to design a new classifier, but to better understand the interaction between gender and ethnicity. We have studied the behavior of ethnicity-specific gender classifiers and gender-specific ethnicity classifiers, building them according to different algorithms and comparing performances with generic classifiers.

III. Computer Vision and Human Perception

For a human brain it is, generally, extremely easy to distinguish male from female; however, the same task, can be hard for an automatic system.

Fig. 2. Illusion of Sex: the face with more contrast is perceived as female [22].

People are not consciously aware of which features let them to disambiguate the gender. Some sexual differences have been found in the face shape, however geometry is not the only information we use. A very interesting work is provided by Russel [23]: by experimenting with an androgynous face, he learnt that faces can be manipulated to appear female by increasing facial contrast around eyes and lips or to appear male by decreasing the contrast. Figure 2 shows the 'illusion of sex' experiment.

To solve the gender classification task, human brain and computers do not work in the same way. A remarkable prove of that is given by the importance of the hair style; Makinen at al. in [7] demonstrated that larger face areas with hair do not produce better classification on FERET database; however, for human perception, this information is very important. To prove this assessment we invite the readers to try the following short experiment: Figure 3 shows faces cropped from FERET database; try to assign a sex to the subjects; now, please move at the end of the paper and look at entire images displayed in Figure 6; are your previous choices correct? Most of the people are not able to classify correctly the gender looking at the cropped version.

Fig. 3. Gender perception experiment. Try to assign a sex to the subjects; now, please move at the end of the paper and look at entire images displayed in Figure 6; are your previous choices correct?

Regarding the ethnicity classification, the problem is even more complex. The first sensitive point is in the definition of different ethnic groups. This problem has been studied for years by anthropologies and it seems to be quite impossible to find a commonly accepted solution. The fact is that boundaries between different ethnicities are fuzzy (Figure 4), it is necessary to impose a threshold on the level of differences required to classify people as belonging at different ethnicities [24].

TABLE I
SUMMARY AND COMPARISON OF GENDER CLASSIFICATION TECHNIQUES

Related work	Database	Data description		Features	Method	(Max) Classification Rate
		Subject #	Image #			
Golomb et al. 1990 [4]	Private		90	Pixel-based	Neural Network	91.9%
Brunelli et al. 1995 [10]	Private	40(20M,20F)	168	Geometrical Fetures	HyperBF Network	79%
Abdi et al. 1995 [16]	Private			Pixel-based	RBF Network	90%
Gutta et al. 1998 [17]	FERET	1009	3006	Binary code given by pixel comparison	Neural Network plus Decision Tree	96%
Moghaddam 2002 [5]	FERET		1755 (1044M,711F)	pixel-based	SVM+RBF	95.3%
Sun et al. 2002 [9]	Private	400	400	GA is used to select select a subsets features	Bayes classifier Neural Network SVM LDA	86.7% 88.7% 95.3% 91%
Baluja et al. 2005 [8]	FERET	994(591M, 403F)	2409	Pixel-based	SVM Adaboost	93,50% 94,40%
Lian et al. 2006 [18]	CASPEAL	1040	14384	LBP	SVM+RBF	96.75%
Lu et al. 2006 [19]	UND MSU	276 100	944 296	Pixel-based	LDA	86% (UND+MSU)
Yang et al. 2007 [12]	FERET PIE	1196 68	3540 696	LBP	Adaboost	88.1% (FERET)
Makinen et al. 2008 [11] [7]	FERET WWW	900(450M,450F) 4720(2360M,2360F)	900 4720	Pixel-based LBP Haar-like	SVM+RBF Mean Adaboost LUT Adaboost T-Adaboost Neural Network	92.00% (FERET) 90.00% (FERET) 93.33% (FERET) 90.00% (FERET) 92.22% (FERET)
Demirkus et al. 2010 [1]	Different public face databases and WWW	1458(635 F,823 M)	1458	pixel-based BIM	SVM	89.4%
Alexandre 2010 [20]	New taken by: FERET-UND	487(321M,186F)	487	LBP	SVM	99.07% (FERET)
Chu et al. 2010 [21]	FERET MORPHO	900(450M,450F) 418	900 55843	pixel-based	Similarity between subspaces with LDA	95.5% (FERET)
Bekios-Calfa et al. 2011 [14]	UCN FERET PAL	10669(5,628M,5,041F) 994(591 M, 403 F) 576(219M,357M)	10669 994 576	pixel-based	LDA	88.72% (UCN/FERET)
Ylioinas et al. 2011 [6]	FRGC2.0 FERET XM2VTS		28183 3083 2312	LBP+VAR	SVM	97.61%

