Gender and Age Estimation Using Face Images

Fatma Guney
Bogazici University
Computer Engineering
Istanbul, Turkey 34342
Email: guneyftm@gmail.com

***Abstract*—In this study, I present a real-time gender and age estimation approach that I developed for the course project of Artificial Neural Networks. For these tasks, I employed a local appearance-based representation using Discrete Cosine Transform (DCT), which has been shown to be very effective in real-time processing and robust against changes caused by light and facial expressions. Using these features, I trained support vector machine classifiers. Binary support vector machine classifier is used to discriminate between males and females. In case of age estimation, a two-step classifier is used. First, a support vector machine classifier is used to discriminate between youths and adults. Then, three separate support vector machine regression functions are used for youths, adults and for those that are close to the youth-adult decision border.**

I. Introduction

Automatic gender and age estimation from face data are two important research subjects in the field of facial image analysis. There are many applications that could benefit from the advances of these research. First of all, age and gender estimation could be employed in a human-computer interaction (HCI) device, so that system can be configured according to the knowledge extracted from the user. Automatic marketing which utilizes age and gender information of shopping customers is one of the many related commercial areas. Surveillance cameras that are used to control age limitations in traffic or in restricted products such as alcohol, cigarette also benefit from the advances in the subject. Other research areas such as face verification and recognition can also be improved with the extra knowledge on gender and age of the subject.

The objective of this study is to get an overview of the current methods and to develop a real-time application based on proposed solutions. The approach of [1] and [2] have been benefited during the study.

There are many studies on both gender and age estimation in the literature [3]. Studies can be classified according to the features and classifiers they employ. In general, there are two types of features; geometric features and appearance-based features. Geometric features captures the shape information related to the parts of face including mouth, eyes, brows while appearance-based features are based on the texture of the face. Feature extraction schemes can be global, that is can be applied to entire face or local, that is can be performed on the overlapping or non-overlapping blocks of face [5]. Local binary patterns (LBP), gabor wavelets and discrete cosine transform (DCT) are appearance-based features frequently employed in gender and age estimation systems. On the other hand, active appearance models (AAM) are frequently used in especially age estimation studies since they provide shape and texture information at the same time [1]. Principle component analysis (PCA), linear discriminant analysis (LDA) and Adaboost are used in some applications to perform feature selection before classfication. In many of the recent approaches, support vector machines (SVM) are used to model estimation problems as both classification and regression [4].

In this study, similar approaches are developed for gender and age estimation which could be exploited to develop a more general system that can perform both tasks. First, face and eye detection based on modified census transform (MCT) are performed on the input image. After detection, alignment based on eye coordinates of the detected face image is applied to scale and translate face and reduce s in feature space. Aligned face image is divided into local blocks and discrete cosine transform is performed on these local blocks. Concatenating features of each block, an overall feature vector is obtained. In gender estimation, SVM classifier is used for binary classification between female and male. In age estimation, a two step classifier is used. In the first step, SVM classifier is used to discriminate between youth and adult and in the second step, support vector regression (SVR) is used for youth, adult and global age estimation to determine the specific age.

II. Methodology

A. Face Detection and Alignment

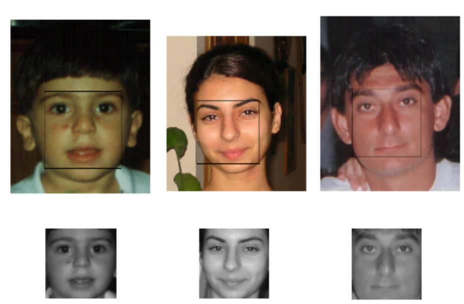

Fig. 1. Face detection and alignment examples

The face and eyes are automatically detected using a modified census transformation (MCT) based face and eye detector [10]. Alignment of detected face is necessary to decrease the variation in feature space caused by pose, angle and scale changes. Alignment is based on transformation of face in euclidean space by using the detected eye coordinates. Face is cut out and scaled according to the fixed distance between two eyes and eyes are at the same fixed position for all images. Some examples of detection and alignment are shown in Fig.1

B. Local Appearance-based Face Representation

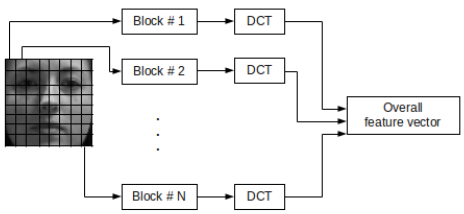

Fig. 2. Local appearance-based feature extraction scheme using DCT.