TABLE II
SUMMARY AND COMPARISON OF ETHNICITY CLASSIFICATION TECHNIQUES

Related work	Database	Data description		Label	Features	Method	(Max) Classification Rate
		Subject #	Image #				
Gutta et al. 1998 [17]	FERET	1009	3006	Caucasian, Asian Oriental, African	Binary code given by pixel comparison	Neural Network plus Decision Tree	94%
Lu et al. 2004 [13]	Yale+AR+NLPR+ AsianPF01	263	2630	Asian, non Asian	Pixel-based	NN LDA	93.8% 96.0%
Lu et al. 2006 [19]	UND MSU	276 100	944 296	Asian, non Asian	Pixel-based	LDA	96.8%
Yang et al. 2007 [12]	FERET PIE	1196 68	3540 696	Asian, non Asian	LBP	Adaboost	92,89%
Demirkus et al. 2010 [1]	Different public face databases and WWW	600 (200 each class)	600	African American Caucasian Asian	pixel-based BIM	SVM	92%

Maybe the oldest and the most widely used classification is the one which divide the population in three broad groups: 'Caucasoid', 'Mongoloid' and 'Negroid', however the within-group to between-group variation is very high, that means that individuals from one race may be more similar to individuals from other races than to other individuals belonging to the same race [25].

Under an anthropological point of view, except people closed to the boundaries, the differences among ethnic groups are evident. Some of these differences involve characteristics that are independent from the gender, as the skin color; for some others, as the geometrical proportions among facial points, there is a correlation with gender.

Humans are affected by the well know cross-race effect which is the tendency for people of one race to have difficulty recognizing and processing faces from other ethnic groups [26]. In computer vision, the most widely used approach, for both gender and ethnicity classification in 2D, involves the usage of learning machines with global descriptors. It is reasonable to consider the possibility that, training and testing the machine selecting one trait while evaluating another one (for example selecting a specific ethnicity while evaluating the gender), performances can improve. However, it is not so obvious: it will depend on the criteria used by the learning machine, which is, in most cases, unknown using global descriptors (as pixel-based features). The point is, if a large variability in ethnicity has a negative impact on gender classification, to manage the problem correctly we should take it into consideration. The same goes for ethnicity classification: if gender factor affects the classification we should consider somehow this point. Otherwise, if a classifier does not suffer for these kind of lateral constraints, than the two problems can be solved independently.

IV. Technical Background

The classification methods used in the experiments are described next. Three different sort of features are selected and used as input for an SVM. The features, selected according to state-of-art algorithms, are Pixel-Based (PB), Local Binary Patterns (LBPs) and Histogram of Oriented Gradients (HOG).

A. The Features

- **Pixel-Based (PB).** The first category of features, used in our tests, is the simple pixel-intensity values [5]. The raw value of pixels is extracted for different image resolutions: 64x64, 48x48 and 32x32.

- **Local Binary Patterns.** LBP [28] is a gray-scale invariant texture operator. It is obtained by thresholding the neighborhood of each pixel with the value of the central one. Originally introduced for texture classification, it has been largely used in face recognition and classification. In our experiments the LBP features are extracted from 64x64 images. Each image is divided in 8x8-blocks and filtered with the basic LBP operator with four neighbors at the radius one ($LBP_{4,1}$). This procedure generates a 16-bin histogram for each block. The whole image is, also, filtered with the uniform LBP with eight neighbors at the radius one ($LBP_{8,1}$), generating a 59-bin histogram. Then, all the obtained histograms are concatenated together. This method differs from the one used by Lian and Lu [18] for the presence of the LBP filter on the whole image; this variation has been proposed by Makinen and Raisamo in [7].

- **Histogram of Oriented Gradients.** Histogram of the Oriented Gradients (HOG) has been proposed by Dalal and Triggs for pedestrian detection in static images [29]. It is broadly used in computer vision, especially in the field of object detection. HOG has been recently used by Guo et al. [30] for gender classification purpose. In our experiment the HOG over 64x64 images is evaluated using 8 cells and 9 histogram channels.

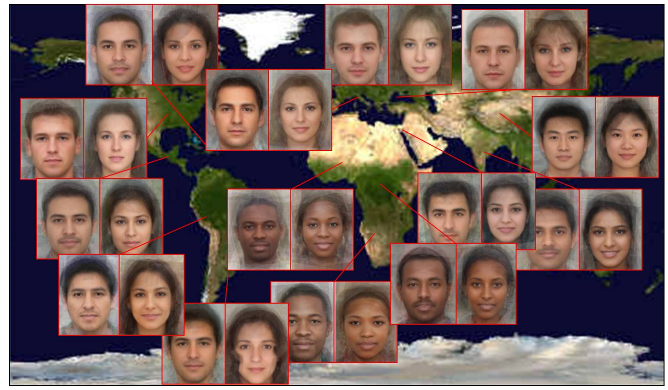

Fig. 4. Average faces created for the 'World of Averages' project [27] ; these faces have been obtained assembling several images to create computer composites of the average face of citizens of different parts of the world.

B. The SVM Classifier

It has been proved that SVM classifier is the most efficient technique for demographic classification tasks [11]; Adaboost and neural network classifiers may be preferred when the computational speed is mandatory, while LDA has been shown to be efficient when the amount of data is limited [14]. The idea of the SVM is that different classes can be disambiguating in a higher dimensional space using transformed features. Many different kernel functions have been proposed for the transformation, but the RBF is one of the most widely used.

V. Experimental Set Up

A. Databases

The following databases have been used for the experiments:

- *FERET ([31],[32]).* It is a database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office. This database contains several gray scaled and

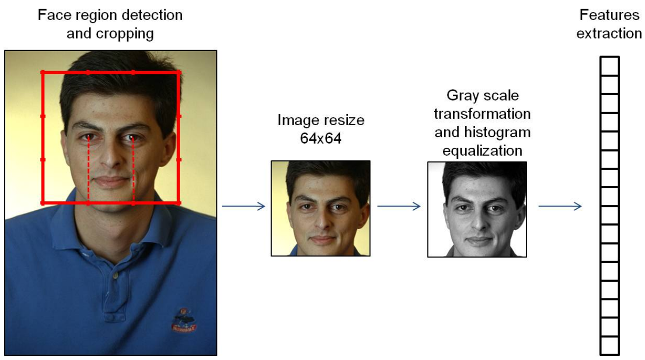

Fig. 5. Pre-processing and feature extraction procedure.

colored images of 994 individuals; a ground truth provides several information as facial points, date of birth, ethnicity, gender, light conditions and more.

- *TRECVID ([3]).* The TREC conference series is sponsored by the National Institute of Standards and Technology (NIST) with additional support from other U.S. government agencies. The TRECVID database consists of numerous video shots taken from the web. Two-hundred face images have been taken from TRECVID database and used to tune the SVMs.

As a general rule for our experiments, we selected one frontal image for each individual. It is worth to notice, also, that we chose to use TRECVID for the tuning to cover a significant variations in terms of resolution and illumination conditions. Instead, to test the SVMs, we have decided to use the FERET benchmark database, to make the experiments easier to reproduce.

B. Pre-processing

Prior to extract the features, some pre-processing steps are needed. First of all, for each image, the face region is cropped. The face region is a square, automatically cropped according to the face geometry: it is, indeed, centered relatively to the eyes and the borders size are three times the distance between them. This distance is evaluated as difference between the two eyes x-coordinates saved in the ground truth.

After the face extraction, all the images are resized at 64x64 resolution, converted in a gray level color map and histogram equalized as in [11] and [7] . Figure 5 shows pre-processing and feature extraction procedure.