For face representation a local appearance-based approach is used. In local appearance-based approach, face is divided into non-overlapping blocks and feature extraction is performed on these blocks instead of whole face as can be seen in Fig.2. When there is a change in appearance of face due to the facial expression or occlusion, using local blocks provides advantage because only the related region of block or blocks are affected. In case of applying feature extraction on entire face, entire representation is affected by changes [6].

For feature extraction on local regions, discrete cosine transform (DCT) is used. DCT is a signal analysis tool which is frequently used in facial image analysis because it provides frequency information in a compact representation. DCT is also frequently preferred in real-time applications due to its fast computation. DCT representation is shown to be robust to lighting changes and scaling variations due to its decomposition capability, that is elements sensitive to lighting changes and scaling variations can be removed. For example, first coefficient represent the average intensity value of the face image which can be directly effected by illumination variations ł [5].

A detected and aligned face image is divided into blocks of 8x8 pixels size. Each block is then represented by its DCT coefficients. 8x8 is chosen as size in face recognition applications, because it is small enough to provide stationary within the block with a simple transform complexity and big enough to provide sufficient compression [7].

After selecting coefficients a two-step normalization is applied to features of each block. Firstly, blocks with different brightness levels may have DCT coefficients with different value levels. Local feature vector's magnitude is normalized to unit norm. Secondly, first coefficients have higher magnitudes, therefore each coefficient is divided by its standard deviation to balance the contribution of coefficients [6].

Top-left DCT coefficient which is the first one, is removed from the representation since it only represents the average intensity value of the block. From the remaining DCT coefficients, five coefficients which contain the highest information are extracted using zig-zag scanning as shown in Fig.3. Finally, the DCT coefficients extracted from each block are concatenated to construct the overall feature vector [6], [5], [7].

0	1	5	6	14	15	27	28
2	4	7	13	16	26	29	42
3	8	12	17	25	30	41	43
9	11	18	24	31	40	44	53
10	19	23	32	39	45	52	54
20	22	33	38	46	51	55	60
21	34	37	47	50	56	59	61
35	36	48	49	57	58	62	63

Fig. 3. Zig-zag scanning order used to select the DCT coefficients.

C. Gender Estimation

In this part, the goal is to determine the gender of the faces detected in the images. Gender estimation is a binary classification problem with two classes, male and female.

Scaling before applying SVM is very important to avoid features with huge ranges from dominating others. The biggest and smallest magnitude values are determined for each feature index and saved during training. Then, in both training and testing, each feature vector is scaled linearly into the range $[1, +1]$ as recommended in [9].

In gender estimation SVM with a radial basis function (RBF) kernel is used. RBF kernel transforms the features into a higher dimensional space, so that they can be then linearly separated by a hyperplane.

For the optimization of parameters slack variable C and kernel parameter γ, five-fold cross validation is used. Grid-search on C and γ using cross-validation is performed as recommended in [9]. In grid search, pairs of (C, γ) are tried and the one with the best cross-validation accuracy is picked.

D. Age Estimation

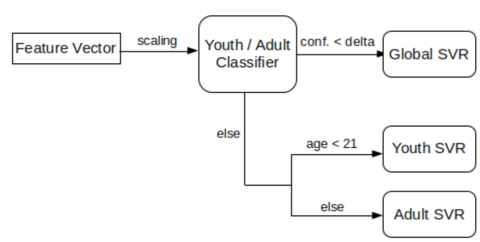

Fig. 4. Age estimation approach using two-step classification.

In this study, a two-step classifier is used for age estimation as as shown in Fig.4 [2]. First, support vector machine classifier is used to discriminate between youths and adults. Youth is defined as person whose age under 21 and adult is defined as age 21 and above. According to research on aging of human face, 20 is an approximate turning point. Shape of human face changes until 20, until then changes generally occur in the texture of the face. Therefore, training two different support vector regression functions for youth and adult faces provides a better way of modelling age estimation problem.