C. Experimental Protocol

As discussed before, differently from gender classification, no clear division for ethnicity perception exists. In our experiments we decided to split the database in Caucasian people and non-Caucasian people, considering ethnicity classification as a binary problem similar to [13], [19] and [12]. We chose this splitting rule in order to guarantee a similar distribution among the two classes. Indeed if we had split into Asian and non-Asian, the number of Asian samples would not have been big enough, on the other hand if we were added another database with more Asian faces, ethnicity selection would have been turn into a database selection compromising the reliability of the experiments. Table III shows the distribution of images among the classes.

TABLE III
DISTRIBUTION OF IMAGES AMONG CLASSES.

	# Male	# Female	# tot
# Caucasian	316	234	550
# non-Caucasian	172	127	299
# tot	488	361	

Three kinds of features are extracted: PB, LBP and HOG. The pixel intensity values are extracted from different image size: 64x64, 48x48 and 32x32. All these five set of features (PB in three resolutions, LBP and HOG) are used as input for an SVM with RBF kernel.

In our experiments LIBSVM [33] is used to train and test SVMs following Leave One Out protocol. LOO strategy is the best choice when the number of available samples is limited.

The best parameters for gender classifiers and ethnicity classifiers have been evaluated using the grid search system provided in LIBSVM. Two hundred face images from TRECVID have been used for this purpose. Table IV shows the tuning step results, i.e. the parameters that have been used for gender and ethnicity classifiers, for each algorithm tested. For more datails on parameters see [33].

TABLE IV
OPTIMUM PARAMETERS ESTIMATED DURING THE TUNING STEP.

	Gender classifier (100M,100F)	Ethnicity classifier (100C,100NC)
PB : 64x64	c= 2 g=0,0078	c= 32 g=0,00012
PB : 48x48	c= 2 g=0,0078	c= 8 g=0,00048
PB : 32x32	c= 2 g=0,0312	c= 32 g=0,00048
LBP	c= 8 g=0,0125	c= 8 g=0,125
HOG	c= 2 g=0,0312	c= 32 g=0,00195

1) Ethnicity impact on gender classification: In order to study if ethnicity affects the gender classification we run two experiments. In the first experiment we provide a comparison between Caucasian-specific gender classifier and generic gender classifiers; the second experiment provides a comparison between non-Caucasian-specific gender classifier and generic gender classifiers. We systematically follow the same procedure for the experiments.

For example, to test the Caucasian-specific gender classifier, Caucasian faces are selected; in this selection we have a total of 316 males and 234 females. The classification algorithms are run over this partition. To provide a comparison we evaluate the performances of twenty generic gender classifiers; these classifiers are trained and tested using subjects from different ethnicities. The images are extracted randomly but the number of males and females is forced to be 316 and 234 respectively. In that way we secure the independence with respect to the number of samples. The same protocol is used to evaluate non-Caucasian-specific classifiers.

2) Gender impact on ethnicity classification: Two experiments are run in order to study the impact of gender on ethnicity classification. The first provides a comparison between Male-specific ethnicity classifier and generic ethnicity classifiers; in the second we provide a comparison between Female-specific ethnicity classifier and generic ethnicity classifiers.

To test the Male-specific ethnicity classifier, Male faces are selected and the ethnicity classification algorithms are run over this partition. We, also, create twenty generic ethnicity classifiers trained and tested with both Female and Male; the number of Caucasian and non-Caucasian is forced to be the same that we have in the Male-specific classifier. With similar procedure we evaluate how a Female-specific ethnicity classifiers work with respect to a generic one.

VI. Results

Table V and Table VI show the results for gender and ethnicity classification respectively. For the generic classifiers an accuracy range is reported, i.e. the lowest and the highest accuracies obtained among the twenty generic classifiers tested in each experiment. In the brackets we have reported the number of images for each class to disambiguate, i.e. male (M) and female (F) for gender classifiers, and Caucasian (C) and non-Caucasian (NC) for the ethnicity classifiers. The accuracy is reported for the three algorithms that have been tested: Pixel-based with trhee different resolutions (PB: 64x64, PB: 48x48, PB: 32x32), Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP).

Looking at Table V we notice that, using generic classifiers, the performances do not decrease. These results are consistent with the ones obtained by Gao et al. using Probabilisting Boosting Tree [15]. Selecting Caucasian people, the best accuracy, obtained using LBP features, is 93.8%. The LBP average accuracy, over the twenty generic classifiers used to provide the comparison, is 94.8%. Selecting non-Caucasian people the best accuracy, obtained using HOG features, is 92%. The HOG average accuracy, over the twenty generic classifiers used to provide the comparison, is 93.4%. For all the other methods tested, we have similar results: looking at the resulting performances, it is clear that a large variation in ethnicity do not negatively affect the gender classification, but even, it seems it could enrich the training information content, helping in the gender classification task.

Looking at Table VI we observe that, for ethnicity classification the gender selection is not relevant for the classifiers: the accuracies achieved are comparable with the ones obtained using generic classifiers with both male and female. Selecting Male the best accuracy, obtained with HOG, is 87% and the average accuracy over the twenty generic classifiers using the same method is 87.6%. For the Female selection, PB method work better with an accuracy equal to 82% using images of size 64x64. The average accuracy obtained with the generic classifiers is 83.6%.

In both gender and ethnicity classification it is possible to deduce that when less images are used for the experiment, the accuracy decrease significantly. For gender classification experiments we use a total of 550 images in the first experiment (Caucasian-specific vs generic) and 299 for the second experiment (non Caucasian-specific vs generic); in average we have an accuracy loss of 4%. For ethnicity classification experiments we use a total of 488 images in the first experiment (Male-specific vs generic) and 361 for the second experiment (Female- specific vs generic); the accuracy loss is around 2%.

Looking at the results it is possible to notice that, using these methods, it is easier to classify the gender than the ethnicity. The best average results for gender is obtained using LBP and is 94.8%, for the ethnicity the highest accuracy, obtained using LBP, is 88.8%. Two possible reasons are the follows. First, we have to consider that all the images have been gray scaled in order to use all the FERET images without selecting just the colored ones. This procedure is widely used in the state of the art at least for gender classification ([11], [7], [6]). The color information can be important for the ethnicity, however our purpose was to study the impact of gender on ethnicity classification and, as we know, the skin color can disambiguate the race regardless the gender; for these reason the color information removal does not have any impact on the effectiveness of our study. The second reason is given by the high ethnicity within-group to between-group variation: individuals from one class may be more similar to individuals from the other class than to other individuals belonging to the same class ([25]). This problem cannot be solved because is intrinsic to the definition of ethnicity.

LBP and HOG have similar performances and they are more effective respect to the PB methods for both gender and ethnicity classification. In any case, the differences in the classification rates between the methods are small. This evidence is consistent with the results obtained by Makinen et al. in [7].

The fact that specific and generic classifiers perform equally is unexpected and very interesting. It means that, at least for the algorithms that have been tested, gender and ethnicity do not affect each other during the classification. It is worth to remember that we chose to implement some of the most widely used algorithms; these results suggest that, with these methods, it is not required to take care of ethnicity while evaluating the gender and, vice versa, to classify the ethnicity it is not necessary to estimate the gender. The two problems can be solved independently, since we have demonstrated that there is no interaction.

From a human point of view these results might appear surprising, but it is not if we consider that we are using machine learning with texture features. We have already discussed the fact that some characteristic are important for

TABLE V
ACCURACY FOR GENDER CLASSIFICATION USING SPECIFIC AND GENERIC CLASSIFIERS.

	PB: 64x64	PB: 48x48	PB: 32x32	LBP	HOG
Caucasian-Specific Gender classifier (316M,234F)	89.3%	90.2%	89.3%	93.8%	92.9%
Generic Gender classifiers (316M,234F)	[94.0, 96.9]%	[94.3, 97.1]%	[93.6, 96.7]%	[93.2, 97.4]%	[94.9, 97.8]%
non-Caucasian-Specific Gender classifier (172M,127F)	86.3%	88.3%	87.3%	86.6%	92.0%
Generic Gender classifiers (172M,127F)	[86.6, 92.3]%	[89.6, 94.3]%	[87.3, 92.9]%	[90.3,94.9]%	[92.6, 95.3]%

TABLE VI
ACCURACY FOR ETHNICITY CLASSIFICATION USING SPECIFIC AND GENERIC CLASSIFIERS.