If an adult face is classified as youth, wrong regression function is used to estimate the age. To prevent the problems caused by this situation, a soft decision is made by the first classifier. Distance to hyper-plane can be used as a confidence value for the first classifier. By putting a threshold on this distance, a third, global support vector regressor is used in addition to youth and adult regression functions.

Scaling procedure and kernel function selection is the same as the gender estimation. Approach on parameter optimization is also same with additional two parameters, one is ϵ in loss function of epsilon-SVR and the other is δ which is the threshold on the distance to hyper-plane [2].

III. Experiments and Results

A. Experimental Setup

Face Recognition Grand Challenge (FRGC) dataset is used for the experiments in gender estimation. It contains face images of 345 subjects, 153 female and 192 male. In total there are 2113 images, but in 2095 of them a face is detected [**?**].

The FG-NET aging database is used for the experiments in age estimation. It contains face images of varying ages between 0 and 69 for each 82 subjects. In total there are 1002 images [8].

For both gender and age estimation, each detected face is scaled to 64x64 pixels and aligned so that eye row is 18 and distance between eyes is 32 pixels [2], [10]. DCT is performed on blocks of 8x8 pixels. For each block the first 5 coefficients in the zig-zag scanning order are kept, leading to a 8x8x5 = 320 dimensional feature vector. After scaling the overall feature vector, SVM parameter optimization is performed by grid-search using five-fold cross validation. When optimizing the SVM parameters via cross validation on the training set, the subjects instead of the single images are chosen to prevent classifiers from learning subject related information rather than age and gender related ones.

For the evaluation of gender estimation, five-fold cross validation is used on FRGC dataset. To prevent the algorithm from learning the identity of the subjects in the training set rather than gender, all samples of a single subject are required to be only in one fold at a time, that is samples of a single subject are not in the testing and training set at the same time

For the evaluation of age estimation, leave-one-person-out (LOPO) evaluation scheme is used, since this is a common evaluation scheme for FG-NET, due to its small size [2]. All images of a single subject are used for testing and the remaining samples for training. This is done for all subjects so that each person is once used for testing, resulting in 82 folds.

B. Evaluation Criteria and Results

Different evaluation criteria are used for gender and age estimation. For gender estimation, confusion matrix is filled in each fold of cross validation. From this confusion matrix, some metrics are calculated to evaluate the classifier. First one is the accuracy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

where TP is the number of correctly classified positive samples, FP the number of samples that have been classified incorrectly as positives, TN the number of correctly classified negative samples, and FN the number of samples that have been classified incorrectly as negatives.

This metric specifies the overall accuracy of the classification, but a classifier that has unbalanced accuracies between male and female images can lead to a higher accuracy than a classifier that has more balanced accuracies between genders [11]. In order to overcome this issue, using true positive rate (TPR) and true negative rate (TNR) besides accuracy is recommended. There is also average correct rate (ACR) which is a more robust metric than accuracy in case of imbalances in the dataset [12].

$$TPR = \frac{TP}{TP+FN} \tag{2}$$

$$TNR = \frac{TN}{TN+FP} \tag{3}$$

$$ACR = \frac{TPR+TNR}{2}; \tag{4}$$

TABLE I
Results of gender estimation

	Accuracy	TPR	TNR	ACR
Fold-1	0.965844	0.954733	0.975352	0.965042
Fold-2	0.940945	0.99177	0.89434	0.943055
Fold-3	0.967662	0.950549	0.981818	0.966184
Fold-4	0.939698	0.910714	0.977011	0.943863
Fold-5	0.96732	0.97619	0.959839	0.968015
Average	0.957232	0.956791	0.957672	0.957232

For age estimation, one of the metrics frequently used in evaluation is mean absolute error (MAE) [1]:

$$MAE = \frac{\sum_{i=0}^{n} |\hat{a}_i - a_i|}{n} \tag{5}$$

where a_i is the real age of subject i, $\hat{a}_i$ is predicted age and n is total number of subjects. MAE value of 5.66612 is reached using LOPO in the experiments conducted on FG-NET dataset. In order to evaluate performance over the different age ranges MAEs per decade (MAE/D) metric is defined [1]. MAE/D is MAE calculated for each decade separately.