	PB: 64x64	PB: 48x48	PB: 32x32	LBP	HOG
Male-Specific Ethnicity classifier (316C,172NC)	83.8%	82.4%	83.6%	83.0%	87.0%
Generic Ethnicity classifiers (316C,172NC)	[81.5, 87.5]%	[82.2, 86.1]%	[80.9, 86.5]%	[87.5, 90.2]%	[85.2, 89.9]%
Female-Specific Ethnicity classifier (234C,127NC)	82.0%	80.6%	80.9%	81.2%	79.5%
Generic Ethnicity classifiers (234C,127NC)	[78.9, 86.7]%	[77.8, 86.7]%	[79.5, 86.9]%	[81.2, 89.7]%	[81.7, 88.3]%

gender disambiguation regardless the ethnicity, some other are important for ethnicity perception regardless the gender and some features are discriminative for both gender and ethnicity. Based to the results achieved, we conclude that the machines, to classify the gender, use characteristics that are shared by all the ethnicity and, vice versa, to classify the ethnicity, use discriminative characteristics present in both male and female.

VII. CONCLUSION

This paper has provided an analysis on gender and ethnicity classification task and their interaction. Ethnicity-specific gender classifiers have been trained and tested using faces from a specific ethnicity; the accuracies achieved have been compared with the ones obtained using generic gender classifiers, trained and tested with faces from different ethnicities. With a similar procedure, gender-specific ethnicity classifiers have been trained and tested selecting faces with a specific gender; the accuracies achieved have been compared with the ones obtained using generic ethnicity classifiers, trained and tested with both male and female faces. Three widely used algorithms have been implemented to conduct the experiments: Pixel-based with three different resolutions, HOG and LBP. We have proven that gender and ethnicity do not affect each other during the classification; it means that, using these algorithms, gender and ethnicity classification tasks can be solved separately.

As a future work it would be interesting to extend this study using geometrical features in both 2D and 3D images to verify if gender and ethnicity are still uncorrelated when involving the shape. A natural extension of this work is, also, represented by the study on interdependences with age.

REFERENCES

[1] M. Demirkus, K. Garg, and S. Guler, "Automated person categorization for video surveillance using soft biometrics," in *Biometric Technology for Human Identification VII*, 2010.

[2] A. Dantcheva, C. Velardo, A. D'angelo, and J.-L. Dugelay, "Bag of soft biometrics for person identification : New trends and challenges," *Mutimedia Tools and Applications, Springer*, October 2010.

[3] A. F. Smeaton, P. Over, and W. Kraaij, "Evaluation campaigns and trecvid," in *MIR '06: Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*. New York, NY, USA: ACM Press, 2006, pp. 321–330.

[4] B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski, "Sexnet: A neural network identifies sex from human faces," in *conference on Advances in neural information processing systems 3*, ser. NIPS-3, San Francisco, CA, USA, 1990, pp. 572–577.

[5] B. Moghaddam and M.-H. Yang, "Learning gender with support faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 707 –711, May 2002.

[6] J. Ylioinas, A. Hadid, and M. Pietikinen, "Combining contrast information and local binary patterns for gender classification," in *Image Analysis*, ser. Lecture Notes in Computer Science, A. Heyden and F. Kahl, Eds., vol. 6688. Springer Berlin / Heidelberg, 2011, pp. 676–686.

[7] E. Mäkinen and R. Raisamo, "An experimental comparison of gender classification methods," *Pattern Recogn. Lett.*, vol. 29, pp. 1544–1556, July 2008.

[8] S. Baluja and H. A. Rowley, "Boosting sex identification performance," in *National Conference on Artificial Intelligence*, 2005, pp. 1508–1513.

[9] Z. Sun, G. Bebis, X. Yuan, and S. J. Louis, "Genetic feature subset selection for gender classification: A comparison study," in *IEEE Workshop on Applications of Computer Vision*, 2002, pp. 165–170.

[10] P. Brunelli, "Hyberbf networks for gender classification," 1995.

[11] E. Makinen and R. Raisamo, "Evaluation of gender classification methods with automatically detected and aligned faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 541–547, March 2008.

[12] Z. Yang and H. Ai, "Demographic classification with local binary patterns," in *International Conference on Biometrics*, 2007, pp. 464–473.

[13] X. Lu, , and A. K. Jain, "Ethnicity identification from face images," in *Proceedings of SPIE International Symposium on Defense and Security : Biometric Technology for Human Identification*, 2004, pp. 114–123.

[14] J. Bekios-Calfa, J. Buenaposada, and L. Baumela, "Revisiting linear discriminant techniques in gender recognition," *Pattern Analysis and*

Fig. 6. Gender perception experiment. Try to assign a sex to the subjects shown in Figure 3; here the entire images are displayed.

Machine Intelligence, IEEE Transactions on, vol. 33, no. 4, pp. 858 –864, April 2011.
[15] W. Gao and H. Ai, “Face gender classification on consumer images in a multiethnic environment,” in *Advances in Biometrics*, ser. Lecture Notes in Computer Science, M. Tistarelli and M. Nixon, Eds. Springer Berlin / Heidelberg, 2009, vol. 5558, pp. 169–178.
[16] H. Abdi, D. Valentin, B. Edelman, and A. J. O. Toole, “More about the difference between men and women: Evidence from linear neural networks and the principal-component approach,” 1995.
[17] S. Gutta, H. Wechsler, and P. Phillips, “Gender and ethnic classification of face images,” in *Third IEEE International Conference on Automatic Face and Gesture Recognition.*, April 1998, pp. 194 –199.
[18] H. cheng Lian and B. liang Lu, “Multi-view gender classification using local binary patterns and support vector machines,” in *in Proc. International Symposium on Neural Networks*, 2006, pp. 202–209.
[19] X. Lu, H. Chen, and A. K. Jain, “Multimodal facial gender and ethnicity identification,” in *ICB*, 2006, pp. 554–561.
[20] L. A. Alexandre, “Gender recognition: A multiscale decision fusion approach,” *Pattern Recogn. Lett.*, vol. 31, pp. 1422–1427, August 2010.
[21] W.-S. Chu, C.-R. Huang, and C.-S. Chen, “Identifying gender from unaligned facial images by set classification,” in *20th International Conference on Pattern Recognition (ICPR)*, August 2010, pp. 2636 – 2639.
[22] “Illusion of sex,” http://illusioncontest.neuralcorrelate.com/2009/the-illusion-of-sex/, 2009.
[23] R. Russell, “A sex difference in facial contrast and its exaggeration by cosmetics,” *Perception*, vol. 38, 2009.
[24] A. R. Templeton, “Human races: A genetic and evolutionary perspective.” *American Anthropologist*, vol. 100, pp. 632–650, 1998.
[25] S. O. Y. Keita, R. A. Kittles, C. D. M. Royal, G. E. Bonney, P. Furbert-Harris, G. M. Dunston, and C. N. Rotimi, “Conceptualizing human variation.” *Nature Genetics*, vol. 36, pp. 17–20, 2004. [Online]. Available: http://www.biomedsearch.com/nih/Conceptualizing-human-variation/15507998.html
[26] N. Furl, P. Phillips, and A. J. OToole, “Face recognition algorithms and the other-race effect: computational mechanisms for a developmental contact hypothesis,” *Cognitive Science*, vol. 26, no. 6, pp. 797 – 815, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0364021302000848
[27] “The world of averages,” http://pmsol3.wordpress.com, 2011.
[28] T. Ojala, M. P. inen, and D. Harwood, “A comparative study of texture measures with classification based on featured distributions,” *Pattern Recognition*, vol. 29, pp. 51–59, 1996.
[29] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1 - Volume 01*, Washington, DC, USA, 2005, pp. 886–893.
[30] G. Guo, C. Dyer, Y. Fu, and T. Huang, “Is gender recognition affected by age?” in *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, October 2009, pp. 2032 –2039.
[31] P. Phillips, H. Wechsler, J. Huang, and P. Rauss, “The feret database and evaluation procedure for face recognition algorithms,” *Image and Vision Computing J*, vol. 16, pp. 295–306, 1998.
[32] P. Phillips, H. Moon, S. Rizvi, and P. Rauss, “The feret evaluation methodology for face recognition algorithms,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1090–1104, 2000.
[33] C.-C. Chang and C.-J. Lin, “LIBSVM: A library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.