Another measure is cumulative score (CS) [1]:

$$CS(d) = \frac{N(|\hat{a}_i - a_i| \leq d)}{n} x100 \tag{6}$$

where $N(|\hat{a}_i - a_i| \leq d)$ is number of estimations with an estimation error less than or equal to d. Cumulative scores

TABLE II
RESULTS OF AGE ESTIMATION

Youth/Adult SVC error (%)	22.156
Youth-SVR MAE	2.85706
Adult-SVR MAE	7.8054

TABLE III
MAE/D FOR AGE ESTIMATION

Age range	Image count	MAE
0-9	354	3.26948
10-19	327	3.67529
20-29	138	5.55059
30-39	79	10.5041
40-49	44	19.4343
50-59	15	28.2014
60-69	7	34.3809

obtained on age differences from zero to ten are shown in Fig.5.

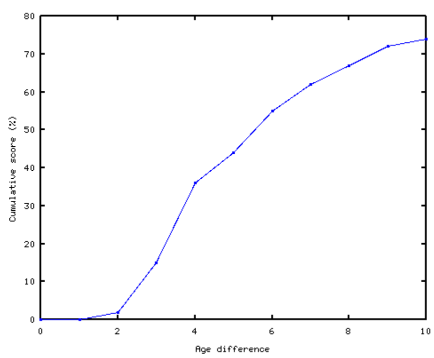

Fig. 5. Cumulative scores obtained from LOPO on FG-NET dataset.

IV. CONCLUSION

A real-time gender and age estimation system using face data is proposed and developed. Given an input image or video sequence, first faces are detected and aligned for the feature extraction. Local appearance-based feature extraction using DCT is applied to each aligned face. Features of each block are concatenated to obtain the overall feature vector and each feature is scaled before applying SVM. For gender estimation binary SVM classifier is trained and used. For age estimation, two-step classifier is used to estimate ages of different age groups. First, feature vector is classified as youth or adult by using binary SVM and then according to confidence and result of this classification, one of three separate SVR functions is used to estimate age. For the evaluation of gender estimation five-fold cross validation is used and for age estimation LOPO method is used. Obtained results show that SVM classification and regression using local appearance-based features give very promising results for gender and age estimation in a common framework.

REFERENCES

[1] Luu, K., Ricanek, K., Bui, T.D., Suen, C.Y.: Age estimation using active appearance models and support vector machine regression. 3r d IEEE Intl. Conf. on Biometrics pp. 314318 (2009)
[2] M. Steiner, ”Facial Image-based Age Estimation”. Study Thesis at the Faculty of Computer Science Institute for Anthropomatics Facial Image Processing and Analysis. (2010)
[3] Fu, Y., G.D.Guo, T.S.Huang: Age synthesis and estimation via faces: A survey. IEEE Trans. on Pattern Analysis and Machine Intelligence 32, 19551976 (2010)
[4] K. Luu, T. Bui, C. Suen, and K. Ricanek, Spectral Regression based age determination, in IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW 2010), no. 1993. IEEE, 2010, pp. 103107.
[5] Ekenel, H.K., Stiefelhagen, R.: Why is facial occlusion a challenging problem? In: Proc. of Intl. Conf. on Biometrics (2009)
[6] Ekenel, H.K., Stiefelhagen, R.: Automatic frequency band selection for illumination robust face recognition. In: Proc. of Intl. Conf. on Pattern Recognition (2010)
[7] Stallkamp, J., Ekenel, H.K., Stiefelhagen, R.: Video-based face recognition on real-world data. In: Proc. of IEEE Intl. Conf. on Computer Vision. pp. 18 (2007)
[8] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, Overview of the face recognition grand challenge, Proc. of CVPR05, no. 1, pp. 947954, June 2005
[9] The FG-NET aging database (2010), http://www.fgnet.rsunit.com/
[10] Chih-Chung Chang, C.J.L.: Libsvm: a library for support vector machines (2001), http://www.csie.ntu.edu.tw/ cjlin/libsvm
[11] CVHCI lab. Okapi (open karlsruhe library for processing images) c++ library.
[12] R. Martin-Felez, R. A. Mollineda, and J. S. Sanchez, A Gender Recognition Experiment on the CASIA Gait Database Dealing with its Imbalanced Nature, in International Conference on Computer Vision Theory and Applications (VISAPP 2010), 2010, pp. 439444.
[13] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan, Toward practical smile detection. IEEE transactions on pattern analysis and machine intelligence, vol. 31, no. 11, pp. 210611, Nov. 2009.